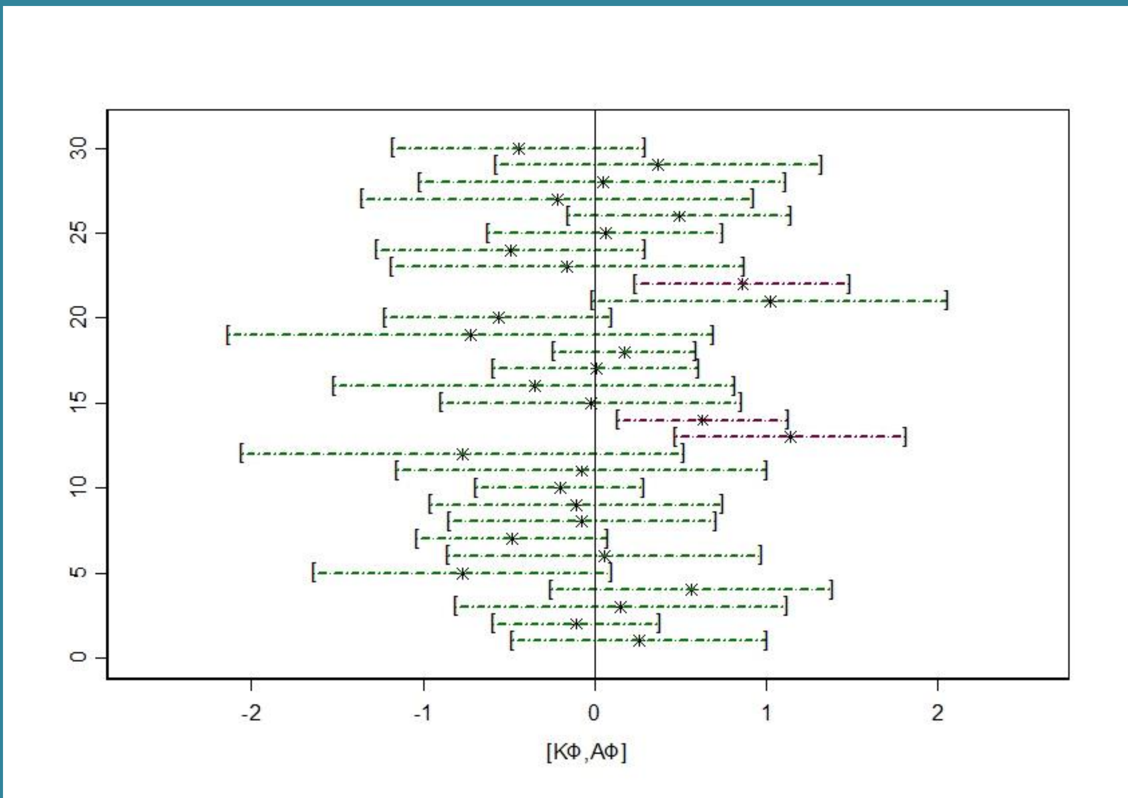
Θέματα Παραμετρικής Στατιστικής Συμπερασματολογίας:

Εκτιμητική και Διαστήματα Εμπιστοσύνης



Σταύρος Κουρούκλης
Κωνσταντίνος Πετρόπουλος, Βιολέττα Πιπερίγκου

Ελληνικά Ακαδημαϊκά Ηλεκτρονικά Συγγράμματα και Βοηθήματα
www.kallipos.gr
HEALLINK
Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών
Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Θέματα
Παραμετρικής Στατιστικής Συμπερασματολογίας:
Εκτιμητική και Διαστήματα Εμπιστοσύνης

Σταύρος Κουρούκλης
Καθηγητής, Τμήμα Μαθηματικών, Πανεπιστήμιο Πατρών

Κωνσταντίνος Πετρόπουλος
Επίκουρος Καθηγητής, Τμήμα Μαθηματικών, Πανεπιστήμιο Πατρών

Βιολέττα Πιπερίγκου
Επίκουρος Καθηγήτρια, Τμήμα Μαθηματικών, Πανεπιστήμιο Πατρών

Ελληνικά Ακαδημαϊκά Ηλεκτρονικά
Συγγράμματα και Βοηθήματα
www.kallipos.gr

Θέματα Παραμετρικής Στατιστικής Συμπερασματολογίας:

Εκτιμητική και Διαστήματα Εμπιστοσύνης

Συγγραφή

Σταύρος Κουρούκλης

Κωνσταντίνος Πετρόπουλος, Βιολέττα Πιπερίγκου

Κριτικός αναγνώστης

Απόστολος Μπατσίδης

Συντελεστές έκδοσης

ΓΛΩΣΣΙΚΗ ΕΠΙΜΕΛΕΙΑ: Ουρανία Γυφτοπούλου

ΣΥΝΔΕΣΜΟΣ ΕΛΛΗΝΙΚΩΝ ΑΚΑΔΗΜΑΪΚΩΝ ΒΙΒΛΙΟΘΗΚΩΝ

Εθνικό Μετσόβιο Πολυτεχνείο

Ηρώων Πολυτεχνείου 9, 15780 Ζωγράφου

www.kallipos.gr

ISBN: 978-960-603-435-0

Αντί προλόγου

Το ηλεκτρονικό βοήθημα «*Θέματα Παραμετρικής Στατιστικής Συμπερασματολογίας: Εκτιμητική και Διαστήματα Εμπιστοσύνης*» αποτυπώνει σε μεγάλο βαθμό την ύλη των παραδόσεων του μαθήματος κορμού «*Στατιστική Συμπερασματολογία Ι*» που διδάσκεται στο Ε΄ εξάμηνο του Προγράμματος Σπουδών του Τμήματος Μαθηματικών του Πανεπιστημίου Πατρών (πέντε ώρες εβδομαδιαία).

Έμφαση έχει δοθεί στην αυστηρή και μαθηματική θεμελίωση δύο βασικών κλάδων της Στατιστικής Συμπερασματολογίας, Εκτιμητικής και Διαστημάτων Εμπιστοσύνης, συγχρόνως όμως και στη διαισθητική ερμηνεία των διαφόρων εννοιών και αποτελεσμάτων. Θεωρούμε ότι μέσα από μία τέτοια μαθηματική θεμελίωση, μπορεί κάποιος να αναγνωρίσει και να εκτιμήσει ότι ο κύριος ρόλος της Στατιστικής στην κοινωνία είναι να δίνει μαθηματικά ακριβείς απαντήσεις σε σύνθετα προβλήματα της ανθρώπινης δραστηριότητας και όχι απλά και μόνον να διαχειρίζεται, περιγραφικά, σύνολα δεδομένων.

Υπάρχουν αρκετά συγγράμματα Στατιστικής Συμπερασματολογίας στην ελληνική και ξένη (κυρίως αγγλική) βιβλιογραφία, κάποια εκ των οποίων δίνονται στη βιβλιογραφία του Κεφαλαίου 2. Το ηλεκτρονικό βοήθημα «*Θέματα Παραμετρικής Στατιστικής Συμπερασματολογίας: Εκτιμητική και Διαστήματα Εμπιστοσύνης*» δεν διεκδικεί θέση στη λίστα των συγγραμμάτων: σκοπό έχει να φανεί χρήσιμο στους φοιτητές, στην πρώτη τους επαφή με τη (Θεωρητική-Μαθηματική) Στατιστική. Έτσι, για να εξυπηρετηθεί αυτός ο σκοπός, περιέχονται και αρκετά λυμένα παραδείγματα.

Μεγάλο μέρος του υλικού που παρουσιάζεται πρωτογράφτηκε σε χειρόγραφη μορφή από τον Σ. Κουρούκλη πριν αρκετά χρόνια, όταν ο Κ. Πετρόπουλος και η Β. Πιπερίγκου ήταν ακόμη φοιτητές, και παρέμεινε έτσι χωρίς να υποστεί σημαντικές επιστημονικές αλλαγές. Τη γραφή του σε LaTeX έκανε για πρώτη φορά ο συνάδελφος κ. Παναγιώτης Μπομποτάς (Department of Mathematical Sciences, University of Southampton), τότε μεταπτυχιακός φοιτητής του Τμήματος Μαθηματικών του Πανεπιστημίου Πατρών, προς τον οποίον επιθυμούμε να εκφράσουμε, και από τη θέση αυτή, τις θερμές ευχαριστίες μας για τον χρόνο που διέθεσε, τη διάθεση που επέδειξε και το αποτέλεσμα που παρήγαγε. Το υλικό αυτό έλαβε την τελική του μορφή ως «*Θέματα Παραμετρικής Στατιστικής Συμπερασματολογίας: Εκτιμητική και Διαστήματα Εμπιστοσύνης*» πρόσφατα στα πλαίσια της δράσης «Κάλλιπος» για τη συγγραφή Ελληνικών Ακαδημαϊκών ηλεκτρονικών συγγραμμάτων και βοηθημάτων.

Θέλουμε επίσης να ευχαριστήσουμε τον συνάδελφο κ. Απόστολο Μπατσίδη (Τμήμα Μαθηματικών, Παναπιστήμιο Ιωαννίνων) κριτικό αναγνώστη αυτού του ηλεκτρονικού βοηθήματος, για την ενδελεχή του μελέτη, καθώς και για τις εύστοχες παρατηρήσεις του. Προϊόν αυτών είναι, μεταξύ άλλων, η προσθήκη των αποδείξεων για τις ασυμπτωτικές ιδιότητες των εκτιμητών μέγιστης πιθανοφάνειας, η παρουσίαση εκτιμητών Bayes ως προς το απόλυτο σφάλμα και τρεις ποσότητες οδηγοί στην Πρόταση 9.2.1.

Ίσως το υλικό που παρουσιάζεται εδώ αποτελέσει, κάποτε, μέρος της κρίσιμης μάζας για τη συγγραφή ενός βιβλίου. Ίδωμεν. ...

Πάτρα, Δεκέμβριος 2015

Περιεχόμενα

Αντί προλόγου **i**

1 Στοιχεία Θεωρίας Κατανομών **1**

1.1 Ορισμός πιθανότητας 1
1.2 Δεσμευμένη πιθανότητα και ανεξαρτησία 5
1.3 Τυχαίες μεταβλητές και κατανομές 7
1.4 Μέση τιμή, διασπορά, ροπές, ροπογεννήτρια 10
1.5 Ανισότητες Markov, Tchebychev, Jensen 12
1.6 Μερικές βασικές κατανομές 14
1.6.1 Μερικές διακριτές κατανομές 14
1.6.2 Μερικές συνεχείς κατανομές 15
1.7 Κατανομή συνάρτησης τυχαίας μεταβλητής 18
1.8 Τυχαία διανύσματα 19
1.9 Δεσμευμένες κατανομές 27
1.10 Οριακά Θεωρήματα 31
Βιβλιογραφία 38

2 Γενικά περί Στατιστικής Συμπερασματολογίας **41**

Βιβλιογραφία 48

3 Γενικά περί Εκτιμητικής **51**

3.1 Περιγραφή του προβλήματος 51
3.2 Κριτήρια επιλογής εκτιμητών 53
3.3 Απλές μέθοδοι εκτίμησης 57

4 Μέσο Τετραγωνικό Σφάλμα και Αμεροληψία **63**
4.1 Μέσο τετραγωνικό σφάλμα 63
4.2 Αμεροληψία 78
4.3 Η μη αποδεκτικότητα της δειγματικής διασποράς με κριτήριο το ΜΤΣ 95
4.4 Ασκήσεις 100
Βιβλιογραφία 101

5 Ανισότητα των Cramér–Rao, Πληροφορία του Fisher και Αποδοτικοί Εκτιμητές **103**
5.1 Το κάτω φράγμα των Cramér–Rao και ο αριθμός πληροφορίας του Fisher 104
5.2 Εκθετική οικογένεια κατανομών και αποδοτικοί εκτιμητές . 117
5.3 Ανισότητα των Cramér–Rao και πίνακας πληροφορίας του Fisher για $\theta = (\theta_1, \ldots, \theta_\kappa) \in \mathbb{R}^\kappa$ 139
5.4 Ασκήσεις 145
Βιβλιογραφία 147

6 Επάρκεια, πληρότητα και ΑΟΕΔ εκτιμητές **151**
6.1 Επάρκεια 152
6.2 Χρήση της επάρκειας στη βελτίωση εκτιμητών-μείωση της διασποράς και του ΜΤΣ 169
6.3 Πληρότητα και ΑΟΕΔ εκτιμητές 172
6.4 Βέλτιστοι εκτιμητές ειδικής μορφής 207
6.5 Αποτίμηση της αμεροληψίας 213
6.6 Ασκήσεις 215
Βιβλιογραφία 219

7 Μέθοδος της μέγιστης πιθανοφάνειας και μέθοδος των ροπών **221**
7.1 Μέθοδος της μέγιστης πιθανοφάνειας 221
7.2 Ιδιότητες των εκτιμητών μέγιστης πιθανοφάνειας 240
7.3 Αποτίμηση των εκτιμητών μέγιστης πιθανοφάνειας 273

7.4 Μέθοδος των ροπών 274
7.5 Ασκήσεις 280
Βιβλιογραφία 283

8 Εκτιμητές Bayes και εκτιμητές minimax 287
8.1 Εκτιμητές Bayes 288
8.2 Εκτιμητές minimax 304
8.3 Ασκήσεις 307
Βιβλιογραφία 308

9 Διαστήματα εμπιστοσύνης 311
9.1 Ορισμός και ορολογία 312
9.2 Κατασκευή διαστημάτων εμπιστοσύνης μέσω ποσότητας οδηγού 318
9.3 Κατασκευή διαστημάτων εμπιστοσύνης μέσω της συνάρτησης κατανομής στατιστικής συνάρτησης 327
9.4 Εφαρμογές σε κανονικούς πληθυσμούς 334
9.4.1 Ένας κανονικός πληθυσμός 335
Διάστημα εμπιστοσύνης για τη μέση τιμή 335
Διάστημα εμπιστοσύνης για τη διασπορά 343
9.4.2 Δύο κανονικοί πληθυσμοί 349
Διάστημα εμπιστοσύνης για τη διαφορά των μέσων τιμών 349
Διάστημα εμπιστοσύνης για το λόγο των διασπορών . 353
9.5 Ασυμπτωτικά διαστήματα εμπιστοσύνης 356
9.5.1 Ασυμπτωτικό διάστημα εμπιστοσύνης για τη μέση τιμή 356
9.5.2 Ασυμπτωτικό διάστημα εμπιστοσύνης για τη διαφορά δύο μέσων τιμών 357
9.5.3 Ασυμπτωτικό διάστημα εμπιστοσύνης για ποσοστό . 358
9.5.4 Ασυμπτωτικό διάστημα εμπιστοσύνης για τη διαφορά δύο ποσοστών 361
9.6 Ασυμπτωτικά διαστήματα εμπιστοσύνης και εκτιμητές μέγιστης πιθανοφάνειας 363
9.7 Μπεϋζιανά Διαστήματα 367

9.8 Ασκήσεις 371
Βιβλιογραφία 375

Αγγλική Ορολογία **377**

Ευρετήριο **381**

Πίνακας Συντμήσεων

Α.Δ.Ε.	Ασυμπτωτικό Διάστημα Εμπιστοσύνης
ΑΟΕΔ	Αμερόληπτος Ομοιομόρφως Ελάχιστης Διασποράς
ΑΝΜΑ	Ασθενής Νόμος των Μεγάλων Αριθμών
Δ.Ε.	Διάστημα Εμπιστοσύνης
ε.μ.π.	εκτιμητής μέγιστης πιθανοφάνειας
ε.μ.ρ.	εκτιμητής μεθόδου ροπών
ΙΝΜΑ	Ισχυρός Νόμος των Μεγάλων Αριθμών
Κ.Φ. C-R	Κάτω Φράγμα των Cramér - Rao
ΚΟΘ	Κεντρικό Οριακό Θεώρημα
ΜΤΣ	Μέσο Τετραγωνικό Σφάλμα
Μ.Ε.Ο.Κ.	Μονοπαραμετρική Εκθετική Οικογένεια Κατανομών
Σ.Σ.	Στατιστική Συμπερασματολογία
σ.ε.	συντελεστής εμπιστοσύνης
ΤΣ	Τυπικό Σφάλμα

Κεφάλαιο 1

Στοιχεία Θεωρίας Κατανομών

Το κεφάλαιο αυτό περιέχει, πολύ συνοπτικά, στοιχεία Θεωρίας Πιθανοτήτων και Θεωρίας Κατανομών, όπως η έννοια της πιθανότητας, τυχαίες μεταβλητές και κατανομές τους, μέση τιμή και διασπορά κατανομής, ειδικές διακριτές και συνεχείς τυχαίες μεταβλητές, από κοινού κατανομές, δεσμευμένες κατανομές, ανεξαρτησία, κατανομές μετασχηματισμών και οριακά θεωρήματα.

1.1 Ορισμός πιθανότητας

Τυχαίο (ή στοχαστικό) πείραμα ή πείραμα τύχης (random experiment) είναι ένα πείραμα με τα εξής δύο χαρακτηριστικά:

α. Το αποτέλεσμα του πειράματος δεν μπορεί να προβλεφθεί εκ των προτέρων (δηλαδή προτού εκτελεστεί το πείραμα).

β. Το σύνολο των αποτελεσμάτων του πειράματος είναι γνωστό εκ των προτέρων.

Το σύνολο αυτό, Ω, ονομάζεται *δειγματικός χώρος* του τυχαίου πειράματος. Κάθε ένα από τα στοιχεία του Ω (δηλαδή κάθε ένα από τα αποτελέσματα) αναφέρεται ως απλό ή στοιχειώδες ενδεχόμενο. *Ενδεχόμενα* του τυχαίου πειράματος είναι συγκεκριμένα υποσύνολα A του δειγματικού χώρου Ω. Το ενδεχόμενο A *πραγματοποιείται*, όταν και εάν το αποτέλεσμα του τυχαίου πειράματος είναι στοιχείο του A. Σε αυτήν την περίπτωση το αποτέλεσμα αναφέρεται ως *ευνοϊκό* για το A. Για παράδειγμα, στο τυχαίο πείραμα της ρίψης ενός ζαριού, ο δειγματικός χώρος είναι $\Omega = \{1, 2, 3, 4, 5, 6\}$ και το ενδεχόμενο $A = \{2, 4, 6\} = \{\text{άρτιος αριθμός}\}$ πραγματοποιείται, όταν και εάν το αποτέλεσμα της ρίψης είναι ή 2 ή 4 ή 6. Τα αποτελέσματα 2, 4, 6 είναι ευνοϊκά για το A. Το ενδεχόμενο A είναι το σύνολο των ευνοϊκών του αποτελεσμάτων.

Επειδή υπάρχει αβεβαιότητα ως προς το αποτέλεσμα του τυχαίου πειράματος (χαρακτηριστικό (α)), υπάρχει κατά συνέπεια και αβεβαιότητα ως προς την πραγματοποίηση ενός ενδεχομένου A. Αυτή η αβεβαιότητα μετράται μέσω της *πιθανότητας του* A, $\mathbb{P}(A)$.

Κλασικός ορισμός πιθανότητας (De Moivre 1711, Laplace 1812). Ας υποθέσουμε ότι ο δειγματικός χώρος Ω του τυχαίου πειράματος είναι πεπερασμένο σύνολο. Η πιθανότητα του ενδεχομένου $A \subset \Omega$ ορίζεται ως εξής:

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{\text{αριθμός των ευνοϊκών αποτελεσμάτων για το } A}{\text{αριθμός όλων των αποτελεσμάτων}}$$

($|A| =$ αριθμός των στοιχείων του A). Από τον ορισμό προκύπτουν αμέσως οι εξής ιδιότητες.

1. $\mathbb{P}(A) \geq 0$.
2. $\mathbb{P}(\Omega) = 1$.
3. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$, εάν $A \cap B = \emptyset$.

Ο κλασικός ορισμός παρουσιάζει δύο μειονεκτήματα:

α. απαιτεί ο δειγματικός χώρος να είναι *πεπερασμένο σύνολο*, συνεπώς δεν μπορεί να εφαρμοστεί σε τυχαία πειράματα με μη πεπερασμένο δειγματικό χώρο (π.χ. ρίψη νομίσματος μέχρι να εμφανιστεί η όψη «κεφαλή»).

β. απαιτεί όλα τα αποτελέσματα να είναι *ισοπίθανα*, γιατί, αν $\omega \in \Omega$, τότε $\mathbb{P}(\{\omega\}) = \frac{|\{\omega\}|}{|\Omega|} = \frac{1}{|\Omega|}$. Συνεπώς δεν μπορεί να εφαρμοστεί σε τυχαία πειράματα με ανισοπίθανα αποτελέσματα (π.χ. ρίψη κάλπικου νομίσματος).

Εμπειρικός ορισμός πιθανότητας (von Mises 1917). Ας υποθέσουμε ότι ένα τυχαίο πείραμα με δειγματικό χώρο Ω μπορεί να επαναληφθεί πολλές φορές υπό τις ίδιες ακριβώς συνθήκες. Για κάθε ενδεχόμενο A ας ορίσουμε ως $\nu(A)$ τον αριθμό των φορών που πραγματοποιείται το A κατά τις ν επαναλήψεις του τυχαίου πειράματος. Ο λόγος $\nu(A)/\nu$ λέγεται *σχετική συχνότητα* του A και η πιθανότητα του A ορίζεται ως το όριο της σχετικής συχνότητας για $\nu \to \infty$, δηλαδή

$$\mathbb{P}(A) = \lim_{\nu \to \infty} \frac{\nu(A)}{\nu} \; .$$

Από τον ορισμό προκύπτουν αμέσως οι εξής ιδιότητες.

1. $\mathbb{P}(A) \geq 0$.

2. $\mathbb{P}(\Omega) = 1$.

3. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$, εάν $A \cap B = \emptyset$.

Ο εμπειρικός ορισμός παρουσιάζει δύο μειονεκτήματα:

α. απαιτεί επανάληψη του τυχαίου πειράματος πολλές φορές (υπό τις ίδιες συνθήκες), το οποίο είναι συχνά δύσκολο ή και αδύνατο να γίνει.

β. απαιτεί τον υπολογισμό του ορίου της σχετικής συχνότητας,

Στην πράξη, ως εκτίμηση (προσέγγιση) της πιθανότητας του A λαμβάνεται η σχετική συχνότητα για ν «μεγάλο», δηλαδή $\mathbb{P}(A) \approx \frac{\nu(A)}{\nu}$.

Οι προηγούμενοι ορισμοί πιθανότητας παρουσιάζουν λοιπόν μειονεκτήματα, αλλά, αν και εντελώς διαφορετικοί, χαρακτηρίζονται από κοινές ιδιότητες. Επιπλέον οι ιδιότητες αυτές διαισθητικά φαίνονται πολύ λογικές. Ο επόμενος αξιωματικός ορισμός πιθανότητας δεν έχει τα μειονεκτήματα των προηγούμενων και τους ενοποιεί διατηρώντας τις ιδιότητες τους ως αξιώματα. Ο ορισμός αυτός προτάθηκε από τον Kolmogorov στις αρχές της δεκαετίας του 1930.

Αξιωματικός ορισμός πιθανότητας (Kolmogorov 1933). Ένας *χώρος πιθανότητας* είναι μία τριάδα $(\Omega, \mathscr{A}, \mathbb{P})$, όπου:

1. Ω είναι ένα μη κενό σύνολο, ο δειγματικός χώρος.

2. $\mathscr{A}$ είναι μία σ–άλγεβρα, δηλαδή μία οικογένεια υποσυνόλων του Ω που

   α. περιέχει το Ω

   β. $\forall\ A \in \mathscr{A} \Rightarrow A^c \in \mathscr{A}$

   γ. $\forall\ A_i \in \mathscr{A}, i = 1, 2, \ldots\ \Rightarrow \cup_{i=1}^{\infty} A_i \in \mathscr{A}$.

3. $\mathbb{P}$ είναι μία συνάρτηση, $\mathbb{P}\colon \mathscr{A} \longrightarrow \mathbb{R}$, που ονομάζεται *πιθανότητα* και ικανοποιεί τις εξής ιδιότητες (αξιώματα Kolmogorov)

   α. $\mathbb{P}(A) \geq 0, \forall\ A \in \mathscr{A}$.

   β. $\mathbb{P}(\Omega) = 1$.

   γ. $\mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i), \forall\ A_i \in \mathscr{A},\ i = 1, 2, \ldots$ με $A_i \cap A_j = \emptyset,\ i \neq j$.

Σημειώνουμε ότι με τον αξιωματικό ορισμό η πιθανότητα είναι μία συνάρτηση που δεν ορίζεται μέσω κάποιου τύπου, αλλά μέσω συγκεκριμένων *αξιωμάτων*.

Από τον ορισμό μπορεί να δειχθούν οι εξής ιδιότητες.

1. $\emptyset \in \mathscr{A},\ \mathbb{P}(\emptyset) = 0$.

2. $\mathbb{P}(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mathbb{P}(A_i),\ \forall\ A_i \in \mathscr{A},\ i = 1, \ldots, n$ με $A_i \cap A_j = \emptyset$, $i \neq j$.

3. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$, $\forall\, A \in \mathscr{A}$.

4. $0 \leq \mathbb{P}(A) \leq 1$, $\forall\, A \in \mathscr{A}$.

5. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$, $\forall\, A, B \in \mathscr{A}$.

6. $A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$, $\forall\, A, B \in \mathscr{A}$.

7. $\mathbb{P}(\cup_{i=1}^n A_i) \leq \sum_{i=1}^n \mathbb{P}(A_i)$, $\forall\, A_i \in \mathscr{A}$, $i = 1, \ldots, n$.

1.2 Δεσμευμένη πιθανότητα και ανεξαρτησία

Η δεσμευμένη (ή υπό συνθήκη) πιθανότητα του ενδεχομένου A εκφράζει πιθανότητα πραγματοποίησης του A λαμβάνοντας υπόψη κάποια μερική πληροφορία για το αποτέλεσμα του τυχαίου πειράματος. Έτσι, ενώ η πιθανότητα «άσσου» στη ρίψη αμερόληπτου ζαριού είναι $1/6$, η (δεσμευμένη) πιθανότητα «άσσου» δοθείσης της πληροφορίας ότι η ρίψη απέφερε περιττό αποτέλεσμα είναι $1/3$. Η (μερική) πληροφορία για το αποτέλεσμα του τυχαίου πειράματος εκφράζεται από κάποιο άλλο ενδεχόμενο E, στο παράδειγμα, $E = \{1, 3, 5\}$. Το E είναι ο «νέος» δειγματικός χώρος του τυχαίου πειράματος.

Γενικά, εάν $(\Omega, \mathscr{A}, \mathbb{P})$ είναι ένας χώρος πιθανότητας, η *δεσμευμένη πιθανότητα* του $A \in \mathscr{A}$ δοθέντος του $E \in \mathscr{A}$, $\mathbb{P}(A \mid E)$, ορίζεται από τη σχέση

$$\mathbb{P}(A \mid E) = \frac{\mathbb{P}(A \cap E)}{\mathbb{P}(E)}$$

υπό τον όρο ότι $\mathbb{P}(E) > 0$.

Από τον ορισμό μπορεί να δειχθεί ότι η δεσμευμένη πιθανότητα $\mathbb{P}(A \mid E)$ ικανοποιεί τα αξιώματα Kolmogorov (για σταθερό E) και επομένως έχει και όλες τις άλλες ιδιότητες της πιθανότητας, π.χ. $\mathbb{P}(A \cup B \mid E) = \mathbb{P}(A \mid E) + \mathbb{P}(B \mid E)$, για $A \cap B = \emptyset$.

Γενικά, η πληροφορία (το ενδεχόμενο) E μεταβάλλει την πιθανότητα του ενδεχομένου A, δηλαδή $\mathbb{P}(A \mid E) \neq \mathbb{P}(A)$. Εάν όμως συμβεί $\mathbb{P}(A \mid E) = \mathbb{P}(A)$, αυτό σημαίνει ότι το ενδεχόμενο E δεν επηρεάζει την πραγματοποίηση του ενδεχομένου A και δια τούτο τα A και E λέγονται *ανεξάρτητα ενδεχόμενα*. Επειδή η τελευταία σχέση δεν είναι συμμετρική ως προς A και E και ισχύει υπό τον όρο $\mathbb{P}(E) > 0$, δια τούτο ως ορισμός ανεξαρτησίας λαμβάνεται η σχέση

$$\mathbb{P}(A \cap E) = \mathbb{P}(A)\mathbb{P}(E)$$

που προκύπτει από την $\mathbb{P}(A \mid E) = \mathbb{P}(A)$ αντικαθιστώντας $\mathbb{P}(A \mid E) = \frac{\mathbb{P}(A \cap E)}{\mathbb{P}(E)}$.

Η έννοια της ανεξαρτησίας μπορεί να γενικευθεί σε περισσότερα από δύο ενδεχόμενα ως εξής.

Ορισμός 1.2.1. *Τα ενδεχόμενα $A_1, \dots, A_n$ ονομάζονται ανεξάρτητα εάν για κάθε υποομάδα $A_{i_1}, \dots, A_{i_\kappa}$, $2 \le \kappa \le n$, με $i_1 \neq \cdots \neq i_\kappa$ ισχύει*

$$\mathbb{P}(\cap_{j=1}^{\kappa} A_{ij}) = \prod_{j=1}^{\kappa} \mathbb{P}(A_{ij}).$$

Ορισμός 1.2.2. *Μία ακολουθία ενδεχομένων A_n, $n = 1, 2, \dots$ ονομάζεται ακολουθία ανεξάρτητων ενδεχομένων, εάν τα ενδεχόμενα κάθε πεπερασμένης υποακολουθίας της είναι ανεξάρτητα.*

Πρόταση 1.2.1. *Έστω ότι B_i, $i = 1, \dots, n$ είναι μια διαμέριση του δειγματικού χώρου με $\mathbb{P}(B_i) > 0$, $i = 1, \dots, n$.*

(α) (Θεώρημα Ολικής Πιθανότητας) Για κάθε ενδεχόμενο A, ισχύει

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i).$$

(β) (τύπος Bayes, εκ των υστέρων πιθανότητα) Ισχύει

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i)}.$$

1.3 Τυχαίες μεταβλητές και κατανομές

Πολλές φορές σε ένα τυχαίο πείραμα μάς ενδιαφέρει κάποια «ποσότητα» (συνάρτηση) η τιμή της οποίας εξαρτάται από το αποτέλεσμα του τυχαίου πειράματος. Αυτή η «ποσότητα» λέγεται *τυχαία μεταβλητή*. Επειδή υπάρχει αβεβαιότητα ως προς το αποτέλεσμα του τυχαίου πειράματος, υπάρχει κατά συνέπεια και αβεβαιότητα ως προς την τιμή της τυχαίας μεταβλητής. Έτσι έχει έννοια να μιλάμε για το *σύνολο των δυνατών τιμών* της τυχαίας μεταβλητής (που θα αναφέρεται απλά ως σύνολο τιμών), για την πιθανότητα η τιμή της τυχαίας μεταβλητής να είναι x ή να ανήκει στο διάστημα $(\alpha,\ \beta)$ κ.λπ. Ο αυστηρός ορισμός της τυχαίας μεταβλητής έχει ως εξής.

Ορισμός 1.3.1. *Έστω $(\Omega, \mathscr{A}, \mathbb{P})$ ένας χώρος πιθανότητας. Μία συνάρτηση $X\colon \Omega \longrightarrow \mathbb{R}$ ονομάζεται τυχαία μεταβλητή (τ.μ.) εάν το σύνολο $\{\omega \in \Omega\colon X(\omega) \leq x\}$ ανήκει στη σ-άλγεβρα $\mathscr{A}$ για κάθε $x \in \Re$.*

Η συνθήκη $\{\omega \in \Omega\colon X(\omega) \leq x\} \in \mathscr{A},\ \forall\ x \in \mathbb{R}$ είναι ισοδύναμη προς τη συνθήκη $X^{-1}(B) = \{\omega \in \Omega\colon X(\omega) \in B\} \in \mathscr{A}$ για κάθε σύνολο Borel B της πραγματικής ευθείας και μας επιτρέπει να ορίσουμε πιθανότητες, που αφορούν την τυχαία μεταβλητή X. Σύνολα Borel είναι τα διαστήματα και όλα τα σύνολα που παράγονται από αριθμήσιμες το πλήθος πράξεις (ενώσεις, τομές, διαφορές) μεταξύ διαστημάτων. Έτσι η πιθανότητα η τιμή της τυχαίας μεταβλητής X να είναι μικρότερη ή ίση του x, συμβολικά $\mathbb{P}(X \leq x)$, ορίζεται από τη σχέση $\mathbb{P}(X \leq x) = \mathbb{P}(\{\omega \in \Omega\colon X(\omega) \leq x\})$. Ανάλογη έννοια και ορισμό έχουν οι συμβολισμοί $\mathbb{P}(X = x)$, $\mathbb{P}(X < x)$, $\mathbb{P}(X > x)$, $\mathbb{P}(X \geq x)$, $\mathbb{P}(\alpha < X \leq \beta)$, $\mathbb{P}(X \in B)$, κ.λ.π.

Η κατανομή πιθανότητας μιας τυχαίας μεταβλητής X καθορίζεται πλήρως από την (αθροιστική) *συνάρτηση κατανομής*

$$F(x) = \mathbb{P}(X \leq x),\ x \in \mathbb{R}.$$

Πιθανότητες που αφορούν την τυχαία μεταβλητή X υπολογίζονται από τη συνάρτηση κατανομής βάσει των τύπων:

α. $\mathbb{P}(X = x) = F(x) - F(x-)$.

β. $\mathbb{P}(X < x) = F(x-)$.

γ. $\mathbb{P}(X > x) = 1 - F(x)$.

δ. $\mathbb{P}(X \geq x) = 1 - F(x-)$.

ε. $\mathbb{P}(\alpha < X \leq \beta) = F(\beta) - F(\alpha)$.

Η συνάρτηση κατανομής έχει τις εξής χαρακτηριστικές ιδιότητες:

α. Είναι αύξουσα.

β. Είναι δεξιά συνεχής.

γ. $\lim_{x\to\infty} F(x) = 1$, $\lim_{x\to-\infty} F(x) = 0$.

Οι τυχαίες μεταβλητές που θα μελετήσουμε είναι *διακριτές* ή (απολύτως) *συνεχείς*.

Ορισμός 1.3.2. *Η τυχαία μεταβλητή X ονομάζεται διακριτή, εάν το σύνολο τιμών της, $\mathcal{S}$, είναι το πολύ αριθμήσιμο (δηλαδή πεπερασμένο ή άπειρο αλλά αριθμήσιμο).*

Η *συνάρτηση πιθανότητας* $p(x)$ της διακριτής τυχαίας μεταβλητής X ορίζεται από τη σχέση $p(x) = \mathbb{P}(X = x)$, $x \in \mathbb{R}$ και καθορίζει (όπως και η συνάρτηση κατανομής) πλήρως την κατανομή πιθανότητας της τυχαίας μεταβλητής X. Στα επόμενα κεφάλαια και για λόγους ομοιομορφίας με τις συνεχείς τυχαίες μεταβλητές, η συνάρτηση πιθανότητας θα συμβολίζεται $f(x)$ και θα αναφέρεται ως πυκνότητα της X.

Ιδιότητες της συνάρτησης πιθανότητας

1. $p(x) \geq 0$, $\forall\, x \in \mathbb{R}$.
2. $\sum_{x\in\mathcal{S}} p(x) = 1$.

Ορισμός 1.3.3. *Η τυχαία μεταβλητή X με συνάρτηση κατανομής $F(x)$ ονομάζεται (απολύτως) συνεχής, εάν υπάρχει μία συνάρτηση $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ με $f(x) \geq 0$, $\forall\, x \in \mathbb{R}$ έτσι ώστε $F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$, $\forall\, x \in \mathbb{R}$. Η f ονομάζεται πυκνότητα πιθανότητας της X ή απλά πυκνότητα της X.*

Από τον ορισμό προκύπτουν οι εξής ιδιότητες.

Ιδιότητες συνεχούς τυχαίας μεταβλητής

1. $1 = \lim_{x\to\infty} F(x) = \lim_{x\to\infty} \int_{-\infty}^{x} f(t)\,\mathrm{d}t = \int_{-\infty}^{\infty} f(t)\,\mathrm{d}t$. Επομένως η συνάρτηση $f(x)$ έχει τις ιδιότητες:

   α. $f(x) \geq 0,\ \forall\, x \in \mathbb{R}$.

   β. $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$.

2. Η F είναι συνεχής στο $\mathbb{R}$ και επιπλέον $F'(x) = f(x)$ για κάθε σημείο συνέχειας x της $f(x)$.

3. $\mathbb{P}(X = x) = 0,\ \forall\, x \in \mathbb{R}$, γιατί $\mathbb{P}(X = x) = F(x) - F(x-)$ και η F είναι συνεχής.

4. Το σύνολο τιμών της τυχαίας μεταβλητής X, έστω $\mathcal{S}$, είναι *μη αριθμήσιμο*. Πράγματι, εάν ήταν αριθμήσιμο, $\mathcal{S} = \{x_1, x_2, \dots\}$, θα είχαμε $1 = \mathbb{P}(X \in \mathcal{S}) = \mathbb{P}((X = x_1) \cup (X = x_2) \cup \dots) = \sum_i \mathbb{P}(X = x_i) = 0$, γιατί $\mathbb{P}(X = x_i) = 0,\ \forall\, i$ (ιδιότητα 3).

5. $\mathbb{P}(\alpha < X \leq \beta) = F(\alpha) - F(\beta) = \int_{-\infty}^{\beta} f(x)\,\mathrm{d}x - \int_{-\infty}^{\alpha} f(x)\,\mathrm{d}x = \int_{\alpha}^{\beta} f(x)\,\mathrm{d}x$. Επίσης λόγω της ιδιότητας 3 έχουμε $\mathbb{P}(\alpha \leq X < \beta) = \mathbb{P}(\alpha < X < \beta) = \mathbb{P}(\alpha \leq X \leq \beta) = \mathbb{P}(\alpha < X \leq \beta) = \int_{\alpha}^{\beta} f(x)\,\mathrm{d}x$.
   Ανάλογα έχουμε $\mathbb{P}(X \leq \alpha) = \mathbb{P}(X < \alpha) = \int_{-\infty}^{\alpha} f(x)\,\mathrm{d}x$ και $\mathbb{P}(X \geq \alpha) = \mathbb{P}(X > \alpha) = \int_{\alpha}^{\infty} f(x)\,\mathrm{d}x$.
   Γενικά μπορεί να δειχθεί ότι $\mathbb{P}(X \in B) = \int_B f(x)\,\mathrm{d}x$ για κάθε σύνολο Borel B. Οι παραπάνω τύποι είναι πολύ χρήσιμοι στον υπολογισμό πιθανοτήτων, που αφορούν την τυχαία μεταβλητή X, μέσω της πυκνότητας.

6. (Ερμηνεία της πυκνότητας πιθανότητας $f(x)$) Έστω $x \in \mathbb{R}$ με $f(t) > 0$ για t σε μία περιοχή $(x - \varepsilon,\ x + \varepsilon)$, όπου $\varepsilon > 0$ και «μικρό». Τότε $\mathbb{P}(x-\varepsilon < X < x+\varepsilon) = \int_{x-\varepsilon}^{x+\varepsilon} f(t)\,\mathrm{d}t \approx f(x)(2\varepsilon) > 0$. Παρατηρούμε λοιπόν ότι, ενώ $\mathbb{P}(X = x) = 0$, η πιθανότητα η τιμή της τυχαίας

μεταβλητής να είναι σε μία «μικρή» περιοχή γύρω από το x είναι θετική και (κατά προσέγγιση) ανάλογη της τιμής $f(x)$. Συνεπώς η $f(x)$ εκφράζει πόσο μεγάλη (πυκνή) είναι η πιθανότητα η τιμή της τυχαίας μεταβλητής X να είναι σε μία μικρή περιοχή γύρω από το x.

7. Το σύνολο τιμών της τυχαίας μεταβλητής X καθορίζεται από το σύνολο $\mathcal{S} = \{x \in \mathbb{R} \colon f(x) > 0\}$, με την έννοια ότι $\mathbb{P}(X \in \mathcal{S}) = 1$ και $\mathbb{P}(X \in \mathcal{S}_1) < 1$ για κάθε υποσύνολο $\mathcal{S}_1$ του $\mathcal{S}$ που έχει θετικό μήκος.

1.4 Μέση τιμή, διασπορά, ροπές, ροπογεννήτρια

Ορισμός 1.4.1. *Η μέση τιμή μιας τυχαίας μεταβλητής X ορίζεται από τη σχέση*

$$\mu = \mathbb{E}X = \begin{cases} \sum_{x \in \mathcal{S}} x p(x) & \text{, εάν η τ.μ. } X \text{ είναι διακριτή} \\ \int_{-\infty}^{\infty} x f(x) \, \mathrm{d}x & \text{, εάν η τ.μ. } X \text{ είναι συνεχής} \end{cases}.$$

Η μέση τιμή παριστάνει μία μορφή σταθμισμένου μέσου όρου των τιμών της X. Πράγματι, ας υποθέσουμε ότι οι τιμές της X είναι $x_1, \ldots, x_n$ και ότι είναι ισοπίθανες, δηλαδή $p(x_i) = \mathbb{P}(X = x_i) = \frac{1}{n}$ $\forall$ $i = 1, \ldots, n$. Τότε, $\mathbb{E}X = \frac{x_1 + \cdots + x_n}{n}$, δηλαδή η μέση τιμή της X είναι ο μέσος όρος των τιμών της. Γενικότερα, εάν $p(x_i) = p_i$ τότε $\mathbb{E}X = p_1 x_1 + \cdots + p_n x_n$ και οι πιθανότητες p_i μπορούν να θεωρηθούν ως «συντελεστές βαρύτητας» για τις αντίστοιχες τιμές x_i. Έτσι, η $\mathbb{E}X$ μπορεί να θεωρηθεί ως μία γενικευμένη μορφή μέσου όρου των τιμών της X.

Η μέση τιμή παρέχει κάποια (αμυδρή έστω) πληροφορία για τη «θέση» των τιμών της X πάνω στην πραγματική ευθεία. Δια τούτο αναφέρεται ως *παράμετρος θέσης* για την κατανομή της X.

Εάν $h \colon \mathbb{R} \longrightarrow \mathbb{R}$ είναι μία συνάρτηση, η μέση τιμή της τυχαίας μεταβλητής $h(X)$ δίνεται από τη σχέση

$$\mathbb{E}h(X) = \begin{cases} \sum_{x \in \mathcal{S}} h(x) p(x) & \text{, εάν η τ.μ. } X \text{ είναι διακριτή} \\ \int_{-\infty}^{\infty} h(x) f(x) \, \mathrm{d}x & \text{, εάν η τ.μ. } X \text{ είναι συνεχής} \end{cases}.$$

Ορισμός 1.4.2. *Η διασπορά μιας τυχαίας μεταβλητής X ορίζεται από τη σχέση*

$$\mathbb{V}ar X = \mathbb{E}(X - \mathbb{E}X)^2.$$

Όπως φαίνεται από τον ορισμό, η διασπορά εκφράζει τη μέση τετραγωνική απόκλιση της X από τη μέση τιμή της $\mathbb{E}X$. Είναι λοιπόν ένα μέτρο μεταβλητότητας των τιμών της X γύρω από την $\mathbb{E}X$. Δια τούτο αναφέρεται ως *παράμετρος μεταβλητότητας* για την κατανομή της X.

Ο υπολογισμός της διασποράς συχνά γίνεται χρησιμοποιώντας την επόμενη πρόταση.

Πρόταση 1.4.1. *Ισχύουν οι εξής σχέσεις.*

α. $\mathbb{V}ar X = \mathbb{E}(X^2) - (\mathbb{E}X)^2$.

β. $\mathbb{E}(X^2) = \mathbb{V}ar X + (\mathbb{E}X)^2$.

Στην επόμενη πρόταση καταγράφονται μερικές ιδιότητες της μέσης τιμής και της διασποράς.

Πρόταση 1.4.2. *Ισχύουν οι εξής σχέσεις.*

1. $\mathbb{E}c = c$, *όπου c σταθερά.*
2. $\mathbb{E}(\alpha X + \beta) = \alpha \mathbb{E}X + \beta$.
3. $\mathbb{V}ar c = 0$, *όπου c σταθερά.*
4. $\mathbb{V}ar(\alpha X + \beta) = \alpha^2 \mathbb{V}ar X$.
5. *Αν* $X \geqslant 0$ *και* $\mathbb{E}X = 0$, *τότε* $\mathbb{P}(X = 0) = 1$.
6. *Αν* $X \geqslant \alpha$ *(α σταθερά) τότε* $\mathbb{E}X \geqslant \alpha$.
7. *Αν* $X \geq Y$ *τότε* $\mathbb{E}X \geqslant \mathbb{E}Y$. *Αν, επιπλέον,* $\mathbb{P}(X > Y) > 0$ *και η* $\mathbb{E}Y$ *είναι πεπερασμένη, τότε* $\mathbb{E}X > \mathbb{E}Y$.
8. $\mathbb{E}(X^2) \geqslant (\mathbb{E}X)^2$ *και η ισότητα ισχύει αν και μόνον αν* $\mathbb{P}(X = \mathbb{E}X) = 1$.

Ορισμός 1.4.3. *Η ροπογεννήτρια μίας τυχαίας μεταβλητής X ορίζεται από τη σχέση*

$$m(t) = \mathbb{E}(e^{tX}),\ t \in \mathbb{R}.$$

Η ροπογεννήτρια παρότι δεν έχει φυσική ερμηνεία, όμως χρησιμεύει στην απόδειξη πολλών αποτελεσμάτων της Θεωρίας Πιθανοτήτων.

Πρόταση 1.4.3. *1.* $m(0) = 1$.

2. Εάν η ροπογεννήτρια $m(t)$ είναι πεπερασμένη για κάθε t σε μία περιοχή του 0, τότε

α. καθορίζει την κατανομή της τυχαίας μεταβλητής,

β. $\frac{\mathrm{d}^\kappa}{\mathrm{d}t^\kappa} m(t)\big|_{t=0} = \mathbb{E}(X^\kappa)$ *(ροπή κ τάξης).*

3. Εάν $m_{\alpha X+\beta}(t)$ είναι η ροπογεννήτρια της τυχαίας μεταβλητής $\alpha X+\beta$ και $m_X(t)$ η ροπογεννήτρια της X, τότε ισχύει

$$m_{\alpha X+\beta}(t) = e^{\beta t} m_X(\alpha t).$$

1.5 Ανισότητες Markov, Tchebychev, Jensen

Πρόταση 1.5.1. *(i) (ανισότητα Markov). Αν Y είναι μη αρνητική τυχαία μεταβλητή, τότε ισχύει η σχέση*

$$\mathbb{P}(Y > \varepsilon) \leqslant \frac{\mathbb{E}Y}{\varepsilon}\ ,\ \ \forall\, \varepsilon > 0.$$

(ii) (ανισότητα Tchebychev). Αν X είναι τυχαία μεταβλητή με $\mathbb{E}X = \mu$ και $\mathbb{V}arX = \sigma^2$ πεπερασμένη, τότε ισχύει η σχέση

$$\mathbb{P}(|X-\mu| > \varepsilon) \leqslant \frac{\sigma^2}{\varepsilon^2}\ ,\ \ \forall\, \varepsilon > 0.$$

Απόδειξη. (i) Ορίζουμε την τυχαία μεταβλητή X από τη σχέση

$$X = \begin{cases} 0 & , \quad 0 \leqslant Y \leqslant \varepsilon \\ \varepsilon & , \quad Y > \varepsilon. \end{cases}$$

Τότε έχουμε $X \leqslant Y$ και επομένως $\mathbb{E}X \leqslant \mathbb{E}Y$ (Πρόταση 1.4.2). Όμως, $\mathbb{E}X = 0\mathbb{P}(X=0) + \varepsilon\mathbb{P}(X=\varepsilon) = \varepsilon\mathbb{P}(Y>\varepsilon)$, οπότε παίρνουμε $\varepsilon\mathbb{P}(Y > \varepsilon) \leqslant \mathbb{E}Y$, δηλαδή $\mathbb{P}(Y>\varepsilon) \leqslant \dfrac{\mathbb{E}Y}{\varepsilon}$.

(ii) Έχουμε $\mathbb{P}(|X-\mu|>\varepsilon) = \mathbb{P}((X-\mu)^2 > \varepsilon^2) \leqslant \dfrac{\mathbb{E}(X-\mu)^2}{\varepsilon^2} = \dfrac{\sigma^2}{\varepsilon^2}$, εφαρμόζοντας την ανισότητα Markov για την τυχαία μεταβλητή $Y=(X-\mu)^2$. □

Μία συνάρτηση h, ορισμένη σε ένα διάστημα, ονομάζεται κυρτή, αν $h(\lambda x + (1-\lambda)y) \leqslant \lambda h(x) + (1-\lambda)h(y)$ για κάθε x, y και $0<\lambda<1$. Αν η ανισότητα είναι γνήσια για κάθε x, y με $x \neq y$, τότε η h ονομάζεται γνησίως κυρτή. Η h είναι κοίλη (γνησίως κοίλη), αν η $-h$ είναι κυρτή (γνησίως κυρτή). Μία κυρτή συνάρτηση «βρίσκεται» πάνω από όλες τις εφαπτομένες της. Αν δηλαδή η ευθεία $\alpha x + \beta$ είναι εφαπτόμενη της γραφικής παράστασης της h, τότε ισχύει $h(x) \geqslant \alpha x + \beta$. Αν, περαιτέρω, η h είναι γνησίως κυρτή, τότε ισχύει $h(x) > \alpha x + \beta$, εκτός από το (κοινό) σημείο επαφής. Υπό την προϋπόθεση ότι υπάρχουν και είναι πεπερασμένες οι μέσες τιμές $\mathbb{E}X$ και $\mathbb{E}h(X)$, ισχύει η εξής ανισότητα για κυρτές συναρτήσεις.

Πρόταση 1.5.2. *(Ανισότητα Jensen) Αν X είναι τυχαία μεταβλητή και h είναι μια κυρτή συνάρτηση, τότε έχουμε*

$$\mathbb{E}h(X) \geqslant h(\mathbb{E}X). \tag{1.1}$$

Αν επιπλέον, η X δεν είναι σταθερά και η h είναι γνησίως κυρτή τότε η ανισότητα (1.1) *είναι γνήσια.*

Απόδειξη. Έστω $\alpha x + \beta$ η εφαπτομένη της h στο σημείο $\mathbb{E}X$. Τότε $h(x) \geqslant \alpha x + \beta$ για κάθε x, σύμφωνα με τα παραπάνω, και $h(\mathbb{E}X) = \alpha\mathbb{E}X + \beta$ αφού $\mathbb{E}X$ είναι σημείο επαφής. Επομένως $h(X) \geqslant \alpha X + \beta$, οπότε παίρνουμε $\mathbb{E}h(X) \geqslant \mathbb{E}(\alpha X + \beta)$ λόγω της Πρότασης 1.4.2(7). Όμως $\mathbb{E}(\alpha X + \beta) = \alpha\mathbb{E}X + \beta = h(\mathbb{E}X)$, οπότε συμπεραίνουμε ότι $\mathbb{E}h(X) \geqslant h(\mathbb{E}X)$. Αν, επιπλέον, η h είναι γνησίως κυρτή και η X δεν είναι σταθερά, τότε έχουμε $\mathbb{P}(h(X) > \alpha X + \beta) > 0$, το οποίο λόγω της Πρότασης 1.4.2(7) συνεπάγεται ότι $\mathbb{E}h(X) > \mathbb{E}(\alpha X + \beta) = h(\mathbb{E}X)$. □

1.6 Μερικές βασικές κατανομές

1.6.1 Μερικές διακριτές κατανομές

Κατανομή Bernoulli

Συμβολισμός: $X \sim \mathcal{B}(1,p)$, $0 < p < 1$.
Συνάρτηση πιθανότητας:

$$p(x) = p^x(1-p)^{1-x},\ x = 0, 1.$$

Μέση τιμή και διασπορά: $\mathbb{E}X = p$, $\mathbb{V}ar X = p(1-p)$.
Ροπογεννήτρια: $m(t) = 1 - p + pe^t$, $t \in \mathbb{R}$.

Διωνυμική κατανομή

Συμβολισμός: $X \sim \mathcal{B}(n,p)$, $0 < p < 1$.
Συνάρτηση πιθανότητας:

$$p(x) = \binom{n}{x} p^x(1-p)^{n-x},\ x = 0, 1, \ldots, n.$$

Μέση τιμή και διασπορά: $\mathbb{E}X = np$, $\mathbb{V}ar X = np(1-p)$.
Ροπογεννήτρια: $m(t) = (1 - p + pe^t)^n$, $t \in \mathbb{R}$.

Γεωμετρική κατανομή

Συμβολισμός: $X \sim \mathcal{G}e(p)$, $0 < p < 1$.
Συνάρτηση πιθανότητας:

$$p(x) = (1-p)^{x-1}p,\ x = 1, 2, \ldots$$

Μέση τιμή και διασπορά: $\mathbb{E}X = \frac{1}{p}$, $\mathbb{V}ar X = \frac{1-p}{p^2}$.
Ροπογεννήτρια: $m(t) = \frac{pe^t}{1-(1-p)e^t}$, $t < -\ln(1-p)$.

Κατανομή Poisson

Συμβολισμός: $X \sim \mathcal{P}(\lambda)$, $\lambda > 0$.
Συνάρτηση πιθανότητας:

$$p(x) = e^{-\lambda}\frac{\lambda^x}{x!},\ x = 0, 1, \ldots .$$

Μέση τιμή και διασπορά: $\mathbb{E}X = \lambda$, $\mathbb{V}arX = \lambda$.
Ροπογεννήτρια: $m(t) = e^{\lambda(e^t-1)}$, $t \in \mathbb{R}$.

1.6.2 Μερικές συνεχείς κατανομές

Ομοιόμορφη κατανομή στο $(\alpha,\, \beta)$

Συμβολισμός: $X \sim \mathcal{U}(\alpha,\, \beta)$, $\alpha < \beta$.
Συνάρτηση πυκνότητας πιθανότητας:

$$f(x) = \frac{1}{\beta - \alpha},\ \alpha < x < \beta.$$

Μέση τιμή και διασπορά: $\mathbb{E}X = \frac{\alpha+\beta}{2}$, $\mathbb{V}arX = \frac{(\beta-\alpha)^2}{12}$.
Ροπογεννήτρια: $m(t) = \frac{e^{t\beta}-e^{t\alpha}}{t(\beta-\alpha)}$, $t \neq 0$, $m(0) = 1$.

Εκθετική κατανομή

Συμβολισμός: $X \sim \mathcal{E}(\theta)$, $\theta > 0$.
Συνάρτηση πυκνότητας πιθανότητας:

$$f(x) = \frac{1}{\theta}e^{-\frac{x}{\theta}},\ x > 0.$$

Μέση τιμή και διασπορά: $\mathbb{E}X = \theta$, $\mathbb{V}arX = \theta^2$.
Ροπογεννήτρια: $m(t) = \frac{1}{1-\theta t}$, $t < \frac{1}{\theta}$.

Γάμμα κατανομή

Συμβολισμός: $X \sim \mathcal{G}(\alpha,\, \beta)$, $\alpha, \beta > 0$.
Συνάρτηση πυκνότητας πιθανότητας:

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-\frac{x}{\beta}},\ x > 0.$$

Μέση τιμή και διασπορά: $\mathbb{E}X = \alpha\beta$, $\mathbb{V}arX = \alpha\beta^2$.
Ροπογεννήτρια: $m(t) = \frac{1}{(1-\beta t)^\alpha}$, $t < \frac{1}{\beta}$.

Κατανομή Weibull

Συμβολισμός: $X \sim W(\alpha,\, \beta)$, $\alpha, \beta > 0$.
Συνάρτηση πυκνότητας πιθανότητας:

$$f(x) = \frac{\alpha x^{\alpha-1}}{\beta^\alpha} e^{-\left(\frac{x}{\beta}\right)^\alpha},\ x > 0.$$

Μέση τιμή και διασπορά: $\mathbb{E}X = \beta\Gamma\left(1+\frac{1}{\alpha}\right)$, $\mathbb{V}arX = \beta^2\left[\Gamma\left(1+\frac{2}{\alpha}\right) - \left(\Gamma\left(1+\frac{1}{\alpha}\right)\right)^2\right]$.
Ροπογεννήτρια: $m(t) = \displaystyle\sum_{k=0}^{+\infty} \frac{t^k\beta^k}{k!}\Gamma\left(1+\frac{k}{\alpha}\right)$.

χι–τετράγωνο κατανομή με n βαθμούς ελευθερίας

Συμβολισμός: $X \sim \chi^2_n$.
Συνάρτηση πυκνότητας πιθανότητας:

$$f(x) = \frac{1}{\Gamma(n/2)2^{n/2}}\, x^{n/2-1} e^{-\frac{x}{2}},\ x > 0.$$

Μέση τιμή και διασπορά: $\mathbb{E}X = n$, $\mathbb{V}arX = 2n$.
Ροπογεννήτρια: $m(t) = (1-2t)^{-n/2}$, $t < 1/2$.
Η παράμετρος n αναφέρεται ως *βαθμοί ελευθερίας* της κατανομής.

Βήτα κατανομή

Συμβολισμός: $X \sim \mathcal{B}eta(\alpha,\, \beta)$, $\alpha, \beta > 0$.
Συνάρτηση πυκνότητας πιθανότητας:

$$f(x) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1},\ 0 < x < 1.$$

Μέση τιμή και διασπορά: $\mathbb{E}X = \frac{\alpha}{\alpha+\beta}$, $\mathbb{V}arX = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$.

Κατανομή t με n βαθμούς ελευθερίας

Συμβολισμός: $X \sim t_n$.
Συνάρτηση πυκνότητας πιθανότητας:

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \; x \in \mathbb{R}.$$

Μέση τιμή και διασπορά: $\mathbb{E}X = 0$ ($n \geqslant 2$), $\mathbb{V}arX = \dfrac{n}{n-2}$ ($n \geqslant 3$).
Η παράμετρος n αναφέρεται ως *βαθμοί ελευθερίας* της κατανομής.

Κατανομή Cauchy

Συμβολισμός: $X \sim \mathcal{C}(\theta)$, $\theta \in \mathbb{R}$.
Συνάρτηση πυκνότητας πιθανότητας:

$$f(x) = \frac{1}{\pi(1 + (x-\theta)^2)}, \; x \in \mathbb{R}.$$

Μέση τιμή και διασπορά: δεν υπάρχουν.

Κατανομή $\mathcal{F}$ με n και m βαθμούς ελευθερίας

Συμβολισμός: $X \sim \mathcal{F}_{n,m}$.
Συνάρτηση πυκνότητας πιθανότητας:

$$f(x) = \frac{\Gamma(\frac{n+m}{2})(\frac{n}{m})^{n/2}\, x^{n/2-1}}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})(1 + \frac{n}{m}\, x)^{\frac{n+m}{2}}}, \; x > 0.$$

Μέση τιμή και διασπορά: $\mathbb{E}X = \dfrac{m}{m-2}$ ($m \geqslant 3$),
$\mathbb{V}arX = \dfrac{2m^2(n+m-2)}{n(m-2)^2(m-4)}$ ($m \geqslant 5$).
Οι παράμετροι n και m αναφέρονται ως *βαθμοί ελευθερίας* της κατανομής.

Κανονική κατανομή

Συμβολισμός: $X \sim \mathcal{N}(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma > 0$.
Συνάρτηση πυκνότητας πιθανότητας:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \; x \in \mathbb{R}.$$

Μέση τιμή και διασπορά: $\mathbb{E}X = \mu$, $\mathbb{V}arX = \sigma^2$.
Ροπογεννήτρια: $m(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$, $t \in \mathbb{R}$.

Πρόταση 1.6.1. ***(Ιδιότητες της κανονικής κατανομής)***

1. *Εάν* $X \sim \mathcal{N}(\mu, \sigma^2)$, *τότε* $\alpha X + \beta \sim \mathcal{N}(\alpha\mu + \beta, \alpha^2\sigma^2)$, $\alpha \neq 0$.

2. *Εάν* $X \sim \mathcal{N}(\mu, \sigma^2)$, *τότε* $\frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$ *(τυπική κανονική).*

3. *Εάν* $X \sim \mathcal{N}(0, 1)$, *τότε* $X^2 \sim \chi^2_1$.

4. *Εάν* $X \sim \mathcal{N}(\mu, \sigma^2)$, *τότε* $\left(\frac{X-\mu}{\sigma}\right)^2 \sim \chi^2_1$.

1.7 Κατανομή συνάρτησης τυχαίας μεταβλητής

Έστω X μία τυχαία μεταβλητή με συνάρτηση πιθανότητας $p(x)$ ή συνάρτηση πυκνότητας πιθανότητας $f(x)$. Ακόμη έστω $g\colon \mathbb{R} \longrightarrow \mathbb{R}$ μία συνάρτηση. Τότε η $Y = g(X)$ είναι τυχαία μεταβλητή και η κατανομή της μπορεί να βρεθεί ως εξής.

Εάν η X είναι διακριτή, τότε και η Y είναι διακριτή με συνάρτηση πιθανότητας που δίνεται από τον τύπο

$$p_Y(y) = \mathbb{P}(Y = y) = \mathbb{P}(g(X) = y) = \sum_{\{x\colon g(x)=y\}} p(x).$$

Για συνεχή τυχαία μεταβλητή, η κατανομή της Y δίνεται στην ακόλουθη πρόταση.

Πρόταση 1.7.1. *Εάν η X είναι συνεχής και η $g(x)$ είναι παραγωγίσιμη με $g'(x) \neq 0$ για κάθε x στο σύνολο των τιμών της X, τότε και η Y είναι συνεχής με συνάρτηση πυκνότητας πιθανότητας που δίνεται από τον τύπο*

$$f_Y(y) = f(g^{-1}(y)) \left| \frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y) \right|.$$

1.8 Τυχαία διανύσματα

Έστω $X_1, X_2, \ldots, X_n$ τυχαίες μεταβλητές ορισμένες στον ίδιο χώρο πιθανότητας, $(\Omega, \mathscr{A}, \mathbb{P})$. Κάθε μία από αυτές τις τυχαίες μεταβλητές έχει κάποια κατανομή, αλλά συχνά μας ενδιαφέρει να τις μελετήσουμε συγχρόνως. Σε αυτές τις περιπτώσεις λοιπόν μας ενδιαφέρει η *από κοινού κατανομή* των $X_1, X_2, \ldots, X_n$ την οποία θα ορίσουμε στη συνέχεια.

Ορισμός 1.8.1. *Το διάνυσμα $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ονομάζεται n-διάστατο τυχαίο διάνυσμα.*

Η *κατανομή* πιθανότητας του $\underset{\sim}{X}$ ή *από κοινού κατανομή* των $X_1, \ldots, X_n$ καθορίζεται από την *από κοινού συνάρτηση κατανομής*

$$F(\underset{\sim}{x}) = F(x_1, \ldots, x_n) = \mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n) ,$$

όπου $\underset{\sim}{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$.

Η $F(x_1, \ldots, x_n)$ έχει ανάλογες ιδιότητες όπως η συνάρτηση κατανομής μίας τυχαίας μεταβλητής, για παράδειγμα,

$$\lim_{x_1 \to \infty, \ldots, x_n \to \infty} F(x_1, \ldots, x_n) = 1, \qquad \lim_{x_1 \to -\infty, \ldots, x_n \to -\infty} F(x_1, \ldots, x_n) = 0.$$

Τα τυχαία διανύσματα που θα μελετήσουμε είναι *διακριτά* ή (απολύτως) *συνεχή*.

Ορισμός 1.8.2. *Το τυχαίο διάνυσμα $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ονομάζεται διακριτό, εάν το σύνολο τιμών του είναι το πολύ αριθμήσιμο υποσύνολο του $\mathbb{R}^n$.*

Από τον ορισμό είναι προφανές ότι το $\underset{\sim}{X}$ είναι διακριτό, εάν και μόνον εάν οι συνιστώσες X_i, $i = 1, \ldots, n$ είναι διακριτές τυχαίες μεταβλητές.

Η *συνάρτηση πιθανότητας* $p(\underset{\sim}{x}) = p(x_1, \ldots, x_n)$ του διακριτού $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ορίζεται από τη σχέση $p(\underset{\sim}{x}) = p(x_1, \ldots, x_n) = \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n)$, $\underset{\sim}{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ και καθορίζει (όπως και η από κοινού συνάρτηση κατανομής) την κατανομή του $\underset{\sim}{X}$.

Ιδιότητες

1. $p(\underset{\sim}{x}) \geq 0$, $\forall\, \underset{\sim}{x}$.
2. $\sum_{\underset{\sim}{x}} p(\underset{\sim}{x}) = 1$.

Η συνάρτηση πιθανότητας $p_{X_i}(x_i)$ της συνιστώσας X_i προσδιορίζεται από τη συνάρτηση πιθανότητας $p(\underset{\sim}{x}) = p(x_1, \ldots, x_n)$ του $\underset{\sim}{X}$ μέσω του τύπου

$$p_{X_i}(x_i) = \sum_{x_j,\, j \neq i} p(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n).$$

Ορισμός 1.8.3. *Το τυχαίο διάνυσμα* $\underset{\sim}{X} = (X_1, \ldots, X_n)$ *με συνάρτηση κατανομής* $F(\underset{\sim}{x}) = F(x_1, \ldots, x_n)$ *ονομάζεται (απολύτως) συνεχές εάν υπάρχει συνάρτηση* $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ *με* $f(\underset{\sim}{x}) = f(x_1, \ldots, x_n) \geq 0$, $\forall\, \underset{\sim}{x} \in \mathbb{R}^n$ *έτσι ώστε*

$$F(\underset{\sim}{x}) = F(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(t_1, \ldots, t_n)\, \mathrm{d}t_n \ldots \mathrm{d}t_1, \quad \forall\, \underset{\sim}{x} \in \mathbb{R}^n.$$

Η συνάρτηση $f(\underset{\sim}{x})$ λέγεται *συνάρτηση πυκνότητας πιθανότητας* του $\underset{\sim}{X}$ ή (απλώς) *πυκνότητα* του $\underset{\sim}{X}$ ή *από κοινού πυκνότητα* των $X_1, \ldots, X_n$ και έχει τις εξής χαρακτηριστικές ιδιότητες.

Ιδιότητες

1. $f(\underset{\sim}{x}) \geq 0$, $\forall\, \underset{\sim}{x} \in \mathbb{R}^n$.
2. $\int_{\mathbb{R}^n} f(\underset{\sim}{x})\, \mathrm{d}\underset{\sim}{x} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_n)\, \mathrm{d}x_n \ldots \mathrm{d}x_1 = 1$.

Η πυκνότητα $f_{X_i}(x_i)$ της συνιστώσας X_i προσδιορίζεται από την πυκνότητα $f(\underset{\sim}{x})$ του $\underset{\sim}{X}$ μέσω του τύπου

$$f_{X_i}(x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n) \times$$
$$\times\, \mathrm{d}x_n \ldots \mathrm{d}x_{i+1}\, \mathrm{d}x_{i-1} \ldots \mathrm{d}x_1.$$

Εάν $\underset{\sim}{X} = (X_1, \ldots, X_n)$ είναι ένα τυχαίο διάνυσμα και $g\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ μία συνάρτηση, τότε η μέση τιμή της τυχαίας μεταβλητής $g(\underset{\sim}{X}) = g(X_1, \ldots, X_n)$ ορίζεται από τη σχέση

$$\mathbb{E}g(\underset{\sim}{X}) = \begin{cases} \sum_{x_1} \cdots \sum_{x_n} g(\underset{\sim}{x})p(\underset{\sim}{x}), & \text{εάν το } \underset{\sim}{X} \text{ είναι διακριτό,} \\ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(\underset{\sim}{x})f(\underset{\sim}{x})\,\mathrm{d}x_n \ldots \mathrm{d}x_1, & \text{εάν το } \underset{\sim}{X} \text{ είναι συνεχές.} \end{cases}$$

Για $n = 2$ και $g(x_1, x_2) = \big(x_1 - \mathbb{E}X_1\big)\big(x_2 - \mathbb{E}X_2\big)$ η $\mathbb{E}\big(g(X_1, X_2)\big) = \mathbb{E}\big[\big(X_1 - \mathbb{E}X_1\big)\big(X_2 - \mathbb{E}X_2\big)\big]$ λέγεται *συνδιασπορά* των τυχαίων μεταβλητών X_1, X_2 και συμβολίζεται $\mathbb{C}ov(X_1, X_2)$.

Πρόταση 1.8.1. *Ισχύουν οι εξής ιδιότητες.*

1. $\mathbb{C}ov(X_1, X_2) = \mathbb{E}(X_1X_2) - \mathbb{E}(X_1)\,\mathbb{E}(X_2)$.
2. $\mathbb{C}ov(X_1, X_2) = \mathbb{C}ov(X_2, X_1)$.
3. $\mathbb{C}ov(\alpha X_1, \beta X_2) = \alpha\beta\,\mathbb{C}ov(X_1, X_2)$, α, β *σταθερές.*
4. $\mathbb{C}ov(X_1 + \alpha, X_2 + \beta) = \mathbb{C}ov(X_1, X_2)$.
5. $\mathbb{C}ov(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j) = \sum_{i=1}^n \sum_{j=1}^m \mathbb{C}ov(X_i, Y_j)$.
6. $|\mathbb{C}ov(X_1, X_2)| \le \sqrt{\mathbb{V}ar(X_1)\,\mathbb{V}ar(X_2)}$ *(βλέπε Πρόταση 1.8.2)*
7. $\mathbb{C}ov(X_1, X_2) = 0$, *εάν και μόνον εάν* $\mathbb{E}(X_1X_2) = \mathbb{E}(X_1)\,\mathbb{E}(X_2)$. *Σε αυτή την περίπτωση οι τυχαίες μεταβλητές* X_1, X_2 *ονομάζονται ασυσχέτιστες.*
8. $\mathbb{C}ov(X_1, X_1) = \mathbb{V}ar(X_1)$.
9. $\mathbb{V}ar(X_1 + \cdots + X_n) = \sum_{i=1}^n \mathbb{V}ar(X_i) + 2\sum_{i<j} \mathbb{C}ov(X_i, X_j)$.
10. $\mathbb{V}ar(X_1 + \cdots + X_n) = \sum_{i=1}^n \mathbb{V}ar(X_i)$ *εάν και μόνον εάν οι τυχαίες μεταβλητές* $X_1, \ldots, X_n$ *είναι ανά δύο ασυσχέτιστες.*

Πρόταση 1.8.2. (Ανισότητα Cauchy–Schwarz) *Έστω* Y *και* Z *τυχαίες μεταβλητές με* $\mathbb{V}arY < \infty$ *και* $\mathbb{V}arZ < \infty$. *Τότε ισχύουν τα εξής.*

1. *$\mathbb{C}ov^2(Y,Z) \le \mathbb{V}arY \cdot \mathbb{V}arZ$.*

2. *Εάν Y είναι σταθερά ή Z είναι σταθερά, τότε η 1 ισχύει ως ισότητα.*

3. *Εάν καμία από τις Y και Z δεν είναι σταθερά και η 1 ισχύει ως ισότητα, τότε υπάρχουν σταθερές c_1 και c_2 με $c_1 \neq 0$, τέτοιες ώστε $\mathbb{P}(Z = c_1Y + c_2) = 1$.*

4. *(Αντίστροφο της 3) Εάν οι Z και Y συνδέονται γραμμικά, δηλαδή $Z = c_1Y + c_2$, τότε η ανισότητα 1 ισχύει ως ισότητα.*

Απόδειξη. 1. Εάν $Y = c$ (σταθερά), τότε και τα δύο μέλη της 1 είναι 0 και άρα η 1 ισχύει (ως ισότητα).
Έστω ότι Y δεν είναι σταθερά. Τότε έχουμε $\mathbb{V}arY > 0$. Επιπλέον, για κάθε $t \in \mathbb{R}$ ισχύει $\mathbb{V}ar(tY + Z) \ge 0$ (η διασπορά είναι μη αρνητική). Επειδή όμως $\mathbb{V}ar(tY + Z) = \mathbb{E}(tY + Z)^2 - \big[\mathbb{E}(tY + Z)\big]^2 = \mathbb{E}(t^2Y^2 + 2tYZ + Z^2) - (t\mathbb{E}Y + \mathbb{E}Z)^2 = t^2\mathbb{V}arY + \mathbb{V}arZ + 2t\mathbb{C}ov(Y,Z)$, προκύπτει ότι $t^2\mathbb{V}arY + 2t\mathbb{C}ov(Y,Z) + \mathbb{V}arZ \ge 0$, $\forall\, t \in \mathbb{R}$. Η τελευταία σχέση δηλώνει ότι ένα δευτεροβάθμιο ως προς t τριώνυμο της μορφής $\alpha t^2 + \beta t + \gamma$ με $\alpha > 0$ έχει μη αρνητικές τιμές για κάθε $t \in \mathbb{R}$ και επομένως η διακρίνουσα $\mathscr{D} = \beta^2 - 4\alpha\gamma \le 0$. Άρα, $4\mathbb{C}ov^2(Y,Z) - 4\mathbb{V}arY \cdot \mathbb{V}arZ \le 0$, δηλαδή η 1 ισχύει.
2. Έχει ήδη δειχθεί στο 1.
3. Έστω ότι Y και Z δεν είναι σταθερές και ισχύει η 1 ως ισότητα. Άρα, $\mathscr{D} = 0$ και επομένως το τριώνυμο $t^2\mathbb{V}arY + 2t\mathbb{C}ov(Y,Z) + \mathbb{V}arZ = \mathbb{V}ar(tY + Z)$ έχει διπλή ρίζα, έστω t_0. Τότε έχουμε ότι $\mathbb{V}ar(t_0Y + Z) = 0$ και συνεπώς $\mathbb{P}(t_0Y + Z = c) = 1$, όπου c σταθερά. Περαιτέρω είναι $t_0 \neq 0$, γιατί $t_0 = 0$ συνεπάγεται $\mathbb{V}arZ = 0$ που αποκλείεται, αφού η Z δεν είναι σταθερά. Τελικά, έχουμε $\mathbb{P}(Z = -t_0Y + c) = 1$.
4. Από τις ιδιότητες της συνδιασποράς (Πρόταση 1.8.1) έχουμε

$$\begin{aligned}\mathbb{C}ov(Y,Z) &= \mathbb{C}ov(Y, c_1Y + c_2)\\ &= \mathbb{C}ov(Y, c_1Y) = c_1\mathbb{C}ov(Y,Y) = c_1\mathbb{V}arY.\end{aligned}$$

Συνεπώς, $\mathbb{C}ov^2(Y,Z) = c_1^2(\mathbb{V}arY)^2 = \mathbb{V}arY \cdot \mathbb{V}arZ$. $\square$

Ορισμός 1.8.4. *Οι τυχαίες μεταβλητές $X_1, \ldots, X_n$ (ορισμένες στον ίδιο χώρο $(\Omega, \mathscr{A}, \mathbb{P})$) λέγονται ανεξάρτητες εάν για κάθε σύνολο Borel $B_i \subset \mathbb{R}$, $i = 1, \ldots, n$, ισχύει*

$$\mathbb{P}(X_1 \in B_1, \ldots, X_n \in B_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \in B_i).$$

Ανάλογα ορίζεται η έννοια ανεξαρτησίας τυχαίων διανυσμάτων. Από τον ορισμό προκύπτει αμέσως ότι, εάν οι τυχαίες μεταβλητές $X_1, \ldots, X_n$ είναι ανεξάρτητες τότε και οι τυχαίες μεταβλητές $X_{i_1}, \ldots, X_{i_\kappa}$, $2 \leq \kappa \leq n$, $\{i_1, \ldots, i_\kappa\} \subset \{1, 2, \ldots, n\}$ είναι επίσης ανεξάρτητες.
Γενικά, η κατανομή του $\underset{\sim}{X} = (X_1, \ldots, X_n)$ δεν μπορεί να προσδιοριστεί από τις κατανομές των $X_1, \ldots, X_n$. Μπορεί όμως να προσδιοριστεί στην ειδική (και σημαντική) περίπτωση που οι τυχαίες μεταβλητές $X_1, \ldots, X_n$ είναι ανεξάρτητες.

Πρόταση 1.8.3. *Έστω $X_1, \ldots, X_n$ ανεξάρτητες τυχαίες μεταβλητές με συναρτήσεις κατανομής $F_{X_i}(x_i)$, $i = 1, \ldots, n$. Τότε έχουμε τα εξής.*

1. $F_{\underset{\sim}{X}}(\underset{\sim}{x}) = \prod_{i=1}^{n} F(x_i), \quad \forall\, \underset{\sim}{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n.$

2. *Εάν οι τυχαίες μεταβλητές $X_1, \ldots, X_n$ είναι διακριτές με συναρτήσεις πιθανότητας $p_{X_i}(x_i)$, $i = 1, \ldots, n$, αντίστοιχα, τότε*

$$p_{\underset{\sim}{X}}(\underset{\sim}{x}) = \prod_{i=1}^{n} p_{X_i}(x_i), \quad \forall\, \underset{\sim}{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

3. *Εάν οι τυχαίες μεταβλητές $X_1, \ldots, X_n$ είναι συνεχείς με συναρτήσεις πυκνότητας $f_{X_i}(x_i)$, $i = 1, \ldots, n$, αντίστοιχα, τότε*

$$f_{\underset{\sim}{X}}(\underset{\sim}{x}) = \prod_{i=1}^{n} f_{X_i}(x_i), \quad \forall\, \underset{\sim}{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

Πρόταση 1.8.4.

1. *Εάν οι τυχαίες μεταβλητές $X_1, \ldots, X_n$ είναι ανεξάρτητες και $g_i : \mathbb{R} \longrightarrow \mathbb{R}$ είναι συναρτήσεις τότε και οι τυχαίες μεταβλητές $g_1(X_1)$, $\ldots, g_n(X_n)$ είναι επίσης ανεξάρτητες τυχαίες μεταβλητές. Ανάλογη ιδιότητα ισχύει και για τυχαία διανύσματα.*

2. *Εάν οι τυχαίες μεταβλητές* $X_1, \ldots, X_n$ *είναι ανεξάρτητες, τότε*

$$\mathbb{E}(X_1 \cdots X_n) = \prod_{i=1}^{n} \mathbb{E}X_i.$$

3. *Εάν οι τυχαίες μεταβλητές* X_1, X_2 *είναι ανεξάρτητες, τότε είναι και ασυσχέτιστες (το αντίστροφο όμως δεν ισχύει).*

4. *Εάν οι τυχαίες μεταβλητές* $X_1, \ldots, X_n$ *είναι ανεξάρτητες τότε* $\mathbb{V}ar(X_1 + \ldots + X_n) = \sum_{i=1}^{n} \mathbb{V}ar X_i$.

5. *Εάν οι τυχαίες μεταβλητές* $X_1, \ldots, X_n$ *είναι ανεξάρτητες και* $m_{X_i}(t)$, $i = 1, \ldots, n$, $m_{\sum_{i=1}^{n} X_i}(t)$ *είναι οι ροπογεννήτριες των* X_i, $i = 1, \ldots, n$ *και* $\sum_{i=1}^{n} X_i$ *αντίστοιχα, τότε*

$$m_{\sum_{i=1}^{n} X_i}(t) = \prod_{i=1}^{n} m_{X_i}(t), \quad \forall\, t.$$

Χρησιμοποιώντας την τελευταία ιδιότητα και την Πρόταση 1.4.3 μπορούμε να αποδείξουμε την επόμενη πρόταση.

Πρόταση 1.8.5. *Έστω* $X_1, \ldots, X_n$ *ανεξάρτητες τυχαίες μεταβλητές.*

1. *Εάν* $X_i \sim \mathcal{B}(1, p)$ *(Bernoulli), τότε*

$$\sum_{i=1}^{n} X_i \sim \mathcal{B}\left(n, p\right).$$

2. *Εάν* $X_i \sim \mathcal{B}(n_i, p)$ *(διωνυμική), τότε*

$$\sum_{i=1}^{n} X_i \sim \mathcal{B}\left(\sum_{i=1}^{n} n_i, p\right).$$

3. *Εάν* $X_i \sim \mathcal{P}(\lambda_i)$ *(Poisson), τότε*

$$\sum_{i=1}^{n} X_i \sim \mathcal{P}\left(\sum_{i=1}^{n} \lambda_i\right).$$

4. *Εάν* $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ *(κανονική), τότε*

$$\sum_{i=1}^{n} X_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right).$$

5. *Εάν* $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ *(κανονική) και* $a_1, \ldots, a_n$, β *είναι σταθερές τότε*

$$\sum_{i=1}^{n} a_i X_i + \beta \sim \mathcal{N}\left(\mu, \sigma^2\right),$$

   όπου $\mu = \sum_{i=1}^{n} a_i \mu_i + \beta$ *και* $\sigma^2 = \sum_{i=1}^{n} a_i^2 \sigma_i^2$.

6. *Εάν* $X_i \sim \mathcal{E}(\theta)$ *(εκθετική), τότε*

$$\sum_{i=1}^{n} X_i \sim \mathcal{G}(n, \theta) \quad \text{(Γάμμα)}.$$

7. *Εάν* $X_i \sim \mathcal{G}(\alpha_i, \beta)$ *(Γάμμα), τότε*

$$\sum_{i=1}^{n} X_i \sim \mathcal{G}\left(\sum_{i=1}^{n} \alpha_i, \beta\right).$$

8. *Εάν* $X_i \sim \chi^2_{n_i}$ *(χι-τετράγωνο), τότε*

$$\sum_{i=1}^{n} X_i \sim \chi^2_{\sum_{i=1}^{n} n_i}.$$

9. *Εάν* $X \sim \mathcal{G}(\alpha, \beta)$ *και* $c > 0$, *τότε* $cX \sim \mathcal{G}(\alpha, c\beta)$.

10. *Εάν* $X \sim \mathcal{G}(n, \beta)$, *τότε* $\frac{2}{\beta} X \sim \chi^2_{2n}$.

11. *Εάν* $X_1 \sim \mathcal{N}(0,1)$ *και* $X_2 \sim \chi^2_n$, *τότε* $\dfrac{X_1}{\sqrt{X_2/n}} \sim t_n$.

12. *Εάν* $X_1 \sim \chi^2_p$, $X_2 \sim \chi^2_q$, *τότε* $\dfrac{X_1/p}{X_2/q} \sim \mathcal{F}_{p,q}$.

Πρόταση 1.8.6. *Έστω* $X_1, \ldots, X_n$, $n \geq 2$, *ανεξάρτητες τυχαίες μεταβλητές με κοινή κανονική κατανομή* $\mathcal{N}(\mu, \sigma^2)$. *Τότε ισχύουν τα εξής.*

1. $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ *έχει κανονική κατανομή* $\mathcal{N}(\mu, \frac{\sigma^2}{n})$.

2. $\bar{X}$, $(X_1 - \bar{X}, \ldots, X_n - \bar{X})$ *είναι ανεξάρτητα.*

3. $\bar{X}$, $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ *είναι ανεξάρτητες τυχαίες μεταβλητές.*

4. $\displaystyle\sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2}$ *έχει κατανομή* χ^2_n.

5. $\displaystyle\sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{\sigma^2}$ *έχει κατανομή* χ^2_{n-1}.

Σε πολλά θέματα Πιθανοτήτων και Στατιστικής η μέγιστη ή και η ελάχιστη μεταξύ τυχαίων μεταβλητών $X_1, X_2, \ldots, X_n$ παίζουν σημαντικό ρόλο. Η Πρόταση 1.8.7 δίνει την πυκνότητα αυτών των τυχαίων μεταβλητών στην περίπτωση που οι αρχικές τυχαίες μεταβλητές είναι συνεχείς, ανεξάρτητες και έχουν κοινή κατανομή.

Πρόταση 1.8.7. *Έστω* $X_1, \ldots, X_n$ *ανεξάρτητες τυχαίες μεταβλητές με συνεχή κατανομή, κοινή συνάρτηση κατανομής* F *και κοινή πυκνότητα* f*. Θέτουμε* $X_{(n)} = \max\{X_1, \ldots, X_n\}$ *και* $X_{(1)} = \min\{X_1, \ldots, X_n\}$*. Τότε οι πυκνότητες των* $X_{(n)}$ *και* $X_{(1)}$ *δίνονται, αντίστοιχα, από τους τύπους*

$$f_{X_{(n)}}(x) = n[F(x)]^{n-1}f(x)\ , \quad \textit{και} \quad f_{X_{(1)}}(x) = n[1 - F(x)]^{n-1}f(x).$$

Η ακόλουθη πρόταση δίνει μία χαρακτηριστική ιδιότητα της κατανομής Γάμμα.

Πρόταση 1.8.8. *(i) Εάν* X_1 *και* X_2 *είναι ανεξάρτητες τυχαίες μεταβλητές με Γάμμα κατανομές* $\mathcal{G}(\alpha_1, \beta)$ *και* $\mathcal{G}(\alpha_2, \beta)$*, αντίστοιχα, τότε οι* $X_1 + X_2$ *και* $\dfrac{X_1}{X_1 + X_2}$ *είναι ανεξάρτητες με κατανομές Γάμμα* $\mathcal{G}(\alpha_1 + \alpha_2, \beta)$ *και Βήτα* $\mathcal{B}eta(\alpha_1, \alpha_2)$*, αντίστοιχα.*

(ii) Εάν $X_1, \ldots, X_n$ *είναι ανεξάρτητες τυχαίες μεταβλητές με Γάμμα κατανομές* $\mathcal{G}(\alpha_i, \beta)$*,* $i = 1, \ldots, n$*, αντίστοιχα τότε οι* $X_1 + \ldots + X_n$ *και* $\dfrac{X_1}{X_1 + \ldots + X_n}$ *είναι ανεξάρτητες με κατανομές Γάμμα* $\mathcal{G}(\alpha, \beta)$ *και Βήτα* $\mathcal{B}eta(\alpha_1, \alpha)$*, αντίστοιχα, όπου* $\alpha = \sum_{i=1}^{n}\alpha_i$.

Απόδειξη. Bickel and Doksum (1977, σελ. 13–14). □

Πρόταση 1.8.9. *(i) Για $\kappa \in \{1, 2, \ldots, n\}$ και $\theta \in (0, 1)$ ισχύει η σχέση*

$$\sum_{x=\kappa}^{n} \binom{n}{x} \theta^x (1-\theta)^{n-x} = \frac{1}{B(\kappa, n-\kappa+1)} \int_0^\theta t^{\kappa-1}(1-t)^{n-\kappa} dt. \qquad (1.2)$$

(ii) Αν X και Y είναι τυχαίες μεταβλητές με διωνυμική $\mathcal{B}(n, \theta)$ και Βήτα $\mathcal{B}eta(\kappa, n-\kappa+1)$ κατανομή, αντίστοιχα, τότε ισχύει

$$\mathbb{P}(X \geqslant \kappa) = \mathbb{P}(Y \leqslant \theta). \qquad (1.3)$$

Απόδειξη. (i) Η (1.2) είναι γνωστή σχέση μεταξύ των διωνυμικών πιθανοτήτων και της μη πλήρους συνάρτησης Βήτα που ορίζεται από τη σχέση $I_{\alpha,\beta}(\theta) = \int_0^\theta t^{\alpha-1}(1-t)^{\beta-1}dt$, $0 \leqslant \theta \leqslant 1$. Μπορεί να αποδειχθεί με διαδοχικές παραγοντικές ολοκληρώσεις. (Για $\theta = 1$, $I_{\alpha,\beta}(1) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt = B(\alpha, \beta)$)
(ii) Η (1.2) είναι απλώς ένας άλλος (πιθανοτικός) τρόπος γραφής της (1.2). □

Πρόταση 1.8.10. *Αν η τυχαία μεταβλητή V έχει κατανομή $\mathcal{F}_{p,q}$, τότε η τυχαία μεταβλητή $U = \dfrac{(p/q)V}{1+(p/q)V}$ έχει κατανομή Βήτα $\mathcal{B}eta(p/2, q/2)$.*

Απόδειξη. Η V, εξ' ορισμού της $\mathcal{F}_{p,q}$ (Πρόταση 1.8.5 ή Κεφάλαιο 9), έχει την ίδια κατανομή με την $\dfrac{q}{p}\dfrac{X}{Y}$ όπου X, Y είναι ανεξάρτητες τυχαίες μεταβλητές με κατανομές χ^2_p και χ^2_q αντίστοιχα. Επομένως η U έχει την ίδια κατανομή με την τυχαία μεταβλητή $\dfrac{X/Y}{1+X/Y} = \dfrac{X}{X+Y}$. Η X ως χ^2_p είναι $\mathcal{G}(p/2, 2)$ και η Y ως χ^2_q είναι $\mathcal{G}(q/2, 2)$. Επομένως, το αποτέλεσμα προκύπτει από την Πρόταση 1.8.8. □

1.9 Δεσμευμένες κατανομές

Θα περιορισθούμε σε δύο τυχαίες μεταβλητές X_1, X_2. Η γενίκευση σε περισσότερες από δύο τυχαίες μεταβλητές είναι άμεση. Κάθε μία από τις τυχαίες μεταβλητές έχει κάποια κατανομή. Είναι δυνατό γνώση της τιμής της μίας τυχαίας μεταβλητής να επηρεάσει την κατανομή της άλλης. Έτσι με δεδομένη την τιμή $X_2 = x_2$ η κατανομή της X_1 είναι, γενικά, διαφορετική από την αρχική κατανομή της. Αυτή η δεύτερη κατανομή λέγεται *δεσμευμένη κατανομή*. Εδώ σημειώνουμε ότι, εάν οι τυχαίες μεταβλητές X_1, X_2 είναι ανεξάρτητες, γνώση της τιμής $X_2 = x_2$ δεν επηρεάζει την κατανομή της X_1 και συνεπώς η δεσμευμένη κατανομή συμπίπτει με την αρχική κατανομή.

Για διακριτές τυχαίες μεταβλητές η δεσμευμένη κατανομή της X_1 δοθέντος ότι $X_2 = x_2$ καθορίζεται από τη *δεσμευμένη συνάρτηση πιθανότητας*

$$\begin{aligned} p(x_1 \mid x_2) &= \mathbb{P}(X_1 = x_1 \mid X_2 = x_2) \\ &= \frac{\mathbb{P}(X_1 = x_1, X_2 = x_2)}{\mathbb{P}(X_2 = x_2)} \\ &= \frac{p(x_1, x_2)}{p_{X_2}(x_2)}, \end{aligned}$$

όπου $p(x_1, x_2), p_{X_2}(x_2)$ είναι οι συναρτήσεις πιθανότητας των $\underset{\sim}{X} = (X_1, X_2)$ και X_2 αντίστοιχα. Η δεσμευμένη συνάρτηση πιθανότητας πληροί τις ιδιότητες μίας συνάρτησης πιθανότητας, δηλαδή:

1. $p(x_1 \mid x_2) \geq 0, \; \forall \; x_1$.
2. $\sum_{x_1} p(x_1 \mid x_2) = \frac{\sum_{x_1} p(x_1, x_2)}{p_{X_2}(x_2)} = \frac{p_{X_2}(x_2)}{p_{X_2}(x_2)} = 1$.

Για συνεχείς τυχαίες μεταβλητές η δεσμευμένη κατανομή της X_1 δοθέντος ότι $X_2 = x_2$ καθορίζεται από τη *δεσμευμένη συνάρτηση πυκνότητας*

$$f(x_1 \mid x_2) = \frac{f(x_1, x_2)}{f_{X_2}(x_2)},$$

όπου $f(x_1, x_2), f_{X_2}(x_2)$ είναι οι συναρτήσεις πυκνότητας των $\underset{\sim}{X} = (X_1, X_2)$ και X_2 αντίστοιχα. Η δεσμευμένη συνάρτηση πυκνότητας πληροί τις ιδιότητες μίας συνάρτησης πυκνότητας, δηλαδή:

1. $f(x_1 \mid x_2) \geq 0, \; \forall \; x_1$.
2. $\int_{-\infty}^{\infty} f(x_1 \mid x_2) \, \mathrm{d}x_1 = \frac{\int_{-\infty}^{\infty} f(x_1, x_2) \, \mathrm{d}x_1}{f_{X_2}(x_2)} = \frac{f_{X_2}(x_2)}{f_{X_2}(x_2)} = 1$.

Η μέση τιμή και διασπορά που αντιστοιχούν στη δεσμευμένη κατανομή λέγονται *δεσμευμένη μέση τιμή* και *δεσμευμένη διασπορά* δοθέντος $X_2 = x_2$ και ορίζονται από τις σχέσεις:

$$\mu(x_2) = \mathbb{E}(X_1 \mid X_2 = x_2)$$
$$= \begin{cases} \sum_{x_1} x_1 p(x_1 \mid x_2), & \text{για διακριτό } \underset{\sim}{X} = (X_1, X_2), \\ \int_{-\infty}^{\infty} x_1 f(x_1 \mid x_2)\,\mathrm{d}x_1, & \text{για συνεχές } \underset{\sim}{X} = (X_1, X_2). \end{cases}$$

$$\sigma^2(x_2) = \mathbb{V}ar(X_1 \mid X_2 = x_2) = \mathbb{E}\big[\big(X_1 - \mu(x_2)\big)^2 \mid X_2 = x_2\big]$$
$$= \mathbb{E}(X_1^2 \mid X_2 = x_2) - \big(\mathbb{E}(X_1 \mid X_2 = x_2)\big)^2.$$

Πιο γενικά, η δεσμευμένη μέση τιμή της $g(X_1)$ δοθέντος ότι $X_2 = x_2$ ορίζεται από τη σχέση

$$\mathbb{E}\big(g(X_1) \mid X_2 = x_2\big)$$
$$= \begin{cases} \sum_{x_1} g(x_1) p(x_1 \mid x_2), & \text{για διακριτό } \underset{\sim}{X} = (X_1, X_2), \\ \int_{-\infty}^{\infty} g(x_1) f(x_1 \mid x_2)\,\mathrm{d}x_1, & \text{για συνεχές } \underset{\sim}{X} = (X_1, X_2). \end{cases}$$

Οι συναρτήσεις $\mu(x_2)$, $\sigma^2(x_2)$ ορίζουν τις τυχαίες μεταβλητές $\mu(X_2)$, $\sigma^2(X_2)$ που συμβολίζονται $\mathbb{E}(X_1 \mid X_2)$, $\mathbb{V}ar(X_1 \mid X_2)$ και λέγονται *δεσμευμένη μέση τιμή* και *δεσμευμένη διασπορά* της X_1 δοθείσης της X_2. Όμοια ορίζεται η $\mathbb{E}\big(g(X_1) \mid X_2\big)$. Τονίζουμε ότι οι τυχαίες μεταβλητές $\mathbb{E}(X_1 \mid X_2)$, $\mathbb{V}ar(X_1 \mid X_2)$ και $\mathbb{E}(g(X_1) \mid X_2)$ είναι συναρτήσεις της X_2, με τιμές, όταν $X_2 = x_2$, $\mathbb{E}(X_1 \mid X_2 = x_2)$, $\mathbb{V}ar(X_1 \mid X_2 = x_2)$, $\mathbb{E}(g(X_1) \mid X_2 = x_2)$, αντίστοιχα.

Πρόταση 1.9.1.

1. *Αν η X_1 είναι συνάρτηση της X_2, $X_1 = h(X_2)$, τότε $\mathbb{E}(X_1 \mid X_2) = X_1$.*

2. $\mathbb{V}ar(X_1 \mid X_2) \geqslant 0$.

3. *Αν* $\mathbb{V}ar(X_1 \mid X_2) = 0$, *τότε η* X_1 *είναι συνάρτηση της* X_2 *και αντίστροφα.*

Απόδειξη. 1. Για κάθε τιμή x_2 της X_2 έχουμε $\mathbb{E}(X_1 \mid X_2 = x_2) = \mathbb{E}(h(X_2) \mid X_2 = x_2) = \mathbb{E}(h(x_2) \mid X_2 = x_2) = h(x_2)$, λόγω της Πρότασης 1.4.2(1), επειδή η $h(x_2)$ είναι σταθερά. Άρα από τον ορισμό της $\mathbb{E}(X_1 \mid X_2)$, προκύπτει ότι $\mathbb{E}(X_1 \mid X_2) = h(X_2) = X_1$.
2. Για κάθε τιμή x_2 της X_2 έχουμε $\mathbb{V}ar(X_1 \mid X_2 = x_2) \geqslant 0$ ως διασπορά. Άρα από τον ορισμό της $\mathbb{V}ar(X_1 \mid X_2)$, προκύπτει ότι $\mathbb{V}ar(X_1 \mid X_2) \geqslant 0$.
3. Η σχέση $\mathbb{V}ar(X_1 \mid X_2) = 0$ ισοδύναμα σημαίνει $\mathbb{V}ar(X_1 \mid X_2 = x_2) = 0$, για κάθε τιμή x_2 της X_2. Επειδή η διασπορά της X_1 δοθέντος ότι $X_2 = x_2$ είναι 0, συμπεραίνουμε ότι η X_1 δοθέντος ότι $X_2 = x_2$ είναι σταθερά, εν γένει εξαρτώμενη από το x_2, έστω $h(x_2)$. Καταλήγουμε λοιπόν ότι η X_1 δοθέντος ότι $X_2 = x_2$ είναι ίση προς $h(x_2)$. Άρα, η X_1 είναι συνάρτηση της X_2, $X_1 = h(X_2)$. Αντίστροφα, αν $X_1 = h(X_2)$, τότε για κάθε τιμή x_2 της X_2 έχουμε $\mathbb{V}ar(X_1|X_2 = x_2) = \mathbb{V}ar(h(X_2)|X_2 = x_2) = \mathbb{V}ar(h(x_2)|X_2 = x_2) = 0$. Επομένως, $\mathbb{V}ar(X_1 \mid X_2) = 0$. ☐

Πρόταση 1.9.2.

1. $\mathbb{E}\big[\mathbb{E}(X_1 \mid X_2)\big] = \mathbb{E}X_1$ *και γενικότερα* $\mathbb{E}\big[\mathbb{E}(g(X_1) \mid X_2)\big] = \mathbb{E}g(X_1)$.

2. $\mathbb{V}arX_1 = \mathbb{E}\big(\mathbb{V}ar(X_1 \mid X_2)\big) + \mathbb{V}ar\big(\mathbb{E}(X_1 \mid X_2)\big)$ *και γενικότερα* $\mathbb{V}ar g(X_1) = \mathbb{E}\big(\mathbb{V}ar(g(X_1) \mid X_2)\big) + \mathbb{V}ar\big(\mathbb{E}(g(X_1) \mid X_2)\big)$.

Απόδειξη. (Για διακριτές τυχαίες μεταβλητές)

1. Έχουμε

$$\begin{aligned}
\mathbb{E}\big[\mathbb{E}(X_1 \mid X_2)\big] &= \mathbb{E}\big(\mu(X_2)\big) = \sum_{x_2} \mu(x_2) p_{X_2}(x_2) \\
&= \sum_{x_2} \Big(\sum_{x_1} x_1 p(x_1 \mid x_2) \Big) p_{X_2}(x_2) \\
&= \sum_{x_1} \sum_{x_2} x_1 p(x_1 \mid x_2) p_{X_2}(x_2) \\
&= \sum_{x_1} x_1 \sum_{x_2} p(x_1, x_2) \\
&= \sum_{x_1} x_1 p_{X_1}(x_1) = \mathbb{E}X_1.
\end{aligned}$$

2. Έχουμε

$$\begin{aligned}
&\mathbb{E}\big(\mathbb{V}ar(X_1 \mid X_2)\big) + \mathbb{V}ar\big(\mathbb{E}(X_1 \mid X_2)\big) \\
&\quad = \mathbb{E}\big\{\mathbb{E}(X_1^2 \mid X_2) - \big(\mathbb{E}(X_1 \mid X_2)\big)^2\big\} + \mathbb{V}ar\big(\mathbb{E}(X_1 \mid X_2)\big) \\
&\quad = \mathbb{E}\big(\mathbb{E}(X_1^2 \mid X_2)\big) - \mathbb{E}\big(\mu(X_2)\big)^2 + \mathbb{V}ar\big(\mathbb{E}(X_1 \mid X_2)\big) \\
&\quad = \mathbb{E}(X_1^2) - \mathbb{E}\big(\mu(X_2)\big)^2 + \mathbb{E}\big(\mu(X_2)\big)^2 - \big(\mathbb{E}\mu(X_2)\big)^2 \\
&\quad = \mathbb{E}(X_1^2) - \big\{\mathbb{E}\big(\mathbb{E}(X_1 \mid X_2)\big)\big\}^2 \\
&\quad = \mathbb{E}(X_1^2) - \big(\mathbb{E}X_1\big)^2 = \mathbb{V}arX_1.
\end{aligned}$$

□

1.10 Οριακά Θεωρήματα

Έστω $(\Omega, \mathscr{A}, \mathbb{P})$ ένας χώρος πιθανότητας και X_n, $n = 1, 2, \ldots$, μία ακολουθία τυχαίων μεταβλητών ορισμένων στο Ω. Ειδικά, η ακολουθία X_n, $n = 1, 2, \ldots$, είναι μία ακολουθία συναρτήσεων και συνεπώς μπορούμε να μελετήσουμε την οριακή συμπεριφορά της από την πλευρά της Ανάλυσης. Για παράδειγμα, αν υπάρχει συνάρτηση X ορισμένη στο Ω, τέτοια ώστε $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ για κάθε $\omega \in \Omega$, τότε λέμε ότι η ακολουθία X_n, $n = 1, 2, \ldots$, συγκλίνει κατά σημείο στην X. Εάν περαιτέρω $\lim_{n\to\infty} \sup_{\omega\in\Omega} |X_n(\omega) - X(\omega)| = 0$, τότε λέμε ότι η ακολουθία X_n, $n = 1, 2, \ldots$ συγκλίνει ομοιόμορφα στην X.

Πέραν των παραπάνω, στη Θεωρία Πιθανοτήτων και τη Στατιστική, οι επόμενες τρεις μορφές σύγκλισης διαδραματίζουν θεμελιακό ρόλο στην απόδειξη πληθώρας αποτελεσμάτων. Ειδικές περιπτώσεις αυτών των μορφών σύγκλισης είναι ο Ασθενής Νόμος των Μεγάλων Αριθμών, ο Ισχυρός Νόμος των Μεγάλων Αριθμών και το Κεντρικό Οριακό Θεώρημα, που αποτελούν ακρογωνιαίους λίθους της Ασυμπτωτικής Στατιστικής ή, όπως διαφορετικά αναφέρεται, της Θεωρίας Μεγάλων Δειγμάτων (Large Sample Theory), δηλαδή της Στατιστικής Ανάλυσης μεγάλου αριθμού δεδομένων. Αυτές οι μορφές σύγκλισης ορίζονται ως ακολούθως.

Ορισμός 1.10.1. *Η ακολουθία X_n, $n = 1, 2, \ldots$ συγκλίνει κατά πιθανότητα στην τυχαία μεταβλητή X, συμβολικά $X_n \xrightarrow{\mathbb{P}} X$, εάν*

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0 \ , \quad \forall\, \varepsilon > 0. \tag{1.4}$$

Η ερμηνεία της (1.4) είναι ότι η πιθανότητα να αποκλίνει (να διαφέρει) η X_n από την X περισσότερο από ε, συγκλίνει, καθώς $n \to \infty$, στο μηδέν όσο «μικρό» και αν είναι το ε. Η περίπτωση που η X είναι σταθερά, έστω c, και ισχύει $X_n \xrightarrow{\mathbb{P}} c$, εμφανίζεται συχνά στην Ασυμπτωτική Στατιστική και ταυτίζεται με την ιδιότητα της ασθενούς συνέπειας εκτιμητών που θα μελετήσουμε στην Ενότητα 7.2. Στη Θεωρία Πιθανοτήτων, η πλέον διάσημη περίπτωση σύγκλισης κατά πιθανότητα σε σταθερά είναι ο Ασθενής Νόμος των Μεγάλων Αριθμών.

Η πρώτη απόδειξη του έγινε από τον J. Bernoulli (1654 – 1705) στην ειδική περίπτωση που οι τυχαίες μεταβλητές X_n, $n = 1, 2, \ldots$ είναι ανεξάρτητες με κοινή δίτιμη κατανομή και τιμές 1 ή 0 (δηλαδή κατανομή Bernoulli!!!). Σε αυτή τη μορφή, ο Ασθενής Νόμος των Μεγάλων Αριθμών δικαιολογεί την ύπαρξη του ορίου στον εμπειρικό ορισμό πιθανότητας του von Mises, ως όριο της σχετικής συχνότητας. Ο Tchebychev (1821 – 1894) τον γενίκευσε για οποιαδήποτε κατανομή με πεπερασμένη διασπορά, ενώ ο Khintchine (1894 – 1959) τον απέδειξε με την υπόθεση, μόνον, πεπερασμένης μέσης τιμής.

Θεώρημα 1.10.1. (Ασθενής Νόμος των Μεγάλων Αριθμών – ΑΝΜΑ – Khintchine) *Έστω X_n, $n = 1, 2, \ldots$ μία ακολουθία τυχαίων μεταβλητών ανεξάρτητων και ισόνομων (δηλαδή, με κοινή κατανομή) με πεπερασμένη* *μέση τιμή* $\mathbb{E}X_1$. *Τότε ισχύει*

$$\frac{X_1 + \ldots + X_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}X_1.$$

Απόδειξη. Αν η κοινή διασπορά των X_n, σ^2, είναι πεπερασμένη, η απόδειξη βασίζεται στην ανισότητα Tchebychev. Θέτουμε $\bar{X}_n = \dfrac{X_1 + \ldots + X_n}{n}$ και παρατηρούμε ότι $\mathbb{E}\bar{X}_n = \mathbb{E}\left(\dfrac{X_1 + \ldots + X_n}{n}\right) = \dfrac{1}{n}\sum_{i=1}^n \mathbb{E}X_i = \dfrac{1}{n}\sum_{i=1}^n \mathbb{E}X_1 = \mathbb{E}X_1$ και $\mathbb{V}ar(\bar{X}_n) = \dfrac{1}{n^2}\mathbb{V}ar(X_1 + \ldots + X_n) = \dfrac{1}{n^2}\sum_{i=1}^n \mathbb{V}arX_i = \dfrac{1}{n^2}\sum_{i=1}^n \sigma^2 = \dfrac{\sigma^2}{n}$. Επομένως, από την ανισότητα Tchebychev (Πρόταση 1.5.1) παίρνουμε

$$\mathbb{P}(|\bar{X}_n - \mathbb{E}X_1| > \varepsilon) \leqslant \frac{\mathbb{V}ar(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \to 0 \text{ , καθώς } n \to \infty.$$

Για τη γενική περίπτωση παραπέμπουμε τον αναγνώστη στον Rohatgi (1976, σελ. 261). □

Στη Στατιστική ορολογία ο αριθμητικός μέσος $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ αναφέρεται ως δειγματικός μέσος και ο ΑΝΜΑ λέει ότι ο δειγματικός μέσος είναι ασθενώς συνεπής εκτιμητής της μέσης τιμής της κοινής κατανομής των X_n (Παράδειγμα 7.2.2).

Ορισμός 1.10.2. *Η ακολουθία* X_n, $n = 1, 2, \ldots$ *συγκλίνει με πιθανότητα 1 προς την τυχαία μεταβλητή* X, *συμβολικά* $X_n \xrightarrow{\mu.\pi.1} X$, *εάν*

$$\mathbb{P}(\left\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}) = 1. \tag{1.5}$$

Η ερμηνεία της (1.5) είναι ότι, με πιθανότητα 1, η ακολουθία X_n, $n = 1, 2, \ldots$ συγκλίνει κατά σημείο στην X (με την παραπάνω έννοια της Ανάλυσης). Ενώ η σύγκλιση κατά πιθανότητα έχει να κάνει με τη σύγκλιση της ακολουθίας των πιθανοτήτων $a_n = \mathbb{P}(|X_n - X| > \varepsilon)$, η σύγκλιση με πιθανότητα 1 αναφέρεται στην πιθανότητα του ενδεχομένου η ακολουθία των αριθμών $X_n(\omega)$ να συγκλίνει, απαιτώντας αυτό το ενδεχόμενο να έχει πιθανότητα 1. Η σύγκλιση με πιθανότητα 1 είναι πιο ισχυρή από (δηλαδή συνεπάγεται) την σύγκλιση κατά πιθανότητα (Πρόταση 1.10.4) και για αυτόν το λόγο αναφέρονται συχνά στη βιβλιογραφία ως ισχυρή και ασθενής σύγκλιση, αντίστοιχα. Η περίπτωση που η X είναι σταθερά, έστω c, και ισχύει $X_n \xrightarrow{\text{μ.π.1}} c$, εμφανίζεται συχνά στην Ασυμπτωτική Στατιστική και ταυτίζεται με την ιδιότητα της ισχυρής συνέπειας εκτιμητών που θα μελετήσουμε στην Ενότητα 7.2. Στη Θεωρία Πιθανοτήτων η πλέον διάσημη περίπτωση σύγκλισης με πιθανότητα 1 σε σταθερά είναι ο Ισχυρός Νόμος των Μεγάλων Αριθμών. Η απόδειξη του οφείλεται στον Kolmogorov.

Θεώρημα 1.10.2. (Ισχυρός Νόμος των Μεγάλων Αριθμών – ΙΝΜΑ – Kolmogorov).
Έστω X_n, $n = 1, 2, \ldots$ μία ακολουθία τυχαίων μεταβλητών ανεξάρτητων και ισόνομων με πεπερασμένη μέση τιμή $\mathbb{E}X_1$ (ενδεχομένως $+\infty$ ή $-\infty$). Τότε ισχύει

$$\frac{X_1 + \ldots + X_n}{n} \xrightarrow{\text{μ.π.1}} \mathbb{E}X_1.$$

Απόδειξη. Rohatgi (1976, σελ. 274), Chow and Teicher (1978, σελ. 122) . □

Στη Στατιστική ορολογία ο ΙΝΜΑ λέει ότι ο δειγματικός μέσος είναι ισχυρά συνεπής εκτιμητής της μέσης τιμής της κοινής κατανομής των X_n.

Οι παραπάνω συγκλίσεις αφορούν τις τιμές της ακολουθίας των τυχαίων μεταβλητών, X_n, $n = 1, 2, \ldots$. Η επόμενη σύγκλιση (ασθενέστερη, αλλά πολύ χρήσιμη στις εφαρμογές) έχει να κάνει με προσέγγιση πιθανοτήτων των X_n.

Ορισμός 1.10.3. *Έστω F_n, $n = 1, 2, \ldots$ η συνάρτηση κατανομής της τυχαίας μεταβλητής X_n, $n = 1, 2, \ldots$ αντίστοιχα και F η συνάρτηση κατανομής της τυχαίας μεταβλητής X. Η ακολουθία X_n, $n = 1, 2, \ldots$ συγκλίνει κατά κατανομή στην τυχαία μεταβλητή X, συμβολικά $X_n \xrightarrow{\mathcal{L}} X$, εάν*

$$\lim_{n\to\infty} F_n(x) = F(x). \tag{1.6}$$

για κάθε σημείο $x \in \Re$ που είναι σημείο συνέχειας της F.

Η ερμηνεία της (1.6) είναι ότι η πιθανότητα $\mathbb{P}(X_n \leqslant x) = F_n(x)$ συγκλίνει, καθώς, $n \to \infty$, προς την πιθανότητα $\mathbb{P}(X \leqslant x) = F(x)$. Συνεπώς για «μεγάλο» n ισχύει η προσέγγιση $\mathbb{P}(X_n \leqslant x) \approx \mathbb{P}(X \leqslant x)$ και ως εκ τούτου πιθανότητες που αφορούν την X_n μπορούν να προσεγγιστούν από αντίστοιχες πιθανότητες της X, π.χ. $\mathbb{P}(\alpha < X_n \leqslant \beta) = \mathbb{P}(X_n \leqslant \beta) - \mathbb{P}(X_n \leqslant \alpha) \approx \mathbb{P}(X \leqslant \beta) - \mathbb{P}(X \leqslant \alpha) = \mathbb{P}(\alpha < X \leqslant \beta)$. Η πλέον διάσημη περίπτωση σύγκλισης κατά κατανομή είναι το Κεντρικό Οριακό Θεώρημα, η κορωνίδα των οριακών θεωρημάτων της Θεωρίας Πιθανοτήτων. Η εξέλιξη του μέχρι να λάβει την πρώτη γενική μορφή του πέρασε από διάφορα στάδια (De Moivre 1733 για τυχαίες μεταβλητές Bernoulli, $\mathcal{B}(1, p)$ με $p = 1/2$, Laplace 1812 για κάθε $0 < p < 1$, Tchebychev, Markov, Liapounov 1900, Lindeberg 1922, Feller 1935). Στην κλασική του μορφή διατυπώνεται ως ακολούθως.

Θεώρημα 1.10.3. (Κεντρικό Οριακό Θεώρημα – ΚΟΘ – Tchebychev – Markov – Liapounov).
Έστω X_n, $n = 1, 2, \ldots$ μία ακολουθία ανεξάρτητων και ισόνομων τυχαίων μεταβλητών με $\mathbb{E}X_1 = \mu$ και $\mathbb{V}arX_1 = \sigma^2 < \infty$. Θεωρούμε την ακολουθία (των τυποποιημένων δειγματικών μέσων) $Z_n = \dfrac{\sqrt{n}(\bar{X} - \mu)}{\sigma}$, όπου $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, $n = 1, 2, \ldots$. Τότε ισχύει

$$Z_n \xrightarrow{\mathcal{L}} Z,$$

ή ισοδύναμα

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \xrightarrow{\mathcal{L}} Z,$$

όπου Z είναι μία τυχαία μεταβλητή με τυπική κανονική κατανομή $\mathcal{N}(0, 1)$, δηλαδή

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \leqslant x\right) = \Phi(x)\ , \quad \forall\, x \in \Re. \tag{1.7}$$

ή ισοδύναμα

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \leqslant x\right) = \Phi(x)\ , \quad \forall\, x \in \Re. \tag{1.8}$$

(Φ είναι η συνάρτηση κατανομής της $\mathcal{N}(0,1)$). Περαιτέρω, η σύγκλιση στις (1.7) *και* (1.8) *είναι ομοιόμορφη ως προς $x \in \Re$.*

Απόδειξη. Casella and Berger (2002, σελ. 236) για την περίπτωση που η ροπογεννήτρια είναι πεπερασμένη σε περιοχή του 0. Chow and Teicher (1978, σελ. 291), Rohatgi (1976, σελ. 282) για τη γενική περίπτωση. □

Η ερμηνεία του ΚΟΘ είναι ότι η τυχαία μεταβλητή $\dfrac{\sqrt{n}(\bar{X}-\mu)}{\sigma}$ έχει, για «μεγάλο» n, κατά προσέγγιση κατανομή $\mathcal{N}(0,1)$,

$$\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \overset{\mathcal{L}}{\approx} \mathcal{N}(0,1),$$

ή ισοδύναμα ότι

$$\bar{X}_n \overset{\mathcal{L}}{\approx} \mathcal{N}(\mu, \sigma^2/n), \tag{1.9}$$

δηλαδή ο δειγματικός μέσος, $\bar{X}_n$, ανεξάρτητων και ισόνομων τυχαίων μεταβλητών συμπεριφέρεται, κατά προσέγγιση, όπως μία κανονική τυχαία μεταβλητή $\mathcal{N}(\mu, \sigma^2/n)$ ή ισοδύναμα ότι

$$\sum_{i=1}^{n} X_i \overset{\mathcal{L}}{\approx} \mathcal{N}(n\mu, n\sigma^2), \tag{1.10}$$

δηλαδή ότι το άθροισμα $\sum_{i=1}^{n} X_i$ ανεξάρτητων και ισόνομων τυχαίων μεταβλητών συμπεριφέρεται, κατά προσέγγιση, όπως μία κανονική τυχαία μεταβλητή $\mathcal{N}(n\mu, n\sigma^2)$. Για παράδειγμα ισχύει η προσέγγιση

$$\mathbb{P}(\alpha \leqslant \sum_{i=1}^{n} X_i \leqslant \beta) \approx \Phi\left(\frac{\beta - n\mu}{\sqrt{n}\sigma}\right) - \Phi\left(\frac{\alpha - n\mu}{\sqrt{n}\sigma}\right).$$

Η ισχύς του ΚΟΘ είναι όντως εκπληκτική, αν λάβουμε υπ' όψη ότι καλύπτει τυχαίες μεταβλητές με αυθαίρετη κατανομή. Σημειώνουμε επίσης ότι το ΚΟΘ παρέχει την ταχύτητα σύγκλισης στον ΑΝΜΑ, $\bar{X}_n \xrightarrow{\mathbb{P}} \mu$, που είναι της τάξης $1/\sqrt{n}$. Ακόμη, στην περίπτωση που οι τυχαίες μεταβλητές X_n, $n = 1, 2, \ldots$ έχουν κανονική κατανομή $\mathcal{N}(\mu, \sigma^2)$, τότε οι (1.7), (1.9) και (1.10) είναι ακριβείς, ως γραμμικοί συνδυασμοί ανεξάρτητων κανονικών κατανομών (Πρόταση 1.8.6(4)).

Η σχέση μεταξύ των τριών συγκλίσεων δίνεται στην επόμενη πρόταση. Η πιο ισχυρή είναι η σύγκλιση με πιθανότητα 1 (που ουσιαστικά, όπως αναφέρθηκε, συμπίπτει με την κατά σημείο σύγκλιση της Ανάλυσης).

Πρόταση 1.10.4.
(i) Ισχύουν οι σχέσεις

$$X_n \xrightarrow{\mu.\pi.1} X \Rightarrow X_n \xrightarrow{\mathbb{P}} X \Rightarrow X_n \xrightarrow{\mathcal{L}} X.$$

(ii) Αν c είναι σταθερά και $X_n \xrightarrow{\mu.\pi.1} c$, τότε $X_n \xrightarrow{\mathbb{P}} c$.
(iii) Ο ΙΝΜΑ συνεπάγεται τον ΑΝΜΑ.
(iv) Αν c είναι σταθερά και $X_n \xrightarrow{\mathcal{L}} c$, τότε $X_n \xrightarrow{\mathbb{P}} c$.

Απόδειξη. (i) Rohatgi (1976, σελ. 246, 250).
(ii), (iii) προκύπτουν αμέσως από το (i).
(iv) Rohatgi (1976, σελ. 246). □

Μερικές ιδιότητες των τριών συγκλίσεων συνοψίζονται στην επόμενη πρόταση.

Πρόταση 1.10.5.
(i) Αν $X_n \xrightarrow{\mathbb{P}} X$ και g είναι συνεχής συνάρτηση, τότε $g(X_n) \xrightarrow{\mathbb{P}} g(X)$.
(ii) Αν $X_n \xrightarrow{\mu.\pi.1} X$ και g είναι συνεχής συνάρτηση, τότε $g(X_n) \xrightarrow{\mu.\pi.1} g(X)$.
(iii) Αν $X_n \xrightarrow{\mathbb{P}} X$ και $Y_n \xrightarrow{\mathbb{P}} Y$, τότε $X_n + Y_n \xrightarrow{\mathbb{P}} X + Y$, $X_nY_n \xrightarrow{\mathbb{P}} XY$, $X_n/Y_n \xrightarrow{\mathbb{P}} X/Y$ (υπό τη συνθήκη $\mathbb{P}(Y = 0) = 0$).
(iv) Αν $X_n \xrightarrow{\mu.\pi.1} X$ και $Y_n \xrightarrow{\mu.\pi.1} Y$, τότε $X_n + Y_n \xrightarrow{\mu.\pi.1} X + Y$, $X_nY_n \xrightarrow{\mu.\pi.1} XY$, $X_n/Y_n \xrightarrow{\mu.\pi.1} X/Y$ (υπό τη συνθήκη $\mathbb{P}(Y = 0) = 0$).

Απόδειξη. Rohatgi (1976, σελ. 244–245), Chow and Teicher (1978, σελ. 68, 73). □

Το επόμενο σημαντικό αποτέλεσμα χρησιμοποιείται συχνά στην Στατιστική για την εύρεση ασυμπτωτικών κατανομών.

Θεώρημα 1.10.6. (Slutsky)
(i) Αν $X_n - Y_n \xrightarrow{\mathbb{P}} 0$ και $Y_n \xrightarrow{\mathcal{L}} X$, τότε $X_n \xrightarrow{\mathcal{L}} X$.

(ii) Αν $X_n \xrightarrow{\mathcal{L}} X$, $Y_n \xrightarrow{\mathbb{P}} a$, $Z_n \xrightarrow{\mathbb{P}} b$, *τότε* $Y_n X_n + Z_n \xrightarrow{\mathcal{L}} aX + b$.
(iii) Αν $X_n \xrightarrow{\mathcal{L}} X$, $a_n \to a$, $b_n \to b$, *όπου* a_n, b_n, $n = 1, 2, \ldots$ *είναι ακολουθίες πραγματικών αριθμών, τότε* $a_n X_n + b_n \xrightarrow{\mathcal{L}} aX + b$.
(iv) Αν $X_n \xrightarrow{\mathcal{L}} X$, $Z_n \xrightarrow{\mathbb{P}} 0$, *τότε* $X_n + Z_n \xrightarrow{\mathcal{L}} X$.
(v) Αν $X_n \xrightarrow{\mathcal{L}} X$ *και* $Y_n \xrightarrow{\mathbb{P}} 0$, *τότε* $X_n Y_n \xrightarrow{\mathbb{P}} 0$.

Απόδειξη. (i) Chow and Teicher (1978, σελ. 249).
(ii) Chow and Teicher (1978, σελ. 249).
(iii) Θέτουμε $Y_n = a_n$ και $Z_n = b_n$. Τότε $Y_n \xrightarrow{\mathbb{P}} a$, $Z_n \xrightarrow{\mathbb{P}} b$ και εφαρμόζουμε το (ii).
(iv) Προκύπτει αμέσως από τα (ii) και (iii).
(v) Προκύπτει αμέσως από το (ii) και την Πρόταση 1.10.4(iv). □

Μία ακολουθία X_n, $n = 1, 2, \ldots$ είναι φραγμένη κατά πιθανότητα, συμβολικά γράφουμε $X_n = O_{\mathbb{P}}(1)$, αν για κάθε $\varepsilon > 0$ υπάρχουν M_ε και n_ε τέτοια, ώστε $\mathbb{P}(|X_n| < M_\varepsilon)$ για $n > n_\varepsilon$.

Από τον ορισμό και τους αντίστοιχους της σύγκλισης κατά πιθανότητα και της σύγκλισης κατά κατανομή προκύπτουν εύκολα τα εξής αποτελέσματα. Ειδικά, το τελευταίο εξ αυτών παραπέμπει στη γνωστή ιδιότητα ακολουθιών πραγματικών αριθμών ότι «το γινόμενο μηδενικής ακολουθίας με φραγμένη ακολουθία είναι μηδενική ακολουθία».

Πρόταση 1.10.7.
(i) $Y_n \xrightarrow{\mathcal{L}} Y \Rightarrow Y_n = O_{\mathbb{P}}(1)$.
(ii) $Y_n \xrightarrow{\mathbb{P}} c$, *όπου* c *σταθερά* $\Rightarrow Y_n = O_{\mathbb{P}}(1)$.
(iii) $|Z_n| \leqslant Y_n$ *και* $Y_n = O_{\mathbb{P}}(1) \Rightarrow Z_n = O_{\mathbb{P}}(1)$.
(iv) $X_n \xrightarrow{\mathbb{P}} 0$ *και* $Z_n = O_{\mathbb{P}}(1) \Rightarrow Z_n X_n \xrightarrow{\mathbb{P}} 0$.

Βιβλιογραφία

1. Bickel, P.J. and Doksum, K.A. (1977). *Mathematical Statistics, Basic Ideas and Selected Topics, Vol. 1.* Holden-Day.

2. Casella, G. and Berger, R.L. (2002). *Statistical Inference.* Duxbury Press; 2nd edition.

3. Chow, Y. and H. Teicher (1978). *Probability theory.* Springer-Verlag, New York.

4. Feller, W. (1935). Uber den Zentralen Grenzwertsatz der wahrscheinlichkeitsrechnung. *Math, Zeit.*, **40**, 521–559.

5. Lindeberg, J.W. (1922). Eine neue Herleitung des Exponentialgesetz in der Wahrscheinlichkeitsrechnung. *Math. Zeit.*, **15**, 211–225.

6. Rohatgi, V.K. (1976). *An Introduction to Probability Theory and Mathematical Statistics.* John Wiley and Sons, New York.

Κεφάλαιο 2

Γενικά περί Στατιστικής Συμπερασματολογίας

Σε πάρα πολλούς τομείς της ανθρώπινης δραστηριότητας δεδομένα συλλέγονται και αναλύονται με σκοπό την απόκτηση γνώσης, την εξαγωγή συμπερασμάτων και σε πολλές περιπτώσεις τη λήψη αποφάσεων. Ενδεικτικά παραδείγματα αποτελούν:

- Η πρόβλεψη εκλογικών αποτελεσμάτων.
- Ο έλεγχος ποιότητας ενός προϊόντος, πριν διατεθεί μαζικά στην αγορά.
- Η σύγκριση δύο ανταγωνιστικών «προϊόντων», όπως θεραπευτικών αγωγών, φαρμάκων, μεθόδων διδασκαλίας.
- Η αναζήτηση ποιοτικής ή ποσοτικής σχέσης (μαθηματικού μοντέλου) μεταξύ καθορισμένων επιβαρυντικών παραγόντων και συγκεκριμένης ασθένειας.

Σε αδρές γραμμές, *Στατιστική* είναι η επιστήμη που ασχολείται με τη συλλογή δεδομένων, την ανάλυση τους και την ερμηνεία τους. Στα αρχικά στάδια της ανάπτυξης της η Στατιστική περιοριζόταν στην απλή καταγραφή στοιχείων και τη συνοπτική παρουσίαση τους μέσω πινάκων, σχεδιαγραμμάτων και υπολογισμών απλών δεικτών (μέσοι όροι, ποσοστά κ.λ.π.). Αρχικά λοιπόν η Στατιστική είχε εμπειρικό (περιγραφικό) χαρακτήρα (*Περιγραφική Στατιστική*). Από τις αρχές του 20ου αιώνα όμως, με την ανάπτυξη και τη μαθηματική θεμελίωση της Θεωρίας Πιθανοτήτων, η Στατιστική – εκτός από την αρχική περιγραφική της μορφή (που εξακολουθεί να διατηρεί) – άρχισε παράλληλα να προσλαμβάνει αυστηρή και μαθηματική μορφή (*Μαθηματική Στατιστική*) και να αναδεικνύεται ως ένας νέος αυτοδύναμος κλάδος των Θετικών Επιστημών. Κατά τα τελευταία είκοσι χρόνια περίπου, σε συνδυασμό με τη διαθέσιμη και ολοένα αυξανόμενη ισχύ των υπολογιστών, έχει παρατηρηθεί μεγάλη ανάπτυξη της Στατιστικής προς την κατεύθυνση της θεωρίας και μεθοδολογίας ανάλυσης μεγάλου έως τεράστιου μεγέθους δεδομένων (Big Data Statistics).

Σήμερα η Στατιστική (Περιγραφική ή Μαθηματική) έχει εφαρμογές στις οικονομικές επιστήμες (Economic and Business Statistics), στις φυσικές επιστήμες (Physical and Engineering Sciences Statistics), στις κοινωνικές επιστήμες (Social Statistics), στις επιστήμες υγείας (Medical Statistics, Biostatistics, Statistical Genetics), στη Βιομηχανία (Industrial Statistics), στην Ψυχολογία (Psychometrics), στη Λογοτεχνία (Stylometry, the statistical analysis of literary style), και αλλού.

Στη Μαθηματική Στατιστική τα προς ανάλυση δεδομένα θεωρούνται τιμές τυχαίων μεταβλητών, έχουν δηλαδή προκύψει από ένα *πείραμα τύχης* (random experiment, Ενότητα 1.1) σύμφωνα με κάποια *κατανομή πιθανότητας*. Ανάλογα με το υπό μελέτη φυσικό πρόβλημα, αυτή η κατανομή πιθανότητας είναι εντελώς ή εν μέρει άγνωστη. (στο Παράδειγμα 2.1 παρακάτω η κατανομή είναι εν μέρει άγνωστη.) Ο πυρήνας της Μαθηματικής Στατιστικής είναι η *Στατιστική Συμπερασματολογία*, που έχει ως αντικείμενο την εξαγωγή συμπερασμάτων από τα δεδομένα, για μία παράμετρο (κάποιο χαρακτηριστικό) της προαναφερθείσης κατανομής πιθανότητας. Αυτή η παράμετρος έχει άγνωστη τιμή, γι' αυτό θα αναφέρεται στη συνέχεια ως άγνωστη παράμετρος και συνήθως, στην πράξη, έχει φυσική ερμηνεία, δηλαδή μπορεί να είναι π.χ. η μέση τιμή, η διασπορά, ένα ποσοστό της κατανομής ή ακόμη και η ίδια η συνάρτηση κατανομής. Στην τελευταία περίπτωση, τα συμπεράσματα αφορούν την ίδια την άγνωστη κατανομή πιθανότητας. Τέλος τα συμπεράσματα αυτά μπορούν εν συνεχεία να χρησιμοποιηθούν για τη λήψη αποφάσεων στο υπό μελέτη φυσικό πρόβλημα.

Η Στατιστική Συμπερασματολογία (Σ.Σ.) περιλαμβάνει κυρίως τρεις κλάδους.

1. *Εκτιμητική ή Σημειοεκτιμητική* (εκτίμηση με μία τιμή, ένα σημείο). Είναι ο κλάδος της Σ.Σ. που ασχολείται με μεθόδους εκτίμησης της άγνωστης παραμέτρου. Η εκτίμηση γίνεται με τον προσδιορισμό μιας κατάλληλης συνάρτησης, *του εκτιμητή*, οι τιμές του οποίου αναμένουμε να είναι «κοντά» στην τιμή της άγνωστης παραμέτρου.

2. *Διαστήματα Εμπιστοσύνης.* Είναι ο κλάδος της Σ.Σ. που ασχολείται με τον προσδιορισμό διαστήματος (περιοχής γενικότερα) που περιέχει με «μεγάλη» πιθανότητα την άγνωστη παράμετρο. Αντί λοιπόν μιας τιμής που παρέχει ένας εκτιμητής, δίνεται ένα διάστημα τιμών για την άγνωστη παράμετρο. Σε πολλές περιπτώσεις, ένα διάστημα εμπιστοσύνης παρέχει, επί πλέον, πληροφορίες για το σφάλμα της εκτίμησης της άγνωστης παραμέτρου από ένα συγκεκριμένο εκτιμητή.

3. *Έλεγχος Στατιστικών Υποθέσεων.* Είναι ο κλάδος της Σ.Σ. που ασχολείται με τον έλεγχο ισχυρισμών (υποθέσεων) για την τιμή της άγνωστης παραμέτρου, η οποία, όπως ήδη αναφέρθηκε, μπορεί να είναι ακόμη και η συνάρτηση κατανομής. Σε αυτήν την περίπτωση, ο ισχυρισμός μπορεί να είναι ότι τα δεδομένα έχουν προκύψει από («υπακούουν» σε) μια κατανομή πιθανότητας, συγκεκριμένης παραμετρικής μορφής.

Για να εξηγήσουμε καλύτερα τα παραπάνω, ας εξετάσουμε ένα παράδειγμα.

Παράδειγμα 2.1. Έστω ότι μια βιομηχανία, προκειμένου να αποφασίσει, εάν θα θέσει σε μαζική παραγωγή ένα νέο τύπο ηλεκτρικών λαμπτήρων, ενδιαφέρεται να μελετήσει προκαταβολικά τον μέσο χρόνο ζωής τους, θ. Καθώς το νέο αυτό προϊόν δεν έχει ακόμα παραχθεί και δοκιμαστεί, ο μέσος χρόνος ζωής θ είναι άγνωστος. Από προηγούμενη εμπειρία η

Βιομηχανία θεωρεί ότι ο χρόνος ζωής X ενός τέτοιου λαμπτήρα μπορεί να περιγραφεί ικανοποιητικά από μια τυχαία μεταβλητή που ακολουθεί την εκθετική κατανομή. (Μία συζήτηση για την υιοθέτηση της εκθετικής κατανομής γίνεται στο τέλος του κεφαλαίου.) Επειδή από τον ορισμό του θ έχουμε $\mathbb{E}X = \theta$, η πυκνότητα της τυχαίας μεταβλητής X είναι $f(x;\theta) = (1/\theta)e^{-x/\theta}$, $x > 0$, $\theta > 0$. Το θ είναι στην προκειμένη περίπτωση η άγνωστη παράμετρος. Συμπεράσματα για το θ μπορούν να εξαχθούν, εκτελώντας ένα απλό πείραμα τύχης: δείγμα n λαμπτήρων τίθεται σε λειτουργία και για καθέναν καταγράφεται (παρατηρείται) ο χρόνος ζωής του, έστω, $x_i, i = 1, 2, \ldots, n$. Οι αριθμητικές τιμές $x_1, x_2, \ldots, x_n$ αποτελούν τα προς ανάλυση δεδομένα. Έστω ακόμη $X_i, i = 1, 2, \ldots, n$, η τυχαία μεταβλητή που εκφράζει το χρόνο ζωής του i-οστού λαμπτήρα, πριν την εκτέλεση του πειράματος. Τα $x_1, x_2, \ldots, x_n$ είναι λοιπόν, αντίστοιχα, οι *παρατηρηθείσες* τιμές των τυχαίων μεταβλητών $X_1, X_2, \ldots, X_n$, η κατανομή των οποίων, δηλαδή η παραπάνω εκθετική κατανομή, είναι εν μέρει άγνωστη, γιατί έχει μεν δεδομένο μαθηματικό τύπο, εξαρτάται όμως από την άγνωστη παράμετρο θ.
Ας δούμε τώρα τι μπορεί να σημαίνει καθένας από τους τρεις αυτούς κλάδους στο συγκεκριμένο παράδειγμα.

1. Λόγω της φυσικής ερμηνείας του θ ως μέσου χρόνου ζωής όλων των λαμπτήρων, η συνάρτηση $T(X_1, \ldots, X_n) = (X_1 + \cdots + X_n)/n$, δηλαδή ο μέσος χρόνος ζωής των λαμπτήρων του δείγματος, είναι λογικό να ληφθεί ως εκτιμητής του θ. Η τιμή του εκτιμητή, $T(x_1, \ldots, x_n) = (x_1 + \cdots + x_n)/n$, για τις παρατηρηθείσες τιμές $x_1, \ldots, x_n$ των $X_1, \ldots, X_n$, ανακοινώνεται ως εκτίμηση (της τιμής) του θ. Αυστηρά, η χρησιμοποίηση του $T(\underset{\sim}{X})$ μπορεί να δικαιολογηθεί ως εξής. Επειδή από τον Ισχυρό Νόμο των Μεγάλων Αριθμών, Θεώρημα 1.10.2,
$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{\text{μ.π.1}} \mathbb{E}_\theta X_1 = \theta, \text{ καθώς } n \to \infty,$$
αναμένουμε η εκτίμηση $(x_1 + \cdots + x_n)/n$ να είναι «κοντά» στην άγνωστη τιμή του θ, τουλάχιστον για «μεγάλο» n.

2. Ένα διάστημα εμπιστοσύνης για το θ είναι το διάστημα της μορφής $(T - c_1, T + c_2)$, όπου c_1, c_2 θετικές, γνωστές, σταθερές και T ο παραπάνω εκτιμητής. Σημειώνουμε ότι το διάστημα ορίστηκε με το σκεπτικό να περιέχει τον εκτιμητή. Εάν π.χ. $(x_1 + \cdots + x_n)/n = 720$ ώρες και $c_1 = c_2 = 40$, τότε η εκτίμηση του θ είναι 720 ώρες με διάστημα εμπιστοσύνης (όρια κύμανσης του θ) $(680,\ 760)$. Οι σταθερές c_1 και c_2 εδώ ελήφθησαν αυθαίρετα, για χάρη απλότητας, γενικά όμως, ο προσδιορισμός τους αποτελεί μέρος της κατασκευής του διαστήματος εμπιστοσύνης. Εναλλακτικά το διάστημα αυτό μπορεί να δοθεί στη μορφή 720 ± 40, δηλώνοντας έτσι ότι το (μέγιστο) σφάλμα της εκτίμησης είναι (κατ' εκτίμηση) 40 ώρες. Ένα άλλο διάστημα εμπιστοσύνης για το θ είναι το διάστημα της μορφής (c_3T, c_4T), όπου $0 < c_3 < 1 < c_4$ είναι γνωστές (προσδιορίσιμες) σταθερές. Η συνθήκη, την οποία ικανοποιούν οι σταθερές, εξασφαλίζει ξανά τη λογική απαίτηση, ώστε το διάστημα να περιέχει τον εκτιμητή T.

3. Η Βιομηχανία ισχυρίζεται ότι ο μέσος χρόνος ζωής των λαμπτήρων είναι τουλάχιστον θ_0, δηλαδή $\theta \geq \theta_0$, όπου θ_0 είναι κάποια γνωστή σταθερά. Ο έλεγχος αυτού του ισχυρισμού μπορεί να γίνει χρησιμοποιώντας την εκτίμηση του θ, $T(x_1, \ldots, x_n) = (x_1 + \cdots + x_n)/n$. Ο ισχυρισμός μπορεί να απορριφθεί, εάν π.χ. η εκτίμηση $T(x_1, \ldots, x_n)$ είναι «αρκετά» μικρότερη του θ_0.

Σημειώνουμε ότι, αν, αντί για εκθετική, η κατανομή του χρόνου ζωής των λαμπτήρων θεωρηθεί εντελώς άγνωστη, η συνάρτηση $T(X_1, \ldots, X_n) = (X_1 + \cdots + X_n)/n$ μπορεί πάλι να θεωρηθεί ως εκτιμητής του θ, λόγω της φυσικής ερμηνείας του θ και του ΙΝΜΑ. Όμως τότε απολαμβάνει μόνον όσες ιδιότητες δεν συνδέονται με την εκθετική κατανομή, π.χ. την ιδιότητα της αμεροληψίας (Πρόταση 4.2.3) και την ιδιότητα της συνέπειας (Ενότητα 7.2). Υιοθετώντας την εκθετική κατανομή ο $T(\underset{\sim}{X})$ είναι, επί πλέον, αποδοτικός εκτιμητής (Παράδειγμα 5.2.8).

Από το Παράδειγμα 2.1 γίνεται φανερό ότι η Σ.Σ. έχει *επαγωγικό* χαρακτήρα (Επαγωγική Στατιστική), αφού χρησιμοποιεί το δείγμα, δηλαδή

Σχήμα 2.1: Στατιστική και Θεωρία Πιθανοτήτων

ένα μέρος του υπό μελέτη «πληθυσμού», για να εξάγει συμπεράσματα για (ολόκληρο) τον «πληθυσμό». Στο Παράδειγμα 2.1, ο «πληθυσμός» είναι το σύνολο των υπό παραγωγή λαμπτήρων. Συμπεράσματα για το άγνωστο πιθανοθεωρητικό μοντέλο (πυκνότητα) του «πληθυσμού», $f(x;\theta) = (1/\theta)e^{-x/\theta}$, $x > 0$, μπορούν να εξαχθούν εκτιμώντας το με το μοντέλο $f(x;\hat{\theta}) = (1/\hat{\theta})e^{-x/\hat{\theta}}$, $x > 0$, όπου $\hat{\theta} = T(x_1, \ldots, x_n) = (x_1 + \cdots + x_n)/n$. Αντίθετα, στη Θεωρία Πιθανοτήτων το πιθανοθεωρητικό μοντέλο του πληθυσμού θεωρείται πλήρως γνωστό ή δεδομένο και επομένως συμπεράσματα μπορούν να εξαχθούν (από τον πληθυσμό) για οποιοδήποτε μέρος (δείγμα) αυτού. Αν, για παράδειγμα, γνωρίζαμε ότι $\theta = 700$(ώρες), το πιθανοθεωρητικό μοντέλο (η πυκνότητα) του χρόνου ζωής των λαμπτήρων θα ήταν $f(x;700) = \frac{1}{700}e^{-x/700}$, $x > 0$, οπότε, π.χ., η πιθανότητα ένας λαμπτήρας να έχει διάρκεια ζωής μεγαλύτερη από 1000 ώρες θα υπολογιζόταν ως $\mathbb{P}(X > 1000) = \int_{1000}^{\infty} \frac{1}{700}e^{-x/700}dx = e^{-10/7} \simeq 0.24$. Έτσι, η Θεωρία Πιθανοτήτων έχει *απαγωγικό* χαρακτήρα (βλ. Σχήμα 2.1).

Στο Παράδειγμα 2.1, η συναρτησιακή μορφή του πιθανοθεωρητικού μοντέλου (της πυκνότητας), $f(x;\theta)$ είναι δεδομένη, και άγνωστη είναι μόνον η τιμή της παραμέτρου θ. Σε αυτές τις περιπτώσεις η Στατιστική Συμπερασματολογία αναφέρεται ως *Παραμετρική Στατιστική Συμπερασματολογία*. Εύλογα όμως μπορεί να τεθεί το ερώτημα: πάνω σε ποια βάση, υποθέτουμε ότι η πυκνότητα ενός πληθυσμού έχει συγκεκριμένη παραμετρική μορφή; Ας μη λησμονούμε ότι πολλά παραμετρικά μοντέλα κατανομών πιθανότητας είναι προϊόντα βασικής έρευνας με κίνητρο την ερμηνεία και την πρόβλεψη της συμπεριφοράς αντίστοιχων τυχαίων φαινομένων. Η αναζήτηση τέτοιων μοντέλων είναι σε ένα μεγάλο βαθμό το αντικείμενο της Στοχαστικής Μοντελοποίησης, ως κλάδου της Θεωρίας Πιθανοτήτων.

Για παράδειγμα, δεδομένα ορισμένων τροχαίων ατυχημάτων ή εκπομπών ραδιενεργών σωματιδίων από ραδιενεργό υλικό (είναι γνωστό ότι) μπορούν να μοντελοποιηθούν μέσω της κατανομής Poisson $\mathcal{P}(\theta)$, $\theta > 0$. Επίσης, δεδομένα που αφορούν διάρκεια ζωής, π.χ. ηλεκτρονικών εξαρτημάτων, περιγράφονται σε ορισμένες περιπτώσεις ικανοποιητικά από την εκθετική $\mathcal{E}(\theta)$, $\theta > 0$ ή την Γάμμα $\mathcal{G}(\alpha, \beta)$, $\theta = (\alpha, \beta)$ ή την Weibull κατανομή, $W(\alpha, \beta)$, $\theta = (\alpha, \beta)$. Δεδομένα, όπως ύψους ή βάρους πληθυσμού, που προκύπτουν ως αθροιστικό (συσσωρευτικό) αποτέλεσμα πολλών παραγόντων, γενετικών, περιβαλλοντικών κλπ, μπορούν να προσεγγιστούν από την κανονική κατανομή $\mathcal{N}(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2)$, λόγω του Κεντρικού Οριακού Θεωρήματος. Περαιτέρω, σε δημοσκόπηση πρόθεσης ψήφου, χρησιμοποιώντας ένα μεγάλο και αντιπροσωπευτικό (τυχαίο) δείγμα n ψηφοφόρων, δεν θα υπήρχε, μάλλον, καμία αντίρρηση ο αριθμός X των ψήφων συγκεκριμένου κόμματος σε επικείμενες βουλευτικές εκλογές να προσεγγιστεί από τη διωνυμική κατανομή, $\mathcal{B}(n, \theta)$, όπου θ είναι το ποσοστό των ψήφων του κόμματος στο σύνολο του εκλογικού σώματος. Σε τελική ανάλυση, η υπόθεση πυκνότητας δοθείσης παραμετρικής μορφής μπορεί να ελεγχθεί με μεθοδολογίες του *Ελέγχου Στατιστικών Υποθέσεων.* Υπάρχουν (επίσης) περιπτώσεις κατά τις οποίες, με βάση προηγούμενη ή άλλη γνώση, μπορεί να τεκμηριωθεί η υιοθέτηση ενός γνωστού πιθανοθεωρητικού μοντέλου που περιέχει όμως κάποια άγνωστη παράμετρο.

Εάν όμως η συναρτησιακή μορφή του πιθανοθεωρητικού μοντέλου του πληθυσμού είναι εντελώς άγνωστη, τότε η Στατιστική Συμπερασματολογία αναφέρεται ως *Απαραμετρική ή Μη Παραμετρική Στατιστική Συμπερασματολογία.* Ενδεικτικά, οι Προτάσεις 4.2.3, 4.2.4, 4.2.5 και 4.2.6, καθώς και η Ενότητα 4.3 αφορούν σε αυτόν τον κλάδο της Στατιστικής Συμπερασματολογίας. Το ηλεκτρονικό αυτό βοήθημα πραγματεύεται κυρίως θέματα Παραμετρικής Στατιστικής Συμπερασματολογίας.
Τα θέματα αυτά είναι το αντικείμενο μελέτης πολλών ελληνικών και ξένων συγγραμμάτων. Στη βιβλιογραφία που ακολουθεί παρατίθενται ορισμένα από αυτά στα οποία ο αναγνώστης μπορεί να ανατρέξει.

Βιβλιογραφία

1. Bickel, P.J. and Doksum, K.A. (1977). *Mathematical Statistics: Basic Ideas and Selected Topics*. San Francisco: Holden-Day.

2. Casella, G. and Berger, R.L. (2002). *Statistical Inference*. Duxbury Press; 2nd edition.

3. DeGroot, M.H. and Schervish, M.J. (2010). *Probability and Statistics*. Pearson Education; 4th edition.

4. Hogg, R.V., McKean, J. and Craig, A.T. (2004). *Introduction to Mathematical Statistics*. Pearson; 6th edition.

5. Larsen, R.J. and Marx, M.L. (2012). *An Introduction to Mathematical Statistics and Its Applications*. Pearson; 5th edition.

6. Lehmann, E.L. and Casella, G. (1998). *Theory of point estimation*. Springer; 2nd edition.

7. Mood, A.M., Graybill, F.A. and Boes, D.C. (1974). *Introduction to the Theory of Statistics*. McGraw-Hill; 3rd edition.

8. Rice, J.A. (1995). *Mathematical Statistics and Data Analysis*. Duxbury Press; 2nd edition.

9. Rohatgi, V.K. (1976). *An Introduction to Probability Theory and Mathematical Statistics*. John Wiley and Sons, New York.

10. Roussas, G.G. (1997). *A Course in Mathematical Statistics*. Academic Press; 2nd edition.

11. Roussas, G.G. (2003). *An Introduction to Probability and Statistical Inference*. Academic Press; 1st edition.

12. Samaniego, F.J. (2014). *Stochastic Modeling and Mathematical Statistics: A Text for Statisticians and Quantitative Scientists*.Chapman and Hall.

13. Δαμιανού, Χ. και Κούτρας Μ. (2003). *Εισαγωγή στη Στατιστική. Μέρος Ι.* Σ. Αθανασόπουλος & Σια Ο.Ε.

14. Ηλιόπουλος, Γ. (2013). *Βασικές Μέθοδοι Εκτίμησης Παραμέτρων.* Εκδόσεις Σταμούλη; 2η έκδοση.

15. Κάκουλλος, Θ. Ν. (1972). *Στατιστική Θεωρία και Εφαρμογαί.* Αθήνα.

16. Κολυβά - Μαχαίρα, Φ. (1998). *Μαθηματική Στατιστική. Τόμος Ι: Εκτιμητική.* Ζήτη Πελαγία & Σια Ο.Ε.

17. Παπαϊωάννου, Π. και Φερεντίνος, Κ. (2001). *Μαθηματική Στατιστική, Εκτιμητική - Έλεγχος Υποθέσεων - Εφαρμογές.* Σταμούλης; 2η έκδοση.

18. Ρούσσας, Γ. (1994). *Στατιστική συμπερασματολογία, Τόμος Ι: Εκτιμητική.* Ζήτη Πελαγία & Σια Ο.Ε.; 2η έκδοση.

Κεφάλαιο 3

Γενικά περί Εκτιμητικής

Στο κεφάλαιο αυτό παρουσιάζουμε το πρόβλημα της Εκτιμητικής εισάγοντας συγχρόνως τη σχετική ορολογία. Επιπλέον, σκιαγραφούμε ορισμένα σημαντικά κριτήρια σύγκρισης και αξιολόγησης εκτιμητών και μελετάμε εν συντομία μερικές απλές μεθόδους εκτίμησης.

3.1 Περιγραφή του προβλήματος

Έχοντας ως οδηγό το Παράδειγμα 2.1, μπορούμε να περιγράψουμε ένα πρόβλημα Εκτιμητικής ως εξής.

1. Σε έναν πληθυσμό (σύνολο μονάδων), υπάρχει κάποιο χαρακτηριστικό, θ, με μοναδική, αλλά άγνωστη τιμή, η οποία ανήκει σε ένα γνωστό σύνολο Θ. Το θ λέγεται *άγνωστη παράμετρος* και το Θ *παραμετρικός χώρος* . Μας ενδιαφέρει να εκτιμήσουμε την άγνωστη τιμή του θ ή γενικότερα την (άγνωστη) τιμή $g(\theta)$, όπου $g\colon \Theta \to \mathbb{R}^m$ είναι δεδομένη συνάρτηση. (στο Παράδειγμα 2.1 $\Theta = \{\theta : \theta > 0\}$ και $g(\theta) = \theta$.) Για το Θ υποθέτουμε ότι αυτό έχει τουλάχιστον δύο στοιχεία, διαφορετικά το θ είναι γνωστό, ενώ για την g, υποθέτουμε ότι δεν είναι η σταθερή συνάρτηση, διαφορετικά, αν π.χ. $g(\theta) = 5$, $\forall\theta \in \Theta$, δεν υπάρχει άγνωστη τιμή $g(\theta)$ προς εκτίμηση. Επίσης, για την g συμβατικά θα θεωρούμε ότι είναι πραγματική συνάρτηση ($m = 1$). Στις περιπτώσεις όμως που δεν είναι ($m \geqslant 2$), θα φαίνεται αυτό από τον ορισμό της.

2. Έχουμε στη διάθεση μας *δεδομένα* $\underset{\sim}{x} = (x_1, \ldots, x_n)$ τα οποία είναι *παρατηρηθείσες τιμές* τυχαίων μεταβλητών (ή τυχαίων διανυσμάτων) $X_1, \ldots, X_n$ με από κοινού πυκνότητα στο σημείο $\underset{\sim}{x} = (x_1, \ldots, x_n)$ $f(\underset{\sim}{x}; \theta)$ που εξαρτάται από την άγνωστη παράμετρο θ. Από μαθηματικής σκοπιάς, λοιπόν, η αλληλοεξάρτηση δεδομένων και άγνωστης παραμέτρου δηλώνεται μέσω της πυκνότητας $f(\underset{\sim}{x}; \theta)$ που είναι συνάρτηση του θ (πέραν του $\underset{\sim}{x}$). Καταχρηστικά, και οι τυχαίες μεταβλητές $X_1, \ldots, X_n$ και το τυχαίο διάνυσμα $\underset{\sim}{X} = (X_1, \ldots, X_n)$ αναφέρονται ως *δεδομένα ή παρατηρήσεις ή δείγμα*. Εάν οι παρατηρήσεις $X_1, \ldots, X_n$ είναι ανεξάρτητες τυχαίες μεταβλητές με την ίδια κατανομή πιθανότητας, τότε λέμε ότι αποτελούν ένα *τυχαίο δείγμα* από αυτήν την κατανομή. Ο αριθμός των δεδομένων, n, λέγεται *μέγεθος του δείγματος*. (Στο Παράδειγμα 2.1, η πυκνότητα του $\underset{\sim}{X} = (X_1, \ldots, X_n)$ είναι $f(\underset{\sim}{x}; \theta) = \prod_{i=1}^{n} (1/\theta) e^{-x_i/\theta} = \frac{1}{\theta^n} e^{-\sum x_i/\theta}$, $x_i > 0$, $i = 1, \ldots, n$.)

3. Η εκτίμηση της τιμής $g(\theta)$ βασίζεται σε συναρτήσεις των δεδομένων $\underset{\sim}{X}$. Κάθε συνάρτηση του $\underset{\sim}{X} = (X_1, \ldots, X_n)$, που δεν εξαρτάται από την άγνωστη παράμετρο θ, λέγεται *στατιστική συνάρτηση* . Εξ ορισμού, λοιπόν, ο (μαθηματικός) τύπος μιας στατιστικής συνάρτησης δεν περιέχει την άγνωστη παράμετρο θ. Ειδικά, μία στατιστική συνάρτηση, που χρησιμοποιείται για την εκτίμηση της τιμής $g(\theta)$, λέγεται *εκτιμητής* της τιμής $g(\theta)$ ή πιο απλά εκτιμητής του $g(\theta)$. Το πρόβλημα είναι ο προσδιορισμός ενός «κατάλληλου» εκτιμητή του $g(\theta)$, δηλαδή μιας στατιστικής συνάρτησης $T(\underset{\sim}{X})$ με την επιθυμητή ιδιότητα οι τιμές του $T(\underset{\sim}{X})$ να είναι «κοντά» στην άγνωστη τιμή $g(\theta)$. Η τιμή του εκτιμητή για τη παρατηρηθείσα τιμή $\underset{\sim}{x}$ του $\underset{\sim}{X}$, δηλαδή $T(\underset{\sim}{x})$, λέγεται *εκτίμηση* της τιμής $g(\theta)$ ή πιο απλά εκτίμηση του $g(\theta)$. Συχνά, για λόγους απλότητας στο συμβολισμό, θα γράφουμε T αντί $T(\underset{\sim}{X})$.

3.2 Κριτήρια επιλογής εκτιμητών

Πώς θα μπορούσαμε να αποφανθούμε για την καταλληλότητα μιας στατιστικής συνάρτησης $T(\underset{\sim}{X})$ ως εκτιμητή του $g(\theta)$; Φαινομενικά, το απόλυτο σφάλμα εκτίμησης, $|T(\underset{\sim}{X}) - g(\theta)|$, δίνει απάντηση σε αυτό το ερώτημα: Όσο πιο «μικρό», τόσο πιο «κοντά» είναι ο $T(\underset{\sim}{X})$ στο $g(\theta)$. Όμως, οφείλουμε να αναγνωρίσουμε ότι η παράσταση $|T(\underset{\sim}{X}) - g(\theta)|$ έχει *άγνωστη* και *τυχαία* τιμή· άγνωστη, επειδή περιέχει την άγνωστη τιμή $g(\theta)$, και τυχαία επειδή, είναι τυχαία μεταβλητή ως συνάρτηση του $\underset{\sim}{X}$. Αν λοιπόν $\mathcal{X}$ είναι το σύνολο τιμών του $\underset{\sim}{X}$, οποιαδήποτε από τις τιμές $|T(\underset{\sim}{y}) - g(\theta)|$, $\underset{\sim}{y} \in \mathcal{X}$, $\theta \in \Theta$ θα μπορούσε να είναι η τιμή του απολύτου σφάλματος. Δεν γνωρίζουμε όμως ποια είναι. Επομένως, το απόλυτο σφάλμα, ως μη «μετρήσιμο», δεν εξυπηρετεί ως κριτήριο επιλογής ή σύγκρισης εκτιμητών. Προφανώς, στο ίδιο συμπέρασμα θα καταλήξουμε, αν αποπειραθούμε να αξιολογήσουμε τον εκτιμητή $T(\underset{\sim}{X})$, χρησιμοποιώντας παρόμοιες ποσότητες, όπως το τετραγωνικό σφάλμα εκτίμησης $(T(\underset{\sim}{X}) - g(\theta))^2$ ή ακόμη μια δύναμη του απολύτου σφάλματος $|T(\underset{\sim}{X}) - g(\theta)|^\kappa$, $\kappa > 0$.

Από την παραπάνω ανάλυση γίνεται αντιληπτό ότι, για να καταλήξουμε σε ένα *χρήσιμο* κριτήριο, θα πρέπει, κατά κάποιο τρόπο, να *διαχειριστούμε* την τυχαιότητα του σφάλματος εκτίμησης, αλλά και το άγνωστο (του) θ. Προς αυτήν την κατεύθυνση μερικά σημαντικά κριτήρια είναι τα εξής.

1. Μέσο απόλυτο σφάλμα - Μέσο τετραγωνικό σφάλμα - Μέση ζημία (Διαχείριση της τυχαιότητας του σφάλματος)

Αντί της τυχαίας τιμής (τυχαίας μεταβλητής) $|T(\underset{\sim}{X}) - g(\theta)|$, θεωρούμε την κατά μέσο όρο τιμή της ως προς την κατανομή του $\underset{\sim}{X}$, δηλαδή λαμβάνουμε ως κριτήριο το μέσο απόλυτο σφάλμα $\mathbb{E}_\theta|T(\underset{\sim}{X}) - g(\theta)|$, το οποίο εξαρτάται μόνον από το θ. Τώρα, ναι μεν το θ έχει μία και μοναδική τιμή στο Θ, όμως αυτή μπορεί να είναι οποιοδήποτε στοιχείο του Θ. Επομένως, για να μελετήσουμε το μέσο απόλυτο σφάλμα, επιβάλλεται να το θεωρήσουμε ως *συνάρτηση του* θ, αφήνοντας το θ να διατρέχει το Θ (δηλαδή το πεδίο ορισμού της συνάρτησης είναι το Θ). Μάλιστα, ο δείκτης στην παραπάνω μέση τιμή τονίζει την εξάρτησή της από το θ. Στη συνέχεια, λοιπόν, χωρίς

να υπάρχει κίνδυνος παρερμηνείας, θα συμβολίζουμε με θ, όχι μόνον την άγνωστη παράμετρο (που έχει μία και μοναδική τιμή), αλλά και οποιοδήποτε στοιχείο του Θ.
Με το ίδιο σκεπτικό, ως κριτήριο μπορεί να ληφθεί το *μέσο τετραγωνικό σφάλμα* $\mathbb{E}_\theta(T(\underset{\sim}{X})-g(\theta))^2$ ή ακόμη $\mathbb{E}_\theta|T(\underset{\sim}{X})-g(\theta)|^\kappa$, $\kappa>0$, καθένα μάλιστα θεωρούμενο ως συνάρτηση του $\theta\in\Theta$. Ακόμη γενικότερα μπορούμε να θεωρήσουμε μία συνάρτηση $L(t,\theta)$, που την ονομάζουμε *συνάρτηση ζημίας* και η οποία παριστάνει, ποσοτικά, τη «ζημία μας» (ή την ποινή ή το σφάλμα), εάν εκτιμήσουμε το $g(\theta)$ με την τιμή t. Συνήθως η συνάρτηση $L(t,\theta)$ λαμβάνεται τέτοια ώστε

$$\begin{aligned} L(t,\theta) &\geq 0, \quad \forall\, \theta, t \\ L(g(\theta),\theta) &= 0, \quad \forall\, \theta\in\Theta. \end{aligned}$$

Οι ποιοτικές αυτές ιδιότητες δηλώνουν ότι η «ζημία μας» είναι μηδέν ή θετική και ειδικά, στην περίπτωση που η εκτίμηση t συμπέσει με την τιμή $g(\theta)$, μηδενίζεται. Εάν $T(\underset{\sim}{X})$ είναι εκτιμητής του $g(\theta)$, η συνάρτηση (του θ)

$$R(T,\theta)=\mathbb{E}_\theta(L(T(\underset{\sim}{X}),\theta))$$

λέγεται *συνάρτηση κινδύνου* του εκτιμητή $T(\underset{\sim}{X})$ και παριστάνει τη «μέση ζημία μας» από τη χρησιμοποίηση του $T(\underset{\sim}{X})$ ως εκτιμητή του $g(\theta)$. Προφανώς, «μικρές» τιμές $R(T,\theta)$, $\theta\in\Theta$ συνηγορούν υπέρ της καταλληλότητας του εκτιμητή T. Σημειώνουμε ότι για $L(t,\theta)=|t-g(\theta)|^r$ με $r=1,2,\kappa$ προκύπτουν τα παραπάνω αναφερθέντα κριτήρια επιλογής. Ιδανικά, θα θέλαμε να βρούμε έναν εκτιμητή $T(\underset{\sim}{X})$ που ελαχιστοποιεί το $R(T,\theta)$ οποιαδήποτε και εάν είναι η τιμή του θ, δηλαδή για κάθε $\theta\in\Theta$. Όμως, όπως θα δούμε στο Κεφάλαιο 4 (μετά το Παράδειγμα 4.1.6), αυτό είναι αδύνατο.

Για λόγους απλότητας και ευκολίας στον υπολογισμό του, θα ασχοληθούμε, σε επόμενα κεφάλαια, κυρίως με το μέσο τετραγωνικό σφάλμα $\mathbb{E}_\theta(T(\underset{\sim}{X})-g(\theta))^2$, το οποίο είναι ίσως το πιο ευρέως χρησιμοποιούμενο στην πράξη κριτήριο σύγκρισης εκτιμητών.

2. **Bayes εκτιμητής και minimax εκτιμητής (Διαχείριση της τυχαιότητας του σφάλματος και του άγνωστου θ)**

α. Μπορούμε να έχουμε μια συνοπτική εικόνα της μέσης ζημίας του εκτιμητή $T(\underset{\sim}{X})$, θεωρώντας κάποιο «μέσο όρο» των τιμών της συνάρτησης $R(T;\theta)$, $\theta \in \Theta$. Πιο συγκεκριμένα σε κάθε στοιχείο $\theta \in \Theta$ μπορούμε να αντιστοιχίσουμε ένα συντελεστή βαρύτητας $\pi(\theta)$, που υποδεικνύει τη σημαντικότητα αυτού του στοιχείου, και να θεωρήσουμε ως κριτήριο τη σταθμισμένη μέση τιμή

$$BR(T) = \sum_{\theta \in \Theta} R(T,\theta)\pi(\theta) \quad \text{ή} \quad BR(T) = \int_{\Theta} R(T,\theta)\pi(\theta)d\theta$$

για Θ αριθμήσιμο - πεπερασμένο σύνολο ή Θ μη αριθμήσιμο σύνολο, π.χ. διάστημα της πραγματικής ευθείας, αντίστοιχα, που ονομάζεται *κίνδυνος Bayes του T*. Συχνά η συνάρτηση βαρύτητας επιλέγεται, έτσι ώστε να έχει τις χαρακτηριστικές ιδιότητες της πυκνότητας μιας τυχαίας μεταβλητής με σύνολο τιμών Θ, δηλαδή $\pi(\theta) \geqslant 0$, για κάθε $\theta \in \Theta$ (κανείς δεν θα αμφισβητούσε αυτή την ιδιότητα) και $\sum_{\theta \in \Theta} \pi(\theta) = 1$ (Θ αριθμήσιμο ή πεπερασμένο) ή $\int_{\Theta} \pi(\theta)d\theta = 1$ (Θ μη αριθμήσιμο). Η δεύτερη ιδιότητα έχει την ερμηνεία ότι συνολική βαρύτητα μιας μονάδας κατανέμεται μέσω της συνάρτησης $\pi(\theta)$ στα στοιχεία του Θ, καθένα εκ των οποίων μπορεί να είναι η άγνωστη τιμή του θ. Σε αυτή την περίπτωση, $BR(T)$ είναι και τυπικά η μέση τιμή της τυχαίας μεταβλητής $R(T,\boldsymbol{\theta})$, όπου $\boldsymbol{\theta}$ είναι τυχαία μεταβλητή με πυκνότητα $\pi(\theta)$, $\theta \in \Theta$. Έτσι λοιπόν για τον εκτιμητή $T(\underset{\sim}{X})$, η κατά μέσο όρο ως προς $\underset{\sim}{X} \in \mathcal{X}$ και ως προς $\theta \in \Theta$ «ζημία μας» απεικονίζεται στον γνωστό (υπολογίσιμο) πραγματικό αριθμό (ή ∞, αν είμαστε τόσο άτυχοι στην επιλογή μας) $BR(T)$, μικρές τιμές του οποίου είναι επιθυμητές. Ένας εκτιμητής T^* (αν υπάρχει) που ελαχιστοποιεί τον κίνδυνο Bayes, δηλαδή ικανοποιεί τη σχέση $BR(T^*) \leqslant BR(T)$ για κάθε άλλο εκτιμητή T, ονομάζεται Bayes εκτιμητής του $g(\theta)$. Τέλος, δύο τυχόντες εκτιμητές T_1 και T_2 είναι συγκρίσιμοι με το κριτήριο Bayes, αφού $BR(T_1) < BR(T_2)$ ή

$BR(T_1) = BR(T_2)$ ή $BR(T_1) > BR(T_2)$. Η ονομασία εκτιμητής Bayes οφείλεται στο γεγονός ότι στην εύρεση του T_0 μπορεί να χρησιμοποιηθεί ο τύπος του Bayes (βλέπε Πρόταση 1.2.1). Οι εκτιμητές Bayes μελετώνται στο Κεφάλαιο 8.

β. (Απαισιόδοξα σκεπτόμενοι) θεωρούμε ως κριτήριο επιλογής ενός εκτιμητή τη μέγιστη ως προς $\theta \in \Theta$ μέση ζημία του εκτιμητή $T(\underset{\sim}{X})$, δηλαδή

$$\sup_{\theta \in \Theta} R(T, \theta) = \sup_{\theta \in \Theta} \mathbb{E}_\theta L(T(\underset{\sim}{X}), \theta).$$

Ένας εκτιμητής T_* (αν υπάρχει) που ικανοποιεί τη σχέση $\sup\limits_{\theta \in \Theta} R(T_*, \theta) \leqslant \sup\limits_{\theta \in \Theta} R(T, \theta)$ για κάθε άλλο εκτιμητή $T(\underset{\sim}{X})$ ονομάζεται minimax εκτιμητής του $g(\theta)$. Η ονομασία δικαιολογείται από το γεγονός ότι ο T_* ελαχιστοποιεί (mini) τη μέγιστη (max) μέση ζημία. Είναι επίσης προφανές ότι δύο τυχόντες εκτιμητές T_1 και T_2 είναι συγκρίσιμοι με το κριτήριο minimax, αφού $\sup\limits_{\theta \in \Theta} R(T_1, \theta) < \sup\limits_{\theta \in \Theta} R(T_2, \theta)$ ή $\sup\limits_{\theta \in \Theta} R(T_1, \theta) = \sup\limits_{\theta \in \Theta} R(T_2, \theta)$ ή $\sup\limits_{\theta \in \Theta} R(T_1, \theta) > \sup\limits_{\theta \in \Theta} R(T_2, \theta)$. Οι εκτιμητές minimax μελετώνται επίσης στο Κεφάλαιο 8.

Ανακεφαλαιώνοντας, το κριτήριο της μέσης ζημίας είναι ισχυρότερο από τα άλλα δύο, γιατί αποτυπώνει την καταλληλότητα του εκτιμητή μέσω μιας συνάρτησης, της συνάρτησης κινδύνου $R(T, \theta)$, $\theta \in \Theta$, και όχι μέσω μίας σταθεράς. Όμως, έχει τη δυσκολία ότι δεν είναι πάντοτε καταληκτικό, αφού ανάγει τη σύγκριση δύο εκτιμητών στη σύγκριση δύο συναρτήσεων. Τα άλλα δύο κριτήρια «συμπιέζουν» ή «συνοψίζουν» τη συνάρτηση $R(T, \theta)$ σε μία σταθερά, στον κίνδυνο Bayes $BR(T)$ και και στη μέγιστη ως προς $\theta \in \Theta$ μέση ζημία $R(T, \theta)$, αντίστοιχα, είναι καταληκτικά, αλλά σαφώς ασθενέστερα (ειδικά το κριτήριο minimax), αφού λογικό είναι να αποκρυβεί σημαντική πληροφορία για την καταλληλότητα του εκτιμητή κατά τη «συμπίεση». Για αυτό τον λόγο, συχνά τα δύο αυτά κριτήρια συνδυάζονται μεταξύ τους ή με άλλα κριτήρια, π.χ. αναζητούμε εκτιμητή που είναι συγχρόνως Bayes και minimax, ή Bayes και αποδεκτός, ή minimax και αποδεκτός. Η ιδιότητα της αποδεκτικότητας ορίζεται στην Ενότητα 4.1.

3.3 Απλές μέθοδοι εκτίμησης

Τα κριτήρια εκτίμησης που σκιαγραφήσαμε στην προηγούμενη ενότητα (α) εξυπηρετούν στη σύγκριση δύο *δοθέντων* εκτιμητών και (β) θεσπίζουν απαιτήσεις για την ανάδειξη βέλτιστου εκτιμητή. Η κατασκευή βέλτιστου εκτιμητή είναι μία διαδικασία που ακολουθεί ειδική τεχνική ανάλογα με το κριτήριο. Αυτές τις τεχνικές θα τις μελετήσουμε διεξοδικά σε επόμενα κεφάλαια. Σε πολλές όμως περιπτώσεις μπορούμε να κατασκευάσουμε εκτιμητές βασιζόμενοι σε απλές, λογικοφανείς, κατανοητές και εύκολα εφαρμόσιμες ιδέες. Μάλιστα, κάποιες από αυτές παράγουν εκτιμητές που τελικά αποδεικνύεται ότι έχουν πολύ καλές ιδιότητες. Θα περιγράψουμε, εν συντομία, μερικές μεθόδους που υλοποιούν αυτές τις ιδέες. Για λόγους απλότητας, θεωρούμε ότι οι παρατηρήσεις $X_1, X_2, \ldots, X_n$ αποτελούν τυχαίο δείγμα από κάποιο πληθυσμό.

1. **Δειγματικά ανάλογα παραμέτρων.** Σε πολλές περιπτώσεις πρακτικού ενδιαφέροντος, το άγνωστο $g(\theta)$ έχει φυσική ερμηνεία. Τότε, αβίαστα, μπορούμε να θεωρήσουμε ως εκτιμητή του το δειγματικό ανάλογό του, δηλαδή την αντίστοιχη «ποσότητα» του δείγματος $X_1, X_2, \ldots, X_n$. Ας δούμε μερικές ειδικές περιπτώσεις.

α. **Εκτίμηση της μέσης τιμής, $g(\theta) = \mathbb{E}_\theta X_1$.**
Αφού το $g(\theta)$ παριστάνει τη μέση τιμή του πληθυσμού, είναι διαισθητικά λογικό να την εκτιμήσουμε με τη μέση τιμή του δείγματος $X_1, X_2, \ldots, X_n$, δηλαδή με τη στατιστική συνάρτηση $\bar{X} = \frac{X_1+\ldots+X_n}{n}$. Μάλιστα, από τον (Ασθενή) Νόμο των Μεγάλων Αριθμών (ΑΝΜΑ) γνωρίζουμε ότι $\bar{X} \xrightarrow{P} \mathbb{E}_\theta X_1$, για κάθε $\theta \in \Theta$, άρα, τουλάχιστον για «μεγάλο» μέγεθος δείγματος n, αναμένουμε $\bar{X} \approx g(\theta)$. Η στατιστική συνάρτηση $\bar{X}$ ονομάζεται *δειγματικός μέσος* και είναι το δειγματικό ανάλογο του $g(\theta) = \mathbb{E}_\theta X_1$.

β. **Εκτίμηση της μέσης τιμής τετραγώνου, $g(\theta) = \mathbb{E}_\theta(X_1^2)$.**
Επειδή το $g(\theta)$ είναι η μέση τιμή της κοινής κατανομής των $X_1^2, X_2^2, \ldots, X_n^2$, κατ' αναλογία με το ($\alpha$) μπορούμε να θεωρήσουμε ως εκτιμητή του τη στατιστική συνάρτηση $\dfrac{1}{n}\sum\limits_{i=1}^{n} X_i^2$.

γ. **Εκτίμηση της διασποράς με γνωστή μέση τιμή μ, $g(\theta) = \mathbb{E}_\theta(X_1-$**

$\boldsymbol{\mu)^2}$**.**

Επειδή το $g(\theta)$ είναι η μέση τιμή της κοινής κατανομής των $(X_1-\mu)^2, (X_2-\mu)^2, \ldots, (X_n-\mu)^2$, κατ' αναλογία με το ($\alpha$) μπορούμε να εκτιμήσουμε το $g(\theta)$ με τον εκτιμητή $\frac{1}{n}\sum\limits_{i=1}^{n}(X_i-\mu)^2$.

δ**. Εκτίμηση της συνάρτησης κατανομής,** $\boldsymbol{g(\theta) = F(x;\theta) = \mathbb{P}_\theta(X_1 \leqslant x)}$**,** $\boldsymbol{x \in \mathbb{R}}$**.**

Η συνάρτηση κατανομής στο σημείο x, $F(x;\theta)$, ερμηνεύεται διαισθητικά ως το ποσοστό των τιμών της κατανομής (του πληθυσμού) που δεν υπερβαίνουν το x. Συνεπώς, μπορούμε να θεωρήσουμε ως εκτιμητή το αντίστοιχο ποσοστό των παρατηρήσεων $X_1, X_2, \ldots, X_n$, δηλαδή, τη στατιστική συνάρτηση $\hat{F}(x) = \frac{1}{n} \times \{$ αριθμός των X_i που ικανοποιούν τη σχέση $X_i \leqslant x$, $i = 1, \ldots, n\} = \frac{1}{n}\sum\limits_{i=1}^{n} I_{(-\infty,x]}(X_i)$, όπου I_A, $A \subset \mathbb{R}$, είναι η δείκτρια συνάρτηση με τύπο $I_A(y) = 1$ ή 0 αν $y \in A$ ή $y \notin A$, αντίστοιχα. Ο εκτιμητής $\hat{F}(x)$ ονομάζεται *εμπειρική συνάρτηση κατανομής* και είναι το δειγματικό ανάλογο της συνάρτησης κατανομής $F(x;\theta)$.

ϵ**. Εκτίμηση ενός ποσοστού,** $\boldsymbol{g(\theta) = \mathbb{P}_\theta(X_1 \in B)}$**,** B **δοθέν υποσύνολο του συνόλου τιμών της** X_1**.**

Όπως στο (δ), το δειγματικό ανάλογο του $g(\theta)$ είναι η στατιστική συνάρτηση $\frac{1}{n} \times \{$ αριθμός των X_i που ικανοποιούν τη σχέση $X_i \in B, i = 1, \ldots, n\}$ $= \frac{1}{n}\sum\limits_{i=1}^{n} I_B(X_i)$.

2. **Αρχή της αντικατάστασης.** Ας θεωρήσουμε προς στιγμή $g(\theta) = (\mathbb{E}_\theta X_1)^2$. Λόγω του ($\alpha$) είναι λογικό να θεωρήσουμε ως εκτιμητή του $g(\theta)$ τη στατιστική συνάρτηση $\bar{X}^2$. (αφού εκτιμούμε τη $\mathbb{E}_\theta X_1$ με $\bar{X}$ γιατί να μην εκτιμήσουμε το τετράγωνό της με $\bar{X}^2$;) Ας γενικεύσουμε τώρα αυτήν την ιδέα. Έστω ότι το $g(\theta)$ μπορεί να παρασταθεί ως συνάρτηση άγνωστων «ποσοτήτων» $v_1, v_2, \ldots, v_\kappa$, δηλαδή, $g(\theta) = h(v_1, v_2, \ldots, v_\kappa)$, όπου η h είναι «ομαλή» (συνεχής) συνάρτηση. Έστω ακόμη ότι διαθέτουμε εκτιμητές $\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_\kappa$ αυτών των «ποσοτήτων», αντίστοιχα. Τότε *αντικαθιστώντας* στον τύπο της h τα v_i με τα $\hat{v}_i$, μπορούμε να θεωρήσουμε ως εκτιμητή του $g(\theta)$ τη στατιστική συνάρτηση $\hat{g}(\theta) = h(\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_\kappa)$. Ως εφαρμογή ας δούμε δύο παραδείγματα.

α**. Εκτίμηση της διασποράς,** $\boldsymbol{g(\theta) = \mathbb{V}ar_\theta X_1}$**.**

Έχουμε $g(\theta) = \mathbb{E}_\theta(X_1^2) - (\mathbb{E}_\theta X_1)^2 = v_2 - v_1^2 = h(v_1, v_2)$, όπου $v_1 = \mathbb{E}_\theta X_1$ και $v_2 = \mathbb{E}_\theta(X_1^2)$. Από τα (1$\alpha$) και (1$\beta$) προκύπτουν οι εκτιμητές $\hat{v}_1 = \bar{X}$ και $\hat{v}_2 = \frac{1}{n}\sum\limits_{i=1}^{n} X_i^2$ και επομένως ο εκτιμητής της διασποράς βάσει της αρχής της αντικατάστασης είναι

$$\hat{g}(\theta) = \hat{v}_2 - \hat{v}_1^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

β. **Εκτίμηση της τυπικής απόκλισης, $g(\theta) = \sqrt{\mathbb{V}ar_\theta X_1}$.**
Έχουμε $g(\theta) = \sqrt{v_2 - v_1^2}$, οπότε ο εκτιμητής του $g(\theta)$ βάσει της αρχής της αντικατάστασης είναι

$$\hat{g}(\theta) = \sqrt{\hat{v}_2 - \hat{v}_1^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

3. **Μέθοδος της μέγιστης πιθανοφάνειας.** Θα περιγράψουμε αυτή τη μέθοδο, πρώτα, μέσω ενός παραδείγματος. Έστω ότι σε κάθε εκτέλεση ενός πειράματος τύχης προκύπτει (παρατηρείται) μία τιμή τυχαίας μεταβλητής X, της οποίας η κατανομή πιθανότητας συνοψίζεται στον πίνακα:

	$X=0$	$X=1$	$X=2$
$\theta=\theta_1$	0.02	0.95	0.03
$\theta=\theta_2$	0.90	0.05	0.05
$\theta=\theta_3$	0.23	0.06	0.71

Δηλαδή, η X είναι διακριτή με σύνολο τιμών $\{0,1,2\}$, η άγνωστη παράμετρος θ έχει τιμή που ανήκει στο σύνολο $\Theta = \{\theta_1, \theta_2, \theta_3\}$ και κάθε στοιχείο του πίνακα είναι η πιθανότητα της τιμής της X για την αντίστοιχη τιμή του θ, $f(x;\theta) = \mathbb{P}_\theta(X = x)$, π.χ. $f(0;\theta_2) = 0.90$. Έστω ακόμη ότι σε μία εκτέλεση του πειράματος παρατηρήθηκε η τιμή $x = 1$ και μας ζητείται με μόνο δεδομένο αυτήν την τιμή να εκτιμήσουμε το θ. Σκεφτόμαστε ως εξής: τι πιθανότητα είχε η παρατηρηθείσα τιμή $x = 1$ να προκύψει στη συγκεκριμένη εκτέλεση του πειράματος; Η απάντηση εξαρτάται από την άγνωστη τιμή του θ και είναι $\mathbb{P}_\theta(X = 1) = 0.95$ αν $\theta = \theta_1$, ή $\mathbb{P}_\theta(X = 1) = 0.05$ αν $\theta = \theta_2$ ή $\mathbb{P}_\theta(X = 1) = 0.06$ αν $\theta = \theta_3$. Βλέπουμε λοιπόν ότι η παρατηρηθείσα τιμή $x = 1$ είχε τη *μεγαλύτερη πιθανότητα να* *προκύψει*, αν $\theta = \theta_1$ (και όχι αν $\theta = \theta_2$ ή $\theta = \theta_3$). Άρα είναι λογικά ορθό να εκτιμήσουμε το θ με την τιμή θ_1 (και όχι με κάποια εκ των θ_2, θ_3). Με το ίδιο σκεπτικό, αν είχε παρατηρηθεί η τιμή $x = 0$, η εκτίμηση του θ θα ήταν θ_2, ενώ για $x = 2$ η εκτίμηση θα ήταν θ_3.

Γενικά λοιπόν αν σε μια εκτέλεση του πειράματος τύχης παρατηρηθεί η τιμή $x \in \{0, 1, 2\}$, ως εκτίμηση του θ λαμβάνουμε την τιμή $\hat{\theta}(x) \in \Theta = \{\theta_1, \theta_2, \theta_3\}$ που *μεγιστοποιεί* ως προς $\theta \in \Theta$ την πιθανότητα να προκύψει η παρατηρηθείσα τιμή x, $\mathbb{P}_\theta(X = x)$, η οποία εδώ συμπίπτει με την πυκνότητα $f(x; \theta)$, λόγω διακριτής κατανομής. Σημειώνουμε ότι η μεγιστοποίηση της πυκνότητας τεκμηριώνεται ακόμη και στην περίπτωση συνεχούς κατανομής, αφού η τιμή της πυκνότητας στο σημείο x, $f(x; \theta)$, υποδηλώνει πόσο μεγάλη είναι η πιθανότητα η τυχαία μεταβλητή να πάρει τιμή σε μια «μικρή» περιοχή του x : για «μικρό» $\epsilon > 0$, έχουμε

$$\mathbb{P}(x - \epsilon < X < x + \epsilon) = \int_{x-\epsilon}^{x+\epsilon} f(t; \theta) dt \approx 2\epsilon f(x; \theta).$$

Αυτή η ιδέα της μεγιστοποίησης της πυκνότητας ως προς την άγνωστη παράμετρο αναφέρεται ως *αρχή της μέγιστης πιθανοφάνειας* .

Γενικότερα, σε ένα πρόβλημα εκτίμησης παραμέτρου $\theta \in \Theta$ και σύμφωνα με το παράδειγμα αυτό, έστω $\underset{\sim}{x} = (x_1, \ldots, x_n)$ η παρατηρηθείσα τιμή του $\underset{\sim}{X} = (X_1, \ldots, X_n)$ και $\hat{\theta}(\underset{\sim}{x}) \in \Theta$ η μεγιστοποιούσα την πυκνότητα του $\underset{\sim}{X}$, $f(\underset{\sim}{x}; \theta)$, τιμή του $\theta \in \Theta$ (αν υπάρχει), δηλαδή εξ' ορισμού, $f(\underset{\sim}{x}; \hat{\theta}(\underset{\sim}{x})) = \max\{f(\underset{\sim}{x}; \theta) : \theta \in \Theta\}$. Τότε η τιμή $\hat{\theta}(\underset{\sim}{x})$ αναφέρεται ως η εκτίμηση μέγιστης πιθανοφάνειας και η στατιστική συνάρτηση $\hat{\theta}(\underset{\sim}{X})$ ως ο εκτιμητής μέγιστης πιθανοφάνειας του θ.

Σημειώνουμε ότι, πέραν της διαισθητικής ερμηνείας του $\hat{\theta}(\underset{\sim}{x})$, και η διαδικασία εύρεσης της εκτίμησης μέγιστης πιθανοφάνειας είναι σχετικά απλή, αφού απαιτεί μόνον τεχνικές μεγιστοποίησης, αναλυτικές ή αριθμητικές, που υπάρχουν διαθέσιμες στη βιβλιογραφία. Τους εκτιμητές μέγιστης πιθανοφάνειας θα μελετήσουμε λεπτομερώς στο Κεφάλαιο 7 και όπως θα δούμε έχουν, εν γένει, πολύ καλές ιδιότητες.

4. **Μέθοδος των ροπών.** Η ροπή k τάξης της (κοινής) κατανομής των παρατηρήσεων $X_1, \ldots, X_n$ ορίζεται από τη σχέση $\mu_k(\theta) = \mathbb{E}_\theta(X_1^k)$,

$k = 1, 2, \ldots$, ενώ η δειγματική ροπή k τάξης (το δειγματικό ανάλογο του $\mu_k(\theta)$) είναι η στατιστική συνάρτηση $m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$. Από τον ΑΝΜΑ γνωρίζουμε ότι $m_k \xrightarrow{P} \mu_k(\theta)$ καθώς $n \to \infty$, άρα τουλάχιστον για «μεγάλο» n αναμένουμε $m_k \approx \mu_k(\theta)$. Έτσι δικαιολογούμαστε, εν μέρει, να *εξισώσουμε* τα δύο μέλη αυτής της προσεγγιστικής σχέσης, δηλαδή να θέσουμε $m_k = \mu_k(\theta)$. Η λύση της εξίσωσης ως προς θ δίνει μια στατιστική συνάρτηση $\hat{\hat{\theta}}$ που ονομάζεται εκτιμητής μεθόδου ροπών του θ. Αν τώρα η παράμετρος είναι διάνυσμα $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_r)$, η λύση του συστήματος των εξισώσεων $m_1 = \mu_1(\theta_1, \ldots, \theta_r)$, ..., $m_r = \mu_r(\theta_1, \ldots, \theta_r)$ ως προς $\theta_1, \ldots, \theta_r$ δίνει στατιστικές συναρτήσεις $\hat{\hat{\theta}}_1, \ldots, \hat{\hat{\theta}}_r$, οπότε η στατιστική συνάρτηση $\hat{\hat{\underset{\sim}{\theta}}} = (\hat{\hat{\theta}}_1, \ldots, \hat{\hat{\theta}}_r)$ είναι ο εκτιμητής μεθόδου ροπών του $\underset{\sim}{\theta}$. Τη μέθοδο των ροπών θα τη μελετήσουμε λεπτομερώς στο Κεφάλαιο 7.

5. **Μέθοδος των ελαχίστων τετραγώνων (ή μέθοδος Gauss).** Ας θεωρήσουμε πρώτα την εξής ειδική περίπτωση. Έστω ότι η άγνωστη παράμετρος θ παριστάνει μια αριθμητική ποσότητα, η τιμή της οποίας μπορεί να μετρηθεί μόνον πειραματικά με κάποιο όργανο (π.χ. αρτηριακή πίεση - πιεσόμετρο). Η μέτρηση όμως υπόκειται σε τυχαίο σφάλμα, δηλαδή κάθε φορά που χρησιμοποιείται το όργανο, η ένδειξή του, x, είναι μία τιμή τυχαίας μεταβλητής X, ενώ το σφάλμα της μέτρησης $\epsilon = x - \theta$ είναι άγνωστο. Για λόγους λοιπόν αξιοπιστίας, η μέτρηση του θ επαναλαμβάνεται n φορές και καταγράφονται οι ενδείξεις του οργάνου $x_1, \ldots, x_n$ που είναι οι παρατηρηθείσες τιμές αντίστοιχων τυχαίων μεταβλητών $X_1, \ldots, X_n$. Για αυτές τις ενδείξεις, τα σφάλματα είναι $\epsilon_i = x_i - \theta$, $i = 1, \ldots, n$ και ως εκτίμηση του θ λαμβάνεται η τιμή που ελαχιστοποιεί το άθροισμα των τετραγώνων των σφαλμάτων $\sum_{i=1}^{n}\epsilon_i^2 = \sum_{i=1}^{n}(x_i - \theta)^2$ ως προς θ. Πολύ εύκολα (π.χ. με παραγώγιση) μπορεί να δειχθεί ότι η εκτίμηση του θ είναι $\hat{\theta}_\epsilon = \bar{x} = (x_1 + \ldots + x_n)/n$, δηλαδή ο μέσος όρος των ενδείξεων (διόλου παράξενο). Γενικότερα, αν η τυχαία μεταβλητή X_i παριστάνει την (πειραματική) μέτρηση μιας αριθμητικής ποσότητας $\tau_i(\theta)$, όπου το θ είναι άγνωστη παράμετρος, τ_i είναι γνωστή συνάρτηση και x_i είναι η παρατηρηθείσα τιμή της X_i, $i = 1, \ldots, n$, η εκτίμηση ελαχίστων τετραγώνων του θ είναι η τιμή που ελαχιστοποιεί την παράσταση $\sum_{i=1}^{n}\epsilon_i^2 = \sum_{i=1}^{n}(x_i - \tau_i(\theta))^2$

ως προς θ.

Η μέθοδος των ελαχίστων τετραγώνων θεμελιώθηκε από τον (μεγάλο Γερμανό μαθηματικό) Gauss σε αστρονομικές μελέτες του και χρησιμοποιείται ευρέως σε περιοχές της Στατιστικής, όπως η Ανάλυση Παλινδρόμησης και Ανάλυση Διασποράς (που είναι πέραν του σκοπού αυτών των σημειώσεων). Σημειώνουμε τέλος ότι στον αντίποδα της μεθόδου Gauss βρίσκεται η εκτίμηση του θ, που ελαχιστοποιεί το άθροισμα των απολύτων σφαλμάτων $\sum_{i=1}^{n}|\epsilon_i| = \sum_{i=1}^{n}|x_i - \theta|$ ως προς θ. Μπορεί να δειχθεί ότι η ελαχιστοποιούσα τιμή του θ είναι οποιαδήποτε διάμεσος των x_i, $i = 1, \ldots, n$ (αλλά η απόδειξη κάθε άλλο παρά εύκολη είναι).

Κεφάλαιο 4

Μέσο Τετραγωνικό Σφάλμα και Αμεροληψία

Στο κεφάλαιο αυτό μελετάμε το μέσο τετραγωνικό σφάλμα, το πιο γνωστό και συνάμα ευρέως χρησιμοποιούμενο, στη θεωρία και στις εφαρμογές της Στατιστικής, κριτήριο σύγκρισης και αξιολόγησης εκτιμητών. Επιπροσθέτως, μέσω αυτού του κριτηρίου, παρουσιάζουμε μια ιδιότητα των εκτιμητών, την αμεροληψία. Εκτιμητές που έχουν αυτήν την ιδιότητα αναφέρονται ως αμερόληπτοι εκτιμητές. Ακολούθως δίνουμε αμερόληπτους εκτιμητές σημαντικών παραμέτρων μιας κατανομής, όπως η μέση τιμή και η διασπορά της.

4.1 Μέσο τετραγωνικό σφάλμα

Το μέσο τετραγωνικό σφάλμα (ΜΤΣ) εκτιμητή $T(\underset{\sim}{X})$ της τιμής $g(\theta)$ ορίστηκε στην Ενότητα 3.2 και δίνεται από τη σχέση

$$\text{ΜΤΣ}(T(\underset{\sim}{X}),\theta) = \mathbb{E}_\theta\Big(T(\underset{\sim}{X}) - g(\theta)\Big)^2. \tag{4.1}$$

Επίσης, το τυπικό σφάλμα (ΤΣ) του $T(\underset{\sim}{X})$ ορίζεται από τη σχέση

$$\text{ΤΣ}(T(\underset{\sim}{X}),\theta) = \sqrt{\text{ΜΤΣ}(T(\underset{\sim}{X}),\theta)}.$$

Σε αριθμητικούς υπολογισμούς, θεωρείται προτιμότερο να ανακοινώνεται η τιμή του τυπικού σφάλματος αντί της τιμής του μέσου τετραγωνικού σφάλματος. Αυτό δικαιολογείται επειδή το ΜΤΣ αξιολογεί τις *τετραγωνικές* αποκλίσεις των τιμών του $T(\underset{\sim}{X})$ από το $g(\theta)$, ενώ υπολογίζοντας την τετραγωνική ρίζα του ΜΤΣ, το ΤΣ είναι άμεσα συγκρίσιμο με την εκτίμηση του $g(\theta)$ και επιπλέον έχει τις ίδιες μονάδες μέτρησης με το $g(\theta)$. Σημειώνουμε επίσης ότι το ΤΣ ή ένας εκτιμητής του ΤΣ καθορίζει σε ορισμένες περιπτώσεις το εύρος του διαστήματος εμπιστοσύνης για το $g(\theta)$.

Ο συμβολισμός ΜΤΣ$(T(\underset{\sim}{X}), \theta)$ ή πιο απλά ΜΤΣ(T, θ) δηλώνει ότι το ΜΤΣ εξαρτάται, εκτός φυσικά από τον συγκεκριμένο εκτιμητή $T(\underset{\sim}{X})$, εν γένει, και από την άγνωστη παράμετρο $\theta \in \Theta$, είναι δηλαδή συνάρτηση του $\theta \in \Theta$. Η φυσική του ερμηνεία είναι έκδηλη στον ορισμό του: παριστάνει την «κατά μέσο όρο» τετραγωνική απόκλιση του εκτιμητή από την τιμή $g(\theta)$, που έχει κληθεί να εκτιμήσει. Η καθιέρωση του ΜΤΣ ως κριτηρίου οφείλεται σε ένα μεγάλο βαθμό στο γεγονός ότι μπορεί να υπολογιστεί σχετικά εύκολα μέσω ενός απλού και χρήσιμου τύπου ο οποίος ισχύει για οποιονδήποτε εκτιμητή. Πιο συγκεκριμένα, στην επόμενη πρόταση παρατηρούμε ότι ο υπολογισμός του ΜΤΣ (ως συνάρτησης του θ) απαιτεί μόνον την εύρεση δύο βασικών παραμέτρων, της μέσης τιμής και της διασποράς του εκτιμητή και όχι κατ' ανάγκη τη γνώση της κατανομής του. Ας προσθέσουμε ακόμη ότι η άμεση σύνδεση του ΜΤΣ με τις δύο αυτές παραμέτρους, γνωστές σε κάθε χρήστη Στατιστικής, συνεισφέρει από μόνη της στη δημοφιλία και στην κατανόηση του ΜΤΣ. Μάλιστα, για εκτιμητές που έχουν την ιδιότητα της αμεροληψίας, το ΜΤΣ *συμπίπτει* με τη διασπορά τους, όπως θα δούμε στην Ενότητα 4.2. Τονίζουμε εδώ ότι το πλεονέκτημα του γενικού απλού τύπου υπολογισμού δεν το έχουν άλλα επίσης ορθολογικά κριτήρια, όπως, π.χ. το μέσο απόλυτο σφάλμα $\mathbb{E}_\theta|T(\underset{\sim}{X}) - g(\theta)|$ (βλέπε το Παράδειγμα 4.1.1). Τέλος, όσον αφορά τον συμβολισμό, ας μην ξεχνάμε ότι $T(\underset{\sim}{X})$ ή T παριστάνουν τον ίδιο εκτιμητή.

Πρόταση 4.1.1. *Ισχύει ότι*

$$\text{ΜΤΣ}(T, \theta) = \mathbb{V}ar_\theta T(\underset{\sim}{X}) + \Big(\mathbb{E}_\theta T(\underset{\sim}{X}) - g(\theta)\Big)^2, \quad \theta \in \Theta. \tag{4.2}$$

Απόδειξη. Χρησιμοποιώντας τη σχέση $\mathbb{V}arY = \mathbb{E}Y^2 - (\mathbb{E}Y)^2$ προκύπτει ότι $\mathbb{E}Y^2 = \mathbb{V}arY + (\mathbb{E}Y)^2$. Θέτοντας $Y = T(\underset{\sim}{X}) - g(\theta)$, έχουμε

$$\begin{aligned} \text{ΜΤΣ}(T,\theta) &= \mathbb{E}_\theta \Big(T(\underset{\sim}{X}) - g(\theta) \Big)^2 \\ &= \mathbb{V}ar_\theta \Big(T(\underset{\sim}{X}) - g(\theta) \Big) + \Big(\mathbb{E}_\theta \Big(T(\underset{\sim}{X}) - g(\theta) \Big) \Big)^2 \\ &= \mathbb{V}ar_\theta T(\underset{\sim}{X}) + \Big(\mathbb{E}_\theta T(\underset{\sim}{X}) - g(\theta) \Big)^2, \end{aligned}$$

επειδή $\mathbb{V}ar(T+\alpha) = \mathbb{V}ar\,T$ και $\mathbb{E}(T+\alpha) = \mathbb{E}\,T+\alpha$, όπου α σταθερά. □

Η σχέση (4.2) μας δίνει αφορμή να ορίσουμε την έννοια της μεροληψίας ενός εκτιμητή.

Ορισμός 4.1.1. *Η διαφορά*

$$b(T,\theta) = \mathbb{E}_\theta T(\underset{\sim}{X}) - g(\theta), \quad \theta \in \Theta$$

ονομάζεται μεροληψία του $T(\underset{\sim}{X})$ ως εκτιμητή του $g(\theta)$.

Γράφοντας $\mathbb{E}_\theta T(\underset{\sim}{X}) - g(\theta) = \mathbb{E}_\theta \Big(T(\underset{\sim}{X}) - g(\theta) \Big)$, βλέπουμε ότι η μεροληψία παριστάνει την «κατά μέσο όρο» απόκλιση του εκτιμητή από την εκτιμώμενη τιμή $g(\theta)$. Από την Πρόταση 4.1.1 και τον Ορισμό 4.1.1, προκύπτει αμέσως η σχέση

$$\text{ΜΤΣ}(T,\theta) = \mathbb{V}ar_\theta T + b^2(T,\theta), \quad \theta \in \Theta. \tag{4.3}$$

Παραθέτουμε δύο παραδείγματα υπολογισμού του ΜΤΣ.

Παράδειγμα 4.1.1. (κανονική κατανομή με γνωστή διασπορά - εκτίμηση της μέσης τιμής) Έστω $\underset{\sim}{X} = (X_1,\ldots,X_n)$ ένα τυχαίο δείγμα από την κανονική κατανομή $\mathcal{N}(\theta,\sigma^2)$, όπου θ είναι άγνωστη παράμετρος, $\theta \in \Theta = \mathbb{R}$ και σ^2 γνωστή θετική σταθερά. Για την εκτίμηση της μέσης τιμής θ (δηλαδή, εδώ, $g(\theta) = \theta$) θεωρούμε ως εκτιμητή τον δειγματικό μέσο $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ (βλέπε Ενότητα 3.3.1α). Θα υπολογίσουμε το ΜΤΣ και το μέσο απόλυτο σφάλμα του $\bar{X}$. Χρησιμοποιώντας (μόνον) ιδιότητες της μέσης τιμής και της διασποράς, προκύπτει ότι

$$\mathbb{V}ar_\theta \bar{X} = \mathbb{V}ar_\theta \left(\frac{1}{n} \sum_{i=1}^{n} X_i \right) = \frac{1}{n^2} \sum_{i=1}^{n} \mathbb{V}ar_\theta X_i = \frac{1}{n^2} \sum_{i=1}^{n} \sigma^2 = \frac{\sigma^2}{n}$$

και

$$b(\bar{X},\theta) = \mathbb{E}_\theta \bar{X} - \theta = \mathbb{E}_\theta \left(\frac{1}{n} \sum_{i=1}^{n} X_i \right) - \theta = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}_\theta X_i - \theta$$
$$= \frac{1}{n} \sum_{i=1}^{n} \theta - \theta = \theta - \theta = 0.$$

Άρα, από τη σχέση (4.3), έχουμε

$$\text{ΜΤΣ}(\bar{X},\theta) = \mathbb{V}ar_\theta \bar{X} + b^2(\bar{X},\theta) = \frac{\sigma^2}{n}, \quad \theta \in \mathbb{R}.$$

(Ας παρατηρήσουμε ότι το ΜΤΣ$(\bar{X},\theta)$ είναι στην προκειμένη περίπτωση σταθερά συνάρτηση ως προς θ.) Ο υπολογισμός του μέσου απόλυτου σφάλματος απαιτεί γνώση της κατανομής του $\bar{X}$. Από τις ιδιότητες της κανονικής κατανομής (βλέπε Πρόταση 1.8.6) γνωρίζουμε ότι ο $\bar{X}$ έχει κανονική κατανομή $\mathcal{N}(\theta,\sigma^2/n)$, με πυκνότητα $f(x) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} e^{-\frac{n}{2\sigma^2}(x-\theta)^2}$. Επομένως,

$$\mathbb{E}_\theta|\bar{X} - \theta| = \int_{-\infty}^{\infty} |x-\theta| \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} e^{-\frac{n}{2\sigma^2}(x-\theta)^2} dx = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} |y| e^{-\frac{n}{2\sigma^2}y^2} dy$$
$$= \frac{2\sqrt{n}}{\sigma\sqrt{2\pi}} \int_0^{\infty} y e^{-\frac{n}{2\sigma^2}y^2} dy,$$

επειδή η συνάρτηση $|y|e^{-\frac{n}{2\sigma^2}y^2}$ είναι άρτια. Όμως,

$$\int_0^{\infty} y e^{-\frac{n}{2\sigma^2}y^2} dy = -\frac{\sigma^2}{n} \int_0^{\infty} \left(e^{-\frac{n}{2\sigma^2}y^2} \right)' dy = -\frac{\sigma^2}{n} \left[e^{-\frac{n}{2\sigma^2}y^2} \right]_0^{\infty} = \frac{\sigma^2}{n}$$

και συνεπώς έχουμε $\mathbb{E}_\theta|\bar{X} - \theta| = \frac{\sqrt{2}}{\sqrt{\pi n}}\,\sigma \left(= \frac{\sqrt{2}}{\sqrt{\pi}} \,\text{ΤΣ}(\bar{X},\theta) \right), \quad \theta \in \mathbb{R}.$

Βλέπουμε λοιπόν πόσο πιο απλός και πιο εύκολος είναι ο υπολογισμός του ΜΤΣ σε σχέση με τον υπολογισμό του μέσου απόλυτου σφάλματος.

Παράδειγμα 4.1.2. (κατανομή Bernoulli - εκτίμηση της πιθανότητας «επιτυχίας») Έστω $\underset{\sim}{X} = (X_1,\ldots,X_n)$ ένα τυχαίο δείγμα από την κατανομή Bernoulli με πυκνότητα

$$f_1(x;\theta) = \theta^x(1-\theta)^{1-x}, \;\; x = 0,1, \;\theta \in \Theta = (0,1).$$

Γνωρίζουμε ότι αυτή η κατανομή μοντελοποιεί ένα πείραμα τύχης, το αποτέλεσμα του οποίου καταλήγει σε «επιτυχία» με πιθανότητα θ ή σε «αποτυχία» με πιθανότητα $1-\theta$, ενώ οι τιμές 1 και 0 παριστάνουν την «επιτυχία» και την «αποτυχία», αντίστοιχα. Με αυτήν την ερμηνεία, το άθροισμα $\sum_{i=1}^{n} X_i$ εκφράζει τον αριθμό των «επιτυχιών» σε n ανεξάρτητες επαναλήψεις του πειράματος τύχης, ενώ $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ είναι το ποσοστό των «επιτυχιών» στο δείγμα αυτών των n επαναλήψεων. Θεωρούμε εκτίμηση του θ (δηλαδή, εδώ, το $g(\theta)=\theta$). Από τα παραπάνω, δικαιολογείται η εκτίμηση της πιθανότητας «επιτυχίας» θ με το ποσοστό των «επιτυχιών» του δείγματος των n επαναλήψεων, $\bar{X}$. Έχουμε,

$$\begin{aligned}\mathbb{V}ar_\theta \bar{X} &= \mathbb{V}ar_\theta\Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big) = \frac{1}{n^2}\sum_{i=1}^{n} \mathbb{V}ar_\theta X_i = \frac{1}{n^2}\sum_{i=1}^{n} \theta(1-\theta) \\ &= \frac{\theta(1-\theta)}{n}\end{aligned}$$

και

$$\begin{aligned}b(\bar{X},\theta) &= \mathbb{E}_\theta \bar{X} - \theta = \mathbb{E}_\theta\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \theta = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_\theta X_i - \theta \\ &= \frac{1}{n}\sum_{i=1}^{n} \theta - \theta = \theta - \theta = 0.\end{aligned}$$

Συνεπώς, από τη σχέση (4.3) προκύπτει ότι

$$\text{ΜΤΣ}(\bar{X},\theta) = \mathbb{V}ar_\theta \bar{X} + b^2(\bar{X},\theta) = \frac{\theta(1-\theta)}{n}, \quad 0<\theta<1.$$

Επιπλέον, επειδή $\theta(1-\theta) \leqslant \frac{1}{4}$, για κάθε $\theta \in (0,1)$, παρατηρούμε ότι

$$\text{ΜΤΣ}(\bar{X},\theta) \leqslant \frac{1}{4n}\,, \quad 0<\theta<1.$$

Το παραπάνω άνω φράγμα του ΜΤΣ$(\bar{X},\theta)$ έχει την εξής ενδιαφέρουσα ερμηνεία. Αν θ είναι το ποσοστό των ψήφων ενός κόμματος στις επόμενες βουλευτικές εκλογές, οπότε «επιτυχία» σημαίνει θετική ψήφος, και εκτιμηθεί το θ με $\bar{X}$, το ποσοστό δηλαδή των ψήφων του κόμματος σε τυχαίο δείγμα, έστω, $n = 1000$ ψηφοφόρων, τότε το ΜΤΣ του $\bar{X}$ ικανοποιεί τη σχέση

$$\text{ΜΤΣ}(\bar{X},\theta) \leqslant \frac{1}{4000} = 0.00025$$

ενώ για το τυπικό σφάλμα του $\bar{X}$ έχουμε

$$\text{ΤΣ}(\bar{X},\theta) = \sqrt{\text{ΜΤΣ}(\bar{X},\theta)} \leqslant \sqrt{\frac{1}{4000}} = 0.0158 (\simeq 1.6\%),$$

για κάθε $\theta \in (0,1)$, δηλαδή όποια και αν είναι η (άγνωστη) τιμή του θ. Διαπιστώνουμε λοιπόν ότι η μέγιστη τιμή του τυπικού σφάλματος, της τάξης του 1.6%, είναι αξιοσημείωτα μικρή για μέτριες ή μεγάλες τιμές του ποσοστού θ (π.χ. 25% ή μεγαλύτερες). Επίσης, αυτή η μέγιστη τιμή του τυπικού σφάλματος καθορίζει το μέγιστο εύρος του διαστήματος εμπιστοσύνης για το θ, όπως θα δούμε στο Κεφάλαιο 8. Για σύγκριση, με δείγμα $n = 10000$ ψηφοφόρων (συνήθης τιμή σε exit polls) η μέγιστη τιμή του $\text{ΤΣ}(\bar{X},\theta)$ είναι $\sqrt{\frac{1}{4n}} = \sqrt{\frac{1}{40000}} = 0.005 (0.5\%)$!!!
Ας θεωρήσουμε τώρα τον εκτιμητή $T(\underset{\sim}{X}) = 0.25$, ο οποίος ουσιαστικά αγνοεί το δείγμα $\underset{\sim}{X}$, αφού εκτιμά το θ με την τιμή 0.25, πάντα, όποια και αν είναι η παρατηρηθείσα τιμή του $\underset{\sim}{X}$. Προφανώς, θα ήταν παράδοξο να χρησιμοποιηθεί ο $T(\underset{\sim}{X})$, εκτός αν είχαμε λόγους να πιστεύουμε ότι το θ είναι περίπου 0.25. Παραβλέποντας αυτό το παράδοξο του $T(\underset{\sim}{X})$, θα τον συγκρίνουμε με τον $\bar{X}$. Το ΜΤΣ του κατ΄ ευθείαν από τον ορισμό του, σχέση (4.1), είναι

$$\text{ΜΤΣ}(T(\underset{\sim}{X}),\theta) = \mathbb{E}_\theta(0.25-\theta)^2 = (\theta - 0.25)^2.$$

Καταλαβαίνουμε ότι, αν, όντως, η τιμή του θ είναι 0.25 , που δεν το ξέρουμε, έχουμε επιτύχει την ιδανική εκτίμηση με την επιλογή του εκτιμητή $T(\underset{\sim}{X}) = 0.25$. Μάλιστα αυτή η επιλογή έχει ΜΤΣ μηδέν, γιατί

$$\text{ΜΤΣ}(T(\underset{\sim}{X}),0.25) = (0.25-0.25)^2 = 0.$$

Αν όμως η (άγνωστη) τιμή του θ είναι, π.χ., 0.35, και επιλέξουμε τον εκτιμητή $T(\underset{\sim}{X}) = 0.25$, τότε

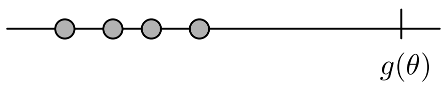
$$\text{ΜΤΣ}(T(\underset{\sim}{X}),0.35) = (0.35-0.25)^2 = 0.01,$$

με τυπικό σφάλμα ίσο προς $\sqrt{0.01} = 0.1$ (10%), ενώ επιλέγοντας τον εκτιμητή $\bar{X}$ έχουμε

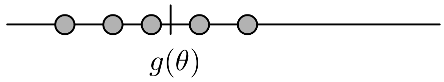
$$\text{ΜΤΣ}(\bar{X},\theta) = \frac{0.35 \times (1-0.35)}{1000} \backsimeq 0.0002275$$

με τυπικό σφάλμα ίσο προς $\sqrt{0.0002275} \backsimeq 0.015$ (1.51%), δηλαδή περίπου επτά φορές μικρότερο από το τυπικό σφάλμα του $T(\underset{\sim}{X})$. Για $n = 10000$, η αντίστοιχη τιμή του τυπικού σφάλματος του $\bar{X}$ είναι 0.0047, περίπου είκοσι φορές μικρότερη. Από τα παραπάνω συμπεραίνουμε ότι ο $T(\underset{\sim}{X}) = 0.25$ δεν μπορεί να «αποκλειστεί» ως εκτιμητής του θ, με κριτήριο το ΜΤΣ, παρότι όπως αναμενόταν έχει «μεγάλο» ΜΤΣ αν η (άγνωστη) τιμή του θ απέχει «αρκετά» από 0.25. Αυτός ο «μη αποκλεισμός» του $T(\underset{\sim}{X}) = 0.25$, όπως θα δούμε στο Παράδειγμα 4.1.6, υποδηλώνει σε αυστηρή ορολογία, την αποδεκτικότητά του.

Οι σχέσεις (4.2) και (4.3) αποκαλύπτουν ότι η διασπορά και η μεροληψία του εκτιμητή είναι οι δύο παράγοντες που διαμορφώνουν το ΜΤΣ. Με κριτήριο το ΜΤΣ ένας καλός εκτιμητής επιβάλλεται να έχει μικρή διασπορά και *συγχρόνως* μικρή μεροληψία σε απόλυτη τιμή. Σε μια πιο χαλαρή διατύπωση, ο εκτιμητής απαιτείται να είναι ακριβής (μικρή διασπορά), αλλά και ορθός ως προς το στόχο του, το $g(\theta)$ (βλέπε Σχήματα 4.1 και 4.2). Πάντως πρέπει να ομολογήσουμε ότι αυτή η απαίτηση είναι, εξ αρχής, πολύ λογική: λόγω μικρής διασποράς, περιμένουμε οι τιμές του εκτιμητή να μη διαφέρουν πολύ από τη μέση τιμή $\mathbb{E}_\theta T$, ενώ η μικρή (σε απόλυτη τιμή) μεροληψία εξασφαλίζει ότι η μέση τιμή $\mathbb{E}_\theta T$ δε διαφέρει πολύ από το $g(\theta)$. Τελικά, λοιπόν, περιμένουμε οι τιμές ενός τέτοιου εκτιμητή να είναι «κοντά» στην άγνωστη τιμή $g(\theta)$, το οποίο είναι ευθύς εξ αρχής το ζητούμενο.



Σχήμα 4.1: Ακριβής εκτιμητής, αλλά μη ορθός



Σχήμα 4.2: Ακριβής και ορθός εκτιμητής

Ας εξετάσουμε τώρα καθέναν από τους δύο αυτούς παράγοντες χωριστά, με στόχο να διαπιστώσουμε τελικά πως αλληλοεπιδρούν. Ξεκινώντας με τον πρώτο, η διασπορά $\mathbb{V}ar_\theta T$ στη σχέση (4.3) μπορεί να ελαχιστοποιηθεί αν μηδενιστεί, δηλαδή αν επιλεγεί μια σταθερά ως εκτιμητής του $g(\theta)$. Όμως, μια τέτοια επιλογή θα ήταν εντελώς ανόητη και παράλογη, γιατί θα αγνοούσε τα δεδομένα $\underset{\sim}{X}$ και, εκτός αυτού, θα έκανε ανεξέλεγκτη τη μεροληψία, άρα και το ΜΤΣ: π.χ., αν $T(\underset{\sim}{X}) = 5$ και $g(\theta) = \theta \in \Theta = \mathbb{R}$, έχουμε ΜΤΣ$(T, \theta) = (\theta - 5)^2$ με σύνολο τιμών στο $[0, \infty)$. Συνεχίζοντας με τον δεύτερο παράγοντα, η μεροληψία $b(T, \theta)$ στη σχέση (4.3) μπορεί να ελαχιστοποιηθεί, κατά απόλυτη τιμή, αν και αυτή μηδενιστεί, οπότε $b(T, \theta) = 0$, για κάθε $\theta \in \Theta$ συνεπάγεται $\mathbb{E}_\theta T(\underset{\sim}{X}) = g(\theta)$, για κάθε $\theta \in \Theta$. Επιλογή εκτιμητή που ικανοποιεί αυτή τη συνθήκη (αμεροληψίας, όπως θα την ονομάσουμε και μελετήσουμε στην επόμενη ενότητα) είναι, κατ' αρχάς τουλάχιστον, αναντίρρητα λογική: επιβάλλει στον εκτιμητή $T(\underset{\sim}{X})$ να παίρνει «κατά μέσο όρο» τιμή ίση με την υπό εκτίμηση «ποσότητα» $g(\theta)$. Όμως, η επιβολή της συνθήκης μηδενικής μεροληψίας δημιουργεί, ενδεχομένως, αυξημένη διασπορά $\mathbb{V}ar_\theta T$. Δεν αποκλείεται δηλαδή ένας άλλος εκτιμητής με μη μηδενική (αλλά σχετικά μικρή) μεροληψία να έχει μικρότερη διασπορά και τελικά μικρότερο ΜΤΣ. Γενικά διασπορά και μεροληψία αλληλοεπιδρούν ανταγωνιστικά· και προς επαλήθευση αυτού του φαινομένου ας δούμε το επόμενο παράδειγμα.

Παράδειγμα 4.1.3. (**κανονική κατανομή με γνωστό συντελεστή μεταβλητότητας - εκτίμηση της μέσης τιμής**) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την κανονική κατανομή $\mathcal{N}(\theta, \kappa\theta^2)$, όπου $\kappa > 0$ γνωστή σταθερά και $\theta \neq 0$ άγνωστη παράμετρος, $\theta \in \Theta = \mathbb{R} \setminus \{0\}$. Ο συντελεστής μεταβλητότητας (δείκτης του μεγέθους της τυπικής απόκλισης σε σχέση με τη μέση τιμή) είναι ο ελεύθερος μονάδων αριθμός $\sqrt{\mathbb{V}ar_\theta X_1}/|\mathbb{E}_\theta X_1|$, που εδώ συμπίπτει με το κ. Θεωρούμε εκτίμηση του θ (δηλαδή $g(\theta) = \theta$ στην προκειμένη περίπτωση) με κριτήριο το ΜΤΣ. Επειδή το θ παριστάνει τη μέση τιμή της κατανομής, $\mathbb{E}_\theta X_1 = \theta$, όπως ήδη ξέρουμε (Ενότητα 3.3.1α), ένας εκτιμητής του είναι το δειγματικό του ανάλογο, ο δειγματικός μέσος $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Θα υπολογίσουμε, αρχικά, το ΜΤΣ του $\bar{X}$ μέσω της σχέσης (4.3). Έχουμε,

$$\mathbb{V}ar_\theta \bar{X} = \mathbb{V}ar_\theta \left(\frac{1}{n} \sum_{i=1}^{n} X_i \right) = \frac{1}{n^2} \sum_{i=1}^{n} \mathbb{V}ar_\theta X_i = \frac{1}{n^2} \sum_{i=1}^{n} \kappa\theta^2 = \frac{\kappa\theta^2}{n},$$

ενώ

$$\begin{aligned} b(\bar{X}, \theta) = \mathbb{E}_\theta \bar{X} - \theta = \mathbb{E}_\theta \left(\frac{1}{n} \sum_{i=1}^{n} X_i \right) - \theta = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}_\theta X_i - \theta \\ = \frac{1}{n} \sum_{i=1}^{n} \theta - \theta = \theta - \theta = 0 \ \text{ (μηδενική μεροληψία).} \end{aligned}$$

Άρα από τη σχέση (4.3), προκύπτει ότι

$$\text{ΜΤΣ}(\bar{X}, \theta) = \mathbb{V}ar_\theta \bar{X} + b^2(\bar{X}, \theta) = \mathbb{V}ar_\theta \bar{X} = \frac{\kappa\theta^2}{n}.$$

Σύμφωνα με τη συζήτηση, που προηγήθηκε αυτού του παραδείγματος, ο δειγματικός μέσος $\bar{X}$ εκπροσωπεί την επιλογή εκτιμητή μηδενικής μεροληψίας.

Ας εξετάσουμε τώρα κατά πόσον τροποποιώντας τον $\bar{X}$ μπορούμε να μειώσουμε τη διασπορά του με την ελπίδα να μην αυξηθεί σημαντικά η μεροληψία, ώστε το τελικό «προϊόν» να είναι εκτιμητής με μικρότερο ΜΤΣ από αυτό του $\bar{X}$. Δεν είναι καθόλου προφανές, όμως δεν θα είχε κανείς αντίρρηση να «δοκιμάσουμε» εκτιμητές της μορφής $T_c = c\bar{X}$, όπου c σταθερά με $c \neq 0$ και $|c| < 1$, γιατί, τότε

$$\mathbb{V}ar_\theta(c\bar{X}) = c^2 \mathbb{V}ar_\theta \bar{X} < \mathbb{V}ar_\theta \bar{X}.$$

Όντως λοιπόν οι εκτιμητές της μορφής T_c επιτυγχάνουν μείωση της διασποράς. Από την άλλη πλευρά, εισάγουν μεροληψία, αφού

$$b(T_c, \theta) = \mathbb{E}_\theta T_c - \theta = \mathbb{E}_\theta(c\bar{X}) - \theta = c\,\mathbb{E}_\theta \bar{X} - \theta = c\theta - \theta = (c-1)\theta \neq 0$$

με απόλυτη τιμή $(1-c)|\theta|$. Επιπροσθέτως, κατανοούμε ότι για να μην αυξηθεί σημαντικά η μεροληψία, το c πρέπει να επιλεγεί «κοντά» στο 1, για αυτό το λόγο ας θεωρήσουμε προς στιγμή ότι $0 < c < 1$ (ας μη ξεχνάμε ότι $|c| < 1$). Τότε η διασπορά $\mathbb{V}ar_\theta T_c = c^2 \mathbb{V}ar_\theta \bar{X}$ είναι αύξουσα συνάρτηση του c, μεταβαλλόμενη από 0 για $c \to 0+$ έως $\mathbb{V}ar_\theta \bar{X}$ για $c \to 1-$, ενώ η απόλυτη τιμή της μεροληψίας, $|b(T_c, \theta)| = (1-c)|\theta|$, είναι φθίνουσα ως προς c, μεταβαλλόμενη από 0 για $c \to 1-$ έως $|\theta|$

για $c \to 0+$. Επιβεβαιώνεται δηλαδή το φαινόμενο η διασπορά και η μεροληψία να μεταβάλλονται προς αντίθετες κατευθύνσεις. Είμαστε τώρα σε θέση να μελετήσουμε το ΜΤΣ των εκτιμητών της μορφής T_c και να εξετάσουμε κατά πόσον κάποιος εξ αυτών έχει μικρότερο ΜΤΣ από τον $\bar{X}$. Για κάθε c, έχουμε

$$\text{ΜΤΣ}(T_c, \theta) = \mathbb{V}ar_\theta T_c + b^2(T_c, \theta) = c^2\kappa\frac{\theta^2}{n} + (c-1)^2\theta^2,$$

οπότε

$$\text{ΜΤΣ}(T_c, \theta) < \text{ΜΤΣ}(\bar{X}, \theta), \quad \forall\, \theta \neq 0$$

εάν και μόνον εάν

$$c^2\kappa\frac{\theta^2}{n} + (c-1)^2\theta^2 < \kappa\frac{\theta^2}{n}, \quad \forall\, \theta \neq 0.$$

Αυτή η δευτεροβάθμια ως προς c ανίσωση ισοδύναμα γράφεται

$$\left(1 + \frac{\kappa}{n}\right)c^2 - 2c + 1 - \frac{\kappa}{n} < 0.$$

Προκύπτει εύκολα ότι παραπάνω ανισότητα αληθεύει για $\frac{n-\kappa}{n+\kappa} < c < 1$ και αυτές οι τιμές του c ικανοποιούν επίσης και την αρχική συνθήκη $|c| < 1$. Συμπερασματικά, για κάθε σταθερά c (που δεν εξαρτάται από το θ, γιατί διαφορετικά, εξ' ορισμού, ο T_c δεν θα ήταν στατιστική συνάρτηση) με $c \neq 0$ και $\frac{n-\kappa}{n+\kappa} < c < 1$, ο εκτιμητής $T_c = c\bar{X}$ έχει μεν μη μηδενική μεροληψία, αλλά μικρότερη διασπορά και μικρότερο ΜΤΣ από τον μηδενικής μεροληψίας κλασικό εκτιμητή της μέσης τιμής θ, τον δειγματικό μέσο $\bar{X}$.

Ένα σχόλιο πριν κλείσουμε αυτό το παράδειγμα. Όλοι οι εκτιμητές της μορφής $T_c = c\bar{X}$ με $c \neq 0$ και $\frac{n-\kappa}{n+\kappa} < c < 1$ έχουν μικρότερο ΜΤΣ από τον $\bar{X}$. Ποιος όμως από αυτούς έχει το μικρότερο; Η απάντηση μπορεί εύκολα να δοθεί ελαχιστοποιώντας ως προς c το

$$\text{ΜΤΣ}(T_c, \theta) = c^2\kappa\frac{\theta^2}{n} + (c-1)^2\theta^2.$$

Η ελαχιστοποιούσα τιμή του c είναι $c_0 = \frac{n}{n+\kappa}$ και, επειδή $\frac{n-\kappa}{n+\kappa} < c_0 < 1$, ο εκτιμητής $T_{c_0} = \frac{n}{n+\kappa}\bar{X}$ έχει το ελάχιστο ΜΤΣ ίσο προς $\frac{\kappa}{n+\kappa}\theta^2$. □

Το προηγούμενο παράδειγμα μπορεί εύκολα να γενικευθεί ως εξής.

Παράδειγμα 4.1.4. (εκτιμητές ειδικής μορφής) Έστω T εκτιμητής της παραμέτρου θ με $\mathbb{E}_\theta T = \theta$ και $\mathbb{V}ar_\theta\ T = a\theta^2$, όπου $a > 0$ γνωστή σταθερά και θ άγνωστη παράμετρος, $\theta \in \Theta \subset \mathbb{R} - \{0\}$. Τότε, θεωρώντας εκτιμητές της μορφής $T_c = cT$, όπου c σταθερά, το ΜΤΣ του T_c, λόγω της (4.3), είναι

$$\text{ΜΤΣ}(T_c, \theta) = \mathbb{V}ar_\theta T_c + b^2(T_c, \theta) = c^2 a\theta^2 + (c-1)^2\theta^2,$$

που ελαχιστοποιείται ως προς c για την τιμή $c_0 = \frac{1}{1+a}$. Από τον ορισμό του c_0, έχουμε, ειδικά, ότι

$$\text{ΜΤΣ}(T_{c_0}, \theta) < \text{ΜΤΣ}(T_1, \theta) = \text{ΜΤΣ}(T, \theta), \quad \forall\, \theta \neq 0.$$

Το Παράδειγμα 4.1.3 αντιστοιχεί στην περίπτωση $T = \bar{X}$ και $a = \frac{\kappa}{n}$. Μία άλλη ειδική περίπτωση είναι η ακόλουθη: Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την εκθετική κατανομή $\mathcal{E}(\theta)$, όπου $\theta \in \Theta = (0, \infty)$ άγνωστη παράμετρος. Επειδή $\mathbb{E}_\theta X_1 = \theta$, θεωρούμε ως εκτιμητή του θ τον δειγματικό μέσο $T = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ (Ενότητα 3.3.1α) και παρατηρούμε ότι

$$\mathbb{E}_\theta T = \mathbb{E}_\theta\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_\theta X_i = \frac{1}{n}\sum_{i=1}^{n}\theta = \theta,$$

ενώ

$$\mathbb{V}ar_\theta T = \mathbb{V}ar_\theta\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}ar_\theta X_i = \frac{1}{n^2}\sum_{i=1}^{n}\theta^2 = \frac{\theta^2}{n}.$$

Επομένως, το παραπάνω γενικό αποτέλεσμα ισχύει με $a = \frac{1}{n}$ και $c_0 = \frac{1}{1+1/n} = \frac{n}{n+1}$, δηλαδή ο εκτιμητής $T_{c_0} = \frac{n}{n+1}\bar{X} = \frac{1}{n+1}\sum_{i=1}^{n} X_i$ έχει μικρότερο ΜΤΣ από τον $\bar{X}$. □

Όπως είπαμε παραπάνω και όπως διαπιστώνεται από τα Παραδείγματα 4.1.2 - 4.1.4, το ΜΤΣ είναι εν γένει συνάρτηση του $\theta \in \Theta$. Έτσι, για να «ανακηρύξουμε» έναν εκτιμητή καλύτερο ενός άλλου, και επειδή δεν γνωρίζουμε την τιμή του θ, «αναγκαζόμαστε» να τους συγκρίνουμε *για κάθε* δυνατή τιμή του θ στον παραμετρικό χώρο Θ. Δίνεται λοιπόν ο εξής ορισμός.

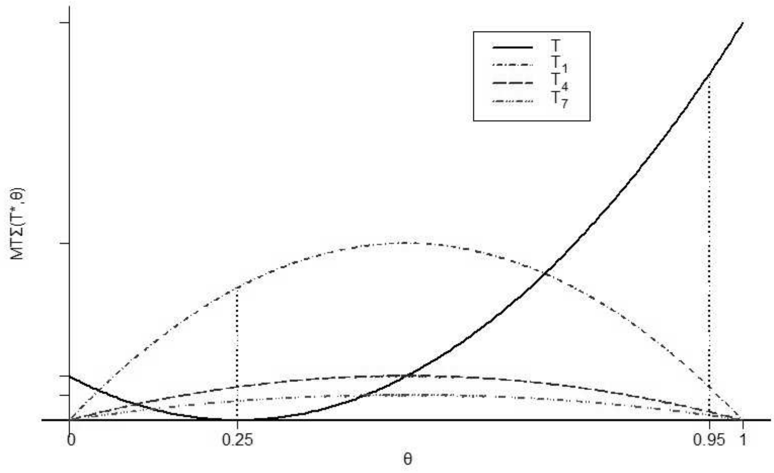
Ορισμός 4.1.2. *Για την εκτίμηση του $g(\theta)$, ένας εκτιμητής $T_1(\underset{\sim}{X})$ λέγεται καλύτερος από τον εκτιμητή $T_2(\underset{\sim}{X})$ με κριτήριο το ΜΤΣ, αν*

(α) ΜΤΣ$(T_1, \theta) \leq$ ΜΤΣ(T_2, θ), $\forall\, \theta \in \Theta$

(β) ΜΤΣ$(T_1, \theta_0) <$ ΜΤΣ(T_2, θ_0), για κάποιο $\theta_0 \in \Theta$.

Σε αυτήν την περίπτωση ο $T_2(\underset{\sim}{X})$ λέγεται μη αποδεκτός εκτιμητής. Ένας εκτιμητής $T(\underset{\sim}{X})$ λέγεται αποδεκτός, εάν δεν υπάρχει καλύτερος του εκτιμητής.

Είναι κατανοητό ότι δύο εκτιμητές $T_1(\underset{\sim}{X})$ και $T_2(\underset{\sim}{X})$ μπορεί να μην είναι συγκρίσιμοι με βάση τον Ορισμό 4.1.2. Αυτό θα συμβεί, αν το ΜΤΣ του $T_1(\underset{\sim}{X})$ είναι μικρότερο από το ΜΤΣ του $T_2(\underset{\sim}{X})$ για κάποιες τιμές $\theta \in \Theta$ και το ΜΤΣ του T_2 μικρότερο από το ΜΤΣ του $T_1(\underset{\sim}{X})$ για κάποιες άλλες.



Σχήμα 4.3: ΜΤΣ των $T(\underset{\sim}{X}) = 0.25$ και $T_n(\underset{\sim}{X}) = \bar{X}$, $n = 1, 4, 7$

Παράδειγμα 4.1.5. (Συνέχεια του Παραδείγματος 4.1.2 - μη συγκρίσιμοι εκτιμητές) Στο Παράδειγμα 4.1.2, θέτοντας $T_1(\underset{\sim}{X}) = \bar{X}$ και $T_2(\underset{\sim}{X}) = 0.25$, για $\theta = 0.25$, έχουμε

$$\text{ΜΤΣ}(\bar{X}, 0.25) = \frac{1}{4n} \text{ και } \text{ΜΤΣ}(T(\underset{\sim}{X}), 0.25) = 0,$$

ενώ για $\theta = 0.95$,

$$\text{ΜΤΣ}(\bar{X}, 0.95) = \frac{0.0475}{n}$$

και

$$\text{ΜΤΣ}(T(\underset{\sim}{X}), 0.95) = (0.25 - 0.95)^2 = 0.49 > \frac{0.475}{n}, \ \forall\, n = 1, 2, \ldots,$$

δηλαδή ο $\bar{X}$ και $T(\underset{\sim}{X}) = 0.25$ είναι μη συγκρίσιμοι (βλέπε Σχήμα 4.3). $\square$

Αναφερόμενοι στο Παράδειγμα 4.1.1, ο $\bar{X}$ είναι αποδεκτός εκτιμητής του θ, όμως η απόδειξη αυτού του αποτελέσματος είναι πολύ πέραν του σκοπού αυτών των σημειώσεων (βλέπε Lehmann and Casella, 1998, σελ. 335). Επίσης, στο Παράδειγμα 4.1.3 ο $\bar{X}$ δεν είναι αποδεκτός εκτιμητής του θ, αφού ο $T_c = c\bar{X}$, $c \neq 0$ και $\frac{n-\kappa}{n+\kappa} < c < 1$ έχει μικρότερο ΜΤΣ. Παρόμοια, στο Παράδειγμα 4.1.4, ο $\frac{1}{\alpha+1}T$ είναι καλύτερος από τον T με κριτήριο το ΜΤΣ. Η αποδεκτικότητα ενός εκτιμητή είναι μια ελάχιστη ιδιότητα-απαίτηση που θα ζητούσαμε από έναν υποψήφιο εκτιμητή με κριτήριο το ΜΤΣ. Πρακτικά, η αποδεκτικότητα σημαίνει απλά και μόνον ότι ο εκτιμητής δεν μπορεί να «αποκλειστεί» ή να «εξαιρεθεί» από το σύνολο των εκτιμητών του $g(\theta)$. Για αυτό τον λόγο η αποδεκτικότητα συχνά συνδυάζεται και με την αναζήτηση άλλων ιδιοτήτων, όπως να είναι ο εκτιμητής Bayes ή minimax (βλέπε σχετικό σχόλιο στην Ενότητα 3.2). Από την άλλη πλευρά, ένας μη αποδεκτός εκτιμητής απορρίπτεται με κριτήριο το ΜΤΣ, αφού υπάρχει καλύτερός του. Ας επιστρέψουμε τώρα στο Παράδειγμα 4.1.2, για να συζητήσουμε την αποδεκτικότητα των εκτιμητών $\bar{X}$ και $T(\underset{\sim}{X}) = 0.25$.

Παράδειγμα 4.1.6. (συνέχεια του Παραδείγματος 4.1.2- αποδεκτικότητα σταθεράς) Θα δείξουμε ότι ο $T(\underset{\sim}{X}) = 0.25$, παρότι παράδοξος εκτιμητής, είναι αποδεκτός εκτιμητής του θ, χρησιμοποιώντας άτοπο απαγωγή. Έστω ότι ο $T(\underset{\sim}{X}) = 0.25$ δεν είναι αποδεκτός. Τότε από τον Ορισμό 4.1.2 υπάρχει εκτιμητής $T_1(\underset{\sim}{X})$, που δεν συμπίπτει με τον $T(\underset{\sim}{X}) = 0.25$, τέτοιος ώστε

$$\text{ΜΤΣ}(T_1(\underset{\sim}{X}), \theta) \leqslant \text{ΜΤΣ}(T(\underset{\sim}{X}), \theta), \quad \forall\, \theta \in (0, 1).$$

Επειδή $\text{ΜΤΣ}(T(\underset{\sim}{X}), \theta) = (\theta - 0.25)^2$, για $\theta = 0.25$ η ανισότητα γίνεται

$$\text{ΜΤΣ}(T_1(\underset{\sim}{X}), 0.25) \leqslant (0.25-0.25)^2 = 0 \ \text{ δηλαδή } \ \mathbb{E}_{\theta=0.25}(T_1(\underset{\sim}{X}) - 0.25)^2 \leqslant 0,$$

το οποίο σημαίνει ότι

$$\mathbb{E}_{0.25}(T_1(\underset{\sim}{X}) - 0.25)^2 = 0 \ \text{ και επομένως } \ \mathbb{P}_{\theta=0.25}(T_1(\underset{\sim}{X}) = 0.25) = 1.$$

Η τελευταία ισότητα προκύπτει επειδή η σχέση $\mathbb{E}\, Y^2 = 0$ συνεπάγεται ότι $\mathbb{P}(Y = 0) = 1$. Καταλήξαμε λοιπόν ότι, για $\theta = 0.25$, ο εκτιμητής $T_1(\underset{\sim}{X})$ συμπίπτει με τη σταθερά 0.25, δηλαδή με τον εκτιμητή $T(\underset{\sim}{X})$, όπερ άτοπο. Με το ίδιο ακριβώς σκεπτικό, οποιαδήποτε άλλη σταθερά $\theta_0 \in (0, 1)$ είναι επίσης αποδεκτός εκτιμητής του θ. Για τον $\bar{X}$ αναφέρουμε ότι είναι αποδεκτός, όμως η απόδειξη, όπως και στην περίπτωση του Παραδείγματος 4.1.1, ξεπερνά κατά πολύ τα όρια αυτών των σημειώσεων (βλέπε Lehmann and Casella, 1998).

Ιδανικά, θα θέλαμε να βρούμε ένα βέλτιστο εκτιμητή, δηλαδή έναν εκτιμητή που είναι καλύτερος από κάθε άλλον με το κριτήριο του Ορισμού 4.1.2. Όμως, ένας τέτοιος εκτιμητής, έστω $T^*(\underset{\sim}{X})$ αν υπήρχε, θα έπρεπε να ικανοποιεί τη σχέση

$$\mathbb{P}_\theta \Big(T^*(\underset{\sim}{X}) = g(\theta) \Big) = 1 \ , \quad \forall\, \theta \in \Theta, \tag{4.4}$$

δηλαδή να εκτιμά το $g(\theta)$ χωρίς σφάλμα, το οποίο είναι αδύνατο να συμβεί, εκτός από κάποιες τετριμμένες περιπτώσεις, όπως αυτή του Παραδείγματος 4.1.7 που δίνεται παρακάτω. Ας δούμε γιατί πρέπει να ισχύει η (4.4). Θεωρούμε αυθαίρετο $\theta_0 \in \Theta$ και τον εκτιμητή $T(\underset{\sim}{X}) = g(\theta_0)$, ο οποίος αγνοώντας τα δεδομένα $\underset{\sim}{X}$ εκτιμά το $g(\theta)$ με την τιμή $g(\theta_0)$. Συγκρίνοντας το βέλτιστο εκτιμητή $T^*(\underset{\sim}{X})$ με τον $T(\underset{\sim}{X})$ έχουμε ότι

$$\text{ΜΤΣ}(T^*(\underset{\sim}{X}), \theta) \leqslant \text{ΜΤΣ}(T(\underset{\sim}{X}), \theta) \quad \forall\, \theta \in \Theta$$

ή, ισοδύναμα

$$\mathbb{E}_\theta \big(T^*(\underset{\sim}{X}) - g(\theta) \big)^2 \leqslant \mathbb{E}_\theta (g(\theta_0) - g(\theta))^2 = (g(\theta_0) - g(\theta))^2 \ , \ \ \forall\, \theta \in \Theta.$$

Για $\theta = \theta_0$, η τελευταία ανισότητα γίνεται

$$\mathbb{E}_{\theta_0} (T^*(\underset{\sim}{X}) - g(\theta_0))^2 \leqslant 0, \ \ \text{άρα, κατ' ανάγκη,} \ \ \mathbb{E}_{\theta_0} (T^*(\underset{\sim}{X}) - g(\theta_0))^2 = 0,$$

οπότε, όπως στο Παράδειγμα 4.1.6, έχουμε

$$\mathbb{P}_{\theta_0} \Big(T^*(\underset{\sim}{X}) = g(\theta_0) \Big) = 1$$

που είναι η (4.4), αφού το θ_0 είναι οποιοδήποτε σημείο του Θ.

Παράδειγμα 4.1.7. (ύπαρξη βέλτιστου εκτιμητή) Έστω ότι $\theta \in \Theta = \{0, 1\}$, $g(\theta) = \theta$ και ότι τα δεδομένα είναι μία μόνον παρατήρηση, X, με κατανομή $\mathbb{P}_\theta(X = \theta) = 1$, $\theta \in \Theta$. Τότε, ο εκτιμητής $T^*(X) = X$ είναι βέλτιστος εκτιμητής του θ, αφού $\mathbb{P}_\theta(T^*(X) = \theta) = 1$, $\theta \in \Theta$.

Το Παράδειγμα 4.1.7 είναι μία ακραία και μη ρεαλιστική κατάσταση, αφού τα δεδομένα X παίρνουν μία και μόνον τιμή που συμπίπτει με την υπό εκτίμηση παράμετρο θ και αποτελεί μη ενδιαφέρουσα εξαίρεση του κανόνα μη ύπαρξης βέλτιστου εκτιμητή. Ήταν ευθύς εξ αρχής, πολύ φιλόδοξη η ιδέα αναζήτησης βέλτιστου εκτιμητή μεταξύ *όλων* των εκτιμητών: το ΜΤΣ είναι συνάρτηση του $\theta \in \Theta$, άρα η γραφική παράσταση του ΜΤΣ του βέλτιστου εκτιμητή θα έπρεπε να είναι *κάτω* από την αντίστοιχη *οποιουδήποτε* άλλου εκτιμητή, πράγμα αδύνατο λόγω της *πληθώρας* των διαθέσιμων εκτιμητών. Διατηρώντας το κριτήριο του Ορισμού 4.1.1, ένας τρόπος να αντιμετωπίσουμε το πρόβλημα της μη ύπαρξης βέλτιστου εκτιμητή είναι να περιορίσουμε την κλάση των εκτιμητών, απαιτώντας ο εκτιμητής $T(\utilde{X})$ του $g(\theta)$ να έχει ειδική μορφή, όπως π.χ. οι εκτιμητές T_c στα Παραδείγματα 4.1.3 και 4.1.4 ή να ικανοποιεί κάποια «λογική» συνθήκη και μέσα στην μικρότερη αυτή κλάση εκτιμητών να αναζητήσουμε τον καλύτερο με κριτήριο το ΜΤΣ. Μία τέτοια συνθήκη που πρωτοείδαμε ως συνθήκη μηδενικής μεροληψίας είναι η

$$\mathbb{E}_\theta T(\utilde{X}) = g(\theta), \quad \forall\, \theta \in \Theta.$$

Η συνθήκη απαιτεί από τον εκτιμητή $T(\utilde{X})$ να παίρνει «κατά μέσο όρο» τιμή ίση προς το $g(\theta)$, δηλαδή ίση προς την τιμή που έχει κληθεί να εκτιμήσει, και αυτό να συμβαίνει όποια και αν είναι η άγνωστη τιμή θ. Η συνθήκη της αμεροληψίας αποκλείει, π.χ., τις σταθερές ως εκτιμητές του $g(\theta)$ (όπως αυτήν των Παραδειγμάτων 4.1.2 και 4.1.6) τις οποίες ούτως ή άλλως δεν θα θέλαμε να χρησιμοποιήσουμε. Αυτήν τη συνθήκη τη μελετάμε στην αμέσως επόμενη ενότητα. Ακολούθως στα Κεφάλαια 4 και 5, η αμεροληψία συνδυάζεται με το κριτήριο του ΜΤΣ με σκοπό την εύρεση του καλύτερου εκτιμητή, μεταξύ όλων των αμερόληπτων εκτιμητών.

4.2 Αμεροληψία

Η μεροληψία $b(T, \theta) = \mathbb{E}_\theta T(\underset{\sim}{X}) - g(\theta)$ ενός εκτιμητή $T(\underset{\sim}{X})$ της τιμής $g(\theta)$ μπορεί να είναι για δοθέν $\theta \in \Theta$, αρνητική, θετική ή μηδέν. Αρνητική μεροληψία, $\mathbb{E}_\theta T(\underset{\sim}{X}) < g(\theta)$, ερμηνεύεται ως «τάση» του εκτιμητή $T(\underset{\sim}{X})$ να υποεκτιμά το $g(\theta)$, ενώ αντίθετα θετική μεροληψία, $\mathbb{E}_\theta T(\underset{\sim}{X}) > g(\theta)$, υποδηλώνει την «τάση» του εκτιμητή $T(\underset{\sim}{X})$ να υπερεκτιμά το $g(\theta)$. Όπως αναφέρεται στην Ενότητα 4.1, η μηδενική μεροληψία ή αμεροληψία , δηλαδή $\mathbb{E}_\theta T(\underset{\sim}{X}) = g(\theta)$, έχει την ερμηνεία ότι «κατά μέσο όρο» ο εκτιμητής $T(\underset{\sim}{X})$ παίρνει τιμή που συμπίπτει με την προς εκτίμηση τιμή $g(\theta)$. Η μεροληψία, αρνητική ή θετική, σε ορισμένες περιπτώσεις πρέπει να αποφεύγεται (π.χ. η υπερεκτίμηση της πιθανότητας επιτυχίας μιας χειρουργικής επέμβασης μπορεί να αποβεί ολέθρια για τον ασθενή), ενώ σε άλλες περιπτώσεις μπορεί να είναι χρήσιμη, όπως αν ο επιδιωκόμενος σκοπός είναι η μείωση του ΜΤΣ (βλέπε τα Παραδείγματα 4.1.3, 4.1.4, αλλά και την Πρόταση 4.3.1). Από την άλλη πλευρά, η αμεροληψία φαντάζει ως μια λογική και ουδέτερη συμπεριφορά του εκτιμητή, η οποία σε κάποιες περιπτώσεις είναι επιθυμητή είτε λόγω του υποκείμενου φυσικού προβλήματος είτε υπό τον «φόβο» ενδεχόμενων αρνητικών συνεπειών από τη χρήση μεροληπτικού εκτιμητή. Αυτήν την ομολογουμένως ελκυστική σε ονομασία ιδιότητα της ουδετερότητας ορίζουμε και μελετάμε σε αυτήν την ενότητα. Η τελική αποτίμησή της, όμως, γίνεται στο τέλος του Κεφαλαίου 6 μετά την ολοκλήρωση της μελέτης των αμερόληπτων εκτιμητών ελάχιστης διασποράς.

Ορισμός 4.2.1. *Ένας εκτιμητής $T(\underset{\sim}{X})$ ονομάζεται αμερόληπτος εκτιμητής του $g(\theta)$, εάν*

$$\mathbb{E}_\theta T(\underset{\sim}{X}) = g(\theta), \ \ \forall\, \theta \in \Theta.$$

Από τον ορισμό προκύπτει αμέσως ότι κάθε στατιστική συνάρτηση είναι αμερόληπτος εκτιμητής της μέσης τιμής της (εφόσον αυτή εξαρτάται από το θ). Για παράδειγμα, αν $\underset{\sim}{X} = (X_1, \ldots, X_n)$ είναι ένα τυχαίο δείγμα από την κανονική κατανομή $\mathcal{N}(\theta, 1)$, $\theta \in \Theta = \mathbb{R}$, επειδή

$$\mathbb{E}_\theta X_i = \theta \quad \forall\, \theta \in \Theta,$$

συμπεραίνουμε ότι οι στατιστικές συναρτήσεις

$$T_1(\underset{\sim}{X}) = X_1,\ T_2(\underset{\sim}{X}) = X_2, \ldots, T_n(\underset{\sim}{X}) = X_n,$$

είναι αμερόληπτοι εκτιμητές του θ. Ομοίως, αν X είναι μια παρατήρηση από την κανονική κατανομή $\mathcal{N}(0,\theta^2)$, $\theta \in \Theta = (0,\infty)$, επειδή

$$\mathbb{E}_\theta X^2 = \mathbb{V}ar_\theta X + (\mathbb{E}_\theta X)^2 = \theta^2 \quad \forall\, \theta \in \Theta,$$

συμπεραίνουμε ότι η στατιστική συνάρτηση X^2 εκτιμά αμερόληπτα τη διασπορά θ^2. Επιπλέον, για να απαντήσουμε στο ερώτημα ποια τιμή εκτιμά αμερόληπτα η στατιστική συνάρτηση $\sqrt{X^2} = |X|$ (μήπως την τυπική απόκλιση $\sqrt{\theta^2} = \theta$;) δεν έχουμε παρά να βρούμε τη μέση τιμή $\mathbb{E}_\theta|X|$. Έχουμε λοιπόν

$$\mathbb{E}_\theta|X| = \int_{-\infty}^{\infty} |x| \frac{1}{\theta\sqrt{2\pi}} e^{-\frac{x^2}{2\theta^2}} dx = 2\int_0^{\infty} x \frac{1}{\theta\sqrt{2\pi}} e^{-\frac{x^2}{2\theta^2}} dx = \frac{\sqrt{2}}{\sqrt{\pi}}\theta, \quad \forall\, \theta \in \Theta$$

(ο υπολογισμός του ολοκληρώματος έχει γίνει στο Παράδειγμα 4.1.1), δηλαδή η στατιστική συνάρτηση $|X|$ εκτιμά αμερόληπτα την τιμή $\frac{\sqrt{2}}{\sqrt{\pi}}\theta$ (και όχι την τυπική απόκλιση θ). Αντίστροφα, δοθείσης συνάρτησης g, μεγαλύτερο ενδιαφέρον παρουσιάζει το ερώτημα κατά πόσον υπάρχει αμερόληπτος εκτιμητής της τιμής $g(\theta)$ και αν υπάρχει, πόσοι άλλοι αμερόληπτοι εκτιμητές υπάρχουν. Η απάντηση δίνεται στην επόμενη πρόταση.

Πρόταση 4.2.1. *(α) Εάν υπάρχουν δύο αμερόληπτοι εκτιμητές του $g(\theta)$, τότε υπάρχει ένα μη αριθμήσιμο πλήθος αμερόληπτων εκτιμητών του $g(\theta)$.* *(β) Το σύνολο των αμερόληπτων εκτιμητών του $g(\theta)$ είναι ή κενό ή μονοσύνολο ή μη αριθμήσιμο.*

Απόδειξη. (α) Έστω $T_1 \neq T_0$ δύο αμερόληπτοι εκτιμητές του $g(\theta)$. Για κάθε σταθερά $\alpha \in [0,1]$, ορίζουμε $T_\alpha = \alpha T_1 + (1-\alpha)T_0$ και παρατηρούμε ότι

$$\begin{aligned}
\mathbb{E}_\theta T_\alpha &= \mathbb{E}_\theta\left(\alpha T_1 + (1-\alpha)T_0\right) = \alpha\mathbb{E}_\theta T_1 + (1-\alpha)\mathbb{E}_\theta T_0 \\
&= \alpha g(\theta) + (1-\alpha)g(\theta) = g(\theta) \quad \forall\, \theta \in \Theta,
\end{aligned}$$

δηλαδή ο T_α είναι αμερόληπτος εκτιμητής του $g(\theta)$. Επιπροσθέτως, για $\alpha, \beta \in [0,1]$ με $\alpha \neq \beta$, έχουμε ότι $T_\alpha \neq T_\beta$. Συμπεραίνουμε λοιπόν ότι, αν υπάρχουν δύο αμερόληπτοι εκτιμητές, τότε υπάρχει ένα μη αριθμήσιμο σύνολο αμερόληπτων εκτιμητών, το

$$\{T_\alpha = \alpha T_1 + (1-\alpha)T_0 : \alpha \in [0,1]\}.$$

(β) Αν δεν υπάρχουν δύο αμερόληπτοι εκτιμητές του $g(\theta)$, τότε ή δεν υπάρχει κανένας ή υπάρχει μόνον ένας αμερόληπτος εκτιμητής του $g(\theta)$ και έτσι ολοκληρώνεται η απόδειξη της πρότασης. $\square$

Σημειώνουμε ότι η πλέον ενδιαφέρουσα περίπτωση από τις τρεις της Πρότασης 4.2.1 είναι η τρίτη, γιατί θέτει αμέσως το θέμα επιλογής του καλύτερου μεταξύ όλων των, μη αριθμήσιμα πολλών, αμερόληπτων εκτιμητών. Τα επόμενα δύο κεφάλαια ασχολούνται ακριβώς με αυτό το θέμα. Οι άλλες δύο περιπτώσεις καλύπτονται στα Παραδείγματα 4.2.3 και 4.2.5.

Είδαμε παραπάνω ότι, αν $X \sim \mathcal{N}(0,\theta^2)$, X^2 είναι αμερόληπτος εκτιμητής του θ^2, ενώ $\sqrt{X^2} = |X|$ δεν είναι αμερόληπτος εκτιμητής του $\sqrt{\theta^2} = \theta$ (αλλά του $\frac{\sqrt{2}}{\sqrt{\pi}}\theta$). Γενικά, λοιπόν, πρέπει να γνωρίζουμε ότι η αμεροληψία δεν μεταβιβάζεται σε μετασχηματισμούς του εκτιμητή ή του $g(\theta)$. Ο $T(\underset{\sim}{X})$ μπορεί να είναι αμερόληπτος εκτιμητής του $g(\theta)$, αλλά ο $h\big(T(\underset{\sim}{X})\big)$ μεροληπτικός για το $h\left(g(\theta)\right)$. Τέτοιοι μεροληπτικοί μετασχηματισμοί αμερόληπτων εκτιμητών δίνονται παρακάτω στην Παρατήρηση 4.2.1. Μια, όμως, σημαντική περίπτωση μετασχηματισμών, όπου μεταβιβάζεται η αμεροληψία, είναι αυτή των γραμμικών μετασχηματισμών και αυτό αποδεικνύεται στην ακόλουθη πρόταση.

Πρόταση 4.2.2. *(α) Έστω $T(\underset{\sim}{X})$ αμερόληπτος εκτιμητής του $g(\theta)$ και α, β σταθερές (μη εξαρτώμενες από το θ). Τότε $\alpha T(\underset{\sim}{X}) + \beta$ είναι αμερόληπτος εκτιμητής του $\alpha g(\theta) + \beta$.*
(β) Έστω $T_i(\underset{\sim}{X})$ αμερόληπτος εκτιμητής του $g_i(\theta)$, $i = 1,\ldots,\kappa$ και α_i, $i = 1,\ldots,\kappa$ σταθερές (μη εξαρτώμενες από το θ). Τότε $\alpha_1 T_1(\underset{\sim}{X}) + \cdots + \alpha_\kappa T_\kappa(\underset{\sim}{X})$ είναι αμερόληπτος εκτιμητής του $\alpha_1 g_1(\theta) + \cdots + \alpha_\kappa g_\kappa(\theta)$.

Απόδειξη. (α) Λόγω της γραμμικής ιδιότητας της μέσης τιμής έχουμε

$$\mathbb{E}_\theta(\alpha T(\underset{\sim}{X}) + \beta) = \alpha \mathbb{E}_\theta T(\underset{\sim}{X}) + \beta = \alpha g(\theta) + \beta, \quad \forall\, \theta \in \Theta.$$

(β) Λόγω της προσθετικής και της γραμμικής ιδιότητας της μέσης τιμής έχουμε

$$\begin{aligned}\mathbb{E}_\theta(\alpha_1 T_1(\underset{\sim}{X}) + \cdots + \alpha_\kappa T_\kappa(\underset{\sim}{X})) &= \alpha_1 \mathbb{E}_\theta T_1(\underset{\sim}{X}) + \cdots + \alpha_\kappa \mathbb{E}_\theta T_\kappa(\underset{\sim}{X}) \\ &= \alpha_1 g_1(\theta) + \cdots + \alpha_\kappa g_\kappa(\theta), \quad \forall\, \theta \in \Theta. \;\square\end{aligned}$$

Η αξία της Πρότασης 4.2.2 έγκειται στο γεγονός ότι παρέχει, κατ' ευθείαν, αμερόληπτο εκτιμητή οποιουδήποτε γραμμικού μετασχηματισμού του $g(\theta)$, έχοντας διαθέσιμο έναν αμερόληπτο εκτιμητή του $g(\theta)$. Επίσης, στην περίπτωση που το $g(\theta)$ είναι άθροισμα τιμών, $g(\theta) = \sum_{i=1}^{\kappa} \alpha_i g_i(\theta)$, η Πρόταση 4.2.2 ανάγει την εύρεση αμερόληπτου εκτιμητή του $g(\theta)$ στην εύρεση αμερόληπτων εκτιμητών των προσθετέων $g_i(\theta)$, που μπορεί να είναι πιο εύκολο να εκτιμηθούν αμερόληπτα. Σημειώνουμε ότι είναι απαραίτητο οι σταθερές να μην εξαρτώνται από την άγνωστη παράμετρο θ. Έτσι, αν $T(\underset{\sim}{X})$ είναι αμερόληπτος εκτιμητής του $g(\theta)$, τότε $2T(\underset{\sim}{X}) + 1$ είναι αμερόληπτος εκτιμητής του $2g(\theta) + 1$. Όμως, ο $\theta T(\underset{\sim}{X}) + 1$ δεν είναι καν εκτιμητής του $\theta g(\theta) + 1$ επειδή δεν είναι στατιστική συνάρτηση, παρότι η τυχαία μεταβλητή $\theta T(\underset{\sim}{X}) + 1$ ικανοποιεί τη σχέση

$$\mathbb{E}_\theta(\theta T(\underset{\sim}{X}) + 1) = \theta g(\theta) + 1, \quad \forall\, \theta \in \Theta.$$

Παρατήρηση 4.2.1. Έστω $T(\underset{\sim}{X})$ αμερόληπτος εκτιμητής του $g(\theta)$, οπότε ο $T(\underset{\sim}{X})$ δεν είναι σταθερά (αν ήταν, π.χ. $T(\underset{\sim}{X}) = 5$, τότε $g(\theta) = \mathbb{E}_\theta T(\underset{\sim}{X}) = 5$ και δεν θα είχε νόημα το πρόβλημα εκτίμησης του $g(\theta)$). Θεωρώντας τον $T^2(\underset{\sim}{X})$ ως εκτιμητής του $g^2(\theta)$ παρατηρούμε ότι η μεροληψία του είναι

$$\begin{aligned}b(T^2, \theta) &= \mathbb{E}_\theta T^2(\underset{\sim}{X}) - g^2(\theta) = \mathbb{E}_\theta T^2(\underset{\sim}{X}) - \Big(\mathbb{E}_\theta T(\underset{\sim}{X})\Big)^2 \\ &= \mathbb{V}ar_\theta T(\underset{\sim}{X}) > 0, \quad \forall\, \theta \in \Theta.\end{aligned}$$

Ενώ δηλαδή ο $T(\underset{\sim}{X})$ είναι αμερόληπτος εκτιμητής του $g(\theta)$, ο $T^2(\underset{\sim}{X})$ είναι θετικά μεροληπτικός εκτιμητής του $g^2(\theta)$, για κάθε $\theta \in \Theta$. Γενικότερα, αν h είναι μια αυστηρά κυρτή συνάρτηση, από την ανισότητα Jensen (Πρόταση 1.5.2) έχουμε

$$\mathbb{E}_\theta h\big(T(\underset{\sim}{X})\big) > h\Big(\mathbb{E}_\theta T(\underset{\sim}{X})\Big) = h\left(g(\theta)\right),$$

δηλαδή ο $h\big(T(\underset{\sim}{X})\big)$ είναι θετικά μεροληπτικός εκτιμητής του $h\left(g(\theta)\right)$. Ομοίως, αν h είναι μια αυστηρά κοίλη συνάρτηση, ο $h\big(T(\underset{\sim}{X})\big)$ είναι αρνητικά μεροληπτικός εκτιμητής του $h\left(g(\theta)\right)$. Η Πρόταση 4.2.2 είναι λοιπόν μια εξαίρεση του κανόνα ότι η αμεροληψία δεν μεταβιβάζεται σε μετασχηματισμούς. □

Στις επόμενες προτάσεις παραθέτουμε αμερόληπτους εκτιμητές σημαντικών παραμέτρων μιας κατανομής, όπως η μέση τιμή και η διασπορά, αλλά και της συνάρτησης κατανομής καθώς και της πιθανότητας δοθέντος ενδεχομένου. Οι προτάσεις αυτές είναι μη παραμετρικού χαρακτήρα, ανήκουν δηλαδή στην περιοχή της Μη Παραμετρικής Στατιστικής Συμπερασματολογίας, αφού δεν γίνεται καμμία υπόθεση για τον (μαθηματικό) τύπο της πυκνότητας της κατανομής.

Πρόταση 4.2.3. *(**αμερόληπτος εκτιμητής της μέσης τιμής μιας κατανομής**) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από μία αυθαίρετη κατανομή με πυκνότητα $f_1(x;\theta)$, $\theta \in \Theta$ και $g(\theta) = \mathbb{E}_\theta X_1 = \mu$ η μέση τιμή της κατανομής.*
(α) Για κάθε $n \geq 1$ ο δειγματικός μέσος $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ είναι αμερόληπτος εκτιμητής του μ, δηλαδή $\mathbb{E}_\theta \bar{X} = \mu$ για κάθε $\theta \in \Theta$.
(β) $\mathbb{V}ar_\theta \bar{X} = \frac{\sigma^2}{n}$, όπου $\sigma^2 = \mathbb{V}ar_\theta X_1$ είναι η διασπορά της κατανομής.

Απόδειξη. Από τις ιδιότητες της μέσης τιμής έχουμε

$$\begin{aligned}\mathbb{E}_\theta \bar{X} &= \mathbb{E}_\theta \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\mathbb{E}_\theta \left(\sum_{i=1}^{n} X_i\right) \\ &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_\theta X_i = \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{1}{n} n\mu = \mu, \quad \forall\, \theta \in \Theta.\end{aligned}$$

Ανάλογα,

$$\mathbb{V}ar_\theta \bar{X} = \mathbb{V}ar_\theta\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\mathbb{V}ar_\theta\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}ar_\theta X_i$$
$$= \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}. \qquad \square$$

Παρατήρηση 4.2.2. Για την απόδειξη της αμεροληψίας του $\bar{X}$ χρησιμοποιήσαμε μόνο το γεγονός ότι $\mathbb{E}_\theta X_i = \mu,\ i = 1, \ldots, n$, ενώ δεν χρειάστηκε η ανεξαρτησία ή η κοινή κατανομή των $X_1, \ldots, X_n$. Επομένως, ο δειγματικός μέσος $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ είναι αμερόληπτος εκτιμητής του μ, ακόμη και αν οι παρατηρήσεις $X_1, \ldots, X_n$ είναι εξαρτημένες με διαφορετικές κατανομές αλλά με κοινή μέση τιμή μ.

Πρόταση 4.2.4. *(***αμερόληπτος εκτιμητής της διασποράς μιας κατανομής***) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$, ένα τυχαίο δείγμα από μία αυθαίρετη κατανομή με πυκνότητα $f_1(x;\theta)$, $\theta \in \Theta$ και $g(\theta) = \mathbb{E}_\theta\left(X_1 - \mathbb{E}_\theta X_1\right)^2 = \sigma^2$ η διασπορά της κατανομής.*

(α) Για κάθε $n \geq 1$, εάν η μέση τιμή $\mathbb{E}_\theta X_1 = \mu$ είναι γνωστή σταθερά, τότε $S_\mu^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$ είναι αμερόληπτος εκτιμητής του σ^2, δηλαδή $\mathbb{E}_\theta S_\mu^2 = \sigma^2$ για κάθε $\theta \in \Theta$.

(β) Για κάθε $n \geq 2$, η δειγματική διασπορά που ορίζεται από τη σχέση $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$ είναι αμερόληπτος εκτιμητής του σ^2, δηλαδή $\mathbb{E}_\theta S^2 = \sigma^2$ για κάθε $\theta \in \Theta$.

(γ) $\mathbb{V}ar_\theta S^2 = \frac{\mathbb{E}_\theta\left[(X_1 - \mathbb{E}_\theta X_1)^4\right]}{n} - \frac{(n-3)\sigma^4}{n(n-1)}$ $(n \geq 2)$.

Απόδειξη. (α) Επειδή το μ είναι γνωστή σταθερά (άρα μη εξαρτώμενη από την άγνωστη παράμετρο θ), η τυχαία μεταβλητή $S_\mu^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$ είναι στατιστική συνάρτηση, συνεπώς μπορεί να χρησιμοποιηθεί ως εκτιμητής του σ^2 και μάλιστα είναι το δειγματικό ανάλογο του σ^2 (βλέπε Ενότητα

3.3.1γ). Έχουμε ότι

$$
\begin{aligned}
\mathbb{E}_\theta S_\mu^2 &= \mathbb{E}_\theta \left(\frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2 \right) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}_\theta \left(X_i - \mu\right)^2 \\
&= \frac{1}{n} \sum_{i=1}^{n} \sigma^2 = \frac{1}{n} n\sigma^2 = \sigma^2, \forall\, \theta \in \Theta.
\end{aligned}
$$

(β) Ισχύει

$$
\begin{aligned}
\sum_{i=1}^{n} (X_i - \bar{X})^2 &= \sum_{i=1}^{n} (X_i^2 + \bar{X}^2 - 2\bar{X}X_i) = \sum_{i=1}^{n} X_i^2 + n\bar{X}^2 - 2\bar{X} \sum_{i=1}^{n} X_i \\
&= \sum_{i=1}^{n} X_i^2 + n\bar{X}^2 - 2\bar{X}n\bar{X} = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2.
\end{aligned}
$$

Χρησιμοποιώντας τη σχέση

$$
\mathbb{E}Y^2 = \mathbb{V}arY + (\mathbb{E}Y)^2,
$$

βρίσκουμε, από την Πρόταση 4.2.3, ότι

$$
\mathbb{E}X_i^2 = \mathbb{V}arX_i + (\mathbb{E}X_i)^2 = \sigma^2 + \mu^2 \quad \text{και} \quad \mathbb{E}\bar{X}^2 = \mathbb{V}ar\bar{X} + (\mathbb{E}\bar{X})^2 = \frac{\sigma^2}{n} + \mu^2.
$$

Επομένως, έχουμε

$$
\begin{aligned}
\mathbb{E}_\theta S^2 &= \frac{1}{n-1} \mathbb{E}_\theta \left(\sum_{i=1}^{n} (X_i - \bar{X})^2 \right) = \frac{1}{n-1} \mathbb{E}_\theta \left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2 \right) \\
&= \frac{1}{n-1} \left\{ \sum_{i=1}^{n} \mathbb{E}_\theta X_i^2 - n\mathbb{E}_\theta \bar{X}^2 \right\} \\
&= \frac{1}{n-1} \left\{ \sum_{i=1}^{n} (\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2 \right) \right\} \\
&= \frac{1}{n-1} \left\{ n(\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2 \right) \right\} \\
&= \frac{1}{n-1} (n-1)\sigma^2 = \sigma^2, \quad \forall\, \theta \in \Theta.
\end{aligned}
$$

(γ) Η απόδειξη παραλείπεται επειδή υπερβαίνει το επίπεδο αυτών των σημειώσεων. Ως άσκηση, με υποδείξεις, δίνεται στο βιβλίο Γ. Ηλιόπουλος (2013, σελ. 51). □

Παρατήρηση 4.2.3. Στην περίπτωση που η μέση τιμή $\mathbb{E}_\theta X_i = \mu$ εξαρτάται από την άγνωστη παράμετρο θ, η τυχαία μεταβλητή $\frac{1}{n}\sum\limits_{i=1}^{n}(X_i-\mu)^2$ δεν είναι στατιστική συνάρτηση, άρα πολύ περισσότερο δεν είναι εκτιμητής του σ^2. Πάντως, το γεγονός αυτό δεν αναιρεί τη σχέση

$$\mathbb{E}_\theta\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2\right) = \sigma^2, \quad \forall\, \theta \in \Theta,$$

που εξακολουθεί να ισχύει ως μέση τιμή της τυχαίας μεταβλητής $\frac{1}{n}\sum\limits_{i=1}^{n}(X_i-\mu)^2$. Η δειγματική διασπορά, S^2, μπορεί να θεωρηθεί ότι προκύπτει από την τυχαία μεταβλητή $\frac{1}{n}\sum\limits_{i=1}^{n}(X_i-\mu)^2$, αντικαθιστώντας το άγνωστο μ με τον εκτιμητή του, $\bar{X}$. Αυτή η αντικατάσταση, όμως, μεταβάλλει τη μέση τιμή σε

$$\mathbb{E}_\theta\left(\sum_{i=1}^{n}(X_i-\bar{X})^2\right) = (n-1)\sigma^2 \;\text{ από }\; \mathbb{E}_\theta\left(\sum_{i=1}^{n}(X_i-\mu)^2\right) = n\sigma^2,$$

οπότε η συνθήκη της αμεροληψίας επιβάλλει τον διαιρέτη $n-1$ αντί για n. Έτσι τελικά προκύπτει η δειγματική διασπορά $S^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\bar{X})^2$ ως αμερόληπτος εκτιμητής του σ^2, ενώ, αν χρησιμοποιήσουμε τον $S_*^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i-\bar{X})^2$, τότε

$$\mathbb{E}_\theta S_*^2 = \mathbb{E}_\theta\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2\right) = \frac{1}{n}(n-1)\sigma^2 < \sigma^2,$$

δηλαδή ο $S_*^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i-\bar{X})^2$ έχει αρνητική μεροληψία ,

$$b(S_*^2,\theta) = \mathbb{E}_\theta S_*^2 - \sigma^2 = \frac{n-1}{n}\sigma^2 - \sigma^2 = -\frac{\sigma^2}{n},$$

και συνεπώς την «τάση» να υποεκτιμά το σ^2. Προκειμένου λοιπόν να διορθωθεί αυτή η «τάση» υποεκτίμησης του σ^2, αυξάνουμε τον εκτιμητή S_*^2 διαιρώντας το $\sum\limits_{i=1}^{n}(X_i-\bar{X})^2$ με σταθερά μικρότερη του n και η «σωστή» σταθερά, ώστε να επιτευγχθεί η αμεροληψία είναι $n-1$.

Παραθέτουμε τώρα μερικά παραδείγματα εφαρμογής των Προτάσεων 4.2.3 και 4.2.4.

Παράδειγμα 4.2.1. (ομοιόμορφη κατανομή, συμμετρική γύρω από το μηδέν) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την ομοιόμορφη κατανομή στο διάστημα $(-\theta, \theta)$, $\mathcal{U}(-\theta, \theta)$, όπου $\theta \in \Theta = (0, \infty)$ είναι άγνωστη παράμετρος. Έχουμε ότι $\mathbb{E}_\theta X_1 = 0$ και $\mathbb{V}ar_\theta X_1 = \frac{\theta^2}{3}$. Η Πρόταση 4.2.4 ισχύει για αυθαίρετη κατανομή, άρα και για την $\mathcal{U}(-\theta, \theta)$. Επομένως οι εκτιμητές $S_0^2 = \frac{1}{n}\sum\limits_{i=1}^{n} X_i^2$ και $S^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \bar{X})^2$ είναι αμερόληπτοι εκτιμητές του $\frac{\theta^2}{3}$ και λόγω της Πρότασης 4.2.2 οι $3S_0^2$ και $3S^2$ είναι αμερόληπτοι εκτιμητές του θ^2. Εναλλακτικά, μπορούμε να πούμε ότι λόγω αμεροληψίας

$$\mathbb{E}_\theta S_0^2 = \frac{\theta^2}{3} \text{ ή } 3\,\mathbb{E}_\theta S_0^2 = \theta^2 \text{ ή } \mathbb{E}_\theta(3S_0^2) = \theta^2, \quad \forall\, \theta \in \Theta,$$

που σημαίνει ότι $3S_0^2$ είναι αμερόληπτος εκτιμητής του θ^2. Ανάλογα καταλήγουμε ότι $3S^2$ είναι επίσης αμερόληπτος εκτιμητής του θ^2. Σημειώνουμε ότι υπάρχουν καλύτεροι αμερόληπτοι και μη αμερόληπτοι εκτιμητές από τους S_0^2 και S^2, με κριτήριο το ΜΤΣ, δηλαδή οι S_0^2 και S^2 είναι μη αποδεκτοί (βλέπε Άσκηση 6.18).

Παράδειγμα 4.2.2. (κανονικές κατανομές με κοινή μέση τιμή)
(α) Θεωρούμε παρατηρήσεις $X_1, \ldots, X_n$ με κανονικές κατανομές $\mathcal{N}(\theta, \sigma_1^2)$, $\ldots, \mathcal{N}(\theta, \sigma_n^2)$, αντίστοιχα, και όχι κατ' ανάγκη ανεξάρτητες, όπου $\theta \in \Theta = \mathbb{R}$ είναι άγνωστη παράμετρος και σ_i^2, $i = 1 \ldots, n$ είναι γνωστές σταθερές. Επειδή, $\mathbb{E}_\theta X_i = \theta$, $\; i = 1 \ldots, n$, από την Παρατήρηση 4.2.2 προκύπτει ότι ο δειγματικός μέσος $\bar{X} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$ είναι αμερόληπτος εκτιμητής του θ.
(β) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$, $n \geq 2$, ένα τυχαίο δείγμα, από την κανονική κατανομή $\mathcal{N}(\theta, \kappa\theta^2)$, όπου $\theta \in \Theta = \mathbb{R} \setminus \{0\}$ είναι άγνωστη παράμετρος και $\kappa > 0$, γνωστή σταθερά (βλέπε επίσης το Παράδειγμα 4.1.3). Επειδή το θ είναι η μέση τιμή της κατανομής, $\mathbb{E}_\theta X_1 = \theta$, από την Πρόταση 4.2.3 συμπεραίνουμε ότι ο δειγματικός μέσος $\bar{X} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$ είναι αμερόληπτος εκτιμητής του θ. Επίσης, επειδή $\mathbb{V}ar_\theta X_1 = \kappa\theta^2$, από την Πρόταση 4.2.4, προκύπτει ότι η δειγματική διασπορά $S^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \bar{X})^2$ είναι αμερόληπτος εκτιμητής του $\kappa\theta^2$ και συνεπώς από την Πρόταση 4.2.1, $\frac{S^2}{\kappa}$ είναι αμερόληπτος εκτιμητής του θ^2.

(γ) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την κανονική κατανομή $\mathcal{N}(2\theta+1, 1)$, όπου $\theta \in \Theta = \mathbb{R}$ είναι άγνωστη παράμετρος. Επειδή η μέση τιμή της κατανομής είναι $2\theta+1$, $\mathbb{E}_\theta X_1 = 2\theta+1$, από την Πρόταση 4.2.3, ο δειγματικός μέσος $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ είναι αμερόληπτος εκτιμητής του $2\theta+1$. Θα βρούμε τώρα αμερόληπτο εκτιμητή του θ ως ακολούθως. Θέτουμε $g(\theta) = 2\theta+1$ και λύνουμε ως προς θ, $\theta = \frac{1}{2}g(\theta) - \frac{1}{2}$, οπότε από την Πρόταση 4.2.2, με $\alpha = \frac{1}{2}$, $\beta = -\frac{1}{2}$, προκύπτει ο $\frac{1}{2}\bar{X} - \frac{1}{2}$ ως αμερόληπτος εκτιμητής του θ. Εναλλακτικά, από τον ορισμό αμεροληψίας, έχουμε $\mathbb{E}_\theta \bar{X} = 2\theta+1$ και λύνουμε πάλι ως προς θ,

$$\theta = \frac{1}{2}\mathbb{E}_\theta \bar{X} - \frac{1}{2} = \mathbb{E}_\theta\left(\frac{1}{2}\bar{X} - \frac{1}{2}\right), \text{ δηλαδή } \mathbb{E}_\theta\left(\frac{1}{2}\bar{X} - \frac{1}{2}\right) = \theta, \ \forall\, \theta \in \Theta,$$

που εξ ορισμού σημαίνει ότι $\frac{1}{2}\bar{X} - \frac{1}{2}$ είναι αμερόληπτος εκτιμητής του θ.
□

Οι επόμενες δύο προτάσεις παρέχουν αμερόληπτους εκτιμητές της συνάρτησης κατανομής και της πιθανότητας ενός ενδεχομένου (σε πιο αδρές γραμμές, ενός ποσοστού που αναφέρεται σε αυτήν την κατανομή). Υπενθυμίζουμε ότι στην Ενότητα 3.3.1.δ ορίστηκε η εμπειρική συνάρτηση κατανομής και δικαιολογήθηκε ότι έχει έννοια να χρησιμοποιείται για την εκτίμηση της συνάρτησης κατανομής, ως το δειγματικό ανάλογό της. Η επόμενη πρόταση αποδεικνύει την αμεροληψία της. Η δείκτρια συνάρτηση, που εμφανίζεται στην πρόταση έχει οριστεί στην Ενότητα 3.3.1.δ .

Πρόταση 4.2.5. (**αμερόληπτος εκτιμητής της συνάρτησης κατανομής σε δοθέν σημείο**) *Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από μία αυθαίρετη κατανομή με πυκνότητα $f_1(x;\theta)$, $\theta \in \Theta$, και συνάρτηση κατανομής $F(x;\theta) = \mathbb{P}_\theta(X_1 \le x)$, $x \in \mathbb{R}$.*
(α) Για κάθε $n \ge 1$, η εμπειρική συνάρτηση κατανομής $\widehat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}_{(-\infty,x]}(X_i)$ είναι αμερόληπτος εκτιμητής της συνάρτησης κατανομής $F(x;\theta)$, δηλαδή

$$\mathbb{E}_\theta \widehat{F}(x) = F(x;\theta), \ \forall\, \theta \in \Theta, \ \forall\, x \in \mathbb{R}.$$

(β) $\mathbb{V}ar_\theta \widehat{F}(x) = \dfrac{F(x;\theta)\,(1 - F(x;\theta))}{n}, \ \forall\, \theta \in \Theta, \ \forall\, x \in \mathbb{R}.$

Απόδειξη. (α) Θέτουμε $Y_i = \mathrm{I}_{(-\infty,x]}(X_i)$, $i = 1\dots,n$, και παρατηρούμε ότι οι τυχαίες μεταβλητές $Y_1,\dots,Y_n$ είναι ανεξάρτητες, επειδή κάθε μία είναι συνάρτηση μιας και μόνον εκ των $X_1,\dots,X_n$, έχουν κοινή κατανομή, αφού οι $X_1,\dots,X_n$ έχουν κοινή κατανομή, ενώ οι τιμές τους είναι 1 ή 0 εξ ορισμού της δείκτριας συνάρτησης. Επομένως, οι $Y_1,\dots,Y_n$ αποτελούν ένα τυχαίο δείγμα από την κατανομή Bernoulli με παράμετρο

$$p = \mathbb{P}_\theta(Y_1 = 1) = \mathbb{P}_\theta\left(X_1 \in (-\infty, x]\right) = \mathbb{P}_\theta(X_1 \le x) = F(x;\theta).$$

Η μέση τιμή της κατανομής Bernoulli είναι ίση με την παράμετρό της, δηλαδή στην προκειμένη περίπτωση έχουμε ότι $\mathbb{E}_\theta Y_1 = p = F(x;\theta)$. Συνεπώς, η προς εκτίμηση παράμετρος $F(x;\theta)$ συμπίπτει με τη μέση τιμή της κοινής κατανομής του τυχαίου δείγματος $Y_1,\dots,Y_n$ και επομένως, από την Πρόταση 4.2.3, ο δειγματικός μέσος των $Y_1,\dots,Y_n$, $\bar{Y} = \frac{1}{n}\sum\limits_{i=1}^{n} Y_i$ είναι αμερόληπτος εκτιμητής της. Άρα έχουμε

$$\mathbb{E}_\theta \bar{Y} = F(x;\theta), \ \ \forall\, \theta \in \Theta, \ \ \forall\, x \in \mathbb{R},$$

και αντικαθιστώντας τα Y_i στην τελευταία σχέση,
$\mathbb{E}_\theta\left(\frac{1}{n}\sum_{i=1}^{n} \mathrm{I}_{(-\infty,x]}(X_i)\right) = F(x;\theta)$, δηλαδή $\mathbb{E}_\theta \widehat{F}(x) = F(x;\theta)$, $\forall\, \theta \in \Theta$, $\forall\, x \in \mathbb{R}$. Εναλλακτικά, το τελευταίο μέρος της απόδειξης μπορεί να παραχθεί και ως ακολούθως. Επειδή $\mathbb{E}_\theta Y_i = F(x;\theta)$, $i = 1\dots,n$, έχουμε

$$\begin{aligned}
\mathbb{E}_\theta \widehat{F}(x) &= \mathbb{E}_\theta\left(\frac{1}{n}\sum_{i=1}^{n} \mathrm{I}_{(-\infty,x]}(X_i)\right) = \mathbb{E}_\theta\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_\theta Y_i \\
&= \frac{1}{n}\sum_{i=1}^{n} F(x;\theta) = \frac{1}{n} n F(x;\theta) = F(x;\theta), \quad \forall\, \theta \in \Theta, \quad \forall\, x \in \mathbb{R}.
\end{aligned}$$

(β) Η κατανομή Bernoulli με παράμετρο p έχει διασπορά $\sigma^2 = p(1-p)$, άρα στην περίπτωση των $Y_1,\dots,Y_n$ έχουμε

$$\mathbb{V}ar_\theta Y_i = \sigma^2 = p(1-p) = F(x;\theta)\left(1 - F(x;\theta)\right).$$

Συνεπώς, από την Πρόταση 4.2.3, βλέποντας τα Y_i ως X_i, προκύπτει

$$\mathbb{V}ar_\theta \widehat{F}(x) = \mathbb{V}ar_\theta(\bar{Y}) = \frac{\sigma^2}{n} = \frac{F(x;\theta)\left(1 - F(x;\theta)\right)}{n}, \quad \forall\, \theta \in \Theta, \quad \forall\, x \in \mathbb{R}.$$

□

Για την επόμενη πρόταση θεωρούμε ένα τυχαίο δείγμα $\underset{\sim}{X} = (X_1, \ldots, X_n)$ από μία κατανομή με πυκνότητα $f_1(x;\theta)$, $\theta \in \Theta$, και έστω B δοθέν υποσύνολο του συνόλου τιμών της. Έστω προς εκτίμηση η πιθανότητα του ενδεχομένου B, δηλαδή $g(\theta) = \mathbb{P}_\theta(X_1 \in B)$. Αν θεωρήσουμε ως «επιτυχία» το ενδεχόμενο $X_1 \in B$, τότε το $g(\theta)$ είναι η πιθανότητα «επιτυχίας» δηλαδή η παράμετρος μιας κατανομής Bernoulli. Διαφορετικά, το $g(\theta)$ μπορεί να ερμηνευθεί ως το ποσοστό των τιμών της κατανομής που ανήκουν στο σύνολο B. Στην Ενότητα 3.3.1.ε, δόθηκε ως εκτιμητής του $g(\theta)$ το δειγματικό ανάλογό του, $\frac{1}{n}\sum_{i=1}^{n} \mathrm{I}_B(X_i)$. Η επόμενη πρόταση αποδεικνύει την αμεροληψία αυτού του εκτιμητή. Επιπροσθέτως, αποτελεί γενίκευση της Πρότασης 4.2.5, αφού για $B = (-\infty, x]$ προκύπτει ακριβώς αυτή η πρόταση.

Πρόταση 4.2.6. *(***αμερόληπτος εκτιμητής της πιθανότητας ενός ενδεχομένου***) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από αυθαίρετη κατανομή με πυκνότητα $f_1(x;\theta)$, $\theta \in \Theta$ και B υποσύνολο του συνόλου των τιμών της.*

(α). Για κάθε $n \geq 1$, η στατιστική συνάρτηση $\frac{1}{n}\sum_{i=1}^{n} \mathrm{I}_B(X_i)$ είναι αμερόληπτος εκτιμητής της πιθανότητας $\mathbb{P}_\theta(X_1 \in B)$, δηλαδή

$$\mathbb{E}_\theta\left(\frac{1}{n}\sum_{i=1}^{n} \mathrm{I}_B(X_i)\right) = \mathbb{P}_\theta(X_1 \in B), \ \ \forall\, \theta \in \Theta.$$

(β). $\mathbb{V}ar_\theta\left(\frac{1}{n}\sum_{i=1}^{n} \mathrm{I}_B(X_i)\right) = \dfrac{\mathbb{P}_\theta(X_1 \in B)\,(1 - \mathbb{P}_\theta(X_1 \in B))}{n}.$

Απόδειξη. Είναι εντελώς παρόμοια με την απόδειξη της Πρότασης 4.2.5, αρκεί να θέσουμε $Z_i = \mathrm{I}_B(X_i)$, $i = 1, \ldots, n$, και να παρατηρήσουμε ότι οι $Z_1, \ldots, Z_n$ αποτελούν τυχαίο δείγμα από την κατανομή Bernoulli με παράμετρο $p = \mathbb{P}_\theta(X_1 \in B)$. □

Παραθέτουμε μερικά ακόμη παραδείγματα εμπέδωσης της έννοιας της αμεροληψίας.

Παράδειγμα 4.2.3. (κατανομή Bernoulli-ικανή και αναγκαία συνθήκη για την ύπαρξη αμερόληπτου εκτιμητή του g(θ)) Έστω $\underset{\sim}{X} =$

$(X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την κατανομή Bernoulli $\mathcal{B}(1, \theta)$, όπου $\theta \in \Theta = (0, 1)$. Για την εκτίμηση του θ, παρατηρούμε ότι $\theta = \mathbb{P}_\theta(X_1 = 1)$, άρα εφαρμόζοντας την Πρόταση 4.2.6 με $B = \{1\}$, προκύπτει ο αμερόληπτος εκτιμητής του θ,

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{I}_{\{1\}}(X_i) = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X},$$

με την προτελευταία ισότητα να ισχύει επειδή $X_i = 1$ ή 0, $i = 1 \ldots, n$. Στον ίδιο εκτιμητή θα καταλήξουμε, αν λάβουμε υπ' όψη ότι η παράμετρος θ είναι η μέση τιμή της κατανομής, $\mathbb{E}_\theta X_1 = \theta$, και εφαρμόσουμε την Πρόταση 4.2.3.

Θα αναζητήσουμε τώρα αμερόληπτο εκτιμητή του θ^2. Εφαρμόζοντας την αρχή της αντικατάστασης (βλέπε Ενότητα 3.3.2), θεωρούμε ως εκτιμητή του θ^2 τον $\bar{X}^2$. Επειδή μας ενδιαφέρει η αμεροληψία του $\bar{X}^2$, υπολογίζουμε την $\mathbb{E}_\theta \bar{X}^2$ από τη σχέση

$$\mathbb{E}_\theta \bar{X}^2 = \mathbb{V}ar_\theta \bar{X} + \left(\mathbb{E}_\theta \bar{X}\right)^2.$$

Λόγω αμεροληψίας του $\bar{X}$, έχουμε $\mathbb{E}_\theta \bar{X} = \theta$, ενώ από την Πρόταση 4.2.3 (β), επειδή $\mathbb{V}ar_\theta X_1 = \theta(1-\theta)$, έχουμε

$$\mathbb{V}ar_\theta \bar{X} = \frac{\theta(1-\theta)}{n}.$$

Συνεπώς,

$$\mathbb{E}_\theta \bar{X}^2 = \frac{\theta(1-\theta)}{n} + \theta^2 = \frac{n-1}{n}\theta^2 + \frac{\theta}{n},$$

το οποίο σημαίνει ότι ο $\bar{X}^2$ δεν είναι αμερόληπτος εκτιμητής του θ^2 (παρότι ο $\bar{X}$ είναι για το θ - βλέπε, ξανά, την Παρατήρηση 4.2.1). Αν στο δεύτερο μέλος της τελευταίας ισότητας δεν υπήρχε ο όρος $\frac{\theta}{n}$ ή υπήρχε στη θέση του κάποια σταθερά μη εξαρτώμενη από το θ, θα ήταν πιο εύκολο να καταλήξουμε σε έναν αμερόληπτο εκτιμητή του θ^2, λύνοντας την αντίστοιχη ισότητα ως προς θ^2. Αυτό θα κάνουμε και τώρα, αφού προηγηθεί ένα σημαντικό ενδιάμεσο βήμα, το ακόλουθο. Επειδή $\mathbb{E}_\theta \bar{X} = \theta$, προκύπτει ότι

$$\mathbb{E}_\theta \bar{X}^2 = \frac{n-1}{n}\theta^2 + \frac{\mathbb{E}_\theta \bar{X}}{n}.$$

Λύνοντας ως προς θ^2, για $n \geq 2$, καταλήγουμε ότι

$$\theta^2 = n\left(\mathbb{E}_\theta \bar{X}^2 - \frac{1}{n}\mathbb{E}_\theta \bar{X}\right) / (n-1) = \mathbb{E}_\theta\left[\frac{n-1}{n}\left(\bar{X}^2 - \frac{1}{n}\bar{X}\right)\right]$$

ή
$$\mathbb{E}_\theta\left[\frac{n}{n-1}\bar{X}\left(\bar{X}-\frac{1}{n}\right)\right]=\theta^2,\ \ \forall\, 0<\theta<1\ \ (n\geq 2),$$
δηλαδή ο εκτιμητής $\frac{n}{n-1}\bar{X}\left(\bar{X}-\frac{1}{n}\right)$ είναι αμερόληπτος εκτιμητής του θ^2. Συνεχίζοντας, θα θέσουμε και θα απαντήσουμε το ερώτημα: Για ποιες συναρτήσεις g υπάρχει αμερόληπτος εκτιμητής της τιμής $g(\theta)$; Έστω $T(\underset{\sim}{X})$ αμερόληπτος εκτιμητής του $g(\theta)$, οπότε
$$\mathbb{E}_\theta T(\underset{\sim}{X})=g(\theta),\ \ \forall\,\theta\in(0,1).$$
Η πυκνότητας της παρατήρησης X_i είναι
$$f_1(x_i;\theta)=\theta^{x_i}(1-\theta)^{1-x_i},\ \ x_i=1 \text{ ή } 0,\ \ i=1,\ldots,n$$
και επομένως, λόγω ανεξαρτησίας, η πυκνότητα του $\underset{\sim}{X}=(X_1,\ldots,X_n)$ είναι
$$f(\underset{\sim}{x};\theta)=\prod_{i=1}^{n}f_1(x_i;\theta)=\theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i},$$
$\underset{\sim}{x}=(x_1,\ldots,x_n)$ με $x_i=1$ ή 0. Τότε, από τον ορισμό της μέσης τιμής, η συνθήκη της αμεροληψίας γράφεται $\sum_{\underset{\sim}{x}}T(\underset{\sim}{x})f(\underset{\sim}{x};\theta)=g(\theta)$ ή
$$\sum_{\underset{\sim}{x}}T(\underset{\sim}{x})\theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}=g(\theta),\ \ \forall\,\theta\in(0,1).$$
Παρατηρούμε τώρα ότι το άθροισμα $\sum\limits_{i=1}^{n}x_i$ παίρνει τιμές $0,1,\ldots,n$ και επομένως το πρώτο μέλος της τελευταίας σχέσης είναι πολυώνυμο ως προς θ βαθμού το πολύ n, αφού για κάθε $\underset{\sim}{x}$ το $T(\underset{\sim}{x})$, ως τιμή του εκτιμητή, είναι σταθερά μη εξαρτώμενη από το θ. Άρα και το δεύτερο μέλος αυτής της σχέσης, δηλαδή η συνάρτηση $g(\theta)$, είναι ένα πολυώνυμο ως προς θ, βαθμού το πολύ n. Αντίστροφα, έστω
$$g(\theta)=\alpha_0+\alpha_1\theta+\cdots+\alpha_n\theta^n$$
με α_i, $i=0,\ldots,n$, σταθερές όχι όμως όλες μηδέν. Εφαρμόζοντας την Πρόταση **4.2.2**, για να κατασκευάσουμε αμερόληπτο εκτιμητή του $g(\theta)$ αρκεί να βρούμε αμερόληπτους εκτιμητές των (δυνάμεων) $g_i(\theta)=\theta^i$, $i=1,\ldots,n$. Παρατηρούμε (αν και διόλου προφανές!!!) ότι, λόγω της ανεξαρτησίας και της κοινής κατανομής των παρατηρήσεων,
$$\mathbb{P}_\theta(X_1=1,X_2=1,\ldots,X_i=1)=\theta^i=g_i(\theta),\ \ \forall\, i=1,\ldots,n.$$

Έχοντας εκφράσει το $g_i(\theta)$ ως πιθανότητα ενδεχομένου, η δείκτρια συνάρτηση αυτού του ενδεχομένου είναι αμερόληπτος εκτιμητής του $g_i(\theta)$. Πράγματι, ορίζουμε τη στατιστική συνάρτηση $T_i(\underset{\sim}{X})$ με τιμές

$$T_i(\underset{\sim}{X}) = \begin{cases} 1, & \text{αν } X_1 = X_2 = \cdots = X_i = 1, \\ 0, & \text{αν κάποια από τις } X_1, X_2, \ldots, X_i \text{ είναι } 0. \end{cases}$$

Εξ ορισμού, η $T_i(\underset{\sim}{X})$ έχει κατανομή Bernoulli με παράμετρο

$$p = \mathbb{P}_\theta(T_i(\underset{\sim}{X}) = 1) = \mathbb{P}_\theta(X_1 = X_2 = \cdots = X_i = 1) = g_i(\theta)$$

και επομένως η μέση τιμή της είναι η παράμετρός της, δηλαδή

$$\mathbb{E}_\theta T_i(\underset{\sim}{X}) = p = g_i(\theta), \ \ \forall\, \theta \in (0,1), \ \ i = 1, \ldots, n,$$

που σημαίνει ότι η στατιστική συνάρτηση $T_i(\underset{\sim}{X})$ είναι αμερόληπτος εκτιμητής του $g_i(\theta) = \theta^i$. Εναλλακτικά, παρατηρούμε ότι $T_i(\underset{\sim}{X}) = X_1 X_2 \cdots X_i$, οπότε λόγω ανεξαρτησίας των παρατηρήσεων,

$$\mathbb{E}_\theta T_i(\underset{\sim}{X}) = \mathbb{E}_\theta X_1 \mathbb{E}_\theta X_2 \cdots \mathbb{E}_\theta X_i = \theta^i.$$

Συνεπώς, λόγω της Πρότασης 4.2.2, η στατιστική συνάρτηση

$$\alpha_0 + \alpha_1 T_1(\underset{\sim}{X}) + \cdots + \alpha_n T_n(\underset{\sim}{X})$$

είναι αμερόληπτος εκτιμητής του

$$\alpha_0 + \alpha_1 g_1(\theta) + \cdots + \alpha_n g_n(\theta) = \alpha_0 + \alpha_1\theta + \cdots + \alpha_n\theta^n = g(\theta).$$

Καταλήξαμε λοιπόν στο συμπέρασμα: Δοθέντος τυχαίου δείγματος μεγέθους n από την κατανομή Bernoulli $\mathcal{B}(1,\theta)$, $\theta \in (0,1)$, η τιμή $g(\theta)$ μπορεί να εκτιμηθεί αμερόληπτα αν και μόνον αν η συνάρτηση g είναι πολυώνυμο βαθμού το πολύ n ως προς θ. Επομένως δεν υπάρχει αμερόληπτος εκτιμητής του $g(\theta) = \theta^{n+1}$ ή του $g(\theta) = \frac{\theta}{1-\theta}$.Ο λόγος της πιθανότητας «επιτυχίας» προς την πιθανότητα «αποτυχίας», $\frac{\theta}{1-\theta}$, ονομάζεται «odds ratio», και έχει ενδιαφέρον, για παράδειγμα, σε στοιχηματικά παιχνίδια, αν π.χ. $\theta = 1/5$, το odds ratio είναι $1:4$, δηλαδή η πιθανότητα να χάσει ο παίκτης το στοίχημα είναι τετραπλάσια από την πιθανότητα να το κερδίσει.

Παράδειγμα 4.2.4. (κατανομή Poisson-εκτίμηση της μέσης τιμής και μιας πιθανότητας) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$, $n \geq 2$, ένα τυχαίο δείγμα από την κατανομή Poisson, $\mathcal{P}(\theta)$, με πυκνότητα

$$f_1(x;\theta) = \mathbb{P}_\theta(X_1 = x) = e^{-\theta}\frac{\theta^x}{x!}, \;\; x = 0, 1, \ldots, \;\; \theta \in \Theta = (0, \infty).$$

Η κατανομή Poisson, $\mathcal{P}(\theta)$, μπορεί να χρησιμοποιηθεί ως μοντέλο του αριθμού (τροχαίων ή άλλων) ατυχημάτων, που συμβαίνουν σε καθορισμένο χρονικό διάστημα. Με αυτήν την ερμηνεία, η μέση τιμή της θ παριστάνει τον μέσο αριθμό ατυχημάτων. Για την εκτίμηση του θ, επειδή $\mathbb{E}_\theta X_1 = \mathbb{V}ar_\theta X_1 = \theta$ από τις Προτάσεις 4.2.3 και 4.2.4 αντίστοιχα προκύπτουν οι αμερόληπτοι εκτιμητές $\bar{X} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$ και $S^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}\big(X_i - \bar{X}\big)^2$. Στο Κεφάλαιο 6 θα διαπιστώσουμε ότι με κριτήριο το ΜΤΣ ο δειγματικός μέσος $\bar{X}$ είναι καλύτερος εκτιμητής από τη δειγματική διασπορά S^2. Ας θεωρήσουμε, επιπλέον, εκτίμηση του $g(\theta) = e^{-\theta}$. Παρατηρούμε ότι $\mathbb{P}_\theta(X_1 = 0) = e^{-\theta}$. Άρα με την παραπάνω ερμηνεία του μοντέλου ατυχημάτων, $e^{-\theta}$ είναι το ποσοστό αυτών των χρονικών διαστημάτων με μηδέν ατυχήματα. Από την Πρόταση 4.2.6 ένας αμερόληπτος εκτιμητής του $e^{-\theta}$ είναι ο $S_1(\underset{\sim}{X}) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \mathrm{I}_{\{0\}}(X_i)$, δηλαδή το ποσοστό των μηδενικών τιμών του δείγματος $\underset{\sim}{X} = (X_1, \ldots, X_n)$. Στο Κεφάλαιο 6 θα βρούμε έναν άλλο αμερόληπτο εκτιμητή με μικρότερο ΜΤΣ.

Παράδειγμα 4.2.5. (κατανομή Poisson-μοναδικός αμερόληπτος εκτιμητής) Έστω X μια παρατήρηση από την κατανομή Poisson, $\mathcal{P}(\theta)$, $\theta \in \Theta = (0, \infty)$. Επειδή $\mathbb{E}_\theta X = \theta$, για κάθε $\theta \in \Theta$, η X είναι αμερόληπτος εκτιμητής του θ. Θα δείξουμε ότι είναι και μοναδικός. Έστω $T(X)$ ένας αμερόληπτος εκτιμητής του θ, επομένως έχουμε

$$\mathbb{E}_\theta T(X) = \theta, \quad \forall\, \theta \in \Theta \text{ ή } \sum_{x=0}^{\infty} T(x) e^{-\theta}\frac{\theta^x}{x!} = \theta, \quad \forall\, \theta \in \Theta$$

$$\text{ή } \sum_{x=0}^{\infty} T(x)\frac{\theta^x}{x!} = \theta e^{\theta} = \theta\sum_{x=0}^{\infty}\frac{\theta^x}{x!} = \sum_{x=0}^{\infty}\frac{\theta^{x+1}}{x!}, \quad \forall\, \theta \in \Theta,$$

$$\text{δηλαδή } \sum_{x=0}^{\infty} T(x)\frac{\theta^x}{x!} = \sum_{x=1}^{\infty}\frac{\theta^x}{(x-1)!}, \quad \forall\, \theta \in \Theta.$$

Επειδή οι δύο αυτές δυναμοσειρές συμπίπτουν για κάθε $\theta > 0$, οι συντελεστές των ιδίων δυνάμεων του θ είναι ίσοι και επομένως

$$T(0) = 0, \quad \frac{T(x)}{x!} = \frac{1}{(x-1)!}, \quad x = 1, 2, \ldots,$$

οπότε

$$T(x) = x, \quad x = 0, 1, 2, \ldots .$$

Η τελευταία σχέση σημαίνει ότι $T(X) = X$ (αφού η X, ως τυχαία μεταβλητή Poisson, έχει σύνολο τιμών $\{0, 1, 2, \ldots\}$).

Παράδειγμα 4.2.6. (βέλτιστος αμερόληπτος εκτιμητής ειδικής μορφής) Θεωρούμε ανεξάρτητα τυχαία δείγματα $\underset{\sim}{X} = (X_1, X_2, \ldots, X_n)$ και $\underset{\sim}{Y} = (Y_1, Y_2, \ldots, Y_n)$ με την ίδια άγνωστη μέση τιμή $\theta \in \Theta$ και γνωστές διασπορές σ_1^2, σ_2^2, αντίστοιχα. Στην πράξη τα $\underset{\sim}{X}$ και $\underset{\sim}{Y}$ μπορεί να παριστάνουν πειραματικές μετρήσεις μιας άγνωστης «ποσότητας» θ με δύο μεθόδους γνωστής, αλλά διαφορετικής ακρίβειας. Θα δείξουμε ότι η στατιστική συνάρτηση $T_c = T_c(\underset{\sim}{X}, \underset{\sim}{Y}) = c\bar{X} + (1-c)\bar{Y}$, όπου $c \in \mathbb{R}$ σταθερά μη εξαρτώμενη από το θ, είναι αμερόληπτος εκτιμητής του θ. Πράγματι, έχουμε

$$\mathbb{E}_\theta T_c = \mathbb{E}_\theta(c\bar{X} + (1-c)\bar{Y}) = c\mathbb{E}_\theta\bar{X} + (1-c)\mathbb{E}_\theta\bar{Y} = c\theta + (1-c)\theta = \theta, \ \forall\ \theta \in \Theta.$$

Συνεπώς, ο T_c είναι αμερόληπτος εκτιμητής του θ, για κάθε $c \in \mathbb{R}$. Στη συνέχεια θα βρούμε τον καλύτερο εκτιμητή της μορφής T_c, ως προς το ΜΤΣ. Επειδή οι εκτιμητές T_c, $c \in \mathbb{R}$ ορίζουν μία κλάση εκτιμητών του θ, αναζητούμε εκείνη την τιμή του c η οποία ελαχιστοποιεί το ΜΤΣ(T_c, θ). Λόγω όμως της αμεροληψίας του T_c, από τη σχέση (4.2) ή (4.3) έχουμε

$$\begin{aligned}\text{ΜΤΣ}(T_c, \theta) &= \mathbb{V}ar_\theta T_c = \mathbb{V}ar_\theta(c\bar{X} + (1-c)\bar{Y}) = c^2\mathbb{V}ar_\theta\bar{X} + (1-c)^2\mathbb{V}ar_\theta\bar{Y} \\ &= c^2\frac{\sigma_1^2}{n} + (1-c)^2\frac{\sigma_2^2}{n} = \frac{\sigma_1^2 + \sigma_2^2}{n}c^2 - 2\frac{\sigma_2^2}{n}c + \frac{\sigma_2^2}{n}.\end{aligned}$$

Η τρίτη ισότητα ισχύει λόγω της ανεξαρτησίας των δειγμάτων $\underset{\sim}{X}$ και $\underset{\sim}{Y}$. Παρατηρούμε ότι το ΜΤΣ(T_c, θ) είναι ένα τριώνυμο ως προς c (της μορφής $ac^2 + bc + d$, με $a > 0$), το οποίο ελαχιστοποιείται για

$$c_0 = -\frac{b}{2a} = -\frac{-2\frac{\sigma_2^2}{n}}{2\frac{\sigma_1^2+\sigma_2^2}{n}} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}.$$

Συνεπώς, ο καλύτερος εκτιμητής στην κλάση των εκτιμητών του θ, $\{T_c : c \in \mathbb{R}\}$, είναι ο σταθμισμένος μέσος των $\bar{X}$ και $\bar{Y}$,

$$T_{c_0} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\bar{X} + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}\bar{Y}.$$

Στην ειδική περίπτωση $\sigma_1 = \sigma_2$, ο T_{c_0} συμπίπτει με τον μέσο όρο των δειγματικών μέσων.

4.3 Η μη αποδεκτικότητα της δειγματικής διασποράς με κριτήριο το ΜΤΣ

Η ενότητα αυτή αναφέρεται σε μια απρόσμενη συμπεριφορά του ΜΤΣ της δειγματικής διασποράς $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\big(X_i - \bar{X}\big)^2$ ως εκτιμητή της διασποράς, σ^2, μιας κατανομής. Το περιεχόμενό της είναι αρκετά εξειδικευμένο και θα μπορούσε και να παραληφθεί σε πρώτη ανάγνωση.

Έστω $\underset{\sim}{X} = (X_1, X_2, \ldots, X_n)$, $n \geq 2$, ένα τυχαίο δείγμα από μία αυθαίρετη κατανομή με πυκνότητα $f_1(x;\theta)$, $\theta \in \Theta$. Υποθέτουμε ότι η X_1 δεν παίρνει μόνο μία τιμή, οπότε η διασπορά της, σ^2, είναι γνησίως θετική και (για τεχνικούς λόγους) υποθέτουμε ότι η X_1 έχει πεπερασμένη τέταρτη ροπή, $\mathbb{E}_\theta X_1^4 < \infty$. Οι δύο πλέον γνωστοί και χρησιμοποιούμενοι στις εφαρμογές εκτιμητές του σ^2 είναι η δειγματική διασπορά $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\big(X_i - \bar{X}\big)^2$ και η στατιστική συνάρτηση $S_*^2 = \frac{1}{n}\sum_{i=1}^{n}\big(X_i - \bar{X}\big)^2$, που, όπως είδαμε στην Ενότητα 3.3.2, προκύπτει από την αρχή της αντικατάστασης και επιπροσθέτως είναι ο εκτιμητής μεθόδου ροπών του σ^2, όπως θα δούμε στο Κεφάλαιο 7. Από την Πρόταση 4.2.4 γνωρίζουμε ότι S^2 είναι αμερόληπτος εκτιμητής του σ^2, ενώ στην Παρατήρηση 4.2.3 διαπιστώσαμε ότι S_*^2 δεν είναι αμερόληπτος εκτιμητής του σ^2.

Σχετικά πρόσφατα, ο Yatracos (2005) κατασκεύασε τον εκτιμητή του σ^2,

$$S_2^2 = c_2 S^2, \quad \text{όπου} \quad c_2 = \frac{(n+2)(n-1)}{n(n+1)},$$

και απέδειξε ότι ο S_2^2 έχει μικρότερο ΜΤΣ από τη δειγματική διασπορά S^2, οποιαδήποτε και αν είναι η κατανομή των παρατηρήσεων $X_1, \ldots, X_n$,

για οποιαδήποτε μέγεθος δείγματος $n \geq 2$, αρκεί μόνον να ισχύει ότι $\mathbb{E}_\theta X_1^4 < \infty$. Έτσι λοιπόν για *τυχούσα* κατανομή ο S^2 είναι μη αποδεκτός εκτιμητής του σ^2 με κριτήριο το ΜΤΣ και καλύτερος είναι ο S_2^2.
Η μη αποδεκτικότητα του S^2 ήταν γνωστή πριν το 2005 σε ορισμένες ειδικές περιπτώσεις, όπως, π.χ., στην περίπτωση της κανονικής κατανομής με (επίσης) άγνωστη μέση τιμή. Σε αυτήν την περίπτωση, όπως θα δούμε στο Παράδειγμα 6.4.2, ο εκτιμητής $S_3^2 = \frac{1}{n+1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$ έχει μικρότερο ΜΤΣ από τον S^2, ενώ εκτιμητές με μικρότερο ΜΤΣ από αυτό του S_3^2 έχουν κατασκευάσει οι Stein (1964), Brown (1968), Brewster and Zidek (1974), Strawderman (1974), Maruyama (1998) και Maruyama and Strawderman (2006). Όμως, η καθολικότητα του αποτελέσματος του Yatracos (2005) για οποιαδήποτε κατανομή ήταν απρόσμενη και εντυπωσιακή. Με κίνητρο αυτήν την εργασία, ο Kourouklis (2012) κατασκεύασε τον εκτιμητή

$$S_1^2 = c_1 S^2, \quad \text{όπου} \quad c_1 = \frac{n(n-1)}{n(n-1)+2}$$

και απέδειξε ότι, για οποιαδήποτε κατανομή και κάθε $n \geq 2$, ο S_1^2 έχει μικρότερο ΜΤΣ από τη δειγματική διασπορά S^2, αλλά και από τον S_2^2, οπότε και ο S_2^2 είναι μη αποδεκτός. Η μέθοδος κατασκευής του S_2^2 βασίστηκε σε μια πρωτοποριακή για την εποχή της, αλλά και κλασική πια ιδέα του Stein (1964). Ας δούμε τώρα βήμα προς βήμα την κατασκευή του $S_1^2 = c_1 S^2$. Κατ' αρχάς παρατηρούμε ότι οι εκτιμητές του σ^2, S^2, S_o^2 και S_2^2 είναι και οι τρεις ειδικές περιπτώσεις της γενικής μορφής cS^2, $c > 0$, με $c = 1$, $c = \frac{n-1}{n}$ και $c = c_2$ αντίστοιχα. Θεωρούμε λοιπόν όλους τους εκτιμητές της μορφής cS^2, $c > 0$, με στόχο να επιλέξουμε τη σταθερά c, μη εξαρτώμενη από το θ, ώστε να ελαχιστοποιηθεί το ΜΤΣ του cS^2 ως εκτιμητή του σ^2. Θα υπολογίσουμε το ΜΤΣ του cS^2 και ακολούθως θα το μελετήσουμε ως συνάρτηση του c. Από τη σχέση (4.2) και την Πρόταση

4.2.4, θέτοντας $\mu_4 = \mathbb{E}_\theta\left(X_1 - \mathbb{E}_\theta X_1\right)^4$ και $\alpha_4 = \frac{\mu_4}{\sigma^4}$ έχουμε

$$\begin{aligned}
\text{ΜΤΣ}(cS^2, \theta) &= \mathbb{V}ar_\theta\big(cS^2\big) + b^2(cS^2, \theta) = c^2 \mathbb{V}ar_\theta S^2 + \big(\mathbb{E}_\theta(cS^2) - \sigma^2\big)^2 \\
&= c^2\left\{\frac{\mu_4}{n} - \frac{(n-3)\sigma^4}{n(n-1)}\right\} + (c\sigma^2 - \sigma^2) \\
&= c^2\left\{\frac{\mu_4}{n} - \frac{(n-3)\sigma^4}{n(n-1)}\right\} + (c-1)^2\sigma^4
\end{aligned}$$

$$\begin{aligned}
&= c^2\left\{\frac{\mu_4}{n} - \frac{(n-3)\sigma^4}{n(n-1)} + \sigma^4\right\} - 2c\sigma^4 + \sigma^4 \\
&= c^2\left\{\frac{\alpha_4\sigma^4}{n} - \frac{(n-3)\sigma^4}{n(n-1)} + \sigma^4\right\} - 2c\sigma^4 + \sigma^4.
\end{aligned}$$

Σημειώνουμε ότι η συνθήκη $\mathbb{E}_\theta X_1^4 < \infty$ διασφαλίζει ότι το ΜΤΣ(cS^2, θ) δεν απειρίζεται, οπότε ως συνάρτηση του c είναι ένα τριώνυμο της μορφής $\alpha c^2 + \beta c + \gamma$ με

$$\alpha = \left\{\frac{\alpha_4}{n} - \frac{(n-3)}{n(n-1)} + 1\right\}\sigma^4 = \frac{n(n-1) + 2 + (n-1)(\alpha_4 - 1)}{n(n-1)}\sigma^4.$$

Θα δείξουμε τώρα ότι $\alpha_4 \geq 1$, το οποίο θα μας εξασφαλίσει ότι $\alpha > 0$. Αρκεί να δείξουμε ότι $\mu_4 > \sigma^4$. Πράγματι, θέτουμε $W = (X_1 - \mathbb{E}_\theta X_1)^2$ και συνεπώς έχουμε ότι $\mathbb{E}W = \sigma^2$ και $\mathbb{E}(W^2) = \mu_4$. Τέλος, επειδή $\mathbb{V}ar(W) \geq 0$ συνάγουμε ότι $\mathbb{E}(W^2) - (\mathbb{E}W)^2 \geq 0$ ή $\mu_4 - \sigma^4 \geq 0$, δηλαδή $\alpha_4 \geq 1$.

Επειδή $\alpha > 0$ από τις ιδιότητες του τριωνύμου, το ΜΤΣ(cS^2, θ) παρουσιάζει ελάχιστο ως προς c στο σημείο

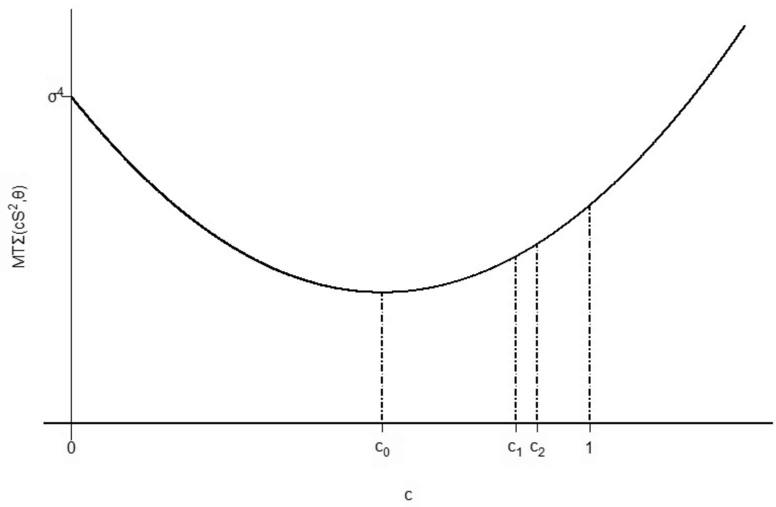
$$c_0 = \frac{n(n-1)}{n(n-1) + 2 + (n-1)(\alpha_4 - 1)}.$$

Όμως, εν γένει, η σταθερά α_4 άρα και η c_0 εξαρτώνται από την άγνωστη παράμετρο θ και σε αυτήν την περίπτωση η $c_0 S^2$ δεν μπορεί να χρησιμοποιηθεί ως εκτιμητής του σ^2. Ο αρχικός λοιπόν στόχος ελαχιστοποίησης του ΜΤΣ από εκτιμητή της μορφής cS^2 δεν είναι δυνατόν να επιτευχθεί. Σε αυτό το σημείο χρησιμοποιείται η ιδέα του Stein (1964): επειδή

ΜΤΣ(cS^2, θ) είναι γνησίως αύξουσα συνάρτηση του c για $c \geq c_0$, όσο μικρότερο είναι το c, αλλά μεγαλύτερο του c_0, τόσο μικρότερο είναι και το ΜΤΣ του cS^2 (βλέπε Σχήμα 4.4). Έτσι θέτουμε ως νέο στόχο να βρούμε τη μικρότερη σταθερά c, που ικανοποιεί τη σχέση $c \geq c_0$ και δεν εξαρτάται από την άγνωστη παράμετρο θ. Περιέργως, ίσως, αυτός ο στόχος μπορεί να επιτευχθεί σχετικά εύκολα. Πράγματι, επειδή $\alpha_4 \geq 1$, έχουμε ότι

$$c_0 = \frac{n(n-1)}{n(n-1) + 2 + (n-1)(\alpha_4 - 1)} \leq \frac{n(n-1)}{n(n-1)+2}, \quad \forall\, \theta \in \Theta.$$

Θέτουμε τελικά $c_1 = \frac{n(n-1)}{n(n-1)+2}$ και αναγνωρίζουμε ότι όντως η σταθερά c_1 είναι το μικρότερο άνω φράγμα για το c_0, αφού $c_0 = c_1$ όταν $\alpha_4 = 1$. Έτσι λοιπόν κατασκευάστηκε ο εκτιμητής $S_1^2 = c_1 S^2$ και η ανωτερότητα με κριτήριο το ΜΤΣ έναντι των S^2 και S_2^2 αποδεικνύεται πολύ εύκολα, μετά την ανάλυση που προηγήθηκε.



Σχήμα 4.4: ΜΤΣ του εκτιμητή cS^2

Πρόταση 4.3.1. *Ο εκτιμητής* $S_1^2 = c_1 S^2$ *έχει μικρότερο ΜΤΣ από τον* $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$ *και τον* $S_2^2 = c_2 S^2$*, για οποιαδήποτε κατανομή* *με $\mathbb{E}_\theta X_1^4 < \infty$ και για κάθε $n \geq 2$. Συνεπώς, οι εκτιμητές S^2 και S_2^2 είναι μη αποδεκτοί με κριτήριο το ΜΤΣ.*

Απόδειξη. Όπως ειπώθηκε προηγουμένως, το ΜΤΣ(cS^2, θ) είναι γνησίως αύξουσα συνάρτηση του c για $c \geq c_0$. Γνωρίζουμε ήδη ότι $c_0 \leq c_1$ και είναι πολύ εύκολο να διαπιστώσουμε ότι $c_1 < c_2 < 1$. Επομένως, έχουμε

$$\text{ΜΤΣ}(c_1 S^2, \theta) < \text{ΜΤΣ}(c_2 S^2, \theta) < \text{ΜΤΣ}(S^2, \theta) \quad \forall\, \theta \in \Theta. \qquad \square$$

Παρατήρηση 4.3.1. Το φαινόμενο η διασπορά και η μεροληψία να μεταβάλλονται προς αντίθετες κατευθύνσεις, που αναφέρθηκε στη Ενότητα 4.1 και το Παράδειγμα 4.1.3, εμφανίζεται και στους εκτιμητές S_1^2, S_2^2 και S^2. Οι δύο πρώτοι είναι αρνητικά μεροληπτικοί και συνεπώς έχουν την «τάση» να υποεκτιμούν το σ^2, ενώ ο S^2 είναι αμερόληπτος. Όμως οι S_1^2 και S_2^2 έχουν «αρκετά» μικρότερη διασπορά από τον S^2, ώστε τελικά να έχουν και μικρότερο ΜΤΣ από τον S^2.

Παρατήρηση 4.3.2. Προς αποφυγή παρερμηνείας, το συμπέρασμα που προκύπτει από το αποτέλεσμα του Yatracos (2005) και την Πρόταση 4.3.1 είναι ότι: αν κάποιος *υιοθετήσει* το ΜΤΣ ως κριτήριο επιλογής, τότε υπάρχουν καλύτεροι εκτιμητές από τη δειγματική διασπορά S^2 για την εκτίμηση της διασποράς, σ^2, τυχούσας κατανομής με $\mathbb{E}X_1^4 < \infty$. Όμως, όπως επισημαίνουν οι Stein (1964) και Casella and Berger (2002, σελ. 332) για την περίπτωση κανονικής κατανομής, αλλά το σκεπτικό τους έχει ισχύ και για οποιαδήποτε άλλη κατανομή, μπορεί κάποιος να ισχυριστεί ότι το ΜΤΣ δεν είναι το πλέον κατάλληλο μέτρο αξιολόγησης εκτιμητών του σ^2 για τον εξής λόγο. Επειδή $\sigma^2 > 0$ είναι πολύ λογικό να απαιτήσουμε από κάθε εκτιμητή του, T, να ικανοποιεί την συνθήκη $T > 0$. Έτσι, για έναν τέτοιο εκτιμητή, που όμως υποεκτιμά «υπερβολικά» το σ^2, δηλαδή $T \to 0^+$, αναμένουμε $\mathbb{E}(T - \sigma^2)^2 \to \sigma^4$, ενώ, αν υπερεκτιμά «υπερβολικά» το σ^2, δηλαδή $T \to \infty$, αναμένουμε $\mathbb{E}(T - \sigma^2)^2 \to \infty$. Το ΜΤΣ, λοιπόν, ως κριτήριο επιλογής δεν αντιμετωπίζει ισοδύναμα υποεκτίμηση και υπερεκτίμηση του σ^2 και φαίνεται να «μεροληπτεί» υπέρ της υποεκτίμησης, αφού το ΜΤΣ υποεκτίμησης είναι πάντοτε πεπερασμένο, ενώ το ΜΤΣ υπερεκτίμησης μπορεί να γίνει αυθαίρετα μεγάλο.

Υπό το φως αυτής της διαπίστωσης των Stein (1964) και Casella and Berger (2002, σελ. 332) ας επανέλθουμε στους εκτιμητές S_2^2 και S_1^2. Όπως αναφέρεται στην Παρατήρηση 4.3.1, επειδή οι S_2^2 και S_1^2 έχουν την «τάση» να υποεκτιμούν το σ^2, ίσως λοιπόν η υπεροχή τους έναντι της δειγματικής διασποράς S^2 να οφείλεται ακριβώς σε αυτήν την «μεροληπτική» συμπεριφορά του ΜΤΣ υπέρ της υποεκτίμησης!!!

Κλείνοντας αυτήν την ενότητα, σημειώνουμε ότι, για $n \geq 10$, οι σταθερές $c_2 = \frac{(n+2)(n-1)}{n(n+1)}$, $c_1 = \frac{n(n-1)}{n(n-1)+2}$ και $\frac{1}{n-1}$ διαφέρουν πολύ λίγο μεταξύ τους, οπότε στην πράξη, για τέτοια δειγματικά μεγέθη, οι εκτιμητές S_2^2, S_1^2 και S^2 είναι σχεδόν ίσοι.

4.4 Ασκήσεις

4.1. Εάν $\underset{\sim}{X} = (X_1, \ldots, X_n)$ είναι ένα τυχαίο δείγμα από την ομοιόμορφη κατανομή $\mathcal{U}(0, \theta)$, $\theta \in \Theta = (0, \infty)$, να δειχθεί ότι $T(\underset{\sim}{X}) = \frac{2}{n}\sum_{i=1}^{n} X_i$ είναι αμερόληπτος εκτιμητής του θ και να βρεθεί το ΜΤΣ του.

4.2. Εάν $\underset{\sim}{X} = (X_1, \ldots, X_n)$ είναι ένα τυχαίο δείγμα από την ομοιόμορφη κατανομή $\mathcal{U}(0, \theta)$, $\theta \in \Theta = (0, \infty)$, να βρεθεί ένας αμερόληπτος εκτιμητής του θ^2.

4.3. Εάν X είναι μία παρατήρηση από την κατανομή Poisson $\mathcal{P}(\theta)$, $\theta \in \Theta = (0, \infty)$, να βρεθεί ένας αμερόληπτος εκτιμητής του θ^2.

4.4. Εάν $\underset{\sim}{X} = (X_1, \ldots, X_n)$ είναι ένα τυχαίο δείγμα από την κατανομή Bernoulli $\mathcal{B}(1, \theta)$, $\theta \in \Theta = (0, 1)$, να βρεθεί ένας αμερόληπτος εκτιμητής του $\theta(1 - \theta)$.

4.5. Έστω $X_1, X_2, \ldots, X_n$ είναι ανεξάρτητες τυχαίες μεταβλητές, όπου κάθε $X_i \sim \mathcal{B}(n_i, \theta)$, $i = 1, 2, \ldots, n$. Να βρεθεί ένας αμερόληπτος εκτιμητής του θ και να υπολογιστεί το ΜΤΣ.

4.6. Εάν X είναι μία παρατήρηση από την κατανομή Poisson $\mathcal{P}(\theta)$, $\theta \in \Theta = (0, \infty)$, να βρεθεί ένας αμερόληπτος εκτιμητής του $g(\theta) = \mathbb{P}_\theta(X = k) = e^{-\theta}\frac{\theta^k}{k!}$ για $k = 0, 1, 2, \ldots$.

4.7. Εάν $\underset{\sim}{X} = (X_1, X_2)$ είναι ένα τυχαίο δείγμα από μια κατανομή με διασπορά σ^2, να δειχθεί ότι η στατιστική συνάρτηση $\frac{(X_1-X_2)^2}{2}$ είναι αμερόληπτος εκτιμητής του σ^2.

4.8. Δίνεται δείγμα $\underset{\sim}{X} = (X_1, X_2, \ldots, X_n)$ και οι στατιστικές συναρτήσεις $T_1(\underset{\sim}{X})$, $T_2(\underset{\sim}{X})$ και $T_3(\underset{\sim}{X})$, οι οποίες είναι ανεξάρτητες μεταξύ τους. Οι T_1, T_2 και $\frac{T_3}{2}$ είναι αμερόληπτοι εκτιμητές του θ και έχουν την ίδια διασπορά. Θεωρούμε την κλάση εκτιμητών του θ, $\mathcal{C} = \{cT_1+cT_2+(1-2c)T_3 : c \in \mathbb{R}\}$.

1. Να βρεθεί το στοιχείο της κλάσης $\mathcal{C}$ που είναι αμερόληπτος εκτιμητής του θ.

2. Να βρεθεί το στοιχείο της κλάσης $\mathcal{C}$ με την ελάχιστη διασπορά.

3. Να δειχθεί ότι η στατιστική συνάρτηση T_3 δεν είναι αποδεκτή με κριτήριο το ΜΤΣ.

4.9. Εάν $\underset{\sim}{X} = (X_1, X_2, \ldots, X_n)$ είναι ένα τυχαίο δείγμα από κάποια κατανομή με μέση τιμή μ και διασπορά σ^2, να δειχθεί ότι,

1. Η στατιστική συνάρτηση $\sum_{i=1}^{n} a_i X_i$ είναι αμερόληπτος εκτιμητής του μ, όταν οι σταθερές a_i ικανοποιούν τη συνθήκη $\sum_{i=1}^{n} a_i = 1$.

2. Αν $\sum_{i=1}^{n} a_i = 1$, βρείτε για ποια a_i ελαχιστοποιείται η $\mathbb{V}ar_\theta\left(\sum_{i=1}^{n} a_i X_i\right)$.

4.10. Εάν $T(\underset{\sim}{X})$ είναι αμερόληπτος εκτιμητής του θ, να δειχθεί ότι η στατιστική συνάρτηση $T(\underset{\sim}{X}) + b$ είναι μη αποδεκτός εκτιμητής του θ με κριτήριο το ΜΤΣ.

4.11. Εάν οι εκτιμητές του θ, T_1 και T_2 έχουν ίσα ΜΤΣ για κάθε $\theta \in \Theta$, να δειχθεί ότι ο εκτιμητής $T = aT_1 + (1-a)T_2$, όπου a σταθερά με $0 < a < 1$, είναι καλύτερος και από τους δύο (με κριτήριο το ΜΤΣ).

Βιβλιογραφία

1. Brewster, J. F. and Zidek, J. V. (1974). Improving on equivariant estimators. *Ann. Statist.*, **2**, 21–38.

2. Brown, L.D. (1968). Inadmissiblity of the usual estimators of scale parameters in problems with unknown location and scale parameters. *Ann. Math. Statist.*, **39**, 29–48.

3. Casella, G. and Berger, R.L. (2002). *Statistical Inference.* Duxbury Press; 2nd edition.

4. Lehmann, E.L. and Casella, G. (1998). *Theory of point estimation.* Springer; 2nd edition.

5. Kourouklis, S. (2012). A new estimator of the variance based on minimizing the mean squared error. *The American Statistician*, **66**, 234–236.

6. Maruyama, Y. (1998). Minimax estimators of a normal variance. *Metrika*, **48**, 209–214.

7. Maruyama, Y. and Strawderman, W. E. (2006). A new class of minimax generalized Bayes estimators of a normal variance. *J. Statist. Plann. Inference* **136**, 3822–3836.

8. Stein, C. (1964). Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. *Ann. Inst. Statist. Math.*, **16**, 155–160.

9. Yatracos, Y.G. (2005). Artificially augmented samples, shrinkage, and mean squared error reduction. *Journal of the American Statistical Association*, **100**, 1168–1175.

10. Ηλιόπουλος, Γ. (2013). *Βασικές Μέθοδοι Εκτίμησης Παραμέτρων.* Εκδόσεις Σταμούλη; 2η έκδοση.

Κεφάλαιο 5

Ανισότητα των Cramér–Rao, Πληροφορία του Fisher και Αποδοτικοί Εκτιμητές

Στο Κεφάλαιο 4 είδαμε ότι ένας τρόπος να αντιπαρέλθουμε το πρόβλημα της μη ύπαρξης βέλτιστου εκτιμητή μεταξύ όλων των εκτιμητών, με κριτήριο το ΜΤΣ, είναι να περιορίσουμε την κλάση των εκτιμητών, απαιτώντας οι εκτιμητές να έχουν ειδική μορφή ή να ικανοποιούν κάποια «λογική» συνθήκη. Εκτιμητές ειδικής μορφής μελετήσαμε στα Παραδείγματα 4.1.3, 4.1.4, 4.2.6 και μάλιστα καταλήξαμε ότι μεταξύ αυτών υπάρχει καλύτερος ως προς το ΜΤΣ. Αναφέρουμε εδώ ότι με εκτιμητές ειδικής μορφής θα ασχοληθούμε πιο συστηματικά στην Ενότητα 6.4. Από την άλλη πλευρά, μέσα από τη μελέτη του ΜΤΣ αναδύθηκε η συνθήκη-ιδιότητα της αμεροληψίας ενός εκτιμητή. Από τη σχέση (4.2) ή (4.3), ένας αμερόληπτος εκτιμητής έχει ΜΤΣ ίσο προς τη διασπορά του. Το σύνολο των αμερόληπτων εκτιμητών, αν δεν είναι κενό ή μονοσύνολο, όπως αποδείχθηκε στην Πρόταση 4.2.1, είναι μη αριθμήσιμο. Σε αυτήν την περίπτωση, μεταξύ όλων αυτών των αμερόληπτων εκτιμητών, καλύτερος ως προς το ΜΤΣ είναι εκείνος, αν υπάρχει, που έχει ελάχιστη διασπορά για κάθε $\theta \in \Theta$. Στο κεφάλαιο αυτό θα ασχοληθούμε με την ύπαρξη ενός κάτω φράγματος για τη διασπορά ενός αμερόληπτου εκτιμητή. Στη συνέχεια θα μελετήσουμε πώς αυτό το κάτω φράγμα μπορεί να χρησιμοποιηθεί, προκειμένου να βρεθεί αμερόληπτος εκτιμητής ελάχιστης διασποράς και μάλιστα ίσης προς το κάτω φράγμα. Ένας τέτοιος εκτιμητής αναφέρεται ως *αποδοτικός εκτιμητής* (efficient estimator). Θα δείξουμε, συγκεκριμένα, ότι αποδοτικοί εκτιμητές υπάρχουν μόνον, όταν η κατανομή του δείγματος $\underset{\sim}{X}$ ανήκει σε μια ειδική οικογένεια κατανομών, την εκθετική οικογένεια κατανομών. Αυτό το κάτω φράγμα προκύπτει από μια διάσημη πλέον ανισότητα που παρουσιάστηκε την ίδια περίπου χρονική περίοδο (στα μέσα της δεκαετίας 1940-1950, όταν η μετάδοση και διακίνηση πληροφορίας γινόταν με άλλους ρυθμούς από τους σημερινούς), ανεξάρτητα, από τους Frechét (1943), Rao (1945), Darmois (1945), Cramér (1946) και αναφέρεται στη βιβλιογραφία ως ανισότητα των Cramér - Rao (η επικρατέστερη ονομασία) ή ανισότητα των Frechét - Cramér - Rao (π.χ. Rohatgi (1976)) ή Information Inequality (π.χ. Lehmann and Casella (1998), Bickel and Doksum (1977)) ή Cramér - Rao (Information) Inequality (DeGroot and Schervish (2010)). Επιπροσθέτως, το κάτω φράγμα διαδραματίζει σημαντικό ρόλο στη μελέτη των ασυμπτωτικών ιδιοτήτων των εκτιμητών μέγιστης πιθανοφάνειας (όταν δηλαδή το μέγεθος του δείγματος $n \to \infty$). Μεταγενέστερα, η ανισότητα έχει χρησιμοποιηθεί και ως «εργαλείο» για την απόδειξη αποδεκτικότητας (admissibility) και minimaxity εκτιμητών (βλέπε Lehmann and Casella, 1998, Section 5.2). Τέλος, περιέχει, ως συνιστώσα της, μία θεμελιακή στατιστική «ποσότητα», τον αριθμό πληροφορίας του Fisher, κάτι που της προσδίδει ιδιαίτερο ενδιαφέρον και δικαιολογεί μία από τις ονομασίες της.

5.1 Το κάτω φράγμα των Cramér–Rao και ο αριθμός πληροφορίας του Fisher

Υποθέτουμε ότι τα δεδομένα $\underset{\sim}{X} = (X_1, \dots, X_n)$ έχουν πυκνότητα $f(\underset{\sim}{x}; \theta)$, $\theta \in \Theta \subset \mathbb{R}$. Θα χρειαστούμε τις εξής συνθήκες που διατυπώνονται για συνεχή κατανομή του $\underset{\sim}{X}$. Ανάλογα, για διακριτό $\underset{\sim}{X}$ με αριθμήσιμο σύνολο τιμών, τα ολοκληρώματα που υπάρχουν στις συνθήκες αντικαθίστανται με σειρές, ενώ για διακριτό $\underset{\sim}{X}$ με πεπερασμένο σύνολο τιμών οι αντίστοιχες συνθήκες περιέχουν πεπερασμένα αθροίσματα και ισχύουν, όπως θα διαπιστώσουμε παρακάτω.

I1. Το Θ είναι ένα ανοικτό σύνολο του $\mathbb{R}$.

I2. Το σύνολο τιμών του $\underset{\sim}{X}$, $\mathcal{S} = \big\{\underset{\sim}{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n : f(\underset{\sim}{x};\theta) > 0\big\}$ δεν εξαρτάται από το θ. Για κάθε $\underset{\sim}{x} \in \mathcal{S}$ και $\theta \in \Theta$, η παράγωγος $\frac{\partial}{\partial\theta} f(\underset{\sim}{x};\theta)$ υπάρχει (και είναι πεπερασμένη).

I3. $\displaystyle\int_{\mathcal{S}} \frac{\partial}{\partial\theta} f(\underset{\sim}{x};\theta)\,\mathrm{d}\underset{\sim}{x} = \frac{\partial}{\partial\theta}\int_{\mathcal{S}} f(\underset{\sim}{x};\theta)\,\mathrm{d}\underset{\sim}{x}$ (= 0), $\forall\, \theta \in \Theta$.

I4. $\displaystyle\int_{\mathcal{S}} T(\underset{\sim}{x}) \frac{\partial}{\partial\theta} f(\underset{\sim}{x};\theta)\,\mathrm{d}\underset{\sim}{x} = \frac{\partial}{\partial\theta}\int_{\mathcal{S}} T(\underset{\sim}{x}) f(\underset{\sim}{x};\theta)\,\mathrm{d}\underset{\sim}{x}$, $\forall\, \theta \in \Theta$, όπου $T(\underset{\sim}{X})$ είναι μία στατιστική συνάρτηση.

I5. $0 < I(\theta) < \infty$, $\forall\, \theta \in \Theta$, όπου $I(\theta) = \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta)\right)^2\right]$.

Θεώρημα 5.1.1. **(Ανισότητα των Cramér–Rao)** *Έστω ότι ισχύουν οι συνθήκες* I1–I5 *και ας θέσουμε* $\tau(\theta) = \mathbb{E}_\theta T(\underset{\sim}{X})$ *και* $\tau'(\theta) = \frac{\partial}{\partial\theta}\mathbb{E}_\theta T(\underset{\sim}{X})$. *Τότε έχουμε*

$$\mathbb{V}ar_\theta T(\underset{\sim}{X}) \geq \frac{\big(\tau'(\theta)\big)^2}{I(\theta)}, \quad \forall\, \theta \in \Theta. \tag{5.1}$$

Εάν επιπλέον η στατιστική συνάρτηση $T(\underset{\sim}{X})$ *είναι αμερόληπτος εκτιμητής του* $g(\theta)$, *τότε έχουμε*

$$\mathbb{V}ar_\theta T(\underset{\sim}{X}) \geq \frac{\big(g'(\theta)\big)^2}{I(\theta)}, \quad \forall\, \theta \in \Theta, \tag{5.2}$$

όπου $g'(\theta)$ *είναι η παράγωγος της συνάρτησης* $g(\theta)$.

Απόδειξη. Θεωρούμε ότι το $\underset{\sim}{X}$ έχει συνεχή κατανομή. Η απόδειξη στη διακριτή περίπτωση είναι ανάλογη, αντικαθιστώντας τα ολοκληρώματα με αθροίσματα ή σειρές. Η απόδειξη βασίζεται στην εφαρμογή της ανισότητας Cauchy-Schwarz (βλέπε Πρόταση 1.8.2) για τις τυχαίες μεταβλητές $\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta)$ και $T(\underset{\sim}{X})$. Αρχικά, ως προετοιμασία για την εφαρμογή της θα δείξουμε τις σχέσεις

$$\mathbb{E}_\theta\left(\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta)\right) = 0\ , \quad \forall\theta \in \Theta\ , \tag{5.3}$$

$$\mathbb{V}ar_\theta\left(\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta)\right) = I(\theta)\ , \quad \forall\theta \in \Theta\ , \tag{5.4}$$

$$\mathbb{C}ov_\theta\left(\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta), T(\underset{\sim}{X})\right) = \tau'(\theta), \quad \forall\, \theta \in \Theta. \tag{5.5}$$

Παρατηρούμε ότι, λόγω της I3,

$$\mathbb{E}_\theta\left(\tfrac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta)\right) = \int_{\mathcal{S}} \left(\tfrac{\partial}{\partial\theta}\ln f(\underset{\sim}{x};\theta)\right)\cdot f(\underset{\sim}{x};\theta)\,\mathrm{d}\underset{\sim}{x} = \int_{\mathcal{S}} \tfrac{\partial}{\partial\theta} f(\underset{\sim}{x};\theta)\,\mathrm{d}\underset{\sim}{x}$$
$$= \tfrac{\partial}{\partial\theta}\int_{\mathcal{S}} f(\underset{\sim}{x};\theta)\,\mathrm{d}\underset{\sim}{x} = \tfrac{\partial}{\partial\theta}(1) = 0,$$

δηλαδή, η (5.3) ισχύει. Ακολούθως, από το γενικό τύπο $\mathbb{V}arY = \mathbb{E}(Y^2) - (\mathbb{E}Y)^2$, θέτοντας $Y = \frac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta)$ και χρησιμοποιώντας την (5.3) έχουμε

$$\mathbb{V}ar_\theta\left(\tfrac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta)\right) = \mathbb{E}_\theta\left[\left(\tfrac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta)\right)^2\right] = I(\theta)\ ,$$

δηλαδή η (5.4) ισχύει.
Λόγω της (5.3) και της I4, προκύπτει ότι η συνδιασπορά των $\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta)$ και $T(\underset{\sim}{X})$ είναι

$$\begin{aligned}
&\mathbb{C}ov_\theta\left(\tfrac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta), T(\underset{\sim}{X})\right)\\
&= \mathbb{E}_\theta\left(\tfrac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta)\cdot T(\underset{\sim}{X})\right) - \mathbb{E}_\theta\left(\tfrac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta)\right)\cdot \mathbb{E}_\theta T(\underset{\sim}{X})\\
&= \mathbb{E}_\theta\left(\tfrac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta)\cdot T(\underset{\sim}{X})\right) = \int_{\mathcal{S}} \tfrac{\partial}{\partial\theta}\ln f(\underset{\sim}{x};\theta)\cdot T(\underset{\sim}{x})\cdot f(\underset{\sim}{x};\theta)\,\mathrm{d}\underset{\sim}{x}\\
&= \int_{\mathcal{S}} T(\underset{\sim}{x})\cdot \tfrac{\partial}{\partial\theta} f(\underset{\sim}{x};\theta)\,\mathrm{d}\underset{\sim}{x} = \tfrac{\partial}{\partial\theta}\int_{\mathcal{S}} T(\underset{\sim}{x})\cdot f(\underset{\sim}{x};\theta)\,\mathrm{d}\underset{\sim}{x}\\
&= \tfrac{\partial}{\partial\theta}\mathbb{E}_\theta T(\underset{\sim}{X}) = \tau'(\theta),
\end{aligned}$$

δηλαδή, ισχύει και η (5.5).
Εφαρμόζοντας τώρα την ανισότητα Cauchy–Schwarz για τις $\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta)$ και $T(\underset{\sim}{X})$ και χρησιμοποιώντας την (5.4), από την (5.5), παίρνουμε

$$\left(\tau'(\theta)\right)^2 \le \mathbb{V}ar_\theta\left(T(\underset{\sim}{X})\right)\cdot I(\theta)$$

και επομένως, λόγω της I5,

$$\mathbb{V}ar_\theta T(\underset{\sim}{X}) \ge \frac{\left(\tau'(\theta)\right)^2}{I(\theta)}, \quad \forall\ \theta\in\Theta.$$

Για την ολοκλήρωση της απόδειξης του θεωρήματος, παρατηρούμε ότι η (5.2) προκύπτει αμέσως από την (5.1) και τον ορισμό της αμεροληψίας, επειδή $\tau(\theta) = \mathbb{E}_\theta T(\underset{\sim}{X}) = g(\theta)$, για κάθε $\theta\in\Theta$. □

Παραθέτουμε μερικά διευκρινιστικά σχόλια για το ρόλο των συνθηκών I1 – I5, αλλά και για την καλύτερη κατανόησή τους. Η συνθήκη I1 διασφαλίζει ότι κάθε σημείο $\theta \in \Theta$ είναι εσωτερικό σημείο και συνεπώς μπορεί να οριστεί η παράγωγος ως προς θ (ως όριο). Δεν χρειάζεται πουθενά αλλού στην απόδειξη του Θεωρήματος 5.1.1. Για αυτό το λόγο, η I1 μπορεί να παραληφθεί και στις συνθήκες I2 – I5, όπως και στο Θεώρημα 5.1.1, το Θ να αντικατασταθεί με το Θ^{o}, το σύνολο των εσωτερικών σημείων του Θ.

Η συνθήκη ότι το σύνολο $\mathcal{S} = \{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : f(\underset{\sim}{x};\theta) > 0\}$ δεν εξαρτάται από το θ (μέρος της συνθήκης I2) δεν φαίνεται ρητά να χρησιμοποιείται στην απόδειξη του Θεωρήματος 5.1.1, όμως αποτελεί εν γένει απαραίτητη προϋπόθεση προκειμένου να ικανοποιούνται οι I3 και I4. Προς διευκρίνιση αυτού, έστω $n = 1$, δηλαδή $\underset{\sim}{X} = X_1$ και ότι η πυκνότητα της παρατήρησης X_1 είναι $f(x_1;\theta) = e^{-(x_1-\theta)}$, $x_1 \geqslant \theta$, $\theta \in \Theta = \mathbb{R}$, οπότε $\mathcal{S} = [\theta, \infty)$ που εξαρτάται από το θ. Τότε $\int_{\mathcal{S}} \frac{\partial}{\partial\theta} f(x_1;\theta)dx_1 = \int_\theta^\infty e^{-x_1+\theta}dx_1 = 1$, ενώ $\frac{\partial}{\partial\theta}\int_{\mathcal{S}} f(x_1;\theta)dx_1 = \frac{\partial}{\partial\theta}(1) = 0$, δηλαδή η I3 δεν ισχύει. Όσον αφορά τη συνθήκη I4, αυτή τετριμμένα ισχύει, αν $T(X_1) = 0$ με πιθανότητα 1 (και τότε η (5.1) ισχύει ως ισότητα με τιμή μηδέν). Θα δείξουμε ότι δεν υπάρχει άλλη στατιστική συνάρτηση $T(X_1)$ που ικανοποιεί την I4. Επειδή $\frac{\partial}{\partial\theta} f(x_1;\theta) = e^{-(x_1-\theta)}$, $x_1 > \theta$, η I4 ισοδύναμα γράφεται

$$\int_\theta^\infty T(x_1)e^{-(x_1-\theta)}dx_1 = \frac{\partial}{\partial\theta}\int_\theta^\infty T(x_1)e^{-(x_1-\theta)}dx_1 \text{ ή } \tau(\theta) = \tau'(\theta),\ \theta \in \mathbb{R},$$

όπου $\tau(\theta) = \mathbb{E}_\theta T(X_1)$. Η γενική λύση αυτής της διαφορικής εξίσωσης είναι $\tau(\theta) = ce^\theta$, όπου c σταθερά, δηλαδή $\int_\theta^\infty T(x_1)e^{-(x_1-\theta)}dx_1 = ce^\theta$, $\theta \in \mathbb{R}$, οπότε $\int_\theta^\infty T(x_1)e^{-x_1}dx_1 = c$, για κάθε $\theta \in \mathbb{R}$. Από τη θεωρία ολοκλήρωσης, η τελευταία σχέση συνεπάγεται $T(X_1) = 0$ με πιθανότητα 1. Μία πειστική εναλλακτική απόδειξη μπορεί να δοθεί, αν θεωρήσουμε ότι η T είναι συνεχής συνάρτηση. Σε αυτήν την περίπτωση, παραγωγίζοντας ως προς θ τη σχέση $\int_\theta^\infty T(x_1)e^{-x_1}dx_1 = c$ παίρνουμε $-T(\theta)e^{-\theta} = 0$, για κάθε $\theta \in \mathbb{R}$, δηλαδή $T(\theta) = 0$, για κάθε $\theta \in \mathbb{R}$ και άρα $T(X_1) = 0$. Ανεξάρτητα από το παράδειγμα, έστω $n = 1$, και ότι το σύνολο $\mathcal{S}$ είναι ένα διάστημα $(\alpha(\theta), \beta(\theta))$, όπου $\alpha(\theta)$ και $\beta(\theta)$ είναι παραγωγίσιμες συναρτήσεις. Τότε, από τον τύπο του Leibnitz

($\frac{\partial}{\partial\theta}\int_{a(\theta)}^{b(\theta)} h(x,\theta)dx = \int_{a(\theta)}^{b(\theta)} \frac{\partial}{\partial\theta}h(x,\theta)dx + b'(\theta)h(b(\theta),\theta) - a'(\theta)h(a(\theta),\theta)$, βλέπε Apostol (1969, σελ. 219 – 220)), προκύπτει ότι

$$\begin{aligned}\frac{\partial}{\partial\theta}\int_{\mathcal{S}} T(x_1)f(x_1;\theta)dx_1 &= \frac{\partial}{\partial\theta}\int_{\alpha(\theta)}^{\beta(\theta)} T(x_1)f(x_1;\theta)dx_1 \\ &= \int_{\alpha(\theta)}^{\beta(\theta)} T(x_1)\tfrac{\partial}{\partial\theta}f(x_1;\theta)dx_1 \\ &\quad + \beta'(\theta)T(\beta(\theta))f(\beta(\theta);\theta) - \alpha'(\theta)T(\alpha(\theta))f(\alpha(\theta);\theta).\end{aligned}$$

Παρατηρούμε λοιπόν ότι, για να ισχύουν η I3 (που είναι η I4 με $T(x_1) = 1$) και η I4, πρέπει $\beta'(\theta)T(\beta(\theta))f(\beta(\theta);\theta) - \alpha'(\theta)T(\alpha(\theta))f(\alpha(\theta);\theta) = 0$, για κάθε $\theta \in \Theta$, το οποίο εξασφαλίζεται (για τυχούσα στατιστική συνάρτηση $T(X_1)$), αν οι συναρτήσεις $\alpha(\theta)$ και $\beta(\theta)$ είναι σταθερές ως προς θ, γιατί τότε $\alpha'(\theta) = \beta'(\theta) = 0$, $\theta \in \Theta$. Με $\alpha(\theta) = c_1$ και $\beta(\theta) = c_2$, όπου c_1 και c_2 σταθερές μη εξαρτώμενες από το θ (ενδεχομένως $c_1 = -\infty$ ή/και $c_2 = +\infty$), το $\mathcal{S} = (c_1, c_2)$ όντως δεν εξαρτάται από το θ, και οι I3, I4 είναι δυνητικά σε ισχύ.

Ακόμη και όταν ισχύει η I2, η επαλήθευση των συνθηκών I3 και I4 δεν είναι γενικά εύκολη και συνήθως ανάγεται στην επαλήθευση άλλων ικανών συνθηκών που επιτρέπουν την εναλλαγή στη σειρά εκτέλεσης παραγώγισης και ολοκλήρωσης στη συνεχή περίπτωση ή παραγώγισης και σειράς στη διακριτή. Τέτοιες ικανές συνθήκες μελετώνται σε βιβλία Ανάλυσης ή Θεωρίας Μέτρου (π.χ. Apostol (1969), Billingsley (1995)). Παραπέμπουμε όμως τον αναγνώστη στο σύγγραμμα Πιθανοτήτων -Στατιστικής Casella and Berger (2002, σελ. 68-75) που περιέχει μία ενδελεχή παρουσίαση αυτών των συνθηκών καθώς και σε αυτό του Rohatgi (1976, σελ. 12).
Εξαίρεση αποτελεί η περίπτωση που τα δεδομένα $\underset{\sim}{X}$ έχουν διακριτή κατανομή με *πεπερασμένο* σύνολο τιμών $\mathcal{S}$ (π.χ. τυχαίο δείγμα από διωνυμική κατανομή). Τότε, στις συνθήκες I3 και I4, αντί ολοκληρωμάτων, έχουμε πεπερασμένα αθροίσματα και φυσικά ισχύουν οι I3 και I4, αφού η παράγωγος (πεπερασμένου) αθροίσματος συναρτήσεων είναι ίση προς το άθροισμα των παραγώγων αυτών των συναρτήσεων. Για παράδειγμα, έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την κατανομή Bernoulli $\mathcal{B}(1,\theta)$

με πυκνότητα $f_1(x;\theta) = \theta^x(1-\theta)^{1-x}$, $x = 0,1$, $\theta \in \Theta = (0,1)$. Τότε η πυκνότητα του $\underset{\sim}{X}$ είναι $f(\underset{\sim}{x};\theta) = \prod_{i=1}^n f_1(x_i;\theta) = \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum\limits_{i=1}^{n} x_i}$, $\underset{\sim}{x} = (x_1,\ldots,x_n) \in \mathcal{S} = \{0,1\} \times \{0,1\} \times \ldots \times \{0,1\}$. Το σύνολο $\mathcal{S}$ προφανώς δεν εξαρτάται από το θ και είναι πεπερασμένο (με 2^n στοιχεία). Η I3 έχει τη μορφή, λόγω διακριτού $\underset{\sim}{X}$,

$$\sum_{\underset{\sim}{x}\in\mathcal{S}} \tfrac{\partial}{\partial\theta} f(\underset{\sim}{x};\theta) = \tfrac{\partial}{\partial\theta} \sum_{\underset{\sim}{x}\in\mathcal{S}} f(\underset{\sim}{x};\theta),$$

ενώ η I4 για τυχούσα στατιστική συνάρτηση $T(\underset{\sim}{X})$ γράφεται ως

$$\sum_{\underset{\sim}{x}\in\mathcal{S}} T(\underset{\sim}{x})\tfrac{\partial}{\partial\theta} f(\underset{\sim}{x};\theta) = \sum_{\underset{\sim}{x}\in\mathcal{S}} \tfrac{\partial}{\partial\theta}\big(T(\underset{\sim}{x})f(\underset{\sim}{x};\theta)\big) = \tfrac{\partial}{\partial\theta} \sum_{\underset{\sim}{x}\in\mathcal{S}} T(\underset{\sim}{x}) f(\underset{\sim}{x};\theta).$$

Σε κάθε μία από αυτές τις σχέσεις το αριστερό μέλος είναι (πεπερασμένο) άθροισμα παραγώγων συναρτήσεων, ενώ το δεξιό μέλος είναι η παράγωγος του αντίστοιχου (πεπερασμένου) αθροίσματος συναρτήσεων, συνεπώς τα δύο μέλη είναι ίσα, το οποίο σημαίνει ότι η I3 και η I4 αληθεύουν. Συμπληρωματικά, αναφέρουμε ότι η συνθήκη I3 είναι ισοδύναμη με την

Ι′ 3. $\int_{\mathcal{S}} \frac{\partial}{\partial\theta} f(\underset{\sim}{x};\theta) d\underset{\sim}{x} = 0$, $\forall\theta \in \Theta$,

επειδή $\int_{\mathcal{S}} f(\underset{\sim}{x};\theta)d\underset{\sim}{x} = 1$ και επομένως, $\frac{\partial}{\partial\theta}\int_{\mathcal{S}} f(\underset{\sim}{x};\theta)d\underset{\sim}{x} = 0$, για κάθε $\theta \in \Theta$. Για διακριτό $\underset{\sim}{X}$, το ολοκλήρωμα αντικαθίσταται με σειρά ή πεπερασμένο άθροισμα και στην τελευταία περίπτωση η συνθήκη ισχύει (όπως και η I3). Σε μεμονωμένες περιπτώσεις, η Ι′3 μπορεί να επαληθευτεί με υπολογισμό του ολοκληρώματος (ή της αντίστοιχης σειράς). Για παράδειγμα, έστω $\underset{\sim}{X} = (X_1,\ldots,X_n)$ ένα τυχαίο δείγμα από την εκθετική κατανομή $\mathcal{E}(1/\theta)$ με πυκνότητα $f_1(x;\theta) = \theta e^{-\theta x}$, $x > 0$, $\theta \in \Theta = (0,\infty)$ και $\mathbb{E}_\theta X_i = \frac{1}{\theta}$. Επειδή $f(\underset{\sim}{x};\theta) = \prod\limits_{i=1}^{n} f_1(x_i;\theta) = \theta^n e^{-\theta\sum\limits_{i=1}^{n} x_i}$, $x_i > 0$, $i = 1,2,\ldots,n$ και $\frac{\partial}{\partial\theta} f(\underset{\sim}{x};\theta) = n\theta^{n-1}e^{-\theta\sum\limits_{i=1}^{n} x_i} - (\sum\limits_{i=1}^{n} x_i)\theta^n e^{-\theta\sum\limits_{i=1}^{n} x_i}$, με

$\mathcal{S} = (0, \infty)^n$, έχουμε

$$\begin{aligned}
\int_{\mathcal{S}} \tfrac{\partial}{\partial\theta} f(\underset{\sim}{x};\theta)d\underset{\sim}{x} &= \int_{\mathcal{S}} n\theta^{n-1} e^{-\theta \sum\limits_{i=1}^{n} x_i} d\underset{\sim}{x} - \int_{\mathcal{S}} (\sum_{i=1}^{n} x_i)\theta^n e^{-\theta \sum\limits_{i=1}^{n} x_i} d\underset{\sim}{x} \\
&= \frac{n}{\theta}\int_{\mathcal{S}} f(\underset{\sim}{x};\theta)d\underset{\sim}{x} - \int_{\mathcal{S}} (\sum_{i=1}^{n} x_i) f(\underset{\sim}{x};\theta)d\underset{\sim}{x} = \frac{n}{\theta} - \mathbb{E}_\theta\left(\sum_{i=1}^{n} X_i\right) \\
&= \frac{n}{\theta} - \sum_{i=1}^{n}\mathbb{E}_\theta X_i = \frac{n}{\theta} - \sum_{i=1}^{n}\frac{1}{\theta} = 0,
\end{aligned}$$

για κάθε $\theta \in \Theta$, δηλαδή η Ι΄3 ισχύει (άρα και η Ι3).

Η συνθήκη Ι4 εξαρτάται από τη στατιστική συνάρτηση $T(\underset{\sim}{X})$. Στις εφαρμογές του Θεωρήματος 5.1.1 μας ενδιαφέρει να αληθεύει για οποιοδήποτε αμερόληπτο εκτιμητή του $g(\theta)$, προκειμένου να αναζητήσουμε αποδοτικό εκτιμητή του, διαδικασία που μελετάμε στην επόμενη ενότητα. Πάντως, δοθείσης της $T(\underset{\sim}{X})$, η Ι4 μπορεί να επαληθευτεί υπολογίζοντας χωριστά καθένα από τα δύο μέλη της. Για παράδειγμα, στην περίπτωση τυχαίου δείγματος $\underset{\sim}{X} = (X_1, \ldots, X_n)$ από την εκθετική κατανομή $\mathcal{E}(1/\theta)$ και για $T(\underset{\sim}{X}) = \sum\limits_{i=1}^{n} X_i$, παρόμοιοι υπολογισμοί με τους παραπάνω δίνουν

$$\begin{aligned}
\int_{\mathcal{S}} (\sum_{i=1}^{n} x_i)\tfrac{\partial}{\partial\theta} f(\underset{\sim}{x};\theta)d\underset{\sim}{x} &= \frac{n}{\theta}\mathbb{E}_\theta\left(\sum_{i=1}^{n} X_i\right) - \mathbb{E}_\theta\left(\sum_{i=1}^{n} X_i\right)^2 \\
&= \left(\frac{n}{\theta}\right)^2 - \left\{\mathbb{V}ar_\theta\left(\sum_{i=1}^{n} X_i\right) + \left(\mathbb{E}_\theta\left(\sum_{i=1}^{n} X_i\right)^2\right)\right\} \\
&= \left(\frac{n}{\theta}\right)^2 - \left\{\frac{n}{\theta^2} + \left(\frac{n}{\theta}\right)^2\right\} = -\frac{n}{\theta^2},
\end{aligned}$$

ενώ $\frac{\partial}{\partial\theta}\int_{\mathcal{S}}(\sum i=1^n x_i) f(\underset{\sim}{x};\theta)d\underset{\sim}{x} = \frac{\partial}{\partial\theta}\mathbb{E}_\theta\left(\sum\limits_{i=1}^{n} X_i\right) = \frac{\partial}{\partial\theta}\left(\frac{n}{\theta}\right) = -\frac{n}{\theta^2}$ και επομένως η Ι4 ισχύει.

Οι Ι3 και Ι4 μπορούν επίσης να διατυπωθούν παριστάνοντας τα ολοκληρώματα (και αντίστοιχα, τις σειρές ή τα αθροίσματα) ως μέσες τιμές. Μάλιστα, μια τέτοια συνεκτική παράσταση ενοποιεί τη συνεχή και τη διακριτή περίπτωση. Η Ι3 είναι ισοδύναμη με την Ι΄3, δηλαδή $\int_{\mathcal{S}} \frac{\partial}{\partial\theta} f(\underset{\sim}{x};\theta) d\underset{\sim}{x} = 0$, ενώ $\mathbb{E}_\theta \frac{\partial}{\partial\theta} \ln f(\underset{\sim}{X};\theta) = \int_{\mathcal{S}} \frac{\partial}{\partial\theta} \ln f(\underset{\sim}{x};\theta)\ f(\underset{\sim}{x};\theta) d\underset{\sim}{x} = \int_{\mathcal{S}} \frac{\partial}{\partial\theta} f(\underset{\sim}{x};\theta) d\underset{\sim}{x}$. Κατά συνέπεια, η Ι3 μπορεί να αντικατασταθεί με την

Ι6. $\mathbb{E}_\theta \frac{\partial}{\partial\theta} \ln f(\underset{\sim}{X};\theta) = 0$, $\forall\theta \in \Theta$ (που είναι η (5.3)).

Επίσης, το αριστερό μέλος της Ι4 γράφεται

$$\begin{aligned}\int_{\mathcal{S}} T(\underset{\sim}{x}) \tfrac{\partial}{\partial\theta} f(\underset{\sim}{x};\theta) d\underset{\sim}{x} &= \int_{\mathcal{S}} T(\underset{\sim}{x}) \tfrac{\partial}{\partial\theta} \ln f(\underset{\sim}{x};\theta)\ f(\underset{\sim}{x};\theta) d\underset{\sim}{x} \\ &= \mathbb{E}_\theta \left(T(\underset{\sim}{X}) \tfrac{\partial}{\partial\theta} \ln f(\underset{\sim}{X};\theta) \right)\end{aligned}$$

ενώ το δεξιό της μέλος είναι $\frac{\partial}{\partial\theta} \mathbb{E}_\theta T(\underset{\sim}{X})$. Επομένως, η Ι4 μπορεί να αντικατασταθεί με την

Ι7. $\mathbb{E}_\theta \left(T(\underset{\sim}{X}) \frac{\partial}{\partial\theta} \ln f(\underset{\sim}{X};\theta) \right) = \frac{\partial}{\partial\theta} \mathbb{E}_\theta T(\underset{\sim}{X})$, $\forall\theta \in \Theta$.

Το Θεώρημα 5.1.1 ισχύει λοιπόν και υπό τις συνθήκες Ι1, Ι2, Ι5, Ι6, Ι7. Όμως, σε κάθε ανάγνωση (και όχι σε πρώτη ούτε σε δεύτερη ...) οι Ι6 και Ι7 υπολείπονται σε διαισθητική ερμηνεία των Ι3 και Ι4, και για αυτό τον λόγο δεν χρησιμοποιήθηκαν εξ' αρχής.

Για ορισμένες όμως γνωστές κατανομές, όπως η κανονική, η Poisson, η εκθετική και άλλες, μπορεί να δειχθεί ότι οι συνθήκες Ι3 και Ι4 ισχύουν και μάλιστα η Ι4 είναι αληθής για οποιαδήποτε στατιστική συνάρτηση $T(\underset{\sim}{X})$. Αυτό προκύπτει από ένα γενικό αποτέλεσμα, την Πρόταση 5.2.2, που δίνεται στην επόμενη ενότητα.

Η συνάρτηση $I(\theta)$, που εμφανίζεται στη συνθήκη Ι5, είναι εξ ορισμού μη αρνητική, ως μέση τιμή μη αρνητικής τυχαίας μεταβλητής. Η Ι5 απλώς εξασφαλίζει ότι το δεξιό μέλος των (5.1) και (5.2) είναι ένας καλώς ορισμένος πραγματικός (θετικός) αριθμός (δεν θα ήταν, αν $I(\theta) = 0$ ή $I(\theta) = \infty$). Η επαλήθευσή της γίνεται υπολογίζοντας το $I(\theta)$. Όπως θα διαπιστώσουμε στη συνέχεια, υπό κάποιες συνθήκες που συνήθως ισχύουν, υπάρχουν διάφοροι εναλλακτικοί τρόποι υπολογισμού του $I(\theta)$, πιο εύκολοι από τον ορισμό.

Η ανισότητα (5.2) είναι ειδική περίπτωση της (5.1). Κάθε μία από τις (5.1) και (5.2) είναι γνωστή στη βιβλιογραφία ως ανισότητα των Cramér–Rao ή Information Inequality ή Cramér–Rao (Information) Inequality. Σημειώνουμε επιπροσθέτως ότι το δεξιό μέλος της σχέσης (5.1) εξαρτάται εν γένει από τον εκτιμητή $T(\utilde{X})$, λόγω της παραγώγου $\tau'(\theta) = \frac{\partial}{\partial\theta}\mathbb{E}_\theta T(\utilde{X})$, ενώ το δεξιό μέλος της (5.2), $[g'(\theta)]^2/I(\theta)$, είναι ανεξάρτητο του συγκεκριμένου αμερόληπτου εκτιμητή $T(\utilde{X})$. Συνεπώς, η (5.2) (και όχι η (5.1)) παρέχει ένα κάτω φράγμα ως προς $T(\utilde{X})$ για τη $\mathbb{V}ar_\theta T(\utilde{X})$, το οποίο αναφέρεται ως *κάτω φράγμα των Cramér-Rao για τη διασπορά αμερόληπτων εκτιμητών του* $g(\theta)$ και γράφεται ως

$$\text{Κ.Φ. C-R } = \big(g'(\theta)\big)^2/I(\theta),\ \theta \in \Theta\ , \tag{5.6}$$

(παραλείποντας στο συμβολισμό την εξάρτησή του από το θ).

Παρατηρούμε ότι το Κ.Φ. C-R καθορίζεται πλήρως από τη συνάρτηση g, την τιμή της οποίας, $g(\theta)$, επιθυμούμε να εκτιμήσουμε και από ένα «χαρακτηριστικό» της πυκνότητας $f(\utilde{x};\theta)$ του δείγματος $\utilde{X}$, το $I(\theta)$. Το «χαρακτηριστικό» αυτό (συνάρτηση του $\theta \in \Theta$ ως μαθηματικό αντικείμενο)

$$I(\theta) = \mathbb{E}_\theta\left[\left(\tfrac{\partial}{\partial\theta}\ln f(\utilde{X};\theta)\right)^2\right]$$

ονομάζεται *αριθμός ή μέτρο πληροφορίας ή απλά πληροφορία του Fisher που περιέχεται στο δείγμα* $\utilde{X}$ *για την άγνωστη παράμετρο* θ. Ο συμβολισμός $I(\theta)$, όπως και των συνθηκών I1 – I5, παραπέμπει στο Information (Πληροφορία), ενώ λόγω της ύπαρξης του $I(\theta)$ στις ανισότητες (5.1) και (5.2) δικαιολογείται η εναλλακτική ονομασία κάθε μίας ως Information Inequality.

Μία ενδιαφέρουσα ερμηνεία του $I(\theta)$ που δίνουν οι Lehmann and Casella (1998, σελ. 115) είναι η ακόλουθη. Η παράγωγος $\frac{\partial}{\partial\theta}\ln f(\utilde{x};\theta) = \frac{\partial}{\partial\theta}f(\utilde{x};\theta)/f(\utilde{x};\theta)$ παριστάνει το σχετικό ρυθμό μεταβολής ως προς θ της πυκνότητας $f(\utilde{x};\theta)$. Άρα $I(\theta)$ είναι ο κατά μέσο όρο (ως προς την κατανομή του $\utilde{X}$) τετραγωνικός σχετικός ρυθμός μεταβολής της $f(\utilde{x};\theta)$ στο σημείο θ. Συνεπώς, μία «μεγάλη» τιμή της συνάρτησης $I(\theta)$ στο σημείο $\theta = \theta_0$, υποδηλώνει ταχεία μεταβολή ως προς θ της πυκνότητας $f(\utilde{x};\theta)$

σε μια μικρή περιοχή του θ_0, γεγονός που καθιστά πιο εύκολο να ξεχωρίσουμε το θ_0 από γειτονικά του σημεία και επομένως να εκτιμήσουμε το θ με μεγαλύτερη ακρίβεια αν η τιμή του είναι θ_0. Εν κατακλείδι, για κάθε $\theta \in \Theta$, ο αριθμός $I(\theta)$ παρέχει πληροφορία για την ακρίβεια με την οποία μπορεί να εκτιμηθεί το θ από το δείγμα $\underset{\sim}{X}$, με «μεγάλες» τιμές του $I(\theta)$ να είναι επιθυμητές.

Σε αυτό το σημείο προσθέτουμε ότι για $g(\theta) = \theta$, από την (5.2) παίρνουμε $\mathbb{V}ar_\theta T(\underset{\sim}{X}) \geqslant 1/I(\theta) =$ Κ.Φ. C-R , για κάθε $\theta \in \Theta$ και για κάθε αμερόληπτο εκτιμητή $T(\underset{\sim}{X})$ του θ, που ικανοποιεί την I4. Αν λοιπόν υπάρχει αμερόληπτος εκτιμητής $T_0(\underset{\sim}{X})$, τέτοιος ώστε $\mathbb{V}ar_\theta T_0(\underset{\sim}{X}) = 1/I(\theta)$, $\theta \in \Theta$, τότε επιβεβαιώνεται η παραπάνω ερμηνεία του $I(\theta)$, αφού χρησιμοποιώντας τον $T_0(\underset{\sim}{X})$, για να εκτιμήσουμε το θ μία μεγάλη τιμή της πληροφορίας $I(\theta)$ συνεπάγεται μικρή διασπορά $\mathbb{V}ar_\theta T_0(\underset{\sim}{X})$ και συνεπώς ακριβή εκτίμηση του θ. Με τη μελέτη εκτιμητών όπως ο $T_0(\underset{\sim}{X})$ (που θα τον ονομάσουμε αποδοτικό εκτιμητή) θα ασχοληθούμε στην επόμενη ενότητα.

Η έννοια της πληροφορίας εισήχθηκε από τον Edgeworth (1908), όμως ο αριθμός πληροφορίας $I(\theta)$ μελετήθηκε συστηματικά από τον Fisher (1922, 1925, 1934), και για αυτό, τιμητικά, φέρει το όνομα του. Η έννοια του αριθμού πληροφορίας μπορεί να γενικευθεί, έτσι ώστε να αναφέρεται σε οποιαδήποτε τυχαία μεταβλητή, στατιστική συνάρτηση ή τυχαίο διάνυσμα Y (και όχι μόνον στο δείγμα $\underset{\sim}{X}$). Αν η πυκνότητα του Y είναι $h(y;\theta)$, $\theta \in \Theta$, τότε ο αριθμός ή μέτρο πληροφορίας ή απλά πληροφορία του Fisher που περιέχεται στο Y για την άγνωστη παράμετρο θ ορίζεται από τη σχέση

$$I_Y(\theta) = \mathbb{E}_\theta\left[\left(\tfrac{\partial}{\partial\theta}\ln h(Y;\theta)\right)^2\right]\ ,\ \ \theta \in \Theta. \tag{5.7}$$

Για παράδειγμα, αν $\underset{\sim}{X} = (X_1, \ldots, X_n)$ και η πυκνότητα της παρατήρησης X_i είναι $f_i(x_i;\theta)$ τότε

$$I_{X_i}(\theta) = \mathbb{E}_\theta\left[\left(\tfrac{\partial}{\partial\theta}\ln f_i(X_i;\theta)\right)^2\right]\ ,\ \ \theta \in \Theta$$

είναι η πληροφορία του Fisher που περιέχεται στην X_i για το θ.

Η επόμενη πρόταση παρέχει εναλλακτικούς τύπους και ιδιότητες του $I(\theta)$, που, πέραν από το θεωρητικό ενδιαφέρον που παρουσιάζουν, χρησιμεύουν επίσης και στον υπολογισμό του. Οι συνθήκες της πρότασης διατυπώνονται για συνεχή κατανομή του $\underset{\sim}{X}$. Όπως και στο Θεώρημα 5.1.1, στην περίπτωση διακριτής κατανομής του $\underset{\sim}{X}$ με αριθμήσιμο σύνολο τιμών απαιτούνται αντίστοιχες συνθήκες που προκύπτουν αντικαθιστώντας τα ολοκληρώματα με σειρές. Τέλος, για διακριτή κατανομή του $\underset{\sim}{X}$ με πεπερασμένο σύνολο τιμών οι αντίστοιχες συνθήκες, που προκύπτουν αντικαθιστώντας τα ολοκληρώματα με (πεπερασμένα) αθροίσματα, ισχύουν.

Πρόταση 5.1.2. (διαφορετικές εκφράσεις του $\mathbf{I}(\theta)$) *(i) Έστω ότι ισχύουν οι συνθήκες I1, I2, I3. Τότε έχουμε,*

$$I(\theta) = \mathbb{V}ar_\theta\left(\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta)\right) \ , \ \ \forall\theta\in\Theta. \tag{5.8}$$

(ii) Έστω ότι ισχύουν οι συνθήκες I1, I2 και

I8 Για κάθε $\underset{\sim}{x}\in\mathcal{S}=\left\{\underset{\sim}{x}=(x_1,\ldots,x_n)\in\mathbb{R}^n : f(\underset{\sim}{x};\theta)>0\right\}$ και $\theta\in\Theta$, η δεύτερη παράγωγος $\frac{\partial^2}{\partial\theta^2}\ln f(\underset{\sim}{x};\theta)$ υπάρχει, είναι πεπερασμένη και

$$\int_{\mathcal{S}}\frac{\partial^2}{\partial\theta^2}\ln f(\underset{\sim}{x};\theta)d\underset{\sim}{x} = \frac{\partial^2}{\partial\theta^2}\int_{\mathcal{S}}\ln f(\underset{\sim}{x};\theta)d\underset{\sim}{x} \ (=0) \ , \ \ \forall\theta\in\Theta.$$

Τότε έχουμε,

$$I(\theta) = -\mathbb{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\ln f(\underset{\sim}{X};\theta)\right) \ , \ \ \forall\theta\in\Theta. \tag{5.9}$$

(iii) Έστω ότι το δείγμα $\underset{\sim}{X}=(X_1,\ldots,X_n)$ αποτελείται από ανεξάρτητες παρατηρήσεις X_i με πυκνότητες $f_i(x_i;\theta)$, $i=1,\ldots,n$, αντίστοιχα. Έστω ακόμη ότι ισχύουν οι συνθήκες I1 και οι αντίστοιχες συνθήκες των I2 και I3 για κάθε μία από τις X_i, $i=1,\ldots,n$, (δηλαδή στις I2 και I3, η $f(\underset{\sim}{x};\theta)$ αντικαθίσταται από την $f_i(x_i;\theta)$). Τότε έχουμε,

$$I(\theta) = \sum_{i=1}^{n} I_i(\theta) \ , \ \ \forall\theta\in\Theta. \tag{5.10}$$

(προσθετική ιδιότητα του αριθμού πληροφορίας του Fisher), όπου

$$I_i(\theta) = \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f_i(X_i;\theta)\right)^2\right] = \mathbb{V}ar_\theta\left(\frac{\partial}{\partial\theta}\ln f_i(X_i;\theta)\right) \ , \ \ \forall\theta\in\Theta. \tag{5.11}$$

είναι ο αριθμός πληροφορίας του Fisher που περιέχεται στην παρατήρηση X_i, $i=1,\ldots,n$, για την παράμετρο θ.

(iv) Εάν επιπλέον από τις υποθέσεις της (iii), οι παρατηρήσεις X_i, $i = 1, \ldots, n$ έχουν κοινή κατανομή, τότε

$$I(\theta) = nI_1(\theta)\ , \quad \forall \theta \in \Theta. \tag{5.12}$$

(v) Έστω ότι ισχύουν οι συνθήκες I1 και οι αντίστοιχες συνθήκες των I2 και I8 για κάθε μία από τις X_i, $i = 1, \ldots, n$. Τότε έχουμε,

$$I_i(\theta) = -\mathbb{E}_\theta \left(\tfrac{\partial^2}{\partial \theta^2} \ln f_i(X_i;\theta) \right)\ , \quad \forall \theta \in \Theta. \tag{5.13}$$

Απόδειξη. Υποθέτουμε ότι το $\underset{\sim}{X}$ έχει συνεχή κατανομή. Η απόδειξη στη διακριτή περίπτωση είναι ανάλογη.

(i) Έχει δειχθεί κατά την απόδειξη του Θεωρήματος 5.1.1 (σχέση (5.4)).

(ii) Έχουμε,

$$\begin{aligned}
\mathbb{E}_\theta \left(\tfrac{\partial^2}{\partial \theta^2} \ln f(\underset{\sim}{X};\theta) \right) &= \mathbb{E}_\theta \left(\tfrac{\partial}{\partial \theta} \left(\frac{\frac{\partial}{\partial \theta} f(\underset{\sim}{X};\theta)}{f(\underset{\sim}{X};\theta)} \right) \right) \\
&= \mathbb{E}_\theta \left[\frac{\left\{ \frac{\partial^2}{\partial \theta^2} f(\underset{\sim}{X};\theta) \cdot f(\underset{\sim}{X};\theta) - \left(\frac{\partial}{\partial \theta} f(\underset{\sim}{X};\theta) \right)^2 \right\}}{f^2(\underset{\sim}{X};\theta)} \right] \\
&= \mathbb{E}_\theta \left(\tfrac{\partial^2}{\partial \theta^2} f(\underset{\sim}{X};\theta) / f(\underset{\sim}{X};\theta) \right) - \mathbb{E}_\theta \left(\frac{\frac{\partial}{\partial \theta} f(\underset{\sim}{X};\theta)}{f(\underset{\sim}{X};\theta)} \right)^2 \\
&= \int_{\mathcal{S}} \frac{\frac{\partial^2}{\partial \theta^2} f(\underset{\sim}{x};\theta)}{f(\underset{\sim}{x};\theta)} \cdot f(\underset{\sim}{x};\theta)\, \mathrm{d}\underset{\sim}{x} - \mathbb{E}_\theta \left(\tfrac{\partial}{\partial \theta} \ln f(\underset{\sim}{X};\theta) \right)^2 \\
&= \int_{\mathcal{S}} \tfrac{\partial^2}{\partial \theta^2} f(\underset{\sim}{x};\theta)\, \mathrm{d}\underset{\sim}{x} - I(\theta) = \tfrac{\partial^2}{\partial \theta^2} \int_{\mathcal{S}} f(\underset{\sim}{x};\theta)\, \mathrm{d}\underset{\sim}{x} - I(\theta) \\
&= \tfrac{\partial^2}{\partial \theta^2}(1) - I(\theta) = -I(\theta).
\end{aligned}$$

Επομένως, $I(\theta) = -\mathbb{E}_\theta \left(\frac{\partial^2}{\partial \theta^2} \ln f(\underset{\sim}{X};\theta) \right)$.

(iii) Η πρώτη ισότητα στην (5.11) είναι ο ορισμός του $I_i(\theta)$. Η απόδειξη ότι $I_i(\theta) = \mathbb{V}ar_\theta \left(\frac{\partial}{\partial \theta} \ln f_i(X_i;\theta) \right)$ είναι πανομοιότυπη με την απόδειξη της (5.4), επειδή $\mathbb{E}_\theta \left(\frac{\partial}{\partial \theta} \ln f_i(X_i;\theta) \right) = 0$ που αποδεικνύεται όπως η (5.3). Λόγω ανεξαρτησίας των X_i, $i = 1, \ldots, n$, $f(\underset{\sim}{x};\theta) = \displaystyle\prod_{i=1}^{n} f_i(x_i;\theta)$, οπότε

$\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{x};\theta) = \sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln f_i(x_i;\theta)$ και επομένως έχουμε

$$\begin{aligned}
I(\theta) &= \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta)\right)^2\right] = \mathbb{E}_\theta\left[\left(\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln f_i(X_i;\theta)\right)^2\right]\\
&= \mathbb{E}_\theta\left[\sum_{i=1}^{n}\left(\frac{\partial}{\partial\theta}\ln f_i(X_i;\theta)\right)^2 + \sum_{i\neq j}\frac{\partial}{\partial\theta}\ln f_i(X_i;\theta)\frac{\partial}{\partial\theta}\ln f_j(X_j;\theta)\right]\\
&= \sum_{i=1}^{n}\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f_i(X_i;\theta)\right)^2\right]\\
&+ \sum_{i\neq j}\mathbb{E}_\theta\left(\frac{\partial}{\partial\theta}\ln f_i(X_i;\theta)\right)\ \mathbb{E}_\theta\left(\frac{\partial}{\partial\theta}\ln f_j(X_j;\theta)\right)\\
&= \sum_{i=1}^{n} I_i(\theta).
\end{aligned}$$

Η προτελευταία ισότητα ισχύει λόγω της ανεξαρτησίας των X_i και X_j, $i \neq j$, και η τελευταία επειδή $\mathbb{E}_\theta\left(\frac{\partial}{\partial\theta}\ln f_i(X_i;\theta)\right) = 0$.
(iv) Επειδή οι X_i, $i = 1,\ldots,n$, έχουν την ίδια κατανομή, είναι $f_i(x_i;\theta) = f_1(x_i;\theta)$ και άρα

$$I_i(\theta) = \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f_1(X_i;\theta)\right)^2\right]. \tag{5.14}$$

Επιπλέον, οι τυχαίες μεταβλητές $\left(\frac{\partial}{\partial\theta}\ln f_1(X_i;\theta)\right)^2$, $i = 1,\ldots,n$, έχουν την ίδια κατανομή και συνεπώς την ίδια μέση τιμή, δηλαδή

$$\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f_1(X_i;\theta)\right)^2\right] = \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f_1(X_1;\theta)\right)^2\right]\ ,$$

που λόγω της (5.14) γράφεται $I_i(\theta) = I_1(\theta)$, $i = 1,\ldots,n$. Εξ υποθέσεως, οι X_i είναι ανεξάρτητες, οπότε από την (5.10) παίρνουμε

$$I(\theta) = \sum_{i=1}^{n} I_i(\theta) = \sum_{i=1}^{n} I_1(\theta) = nI_1(\theta).$$

(v) Προκύπτει εντελώς ανάλογα, όπως η (5.9). □

Η προσθετική ιδιότητα (5.10) έχει την προφανή ερμηνεία ότι στην περίπτωση που το δείγμα $\underset{\sim}{X} = (X_1, \ldots, X_n)$ αποτελείται από ανεξάρτητες παρατηρήσεις, η (συνολική) πληροφορία για το θ είναι το αθροιστικό αποτέλεσμα των πληροφοριών που εμπεριέχονται στις X_i, $i = 1, \ldots, n$. Δεν υπάρχει δηλαδή «επικάλυψη» πληροφορίας μεταξύ δύο διαφορετικών παρατηρήσεων. Επιπλέον, η (5.10) παρέχει ένα εναλλακτικό και χρήσιμο τρόπο υπολογισμού του $I(\theta)$: είναι σχετικά πιο εύκολο να υπολογιστεί το $I_i(\theta)$ χρησιμοποιώντας τη μονομεταβλητή (περιθώρια) πυκνότητα της X_i, $f_i(x_i; \theta)$, μέσω της (5.11) ή της (5.13), και μετά να εφαρμοστεί η (5.10), παρά να χρησιμοποιηθεί μία από τις (5.8) ή (5.9) που απαιτούν τον υπολογισμό της πολυμεταβλητής πυκνότητας του δείγματος $\underset{\sim}{X} = (X_1, \ldots, X_n)$, $f(\underset{\sim}{x}; \theta)$. Η (5.11) συνεπάγεται ότι $I(\theta) \to \infty$ καθώς $n \to \infty$, δηλαδή «άπειρες» παρατηρήσεις δημιουργούν «άπειρη» πληροφορία, κάτι που είναι διαισθητικά αναμενόμενο.

Κλείνοντας την ενότητα, αναφέρουμε ότι έχουν οριστεί και μελετηθεί από πολλούς ερευνητές αρκετά άλλα μέτρα στατιστικής πληροφορίας, εκτός του αριθμού πληροφορίας του Fisher. Κοινός τόπος αυτών των μέτρων είναι ότι ικανοποιούν ορισμένες βασικές ιδιότητες, όπως, π.χ. η προσθετική ιδιότητα. Η περιοχή της Στατιστικής που ασχολείται με τη θεωρία και τις εφαρμογές των μέτρων στατιστικής πληροφορίας είναι γνωστή με την ονομασία Στατιστική Θεωρία Πληροφοριών (Statistical Information Theory), (βλέπε για παράδειγμα Kullback (1997)).

5.2 Εκθετική οικογένεια κατανομών και αποδοτικοί εκτιμητές

Στην ενότητα αυτή θα μελετήσουμε μία τεχνική για την εύρεση αμερόληπτου εκτιμητή ελάχιστης διασποράς, που βασίζεται στην ανισότητα των Cramér–Rao. Ορίζουμε πρώτα την μονοπαραμετρική εκθετική οικογένεια κατανομών.

Ορισμός 5.2.1. *Η οικογένεια κατανομών του* $\underset{\sim}{X} = (X_1, \ldots, X_n)$, $\{f(\underset{\sim}{x}; \theta):$

$\theta \in \Theta\}$, $\Theta \subset \mathbb{R}$, *ανήκει στη Μονοπαραμετρική Εκθετική Οικογένεια Κατανομών (Μ.Ε.Ο.Κ.) εάν:*

(α) *Το σύνολο τιμών του* $\underset{\sim}{X}$, $\mathcal{S} = \big\{\underset{\sim}{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n \colon f(\underset{\sim}{x};\theta) > 0\big\}$ *δεν εξαρτάται από το* θ.

(β) *Η πυκνότητα* $f(\underset{\sim}{x};\theta)$ *έχει τη μορφή*

$$f(\underset{\sim}{x};\theta) = e^{A(\theta)+B(\underset{\sim}{x})+C(\theta)D(\underset{\sim}{x})}, \qquad \underset{\sim}{x} \in \mathcal{S},\ \theta \in \Theta.$$

Αντί της ορολογίας η οικογένεια $\big\{f(\underset{\sim}{x};\theta)\colon \theta \in \Theta\big\}$ ανήκει στην Μ.Ε.Ο.Κ., θα χρησιμοποιούμε, αδιακρίτως, και την έκφραση η οικογένεια $\{f(\underset{\sim}{x};\theta) : \theta \in \Theta\}$ είναι μία Μ.Ε.Ο.Κ.. Η ορολογία «μονοπαραμετρική» δικαιολογείται από ότι το θ είναι πραγματική παράμετρος (και όχι π.χ. ζεύγος πραγματικών παραμέτρων, $\theta = (\theta_1, \theta_2)$). Σημειώνουμε επίσης ότι οι συναρτήσεις $A(\theta)$, $B(\underset{\sim}{x})$, $C(\theta)$ και $D(\underset{\sim}{x})$ δεν ορίζονται μονοσήμαντα, αφού για παράδειγμα $f(\underset{\sim}{x};\theta) = e^{A_1(\theta)+B_1(\underset{\sim}{x})+C_1(\theta)D_1(\underset{\sim}{x})}$ με $A_1(\theta) = A(\theta) - 1$, $B_1(\underset{\sim}{x}) = B(\underset{\sim}{x}) + 1$, $C_1(\theta) = C(\theta)/2$, $D_1(\underset{\sim}{x}) = 2D(\underset{\sim}{x})$. Από τον ορισμό προκύπτει αμέσως η επόμενη πρόταση.

Πρόταση 5.2.1. *Αν* $\underset{\sim}{X} = (X_1, \ldots, X_n)$ *είναι ένα τυχαίο δείγμα και η κοινή οικογένεια κατανομών των* X_i *είναι μία Μ.Ε.Ο.Κ. με πυκνότητα*

$$f_1(x;\theta) = e^{A(\theta)+B(x)+C(\theta)D(x)}, \quad x \in \mathcal{S}_1, \theta \in \Theta,$$

τότε η οικογένεια κατανομών του $\underset{\sim}{X} = (X_1, \ldots, X_n)$, $\big\{f(\underset{\sim}{x};\theta)\colon \theta \in \Theta\big\}$, *είναι επίσης μία Μ.Ε.Ο.Κ., διατηρώντας μάλιστα την ίδια συνάρτηση* $C(\theta)$, $\theta \in \Theta$.

Απόδειξη. Έχουμε,

$$\begin{aligned} f(\underset{\sim}{x};\theta) = f(x_1, \ldots, x_n;\theta) &= \prod_{i=1}^{n} f_1(x_i;\theta) = e^{nA(\theta)+\sum\limits_{i=1}^{n} B(x_i)+C(\theta)\sum\limits_{i=1}^{n} D(x_i)} \\ &= e^{A^*(\theta)+B^*(\underset{\sim}{x})+C(\theta)D^*(\underset{\sim}{x})}, \quad \underset{\sim}{x} = (x_1, \ldots, x_n) \in \mathcal{S} = \mathcal{S}_1 \times \cdots \times \mathcal{S}_1, \end{aligned}$$

όπου $A^*(\theta) = nA(\theta)$, $B^*(\underset{\sim}{x}) = \sum\limits_{i=1}^{n} B(x_i)$, $D^*(\underset{\sim}{x}) = \sum\limits_{i=1}^{n} D(x_i)$. □

Μερικά παραδείγματα Μ.Ε.Ο.Κ. είναι τα ακόλουθα. Σε όλα θεωρούμε ότι το $\underset{\sim}{X} = (X_1, \ldots, X_n)$ είναι ένα τυχαίο δείγμα από μία κατανομή με πυκνότητα $f_1(x;\theta)$, οπότε η απόδειξη για την οικογένεια πυκνοτήτων $\{f(\underset{\sim}{x};\theta) : \theta \in \Theta\}$ ανάγεται στην αντίστοιχη για την οικογένεια $\{f_1(x;\theta) : \theta \in \Theta\}$.

Παράδειγμα 5.2.1. Η οικογένεια των κανονικών κατανομών $\mathcal{N}(\theta, 1)$, $\theta \in \Theta = \mathbb{R}$ είναι μία Μ.Ε.Ο.Κ. γιατί $\mathcal{S}_1 = \mathbb{R}$ και

$$\begin{aligned} f_1(x;\theta) &= \tfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta)^2} = e^{-\frac{1}{2}\theta^2 - \frac{1}{2}\ln(2\pi) - \frac{1}{2}x^2 + \theta x}, \\ &= e^{A(\theta)+B(x)+C(\theta)D(x)}, \quad x \in \mathbb{R},\ \theta \in \mathbb{R}, \end{aligned}$$

με $A(\theta) = -\frac{1}{2}\theta^2 - \frac{1}{2}\ln(2\pi)$, $B(x) = -\frac{1}{2}x^2$, $c(\theta) = \theta$, $D(x) = x$.

Παράδειγμα 5.2.2. Η οικογένεια των κανονικών κατανομών $\mathcal{N}(0, \theta^2)$, $\theta \in \Theta = (0, \infty)$ είναι μία Μ.Ε.Ο.Κ. γιατί $\mathcal{S}_1 = \mathbb{R}$ και

$$\begin{aligned} f_1(x;\theta) &= \tfrac{1}{\theta\sqrt{2\pi}} e^{-\frac{x^2}{2\theta^2}} = e^{-\ln\theta - \frac{1}{2}\ln(2\pi) - \frac{1}{2\theta^2}x^2}, \\ &= e^{A(\theta)+B(x)+C(\theta)D(x)}, \quad x \in \mathbb{R},\ \theta > 0, \end{aligned}$$

με $A(\theta) = -\ln\theta - \frac{1}{2}\ln(2\pi)$, $B(x) = 0$, $c(\theta) = -\frac{1}{2\theta^2}$, $D(x) = x^2$.

Παράδειγμα 5.2.3. Η οικογένεια των κατανομών Poisson, $\mathcal{P}(\theta)$, $\theta \in \Theta = (0, \infty)$ είναι μία Μ.Ε.Ο.Κ. γιατί $\mathcal{S}_1 = \{0, 1, 2, \ldots\}$ και

$$\begin{aligned} f_1(x;\theta) &= e^{-\theta}\tfrac{\theta^x}{x!} = e^{-\theta - \ln x! + (\ln\theta)\cdot x}, \\ &= e^{A(\theta)+B(x)+C(\theta)D(x)}, \quad x \in \mathcal{S}_1,\ \theta > 0, \end{aligned}$$

με $A(\theta) = -\theta$, $B(x) = -\ln x!$, $c(\theta) = \ln\theta$, $D(x) = x$.

Παράδειγμα 5.2.4. Η οικογένεια των διωνυμικών κατανομών $\mathcal{B}(n, \theta)$, $\theta \in \Theta = (0, 1)$ είναι μία Μ.Ε.Ο.Κ. γιατί $\mathcal{S}_1 = \{0, 1, \ldots, n\}$ και

$$\begin{aligned} f_1(x;\theta) &= \tbinom{n}{x}\theta^x(1-\theta)^{n-x} = \tbinom{n}{x}\left(\tfrac{\theta}{1-\theta}\right)^x (1-\theta)^n \\ &= e^{n\ln(1-\theta) + \ln\binom{n}{x} + \ln\left(\frac{\theta}{1-\theta}\right)\cdot x}, \\ &= e^{A(\theta)+B(x)+C(\theta)D(x)}, \quad x \in \mathcal{S}_1,\ \theta \in (0, 1), \end{aligned}$$

με $A(\theta) = n\ln(1-\theta)$, $B(x) = \ln\binom{n}{x}$, $c(\theta) = \ln\left(\frac{\theta}{1-\theta}\right)$, $D(x) = x$.

Παράδειγμα 5.2.5. Η οικογένεια των ομοιόμορφων κατανομών $\mathcal{U}(0,\theta)$, $\theta \in \Theta = (0,\infty)$ δεν είναι μία Μ.Ε.Ο.Κ. γιατί

$$f_1(x;\theta) = \begin{cases} \frac{1}{\theta}, & 0 < x < \theta, \\ 0, & \text{αλλού}, \end{cases}$$

οπότε $\mathcal{S}_1 = (0,\theta)$ το οποίο εξαρτάται από το θ.

Παράδειγμα 5.2.6. Η οικογένεια των κατανομών Cauchy με πυκνότητα

$$f_1(x;\theta) = \frac{1}{\pi\big(1+(x-\theta)^2\big)}, \quad -\infty < x < \infty,\ \theta \in \Theta = \mathbb{R},$$

δεν είναι μία Μ.Ε.Ο.Κ., γιατί δεν μπορεί να γραφεί στη μορφή $f_1(x;\theta) = e^{A(\theta)+B(x)+C(\theta)D(x)}$.

Η επόμενη πρόταση καθιστά πολύ εύκολη την επαλήθευση των συνθηκών Ι2, Ι3 και Ι4 για μία Μ.Ε.Ο.Κ.

Πρόταση 5.2.2. *Εάν η οικογένεια κατανομών του* $\underset{\sim}{X}$, $\{f(\underset{\sim}{x};\theta)\colon \theta \in \Theta \subset \mathbb{R}\}$ *με* Θ *ανοικτό σύνολο, είναι μία Μ.Ε.Ο.Κ. με* $f(\underset{\sim}{x};\theta) = e^{A(\theta)+B(\underset{\sim}{x})+C(\theta)D(\underset{\sim}{x})}$, $\underset{\sim}{x} \in \mathcal{S}$ *και η συνάρτηση* $C(\theta)$ *έχει συνεχή και μη μηδενική παράγωγο στο* Θ, *τότε οι συνθήκες Ι2, Ι3 και Ι4 ισχύουν και η Ι4 ισχύει για κάθε στατιστική συνάρτηση* $T(\underset{\sim}{X})$.

Απόδειξη. Παραλείπεται, επειδή υπερβαίνει το επίπεδο αυτών των σημειώσεων. Παραπέμπουμε, όμως, τον αναγνώστη στους Bickel and Doksum (1977, σελ. 130 και 147 – 148) που παρέχουν υποδείξεις για την απόδειξη. □

Στην περίπτωση που τα δεδομένα $\underset{\sim}{X} = (X_1,\ldots,X_n)$ αποτελούν ένα τυχαίο δείγμα από μία Μ.Ε.Ο.Κ., αντί της Πρότασης 5.2.2, μπορεί να χρησιμοποιηθεί και η επόμενη πρόταση για την επαλήθευση των Ι2, Ι3 και Ι4.

Πρόταση 5.2.3. *Εάν* $\underset{\sim}{X} = (X_1,\ldots,X_n)$ *είναι ένα τυχαίο δείγμα από μία Μ.Ε.Ο.Κ.*

$$f_1(x_1;\theta) = e^{A(\theta)+B(x_1)+C(\theta)D(x_1)}, \quad x_1 \in \mathcal{S}_1,\ \theta \in \Theta \subset \mathbb{R}$$

με Θ ανοικτό σύνολο και η συνάρτηση $C(\theta)$ έχει συνεχή και μη μηδενική παράγωγο στο Θ, τότε ισχύουν οι συνθήκες I2, I3 και I4 (για κάθε στατιστική συνάρτηση $T(\underset{\sim}{X})$) για την οικογένεια κατανομών $\{f(\underset{\sim}{x};\theta)\colon \theta \in \Theta\}$ του $\underset{\sim}{X}$.

Απόδειξη. Άμεση από τις Προτάσεις 5.2.1 και 5.2.2. □

Παράδειγμα 5.2.7. Για την Μ.Ε.Ο.Κ., που δόθηκε στο Παράδειγμα 5.2.2, έχουμε $\Theta = (0,\infty)$, $C(\theta) = -\frac{1}{2\theta^2}$, $C'(\theta) = \frac{1}{\theta^3}$. Συνεπώς, βάσει της Πρότασης 5.2.3 οι συνθήκες I2, I3 και I4 ισχύουν. □

Η ανισότητα (5.2) των Cramér–Rao μπορεί να χρησιμοποιηθεί για να αποδειχθεί ότι ένας αμερόληπτος εκτιμητής έχει ελάχιστη διασπορά ως εξής. Εάν ισχύουν οι συνθήκες I1–I5 και η I4 ισχύει για οποιοδήποτε αμερόληπτο εκτιμητή του $g(\theta)$, τότε από την (5.2) και την (5.6) παίρνουμε

$$\mathbb{V}ar_\theta T(\underset{\sim}{X}) \geqslant \frac{(g'(\theta))^2}{I(\theta)} = \text{Κ.Φ. C-R}$$

για κάθε $\theta \in \Theta$ και για κάθε αμερόληπτο εκτιμητή $T(\underset{\sim}{X})$. Εάν εν συνεχεία επιτύχουμε να βρούμε έναν αμερόληπτο εκτιμητή $T^*(\underset{\sim}{X})$ του οποίου η διασπορά είναι ίση προς το κάτω φράγμα $\big(g'(\theta)\big)^2/I(\theta)$, δηλαδή $\mathbb{V}ar_\theta T^*(\underset{\sim}{X}) = \text{Κ.Φ. C-R}$ για κάθε $\theta \in \Theta$, τότε ισχύει $\mathbb{V}ar_\theta T(\underset{\sim}{X}) \geqslant \mathbb{V}ar_\theta T^*(\underset{\sim}{X})$, για κάθε $\theta \in \Theta$, δηλαδή ο $T^*(\underset{\sim}{X})$ έχει την ελάχιστη δυνατή διασπορά και άρα το ελάχιστο δυνατό ΜΤΣ μεταξύ όλων των αμερόληπτων εκτιμητών. Για αυτόν τον εκτιμητή δίνουμε τον εξής ορισμό.

Ορισμός 5.2.2. *Ο εκτιμητής $T^*(\underset{\sim}{X})$ του $g(\theta)$ ονομάζεται αποδοτικός, εάν ισχύουν οι συνθήκες I1-I5 και*

(α) είναι αμερόληπτος,

(β) $\mathbb{V}ar_\theta T^*(\underset{\sim}{X}) = \text{Κ.Φ. C-R} = \big(g'(\theta)\big)^2/I(\theta), \quad \forall\, \theta \in \Theta$.

Παράδειγμα 5.2.8. (εκθετική κατανομή - αποδοτικός εκτιμητής της μέσης τιμής) Έστω $\underset{\sim}{X} = (X_1,\ldots,X_n)$ ένα τυχαίο δείγμα από την εκθετική κατανομή

$$f_1(x_1;\theta) = \tfrac{1}{\theta}e^{-\frac{x_1}{\theta}},\ \ x_1 > 0,\ \theta \in \Theta = (0,\infty).$$

Υπενθυμίζουμε ότι στο (εισαγωγικό) Παράδειγμα 2.1 είχαμε χρησιμοποιήσει την εκθετική κατανομή ως μοντέλο του χρόνου ζωής ηλεκτρικών λαμπτήρων και είχαμε αναδείξει τον δειγματικό μέσο $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ ως ένα λογικό εκτιμητή του θ. Θα αναζητήσουμε αποδοτικό εκτιμητή του θ, δηλαδή εδώ $g(\theta) = \theta$. Υπενθυμίζουμε ακόμη ότι $\mathbb{E}_\theta X_1 = \theta$ και $\mathbb{V}ar_\theta X_1 = \theta^2$. Επαληθεύουμε πρώτα τις συνθήκες I1–I5. Επειδή $\Theta = (0, \infty)$ η I1 προφανώς ισχύει. Για τις I2, I3, I4 παρατηρούμε ότι η οικογένεια $\{f_1(x_1;\theta) : \theta \in \Theta\}$ είναι μία Μ.Ε.Ο.Κ. με $C(\theta) = -\frac{1}{\theta}$ και παράγωγο $C'(\theta) = \frac{1}{\theta^2} \neq 0$ και συνεχή για κάθε $\theta \in \Theta$. Άρα, οι I2, I3, I4 ισχύουν (και η I4 είναι αληθής για κάθε στατιστική συνάρτηση $T(\underset{\sim}{X})$) βάσει της Πρότασης 5.2.3. Για την I5 παρατηρούμε ότι

$$\frac{\partial}{\partial\theta} \ln f_1(x_1;\theta) = -\frac{1}{\theta} + \frac{x_1}{\theta^2}.$$

Συνεπώς, από τις (5.12) και (5.11),

$$\begin{aligned} I(\theta) &= nI_1(\theta) = n\mathbb{V}ar_\theta\left(\tfrac{\partial}{\partial\theta}\ln f_1(X_1;\theta)\right) = n\mathbb{V}ar_\theta\left(-\tfrac{1}{\theta} + \tfrac{X_1}{\theta^2}\right) \\ &= \frac{n}{\theta^4}\mathbb{V}ar_\theta\,(X_1) = \frac{n}{\theta^4}\theta^2 = \frac{n}{\theta^2}, \end{aligned}$$

οπότε η I5 προφανώς ισχύει. Αφού λοιπόν ισχύουν οι συνθήκες I1–I5 και $g(\theta) = \theta$, από την (5.2) έχουμε

$$\mathbb{V}ar_\theta T(\underset{\sim}{X}) \geq \frac{1}{I(\theta)} = \frac{\theta^2}{n}, \quad \forall\, \theta \in \Theta$$

για κάθε αμερόληπτο εκτιμητή $T(\underset{\sim}{X})$ του θ, ενώ από την (5.6) το κάτω φράγμα των Cramér-Rao είναι Κ.Φ. C-R $= \frac{1}{I(\theta)} = \frac{\theta^2}{n}$. Παίρνοντας $T^*(\underset{\sim}{X}) = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$ συμπεραίνουμε ότι ο $T^*(\underset{\sim}{X}) = \bar{X}$ είναι αποδοτικός εκτιμητής του θ, αφού ο $\bar{X}$ είναι αμερόληπτος με $\mathbb{V}ar_\theta(\bar{X}) = \frac{\theta^2}{n} =$ Κ.Φ. C-R (βλέπε Πρόταση 4.2.3).

Οι επόμενες δύο προτάσεις χαρακτηρίζουν τις οικογένειες κατανομών $\{f(\underset{\sim}{x};\theta) : \theta \in \Theta \subset \mathbb{R}\}$ και εκείνες τις συναρτήσεις g για τις οποίες υπάρχουν αποδοτικοί εκτιμητές του $g(\theta)$, παρέχοντας συγχρόνως τους εκτιμητές αυτούς.

Πρόταση 5.2.4. *(Ικανή συνθήκη για αποδοτικότητα) Έστω ότι η οικογένεια κατανομών του* $\underset{\sim}{X}$, $\{f(\underset{\sim}{x};\theta) : \theta \in \Theta\}$, *ανήκει στη Μ.Ε.Ο.Κ., δηλαδή*

$$f(\underset{\sim}{x};\theta) = e^{A(\theta)+B(\underset{\sim}{x})+C(\theta)\cdot D(\underset{\sim}{x})}, \quad \underset{\sim}{x} \in \mathcal{S},$$

και ικανοποιούνται οι συνθήκες:

(α) *Το* Θ *είναι ένα ανοικτό σύνολο του* $\mathbb{R}$.

(β) *Η παράγωγος* $C'(\theta)$ *υπάρχει, είναι συνεχής και μη μηδενική, για κάθε* $\theta \in \Theta$.

(γ) $0 < I(\theta) < \infty, \forall\, \theta \in \Theta$.

Θέτουμε $g(\theta) = \mathbb{E}_\theta D(\underset{\sim}{X})$. *Τότε ισχύουν τα εξής:*

(i) $D(\underset{\sim}{X})$ *είναι αποδοτικός εκτιμητής του* $g(\theta)$.

(ii) $\alpha D(\underset{\sim}{X}) + \beta$ *είναι αποδοτικός εκτιμητής του* $\alpha g(\theta) + \beta$, *όπου* $\alpha \neq 0$ *και* β *αυθαίρετες σταθερές (μη εξαρτώμενες από το* θ*).*

Απόδειξη. (i) Από τις συνθήκες και την Πρόταση 5.2.2 προκύπτει ότι ικανοποιούνται οι I1–I5 της ανισότητας των Cramér–Rao και επομένως

$$\mathbb{V}ar_\theta T(\underset{\sim}{X}) \geq \frac{\big(g'(\theta)\big)^2}{I(\theta)}, \quad \forall\, \theta \in \Theta,$$

για κάθε αμερόληπτο εκτιμητή $T(\underset{\sim}{X})$ του $g(\theta)$. Από τον Ορισμό 5.2.2, ο $D(\underset{\sim}{X})$ θα είναι αποδοτικός εκτιμητής εάν

$$\mathbb{E}_\theta D(\underset{\sim}{X}) = g(\theta), \quad \forall\, \theta \in \Theta, \tag{5.15}$$

$$\mathbb{V}ar_\theta D(\underset{\sim}{X}) = \frac{\big(g'(\theta)\big)^2}{I(\theta)}, \quad \forall\, \theta \in \Theta. \tag{5.16}$$

Η (5.15) είναι ο ορισμός του $g(\theta)$. Για την (5.16) παρατηρούμε τα εξής. Κατ' αρχάς

$$\ln f(\underset{\sim}{x};\theta) = A(\theta) + B(\underset{\sim}{x}) + C(\theta)\cdot D(\underset{\sim}{x}), \quad \forall\, \underset{\sim}{x} \in \mathcal{S},\ \forall\, \theta \in \Theta$$

και

$$\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{x};\theta) = A'(\theta) + C'(\theta)\cdot D(\underset{\sim}{x}), \quad \forall\, \underset{\sim}{x}\in\mathcal{S},\ \forall\,\theta\in\Theta.$$

Η τελευταία σχέση δηλώνει ότι οι τυχαίες μεταβλητές $\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta)$ και $D(\underset{\sim}{X})$ συνδέονται γραμμικά μεταξύ τους (με πιθανότητα 1, για κάθε $\theta\in\Theta$) και επομένως η ανισότητα Cauchy–Schwarz ισχύει ως ισότητα (βλέπε Πρόταση 1.8.2). Συνεπώς έχουμε,

$$\mathbb{C}ov_\theta^2\left(\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta), D(\underset{\sim}{X})\right) = \mathbb{V}ar_\theta\left(\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta)\right)\cdot \mathbb{V}ar_\theta D(\underset{\sim}{X}), \quad \forall\,\theta\in\Theta. \quad (5.17)$$

Όμως, από την (5.5) με $T(\underset{\sim}{X}) = D(\underset{\sim}{X})$ και την (5.15) συμπεραίνουμε ότι

$$\mathbb{C}ov_\theta\left(\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta), D(\underset{\sim}{X})\right) = g'(\theta), \quad \forall\,\theta\in\Theta,$$

ενώ από την (5.8)

$$\mathbb{V}ar_\theta\left(\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta)\right) = I(\theta), \quad \forall\,\theta\in\Theta.$$

Αντικατάσταση στην (5.17) δίνει $\big(g'(\theta)\big)^2 = I(\theta)\cdot\mathbb{V}ar_\theta D(\underset{\sim}{X})$, για κάθε $\theta\in\Theta$, που είναι η (5.16).

(ii) Θα ανάγουμε την απόδειξη στο πρώτο μέρος που έχει ήδη δειχθεί. Πράγματι, θέτοντας $D_1(\underset{\sim}{X}) = \alpha D(\underset{\sim}{X}) + \beta$, λύνοντας ως προς $D(\underset{\sim}{X})$ και αντικαθιστώντας στον τύπο της $f(\underset{\sim}{x};\theta)$ παίρνουμε

$$f(\underset{\sim}{x};\theta) = e^{A_1(\theta)+B(\underset{\sim}{x})+C_1(\theta)\cdot D_1(\underset{\sim}{x})},$$

όπου, $A_1(\theta) = A(\theta) - \frac{\beta}{\alpha}\,C(\theta)$, $C_1(\theta) = \frac{C(\theta)}{\alpha}$. Προφανώς, για αυτήν τη νέα έκφραση της πυκνότητας $f(\underset{\sim}{x};\theta)$ ικανοποιούνται οι συνθήκες (α), (β) και (γ) του πρώτου μέρους και επομένως ο $D_1(\underset{\sim}{X})$ είναι αποδοτικός εκτιμητής του $g_1(\theta) = \mathbb{E}_\theta D_1(\underset{\sim}{X}) = \alpha\mathbb{E}_\theta D(\underset{\sim}{X}) + \beta$. □

Η επόμενη πρόταση σε συνδυασμό με την Παρατήρηση 5.2.1 που α-κολουθεί είναι, ουσιαστικά, η αντίστροφη της Πρότασης 5.2.4.

Πρόταση 5.2.5. *(Αναγκαία συνθήκη για αποδοτικότητα) Έστω ότι ισχύουν οι συνθήκες I1-I5, $g(\theta)$ δεν είναι σταθερά ως συνάρτηση του θ και $T(\underset{\sim}{X})$ είναι αποδοτικός εκτιμητής του $g(\theta)$. Τότε, η οικογένεια κατανομών του $\underset{\sim}{X}$, $\{f(\underset{\sim}{x};\theta) : \theta \in \Theta\}$, με πιθανότητα 1, ικανοποιεί τη σχέση*

$$\frac{\partial}{\partial\theta} \ln f(\underset{\sim}{X};\theta) = c_1(\theta)T(\underset{\sim}{X}) + c_2(\theta) \ , \quad \forall\, \theta \in \Theta \ , \tag{5.18}$$

όπου $c_1(\theta)$ και $c_2(\theta)$ είναι συναρτήσεις με $c_1(\theta) \neq 0$ για κάθε $\theta \in \Theta$ ή ισοδύναμα

$$\frac{\partial}{\partial\theta} \ln f(\underset{\sim}{X};\theta) = c_1(\theta)(T(\underset{\sim}{X}) - g(\theta)) \ , \quad \forall\, \theta \in \Theta \ . \tag{5.19}$$

Εάν, επιπλέον, $g_1(\theta)$ δεν είναι σταθερά ως προς θ και $T_1(\underset{\sim}{X})$ είναι αποδοτικός εκτιμητής του $g_1(\theta)$, τότε υπάρχουν σταθερές $a \neq 0$ και β μη εξαρτώμενες από το θ, τέτοιες ώστε

$$T_1(\underset{\sim}{X}) = aT(\underset{\sim}{X}) + \beta \ \text{ και } \ g_1(\theta) = ag(\theta) + \beta.$$

Απόδειξη. Αποδοτικότητα του $T(\underset{\sim}{X})$ σημαίνει

$$\mathbb{E}_\theta T(\underset{\sim}{X}) = g(\theta), \quad \forall\, \theta \in \Theta, \tag{5.20}$$

$$\mathbb{V}ar_\theta T(\underset{\sim}{X}) = \frac{\big(g'(\theta)\big)^2}{I(\theta)}, \quad \forall\, \theta \in \Theta. \tag{5.21}$$

Όμως, από την (5.8) έχουμε

$$\mathbb{V}ar_\theta \left(\frac{\partial}{\partial\theta} \ln f(\underset{\sim}{X};\theta)\right) = I(\theta), \quad \forall\, \theta \in \Theta, \tag{5.22}$$

ενώ από τις (5.5) και (5.20)

$$\mathbb{C}ov_\theta \left(\frac{\partial}{\partial\theta} \ln f(\underset{\sim}{X};\theta), T(\underset{\sim}{X})\right) = g'(\theta), \quad \forall\, \theta \in \Theta. \tag{5.23}$$

Άρα η (5.21) γράφεται

$$\begin{aligned} &\mathbb{C}ov_\theta^2 \left(\frac{\partial}{\partial\theta} \ln f(\underset{\sim}{X};\theta), T(\underset{\sim}{X})\right) \\ &\qquad = \mathbb{V}ar_\theta \left(\frac{\partial}{\partial\theta} \ln f(\underset{\sim}{X};\theta)\right) \cdot \mathbb{V}ar_\theta T(\underset{\sim}{X}), \quad \forall\, \theta \in \Theta. \end{aligned} \tag{5.24}$$

Η τελευταία σχέση δηλώνει ότι η ανισότητα Cauchy–Schwarz για τις τυχαίες μεταβλητές $\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta)$ και $T(\underset{\sim}{X})$ ισχύει ως ισότητα (βλέπε Πρόταση 1.8.2). Επιπλέον, καμία από τις $\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta)$ και $T(\underset{\sim}{X})$ δεν είναι σταθερά (ως συνάρτηση του $\underset{\sim}{X}$), η μεν πρώτη επειδή $\mathbb{V}ar_\theta\left(\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta)\right) > 0$ λόγω της (5.22) και της Ι5, ενώ, αν $T(\underset{\sim}{X}) = c$, σταθερά μη εξαρτώμενη από το θ, λόγω αμεροληψίας $\mathbb{E}_\theta T(\underset{\sim}{X}) = g(\theta)$ ή $c = g(\theta)$ για κάθε $\theta \in \Theta$ που αποκλείεται εξ υποθέσεως. Επομένως, η Πρόταση 1.8.2 συνεπάγεται ότι $\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta)$ και $T(\underset{\sim}{X})$ είναι γραμμικά εξαρτημένες, δηλαδή για κάθε $\theta \in \Theta$ υπάρχουν σταθερές $c_1(\theta) \neq 0$ και $c_2(\theta)$ εξαρτώμενες εν γένει από το θ, έτσι ώστε να ισχύει η σχέση

$$\mathbb{P}_\theta\left(\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta) = c_1(\theta)\,T(\underset{\sim}{X}) + c_2(\theta)\right) = 1. \qquad (5.25)$$

Η (5.6) σημαίνει ότι, με πιθανότητα 1, ισχύει η (5.18). Επιπλέον, από τις (5.3) και (5.18) προκύπτει ότι

$$0 = \mathbb{E}_\theta\left(\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta)\right) = \mathbb{E}_\theta\left[c_1(\theta)\,T(\underset{\sim}{X}) + c_2(\theta)\right] = c_1(\theta)\cdot g(\theta) + c_2(\theta),$$

δηλαδή

$$g(\theta) = -\frac{c_2(\theta)}{c_1(\theta)}\ ,$$

οπότε αντικαθιστώντας $c_2(\theta) = -c_1(\theta)g(\theta)$ στην (5.18) προκύπτει η (5.19).

Αν τώρα $T_1(\underset{\sim}{X})$ είναι ένας αποδοτικός εκτιμητής του $g_1(\theta)$, τότε αντίστοιχα με την (5.18), με πιθανότητα 1, ισχύει

$$\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta) = c_3(\theta)T_1(\underset{\sim}{X}) + c_4(\theta)\ . \qquad (5.26)$$

Από τις (5.18) και (5.26) συμπεραίνουμε ότι (με πιθανότητα 1)

$$c_1(\theta)T(\underset{\sim}{X}) + c_2(\theta) = c_3(\theta)T_1(\underset{\sim}{X}) + c_4(\theta),$$

δηλαδή

$$T_1(\underset{\sim}{X}) = \tfrac{c_1(\theta)}{c_3(\theta)}T(\underset{\sim}{X}) + \tfrac{c_2(\theta)-c_4(\theta)}{c_3(\theta)}.$$

Επειδή, όμως, $T_1(\underset{\sim}{X})$ είναι στατιστική συνάρτηση, ο συντελεστής του $T(\underset{\sim}{X})$ και ο σταθερός όρος δεν πρέπει να εξαρτώνται από το θ, άρα $T_1(\underset{\sim}{X}) = aT(\underset{\sim}{X}) + \beta$, όπου $a \neq 0$ και β σταθερές. Τέλος, $g_1(\theta) = \mathbb{E}_\theta T_1(\underset{\sim}{X}) = a\mathbb{E}_\theta T(\underset{\sim}{X}) + \beta$. □

Παρατήρηση 5.2.1. Η (5.18) συνεπάγεται ότι η οικογένεια κατανομών $\{f(\underset{\sim}{x};\theta) : \theta \in \Theta\}$ ανήκει στη Μ.Ε.Ο.Κ.. Μία αυστηρή απόδειξη δίνεται από τους Wijsman (1973), Joshi (1976) και Müller-Funk, Pukelsheim and Witting (1989). Σε αδρές γραμμές η απόδειξη έχει ως εξής. Από την (5.18) ουσιαστικά για κάθε $\underset{\sim}{x}$ στο σύνολο τιμών του $\underset{\sim}{X}$ παίρνουμε $\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{x};\theta) = c_1(\theta)\,T(\underset{\sim}{x}) + c_2(\theta)$ και συνεπώς $\ln f(\underset{\sim}{x};\theta) = T(\underset{\sim}{x})\int c_1(\theta)d\theta + \int c_2(\theta)d\theta + B(\underset{\sim}{x})$. Θέτοντας $A(\theta) = \int c_1(\theta)d\theta$ και $C(\theta) = \int c_2(\theta)d\theta$ έχουμε

$$f(\underset{\sim}{x};\theta) = e^{A(\theta)+B(\underset{\sim}{x})+C(\theta)T(\underset{\sim}{x})}.$$

Αυτή, όμως είναι η μορφή της πυκνότητας μιας Μ.Ε.Ο.Κ., με $D(\underset{\sim}{X}) = T(\underset{\sim}{X})$, τον αποδοτικό εκτιμητή. Οι Πρότασεις 5.2.4 και 5.2.5 δηλώνουν λοιπόν ότι για την οικογένεια κατανομών $\{f(\underset{\sim}{x};\theta) : \theta \in \Theta \subset \mathbb{R}\}$ υπάρχουν αποδοτικοί εκτιμητές αν και μόνον αν αυτή ανήκει στη Μ.Ε.Ο.Κ. $f(\underset{\sim}{x};\theta) = e^{A(\theta)+B(\underset{\sim}{x})+C(\theta)\cdot D(\underset{\sim}{x})}$. Αποδοτικοί εκτιμητές είναι οι $D(\underset{\sim}{X})$ και $\alpha D(\underset{\sim}{X}) + \beta$ (Πρόταση 5.2.4) και μόνον αυτοί (Πρόταση 5.2.5). Αποδοτικά εκτιμούνται οι τιμές $g(\theta) = \mathbb{E}_\theta D(\underset{\sim}{X})$ και $g_1(\theta) = \alpha g(\theta) + \beta$ (Πρόταση 5.2.4) και μόνον αυτές (Πρόταση 5.2.5). Τέλος, υπό τις προϋποθέσεις της Πρότασης 5.2.5, ο αποδοτικός εκτιμητής του $g(\theta)$ είναι μοναδικός. Αν υπήρχαν δύο, ο $T(\underset{\sim}{X})$ και ο $\alpha T(\underset{\sim}{X})+\beta$ αποδοτικοί για το $g(\theta)$, θα είχαμε $g(\theta) = \alpha g(\theta)+\beta$, οπότε $\alpha = 1$ και $\beta = 0$ αφού $g(\theta)$ δεν είναι σταθερά ως προς θ.

Παρατήρηση 5.2.2. Ακολουθώντας την αντίστροφη πορεία της απόδειξης της Πρότασης 5.2.5, είναι εύκολο να δειχθεί ότι σε μία Μ.Ε.Ο.Κ. που ικανοποιεί τις I1–I5, η (5.19) είναι και ικανή συνθήκη προκειμένου ο $T(\underset{\sim}{X})$ να είναι αποδοτικός εκτιμητής του $g(\theta)$. Συνεπώς, φέρνοντας την τυχαία μεταβλητή $\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta)$ στη μορφή (5.19) συμπεραίνουμε αμέσως ότι ο $T(\underset{\sim}{X})$ είναι αποδοτικός για το $g(\theta)$, χωρίς να χρειαστεί να υπολογίσουμε τη μέση τιμή $\mathbb{E}_\theta T(\underset{\sim}{X})$, που αν την υπολογίσουμε θα προκύψει φυσικά ότι είναι $g(\theta)$.

Παράδειγμα 5.2.9. (κατανομή Bernoulli - αποδοτικοί εκτιμητές) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την κατανομή Bernoulli $\mathcal{B}(1,\theta)$ με πυκνότητα $f_1(x;\theta) = \theta^x(1-\theta)^{1-x}$, $x = 0, 1$, $\theta \in \Theta = (0,1)$.

Τότε, η οικογένεια κατανομών του $\underset{\sim}{X}$ είναι μία Μ.Ε.Ο.Κ. με πυκνότητα

$$
\begin{aligned}
f(\underset{\sim}{x};\theta) &= \prod_{i=1}^{n} f_1(x_i;\theta) = \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i} \\
&= e^{n\ln(1-\theta)+\ln\frac{\theta}{1-\theta}\cdot\sum\limits_{i=1}^{n} x_i}, \quad \underset{\sim}{x} = (x_1,\ldots,x_n) \in \mathcal{S} = \{0,1\}\times\cdots\times\{0,1\}.
\end{aligned}
$$

Οι συνθήκες Ι1 και Ι2 προφανώς ισχύουν. Οι Ι3 και Ι4 επίσης ικανοποιούνται επειδή το $\mathcal{S}$ είναι πεπερασμένο ή διαφορετικά επειδή ισχύει η συνθήκη (β) της Πρότασης 5.2.4 καθώς $c(\theta) = \ln(\theta/(1-\theta))$ και $c'(\theta) = 1/(\theta(1-\theta))$. Τέλος, από την (5.8), ο αριθμός πληροφορίας του Fisher είναι

$$
\begin{aligned}
I(\theta) &= \mathbb{V}ar_\theta\left(\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta)\right) = \frac{1}{\theta^2(1-\theta)^2}\mathbb{V}ar_\theta\left(\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{\theta^2(1-\theta)^2}\sum_{i=1}^{n}\mathbb{V}ar_\theta X_i = \frac{1}{\theta^2(1-\theta)^2}\sum_{i=1}^{n}\theta(1-\theta) = \frac{n}{\theta(1-\theta)},
\end{aligned}
$$

δηλαδή ισχύει η συνθήκη Ι5.

Επομένως, από την Πρόταση 5.2.4 η στατιστική συνάρτηση $D(\underset{\sim}{X}) = \sum\limits_{i=1}^{n} X_i$ είναι αποδοτικός εκτιμητής του $g(\theta) = \mathbb{E}_\theta D(\underset{\sim}{X}) = n\theta$ και η στατιστική συνάρτηση $aD(\underset{\sim}{X})+\beta$ είναι αποδοτικός εκτιμητής του $ag(\theta)+\beta = a(n\theta)+\beta$. Έτσι, με $a = \frac{1}{n}$, $\beta = 0$, ο εκτιμητής $\frac{D(\underset{\sim}{X})}{n} = \bar{X}$ είναι αποδοτικός για το θ. Ακόμη, παρατηρούμε ότι η διασπορά της κατανομής, $g_1(\theta) = \theta(1-\theta)$, δεν μπορεί να γραφεί στη μορφή $g_1(\theta) = ag(\theta)+\beta = a(n\theta)+\beta$, για κάθε $\theta \in (0,1)$, με a και β σταθερές μη εξαρτώμενες από το θ (το πρώτο μέλος είναι συνάρτηση δεύτερου βαθμού ως προς θ ενώ το δεύτερο πρώτου βαθμού) και επομένως δεν υπάρχει αποδοτικός εκτιμητής του $\theta(1-\theta)$. Εναλλακτικά, χρησιμοποιώντας τη διαδικασία που περιγράφεται στην Παρατήρηση 5.2.2, είναι πολύ εύκολο να φέρουμε την τυχαία μεταβλητή $\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta)$ στις ισοδύναμες μορφές

$$\begin{aligned}\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta) &= \frac{1}{\theta(1-\theta)}\left(\sum_{i=1}^{n} X_i - n\theta\right) = \frac{n}{\theta(1-\theta)}(\bar{X}-\theta)\\ &= \frac{n}{a\theta(1-\theta)}(a\bar{X}+\beta-(a\theta+\beta))\end{aligned}$$

απ' όπου προκύπτει ότι οι στατιστικές συναρτήσεις $\sum_{i=1}^{n} X_i$, $\bar{X}$, $a\bar{X}+\beta$ είναι αποδοτικοί εκτιμητές των $n\theta$, θ, $a\theta+\beta$, αντίστοιχα. Όλοι οι αποδοτικοί εκτιμητές είναι οι γραμμικοί μετασχηματισμοί $a\bar{X}+\beta$.

Παράδειγμα 5.2.10. (κατανομή Poisson - πληροφορία $\mathbf{I}(\theta)$) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την κατανομή Poisson $\mathcal{P}(\theta)$, $\theta \in \Theta = (0, \infty)$ με πυκνότητα

$$f_1(x;\theta) = e^{-\theta}\frac{\theta^x}{x!},\quad x = 0, 1, 2, \ldots\ .$$

Υπενθυμίζουμε ότι $\mathbb{E}_\theta X_1 = \mathbb{V}ar_\theta X_1 = \theta$.
Τότε

$$f(\underset{\sim}{x};\theta) = \prod_{i=1}^{n} f_1(x_i;\theta) = e^{-n\theta}\cdot\frac{\theta^{\sum_{i=1}^{n}x_i}}{x_1!\cdots x_n!}$$

και

$$\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{x};\theta) = -n + \frac{\sum_{i=1}^{n}x_i}{\theta}.$$

Επομένως, από την (5.8) προκύπτει

$$\begin{aligned}I(\theta) &= \mathbb{V}ar_\theta\left(\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta)\right) = \mathbb{V}ar_\theta\left(-n + \sum_{i=1}^{n} X_i/\theta\right) = \frac{1}{\theta^2}\mathbb{V}ar_\theta\left(\sum_{i=1}^{n} X_i\right)\\ &= \frac{1}{\theta^2}\sum_{i=1}^{n}\mathbb{V}ar_\theta X_i = \frac{1}{\theta^2}\sum_{i=1}^{n}\theta = \frac{1}{\theta^2}n\theta = \frac{n}{\theta}.\end{aligned}$$

Παρατηρούμε ότι η πληροφορία $I(\theta)$ είναι αντιστρόφως ανάλογη της διασποράς, θ, της κατανομής. Όσο λοιπόν μικρότερη είναι η διασπορά, τόσο περισσότερη πληροφορία περιέχεται στο δείγμα $\underset{\sim}{X}$ για το θ, κάτι που είναι διαισθητικά αναμενόμενο.

Σύμφωμα με την Πρόταση 5.1.2, εφαρμογή της (5.8), $I(\theta) = \mathbb{V}ar_\theta\left(\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta)\right)$, προϋποθέτει την επαλήθευση των I1, I2, I3 που γίνεται εύκολα μέσω της Πρότασης 5.2.2. Εναλλακτικά, $\mathbb{E}_\theta\left(\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta)\right) = \mathbb{E}_\theta\left(-n + \sum_{i=1}^n X_i/\theta\right) = -n + \frac{1}{\theta}\sum_{i=1}^n \mathbb{E}_\theta X_i = -n + \frac{1}{\theta}(n\theta) = 0$, δηλαδή η (5.3) ή ισοδύναμα η I6 ισχύουν, οπότε έχουμε

$$I(\theta) = \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta)\right)^2\right] = \mathbb{V}ar_\theta\left(\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta)\right).$$

Παράδειγμα 5.2.11. (κανονική κατανομή με γνωστή διασπορά - πληροφορία $\mathbf{I}(\theta)$ και αποδοτικοί εκτιμητές) Έστω $\underset{\sim}{X} = (X_1,\ldots,X_n)$ ένα τυχαίο δείγμα από την κανονική κατανομή $\mathcal{N}(\theta,\sigma^2)$ με πυκνότητα

$$f_1(x;\theta) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2\sigma^2}(x-\theta)^2}, x\in\mathbb{R},$$

όπου σ^2 είναι γνωστό και $\theta\in\Theta=\mathbb{R}$ άγνωστο. Τότε

$$f(\underset{\sim}{x};\theta) = \frac{1}{\sigma^n(2\pi)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum\limits_{i=1}^n (x_i-\theta)^2} = \frac{1}{\sigma^n(2\pi)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum\limits_{i=1}^n x_i^2 + \frac{\theta}{\sigma^2}\sum\limits_{i=1}^n x_i - \frac{n\theta^2}{2\sigma^2}}$$

και

$$\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{x};\theta) = \frac{1}{\sigma^2}\left(\sum_{i=1}^n x_i - n\theta\right). \tag{5.27}$$

Οι συνθήκες της Πρότασης 5.2.2 επαληθεύονται πολύ εύκολα (βλέπε και το Παράδειγμα 5.2.1). Άρα, από την (5.8),

$$I(\theta) = \mathbb{V}ar_\theta\left(\frac{1}{\sigma^2}\sum_{i=1}^n X_i - n\theta\right) = \frac{1}{\sigma^4}\mathbb{V}ar_\theta\left(\sum_{i=1}^n X_i\right) = \frac{1}{\sigma^4}n\sigma^2 = \frac{n}{\sigma^2}.$$

Επίσης, μπορούμε να υπολογίσουμε το $I(\theta)$ από τη σχέση (5.12), $I(\theta) = nI_1(\theta)$. Πράγματι, έχουμε

$$f_1(x_1;\theta) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2\sigma^2}(x_1-\theta)^2}$$

και $\frac{\partial}{\partial\theta}\ln f_1(x_1;\theta) = \frac{1}{\sigma^2}(x_1-\theta)$. Επομένως, από την (5.11),

$$I_1(\theta) = \mathbb{E}_\theta\left[\frac{1}{\sigma^4}(X_1-\theta)^2\right] = \frac{1}{\sigma^4}\mathbb{E}_\theta(X_1-\theta)^2 = \frac{1}{\sigma^4}\mathbb{V}ar_\theta X_1 = \frac{1}{\sigma^4}\,\sigma^2 = \frac{1}{\sigma^2}$$

που συνεπάγεται
$$I(\theta) = nI_1(\theta) = \frac{n}{\sigma^2}.$$
Εναλλακτικά, χρησιμοποιώντας τη σχέση (5.9)
$$I(\theta) = -\mathbb{E}_\theta\Big(\frac{\partial^2}{\partial\theta^2}\ln f(\underset{\sim}{X};\theta)\Big)$$
και επειδή $\frac{\partial^2}{\partial\theta^2}\ln f(\underset{\sim}{X};\theta) = -\frac{n}{\sigma^2}$, προκύπτει ότι $I(\theta) = -\mathbb{E}_\theta(-\frac{n}{\sigma^2}) = \frac{n}{\sigma^2}$. Έχει ενδιαφέρον να επισημάνουμε ότι η πληροφορία $I(\theta)$ είναι σταθερή ως προς θ, αλλά και αντιστρόφως ανάλογη της διασποράς, σ^2, της κατανομής. Όπως λοιπόν και στο Παράδειγμα 5.2.10, όσο μικρότερη είναι η διασπορά τόσο μεγαλύτερη είναι η πληροφορία που παρέχει το δείγμα $\underset{\sim}{X}$ για την άγνωστη παράμετρο θ.
Λόγω της (5.27), από την Παρατήρηση 5.2.2, συμπεραίνουμε αμέσως ότι η στατιστική συνάρτηση $D(\underset{\sim}{X}) = \sum_{i=1}^{n} X_i$ είναι αποδοτικός εκτιμητής του $g(\theta) = n\theta$. Έχοντας βρει έναν αποδοτικό εκτιμητή, τον $D(\underset{\sim}{X})$, από την Παρατήρηση 5.2.1, αποδοτικοί εκτιμητές είναι οι γραμμικοί μετασχηματισμοί του, $T(\underset{\sim}{X}) = aD(\underset{\sim}{X}) + \beta = a\sum_{i=1}^{n} X_i + \beta$ και μόνον αυτοί. Οι αντίστοιχες τιμές που εκτιμούνται αποδοτικά είναι
$$g_1(\theta) = \mathbb{E}_\theta T(\underset{\sim}{X}) = \mathbb{E}_\theta(aD(\underset{\sim}{X} + \beta) = ag(\theta) + \beta = a(n\theta) + \beta,$$
δηλαδή οι γραμμικοί μετασχηματισμοί του θ και μόνον αυτοί. Ως ειδική περίπτωση, για $a = \frac{1}{n}$ και $\beta = 0$, ο $\bar{X}$ είναι αποδοτικός εκτιμητής του θ.

Παράδειγμα 5.2.12. (κανονική κατανομή με γνωστή μέση τιμή - πληροφορία $\mathbf{I}(\theta)$ και αποδοτικοί εκτιμητές) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ είναι ένα τυχαίο δείγμα από την κανονική κατανομή $\mathcal{N}(\mu, \theta^2)$, όπου μ είναι γνωστό και $\theta \in \Theta = (0, \infty)$ άγνωστο. Θα υπολογίσουμε τον αριθμό πληροφορίας του Fisher $I(\theta)$, το Κάτω Φράγμα των Cramér-Rao (Κ.Φ.C-R) για τη διασπορά των αμερόληπτων εκτιμητών του θ^2 καθώς και του θ και θα βρούμε όλους τους αποδοτικούς εκτιμητές και τις αντίστοιχες τιμές που εκτιμούν.
Επαληθεύουμε πρώτα τις συνθήκες I1–I5 του Θεωρήματος 5.1.1.
I1. Ο παραμετρικός χώρος $\Theta = (0, \infty)$ είναι ανοικτό διάστημα του $\mathbb{R}$.
I2 - I4. Θα χρησιμοποιήσουμε την Πρόταση 5.2.2. Αρχικά δείχνουμε ότι

η οικογένεια κατανομών του $\underset{\sim}{X}$ ανήκει στην Μ.Ε.Ο.Κ..
1. Το σύνολο $\mathcal{S} = \{\underset{\sim}{x} : f(\underset{\sim}{x}, \theta) > 0\} = \mathbb{R}^n$ δεν εξαρτάται από το θ.

2.
$$\begin{aligned} f(\underset{\sim}{x}, \theta) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\theta}\, e^{-\frac{(x_i-\mu)^2}{2\theta^2}} = \frac{1}{(2\pi)^{n/2}\theta^n}\, e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\theta^2}} \\ &= \exp\left\{-\frac{n}{2}\ln(2\pi) - n\ln\theta - \frac{1}{2\theta^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\} \\ &= e^{A(\theta)+B(\underset{\sim}{x})+C(\theta)D(\underset{\sim}{x})}, \end{aligned}$$

όπου,

$$A(\theta) = -\frac{n}{2}\ln(2\pi) - n\ln\theta,\ B(\underset{\sim}{x}) = 0,\ C(\theta) = -\frac{1}{2\theta^2},\ D(\underset{\sim}{x}) = \sum_{i=1}^{n}(x_i-\mu)^2$$

και επομένως η οικογένεια κατανομών του $\underset{\sim}{X}$ ανήκει στην Μ.Ε.Ο.Κ..
Επειδή η συνάρτηση $C(\theta) = -\frac{1}{2\theta^2}$ έχει συνεχή και μη μηδενική παράγωγο, $C'(\theta) = \frac{1}{\theta^3}$, για $\theta \in (0, \infty)$, οι συνθήκες Ι2–Ι4 ισχύουν, λόγω της Πρότασης 5.2.2.
Ι5. Από τις (5.12) και (5.13) παίρνουμε,

$$I(\theta) = nI_1(\theta), \quad \text{όπου} \quad I_1(\theta) = -\mathbb{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\ln f_1(X_1;\theta)\right).$$

Έχουμε λοιπόν,

$$\begin{aligned} f_1(x_1;\theta) &= \frac{1}{\sqrt{2\pi}\theta}\, e^{-\frac{(x_1-\mu)^2}{2\theta^2}}, \\ \ln f_1(x_1;\theta) &= -\frac{1}{2}\ln(2\pi) - \ln\theta - \frac{(x_1-\mu)^2}{2\theta^2}, \\ \frac{\partial}{\partial\theta}\ln f_1(x_1;\theta) &= -\frac{1}{\theta} + \frac{(x_1-\mu)^2}{\theta^3}, \\ \frac{\partial^2}{\partial\theta^2}\ln f_1(x_1;\theta) &= \frac{1}{\theta^2} - \frac{3(x_1-\mu)^2}{\theta^4}. \end{aligned}$$

Άρα,

$$\begin{aligned} I_1(\theta) &= -\mathbb{E}_\theta\left\{\frac{1}{\theta^2} - \frac{3(X_1-\mu)^2}{\theta^4}\right\} = -\frac{1}{\theta^2} + \frac{3}{\theta^4}\mathbb{E}_\theta(X_1-\mu)^2 \\ &= -\frac{1}{\theta^2} + \frac{3}{\theta^4}\mathbb{V}ar_\theta X_1 = -\frac{1}{\theta^2} + \frac{3}{\theta^4}\theta^2 = \frac{2}{\theta^2}, \end{aligned}$$

οπότε, $I(\theta) = nI_1(\theta) = \frac{2n}{\theta^2}$, δηλαδή ισχύει και η I5.
Επομένως, από το Θεώρημα 5.1.1, το κάτω φράγμα των Cramér–Rao για τη διασπορά των αμερόληπτων εκτιμητών του $g(\theta) = \theta^2$ ή $g(\theta) = \theta$ είναι αντίστοιχα,

$$\text{Κ.Φ. C-R} = \frac{(g'(\theta))^2}{I(\theta)} = \frac{4\theta^2}{2n/\theta^2} = \frac{2\theta^4}{n}\ , \quad \text{Κ.Φ. C-R} = \frac{(g'(\theta))^2}{I(\theta)} = \frac{1}{2n/\theta^2} = \frac{\theta^2}{2n}.$$

Εφαρμόζοντας την Πρόταση 5.2.4 σε συνδυασμό με την Παρατήρηση 5.2.1, συμπεραίνουμε ότι όλοι οι αποδοτικοί εκτιμητές είναι οι γραμμικοί μετασχηματισμοί του $D(\underset{\sim}{X})$, δηλαδή

$$T(\underset{\sim}{X}) = aD(\underset{\sim}{X}) + \beta = a\sum_{i=1}^{n}(X_i - \mu)^2 + \beta.$$

Οι αντίστοιχες τιμές που εκτιμούνται αποδοτικά είναι μόνον οι γραμμικοί μετασχηματισμοί του $g(\theta) = \mathbb{E}_\theta D(\underset{\sim}{X}) = \mathbb{E}_\theta \sum_{i=1}^{n}(X_i - \mu)^2 = n\theta^2$, δηλαδή $g_1(\theta) = a(n\theta^2) + \beta$. Ειδικά για $a = \frac{1}{n}$ και $\beta = 0$, $\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$ είναι αποδοτικός εκτιμητής της διασποράς θ^2, ενώ δεν υπάρχει αποδοτικός εκτιμητής της τυπικής απόκλισης θ.

Παράδειγμα 5.2.13. (Συνέχεια του Παραδείγματος 5.2.12 - αλλαγή παραμέτρου) Η κανονική κατανομή καθορίζεται πλήρως από τη μέση τιμή και τη διασπορά της. Δικαιολογείται λοιπόν (και από στατιστικής σκοπιάς) να θεωρήσουμε στο Παράδειγμα 5.2.12 ως άγνωστη παράμετρο τη διασπορά $\eta = \theta^2$ αντί της τυπικής απόκλισης θ. Τότε η κοινή κατανομή των X_i, $i = 1, \ldots, n$, είναι $\mathcal{N}(\mu, \eta)$ με πυκνότητα

$$f_1^*(x;\eta) = \frac{1}{\sqrt{2\pi\eta}}e^{-\frac{1}{2\eta}(x-\mu)^2}\ , x \in \mathbb{R}\ ,\ \ \eta \in H = (0, \infty)$$

και μ γνωστή σταθερά. Με ανάλογους υπολογισμούς η πληροφορία του Fisher που περιέχεται στο $\underset{\sim}{X}$ για το η βρίσκουμε ότι είναι

$$I^*(\eta) = \frac{n}{2\eta^2} \quad \left(= \frac{n}{2\theta^4}\right)\ , \tag{5.28}$$

ενώ, όπως είδαμε στο Παράδειγμα 5.2.12

$$I(\theta) = \frac{2n}{\theta^2} \, . \tag{5.29}$$

Ένα πρώτο συμπέρασμα που συνάγεται από τις (5.28) και (5.29) είναι ότι η πληροφορία, $I(\theta)$, που παρέχει το $\underset{\sim}{X}$ για την τυπική απόκλιση θ διαφέρει από την πληροφορία, $I^*(\eta)$, που παρέχει για τη διασπορά η. Η διαπίστωση αυτή δεν είναι συμπτωματική, αντίθετα, επιβεβαιώνει το γενικό κανόνα ότι η πληροφορία του Fisher είναι συνυφασμένη με τη συγκεκριμένη άγνωστη παράμετρο, μέσω της οποίας έχουμε επιλέξει να παραμετροποιήσουμε την οικογένεια κατανομών του δείγματος $\underset{\sim}{X}$. Εξ άλλου η συσχέτιση της πληροφορίας του Fisher με την άγνωστη παράμετρο θ είναι εγγενής στον ορισμό της, $I(\theta) = \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta)\right)^2\right]$, αφού η παραγώγιση γίνεται *ως προς* θ. Αλλαγή παραμέτρου από θ σε $\eta = \theta^2$, επιβάλλει παραγώγιση ως προς η της συνάρτησης

$$\ln f(\underset{\sim}{x};\theta) = \ln f(\underset{\sim}{x};\eta^{1/2}) = \ln f^*(\underset{\sim}{x};\eta) \; ,$$

όπου $f^*(\underset{\sim}{x};\eta)$ είναι η πυκνότητα του δείγματος $\underset{\sim}{X}$ με παράμετρο το η. Από τον τύπο παραγώγισης σύνθετης συνάρτησης έχουμε

$$\begin{aligned}
\frac{\partial}{\partial\eta}\ln f^*(\underset{\sim}{x};\eta) &= \frac{\partial}{\partial\theta}\ln f(\underset{\sim}{x};\theta)\Big|_{\theta=\eta^{1/2}}\frac{\partial\theta}{\partial\eta} \\
&= \frac{\partial}{\partial\theta}\ln f(\underset{\sim}{x};\theta)\Big|_{\theta=\eta^{1/2}}\frac{\partial\eta^{1/2}}{\partial\eta} \\
&= \frac{\partial}{\partial\theta}\ln f(\underset{\sim}{x};\theta)\Big|_{\theta=\eta^{1/2}}\frac{1}{2\sqrt{\eta}}.
\end{aligned}$$

Επομένως,

$$\begin{aligned}
I^*(\eta) = \mathbb{E}_\eta\left[\left(\frac{\partial}{\partial\eta}\ln f^*(\underset{\sim}{x};\eta)\right)^2\right] &= \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{x};\theta)\Big|_{\theta=\eta^{1/2}}\right)^2\frac{1}{4\eta}\right] \\
&= I(\eta^{1/2})\frac{1}{4\eta} = \frac{2n}{(\eta^{1/2})^2}\frac{1}{4\eta} = \frac{n}{2\eta^2},
\end{aligned}$$

που είναι η (5.28). Στην Άσκηση 5.13 ζητείται να δειχθεί μία γενική σχέση που συνδέει τους αριθμούς πληροφορίας του Fisher για δύο διαφορετικές παραμετροποιήσεις της πυκνότητας του $\underset{\sim}{X}$.

Ως δεύτερο συμπέρασμα παρατηρούμε ότι οι πληροφορίες $I^*(\eta)$ και $I(\theta)$ είναι φθίνουσες συναρτήσεις της αντίστοιχης παραμέτρου και συνεπώς η εκτίμηση της διασποράς η (αντίστοιχα, της τυπικής απόκλισης θ) μπορεί να γίνει με μεγαλύτερη ακρίβεια όταν η πραγματική τιμή της είναι μικρή. Τέλος έχουμε

$$\frac{I^*(\eta)}{I(\theta)} = \frac{1}{4\eta} \to \infty \text{ καθώς } \eta \to 0,$$

που ερμηνεύεται ότι η διασπορά μπορεί να εκτιμηθεί με πολύ μεγαλύτερη ακρίβεια από την τυπική απόκλιση, όταν η πραγματική τους τιμή είναι «πολύ μικρή», ενώ το αντίστροφο ισχύει όταν η πραγματική τους τιμή είναι «πολύ μεγάλη».

Το Κ.Φ.C-R είναι ανεξάρτητο από την επιλεγείσα παραμετροποίηση. Πράγματι, με παράμετρο το θ, το Κ.Φ.C-R για τη διασπορά των αμερόληπτων εκτιμητών του θ είναι, όπως είδαμε στο Παράδειγμα 5.2.12,

$$\text{Κ.Φ. C-R} = \frac{(g'(\theta))^2}{I(\theta)} = \frac{1}{I(\theta)} = \frac{\theta^2}{2n},$$

επειδή $g(\theta) = \theta$. Το ίδιο κάτω φράγμα, με παράμετρο το η είναι

$$\text{Κ.Φ. C-R} = \frac{(g_1'(\eta))^2}{I^*(\eta)} = \frac{\eta}{2n} \quad \left(= \frac{\theta^2}{2n}\right),$$

επειδή $g_1(\eta) = \eta^{1/2} = \theta$ και $g_1'(\eta) = \frac{1}{2\sqrt{\eta}}$.

Παρόμοια, είτε με τη μία είτε με την άλλη παραμετροποίηση, το Κ.Φ.C-R για τη διασπορά των αμερόληπτων εκτιμητών του θ^2 είναι (Παράδειγμα 5.2.12)

$$\text{Κ.Φ. C-R} = \frac{2\theta^4}{n} = \frac{2\eta^2}{n}.$$

Στην Άσκηση 5.14 ζητείται να δειχθεί η μη εξάρτηση του Κ.Φ.C-R από οποιαδήποτε παραμετροποίηση της πυκνότητας του $\underset{\sim}{X}$.

Παράδειγμα 5.2.14. (απλό γραμμικό μοντέλο - πληροφορία $\mathbf{I}(\theta)$) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ανεξάρτητες παρατηρήσεις με κανονικές κατανομές $\mathcal{N}(\theta t_i, 1)$, $i = 1, \ldots, n$ αντίστοιχα, όπου t_i είναι γνωστές σταθερές και

$\theta \in \Theta = \mathbb{R}$ άγνωστο. Τότε

$$f(\underset{\sim}{x};\theta) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i - \theta t_i)^2}$$

και

$$\frac{\partial}{\partial\theta} \ln f(\underset{\sim}{x};\theta) = \sum_{i=1}^{n}(x_i - \theta t_i)t_i = \sum_{i=1}^{n} x_i t_i - \theta \sum_{i=1}^{n} t_i^2.$$

Επομένως, έχουμε

$$\begin{aligned} I(\theta) &= \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta)\right)^2\right] = \mathbb{E}_\theta\left(\sum_{i=1}^{n} t_i X_i - \theta\sum_{i=1}^{n} t_i^2\right)^2 \\ &= \mathbb{V}ar_\theta\left(\sum_{i=1}^{n} t_i X_i\right) = \sum_{i=1}^{n} t_i^2 \mathbb{V}ar_\theta X_i = \sum_{i=1}^{n} t_i^2. \end{aligned}$$

Η τρίτη ισότητα είναι ο ορισμός της διασποράς της τυχαίας μεταβλητής $\sum_{i=1}^{n} t_i X_i$, επειδή

$$\mathbb{E}_\theta \sum_{i=1}^{n} t_i X_i = \sum_{i=1}^{n} t_i \mathbb{E}_\theta X_i = \sum_{i=1}^{n} \theta t_i^2 = \theta \sum_{i=1}^{n} t_i^2.$$

Παράδειγμα 5.2.15. (κατανομή Βήτα - πληροφορία $I(\theta)$ και αποδοτικοί εκτιμητές) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ τυχαίο δείγμα από την κατανομή Βήτα, $Beta(\theta, 1)$, με πυκνότητα

$$f_1(x;\theta) = \theta x^{\theta-1}, \;\; 0 < x < 1, \; \theta \in \Theta = (0, \infty).$$

Οι συνθήκες I1–I4 επαληθεύονται εύκολα, μέσω της Πρότασης 5.2.2. Τότε έχουμε

$$f(\underset{\sim}{x};\theta) = \theta^n \left(\prod_{i=1}^{n} x_i\right)^{\theta-1} = e^{n\ln\theta + (\theta-1)\sum\limits_{i=1}^{n}\ln x_i},$$

$$\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{x};\theta) = \frac{n}{\theta} + \sum_{i=1}^{n}\ln x_i \;\; \text{και} \;\; \frac{\partial^2}{\partial\theta^2}\ln f(\underset{\sim}{x};\theta) = -\frac{n}{\theta^2},$$

οπότε, από την (5.9), προκύπτει

$$I(\theta) = -\mathbb{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\ln f(\underset{\sim}{X};\theta)\right) = -\mathbb{E}_\theta\left(-\frac{n}{\theta^2}\right) = \frac{n}{\theta^2}.$$

Για σύγκριση, προκειμένου να χρησιμοποιήσουμε τον ορισμό του $I(\theta)$, θα βρούμε την κατανομή του $\ln X_i$. Από τον τύπο της Ενότητας 1.7 προκύπτει ότι η τυχαία μεταβλητή $Y_i = -\ln X_i$ έχει εκθετική κατανομή με πυκνότητα $\theta e^{-\theta y}, y > 0$, συνεπώς, $\mathbb{E}_\theta Y_i = \frac{1}{\theta}$, $\mathbb{V}ar_\theta Y_i = \frac{1}{\theta^2}$ και $\mathbb{E}_\theta \sum_{i=1}^{n} Y_i = \dfrac{n}{\theta}$. Επομένως,

$$\begin{aligned} I(\theta) &= \mathbb{E}_\theta \left[\left(\tfrac{\partial}{\partial\theta} \ln f(\underset{\sim}{X};\theta) \right)^2 \right] = \mathbb{E}_\theta \left(\sum_{i=1}^{n} Y_i - \frac{n}{\theta} \right)^2 \\ &= \mathbb{V}ar_\theta \left(\sum_{i=1}^{n} Y_i \right) = \sum_{i=1}^{n} \mathbb{V}ar_\theta Y_i = \frac{n}{\theta^2}. \end{aligned}$$

Επειδή (με πιθανότητα 1)

$$\tfrac{\partial}{\partial\theta} \ln f(\underset{\sim}{X};\theta) = \frac{n}{\theta} + \sum_{i=1}^{n} \ln X_i = -\left(-\sum_{i=1}^{n} \ln X_i - \frac{n}{\theta} \right) ,$$

από την Παρατήρηση 5.2.2 συμπεραίνουμε αμέσως ότι η στατιστική συνάρτηση $D(\underset{\sim}{X}) = -\sum_{i=1}^{n} \ln X_i$ είναι αποδοτικός εκτιμητής του $g(\theta) = \frac{n}{\theta}$ και από την Παρατήρηση 5.2.1 ότι όλοι οι αποδοτικοί εκτιμητές είναι $aD(\underset{\sim}{X}) + \beta = a\left(-\sum_{i=1}^{n} \ln X_i \right) + \beta$ με αντίστοιχες τιμές που εκτιμούνται αποδοτικά, τις $g_1(\theta) = ag(\theta) + \beta = a\frac{n}{\theta} + \beta$. Ειδικά για $\beta = 0$ ή 1 και $a = \frac{1}{n}$, οι αποδοτικοί εκτιμητές του $\frac{1}{\theta}$ και του $\frac{1}{\theta} + 1 = \frac{\theta+1}{\theta}$ είναι $-\dfrac{1}{n}\sum_{i=1}^{n} \ln X_i$ και $-\dfrac{1}{n}\sum_{i=1}^{n} \ln X_i + 1$ αντίστοιχα, ενώ δεν υπάρχει αποδοτικός εκτιμητής του θ.

Παράδειγμα 5.2.16. (ανεξάρτητες Poisson - πληροφορία $\mathbf{I}(\theta)$, Κ.Φ. C-R και αποδοτικός εκτιμητής) Έστω ότι $\underset{\sim}{X} = (X_1, \ldots, X_n)$ είναι ανεξάρτητες παρατηρήσεις από Poisson κατανομές $\mathcal{P}(a_i\theta)$, με γνωστά a_i, $i = 1, 2, \ldots, n$ και άγνωστο $\theta \in \Theta = (0, \infty)$. Θα υπολογίσουμε τον αριθμό πληροφορίας του Fisher $I(\theta)$ και το Κάτω Φράγμα των Cramér-Rao για τη διασπορά των αμερόληπτων εκτιμητών του $g(\theta) = \theta$.

Επαληθεύουμε πρώτα τις συνθήκες I1–I5 του Θεωρήματος 5.1.1.

I1. Ο παραμετρικός χώρος $\Theta = (0, \infty)$ είναι ανοικτό διάστημα του $\mathbb{R}$.

I2 - I4. Η οικογένεια κατανομών του $\underset{\sim}{X}$ ανήκει στην Μ.Ε.Ο.Κ.:

1. Το σύνολο $\mathcal{S} = \{\underset{\sim}{x} : f(\underset{\sim}{x}, \theta) > 0\} = \mathbb{N}^n$ δεν εξαρτάται από το θ.

2. $$\begin{aligned} f(\underset{\sim}{x}, \theta) &= \prod_{i=1}^{n} e^{-a_i\theta} \frac{(a_i\theta)^{x_i}}{x_i!} = e^{-\sum_{i=1}^n a_i\theta} \left(\prod_{i=1}^{n} a_i^{x_i} \right) \frac{\theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} \\ &= \exp\left\{ -(\sum_{i=1}^{n} a_i)\theta - \sum_{i=1}^{n} \ln x_i! - \sum_{i=1}^{n} x_i \ln a_i + \ln\theta \sum_{i=1}^{n} x_i \right\} \\ &= e^{A(\theta)+B(\underset{\sim}{x})+C(\theta)D(\underset{\sim}{x})}, \end{aligned}$$

όπου,
$$A(\theta) = -(\sum_{i=1}^{n} a_i)\theta,\ B(\underset{\sim}{x}) = \sum_{i=1}^{n} \ln x_i! - \sum_{i=1}^{n} x_i \ln a_i,\ C(\theta) = \ln\theta,\ \text{και}$$
$D(\underset{\sim}{x}) = \sum_{i=1}^{n} x_i$. Επειδή η συνάρτηση $C(\theta) = \ln\theta$ έχει συνεχή και μη μηδενική παράγωγο, $C'(\theta) = \frac{1}{\theta}$, για $\theta \in (0, \infty)$, ισχύουν οι συνθήκες I2–I4, λόγω της Πρότασης 5.2.2.

I5. Από τις (5.10) και (5.13) παίρνουμε,
$$I(\theta) = \sum_{i=1}^{n} I_i(\theta), \quad \text{όπου} \quad I_i(\theta) = -\mathbb{E}_\theta\left(\frac{\partial^2}{\partial\theta^2} \ln f_i(X_i;\theta) \right).$$
Έχουμε λοιπόν,
$$\begin{aligned} f_i(x_i;\theta) &= e^{-a_i\theta} \frac{(a_i\theta)^{x_i}}{x_i!}, \\ \ln f_i(x_i;\theta) &= -a_i\theta + x_i \ln a_i + x_i \ln\theta - \ln x_i!, \\ \frac{\partial}{\partial\theta} \ln f_i(x_i;\theta) &= -a_i + \frac{x_i}{\theta}, \frac{\partial^2}{\partial\theta^2} \ln f_i(x_i;\theta) = -\frac{x_i}{\theta^2}, \\ I_i(\theta) &= -\mathbb{E}_\theta\left(-\frac{X_i}{\theta^2} \right) = \frac{1}{\theta^2}\mathbb{E}_\theta X_i = \frac{1}{\theta^2} a_i\theta = \frac{a_i}{\theta}, \end{aligned}$$
οπότε $I(\theta) = \frac{1}{\theta} \sum_{i=1}^{n} a_i > 0$, δηλαδή ισχύει και η I5.

Επομένως, από το Θεώρημα 5.1.1, το κάτω φράγμα των Cramér–Rao για τη διασπορά των αμερόληπτων εκτιμητών του $g(\theta) = \theta$ είναι
$$\text{Κ.Φ. C-R} = \frac{(g'(\theta))^2}{I(\theta)} = \frac{1}{(\sum_{i=1}^n a_i)/\theta} = \frac{\theta}{\sum_{i=1}^n a_i}.$$

Επί πλέον, μέσω της Πρότασης 5.2.4, είναι εύκολο να δειχθεί ότι ο αποδοτικός εκτιμητής του θ είναι $\sum_{i=1}^{n} X_i / \sum_{i=1}^{n} a_i$.

5.3 Ανισότητα των Cramér–Rao και πίνακας πληροφορίας του Fisher για $\theta = (\theta_1, \ldots, \theta_\kappa) \in \mathbb{R}^\kappa$

Η ανισότητα των Cramér–Rao και ο αριθμός πληροφορίας του Fisher μπορούν να επεκταθούν και στην περίπτωση που η άγνωστη παράμετρος είναι διανυσματική, δηλαδή $\theta = (\theta_1, \ldots, \theta_\kappa) \in \Theta \subset \mathbb{R}^\kappa$, $\kappa > 1$. Θα περιγράψουμε, εν συντομία, την επέκτασή τους παραλείποντας αποδείξεις και παραπέμποντας τον αναγνώστη στους Rao (1973, σελ. 326 – 329) και Lehmann and Casella (1998, Section 2.6) για περισσότερες λεπτομέρειες.

Υποθέτουμε ότι το Θ είναι ανοικτό υποσύνολο του $\mathbb{R}^\kappa$ (διαφορετικά, τα παρακάτω αφορούν το σύνολο των εσωτερικών σημείων του Θ), το σύνολο τιμών του δείγματος $\underset{\sim}{X} = (X_1, \ldots, X_n)$ $\mathcal{S} = \{\underset{\sim}{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n : f(\underset{\sim}{x}; \theta) > 0\}$ δεν εξαρτάται από το θ, ότι υπάρχουν και είναι πεπερασμένες οι μερικές παράγωγοι $\frac{\partial}{\partial \theta_i} \ln f(\underset{\sim}{x}; \theta)$, $i = 1, \ldots, \kappa$ για κάθε $\underset{\sim}{x} \in \mathcal{S}$, $\theta \in \Theta$ και επιτρέπεται η αλλαγή σειράς παραγώγισης και ολοκλήρωσης ή άθροισης όπως στις συνθήκες I3 και I4. Τότε ο πίνακας πληροφορίας ή απλά πληροφορία του Fisher που περιέχεται στο $\underset{\sim}{X}$ για το θ ορίζεται ως ο $\kappa \times \kappa$ πίνακας $I(\theta) = (I_{ij}(\theta))$ με γενικό στοιχείο

$$\begin{aligned} I_{ij}(\theta) &= \mathbb{E}_\theta \left[\tfrac{\partial}{\partial \theta_i} \ln f(\underset{\sim}{X}; \theta) \; \cdot \; \tfrac{\partial}{\partial \theta_j} \ln f(\underset{\sim}{X}; \theta) \right] \\ &= \mathbb{C}ov_\theta \left(\tfrac{\partial}{\partial \theta_i} \ln f(\underset{\sim}{X}; \theta), \tfrac{\partial}{\partial \theta_j} \ln f(\underset{\sim}{X}; \theta) \right) \quad i, j = 1, \ldots, \kappa. \end{aligned}$$

Η πρώτη ισότητα είναι ο ορισμός του $I_{ij}(\theta)$, ενώ η δεύτερη προκύπτει αμέσως από τον τύπο $\mathbb{C}ov(Y, Z) = \mathbb{E}(YZ) - \mathbb{E}Y\mathbb{E}Z$, επειδή

$$\mathbb{E}_\theta \left(\tfrac{\partial}{\partial \theta_i} \ln f(\underset{\sim}{X}; \theta) \right) = \mathbb{E}_\theta \left(\tfrac{\partial}{\partial \theta_j} \ln f(\underset{\sim}{X}; \theta) \right) = 0$$

κατ' αντιστοιχία με την (5.3). Ο πίνακας $I(\theta)$ είναι συμμετρικός, $I_{ij}(\theta) = I_{ji}(\theta)$, με διαγώνια στοιχεία τις διασπορές $I_{ii}(\theta) = \mathbb{V}ar_\theta \left(\frac{\partial}{\partial \theta_i} \ln f(\underset{\sim}{X}; \theta) \right)$, $i = 1, \ldots, \kappa$. Μπορεί να δειχθεί ότι ο $I(\theta)$ είναι θετικά ημιορισμένος (positive semidefinite) πίνακας, αλλά υποθέτουμε, επιπλέον, ότι είναι θετικά ορισμένος (positive definite) (συνθήκη αντίστοιχη της I5). Η υπόθεση αυτή είναι πολύ ήπια και γενικά ισχύει εκτός εάν οι τυχαίες μεταβλητές $\frac{\partial}{\partial \theta_i} \ln f(\underset{\sim}{X};\theta)$ είναι γραμμικά εξαρτημένες μεταξύ τους, μία ακραία περίπτωση που δεν θα μας αποσχολήσει.

Κατ' αναλογία με την (5.7), αν Y είναι μία τυχαία «ποσότητα» με πυκνότητα $h(y;\theta)$, $\theta = (\theta_1, \ldots, \theta_\kappa) \in \Theta \subset \mathbb{R}^\kappa$, τότε ο πίνακας πληροφορίας ή πληροφορία του Fisher που περιέχεται στο Y για το θ είναι ο $\kappa \times \kappa$ πίνακας που προκύπτει από τον $I(\theta)$ θέτοντας $h(Y;\theta)$ αντί $f(\underset{\sim}{X};\theta)$. Ειδικά, για $Y = X_i$ η πληροφορία του Fisher που περιέχεται στην παρατήρηση X_i συμβολίζεται με $I_i(\theta)$, $i = 1, \ldots, n$, οπότε στην περίπτωση ανεξαρτησίας των X_i ισχύει η προσθετική ιδιότητα,

$$I(\theta) = \sum_{i=1}^{n} I_i(\theta) \ ,$$

ενώ, αν επιπλέον οι X_i, $i = 1, \ldots, n$, έχουν την ίδια κατανομή, τότε

$$I(\theta) = nI_1(\theta). \tag{5.30}$$

Έστω τώρα $T(\underset{\sim}{X})$ ένας εκτιμητής του $g(\theta)$ με μέση τιμή $\mathbb{E}_\theta T(\underset{\sim}{X}) = \tau(\theta)$ για την οποία θεωρούμε ότι υπάρχουν και είναι πεπερασμένες οι μερικές παράγωγοι $\tau_i(\theta) = \frac{\partial}{\partial \theta_i}\tau(\theta)$, $i = 1, \ldots, \kappa$. Συμβολίζουμε

$$\left(\tfrac{\partial}{\partial \theta}\tau(\theta)\right)' = (\tau_1(\theta), \ldots, \tau_\kappa(\theta))$$

τον $1 \times \kappa$ πίνακα με γενικό στοιχείο $\tau_i(\theta)$ και διευκρινίζουμε ότι παριστάνει εδώ τον ανάστροφο πίνακα, οπότε $\frac{\partial}{\partial \theta}\tau(\theta)$ είναι ο $\kappa \times 1$ πίνακας με το ίδιο γενικό στοιχείο. Υπό τις παραπάνω συνθήκες μπορεί να δειχθεί ότι

$$\mathbb{V}ar_\theta T(\underset{\sim}{X}) \geqslant \left(\tfrac{\partial}{\partial \theta}\tau(\theta)\right)' I^{-1}(\theta) \left(\tfrac{\partial}{\partial \theta}\tau(\theta)\right) \ , \ \ \forall \theta \in \Theta, \tag{5.31}$$

όπου $I^{-1}(\theta)$ είναι ο αντίστροφος του πίνακα $I(\theta)$. Σημειώνουμε ότι η ύπαρξη του $I^{-1}(\theta)$ εξασφαλίζεται από την υπόθεση ότι ο $I(\theta)$ είναι θετικά ορισμένος. Προς διευκρίνιση, το δεξιό μέλος της (5.31) είναι όντως αριθμός (και μάλιστα μη αρνητικός επειδή ο $I^{-1}(\theta)$ είναι θετικά ορισμένος) ως γινόμενο τριών πινάκων $1\times\kappa$, $\kappa\times\kappa$ και $\kappa\times 1$, αντίστοιχα. Αν ο $T(\underset{\sim}{X})$ είναι αμερόληπτος εκτιμητής του $g(\theta)$, τότε $\tau(\theta)=g(\theta)$ και η (5.31) γίνεται

$$\mathbb{V}ar_\theta T(\underset{\sim}{X}) \geqslant \left(\tfrac{\partial}{\partial\theta}g(\theta)\right)' I^{-1}(\theta)\left(\tfrac{\partial}{\partial\theta}g(\theta)\right)\ ,\ \ \forall\theta\in\Theta, \tag{5.32}$$

όπου

$$\left(\tfrac{\partial}{\partial\theta}g(\theta)\right)' = (g_1(\theta),\ldots,g_\kappa(\theta)) \ \text{ και }\ g_i(\theta)=\tfrac{\partial}{\partial\theta_i}g(\theta).$$

Κάθε μία από τις (5.31) και (5.32) είναι μία επέκταση της ανισότητας των Cramér–Rao για διανυσματική παράμετρο θ, ενώ το δεξιό μέλος της (5.32) είναι το κάτω φράγμα των Cramér–Rao για τη διασπορά αμερόληπτων εκτιμητών του $g(\theta)$,

$$\text{Κ.Φ. C-R } = \left(\tfrac{\partial}{\partial\theta}g(\theta)\right)' I^{-1}(\theta)\left(\tfrac{\partial}{\partial\theta}g(\theta)\right). \tag{5.33}$$

Περαιτέρω, ο $T(\underset{\sim}{X})$ είναι αποδοτικός εκτιμητής του $g(\theta)$, αν είναι αμερόληπτος και η (5.32) ισχύει ως ισότητα, δηλαδή η διασπορά $\mathbb{V}ar_\theta T(\underset{\sim}{X})$ συμπίπτει με το Κ.Φ.C-R για κάθε $\theta\in\Theta$. Επιπροσθέτως, μπορεί να δειχθεί ότι η (5.32) ισχύει ως ισότητα, εάν και μόνον εάν (με πιθανότητα 1)

$$T(\underset{\sim}{X}) = \sum_{i=1}^{\kappa} c_i(\theta)\tfrac{\partial}{\partial\theta_i}\ln f(\underset{\sim}{X};\theta)+c_{\kappa+1}(\theta), \tag{5.34}$$

όπου $c_i(\theta)$ είναι συναρτήσεις του θ. Η σχέση (5.34) είναι η επέκταση της (5.18) (αν η (5.18) λυθεί ως προς $T(\underset{\sim}{X})$). Επομένως, η (5.34) δίνει τη γενική μορφή ενός αποδοτικού εκτιμητή στην περίπτωση διανυσματικής παραμέτρου θ, από την οποία μάλιστα προκύπτει ότι $g(\theta)=\mathbb{E}_\theta T(\underset{\sim}{X}) = c_{\kappa+1}(\theta)$. Αποδοτικοί εκτιμητές υπάρχουν μόνον σε εκθετικές οικογένειες κατανομών τάξης κ, τις οποίες θα ορίσουμε και θα μελετήσουμε στο επόμενο κεφάλαιο και οι οποίες για $\kappa=1$ είναι οι Μ.Ε.Ο.Κ. .

Παράδειγμα 5.3.1. (κανονική κατανομή με άγνωστη μέση τιμή και διασπορά) Έστω $\underset{\sim}{X}=(X_1,\ldots,X_n)$ ένα τυχαίο δείγμα από την κανονική κατανομή $\mathcal{N}(\mu,\sigma^2)$, όπου μ και σ^2 είναι άγνωστες παράμετροι. Θέτουμε $\theta_1=\mu$, $\theta_2=\sigma^2$ και $\Theta=\left\{(\theta_1,\theta_2)\in\mathbb{R}^2:\theta_1\in\mathbb{R},\theta_2>0\right\}$. Τότε η κοινή πυκνότητα των X_i είναι

$$f_1(x;\theta) = \frac{1}{\sqrt{2\pi\theta_2}}\, e^{-\frac{1}{2\theta_2}(x-\theta_1)^2} \ , \quad x \in \mathbb{R}\ ,$$

και επομένως

$$\ln f_1(x;\theta) = -\frac{1}{2}\ln(2\pi\theta_2) - \frac{1}{2\theta_2}(x-\theta_1)^2,$$

$$\frac{\partial}{\partial\theta_1}\ln f_1(x;\theta) = \frac{1}{\theta_2}(x-\theta_1),$$

$$\frac{\partial}{\partial\theta_2}\ln f_1(x;\theta) = -\frac{1}{2\theta_2} + \frac{1}{2\theta_2^2}(x-\theta_1)^2.$$

Αρχικά, υπολογίζουμε τον 2×2 πίνακα $I_1(\theta) = (I_{ij}^{(1)}(\theta))$, ώστε εν συνεχεία να χρησιμοποιήσουμε την (5.30) για την εύρεση του $I(\theta)$. Έχουμε,

$$I_{11}^{(1)}(\theta) = \mathbb{V}ar_\theta\left(\frac{X_1-\theta_1}{\theta_2}\right) = \frac{1}{\theta_2^2}\theta_2 = \frac{1}{\theta_2}\ ,$$

$$I_{22}^{(1)}(\theta) = \mathbb{V}ar_\theta\left(-\frac{1}{2\theta_2} + \frac{1}{2\theta_2^2}(X_1-\theta_1)^2\right) = \frac{1}{4\theta_2^2}\mathbb{V}ar(\mathcal{X}_1^2) = \frac{1}{2\theta_2^2}\ ,$$

επειδή $(X_1-\theta_1)^2/\theta_2 \sim \mathcal{X}_1^2$ και $\mathbb{V}ar(\mathcal{X}_1^2) = 2$,

$$\begin{aligned} I_{12}^{(1)}(\theta) = I_{21}^{(1)}(\theta) &= \mathbb{C}ov_\theta\left(\frac{1}{\theta_2}(X_1-\theta_1), -\frac{1}{2\theta_2} + \frac{1}{2\theta_2^2}(X_1-\theta_1)^2\right) \\ &= \frac{1}{2\theta_2^3}\mathbb{C}ov_\theta(X_1-\theta_1, (X-\theta_1)^2) \\ &\quad \text{(βλέπε τις ιδιότητες της συνδιασποράς στην ενότητα 1)} \\ &= \frac{1}{2\theta_2^3}\mathbb{E}_\theta(X_1-\theta_1)^3 = 0, \end{aligned}$$

ως κεντρική ροπή περιττής τάξης κανονικής κατανομής.

Επομένως, από την (5.30), προκύπτει

$$I(\theta) = \begin{pmatrix} \frac{n}{\theta_2} & 0 \\ 0 & \frac{n}{2\theta_2^2} \end{pmatrix} \text{ με αντίστροφο πίνακα } I^{-1}(\theta) = \begin{pmatrix} \frac{\theta_2}{n} & 0 \\ 0 & \frac{2\theta_2^2}{n} \end{pmatrix}.$$

Περαιτέρω, για κάθε αμερόληπτο εκτιμητή $T(\underset{\sim}{X})$ της μέσης τιμής $g(\theta) = \theta_1 = \mu$, από τις (5.32) και (5.33), έχουμε

$$\mathbb{V}ar_\theta T(\underset{\sim}{X}) \geqslant \text{Κ.Φ. C-R } = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} \frac{\theta_2}{n} & 0 \\ 0 & \frac{2\theta_2^2}{n} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{\theta_2}{n},$$

επειδή δε $\mathbb{V}ar_\theta \bar{X} = \dfrac{\theta_2}{n}$, συμπεραίνουμε ότι $\bar{X}$ είναι αποδοτικός εκτιμητής του μ ακόμη και στην περίπτωση άγνωστης διασποράς (βλέπε επίσης το Παράδειγμα 5.2.11).

Από την άλλη πλευρά, για τους αμερόληπτους εκτιμητές $T_1(\underset{\sim}{X})$ της διασποράς $g(\theta) = \theta_2 = \sigma^2$, παίρνουμε

$$\mathbb{V}ar_\theta T_1(\underset{\sim}{X}) \geqslant \text{Κ.Φ. C-R } = \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} \dfrac{\theta_2}{n} & 0 \\ 0 & \dfrac{2\theta_2^2}{n} \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{2\theta_2^2}{n}.$$

Θέτοντας $T_1(\underset{\sim}{X}) = S^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^n (X_i - \bar{X})^2$, γνωρίζουμε, από την Πρόταση 4.2.4, ότι S^2 είναι αμερόληπτος εκτιμητής με

$$\begin{aligned}
\mathbb{V}ar_\theta S^2 &= \frac{\mathbb{E}_\theta (X_1 - \theta_1)^4}{n} - \frac{(n-3)}{n(n-1)}\theta_2^2 = \frac{\theta_2^2}{n}\mathbb{E}[(\mathcal{X}_1^2)^2] - \frac{(n-3)\theta_2^2}{n(n-1)} \\
&= \frac{\theta_2^2}{n}\left[\mathbb{V}ar(\mathcal{X}_1^2) + \big(\mathbb{E}(\mathcal{X}_1^2)\big)^2\right] - \frac{(n-3)\theta_2^2}{n(n-1)} \\
&= \frac{\theta_2^2}{n}(2+1) - \frac{(n-3)\theta_2^2}{n(n-1)} = \frac{2\theta_2^2}{n-1} > \frac{2\theta_2^2}{n} = \text{Κ.Φ. C-R} \;.
\end{aligned}$$

Συμπεραίνουμε λοιπόν ότι S^2 δεν είναι αποδοτικός εκτιμητής του σ^2. Στο Κεφάλαιο 6, πάντως θα δειχθεί ότι S^2 έχει την ελάχιστη διασπορά μεταξύ όλων των αμερόληπτων εκτιμητών του σ^2.

Γενικότερα, σύμφωνα με την (5.34), οι αποδοτικοί εκτιμητές είναι

$$\begin{aligned}
T(\underset{\sim}{X}) &= c_1(\theta)\left\{\sum_{i=1}^n \frac{1}{\theta_2}(X_i - \theta_1)\right\} \\
&\quad + c_2(\theta)\left\{\sum_{i=1}^n \left(-\frac{1}{\theta_2} + \frac{1}{2\theta_2^2}(X_i - \theta_1)^2\right)\right\} + c_3(\theta) \\
&= c_1^*(\theta)\sum_{i=1}^n X_i + c_2^*(\theta)\sum_{i=1}^n X_i^2 + c_3^*(\theta),
\end{aligned}$$

όπου $c_1^*(\theta) = \frac{c_1(\theta)}{\theta_2} - \frac{c_2(\theta)\theta_1}{\theta_2}$, $c_2^*(\theta) = \frac{1}{2\theta_2^2}c_2(\theta)$, $c_3^*(\theta) = -\frac{nc_1(\theta)\theta_1}{\theta_2} - \frac{nc_2(\theta)}{2\theta_2} + \frac{nc_2(\theta)\theta_1^2}{2\theta_2^2} + c_3(\theta)$.

Όμως, $T(\underset{\sim}{X})$ είναι στατιστική συνάρτηση, άρα τελικά οι σταθερές $c_1^*(\theta)$, $c_2^*(\theta)$, $c_3^*(\theta)$ δεν εξαρτώνται από το θ, οπότε $T(\underset{\sim}{X}) = \alpha \sum\limits_{i=1}^{n} X_i + \beta \sum\limits_{i=1}^{n} X_i^2 + \gamma$ με αντίστοιχες τιμές $g(\theta)$ που εκτιμώνται αποδοτικά, τις $g(\theta) = \mathbb{E}_\theta T(\underset{\sim}{X}) = \alpha(n\mu) + \beta(\mu^2 + \sigma^2) + \gamma$. $\square$

Το Θεώρημα 5.1.1 μπορεί να χρησιμοποιηθεί και στην περίπτωση διανυσματικής παραμέτρου $\theta = (\theta_1, \ldots, \theta_\kappa)$, αρκεί η παραγώγιση να γίνει ως προς μία συγκεκριμένη συνιστώσα θ_i, $i = 1, \ldots, \kappa$. Έτσι είναι εύκολο να δειχθεί (ακολουθώντας την απόδειξη του Θεωρήματος 5.1.1) ότι

$$\mathbb{V}ar_\theta T(\underset{\sim}{X}) \geqslant \frac{(\tau_i(\theta))^2}{I_{ii}(\theta)}\ , \quad \forall \theta \in \Theta, \tag{5.35}$$

ενώ για την περίπτωση αμερόληπτου εκτιμητή του $g(\theta)$ η (5.35) γίνεται

$$\mathbb{V}ar_\theta T(\underset{\sim}{X}) \geqslant \frac{(g_i(\theta))^2}{I_{ii}(\theta)}\ , \quad \forall \theta \in \Theta. \tag{5.36}$$

Μπορούμε λοιπόν να παράγουμε αμέσως ανισότητες τύπου Crámer - Rao προσποιούμενοι ότι έχουμε μια μόνον (πραγματική) άγνωστη παράμετρο θ_i και «κρατώντας» τις υπόλοιπες «σταθερές». Μεγαλύτερα, άρα και πιο ακριβή, είναι πάντως τα φράγματα (5.32) και (5.33) και μόνον αν ο $I(\theta)$ είναι διαγώνιος και $\tau(\theta)$ είναι συνάρτηση μόνον του θ_i, τότε τα (5.32) και (5.33) συμπίπτουν αντίστοιχα με τα (5.35) και (5.36). Αυτή ακριβώς είναι η περίπτωση του Παραδείγματος 5.3.1.

Θα δείξουμε ότι το φράγμα (5.33) είναι μεγαλύτερο του φράγματος (5.36) στην περίπτωση $\theta = (\theta_1, \theta_2)$ και $g(\theta) = \theta_1$. Η απόδειξη στη γενική περίπτωση δίνεται από τους Lehmann and Casella (1998, σελ. 128). Θέτουμε $I(\theta) = \begin{pmatrix} I_{11}(\theta) & I_{12}(\theta) \\ I_{21}(\theta) & I_{22}(\theta) \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$ με αντίστροφο πίνακα $I^{-1}(\theta) = \dfrac{1}{\sigma_1^2\sigma_2^2 - \sigma_{12}^2} \begin{pmatrix} \sigma_2^2 & -\sigma_{12} \\ -\sigma_{12} & \sigma_1^2 \end{pmatrix}$.

Το φράγμα (5.36) για $i = 1$ είναι $\dfrac{1}{\sigma_1^2}$, αφού $g_1(\theta) = 1$, ενώ τετριμμένα για

$i = 2$ είναι 0, αφού $g_2(\theta) = 0$. Το φράγμα (5.33) είναι

$$(\ 1 \quad 0 \ ) \, I^{-1}(\theta) \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{\sigma_2^2}{\sigma_1^2\sigma_2^2 - \sigma_{12}^2} \geqslant \frac{1}{\sigma_1^2},$$

επειδή από την ανισότητα Cauchy - Schwarz $\sigma_1^2\sigma_2^2 - \sigma_{12}^2 > 0$, ενώ η ισότητα ισχύει μόνον εάν $\sigma_{12} = 0$, δηλαδή ο $I(\theta)$ είναι διαγώνιος.

5.4 Ασκήσεις

5.1. Εάν $\underset{\sim}{X} = (X_1, \ldots, X_n)$ είναι ένα τυχαίο δείγμα από την κατανομή Γάμμα $\mathcal{G}(\alpha, \theta)$, με α γνωστό και $\theta \in \Theta = (0, \infty)$ άγνωστο, να υπολογιστεί ο αριθμός πληροφορίας $I(\theta)$ και το κάτω φράγμα των Cramér–Rao για τη διασπορά αμερολήπτων εκτιμητών του θ και του θ^2. Για ποιές συναρτήσεις g μπορούμε να βρούμε αποδοτικό εκτιμητή της τιμής $g(\theta)$;

5.2. Εάν $\underset{\sim}{X} = (X_1, \ldots, X_n)$ είναι ανεξάρτητες παρατηρήσεις με διωνυμικές κατανομές $\mathcal{B}(n_i, \theta)$, $i = 1, \ldots, n$ αντίστοιχα, $\theta \in \Theta = (0, 1)$, να υπολογιστεί ο αριθμός πληροφορίας $I(\theta)$. Επιπλέον, να βρεθεί ο αποδοτικός εκτιμητής του θ.

5.3. Ποιές από τις παρακάτω οικογένειες κατανομών είναι Μ.Ε.Ο.Κ. ;

α. $f(x; \theta) = e^{-(x-\theta)}, \quad x \geq \theta, \ \theta \in \Theta = (0, \infty)$.

β. $f(x; \theta) = \theta(1 - x)^{\theta - 1}, \quad 0 < x < 1, \ \theta \in \Theta = (0, \infty)$.

γ. $f(x; \theta) = (1 - \theta)^{x-1} \cdot \theta, \quad x = 1, 2, \ldots, \ \theta \in \Theta = (0, 1)$.

δ. $f(x; \theta) = \theta e^{-\theta(x-1)}, \quad x \geq 1, \ \theta \in \Theta = (0, \infty)$.

5.4. Δίνεται τυχαίο δείγμα $\underset{\sim}{X} = (X_1, X_2, \ldots, X_n)$ από την εκθετική κατανομή $\mathcal{E}(\theta^2)$, με άγνωστο $\theta > 0$. Να υπολογιστεί το κάτω φράγμα των Cramér-Rao για τη διασπορά αμερόληπτων εκτιμητών των θ και θ^2. Για ποιες συναρτήσεις g μπορούμε να βρούμε αποδοτικούς εκτιμητές του $g(\theta)$;

5.5. Έστω $\underset{\sim}{X} = (X_1, X_2)$, όπου X_1 και X_2 είναι ανεξάρτητες παρατηρήσεις με κανονικές κατανομές $X_1 \sim \mathcal{N}(\theta, \sigma_1^2)$, $X_2 \sim \mathcal{N}(\theta, \sigma_2^2)$, σ_1^2 και σ_2^2 είναι γνωστές σταθερές και $\theta \in \mathbb{R}$ άγνωστο. Να βρεθεί αποδοτικός εκτιμητής του θ.

5.6. Αν $\underset{\sim}{X} = (X_1, X_2, \ldots, X_n)$, είναι ένα τυχαίο δείγμα από τη γεωμετρική κατανομή $Ge(\theta)$, με $\theta \in (0,1)$ άγνωστο, να βρεθεί ο αποδοτικός εκτιμητής του $g(\theta) = \dfrac{1}{\theta}$.

5.7. Δίνεται ένα τυχαίο δείγμα $\underset{\sim}{X} = (X_1, X_2, \ldots, X_n)$ από την κανονική κατανομή $\mathcal{N}(\theta+1, 1)$, με άγνωστο $\theta \in \mathbb{R}$. Να βρεθεί αποδοτικός εκτιμητής του $3\theta + 2$.

5.8. Έστω ότι $\underset{\sim}{X} = (X_1, X_2, \ldots, X_n)$, είναι ένα τυχαίο δείγμα από την κατανομή με πυκνότητα $f_1(x;\theta) = \dfrac{\theta}{x^{\theta+1}}$, $x > 1$, όπου $\theta > 0$ είναι άγνωστη παράμετρος. Να βρεθεί αποδοτικός εκτιμητής του $g(\theta) = \frac{\theta+1}{\theta}$.

5.9. Έστω $\underset{\sim}{X} = (X_1, X_2, \ldots, X_n)$, όπου X_i είναι ανεξάρτητες παρατηρήσεις με εκθετικές κατανομές $X_i \sim \mathcal{E}(i\theta^2)$, $i = 1, \ldots, n$, και άγνωστη παράμετρο $\theta > 0$. Να βρεθεί το κάτω φράγμα των Cramér-Rao για τη διασπορά των αμερόληπτων εκτιμητών του θ^2. Για ποιες συναρτήσεις g μπορούμε να βρούμε αποδοτικούς εκτιμητές του $g(\theta)$;

5.10. Δίνεται ένα τυχαίο δείγμα $\underset{\sim}{X} = (X_1, X_2, \ldots, X_n)$, από την κανονική κατανομή $\mathcal{N}(\theta, \theta)$, $\theta > 0$. Να βρεθεί αποδοτικός εκτιμητής για το $g(\theta) = \theta + \theta^2$.

5.11. Δίνεται ένα δείγμα $\underset{\sim}{X} = (X_1, X_2, \ldots, X_{n-1}, X_n)$, $n \geqslant 2$, αποτελούμενο από ανεξάρτητες παρατηρήσεις, όπου οι $X_1, X_2, \ldots, X_{n-1}$ προέρχονται από την κατανομή Poisson $\mathcal{P}(\theta)$ και η X_n από την $\mathcal{P}(n\theta)$.
α. Να δειχθεί ότι η κατανομή του $\underset{\sim}{X}$ ανήκει στην Μ.Ε.Ο.Κ. και να δοθούν οι συναρτήσεις του ορισμού της οικογένειας.
β. Να δειχθεί ότι η στατιστική συνάρτηση $T(\underset{\sim}{X}) = \frac{X_1+X_2+\ldots+X_{n-1}+X_n}{2n-1}$ είναι αποδοτικός εκτιμητής του θ.
γ. Υπάρχει αποδοτικός εκτιμητής του θ^2;

5.12. Δίνεται ένα τυχαίο δείγμα $\underset{\sim}{X} = (X_1, X_2, \ldots, X_n)$ από την κατανομή με πυκνότητα

$$\frac{1}{2\theta}e^{-\frac{|x|}{\theta}}, \qquad x \in R, \quad \theta > 0.$$

Να δειχθεί ότι η κατανομή του $\underset{\sim}{X}$ ανήκει στην Μ.Ε.Ο.Κ.και να βρεθεί το κάτω φράγμα των Cramér-Rao για τη διασπορά των αμερόληπτων εκτιμητών του θ. Να βρεθεί αποδοτικός εκτιμητής του $g(\theta) = \mathbb{E}_\theta|X_1|$.

5.13. (πληροφορία Fisher και αλλαγή παραμέτρου). Έστω $I(\theta)$ ο αριθμός πληροφορίας Fisher για το θ, που περιέχεται στο δείγμα $\underset{\sim}{X}$ με πυκνότητα $f(\underset{\sim}{x};\theta)$, $\theta \in \Theta$, όπου Θ είναι ανοικτό υποσύνολο του $\Re$. Έστω ακόμη h μία $1-1$ συνάρτηση ορισμένη στο Θ, παραγωγίσιμη και $h'(\theta) \neq 0$. Θέτουμε $\eta = h(\theta)$, $\theta \in \Theta$, οπότε $\theta = h^{-1}(\eta)$ και αντικαθιστούμε το θ στον τύπο της $f(\underset{\sim}{x};\theta)$ με $h^{-1}(\eta)$. Θέτουμε $f^*(\underset{\sim}{x};\eta) = f(\underset{\sim}{x};h^{-1}(\eta))$ και έτσι αλλάζουμε την παράμετρο της πυκνότητας του $\underset{\sim}{X}$ από θ σε η. Έστω $I^*(\eta)$ ο αριθμός πληροφορίας του Fisher για το η, που περιέχεται στο δείγμα $\underset{\sim}{X}$ με πυκνότητα $f^*(\underset{\sim}{x};\eta)$. Να δειχθεί ότι $I^*(\eta) = I(h^{-1}(\eta))(\frac{d}{d\eta}h^{-1}(\eta))^2$.

5.14. (Κ.Φ. C-R και αλλαγή παραμέτρου). Υπό τις προϋποθέσεις της Άσκησης 5.13, να δειχθεί ότι το Κ.Φ. C-R, για τη διασπορά αμερόληπτων εκτιμητών του $g(\theta)$ παραμένει αναλλοίωτο όταν η παράμετρος θ αντικατασταθεί με την παράμετρο η.

Βιβλιογραφία

1. Apostol, T.M. (1969). *Mathematical Analysis.* Addison Wesley; 3rd printing.

2. Bickel, P.J. and Doksum, K.A. (1977). *Mathematical Statistics, Basic Ideas and Selected Topics, Vol. 1.* Holden-Day

3. Billingsley P. (1995). *Probability and measure.* Wiley; 3rd edition.

4. Casella, G. and Berger, R.L. (2002). *Statistical Inference.* Duxbury Press; 2nd edition.

5. Cramér, H. (1946). *Mathematical Methods of Statistics.* Princeton University Press.

6. Darmois, G.(1945). Sur les lois limites de la dispersion de certaines estimations. *Rev. Inst. Internat. Statist.*, **13**, 9–15.

7. DeGroot, M.H. and Schervish, M.J. (2010). *Probability and Statistics.* Pearson Education; 4th edition.

8. Edgeworth, F.Y. (1908). On the probable errors of frequency constants. *J. Roy. Statist. Soc.*, **71**, 381–397.

9. Fisher, R.A. (1922). On the mathematical foundations of theoretical statistics. *Philos. Trans. Roy. Soc. London Ser. A*, **222**, 309–368.

10. Fisher, R.A. (1925). Theory of statistical estimation. *Proc. Camb. Phil. Soc.*, **22**, 700–725.

11. Fisher, R.A. (1934). Two new properties of mathematical likelihood. *Proc. Roy. Soc. Ser. A*, **144**, 285–307.

12. Frechét, M. (1943). Sur lextension de certaines evaluations statistiques au cas de petits echantillons. *Revue del Institut International de Statistique / Review of the International Statistical Institute*, **11**, 182–205.

13. Joshi, V.M. (1976). On the attainment of the Cramér-Rao lower bound. *Ann.Statist.*, **4**, 998–1002.

14. Kullback, S. (1997). *Information theory and Statistics.* Dover Publications, Inc.

15. Lehmann, E.L. and Casella, G. (1998). *Theory of point estimation.* Springer; 2nd edition.

16. Müller-Funk, U., Pukelsheim, F. and Witting, H. (1989). On the attainment of the Cramér-Rao bound in $\mathcal{L}_r$–differentiable families of distributions. *Ann. Statist.*, **17**, 1742–1748.

17. Rao, C. R. (1945). Information and the accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.*. **37**, 81–89.

18. Rao, C. R. (1973). *Linear Statistical Inference and its Applications.* Wiley: 2nd edition.

19. Rohatgi, V.K. (1976). *An introduction to probability theory and mathematical statistics.* Wiley

Κεφάλαιο 6

Επάρκεια, πληρότητα και ΑΟΕΔ εκτιμητές

Στο Κεφάλαιο 5 μελετήσαμε, μέσω της ανισότητας των Cramér–Rao, την ύπαρξη και εύρεση αποδοτικών εκτιμητών αγνώστων τιμών $g(\theta)$, δηλαδή εκτιμητών που, εξ ορισμού,

(α) είναι αμερόληπτοι,

(β) έχουν ελάχιστη διασπορά μεταξύ όλων των αμερόληπτων εκτιμητών του $g(\theta)$ (άρα και ελάχιστο ΜΤΣ λόγω του α) και

(γ) η (ελάχιστη) διασπορά τους συμπίπτει με το αντίστοιχο Κ.Φ. C-R.

Καταλήξαμε μάλιστα ότι στην περίπτωση πραγματικής παραμέτρου θ αποδοτικοί εκτιμητές υπάρχουν μόνον στις Μ.Ε.Ο.Κ. για την εκτίμηση τιμών $g(\theta)$, σε ειδικές περιπτώσεις συναρτήσεων g. Στο κεφάλαιο αυτό αναπτύσσεται μία γενική μεθοδολογία κατασκευής εκτιμητή του $g(\theta)$, για πραγματική ή διανυσματική παράμετρο θ, για δοθείσα αλλά αυθαίρετη πραγματική συνάρτηση g, ο οποίος ικανοποιεί τα (α) και (β), αλλά όχι κατ' ανάγκη το (γ). Ένας τέτοιος εκτιμητής, αν υπάρχει, ονομάζεται *αμερόληπτος εκτιμητής ελάχιστης διασποράς*. Αρχικά παρουσιάζεται η έννοια της επάρκειας μιας στατιστικής συνάρτησης και η χρήση της στη βελτίωση δοθέντος εκτιμητή, δηλαδή στη μείωση της διασποράς του και του ΜΤΣ του. Ακολούθως και σε συνδυασμό με την ιδιότητα της πληρότητας ένας βελτιωμένος αμερόληπτος εκτιμητής αναδεικνύεται ως αμερόληπτος εκτιμητής ελάχιστης διασποράς. Η μεθοδολογία αυτή αναπτύχθηκε από τους Lehmann and Scheffé (1955).

6.1 Επάρκεια

Η έννοια της επάρκειας είναι μία από τις πιο σημαντικές έννοιες της Στατιστικής Συμπερασματολογίας. Εισήχθηκε από τον Fisher στις αρχές του 20ου αιώνα και χαρακτηρίζει εκείνες τις στατιστικές συναρτήσεις $T(\underset{\sim}{X})$ που έχουν την ιδιότητα να περιέχουν όλες τις πληροφορίες για την άγνωστη παράμετρο θ (και κατ' επέκταση για το $g(\theta)$) που περιέχει και το δείγμα $\underset{\sim}{X} = (X_1, \ldots, X_n)$. Η επάρκεια συνεπάγεται ότι γνώση της παρατηρηθείσης τιμής $\underset{\sim}{x}$ του $\underset{\sim}{X}$ δεν προσφέρει τίποτε περισσότερο, όσον αφορά την εξαγωγή συμπερασμάτων για το θ ή γενικότερα για το $g(\theta)$, απ' ό,τι προσφέρει η γνώση, μόνον, της τιμής $T(\underset{\sim}{x})$ της επαρκούς στατιστικής συνάρτησης $T(\underset{\sim}{X})$. Επιπλέον, η επαρκής στατιστική συνάρτηση $T(\underset{\sim}{X})$ παρουσιάζει το σημαντικό πλεονέκτημα έναντι του $\underset{\sim}{X}$ ότι συνήθως έχει πολύ μικρότερη διάσταση από τη διάσταση n του $\underset{\sim}{X} = (X_1, \ldots, X_n)$. Με αυτήν την έννοια, η επάρκεια επιφέρει επιθυμητή *σύμπτυξη των προς ανάλυση δεδομένων χωρίς απώλεια πληροφορίας* για το θ. Για παράδειγμα, εάν $\underset{\sim}{X} = (X_1, \ldots, X_n)$ είναι ένα τυχαίο δείγμα από την κατανομή Bernoulli $\mathcal{B}(1, \theta)$, $\theta \in \Theta = (0, 1)$, τότε είναι ευνόητο ότι για να εκτιμήσουμε την πιθανότητα «επιτυχίας» θ, η γνώση των τιμών $x_1, \ldots, x_n$ των $X_1, \ldots, X_n$, δηλαδή ποιες δοκιμές Bernoulli κατέληξαν σε «επιτυχία» και ποιες σε «αποτυχία» δεν προσφέρει τίποτε περισσότερο απ' ό,τι προσφέρει η γνώση της τιμής $T(\underset{\sim}{x}) = \sum_{i=1}^{n} x_i$, *πόσες* δηλαδή δοκιμές Bernoulli κατέληξαν σε «επιτυχία». Έτσι, η στατιστική συνάρτηση $T(\underset{\sim}{X}) = \sum_{i=1}^{n} X_i$ είναι επαρκής και εάν π.χ. $n = 3$ και οι παρατηρηθείσες τιμές των X_1, X_2, X_3 είναι αντίστοιχα $x_1 = 1, x_2 = 1, x_3 = 0$, τα δεδομένα $(1, 1, 0)$ διάστασης τρία συμπτύσσονται λόγω επάρκειας στη τιμή $T(\underset{\sim}{x}) = \sum_{i=1}^{3} x_i = 2$ (διάστασης ένα) χωρίς απώλεια πληροφορίας για το θ.

Στην επόμενη ενότητα αυτού του κεφαλαίου θα δούμε μία άμεση εφαρμογή της επάρκειας (Θεώρημα των Rao–Blackwell). Δίνουμε τώρα τον αυστηρό ορισμό επαρκούς στατιστικής συνάρτησης.

Ορισμός 6.1.1. *Η στατιστική συνάρτηση $T(\underset{\sim}{X})$ ονομάζεται επαρκής ή επαρκής για το $\theta \in \Theta$ εάν η δεσμευμένη κατανομή του $\underset{\sim}{X}$ δοθέντος ότι*

$T(\underset{\sim}{X}) = t$ δεν εξαρτάται από το θ (είναι δηλαδή σταθερή ως προς θ), για κάθε τιμή t της $T(\underset{\sim}{X})$.

Η ουσία του ορισμού είναι ότι άπαξ και δοθεί η τιμή t της επαρκούς στατιστικής συνάρτησης $T(\underset{\sim}{X})$, το δείγμα $\underset{\sim}{X}$ δεν περιέχει καμία περαιτέρω πληροφορία για το θ αφού η κατανομή του είναι η ίδια, όποια και αν είναι η άγνωστη τιμή τού θ. Συνεπώς, όλες οι πληροφορίες για το θ περιέχονται στην επαρκή στατιστική συνάρτηση $T(\underset{\sim}{X})$. Στην περίπτωση που το $\underset{\sim}{X}$ έχει διακριτή κατανομή, η δεσμευμένη κατανομή του $\underset{\sim}{X}$ δοθέντος ότι $T(\underset{\sim}{X}) = t$, καθορίζεται από τη δεσμευμένη πυκνότητα (πιθανότητα)$f_{\underset{\sim}{X}|T}(\underset{\sim}{x} \mid t) = \mathbb{P}_\theta(\underset{\sim}{X} = \underset{\sim}{x} \mid T(\underset{\sim}{X}) = t)$ για t τέτοιο ώστε $\mathbb{P}_\theta(T(\underset{\sim}{X}) = t) > 0$. Επομένως,

$$\begin{aligned}
f_{\underset{\sim}{X}|T}(\underset{\sim}{x} \mid t) &= \frac{\mathbb{P}_\theta(\underset{\sim}{X} = \underset{\sim}{x}, T(\underset{\sim}{X}) = t)}{\mathbb{P}_\theta(T(\underset{\sim}{X}) = t)} \\
&= \begin{cases} 0, & \text{αν } T(\underset{\sim}{x}) \neq t \\ \dfrac{\mathbb{P}_\theta(\underset{\sim}{X} = \underset{\sim}{x})}{\mathbb{P}_\theta(T(\underset{\sim}{X}) = t)}, & \text{αν } T(\underset{\sim}{x}) = t \end{cases} \\
&= \begin{cases} 0, & \text{αν } T(\underset{\sim}{x}) \neq t \\ \dfrac{f(\underset{\sim}{x};\theta)}{f_T(t;\theta)}, & \text{αν } T(\underset{\sim}{x}) = t. \end{cases}
\end{aligned}$$

Βάσει λοιπόν του ορισμού, στη διακριτή περίπτωση, επάρκεια της $T(\underset{\sim}{X})$ σημαίνει ότι ο λόγος της πυκνότητας του δείγματος $\underset{\sim}{X}$ προς την πυκνότητα της $T(\underset{\sim}{X})$, $\frac{f(\underset{\sim}{x};\theta)}{f_T(t;\theta)}$, δεν πρέπει να εξαρτάται από το θ, για κάθε $\underset{\sim}{x}$ και t τέτοια ώστε $T(\underset{\sim}{x}) = t$. Στη συνεχή περίπτωση, υπάρχει δυσκολία να οριστεί η δεσμευμένη πυκνότητα του $\underset{\sim}{X}$ δοθέντος ότι $T(\underset{\sim}{X}) = t$, $f_{\underset{\sim}{X}|T}(\underset{\sim}{x} \mid t)$, επειδή η $T(\underset{\sim}{X})$ είναι συνάρτηση του $\underset{\sim}{X}$. Παρακινούμενοι όμως από τον γενικό ορισμό της δεσμευμένης πυκνότητας, ας θέσουμε $f_{\underset{\sim}{X}|T}(\underset{\sim}{x} \mid t) = \frac{f_{\underset{\sim}{X},T}(\underset{\sim}{x},t;\theta)}{f_T(t;\theta)}$, όπου ο αριθμητής είναι η από κοινού πυκνότητα των $\underset{\sim}{X}$ και $T(\underset{\sim}{X})$ και ο παρανομαστής η πυκνότητα της $T(\underset{\sim}{X})$, για t τέτοιο ώστε $f_T(t;\theta) > 0$. Επειδή όμως η $T(\underset{\sim}{X})$ είναι συνάρτηση του $\underset{\sim}{X}$, η κατανομή πιθανότητας στις διάφορες τιμές $\underset{\sim}{x}$ του $\underset{\sim}{X}$ συμπίπτει με την κατανομή πιθανότητας στις διάφορες τιμές $(\underset{\sim}{x}, T(\underset{\sim}{x}))$ του $(\underset{\sim}{X}, T(\underset{\sim}{X}))$, αφού όταν $\underset{\sim}{X} = \underset{\sim}{x}$, τότε $\underset{\sim}{X} = \underset{\sim}{x}$

και $T(\underset{\sim}{X}) = T(\underset{\sim}{x})$ και αντίστροφα. Έτσι ερμηνεύοντας την από κοινού κατανομή πιθανότητας των $\underset{\sim}{X}$ και $T(\underset{\sim}{X})$, θέτουμε

$$f_{\underset{\sim}{X},T}(\underset{\sim}{x}, t; \theta) = \begin{cases} f(\underset{\sim}{x}; \theta), & \text{αν } T(\underset{\sim}{x}) = t \\ 0, & \text{αν } T(\underset{\sim}{x}) \neq t. \end{cases}$$

Ως εκ τούτου, ορίζουμε

$$f_{\underset{\sim}{X}|T}(\underset{\sim}{x} \mid t) = \begin{cases} \frac{f(\underset{\sim}{x};\theta)}{f_T(t;\theta)}, & \text{αν } T(\underset{\sim}{x}) = t \\ 0, & \text{αν } T(\underset{\sim}{x}) \neq t, \end{cases}$$

και αυτός ο λόγος δεν πρέπει να εξαρτάται από το θ, όπως δηλαδή και στη διακριτή περίπτωση, προκειμένου η $T(\underset{\sim}{X})$ να είναι επαρκής. Για την καλύτερη κατανόηση του ορισμού της επάρκειας αναφέρουμε τα εξής παράδειγματα.

Παράδειγμα 6.1.1. (Mood, Graybill and Boes, 1974, σελ. 302, κατανομή Bernoulli–επάρκεια) Έστω $\underset{\sim}{X} = (X_1, X_2, X_3)$ ένα τυχαίο δείγμα μεγέθους $n = 3$ από την κατανομή Bernoulli $\mathcal{B}(1, \theta)$, $\theta \in \Theta = (0, 1)$. Τότε η στατιστική συνάρτηση $T(\underset{\sim}{X}) = \sum_{i=1}^{3} X_i$ έχει σύνολο τιμών το $\{0, 1, 2, 3\}$ και είναι επαρκής γιατί όπως φαίνεται από τον Πίνακα 6.1 για κάθε τιμή $t \in \{0, 1, 2, 3\}$ της $T(\underset{\sim}{X})$ η δεσμευμένη κατανομή του $\underset{\sim}{X} = (X_1, X_2, X_3)$ είναι σταθερή ως προς θ. Προς διευκρίνιση, αν π.χ., $T(\underset{\sim}{X}) = 1$ δηλαδή $\sum_{i=1}^{3} X_i = 1$, τότε οι τιμές του $\underset{\sim}{X} = (X_1, X_2, X_3)$ είναι $(1, 0, 0)$ ή $(0, 1, 0)$ ή $(0, 0, 1)$ και κάθε μία, λόγω συμμετρίας, έχει πιθανότητα $\frac{1}{3}$, ανεξάρτητα από την τιμή του θ.

Παράδειγμα 6.1.2. (εκθετική κατανομή–επάρκεια) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την εκθετική κατανομή $\mathcal{E}(\theta)$ με πυκνότητα $f_1(x; \theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}}$, $x > 0$, $\theta \in \Theta = (0, \infty)$. Θα δείξουμε ότι η $T(\underset{\sim}{X}) = \sum_{i=1}^{n} X_i$ είναι επαρκής. Η πυκνότητα του $\underset{\sim}{X}$ είναι

$$f(\underset{\sim}{x}; \theta) = \prod_{i=1}^{n} f_1(x_i; \theta) = \frac{1}{\theta^n} e^{-\frac{1}{\theta} \sum_{i=1}^{n} x_i}, \quad \underset{\sim}{x} = (x_1, \ldots, x_n) \quad \text{με} \quad x_i > 0.$$

τιμές του $T=\sum_{i=1}^{3}X_i$		τιμές του $\underset{\sim}{X}=(X_1,X_2,X_3)$	δεσμευμένες πιθανότητες
	$t=0$	$(0,0,0)$	1
	$t=1$	$(1,0,0)$	$1/3$
		$(0,1,0)$	$1/3$
		$(0,0,1)$	$1/3$
	$t=2$	$(1,1,0)$	$1/3$
		$(0,1,1)$	$1/3$
		$(1,0,1)$	$1/3$
	$t=3$	$(1,1,1)$	1

Πίνακας 6.1: Δεσμευμένη κατανομή του $\underset{\sim}{X}$, $\mathbb{P}_\theta(\underset{\sim}{X}=\underset{\sim}{x}\mid T=t)$

Η $T(\underset{\sim}{X})$ έχει κατανομή Γάμμα $\mathcal{G}(n,\theta)$ με πυκνότητα

$$f_T(t;\theta)=\frac{1}{\Gamma(n)\theta^n}t^{n-1}e^{-\frac{t}{\theta}},\quad t>0.$$

Για $\underset{\sim}{x}$ και t τέτοια ώστε $T(\underset{\sim}{x})=t$, δηλαδή $\sum\limits_{i=1}^{n}x_i=t$, ο λόγος $\frac{f(\underset{\sim}{x};\theta)}{f_T(t;\theta)}=\frac{\Gamma(n)}{t^{n-1}}$ δεν εξαρτάται από το θ και συνεπώς, από τη συζήτηση που προηγήθηκε του Παραδείγματος 6.1.1, η $T(\underset{\sim}{X})$ είναι επαρκής. Έχει ενδιαφέρον να σημειώσουμε ότι αυτή η δεσμευμένη κατανομή του $\underset{\sim}{X}=(X_1,\ldots,X_n)$, αν και είναι $n-$διάστατη, «ζει» στο υπερεπίπεδο του $\mathbb{R}^n$, $H=\{(x_1,\ldots,x_n)\in\mathbb{R}^n:\ \sum\limits_{i=1}^{n}x_i=t\}$, δηλαδή σε έναν χώρο διάστασης $n-1$. □

Γενικά, η απόδειξη της επάρκειας μιας στατιστικής συνάρτησης από τον ορισμό παρουσιάζει δυσκολίες, ειδικά στις συνεχείς κατανομές, επειδή απαιτείται ο υπολογισμός της δεσμευμένης κατανομής, προκειμένου να διαπιστωθεί η μη εξάρτησή της από την άγνωστη παράμετρο θ. Επίσης μια άλλη δυσκολία είναι ότι πριν εφαρμόσουμε τον ορισμό, πρέπει πρώτα να «μαντέψουμε» ποια στατιστική συνάρτηση είναι υποψήφια για επαρκής, κάτι που γενικά δεν είναι εύκολο. Οι δύο αυτές δυσκολίες μπορούν να ξεπεραστούν με την εφαρμογή μιας απλής ικανής και αναγκαίας συνθήκης που συνήθως αναφέρεται στη βιβλιογραφία ως *παραγοντικό κριτήριο των*

Neyman-Fisher. Αναπτύχθηκε σταδιακά από τους Fisher (1922), Neyman (1935) και Halmos and Savage (1949).

Θεώρημα 6.1.1. (**παραγοντικό κριτήριο των Neyman–Fisher**) *Έστω ότι το δείγμα $\underset{\sim}{X}$ έχει πυκνότητα $f(\underset{\sim}{x};\theta)$, $\theta \in \Theta$. Τότε η στατιστική συνάρτηση $T(\underset{\sim}{X})$ είναι επαρκής εάν και μόνον εάν υπάρχουν μη αρνητικές συναρτήσεις q και h, με την h να μην εξαρτάται από το θ, έτσι ώστε*

$$f(\underset{\sim}{x};\theta) = q\big(T(\underset{\sim}{x}),\theta\big) \cdot h(\underset{\sim}{x}), \quad \forall\, \underset{\sim}{x},\ \forall\, \theta. \tag{6.1}$$

Απόδειξη. Μια αυστηρή απόδειξη απαιτεί γνώσεις Θεωρίας Μέτρου, βλέπε Lehmann and Romano (2005, Ενότητα 2.6). Δίνουμε την απόδειξη για διακριτό $\underset{\sim}{X}$, οπότε $f(\underset{\sim}{x};\theta) = \mathbb{P}_\theta(\underset{\sim}{X} = \underset{\sim}{x})$ και η $T(\underset{\sim}{X})$ έχει επίσης διακριτή κατανομή. Έστω ότι ισχύει η (6.1). Ας συμβολίσουμε με $\mathcal{S}$ το σύνολο τιμών του $\underset{\sim}{X}$. Η δεσμευμένη κατανομή του $\underset{\sim}{X}$ δοθέντος ότι $T(\underset{\sim}{X}) = t$ μπορεί να οριστεί για κάθε t με $\mathbb{P}_\theta(T(\underset{\sim}{X}) = t) > 0$ και εκφράζεται από τη δεσμευμένη πιθανότητα

$$\mathbb{P}_\theta(\underset{\sim}{X} = \underset{\sim}{x} \mid T(\underset{\sim}{X}) = t) = \frac{\mathbb{P}_\theta(\underset{\sim}{X} = \underset{\sim}{x}, T(\underset{\sim}{X}) = t)}{\mathbb{P}_\theta(T(\underset{\sim}{X}) = t)}. \tag{6.2}$$

Έστω, λοιπόν ένα τέτοιο t και $\theta \in \Theta$. Το ενδεχόμενο $\mathcal{S}_t = (T(\underset{\sim}{X}) = t)$ μπορούμε να το περιγράψουμε ως εξής. Έχουμε $\mathcal{S}_t = (T(\underset{\sim}{X}) = t) = \big\{\underset{\sim}{x} \in \mathcal{S} : T(\underset{\sim}{x}) = t\big\}$ το οποίο είναι πεπερασμένο ή αριθμήσιμο και επομένως

$$\begin{aligned}\mathbb{P}_\theta(T(\underset{\sim}{X}) = t) &= \sum_{\underset{\sim}{x} \in S_t} f(\underset{\sim}{x};\theta) = \sum_{\underset{\sim}{x} \in S_t} \mathbb{P}_\theta(\underset{\sim}{X} = \underset{\sim}{x}) = \sum_{\underset{\sim}{x} \in S_t} q\big(T(\underset{\sim}{x}),\theta\big) h(\underset{\sim}{x}) \\ &= \sum_{\underset{\sim}{x} \in S_t} q(t,\theta) h(\underset{\sim}{x}) = q(t,\theta) \sum_{\underset{\sim}{x} \in S_t} h(\underset{\sim}{x}).\end{aligned} \tag{6.3}$$

Συνεπώς, για κάθε $\underset{\sim}{x}$ από την (6.2) παίρνουμε

$$\begin{aligned}
\mathbb{P}_\theta(\underset{\sim}{X} = \underset{\sim}{x} \mid T(\underset{\sim}{X}) = t) &= \begin{cases} 0 & , \text{εάν } \underset{\sim}{x} \notin \mathcal{S}_t \\ \dfrac{\mathbb{P}_\theta(\underset{\sim}{X} = \underset{\sim}{x})}{\mathbb{P}_\theta(T(\underset{\sim}{X}) = t)} & , \text{εάν } \underset{\sim}{x} \in \mathcal{S}_t \end{cases} \\
&= \begin{cases} 0 & , \text{εάν } \underset{\sim}{x} \notin \mathcal{S}_t \\ \dfrac{f(\underset{\sim}{x};\theta)}{q(t,\theta)\displaystyle\sum_{\underset{\sim}{x}\in S_t} h(\underset{\sim}{x})} & , \text{εάν } \underset{\sim}{x} \in \mathcal{S}_t \end{cases} \\
&= \begin{cases} 0 & , \text{εάν } \underset{\sim}{x} \notin \mathcal{S}_t \\ \dfrac{q\big(T(\underset{\sim}{x}),\theta\big)h(\underset{\sim}{x})}{q(t,\theta)\displaystyle\sum_{\underset{\sim}{x}\in S_t} h(\underset{\sim}{x})} & , \text{εάν } \underset{\sim}{x} \in \mathcal{S}_t \end{cases} \\
&= \begin{cases} 0 & , \text{εάν } \underset{\sim}{x} \notin \mathcal{S}_t \\ \dfrac{q(t,\theta)h(\underset{\sim}{x})}{q(t,\theta)\displaystyle\sum_{\underset{\sim}{x}\in S_t} h(\underset{\sim}{x})} & , \text{εάν } \underset{\sim}{x} \in \mathcal{S}_t \end{cases} \\
&= \begin{cases} 0 & , \text{εάν } \underset{\sim}{x} \notin \mathcal{S}_t \\ \dfrac{h(\underset{\sim}{x})}{\displaystyle\sum_{\underset{\sim}{x}\in S_t} h(\underset{\sim}{x})} & , \text{εάν } \underset{\sim}{x} \in \mathcal{S}_t. \end{cases}
\end{aligned}$$

Επομένως, η δεσμευμένη κατανομή του $\underset{\sim}{X}$ δοθέντος ότι $T(\underset{\sim}{X}) = t$ δεν περιέχει το θ και συνεπώς, από τον ορισμό, η στατιστική συνάρτηση $T(\underset{\sim}{X})$ είναι επαρκής.

Αντίστροφα, έστω ότι η $T(\underset{\sim}{X})$ είναι επαρκής. Τότε, θέτοντας $q\big(T(\underset{\sim}{x}),\theta\big) = \mathbb{P}_\theta\big(T(\underset{\sim}{X}) = T(\underset{\sim}{x})\big)$ και $h(\underset{\sim}{x}) = \mathbb{P}\big(\underset{\sim}{X} = \underset{\sim}{x} \mid T(\underset{\sim}{X}) = T(\underset{\sim}{x})\big)$ που δεν εξαρτάται από το θ λόγω επάρκειας, έχουμε

$$\begin{aligned}
f(\underset{\sim}{x};\theta) = \mathbb{P}_\theta(\underset{\sim}{X} = \underset{\sim}{x}) &= \mathbb{P}_\theta\big(\underset{\sim}{X} = \underset{\sim}{x}, T(\underset{\sim}{X}) = T(\underset{\sim}{x})\big) \\
&= \mathbb{P}_\theta\big(T(\underset{\sim}{X}) = T(\underset{\sim}{x})\big) \cdot \mathbb{P}\big(\underset{\sim}{X} = \underset{\sim}{x} \mid T(\underset{\sim}{X}) = T(\underset{\sim}{x})\big) \\
&= q\big(T(\underset{\sim}{x}),\theta\big)h(\underset{\sim}{x}).
\end{aligned}$$

Άρα, η (6.1) ισχύει. □

Το παραγοντικό κριτήριο μας κατευθύνει συγχρόνως προς την αναγνώριση μιας επαρκούς στατιστικής συνάρτησης και την απόδειξη της επάρκειας της. Παραγοντοποιώντας την πυκνότητα του δείγματος $\underset{\sim}{X}$, $f(\underset{\sim}{x};\theta)$, προσπαθούμε να ενσωματώσουμε σε μια συνάρτηση $h(\underset{\sim}{x})$ όρους που δεν περιέχουν την άγνωστη παράμετρο θ και εξαρτώνται μόνον από το $\underset{\sim}{x}$. «Ό,τι απομένει» μετά την ενσωμάτωση, δηλαδή $\dfrac{f(\underset{\sim}{x};\theta)}{h(\underset{\sim}{x})}$, είναι ακριβώς ο όρος $q\big(T(\underset{\sim}{x}),\theta\big)$ στη σχέση (6.1), ο οποίος εξαρτάται από το θ και έμμεσα από το $\underset{\sim}{x}$ μέσω κάποιας τιμής, έστω, $T(\underset{\sim}{x})$. Αυτή η τιμή ταυτοποιεί τη στατιστική συνάρτηση $T(\underset{\sim}{X})$ ως επαρκή. Οι σταθερές, αν υπάρχουν, μη εξαρτώμενες από το θ ή το $\underset{\sim}{x}$, ενσωματώνονται αδιακρίτως στην $h(\underset{\sim}{x})$ ή στην $q\big(T(\underset{\sim}{x}),\theta\big)$. Οι συναρτήσεις q και h δεν είναι μοναδικές (ούτε εξ άλλου απαιτείται αυτό στο Θεώρημα 6.1.1), για παράδειγμα η (6.1) ισχύει αν η q αντικατασταθεί με την $2q$ και η h με την $\dfrac{h}{2}$.
Το παραγοντικό κριτήριο ισχύει χωρίς κανένα περιορισμό για την πυκνότητα του διακριτού ή συνεχούς δείγματος $\underset{\sim}{X}$, $f(\underset{\sim}{x};\theta)$, το θ μπορεί να είναι πραγματική παράμετρος ή διανυσματική παράμετρος, $\theta=(\theta_1,\dots,\theta_\kappa)$, ακόμη και μη Ευκλείδια παράμετρος, ενώ η $T(\underset{\sim}{X})$ μπορεί να είναι πραγματική συνάρτηση ή διανυσματική συνάρτηση, $T(\underset{\sim}{X})=(T_1(\underset{\sim}{X}),\dots,T_m(\underset{\sim}{X}))$. Στην τελευταία περίπτωση, συνήθως, $m=\kappa$, εξαιρέσεις όμως υπάρχουν αρκετές κάποιες εκ των οποίων θα δούμε σε παραδείγματα.

Παράδειγμα 6.1.3. (**κανονική κατανομή - επάρκεια**) Έστω $\underset{\sim}{X}=(X_1,\dots,X_n)$ ένα τυχαίο δείγμα από την κανονική κατανομή $\mathcal{N}(\mu,\sigma^2)$. Διακρίνουμε τρεις περιπτώσεις.
<u>1η Περίπτωση</u>: $\boldsymbol{\sigma^2}$ **γνωστό,** $\boldsymbol{\mu=\theta}$ **άγνωστο**, $\theta\in\Theta=\mathbb{R}$.

Τότε έχουμε,

$$\begin{aligned} f(\underset{\sim}{x};\theta) &= \prod_{i=1}^{n} f_1(x_i;\theta) = \prod_{i=1}^{n} \tfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x_i-\theta)^2} \\ &= \tfrac{1}{\sigma^n(2\pi)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\theta)^2} \\ &= \tfrac{1}{\sigma^n(2\pi)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2 + \frac{\theta}{\sigma^2}\sum_{i=1}^n x_i - \frac{n\theta^2}{2\sigma^2}} \\ &= q\big(\sum_{i=1}^{n} x_i, \theta\big) h(\underset{\sim}{x}), \quad \forall\, \underset{\sim}{x}, \quad \forall\, \theta, \end{aligned}$$

όπου $q\left(\sum_{i=1}^n x_i, \theta\right) = e^{\frac{\theta}{\sigma^2}\sum_{i=1}^n x_i - \frac{n\theta^2}{2\sigma^2}}$ και $h(\underset{\sim}{x}) = \frac{1}{\sigma^n(2\pi)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2}$.
Επομένως, η στατιστική συνάρτηση $T(\underset{\sim}{X}) = \sum\limits_{i=1}^{n} X_i$ είναι επαρκής.
Εκ των υστέρων κρίνοντας, είναι λογικό ότι προέκυψε η $T(\underset{\sim}{X})$ ως επαρκής στατιστική συνάρτηση. Το θ είναι η μέση τιμή της κοινής κατανομής των X_i και συνεπώς μπορεί να εκτιμηθεί με το δειγματικό του ανάλογο, τον δειγματικό μέσο $\overline{X} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i = \frac{1}{n} T(\underset{\sim}{X})$, βλέπε Ενότητα 3.3.1α, που μάλιστα είναι και αποδοτικός εκτιμητής (Παράδειγμα 5.2.11). Βλέπουμε λοιπόν ότι προκειμένου να εκτιμήσουμε το θ δεν είναι απαραίτητο να γνωρίζουμε τις τιμές των παρατηρήσεων $X_1, \ldots, X_n$, αρκεί να καταγραφεί η τιμή του αθροίσματός τους, $T(\underset{\sim}{X})$. Διαισθητικά, αυτό δικαιολογεί την επάρκεια της $T(\underset{\sim}{X})$. Από πρακτικής σκοπιάς, λόγω επάρκειας, δεν χρειάζεται να «αποθηκεύσουμε» τις παρατηρηθείσες τιμές των X_i και μετά να τις προσθέσουμε ώστε να εκτιμήσουμε το θ. Αρκεί κάθε παρατηρηθείσα τιμή να προστίθεται κατ΄ ευθείαν στο μερικό άθροισμα των προηγούμενων τιμών.

2η Περίπτωση: **$\boldsymbol{\mu}$ γνωστό, $\boldsymbol{\sigma^2 = \theta}$ άγνωστο**, $\theta \in \Theta = (0, \infty)$.
Τότε έχουμε,

$$\begin{aligned} f(\underset{\sim}{x};\theta) = \prod_{i=1}^{n} f_1(x_i;\theta) &= \tfrac{1}{\theta^{n/2}(2\pi)^{n/2}} e^{-\frac{1}{2\theta}\sum_{i=1}^n (x_i-\mu)^2} \\ &= q\left(\sum_{i=1}^{n} (x_i-\mu)^2, \theta\right) h(\underset{\sim}{x}), \quad \forall\, \underset{\sim}{x}, \quad \forall\, \theta, \end{aligned}$$

όπου $q\left(\sum\limits_{i=1}^{n}(x_i-\mu)^2,\theta\right) = \frac{1}{\theta^{n/2}(2\pi)^{n/2}}e^{-\frac{1}{2\theta}\sum_{i=1}^{n}(x_i-\mu)^2}$ και $h(\underset{\sim}{x}) = 1$. Επομένως, η στατιστική συνάρτηση $T_1(\underset{\sim}{X}) = \sum\limits_{i=1}^{n}(X_i-\mu)^2$ είναι επαρκής. Στην περίπτωση αυτή, ως εκτιμητής του θ γνωρίζουμε ότι μπορεί να ληφθεί η στατιστική συνάρτηση $\frac{1}{n}\sum\limits_{i=1}^{n}(X_i-\mu)^2 = \frac{1}{n}T_1(\underset{\sim}{X})$, βλέπε Ενότητα 3.3.1γ, γεγονός που δικαιολογεί την επάρκεια της $T_1(\underset{\sim}{X})$. Σημειώνουμε επίσης ότι στην ίδια επαρκή στατιστική συνάρτηση θα καταλήξουμε αν θεωρήσουμε ως άγνωστη παράμετρο την τυπική απόκλιση σ (αντί της διασποράς σ^2).

3η Περίπτωση: $\boldsymbol{\mu, \sigma^2}$ **άγνωστα**, οπότε $\theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times (0, \infty)$. Τότε έχουμε,

$$\begin{aligned} f(\underset{\sim}{x};\theta) &= \prod_{i=1}^{n} f_1(x_i;\theta) = \prod_{i=1}^{n} \frac{1}{\sigma(2\pi)^{1/2}}e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2} \\ &= \frac{1}{\sigma^n(2\pi)^{n/2}}e^{\frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i^2+\frac{\mu}{\sigma^2}\sum_{i=1}^{n}x_i-\frac{n\mu^2}{2\sigma^2}} \\ &= q\Big(\sum_{i=1}^{n}x_i, \sum_{i=1}^{n}x_i^2, \mu, \sigma^2\Big)h(\underset{\sim}{x}), \quad \forall\, \underset{\sim}{x}, \quad \forall\, \theta, \end{aligned}$$

όπου $q\big(\sum\limits_{i=1}^{n}x_i, \sum\limits_{i=1}^{n}x_i^2, \mu, \sigma^2\big) = \frac{1}{\sigma^n(2\pi)^{n/2}}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i^2+\frac{\mu}{\sigma^2}\sum_{i=1}^{n}x_i-\frac{n\mu^2}{2\sigma^2}}$ και $h(\underset{\sim}{x}) = 1$. Επομένως, η στατιστική συνάρτηση $T_2(\underset{\sim}{X}) = (\sum\limits_{i=1}^{n}X_i, \sum\limits_{i=1}^{n}X_i^2) = \Big(T(\underset{\sim}{X}), T_3(\underset{\sim}{X})\Big)$ είναι επαρκής. Παρατηρούμε ότι η επαρκής στατιστική συνάρτηση έχει διάσταση 2, όπως και η παράμετρος θ. Σε αυτήν την περίπτωση λοιπόν αρκεί να γνωρίζουμε την $T_2(\underset{\sim}{X})$ ή ισοδύναμα τις $T(\underset{\sim}{X})$ και $T_3(\underset{\sim}{X})$ για να εκτιμήσουμε τα άγνωστα μ και σ^2. Πράγματι το μ μπορεί να εκτιμηθεί με $\overline{X} = \frac{1}{n}T(\underset{\sim}{X})$ και το σ^2 με εκτιμητή της μορφής

$$c\sum_{i=1}^{n}(X_i-\overline{X})^2 = c\left\{\sum_{i=1}^{n}X_i^2 - n\overline{X}^2\right\} = c\left\{T_3(\underset{\sim}{X}) - \frac{T^2(\underset{\sim}{X})}{n}\right\},$$

όπου c θετική σταθερά (με κλασικές επιλογές τις $c = \frac{1}{n-1}, c = \frac{1}{n}, c = \frac{1}{n+1}$ και μη κλασικές τις $c = c_1, c = c_2$ της Ενότητας 4.3). □

Στην περίπτωση τυχαίου δείγματος, όταν το κοινό σύνολο τιμών των παρατηρήσεων X_i, $S_1 = \{x \in \mathbb{R} : f_1(x;\theta) > 0\}$, είναι γνήσιο υποσύνολο του $\mathbb{R}$, η εφαρμογή του παραγοντικού κριτηρίου διευκολύνεται εισάγοντας τη δείκτρια συνάρτηση $\mathrm{I}_{S_1}(x) = \begin{cases} 1, & x \in S_1 \\ 0, & x \notin S_1 \end{cases}$. Ισχύει, τότε, $f_1(x;\theta) = f_1(x;\theta)\mathrm{I}_{S_1}(x)$, $\forall\, x \in \mathbb{R}$.

Παράδειγμα 6.1.4. (Κατανομή Poisson - επάρκεια) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την κατανομή Poisson $\mathcal{P}(\theta)$, με πυκνότητα $f_1(x;\theta) = e^{-\theta}\frac{\theta^x}{x!}$, $x \in \{0,1,2,\ldots\}$ ή $f_1(x;\theta) = e^{-\theta}\frac{\theta^x}{x!}\mathrm{I}_{\{0,1,2,\ldots\}}(x)$, $\theta \in \Theta = (0,\infty)$. Έχουμε

$$\begin{aligned} f(\underset{\sim}{x};\theta) &= \prod_{i=1}^{n} f_1(x_i;\theta) = \prod_{i=1}^{n} e^{-\theta}\frac{\theta^{x_i}}{x_i!}\mathrm{I}_{\{0,1,2,\ldots\}}(x_i) \\ &= e^{-n\theta}\theta^{\sum_{i=1}^n x_i}\prod_{i=1}^{n}\frac{1}{x_i!}\mathrm{I}_{\{0,1,2,\ldots\}}(x_i) \\ &= q\Big(\sum_{i=1}^{n} x_i, \theta\Big)h(\underset{\sim}{x}), \quad \forall\, \underset{\sim}{x},\ \forall\, \theta, \end{aligned}$$

όπου $q(\sum\limits_{i=1}^{n} x_i, \theta) = e^{-n\theta}\theta^{\sum_{i=1}^n x_i}$ και $h(\underset{\sim}{x}) = \prod_{i=1}^{n}\frac{1}{x_i!}\mathrm{I}_{\{0,1,2,\ldots\}}(x_i)$. Επομένως, η στατιστική συνάρτηση $T(\underset{\sim}{X}) = \sum\limits_{i=1}^{n} X_i$ είναι επαρκής.

Παράδειγμα 6.1.5. (Ομοιόμορφη κατανομή με ένα άκρο άγνωστο - επάρκεια) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την ομοιόμορφη κατανομή $\mathcal{U}(0,\theta)$ με πυκνότητα $f_1(x;\theta) = \frac{1}{\theta}\mathrm{I}_{(0,\theta)}(x)$, $\theta \in \Theta = (0,\infty)$. Τότε έχουμε

$$f(\underset{\sim}{x};\theta) = \prod_{i=1}^{n} f_1(x_i;\theta) = \prod_{i=1}^{n}\frac{1}{\theta}\mathrm{I}_{(0,\theta)}(x_i) = \frac{1}{\theta^n}\prod_{i=1}^{n}\mathrm{I}_{(0,\theta)}(x_i).$$

Παρατηρούμε τώρα ότι

$$\begin{aligned} \prod_{i=1}^{n}\mathrm{I}_{(0,\theta)}(x_i) &= \begin{cases} 1, & 0 < x_i < \theta,\ \forall\, i = 1,\ldots,n \\ 0, & \text{διαφορετικά} \end{cases} \\ &= \begin{cases} 1, & 0 < x_{(1)} \le x_{(n)} < \theta,\ \forall\, i = 1,\ldots,n \\ 0, & \text{διαφορετικά} \end{cases} \\ &= \mathrm{I}_{(0,x_{(n)}]}(x_{(1)})\mathrm{I}_{(0,\theta)}(x_{(n)}), \end{aligned}$$

όπου $x_{(1)} = \min(x_1, \ldots, x_n)$ και $x_{(n)} = \max(x_1, \ldots, x_n)$. Επομένως,

$$f(\underset{\sim}{x};\theta) = \frac{1}{\theta^n}\mathrm{I}_{(0,\theta)}(x_{(n)})\mathrm{I}_{(0,x_{(n)}]}(x_{(1)}) = q(x_{(n)},\theta)h(\underset{\sim}{x}), \quad \forall\ \underset{\sim}{x},\ \forall\ \theta,$$

όπου $q(x_{(n)},\theta) = \frac{1}{\theta^n}\mathrm{I}_{(0,\theta)}(x_{(n)})$ και $h(\underset{\sim}{x}) = \mathrm{I}_{(0,x_{(n)}]}(x_{(1)})$. Συνεπώς, η στατιστική συνάρτηση $T(\underset{\sim}{X}) = X_{(n)} = \max(X_1, \ldots, X_n)$ είναι επαρκής.
Θα πρέπει να ομολογήσουμε ότι αυτή η παραγοντοποίηση δεν είναι προφανής και μάλιστα είναι στοχευμένη έτσι ώστε να αναδυθεί τελικά η (μεταβλητή) $x_{(n)}$ μέσα στον τύπο της πυκνότητας $f(\underset{\sim}{x};\theta)$. Ας δούμε, διαισθητικά, γιατί η $X_{(n)}$ είναι υποψήφια επαρκής στατιστική συνάρτηση. Το κοινό σύνολο τιμών των παρατηρήσεων X_i είναι το διάστημα $(0,\theta)$, οπότε (με πιθανότητα 1) $0 < X_1 < \theta$, $0 < X_2 < \theta$,...,$0 < X_n < \theta$. Αφού λοιπόν το θ είναι μεγαλύτερο από όλες τις παρατηρήσεις, είναι μεγαλύτερο και από τη μέγιστη, $X_{(n)} = \max(X_1, \ldots, X_n)$, και αντίστροφα. Γνωρίζοντας επομένως την τιμή $x_{(n)}$ της $X_{(n)}$ αντλούμε την πληροφορία $\theta > x_{(n)}$, η οποία καλύπτει όλες τις πληροφορίες $\theta > x_1$,$\theta > x_2$,...,$\theta > x_n$ που παρέχει η τιμή $\underset{\sim}{x} = (x_1, \ldots, x_n)$ του $\underset{\sim}{X}$.

Παράδειγμα 6.1.6. (Ομοιόμορφη κατανομή, άγνωστα άκρα - επάρκεια) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την ομοιόμορφη κατανομή $\mathcal{U}(\theta,\theta+1)$ με πυκνότητα $f_1(x;\theta) = \mathrm{I}_{(\theta,\theta+1)}(x_1)$, $\theta \in \Theta = \mathbb{R}$. Τότε ανάλογα με το προηγούμενο παράδειγμα, έχουμε,

$$\begin{aligned} f(\underset{\sim}{x};\theta) &= \prod_{i=1}^{n} f_1(x_i;\theta) = \prod_{i=1}^{n} \mathrm{I}_{(\theta,\theta+1)}(x_i) = \mathrm{I}_{(\theta,x_{(n)}]}(x_{(1)}) \cdot \mathrm{I}_{(\theta,\theta+1)}(x_{(n)}), \\ &= q(x_{(1)},x_{(n)},\theta)h(\underset{\sim}{x}), \quad \forall\ \underset{\sim}{x},\ \forall\ \theta, \end{aligned}$$

όπου $x_{(1)} = \min(x_1, \ldots, x_n)$, $x_{(n)} = \max(x_1, \ldots, x_n)$, $q(x_{(1)},x_{(n)},\theta) = \mathrm{I}_{(\theta,x_{(n)}]}(x_{(1)}) \cdot \mathrm{I}_{(\theta,\theta+1)}(x_{(n)})$ και $h(\underset{\sim}{x}) = 1$. Επομένως, η στατιστική συνάρτηση $T(\underset{\sim}{X}) = (X_{(1)}, X_{(n)})$ είναι επαρκής. Παρατηρούμε εδώ ότι η επαρκής στατιστική συνάρτηση έχει διάσταση 2, ενώ η άγνωστη παράμετρος θ είναι πραγματική (έχει διάσταση 1). Τέτοιες περιπτώσεις, όπου η επαρκής στατιστική συνάρτηση έχει διάσταση μεγαλύτερη από τη διάσταση της παραμέτρου δημιουργούν γενικά δυσκολία στην εύρεση ενός «καλού» εκτιμητή της παραμέτρου.

Μια άλλη παραγοντοποίηση της $f(\underset{\sim}{x};\theta)$ είναι

$$f(\underset{\sim}{x};\theta) = \mathrm{I}_{(\theta,\theta+1)}(x_{(1)}) \cdot \mathrm{I}_{(\theta,\theta+1)}(x_{(n)})$$

που επίσης δίνει την $T(\underset{\sim}{X})$ ως επαρκή στατιστική συνάρτηση, ενώ εύκολα μπορούν να βρεθούν και άλλες. Μια αρχική πληροφορία που περιέχει η $T(\underset{\sim}{X}) = (X_{(1)}, X_{(n)})$ για το θ προκύπτει από το κοινό σύνολο τιμών των X_i, $(\theta, \theta+1)$. Με πιθανότητα 1, $\theta < X_1 < \theta+1$, $\theta < X_2 < \theta+1$,…,$\theta < X_n < \theta + 1$ ή ισοδύναμα $\theta < X_{(1)} \leq X_{(n)} < \theta + 1$, επομένως συμπεραίνουμε ότι $X_{(n)} - 1 < \theta < X_{(1)}$.

Παράδειγμα 6.1.7. (Ομοιόμορφη κατανομή - συμμετρικά άκρα) Έστω X_1 μία παρατήρηση από την ομοιόμορφη κατανομή $\mathcal{U}(-\theta,\theta)$ με πυκνότητα $f_1(x;\theta) = \frac{1}{2\theta}\mathrm{I}_{[0,\theta)}(|x|)$, $\theta \in \Theta = (0,\infty)$. Τότε $f_1(x;\theta) = q(|x|,\theta)h(x)$, όπου $q(|x|,\theta) = \frac{1}{2\theta}\mathrm{I}_{[0,\theta)}(|x|)$ και $h(x) = 1$. Επομένως, η στατιστική συνάρτηση $T(\underset{\sim}{X}) = |X|$ είναι επαρκής.

Παράδειγμα 6.1.8. (Κατανομή Bernoulli - μη επαρκής στατιστική συνάρτηση) Έστω $\underset{\sim}{X} = (X_1, X_2)$ είναι ένα τυχαίο δείγμα από την κατανομή Bernoulli $\mathcal{B}(1,\theta)$, $\theta \in \Theta = (0,1)$. Θα δείξουμε ότι η στατιστική συνάρτηση $S(\underset{\sim}{X}) = X_1$ δεν είναι επαρκής. Βάσει του ορισμού, αρκεί να δείξουμε ότι υπάρχει μία τιμή s της $S(\underset{\sim}{X})$ για την οποία, η δεσμευμένη κατανομή του $\underset{\sim}{X}|S(\underset{\sim}{X}) = s$ εξαρτάται από το θ. Θεωρούμε $s = 1$ και υπολογίζουμε τη δεσμευμένη πιθανότητα,

$$\begin{aligned}
\mathbb{P}_\theta(\underset{\sim}{X} = (1,0)|X_1 = 1) &= \mathbb{P}_\theta(X_1 = 1, X_2 = 0|X_1 = 1) \\
&= \frac{\mathbb{P}_\theta(X_1 = 1, X_2 = 0, X_1 = 1)}{\mathbb{P}_\theta(X_1 = 1)} \\
&= \frac{\mathbb{P}_\theta(X_1 = 1, X_2 = 0)}{\mathbb{P}_\theta(X_1 = 1)} \\
&= \frac{\mathbb{P}_\theta(X_1 = 1)\mathbb{P}_\theta(X_2 = 0)}{\mathbb{P}_\theta(X_1 = 1)} \quad \text{(ανεξαρτησία)} \\
&= \mathbb{P}_\theta(X_2 = 0) = 1 - \theta,
\end{aligned}$$

η οποία εξαρτάται από το θ. Άρα η $S(\underset{\sim}{X})$ δεν είναι επαρκής.

Παράδειγμα 6.1.9. (**Κατανομή Bernoulli - μη επαρκής στατιστική συνάρτηση**) Έστω $\underset{\sim}{X} = (X_1, X_2)$, όπου X_1, X_2 είναι ανεξάρτητες με κατανομές Bernoulli $X_1 \sim \mathcal{B}(1,\theta)$ και $X_2 \sim \mathcal{B}(1,2\theta)$, $\theta \in \Theta = (0,1/2)$. Τότε η στατιστική συνάρτηση $T(\underset{\sim}{X}) = X_1 + X_2$ δεν είναι επαρκής. Όπως στο Παράδειγμα 6.1.8, θα δείξουμε ότι υπάρχει τιμή t της $T(\underset{\sim}{X})$ για την οποία η δεσμευμένη κατανομή του $\underset{\sim}{X}|T(\underset{\sim}{X}) = t$ εξαρτάται από το θ. Θεωρούμε $t = 1$ και υπολογίζουμε τη δεσμευμένη πιθανότητα,

$$\begin{aligned}
\mathbb{P}_\theta(\underset{\sim}{X}) = (1,0)|X_1 + X_2 = 1) &= \mathbb{P}_\theta(X_1 = 1, X_2 = 0|X_1 + X_2 = 1) \\
&= \frac{\mathbb{P}_\theta(X_1 = 1, X_2 = 0, X_1 + X_2 = 1)}{\mathbb{P}_\theta(X_1 + X_2 = 1)} = \frac{\mathbb{P}_\theta(X_1 = 1, X_2 = 0)}{\mathbb{P}_\theta(X_1 + X_2 = 1)} \\
&= \frac{\mathbb{P}_\theta(X_1 = 1)\mathbb{P}_\theta(X_2 = 0)}{\mathbb{P}_\theta(X_1 = 1)\mathbb{P}_\theta(X_2 = 0) + \mathbb{P}_\theta(X_1 = 0)\mathbb{P}_\theta(X_2 = 1)} \quad \text{(ανεξαρτησία)} \\
&= \frac{\theta(1-2\theta)}{(1-\theta)2\theta + (1-2\theta)\theta} = \frac{1-2\theta}{3-4\theta},
\end{aligned}$$

η οποία εξαρτάται από το θ. Συνεπώς η στατιστική συνάρτηση $T(\underset{\sim}{X}) = X_1 + X_2$ δεν είναι επαρκής. □

Αναφέρουμε τώρα μερικές γενικές παρατηρήσεις για την έννοια της επάρκειας.

Παρατήρηση 6.1.1. Το δείγμα $\underset{\sim}{X} = (X_1, \ldots, X_n)$ είναι (τετριμμένα) επαρκής στατιστική συνάρτηση αφού $f(\underset{\sim}{x};\theta) = q(\underset{\sim}{x};\theta)h(\underset{\sim}{x})$ με $q(\underset{\sim}{x};\theta) = f(\underset{\sim}{x};\theta)$ και $h(\underset{\sim}{x}) = 1$.

Παρατήρηση 6.1.2. Εάν $\underset{\sim}{X} = (X_1, \ldots, X_n)$ είναι ένα τυχαίο δείγμα από μία κατανομή με πυκνότητα $f_1(x_1;\theta)$ και $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ είναι οι *διατεταγμένες στατιστικές συναρτήσεις*, δηλαδή

$$\begin{aligned} X_{(1)} &= \min(X_1, \ldots, X_n), \\ X_{(2)} &= 2\text{η μικρότερη παρατήρηση} \\ &\vdots \\ X_{(n-1)} &= 2\text{η μεγαλύτερη παρατήρηση} \\ X_{(n)} &= \max(X_1, \ldots, X_n), \end{aligned}$$

τότε έχουμε

$$\begin{aligned} f(\underset{\sim}{x};\theta) &= \prod_{i=1}^{n} f_1(x_i;\theta) = \prod_{i=1}^{n} f_1(x_{(i)};\theta) \quad \text{(λόγω συμμετρίας)} \\ &= q(x_{(1)}, \ldots, x_{(n)}, \theta) h(\underset{\sim}{x}) \end{aligned}$$

όπου $q(x_{(1)}, \ldots, x_{(n)}, \theta) = \prod_{i=1}^{n} f_1(x_{(i)};\theta)$ και $h(\underset{\sim}{x}) = 1$. Επομένως, η στατιστική συνάρτηση $T(\underset{\sim}{X}) = (X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ είναι επαρκής. Παρατηρούμε ότι αν και η $T(\underset{\sim}{X})$ έχει την ίδια διάσταση n όπως και το δείγμα $\underset{\sim}{X}$, εντούτοις αποτελεί «σύμπτυξη» της αφού είναι συνάρτηση του $\underset{\sim}{X}$.

Παρατήρηση 6.1.3. Έστω ότι η στατιστική συνάρτηση $T_1(\underset{\sim}{X})$ είναι επαρκής και $T_2(\underset{\sim}{X}) = K(T_1(\underset{\sim}{X}))$ είναι ένας $1-1$ μετασχηματισμός της T_1. Τότε $T_1(\underset{\sim}{X}) = K^{-1}\big(T_2(\underset{\sim}{X})\big)$ και η $T_2(\underset{\sim}{X})$ είναι επαρκής γιατί

$$f(\underset{\sim}{x};\theta) = q\big(T_1(\underset{\sim}{x}),\theta\big)\cdot h(\underset{\sim}{x}) = q\Big(K^{-1}\big(T_2(\underset{\sim}{x})\big),\theta\Big)\cdot h(\underset{\sim}{x}) = q_1\big(T_2(\underset{\sim}{x}),\theta\big)\cdot h(\underset{\sim}{x}).$$

Για παράδειγμα, ο γραμμικός μετασχηματισμός $T_2(\underset{\sim}{X}) = \alpha T_1(\underset{\sim}{X}) + \beta$, $\alpha \neq 0$, είναι επαρκής. Στο Παράδειγμα 6.1.3, στην πρώτη περίπτωση, ο εκτιμητής $\overline{X}$ του $\theta = \mu$ είναι επαρκής, στη δεύτερη ο εκτιμητής $\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$ του $\theta = \sigma^2$ είναι επαρκής, ενώ στην τρίτη περίπτωση το ζεύγος των εκτιμητών $(\overline{X}, \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2)$ είναι επαρκής στατιστική συνάρτηση για το $\theta = (\mu, \sigma^2)$ ως $1-1$ συνάρτηση της $T_2(\underset{\sim}{X}) = (\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2)$.

Παρατήρηση 6.1.4. Ας εξετάσουμε το εξής παράδειγμα (Lehmann and Casella, 1998, σελ. 37). Έστω ένα τυχαίο δείγμα $\underset{\sim}{X} = (X_1, \ldots, X_n)$ από την κανονική κατανομή $\mathcal{N}(0, \theta^2)$ με $\theta \in \Theta = (0, \infty)$. Τότε, εφαρμόζοντας το παραγοντικό κριτήριο είναι εύκολο να δειχθεί ότι οι στατιστικές συναρτήσεις

$$\begin{aligned}
T_1(\underset{\sim}{X}) &= (X_1, \ldots, X_n) \,, \\
T_2(\underset{\sim}{X}) &= (X_1^2, \ldots, X_n^2) \,, \\
T_3(\underset{\sim}{X}) &= \Big(\sum_{i=1}^{m} X_i^2, \sum_{i=m+1}^{n} X_i^2\Big) \,, \\
T_4(\underset{\sim}{X}) &= \sum_{i=1}^{n} X_i^2
\end{aligned}$$

είναι επαρκείς.

Παρατηρούμε ότι κάθε μία από τις $T_i(\underset{\sim}{X})$ αποτελεί συνάρτηση όλων των προηγούμενων και φαίνεται διαισθητικά τουλάχιστον ότι η $T_4(\underset{\sim}{X})$, που είναι συνάρτηση όλων των άλλων, αποτελεί τη μεγαλύτερη δυνατή «σύμπτυξη» του $\underset{\sim}{X}$ (χωρίς απώλεια πληροφορίας για το θ^2). Για αυτόν τον λόγο, η $T_4(\underset{\sim}{X}) = \sum_{i=1}^{n} X_i^2$ αναφέρεται ως ελάχιστη επαρκής (minimal sufficient) στατιστική συνάρτηση. Η επωνυμία ελάχιστη αντανακλά τη μικρότερη διάσταση. Γενικά, μία επαρκής στατιστική συνάρτηση λέγεται *ελάχιστη επαρκής* εάν είναι συνάρτηση οποιασδήποτε άλλης επαρκούς στατιστικής συνάρτησης. Από τον ορισμό και την Παρατήρηση 6.1.3 προκύπτει εύκολα ότι $1-1$ μετασχηματισμός ελάχιστης επαρκούς στατιστικής συνάρτησης είναι επίσης ελάχιστη επαρκής. Επιπλέον, αν δύο επαρκείς στατιστικές συναρτήσεις είναι ελάχιστες, τότε κάθε μια είναι συνάρτηση της άλλης. Σε όλα τα παραδείγματα που αναφέραμε προηγουμένως, οι επαρκείς στατιστικές συναρτήσεις ήταν ελάχιστες επαρκείς στατιστικές συναρτήσεις. Υπάρχει σχετική μεθοδολογία απόδειξης ότι μία επαρκής στατιστική συνάρτηση είναι ελάχιστη επαρκής αλλά δεν θα ασχοληθούμε με αυτήν τη μεθοδολογία εδώ, παραπέμπουμε, όμως, τον αναγνώστη στα βιβλία Casella and Berger (2002), Lehmann and Casella (1998) και Ηλιόπουλος (2013).

Η επόμενη πρόταση ερμηνεύει τη συνάρτηση q του παραγοντικού κριτηρίου, δείχνοντας ότι αποτελεί τον έναν από τους δύο όρους παραγοντοποίησης, παρόμοιας προς την (6.1), της πυκνότητας της επαρκούς στατιστικής συνάρτησης $T(\underset{\sim}{X})$. Επιπλέον, τεκμηριώνει ποσοτικά, μέσω του αριθμού πληροφορίας του Fisher, τη διαισθητική ερμηνεία της επάρκειας περί μη απώλειας πληροφορίας.

Πρόταση 6.1.2. *Έστω $T(\underset{\sim}{X})$ επαρκής στατιστική συνάρτηση με πυκνότητα $f_T(t;\theta)$, $\theta \in \Theta$. Θεωρούμε την παραγοντοποίηση της πυκνότητας του δείγματος $\underset{\sim}{X}$, $f(\underset{\sim}{x};\theta)$, που δίνεται στη σχέση* (6.1).

(α) Για κάθε t στο σύνολο τιμών της $T(\underset{\sim}{X})$ και για κάθε $\theta \in \Theta$ ισχύει η σχέση

$$f_T(t;\theta) = q\big(t,\theta\big)\, h_1(t), \tag{6.4}$$

όπου h_1 είναι συνάρτηση μη εξαρτώμενη από το θ.

(β) Για κάθε $\theta \in \Theta$, ο αριθμός πληροφορίας του Fisher, $I_T(\theta)$, που περιέχεται στη στατιστική συνάρτηση $T(\underset{\sim}{X})$ για το θ (όπως ορίστηκε στη σχέση (5.7) *) είναι ίσος με τον αριθμό πληροφορίας του Fisher για το θ που περιέχεται στο δείγμα $\underset{\sim}{X}$, δηλαδή*

$$I_T(\theta) = I(\theta). \tag{6.5}$$

Απόδειξη. (α) Θα θεωρήσουμε, κατ' αρχάς, ότι το δείγμα $\underset{\sim}{X}$ έχει διακριτή κατανομή, οπότε διακριτή είναι και η κατανομή της $T(\underset{\sim}{X})$. Έστω t στο σύνολο τιμών της $T(\underset{\sim}{X})$, άρα

$$f_T(t;\theta) = \mathbb{P}_\theta(T(\underset{\sim}{X}) = t) > 0$$

και από τις (6.1) και (6.3) παίρνουμε

$$f_T(t;\theta) = q\big(t,\theta\big) \sum_{\underset{\sim}{x} \in \mathcal{S}_t} h\left(\underset{\sim}{x}\right), \tag{6.6}$$

όπου $\mathcal{S}_t = \big\{\underset{\sim}{x} \in \mathcal{S} : T(\underset{\sim}{x}) = t\big\}$. Επειδή το σύνολο $\mathcal{S}_t$ καθορίζεται πλήρως από το σημείο t, το άθροισμα ή η σειρά $\sum\limits_{\underset{\sim}{x}\,\in\,\mathcal{S}_t} h\left(\underset{\sim}{x}\right)$ είναι συνάρτηση του t, η οποία δεν εξαρτάται από το θ αφού και η h δεν εξαρτάται από το θ. Θέτουμε $h_1(t) = \sum_{\underset{\sim}{x} \in \mathcal{S}_t} h\left(\underset{\sim}{x}\right)$, οπότε από την (6.6) προκύπτει η (6.4).

Η περίπτωση συνεχούς δείγματος $\underset{\sim}{X}$ είναι τεχνικά δύσκολη, για αυτό και παραθέτουμε μια σκιαγράφηση της απόδειξης. Κατ' αντιστοιχία με τη διακριτή περίπτωση, (όπου προσθέτουμε τιμές της πυκνότητας $f(\underset{\sim}{x};\theta)$ προκειμένου να βρούμε την πυκνότητα της $T(\underset{\sim}{X})$) έχουμε

$$f_T(t;\theta) = \int_{\mathcal{S}_t} f(\underset{\sim}{x};\theta)\, d\underset{\sim}{x},$$

όπου $\mathcal{S}_t = \left\{\underset{\sim}{x} \in \mathcal{S} : T(\underset{\sim}{x}) = t\right\}$ είναι μη αριθμήσιμο σύνολο και λόγω της (6.1),

$$\begin{aligned} f_T(t;\theta) &= \int_{\mathcal{S}_t} q\big(T(\underset{\sim}{x}),\theta\big) h\left(\underset{\sim}{x}\right)\, d\underset{\sim}{x} = \int_{\mathcal{S}_t} q\left(t,\theta\right) h\left(\underset{\sim}{x}\right)\, d\underset{\sim}{x} \\ &= q\left(t,\theta\right) \int_{\mathcal{S}_t} h(\underset{\sim}{x})\, d\underset{\sim}{x}. \end{aligned} \tag{6.7}$$

Η δεύτερη ισότητα ισχύει επειδή για $\underset{\sim}{x} \in \mathcal{S}_t$, εξ ορισμού του $\mathcal{S}_t$, έχουμε $T(\underset{\sim}{x}) = t$ και η τρίτη επειδή $q\left(t,\theta\right)$ είναι σταθερά ως προς τη μεταβλητή της ολοκλήρωσης, $\underset{\sim}{x}$. Το τελευταίο ολοκλήρωμα εξαρτάται μόνον από το σημείο t, είναι δηλαδή συνάρτηση του t, οπότε θέτοντας $h_1(t) = \int_{\mathcal{S}_t} h\left(\underset{\sim}{x}\right)\, d\underset{\sim}{x}$, η (6.4) αληθεύει.

(β) Από τον ορισμό του $I_T(\theta)$ έχουμε

$$I_T(\theta) = \mathbb{E}_\theta\left[\left(\tfrac{\partial}{\partial\theta} \ln f_T(T(\underset{\sim}{X});\theta)\right)^2\right]. \tag{6.8}$$

Επιπλέον από την (6.4) παίρνουμε διαδοχικά,

$$\begin{aligned} \ln f_T(T(\underset{\sim}{X});\theta) &= \ln q\big(T(\underset{\sim}{X}),\theta\big) + \ln h_1\big(T(\underset{\sim}{X})\big), \\ \tfrac{\partial}{\partial\theta} \ln f_T(T(\underset{\sim}{X});\theta) &= \tfrac{\partial}{\partial\theta} \ln q\big(T(\underset{\sim}{X}),\theta\big). \end{aligned} \tag{6.9}$$

Ανάλογα, από την (6.1) προκύπτει ότι

$$\tfrac{\partial}{\partial\theta} \ln f(\underset{\sim}{X};\theta) = \tfrac{\partial}{\partial\theta} \ln q\big(T(\underset{\sim}{X}),\theta\big). \tag{6.10}$$

Συνδυάζοντας τις (6.8), (6.9) και (6.10) συμπεραίνουμε ότι

$$\begin{aligned} I_T(\theta) &= \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln q(T(\underset{\sim}{X});\theta)\right)^2\right] \\ &= \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f(\underset{\sim}{X};\theta)\right)^2\right] = I(\theta). \end{aligned}$$

□

6.2 Χρήση της επάρκειας στη βελτίωση εκτιμητών-μείωση της διασποράς και του ΜΤΣ

Μία σημαντική εφαρμογή της έννοιας της επάρκειας δίνεται στο επόμενο θεώρημα των Rao–Blackwell. Η απόδειξή του παρουσιάστηκε από τον Rao (1945) και, ανεξάρτητα, από τον Blackwell (1947). Το θεώρημα αυτό δείχνει ότι εάν $T(\underset{\sim}{X})$ είναι επαρκής στατιστική συνάρτηση και $S(\underset{\sim}{X})$ είναι ένας εκτιμητής του $g(\theta)$ *που δεν είναι* συνάρτηση της $T(\underset{\sim}{X})$, τότε υπάρχει εκτιμητής *που είναι* συνάρτηση της $T(\underset{\sim}{X})$ και έχει μέσο τετραγωνικό σφάλμα μικρότερο από αυτό του $S(\underset{\sim}{X})$. Συνεπώς, *εκτιμητές που δεν είναι συναρτήσεις της επαρκούς στατιστικής συνάρτησης* $T(\underset{\sim}{X})$ *είναι μη αποδεκτοί, με κριτήριο το ΜΤΣ* (βλέπε τον Ορισμό 4.1.2). Βάσει αυτής της ιδιότητας της επάρκειας, κάλλιστα, μπορούμε να ισχυριστούμε ότι μια επαρκής στατιστική συνάρτηση όχι μόνον «αρκεί για την εξαγωγή συμπερασμάτων για την άγνωστη παράμετρο θ και την άγνωστη τιμή $g(\theta)$» αλλά είναι και αναγκαία, αφού η μη χρησιμοποίησή της οδηγεί σε μη ακριβή συμπεράσματα - μη ακριβή εκτίμηση στην προκειμένη περίπτωση. Θεωρούμε λοιπόν ότι η ονομασία *επαρκής και αναγκαία* θα απέδιδε πιο πιστά τη σημασία της.

Υπενθυμίζουμε ότι η δεσμευμένη μέση τιμή και η δεσμευμένη διασπορά που αναφέρονται στο Θεώρημα 6.2.1 και την απόδειξή του ορίζονται γενικά στην Ενότητα 1.9. Ειδικά, υπενθυμίζουμε ότι η $\mathbb{E}(S \mid T)$ είναι τυχαία μεταβλητή και συνάρτηση της $T(\underset{\sim}{X})$. Για λόγους απλότητας, θα γράφουμε $\mathbb{E}(S \mid T)$ αντί $\mathbb{E}\big(S(\underset{\sim}{X}) \mid T(\underset{\sim}{X})\big)$ και θα χρησιμοποιούμε αδιακρίτως τον συμβολισμό $S^* = \mathbb{E}(S \mid T)$ ή $S^*(T) = \mathbb{E}(S \mid T)$ με τον δεύτερο να τονίζει ότι η δεσμευμένη μέση τιμή είναι συνάρτηση της $T(\underset{\sim}{X})$. Επίσης, ως συνήθως, S ή $S(\underset{\sim}{X})$ παριστάνει την ίδια στατιστική συνάρτηση.

Θεώρημα 6.2.1. *(***Rao–Blackwell***)* *Έστω $T(\underset{\sim}{X})$ μια επαρκής στατιστική συνάρτηση και $S(\underset{\sim}{X})$ ένας εκτιμητής του $g(\theta)$ με $\mathbb{V}ar_\theta S < \infty$ για κάθε $\theta \in \Theta$. Ακόμη, έστω $S^*(T) = \mathbb{E}(S \mid T)$. Τότε έχουμε τα εξής.*

(α) $\mathbb{E}_\theta S^* = \mathbb{E}_\theta S$, *για κάθε $\theta \in \Theta$ και συνεπώς εάν ο S είναι αμερόληπτος εκτιμητής του $g(\theta)$ το ίδιο ισχύει και για τον S^*.*

(β) $\mathbb{V}ar_\theta S^* \le \mathbb{V}ar_\theta S$, *για κάθε $\theta \in \Theta$ και ισχύει γνήσια ανισότητα εκτός εάν ο εκτιμητής S είναι συνάρτηση του T, οπότε $S^* = S$.*

(γ) *ΜΤΣ$(S^*, \theta) \le$ ΜΤΣ(S, θ), για κάθε $\theta \in \Theta$ και ισχύει γνήσια ανισότητα εκτός εάν ο εκτιμητής $S(\underset{\sim}{X})$ είναι συνάρτηση της $T(\underset{\sim}{X})$, οπότε $S^* = S$.*

Απόδειξη. Επειδή η $T(\underset{\sim}{X})$ είναι επαρκής στατιστική συνάρτηση η κατανομή του $\underset{\sim}{X}$, άρα και του $S(\underset{\sim}{X})$ δοθέντος ότι $T(\underset{\sim}{X}) = t$ δεν εξαρτάται από το θ, για κάθε τιμή t της $T(\underset{\sim}{X})$. Επομένως και η δεσμευμένη μέση τιμή $S^*(T) = \mathbb{E}(S \mid T)$ δεν εξαρτάται από το θ. Συνεπώς $S^*(T)$ είναι στατιστική συνάρτηση.

(α) Από την Πρόταση 1.9.2 έχουμε

$$\mathbb{E}_\theta S^* = \mathbb{E}_\theta\big(\mathbb{E}(S \mid T)\big) = \mathbb{E}_\theta S, \quad \text{για κάθε } \theta \in \Theta.$$

(β) Από την Πρόταση 1.9.2 έχουμε

$$\begin{aligned} \mathbb{V}ar_\theta S &= \mathbb{V}ar_\theta\big(\mathbb{E}(S \mid T)\big) + \mathbb{E}_\theta\big(\mathbb{V}ar(S \mid T)\big) \\ &= \mathbb{V}ar_\theta S^* + \mathbb{E}_\theta\big(\mathbb{V}ar(S \mid T)\big). \end{aligned} \tag{6.11}$$

Επειδή $\mathbb{V}ar(S \mid T) \ge 0$ (Πρόταση 1.9.1) έχουμε $\mathbb{E}_\theta\big(\mathbb{V}ar(S \mid T)\big) \ge 0$ (Πρόταση 1.4.2(6)). Άρα από την (6.11) παίρνουμε $\mathbb{V}ar_\theta S \ge \mathbb{V}ar_\theta S^*$, για κάθε $\theta \in \Theta$. Έστω επιπλέον ότι, $\mathbb{V}ar_\theta S = \mathbb{V}ar_\theta S^*$. Τότε από την (6.11) προκύπτει ότι η τελευταία σχέση ισχύει εάν και μόνον εάν $\mathbb{E}_\theta\big(\mathbb{V}ar(S \mid T)\big) = 0$ ή ισοδύναμα $\mathbb{V}ar(S \mid T) = 0$ με πιθανότητα 1 (Πρόταση 1.4.2(5)) ή ισοδύναμα ο S είναι συνάρτηση της $T(\underset{\sim}{X})$ (Πρόταση 1.9.1(3)) οπότε,

από την Πρόταση 1.9.1(1), έχουμε $\mathbb{E}(S \mid T) = S$, δηλαδή $S^* = S$. Τελικά λοιπόν έχουμε $S^* = S$, εάν και μόνον εάν $\mathbb{V}ar_\theta S = \mathbb{V}ar_\theta S^*$.
(γ) Από την Πρόταση 4.1.1 έχουμε

$$\text{ΜΤΣ}(S^*, \theta) = \mathbb{V}ar_\theta S^* + (\mathbb{E}_\theta S^* - g(\theta))^2\,, \tag{6.12}$$

και αντίστοιχα

$$\text{ΜΤΣ}(S, \theta) = \mathbb{V}ar_\theta S + (\mathbb{E}_\theta S - g(\theta))^2\,. \tag{6.13}$$

Το συμπέρασμα (γ) προκύπτει, επομένως, από τις (6.12), (6.13) και τα (α), (β). □

Σημειώνουμε ότι στην απόδειξη του Θεωρήματος 6.2.1, η επάρκεια χρειάστηκε μόνον για να διαπιστώσουμε ότι $S^*(T) = \mathbb{E}(S \mid T)$ είναι στατιστική συνάρτηση. Ο εκτιμητής $S^*(T) = \mathbb{E}(S \mid T)$ αναφέρεται στη βιβλιογραφία ως *Rao-Blackwell βελτίωση* του εκτιμητή S.
Η επόμενη πρόταση καταγράφει, ουσιαστικά, τα (α), (β) και (γ) του Θεωρήματος 6.2.1 με έναν πιο εμφατικό τρόπο. Η απόδειξή της είναι άμεση συνέπεια του Θεωρήματος 6.2.1.

Πρόταση 6.2.2. *Έστω $T(\underset{\sim}{X})$ επαρκής στατιστική συνάρτηση, $S(\underset{\sim}{X})$ εκτιμητής του $g(\theta)$ με $\mathbb{V}ar_\theta S < \infty$ για κάθε $\theta \in \Theta$, ο οποίος δεν είναι συνάρτηση της $T(\underset{\sim}{X})$, και $S^*(T) = \mathbb{E}(S \mid T)$ η Rao-Blackwell βελτίωση του $S(\underset{\sim}{X})$. Τότε ισχύουν τα εξής.*

(α) Ο $S(\underset{\sim}{X})$ είναι μη αποδεκτός εκτιμητής του $g(\theta)$ και ΜΤΣ$(S^*, \theta) <$ ΜΤΣ(S, θ), για κάθε $\theta \in \Theta$.

(β) Εάν, επιπλέον, $S(\underset{\sim}{X})$ είναι αμερόληπτος εκτιμητής του $g(\theta)$ τότε και ο $S^*(\underset{\sim}{X})$ είναι αμερόληπτος με $\mathbb{V}ar_\theta S^* < \mathbb{V}ar_\theta S$, για κάθε $\theta \in \Theta$.

Παρατήρηση 6.2.1. Ο εκτιμητής S^* δεν μπορεί να βελτιωθεί περαιτέρω μέσω της επαρκούς στατιστικής συνάρτησης $T(\underset{\sim}{X})$ γιατί είναι ήδη συνάρτηση της $T(\underset{\sim}{X})$ και επομένως η Rao–Blackwell βελτίωση του, $\mathbb{E}(S^* \mid T)$, είναι ο ίδιος ο εκτιμητής S^*, λόγω του (β) του Θεωρήματος 6.2.1, δηλαδή $\mathbb{E}(S^* \mid T) = S^*$.

Παράδειγμα 6.2.1. (Ομοιόμορφη κατανομή - μη αποδεκτικότητα του δειγματικού μέσου $\bar{X}$) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$, $n \geq 2$, ένα τυχαίο δείγμα από την ομοιόμορφη κατανομή $\mathcal{U}(0, \theta)$, $\theta \in \Theta = (0, \infty)$. Υπενθυμίζουμε ότι $\mathbb{E}_\theta(X_i) = \frac{\theta}{2}$. Γνωρίζουμε ότι η $T(\underset{\sim}{X}) = X_{(n)} = \max(X_1, \ldots, X_n)$ είναι επαρκής στατιστική συνάρτηση (Παράδειγμα 6.1.5) και $S(\underset{\sim}{X}) = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ είναι αμερόληπτος εκτιμητής του $\frac{\theta}{2}$. Ο $S(\underset{\sim}{X})$ είναι μη αποδεκτός αφού δεν είναι συνάρτηση της $T(\underset{\sim}{X})$ και ένας καλύτερος εκτιμητής είναι ο $S^*(T) = \mathbb{E}(S \mid T) = \mathbb{E}(\bar{X} \mid X_{(n)})$. Ο υπολογισμός του $S^*(T)$ σε κλειστή μορφή θα γίνει με μέθοδο της επόμενης ενότητας (βλέπε Παράδειγμα 6.3.3).

6.3 Πληρότητα και ΑΟΕΔ εκτιμητές

Είδαμε ότι εάν $T(\underset{\sim}{X})$ είναι επαρκής στατιστική συνάρτηση και $S(\underset{\sim}{X})$ είναι αμερόληπτος εκτιμητής του $g(\theta)$ τότε $S^*(T) = \mathbb{E}(S \mid T)$, η Rao–Blackwell βελτίωση του $S(\underset{\sim}{X})$, είναι επίσης αμερόληπτος εκτιμητής του $g(\theta)$ και έχει μικρότερη (ή το πολύ ίση) διασπορά προς αυτήν του $S(\underset{\sim}{X})$. Θα δούμε στη συνέχεια ότι εάν η στατιστική συνάρτηση $T(\underset{\sim}{X})$ έχει μία επιπλέον ιδιότητα, την ιδιότητα της πληρότητας, τότε $S^*(T) = \mathbb{E}(S \mid T)$ δεν είναι απλώς καλύτερος εκτιμητής (ως προς τη διασπορά, άρα και το ΜΤΣ) από τον $S(\underset{\sim}{X})$, αλλά είναι καλύτερος από οποιονδήποτε άλλο αμερόληπτο εκτιμητή του $g(\theta)$. Δίνουμε πρώτα τον εξής ορισμό.

Ορισμός 6.3.1. *Ο εκτιμητής $S_0(\underset{\sim}{X})$ ονομάζεται αμερόληπτος ομοιομόρφως ελάχιστης διασποράς (ΑΟΕΔ) εκτιμητής του $g(\theta)$ εάν*

(α) είναι αμερόληπτος,

(β) $\mathbb{V}ar_\theta S_0(\underset{\sim}{X}) \leq \mathbb{V}ar_\theta S_1(\underset{\sim}{X})$, για κάθε $\theta \in \Theta$ και για κάθε αμερόληπτο εκτιμητή $S_1(\underset{\sim}{X})$ του $g(\theta)$.

Ο όρος «ομοιομόρφως» τονίζει το γεγονός ότι η ανισότητα (β) ισχύει για κάθε $\theta \in \Theta$ (όποια δηλαδή και αν είναι η τιμή της άγνωστης παραμέτρου θ).

Από τον Ορισμό 5.2.2, προκύπτει αμέσως ότι ένας αποδοτικός εκτιμητής είναι ΑΟΕΔ εκτιμητής αφού η διασπορά του συμπίπτει με την ελάχιστη διασπορά αμερόληπτων εκτιμητών που παρέχει το Κ.Φ. C–R=

$\big(g'(\theta)\big)^2/I(\theta)$. Από την άλλη πλευρά, ένας ΑΟΕΔ εκτιμητής έχει μεν ελάχιστη διασπορά μεταξύ των αμερόληπτων εκτιμητών, αλλά είναι δυνατόν και αυτή να είναι μεγαλύτερη από το Κ.Φ. C–R. Επομένως, ένας ΑΟΕΔ εκτιμητής δεν είναι κατ' ανάγκη αποδοτικός.

Η έννοια της πληρότητας εισήχθηκε από τους Lehmann and Scheffé (1955) και ορίζεται ως ακολούθως.

Ορισμός 6.3.2. *Η στατιστική συνάρτηση $T(\underset{\sim}{X})$ ονομάζεται πλήρης εάν η σχέση*

$$\mathbb{E}_\theta \phi(T) = 0, \forall\ \theta \in \Theta,$$

όπου η ϕ είναι μια συνάρτηση, με πραγματικές τιμές, ορισμένη στο σύνολο τιμών της $T(\underset{\sim}{X})$ και μη εξαρτώμενη από το θ, συνεπάγεται $\phi(T) = 0$, δηλαδή $\phi(t) = 0$ για κάθε τιμή t της στατιστικής συνάρτησης $T(\underset{\sim}{X})$.

Προφανώς, εάν $\phi(T) = 0$, τότε $\mathbb{E}_\theta \phi(T) = 0$, για κάθε $\theta \in \Theta$. Πληρότητα σημαίνει ότι ισχύει και το *αντίστροφο*. Παρατηρούμε λοιπόν ότι η στατιστική συνάρτηση $T(\underset{\sim}{X})$ είναι πλήρης εάν και μόνον εάν η μοναδική συνάρτηση της $T(\underset{\sim}{X})$ με μέση τιμή μηδέν για κάθε $\theta \in \Theta$ είναι η σταθερή συνάρτηση μηδέν. Μία στατιστική συνάρτηση $\phi(T)$ που ικανοποιεί τη συνθήκη του ορισμού, $\mathbb{E}_\theta \phi(T) = 0$, για κάθε $\theta \in \Theta$, αναφέρεται ως *αμερόληπτος εκτιμητής του μηδενός* (καταχράζοντας τον ορισμό της αμεροληψίας, αφού το μηδέν δεν χρειάζεται να εκτιμηθεί). Συνεπώς, η στατιστική συνάρτηση $T(\underset{\sim}{X})$ είναι πλήρης εάν και μόνον εάν ο μοναδικός αμερόληπτος εκτιμητής του μηδενός είναι η σταθερή συνάρτηση μηδέν.

Παρατηρούμε ακόμη ότι η ιδιότητα της πληρότητας (όπως και της αμεροληψίας) είναι ιδιότητα *της οικογένειας κατανομών* $\{f_T(t;\theta)\colon \theta \in \Theta\}$ της στατιστικής συνάρτησης $T(\underset{\sim}{X})$ και όχι της $T(\underset{\sim}{X})$ αυτής καθ' εαυτής ως συνάρτησης. Για αυτό, είναι πιο σωστό να λέγεται ότι *η οικογένεια κατανομών της στατιστικής συνάρτησης $T(\underset{\sim}{X})$, $\{f_T(t;\theta)\colon \theta \in \Theta\}$, είναι πλήρης*. Παραπέμπουμε μάλιστα τον αναγνώστη στον Rohatgi (1976, σελ.346) όπου η οικογένεια κατανομών $\{f_T(t;\theta)\colon \theta \in \Theta\}$ μιας στατιστικής συνάρτησης $T(\underset{\sim}{X})$ είναι πλήρης, αν όμως αφαιρεθεί ένα οποιοδήποτε μέλος αυτής της οικογένειας, τότε δεν διατηρείται η πληρότητα, δηλαδή η οικογένεια κατανομών $\{f_T(t;\theta)\colon \theta \in \Theta \setminus \{\theta_0\}\}$, όπου $\theta_0 \in \Theta$, δεν είναι πλήρης. Εν τούτοις για λόγους απλότητας θα διατηρήσουμε την ορολογία ότι η στατιστική συνάρτηση $T(\underset{\sim}{X})$ είναι πλήρης. Ένα πιο απλό παράδειγμα που δείχνει ότι η πληρότητα είναι ιδιότητα της οικογένειας κατανομών είναι το ακόλουθο.

Παράδειγμα 6.3.1. (**η πληρότητα ως ιδιότητα της οικογένειας κατανομών**) Έστω X μία παρατήρηση από τη διακριτή κατανομή που δίνεται στον πίνακα της Ενότητας 3.3. Θα δείξουμε ότι η οικογένεια κατανομών της $T(X) = X$, $\{f(x;\theta)\colon \theta \in \Theta\}$, όπου $\Theta = \{\theta_1, \theta_2, \theta_3\}$, είναι πλήρης. Θεωρούμε, σύμφωνα με τον ορισμό της πληρότητας, στατιστική συνάρτηση $\phi(X)$ τέτοια ώστε $\mathbb{E}_\theta \phi(X) = 0$ ή ισοδύναμα $\sum_{x=0}^{2} \phi(x) f(x;\theta) = 0$, για κάθε $\theta \in \{\theta_1, \theta_2, \theta_3\}$. Αναλυτικά, αντικαθιστώντας τις τιμές $f(x;\theta)$ από τον πίνακα προκύπτουν οι εξής τρεις σχέσεις, μία για κάθε τιμή του θ.

$$
\begin{aligned}
0.02\phi(0) + 0.95\phi(1) + 0.03\phi(2) &= 0 \qquad (\theta = \theta_1)\\
0.90\phi(0) + 0.05\phi(1) + 0.05\phi(2) &= 0 \qquad (\theta = \theta_2)\\
0.23\phi(0) + 0.06\phi(1) + 0.71\phi(2) &= 0 \qquad (\theta = \theta_3)
\end{aligned}
$$

Λύνοντας ως προς $\phi(0)$, $\phi(1)$, $\phi(2)$ το *σύστημα* αυτών των εξισώσεων είναι εύκολο να διαπιστώσουμε ότι η μοναδική λύση του είναι $\phi(0) = \phi(1) = \phi(2) = 0$. Τότε όμως $\phi(X) = 0$, αφού η παρατήρηση X έχει σύνολο τιμών $\{0, 1, 2\}$, οπότε εξ ορισμού είναι πλήρης. Περαιτέρω, ας θεωρήσουμε την X με οικογένεια κατανομών $\{f(x;\theta)\colon \theta \in \Theta^*\}$, όπου $\Theta^* = \{\theta_1, \theta_2\}$, αφαιρώντας δηλαδή ένα μέλος της αρχικής οικογένειας, αυτό που αντιστοιχεί στην τιμή $\theta = \theta_3$. Για να εξετάσουμε την πληρότητα της $T(X) = X$, θεωρούμε όπως προηγουμένως στατιστική συνάρτηση $\phi(X)$ τέτοια ώστε $\mathbb{E}_\theta \phi(X) = 0$, για κάθε $\theta \in \Theta^*$. Ανάλογα, προκύπτει ένα σύστημα δύο εξισώσεων με τρεις αγνώστους $\phi(0)$, $\phi(1)$, $\phi(2)$ (που είναι οι δύο πρώτες από τις παραπάνω τρεις), το οποίο έχει άπειρες μη μηδενικές λύσεις. Άρα υπάρχουν άπειρες μη μηδενικές στατιστικές συναρτήσεις $\phi(X)$ που ικανοποιούν τη σχέση $\mathbb{E}_\theta \phi(X) = 0$, για κάθε $\theta \in \Theta^*$, οπότε εξ ορισμού, η X δεν είναι πλήρης. Βλέπουμε λοιπόν ότι μεταβάλλοντας την οικογένεια κατανομών της X, δεν διατηρήθηκε η πληρότητα.

Η σημασία της πληρότητας καταδεικνύεται στην επόμενη πρόταση: δύο στατιστικές συναρτήσεις που είναι συναρτήσεις πλήρους στατιστικής συνάρτησης είναι ίσες, αρκεί να έχουν ίσες μέσες τιμές.

Πρόταση 6.3.1. *Έστω $T(\underset{\sim}{X})$ πλήρης στατιστική συνάρτηση και $S_1(T)$, $S_2(T)$ στατιστικές συναρτήσεις με πραγματικές τιμές που είναι συναρτήσεις της $T(\underset{\sim}{X})$. Αν $\mathbb{E}_\theta S_1(T) = \mathbb{E}_\theta S_2(T)$ για κάθε $\theta \in \Theta$, τότε ισχύει $S_1(T) = S_2(T)$.*

Απόδειξη. Επειδή $S_1(T)$ και $S_2(T)$ είναι συναρτήσεις της $T(\underset{\sim}{X})$, το ίδιο ισχύει και για τη διαφορά τους $\phi(T) = S_1(T) - S_2(T)$. Επιπλέον, έχουμε

$$\mathbb{E}_\theta \phi(T) = \mathbb{E}_\theta\big(S_1(T) - S_2(T)\big) = \mathbb{E}_\theta S_1(T) - \mathbb{E}_\theta S_2(T) = 0, \quad \forall\, \theta \in \Theta.$$

Άρα από τον ορισμό της πληρότητας, $\phi(T) = 0$, δηλαδή $S_1(T) = S_2(T)$. □

Εφαρμόζοντας την Πρόταση 6.3.1 σε αμερόληπτους εκτιμητές προκύπτει ένα αποτέλεσμα καθοριστικής σημασίας για την εύρεση ΑΟΕΔ εκτιμητή.

Πρόταση 6.3.2. *Έστω $T(\underset{\sim}{X})$ πλήρης στατιστική συνάρτηση.*

(α) Αν $S_1(T)$ και $S_2(T)$ είναι αμερόληπτοι εκτιμητές του $g(\theta)$ και συναρτήσεις της $T(\underset{\sim}{X})$, τότε $S_1(T) = S_2(T)$.

(β) Το σύνολο των αμερόληπτων εκτιμητών του $g(\theta)$ που είναι συναρτήσεις της $T(\underset{\sim}{X})$ είναι κενό ή μονοσύνολο, δηλαδή υπάρχει το πολύ ένας αμερόληπτος εκτιμητής του $g(\theta)$ που είναι συνάρτηση πλήρους στατιστικής συνάρτησης.

Απόδειξη. (α) Λόγω αμεροληψίας έχουμε $\mathbb{E}_\theta S_1(T) = g(\theta)$, $\mathbb{E}_\theta S_2(T) = g(\theta)$ και άρα $\mathbb{E}_\theta S_1(T) = \mathbb{E}_\theta S_2(T)$ για κάθε $\theta \in \Theta$. Επομένως, από την Πρόταση 6.3.1 προκύπτει ότι $S_1(T) = S_2(T)$.
(β) Άμεση συνέπεια του (α). □

Θα περιγράψουμε τώρα μια κλασική πλέον τεχνική μέσω της οποίας η επάρκεια σε συνδυασμό με την πληρότητα μπορούν να οδηγήσουν στην εύρεση ΑΟΕΔ εκτιμητή. Η τεχνική αυτή παρουσιάστηκε από τους Lehmann and Scheffé (1955) και φέρει το όνομά τους. Έστω ότι έχουμε στη διάθεσή μας στατιστική συνάρτηση που είναι επαρκής και συγχρόνως πλήρης. Κατ' αρχάς, η επάρκεια, μέσω του Θεωρήματος των Rao–Blackwell, περιορίζει την κλάση των αμερόληπτων εκτιμητών του $g(\theta)$ (αν δεν είναι κενή ή δεν περιέχει έναν μόνον εκτιμητή), απορρίπτοντας όσους δεν είναι συναρτήσεις της $T(\underset{\sim}{X})$. Τότε, όμως, επειδή η $T(\underset{\sim}{X})$ είναι πλήρης, από την Πρόταση 6.3.2 (β) απομένει μόνον ένας αμερόληπτος εκτιμητής που είναι συνάρτηση της $T(\underset{\sim}{X})$ και αυτός είναι ο ΑΟΕΔ εκτιμητής. Οι λεπτομέρειες αυτής της τεχνικής δίνονται στο επόμενο θεώρημα.

Θεώρημα 6.3.3. **(Lehmann–Scheffé)** *Έστω $T(\underset{\sim}{X})$ επαρκής και πλήρης στατιστική συνάρτηση και $S(\underset{\sim}{X})$ αμερόληπτος εκτιμητής του $g(\theta)$. Τότε η Rao-Blackwell βελτίωση του $S(\underset{\sim}{X})$, $S^*(T) = \mathbb{E}(S \mid T)$ είναι ΑΟΕΔ εκτιμητής του $g(\theta)$. Εάν, επιπλέον, $\mathbb{V}ar_\theta S^* < \infty$ για κάθε $\theta \in \Theta$, τότε $S^*(T)$ είναι ο μοναδικός ΑΟΕΔ εκτιμητής του $g(\theta)$.*

Απόδειξη. Πρέπει να δείξουμε ότι

(α) ο $S^*(T)$ είναι αμερόληπτος εκτιμητής του $g(\theta)$,

(β) $Var_\theta S^*(T) \le Var_\theta S_1(\underset{\sim}{X})$, για κάθε $\theta \in \Theta$ και για κάθε αμερόληπτο εκτιμητή $S_1(\underset{\sim}{X})$ του $g(\theta)$,

(γ) δεν υπάρχει άλλος ΑΟΕΔ του $g(\theta)$.

(α) Η αμεροληψία του $S^*(T)$ προκύπτει από την αμεροληψία του $S(\underset{\sim}{X})$ και το Θεώρημα 6.2.1 (α).

(β) Εάν $\mathbb{V}ar_\theta S_1 = \infty$, τότε τετριμμένα το αποτέλεσμα ισχύει. Έστω $\mathbb{V}ar_\theta S_1 < \infty$. Θεωρούμε την Rao–Blackwell βελτίωση του $S_1(\underset{\sim}{X})$, $S_1^*(T) = \mathbb{E}(S_1 \mid T)$. Τότε και ο $S_1^*(T)$ είναι αμερόληπτος εκτιμητής του $g(\theta)$ (για τον ίδιο λόγο, όπως ο $S^*(T)$). Οι εκτιμητές $S^*(T)$ και $S_1^*(T)$ είναι, επιπλέον, συναρτήσεις της πλήρους στατιστικής συνάρτησης $T(\underset{\sim}{X})$, οπότε από την Πρόταση 6.3.2 (α) συμπεραίνουμε ότι $S^*(T) = S_1^*(T)$. Περαιτέρω, από το Θεώρημα 6.2.1 (β), $Var_\theta S_1^*(T) \le Var_\theta S_1(\underset{\sim}{X})$ για κάθε $\theta \in \Theta$, άρα και

$Var_\theta S^*(T) \leq Var_\theta S_1(\underset{\sim}{X})$.

(γ) Έστω ότι υπάρχει και άλλος ΑΟΕΔ εκτιμητής του $g(\theta)$, ο $S_0(\underset{\sim}{X})$, οπότε $\mathbb{V}ar_\theta S_0 = \mathbb{V}ar_\theta S^* < \infty$. Τότε, κατ' ανάγκη, ο $S_0(\underset{\sim}{X})$ είναι συνάρτηση της $T(\underset{\sim}{X})$, διαφορετικά η Rao–Blackwell βελτίωσή του, $S_0^*(T) = \mathbb{E}(S_0 \mid T)$ θα ήταν αμερόληπτος και θα είχε διασπορά γνησίως μικρότερη της διασποράς του $S_0(\underset{\sim}{X})$, λόγω της Πρότασης 6.2.2 (β), το οποίο είναι αδύνατο αφού ο $S_0(\underset{\sim}{X})$ είναι ΑΟΕΔ εκτιμητής. Οι $S^*(T)$ και $S_0(\underset{\sim}{X})$ ως αμερόληπτοι και συναρτήσεις της πλήρους στατιστικής συνάρτησης $T(\underset{\sim}{X})$, συμπίπτουν (Πρόταση 6.3.2 (α)). Άρα ο $S^*(T)$ είναι ο μοναδικός ΑΟΕΔ εκτιμητής του $g(\theta)$. □

Μια εναλλακτική μορφή του Θεωρήματος 6.3.3 δίνεται στην ακόλουθη πρόταση.

Πρόταση 6.3.4. *Έστω $T(\underset{\sim}{X})$ επαρκής και πλήρης στατιστική συνάρτηση και $S(T)$ ένας αμερόληπτος εκτιμητής του $g(\theta)$ που είναι συνάρτηση της $T(\underset{\sim}{X})$, με $\mathbb{V}ar_\theta S(T) < \infty$ για κάθε $\theta \in \Theta$. Τότε $S(T)$ είναι ο μοναδικός ΑΟΕΔ εκτιμητής του $g(\theta)$.*

Απόδειξη. Από το Θεώρημα 6.3.3 έχουμε ότι $S^* = \mathbb{E}(S \mid T)$ είναι ο μοναδικός ΑΟΕΔ εκτιμητής του $g(\theta)$. Επειδή όμως $S(T)$ είναι συνάρτηση της $T(\underset{\sim}{X})$, η Rao–Blackwell βελτίωσή του $S^*(T)$ συμπίπτει με τον $S(T)$. □

Παρατήρηση 6.3.1. Το Θεώρημα 6.3.3 και η Πρόταση 6.3.4 παρέχουν δύο διαφορετικούς τρόπους εφαρμογής της τεχνικής των Lehmann–Scheffé. Αρχικά, σε πρώτο στάδιο, και ο ένας τρόπος αλλά και ο άλλος απαιτούν την εύρεση επαρκούς και πλήρους στατιστικής συνάρτησης. Το Θεώρημα 6.3.3 δηλώνει ότι εάν περαιτέρω βρεθεί ένας *οποιοσδήποτε* αμερόληπτος εκτιμητής $S(\underset{\sim}{X})$ του $g(\theta)$, τότε $S^*(T) = \mathbb{E}(S \mid T)$ είναι ο ΑΟΕΔ εκτιμητής. Ο τρόπος αυτός απαιτεί εν συνεχεία τον υπολογισμό της δεσμευμένης μέσης τιμής. Η Πρόταση 6.3.4 δηλώνει ότι εάν βρεθεί *αμερόληπτος εκτιμητής του $g(\theta)$ που είναι συνάρτηση επαρκούς και πλήρους στατιστικής συνάρτησης* τότε αυτός είναι ο ΑΟΕΔ εκτιμητής του $g(\theta)$.

Δίνουμε στη συνέχεια μερικά παραδείγματα ΑΟΕΔ εκτιμητών. Μερικοί από αυτούς δεν καλύπτονται από την ανισότητα των Cramér–Rao, δηλαδή δεν είναι αποδοτικοί εκτιμητές.

Παράδειγμα 6.3.2. (**κατανομή Bernoulli–ΑΟΕΔ εκτιμητές**) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$, $n \geq 2$, ένα τυχαίο δείγμα από την κατανομή Bernoulli $\mathcal{B}(1,\theta)$, $\theta \in \Theta = (0,1)$. Θα βρούμε ΑΟΕΔ εκτιμητές των θ, θ^2, $\theta(1-\theta)$, θ^r, r ακέραιος, $1 \leq r \leq n$. Αναζητούμε πρώτα επαρκή και πλήρη στατιστική συνάρτηση $T(\underset{\sim}{X})$. Έχουμε

$$
\begin{aligned}
f(\underset{\sim}{x};\theta) = \prod_{i=1}^{n} f_1(x_i;\theta) &= \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}\mathrm{I}_{\{0,1\}}(x_i) \\
&= \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i}\prod_{i=1}^{n}\mathrm{I}_{\{0,1\}}(x_i).
\end{aligned}
$$

Επομένως από το παραγοντικό κριτήριο προκύπτει ότι η $T(\underset{\sim}{X}) = \sum_{i=1}^{n} X_i$ είναι επαρκής στατιστική συνάρτηση. Αποδεικνύουμε στη συνέχεια ότι είναι και πλήρης. Η κατανομή της $T(\underset{\sim}{X}) = \sum_{i=1}^{n} X_i$ είναι διωνυμική $\mathcal{B}(n,\theta)$. Έστω τώρα ότι $\mathbb{E}_\theta \phi(T) = 0$, για κάθε $\theta \in \Theta$. Τότε

$$
\begin{aligned}
\sum_{t=0}^{n} \phi(t)\binom{n}{t}\theta^t(1-\theta)^{n-t} = 0, &\quad \forall\, \theta \in (0,1) \quad \text{ή} \\
\sum_{t=0}^{n} \phi(t)\binom{n}{t}\left(\frac{\theta}{1-\theta}\right)^t = 0, &\quad \forall\, \theta \in (0,1) \quad \text{ή} \\
\sum_{t=0}^{n} \phi(t)\binom{n}{t}\rho^t = 0, &\quad \forall\, \rho > 0,
\end{aligned}
$$

όπου $\rho = \theta/(1-\theta)$ επειδή το σύνολο των τιμών της συνάρτησης $\theta/(1-\theta)$ είναι το $(0,\infty)$. Η τελευταία σχέση δηλώνει ότι κάθε αριθμός $\rho > 0$ είναι ρίζα του πολυωνύμου $\sum_{t=0}^{n}\phi(t)\binom{n}{t}x^t$ που είναι βαθμού το πολύ n. Ένα όμως πολυώνυμο βαθμού το πολύ n έχει το πολύ n διαφορετικές ρίζες, εκτός εάν είναι το μηδενικό πολυώνυμο. Άρα έχουμε

$$
\phi(t)\binom{n}{t} = 0, \quad \text{ή} \quad \phi(t) = 0, \quad \forall\, t = 0,1,\ldots,n,
$$

οπότε $\phi(T) = 0$ αφού το σύνολο τιμών της T είναι το $\{0, 1, \ldots, n\}$. Συνεπώς η στατιστική συνάρτηση $T(\underset{\sim}{X}) = \sum_{i=1}^{n} X_i$ είναι πλήρης.

(α) <u>ΑΟΕΔ εκτιμητής του θ</u>

Παρατηρούμε ότι $\mathbb{E}_\theta \bar{X} = \theta$, για κάθε $\theta \in (0, 1)$, άρα, από την Πρόταση 6.3.4, ο $\bar{X} = \frac{1}{n} T(\underset{\sim}{X})$ ως αμερόληπτος εκτιμητής του θ και συνάρτηση της επαρκούς και πλήρους στατιστικής συνάρτησης είναι ο ΑΟΕΔ εκτιμητής του θ. Επιπλέον, ο $\bar{X}$ είναι αποδοτικός εκτιμητής του θ, όπως είδαμε στο Παράδειγμα 5.2.9.

(β) <u>ΑΟΕΔ εκτιμητής του θ^2</u>

• Α΄ τρόπος (εφαρμογή της Πρότασης 6.3.4)

Παρατηρούμε ότι

$$\mathbb{E}_\theta(\bar{X}^2) = \mathbb{V}ar_\theta \bar{X} + (\mathbb{E}_\theta \bar{X})^2 = \frac{\theta(1-\theta)}{n} + \theta^2 = \frac{\theta}{n} + \frac{n-1}{n}\theta^2 = \frac{\mathbb{E}_\theta \bar{X}}{n} + \frac{n-1}{n}\theta^2.$$

Επομένως έχουμε,

$$\mathbb{E}_\theta \left(\bar{X}^2 - \frac{\bar{X}}{n} \right) = \frac{n-1}{n}\theta^2 \quad \text{και} \quad \mathbb{E}_\theta \big\{ (\bar{X}^2 - \frac{\bar{X}}{n}) \frac{n}{n-1} \big\} = \theta^2.$$

Άρα, $(\bar{X}^2 - \frac{\bar{X}}{n})\frac{n}{n-1} = \frac{T(T-1)}{n(n-1)}$ είναι αμερόληπτος εκτιμητής του θ^2 και συνάρτηση της επαρκούς και πλήρους στατιστικής συνάρτησης $T(\underset{\sim}{X}) = \sum_{i=1}^{n} X_i$. Συνεπώς είναι ο ΑΟΕΔ εκτιμητής του θ^2 (αλλά δεν είναι αποδοτικός εκτιμητής). Σημειώνουμε ότι, όπως ήδη έχει αναφερθεί στο Παράδειγμα 4.2.3, το σκεπτικό για τον υπολογισμό της μέσης τιμής $\mathbb{E}_\theta(\bar{X}^2)$ βασίζεται στην αρχή της αντικατάστασης: ο $\bar{X}^2$ είναι ένας «λογικός» εκτιμητής του θ^2 αφού ο $\bar{X}$ είναι εκτιμητής του θ.

• Β΄ τρόπος (εφαρμογή του Θεωρήματος 6.3.3)

Παρατηρούμε ότι θ^2 είναι η πιθανότητα «επιτυχίας» στις δύο πρώτες «δοκιμές». Ορίζουμε λοιπόν

$$S(\underset{\sim}{X}) = \begin{cases} 1, & \text{εάν } X_1 = X_2 = 1 \\ 0, & \text{διαφορετικά} \end{cases}$$

οπότε

$$\begin{aligned} \mathbb{E}_\theta S &= 1 \cdot \mathbb{P}_\theta(S = 1) + 0 \cdot \mathbb{P}_\theta(S = 0) = \mathbb{P}_\theta(S = 1) = \mathbb{P}_\theta(X_1 = 1, X_2 = 1) \\ &= \mathbb{P}_\theta(X_1 = 1)\mathbb{P}_\theta(X_2 = 1) = \theta \cdot \theta = \theta^2. \end{aligned}$$

Άρα $S(\underset{\sim}{X})$ είναι ένας αμερόληπτος εκτιμητής του θ^2 και επομένως $S^* = \mathbb{E}(S \mid T)$ είναι ο ΑΟΕΔ εκτιμητής του θ^2. Υπολογίζουμε τη δεσμευμένη μέση τιμή, για την τιμή t της $T(\underset{\sim}{X})$, $t \in \{0, 1, \dots, n\}$. Έχουμε,

$$\begin{aligned}
\mathbb{E}(S \mid T = t) &= 1 \cdot \mathbb{P}(S = 1 \mid T = t) + 0 \cdot \mathbb{P}(S = 0 \mid T = t) \\
&= \mathbb{P}(S = 1 \mid T = t) = \frac{\mathbb{P}_\theta(S = 1, T = t)}{\mathbb{P}_\theta(T = t)} \\
&= \frac{\mathbb{P}_\theta(X_1 = 1, X_2 = 1, \sum_{i=1}^n X_i = t)}{\mathbb{P}_\theta(\sum_{i=1}^n X_i = t)} \\
&= \frac{\mathbb{P}_\theta(X_1 = 1, X_2 = 1, \sum_{i=3}^n X_i = t - 2)}{\mathbb{P}_\theta(\sum_{i=1}^n X_i = t)} \\
&= \frac{\mathbb{P}_\theta(X_1 = 1)\mathbb{P}_\theta(X_2 = 1)\mathbb{P}_\theta(\sum_{i=3}^n X_i = t - 2)}{\mathbb{P}_\theta(\sum_{i=1}^n X_i = t)} \quad (n \geq 3) \\
&= \frac{\theta \cdot \theta \binom{n-2}{t-2} \theta^{t-2} (1-\theta)^{(n-2)-(t-2)}}{\binom{n}{t} \theta^t (1-\theta)^{n-t}}, \quad t \geq 2 \\
&= \frac{\binom{n-2}{t-2}}{\binom{n}{t}} = \frac{t(t-1)}{n(n-1)}, \quad t \geq 2.
\end{aligned}$$

Η τρίτη από το τέλος ισότητα ισχύει επειδή $\sum\limits_{i=3}^{n} X_i \sim \mathcal{B}(n-2, \theta)$, $n \geq 3$, ενώ για $t = 0, 1$, $\mathbb{E}(S \mid T = t) = 0$. Άρα $\mathbb{E}(S \mid T = t) = \frac{t(t-1)}{n(n-1)}$, $t = 0, 1, \dots, n$, δηλαδή $S^* = \mathbb{E}(S \mid T) = \frac{T(T-1)}{n(n-1)}$.

(γ) <u>ΑΟΕΔ εκτιμητής της διασποράς $\theta(1-\theta)$</u>

Παρατηρούμε ότι $\theta(1-\theta) = \theta - \theta^2$. Συνεπώς, χρησιμοποιώντας τους ΑΟΕΔ εκτιμητές $\frac{T}{n}$ και $\frac{T(T-1)}{n(n-1)}$ των θ, θ^2 αντίστοιχα έχουμε ότι

$$\begin{aligned}
\mathbb{E}_\theta \left(\frac{T}{n} - \frac{T(T-1)}{n(n-1)} \right) &= \mathbb{E}_\theta \frac{T}{n} - \mathbb{E}_\theta \frac{T(T-1)}{n(n-1)} \\
&= \theta - \theta^2 = \theta(1-\theta), \forall\, \theta \in (0, 1).
\end{aligned}$$

Επομένως, $\frac{T}{n} - \frac{T(T-1)}{n(n-1)} = \frac{n\bar{X}(1-\bar{X})}{n-1}$ ως αμερόληπτος εκτιμητής του $\theta(1-\theta)$ και συνάρτηση της επαρκούς και πλήρους στατιστικής συνάρτησης $T(\underset{\sim}{X}) = \sum\limits_{i=1}^{n} X_i$ είναι ο ΑΟΕΔ εκτιμητής του $\theta(1-\theta)$ (αλλά δεν είναι αποδοτικός εκτιμητής). Σημειώνουμε ότι ο ΑΟΕΔ εκτιμητής συμπίπτει με τη δειγματική διασπορά επειδή

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right\}$$
$$= \frac{1}{n-1}\left\{\sum_{i=1}^{n} X_i - n\bar{X}^2\right\} = \frac{n}{n-1}\bar{X}(1-\bar{X})$$

όπου τέθηκε $X_i^2 = X_i$ αφού $X_i = 0, 1$.

(δ) ΑΟΕΔ εκτιμητής του θ^r

Εργαζόμενοι όπως παραπάνω για την εύρεση του ΑΟΕΔ εκτιμητή του θ^2 με εφαρμογή του Θεωρήματος 6.3.3, μπορούμε να αποδείξουμε ότι ο ΑΟΕΔ εκτιμητής του θ^r είναι $\frac{T(T-1)\cdots(T-r+1)}{n(n-1)\cdots(n-r+1)}$.

Παράδειγμα 6.3.3. (ομοιόμορφη κατανομή με ένα άκρο άγνωστο– ΑΟΕΔ εκτιμητής της μέσης τιμής–Κ.Φ. C-R) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την ομοιόμορφη κατανομή $\mathcal{U}(0,\theta)$, $\theta \in \Theta = (0, \infty)$ με $f_1(x_1;\theta) = \frac{1}{\theta}$, $0 < x_1 < \theta$. Θα βρούμε τον ΑΟΕΔ εκτιμητή της μέσης τιμής $\frac{\theta}{2}$.

Παρατηρούμε πρώτα ότι η στατιστική συνάρτηση $T(\underset{\sim}{X}) = X_{(n)} = \max(X_1, \ldots, X_n)$ είναι επαρκής (βλέπε Παράδειγμα 6.1.5). Θα αποδείξουμε επιπλέον ότι $T(\underset{\sim}{X})$ είναι και πλήρης. Αρχικά, χρειάζεται η πυκνότητα της στατιστικής συνάρτησης $T(\underset{\sim}{X})$ που δίνεται από τον γενικό τύπο (βλέπε Πρόταση 1.8.7)

$$f_T(t;\theta) = n\,[F_1(t;\theta)]^{n-1}\, f_1(t;\theta).$$

Επειδή

$$F_1(t;\theta) = \int_0^t f_1(x_1;\theta)\,\mathrm{d}x_1 = \frac{t}{\theta}, \quad 0 < t < \theta,$$

τελικά έχουμε

$$f_T(t;\theta) = \frac{n}{\theta^n} t^{n-1}, \quad 0 < t < \theta.$$

Έστω τώρα ότι $\mathbb{E}_\theta \phi(T) = 0$, για κάθε $\theta > 0$. Θα δείξουμε ότι $\phi(T) = 0$ δηλαδή $\phi(t) = 0$ για κάθε τιμή t της $T(\underset{\sim}{X})$, δηλαδή $\phi(t) = 0$, για κάθε $t > 0$, επειδή το σύνολο τιμών της $T(\underset{\sim}{X})$ είναι το $(0, \theta)$ και το θ μπορεί να είναι οποιοσδήποτε θετικός αριθμός. Έχουμε

$$\int_0^\theta \phi(t) f_T(t;\theta)\,\mathrm{d}t = 0 \text{ ή } \int_0^\theta \phi(t)\frac{n}{\theta^n}t^{n-1}\,\mathrm{d}t = 0 \text{ ή } \int_0^\theta \phi(t)t^{n-1}\,\mathrm{d}t = 0, \quad \forall\,\theta > 0.$$

Η σχέση αυτή, από τη θεωρία ολοκλήρωσης, συνεπάγεται $\phi(T) = 0$. Θεωρώντας ότι η ϕ είναι συνεχής, μπορεί να δοθεί η εξής απλή απόδειξη της πληρότητας. Παραγωγίζοντας ως προς θ, προκύπτει ότι $\phi(\theta)\theta^{n-1} = 0,\ \theta > 0$ οπότε $\phi(\theta) = 0,\ \theta > 0$, δηλαδή $\phi(t) = 0,\ t > 0$. Επομένως, $T(\underset{\sim}{X}) = X_{(n)}$ είναι επαρκής και πλήρης. Παρατηρούμε ότι

$$\mathbb{E}_\theta X_{(n)} = \int_o^\theta t\frac{n}{\theta^n}t^{n-1}\,\mathrm{d}t = \frac{n}{n+1}\theta$$

και επομένως $\mathbb{E}_\theta\left(\frac{n+1}{2n}T\right) = \frac{\theta}{2}$, για κάθε $\theta > 0$. Συνεπώς, βάσει της Πρότασης 6.3.4, $\frac{n+1}{2n}T = \frac{n+1}{2n}X_{(n)}$ ως αμερόληπτος εκτιμητής του $\frac{\theta}{2}$ και συνάρτηση της επαρκούς και πλήρους στατιστικής συνάρτησης είναι ο ΑΟΕΔ του $\frac{\theta}{2}$. Επιπλέον, ο ΑΟΕΔ εκτιμητής του θ είναι ο $\frac{n+1}{n}T = \frac{n+1}{n}X_{(n)}$. Φαίνεται να ήταν «θέμα τύχης» ότι η μέση τιμή $\mathbb{E}_\theta X_{(n)}$ βρέθηκε να είναι γραμμική συνάρτηση της παραμέτρου θ. Κι' όμως δεν ήταν, αρκεί να παρατηρήσουμε ότι η $\frac{T}{\theta}$ έχει κατανομή που δεν εξαρτάται από το θ, συνεπώς και η μέση τιμή της $\frac{T}{\theta}$ δεν εξαρτάται από το θ. Επομένως, $\mathbb{E}_\theta\left(\frac{T}{\theta}\right) = c$ οπότε $\mathbb{E}_\theta T = c\theta$ δηλαδή η $\mathbb{E}_\theta T$ είναι γραμμική συνάρτηση του θ. Διαφορετικά, η αναζήτηση αμερόληπτου εκτιμητή του $\frac{\theta}{2}$ που είναι συνάρτηση της στατιστικής συνάρτησης $T(\underset{\sim}{X}) = X_{(n)}$ μπορεί να αντιμετωπιστεί με την επίλυση της συναρτησιακής εξίσωσης $\mathbb{E}_\theta S^*(T) = \frac{\theta}{2}$, για κάθε $\theta > 0$ (με άγνωστο, δηλαδή, τη συνάρτηση $S^*(T)$). Για την επίλυση της παρατηρούμε ότι

$$\mathbb{E}_\theta S^*(T) = \theta \text{ ή } \int_0^\theta S^*(t)\frac{n}{\theta^n}t^{n-1}\,\mathrm{d}t = \frac{\theta}{2} \text{ ή } \int_0^\theta S^*(t)t^{n-1}\,\mathrm{d}t = \frac{\theta^{n+1}}{2n},\ \theta > 0.$$

Υποθέτοντας συνέχεια της $S^*(t)$ και παραγωγίζοντας ως προς θ έχουμε

$$S^*(\theta)\theta^{n-1} = (1+n)\frac{\theta^n}{2} \text{ ή } S^*(\theta) = \frac{n+1}{2n}\theta,\ \theta > 0,$$

δηλαδή $S^*(T) = \frac{n+1}{2n}T = \frac{n+1}{2n}X_{(n)}$ (που είναι συνεχής).

Στο Παράδειγμα 6.2.1 είδαμε ότι ο δειγματικός μέσος $\bar{X}$ είναι μη αποδεκτός εκτιμητής της μέσης τιμής $\frac{\theta}{2}$ επειδή δεν είναι συνάρτηση της επαρκούς στατιστικής συνάρτησης $X_{(n)}$ (βλέπε Πρόταση 6.2.2). Καλύτερος εκτιμητής με κριτήριο το ΜΤΣ είναι η Rao–Blackwell βελτίωσή του, $\mathbb{E}(\bar{X} \mid X_{(n)})$ που όμως δεν είχε υπολογιστεί σε κλειστή μορφή. Ο $\bar{X}$ είναι αμερόληπτος εκτιμητής του $\frac{\theta}{2}$, άρα από το Θεώρημα των Lehmann–Scheffé, ο $\mathbb{E}(\bar{X} \mid X_{(n)})$ είναι ο ΑΟΕΔ εκτιμητής του $\frac{\theta}{2}$, τον οποίο έχουμε ήδη βρει. Λόγω λοιπόν της μοναδικότητας του ΑΟΕΔ εκτιμητή έχουμε $\mathbb{E}(\bar{X} \mid X_{(n)}) = \frac{n+1}{2n} X_{(n)}$.
Με το ίδιο σκεπτικό, συμπεραίνουμε αμέσως ότι $\mathbb{E}(X_1 \mid X_{(n)}) = \frac{n+1}{2n} X_{(n)}$. Για σύγκριση, κατ' ευθείαν υπολογισμός της δεσμευμένης μέσης τιμής απαιτεί την εύρεση της δεσμευμένης κατανομής της X_1 δοθέντος ότι $X_{(n)} = t$. Επειδή $X_{(n)}$ είναι η μέγιστη παρατήρηση, αν $X_{(n)} = t$ τότε η X_1 έχει σύνολο τιμών το $(0, t]$. Επιπλέον, $\mathbb{P}(X_1 = t \mid X_{(n)} = t) = \frac{1}{n}$, αφού λόγω συμμετρίας οποιαδήποτε από τις n παρατηρήσεις $X_1, X_2, \ldots, X_n$ έχει την ίδια πιθανότητα να είναι η μέγιστη. Απομένει πιθανότητα $1 - \frac{1}{n}$ που κατανέμεται *ομοιόμορφα* στο ανοικτό διάστημα $(0, t)$. Η ομοιόμορφη κατανομή έχει σταθερή πυκνότητα, άρα

$$f_{X_1|X_{(n)}}(x_1 \mid t) = c, \quad 0 < x_1 < t$$

και επειδή πρέπει

$$\int_0^t f_{X_1|X_{(n)}}(x_1 \mid t)\, dx_1 = 1 - \frac{1}{n} \quad \text{(και όχι 1)}$$

έχουμε

$$\int_0^t c\, dx_1 = 1 - \frac{1}{n} \quad \text{ή} \quad c = \frac{1}{t}\left(1 - \frac{1}{n}\right).$$

Τελικά, η δεσμευμένη κατανομή της X_1 δοθέντος ότι $X_{(n)} = t$ είναι μια μεικτή κατανομή (διακριτή και συνεχής) με πυκνότητα

$$f_{X_1|X_{(n)}}(x_1 \mid t) = \begin{cases} \frac{1}{n}, & x_1 = t \\ \left(1 - \frac{1}{n}\right)\frac{1}{t}, & 0 < x_1 < t \end{cases}$$

Επομένως έχουμε

$$\mathbb{E}(X_1 \mid X_{(n)} = t) = t \cdot \frac{1}{n} + \int_0^t x_1 \frac{1}{t}\left(1 - \frac{1}{n}\right) dx_1 = \frac{n+1}{2n} t, \quad \forall\, t > 0,$$

από όπου προκύπτει $\mathbb{E}(X_1 \mid X_{(n)}) = \frac{n+1}{2n} X_{(n)}$ (βλέπε Ενότητα 1.9).
Η διασπορά του ΑΟΕΔ εκτιμητή $S^* = \frac{n+1}{2n} X_{(n)}$ υπολογίζεται εύκολα και είναι $Var_\theta S^* = \frac{\theta^2}{4n(n+2)}$. Θα συγκρίνουμε τώρα τη διασπορά $Var_\theta S^*$ με το αντίστοιχο Κ.Φ. C-R. Η πυκνότητα του $\underset{\sim}{X}$ είναι

$$f(\underset{\sim}{x};\theta) = \frac{1}{\theta^n}, \quad 0 < x_i < \theta,$$

οπότε έχουμε

$$\ln f(\underset{\sim}{X};\theta) = -n \ln \theta, \quad \tfrac{\partial}{\partial\theta} \ln f(\underset{\sim}{X};\theta) = -\frac{n}{\theta},$$

$$I(\theta) = \mathbb{E}_\theta\left[\left(\tfrac{\partial}{\partial\theta} \ln f(\underset{\sim}{X};\theta)\right)^2\right] = \mathbb{E}\left(\frac{n^2}{\theta^2}\right) = \frac{n^2}{\theta^2},$$

και επομένως το κάτω φράγμα των Cramér–Rao για τη διασπορά αμερόληπτων εκτιμητών του $g(\theta) = \frac{\theta}{2}$ δίνεται από τη σχέση

$$\text{Κ.Φ. C-R} = \frac{\left(g'(\theta)\right)^2}{I(\theta)} = \frac{\theta^2}{4n^2}.$$

Παρατηρούμε λοιπόν ότι $Var_\theta(S^*)$ <Κ.Φ. C-R που δεν αντίκειται όμως στην ανισότητα (5.1) ή (5.2) των Cramér–Rao επειδή η ομοιόμορφη κατανομή $\mathcal{U}(0,\theta)$ δεν ικανοποιεί την Ι2 (ούτε τις Ι3, Ι4).

Παράδειγμα 6.3.4. (εκθετική κατανομή–ΑΟΕΔ εκτιμητής του θ^r) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την εκθετική κατανομή $\mathcal{E}(\theta)$, $\theta \in \Theta = (0,\infty)$. Θα βρούμε τον ΑΟΕΔ εκτιμητή του θ^r, $r > 0$. Υπενθυμίζουμε ότι $\mathbb{E}_\theta X_1 = \theta$ και $Var_\theta X_1 = \theta^2$.
Από το παραγοντικό κριτήριο είναι εύκολο να δειχθεί ότι η $T(\underset{\sim}{X}) = \sum_{i=1}^{n} X_i$ είναι επαρκής στατιστική συνάρτηση. Επιπλέον, θα δείξουμε ότι η $T(\underset{\sim}{X})$ είναι και πλήρης. Η κατανομή της $T(\underset{\sim}{X}) = \sum_{i=1}^{n} X_i$ είναι Γάμμα $\mathcal{G}(n,\theta)$,

δηλαδή η πυκνότητα της $T(\underset{\sim}{X})$ είναι $f_T(t;\theta) = \frac{1}{(n-1)!\theta^n} t^{n-1} e^{-\frac{t}{\theta}},\ t > 0$. Έστω $\mathbb{E}_\theta \phi(T) = 0$, για κάθε $\theta > 0$. Τότε

$$\int_0^\infty \phi(t) f_T(t;\theta)\,\mathrm{d}t = 0,\ \theta > 0 \quad \text{ή} \quad \int_0^\infty \phi(t) \frac{1}{(n-1)!\theta^n} t^{n-1} e^{-\frac{t}{\theta}}\,\mathrm{d}t = 0,\ \theta > 0$$

$$\text{ή} \quad \int_0^\infty \phi(t) t^{n-1} e^{-\frac{t}{\theta}}\,\mathrm{d}t = 0,\ \theta > 0.$$

Η τελευταία σχέση δηλώνει ότι ο μετασχηματισμός Laplace της συνάρτησης $\phi(t)t^{n-1}$ στο σημείο $-\frac{1}{\theta}$ είναι μηδέν. Άρα λόγω της μοναδικότητας του μετασχηματισμού Laplace, η συνάρτηση αυτή είναι 0, δηλαδή $\phi(t)t^{n-1} = 0,\ t > 0$, ή $\phi(t) = 0$, για κάθε t στο σύνολο τιμών της $T(\underset{\sim}{X})$, $(0, \infty)$. Επομένως $T(\underset{\sim}{X}) = \sum_{i=1}^{n} X_i$ είναι πλήρης. Αναζητούμε τώρα συνάρτηση της $T(\underset{\sim}{X})$ που είναι αμερόληπτος εκτιμητής του θ^r. Ας υπολογίσουμε (δοκιμαστικά) τη μέση τιμή $\mathbb{E}_\theta T$. Έχουμε $\mathbb{E}_\theta T = n\theta$, οπότε $\mathbb{E}_\theta\left(\frac{T}{n}\right) = \theta$ άρα $\frac{T}{n} = \bar{X}$ είναι ο ΑΟΕΔ εκτιμητής του θ λόγω της Πρότασης 6.3.4. Αφού $\bar{X}$ εκτιμά το θ, ας δοκιμάσουμε $(\frac{T}{n})^r$ ως εκτιμητή του θ^r. Τότε

$$\begin{aligned}
\mathbb{E}_\theta\Big(\frac{T}{n}\Big)^r = \frac{1}{n^r}\mathbb{E}_\theta T^r &= \frac{1}{n^r}\int_0^\infty t^r \frac{1}{(n-1)!\theta^n} t^{n-1} e^{-\frac{t}{\theta}}\,\mathrm{d}t \\
&= \frac{1}{n^r(n-1)!\theta^n}\int_0^\infty t^{r+n-1} e^{-\frac{t}{\theta}}\,\mathrm{d}t \\
&= \frac{1}{n^r(n-1)!\theta^n}\Gamma(r+n)\theta^{r+n} = \frac{\Gamma(r+n)}{n^r(n-1)!}\theta^r.
\end{aligned}$$

Επομένως, έχουμε $\mathbb{E}_\theta\big(\frac{(n-1)!}{\Gamma(r+n)}T^r\big) = \theta^r$, $\theta > 0$. Συνεπώς, $S(T) = \frac{(n-1)!}{\Gamma(r+n)}T^r$ ως αμερόληπτος εκτιμητής του θ^r και συνάρτηση της επαρκούς και πλήρους στατιστικής συνάρτησης T είναι ο ΑΟΕΔ εκτιμητής του θ^r. Σημειώνουμε, επιπλέον, ότι για $r \neq 1$ ο ΑΟΕΔ εκτιμητής δεν είναι αποδοτικός.

Παράδειγμα 6.3.5. (Διακριτός παραμετρικός χώρος) Η παρατήρηση X έχει κατανομή που δίνεται στον παρακάτω πίνακα για $\theta \in \Theta = \{0, 1\}$.

	$X = 1$	$X = 2$
$\theta = 0$	$2/3$	$1/3$
$\theta = 1$	$1/3$	$2/3$

Θα βρούμε τον ΑΟΕΔ εκτιμητή του θ. Επειδή το δείγμα $\underset{\sim}{X}$ είναι τετριμμένα επαρκής στατιστική συνάρτηση (Παρατήρηση 6.1.1) και στη συγκεκριμένη περίπτωση έχουμε μόνο μία παρατήρηση, αυτή είναι επαρκής, δηλαδή $T(X) = X$ είναι επαρκής στατιστική συνάρτηση.

Επιπλέον, θα δείξουμε ότι η $T(X) = X$ είναι πλήρης. Έστω

$$\mathbb{E}_\theta \phi(X) = 0, \quad \text{για } \theta \in \{0, 1\}$$

ή, ισοδύναμα,

$$\sum_{x=1}^{2} \phi(x) \mathbb{P}_\theta(X = x) = 0 \ \text{ για } \theta \in \{0, 1\}.$$

Για $\theta = 0$, έχουμε

$$\phi(1)\mathbb{P}_0(X = 1) + \phi(2)\mathbb{P}_0(X = 2) = 0 \ \text{ ή } \ \phi(1)\frac{2}{3} + \phi(2)\frac{1}{3} = 0.$$

Ομοίως, για $\theta = 1$ παίρνουμε,

$$\phi(1)\frac{1}{3} + \phi(2)\frac{2}{3} = 0 \ \text{ ή } \ \phi(1)\frac{1}{3} + \phi(2)\frac{2}{3} = 0.$$

Λύνοντας το σύστημα των δύο εξισώσεων ως προς $\phi(1)$ και $\phi(2)$, προκύπτει $\phi(1) = \phi(2) = 0$, δηλαδή $\phi(x) = 0$, για κάθε τιμή x της παρατήρησης X (που έχει σύνολο τιμών το $\{1, 2\}$). Επομένως η $T(X) = X$ είναι πλήρης.

Αναζητούμε τώρα συνάρτηση της παρατήρησης X που είναι αμερόληπτος εκτιμητής του θ. Θα «δοκιμάσουμε» ένα γραμμικό μετασχηματισμό, $\alpha X + \beta$ και θα βρούμε τις σταθερές α και β έτσι ώστε να ισχύει η σχέση

$$\mathbb{E}_\theta(\alpha X + \beta) = \theta, \quad \text{για} \quad \theta \in \{1, 2\}.$$

Για $\theta = 0$, η σχέση αυτή γράφεται

$$\alpha \mathbb{E}_{\theta=0} X + \beta = 0 \ \text{ ή } \ \alpha\left(1 \cdot \frac{2}{3} + 2 \cdot \frac{1}{3}\right) + \beta = 0 \ \text{ ή } \ \frac{4}{3}\alpha + \beta = 0.$$

Ανάλογα, για $\theta = 1$ προκύπτει

$$\alpha\left(1 \cdot \frac{1}{3} + 2 \cdot \frac{2}{3}\right) + \beta = 1 \ \text{ ή } \ \frac{5}{3}\alpha + \beta = 1.$$

Επομένως, $\alpha = 3$, $\beta = -4$, δηλαδή η $3X - 4$ είναι ο ΑΟΕΔ εκτιμητής του θ ως αμερόληπτος και συνάρτηση της επαρκούς και πλήρους στατιστικής συνάρτησης X.

Παράδειγμα 6.3.6. (**$\mathcal{U}(\theta, \theta+1)$-ελάχιστη επαρκής αλλά μη πλήρης στατιστική συνάρτηση**) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την ομοιόμορφη κατανομή $\mathcal{U}(\theta, \theta+1)$ με $f_1(x_1; \theta) = 1$, $\theta < x_1 < \theta + 1$, $\theta \in \Theta = \mathbb{R}$. Τότε $T(\underset{\sim}{X}) = (X_{(1)}, X_{(n)})$ είναι (ελάχιστη) επαρκής (βλέπε Παράδειγμα 6.1.6) αλλά δεν είναι πλήρης. Πράγματι, εάν

$$Y_1 = X_1 - \theta, \ldots, Y_n = X_n - \theta,$$

τότε $\underset{\sim}{Y} = (Y_1, \ldots, Y_n)$ είναι ένα τυχαίο δείγμα από την ομοιόμορφη κατανομή $\mathcal{U}(0, 1)$, όπως προκύπτει αμέσως από την Πρόταση 1.7.1. Επομένως, η κατανομή των $Y_1, \ldots, Y_n$ δεν εξαρτάται από το θ. Παρατηρούμε τώρα ότι $X_1 = Y_1 + \theta, \ldots, X_n = Y_n + \theta$ οπότε $X_{(1)} = Y_{(1)} + \theta, X_{(n)} = Y_{(n)} + \theta$, όπου $Y_{(1)} = \min(Y_1, \ldots, Y_n)$ και $Y_{(n)} = \max(Y_1, \ldots, Y_n)$. Συνεπώς $\mathbb{E}_\theta(X_{(n)} - X_{(1)}) = \mathbb{E}(Y_{(n)} - Y_{(1)}) = c$, με c σταθερά που δεν εξαρτάται από το θ αφού η κατανομή της τυχαίας μεταβλητής $Y_{(n)} - Y_{(1)}$ δεν εξαρτάρται από το θ. Τελικά, έχουμε $\mathbb{E}_\theta(X_{(n)} - X_{(1)} - c) = 0$, για κάθε $\theta \in \Theta$. Άρα η στατιστική συνάρτηση $\phi(T) = X_{(n)} - X_{(1)} - c$ πληροί τη σχέση $\mathbb{E}_\theta \phi(T) = 0$, για κάθε $\theta \in \Theta$ και είναι $\phi(T) \neq 0$. Επομένως η στατιστική συνάρτηση T δεν είναι πλήρης.

Παράδειγμα 6.3.7. (**μετατοπισμένη εκθετική κατανομή-επάρκεια και πληρότητα**) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την κατανομή με πυκνότητα $f_1(x_1; \theta) = e^{-(x_1 - \theta)}$, $x_1 \geq \theta$, $\theta \in \Theta = \mathbb{R}$. Θα δείξουμε ότι $T(\underset{\sim}{X}) = X_{(1)} = \min(X_1, \ldots, X_n)$ είναι επαρκής και πλήρης.
Έχουμε $f_1(x_1; \theta) = e^{-(x_1 - \theta)} \mathrm{I}_{[\theta, \infty)}(x_1)$ (χρησιμοποιώντας τη δείκτρια συνάρτηση). Επομένως, η πυκνότητα του $\underset{\sim}{X} = (X_1, \ldots, X_n)$ είναι

$$f(\underset{\sim}{x}; \theta) = \prod_{i=1}^{n} f_1(x_i; \theta) = e^{-\sum x_i + n\theta} \prod_{i=1}^{n} \mathrm{I}_{[\theta, \infty)}(x_i).$$

Παρατηρούμε ότι

$$\begin{aligned}\prod_{i=1}^{n} \mathrm{I}_{[\theta,\infty)}(x_i) &= \begin{cases} 1, & x_i \geq \theta,\ i = 1, \ldots, n \\ 0, & \text{διαφορετικά} \end{cases} \\ &= \begin{cases} 1, & x_{(1)} \geq \theta, \\ 0, & \text{διαφορετικά} \end{cases} \\ &= \mathrm{I}_{[\theta,\infty)}(x_{(1)}),\end{aligned}$$

όπου $x_{(1)} = \min(x_1, \ldots, x_n)$. Συνεπώς έχουμε

$$f(\underset{\sim}{x}; \theta) = e^{-\sum_i x_i + n\theta} \mathrm{I}_{[\theta,\infty)}(x_{(1)}) = q(x_{(1)}, \theta) h(\underset{\sim}{x}),$$

όπου $q(x_{(1)}, \theta) = e^{n\theta} \mathrm{I}_{[\theta,\infty)}(x_{(1)})$ και $h(\underset{\sim}{x}) = e^{-\sum_i x_i}$. Άρα από το παραγοντικό κριτήριο η στατιστική συνάρτηση $T(\underset{\sim}{X}) = X_{(1)} = \min(X_1, \ldots, X_n)$ είναι επαρκής. Για να δείξουμε ότι η στατιστική συνάρτηση T είναι πλήρης χρειαζόμαστε την πυκνότητα της, $f_T(t; \theta)$, που δίνεται από τον τύπο (βλέπε Πρόταση 1.8.7)

$$f_T(t; \theta) = n \left[1 - F_1(t; \theta)\right]^{n-1} f_1(t; \theta).$$

Είναι τώρα

$$\begin{aligned}F_1(t; \theta) = \mathbb{P}_\theta(X_1 \leq t) &= \int_\theta^t f_1(x_1; \theta) \, \mathrm{d}x_1 \\ &= \int_\theta^t e^{-(x_1 - \theta)} \, \mathrm{d}x_1 = 1 - e^{-(t-\theta)},\ t \geq \theta.\end{aligned}$$

Επομένως

$$f_T(t; \theta) = n e^{-(n-1)(t-\theta)} e^{-(t-\theta)} = n e^{-n(t-\theta)},\ t \geq \theta.$$

Έστω τώρα ότι $\mathbb{E}_\theta \phi(T) = 0$, για κάθε $\theta \in \Theta = \mathbb{R}$. Τότε

$$\int_\theta^\infty \phi(t) n e^{-n(t-\theta)} \, \mathrm{d}t = 0,\ \theta \in \mathbb{R} \text{ ή } \int_\theta^\infty \phi(t) n e^{-nt} \, \mathrm{d}t = 0,\ \theta \in \mathbb{R}.$$

Όπως στο Παράδειγμα 6.3.3, η τελευταία σχέση συνεπάγεται ότι $\phi(t) = 0$, $t \in \mathbb{R}$. Μία απλή απόδειξη αυτού μπορεί να δοθεί θεωρώντας ότι η ϕ είναι συνεχής. Τότε παραγωγίζοντας ως προς θ έχουμε

$$-n\phi(\theta)e^{-n\theta} = 0,\ \forall\ \theta \in \mathbb{R} \text{ ή } \phi(\theta) = 0,\ \forall\ \theta \in \mathbb{R} \text{ ή } \phi(t) = 0,\ \forall\ t \in \mathbb{R},$$

δηλαδή $\phi(t) = 0$ για κάθε τιμή t της στατιστικής συνάρτησης $T(\underset{\sim}{X}) = X_{(1)}$. Επομένως η $T(\underset{\sim}{X}) = X_{(1)}$ είναι πλήρης.

Παράδειγμα 6.3.8. (**το δείγμα $\underset{\sim}{X}$ είναι επαρκής αλλά δεν είναι πλήρης στατιστική συνάρτηση**) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$, $n \geq 2$, ένα τυχαίο δείγμα από την κατανομή Bernoulli $\mathcal{B}(1, \theta)$, $\theta \in \Theta = (0, 1)$. Τότε η στατιστική συνάρτηση $T(\underset{\sim}{X}) = \underset{\sim}{X} = (X_1, \ldots, X_n)$, είναι επαρκής βάσει της Παρατήρησης 6.1.1, αλλά δεν είναι πλήρης. Πράγματι, εάν ορίσουμε $\phi(T) = X_2 - X_1$ τότε

$$\mathbb{E}_\theta\phi(T) = \mathbb{E}_\theta(X_2 - X_1) = \mathbb{E}_\theta X_2 - \mathbb{E}_\theta X_1 = \theta - \theta = 0,\ \forall\ \theta \in (0, 1),$$

ενώ $\phi(T)$ δεν είναι η σταθερά συνάρτηση μηδέν όπως απαιτεί ο ορισμός της πληρότητας. Σημειώνουμε ότι η κατανομή Bernoulli δεν παίζει κανένα ουσιαστικό ρόλο στην απόδειξη και μπορεί να αντικατασταθεί με οποιαδήποτε άλλη.

Παράδειγμα 6.3.9. (**κανονική κατανομή με γνωστό συντελεστή μεταβλητότητας - μη πληρότητα επαρκούς στατιστικής συνάρτησης**) Έστω ένα τυχαίο δείγμα $\underset{\sim}{X} = (X_1, X_2, \ldots, X_n)$, $n \geq 2$, από την κανονική κατανομή $\mathcal{N}(\theta, \theta^2)$, με άγνωστο $\theta \in \Theta = (0, \infty)$. Θα δείξουμε ότι η στατιστική συνάρτηση $T(\underset{\sim}{X}) = \left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$ είναι επαρκής, αλλά δεν είναι πλήρης. Έχουμε

$$\begin{aligned}
f(\underset{\sim}{x}; \theta) &= \prod_{i=1}^{n} f(x_i; \theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\theta} e^{-\frac{(x_i - \theta)^2}{2\theta^2}} \\
&= \frac{1}{(2\pi)^{n/2}\theta^n} \exp\left\{-\frac{1}{2\theta^2}\sum_{i=1}^{n}(x_i - \theta)^2\right\} \\
&= \exp\left\{-\frac{n}{2}\ln 2\pi - n\ln\theta - \frac{1}{2\theta^2}\sum_{i=1}^{n} x_i^2 - \frac{n}{2} + \frac{1}{\theta}\sum_{i=1}^{n} x_i\right\},\ \underset{\sim}{x} \in \mathbb{R}^n.
\end{aligned}$$

Θέτουμε $q(T(\underset{\sim}{x});\theta) = \exp\left\{-n\ln\theta - \frac{1}{2\theta^2}\sum_{i=1}^{n} x_i^2 + \frac{1}{\theta}\sum_{i=1}^{n} x_i\right\}$, όπου $T(\underset{\sim}{x}) = \left(\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i^2\right)$ και $h(\underset{\sim}{x}) = \exp\left\{-\frac{n}{2}\ln 2\pi - \frac{n}{2}\right\}$, οπότε

$$f(\underset{\sim}{x};\theta) = q(T(\underset{\sim}{x});\theta)h(\underset{\sim}{x}),\ \forall \underset{\sim}{x},\ \theta$$

οπότε, σύμφωνα με το παραγοντικό κριτήριο των Neyman-Fisher, η στατιστική συνάρτηση $T(\underset{\sim}{X}) = \left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$ είναι επαρκής. Μάλιστα, μπορεί να δειχθεί ότι η $T(\underset{\sim}{X})$ είναι ελάχιστη επαρκής. Παρατηρούμε λοιπόν ότι ενώ η άγνωστη παράμετρος θ είναι μονοδιάστατη (πραγματική παράμετρος) η ελάχιστη επαρκής στατιστική συνάρτηση είναι διδιάστατη. Τέτοιες περιπτώσεις δημιουργούν, γενικότερα, «δυσκολίες» στην εκτίμηση της παραμέτρου. Μια τέτοια «δυσκολία» είναι η μη πληρότητα της $T(\underset{\sim}{X})$.

Θα βρούμε πρώτα δύο αμερόληπτους εκτιμητές του $g(\theta) = \theta^2$ που είναι συναρτήσεις της στατιστικής συνάρτησης T, έστω $\phi_1(T)$ και $\phi_2(T)$. Στη συνέχεια, θα θεωρήσουμε τη στατιστική συνάρτηση $\phi(T) = \phi_1(T) - \phi_2(T)$, η οποία θα δείξουμε ότι παραβιάζει τον ορισμό της πληρότητας. Επειδή $\mathbb{V}ar_\theta X_i = \theta^2$, ένας αμερόληπτος εκτιμητής του θ^2 είναι η δειγματική διασπορά

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_1 - \bar{X})^2 = \frac{1}{n-1}\{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\}.$$

Επίσης, το θ μπορεί να εκτιμηθεί με $\bar{X}$ αφού είναι η μέση τιμή της κατανομής, οπότε το θ^2 μπορεί να εκτιμηθεί με $\bar{X}^2$. Τότε έχουμε

$$\mathbb{E}_\theta(\bar{X}^2) = \mathbb{V}ar_\theta\bar{X} + (\mathbb{E}_\theta\bar{X})^2 = \frac{\theta^2}{n} + \theta^2 = \frac{n+1}{n}\theta^2.$$

Επομένως $\mathbb{E}_\theta\left(\frac{n}{n+1}\bar{X}^2\right) = \theta^2$. Θέτουμε $\phi_1(T) = \frac{1}{n-1}\{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\}$ και $\phi_2(T) = \frac{n}{n+1}\bar{X}^2$, οπότε για τη στατιστική συνάρτηση

$$\phi(T) = \phi_1(T) - \phi_2(T) = -\frac{2}{(n-1)(n+1)}\left(\sum_{i=1}^{n} X_i\right)^2 + \frac{1}{n-1}\sum_{i=1}^{n} X_i^2$$

έχουμε $\mathbb{E}_\theta\phi(T) = \mathbb{E}_\theta\phi_1(T) - \mathbb{E}_\theta\phi_2(T) = \theta^2 - \theta^2 = 0$, για κάθε $\theta \in \Theta$. Επειδή $\phi(T) \neq 0$ και $\mathbb{E}_\theta\phi(T) = 0$ για κάθε $\theta \in \Theta$, συμπεραίνουμε, από τον ορισμό της πληρότητας, ότι η T δεν είναι πλήρης. Στην Άσκηση 6.15 δίνονται δύο άλλες στατιστικές συναρτήσεις με την ίδια μέση τιμή, μέσω των οποίων μπορεί επίσης να δειχθεί η μη πληρότητα της T.

Η απόδειξη της πληρότητας μιας στατιστικής συνάρτησης, γενικά, δεν είναι εύκολη και απαιτεί γνώσεις Ανάλυσης και Θεωρίας Μέτρου. Δίνουμε στη συνέχεια μία πρόταση που παρέχει πλήρη στατιστική συνάρτηση για *εκθετικές οικογένειες κατανομών τάξης* κ που ορίζονται ως εξής.

Ορισμός 6.3.3. *Η οικογένεια κατανομών του* $\underset{\sim}{X} = (X_1, \ldots, X_n)$, $\big\{f(\underset{\sim}{x};\theta)\colon \theta \in \Theta\big\}$, *ανήκει στην Εκθετική Οικογένεια Κατανομών τάξης* κ *(Ε.Ο.Κ.) , εάν*

(α) *Το σύνολο* $\mathcal{S} = \big\{\underset{\sim}{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n\colon f(\underset{\sim}{x};\theta) > 0\big\}$ *δεν εξαρτάται από το* θ.

(β) *Η πυκνότητα* $f(\underset{\sim}{x};\theta)$ *έχει τη μορφή*

$$f(\underset{\sim}{x};\theta) = \exp\Big\{A(\theta) + B(\underset{\sim}{x}) + \sum_{j=1}^{\kappa} C_j(\theta) D_j(\underset{\sim}{x})\Big\},\ \underset{\sim}{x} \in \mathcal{S},\ \theta \in \Theta.$$

Σημειώνουμε ότι για $\kappa = 1$ και $\theta \in \Theta \subset \mathbb{R}$, ο ορισμός συμπίπτει με αυτόν της Μ.Ε.Ο.Κ. (βλέπε Κεφάλαιο 5). Συνήθως (αλλά όχι πάντοτε) η τάξη κ συμπίπτει με τη διάσταση του θ, δηλαδή είναι $\theta = (\theta_1, \ldots, \theta_\kappa)$. Θεωρούμε ότι στην παραπάνω μορφή της πυκνότητας $f(\underset{\sim}{x};\theta)$, το κ είναι ο ελάχιστος θετικός ακέραιος για τον οποίον ισχύει μια τέτοια παράσταση. Επίσης, αδιακρίτως, θα χρησιμοποιούμε και την ορολογία η οικογένεια κατανομών του $\underset{\sim}{X}$ είναι (αντί ανήκει) μια Ε.Ο.Κ τάξης κ.

Πρόταση 6.3.5. *Έστω ότι η οικογένεια κατανομών του* $\underset{\sim}{X} = (X_1, \ldots, X_n)$, $\big\{f(\underset{\sim}{x};\theta)\colon \theta \in \Theta\big\}$, *ανήκει στην Ε.Ο.Κ. τάξης* κ. *Τότε ισχύουν τα εξής.*

(α) *Η στατιστική συνάρτηση* $T(\underset{\sim}{X}) = \big(D_1(\underset{\sim}{X}), \ldots, D_\kappa(\underset{\sim}{X})\big)$ *είναι επαρκής.*

(β) *Εάν το σύνολο* $\big\{\big(C_1(\theta), \ldots, C_\kappa(\theta)\big) \in \mathbb{R}^\kappa\colon \theta \in \Theta\big\}$ *περιέχει ένα ανοικτό υποσύνολο του* $\mathbb{R}^\kappa$ *τότε η στατιστική συνάρτηση* $T(\underset{\sim}{X})$ *είναι πλήρης.*

Απόδειξη.

(α) Άμεση συνέπεια του παραγοντικού κριτηρίου (Θεώρημα 6.1.1).

(β) Η απόδειξη, ουσιαστικά, βασίζεται στην μοναδικότητα του μετασχηματισμού Laplace της πυκνότητας της στατιστικής συνάρτησης $T(\underset{\sim}{X})$. Για τις τεχνικές λεπτομέρειες της απόδειξής της παραπέμπουμε τον αναγνώστη στους Lehmann and Romano (2005, παράγραφος 4.3). □

Στην περίπτωση τυχαίου δείγματος, η επόμενη πρόταση δίνει την δυνατότητα εύρεσης επαρκούς και πλήρους στατιστικής συνάρτησης μελετώντας μόνον την κοινή κατανομή των παρατηρήσεων.

Πρόταση 6.3.6. *Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την Ε.Ο.Κ. τάξης κ με πυκνότητα*

$$f_1(x;\theta) = \exp\Big\{A(\theta) + B(x) + \sum_{j=1}^{\kappa} C_j(\theta) D_j^*(x)\Big\},\ x \in \mathcal{S}_1,\ \theta \in \Theta.$$

Τότε έχουμε τα εξής.

(α) Η στατιστική συνάρτηση $T(\underset{\sim}{X}) = \big(D_1(\underset{\sim}{X}), \ldots, D\kappa(\underset{\sim}{X})\big)$, όπου $D_j(\underset{\sim}{X}) = \sum_{i=1}^{n} D_j^*(X_i)$, είναι επαρκής.

(β) Εάν το σύνολο $\big\{\big(C_1(\theta), \ldots, C_\kappa(\theta)\big) \in \mathbb{R}^\kappa \colon \theta \in \Theta\big\}$ περιέχει ένα ανοικτό υποσύνολο του $\mathbb{R}^\kappa$ τότε η στατιστική συνάρτηση $T(\underset{\sim}{X})$ είναι πλήρης.

Απόδειξη. Η πυκνότητα του $\underset{\sim}{X} = (X_1, \ldots, X_n)$ είναι

$$\begin{aligned}
f(\underset{\sim}{x};\theta) &= \prod_{i=1}^{n} f_1(x_i;\theta) \\
&= \exp\Big\{nA(\theta) + \sum_{i=1}^{n} B(x_i) + \sum_{j=1}^{\kappa} C_j(\theta) \sum_{i=1}^{n} D_j^*(x_i)\Big\} \\
&= \exp\Big\{nA(\theta) + \sum_{i=1}^{n} B(x_i) + \sum_{j=1}^{\kappa} C_j(\theta) D_j(\underset{\sim}{x})\Big\}, \\
&\underset{\sim}{x} = (x_1, \ldots, x_n) \in \mathcal{S} = \mathcal{S}_1 \times \cdots \times \mathcal{S}_1,\ \theta \in \Theta.
\end{aligned}$$

Επομένως η οικογένεια κατανομών του $\underset{\sim}{X} = (X_1, \ldots, X_n)$ είναι Ε.Ο.Κ. τάξης κ και η απόδειξη της πρότασης είναι άμεση συνέπεια της Πρότασης 6.3.5. □

Η Πρόταση 6.3.5 σε συνδυασμό με την Πρόταση 6.3.4 δίνει ΑΟΕΔ εκτιμητές σε Μ.Ε.Ο.Κ.

Πρόταση 6.3.7. *Έστω ότι η οικογένεια κατανομών του $\underset{\sim}{X} = (X_1, \ldots, X_n)$ είναι μια Μ.Ε.Ο.Κ. με πυκνότητα*

$$f(\underset{\sim}{x};\theta) = \exp\Big\{A(\theta) + B(\underset{\sim}{x}) + C(\theta)D(\underset{\sim}{x})\Big\},\ \underset{\sim}{x} \in \mathcal{S},\ \theta \in \Theta \subseteq \mathbb{R}.$$

και το σύνολο $\{C(\theta)\colon \theta \in \Theta\}$ περιέχει ένα ανοικτό υποσύνολο του $\mathbb{R}$. Τότε ισχύουν τα εξής.

(α) *Η στατιστική συνάρτηση $D(\underset{\sim}{X})$ είναι ο ΑΟΕΔ εκτιμητής του $g(\theta) = \mathbb{E}_\theta D(\underset{\sim}{X})$, εφ' όσον $\mathbb{V}ar_\theta D(\underset{\sim}{X}) < \infty$.*

(β) *Η στατιστική συνάρτηση $\alpha D(\underset{\sim}{X}) + \beta$ είναι ο ΑΟΕΔ εκτιμητής του $\alpha g(\theta) + \beta$, όπου $\alpha \neq 0$ και β σταθερές, εφ' όσον $\mathbb{V}ar_\theta D(\underset{\sim}{X}) < \infty$.*

(γ) *Η στατιστική συνάρτηση $\varphi\big(D(\underset{\sim}{X})\big)$ είναι ο ΑΟΕΔ εκτιμητής του $h(\theta) = \mathbb{E}_\theta \varphi\big(D(\underset{\sim}{X})\big)$, εφ' όσον $\mathbb{V}ar_\theta \varphi\Big(D(\underset{\sim}{X})\Big) < \infty$.*

Απόδειξη. (α) $D(\underset{\sim}{X})$ είναι επαρκής και πλήρης, λόγω της Πρότασης 6.3.5. Επιπλέον είναι αμερόληπτος εκτιμητής του $g(\theta)$, εξ ορισμού του $g(\theta)$. Το συμπέρασμα τώρα προκύπτει από την Πρόταση 6.3.4.

(β) και (γ) Παρόμοια απόδειξη όπως το (α).

□

Η Πρόταση 6.3.7 μας υπενθυμίζει την Πρόταση 5.2.4 που δίνει αποδοτικούς εκτιμητές σε Μ.Ε.Ο.Κ.. Γνωρίζουμε επίσης ότι ένας αποδοτικός εκτιμητής είναι και ΑΟΕΔ εκτιμητής. Συνεπώς, το συμπέρασμα της Πρότασης 5.2.4 είναι ισχυρότερο από αυτό της Πρότασης 6.3.7. Όμως θα πρέπει να αναγνωρίσουμε ότι η συνθήκη της Πρότασης 5.2.4 είναι ισχυρότερη από αυτήν της Πρότασης 6.3.7, ας δούμε γιατί. Η Πρόταση 5.2.4

απαιτεί να υπάρχει η παράγωγος $C'(\theta)$, $\theta \in \Theta$, στην οποία περίπτωση η $C(\theta)$ είναι συνεχής. Θεωρώντας ένα πεπερασμένο διάστημα $\Delta \subset \Theta$, το σύνολο $\{C(\theta) \colon \theta \in \Delta\}$ είναι επίσης διάστημα του $\mathbb{R}$, εφόσον η $C(\theta)$ είναι συνεχής και άρα τετριμμένα, περιέχει ένα ανοικτό υποσύνολο του $\mathbb{R}$. Αυτό το ανοικτό υποσύνολο, προφανώς περιέχεται και στο σύνολο $\{C(\theta) \colon \theta \in \Theta\} \supset \{C(\theta) \colon \theta \in \Delta\}$. Επομένως, η συνθήκη της Πρότασης 5.2.4 συνεπάγεται αυτήν της Πρότασης 6.3.7.

Η αξία λοιπόν της Πρότασης 6.3.7 (σε σχέση με την Πρόταση 5.2.4) είναι ότι όσον αφορά την εύρεση ΑΟΕΔ εκτιμητή, δεν είναι απαραίτητο η $C(\theta)$ να είναι παραγωγίσιμη. Επιπλέον, η Πρόταση 6.3.7 (γ) δίνει ΑΟΕΔ εκτιμητές που δεν είναι κατ' ανάγκη γραμμικοί μετασχηματισμοί του $D(\underset{\sim}{X})$. Σημειώνουμε επίσης ότι όλες οι σημαντικές Μ.Ε.Ο.Κ. (βλέπε τα Παραδείγματα 5.2.1–5.2.4) ικανοποιούν την συνθήκη της Πρότασης 6.3.7. Μία Ε.Ο.Κ. τάξης $\kappa = 2$ που δεν ικανοποιεί αυτήν τη συνθήκη δίνεται στο Παράδειγμα 6.3.9, όπου $C_1(\theta) = \frac{1}{\theta}$ και $C_2(\theta) = -\frac{1}{2\theta^2}$, $\theta \in \Theta = (0, \infty)$. Σε αυτήν την περίπτωση, το σύνολο

$$\begin{aligned}\big\{\big(C_1(\theta), C_2(\theta)\big) \in \mathbb{R}^2 \colon \theta \in \Theta\big\} &= \left\{\Big(\frac{1}{\theta}, -\frac{1}{2\theta^2}\Big) \in \mathbb{R}^2 \colon \theta > 0\right\} \\ &= \left\{\Big(t, -\frac{1}{2}t^2\Big) \in \mathbb{R}^2 \colon t > 0\right\}\end{aligned}$$

είναι μια καμπύλη (τμήμα παραβολής) στο επίπεδο $\mathbb{R}^2$, που προφανώς δεν περιέχει ανοικτό υποσύνολο του $\mathbb{R}^2$.

Στην περίπτωση ενός τυχαίου δείγματος από μία Μ.Ε.Ο.Κ., η Πρόταση 6.3.7 παίρνει την εξής μορφή.

Πρόταση 6.3.8. *Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από μια Μ.Ε.Ο.Κ. με πυκνότητα*

$$f_1(x;\theta) = \exp\Big\{A(\theta) + B(x) + C(\theta)D^*(x)\Big\},\ x \in \mathcal{S}_1,\ \theta \in \Theta \subset \mathbb{R}.$$

Εάν το σύνολο $\{C(\theta) \colon \theta \in \Theta\}$ περιέχει ένα ανοικτό υποσύνολο του $\mathbb{R}$, τότε ισχύουν τα συμπεράσματα της Πρότασης 6.3.7 με $D(\underset{\sim}{X}) = \sum_{i=1}^{n} D^*(X_i)$.

Απόδειξη. Όπως στην Πρόταση 6.3.6, η οικογένεια των κατανομών του $\underset{\sim}{X}$ είναι μια Μ.Ε.Ο.Κ με πυκνότητα

$$\begin{aligned} f(\underset{\sim}{x};\theta) &= \prod_{i=1}^{n} f_1(x_i;\theta) \\ &= \exp\Big\{nA(\theta) + \sum_{i=1}^{n} B(x_i) + C(\theta)\sum_{i=1}^{n} D^*(x_i)\Big\}, \end{aligned}$$

$\underset{\sim}{x} = (x_1,\ldots,x_n) \in \mathcal{S} = \mathcal{S}_1 \times \cdots \times \mathcal{S}_1$, $\theta \in \Theta$. Επομένως το συμπέρασμα της Πρότασης 6.3.8 ισχύει λόγω της Πρότασης 6.3.7. □

Παράδειγμα 6.3.10. (κανονική κατανομή - πληρότητα - ΑΟΕΔ εκτιμητές των μ και σ^2) Έστω $\underset{\sim}{X} = (X_1,\ldots,X_n)$ ένα τυχαίο δείγμα από την κανονική κατανομή $\mathcal{N}(\mu,\sigma^2)$. Θα βρούμε επαρκή και πλήρη στατιστική συνάρτηση.

1η Περίπτωση: $\boldsymbol{\sigma^2}$ **γνωστό,** $\boldsymbol{\mu = \theta}$ **άγνωστο**, $\theta \in \Theta = \mathbb{R}$.

Τότε έχουμε

$$\begin{aligned} f_1(x;\theta) &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\theta)^2} \\ &= \exp\left\{-\frac{\theta^2}{2\sigma^2} - \frac{1}{2}\ln(2\pi\sigma^2) - \frac{x^2}{2\sigma^2} + \frac{\theta}{\sigma^2}x\right\}, \quad x \in \mathbb{R}. \end{aligned}$$

Συνεπώς η οικογένεια $\{f_1(x;\theta)\colon \theta \in \Theta\}$ είναι μία Ε.Ο.Κ. τάξης $\kappa = 1$, δηλαδή Μ.Ε.Ο.Κ. με $A(\theta) = -\frac{\theta^2}{2\sigma^2}$, $B(x) = -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{x^2}{2\sigma^2}$, $C_1(\theta) = \frac{\theta}{\sigma^2}$, $D_1(x) = x$. Επειδή το σύνολο $\{C_1(\theta)\colon \theta \in \Theta\} = \left\{\frac{\theta}{\sigma^2}\colon \theta \in \mathbb{R}\right\} = \mathbb{R}$, προφανώς περιέχει ένα ανοικτό υποσύνολο, από την Πρόταση 6.3.6, η στατιστική συνάρτηση $T(\underset{\sim}{X}) = \sum_{i=1}^{n} D_1(X_i) = \sum_{i=1}^{n} X_i$ είναι επαρκής και πλήρης.

2η Περίπτωση: $\boldsymbol{\mu}$ **γνωστό,** $\boldsymbol{\sigma^2 = \theta}$ **άγνωστο**, $\theta \in \Theta = (0,\infty)$.

Τότε έχουμε

$$\begin{aligned} f_1(x;\theta) &= \frac{1}{\theta^{1/2}\sqrt{2\pi}} e^{-\frac{1}{2\theta}(x-\mu)^2} \\ &= \exp\left\{-\frac{1}{2}\ln(2\pi\theta) - \frac{1}{2\theta}(x-\mu)^2\right\}, \quad x \in \mathbb{R}. \end{aligned}$$

Συνεπώς η οικογένεια $\{f_1(x;\theta)\colon \theta \in \Theta\}$ είναι μία Ε.Ο.Κ. τάξης $\kappa = 1$, δηλαδή Μ.Ε.Ο.Κ. με $A(\theta) = -\frac{1}{2}\ln(2\pi\theta)$, $B(x) = 0$, $C_1(\theta) = -\frac{1}{2\theta}$, $D_1(x) = (x-\mu)^2$. Επειδή το σύνολο $\{C_1(\theta)\colon \theta \in \Theta\} = \big\{-\frac{1}{2\theta}\colon \theta > 0\big\} = (-\infty, 0)$, προφανώς περιέχει ένα ανοικτό υποσύνολο, από την Πρόταση 6.3.6, η στατιστική συνάρτηση $T(\underset{\sim}{X}) = \sum\limits_{i=1}^{n} D_1(X_i) = \sum\limits_{i=1}^{n}(X_i - \mu)^2$ είναι επαρκής και πλήρης.

3η Περίπτωση: $\boldsymbol{\mu, \sigma^2}$ **άγνωστα**, οπότε $\theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times (0, \infty)$. Τότε

$$
\begin{aligned}
f_1(x;\theta) &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \\
&= \exp\left\{-\frac{\mu^2}{2\sigma^2} - \frac{1}{2}\ln(2\pi\sigma^2) - \frac{x^2}{2\sigma^2} + \frac{\mu}{\sigma^2}x\right\}, \quad x \in \mathbb{R}.
\end{aligned}
$$

Συνεπώς η οικογένεια $\{f_1(x;\theta)\colon \theta \in \Theta\}$ είναι Ε.Ο.Κ. τάξης $\kappa = 2$ με $A(\theta) = -\frac{\mu^2}{2\sigma^2} - \frac{1}{2}\ln(2\pi\sigma^2)$, $B(x) = 0$, $C_1(\theta) = \frac{\mu}{\sigma^2}$, $C_2(\theta) = -\frac{1}{2\sigma^2}$, $D_1(x) = x$, $D_2(x) = x^2$. Επειδή το σύνολο $\big\{\big(C_1(\theta), C_2(\theta)\big) \in \mathbb{R}^2\colon \theta \in \Theta\big\}$ $= \big\{\big(\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2}\big)\colon \mu \in \mathbb{R},\, \sigma > 0\big\} = \mathbb{R} \times (-\infty, 0)$, προφανώς περιέχει ένα ανοικτό υποσύνολο (του $\mathbb{R}^2$) από την Πρόταση 6.3.6 η στατιστική συνάρτηση $T(\underset{\sim}{X}) = \big(\sum\limits_{i=1}^{n} D_1(X_i), \sum\limits_{i=1}^{n} D_2(X_i)\big) = \big(\sum\limits_{i=1}^{n} X_i, \sum\limits_{i=1}^{n} X_i^2\big)$ είναι επαρκής και πλήρης.

Θα βρούμε τώρα τους ΑΟΕΔ εκτιμητές των μ, σ^2, μ^2 χρησιμοποιώντας την τεχνική των Lehmann–Scheffé.

Επειδή $\mathbb{E}_\theta \bar{X} = \mu$, για κάθε $\theta = (\mu, \sigma^2)$ έχουμε ότι $\bar{X} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$ είναι αμερόληπτος εκτιμητής του μ και συνάρτηση της επαρκούς και πλήρους στατιστικής συνάρτησης T. Επομένως από την Πρόταση 6.3.4, $\bar{X}$ είναι ο ΑΟΕΔ εκτιμητής του μ. Για τον ΑΟΕΔ εκτιμητή του σ^2 παρατηρούμε ότι η δειγματική διασπορά

$$
S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\Big\{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\Big\}
$$

είναι αμερόληπτος εκτιμητής του σ^2 (βλέπε Πρόταση 4.2.4) και συνάρτηση της επαρκούς και πλήρους στατιστικής συνάρτησης T. Επομένως από την

Πρόταση 6.3.4, S^2 είναι ο ΑΟΕΔ του σ^2. Τέλος, παρατηρούμε ότι

$$\mathbb{E}_\theta(\bar{X}^2) = \mathbb{V}ar_\theta\bar{X} + (\mathbb{E}_\theta\bar{X})^2 = \frac{\sigma^2}{n} + \mu^2 = \frac{\mathbb{E}_\theta S^2}{n} + \mu^2 = \mathbb{E}_\theta(\frac{S^2}{n}) + \mu^2.$$

Επομένως, $\mathbb{E}_\theta(\bar{X}^2 - \frac{S^2}{n}) = \mu^2$, για κάθε $\theta = (\mu, \sigma^2)$. Άρα $\bar{X}^2 - \frac{S^2}{n}$ είναι αμερόληπτος εκτιμητής του μ^2 και συνάρτηση της επαρκούς και πλήρους στατιστικής συνάρτησης T. Συνεπώς, από την Πρόταση 6.3.4, $\bar{X}^2 - \frac{S^2}{n}$ είναι ο ΑΟΕΔ εκτιμητής του μ^2. Υπενθυμίζουμε ότι $\bar{X}$ είναι, επιπλέον, αποδοτικός εκτιμητής του μ, ενώ S^2 είναι ΑΟΕΔ εκτιμητής αλλά μη αποδοτικός (βλέπε Παράδειγμα 5.3.1). Επίσης ο ΑΟΕΔ του μ^2 δεν είναι αποδοτικός.

Παρατήρηση 6.3.2. Η ύπαρξη επαρκούς και πλήρους στατιστικής συνάρτησης εξασφαλίζει, μέσω του Θεωρήματος 6.3.3 (Lehmann–Scheffé), την ύπαρξη ΑΟΕΔ εκτιμητή εφόσον υπάρχει κάποιος αμερόληπτος εκτιμητής. Ο πλέον λοιπόν τυπικός και συνηθισμένος τρόπος αναζήτησης ΑΟΕΔ εκτιμητή είναι μέσω του Θεωρήματος 6.3.3 (Lehmann–Scheffé). Φυσικά υπάρχει και ο τρόπος εύρεσης ΑΟΕΔ εκτιμητή μέσω της ανισότητας Cramer–Rao (αφού ο αποδοτικός εκτιμητής είναι και ΑΟΕΔ εκτιμητής), αλλά αυτός ο τρόπος καλύπτει ειδικές μόνον περιπτώσεις (οι οποίες όμως καλύπτονται και από το Θεώρημα 6.3.3 (Lehmann–Scheffé)). Σημειώνουμε εδώ ότι η ύπαρξη πλήρους στατιστικής συνάρτησης δεν είναι αναγκαία για την ύπαρξη ΑΟΕΔ εκτιμητή, και είναι δυνατόν να υπάρχει ΑΟΕΔ εκτιμητής και όταν ακόμη δεν υπάρχει πλήρης στατιστική συνάρτηση, βλέπε Rao (1973, σελ. 313). Έτσι υπάρχουν περιθώρια μελέτης ΑΟΕΔ εκτιμητών και πέραν του Θεωρήματος 6.3.3 (Lehmann–Scheffé). Συγκεκριμένα, θα δείξουμε ότι αν S είναι αμερόληπτος εκτιμητής του $g(\theta)$ με $\mathbb{V}ar_\theta S < \infty$ και ισχύει η σχέση

$$\mathbb{C}ov_\theta(S, U) = 0, \ \forall\ \theta \in \Theta \tag{6.14}$$

για κάθε στατιστική συνάρτηση U, με $\mathbb{V}ar_\theta U < \infty$, που ικανοποιεί τη συνθήκη $\mathbb{E}_\theta U = 0$ για κάθε $\theta \in \Theta$, τότε S είναι ΑΟΕΔ εκτιμητής του $g(\theta)$ και αντίστροφα. Πράγματι, έστω S_1 (αυθαίρετος) αμερόληπτος εκτιμητής του $g(\theta)$ με $\mathbb{V}ar_\theta S_1 < \infty$. Θέτοντας $U = S - S_1$, προκύπτει $\mathbb{E}_\theta U =$

$\mathbb{E}_\theta S - \mathbb{E}_\theta S_1 = g(\theta) - g(\theta) = 0$, οπότε από τη σχέση (6.14) παίρνουμε $\mathbb{C}ov_\theta(S, S - S_1) = 0$. Από τις ιδιότητες της συνδιασποράς, Πρόταση 1.8.1, η τελευταία σχέση συνεπάγεται ότι $\mathbb{V}ar_\theta S - \mathbb{C}ov_\theta(S, S_1) = 0$, δηλαδή $\mathbb{V}ar_\theta S = \mathbb{C}ov_\theta(S, S_1)$. Τότε όμως έχουμε

$$0 \leq \mathbb{V}ar_\theta(S - S_1) = \mathbb{V}ar_\theta S + \mathbb{V}ar_\theta S_1 - 2\mathbb{C}ov_\theta(S, S_1) = \mathbb{V}ar_\theta S_1 - \mathbb{V}ar_\theta S$$

και επομένως $\mathbb{V}ar_\theta S \leq \mathbb{V}ar_\theta S_1$. Συνεπώς S είναι ΑΟΕΔ εκτιμητής του $g(\theta)$. Αντίστροφα, έστω S ΑΟΕΔ εκτιμητής του $g(\theta)$ και U στατιστική συνάρτηση με $\mathbb{E}_\theta U = 0$. Τότε για κάθε $\lambda \in \mathbb{R}$, $S_1 = S + \lambda U$ είναι αμερόληπτος εκτιμητής του $g(\theta)$ και επομένως

$$\mathbb{V}ar_\theta S \leq \mathbb{V}ar_\theta S_1 = \mathbb{V}ar_\theta(S + \lambda U) = \mathbb{V}ar_\theta S + \lambda^2 \mathbb{V}ar_\theta U + 2\lambda \mathbb{C}ov_\theta(S, U),$$

που συνεπάγεται ότι $\Delta = \lambda^2 \mathbb{V}ar_\theta U + 2\lambda \mathbb{C}ov_\theta(S, U) \geq 0$. Αν τώρα $\mathbb{V}ar_\theta U = 0$, τότε σε συνδυασμό με τη σχέση $\mathbb{E}_\theta U = 0$ έχουμε $U = 0$ με πιθανότητα 1 και η (6.14) ισχύει τετριμμένα. Αν $\mathbb{V}ar_\theta U > 0$, τότε για $\lambda = -\frac{\mathbb{C}ov_\theta(S,U)}{\mathbb{V}ar_\theta U}$ προκύπτει $\Delta = -\frac{\mathbb{C}ov^2_\theta(S,U)}{\mathbb{V}ar_\theta U}$ που αντίκειται στη σχέση $\Delta \geq 0$ εκτός αν $\mathbb{C}ov_\theta(S, U) = 0$.

Με διαφορετική διατύπωση, η (6.14) σημαίνει ότι ο S είναι ΑΟΕΔ εκτιμητής του $g(\theta) = \mathbb{E}_\theta S$ εάν και μόνον εάν είναι ασυσχέτιστος προς κάθε αμερόληπτο εκτιμητή του μηδενός. Η σχέση λοιπόν (6.14) δίνει μία ικανή και αναγκαία συνθήκη για να είναι ένας εκτιμητής ΑΟΕΔ εκτιμητής. Ως εφαρμογή αυτής της συνθήκης αποδεικνύουμε την μοναδικότητα του ΑΟΕΔ εκτιμητή, χωρίς την επίκληση επάρκειας και πληρότητας.

Πρόταση 6.3.9. *Εάν T_1, T_2 είναι ΑΟΕΔ εκτιμητές του $g(\theta)$ με $\mathbb{V}ar_\theta T_1 = \mathbb{V}ar_\theta T_2 < \infty$, τότε με πιθανότητα 1, $T_1 = T_2$.*

Απόδειξη. Θέτουμε $U = T_1 - T_2$ και παρατηρούμε ότι

$$\mathbb{E}_\theta U = \mathbb{E}_\theta(T_1 - T_2) = \mathbb{E}_\theta T_1 - \mathbb{E}_\theta T_2 = g(\theta) - g(\theta) = 0, \quad \forall\, \theta \in \Theta.$$

Λόγω της (6.14) επειδή ο T_1 είναι ΑΟΕΔ εκτιμητής του $g(\theta)$, έχουμε $\mathbb{C}ov_\theta(T_1, U) = 0$, δηλαδή

$$\mathbb{C}ov_\theta(T_1, T_1 - T_2) = 0. \tag{6.15}$$

Όμως από την Πρόταση 1.8.1 (5,8),

$$\mathbb{C}ov_\theta(T_1, T_1 - T_2) = \mathbb{C}ov_\theta(T_1, T_1) - \mathbb{C}ov_\theta(T_1, T_2) = \mathbb{V}ar_\theta T_1 - \mathbb{C}ov_\theta(T_1, T_2)$$

οπότε η (6.15) δίνει $\mathbb{V}ar_\theta T_1 = \mathbb{C}ov_\theta(T_1, T_2)$. Επομένως,

$$\begin{aligned}\mathbb{V}ar_\theta(T_1 - T_2) &= \mathbb{V}ar_\theta T_1 + \mathbb{V}ar_\theta T_2 - 2\mathbb{C}ov_\theta(T_1, T_2)\\ &= 2\mathbb{V}ar_\theta T_1 - 2\mathbb{V}ar_\theta T_1 = 0\end{aligned}$$

το οποίο συνεπάγεται ότι με πιθανότητα 1, $T_1 - T_2$ είναι σταθερά. Επειδή όμως $\mathbb{E}_\theta(T_1 - T_2) = 0$, αυτή η σταθερά είναι κατ' ανάγκη το μηδέν, δηλαδή $T_1 - T_2 = 0$. □

Γνωρίζουμε ότι η αμεροληψία μεταβιβάζεται σε γραμμικούς μετασχηματισμούς. Ανάλογη ιδιότητα ισχύει και για γραμμικούς μετασχηματισμούς ΑΟΕΔ εκτιμητών.

Πρόταση 6.3.10. *Εάν $S(\underset{\sim}{X})$ είναι ο ΑΟΕΔ εκτιμητής του $g(\theta)$, τότε $\alpha S(\underset{\sim}{X}) + \beta$ είναι ο ΑΟΕΔ εκτιμητής του $\alpha g(\theta) + \beta$, όπου $\alpha \neq 0$ και β σταθερές (μη εξαρτώμενες από το θ).*

Απόδειξη. Θα χρησιμοποιήσουμε την ικανή και αναγκαία συνθήκη (6.14). Έστω U αμερόληπτος εκτιμητής του μηδενός, $\mathbb{E}_\theta U = 0$. Τότε, από την Πρόταση 1.8.1 έχουμε

$$\mathbb{C}ov_\theta(\alpha S + \beta, U) = \alpha \mathbb{C}ov_\theta(S, U) = 0, \quad \forall\, \theta \in \Theta.$$

□

Στο επόμενο θεώρημα δίνουμε μία ακόμη χρήση των εννοιών της επάρκειας και πληρότητας. Το θεώρημα δείχνει ότι σε ορισμένες περιπτώσεις η ανεξαρτησία μεταξύ τυχαίων μεταβλητών μπορεί να αποδειχθεί μέσω επάρκειας και πληρότητας.

Θεώρημα 6.3.11. *(**Basu**) Έστω ότι η στατιστική συνάρτηση $T(\underset{\sim}{X})$ είναι επαρκής και πλήρης και $S(\underset{\sim}{X})$ είναι μία στατιστική συνάρτηση με κατανομή που δεν εξαρτάται από την άγνωστη παράμετρο $\theta \in \Theta$. Τότε $S(\underset{\sim}{X})$ και $T(\underset{\sim}{X})$ είναι ανεξάρτητες στατιστικές συναρτήσεις.*

Απόδειξη. Έστω A ένα υποσύνολο του συνόλου τιμών της στατιστικής συνάρτησης $S(\underset{\sim}{X})$. Αρκεί να δείξουμε ότι $\mathbb{P}(S \in A \mid T = t) = \mathbb{P}(S \in A)$ για κάθε τιμή t της στατιστικής συνάρτησης $T(\underset{\sim}{X})$. Παρατηρούμε τώρα ότι εάν $\mathrm{I}_A(S)$ είναι η δείκτρια συνάρτηση του συνόλου A, δηλαδή

$$\mathrm{I}_A(S) = \begin{cases} 1, & \text{εάν } S(\underset{\sim}{X}) \in A \\ 0, & \text{εάν } S(\underset{\sim}{X}) \notin A \end{cases}$$

τότε

$$\begin{aligned} \mathbb{E}\big(\mathrm{I}_A(S) \mid T\big) &= 1 \cdot \mathbb{P}\big(\mathrm{I}_A(S) = 1 \mid T\big) + 0 \cdot \mathbb{P}\big(\mathrm{I}_A(S) = 0 \mid T\big) \\ &= \mathbb{P}\big(\mathrm{I}_A(S) = 1 \mid T\big) = \mathbb{P}(S \in A \mid T). \end{aligned}$$

Επομένως,

$$\begin{aligned} \mathbb{E}_\theta\big(\mathbb{P}(S \in A \mid T) - \mathbb{P}(S \in A)\big) &= \mathbb{E}_\theta\big[\mathbb{E}\big(\mathrm{I}_A(S) \mid T\big) - \mathbb{P}(S \in A)\big] \\ &= \mathbb{E}_\theta\big[\mathbb{E}\big(\mathrm{I}_A(S) \mid T\big)\big] - \mathbb{P}(S \in A) \\ &= \mathbb{E}\,\mathrm{I}_A(S) - \mathbb{P}(S \in A) \\ &= 1 \cdot \mathbb{P}\big(\mathrm{I}_A(S) = 1\big) + 0 \cdot \mathbb{P}\big(\mathrm{I}_A(S) = 0\big) - \mathbb{P}(S \in A) \\ &= \mathbb{P}\big(\mathrm{I}_A(S) = 1\big) - \mathbb{P}(S \in A) \\ &= \mathbb{P}(S \in A) - \mathbb{P}(S \in A) = 0,\ \theta \in \Theta. \end{aligned}$$

Άρα με $\phi(T) = \mathbb{P}(S \in A \mid T) - \mathbb{P}(S \in A)$ έχουμε ότι

$$\mathbb{E}_\theta \phi(T) = 0,\ \forall\, \theta \in \Theta. \tag{6.16}$$

Τώρα, η πιθανότητα $\mathbb{P}(S \in A)$ δεν εξαρτάται από το θ αφού εξ υποθέσεως η κατανομή του S δεν εξαρτάται από το θ. Ακόμη, λόγω επάρκειας, η δεσμευμένη πιθανότητα $\mathbb{P}(S \in A \mid T)$ δεν εξαρτάται από το θ. Συνεπώς η $\phi(T)$ είναι στατιστική συνάρτηση και συνάρτηση της πλήρους στατιστικής συνάρτησης T. Επομένως, λόγω της (6.16) και του ορισμού πληρότητας, έχουμε $\phi(T) = 0$, δηλαδή $\phi(t) = 0$ για κάθε τιμή t της T. Άρα $\mathbb{P}(S \in A \mid T = t) - \mathbb{P}(S \in A) = 0$ για κάθε τιμή t της T. Συνεπώς, $\mathbb{P}(S \in A \mid T = t) = \mathbb{P}(S \in A)$. □

Ως εφαρμογή του Θεωρήματος 6.3.11 (Basu) αποδεικνύουμε την ακόλουθη πρόταση.

Πρόταση 6.3.12. *Έστω* $\underset{\sim}{X} = (X_1, \ldots, X_n)$ *ένα τυχαίο δείγμα από την κανονική κατανομή* $\mathcal{N}(\mu, \sigma^2)$. *Τότε* $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ *και* $(X_1 - \bar{X}, \ldots, X_n - \bar{X})$ *είναι ανεξάρτητες στατιστικές συναρτήσεις.*

Απόδειξη. Θα πρέπει να αποδείξουμε την ανεξαρτησία για οποιαδήποτε τιμή του ζεύγους (μ, σ^2). Θεωρούμε κατ' αρχάς μία δεδομένη τιμή του σ^2, έστω $\sigma^2 = \sigma_0^2$, και λαμβάνουμε το μ ως άγνωστη παράμετρο θ, δηλαδή $\theta = \mu \in \Theta = \mathbb{R}$. Τότε γνωρίζουμε (βλέπε Παράδειγμα 6.3.10) ότι η στατιστική συνάρτηση $T = \sum_{i=1}^{n} X_i$ είναι επαρκής και πλήρης. Ακόμη, μπορούμε να θέσουμε $X_i = Z_i + \theta$, $i = 1, \ldots, n$, οπότε $\underset{\sim}{Z} = (Z_1, \ldots, Z_n)$ είναι ένα τυχαίο δείγμα από την κατανομή $\mathcal{N}(0, \sigma_0^2)$ και $\bar{X} = \bar{Z} + \theta$, όπου $\bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i$. Παρατηρούμε τότε ότι η κατανομή των $Z_1, \ldots, Z_n$ δεν εξαρτάται από το θ και συνεπώς και η κατανομή της στατιστικής συνάρτησης $S = (Z_1 - \bar{Z}, \ldots, Z_n - \bar{Z})$ δεν εξαρτάται από το θ. Αλλά $S = (Z_1 - \bar{Z}, \ldots, Z_n - \bar{Z}) = (X_1 - \bar{X}, \ldots, X_n - \bar{X})$. Άρα $T = \sum_{i=1}^{n} X_i$ είναι επαρκής και πλήρης και $S = (X_1 - \bar{X}, \ldots, X_n - \bar{X})$ έχει κατανομή που δεν εξαρτάται από το θ. Επομένως $\sum_{i=1}^{n} X_i$ (και άρα $\bar{X}$) και $(X_1 - \bar{X}, \ldots, X_n - \bar{X})$ είναι ανεξάρτητες στατιστικές συναρτήσεις για κάθε τιμή του $\theta = \mu$, από το Θεώρημα 6.3.11 (Basu), δηλαδή για κάθε τιμή (μ, σ_0^2). Επειδή τώρα η τιμή $\sigma^2 = \sigma_0^2$ είναι τυχούσα, συμπεραίνουμε ότι η ανεξαρτησία ισχύει για κάθε τιμή του ζεύγους (μ, σ^2). □

Ως μία ακόμη ενδιαφέρουσα εφαρμογή του Θεωρήματος του Basu και του Θεωρήματος των Lehmann και Scheffé δίνουμε το εξής παράδειγμα.

Παράδειγμα 6.3.11. (εκθετική κατανομή - ΑΟΕΔ εκτιμητές της συνάρτησης αξιοπιστίας) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την εκθετική κατανομή $\mathcal{E}(\theta)$ με πυκνότητα $f_1(x; \theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}}$, $x > 0$, $\theta \in \Theta = (0, \infty)$. Θα βρούμε τον ΑΟΕΔ εκτιμητή του

$$g(\theta) = \mathbb{P}_\theta(X_1 > x_0) = \int_{x_0}^{\infty} \frac{1}{\theta} e^{-\frac{x}{\theta}}\, dx = e^{-\frac{x_0}{\theta}},$$

όπου $x_0 > 0$ δοθείσα σταθερά. Έχουμε αναφέρει, στο Κεφάλαιο 2 (Παράδειγμα 2.1) ότι η εκθετική κατανομή μπορεί να χρησιμοποιηθεί ως μοντέλο του χρόνου ζωής ενός «συστήματος». Σε αυτήν την περίπτωση, το $g(\theta)$ είναι η πιθανότητα ένα τέτοιο σύστημα να έχει χρόνο ζωής τουλάχιστον x_0 ή με άλλα λόγια εκφράζει το ποσοστό τέτοιων «συστημάτων» με χρόνο ζωής μεγαλύτερο του x_0. Στη Θεωρία Αξιοπιστίας (Reliability Theory) η πιθανότητα $\mathbb{P}_\theta(X_1 > x_0)$ αναφέρεται ως αξιοπιστία του «συστήματος», την χρονική στιγμή x_0. Αν το x_0 θεωρηθεί ως η χρονική εγγύηση που δίνει ο κατασκευαστής του «συστήματος» υποχρεούμενος να το αντικαταστήσει στην περίπτωση που ο χρόνος ζωής του δεν υπερβεί την εγγύηση x_0, τότε $1 - \mathbb{P}_\theta(X_1 > x_0) = \mathbb{P}_\theta(X_1 \le x_0)$ ερμηνεύεται ως το ποσοστό των «συστημάτων» που ο κατασκευαστής υποχρεούται να αντικαταστήσει. Οι παρατηρήσεις X_i, $i = 1, \ldots, n$, είναι οι χρόνοι ζωής ενός δείγματος n «συστημάτων». Από το Παράδειγμα 6.3.4, γνωρίζουμε ήδη ότι η στατιστική συνάρτηση $T(\underset{\sim}{X}) = \sum_{i=1}^{n} X_i$ (δηλαδή ο συνολικός χρόνος ζωής του δείγματος) είναι επαρκής και πλήρης. Σε αυτό το σημείο, προσθέτουμε ότι η οικογένεια κατανομών του $\underset{\sim}{X}$ είναι μία Μ.Ε.Ο.Κ, οπότε η επάρκεια και πληρότητα της $T(\underset{\sim}{X})$ μπορεί να διαπιστωθεί και μέσω της Πρότασης 6.3.7. Επιπλέον, επειδή το $g(\theta)$ εκφράζει πιθανότητα ενδεχομένου, ένας αμερόληπτος εκτιμητής του είναι η δείκτρια αυτού του ενδεχομένου. Συγκεκριμένα, ορίζουμε

$$S(\underset{\sim}{X}) = \mathrm{I}_{(x_0,\infty)}(X_1) = \begin{cases} 1, & \text{αν } X_1 > x_0, \\ 0, & \text{αν } X_1 \le x_0 \end{cases}$$

Τότε έχουμε

$$\begin{aligned} \mathbb{E}_\theta S(\underset{\sim}{X}) &= 1 \cdot \mathbb{P}_\theta(S(\underset{\sim}{X}) = 1) + 0 \cdot \mathbb{P}_\theta(S(\underset{\sim}{X}) = 0) \\ &= \mathbb{P}_\theta(X_1 > x_0) = g(\theta), \quad \forall\, \theta \in \Theta \end{aligned}$$

και συνεπώς η $S(\underset{\sim}{X})$ είναι αμερόληπτος εκτιμητής του $g(\theta)$. Το Θεώρημα Lehmann–Scheffé εξασφαλίζει ότι η Rao–Blackwell βελτίωση $S^*(T) = \mathbb{E}(S \mid T)$ είναι ο ΑΟΕΔ εκτιμητής του $g(\theta)$. Για να βρούμε τον τύπο του

$S^*(T)$ έστω $t > 0$ μια τυχούσα τιμή της $T(\underset{\sim}{X})$. Τότε

$$\begin{aligned} S^*(t) &= \mathbb{E}(S \mid T = t) = 1 \cdot \mathbb{P}_\theta(S = 1 \mid T = t) + 0 \cdot \mathbb{P}_\theta(S = 0 \mid T = t) \\ &= \mathbb{P}_\theta(X_1 > x_0 \mid T = t). \end{aligned}$$

Σε αυτό το σημείο, χρειαζόμαστε τη δεσμευμένη κατανομή της X_1 δοθέντος ότι $T = t$, την οποία θα βρούμε χρησιμοποιώντας (ευφυώς) το Θεώρημα Basu. Θέτουμε $U = \frac{X_1}{T}$, και θα δείξουμε ότι οι U, T είναι ανεξάρτητες. Επειδή η T είναι επαρκής και πλήρης, αρκεί να δείξουμε ότι η κατανομή της U δεν εξαρτάται από το θ. Πράγματι, αφού $X_1 \sim \mathcal{E}(\theta)$, έχουμε $\frac{X_1}{\theta} \sim \mathcal{E}(1)$, ενώ $T \sim \mathcal{G}(n, \theta)$ συνεπάγεται ότι $\frac{T}{\theta} \sim \mathcal{G}(n, 1)$ λόγω της Πρότασης 1.8.5 (10). Κατά συνέπεια $U = \frac{X_1/\theta}{T/\theta}$ έχει κατανομή που δεν εξαρτάται από το θ. Η X_1 λοιπόν γράφεται ως γινόμενο ανεξαρτήτων στατιστικών συναρτήσεων, $X_1 = U \cdot T$. Τώρα, δοθέντος ότι $T = t$, λόγω ανεξαρτησίας η κατανομή της U παραμένει αμετάβλητη, ενώ η X_1 γίνεται $X_1 = tU$. Εν κατακλείδι, η δεσμευμένη κατανομή της X_1 δοθέντος ότι $T = t$ είναι η κατανομή της tU, οπότε

$$S^*(t) = \mathbb{P}_\theta(X_1 > x_0 \mid T = t) = \mathbb{P}_\theta(tU > x_0) = \mathbb{P}_\theta(U > \frac{x_0}{t}).$$

Από την Πρόταση 1.8.8, η κατανομή της U είναι $\mathcal{B}eta(\alpha, \beta)$ με $\alpha = 1$ και $\beta = n - 1$, $n \geq 2$, δηλαδή έχει πυκνότητα $(n-1)(1-u)^{n-2}$, $0 < u < 1$. Τελικά, λοιπόν,

$$S^*(t) = \begin{cases} \displaystyle\int_{\frac{x_0}{t}}^{\infty} (n-1)(1-u)^{n-2}\, du = \Big(1 - \frac{x_0}{t}\Big)^{n-1}, & \text{αν } \ x_0 < t, \\ 0, & \text{αν } \ x_0 \geq t \end{cases}$$

το οποίο συνεπάγεται ότι

$$S^*(T) = \begin{cases} \Big(1 - \dfrac{x_0}{T}\Big)^{n-1}, & \text{αν } \ T > x_0, \\ 0, & \text{αν } \ T \leq x_0 \end{cases} \qquad (6.17)$$

είναι ο ΑΟΕΔ της πιθανότητας $g(\theta) = \mathbb{P}_\theta(X_1 > x_0)$, $n \geq 2$. Για $n = 1$, $\underset{\sim}{X} = X_1$ και ο ΑΟΕΔ εκτιμητής είναι $S(X_1)$ από την Πρόταση 6.3.4.

Όμως, θέτοντας $n = 1$ στην (6.17) προκύπτει $S^*(T) = S(X_1)$. Συνεπώς η (6.17) δίνει τον ΑΟΕΔ του $g(\theta)$, για κάθε $n \geq 1$. Επιπλέον, ο ΑΟΕΔ εκτιμητής της συνάρτησης κατανομής στο σημείο x_0 $F(x_0;\theta) = \mathbb{P}_\theta(X_1 \leq x_0) = 1 - \mathbb{P}_\theta(X_1 > x_0)$ είναι ο $1 - S^*(T)$.

Παρατηρούμε ότι εάν ο συνολικός χρόνος ζωής των n «συστημάτων» του δείγματος T, είναι μικρότερος του x_0, η εκτίμηση που δίνει ο $S^*(T)$ για το $g(\theta)$ είναι 0, παρότι $0 < g(\theta) < 1$. Αυτό το γεγονός είναι ένα μειονέκτημα του $S^*(T)$, όμως για κάθε πρακτικά ορθολογική τιμή της εγγύησης x_0, αναμένουμε ο συνολικός χρόνος ζωής του δείγματος να είναι μεγαλύτερος του x_0, δηλαδή $T > x_0$, οπότε

$$S^*(T) = \Big(1 - \frac{x_0}{T}\Big)^{n-1} \quad \text{και} \quad 0 < S^*(T) < 1.$$

Ας προσθέσουμε ότι για $n \to \infty$ από τον (Ασθενή) Νόμο των Μεγάλων Αριθμών,

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{T}{n} \xrightarrow{\mathbb{P}_\theta} \mathbb{E}_\theta X_1 = \theta$$

(και άρα $T \xrightarrow{\mathbb{P}_\theta} \infty$), το οποίο συνεπάγεται ότι

$$S^*(T) = \Big(1 - \frac{x_0}{T}\Big)^{n-1} \approx \Big(1 - \frac{x_0}{n\theta}\Big)^{n-1} \to e^{-\frac{x_0}{\theta}} = g(\theta).$$

Επομένως για μεγάλο μέγεθος δείγματος $S^*(T) \approx g(\theta)$, γεγονός καθησυχαστικό για τη συμπεριφορά του $S^*(T)$.
Ως εφαρμογή ας θεωρήσουμε ότι είναι $x_0 = 1$ (έτος-συνήθης εγγύηση π.χ. σε φορητούς υπολογιστές), $n = 20$ και ότι κατά μέσο όρο στο δείγμα ο χρόνος ζωής είναι 5 (έτη), δηλαδή $\bar{x} = 5$ οπότε $\sum\limits_{i=1}^{20} x_i = n\bar{x} = 100$ (λόγω επάρκειας, δεν έχουν ιδιαίτερη σημασία οι παρατηρηθέντες χρόνοι ζωής x_i, $i = 1, \ldots, 20$). Τότε από την (6.17) προκύπτει ως εκτίμηση του ποσοστού των φορητών υπολογιστών με χρόνο ζωής μεγαλύτερο του ενός έτους η τιμή $s^* = \big(1 - \frac{1}{100}\big)^{19} \simeq 0.8261$ ή 83%.
Από την Ενότητα 3.3 (ε), για $B = (x_0, \infty)$, ο εκτιμητής $\frac{1}{n}\sum\limits_{i=1}^{n} \mathrm{I}_{(x_0,\infty)}(X_i)$ είναι αμερόληπτος για το $g(\theta) = \mathbb{P}_\theta(X_1 > x_0)$, όμως είναι μη αποδεκτός με κριτήριο το ΜΤΣ επειδή δεν είναι συνάρτηση της επαρκούς στατιστικής

συνάρτησης T (Πρόταση 6.2.2). Ανάλογο συμπέρασμα ισχύει και για την εμπειρική συνάρτηση κατανομής $\widehat{F}(x_0) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}_{(0,x_0]}(X_i)$, που σύμφωνα με την Πρόταση 4.2.5 είναι αμερόληπτος εκτιμητής της συνάρτησης κατανομής της X_i στο x_0, $F(x_0;\theta) = 1 - e^{-\frac{x_0}{\theta}}$. □

Κλείνουμε αυτήν την ενότητα με μερικά ακόμη παραδείγματα.

Παράδειγμα 6.3.12. (**ανεξάρτητες Γάμμα παρατηρήσεις**) Έστω $\underset{\sim}{X} = (X_1, X_2, \ldots, X_n)$, όπου οι X_i είναι ανεξάρτητες παρατηρήσεις, $X_i \sim \mathcal{G}(i,\theta)$, $i = 1,\ldots,n$, $\theta \in \Theta = (0,\infty)$. Θα βρούμε τους ΑΟΕΔ εκτιμητές των θ και θ^r (εφόσον υπάρχει ο δεύτερος).

Αρχικά θα βρούμε επαρκή και πλήρη στατιστική συνάρτηση. Η οικογένεια κατανομών του $\underset{\sim}{X}$ ανήκει στην Μ.Ε.Ο.Κ. διότι,

1. Το σύνολο $\mathcal{S} = \{\underset{\sim}{x} : f(\underset{\sim}{x};\theta) > 0\} = \mathbb{R}_+^n$ δεν εξαρτάται από το θ.
2. $$f(\underset{\sim}{x},\theta) = \prod_{k=1}^{n} \frac{1}{\Gamma(k)\theta^k} x_k^{k-1} e^{-\frac{x_k}{\theta}} = \frac{1}{(\prod_{k=1}^{n}\Gamma(k))\theta^{n(n+1)/2}} \prod_{k=1}^{n} x_k^{k-1} e^{-\frac{\sum x_k}{\theta}}$$
$$= \exp\left\{-\sum_{k=1}^{n}\ln\Gamma(k) - \frac{n(n+1)}{2}\ln\theta + \sum_{k=1}^{n}(k-1)\ln x_k - \frac{1}{\theta}\sum_{k=1}^{n} x_k\right\}.$$

Θέτουμε
$A(\theta) = -\sum_{k=1}^{n}\ln\Gamma(k) - \dfrac{n(n+1)}{2}\ln\theta$, $B(\underset{\sim}{x}) = \sum_{k=1}^{n}(k-1)\ln x_k$, $C(\theta) = -\dfrac{1}{\theta}$ και $D(\underset{\sim}{x}) = \sum_{k=1}^{n} x_k$ καταλήγουμε στο συμπέρασμα ότι η οικογένεια κατανομών του $\underset{\sim}{X}$ είναι μια Μ.Ε.Ο.Κ.

Κάνοντας χρήση της Πρότασης 6.3.5, η στατιστική συνάρτηση $D(\underset{\sim}{X}) = \sum_{k=1}^{n} X_k$ είναι επαρκής και επειδή το σύνολο τιμών της συνάρτησης $C(\theta) = -\dfrac{1}{\theta}$ είναι το $(-\infty, 0)$, δηλαδή ανοικτό υποσύνολο του $\mathbb{R}$, είναι και πλήρης.

Η εύρεση ΑΟΕΔ εκτιμητή του θ, ανάγεται πλέον στο πρόβλημα ύπαρξης αμερόληπτου εκτιμητή του θ ο οποίος είναι συνάρτηση της επαρκούς και πλήρους στατιστικής συνάρτησης $D(\underset{\sim}{X}) = \sum_{k=1}^{n} X_k$. Υπολογίζοντας την $\mathbb{E}_\theta D(\underset{\sim}{X})$ προκύπτει ότι

$$\mathbb{E}_\theta D(\underset{\sim}{X}) = \mathbb{E}_\theta \sum_{k=1}^{n} X_k = \sum_{k=1}^{n}\mathbb{E}_\theta X_k = \sum_{k=1}^{n} k\theta = \frac{n(n+1)}{2}\theta,$$

οπότε $\mathbb{E}_\theta\left\{\frac{2}{n(n+1)}\sum\limits_{k=1}^{n}X_k\right\}=\theta$.

Συνεπώς, ο εκτιμητής $\dfrac{2}{n(n+1)}\sum\limits_{k=1}^{n}X_k$, ως αμερόληπτος εκτιμητής του θ και συνάρτηση της επαρκούς και πλήρους στατιστικής συνάρτησης $D(\underset{\sim}{X})$, είναι ο ΑΟΕΔ εκτιμητής του θ (βλέπε Πρόταση 6.3.4).

Εφ' όσον η στατιστική συνάρτηση $D=\sum\limits_{k=1}^{n}X_k$ χρησιμοποιήθηκε για την εύρεση ΑΟΕΔ εκτιμητή του θ, είναι λογικό να δουλέψουμε με τον D^r για την εκτίμηση του θ^r. Από την Πρόταση 1.8.5 (7) προκύπτει

$$D(\underset{\sim}{X})=\sum_{k=1}^{n}X_k\sim\mathcal{G}\left(\sum_{k=1}^{n}k,\theta\right)\equiv\mathcal{G}\left(\frac{n(n+1)}{2},\theta\right).$$

Έχουμε λοιπόν,

$$\mathbb{E}_\theta D^r=\int_0^\infty x^r\frac{1}{\Gamma(n(n+1)/2)\theta^{n(n+1)/2}}x^{n(n+1)/2-1}e^{-\frac{x}{\theta}}dx=\frac{\Gamma(n(n+1)/2+r)}{\Gamma(n(n+1)/2)}\theta^r$$

ή $\mathbb{E}_\theta\left\{\frac{\Gamma(n(n+1)/2)}{\Gamma(n(n+1)/2+r)}D^r\right\}=\theta^r$, για κάθε $\theta>0$ και $r>-\dfrac{n(n+1)}{2}$.
Επομένως, σύμφωνα με την Πρόταση 6.3.4, η στατιστική συνάρτηση $\frac{\Gamma(n(n+1)/2)}{\Gamma(n(n+1)/2+r)}D^r$ είναι ο ΑΟΕΔ εκτιμητής του θ^r.

Παράδειγμα 6.3.13. (**ανεξάρτητες Poisson παρατηρήσεις**) Έστω ότι $\underset{\sim}{X}=(X_1,\ldots,X_n)$ όπου οι X_i είναι ανεξάρτητες παρατηρήσεις, $X_i\sim\mathcal{P}(\theta t_i)$, $i=1,2,\ldots,n$, $\theta\in\Theta=(0,\infty)$ και $t_i>0$, γνωστές σταθερές $i=1,2,\ldots,n$. Θα βρούμε ΑΟΕΔ εκτιμητές των θ και θ^2.

Αρχικά θα βρούμε επαρκή και πλήρη στατιστική συνάρτηση. Η οικογένεια κατανομών του $\underset{\sim}{X}$ ανήκει στην Μ.Ε.Ο.Κ. διότι,

1. Το σύνολο $\mathcal{S}=\{\underset{\sim}{x}:f(\underset{\sim}{x};\theta)>0\}=\{0,1,\ldots\}^n$ δεν εξαρτάται από το θ.
2. $$f(\underset{\sim}{x},\theta)=\prod_{i=1}^{n}e^{-\theta t_i}\frac{(\theta t_i)^{x_i}}{x_i!}=e^{-\theta\sum_{i=1}^n t_i}\theta^{\sum_{i=1}^n x_i}\prod_{i=1}^{n}t_i^{x_i}\prod_{i=1}^{n}\frac{1}{x_i!}$$
$$=\exp\left\{-\theta\sum_{i=1}^{n}t_i+\sum_{i=1}^{n}x_i\ln t_i-\sum_{i=1}^{n}\ln x_i!+\ln\theta\sum_{i=1}^{n}x_i\right\}.$$

Θέτουμε,

$A(\theta)=-\theta\sum\limits_{i=1}^{n}t_i$, $B(\underset{\sim}{x})=\sum\limits_{i=1}^{n}x_i\ln t_i-\sum\limits_{i=1}^{n}\ln x_i!$, $C(\theta)=\ln\theta$ και

$D(\underset{\sim}{x}) = \sum_{i=1}^{n} x_i$ και καταλήγουμε στο συμπέρασμα ότι η οικογένεια κατανομών του $\underset{\sim}{X}$ είναι μία Μ.Ε.Ο.Κ..
Κάνοντας χρήση της Πρότασης 6.3.5, η στατιστική συνάρτηση $D(\underset{\sim}{X}) = \sum_{i=1}^{n} X_i$ είναι επαρκής και επειδή το σύνολο τιμών της συνάρτησης $C(\theta) = -\ln\theta$ είναι το $(-\infty, \infty)$, είναι και πλήρης.

Από την Πρόταση 1.8.5 προκύπτει

$$D = D(\underset{\sim}{X}) = \sum_{i=1}^{n} X_i \sim \mathcal{P}(\theta \sum_{i=1}^{n} t_i),$$

άρα

$$\mathbb{E}_\theta D = \theta \sum_{i=1}^{n} t_i \quad \text{ή} \quad \mathbb{E}_\theta \left\{ \frac{1}{\sum_{i=1}^{n} t_i} D \right\} = \theta, \ \ \forall\, \theta > 0. \qquad (6.18)$$

Κάνοντας χρήση της Πρότασης 6.3.4 προκύπτει ότι ο αμερόληπτος εκτιμητής $\frac{1}{\sum_{i=1}^{n} t_i} D$ του θ, είναι ο ΑΟΕΔ εκτιμητής του θ, αφού είναι συνάρτηση της επαρκούς και πλήρους στατιστικής συνάρτησης D.
Ομοίως,

$$\mathbb{E}_\theta D^2 = \mathbb{V}ar_\theta D + (\mathbb{E}_\theta D)^2 = \theta \sum_{i=1}^{n} t_i + (\sum_{i=1}^{n} t_i)^2 \theta^2.$$

Η παραπάνω σχέση, λόγω της σχέσης (6.18), γίνεται

$$\mathbb{E}_\theta D^2 = \mathbb{E}_\theta D + (\sum_{i=1}^{n} t_i)^2 \theta^2 \quad \text{ή} \quad \mathbb{E}_\theta \left\{ \frac{D(D-1)}{(\sum_{i=1}^{n} t_i)^2} \right\} = \theta^2, \quad \text{για κάθε} \quad \theta > 0,$$

οπότε, η στατιστική συνάρτηση $\frac{D(D-1)}{(\sum_{i=1}^{n} t_i)^2}$ είναι ο ΑΟΕΔ εκτιμητής του θ^2.

6.4 Βέλτιστοι εκτιμητές ειδικής μορφής

Από τη μελέτη της επάρκειας και συγκεκριμένα από την Πρόταση 6.2.2 γνωρίζουμε ότι για την εκτίμηση του $g(\theta)$, αρκεί να περιοριστούμε σε εκτιμητές που είναι συναρτήσεις (ελάχιστης) επαρκούς στατιστικής συνάρτησης–

οι υπόλοιποι εκτιμητές είναι μη αποδεκτοί. Όλοι οι πρώτοι εκτιμητές συνιστούν μία κατηγορία εκτιμητών ειδικής (συναρτησιακής) μορφής. Σε ορισμένες περιπτώσεις, η κατανομή του δείγματος $\underset{\sim}{X} = (X_1, \ldots, X_n)$ και ο τύπος της g τεκμηριώνουν περαιτέρω μείωση του αριθμού των εν δυνάμει εκτιμητών, γεγονός που καθιστά τεχνικά πιο εύκολη την εύρεση βέλτιστου εκτιμητή μεταξύ αυτών. Μία τέτοια περίπτωση περιγράφεται παρακάτω.
Έστω ότι η πυκνότητα στατιστικής συνάρτησης $T(\underset{\sim}{X})$ έχει τη μορφή $f_T(t;\theta) = h(t-\theta)$, $t \in \mathbb{R}$, $\theta \in \Theta = \mathbb{R}$, είναι δηλαδή συνάρτηση της διαφοράς $t-\theta$. Στις εφαρμογές η $T(\underset{\sim}{X})$ είναι επαρκής ή συνάρτηση επαρκούς στατιστικής συνάρτησης. Από την Πρόταση 1.7.1 προκύπτει ότι, για κάθε $c \in \mathbb{R}$, η $T^* = T + c$ έχει πυκνότητα $f_{T+c}(t;\theta) = f_T(t;\theta+c)$, δηλαδή η $T+c$ έχει πυκνότητα της ίδιας μορφής όπως η T αλλά με παράμετρο $\theta + c$ αντί θ. Συμβολικά, θα γράφουμε $T \sim f_T(t;\theta)$ και $T + c \sim f_T(t;\theta+c)$. Σε αυτήν την περίπτωση, το θ αναφέρεται ως *παράμετρος θέσης* (location parameter) για την οικογένεια των κατανομών της T. Ο μετασχηματισμός $T \to T + c$ αλλάζει, λοιπόν, τη «θέση» της κατανομής της T, από θ σε $\theta + c$.
Έστω τώρα ότι για την εκτίμηση του $g(\theta) = \theta$ υιοθετούμε τον εκτιμητή $S(T)$, όπου ας μην ξεχνάμε ότι $T \sim f_T(t;\theta)$. Επομένως, το γεγονός ότι $T^* \sim f_T(t;\theta+c)$ δικαιολογεί την υιοθέτηση του $S(T^*)$ για την εκτίμηση του $\theta + c$. Όμως, το $\theta + c$ μπορεί επίσης να εκτιμηθεί με τον εκτιμητή $S(T) + c$ με βάση την αρχή της αντικατάστασης (Ενότητα 3.3). Περαιτέρω είναι λογικά ορθό να απαιτήσουμε να συμπίπτουν οι δύο αυτοί εκτιμητές του $\theta + c$, αφού έχουν προκύψει από τον *ίδιο* εκτιμητή του θ, τον $S(T)$. Απαιτούμε δηλαδή ο $S(T)$ να ικανοποιεί *τη συνθήκη*

$$S(T^*) = S(T) + c \;\text{ ή ισοδύναμα }\; S(T+c) = S(T) + c, \quad \text{για κάθε } \; c \in \mathbb{R}.$$

Η τελευταία σχέση συνεπάγεται $S(t+c) = S(t) + c$, για κάθε τιμή t της T, οπότε θέτοντας $c = -t$ προκύπτει ότι $S(0) = S(t) - t$ ή $S(t) = t + S(0)$ που ισοδυναμεί με $S(T) = T + S(0)$. Το $S(0)$ είναι μία πραγματική σταθερά, ας τη συμβολίσουμε με c_0 (μόνον σημειολογική σημασία έχει αυτός ο συμβολισμός), οπότε $S(T) = T + c_0$.

Εν κατακλείδι, όταν $T \sim f_T(t;\theta) = h(t-\theta)$, $t \in \mathbb{R}$, $\theta \in \Theta = \mathbb{R}$, για την εκτίμηση του θ τεκμηριώνεται σύμφωνα με το παραπάνω σκεπτικό η χρησιμοποίηση εκτιμητών της μορφής

$$S(T) = T + c_0, \tag{6.19}$$

όπου $c_0 = S(0)$ είναι τυχούσα πραγματική σταθερά. Η (6.19) ορίζει λοιπόν μία κλάση εκτιμητών του θ και η επόμενη πρόταση δίνει τον βέλτιστο εκτιμητή μέσα σε αυτήν.

Πρόταση 6.4.1. *Έστω $T \sim f_T(t;\theta) = h(t-\theta)$, $t \in \mathbb{R}$, $\theta \in \Theta = \mathbb{R}$. Θεωρούμε την κλάση εκτιμητών του θ, $C = \{T + c_0 \colon c_0 \in \mathbb{R}\}$. Τότε, με κριτήριο το ΜΤΣ, ο βέλτιστος εκτιμητής στην κλάση C είναι $T + c^*$, όπου $c^* = -\mathbb{E}_{\theta=0}T$.*

Απόδειξη. Επειδή, όπως είδαμε παραπάνω $T + c \sim f_T(t;\theta + c)$ για κάθε $c \in \mathbb{R}$, θέτοντας $c = -\theta$ συμπεραίνουμε ότι $T - \theta \sim f_T(t;0)$. Άρα η τυχαία μεταβλητή $T - \theta$ έχει κατανομή που δεν εξαρτάται από το θ και συνεπώς συμπίπτει με την κατανομή της T όταν $\theta = 0$. Επομένως, το ΜΤΣ του $T + c_0$ είναι

$$\text{ΜΤΣ}(T + c_0, \theta) = \mathbb{E}_\theta(T + c_0 - \theta)^2 = \mathbb{E}_{\theta=0}(T + c_0)^2 = c_0^2 + 2c_0\mathbb{E}_0T + \mathbb{E}_0T^2,$$

το οποίο ελαχιστοποιείται ως προς c_0 για την τιμή $c_0 = c^* = -\mathbb{E}_0T$. □

Παράδειγμα 6.4.1. (μετατοπισμένη εκθετική κατανομή-συνέχεια του Παραδείγματος 6.3.7) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την κατανομή με πυκνότητα $f_1(x;\theta) = e^{-(x-\theta)}$, $x \geq \theta$, $\theta \in \Theta = \mathbb{R}$. Στο Παράδειγμα 6.3.7 δείξαμε ότι $T(\underset{\sim}{X}) = X_{(1)} = \min(X_1, \ldots, X_n)$ είναι επαρκής και πλήρης με πυκνότητα $f_T(t;\theta) = ne^{-n(t-\theta)}$, $t \geq \theta$. Περαιτέρω,

$$\mathbb{E}_\theta T = \int_\theta^\infty t\, ne^{-n(t-\theta)}\, dt = \theta + \frac{1}{n} \tag{6.20}$$

οπότε $\mathbb{E}_\theta\left(T - \frac{1}{n}\right) = \theta$, για κάθε $\theta \in \Theta$. Από την Πρόταση 6.3.4, συμπεραίνουμε ότι $T - \frac{1}{n}$ είναι ο ΑΟΕΔ του θ ως αμερόληπτος και συνάρτηση της επαρκούς και πλήρους στατιστικής συνάρτησης T. Παρατηρούμε, επιπλέον, ότι η πυκνότητα της T μπορεί να γραφεί στη μορφή

$$f_T(t;\theta) = ne^{-n(t-\theta)}\mathrm{I}_{[0,\infty)}(t-\theta) = h(t-\theta),$$

όπου $h(x) = ne^{-nx}\mathrm{I}_{[0,\infty)}(x)$. Οι προϋποθέσεις της Πρότασης 6.4.1 λοιπόν ικανοποιούνται, συνεπώς ο βέλτιστος εκτιμητής της κλάσης $C = \{T + c_0 \colon c_0 \in \mathbb{R}\}$ είναι $T + c^*$, όπου $c^* = -\mathbb{E}_{\theta=0}T = -\frac{1}{n}$ λόγω της (6.20) δηλαδή ο ΑΟΕΔ εκτιμητής. Συμπερασματικά ο $T - \frac{1}{n}$ είναι ο βέλτιστος μεταξύ όλων των αμερόληπτων εκτιμητών του θ καθώς επίσης μεταξύ όλων των μη αμερόληπτων της μορφής $T + c_0$, $c_0 \in \mathbb{R}$.

Σε αυτό το σημείο θα προσθέσουμε ότι δεν ήταν απαραίτητη η επίκληση της Πρότασης 6.4.1 προκειμένου να διαπιστώσουμε ότι ο ΑΟΕΔ του Παραδείγματος 6.4.1 είναι ο βέλτιστος της κλάσης C. Αρκεί το γεγονός ότι είναι μέλος της κλάσης C (με $c_0 = -\frac{1}{n}$) όπως αποδεικνύεται στην επόμενη πρόταση.

Πρόταση 6.4.2. *Θεωρούμε την κλάση εκτιμητών του $g(\theta)$, $C = \{T + c_0 \colon c_0 \in \mathbb{R}\}$, όπου T είναι τυχούσα στατιστική συνάρτηση. Εάν υπάρχει αμερόληπτος εκτιμητής του $g(\theta)$ που ανήκει στην C, τότε αυτός είναι ο βέλτιστος εκτιμητής στην C με κριτήριο το ΜΤΣ.*

Απόδειξη. Έστω $T + c_1$, $c_1 \in \mathbb{R}$ αμερόληπτος εκτιμητής του $g(\theta)$ που ανήκει στην κλάση C. Τότε $\mathbb{E}_\theta(T + c_1) = g(\theta)$ και

$$\text{ΜΤΣ}(T + c_1, \theta) = \mathbb{V}ar_\theta(T + c_1) = \mathbb{V}ar_\theta T.$$

Επιπλέον από την (4.2), για κάθε άλλον εκτιμητή $T + c_0$, $c_0 \neq c_1$, προκύπτει ότι

$$\begin{aligned}
\text{ΜΤΣ}(T + c_0, \theta) &= \mathbb{V}ar_\theta(T + c_0) + (\mathbb{E}_\theta(T + c_0) - g(\theta))^2 \\
&= \mathbb{V}ar_\theta T + (\mathbb{E}_\theta(T + c_1) - g(\theta) + c_0 - c_1)^2 \\
&= \mathbb{V}ar_\theta T + (c_0 - c_1)^2 \\
&> \mathbb{V}ar_\theta T = \text{ΜΤΣ}(T + c_1, \theta).
\end{aligned}$$

□

Μία ακόμη ενδιαφέρουσα περίπτωση ύπαρξης βέλτιστου εκτιμητή είναι η ακόλουθη. Έστω ότι η πυκνότητα στατιστικής συνάρτησης $T(\underset{\sim}{X})$ έχει τη μορφή $f_T(t;\theta) = \frac{1}{\theta}h(\frac{t}{\theta})$, $t > 0$, $\theta \in \Theta = (0, \infty)$, είναι δηλαδή συνάρτηση του λόγου $\frac{t}{\theta}$. Στις εφαρμογές η $T(\underset{\sim}{X})$ είναι επαρκής ή συνάρτηση επαρκούς στατιστικής συνάρτησης. Από την Πρόταση 1.7.1 προκύπτει ότι, για κάθε $c > 0$, $T^* = cT \sim f_T(t; c\theta)$, δηλαδή η cT έχει πυκνότητα της ίδιας μορφής όπως η T αλλά με παράμετρο $c\theta$ αντί θ. Σε αυτήν την περίπτωση, το θ αναφέρεται ως *παράμετρος κλίμακος* (scale parameter) για την οικογένεια κατανομών της T. Ο μετασχηματισμός $T \to cT$ μπορεί να θεωρηθεί ότι αλλάζει κλίμακα μέτρησης (μονάδα μέτρησης) της T.
Έστω τώρα $S(T)$ εκτιμητής του θ. Τότε, σκεπτόμενοι παρόμοια με την περίπτωση παραμέτρου θέσης, για την εκτίμηση του $c\theta$ δικαιολογείται υιοθέτηση του εκτιμητή $S(T^*)$, επειδή $T^* \sim f_T(t; c\theta)$, αλλά και του $cS(T)$ από την αρχή της αντικατάστασης (Ενότητα 3.3). Για αυτούς τους δύο εκτιμητές του $c\theta$, που έχουν προκύψει από τον ίδιο αρχικό εκτιμητή του θ, τον $S(T)$, είναι εύλογο να απαιτήσουμε να συμπίπτουν, με άλλα λόγια να ικανοποιούν τη συνθήκη

$$S(T^*) = c\,S(T) \ \text{ ή ισοδύναμα } \ S(ct) = c\,S(t),$$

για κάθε τιμή $t > 0$ της T και κάθε $c > 0$. Θέτοντας $c = \frac{1}{t}$, προκύπτει ότι $S(1) = \frac{1}{t}S(t)$ ή $S(t) = S(1)t$ που ισοδυναμεί με $S(T) = S(1)T$ ή $S(T) = c_0T$ όπου $c_0 = S(1)$.
Συμπερασματικά όταν $T \sim f_T(t;\theta) = \frac{1}{\theta}h(\frac{t}{\theta})$, $t > 0$, $\theta \in \Theta = (0, \infty)$, για την εκτίμηση του θ τεκμηριώνεται η χρησιμοποίηση εκτιμητών της μορφής

$$S(T) = c_0T, \tag{6.21}$$

όπου $c_0 = S(1)$ είναι τυχούσα θετική σταθερά. Η (6.21) ορίζει λοιπόν μία κλάση εκτιμητών του θ και η επόμενη πρόταση δίνει τον βέλτιστο εκτιμητή εξ αυτών.

Πρόταση 6.4.3. *Έστω $T \sim f_T(t;\theta) = \frac{1}{\theta}h(\frac{t}{\theta})$, $t > 0$, $\theta \in \Theta = (0, \infty)$. Θεωρούμε την κλάση εκτιμητών του θ, $C = \{c_0T \colon c_0 > 0\}$. Τότε, με κριτήριο το ΜΤΣ, ο βέλτιστος εκτιμητής στην κλάση C είναι c^*T, όπου $c^* = \frac{\mathbb{E}_{\theta=1}T}{\mathbb{E}_{\theta=1}T^2}$.*

Απόδειξη. Επειδή, όπως είδαμε παραπάνω, $cT \sim f_T(t; c\theta)$ για κάθε $c > 0$, θέτοντας $c = \frac{1}{\theta}$ συμπεραίνουμε ότι $\frac{T}{\theta} \sim f_T(t; 1)$. Άρα η τυχαία μεταβλητή $\frac{T}{\theta}$ έχει κατανομή που δεν εξαρτάται από το θ και συνεπώς συμπίπτει με την κατανομή της T όταν $\theta = 1$. Επομένως, το ΜΤΣ του $c_0 T$ είναι

$$\begin{aligned}\text{ΜΤΣ}(c_0T, \theta) &= \mathbb{E}_\theta(c_0T - \theta)^2 = \theta^2 \mathbb{E}_\theta(c_0\frac{T}{\theta} - 1)^2 \\ &= \theta^2 \mathbb{E}_{\theta=1}(c_0T - 1)^2 = \theta^2 \left\{c_0^2 \mathbb{E}_1 T^2 - 2c_0 \mathbb{E}_1 T + 1\right\},\end{aligned}$$

το οποίο ελαχιστοποιείται ως προς c_0 για την τιμή $c_0 = c^* = \frac{\mathbb{E}_1 T}{\mathbb{E}_1 T^2}$. □

Παραθέτουμε στη συνέχεια μερικά παραδείγματα εφαρμογής της Πρότασης 6.4.3.

Παράδειγμα 6.4.2. (συνέχεια του Παραδείγματος 6.3.10, κανονική κατανομή με άγνωστη μέση τιμή και διασπορά– μη αποδεκτικότητα της δειγματικής διασποράς) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$, $n \geq 2$, ένα τυχαίο δείγμα από την κανονική κατανομή $\mathcal{N}(\mu, \sigma^2)$, $\underset{\sim}{\theta} = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+$. Ο ΑΟΕΔ εκτιμητής του σ^2 είναι η δειγματική διασπορά $S^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \bar{X})^2$, όπως είδαμε στο Παράδειγμα 6.3.10. Όμως, από την Πρόταση 4.3.1, ο εκτιμητής $S_1^2 = c_1 S^2$, $c_1 = \frac{n(n-1)}{n(n-1)+2}$ έχει μικρότερο ΜΤΣ από τον S^2 για κάθε κατανομή, άρα και για την $\mathcal{N}(\mu, \sigma^2)$. Επομένως ο S^2 είναι μη αποδεκτός με κριτήριο το ΜΤΣ. Στην προκειμένη μάλιστα περίπτωση, μέσω της Πρότασης 6.4.3, θα δείξουμε ότι ο εκτιμητής $\frac{1}{n+1}\sum_{i=1}^n (X_i - \bar{X})^2$ είναι ο καλύτερος εκτιμητής του σ^2 με κριτήριο το ΜΤΣ μεταξύ όλων των εκτιμητών της μορφής $c\sum_{i=1}^n (X_i - \bar{X})^2$, όπου c θετική σταθερά, άρα και από τον S^2 και τον S_1^2. Η απόδειξη έχει ως εξής.
Θέτουμε $T = \sum\limits_{i=1}^{n}(X_i - \bar{X})^2$, οπότε από την Πρόταση 1.8.6 γνωρίζουμε ότι $\frac{T}{\sigma^2} \sim \chi^2_{n-1}$ με πυκνότητα

$$h(x) = \frac{1}{\Gamma\left(\frac{n-1}{2}\right) 2^{\frac{n-1}{2}}}\, x^{\frac{n-1}{2}-1} e^{-\frac{x}{2}},\ x > 0.$$

Από την Πρόταση 1.7.1 η πυκνότητα της T είναι

$$f(t; \sigma^2) = \frac{1}{\sigma^2} h\left(\frac{t}{\sigma^2}\right), \quad t > 0.$$

Εφαρμόζοντας λοιπόν την Πρόταση 6.4.3 (με θ το σ^2) έχουμε

$$c^* = \frac{\mathbb{E}_{\sigma^2=1}T}{\mathbb{E}_{\sigma^2=1}T^2} = \frac{\mathbb{E}_{\sigma^2=1}T}{\mathbb{V}ar_{\sigma^2=1}T + (\mathbb{E}_{\sigma^2=1}T)^2} = \frac{n-1}{2(n-1)+(n-1)^2} = \frac{1}{n+1},$$

δηλαδή ο $\frac{1}{n+1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ είναι ο βέλτιστος στην κλάση $c\sum_{i=1}^{n}(X_i - \bar{X})^2$.

Παράδειγμα 6.4.3. (συνέχεια του Παραδείγματος 6.3.3, ομοιόμορφη κατανομή– μη αποδεκτικότητα του ΑΟΕΔ εκτιμητή) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την ομοιόμορφη κατανομή $\mathcal{U}(0, \theta)$, $\theta \in \Theta = (0, \infty)$. Η πυκνότητα της επαρκούς και πλήρους στατιστικής συνάρτησης $T(\underset{\sim}{X}) = X_{(n)}$ είναι, όπως είδαμε στο Παράδειγμα 6.3.3,

$$f_T(t;\theta) = \frac{n}{\theta^n}t^{n-1}\mathrm{I}_{(0,\theta)}(t) = \frac{1}{\theta}h\left(\frac{t}{\theta}\right),$$

όπου $h(x) = nx^{n-1}\mathrm{I}_{(0,1)}(x)$. Η Πρόταση 6.4.3, συνεπώς, συνεπάγεται ότι μεταξύ των εκτιμητών του θ, που έχουν τη μορφή cT, $c > 0$, ο βέλτιστος εκτιμητής ως προς το ΜΤΣ είναι ο c^*T, όπου

$$c^* = \frac{\mathbb{E}_{\theta=1}T}{\mathbb{E}_{\theta=1}T^2} = \frac{\int_0^1 nt^n dt}{\int_0^1 nt^{n+1}dt} = \frac{n+2}{n+1},$$

δηλαδή ο $\frac{n+2}{n+1}T$. Ειδικά, ο $\frac{n+2}{n+1}T$ είναι καλύτερος από τον ΑΟΕΔ εκτιμητή $\frac{n+1}{n}T$ (βλέπε το Παράδειγμα 6.3.3). Επίσης μη αποδεκτός είναι και ο ΑΟΕΔ της μέσης τιμής $\frac{\theta}{2}$, $\frac{n+1}{2n}T$, επειδή ο $\frac{n+2}{2(n+1)}T$ έχει μικρότερο ΜΤΣ.

6.5 Αποτίμηση της αμεροληψίας

Η ιδιότητα της αμεροληψίας μελετήθηκε ενδελεχώς στα Κεφάλαια 4, 5 και 6. Είναι μια λογική ιδιότητα–συνθήκη που εκφράζει «κατάσταση ουδετερότητας» από την πλευρά του εκτιμητή. Συνδέεται άμεσα με το ΜΤΣ και μάλιστα το ερμηνεύει αφού για αμερόληπτους εκτιμητές το ΜΤΣ συμπίπτει με τη διασπορά του. Είναι μια εύπεπτη και καλοδεχούμενη ιδιότητα σε ευρεία κλίμακα χρηστών Στατιστικής Μεθοδολογίας. Εστιάζοντας σε αμερόληπτους εκτιμητές, είναι δυνατόν σε πολλές περιπτώσεις να βρεθεί ο βέλτιστος μεταξύ αυτών ως προς τη διασπορά, δηλαδή ο ΑΟΕΔ εκτιμητής. Ένα σημείο άσκησης κριτικής για τους ΑΟΕΔ εκτιμητές είναι το γεγονός ότι το σύνολο τιμών τους δεν συμπίπτει κατ' ανάγκη με το σύνολο των εν δυνάμει τιμών της εκτιμώμενης «ποσότητας». Αν και αυτό δεν συμβαίνει στις περισσότερες των πρακτικών εφαρμογών, ωστόσο δεν μπορεί να αποκλειστεί. Ως ένα (θεωρητικό) παράδειγμα, στην περίπτωση τυχαίου δείγματος από την κανονική κατανομή $\mathcal{N}(\theta, 1)$, $\theta \in \Theta = \mathbb{R}$, ο ΑΟΕΔ εκτιμητής του θ είναι ο δειγματικός μέσος $\bar{X}$ που έχει (δυνητικά) σύνολο τιμών επίσης το $\mathbb{R}$, όμως ο ΑΟΕΔ εκτιμητής του θ^2 είναι $\bar{X}^2 - \frac{1}{n}$, ο οποίος, με θετική πιθανότητα, μπορεί να πάρει αρνητικές τιμές ενώ $\theta^2 \in [0, \infty)$. Ειδικά στην περίπτωση που η πραγματική (αλλά άγνωστη) τιμή του θ είναι 0, η πιθανότητα αρνητικής εκτίμησης του θ^2 είναι πολύ μεγάλη καθώς $\bar{X} \sim \mathcal{N}\left(0, \frac{1}{n}\right)$, οπότε $Z = \sqrt{n}\bar{X} \sim \mathcal{N}(0, 1)$ και επομένως

$$
\begin{aligned}
\mathbb{P}_{\theta=0}\left(\left|\bar{X}\right| < \frac{1}{\sqrt{n}}\right) &= \mathbb{P}_0\left(\left|\sqrt{n}\bar{X}\right| < 1\right) = \mathbb{P}\left(|Z| < 1\right) \\
&= \mathbb{P}\left(-1 < Z < 1\right) = \Phi(1) - \Phi(-1) = \Phi(1) - (1 - \Phi(1)) \\
&= 2\Phi(1) - 1 \simeq 2 \cdot 0.84 - 1 = 0.68(!!!),
\end{aligned}
$$

όσο μεγάλο και αν είναι το μέγεθος δείγματος n.

Μια άλλη δυσκολία είναι ότι σε αρκετές πριπτώσεις, η «επιβολή» της αμεροληψίας μπορεί να προκαλέσει την «εισαγωγή πρόσθετης διασποράς», ακόμη και αν πρόκειται για τον ΑΟΕΔ εκτιμητή, με τελικό προϊόν τη δημιουργία σχετικά μεγάλου ΜΤΣ. Η σχέση αμεροληψίας και διασποράς μελετήθηκε διεξοδικά στο Παράδειγμα 4.1.3. Η πιο ηχηρή πάντως περίπτωση είναι αυτή του Παραδείγματος 6.4.2 όπου ο ΑΟΕΔ εκτιμητής της διασποράς κανονικής κατανομής με (επίσης) άγνωστη μέση τιμή, $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ έχει μεγαλύτερο ΜΤΣ (άρα είναι μη αποδεκτός) από τον εκτιμητή $S_1^2 = \frac{1}{n+1}\sum_{i=1}^{n}(X_i - \overline{X})^2$. Μία ακόμη περίπτωση είναι αυτή του Παραδείγματος 6.4.3

Στοιχείο σκεπτικισμού αποτελεί επίσης η εγγενής αδυναμία «μεταβίβασης» της αμεροληψίας σε μετασχηματισμούς αμερόληπτων εκτιμητών, με εξαίρεση τους γραμμικούς μετασχηματισμούς. Η αδυναμία αυτή έχει σχολιαστεί στην Παρατήρηση 4.2.1.

Κλείνοντας την αποτίμηση της αμεροληψίας, σημειώνουμε ότι το κατά πόσον θα επιδιώξουμε τη χρησιμοποίηση αμερόληπτου εκτιμητή σε ένα πρόβλημα εκτίμησης δεν είναι ένα ερώτημα που μπορεί να απαντηθεί γενικά με ένα ναι ή ένα όχι. Η απόφαση πρέπει να ληφθεί λαμβάνοντας πρωτίστως υπόψη το φυσικό πρόβλημα, τον σκοπό της εκτίμησης και τις συνέπειές της.

6.6 Ασκήσεις

6.1. Εάν $\underset{\sim}{X} = (X_1, \ldots, X_n)$, $n \geq 2$, είναι ένα τυχαίο δείγμα από την κανονική κατανομή $\mathcal{N}(\mu, \sigma^2)$ με σ^2 άγνωστο, $\sigma^2 \in (0, \infty)$, να βρεθεί ο ΑΟΕΔ εκτιμητής του σ^κ, $\kappa > 0$. (2 περιπτώσεις)

6.2. Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την κατανομή Γάμμα $\mathcal{G}(\alpha, \theta)$ με α γνωστό και $\theta \in \Theta = (0, \infty)$. Να βρεθεί ο ΑΟΕΔ εκτιμητής του θ.

6.3. Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από τη γεωμετρική κατανομή $\mathcal{G}e(\theta)$, $\theta \in \Theta = (0, 1)$. Να δειχθεί ότι η κατανομή αυτή ανήκει στην Μ.Ε.Ο.Κ., να βρεθεί επαρκής και πλήρης στατιστική συνάρτηση καθώς και ο ΑΟΕΔ εκτιμητής του $\frac{1}{\theta}$.

6.4. Έστω X μία παρατήρηση από την ομοιόμορφη κατανομή $\mathcal{U}(-\theta, \theta)$, $\theta \in \Theta = (0, \infty)$. Να δειχθεί ότι η X είναι επαρκής αλλά δεν είναι πλήρης. Να ευρεθεί ο ΑΟΕΔ εκτιμητής του θ.

6.5. Ποιες από τις εξής οικογένειες κατανομών είναι Ε.Ο.Κ.;

α. Κατανομή Γάμμα $\mathcal{G}(\alpha, \beta)$, $f(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}$, $x > 0$. (3 περιπτώσεις)

β. Κατανομή Βήτα $\mathcal{B}eta(\alpha, \beta)$, $f(x; \alpha, \beta) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-z)^{\beta-1}$, $0 < x < 1$. (3 περιπτώσεις)

γ. $f(x; \theta, \gamma) = \frac{1}{\theta} e^{-\frac{1}{\theta}(x-\gamma)}$, $x \geq \gamma$. (3 περιπτώσεις)

Σε όσες περιπτώσεις οι κατανομές ανήκουν στην Ε.Ο.Κ. να βρεθεί επαρκής και πλήρης στατιστική συνάρτηση με βάση ένα τυχαίο δείγμα από την κατανομή.

6.6. Η παρατήρηση X έχει κατανομή που δίνεται στον πίνακα για $\theta \in \Theta = \{-1, 0, 1\}$. Είναι η στατιστική συνάρτηση $T(X) = X$ πλήρης;

	$X = 1$	$X = 2$	$X = 3$
$\theta = -1$	$1/6$	$1/6$	$2/3$
$\theta = 0$	$1/3$	$1/3$	$1/3$
$\theta = 1$	$2/3$	$1/6$	$1/6$

6.7. Έστω X μία παρατήρηση από την κατανομή Poisson, $\mathcal{P}(\theta)$, $\theta \in \Theta = (0, \infty)$. Να δειχθεί ότι ο ΑΟΕΔ εκτιμητής του $g(\theta) = e^{-2\theta}$ είναι $T(X) = (-1)^X$. (Ο εκτιμητής T είναι πολύ «άστοχος» εκτιμητής αφού παίρνει τιμές -1 ή 1 ενώ $0 < e^{-2\theta} < 1$.)

6.8. Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$, $n \geq 2$, ένα τυχαίο δείγμα από την διωνυμική κατανομή $\mathcal{B}(2, \theta)$, $0 < \theta < 1$. Να βρεθεί ο ΑΟΕΔ εκτιμητής του $g(\theta) = \mathbb{P}_\theta(X_1 \leq 1)$.

6.9. Εάν $\underset{\sim}{X} = (X_1, \ldots, X_n)$ είναι ένα τυχαίο δείγμα από την ομοιόμορφη κατανομή $\mathcal{U}(\theta - 1/2, \theta + 1/2)$ να βρεθεί (ελάχιστη) επαρκής στατιστική συνάρτηση.

6.10. Εάν $\underset{\sim}{X} = (X_1, \ldots, X_n)$ είναι τυχαίο δείγμα από την κανονική κατανομή $\mathcal{N}(\mu, \sigma^2)$ με μ, σ^2 άγνωστα, να βρεθεί ο ΑΟΕΔ εκτιμητής του μ/σ.

6.11. Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ και $\underset{\sim}{Y} = (Y_1, \ldots, Y_m)$ δύο ανεξάρτητα τυχαία δείγματα με την ίδια άγνωστη μέση τιμή θ και γνωστές διασπορές σ_1^2 και σ_2^2 αντίστοιχα. Να δειχθεί ότι η στατιστική συνάρτηση $T_c = c\bar{X} + (1-c)\bar{Y}$, $0 \leq c \leq 1$ είναι αμερόληπτος εκτιμητής του θ. Να βρεθεί η τιμή του c για την οποία η διασπορά του T_c είναι ελάχιστη. Είναι οι εκτιμητές $\bar{X}$ και $\bar{Y}$ αποδεκτοί ως εκτιμητές του θ;

6.12. Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$, $n \geq 2$, ένα τυχαίο δείγμα από την ομοιόμορφη κατανομή $\mathcal{U}(\theta, 2\theta)$, $\theta \in \Theta = (0, \infty)$. Να δειχθεί ότι $T = \frac{2}{3}\bar{X}$ είναι αμερόληπτος εκτιμητής του θ αλλά μη αποδεκτός.

6.13. Εάν T_1, T_2 είναι ΑΟΕΔ εκτιμητές των $g_1(\theta)$, $g_2(\theta)$ αντίστοιχα, να δειχθεί ότι $T_1 + T_2$ είναι ΑΟΕΔ εκτιμητής του $g_1(\theta) + g_2(\theta)$.

6.14. Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την εκθετική κατανομή $\mathcal{E}(\theta)$. Να βρεθεί με κριτήριο το ΜΤΣ, ο βέλτιστος εκτιμητής του θ, στην κλάση $C = \left\{c\bar{X} : c > 0\right\}$.

6.15. Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$, $n \geq 2$, ένα τυχαίο δείγμα από την κανονική κατανομή $\mathcal{N}(\theta, \alpha\theta^2)$, με $\theta \in \Theta = (0, \infty)$ και α γνωστή σταθερά.

α. Να βρεθεί αμερόληπτος εκτιμητής του θ της μορφής cS, όπου $S^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \bar{X})^2$ και $c > 0$.

β. Να βρεθεί αμερόληπτος εκτιμητής του θ, της μορφής $c\bar{X}$.

γ. Από τα α. και β. να δειχθεί ότι η επαρκής στατιστική συνάρτηση του Παραδείγματος 6.3.9 δεν είναι πλήρης.

δ. (Khan, 1968) Μεταξύ των αμερόληπτων εκτιμητών του θ της μορφής $\alpha\bar{X} + \beta S$ να βρεθεί αυτός που έχει την ελάχιστη διασπορά.

ε. Να δειχθεί ότι ο $\bar{X}$ είναι μη αποδεκτός εκτιμητής του θ με κριτήριο το ΜΤΣ.

6.16. Έστω T στατιστική συνάρτηση και θ παράμετρος κλίμακας για την οικογένεια κατανομών της T, δηλαδή η T έχει πυκνότητα $f_T(t;\theta) = \frac{1}{\theta}h\left(\frac{t}{\theta}\right)$, $t > 0$, $\theta \in \Theta = (0, \infty)$. Θεωρούμε την κλάση των εκτιμητών του θ, $C = \{cT : c > 0\}$ και τη συνάρτηση ζημίας $L(t, \theta) = \frac{t}{\theta} - \ln\frac{t}{\theta} - 1$ (γνωστή ως entropy loss ή Stein loss).

α. Να δειχθεί ότι το μέλος της κλάσης C, c_0T όπου $c_0 = \frac{1}{\mathbb{E}_{\theta=1}T}$ είναι ο αμερόληπτος εκτιμητής του θ.

β. Ο c_0T (εκτός από αμερόληπτος) είναι επιπλέον ο βέλτιστος εκτιμητής της κλάσης C με κριτήριο τη μέση ζημία $\mathbb{E}L(cT, \theta)$.

6.17. (Η δειγματική διασπορά «αντεπιτίθεται». Στο Παράδειγμα 6.4.2 αλλά και στην Ενότητα 4.3 είδαμε ότι στην περίπτωση κανονικής κατανομής με άγνωστη μέση τιμή και άγνωστη διασπορά, υπάρχουν εκτιμητές της μορφής $c\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2$ με μικρότερο ΜΤΣ από αυτό του ΑΟΕΔ εκτιμητή $S^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2$. Η άσκηση αυτή δείχνει μία βέλτιστη ιδιότητα του S^2 ως προς τη συνάρτηση ζημίας του Stein.)
Έστω $\underset{\sim}{X}=(X_1,\ldots,X_n)$, $n\geq 2$, ένα τυχαίο δείγμα από την κανονική κατανομή $\mathcal{N}(\mu,\sigma^2)$, με μ και σ^2 άγνωστα. Για την εκτίμηση του σ^2 θεωρούμε την κλάση των εκτιμητών $C=\left\{T_c=c\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2 : c>0\right\}$ και την συνάρτηση ζημίας του Stein $L(t,\sigma^2)=\frac{t}{\sigma^2}-\ln\frac{t}{\sigma^2}-1$. Να δειχθεί ότι ο βέλτιστος εκτιμητής στην κλάση C με κριτήριο τη μέση ζημία $\mathbb{E}L(T_c,\theta)$ είναι ο ΑΟΕΔ εκτιμητής του σ^2, δηλαδή η δειγματική διασπορά $S^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2$.

6.18. Έστω $\underset{\sim}{X}=(X_1,\ldots,X_n)$, $n\geq 2$, ένα τυχαίο δείγμα από την ομοιόμορφη κατανομή $\mathcal{U}(-\theta,\theta)$, $\theta\in\Theta=(0,\infty)$. Θέτουμε $Y_{(n)}=\max\{Y_1,\ldots,Y_n\}$ όπου $Y_i=|X_i|$, $i=1,\ldots,n$.

α. Να δειχθεί ότι η $Y_{(n)}$ είναι επαρκής και πλήρης με πυκνότητα

$$n\frac{y^{n-1}}{\theta^n},\quad 0\leq y<\theta.$$

β. Να βρεθεί ο ΑΟΕΔ εκτιμητής της διασποράς της κατανομής, $\frac{\theta^2}{3}$.

γ. Να βρεθεί ο βέλτιστος εκτιμητής του $\frac{\theta^2}{3}$ στην κλάση των εκτιμητών $C=\left\{cY_{(n)}^2 : c>0\right\}$ με κριτήριο το ΜΤΣ.

δ. Να δειχθεί ότι η δειγματική διασπορά $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ και $S_0^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2$ είναι μη αποδεκτοί και να βρεθούν καλύτεροι εκτιμητές από αυτούς.

ε. Να δειχθεί ότι ο ΑΟΕΔ εκτιμητής του $\frac{\theta^2}{3}$ είναι μη αποδεκτός βρίσκοντας έναν καλύτερο εκτιμητή.

Βιβλιογραφία

1. Blackwell, D. (1947). Conditional expectation and unbiased sequential estimation. *Ann. Math. Statistics*, **18**, 105–110.

2. Casella, G. and Berger, R.L. (2002). *Statistical Inference*. Duxbury Press; 2nd edition.

3. Fisher, R.A. (1922). On the mathematical foundations of theoretical statistics. *Philos. Trans. Roy. Soc. London Ser. A*, **222**, 309–368.

4. Halmos, P.R. and Savage, L.J. (1949). Application of the Radon-Nikodym theorem to the theory of sufficient statistics. *Ann. Math. Statistics*, **20**, 225–241.

5. Khan, R.A. (1968). A note on estimating the mean of a normal distribution with known coefficient of variation. *J. Am. Stat. Assoc.*, **63**, 1039–1041.

6. Lehmann, E.L. and Romano, J.P. (2005). *Testing statistical hypotheses*. Springer; Third edition, New York.

7. Lehmann, E.L. and Casella, G. (1998). *Theory of point estimation*. Springer; 2nd edition.

8. Lehmann, E.L. and Scheffé, H. (1955). Completeness, similar regions, and unbiased estimation. II. *Sankhyã*, **15**, 219–236.

9. Mood, A., Graybill, F. and Boes, D. (1974). *Introduction to the Theory of Statistics*. McGraw-Hill; 3rd edition.

10. Neyman, J. (1935). Sur la verification des hypotheses statistiques composees. *Bull. Soc. Math. France*, **63**, 246–266.

11. Rao, C. R. (1945). Information and the accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.*, **37**, 81–89.

12. Rao, C. R. (1973). *Linear Statistical Inference and its Applications*. Wiley: 2nd edition.

13. Rohatgi, V.K. (1976). *An Introduction to Probability Theory and Mathematical Statistics*. Wiley.

14. Ηλιόπουλος, Γ. (2013). *Βασικές Μέθοδοι Εκτίμησης Παραμέτρων*. Εκδόσεις Σταμούλη; 2η έκδοση.

Κεφάλαιο 7

Μέθοδος της μέγιστης πιθανοφάνειας και μέθοδος των ροπών

Στα Κεφάλαια 4, 5 και 6 δόθηκε έμφαση στους αποδοτικούς εκτιμητές και τους ΑΟΕΔ εκτιμητές, η αναζήτηση των οποίων έχει ως αφετηρία το κριτήριο του ΜΤΣ, ενώ βασίζεται σε δύο λογικοφανείς ιδιότητες: την αμεροληψία και την ελάχιστη διασπορά. Με άλλα λόγια, πρώτα «επιβλήθηκαν» στους εκτιμητές οι ιδιότητες αυτές και κατόπιν κατασκευάστηκαν οι εκτιμητές που τις ικανοποιούν. Αντίθετα, οι εκτιμητές που παρουσιάζονται σε αυτό το κεφάλαιο βασίζονται σε δύο απλές στατιστικές αρχές, που τις είδαμε εν συντομία στην Ενότητα 3.3, αυτές της μέγιστης πιθανοφάνειας και της αντικατάστασης και αναφέρονται, αντίστοιχα, ως εκτιμητές μέγιστης πιθανοφάνειας και εκτιμητές ροπών. Η μελέτη των ιδιοτήτων αυτών των εκτιμητών έπεται της κατασκευής τους. Ένα σημαντικό πλεονέκτημά τους είναι ότι η κατασκευή τους δεν απαιτεί προηγούμενη στατιστική γνώση (όπως επάρκεια ή πληρότητα).

7.1 Μέθοδος της μέγιστης πιθανοφάνειας

Η μέθοδος της μέγιστης πιθανοφάνειας αποτελεί μία γενική τεχνική κατασκευής εκτιμητή για μία άγνωστη παράμετρο θ, πραγματική ή διανυσματική και γενικότερα για μία άγνωστη τιμή $g(\theta)$. Σε ειδικές περιπτώσεις, η μέθοδος χρησιμοποιήθηκε από τον Gauss, αλλά και νωρίτερα από τον Laplace (γνωστό και από τον τύπο της κλασικής πιθανότητας). Όμως, ως γενική μέθοδος εκτίμησης προτάθηκε, ονομάστηκε και καθιερώθηκε από τον Fisher σε μια σειρά εργασιών του, Fisher (1912, 1922, 1925, 1934), όπου μελέτησε τις ιδιότητές της. Η μέθοδος της μέγιστης πιθανοφάνειας έχει πλέον ταυτιστεί με το όνομα του Fisher, μπορεί να εφαρμοστεί εύκολα σε πάρα πολλά προβλήματα εκτίμησης, ερμηνεύεται διαισθητικά πολύ απλά και γενικά παράγει καλούς εκτιμητές ειδικά για μεγάλο μέγεθος δείγματος. Δεν είναι υπερβολή να ισχυριστούμε ότι είναι η πλέον γνωστή και η πλέον χρησιμοποιούμενη στις εφαρμογές μέθοδος εκτίμησης. Όπως αναλύθηκε στην Ενότητα 3.3, η μέθοδος στηρίζεται στην απλή και βασική αρχή ότι, εάν $\underset{\sim}{x} = (x_1, \ldots, x_n)$ είναι η παρατηρηθείσα τιμή του $\underset{\sim}{X} = (X_1, \ldots, X_n)$, (δηλαδή $\underset{\sim}{x}$ είναι τα προς ανάλυση δεδομένα σύμφωνα με την Ενότητα 3.1), τότε επιλέγεται ως εκτίμηση του θ η τιμή $\hat{\theta}(\underset{\sim}{x}) \in \Theta$, που μεγιστοποιεί ως προς $\theta \in \Theta$ την πιθανοφάνεια του $\underset{\sim}{x}$ (*αρχή της μέγιστης πιθανοφάνειας*) . Για διακριτό $\underset{\sim}{X}$, η πιθανοφάνεια του $\underset{\sim}{x}$ είναι η πιθανότητα να προκύψει η ήδη παρατηρηθείσα τιμή $\underset{\sim}{x}$, δηλαδή $\mathbb{P}_\theta(\underset{\sim}{X} = \underset{\sim}{x})$, που συμπίπτει με την πυκνότητα του $\underset{\sim}{X}$ στο σημείο $\underset{\sim}{x}$. Γενικά, εάν το δείγμα $\underset{\sim}{X}$ (διακριτό ή συνεχές) έχει πυκνότητα $f(\underset{\sim}{x};\theta)$, $\theta \in \Theta$, η *συνάρτηση πιθανοφάνειας* (ή απλά *πιθανοφάνεια του* $\underset{\sim}{x}$) ορίζεται από τη σχέση

$$L(\theta \mid \underset{\sim}{x}) = f(\underset{\sim}{x};\theta), \ \theta \in \Theta.$$

Είναι, δηλαδή, η συνάρτηση πιθανοφάνειας απλά και μόνον η πυκνότητα του $\underset{\sim}{X}$, $f(\underset{\sim}{x};\theta)$, υπολογιζόμενη στην παρατηρηθείσα τιμή $\underset{\sim}{x}$ του $\underset{\sim}{X}$ και θεωρούμενη ως συνάρτηση του θ (με σταθερό $\underset{\sim}{x}$). Συχνά η συνάρτηση $L(\theta \mid \underset{\sim}{x})$ θα συμβολίζεται απλά με $L(\theta)$.

Η διαισθητική ερμηνεία της μεγιστοποίησης ως προς θ της πιθανοφάνειας $L(\theta \mid \underset{\sim}{x}) = f(\underset{\sim}{x};\theta)$ δόθηκε μέσω ενός παραδείγματος στην Ενότητα 3.3. Μια πιο αυστηρή αιτιολόγηση παρουσιάζεται στην Πρόταση 7.1.1. Θεωρούμε ότι το $\underset{\sim}{X} = (X_1, \ldots, X_n)$ είναι ένα τυχαίο δείγμα από την κατανομή με πυκνότητα $f_1(x;\theta)$, $\theta \in \Theta$ και υποθέτουμε τα εξής.

Σ1. Το κοινό σύνολο τιμών των X_i, $S_1 = \{x : f_1(x;\theta) > 0\}$, δεν εξαρτάται από το θ.

Σ2. Για κάθε $\theta_1 \neq \theta_2$, σημεία του Θ, ισχύει $f_1(x;\theta_1) \neq f_1(x;\theta_2)$, $x \in S_2 \subset S_1$ με $\mathbb{P}(S_2) > 0$, σε κάθε δηλαδή σημείο του Θ αντιστοιχεί διαφορετική κατανομή πιθανότητας. [1]

Επίσης, συμβολίζουμε με θ_0 την άγνωστη, αλλά μοναδική τιμή της παραμέτρου θ, προς διάκριση από τα υπόλοιπα σημεία του παραμετρικού χώρου Θ, και θα την αναφέρουμε εμφατικά ως *αληθή τιμή* του θ. Τα δεδομένα $\underset{\sim}{X}$ έχουν δηλαδή παραχθεί από την κατανομή πιθανότητας, που αντιστοιχεί στο θ_0 (και μόνο σε αυτό) και την οποία συμβολίζουμε $\mathbb{P}_{\theta_0}$.

Πρόταση 7.1.1. *Υπό τις Σ1 και Σ2 ισχύει ότι*

$$\mathbb{P}_{\theta_0}\left(\lim_{n\to\infty}\frac{1}{n}\ln\frac{L(\theta_0 \mid \underset{\sim}{X})}{L(\theta \mid \underset{\sim}{X})} > 0\right) = 1\ , \tag{7.1}$$

για κάθε $\theta \in \Theta$, $\theta \neq \theta_0$.

Απόδειξη. Έχουμε $L(\theta \mid \underset{\sim}{x}) = f(\underset{\sim}{x};\theta) = \prod_{i=1}^n f_1(x_i;\theta)$ για κάθε $\theta \in \Theta$, οπότε η (7.1) ισοδύναμα γράφεται

$$\mathbb{P}_{\theta_0}\left(\lim_{n\to\infty}\frac{1}{n}\ln\prod_{i=1}^{n}\frac{f_1(X_i;\theta_0)}{f_1(X_i;\theta)} > 0\right) = 1$$

ή

$$\mathbb{P}_{\theta_0}\left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}Y_i > 0\right) = 1\ , \tag{7.2}$$

όπου

$$Y_i = \ln\frac{f_1(X_i;\theta_0)}{f_1(X_i;\theta)}\ ,\ \ i = 1,\ldots,n.$$

Θα πρέπει λοιπόν να δείξουμε την (7.2).

[1]Υποσημείωση: Όταν η οικογένεια πυκνοτήτων $\{f_1(x;\theta)\ :\ \theta \in \Theta\}$ ικανοποιεί τη συνθήκη Σ2, η παράμετρος θ λέγεται αναγνωρίσιμη ή ταυτοποιήσιμη (identifiable). Για παράδειγμα, στην οικογένεια κανονικών κατανομών $\mathcal{N}(\theta,1)$, $\theta \in \Re$, το θ είναι αναγνωρίσιμο, ενώ στην οικογένεια κανονικών κατανομών $\mathcal{N}(\cos\theta,1)$, $\theta \in \Re$, το θ δεν είναι αναγνωρίσιμο, γιατί όλα τα $\theta = 2\kappa\pi + \frac{\pi}{3}$, $\kappa = 0,1,\ldots$ παράγουν την ίδια κατανομή $\mathcal{N}(1/2,1)$. Ουσιαστικά, δεν έχει νόημα η εκτίμηση μη αναγνωρίσιμης παραμέτρου.

Η υπόθεση Σ1 εξασφαλίζει ότι $0 < \frac{f_1(X_i;\theta_0)}{f_1(X_i;\theta)} < \infty$ με πιθανότητα 1 ως προς την κατανομή $\mathbb{P}_{\theta_0}$ (αφού ο αριθμητής και ο παρανομαστής είναι θετικοί για X_i στο ίδιο σύνολο, το S_1, και $\mathbb{P}_{\theta_0}(X_i \in S_1) = 1$). Επομένως οι τυχαίες μεταβλητές Y_i, $i = 1, \ldots, n$ είναι καλώς ορισμένες. Επιπλέον, οι Y_i, $i = 1, \ldots, n$ είναι ανεξάρτητες με κοινή κατανομή ως συναρτήσεις των X_i, $i = 1, \ldots, n$, αντίστοιχα. Συνεπώς, από τον Ισχυρό Νόμο των Μεγάλων Αριθμών (Θεώρημα 1.10.2) έχουμε $\mathbb{P}_{\theta_0}\left(\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} Y_i = \mathbb{E}_{\theta_0}Y_1\right) = 1$ και η (7.2) θα ισχύει αν δείξουμε ότι

$$\mathbb{E}_{\theta_0}Y_1 = \mathbb{E}_{\theta_0} \ln \frac{f_1(X_1;\theta_0)}{f_1(X_1;\theta)} > 0 \text{ , για } \theta \neq \theta_0. \qquad (7.3)$$

Η (7.3) είναι ένα κλασικό αποτέλεσμα που περιέχεται στον Wald (1949). (Ενδεχομένως να ήταν και γνωστό νωρίτερα σε κύκλους της Στατιστικής Θεωρίας Πληροφοριών. Μάλιστα, το δεύτερο μέλος της ισότητας στην (7.3) ονομάζεται *Kullback-Leibler απόσταση* μεταξύ των πυκνοτήτων (κατανομών) $f_1(x;\theta_0)$ και $f_1(x;\theta)$, χρησιμεύει ως αριθμητικός δείκτης της διαφορετικότητας (discrimination) τους (βλέπε Kullback and Leibler, 1951), ενώ επίσης έχει εφαρμογές και σε Ελέγχους Στατιστικών Υποθέσεων για $n \to \infty$, βλέπε π.χ. Bahadur (1960).) Η απόδειξή της (7.3) χρησιμοποιεί την ανισότητα Jensen (βλέπε Πρόταση 1.5.2). Από την υπόθεση Σ2, για $\theta \neq \theta_0$ η τυχαία μεταβλητή $\dfrac{f_1(X_i;\theta_0)}{f_1(X_i;\theta)}$ δεν είναι σταθερά (= 1, αν ήταν) και επειδή η συνάρτηση $-\ln x$ είναι γνησίως κυρτή έχουμε

$$\begin{aligned}\mathbb{E}_{\theta_0}Y_1 &= \mathbb{E}_{\theta_0}\left(-\ln \frac{f_1(X_i;\theta)}{f_1(X_i;\theta_0)}\right) > -\ln \mathbb{E}_{\theta_0}\frac{f_1(X_i;\theta)}{f_1(X_i;\theta_0)} = \\ &= -\ln \int_{S_1} \frac{f_1(x;\theta)}{f_1(x;\theta_0)} f_1(x;\theta_0)dx = -\ln \int_{S_1} f_1(x;\theta)dx = -\ln 1 = 0.\end{aligned}$$

(Στην περίπτωση διακριτής κατανομής, το ολοκλήρωμα αντικαθίσταται με σειρά ή άθροισμα.) $\square$

Η (7.1) συνεπάγεται ότι με πιθανότητα 1, $\dfrac{1}{n}\ln\dfrac{L(\theta_0 \mid \underset{\sim}{X})}{L(\theta \mid \underset{\sim}{X})} > 0$ για $n \to \infty$.

Ως εκ τούτου έχουμε

$$\frac{L(\theta_0 \mid \underset{\sim}{X})}{L(\theta \mid \underset{\sim}{X})} > 1 \text{ ή } L(\theta_0 \mid \underset{\sim}{X}) > L(\theta \mid \underset{\sim}{X}) \ , \ \ \forall \theta \neq \theta_0$$

και συνεπώς το σημείο θ_0, η αληθής τιμή της παραμέτρου, μεγιστοποιεί την πιθανοφάνεια του $\underset{\sim}{X} = (X_1, \ldots, X_n)$, $L(\theta \mid \underset{\sim}{X})$, καθώς $n \to \infty$, αν δηλαδή είχαμε «άπειρα» το πλήθος διαθέσιμα δεδομένα. Λόγω αυτής της οριακής – «πληθυσμιακής» ιδιότητας του θ_0 επιλέγουμε ως εκτίμηση του το *δειγματικό του ανάλογο*, συγκεκριμένα, εκείνη την τιμή $\theta(\underset{\sim}{x}) \in \Theta$, που έχει *την ίδια ακριβώς ιδιότητα στο πεπερασμένο δείγμα* $\underset{\sim}{x} = (x_1, \ldots, x_n)$, δηλαδή μεγιστοποιεί ως προς $\theta \in \Theta$ τη συνάρτηση $L(\theta \mid \underset{\sim}{x})$. Η τιμή $\hat{\theta}(\underset{\sim}{x})$ είναι η εκτίμηση μέγιστης πιθανοφάνειας και η στατιστική συνάρτηση $\hat{\theta}(\underset{\sim}{X})$ είναι ο εκτιμητής μέγιστης πιθανοφάνειας του θ για κάθε πεπερασμένο μέγεθος δείγματος n. Σημειώνουμε ότι ο ιδιαίτερος συμβολισμός $\hat{\theta}(\underset{\sim}{X})$, αντί π.χ. $T(\underset{\sim}{X})$ ή $S(\underset{\sim}{X})$ κλπ, έχει καθιερωθεί στη διεθνή βιβλιογραφία. Ο εκτιμητής μέγιστης πιθανοφάνειας του θ ορίζεται λοιπόν ως εξής.

Ορισμός 7.1.1. *Ο εκτιμητής $\hat{\theta}(\underset{\sim}{X})$ ονομάζεται εκτιμητής μέγιστης πιθανοφάνειας (ε.μ.π.) του θ εάν για κάθε τιμή $\underset{\sim}{x}$ του $\underset{\sim}{X}$ ισχύει*

$$L\big(\hat{\theta}(\underset{\sim}{x})\big) = \max_{\theta \in \Theta} L(\theta \mid \underset{\sim}{x}). \tag{7.4}$$

Η συντομογραφία ε.μ.π. θα χρησιμοποιείται αδιακρίτως, λόγω απλότητας, για τον εκτιμητή $\hat{\theta}(\underset{\sim}{X})$, αλλά και την εκτίμηση $\hat{\theta}(\underset{\sim}{x})$, προτάσσοντας το αντίστοιχο άρθρο (ο ή η). Η εύρεση λοιπόν του ε.μ.π. του θ ανάγεται στην εύρεση της τιμής $\theta \in \Theta$, που μεγιστοποιεί ολικά τη συνάρτηση πιθανοφάνειας $L(\theta \mid \underset{\sim}{x})$. Το ολικό μέγιστο μπορεί να επιτυγχάνεται σε μία τιμή $\hat{\theta}(\underset{\sim}{x})$ (η πλέον τυπική περίπτωση) ή σε περισσότερες από μία τιμές ή να μην υπάρχει. Ανάλογα, ο ε.μ.π. είναι μοναδικός ή υπάρχουν πολλοί ε.μ.π. ή δεν υπάρχει ε.μ.π.. Εάν η συνάρτηση $L(\theta \mid \underset{\sim}{x})$ παραγωγίζεται ως προς θ, το μέγιστο υπάρχει και επιτυγχάνεται σε εσωτερικό σημείο του Θ τότε μπορεί να βρεθεί με παραγώγιση. Σε αυτές τις περιπτώσεις λόγω της μορφής της $L(\theta \mid \underset{\sim}{x})$, είναι συχνά πιο εύκολο να μεγιστοποιήσουμε τον (νεπέριο) λογάριθμο $\ln L(\theta \mid \underset{\sim}{x})$. Προφανώς, κάθε τιμή του θ

που μεγιστοποιεί τη συνάρτηση $\ln L(\theta \mid \underset{\sim}{x})$ επίσης μεγιστοποιεί και τη συνάρτηση $L(\theta \mid \underset{\sim}{x})$, γιατί ο λογάριθμος είναι γνησίως αύξουσα συνάρτηση. Στην περίπτωση τυχαίου δείγματος $\underset{\sim}{X} = (X_1, \ldots, X_n)$ από κατανομή με πυκνότητα $f_1(x;\theta)$, η συνάρτηση πιθανοφάνειας έχει τη μορφή

$$L(\theta) = L(\theta \mid \underset{\sim}{x}) = \prod_{i=1}^{n} f_1(x_i;\theta), \quad \underset{\sim}{x} = (x_1, \ldots, x_n),$$

οπότε

$$\ln L(\theta) = \sum_{i=1}^{n} \ln f_1(x_i;\theta).$$

Αν το θ είναι πραγματικός αριθμός και το μέγιστο μπορεί να βρεθεί με παραγώγιση, ο ε.μ.π. είναι λύση ως προς θ της εξίσωσης

$$0 = \frac{\partial}{\partial\theta} \ln L(\theta) = \sum_{i=1}^{n} \frac{\frac{\partial}{\partial\theta} f_1(x_i;\theta)}{f_1(x_i;\theta)}\ , \tag{7.5}$$

η οποία αναφέρεται ως *εξίσωση πιθανοφάνειας* . Στην περίπτωση που το θ είναι διανυσματική παράμετρος $\theta = (\theta_1, \ldots, \theta_r)$ και το μέγιστο μπορεί να βρεθεί με παραγώγιση, ο ε.μ.π. είναι λύση ως προς $\theta_1, \ldots, \theta_r$ του συστήματος των εξισώσεων πιθανοφάνειας

$$0 = \sum_{i=1}^{n} \frac{\frac{\partial}{\partial\theta_j} f_1(x_i;\theta)}{f_1(x_i;\theta)}, \quad j = 1, \ldots, r. \tag{7.6}$$

Λύνοντας την (7.5) ή τις (7.6), πρέπει περαιτέρω να διαπιστώνεται ότι η λύση αντιστοιχεί σε *ολικό μέγιστο.* Ανάλογα με τη μορφή της $L(\theta \mid \underset{\sim}{x})$, οι εξισώσεις (7.5) και (7.6) είναι δυνατόν να επιδέχονται λύση σε κλειστή μορφή ως συναρτήσεις του $\underset{\sim}{x}$ ή μόνον αριθμητική επίλυση.

Δίνουμε στη συνέχεια μερικά παραδείγματα υπολογισμού εκτιμητών μέγιστης πιθανοφάνειας.

Παράδειγμα 7.1.1. (**Διακριτός παραμετρικός χώρος**) Δίνεται μία παρατήρηση X από την κατανομή που φαίνεται στον πίνακα. Το σύνολο τιμών του X είναι $\{1, 2, 3\}$, τα στοιχεία του πίνακα είναι οι αντίστοιχες πιθανότητες αυτών των τιμών και $\theta \in \Theta = \{0, 1\}$.

	$X=1$	$X=2$	$X=3$
$\theta=0$	0.1	0.7	0.2
$\theta=1$	0.9	0.05	0.05

Εάν $X = 1$, η πιθανοφάνεια $L(\theta)$ έχει τιμές $L(0) = 0.1$, $L(1) = 0.9$ και μεγιστοποιείται για $\theta = 1$. Ανάλογα, εάν $X = 2$, η πιθανοφάνεια μεγιστοποιείται για $\theta = 0$ και εάν $X = 3$ η πιθανοφάνεια μεγιστοποιείται για $\theta = 0$. Επομένως, ο ε.μ.π. του θ είναι $\hat{\theta}(X)$ με $\hat{\theta}(1) = 1$, $\hat{\theta}(2) = 0$, $\hat{\theta}(3) = 0$.

Παράδειγμα 7.1.2. (**Κανονική κατανομή - ε.μ.π.**) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$, $n \geq 2$, ένα τυχαίο δείγμα από την κανονική κατανομή $\mathcal{N}(\mu, \sigma^2)$.
1η Περίπτωση: $\boldsymbol{\sigma^2}$ **γνωστό,** $\boldsymbol{\mu = \theta}$ **άγνωστο**, $\theta \in \Theta = \mathbb{R}$.
Τότε

$$f_1(x;\theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\theta)^2}$$

και επομένως έχουμε

$$L(\theta) = f(\underset{\sim}{x};\theta) = \prod_{i=1}^{n} f_1(x_i;\theta) = \frac{1}{\sigma^n(2\pi)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2},$$

$$\ln L(\theta) = -\frac{n}{2}\ln\sigma^2 - \frac{n}{2}\ln(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2,$$

$$\frac{\partial}{\partial\theta}\ln L(\theta) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\theta).$$

Από την εξίσωση πιθανοφάνειας $\frac{\partial}{\partial\theta}\ln L(\theta) = 0$, προκύπτει $\sum_{i=1}^{n}(x_i-\theta) = 0$, δηλαδή $\theta = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$. Η λύση $\bar{x}$ ανήκει στο Θ, είναι μοναδική και αντιστοιχεί σε ολικό μέγιστο, επειδή $\frac{\partial^2}{\partial\theta^2}\ln L(\theta) = -\frac{n}{\sigma^2} < 0$ για κάθε $\theta \in \Theta$. Άρα ο ε.μ.π. του θ είναι $\hat{\theta}(\underset{\sim}{X}) = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$. Παρατηρούμε ότι ο ε.μ.π. συμπίπτει με τον αποδοτικό εκτιμητή του θ (βλέπε Παράδειγμα 5.2.11).
2η Περίπτωση: $\boldsymbol{\mu}$ **γνωστό,** $\boldsymbol{\sigma^2 = \theta}$ **άγνωστο**, $\theta \in \Theta = (0, \infty)$.

Τότε

$$f_1(x;\theta) = \frac{1}{\theta^{1/2}\sqrt{2\pi}} e^{-\frac{1}{2\theta}(x-\mu)^2}$$

και επομένως έχουμε

$$\begin{aligned} L(\theta) = f(\underset{\sim}{x};\theta) = \prod_{i=1}^{n} f_1(x_i;\theta) &= \frac{1}{\theta^{n/2}(2\pi)^{n/2}} e^{-\frac{1}{2\theta}\sum_{i=1}^{n}(x_i-\mu)^2}, \\ \ln L(\theta) &= -\frac{n}{2}\ln\theta - \frac{n}{2}\ln(2\pi) - \frac{1}{2\theta}\sum_{i=1}^{n}(x_i-\mu)^2, \\ \frac{\partial}{\partial\theta}\ln L(\theta) &= -\frac{n}{2\theta} + \frac{1}{2\theta^2}\sum_{i=1}^{n}(x_i-\mu)^2. \end{aligned}$$

Χωρίς βλάβη της γενικότητας θεωρούμε $\sum_{i=1}^{n}(x_i-\mu)^2 > 0$ (λόγω της συνεχούς κατανομής των X_i). Τότε, λύνοντας την εξίσωση πιθανοφάνειας $\frac{\partial}{\partial\theta}\ln L(\theta) = 0$, προκύπτει $-n + \frac{1}{\theta}\sum_{i=1}^{n}(x_i-\mu)^2 = 0$, δηλαδή $\theta = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 = \hat{\theta}$ (έστω). Η λύση είναι μοναδική, αντιστοιχεί σε μέγιστο, επειδή $\frac{\partial^2}{\partial\theta^2}\ln L(\theta)\mid_{\theta=\hat{\theta}} = \frac{n}{\hat{\theta}^2} - \frac{1}{\hat{\theta}^3}\sum_{i=1}^{n}(x_i-\mu)^2 = -\frac{n}{2\hat{\theta}^2} < 0$, το οποίο είναι ολικό μέγιστο, αφού $\lim\limits_{\theta\to 0^+}\ln L(\theta) = \lim\limits_{\theta\to\infty}\ln L(\theta) = -\infty$. Άρα ο ε.μ.π. του θ είναι $\hat{\theta}(\underset{\sim}{X}) = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$. Παρατηρούμε ότι ο ε.μ.π. συμπίπτει με τον αποδοτικό εκτιμητή του θ (βλέπε Παράδειγμα 5.2.12). Αν θεωρήσουμε το σ ως άγνωστη παράμετρο αντί του σ^2 και ακολουθήσουμε την παραπάνω διαδικασία (οπότε τώρα θα παραγωγίσουμε ως προς σ), θα βρούμε ως ε.μ.π. τη στατιστική συνάρτηση $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2} = \sqrt{\hat{\theta}(\underset{\sim}{X})}$. Στην Ενότητα 7.2, όπου μελετάμε γενικές ιδιότητες των ε.μ.π. θα διαπιστώσουμε ότι, έχοντας ήδη βρει τον ε.μ.π. του σ^2, $\hat{\theta}(\underset{\sim}{X})$, μπορούμε να βρούμε τον ε.μ.π. του σ παίρνοντας κατ' ευθείαν την τετραγωνική ρίζα $\sqrt{\hat{\theta}(\underset{\sim}{X})}$.

<u>3η Περίπτωση:</u> $\boldsymbol{\mu, \sigma^2}$ **άγνωστα**, οπότε $\theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times (0,\infty)$. Τότε

$$f_1(x;\theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

και επομένως έχουμε

$$L(\theta) = f(\underset{\sim}{x};\theta) = \prod_{i=1}^{n} f_1(x_i;\theta) = \frac{1}{\sigma^n (2\pi)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2},$$

$$\ln L(\theta) = -\frac{n}{2}\ln\sigma^2 - \frac{n}{2}\ln(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2,$$

$$\frac{\partial}{\partial\sigma^2}\ln L(\theta) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2,$$

$$\frac{\partial}{\partial\mu}\ln L(\theta) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu).$$

Ανάλογα με την προηγούμενη περίπτωση, θεωρούμε $\sum_{i=1}^{n}(x_i-\mu)^2 > 0$ και $\sum_{i=1}^{n}(x_i-\bar{x})^2 > 0$. Τότε, λύνοντας το σύστημα των εξισώσεων πιθανοφάνειας $\frac{\partial}{\partial\sigma^2}\ln L(\theta) = 0$ και $\frac{\partial}{\partial\mu}\ln L(\theta) = 0$ έχουμε $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$ και $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$. Η λύση είναι μοναδική και αντιστοιχεί σε ολικό μέγιστο (δικαιολογείται παρακάτω). Άρα οι ε.μ.π. των μ, σ^2 είναι αντίστοιχα $\hat{\mu} = \bar{X}$, $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$. Παρατηρούμε ότι ο ε.μ.π. του σ^2 διαφέρει από τον ΑΟΕΔ εκτιμητή που είναι $\frac{1}{n-1}\sum_i(X_i-\bar{X})^2$ (βλέπε Παράδειγμα 6.3.10). Θα αποδείξουμε τώρα ότι η τιμή $\hat{\theta} = (\hat{\mu}, \hat{\sigma^2})$ όντως μεγιστοποιεί ολικά τη συνάρτηση $\ln L(\theta)$. Κατ' αρχάς, ας συμβολίσουμε τη συνάρτηση πιθανοφάνειας $L(\mu,\sigma^2)$ αντί $L(\theta)$, αφού εξ άλλου $\theta = (\mu,\sigma^2)$. Τότε για κάθε $\sigma^2 > 0$, έχουμε

$$\ln L(\mu,\sigma^2) \leqslant \ln L(\bar{x},\sigma^2)\ ,\quad \forall\mu\in\Re, \tag{7.7}$$

επειδή, όπως είδαμε στην 1η περίπτωση, $\bar{x}$ είναι η ε.μ.π. του μ για κάθε δεδομένο (γνωστό), αλλά οποιοδήποτε $\sigma^2 > 0$. Περαιτέρω και κατ' αναλογία με την 2η περίπτωση, η συνάρτηση $\ln L(\bar{x},\sigma^2) = -\frac{n}{2}\ln\sigma^2 - \frac{n}{2}\ln(2\pi) - \frac{1}{2\sigma^2}\sum\limits_{i=1}^{n}(x_i-\bar{x})^2$, μεγιστοποιείται μοναδικά για $\sigma^2 = \hat{\sigma}^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(x_i-\bar{x})^2$. Επομένως, έχουμε

$$\ln L(\bar{x},\sigma^2) \leqslant \ln L(\bar{x},\hat{\sigma}^2)\ ,\quad \forall\sigma^2 > 0. \tag{7.8}$$

Τελικά, συνδυάζοντας τις (7.7) και (7.8), προκύπτει ότι

$$\ln L(\mu,\sigma^2) \leqslant \ln L(\bar{x},\hat{\sigma}^2)\ ,\quad \forall(\mu,\sigma^2)\in \Re\times(0,\infty)\ ,$$

δηλαδή το σημείο $(\bar{x},\hat{\sigma}^2)$ μεγιστοποιεί μοναδικά τη συνάρτηση $\ln L(\mu,\sigma^2)$. Σημειώνουμε, ότι αυτός ο τρόπος απόδειξης, που αποφεύγει τη χρησιμοποίηση του πίνακα των δεύτερων παραγώγων, είναι γενικός, δεν εξαρτάται δηλαδή από τη συγκεκριμένη συνάρτηση πιθανοφάνειας και θα τον χρησιμοποιήσουμε και στο Παράδειγμα 7.1.8.

Παράδειγμα 7.1.3. (Κατανομή Poisson - ε.μ.π.) Έστω $\underset{\sim}{X}=(X_1,\ldots,X_n)$, ένα τυχαίο δείγμα από την κατανομή Poisson, $\mathcal{P}(\theta)$, $\theta\in\Theta=(0,\infty)$. Έχουμε

$$f_1(x;\theta)=e^{-\theta}\frac{\theta^x}{x!},\ x=0,1,2,\ldots$$

και επομένως

$$L(\theta)=f(\underset{\sim}{x};\theta)=\prod_{i=1}^n f_1(x_i;\theta)=e^{-n\theta}\frac{\theta^{\sum_{i=1}^n x_i}}{x_1!\cdots x_n!}.$$

Παρατηρούμε, ότι εάν $\sum_{i=1}^n x_i=0\Leftrightarrow x_i=0,\ i=1,\ldots,n$, τότε $L(\theta\mid\underset{\sim}{0})=e^{-n\theta}$ και $\sup_{\theta\in\Theta}L(\theta\mid\underset{\sim}{0})=1$, το οποίο όμως δεν επιτυγχάνεται για καμία τιμή του $\theta\in(0,\infty)$, αφού $0<e^{-n\theta}<1$, για κάθε $\theta>0$. Άρα, εάν η παρατηρηθείσα τιμή του $\underset{\sim}{X}=(X_1,\ldots,X_n)$ είναι $\underset{\sim}{0}=(0,\ldots,0)$, τότε δεν υπάρχει $\max_{\theta\in\Theta}L(\theta)$, δηλαδή για $\underset{\sim}{X}=\underset{\sim}{0}$ δεν υπάρχει ε.μ.π. του θ. Εάν όμως $\sum_{i=1}^n x_i>0$, δηλαδή η παρατηρηθείσα τιμή του $\underset{\sim}{X}$ είναι διάφορη του $\underset{\sim}{0}$, τότε

$$\ln L(\theta)=-n\theta+\big(\sum_{i=1}^n x_i\big)\ln\theta-\ln(x_1!\cdots x_n!)\,,$$

$$\frac{\partial}{\partial\theta}\ln L(\theta)=-n+\frac{1}{\theta}\sum_{i=1}^n x_i\,.$$

Λύνοντας την εξίσωση πιθανοφάνειας $\frac{\partial}{\partial\theta}\ln L(\theta)=0$ ως προς θ, έχουμε $\theta=\frac{1}{n}\sum_{i=1}^n x_i=\bar{x}$. Η λύση είναι μοναδική, αντιστοιχεί σε ολικό μέγιστο, επειδή $\frac{\partial^2}{\partial\theta^2}\ln L(\theta) = -\frac{1}{\theta^2}\sum_{i=1}^n x_i < 0$ για κάθε $\theta \in \Theta$ και ανήκει στο $\Theta = (0, +\infty)$, αφού $\sum_{i=1}^n x_i > 0$. Τελικά έχουμε ότι ο ε.μ.π. του θ είναι

$$\hat{\theta}(\underset{\sim}{X}) = \begin{cases} \text{δεν υπάρχει} & , \text{ εάν } \underset{\sim}{X} = \underset{\sim}{0} \\ \bar{X} & , \text{ εάν } \underset{\sim}{X} \neq \underset{\sim}{0} \end{cases}.$$

Η πιθανότητα μη ύπαρξης του ε.μ.π. του θ είναι

$$\begin{aligned} \mathbb{P}_\theta(\text{δεν υπάρχει ο ε.μ.π. του } \theta) &= \mathbb{P}_\theta(\underset{\sim}{X} = \underset{\sim}{0}) = \mathbb{P}_\theta(X_1 = 0, \ldots, X_n = 0) \\ &= \prod_{i=1}^n \mathbb{P}_\theta(X_i = 0) = \prod_{i=1}^n e^{-\theta} = e^{-n\theta} > 0, \end{aligned}$$

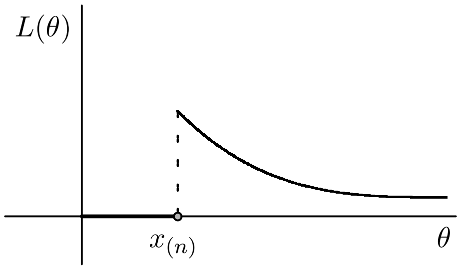
με $\lim_{n\to\infty} \mathbb{P}_\theta(\text{δεν υπάρχει ο ε.μ.π. του } \theta) = 0$. Στο παράδειγμα λοιπόν αυτό υπάρχει μεν θετική πιθανότητα να μην υπάρχει ο ε.μ.π. της παραμέτρου θ, αλλά το «παρήγορο» είναι ότι η πιθανότητα αυτή συγκλίνει στο μηδέν με εκθετική ταχύτητα, καθώς το μέγεθος του δείγματος $n \to \infty$. Με άλλα λόγια, ο ε.μ.π. του θ υπάρχει με πιθανότητα που τείνει στο 1, όταν $n \to \infty$.

Παράδειγμα 7.1.4. (Ομοιόμορφη κατανομή με ένα άκρο γνωστό - ε.μ.π.) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την ομοιόμορφη κατανομή $\mathcal{U}[0, \theta]$, $\theta \in \Theta = (0, \infty)$.
Έχουμε

$$f_1(x; \theta) = \begin{cases} \frac{1}{\theta} & , 0 \leqslant x \leqslant \theta \\ 0 & , \text{ διαφορετικά} \end{cases},$$

οπότε

$$\begin{aligned} L(\theta) = f(\underset{\sim}{x}; \theta) &= \prod_{i=1}^n f_1(x_i; \theta) \\ &= \begin{cases} \frac{1}{\theta^n} & , \ 0 \leq x_i \leq \theta, i = 1, \ldots, n \\ 0 & , \text{ διαφορετικά} \end{cases} \\ &= \begin{cases} \frac{1}{\theta^n} & , \ \theta \geq x_{(n)} = \max(x_1, \ldots, x_n) \\ 0 & , \text{ διαφορετικά.} \end{cases} \end{aligned}$$



Σχήμα 7.1: Συνάρτηση πιθανοφάνειας για την ομοιόμορφη κατανομή

Το μέγιστο της $L(\theta)$ θα αναζητηθεί, προφανώς, στον θετικό κλάδο, δηλαδή είναι το μέγιστο της συνάρτησης $\frac{1}{\theta^n}$ για $\theta \geq x_{(n)}$, που είναι γνησίως φθίνουσα ως προς θ. Συνεπώς, το μέγιστο επιτυγχάνεται, ό-ταν $\theta = x_{(n)}$ (βλέπε Σχήμα 7.1). Άρα ο ε.μ.π. του θ είναι $\hat{\theta}(\underset{\sim}{X}) = X_{(n)} = \max(X_1, \ldots, X_n)$. Σημειώνουμε ότι ο ΑΟΕΔ εκτιμητής του θ είναι $\frac{n+1}{n}X_{(n)}$ (βλέπε Παράδειγμα 6.3.3). Επίσης παρατηρούμε ότι $\mathbb{P}_\theta(\hat{\theta}(\underset{\sim}{X}) \leqslant \theta) = \mathbb{P}_\theta(X_{(n)} \leqslant \theta) = \mathbb{P}_\theta(X_1 \leqslant \theta, X_2 \leqslant \theta, \ldots, X_n \leqslant \theta) = \mathbb{P}_\theta(X_1 \leqslant \theta)\mathbb{P}_\theta(X_2 \leqslant \theta) \ldots \mathbb{P}_\theta(X_n \leqslant \theta) = [\mathbb{P}_\theta(X_1 \leqslant \theta)]^n = 1$, δηλαδή με πιθανότη-τα 1, ισχύει $\hat{\theta}(\underset{\sim}{X}) \leqslant \theta$, το οποίο σημαίνει ότι στην προκειμένη περίπτωση ο ε.μ.π. *υποεκτιμά* την άγνωστη παράμετρο θ. Η χρησιμοποίηση, εδώ, του ε.μ.π. γνωρίζοντας ότι το θ είναι με πιθανότητα 1 μεγαλύτερο από την εκτίμηση του δεν πρέπει γενικά να περάσει απαρατήρητη, αντίθετα πρέπει να προβληματίσει τον χρήστη, ανάλογα με το υπό μελέτη φυσικό πρόβλημα. Περαιτέρω, είναι εύκολο να δειχθεί, χρησιμοποιώντας τις σχέ-σεις $\mathbb{E}_\theta X_{(n)} = \frac{n}{n+1}\theta$ και $\mathbb{V}ar_\theta X_{(n)} = \dfrac{n}{(n+2)(n+1)^2}\theta^2$, (η απόδειξη των οποίων βασίζεται στην κατανομή της $X_{(n)}$, βλέπε Παράδειγμα 6.3.3), ότι

$$\text{ΜΤΣ}(\frac{n+1}{n}X_{(n)}; \theta) = \mathbb{V}ar_\theta(\frac{n+1}{n}X_{(n)}) = \frac{1}{n(n+2)}\theta^2,$$

ενώ

$$\text{ΜΤΣ}(\hat{\theta}(\underset{\sim}{X}); \theta) = \mathbb{V}ar_\theta(X_{(n)}) + (\mathbb{E}_\theta X_{(n)} - \theta)^2 = \frac{2}{(n+1)(n+2)}\theta^2.$$

Επομένως ο ΑΟΕΔ εκτιμητής είναι καλύτερος από τον ε.μ.π. με κριτήριο το ΜΤΣ. Εάν τώρα θεωρήσουμε την ομοιόμορφη κατανομή $\mathcal{U}(0, \theta)$, δηλαδή στο ανοικτό διάστημα $(0, \theta)$, τότε έχουμε

$$L(\theta) = \begin{cases} \frac{1}{\theta^n} & , \ \theta > x_{(n)} \\ 0 & , \ \text{διαφορετικά.} \end{cases}$$

Σε αυτήν την περίπτωση $\sup\limits_{\theta > x_{(n)}} L(\theta) = \dfrac{1}{x_{(n)}^n}$, το οποίο όμως δεν επιτυγχάνεται για καμία τιμή του $\theta \in (x_{(n)}, \infty)$, άρα δεν υπάρχει ο ε.μ.π. του θ. Βλέπουμε, εδώ, ότι μία μη ουσιαστική αλλαγή στον ορισμό της (συνεχούς) πυκνότητας των δεδομένων, επέφερε δραστική μεταβολή στην εκτίμηση του θ με τη μέθοδο της μέγιστης πιθανοφάνειας.

Παράδειγμα 7.1.5. (**Ομοιόμορφη κατανομή με άγνωστα άκρα - ε.μ.π.**)
Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την ομοιόμορφη κατανομή $\mathcal{U}[\theta, \theta + 1]$, $\theta \in \Theta = \mathbb{R}$.
Έχουμε

$$f_1(x; \theta) = \begin{cases} 1 & , \ \theta \leq x \leq \theta + 1 \\ 0 & , \ \text{διαφορετικά.} \end{cases}$$

Επομένως,

$$\begin{aligned} L(\theta) = f(\underset{\sim}{x}; \theta) &= \prod_{i=1}^{n} f_1(x_i; \theta) \\ &= \begin{cases} 1 & , \ \theta \leq x_i \leq \theta + 1 \ , \ \ i = 1, \ldots, n \\ 0 & , \ \text{διαφορετικά} \end{cases} \\ &= \begin{cases} 1 & , \ x_{(n)} - 1 \leq \theta \leq x_{(1)} \\ 0 & , \ \text{διαφορετικά} \end{cases}, \end{aligned}$$

όπου $x_{(1)} = \min(x_1, \ldots, x_n)$ και $x_{(n)} = \max(x_1, \ldots, x_n)$. Προφανώς, $\max_{\theta \in \Theta} L(\theta) = 1$, το οποίο επιτυγχάνεται για οποιαδήποτε τιμή του θ στο διάστημα $[x_{(n)} - 1, x_{(1)}]$. Συνεπώς, κάθε εκτιμητής $\hat{\theta}(\underset{\sim}{X})$ που ικανοποιεί τη σχέση $X_{(n)} - 1 \leq \hat{\theta}(\underset{\sim}{X}) \leq X_{(1)}$ είναι ε.μ.π. του θ. Συνεπώς υπάρχουν άπειροι εκτιμητές μέγιστης πιθανοφάνειας του θ. Μερικοί από αυτούς είναι $X_{(1)}$, $X_{(n)}-1$, $c(X_{(n)}-1)+(1-c)X_{(1)}$, όπου c σταθερά, $0 \le c \le 1$, $c(\underset{\sim}{X})(X_{(n)}-1)+\big(1-c(\underset{\sim}{X})\big)X_{(1)}$, όπου $c(\underset{\sim}{X})$ είναι στατιστική συνάρτηση με $0 \le c(\underset{\sim}{X}) \le 1$ (π.χ. $c(\underset{\sim}{X}) = \cos^2 X_1$).

Παράδειγμα 7.1.6. (Κατανομή Γάμμα - ε.μ.π.) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$, $n \ge 2$, ένα τυχαίο δείγμα από την κατανομή Γάμμα $\mathcal{G}(\alpha, \beta)$ με α, β άγνωστα, οπότε $\theta = (\alpha, \beta) \in (0, \infty) \times (0, \infty)$.
Έχουμε

$$f_1(x;\theta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}},\ x > 0.$$

Επομένως,

$$\begin{aligned} L(\theta) = f(\underset{\sim}{x};\theta) &= \prod_{i=1}^{n} f_1(x_i;\theta) \\ &= \frac{1}{\big[\Gamma(\alpha)\big]^n \beta^{n\alpha}} \big(\prod_{i=1}^{n} x_i\big)^{\alpha-1} e^{-\frac{1}{\beta}\sum_{i=1}^{n} x_i}, \end{aligned}$$

$$\ln L(\theta) = -n \ln \Gamma(\alpha) - n\alpha \ln \beta + (\alpha - 1)\sum_{i=1}^{n} \ln x_i - \frac{1}{\beta}\sum_{i=1}^{n} x_i,$$

$$\begin{aligned} \frac{\partial}{\partial \alpha} \ln L(\theta) &= -n\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} - n \ln \beta + \sum_{i=1}^{n} \ln x_i, \\ \frac{\partial}{\partial \beta} \ln L(\theta) &= -n\alpha\frac{1}{\beta} + \frac{1}{\beta^2}\sum_{i=1}^{n} x_i. \end{aligned}$$

Θέτοντας $\frac{\partial}{\partial \alpha} \ln L(\theta) = 0$ και $\frac{\partial}{\partial \beta} \ln L(\theta) = 0$, έχουμε

$$\begin{gathered} -n\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} - n \ln \beta + \sum_{i=1}^{n} \ln x_i = 0, \\ \alpha\beta = \bar{x}. \end{gathered}$$

Είναι δυνατόν να δειχθεί ότι το σύστημα των δύο αυτών εξισώσεων πιθανοφάνειας έχει ακριβώς μία λύση και η λύση αυτή αντιστοιχεί σε ολικό μέγιστο της συνάρτησης $L(\theta)$ (βλέπε Ηλιόπουλος, 2013, σελ. 151). Επομένως υπάρχουν οι ε.μ.π. των α, β και είναι μοναδικοί. Για τη λύση του συστήματος παρατηρούμε ότι

$$\beta = \frac{\bar{x}}{\alpha},$$

οπότε

$$-n\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} + n\ln\alpha - n\ln\bar{x} + \sum_{i=1}^{n}\ln x_i = 0.$$

Η λύση της τελευταίας εξίσωσης δεν υπάρχει σε αναλυτική μορφή, μπορεί όμως η εξίσωση να λυθεί με μεθόδους αριθμητικής ανάλυσης και έτσι να υπολογιστεί (κατά προσέγγιση) η τιμή $\hat{\alpha}(x_1, \ldots, x_n)$, που ικανοποιεί την εξίσωση, δοθέντων των $x_1, \ldots, x_n$. Αυτή η τιμή είναι η ε.μ.π. του α. Κατόπιν, η ε.μ.π. του β για $\underset{\sim}{X} = \underset{\sim}{x}$ είναι $\hat{\beta}(x_1, \ldots, x_n) = \frac{\bar{x}}{\hat{\alpha}(x_1,\ldots,x_n)}$.

Παράδειγμα 7.1.7. (Ομοιόμορφη κατανομή με δύο άγνωστες παραμέτρους - ε.μ.π.) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$, $n \geq 2$, ένα τυχαίο δείγμα από την ομοιόμορφη κατανομή $\mathcal{U}[\theta_1, \theta_2]$ με θ_1, θ_2 άγνωστα, οπότε $\theta = (\theta_1, \theta_2) \in \Theta = \big\{(\theta_1, \theta_2) \colon \theta_i \in \mathbb{R}, \theta_1 < \theta_2\big\}$.
Έχουμε

$$f_1(x;\theta) = \begin{cases} \frac{1}{\theta_2 - \theta_1} & , \ \theta_1 \leq x \leq \theta_2 \\ 0 & , \ \text{διαφορετικά} \end{cases}.$$

Επομένως,

$$\begin{aligned} L(\theta) = f(\underset{\sim}{x};\theta) &= \prod_{i=1}^{n} f_1(x_i;\theta) \\ &= \begin{cases} \frac{1}{(\theta_2-\theta_1)^n} & , \ \theta_1 \leq x_i \leq \theta_2, \ i = 1, \ldots, n \\ 0 & , \ \text{διαφορετικά} \end{cases} \\ &= \begin{cases} \frac{1}{(\theta_2-\theta_1)^n} & , \ \theta_1 \leq x_{(1)} \leq x_{(n)} \leq \theta_2 \\ 0 & , \ \text{διαφορετικά} \end{cases}, \end{aligned}$$

όπου $x_{(1)} = \min(x_1, \dots, x_n)$ και $x_{(n)} = \max(x_1, \dots, x_n)$. Παρατηρούμε ότι για να μεγιστοποιήσουμε τη συνάρτηση πιθανοφάνειας $L(\theta)$ πρέπει να μεγιστοποιήσουμε τη συνάρτηση $\frac{1}{(\theta_2-\theta_1)^n}$ για $\theta_1 \leq x_{(1)}$, $\theta_2 \geq x_{(n)}$. Επειδή η συνάρτηση $\frac{1}{(\theta_2-\theta_1)^n}$ είναι γνησίως φθίνουσα ως προς θ_2 και γνησίως αύξουσα ως προς θ_1, το μέγιστο της επιτυγχάνεται για $\theta_2 = x_{(n)}$, $\theta_1 = x_{(1)}$. Άρα οι ε.μ.π. των θ_1, θ_2 είναι $\hat{\theta}_1(\underset{\sim}{X}) = X_{(1)} = \min(X_1, \dots, X_n)$ και $\hat{\theta}_2(\underset{\sim}{X}) = X_{(n)} = \max(X_1, \dots, X_n)$. Σημειώνουμε ότι ο $\hat{\theta}_1(\underset{\sim}{X})$ υπερεκτιμά το θ_1, ενώ ο $\hat{\theta}_2(\underset{\sim}{X})$ υποεκτιμά το θ_2 (όπως στο Παράδειγμα 7.1.4).

Παράδειγμα 7.1.8. (**Διπαραμετρική εκθετική κατανομή - ε.μ.π.**) Έστω $\underset{\sim}{X} = (X_1, \dots, X_n)$, $n \geq 2$, ένα τυχαίο δείγμα από τη διπαραμετρική εκθετική κατανομή με πυκνότητα $\dfrac{1}{\sigma}e^{-\frac{x-\mu}{\sigma}}$ για $x \geqslant \mu$ και 0 για $x < \mu$, όπου $\mu \in \Re$ και $\sigma > 0$ είναι σταθερές. Η κατανομή αυτή αναφέρεται και ως μετατοπισμένη εκθετική, επειδή, θέτοντας $Y_i = X_i - \mu$, η Y_i έχει εκθετική κατανομή $\mathcal{E}(\sigma)$, οπότε $X_i = Y_i + \mu$ είναι η «μετατόπιση» της Y_i κατά μ. Από αυτήν την παράσταση της X_i, προκύπτει αμέσως ότι $\mathbb{E}X_i = \mathbb{E}Y_i + \mu = \sigma + \mu$ και $\mathbb{V}arX_i = \mathbb{V}ar(Y_i + \mu) = \mathbb{V}arY_i = \sigma^2$. Στην πράξη η κατανομή βρίσκει εφαρμογές ως μοντέλο χρόνου ζωής συστήματος, (όπως εξ' άλλου και η εκθετική κατανομή). Σε αυτήν την περίπτωση, επειδή $\mathbb{P}_\theta(X_i \geqslant \mu) = 1$, η σταθερά μ παριστάνει τον ελάχιστο χρόνο ζωής του συστήματος, ενώ σε κάθε περίπτωση η σταθερά σ είναι η τυπική απόκλιση της κατανομής.

1η Περίπτωση: ***σ* γνωστό, *$\mu = \theta$* άγνωστο**, $\theta \in \Theta = \mathbb{R}$.
Τότε

$$f_1(x;\theta) = \begin{cases} \frac{1}{\sigma}e^{-\frac{x-\theta}{\sigma}} & , \ x \geqslant \theta \\ 0 & , \ \text{διαφορετικά} \end{cases},$$

οπότε έχουμε

$$\begin{aligned} L(\theta) = f(\underset{\sim}{x};\theta) &= \prod_{i=1}^{n} f_1(x_i;\theta) \\ &= \begin{cases} \frac{1}{\sigma^n}e^{-\frac{1}{\sigma}\sum\limits_{i=1}^{n}(x_i-\theta)} & , \ x_i \geqslant \theta \ , \ \ i = 1, \dots, n \\ 0 & , \ \text{διαφορετικά} \end{cases}, \end{aligned}$$

$$= \begin{cases} \frac{1}{\sigma^n} e^{-\frac{1}{\sigma}\sum\limits_{i=1}^{n} x_i + \frac{n\theta}{\sigma}} & , \ x_{(1)} \geqslant \theta \\ 0 & , \ \text{διαφορετικά} \end{cases},$$

όπου $x_{(1)} = \min(x_1, \ldots, x_n)$. Προφανώς, το μέγιστο της $L(\theta)$ θα αναζητηθεί στο θετικό κλάδο, $\frac{1}{\sigma^n} e^{-\frac{1}{\sigma}\sum\limits_{i=1}^{n} x_i + \frac{n\theta}{\sigma}}$ για $\theta \in (-\infty, x_{(1)}]$. Η συνάρτηση αυτή είναι γνησίως αύξουσα ως προς θ, (επειδή $\sigma > 0$) και επομένως το μέγιστό της επιτυγχάνεται για $\theta = x_{(1)}$. Συνεπώς, ο ε.μ.π. του θ είναι $\hat{\theta}(\underset{\sim}{X}) = X_{(1)}$. Σημειώνουμε ότι $\theta \leqslant X_{(1)}$ (με πιθανότητα 1) δηλαδή στην προκειμένη περίπτωση ο ε.μ.π. *υπερεκτιμά* την άγνωστη παράμετρο θ, γεγονός που στην πράξη χρήζει προσοχής, ανάλογα με το υπό μελέτη φυσικό πρόβλημα.

2η Περίπτωση: $\boldsymbol{\mu}$ **γνωστό,** $\boldsymbol{\sigma = \theta}$ **άγνωστο**, $\theta \in \Theta = (0, \infty)$.
Τότε

$$f_1(x;\theta) = \begin{cases} \frac{1}{\theta} e^{-\frac{x-\mu}{\theta}} & , \ x \geqslant \mu \\ 0 & , \ \text{διαφορετικά} \end{cases},$$

οπότε έχουμε

$$L(\theta) = f(\underset{\sim}{x};\theta) = \prod_{i=1}^{n} f_1(x_i;\theta)$$

$$= \begin{cases} \frac{1}{\theta^n} e^{-\frac{1}{\theta}\sum\limits_{i=1}^{n}(x_i - \mu)} & , \ x_{(1)} \geqslant \mu \\ 0 & , \ \text{διαφορετικά.} \end{cases}$$

Για την εύρεση του ε.μ.π. του θ, πρέπει να μεγιστοποιήσουμε τη συνάρτηση $L_1(\theta) = \frac{1}{\theta^n} e^{-\frac{1}{\theta}\sum\limits_{i=1}^{n}(x_i-\mu)}$ ή ισοδύναμα το λογάριθμο της,

$$\ln L_1(\theta) = -n \ln \theta - \frac{1}{\theta} \sum_{i=1}^{n} (x_i - \mu)$$

για $\theta \in \Theta$. Χωρίς βλάβη της γενικότητας, θεωρούμε $\sum\limits_{i=1}^{n}(x_i - \mu) > 0$, αφού $\mathbb{P}(X_i \geqslant \mu) = 1$ και η κατανομή των X_i είναι συνεχής. Θέτοντας

$\frac{\partial}{\partial\theta}\ln L_1(\theta)=0$, προκύπτει η εξίσωση $-\frac{n}{\theta}+\frac{1}{\theta^2}\sum\limits_{i=1}^{n}(x_i-\mu)=0$, η οποία έχει μοναδική λύση $\hat{\theta}(\underset{\sim}{x})=\frac{1}{n}\sum\limits_{i=1}^{n}(x_i-\mu)$. Η λύση αυτή αντιστοιχεί σε μέγιστο, επειδή $\frac{\partial^2}{\partial\theta^2}\ln L_1(\theta)\ \Big|_{\theta=\hat{\theta}(\underset{\sim}{x})}=-\dfrac{n}{\hat{\theta}^2(\underset{\sim}{x})}<0$. Επιπροσθέτως, $\lim\limits_{\theta\to 0}\ln L_1(\theta)=\lim\limits_{\theta\to\infty}\ln L_1(\theta)=-\infty$ και συνεπώς η λύση $\hat{\theta}(\underset{\sim}{x})$ είναι θέση ολικού μεγίστου. Άρα ο ε.μ.π. του θ είναι $\hat{\theta}(\underset{\sim}{X})=\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)$.

<u>3η Περίπτωση</u>: $\boldsymbol{\mu}$**,** $\boldsymbol{\sigma}$ **άγνωστα**, $\theta=(\mu,\sigma)\in\Theta=\Re\times(0,\infty)$.

Τότε

$$f_1(x;\theta)=\begin{cases}\frac{1}{\sigma}e^{-\frac{x-\mu}{\sigma}} & ,\ x\geqslant\mu\\ 0 & ,\ \text{διαφορετικά}\end{cases},$$

οπότε έχουμε

$$\begin{aligned}L(\theta)=f(\underset{\sim}{x};\theta)&=\prod_{i=1}^{n}f_1(x_i;\theta)\\ &=\begin{cases}\frac{1}{\sigma^n}e^{-\frac{1}{\sigma}\sum\limits_{i=1}^{n}x_i+\frac{n\mu}{\sigma}} & ,\ x_{(1)}\geqslant\mu\\ 0 & ,\ \text{διαφορετικά.}\end{cases}\end{aligned}$$

Ανάλογα με τις προηγούμενες περιπτώσεις πρέπει να μεγιστοποιήσουμε τη συνάρτηση $L_1(\theta)=\frac{1}{\sigma^n}e^{-\frac{1}{\sigma}\sum\limits_{i=1}^{n}x_i+\frac{n\mu}{\sigma}}$ ως προς μ και σ, για $\mu\in[x_{(1)},\infty)$ και $\sigma\in(0,\infty)$. Παρατηρούμε ότι η συνάρτηση $L_1(\theta)$ είναι γνησίως αύξουσα ως προς μ και επομένως για κάθε $\sigma>0$,

$$L_1(\theta)=L_1(\mu,\sigma)\leqslant L_1(x_{(1)},\sigma)=\frac{1}{\sigma^n}e^{-\frac{1}{\sigma}\sum\limits_{i=1}^{n}x_i+\frac{nx_{(1)}}{\sigma}}=\frac{1}{\sigma^n}e^{-\frac{1}{\sigma}\sum\limits_{i=1}^{n}(x_i-x_{(1)})}. \tag{7.9}$$

Περαιτέρω, η συνάρτηση $L_1(x_{(1)},\sigma)$ μπορεί να μεγιστοποιηθεί ως προς σ, ακριβώς όπως η συνάρτηση $L_1(\theta)$ της 2ης Περίπτωσης ως προς θ, αφού είναι η ίδια συνάρτηση με μεταβλητή σ αντί θ και $x_{(1)}$ αντί μ. Κατ' αναλογία με την προηγούμενη περίπτωση θεωρούμε $\sum\limits_{i=1}^{n}(x_i-x_{(1)})>0$. Επομένως το μέγιστο της $L_1(x_{(1)},\sigma)$ επιτυγχάνεται για $\sigma=\frac{1}{n}\sum\limits_{i=1}^{n}(x_i-x_{(1)})=\hat{\sigma}$ (έστω), οπότε έχουμε

$$L_1(x_{(1)},\sigma)\leqslant L_1(x_{(1)},\hat{\sigma})\ ,\ \ \forall\sigma>0. \tag{7.10}$$

Συνδυάζοντας τις (7.9) και (7.10) προκύπτει ότι

$$L_1(\mu, \sigma) \leqslant L_1(x_{(1)}, \hat{\sigma}) \ , \quad \forall \mu \in [x_{(1)}, \infty) \ , \quad \sigma > 0$$

και η ισότητα ισχύει αν και μόνον αν $\mu = x_{(1)}$ και $\sigma = \hat{\sigma}$, το οποίο τελικά σημαίνει ότι το μέγιστο της $L_1(\mu, \sigma)$ επιτυγχάνεται για $(\mu, \sigma) = (x_{(1)}, \hat{\sigma})$. Άρα ο ε.μ.π. των μ και σ είναι $\hat{\mu} = X_{(1)}$ (που υπερεκτιμά το μ) και $\hat{\sigma} = \frac{1}{\sigma} \sum_{i=1}^{n} (X_i - X_{(1)})$.

Παράδειγμα 7.1.9. (Κανονική κατανομή με περιορισμένο παραμετρικό χώρο - ε.μ.π.) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$, $n \geq 2$, ένα τυχαίο δείγμα από την κανονική κατανομή $\mathcal{N}(\theta, \sigma^2)$, όπου σ^2 είναι γνωστό, θ είναι άγνωστο και $\theta \in \Theta = [\theta_1, \theta_2]$, με θ_1 και θ_2 δοθείσες (γνωστές) σταθερές. Η περίπτωση αυτή διαφέρει από την 1η Περίπτωση του Παραδείγματος 7.1.2, γιατί εδώ η άγνωστη παράμετρος (μέση τιμή) θ «περιορίζεται» στο διάστημα $[\theta_1, \theta_2]$ αντί να έχει τιμή στο $(-\infty, \infty)$. Στην πράξη, τα δεδομένα $\underset{\sim}{X}$ μπορούν να αφορούν ύψη ενός δείγματος από έναν πληθυσμό, οπότε το θ_1 παριστάνει ένα κάτω φράγμα του μέσου ύψους αυτού του πληθυσμού και αντίστοιχα το θ_2 είναι ένα άνω φράγμα του μέσου ύψους.

Όπως στο Παράδειγμα 7.1.2 (1η Περίπτωση) πρέπει να μεγιστοποιήσουμε τη συνάρτηση

$$L(\theta) = \frac{1}{\sigma^n (2\pi)^{n/2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \theta)^2}$$

ή ισοδύναμα το λογάριθμο της

$$\ln L(\theta) = -\frac{n}{2} \ln \sigma^2 - \frac{n}{2} \ln(2\pi) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \theta)^2 \ ,$$

όμως, για $\theta \in [\theta_1, \theta_2]$ και όχι για $\theta \in (-\infty, \infty)$.

Θέτοντας $\frac{\partial}{\partial \theta} \ln L(\theta) = 0$ προκύπτει η εξίσωση $\sum_{i=1}^{n} (x_i - \theta) = 0$ με μοναδική λύση $\theta = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}$. Αυτή η λύση αντιστοιχεί σε μέγιστο της $L(\theta)$ για $\theta \in (-\infty, \infty)$, όπως διαπιστώθηκε στο Παράδειγμα 7.1.2 (1η Περίπτωση), και προφανώς αντιστοιχεί σε μέγιστο της $L(\theta)$ για $\theta \in [\theta_1, \theta_2]$, *αν* $\bar{x} \in [\theta_1, \theta_2]$. Αν όμως $\bar{x} \notin [\theta_1, \theta_2]$, τότε πρέπει να αναζητηθεί σημείο του $[\theta_1, \theta_2]$ που μεγιστοποιεί την $L(\theta)$.

Κατ' αρχάς παρατηρούμε ότι η $\ln L(\theta)$ είναι γνησίως αύξουσα στο $(-\infty, \bar{x}]$ και γνησίως φθίνουσα στο $[\bar{x}, \infty)$. Αν $\bar{x} < \theta_1$, τότε η $\ln L(\theta)$, ως γνησίως φθίνουσα στο $[\theta_1, \theta_2] \subset [\bar{x}, \infty)$ έχει μέγιστο για $\theta = \theta_1$. Ανάλογα, αν $\bar{x} > \theta_2$ η $\ln L(\theta)$, ως γνησίως αύξουσα στο $[\theta_1, \theta_2] \subset (-\infty, \bar{x}]$ έχει μέγιστο για $\theta = \theta_2$. Συμπεραίνουμε λοιπόν ότι ο ε.μ.π. του θ είναι

$$\hat{\theta}(\underset{\sim}{X}) = \begin{cases} \theta_1 & , \bar{X} < \theta_1 \\ \bar{x} & , \theta_1 \leqslant \bar{X} \leqslant \theta_2 \\ \theta_2 & , \bar{X} > \theta_2 . \end{cases}$$

7.2 Ιδιότητες των εκτιμητών μέγιστης πιθανοφάνειας

Στην ενότητα αυτή μελετάμε γενικές ιδιότητες των εκτιμητών μέγιστης πιθανοφάνειας.

1. Είναι σημαντικό για έναν εκτιμητή να είναι συνάρτηση της (ελάχιστης) επαρκούς στατιστικής συνάρτησης, γιατί, αν δεν είναι, τότε είναι μη αποδεκτός και η Rao - Blackwell βελτίωσή του είναι καλύτερος εκτιμητής με κριτήριο το ΜΤΣ (Πρόταση 6.2.1). Υπό αυτό το πρίσμα, η αξία της αρχής της μέγιστης πιθανοφάνειας θα ετίθετο τουλάχιστον υπό αμφισβήτηση, αν ο ε.μ.π. δεν ήταν συνάρτηση της (ελάχιστης) επαρκούς στατιστικής συνάρτησης. Ως πρώτη και βασική ιδιότητα, στην επόμενη πρόταση, αποδεικνύουμε ότι ο ε.μ.π. είναι συνάρτηση της (ελάχιστης) επαρκούς στατιστικής συνάρτησης, εφ' όσον είναι μοναδικός.

Πρόταση 7.2.1. *Έστω $T(\underset{\sim}{X})$ επαρκής στατιστική συνάρτηση και $\hat{\theta}(\underset{\sim}{X})$ ε.μ.π. του θ, ο οποίος είναι μοναδικός. Τότε $\hat{\theta}(\underset{\sim}{X})$ είναι συνάρτηση του $T(\underset{\sim}{X})$.*

Απόδειξη. Από το παραγοντικό κριτήριο, η πυκνότητα του $\underset{\sim}{X}$, $f(\underset{\sim}{x}; \theta)$, επιδέχεται την παραγοντοποίηση

$$f(\underset{\sim}{x}; \theta) = q\big(T(\underset{\sim}{x}), \theta\big) h(\underset{\sim}{x}) \ , \ \ \forall \underset{\sim}{x} \ , \ \ \forall \theta \in \Theta \ , \tag{7.11}$$

όπου q και h είναι μη αρνητικές συναρτήσεις και η h δεν εξαρτάται από το θ. Επειδή ο ε.μ.π. $\hat{\theta}(\underset{\sim}{X})$ είναι μοναδικός, για κάθε τιμή $\underset{\sim}{x}$ του $\underset{\sim}{X}$, η εκτίμηση $\hat{\theta}(\underset{\sim}{x})$ μεγιστοποιεί μοναδικά, ως προς θ, τη συνάρτηση πιθανοφάνειας $L(\theta) = f(\underset{\sim}{x}; \theta)$ ή ισοδύναμα, λόγω της (7.11), μεγιστοποιεί μοναδικά την $q(T(\underset{\sim}{x}), \theta)$ αφού $h(\underset{\sim}{x}) \geqslant 0$ και η h δεν εξαρτάται από το θ. Έχουμε λοιπόν ότι

$$q(T(\underset{\sim}{x}), \hat{\theta}(\underset{\sim}{x})) = \max_{\theta \in \Theta} q(T(\underset{\sim}{x}), \theta) \ , \ \ \forall \underset{\sim}{x}$$

ή θέτοντας $T(\underset{\sim}{x}) = t$

$$q(t, \hat{\theta}(\underset{\sim}{x})) = \max_{\theta \in \Theta} q(t, \theta) \ , \ \ \forall \underset{\sim}{x}. \tag{7.12}$$

Για να δείξουμε ότι ο $\hat{\theta}(\underset{\sim}{X})$ είναι συνάρτηση του $T(\underset{\sim}{X})$, αρκεί να δείξουμε ότι, αν $\underset{\sim}{x}_1$ και $\underset{\sim}{x}_2$ είναι τιμές του $\underset{\sim}{X}$, τέτοιες ώστε $T(\underset{\sim}{x}_1) = T(\underset{\sim}{x}_2)$, τότε ισχύει $\hat{\theta}(\underset{\sim}{x}_1) = \hat{\theta}(\underset{\sim}{x}_2)$. Έστω λοιπόν τέτοια $\underset{\sim}{x}_1$ και $\underset{\sim}{x}_2$ και ας θέσουμε $t_0 = T(\underset{\sim}{x}_1) = T(\underset{\sim}{x}_2)$. Τότε από την (7.12) για $\underset{\sim}{x} = \underset{\sim}{x}_1$ και $\underset{\sim}{x} = \underset{\sim}{x}_2$ παίρνουμε, αντίστοιχα,

$$q(t_0, \hat{\theta}(\underset{\sim}{x}_1)) = \max_{\theta \in \Theta} q(t_0, \theta) \ \text{ και } \ q(t_0, \hat{\theta}(\underset{\sim}{x}_2)) = \max_{\theta \in \Theta} q(t_0, \theta),$$

δηλαδή η συνάρτηση $q(t_0, \theta)$ μεγιστοποιείται ως προς θ στα σημεία $\hat{\theta}(\underset{\sim}{x}_1)$ και $\hat{\theta}(\underset{\sim}{x}_2)$. Όμως η $q(t_0, \theta)$ μεγιστοποιείται μοναδικά ως προς θ, άρα $\hat{\theta}(\underset{\sim}{x}_1) = \hat{\theta}(\underset{\sim}{x}_2)$. □

Παρατήρηση 7.2.1. Η συνθήκη της μοναδικότητας είναι απαραίτητη για την ισχύ της Πρότασης 7.2.1. Στο Παράδειγμα 7.1.5, όπου υπάρχουν άπειροι ε.μ.π. του θ, η (ελάχιστη) επαρκής στατιστική συνάρτηση είναι $T(\underset{\sim}{X}) = (X_{(1)}, X_{(n)})$, διδιάστατη ενώ η παράμετρος είναι πραγματική. Οι ε.μ.π. $X_{(1)}$, $X_{(n)} - 1$, $\frac{1}{2}X_{(1)} + \frac{1}{2}(X_{(n)} - 1)$ είναι συναρτήσεις του $T(\underset{\sim}{X})$, ενώ ο ε.μ.π. $(\cos^2 X_1)X_{(1)} + (\sin^2 X_1)(X_{(n)} - 1)$ δεν είναι συνάρτηση του $T(\underset{\sim}{X})$. Σε όλα τα υπόλοιπα παραδείγματα της Ενότητας 7.1, ο ε.μ.π. του θ είναι μοναδικός και όντως είναι συνάρτηση της (ελάχιστης) επαρκούς στατιστικής συνάρτησης.

2. Θα παρατηρήσατε ότι ο ε.μ.π. ορίστηκε (μόνον) για την άγνωστη παράμετρο θ και όχι γενικότερα για μία άγνωστη τιμή $g(\theta)$. Σύμφωνα, όμως, με τον Ορισμό 7.1.1 για να έχει έννοια ο ε.μ.π. του $\xi = g(\theta)$, θα πρέπει στην πυκνότητα των δεδομένων $\underset{\sim}{X}$, $f(\underset{\sim}{x};\theta)$ $\theta \in \Theta$, να γίνει αλλαγή παραμέτρου από θ σε ξ και εν συνεχεία να μεγιστοποιηθεί η συνάρτηση πιθανοφάνειας ως προς ξ. Αυτή η αλλαγή παραμέτρου μπορεί να πραγματοποιηθεί αμέσως στην περίπτωση που η g είναι $1-1$ συνάρτηση και αυτή θα εξετάσουμε αρχικά. Έστω λοιπόν $\Xi = g(\Theta)$ ο (νέος) παραμετρικός χώρος για την (νέα) παράμετρο $\xi = g(\theta)$, δηλαδή το σύνολο Ξ είναι η εικόνα του Θ μέσω της g. Επειδή η g είναι $1-1$, αντιστρέφεται, άρα $\theta = g^{-1}(\xi)$ και η συνάρτηση πιθανοφάνειας γίνεται

$$L(\theta) = f(\underset{\sim}{x};\theta) = f(\underset{\sim}{x};g^{-1}(\xi)) = f^*(\underset{\sim}{x};\xi) = L^*(\xi) \ , \tag{7.13}$$

όπου $f^*(\underset{\sim}{x};\xi)$ είναι η πυκνότητα των δεδομένων $\underset{\sim}{X}$ με παράμετρο $\xi \in \Xi$ και $L^*(\xi)$ η αντίστοιχη συνάρτηση πιθανοφάνειας, η μεγιστοποίηση της οποίας ως προς $\xi \in \Xi$ παρέχει τον ε.μ.π. του ξ, έστω $\hat{\xi}(\underset{\sim}{X})$. Η επόμενη πρόταση αποδεικνύει ότι ο εκτιμητής $\hat{\xi}(\underset{\sim}{X})$ υπολογίζεται κατ' ευθείαν από τη σχέση $\hat{\xi}(\underset{\sim}{X}) = g(\hat{\theta}(\underset{\sim}{X}))$, όπου $\hat{\theta}(\underset{\sim}{X})$ είναι ε.μ.π. του θ, αντικαθιστώντας δηλαδή στον τύπο της $g(\theta)$, το θ με $\hat{\theta}$ (χωρίς να είναι απαραίτητη η μεγιστοποίηση της $L^*(\xi)$ ως προς ξ). Η ιδιότητα αυτή δηλώνει ότι η αρχή της μέγιστης πιθανοφάνειας είναι συμβατή με την αρχή της αντικατάστασης (βλέπε Ενότητα 3.3).

Πρόταση 7.2.2. *Έστω $\hat{\theta}(\underset{\sim}{X})$ ε.μ.π. του θ και g μία $1-1$ συνάρτηση ορισμένη στο Θ. Τότε ο εκτιμητής $g(\hat{\theta}(\underset{\sim}{X}))$ είναι ε.μ.π. του $g(\theta)$.*

Απόδειξη. Σύμφωνα με τον Ορισμό 7.1.1 του ε.μ.π., αρκεί να δείξουμε ότι ισχύει η σχέση

$$L^*(g(\hat{\theta}(\underset{\sim}{x}))) = \max_{\xi\in\Xi} L^*(\xi) \ , \quad \forall \underset{\sim}{x} \tag{7.14}$$

όπου, όπως παραπάνω, $\xi = g(\theta)$ και $\Xi = g(\Theta)$. Από την (7.13) παίρνουμε

$$L^*(g(\hat{\theta}(\underset{\sim}{x}))) = L(\hat{\theta}(\underset{\sim}{x})) \ ,$$

ενώ από τον ορισμό του $\hat{\theta}(\underset{\sim}{x})$, σχέση (7.4), έχουμε

$$L(\hat{\theta}(\underset{\sim}{x})) = \max_{\theta \in \Theta} L(\theta)$$

και συνεπώς

$$L^*(g(\hat{\theta}(\underset{\sim}{x}))) = \max_{\theta \in \Theta} L(\theta) \tag{7.15}$$

Επίσης από την (7.13) προκύπτει ότι

$$\max_{\theta \in \Theta} L(\theta) = \max_{\xi \in \Xi} L^*(\xi), \tag{7.16}$$

αφού $L(\theta) = L^*(\xi)$ και το Θ απεικονίζεται στο Ξ μέσω του μετασχηματισμού $\xi = g(\theta)$. Συνδυάζοντας την (7.15) με την (7.16), παίρνουμε την (7.14). □

Στην περίπτωση που η g δεν είναι $1-1$, η εξίσωση $\xi = g(\theta)$ δεν έχει μοναδική λύση ως προς θ και άρα δεν μπορεί να γίνει αλλαγή μεταβλητής από θ σε ξ στην πυκνότητα $f(\underset{\sim}{x};\theta)$. Αυτό συνεπάγεται ότι δεν έχει έννοια η πιθανοφάνεια, ως συνάρτηση του ξ. Συνεπώς για να οριστεί ο ε.μ.π. του $\xi = g(\theta)$ για αυθαίρετη συνάρτηση του g, απαιτείται να δοθεί πρώτα ένας γενικότερος ορισμός της πιθανοφάνειας ως συνάρτησης του $\xi = g(\theta)$. Έστω λοιπόν $\Xi = g(\Theta)$ και για κάθε $\xi \in \Xi$ ορίζουμε $\Theta_\xi = \{\theta \in \Theta : g(\theta) = \xi\}$ και

$$L^*(\xi) = \sup_{\theta \in \Theta_\xi} L(\theta) = \sup_{\theta \in \Theta_\xi} f(\underset{\sim}{x};\theta). \tag{7.17}$$

Η συνάρτηση $L^*(\xi)$, $\xi \in \Xi$, λέγεται *γενικευμένη συνάρτηση πιθανοφάνειας* ως προς ξ και ο ε.μ.π. του $\xi = g(\theta)$ είναι η τιμή $\hat{\xi}$ που μεγιστοποιεί τη συνάρτηση $L^*(\xi)$ για $\xi \in \Xi$. Έχουμε δηλαδή τον εξής γενικό ορισμό του ε.μ.π.

Ορισμός 7.2.1. *Έστω $\xi = g(\theta)$, όπου g αυθαίρετη συνάρτηση. Ο εκτιμητής $\hat{\xi}(\underset{\sim}{X})$ ονομάζεται εκτιμητής μέγιστης πιθανοφάνειας (ε.μ.π.) του ξ, εάν για κάθε τιμή $\underset{\sim}{x}$ του $\underset{\sim}{X}$ ισχύει η σχέση*

$$L^*\big(\hat{\xi}(\underset{\sim}{x})\big) = \max_{\xi \in \Xi} L^*(\xi). \tag{7.18}$$

Σημειώνουμε ότι στην περίπτωση $g(\theta) = \theta$, ο Ορισμός 7.2.1 συμπίπτει με τον Ορισμό 7.1.1, ενώ, εάν γενικότερα η g είναι $1-1$, τότε η $L^*(\xi)$ στην (7.17) συμπίπτει με την $L^*(\xi)$ στην (7.13) αφού $\Theta_\xi = \{g^{-1}(\xi)\}$, και ο Ορισμός 7.2.1 ανάγεται στον Ορισμό 7.1.1. Η επόμενη πρόταση επεκτείνει την Πρόταση 7.2.2 για συναρτήσεις g που δεν είναι κατ' ανάγκη $1-1$.

Πρόταση 7.2.3. *Έστω $\hat{\theta}(\underset{\sim}{X})$ ε.μ.π. του θ και g αυθαίρετη συνάρτηση ορισμένη στο Θ. Τότε, ο εκτιμητής $g(\hat{\theta}(\underset{\sim}{X}))$ είναι ε.μ.π. του $g(\theta)$.*

Απόδειξη. Σύμφωνα με τον Ορισμό 7.2.1, αρκεί να δείξουμε ότι ισχύει η σχέση

$$L^*(g(\hat{\theta}(\underset{\sim}{x}))) = \max_{\xi\in\Xi} L^*(\xi)\ ,\ \ \forall \underset{\sim}{x}\ , \tag{7.19}$$

όπου, $\xi = g(\theta)$ και L^* είναι η συνάρτηση στη σχέση (7.17). Από την (7.17) έχουμε

$$L^*(g(\hat{\theta}(\underset{\sim}{x}))) = \sup_{\theta\in\Theta_{g(\hat{\theta}(\underset{\sim}{x}))}} L(\theta). \tag{7.20}$$

Προφανώς, $\hat{\theta}(\underset{\sim}{x}) \in \Theta_{g(\hat{\theta}(\underset{\sim}{x}))} = \{\theta\in\Theta : g(\theta) = g(\hat{\theta}(\underset{\sim}{x}))\}$ και επειδή η τιμή $\hat{\theta}(\underset{\sim}{x})$ μεγιστοποιεί την συνάρτηση $L(\theta)$ για $\theta\in\Theta$, ειδικά μεγιστοποιεί την $L(\theta)$ και για $\theta \in \Theta_{g(\hat{\theta}(\underset{\sim}{x}))}$, δηλαδή

$$\sup_{\theta\in\Theta_{g(\hat{\theta}(\underset{\sim}{x}))}} L(\theta) = L(\hat{\theta}(\underset{\sim}{x})) = \max_{\theta\in\Theta} L(\theta). \tag{7.21}$$

Από τις (7.20) και (7.21) συνάγουμε ότι

$$L^*(g(\hat{\theta}(\underset{\sim}{x}))) = \max_{\theta\in\Theta} L(\theta). \tag{7.22}$$

Όμως, από τον ορισμό της $L^*(\xi)$ στην (7.17) προκύπτει ότι οι συναρτήσεις $L^*(\xi)$, $\xi\in\Xi$ και $L(\theta)$, $\theta\in\Theta$ έχουν το ίδιο supremum, δηλαδή

$$\sup_{\xi\in\Xi} L^*(\xi) = \sup_{\theta\in\Theta} L(\theta) = \max_{\theta\in\Theta} L(\theta). \tag{7.23}$$

Άρα, από τις (7.22) και (7.23) έχουμε

$$L^*(g(\hat{\theta}(\underset{\sim}{X}))) = \sup_{\xi\in\Xi} L^*(\xi)$$

που είναι η σχέση (7.19) αφού $g(\hat{\theta}(\underset{\sim}{x})) \in \Xi$. $\square$

Παρατήρηση 7.2.2. Ορισμένοι ερευνητές ορίζουν κατ' ευθείαν ως ε.μ.π. του $\xi = g(\theta)$ τον εκτιμητή $\hat{\xi} = g(\hat{\theta})$, όπου $\hat{\theta}$ είναι ε.μ.π. του θ, για κάθε συνάρτηση g. Ο ορισμός αυτός είναι συμβατός με τις Προτάσεις 7.2.1 και 7.2.2. Για περαιτέρω μελέτη όσον αφορά τον ορισμό του ε.μ.π. παραπέμπουμε στους Zehna (1966), Berk (1967), Scholz (1980, 2006) και Efron (1982).

Παράδειγμα 7.2.1. (κατανομή Bernoulli - ε.μ.π. του odds ratio $\frac{\theta}{1-\theta}$) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$, $n \geq 2$, ένα τυχαίο δείγμα από την κατανομή Bernoulli $\mathcal{B}(1, \theta)$, $\theta \in \Theta = (0, 1)$. Έχουμε $f_1(x; \theta) = \theta^x (1-\theta)^{1-x}$, $x = 0, 1$ και επομένως

$$
\begin{aligned}
L(\theta) = f(\underset{\sim}{x}; \theta) &= \prod_{i=1}^{n} f_1(x_i; \theta) \\
&= \begin{cases} \theta^{\sum\limits_{i=1}^{n} x_i} (1-\theta)^{n - \sum\limits_{i=1}^{n} x_i} & , \; x_i = 0, 1 \; , \; i = 1, \ldots, n \\ 0 & , \; \text{διαφορετικά.} \end{cases}
\end{aligned}
$$

Για $x_i = 0, 1$, έχουμε $\sum\limits_{i=1}^{n} x_i \in \{0, 1, \ldots, n\}$. Παρατηρούμε ότι, αν $\sum\limits_{i=1}^{n} x_i = n$ (δηλαδή $x_i = 1$, $i = 1, \ldots, n$), τότε $\sup\limits_{\theta \in \Theta} L(\theta) = \sup\limits_{\theta \in \Theta} \theta^n = 1$, το οποίο όμως δεν επιτυγχάνεται για καμία τιμή του $\theta \in \Theta = (0, 1)$, αφού $0 < \theta^n < 1$. Επομένως, δεν υπάρχει το $\max\limits_{\theta \in \Theta} L(\theta)$ και συνεπώς δεν υπάρχει ο ε.μ.π. του θ, αν $\sum\limits_{i=1}^{n} x_i = n$. Το ίδιο φαινόμενο παρατηρείται και αν $\sum\limits_{i=1}^{n} x_i = 0$ (δηλαδή $x_i = 0$, $i = 1, \ldots, n$) επειδή $\sup\limits_{\theta \in \Theta} L(\theta) = \sup\limits_{\theta \in \Theta} (1-\theta)^n = 1$, αλλά $0 < (1-\theta)^n < 1$ για κάθε $\theta \in (0, 1)$. Αυτή η «παθολογική» συμπεριφορά που εξάλλου έχει πιθανότητα που τείνει στο 0, μπορεί να διορθωθεί, αν θεωρήσουμε ως παραμετρικό χώρο το κλειστό διάστημα $\Theta^* = [0, 1]$, δηλαδή θεωρήσουμε ότι η πιθανότητα «επιτυχίας» θ μπορεί να είναι και 0 ή 1. Σε αυτήν την περίπτωση αν $\sum\limits_{i=1}^{n} x_i = n$, έχουμε $\max\limits_{0 \leqslant \theta \leqslant 1} L(\theta) = \max\limits_{0 \leqslant \theta \leqslant 1} \theta^n = 1 = L(1)$ και ανάλογα αν $\sum\limits_{i=1}^{n} x_i = 0$, έχουμε $\max\limits_{0 \leqslant \theta \leqslant 1} L(\theta) =$

$\max\limits_{0\leqslant\theta\leqslant 1}(1-\theta)^n = 1 = L(0)$. Απομένει λοιπόν να μεγιστοποιήσουμε την συνάρτηση $L(\theta)$ και συγκεκριμένα τον κλάδο $L_1(\theta) = \theta^{\sum\limits_{i=1}^{n} x_i}(1-\theta)^{n-\sum\limits_{i=1}^{n} x_i}$ ως προς $\theta \in \Theta^*$ για $0 < \sum\limits_{i=1}^{n} x_i < n$. Προφανώς το μέγιστο δεν επιτυγχάνεται για $\theta = 0$ ή $\theta = 1$, γιατί $L(0) = L(1) = 0$, άρα παρακάτω θεωρούμε $\theta \in (0,1)$. Έχουμε

$$\ln L_1(\theta) = \sum_{i=1}^{n} x_i \ln\theta + (n - \sum_{i=1}^{n} x_i)\ln(1-\theta)$$

και

$$\frac{\partial}{\partial\theta}\ln L_1(\theta) = \frac{\sum_{i=1}^{n} x_i}{\theta} - \frac{n-\sum_{i=1}^{n} x_i}{1-\theta}. \tag{7.24}$$

Η λύση της εξίσωσης πιθανοφάνειας $\frac{\partial}{\partial\theta}\ln L_1(\theta) = 0$ είναι μοναδική, $\theta = \frac{1}{n}\sum\limits_{i=1}^{n} x_i = \bar{x}$ και όντως αντιστοιχεί σε ολικό μέγιστο (επειδή $\frac{\partial^2}{\partial\theta^2}\ln L_1(\theta)\big|_{\theta=\bar{x}} = -\dfrac{n}{\bar{x}(1-\bar{x})} < 0$ και $\ln \underset{\theta\to 0}{L_1}(\theta) = \ln \underset{\theta\to 1}{L_1}(\theta) = -\infty$). Επομένως ο ε.μ.π. του $\theta \in \Theta^*$ είναι

$$\hat{\theta}(\underset{\sim}{X}) = \begin{cases} 0 & , \sum_{i=1}^{n} X_i = 0 \\ \bar{X} & , 0 < \sum_{i=1}^{n} X_i < n \\ 1 & , \sum_{i=1}^{n} X_i = n \end{cases},$$

δηλαδή $\hat{\theta}(\underset{\sim}{X}) = \bar{X}$. Επιπλέον ο ε.μ.π. του odds ratio $g(\theta) = \frac{\theta}{1-\theta}$, $\theta \in \Theta$ (βλέπε Παράδειγμα 4.2.3) σύμφωνα με την Πρόταση 7.2.2 είναι

$$g(\hat{\theta}(\underset{\sim}{X})) = \begin{cases} \dfrac{\hat{\theta}(\underset{\sim}{X})}{1-\hat{\theta}(\underset{\sim}{X})} & , 0 < \sum\limits_{i=1}^{n} X_i < n \\ \text{δεν υπάρχει} & , \sum\limits_{i=1}^{n} X_i = 0 \text{ ή } \sum\limits_{i=1}^{n} X_i = n \end{cases}$$

δηλαδή

$$g(\hat{\theta}(\underset{\sim}{X})) = \begin{cases} \dfrac{\bar{X}}{1-\bar{X}} & , 0 < \sum\limits_{i=1}^{n} X_i < n \\ \text{δεν υπάρχει} & , \sum\limits_{i=1}^{n} X_i = 0 \text{ ή } \sum\limits_{i=1}^{n} X_i = n. \end{cases}$$

Αν τώρα, ο παραμετρικός χώρος περιοριστεί στο σύνολο, έστω, $\Theta_1 = [0.25, 0.35]$, π.χ. το θ παριστάνει το ποσοστό που θα πάρει στις επόμενες βουλευτικές εκλογές ένα κόμμα με πολλούς υποστηρικτές, ο ε.μ.π. του θ δεν είναι κατ' ανάγκη $\bar{X}$, όπως όταν $\theta \in (0,1)$ (αγνοώντας τις μη ρεαλιστικές τιμές $\underset{\sim}{x} = (0,\ldots,0)$ και $\underset{\sim}{x} = (1,\ldots,1)$), αφού η μεγιστοποίηση της $L_1(\theta)$ πρέπει να γίνει για $\theta \in \Theta_1$. Από την (7.24), έχουμε $\frac{\partial}{\partial\theta}\ln L_1(\theta) > 0$ εάν και μόνον εάν

$$\frac{\sum_{i=1}^{n} x_i}{\theta} - \frac{n - \sum_{i=1}^{n} x_i}{1-\theta} > 0 \Leftrightarrow \theta < \bar{x},$$

το οποίο συνεπάγεται ότι η $L_1(\theta)$ είναι γνησίως αύξουσα στο διάστημα $(0, \bar{x})$, γνησίως φθίνουσα στο διάστημα $(\bar{x}, 1)$ και έχει ολικό μέγιστο στο $\Theta = (0,1)$ για $\theta = \bar{x}$. Αν λοιπόν $\bar{x} \in [0.25, 0.35]$ τότε η τιμή $\bar{x}$ μεγιστοποιεί την $L_1(\theta)$ και για $\theta \in \Theta_1$. Έστω ότι $\bar{x} < 0.25$. Τότε η $L_1(\theta)$ ως γνησίως φθίνουσα στο $(\bar{x}, 1)$, είναι ειδικά γνησίως φθίνουσα και στο $\Theta_1 = [0.25, 0.35] \subset (\bar{x}, 1)$, οπότε μεγιστοποιείται για $\theta = 0.25$. Έστω ότι $\bar{x} > 0.35$. Τότε η $L_1(\theta)$ ως γνησίως αύξουσα στο $(0, \bar{x})$, είναι ειδικά γνησίως αύξουσα και στο $[0.25, 0.35] \subset (0, \bar{x})$, οπότε μεγιστοποιείται για $\theta = 0.35$. Τελικά ο ε.μ.π. του θ για $\theta \in [0.25, 0.35]$ είναι

$$\hat{\theta}(\underset{\sim}{X}) = \begin{cases} 0.25 & , \bar{X} < 0.25 \\ \bar{X} & , 0.25 \leqslant \bar{X} \leqslant 0.35 \\ 0.35 & , \bar{X} > 0.35\ . \end{cases}$$

Οι εκτιμήσεις του ποσοστού θ, 0.25 και 0.35 ερμηνεύονται και διαισθητικά. Αν δηλαδή ο $\bar{X}$, το ποσοστό του δείγματος των ψηφοφόρων που υποστηρίζουν το κόμμα, «αποκλίνει» λίγο από τα άκρα του $\Theta_1 = [0.25, 0.35]$ είναι λογικό να θεωρηθεί ως εκτίμηση του θ το ένα ή το άλλο άκρο, αντίστοιχα εφ' όσον είναι δεδομένο ότι $\theta \in [0.25, 0.35]$. Αν ο $\bar{X}$ «αποκλίνει» πολύ από τα άκρα αυτά, τότε η εκτίμηση είναι μεν πάλι το αντίστοιχο άκρο, όμως σε αυτήν την περίπτωση θα πρέπει τουλάχιστον να επανεξεταστεί η διαδικασία συλλογής των συγκεκριμένων δεδομένων ή/και η θεώρηση (υπόθεση) ότι $\theta \in [0.25, 0.35]$.

3. Ο εκτιμητής μέγιστης πιθανοφάνειας είναι υπό ορισμένες συνθήκες (που δίνονται στη συνέχεια) *συνεπής εκτιμητής*. Η συνέπεια είναι ασυμπτωτική ιδιότητα, αναφέρεται δηλαδή σε μέγεθος δείγματος $n \to \infty$ (τέτοιες ιδιότητες λέγονται και ιδιότητες μεγάλου δείγματος - large sample properties).

Ένας εκτιμητής $T_n(X_1, \ldots, X_n)$ του $g(\theta)$ ονομάζεται *ασθενώς συνεπής*, εάν συγκλίνει κατά πιθανότητα στο $g(\theta)$, δηλαδή

$$T_n(X_1, \ldots, X_n) \xrightarrow{\mathbb{P}_\theta} g(\theta) \text{ καθώς } n \to \infty \ , \ \ \forall \theta \in \Theta \tag{7.25}$$

ή ισοδύναμα

$$\mathbb{P}_\theta\big(|T_n(X_1, \ldots, X_n) - g(\theta)| > \varepsilon\big) \to 0 \text{ καθώς } n \to \infty \ , \ \ \forall \varepsilon > 0 \ , \ \ \forall \theta \in \Theta.$$

Αντίστοιχα, ο $T_n(X_1, \ldots, X_n)$ ονομάζεται *ισχυρά συνεπής*, εάν συγκλίνει με πιθανότητα 1 στο $g(\theta)$, δηλαδή

$$T_n(X_1, \ldots, X_n) \xrightarrow{\text{μ.π. } 1} g(\theta) \text{ καθώς } n \to \infty \ , \ \ \forall \theta \in \Theta. \tag{7.26}$$

ή ισοδύναμα

$$\mathbb{P}_\theta(\lim_{n \to \infty} T_n(X_1, \ldots, X_n) = g(\theta)) = 1 \ , \ \ \forall \ \theta \in \Theta.$$

Σε αδρές γραμμές, και στις δύο περιπτώσεις, συνέπεια σημαίνει ότι, όταν υπάρχει διαθέσιμο μεγάλο πλήθος δεδομένων, n, ο εκτιμητής $T_n(X_1, \ldots, X_n)$ προσεγγίζει (συγκλίνει προς) την άγνωστη τιμή $g(\theta)$. Φυσικά, δεν χρειάζεται να επιχειρηματολογήσουμε ότι ένας καλός εκτιμητής «επιβάλλεται» να συγκλίνει προς την υπό εκτίμηση τιμή όταν $n \to \infty$. Με άλλα λόγια, η συνέπεια είναι μία ελάχιστη ασυμπτωτική ιδιότητα (minimal asymptotic property) που πρέπει να ικανοποιεί ένας υποψήφιος εκτιμητής. Μάλιστα, θα μπορούσαμε να την παραλληλίσουμε με την αποδεκτικότητα που επίσης είναι ελάχιστη ιδιότητα ως προς το κριτήριο του ΜΤΣ για σταθερό (πεπερασμένο) μέγεθος δείγματος n.

Είναι κατανοητό ότι η συνέπεια (ασθενής ή ισχυρή) είναι ιδιότητα της ακολουθίας των εκτιμητών $\{T_n(X_1, \ldots, X_n) : n = 1, 2, \ldots\}$ και όχι του αντιπροσωπευτικού μέλους της $T_n(X_1, \ldots, X_n)$, όμως χάριν απλότητας θα διατηρήσουμε την παραπάνω ορολογία. Επίσης, επειδή η σύγκλιση με πιθανότητα 1 συνεπάγεται τη σύγκλιση κατά πιθανότητα (Πρόταση 1.10.4), ένας εκτιμητής ισχυρά συνεπής είναι και ασθενώς συνεπής. Γενικά, όμως, είναι πιο εύκολο να αποδείξουμε ασθενή συνέπεια.

Διαισθητικά, ο ε.μ.π. του θ αναμένεται να είναι συνεπής λόγω της ερμηνείας του (βάσει της Πρότασης 7.1.1) ως το δειγματικό ανάλογο του θ. Εγγενής ιδιότητα εκτιμητή, που κατασκευάζεται ως δειγματικό ανάλογο παραμέτρου, είναι να προσεγγίζει την παράμετρο, καθώς αυξάνει το μέγεθος δείγματος, και αυτό ακριβώς σημαίνει, κατ' ουσίαν, συνέπεια. Ο Fisher (1912, 1922, 1925) ισχυρίστηκε ότι ο ε.μ.π. είναι «πάντοτε» συνεπής, χωρίς να δώσει όμως αυστηρή απόδειξη. Αυτός ο ισχυρισμός ήταν αφετηρία για έρευνα που και ακόμη επί των ημερών μας έχει ενδιαφέρον. Η πρόκληση για τη μελέτη της συνέπειας δεν είναι τόσο πια η απόδειξη της συνέπειας, αλλά κυρίως η εύρεση των πιο γενικών συνθηκών, υπό τις οποίες ισχύει και η κάλυψη περιπτώσεων που δεν έχουν απαντηθεί από προηγούμενη έρευνα. Μεμονωμένες περιπτώσεις μη συνεπούς ε.μ.π. υπάρχουν, (βλέπε Bahadur (1958), Le Cam (1979, 1990), Lehmann and Casella (1998, σελ. 495), Neymann and Scott (1948)).

Η επόμενη πρόταση αποδεικνύει υπό ορισμένες (απλές) συνθήκες την ισχυρή συνέπεια του ε.μ.π. του θ. Όπως στην Πρόταση 7.1.1, θεωρούμε ότι το $\underset{\sim}{X} = (X_1, \ldots, X_n)$ είναι ένα τυχαίο δείγμα από την κατανομή $f_1(x;\theta)$, με θ_0 συμβολίζουμε την αληθή τιμή του θ, με S_1 το σύνολο $\{x : f_1(x;\theta) > 0\}$ και με $L_n(\theta|\underset{\sim}{x}) = \prod_{i=1}^{n} f_1(x_i;\theta)$ τη συνάρτηση πιθανοφάνειας. Οι συνθήκες Σ1 και Σ2 που αναφέρονται στην Πρόταση 7.2.4 είναι αυτές της Πρότασης 7.1.1. Υποθέτουμε τα εξής.

Σ0. Ο παραμετρικός χώρος Θ είναι ανοικτό υποσύνολο του $\Re = (-\infty, \infty)$.

Σ3. Για κάθε $x \in S_1$ και $\theta \in \Theta$, υπάρχει η παράγωγος $\frac{\partial}{\partial\theta} f_1(x;\theta)$ και είναι πεπερασμένη.

Πρόταση 7.2.4. *Έστω ότι ισχύουν οι συνθήκες Σ0, Σ1, Σ2, Σ3. Τότε έχουμε τα εξής.*

(i). Υπάρχει μία ακολουθία εκτιμητών $\{\hat{\theta}_n(\underset{\sim}{X}) : n = 1, 2, \ldots\}$, η οποία έχει τις εξής ιδιότητες.

α. Με πιθανότητα 1 ως προς την κατανομή $\mathbb{P}_{\theta_0}$, $\hat{\theta}_n(\underset{\sim}{X})$ μεγιστοποιεί τοπικά τη συνάρτηση πιθανοφάνειας $L_n(\theta|\underset{\sim}{X})$ και είναι λύση της εξίσωσης πιθανοφάνειας $\frac{\partial}{\partial\theta}L_n(\theta|\underset{\sim}{X}) = 0$, για κάθε $n \geqslant n_0$, όπου το n_0 εξαρτάται εν γένει από το $\underset{\sim}{X}$,

β. $\hat{\theta}_n(\underset{\sim}{X}) \xrightarrow{\mu.\pi.\ 1} \theta_0$, ως προς την κατανομή $\mathbb{P}_{\theta_0}$.

(ii). Εάν υπάρχει ο ε.μ.π. του θ και η εξίσωση πιθανοφάνειας $\frac{\partial}{\partial\theta}L_n(\theta|\underset{\sim}{X}) = 0$ έχει μοναδική λύση, τότε ο ε.μ.π. του θ είναι ισχυρά συνεπής εκτιμητής του θ.

Απόδειξη. (i). (Σύμφωνα με τον Cramer (1946, σελ. 500-504) και τον Serfling (1980, σελ. 147-148)) Έστω m θετικός ακέραιος τέτοιος, ώστε $\theta_0 - \frac{1}{m}$ και $\theta_0 + \frac{1}{m} \in \Theta$. Η ύπαρξη του m εξασφαλίζεται από το γεγονός ότι το Θ είναι ανοικτό και άρα το θ_0 είναι εσωτερικό σημείο του. Θεωρούμε προς το παρόν ακέραιο $\kappa \geqslant m$, οπότε $\theta_0 \pm \frac{1}{\kappa} \in \Theta$. Από την Πρόταση 7.1.1, για $\theta = \theta_0 - \frac{1}{\kappa}$ και $\theta = \theta_0 + \frac{1}{\kappa}$ με πιθανότητα 1 ως προς $\mathbb{P}_{\theta_0}$ ισχύει $\lim_{n\to\infty} \frac{1}{n} \ln \frac{L_n(\theta_0 \mid \underset{\sim}{X})}{L_n(\theta_0 \pm \frac{1}{\kappa} \mid \underset{\sim}{X})} > 0$ και συνεπώς

$$L_n(\theta_0 \mid \underset{\sim}{x}) > L_n(\theta_0 - \frac{1}{\kappa} \mid \underset{\sim}{x}) \text{ και } L_n(\theta_0 \mid \underset{\sim}{x}) > L_n(\theta_0 + \frac{1}{\kappa} \mid \underset{\sim}{x}) \qquad (7.27)$$

για κάθε $n \geqslant N_\kappa(\underset{\sim}{x})$ και για κάθε $\underset{\sim}{x} \in S_\kappa$ με $\mathbb{P}_{\theta_0}(S_\kappa) = 1$. Από την Σ3, η $L_n(\theta|\underset{\sim}{X})$ είναι παραγωγίσιμη ως προς θ και άρα είναι συνεχής στο $[\theta_0 - \frac{1}{\kappa}, \theta_0 + \frac{1}{\kappa}]$. Λόγω της συνέχειας και της (7.27) παρουσιάζει τοπικό μέγιστο για $\theta = \hat{\theta}_{n\kappa}(\underset{\sim}{x})$, έστω, στο διάστημα $(\theta_0 - \frac{1}{\kappa}, \theta_0 + \frac{1}{\kappa})$. Περαιτέρω, η παραγωγισιμότητα συνεπάγεται ότι το $\hat{\theta}_{n\kappa}(\underset{\sim}{x})$ είναι λύση της εξίσωσης πιθανοφάνειας $\frac{\partial}{\partial\theta}L_n(\theta|\underset{\sim}{x}) = 0$. Ορίζουμε $S_0 = \bigcap_{\kappa=m}^{\infty} S_\kappa$, οπότε $\mathbb{P}_{\theta_0}(S_0) = 1$. [2] Έστω $\underset{\sim}{x} \in S_0$. Τότε $\underset{\sim}{x} \in S_\kappa$ για όλα τα $\kappa \geqslant m$ και χωρίς βλάβη

[2]Υποσημείωση: Εάν για μια ακολουθία ενδεχομένων A_n, $n = 1, 2, \ldots$ ισχύει $\mathbb{P}(A_n) =$

της γενικότητας μπορούμε να υποθέσουμε ότι η ακολουθία $N_\kappa(\underset{\sim}{x}), \kappa = m, m+1, \ldots$ είναι γνησίως αύξουσα ακολουθία θετικών ακεραίων. Για $n = 1, 2, \ldots$ ορίζουμε $\hat{\theta}_n(\underset{\sim}{x}) = \hat{\theta}_{n\kappa}(\underset{\sim}{x})$ αν $N_\kappa(\underset{\sim}{x}) \leqslant n < N_{\kappa+1}(\underset{\sim}{x})$ για κάποιο $\kappa = m, m+1, \ldots$ και $\hat{\theta}_n(\underset{\sim}{x}) = 0$ (αυθαίρετα) αν $n < N_m(\underset{\sim}{x})$. Θέτουμε $n_0 = N_m(\underset{\sim}{x})$. Επίσης, αν $\underset{\sim}{x} \notin S_0$ θέτουμε $\hat{\theta}_n(\underset{\sim}{x}) = 0$ (αυθαίρετα) για όλα τα $n = 1, 2, \ldots$. Τότε με πιθανότητα 1 ως προς $\mathbb{P}_{\theta_0}$, για $n \geqslant n_0$, το $\hat{\theta}_n(\underset{\sim}{X})$ μεγιστοποιεί τοπικά την $L_n(\theta|\underset{\sim}{X})$, αφού $\hat{\theta}_n(\underset{\sim}{X}) = \hat{\theta}_{n\kappa}(\underset{\sim}{X})$. Επί πλέον, από την κατασκευή του $\hat{\theta}_n(\underset{\sim}{x})$ για $\underset{\sim}{x} \in S_0$ έχουμε, $\hat{\theta}_n(\underset{\sim}{x}) \in (\theta_0 - \frac{1}{\kappa}, \theta_0 + \frac{1}{\kappa})$ για όλα τα $n \geqslant N_\kappa(\underset{\sim}{x})$ και συνεπώς $|\hat{\theta}_n(\underset{\sim}{x}) - \theta_0| < 1/\kappa$, $n \geqslant N_\kappa(\underset{\sim}{x})$ και για κάθε $\kappa \geqslant m$. Για $\kappa < m$, είναι $|\hat{\theta}_n(\underset{\sim}{x}) - \theta_0| < \frac{1}{m} < \frac{1}{\kappa}$, $n \geqslant N_m(\underset{\sim}{x})$.

Τελικά για κάθε $\kappa > 0$ και $\underset{\sim}{x} \in S_0$ υπάρχει $N_\kappa^*(\underset{\sim}{x})$ ($= N_\kappa(\underset{\sim}{x})$ ή $N_m(\underset{\sim}{x})$), έτσι ώστε $|\hat{\theta}_n(\underset{\sim}{x}) - \theta_0| < 1/\kappa$, $n \geqslant N_\kappa^*(\underset{\sim}{x})$. Επομένως, για $\underset{\sim}{x} \in S_0$ έχουμε $\hat{\theta}_n(\underset{\sim}{X}) \to \theta_0$ καθώς $n \to \infty$, δηλαδή $\hat{\theta}_n(\underset{\sim}{X}) \xrightarrow{\text{μ.π. } 1} \theta_0$ ως προς την κατανομή $\mathbb{P}_{\theta_0}$.

(ii). Εφ' όσον υπάρχει ο ε.μ.π., το Θ είναι ανοικτό και η $L_n(\theta|\underset{\sim}{X})$ είναι παραγωγίσιμη, τότε ο ε.μ.π. είναι η λύση της εξίσωσης πιθανοφάνειας $\frac{\partial}{\partial\theta} L_n(\theta|\underset{\sim}{X}) = 0$. Επειδή η λύση είναι μοναδική, ο ε.μ.π. συμπίπτει με τη λύση $\hat{\theta}_n(\underset{\sim}{X})$ του *i(α)* και η ισχυρή συνέπεια του προκύπτει από το *i(β)*. $\square$

Μερικά σχόλια για την Πρόταση 7.2.4 έχουν ως εξής. Ο αναλυτικός εντοπισμός του ολικού μεγίστου της συνάρτησης πιθανοφάνειας είναι γενικά μια δύσκολη διαδικασία, όταν υπάρχουν περισσότερα του ενός τοπικά μέγιστα και οι λύσεις της εξίσωσης πιθανοφάνειας δεν είναι δυνατόν να βρεθούν σε αναλυτική μορφή (σύνηθες φαινόμενο σε σύνθετα προβλήματα). Είναι όμως πιο εύκολο να αποδειχθεί η ύπαρξη ενός τοπικού μεγίστου (που είναι, άρα, υποψήφιο για ολικό μέγιστο), η θέση του οποίου βρίσκεται ακολούθως λύνοντας (αναλυτικά ή αριθμητικά) την εξίσωση πιθανοφάνειας. Για αυτούς τους λόγους, η διεθνής βιβλιογραφία και έρευνα έχει επικεντρωθεί προς την κατεύθυνση ύπαρξης συνεπούς λύσης της εξί-

---

1, τότε $\mathbb{P}(A) = 1$ όπου $A = \bigcap_{n=1}^{\infty} A_n$. Πράγματι $\mathbb{P}(A^c) = \mathbb{P}(\cup_{n=1}^{\infty} A_n) \leqslant \sum_{n=1}^{\infty} \mathbb{P}(A_n^c) = 0$, λόγω της υποπροσθετικής ιδιότητας (βλέπε Ενότητα 1.1). Άρα $\mathbb{P}(A^c) = 0$, που συνεπάγεται $\mathbb{P}(A) = 1$.

σωσης πιθανοφάνειας, παρά στο τεχνικά πιο δύσκολο πρόβλημα ύπαρξης και συνέπειας του ε.μ.π.. Ωστόσο, στις περιπτώσεις που η ύπαρξη και μοναδικότητα του ε.μ.π. μπορεί να διασφαλιστεί, η συνέπεια του προκύπτει από την Πρόταση 7.2.4(ii), εφ' όσον ικανοποιούνται οι υποθέσεις της.

Η σύγκλιση με πιθανότητα 1 ως προς $\mathbb{P}_{\theta_0}$, $\hat{\theta}_n(\underset{\sim}{X}) \to \theta_0$, και το γεγονός ότι η αληθής τιμή θ_0 μπορεί να είναι οποιοδήποτε σημείο του Θ, όντως δηλώνουν ότι ο $\hat{\theta}_n(\underset{\sim}{X})$ είναι ισχυρά συνεπής εκτιμητής του θ. Οπωσδήποτε δεν πρέπει να περάσει απαρατήρητο ότι οι συνθήκες Σ0 - Σ3 είναι πολύ ήπιες και ισχύουν σε πολλές πρακτικές εφαρμογές. Μάλιστα, το Θ δεν χρειάζεται να είναι ανοικτό, αρκεί το θ_0 να είναι εσωτερικό του σημείο. Ανάλογα, η παραγωγισιμότητα της $f_1(x;\theta)$ ως προς θ αρκεί να ισχύει σε μια περιοχή του θ_0. Το θέμα όμως είναι ότι δεν γνωρίζουμε «που μέσα στο Θ είναι το θ_0» και, για να ξεπεράσουμε αυτή τη δυσκολία, επιβάλλουμε τη συνθήκη ανοικτότητας του Θ και παραγωγισιμότητας σε όλο το Θ.

Η Πρόταση 7.2.4 απαιτεί τα δεδομένα $\underset{\sim}{X} = (X_1, \ldots, X_n)$ να αποτελούν τυχαίο δείγμα. Η γενίκευση της για οποιοδήποτε $\underset{\sim}{X}$ με παρατηρήσεις X_i ανεξάρτητες ή μη ανεξάρτητες και κατανομή κοινή ή μη κοινή έχει δοθεί από τον Kourouklis (1987) υπό επίσης ήπιες συνθήκες. Για την επέκταση της Πρότασης 7.2.4 σε διανυσματική παράμετρο $\theta = (\theta_1, \ldots, \theta_r)$ παραπέμπουμε στον Cramer (1946, σελ. 504) και τον Serfling (1980, σελ. 148).

Κλείνοντας τα σχόλια για την Πρόταση 7.2.4, μέσα στην πληθώρα της έρευνας για τη συνέπεια, θα ήταν παράλειψη, αν δεν ξεχωρίζαμε την σπουδαία και κλασική εργασία του Wald (1949), όπου δίνονται άλλου τύπου συνθήκες - χωρίς την απαίτηση παραγωγισιμότητας - για την ισχυρή συνέπεια του ε.μ.π. (και όχι λύσης της εξίσωσης πιθανοφάνειας). Η εργασία αυτή ξεπερνάει τα όρια αυτών των σημειώσεων. Ενδεικτικά, αναφέρουμε επίσης τους Hotelling (1930), Doob (1934, 1936), Bahadur (1958), Huber (1967), Hoadley (1971), Le Cam (1979, 1990), Bai and Fu (1987) και Wasserman (2003).

Θα ασχοληθούμε στη συνέχεια με την ιδιότητα της ασθενούς συνέπειας του ε.μ.π. και γενικότερα ενός εκτιμητή T_n του $g(\theta)$. Η επόμενη πρόταση καταδεικνύει ότι, μελετώντας μόνον τη μέση τιμή και τη διασπορά του εκτιμητή, μπορούμε να οδηγηθούμε στην ασθενή συνέπειά του.

Πρόταση 7.2.5. *Ο εκτιμητής T_n είναι ασθενώς συνεπής εκτιμητής του $g(\theta)$ εάν*

(α) $\mathbb{E}_\theta T_n \to g(\theta)$ *καθώς* $n \to \infty$, $\forall \theta \in \Theta$ *(ασυμπτωτική αμεροληψία)*

(β) $\mathbb{V}ar_\theta T_n \to 0$ *καθώς* $n \to \infty$, $\forall \theta \in \Theta$.

Απόδειξη. Για την τυχαία μεταβλητή Y από την ανισότητα Markov (Πρόταση 1.5.1), έχουμε $\mathbb{P}(|Y| > \varepsilon) = \mathbb{P}(Y^2 > \varepsilon^2) \leqslant \frac{\mathbb{E}Y^2}{\varepsilon^2}$ για κάθε $\varepsilon > 0$. Θέτουμε $Y = T_n - g(\theta)$, οπότε παίρνουμε

$$\begin{aligned}\mathbb{P}_\theta\big(|T_n - g(\theta)| > \varepsilon\big) &\leqslant \frac{\mathbb{E}_\theta\big(T_n - g(\theta)\big)^2}{\varepsilon^2}\\ &= \frac{\mathbb{V}ar_\theta T_n + \big(\mathbb{E}_\theta T_n - g(\theta)\big)^2}{\varepsilon^2} \to 0\end{aligned}$$

καθώς $n \to \infty$, βάσει των (α), (β). Επομένως, $\mathbb{P}_\theta\big(|T_n - g(\theta)| > \varepsilon\big) \to 0$ καθώς $n \to \infty$, το οποίο εξ' ορισμού σημαίνει ότι ο T_n είναι ασθενώς συνεπής εκτιμητής του $g(\theta)$. $\square$

Παρατήρηση 7.2.3. Οι συνθήκες της Πρότασης 7.2.5 έχουν μια πολύ απλή διαισθητική ερμηνεία. Όταν η διασπορά συγκλίνει στο 0, ο εκτιμητής «τείνει να γίνει σταθερά», αυτή η σταθερά, όμως, θα είναι και η μέση τιμή του. Αφού η μέση τιμή τείνει στο $g(\theta)$, αυτή η σταθερά θα είναι το $g(\theta)$. Τελικά λοιπόν, ο εκτιμητής «τείνει να γίνει $g(\theta)$», το οποίο σημαίνει ότι ο εκτιμητής είναι συνεπής.

Παρατήρηση 7.2.4. Η συνθήκη (β) μπορεί να αντικατασταθεί με την

(γ) $\mathbb{E}_\theta T_n^2 \to g^2(\theta)$, καθώς $n \to \infty$, $\forall \theta \in \Theta$.

Πράγματι, αν ισχύει η (γ), τότε σε συνδυασμό με την (α) συμπεραίνουμε ότι

$$\mathbb{V}ar_\theta T_n = \mathbb{E}_\theta T_n^2 - (\mathbb{E}_\theta T_n)^2 \to g^2(\theta) - g^2(\theta) = 0,$$

καθώς $n \to \infty$, για κάθε $\theta \in \Theta$, δηλαδή ισχύει και η (β), οπότε ισχύει και το συμπέρασμα της Πρότασης 7.2.5. Η συνθήκη (γ) είναι πρακτικά χρήσιμη στις περιπτώσεις που η επαλήθευση της (β) διέρχεται μέσα από τον υπολογισμό της $\mathbb{E}_\theta T_n^2$ και εν συνεχεία εφαρμογή του τύπου $\mathbb{V}ar_\theta T_n = \mathbb{E}_\theta T_n^2 - (\mathbb{E}_\theta T_n)^2$ για τον υπολογισμό της $\mathbb{V}ar_\theta T_n$.

Παραθέτουμε μερικά παραδείγματα εφαρμογής της Πρότασης 7.2.5.

Παράδειγμα 7.2.2. (**ασθενής συνέπεια των δειγματικών ροπών**) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από μία κατανομή με (πεπερασμένες) ροπές κ τάξης $\mu_\kappa = \mathbb{E}_\theta X_1^\kappa$, $\kappa = 1, 2, \ldots$, $\theta \in \Theta$. Θεωρούμε τα αντίστοιχα δειγματικά ανάλογα, δηλαδή τις δειγματικές ροπές κ τάξης $m_\kappa = \frac{1}{n}\sum_{i=1}^n X_i^\kappa$. Τότε οι στατιστικές συναρτήσεις m_κ είναι ασθενώς συνεπείς εκτιμητές των μ_κ, $\kappa = 1, 2, \ldots$, αντίστοιχα (εδώ, δηλαδή, $g(\theta) = \mu_\kappa$). Πράγματι,

$$\mathbb{E}_\theta(m_\kappa) = \mathbb{E}_\theta\left(\frac{1}{n}\sum_{i=1}^n X_i^\kappa\right) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}_\theta X_i^\kappa = \frac{1}{n}\sum_{i=1}^n \mu_\kappa = \mu_\kappa$$

και επομένως η συνθήκη *(α)* της Πρότασης 7.2.5 ισχύει τετριμμένα. Επιπλέον,

$$\begin{aligned}\mathbb{V}ar_\theta(m_\kappa) = \mathbb{V}ar_\theta\left(\frac{1}{n}\sum_{i=1}^n X_i^\kappa\right) &= \frac{1}{n^2}\sum_{i=1}^n \mathbb{V}ar_\theta X_i^\kappa = \\ &= \frac{1}{n^2}\sum_{i=1}^n (\mathbb{E}_\theta X_i^{2\kappa} - (\mathbb{E}_\theta X_i^\kappa)^2) \\ &= \frac{1}{n^2}\sum_{i=1}^n (\mu_{2\kappa} - \mu_\kappa^2) = \frac{1}{n}(\mu_{2\kappa} - \mu_\kappa^2) \to 0,\end{aligned}$$

καθώς $n \to \infty$, δηλαδή ισχύει και η συνθήκη *(β)*.

Σημειώνουμε, ότι η ασθενής συνέπεια των m_κ μπορεί επίσης να προκύψει κατευθείαν από τον ΑΝΜΑ (όπως και η ισχυρή συνέπεια, από τον ΙΝΜΑ).

Παράδειγμα 7.2.3. (**Ομοιόμορφη κατανομή - ασθενής συνέπεια του ε.μ.π.**) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$, $n \geqslant 1$, ένα τυχαίο δείγμα από την ομοιόμορφη κατανομή $\mathcal{U}[0,\theta]$, $\theta \in \Theta = (0,\infty)$. Στο Παράδειγμα 7.1.4, είδαμε ότι ο ε.μ.π. του θ είναι $\hat{\theta}_n(\underset{\sim}{X}) = X_{(n)}$. Η ισχυρή συνέπεια του $X_{(n)}$, αν και ισχύει, δεν μπορεί να αποδειχθεί μέσω της Πρότασης 7.2.4, γιατί δεν ικανοποιείται η συνθήκη Σ1, αφού το σύνολο $S_1 = \{x : f_1(x;\theta) > 0\} = [0,\theta]$ εξαρτάται από το θ. Θα αποδείξουμε, όμως, ασθενή συνέπεια επαληθεύοντας τις συνθήκες της Πρότασης 7.2.5 με $T_n = X_{(n)}$ και $g(\theta) = \theta$. Έχουμε

$$f_1(x;\theta) = \begin{cases} \frac{1}{\theta} & , 0 \le x \le \theta \\ 0 & , \text{διαφορετικά} \end{cases}.$$

Χρησιμοποιώντας την πυκνότητα του $X_{(n)}$ (βλέπε Παράδειγμα 6.3.3)

$$f_{X_{(n)}}(t;\theta) = \begin{cases} \frac{n}{\theta^n} t^{n-1} & , 0 \le t \le \theta \\ 0 & , \text{διαφορετικά} \end{cases}$$

παίρνουμε

$$\mathbb{E}_\theta X_{(n)} = \frac{n}{n+1}\,\theta \quad \text{και} \quad \mathbb{E}_\theta X_{(n)}^2 = \frac{n}{n+2}\,\theta^2,$$

οπότε $\mathbb{V}ar_\theta X_{(n)} = \mathbb{E}_\theta X_{(n)}^2 - (\mathbb{E}_\theta X_{(n)})^2 = \dfrac{n}{(n+2)(n+1)^2}\theta^2.$

Επομένως, $\mathbb{E}_\theta X_{(n)} \to \theta$ και $\mathbb{V}ar_\theta X_{(n)} \to 0$ καθώς $n \to \infty$ και συνεπώς $X_{(n)}$ είναι ασθενώς συνεπής εκτιμητής του θ. Σημειώνουμε ότι δεν χρειαζόταν ο υπολογισμός της $\mathbb{V}ar_\theta X_{(n)}$, γιατί ισχύει η συνθήκη (γ) της Παρατήρησης 7.2.4 αφού $\mathbb{E}_\theta X_{(n)}^2 \to \theta^2$.

Μία άλλη απόδειξη της ασθενούς συνέπειας του ε.μ.π. $X_{(n)}$, η οποία βασίζεται στον ορισμό της είναι η ακόλουθη. Για $\varepsilon > 0$ και, επειδή $\mathbb{P}_\theta(X_{(n)} \leqslant \theta) = 1$, έχουμε

$$\mathbb{P}_\theta(|X_{(n)} - \theta| > \varepsilon) = \mathbb{P}_\theta(\theta - X_{(n)} > \varepsilon) = \mathbb{P}_\theta(X_{(n)} < \theta - \varepsilon).$$

Αν $\varepsilon \geqslant \theta$, η τελευταία πιθανότητα είναι 0, αφού $\mathbb{P}_\theta(X_{(n)} \geqslant 0) = 1$. Αν $\varepsilon < \theta$, τότε $\mathbb{P}_\theta(X_{(n)} < \theta - \varepsilon) = \mathbb{P}_\theta(X_1 < \theta - \varepsilon, \ldots, X_n < \theta - \varepsilon) = \mathbb{P}_\theta(X_1 < \theta - \varepsilon) \ldots \mathbb{P}_\theta(X_n < \theta - \varepsilon) = \left(\frac{\theta-\varepsilon}{\theta}\right)^n = \left(1 - \frac{\varepsilon}{\theta}\right)^n \to 0$, καθώς $n \to \infty$. Συμπεραίνουμε λοιπόν ότι $\mathbb{P}_\theta(|X_{(n)} - \theta| > \varepsilon) \to 0$, καθώς $n \to \infty$, $\forall\, \theta \in \Theta$, $\forall\, \varepsilon > 0$, που είναι ο ορισμός της ασθενούς συνέπειας.

4. Ο εκτιμητής μέγιστης πιθανοφάνειας είναι υπό ορισμένες συνθήκες *ασυμπτωτικά κανονικός και αποδοτικός εκτιμητής* . Πιο συγκεκριμένα, εάν το θ είναι πραγματική παράμετρος, $\underset{\sim}{X} = (X_1, \ldots, X_n)$ είναι ένα τυχαίο δείγμα και $\hat{\theta}_n$ είναι ο ε.μ.π. του θ, θα δείξουμε ότι για «μεγάλο» μέγεθος δείγματος n, η κατανομή του $\hat{\theta}_n$ είναι κατά προσέγγιση κανονική $\mathcal{N}(\theta, \frac{1}{nI_1(\theta)})$, όπου $I_1(\theta)$ είναι η πληροφορία Fisher, που περιέχεται σε μία παρατήρηση X_i. Λαμβάνοντας υπ' όψη ότι $nI_1(\theta) = I(\theta)$ είναι η πληροφορία Fisher που περιέχεται στο δείγμα $\underset{\sim}{X}$ συμπεραίνουμε ότι $\frac{1}{nI_1(\theta)}$ είναι το κάτω φράγμα των Cramér–Rao, Κ.Φ. C-R, για τη διασπορά αμερόληπτων εκτιμητών του θ (βλέπε (5.6) και (5.12)). Περαιτέρω, λόγω της προσεγγιστικής κανονικής κατανομής, έχουμε

$$\mathbb{E}_\theta \hat{\theta}_n \approx \theta \quad \text{και} \quad \mathbb{V}ar_\theta \hat{\theta}_n \approx \frac{1}{nI_1(\theta)}.$$

Άρα, ο $\hat{\theta}_n$ είναι κατά προσέγγιση αμερόληπτος εκτιμητής του θ με διασπορά κατά προσέγγιση ίση προς το Κ.Φ. C-R. Σύμφωνα με τον ορισμό αποδοτικού εκτιμητή (βλέπε Ορισμό 5.2.2), δικαιολογούμαστε να χαρακτηρίσουμε τον ε.μ.π. $\hat{\theta}_n$ ως *ασυμπτωτικά ή κατά προσέγγιση αποδοτικό εκτιμητή* . Αυτή η ιδιότητα της (έστω και) ασυμπτωτικής αποδοτικότητας είναι όντως εκπληκτική!!! Την ασυμπτωτική αμεροληψία την προϊδεάζει η συνέπεια κατά κάποιο τρόπο: επειδή, εν γένει, $\hat{\theta}_n \to \theta$ καθώς $n \to \infty$, δεν θα ήταν παράξενο αν $\mathbb{E}_\theta \hat{\theta}_n \to \mathbb{E}_\theta \theta = \theta$ (προσοχή, δεν ισχυριζόμαστε ότι ισχύει πάντα αυτή η σχέση) οπότε $\mathbb{E}_\theta \hat{\theta}_n \approx \theta$. Το ότι $\mathbb{V}ar_\theta \hat{\theta}_n \approx 0$, επίσης προϊδεάζεται από τη συνέπεια (όπως αφήνεται να εννοηθεί στην Πρόταση 7.2.5). Όμως, ότι $\mathbb{V}ar_\theta \hat{\theta}_n \approx \frac{1}{nI_1(\theta)} =$ Κ.Φ. C-R δεν διαφαίνεται εξ' αρχής, εκτός αν διερευνηθεί το ερώτημα της ασυμπτωτικής κατανομής του ε.μ.π. $\hat{\theta}_n$. Η ιδιότητα της ασυμπτωτικής κανονικότητας και αποδοτικότητας είναι το «ισχυρό χαρτί» στα χέρια των υποστηρικτών της μεθόδου μέγιστης πιθανοφάνειας. Ξεκινώντας, δηλαδή, από μία απλή και διαισθητικά κατανοητή αρχή, αυτήν της μεγιστοποίησης της πιθανοφάνειας, κατασκευάζεται ένας εκτιμητής, ο ε.μ.π., που έστω και οριακά - ασυμπτωτικά ικανοποιεί τις βασικές ιδιότητες της αμεροληψίας και της ελάχιστης διασποράς, ίσης μάλιστα με το Κ.Φ. C-R!!!

Μετά την παραπάνω συζήτηση, ας δούμε τις τεχνικές λεπτομέρειες της ασυμπτωτικής κανονικότητας του ε.μ.π. του θ. Θεωρούμε ξανά την περίπτωση τυχαίου δείγματος $\underset{\sim}{X} = (X_1, \ldots, X_n)$ από την κατανομή με πυκνότητα $f_1(x;\theta)$, $\theta \in \Theta \subset \Re$, με $S_1 = \{x : f_1(x;\theta) > 0\}$, θ_0 είναι η αληθής τιμή του θ και επιβάλλουμε τις εξής συνθήκες που διατυπώνονται για συνεχή κατανομή. Στη διακριτή περίπτωση, τα ολοκληρώματα αντικαθίστανται με αθροίσματα ή σειρές. Κάποιες από τις ιδιότητες είναι παρόμοιες αυτών της ανισότητας των Cramér–Rao (Θεώρημα 5.1.1).

Σ4. Για κάθε $x \in S_1$, η πυκνότητα $f_1(x;\theta)$ είναι τρεις φορές παραγωγίσιμη ως προς θ και οι παράγωγοι είναι πεπερασμένες.

Σ5. $\displaystyle\int_{S_1} \frac{\partial}{\partial\theta} f_1(x;\theta)\,\mathrm{d}x = \frac{\partial}{\partial\theta}\int_{S_1} f_1(x;\theta)\,\mathrm{d}x$ $(= 0)$ και
$\displaystyle\int_{S_1} \frac{\partial^2}{\partial\theta^2} f_1(x;\theta)\,\mathrm{d}x = \frac{\partial^2}{\partial\theta^2}\int_{S_1} f_1(x;\theta)\,\mathrm{d}x$ $(= 0)$.

Σ6. $0 < I_1(\theta) < \infty$, για κάθε $\theta \in \Theta$, όπου $I_1(\theta) = \mathbb{E}_\theta\left(\frac{\partial}{\partial\theta}\ln f_1(X_1;\theta)\right)^2$ είναι ο αριθμός πληροφορίας του Fisher που περιέχεται στην παρατήρηση X_1.

Σ7. $\left|\frac{\partial^3}{\partial\theta^3}\ln f_1(x;\theta)\,\mathrm{d}x\right| \leqslant M(x)$ για κάθε $x \in S_1$ και $\theta_0 - \varepsilon < \theta < \theta_0 + \varepsilon$, με $\varepsilon > 0$ τέτοιο ώστε $\theta_0 \pm \varepsilon \in \Theta$, και $\mathbb{E}_{\theta_0} M(X_1) < \infty$.

Πρόταση 7.2.6. *Έστω ότι ισχύουν οι συνθήκες Σ0, Σ1, Σ4, Σ5, Σ6 και Σ7.*
(i). Αν $\{\hat{\theta}_n : n = 1, 2, \ldots\}$ είναι μία ακολουθία λύσεων της εξίσωσης πιθανοφάνειας $\frac{\partial}{\partial\theta}L_n(\theta) = 0$, η οποία είναι (ασθενώς) συνεπής, τότε έχουμε

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}_{\theta_0}} \mathcal{N}(\theta, \frac{1}{I_1(\theta_0)})\ , \tag{7.28}$$

όπου $\mathcal{L}_{\theta_0}$ συμβολίζει σύγκλιση κατά κατανομή ως προς $\mathbb{P}_{\theta_0}$.
(ii). Αν υπάρχει ο ε.μ.π. του θ, ικανοποιεί την εξίσωση πιθανοφάνειας και είναι (ασθενώς) συνεπής, τότε η ασυμπτωτική του κατανομή είναι αυτή της σχέσης (7.28).

Απόδειξη. (i). (Σύμφωνα με τους Cramér (1946, σελ. 500-504) και Lehmann and Casella (1998, σελ. 449-450).) Θέτουμε

$$l_n(\theta) = \ln L_n(\theta) = \ln \prod_{i=1}^{n} f_1(X_i;\theta) = \sum_{i=1}^{n} \ln f_1(X_i;\theta), \tag{7.29}$$

οπότε

$$\begin{aligned} l_n'(\theta) &= \tfrac{\partial}{\partial\theta} \ln L_n(\theta) = \textstyle\sum_{i=1}^n \tfrac{\partial}{\partial\theta} \ln f_1(X_i;\theta), \\ l_n''(\theta) &= \textstyle\sum_{i=1}^n \tfrac{\partial^2}{\partial\theta^2} \ln f_1(X_i;\theta), \\ l_n'''(\theta) &= \textstyle\sum_{i=1}^n \tfrac{\partial^3}{\partial\theta^3} \ln f_1(X_i;\theta). \end{aligned} \tag{7.30}$$

Θεωρώντας το ανάπτυγμα Taylor της συνάρτησης $l_n'(\theta)$ γύρω από το θ_0 έχουμε,

$$l_n'(\hat\theta_n) = l_n'(\theta_0) + (\hat\theta_n - \theta_0) l_n''(\theta_0) + \frac{1}{2}(\hat\theta_n - \theta_0)^2 l_n'''(\theta_n^*) \ , \tag{7.31}$$

όπου θ_n^* είναι μεταξύ $\hat\theta_n$ και θ_0. Εξ υποθέσεως, το πρώτο μέλος της (7.31) είναι μηδέν, οπότε έχουμε

$$\sqrt{n}(\hat\theta_n - \theta_0) = \frac{(1/\sqrt{n}) l_n'(\theta_0)}{-(1/n) l_n''(\theta_0) - (1/(2n))(\hat\theta_n - \theta_0) l_n'''(\theta_n^*)}. \tag{7.32}$$

Θα εξετάσουμε καθένα όρο του δεύτερου μέλους της (7.32) χωριστά. Από την (7.30),

$$l_n'(\theta_0) = \sum_{i=1}^{n} \tfrac{\partial}{\partial\theta} \ln f_1(X_i;\theta) \mid_{\theta=\theta_0}$$

είναι άθροισμα των ανεξάρτητων και με κοινή κατανομή τυχαίων μεταβλητών $\frac{\partial}{\partial\theta} \ln f_1(X_i;\theta) \mid_{\theta=\theta_0}$, $i = 1, 2, \ldots, n$. Λόγω της πρώτης από τις δύο σχέσεις στην Σ5, $\mathbb{E}_{\theta_0}\left(\frac{\partial}{\partial\theta} \ln f_1(X_i;\theta) \mid_{\theta=\theta_0}\right) = 0$. Τότε, λόγω της Σ6, προκύπτει ότι $\mathbb{V}ar_{\theta_0}\left(\frac{\partial}{\partial\theta} \ln f_1(X_i;\theta) \mid_{\theta=\theta_0}\right) = I_1(\theta_0)$. Επομένως, από το Κεντρικό Οριακό Θεώρημα (Θεώρημα 1.10.3), συμπεραίνουμε ότι

$$\frac{l_n'(\theta_0)}{\sqrt{nI_1(\theta_0)}} \xrightarrow{\mathcal{L}_{\theta_0}} \mathcal{N}(0,1)$$

ή ισοδύναμα

$$(1/\sqrt{n}) l_n'(\theta_0) \xrightarrow{\mathcal{L}_{\theta_0}} \mathcal{N}(0, I_1(\theta_0)). \tag{7.33}$$

Επίσης από την (7.30) έχουμε ότι $\frac{1}{n}l''_n(\theta_0) = \frac{1}{n}\sum_{i=1}^n \frac{\partial^2}{\partial\theta^2}\ln f_1(X_i;\theta)\ |_{\theta=\theta_0}$ είναι ο δειγματικός μέσος των ανεξάρτητων και με κοινή κατανομή τυχαίων μεταβλητών $\frac{\partial^2}{\partial\theta^2}\ln f_1(X_i;\theta)\ |_{\theta=\theta_0}$. Συνεπώς, από τον Ασθενή Νόμο των Μεγάλων Αριθμών (ΑΝΜΑ), Θεώρημα 1.10.1, έχουμε

$$\frac{1}{n}l''_n(\theta_0) \xrightarrow{\mathbb{P}_{\theta_0}} \mathbb{E}_{\theta_0}\left(\frac{\partial^2}{\partial\theta^2}\ln f_1(X_i;\theta)\ |_{\theta=\theta_0}\right).$$

Λόγω των Σ5 και Σ6,

$$\mathbb{E}_{\theta_0}\left(\frac{\partial^2}{\partial\theta^2}\ln f_1(X_i;\theta)\ |_{\theta=\theta_0}\right) = -I_1(\theta_0)\ ,$$

οπότε

$$\frac{1}{n}l''_n(\theta_0) \xrightarrow{\mathbb{P}_{\theta_0}} -I_1(\theta_0). \tag{7.34}$$

Θα μελετήσουμε τώρα τη συμπεριφορά της ακολουθίας των τυχαίων μεταβλητών

$$\frac{1}{n}l'''_n(\theta^*_n) = \frac{1}{n}\sum_{i=1}^n \frac{\partial^3}{\partial\theta^3}\ln f_1(X_i;\theta)\ |_{\theta=\theta^*_n}\ .$$

Αυτό το σημείο είναι το πιο λεπτό σημείο της απόδειξης. Κατ' αρχάς, επειδή $\hat{\theta}_n \xrightarrow{\mathbb{P}_{\theta_0}} \theta_0$ και $|\theta^*_n - \theta_0| < |\hat{\theta}_n - \theta_0|$ έχουμε $\theta^*_n \xrightarrow{\mathbb{P}_{\theta_0}} \theta_0$ και συνεπώς με πιθανότητα που τείνει στο 1, $\theta_0 - \varepsilon < \theta^*_n < \theta_0 + \varepsilon$. Τότε από την Σ7 και τον ΑΝΜΑ παίρνουμε

$$\begin{aligned}\left|\frac{1}{n}l'''_n(\theta^*_n)\right| &= \left|\frac{1}{n}\sum_{i=1}^n \frac{\partial^3}{\partial\theta^3}\ln f_1(X_i;\theta)\ |_{\theta=\theta^*_n}\right| \\ &\leqslant \frac{1}{n}\sum_{i=1}^n M(X_i) \xrightarrow{\mathbb{P}_{\theta_0}} \mathbb{E}_{\theta_0}M(X_1) < \infty\end{aligned}$$

και επομένως, λόγω της Πρότασης 1.10.7(ii, iii), η ακολουθία $\frac{1}{n}l'''_n(\theta^*_n)$ είναι φραγμένη κατά πιθανότητα ως προς $\mathbb{P}_{\theta_0}$, δηλαδή $\frac{1}{n}l'''_n(\theta^*_n) = O_{\mathbb{P}}(1)$. Σε συνδυασμό με το γεγονός ότι $\hat{\theta}_n - \theta_0 \xrightarrow{\mathbb{P}_{\theta_0}} 0$ και την Πρόταση 1.10.7(iv), καταλήγουμε στη σχέση

$$(\hat{\theta}_n - \theta_0)(\frac{1}{n}l'''_n(\theta^*_n)) \xrightarrow{\mathbb{P}_{\theta_0}} 0. \tag{7.35}$$

Από τις (7.34) και (7.35) έχουμε

$$B_n = -\frac{1}{n}l''_n(\theta_0) - \frac{1}{2}(\hat{\theta}_n - \theta_0)(\frac{1}{n}l'''_n(\theta^*_n)) \xrightarrow{\mathbb{P}_{\theta_0}} I_1(\theta_0),$$

οπότε από την (7.32) και το Θεώρημα Slutsky (Θεώρημα 1.10.6(ii)) προκύπτει ότι

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{1}{B_n}(\frac{1}{\sqrt{n}}l'_n(\theta_0)) \xrightarrow{\mathcal{L}_{\theta_0}} \frac{1}{I_1(\theta_0)}\mathcal{N}(0, I_1(\theta_0)).$$

Τελικά λοιπόν ισχύει

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}_{\theta_0}} \mathcal{N}(0, 1/I_1(\theta_0)).$$

(ii). Προκύπτει αμέσως από το πρώτο μέρος. □

Η χρησιμοποίηση του αναπτύγματος Taylor είναι μια κλασική πρακτική στη μελέτη της ασυμπτωτικής κατανομής στατιστικών συναρτήσεων. Από την αποδεικτική διαδικασία είναι φανερό (πια) γιατί ένας συνεπής ε.μ.π. του θ, $\hat{\theta}_n$, έχει ασυμπτωτική κανονική κατανομή. Από τις σχέσεις (7.30) και (7.32), βλέπουμε ότι

$$\hat{\theta}_n = \theta_0 + \frac{1}{nB_n}\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ln f_1(X_i; \theta) \mid_{\theta=\theta_0} .$$

Λόγω συνέπειας, όπως διαπιστώθηκε αυστηρά, ασυμπτωτικά έχουμε

$$B_n \approx I_1(\theta_0)$$

και άρα

$$\hat{\theta}_n \approx \theta_0 + \frac{1}{nI_1(\theta_0)}\left(\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ln f_1(X_i; \theta) \mid_{\theta=\theta_0}\right).$$

Συμπεραίνουμε λοιπόν ότι ασυμπτωτικά ο ε.μ.π. $\hat{\theta}_n$ συμπεριφέρεται ως μετατοπισμένος κατά θ_0 δειγματικός μέσος ανεξάρτητων τυχαίων μεταβλητών με κοινή κατανομή, οπότε το Κεντρικό Οριακό Θεώρημα αναλαμβάνει να διεκπεραιώσει ό,τι απομένει για την εύρεση της ασυμπτωτικής κατανομής του $\hat{\theta}_n$. Συγκεκριμένα,

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ln f_1(X_i; \theta) \mid_{\theta=\theta_0} \overset{\mathcal{L}_{\theta_0}}{\approx} \mathcal{N}(0, I_1(\theta_0)/n),$$

το οποίο συνεπάγεται ότι

$$\hat{\theta}_n \overset{\mathcal{L}_{\theta_0}}{\approx} \theta_0 + \frac{1}{I_1(\theta_0)}\mathcal{N}(0, I_1(\theta_0)/n) = \mathcal{N}(\theta_0, \frac{1}{nI_1(\theta_0)}).$$

Η Πρόταση 7.2.6 μπορεί να επεκταθεί και στην περίπτωση διανυσματικής παραμέτρου $\theta = (\theta_1, \ldots, \theta_r)$, βλέπε π.χ. Lehmann and Casella (1998, σελ. 461-465). Επίσης, υπάρχει πληθώρα βιβλιογραφίας, που αφορά σε γενικεύσεις ή εναλλακτικές υποθέσεις της Πρότασης 7.2.6 (π.χ. περιπτώσεις δεδομένων $\underset{\sim}{X}$ που δεν αποτελούν τυχαίο δείγμα), όπως Le Cam (1970, 1990).

Θα περιγράψουμε τώρα μια εναλλακτική μεθοδολογία, που οδηγεί στην ασυμπτωτική κανονικότητα χωρίς τη χρήση της τρίτης παραγώγου (συνθήκη Σ7). Η μεθοδολογία αυτή αναπτύχθηκε από τον Le Cam (1956) και αργότερα από τον Inagaki (1973). Χρησιμοποιώντας μόνον δύο (αντί τρεις) όρους στο ανάπτυγμα Taylor της $l'_n(\theta)$ γύρω από το θ_0 έχουμε

$$l'_n(\hat{\theta}_n) = l'_n(\theta_0) + (\hat{\theta}_n - \theta_0)l''_n(\theta^{**}_n), \tag{7.36}$$

όπου θ^{**}_n είναι μεταξύ $\hat{\theta}_n$ και θ_0, δηλαδή $|\theta^{**}_n - \theta_0| < |\hat{\theta}_n - \theta_0|$. Όπως και στη σχέση (7.31), το πρώτο μέλος της (7.36) είναι μηδέν, οπότε έχουμε

$$0 = \sqrt{n}(\frac{1}{n}l'_n(\theta_0)) + \sqrt{n}(\hat{\theta}_n - \theta_0)\frac{1}{n}l''_n(\theta^{**}_n). \tag{7.37}$$

Ο όρος $\sqrt{n}(\frac{1}{n}l'_n(\theta_0))$, όπως είδαμε στην (7.33), έχει ασυμπτωτική κανονική κατανομή $\mathcal{N}(0, I_1(\theta_0))$. Άρα η ασυμπτωτική συμπεριφορά του $\sqrt{n}(\hat{\theta}_n - \theta_0)$, θα καθοριστεί από την ασυμπτωτική συμπεριφορά του όρου $\frac{1}{n}l''_n(\theta^{**}_n)$. Η συνέπεια του $\hat{\theta}_n$ σημαίνει $\hat{\theta}_n \xrightarrow{\mathbb{P}_{\theta_0}} \theta_0$ και επομένως $\theta^{**}_n \xrightarrow{\mathbb{P}_{\theta_0}} \theta_0$, αφού $|\theta^{**}_n - \theta_0| < |\hat{\theta}_n - \theta_0|$. Περαιτέρω η σχέση, $\theta^{**}_n \xrightarrow{\mathbb{P}_{\theta_0}} \theta_0$ μας προδιαθέτει να αποδεχθούμε ότι $\frac{1}{n}l''_n(\theta^{**}_n) \approx \frac{1}{n}l''_n(\theta_0)$. Όμως, $\frac{1}{n}l''_n(\theta_0) = \frac{1}{n}\sum_{i=1}^n \frac{\partial^2}{\partial\theta^2}\ln f_1(X_i;\theta) \mid_{\theta=\theta_0} \approx \mathbb{E}\frac{\partial^2}{\partial\theta^2}\ln f_1(X_i;\theta) \mid_{\theta=\theta_0} = -I_1(\theta_0)$, λόγω του ΑΝΜΑ, το οποίο συνεπάγεται ότι

$$\frac{1}{n}l''_n(\theta^{**}_n) \approx -I_1(\theta_0).$$

Τότε, από την (7.37) παίρνουμε

$$0 \overset{\mathcal{L}_{\theta_0}}{\approx} \mathcal{N}(0, I_1(\theta_0)) + \sqrt{n}(\hat{\theta}_n - \theta_0)(-I_1(\theta_0)),$$

δηλαδή

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \overset{\mathcal{L}_{\theta_0}}{\approx} \mathcal{N}(0, 1/I_1(\theta_0))$$

ή

$$\hat{\theta}_n \overset{\mathcal{L}_{\theta_0}}{\approx} \mathcal{N}(\theta_0, \frac{1}{nI_1(\theta_0)}).$$

Η συνεπαγωγή $\theta_n^{**} \xrightarrow{\mathbb{P}_{\theta_0}} \theta_0 \Rightarrow \frac{1}{n}l_n''(\theta_n^{**}) \approx \frac{1}{n}l_n''(\theta_0)$, για να τεκμηριωθεί αυστηρά απαιτεί την ομοιόμορφη σύγκλιση της ακολουθίας $\frac{1}{n}l_n''(\theta)$ ως προς θ σε μια περιοχή του θ_0, και αυτήν την υπόθεση εισήγαγε, κατ' ουσίαν ο Le Cam (1956), προκειμένου να αποφύγει τη συνθήκη Σ7 και να καλύψει περιπτώσεις που αυτή δεν ισχύει. Στο ίδιο πνεύμα είναι και οι συνθήκες που έχουν προταθεί από τον Kulldorf (1957).

Παράδειγμα 7.2.4. (Μείξη κατανομών – συνέπεια και ασυμπτωτική κανονικότητα ε.μ.π.) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$, $n \geqslant 1$, ένα τυχαίο δείγμα από την κατανομή με πυκνότητα $f_1(x;\theta) = \frac{1}{2}\theta(1+\theta x)e^{-\theta x}$, $x > 0$, $\theta \in \Theta = (0, \infty)$. Η κατανομή αυτή προκύπτει εάν η παρατήρηση X_i επιλεγεί από την εκθετική κατανομή $\mathcal{E}(1/\theta)$, που είναι η κατανομή Γάμμα $\mathcal{G}(1, 1/\theta)$, με πιθανότητα $\frac{1}{2}$ ή από την κατανομή Γάμμα $\mathcal{G}(2, 1/\theta)$, με πιθανότητα επίσης $\frac{1}{2}$. Τέτοιες κατανομές αναφέρονται ως μείξεις, στη συγκεκριμένη περίπτωση έχουμε μείξη των δύο παραπάνω κατανομών Γάμμα με συντελεστές μείξης $1/2$ για κάθε μία. Η μέση τιμή της κατανομής είναι εύκολο να δειχθεί ότι είναι ο γραμμικός συνδυασμός (στην πραγματικότητα, ο κυρτός συνδυασμός) των μέσων τιμών των δύο κατανομών Γάμμα με συντελεστές τους αντίστοιχους συντελεστές μείξης δηλαδή $\mathbb{E}_\theta X_1 = \frac{1}{2}\frac{1}{\theta} + \frac{1}{2}\frac{2}{\theta} = \frac{3}{2\theta}$. Θα αναζητήσουμε ε.μ.π. του θ (από τον οποίο, αμέσως προκύπτει και ο ε.μ.π. της μέσης τιμής). Έχουμε,

$$L(\theta) = f(\underset{\sim}{x};\theta) = \frac{1}{2^n}\,\theta^n \prod_{i=1}^{n}(1+\theta x_i)e^{-\theta\sum_{i=1}^{n} x_i}\ ,$$

$$\ln L(\theta) = -n\ln 2 + n\ln\theta + \sum_{i=1}^{n}\ln(1+\theta x_i) - \theta\sum_{i=1}^{n}x_i\ ,$$

$$\frac{\partial}{\partial\theta}\ln L(\theta) = \frac{n}{\theta} + \sum_{i=1}^{n}\frac{x_i}{1+\theta x_i} - \sum_{i=1}^{n}x_i, \tag{7.38}$$

$$\frac{\partial^2}{\partial\theta^2}\ln L(\theta) = -\frac{n}{\theta^2} - \sum_{i=1}^{n}\frac{x_i^2}{(1+\theta x_i)^2} < 0,\ \ \forall\,\theta\in(0,\infty). \tag{7.39}$$

Από την (7.39) συμπεραίνουμε ότι η συνάρτηση $\frac{\partial}{\partial\theta}\ln L(\theta)$ είναι γνησίως φθίνουσα στο $(0,\infty)$, ενώ από την (7.38) ότι

$$\lim_{\theta\to 0^+}\frac{\partial}{\partial\theta}\ln L(\theta) = \infty \text{ και } \lim_{\theta\to\infty}\frac{\partial}{\partial\theta}\ln L(\theta) = -\sum_{i=1}^{n}x_i < 0.$$

Επομένως, λόγω συνέχειας, η συνάρτηση $\frac{\partial}{\partial\theta}\ln L(\theta)$ μηδενίζεται για μία και μοναδική τιμή στο $(0,\infty)$, έστω $\hat\theta(\underset{\sim}{x})$. Περαιτέρω, λόγω της μονοτονίας της $\frac{\partial}{\partial\theta}\ln L(\theta)$ για $\theta < \hat\theta(\underset{\sim}{x})$ έχουμε $\frac{\partial}{\partial\theta}\ln L(\theta) > \frac{\partial}{\partial\theta}\ln L(\theta)\ |_{\theta=\hat\theta(\underset{\sim}{x})} = 0$, το οποίο συνεπάγεται ότι η συνάρτηση $\ln L(\theta)$ είναι γνησίως αύξουσα στο διάστημα $(0,\hat\theta(\underset{\sim}{x}))$. Κατ' αναλογία, για $\theta > \hat\theta(\underset{\sim}{x})$ έχουμε $\frac{\partial}{\partial\theta}\ln L(\theta) < \frac{\partial}{\partial\theta}\ln L(\theta)\ |_{\theta=\hat\theta(\underset{\sim}{x})} = 0$, συνεπώς η συνάρτηση $\ln L(\theta)$ είναι γνησίως φθίνουσα στο διάστημα $(\hat\theta(\underset{\sim}{x}),\infty)$. Από τα παραπάνω καταλήγουμε ότι το σημείο $\hat\theta(\underset{\sim}{x})$ είναι η μοναδική θέση ολικού μεγίστου της $\ln L(\theta)$, δηλαδή ο ε.μ.π. του θ είναι $\hat\theta(\underset{\sim}{X})$.

Όπως δείξαμε, η ε.μ.π. $\hat\theta(\underset{\sim}{x})$ είναι η μοναδική λύση της εξίσωσης πιθανοφάνειας $\frac{\partial}{\partial\theta}\ln L(\theta) = 0$ στο $(0,\infty)$. Για $n=1$, από την (7.38) η εξίσωση πιθανοφάνειας γράφεται

$$\frac{1}{\theta} + \frac{x_1}{1+\theta x_1} - x_1 = 0$$

ή ισοδύναμα

$$\theta^2 x_1^2 - \theta x_1 - 1 = 0.$$

Οι λύσεις είναι $\theta = \dfrac{(1\pm\sqrt{5})x_1}{2}$ και καθώς στο $(0,\infty)$ ανήκει μόνον η $\dfrac{(1+\sqrt{5})x_1}{2}$, προκύπτει ότι η ε.μ.π. είναι $\hat\theta(X_1) = \dfrac{(1+\sqrt{5})X_1}{2}$. Για

$n \geqslant 2$, από την εξίσωση πιθανοφάνειας προκύπτει ισοδύναμη εξίσωση βαθμού $n+1$ ως προς θ, η θετική λυση της οποίας δεν είναι δυνατόν να βρεθεί σε κλειστή μορφή, αλλά μόνον με αριθμητική επίλυση.

Οι συνθήκες Σ0, Σ1, Σ2, Σ3 της Πρότασης 7.2.4 προφανώς ισχύουν, οπότε ο $\hat{\theta}(\underset{\sim}{X})$ είναι ισχυρά συνεπής εκτιμητής του θ. Η $f_1(x;\theta)$ έχει παραγώγους κάθε τάξης ως προς θ, άρα ισχύει και η συνθήκη Σ4. Επίσης, οι σχέσεις $\int_0^\infty \frac{\partial}{\partial\theta} f_1(x;\theta)dx = 0$ και $\int_0^\infty \frac{\partial^2}{\partial\theta^2} f_1(x;\theta)dx = 0$ επαληθεύονται εύκολα υπολογίζοντας τα αντίστοιχα ολοκληρώματα, άρα ισχύει η συνθήκη Σ5. Ο αριθμός πληροφορίας του Fisher $I_1(\theta)$ μπορεί να υπολογιστεί μέσω της σχέσης

$$\begin{aligned} I_1(\theta) = -\mathbb{E}_\theta \frac{\partial^2}{\partial\theta^2} \ln f_1(X_1;\theta) &= \mathbb{E}_\theta \left(\frac{1}{\theta^2} + \frac{X_1^2}{(1+\theta X_1)^2} \right) \\ &= \frac{1}{\theta^2} + \mathbb{E}_\theta \frac{X_1^2}{(1+\theta X_1)^2}. \end{aligned}$$

Έχουμε, $I_1(\theta) > \frac{1}{\theta^2}$ και $I_1(\theta) = \frac{1}{\theta^2} + \frac{1}{\theta^2}\mathbb{E}_\theta \left(\frac{\theta X_1}{1+\theta X_1} \right)^2 < \frac{2}{\theta^2}$, αφού $0 < \frac{\theta X_1}{1+\theta X_1} < 1$. Συνεπώς, $\frac{1}{\theta^2} < I_1(\theta) < \frac{2}{\theta^2}$, οπότε ισχύει και η συνθήκη Σ6. Τέλος, $\frac{\partial^3}{\partial\theta^3} \ln f_1(x;\theta) = \frac{2}{\theta^3} + \frac{2x^3}{(1+\theta x)^3} = \frac{2}{\theta^3} + \frac{2}{\theta^3} \left(\frac{\theta x}{1+\theta x} \right)^3 < \frac{4}{\theta^3}$ και επομένως ισχύει η συνθήκη Σ7. Βάσει της Πρότασης 7.2.6 συμπεραίνουμε ότι ο ε.μ.π. $\hat{\theta}(\underset{\sim}{X})$ έχει ασυμπτωτική κατανομή

$$\hat{\theta}(\underset{\sim}{X}) \overset{\mathcal{L}_\theta}{\approx} \mathcal{N}(\theta, \frac{1}{nI_1(\theta)}).$$

Για τον υπολογισμό του $I_1(\theta)$, έχουμε

$$\begin{aligned} \mathbb{E}_\theta \left(\frac{\theta X_1}{1+\theta X_1} \right)^2 &= \int_0^\infty \left(\frac{\theta x}{1+\theta x} \right)^2 \frac{1}{2}\, \theta(1+\theta x) e^{-\theta x}\, dx = \\ &= \frac{1}{2} \int_0^\infty \frac{\theta^2 x^2}{1+\theta x}\, \theta e^{-\theta x}\, dx = \\ &= \frac{1}{2} \int_0^\infty \frac{y^2}{1+y}\, e^{-y}\, dy = \frac{1}{2}\, e \int_1^\infty \frac{(u-1)^2}{u}\, e^{-u}\, du \\ &= \frac{1}{2}\, e \left\{ \int_1^\infty u\, e^{-u}\, du - 2\int_1^\infty e^{-u}\, du + \int_1^\infty \frac{e^{-u}}{u}\, du \right\} \end{aligned}$$

Όμως, $\int_1^\infty u\, e^{-u}\, du = 2\int_1^\infty e^{-u}\, du$, οπότε

$$\begin{aligned}\mathbb{E}_\theta\left(\frac{\theta X_1}{1+\theta X_1}\right)^2 &= \frac{1}{2}\, e\int_1^\infty \frac{e^{-u}}{u}\, du = \frac{1}{2}\, e\int_1^\infty e^{-u}\, d\ln u\\ &= \frac{1}{2}\, e\left\{e^{-u}\ln u\,|_1^\infty + \int_1^\infty e^{-u}\ln u\, du\right\} =\\ &= \frac{1}{2}\, e\int_1^\infty e^{-u}\ln u\, du,\end{aligned}$$

οπότε

$$I_1(\theta) = \frac{1}{\theta^2}\left(1+\frac{e}{2}\int_1^\infty e^{-u}\ln u\, du\right) = \frac{1}{a\theta^2}\ ,$$

όπου $\dfrac{1}{a} = 1+\dfrac{e}{2}\int_1^\infty e^{-u}\ln u\, du \approx 1.298$. Τελικά λοιπόν

$$\hat{\theta}(\underset{\sim}{X}) \overset{\mathcal{L}_\theta}{\approx} \mathcal{N}(\theta, \frac{a\theta^2}{n}).$$

5. Οι ιδιότητες της συνέπειας και της ασυμπτωτικής κανονικότητας μεταφέρονται πολύ εύκολα από την περίπτωση εκτίμησης του θ στη γενικότερη περίπτωση εκτίμησης του $g(\theta)$. Οι λεπτομέρειες δίνονται στην επόμενη πρόταση. Είναι αξιοσημείωτο ότι η πρόταση ισχύει για αυθαίρετο εκτιμητή T_n και άρα, ειδικά για τον ε.μ.π. του θ, $\hat{\theta}_n$ (βλέπε Πρόταση 7.2.8).

Πρόταση 7.2.7. *Έστω T_n (ασθενώς ή ισχυρά) συνεπής εκτιμητής του $\theta \in \Theta$ και g συνεχής συνάρτηση ορισμένη στο Θ. Τότε ισχύουν τα εξής.*

(ι) Ο $g(T_n)$ είναι (ασθενώς ή ισχυρά) συνεπής εκτιμητής του $g(\theta)$.

(ιι) (Μέθοδος Δέλτα) Εάν

$$\sigma_n(T_n-\theta_0) \xrightarrow{\mathcal{L}_{\theta_0}} \mathcal{N}(0,\sigma^2)\ ,$$

όπου $\sigma > 0$, $\sigma_n \to \infty$ και η παράγωγος $g'(\theta_0)$ υπάρχει, είναι πεπερασμένη και $g'(\theta_0) \neq 0$ τότε

$$\sigma_n(g(T_n)-g(\theta_0)) \xrightarrow{\mathcal{L}_{\theta_0}} \mathcal{N}(0,\sigma^2[g'(\theta_0)]^2). \tag{7.40}$$

Απόδειξη. (ι) Προκύπτει αμέσως από τις ιδιότητες της ασθενούς και της ισχυρής σύγκλισης ακολουθιών τυχαίων μεταβλητών (βλέπε Πρόταση 1.10.5(i, ii)).
(ιι) Θα αποδείξουμε, κατ' αρχάς, ότι

$$\frac{\sigma_n(g(T_n)-g(\theta_0))}{g'(\theta_0)} - \sigma_n(T_n-\theta_0) \xrightarrow{\mathbb{P}_{\theta_0}} 0. \tag{7.41}$$

Θεωρούμε τη συνάρτηση h ορισμένη στο Θ με τύπο

$$h(x) = \begin{cases} \dfrac{g(x)-g(\theta_0)}{x-\theta_0} - g'(\theta_0) & , x \neq \theta_0 \\ 0 & , x = \theta_0 \, . \end{cases}$$

Επειδή η g είναι συνεχής στο Θ και παραγωγίσιμη στο θ_0, η h είναι συνεχής στο Θ. Θέτουμε $\sigma_n(T_n-\theta_0) = Z_n$, οπότε $T_n = \theta_0 + \frac{1}{\sigma_n} Z_n$. Επειδή $\sigma_n \to \infty$ και $Z_n \xrightarrow{\mathcal{L}_{\theta_0}} \mathcal{N}(0,\sigma^2)$, προκύπτει ότι $\frac{1}{\sigma_n} Z_n \xrightarrow{\mathbb{P}_{\theta_0}} 0$ (Θεώρημα 1.10.6(v)) και επομένως $T_n \xrightarrow{\mathbb{P}_{\theta_0}} \theta_0$, λόγω της Πρότασης 1.10.5(*iii*). Η συνέχεια της h και η Πρόταση 1.10.5(*i*) εγγυώνται τότε ότι $h(T_n) \xrightarrow{\mathbb{P}_{\theta_0}} h(\theta_0) = 0$.
Τώρα έχουμε

$$\frac{\sigma_n(g(T_n)-g(\theta_0))}{g'(\theta_0)} - \sigma_n(T_n-\theta_0) = \begin{cases} \dfrac{h(T_n)\sigma_n(T_n-\theta_0)}{g'(\theta_0)} & , T_n \neq \theta_0 \\ 0 & , T_n = \theta_0 \, . \end{cases} \xrightarrow{\mathbb{P}_{\theta_0}} 0 \, ,$$

επειδή $h(T_n) \xrightarrow{\mathbb{P}_{\theta_0}} 0$ και $\sigma_n(T_n-\theta_0) \xrightarrow{\mathcal{L}_{\theta_0}} \mathcal{N}(0,\sigma^2)$, εφαρμόζοντας άλλη μια φορά το Θεώρημα 1.10.6(v). Επομένως ισχύει η (7.41), οπότε από το Θεώρημα Slutsky (Θεώρημα 1.10.6(i)), η ασυμπτωτική κατανομή της ακολουθίας $\dfrac{\sigma_n(g(T_n)-g(\theta_0))}{g'(\theta_0)}$ συμπίπτει με αυτήν της ακολουθίας $\sigma_n(T_n-\theta_0)$ που είναι $\mathcal{N}(0,\sigma^2)$. Άρα τελικά, από το Θεώρημα 1.10.6(iii) έχουμε,

$$\sigma_n(g(T_n)-g(\theta_0)) \xrightarrow{\mathcal{L}_{\theta_0}} \mathcal{N}(0,\sigma^2[g'(\theta_0)]^2).$$

□

Εφαρμόζοντας την Πρόταση 7.2.7 στην περίπτωση που ο T_n είναι ο ε.μ.π. του θ, $\hat{\theta}_n$, έχουμε αμέσως το εξής αποτέλεσμα για τον ε.μ.π. του $g(\theta)$, $g(\hat{\theta}_n)$.

Πρόταση 7.2.8. *Έστω $\hat{\theta}_n$ ο ε.μ.π. του θ και g παραγωγίσιμη συνάρτηση στο Θ με $g'(\theta_0) \neq 0$. Τότε ισχύουν τα εξής.*

(ι) Εάν ο $\hat{\theta}_n$ είναι (ασθενώς ή ισχυρά) συνεπής εκτιμητής του θ, τότε και ο ε.μ.π. του $g(\theta)$, $g(\hat{\theta}_n)$, είναι (ασθενώς ή ισχυρά) συνεπής.

(ιι) Εάν

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}_{\theta_0}} \mathcal{N}(0, 1/I_1(\theta_0)) \ ,$$

τότε

$$\sqrt{n}(g(\hat{\theta}_n) - g(\theta_0)) \xrightarrow{\mathcal{L}_{\theta_0}} \mathcal{N}(0, [g'(\theta_0)]^2/I_1(\theta_0)).$$

Η Πρόταση 7.2.8 αποκτά ιδιαίτερο ενδιαφέρον στην περίπτωση που ο ε.μ.π. $\hat{\theta}_n$ δεν μπορεί να βρεθεί σε κλειστή μορφή. Σημειώνουμε εδώ ότι η ασυμπτωτική κανονικότητα του $\hat{\theta}_n$ μπορεί να διασφαλιστεί μέσω της Πρότασης 7.2.6. Ωστόσο, εάν ο $\hat{\theta}_n$ υπάρχει σε κλειστή μορφή, τότε ενδεχομένως το αποτέλεσμα της Πρότασης 7.2.8(ii) μπορεί να προκύψει κατ' ευθείαν από τον μαθηματικό τύπο του $g(\hat{\theta}_n)$ και το Κεντρικό Οριακό Θεώρημα.

Παρατήρηση 7.2.5. Η Πρόταση 7.2.8(ii) δηλώνει ότι κατά προσέγγιση η κατανομή του $g(\hat{\theta}_n)$ είναι κανονική

$$g(\hat{\theta}_n) \overset{\mathcal{L}_{\theta_0}}{\approx} \mathcal{N}(g(\theta_0), [g'(\theta_0)]^2/(nI_1(\theta_0)))$$

με $\mathbb{E}_{\theta_0} g(\hat{\theta}_n) \approx g(\theta_0)$ και $\mathbb{V}ar_{\theta_0} g(\hat{\theta}_n) \approx [g'(\theta_0)]^2/(nI_1(\theta_0))$. Όμως $[g'(\theta_0)]^2/(nI_1(\theta_0))$ είναι το Κ.Φ. C-R για τη διασπορά αμερόληπτων εκτιμητών του $g(\theta)$. Συνεπώς οι παραπάνω σχέσεις σημαίνουν ότι ο ε.μ.π. $g(\hat{\theta}_n)$ είναι ασυμπτωτικά αμερόληπτος και ασυμπτωτικά αποδοτικός εκτιμητής του $g(\theta)$.

Παρατήρηση 7.2.6. Η τεχνική της μεθόδου Δέλτα, της απόδειξης δηλαδή της ασυμπτωτικής κανονικότητας του εκτιμητή $g(T_n)$ βασίζεται, στην ουσία, στην τοπική γραμμικοποίηση της συνάρτησης g χρησιμοποιώντας έναν όρο του αναπτύγματος Taylor. Πιο συγκεκριμένα, έχουμε

$$g(\theta) = g(\theta_0) + (\theta - \theta_0)g'(\theta^*) \text{ (τοπική γραμμικοποίηση ως προς } \theta),$$

όπου θ^* σημείο μεταξύ θ και θ_0. Ειδικά, λοιπόν,

$$g(T_n) = g(\theta_0) + (T_n - \theta_0)g'(\theta_n^*)$$

όπου θ^* είναι μεταξύ T_n και θ_0. Τώρα, λόγω συνέπειας, $T_n \to \theta_0$ και επειδή $|\theta_n^* - \theta_0| < |T_n - \theta_0| \to 0$, προκύπτει αμέσως ότι $\theta_n^* \to \theta_0$. Αυτό με τη σειρά του συνεπάγεται ότι τελικά (για $n \to \infty$), η $g(T_n)$ «οφείλει» να συμπεριφέρεται ως $g(T_n) \approx g(\theta_0) + (T_n - \theta_0)g'(\theta_0)$, δηλαδή να είναι γραμμική συνάρτηση της $T_n - \theta_0$ (Μάλιστα, αν η g' είναι συνεχής ισχύει, αυστηρά, η παραπάνω προσέγγιση). Η $T_n - \theta_0$ έχει ασυμπτωτική κανονική κατανομή $\mathcal{N}(\theta, \sigma^2/\sigma_n^2)$. Επομένως,

$$g(T_n) \overset{\mathcal{L}_{\theta_0}}{\approx} \mathcal{N}(g(\theta_0), [g'(\theta_0)]^2\sigma^2/\sigma_n^2)$$

που είναι το συμπέρασμα της Πρότασης 7.2.7(ii).

Παρατήρηση 7.2.7. Η παραπάνω μεθοδολογία καταδεικνύει γιατί απαιτείται $g'(\theta_0) \neq 0$, αλλά επίσης υποδηλώνει πώς πρέπει να χειριστούμε την περίπτωση που $g'(\theta_0) = 0$: θα χρησιμοποιήσουμε έναν ακόμη όρο του αναπτύγματος Taylor. Σε αυτήν την περίπτωση, έχουμε

$$\begin{aligned} g(\theta) &= g(\theta_0) + (\theta - \theta_0)g'(\theta_0) + \frac{1}{2}(\theta - \theta_0)^2 g''(\theta^{**}) \\ &= g(\theta_0) + \frac{1}{2}(\theta - \theta_0)^2 g''(\theta^{**}). \end{aligned}$$

Περαιτέρω,

$$g(T_n) = g(\theta_0) + \frac{1}{2}(T_n - \theta_0)^2 g''(\theta_n^{**}) \ ,$$

οπότε επαναλαμβάνοντας το προηγούμενο σκεπτικό,

$$g(T_n) \approx g(\theta_0) + \frac{1}{2}(T_n - \theta_0)^2 g''(\theta_0) \ ,$$

ή
$$\frac{2(g(T_n) - g(\theta_0))}{g''(\theta_0)} \approx (T_n - \theta_0)^2$$
(εφ' όσον $g''(\theta_0) > 0$). Η $T_n - \theta_0$ έχει (ασυμπτωτική) κανονική κατανομή με μέση τιμή 0, εξ υποθέσεως, άρα το τετράγωνο της παραπέμπει σε $\mathcal{X}_1^2$ και συγκεκριμένα
$$\frac{\sigma_n^2(T_n - \theta_0)^2}{\sigma^2} \xrightarrow{\mathcal{L}_{\theta_0}} \mathcal{X}_1^2$$
που συνεπάγεται
$$\frac{2\sigma_n^2(g(T_n) - g(\theta_0))}{\sigma^2 g''(\theta_0)} \xrightarrow{\mathcal{L}_{\theta_0}} \mathcal{X}_1^2.$$

Παραθέτουμε ένα παράδειγμα εφαρμογής της Πρότασης 7.2.8.

Παράδειγμα 7.2.5. (ασυμπτωτική κατανομή του ε.μ.π. του odds ratio $\frac{\theta}{1-\theta}$) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$, $n \geq 2$, ένα τυχαίο δείγμα από την κατανομή Bernoulli $\mathcal{B}(1, \theta)$, $\theta \in \Theta = (0, 1)$. Στο Παράδειγμα 7.2.1 είδαμε ότι ο ε.μ.π. του θ είναι $\hat{\theta}_n = \bar{X}$ (με πιθανότητα που τείνει στο 1) και ο ε.μ.π. του odds ratio $g(\theta) = \dfrac{\theta}{1-\theta}$ δίνεται από τη σχέση $g(\hat{\theta}_n) = \dfrac{\hat{\theta}_n}{1-\hat{\theta}_n} = \dfrac{\bar{X}}{1-\bar{X}}$. Από τους Νόμους των Μεγάλων Αριθμών προκύπτει αμέσως η ασθενής και ισχυρή συνέπεια του $\hat{\theta}_n$, οπότε από την Πρόταση 7.2.8(i) συνάγουμε και ότι ο ε.μ.π. $g(\hat{\theta}_n)$ είναι ασθενώς και ισχυρά συνεπής. Επιπλέον, κατευθείαν από το Κεντρικό Οριακό Θεώρημα έχουμε ότι
$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}_{\theta_0}} \mathcal{N}(0, \theta_0(1-\theta_0))$$
και επομένως από την Πρόταση 7.2.8(ii) συμπεραίνουμε ότι
$$\sqrt{n}(g(\hat{\theta}_n) - g(\theta_0)) \xrightarrow{\mathcal{L}_{\theta_0}} \mathcal{N}(0, \frac{\theta_0}{(1-\theta_0)^3}) \ ,$$
αφού $g'(\theta) = 1/(1-\theta)^2$. Κατά προσέγγιση λοιπόν έχουμε
$$\frac{\bar{X}}{1-\bar{X}} \overset{\mathcal{L}_{\theta_0}}{\approx} \mathcal{N}\left(\frac{\theta_0}{1-\theta_0}, \frac{\theta_0}{n(1-\theta_0)^3}\right).$$
Σημειώνουμε ότι η επαλήθευση της συνέπειας και της ασυμπτωτικής κανονικότητας του $\hat{\theta}_n = \bar{X}$ μπορεί να γίνει και μέσω των γενικών συνθηκών των Προτάσεων 7.2.4 και 7.2.6. Οι συνθήκες Σ0–Σ6 επαληθεύονται

πολύ εύκολα, τις περισσότερες εξ αυτών τις έχουμε ήδη συναντήσει και στην ανισότητα των Cramér–Rao (Πρόταση 5.1.1). Για την Σ7, έχουμε $f_1(x;\theta) = \theta^x(1-\theta)^{1-x}$, $x = 0,1$ και $\theta \in (0,1)$, οπότε $\frac{\partial^3}{\partial\theta^3}\ln f_1(X_1;\theta) = \dfrac{2X_1}{\theta^3} - \dfrac{2(1-X_1)}{(1-\theta)^3}$. Έστω τώρα $\varepsilon > 0$, τέτοιο ώστε $\theta_0 \pm \varepsilon \in (0,1)$, όπου $\theta_0 \in (0,1)$ η αληθής τιμή του θ. Τότε για κάθε θ με $\theta_0 - \varepsilon < \theta < \theta_0 + \varepsilon$, παρατηρούμε ότι

$$\left|\frac{\partial^3}{\partial\theta^3}\ln f_1(X_1;\theta)\right| \leqslant \frac{2X_1}{\theta^3}+\frac{2(1-X_1)}{(1-\theta)^3} < \frac{2X_1}{(\theta_0-\varepsilon)^3}+\frac{2(1-X_1)}{(1-\theta_0-\varepsilon)^3} = M(X_1)$$

με $\mathbb{E}_{\theta_0}M(X_1) = \dfrac{2\theta_0}{(\theta_0-\varepsilon)^3} + \dfrac{2(1-\theta_0)}{(1-\theta_0-\varepsilon)^3}$ και συνεπώς η Σ7 ισχύει.

Στο σημείο αυτό, κλείνοντας με τις ασυμπτωτικές ιδιότητες των ε.μ.π. επισημαίνουμε ότι η έρευνα σε συνέπεια και ασυμπτωτική κανονικότητα είναι ανεξάντλητη, πάντα επίκαιρη και αδύνατον να παρουσιαστεί στον περιορισμένο χώρο αυτών των σημειώσεων.

6. Είδαμε στις προηγούμενες ιδιότητες ότι ο ε.μ.π. είναι υπό ορισμένες συνθήκες ασυμπτωτικά αμερόληπτος και ασυμπτωτικά αποδοτικός. Θα εξετάσουμε τώρα την αμεροληψία και την αποδοτικότητα του ε.μ.π., για σταθερό (πεπερασμένο) μέγεθος δείγματος n.

Γενικά ο ε.μ.π. δεν είναι αμερόληπτος. Ακόμη και αν ο ε.μ.π. του θ, $\hat{\theta}_n$, είναι αμερόληπτος, ο ε.μ.π. του $g(\theta)$, $g(\hat{\theta}_n)$, δεν είναι αμερόληπτος, αφού η σχέση $\mathbb{E}_\theta\hat{\theta}_n = \theta$ εν γένει δεν συνεπάγεται τη σχέση $\mathbb{E}_\theta g(\hat{\theta}_n) = g(\theta)$ (εκτός φυσικά της περίπτωσης που η g είναι γραμμική συνάρτηση.). Από ορισμένους ερευνητές η μεροληψία του ε.μ.π. θεωρείται ως παθογένεια της μεθόδου, την οποία προσπαθούν να «διορθώσουν», επινοώντας τρόπους μείωσης της, π.χ. με τη μεθοδολογία jacknife. Παραπέμπουμε τον αναγνώστη στους Quenouille (1956) και την πρόσφατη εργασία Yatracos (2015) για περαιτέρω μελέτη του θέματος αυτού.

Στην περίπτωση, όμως, που ο ε.μ.π. είναι αμερόληπτος, τότε κατά κανόνα συμπίπτει με τον ΑΟΕΔ εκτιμητή. Πιο συγκεκριμένα, αν ο ε.μ.π. είναι μοναδικός τότε γνωρίζουμε πως είναι συνάρτηση της (ελάχιστης) επαρκούς στατιστικής συνάρτησης. Αν, περαιτέρω, αυτή είναι πλήρης, τότε από το Θεώρημα των Lehmann - Scheffé προκύπτει όντως ότι ο ε.μ.π. είναι και ΑΟΕΔ εκτιμητής. Τώρα, θα αντιστρέψουμε τους ρόλους ΑΟΕΔ εκτιμητή και ε.μ.π. και θα θέσουμε το ερώτημα: είναι ο ΑΟΕΔ εκτιμητής, ε.μ.π.; Την απάντηση ουσιαστικά την έχουμε δώσει παραπάνω και είναι εν γένει αρνητική, αφού ο ε.μ.π. είναι συνήθως μη αμερόληπτος. Το ερώτημα τέθηκε περισσότερο ρητορικά, ως εισαγωγή για τη διερεύνηση της σχέσης αποδοτικού εκτιμητή και ε.μ.π. (επαναλαμβάνουμε, για σταθερό μέγεθος δείγματος n). Στην επόμενη πρόταση αποδεικνύεται ότι ο αποδοτικός εκτιμητής είναι ουσιαστικά και ε.μ.π.. Διαπιστώνουμε λοιπόν ξανά την ύπαρξη αυτής της εκπληκτικής σχέσης μεταξύ αμεροληψίας – αποδοτικότητας και της αρχής της μέγιστης πιθανοφάνειας, αυτήν τη φορά όμως από την αντίστροφη κατεύθυνση. Περαιτέρω, γνωρίζουμε ότι για $\theta \in \Theta \subset \Re$, αποδοτικοί εκτιμητές υπάρχουν μόνον στις Μ.Ε.Ο.Κ. και για συγκεκριμένες συναρτήσεις $g(\theta)$. Γι' αυτές λοιπόν τις $g(\theta)$ ο αποδοτικός εκτιμητής συμπίπτει με τον ε.μ.π.

Πρόταση 7.2.9. *Έστω ότι ισχύουν οι συνθήκες* I1–I5 *της Πρότασης 5.2.5, $g(\theta)$ δεν είναι σταθερά ως συνάρτηση του θ, υπάρχει η παράγωγος $g'(\theta)$ και $T(\underset{\sim}{X})$ είναι αποδοτικός εκτιμητής του $g(\theta)$, του οποίου το σύνολο τιμών είναι υποσύνολο του $g(\Theta)$. Έστω ακόμη ότι η δεύτερη παράγωγος $\frac{\partial^2}{\partial\theta^2} f(\underset{\sim}{x};\theta)$ υπάρχει και είναι πεπερασμένη για κάθε $\underset{\sim}{x} \in \mathcal{S} = \{\underset{\sim}{x} : f(\underset{\sim}{x};\theta) > 0\}$. Τότε η εξίσωση πιθανοφάνειας*

$$\frac{\partial}{\partial\theta} \ln f(\underset{\sim}{X};\theta) = 0$$

έχει λύση $\hat{\theta}(\underset{\sim}{X})$ που μεγιστοποιεί τη συνάρτηση πιθανοφάνειας και ικανοποιεί τη σχέση

$$T(\underset{\sim}{X}) = g(\hat{\theta}(\underset{\sim}{X})).$$

Αν επί πλέον η g είναι 1-1 τότε η λύση $\hat{\theta}(\underset{\sim}{X})$ είναι μοναδική.

Απόδειξη. Επειδή ο $T(\underset{\sim}{X})$ είναι αποδοτικός εκτιμητής του $g(\theta)$, από την Πρόταση 5.2.5, έχουμε (με πιθανότητα 1)

$$\frac{\partial}{\partial\theta} \ln f(\underset{\sim}{X};\theta) = c_1(\theta)(T(\underset{\sim}{X}) - g(\theta)) \ , \ \ \forall\theta \in \Theta \tag{7.42}$$

όπου $c_1(\theta) \neq 0$ για κάθε $\theta \in \Theta$. Αφού $T(\underset{\sim}{X}) \in g(\Theta)$, υπάρχει $\hat{\theta}(\underset{\sim}{X}) \in \Theta$ τέτοιο ώστε

$$T(\underset{\sim}{X}) = g(\hat{\theta}(\underset{\sim}{X})). \tag{7.43}$$

και το οποίο, λόγω της (7.42) είναι λύση της εξίσωσης πιθανοφάνειας. Θα δείξουμε στη συνέχεια ότι η λύση $\hat{\theta}(\underset{\sim}{X})$ αντιστοιχεί σε μέγιστο της συνάρτησης πιθανοφάνειας. Παραγωγίζοντας την (7.42) προκύπτει

$$\tfrac{\partial^2}{\partial\theta^2} \ln f(\underset{\sim}{X};\theta) = c_1'(\theta)(T(\underset{\sim}{X}) - g(\theta)) - c_1(\theta)g'(\theta) ,$$

οπότε

$$\tfrac{\partial^2}{\partial\theta^2} \ln f(\underset{\sim}{X};\theta) \mid_{\theta=\hat{\theta}(\underset{\sim}{X})} = -c_1(\hat{\theta}(\underset{\sim}{X}))g'(\hat{\theta}(\underset{\sim}{X})). \tag{7.44}$$

Αρκεί λοιπόν να δείξουμε ότι το δεύτερο μέλος της (7.44) είναι αρνητικό.

Από τη σχέση (5.23) έχουμε

$$\mathbb{C}ov_\theta\left(\tfrac{\partial}{\partial\theta} \ln f(\underset{\sim}{X};\theta), T(\underset{\sim}{X})\right) = g'(\theta), \quad \forall\ \theta \in \Theta. \tag{7.45}$$

Αντικαθιστώντας στην (7.45), $T(\underset{\sim}{X}) = \dfrac{1}{c_1(\theta)}\tfrac{\partial}{\partial\theta} \ln f(\underset{\sim}{X};\theta) + g(\theta)$ που προκύπτει από την (7.42) και χρησιμοποιώντας ιδιότητες της συνδιασποράς (βλέπε Ενότητα 1.4), παίρνουμε

$$\frac{1}{c_1(\theta)}\mathbb{V}ar_\theta\left(\tfrac{\partial}{\partial\theta} \ln f(\underset{\sim}{X};\theta)\right) = g'(\theta)\ ,\ \ \forall\theta \in \Theta$$

Επομένως,

$$c_1(\theta)g'(\theta) = \mathbb{V}ar_\theta\left(\tfrac{\partial}{\partial\theta} \ln f(\underset{\sim}{X};\theta)\right) = I(\theta) > 0\ ,\ \ \forall\theta \in \Theta, \tag{7.46}$$

όπου $I(\theta)$ είναι ο αριθμός πληροφορίας του Fisher. Θέτοντας $\theta = \hat{\theta}(\underset{\sim}{X})$ στην (7.46) έχουμε

$$c_1(\hat{\theta}(\underset{\sim}{X}))g'(\hat{\theta}(\underset{\sim}{X})) > 0$$

και αυτό θέλαμε να δείξουμε.

Τέλος, αν η g είναι 1-1, η εξίσωση $T(\underset{\sim}{X}) = g(\theta)$ έχει μοναδική λύση ως προς θ, $\hat{\theta}(\underset{\sim}{X})$, η οποία λόγω της (7.46) είναι και μοναδική λύση της εξίσωσης πιθανοφάνειας. □

Παρατήρηση 7.2.8. Ας επικεντρώσουμε το ενδιαφέρον μας σε συναρτήσεις g που είναι 1-1 (ειδικά, καλύπτουμε και την περίπτωση $g(\theta) = \theta$). Τότε, λόγω της μοναδικότητας του $\hat{\theta}(\underset{\sim}{X})$, η συνάρτηση πιθανοφάνειας δεν παρουσιάζει άλλο ακρότατο παρά μόνον στο $\theta = \hat{\theta}(\underset{\sim}{X})$ στο Θ, το οποίο μάλιστα, όπως δείχθηκε, αντιστοιχεί σε μέγιστο. Επομένως, $\hat{\theta}(\underset{\sim}{X})$ είναι ο ε.μ.π. του θ, εκτός της ακραίας περίπτωσης κατά την οποία το supremum της συνάρτησης πιθανοφάνειας επιτυγχάνεται στα άκρα του Θ. Ως επακόλουθο, $g(\hat{\theta}(\underset{\sim}{X}))$ είναι ο ε.μ.π. του $g(\theta)$ και το συμπέρασμα της Πρότασης 7.2.9

$$T(\underset{\sim}{X}) = g(\hat{\theta}(\underset{\sim}{X}))$$

σημαίνει ότι η ύπαρξη αποδοτικού εκτιμητή εξασφαλίζει επίσης την ύπαρξη του ε.μ.π. και ότι οι δύο εκτιμητές συμπίπτουν!!!

Παράδειγμα 7.2.6. (Κανονική κατανομή – αποδοτικός εκτιμητής και ε.μ.π.) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$, $n \geq 2$, ένα τυχαίο δείγμα από την κανονική κατανομή $\mathcal{N}(\theta, \sigma^2)$, όπου $\theta \in \Theta = (-\infty, \infty)$ είναι άγνωστο και σ^2 γνωστή (θετική) σταθερά. Στο Παράδειγμα 5.2.11 είδαμε ότι ο δειγματικός μέσος $\bar{X}$ είναι αποδοτικός εκτιμητής του θ, ενώ στο Παράδειγμα 7.1.2 δείξαμε ότι είναι επίσης ε.μ.π. του θ. Η ταύτιση αυτή καθόλου συμπτωματική δεν είναι, αφού προκύπτει ως απόρροια της Πρότασης 7.2.9.

7. Οι ε.μ.π. σχετίζονται και με τους εκτιμητές Bayes, τους οποίους θα μελετήσουμε στο Κεφάλαιο 8. Για λόγους πληρότητας και συνολικής εικόνας των ιδιοτήτων των ε.μ.π., αναφέρουμε ότι ο ε.μ.π. είναι κατά προσέγγιση (δηλαδή για μεγάλο μέγεθος δείγματος) εκτιμητής Bayes. Περισσότερο, όμως, θα ασχοληθούμε με αυτήν την ιδιότητα στο επόμενο κεφάλαιο.

7.3 Αποτίμηση των εκτιμητών μέγιστης πιθανοφάνειας

Κλείνοντας την μελέτη των ε.μ.π., παραθέτουμε μία γενική αποτίμηση του ρόλου και της χρησιμότητάς τους.

1. Όπως αναφέρθηκε και στον πρόλογο αυτού του κεφαλαίου, η μέθοδος της μέγιστης πιθανοφάνειας είναι, ίσως, η πλέον γνωστή και ευρέως χρησιμοποιούμενη στις εφαρμογές μέθοδος εκτίμησης παραμέτρων. Είναι ιδιαίτερα απλή στην υλοποίησή της, αφού σε τελική ανάλυση η εύρεση του ε.μ.π. ανάγεται στην εύρεση της θέσης μεγίστου μιας συνάρτησης (εν προκειμένω, της συνάρτησης πιθανοφάνειας). Η δημοφιλία της οφείλεται και στο γεγονός ότι δεν απαιτεί εξειδικευμένη στατιστική γνώση, «οποιοσδήποτε» χρήστης μπορεί να την εφαρμόσει. Οι ε.μ.π. υπάρχουν πολύ συχνά και εξυπηρετούν περίπλοκα στατιστικά μοντέλα με πολλές παραμέτρους.
2. Εξ ορισμού, οι ε.μ.π. παίρνουν τιμές μέσα στο σύνολο τιμών των αντίστοιχων παραμέτρων.
3. Η μέθοδος είναι συμβατή με την αρχή της αντικατάστασης (Ενότητα 3.3), δηλαδή η εύρεση του ε.μ.π. του θ, $\hat{\theta}$, προσφέρει αμέσως και τον ε.μ.π. του $g(\theta)$, $g(\hat{\theta})$, για αυθαίρετη συνάρτηση g.
4. Υπάρχουν περιπτώσεις που οι ε.μ.π. «μονίμως» υποεκτιμούν ή υπερεκτιμούν τις αντίστοιχες παραμέτρους και έχουν κακή συμπεριφορά με κριτήριο το ΜΤΣ (Παραδείγματα 7.1.4, 7.1.7 και 7.1.8.)
5. Οι ε.μ.π. γενικά έχουν πολύ καλή συμπεριφορά, όταν υπάρχει διαθέσιμο μεγάλο πλήθος δεδομένων. Συγκεκριμένα, σε τυπικές περιπτώσεις είναι συνεπείς, ασυμπτωτικά αμερόληπτοι και έχουν ασυμπτωτικά κανονική κατανομή με την ελάχιστη δυνατή διασπορά (ασυμπτωτική αποδοτικότητα).
6. Οι ε.μ.π., εν γένει, δεν είναι αμερόληπτοι, όταν όμως είναι, τότε κατά κανόνα, συμπίπτουν με τον αποδοτικό εκτιμητή.
7. Οι ε.μ.π. είναι κατά προσέγγιση εκτιμητές Bayes.

7.4 Μέθοδος των ροπών

Εάν X_1 είναι μία τυχαία μεταβλητή, η *ροπή κ τάξης* της X_1 (ή της κατανομής της X_1), μ_κ, ορίζεται από τη σχέση

$$\mu_\kappa = \mathbb{E}_\theta X_1^\kappa,\ \kappa = 1, 2, \dots .$$

Εάν τώρα $\underset{\sim}{X} = (X_1, \ldots, X_n)$ είναι ένα τυχαίο δείγμα από την κατανομή με πυκνότητα $f_1(x;\theta)$, $\theta \in \Theta$, η *δειγματική ροπή κ τάξης* , m_κ, ορίζεται από τη σχέση

$$m_\kappa = \frac{1}{n}\sum_{i=1}^{n} X_i^\kappa, \ \kappa = 1, 2, \ldots .$$

Είναι προφανές ότι μ_κ είναι συναρτήσεις της παραμέτρου θ, ενώ m_κ είναι συναρτήσεις του δείγματος $\underset{\sim}{X}$, δηλαδή στατιστικές συναρτήσεις. Επιπλέον, οι δειγματικές ροπές m_κ είναι τα δειγματικά ανάλογα των ροπών της κατανομής, μ_κ, και επομένως μπορούν να χρησιμοποιηθούν ως εκτιμητές τους (βλέπε Ενότητα 3.3.1α). Οι δειγματικές ροπές m_κ ως εκτιμητές των μ_κ έχουν τις εξής δύο βασικές ιδιότητες.

1. $\mathbb{E}_\theta m_\kappa = \mu_\kappa$, $\forall\ \theta \in \Theta$, $\kappa = 1, 2, \ldots$, δηλαδή η δειγματική ροπή κ τάξης είναι αμερόληπτος εκτιμητής της ροπής κ τάξης της κατανομής. Πράγματι,
   $\mathbb{E}_\theta m_\kappa = \mathbb{E}_\theta\left(\frac{1}{n}\sum\limits_{i=1}^{n} X_i^\kappa\right) = \frac{1}{n}\sum\limits_{i=1}^{n}\mathbb{E}_\theta X_i^\kappa = \frac{1}{n}\sum\limits_{i=1}^{n}\mu_\kappa = \mu_\kappa.$
2. $m_\kappa \to \mu_\kappa$, καθώς $n \to \infty$, $\kappa = 1, 2, \ldots$, όπου η σύγκλιση είναι είτε κατά πιθανότητα είτε με πιθανότητα 1 και προκύπτει κατ' ευθείαν είτε από τον ΑΝΜΑ είτε από τον ΙΝΜΑ. Με άλλα λόγια, οι δειγματικές ροπές είναι (ασθενώς και ισχυρά) συνεπείς εκτιμητές των αντίστοιχων ροπών της κατανομής.

Στη δεύτερη ιδιότητα στηρίζεται η μέθοδος των ροπών για την εκτίμηση παραμέτρων. Η μέθοδος των ροπών είναι, ίσως, η πιο απλή μέθοδος εκτίμησης και, ιστορικά, από τις παλαιότερες αφού χρονολογείται από την εποχή του Άγγλου στατιστικού Carl Pearson στα τέλη του 1800. Οι εκτιμητές που προκύπτουν με αυτήν τη μέθοδο λέγονται *εκτιμητές μεθόδου ροπών* (ε.μ.ρ.) . Η μέθοδος μπορεί να περιγραφεί ως εξής.

Α. Έστω ότι η άγνωστη παράμετρος θ είναι διανυσματική διάστασης r, δηλαδή $\theta = (\theta_1, \ldots, \theta_r)$. Υπολογίζουμε r ροπές της κατανομής, συνήθως για χάρη απλότητας τις r πρώτες ροπές. Επειδή η κατανομή εξαρτάται από τα $\theta_1, \ldots, \theta_r$, το ίδιο ισχύει γενικά και για τις ροπές της. Έτσι γράφουμε $\mu_\kappa(\theta_1, \ldots, \theta_r)$ αντί μ_κ και επομένως έχουμε

$$
\begin{aligned}
\mu_1(\theta_1, \ldots, \theta_r) &= \mathbb{E}_\theta X_1, \\
\mu_2(\theta_1, \ldots, \theta_r) &= \mathbb{E}_\theta X_1^2, \\
&\vdots \\
\mu_r(\theta_1, \ldots, \theta_r) &= \mathbb{E}_\theta X_1^r.
\end{aligned}
$$

B. Λόγω της ιδιότητας 2, οι δειγματικές ροπές είναι προσεγγίσεις των ροπών μ_κ, δηλαδή

$$\mu_\kappa(\theta_1, \ldots, \theta_r) \approx m_\kappa$$

για μεγάλο n. Τότε δικαιολογούμαστε, εν μέρει, αν *απαιτήσουμε* αυτές οι προσεγγιστικές σχέσεις να ισχύουν ως *ισότητες*. Θέτουμε λοιπόν

$$
\begin{aligned}
\mu_1(\theta_1, \ldots, \theta_r) &= m_1, \\
\mu_2(\theta_1, \ldots, \theta_r) &= m_2, \\
&\vdots \\
\mu_r(\theta_1, \ldots, \theta_r) &= m_r,
\end{aligned}
$$

και λύνουμε το σύστημα των r εξισώσεων με r αγνώστους ως προς $\theta_1, \ldots, \theta_r$. Η λύση του συστήματος $\hat{\hat{\theta}}_1, \ldots, \hat{\hat{\theta}}_r$ είναι εξ ορισμού οι ε.μ.ρ. των $\theta_1, \ldots, \theta_r$. Γενικότερα, ως ε.μ.ρ. του $g(\theta_1, \ldots, \theta_r)$ ορίζεται η στατιστική συνάρτηση $g(\hat{\hat{\theta}}_1, \ldots, \hat{\hat{\theta}}_r)$, ορισμός συμβατός με την αρχή της αντικατάστασης (Ενότητα 3.3).

Μία παραλλαγή αυτής της περιγραφής της μεθόδου των ροπών, η οποία βασίζεται στην αρχή της αντικατάστασης είναι η ακόλουθη. Διατηρούμε το βήμα Α, αλλά αντί του βήματος Β λύνουμε το σύστημα των εξισώσεων του Α ως προς $\theta_1, \ldots, \theta_r$. Η λύση αυτή εκφράζει τα $\theta_1, \ldots, \theta_r$ ως συναρτήσεις των $\mathbb{E}_\theta X_1, \mathbb{E}_\theta X_1^2, \ldots, \mathbb{E}_\theta X_1^r$, δηλαδή

$$\begin{aligned}
\theta_1 &= h_1(\mathbb{E}_\theta X_1, \ldots, \mathbb{E}_\theta X_1^r),\\
\theta_2 &= h_2(\mathbb{E}_\theta X_1, \ldots, \mathbb{E}_\theta X_1^r),\\
&\vdots\\
\theta_r &= h_r(\mathbb{E}_\theta X_1, \ldots, \mathbb{E}_\theta X_1^r).
\end{aligned}$$

Οι ροπές $\mathbb{E}_\theta X_1, \ldots, \mathbb{E}_\theta X_1^r$ εκτιμώνται, αντίστοιχα, με τις δειγματικές ροπές, $m_1, \ldots, m_r$, οπότε επικαλούμενοι την αρχή της αντικατάστασης, οι ε.μ.ρ. των $\theta_1, \ldots, \theta_r$ είναι, αντίστοιχα,

$$\begin{aligned}
\hat{\hat{\theta}}_1 &= h_1(m_1, \ldots, m_r),\\
\hat{\hat{\theta}}_2 &= h_2(m_1, \ldots, m_r),\\
&\vdots\\
\hat{\hat{\theta}}_r &= h_r(m_1, \ldots, m_r).
\end{aligned}$$

που προφανώς συμπίπτουν με τη λύση του συστήματος του βήματος Β. Γενικότερα, εάν το $g(\theta_1, \ldots, \theta_r)$ εκφραστεί ως συνάρτηση των $\mathbb{E}_\theta X_1, \ldots, \mathbb{E}_\theta X_1^r$, δηλαδή $g(\theta_1, \ldots, \theta_r) = h(\mathbb{E}_\theta X_1, \ldots, \mathbb{E}_\theta X_1^r)$, τότε με βάση την αρχή της αντικατάστασης, ως ε.μ.ρ. του $g(\theta_1, \ldots, \theta_r)$ ορίζεται η στατιστική συνάρτηση $h(m_1, \ldots, m_r)$ που δεν είναι άλλη από την $g(\hat{\hat{\theta}}_1, \ldots, \hat{\hat{\theta}}_r)$.

Δίνουμε στη συνέχεια μερικά παραδείγματα υπολογισμού του ε.μ.ρ., ακολουθώντας τα βήματα Α και Β. Από τα παραδείγματα αυτά φαίνεται πόσο απλή είναι η μέθοδος.

Παράδειγμα 7.4.1. (Κανονική κατανομή - ε.μ.ρ.) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$, $n \geq 2$, ένα τυχαίο δείγμα από την κανονική κατανομή $\mathcal{N}(\mu, \sigma^2)$.
1η Περίπτωση: $\boldsymbol{\sigma^2}$ **γνωστό,** $\boldsymbol{\mu = \theta}$ **άγνωστο**, $\theta \in \Theta = \mathbb{R}$. Έχουμε $\mu_1(\theta) = \mathbb{E}_\theta X_1 = \theta$ (βήμα Α) και θέτουμε $\mu_1(\theta) = m_1$, δηλαδή $\theta = \bar{X}$ (βήμα Β). Άρα ο ε.μ.ρ. του θ είναι $\hat{\hat{\theta}} = \bar{X}$ που συμπίπτει με τον ε.μ.π. και τον αποδοτικό εκτιμητή.
2η Περίπτωση: $\boldsymbol{\mu}$ **γνωστό,** $\boldsymbol{\sigma^2 = \theta}$ **άγνωστο**, $\theta \in \Theta = (0, \infty)$. Έχουμε

$\mu_1(\theta) = \mathbb{E}_\theta X_1 = \mu$. Όμως το μ, ως γνωστό, δεν εξαρτάται από το θ, άρα η πρώτη ροπή είναι άχρηστη για την εκτίμηση του θ. Γι' αυτό το λόγο, χρησιμοποιούμε τη δεύτερη ροπή. Έχουμε

$$\mu_2(\theta) = \mathbb{E}_\theta X_1^2 = \mathbb{V}ar_\theta X_1 + (\mathbb{E}_\theta X_1)^2 = \theta + \mu^2.$$

Τα πράγματα είναι καλύτερα τώρα, γιατί η δεύτερη ροπή προέκυψε (μη σταθερή) συνάρτηση της άγνωστης παραμέτρου θ, όπως θα θέλαμε. Αυτό είναι το βήμα Α της διαδικασίας. Σύμφωνα με το βήμα Β, θέτουμε $\mu_2(\theta) = m_2$ δηλαδή $\theta + \mu^2 = \frac{1}{n}\sum\limits_{i=1}^{n} X_i^2$ και λύνουμε ως προς θ. Άρα ο ε.μ.ρ. είναι $\hat{\hat{\theta}} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i^2 - \mu^2$. Για τον $\hat{\hat{\theta}}$ παρατηρούμε ότι δεν είναι συνάρτηση της επαρκούς στατιστικής συνάρτησης $\sum\limits_{i=1}^{n}(X_i - \mu)^2$ (βλέπε Παράδειγμα 6.3.10) άρα είναι μη αποδεκτός με κριτήριο το ΜΤΣ. Επιπροσθέτως, μπορεί να πάρει και αρνητικές τιμές, αν και $\mathbb{P}_\theta(\hat{\hat{\theta}} < 0) = \mathbb{P}_\theta\left(\frac{1}{n}\sum\limits_{i=1}^{n} X_i^2 < \mu^2\right) \to 0$, καθώς $n \to \infty$, λόγω του ΑΝΜΑ $\frac{1}{n}\sum\limits_{i=1}^{n} X_i^2 \xrightarrow{\mathbb{P}_\theta} \mathbb{E}_\theta X_1^2 = \theta + \mu^2 > \mu^2$.

<u>3η Περίπτωση</u>: $\boldsymbol{\mu}$**,** $\boldsymbol{\sigma^2}$ **άγνωστα**, $\theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times \infty$. Έχουμε $\mu_1(\mu, \sigma^2) = \mathbb{E}_\theta X_1 = \mu$ και $\mu_2(\mu, \sigma^2) = \mathbb{E}_\theta X_1^2 = \mu^2 + \sigma^2$ (Βήμα Α). Θέτουμε λοιπόν

$$\mu_1(\mu, \sigma^2) = m_1 \ \text{ και } \mu_2(\mu, \sigma^2) = m_2,$$

δηλαδή

$$\mu = \bar{X} \text{ και } \mu^2 + \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

και λύνουμε το σύστημα ως προς μ και σ^2 (Βήμα Β). Η λύση του συστήματος, δηλαδή οι ε.μ.ρ., είναι $\hat{\hat{\mu}} = \bar{X}$, $\hat{\hat{\sigma}}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Παρατηρούμε ότι οι ε.μ.ρ. συμπίπτουν με τους ε.μ.π..

Παράδειγμα 7.4.2. (**Ομοιόμορφη κατανομή - ε.μ.ρ.**) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$, $n \geq 1$, ένα τυχαίο δείγμα από την ομοιόμορφη κατανομή $\mathcal{U}[0, \theta]$, $\theta \in \Theta = (0, \infty)$. Έχουμε $\mu_1(\theta) = \mathbb{E}_\theta X_1 = \frac{\theta}{2}$. Θέτουμε λοιπόν $\mu_1(\theta) = m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$, οπότε $\frac{\theta}{2} = \bar{X}$ και $\theta = 2\bar{X}$. Άρα ο ε.μ.ρ.

του θ είναι $\hat{\hat{\theta}} = 2\bar{X}$. Παρατηρούμε ότι για $n \geq 2$, ο εκτιμητής $\hat{\hat{\theta}}$ δεν είναι συνάρτηση της επαρκούς στατιστικής συνάρτησης $X_{(n)}$ και συνεπώς είναι μη αποδεκτός εκτιμητής με κριτήριο το ΜΤΣ.

Παράδειγμα 7.4.3. (Κατανομή Γάμμα – ε.μ.ρ.) Έστω $\underset{\sim}{X} = (X_1, \dots, X_n)$, $n \geq 2$, ένα τυχαίο δείγμα από την κατανομή Γάμμα $\mathcal{G}(\alpha, \beta)$ με α, β άγνωστα, οπότε $\theta = (\alpha, \beta) \in \Theta = (0, \infty) \times (0, \infty)$. Έχουμε $\mu_1(\theta) = \mathbb{E}_\theta X_1 = \alpha\beta$ και $\mu_2(\theta) = \mathbb{E}_\theta X_1^2 = \mathbb{V}ar_\theta X_1 + (\mathbb{E}_\theta X_1)^2 = \alpha\beta^2 + (\alpha\beta)^2$. Θέτουμε λοιπόν $\mu_1(\alpha, \beta) = m_1 = \frac{1}{n}\sum_{i=1}^n X_i$ και $\mu_2(\alpha, \beta) = m_2 = \frac{1}{n}\sum_{i=1}^n X_i^2$, οπότε προκύπτει το σύστημα ως προς α, β

$$\alpha\beta = \bar{X},$$
$$\alpha\beta^2 + \alpha^2\beta^2 = \frac{1}{n}\sum_{i=1}^n X_i^2.$$

Η λύση του συστήματος, δηλαδή οι ε.μ.ρ., είναι

$$\hat{\hat{\alpha}} = \frac{\bar{X}^2}{\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2} \quad \text{και} \quad \hat{\hat{\beta}} = \frac{\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2}{\bar{X}}.$$

Υπενθυμίζουμε ότι οι ε.μ.π. των α, β υπάρχουν μεν, αλλά όχι σε αναλυτική μορφή (Παράδειγμα 7.1.6). Σημειώνουμε επίσης ότι η τιμή του $\hat{\hat{\alpha}}$ μπορεί να χρησιμοποιηθεί ως αρχική τιμή στην αριθμητική επίλυση της εξίσωσης πιθανοφάνειας για την εύρεση του ε.μ.π. του α. Με αυτή την έννοια ο ε.μ.ρ. $\hat{\hat{\alpha}}$ μπορεί να θεωρηθεί ως «προκαταρτικός» εκτιμητής για τον υπολογισμό του ε.μ.π. $\hat{\alpha}$.

Κλείνουμε αυτήν την ενότητα αναφέροντας μερικές γενικές ιδιότητες της μεθόδου των ροπών.

1. Η μέθοδος είναι πολύ απλή και οι ε.μ.ρ. υπολογίζονται συνήθως σε αναλυτική μορφή.

2. Οι ε.μ.ρ. χρησιμοποιούνται συχνά ως «προκαταρτικοί» εκτιμητές για τον υπολογισμό καλύτερων εκτιμητών, π.χ. του ε.μ.π..

3. Οι ε.μ.ρ. είναι συνεπείς εκτιμητές.

4. Οι ε.μ.ρ. δεν είναι κατ' ανάγκη συναρτήσεις της (ελάχιστης) επαρκούς στατιστικής συνάρτησης.

5. Οι ε.μ.ρ. δεν παίρνουν κατ' ανάγκη τιμές μέσα στο σύνολο τιμών της παραμέτρου που εκτιμούν.

6. Η λύση του συστήματος των εξισώσεων του βήματος Β δεν είναι κατ' ανάγκη μοναδική.

7. Δυνητικά, στο βήμα Α μπορούν να επιλεγούν προς υπολογισμό οποιεσδήποτε r ροπές της κατανομής, αριθμός δηλαδή ίσος προς τον αριθμό των άγνωστων παραμέτρων $\theta_1, \ldots, \theta_r$, και όχι κατ' ανάγκη οι r πρώτες.

7.5 Ασκήσεις

7.1. Δίνεται μία παρατήρηση X από την κατανομή που φαίνεται στον πίνακα. Να βρεθεί ο ε.μ.π. του $\theta \in \Theta = \{0, 1, -1\}$.

	$X=1$	$X=2$	$X=3$
$\theta=0$	$1/6$	$1/6$	$4/6$
$\theta=1$	$2/6$	$2/6$	$2/6$
$\theta=-1$	$3/6$	$3/6$	0

7.2. Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την κατανομή με πυκνότητα $f_1(x;\theta)$, $x \in \mathcal{S}_1 \subset \Re$, $\theta \in \Theta = \Re$, όπου το $\mathcal{S}_1$ δεν εξαρτάται από το θ. Υποθέτουμε τα εξής.

α) $\frac{\partial^2}{\partial\theta^2} f_1(x;\theta) < 0$ για κάθε $\theta \in \Theta$ και $x \in \mathcal{S}_1$.

β) $\lim\limits_{\theta\to-\infty} \dfrac{\partial}{\partial\theta} f_1(x;\theta) = \alpha(x)$, όπου $\alpha(x) > 0$ ή $\alpha(x) = \infty$,
$\lim\limits_{\theta\to+\infty} \dfrac{\partial}{\partial\theta} f_1(x;\theta) = \beta(x)$, όπου $\beta(x) < 0$ ή $\beta(x) = -\infty$, για $x \in \mathcal{S}_1$.

Να δειχθεί ότι υπάρχει ο ε.μ.π. του θ, είναι μοναδικός και ικανοποιεί την εξίσωση πιθανοφάνειας.

7.3. Εάν $\underset{\sim}{X} = (X_1, \ldots, X_n)$ είναι ένα τυχαίο δείγμα από την λογιστική κατανομή με πυκνότητα

$$f_1(x;\theta) = \frac{e^{-(x-\theta)}}{[1+e^{-(x-\theta)}]^2},\ x \in \mathbb{R},\ \theta \in \Theta = \mathbb{R},$$

να δειχθεί ότι υπάρχει ο ε.μ.π. του θ και είναι μοναδικός. Να δοθεί η εξίσωση πιθανοφάνειας και να εξεταστούν η ισχυρή συνέπεια και η ασυμπτωτική κανονικότητα του ε.μ.π.

7.4. Εάν $X_1, \ldots, X_n$ είναι ανεξάρτητες παρατηρήσεις με κατανομές αντίστοιχα διωνυμικές $\mathcal{B}(\kappa_i, \theta)$, $\theta \in \Theta = [0, 1]$, να βρεθεί ο ε.μ.π. του θ και να εξεταστεί εάν είναι ασθενώς συνεπής, υποθέτοντας ότι $\sum_{i=1}^{n} \kappa_i \to \infty$ καθώς $n \to \infty$.

7.5. Να δειχθεί ότι ο ε.μ.ρ. του θ από την ομοιόμορφη κατανομή $\mathcal{U}(0, \theta)$ είναι ασθενώς συνεπής.

7.6. Εάν $\underset{\sim}{X} = (X_1, \ldots, X_n)$ είναι ένα τυχαίο δείγμα από την εκθετική κατανομή $\mathcal{E}(\theta)$, $\theta \in \Theta = (0, \infty)$, να βρεθεί ο ε.μ.π. της $\mathbb{P}_\theta(X_1 > c)$, όπου c δοθείσα (γνωστή) σταθερά.

7.7. Το ύψος σε έναν πληθυσμό ανδρών ακολουθεί την κανονική κατανομή $\mathcal{N}(\mu, \sigma^2)$. Με δεδομένα τα ύψη 10 ανδρών $1.72, 1.77$, $1.78, 1.67$, $1.82, 1.69$, 1.75, 1.70, 1.80, 1.68, να υπολογιστεί η εκτίμηση μέγιστης πιθανοφάνειας του ποσοστού των ανδρών (στο συνολικό πληθυσμό) που είναι ψηλότεροι από 1.73.

7.8. Ο χρόνος ζωής (σε ώρες) ηλεκτρικής λυχνίας ακολουθεί την κατανομή με πυκνότητα

$$f_1(x;\theta) = \frac{1}{\theta^2} x e^{-\frac{x}{\theta}},\ x > 0,\ \theta \in \Theta = (0, \infty).$$

Με δεδομένα τους χρόνους ζωής 10 λυχνιών $540h$, $620h$, $570h$, $552h$, $605h$, $580h$, $590h$, $557h$, $563h$, $547h$, να υπολογισθεί η εκτίμηση μέγιστης πιθανοφάνειας του μέσου χρόνου ζωής των λυχνιών.

7.9. Τα τυχαία σφάλματα $X_1, \ldots, X_n$ σε n ανεξάρτητες μετρήσεις κάποιας «ποσότητας» ακολουθούν κανονική κατανομή $\mathcal{N}(\theta, 4)$ με θ άγνωστο. Να βρεθεί ο ε.μ.π. του θ χρησιμοποιώντας ως δεδομένα μόνον το πλήθος των παρατηρήσεων X_i, $i = 1, \ldots, n$ που είναι θετικές.

7.10. Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ είναι ένα τυχαίο δείγμα από την ομοιόμορφη κατανομή $\mathcal{U}[-\theta, \theta]$, $\theta \in \Theta = (0, \infty)$.

α. Να δειχθεί ότι ο ε.μ.π. του θ είναι $\hat{\theta} = \max\{|X_1|, \ldots, |X_n|\}$.

β. Να βρεθεί η κατανομή του $\hat{\theta}$.

γ. Να δειχθεί ότι ο $\hat{\theta}$ είναι ασθενώς συνεπής.

δ. Θεωρούμε την κλάση εκτιμητών του θ, $\mathcal{C} = \{c\hat{\theta} : c \text{ θετική σταθερά}\}$. Να βρεθεί ο εκτιμητής της κλάσης $\mathcal{C}$ με το μικρότερο ΜΤΣ.

ε. Να δειχθεί ότι ο ε.μ.π. $\hat{\theta}$ είναι μη αποδεκτός με κριτήριο το ΜΤΣ.

στ. Να βρεθεί ο ε.μ.ρ. του θ, χρησιμοποιώντας την ροπή δεύτερης τάξης.

7.11. Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ είναι ένα τυχαίο δείγμα από την κατανομή με πυκνότητα $f_1(x; \mu, \sigma) = \frac{1}{\sigma} e^{-(x-\mu)/\sigma}$, $x \geqslant \mu$, $\mu \in \Re$, $\sigma > 0$.

α. Με σ γνωστό και μ άγνωστο, να βρεθεί ο ε.μ.ρ. του μ και να δειχθεί ότι είναι ασθενώς συνεπής.

β. Με μ γνωστό και σ άγνωστο, να βρεθεί ο ε.μ.ρ. του σ και να δειχθεί ότι είναι ασθενώς συνεπής.

γ. Με μ και σ άγνωστα, να βρεθούν ο ε.μ.ρ. των μ και σ.

7.12. Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ είναι ένα τυχαίο δείγμα από την κατανομή με πυκνότητα $f_1(x; \theta) = \dfrac{2}{\theta^2 y^3}$, $y \geqslant \theta$, $\theta \in \Theta = (0, \infty)$.

α. Να δειχθεί ότι ο ε.μ.π. του θ είναι $\hat{\theta} = \max\{X_1, \ldots, X_n\}$.

β. Να βρεθεί η κατανομή του $\hat{\theta}$.

γ. Να δειχθεί ότι ο ε.μ.π. $\hat{\theta}$ είναι ασθενώς συνεπής εκτιμητής του θ.

δ. Να βρεθεί ο ε.μ.π. της μέσης τιμής της κατανομής $f_1(x;\theta)$.

ε. Να βρεθεί ο ε.μ.ρ. του θ και να δειχθεί ότι είναι μη αποδεκτός με κριτήριο το ΜΤΣ για $n \geqslant 2$.

7.13. Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την Weibull κατανομή με πυκνότητα $f_1(x;\theta) = \theta x^{\theta-1} e^{-x^{\theta}}$, $x > 0$, $\theta \in \Theta = (0, \infty)$. Να δειχθεί ότι υπάρχει ο ε.μ.π. του θ είναι μοναδικός και ικανοποιεί την εξίσωση πιθανοφάνειας.

7.14. Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την κατανομή Pareto με πυκνότητα $f_1(x;\theta) = \dfrac{a\mu^a}{x^{a+1}}$, $x \geqslant \mu$, $\mu > 0$, $a > 0$. Να βρεθούν οι ε.μ.π. και ε.μ.ρ. των αγνώστων παραμέτρων στις εξής περιπτώσεις: (α) a γνωστό και μ άγνωστο, (β) μ γνωστό και a άγνωστο, (γ) a και μ άγνωστα. Επί πλέον στην (α) περίπτωση να δειχθεί ότι ο ε.μ.π. του μ είναι ασθενώς συνεπής ενώ στη (β) περίπτωση να βρεθεί ο ε.μ.π. του $\frac{1}{a}$ και να εξεταστεί η ισχυρή συνέπεια και ασυμπτωτική κανονικότητα του.

Βιβλιογραφία

1. Ηλιόπουλος, Γ. (2013). *Βασικές Μέθοδοι Εκτίμησης Παραμέτρων.* Εκδόσεις Αθ. Σταμούλης.

2. Bahadur, R.R. (1958). Examples of inconsistency of maximum likelihood estimates. *Sankhya*, **20**, 207–210.

3. Bahadur, R.R. (1960). Stochastic comparison of tests. *Ann. Math. Statist.*, **31**, 276–295.

4. Bai, Z.D. and Fu, J.C. (1987). On the maximum-likelihood estimator for the location parameter of a Cauchy distribution. *Can. J. Statist.*, **15**, 137–146.

5. Berk, R.H.(1967). Zehna, P.W.: Invariance of maximum likelihood estimators, Review #1922. *Mathematical Reviews*, **33**, 342–343.

6. Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton Univ. Press.

7. Doob, J.L. (1934). Probability and statistics. *Trans. Amer. Math. Soc.*, **36**, 759–775.

8. Doob, J.L. (1936). Statistical estimation. *Trans. Amer. Math. Soc.*, **39**, 410–421.

9. Efron, B. (1982). Maximum likelihood and decision theory. *Ann. Statist.*, **10**, 340–356.

10. Fisher, R.A. (1912). On the absolute criterion for fitting frequency curves. *Messenger of Mathematics*, **41**, 155.

11. Fisher, R.A. (1922). On the mathematical foundations of theoretical statistics. *Philos. Trans. Roy. Soc. London Ser. A*, **222**, 309–368.

12. Fisher, R.A. (1925). Theory of statistical estimation. *Proc. Camb. Phil. Soc.*, **22**, 700–725.

13. Fisher, R.A. (1934). Two new properties of mathematical likelihood. *Proc. Roy. Soc. Ser. A*, **144**, 285–307.

14. Hoadley, B. (1971). Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case. *Ann. Math Statist.*, **42**, 1977–1991.

15. Hotelling, H. (1930). The consistency and ultimate distribution of optimum statistics. *Trans. Amer. Math. Soc.*, **32**, 847–859.

16. Huber, P. (1967). The behavior of the maximum likelihood estimates under nonstandard conditions. *In Proceedings of the Fifth*

Berkeley Symposium on Mathematical Statistics and Probability, **1**, 221–233.

17. Inagaki, N. (1973). Asymptotic relations between the likelihood estimating function and the maximum likelihood estimator. *Ann. Inst. Statist. Math.*, **265**, 1–26.

18. Kourouklis, S. (1987). On the strong consistency of a solution to the likelihood equation. *Stat. Prob. Letters*, **5**, 23–27.

19. Kullback, S. and Leibler, R.A. (1951). On information and sufficiency. *Ann. Math Statist.*, **22**, 79–86.

20. Kulldorff, G. (1957). On the conditions for consistency and asymptotic efficiency of maximum likelihood estimates. *Skand Aktuartidskr.*, **40**, 129–144.

21. Le Cam, L. (1956). On the Asymptotic Theory of Estimation and Testing Hypotheses. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, 129–156. University of California Press, Berkeley, Calif. .

22. Le Cam, L. (1970). On the assumptions used to prove asymptotic normality of maximum likelihood estimators. *Ann. Math Statist.*, **41**, 802–828.

23. Le Cam, L. (1979). *Maximum Likelihood: An Introduction.* Lecture Notes in Statist. **18**, Springer, Berlin.

24. Le Cam, L. (1990). Maximum likelihood: an introduction. *Internat. Statist. Rev.*, **58**, 153–171.

25. Lehmann, E.L. and Casella, G. (1998). *Theory of Point Estimation.* Springer–Verlag, New York.

26. Neyman, J. and Scott, E.L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, **16**, 1–32.

27. Quenouille, M.H. (1956). Notes on Bias in Estimation. *Biometrika*, **43**, 353–360.

28. Scholz, F.W. (1980). Towards a unified definition of maximum likelihood. *Canad. J. Statist.*, **8**, 193–203.

29. Scholz, F.W. (2006). Maximum Likelihood Estimation. *Encyclopedia of Statistical Sciences*, bf 7.

30. Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

31. Wald, A. (1949). A Note on the Consistency of the Maximum Likelihood Estimate. *Ann. Math. Statist.*, **20**, 595–601.

32. Wasserman, L. (2004). *All of statistics: A concise course in statistical inference*. Springer, New York.

33. Yatracos, Y.G. (2015). MLE's bias pathology, model updated MLE, and Wallace's minimum message length method. *Information Theory, IEEE Transactions on*, **61**, 1426–1431.

34. Zehna, P.W. (1966). Invariance of maximum likelihood estimation. *Ann. Math Statist.*, **37**, 744.

Κεφάλαιο 8

Εκτιμητές Bayes και εκτιμητές minimax

Σε αυτό το κεφάλαιο μελετάμε την κατασκευή εκτιμητών χρησιμοποιώντας ως κριτήριο επιλογής το κριτήριο Bayes ή το κριτήριο minimax. Όπως εν συντομία αναφέρθηκε στην Ενότητα 3.2 και τα δύο κριτήρια συνοψίζουν τη συμπεριφορά ενός εκτιμητή σε μια σταθερά (τον κίνδυνο Bayes και τη μέγιστη μέση ζημία, αντίστοιχα). Για σύγκριση, υπενθυμίζουμε ότι το κριτήριο του ΜΤΣ αξιολογεί τον εκτιμητή $T(\underset{\sim}{X})$ του $g(\theta)$ μέσω της συνάρτησης $\mathbb{E}_\theta(T(\underset{\sim}{X}) - g(\theta))^2$, $\theta \in \Theta$. Το πλεονέκτημα των δύο κριτηρίων είναι ότι είτε με το ένα είτε με το άλλο μπορούμε πάντοτε να συγκρίνουμε δύο εκτιμητές, συγκρίνοντας τις αντίστοιχες σταθερές, αλλά και να βρούμε, συχνά, βέλτιστο εκτιμητή μεταξύ *όλων* των εκτιμητών. Σε αντίθεση, με το ΜΤΣ δύο εκτιμητές δεν είναι κατ' ανάγκη συγκρίσιμοι (π.χ., βλέπε το Παράδειγμα 4.1.5) και για να βρούμε βέλτιστο εκτιμητή αναγκαζόμαστε να περιοριστούμε σε συγκεκριμένη κλάση εκτιμητών (π.χ., στην κλάση των αμερόληπτων εκτιμητών).

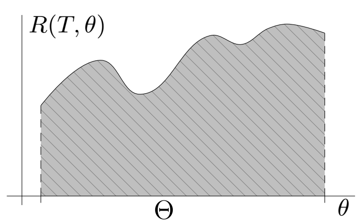
Από την άλλη πλευρά, και τα δύο κριτήρια είναι πιο ασθενή από το ΜΤΣ και δεν παρέχουν πλήρη εικόνα για τον εκτιμητή, αφού αποτυπώνουν τη συμπεριφορά του πολύ αδρά, σε μία σταθερά. Για αυτόν το λόγο, συχνά αναζητούνται εκτιμητές που ικανοποιούν και τα δύο κριτήρια συγχρόνως ή ακόμη και κάποια τρίτη ιδιότητα, όπως αποδεκτικότητα ή βελτίωση άλλου γνωστού εκτιμητή.

8.1 Εκτιμητές Bayes

Στην Ενότητα 3.1, για τον εκτιμητή $T(\underset{\sim}{X})$ της άγνωστης τιμής $g(\theta)$ ορίστηκε η *συνάρτηση κινδύνου*

$$R(T,\theta) = \mathbb{E}_\theta L\big(T(\underset{\sim}{X}),\theta\big), \quad \theta \in \Theta,$$

όπου $L(t,\theta)$ είναι η *συνάρτηση ζημίας* , η οποία ικανοποιεί τις συνθήκες $L(t,\theta) \geqslant 0$ και $L(g(\theta),\theta) = 0$. Μια ερμηνεία της $L(t,\theta)$ είναι ότι παριστάνει την «ζημία», εάν εκτιμήσουμε το $g(\theta)$ με την τιμή t. Άρα, η συνάρτηση κινδύνου $R(T,\theta)$ παριστάνει τη μέση ζημία, εάν χρησιμοποιηθεί ο $T(\underset{\sim}{X})$, για να εκτιμήσουμε το $g(\theta)$. Μεταξύ δύο εκτιμητών καλύτερος είναι αυτός που έχει τη μικρότερη συνάρτηση κινδύνου – μέση ζημία – για κάθε $\theta \in \Theta$. Ιδανικά θα επιθυμούσαμε να βρούμε έναν εκτιμητή $T(\underset{\sim}{X})$ με την ελάχιστη συνάρτηση κινδύνου $R(T,\theta)$ για κάθε $\theta \in \Theta$, μεταξύ όλων των εκτιμητών. Αυτό όμως είναι αδύνατο λόγω της πληθώρας των διαθέσιμων εκτιμητών. Μάλιστα, στην περίπτωση που η συνάρτηση ζημίας είναι το τετραγωνικό σφάλμα, δηλαδή $L(t,\theta) = (t - g(\theta))^2$ – οπότε η συνάρτηση κινδύνου είναι το ΜΤΣ – δείξαμε στη σχέση (4.4) ότι ένας τέτοιος εκτιμητής θα εκτιμούσε την άγνωστη τιμή $g(\theta)$ χωρίς σφάλμα, πράγμα αδύνατο εκτός κάποιων τετριμμένων καταστάσεων, όπως αυτή του Παραδείγματος 4.1.6. Μία λύση στο πρόβλημα αυτό είναι να αναζητήσουμε εκτιμητή $T(\underset{\sim}{X})$ που ελαχιστοποιεί το εμβαδόν κάτω από την καμπύλη $R(T,\theta)$, $\theta \in \Theta$, (βλέπε Σχήμα 8.1) δηλαδή το ολοκλήρωμα $\int_\Theta R(T,\theta)\,\mathrm{d}\theta$ (ή το άθροισμα – σειρά $\sum\limits_{\theta\in\Theta} R(T,\theta)$, εάν το Θ είναι αριθμήσιμο). Το σκεπτικό εδώ είναι ότι εάν



Σχήμα 8.1: $\int_\Theta R(T,\theta)\,\mathrm{d}\theta$

το ολοκλήρωμα (εμβαδόν) αυτό γίνεται ελάχιστο τότε αναμένεται ότι η συνάρτηση κινδύνου $R(T, \theta)$ δεν θα έχει «πολύ μεγάλη» τιμή σε «μεγάλα» διαστήματα τιμών του $\theta \in \Theta$ και συνεπώς ο εκτιμητής $T(\underset{\sim}{X})$ μπορεί να θεωρηθεί ως ικανοποιητικός εκτιμητής του $g(\theta)$. Πιο γενικά, μπορούμε να προκαθορίσουμε μία συνάρτηση $\pi(\theta)$, $\theta \in \Theta$, με τις εξής δύο ιδιότητες

(α) $\pi(\theta) \geq 0$, $\theta \in \Theta$

(β) $\int_\Theta \pi(\theta)\, \mathrm{d}\theta = 1$ (ή $\sum\limits_{\theta \in \Theta} \pi(\theta) = 1$ για Θ αριθμήσιμο)

και να αναζητήσουμε εκτιμητή $T(\underset{\sim}{X})$ που ελαχιστοποιεί το ολοκλήρωμα $\int_\Theta R(T, \theta)\pi(\theta)\, \mathrm{d}\theta$ (ή $\sum\limits_{\theta \in \Theta} R(T, \theta)\pi(\theta)$ για Θ αριθμήσιμο). Η συνάρτηση $\pi(\theta)$, $\theta \in \Theta$ λέγεται συνάρτηση βαρύτητας ή *εκ των προτέρων πυκνότητα* ή καταχρηστικά *εκ των προτέρων κατανομή* του θ και έχει, λόγω των (α) και (β), όντως τις ιδιότητες πυκνότητας κατανομής. Τα παραπάνω οδηγούν στον εξής ορισμό του εκτιμητή Bayes.

Ορισμός 8.1.1. *Ο εκτιμητής $T_*(\underset{\sim}{X})$ ονομάζεται εκτιμητής Bayes του $g(\theta)$ ως προς τη συνάρτηση ζημίας $L(t, \theta)$ και την εκ των προτέρων κατανομή $\pi(\theta)$ εάν*

$$\int_\Theta R(T_*, \theta)\pi(\theta)\, \mathrm{d}\theta \leq \int_\Theta R(T, \theta)\pi(\theta)\, \mathrm{d}\theta$$

$$\left(\text{ή } \sum_{\theta \in \Theta} R(T_*, \theta)\pi(\theta) \leq \sum_{\theta \in \Theta} R(T, \theta)\pi(\theta) \text{ για } \Theta \text{ αριθμήσιμο}\right)$$

για κάθε εκτιμητή $T(\underset{\sim}{X})$.

Για κάθε εκτιμητή $T(\underset{\sim}{X})$, η παράσταση $\int_\Theta R(T, \theta)\pi(\theta)\, \mathrm{d}\theta = BR(T)$ αναφέρεται ως *κίνδυνος Bayes* του εκτιμητή $T(\underset{\sim}{X})$. Επομένως ο εκτιμητής Bayes $T_*(\underset{\sim}{X})$ έχει τον ελάχιστο κίνδυνο Bayes μεταξύ όλων των εκτιμητών.

Οι δύο ονομασίες της συνάρτησης $\pi(\theta)$, $\theta \in \Theta$ υποδηλώνουν δύο διαφορετικές ερμηνείες της. Ως συνάρτηση βαρύτητας, η αριθμητική τιμή $\pi(\theta)$ εκφράζει τη *σημαντικότητα* του σημείου θ σε σχέση με τα υπόλοιπα στοιχεία του Θ. Αν π.χ. $\Theta = \{\theta_1, \theta_2, \theta_3\}$ και ορίσουμε $\pi(\theta_1) = \frac{5}{10}$,

$\pi(\theta_2) = \frac{2}{10}$ και $\pi(\theta_3) = \frac{3}{10}$, τότε θεωρούμε ότι το θ_1 είναι πιο σημαντικό από τα θ_2 και θ_3. Από την άλλη πλευρά, οι σχέσεις $\pi(\theta_1) = \pi(\theta_2) = \pi(\theta_3) = \frac{1}{3}$ δηλώνουν ότι κανένα από τα τρία σημεία δεν είναι πιο σημαντικό από τα υπόλοιπα. Με αυτή την παραδοχή για τη συνάρτηση $\pi(\theta)$, $\theta \in \Theta$, η αλγεβρική ερμηνεία της παράστασης $\int_\Theta R(T,\theta)\pi(\theta)\,\mathrm{d}\theta$ ή $\sum_{\theta\in\Theta} R(T,\theta)\pi(\theta)$ είναι αυτή του σταθμισμένου μέσου όρου των τιμών της συνάρτησης $R(T,\theta)$, $\theta \in \Theta$ με συντελεστές βαρύτητας (στάθμες) τις τιμές $\pi(\theta)$, $\theta \in \Theta$.

Ως εκ των προτέρων κατανομή, η αριθμητική τιμή $\pi(\theta)$ εκφράζει την (συνήθως υποκειμενική) αντίληψη για το πόσο *πιθανόν* είναι η αντίστοιχη τιμή θ να είναι η (αληθής) τιμή της άγνωστης παραμέτρου, εκ των προτέρων – πριν δηλαδή τη συλλογή των δεδομένων $\underset{\sim}{X} = (X_1, \ldots, X_n)$. Στο παραπάνω παράδειγμα με $\Theta = \{\theta_1, \theta_2, \theta_3\}$, ορίζοντας $\pi(\theta_1) = \frac{5}{10}$ εκφράζουμε την αντίληψη πως η πιθανότητα η άγνωστη παράμετρος να έχει τιμή θ_1 είναι 50%. Ουσιαστικά, με αυτήν τη θεώρηση, η άγνωστη παράμετρος είναι *τυχαία μεταβλητή* με σύνολο τιμών Θ και πυκνότητα $\pi(\theta)$, $\theta \in \Theta$. Τότε η πιθανοτική ερμηνεία της παράστασης $\int_\Theta R(T,\theta)\pi(\theta)\,\mathrm{d}\theta$ ή $\sum_{\theta\in\Theta} R(T,\theta)\pi(\theta)$ είναι αυτή της μέσης τιμής της τυχαίας μεταβλητής $R(T,\boldsymbol{\theta})$, όπου $\boldsymbol{\theta}$ έχει πυκνότητα $\pi(\theta)$.

Η αντίληψη («φιλοσοφία») ότι η άγνωστη παράμετρος είναι τυχαία μεταβλητή (και όχι σταθερά) αποτελεί το θεμέλιο λίθο της Μπεϋζιανής Στατιστικής (Bayesian Statistics). Οι Μπεϋζιανοί Στατιστικοί χρησιμοποιούν τις εκ των προτέρων πληροφορίες για την παράμετρο που ενσωματώνονται και ποσοτικοποιούνται στην εκ των προτέρων κατανομή και τις πληροφορίες που παρέχουν τα δεδομένα $\underset{\sim}{X} = (X_1, \ldots, X_n)$, για να επικαιροποιήσουν (update) τις γνώσεις τους για την παράμετρο και τελικά να εξαγάγουν συμπεράσματα για αυτήν. Ως ενδεικτική βιβλιογραφία Μπεϋζιανής Στατιστικής αναφέρουμε τους Lee (2004), Gelman et al. (2004), Bolstad (2007) και Hoff (2009).

Οι κλασικοί (μη Μπεϋζιανοί) Στατιστικοί αντιμετωπίζουν την άγνωστη παράμετρο ως σταθερά και χρησιμοποιούν τη συνάρτηση $\pi(\theta)$, $\theta \in \Theta$

στην καλύτερη περίπτωση ως συνάρτηση βαρύτητας, ενώ γενικότερα ως ένα μαθηματικό εργαλείο χωρίς συγκεκριμένη ερμηνεία με σκοπό την κατασκευή (κλάσεων) εκτιμητών. Αυτήν την κλασική προσέγγιση έχουν υιοθετήσει αυτές οι σημειώσεις.

Θα μελετήσουμε στη συνέχεια τη διαδικασία εύρεσης του εκτιμητή Bayes $T_*(\underset{\sim}{X})$. (Σε ό,τι ακολουθεί για διακριτές κατανομές αντικαθιστούμε το $\int$ με $\sum$.)

Έστω $f(\underset{\sim}{x};\theta)$ η πυκνότητα του δείγματος $\underset{\sim}{X} = (X_1, \ldots, X_n)$. Θέτουμε

$$f(\underset{\sim}{x}) = \int_\Theta f(\underset{\sim}{x};\theta)\pi(\theta)\,\mathrm{d}\theta \tag{8.1}$$

και

$$\pi(\theta \mid \underset{\sim}{x}) = \frac{f(\underset{\sim}{x};\theta)\pi(\theta)}{f(\underset{\sim}{x})} \tag{8.2}$$

και παρατηρούμε τα εξής. Είδαμε προηγουμένως ότι τυπικά ή ουσιαστικά η άγνωστη παράμετρος θ μπορεί να θεωρηθεί ως τυχαία μεταβλητή $\boldsymbol{\theta}$ με σύνολο τιμών το Θ και πυκνότητα $\pi(\theta)$. Τότε η πυκνότητα $f(\underset{\sim}{x};\theta)$ ως συνάρτηση του $\underset{\sim}{x}$, για δεδομένο $\theta \in \Theta$, μπορεί να θεωρηθεί ως η δεσμευμένη πυκνότητα του $\underset{\sim}{X}$, δοθέντος ότι $\boldsymbol{\theta} = \theta$. Συνεπώς το γινόμενο $f(\underset{\sim}{x};\theta)\pi(\theta)$ είναι η από κοινού πυκνότητα των δεδομένων $\underset{\sim}{X}$ και της τυχαίας μεταβλητής $\boldsymbol{\theta}$. Περαιτέρω, η $f(\underset{\sim}{x})$ στην (8.1) είναι η περιθώρια πυκνότητα του $\underset{\sim}{X}$ και η $\pi(\theta \mid \underset{\sim}{x})$ στην (8.2) είναι η δεσμευμένη πυκνότητα του $\boldsymbol{\theta}$, δοθέντος ότι $\underset{\sim}{X} = \underset{\sim}{x}$ (δια τούτο και ο συμβολισμός $\pi(\theta \mid \underset{\sim}{x})$). Η συνάρτηση $\pi(\theta \mid \underset{\sim}{x})$ με μεταβλητή το θ και σταθερό $\underset{\sim}{x}$ λέγεται *εκ των υστέρων πυκνότητα* ή καταχρηστικά *εκ των υστέρων κατανομή* του $\boldsymbol{\theta}$ και έχει όντως τις ιδιότητες πυκνότητας κατανομής, δηλαδή $\pi(\theta \mid \underset{\sim}{x}) \geq 0$, για κάθε $\theta \in \Theta$ και $\int_\Theta \pi(\theta \mid \underset{\sim}{x})\,\mathrm{d}\theta = 1$, όπως προκύπτει εύκολα από τις (8.1) και (8.2). Η εκ των υστέρων κατανομή $\pi(\theta \mid \underset{\sim}{x})$ μπορεί να θεωρηθεί ότι συνοψίζει τις εκ των υστέρων πληροφορίες για την άγνωστη παράμετρο, δηλαδή πληροφορίες μετά τη συλλογή των δεδομένων $\underset{\sim}{x}$. Με άλλα λόγια, οι αρχικές πληροφορίες για το θ, που περιέχονται στην εκ των προτέρων κατανομή $\pi(\theta)$, μετά τη συλλογή των δεδομένων, τροποποιούνται – επικαιροποιούνται και εκφράζονται από την εκ των υστέρων κατανομή $\pi(\theta \mid \underset{\sim}{x})$. Η σχέση

(8.2), λόγω της (8.1), γράφεται

$$\pi(\theta \mid \underset{\sim}{x}) = \frac{f(\underset{\sim}{x};\theta)\pi(\theta)}{\int_\Theta f(\underset{\sim}{x};\theta)\pi(\theta)\,\mathrm{d}\theta}. \tag{8.3}$$

Με μια προσεκτική ματιά, η (8.3) είναι το συνεχές ανάλογο του τύπου Bayes για τον υπολογισμό της εκ των υστέρων πιθανότητας ενός ενδεχομένου (βλέπε Πρόταση 1.2.1(β)). Η ονομασία των εκτιμητών Bayes οφείλεται στον γενικευμένο τύπο Bayes της σχέσης (8.3) και τον ρόλο που παίζει η $\pi(\theta \mid \underset{\sim}{x})$ στην εύρεση αυτών των εκτιμητών.

Η επόμενη πρόταση δείχνει πως μπορεί να υπολογιστεί ο εκτιμητής Bayes μέσω της $\pi(\theta \mid \underset{\sim}{x})$.

Πρόταση 8.1.1. *Για $\underset{\sim}{X} = \underset{\sim}{x}$, ο εκτιμητής Bayes $T_*(\underset{\sim}{X})$ του $g(\theta)$ ως προς τη συνάρτηση ζημίας $L(t,\theta)$ και την εκ των προτέρων κατανομή $\pi(\theta)$ έχει τιμή $T_*(\underset{\sim}{x}) = t_*$, όπου t_* είναι η τιμή του t (αν υπάρχει) που ελαχιστοποιεί τη συνάρτηση $h(t) = \int_\Theta L(t,\theta)\pi(\theta \mid \underset{\sim}{x})\,\mathrm{d}\theta$.*

Απόδειξη. Για κάθε εκτιμητή $T(\underset{\sim}{X})$ του $g(\theta)$ έχουμε

$$R(T,\theta) = \mathbb{E}_\theta L\big(T(\underset{\sim}{X}),\theta\big) = \int_{\mathfrak{X}} L\big(T(\underset{\sim}{x}),\theta\big) f(\underset{\sim}{x};\theta)\,\mathrm{d}\underset{\sim}{x},$$

όπου $\mathfrak{X}$ είναι το σύνολο τιμών του $\underset{\sim}{X}$.
Τότε ο κίνδυνος Bayes του $T(\underset{\sim}{X})$ είναι

$$\begin{aligned}
BR(T) &= \int_\Theta R(T,\theta)\pi(\theta)\,\mathrm{d}\theta \\
&= \int_\Theta \int_{\mathfrak{X}} L\big(T(\underset{\sim}{x}),\theta\big) f(\underset{\sim}{x};\theta)\,\mathrm{d}\underset{\sim}{x}\,\pi(\theta)\,\mathrm{d}\theta \\
&= \int_{\mathfrak{X}} \int_\Theta L\big(T(\underset{\sim}{x}),\theta\big) f(\underset{\sim}{x};\theta)\pi(\theta)\,\mathrm{d}\theta\,\mathrm{d}\underset{\sim}{x} \\
&= \int_{\mathfrak{X}} \int_\Theta L\big(T(\underset{\sim}{x}),\theta\big) \pi(\theta \mid \underset{\sim}{x}) f(\underset{\sim}{x})\,\mathrm{d}\theta\,\mathrm{d}\underset{\sim}{x} \\
&= \int_{\mathfrak{X}} f(\underset{\sim}{x}) \left[\int_\Theta L\big(T(\underset{\sim}{x}),\theta\big)\pi(\theta \mid \underset{\sim}{x})\,\mathrm{d}\theta\right] \mathrm{d}\underset{\sim}{x}.
\end{aligned}$$

Ειδικά για τον εκτιμητή $T_*(\underset{\sim}{X})$ προκύπτει ότι

$$\begin{aligned}
BR(T_*) &= \int_{\mathfrak{X}} f(\underset{\sim}{x}) \left[\int_{\Theta} L\big(T_*(\underset{\sim}{x}), \theta\big) \pi(\theta \mid \underset{\sim}{x})\, \mathrm{d}\theta \right] \mathrm{d}\underset{\sim}{x} \\
&= \int_{\mathfrak{X}} f(\underset{\sim}{x}) \left[\int_{\Theta} L(t_*, \theta) \pi(\theta \mid \underset{\sim}{x})\, \mathrm{d}\theta \right] \mathrm{d}\underset{\sim}{x} \\
&\le \int_{\mathfrak{X}} f(\underset{\sim}{x}) \int_{\Theta} L\big(T(\underset{\sim}{x}), \theta\big) \pi(\theta \mid \underset{\sim}{x})\, \mathrm{d}\theta\, \mathrm{d}\underset{\sim}{x} \quad \text{(από τον ορισμό του } t_*\text{)} \\
&= BR(T).
\end{aligned}$$

Επομένως, ο εκτιμητής T_* έχει τον ελάχιστο κίνδυνο Bayes μεταξύ όλων των εκτιμητών και συνεπώς είναι, εξ ορισμού, εκτιμητής Bayes. □

Η Πρόταση 8.1.1 αφορά τυχούσα συνάρτηση ζημίας $L(t, \theta)$. Στη σημαντική περίπτωση, που η συνάρτηση ζημίας για την εκτίμηση του $g(\theta)$ είναι το *τετραγωνικό σφάλμα*, $L(t, \theta) = \big(t - g(\theta)\big)^2$, ο εκτιμητής Bayes του $g(\theta)$ υπολογίζεται εύκολα και είναι η μέση τιμή $\mathbb{E}g(Y)$, όπου Y είναι τυχαία μεταβλητή με κατανομή την εκ των υστέρων κατανομή $\pi(\theta \mid \underset{\sim}{x})$. Αυτό το αποτέλεσμα αποδεικνύεται χρησιμοποιώντας τις επόμενες προτάσεις.

Πρόταση 8.1.2. *Εάν Z είναι μία τυχαία μεταβλητή με $\mathbb{V}ar Z < \infty$, τότε η συνάρτηση $\mathbb{E}(Z - t)^2$, $t \in \mathbb{R}$, ελαχιστοποιείται για $t = \mathbb{E}Z$.*

Απόδειξη. Από την Πρόταση 1.4.1(β), έχουμε $\mathbb{E}(Z - t)^2 = \mathbb{V}ar Z + (\mathbb{E}Z - t)^2 \ge \mathbb{V}ar Z$. Επομένως το ελάχιστο ως προς t είναι $\mathbb{V}ar Z$ και επιτυγχάνεται, όταν $(\mathbb{E}Z - t)^2 = 0$, δηλαδή $t = \mathbb{E}Z$. □

Πρόταση 8.1.3. *Έστω ότι η συνάρτηση ζημίας για την εκτίμηση του $g(\theta)$ είναι το τετραγωνικό σφάλμα $L(t, \theta) = \big(t - g(\theta)\big)^2$. Τότε για $\underset{\sim}{X} = \underset{\sim}{x}$, ο εκτιμητής Bayes $T_*(\underset{\sim}{X})$ του $g(\theta)$ έχει τιμή $T_*(\underset{\sim}{x}) = \mathbb{E}g(Y)$, όπου Y είναι τυχαία μεταβλητή με κατανομή την εκ των υστέρων κατανομή $\pi(\theta \mid \underset{\sim}{x})$, για την οποία υποθέτουμε ότι $\mathbb{V}ar\, g(Y) < \infty$. Ειδικά, για $g(\theta) = \theta$, η εκτίμηση Bayes $T_*(\underset{\sim}{x})$ συμπίπτει με τη μέση τιμή της εκ των υστέρων κατανομής.*

Απόδειξη. Από την Πρόταση 8.1.1, η τιμή του εκτιμητή Bayes $T_*(\underset{\sim}{X})$ για $\underset{\sim}{X} = \underset{\sim}{x}$, $T_*(\underset{\sim}{x})$, είναι η τιμή του t που ελαχιστοποιεί τη συνάρτηση

$$\begin{aligned} h(t) &= \int_\Theta L(t,\theta)\pi(\theta \mid \underset{\sim}{x})\,\mathrm{d}\theta \\ &= \int_\Theta (g(\theta)-t)^2\pi(\theta \mid \underset{\sim}{x})\,\mathrm{d}\theta = \mathbb{E}(g(Y)-t)^2. \end{aligned}$$

Από την Πρόταση 8.1.2 όμως, με $Z = g(Y)$, η συνάρτηση $h(t)$ ελαχιστοποιείται για $t = \mathbb{E}g(Y)$. Επομένως $T_*(\underset{\sim}{x}) = \mathbb{E}g(Y)$. Τέλος, για $g(\theta) = \theta$, έχουμε $T_*(\underset{\sim}{x}) = \mathbb{E}Y$. □

Θα μελετήσουμε μία ακόμη ειδική περίπτωση της Πρότασης 8.1.1, θεωρώντας ως συνάρτηση ζημίας για την εκτίμηση του $g(\theta)$ *το απόλυτο σφάλμα* , $L(t,\theta) = |t - g(\theta)|$. Κατ' αντιστοιχία με την Πρόταση 8.1.3, θα δείξουμε ότι εκτιμητής Bayes του $g(\theta)$ είναι η διάμεσος της $g(Y)$, όπου Y είναι τυχαία μεταβλητή με κατανομή την εκ των υστέρων κατανομή $\pi(\theta|\underset{\sim}{x})$. Ως διάμεσος της τυχαίας μεταβλητής Z ορίζεται μια σταθερά m που ικανοποιεί τις σχέσεις $\mathbb{P}(Z \geqslant m) \geqslant 1/2$ και $\mathbb{P}(Z \leqslant m) \geqslant 1/2$. Η διάμεσος της Z δεν είναι κατ' ανάγκη μοναδική. Αν όμως η Z έχει συνεχή κατανομή τότε η διάμεσος m είναι η μοναδική λύση της εξίσωσης

$$\mathbb{P}(Z \leqslant m) = 1/2 \tag{8.4}$$

Αρχικά, ως ανάλογο της Πρότασης 8.1.2 ισχύει το ακόλουθο αποτέλεσμα.

Πρόταση 8.1.4. *Εάν Z είναι μία τυχαία μεταβλητή με $\mathbb{E}|Z| < \infty$, τότε η συνάρτηση $\mathbb{E}|Z-t|$, $t \in \Re$ ελαχιστοποιείται για $t = m$, όπου m είναι (οποιαδήποτε) διάμεσος της Z.*

Απόδειξη. Για λόγους ευκολίας δίνουμε την απόδειξη στην περίπτωση που η Z έχει συνεχή κατανομή με πυκνότητα $v(\xi)$. Για τη γενική περίπτωση παραπέμπουμε τον αναγνώστη στους Bickel and Doksum (1977, σελ. 54).

Έστω, αρχικά, ότι $t > m$. Τότε έχουμε

$$\begin{aligned}
&\mathbb{E}|Z-t| - \mathbb{E}|Z-m| = \int_{-\infty}^{+\infty} |\xi - t| v(\xi)\,\mathrm{d}\xi - \int_{-\infty}^{+\infty} |\xi - m| v(\xi)\,\mathrm{d}\xi \\
&= \int_{-\infty}^{m} (t-m) v(\xi)\,\mathrm{d}\xi + \int_{m}^{t} (t+m-2\xi) v(\xi)\,\mathrm{d}\xi + \int_{t}^{+\infty} (m-t) v(\xi)\,\mathrm{d}\xi \\
&\geqslant \int_{-\infty}^{m} (t-m) v(\xi)\,\mathrm{d}\xi + \int_{m}^{t} (m-t) v(\xi)\,\mathrm{d}\xi + \int_{t}^{+\infty} (m-t) v(\xi)\,\mathrm{d}\xi \\
&(\text{επειδή } t+m-2\xi \geqslant m-t \text{ για } \xi \in [m,t]) \\
&= \int_{-\infty}^{m} (t-m) v(\xi)\,\mathrm{d}\xi + \int_{m}^{+\infty} (m-t) v(\xi)\,\mathrm{d}\xi \\
&= (t-m)\left\{\mathbb{P}(Z \leqslant m) - \mathbb{P}(Z \geqslant m)\right\} = 0,
\end{aligned}$$

αφού λόγω της (8.4) και της συνεχούς κατανομής της Z, $\mathbb{P}(Z \geqslant m) = 1/2 = \mathbb{P}(Z \leqslant m)$. Συμπεραίνουμε λοιπόν ότι $\mathbb{E}|Z-m| \leqslant |Z-t|$. Ανάλογη είναι η απόδειξη και στην περίπτωση $t < m$. □

Πρόταση 8.1.5. *Έστω ότι η συνάρτηση ζημίας για την εκτίμηση του $g(\theta)$ είναι το απόλυτο σφάλμα $L(t,\theta) = |t - g(\theta)|$. Τότε για $\underset{\sim}{X} = \underset{\sim}{x}$, ο εκτιμητής Bayes $T_*(\underset{\sim}{X})$ του $g(\theta)$ έχει τιμή $T_*(\underset{\sim}{x}) = m_{g(Y)}$, όπου $m_{g(Y)}$ είναι μία διάμεσος της $g(Y)$ και Y είναι τυχαία μεταβλητή με κατανομή την εκ των υστέρων κατανομή $\pi(\theta \mid \underset{\sim}{x})$, για την οποία υποθέτουμε ότι $\mathbb{E}|g(Y)| < \infty$. Ειδικά, για $g(\theta) = \theta$, η εκτίμηση Bayes $T_*(\underset{\sim}{x})$ είναι μία διάμεσος της εκ των υστέρων κατανομής.*

Απόδειξη. Από την Πρόταση 8.1.1, η τιμή του εκτιμητή Bayes $T_*(\underset{\sim}{X})$ για $\underset{\sim}{X} = \underset{\sim}{x}$, $T_*(\underset{\sim}{x})$, είναι η τιμή του t που ελαχιστοποιεί τη συνάρτηση

$$\begin{aligned}
h(t) &= \int_{\Theta} L(t,\theta)\pi(\theta \mid \underset{\sim}{x})\,\mathrm{d}\theta \\
&= \int_{\Theta} |g(\theta) - t| \pi(\theta \mid \underset{\sim}{x})\,\mathrm{d}\theta = \mathbb{E}|g(Y) - t|.
\end{aligned}$$

Από την Πρόταση 8.1.4 όμως, με $Z = g(Y)$, η συνάρτηση $h(t)$ ελαχιστοποιείται για $t = m_{g(Y)}$. Επομένως $T_*(\underset{\sim}{x}) = m_{g(Y)}$. Τέλος, για $g(\theta) = \theta$, έχουμε $T_*(\underset{\sim}{x}) = m_Y$. □

Δίνουμε στη συνέχεια μερικά παραδείγματα υπολογισμού εκτιμητών Bayes.

Παράδειγμα 8.1.1. (κατανομή Bernoulli – εκτιμητής Bayes της πιθανότητας «επιτυχίας» ως προς το τετραγωνικό σφάλμα) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την κατανομή Bernoulli $\mathcal{B}(1,\theta)$ με πυκνότητα $f_1(x;\theta) = \theta^x(1-\theta)^{1-x}$, $x = 0, 1$, $\theta \in \Theta = (0,1)$. Θα βρούμε τον εκτιμητή Bayes του θ (δηλαδή, εδώ, $g(\theta) = \theta$), όταν η συνάρτηση ζημίας είναι το τετραγωνικό σφάλμα $L(t,\theta) = (t-\theta)^2$ και η εκ των προτέρων κατανομή του θ είναι Βήτα $\mathcal{B}eta(\alpha,\beta)$, δηλαδή

$$\pi(\theta) = \frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1},\ 0 < \theta < 1.$$

Έχουμε

$$\begin{aligned} f(\underset{\sim}{x};\theta) &= \prod_{i=1}^n f_1(x_i;\theta) \\ &= \prod_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}. \end{aligned}$$

Επομένως η εκ των υστέρων κατανομή του θ είναι

$$\begin{aligned} \pi(\theta \mid \underset{\sim}{x}) &= \frac{f(\underset{\sim}{x};\theta)\pi(\theta)}{f(\underset{\sim}{x})} \\ &= \frac{1}{f(\underset{\sim}{x})B(\alpha,\beta)}\theta^{\sum_{i=1}^n x_i+\alpha-1}(1-\theta)^{n-\sum_{i=1}^n x_i+\beta-1},\ 0 < \theta < 1. \end{aligned}$$

Συγκρίνοντας την πυκνότητα $\pi(\theta \mid \underset{\sim}{x})$ με τη γενική μορφή της πυκνότητας μιας κατανομής Βήτα (βλέπε Ενότητα 1.6) προκύπτει ότι η $\pi(\theta \mid \underset{\sim}{x})$ είναι Βήτα $\mathcal{B}eta(\sum_{i=1}^n x_i + \alpha, n - \sum_{i=1}^n x_i + \beta)$ (η σταθερά $f(\underset{\sim}{x})B(\alpha,\beta)$ κατ' ανάγκην είναι $B(\sum_{i=1}^n x_i + \alpha, n - \sum_{i=1}^n x_i + \beta)$ και δεν χρειάζεται να υπολογισθεί το $f(\underset{\sim}{x}) = \int_\Theta f(\underset{\sim}{x};\theta)\pi(\theta)\,\mathrm{d}\theta$, για να διαπιστωθεί αυτό). Επειδή $g(\theta) = \theta$ και η συνάρτηση ζημίας είναι το τετραγωνικό σφάλμα, από την Πρόταση 8.1.3 η τιμή του εκτιμητή Bayes $T_*(\underset{\sim}{X})$ για $\underset{\sim}{X} = \underset{\sim}{x}$, είναι $T_*(\underset{\sim}{x}) = \mathbb{E}g(Y) = \mathbb{E}Y$,

όπου η τυχαία μεταβλητή Y έχει κατανομή $\mathcal{B}eta(\sum\limits_{i=1}^{n} x_i + \alpha, n - \sum\limits_{i=1}^{n} x_i + \beta)$.
Άρα,

$$T_*(\underset{\sim}{x}) = \frac{\sum_{i=1}^{n} x_i + \alpha}{\sum_{i=1}^{n} x_i + \alpha + n - \sum_{i=1}^{n} x_i + \beta} = \frac{\sum_{i=1}^{n} x_i + \alpha}{n + \alpha + \beta},$$

επειδή η μέση τιμή της κατανομής $\mathcal{B}eta(\alpha_1, \beta_1)$ είναι $\frac{\alpha_1}{\alpha_1+\beta_1}$. Επομένως ο εκτιμητής Bayes είναι

$$T_*(\underset{\sim}{X}) = \frac{\sum_{i=1}^{n} X_i + \alpha}{n + \alpha + \beta}.$$

Επισημαίνουμε ότι, επειδή τα α, β είναι αυθαίρετες (θετικές) σταθερές, στην πραγματικότητα κατασκευάστηκε όχι μόνον ένας εκτιμητής, αλλά μία κλάση εκτιμητών Bayes για το θ.

Παράδειγμα 8.1.2. (Συνέχεια του Παραδείγματος 8.1.1) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την κατανομή Bernoulli $\mathcal{B}(1, \theta)$, $\theta \in \Theta = (0, 1)$. Θα βρούμε τον εκτιμητή Bayes του θ, όταν η συνάρτηση ζημίας είναι $L(t, \theta) = \frac{(t-\theta)^2}{\theta(1-\theta)}$ και η εκ των προτέρων κατανομή του θ είναι ομοιόμορφη $\mathcal{U}(0, 1)$, δηλαδή $\pi(\theta) = 1$, $0 < \theta < 1$.
Έχουμε, όπως στο Παράδειγμα 8.1.1,

$$f(\underset{\sim}{x}; \theta) = \theta^{\sum\limits_{i=1}^{n} x_i} (1-\theta)^{n - \sum\limits_{i=1}^{n} x_i}$$

και

$$\pi(\theta \mid \underset{\sim}{x}) = \frac{f(\underset{\sim}{x}; \theta)\pi(\theta)}{f(\underset{\sim}{x})} = \frac{\theta^{\sum\limits_{i=1}^{n} x_i} (1-\theta)^{n - \sum\limits_{i=1}^{n} x_i}}{f(\underset{\sim}{x})}, \ 0 < \theta < 1.$$

Αναγνωρίζουμε την εκ των υστέρων κατανομή ως κατανομή $\mathcal{B}eta(\sum\limits_{i=1}^{n} x_i + 1, n - \sum\limits_{i=1}^{n} x_i + 1)$, οπότε $f(\underset{\sim}{x}) = B(\sum\limits_{i=1}^{n} x_i + 1, n - \sum\limits_{i=1}^{n} x_i + 1)$. Για τον υπολογισμό του εκτιμητή Bayes, βάσει της Πρότασης 8.1.1, ελαχιστοποιούμε τη συνάρτηση

$$\begin{aligned}
h(t) &= \int_\Theta L(t,\theta)\pi(\theta \mid \underset{\sim}{x})\,\mathrm{d}\theta \\
&= \int_0^1 \frac{(t-\theta)^2}{\theta(1-\theta)}\frac{1}{f(\underset{\sim}{x})}\theta^{\sum\limits_{i=1}^n x_i}(1-\theta)^{n-\sum\limits_{i=1}^n x_i}\,\mathrm{d}\theta \\
&= \frac{1}{f(\underset{\sim}{x})}B(\sum_{i=1}^n x_i, n-\sum_{i=1}^n x_i)\int_0^1 (t-\theta)^2 \frac{\theta^{\sum\limits_{i=1}^n x_i - 1}(1-\theta)^{n-\sum\limits_{i=1}^n x_i - 1}}{B(\sum\limits_{i=1}^n x_i, n-\sum\limits_{i=1}^n x_i)}\,\mathrm{d}\theta \\
&= \frac{1}{f(\underset{\sim}{x})}B(\sum_{i=1}^n x_i, n-\sum_{i=1}^n x_i)\,\mathbb{E}(Y-t)^2,
\end{aligned}$$

όπου Y είναι τυχαία μεταβλητή με κατανομή Βήτα $\mathcal{B}eta(\sum\limits_{i=1}^n x_i, n-\sum\limits_{i=1}^n x_i)$, εφ' όσον $0<\sum\limits_{i=1}^n X_i<n$. Από την Πρόταση 8.1.2, η $h(t)$ ελαχιστοποιείται για

$$t = \mathbb{E}Y = \frac{\sum_{i=1}^n x_i}{\sum_{i=1}^n x_i + n - \sum_{i=1}^n x_i} = \frac{\sum_{i=1}^n x_i}{n} = \bar{x}.$$

Για $\sum\limits_{i=1}^n x_i = 0$ ή $\sum\limits_{i=1}^n x_i = n$ είναι $h(t)=\infty$ για κάθε t, εκτός αν $t=0$ ή 1 αντίστοιχα, οπότε η $h(t)$ ελαχιστοποιείται για $t=\bar{x}=0$ ή 1, αντίστοιχα. Άρα ο εκτιμητής Bayes του θ είναι $T_*(\underset{\sim}{X}) = \bar{X}$.

Παράδειγμα 8.1.3. (κατανομή Poisson – εκτιμητής Bayes της μέσης τιμής) Έστω $\underset{\sim}{X}=(X_1,\ldots,X_n)$ ένα τυχαίο δείγμα από την κατανομή Poisson $\mathcal{P}(\theta)$. Με συνάρτηση ζημίας το τετραγωνικό σφάλμα $L(t,\theta)=(t-\theta)^2$ και εκ των προτέρων κατανομή την Γάμμα $\mathcal{G}(\alpha,\beta)$, δηλαδή $\pi(\theta)=\frac{1}{\Gamma(\alpha)\beta^\alpha}\theta^{\alpha-1}e^{-\theta/\beta}$, $\theta>0$, θα βρούμε τον εκτιμητή Bayes του θ.
Έχουμε

$$f(\underset{\sim}{x};\theta) = \prod_{i=1}^n e^{-\theta}\frac{\theta^{x_i}}{x_i!} = e^{-n\theta}\frac{\theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}$$

και

$$\pi(\theta \mid \underset{\sim}{x}) = \frac{f(\underset{\sim}{x};\theta)\pi(\theta)}{f(\underset{\sim}{x})} = \frac{1}{f(\underset{\sim}{x})\prod\limits_{i=1}^n x_i!\Gamma(\alpha)\beta^\alpha}\theta^{\sum\limits_{i=1}^n x_i+\alpha-1}e^{-(n+1/\beta)\theta},\ \theta>0.$$

Συγκρίνοντας την πυκνότητα $\pi(\theta \mid \underset{\sim}{x})$ με τη γενική μορφή της πυκνότητας κατανομής Γάμμα, προκύπτει ότι η $\pi(\theta \mid \underset{\sim}{x})$ είναι Γάμμα $\mathcal{G}(\sum\limits_{i=1}^{n} x_i + \alpha, \frac{1}{n+1/\beta})$ (η σταθερά $f(\underset{\sim}{x}) \prod_{i=1}^{n} x_i! \Gamma(\alpha) \beta^{\alpha}$ είναι κατ' ανάγκην $\Gamma(\sum\limits_{i=1}^{n} x_i + \alpha) \cdot (n+1/\beta)^{-(\sum\limits_{i=1}^{n} x_i + \alpha)}$ και δεν χρειάζεται να υπολογιστεί το $f(\underset{\sim}{x})$ για να διαπιστωθεί αυτό). Τότε από την Πρόταση 8.1.3, επειδή $g(\theta) = \theta$, ο εκτιμητής Bayes του θ για $\underset{\sim}{X} = \underset{\sim}{x}$ είναι $T_*(\underset{\sim}{x}) = \mathbb{E}g(Y) = \mathbb{E}Y$, όπου η τυχαία μεταβλητή Y έχει κατανομή $\pi(\theta \mid \underset{\sim}{x})$. Άρα $T_*(\underset{\sim}{x}) = \frac{\sum_{i=1}^{n} x_i + \alpha}{n+1/\beta}$. Επομένως ο εκτιμητής Bayes είναι

$$T_*(\underset{\sim}{X}) = \frac{\sum_{i=1}^{n} X_i + \alpha}{n + 1/\beta}.$$

Σημειώνουμε ότι, όπως στο Παράδειγμα 8.1.1, επειδή τα α και β είναι αυθαίρετες σταθερές, κατ' ουσίαν κατασκευάστηκε μία κλάση εκτιμητών Bayes για το θ.

Παράδειγμα 8.1.4. (Κανονική κατανομή με γνωστή διασπορά – εκτιμητής Bayes της μέσης τιμής) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την κανονική κατανομή $\mathcal{N}(\theta, 1)$, $\theta \in \Theta = \Re$. Με συνάρτηση ζημίας το τετραγωνικό σφάλμα $L(t, \theta) = (t - \theta)^2$ και εκ των προτέρων κατανομή του θ την κανονική $\mathcal{N}(0, 1)$, θα βρούμε τον εκτιμητή Bayes του θ.

Έχουμε

$$f(\underset{\sim}{x}; \theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_i - \theta)^2} = (2\pi)^{-n/2} e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i - \theta)^2}$$

και

$$\pi(\theta) = \frac{1}{\sqrt{2\pi}} e^{-\theta^2/2},$$

οπότε

$$\pi(\theta \mid \underset{\sim}{x}) = \frac{f(\underset{\sim}{x}; \theta)\pi(\theta)}{f(\underset{\sim}{x})} = \frac{(2\pi)^{-\frac{n+1}{2}}}{f(\underset{\sim}{x})} e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i - \theta)^2 - \frac{\theta^2}{2}}, \ \theta \in \mathbb{R}.$$

Ο εκθέτης γίνεται

$$
\begin{aligned}
\sum_{i=1}^{n}(x_i-\theta)^2+\theta^2 &= \sum_{i=1}^{n}x_i^2+(n+1)\theta^2-2\theta\sum_{i=1}^{n}x_i \\
&= \sum_{i=1}^{n}x_i^2+(n+1)\Big\{\theta^2-2\theta\frac{\sum_{i=1}^n x_i}{n+1}+\Big(\frac{\sum_{i=1}^n x_i}{n+1}\Big)^2-\Big(\frac{\sum_{i=1}^n x_i}{n+1}\Big)^2\Big\} \\
&= \sum_{i=1}^{n}x_i^2-\frac{(\sum_{i=1}^n x_i)^2}{n+1}+(n+1)\Big(\theta-\frac{\sum_{i=1}^n x_i}{n+1}\Big)^2.
\end{aligned}
$$

Επομένως,

$$
\pi(\theta \mid \underset{\sim}{x}) = \frac{(2\pi)^{-\frac{n+1}{2}}e^{-\frac{1}{2}\sum_{i=1}^n x_i^2+\frac{(\sum_{i=1}^n x_i)^2}{2(n+1)}}}{f(\underset{\sim}{x})}\,e^{-\frac{n+1}{2}\left(\theta-\frac{\sum_{i=1}^n x_i}{n+1}\right)^2},\ \theta\in\mathbb{R}.
$$

Συγκρίνοντας τώρα την πυκνότητα $\pi(\theta \mid \underset{\sim}{x})$ με τη γενική μορφή πυκνότητας κανονικής κατανομής, συμπεραίνουμε ότι η $\pi(\theta \mid \underset{\sim}{x})$ είναι κανονική $\mathcal{N}\big(\frac{\sum x_i}{n+1},\frac{1}{n+1}\big)$ (η σταθερά πριν το $e^{-\frac{n+1}{2}\left(\theta-\frac{\sum x_i}{n+1}\right)^2}$ κατ' ανάγκην είναι $\sqrt{n+1}/\sqrt{2\pi}$ και δεν χρειάζεται να υπολογιστεί το $f(\underset{\sim}{x})$ για να διαπιστωθεί αυτό). Τότε από την Πρόταση 8.1.3 ο εκτιμητής Bayes του θ για $\underset{\sim}{X}=\underset{\sim}{x}$ είναι $T_*(\underset{\sim}{x})=\mathbb{E}g(Y)=\mathbb{E}Y$, όπου η τυχαία μεταβλητή Y έχει κατανομή $\pi(\theta \mid \underset{\sim}{x})$. Άρα $T_*(\underset{\sim}{x})=\dfrac{\sum_{i=1}^n x_i}{n+1}$. Επομένως ο εκτιμητής Bayes είναι

$$
T_*(\underset{\sim}{X})=\frac{\sum_{i=1}^n X_i}{n+1}.
$$

Σημειώνουμε επίσης ότι αν θεωρήσουμε ως συνάρτηση ζημίας το απόλυτο σφάλμα, τότε από την Πρόταση 8.1.5 ο εκτιμητής Bayes του θ, $T_{1*}(\underset{\sim}{X})$, είναι η διάμεσος της εκ των υστέρων κατανομής $\mathcal{N}\left(\frac{\sum_{i=1}^n X_i}{n+1},\frac{1}{n+1}\right)$ που λόγω συμμετρίας της κανονικής κατανομής συμπίπτει με τη μέση τιμή και επομένως $T_{1*}(\underset{\sim}{X})=T_*(\underset{\sim}{X})$. □

Παρατήρηση 8.1.1. Σε όλα τα παραπάνω Παραδείγματα 8.1.1–8.1.4, η εκ των υστέρων κατανομή ανήκει στην ίδια οικογένεια κατανομών με την εκ των προτέρων κατανομή. Μία οικογένεια κατανομών $\mathcal{F}$ επί του παραμετρικού χώρου Θ με την ιδιότητα αν η $\pi(\theta) \in \mathcal{F}$ επιλεγεί ως εκ των προτέρων κατανομή του θ τότε και η εκ των υστέρων κατανομή $\pi(\theta|\underset{\sim}{x}) \in \mathcal{F}$, ονομάζεται συζυγής οικογένεια εκ των προτέρων κατανομών για την οικογένεια κατανομών του δείγματος $\underset{\sim}{X}$, $\{f(\underset{\sim}{x};\theta) : \theta \in \Theta\}$. Έτσι, από το Παράδειγμα 8.1.3 προκύπτει ότι η οικογένεια των κατανομών Γάμμα, $\mathcal{F} = \{G(\alpha,\beta) : \alpha > 0, \beta > 0\}$ είναι συζυγής για την οικογένεια των κατανομών Poisson, $\mathcal{P}(\theta)$ με $\theta \in \Theta = (0,\infty)$.

Παράδειγμα 8.1.5. (μετατοπισμένη εκθετική – εκτιμητής Bayes) Έστω $\underset{\sim}{X} = (X_1,\ldots,X_n)$ ένα τυχαίο δείγμα από την κατανομή με πυκνότητα $f_1(x;\theta) = e^{-(x-\theta)}$, $x \geqslant \theta$, $\theta \in \Theta = (0,\infty)$. Θα βρούμε τον εκτιμητή Bayes του θ όταν η συνάρτηση ζημίας είναι το τετραγωνικό σφάλμα ή το απόλυτο σφάλμα και η εκ των προτέρων κατανομή του θ είναι εκθετική $\mathcal{E}(1)$, δηλαδή $\pi(\theta) = e^{-\theta}$, $\theta > 0$.

Έχουμε $f(\underset{\sim}{x};\theta) = \prod_{i=1}^{n} e^{-(x_i-\theta)} = e^{-\sum_{i=1}^n x_i + n\theta}$, αν $x_i \geqslant \theta$ για κάθε $i = 1,\ldots,n$ και $f(\underset{\sim}{x};\theta) = 0$ αλλού ή ισοδύναμα

$$f(\underset{\sim}{x};\theta) = \begin{cases} e^{-\sum_{i=1}^n x_i + n\theta} & , \quad x_{(1)} \geqslant \theta \\ 0 & , \quad x_{(1)} < \theta, \end{cases}$$

όπου $x_{(1)} = \min\{x_1,\ldots,x_n\}$. Η εκ των υστέρων κατανομή του θ είναι

$$\pi(\theta|\underset{\sim}{x}) = \frac{f(\underset{\sim}{x};\theta)\pi(\theta)}{f(\underset{\sim}{x})} = \begin{cases} \dfrac{e^{-\sum_{i=1}^n x_i + (n-1)\theta}}{f(\underset{\sim}{x})} & , \quad 0 < \theta \leqslant x_{(1)} \\ 0 & , \quad \text{αλλού}, \end{cases}$$

με $f(\underset{\sim}{x}) = \int_0^{x_{(1)}} e^{-\sum_{i=1}^n x_i + (n-1)\theta} d\theta = e^{-\sum_{i=1}^n x_i} \int_0^{x_{(1)}} e^{(n-1)\theta} d\theta$, οπότε

$$f(\underset{\sim}{x}) = e^{-\sum_{i=1}^n x_i} \, \frac{e^{(n-1)x_{(1)}} - 1}{n-1}.$$

Επομένως έχουμε

$$\pi(\theta|\underset{\sim}{x}) = \begin{cases} \dfrac{(n-1)e^{(n-1)\theta}}{e^{(n-1)x_{(1)}} - 1} & , \quad 0 < \theta \leqslant x_{(1)} \\ 0 & , \quad \text{αλλού}. \end{cases}$$

Σημειώνουμε εδώ ότι, επειδή $\mathbb{P}_\theta(X_i \geqslant \theta) = 1$ και $\theta > 0$, θεωρούμε $x_i > 0$, οπότε και $x_{(1)} > 0$. Η $\pi(\theta|\underset{\sim}{x})$ δεν είναι πυκνότητα κάποιας γνωστής κατανομής, όμως το γεγονός αυτό δεν επηρεάζει την εύρεση του εκτιμητή Bayes. Από την Πρόταση 8.1.3 με $g(\theta) = \theta$, η εκτίμηση Bayes είναι η μέση τιμή της εκ των υστέρων κατανομής, δηλαδή

$$\begin{aligned} T_*(\underset{\sim}{x}) &= \int_0^{x_{(1)}} \theta\pi(\theta|\underset{\sim}{x})d\theta = \frac{n-1}{e^{(n-1)x_{(1)}}-1}\int_0^{x_{(1)}} \theta e^{(n-1)\theta}d\theta \\ &= \frac{e^{(n-1)x_{(1)}}\left[x_{(1)} - \frac{1}{n-1}\right] + \frac{1}{n-1}}{e^{(n-1)x_{(1)}}-1}. \end{aligned}$$

Για σύγκριση αναφέρουμε ότι ο ΑΟΕΔ εκτιμητής του θ είναι $X_{(1)} - \frac{1}{n}$, ο ε.μ.π. είναι $\hat{\theta} = X_{(1)}$ και ο ε.μ.ρ., $\tilde{\theta} = \bar{X} - 1$. Χρησιμοποιώντας την Πρόταση 7.2.5 και την κατανομή του $X_{(1)}$, βλέπε Παράδειγμα 6.3.7, είναι εύκολο να δειχθεί ότι και οι τρεις αυτοί εκτιμητές είναι (ασθενώς) συνεπείς εκτιμητές του θ, ειδικά λοιπόν $X_{(1)} \xrightarrow{\mathbb{P}_\theta} \theta$. Τότε γράφοντας τον εκτιμητή Bayes στη μορφή

$$T_*(\underset{\sim}{X}) = \frac{e^{(n-1)X_{(1)}}}{e^{(n-1)X_{(1)}}-1}\left\{X_{(1)} - \frac{1}{n-1}\right\} + \frac{1}{(n-1)\left\{e^{(n-1)X_{(1)}}-1\right\}}$$

και εφαρμόζοντας την Πρόταση 1.10.5(i), προκύπτει ότι $T_*(\underset{\sim}{X}) \xrightarrow{\mathbb{P}_\theta} \theta$, δηλαδή και ο $T_*(\underset{\sim}{X})$ είναι (ασθενώς) συνεπής εκτιμητής του θ.

Όσον αφορά το απόλυτο σφάλμα, από την Πρόταση 8.1.5, η εκτίμηση Bayes είναι η διάμεσος της εκ των υστέρων κατανομής $\pi(\theta|\underset{\sim}{x})$, δηλαδή η λύση της εξίσωσης

$$\int_0^m \pi(\theta|\underset{\sim}{x})d\theta = 1/2, \tag{8.5}$$

βλέπε (8.4). Αντικαθιστώντας την $\pi(\theta|\underset{\sim}{x})$, από την (8.5) διαδοχικά παίρνουμε

$$\begin{aligned} \int_0^m (n-1)e^{(n-1)\theta}d\theta &= \frac{1}{2}\{e^{(n-1)x_{(1)}}-1\}, \\ e^{(n-1)m} - 1 &= \frac{1}{2}\left\{e^{(n-1)x_{(1)}}-1\right\}, \\ m &= \frac{1}{n-1}\ln\left\{e^{(n-1)x_{(1)}}+1\right\} - \frac{\ln 2}{n-1}. \end{aligned}$$

Επομένως ο εκτιμητής Bayes ως προς το απόλυτο σφάλμα είναι

$$T_{1*}(\underset{\sim}{X}) = \frac{1}{n-1} \ln \left\{ e^{(n-1)X_{(1)}} + 1 \right\} - \frac{\ln 2}{n-1}$$

που είναι επίσης (ασθενώς) συνεπής εκτιμητής του θ.

Αναφέρουμε στη συνέχεια μερικές γενικές ιδιότητες των εκτιμητών Bayes, χωρίς όμως να εμβαθύνουμε σε αποδείξεις.

1. Ο εκτιμητής Bayes είναι συνάρτηση της (ελάχιστης) επαρκούς στατιστικής συνάρτησης (Άσκηση 8.1).

2. Εάν ο εκτιμητής Bayes είναι μοναδικός, τότε είναι αποδεκτός εκτιμητής ως προς το κριτήριο της μέσης ζημίας (Άσκηση 8.2).

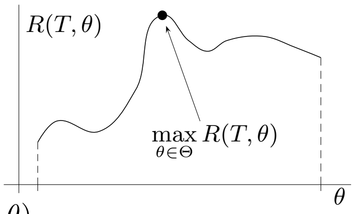
3. Ο εκτιμητής Bayes είναι ασθενώς συνεπής εκτιμητής υπό γενικές συνθήκες και για ευρείες κλάσεις συναρτήσεων ζημίας και εκ των προτέρων κατανομών.

4. Κατά κανόνα, καθώς το μέγεθος δείγματος αυξάνει, η επίδραση της εκ των προτέρων κατανομής εξασθενεί και τα δεδομένα $\underset{\sim}{X} = (X_1, \ldots, X_n)$ καθορίζουν, κυρίως, τη συμπεριφορά του εκτιμητή Bayes. Αυτή η ιδιότητα είναι φανερή στα Παραδείγματα 8.1.1 και 8.1.3. Στο πρώτο εξ αυτών, ο εκτιμητής Bayes του ποσοστού θ γράφεται στη μορφή

$$\begin{aligned} T_*(\underset{\sim}{X}) &= \frac{\sum_{i=1}^n X_i + \alpha}{n + 1/\beta} \\ &= \frac{n}{n+\alpha+\beta} \frac{1}{n} \sum_{i=1}^n X_i + \frac{\alpha}{n+\alpha+\beta}, \end{aligned}$$

οπότε $T_*(\underset{\sim}{X}) \approx \bar{X}$ καθώς $n \to \infty$. Συνεπώς, όταν υπάρχει μεγάλο πλήθος διαθέσιμων δεδομένων, η εκ των προτέρων κατανομή ουσιαστικά δεν συμμετέχει στην συμπεριφορά του $T_*(\underset{\sim}{X})$. Το γεγονός αυτό μόνον θετικά μπορεί να σχολιαστεί: όποια και αν είναι τα α και β, ακόμη και η εσφαλμένη επιλογή τους (άρα εσφαλμένη υποκειμενική αντίληψη για την παράμετρο) δεν πρόκειται να επηρεάσει σημαντικά την εκτίμηση του ποσοστού θ.

5. Ένα κοινό γνώρισμα των εκτιμητών Bayes με τους ε.μ.π. είναι ότι και οι δύο εκτιμητές εξαρτώνται από τα δεδομένα μέσω της συνάρτησης



Σχήμα 8.2: $\max\limits_{\theta\in\Theta} R(T,\theta)$

πιθανοφάνειας $L(\theta \mid \underset{\sim}{x}) = f(\underset{\sim}{x};\theta)$. Πράγματι, γνωρίζουμε ότι ο ε.μ.π. του θ προκύπτει από τη μεγιστοποίηση της $L(\theta \mid \underset{\sim}{x})$ ως προς θ, ενώ ο εκτιμητής Bayes καθορίζεται από την εκ των υστέρων κατανομή

$$\pi(\theta \mid \underset{\sim}{x}) = \frac{f(\underset{\sim}{x};\theta)\pi(\theta)}{f(\underset{\sim}{x})} = \frac{L(\theta \mid \underset{\sim}{x})\pi(\theta)}{f(\underset{\sim}{x})} ,$$

που περιέχει ως συνιστώσα της την $L(\theta \mid \underset{\sim}{x})$.

Η εκ των υστέρων κατανομή υπό γενικές συνθήκες (βλέπε Schervish (1995), Κεφάλαιο 7), καθώς $n \to \infty$ έχει κατά προσέγγιση κανονική κατανομή με μέση τιμή $\hat{\theta}$, όπου $\hat{\theta}$ είναι ο ε.μ.π. του θ. Περαιτέρω, αν η συνάρτηση ζημίας για την εκτίμηση του θ είναι το τετραγωνικό σφάλμα, τότε ο εκτιμητής Bayes του θ είναι η μέση τιμή της εκ των υστέρων κατανομής (Πρόταση 8.1.3). Συνεπώς, ο εκτιμητής Bayes είναι κατά προσέγγιση ίσος προς τον ε.μ.π. .

8.2 Εκτιμητές minimax

Μία άλλη λύση στο πρόβλημα μη ύπαρξης εκτιμητή $T(\underset{\sim}{X})$ με ελάχιστη συνάρτηση κινδύνου $R(T,\theta)$ για κάθε $\theta \in \Theta$ είναι η αναζήτηση εκτιμητή $T(\underset{\sim}{X})$, που ελαχιστοποιεί τη μέγιστη τιμή της συνάρτησης κινδύνου, $\max_{\theta\in\Theta} R(T,\theta)$, δηλαδή την μέγιστη (ως προς θ) μέση ζημία. Μία τέτοια θεώρηση εκφράζει έναν «απαισιόδοξο» στατιστικό, ο οποίος προσπαθεί να προστατεύσει τον εαυτό του από το «χειρότερο που μπορεί να συμβεί», δηλαδή η άγνωστη τιμή του θ να είναι αυτή που αντιστοιχεί στη μέγιστη μέση ζημία $\max_{\theta\in\Theta} R(T,\theta)$, βλέπε Σχήμα 8.2. Τα παραπάνω οδηγούν στον εξής ορισμό του εκτιμητή minimax.

Ορισμός 8.2.1. *Ο εκτιμητής $T_*(\underset{\sim}{X})$ ονομάζεται εκτιμητής minimax του $g(\theta)$ ως προς τη συνάρτηση ζημίας $L(t,\theta)$, εάν*

$$\sup_{\theta\in\Theta} R(T_*,\theta) \le \sup_{\theta\in\Theta} R(T,\theta)$$

για κάθε εκτιμητή $T(\underset{\sim}{X})$.

Η επόμενη πρόταση δίνει μία τεχνική κατασκευής εκτιμητή minimax της τιμής $g(\theta)$ ως προς τη συνάρτηση ζημίας $L(t,\theta)$. Η τεχνική αυτή απαιτεί ο εκτιμητής να είναι εκτιμητής Bayes με σταθερή συνάρτηση κινδύνου.

Πρόταση 8.2.1. *Εάν ο εκτιμητής $T_*(\underset{\sim}{X})$ είναι εκτιμητής Bayes του $g(\theta)$ ως προς τη συνάρτηση ζημίας $L(t,\theta)$ και την εκ των προτέρων κατανομή $\pi(\theta)$ και έχει σταθερή συνάρτηση κινδύνου, δηλαδή $R(T_*,\theta)=c$, για κάθε $\theta\in\Theta$, τότε είναι εκτιμητής minimax του $g(\theta)$ (ως προς την $L(t,\theta)$).*

Απόδειξη. (για μη αριθμήσιμο Θ)
Για κάθε εκτιμητή $T(\underset{\sim}{X})$ με $\sup_{\theta\in\Theta} R(T,\theta)<\infty$ έχουμε

$$\begin{aligned}
\sup_{\theta\in\Theta} R(T_*,\theta) = c &= \int_\Theta c\cdot\pi(\theta)\,\mathrm{d}\theta \quad \text{(ολοκλήρωμα πυκνότητας} = 1)\\
&= \int_\Theta R(T_*,\theta)\pi(\theta)\,\mathrm{d}\theta\\
&\le \int_\Theta R(T,\theta)\pi(\theta)\,\mathrm{d}\theta \quad \text{(ορισμός εκτιμητή Bayes)}\\
&\le \int_\Theta \sup_{\theta\in\Theta} R(T,\theta)\pi(\theta)\,\mathrm{d}\theta \quad \text{(ορισμός supremum)}\\
&= \sup_{\theta\in\Theta} R(T,\theta)\int_\Theta \pi(\theta)\,\mathrm{d}\theta = \sup_{\theta\in\Theta} R(T,\theta).
\end{aligned}$$

Άρα,

$$\sup_{\theta\in\Theta} R(T_*,\theta) \le \sup_{\theta\in\Theta} R(T,\theta)$$

για κάθε εκτιμητή T και επομένως ο εκτιμητής T_* είναι εκτιμητής minimax. □

Βάσει λοιπόν της Πρότασης 8.2.1, για να δείξουμε ότι ένας εκτιμητής $T_*(\underset{\sim}{X})$ είναι εκτιμητής minimax του $g(\theta)$ ως προς τη συνάρτηση ζημίας $L(t,\theta)$, αρκεί να βρούμε κάποια εκ των προτέρων κατανομή $\pi(\theta)$, έτσι ώστε ο $T_*(\underset{\sim}{X})$ να είναι εκτιμητής Bayes του $g(\theta)$ ως προς $L(t,\theta)$ και $\pi(\theta)$ με σταθερή συνάρτηση κινδύνου. Τότε ο $T_*(\underset{\sim}{X})$ είναι εκτιμητής minimax του $g(\theta)$ ως προς $L(t,\theta)$. Ως εφαρμογή των παραπάνω δίνουμε το εξής παράδειγμα.

Παράδειγμα 8.2.1. (κατανομή Bernoulli – minimax εκτιμητής της πιθανότητας «επιτυχίας») Έστω $\underset{\sim}{X} = (X_1,\ldots,X_n)$ ένα τυχαίο δείγμα από την κατανομή Bernoulli $\mathcal{B}(1,\theta)$, $\theta \in (0,1)$. Θα βρούμε εκτιμητή minimax του θ ως προς τη συνάρτηση ζημίας $L(t,\theta) = (t-\theta)^2$.
Από το Παράδειγμα 8.1.1 έχουμε ότι ο εκτιμητής $T_*(\underset{\sim}{X}) = \frac{\sum X_i+\alpha}{n+\alpha+\beta}$ είναι εκτιμητής Bayes του θ ως προς τη συνάρτηση ζημίας $L(t,\theta) = (t-\theta)^2$ και εκ των προτέρων κατανομή $\mathcal{B}eta(\alpha,\beta)$. Θα αναζητήσουμε τώρα τα α,β έτσι ώστε η συνάρτηση κινδύνου του $T_*(\underset{\sim}{X})$ να είναι σταθερή ως προς θ. Έχουμε

$$\begin{aligned}
R(T_*,\theta) &= \mathbb{E}_\theta L(T_*,\theta) = \mathbb{E}_\theta (T_*-\theta)^2\\
&= \mathbb{E}_\theta\Big(\frac{\sum X_i+\alpha}{n+\alpha+\beta}-\theta\Big)^2\\
&= \mathbb{V}ar_\theta\Big(\frac{\sum X_i+\alpha}{n+\alpha+\beta}\Big) + \Big(\mathbb{E}_\theta\frac{\sum X_i+\alpha}{n+\alpha+\beta}-\theta\Big)^2\\
&= \frac{1}{(n+\alpha+\beta)^2}\big\{\big[(\alpha+\beta)^2-n\big]\theta^2 + \big[n-2a(\alpha+\beta)\big]\theta+\alpha^2\big\}.
\end{aligned}$$

Για να είναι λοιπόν η συνάρτηση $R(T_*,\theta)$ σταθερή ως προς θ, πρέπει

$$(\alpha+\beta)^2 - n = 0 \qquad \text{και} \qquad n - 2\alpha(\alpha+\beta) = 0.$$

Από τις σχέσεις αυτές (και επειδή τα α,β είναι θετικά) προκύπτει ότι $\alpha = \beta = \frac{\sqrt{n}}{2}$. Επομένως ο εκτιμητής

$$T_*(\underset{\sim}{X}) = \frac{\sum_{i=1}^n X_i + \frac{\sqrt{n}}{2}}{n+\sqrt{n}}$$

είναι εκτιμητής Bayes του θ με σταθερή συνάρτηση κινδύνου και συνεπώς, από την Πρόταση 8.1.4, είναι και εκτιμητής minimax του θ.

8.3 Ασκήσεις

8.1. Να δειχθεί ότι ο εκτιμητής Bayes είναι συνάρτηση της (ελάχιστης) επαρκούς στατιστικής συνάρτησης.

8.2. Να δειχθεί ότι, εάν ο εκτιμητής Bayes είναι μοναδικός, τότε είναι αποδεκτός με κριτήριο την αντίστοιχη μέση ζημία.

8.3. Να δειχθεί ότι ένας εκτιμητής με σταθερή συνάρτηση κινδύνου και αποδεκτός ως προς την αντίστοιχη μέση ζημία είναι εκτιμητής minimax.

8.4. Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την κατανομή Bernoulli $\mathcal{B}(1, \theta)$, $\theta \in (0, 1)$. Να βρεθεί ο εκτιμητής Bayes του $\theta(1-\theta)$ (διασπορά της κατανομής) ως προς τη συνάρτηση ζημίας $L(t, \theta) = \big(t - \theta(1-\theta)\big)^2$ και εκ των προτέρων κατανομή

α. $\pi(\theta) = 1,\ 0 < \theta < 1,$

β. $\pi(\theta) = 2\theta,\ 0 < \theta < 1.$

8.5. Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την κανονική κατανομή $\mathcal{N}(\theta, 1)$, $\theta \in \Theta = \mathbb{R}$. Να βρεθεί ο εκτιμητής Bayes του θ, εάν η συνάρτηση ζημίας είναι το τετραγωνικό σφάλμα, $L(t, \theta) = (t - \theta)^2$, και εκ των προτέρων κατανομή είναι κανονική $\mathcal{N}(\theta_0, 1)$, όπου θ_0 δεδομένη σταθερά.

8.6. Έστω X μία παρατήρηση από τη διωνυμική κατανομή $\mathcal{B}(n, \theta)$, $\theta \in (0, 1)$. Να βρεθεί ο εκτιμητής Bayes του θ, εάν η συνάρτηση ζημίας είναι το τετραγωνικό σφάλμα και η εκ των προτέρων κατανομή είναι ομοιόμορφη $\mathcal{U}(0, 1)$.

8.7. Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την ομοιόμορφη κατανομή $\mathcal{U}(0, \theta)$, $\theta \in (0, \infty)$. Να βρεθεί ο εκτιμητής Bayes του θ με συνάρτηση ζημίας $L(t, \theta) = \frac{(t-\theta)^2}{\theta^2}$ και εκ των προτέρων κατανομή $\pi(\theta) = 1$, $0 < \theta < 1$.

8.8. Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την κατανομή με πυκνότητα $f_1(x;\theta) = \theta e^{-\theta x}$, $x > 0$, $\theta > 0$. Η συνάρτηση ζημίας είναι το τετραγωνικό σφάλμα $L(t,\theta) = (t - g(\theta))^2$ και η εκ των προτέρων κατανομή είναι Γάμμα $\mathcal{G}(\alpha, \beta)$.

α. Να βρεθεί και να αναγνωριστεί η εκ των υστέρων κατανομή του θ.

β. Να βρεθεί ο εκτιμητής Bayes του θ.

γ. Να βρεθεί ο εκτιμητής Bayes του $1/\theta$ (μέση τιμή της κατανομής της X_1) και να εξεταστεί εάν είναι ασθενώς συνεπής.

8.9. Να υπολογιστεί ο κίνδυνος Bayes των εκτιμητών Bayes των Παραδειγμάτων 8.1.1 και 8.1.3.

8.10. Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την κανονική κατανομή $\mathcal{N}(0, \theta)$, $\theta \in \Theta = (0, \infty)$. Για την εκτίμηση της διασποράς θ, θεωρούμε τη συνάρτηση ζημίας του Stein την $L(t,\theta) = \frac{t}{\theta} - \ln\frac{t}{\theta} - 1$. Η εκ των προτέρων κατανομή του θ είναι $\pi(\theta) = \frac{1}{\Gamma(\alpha)\beta^\alpha \theta^{\alpha+1}} e^{-\frac{1}{\beta\theta}}$, $\theta > 0$, $\alpha > 0$, $\beta > 0$ (που αναφέρεται ως αντίστροφη Γάμμα, γιατί η $\frac{1}{\theta}$ έχει Γάμμα $\mathcal{G}(\alpha,\beta)$ κατανομή). Να δειχθούν τα εξής.

α. Η εκτίμηση Bayes του θ είναι $T_*(\underset{\sim}{x}) = \dfrac{1}{\mathbb{E}(Y^{-1})}$, όπου η τυχαία μεταβλητή Y έχει κατανομή την εκ των υστέρων κατανομή $\pi(\theta|\underset{\sim}{x})$, ανεξάρτητα από την εκ των προτέρων κατανομή.

β. Για τη δοθείσα $\pi(\theta)$, η εκτίμηση Bayes είναι $T_*(\underset{\sim}{x}) = \dfrac{\beta\sum_{i=1}^n x_i^2 + 2}{(\alpha+n)\beta}$.

Βιβλιογραφία

1. Bickel, P.J. and Doksum, K.A. (1977). *Mathematical Statistics, Basic Ideas and Selected Topics, Vol. 1.* Holden-Day.

2. Bolstad, W. M. (2007). *Introduction to Bayesian Statistics*, 2nd ed., Wiley.

3. Gelman, A. , Carlin, J. B. , Stern H. S. , and Rubin, D. B. (2004). Bayesian Data Analysis, 3rd ed., Chapman and Hall.

4. Hoff, P.D. (2009). *A First Course in Bayesian Statistical Methods*, Springer.

5. Lee, P.M. (2012). *Bayesian Statistics: An Introduction*, 4th ed., Wiley.

6. Schervish, M. (1995). *Theory of Statistics*. Springer.

Κεφάλαιο 9

Διαστήματα εμπιστοσύνης

Τα Διαστήματα Εμπιστοσύνης (Δ.Ε.) (ή πιο γενικά οι περιοχές εμπιστοσύνης) αποτελούν τον δεύτερο βασικό κλάδο της Στατιστικής Συμπερασματολογίας. Όπως υποδηλώνει η ονομασία τους, οι πληροφορίες που παρέχουν για την άγνωστη τιμή $g(\theta)$ είναι υπό μορφή διαστήματος (ή περιοχής) τιμών αντί μιας μόνον τιμής, της εκτίμησης. Προσδιορίζοντας ένα Δ.Ε., αναμένουμε αυτό να περιέχει την άγνωστη τιμή $g(\theta)$ με «μεγάλη» πιθανότητα.

Σε ορισμένες περιπτώσεις, ένα Δ.Ε. παρέχει επίσης πληροφορίες για το σφάλμα της εκτίμησης του $g(\theta)$ από έναν συγκεκριμένο εκτιμητή, υπηρετώντας ρόλο συμπληρωματικό της Εκτιμητικής. Σε κάθε περίπτωση, προσφέροντας μαζί με την εκτίμηση και ένα Δ.Ε., παρέχουμε πληρέστερη εικόνα για την άγνωστη τιμή $g(\theta)$.

Σε αυτό το κεφάλαιο, αφού ορίσουμε την έννοια του Δ.Ε., μελετάμε δύο γενικές μεθόδους κατασκευής Δ.Ε., αυτήν της «ποσότητας οδηγού» – pivotal quantity – και αυτήν της συνάρτησης κατανομής. Ακολούθως, εφαρμόζουμε την πρώτη σε δεδομένα τα οποία προέρχονται από μία κανονική κατανομή. Για να καλύψουμε περιπτώσεις που δεν μπορεί να χρησιμοποιηθεί καμία από αυτές αλλά και ανεξάρτητα, ασχολούμαστε με την κατασκευή ασυμπτωτικών διαστημάτων εμπιστοσύνης, δηλαδή Δ.Ε. που μπορούν να χρησιμοποιηθούν, όταν υπάρχει διαθέσιμος μεγάλος αριθμός δεδομένων. Τέλος, γίνεται μια σύντομη αναφορά σε Μπεϋζιανά διαστήματα.

9.1 Ορισμός και ορολογία

Για να ορίσουμε την έννοια του διαστήματος εμπιστοσύνης ας εξετάσουμε τα εξής. Έστω $T(\underset{\sim}{X})$ ένας εκτιμητής του $g(\theta)$. Σε πολλές περιπτώσεις υπολογισμός της εκτίμησης $T(\underset{\sim}{x})$ - όπου $\underset{\sim}{x}$ είναι η παρατηρηθείσα τιμή του δείγματος $\underset{\sim}{X}$ - και ανακοίνωση μόνον αυτής δεν αρκεί. Πρέπει να δοθεί συγχρόνως και κάποια πληροφορία για την ακρίβεια ή το σφάλμα της εκτίμησης $T(\underset{\sim}{x})$. Έτσι συχνά στις εφαρμογές ανακοινώνεται η εκτίμηση του $g(\theta)$ υπό τη μορφή $T(\underset{\sim}{x}) \pm \varepsilon$, όπου ε ένας θετικός αριθμός που εκφράζει το σφάλμα. Για παράδειγμα, οι προβλέψεις εκλογικών αποτελεσμάτων με τη διαδικασία των exit polls παρουσιάζονται υπό τη μορφή $33\% \pm 2\%$, $5.5\% \pm 0.5\%$ κ.λπ.. Οι τιμές $T(\underset{\sim}{x}) - \varepsilon$ και $T(\underset{\sim}{x}) + \varepsilon$ μπορούν να θεωρηθούν, αντίστοιχα, ως άνω και κάτω όρια (φράγματα) της άγνωστης τιμής $g(\theta)$. Εν τούτοις, δεν υπάρχει 100% βεβαιότητα ότι το διάστημα $[T(\underset{\sim}{X}) - \varepsilon, T(\underset{\sim}{X}) + \varepsilon]$ περιέχει την άγνωστη τιμή $g(\theta)$. Στην πραγματικότητα, στα περισσότερα προβλήματα, η πιθανότητα το διάστημα $[T(\underset{\sim}{X}) - \varepsilon, T(\underset{\sim}{X}) + \varepsilon]$ να μην περιέχει το $g(\theta)$ είναι θετική για κάθε (συμβατή τιμή του) $\varepsilon > 0$.

Για παράδειγμα, ας θεωρήσουμε ότι τα δεδομένα $\underset{\sim}{X} = (X_1, \ldots, X_n)$ είναι μετρήσεις μιας «ποσότητας» θ που γίνονται με κάποιο όργανο γνωστής ακρίβειας. Ως υπόθεση εργασίας, θεωρούμε ότι οι μετρήσεις είναι ανεξάρτητες παρατηρήσεις από την κανονική κατανομή με μέση τιμή την άγνωστη «ποσότητα» θ και διασπορά σ^2, όπου σ^2 γνωστή σταθερά, δηλαδή $\mathcal{N}(\theta, \sigma^2)$. Τότε είναι φυσικό να εκτιμήσουμε το θ με τον εκτιμητή $T(\underset{\sim}{X}) = \bar{X}$ που έχει *τυπικό σφάλμα* $\sqrt{\mathrm{MT\Sigma}(\bar{X}, \theta)} = \sqrt{\mathbb{E}_\theta(\bar{X} - \theta)^2} = \sqrt{\mathbb{V}ar_\theta \bar{X}} = \frac{\sigma}{\sqrt{n}}$ (Πρόταση 4.2.3). Ας δούμε τι σημαίνει να ανακοινώσουμε ότι η τιμή του θ κυμαίνεται μέσα στα όρια $\bar{X} \pm \frac{\sigma}{\sqrt{n}}$. Αρχικά, θα υπολογίσουμε την πιθανότητα το διάστημα $[\bar{X} - \varepsilon, \bar{X} + \varepsilon]$ να μην περιέχει το θ με δεδομένη τη σταθερά $\varepsilon > 0$. Από τις Προτάσεις 1.8.6(1) και 1.6.1(2) γνωρίζουμε ότι η τυχαία μεταβλητή $Z = \frac{\sqrt{n}(\bar{X} - \theta)}{\sigma}$ έχει τυπική κανονική κατανομή $\mathcal{N}(0, 1)$.

Επομένως έχουμε

$$\begin{aligned}\mathbb{P}_\theta(\theta \notin [\bar{X}-\varepsilon, \bar{X}+\varepsilon]) &= \mathbb{P}(|\bar{X}-\theta| > \varepsilon) \\ &= \mathbb{P}(|Z| > \tfrac{\sqrt{n}}{\sigma}\varepsilon) = 2\mathbb{P}(Z > \tfrac{\sqrt{n}}{\sigma}\varepsilon),\end{aligned}$$

λόγω της συμμετρίας της $\mathcal{N}(0,1)$.

Έτσι, εάν θέσουμε $\varepsilon = \frac{\sigma}{\sqrt{n}}$, η πιθανότητα το διάστημα $[\bar{X}-\frac{\sigma}{\sqrt{n}}, \bar{X}+\frac{\sigma}{\sqrt{n}}]$ να μην περιέχει το θ είναι

$$2\mathbb{P}(Z > \tfrac{\sqrt{n}}{\sigma}\varepsilon) = 2\mathbb{P}(Z > 1) = 2(1-\mathbb{P}(Z \leqslant 1)) \approx 0.32,$$

όπως προκύπτει από πίνακα της τυπικής κανονικής κατανομής (ή κάποιο στατιστικό λογισμικό). Συνεπώς τα όρια $\bar{X} \pm \frac{\sigma}{\sqrt{n}}$ κάθε άλλο παρά ικανοποιητικά είναι ως όρια της τιμής του θ. Από την άλλη πλευρά είναι προφανές ότι, επειδή $2\mathbb{P}(Z > \frac{\sqrt{n}}{\sigma}\varepsilon) \to 0$ καθώς $\varepsilon \to \infty$, με κατάλληλη επιλογή του ε η πιθανότητα $2\mathbb{P}(Z > \frac{\sqrt{n}}{\sigma}\varepsilon)$ μπορεί να γίνει αυθαίρετα «μικρή». Τα παραπάνω συνιστούν ότι, αντί να πάρουμε $\varepsilon = \frac{\sigma}{\sqrt{n}}$, μπορούμε να προκαθορίσουμε ένα «μικρό» αριθμό α, $0 < \alpha < 1$, και κατόπιν να προσδιορίσουμε (π.χ. από πίνακα της τυπικής κανονικής κατανομής) την τιμή της σταθεράς ε, έστω ε_0, έτσι ώστε

$$\mathbb{P}_\theta(\theta \notin [\bar{X}-\varepsilon_0, \bar{X}+\varepsilon_0]) = 2\mathbb{P}(Z > \tfrac{\sqrt{n}}{\sigma}\varepsilon_0) = \alpha.$$

Ισοδύναμα, έχουμε

$$\mathbb{P}_\theta(\bar{X}-\varepsilon_0 \le \theta \le \bar{X}+\varepsilon_0) = 1-\alpha$$

και μπορούμε να ανακοινώσουμε ότι το διάστημα $[\bar{X}-\varepsilon_0, \bar{X}+\varepsilon_0]$ περιέχει την άγνωστη τιμή θ με πιθανότητα $1-\alpha$ ή ότι είμαστε $100(1-\alpha)\%$ βέβαιοι ότι η άγνωστη τιμή θ περιέχεται στο διάστημα $[\bar{X}-\varepsilon_0, \bar{X}+\varepsilon_0]$. Ο αριθμός α είναι δικής μας επιλογής και κατά κάποιο τρόπο εκφράζει το βαθμό «ανεκτικότητας» μας, υποδεικνύοντας ότι είμαστε διατεθειμένοι να «ανεχθούμε» ή να «ρισκάρουμε» $100\,\alpha\%$ πιθανότητα το διάστημα να μην περιέχει την άγνωστη τιμή του θ.

Γενικεύοντας τα παραπάνω, ορίζουμε την έννοια του διαστήματος εμπιστοσύνης ως εξής.

Ορισμός 9.1.1. *Εάν* $T_1(\underset{\sim}{X})$ *και* $T_2(\underset{\sim}{X})$ *είναι στατιστικές συναρτήσεις με* $T_1(\underset{\sim}{X}) < T_2(\underset{\sim}{X})$, *το τυχαίο διάστημα* $[T_1(\underset{\sim}{X}), T_2(\underset{\sim}{X})]$ *ονομάζεται διάστημα εμπιστοσύνης για το* $g(\theta)$ *με συντελεστή εμπιστοσύνης (σ.ε.)* $100(1-\alpha)\%$ *εάν*

$$\mathbb{P}_\theta\big(T_1(\underset{\sim}{X}) \le g(\theta) \le T_2(\underset{\sim}{X})\big) = 1-\alpha,\ \forall\, \theta \in \Theta. \tag{9.1}$$

Επιπλέον, η πιθανότητα στο πρώτο μέλος της (9.1) *ονομάζεται πιθανότητα κάλυψης του* $g(\theta)$.

Σημειώνουμε ότι σε ορισμένες περιπτώσεις, ειδικά σε διακριτές κατανομές, δεν είναι πάντοτε δυνατόν (δηλαδή για κάθε α) να βρούμε ένα διάστημα $[T_1(\underset{\sim}{X}), T_2(\underset{\sim}{X})]$ έτσι ώστε $\mathbb{P}_\theta(T_1 \le g(\theta) \le T_2) = 1-\alpha$. Σε αυτές τις περιπτώσεις προσπαθούμε να βρούμε στατιστικές συναρτήσεις T_1, T_2, έτσι ώστε η πιθανότητα $\mathbb{P}_\theta(T_1(\underset{\sim}{X}) \le g(\theta) \le T_2(\underset{\sim}{X}))$ να είναι τουλάχιστον $1-\alpha$ και όσον το δυνατόν πλησιέστερα στο $1-\alpha$. Έτσι, λαμβάνοντας υπόψη αυτές τις περιπτώσεις, ένα διάστημα εμπιστοσύνης με σ.ε. τουλάχιστον $100(1-\alpha)\%$ μπορεί να οριστεί και από τη σχέση

$$\mathbb{P}_\theta\big(T_1(\underset{\sim}{X}) \le g(\theta) \le T_2(\underset{\sim}{X})\big) \ge 1-\alpha,\ \forall\, \theta \in \Theta$$

και στην περίπτωση που ισχύει η ισότητα (δηλαδή η (9.1)) αναφέρεται ως ακριβές (exact) Δ.Ε.. Στη συνέχεια, πάντως, χάριν απλότητας θα παραλείπεται ο προσδιορισμός «ακριβές». Επίσης σε περιπτώσεις, όπως παραπάνω ή άλλες που δεν είναι δυνατόν να προσδιορίσουμε διάστημα εμπιστοσύνης με συντελεστή εμπιστοσύνης (ακριβώς) $100(1-\alpha)\%$ (π.χ. μπορεί η κατανομή των δεδομένων να είναι εντελώς άγνωστη), συχνά προσπαθούμε να βρούμε στατιστικές συναρτήσεις $T_1(\underset{\sim}{X}), T_2(\underset{\sim}{X})$ εφόσον το μέγεθος του δείγματος n είναι «μεγάλο», έτσι ώστε

$$\mathbb{P}_\theta\big(T_1(\underset{\sim}{X}) \le g(\theta) \le T_2(\underset{\sim}{X})\big) \approx 1-\alpha,\ \forall\, \theta \in \Theta.$$

Η σχέση αυτή αυστηρά σημαίνει ότι $\lim_{n\to\infty} \mathbb{P}_\theta\big(T_1(\underset{\sim}{X}) \le g(\theta) \le T_2(\underset{\sim}{X})\big) = 1-\alpha$, για κάθε $\theta \in \Theta$. Ένα τέτοιο διάστημα $[T_1(\underset{\sim}{X}), T_2(\underset{\sim}{X})]$ λέγεται *ασυμπτωτικό διάστημα εμπιστοσύνης* (Α.Δ.Ε.) για το $g(\theta)$ με συντελεστή εμπιστοσύνης $100(1-\alpha)\%$ ή *διάστημα εμπιστοσύνης* για το $g(\theta)$ με συντελεστή εμπιστοσύνης *κατά προσέγγιση* $100(1-\alpha)\%$.

Αναφέρουμε ακόμη ότι άλλες φορές ενδιαφερόμαστε να βρούμε ένα κάτω φράγμα ή ένα άνω φράγμα για το $g(\theta)$ αντί διαστήματος εμπιστοσύνης. Για παράδειγμα, εάν το $g(\theta)$ παριστάνει τον μέσο χρόνο ζωής κάποιου συστήματος, τότε ένα κάτω φράγμα μπορεί να θεωρηθεί ως «εγγύηση» για την διάρκεια ζωής του συστήματος. Ένα *κάτω φράγμα*, $L(\underset{\sim}{X})$ για το $g(\theta)$, *με συντελεστή εμπιστοσύνης* $100(1-\alpha)\%$ ορίζεται από τη σχέση

$$\mathbb{P}_\theta\big(g(\theta) \geqslant L(\underset{\sim}{X})\big) = 1-\alpha,\ \forall\, \theta \in \Theta.$$

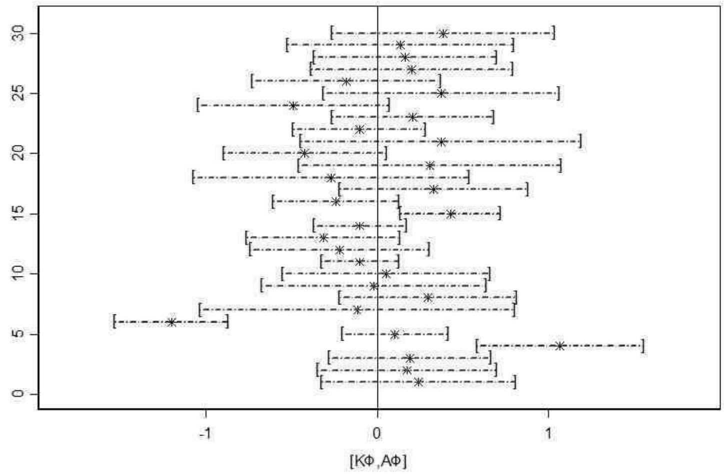
Ανάλογα ορίζονται οι έννοιες του άνω φράγματος, του *ασυμπτωτικού κάτω φράγματος* και του *ασυμπτωτικού άνω φράγματος*, από τις σχέσεις

$$\mathbb{P}_\theta\big(g(\theta) \leqslant U(\underset{\sim}{X})\big) = 1-\alpha,\ \forall\, \theta \in \Theta\ ,$$

$$\lim_{n\to\infty} \mathbb{P}_\theta\big(g(\theta) \geqslant L(\underset{\sim}{X})\big) = 1-\alpha,\ \forall\, \theta \in \Theta,$$
$$\lim_{n\to\infty} \mathbb{P}_\theta\big(g(\theta) \leqslant U(\underset{\sim}{X})\big) = 1-\alpha,\ \forall\, \theta \in \Theta\ ,$$

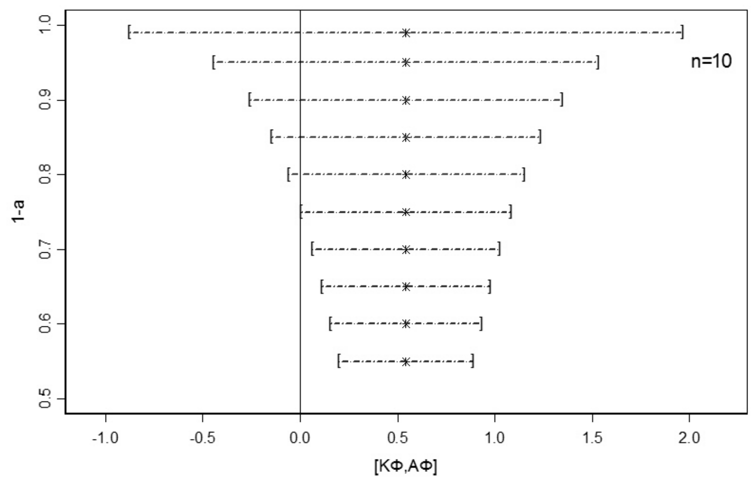
αντίστοιχα.

Παρατήρηση 9.1.1. Η σχέση (9.1) παραπέμπει στην κατανομή πιθανότητας του δείγματος $\underset{\sim}{X}$ και διαβάζεται ως «η πιθανότητα το (τυχαίο) διάστημα $[T_1(\underset{\sim}{X}), T_2(\underset{\sim}{X})]$ να περιέχει το $g(\theta)$ είναι $1-\alpha$». Επισημαίνουμε ότι εάν $\underset{\sim}{x}$ είναι η παρατηρηθείσα τιμή του δείγματος $\underset{\sim}{X}$, τότε το αριθμητικό διάστημα $[T_1(\underset{\sim}{x}), T_2(\underset{\sim}{x})]$ που προκύπτει από το τυχαίο διάστημα $[T_1(\underset{\sim}{X}), T_2(\underset{\sim}{X})]$ αντικαθιστώντας $\underset{\sim}{X} = \underset{\sim}{x}$ ή περιέχει την άγνωστη (αλλά σταθερή) τιμή $g(\theta)$ ή δεν την περιέχει, οπότε $\mathbb{P}_\theta(T_1(\underset{\sim}{x}) \leq g(\theta) \leq T_2(\underset{\sim}{x})) = 1$ ή 0, και όχι $1-\alpha$. Τονίζουμε λοιπόν ότι στον Ορισμό 9.1.1 το $g(\theta)$ είναι μια άγνωστη σταθερά και όχι τυχαία μεταβλητή. Για την περίπτωση που το θ και το $g(\theta)$ είναι τυχαίες μεταβλητές (Μπεϋζιανή προσέγγιση), θα αναφερθούμε στην Ενότητα 9.7. Η ερμηνεία του Δ.Ε. με συντελεστή εμπιστοσύνης $100(1-\alpha)\%$, όπως προκύπτει από τη σχέση (9.1) και τον εμπειρικό ορισμό πιθανότητας (Ενότητα 1.1), είναι η εξής· Εάν η συλλογή δεδομένων $\underset{\sim}{X}$ επαναληφθεί πολλές φορές και υπό τις ίδιες συνθήκες και κάθε φορά υπολογιστεί το διάστημα $[T_1(\underset{\sim}{X}), T_2(\underset{\sim}{X})]$, τότε εξ όλων αυτών των διαστημάτων, αναμένουμε περίπου τα $100(1-\alpha)\%$ να περιέχουν την άγνωστη τιμή $g(\theta)$. Για την συγκεκριμένη παρατηρηθείσα τιμή $\underset{\sim}{x}$ του δείγματος $\underset{\sim}{X}$, έχουμε την «ελπίδα» ότι το διάστημα $[T_1(\underset{\sim}{x}), T_2(\underset{\sim}{x})]$ είναι ένα από αυτά τα (πολλά) $100(1-\alpha)\%$ διαστήματα που περιέχουν το $g(\theta)$. Για παράδειγμα, το Σχήμα 9.1 απεικονίζει 30 Δ.Ε., με συντελεστή εμπιστοσύνης 90%, για τη μέση τιμή κανονικής με άγνωστη διασπορά. Τα δείγματα, μεγέθους $n = 10$, παρήχθησαν από την κατανομή $\mathcal{N}(0,1)$. Εξ αυτών των Δ.Ε., τα 27 περιέχουν τη μέση τιμή 0 ενώ τα υπόλοιπα 3 δεν την περιέχουν. Το μέσο των διαστημάτων αυτών (το σημείο $*$) είναι ο δειγματικός μέσος $\bar{x}$ του κάθε δείγματος. Η κατασκευή του συγκεκριμένου Δ.Ε. δίνεται στην Ενότητα 9.4.1.



Σχήμα 9.1: Δ.Ε. για τη μέση τιμή κανονικής με άγνωστη διασπορά, **[ΚΦ,ΑΦ]**=$\left[\bar{X} - t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}, \bar{X} + t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}\right]$

Όπως θα διαπιστώσουμε στις επόμενες ενότητες, είναι δυνατόν να υπάρχουν πολλά διαστήματα εμπιστοσύνης με τον ίδιο συντελεστή $100(1-\alpha)\%$. Είναι φανερό τότε ότι όσο μικρότερο είναι το μήκος του διαστήματος, τόσο καλύτερο είναι το διάστημα γιατί «εγκλωβίζει» την άγνωστη τιμή $g(\theta)$ μεταξύ πιο «κοντινών ορίων» από κάποιο άλλο με μεγαλύτερο μήκος. Δια τούτο, ως *κριτήριο επιλογής* μεταξύ διαστημάτων εμπιστοσύνης με τον ίδιο συντελεστή εμπιστοσύνης, συχνά λαμβάνεται *το μήκος* τους ή *η μέση τιμή του μήκους* τους, γιατί το μήκος είναι (γενικά) τυχαία μεταβλητή. Μεταξύ δύο Δ.Ε. με διαφορετικούς σ.ε., (για παράδειγμα 90% και 95%), αναμένουμε το δεύτερο, που έχει μεγαλύτερο σ.ε., να έχει και μεγαλύτερο μήκος από το πρώτο. Είναι λογικό, εξ άλλου, ότι, εάν απαιτήσουμε 95% «βεβαιότητα» το διάστημα να περιέχει την άγνωστη τιμή, τότε είναι πιθανόν να χρειαστεί να συμβιβαστούμε με ευρύτερα όρια κύμανσης αυτής της τιμής, από αυτά του διαστήματος με σ.ε. 90% (που όμως προσφέρει μικρότερη «βεβαιότητα»). Για παράδειγμα, το Σχήμα 9.2 απεικονίζει Δ.Ε., με διαφορετικό συντελεστή εμπιστοσύνης από 55% έως 99%, για τη μέση τιμή κανονικής με άγνωστη διασπορά. Τα Δ.Ε. κατασκευάστηκαν από ένα μόνο δείγμα, μεγέθους $n = 10$, που παρήχθη από την κατανομή $\mathcal{N}(0,1)$. Εξ αυτών των Δ.Ε., αυτά με συντελεστή εμπιστοσύνης μεγαλύτερο του 75% περιέχουν τη μέση τιμή, ενώ αυτά με μικρότερο του 75% δεν την περιέχουν. Το κοινό μέσο των διαστημάτων αυτών (το σημείο $*$) είναι ο δειγματικός μέσος $\bar{x}$ του δείγματος.



Σχήμα 9.2: Δ.Ε. για τη μέση τιμή κανονικής με άγνωστη διασπορά, [ΚΦ,ΑΦ]=$\left[\bar{X} - t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}, \bar{X} + t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}\right]$

9.2 Κατασκευή διαστημάτων εμπιστοσύνης μέσω ποσότητας οδηγού

Σε αυτήν την ενότητα παρουσιάζουμε μία γενική μέθοδο κατασκευής Δ.Ε. για το $g(\theta)$ με δεδομένο συντελεστή εμπιστοσύνης $100(1-\alpha)\%$, $0 < a < 1$. Η μέθοδος αυτή μπορεί να περιγραφεί ως εξής:

1. Προσπαθούμε να βρούμε μία τυχαία μεταβλητή $T(\underset{\sim}{X}, \theta)$, συνάρτηση των $\underset{\sim}{X}$ και θ, της οποίας η κατανομή δεν εξαρτάται από το θ. Αυτή η τυχαία μεταβλητή αναφέρεται ως *ποσότητα οδηγός (pivotal quantity) ή αντιστρεπτή ποσότητα.* Όταν το θ είναι πραγματική παράμετρος, η ποσότητα οδηγός είναι συχνά ένας μετασχηματισμός εκτιμητή του θ (που περιέχει και το θ).

2. Προσδιορίζουμε σταθερές $c_1 < c_2$ (που εξαρτώνται από την κατανομή της T και το α, αλλά όχι το θ) έτσι ώστε
$$\mathbb{P}_\theta\big(c_1 \leq T(\underset{\sim}{X}, \theta) \leq c_2\big) = 1 - \alpha, \ \forall\, \theta \in \Theta.$$

3. Λύνουμε (αντιστρέφουμε) την παραπάνω διπλή ανισότητα ως προς θ ή $g(\theta)$ και εφόσον αυτό είναι δυνατόν καταλήγουμε στη σχέση
$$\mathbb{P}_\theta\big(T_1(\underset{\sim}{X}) \leq g(\theta) \leq T_2(\underset{\sim}{X})\big) = 1 - \alpha, \ \forall\, \theta \in \Theta,$$
για κάποιες στατιστικές συναρτήσεις $T_1(\underset{\sim}{X}), T_2(\underset{\sim}{X})$. Τότε το διάστημα $[T_1(\underset{\sim}{X}), T_2(\underset{\sim}{X})]$ είναι Δ.Ε. για το $g(\theta)$ με σ.ε. $100(1-\alpha)\%$. Σημειώνουμε εδώ ότι, αν η g είναι γνησίως μονότονη συνάρτηση, από ένα Δ.Ε. για το θ, μπορούμε να παράγουμε αμέσως αντίστοιχο Δ.Ε. για το $g(\theta)$. Επίσης για την κατασκευή άνω ή κάτω φράγματος εμπιστοσύνης με σ.ε. $100(1-\alpha)\%$, είναι προφανές ότι αρκεί να προσδιοριστεί μια σταθερά c_* ή c^* έτσι ώστε $\mathbb{P}(T(\underset{\sim}{X}, \theta) \leqslant c^*) = 1-\alpha$ ή $\mathbb{P}(T(\underset{\sim}{X}, \theta) \geqslant c_*) = 1 - \alpha$.

Η επόμενη πρόταση δίνει ποσότητες οδηγούς στην περίπτωση που το $\underset{\sim}{X} = (X_1, \ldots, X_n)$ είναι ένα τυχαίο δείγμα από συνεχή κατανομή. Οι ποσότητες αυτές μπορούν να χρησιμοποιηθούν στην εύρεση Δ.Ε. εφ' όσον θεωρούμενες ως συναρτήσεις του θ ή του $g(\theta)$ είναι αντιστρέψιμες. Αναφέρουμε όμως ότι ακόμη και οι ποσότητες οδηγοί $T(\underset{\sim}{X}, \theta)$ και $\tilde{T}(\underset{\sim}{X}, \theta)$ της Πρότασης 9.2.1, αν και συναρτήσεις όλου του δείγματος $\underset{\sim}{X}$, δεν είναι κατ' ανάγκη συναρτήσεις της (ελάχιστης) επαρκούς στατιστικής συνάρτησης οπότε τα Δ.Ε. που θα προκύψουν είναι αμφιβόλου αξίας. Σε αυτές τις περιπτώσεις συνιστάται να αναζητούνται άλλες ποσότητες οδηγοί που είναι συναρτήσεις της (ελάχιστης) επαρκούς στατιστικής συνάρτησης (βλέπε τα Παραδείγματα 9.2.1, 9.2.2).

Πρόταση 9.2.1. *Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από μία συνεχή κατανομή με συνάρτηση κατανομής $F(x;\theta)$, $x \in \mathbb{R}$, $\theta \in \Theta$. Τότε έχουμε τα εξής:*

1. *Η $Y_i = F(X_i;\theta)$ έχει ομοιόμορφη κατανομή $\mathcal{U}(0,1)$.*
2. *Η $Z_i = -2\ln F(X_i;\theta)$ έχει κατανομή χ^2_2.*
3. *Η $T(\underset{\sim}{X}, \theta) = -2\sum_{i=1}^{n} \ln F(X_i;\theta)$ έχει κατανομή χ^2_{2n}.*
4. *Η $U_i = 1 - F(X_i;\theta)$ έχει ομοιόμορφη κατανομή $\mathcal{U}(0,1)$.*
5. *Η $V_i = -2\ln(1 - F(X_i;\theta))$ έχει κατανομή χ^2_2.*
6. *Η $\tilde{T}(\underset{\sim}{X}, \theta) = -2\sum_{i=1}^{n} \ln(1 - F(X_i;\theta))$ έχει κατανομή χ^2_{2n}.*
7. *Οι Y_i, Z_i, U_i, V_i, $i = 1, \ldots, n$, $T(\underset{\sim}{X}, \theta)$ και $\tilde{T}(\underset{\sim}{X}, \theta)$ είναι ποσότητες οδηγοί.*

Απόδειξη. 1. (Η πρόταση είναι ένα γνωστό αποτέλεσμα που αναφέρεται στην αγγλική βιβλιογραφία ως probability integral transform και ισχύει γενικότερα, αρκεί η $F(x;\theta)$ να είναι συνεχής ως συνάρτηση του x.) Έχουμε $0 < F(X_i;\theta) < 1$ με πιθανότητα 1, άρα η τυχαία μεταβλητή Y_i παίρνει τιμές στο διάστημα $(0,1)$. Επιπλέον η συνάρτηση $Y_i = F(X_i;\theta)$ είναι γνησίως αύξουσα συνάρτηση του X_i και άρα είναι αντιστρέψιμη με αντίστροφη την $F^{-1}(Y_i;\theta)$, η οποία είναι επίσης γνησίως αύξουσα. Τότε, εάν $F_{Y_i}(y)$ είναι η συνάρτηση κατανομής της τυχαίας μεταβλητής Y_i, για $y \in (0,1)$ έχουμε

$$\begin{aligned} F_{Y_i}(y) &= \mathbb{P}(Y_i \le y) = \mathbb{P}\big(F(X_i;\theta) \le y\big) \\ &= \mathbb{P}\big(X_i \le F^{-1}(y;\theta)\big) = F\big(F^{-1}(y;\theta);\theta\big) = y. \end{aligned}$$

Παραγωγίζοντας, η πυκνότητα της τυχαίας μεταβλητής Y_i είναι

$$f_{Y_i}(y) = 1,\ 0 < y < 1$$

και συνεπώς η τυχαία μεταβλητή Y_i έχει ομοιόμορφη κατανομή $\mathcal{U}(0,1)$.

2. Έχουμε $Z_i = -2\ln Y_i = h(Y_i)$ και επομένως από την Πρόταση 1.7.1 η πυκνότητα της τυχαίας μεταβλητής Z_i είναι

$$f_{Z_i}(\zeta) = f_{Y_i}\big(h^{-1}(\zeta)\big)\left|\tfrac{\mathrm{d}}{\mathrm{d}\zeta}\,h^{-1}(\zeta)\right| = \frac{1}{2}e^{-\zeta/2},\ 0 < \zeta < \infty$$

που είναι η πυκνότητα της κατανομής χ^2_2.

3. Οι τυχαίες μεταβλητές $Z_i = -2\ln F(X_i;\theta)$ είναι ανεξάρτητες (επειδή οι τυχαίες μεταβλητές X_i είναι ανεξάρτητες) με κατανομή χ^2_2. Επομένως το άθροισμα τους έχει κατανομή χι–τετράγωνο με βαθμούς ελευθερίας το άθροισμα των βαθμών ελευθερίας (Πρόταση 1.8.5(8)). Άρα η τυχαία μεταβλητή $T(\underset{\sim}{X},\theta)$ έχει κατανομή χ^2_{2n}.

4.5.6. Η (4) προκύπτει από την (1) και το γεγονός ότι, όταν $Y_i \sim \mathcal{U}(0,1)$ τότε και $1-Y_i \sim \mathcal{U}(0,1)$, η (5) είναι ουσιαστικά η (2), επειδή $U_i \sim Y_i$ και η (6) είναι ουσιαστικά η (3), επειδή επίσης οι $U_i \sim Y_i$.

7. Οι τυχαίες μεταβλητές Y_i, Z_i, $T(\underset{\sim}{X},\theta)$ και $\tilde{T}(\underset{\sim}{X},\theta)$ είναι συναρτήσεις των $\underset{\sim}{X}$ και θ με κατανομές που δεν εξαρτώνται από το θ και συνεπώς είναι ποσότητες οδηγοί. $\square$

Παρατήρηση 9.2.1. Από τις Y_i, Z_i, U_i, V_i, $T(\underset{\sim}{X},\theta)$ και $\tilde{T}(\underset{\sim}{X},\theta)$ προτιμητέες είναι οι $T(\underset{\sim}{X},\theta)$ και $\tilde{T}(\underset{\sim}{X},\theta)$ επειδή χρησιμοποιούν όλο το δείγμα $\underset{\sim}{X}$.

Δίνουμε τώρα μερικά παραδείγματα κατασκευής Δ.Ε. ως εφαρμογές της προηγούμενης πρότασης και της μεθόδου της ποσότητας οδηγού.

Παράδειγμα 9.2.1. (κατανομή Βήτα – Δ.Ε. ίσων ουρών) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την κατανομή Βήτα $\mathcal{B}eta(1, \theta)$ με πυκνότητα $f(x; \theta) = \theta x^{\theta-1}$, $0 < x < 1$, $\theta \in \Theta = (0, \infty)$. Θα κατασκευάσουμε ένα Δ.Ε. με σ.ε. $100(1-\alpha)\%$ για το θ.
Η συνάρτηση κατανομής των X_i είναι

$$F(x; \theta) = \begin{cases} 0, & x \leq 0 \\ x^{\theta}, & 0 < x < 1 \\ 1, & x \geq 1. \end{cases}$$

Επειδή $0 < X_i < 1$ (με πιθανότητα 1) έχουμε ότι $F(X_i; \theta) = X_i^{\theta}$ και επομένως, από την Πρόταση 9.2.1, η τυχαία μεταβλητή

$$T(\underset{\sim}{X}, \theta) = -2 \sum_{i=1}^{n} \ln F(X_i; \theta) = -2\theta \sum_{i=1}^{n} \ln X_i$$

είναι ποσότητα οδηγός με κατανομή χ^2_{2n}. Η ποσότητα οδηγός T μπορεί τώρα να χρησιμοποιηθεί για την κατασκευή ενός Δ.Ε. για το θ ως εξής. Για $0 < \alpha < 1$ υπάρχουν σταθερές $c_1 < c_2$ έτσι ώστε

$$\mathbb{P}_{\theta}(c_1 \leq T \leq c_2) = 1 - \alpha, \ \forall\, \theta \in \Theta. \tag{9.2}$$

Οι σταθερές c_1, c_2 εξαρτώνται από την κατανομή της τυχαίας μεταβλητής T και το α. Επομένως,

$$\mathbb{P}_{\theta}\Big(c_1 \leq -2\theta \sum_{i=1}^{n} \ln X_i \leq c_2\Big) = 1 - \alpha, \ \forall\, \theta \in \Theta,$$

ή

$$\mathbb{P}_{\theta}\bigg(- c_1 \frac{1}{2\sum_{i=1}^{n} \ln X_i} \leq \theta \leq -c_2 \frac{1}{2\sum_{i=1}^{n} \ln X_i}\bigg) = 1 - \alpha, \ \forall\, \theta \in \Theta.$$

Συνεπώς το διάστημα

$$\left[-\frac{c_1}{2\sum_{i=1}^n \ln X_i}, -\frac{c_2}{2\sum_{i=1}^n \ln X_i}\right]$$

είναι Δ.Ε. για το θ με σ.ε. $100(1-\alpha)\%$. Υπάρχουν άπειρα ζεύγη c_1, c_2 που ικανοποιούν την (9.2) και κάθε ένα από αυτά παράγει ένα Δ.Ε. για το θ. Μία από αυτές τις επιλογές των c_1, c_2 είναι η εξής. Από την (9.2) έχουμε

$$\mathbb{P}_\theta(T < c_1) + \mathbb{P}_\theta(T > c_2) = \alpha$$

και τα c_1, c_2 μπορούν να προσδιοριστούν έτσι ώστε

$$\mathbb{P}_\theta(T < c_1) = \alpha/2 \text{ και } \mathbb{P}_\theta(T > c_2) = \alpha/2. \tag{9.3}$$

Εάν χρησιμοποιήσουμε το συμβολισμό $c = \chi^2_{\kappa,\alpha}$ για το σημείο c με $\mathbb{P}(\chi^2_\kappa > c) = \alpha$ τότε $c_2 = \chi^2_{2n,\alpha/2}$ και $c_1 = \chi^2_{2n,1-\alpha/2}$. Δοθέντος του n και του α, τα c_1, c_2 μπορούν να υπολογιστούν από πίνακες της κατανομής χ^2, οπότε το Δ.Ε. παίρνει τη μορφή

$$\left[-\frac{\chi^2_{2n,1-\alpha/2}}{2\sum_{i=1}^n \ln X_i}, -\frac{\chi^2_{2n,\alpha/2}}{2\sum_{i=1}^n \ln X_i}\right],$$

και αναφέρεται ως διάστημα εμπιστοσύνης *ίσων ουρών* λόγω της (9.3). Σημειώνουμε ότι στο παράδειγμα αυτό η (ελάχιστη) επαρκής στατιστική συνάρτηση είναι $\prod_{i=1}^n X_i$ και η ποσότητα οδηγός έχει την επιθυμητή ιδιότητα να εξαρτάται από τα δεδομένα $\underset{\sim}{X}$ μέσω της (ελάχιστης) επαρκούς στατιστικής συνάρτησης. (Εδώ μάλιστα η $2\sum\limits_{i=1}^n \ln X_i$ είναι επίσης ελάχιστη επαρκής – Παρατήρηση 6.1.3) Ως αποτέλεσμα, το *Δ.Ε. που προέκυψε είναι συνάρτηση της (ελάχιστης) επαρκούς στατιστικής συνάρτησης.*

Παρατήρηση 9.2.2. Από την (9.3), αντικαθιστώντας $T = -2\theta\sum\limits_{i=1}^n \ln X_i$ και $c_1 = \chi^2_{2n,1-\alpha/2}$, $c_2 = \chi^2_{2n,\alpha/2}$ προκύπτουν οι σχέσεις

$$\mathbb{P}\Big(\theta \geqslant -\frac{\chi^2_{2n,1-\alpha/2}}{2\sum\limits_{i=1}^n \ln X_i}\Big) = 1-\alpha/2 \text{ και } \mathbb{P}\Big(\theta \leqslant -\frac{\chi^2_{2n,\alpha/2}}{2\sum\limits_{i=1}^n \ln X_i}\Big) = 1-\alpha/2 \;\forall\, \theta \in \Theta$$

που δηλώνουν ότι τα άκρα του Δ.Ε. ίσων ουρών $L(\underset{\sim}{X}) = -\dfrac{\chi^2_{2n,1-\alpha/2}}{2\sum\limits_{i=1}^{n} \ln X_i}$ και $U(\underset{\sim}{X}) = -\dfrac{\chi^2_{2n,\alpha/2}}{2\sum\limits_{i=1}^{n} \ln X_i}$ είναι κάτω και άνω φράγμα εμπιστοσύνης για το θ με σ.ε. $100(1-\alpha/2)\%$, αντίστοιχα. Αφού οι παραπάνω σχέσεις ισχύουν για κάθε α, αλλάζοντας το $\alpha/2$ σε α στις εκφράσεις των $L(\underset{\sim}{X})$ και $U(\underset{\sim}{X})$, παίρνουμε $L_1(\underset{\sim}{X}) = -\dfrac{\chi^2_{2n,1-\alpha}}{2\sum\limits_{i=1}^{n} \ln X_i}$ και $U_1(\underset{\sim}{X}) = -\dfrac{\chi^2_{2n,\alpha}}{2\sum\limits_{i=1}^{n} \ln X_i}$ που είναι κάτω και άνω φράγμα εμπιστοσύνης για το θ με σ.ε. $100(1-\alpha)\%$. Η διαδικασία αυτή είναι γενικός κανόνας για την παραγωγή άνω και κάτω φράγματος από ένα Δ.Ε. ίσων ουρών.

Παράδειγμα 9.2.2. (ομοιόμορφη κατανομή – Δ.Ε.) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την ομοιόμορφη κατανομή $\mathcal{U}(0,\theta)$, $\theta \in \Theta = (0,\infty)$. Θα κατασκευάσουμε ένα Δ.Ε. για το θ με σ.ε. $100(1-a)\%$.
Η συνάρτηση κατανομής των X_i είναι

$$F(x;\theta) = \begin{cases} 0, & x \le 0 \\ \frac{x}{\theta}, & 0 < x < \theta \\ 1, & x \ge \theta. \end{cases}$$

Επειδή $0 < X_i < 1$ (με πιθανότητα 1) έχουμε ότι $F(X_i;\theta) = \frac{X_i}{\theta}$ και επομένως, από την Πρόταση 9.2.1, η τυχαία μεταβλητή

$$T(\underset{\sim}{X},\theta) = -2\sum_{i=1}^{n} \ln F(X_i;\theta) = -2\sum_{i=1}^{n} \ln X_i + 2n\ln\theta$$

είναι ποσότητα οδηγός με κατανομή χ^2_{2n}. Εφαρμόζοντας τώρα παρόμοια διαδικασία όπως στο Παράδειγμα 9.2.1 καταλήγουμε ότι το διάστημα

$$\left[e^{\frac{\chi^2_{2n,1-\alpha/2}}{2n}} \Big(\prod_{i=1}^{n} X_i\Big)^{1/n}, e^{\frac{\chi^2_{2n,\alpha/2}}{2n}} \Big(\prod_{i=1}^{n} X_i\Big)^{1/n} \right]$$

είναι Δ.Ε. ίσων ουρών για το θ με σ.ε. $100(1-\alpha)\%$.
Σημειώνουμε ότι η (ελάχιστη) επαρκής στατιστική συνάρτηση είναι $X_{(n)} =$

$\max\{X_1, \ldots, X_n\}$, (βλέπε Παράδειγμα 6.1.5,) και η ποσότητα οδηγός δεν είναι συνάρτηση της (ελάχιστης) επαρκούς στατιστικής συνάρτησης για $n \geq 2$. Ως αποτέλεσμα, το *Δ.Ε. που προέκυψε δεν βασίζεται στην (ελάχιστη) επαρκή στατιστική συνάρτηση* και δια τούτο δεν προτείνεται. Ένα Δ.Ε. για το θ που χρησιμοποιεί την (ελάχιστη) επαρκή στατιστική συνάρτηση, $X_{(n)}$, δίνεται στο επόμενο παράδειγμα.

Παράδειγμα 9.2.3. (Συνέχεια του Παραδείγματος 9.2.2) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την ομοιόμορφη κατανομή $\mathcal{U}(0, \theta)$, $\theta \in \Theta = (0, \infty)$. Θα κατασκευάσουμε ένα Δ.Ε. για το θ χρησιμοποιώντας διαφορετική ποσότητα οδηγό από αυτήν του Παραδείγματος 9.2.2. Η νέα ποσότητα οδηγός θα προκύψει μετασχηματίζοντας κατάλληλα έναν εκτιμητή του θ (βλέπε το πρώτο βήμα της περιγραφής της μεθόδου, στην αρχή αυτής της ενότητας). Ο ε.μ.π. του θ είναι $X_{(n)}$. Θέτουμε $Y = X_{(n)}$. Η πυκνότητα της τυχαίας μεταβλητής Y είναι (βλέπε Παράδειγμα 6.3.3)

$$f_Y(y;\theta) = \begin{cases} \dfrac{n}{\theta^n} y^{n-1}, & 0 < y < \theta \\ 0, & \text{διαφορετικά.} \end{cases}$$

Παρατηρούμε ότι το θ είναι παράμετρος κλίμακας για την κατανομή της Y (βλέπε Ενότητα 6.4), οπότε η $T = \frac{Y}{\theta}$ έχει κατανομή που δεν εξαρτάται από το θ. Συγκεκριμένα από την Πρόταση 1.7.1, η πυκνότητα της τυχαίας μεταβλητής $T = \frac{Y}{\theta}$ είναι

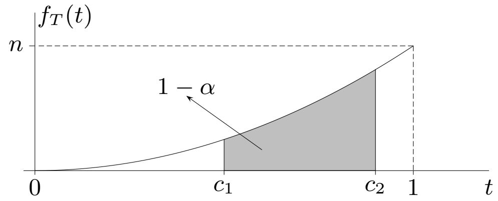
$$f_T(t) = \begin{cases} n t^{n-1}, & 0 < t < 1 \\ 0, & \text{διαφορετικά.} \end{cases}$$

Άρα η τυχαία μεταβλητή $T = \frac{X_{(n)}}{\theta}$ είναι ποσότητα οδηγός. Συνεπώς υπάρχουν σταθερές $c_1 < c_2$ που εξαρτώνται από το α, $0 < \alpha < 1$, (βλέπε Σχήμα **;;**) έτσι ώστε

$$\mathbb{P}_\theta\big(c_1 \leq T \leq c_2\big) = 1 - \alpha, \ \forall\, \theta > 0 \tag{9.4}$$

ή

$$\mathbb{P}_\theta\Big(c_1 \leq \frac{X_{(n)}}{\theta} \leq c_2\Big) = 1 - \alpha, \ \forall\, \theta > 0$$



Σχήμα 9.3: $\mathbb{P}_\theta(c_1 \le T \le c_2)$

ή

$$\mathbb{P}_\theta\Big(\frac{X_{(n)}}{c_2} \le \theta \le \frac{X_{(n)}}{c_1}\Big) = 1-\alpha, \ \forall\, \theta > 0.$$

Επομένως το διάστημα $\big[\frac{X_{(n)}}{c_2}, \frac{X_{(n)}}{c_1}\big]$ είναι Δ.Ε. για το θ με σ.ε. $100(1-\alpha)\%$. Ένας τρόπος προσδιορισμού των c_1 και c_2 είναι ο εξής: η (9.4) ικανοποιείται εάν $\mathbb{P}_\theta(T < c_1) = \alpha/2$ και $\mathbb{P}_\theta(T > c_2) = \alpha/2$. Σε αυτήν την περίπτωση, το διάστημα είναι Δ.Ε. ίσων ουρών σύμφωνα με την ορολογία που χρησιμοποιήθηκε στο Παράδειγμα 9.2.1. Επειδή για $x \in (0,1)$,

$$\mathbb{P}_\theta(T < x) = \int_0^x f_T(t)\,\mathrm{d}t = \int_0^x nt^{n-1}\,\mathrm{d}t = x^n,$$

έχουμε $c_1^n = \alpha/2$ και $1 - c_2^n = \alpha/2$, οπότε $c_1 = \sqrt[n]{\alpha/2}$ και $c_2 = \sqrt[n]{1-\alpha/2}$. Άρα, το Δ.Ε. ίσων ουρών για το θ με σ.ε. $100(1-\alpha)\%$ είναι

$$\Big[\frac{X_{(n)}}{\sqrt[n]{1-\alpha/2}}, \frac{X_{(n)}}{\sqrt[n]{\alpha/2}}\Big].$$

Εναλλακτικά τα c_1 και c_2 μπορούν να προσδιορισθούν έτσι ώστε το Δ.Ε. να έχει ελάχιστο μήκος. Το μήκος του διαστήματος $\big[\frac{X_{(n)}}{c_2}, \frac{X_{(n)}}{c_1}\big]$ είναι $\ell = X_{(n)}\big(\frac{1}{c_1} - \frac{1}{c_2}\big)$, οπότε για την ελαχιστοποίηση του ως προς c_1 και c_2 αρκεί να ελαχιστοποιήσουμε τη συνάρτηση $\ell^* = \frac{1}{c_1} - \frac{1}{c_2}$ υπό τον περιορισμό (9.4) που είναι ισοδύναμος προς τη σχέση

$$c_2^n - c_1^n = 1 - \alpha, \tag{9.5}$$

με $0 \le c_1 < c_2 \le 1$. Από την (9.5) παίρνουμε $c_2 = (c_1^n + 1 - \alpha)^{1/n}$. Τότε η συνάρτηση ℓ^* γράφεται $\ell^* = \frac{1}{c_1} - \frac{1}{(c_1^n+1-\alpha)^{1/n}}$ και έχει παράγωγο

$\frac{\mathrm{d}\ell^*}{\mathrm{d}c_1} = \frac{c_1^{n+1} - (c_1^n + 1 - \alpha)^{\frac{n+1}{n}}}{c_1^2 (c_1^n + 1 - \alpha)^{\frac{n+1}{n}}} < 0$, αφού $(c_1^n + 1 - \alpha)^{\frac{n+1}{n}} > (c_1^n)^{\frac{n+1}{n}} = c_1^{n+1}$. Επομένως η συνάρτηση ℓ^* είναι γνησίως φθίνουσα συνάρτηση του c_1 και επειδή $c_1^n = c_2^n - 1 + \alpha \le 1 - 1 + \alpha = \alpha$, δηλαδή $c_1 \le \sqrt[n]{\alpha}$, το ελάχιστο της ℓ^* επιτυγχάνεται για $c_1 = \sqrt[n]{\alpha}$ οπότε από την (9.5) $c_2 = 1$. Άρα το Δ.Ε. ελαχίστου μήκους για το θ με σ.ε. $100(1-\alpha)\%$ που βασίζεται στην (ελάχιστη) επαρκή στατιστική συνάρτηση $X_{(n)}$ είναι

$$\Big[X_{(n)}, \frac{X_{(n)}}{\sqrt[n]{\alpha}}\Big].$$

Από το αριστερό άκρο του, $X_{(n)}$, δεν αντλούμε καμία πληροφορία για το θ, αφού γνωρίζουμε ευθύς εξ' αρχής (δηλαδή, ανεξάρτητα από την κατασκευή του Δ.Ε.) ότι $\theta > X_{(n)}$, με πιθανότητα 1. Ως επακόλουθο, το δεξιό άκρο $\frac{X_{(n)}}{\sqrt[n]{\alpha}}$ είναι άνω φράγμα εμπιστοσύνης με σ.ε. $100(1-\alpha)\%$ αφού

$$\mathbb{P}_\theta\Big(\theta < \frac{X_{(n)}}{\sqrt[n]{\alpha}}\Big) = \mathbb{P}\Big(X_{(n)} < \theta < \frac{X_{(n)}}{\sqrt[n]{\alpha}}\Big) = 1 - \alpha.$$

Συνοψίζοντας, στην προκειμένη περίπτωση, το Δ.Ε. ελαχίστου μήκους περιέχει ακριβώς την ίδια πληροφορία για το θ όπως το άνω φράγμα.

Έχει ενδιαφέρον να αναφέρουμε ότι για $0 < \alpha < \frac{2}{3}$ (που καλύπτει όλες τις πρακτικές εφαρμογές) τα Δ.Ε. ίσων ουρών και ελαχίστου μήκους έχουν τομή το διάστημα $\Big[\frac{X_{(n)}}{\sqrt[n]{1-\alpha/2}}, \frac{X_{(n)}}{\sqrt[n]{\alpha}}\Big]$, το οποίο εύκολα αποδεικνύεται ότι είναι Δ.Ε. για το θ με σ.ε. $100(1-3\alpha/2)\%$.

Σύμφωνα με την Παρατήρηση 9.2.2, από το Δ.Ε. ίσων ουρών, αλλάζοντας το $\alpha/2$ σε α παίρνουμε το κάτω και άνω φράγμα εμπιστοσύνης για το θ με σ.ε. $100(1-\alpha)\%$, $L(\underset{\sim}{X}) = \frac{X_{(n)}}{\sqrt[n]{1-\alpha}}$ και $U(\underset{\sim}{X}) = \frac{X_{(n)}}{\sqrt[n]{\alpha}}$.

Παράδειγμα 9.2.4. (Εκθετική κατανομή – Δ.Ε.) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την εκθετική κατανομή $\mathcal{E}(\theta)$ με πυκνότητα $f_1(x;\theta) = \frac{1}{\theta}\, e^{-x/\theta}$, $x > 0$, $\theta \in \Theta = (0, \infty)$ και συνάρτηση κατανομής $F(x;\theta) = 1 - e^{-x/\theta}$, $x > 0$. Άρα, από την Πρόταση 9.2.1(6), η τυχαία μεταβλητή $\tilde{T}(\underset{\sim}{X}, \theta) = -2\sum_{i=1}^{n} \ln(1 - F(X_i;\theta)) = \frac{2}{\theta}\sum_{i=1}^{n} X_i$ έχει κατανομή χ^2_{2n}, είναι προφανώς αντιστρέψιμη ως προς θ και επιπλέον είναι συνάρτηση της (ελάχιστης) επαρκούς στατιστικής συνάρτησης $\sum_{i=1}^{n} X_i$. Η κατασκευή Δ.Ε. για το θ χρησιμοποιώντας την $\tilde{T}(\underset{\sim}{X}, \theta)$ είναι η Άσκηση 9.5. Αξίζει να αναφέρουμε ότι, σε αντίθεση, η $T(\underset{\sim}{X}, \theta) = -2\sum_{i=1}^{n} \ln F(X_i; \theta) = -2\sum_{i=1}^{n} \ln(1 - e^{-X_i/\theta})$ δεν είναι αντιστρέψιμη ως προς θ ($n \geqslant 2$), συνεπώς δεν χρησιμεύει για την κατασκευή Δ.Ε. για το θ.

Κλείνουμε αυτήν την ενότητα με την ακόλουθη παρατήρηση.

Παρατήρηση 9.2.3. Αν $\tilde{\theta}$ είναι στατιστική συνάρτηση (ή ειδικά, εκτιμητής του θ) και το θ είναι παράμετρος θέσης για την κατανομή του $\tilde{\theta}$, τότε η $T = \tilde{\theta} - \theta$ έχει κατανομή που δεν εξαρτάται από το θ, βλέπε Ενότητα 6.4. Άρα η T είναι ποσότητα οδηγός και μάλιστα (τετριμμένα) αντιστρέψιμη ως προς θ, οπότε μπορεί να χρησιμοποιηθεί για την κατασκευή Δ.Ε. για το θ. Η Άσκηση 9.17, στην οποία ζητείται Δ.Ε. για την παράμετρο θέσης μιας κατανομής Cauchy, είναι εφαρμογή αυτής της παρατήρησης. Αντίστοιχα, αν το θ είναι παράμετρος κλίμακας, τότε ποσότητα οδηγός είναι η $T = \frac{\tilde{\theta}}{\theta}$. Στο Παράδειγμα 9.2.3, προτάθηκε η $T = \frac{X_{(n)}}{\theta}$ ως ποσότητα οδηγός ακριβώς επειδή το θ είναι παράμετρος κλίμακας για την κατανομή του ε.μ.π. του θ, $X_{(n)}$.

9.3 Κατασκευή διαστημάτων εμπιστοσύνης μέσω της συνάρτησης κατανομής στατιστικής συνάρτησης

Θα παρουσιάσουμε μια άλλη μέθοδο κατασκευής διαστήματος εμπιστοσύνης για την παράμετρο θ που βασίζεται στη συνάρτηση κατανομής κατάλληλης στατιστικής συνάρτησης (ενώ η ποσότητα οδηγός δεν είναι στατιστική συνάρτηση). Συνήθως αυτή είναι ένας εκτιμητής του θ ή μια μονότονη συνάρτηση του εκτιμητή. Εν συνεχεία, από το Δ.Ε. για το θ μπορούμε αμέσως να κατασκευάσουμε ένα διάστημα εμπιστοσύνης για το $g(\theta)$, όταν η g είναι μονότονη συνάρτηση. Για καλύτερη κατανόηση της μεθόδου, παραθέτουμε το ακόλουθο παράδειγμα.

Παράδειγμα 9.3.1. (Κατανομή Bernoulli–Δ.Ε. για την πιθανότητα «επιτυχίας») Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την κατανομή Bernoulli $\mathcal{B}(1, \theta)$. Θα χρησιμοποιήσουμε ως σημείο εκκίνησης της μεθόδου τη στατιστική συνάρτηση $T = \sum_{i=1}^{n} X_i = n\bar{X}$ που παριστάνει τον αριθμό των «επιτυχιών» στις n δοκιμές Bernoulli. Η T είναι γνησίως αύξουσα συνάρτηση του κλασικού εκτιμητή του θ, $\bar{X}$ (ποσοστό «επιτυχιών» στο δείγμα $\underset{\sim}{X}$). Είναι λογικό να αναμένουμε ότι «μεγάλο» θ προκαλεί «μεγάλο» αριθμό «επιτυχιών», συνεπώς αναμένουμε ότι για κάθε δοθέν (σταθερό) $t \in \{1, 2, \ldots, n\}$, το ενδεχόμενο $T \geq t$ έχει αυξανόμενη πιθανότητα καθώς αυξάνει το θ, δηλαδή $\mathbb{P}_\theta(T \geq t)$ είναι γνησίως αύξουσα συνάρτηση του θ. Όντως, αυτή η ιδιότητα ισχύει (και συσχετίζει τη διωνυμική κατανομή με την κατανομή Βήτα, βλέπε (9.13)). Ανάλογα, η πιθανότητα $\mathbb{P}_\theta(T \leq t)$, $t \in \{0, 1, \ldots, n-1\}$ είναι (γνησίως) φθίνουσα συνάρτηση του θ.

Στη συνέχεια, εξαιρώντας, για χάρη απλότητας, τις ακραίες περιπτώσεις $t = 0$ και $t = n$ (δηλαδή, 0 και n «επιτυχίες», αντίστοιχα, στις n δοκιμές), θεωρούμε $t \in \{1, 2, \ldots, n-1\}$. Τότε, λόγω μονοτονίας και συνέχειας, για $0 < \alpha_1 < 1$ και $0 < \alpha_2 < 1$ (έχοντας κατά νου «μικρά» α_1 και α_2) υπάρχει ακριβώς ένα σημείο $\theta_1(t) \in \Theta = (0, 1)$ και ακριβώς ένα σημείο $\theta_2(t) \in \Theta = (0, 1)$ τέτοια ώστε

$$\mathbb{P}_{\theta_1(t)}(T \geq t) = \alpha_1, \qquad \mathbb{P}_{\theta_2(t)}(T \leq t) = \alpha_2. \tag{9.6}$$

Για «μικρό» α_1, επειδή $\mathbb{P}_\theta(T \geq t)$ είναι (γνησίως) αύξουσα ως προς θ, αναμένουμε ότι και το $\theta_1(t)$ είναι «μικρό». Ανάλογα, για «μικρό» α_2, αναμένουμε ότι το $\theta_2(t)$ είναι «μεγάλο». Δικαιολογείται έτσι ο ισχυρισμός $\theta_1(t) < \theta_2(t)$ που είναι αληθής αν περαιτέρω επιλέξουμε τα α_1 και α_2 έτσι ώστε $\alpha_1 + \alpha_2 < 1$. Θα επανέλθουμε σε αυτόν τον ισχυρισμό παρακάτω, αλλά ας τον δεχτούμε προς το παρόν.

Οι συναρτήσεις $\theta_1(t)$ και $\theta_2(t)$ ορίζουν το τυχαίο διάστημα $[\theta_1(T), \theta_2(T)]$ για το οποίο θα δείξουμε ότι ισχύει

$$\mathbb{P}_\theta(\theta_1(T) \leq \theta \leq \theta_2(T)) \geq 1 - \alpha, \qquad \alpha = \alpha_1 + \alpha_2, \quad \forall\, \theta \in \Theta \tag{9.7}$$

δηλαδή $[\theta_1(T), \theta_2(T)]$ είναι Δ.Ε. για το θ με σ.ε. τουλάχιστον $100(1-\alpha)\%$.

Κατ' αρχάς, παρατηρούμε ότι

$$\{t : \theta_2(t) \geq \theta\} = \{t : F(t;\theta) \geq \alpha_2\}, \tag{9.8}$$

όπου $F(t;\theta) = \mathbb{P}_\theta(T \leq t)$. Πράγματι, επειδή η $F(t;\theta)$ είναι γνησίως φθίνουσα ως προς θ, έχουμε

$$\theta_2(t) \geq \theta \Leftrightarrow F(t;\theta_2(t)) \leq F(t;\theta) \Leftrightarrow \mathbb{P}_{\theta_2(t)}(T \leq t) \leq F(t;\theta) \Leftrightarrow \alpha_2 \leq F(t;\theta),$$

λόγω της (9.6). Ανάλογα προκύπτει ότι

$$\{t : \theta_1(t) \leq \theta\} = \left\{t : \bar{F}(t;\theta) \geq \alpha_1\right\}, \tag{9.9}$$

όπου $\bar{F}(t;\theta) = \mathbb{P}_\theta(T \geq t)$. Επομένως λόγω των (9.8) και (9.9) έχουμε

$$\begin{aligned}
\mathbb{P}_\theta(\theta_1(T) \leq \theta \leq \theta_2(T)) &= \mathbb{P}_\theta(\theta \leq \theta_2(T)) - \mathbb{P}_\theta(\theta < \theta_1(T)) \\
&= \mathbb{P}_\theta(F(T;\theta) \geq \alpha_2) - 1 + \mathbb{P}_\theta(\theta \geq \theta_1(T)) \\
&= \mathbb{P}_\theta(F(T;\theta) \geq \alpha_2) - 1 + \mathbb{P}_\theta(\bar{F}(T;\theta) \geq \alpha_1) \\
&= -\mathbb{P}_\theta(F(T;\theta) < \alpha_2) + \mathbb{P}_\theta(\bar{F}(T;\theta) \geq \alpha_1)
\end{aligned} \tag{9.10}$$

Επειδή η συνάρτηση κατανομής $F(t;\theta)$ είναι αύξουσα ως προς t η σχέση $F(T;\theta) < \alpha_2$ είναι ισοδύναμη με $T \leq t_0$ για t_0 τέτοιο ώστε $\mathbb{P}_\theta(T \leq t_0) \leq \alpha_2$, βλέπε Σχήμα 9.4. Επομένως, έχουμε

$$\mathbb{P}_\theta(F(T;\theta) < \alpha_2) = \mathbb{P}_\theta(T \leq t_0) \leq \alpha_2. \tag{9.11}$$

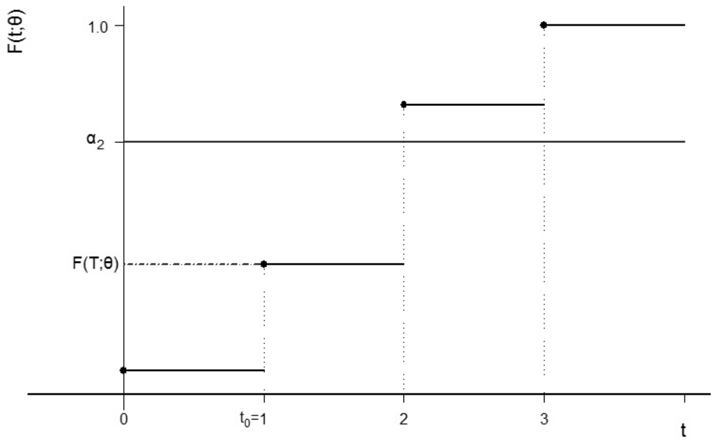
Ανάλογα, η συνάρτηση $\bar{F}(t;\theta)$ είναι φθίνουσα ως προς t, συνεπώς η σχέση $\bar{F}(T;\theta) \geq \alpha_1$ είναι ισοδύναμη με $T \leq t_1$ για t_1 τέτοιο ώστε $\mathbb{P}_\theta(T \leq t_1) \geq 1 - \alpha_1$. Επομένως, έχουμε

$$\mathbb{P}_\theta(\bar{F}(T;\theta) \geq \alpha_1) = \mathbb{P}_\theta(T \leq t_1) \geq 1 - \alpha_1. \tag{9.12}$$

Συνδυάζοντας τις (9.11) και (9.12), από την (9.10) παίρνουμε

$$\mathbb{P}_\theta(\theta_1(T) \leq \theta \leq \theta_2(T)) \geq 1 - \alpha_1 - \alpha_2 = 1 - \alpha, \quad \forall\, \theta \in \Theta.$$

Για την ολοκλήρωση της απόδειξης της (9.7) θα δικαιολογήσουμε τη σχέση



Σχήμα 9.4: $F(t;\theta) = \mathbb{P}_\theta(T \leq t)$ για $n = 3$

$\theta_1(t) < \theta_2(t)$ υπό τη συνθήκη $\alpha_1 + \alpha_2 < 1$. Έστω $\theta_1(t) \geq \theta_2(t)$. Τότε από τη μονοτονία της $\mathbb{P}_\theta(T \geq t)$ ως προς θ και την (9.6) έχουμε

$$\begin{aligned}\alpha_1 = \mathbb{P}_{\theta_1(t)}(T \geq t) \geq \mathbb{P}_{\theta_2(t)}(T \geq t) &= 1 - \mathbb{P}_{\theta_2(t)}(T < t)\\ &\geq 1 - \mathbb{P}_{\theta_2(t)}(T \leq t) = 1 - \alpha_2,\end{aligned}$$

δηλαδή προκύπτει $\alpha_1 \geq 1 - \alpha_2$, άτοπο.

Από την (9.6) συμπεραίνουμε ότι τα $\theta_1(t)$ και $\theta_2(t)$ είναι, αντίστοιχα, οι λύσεις ως προς θ των εξισώσεων

$$\sum_{x=t}^{n} \binom{n}{x} \theta^x (1-\theta)^{n-x} = \alpha_1 \quad \text{και} \quad \sum_{x=0}^{t} \binom{n}{x} \theta^x (1-\theta)^{n-x} = \alpha_2.$$

Για $n \geq 3$ και $t \in \{1, 2, \ldots, n-1\}$ τα $\theta_1(t)$ και $\theta_2(t)$ δεν είναι δυνατόν να βρεθούν σε κλειστή μορφή, μπορούν όμως να υπολογιστούν αριθμητικά χρησιμοποιώντας μια αριθμητική μέθοδο επίλυσης εξισώσεων (με δεδομένη την παρατηρηθείσα τιμή t της στατιστικής συνάρτησης $T = \sum_{i=1}^{n} X_i$).

Εναλλακτικά, τα $\theta_1(t)$ και $\theta_2(t)$ μπορούν να προσδιοριστούν με την ακόλουθη διαδικασία, που βασίζεται στη σχέση μεταξύ των κατανομών $\mathcal{B}(n,\theta)$, $\mathcal{B}eta(\alpha,\,\beta)$ και $\mathcal{F}_{p,q}$. Η T έχει διωνυμική κατανομή $\mathcal{B}(n,\theta)$ και επομένως από την Πρόταση 1.8.9, για $t \in \{1,2,\ldots,n\}$

$$\mathbb{P}_\theta(T \geq t) = \mathbb{P}(Y \leq \theta), \tag{9.13}$$

όπου η Y έχει Βήτα κατανομή $\mathcal{B}eta(t, n-t+1)$. Η (9.13) δείχνει ότι η πιθανότητα $\mathbb{P}_\theta(T \geq t)$ είναι γνησίως αύξουσα συνάρτηση ως προς θ αφού συμπίπτει με τη συνάρτηση κατανομής συνεχούς τυχαίας μεταβλητής. Επιπλέον, από την Πρόταση 1.8.9, αν η V έχει κατανομή $\mathcal{F}_{p,q}$ τότε η $\frac{(p/q)V}{1+(p/q)V}$ έχει κατανομή Βήτα $\mathcal{B}eta(p/2,\,q/2)$. Θέτοντας $t = p/2$ και $n-t+1 = q/2$ η Y έχει την ίδια κατανομή όπως η $\frac{(t/(n-t+1))V}{1+(t/(n-t+1))V}$, όπου V έχει κατανομή $\mathcal{F}_{2t,2(n-t+1)}$. Τότε από την (9.13) παίρνουμε

$$\begin{aligned}\mathbb{P}_\theta(T \geq t) = \mathbb{P}(Y \leq \theta) &= \mathbb{P}\Big(\frac{(t/(n-t+1))V}{1+(t/(n-t+1))V} \leq \theta\Big)\\ &= \mathbb{P}\Big(V \leq \frac{n-t+1}{t}\frac{\theta}{1-\theta}\Big).\end{aligned}$$

Τελικά, από την (9.6), $\theta_1(t)$ είναι η λύση ως προς θ της εξίσωσης

$$\mathbb{P}_\theta(T \geq t) = \mathbb{P}\Big(V \leq \frac{n-t+1}{t}\frac{\theta}{1-\theta}\Big) = \alpha_1. \tag{9.14}$$

Ορίζοντας το β-ποσοστιαίο σημείο της κατανομής $\mathcal{F}$, $\mathcal{F}_{p,q,\beta}$, από τη σχέση $\mathsf{P}\big(V > \mathcal{F}_{p,q,\beta}\big) = \beta$ (το οποίο μπορεί να βρεθεί σε διαθέσιμο πίνακα ποσοστιαίων σημείων της κατανομής $\mathcal{F}$ ή να υπολογιστεί από ένα στατιστικό λογισμικό), από την (9.14) παίρνουμε $\frac{n-t+1}{t}\frac{\theta}{1-\theta} = \mathcal{F}_{2t,2(n-t+1),1-\alpha_1}$, δηλαδή

$$\theta_1(t) = \frac{\frac{t}{n-t+1}\mathcal{F}_{2t,2(n-t+1),1-\alpha_1}}{1+\frac{t}{n-t+1}\mathcal{F}_{2t,2(n-t+1),1-\alpha_1}}.$$

Για τον υπολογισμό του $\theta_2(t)$, από την (9.6) παρατηρούμε ότι η σχέση $\mathbb{P}_\theta(T \leq t) = \alpha_2$ είναι ισοδύναμη με την $\mathbb{P}_\theta(T \geq t+1) = 1-\alpha_2$ και επομένως $\theta_2(t) = \theta_1(t+1)$ με α_1 το $1-\alpha_2$ (και $t \in \{1,2,\ldots,n-1\}$). Συνεπώς έχουμε

$$\theta_2(t) = \frac{\frac{t+1}{n-t}\mathcal{F}_{2(t+1),2(n-t),\alpha_2}}{1+\frac{t+1}{n-t}\mathcal{F}_{2(t+1),2(n-t),\alpha_2}}.$$

Τέλος αν η παρατηρηθείσα τιμή είναι $t = 0$ (ένδειξη ότι το θ είναι «μικρό»), το $\theta_1(t)$ δεν ορίζεται και συμβατικά θέτουμε $\theta_1(t) = 0$, ενώ αν $t = n$ (ένδειξη ότι το θ είναι «μεγάλο») το $\theta_2(t)$ δεν ορίζεται και συμβατικά θέτουμε $\theta_2(t) = 1$. Το διάστημα $[\theta_1(T), \theta_2(T)]$ κατασκευάστηκε από τους Clopper and Pearson (1934). □

Από την παραπάνω διαδικασία φαίνεται ξεκάθαρα, ότι η μονοτονία της συνάρτησης $\mathbb{P}_\theta(T \geq t)$ ως προς θ (όπου $T = \sum_{i=1}^{n} X_i$) παίζει καθοριστικό ρόλο στην κατασκευή του διαστήματος εμπιστοσύνης. Επιπλέον, για κάθε στατιστική συνάρτηση αν η συνάρτηση κατανομής $F(t;\theta) = \mathbb{P}_\theta(T \leq t)$ είναι συνεχής ως προς t (το οποίο ισχύει όταν η T έχει συνεχή κατανομή), τότε η τυχαία μεταβλητή $U = F(T;\theta)$ έχει ομοιόμορφη κατανομή $\mathcal{U}(0,\, 1)$, βλέπε Πρόταση 9.2.1, και επομένως

$$\mathbb{P}_\theta(F(T;\theta) < \alpha_2) = \mathbb{P}(U < \alpha_2) = \alpha_2. \tag{9.15}$$

Παρόμοια, έχουμε

$$\bar{F}(t;\theta) = \mathbb{P}_\theta(T \geq t) = 1 - \mathbb{P}_\theta(T < t) = 1 - F(t;\theta),$$

οπότε $\bar{F}(T;\theta) = 1 - F(T;\theta) = 1 - U$. Συνεπώς, παίρνουμε

$$\mathbb{P}_\theta(\bar{F}(T;\theta) \geq \alpha_1) = \mathbb{P}(1 - U \geq \alpha_1) = \mathbb{P}(U \leq 1 - \alpha_1) = 1 - \alpha_1. \tag{9.16}$$

Οι (9.15) και (9.16) σε συνδυασμό με την (9.10) δίνουν

$$\begin{aligned}\mathbb{P}_\theta(\theta_1(T) \leq \theta \leq \theta_2(T)) &= \mathbb{P}_\theta(\bar{F}(T;\theta) \geq \alpha_1) - \mathbb{P}_\theta(F(T;\theta) < \alpha_2)\\ &= 1 - \alpha_1 - \alpha_2 = 1 - \alpha,\end{aligned}$$

δηλαδή προκύπτει Δ.Ε. με σ.ε. (ακριβώς) $1 - \alpha$, αν δε επιλέξουμε $\alpha_1 = \frac{\alpha}{2} = \alpha_2$, τότε έχουμε κατασκευάσει Δ.Ε. ίσων ουρών.

Η επόμενη πρόταση καταγράφει την κατασκευή Δ.Ε. για την παράμετρο θ χρησιμοποιώντας στατιστική συνάρτηση με μονότονη, ως προς θ, συνάρτηση $\mathbb{P}_\theta(T \geq t)$ και σταθερό t. Αν και παρουσιάζεται μόνον η περίπτωση (γνησίως) αύξουσας συνάρτησης $\mathbb{P}_\theta(T \geq t)$ ανάλογο αποτέλεσμα ισχύει και για $\mathbb{P}_\theta(T \geq t)$ (γνησίως) φθίνουσα, βλέπε Casella and Berger (2002).

Πρόταση 9.3.1. *Έστω T στατιστική συνάρτηση τέτοια ώστε η $\mathbb{P}_\theta(T \geq t)$ είναι γνησίως αύξουσα συνάρτηση ως προς $\theta \in \Theta$, για κάθε t στο σύνολο τιμών της T (με εξαίρεση, ίσως, ορισμένων ακραίων τιμών). Υποθέτουμε ότι υπάρχουν $\theta_1(t)$ και $\theta_2(t)$ που ικανοποιούν τις σχέσεις*

$$\mathbb{P}_{\theta_1(t)}(T \geq t) = \alpha_1 \quad \textit{και} \quad \mathbb{P}_{\theta_2(t)}(T \leq t) = \alpha_2 \tag{9.17}$$

όπου α_1, α_2 θετικές σταθερές με $\alpha_1 + \alpha_2 < 1$. Τότε ισχύουν τα εξής.

(α) Το διάστημα $[\theta_1(T),\, \theta_2(T)]$ είναι Δ.Ε. με σ.ε. (τουλάχιστον) $100(1-\alpha)\%$, δηλαδή

$$\mathbb{P}_\theta(\theta_1(T) \leq \theta \leq \theta_2(T)) \geq 1 - \alpha, \quad \forall\, \theta \in \Theta. \tag{9.18}$$

(β) Αν, επιπλέον, η συνάρτηση κατανομής $F(t;\theta) = \mathbb{P}_\theta(T \leq t)$ είναι συνεχής ως προς t (για κάθε $\theta \in \Theta$), τότε η (9.18) ισχύει ως ισότητα.

Απόδειξη. (α) Όπως στο Παράδειγμα 9.3.1.
(β) Όπως η ανάλυση που έπεται του Παραδείγματος 9.3.1. □

Παρατήρηση 9.3.1. Αν η $\mathbb{P}_\theta(T \geq t)$ είναι συνεχής ως προς θ με σύνολο τιμών το $(0,1)$, όπως συμβαίνει στο Παράδειγμα 9.3.1, τότε υπάρχουν τα $\theta_1(t)$ και $\theta_2(t)$ που ικανοποιούν την (9.17) και είναι μοναδικά.

Δίνουμε ένα παράδειγμα εφαρμογής της Πρότασης 9.3.1 από συνεχή κατανομή.

Παράδειγμα 9.3.2. Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την κατανομή με πυκνότητα

$$f_1(x;\theta) = \frac{\theta}{x^2}, \quad x \geq \theta, \quad \theta \in \Theta = (0, \infty).$$

είναι εύκολο να δειχθεί ότι ο ε.μ.π. του θ είναι $X_{(1)}$. Θέτουμε $T = X_{(1)}$. Τότε, για $t > 0$,

$$\mathbb{P}_\theta(T \geq t) = \mathbb{P}_\theta(X_{(1)} \geq t) = \big[\mathbb{P}_\theta(X_1 \geq t)\big]^n = \left(\frac{\theta}{t}\right)^n, \quad 0 < \theta \leq t,$$

που είναι συνεχής και γνησίως αύξουσα συνάρτηση ως προς θ. Για $0 < \alpha < 1$, θέτοντας $\mathbb{P}_\theta(X_{(1)} \geq t) = \frac{\alpha}{2}$ έχουμε $\left(\frac{\theta}{t}\right)^n = \frac{\alpha}{2}$ με λύση ως προς θ, $\theta_1(t) = t\sqrt[n]{\frac{\alpha}{2}}$. Ανάλογα, θέτοντας $\mathbb{P}_\theta(X_{(1)} \leq t) = \frac{\alpha}{2}$ έχουμε $1 - \left(\frac{\theta}{t}\right)^n = \frac{\alpha}{2}$ με λύση ως προς θ, $\theta_2(t) = t\sqrt[n]{1-\frac{\alpha}{2}}$. Επομένως από την Πρόταση 9.3.1, το Δ.Ε. ίσων ουρών για το θ με σ.ε. $100(1-\alpha)\%$ είναι $\left[X_{(1)}\sqrt[n]{\frac{\alpha}{2}},\, X_{(1)}\sqrt[n]{1-\frac{\alpha}{2}}\right]$.

9.4 Εφαρμογές σε κανονικούς πληθυσμούς

Ως εφαρμογές της μεθόδου κατασκευής Δ.Ε. μέσω κατάλληλης ποσότητας οδηγού παρουσιάζουμε σε αυτήν την ενότητα διαστήματα εμπιστοσύνης για τις παραμέτρους κανονικών πληθυσμών $\mathcal{N}(\mu, \sigma^2)$. Αρχικά, δίνουμε την επόμενη πρόταση των Ferentinos and Kourouklis (1990) που είναι γενικά χρήσιμη στην κατασκευή διαστημάτων ελαχίστου μήκους (και όχι μόνον για τις παραμέτρους της κανονικής κατανομής). Θεωρούμε ότι τα ολοκληρώματα παρακάτω υπάρχουν και είναι πεπερασμένα. Σε πιο ειδική μορφή η πρόταση έχει αποδειχθεί και από τον Juola (1993).

Πρόταση 9.4.1. *Έστω f_1, f_2, δ πραγματικές συναρτήσεις με $0 \leqslant \delta(x) \leqslant 1$ για όλα τα x. Τότε, για δοθείσα πραγματική σταθερά c, το ολοκλήρωμα $\int_{-\infty}^{+\infty} \delta(x) f_1(x)\,\mathrm{d}x$ ελαχιστοποιείται ως προς δ υπό τον περιορισμό*

$$\int_{-\infty}^{+\infty} \delta(x) f_2(x)\,\mathrm{d}x = c \tag{9.19}$$

για

$$\delta^*(x) = \begin{cases} 1 & , \quad f_1(x) \leqslant \kappa f_2(x) \\ 0 & , \quad f_1(x) > \kappa f_2(x)\,, \end{cases}$$

όπου η σταθερά κ προσδιορίζεται έτσι ώστε να ικανοποιείται η (9.19) για την δ^*.

Απόδειξη. Η διαδικασία της απόδειξης είναι παρόμοια αυτής ενός κλασικού αποτελέσματος της Θεωρίας Ελέγχου Στατιστικών Υποθέσεων, του Λήμματος Neymann – Pearson (βλέπε, π.χ. Παπαϊωάννου και Φερεντίνος, 2001, σελ. 117, ή Casella and Berger, 2002, σελ. 388). Από τον ορισμό της δ^*, για κάθε συνάρτηση δ με $0 \leqslant \delta(x) \leqslant 1$ που ικανοποιεί την (9.19) και για όλα τα x ισχύει

$$\delta(x)\{f_1(x) - \kappa f_2(x)\} \geqslant \delta^*(x)\{f_1(x) - \kappa f_2(x)\}$$

και επομένως

$$\{\delta(x) - \delta^*(x)\}f_1(x) \geqslant \kappa\{\delta(x) - \delta^*(x)\}f_2(x).$$

Ολοκληρώνοντας τη τελευταία σχέση έχουμε

$$\int_{-\infty}^{+\infty} \delta(x) f_1(x)\,\mathrm{d}x - \int_{-\infty}^{+\infty} \delta^*(x) f_1(x)\,\mathrm{d}x \geqslant$$
$$\kappa\left\{\int_{-\infty}^{+\infty} \delta(x) f_2(x)\,\mathrm{d}x - \int_{-\infty}^{+\infty} \delta^*(x) f_2(x)\,\mathrm{d}x\right\} = \kappa\{c - c\} = 0.$$

Επομένως

$$\int_{-\infty}^{+\infty} \delta(x) f_1(x)\,\mathrm{d}x \geqslant \int_{-\infty}^{+\infty} \delta^*(x) f_1(x)\,\mathrm{d}x$$

και αυτό έπρεπε να δείξουμε. □

9.4.1 Ένας κανονικός πληθυσμός

Διάστημα εμπιστοσύνης για τη μέση τιμή

1η Περίπτωση: σ^2 γνωστό. Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την κανονική κατανομή $\mathcal{N}(\theta, \sigma^2)$ με θ άγνωστο, $\theta \in \Theta = \mathbb{R}$ και σ^2 γνωστή σταθερά. Η στατιστική συνάρτηση (και εκτιμητής του θ) $\bar{X}$ έχει κατανομή $\mathcal{N}(\theta, \frac{\sigma^2}{n})$, Πρόταση 1.8.6(1), επομένως από την Πρόταση 1.6.1(2), προκύπτει η

$$T = \frac{\bar{X} - \theta}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0, 1). \tag{9.20}$$

Άρα η τυχαία μεταβλητή T είναι ποσότητα οδηγός αφού η κατανομή της, $\mathcal{N}(0, 1)$, προφανώς δεν εξαρτάται από την άγνωστη παράμετρο θ. Επομένως, για κάθε α, $0 < \alpha < 1$ υπάρχουν σταθερές $c_1 < c_2$ που εξαρτώνται από την κατανομή $\mathcal{N}(0, 1)$ και το α, έτσι ώστε

$$\mathbb{P}_\theta\big(c_1 \leq T \leq c_2\big) = 1 - \alpha, \ \forall\, \theta \in \Theta \tag{9.21}$$

ή

$$\mathbb{P}_\theta\Big(c_1 \leq \frac{\bar{X}-\theta}{\frac{\sigma}{\sqrt{n}}} \leq c_2\Big) = 1-\alpha,\ \forall\, \theta \in \Theta$$

ή

$$\mathbb{P}_\theta\Big(\bar{X} - c_2\frac{\sigma}{\sqrt{n}} \leq \theta \leq \bar{X} - c_1\frac{\sigma}{\sqrt{n}}\Big) = 1-\alpha,\ \forall\, \theta \in \Theta. \tag{9.22}$$

Επομένως το διάστημα $[\bar{X} - c_2\sigma/\sqrt{n}, \bar{X} - c_1\sigma/\sqrt{n}]$ είναι Δ.Ε. για το θ με σ.ε. $100(1-\alpha)\%$. Είναι προφανές τώρα ότι υπάρχουν άπειρα ζεύγη (c_1, c_2) που ικανοποιούν την (9.21), άρα και την (9.22). Πράγματι, η (9.21) είναι ισοδύναμη προς την

$$\mathbb{P}_\theta(T < c_1) + \mathbb{P}_\theta(T > c_2) = \alpha$$

ή την

$$\Phi(c_1) + 1 - \Phi(c_2) = \alpha, \tag{9.23}$$

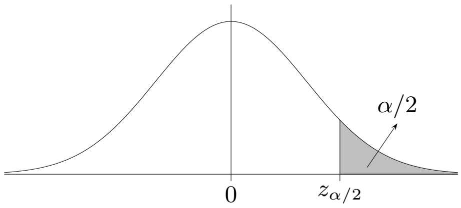
όπου Φ είναι η συνάρτηση κατανομής της τυπικής κανονικής $\mathcal{N}(0,1)$), και άρα η (9.21) ισχύει εάν για οποιοδήποτε α_1, $0 < \alpha_1 < \alpha$, ορίσουμε $c_1 = \Phi^{-1}(\alpha_1)$ και $c_2 = \Phi^{-1}\big(1-(\alpha-\alpha_1)\big)$. Ειδικά για $\alpha_1 = \alpha/2$ προκύπτει Δ.Ε. ίσων ουρών με $c_2 = \Phi^{-1}(1-\alpha/2)$ και $c_1 = -c_2$, λόγω συμμετρίας. Θέτοντας περαιτέρω (για λόγους συμβολισμού) $z_{\alpha/2} = \Phi^{-1}(1-\alpha/2)$, δηλαδή ορίζοντας το σημείο $z_{\alpha/2}$, βλέπε Σχήμα 9.5, από τη σχέση

$$\mathbb{P}(Z > z_{\alpha/2}) = \alpha/2,$$

έχουμε $c_2 = z_{\alpha/2}$ και $c_1 = -z_{\alpha/2}$, οπότε το Δ.Ε. ίσων ουρών για το θ γίνεται

$$\Big[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\Big].$$

Από τα παραπάνω, προκύπτει επίσης η σχέση $\mathbb{P}_\theta(|\bar{X}-\theta| \leqslant z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1-\alpha$, για κάθε $\theta \in \Theta$ που δηλώνει ότι η εκτίμηση του θ με τον εκτιμητή $\bar{X}$ έχει, με πιθανότητα $1-\alpha$, σφάλμα το πολύ ίσο προς $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$. Υπό το



Σχήμα 9.5: $\alpha/2$–ποσοστιαίο σημείο της $\mathcal{N}(0,1)$.

πρίσμα αυτής της παρατήρησης, το διάστημα αυτό μπορεί να γραφεί στη συμβολική μορφή

$$\bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Θα δείξουμε στη συνέχεια ότι το παραπάνω διάστημα έχει ελάχιστο μήκος μεταξύ όλων των διαστημάτων της μορφής $[\bar{X} - c_2\frac{\sigma}{\sqrt{n}}, \bar{X} - c_1\frac{\sigma}{\sqrt{n}}]$ με συντελεστή εμπιστοσύνης $100(1-\alpha)\%$. Πράγματι, το μήκος ενός τέτοιου διαστήματος είναι $\ell = \frac{\sigma}{\sqrt{n}}(c_2 - c_1)$, όπου τα c_1, c_2 ικανοποιούν τη συνθήκη (9.23), που ισοδύναμα γράφεται

$$\int_{c_1}^{c_2} \phi(x)\,\mathrm{d}x = 1 - \alpha \tag{9.24}$$

ή

$$\int_{-\infty}^{+\infty} I_{[c_1,c_2]}(x)\phi(x)\,\mathrm{d}x = \int_{-\infty}^{+\infty} \delta(x)\phi(x)\,\mathrm{d}x = 1 - \alpha \tag{9.25}$$

όπου $\delta(x) = I_{[c_1,c_2]}(x)$ είναι η δείκτρια συνάρτηση του $[c_1, c_2]$ και $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ είναι η πυκνότητα της $\mathcal{N}(0,1)$. Συνεπώς ελαχιστοποίηση του μήκους σημαίνει ελαχιστοποίηση του $\ell^* = c_2 - c_1$ υπό τον περιορισμό (9.25).

Περαιτέρω, παρατηρούμε ότι $\ell^* = c_2 - c_1 = \int_{c_1}^{c_2} \mathrm{d}x$, δηλαδή

$$\ell^* = \int_{-\infty}^{+\infty} I_{[c_1,c_2]}(x)\,\mathrm{d}x = \int_{-\infty}^{+\infty} \delta(x)\,\mathrm{d}x \tag{9.26}$$

Τελικά, ζητάμε ελαχιστοποίηση, ως προς τη συνάρτηση $\delta(x) = I_{[c_1,c_2]}(x)$ του ολοκληρώματος (9.26) υπό τον περιορισμό (9.25). Η λύση αυτού του προβλήματος ελαχιστοποίησης δίνεται από τη συνάρτηση $\delta^*(x) =$

$I_{\{x:f_1(x)\leqslant\kappa f_2(x)\}}(x)$ της Πρότασης 9.4.1 με $f_1(x) = 1$ και $f_2(x) = \phi(x)$ εφ' όσον η $\delta^*(x)$ είναι δείκτρια πεπερασμένου διαστήματος, δηλαδή εφ' όσον το σύνολο $\{x : 1 \leqslant \kappa\phi(x)\}$ είναι διάστημα της μορφής $[c_1, c_2]$. Πράγματι, η σχέση $1 \leqslant \kappa\phi(x)$ σημαίνει $\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \geqslant \frac{1}{\kappa}$ που είναι ισοδύναμη με τη σχέση $|x| \leqslant c$ επειδή η συνάρτηση $e^{-\frac{x^2}{2}}$ είναι γνησίως φθίνουσα ως προς $|x|$. (Τα c και κ ικανοποιούν την $\frac{1}{\sqrt{2\pi}}e^{-\frac{c^2}{2}} = \frac{1}{\kappa}$ που όμως δεν έχει καμία χρησιμότητα.) Καταλήγουμε λοιπόν ότι $\delta^*(x) = I_{\{x:|x|\leqslant c\}}(x) = I_{[-c,c]}(x)$, άρα η συνάρτηση ℓ^* ελαχιστοποιείται όταν $c_2 = c$, $c_1 = -c$ και η σταθερά c ικανοποιεί τον περιορισμό (9.24), $\int_{-c}^{c} \phi(x)\,\mathrm{d}x = 1 - \alpha$. Η τελευταία σχέση συνεπάγεται

$$\begin{aligned}
\Phi(c) - \Phi(-c) &= 1 - \alpha \Rightarrow \quad (\text{επειδή } \Phi(x) = 1 - \Phi(-x)) \\
\Phi(c) - \{1 - \Phi(c)\} &= 1 - \alpha \Rightarrow \\
\Phi(c) &= 1 - \alpha/2 \Rightarrow \\
c &= \Phi^{-1}(1 - \alpha/2) = z_{\alpha/2}.
\end{aligned}$$

Επομένως το Δ.Ε. ελαχίστου μήκους έχει $c_2 = c = z_{\alpha/2}$, $c_1 = -c = -z_{\alpha/2}$ και όντως συμπίπτει με το Δ.Ε. ίσων ουρών.

Γεωμετρικά, η επιλογή $c_1 = -c_2$ ώστε να ελαχιστοποιηθεί η διαφορά $\ell^* = c_2 - c_1$ υπό τον περιορισμό $\int_{c_1}^{c_2} \phi(x)\,\mathrm{d}x = 1 - \alpha$ είναι σχεδόν προφανής λόγω συμμετρίας της $\phi(x)$ και επειδή η τυπική κανονική κατανομή συγκεντρώνει περισσότερη πιθανότητα γύρω από το 0 (την κορυφή της). Σημειώνουμε ακόμη ότι το διάστημα

$$\left[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

έχει την επιπλέον επιθυμητή ιδιότητα να βασίζεται στην (ελάχιστη) επαρκή στατιστική συνάρτηση $\bar{X}$. Τέλος αναφέρουμε ότι η τιμή του ποσοστιαίου σημείου $z_{\alpha/2}$ για διάφορες τιμές του α μπορεί να βρεθεί από πίνακες της κατανομής $\mathcal{N}(0, 1)$.

Παράδειγμα 9.4.1. Ένα τυχαίο δείγμα 25 παρατηρήσεων από την κατανομή $\mathcal{N}(\theta, 100)$ έδωσε $\bar{X} = 73.2$. Θα υπολογίσουμε το Δ.Ε. ελαχίστου

μήκους για το θ με σ.ε. 90%.
Έχουμε $\sigma = 10$ και $1-\alpha = 0.9$, οπότε $\alpha = 0.1$. Από πίνακες της $\mathcal{N}(0,1)$ έχουμε $z_{\alpha/2} = z_{0.05} = 1.645$ (σε προσέγγιση χιλιοστού). Άρα το Δ.Ε. είναι

$$\Big[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\Big]$$
$$= \big[73.2 - 1.645 \times \frac{10}{5}, 73.2 + 1.645 \times \frac{10}{5}\big]$$
$$= 73.2 \pm 2.29 = [69.91, 76.49].$$

Λαμβάνοντας υπόψη ότι η μέση τιμή θ εκτιμάται με το δειγματικό μέσο $\bar{X}$, από τα δεδομένα προκύπτει λοιπόν εκτίμηση του θ ίση προς 73.2 με σφάλμα το πολύ 2.29, σε βαθμό εμπιστοσύνης 90% (και υπό το πρίσμα της Παρατήρησης 9.1.1).

2η Περίπτωση: σ^2 άγνωστο. Υπενθυμίζουμε αρχικά τον ορισμό της κατανομής t *με* m *βαθμούς ελευθερίας*, που δόθηκε στην Ενότητα 1.6.2 και την Πρόταση 1.8.5, η οποία θα χρησιμοποιηθεί στην κατασκευή του Δ.Ε.

Ορισμός 9.4.1. *Έστω Z μία τυχαία μεταβλητή με κατανομή $\mathcal{N}(0,1)$ και Y μία τυχαία μεταβλητή με κατανομή χ^2_m. Εάν Z και Y είναι ανεξάρτητες, τότε η κατανομή της τυχαίας μεταβλητής*

$$T = \frac{Z}{\sqrt{\frac{Y}{m}}}$$

ονομάζεται κατανομή t με m βαθμούς ελευθερίας . Συμβολικά γράφουμε $T \sim t_m$ ή ακόμη $t_m = \frac{\mathcal{N}(0,1)}{\sqrt{\frac{\chi^2_m}{m}}}$.

Μπορεί να δειχθεί ότι η πυκνότητα της κατανομής t_m είναι

$$f(t) = \frac{\Gamma(\frac{m+1}{2})}{\sqrt{\pi m}\,\Gamma(\frac{m}{2})\big(1 + \frac{t^2}{m}\big)^{\frac{m+1}{2}}}, \ t \in \mathbb{R}.$$

Η γραφική παράσταση της $f(t)$ είναι «καμπανοειδής», όπως της $\mathcal{N}(0,1)$, αλλά με πιο «βαριές ουρές» δηλαδή η $f(t)$ συγκλίνει στο μηδέν για $t \to \pm\infty$, πιο αργά όμως από την $\phi(t)$, (βλέπε Σχήμα 9.6.)

Μερικές βασικές ιδιότητες της κατανομής t_m είναι οι εξής.

1. Η κατανομή είναι συμμετρική γύρω από το μηδέν, αφού $f(t) = f(-t)$.
2. Για $m \to \infty$ η κατανομή t_m συγκλίνει προς την κατανομή $\mathcal{N}(0, 1)$. Πρακτικά, για $m > 30$ η κατανομή t_m πολύ λίγο διαφέρει από την κατανομή $\mathcal{N}(0, 1)$.
3. $\mathbb{E}T = 0$ για $m \geq 2$.
4. $\mathbb{V}arT = \frac{m}{m-2}$ για $m \geq 3$.
5. Για $m = 1$, η t_1 συμπίπτει με την κατανομή Cauchy (της οποίας δεν υπάρχει η μέση τιμή).

Κατασκευάζουμε τώρα Δ.Ε. για το μ με δεδομένο σ.ε. $100(1 - \alpha)\%$. Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$, $n \geq 2$, ένα τυχαίο δείγμα από την κανονική κατανομή $\mathcal{N}(\mu, \sigma^2)$ με $\mu \in \mathbb{R}$ και $\sigma^2 > 0$ άγνωστα. Θεωρούμε την τυχαία μεταβλητή

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}, \tag{9.27}$$

όπου $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$ είναι (ο ΑΟΕΔ) εκτιμητής του σ^2. Η τυχαία μεταβλητή T μπορεί να θεωρηθεί ότι προκύπτει από την (9.20) εάν αντικαταστήσουμε το άγνωστο τώρα σ με τον εκτιμητή του S. Γράφοντας την τυχαία μεταβλητή T στη μορφή

$$T = \frac{\sqrt{n}(\bar{X} - \mu)/\sigma}{\sqrt{S^2/\sigma^2}} = \frac{\sqrt{n}(\bar{X} - \mu)/\sigma}{\sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2/\sigma^2}}$$

και χρησιμοποιώντας τις Προτάσεις 1.8.6 και 1.6.1 παρατηρούμε ότι η τυχαία μεταβλητή $Z = \sqrt{n}(\bar{X} - \mu)/\sigma$ έχει κατανομή $\mathcal{N}(0, 1)$, η τυχαία μεταβλητή $Y = \sum_{i=1}^{n} (X_i - \bar{X})^2/\sigma^2$ έχει κατανομή χ^2_{n-1} και οι τυχαίες μεταβλητές Z, Y είναι ανεξάρτητες. Άρα η τυχαία μεταβλητή T στην (9.27) έχει κατανομή t_{n-1}, δηλαδή

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$$

επομένως είναι ποσότητα οδηγός, αφού η κατανομή της προφανώς δεν εξαρτάται από την άγνωστη παράμετρο $\theta = (\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)$. Άρα υπάρχουν σταθερές $c_1 < c_2$ που εξαρτώνται από την κατανομή t_{n-1} και το α, έτσι ώστε

$$\mathbb{P}_\theta\big(c_1 \le T \le c_2\big) = 1 - \alpha, \ \forall\, \theta \in \Theta \tag{9.28}$$

ή

$$\mathbb{P}_\theta\left(c_1 \le \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \le c_2\right) = 1 - \alpha, \ \forall\, \theta \in \Theta$$

ή

$$\mathbb{P}_\theta\Big(\bar{X} - c_2\frac{S}{\sqrt{n}} \le \mu \le \bar{X} - c_1\frac{S}{\sqrt{n}}\Big) = 1 - \alpha, \ \forall\, \theta \in \Theta. \tag{9.29}$$

Επομένως, το διάστημα $[\bar{X} - c_2 S/\sqrt{n}, \bar{X} - c_1 S/\sqrt{n}]$ είναι Δ.Ε. για το μ με σ.ε. $100(1-\alpha)\%$. Όπως στην περίπτωση με σ^2 γνωστό, είναι προφανές ότι υπάρχουν άπειρα ζεύγη (c_1, c_2) που ικανοποιούν την (9.28), άρα και την (9.29). Για κάθε τέτοιο ζεύγος έχουμε

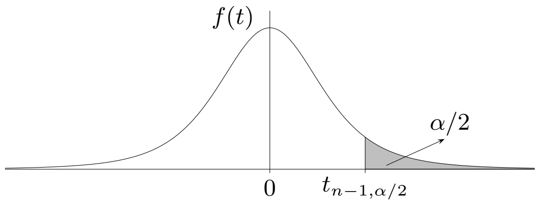
$$\mathbb{P}_\theta(T < c_1) + \mathbb{P}_\theta(T > c_2) = \alpha$$

ή

$$F(c_1) + 1 - F(c_2) = \alpha,$$

όπου F είναι η συνάρτηση κατανομής της t_{n-1}. Ειδικά, για Δ.Ε. ίσων ουρών θέτουμε $F(c_1) = \alpha/2 = 1 - F(c_2)$, οπότε λόγω της συμμετρίας της κατανομής t_{n-1}, έχουμε $c_1 = -c_2 = -F^{-1}(1-\alpha/2) = F^{-1}(\alpha/2)$. Θέτοντας (για λόγους συμβολισμού) $t_{n-1,\alpha/2} = F^{-1}(1-\alpha/2)$, βλέπε Σχήμα 9.6, δηλαδή ορίζοντας το σημείο $t_{n-1,\alpha/2}$ από τη σχέση

$$\mathbb{P}(t_{n-1} > t_{n-1,\alpha/2}) = \alpha/2 \ ,$$



Σχήμα 9.6: $\alpha/2$–ποσοστιαίο σημείο της t_{n-1}.

έχουμε $c_2 = t_{n-1,\alpha/2}$ και $c_1 = -t_{n-1,\alpha/2}$, οπότε το Δ.Ε. ίσων ουρών για το μ είναι

$$\left[\bar{X} - t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}, \bar{X} + t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}\right].$$

Ακολουθώντας την ίδια διαδικασία όπως στην περίπτωση με σ^2 γνωστό και λαμβάνοντας υπόψη τη συμμετρία της κατανομής t_{n-1} καθώς και ότι η πυκνότητα της $f(t)$, είναι γνησίως φθίνουσα ως προς $|t|$ είναι εύκολο να δειχθεί ότι το παραπάνω διάστημα έχει επίσης ελάχιστο μήκος μεταξύ όλων των διαστημάτων της μορφής $[\bar{X} - c_2 S/\sqrt{n}, \bar{X} - c_1 S/\sqrt{n}]$ με σ.ε. $100(1-\alpha)\%$. Σημειώνουμε ακόμη ότι το διάστημα αυτό έχει την επιπλέον επιθυμητή ιδιότητα να βασίζεται στην (ελάχιστη) επαρκή στατιστική συνάρτηση $(\bar{X}, S^2)$. Τέλος αναφέρουμε ότι η τιμή του $t_{n-1,\alpha/2}$ για διάφορες τιμές του n και του α μπορεί να βρεθεί από πίνακες της κατανομής t_m και ότι για $n > 30$ είναι $t_{n-1,\alpha/2} \approx z_{\alpha/2}$.

Παράδειγμα 9.4.2. Ένα τυχαίο δείγμα 15 παρατηρήσεων από την κατανομή $\mathcal{N}(\mu, \sigma^2)$ έδωσε $\bar{X} = 27.38$ και $S^2 = 5.10$. Θα βρούμε το Δ.Ε. ελαχίστου μήκους για το μ με σ.ε. 90%.
Έχουμε $1-\alpha = 0.9$ οπότε $\alpha = 0.1$. Από πίνακες της κατανομής t_m παίρνουμε $t_{n-1,\alpha/2} = t_{14,0.05} = 1.761$ (σε προσέγγιση χιλιοστού). Άρα το Δ.Ε.

είναι

$$\begin{aligned}&\left[\bar{X}-t_{n-1,\alpha/2}\frac{S}{\sqrt{n}},\bar{X}+t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}\right]\\&\qquad=\left[27.38-1.761\times\frac{\sqrt{5.10}}{\sqrt{15}},27.38+1.761\times\frac{\sqrt{5.10}}{\sqrt{15}}\right]\\&\qquad\qquad=27.38\pm 1.03=[26.35,28.41].\end{aligned}$$

Από τα δεδομένα λοιπόν προκύπτει εκτίμηση του μ ίση προς 27.38 με όρια κύμανσης από 26.35 έως 28.41, σε βαθμό εμπιστοσύνης 90% (και υπό το πρίσμα της Παρατήρησης 9.1.1).

Διάστημα εμπιστοσύνης για τη διασπορά

1η Περίπτωση: $\boldsymbol{\mu}$ γνωστό. Έστω $\underset{\sim}{X}=(X_1,\ldots,X_n)$ ένα τυχαίο δείγμα από την κανονική κατανομή $\mathcal{N}(\mu,\theta^2)$ με θ άγνωστο, $\theta\in\Theta=(0,\infty)$ και μ γνωστή σταθερά.
Θεωρούμε τον (ΑΟΕΔ και μέγιστης πιθανοφάνειας) εκτιμητή του θ^2

$$\hat{\theta}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$$

και παρατηρούμε ότι η τυχαία μεταβλητή

$$T=\frac{n\hat{\theta}^2}{\theta^2}=\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\theta^2}$$

έχει κατανομή χ^2_n (βλέπε Πρόταση 1.8.6), δηλαδή

$$T=\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\theta^2}\sim\chi^2_n.$$

Άρα η τυχαία μεταβλητή T είναι ποσότητα οδηγός αφού η κατανομή της προφανώς δεν εξαρτάται από την άγνωστη παράμετρο θ. Συνεπώς, υπάρχουν σταθερές $c_1<c_2$ που εξαρτώνται από την κατανομή χ^2_n και το α, $0<\alpha<1$, έτσι ώστε

$$\mathbb{P}_\theta\big(c_1\leq T\leq c_2\big)=1-\alpha,\ \forall\,\theta\in\Theta \tag{9.30}$$

ή

$$\mathbb{P}_\theta\bigg(c_1 \le \frac{\sum_{i=1}^n (X_i-\mu)^2}{\theta^2} \le c_2\bigg) = 1-\alpha,\ \forall\, \theta\in\Theta$$

ή

$$\mathbb{P}_\theta\bigg(\frac{\sum_{i=1}^n (X_i-\mu)^2}{c_2} \le \theta^2 \le \frac{\sum_{i=1}^n (X_i-\mu)^2}{c_1}\bigg) = 1-\alpha,\ \forall\, \theta\in\Theta. \quad (9.31)$$

Επομένως, το διάστημα $\big[\sum_{i=1}^n (X_i-\mu)^2/c_2, \sum_{i=1}^n (X_i-\mu)^2/c_1\big]$ είναι Δ.Ε. για το θ^2 με σ.ε. $100(1-\alpha)\%$. Είναι προφανές ότι υπάρχουν άπειρα ζεύγη (c_1, c_2) που ικανοποιούν την (9.30), άρα και την (9.31). Για κάθε τέτοιο ζεύγος έχουμε

$$\mathbb{P}_\theta(T < c_1) + \mathbb{P}_\theta(T > c_2) = \alpha, \quad \forall\, \theta\in\Theta$$

ή

$$F(c_1) + 1 - F(c_2) = \alpha, \quad \forall\, \theta\in\Theta, \quad (9.32)$$

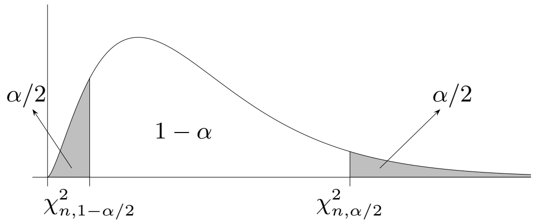
όπου F είναι η συνάρτηση κατανομής της χ^2_n. Ειδικά επιλέγοντας τα c_1, c_2 έτσι ώστε $F(c_1) = \alpha/2$ και $1 - F(c_2) = \alpha/2$ το διάστημα που προκύπτει είναι *διάστημα εμπιστοσύνης ίσων ουρών* για το θ^2. Χρησιμοποιώντας το συμβολισμό του Παραδείγματος 9.2.1 $\chi^2_{n,\alpha/2} = F^{-1}(1-\alpha/2)$, δηλαδή ορίζοντας το σημείο $\chi^2_{n,\alpha/2}$ από τη σχέση

$$\mathbb{P}(\chi^2_n > \chi^2_{n,\alpha/2}) = \alpha/2,$$

έχουμε $c_2 = \chi^2_{n,\alpha/2}$ και $c_1 = \chi^2_{n,1-\alpha/2}$, βλέπε Σχήμα 9.7, οπότε το Δ.Ε. ίσων ουρών για το θ^2 γίνεται

$$\bigg[\frac{\sum_{i=1}^n (X_i-\mu)^2}{\chi^2_{n,\alpha/2}}, \frac{\sum_{i=1}^n (X_i-\mu)^2}{\chi^2_{n,1-\alpha/2}}\bigg].$$

Σημειώνουμε ότι η τιμή του $\chi^2_{n,\alpha/2}$ για διάφορες τιμές του n και του α μπορεί να βρεθεί από πίνακες της κατανομής χ^2_n.



Σχήμα 9.7: $\alpha/2$ και $1-\alpha/2$ ποσοστιαία σημεία της χ^2_n.

Στη συνέχεια θα αναζητήσουμε Δ.Ε. ελαχίστου μήκους για το θ^2 με σ.ε. $100(1-\alpha)\%$. Το μήκος του Δ.Ε. είναι $\ell = \sum_{i=1}^{n}(X_i-\mu)^2\big(\frac{1}{c_1}-\frac{1}{c_2}\big)$, όπου τα c_1, c_2 ικανοποιούν τη συνθήκη (9.32). Συνεπώς ελαχιστοποίηση του μήκους σημαίνει ελαχιστοποίηση του $\ell^* = \frac{1}{c_1}-\frac{1}{c_2}$ υπό τον περιορισμό $F(c_2)-F(c_1)=1-\alpha$, που ισοδύναμα γράφεται

$$\int_{c_1}^{c_2} f(x)\,\mathrm{d}x = 1-\alpha \tag{9.33}$$

ή

$$\int_{-\infty}^{+\infty} I_{[c_1,c_2]}(x)f(x)\,\mathrm{d}x = \int_{-\infty}^{+\infty} \delta(x)f(x)\,\mathrm{d}x = 1-\alpha \tag{9.34}$$

όπου $f(x)$ είναι η πυκνότητα της χ^2_n και $\delta(x) = I_{[c_1,c_2]}(x)$. Περαιτέρω παρατηρούμε ότι $\ell^* = \frac{1}{c_1}-\frac{1}{c_2} = \int_{c_1}^{c_2}\frac{1}{x^2}\,\mathrm{d}x$, δηλαδή

$$\ell^* = \int_{-\infty}^{+\infty} I_{[c_1,c_2]}(x)\frac{1}{x^2}\,\mathrm{d}x = \int_{-\infty}^{+\infty} \delta(x)\frac{1}{x^2}\,\mathrm{d}x. \tag{9.35}$$

Τελικά, ζητάμε ελαχιστοποίηση του ολοκληρώματος (9.35), ως προς τη συνάρτηση $\delta(x) = I_{[c_1,c_2]}(x)$, υπό τον περιορισμό (9.34). Η λύση αυτού του προβλήματος ελαχιστοποίησης δίνεται από τη συνάρτηση $\delta^*(x) = I_{\{x: f_1(x)\leqslant \kappa f_2(x)\}}(x)$ της Πρότασης 9.4.1 με $f_1(x) = \frac{1}{x^2}$ και $f_2(x) = f(x)$ εφ' όσον το σύνολο $\{x : 1/x^2 \leqslant \kappa f(x)\}$ είναι διάστημα της μορφής $[c_1, c_2]$. Πράγματι, η σχέση $\frac{1}{x^2} \leqslant \kappa f(x)$ είναι ισοδύναμη με την $x^2 f(x) \geqslant \frac{1}{\kappa}$. Αντικαθιστώντας $f(x) = \frac{1}{\Gamma(n/2)2^{n/2}}x^{n/2-1}e^{-x/2}$, προκύπτει η σχέση

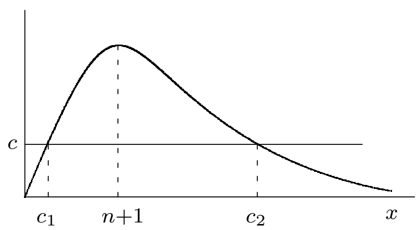
$$x^{n/2+1}e^{-x/2} \geqslant c \tag{9.36}$$

όπου $c = \Gamma(n/2)2^{n/2}/\kappa$. Είναι εύκολο να δειχθεί ότι η συνάρτηση $x^{n/2+1}e^{-x/2}$ είναι γνησίως αύξουσα στο $(0, n+1]$ και γνησίως φθίνουσα στο $[n+1, \infty)$.

Επομένως, η (9.36) είναι ισοδύναμη με την $c_1 \leqslant x \leqslant c_2$, όπου τα c_1 και c_2 ικανοποιούν τη σχέση

$$c_1^{n/2+1} e^{-c_1/2} = c_2^{n/2+1} e^{-c_2/2} (= c). \tag{9.37}$$

(βλέπε Σχήμα 9.8.) Καταλήγουμε λοιπόν ότι $\delta^* = I_{[c_1,c_2]}(x)$, άρα η συνάρτηση $\ell^* = \frac{1}{c_1} - \frac{1}{c_2}$ ελαχιστοποιείται όταν τα c_1, c_2 ικανοποιούν το σύστημα των εξισώσεων (9.33) και (9.37). Θα δείξουμε στη συνέχεια ότι το σύστημα αυτό έχει μία και μοναδική λύση. Για κάθε $c_1 \in (0, n+1)$ υπάρχει ακριβώς ένα $c_2 \in (n+1, \infty)$ ώστε να ισχύει η (9.37) και μάλιστα αυτό το c_2 είναι γνησίως φθίνουσα συνάρτηση του c_1, βλέπε Σχήμα 9.8. Ως εκ τούτου το ολοκλήρωμα $\int_{c_1}^{c_2} f(x)\,\mathrm{d}x$ είναι γνησίως φθίνουσα συνάρτηση του c_1. Επιπλέον είναι συνεχής με σύνολο τιμών το $[0,1]$. Άρα για την ενδιάμεση τιμή $1-\alpha \in [0,1]$ υπάρχει ακριβώς ένα $c_1 \in (0, n+1)$ με αντίστοιχο c_2 από την (9.37) ώστε $\int_{c_1}^{c_2} f(x)\,\mathrm{d}x = 1-\alpha$. Το σύστημα των (9.33) και (9.37) μπορεί να λυθεί μόνον με αριθμητικές μεθόδους. Η λύση του καθορίζει τα c_1, c_2 και εν συνεχεία το Δ.Ε. ελαχίστου μήκους για το σ^2. Τέλος αναφέρουμε ότι και τα δύο διαστήματα έχουν την επιθυμητή ιδιότητα να εξαρτώνται από τα δεδομένα $\underset{\sim}{X}$ μέσω της (ελάχιστης) επαρκούς στατιστικής συνάρτησης $\sum\limits_{i=1}^{n}(X_i-\mu)^2$.



Σχήμα 9.8: Γραφική παράσταση της $x^{n/2+1}e^{-x/2}$.

Παρατήρηση 9.4.1. Από την (9.31) έχουμε

$$\mathbb{P}_\theta\left(\left(\frac{\sum_{i=1}^n (X_i-\mu)^2}{c_2}\right)^{1/2} \leq \theta \leq \left(\frac{\sum_{i=1}^n (X_i-\mu)^2}{c_1}\right)^{1/2}\right) = 1-\alpha,\ \forall\, \theta \in \Theta$$

και επομένως το διάστημα

$$\left[\sqrt{\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{c_2}},\ \sqrt{\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{c_1}}\right]$$

είναι το Δ.Ε. για την τυπική απόκλιση θ με σ.ε. $100(1-\alpha)\%$. Το αντίστοιχο διάστημα ίσων ουρών είναι

$$\left[\sqrt{\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{n,\alpha/2}}},\ \sqrt{\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{n,1-\alpha/2}}}\right].$$

2η Περίπτωση: μ άγνωστο. Έστω $\underset{\sim}{X} = (X_1,\ldots,X_n)$, $n \geq 2$, ένα τυχαίο δείγμα από την κανονική κατανομή $\mathcal{N}(\mu,\sigma^2)$ με μ,σ^2 άγνωστα, $\theta = (\mu,\sigma^2) \in \Theta = \mathbb{R}\times(0,\infty)$.
Θεωρούμε τον (ΑΟΕΔ) εκτιμητή του σ^2, $S^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\bar{X})^2$ και παρατηρούμε ότι η τυχαία μεταβλητή

$$T = \frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2}$$

έχει κατανομή χ^2_{n-1} (βλέπε Πρόταση (1.8.6)), δηλαδή

$$T = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Άρα η τυχαία μεταβλητή T είναι ποσότητα οδηγός αφού η κατανομή της προφανώς δεν εξαρτάται από την άγνωστη παράμετρο σ^2. Επομένως υπάρχουν σταθερές $c_1 < c_2$ που εξαρτώνται από την κατανομή χ^2_{n-1} και το α, $0<\alpha<1$, έτσι ώστε

$$\mathbb{P}_\theta\big(c_1 \leq T \leq c_2\big) = 1-\alpha,\ \forall\, \theta\in\Theta$$

ή

$$\mathbb{P}_\theta\left(c_1 \leq \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2} \leq c_2\right) = 1-\alpha,\ \forall\, \theta\in\Theta$$

ή

$$\mathbb{P}_\theta\left(\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{c_2} \leq \sigma^2 \leq \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{c_1}\right) = 1-\alpha,\ \forall\, \theta\in\Theta.$$

Επομένως το διάστημα $[\sum_{i=1}^n (X_i - \bar{X})^2/c_2,\ \sum_{i=1}^n (X_i - \bar{X})^2/c_1]$ είναι Δ.Ε. για το σ^2 με σ.ε. $100(1-\alpha)\%$. Ακολουθώντας την ίδια διαδικασία όπως στην προηγούμενη περίπτωση (με μ γνωστό) βρίσκουμε ότι το *Δ.Ε. ίσων ουρών* για το σ^2 είναι

$$\left[\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\chi^2_{n-1,\alpha/2}},\ \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\chi^2_{n-1,1-\alpha/2}}\right],$$

ενώ το Δ.Ε. ελαχίστου μήκους για το σ^2 αντιστοιχεί σε $c_1 < c_2$ που ικανοποιούν τις συνθήκες

$$c_1^{\frac{n-1}{2}+1} e^{-c_1/2} = c_2^{\frac{n-1}{2}+1} e^{-c_2/2}$$

και

$$\int_{c_1}^{c_2} \frac{1}{\Gamma(\frac{n-1}{2})2^{\frac{n-1}{2}}} x^{\frac{n-1}{2}-1} e^{-x/2}\,\mathrm{d}x = 1-\alpha.$$

Το παραπάνω σύστημα ως προς c_1, c_2 μπορεί να λυθεί μόνον με αριθμητικές μεθόδους. Σημειώνεται ξανά ότι και τα δύο Δ.Ε. έχουν την επιθυμητή ιδιότητα να χρησιμοποιούν την (ελάχιστη) επαρκή στατιστική συνάρτηση $(\bar{X}, S^2)$.

Παρατήρηση 9.4.2. Όπως στην περίπτωση με μ γνωστό, το διάστημα

$$\left[\sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\chi^2_{n-1,\alpha/2}}},\ \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\chi^2_{n-1,1-\alpha/2}}}\right]$$

είναι Δ.Ε. ίσων ουρών για το θ με σ.ε. $100(1-\alpha)\%$.

Παράδειγμα 9.4.3. Ένα τυχαίο δείγμα 13 παρατηρήσεων από την κανονική κατανομή $\mathcal{N}(\mu, \sigma^2)$ έδωσε $\sum_{i=1}^{13} (X_i - \bar{X})^2 = 128.41$. Θα υπολογίσουμε Δ.Ε. ίσων ουρών για τα σ^2 και σ με σ.ε. 90%.
Έχουμε $1-\alpha = 0.90$ οπότε $\alpha = 0.1$. Από πίνακες της κατανομής χ^2 βρίσκουμε $\chi^2_{n-1,\alpha/2} = \chi^2_{12,0.05} = 21.030$ και $\chi^2_{n-1,1-\alpha/2} = \chi^2_{12,0.95} = 5.226$ (σε προσέγγιση χιλιοστού). Επομένως το Δ.Ε. ίσων ουρών για το σ^2 είναι

$$\left[\frac{128.41}{21.03},\ \frac{128.41}{5.226}\right] = [6.11,\ 24,57]$$

και για το σ είναι

$$\left[\sqrt{6.11},\ \sqrt{24.57}\right] = [2.47,\ 4.96].$$

9.4.2 Δύο κανονικοί πληθυσμοί

Διάστημα εμπιστοσύνης για τη διαφορά των μέσων τιμών

1η Περίπτωση: σ_1^2, σ_2^2 γνωστά. Έστω $\underset{\sim}{X} = (X_1, \ldots, X_{n_1})$ ένα τυχαίο δείγμα από την κανονική κατανομή $\mathcal{N}(\mu_1, \sigma_1^2)$ και $\underset{\sim}{Y} = (Y_1, \ldots, Y_{n_2})$ ένα τυχαίο δείγμα από την κανονική κατανομή $\mathcal{N}(\mu_2, \sigma_2^2)$. Υποθέτουμε ότι τα δύο δείγματα είναι ανεξάρτητα και θέτουμε $\bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i$, $\bar{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i$. Τα μ_1, μ_2 είναι άγνωστα και τα σ_1^2, σ_2^2 γνωστά. Από τις Προτάσεις 1.8.6(1) και 1.8.5(4) προκύπτει ότι η στατιστική συνάρτηση (και εκτιμητής του $\mu_1 - \mu_2$) $\bar{X} - \bar{Y}$ έχει κατανομή $\mathcal{N}(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$. Τότε η τυχαία μεταβλητή

$$T = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \tag{9.38}$$

έχει κατανομή $\mathcal{N}(0, 1)$, δηλαδή

$$T = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathcal{N}(0, 1)$$

και επομένως είναι ποσότητα οδηγός αφού η κατανομή της προφανώς δεν εξαρτάται από την άγνωστη παράμετρο $\theta = (\mu_1, \mu_2) \in \Theta = \mathbb{R} \times \mathbb{R}$. Άρα υπάρχουν σταθερές $c_1 < c_2$ που εξαρτώνται από την κατανομή $\mathcal{N}(0, 1)$ και το α, $0 < \alpha < 1$, έτσι ώστε

$$\mathbb{P}_\theta\big(c_1 \le T \le c_2\big) = 1 - \alpha,\ \forall\, \theta \in \Theta \tag{9.39}$$

ή

$$\mathbb{P}_\theta\left(c_1 \le \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \le c_2\right) = 1 - \alpha,\ \forall\, \theta \in \Theta$$

ή

$$\mathbb{P}_\theta\Big(\bar{X}-\bar{Y}-c_2\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\le \mu_1-\mu_2\le \bar{X}-\bar{Y}-c_1\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\Big)$$
$$=1-\alpha,\ \forall\,\theta\in\Theta. \quad (9.40)$$

Επομένως το διάστημα $\big[\bar{X}-\bar{Y}-c_2\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}},\ \bar{X}-\bar{Y}-c_1\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\big]$ είναι Δ.Ε. για το $\mu_1-\mu_2$ με σ.ε. $100(1-\alpha)\%$. Όπως και σε προηγούμενες περιπτώσεις είναι προφανές ότι υπάρχουν άπειρα ζεύγη (c_1,c_2) που ικανοποιούν την (9.39) και άρα την (9.40). Κάθε τέτοιο ζεύγος ικανοποιεί την σχέση

$$\Phi(c_1)+1-\Phi(c_2)=\alpha,$$

όπου Φ είναι η συνάρτηση κατανομής της $\mathcal{N}(0,1)$. Ειδικά, για $\Phi(c_1)=\alpha/2=1-\Phi(c_2)$ προκύπτει *διάστημα ίσων ουρών* με συμμετρικά γύρω από το μηδέν c_1,c_2, όπου $c_1=-z_{\alpha/2}$, $c_2=z_{\alpha/2}$. Το διάστημα αυτό είναι

$$\left[\bar{X}-\bar{Y}-z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}},\ \bar{X}-\bar{Y}+z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right].$$

Λόγω της συμμετρίας της $\mathcal{N}(0,1)$ μπορεί ακόμη να δειχθεί ότι το διάστημα αυτό είναι και διάστημα εμπιστοσύνης ελαχίστου μήκους για το $\mu_1-\mu_2$. Επιπλέον, βασίζεται στην (ελάχιστη) επαρκή στατιστική συνάρτηση $(\bar{X},\bar{Y})$.

Παράδειγμα 9.4.4. Ένα τυχαίο δείγμα 15 παρατηρήσεων από την κατανομή $\mathcal{N}(\mu_1,60)$ έδωσε $\bar{X}=70.1$ και ένα άλλο ανεξάρτητο από αυτό με 8 παρατηρήσεις από την κατανομή $\mathcal{N}(\mu_2,40)$ έδωσε $\bar{Y}=75.3$. Θα βρούμε ένα Δ.Ε. με σ.ε. 90% για τη διαφορά $\mu_1-\mu_2$.

Έχουμε $1-\alpha=0.90$, οπότε $\alpha=0.10$. Από πίνακες της κατανομής $\mathcal{N}(0,1)$ προκύπτει $z_{\alpha/2}=z_{0.05}=1.645$. Επομένως το Δ.Ε. για το $\mu_1-\mu_2$ είναι

$$\left[70.1-75.3-1.645\times\sqrt{\frac{60}{15}+\frac{40}{8}},\ 70.1-75.3+1.645\times\sqrt{\frac{60}{15}+\frac{40}{8}}\right]$$
$$=[-10.135,\ -0.265].$$

2η Περίπτωση: $\sigma_1^2 = \sigma_2^2 = \sigma^2$ με σ^2 άγνωστο. Έστω $\underset{\sim}{X} = (X_1, \ldots, X_{n_1})$, $n_1 \geq 2$, ένα τυχαίο δείγμα από την κατανομή $\mathcal{N}(\mu_1, \sigma^2)$ και $\underset{\sim}{Y} = (Y_1, \ldots, Y_{n_2})$, $n_2 \geq 2$, ένα τυχαίο δείγμα από την κατανομή $\mathcal{N}(\mu_2, \sigma^2)$. Υποθέτουμε ότι τα δύο δείγματα είναι ανεξάρτητα, τα μ_1, μ_2, σ^2 είναι άγνωστα και θέτουμε $\theta = (\mu_1, \mu_2, \sigma^2) \in \mathbb{R}^2 \times (0, \infty)$. Έστω ακόμη $\bar{X} = \frac{1}{n_1} \sum\limits_{i=1}^{n_1} X_i$, $\bar{Y} = \frac{1}{n_2} \sum\limits_{i=1}^{n_2} Y_i$, $S_1^2 = \frac{1}{n_1 - 1} \sum\limits_{i=1}^{n_1} (X_i - \bar{X})^2$, $S_2^2 = \frac{1}{n_2 - 1} \sum\limits_{i=1}^{n_2} (Y_i - \bar{Y})^2$. Εκτιμούμε, για λόγους που θα γίνουν κατανοητοί στη συνέχεια, την κοινή διασπορά σ^2. Από την Πρόταση 1.8.6 έχουμε ότι

$$\frac{(n_1 - 1)S_1^2}{\sigma^2} = \frac{\sum_{i=1}^{n_1} (X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{n_1 - 1}$$

και

$$\frac{(n_2 - 1)S_2^2}{\sigma^2} = \frac{\sum_{i=1}^{n_2} (Y_i - \bar{Y})^2}{\sigma^2} \sim \chi^2_{n_2 - 1}.$$

Επομένως

$$\mathbb{E}_\theta \left[\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{\sigma^2} \right] = n_1 + n_2 - 2, \forall\, \theta \in \Theta$$

και συνεπώς

$$\mathbb{E}_\theta \left[\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \right] = \sigma^2, \forall\, \theta \in \Theta$$

δηλαδή η στατιστική συνάρτηση $S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$ είναι αμερόληπτος εκτιμητής του σ^2. Επιπλέον, λόγω ανεξαρτησίας των δύο τυχαίων δειγμάτων, από την Πρόταση 1.8.5(6) προκύπτει ότι

$$\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{\sigma^2} \sim \chi^2_{n_1 + n_2 - 2}.$$

Έστω τώρα η τυχαία μεταβλητή

$$T = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{9.41}$$

που μπορεί να θεωρηθεί ότι προκύπτει από την (9.38) εάν θέσουμε $\sigma_1^2 = \sigma_2^2 = \sigma^2$ και εν συνεχεία αντικαταστήσουμε το σ^2 (που είναι άγνωστο) με τον εκτιμητή του S_p^2. Γράφοντας την τυχαία μεταβλητή T στη μορφή

$$T = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{S_p^2}{\sigma^2}}\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}}$$

παρατηρούμε ότι

$$Z = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} \sim \mathcal{N}(0,1),$$
$$Y = \frac{S_p^2}{\sigma^2} \sim \frac{\chi^2_{n_1+n_2-2}}{n_1+n_2-2}$$

και οι τυχαίες μεταβλητές Z, Y είναι ανεξάρτητες (βλέπε Πρόταση 1.8.6(3)). Άρα η τυχαία μεταβλητή T στην (9.41) έχει κατανομή $t_{n_1+n_2-2}$, δηλαδή

$$T = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

και επομένως είναι ποσότητα οδηγός αφού η κατανομή της προφανώς δεν εξαρτάται από την άγνωστη παράμετρο $\theta = (\mu_1, \mu_2, \sigma^2)$. Άρα υπάρχουν σταθερές $c_1 < c_2$ που εξαρτώνται από την κατανομή $t_{n_1+n_2-2}$ και το α, $0 < \alpha < 1$, έτσι ώστε

$$\mathbb{P}_\theta\big(c_1 \leq T \leq c_2\big) = 1 - \alpha, \ \forall\, \theta \in \Theta \tag{9.42}$$

ή

$$\mathbb{P}_\theta\left(c_1 \leq \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \leq c_2\right) = 1 - \alpha, \ \forall\, \theta \in \Theta$$

ή

$$\mathbb{P}_\theta\Big(\bar{X} - \bar{Y} - c_2 S_p\sqrt{\tfrac{1}{n_1} + \tfrac{1}{n_2}} \leq \mu_1 - \mu_2 \leq \bar{X} - \bar{Y} - c_1 S_p\sqrt{\tfrac{1}{n_1} + \tfrac{1}{n_2}}\Big) = 1 - \alpha, \ \forall\, \theta \in \Theta. \tag{9.43}$$

Όπως και σε προηγούμενες περιπτώσεις είναι προφανές ότι υπάρχουν άπειρα ζεύγη (c_1, c_2) που ικανοποιούν την (9.42), άρα και την (9.43). Κάθε τέτοιο ζεύγος ικανοποιεί τη σχέση

$$F(c_1) + 1 - F(c_2) = \alpha, \quad \forall\, \theta \in \Theta$$

όπου F είναι η συνάρτηση κατανομής της $t_{n_1+n_2-2}$. Ειδικά, για $F(c_1) = 1 - F(c_2) = \alpha/2$ προκύπτει διάστημα ίσων ουρών με συμμετρικά γύρω από το μηδέν c_1, c_2, όπου $c_1 = -t_{n_1+n_2-2,\alpha/2}$ και $c_2 = t_{n_1+n_2-2,\alpha/2}$. Το διάστημα αυτό είναι

$$\left[\bar{X} - \bar{Y} - t_{n_1+n_2-2,\,\alpha/2} S_p \sqrt{\tfrac{1}{n_1} + \tfrac{1}{n_2}},\ \bar{X} - \bar{Y} + t_{n_1+n_2-2,\,\alpha/2} S_p \sqrt{\tfrac{1}{n_1} + \tfrac{1}{n_2}}\right].$$

Λόγω συμμετρίας της $t_{n_1+n_2-2,\alpha/2}$ μπορεί ακόμη να δειχθεί ότι το διάστημα αυτό είναι και διάστημα εμπιστοσύνης ελαχίστου μήκους για το $\mu_1 - \mu_2$. Επιπλέον, βασίζεται στην (ελάχιστη) επαρκή στατιστική συνάρτηση $(\bar{X}, \bar{Y}, S_p^2)$.

3η Περίπτωση: σ_1^2, σ_2^2 άγνωστα (και διάφορα μεταξύ των). Στην περίπτωση αυτή δεν είναι δυνατόν να βρεθεί κατάλληλη ποσότητα οδηγός, εκτός αν ο λόγος σ_1^2/σ_2^2 είναι γνωστός (Άσκηση 9.13). Ένα ασυμπτωτικό διάστημα εμπιστοσύνης για το $\mu_1 - \mu_2$ δίνεται στην επόμενη ενότητα.

Διάστημα εμπιστοσύνης για το λόγο των διασπορών

Υπενθυμίζουμε αρχικά τον ορισμό της κατανομής $\mathcal{F}$ *με* n *και* m *βαθμούς ελευθερίας*, που δόθηκε στην Ενότητα 1.6.2 και την Πρόταση 1.8.5, η οποία θα χρησιμοποιηθεί στην κατασκευή του Δ.Ε..

Ορισμός 9.4.2. *Έστω X μία τυχαία μεταβλητή με κατανομή χ_n^2 και Y μία τυχαία μεταβλητή με κατανομή χ_m^2 . Εάν οι τυχαίες μεταβλητές X και Y είναι ανεξάρτητες, τότε η κατανομή της τυχαίας μεταβλητής*

$$T = \frac{X/n}{Y/m}$$

ονομάζεται κατανομή $\mathcal{F}$ με n και m βαθμούς ελευθερίας. Συμβολικά γράφουμε $T \sim \mathcal{F}_{n,m}$, ή ακόμη $\mathcal{F}_{n,m} = \dfrac{\chi^2_n/n}{\chi^2_m/m}$.

Αναφέρουμε ότι η ονομασία και ο συμβολισμός $\mathcal{F}$ προτάθηκε από τον (Αμερικανό) Στατιστικό G. Snedecor προς τιμήν (του Άγγλου) στατιστικού Fisher ($\mathcal{F}$ από το Fisher). Μπορεί να δειχθεί ότι η πυκνότητα της κατανομής $\mathcal{F}_{n,m}$ είναι

$$f(t) = \frac{\Gamma(\frac{n+m}{2})(\frac{n}{m})^{n/2}\, t^{n/2-1}}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})(1+\frac{n}{m}\,t)^{\frac{n+m}{2}}},\ t > 0.$$

Μερικές βασικές ιδιότητες της κατανομής $\mathcal{F}_{n,m}$ είναι οι εξής.

1. Εάν $T \sim \mathcal{F}_{n,m}$ τότε $\dfrac{1}{T} \sim \mathcal{F}_{m,n}$. Συμβολικά $\dfrac{1}{\mathcal{F}_{m,n}} \sim \mathcal{F}_{n,m}$.
2. Εάν $T \sim t_m$ τότε $T^2 \sim \mathcal{F}_{1,m}$ (σχέση των κατανομών t, $\mathcal{F}$).
3. Εάν $T \sim \mathcal{F}_{n,m}$ τότε $\mathbb{E}\,T = \dfrac{m}{m-2}$, $m \geq 3$.
4. Εάν $T \sim \mathcal{F}_{n,m}$ τότε $\mathbb{V}arT = \dfrac{2m^2(n+m-2)}{n(m-2)^2(m-4)}$, $m \geq 5$.

1η Περίπτωση: $\boldsymbol{\mu_1}$, $\boldsymbol{\mu_2}$ άγνωστα. Έστω $\underset{\sim}{X} = (X_1, \ldots, X_{n_1})$, $n_1 \geq 2$, ένα τυχαίο δείγμα από την κανονική κατανομή $\mathcal{N}(\mu_1, \sigma_1^2)$ και $\underset{\sim}{Y} = (Y_1, \ldots, Y_{n_2})$, $n_2 \geq 2$, ένα τυχαίο δείγμα από την κανονική κατανομή $\mathcal{N}(\mu_2, \sigma_2^2)$. Υποθέτουμε ότι τα δύο δείγματα είναι ανεξάρτητα, τα μ_1, μ_2, σ_1^2, σ_2^2 είναι άγνωστα και θέτουμε $\theta = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) \in \mathbb{R}^2 \times (0,\infty)^2$. Έστω ακόμη $\bar{X} = \frac{1}{n_1}\sum\limits_{i=1}^{n_1} X_i$, $\bar{Y} = \frac{1}{n_2}\sum\limits_{i=1}^{n_2} Y_i$, $S_1^2 = \frac{1}{n_1-1}\sum\limits_{i=1}^{n_1}(X_i - \bar{X})^2$, $S_2^2 = \frac{1}{n_2-1}\sum\limits_{i=1}^{n_2}(Y_i - \bar{Y})^2$. Θεωρούμε την στατιστική συνάρτηση (και εκτιμητή του σ_1^2/σ_2^2) S_1^2/S_2^2 και παρατηρούμε ότι η τυχαία μεταβλητή

$$T = \frac{\sigma_2^2}{\sigma_1^2}\,\frac{S_1^2}{S_2^2}$$

έχει κατανομή $\mathcal{F}_{n_1-1,\,n_2-1}$ αφού λόγω της Πρότασης 1.8.6(3)

$$\frac{\sum_{i=1}^{n_1}(X_i - \bar{X})^2}{\sigma_1^2} \sim \chi^2_{n_1-1}, \qquad \frac{\sum_{i=1}^{n_2}(Y_i - \bar{Y})^2}{\sigma_2^2} \sim \chi^2_{n_2-1}$$

και οι δύο αυτές στατιστικές συναρτήσεις είναι ανεξάρτητες. Επομένως,

$$T = \frac{\sigma_2^2}{\sigma_1^2}\frac{S_1^2}{S_2^2} \sim \mathcal{F}_{n_1-1,n_2-1},$$

δηλαδή η τυχαία μεταβλητή T είναι ποσότητα οδηγός αφού προφανώς η κατανομή της δεν εξαρτάται από την άγνωστη παράμετρο $\theta = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$. Άρα, υπάρχουν σταθερές $c_1 < c_2$ που εξαρτώνται από την κατανομή $\mathcal{F}_{n_1-1,n_2-1}$ και το α, $0 < \alpha < 1$, έτσι ώστε

$$\mathbb{P}_\theta\big(c_1 \le T \le c_2\big) = 1 - \alpha,\ \forall\, \theta \in \Theta \tag{9.44}$$

ή

$$\mathbb{P}_\theta\Big(c_1 \le \frac{\sigma_2^2}{\sigma_1^2}\frac{S_1^2}{S_2^2} \le c_2\Big) = 1 - \alpha,\ \forall\, \theta \in \Theta$$

ή

$$\mathbb{P}_\theta\Big(\frac{1}{c_2}\frac{S_1^2}{S_2^2} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{1}{c_1}\frac{S_1^2}{S_2^2}\Big) = 1 - \alpha,\ \forall\, \theta \in \Theta. \tag{9.45}$$

Όπως σε προηγούμενες περιπτώσεις, υπάρχουν άπειρα ζεύγη (c_1, c_2) που ικανοποιούν την (9.44), άρα και την (9.45). Κάθε τέτοιο ζεύγος ικανοποιεί τη σχέση

$$F(c_1) + 1 - F(c_2) = \alpha, \quad \forall\, \theta \in \Theta$$

όπου F είναι η συνάρτηση κατανομής της $\mathcal{F}_{n_1-1,n_2-1}$ και καθορίζει ένα Δ.Ε. για το σ_1^2/σ_2^2 με σ.ε. $100(1-\alpha)\%$. Ειδικά για $F(c_1) = \alpha/2 = 1 - F(c_2)$ προκύπτει διάστημα εμπιστοσύνης ίσων ουρών. Θέτοντας (για λόγους συμβολισμού) $\mathcal{F}_{n_1-1,n_2-1,\,\alpha/2} = F^{-1}(1-\alpha/2)$, δηλαδή ορίζοντας το σημείο $\mathcal{F}_{n_1-1,n_2-1,\,\alpha/2}$ από τη σχέση

$$\mathbb{P}\big(\mathcal{F}_{n_1-1,n_2-1} > \mathcal{F}_{n_1-1,n_2-1,\,\alpha/2}\big) = \alpha/2,$$

έχουμε $c_1 = \mathcal{F}_{n_1-1,n_2-1,\,1-\alpha/2}$, $c_2 = \mathcal{F}_{n_1-1,n_2-1,\,\alpha/2}$, οπότε το Δ.Ε. ίσων ουρών για το σ_1^2/σ_2^2 είναι

$$\left[\frac{1}{\mathcal{F}_{n_1-1,n_2-1,\,\alpha/2}}\frac{S_1^2}{S_2^2},\ \frac{1}{\mathcal{F}_{n_1-1,n_2-1,\,1-\alpha/2}}\frac{S_1^2}{S_2^2}\right].$$

Μπορούμε ακόμη να αναζητήσουμε τα c_1, c_2 που αντιστοιχούν σε Δ.Ε. ελαχίστου μήκους. Σημειώνουμε ότι αυτά τα c_1, c_2 μπορούν να υπολογισθούν μόνον με αριθμητικές μεθόδους. Αναφέρουμε ακόμη ότι και τα δύο διαστήματα έχουν την επιθυμητή ιδιότητα να βασίζονται στην (ελάχιστη) επαρκή στατιστική συνάρτηση $(\bar{X}, \bar{Y}, S_1^2, S_2^2)$.

2η Περίπτωση: $\boldsymbol{\mu_1}$, $\boldsymbol{\mu_2}$ γνωστά. Στην περίπτωση αυτή ένα Δ.Ε. για το σ_1^2/σ_2^2 μπορεί να κατασκευασθεί χρησιμοποιώντας ως ποσότητα οδηγό τη τυχαία μεταβλητή

$$T = \frac{\sigma_2^2}{\sigma_1^2}\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2},$$

όπου $\hat{\sigma}_1^2 = \frac{1}{n_1}\sum_{i=1}^{n_1}(X_i - \mu_1)^2$, $\hat{\sigma}_2^2 = \frac{1}{n_2}\sum_{i=1}^{n_2}(X_i - \mu_2)^2$ είναι εκτιμητές των σ_1^2 και σ_2^2 αντίστοιχα. Η τυχαία μεταβλητή T έχει κατανομή $\mathcal{F}_{n_1,n_2}$ και με παρόμοιο τρόπο όπως παραπάνω το Δ.Ε. ίσων ουρών για το σ_1^2/σ_2^2 είναι

$$\left[\frac{1}{\mathcal{F}_{n_1,n_2,\,\alpha/2}}\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}, \frac{1}{\mathcal{F}_{n_1,n_2,\,1-\alpha/2}}\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}\right].$$

3η Περίπτωση: $\boldsymbol{\mu_1}$ γνωστό, $\boldsymbol{\mu_2}$ άγνωστο. (βλέπε Άσκηση 9.11.)

4η Περίπτωση: $\boldsymbol{\mu_1}$ άγνωστο, $\boldsymbol{\mu_2}$ γνωστό. (βλέπε Άσκηση 9.12.)

9.5 Ασυμπτωτικά διαστήματα εμπιστοσύνης

9.5.1 Ασυμπτωτικό διάστημα εμπιστοσύνης για τη μέση τιμή

Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από μία κατανομή με άγνωστη μέση τιμή μ και άγνωστη διασπορά σ^2. Θα κατασκευάσουμε ένα ασυμπτωτικό διάστημα εμπιστοσύνης (Α.Δ.Ε.) για το μ. Το διάστημα αυτό χρησιμεύει ειδικά στην περίπτωση που η κατανομή των δεδομένων είναι άγνωστη. Θέτουμε $\bar{X} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$ και $S^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \bar{X})^2$. Τότε από το Κεντρικό Οριακό Θεώρημα, Θεώρημα 1.10.3, έχουμε ότι η κατανομή της τυχαίας μεταβλητής $\sqrt{n}(\bar{X} - \mu)/\sigma$ είναι κατά προσέγγιση (για n «μεγάλο») κανονική $\mathcal{N}(0, 1)$. Επιπλέον, η δειγματική διασπορά S^2 είναι (ασθενώς) συνεπής εκτιμητής του σ^2, δηλαδή $S^2 \xrightarrow{\mathbb{P}} \sigma^2$. Άρα από το Θεώρημα Slutsky, Θεώρημα 1.10.6(ii), προκύπτει ότι $\sqrt{n}(\bar{X}-\mu)/S \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$. Αυτό σημαίνει ότι

$$\lim_{n\to\infty} \mathbb{P}\big(\frac{\sqrt{n}(\bar{X}-\mu)}{S} \leq x\big) \longrightarrow \mathbb{P}(Z \leq x),\ \forall\, x \in \mathbb{R},$$

όπου Z είναι τυχαία μεταβλητή με κανονική κατανομή $\mathcal{N}(0,1)$.

Για την κατασκευή του Α.Δ.Ε. για το μ με σ.ε. $100(1-\alpha)\%$, εργαζόμαστε ως εξής. Από τον ορισμό του σημείου $z_{\alpha/2}$ έχουμε

$$\mathbb{P}(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = \mathbb{P}(Z \leq z_{\alpha/2}) - \mathbb{P}(Z \leq -z_{\alpha/2}) = 1-\alpha,$$

οπότε

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}\Big(-z_{\alpha/2} \leq \frac{\sqrt{n}(\bar{X}-\mu)}{S} \leq z_{\alpha/2}\Big) \\
&\quad = \lim_{n\to\infty} \Big\{\mathbb{P}\Big(\frac{\sqrt{n}(\bar{X}-\mu)}{S} \leq z_{\alpha/2}\Big) - \mathbb{P}\Big(\frac{\sqrt{n}(\bar{X}-\mu)}{S} \leq -z_{\alpha/2}\Big)\Big\} \\
&\qquad\qquad = \mathbb{P}(Z \leq z_{\alpha/2}) - \mathbb{P}(Z \leq -z_{\alpha/2}) = 1-\alpha.
\end{aligned}$$

Επομένως, λύνοντας ως προς μ,

$$\lim_{n\to\infty} \mathbb{P}\Big(\bar{X} - z_{\alpha/2}\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2}\frac{S}{\sqrt{n}}\Big) = 1-\alpha$$

δηλαδή το διάστημα

$$\Big[\bar{X} - z_{\alpha/2}\frac{S}{\sqrt{n}},\ \bar{X} + z_{\alpha/2}\frac{S}{\sqrt{n}}\Big]$$

είναι ένα Α.Δ.Ε. για το μ με συντελεστή εμπιστοσύνης $100(1-\alpha)\%$.

9.5.2 Ασυμπτωτικό διάστημα εμπιστοσύνης για τη διαφορά δύο μέσων τιμών

Έστω $\underset{\sim}{X} = (X_1,\ldots,X_{n_1})$ ένα τυχαίο δείγμα από μία κατανομή με άγνωστη μέση τιμή μ_1 και άγνωστη διασπορά σ_1^2. Έστω ακόμη $\underset{\sim}{Y} = (Y_1,\ldots,Y_{n_2})$ ένα τυχαίο δείγμα από μία άλλη κατανομή με άγνωστη μέση τιμή μ_2 και άγνωστη διασπορά σ_2^2. Τα δύο δείγματα είναι ανεξάρτητα. Θα κατασκευάσουμε ένα Α.Δ.Ε. για τη διαφορά $\mu_1 - \mu_2$. Το διάστημα αυτό χρησιμεύει ειδικά στην περίπτωση που οι κατανομές των δεδομένων είναι άγνωστες. Θέτουμε $\bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i$, $\bar{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i$, $S_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_i - \bar{X})^2$, $S_2^2 = \frac{1}{n_2-1}\sum_{i=1}^{n_2}(Y_i - \bar{Y})^2$.

Από το Κεντρικό Οριακό Θεώρημα και λόγω της ανεξαρησίας των δύο δειγμάτων έχουμε ότι καθώς $n_1, n_2 \to \infty$

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

Επιπλέον, επειδή οι δειγματικές διασπορές S_1^2, S_2^2 είναι (ασθενώς) συνεπείς εκτιμητές των σ_1^2, σ_2^2, από το Θεώρημα Slutsky, Θεώρημα 1.10.6(ii) προκύπτει ότι

$$T = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

Αυτό σημαίνει ότι

$$\lim_{\substack{n_1\to\infty \\ n_2\to\infty}} \mathbb{P}(T \le x) = \mathbb{P}(Z \le x), \ \forall\, x \in \mathbb{R},$$

όπου Z είναι τυχαία μεταβλητή με κατανομή $\mathcal{N}(0,1)$. Ακολουθώντας την ίδια μεθοδολογία όπως στην προηγούμενη παράγραφο, καταλήγουμε ότι το διάστημα

$$\left[\bar{X} - \bar{Y} - z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}},\ \bar{X} - \bar{Y} + z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right].$$

είναι Α.Δ.Ε. για το $\mu_1 - \mu_2$ με συντελεστή εμπιστοσύνης $100(1-\alpha)\%$.

9.5.3 Ασυμπτωτικό διάστημα εμπιστοσύνης για ποσοστό

Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την κατανομή Bernoulli $\mathcal{B}(1,p)$. Θα κατασκευάσουμε ένα Α.Δ.Ε. για το p. Θέτουμε $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

(που είναι εκτιμητής του p). Από το Κεντρικό Οριακό Θεώρημα, Θεώρημα 1.10.3, προκύπτει ότι

$$\frac{\bar{X}-p}{\sqrt{\frac{p(1-p)}{n}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

Αυτό σημαίνει ότι

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\bar{X}-p}{\sqrt{\frac{p(1-p)}{n}}} \le x\right) = \mathbb{P}(Z \le x),\ \forall\, x \in \mathbb{R},$$

όπου Z είναι τυχαία μεταβλητή με κατανομή $\mathcal{N}(0,1)$.

Για την κατασκευή του Α.Δ.Ε. για το p με συντελεστή εμπιστοσύνης $100(1-\alpha)\%$ εργαζόμαστε ως εξής: από τον ορισμό του σημείου $z_{\alpha/2}$ έχουμε

$$\mathbb{P}(-z_{\alpha/2} \le Z \le z_{\alpha/2}) = \mathbb{P}(Z \le z_{\alpha/2}) - \mathbb{P}(Z \le -z_{\alpha/2}) = 1-\alpha.$$

Επομένως,

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}\left(-z_{\alpha/2} \le \frac{\bar{X}-p}{\sqrt{\frac{p(1-p)}{n}}} \le z_{\alpha/2}\right) \\
&\quad = \lim_{n\to\infty}\left\{\mathbb{P}\left(\frac{\bar{X}-p}{\sqrt{\frac{p(1-p)}{n}}} \le z_{\alpha/2}\right) - \mathbb{P}\left(\frac{\bar{X}-p}{\sqrt{\frac{p(1-p)}{n}}} \le -z_{\alpha/2}\right)\right\} \\
&\quad = \mathbb{P}(Z \le z_{\alpha/2}) - \mathbb{P}(Z \le -z_{\alpha/2}) = 1-\alpha.
\end{aligned}$$

Άρα,

$$\lim_{n\to\infty} \mathbb{P}\left(-z_{\alpha/2} \le \frac{\bar{X}-p}{\sqrt{\frac{p(1-p)}{n}}} \le z_{\alpha/2}\right) = 1-\alpha. \tag{9.46}$$

Επιλύουμε τώρα τη διπλή ανισότητα ως προς p. Έχουμε

$$\begin{aligned}
&-z_{\alpha/2} \le \frac{\bar{X}-p}{\sqrt{\frac{p(1-p)}{n}}} \le z_{\alpha/2} \Longleftrightarrow \frac{n(\bar{X}-p)^2}{p(1-p)} \le z_{\alpha/2}^2 \\
&\quad \Longleftrightarrow (n+z_{\alpha/2}^2)p^2 - (z_{\alpha/2}^2 + 2n\bar{X})p + n\bar{X}^2 \le 0 \Longleftrightarrow p_1 \le p \le p_2,
\end{aligned}$$

όπου p_1, p_2 είναι οι ρίζες του τριωνύμου ως προς p. Άρα,

$$p_{1,2} = \frac{n\bar{X}}{n + z_{\alpha/2}^2} + \frac{z_{\alpha/2}^2 \pm z_{\alpha/2}\sqrt{z_{\alpha/2}^2 + 4n\bar{X}(1-\bar{X})}}{2(n + z_{\alpha/2}^2)},$$

οπότε η (9.46) γίνεται $\lim_{n\to\infty} \mathbb{P}(p_1 \le p \le p_2) = 1 - \alpha$. Άρα, για n «μεγάλο» έχουμε

$$\mathbb{P}(p_1 \le p \le p_2) \approx 1 - \alpha$$

και συνεπώς το διάστημα $[p_1,\, p_2]$ είναι Α.Δ.Ε. για το p με συντελεστή εμπιστοσύνης $100(1-\alpha)\%$. Για «μεγάλες» τιμές του n οι ρίζες p_1, p_2 μπορούν να προσεγγισθούν από τις τιμές $\bar{X} \pm z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}$, οπότε το Α.Δ.Ε. για το p γίνεται

$$\left[\bar{X} - z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}},\, \bar{X} + z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right] \tag{9.47}$$

Ένας άλλος τρόπος να καταλήξουμε στο Α.Δ.Ε. για το p είναι ο εξής: η τυχαία μεταβλητή

$$\frac{\bar{X} - p}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}},$$

που προκύπτει από την

$$\frac{\bar{X} - p}{\sqrt{\frac{p(1-p)}{n}}},$$

αντικαθιστώντας το p με τον συνεπή εκτιμητή του $\bar{X}$, έχει κατά προσέγγιση κανονική κατανομή $\mathcal{N}(0,1)$. Άρα, για n «μεγάλο» έχουμε

$$\mathbb{P}\left(-z_{\alpha/2} \le \frac{\bar{X} - p}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \le z_{\alpha/2}\right) \approx \mathbb{P}(-z_{\alpha/2} \le Z \le z_{\alpha/2}) = 1 - \alpha.$$

Επομένως,

$$\mathbb{P}\left(\bar{X} - z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \le p \le \bar{X} + z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right) \approx 1 - \alpha,$$

δηλαδή το Α.Δ.Ε. στην (9.47).

Παράδειγμα 9.5.1. 32 άτομα από ένα τυχαίο δείγμα 112 ψηφοφόρων απάντησαν θετικά στο ερώτημα εάν θα ψηφίσουν υπέρ κάποιου κόμματος. Θα υπολογίσουμε ένα ασυμπτωτικό Α.Δ.Ε. για το ποσοστό p του κόμματος με σ.ε. 95%.
Έχουμε $\hat{p} = \bar{x} = \frac{32}{112} = 0.286$ και $1 - \alpha = 0.95$, οπότε $\alpha = 0.05$ και $z_{\alpha/2} = z_{0.05} = 1.96$. Άρα ένα Δ.Ε. για το p με σ.ε. κατά προσέγγιση 95% είναι

$$\Big[0.286 - 1.96\sqrt{\frac{(0.286)(1-0.286)}{112}},\, 0.286 + 1.96\sqrt{\frac{(0.286)(1-0.286)}{112}}\,\Big]$$
$$= [0.286 - 0.084, 0.286 + 0.084] = [0.202,\ 0.370].$$

Το διάστημα μπορεί να γραφεί στη συμβολική μορφή 0.286 ± 0.084 που δηλώνει ότι η εκτίμηση του p είναι 0.286 με μέγιστο σφάλμα, κατ' εκτίμηση και υπό το πρίσμα της Παρατήρησης 9.1.1, 0.084. Για τα συγκεκριμένα δεδομένα, αναφέρουμε ότι το εύρος του διαστήματος είναι αρκετά μεγάλο, ενώ θα μπορούσε να βελτιωθεί παίρνοντας μεγαλύτερο δείγμα.

9.5.4 Ασυμπτωτικό διάστημα εμπιστοσύνης για τη διαφορά δύο ποσοστών

Έστω $\underset{\sim}{X} = (X_1, \ldots, X_{n_1})$ ένα τυχαίο δείγμα από την κατανομή Bernoulli $\mathcal{B}(1, p_1)$ και $\underset{\sim}{Y} = (Y_1, \ldots, Y_{n_2})$ ένα τυχαίο δείγμα από την κατανομή Bernoulli $\mathcal{B}(1, p_2)$. Τα δύο δείγματα είναι ανεξάρτητα. Θα κατασκευάσουμε ένα Α.Δ.Ε. για τη διαφορά $p_1 - p_2$. Θέτουμε $\hat{p}_1 = \bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i$, $\hat{p}_2 = \bar{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i$. Τα $\hat{p}_1$, $\hat{p}_2$ είναι εκτιμητές των p_1, p_2 αντίστοιχα. Από το Κεντρικό Οριακό Θεώρημα προκύπτει ότι η κατανομή της τυχαίας μεταβλητής

$$\frac{\bar{X} - p_1}{\sqrt{\frac{p_1(1-p_1)}{n_1}}}$$

είναι κατά προσέγγιση (n_1 «μεγάλο») κανονική κατανομή $\mathcal{N}(0, 1)$. Ισοδύναμα, η κατανομή του εκτιμητή $\hat{p}_1 = \bar{X}$ είναι κατά προσέγγιση

$\mathcal{N}\big(p_1, \frac{p_1(1-p_1)}{n_1}\big)$. Ανάλογα, η κατανομή του εκτιμητή $\hat{p}_2 = \bar{Y}$ είναι κατά προσέγγιση (για n_2 «μεγάλο») κανονική $\mathcal{N}\big(p_2, \frac{p_2(1-p_2)}{n_2}\big)$. Λόγω της ανεξαρτησίας των δύο δειγμάτων, ο εκτιμητής του $p_1 - p_2$, $\hat{p}_1 - \hat{p}_2$, έχει κατά προσέγγιση (για n_1,n_2 «μεγάλα») κανονική κατανομή $\mathcal{N}\big(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\big)$. Επομένως η τυχαία μεταβλητή

$$\frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

έχει κατά προσέγγιση κατανομή $\mathcal{N}(0,1)$. Εάν τώρα αντικαταστήσουμε στον παρονομαστή του κλάσματος τα p_1, p_2 με τους συνεπείς εκτιμητές των $\hat{p}_1$ και $\hat{p}_2$ αντίστοιχα, η κατανομή του νέου κλάσματος θα είναι επίσης κατά προσέγγιση $\mathcal{N}(0,1)$, λόγω του Θεωρήματος Slutsky, Θεώρημα 1.10.6(ii). Δηλαδή,

$$T = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \approx Z \sim \mathcal{N}(0,1).$$

Για την κατασκευή του Α.Δ.Ε. για το $p_1 - p_2$ με συντελεστή εμπιστοσύνης $100(1-\alpha)\%$ εργαζόμαστε ως εξής: από τον ορισμό του σημείου $z_{\alpha/2}$ έχουμε ότι για «μεγάλα» n_1, n_2

$$\mathbb{P}\big(-z_{\alpha/2} \le T \le z_{\alpha/2}\big) \approx \mathbb{P}\big(-z_{\alpha/2} \le Z \le z_{\alpha/2}\big) = 1 - \alpha$$

ή

$$\mathbb{P}\left(-z_{\alpha/2} \le \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \le z_{\alpha/2}\right) \approx 1 - \alpha$$

ή

$$\begin{aligned}\mathbb{P}\Big(\hat{p}_1 - \hat{p}_2 - z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \\ \le p_1 - p_2 \le \\ \hat{p}_1 - \hat{p}_2 + z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\Big) \approx 1 - \alpha.\end{aligned}$$

Άρα, το διάστημα

$$\left[\hat{p}_1 - \hat{p}_2 - z_{\alpha/2}\sqrt{\tfrac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \tfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}},\, \hat{p}_1 - \hat{p}_2 + z_{\alpha/2}\sqrt{\tfrac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \tfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\right]$$

είναι ένα Α.Δ.Ε. για το $p_1 - p_2$ με σ.ε. κατά προσέγγιση $100(1-\alpha)\%$.

Παράδειγμα 9.5.2. Δύο απορρυπαντικά δοκιμάστηκαν για την ικανότητα τους να καθαρίζουν λεκέδες ενός ορισμένου τύπου. Το πρώτο αποδείχθηκε αποτελεσματικό σε 63 από 91 ανεξάρτητες δοκιμές και το δεύτερο σε 42 από 79 ανεξάρτητες δοκιμές. Θα κατασκευάσουμε ένα Δ.Ε. για τη διαφορά $p_1 - p_2$ των ποσοστών αποτελεσματικότητας των δύο απορρυπαντικών με σ.ε. κατά προσέγγιση 90%. Οι εκτιμητές των p_1 και p_2 είναι $\hat{p}_1 = \frac{63}{91} = 0.692$ και $\hat{p}_2 = \frac{42}{79} = 0.532$, αντίστοιχα. Επιπλέον έχουμε $1-\alpha = 0.90$, οπότε $\alpha = 0.10$ και $z_{\alpha/2} = z_{0.05} = 1.645$. Άρα η εκτίμηση της διαφοράς $p_1 - p_2$ είναι $\hat{p}_1 - \hat{p}_2 = 0.692 - 0.532 = 0.160$ με Α.Δ.Ε.

$$\begin{aligned}&\Big[0.692 - 0.532 - 1.645\sqrt{\tfrac{(0.692)(0.308)}{91} + \tfrac{(0.532)(0.468)}{79}},\\ &\qquad 0.692 - 0.532 + 1.645\sqrt{\tfrac{(0.692)(0.308)}{91} + \tfrac{(0.532)(0.468)}{79}}\Big]\\ &\qquad\qquad = [0.038,\, 0.282].\end{aligned}$$

ή σε συμβολική μορφή 0.160 ± 0.038. Επειδή το Δ.Ε. περιέχει μόνον θετικές τιμές, η ένδειξη είναι ότι το πρώτο απορρυπαντικό είναι πιο αποτελεσματικό από το δεύτερο.

9.6 Ασυμπτωτικά διαστήματα εμπιστοσύνης και εκτιμητές μέγιστης πιθανοφάνειας

Όπως είδαμε στην Ενότητα 7.2, στην περίπτωση τυχαίου δείγματος και υπό ορισμένες συνθήκες ο ε.μ.π. του θ, $\hat{\theta}_n$, έχει, για «μεγάλο» n, κατά προσέγγιση κανονική κατανομή

$$\hat{\theta}_n \overset{\mathcal{L}_\theta}{\approx} \mathcal{N}(\theta, \frac{1}{nI_1(\theta)}) \tag{9.48}$$

και είναι, τουλάχιστον, ασθενώς συνεπής εκτιμητής του θ, $\hat{\theta}_n \xrightarrow{\mathbb{P}_\theta} \theta$. Ισοδύναμα, η παραπάνω προσέγγιση γράφεται

$$\sqrt{nI_1(\theta)}\,(\hat{\theta}_n - \theta) \overset{\mathcal{L}_\theta}{\approx} Z, \tag{9.49}$$

όπου Z είναι μια τυχαία μεταβλητή με τυπική κανονική κατανομή $\mathcal{N}(0,1)$. Θεωρώντας ότι η συνάρτηση $I_1(\theta)$ είναι συνεχής, από την Πρόταση 1.10.5 έχουμε $I_1(\hat{\theta}_n) \xrightarrow{\mathbb{P}_\theta} I_1(\theta)$ και επομένως από το Θεώρημα Slutsky (Θεώρημα 1.10.6) και την (9.49) προκύπτει ότι

$$\sqrt{nI_1(\hat{\theta}_n)}\,(\hat{\theta}_n - \theta) \overset{\mathcal{L}_\theta}{\approx} Z, \tag{9.50}$$

Η (9.50) δηλώνει ότι η τυχαία μεταβλητή $\sqrt{nI_1(\hat{\theta}_n)}\,(\hat{\theta}_n - \theta)$ είναι κατά προσέγγιση ποσότητα οδηγός και επομένως μπορεί να χρησιμοποιηθεί για την κατασκευή ασυμπτωτικού διαστήματος εμπιστοσύνης για το θ με σ.ε. $100(1-\alpha)\%$. Πράγματι, από τον ορισμό του ποσοστιαίου σημείου $z_{\alpha/2}$ της $\mathcal{N}(0,1)$ έχουμε

$$\mathbb{P}\big(-z_{\alpha/2} \le Z \le z_{\alpha/2}\big) = 1-\alpha$$

και επομένως, λόγω της (9.50)

$$\mathbb{P}\Big(-z_{\alpha/2} \le \sqrt{nI_1(\hat{\theta}_n)}\,(\hat{\theta}_n - \theta) \le z_{\alpha/2}\Big) \approx \mathbb{P}\big(-z_{\alpha/2} \le Z \le z_{\alpha/2}\big) = 1-\alpha,$$

που συνεπάγεται ότι

$$\mathbb{P}\Big(\hat{\theta}_n - \frac{1}{\sqrt{nI_1(\hat{\theta}_n)}} z_{\alpha/2} \le \theta \le \hat{\theta}_n + \frac{1}{\sqrt{nI_1(\hat{\theta}_n)}} z_{\alpha/2}\Big) \approx 1-\alpha, \quad \forall\, \theta \in \Theta.$$

Συμπεραίνουμε λοιπόν, ότι το διάστημα

$$\Big[\hat{\theta}_n - \frac{1}{\sqrt{nI_1(\hat{\theta}_n)}} z_{\alpha/2}, \hat{\theta}_n + \frac{1}{\sqrt{nI_1(\hat{\theta}_n)}} z_{\alpha/2}\Big] \tag{9.51}$$

ή συμβολικά

$$\hat{\theta}_n \pm \frac{1}{\sqrt{nI_1(\hat{\theta}_n)}} z_{\alpha/2}$$

είναι Α.Δ.Ε. για το θ με σ.ε. $100(1-\alpha)\%$.

Παράδειγμα 9.6.1. Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την κατανομή Poisson, $\mathcal{P}(\theta)$, $\theta \in \Theta = (0, \infty)$. Όπως είδαμε στο Παράδειγμα 7.1.3, με πιθανότητα που τείνει στο 1, ο ε.μ.π. του θ είναι $\hat{\theta}_n = \bar{X}$. Επιπλέον, $I_1(\theta) = \frac{1}{\theta}$ (Παράδειγμα 5.2.10). Άρα ένα Α.Δ.Ε. για το θ είναι

$$\Big[\bar{X} - \sqrt{\frac{\bar{X}}{n}} z_{\alpha/2}, \bar{X} + \sqrt{\frac{\bar{X}}{n}} z_{\alpha/2}\Big].$$

□

Ανάλογα με τα παραπάνω και χρησιμοποιώντας την Πρόταση 7.2.7, μπορούμε να κατασκευάσουμε ένα Α.Δ.Ε. για την άγνωστη τιμή $g(\theta)$. Υπό τις προϋποθέσεις της Πρότασης 7.2.7 έχουμε

$$\sqrt{n}\left(g(\hat{\theta}_n) - g(\theta)\right) \overset{\mathcal{L}_\theta}{\approx} \mathcal{N}\left(0, \frac{[g'(\theta)]^2}{I_1(\theta)}\right).$$

Εάν επιπλέον θεωρήσουμε ότι η συνάρτηση $g'(\theta)$ είναι συνεχής, τότε η συνέπεια του ε.μ.π. $\hat{\theta}_n$ και το Θεώρημα Slutsky εγγυώνται ότι

$$\frac{\sqrt{nI_1(\hat{\theta}_n)}}{|g'(\hat{\theta}_n)|}\left(g(\hat{\theta}_n) - g(\theta)\right) \overset{\mathcal{L}_\theta}{\approx} Z.$$

Ακολουθώντας την ίδια διαδικασία, όπως προηγουμένως, κατασκευάζουμε το διάστημα

$$\Big[g(\hat{\theta}_n) - \frac{|g'(\hat{\theta}_n)|}{\sqrt{nI_1(\hat{\theta}_n)}} z_{\alpha/2}, g(\hat{\theta}_n) + \frac{|g'(\hat{\theta}_n)|}{\sqrt{nI_1(\hat{\theta}_n)}} z_{\alpha/2}\Big] \tag{9.52}$$

ή συμβολικά,

$$g(\hat{\theta}_n) \pm \frac{|g'(\hat{\theta}_n)|}{\sqrt{nI_1(\hat{\theta}_n)}} z_{\alpha/2}$$

που είναι Α.Δ.Ε. για το $g(\theta)$ με σ.ε. $100(1-\alpha)\%$. Υπό το πρίσμα της Παρατήρησης 9.1.1, η ερμηνεία του είναι ότι ως εκτίμηση του $g(\theta)$ δίνεται η ε.μ.π. $g(\hat{\theta}_n)$ και το μέγιστο σφάλμα αυτής της εκτίμησης είναι (κατ' εκτίμηση) $\frac{|g'(\hat{\theta}_n)|}{\sqrt{nI_1(\hat{\theta}_n)}} z_{\alpha/2}$ με βαθμό εμπιστοσύνης κατά προσέγγιση $100(1-\alpha)\%$.

Το Α.Δ.Ε. στη σχέση (9.52) έχει μία βέλτιστη ασυμπτωτική ιδιότητα που σε αδρές γραμμές περιγράφουμε στη συνέχεια και η οποία πηγάζει από την ιδιότητα του ε.μ.π. $\hat{\theta}_n$ να είναι ασυμπτωτικά αμερόληπτος και ασυμπτωτικά αποδοτικός εκτιμητής (Ενότητα 7.2) με διασπορά κατά προσέγγιση ίση προς Κ.Φ. C-R $= \frac{1}{nI_1(\theta)}$. Έστω T_n ένας άλλος (ασθενώς) συνεπής εκτιμητής του θ, και ασυμπτωτικά κανονικός, δηλαδή κατ' αντιστοιχία με την (9.48) ισχύει

$$T_n \overset{\mathcal{L}_\theta}{\approx} \mathcal{N}\left(\theta, \frac{\sigma^2(\theta)}{n}\right). \tag{9.53}$$

Τότε αναμένουμε (και όντως ισχύει υπό ορισμένες συνθήκες) η ασυμπτωτική διασπορά του T_n, $\frac{\sigma^2(\theta)}{n}$, να είναι μεγαλύτερη (ή έστω ίση) από το Κ.Φ. C-R, δηλαδή

$$\frac{\sigma^2(\theta)}{n} \geq \frac{1}{nI_1(\theta)}. \tag{9.54}$$

Χρησιμοποιώντας τον εκτιμητή T_n και την (9.53) μπορούμε να κατασκευάσουμε το Α.Δ.Ε για το θ με σ.ε. $100(1-\alpha)\%$

$$T_n \pm \frac{|\sigma(T_n)|}{\sqrt{n}} z_{\alpha/2}. \tag{9.55}$$

Το μήκος αυτού του διαστήματος είναι

$$\begin{aligned}
2\frac{|\sigma(T_n)|}{\sqrt{n}} z_{\alpha/2} &\approx 2\frac{|\sigma(\theta)|}{\sqrt{n}} z_{\alpha/2} && (\text{συνέπεια του } T_n) \\
&\geq \frac{2}{\sqrt{nI_1(\theta)}} z_{\alpha/2} && (\text{λόγω της (9.55)}) \\
&\approx \frac{2}{\sqrt{nI_1(\hat{\theta}_n)}} z_{\alpha/2}, && (\text{συνέπεια του } \hat{\theta}_n)
\end{aligned}$$

που είναι το μήκος του Α.Δ.Ε. για το θ στη σχέση (9.51). Τελικά λοιπόν αναμένουμε το Α.Δ.Ε. που κατασκευάζεται χρησιμοποώντας τον ε.μ.π. του θ να έχει μικρότερο μήκος από το Α.Δ.Ε. που κατασκευάζεται χρησιμοποιώντας οποιονδήποτε άλλο συνεπή και ασυμπτωτικά κανονικό εκτιμητή. Η ιδιότητα αυτή σε ένα κάπως διαφορετικό πλαίσιο έχει δειχθεί από τον Wilks (1938) και περιέχεται στους Stuart et al (2004, τόμος 2, σελ. 134-135).

9.7 Μπεϋζιανά Διαστήματα

Σε αυτήν την ενότητα, θα αναφερθούμε πολύ σύντομα στην κατασκευή διαστημάτων για το θ ή το $g(\theta)$, που βασίζονται στην Μπεϋζιανή προσέγγιση. Όπως αναφέρθηκε στο Κεφάλαιο 8, σύμφωνα με την Μπεϋζιανή αντίληψη, η άγνωστη παράμετρος θ είναι τυχαία μεταβλητή με (εκ των προτέρων) πυκνότητα $\pi(\theta)$, $\theta \in \Theta$, ενώ η πυκνότητα των δεδομένων $\underset{\sim}{X} = (X_1, \ldots, X_n)$, $f(\underset{\sim}{x}; \theta)$ παριστάνει την δεσμευμένη πυκνότητα του $\underset{\sim}{X}$ δοθείσης της τιμής θ. Περαιτέρω, $f(\underset{\sim}{x}) = \int_\Theta f(\underset{\sim}{x}; \theta)\pi(\theta)\,\mathrm{d}\theta$ (ή στη διακριτή περίπτωση $f(\underset{\sim}{x}) = \sum\limits_{\theta \in \Theta} f(\underset{\sim}{x}; \theta)\pi(\theta)$) είναι η περιθώρια πυκνότητα του $\underset{\sim}{X}$ και

$$\pi(\theta \mid \underset{\sim}{x}) = \frac{f(\underset{\sim}{x}; \theta)\pi(\theta)}{\int_\Theta f(\underset{\sim}{x}; \theta)\pi(\theta)\,\mathrm{d}\theta}$$

είναι η δεσμευμένη πυκνότητα του θ δοθέντος του $\underset{\sim}{x}$, που στην Μπεϋζιανή ορολογία (αλλά και γενικότερα) αναφέρεται, όπως είδαμε, ως εκ των υστέρων πυκνότητα του θ. Υπενθυμίζουμε ότι η $\pi(\theta \mid \underset{\sim}{x})$ περιέχει τις επικαιροποιημένες πληροφορίες για το θ, δηλαδή τις πληροφορίες που παρέχουν τα δεδομένα $\underset{\sim}{X}$ σε συνδυασμό με τις αρχικές που περιέχει η $\pi(\theta)$.

Επειδή το θ είναι τυχαία μεταβλητή έχει έννοια να θεωρήσουμε πιθανοτικές παραστάσεις, όπως $\mathbb{P}(\alpha < \theta < \beta)$ ή $\mathbb{P}(\gamma < g(\theta) < \delta)$, όπου τα $\alpha, \beta, \gamma, \delta$ είναι σταθερές μη εξαρτώμενες από τα δεδομένα $\underset{\sim}{X}$ ή σταθερές που εξαρτώνται από την παρατηρηθείσα τιμή του $\underset{\sim}{X}$, $\underset{\sim}{x}$, ή ακόμη και συναρτήσεις του $\underset{\sim}{X}$. Ανάλογα αυτές οι πιθανότητες μπορούν να υπολογιστούν ως προς την $\pi(\theta)$ ή την $\pi(\theta \mid \underset{\sim}{x})$ ή την από κοινού πυκνότητα των $\underset{\sim}{X}$ και θ, $f(\underset{\sim}{x}; \theta)\pi(\theta)$.

Στη συνέχεια, θα ασχοληθούμε με διαστήματα που κατασκευάζονται χρησιμοποιώντας την εκ των υστέρων πυκνότητα $\pi(\theta \mid \underset{\sim}{x})$. Εάν $\underset{\sim}{x}$ είναι η παρατηρηθείσα τιμή του $\underset{\sim}{X}$ και $\alpha(\underset{\sim}{x}) \le \beta(\underset{\sim}{x})$ είναι σταθερές εξαρτώμενες από το $\underset{\sim}{x}$, το διάστημα $\big[\alpha(\underset{\sim}{x}), \beta(\underset{\sim}{x})\big]$ ονομάζεται Μπεϋζιανό διάστημα για το $g(\theta)$ πιθανότητας $100(1-\alpha)\%$ εάν

$$\mathbb{P}_{\theta \mid \underset{\sim}{x}}\big(\alpha(\underset{\sim}{x}) \le g(\theta) \le \beta(\underset{\sim}{x})\big) = 1 - \alpha, \tag{9.56}$$

όπου η πιθανότητα υπολογίζεται ως προς θ με πυκνότητα την $\pi(\theta \mid \underset{\sim}{x})$.

Σε αντίθεση με τα διαστήματα εμπιστοσύνης της κλασικής Στατιστικής, το Μπεϋζιανό διάστημα δεν έχει καμία δυσκολία να ερμηνευθεί: Η σχέση (9.56) σημαίνει ό,τι ακριβώς βλέπουμε, δηλαδή η πιθανότητα η τυχαία μεταβλητή $g(\theta)$ να βρίσκεται μεταξύ $\alpha(\underset{\sim}{x})$ και $\beta(\underset{\sim}{x})$ είναι $1-\alpha$. Λόγω της διαφορετικής ερμηνείας του και προς διάκριση από το κλασικό διάστημα εμπιστοσύνης, το $\left[\alpha(\underset{\sim}{x})\,,\,\beta(\underset{\sim}{x})\right]$ στην (9.56) αναφέρεται ως αξιόπιστο διάστημα (credible interval).

Παράδειγμα 9.7.1. (Κανονική κατανομή με γνωστή διασπορά – Μπεϋζιανό διάστημα για τη μέση τιμή) Έστω $\underset{\sim}{X}=(X_1,\ldots,X_n)$ ένα τυχαίο δείγμα από την κανονική κατανομή $\mathcal{N}(\theta,1)$, $\theta\in\Theta=\mathbb{R}$. Έστω ακόμη ότι η εκ των προτέρων πυκνότητα του θ, $\pi(\theta)$, είναι τυπική κανονική $\mathcal{N}(0,1)$. Στο Παράδειγμα 8.1.4 είδαμε ότι η εκ των υστέρων πυκνότητα $\pi(\theta\mid\underset{\sim}{x})$ είναι κανονική $\mathcal{N}\big(\frac{\sum x_i}{n+1},\frac{1}{n+1}\big)$. Θεωρώντας λοιπόν το θ ως τυχαία μεταβλητή με αυτήν την κατανομή έχουμε ότι η

$$Z=\sqrt{n+1}\left(\theta-\frac{\sum_{i=1}^{n}x_i}{n+1}\right)$$

είναι τυπική κανονική $\mathcal{N}(0,1)$. Επομένως από τη σχέση $\mathbb{P}\big(-z_{\alpha/2}\le Z\le z_{\alpha/2}\big)=1-\alpha$ διαδοχικά παίρνουμε

$$\mathbb{P}\Big(-z_{\alpha/2}\le\sqrt{n+1}\left(\theta-\frac{\sum_{i=1}^{n}x_i}{n+1}\right)\le z_{\alpha/2}\Big)=1-\alpha,$$

$$\mathbb{P}\Big(\frac{\sum_{i=1}^{n}x_i}{n+1}-\frac{1}{\sqrt{n+1}}z_{\alpha/2}\le\theta\le\frac{\sum_{i=1}^{n}x_i}{n+1}+\frac{1}{\sqrt{n+1}}z_{\alpha/2}\Big)=1-\alpha.$$

Σύμφωνα με την (9.56), το Μπεϋζιανό διάστημα για το θ πιθανότητας $100(1-\alpha)\%$ είναι

$$\Big[\frac{\sum_{i=1}^{n}x_i}{n+1}-\frac{1}{\sqrt{n+1}}z_{\alpha/2}\,,\,\frac{\sum_{i=1}^{n}x_i}{n+1}+\frac{1}{\sqrt{n+1}}z_{\alpha/2}\Big]. \tag{9.57}$$

Συγκρίνοντας με το κλασικό διάστημα $\left[\bar{X}-\frac{1}{\sqrt{n}}z_{\alpha/2}\,,\,\bar{X}+\frac{1}{\sqrt{n}}z_{\alpha/2}\right]$, βλέπε Ενότητα 9.4.1, παρατηρούμε ότι τα δύο διαστήματα έχουν διαφορετικό

μέσο (ένας Μπεϋζιανός Στατιστικός θα εκτιμούσε το θ με $\frac{\sum_{i=1}^{n} x_i}{n+1}$, ενώ ένας κλασικός Στατιστικός θα χρησιμοποιούσε τον δειγματικό μέσο $\bar{x}$) και ακόμη ότι το Μπεϋζιανό διάστημα έχει μικρότερο μήκος, γεγονός αναμενόμενο αφού η κατασκευή του βασίζεται όχι μόνον στα δεδομένα αλλά και στις πληροφορίες που παρέχει η εκ των προτέρων κατανομή. Πάντως, καθώς $n \to \infty$, $n+1 \approx n$ και τα δύο διαστήματα σχεδόν συμπίπτουν, το οποίο σημαίνει ότι η επίδραση της $\pi(\theta)$ εξασθενεί όταν υπάρχει μεγάλος αριθμός δεδομένων. Στη Άσκηση 9.16 ζητείται να δειχθεί ότι το ίδιο φαινόμενο ισχύει και στην πιο γενική περίπτωση που η $\pi(\theta)$ είναι $\mathcal{N}(0, \tau^2)$, $\tau > 0$.
Το διάστημα στην (9.57) έχει, επιπλέον, ελάχιστο μήκος μεταξύ (όλων) των Μπεϋζιανών διαστημάτων

$$\Big[\frac{\sum_{i=1}^{n} x_i}{n+1} + \frac{1}{\sqrt{n+1}} c_1 \, , \, \frac{\sum_{i=1}^{n} x_i}{n+1} + \frac{1}{\sqrt{n+1}} c_2\Big],$$

όπου οι σταθερές c_1 και c_2 είναι τέτοιες ώστε $\mathbb{P}\big(c_1 \le Z \le c_2\big) = 1 - \alpha$. Η απόδειξη αυτής της βέλτιστης ιδιότητας βασίζεται στην Πρόταση 9.4.1 και είναι ακριβώς η ίδια με την αντίστοιχη του κλασικού διαστήματος $\Big[\bar{X} - \frac{1}{\sqrt{n}} z_{\alpha/2} \, , \, \bar{X} + \frac{1}{\sqrt{n}} z_{\alpha/2}\Big]$.

Η επόμενη πρόταση αντιμετωπίζει γενικά το πρόβλημα κατασκευής Μπεϋζιανού διαστήματος ελαχίστου μήκους.

Πρόταση 9.7.1. *Έστω ότι η εκ των υστέρων πυκνότητα $\pi(\theta \mid \underset{\sim}{x})$ είναι συνεχής ως συνάρτηση του θ και μονοκόρυφη (unimodal), δηλαδή αρχικά γνησίως αύξουσα και μετά γνησίως φθίνουσα. Τότε το διάστημα $\big[\alpha(\underset{\sim}{x}) \, , \, \beta(\underset{\sim}{x})\big]$ όπου*

$$\int_{\alpha(\underset{\sim}{x})}^{\beta(\underset{\sim}{x})} \pi(\theta \mid \underset{\sim}{x}) d\theta = 1 - \alpha \quad \textit{και} \quad \pi(\alpha(\underset{\sim}{x}) \mid \underset{\sim}{x}) = \pi(\beta(\underset{\sim}{x}) \mid \underset{\sim}{x})$$

είναι Μπεϋζιανό διάστημα ελαχίστου μήκους για το θ πιθανότητας $100(1 - \alpha)\%$.

Απόδειξη. Ένα Μπεϋζιανό διάστημα για το θ πιθανότητας $100(1-\alpha)\%$ έχει τη μορφή $\big[c_1, c_2\big]$ όπου $\int_{c_1}^{c_2}\pi(\theta \mid \underset{\sim}{x})d\theta = 1-\alpha$ (και οι σταθερές c_1, c_2 εξαρτώνται, εν γένει, από το $\underset{\sim}{x}$). Το μήκος του είναι

$$\ell = c_2 - c_1 = \int_{c_1}^{c_2} \mathrm{I}_{[c_1,c_2]}(\theta)d\theta = \int_{c_1}^{c_2}\delta(\theta)d\theta,$$

όπου $\delta(\theta) = \mathrm{I}_{[c_1,c_2]}(\theta)$. Σύμφωνα με την Πρόταση 9.4.1, το ολοκλήρωμα ℓ ελαχιστοποιείται ως προς $\delta(\theta)$ υπό τον περιορισμό

$$\int_{c_1}^{c_2}\pi(\theta \mid \underset{\sim}{x})d\theta = \int_{-\infty}^{\infty}\delta(\theta)\pi(\theta \mid \underset{\sim}{x})d\theta = 1-\alpha$$

όταν $\delta(\theta) = \mathrm{I}_{\left\{\theta:1\le\kappa\,\pi(\theta|\underset{\sim}{x})\right\}}(\theta) = \mathrm{I}_{\left\{\theta:\pi(\theta|\underset{\sim}{x})\ge c\right\}}(\theta)$. όπου $c=\frac{1}{\kappa}$, εφόσον το σύνολο $\big\{\theta : \pi(\theta \mid \underset{\sim}{x}) \ge c\big\}$ είναι διάστημα της μορφής $\big[c_1, c_2\big]$. Πράγματι, επειδή η $\pi(\theta \mid \underset{\sim}{x})$ είναι συνεχής και μονοκόρυφη, η σχέση $\pi(\theta \mid \underset{\sim}{x}) \ge c$ είναι ισοδύναμη με την

$$c_1 \le \theta \le c_2 \quad \text{με} \quad \pi(c_1 \mid \underset{\sim}{x}) = \pi(c_2 \mid \underset{\sim}{x})(=c).$$

Επομένως, το μήκος $\ell = c_2 - c_1$ ελαχιστοποιείται για c_1 και c_2 που ικανοποιούν τις σχέσεις

$$\int_{c_1}^{c_2}\pi(\theta \mid \underset{\sim}{x})d\theta = 1-\alpha \quad \text{και} \quad \pi(c_1 \mid \underset{\sim}{x}) = \pi(c_2 \mid \underset{\sim}{x}). \tag{9.58}$$

Το σύστημα των σχέσεων (9.58) έχει μοναδική λύση ως προς c_1 και c_2 (με το ίδιο σκεπτικό όπως αυτό της κατασκευής Δ.Ε. ελαχίστου μήκους για το σ^2 στην Ενότητα 9.4.1). Θέτουμε $\alpha(\underset{\sim}{x}) = c_1$, $\beta(\underset{\sim}{x}) = c_2$ και η απόδειξη είναι πλήρης. □

Παρατήρηση 9.7.1. Επειδή τα σημεία θ του $\big[\alpha(\underset{\sim}{x})\,,\,\beta(\underset{\sim}{x})\big]$ ικανοποιούν τη σχέση $\pi(\theta \mid \underset{\sim}{x}) \ge c$, δηλαδή αποδίδουν «μεγάλη τιμή» στην εκ των υστέρων πυκνότητα, δια τούτο το $\big[\alpha(\underset{\sim}{x})\,,\,\beta(\underset{\sim}{x})\big]$ αναφέρεται ως διάστημα μέγιστης εκ των υστέρων πυκνότητας (highest posterior density interval, Casella and Berger, 2002, Παπαϊωάννου και Φερεντίνος, 2001).

Παράδειγμα 9.7.2. (Κατανομή Poisson – Μπεϋζιανό διάστημα ελαχίστου μήκους για τη μέση τιμή) Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την κατανομή Poisson, $\mathcal{P}(\theta)$, $\theta \in \Theta = (0, \infty)$. Στο Παράδειγμα 8.1.3 είδαμε ότι χρησιμοποιώντας ως εκ των προτέρων κατανομή την Γάμμα $\mathcal{G}(\alpha, \beta)$, η εκ των υστέρων πυκνότητα $\pi(\theta \mid \underset{\sim}{x})$ είναι (επίσης) Γάμμα $\mathcal{G}(\sum\limits_{i=1}^{n} x_i + \alpha, \frac{1}{n+1/\beta})$, δηλαδή

$$\pi(\theta \mid \underset{\sim}{x}) = \frac{1}{\Gamma(\alpha_1)\beta_1^{\alpha_1}} \theta^{\alpha_1 - 1} e^{-\theta/\beta_1}, \ \theta > 0.$$

όπου $\alpha_1 = \sum\limits_{i=1}^{n} x_i + \alpha$, $\beta_1 = \frac{1}{n+1/\beta}$. Είναι εύκολο να δειχθεί ότι η $\pi(\theta \mid \underset{\sim}{x})$ είναι γνησίως αύξουσα στο $(0, m)$ και γνησίως φθίνουσα στο (m, ∞) όπου $m = (\alpha_1 - 1)\beta_1$. Επομένως από την Πρόταση 9.7.1, το Μπεϋζιανό διάστημα ελαχίστου μήκους για το θ πιθανότητας $100(1-\alpha)\%$ είναι $\big[\alpha(\underset{\sim}{x})\,,\, \beta(\underset{\sim}{x})\big]$ όπου $c_1 = \alpha(\underset{\sim}{x})$ και $c_2 = \beta(\underset{\sim}{x})$ είναι η λύση του συστήματος

$$c_1^{\alpha_1 - 1} e^{-c_1/\beta_1} = c_2^{\alpha_1 - 1} e^{-c_2/\beta_1} \quad \text{και} \quad \int_{c_1}^{c_2} \frac{1}{\Gamma(\alpha_1)\beta_1^{\alpha_1}} \theta^{\alpha_1 - 1} e^{-\theta/\beta_1} d\theta = 1 - \alpha$$

(η οποία μπορεί να βρεθεί μόνον αριθμητικά).

9.8 Ασκήσεις

9.1. Έστω X μία παρατήρηση από την κατανομή Βήτα

$$f(x; \theta) = \theta x^{\theta - 1}, \ 0 < x < 1, \ \theta > 0.$$

Να κατασκευαστεί Δ.Ε. ελαχίστου μήκους για το θ με σ.ε. $100(1-\alpha)\%$. (Υπόδειξη: βλέπε Παράδειγμα 9.2.1)

9.2. Να βρεθεί η ελάχιστη τιμή του μεγέθους δείγματος n έτσι ώστε το 90% Δ.Ε. για τη μέση τιμή θ της κανονικής κατανομής $\mathcal{N}(\theta, \sigma^2)$ με σ^2 γνωστό να έχει μήκος το πολύ $\sigma/5$.

9.3. Να υπολογιστεί η πιθανότητα το διάστημα $\bar{X} \pm 8.2/\sqrt{n}$ να περιέχει τη μέση τιμή θ της κατανομής $\mathcal{N}(\theta, 25)$.

9.4. Έστω X μία παρατήρηση από την εκθετική κατανομή $\mathcal{E}(\theta)$. Να υπολογιστεί η σταθερά κ ώστε το διάστημα $(0, \kappa X)$ να είναι Δ.Ε. για το θ με σ.ε. $100(1-\alpha)\%$.

9.5. Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την εκθετική κατανομή $\mathcal{E}(\theta)$. Να δειχθεί ότι η τυχαία μεταβλητή $T = \frac{2}{\theta}\sum_{i=1}^{n} X_i$ είναι ποσότητα οδηγός με κατανομή χ^2_{2n} και να κατασκευαστούν τα Δ.Ε. ίσων ουρών και ελαχίστου μήκους για το θ με σ.ε. $100(1-\alpha)\%$ χρησιμοποιώντας την T. Να υπολογιστεί το Δ.Ε. ίσων ουρών για $n = 7$, $\bar{x} = 93.6$, $\alpha = 0.10$.

9.6. Για να εξαχθούν συμπεράσματα για το μέσο χρόνο ζωής ενός νέου τύπου ηλεκτρικών λυχνιών, καταγράφηκαν οι χρόνοι ζωής σε ώρες ενός τυχαίου δείγματος 25 λυχνιών: 713, 721, 699, 701, 685, 711, 708, 720, 695, 687, 700, 712, 690, 697, 701, 705, 708, 715, 709, 725, 720, 730, 707, 691, 696. Να εκτιμηθεί ο μέσος χρόνος ζωής θ των λυχνιών και να κατασκευαστεί ένα Δ.Ε. για το θ με σ.ε. κατά προσέγγιση 95%.

9.7. Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την ομοιόμορφη κατανομή $\mathcal{U}(-\theta, \theta)$, $\theta \in \Theta = (0, \infty)$. Να κατασκευαστούν Δ.Ε. ίσων ουρών και ελαχίστου μήκους για το θ με σ.ε. $100(1-\alpha)\%$ χρησιμοποιώντας την (ελάχιστη) επαρκή στατιστική συνάρτηση.

9.8. Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την κατανομή

$$f(x;\theta) = \begin{cases} e^{-(x-\theta)}, & x \geq \theta \\ 0, & \text{διαφορετικά} \end{cases}$$

με $\theta \in \Theta = \mathbb{R}$. Να δειχθεί ότι η τυχαία μεταβλητή $T = X_{(1)} - \theta$ είναι ποσότητα οδηγός και να κατασκευαστεί Δ.Ε. ελαχίστου μήκους για το θ με σ.ε. $100(1-\alpha)\%$.

9.9. Ένα τυχαίο δείγμα 5 παρατηρήσεων από την κατανομή $\mathcal{N}(\mu_1, \sigma_1^2)$ έδωσε $S_1^2 = 1759$ και ένα ανεξάρτητο προς αυτό τυχαίο δείγμα 7 παρατηρήσεων από την κατανομή $\mathcal{N}(\mu_2, \sigma_2^2)$ έδωσε $S_2^2 = 2273$. Να κατασκευαστεί ένα Δ.Ε. για το λόγο σ_1^2/σ_2^2 με σ.ε. 98%. Ακόμη να κατσκευαστεί ένα Δ.Ε. για το σ_1/σ_2 με το ίδιο σ.ε. (μ_1, μ_2 είναι άγνωστα).

9.10. Έστω $\underset{\sim}{X} = (X_1, \ldots, X_{n_1})$ ένα τυχαίο δείγμα από την εκθετική κατανομή $\mathcal{E}(\theta_1)$ και $Y_1, \ldots, Y_{n_2}$ ένα τυχαίο δείγμα από την εκθετική κατανομή $\mathcal{E}(\theta_2)$. Τα δύο δείγματα είναι ανεξάρτητα. Να κατασκευαστεί ένα Δ.Ε. ίσων ουρών για το λόγο θ_1/θ_2.
(Υπόδειξη: βλέπε Άσκηση 9.5)

9.11. Να κατασκευαστεί ένα Δ.Ε. για το λόγο σ_1^2/σ_2^2 των διασπορών δύο κανονικών κατανομών $\mathcal{N}(\mu_1, \sigma_1^2)$ και $\mathcal{N}(\mu_2, \sigma_2^2)$ με μ_1 γνωστό και μ_2 άγνωστο, χρησιμοποιώντας ανεξάρτητα τυχαία δείγματα από κάθε μία κατανομή.

9.12. Όπως στην Άσκηση 9.11 αλλά με μ_1 άγνωστο και μ_2 γνωστό.

9.13. Έστω $\bar{X}$, $\bar{Y}$, S_1^2, S_2^2 οι δειγματικοί μέσοι και δειγματικές διασπορές από δύο ανεξάρτητα τυχαία δείγματα μεγέθους n_1, n_2 αντίστοιχα από τις κατανομές $\mathcal{N}(\mu_1, \sigma_1^2)$, $\mathcal{N}(\mu_2, \sigma_2^2)$ με μ_1, μ_2, σ_1^2, σ_2^2 άγνωστα αλλά με $\sigma_1^2/\sigma_2^2 = d$, όπου d γνωστή σταθερά. Να δειχθούν τα εξής.

1. $$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{d\sigma_2^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathcal{N}(0, 1).$$

2. $$\frac{(n_1 - 1)S_1^2}{d\sigma_2^2} + \frac{(n_2 - 1)S_2^2}{\sigma_2^2} \sim \chi^2_{n_1+n_2-2}\ .$$

3. Οι τυχαίες μεταβλητές στα 1 και 2 είναι ανεξάρτητες.

4. Χρησιμοποιώντας τα παραπάνω, να κατασκευαστεί μία τυχαία μεταβλητή (που δεν εξαρτάται από το σ_2^2) με κατανομή $t_{n_1+n_2-2}$ και να κατασκευαστεί Δ.Ε. για το $\mu_1 - \mu_2$ χρησιμοποιώντας αυτήν την τυχαία μεταβλητή ως ποσότητα οδηγό.

9.14. Σε ένα τυχαίο δείγμα 222 ατόμων που διαβάζουν τακτικά εφημερίδα βρέθηκε ότι 38 από αυτά διαβάζουν μία συγκεκριμένη εφημερίδα. Να εκτιμηθεί το ποσοστό των ατόμων που διαβάζουν αυτήν την εφημερίδα και να κατασκευαστεί ένα Δ.Ε. για αυτό με σ.ε. κατά προσέγγιση 95%.

9.15. Έστω X μία παρατήρηση από την κατανομή

$$f(x;\theta) = \begin{cases} \frac{2}{\theta}(1-\frac{x}{\theta}), & 0 < x < \theta \\ 0, & \text{διαφορετικά}, \end{cases}$$

όπου $\theta > 0$ άγνωστη παράμετρος. Να κατασκευαστεί ένα Δ.Ε. για το θ με σ.ε. $100(1-\alpha)\%$.

9.16. Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την κανονική κατανομή $\mathcal{N}(\theta, 1)$, $\theta \in \Theta = \mathbb{R}$. Η εκ των προτέρων κατανομή του θ είναι $\mathcal{N}(0, \tau^2)$, $\tau > 0$.

α. Να δειχθεί ότι η εκ των υστέρων κατανομή είναι επίσης κανονική και να προσδιοριστούν οι παράμετροι της (μέση τιμή και διασπορά).

β. Να κατασκευαστεί ένα Μπεϋζιανό διάστημα για το θ πιθανότητας $100(1-\alpha)\%$.

γ. Να δειχθεί ότι το Μπεϋζιανό διάστημα του (β) δεν διαφέρει σημαντικά από το κλασικό διάστημα καθώς $n \to \infty$.

9.17. Έστω $\underset{\sim}{X} = (X_1, \ldots, X_n)$ ένα τυχαίο δείγμα από την κατανομή Cauchy με παράμετρο θέσης $\theta \in \Theta = \mathbb{R}$ και πυκνότητα $f_1(x;\theta) = \dfrac{1}{\pi(1+(x-\theta)^2)}$, $x \in \mathbb{R}$. Μπορεί να δειχθεί ότι η τυχαία μεταβλητή $\sum_{i=1}^n X_i$ έχει την ίδια πυκνότητα όπως κάθε παρατήρηση X_i, βλέπε Johnson, Kotz and Balakrishnan (1994, τόμος 1, σελ. 301).

α. Να δειχθεί ότι η $T(\underset{\sim}{X}, \theta) = \sum_{i=1}^n X_i - \theta$ είναι ποσότητα οδηγός με πυκνότητα t_1 (t με ένα βαθμό ελευθερίας).

β. Να κατασκευαστεί Δ.Ε. ίσων ουρών για το θ με σ.ε. $100(1-\alpha)\%$ και να δειχθεί ότι συμπίπτει με το Δ.Ε. ελαχίστου μήκους.

γ. Να κατασκευαστούν κάτω και άνω φράγμα εμπιστοσύνης για το θ με σ.ε. $100(1-\alpha)\%$.

Βιβλιογραφία

1. Casella, G. and Berger, R.L. (2002). *Statistical inference*. Pacific Grove, CA: Wadsworth/Brooks Cole; 2nd edition.

2. Clopper, C.J. and Pearson E.S. (1934). The use of confidence or fiducial limits illustrated in the case of the Binomial, *Biometrika*, **26**, 404–413.

3. Ferentinos, K. and Kourouklis, S. (1990). Shortest confidence interval estimation for families of distributions involving two truncation parameters. *Metrika*, **37**, 353–363.

4. Johnson, N.L., Kotz, S. and Balakrishnan, N. (1994). *Continuous univariate distributions. Vol.1.* John Wiley & Sons; 2nd edition.

5. Juola, R.C. (1993). More on shortest confidence intervals. *The American Statistician*, **47**, 117–119.

6. Stuart A., Ord, J,K. and Arnold, S. (2004). *Kendall's advanced theory of statistics. Vol.2A. Classical inference and the linear model.* John Wiley & Sons; 6th edition.

7. Wilks, S.S. (1938). Shortest average confidence intervals from large samples, *The Annals of Mathematical Statistics*, **9**, 166–175.

8. Παπαϊωάννου, Π. και Φερεντίνος, Κ. (2001). *Μαθηματική Στατιστική, Εκτιμητική - Έλεγχος Υποθέσεων - Εφαρμογές*. Σταμούλης· 2η έκδοση.

Αγγλική Ορολογία

Statistical Inference	Στατιστική Συμπερασματολογία
Point Estimation	Εκτιμητική
Confidence Interval	Διάστημα Εμπιστοσύνης
Test of Statistical Hypotheses	Έλεγχος Στατιστικών Υποθέσεων
parameter space	παραμετρικός χώρος
statistic	στατιστική συνάρτηση
estimator	εκτιμητής
estimation	εκτίμηση
optimal estimator	βέλτιστος εκτιμητής
mean squared error	μέσο τετραγωνικό σφάλμα
mean absolute error	μέσο απόλυτο σφάλμα
loss function	συνάρτηση ζημίας
risk function	συνάρτηση κινδύνου
sample mean	δειγματικός μέσος
sample variance	δειγματική διασπορά
substitution principle	αρχή της αντικατάστασης
maximum likelihood estimator	εκτιμητής μέγιστης πιθανοφάνειας
maximum likelihood principle	αρχή της μέγιστης πιθανοφάνειας
method of moments	μέθοδος των ροπών
method of moments estimator	εκτιμητής μεθόδου ροπών
standard error	τυπικό σφάλμα
standard deviation	τυπική απόκλιση
bias	μεροληψία
biased estimator	μεροληπτικός εκτιμητής
unbiasedness	αμεροληψία

unbiased estimator	αμερόληπτος εκτιμητής
admissible estimator	αποδεκτός εκτιμητής
inadmissible estimator	μη αποδεκτός εκτιμητής
admissibility	αποδεκτικότητα
underestimation	υποεκτίμηση
overestimation	υπερεκτίμηση
Cramér–Rao inequality	ανισότητα των Cramér–Rao
Cramér–Rao lower bound	κάτω φράγμα των Cramér–Rao
Fisher information	πληροφορία του Fisher
efficiency	αποδοτικότητα
efficient estimator	αποδοτικός εκτιμητής
information number	αριθμός πληροφορίας
one parameter exponential family of distributions	μονοπαραμετρική εκθετική οικογένεια κατανομών
sufficiency	επάρκεια
sufficient statistic	επαρκής στατιστική συνάρτηση
completeness	πληρότητα
complete statistic	πλήρης στατιστική συνάρτηση
uniformly minimum variance unbiased estimator (UMVUE)	αμερόληπτος εκτιμητής ομοιομόρφως ελάχιστης διασποράς (ΑΟΕΔ εκτιμητής)
minimal sufficient statistic	ελάχιστη επαρκής στατιστική συνάρτηση
Rao–Blackwell improvement	Rao–Blackwell βελτίωση
location parameter	παράμετρος θέσης
scale parameter	παράμετρος κλίμακας
likelihood function	συνάρτηση πιθανοφάνειας
(weak or strong) consistency	(ασθενής ή ισχυρή) συνέπεια
strongly consistent estimator	ισχυρά συνεπής εκτιμητής
weakly consistent estimator	ασθενώς συνεπής εκτιμητής
asymptotic unbiasedness	ασυμπτωτική αμεροληψία
asymptotic distribution	ασυμπτωτική κατανομή
Weak Law of Large Numbers	Ασθενής Νόμος των Μεγάλων Αριθμών (ΑΝΜΑ)

Strong Law of Large Numbers	Ισχυρός Νόμος των Μεγάλων Αριθμών (ΙΝΜΑ)
Central Limit Theorem	Κεντρικό Οριακό Θεώρημα
Convergence in Probability	Σύγκλιση κατά Πιθανότητα
(or Weak Convergence)	(ή Ασθενής Σύγκλιση)
Convergence with Probability 1	Σύγκλιση με Πιθανότητα 1
(or Strong Convergence)	(ή Ισχυρή Σύγκλιση)
Convergence in Law	Σύγκλιση κατά Κατανομή
Bayes estimator	εκτιμητής Bayes
prior distribution	εκ των προτέρων κατανομή
(or density)	(ή πυκνότητα)
posterior distribution	εκ των υστέρων κατανομή
(or density)	(ή πυκνότητα)
Bayes risk	κίνδυνος Bayes
minimax estimator	εκτιμητής minimax
confidence coefficient	συντελεστής εμπιστοσύνης
pivotal quantity	ποσότητα οδηγός
equal tails confidence interval	διάστημα εμπιστοσύνης ίσων ουρών
asymptotic confidence interval	ασυμπτωτικό διάστημα εμπιστοσύνης

Ευρετήριο

άγνωστη παράμετρος, 51
άνω φράγμα εμπιστοσύνης, 315
Rao–Blackwell βελτίωση, 171

αληθής τιμή του θ, 225
αμερόληπτος εκτιμητής, 78
αμεροληψία, 78
ανισότητα Cauchy–Schwarz, 21
ανισότητα Jensen, 13
ανισότητα Markov, 12
ανισότητα Tchebychev, 12
ανισότητα των Cramér–Rao, 105, 112
ΑΟΕΔ εκτιμητής, 172
απόλυτο σφάλμα, 294
αποδεκτός εκτιμητής με κριτήριο το ΜΤΣ, 74
αποδοτικός εκτιμητής, 121
αριθμός πληροφορίας του Fisher, 112
αρνητική μεροληψία, 85
αρχή της αντικατάστασης, 58
αρχή της μέγιστης πιθανοφάνειας, 60, 222
Ασθενής Νόμος των Μεγάλων Αριθμών (ΑΝΜΑ), 32
ασθενώς συνεπής εκτιμητής, 248
ασυμπτωτικά αποδοτικός εκτιμητής, 256
ασυμπτωτικά κανονικός εκτιμητής, 256
ασυμπτωτική αμεροληψία, 253
ασυμπτωτικό διάστημα εμπιστοσύνης (Α.Δ.Ε.), 314, 356

βέλτιστος εκτιμητής ειδικής μορφής, 207
βελτίωση εκτιμητών, 169

δειγματική διασπορά, 83
δειγματική ροπή, 275
δειγματικό ανάλογο, 57
δειγματικός μέσος, 57, 82
δεσμευμένη διασπορά, 29
δεσμευμένη μέση τιμή, 29
διάστημα εμπιστοσύνης ίσων ουρών, 322
διάστημα εμπιστοσύνης (Δ.Ε.), 314
διάστημα εμπιστοσύνης ελαχίστου μήκους, 326,

338

εκ των προτέρων κατανομή (ή πυκνότητα), 289
εκ των υστέρων κατανομή (ή πυκνότητα), 291
Εκθετική Οικογένεια Κατανομών τάξης κ (Ε.Ο.Κ.), 191
εκτίμηση, 52
εκτιμητής, 52
εκτιμητής Bayes, 289
εκτιμητής minimax, 305
εκτιμητής μέγιστης πιθανοφάνειας, 225, 243
εκτιμητής μεθόδου ροπών (ε.μ.ρ.), 275
επάρκεια, 152
επαρκής στατιστική συνάρτηση, 152
εξίσωση πιθανοφάνειας, 226

γενικευμένη συνάρτηση πιθανοφάνειας, 243

ισχυρά συνεπής εκτιμητής, 248
Ισχυρός Νόμος των Μεγάλων Αριθμών (ΙΝΜΑ), 34

Θεώρημα Slutsky, 37
Θεώρημα Basu, 199
Θεώρημα Lehmann–Scheffé, 176

κάτω φράγμα εμπιστοσύνης, 315
κάτω φράγμα των Cramér–Rao (Κ.Φ. C-R), 112
κίνδυνος Bayes, 55
κίνδυνος Bayes, 289
καλύτερος εκτιμητής με κριτήριο το ΜΤΣ, 73
κατανομή $\mathcal{F}_{n,m}$, 354
κατανομή t_m, 339
Κεντρικό Οριακό Θεώρημα (ΚΟΘ), 35
κριτήριο του ΜΤΣ, 73

Μέθοδος Δέλτα, 265
μέσο απόλυτο σφάλμα, 53
μέσο τετραγωνικό σφάλμα (ΜΤΣ), 54, 63
μεροληπτικός εκτιμητής, 80
μεροληψία, 65
μη αποδεκτός εκτιμητής με κριτήριο το ΜΤΣ, 74
Μονοπαραμετρική Εκθετική Οικογένεια Κατανομών (Μ.Ε.Ο.Κ.), 118
Μπεϋζιανή Στατιστική, 290

πίνακας πληροφορίας του Fisher, 139
παράμετρος θέσης, 208
παράμετρος κλίμακος, 211
παραγοντικό κριτήριο των Neyman–Fisher, 156

Παραμετρική Στατιστική Συμπερασματολογία, 42
παραμετρικός χώρος, 51
παρατηρηθείσα τιμή, 52
πλήρης στατιστική συνάρτηση, 173
πληρότητα, 173
ποσότητα οδηγός, 318

ροπή κ τάξης κατανομής, 274

σύγκλιση
- κατά κατανομή, 34
- κατά πιθανότητα, 32
- με πιθανότητα 1, 33

στατιστική συνάρτηση, 52
συνάρτηση κατανομής, 7
συνάρτηση κινδύνου, 54, 288
συνάρτηση πιθανοφάνειας (ή πιθανοφάνεια), 222
συνάρτηση ζημίας, 54, 288
συνέπεια, 248
συνδιασπορά, 21
συντελεστής εμπιστοσύνης (σ.ε.), 314
συζυγής οικογένεια εκ των προτέρων κατανομών, 301

τυπικό σφάλμα, 63
τυχαίο δείγμα, 52
τυχαίο πείραμα, 1

υπερεκτίμηση, 99
υποεκτίμηση, 85, 99