Εφαρμοσμένη Στατιστική

Δημήτριος Μπάγκαβος

Τμήμα Μαθηματικών και Εφαρμοσμένων Μαθηματικών
Πανεπιστήμιο Κρήτης

28 Φεβρουαρίου 2018

Βασικοί ορισμοί.

Ορισμός 1: Τυχαίο δείγμα.

Τυχαίο δείγμα μεγέθους n από την κατανομή $F(x;\boldsymbol{\theta})$ (ή σ.π.π. $f(x;\boldsymbol{\theta})$) καλείται μια συλλογή ανεξάρτητων και ισόνομων τυχαίων μεταβλητών $X_1, X_2, \ldots, X_n \sim F(x;\boldsymbol{\theta})$ (ή ισοδύναμα $\sim f(x;\boldsymbol{\theta})$).

- Το τυχαίο δείγμα είναι στην ουσία μετρήσεις μιας τυχαίας μεταβλητής $X \sim f(x;\boldsymbol{\theta})$.
- Π.χ. αν μας ενδιαφέρει το εισόδημα X των νοικοκυριών της Ελλάδας, η $f(x;\boldsymbol{\theta})$ είναι η κατανομή των εισοδημάτων όλων των νοικοκυριών της χώρας και οι $X_1, X_2, \ldots, X_n$ είναι μετρήσεις που πήραμε ρωτώντας στην τύχη (αντιπροσωπευτικά) n νοικοκυριά.

Ορισμός 2: Δειγματοληπτικός χώρος.

Δειγματοληπτικός χώρος καλείται το σύνολο των δυνατών τιμών του δείγματος (π.χ. αν $X_i \in \mathbb{R}$, τότε ο δειγματοληπτικός χώρος είναι ο $\mathbb{R}^n$). Επίσης, παραμετρικός χώρος Θ καλείται το σύνολο των επιτρεπτών τιμών των παραμέτρων $\boldsymbol{\theta}$ (π.χ. αν $\boldsymbol{\theta} = (\mu, \sigma^2)$ τότε ο παραμετρικός χώρος είναι ο $\mathbb{R} \times (0, \infty)$).

Βασικοί ορισμοί.

Ορισμός 3: Στατιστική (ή δειγματική) συνάρτηση.

Έστω τυχαίο δείγμα $X_1, X_2, \ldots, X_n \sim F(x; \boldsymbol{\theta})$. Στατιστική (ή δειγματική) συνάρτηση (σ.σ.) καλείται κάθε μετρήσιμη συνάρτηση $T = T(\mathbf{X}) = T(X_1, X_2, \ldots, X_n)$ των $X_1, X_2, \ldots, X_n$, που δεν εξαρτάται από άγνωστες παραμέτρους.

Οι πλέον συνήθεις στατιστικές συναρτήσεις που εμφανίζονται σε προβλήματα της Στατιστικής είναι οι επόμενες:

$$\bar{X} = n^{-1}\sum_{i=1}^{n} X_i,\ m_r = n^{-1}\sum_{i=1}^{n} X_i^r,\ \sigma^2 = n^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

- Π.χ. αν μας ενδιαφέρει να εκτιμήσουμε το χρόνο ζωής των παραγόμενων λαμπών μιας εταιρείας τότε είναι λογικό να πάρουμε δείγμα n λαμπών, και να χρησιμοποιήσουμε το μέσο όρο τους $\bar{X}$ ως εκτιμήτρια του χρόνου ζωής μιας οποιασδήποτε λάμπας.
- Προφανώς, μία εκτιμήτρια συνάρτηση $T(X)$ είναι και αυτή μία τυχαία μεταβλητή (κάθε φορά που παίρνουμε ένα άλλο τυχαίο δείγμα X η T δίνει διαφορετική τιμή).

Βασικοί ορισμοί.

- Σημαντική παρατήρηση: Πως συνδέονται οι στατιστικές συναρτήσεις ($\bar{X}, m_r, \sigma^2$, κ.λ.π.) που μελετάμε με το γεγονός ότι οι παρατηρήσεις $X_1, X_2, \ldots, X_n$ που τις απαρτίζουν θεωρούνται ανεξάρτητες και ισόνομες προερχόμενες από κατανομή $f(x; \boldsymbol{\theta})$;
- Απάντηση: Καταρχήν παρατηρούμε ότι οι εν λόγω συναρτήσεις ορίζονται (όλες) ως κάποιας μορφής άθροισμα των $X_1, X_2, \ldots, X_n$.
- Ιδιότητα ανεξαρτησίας: Το ότι οι $X_1, X_2, \ldots, X_n$ είναι ανεξάρτητες σημαίνει ότι

$$\mathbb{V}\left\{\sum_{i=1}^{n} f(X_i)\right\} = \sum_{i=1}^{n} \mathbb{V}(f(X_i)) \tag{1}$$

όπου f κάποια συνάρτηση των X_i: για παράδειγμα:

  - στην περίπτωση του $\bar{X}$, $f(x) = x$
  - στην περίπτωση του m_r, $f(x) = x^r$
  - στην περίπτωση του σ^2, $f(x) = (x - \bar{X})^2$

Η σχέση

$$\mathbb{E}\sum_{i=1}^{n} f(X_i) = \sum_{i=1}^{n} \mathbb{E}f(X_i), \tag{2}$$

- ισχύει ανεξάρτητα του αν οι $X_1, X_2, \ldots, X_n$ είναι ανεξάρτητες ή όχι.

Βασικοί ορισμοί.

- Η (1) είναι συνέπεια του ότι για 2 ανεξάρτητες τ.μ. X, Y, $\mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y)$. Τι ισχύει αν X, Y δεν είναι ανεξάρτητες;
- Ιδιότητα ισονομίας: Το ότι οι $X_1, X_2, \ldots, X_n$ είναι ισόνομες σημαίνει ότι η κάθε μία από τις X_i ακολουθεί ίδια ακριβώς κατανομή (με ίδιες παραμέτρους) με τις άλλες.
- Δηλαδή η κάθε μια από τις X_i έχει ίδια μέση τιμή και διακύμανση με τις άλλες. Με άλλα λόγια

$$\mathbb{E}X_1 = \mathbb{E}X_2 = \cdots = \mathbb{E}X_n = \mu \tag{3}$$

όπου μ είναι η μέση τιμή της κοινής τους κατανομής και μπορεί να είναι είτε παράμετρος της κατανομής (όπως στην περίπτωση της κανονικής) είτε να υπολογίζεται συναρτήσει των παραμέτρων της κατανομής όπως στην περίπτωση της ομοιόμορφης. Φυσικά, ισχύει και ότι

$$\mathbb{V}(X_1) = \mathbb{V}(X_2) = \cdots = \mathbb{V}(X_n) = \sigma^2 \tag{4}$$

- Οι (1), (3), (4) δεν ισχύουν αν οι $X_1, X_2, \ldots, X_n$ δεν είναι ισόνομες. Μη ισονομία σημαίνει ότι οι X_i προέρχονται από την ίδια κατανομή αλλά με διαφορετικές παραμέτρους: $X_i \sim f(x; \theta_i), i = 1, \ldots, n$ και $\theta_1 \neq \theta_2 \cdots \neq \theta_n$.

Βασικοί ορισμοί.

- Οι τιμές που παίρνει μια στατιστική συνάρτηση θα πρέπει να βρίσκονται εντός του παραμετρικού χώρου Θ στον οποίο κινείται η παράμετρος θ

Ορισμός 4. Εκτιμήτρια συνάρτηση.

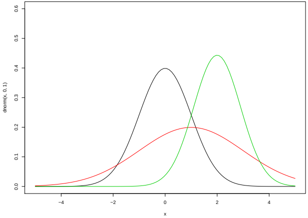
Εκτιμήτρια συνάρτηση της παραμέτρου θ (ή των παραμέτρων $\boldsymbol{\theta}$ ή μιας παραμετρικής συνάρτησης $g(\boldsymbol{\theta})$) θα καλείται μία σ.σ. $T = T(\mathbf{X})$ η οποία χρησιμοποιείται για την εκτίμηση του θ (ή των $\boldsymbol{\theta}$ ή του $g(\boldsymbol{\theta})$).

- Για ένα τυχαίο δείγμα $\mathbf{X}$ μπορούμε να κατασκευάσουμε πολλές εκτιμήτριες για μια παράμετρο θ (π.χ. οι $\bar{X}, X_1, (X_{(n)} - X_{(1)})/2$, θα μπορούσαν να προταθούν ως εκτιμήτριες του μέσου μ ενός κανονικού πληθυσμού).
- Το ερώτημα είναι ποιες θεωρούνται «καλές» ή ποια είναι η «καλύτερη» από όλες.
- Διαισθητικά, αναμένουμε ότι μια «καλή εκτιμήτρια» T του θ θα λαμβάνει τιμές
  1. «γύρω» από το θ και
  2. «κοντά» στο θ με «μεγάλη» πιθανότητα.

Ιδιότητες εκτιμητών.

- Για το 1 απαιτούμε η τυχαία μεταβλητή T να έχει μέση τιμή θ ή «σχεδόν» θ
- Για το 2 απαιτούμε η τυχαία μεταβλητή T να έχει «μικρή» διασπορά.

- Π.χ. η δίπλα εικόνα έχει την κατανομή των τιμών τριών εκτιμητών της μέσης τιμής ενός πληθυσμού.
- Η πραγματική μέση τιμή είναι 0. Ποιόν εκτιμητή θα επιλέγατε;



Πρακτική Άσκηση 1: Παράγετε 20 δείγματα 20 παρατηρήσεων έκαστο από την $N(2, 2^2)$ κατανομή.

1. εκτιμήστε την μέση τιμή κάθε δείγματος
2. κάντε τη γραφική παράσταση της κατανομής των τιμών
3. επαναλάβετε την άσκηση χρησιμοποιώντας ως εκτιμητή της μέσης τιμής την συνάρτηση $(X_{\text{min}} + X_{\text{max}})/2$
4. συγκρίνετε τις δύο κατανομές και διατυπώστε τα συμπεράσματα σας

Αμερόληπτες εκτιμήτριες.

Ορισμός 5: Αμερόληπτη εκτιμήτρια.

Μία εκτιμήτρια συνάρτηση T της $g(\boldsymbol{\theta})$ καλείται αμερόληπτη εκτιμήτρια (α.ε.) εάν $\mathbb{E}(T) = \mathbb{E}(T(\mathbf{X})) = g(\theta)$ για κάθε θ.

Η T θα λέγεται ασυμπτωτικά αμερόληπτη αν

$$\lim_{n\to+\infty} \mathbb{E}(T(\mathbf{X})) = g(\theta) \text{ για κάθε } \theta.$$

- Η διαφορά $\text{bias}(T) = \mathbb{E}(T) - g(\theta)$ καλείται μεροληψία της εκτιμήτριας T. Η μεροληψία μιας α.ε. είναι 0.

Παράδειγμα 1: Έστω δείγμα $X_1, \ldots, X_n$ από την ομοιόμορφη κατανομή $U(0, \theta)$, $\theta \in (0, +\infty)$. Δείξτε ότι μια αμερόληπτη εκτιμήτρια του θ είναι η

$$T(\mathbf{X}) = 2n^{-1} \sum_{i=1}^{n} X_i.$$

Απάντηση: Έχουμε

$$\mathbb{E}T(X) = 2n^{-1}\mathbb{E}\sum_{i=1}^{n} X_i = 2n^{-1}\sum_{i=1}^{n}\mathbb{E}X_i \stackrel{(3)}{=} 2\mathbb{E}X_1 \stackrel{(*)}{=} 2(\theta+0)/2 = \theta$$

$(*)$ για $X \sim U(a, b)$, $\mathbb{E}X = (a+b)/2$.

Παράδειγμα.

Παράδειγμα 2: Εάν $\mathbf{X} = (X_1, X_2)$ είναι ένα τυχαίο δείγμα από μια κατανομή με διασπορά σ^2, να δειχθεί ότι η στατιστική συνάρτηση $(X_1 - X_2)^2/2$ είναι αμερόληπτος εκτιμητής του σ^2.
Απάντηση: Έχουμε

$$\mathbb{E}T(X) = \frac{1}{2}\mathbb{E}(X_1 - X_2)^2 = \frac{1}{2}\mathbb{E}(X_1^2 + X_2^2 - 2X_1X_2) \overset{(*)}{=} \frac{1}{2}\mathbb{E}X_1^2 + \frac{1}{2}\mathbb{E}X_2^2 - (\mathbb{E}X_1)^2 \overset{(**)}{=} \mathbb{E}X_1^2 - (\mathbb{E}X_1)^2 = \mathbb{V}(X_1) = \sigma^2.$$

Ορισμός διακύμ.

$(*)$ Οι X_1, X_2 είναι ανεξάρτητες άρα $\mathbb{E}(X_1X_2) = (\mathbb{E}X_1)(\mathbb{E}X_2) = (\mathbb{E}X_1)^2$ και $(**)$ $\mathbb{E}X_1^2 = \mathbb{E}X_2^2$ (γιατί;).

- Αν έχουμε ένα τυχαίο δείγμα $X_1, X_2, \ldots, X_n \sim F(x;\theta)$ με μέση τιμή μ και διασπορά σ^2, δηλ. σε αυτή την περίπτωση $\boldsymbol{\theta} = (\mu, \sigma^2)$ τότε αποδεικνύονται οι ακόλουθες προτάσεις:

Πρόταση 1.

Ο δειγματικός μέσος $\bar{X}$ είναι α.ε. της μέσης τιμής μ με διασπορά $\mathbb{V}(\bar{X}) = n^{-1}\sigma^2$.

Αμερόληπτες εκτιμήτριες.

Απόδειξη:

- Για την αμεροληψία, η βασική εξίσωση της απόδειξης είναι:

$$\mathbb{E}(\bar{X}) = \mathbb{E}\frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}X_i = \mathbb{E}X_1 = \mu \tag{5}$$

- Ερώτηση: Τι πρέπει να υποθέσουμε για το δείγμα: $X_1, X_2, \ldots, X_n$ για να ισχύει η απόδειξη;
- Βοήθεια: Χρησιμοποιούμε πουθενά στην (5) την ανεξαρτησία του δείγματος; Την ισονομία; Είναι απαραίτητο να υποθέσουμε είτε ότι $X_i \sim F(x; \mu, \sigma^2)$ ή $X_i \sim F(x; \theta), i = 1, \ldots, n$ ή μήπως δεν χρειαζόμαστε τίποτα από τα δύο;
- Για την διακύμανση,

$$\mathbb{V}(\bar{X}) = \mathbb{V}\left\{\frac{1}{n}\sum_{i=1}^{n} X_i\right\} \overset{\text{γιατί;}}{=} \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}(X_i) \overset{\text{γιατί;}}{=} n^{-1}\sigma^2$$

Παράδειγμα.

Παράδειγμα 3: Έστω δείγμα $X_1, \ldots, X_n$ από την ομοιόμορφη κατανομή $U(0,\theta)$, $\theta \in (0,+\infty)$. Να βρεθεί αμερόληπτος εκτιμητής του θ^2.
Απάντηση: Με βάση την παραπάνω πρόταση, και βάση του ότι η διακύμανση της ομοιόμορφης κατανομής είναι $\theta^2/12$ περιμένουμε ο ζητούμενος εκτιμητής να είναι κάποια συνάρτηση του $\hat{\sigma}^2$. Πράγματι, για $\mathbf{X} \sim U(0,\theta)$

$$\mathbb{E}\hat{\sigma}^2 = \mathbb{E}\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n}\mathbb{E}\left(\sum_{i=1}^{n} X_i^2 - 2\frac{n}{n}\bar{X}\sum_{i=1}^{n} X_i + n\bar{X}^2\right) =$$

$$\frac{1}{n}\mathbb{E}\left(\sum_{i=1}^{n} X_i^2 - 2n\bar{X}^2 + n\bar{X}^2\right) = \frac{1}{n}\left(\sum_{i=1}^{n}\mathbb{E}X_i^2 - n\mathbb{E}\bar{X}^2\right) = \mathbb{E}X_1^2 - \mathbb{E}\bar{X}^2. \quad (6)$$

Εφόσον το δείγμα είναι από ομοιόμορφη κατανομή, έχουμε $\mathbb{E}X_1 = \theta/2$ και $\mathbb{V}(X_1) = \frac{\theta^2}{12}$ οπότε $\mathbb{V}(X_1) = \mathbb{E}X_1^2 - (\mathbb{E}X_1)^2 \Leftrightarrow \mathbb{E}X_1^2 = \frac{\theta^2}{3}$. Ομοίως, $\mathbb{V}(\bar{X}) = \mathbb{E}\bar{X}^2 - (\mathbb{E}\bar{X})^2$ οπότε $\mathbb{E}\bar{X}^2 = \frac{\theta^2}{12n} + \frac{\theta^2}{4} = \frac{\theta^2(1+3n)}{12n}$. Η (6) δίνει,

$$\mathbb{E}\hat{\sigma}^2 = \frac{\theta^2}{3} - \frac{\theta^2(1+3n)}{12n} = \frac{4n\theta^2 - \theta^2 - 3n\theta^2}{12n} = \frac{\theta^2(n-1)}{12n}$$

Οπότε, αμέσως προκύπτει ότι $\mathbb{E}(12n(n-1)^{-1}\hat{\sigma}^2) = \theta^2$ που είναι ο ζητούμενος α.ε.

Αμερόληπτες εκτιμήτριες.

Πρόταση 2.

Η δειγματική διασπορά S^2 είναι α.ε. της διασποράς σ^2.

$$\mathbb{E}(S^2) = \mathbb{E}\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\mathbb{E}\left(\sum_{i=1}^{n} X_i^2 - 2\frac{n}{n}\bar{X}\sum_{i=1}^{n} X_i + n\bar{X}^2\right) =$$

$$\frac{1}{n-1}\mathbb{E}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right) = \frac{1}{n-1}\left(\sum_{i=1}^{n}\mathbb{E}X_i^2 - n\mathbb{E}\bar{X}^2\right) = \frac{n}{n-1}(\mathbb{E}X_1^2 - \mathbb{E}\bar{X}^2),$$

γιατί εφόσον $X_1, X_2, \ldots, X_n \sim F(x;\theta)$ και ισόνομες, κάθε X_i έχει μέση τιμή μ διακύμανση σ^2. Επειδή $\mathbb{V}(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2$ έχουμε ότι $\mathbb{E}X_1^2 = \mathbb{V}(X_1) + (\mathbb{E}X_1)^2 = \sigma^2 + \mu^2, \mathbb{E}\bar{X}^2 = \mathbb{V}(\bar{X}) + (\mathbb{E}\bar{X})^2 = n^{-1}\sigma^2 + \mu^2$. Άρα τελικά,

$$\mathbb{E}(S^2) = \frac{n}{n-1}(\sigma^2 + \mu^2 - n^{-1}\sigma^2 - \mu^2) = \frac{n}{n-1}(1 - n^{-1})\sigma^2 = \sigma^2.$$

- Αν δεν χρησιμοποιούμε αμερόληπτες ή σχεδόν αμερόληπτες εκτιμήτριες θα έχουμε υποεκτίμηση ή υπερεκτίμηση του $g(\boldsymbol{\theta})$.
- Μπορεί να υπάρχουν πολλές αμερόληπτες εκτιμήτριες για μία παράμετρο θ, όλες με μέση τιμή θ.

Συνέπεια.

- Μια ιδιότητα που διαισθητικά αναμένουμε να έχει μια «καλή» εκτιμήτρια είναι αυτή της συνέπειας.
- Μια συνεπής εκτιμήτρια βελτιώνεται με την αύξηση του μεγέθους του δείγματος, και για πολύ μεγάλο δείγμα γίνεται πρακτικά ίση με την υπό εκτίμηση ποσότητα:

Ορισμός 10: Συνέπεια.

Μία εκτιμήτρια $T_n = T(X_1, X_2, \ldots, X_n)$ μιας παραμετρικής συνάρτησης $g(\boldsymbol{\theta})$ καλείται συνεπής αν ισχύει

$$\lim_{n \to +\infty} P((T_n - g(\boldsymbol{\theta})) < \varepsilon) = 1 \text{ για κάθε } \varepsilon > 0$$

δηλαδή με άλλα λόγια η T_n συγκλίνει κατά πιθανότητα στην $g(\boldsymbol{\theta})$ ($T_n \to g(\boldsymbol{\theta})$ καθώς $n \to +\infty$.)

- Η ποσότητα $|T_n - g(\boldsymbol{\theta})|$ ονομάζεται σφάλμα του εκτιμητή T.

Συνέπεια.

Ένα απλό κριτήριο για τη συνέπεια μιας εκτιμήτριας δίνεται στην ακόλουθη πρόταση.

Πρόταση 3.

Μία εκτιμήτρια $T_n = T(X_1, X_2, \ldots, X_n)$ μιας παραμετρικής συνάρτησης $g(\boldsymbol{\theta})$ είναι συνεπής αν ισχύει

$$\lim_{n\to+\infty} \mathbb{E}(T_n) = g(\boldsymbol{\theta}), \quad \lim_{n\to+\infty} \mathbb{V}(T_n) = 0.$$

Παράδειγμα: Ο δειγματικός μέσος $\bar{X}$ ενός τυχαίου δείγματος $X_1, X_2, \ldots, X_n \sim F(x;\mu)$, είναι συνεπής εκτιμήτρια του μέσου μ.
Απάντηση: Πράγματι, από την Πρόταση 1 έχουμε

$$\lim_{n\to+\infty} \mathbb{E}(\bar{X}) = \lim_{n\to+\infty} \mathbb{E}X_1 = \mu$$
$$\lim_{n\to+\infty} \mathbb{V}(\bar{X}) = \lim_{n\to+\infty} n^{-1}\sigma^2 = 0,$$

άρα ικανοποιούνται οι συνθήκες του ορισμού της συνέπειας.

Μέσο τετραγωνικό σφάλμα.

- Είναι πιο σύνηθες να χρησιμοποιείται το τετράγωνο του σφάλματος γιατί έχει καλύτερες ιδιότητες.

Ορισμός 11: Μέσο τετραγωνικό σφάλμα.

Αν T είναι μια εκτιμήτρια του $g(\boldsymbol{\theta})$, η ποσότητα

$$\mathrm{MSE}(T) = \mathbb{E}(T - g(\boldsymbol{\theta}))^2 = \mathbb{V}(T) + \mathrm{bias}^2(T)$$

καλείται μέσο τετραγωνικό σφάλμα της T από την $g(\boldsymbol{\theta})$.

Ορισμός 12.

Αν T_1, T_2 είναι δύο εκτιμήτριες της $g(\boldsymbol{\theta})$, η T_1 θα καλείται αποτελεσματικότερη της T_2 εάν ισχύει ότι $\mathrm{MSE}(T_1) < \mathrm{MSE}(T_2)$.

- Από ένα σύνολο εκτιμητριών του $g(\boldsymbol{\theta})$, καλύτερη θεωρείται αυτή που έχει το μικρότερο μέσο τετραγωνικό σφάλμα.
- Το ζητούμενο είναι να βρεθεί μια εκτιμήτρια συνάρτηση T, η οποία θα ελαχιστοποιεί το μέσο τετραγωνικό σφάλμα.

Μέσο τετραγωνικό σφάλμα.

Παράδειγμα: Να βρεθεί το μέσο τετραγωνικό σφάλμα του εκτιμητή $T(\mathbf{X}) = 12n(n-1)^{-1}\hat{\sigma}^2$ του παραδείγματος 3.
Απάντηση:

T αμερόληπτος, άρα;

$$\mathrm{MSE}\{T(\mathbf{X})\} = \mathrm{bias}^2(T(\mathbf{X})) + \mathbb{V}(T(\mathbf{X})) = \mathbb{V}(12n(n-1)^{-1}\hat{\sigma}^2) =$$

ορισμός διακύμανσης

$$\mathbb{V}\left(\frac{12n}{(n-1)^2}\sum_{i=1}^{n}(X_i-\bar{X})^2\right) \overset{(1)}{=} \frac{144n^2}{(n-1)^4}\sum_{i=1}^{n}\mathbb{V}(X_i-\bar{X}) =$$

$$= \frac{144n^2}{(n-1)^4}\sum_{i=1}^{n}\left\{\mathbb{E}(X_i-\bar{X})^2 - \left[\mathbb{E}(X_i-\bar{X})\right]^2\right\}$$

γιατί κάνει 0;

$$= \frac{144n^3}{(n-1)^4}\left\{\mathbb{E}(X_1^2) - \mathbb{E}\bar{X}^2\right\} - \frac{144n^3}{(n-1)^4}\sum_{i=1}^{n}\left[\mathbb{E}(X_i\text{-}\bar{X})\right]^2$$

$$= \frac{144n^3}{(n-1)^4}\frac{\theta^2}{12} = \frac{12n^3}{(n-1)^4}\theta^2.$$

Η μέθοδος της μέγιστης πιθανοφάνειας.

- Η πιο γνωστή μέθοδος εύρεσης μιας εκτιμήτριας για τις παραμέτρους $\boldsymbol{\theta}$ μιας κατανομής F είναι η μέθοδος μέγιστης πιθανοφάνειας.
- Η μέθοδος αυτή είναι αρκετά ισχυρή διότι, με μία σχετικά απλή διαδικασία, οδηγεί σε εκτιμήτριες με πολύ καλές ιδιότητες - βασικά παράγει την πιο πιθανή να συμβεί στην πράξη εκτίμηση.
- Θεωρούμε ένα τυχαίο δείγμα $X_1, X_2, \ldots, X_n$ από μία κατανομή με συνάρτηση σ.π. $f(x; \boldsymbol{\theta})$ που εξαρτάται από τις παραμέτρους $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$, και επιθυμούμε να εκτιμήσουμε το $\boldsymbol{\theta}$.

Ορισμός 12: Συνάρτηση πιθανοφάνειας.

Συνάρτηση πιθανοφάνειας ή πιθανοφάνεια του δείγματος $X_1, X_2, \ldots, X_n$ καλείται η από κοινού σ.π. των $X_1, X_2, \ldots, X_n$ θεωρούμενη ως συνάρτηση του $\boldsymbol{\theta}$, δηλαδή η

$$L_{\mathbf{X}}(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(X_i; \boldsymbol{\theta})$$

Η μέθοδος της μέγιστης πιθανοφάνειας.

Ορισμός 13: εκτιμήτρια μέγιστης πιθανοφάνειας.

Μία στατιστική συνάρτηση $\hat{\boldsymbol{\theta}}$ καλείται εκτιμήτρια μέγιστης πιθανοφάνειας (ΕΜΠ) των παραμέτρων $\boldsymbol{\theta}$ αν μεγιστοποιεί την πιθανοφάνεια $L_{\mathbf{X}}(\boldsymbol{\theta})$, δηλαδή, αν ισχύει ότι

$$L_{\mathbf{X}}(\hat{\boldsymbol{\theta}}) = \sup_{\boldsymbol{\theta} \in \Theta} L_{\mathbf{X}}(\boldsymbol{\theta})$$

- Ισοδύναμα, αντί να αναζητούμε το σημείο μεγίστου της $L_{\mathbf{X}}(\boldsymbol{\theta})$, είναι πιο εύκολο να αναζητούμε το σημείο μεγίστου της $l(\boldsymbol{\theta}) = \ln(L_{\mathbf{X}}(\boldsymbol{\theta}))$ (έχουν το ίδιο σημείο μεγίστου διότι η συνάρτηση ln είναι αύξουσα).
- Για $k = 1$ (δηλ. $\boldsymbol{\theta} = \theta$), η συνάρτηση $l(\boldsymbol{\theta})$ παραγωγίζεται σε ολόκληρο τον παραμετρικό χώρο Θ οπότε μπορούμε να βρούμε το σημείο μεγίστου μέσα από τη λύση της εξίσωσης $l'(\boldsymbol{\theta}) = 0$, ελέγχοντας παράλληλα ότι η δεύτερη παράγωγος $l''(\boldsymbol{\theta}) < 0$.
- Για $k \geq 2$, δηλ. όταν $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ έχουμε ένα πρόβλημα βελτιστοποίησης στο $\mathbb{R}^k$.

Η μέθοδος της μέγιστης πιθανοφάνειας.

- Δεδομένου ότι η $l(\boldsymbol{\theta})$ είναι παραγωγίσιμη στο $\Theta \subseteq \mathbb{R}^k$, σχηματίζουμε τις εξισώσεις

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \theta_1} = 0, \ldots, \frac{\partial l(\boldsymbol{\theta})}{\partial \theta_k} = 0$$

- η λύση του πιο πάνω συστήματος δίνει τον εκτιμητή μέγιστης πιθανόφάνειας $\hat{\boldsymbol{\theta}}$ της $\boldsymbol{\theta}$, υπό τον όρο ότι αυτός ο εκτιμητής ικανοποιεί και την

$$\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \theta_1^2} < 0, \ldots, \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \theta_k^2} < 0.$$

- Είναι πιθανό στη λύση ενός τέτοιου προβλήματος να μην υπάρχει πεπερασμένο μέγιστο, να υπάρχουν παραπάνω από ένα μέγιστα, η να έχουμε ακριβώς ένα μέγιστο.
- Αν $\hat{\boldsymbol{\theta}}$ είναι η ΕΜΠ εκτιμήτρια της $\boldsymbol{\theta}$ και g είναι αμφιμονοσήμαντη συνάρτηση, τότε η $g(\hat{\boldsymbol{\theta}})$ είναι η ΕΜΠ εκτιμήτρια της $g(\boldsymbol{\theta})$.
- Δείτε και τα παραδείγματα 1.11–1.14 από το βιβλίο 'Μαθηματική Στατιστική' στο moodle για εφαρμογές.

Εκτίμηση με τη μέθοδο της μέγιστης πιθανοφάνειας.

Άσκηση 1: Έστω τυχαίο δείγμα $X_1, X_2, \ldots, X_n$ από την εκθετική κατανομή με σ.π.π. $f(x;\theta) = \theta^{-1}\exp\{x\theta^{-1}\}, 0 \leq x \leq +\infty, \theta \in (0, +\infty)$ Να υπολογισθεί ο Ε.Μ.Π. για την παράμετρο θ.

Λύση: Η συνάρτηση πιθανοφάνειας του δείγματος είναι:

$$L(\theta) = \prod_{i=1}^{n} f(X_i;\theta) = \prod_{i=1}^{n} \frac{1}{\theta} e^{-\frac{X_i}{\theta}} = \frac{1}{\theta^n} e^{-\frac{1}{\theta}\sum_{i=1}^{n} X_i}$$

Ο λογάριθμος της συνάρτησης πιθανοφάνειας είναι:

$$\ln L(\theta) = -n\ln\theta - \frac{1}{\theta}\sum_{i=1}^{n} X_i$$

Παραγωγίζοντας και εξισώνοντας με το 0 παίρνουμε

$$\frac{\partial \ln L(\theta)}{\partial\theta} = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} X_i = 0 \Rightarrow \theta = \bar{X}.$$

Για να επιβεβαιώσουμε ότι έχουμε πράγματι μέγιστο,

$$\frac{\partial^2 \ln L(\theta)}{\partial\theta^2} = \frac{n}{\theta^2} - \frac{2}{\theta^3}\sum_{i=1}^{n} X_i \overset{\theta=\bar{X}}{=} \frac{n}{\bar{X}^2} - \frac{2n\bar{X}}{\bar{X}^2} < 0.$$

Εκτίμηση με τη μέθοδο της μέγιστης πιθανοφάνειας.

Άσκηση 2: Έστω τυχαίο δείγμα $X_1, X_2, \ldots, X_n$ από την κανονική κατανομή με σ.π.π.

$$f(x;\theta_1,\theta_2) = \frac{1}{\sqrt{\theta_2}}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(x-\theta_1)^2}{2\theta_2}\right\}, -\infty \le \theta_1 \le +\infty, \theta_2 \in (0,+\infty)$$

Να υπολογισθεί ο Ε.Μ.Π. για τις παραμέτρους $\boldsymbol{\theta} = (\theta_1, \theta_2)$.

Λύση: Η συνάρτηση πιθανοφάνειας του δείγματος είναι:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(X_i;\boldsymbol{\theta}) = \prod_{i=1}^{n}\frac{1}{\sqrt{\theta_2}}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(X_i-\theta_1)^2}{2\theta_2}\right\}$$

$$= \frac{1}{\sqrt{\theta_2^n}}\frac{1}{\sqrt{(2\pi)^n}}\exp\left\{-\frac{1}{2\theta_2}\sum_{i=1}^{n}(X_i-\theta_1)^2\right\}.$$

Ο λογάριθμος της συνάρτησης πιθανοφάνειας είναι:

$$\ln L(\boldsymbol{\theta}) = -\frac{n}{2}\ln\theta_2 - \frac{n}{2}\ln(2\pi) - \frac{1}{2\theta_2}\sum_{i=1}^{n}(X_i-\theta_1)^2.$$

Εκτίμηση με τη μέθοδο της μέγιστης πιθανοφάνειας.

Άσκηση 2 συνέχεια: Παραγωγίζοντας και εξισώνοντας με το 0 παίρνουμε

$$\frac{\partial \ln L(\boldsymbol{\theta})}{\partial \theta_1} = -\frac{2(-1)\sum_{i=1}^{n}(X_i - \theta_1)}{\theta_2} = 0 \Rightarrow \sum_{i=1}^{n}(X_i - \theta_1) = 0 \Rightarrow \hat{\theta}_1 = \bar{X}$$

Επίσης,

$$\frac{\partial \ln L(\boldsymbol{\theta})}{\partial \theta_2} = -\frac{n}{2\theta_2} + \frac{\sum_{i=1}^{n}(X_i - \theta_1)^2}{2\theta_2^2} = 0 \Rightarrow -n\theta_2 + \sum (X_i - \theta_1)^2 = 0$$

από όπου παίρνουμε

$$\hat{\theta}_2 = \hat{\sigma}^2 = n^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Άσκηση: Επιβεβαιώστε ότι οι δεύτερες μερικές παράγωγοι ως προς θ_1 και θ_2 του λογάριθμου της συνάρτησης πιθανοφάνειας είναι αρνητικές οπότε όντως έχουμε βρει τα μέγιστα.

Εκτίμηση με τη μέθοδο της μέγιστης πιθανοφάνειας.

Άσκηση 3:

- Έχουμε παρατηρήσεις $X_1, X_2, \ldots, X_n$ οι οποίες αφορούν ώρες επιβίωσης συγκεκριμένων πλοίων σε συνθήκες πίεσης (τα δεδομένα δίνονται στο αρχείο data.csv στο moodle).
- Ένα συγκεκριμένο μοντέλο που χρησιμοποιείται για ανάλυση δεδομένων επιβίωσης είναι η κατανομή Weibull.

$$f(y;\lambda;\theta) = \frac{\lambda y^{\lambda-1}}{\theta^{\lambda}} e^{-(\frac{y}{\theta})^{\lambda}}, y > 0,$$

όπου λ παράμετρος που καθορίζει το σχήμα της κατανομής, και θ παράμετρος που καθορίζει την κλίμακα.

- Για το συγκεκριμένο παράδειγμα ξέρουμε ότι $\lambda = 2$, δηλ. $X_1, X_2, \ldots, X_n \sim \text{Weibull}(2, \theta)$.
- Η συνάρτηση πιθανοφάνειας είναι η

$$f(X_1, \ldots, X_n; \theta) = \prod_{i=1}^{n} \frac{\lambda X_i^{\lambda-1}}{\theta^{\lambda}} e^{-\left(\frac{X_i}{\theta}\right)^{\lambda}}. \tag{7}$$

Εκτίμηση με τη μέθοδο της μέγιστης πιθανοφάνειας.

Ισοδύναμα μπορούμε να μεγιστοποιήσουμε το λογάριθμο της (7). Έχουμε

$$L(\theta) = \log f(X_1, \ldots, X_n; \theta) = \sum_{i=1}^{n} \left\{ (\lambda - 1) \log X_i + \log \lambda - \lambda \log \theta - \left(\frac{X_i}{\theta} \right)^{\lambda} \right\}.$$

Παραγωγίζοντας,

$$\frac{d\,L(\theta)}{d\theta} = 0 \Rightarrow \sum_{i=1}^{n} \left\{ -\frac{\lambda}{\theta} + \frac{\lambda X_i^{\lambda}}{\theta^{\lambda+1}} \right\} = -\frac{\lambda n}{\theta} + \frac{\lambda \sum_{i=1}^{n} X_i^{\lambda}}{\theta^{\lambda+1}} = 0$$

$$\overset{\lambda=2}{\Rightarrow} -\frac{2n}{\theta} + \frac{2\sum_{i=1}^{n} X_i^2}{\theta^3} = 0 \Rightarrow \frac{n}{\theta} = \frac{\sum_{i=1}^{n} X_i^2}{\theta^3} \Rightarrow$$

$$\theta^2 = n^{-1} \sum_{i=1}^{n} X_i^2 \Rightarrow \theta = \sqrt{n^{-1} \sum_{i=1}^{n} X_i^2}.$$

Εκτίμηση με τη μέθοδο της μέγιστης πιθανοφάνειας.

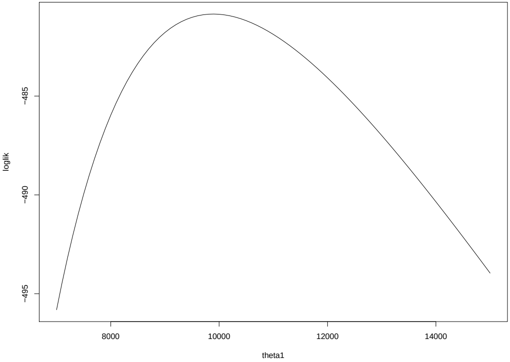
Για να υλοποιήσουμε τη διαδικασία εκτίμησης στην R:

```
lifetimes<-as.matrix(read.csv("data.csv", header = FALSE, sep = ""))
thetause<-mean(lifetimes) #MLE estimate of theta

loglikelihood <- function(data, theta, lambda=2) #define the log-
     likelihood function
  {
   logl <- sum((lambda-1)*log(data)+log(lambda)-lambda*log(theta)-(data/
     theta)^lambda)
   return(logl)
  }

theta1 <- seq(7000, 15000, by=100) # define a range of parameters
loglik <- rep(NA, length(theta1)) # define a vector to store results

for (i in 1:length(theta1)) #calculate log-likelihood for each theta
  {
   loglik[i] <- loglikelihood(lifetimes, theta1[i])
}

plot(theta1, loglik, type="l") # plot the results
```

Ερώτηση: Ποιος είναι ο εκτιμητής μέγιστης πιθανοφάνειας της θ; (τι νούμερο υπολογίσαμε στην R; - αντιστοιχεί στο μέγιστο που φαίνεται στο γράφημα της επόμενης σελίδας;

Εκτίμηση με τη μέθοδο της μέγιστης πιθανοφάνειας.

Το γράφημα της παρουσιάζεται στο πιο κάτω σχήμα και παρατηρούμε ότι υπάρχει τιμή της θ η οποία μεγιστοποιεί την $L(\theta)$.



Η διαδικασία που δείχνουμε εδώ είναι χρήσιμη όταν δεν μπορούμε να βρούμε μέγιστο με αναλυτικό τρόπο οπότε χρησιμοποιούμε αριθμητικές μεθόδους για μεγιστοποίηση της συνάρτησης πιθανοφάνειας.

Εκτίμηση με τη μέθοδο των ροπών.

- Η μέθοδος των ροπών, όπως και η μέθοδος μέγιστης πιθανοφάνειας αποσκοπεί στον προσδιορισμό άγνωστων παραμέτρων μιας κατανομής.
- εφαρμόζεται σε κατανομές που υπάρχουν ροπές k-τάξης και περιγράφεται ως εξής: Έστω τ.δ. $X_1, X_2, \ldots, X_n \sim f(x; \boldsymbol{\theta})$ με θεωρητικές ροπές

$$\mu_1 = \mathbb{E}X,\ \mu_2 = \mathbb{E}X^2, \ldots, \mu_k = \mathbb{E}X^k.$$

- Προσέξτε την αντιστοιχία με τις παραμέτρους μιας κατανομής: $\mu = \mu_1 = \mathbb{E}X$, $\sigma^2 = \mathbb{E}X^2 - \mu^2 = \mu_2 - \mu^2$, κ.λ.π.
- Υπενθυμίζεται ότι οι δειγματικές ροπές αντίστοιχα είναι οι

$$m_1 = n^{-1}\sum_{i=1}^{n} X_i,\ m_2 = n^{-1}\sum_{i=1}^{n} X_i^2, \ldots, m_k = n^{-1}\sum_{i=1}^{n} X_i^k.$$

Θεώρημα

Οι δειγματικές ροπές είναι αμερόληπτοι εκτιμητές των θεωρητικών ροπών.

Εκτίμηση με τη μέθοδο των ροπών.

- Το παραπάνω θεώρημα δείχνει τον τρόπο εκτίμησης με τη μέθοδο των ροπών: εξισώνονται οι θεωρητικές ροπές με τις δειγματικές ροπές ίσης τάξης.
- Με τον τρόπο αυτό συνδέονται οι εκτιμώμενες παράμετροι με στατιστικές συναρτήσεις και από τη λύση των εξισώσεων που προκύπτουν, υπολογίζονται οι εκτιμητές.

Παράδειγμα 1: Να εκτιμηθεί η παράμετρος θ της σ.π.π. $f(x;\theta) = \theta x^{\theta-1}, 0 < x < 1, 0 < \theta < +\infty$ δοθέντος δείγματος μεγέθους n από την κατανομή. Λύση: Έχουμε να εκτιμήσουμε μόνο μία παράμετρο οπότε χρειαζόμαστε μόνο μία εξίσωση. Η ροπή πρώτης τάξης είναι

$$\mu_1 = \mathbb{E}X = \int_0^1 x\theta x^{\theta-1}\, dx = \frac{\theta}{\theta+1}.$$

Η δειγματική ροπή πρώτης τάξης είναι η $m_1 = \bar{X}$ οπότε εξισώνοντας τις μ_1 και m_1 έχουμε:

$$\bar{X} = \frac{\theta}{\theta+1} \Leftrightarrow \hat{\theta} = \frac{\bar{X}}{1-\bar{X}}.$$

Εκτίμηση με τη μέθοδο των ροπών.

Παράδειγμα 2: Να εκτιμηθούν οι παράμετροι μ, σ^2 της $N(x; \mu, \sigma^2)$, δοθέντος δείγματος μεγέθους n από την κατανομή.

Λύση: Εδώ τώρα έχουμε να εκτιμήσουμε δύο παραμέτρους οπότε χρειαζόμαστε δυο εξισώσεις. Έχουμε

$$m_1 = \mu_1 \Rightarrow \bar{X} = \mu, \; m_2 = \mu_2 \Rightarrow \mathbb{E}X^2 = n^{-1} \sum_{i=1}^{n} X_i^2.$$

Χρησιμοποιώντας ότι $\mathbb{E}X = \mu, \mathbb{E}X^2 = \sigma^2 + \mu^2$, παίρνουμε:

$$\mu = n^{-1} \sum_{i=1}^{n} X_i, \; \sigma^2 + \mu^2 = n^{-1} \sum_{i=1}^{n} X_i^2$$

οπότε καταλήγουμε στους εκτιμητές

$$\hat{\mu} = n^{-1} \sum_{i=1}^{n} X_i, \; \hat{\sigma}^2 = n^{-1} \sum_{i=1}^{n} X_i^2 - \mu^2 = n^{-1} \sum_{i=1}^{n} X_i^2 - \bar{X}^2.$$

Ερώτηση: Τι σχέση έχουν με τους εκτιμητές μέγιστης πιθανοφάνειας που υπολογίσαμε πιο πριν;