# Εφαρμοσμένη Στατιστική

Δημήτριος Μπάγκαβος

Τμήμα Μαθηματικών και Εφαρμοσμένων Μαθηματικών
Πανεπιστήμιο Κρήτης

Φεβρουάριος, 2018

# Περιεχόμενα

# Τυχαίες μεταβλητές.

- Δειγματικός χώρος $\Omega$: το σύνολο των δυνατών αποτελεσμάτων ενός πειράματος τύχης. Τα δείγματα δεδομένων που αναλύουμε παίρνουν τιμές στον δειγματικό χώρο.
- Παράδειγμα: στη ρίψη ενός ζαριού, $\Omega = \{1, 2, 3, 4, 5, 6\}$ (δηλαδή κάθε σημείο του δειγματικού χώρου είναι πιθανό αποτέλεσμα της ρίψης).
- Στη ρίψη ενός νομίσματος 3 φορές ο δ.χ. είναι $\Omega = \{$ΚΚΚ, ΚΚΓ, ΚΓΚ, ΓΚΚ, ΓΚΓ, ΓΓΚ, ΚΓΓ, ΓΓΓ$\}$.
- Αν με ενδιαφέρει πόσες φορές εμφανίσθηκε $K$ τότε ο δ.χ. είναι $\Omega = \{3, 2, 1, 0\}$, ανάλογα με το πόσες φορές εμφανίζεται ο $K$. Κατά τη μελέτη ενός πειράματος τύχης μπορούμε να αντιστοιχίσουμε σε κάθε δειγματικό σημείο έναν αριθμό χρησιμοποιώντας έναν προκαθορισμένο κανόνα αντιστοίχησης.
- Παράδειγμα: Ρίψη νομίσματος μέχρι να έρθει γράμματα για πρώτη φορά.
- Στην περίπτωση αυτή ο δ.χ. είναι $\Omega = \{$Γ, ΚΓ,ΚΚΓ,ΚΚΚΓ,ΚΚ $\ldots$ ΚΓ$\}$.
- Έστω ότι μας ενδιαφέρει ο αριθμός των ρίψεων μέχρι να έρθει γράμματα. Ο νέος δ.χ. θα είναι τότε $\Omega = \{1, 2, 3, 4, \ldots\}$.

# Τυχαίες μεταβλητές.

- Δηλαδή έχουμε μία αντιστοιχία της μορφής

$$\begin{array}{ccccccc} \Gamma & K\Gamma & KK\Gamma & KKK\Gamma & KKKK\Gamma & \dots & KKKK\dots\Gamma \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & & \downarrow \\ 1 & 2 & 3 & 4 & 5 & & \dots \end{array}$$

Από τον πολ/στικό κανόνα και την ανεξαρτησία,

$$\begin{aligned} P(X=1) &= P(\Gamma) = p \\ P(X=2) &= P(K\Gamma) = p(1-p) \\ &\dots \\ P(X=n) &= P(K\dots K\Gamma) = (1-p)^{n-1}p. \end{aligned}$$

## Ορισμός

Διακριτή τυχαία μεταβλητή $X$ σε ένα δ.χ. $\Omega$ είναι μια συνάρτηση $X$ με πεδίο ορισμού το $\Omega$ και πεδίο τιμών ένα υποσύνολό του που αποτελείται από ακεραίους $(x_1, x_2, \dots, x_n)$. Η συνάρτηση $X$ έχει την ιδιότητα πως κάθε $x_i, i = 1, \dots, n$ είναι ενδεχόμενο.

- Έστω η τυχαία μεταβλητή $X$ =χρόνος ζωής ενός ηλεκτρικού λαμπτήρα. Το πεδίο τιμών της $X$ είναι το σύνολο $[0, +\infty)$

# Τυχαίες μεταβλητές: Παραδείγματα.

- Έστω το τυχαίο πείραμα της ρίψης δύο ζαριών.
  1. Ο δειγματικός χώρος είναι $\Omega = \{(i,j) | i,j = 1,\dots,6\}$.
  2. Ορίζουμε την τυχαία μεταβλητή $X : \Omega \to R$ τέτοια ώστε $X(i,j) = i + j$, δηλ η $X$ είναι το άθροισμα των ενδείξεων των δύο ζαριών.
- Έστω το τυχαίο πείραμα της ρίψης ενός νομίσματος $n$ φορές. Ορίζουμε την τυχαία μεταβλητή $X =$ αριθμός των εμφανιζομένων κεφαλών στις $n$ ρίψεις.
  1. Ο δειγματικός χώρος $\Omega$ του πειράματος αποτελείται από $n$–άδες με ενδείξεις Κ για τις κεφαλές και Γ για τα γράμματα.
  2. Αν ένα στοιχείο $\omega$ του δειγματικού χώρου είναι μία $n$–άδα με πέντε ενδείξεις κεφαλής τότε $X(\omega) = 5$.
- Η έννοια της διακριτής τ.μ. παρέχει έναν εύκολο τρόπο να περιγράψουμε ένα τυχαίο πείραμα με πεπερασμένο πλήθος αποτελεσμάτων.
- Απλά ορίζουμε μια μεταβλητή που παίρνει τιμές $x_1, x_2, \dots, x_n$ ώστε να ισχύει $X = i$ αν και μόνο αν το πείραμα δίνει το $i$ αποτέλεσμα.
- αν τραβήξουμε ένα χαρτί (από $n$ χαρτιά) τότε θέτουμε $X = i$ αν επιλεγεί το $i$ χαρτί και $P(X = i) = n^{-1}$.

# Τυχαίες μεταβλητές: Παραδείγματα.

- Σε αυτήν την περίπτωση λέμε ότι έχουμε μια τ.μ. $X$ με ακέραιες τιμές $1, 2, \ldots, n$ και έχει συνάρτηση πυκνότητας $p(x) = n^{-1}$ για $x = 1, 2, \ldots, n$ και $p(x) = 0$ αλλιώς.
- Παράδειγμα: Ένας παίκτης σε κάθε ρίψη του νομίσματος κερδίζει 1 μονάδα αν έρθει γράμματα και χάνει 1 μονάδα αν έρθει κορώνα. Ποιο το κέρδος μετά τρεις ρίψεις και ποιες οι αντίστοιχες πιθανότητες;
- Λύση: $S = \{\mathrm{KKK}, \mathrm{KK\Gamma}, \mathrm{K\Gamma K}, \mathrm{\Gamma KK}, \mathrm{\Gamma K\Gamma}, \mathrm{\Gamma\Gamma K}, \mathrm{K\Gamma\Gamma}, \mathrm{\Gamma\Gamma\Gamma}\}$ είναι ο δ.χ. και με ενδιαφέρει το κέρδος, έστω $X$.
- Για το πρώτο αποτέλεσμα, το κέρδος θα είναι -3. Για το δεύτερο, το τρίτο και το τέταρτο, το κέρδος θα είναι -1.
- Για το πέμπτο, έκτο, έβδομο το κέρδος θα είναι 1. Για το τελευταίο, το κέρδος θα είναι 3. Άρα το κέρδος θα είναι τ.μ. γιατί ορίζει μία συνάρτηση με πεδίο ορισμού $S$ και τιμές $\{-3, -1, 1, 3\}$.
- Οι αντίστοιχες πιθανότητες θα είναι:

$$P(X = -3) = P(KKK) = \frac{1}{8}, \quad P(X = -1) = P(KK\Gamma \text{ ή } K\Gamma K \text{ ή } \Gamma KK) = \frac{3}{8}$$

$$P(X = 1) = P(K\Gamma\Gamma \text{ ή } \Gamma K\Gamma \text{ ή } \Gamma\Gamma K) = \frac{3}{8}, P(X = 3) = P(\Gamma\Gamma\Gamma) = \frac{1}{8}.$$

# Διακριτές τυχαίες μεταβλητές: Συνάρτηση πυκνότητας πιθανότητας και αθροιστική συνάρτηση κατανομής.

- Για μία διακριτή (τ.μ.) $X$ το πεδίο τιμών της είναι της μορφής $\{x_1, x_2, \ldots, x_n\}$, $x_i \in \mathbb{N}$.
- Η $p(x) = P(X = x)$ ονομάζεται συνάρτηση πυκνότητας πιθανότητας (σ.π.π.) της τ.μ. $X$ αν έχει τις κάτωθι ιδιότητες:
  1. $p(x_i) \geq 0, \forall\, i = 1, 2, \ldots, n$
  2. $$\sum_{i=1}^{n} p(x_i) = 1.$$
- Η αθροιστική συνάρτηση κατανομής (α.σ.κ.) της τ.μ. $X$ συμβολίζεται με $P_X(x)$ και δηλώνει την πιθανότητα η τ.μ. $X$ να πάρει τιμές μικρότερες ή ίσες από κάποια τιμή $x$. Για διακριτές τ.μ.:

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} p(x_i).$$

- Για διακριτές τ.μ. η πιθανότητα μεμονωμένου ενδεχομένου $X = x_i$ δίνεται από την σ.π.π. της $X$: $P(X = x_i) = p(x_i)$.
- Με άλλα λόγια το αποτέλεσμα $p(x_i)$ είναι πάντα πιθανότητα.

# Διακριτές τυχαίες μεταβλητές: Συνάρτηση πυκνότητας πιθανότητας και αθροιστική συνάρτηση κατανομής.

- Παράδειγμα: Ο αριθμός των αυτοκινήτων που πουλάει μία έκθεση σε μία εβδομάδα είναι τυχαία μεταβλητή $X$ με συνάρτηση πιθανότητας που δίνεται από τον τύπο: $p(x) = cx, x = 1, 2, 3, 4, 5$ και $p(x) = c(10 - x)$, $x = 6, 7, 8, 9$. (α) Να βρεθεί η τιμή της σταθεράς $c$. (β) Ποια είναι η πιθανότητα να πουληθούν σε μία εβδομάδα 1) λιγότερα από 4 αυτοκίνητα; 2) περισσότερα από 5 αυτοκίνητα γνωρίζοντας ότι έχουν πουληθεί τουλάχιστον 3;
- Απάντηση: Για το (α) έχουμε:

$$\sum_{x=1}^{9} p(x) = 1 \Leftrightarrow 25c = 1 \Leftrightarrow c = 1/25.$$

- Για το (β) έχουμε:

$$P(X < 4) = \sum_{x=1}^{3} P(X = x) = (1 + 2 + 3)/25,$$

$$P(X > 5|X \geq 3) = \frac{P(X > 5, X \geq 3)}{P(X \geq 3)} = \frac{P(X > 5)}{P(X \geq 3)} = \frac{p(6) + \cdots + p(9)}{1 - p(1) - p(2)}.$$

# Διακριτές τυχαίες μεταβλητές: Συνάρτηση πυκνότητας πιθανότητας και αθροιστική συνάρτηση κατανομής.

- Παράδειγμα: Η πυκνότητα μιας τ.μ. $X$ δίνεται από τον ακόλουθο πίνακα:

| $x$ | -3 | -1 | 0 | 1 | 2 | 3 | 5 | 8 |
|---|---|---|---|---|---|---|---|---|
| $p(x)$ | 0.1 | 0.2 | 0.15 | 0.2 | 0.1 | 0.15 | 0.05 | 0.05 |

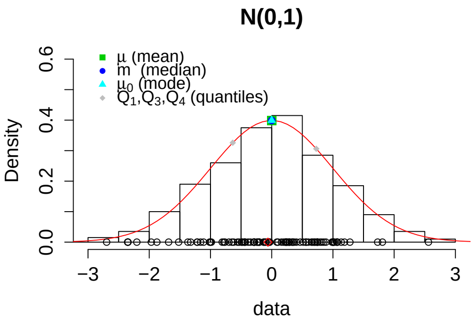
- Να υπολογίσετε τις πιθανότητες:
  - η $X$ να είναι αρνητική
  - η $X$ να είναι άρτια
  - η $X$ να έχει τιμή από 1 ως 8
  - $P(X = -3 | X \leq 0)$
  - $P(X \geq 3 | X \geq 0)$
- Παράδειγμα: Έχουμε ένα δοχείο με 6 κόκκινες και 4 μαύρες μπάλες.
- Επιλέγουμε $n$ μπάλες και έστω $X$ η μεταβλητή που συμβολίζει τον αριθμό των κόκκινων.
- Βρείτε την συνάρτηση πυκνότητας της $X$ όταν η επιλογή γίνεται χωρίς επανατοποθέτηση.

# Συνεχείς τυχαίες μεταβλητές: συνάρτηση πυκνότητας πιθανότητας και αθροιστική συνάρτηση κατανομής.

- Η τυχαία μεταβλητή (τ.μ.) $X$ ονομάζεται συνεχής αν υπάρχει $f : \mathbb{R} \to \mathbb{R}$ :
$$P(X \in B) = \int_B f(x)\,dx$$
- Για να είναι η $f(x)$ η σ.π.π. της $X$, θα πρέπει:
$$(i)\ f(x) \geq 0,\ \forall\ x \in \mathbb{R},\quad (ii)\ \int_{-\infty}^{+\infty} f(x)\,dx = 1.$$
- Για συνεχή τ.μ. η αθροιστική συνάρτηση κατανομής είναι:
$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(u)du.$$
- Η $f(x)$ δεν εκφράζει πιθανότητα, δηλαδή $f(x) \neq P(X = x)$. Για συνεχείς τ.μ., $P(X = x) = 0$ πάντα. Η $f(x)$ εκφράζει πυκνότητα δηλ. όσο μεγαλύτερη είναι η τιμή $f(x)$ τόσο περισσότερο πιθανό είναι να πάρει η μεταβλητή $X$ τιμές κοντά στο $x$. Για σταθερά $c$,
$$P(X = c) = \int_c^c f(x)\,dx = 0 \text{ (γιατί;)}.$$

# Σχέση περιγραφικών μέτρων και κατανομών

Με τον όρο δείγμα από την κατανομή εννοούμε τιμές στον άξονα $x$ (συμβολίζονται με $X_1, X_2, \ldots, X_{100}$) οι οποίες εκτείνονται περίπου από το -3 μέχρι το 3 (θεωρητικά από το $-\infty$ μέχρι το $+\infty$).



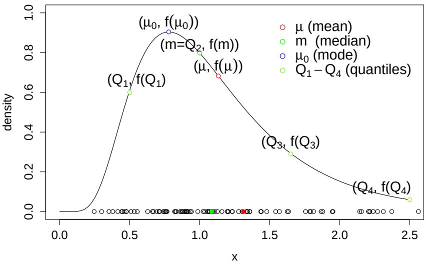

Η καμπύλη της κατανομής δείχνει τον τρόπο με τον οποίο κατανέμονται οι τιμές του δείγματος στον άξονα $x$: Σε αυτό το παράδειγμα οι πιο πολλές τιμές συγκεντρώνονται στη μέση. Όσο απομακρυνόμαστε από τη μέση, τόσο αραιώνουν.

Με τις τιμές του δείγματος μπορούμε να προσεγγίσουμε το σχήμα της κατανομής με το ιστόγραμμα.

Βλέπουμε ότι το ιστόγραμμα ακολουθεί την κατανομή των σημείων στον άξονα $x$.

# Παράμετροι κατανομών.

- Με τα περιγραφικά μέτρα (μέση τιμή, διακύμανση, κ.λ.π.), έχουμε μία ένδειξη για το που κυμαίνονται οι τιμές της μεταβλητής X, δεν έχουμε όμως πληροφορία για το πόσο πιθανές είναι αυτές οι τιμές (μόνο εμπειρικά).




- Το πιο πάνω σχήμα δείχνει ότι όταν είναι γνωστή η κατανομή, τότε έχουμε τις πληροφορίες τόσο των περιγραφικών μέτρων, όσο και αυτές που περιέχονται στο ιστόγραμμα και επιπλέον έχουμε και ακρίβεια στους υπολογισμούς.

## Ορισμός

Για διακριτή τ.μ. $X$ και για συνάρτηση $g(X)$ αυτής

$$\mathbb{E}X = \sum xP(X = x), \; \mathbb{E}g(X) = \sum g(X)P(X = x).$$

# Παράμετροι κατανομών.

## Ορισμός

Για συνεχή τ.μ. $X$ και για συνάρτηση $g(X)$ αυτής

$$\mathbb{E}X = \int_A x f(x)\,dx, \ \mathbb{E}g(X) = \int_A g(X) f(x)\,dx.$$

## Ορισμός

Για διακριτή/συνεχή τ.μ. $X$ καθώς και για συνάρτηση $g(X)$ αυτής

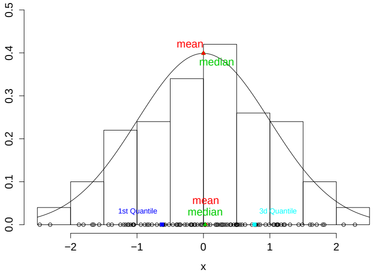
$$\mathbb{V}(X) = \mathbb{E}(X - \mathbb{E}(X))^2, \ \mathbb{V}(g(X)) = \mathbb{E}(g(X) - \mathbb{E}(g(X)))^2.$$

- Η μέση τιμή του δείγματος, $\bar{X}$ (όπως και τα άλλα περιγραφικά μέτρα), είναι προσέγγιση της πραγματικής μέσης τιμής της κατανομής του δείγματος.
- Αυτό φαίνεται και από το σχήμα όπου η δειγματική μέση τιμή ($\bar{X}$, η κόκκινη κουκκίδα στον άξονα $x$) είναι περίπου αλλά όχι ακριβώς στο ίδιο ύψος με το σημείο $(\mu, f(\mu))$ που είναι η πραγματική μέση τιμή (και η ένδειξη πιθανότητας εμφάνισης της)

# Παράμετροι κατανομών.

- Το ίδιο φυσικά ισχύει και για τη διάμεσο που τη βλέπουμε με το πράσινο χρώμα στον άξονα $x$.
- Σε αντίθεση με την σ.π.π. τα περιγραφικά μέτρα υπολογίζονται πάντα γιατί ορίζονται μόνο ως συνάρτηση των τιμών του δείγματος.

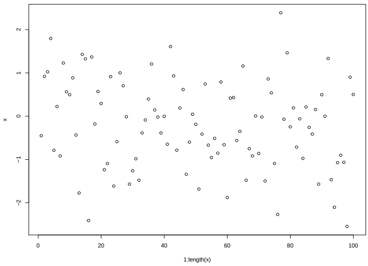
Η σχέση μεταξύ περιγραφικών μέτρων, ιστογράμματος και σ.π.π. είναι ακόμα πιό ξεκάθαρη:
Τα περιγραφικά μέτρα είναι μία πρώτη περιγραφή του δείγματος
Τα περιγραφικά μέτρα είναι η βάση για τη δημιουργία του ιστογράμματος το οποίο είναι μία προσέγγιση της σ.π.π.
Η περισσότερη πληροφορία που μπορούμε να έχουμε για έναν πληθυσμό είναι η σ.π.π. του η οποία ποτέ δεν είναι γνωστή ακριβώς παρά μόνο κατ΄ εκτίμηση.



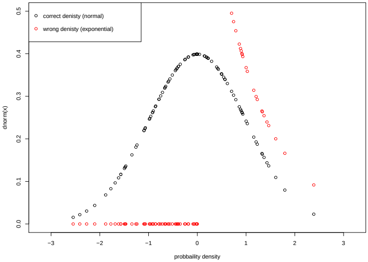

Στο πάνω γράφημα βλέπουμε ότι με το ιστόγραμμα η οπτικοποίηση του δείγματος γίνεται πολύ πιο εύκολα. Παρ΄ όλα αυτά η πληροφορία που παρέχεται από την σ.π.π. είναι ακόμα πιο ακριβής.

# Τυχαίες μεταβλητές και κατανομές πιθανότητας.

Για να διαπιστώσουμε αν οι τιμές $X_1, \dots, X_n$ ενός δείγματος αντιστοιχούν σε τ.μ. κάνουμε τη γρ. παράσταση $(i, X_i), i = 1, \dots, n$.
Αν δεν σχηματίζεται κάποια πατέντα στη γρ. παράσταση (αν τα σημεία είναι τυχαία κατανεμημένα) τότε η μεταβλητή είναι όντως τυχαία.




Η συνάρτηση πυκνότητας $f(X_i)$ αντικατοπτρίζει την πιθανότητα ή ένδειξη πιθανότητας να πάρω την παρατήρηση $X_i$ στο σημείο $X_i$ στον άξονα $x$.
Το γράφημα $(X_i, f(X_i)), i = 1, \dots, n$ έχει νόημα μόνο αν χρησιμοποιήσω τη σωστή σ.π.π. ως $f(x)$ όπως φαίνεται και δίπλα στη εικόνα. Αν χρησιμοποιήσω τη λάθος $f$ (κόκκινα σημεία) τότε το αποτέλεσμα δεν έχει νόημα.



# Διακριτές Κατανομές: Κατανομή Bernoulli

- Μία ακολουθία πειραμάτων στην οποία σε κάθε επανάληψη μπορούν να εμφανισθούν μόνο δύο δυνατά αποτελέσματα (επιτυχία ή αποτυχία),
  - ανεξάρτητα μεταξύ τους έτσι ώστε το αποτέλεσμα οποιουδήποτε πειράματος δεν επηρεάζει τα αποτελέσματα των επόμενων,
  - και η πιθανότητα επιτυχίας (και αποτυχίας) δεν μεταβάλλεται από πείραμα σε πείραμα

  λέγεται πείραμα Bernoulli.
- Μαθηματικά, εκφράζεται με μια τ. μ. X με πεδίο τιμών το 0,1 (συμβολίζει αποτυχία / επιτυχία σε κάθε επανάληψη του πειράματος).
- Η σ.π.π. της X είναι:

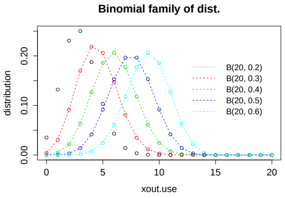
$$f(x) = P(X = x) = \begin{cases} p & \text{αν } x = 1, \\ 1 - p & \text{αν } x = 0. \end{cases}$$

Παραδείγματα: Η ρίψη ενός νομίσματος είναι ένα πείραμα Bernoulli.

- Ο έλεγχος ενός αντικειμένου για το αν είναι ελαττωματικό ή όχι.

# Διακριτές Κατανομές: Διωνυμική Κατανομή

- Αν σε μία ακολουθία Bernoulli ορίσουμε την τ.μ. $X$ να μετράει τον αριθμό των επιτυχιών, τότε η $X$ ακολουθεί την διωνυμική κατανομή με παραμέτρους $n$ (τον αριθμό επαναλήψεων του πειράματος) και $p$ την πιθανότητα επιτυχίας.




- Συμβολικά, $X \sim B(n, p)$. Η σ.π.π. της X είναι:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \; x = 0, 1, \ldots, n, \; 0 < p < 1$$

- Είναι σημαντικό να τονίσουμε ότι σε κάθε επανάληψη η πιθανότητα επιτυχίας $p$ δεν πρέπει να αλλάζει.
- Χρήση: Χρησιμοποιείται σε πειράματα που εμπεριέχουν επαναλαμβανόμενες δοκιμές με δυο πιθανά αποτελέσματα και η πιθανότητα επιτυχίας από επανάληψη σε επανάληψη παραμένει σταθερή π.χ.:
  - εμφάνιση συγκεκριμένης ένδειξης σε ρίψη ζαριού $n$ φορές,
  - αριθμός εμφάνισης συγκεκριμένης ένδειξης (Κ η Γ) σε ρίψη νομίσματος $n$ φορές.

# Διακριτές Κατανομές: Διωνυμική Κατανομή

Παραδείγματα:

- Αριθμός κεφαλών σε ρίψη νομίσματος 5 φορές, αριθμός ελαττωματικών και μη αντικειμένων σε επαναλαμβανόμενη επιλογή, επιτυχία ή αποτυχία $n$ εξεταζόμενων στις εξετάσεις.

Εκτίμηση πιθανότητας επιτυχίας $p$: Ανάλογα με το πρόβλημα μέσω εμπειρικού ορισμού ή κάποιου κατάλληλου κανόνα.

## Χαρακτηρισμοί:

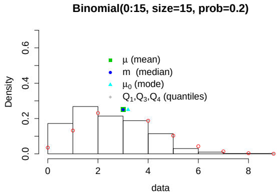
- Ο βασικός χαρακτηρισμός διωνυμικής είναι η σταθερότητα πολυωνυμικής παλινδρόμησης ως προς τη μέση τιμή του δείγματος.
- Αν η δεσμευμένη σ.π.π της $X$ δεδομένης της $X+Y$ είναι υπεργεωμετρική με παραμέτρους $m$ και $n$, τότε $X \sim B(m, \theta), Y \sim B(n, \theta)$ όπου $\theta$ η πιθανότητα επιτυχίας.
- Η μη αρνητική τ.μ. $X$ με σ.π.π $p(x)$ ακολουθεί την διωνυμική κατανομή με παραμέτρους $n, p$ αν και μόνο αν

$$\mathbb{E}(X|X \geq x) = np + \frac{(1-p)xp(x)}{\sum_{j=x}^{n} p(j)}$$

# Διακριτές Κατανομές: Διωνυμική Κατανομή

Ιδιότητες:

$$\mathbb{E}X = np,$$
$$\mathrm{Median}(X) = \lceil np \rceil, \text{ή } \lfloor np \rfloor$$
$$\mathrm{Mode(X)} = \lfloor (n+1)p \rfloor, \text{ή } \lfloor (n+1)p \rfloor - 1,$$
$$\mathbb{V}\mathrm{ar}(X) = np(1-p)$$




Άσκηση 1: Πόσα παιδιά πρέπει να αποκτήσει μία οικογένεια ώστε να έχει με πιθανότητα μεγαλύτερη ή ίση του 0.9 τουλάχιστον ένα αγόρι και τουλάχιστον ένα κορίτσι; Υποθέτουμε ότι σε κάθε γέννηση είναι εξίσου πιθανό να γεννηθεί αγόρι ή κορίτσι.

Λύση: Έστω $n$ ο ζητούμενος αριθμός παιδιών. Έστω $X =$ αριθμός αγοριών, $Y =$ αριθμός κοριτσιών (σε σύνολο $n$ παιδιών). Ισχύει ότι $X \sim Bin(n, 1/2)$ και $Y \sim Bin(n, 1/2)$. Επιπλέον $X + Y = n$. Είναι

$$P(X \geq 1,\ Y \geq 1) = P(X \geq 1,\ n - X \geq 1) = P(1 \leq X \leq n-1) =$$
$$1 - P(X=0) - P(X=n) = 1 - \binom{n}{0}\left(\frac{1}{2}\right)^n - \binom{n}{n}\left(\frac{1}{2}\right)^n = 1 - \frac{1}{2^{n-1}}.$$

Πρέπει $1 - \frac{1}{2^{n-1}} \geq 0.9 \Rightarrow n \geq 1 - \frac{\ln 0.1}{\ln 2} \approx 4.32$, δηλ. τουλ. 5 παιδιά.

# Ασκήσεις στη Διωνυμική Κατανομή

Άσκηση 2: Έστω ότι το 25% από εκείνους που εξετάζονται για την απόκτηση διπλώματος οδηγού αυτοκινήτου αποτυγχάνουν. Έστω ότι η τυχαία μεταβλητή δίνει τον αριθμό των αποτυχόντων ανάμεσα σε είκοσι πέντε εξεταζόμενους. Να υπολογιστούν οι πιθανότητες

1. $P(X \geq 1)$
2. $P(X \leq 20)$
3. $P(5 \leq X \leq 20)$.

Λύση: Ισχύει ότι $X \sim Bin(25, 0.25)$.

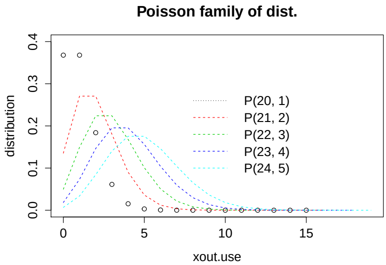
1. $$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - \binom{25}{0} 0.25^0 (1 - 0.25)^{25-0}$$
$$= 1 - 0.75^{25} = 1 - 0.0008 = 0.9992.$$

2. $$P(X \leq 20) = \sum_{x=0}^{20} \binom{25}{x} 0.25^x 0.75^{25-x} \approx 1.$$

3. $P(5 < X \leq 20) = P(X \leq 20) - P(X < 5) = P(X \leq 20) - P(X \leq 4) \approx 0.7863$.

# Διακριτές Κατανομές: Κατανομή Poisson

- Η Poisson θεωρείται κατάλληλη για την περιγραφή τ.μ. που μετράνε εμφάνιση τυχαίων γεγονότων στο χρόνο (η στο χώρο).




- Χρήση: Δομή της Poisson στην απλούστερη περίπτωση όπου τα 'γεγονότα' κατανέμονται τυχαία στο διάστημα $(-\infty, +\infty)$ :
- Οι αριθμοί των 'γεγονότων' που συμβαίνουν σε δύο ξένα μεταξύ τους διαστήματα κατανέμονται ανεξάρτητα ο ένας από τον άλλο,
- ο αναμενόμενος αριθμός 'γεγονότων' σε ένα πεπερασμένο διάστημα $I$ είναι πεπερασμένος και ανάλογος του μήκους του διαστήματος (έστω $\lambda|I|$, $\lambda > 0$),
- η πιθανότητα να συμβούν περισσότερα από ένα 'γεγονότα' στο $I$ τείνει στο 0 ταχύτερα από ότι το $|I|$ όταν $|I| \to 0$.
- Μπορεί να θεωρηθεί ως το όριο διωνυμικής κατανομής για γεγονότα που συμβαίνουν σπάνια και μετά από μεγάλο αριθμό δοκιμών.

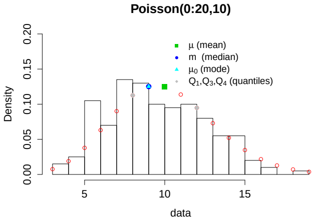
# Διακριτές Κατανομές: Κατανομή Poisson

Λέμε ότι η τυχαία μεταβλητή $X$ ακολουθεί την Poisson κατανομή ή ή $X$ είναι μια Poisson τυχαία μεταβλητή ($x \sim P(\lambda)$), εάν:

$$f(x) = P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!}$$

Ιδιότητες:

$$\mathbb{E}X = \lambda,$$
$$\text{Median}(X) \approx \lfloor \lambda + 1/3 - 0.02\lambda \rfloor$$
$$\text{Mode(X)} = \lfloor \lambda \rfloor, \text{ή } \lceil \lambda \rceil - 1$$
$$\mathbb{V}\text{ar}(X) = \lambda$$




Παρατήρηση: Η παράμετρος $\lambda$ της κατανομής Poisson εκφράζει τον μέσο αριθμό των 'γεγονότων' στην μονάδα του χρόνου.
Παρατήρηση: Για υπολογισμό διαδοχικών πιθανοτήτων μπορεί να χρησιμοποιηθεί ο τύπος

$$\frac{P(x)}{P(x-1)} = \frac{\lambda}{x}.$$

# Εφαρμογές της Poisson

- Πλήθος πελατών που φτάνουν σε ένα κατάστημα, αριθμός ατυχημάτων σε ένα ορισμένο χρονικό διάστημα σε μία συγκεκριμένη περιοχή,
- Αριθμός αιτήσεων αποζημίωσης σε ασφαλιστική εταιρία σε κάποιο χρονικό διάστημα,
- Αριθμός τηλεφωνημάτων που φθάνουν σε ένα τηλεφωνικό κέντρο σε μία ορισμένη χρονική περίοδο της ημέρας,
- Αριθμός των τυπογραφικών λαθών σε μία σελίδα βιβλίου.
- Αριθμός τροχαίων ατυχημάτων σε κάποιο τμήμα του οδικού δικτύου, η για κάποιο χρονικό διάστημα
- Αριθμός επιβατών αεροπορικής πτήσης που δεν εμφανίζονται την ώρα αναχώρησης ενώ έχουν κάνει κράτηση εισιτηρίου
- Αριθμός σωματιδίων που φτάνουν σε συγκεκριμένη επιφάνεια σε συγκεκριμένη χρονική στιγμή.

Εκτίμηση της παραμέτρου $\lambda$: Με δεδομένο ένα δείγμα $X_1, X_2, \ldots, X_n$ από την $P(\lambda)$, η άγνωστη παράμετρος $\lambda$ μπορεί να εκτιμηθεί από την εκτιμήτρια συνάρτηση:

$$\hat{\lambda} = n^{-1} \sum_{i=1}^{n} X_i.$$

# Ασκήσεις στην Κατανομή Poisson

**Χαρακτηρισμοί:**

- Αν ένα δείγμα προέρχεται από την Poisson, τότε η μέση τιμή του ισούται με τη διακύμανσή του. (θα δούμε αργότερα πως μπορεί να ελεγχθεί η υπόθεση).
- Αν οι $X, Y$ είναι ανεξάρτητες και η $X + Y$ ακολουθεί την Poisson τότε και η κάθε μία εκ των $X, Y$ ακολουθεί την Poisson

Άσκηση 1: Τα ελαττώματα κατασκευής σε μεγάλα φύλλα ενός μετάλλου εμφανίζονται τυχαία και με συχνότητα, κατά μέσο όρο, $2.56/100m^2$.

- Να υπολογισθεί η πιθανότητα ανά φύλλο διαστάσεων 4μ x 8μ να μην υπάρχουν καθόλου ελαττώματα.
- Πόσα φύλλα της μορφής αυτής από μία παρτίδα των 100 αναμένεται να έχουν δύο ή περισσότερα ελαττώματα;

Λύση:
$X \sim P(2.56x0.32) = P(0.819), P(X = 0) = e^{-0.819} = 0.4408$=dpois(0, 0.819), $P(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 0.1982$=1-ppois(1, 0.819). Άρα περιμένουμε $0.1982 \times 100 \simeq 20$ ελαττωματικά προϊόντα.

# Ασκήσεις στην Κατανομή Poisson

Άσκηση 2: Έστω ότι ο αριθμός των θανάτων σε ένα νοσοκομείο των Αθηνών σε ένα μήνα ακολουθεί την κατανομή Poisson. Αν η πιθανότητα να συμβεί το πολύ ένας θάνατος σε ένα μήνα είναι τετραπλάσια της πιθανότητας να συμβούν δύο ακριβώς θάνατοι σε ένα μήνα να υπολογιστεί η πιθανότητα

1. να μη συμβεί θάνατος σε ένα μήνα
2. να συμβούν το πολύ δύο θάνατοι σε ένα μήνα.

Λύση: Έστω ότι η τυχαία μεταβλητή $X$ αναπαριστά τον αριθμό των θανάτων σε ένα μήνα. Τότε $X \sim Poisson(\lambda)$. Από τα δεδομένα προκύπτει ότι $P(X \leq 1) = 4P(X = 2)$ δηλαδή

$$e^{-\lambda} + e^{-\lambda}\lambda = 4e^{-\lambda}\frac{\lambda^2}{2!} \Rightarrow 1 + \lambda = 2\lambda^2.$$

Έπεται ότι $\lambda = 1$ ή $\lambda = -1/2$ η οποία είναι μία μη αποδεκτή λύση. Άρα $X \sim Poisson(1)$.

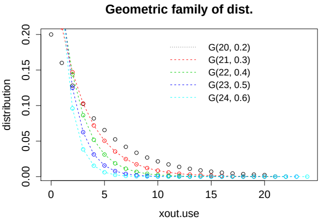
1. $P(X = 0) = e^{-1} \simeq 0.37$.
2. $P(X \leq 2) = \sum_{x=0}^{2} e^{-1}\frac{1^x}{x!} = 2.5e^{-1} \simeq 0.92$.

# Διακριτές Κατανομές: Γεωμετρική Κατανομή

Θεωρούμε μία ακολουθία δοκιμών Benoulli. Έστω Χ ο αριθμός των αποτυχιών πριν εμφανισθεί η πρώτη επιτυχία. Τότε

$$f(x) = P(X = x) = p(1-p)^x, x = 0, 1, \ldots.$$

Ισοδύναμα $X \sim G(q = 1 - p)$.




- Με άλλα λόγια, η γεωμετρική είναι μια διαφορετική θεώρηση ακολουθίας δοκιμών Bernoulli στην οποία ο αριθμός των δοκιμών δεν είναι προκαθορισμένος.
- Σημείωση: Αν σε μια ακολουθία δοκιμών Bernoulli η Χ είναι ο αριθμός των αποτυχιών μέχρις ότου εμφανισθούν $r$ επιτυχίες τότε έχουμε την αρνητική διωνυμική κατανομή με παραμέτρους $r$ και $p$. Εκτίμηση παραμέτρου $p$: Ανάλογα με το πρόβλημα μέσω εμπειρικού ορισμού ή κάποιου κατάλληλου κανόνα.
- Παράδειγμα: Έστω ότι ρίχνουμε ένα ζάρι μέχρι να εμφανιστεί το 1 πρώτη φορά. Η κατανομή πιθανότητας για το πόσες φορές πρέπει να ρίξουμε το ζάρι ($X = 1, 2, 3, \ldots$) είναι γεωμετρική με p=1/6, οπότε

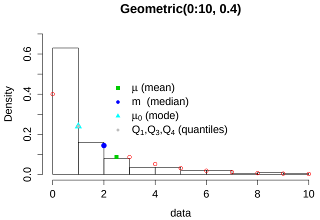
# Διακριτές Κατανομές: Γεωμετρική Κατανομή

$$P(X=1) = 6^{-1}(1-6^{-1})^{1-1} = 1/6$$
$$P(X=2) = 6^{-1}(1-6^{-1})^{2-1} = 5/36$$
$$P(X=3) = 6^{-1}(1-6^{-1})^{3-1} = 25/(6*36)\dots$$

Παρατήρηση: Παρατηρούμε ότι οι πιθανότητες στη γεωμετρική κατανομή είναι όροι μιας γεωμετρικής σειράς με λόγο $q = 1 - p$.

Ιδιότητες:

$$\mathbb{E}X = 1/p,$$
$$\mathrm{Median}(X) = \lceil -\log_2(1-p) \rceil$$
$$\mathrm{Mode(X)} = 1$$
$$\mathbb{V}\mathrm{ar}(X) = (1-p)p^{-2}$$




Σημείωση: Στον ορισμό της μέσης τιμής: έχει έννοια εφόσον η σειρά που ορίζει την μέση τιμή συγκλίνει. Στη γεωμετρική κατανομή υπάρχουν περιπτώσεις που η μέση τιμή δεν υπάρχει.

# Διακριτές Κατανομές: Γεωμετρική Κατανομή

Παραδείγματα:

- Μία τράπεζα αίματος χρειάζεται αίμα ομάδας Β ρέζους αρνητικού και συνεχίζει να αγοράζει αίμα από ιδιώτες μέχρις ότου εμφανισθεί κάποιος με αυτή την ομάδα αίματος. Αν οι αγορές αίματος γίνονται ανεξάρτητα η μία από την άλλη, ο αριθμός X των αγορών που θα γίνουν πριν εμφανισθεί κάποιος με την συγκεκριμένη ομάδα αίματος ακολουθεί την γεωμετρική κατανομή.
- Τυχερά παιχνίδια. Ένας παίκτης ρουλέτας στοιχηματίζει το ίδιο ποσόν α στον ίδιο αριθμό μέχρις ότου κερδίσει για πρώτη φορά. Αν οι στροφές της ρουλέτας γίνονται ανεξάρτητα η μία από την άλλη, ο αριθμός X των φορών που ο παίκτης θα χάσει πριν κερδίσει για πρώτη φορά ακολουθεί την γεωμετρική κατανομή.
- Δειγματοληψία από κάλπη: Σε μία κάλπη υπάρχουν άσπρα και μαύρα σφαιρίδια σε αναλογία $p$ μαύρα και $1-p$ άσπρα. Επιλέγουμε στην τύχη σφαιρίδια από την κάλπη με επανάθεση. (Σε κάθε επιλογή σημειώνουμε το χρώμα του σφαιριδίου και το επανατοποθετούμε στην κάλπη πριν από την επόμενη επιλογή). Ο αριθμός των άσπρων σφαιριδίων που θα επιλεγούν πριν επιλεγεί το πρώτο μαύρο σφαιρίδιο ακολουθεί την γεωμετρική κατανομή με παράμετρο $p$.

# Χαρακτηρισμοί Γεωμετρικής Κατανομής

**Χαρακτηρισμοί:**

- Η Γεωμετρική κατανομή χαρακτηρίζεται από την Μαρκοβιανή ιδιότητα,
$$P(X = x + y|X \geq y) = P(X = x)$$
- Επίσης, αν η $X$ είναι θετική διακριτή τ.μ., τότε
$$P(X = a + b|X \geq a) = P(X > [a + b] - [a])$$
αν και μόνο αν η $X$ ακολουθεί τη Γεωμετρική κατανομή.
- Επίσης η $X$ ακολουθεί τη Γεωμετρική κατανομή αν και μόνο αν
$$\mathbb{E}(X|X > y) = E(X) + [y] + 1$$
όπου [] συμβολίζει το ακέραιο μέρος.

Πολλοί άλλοι χαρακτηρισμοί της γεωμετρικής είναι αντίστοιχοι χαρακτηρισμών της εκθετικής κατανομής.

# Ασκήσεις στη Γεωμετρική Κατανομή

Άσκηση 1: (Ρωσική ρουλέτα): Έστω ότι έχουμε ένα εξάσφαιρο περίστροφο το οποίο έχει μόνο μία σφαίρα.

1. Να βρεθεί η πιθανότητα να εκπυρσοκροτήσει το περίστροφο με την πρώτη δοκιμή.
2. Να βρεθεί η πιθανότητα να εκπυρσοκροτήσει το περίστροφο πριν την τέταρτη δοκιμή.

Λύση:

1. $P(X = 0) = 6^{-1}(1 - 6^{-1})^0 = 1/6 = \text{dgeom}(0, 1/6)$
2. $P(X <= 3) = 1 - (1 - p)^3 = 1 - (5/6)^3 = \text{pgeom}(3\text{-}1, 1/6)$

Άσκηση 2: Από έρευνες έχει διαπιστωθεί ότι οι μαθητές της Γ τάξης του Γυμνασίου καπνίζουν σε ποσοστό 4%. Αν αρχίσουμε και ρωτάμε τον ένα μετά τον άλλον μαθητές της Γ τάξης του Γυμνασίου να απαντήσουν στο ερώτημα αν καπνίζουν ή όχι μέχρις ότου λάβουμε την πρώτη θετική απάντηση, να υπολογιστεί η πιθανότητα να κάνουμε

1. άρτιο αριθμό ερωτήσεων
2. περισσότερες από 8 ερωτήσεις και λιγότερες από 13.

# Ασκήσεις στη Γεωμετρική Κατανομή

Λύση: Έστω $X$ η τυχαία μεταβλητή που αναπαριστά τον αριθμό των ερωτήσεων που θα κάνουμε μέχρις ότου λάβουμε για πρώτη φορά θετική απάντηση. Τότε $X \sim G(p = 4/100)$ Για $x = 1, \ldots$ έχουμε

$$f(x) = P(X = x) = \left(\frac{96}{100}\right)^{x-1} \frac{4}{100}, \; F(x) = P(X \leq x) = 1 - \left(\frac{96}{100}\right)^{x}.$$

1. Η ζητούμενη πιθανότητα είναι ίση με

$$f(2) + f(4) + \ldots = \sum_{x=1}^{\infty} f(2x) = \sum_{x=1}^{\infty} \left(\frac{96}{100}\right)^{2x-1} \frac{4}{100}$$

$$= \frac{4}{100} \frac{100}{96} \sum_{x=1}^{\infty} \left[\left(\frac{96}{100}\right)^{2}\right]^{x} = \frac{24}{49}.$$

2. Η ζητούμενη πιθανότητα είναι ίση με

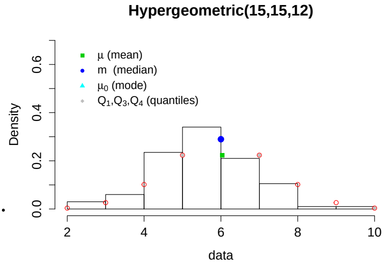
$$P(8 < X < 13) = P(8 < X \leq 12) = F(12) - F(8) =$$

$$\left[1 - \left(\frac{96}{100}\right)^{12}\right] - \left[1 - \left(\frac{96}{100}\right)^{8}\right] \simeq 0.11.$$

# Διακριτές Κατανομές: Υπεργεωμετρική Κατανομή

Η τυχαία μεταβλητή Χ ακολουθεί τη υπεργεωμετρική κατανομή ή Χ είναι μια υπεργεωμετρική τυχαία μεταβλητή (συμβολικά $X \sim H(x : n, m, r)$), εάν:

$$f(x) = P(X = x) = \frac{\binom{m}{x}\binom{n}{r-x}}{\binom{m+n}{r}}, x = 0, 1, \ldots$$




Παράδειγμα: Μέσα σε ένα δοχείο υπάρχουν m μαύρα και n άσπρα σφαιρίδια. Παίρνουμε r σφαιρίδια χωρίς επανατοποθέτηση, τότε η τ.μ. Χ μετράει τον αριθμό των μαύρων σφαιριδίων από τα r.

Εφαρμογές:

- Στην βιομηχανία, όπως είναι γνωστό, ενδιαφέρει ο καθορισμός του ποσοστού των ελαττωματικών αντικειμένων που φθάνουν στην αγορά, και ο περιορισμός τους σε προκαθορισμένα 'ανεκτά' επίπεδα.
- Σε πολλές περιπτώσεις δεν είναι δυνατόν να ελεγχθεί κάθε αντικείμενο που παράγεται, είτε διότι ο έλεγχος είναι πολύ δαπανηρός, είτε διότι συνεπάγεται καταστροφή του ελεγχόμενου προϊόντος.

# Εφαρμογές

- Σε τέτοιες περιπτώσεις χρησιμοποιείται μια μέθοδος δειγματικού ελέγχου που χρησιμοποιεί την έννοια της υπεργεωμετρικής κατανομής.

Παράδειγμα: Έστω ότι τα νέα προϊόντα μιας παραγωγής φθάνουν για έλεγχο σε πακέτα των $N = 50$ αντικειμένων και ότι ένα τυχαίο δείγμα $n = 10$ από αυτά περνά από έλεγχο. Το πακέτο θεωρείται αποδεκτό αν το δείγμα περιέχει το πολύ ένα ελαττωματικό αντικείμενο. Να καθορισθούν οι πιθανότητες αποδοχής του πακέτου ως συναρτήσεις του αριθμού $m$ των ελαττωματικών αντικειμένων σ΄ αυτό.
Λύση: Ζητάμε την πιθανότητα ένα τυχαίο δείγμα μεγέθους 10 να περιέχει ακριβώς $x$ ελαττωματικά προϊόντα. Η περιγραφή ταιριάζει στο μοντέλο της υπεργεωμετρικής οπότε

$$P(X = x) = \binom{m}{x}\binom{50-m}{10-x}\binom{50}{10}^{-1}.$$

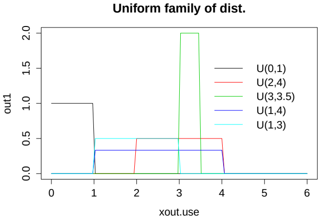
Οπότε η πιθανότητα αποδοχής του πακέτου είναι

$$P(X \leq 1) = P(X = 0) + P(X = 1) = \frac{\binom{m}{0}\binom{50-m}{10}}{\binom{50}{10}} + \frac{\binom{m}{1}\binom{50-m}{10-1}}{\binom{50}{10}}.$$

# Συνεχείς Κατανομές: Ομοιόμορφη Κατανομή

Η απλούστερη μορφή συνεχούς κατανομής πιθανότητας είναι η ομοιόμορφη. Η τ. μ. X ακολουθεί την ομοιόμορφη κατανομή ή X είναι μια ομοιόμορφη τ. μ. (συμβολικά $X \sim U(\alpha,\beta)$), εάν:

$$f(x) = \begin{cases} \frac{1}{\beta-\alpha} & \text{αν } x \epsilon [\alpha,\beta], \\ 0 & \text{αν } x \notin [\alpha,\beta] \end{cases}$$




Προέλευση και χρήση:

- Η ομοιόμορφη κατανομή μπορεί να θεωρηθεί ότι προκύπτει στο πλαίσιο ενός τυχαίου πειράματος κατά το οποίο επιλέγεται τυχαία ένα σημείο από το διάστημα $[a, b]$ με “δίκαιο” (ομοιόμορφο) τρόπο, δηλ. η πιθανότητα παραμένει σταθερή σε όλο το μήκος του διαστήματος του πειράματος.
- Χρησιμοποιείται κατά κόρον στην παραγωγή τυχαίων αριθμών
- Χρησιμοποιείται στην αναπαράσταση σφαλμάτων στρογγυλοποίησης σε $k$ δεκαδικά ψηφία στην αριθμητική ανάλυση.

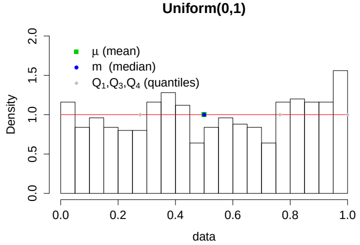
# Συνεχείς Κατανομές: Ομοιόμορφη Κατανομή

## Ιδιότητες:

$$\mathbb{E}X = \frac{\alpha+\beta}{2}, \ \mathbb{V}\mathrm{ar}(X) = \frac{(\beta-\alpha)^2}{12}$$

$$\mathrm{Median}(X) = \frac{\alpha+\beta}{2},$$

$$\mathrm{Mode(X)} = \text{οποιαδήποτε τιμή},$$




Άσκηση 1: Ένας αριθμός $X$ εκλέγεται στο δ/μα $[0,1]$. Να υπολογιστούν οι πιθανότητες (α) το πρώτο δεκαδικό ψηφίο του αριθμού να είναι το 2 ή το 5, (β) αν γνωρίζουμε ότι $X > 0.5$ ποια η πιθανότητα τελικά $X < 0.6$;

Λύση: Η α.σ.κ της $X$ είναι $F(x) = \frac{x-0}{1-0} = x, \ x \in [0,1]$. Για το (α) πρέπει να βρούμε την $P(0.2 \le X \le 0.3$ ή $0.5 \le X \le 0.6) = P(0.2 \le X \le 0.3) + P(0.5 \le X \le 0.6) = 0.3 - 0.2 + 0.6 - 0.5 = 0.2$

Για το (β) έχουμε

$$P(X < 0.6 | X > 0.5) = \frac{P(0.5 < X < 0.6)}{P(X > 0.5)} = \frac{F(0.6) - F(0.5)}{1 - F(0.5)} = 0.2.$$

# Χαρακτηρισμοί ομοιόμορφης κατανομής

## Χαρακτηρισμοί:

- Η τ.μ. $X$ ακολουθεί την ομοιόμορφη κατανομή στο δ/μα $(0, 1)$ αν και μόνο αν
$$\mathbb{E}(-\log(1-X)|X > y) = -\log(1-y) + 1, \ y \in [0, 1).$$
- Ένας εναλλακτικός χαρακτηρισμός, για γνωστές συναρτήσεις $h(X), g(X)$ είναι
$$\mathbb{E}(h(X)|X \geq x) = g(x).$$
- Η τ.μ. $X$ ακολουθεί την ομοιόμορφη κατανομή στο δ/μα $(0, 1)$ αν και μόνο αν
$$\mathbb{E}(X^{-a}|X < y) = \frac{y^{-a}}{1-a}, \ y \in [0, 1).$$

Μία πολύ χρήσιμη ιδιότητα και με πολλές εφαρμογές είναι ότι για οποιαδήποτε τ.μ. $Y$, έχουμε $X = e^{-Y} \sim U(0, 1)$.
Εκτίμηση παραμέτρων: Για δείγμα $X_1, X_2, \ldots, X_n$ από την $U(0, b)$, η άγνωστη παράμετρος $b$ μπορεί να εκτιμηθεί από την $\max\{X_1, \ldots, X_n\}$.

# Ασκήσεις στην ομοιόμορφη κατανομή

Άσκηση 2: Ας θεωρήσουμε ένα όργανο μέτρησης με ακρίβεια τριών δεκαδικών ψηφίων. Το παρεχόμενο από το όργανο αυτό τέταρτο δεκαδικό ψηφίο 109 αποτελεί στρογγυλοποίηση προς τον πλησιέστερο ακέραιο. Τα σφάλματα που προκύπτουν από την στρογγυλοποίηση της μέτρησης δύνανται να θεωρηθούν ότι έχουν την ομοιόμορφη κατανομή $U(\alpha, \beta)$ με $\alpha = -10^{-4}/2, \beta = 10^{-4}/2$. Να υπολογισθούν

1. η πιθανότητα όπως το σφάλμα μέτρησης μιας ποσότητας είναι κατ' απόλυτη τιμή μεγαλύτερο του $10^{-4}/3$
2. η μέση τιμή και η διασπορά του σφάλματος μέτρησης.

Λύση:

1. 

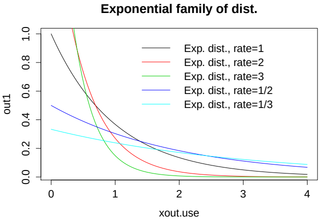
$$P(|X| > 10^{-4}/3) = 1 - P(|X| \leq 10^{-4}/3) =$$
$$1 - P(-10^{-4}/3 \leq X \leq 10^{-4}/3) =$$
$$1 - \{F(10^{-4}/3, -10^{-4}/2, 10^{-4}/2) - F(-10^{-4}/3, -10^{-4}/2, 10^{-4}/2)\} =$$
$$= 1 - 2/3 = 1/3$$

2. $\mathbb{E}(X) = 0, \mathbb{V}\text{ar}(X) = 10^{-8}/12$

# Συνεχείς Κατανομές: Εκθετική Κατανομή

Η τυχαία μεταβλητή Χ ακολουθεί την εκθετική κατανομή ή Χ είναι μια εκθετική τυχαία μεταβλητή (συμβολικά $X \sim exp(\lambda)$), εάν:

$$f(x) = \lambda e^{-\lambda x}, \; x > 0, \; \lambda > 0$$




Χρήσεις:

- Χρησιμοποιείται για την μοντελοποίηση της πιθανότητας να συμβεί ένα τυχαίο γεγονός μέσα σε ένα χρονικό δ/μα.
- Π.χ. ο χρόνος αναμονής μέχρις ότου πραγματοποιηθεί το πρώτο 'γεγονός' σε μια διαδικασία Poisson με παράμετρο $\lambda$ ακολουθεί την εκθετική κατανομή με παράμετρο $1/\lambda$.
- Επίσης χρησιμοποιείται για την μοντελοποίηση του χρόνου: μεταξύ αφίξεων διαφορετικών πελατών σε κατάστημα, συνομιλίας σε μια τηλεφωνική επικοινωνία, ζωής ηλεκτρονικών κομματιών σε μία συσκευή, αναμονής σε τράπεζα, μεταξύ δύο διαδοχικών ισχυρών σεισμών, μεταξύ δύο διαδοχικών κλήσεων σε ένα τηλεφωνικό κέντρο, μεταξύ δύο διαδοχικών ατυχημάτων σε μια διασταύρωση κ.λ.π.

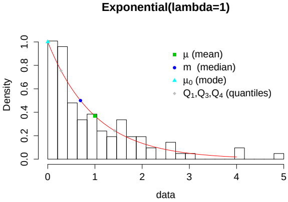
# Ιδιότητες και εκτίμηση στην Εκθετική Κατανομή

Ιδιότητες:

$$\mathbb{E}X = \lambda^{-1}, \ \mathbb{V}\mathrm{ar}(X) = \lambda^{-2}$$

$$\mathrm{Median}(X) = \lambda^{-1}\ln(2),$$

$$\mathrm{Mode(X)} = 0,$$




Γενικά, ο χρόνος που μεσολαβεί μεταξύ δύο διαδοχικών 'γεγονότων' μιας διαδικασίας Poisson ακολουθεί την εκθετική κατανομή.

Εκτίμηση του $\lambda$: Με δεδομένο ένα δείγμα $X_1, X_2, \ldots, X_n$ από την $\exp(\lambda)$, η άγνωστη παράμετρος $\lambda$ μπορεί να εκτιμηθεί από την εκτιμήτρια συνάρτηση:

$$\hat{\lambda} = n^{-1}\sum_{i=1}^{n} X_i,$$

Πρακτική άσκηση 1: Δημιουργήστε ένα τυχαίο δείγμα 20 παρατηρήσεων στην R από την exp(2) (πως;) και εκτιμήστε το $\lambda$.

Πρακτική άσκηση 2: Επαναλάβετε την παραπάνω άσκηση για δείγμα 100, 400, 1000, 5000 παρατηρήσεων. Τι παρατηρείτε για την ακολουθία τιμών του $\lambda$;

(Παρατηρήστε στην βοήθεια της R ότι για την εκθετική rate=1/mean)

# Χαρακτηρισμοί Εκθετικής Κατανομής

## Χαρακτηρισμοί:

- Οι ακόλουθες ιδιότητες είναι χαρακτηριστικές εκθετικών τ.μ.

$$(1-F_X(x_1))(1-F_X(x_2)) = 1-F_X(x_1+x_2), x_1, x_2 > 0,$$
$$1-F(nx) = (1-F(x))^n, n \geq 1, x > 0,$$
$$(1-F(x))' = c(1-F(x)), x > 0.$$

- Η ιδιότητα της αμνησίας (η δεσμευμένη πιθανότητα ο χρόνος αναμονής $X$ να υπερβεί το $x+y$ δοθέντος ότι έχει ήδη υπερβεί το $x$ ισούται με την πιθανότητα το $X$ να υπερβεί το $y$) συχνά χρησιμοποιείται για την ανίχνευση εκθετικού δείγματος. Δηλ. αν η
$$P(X > x+y | X > x) = P(X > y)$$
ισχύει έστω και εμπειρικά τότε αυτό καταδεικνύει εκθετική κατανομή του δείγματος. Με άλλα λόγια αν ο χρόνος που έχει περάσει δεν επηρεάζει τον χρόνο που απαιτείται για την πραγματοποίηση ενός 'γεγονότος' σε μια διαδικασία Poisson τότε έχουμε εκθετική κατανομή.

# Ασκήσεις στην Εκθετική Κατανομή

Άσκηση 1: Ο χρόνος ζωής (σε ώρες) μιας ηλεκτρικής συσκευής ακολουθεί $X \sim Exp(\lambda)$, $\lambda > 0$. Η εταιρεία πουλάει τη συσκευή με καθαρό κέρδος $k$ ευρώ και δίνει στους πελάτες της εγγύηση $\alpha$ ωρών λειτουργίας. Σε περίπτωση που η συσκευή παρουσιάσει βλάβη πριν τη λήξη της εγγύησης, επισκευάζεται δωρεάν από την εταιρεία η οποία και επιβαρύνεται με κόστος επισκευής $k_0$ ευρώ. Αν $Y$ είναι το κέρδος της εταιρείας ανά συσκευή, να υπολογιστεί η ς.π. της $Y$ και να δειχθεί ότι το μέσο κέρδος δίνεται από τον τύπο $\mathbb{E}(Y) = (k - k_0) + k_0 e^{-\lambda\alpha}$.

Λύση: Το κέρδος της εταιρείας ανά συσκευή θα είναι

$$Y = \begin{cases} k, & X > \alpha, \\ k - k_0, & X \leq \alpha \end{cases}$$

οπότε η $Y$ είναι διακριτή τ.μ. που παίρνει 2 τιμές: $k, k - k_0$ με σ.π.π.

$$P(Y = k) = P(X > \alpha) = 1 - F_X(\alpha) = e^{-\lambda\alpha}$$
$$P(Y = k - k_0) = P(X \leq \alpha) = F_X(\alpha) = 1 - e^{-\lambda\alpha}$$

Οπότε η μέση τιμή είναι (Οι πράξεις παραλείπονται ως άσκηση):

$$\mathbb{E}Y = kP(Y = k) + (k - k_0)P(Y = k - k_0)$$

# Ασκήσεις στην Εκθετική Κατανομή

Άσκηση 2: Ας υποθέσουμε ότι η διάρκεια σε λεπτά των υπεραστικών τηλεφωνικών συνδιαλέξεων ακολουθεί την εκθετική κατανομή με μέση τιμή 2 λεπτά. Να βρεθούν οι πιθανότητες η διάρκεια μιας υπεραστικής συνδιάλεξης

- (α) να υπερβεί τα 6 λεπτά,
- (β) να είναι μεταξύ 4 και 6 λεπτών
- (γ) να είναι μικρότερη από 4 λεπτά, και
- (δ) να είναι μικρότερη από 6 λεπτά δεδομένου ότι ήταν μεγαλύτερη από 4 λεπτά.

Λύση : Αν $X$ είναι η διάρκεια σε λεπτά των υπεραστικών τηλεφωνικών συνδιαλέξεων τότε από την εκφώνηση θα ισχύει ότι $\mathbb{E}X = \lambda^{-1} = 2$ και άρα $F_X(x) = P(X \leq x) = 1 - e^{-\lambda x} = 1 - e^{-x/2}$.

- Για το (α): η πιθανότητα να υπερβαίνει η συνομιλία τα 6 λεπτά είναι: $P(X > 6) = 1 - F_X(6) = e^{-6/2} = 0.0498$.
- Για το (β): $P(4 < X < 6) = F_X(6) - F_X(4) = 0.085$.
- Για το (γ): $P(X < 4) = F_X(4) = 0.864$
- Για το (δ):

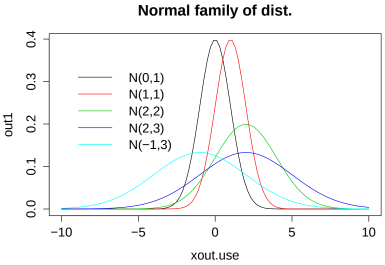
$$P(X < 6|X > 4) = \frac{P(4 < X < 6)}{P(X > 4)} = \frac{F_X(6) - F_X(4)}{1 - F_X(4)}.$$

# Συνεχείς Κατανομές: Κανονική Κατανομή

Η τ.μ. X ακολουθεί την κανονική κατανομή ή X είναι μια κανονική τ.μ. (συμβολικά $X \sim N(\mu, \sigma^2)$), εάν:

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

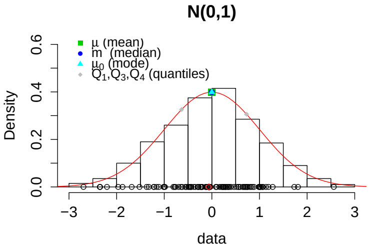
$$\mu \in \mathbb{R}, \ \sigma > 0$$




- Πολλές τ.μ. που περιγράφουν πραγματικά μεγέθη που παίρνουν συνεχείς αριθμητικές τιμές περιγράφονται ικανοποιητικά από την κανονική κατανομή ή από κατανομές που μπορούν να προσεγγισθούν από την κανονική κατανομή.
- Π.χ. οι τ.μ. που μετράνε ανθρώπινα χαρακτηριστικά (ύψος, βάρος).
- Ακόμα, η κανονική προσεγγίζει ικανοποιητικά και πολλές διακριτές κατανομές.
- Ως γενικό κανόνα, είναι η κατάλληλη κατανομή όταν στο υπό μελέτη φαινόμενο οι τιμές τους μαζεύονται συμμετρικά γύρω από μια κεντρική τιμή και 'αραιώνουν' καθώς απομακρύνονται από αυτήν.

# Κανονική Κατανομή: Ιδιότητες

$$\mathbb{E}X = \text{Median}(X) = \text{Mode(X)} = \mu,$$
$$\mathbb{V}\text{ar}(X) = \sigma^2$$




Με δεδομένο ένα δείγμα $X_1, X_2, \ldots, X_n$ από την $N(\mu, \sigma^2)$, οι άγνωστοι παράμετροι $\mu, \sigma^2$ μπορούν να εκτιμηθούν από τις:

$$\hat{\mu} = n^{-1} \sum_{i=1}^{n} X_i, \quad \hat{\sigma}^2 = n^{-1} \sum_{i=1}^{n} (X_i - \hat{\mu})^2$$

Πρακτική άσκηση 1: Δημιουργήστε ένα τυχαίο δείγμα 20 παρατηρήσεων στην R από την $N(0, 1)$ (πως;) και εκτιμήστε τη μέση τιμή και διακύμανση με βάση τους παραπάνω τύπους.

Πρακτική άσκηση 2: Επαναλάβετε την παραπάνω άσκηση για δείγμα 100, 400, 1000, 5000 παρατηρήσεων. Τι παρατηρείτε για την ακολουθία τιμών των $\hat{\mu}, \hat{\sigma}^2$;

# Κανονική Κατανομή: Ιδιότητες

## Χαρακτηρισμοί:

- Η ιδιότητα που χρησιμοποιείται πιο συχνά για την κανονική είναι ότι για δείγμα $X_1, \ldots, X_n$ από οποιαδήποτε κατανομή, για μεγάλο αριθμό παρατηρήσεων,

$$\bar{X} = n^{-1} \sum_{i=1}^{n} X_i \sim N(\mu, n^{-1}\sigma^2)$$

- Αν οι $X_1, \ldots, X_n$ είναι ανά 2 ανεξάρτητες με $\mathbb{E}X_i = 0$, αν

$$L_i = \sum_{j=1}^{n} a_{ij} X_j$$

  είναι γραμμικά ανεξάρτητες και αν $\mathbb{E}(L_1 | L_2, \ldots, L_n) = 0$, τότε τα $X_1, \ldots, X_n$ ακολουθούν την κανονική κατανομή.

- Υπάρχουν ακόμα πολλοί χαρακτηρισμοί της κανονικής κατανομής. Πλήρης λίστα στο βιβλίο: Johnson, Kotz and Balakrishnan: Continuous Univariate Distributions vol. 1, σελ. 100, εκδ. Wiley

# Ασκήσεις στην Κανονική Κατανομή

Άσκηση 1: Σε μία παραλία το βάρος $X$ των χαλικιών ακολουθεί κανονική κατανομή $N(30, 10^2)$. Αν πάρουμε τυχαία ένα χαλίκι (α) Ποια η πιθανότητα να έχει βάρος $X > 30$gr, (β) Ποια η πιθανότητα να έχει βάρος $28 < X < 40$gr, (γ) Πήραμε 10 χαλίκια τυχαία. Ποια η πιθανότητα το συνολικό βάρος τους να είναι πάνω από 300gr;

Λύση: Εφόσον $X \sim N(30, 10^2)$ τότε $Z = (X - 30)/10 \sim N(0, 1)$. (α) $P(X > 30) = P((X - 30)/10 > (30 - 30)/10) = 1 - P(Z \leq 0) = 1 - \Phi(0) = 0.5$ (β) $P(28 < X < 40) = P((28 - 30)/10 < (X - 30)/10 < (40 - 30)/10) = P(-0.2 < Z < 1) = \Phi(1) - \Phi(-0.2) = \Phi(1) - 1 + \Phi(0.2)$ (γ) Έστω $X_1, X_2, \ldots, X_{10}$ το βάρος των 10 χαλικιών. Το άθροισμά τους ακολουθεί πάλι κανονική κατανομή, οπότε για την $W = \sum_{i=1}^{10} X_i$ έχουμε

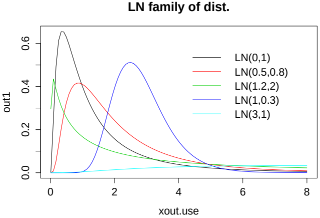
$$\mathbb{E}W = \mathbb{E}\sum_{i=1}^{10} X_i = \sum_{i=1}^{10} \mathbb{E}X_i = 10\mathbb{E}X_1 = 10 \times 30 = 300$$

$$\mathbb{V}(W) = \mathbb{V}(\sum_{i=1}^{10} X_i) = \sum_{i=1}^{10} \mathbb{V}(X_i) = 10 \times 10^2 = 1000.$$

οπότε $W \sim N(300, 1000)$. Θέλουμε την $P(W > 300) =$ αφήνεται ως άσκηση

# Συνεχείς Κατανομές: Λογαριθμοκανονική Κατανομή

Εάν μία τυχαία μεταβλητή X ακολουθεί την κανονική κατανομή, δηλαδή $X \sim N(\mu, \sigma^2)$, τότε η τυχαία μεταβλητή $Y = e^X$ ακολουθεί την λογαριθμοκανονική κατανομή (lognormal) με παραμέτρους $\mu$, $\sigma^2$ και συμβολικά γράφουμε $X \sim LN(\mu, \sigma^2)$.




$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} \quad x \in \mathbb{R},\ \mu \in \mathbb{R},\ \sigma > 0$$

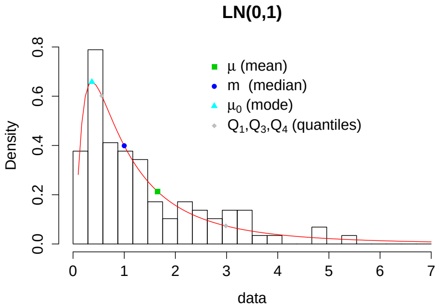
Εφαρμογή: Η λογαριθμοκανονική κατανομή εμφανίζεται σε πολλές εφαρμογές που μια ποσότητα προκύπτει με πολλαπλασιασμό πολλών ανεξάρτητων παραγόντων. Σημειώνεται ότι σε κάποιες περιπτώσεις η Λογαριθμοκανονική κατανομή (με δυο παραμέτρους) είναι πιο ρεαλιστικό μοντέλο για χαρακτηριστικά όπως ύψος / βάρος κ.λ.π. αφού αυτά δεν παίρνουν αρνητικές τιμές.

Παράδειγμα: Η τιμή μίας μετοχή ξεκινάει από τα 100 ευρώ και κάθε μέρα η αξία της μεταβάλλεται είτε +1% είτε -1% με πιθανότητα 1/2. Η κατανομή της μετοχής μετά από 100 ημέρες είναι λογαριθμοκανονική κατανομή.

# Συνεχείς Κατανομές: Λογαριθμοκανονική Κατανομή

Ιδιότητες:

$$\mathbb{E}X = e^{\mu+\frac{1}{2}\sigma^2},$$
$$\text{Median}(X) = e^{\mu},$$
$$\text{Mode(X)} = e^{\mu-\sigma^2},$$
$$\mathbb{V}\text{ar}(X) = (e^{\sigma^2}-1)e^{2\mu+\sigma^2}$$




Με δεδομένο ένα δείγμα $X_1, X_2, \ldots, X_n$ από την $LN(\mu, \sigma^2)$, οι άγνωστοι παράμετροι $\mu, \sigma^2$ της κατανομής μπορούν να εκτιμηθούν από τις εκτιμήτριες συναρτήσεις:

$$\hat{\mu} = n^{-1}\sum_{i=1}^{n} \ln X_i, \quad \hat{\sigma}^2 = n^{-1}\sum_{i=1}^{n} (\ln X_i - \hat{\mu})^2$$

Πρακτική άσκηση: Δημιουργήστε ένα τυχαίο δείγμα 20 παρατηρήσεων στην R από την $LN(0,1)$ (πως;) και εκτιμήστε τη μέση τιμή και διακύμανση με βάση τους παραπάνω τύπους.

# Ασκήσεις στη Λογαριθμοκανονική Κατανομή.

Άσκηση 1: Από μια μελέτη της ποσότητας X του ενζύμου SGPT που περιέχεται στο αίμα των μη φορέων ηπατίτιδας ενός πληθυσμού βρέθηκε ότι $\mathbb{E}(X) = 18.54$ και $\mathbb{V}\mathrm{ar}(X) = 14.03$. Αν είναι γνωστό ότι η τυχαία μεταβλητή X ακολουθεί λογαριθμοκανονική κατανομή να υπολογιστεί το ποσοστό των μη φορέων ηπατίτιδας στους οποίους η ποσότητα του ενζύμου SGPT είναι μικρότερη του 25.

Λύση: Αν συμβολίσουμε με $\mu$ και $\sigma^2$ τις παραμέτρους της λογαριθμοκανονικής κατανομής που ακολουθεί η τ.μ. X, τότε

$$\mathbb{E}X = e^{\mu+\frac{1}{2}\sigma^2} = 18.54$$

$$\mathbb{V}\mathrm{ar}(X) = (e^{\sigma^2} - 1)e^{2\mu+\sigma^2} = (18.54)^2(e^{\sigma^2} - 1) = 14.03$$

Από όπου προκύπτει $\sigma^2 = \log 1.04 = 0.04$, $\mu = 2.9$. Οπότε η ζητούμενη πιθανότητα είναι

$$P(X < 25) = \mathrm{plnorm}(25, 2.9, \mathrm{sqrt}(0.04)) = 0.9445744$$

# Κατανομές στην R

Συναρτήσεις Πυκνότητας Πιθανότητας (σ.π.π.):

- dbinom(x, size, prob): x επιτυχίες σε σύνολο size δοκιμών με πιθανότητα επιτυχίας για κάθε δοκιμή ίση με prob
- dpois(x, lambda): Ακριβώς x επιτυχίες σε μία κατανομή $P(\lambda)$)
- dgeom(x, prob): Ακριβώς x αποτυχίες, Γεωμετρική κατανομή, πιθανότητα επιτυχίας prob.
- dhyper(x, m, n, k): Ακριβώς x επιτυχίες από ένα σύνολο m θετικών και n αρνητικών αποτελεσμάτων, με k δοκιμές κάθε φορά.
- dnorm(x, mean, sd) Κανονική κατανομή με μέση τιμή mean και τυπική απόκλιση sd στο σημείο x του οριζόντιου άξονα.
- dlnorm(x, meanlog, sdlog): Λογαριθμοκανονική κατανομή με παραμέτρους meanlog και sdlog.
- dunif(x, min, max): Ομοιόμορφη κατανομή στο δ/μα (min, max).
- dexp(x, rate): Εκθετική κατανομή με παράμετρο $\lambda$=rate.
- dweibull(x, shape, scale): Κατανομή Weibull με παραμέτρους shape και scale.
- dgamma(x, shape, rate, scale = 1/rate) Κατανομή Γάμμα με παραμέτρους shape και scale.

# Κατανομές στην R

Αθροιστική συνάρτηση κατανομής (σ.π.π.):

- pbinom(x, size, prob): x επιτυχίες σε σύνολο size δοκιμών με πιθανότητα επιτυχίας για κάθε δοκιμή ίση με prob
- ppois(x, lambda): Ακριβώς x επιτυχίες σε μία κατανομή $P(\lambda)$)
- pgeom(x, prob): Ακριβώς x αποτυχίες, Γεωμετρική κατανομή, πιθανότητα επιτυχίας prob.
- phyper(x, m, n, k): Ακριβώς x επιτυχίες από ένα σύνολο m θετικών και n αρνητικών αποτελεσμάτων, με k δοκιμές κάθε φορά.
- pnorm(x, mean, sd) Κανονική κατανομή με μέση τιμή mean και τυπική απόκλιση sd στο σημείο x του οριζόντιου άξονα.
- plnorm(x, meanlog, sdlog): Λογαριθμοκανονική κατανομή με παραμέτρους meanlog και sdlog.
- punif(x, min, max): Ομοιόμορφη κατανομή στο δ/μα (min, max).
- pexp(x, rate): Εκθετική κατανομή με παράμετρο $\lambda$=rate.
- pweibull(x, shape, scale): Κατανομή Weibull με παραμέτρους shape και scale.
- pgamma(x, shape, rate, scale = 1/rate) Κατανομή Γάμμα με παραμέτρους shape και scale.