

Πιθανότητες & Στατιστική

Τυχαίες μεταβλητές - Κατανομές τ.μ.

Δημήτρης Μητσούδης

Παραδείγματα

Ρίχνουμε δύο αμερόληπτα ζάρια. Θεωρούμε την τυχαία μεταβλητή X που ορίζεται ως το άθροισμα των ενδείξεων των δύο ζαριών.

Δειγματικός χώρος Ω :

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

$$\Omega = \{(i, j) \mid i, j = 1, \dots, 6\}$$

$$X(i, j) = i + j.$$

Η τ.μ. X μπορεί να πάρει κάθε ακέραια τιμή μεταξύ 2 και 12.

Μπορούμε να υπολογίσουμε π.χ. τις πιθανότητες:

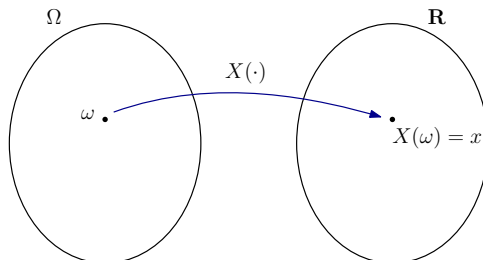
$$P[X = 2] = P[(1, 1)] = 1/36$$

$$P[X = 5] = P[(1, 4), (2, 3), (3, 2), (4, 1)] = 4/36 \dots$$

Τυχαίες Μεταβλητές

Θα ονομάζουμε **τυχαία μεταβλητή (τ.μ.)** μια συνάρτηση $X(\cdot)$ με πεδίο ορισμού το δειγματικό χώρο Ω και τιμές στο $\mathbb{R}$.

Δηλαδή μια τ.μ. απεικονίζει κάθε στοιχείο $\omega \in \Omega$ (αποτέλεσμα του πειράματος) σ' έναν μοναδικό πραγματικό αριθμό $X(\omega) = x$.



Αθροιστική συνάρτηση κατανομής μιας τ.μ.

Η **αθροιστική συνάρτηση κατανομής** (ή συνάρτηση κατανομής) F της τ.μ. X είναι μια συνάρτηση που σε κάθε πραγματικό αριθμό x αντιστοιχεί την πιθανότητα $P[X \leq x]$,

$$F(x) = P[X \leq x].$$

Ιδιότητες αθροιστικής συν. κατανομής

- $0 \leq F(x) \leq 1, -\infty < x < \infty$.
- Η αθροιστική συνάρτηση κατανομής είναι αύξουσα συνάρτηση:

$$F(x_1) \leq F(x_2), \quad -\infty < x_1 \leq x_2 < \infty.$$

Ιδιότητες αθροιστικής συν. κατανομής

Πολλές ερωτήσεις σχετικά με τις πιθανότητες της τ.μ. X μπορούν να απαντηθούν με τη χρήση της αθροιστικής συνάρτησης κατανομής.
Π.χ., έστω ότι θέλουμε να υπολογίσουμε την

$$P[a < X \leq b].$$

Παρατηρήστε ότι: $\{X \leq b\} = \{X \leq a\} \cup \{a < X \leq b\}$ (ξένα ενδεχόμενα).

$$\begin{aligned} \text{Άρα} \quad P[x \leq b] &= P[X \leq a] + P[a < X \leq b] \\ \Rightarrow P[a < X \leq b] &= P[X \leq b] - P[X \leq a] \\ \Rightarrow P[a < X \leq b] &= F(b) - F(a). \end{aligned}$$

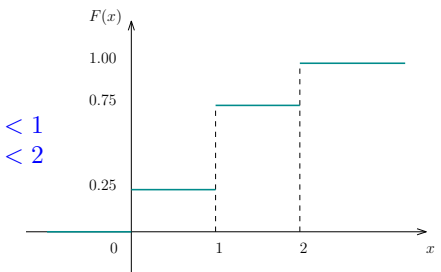
Παράδειγμα

Στρίβουμε ένα αμερόληπτο νόμισμα 2 φορές, και έστω X οι φορές που φέρνω κεφαλή. Τότε η X παίρνει τιμές 0, 1 ή 2. Επίσης

x	0	1	2
$P[X = x]$	0.25	0.50	0.25

Αθρ. συνάρτηση κατανομής:

$$F(x) = P[X \leq x] = \begin{cases} 0, & x < 0 \\ 0.25, & 0 \leq x < 1 \\ 0.75, & 1 \leq x < 2 \\ 1, & x \geq 2 \end{cases}$$



Παράδειγμα

Ας υποθέσουμε ότι η τ.μ. X έχει αθρ. συνάρτηση κατανομής

$$F(x) = \begin{cases} 0 & x \leq 0 \\ 1 - e^{-x^2} & x > 0 \end{cases}$$

Ποια είναι η πιθανότητα $P[X > 1]$;

$$\begin{aligned} P[X > 1] &= 1 - P[X \leq 1] \\ &= 1 - F(1) \\ &= 1 - 1 + e^{-1} \\ &= 0.368 \end{aligned}$$

Τυχαίες μεταβλητές

Οι τ.μ. χωρίζονται σε δύο βασικές κατηγορίες: τις **διακριτές** και τις **συνεχείς**.

Ορισμός

Μια τ.μ. X λέγεται **διακριτή** όταν παίρνει το πολύ ένα αριθμήσιμο πλήθος δυνατών τιμών.

Παράδειγμα: Διακριτή τ.μ.

Ένα άτομο αγοράζει δύο ηλεκτρονικές συσκευές. Η καθεμία είτε λειτουργεί (f) ή είναι ελαττωματική (d).

$$\Omega = \{(d, d), (d, f), (f, d), (f, f)\}.$$

Ας υποθέσουμε επιπλέον ότι τα 4 αυτά πιθανά αποτελέσματα έχουν αντίστοιχες πιθανότητες 0.09, 0.21, 0.21, 0.49.

Αν X είναι η τ.μ. που συμβολίζει τον αριθμό των συσκευών που λειτουργούν, τότε η X παίρνει τις τιμές 0, 1, 2, με αντίστοιχες πιθανότητες

$$P[X = 0] = 0.09, \quad P[X = 1] = 0.42, \quad P[X = 2] = 0.49.$$

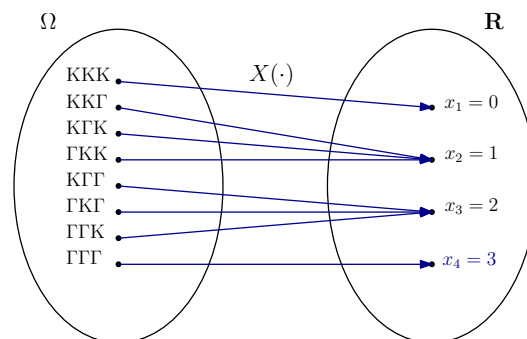
Είναι σύνηθες να εξετάζουμε τις πιθανότητες που σχετίζονται με τις τιμές της X , αντί για τις ισοδύναμές τους στο δειγματικό χώρο Ω .

Παράδειγμα: Διακριτή τ.μ.

Στρίβουμε ένα αμερόληπτο νόμισμα 3 φορές. Ενδιαφερόμαστε για τον αριθμό X των φορών που ήρθε γράμματα. Ο δειγματικός χώρος είναι:

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_8\} = \{KKK, KK\Gamma, K\Gamma K, \Gamma KK, K\Gamma\Gamma, \Gamma K\Gamma, \Gamma\Gamma K, \Gamma\Gamma\Gamma\}.$$

Η τ.μ. X παίρνει τιμές $\{0, 1, 2, 3\}$.



Διακριτές τ.μ.

Μια τ.μ. X που παίρνει διακριτές τιμές x_i , $i = 1, 2, \dots$, λέγεται **διακριτή τ.μ.**

Για μια διακριτή τ.μ. ορίζουμε τη **συνάρτηση πιθανότητας** ως

$$f(x) = P[X = x].$$

Προφανώς:

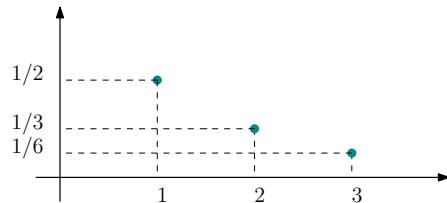
- $f(x_i) \geq 0$, $i = 1, 2, \dots$
- $f(x) = 0$ για κάθε άλλη τιμή του x
- $\sum_i f(x_i) = 1$

Παράδειγμα

Έστω X μια τ.μ που παίρνει τις τιμές $\{1, 2, 3\}$, και έστω f η συνάρτηση πιθανότητας.

Αν γνωρίζουμε ότι $f(1) = 1/2$ και $f(2) = 1/3$, τότε

$$f(3) = 1 - f(1) - f(2) = 1 - 1/2 - 1/3 = 1/6.$$



Συνεχείς τ.μ.

Μια τ.μ. που παίρνει τιμές σε διαστήματα πραγματικών αριθμών θα λέγεται **συνεχής** τ.μ.

Ειδικότερα, θα λέμε **συνεχή** μια τ.μ. αν υπάρχει μια μη αρνητική συνάρτηση $f(x)$ που ορίζεται για κάθε $x \in (-\infty, \infty)$ και έχει την ιδιότητα ότι αν B είναι οποιοδήποτε σύνολο πραγματικών αριθμών, τότε

$$P[X \in B] = \int_B f(t) dt$$

Τότε η f ονομάζεται **συνάρτηση πυκνότητας πιθανότητας** της τ.μ. X .

Η f πρέπει να ικανοποιεί: $1 = P[X \in (-\infty, \infty)] = \int_{-\infty}^{\infty} f(t) dt$

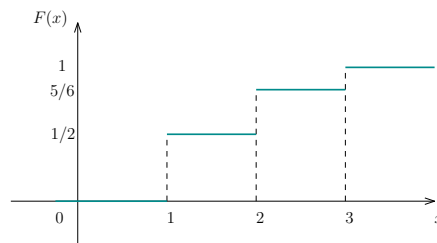
Παράδειγμα (συνέχεια)

Επίσης

$$F(x) = \sum_{t \leq x} f(t)$$

Συνάρτηση κατανομής:

$$F(x) = \begin{cases} 0, & x < 1 \\ 1/2, & 1 \leq x < 2 \\ 5/6, & 2 \leq x < 3 \\ 1, & x \geq 3 \end{cases}$$

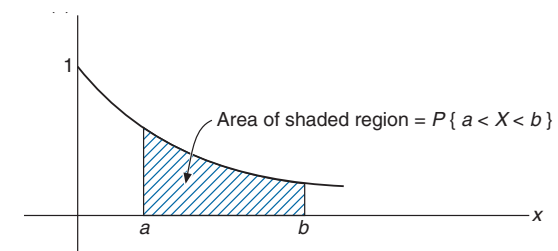


Συνεχείς τ.μ.

Σχέση συν. πυκν. πιθανότητας $f(\cdot)$ και αθρ. συνάρτ. κατανομής $F(\cdot)$:

$$F(x) = P(X \in (-\infty, x]) = \int_{-\infty}^x f(t) dt \Rightarrow \frac{d}{dx} F(x) = f(x)$$

Επιπλέον, $P[a \leq X \leq b] = \int_a^b f(x) dx$



Παράδειγμα - Συνεχείς τ.μ.

- Έστω η τ.μ. X : ο χρόνος ζωής ενός ηλεκτρικού λαμπτήρα.
Η X παίρνει τιμές σ' ολόκληρο το διάστημα $[0, M]$.
- Έστω η τ.μ. X = η διάρκεια ζωής ενός αυτοκινήτου.
Η X παίρνει τιμές σε κάποιο διάστημα $[a, b]$.

Παράδειγμα

Έστω X τ.μ. με σ.π.π.

$$f(x) = \begin{cases} C(4x - 2x^2) & 0 < x < 2 \\ 0 & \text{αλλού} \end{cases}$$

α) Υπολογίστε το C , β) Βρείτε την $P[X > 1]$.

Αφού η f είναι σ.π.π. πρέπει $\int_{-\infty}^{\infty} f(x) dx = 1$.

$$\text{Άρα: } C \int_0^2 (4x - 2x^2) dx = 1 \Rightarrow C \left[2x^2 - \frac{2x^3}{3} \right]_{x=0}^{x=2} = 1 \Rightarrow C = \frac{3}{8}.$$

$$\text{Επίσης: } P[X > 1] = \int_1^{\infty} f(x) dx = \frac{3}{8} \int_1^2 (4x - 2x^2) dx = \frac{1}{2}$$

Χαρακτηριστικές παράμετροι τ.μ.

Έστω X μια **διακριτή** τ.μ. με σ.π. $f(x)$. Η **μέση ή αναμενόμενη τιμή** της X συμβολίζεται με $E[X]$ και ορίζεται ως

$$E[X] = \sum_i x_i f(x_i)$$

Έστω X μια **συνεχής** τ.μ. με σ.π.π. $f(x)$. Η **μέση ή αναμενόμενη τιμή** της X συμβολίζεται με $E[X]$ ή μ και ορίζεται ως

$$E[X] = \mu = \int_{-\infty}^{\infty} x f(x) dx$$

Η μέση τιμή της συνάρτησης μιας διακριτής τ.μ.

Πρόταση

Έστω X μια **διακριτή** τ.μ. που παίρνει τιμές $x_i, i = 1, 2, \dots$, με αντίστοιχες πιθανότητες $p(x_i)$. Τότε για κάθε πραγματική συνάρτηση g ισχύει

$$E[g(X)] = \sum_i g(x_i) p(x_i)$$

$$\text{π.χ. } E[X^2] = \sum_i x_i^2 p(x_i).$$

Η μέση τιμή της συνάρτησης μιας συνεχούς τ.μ.

Πρόταση

Έστω X μια **συνεχής** τ.μ. με σ.π.π. $f(x)$. Τότε για κάθε πραγματική συνάρτηση g ισχύει

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx$$

π.χ. $E[X^2] = \int_{-\infty}^{\infty} x^2 f(x) dx$

Διασπορά συνεχούς τ.μ.

Έστω X μια **συνεχής** τ.μ. με σ.π.π. $f(x)$. Η **διασπορά** της X συμβολίζεται με $\text{Var}[X]$ και ορίζεται ως

$$\text{Var}[X] = E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

Διασπορά διακριτής τ.μ.

Έστω X μια **διακριτή** τ.μ. με σ.π. $f(x)$. Η **διασπορά** της X συμβολίζεται με $\text{Var}[X]$ και ορίζεται ως

$$\text{Var}[X] = E[(X - \mu)^2] = \sum_i (x_i - \mu)^2 f(x_i)$$

Τυπική απόκλιση τ.μ.

Αποδεικνύεται ότι: $\text{Var}(X) = E[X^2] - (E[X])^2$, οπότε εναλλακτικά χρησιμοποιούμε για τη διασπορά μιας τ.μ. τον τύπο:

$$\text{Var}[X] = E[X^2] - (E[X])^2.$$

Ορισμός

Η **τυπική απόκλιση** σ μιας τ.μ. X ορίζεται ως:

$$\sigma = \sqrt{\text{Var}[X]}$$

Ιδιότητες

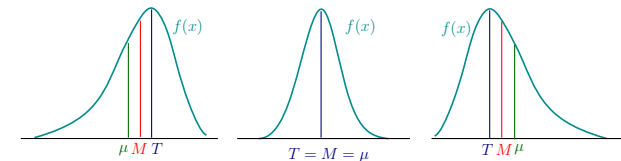
- $E[aX + b] = aE[X] + b$
- $E[X + Y] = E[X] + E[Y]$
- $\text{Var}[aX + b] = a^2 \text{Var}[X]$

p -ποσοστιαίο σημείο, διάμεσος, επικρατούσα τιμή

- Ορίζουμε ως p -ποσοστιαίο σημείο της τ.μ. X κάθε πραγματικό x_p τ.ω.

$$P[X \leq x_p] = p \quad \text{και} \quad P[X \geq x_p] = 1 - p.$$

- Όταν $p = 0.5$ το x_p λέγεται **διάμεσος** και συμβολίζεται με M .
- Ορίζουμε ως **επικρατούσα τιμή** (T) μιας τ.μ. X την τιμή στην οποία λαμβάνει μέγιστο η συνάρτηση πυκνότητας πιθανότητας $f(x)$.



Τυπική τ.μ.

Έστω X τ.μ. με μέση τιμή μ και τυπική απόκλιση σ . Ορίζουμε

$$Z = \frac{X - \mu}{\sigma}.$$

Τότε

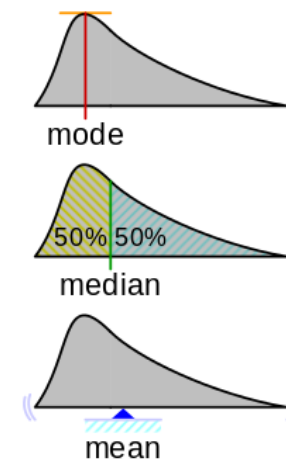
$$E[Z] = E\left[\frac{X - \mu}{\sigma}\right] = \frac{E[X] - \mu}{\sigma} = 0$$

και

$$\text{Var}[Z] = \text{Var}\left[\frac{X - \mu}{\sigma}\right] = \text{Var}\left[\frac{X}{\sigma} - \frac{\mu}{\sigma}\right] = \frac{\text{Var}[X]}{\sigma^2} = 1.$$

Μια τ.μ. με μέση τιμή 0 και τυπική απόκλιση 1 λέγεται **τυπική τυχαία μεταβλητή**.

Επικρατούσα τιμή, Διάμεσος, Αναμενόμενη τιμή



Παράδειγμα 1 (διακριτή τ.μ.)

Ρίχνουμε ένα αμερόληπτο ζάρι, και έστω X ο αριθμός που δείχνει το ζάρι. Τότε η X παίρνει τιμές $1, 2, \dots, 6$. Υπολογίστε τη μέση τιμή της X .

Η σ.π. της X είναι $f(x) = \begin{cases} \frac{1}{6} & x = 1, 2, 3, 4, 5, 6 \\ 0 & \text{αλλού} \end{cases}$

Άρα

$$E[X] = \sum_i x_i f(x_i) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{21}{6} = 3.5$$

Ερμηνεία: Αν εκτελέσουμε ένα μεγάλο αριθμό πειραμάτων, αναμένουμε ο μέσος όρος των παρατηρήσεών μας να προσεγγίζει το 3.5.

$$\text{Var}[X] = \sum_{n=1}^6 (n - 3.5)^2 f(n) = (1 - 3.5)^2 f(1) + \dots + (6 - 3.5)^2 f(6) = 2.9167$$

$$\Rightarrow \sigma = \sqrt{\text{Var}[X]} = \sqrt{2.9167} = 1.7078.$$

Παράδειγμα 2 (συνέχεια)

Βρήκαμε ότι $E[X] = 1.5$. Άρα

$$\begin{aligned} \text{Var}[X] &= \sum_{i=1}^4 (x_i - \mu)^2 f(x_i) \\ &= (0 - 1.5)^2 f(0) + (1 - 1.5)^2 f(1) + (2 - 1.5)^2 f(2) + (3 - 1.5)^2 f(3) \\ &= \frac{3}{4} \\ \Rightarrow \sigma &= \sqrt{\text{Var}[X]} = \frac{\sqrt{3}}{2}. \end{aligned}$$

Παράδειγμα 2 (διακριτή τ.μ.)

Στρίβουμε ένα αμερόληπτο νόμισμα 3 φορές. Έστω X οι φορές που ήρθαν γράμματα. Ο δειγμ. χώρος είναι: $\Omega = \{\text{ΓΓΓ}, \text{ΓΓΚ}, \text{ΓΚΓ}, \text{ΚΓΓ}, \text{ΓΚΚ}, \text{ΚΓΚ}, \text{ΚΚΓ}, \text{ΚΚΚ}\}$.

Η σ.π. της X είναι $f(x) = \begin{cases} 1/8 & x = 0 \\ 3/8 & x = 1 \\ 3/8 & x = 2 \\ 1/8 & x = 3 \\ 0 & \text{αλλού} \end{cases}$

$$\begin{aligned} E[X] &= \sum_i x_i f(x_i) = 0 \cdot f(0) + 1 \cdot f(1) + 2 \cdot f(2) + 3 \cdot f(3) \\ &= 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{12}{8} = 1.5 \end{aligned}$$

Παράδειγμα 3 (συνεχής τ.μ.)

Έστω ότι περιμένουμε ένα μήνυμα μετά τις 5 μ.μ. Από την εμπειρία μας γνωρίζουμε ότι ο χρόνος X σε ώρες, που θα μεσολαβήσει μετά τις 5 μ.μ. μέχρι να φτάσει το μήνυμα είναι μια τ.μ. με σ.π.π.

$$f(x) = \begin{cases} \frac{1}{1.5} & 0 < x < 1.5 \\ 0 & \text{αλλιώς} \end{cases}$$

Τότε:

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx = \int_0^{1.5} \frac{x}{1.5} dx = \frac{2}{3} \left[\frac{x^2}{2} \right]_0^{1.5} = \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{9}{4} = \frac{3}{4}.$$

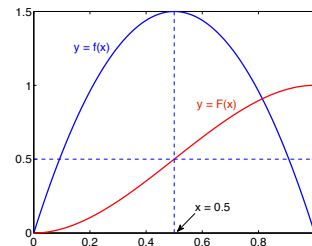
Άρα είναι λογικό να συμπεράνουμε ότι θα χρειαστεί κατά μέσο όρο να περιμένουμε 3/4 της ώρας.

Παράδειγμα 4 (συνεχής τ.μ.)

Η διάμετρος X ενός σύρματος είναι τ.μ. με σ.π.π. $f(x) = 6x(1-x)$, $0 \leq x \leq 1$.

Η αθροιστική συνάρτηση κατανομής θα είναι

$$F(x) = \int_0^x f(t) dt = 6 \left(\frac{x^2}{2} - \frac{x^3}{3} \right).$$



Η μέση τιμή της X είναι:

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx = 6 \int_0^1 (x^2 - x^3)dx = 6 \left[\frac{x^3}{3} - \frac{x^4}{4} \right]_0^1 = \frac{6}{12} = \frac{1}{2}.$$

Παράδειγμα 4 (συνέχεια)

Επίσης

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x)dx = 6 \int_0^1 (x^3 - x^4)dx = 6 \left[\frac{x^4}{4} - \frac{x^5}{5} \right]_0^1 = \frac{6}{20} = \frac{3}{10}.$$

Άρα

$$\text{Var}[X] = E[X^2] - (E[X])^2 = \frac{3}{10} - \frac{1}{4} = \frac{1}{20}.$$

Για την επικρατούσα τιμή κοιτάμε που λαμβάνει μέγιστο η σ.π.π., και βλέπουμε ότι $T = 1/2$.

Όσον αφορά τη διάμεσο ουσιαστικά κοιτάμε για ποιά x η αθροιστική συνάρτηση κατανομής παίρνει την τιμή $1/2$, και βρίσκουμε ότι $M = 1/2$.

Επομένως σε αυτό το παράδειγμα έχουμε: $E[X] = M = T = 1/2$.

Τ.μ. Bernoulli

Έστω ότι εκτελούμε ένα πείραμα το αποτέλεσμα του οποίου μπορεί να χαρακτηριστεί ως «επιτυχία» ή «αποτυχία». Αν $X = 1$ όταν το αποτέλεσμα είναι επιτυχία, και $X = 0$ όταν το αποτέλεσμα είναι αποτυχία, τότε η συνάρτηση πιθανότητας της X είναι

$$P(X = 0) = 1 - p, \quad P(X = 1) = p,$$

όπου p είναι η πιθανότητα το πείραμα να είναι επιτυχές.

Μια τ.μ. λέγεται **Bernoulli** όταν έχει την παραπάνω συνάρτηση πιθανότητας για κάποιο $p \in (0, 1)$.

Διωνυμική κατανομή

Έστω τώρα ότι έχουμε n ανεξάρτητα πειράματα, όπου το αποτέλεσμα του καθενός από αυτά μπορεί να χαρακτηριστεί ως επιτυχία με πιθανότητα p , και ως αποτυχία με πιθανότητα $1 - p$.

Αν X είναι ο αριθμός των επιτυχιών σε n δοκιμές, τότε λέμε ότι η X είναι μια **διωνυμική** τ.μ. με παραμέτρους (n, p) , $X \sim B(n, p)$.

Η συνάρτηση πιθανότητας μιας **διωνυμικής τ.μ. με παραμέτρους n και p** δίνεται από τον τύπο

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \dots, n.$$

Αποδεικνύεται ότι αν $X \sim B(n, p)$, τότε

$$E[X] = np, \quad \text{Var}[X] = np(1 - p).$$

Παράδειγμα

Τα CD που κατασκευάζει μια εταιρεία είναι ελαττωματικά με πιθανότητα 0.01. Η εταιρεία πουλάει CD σε συσκευασίες των 10 και προσφέρει εγγύηση επιστροφής χρημάτων σε περίπτωση που περισσότεροι από 1 στους 10 δίσκους είναι ελαττωματικοί. Σε τι ποσοστό πελατών επιστρέφονται χρήματα; Αν κάποιος αγοράσει τρία πακέτα, ποιά είναι η πιθανότητα να επιστραφεί ακριβώς ένα πακέτο;

Κατανομή Poisson

Μια τ.μ X που παίρνει τις τιμές $0, 1, 2, \dots$, ονομάζεται **τ.μ. Poisson με παράμετρο λ** , αν για κάποιο $\lambda > 0$, ισχύει

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \dots$$

Η κατανομή Poisson χρησιμοποιείται ως προσέγγιση της διωνυμικής κατανομής $B(n, p)$ για n μεγάλο και p μικρό, τέτοια ώστε το np να έχει «μέτρια» τιμή.

Δηλ. αν $X \sim B(n, p)$ και $\lambda = np$, αποδεικνύεται ότι

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \approx e^{-\lambda} \frac{\lambda^k}{k!}.$$

Στην πράξη η προσέγγιση αυτή είναι ικανοποιητική για $n \geq 20$ & $np \leq 10$.

Λύση

Αν X είναι ο αριθμός των ελαττωματικών CD σε ένα πακέτο, τότε η X είναι διωνυμική τ.μ. με παραμέτρους $n = 10$ και $p = 0.01$.

Άρα η πιθανότητα να επιστραφούν χρήματα για κάθε πακέτο είναι

$$\begin{aligned} P(X > 1) &= 1 - P(X = 0 \text{ ή } X = 1) = 1 - P(X = 0) - P(X = 1) \\ &= 1 - \binom{10}{0} (0.01)^0 (0.99)^{10} - \binom{10}{1} (0.01)^1 (0.99)^9 \approx 0.005 \end{aligned}$$

Επομένως, αναμένουμε για το 0.5% των πακέτων να επιστραφούν χρήματα.

Ο αριθμός των πακέτων για τα οποία θα χρειαστεί να επιστραφούν χρήματα είναι διωνυμική τ.μ. με παραμέτρους $n = 3$ και $p = 0.005$. Άρα η πιθανότητα να επιστραφεί ακριβώς ένα στα τρία πακέτα είναι:

$$P(X = 1) = \binom{3}{1} (0.005)^1 (0.995)^2 = 0.015$$

Μέση τιμή και διασπορά κατανομής Poisson

Αποδεικνύεται ότι αν X είναι τ.μ. Poisson με παράμετρο λ , τότε

$$E[X] = \text{Var}[X] = \lambda.$$

Παράδειγμα

Έστω ότι το πλήθος των τυπογραφικών λαθών σε μια σελίδα ενός βιβλίου έχει κατανομή Poisson με παράμετρο $\lambda = 1/2$. Υπολογίστε την πιθανότητα να υπάρχει τουλάχιστον ένα λάθος στη συγκεκριμένη σελίδα.

Λύση: Αν X είναι η τ.μ. του αριθμού των λαθών στη σελίδα έχουμε

$$P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-1/2} \frac{(1/2)^0}{0!} = 1 - e^{-1/2} \approx 0.393.$$

Παράδειγμα

Έστω ότι η παραγωγή ενός προϊόντος γίνεται κάτω από στατιστικό έλεγχο ποιότητας ώστε να πληρούνται οι υποθέσεις των ανεξάρτητων δειγμών Bernoulli. Μια μονάδα του προϊόντος θεωρείται ελαττωματική αν δεν πληροί όλες τις καθορισμένες προδιαγραφές και αυτό συμβαίνει με πιθανότητα $p = 0.01$. Υπολογίστε την πιθανότητα σε ένα κιβώτιο 100 μονάδων του προϊόντος να υπάρχει το πολύ μια ελαττωματική.

Λύση

Αν X είναι η τ.μ. των ελαττωματικών μονάδων του προϊόντος στο κιβώτιο 100 μονάδων, τότε $X \sim B(100, p)$.

Δηλαδή: $P(X = k) = \binom{100}{k} (0.01)^k (0.99)^{100-k}$, $k = 0, 1, 2, \dots, 100$.

Επειδή $n = 100$ (μεγάλο) και $p = 0.01$ (μικρό) έτσι ώστε $\lambda = np = 1 < 10$ η προσέγγιση της διωνυμικής από την Poisson είναι ικανοποιητική, άρα

$$P(X \leq 1) = P(X = 0) + P(X = 1) = e^{-1} \frac{1^0}{0!} + e^{-1} \frac{1^1}{1!} = 2e^{-1} \approx 0.7358.$$

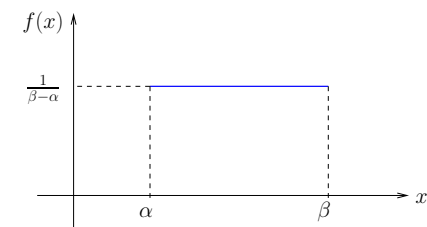
Αν χρησιμοποιήσουμε τη διωνυμική θα έχουμε:

$$P(X \leq 1) = P(X = 0) + P(X = 1) = \dots = 0.3660 + 0.3697 = 0.7357.$$

Ομοιόμορφη κατανομή

Λέμε ότι η τ.μ. X ακολουθεί την ομοιόμορφη κατανομή στο διάστημα $[\alpha, \beta]$ ($X \sim U(\alpha, \beta)$) αν έχει σ.π.π.

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \alpha \leq x \leq \beta \\ 0 & \text{αλλού} \end{cases}$$



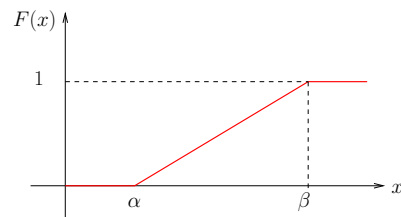
Προφανώς

$$\int_{-\infty}^{\infty} f(x) dx = 1.$$

Ομοιόμορφη κατανομή

Η αθροιστική σ.κ. είναι

$$F(x) = \begin{cases} 0 & x < \alpha \\ \frac{x-\alpha}{\beta-\alpha} & \alpha \leq x \leq \beta \\ 1 & x > \beta \end{cases}$$



Μέση τιμή και διασπορά τ.μ. $X \sim U(\alpha, \beta)$

Μέση τιμή:

$$E[X] = \int_{\alpha}^{\beta} x f(x) dx = \int_{\alpha}^{\beta} \frac{x}{\beta - \alpha} dx = \frac{\beta^2 - \alpha^2}{2(\beta - \alpha)} = \frac{\alpha + \beta}{2}$$

Επίσης

$$E[X^2] = \int_{\alpha}^{\beta} x^2 f(x) dx = \int_{\alpha}^{\beta} \frac{x^2}{\beta - \alpha} dx = \frac{\beta^3 - \alpha^3}{3(\beta - \alpha)} = \frac{\beta^2 + \alpha\beta + \alpha^2}{3}$$

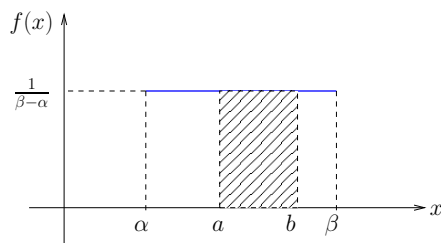
Άρα η διασπορά είναι:

$$\text{Var}[X] = E[X^2] - (E[X])^2 = \frac{\beta^2 + \alpha\beta + \alpha^2}{3} - \left(\frac{\alpha + \beta}{2}\right)^2 = \frac{(\beta - \alpha)^2}{12}$$

Ομοιόμορφη κατανομή

Επίσης αν $[a, b] \subset [\alpha, \beta]$

$$P(a < X < b) = \int_a^b f(x) dx = \frac{1}{\beta - \alpha} \int_a^b dx = \frac{b - a}{\beta - \alpha}$$



Παράδειγμα

Τα λεωφορεία μιας γραμμής φτάνουν σε μια συγκεκριμένη στάση σε 15-λεπτα χρονικά διαστήματα ξεκινώντας από τις 7 π.μ.

(Δηλ. στις 7:00 π.μ., στις 7:15 π.μ., στις 7:30 π.μ., κ.ο.κ.)

Αν ένας επιβάτης φτάσει στη στάση αυτή σε χρόνο που είναι ομοιόμορφα κατανομημένος μεταξύ 7:00 και 7:30 π.μ., βρείτε την πιθανότητα να περιμένει το λεωφορείο

α) λιγότερο από 5 λεπτά;

β) τουλάχιστον 12 λεπτά;

Λύση

Έστω X ο χρόνος σε λεπτά μετά τις 7 π.μ. που ο επιβάτης φτάνει στη στάση.

Αφού $X \sim U(0, 30) \Rightarrow$ θα περιμένει λιγότερο από 5 λεπτά αν φτάσει μεταξύ 7:10 και 7:15 ή μεταξύ 7:25 και 7:30. Άρα

$$P(10 < X < 15) + P(25 < X < 30) = \frac{5}{30} + \frac{5}{30} = \frac{1}{3}.$$

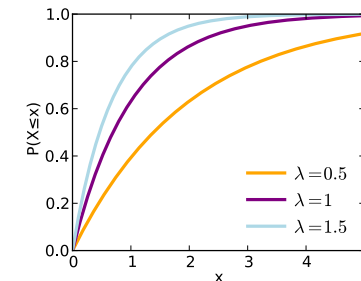
Ανάλογα, θα περιμένει τουλάχιστον 12 λεπτά αν φτάσει μεταξύ 7:00 και 7:03 ή μεταξύ 7:15 και 7:18. Άρα

$$P(0 < X < 3) + P(15 < X < 18) = \frac{3}{30} + \frac{3}{30} = \frac{1}{5}.$$

Εκθετική κατανομή

Η αθροιστική σ.κ. είναι

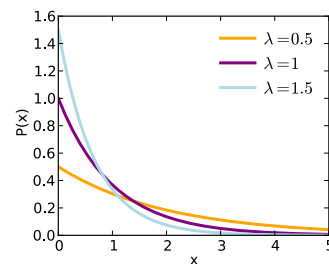
$$F(x) = \begin{cases} 1 - e^{-\lambda x} & 0 \leq x < \infty \\ 0 & x < 0 \end{cases}$$



Εκθετική κατανομή

Λέμε ότι η τ.μ. X ακολουθεί την **εκθετική κατανομή με παράμετρο λ** ($\lambda > 0$) ($X \sim E(\lambda)$) αν έχει σ.π.π.

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & 0 \leq x < \infty \\ 0 & x < 0 \end{cases}$$



Φυσικά

$$\int_{-\infty}^{\infty} f(x) dx = \int_0^{\infty} \lambda e^{-\lambda x} dx = [-e^{-\lambda x}]_0^{\infty} = 1.$$

Μέση τιμή και διασπορά τ.μ. $X \sim E(\lambda)$

Για τη μέση τιμή και τη διασπορά της εκθετικής κατανομής δεν είναι δύσκολο να αποδείξετε ότι:

$$E[X] = \frac{1}{\lambda}, \quad \text{Var}[X] = \frac{1}{\lambda^2}$$

Παράδειγμα εκθετικής κατανομής

Έστω ότι η διάρκεια (σε λεπτά) ενός τηλεφωνήματος σε ένα δημόσιο τηλεφωνικό θάλαμο, ακολουθεί εκθετική κατανομή με μέση τιμή 10 λεπτά. Επίσης, έστω ότι τη στιγμή που κάποιος μπαίνει στον τηλεφωνικό αυτό θάλαμο για ένα τηλεφώνημα ένας άλλος φθάνει εκεί και δεν συναντά κανέναν άλλο να περιμένει. Βρείτε την πιθανότητα ο δεύτερος να περιμένει:

- (α) περισσότερο από 10 λεπτά
- (β) μεταξύ 10 και 20 λεπτών.

Κανονική κατανομή

Θα λέμε ότι μια συνεχής τ.μ. ακολουθεί την **κανονική κατανομή με μέση τιμή μ και διασπορά σ^2** ($X \sim \mathcal{N}(\mu, \sigma^2)$) αν έχει σ.π.π.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty.$$

Όταν

$$X \sim \mathcal{N}(\mu, \sigma^2) \Rightarrow Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1) \quad f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

Η $\mathcal{N}(0, 1)$ λέγεται **τυπική κανονική κατανομή**.

Λύση

Αν X είναι η διάρκεια του τηλεφωνήματος του πρώτου ατόμου έχουμε ότι $X \sim E\left(\frac{1}{10}\right)$, και η αθρ. συν. κατανομής είναι

$$F(x) = \begin{cases} 1 - e^{-\frac{1}{10}x} & x \geq 0 \\ 0 & x < 0 \end{cases}.$$

Επομένως:

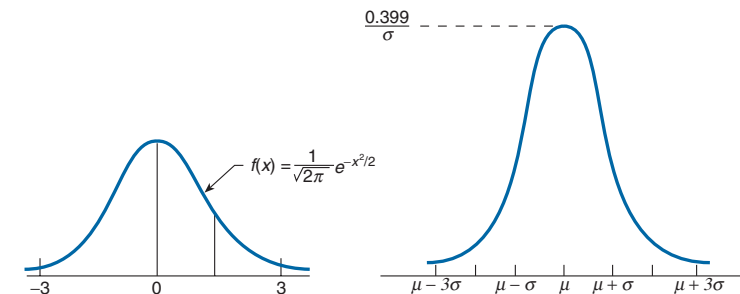
(α)

$$P(X > 10) = 1 - F(10) = e^{-1} = 0.3679.$$

(β)

$$\begin{aligned} P(10 < X \leq 20) &= F(20) - F(10) \\ &= e^{-1} - e^{-2} = 0.3679 - 0.1353 = 0.2326. \end{aligned}$$

Κανονική κατανομή



Κανονική κατανομή

Θα συμβολίζουμε με $\Phi(\cdot)$ την αθροιστική συνάρτηση της τυπικής κανονικής κατανομής. Δηλαδή

$$\Phi(z) = P[Z \leq z] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{t^2}{2}} dt$$

Το γεγονός ότι αν $X \sim \mathcal{N}(\mu, \sigma^2)$, τότε $Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$, μας επιτρέπει αντί να υπολογίζουμε πιθανότητες σχετικά με την τ.μ. X , να υπολογίζουμε πιθανότητες για την τ.μ. Z .

Παράδειγμα

$$P[X \leq b] = P\left[\frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right] = \Phi\left(\frac{b-\mu}{\sigma}\right)$$

ή

$$\begin{aligned} P[a \leq X \leq b] &= P\left[\frac{a-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right] \\ &= P\left[\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right] \\ &= P\left[Z \leq \frac{b-\mu}{\sigma}\right] - P\left[Z < \frac{a-\mu}{\sigma}\right] \\ &= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \end{aligned}$$

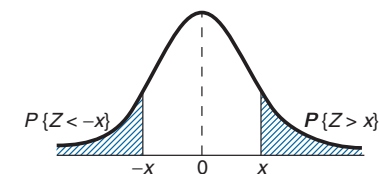
Πίνακας αθρ. συν. τυπ. κανον. κατανομής $\Phi(z)$

	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Αθροιστική συν. τυπικής κανονικής κατανομής

Στον πίνακα δίνονται οι τιμές της $\Phi(z)$ για $z > 0$. Μπορούμε από τον ίδιο πίνακα εκμεταλλευόμενοι τη συμμετρία της $\Phi(z)$ να υπολογίσουμε και τιμές της για $z < 0$. Δηλαδή, αν $z > 0$, τότε

$$\Phi(-z) = P[Z \leq -z] = P[Z \geq z] = 1 - P[Z < z] = 1 - \Phi(z).$$



Παράδειγμα

Αν $X \sim \mathcal{N}(3, 16)$ ($\mu = 3, \sigma^2 = 16$) να υπολογίσετε τις πιθανότητες:

(i) $P[X \leq 11]$, (ii) $P[2 \leq X \leq 7]$, (iii) $P[X > -1]$.

$$(i) \quad P[X \leq 11] = P\left[\frac{X-3}{4} \leq \frac{11-3}{4}\right] = \Phi(2) = 0.9772$$

$$\begin{aligned} (ii) \quad P[2 \leq X \leq 7] &= P\left[\frac{2-3}{4} \leq \frac{X-3}{4} \leq \frac{7-3}{4}\right] \\ &= \Phi(1) - \Phi(-1/4) \\ &= \Phi(1) - (1 - \Phi(1/4)) \\ &= 0.8413 - 1 + 0.5987 = 0.44 \end{aligned}$$

Παράδειγμα

$$\begin{aligned} (iii) \quad P[X > -1] &= P\left[\frac{X-3}{4} > \frac{-1-3}{4}\right] \\ &= P[Z > -1] \\ &= P[Z < 1] = \Phi(1) = 0.8413 \end{aligned}$$

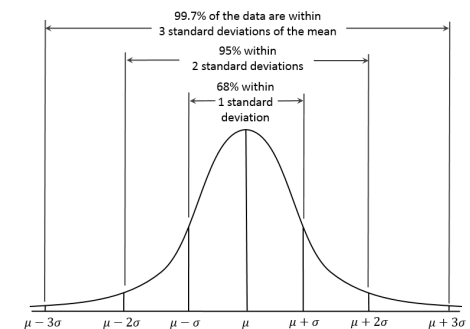
Πίνακας αθρ. συν. τυπ. κανον. κατανομής $\Phi(z)$

	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857

Κανονική κατανομή

Αν κάποιες παρατηρήσεις προέρχονται από την κανονική κατανομή $\mathcal{N}(\mu, \sigma^2)$, τότε το ποσοστό των παρατηρήσεων που απέχει από το μέσο μ λιγότερο από k τυπικές αποκλίσεις μπορεί να υπολογιστεί ως εξής:

$$\begin{aligned} P[|X - \mu| \leq k\sigma] \\ &= P\left[\left|\frac{X - \mu}{\sigma}\right| \leq k\right] \\ &= P[|Z| \leq k] \\ &= P[-k \leq Z \leq k] \\ &= 2\Phi(k) - 1. \end{aligned}$$

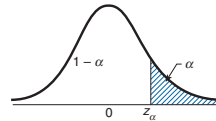


Ποσοστιαία σημεία της Z

Θεωρήστε ένα $0 < \alpha < 1$, και έστω z_α ένας αριθμός τέτοιος ώστε

$$P[Z > z_\alpha] = \alpha.$$

Δηλ., η πιθανότητα η τυπική κανονική τ.μ. Z να είναι μεγαλύτερη από z_α είναι ίση με α .



Προφανώς: $P[Z > z_\alpha] = 1 - P[Z \leq z_\alpha] = 1 - \Phi(z_\alpha)$, άρα

$$\alpha = P[Z > z_\alpha] = 1 - \Phi(z_\alpha).$$

Ποσοστιαία σημεία της Z (συνέχεια)

Τώρα, αν

$$P[Z > z_\alpha] = \alpha \Rightarrow P[Z \leq z_\alpha] = 1 - \alpha,$$

επομένως $100(1 - \alpha)\%$ των περιπτώσεων η τυπική κανονική τ.μ. θα είναι μικρότερη από z_α .

Π.χ. για $\alpha = 0.05$ βρήκαμε ότι

$$P[Z \leq z_{0.05}] = 0.95 \quad \text{και} \quad z_{0.05} = 1.645,$$

άρα κατά 95% η τ.μ. Z θα παίρνει τιμές μικρότερες του 1.645.

Ποσοστιαία σημεία της Z (συνέχεια)

Π.χ., αν $\alpha = 0.05$, τότε

$$P[Z > z_{0.05}] = 0.05 \Leftrightarrow 1 - \Phi(z_{0.05}) = 0.05 \Leftrightarrow \Phi(z_{0.05}) = 0.95,$$

οπότε από τον πίνακα των τιμών της Φ βλέπουμε ότι πρέπει $z_{0.05} = 1.645$ (παρατηρήστε στον πίνακα, ότι $\Phi(1.64) = 0.9495$ και $\Phi(1.65) = 0.9505$, οπότε επειδή ο αριθμός 0.95 βρίσκεται στο μέσον μεταξύ των 0.9495 και 0.9505, βρίσκουμε την τιμή 1.645 ως το μέσο όρο των τιμών 1.64 και 1.65).

Ανάλογα βρίσκουμε ότι

$$z_{0.025} = 1.96 \quad \text{και} \quad z_{0.01} = 2.33.$$