Συνεχείς κατανομές

Πιθανότητες και Στατιστική

Τμήμα Πληροφορικής, Πανεπιστήμιο Πειραιώς

2023-2024

Ομοιόμορφη κατανομή

Μια ΤΜ X η οποία λαμβάνει τιμές στο διάστημα $[a, b]$, όπου $a < b$, λέμε ότι ακολουθεί την **ομοιόμορφη κατανομή** και γράφουμε $X \sim U(a, b)$, ανν η PDF της X δίδεται από τον τύπο

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & \text{αν } x \in [a, b], \\ 0, & \text{αν } x \notin [a, b], \end{cases}$$

οπότε για τη συνάρτηση κατανομής της X ισχύει

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt = \begin{cases} 0, & \text{αν } x < a, \\ \dfrac{x-a}{b-a}, & \text{αν } x \in [a, b], \\ 1, & \text{αν } x > b. \end{cases}$$

Άμεσα αποδεικνύεται ότι η μέση τιμή και η διακύμανση της X είναι αντίστοιχα ίσες με

$$E(X) = \frac{a+b}{2} \quad \text{και} \quad V(X) = \frac{(b-a)^2}{12}.$$

Πράγματι, ισχύει ότι

$$E(X) = \int_a^b x f(x) dx = \int_a^b \frac{x}{b-a} dx = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}.$$

Επίσης,

$$\begin{aligned} E(X^2) &= \int_a^b x^2 f(x) dx = \int_a^b \frac{x^2}{b-a} dx = \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b = \frac{b^3-a^3}{3(b-a)} \\ &= \frac{a^2+ab+b^2}{3}, \end{aligned}$$

οπότε

$$\begin{aligned} V(X) = E(X^2) - E(X)^2 &= \frac{a^2+ab+b^2}{3} - \frac{a^2+2ab+b^2}{4} \\ &= \frac{4a^2+4ab+4b^2-3a^2-6ab-3b^2}{12} = \frac{(b-a)^2}{12}. \end{aligned}$$

Παράδειγμα

Τις βραδινές ώρες οι συρμοί του μετρό διέρχονται κάθε 10 λεπτά από κάποιο σταθμό. Αν κάποιος εισέλθει τυχαία στο σταθμό να βρεθεί:

i) ο μέσος χρόνος αναμονής,

ii) η διακύμανση του χρόνου αναμονής,

iii) η πιθανότητα να περιμένει τουλάχιστον 6 λεπτά για να έρθει ο συρμός.

Λύση

Έστω X ο χρόνος αναμονής στο σταθμό (σε λεπτά).
Η X είναι συνεχής ΤΜ, λαμβάνει τιμές στο διάστημα $[0, 10]$ και ακολουθεί την ομοιόμορφη κατανομή.
i) Ο μέσος χρόνος αναμονής είναι

$$\mu = E(X) = \frac{0+10}{2} = 5 \text{ λεπτά.}$$

Λύση (συνέχεια)

ii) Η διακύμανση του χρόνου αναμονής είναι

$$\sigma^2 = V(X) = \frac{(10-0)^2}{12} = \frac{100}{12} = 8.33 \text{ λεπτά}.$$

iii) Η CDF της X δίδεται από τον τύπο

$$F(x) = P(X \leq x) = \begin{cases} 0, & \text{αν } x < 0, \\ \frac{x}{10}, & \text{αν } x \in [0, 10], \\ 1, & \text{αν } x > 10, \end{cases}$$

οπότε,

$$P(X \geq 6) = 1 - P(x < 6) = 1 - F(6) = 1 - \frac{6}{10} = 0.4.$$

Μια ΤΜ X η οποία λαμβάνει τιμές στο $[0, +\infty)$ λέμε ότι ακολουθεί την *εκθετική κατανομή με παράμετρο* $\lambda > 0$ και γράφουμε $X \sim E(\lambda)$, ανν η PDF της X δίδεται από τον τύπο

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

οπότε η CDF της X δίδεται από τον τύπο

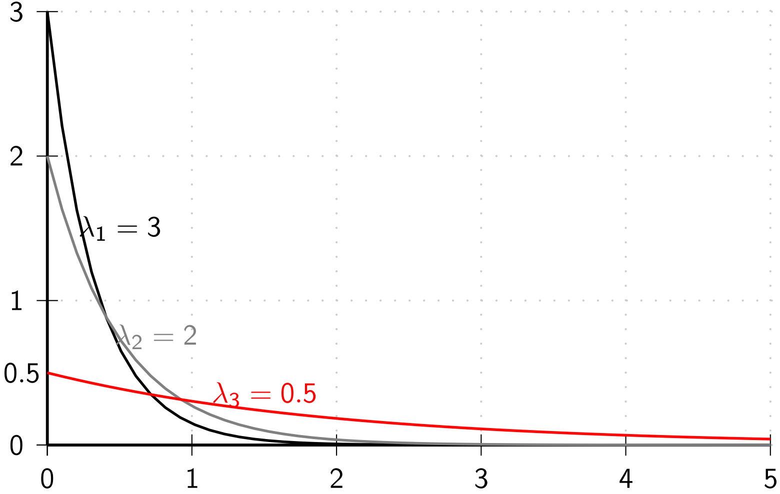
$$F(x) = P(X \leq x) = \int_{-\infty}^{x} \lambda f(t) dt = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Προφανώς, ισχύει ότι

$$P(X > x) = e^{-\lambda x}.$$

Επιπλέον, αποδεικνύεται άμεσα ότι η μέση τιμή και η διακύμανση της X είναι αντίστοιχα ίσες με

$$\mu = E(X) = \frac{1}{\lambda} \text{ και } \sigma^2 = V(X) = \frac{1}{\lambda^2}.$$


3
2
1
0.5
0
0
1
2
3
4
5
$\lambda_1 = 3$
$\lambda_2 = 2$
$\lambda_3 = 0.5$

Εκθετική κατανομή

Η εκθετική κατανομή, όπως και η γεωμετρική κατανομή, έχει την ιδιότητα της έλλειψης μνήμης:

$$P(X > t + s | X > t) = P(X > s) = e^{-\lambda s}, \qquad s, t \geq 0.$$

δηλαδή η πιθανότητα η ΤΜ X να υπερβεί την τιμή $t + s$, όπου $0 < t < t + s$, δοθέντος ότι έχει υπερβεί την τιμή t, είναι ανεξάρτητη του t και ίση προς την αδέσμευτη πιθανότητα να υπερβεί την τιμή s. Πράγματι,

$$\begin{aligned} P(X > t + s | X > t) &= \frac{P(X > t + s)}{P(X > t)} = \frac{1 - P(X \leq t + s)}{1 - P(X \leq t)} = \frac{e^{\lambda(-s-t)}}{e^{-\lambda t}} \\ &= e^{-\lambda s} = 1 - (1 - P(X \leq s)) = P(X > s). \end{aligned}$$

Η εκθετική κατανομή χρησιμοποιείται για την περιγραφή της συμπεριφοράς μεταβλητών όπως π.χ. η διάρκεια τηλεφωνικών συνδιαλέξεων, η διάρκεια λειτουργίας διάφορων μηχανισμών, οι αφίξεις πελατών, κ.ο.κ.

Παράδειγμα

Ο χρόνος ζωής μια μπαταρίας ακολουθεί την εκθετική κατανομή με μέση τιμή 32 μήνες.

i) Ποια είναι η πιθανότητα η μπαταρία να μην χαλάσει στους επόμενους 24 μήνες;

ii) Ποια είναι η πιθανότητα η μπαταρία να χαλάσει μεταξύ του 10ου και του 20ου μήνα;

Λύση

Ο χρόνος X ζωής της μπαταρίας είναι ΤΜ που ακολουθεί την εκθετική κατανομή. Γνωρίζουμε ότι $E(X) = 1/\lambda \Leftrightarrow 32 = 1/\lambda \Leftrightarrow \lambda = 1/32$.

$$i)\ P(X > 24) = 1 - P(X \leq 24) = 1 - (1 - e^{-24\lambda}) = e^{-24/32} = 0.47.$$

$$ii)\ P(10 \leq X \leq 20) = F(20) - F(10) = (1 - e^{-20/32}) - (1 - e^{10/32}) = 0.19.$$

Κατανομή Γάμμα

Υπενθυμίζεται ότι η συνάρτηση Γάμμα ορίζεται ως

$$\Gamma(a) = \int_0^{+\infty} x^{a-1} e^{-x} dx, \qquad \text{για } a > 0.$$

Ειδικά, για κάθε $n \in \mathbb{N}$, ισχύει ότι $\Gamma(n+1) = n!$.
Μια ΤΜ X, η οποία παίρνει τιμές στο $[0, +\infty)$, λέμε ότι ακολουθεί την κατανομή Γάμμα με παραμέτρους $a > 0$ και $\theta > 0$ και γράφουμε $X \sim \Gamma(a, \theta)$, ανν η PDF της X δίδεται από τον τύπο

$$f(x) = \begin{cases} \dfrac{\theta^a}{\Gamma(a)} x^{a-1} e^{-\theta x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Η μέση τιμή και η διακύμανση είναι αντίστοιχα ίσες με a/θ και a/θ^2.
Η εκθετική κατανομή $E(\theta)$ ταυτίζεται με την $\Gamma(1, \theta)$ και η κατανομή $\Gamma(n, \theta)$ ονομάζεται κατανομή Erlang και συμβολίζεται με $E(n, \theta)$.

Αν θεωρήσουμε ένα γεγονός (π.χ. την άφιξη σε μια ουρά αναμονής) που πραγματοποιείται επανειλημμένα σε χρονικό διάστημα $(0, t)$, διαμερίσουμε το διάστημα σε n υποδιαστήματα διάρκειας $\Delta t = t/n$ και θεωρήσουμε ότι η πραγματοποίηση του γεγονότος σε κάθε υποδιάστημα έχει πιθανότητα $p = \lambda \Delta t$ (ανάλογη του χρόνου), είναι ανεξάρτητη από τα υπόλοιπα υποδιαστήματα και είναι πρακτικά απίθανο να συμβεί περισσότερες από μία φορές επειδή το Δt είναι αρκετά μικρό, τότε η μεταβλητή X: πλήθος πραγματοποιήσεων στο διάστημα $(0, t)$ ακολουθεί τη διωνυμική κατανομή. Αν $n \to \infty$, οπότε $\Delta t \to 0$ και $p \to 0$, τότε η κατανομή της X έχει όριο την κατανομή Poisson με παράμετρο $np = \lambda t$.

Διαδικασία Poisson

Αν $N(t)$ είναι η ΤΜ που αντιστοιχεί στο πλήθος αφίξεων σε ένα σύστημα μέχρι τη χρονική στιγμή $t \geq 0$, τότε η οικογένεια $(N(t))_{t\geq 0}$ είναι μια διαδικασία Poisson με παράμετρο λ, αν $N(0) = 0$ και

$$P(N(t_0 + t) - N(t_0) = k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!},$$

αν δηλαδή το πλήθος $N(t_0 + t) - N(t_0)$ των αφίξεων στο χρονικό διάστημα $[t_0, t_0 + t]$ ακολουθεί την κατανομή Poisson.
Η ποσότητα λt εκφράζει τον μέσο αριθμό αφίξεων σε χρονική διάρκεια t. Επομένως, ο μέσος χρόνος μεταξύ δύο διαδοχικών αφίξεων είναι $1/\lambda$, (αφού σε αυτό το χρόνο έχουμε κατά μέσο όρο μία άφιξη).

Διαδικασία Poisson

Αν η τυχαία μεταβλητή $X_n \in [0, +\infty)$ αντιστοιχεί στο χρόνο μέχρι τη n-οστή άφιξη, τότε είναι

$$P(X_n \leq t) = P(N(t) - N(0) \geq n) = \sum_{k \geq n} \frac{e^{-\lambda t}(\lambda t)^k}{k!} = 1 - e^{-\lambda t} \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!}$$

Αν $f(t)$ είναι η PDF της X_n, τότε $P(X_n \leq t) = \int_0^t f(x)dx$, οπότε παραγωγίζοντας κατά μέλη προκύπτει

$$\begin{aligned} f(t) &= \lambda e^{-\lambda t} \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} - \lambda e^{-\lambda t} \sum_{k=1}^{n-1} \frac{k(\lambda t)^{k-1}}{k!} \\ &= \lambda e^{-\lambda t} \left(\sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} - \sum_{k=0}^{n-2} \frac{(\lambda t)^k}{k!} \right) \\ &= \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} \end{aligned}$$

άρα $X_n \sim E(n, \lambda)$, οπότε και $X_1 \sim E(\lambda)$.

Κατανομή Βήτα

Υπενθυμίζεται ότι η συνάρτηση Βήτα ορίζεται ως

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx, \qquad a,b>0$$

και ικανοποιεί τη σχέση $B(a,b) = \dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$.

Μια ΤΜ X, η οποία παίρνει τιμές στο $[0,1]$, λέμε ότι η X ακολουθεί την κατανομή Βήτα με παραμέτρους $a,b>0$ και γράφουμε $X \sim B(a,b)$, αν η συνάρτηση πυκνότητας πιθανότητας της X δίδεται από τον τύπο

$$f(x) = \begin{cases} \dfrac{x^{a-1}(1-x)^{b-1}}{B(a,b)}, & x \in [0,1] \\ 0, & x \notin [0,1] \end{cases}$$

Η μέση τιμή και η διακύμανση είναι αντίστοιχα ίσες με

$$E(X) = \frac{a}{a+b}, \qquad V(X) = \frac{ab}{(a+b)^2(a+b+1)}$$

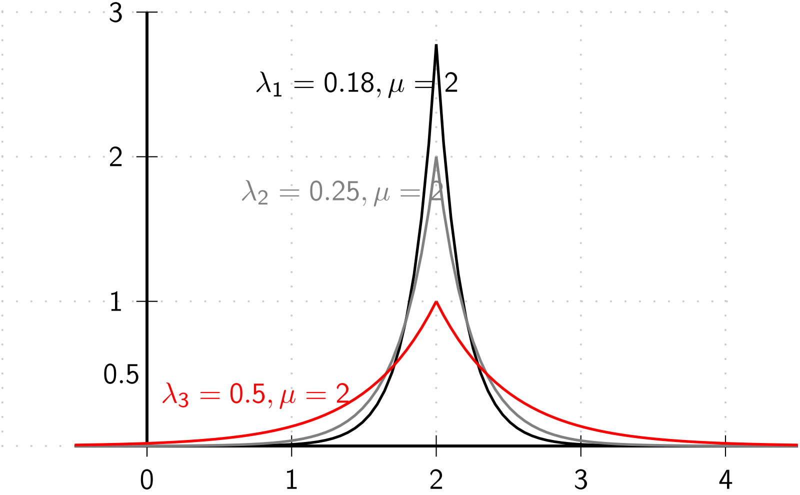
Μια TM X η οποία λαμβάνει τιμές στο $\mathbb{R}$ λέμε ότι ακολουθεί την *κατανομή Laplace* ή διπλή εκθετική με παραμέτρους k και $\lambda > 0$, αν η συνάρτηση πυκνότητας πιθανότητας της X είναι η

$$f(x) = P(X = x) = \frac{1}{2\lambda} e^{-\frac{1}{\lambda}|x-k|}.$$

Η γραφική παράσταση της $f(x)$ είναι συμμετρική ως προς την ευθεία $x = k$. Αποδεικνύεται ότι η μέση τιμή και η διακύμανση της X είναι ίσες με

$$\mu = E(X) = k \text{ και } \sigma^2 = V(X) = 2\lambda^2.$$

Κατανομή Laplace ή διπλή εκθετική



Άσκηση

Ο μέσος χρόνος εξυπηρέτησης ενός πελάτη στα ταμεία μιας τράπεζας είναι 3.5 λεπτά. Να βρεθεί η πιθανότητα ο χρόνος που θα περιμένετε για να εξυπηρετηθεί ο προηγούμενος από εσάς πελάτης να είναι:

i) μεγαλύτερος από πέντε λεπτά.

ii) ανάμεσα σε δύο και τέσσερα λεπτά.

Λύση

Ο χρόνος εξυπηρέτησης X ενός πελάτη ακολουθεί την εκθετική κατανομή με παράμετρο $\lambda = 1/3.5 = 2/7$.

$i)$ $P(X > 5) = 1 - P(X \leq 5) = 1 - (1 - e^{-5 \cdot 2/7}) = e^{-10/7} = 0.239651.$

$ii)$ $P(2 \leq X \leq 4) = P(X \leq 4) - P(X \leq 2) = e^{-2 \cdot 2/7} - e^{-4 \cdot 2/7} = 0.245812.$

Άσκηση

Μια εταιρεία τηλεφώνων εκτιμά ότι η διάρκεια ζωής της μπαταρίας τους ακολουθεί εκθετική κατανομή με μέση τιμή 4 χρόνια. Αν η εταιρεία μπορεί να αντικαταστήσει δωρεάν (μέσα στην εγγύηση) την μπαταρία για το 40% των συσκευών που θα πωληθούν στην αγορά, ποιος πρέπει να είναι ο μέγιστος χρόνος εγγύησης που θα δοθεί για την μπαταρία;

Λύση

Έστω $X \sim E(1/4)$ ο χρόνος ζωής της μπαταρίας και έστω y ο ζητούμενος χρόνος εγγύησης. Η εταιρεία μπορεί να αντικαταστήσει δωρεάν μέχρι το 40% των συσκευών, άρα θα πρέπει $P(X \leq y) \leq 0.4$.

$$P(X \leq y) \leq 0.4 \Leftrightarrow 1 - e^{-y/4} \leq 0.4 \Leftrightarrow 0.6 \leq e^{-y/4} \Leftrightarrow e^{y/4} \leq \frac{5}{3}$$
$$\Leftrightarrow \frac{y}{4} \leq \ln\frac{5}{3} \Leftrightarrow y \leq 4\ln\frac{5}{3} = 2.043.$$

Άρα, ο μέγιστος χρόνος εγγύησης πρέπει να είναι το πολύ 2 χρόνια.

Κανονική κατανομή (Gauss)

Μια ΤΜ X η οποία λαμβάνει τιμές στο $\mathbb{R}$ λέμε ότι ακολουθεί την *κανονική κατανομή με παραμέτρους* μ *και* σ^2 (συμβολισμός $X \sim N(\mu, \sigma^2)$) αν η αθροιστική κατανομή πιθανότητας της X δίδεται από τον τύπο

$$F(x) = P(X \leq x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt, \text{ για κάθε } x \in \mathbb{R}$$

οπότε για τη συνάρτηση πυκνότητας πιθανότητας της X ισχύει

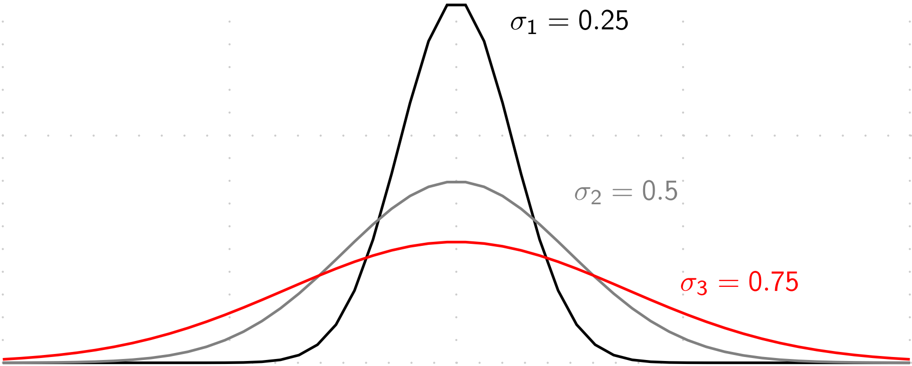
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \text{ για κάθε } x \in \mathbb{R}.$$

Επιπλέον, αποδεικνύεται άμεσα ότι η μέση τιμή και η διακύμανση της X είναι αντίστοιχα ίσες με

$$E(X) = \mu \text{ και } V(X) = \sigma^2$$

Κανονική κατανομή

Η γραφική παράσταση της $f(x)$ έχει κωδωνοειδή μορφή συμμετρική ως προς την κατακόρυφη ευθεία $x = \mu$.



Ως γνωστό, το ολοκλήρωμα της $F(x)$ δεν μπορεί να υπολογισθεί χρησιμοποιώντας στοιχειώδεις συναρτήσεις. Οι τιμές της $F(x)$ υπολογίζονται προσεγγιστικά με αναγωγή στις τιμές της λεγόμενης τυπικής κανονικής κατανομής. Συγκεκριμένα, αν η τ.μ X ακολουθεί την κανονική κατανομή $N(\mu, \sigma^2)$ τότε η ΤΜ

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

ακολουθεί την κανονική κατανομή $N(0, 1)$ (δηλαδή έχει μέσο όρο 0 και διακύμανση 1) η οποία ονομάζεται *τυπική ή τυποποιημένη κανονική κατανομή (standard normal distribution)*.

Στην περίπτωση αυτή, η CDF της Z συμβολίζεται με $\Phi(z)$, ενώ η αντίστοιχη PDF με $\phi(z)$, δηλαδή

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} dt \qquad \text{και} \qquad \phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \quad z \in \mathbb{R}.$$

Κανονική κατανομή

Άμεσα προκύπτουν οι παρακάτω σχέσεις

Ιδιότητες

Για κάθε $z \in \mathbb{R}$ ισχύει η σχέση

$$\Phi(z) + \Phi(-z) = 1.$$

Επιπλέον, αν η $X \sim N(\mu, \sigma^2)$, τότε

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

$$P(a \leq X) = 1 - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

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

Κανονική κατανομή

Παράδειγμα

Έχει διαπιστωθεί ότι ο δείκτης ευφυΐας (IQ) των ανθρώπων ακολουθεί περίπου την κανονική κατανομή $N(105, 20^2)$.

i) Να βρεθεί το ποσοστό των ανθρώπων με δείκτη ευφυΐας α) τουλάχιστον 145, β) το πολύ 90, γ) μεταξύ 80 και 125.

ii) Τι δείκτη ευφυΐας πρέπει να έχει κάποιος για να ανήκει στο ευφυέστερο 1% των ανθρώπων;

Λύση

Έστω $X \sim N(\mu, \sigma^2)$, όπου $\mu = 105$ και $\sigma = 20$, και έστω $Z = (X - \mu)/\sigma$.

$$\begin{aligned}
P(X > 145) &= P((X - 105)/20 > 2) = P(Z > 2) = 1 - \Phi(2) \approx 0.02275, \\
P(X \leq 90) &= P((X - 105)/20 \leq -15/20) = P(Z \leq -0.75) \\
&= 1 - \Phi(0.75) \approx 0.22662, \\
P(80 \leq X \leq 125) &= P(-25/20 \leq (X - 105)/20 \leq 20/20) = P(-1.25 \leq Z \leq 1) \\
&= \Phi(1) - (1 - \Phi(1.25)) \approx 0.7357.
\end{aligned}$$

Λύση (συνέχεια)

ii) Ψάχνουμε μια τιμή x τέτοια ώστε $P(X \geq x) = 0.01$. Είναι

$$P(X \geq x) = 1/100 \Leftrightarrow P(\frac{X-\mu}{\sigma} \geq \frac{x-\mu}{\sigma}) = 0.01 \Leftrightarrow 1 - \Phi(\frac{x-\mu}{\sigma}) = 0.01$$
$$\Leftrightarrow \Phi(\frac{x-\mu}{\sigma}) = 0.99 \Leftrightarrow \frac{x-\mu}{\sigma} = 2.32635$$
$$\Leftrightarrow x = 151.527.$$

Με τον κώδικα της επόμενης διαφάνειας, παίρνουμε την έξοδο:

```
Let X ~ N(mu,sigma^2) be the IQ of a random person, with mu
    = 105, sigma = 20
P(X > 145) = 0.022750131948179195
P(X <= 90) = 0.2266273523768682
P(80 <= X <= 125) = 0.7356949724016876
1% of people have IQ >= 151.5269574808168
or, using the standard normal distribution:
  0.01-percent point z = 2.3263478740408408
  1% of people have IQ >= sigma*zp + mu = 151.5269574808168
```

```
import scipy.stats as st
mu, sigma = 105, 20
rv = st.norm(mu, sigma) #a rv ~ N(mu, sigma^2)
print("Let X ~ N(mu,sigma^2) be the IQ of a random person,
    with mu = %s, sigma = %s"%(mu,sigma))

print("P(X > 145) = %s"%(rv.sf(145)))
#print("P(X>145) = %s"%(1-rv.cdf(145)))
print("P(X <= 90) = %s"%(rv.cdf(90)))
print("P(80 <= X <= 125) = %s"%(rv.cdf(125)-rv.cdf(80)))
print("1% of people have IQ >=", rv.isf(0.01))
#print("1% of people have IQ >=", rv.ppf(0.99))

print("or, using the standard normal distribution:")
zp = st.norm.isf(0.01)
print("\t0.01-percent point z =", zp)
print("\t1% of people have IQ >= sigma*zp+mu =",sigma*zp+mu)
```

Παράδειγμα

Έστω ΤΜ $X \sim N(\mu, \sigma^2)$. Να βρεθούν οι πιθανότητες

i) $P(|X-\mu| \leq \sigma)$

ii) $P(|X-\mu| \leq 2\sigma)$

Λύση

Θέτουμε $Z = (X-\mu)/\sigma$, οπότε $Z \sim N(0,1)$.

i) $P(|X-\mu| \leq \sigma) = P(|Z| \leq 1) = \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 \approx 0.6827$

Δηλαδή, το 68.27% των τιμών της X απέχουν λιγότερο από σ από την μέση τιμή της μ.

ii) $P(|X-\mu| \leq 2\sigma) = P(|Z| \leq 2) = \Phi(2) - \Phi(-2) = 2\Phi(2) - 1 \approx 0.9545$

Δηλαδή, το 95.45% των τιμών της X απέχουν λιγότερο από 2σ από την μέση τιμή της μ.

Παράδειγμα

Ένας μαθητής έδωσε γραπτές και προφορικές εξετάσεις σε κάποιο μάθημα. Στις γραπτές εξετάσεις βαθμολογήθηκε με 71 στα 100, ενώ ο μέσος όρος ήταν 60 με διακύμανση 30. Στις προφορικές εξετάσεις βαθμολογήθηκε με 58 ενώ ο μέσος όρος ήταν 50 με διακύμανση 15. Έστω ότι και στις δύο εξετάσεις οι βαθμολογίες ακολουθούν την κανονική κατανομή, να βρεθεί σε ποια από τις 2 εξετάσεις πήγε καλύτερα;

Λύση

Για κάθε εξέταση θα βρούμε το ποσοστό των συνεξεταζόμενων που έλαβαν μικρότερο ή ίσο βαθμό από αυτόν τον μαθητή. Με άλλα λόγια, αν $X \sim N(60, 30)$ είναι ο βαθμός της πρώτης εξέτασης και $Y \sim N(50, 15)$ ο βαθμός της δεύτερης, θα συγκρίνουμε τις πιθανότητες $P(X \leq 71)$ και $P(Y \leq 58)$ και η μεγαλύτερη θα αντιστοιχεί στην καλύτερη εξέταση.

Λύση (συνέχεια)

Έχουμε λοιπόν ότι

$$P(X \leq 71) = \Phi\left(\frac{71-60}{\sqrt{30}}\right) = \Phi\left(\frac{11}{\sqrt{30}}\right)$$

και

$$P(Y \leq 58) = \Phi\left(\frac{58-50}{\sqrt{15}}\right) = \Phi\left(\frac{32}{\sqrt{30}}\right),$$

άρα τα πήγε καλύτερα στην δεύτερη εξέταση.

Πρόταση

Αν οι ΤΜ X_1, X_2, ..., X_n είναι ανεξάρτητες και ακολουθούν τις κανονικές κατανομές $N(\mu_1, \sigma_1^2)$, $N(\mu_2, \sigma_2^2)$, ..., $N(\mu_n, \sigma_n^2)$ αντίστοιχα, τότε

$$X_1 + X_2 + \cdots + X_n \sim N(\mu_1 + \mu_2 + \cdots + \mu_n, \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2).$$

Παράδειγμα

Το βάρος (σε κιλά) των ατόμων ενός πληθυσμού ακολουθεί την κανονική κατανομή $N(75, 8^2)$. Έστω ότι το κεντρικό ασανσέρ ενός κτηρίου μπορεί να μεταφέρει με ασφάλεια βάρος μέχρι 900 κιλά. Να βρεθούν:

i) Η πιθανότητα 15 άτομα να χρησιμοποιήσουν συγχρόνως το ασανσέρ με ασφάλεια.

ii) Ο μέγιστος αριθμός ατόμων που με πιθανότητα τουλάχιστον 99% μπορούν να χρησιμοποιήσουν συγχρόνως το ασανσέρ.

Λύση

i) Έστω X_i το βάρος του ο i-οστού ατόμου, για κάθε $i \in [15]$ και έστω $X = X_1 + X_2 + \cdots + X_{15}$. Οι ΤΜ X_i ακολουθούν την κανονική κατανομή $N(75, 8^2)$, επομένως η X ακολουθεί την κανονική κατανομή $N(15 \cdot 75, 15 \cdot 8^2) = N(1125, 15 \cdot 8^2)$.

$$P(X \leq 900) = P(\frac{X - 1125}{\sqrt{15} \cdot 8} \leq \frac{900 - 1125}{\sqrt{15} \cdot 8}) = P(Z \leq -\frac{15\sqrt{15}}{8})$$
$$= P(Z \leq -7.26184) \approx 0.$$

ii) Έστω $S_n := \sum_{i=1}^{n} X_i$, οπότε $S_n \sim N(n\mu, n\sigma^2)$ και $Z_n := \dfrac{S_n - n\mu}{\sigma\sqrt{n}} \sim N(0,1)$. Θέτοντας $z_n = \dfrac{900 - n\mu}{\sigma\sqrt{n}}$, είναι

$$P(S_n \leq 900) \geq 0.99 \Leftrightarrow P(Z_n \leq z_n) \geq 0.99 \Leftrightarrow \Phi(z_n) \geq 0.99$$
$$\Leftrightarrow z_n \geq \Phi^{-1}(0.99)$$

Λύση (συνέχεια)

Η τελευταία ισοδυναμία ισχύει διότι η Φ είναι αύξουσα.
Θέτοντας $x = 900$ και $z = \Phi^{-1}(0.99) \approx 2.32635$, είναι

$$\begin{aligned} z_n \geq z &\Leftrightarrow \frac{x - n\mu}{\sigma\sqrt{n}} \geq z \Leftrightarrow \mu n + z\sigma\sqrt{n} - x \leq 0 \\ &\Leftrightarrow \sqrt{n} \leq \frac{-z\sigma + \sqrt{z^2\sigma^2 + 4x\mu}}{2\mu} \approx 3.342251 \\ &\Leftrightarrow n \leq 11.17 \end{aligned}$$

Επομένως, ο μέγιστος αριθμός ατόμων είναι ίσος με 11.

Ο επόμενος κώδικας Python δείχνει τη διαδικασία εύρεσης της τιμής z ώστε $P(Z \leq z) = p = 0.99$.

```python
import numpy as np
from scipy.stats import norm
import matplotlib.pyplot as plt

mu, sigma = 0, 1     #standard normal distribution
rv = norm(loc=mu, scale=sigma) #a normal random variable rv
x = np.linspace(-5,5,100) #range of values for rv
p= 0.99 #desired accuracy
z = rv.ppf(p) #inverse of cdf
print("P(|Z| <= %s) = %s"%(z, rv.cdf(z)))
print("p-percent point for p = %s is z = %s"%(p,rv.ppf(p)))
```

Output

```
P(|Z| <= 2.3263478740408408) = 0.99
p-percent point for p = 0.99 is z = 2.3263478740408408
```

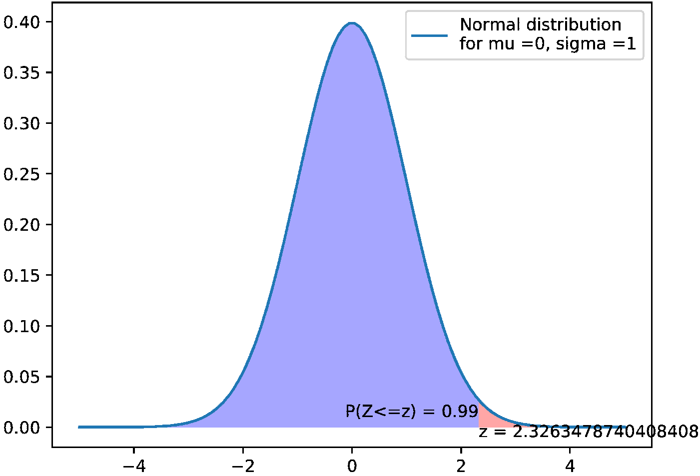
```
#plt.figure(figsize=(12,10))
label1 = "Normal distribution\n"+ "for mu = %s, sigma = %s"
    %(mu,sigma)
plt.plot(x,rv.pdf(x), label=label1)

plt.fill_between(x=np.arange(z,5,0.01),
                 y1= rv.pdf(np.arange(z,5,0.01)),
                 facecolor='red', alpha=0.35)

plt.fill_between(x=np.arange(-5,z,0.01),
                 y1= rv.pdf(np.arange(-5,z,0.01)),
                 facecolor='blue', alpha=0.35)

plt.text(x=z, y=-0.01, s = "z = %s"%(z))
plt.text(x=z, y=0.01, s = "P(Z<=z) = %s"%(p),
    horizontalalignment='right')
plt.legend()
plt.show()
```

Κανονική κατανομή



Άσκηση

Ο χρόνος που χρειάζεται ένας πεζός για να φτάσει από τον σταθμό του Ηλεκτρικού του Πειραιά στο κεντρικό κτήριο του Πα.Πει λόγω φαναριών δεν είναι σταθερός αλλά ακολουθεί την κανονική κατανομή με μέση τιμή 17 λεπτά και διακύμανση 9 λεπτά. Να υπολογισθεί η πιθανότητα ο πεζός να φτάσει:

i) σε λιγότερο από 19 λεπτά,

ii) αργότερα από 22 λεπτά

iii) σε λιγότερο από 21 αλλά περισσότερο από 13 λεπτά.

Λύση

Έστω $X \sim N(17, 9)$ ο χρόνος μετάβασης από τον σταθμό του Ηλεκτρικού στο κεντρικό κτήριο του Πα.Πει.

$$
\begin{aligned}
i)\ P(X \leq 19) &= P\Big(\frac{X-17}{3} \leq \frac{19-17}{3}\Big) = P\Big(Z \leq \frac{2}{3}\Big) = \Phi\Big(\frac{2}{3}\Big) \\
&= \Phi(0.66) = 0.7454. \\
ii)\ P(X \geq 22) &= 1 - P(X \leq 22) = 1 - P\Big(\frac{X-17}{3} \leq \frac{22-17}{3}\Big) \\
&= 1 - P\Big(Z \leq \frac{5}{3}\Big) = 1 - \Phi\Big(\frac{5}{3}\Big) = \Phi\Big(-\frac{5}{3}\Big) = \Phi(-1.66) \\
&= 0.0485. \\
iii)\ P(13 \leq X \leq 21) &= P\Big(\frac{13-17}{3} \leq \frac{X-17}{3} \leq \frac{21-17}{3}\Big) \\
&= P\Big(\frac{-4}{3} \leq Z \leq \frac{4}{3}\Big) = \Phi\Big(\frac{4}{3}\Big) - \Phi\Big(-\frac{4}{3}\Big) = 2\Phi\Big(\frac{4}{3}\Big) - 1 \\
&= 2\Phi(1.33) - 1 = 2 \cdot 0.9082 - 1 = 0.8164.
\end{aligned}
$$

Άσκηση

Ο Λεωνίδας έχει κανονίσει συνάντηση στο σταθμό του μετρό Άγιος Δημήτριος με φίλους του και πρέπει να είναι εκεί σε 20 λεπτά, αν δεν θέλει να τους καθυστερήσει. Έχει τις εξής επιλογές:

1: Να περιμένει το λεωφορείο το οποίο αναμένεται να έρθει σε 7 λεπτά και με χρόνο μετάβασης που ακολουθεί την $N(10, 9)$.

2: Να περπατήσει για 10 λεπτά μέχρι το σταθμό του μετρό. Οι συρμοί έρχονται στο σταθμό κάθε 10 λεπτά και η μετάβαση διαρκεί 4 λεπτά.

3: Να καλέσει ταξί, το οποίο βρίσκεται 10 λεπτά μακριά και ο χρόνος μετάβασης ακολουθεί την $N(6, 4)$.

Με βάση τα παραπάνω στοιχεία, να βρεθεί η πιθανότητα να μην αργήσει στην συνάντηση

i) σε κάθε μια από τις επιλογές,

ii) αν διαλέξει τυχαία μια από τις 3 επιλογές.

iii) Αν ο Λεωνίδας άργησε να πάει στη συνάντηση, ποια είναι η πιθανότητα να μην πήρε ταξί;

Λύση

Έστω $B \sim N(10, 9)$ ο χρόνος μετάβασης με λεωφορείο. Η πιθανότητα να μην αργήσει παίρνοντας το λεωφορείο είναι ίση με

$$P(B \leq 13) = \Phi((13 - 10)/3) = \Phi(1) = 0.84134$$

Έστω $M \sim U(0, 10)$ ο χρόνος αναμονής του συρμού του μετρό. Η πιθανότητα να μην αργήσει παίρνοντας το μετρό είναι ίση με

$$P(M \leq 6) = (6 - 0)/(10 - 0) = 0.6$$

Έστω $T \sim N(6, 4)$ ο χρόνος μετάβασης του ταξί. Η πιθανότητα να μην αργήσει παίρνοντας ταξί είναι ίση με

$$P(T \leq 10) = \Phi((10 - 6)/2) = \Phi(2) = 0.97725$$

Λύση (συνέχεια)

Αν A, B, M, T, είναι αντίστοιχα τα ενδεχόμενα να άργησε, να πήρε λεωφορείο, μετρό, ταξί, τότε $P(B) = P(M) = P(T) = 1/3$ και

$$\begin{aligned} P(\overline{A}) &= P(\overline{A}|B)P(B) + P(\overline{A}|M)P(M) + P(\overline{A}|T)P(T) \\ &= \frac{1}{3}(0.84134 + 0.6 + 0.97725) = 0.8062 \end{aligned}$$

και

$$P(T|A) = \frac{P(A|T)P(T)}{P(A)} = \frac{1}{3} \cdot \frac{1 - 0.97725}{1 - 0.8062} = 0.04 \Rightarrow P(\overline{T}|A) = 0.96.$$