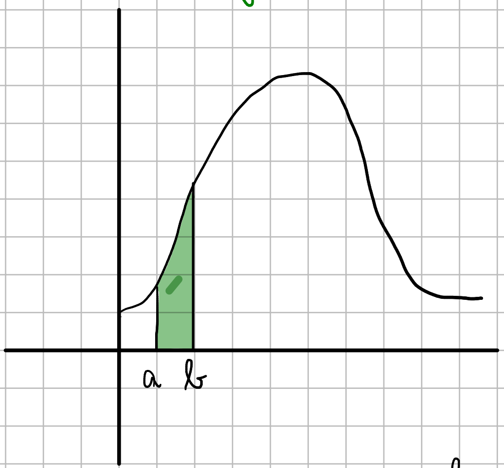
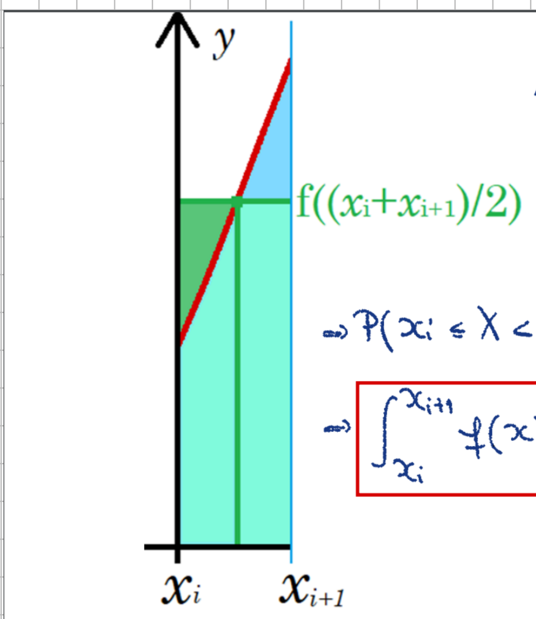
Matematici speciale
(curs 7 - S7)

04.04.2024

1. Histogramă



$P(a \le x \le b) = \int_a^b f(x) = dx$ = arie trapez dreptunghic



$$A = h * f\left(\frac{x_i + x_{i+1}}{2}\right)$$

$$\Rightarrow P(x_i \le X < x_{i+1}) = \int_{x_i}^{x_{i+1}} f(x)\,dx \simeq A$$

$$\Rightarrow \boxed{\int_{x_i}^{x_{i+1}} f(x)\,dx \simeq (x_{i+1} - x_i) \cdot f\left(\frac{x_{i+1} + x_i}{2}\right)}$$

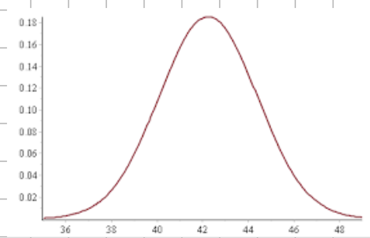
$A = \text{baza} \cdot h = H_i \cdot h$

$A = P(x_i \le x \le x_{i+1}) \simeq \frac{n_i}{N}$

$n_i \rightarrow$ nr. de valori în intervalul $[x_i, x_{i+1}]$

$N \rightarrow$ nr. total de înregistrări $\{x_i, \ldots, x_N\}$

$H_i = \frac{n_i}{Nh}$

V.a. continue (nume)	Notație + parametrii	Densitatea de probabilitate (f)	Proprietăți
Distribuția exponențială	$X \sim Exp(\theta)$ $\theta > 0$	$f(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{\theta} \cdot e^{-\frac{x}{\theta}}, & x \geq 0 \end{cases}$	$M(X) = \theta$ $\sigma^2(X) = \theta^2$ Lipsa de memorie $P(X > s + t \mid X > s) = P(X > t)$
Distribuția uniformă pe un interval $[a,b]$	$X \sim Unif([a,b])$	$f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a,b] \\ 0, & \text{în rest} \end{cases}$	$P(x \leq X < x+L) = \int_x^{x+L} \frac{1}{b-a} dt = \frac{1}{b-a} \cdot t \Big\vert_x^{x+L} = \frac{L}{b-a}$ Probabilitatea ca v.a X sa aiba valori intr-un interval de lungime L.
Distribuția normală (gaussiană)	$X \sim N(m, \sigma^2)$ $m \in \mathbb{R}, \sigma > 0$ mulțimea valorii $X \subset \mathbb{R}$	$f(x; m, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{\frac{-(x-m)^2}{2\sigma^2}}$ $f(m) = \frac{1}{\sigma\sqrt{2\pi}}$ $(m, f(m))$ $\rightarrow$ coord. vârf. clopot	$G_f \longrightarrow$ clopotul lui Gauss $V(m, f(m))$ 
Distribuția normală standard	$m = 0, \sigma = 1$ $X \sim N(0,1)$	$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$	

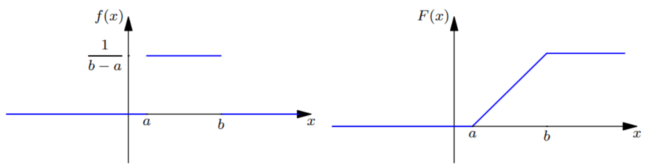
Distribuția exponențială

- Parametru: θ;
- Domeniu de valori $[0, \infty)$;
- Notație:Exp(θ);
- Den de prob. $f(x) = \begin{cases} 0 & \text{dacă } x < 0 \\ \frac{1}{\theta} e^{-x/\theta} & \text{dacă } x \geq 0 \end{cases}, \quad \theta > 0$
- Funcția de repartiție: $F(x) = \begin{cases} 0 & \text{dacă } x < 0 \\ 1 - e^{-x/\theta} & \text{dacă } x \geq 0 \end{cases},$
- $M(X) = \theta$, $\sigma^2(X) = \theta^2$
- Model: timpul de așteptare pentru un proces continuu
- Lipsa de memorie: probabilitatea de a aștepta încă t minute nu este afectată de aceea de a fi așteptat deja s minute fără eveniment.

$$P(X > s + t / X > s) = P(X > t)$$

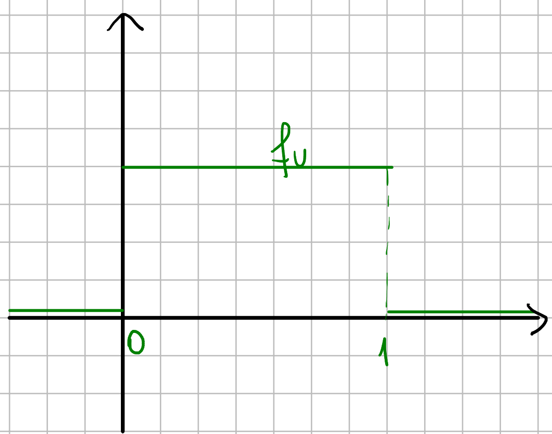
Distribuția uniformă

- Parametrii: a, b
- Domeniu de valori(adică mulțimea de valori posibile unde $f_X(x) > 0$): $[a, b]$
- Notație: Unif[a,]) sau uniform[a,b];
- Model: valorile din domeniu au probabilitate egală de apariție;
- den. de prob. $f(x) = \begin{cases} \frac{1}{b-a} & \text{dacă } x \in [a, b] \\ 0 & \text{dacă } x \notin (a, b) \end{cases}$
- funcție de repartiție $F(x) = \begin{cases} 0, & \text{dacă } \quad x < a \\ \frac{x-a}{b-a} & \text{dacă } \quad a \leq x < b \\ 1 & \text{dacă } \quad x \geq b \end{cases}$



Caz special

$U \sim \text{Unif}([0,1]) \Rightarrow f_U(x) = \begin{cases} 1, & x \in [0,1] \\ 0, & \text{în rest} \end{cases}$



Distribuția normală

- Parametrii: m, σ;
- Domeniu de valori $(-\infty, \infty)$;
- Notație: $N(m, \sigma)$ sau $N(m, \sigma^2)$
- Den de prob: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \frac{-(x-m)^2}{2\sigma^2}, \quad x \in \mathbb{R}$,
- Funcția de repartiție: nu are formulă (valorile se pot determina folosind tabele sau software matematice: `pnorm` in R)
- Dacă $X \sim N(m, \sigma^2)$, atunci $M(X) = m, \sigma^2(X) = \sigma^2$.
- Model: măsurarea erorilor, înălțime, media big data;

Distribuția normală standard

- Parametrii: $m = 0, \sigma = 1$;
- Den de prob: $\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}, \ t \in \mathbb{R}.$;
- φ este pară;
- graficul său se numește clopotul lui Gauss;
- Notație $Z \sim N(0, 1)$;
- Funcția de repartiție: $\Phi(x) = P(Z \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$
- $M(Z) = 0, \sigma^2(Z) = 1$

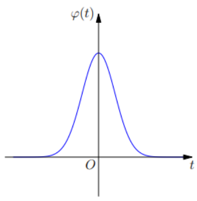


Figure: Densitatea de probabilitate a unei variabile aleatoare $Z \sim N(0, 1)$.

Funcția de repartiție a variabilei aleatoare $Z \sim N(0, 1)$ se notează cu Φ:

$$\Phi(x) = P(Z \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$$

Proprietăți:

- $\Phi(-x) = 1 - \Phi(x), \ x \in \mathbb{R}$
- Valoarea funcției Φ într-un punct x se poate calcula doar prin metode aproximative, dar există tabele pentru aceste valori;

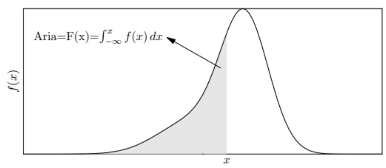


Figure: Ilustrarea semnificației geometrice a valorii funcției de repartiție într-un punct.

Observații legate de parametrii m și σ^2 ai distribuțiilor normale

Se poate arăta că dacă $X \sim N(m, \sigma^2)$ atunci $M(X) = m$

$\sigma^2(X) = \sigma^2$

$$M(X) = \int_{-\infty}^{\infty} x\, f(x; m, \sigma^2)\, dx = m$$

$$\sigma^2(X) = M(X^2) - [M(X)]^2 = \sigma^2$$

Funcția de repartiție: $F_X(x) = P(X \le x) = P(X \in (-\infty, x)) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{\frac{-(t-m)^2}{2\sigma^2}} dt$

Notații clasice folosite:

	Funct. dens. de probabilitate	Funcția de repartiție
X - Distribuție normală	$f(x; m, \sigma^2)$	$F_X(x; m, \sigma^2)$
$Z \sim N(0,1)$ Distribuție normală standard	$\varphi(x, m=0, \sigma^2=1) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$	$\phi(x, m=0, \sigma^2=1) = \phi(x)$

Proprietăți ale funcției de repartiție ϕ $(Z \sim N(0,1))$

1. $\phi(-x) = 1 - \phi(x)$, $\forall x \in \mathbb{R}$

2. pt. $x = 0 \Rightarrow \phi(0) = 1 - \phi(0) \Rightarrow \phi(0) = \frac{1}{2}$

$\phi(0) = P(Z \le 0)$

3. $\exists$ tabel cu val. fct. ϕ

Mediana unei variabile aleatoare X este acea valoare x pentru care:

$P(X \geq x) = P(X \leq x) = 0{,}5$

mediană : $x_1 < x_2 < \ldots < x_N$

m_e = mediană

$F_X(m_e) = 0{,}5$

Definiție mediană

Mediana unei variabile aleatore X este valoarea x pentru care $P(X \leq x) = P(X \geq x) = 0.5$.
Adică, acea valoare x pentru care $F(x) = 0.5$.

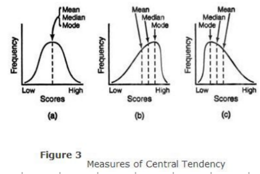


Figure 3 Measures of Central Tendency

De exemplu, dacă $X \sim N(\mu, \sigma^2)$, atunci

$$P(\mu - \sigma \leq X \leq \mu + \sigma) \approx 0.68,$$
$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0.95,$$
$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx 0.99.$$

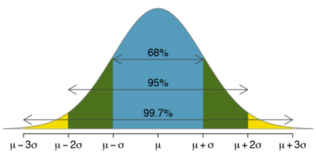


Figure: Regula 3-σ a unei variabile aleatoare $X \sim N(\mu, \sigma)$.

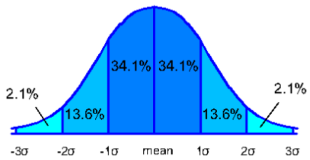


Figure: Regula 3-σ a unei variabile aleatoare $Z \sim N(0, 1)$.

De exemplu, dacă $Z \sim N(0, 1)$, atunci

$$P(-1 \leq Z \leq 1) \approx 0.68,$$
$$P(-2 \leq Z \leq 2) \approx 0.95,$$
$$P(-3 \leq Z \leq 3) \approx 0.99.$$

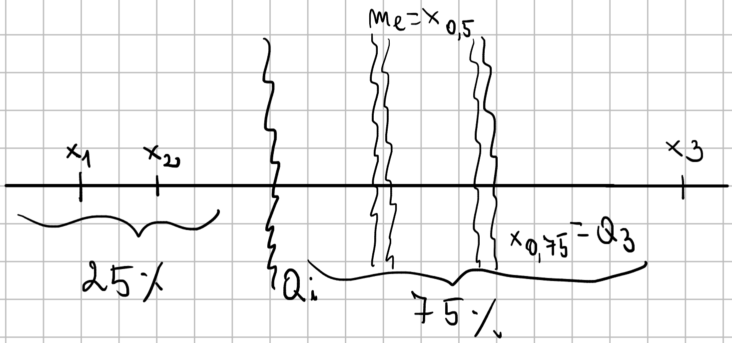
De exemplu, se poate calcula $\phi(1)$. Intr-adevăr,
$\phi(1) = P(Z \leq 1) \approx 0.68 + 0.16 = 0.84.$

Conceptul de α-cuantilă

Notăm x_α și e acea valoare a v.a. X pentru care: $F(x_\alpha) = P(X < x_\alpha) = \alpha$, $\alpha \in [0,1]$

Echivalent, $x_\alpha = F^{-1}(\alpha)$, unde $F : \mathbb{R} \to (0, 1)$, strict crescătoare, inversa $F^{-1} : (0, 1) \to \mathbb{R}$.
Evident, mediana este $x_{0.5}$.
Observație: Pentru $Z \sim N(0, 1)$, notăm z_α.

→ sunt cuantile de ordinul 4 ⇒ ne împart valorile v.a. X în 4 grupe



$F(x_{0,25}) = 0,25 \Rightarrow x_{0,25} = Q_i$

Prop: $z_{1-\alpha} = -z_\alpha$

pt. $\alpha = 0,5 \Rightarrow z_{0,5} = -z_{0,5} \Rightarrow z_{0,5} = 0$

$$\begin{aligned} &\;\; \| \\ &F(0,5) = 0,5 \\ &F(0) = 0,5 \\ &\;\; \| \\ &F(z \le 0) = 0,5 \\ &z \sim N(0,1) \end{aligned}$$

$z_{0,5}$ = mediană

Cum ne propunem sa demonstram propozitia?

Nr. $z_{1-\alpha}$ este nr. pentru care $\phi(z_{1-\alpha}) = 1-\alpha$ $\qquad z \sim N(0,1)$

Nr z_α este nr. pentru care $\phi(z_\alpha) = \alpha$

$z_{1-\alpha} = -z_\alpha$ (Vom folosi $\phi(-z_\alpha) = 1 - \phi(z_\alpha)$)

$\Leftrightarrow \quad 1-\alpha = \phi(z_{1-\alpha})$

Transformarea variabilelor aleatoare

$x \xrightarrow{g} g(x) = y$

ex: $X \sim \text{Unif}\,[0,1]$, $Y = e^X$

$$F_X(x) = \begin{cases} 0, & \text{dacă } x < 0 \\ x, & \text{dacă } 0 \le x < 1 \\ 1, & \text{dacă } x \ge 1 \end{cases}$$

$F_Y(y) = P(Y \le y)$

$X \sim \text{Unif}\,[0,1] \Rightarrow b_X = [0,1]$

$Y = e^X \Rightarrow b_Y = [1,e]$

Vom determina $F_Y(y)$

$y < 1 \Rightarrow F_Y(y) = 0$

$y > e \Rightarrow F_Y(y) = P(Y \le y) = 1$

$y \in [1, e] \Rightarrow F_Y(y) = P(Y \le y) = P(e^X \le y) = P(X \le \ln y) =$

$= F_X(\ln y) = \ln y$

$$F_Y(y) = \begin{cases} 0, & \text{dacă} \quad y < 1 \\ In(y), & \text{dacă} \quad 1 \le y < e \\ 1, & \text{dacă} \quad y \ge e \end{cases}$$

$f_Y(y) = F'_Y(y)$

$M(Y) = M(e^X)$

LOTUS $\int_{-\infty}^{\infty} e^x f(x)\,dx = \int_0^1 e^x \cdot 1\,dx = e^x \Big|_0^1 = e - 1$

$$f_Y(y) = F'_Y(y) = \begin{cases} \frac{1}{y}, & \text{dacă} \quad 1 \le y \le e \\ 0, & \text{altfel} \end{cases}$$

Obs: Prin procedeul de standardizare, transformam o v.a. x $y \sim N(m, \sigma^2)$ intr-o v.a. $Z \sim N(0, 1)$

Proprietate

Fie $X \sim N(m, \sigma^2)$. Atunci, $Z = \frac{X-m}{\sigma}$ are proprietatea că $Z \sim (0, 1)$.

Intr-adevăr, se poate demonstra că

$$f_X(x) = \frac{1}{\sigma}\varphi_Z\left(\frac{x-m}{\sigma}\right), \quad F_X(x) = \Phi\left(\frac{x-m}{\sigma}\right)$$

- $M\left(\frac{X-m}{\sigma}\right) \stackrel{?}{=} 0$

$$\frac{M(X) - m}{\sigma} = \frac{m - m}{\sigma} = 0$$

- $\sigma^2(aX) = a^2\sigma^2(X)$

$$P(a \le X \le b) = P(a < X \le b) = F_X(b) - F_X(a) = ?$$
$$P(a \le X \le b) = P\left(\frac{a-m}{\sigma} \le Z = \frac{X-m}{\sigma} \le \frac{b-m}{\sigma}\right) = \Phi\left(\frac{b-m}{\sigma}\right) - \Phi\left(\frac{a-m}{\sigma}\right)$$

Exemplu:
Fie $X \sim N(1, 0.4^2)$. Să se calculeze: $P(0.75 < X \leq 1.3)$ și să se determine $x \in \mathbb{R}$ astfel încât $P(X \leq x) = 0.95$?

Rezolvare: $P(0.75 < X \leq 1.3) = F_X(1.3) - F_X(0.75) = ?$
Vom standardiza variabila X, transformând-o în variabila $Z = \frac{X-m}{\sigma}$.
Avem: $P(0.75 < X \leq 1.3) = P(\frac{0.75-1}{0.4} < Z \leq \frac{1.3-1}{0.4}) =$
$\Phi((1.3-1)/0.4) - \Phi((0.75-1)/0.4) = \Phi(0.75) - \Phi(-0.625) =$
$\Phi(0.75) - 1 + \Phi(0.625)$.
Avem de calculat $P(X \leq x) = F_X(x) = \Phi((x-1)/0.4) = 0.95$.
Deci $(x-1)/0.4 = \Phi^{-1}(0.95) = z_{0.95}$. Din tabele se știe că 0.95–cvantila distribuției normale standard este $z_{0.95} = 1.64$.
Deci $x = 1 + 0.4 \cdot 1.64$.