MA – curs 4

CAP. 3. Legi (distribuții, repartiții) probabilistice clasice

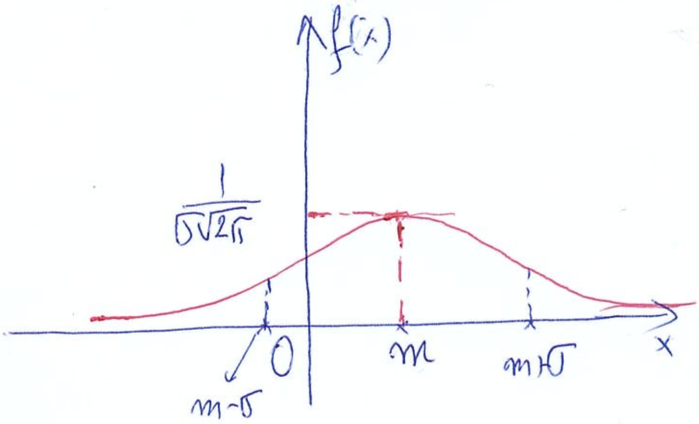
§1. Legea lui Gauss (Distribuția (repartiția) normală)

1. Def. Fie $m \in \mathbb{R}$ și $\sigma > 0$ numere (reale) date. O variabilă aleatoare X are o distribuție (repartiție) normală sau urmează legea lui Gauss de parametri m și σ^2 dacă p.d.f. (X) este

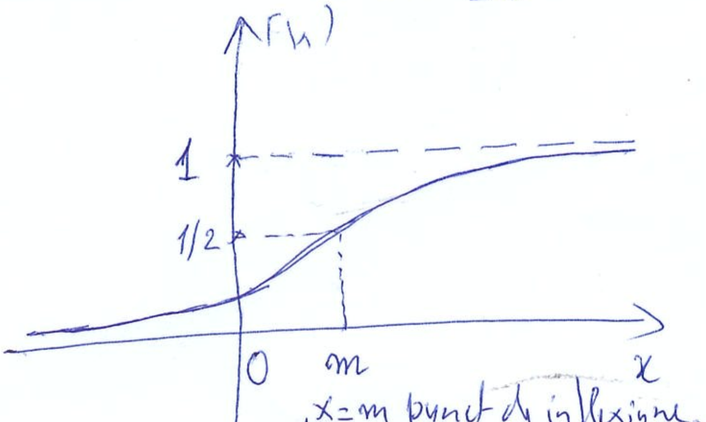
$$f: \mathbb{R} \to \mathbb{R}, \quad f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right], \quad x \in \mathbb{R}$$

Notație $X = N(m; \sigma^2)$

2. Funcția de repartiție $F: \mathbb{R} \to [0,1]$, $F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-m)^2}{2\sigma^2}} dt$



Pdf(X). $x = m \pm \sigma$ puncte de inflexiune.



$x = m$ punct de inflexiune.

Funcția de rep. $F(x) = P(X < x)$

3. Funcția caracteristică $\varphi(t) = M(e^{jtX}) = E(e^{jtX}) = \exp\left(jtm - \frac{t^2\sigma^2}{2}\right)$; $t \in$

4. Valoarea medie $\boxed{M(X) = E(X) = m}$

Dem. $M(X) = \int_{-\infty}^{\infty} x f(x)\,dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x\, e^{-\frac{(x-m)^2}{2\sigma^2}} dx =$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x \cdot e^{-\left(\frac{x-m}{\sigma\sqrt{2}}\right)^2} dx \underset{x = m + t\sigma\sqrt{2}}{\overset{\frac{x-m}{\sigma\sqrt{2}} = t}{=}} \frac{1}{\cancel{\sigma}\sqrt{2\pi}} \int_{-\infty}^{\infty} (m + t\sigma\sqrt{2})\, e^{-t^2} \cdot \sigma\sqrt{2}\,dt$$

$$M(X)=\frac{\cancel{\sigma}\sqrt{2}}{\cancel{\sigma}\sqrt{2\pi}}\Big[m\underbrace{\int_{-\infty}^{\infty}e^{-t^2}dt}_{\sqrt{\pi}\ (\text{Gauss})}+\sigma\sqrt{2}\underbrace{\int_{-\infty}^{\infty}te^{-t^2}dt}_{=0\ (\text{fct. impară})}\Big]$$

$$M(X)=\frac{1}{\sqrt{\pi}}(m\sqrt{\pi}+0)=m.$$

5. Dispersia $\boxed{D^2(X)=Var(X)=\sigma^2}$

$$D^2(X)=M(X^2)-M^2(X) \quad (1)$$

$$M(X^2)=\int_{-\infty}^{\infty}x^2 f(x)\,dx=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}x^2\cdot e^{-\left(\frac{x-m}{\sigma\sqrt{2}}\right)^2}dx \qquad \frac{\frac{x-m}{\sigma\sqrt{2}}=t}{x=m+t\sigma\sqrt{2}}$$

$$=\frac{1}{\cancel{\sigma}\sqrt{2\pi}}\int_{-\infty}^{\infty}(m+t\sigma\sqrt{2})^2e^{-t^2}\cancel{dt}\cdot\cancel{\sigma}\sqrt{2}dt=\frac{1}{\sqrt{\pi}}\Big[m^2\underbrace{\int_{-\infty}^{\infty}e^{-t^2}dt}_{\sqrt{\pi}}+2m\sigma\sqrt{2}\underbrace{\int_{-\infty}^{\infty}te^{-t^2}dt}_{=0}+$$

$$+2\sigma^2\underbrace{\int_{-\infty}^{\infty}t^2e^{-t^2}dt}\Big]=\frac{1}{\sqrt{\pi}}\Big(m^2\sqrt{\pi}+2\sigma^2\cdot 2\underbrace{\int_{0}^{\infty}t^2e^{-t^2}dt}_{t=\sqrt{\tau}}\Big)=\frac{1}{\sqrt{\pi}}\Big(m^2\sqrt{\pi}+$$

$$+4\sigma^2\int_0^\infty \tau e^{-\tau}\frac{1}{2\sqrt{\tau}}d\tau\Big)=\frac{1}{\sqrt{\pi}}\Big(m^2\sqrt{\pi}+2\sigma^2\int_0^\infty\tau^{1/2}e^{-\tau}d\tau\Big)=$$

$$=\frac{1}{\sqrt{\pi}}\Big[m^2\sqrt{\pi}+2\sigma^2\mathcal{L}\{\tau^{1/2}\}(1)\Big]=\frac{1}{\sqrt{\pi}}\Big(m^2\sqrt{\pi}+2\sigma^2\cdot\frac{\Gamma(3/2)}{1^{3/2}}\Big) \qquad \Gamma\left(\tfrac{3}{2}\right)=\tfrac{1}{2}\sqrt{\pi}$$

$$=\frac{1}{\sqrt{\pi}}\Big(m^2\sqrt{\pi}+2\sigma^2\cdot\frac{1}{2}\sqrt{\pi}\Big)=m^2+\sigma^2 \quad (2)$$

Din (1), (2) $\Rightarrow D^2(X)=m^2+\sigma^2-m^2=\sigma^2$.

6. Alte caracteristici statistice (numerice)

$Mo(X)=m$; $Me(X)=m$; $As(X)=0$; $Ex(X)=0$.

OBS. $\mu_n(X)=M\big((X-m)^n\big)=\begin{cases}0; & n \text{ impar}\\ \dfrac{n!\,\sigma^n}{2^{n/2}(n/2)!}; & n \text{ par}\end{cases}$

§2. Legea normală redusă. Funcția lui Laplace

1. Def. Se numește variabilă normală redusă v.a. $X=N(0;1)$; $\begin{cases}m=0\\ \sigma=1\end{cases}$

$$f(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$$

FUNCȚIA ERORILOR (GAUSS)

OBS. Funcția $erf:[0,\infty)\to\mathbb{R}$, $erf(x)=\frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2/2}dt$ se numește

$erf(0)=0$; $erf(\infty)=1$.

2. Funcția lui Laplace. Este funcția de repartiție a legii normale reduse $X = N(0;1)$

Notație: $\boxed{\phi(x)} = P(N(0;1) < x)$; $\phi : \mathbb{R} \to [0,1]$

Formule de calcul: $\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{x/\sqrt{2}} e^{-t^2} dt$; $x \in \mathbb{R}$

3. Proprietățile funcției lui Laplace

(i) $\phi(-x) = 1 - \phi(x)$; $x \in \mathbb{R}$

(ii) $X = N(m;\sigma^2) \Longrightarrow P(a \le X \le b) = \phi\left(\frac{b-m}{\sigma}\right) - \phi\left(\frac{a-m}{\sigma}\right)$, $\forall a, b \in \mathbb{R}$; $a \le b$.

(iii) $X = N(m;\sigma^2)$; $\alpha > 0 \Rightarrow P(|X-m| \le \alpha) = 2\phi\left(\frac{\alpha}{\sigma}\right) - 1$.

4. Legea celor trei sigma

$$X = N(m;\sigma^2) \Longrightarrow \begin{cases} P(|X-m| \le 3\sigma) = 2\phi\left(\frac{3\sigma}{\sigma}\right) - 1 = 2\phi(3) - 1 \simeq 0.9974 \\ P(X \in [m-3\sigma; m+3\sigma]) \simeq 0.9974 = 99,74\% \end{cases}$$

Minimum 99% din valorile unei v.a. $N(m;\sigma^2)$ se află în intervalul $[m-3\sigma, m+3\sigma]$

§3. Teorema limită centrală

Fie $(X_n)_{n\ge 1}$ un șir de v.a. independente, de medii și dispersii egale, i.e. $M(X_n) = m$ și $D^2(X_n) = \sigma^2 \ne 0$ ($\Leftrightarrow \sigma(X_n) = \sigma$), $\forall n \ge 1$. Are loc egalitatea:

$$(1)\quad \lim_{n\to\infty} P\left(\frac{X_1 + X_2 + \dots + X_n - mn}{\sigma\sqrt{n}} < x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt, \quad \forall x \in \mathbb{R}.$$

OBS. Fie $Y_n = \frac{X_1 + X_2 + X_3 + \dots + X_n - mn}{\sigma\sqrt{n}} \Rightarrow \boxed{Y_n = \frac{S_n - M(S_n)}{\sigma(S_n)}; S_n = \sum_{k=1}^{n} X_k}$

$(1) \Longleftrightarrow \lim_{n\to\infty} F_{Y_n}(x) = \phi(x)$, $\forall x \in \mathbb{R} \Longleftrightarrow \lim_{n\to\infty} F_{Y_n}(x) = F_{N(0;1)}(x)$

Deci (i) $Y_n \longrightarrow N(0;1)$ (în sens probabilistic)

(ii) Pentru n suficient de mare, $\boxed{Y_n \simeq N(0;1)}$

OBS $S_n = \sum_{k=1}^{n} X_k \Rightarrow M(S_n) = \sum_{k=1}^{n} M(X_k) = mn$; $D^2(S_n) = \sum_{k=1}^{n} D^2(X_k) = \sigma^2 n \Rightarrow \sigma(S_n) = \sigma\sqrt{n}$

Aplicație (Teoria așteptării) În teoria așteptării, intervalele de timp dintre două evenimente succesive (de exemplu sosirea clienților într-un sistem de servire) sunt descrise de v.a. independente care urmează o lege de probabilitate exponențial-negativă, de medie m (constantă). Să se determine probabilitatea ca evenimentul de rang 2500 să apară în intervalul $[2450m, 2600m]$.

Rezolvare Fie $X_n, n \geq 1$, v.a. asociate intervalelor de timp dintre două evenimente succesive. Notăm $S_n = X_1 + X_2 + X_3 + \dots + X_n \Rightarrow S_n$ este v.a asociată timpului de apariție al evenimentului de rang n.

Din datele problemei avem $M(X_n) = m$. Deoarece X_n este o v.a. de tip exponențial-negativ $\Longrightarrow D^2(X_n) = M^2(X_n) = m^2 \Rightarrow \underline{\sigma(X_n) = m}$.

Mai departe $M(S_n) = M\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} M(X_k) = mn$ și

$D^2(S_n) = D^2\left(\sum_{k=1}^{n} X_k\right) \overset{X_k \text{ v.a. indep.}}{=} \sum_{k=1}^{n} D^2(X_k) = m^2 n$, deci

$M(S_n) = mn$ și $\sigma(S_n) = m\sqrt{n}$

Aplicăm Teorema limită centrală. Fie $Y_n = \dfrac{X_1 + X_2 + \dots + X_n - M(X_n) \cdot n}{\sigma(X_n) \cdot \sqrt{n}} \Rightarrow$

$$\Longrightarrow \quad (2) \quad Y_n = \frac{S_n - mn}{\sigma\sqrt{n}} = \frac{S_n - M(S_n)}{\sigma(S_n)} \text{ și } Y_n \simeq N(0;1)$$

$\boxed{\text{În problema dată se cere să se calculeze } P(2450m \leq S_{2500} \leq 2600m)}$

Din (2) $\Rightarrow S_n = mn + \sigma\sqrt{n} \cdot Y_n = mn + m\sqrt{n}\, Y_n$ (3)

Deci $P(2450m \leq S_{2500} \leq 2600m) = P(2450m \leq 2500m + 50m Y_{2500} \leq 2600m)$

$= P(-50m \leq 50m Y_{2500} \leq 100m) = P(-1 \leq Y_{2500} \leq 2) = \phi(2) - \phi(-1) =$

$\underline{\phi(-x) = 1 - \phi(x)}$ $\phi(2) - (1 - \phi(1)) = \phi(2) + \phi(1) - 1 \simeq 0.9772 + 0.8413 - 1 =$

$= 0.8285 = 82,85\%$

§4. Legea lui Bernoulli (Distribuția sau repartiția binomială)

1. Definiție Fie n, p, q numere reale a.î. $n \in \mathbb{N}^*$, $0 \le p \le 1$, $0 \le q \le 1$, $p+q=1$. O v.a. discretă simplă X are o distribuție (repartiție) binomială sau urmează legea lui Bernoulli de parametri n și p dacă tabloul său de distribuție (repartiție) este:

$$X: \begin{pmatrix} 0 & 1 & 2 & 3 & \dots & n \\ P(n;0) & P(n;1) & P(n;2) & P(n;3) & \dots & P(n;n) \end{pmatrix} = \begin{pmatrix} k \\ P(n;k) \end{pmatrix}_{0 \le k \le n},$$

unde $P(n;k) = C_n^k p^k q^{n-k}$

Notație $X = B(p;n)$

2. Funcția de repartiție $F_n(x) = P(X<x) = \begin{cases} 0; & x \le 0 \\ \sum_{i=0}^{k} P(n;i); & \begin{cases} k < x \le k+1 \\ 0 \le k \le n-1 \end{cases} \\ 1; & x > n \end{cases}$

3. Funcția caracteristică $\varphi_n: \mathbb{R} \to \mathbb{C}$; $\varphi_n(t) = M(e^{itX}) = (pe^{it}+q)^n$; $t \in \mathbb{R}$

4. Valoarea medie $M(X) = E(X) = np$

Dem: $M(X) = \sum_{k=0}^{n} k P(n;k)$ (1)

Avem: $(q+px)^n = \sum_{k=0}^{n} C_n^k p^k q^{n-k} x^k \Rightarrow ()'$

$n(q+px)^{n-1} \cdot p = \sum_{k=0}^{n} C_n^k p^k q^{n-k} \cdot k x^{k-1}$. Luăm $x=1$

$n(q+p)^{n-1} \cdot p = \sum_{k=0}^{n} k \dots \xLeftrightarrow{p+q=1} np = \sum_{k=0}^{n} k \cdot P(n;k)$ (2)

Din (1) + (2) $\Rightarrow$ $\boxed{M(X) = np}$

5. Dispersia (Varianța). Abaterea medie pătratică

$X = B(p;n) \Longrightarrow \boxed{D^2(X) = Var(X) = npq}$, $\boxed{\sigma(X) = \sqrt{npq}}$

6. Moda $Mo(X) = \begin{cases} 1 + [np-q]; & 0 \le p < 1 \text{ și } np-q \in \mathbb{N} \\ np-q \text{ sau } np-q+1; & np-q \in \mathbb{N} \setminus \{n\} \\ 1 & ; p=1 \end{cases}$

7. Asimetria, Excesul $As(X) = \frac{q-p}{\sqrt{npq}}$; $Ex(X) = \frac{1-6p+6p^2}{npq}$

8. Asimilarea (aproximarea) legii lui Bernoulli cu legea normală redusă

Fie $X = B(p; n)$; $m = np$; $\sigma = \sqrt{npq}$

(i) V.a. $Y = \frac{X - np + 0.5}{\sqrt{npq}} = \frac{X - np + 0.5}{\sigma}$ este asimptotic normală $N(0,1)$

Practic pt. $n \geq 50$ și $np \geq 18$, $Y \simeq N(0,1)$

(ii) $m = np$; $\sigma = \sqrt{npq}$; $n \geq 50$; $np \geq 18$; $a, b \in \mathbb{R}$, $a \leq b \Rightarrow$

$$\Rightarrow \quad \boxed{P(a \leq X \leq b) \simeq \phi\left(\frac{b-m}{\sigma}\right) - \phi\left(\frac{a-m}{\sigma}\right)}$$

9. Aplicație (Problemă de sondaj)

Din analize statistice, s-a constatat că 4% din calculatoarele produse de o firmă de profil prezintă defecțiuni. Un magazin comercializează 15000 de calculatoare de acest tip.

1° Să se determine probabilitatea ca în magazin să se afle:
(i) cel mult 630 calculatoare cu defecțiuni
(ii) un număr de calculatoare fără defecțiuni cuprins între 14340 și 14436

2° Să se determine cu o eroare de maximum 3% numărul de calculatoare cu defecțiuni din magazin.

3° Să se determine cu probabilitate de 98% (deci cu o eroare de 2%) numărul de calculatoare fără defecțiuni din magazin.

Rezolvare Fie X v.a. ale cărei valori reprezintă numărul de calculatoare fără defecțiuni din magazin. $\Rightarrow$ X este de tip Bernoulli

Avem $n = 15000$, $p = 96\% = \frac{96}{100} = \frac{24}{25}$; $q = 4\% = \frac{4}{100} = \frac{1}{25} \Rightarrow$

$\Rightarrow m = np = \cancel{15000} \cdot \frac{96}{\cancel{100}} = 14400$; $\sigma = \sqrt{npq} = \sqrt{\cancel{15000} \cdot \frac{96}{\cancel{100}} \cdot \frac{4}{\cancel{100}}} =$

$= \sqrt{\frac{60 \cdot 96}{10}} = \sqrt{6 \cdot 6 \cdot 16} = 6 \cdot 4 = 24 \Rightarrow m = 14400$; $\sigma = 24$.

1° (i) $P(\overline{X} \le 630) = P(X \ge 15000 - 630) = P(X \ge 14370) =$

$\boxed{X + \overline{X} = 15000}$ $= P(14370 \le X < \infty) \simeq \phi\left(\frac{\infty - m}{\sigma}\right) - \phi\left(\frac{14370 - m}{\sigma}\right)$

$= \phi(\infty) - \phi\left(\frac{14370 - 14400}{24}\right) = 1 - \phi\left(\frac{-30}{24}\right) = 1 - \phi(-1.25) =$

$= 1 - (1 - \phi(1.25)) = \phi(1.25) \simeq 0.8944 = 89{,}44\%$; $\underline{\phi(-x) = 1 - \phi(x)}$

(ii) $P(13740 \le X \le 14036) = \phi\left(\frac{14436 - m}{\sigma}\right) - \phi\left(\frac{14340 - m}{\sigma}\right) =$

$= \phi\left(\frac{36}{60}\right) - \phi\left(\frac{-60}{60}\right) = \phi\left(\frac{3}{5}\right) - \phi(-1) = \phi(0.6) - (1 - \phi(1)) =$

$= \phi(0.6) + \phi(1) - 1 \simeq 0.6406 + 0.8413 - 1 = 0.4819 = 48{,}19\%$

2° $\underline{P(|X - m| < \alpha) = 2\phi\left(\frac{\alpha}{\sigma}\right) - 1} \Leftrightarrow \underline{P(X \in (m - \alpha, m + \alpha)) = 2\phi\left(\frac{\alpha}{\sigma}\right) - 1}$

$P(X \in (m - \alpha, m + \alpha)) = 97\% = 0.97 \Rightarrow 2\phi\left(\frac{\alpha}{\sigma}\right) - 1 = 0.97 \Rightarrow$

$\Rightarrow \phi\left(\frac{\alpha}{\sigma}\right) = 0.985 \Rightarrow \frac{\alpha}{24} = \phi^{-1}(0.985) \simeq 2{,}17 \Rightarrow$

$\Rightarrow \alpha \simeq 24 \cdot 2{,}17 = 52.08 \Rightarrow X \in (m - \alpha, m + \alpha) \Rightarrow$

$\Rightarrow X \in (14400 - 52{,}08 ; 14400 + 52{,}08) \Rightarrow \underline{X \in [14347 ; 14453)}$

3° $P(|X - m| < \alpha) = 2\phi\left(\frac{\alpha}{\sigma}\right) - 1 = 98\% = 0.98 \Rightarrow$

$\Rightarrow P(X \in (m - \alpha, m + \alpha)) = 2\phi\left(\frac{\alpha}{\sigma}\right) - 1 = 0{,}98 \Rightarrow \phi\left(\frac{\alpha}{\sigma}\right) = 0.99$

$\Rightarrow \phi\left(\frac{\alpha}{24}\right) = 0.99 \Rightarrow \frac{\alpha}{24} = \phi^{-1}(0.99) = 2.33 \Rightarrow \alpha = 24 \cdot 2{,}33 \Rightarrow$

$\Rightarrow \alpha = 55{,}92 \Rightarrow X \in (14400 - 55{,}92 , 14400 + 55{,}92) \Rightarrow$

$\Rightarrow X \in [14344 ; 14456) \Rightarrow \overline{X} = 15000 - X \in [544 ; 656]$