① Să se demonstreze că în repartiția normală constanta trebuie să fie $\frac{1}{\sqrt{2\pi}}$.

Adică,

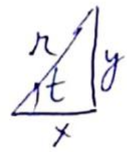
$$I = \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,dx = \sqrt{2\pi}$$

Indicație: să se scrie I^2 ca integrală dublă în coordonate polare.

Deci,

$$I^2 = \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,dx \cdot \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}}\,dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} \cdot e^{-\frac{y^2}{2}}\,dx\,dy =$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}}\,dx\,dy\,;$$



Trecere în coordonate polare: $\begin{cases} x = r\cos t & t \in [0, 2\pi) \\ y = r\sin t & r > 0. \\ dx\,dy = J \cdot dr\,dt = r\,dr\,dt \end{cases}$

$$\Rightarrow I^2 = \int_0^{\infty}\int_0^{2\pi} e^{-\frac{r^2}{2}} \cdot r\,dr\,dt = \int_0^{\infty} r \cdot e^{-\frac{r^2}{2}} \cdot t\Big|_0^{2\pi}\,dr = 2\pi\int_0^{\infty} r \cdot e^{-\frac{r^2}{2}}\,dr =$$

$$= -2\pi e^{-\frac{r^2}{2}}\Big|_0^{\infty} = -2\pi(0-1) = 2\pi.$$

$$\Rightarrow I = \sqrt{2\pi} \quad \text{q.e.d.}$$

Fie X - variabilă al. continuă cu:

mean - μ

variance - σ^2 și

$X \sim N(\mu, \sigma^2)$, atunci:

PDF: $f_X(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} e^{-\left(\frac{x-\mu}{\sigma}\right)^2}$,

CDF: $F_X(x) = P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{u^2}{2}} du =$

$= \Phi\left(\frac{x-\mu}{\sigma}\right)$.

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

Valoarea pasată funcției Φ: „la câte variații standard față de medie"

Faptul că $\Phi(4+) = 0 \Rightarrow$ valorile aflate la mai mult de 4 variații standard față de medie au o probabilitate infimă, care poate fi neglijată.

Law of large numbers

Fie $X_1, X_2, \dots, X_n$ n variabile aleatoare distribuite independent, fiecare având Expected Value $(X_i) = \mu < \infty$.

Atunci, $(\forall)\ \varepsilon > 0$

$$\lim_{n \to \infty} P(|\bar{X} - \mu| \geq \varepsilon) = 0.$$

Adică, pentru un număr foarte mare $(n \to \infty)$ de variabile aleatoare X_i, probabilitatea ca media aritmetică a valorilor mediilor aritmetice să fie diferită de media aritmetică a uneia (oricare) din variabilele aleatoare X_i tinde spre 0.

Adică, pentru un n destul de mare, media aritm. a mediilor aritmetice va tinde spre μ.

Central Limit Theorem

Fie $X_1, X_2, \dots, X_n$ – var. aleat. i.i.d. cu $EX_i = \mu < \infty$ și $Var(X_i) = \sigma^2 < \infty$.

Deci, $\overline{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$ are $E\overline{X} = \mu$. (1)

Dem. (1)

$$\boxed{E\overline{X} = (EX_1 + EX_2 + \dots + EX_n) \cdot \frac{1}{n} \overset{i.i.d}{=} \frac{\cancel{n} EX_i}{\cancel{n}} = EX_i}$$

q.e.d.

și $Var(\overline{X}) = \frac{\sigma^2}{n}$. (2)

Dem. (2)

$$Var(\overline{X}) = Var\left(\frac{(X_1 + X_2 + \dots + X_n)}{n}\right) \underset{\alpha^2 Var(X)}{\overset{Var(\alpha X) =}{=}}$$

$$= \frac{1}{n^2} \cdot \left[\frac{Var(X_1) + Var(X_2) + \dots + Var(X_n)}{1}\right] = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Atunci, CLT spune:

variabila aleatoare:

$$Z_n = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} = \frac{\overline{X} - \mu}{\frac{\sigma^2}{\sqrt{n}}} = \frac{X_1 + \dots + X_n - n\mu}{\sigma \cdot \sqrt{n}}$$

converge ca distribuție la standard normal distribution.

Adică pot aprecia CDF astfel:

$$\lim_{n \to \infty} \left(P(Z_n \leq x)\right) = \Phi(x), \ (\forall) x \in \mathbb{R}$$

Cum aplic?

1. Scriu variabila care mă interesează $Y = X_1 + \dots + X_n$
2. Aflu EY și $Var(Y)$ => țin cont că
$EY = n\mu$ și $Var(Y) = n\sigma^2$
3. Concluzionez din CLT că $\frac{Y - EY}{\sqrt{Var(Y)}} = \frac{Y - n\mu}{\sqrt{n}\sigma}$
are distr. stand. normală.

Deci: $P(y_1 < Y \leq y_2) = P\left(\frac{y_1 - n\mu}{\sqrt{n}\sigma} < Y < \frac{y_2 - n\mu}{\sqrt{n}\sigma}\right)$

$$= \Phi\left(\frac{y_2 - n\mu}{\sqrt{n}\sigma}\right) - \Phi\left(\frac{y_1 - n\mu}{\sqrt{n}\sigma}\right)$$

① A bank teller serves customers standing in the queue one by one. Suppose that the service time X_i for customer i has mean $EX = 2$ (min). and $Var(X_i) = 1$ (min). We assume that service time for different customers are independent. Let Y be the total time the bank teller spends serving 50 customers. What is the prob. that it takes between 90 and 110 minutes?

$$P(90 < Y \leq 110) = P\left(\frac{90 - n\mu}{\sigma\sqrt{n}} < Y \leq \frac{110 - n\mu}{\sqrt{n}\,\sigma}\right) =$$

$$= P\left(\frac{90 - 50 \cdot 2}{\sqrt{50}} < Y \leq \frac{110 - 50 \cdot 2}{\sqrt{50}}\right) = P\left(-\frac{10}{\sqrt{50}} < Y \leq \frac{10}{\sqrt{50}}\right) =$$

$$= \Phi\left(\frac{10}{\sqrt{50}}\right) - \Phi\left(-\frac{10}{\sqrt{50}}\right) = 2\,\Phi\left(\frac{10}{\sqrt{50}}\right) - 1 = 2 \cdot \Phi(\underbrace{\sqrt{2}}_{1.41}) - 1$$

$$\overset{1{,}84146}{=\ 2 \cdot 0.92073} - 1 = 0{,}84146 \approx 84\%$$

2. There are 100 men on a plane. Let X_i be the weight (in pounds) of the ith men. X_i - i.i.d. $EX = 170$, $\sigma_{X_i} = 30$. Find the probability that the total weight of the men on the plane exceeds 18,000 pounds.

$W = X_1 + X_2 + \ldots + X_n$, $n = 100$ (pasageri).

$$P(W > 18.000) \overset{\text{normalizez}}{=} P\left(\frac{W - n\mu_W}{\sqrt{n}\,\sigma_W} > \frac{18.000 - n\mu_W}{\sqrt{n}\,\sigma_W}\right) =$$

$$\times\Big(= 1 - P\left(\frac{W - \cancel{n}\mu_W}{\cancel{\sqrt{n}}\,\sigma_W} < \frac{18.000 - \cancel{n}\mu_W}{\cancel{\sqrt{n}}\,\sigma_W}\right) = 1 - P\left(\frac{W - \cancel{n}\mu_W}{\cancel{\sqrt{n}}\,\sigma_W} < \frac{18k - 100 \cdot 170}{100 \cdot 30}\right)$$

$$= 1 - P\left(\frac{W - n\mu_{X_i}}{\sqrt{Var(W)}} < \frac{18.000 - n\mu_{X_i}}{\sqrt{Var(W)}}\right) = 1 - P\left(\frac{W - 100 \cdot 170}{\sqrt{100\,Var(X_i)}} < \frac{18k - n\mu_{X_i}}{\sigma_W}\right)$$

$$= 1 - P\left(\frac{W - 17000}{10 \cdot 30} < \frac{18.000 - 17000}{300}\right) = 1 - \Phi\left(\frac{10}{3}\right) = 1 - 0,99957 \simeq 0.$$

Problem

You have invited 64 guests to a party. You need to make sandwiches for the guests. You believe that a guest might need 0, 1 or 2 sandwiches with probabilities 1/4, 1/2, and 1/4 respectively. You assume that the number of sandwiches each guest needs is independent from other guests. How many sandwiches should you make so that you are 95% sure that there is no shortage?

3. X_i - numărul de sandvișuri al invitatului i

0 → 25%
1 → 50%
2 → 25%

$$EX_i = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1.$$

$$EX_i^2 = 0^2 \cdot \left(\frac{1}{4}\right) + 1^2 \cdot \frac{1}{2} + 2^2 \cdot \frac{1}{4} = \frac{3}{2}.$$

$$Var(X_i) = EX_i^2 - (EX)^2 = \frac{3}{2} - 1 = \frac{1}{2}.$$

$$\Rightarrow \text{dev. st } \sigma_{X_i} = \frac{\sqrt{2}}{2}.$$

Fie $Y = \sum_{i=1}^{64} X_i = X_1 + X_2 + \dots + X_n$.

$$EY = n EX_i = 64 \cdot 1 = 64.$$

$$\cancel{EY^2 =} \; Var(Y) = \sigma_{X_i} \cdot n = \frac{1}{2} \cdot 64 = 32.$$

$$\Rightarrow \sigma_Y = \sqrt{Var(Y)} = \sqrt{32} = 4\sqrt{2}.$$

$$P(Y \leq y) = 0{,}95 \Leftrightarrow P\Big(\frac{Y - \overset{EY}{n\mu_{X_i}}}{\underset{\sigma_Y}{\sqrt{n}\,\sigma_{X_i}}} \leq \frac{y - n\mu_{X_i}}{\sqrt{n}\,\sigma_{X_i}}\Big) = 0{,}95$$

$$\Leftrightarrow P\left(\frac{Y-64}{4\sqrt{2}} \leq \frac{y-64}{4\sqrt{2}}\right) = 0{,}95 \Leftrightarrow \Phi\left(\frac{y-64}{4\sqrt{2}}\right) = 0{,}95 \;\Big|\; \Phi^{-1}$$

$$\Rightarrow \frac{y-64}{4\sqrt{2}} = \Phi^{-1}(0{,}95) \Leftrightarrow \frac{y-64}{4\sqrt{2}} = 1{,}64.$$

cuantila percentila$_{0.95}$

$$\Rightarrow y = 1{,}64 \cdot 4\sqrt{2} + 64$$

$$\Rightarrow y \approx 73{,}25$$

① X - var. aleatoare cu PDF:

$$f_X(x) = \begin{cases} cx^2 & |x| \leq 1 \\ 0 & \text{altfel} \end{cases}$$

a. Aflați c.

b. Expected value EX. și variance $Var(X)$.

c. $P(X \geq \frac{1}{2})$.

a. $\int\limits_{0} f_X(x)dx = 1 \Leftrightarrow \int_{-1}^{1} cx^2 dx = 1 \Leftrightarrow c\frac{x^3}{3}\Big|_{-1}^{1} = 1.$

$\Leftrightarrow \frac{c}{3}(1+1) = 1 \Rightarrow \boxed{c = \frac{3}{2}}$

b. $EX = \int\limits_{0} x f_X(x) dx$

$\Rightarrow EX = \frac{3}{2}\int_{-1}^{1} x \cdot x^2 dx = \frac{3}{2}\frac{x^4}{4}\Big|_{-1}^{1} = \frac{3}{8}(1-1) = 0.$

de așteptat, grafic simetric

media, mediana, moda = momentul 1.

$Var(X) = M_2[(X-\mu_X)^2] = M[(X-\mu_X)^2] = M[(X-EX)^2].$

$= \int_{-\infty}^{\infty}(x-0)^2 f_X(x) dx \Rightarrow \frac{3}{2}\int x^4 dx = \frac{3}{5}$

sau deviația lui X^2 față de $\mu_X^2 = EX^2 - (EX)^2 = \frac{3}{5}.$

(2) X - var. al.

PDF: $f_X(x) = \frac{1}{2} \cdot e^{-|x|}$, $(\forall)\, x \in \mathbb{R}$.

Dacă $Y = X^2$, aflați CDF al lui Y.

R_Y - domeniul valorilor lui Y. $\Rightarrow R_Y = [0, \infty)$.

$D_Y = [0, \infty)$

$\Rightarrow CDF = F_Y(y) \Rightarrow F_Y(y) = P(Y \leq y)$

$\Leftrightarrow F_Y(y) = P(X^2 \leq y)$

$\Leftrightarrow F_Y(y) = P(-\sqrt{y} \leq X \leq \sqrt{y})$

$\Leftrightarrow F_Y(y) = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{2} e^{-|x|} dx$

$\Leftrightarrow F_Y(y) = \frac{1}{2} \cdot 2 \int_0^{\sqrt{y}} e^{-x} dx$

$\Rightarrow F_Y(y) = -e^{-x} \Big|_0^{\sqrt{y}} = -(e^{-\sqrt{y}} - 1) = 1 - e^{-\sqrt{y}}$.

Deci,

$$CDF_Y = \begin{cases} 1 - e^{-\sqrt{y}}, & y \geq 0. \\ 0, & \text{altfel.} \end{cases}$$

Exerciții

1. $P(X<0)$, dacă $X \sim N(-5,4)$.

$\Rightarrow \mu = -5;\ \sigma = \sqrt{4} = 2.$

$\Rightarrow P(X<0) = F_X(0) = \Phi\left(\frac{0-\mu}{\sigma}\right) = \Phi\left(\frac{0+5}{2}\right) = \Phi(2{,}5) =$

$0{,}99379$

$\simeq 99\%$

2. $P(-7 < X \le -3) = F_X(-3) - F_X(-7) = \Phi\left(\frac{-3+5}{2}\right) - \Phi\left(\frac{-7+5}{2}\right) =$

$= \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 = 2 \cdot 0{,}84134 - 1 = 68\%$

3. $P(X>-3 \mid X>-5)$.

$P(X>-3 \mid X>-5) = \dfrac{P(X>-3 \text{ și } X>-5)}{P(X>-5)} = \dfrac{P(X>-3)}{P(X>-5)}$

$= \dfrac{1-P(X<-3)}{1-P(X<-5)} = \dfrac{1-\Phi\left(\frac{-3+5}{2}\right)}{1-\Phi\left(\frac{-5+5}{2}\right)} = (1-\Phi(1)) \cdot 2 = (1-0{,}84134)\cdot 2$

$\simeq 0{,}32$

$= 32\%$