$$Y = (X-\mu)^2$$

$$\mathbb{P}(Y \geq a^2) \leq \frac{E[Y]}{a^2} = \frac{Var(X)}{a^2}$$

Ales $a = k\sigma$

$$\mathbb{P}(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$$

(Ineg Chernoff)

Fie X v.a., $a > 0$, $t > 0$

$$\mathbb{P}(X \geq a) \leq \frac{E[e^{tx}]}{e^{ta}} \quad \forall a, t$$

Curs 14

Inegalităţi

1) $E[XY] \leq \sqrt{E[X^2]E[Y^2]}$: Cauchy-Schwartz

2) φ convexă $E[\varphi(X)] \geq \varphi(E[X])$ Jensen

φ concavă $E[\varphi(X)] \leq \varphi(E[X])$

3) $X > 0$, $a > 0$

$$\mathbb{P}(X > a) \leq \frac{E[X]}{a}$$ Markov

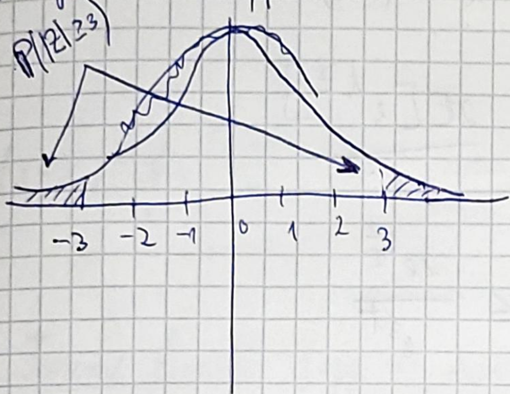
4) X v.a. $E[X] = \mu$, $Var(X) = \sigma^2$ Chebyshev

$$\mathbb{P}(|X-\mu| \geq a) \leq \frac{Var(X)}{a^2}, \forall a > 0$$ (Cebâşev)

5) X v.a., $a > 0$, $t > 0$ atunci $\mathbb{P}(X \geq a) \leq \frac{E[e^{tx}]}{e^{ta}}$, $\forall$ Chernoff

Exp:

$Z \sim N(0,1)$. Vrem să mărginim superior $P(|Z|\geq 3)$ folosind Markov, Chebyshev, Chernoff



Avem prop 68 - 95 - 99.7

$P(|Z|\leq 1) \simeq 0.68$

$P(|Z|\leq 2) \simeq 0.95$

$P(|Z|\leq 3) \simeq 0.997$

Astfel: $P(|Z|\geq 3) \simeq 0.003$

a) Markov

$$P(|Z|\geq 3) \leq \frac{E[|Z|]}{3} \leq \sqrt{\frac{2}{9\pi}}$$

$$tx - \frac{x^2}{2} = x\left(t - \frac{x}{2}\right)$$

$$E[|Z|] = \int |z|\,\varphi(z)\,dz$$

$$= \int |z| \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

$$= 2\int_0^\infty \frac{z}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \sqrt{\frac{2}{\pi}} \int_0^\infty \underbrace{z e^{-\frac{z^2}{2}}}_{\left(-e^{-\frac{z^2}{2}}\right)'} dz = \sqrt{\frac{2}{\pi}} \left(-e^{-\frac{z^2}{2}}\right)\Big|_0^\infty = \sqrt{\frac{2}{\pi}}$$

$$= 0.26$$

b) $$P(|Z|\geq 3) = P(|Z-\underset{\mu}{0}|\geq 3) \leq \frac{Var(Z)}{9} = \frac{1}{9} = 0.11$$

c) Chernoff

$$P(|Z|\geq 3) = 2P(Z\geq 3) \leq \frac{E[e^{tZ}]}{e^{3t}},\ \forall t$$

$$E[e^{tZ}] = \int_{-\infty}^{+\infty} e^{tx} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2} + tx} dx = \frac{1}{\sqrt{2\pi}} \int e^{-\frac{1}{2}(x^2 - 2tx + t^2 - t^2)} dx =$$

$$= \frac{1}{\sqrt{2\pi}} \int e^{-\frac{(x-t)^2}{2} + \frac{t^2}{2}} dx = \frac{e^{\frac{t^2}{2}}}{\sqrt{2\pi}} \int e^{\frac{-(x-t)^2}{2}} dt = e^{\frac{t^2}{2}}$$

Var

$$\to x \sim N(t,1)$$

$$\mathbb{P}(|Z| \geq 3) = 2\,\mathbb{P}(Z \geq 3) \leq \frac{2E[e^{tZ}]}{e^{3t}}$$

$$\mathbb{P}(|Z| \geq 3) \leq 2e^{\frac{t^2}{2} - 3t} \leq \frac{2e^{\frac{t^2}{2}}}{e^{3t}}$$

Pt. $t=3$ avem

$$\mathbb{P}(|Z| \geq 3) \leq 2e^{-\frac{9}{2}} = 0.02$$

Ex: $X \in [a,b]$ $E[X] = \mu$, $Var(X) = \sigma^2$

$$P(|X-\mu| \geq t) \leq \frac{\sigma^2}{t^2}, \ \forall t > 0$$

necunoscut

$$\leq \frac{(b-a)^2}{4t^2}$$

$$\sigma^2 \leq \frac{(b-a)^2}{4}$$

$$g(\gamma) = E[(x-\gamma)^2]$$

$$\gamma = E[X]$$

$$E[(x-\gamma)^2] \geq E[(X-E[X])^2], \ \forall \gamma \in R$$

$$E[(X-\gamma)^2] = E[(X - E[X] + E[X] - \gamma)^2]$$

$$= E[(X-E[X])^2] + 2E[\underbrace{(X-E[X])(E[X]-\gamma)}_{=0}] + (E[X]-\gamma)^2 \geq E[(X-E[X])^2]$$

$$\sigma^2 = Var(X) = E[(X - E[X])^2] \leq E[(X-\gamma)^2], \ \forall \gamma$$

Pt $\gamma = \frac{a+b}{2}$ $\quad E[(X - \frac{a+b}{2})^2] = E[(X-a)(X-b) + \frac{(b-a)^2}{4}]$

$$= E[\underbrace{(X-a)(X-b)}_{\leq 0}] + \frac{(b-a)^2}{4} \leq \frac{(b-a)^4}{4}$$

$$(X-\gamma)^2 = \underset{\leq 0}{A} + \frac{(b-a)^4}{4}$$

Teoreme limită. Legea Nr. mari

Def: Fie $(X_n)_{n\geq 1}$ un șir de v.a. și X o v.a. peste $(\Omega, \mathcal{F}, P)$

Spunem că șirul $(X_n)_n$ conv. la X aproape sigur și notăm

$$X_n \xrightarrow{a.s.} X$$

dacă $P(\underbrace{\lim_n X_n = X}) = 1$

un eveniment

$$A = \{ w \in \Omega \mid \lim_n X_n(w) = X(w) \}$$

Def:

Fie $(X_n)_n$ un șir de v.a. și X o v.a. def. $(\Omega, \mathcal{F}, P)$

Spunem că șirul X_n conv. în prob. la X și notăm

$$X_n \xrightarrow{P} X$$

dacă $\forall \varepsilon > 0$

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

Obs: $\forall \varepsilon > 0, \forall \delta > 0, \ \exists n_0 \in \mathbb{N}$ a.î. $n \geq n_0$

$$P(|X_n - X| \geq \underset{\text{acuratețea}}{\varepsilon}) \leq \delta \leftarrow \text{nivel de încredere}$$

Exp: $X_n \sim \mathcal{U}([0,1])$ indep

$Y_n = \min\{X_1, X_2, \dots, X_n\}$

Ar. $Y_n \xrightarrow{P} 0$

Pentru $\varepsilon > 0$

$$\mathbb{P}(|Y_n - 0| \geq \varepsilon) \xrightarrow[n\to\infty]{} 0$$

$$\mathbb{P}(|Y_n| \geq \varepsilon) = \mathbb{P}(Y_n \geq \varepsilon) \quad (\text{pt că } Y_n \geq 0)$$

$$= \mathbb{P}(X_1 \geq \varepsilon, X_2 \geq \varepsilon, \dots, X_n \leq \varepsilon) =$$

$$\overset{\text{indep}}{=} \mathbb{P}(X_1 \geq \varepsilon)\mathbb{P}(X_2 \geq \varepsilon) \dots \mathbb{P}(X_n \geq \varepsilon)$$

$$= (1 - \mathbb{P}(X_1 < \varepsilon))(1 - \mathbb{P}(X_2 < \varepsilon)) \dots (1 - \mathbb{P}(X_n < \varepsilon)) =$$

$$= (1-\varepsilon) \times (1-\varepsilon) \dots (1-\varepsilon) = (1-\varepsilon)^n$$

pt. $\varepsilon \in [0,1]$

$$\mathbb{P}(|Y_n| \geq \varepsilon) = (1-\varepsilon)^n \text{ pt. } \varepsilon \in (0,1)$$

$\downarrow n\to\infty$

0

Def: Numim eșantion de volum n din populația Q, ~~v.a~~ v.a $X_1, X_2, \dots, X_n$ indep. și identic repartizate (i.i.d) cu ~~repartiție~~ $\mathbb{P}_{X_i} = Q$

Media eșantionului $\overline{X_n} = \dfrac{X_1 + X_2 + \dots + X_n}{n}$

Pp. $X_1, X_2, \dots, X_n$ eșantion de medie μ și varianța σ^2 $(\mathbb{E}[X_i] = \mu, \, Var[X_i] = \sigma^2)$

$$\mathbb{E}[\overline{X_n}] = \mathbb{E}\left[\frac{X_1 + \dots + X_n}{n}\right] = \frac{1}{n}(\mathbb{E}[X_1] + \mathbb{E}[X_2] + \dots + \mathbb{E}[X_n]) = \mu$$

$$Var(\overline{X_n}) = Var\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{1}{n^2}(Var(X_1) + \dots + Var(X_n)) = \frac{\sigma^2}{n}$$

T (LNM) - Legea Nr. Mari (slabă)

Fie $(X_n)_n$ un șir de v.a. i.i.d cu $E[X_1]=\mu<\infty$, $Var[X_1]=\sigma^2<\infty$

Atunci $\overline{X_n} \xrightarrow{P} \mu$

Versiunea Tare

$(X_n)_n$ v.a. i.i.d $E[|X_1|]<\infty$, $E[X_1]=\mu$

$$\overline{X_n} \xrightarrow{a.s.} \mu$$

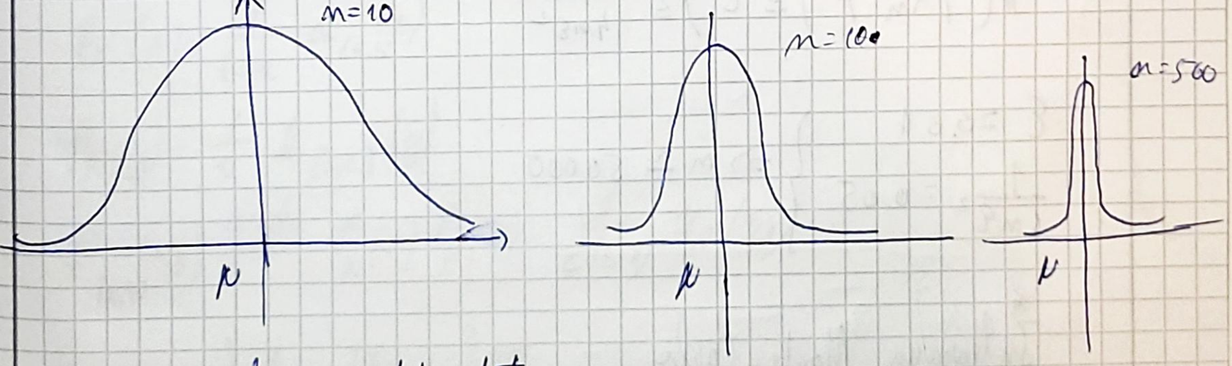
(slabă)

Dem Pt. $\varepsilon>0$

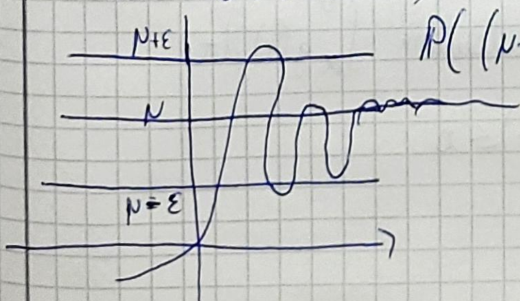
$$P(|\overline{X_n}-\mu|\geq\varepsilon) \to 0$$

Inegalitatea Chebyshev

$$P(|\overline{X_n}-\mu|\geq\varepsilon) \leq \frac{Var(\overline{X_1})}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \xrightarrow[n\to\infty]{} 0$$



Pt. un nivel de acuratețe dat



$$P((\mu-\varepsilon, \mu+\varepsilon) \to \overline{X_n}) \xrightarrow[n\to\infty]{} 1$$

Exp: $(\Omega, \mathcal{F}, \mathbb{P})$ c.p., $A \in \mathcal{F}$

Fie $X_i = \begin{cases} 1, & \omega \in A \\ 0, & \text{altfel} \end{cases}$ $\quad X_i \sim B(p)$

$p = \mathbb{P}(X_i = 1) = \mathbb{P}(A)$

$\overline{X_n} = \dfrac{X_1 + X_2 + \dots + X_n}{n}$ = frecvența relativă de apariție A în n. rep ale exp.

$\overline{X_n} \xrightarrow{\mathbb{P}} E[X_1] = \mathbb{P}(A)$

Exp: Fie p. procentul din pop. care votează cu A.

$X_1, X_2, \dots, X_n \sim B(p)$ indep.

$p(1-p) \leq \frac{1}{4}$

$p - p^2 - \frac{1}{4} \leq 0$

$-p^2 + p - \frac{1}{4} \leq 0$

$\overline{X_n} = \dfrac{X_1 + X_2 + \dots + X_n}{n}$ $\quad$ p - necunoscut

$$\mathbb{P}(|\overline{X_n} - p| \geq \varepsilon) \leq \frac{Var(X_1)}{\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}$$

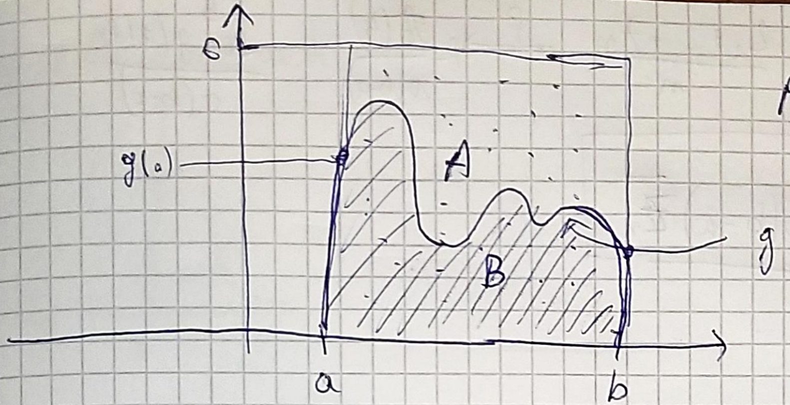
$$\mathbb{P}(|\overline{X_n} - p| \geq \varepsilon) \leq \frac{1}{4n\varepsilon^2}$$

$$\left.\begin{array}{l} \varepsilon = 0.01 \\ \frac{1}{4n\varepsilon^2} = 0.05 \end{array}\right\} \Rightarrow n \simeq 50000$$

Integrarea Monte-Carlo

Pp. că avem o fct g și vrem să calculăm $\int_a^b g(x)\,dx$

Pp. că pe $[a,b]$ avem $0 \leq g(x) \leq c$



$A = [a,b] \times [0,c]$ – dreptunghi

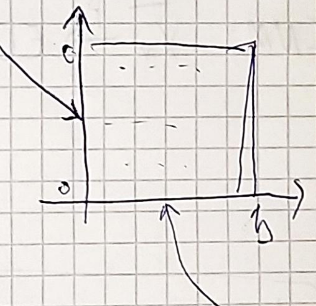
$B = \{(x,y) \mid a \le x \le b,\ 0 \le y \le g(x)\}$

Generăm pct. unif în (A)

$(X_1,Y_1)\dots(X_n,Y_n) \sim \mathcal{U}(A)$

$(X,Y) \sim \mathcal{U}(A)$

$$f_{(X,Y)}(x,y) = \begin{cases} \frac{1}{\mathcal{A}(A)}, & (x,y)\in A \\ 0, & \text{altfel} \end{cases}$$



$\mathcal{A}(A) = c(b-a)$

$X \sim \mathcal{U}([a,b])$ indep.

$Y \sim \mathcal{U}([a,c])$

$f_X(x) = \frac{1}{b-a}\mathbb{1}_{[a,b]}(x)$

$f_Y(y) = \frac{1}{c}\mathbb{1}_{[a,c]}(y)$

$f_{(X,Y)}(x,y) \overset{\text{indep}}{=} f_X(x) f_Y(y) = \frac{1}{c(b-a)}\mathbb{1}_A(x,y)$

Fie $Z_i = \begin{cases} 1, & (X_i,Y_i)\in B \\ 0, & \text{altfel} \end{cases}$, $Z_i \sim B(p)$

$p = \mathbb{P}(Z_i = 1) = \mathbb{P}((X_i,Y_i)\in B)$

$= \iint_B f_{(X,Y)}(x,y)\,dx\,dy = \frac{\mathcal{A}(B)}{\mathcal{A}(A)}$

Din LNM i $\overline{Z_n} = \frac{Z_1 + \dots + Z_n}{n} \xrightarrow{P} p = \frac{\mathcal{A}(B)}{c(b-a)} = \frac{\int_a^b g(x)dx}{c(b-a)}$

$$\boxed{\int_a^b g(x)dx \simeq c(b-a)\overline{Z_n}}$$

Vr 2

Fie $U_1, U_2, \dots, U_n \sim \mathcal{U}[a,b]$ -- ind.

$X_1 = g(U_1),\ X_2 = g(U_2) \dots X_n = g(U_n)$ i.i.d.

LNM: $\overline{X_n} \xrightarrow{P} E[X_1] = E[g(U_1)]$

$$= \int g(x) f_{U_1}(x)dx$$

$$= \int g(x) \frac{1}{b-a} \mathbb{1}_{[a,b]}(x) dx = \frac{1}{b-a}\int_a^b g(x)dx$$

$$\boxed{\int_a^b g(x)dx \simeq (b-a) \cdot \frac{X_1 + \dots + X_n}{n}}$$

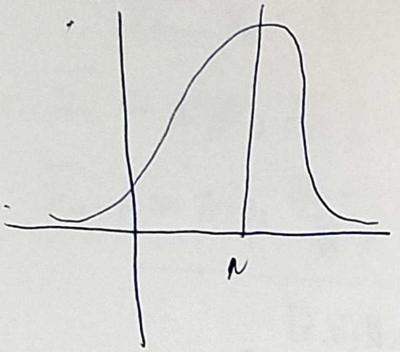
$$\boxed{\int_a^b g(x)dx \simeq (b-a) \cdot \frac{g(U_1) + \dots + g(U_n)}{n} \qquad U_i \sim \mathcal{U}[a,b]}$$

Th Limită Centrală (TLC)

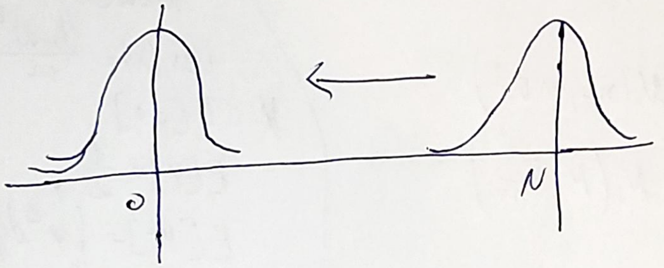
Din LNM $\overline{X_n} \xrightarrow{P} E[X_1]$

Fie $X_1 \dots X_n$ v.a. iid $E[X_1] = \mu$, $Var(X_1) = \sigma^2$

Transf. de locație și de scală



$X - E[X]$



$$Var\left(\frac{X - E[X]}{\sqrt{Var(X)}}\right) = 1$$

$$Z = \frac{X - E[X]}{\sqrt{Var(X)}}$$ – z scor, var. normalizată

$$Z_n = \frac{X_1 + X_2 + \dots + X_n - E[X_1 + \dots + X_n]}{\sqrt{Var(X_1 + \dots + X_n)}}$$

$$= \frac{X_1 + \dots + X_n - n\mu}{\sqrt{n\sigma^2}}$$ – variabila de scor, var. normalizată

$$Z_n = \sqrt{n}\left(\frac{\overline{X}_n - \mu}{\sigma}\right)$$

(T) (Teorema Limită Centrală)

Fie $(X_n)_n$ un șir v.a. i.i.d de medie $E[X_i] = \mu < \infty$, $Var(X_i) = \sigma^2 < \infty$. Atunci

$$\lim_{n} P(Z_n \le x) = \Phi(x), \ \forall x$$

unde $Z_n = \sqrt{n}\,\frac{\overline{X}_n - \mu}{\sigma}$, $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ (fct. de rep. a normalei $\mathcal{N}(0,1)$)

Obs: $X_1, X_2, \dots, X_n$, $E[X_1]=\mu$, $Var(X_1)=\sigma^2$, $S_n = X_1 + \dots + X_n$

$$P(S_n \le c) = P\left(\frac{S_n - E[S_n]}{\sqrt{Var(S_n)}} \le \frac{c - E[S_n]}{\sqrt{Var(S_n)}}\right)$$

$$= P\left(Z_n \le \frac{c - n\mu}{\sqrt{n\sigma^2}}\right) \overset{TLC}{\simeq} \Phi\left(\frac{c - n\mu}{\sigma\sqrt{n}}\right)$$

$X \sim U[a,b]$

$Var(X) = \frac{(b-a)^2}{12}$

$Y \sim U[0,1]$

$E[Y] = \frac{1}{2}$

$E[Y^2] = \int_0^1 y^2 dy = \frac{1}{3}$

$Var(Y) = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$

$X = a + (b-a)Y$

Pt. n suficient de mare

$$\begin{cases} S_n \simeq N(n\mu, n\sigma^2) \\ \bar{X}_n \simeq N\left(\mu, \frac{\sigma^2}{n}\right) \end{cases}$$

Exp: 100 pachete

$U([5,50])$

Care este prob ca greutatea totală ≥ 3000 Kg?

Fie $X_1, \dots, X_{100} \sim U([5,50])$ (greutățile pachetelor)

$S_{100} = X_1 + \dots + X_{100}$ greutatea totală

$$P(S_{100} \ge 3000) = P\bigg(\underbrace{\frac{S_{100} - E[S_{100}]}{\sqrt{Var(S_{100})}}}_{Z_{100} \sim N(0,1)} \ge \frac{3000 - E[S_{100}]}{\sqrt{Var(S_{100})}}\bigg) \overset{TLC}{\simeq} 1 - \Phi\left(\frac{3000 - E[S_{100}]}{\sqrt{Var(S_{100})}}\right)$$

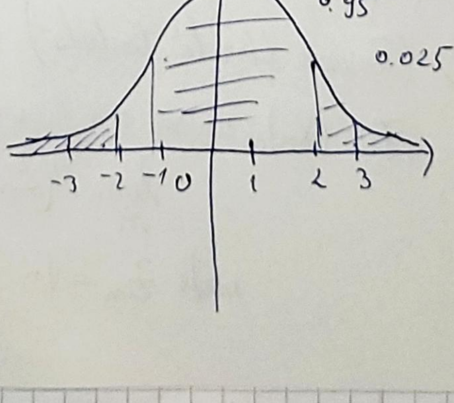
$E[S_{100}] = 100\, E[X_1] = 100 \cdot \frac{55}{2} = 27.5 \cdot 100 = 2750$

$Var(S_{100}) = 100\, Var(S_1) = 100 \cdot \frac{(50-5)^2}{12} = 16875$

$$P(S_{100} \ge 3000) \simeq 1 - \Phi\left(\frac{3000 - 2750}{\sqrt{16875}}\right)$$

$$\simeq 1 - \Phi(1.92) = 0.0274$$

1-pnorm(1.92)



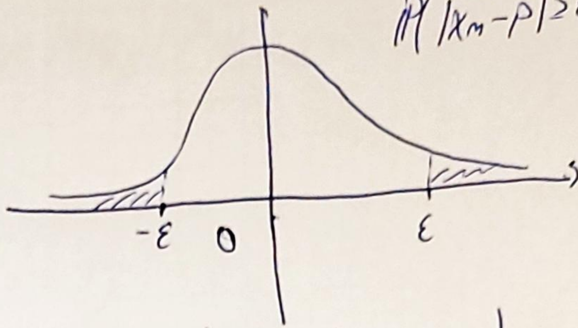
Exp: p pe A în populație

$X_1, \dots X_n \sim B(p)$

TLC: $\overline{X_n} \simeq N(\mu, \frac{\sigma^2}{n}) = N(p, \frac{p(1-p)}{n})$

$\overline{X_n} - p \simeq N(0, \frac{p(1-p)}{n})$

$P(|\overline{X_n} - p| \geq \varepsilon)$



$P(|\overline{X_n} - p| \geq \varepsilon) \simeq 2P(\overline{X_n} - p \geq \varepsilon)$

din simetrie

$\overline{X_n} - p \simeq N(0, \frac{p(1-p)}{n})$

$P(|\overline{X_n} - p| \geq \varepsilon) \simeq 2P(\overline{X_n} - p \geq \varepsilon)$

$\simeq 2P\left(\frac{\overline{X_n} - p}{\sqrt{\frac{p(1-p)}{n}}} \geq \frac{\varepsilon}{\sqrt{\frac{p(1-p)}{n}}}\right)$

$\overset{TLC}{=} 2\left(1 - \Phi\left(\frac{\varepsilon}{\sqrt{\frac{p(1-p)}{n}}}\right)\right) \leq 2(1 - \Phi(2\varepsilon\sqrt{n}))$

$p(1-p) \leq \frac{1}{4} \Rightarrow \frac{p(1-p)}{n} \leq \frac{1}{4n} \Rightarrow \frac{1}{\sqrt{\frac{p(1-p)}{n}}} \geq \frac{1}{\sqrt{\frac{1}{4n}}} \Rightarrow$

$\Rightarrow \frac{\varepsilon}{\sqrt{\frac{p(1-p)}{n}}} \geq 2\varepsilon\sqrt{n}$

$\Rightarrow \Phi\left(\frac{\varepsilon}{\sqrt{\frac{p(1-p)}{n}}}\right) \geq \Phi(2\varepsilon\sqrt{n})$

$P(|\overline{X_n} - p| \geq \varepsilon) \leq \underbrace{2(1 - \Phi(2\varepsilon\sqrt{n}))}_{0.05}$

$\varepsilon = 0.01$

$2(1 - \Phi(2\varepsilon\sqrt{n})) = 0.05$

$\varepsilon = 0.01$

$\Rightarrow \Phi(2 \times 0.01\sqrt{n}) = 0.975$

$2 \times 0.01\sqrt{n} = \underbrace{\Phi^{-1}(0.975)}_{\simeq 2}$

$\sqrt{n} \simeq 100$

$n \simeq 10000$