Cursul # 12

- Teoremă (Legea numerelor mari)

Fie $X_1 \dots X_n$ v.a. $\overset{i.i.d.}{\sim}$ $E[X_1]=\mu$, $Var[X_1]=\sigma^2<+\infty$

Atunci $\overline{S}_n \to \mu$ în probabilitate, i.e.

$$\mathbb{P}(|\overline{S}_n-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \xrightarrow[n\to+\infty]{} 0, \ \forall \varepsilon>0$$

- Observații: Teorema rămâne valabilă și pentru $X_1 \dots X_n$ v.a. cu

i) $\mathbb{E}[X_i] = \mathbb{E}[X_1] := \mu \quad \forall i=\overline{1,n}$

ii) $Var[X_i] \leq \sigma^2 < +\infty \quad \forall i=\overline{1,n}$

iii) $Cov(X_i, X_j) = 0 \quad \forall i \neq j$

- Contraexemplu

Fie $X_1 \dots X_n \overset{i.i.d.}{\sim} Cauchy(x_0, \gamma)$.

Cum $E[X_1]$ nu există, $\overline{S}_n$ nu converge la nici o constantă. În schimb,

$$\overline{S}_n \sim Cauchy(x_0, \gamma)$$

• Exemplu: Precizia sondajelor

Populația României se încrede în capacitatea Facultății de Matematică și Informatică din București de a pregăti tineri informaticieni cu un nivel mediu de încredere $\mu \in [0,1]$ și o varianță de $\sigma^2 = 0.01$. Câți indivizi trebuie să intervievăm pentru a obține un sondaj ce aproximează această valoare μ cu o eroare maximă de $\pm 1\%$ și un nivel de încredere de 95%?

Intervievăm n indivizi:

$X_1, X_2 \dots X_n \in [0,1]$ v.a. i.i.d. cu $E[X_1] = \mu$ este nivelul mediu de încredere <u>teoretic</u> pe care vrem să îl estimăm.

$\overline{S_n} := \frac{X_1 + \dots + X_n}{n}$ este nivelul mediu de încredere empiric pe care îl obținem de la indivizii intervievați.

Inegalitatea Cebâșev

$$\underbrace{\mathbb{P}(|\overline{S_n} - \mu| \geq 0.01)} \leq \frac{\sigma^2}{n(0.01)^2} = \frac{100}{n}$$

Probabilitatea să avem o eroare > 1%

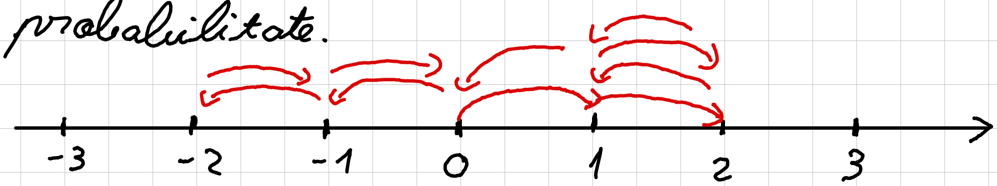
$$\underbrace{\mathbb{P}(|\overline{S_n} - \mu| < 0.01) \geq 1 - \frac{100}{n} > 0.95}$$

Vrem să avem probabilitatea să avem o eroare mai mică de 1% să fie peste 95%, deci alegem n a.î.

$$\frac{100}{n} < 0.05 = \frac{1}{20} \Leftrightarrow \boxed{n > 2000}$$

• Teorema limită centrală: Mersul bețivului

Un bețiv, la ieșirea din crâșmă, face un pas la dreapta cu probabilitatea 50% sau un pas la stânga cu aceeași probabilitate.



$$X_i = \begin{cases} -1, & \text{pasul } i \text{ e făcut la stânga} \\ +1, & \text{pasul } i \text{ e făcut la dreapta} \end{cases}$$

$$X_i \sim \begin{pmatrix} -1 & +1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \quad \text{i.i.d. cu } E[X_i] = 0$$

$$Var[X_i] = 1$$

$S_n := X_1 + X_2 + \ldots + X_n$ ne spune poziția bețivului după n pași.

Ne interesează $S_n \xrightarrow{n \to +\infty} ?$

• Proprietăți

i) $\mathbb{E}[S_n] = n \cdot \mathbb{E}[X_1] = 0$

„În medie, bețivul rămâne în dreptul crâșmei"

ii) $\mathbb{V}ar[S_n] = n \cdot \mathbb{V}ar[X_1] = n$

iii) $\mathbb{P}(X_1 = -1) = \mathbb{P}(X_1 = 1) = \frac{1}{2}$

$$\mathbb{P}(X_2 = -2) = \mathbb{P}(X_2 = -2 \mid X_1 = -1) \cdot \frac{1}{2} + \mathbb{P}(X_2 = -2 \mid X_1 = 1) \cdot \frac{1}{2}$$

$$= \frac{1}{4} = \mathbb{P}(X_2 = 2)$$

$\mathbb{P}(X_2 = 0) = \frac{1}{2}$

...

Putem oare evita aceste calcule, diferite pentru fiecare n?

• Legea numerelor mari ne spune că

$$\overline{S_n} = \frac{X_1 + X_2 + \ldots + X_n}{n} \xrightarrow[n \to +\infty]{} E[X_1] = 0$$

- Dacă aveam

$$\overline{S_n} \xrightarrow{n\to+\infty} \frac{1}{2} \Rightarrow x_1+\dots+x_n \approx \frac{n}{2}$$

dar în cazul nostru avem

$$x_1+x_2+\dots+x_n \approx \underbrace{0\cdot n}_{\text{nedeterminat!}}$$

- Analogie: Ce se întâmplă cu $\sin(x)$ când $x\to 0$?

$$\lim_{x\to 0} \sin(x) = 0$$

$\sin(0.01)\approx 0$, $\sin(0.001)\approx 0$ ... :(

În schimb, dacă folosim $\lim_{x\to 0} \frac{\sin(x)}{x} = 1$

$\sin(0.01)\approx 0.01$, $\sin(0.001)\approx 0.001$ ... :)

- De o astfel de a doua limită avem nevoie și noi pentru a aproxima S_n, și asta ne va da teorema limită centrală.

• Definiție (Convergența în distribuție)

Fie $X_1, \dots X_n \dots$ v. a. reale

Spunem că $X_n \xrightarrow{D} X$ (conv. în distribuție)

dacă $F_n(x) \xrightarrow[n \to +\infty]{} F(x)$ $\forall x$ pct. de continuitate pentru F

unde $F_n(x) = \mathbb{P}(X_n \leq x)$ fct. repartiție X_n

$F(x) = \mathbb{P}(X \leq x)$ fct. repartiție X

• Propoziție

Următoarele afirmații sunt echivalente:

i) $X_n \xrightarrow{D} X$

ii) $\mathbb{E}[f(X_n)] \longrightarrow \mathbb{E}[f(X)]$ $\forall f$ cont. și mărg.

iii) $\varphi_n(t) = \mathbb{E}[e^{itX_n}] \to \varphi(t) = \mathbb{E}[e^{itX}]$

(Convergența funcțiilor caracteristice)

- Teorema limită centrală

Fie $X_1 \dots X_n \dots$ i.i.d. cu $E[X_1] = \mu < +\infty$

$Var[X_1] = \sigma^2 < +\infty$

Atunci

$$\frac{\sqrt{n}}{\sigma}\left(\overline{S}_n - \mu\right) \xRightarrow{D} Z, \text{ unde } Z \sim \mathcal{N}(0,1)$$

Demonstrația se face folosind funcțiile caracteristice.

- Interpretarea:

Pentru n mare, $\frac{\sqrt{n}}{\sigma^2}\left(\overline{S}_n - \mu\right) \overset{\text{aproximativ}}{\sim} \mathcal{N}(0,1)$

- Înapoi la bețivi:

$$\frac{\sqrt{n}}{1}\left(\overline{S}_n - 0\right) \sim \mathcal{N}(0,1)$$

$$\Rightarrow \frac{X_1 + \dots + X_n}{\sqrt{n}} \sim \mathcal{N}(0,1)$$

$$\Rightarrow S_n = X_1 + \dots + X_n \sim \mathcal{N}(0,n)$$

(Poziția bețivului e modelată de $\mathcal{N}(0,n)$)

• Un model mai general: bețivul face pașii conform

$$X_1, \dots X_n \dots \overset{i.i.d.}{\sim} \begin{pmatrix} -1 & 1 \\ 1-p & p \end{pmatrix}$$

$$\mu = 2p - 1 \text{ și } \sigma^2 = 1 - \mu^2 = 4p - 4p^2$$
$$= 4p(1-p)$$

• TLC: $\frac{\sqrt{n}}{\sigma}(\overline{S}_n - \mu) \sim \mathcal{N}(0, 1)$

$$\Rightarrow \frac{S_n}{\sqrt{n}\,\sigma} - \sqrt{n}\,\frac{\mu}{\sigma} \sim \mathcal{N}(0, 1)$$

$$\Rightarrow S_n - n\mu \sim \mathcal{N}(0, n\sigma^2)$$

$$\Rightarrow S_n \sim \mathcal{N}(n\mu, n\sigma^2)$$

• Intuiție TLC: Pentru orice variabile i.i.d de medie μ și varianță $\underline{\sigma^2 < +\infty}$, $X_1 + \dots + X_n \sim \mathcal{N}(n\mu, n\sigma^2)$ pentru n mare