Cursul 13

- Teorema limită centrală

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

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

Atunci

$$\frac{\sqrt{n}}{\sigma}\left(\overline{S_n} - \mu\right) \xRightarrow[n \to +\infty]{D} Z, \text{ unde } Z \sim \mathcal{N}(0,1)$$

- Prin urmare, pentru n mare,

$$\frac{\sqrt{n}}{\sigma}\left(\overline{S_n} - \mu\right) \overset{aprox.}{\sim} \mathcal{N}(0,1) \quad \Big/ \cdot \frac{\sigma}{\sqrt{n}}$$

$$\overline{S_n} - \mu \sim \mathcal{N}\left(0, \frac{\sigma^2}{n}\right) \quad \Big/ \cdot n$$

$$S_n - n\mu \sim \mathcal{N}(0, n\sigma^2) \quad \Big/ + n\mu$$

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

• Exemplu practic

Un pod poate rezista la o greutate de 5000 de tone. Media unei mașini care traversează podul este de 2 tone cu deviație standard de 0.75. Aproximați probabilitatea de colaps a podului când este traversat de 2450 de mașini.

$X_1 \dots X_{2450}$ i.i.d. de medie $\mu = 2$ și

varianță $\sigma^2 = (0.75)^2$

$S_{2450} := X_1 + \dots + X_{2450}$

Ne interesează $\mathbb{P}(S_{2450} > 5000) = ?$

• Dacă dorim doar o margine superioară a probabilității de colaps, folosim Cebâșev

$$\mathbb{P}(S_{2450} \geq 5000) = \mathbb{P}(S_{2450} - 4900 > 100)$$

$$\approx \frac{1}{2} \mathbb{P}(|S_{2450} - 4900| > 100)$$

$$\mathbb{P}(S_{2450} \geq 5000) \leq \frac{1}{2} \frac{\mathbb{V}ar[S_{2450}]}{100^2}$$

$$= \frac{1}{2} \frac{2450}{100^2} \cdot \frac{9}{16}$$

$$\approx 0.069 \quad (6.9\%)$$

(Cam mare!)

- Dar cum vrem să aproximăm probabilitatea, folosim TLC:

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

$$\mathbb{P}(S_{2450} \geq 5000) = \mathbb{P}\left(\overline{S}_{2450} \geq \frac{5000}{2450}\right)$$

$$= \mathbb{P}\left(\overline{S}_{2450} - 2 \geq \frac{100}{49} - 2\right)$$

$$= \mathbb{P}\Bigg(\underbrace{\frac{\sqrt{2450}}{0.75}\left(\overline{S}_{2450} - 2\right)}_{\approx Z \sim \mathcal{N}(0,1)} \geq \underbrace{\frac{\sqrt{2450}}{0.75} \cdot \frac{2}{49}}_{\approx 2.69}\Bigg)$$

$$= \mathbb{P}(Z \geq 2.69) = 1 - \Phi(2.69) = 1 - 0.9964 = 0.0036$$

$$(0.36\%)$$

- Observație:

Exact ca în exemplul precedent, teorema limită centrală poate fi folosită pentru a îmbunătăți acuratețea unei simulări de tip Monte-Carlo.

- Revenim la exemplul „jucărie" cu moneda măsluită din cursul 11:

$X_1 \dots X_n \overset{i.i.d.}{\sim} \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$ reprezintă n aruncări ale monedei măsluite

$$\text{L.N.M.: } \overline{S}_n := \frac{X_1 + \dots + X_n}{n} \longrightarrow E[X_1] = p$$

- Ne interesează $\mathbb{P}(|\overline{S}_n - p| > \varepsilon) \approx ?$

Am putea folosi inegalitatea lui Cebâșev

$$\mathbb{P}(|\overline{S}_n - p| > \varepsilon) \leq \frac{Var[\overline{S}_n]}{\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}$$

dar oare nu putem folosi T.L.C. pentru ceva mai precis?

• Cum $X_1 \dots X_n$ i.i.d. cu $\mu := E[X_1] = p$

$$\sigma^2 := Var[X_1] = p(1-p)$$

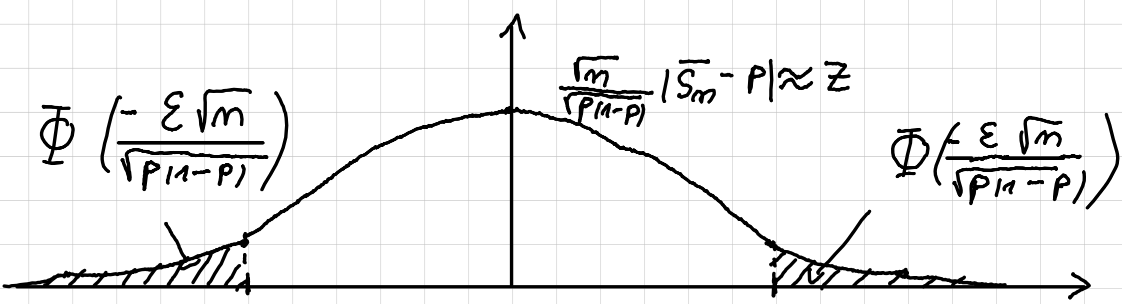
T.L.C.: $\frac{\sqrt{n}}{\sigma}\left(\bar{S}_n - \mu\right) \xrightarrow{D} Z, \quad Z \sim \mathcal{N}(0,1)$

$$\Updownarrow$$

$$\sqrt{\frac{n}{p(1-p)}}\left(\bar{S}_n - p\right) \xrightarrow{D} Z$$

$$\mathbb{P}\left(|\bar{S}_n - p| > \varepsilon\right) = \mathbb{P}\left(\frac{\sqrt{n}}{\sqrt{p(1-p)}}|\bar{S}_n - p| > \frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right)$$

$$\approx 2\,\mathbb{P}\left(Z < \frac{-\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right) = 2\,\Phi\left(\frac{-\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right)$$



• Dacă dorim să avem mai puțin de $\alpha \in (0,1)$ probabilitate să avem o eroare de estimare mai mare ca ε, i.e.

$$\mathbb{P}\left(|\bar{S}_n - p| > \varepsilon\right) < \alpha$$

- Folosind T.L.C.,

$$P(|\overline{S}_n - p| > \varepsilon) \approx 2\,\Phi\left(\frac{-\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right) < \alpha$$

- Cum $p(1-p) \leq \frac{1}{4} \Rightarrow \sqrt{p(1-p)} \leq \frac{1}{2} \Rightarrow$

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

$$\frac{-\varepsilon\sqrt{n}}{\sqrt{p(1-p)}} \leq -2\varepsilon\sqrt{n} \Rightarrow \text{Din } \Phi \text{ crescătoare}$$

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

- Pot impune $2\Phi(-2\varepsilon\sqrt{n}) < \alpha \Leftrightarrow$

$$-2\varepsilon\sqrt{n} < \Phi^{-1}\left(\frac{\alpha}{2}\right) \Leftrightarrow$$

$$\sqrt{n} > -\frac{\Phi^{-1}\left(\frac{\alpha}{2}\right)}{2\varepsilon} \Leftrightarrow$$

$$\boxed{n > \frac{1}{4}\left(\frac{\Phi^{-1}\left(\frac{\alpha}{2}\right)}{\varepsilon}\right)^2}$$

- Pentru $\varepsilon = 0.1$ și nivelul de încredere $95\% \iff \alpha = 0.05$, avem nevoie de

$$n \geq \frac{1}{4\alpha\varepsilon^2} = \frac{1}{4 \cdot \frac{5}{100} \cdot \frac{1}{100}} = \frac{10\,000}{20} = 500.$$

- Cu T. L. C., e de ajuns

$$n \geq \frac{1}{4\varepsilon^2} \cdot \left(\Phi^{-1}(0.025)\right)^2 = 25 \cdot (-2.81)^2 \approx 198$$

- Observație

Inegalitatea Cebâșev rămâne utilă în cazurile când n este prea mic pentru a putea folosi teorema limită centrală.