Curs 16

Teorema limită centrală

$X_1, X_2, \ldots$ iid de medie μ și dispersie $\sigma^2 < \infty$ atunci din LNM avem $\overline{X}_n \xrightarrow{P} \mu$

Definim v.a.

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

$S_n = X_1 + X_2 + \cdots + X_n$

$$Z_n = \frac{S_n - E[S_n]}{\sqrt{Var(S_n)}}$$ – variabila normalizată (z-scor)

$E[Z_n] = 0$ și $Var(Z_n) = 1$

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$ for că $E[S_n] = E[X_1] + \cdots + E[X_n] = n\mu$

$Var(S_n) = Var(X_1) + \cdots + Var(X_n) \underset{\text{indep.}}{=} n\sigma^2$

(T) Fie $X_1, X_2, \ldots$ un șir de variabile aleatoare iid. cu media comună μ și varianța $\sigma^2 < \infty$.

Atunci are loc

$$\lim_{n\to\infty} P(Z_n \le z) = \Phi(z), \forall z \in \mathbb{R}$$

unde $Z_n = \dfrac{S_n - n\mu}{\sigma\sqrt{n}}$ iar $\Phi(z) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{z} e^{-x^2/2}\, dx$

În practică:

Fie $S_n = X_1 + X_2 + \dots + X_n$, cu X_i iid de medie μ și var. σ^2

atunci pt n suficient de mare:

$$\mathbb{P}(S_n \le c) = \mathbb{P}\Big(\underbrace{\frac{S_n - n\mu}{\sigma\sqrt{n}}}_{Z_n \approx N(0,1)} \le \frac{c - n\mu}{\sigma\sqrt{n}}\Big)$$

$$\simeq \mathbb{P}\left(Z \le \frac{c - n\mu}{\sigma\sqrt{n}}\right), \quad Z \sim N(0,1)$$

$$= \Phi\left(\frac{c - n\mu}{\sigma\sqrt{n}}\right) \quad (\text{pnorm}(\cdot))$$

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

TLC: $\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \xrightarrow{d} N(0,1)$ unde "$\xrightarrow{d}$" înseamnă

$$\mathbb{P}\left(\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \le z\right) \longrightarrow \Phi(z), \ \forall z.$$

Exp: Încărcăm un avion cu 100 de pachete a căror greutate este modelată prin intermediul unor v.a. iid repartizate uniform pe $[5, 50]$ kg. Care este probabilitatea ca greutatea totală să depășească 3000 kg?

$X_1, X_2, \dots, X_{100}$ - greutățile pachetelor încărcate în avion

cu $X_i \sim \mathcal{U}([5, 50])$

$S_{100} = X_1 + X_2 + \dots + X_{100}$ - greutatea totală a pachetelor

Vrem să calculăm

$$\mathbb{P}(S_{100} > 3000) = ?$$

$$\mathbb{E}[X_i] = \frac{50+5}{2} = 27.5 \text{ kg}$$

$$Var(X_i) = \frac{(50-5)^2}{12} = \frac{45^2}{12}$$

$$\mathbb{P}(S_{100} > 3000) = \mathbb{P}\left(\frac{S_{100} - 100\mu}{\sigma\sqrt{100}} > \frac{3000 - 100\mu}{\sigma\sqrt{100}}\right)$$

unde $\mu = 27.5$ iar $\sigma = \sqrt{45^2/12} = \frac{45}{2\sqrt{3}}$

Avem

$$\mathbb{P}(S_{100} > 3000) = \mathbb{P}\left(Z_{100} > \frac{3000 - 2750}{\frac{45}{2\sqrt{3}} \cdot 10}\right)$$

Din TLC: $Z_{100} \simeq N(0,1)$

și avem $$\mathbb{P}(S_{100} > 3000) \simeq 1 - \Phi\left(\frac{50\sqrt{3}}{45}\right)$$

$$\simeq 1 - 0.9726 \simeq 0.0274$$

Exp: Un parking al unui imobil nou cu 200 de apartamente este proiectat. Presupunem că nr. de mașini per apartament este de 0, 1 sau 2 cu prob. 0.1, 0.6 și resp. 0.3. Care este nr. minim de locuri de parcare pe care constructorul trebuie să le prevadă dacă vrem să asigurăm cu o prob. de 95% că sunt locuri suficiente pt tot imobilul?

Fie $X \sim \begin{pmatrix} 0 & 1 & 2 \\ 0.1 & 0.6 & 0.3 \end{pmatrix}$ v.a. care corespunde la nr. de mesaje pe zi.

Nr. total de mesaje

$$S_{200} = X_1 + X_2 + \dots + X_{200}$$

Vrem să determinăm y pt care $\mathbb{P}(S_{200} \leq y) \geq 0.95$

Din TLC avem că

$$\frac{S_n - \mathbb{E}[S_n]}{\sqrt{Var(S_n)}} = \frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1)$$

$$\mathbb{P}(S_{200} \leq y) = \mathbb{P}\left(\frac{S_{200} - 200\mu}{\sigma\sqrt{200}} \leq \frac{y - 200\mu}{\sigma\sqrt{200}}\right)$$

unde $\mu = \mathbb{E}[X_1]$ iar $\sigma^2 = Var(X)$

$\mathbb{E}[X] = 1.2 = \mu$

$Var(X) = \sigma^2 = (0^2 \times 0.1 + 1^2 \times 0.6 + 2^2 \times 0.3) - 1.2^2 = 1.8 - 1.44 = 0.36$

$$\mathbb{P}(S_{200} \leq y) = \mathbb{P}\left(\frac{S_{200} - 200 \times 1.2}{\sqrt{200 \times 0.36}} \leq \frac{y - 200 \times 1.2}{\sqrt{200 \times 0.36}}\right)$$

$$\overset{TLC}{\simeq} \Phi\left(\frac{y - 200 \times 1.2}{\sqrt{200 \times 0.36}}\right) = 0.95$$

(qnorm(·))

$$\Rightarrow \frac{y - 240}{6\sqrt{2}} \simeq \Phi^{-1}(0.95) = 1.64$$

$$\Rightarrow y \simeq 240 + 6\sqrt{2} \times 1.64 \simeq 254$$

Aproximarea de Moivre-Laplace a binomialei

Fie $S_n \sim B(n,p)$, S_n poate fi văzută ca o sumă de n v.a. iid Bernoulli de parametru p

$$S_n = X_1 + X_2 + \dots + X_n \quad , X_i \sim B(p) \, , X_i \text{ indep.}$$

$\mu = \mathbb{E}[X_i] = p$

$\sigma^2 = Var(X_i) = p(1-p)$

Vrem să aproximăm $\mathbb{P}(k \le S_n \le l) = ?$

Vom folosi teorema limită centrală:

$$\mathbb{P}(k \le S_n \le l) = \mathbb{P}\left(\frac{k-np}{\sqrt{np(1-p)}} \le \frac{S_n - np}{\sqrt{np(1-p)}} \le \frac{l-np}{\sqrt{np(1-p)}} \right)$$

Din TLC: $\dfrac{S_n - np}{\sqrt{np(1-p)}} \simeq N(0,1)$

$$\mathbb{P}(k \le S_n \le l) \simeq \Phi\left(\frac{l-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k-np}{\sqrt{np(1-p)}}\right)$$

Observăm că pot a calcula $\mathbb{P}(S_n = k)$, $k = l$, obținem $\mathbb{P}(S_n = k) \simeq 0$, ceea ce este fals.

Sugerăm următoarea corecție

$$\mathbb{P}(S_n = k) \simeq \mathbb{P}(k - 1/2 \le S_n \le k + 1/2)$$

Înlocuind în formula de aproximare de mai sus găsim că

$$\boxed{P(k \le S_n \le l) \simeq \Phi\left(\frac{l+1/2-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k-1/2-np}{\sqrt{np(1-p)}}\right)}$$

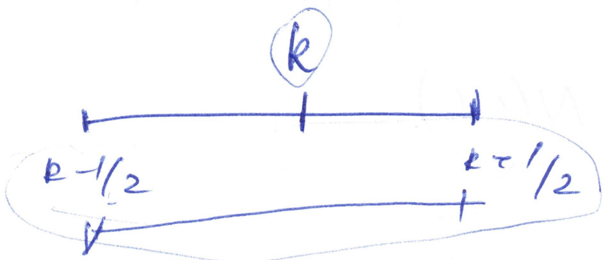
✓ aproximarea normală a binomialei cu factor de corecție

Exp: $S_n \sim B(36, 0.5)$

$P(S_n \le 21) = 0.8785$ (pbinom(·))

aproximăm → $P(S_n \le 21) \simeq \Phi\left(\frac{21-36\times 0.5}{\sqrt{36\times 0.5\times 0.5}}\right) = 0.8413$

$$P(S_n \le 21) \simeq \Phi\left(\frac{21.5-36\times 0.5}{\sqrt{36\times 0.5\times 0.5}}\right) = 0.879$$



Statistică

Def: Numim populație o mulțime de elemente similare care este cercetată din punct de vedere a unei sau mai multor prop.

Elementele populației s.n. indivizi.

Exp: Pb. de prob.: știm că 17% dintre bilele dintr-o urnă, care conține bile albe și negre, sunt albe, extragem 20 de bile cu întoarcere și ne interesează care este prob. să obținem 5 bile negre?

Pb. de statistică: „am extras 20 de bile din urnă (cu întoarcere) și am constatat că 5 sunt de culoare neagră. În ce măsură această informație mă ajută să aproximez (estimez) proporția de bile albe din urnă?

Statistică → frecventistă - parametrul care modelează populația este const. necunoscut

Statistică → bayesiană - parametrul care modelează pop. este v.a.

Def: Numim model statistic o familie de câmpuri de prob. $(\Omega, \mathcal{F}, \mathbb{P}_\theta)$, $\theta \in \Theta \subseteq \mathbb{R}^k$, Θ s.n. sp. parametrilor

Def: Se numește eșantion aleator de volum n (de talie n) un șir de n v.a. independente și identic repartizate. Rep. comună s.n. repartiție mamă (lege mamă) a eșantionului

$X_1, X_2, \dots X_n$ - i.i.d. (eșantion aleator)

Notație: $X_1, X_2, \dots X_n \sim f_\theta$ (sau F_θ) cu $\theta \in \Theta$

- Fie $X_1, X_2, \dots X_n$ un eșantion de volum n dintr-o populație f_θ (cu densitatea / fct de masă f_θ)

$$\begin{array}{ccc} x_1, & x_2 \dots & x_n \\ \uparrow & \uparrow & \uparrow \\ X_1(\omega), & X_2(\omega) \dots & X_n(\omega) \end{array}$$

$x_1, x_2 \dots x_n$ - s.n. eșantion realizat (datele)

<u>Def</u>: Fie $X_1, X_2, \ldots X_n$ un eşantion de volum n. Se numeşte statistică orice funcţie $T(x_1, \ldots x_n)$ cu valori reale sau vect.

$$T_n = T(X_1, X_2, \ldots X_n)$$

<u>Exp</u>: 1) $T(x_1 \ldots x_n) = 1$, $T_n = 1$

2) $T(x_1, \ldots x_n) = x_1$, $T_n = X_1$

3) $T(x_1 \ldots x_n) = x_1 + \ldots + x_n$

<u>Media, varianţa şi momentele empirice</u>

<u>Def</u>: Fie $X_1, X_2 \ldots X_n$ un eşantion dintr-o populaţie f_θ. Se numeşte media empirică (media eşantionului), statistica

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

Se numeşte varianţa empirică, V_n^2,

$$V_n^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$$

varianţa eşantionului, S_n^2,

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$$

Se numeşte moment empiric de ordin r, M_r,

$$M_r = \frac{1}{n} \sum_{i=1}^{n} X_i^r$$

Moment centrat empiric: $M_r' = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X}_n)^r$