Curs 17

Media, varianța și momentele empirice

Fie $X_1, X_2, \dots X_n \sim f_\theta$, $\theta \in \Theta \subseteq \mathbb{R}^k$

(indep. și identic repartizate cu densitatea (resp. fct de masă) f_θ

Exp: a) discret:

- Binomiala $(\underbrace{n,p}_{\theta})$
- Bernoulli $(\underbrace{p}_{\theta})$
- Geometrică $(\underbrace{p}_{\theta})$
- Poisson $(\underbrace{\lambda}_{\theta})$

b) cont.

- Exponențială $(\underbrace{\lambda}_{\theta})$
- Normală $(\underbrace{\mu,\sigma^2}_{\theta})$

media eșantionului (empirică) $\longrightarrow$ $T_n := \overline{X}_n = \dfrac{X_1+X_2+\dots+X_n}{n}$

varianța empirică $\longrightarrow$ $V_n^2 = \dfrac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2$

varianța eșantionului $\longrightarrow$ $S_n^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2$

moment empiric de ordin r $\longrightarrow$ $M_r = \dfrac{1}{n}\sum_{i=1}^{n}X_i^r$

moment centrat empiric de ordin r $\longrightarrow$ $M_r' = \dfrac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X}_n)$

(P) Dacă $X_1, X_2, \ldots X_n \sim f_\theta$ de medie μ și dispersie σ^2 atunci

a) $\mathbb{E}[\bar{X}_n] = \mu$

b) $Var(\bar{X}_n) = \sigma^2/n$

c) $\mathbb{E}[V_n^2] = \frac{n-1}{n}\sigma^2$ și $\mathbb{E}[S_n^2] = \sigma^2$

Dem: a)

$$\mathbb{E}[\bar{X}_n] = \mathbb{E}\left[\frac{X_1+X_2+\cdots+X_n}{n}\right] = \frac{1}{n}\mathbb{E}[X_1+X_2+\cdots+X_n]$$

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

$$= \mu$$

b) $$Var(\bar{X}_n) = Var\left(\frac{X_1+X_2+\cdots+X_n}{n}\right) = \frac{1}{n^2}Var(X_1+X_2+\cdots+X_n)$$

$$\overset{\text{indep}}{=} \frac{1}{n^2}\left(\underbrace{Var(X_1)}_{\sigma^2} + \underbrace{Var(X_2)}_{\sigma^2} + \cdots + \underbrace{Var(X_n)}_{\sigma^2}\right) = \sigma^2/n$$

c) $\mathbb{E}[V_n^2] = ?$

$$V_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 = \frac{1}{n}\sum_{i=1}^{n}\Big(\overbrace{X_i - \mu + \mu - \bar{X}_n}^{(X_i-\mu)-(\bar{X}_n-\mu)}\Big)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left[(X_i-\mu)^2 - 2(X_i-\mu)(\bar{X}_n-\mu) + (\bar{X}_n-\mu)^2\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 - \frac{2}{n}\sum_{i=1}^{n}(X_i-\mu)(\bar{X}_n-\mu) + \frac{1}{n}\sum_{i=1}^{n}(\bar{X}_n-\mu)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 - \frac{2}{n}(\bar{X}_n-\mu)\sum_{i=1}^{n}(X_i-\mu) + (\bar{X}_n-\mu)^2$$

$$\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu) = \bar{X}_n - \mu$$

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

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

$$\mathbb{E}[V_n^2] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 - (\overline{X}_n-\mu)^2\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n}\underbrace{\mathbb{E}[(X_i-\mu)^2]}_{Var(X_i)} - \underbrace{\mathbb{E}[(\overline{X}_n-\mu)^2]}_{Var(\overline{X}_n)}$$

$$= \sigma^2 - \sigma^2/n = \frac{n-1}{n}\sigma^2$$

În mod similar,

$$\mathbb{E}[S_n^2] = \mathbb{E}\left[\frac{n}{n-1}V_n^2\right] = \sigma^2.$$

Obs: $\mathbb{E}(\ ) \longrightarrow \frac{1}{n}\sum_{i=1}^{n}(\ )$ (LNM)

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

$$Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

Def: Considerăm un eșantion de volum n de perechi de observații $(X_1,Y_1), (X_2,Y_2), \ldots (X_n,Y_n)$. Se numește covarianța empirică:

$Cov(X,Y) =$

$$\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X}_n)(Y_i-\overline{Y}_n) \quad (\mathbb{E}[(X-\mathbb{E}(X))(Y-\mathbb{E}(Y$$

Coeficientul de corelație liniară empiric:

$$\frac{\sum_{i=1}^{n}(x_i-\bar{x}_n)(y_i-\bar{y}_n)}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x}_n)^2\sum_{i=1}^{n}(y_i-\bar{y}_n)^2}} \quad \left(\rho(X,Y)=\frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}\right)$$

Cazul în care populația este normală

$X_1, X_2, \dots X_n \sim N(\mu,\sigma^2)$

(P) Fie $X_1, X_2, \dots X_n$ un eșantion de volum n dintr-o populație $N(\mu,\sigma^2)$ și fie $\bar{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i$ și $S_n^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X}_n)^2$

Atunci:

a) $\bar{X}_n \sim N(\mu, \sigma^2/n)$

b) $\bar{X}_n$ și S_n^2 sunt independente

c) $\frac{(n-1)S_n^2}{\sigma^2}$ este repartizată hi pătrat (χ^2) cu $n-1$ grade de libertate

Repartiții derivate din rep. normală

1) Repartiția χ^2 (hi-pătrat)

Def: Dacă $Z_1, Z_2, \dots, Z_\nu$ sunt v.a. i.i.d. de rep. comună $N(0,1)$ atunci v.a.

$$Z = Z_1^2+Z_2^2+\dots+Z_\nu^2 \text{ este repartizată } \chi^2 \text{ cu } \nu \text{ grade de libertate}$$

și se notează $Z \sim \chi^2(\nu)$

Obs: Dacă $Z \sim \chi^2(\nu)$ atunci $E[Z] = \nu$

$Var(Z) = 2\nu$

Obs: Dacă $Z_1 \sim \chi^2(\nu_1)$ și $Z_2 \sim \chi^2(\nu_2)$ iar $Z_1 \perp\!\!\!\perp Z_2$ atunci

$Z_1 + Z_2 \sim \chi^2(\nu_1 + \nu_2)$

Obs: Cum $\frac{(n-1)S_n^2}{\sigma} \sim \chi^2(n-1)$ și $Var(\chi^2(n-1)) = 2(n-1)$

găsim că $Var\left(\frac{n-1}{\sigma^2} S_n^2\right) = 2(n-1) \Rightarrow \boxed{Var(S_n^2) = \frac{2\sigma^4}{n-1}}$

2) Repartiția t-Student

Def: Fie $Z \sim N(0,1)$ și $Q \sim \chi^2(\nu)$ astfel încât Z și Q sunt independente.

Atunci v.a.

$$T = \frac{Z}{\sqrt{\frac{Q}{\nu}}}$$

este rep. Student cu ν grade de libertate

și se notează $T \sim t(\nu)$

(P) (Aplicația fundamentală a rep. Student)

Fie $X_1, X_2, \dots, X_n$ un eșantion de volum n dintr-o pop. normală $N(\mu, \sigma^2)$. Atunci

$$\frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \sim t(n-1)$$

Obs: $X_1, X_2, \dots, X_n \sim N(\mu, \sigma^2) \Rightarrow \overline{X}_n \sim N(\mu, \sigma^2/n)$

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

μ, σ^2 - necunoscute

$$\frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \cdot \frac{\sigma}{S_n} = \frac{\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{S_n^2}{\sigma^2}}}$$

$$= \frac{\overbrace{\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}}^{\sim N(0,1)}}{\sqrt{\frac{\overbrace{\frac{(n-1)S_n^2}{\sigma^2}}^{\sim \chi^2(n-1)}}{(n-1)}}} \sim t(n-1)$$

3) Repartiția Fisher-Snedecor

Def: Fie U și V două v.a. indep. astfel încât $U \sim \chi^2(\nu_1)$ și $V \sim \chi^2(\nu_2)$. Atunci

$$F = \frac{U/\nu_1}{V/\nu_2}$$

este rep. Fisher-Snedecor cu ν_1 grade de libertate la numărător și ν_2 gr. de libertate la numitor.

Se notează $F \sim F(\nu_1, \nu_2)$

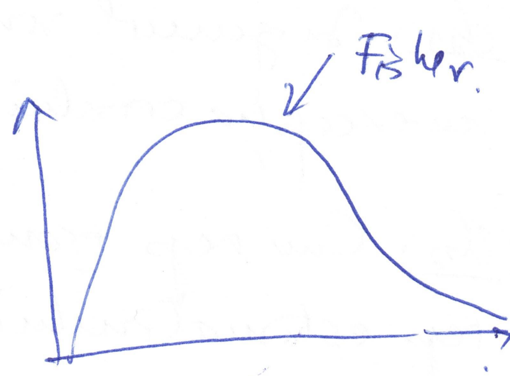
Obs: Dacă $F \sim F(\nu_1, \nu_2)$ at. $1/F \sim F(\nu_2, \nu_1)$

Obs: Apare atunci când studiem raportul a două sume de pătrate de v.a. normale indep.

Fie $X_1, X_2, \dots, X_{n_1} \sim N(\mu_1, \sigma_1^2)$ și $Y_1, \dots, Y_{n_2} \sim N(\mu_2, \sigma_2^2)$

Știm $\dfrac{(n_1-1)S_{n_1}^2}{\sigma_1^2} \sim \chi^2(n_1-1)$

$\dfrac{(n_2-1)S_{n_2}^2}{\sigma_2^2} \sim \chi^2(n_2-1)$



Atunci

$$\frac{\dfrac{\frac{(n_1-1)S_{n_1}^2}{\sigma_1^2}}{n_1-1}}{\dfrac{(n_2-1)S_{n_2}^2/\sigma_2^2}{n_2-1}} = \frac{\dfrac{\chi^2(n_1-1)}{n_1-1}}{\dfrac{\chi^2(n_2-1)}{n_2-1}} \sim F(n_1-1, n_2-1)$$

deci $\dfrac{S_{n_1}^2}{S_{n_2}^2} \cdot \dfrac{\sigma_2^2}{\sigma_1^2} \sim F(n_1-1, n_2-1)$

Estimare punctuală

Fie $X_1, X_2, \dots, X_n \sim f_\theta$, $\theta \in \Theta \subseteq \mathbb{R}^k$

↓ sp. parametrilor

Problema estimării punctuale constă în găsirea unei fct. care depinde de v.s. (eșantion) și care să aproximeze (estimeze) cât mai "bine" parametrul necunoscut.

Def: Fie $X_1, X_2, \dots, X_n \sim f_\theta$, $\theta \in \Theta \subset \mathbb{R}^k$. S.n. estimator punctual al lui θ o statistică $\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \dots, X_n)$ care ia valori în Θ.

Obs: În general vom considera că Θ este un subset din $\mathbb{R}$ cu excepția cazului prop. normale în care $\Theta \subseteq \mathbb{R}^2$.

Obs: Cum rep. eșantionului $X_1, X_2, \dots X_n$ depinde de θ și rep. estimatorului $\hat{\theta}_n$ va depinde de θ.

Proprietăți ale estimatorilor

1) Nedeplasare

Def: Se numește deplasarea estimatorului $\hat{\theta}_n$ față de θ cantitatea

$$b_\theta(\hat{\theta}_n) = \mathbb{E}_\theta[\hat{\theta}_n] - \theta$$ (b_θ - vine de la bias)

Spunem că estimatorul $\hat{\theta}_n$ este nedeplasat dacă

$$b_\theta(\hat{\theta}_n) = 0, \forall \theta.$$

sau $$\mathbb{E}_\theta[\hat{\theta}_n] = \theta, \forall \theta.$$

Ex: $X_1, X_2, \dots X_n$ un eșantion dintr-o pop. de medie μ și dispersie σ^2.

Fie $\bar{X}_n$ și S_n^2 media resp. varianța eșantionului.

Am văzut $\left.\begin{array}{l}\mathbb{E}[\bar{X}_n] = \mu \\ \mathbb{E}[S_n^2] = \sigma^2\end{array}\right\} \Rightarrow \bar{X}_n$ și S_n^2 sunt estimatori nedeplasați pt μ și σ^2

Dacă luăm V_n^2 atunci

$$\mathbb{E}[V_n^2] = \frac{n-1}{n}\sigma^2 \Rightarrow b_{\sigma^2}(V_n^2) = \frac{n-1}{n}\sigma^2 - \sigma^2 = -\frac{1}{n}\sigma^2$$

Obs: Să presupunem că

$$\hat{\theta}_n = \begin{cases} \theta + 1000, & \text{prob. } 1/2 \\ \theta - 1000, & \text{prob. } 1/2 \end{cases}$$

$$\mathbb{P}_\theta(\hat{\theta}_n = \theta - 1000) = \mathbb{P}_\theta(\hat{\theta}_n = \theta + 1000) = 1/2$$

$\mathbb{E}_\theta[\hat{\theta}_n] = \theta$, $\forall \theta \Rightarrow \hat{\theta}_n$ este nedeplasat

Obs: În general, dacă $\hat{\theta}_n$ este un estimator nedeplasat pt θ iar g este o fct. oarecare at. $g(\hat{\theta}_n)$ nu este un estimator nedeplasat pt $g(\theta)$.

Exp: Fie $X \sim Pois(\theta)$ și $\eta = e^{-2\theta}$.

Dacă $\hat{\theta}_1 = X$ at. $\hat{\theta}_1$ este nedeplasat pt θ dar ce se întâmplă cu $\hat{\eta} = e^{-2X}$?

Dar dacă considerăm $\hat{\eta}_2 = (-1)^X$?

2) Consistență

Def: Spunem că estimatorul $\hat{\theta}_n$ este consistent pt θ dacă $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta$

Reamintim: Fie $(X_n)_n$ și X o v.a. atunci $X_n \xrightarrow{\mathbb{P}} X$ dacă $\forall \varepsilon$

$$\lim_n \mathbb{P}(|X_n - X| > \varepsilon) = 0$$

Obs: Intuitiv avem că rep. lui $\hat{\theta}_n$ se concentrează în jurul lui θ.

Exp: Media eșantionului este un estimator consistent pt μ

$\bar{X}_n \xrightarrow{\mathbb{P}} \mu$ (de ce? - din LNM)

Obs: Varianța eșantionului este un estimator consistent pt σ^2

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

$\frac{n}{n-1} \to 1$

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

Din LNM: $\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{\mathbb{P}} \mathbb{E}[X_1^2]$ $\left.\right\}$ $\Rightarrow V_n^2 \xrightarrow{\mathbb{P}} \underbrace{\mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2}_{\sigma^2}$

$\bar{X}_n \xrightarrow{\mathbb{P}} \mathbb{E}[X_1]$

$\frac{n}{n-1} \longrightarrow 1$ pt $n \to \infty$

Astfel $S_n^2 = \frac{n}{n-1} \times V_n^2 \xrightarrow{\mathbb{P}} \sigma^2$

Exp: Fie $X_1, X_2, \ldots X_n \sim N(\theta, 1)$

$$\hat{\theta}_n = \begin{cases} 17, & n < 10^{100} \\ \bar{X}_n, & \text{altfel} \end{cases}$$

$a_n = \mathbb{P}_\theta(|\hat{\theta}_n - \theta| > \varepsilon)$

$b_n = \mathbb{P}_\theta(|\bar{X}_n - \theta| > \varepsilon) \longrightarrow 0$ pt $n \to \infty$ (LNM)

și în plus $a_n = b_n \ \forall n \geq 10^{100} = N$

$\Rightarrow a_n \longrightarrow 0 \Rightarrow \hat{\theta}_n \xrightarrow{\mathbb{P}} \theta$ (este consistent)