7.2 Centralni granični teorem (CGT)

Kažemo da sl. var. X ima normalnu (ili Gaussovu) razdiobu s parametrima μ i σ^2 ($\mu \in \mathbb{R}$, $\sigma > 0$), ako je neprekidna s gustoćom

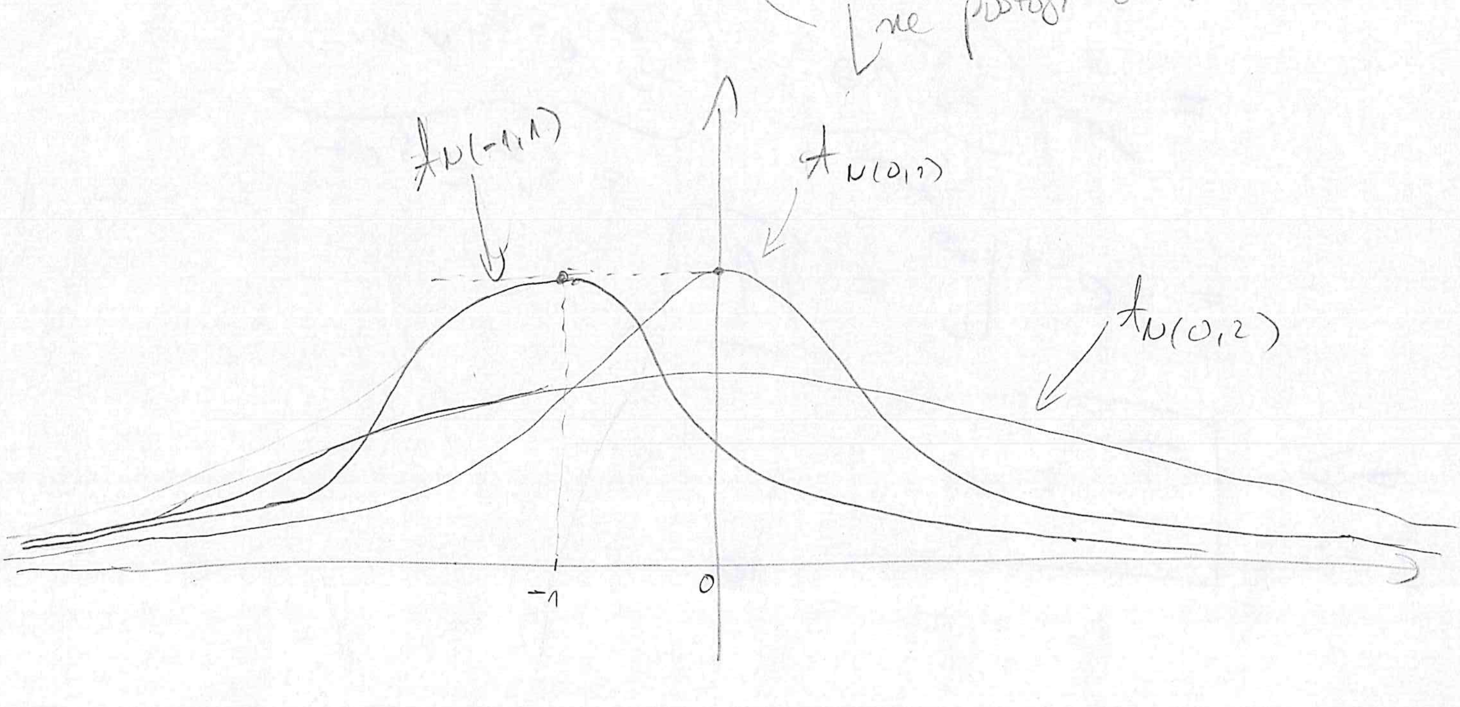
$$f_{N(\mu,\sigma^2)}(t) := \boxed{f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}, \quad t \in \mathbb{R}} \qquad (7.6)$$

(oznaka $X \sim N(\mu, \sigma^2)$)

Razdiobu $N(0,1)$ nazivamo standardna normalna razdioba, a njenu f-ju distribucije obično označavamo s

$$\boxed{\Phi(x) = \int_{-\infty}^{x} \underbrace{\frac{1}{\sqrt{2\pi}} \cdot e^{-t^2/2}}_{= f_{N(0,1)}(t)} dt, \quad x \in \mathbb{R}} \qquad (7.7)$$

[ne postoji za tvoreni oblik!]



Nap. 7.9] Zaista vrijedi

$$\underbrace{\int_{-\infty}^{\infty} f_{N(\mu,\sigma^2)}(t)\,dt}_{=:I} = 1, \quad \forall \mu, \sigma^2.$$

Dokaz.] [Dz]

Imamo

$$I = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\infty} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\,dt$$

$$= \left[z = \frac{t-\mu}{\sigma},\ dz = \frac{dt}{\sigma} \right]$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2}\,dz$$

$$\Rightarrow I^2 = \frac{1}{2\pi} \left(\int_{-\infty}^{\infty} e^{-z^2/2}\,dz \right) \cdot \left(\int_{-\infty}^{\infty} e^{-y^2/2}\,dy \right)$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z^2+y^2)}\,dz\,dy$$

$$= [\text{polarne koordinate}]$$

$$= \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} e^{-r^2/2}\, r\,dr\,d\Theta$$

$$= \frac{1}{\cancel{2\pi}} \underbrace{\left(\int_0^{2\pi} d\Theta \right)}_{=\cancel{2\pi}} \cdot \underbrace{\left(\int_0^{\infty} e^{-r^2/2}\, r\,dr \right)}_{= \int_0^{\infty} e^{-t}\,dt}$$

$$= -e^{-t} \Big|_0^{\infty} = \boxed{1}$$

$$\Rightarrow \boxed{I = 1}$$

$I \geq 0$

$\boxtimes$

Ako je $Z \sim N(0,1)$, imamo ← apsolutno integrabilno!

- $\mathbb{E}[Z] = \int_{-\infty}^{+\infty} \underbrace{t \frac{1}{\sqrt{2\pi}} e^{-t^2/2}}_{\text{neparna f-ja}} dt = \boxed{0}$ (7.8)

- $\mathrm{Var}(Z) = \mathbb{E}[Z^2] = \int_{-\infty}^{+\infty} t^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$

$$= \begin{bmatrix} u = t & dv = t e^{-t^2/2} \\ du = 1 & v = -e^{-t^2/2} \end{bmatrix}$$

$$= \frac{-1}{\sqrt{2\pi}} t e^{-t^2/2} \Big|_{-\infty}^{+\infty} + \overbrace{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-t^2/2} dt}^{=1}$$

$$= -0 + 0 + 1 = \boxed{1} \quad (7.9)$$

Prop. 7.10 (1) Ako je $Z \sim N(0,1)$, $\forall \mu \in \mathbb{R}$ i $\sigma > 0$ vrijedi

$$X := \sigma \cdot Z + \mu \sim N(\mu, \sigma^2) \quad (7.10) \ \square$$

Dokaz: $\forall x \in \mathbb{R}$,

$$P(X \leq x) \overset{\sigma>0}{=} P\left(Z \leq \frac{x-\mu}{\sigma}\right) \overset{Z \sim N(0,1)}{=} \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

$$= \begin{bmatrix} z = \sigma \cdot t + \mu, & \frac{dz}{\sigma} = dt \\ t = \frac{z-\mu}{\sigma} & \end{bmatrix} = \int_{-\infty}^{x} \underbrace{\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(z-\mu)^2}{2\sigma^2}}}_{= f_{N(\mu,\sigma^2)}(z)} dz$$

$\Longrightarrow X \sim N(\mu, \sigma^2)$. $\square$

po def.

(7.8) - (7.10) povlače da za $X \sim N(\mu, \sigma^2)$, vrijedi

$$\boxed{\begin{aligned} \mathbb{E}[X] &= \mu + \sigma \underbrace{\mathbb{E}[Z]}_{=0} = \mu, \\ Var(X) &= \sigma^2 \cdot \underbrace{Var(Z)}_{=1} = \sigma^2, \end{aligned}}$$

te

$$\boxed{\frac{X-\mu}{\sigma} \sim N(0,1)}.$$

Tm. 7.11 (CGT) Neka su $X_1, X_2, \ldots$ njd sl. varijable t.d.

$$\boxed{\sigma^2 := Var(X_1) < \infty},$$

te neka je $\mu := \mathbb{E}[X_1]$, i $S_n = \sum_{i=1}^{n} X_i$, $\forall n \in \mathbb{N}$.

Tada vrijedi

$$(7.11) \quad \lim_{n\to\infty} \mathbb{P}\left(\frac{S_n - \overset{=\mathbb{E}[S_n]}{n\mu}}{\underset{=\sqrt{Var(S_n)}}{\sqrt{n}\cdot\sigma}} \le x \right) = \mathbb{P}(Z \le x), \forall x \in \mathbb{R}$$

(bez dokaza)

gdje je $\boxed{Z \sim N(0,1)}$.

Nap. 7.12 (7.11) pišemo

"konvergira po distribuciji"

$$\frac{S_n - n\mu}{\sqrt{n}\,\sigma} \xrightarrow[n\to\infty]{d} Z \sim N(0,1)$$

Uočimo, za "veliki" n i sve $x \in \mathbb{R}$,

$$\mathbb{P}(S_n \le x) = \mathbb{P}\left(\frac{S_n - n\mu}{\sqrt{n}\cdot\sigma} \le \frac{x-n\mu}{\sqrt{n}\,\sigma}\right)$$

$$\underset{\text{CGT, tj. (7.11)}}{\approx} \mathbb{P}\left(Z \le \frac{x-n\mu}{\sqrt{n}\,\sigma}\right)$$

$$= \mathbb{P}(\underbrace{\sqrt{n}\cdot\sigma Z + n\mu}_{\sim N(n\mu,\, n\sigma^2)} \le x),$$

tj. intuitivno,

"približno po distr."

$$\boxed{S_n \overset{d}{\approx} N(\underset{\mathbb{E}[S_n]}{\overset{\shortparallel}{n\mu}}, \underset{Var(S_n)}{\overset{\shortparallel}{n\sigma^2}})}$$

Nije bitna distrib. od X_n, već samo da je $\sigma^2 < \infty$

<u>Nap. 7.13</u> Može se pokazati da (7.11) povlači

$$\lim_{n\to\infty} \mathbb{P}\left(a \underset{\le}{\overset{<}{\text{ili}}} \frac{S_n - n\mu}{\sqrt{n}\,\sigma} \underset{\le}{\overset{<}{\text{ili}}} b\right) = \mathbb{P}(a \underset{\le}{\overset{<}{\text{ili}}} Z \underset{\le}{\overset{<}{\text{ili}}} b)$$

$$\underset{(7.12)}{=} \Phi(b) - \Phi(a),$$

za sve $-\infty \le a < b \le +\infty$. ∎

<u>Pr. 7.14</u> Neka je $\boxed{Y_n \sim B(n,p)}$, $\forall n \in \mathbb{N}$, za $\boxed{p \in \langle 0,1\rangle \text{ fiksan.}}$

Ako u Tm. 7.11 uzmemo $X_1 \sim \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}$

$\Rightarrow$ $\mu = p$, $\sigma = pq$, te $S_n \sim Y_n$.

$$\underset{\text{CGT}}{\Rightarrow} \quad \frac{Y_n - np}{\sqrt{npq}} \xrightarrow[n\to\infty]{d} Z \sim N(0,1) \qquad (7.13)$$

tzv. $\boxed{\text{de Moivre - Laplaceov tm.}}$ ∎

Uočimo,

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \cdot \frac{1/n}{1/n} = \frac{\sqrt{n}}{\sigma}\left(\frac{S_n}{n} - \mu\right) \qquad (7.14)$$

$\Rightarrow$ CGT daje koliko $\frac{S_n}{n}$ odstupa od μ

(po JZVB, $\frac{S_n}{n} \xrightarrow[n\to\infty]{g.s.} \mu$).

Pr. 7.15 | Nepoznati postotak glasača, p, glasat će za jednu stranku. Slučajno je odabrano n ljudi te neka je

S_n = # ispitanika koji će glasati za tu stranku.

Koliki treba biti n da $\frac{S_n}{n}$ bude na $\pm 3\%$ od p s vjerojatnosti od barem 0.95? [TV prijelaze]

Rj. Uz pretpostavku da su ispitanici nezavisni,

$$S_n \sim B(n, p).$$

Želimo

$$0.95 \le \mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \le 0.03\right)$$

$$= \mathbb{P}\left(\frac{\sqrt{n}}{\sqrt{pq}}\left|\frac{S_n}{n} - p\right| \le \underbrace{\frac{\sqrt{n}}{\sqrt{pq}}\, 0.03}_{=: t_n}\right)$$

$$\underset{(7.12)-(7.14)}{\approx} \Phi(t_n) - \Phi(-t_n)$$

$$= 2\Phi(t_n) - 1,$$

$\Phi(-t_n) = 1 - \Phi(t_n)$

tj. $\Phi(t_n) \geq 0.975$

Iz $\Phi(1.96) \approx 0.975$, imamo

ne znamo p!

$$t_n \geq 1.96,$$

tj. $$\sqrt{n} \geq \frac{1.96}{0.03}\sqrt{pq} = 65.33\sqrt{pq}. \quad (*)$$

Ipak, znamo da $\sqrt{pq} = \sqrt{p(1-p)} \leq \sqrt{\frac{1}{2}(1-\frac{1}{2})} = \frac{1}{2}$, $\forall p \in \langle 0,1\rangle$,

pa odavde je $$n \geq 65.33^2 \cdot \left(\frac{1}{2}\right)^2 = \boxed{1067.11},$$

vrijedi $(*)$ pa onda:

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \leq 0.03\right) \geq 0.95$$

∎