URSPFMF2026-03-19

roka: datumi teoretičnih testov: ob 10:00 ali kasneje:

sreda, 17. 6.
torek, 30. 6.
četrtek, 27. 8.
četrtek, 10. 9.

$S_n = X_1 + \dots + X_n$

$\sigma(X_n) = \sqrt{D(X_n)} = \sqrt{D(X_1)+\dots+D(X_n)} = \sqrt{n\sigma^2} = \sqrt{n}\,\sigma$

centralni limitni izrek: Naj bodo $\{X_n\}$ neodvisne, enako porazdeljene slučajne spremenljivke: $E(X_n) = \mu$, $\sigma(X_n) = \sigma$. Potem za

$$Z_n := \frac{S_n - E(S_n)}{\sigma(S_n)}$$

velja: $\forall x \in \mathbb{R}$: $F_{Z_n}(x) \longrightarrow F_{N(0,1)}(x) =$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\,dt = \frac{1}{2}\,\Phi(x),$$ kjer je $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2}\,dt$

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

torej za velike n je $Z_n \approx N(0,1)$. Zapišimo še drugače:

$$Z_n = \frac{S_n - n\mu \quad /:n}{\sqrt{n}\,\sigma \quad /:n} = \frac{\frac{S_n}{n} - \mu}{\sigma/\sqrt{n}} \approx N(0,1).$$

Def.: Vzorčno povprečje: $\bar{X} := \frac{S_n}{n} = \frac{X_1 + \dots + X_n}{n}$.

Po CLI je zato $\bar{X} \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ (porazdeljena približno). Če je na začetku X_n porazdeljena $N(\mu, \sigma)$, potem $\forall n \in \mathbb{N}$: $\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ (porazdeljena eksaktno)

→ Laplaceova formula je poseben primer CLI:

Opazujemo neodvisne ponovitve postopka, v katerem je možen dogodek A s $P(A) = p$. Naj bo $(X_n = 1)$ dogodek, da se v i-ti ponovitvi postopka zgodi dogodek A; sicer je $X_n = 0$. Potem je $X_n : \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}$, kjer je $q = 1 - p$. To so neodvisne slučajne spremenljivke. $S_n = X_1 + X_2 + \dots + X_n$ je frekvenca dogodka A v prvih n ponovitvah postopka:

$S_n \sim \mathrm{Bin}(n, p)$. Torej je $E(S_n) = n \cdot p$, $D(S_n) = npq$,
$\sigma(S_n) = \sqrt{npq}$. CLI pravi, da je

$$\underbrace{P\left(\frac{S_n - np}{\sqrt{npq}} \le x\right)}_{\parallel} \xrightarrow[n\to\infty]{} \frac{1}{2} + \Phi(x)$$

$$P(S_n \le \underbrace{x\sqrt{npq} + np}_{= \alpha}) = P(S_n \le \alpha)$$

torej $x = \frac{\alpha - np}{\sqrt{npq}}$ in imamo torej

$$P(S_n \le \alpha) \approx \frac{1}{2} + \Phi\left(\frac{\alpha - np}{\sqrt{npq}}\right).$$ Od tod sledi Laplaceova

formula: $P(\alpha \le S_n \le \beta) = P(S_n \le \beta) - P(S_n < \alpha) \approx$

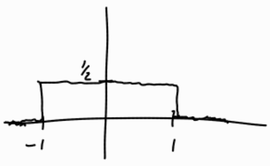
$$\approx P(S_n \le \beta) - P(S_n \le \alpha) \approx \Phi\left(\frac{\beta-np}{\sqrt{npq}}\right) - \Phi\left(\frac{\alpha-np}{\sqrt{npq}}\right) \cancel{+\frac{1}{2}-\frac{1}{2}}$$

Zakaj je veliko stvari v življenju/naravi porazdeljeno približno normalno?

Primer: Prodajalec kostanja pozimi v mestu v škrnicelj vrže točno 20 kostanjev v vrečko ($n=20$). Zanima nas teža škrnicla tega kostanja. Kako je teža porazdeljena?
Teža je $S_{20} = X_1 + X_2 + \dots + X_{20}$, kjer je X_k teža k-tega kostanja. Smiselno je privzeti, da so $\{X_k\}$ neodvisne in enako porazdeljene slučajne spremenljivke. Ne vemo pa, kako so porazdeljene. Po CLI-ju je S_{20} približno normalno porazdeljen. Dodatno, če je že teža vsega kostanja porazdeljena normalno, je S_{20} eksaktno normalno porazdeljena. Velja aditiven efekt: seštevamo veliko neodvisno enako porazdeljenih spremenljivk.

Kako hitro $\{F_{Z_n}(x)\}_{n\in\mathbb{N}}$ pri fiksnem x konvergira k $F_{N(0,1)}(x)$?

Primer: Let $\{X_n\}_{n\in\mathbb{N}}$ neodvisne sl. spr., porazdeljene enakomerno na $[-1,1]$. gostota: p_X:

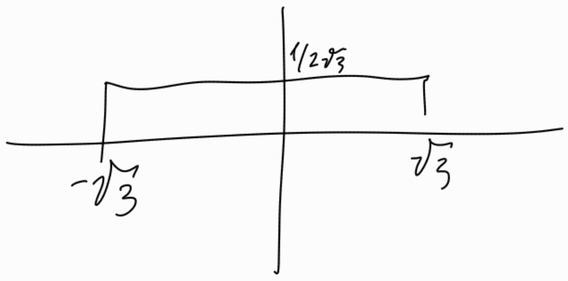


Potem je $E(X_n)=0$ in $D(X_n) = E(X_n^2) - (E(X_n))^2 =$

$= \frac{1}{2}\int_{-1}^{1} x^2\,dx = \frac{1}{2}\,2\int_0^1 x^2\,dx = \frac{x^3}{3}\Big|_0^1 = \frac{1}{3}$. Sledi: $\sigma(X_n) = \frac{1}{\sqrt{3}} = \sigma$.

$$Z_n = \frac{S_n - E(S_n)}{\sigma(S_n)} = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{X_1 + \dots + X_n}{\sqrt{n/3}} = \sqrt{\frac{3}{n}}(X_1 + \dots + X_n)$$

torej $Z_1 = X_1\sqrt{3}$. Ki je porazd. enak. na $[-\sqrt{3}, \sqrt{3}]$.



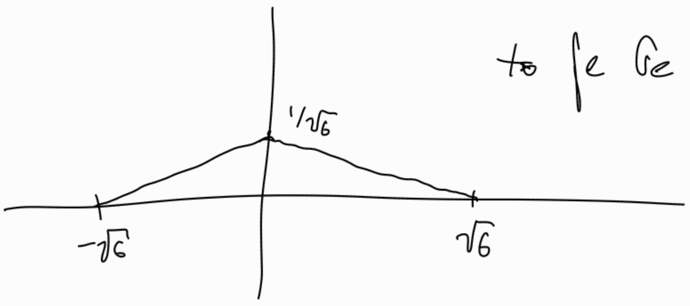
$Z_2 = \sqrt{3/2}\,(X_1+X_2) = \sqrt{3/2}\,Y$; $Y := X_1 + X_2$ ima neničelno gostoto na $[-2,2]$. Očitno je gostota $p_Y(y)$

sedaj: $p_Y(y) = \int_{-\infty}^{\infty} p_{X_1}(x)\, p_{X_2}(y-x)\,dx \overset{DN}{=} \frac{1}{4}(2-y)$. in zato

$p_Y(y) = \begin{cases} \frac{1}{4}(2-|y|) & ; \text{ če } |y| \le 2 \\ 0 & ; \text{ sicer} \end{cases}$. Gostota za Z_2 je

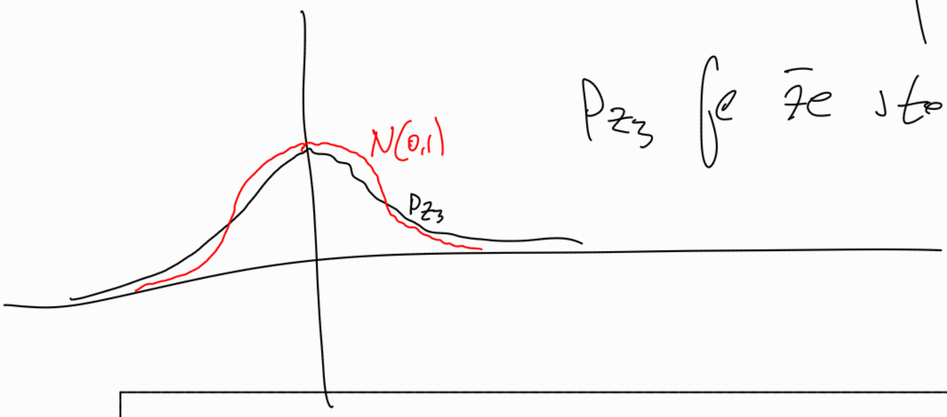
konična na $[-\sqrt{6}, \sqrt{6}]$: $p_{Z_2}(z) \overset{DN}{=} \begin{cases} \frac{1}{6}\left(\sqrt{6} - |z|\right) & ; |z| \le \sqrt{6} \\ 0 & ; \text{sicer} \end{cases}$

to je že bližje std. normalni $N(0,1)$.



$Z_3 = X_1 + X_2 + X_3 = Y + X_3 \quad \overset{DN}{\dashrightarrow} \quad p_{Z_3}(z) \begin{cases} \frac{1}{8}\left(3 - z^2\right) & ; \text{ če } |z| \le 1 \\ \frac{1}{16}\left(3 - |z|\right)^2 & ; \text{ če } 1 \le |z| \le 3 \\ 0 & ; \text{ sicer} \end{cases}$

p_{Z_3} je že skoraj $N(0,1)$



II. STATISTIKA

[1. osnovni pojmi].

Kot vedo statistiko delimo na:

I. <u>opisno statistiko</u>: zbiranje, razvrščanje, prikazovanje podatkov, računanje osnovnih količin: vzorčne povprečje, std. dev.

II. <u>analitična statistika</u>: uporaba podatkov pri sklepanju glede zakonitosti danega področja.

<u>Def.</u>: Populacija je končna ali neskončna množica elementov, pri katerih opazujemo/merimo neko količino.

<u>Zgleda</u>: 1. kontrola kvalitete izdelkov:

množica/serija izdelkov, denimo dnevna proizvodnja izdelkov. merimo lastnosti izdelkov, denimo koliko na je življenska doba žarnice.

2. testiranje oseb: populacija: vsi zaposleni v državi, merimo starost/plačo zaposlenih.

Matematični pogled: na verjetnostnem prostoru (Ω, F) imamo sluč. spremen. X. Praviloma ne moremo izmeriti cele populacije, ampak meritve opravimo na relativno majhnem delu populacije, tako imenovanem „vzorcu". Le-ta mora biti reprezentativen: izbran naključno in dovolj velik.

Vzorec velikosti n je slučajni vektor $(X_1, X_2, \dots, X_n)$, kjer so $X_1, X_2, \dots, X_n$ enako porazdeljene kot X in neodvisne sl. spremen.