

Zapiski pri predmetu Statistika

Minimalni katalog znanja, ki ga bom sproti dopolnjeval. Verjetno bom izpustil kakšen dokaz in pa kakšen zgled.

1 Motivacija

Kako bi "ocenili" verjetnost, da pri metu kovanca pade cifra?

Izvedemo n neodvisnih "enakih" (v istih razmerah, na enak način, pošteno oz. naključno) metov kovanca in iskano verjetnost ocenimo z razmerjem $\frac{\text{število cifer}}{n}$.

Igramo igro, kjer kroglico položimo v eno od treh škatel. Zmešamo škatle med seboj in poskušamo uganiti kje je kroglica. Če uganemo dobimo 10, v nasprotnem primeru pa izgubimo 6.

Kako bi ocenili pričakovano vrednost te igre?

Izvedemo n neodvisnih slučajnih iger in pričakovano vrednost ene igre ocenimo z $\frac{\text{skupni izkupiček}}{n}$.

Zdi se nam, da mora z večjim vzorcem priti boljša ocena.

V 18. stoletju je grof Buffon kovanec vrgel 4040-krat in dobil 2048 cifer. Ocenjena verjetnost cifre je 0.50689.

V 19. stoletju je Pason vrgel kovanec 12000-krat in dobil 6019 cifer. Ocenjena verjetnost je 0.5016.

Aksiome verjetnosti zgradimo tako, da so naša mnenja glede vprašanj upravičena.

2 Konvergenca slučajnih spremenljivk in limitni izrek

DEFINICIJA 2.1. Naj bodo X_1, X_2, X_3, \dots slučajne spremenljivke, definirane na skupnem prostoru Ω .

(1) Pravimo, da zaporedje $\{X_n\}_n$ konvergira k X v porazdelitvi, če

$$\lim_{n \rightarrow \infty} P(X_n \leq x) = P(X \leq x)$$

za vsa tista realna števila x , v katerih je komulativna porazdelitvena funkcija slučajne spremenljivke X zvezna.

(2) Pravimo, da zaporedje $\{X_n\}_n$ konvergira k X v **verjetnosti**, če velja:

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) = 0$$

za vsak $\varepsilon > 0$.

(3) Pravimo, da zaporedje $\{X_n\}_n$ konvergira k X **skoraj gotovo**, če je:

$$P(\{\omega \in \Omega \mid \exists \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}) = 1$$

$$\Longleftrightarrow$$

$$P(\lim_{n \rightarrow \infty} X_n = X) = 1$$

TRDITEV 2.2. Iz konvergence 'skoraj gotovo' sledi konvergenca v verjetnosti.

TRDITEV 2.3. (Neenakost Markova)

Naj bo X slučajna spremenljivka s pričakovano vrednostjo in $a > 0$ pozitivna konstanta. Tedaj je:

$$P(|X| \geq a) \leq \frac{E[|X|]}{a}$$

DOKAZ. Naj bo $a > 0$. Pišemo $A = \{|X| \geq a\} = \{\omega \mid |X(\omega)| \geq a\}$. Tedaj $|X| \geq a \cdot \mathbf{1}_A$. Sledi $E[|X|] \geq a \cdot P(A)$. ■

POSLEDICA 2.4. (Neenakost Čebiševa)

Naj bo X slučajna spremenljivka s (končno) disperzijo. Tedaj velja

$$P(|X - E[X]| \geq \varepsilon) \leq \frac{D(X)}{\varepsilon^2}$$

za vsako pozitivno število ε .

DOKAZ.

$$P(|X - E[X]| \geq \varepsilon) = P((|X - E[X]|)^2 \geq \varepsilon^2) < \frac{E((X - E[X])^2)}{\varepsilon^2} = \frac{D(X)}{\varepsilon^2}$$

■

IZREK 2.5. (Šibki zakon velikih števil)

Naj bodo $X_1, X_2, \dots \Omega \rightarrow \mathbb{R}$ neodvisne in enako porazdeljene slučajne spremenljivke s pričakovano vrednostjo μ in (končnim) odklonom σ . Tedaj zaporedje "vzorčnih povprečij"

$$\frac{X_1 + X_2 + \dots + X_n}{n}$$

konvergira v verjetnosti h konstanti μ .

DOKAZ. Trdimo, da velja $\lim_{n \rightarrow \infty} P(|\frac{X_1 + X_2 + \dots + X_n}{n} - \mu| \geq \varepsilon) = 0$ za vsak pozitiven $\varepsilon > 0$.

Pišimo $\bar{X} = \frac{X_1 + \dots + X_n}{n}$.

$$P(|\bar{X} - \mu| > \varepsilon) \leq P(|\bar{X} - \mu| \geq \varepsilon) \leq \frac{D(\bar{X})}{\varepsilon^2} = \frac{D(\frac{X_1 + \dots + X_n}{n})}{\varepsilon^2} = \frac{1}{n^2 \varepsilon^2} D(X_1) + \dots + D(X_n) = \frac{\sigma^2}{n \varepsilon^2}$$

Sledi, da rezultat konvergira proti 0, ko gre n v neskončnost. ■

OPOMBA 2.6. Verjetnost kateregakoli konkretnega neskončnega zaporedja cifer in grbov je 0, ne glede na to koliko je dejanska verjetnost posameznega meta $p \in (0, 1)$.

OPOMBA 2.7. (Česa šibki zakon velikih števil ne trdi.)

Denimo, da je $p = \frac{1}{2}$. Beležimo število cifer po n poskusih. **Ne velja**, da je število cifer po n poskusih večje od števila grbov po 'približno polovici časa'.

Zlahka namreč trdimo, da je število cifer ves čas večje od števila grbov.

IZREK 2.8. (*Krepki zakon velikih števil*)

Naj bo X_1, X_2, \dots zaporedje neodvisnih in enako porazdeljenih slučajnih spremenljivk s končno pričakovano vrednostjo $E(X_i) \in \mathbb{R}$. Tedaj je zaporedje "vzorčnih povprečij"

$$\frac{X_1 + X_2 + \dots + X_n}{n}$$

konvergira k $E[X_i] =: \mu$ **skoraj gotovo**.

OPOMBA 2.9. Končna pričakovana vrednost pomeni $E[|X_i|] < \infty$

ZGLED 2.10. Ponavljamo Bernulijev poskus z verjetnostjo enice p . Tedaj skoraj gotovo velja:

$$\lim_{n \rightarrow \infty} \frac{\text{št. enic v } n \text{ poskusih}}{n} = p \quad (1)$$

To pomeni: verjetnost tistih neskončnih zaporedij $(\omega_1, \omega_2, \dots)$ za katere (1) velja, je 1.

OPOMBA 2.11. Krepki zakon velikih števil je uzakonitev frekventistične definicije (intuicije) v verjetnosti.

OPOMBA 2.12. Iz izreka 2.8 sledi izrek 2.5

2.1 Centralni limitni izrek

IZREK 2.13. Naj bodo X_1, X_2, \dots neodvisno enako porazdeljene Bernulijeveke $(B(1, p))$. Tedaj zaporedje **standardiziranih povprečij**

$$\frac{\frac{X_1 + X_2 + \dots + X_n}{n} - p}{\frac{\sqrt{p(1-p)}}{\sqrt{n}}} = \frac{\sqrt{n}}{\sqrt{p(1-p)}} \left(\frac{X_1 + X_2 + \dots + X_n}{n} - p \right)$$

konvergira k standardni normalni porazdelitvi.

Z drugimi besedami: Če velja $Y_n \sim \text{Bin}(n, p)$ sledi:

$$\frac{\sqrt{n}}{\sqrt{p(1-p)}} \left(\frac{Y_n}{n} - p \right) \xrightarrow[n \rightarrow \infty]{\text{v porazdelitvi}} \mathcal{N}(0, 1)$$

OPOMBA 2.14. Dokaz bomo izpustili.

Za $p = \frac{1}{2}$ je dokazal leta 1733 De Moivre.

Za splošen p ga je dokazal Laplace.

Uporabljamo ga za aproksimacijo binomskih porazdelitev za velike n z normalnimi porazdelitvami.

Ohlapno lahko rečemo:

$$\text{Bin}(n, p) \sim \mathcal{N}(np, np(1-p))$$

za velike n -je.

IZREK 2.15. (*Centralni limitni izrek*)

Naj bodo X_1, X_2, \dots neodvisne, enako porazdeljene slučajne spremenljivke s končno disperzijo σ^2 in pričakovano vrednostjo μ . Tedaj zaporedje standardiziranih vzorčnih povprečij:

$$\frac{\frac{X_1 + X_2 + \dots + X_n}{n} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

konvergira k porazdelitvi $\mathcal{N}(0, 1)$.

OPOMBA 2.16. V statistiki izrek 2.15 uporabljamo tipično v primerih, ko so X_1, X_2, \dots neodvisne replikacije preučevane slučajne spremenljivke X .

ZGLED 2.17. Ljubljanske mlekarne proizvajajo litrsko plastenko jogurta Mu 3, 2. 'Jamčijo', da ima taka plastenka 'v povprečju' 32g maščob. Privzamemo tudi, da Ljubljanske mlekarne 'jamčijo', da je odklon vsebnosti maščob 1, 5g.

- (1) Ali znamo izračunati (ali oceniti) $P(X \in (31g, 33g))$, če je X zvezna spremenljivka, ki predstavlja maso maščob v slučajno izbrani plastenki?

V splošnem ne znamo odgovoriti, saj ne poznamo porazdelitve.

- (2) Naključno izberemo 100 takih plasten in označimo X_i maso maščob v i -ti plastenki. Ali znamo izračunati (ali oceniti)?

Lahko ocenimo s pomočjo izreka 2.15. Praktične izkušnje kažejo, da je $n = 100$ že dovolj veliko

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{100} \Rightarrow P\left(\frac{\bar{X} - 32}{\frac{3}{2\sqrt{100}}}\right)$$