Verjetnost in statistika - vaje 11

Disperzija, centralni limitni izrek, višji momenti, asimetrija, sploščenost.

Disperzija (varianca):

$$D(X) = E((X - E(X))^2) = E(X^2) - E(X)^2$$

Za poljubno slučajno spremenljivko X velja:

$$D(aX + b) = a^2 D(X)$$

Standardni odklon:

$$\sigma(X) = \sqrt{D(X)}$$

Če sta X in Y neodvisni, velja:

$$E(XY) = E(X)E(Y)$$

$$D(X + Y) = D(X) + D(Y)$$

Pozor! Zakaj ne velja: $D(X + X) = D(X) + D(X) = 2D(X)$?

1. Dani sta neodvisni slučajni spremenljivki $X \sim \chi^2(3)$ in $Y \sim \chi^2(5)$. Izračunajte $D(3X - Y)$ in $E((X + 2Y)^2)$.

Centralni limitni izrek. Naj bodo $X_1, \ldots, X_n$ neodvisne slučajne spremenljivke, za katere velja $D(X_i) < \infty$ $(1 \leq i \leq n)$ in od katerih nobena posebej ne izstopa:

$$E(X_1) \approx \ldots \approx E(X_n) \quad \text{in} \quad D(X_1) \approx \ldots \approx D(X_n).$$

Če je $n \geq 30$, potem je slučajna spremenljivka S, ki predstavlja njihovo vsoto, tj. $S = X_1 + \cdots + X_n$, porazdeljena približno normalno:

$$S \sim N(\mu_S, \sigma_S),$$

kjer je $\mu_S = E(S)$ in $\sigma_S = \sqrt{D(S)}$, torej je:

$$P(a \leq S \leq b) \approx \Phi\left(\frac{b - \mu_S}{\sigma_S}\right) - \Phi\left(\frac{a - \mu_S}{\sigma_S}\right)$$

Če so $X_1, \ldots, X_n$ neodvisne, je seveda:

$$\mu_S = E(X_1) + E(X_2) + \cdots + E(X_n)$$

$$\sigma_S^2 = D(X_1) + D(X_2) + \cdots + D(X_n).$$

Centralni limitni izrek za vzorčenje. Na neki populaciji nas zanima lastnost, ki jo spremlja slučajna spremenljivka X s končnima parametroma $E(X) = \mu$ in $D(X) = \sigma^2$. Iz te populacije zberemo naključni slučajni vzorec $(x_1, \dots, x_n)$. Če so $X_1, \dots, X_n$ slučajne spremenljivke, ki spremljajo elemente tega vzorca po koordinatah, potem so neodvisne in zanje velja

$$E(X_i) \approx \mu \quad \text{in} \quad D(X_i) \approx \sigma^2 \qquad (1 \le i \le n).$$

Če je $n \ge 30$, potem je slučajna spremenljivka, ki predstavlja vzorčno povprečje, tj. $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$, porazdeljena približno normalno:

$$\overline{X} \sim N(\mu_{\overline{X}}, \sigma_{\overline{X}}),$$

za njena parametra pa velja $\mu_{\overline{X}} = \mu$ in $\sigma_{\overline{X}} = \sigma/\sqrt{n}$.

Uporaba: $\quad P(a \le \overline{X} \le b) \approx \Phi\left((b-\mu)\sqrt{n}/\sigma\right) - \Phi\left((a-\mu)\sqrt{n}/\sigma\right).$

2. Slučajne spremenljivke $X_1, \dots, X_{100}$ so neodvisne in porazdeljene diskretno po naslednji shemi:

$$X_i \sim \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0.5 & 0.1 & 0.3 & 0.1 \end{pmatrix}$$

Ocenite $P(170 < X_1 + \cdots + X_{100} < 210)$ in utemeljite odgovor.

3. Slučajna spremenljivka X ima porazdelitev hi kvadrat s 100 prostostnimi stopnjami. Približno izračunajte $P(X > 110)$ in $P(90 < X < 110)$. Odgovor utemeljite!

4. Potapljač se potaplja, dokler ne nabere 80 biserov. Pri vsakem potopu nabere največ en biser, pa še tega le z verjetnostjo 20%. Potopi so med seboj neodvisni.
   (a) Kolikšno je pričakovano število potopov (matematično upanje)?
   (b) Ocenite verjetnost, da se bo moral potopiti več kot 450-krat. Odgovor utemeljite (možni sta dve uporabi centralnega limitnega izreka)!

Višji momenti:
Moment reda $k \in \mathbb{N}$ glede na točko $a \in \mathbb{R}$ imenujemo količino

$$m_k(a) = E\left((X-a)^k\right).$$

Za $a = 0$ dobimo *začetni moment* $z_k = m_k(0)$, za $a = EX$ pa *centralni moment* $m_k = m_k(E(X))$.

Asimetrija:

$$A(X) = \frac{E((X - E(X))^3)}{D(X)^{3/2}}$$

Če sta $a > 0$ in b konstanti, je $A(aX + b) = A(X)$.

Sploščenost (kurtozis):

$$K(X) = \frac{E((X - E(X))^4)}{D(X)^2} - 3$$

Če sta $a \neq 0$ in b konstanti, je $K(aX + b) = K(X)$.

5. Slučajna spremenljivka X je porazdeljena enakomerno $\mathrm{E}[a, b]$. Izračunajte vse njene višje centralne momente $E\left((X - E(X))^k\right)$, za $k = 1, 2, \ldots$, asimetrijo $A(X)$ in sploščenost $K(X)$.

Število q_α je kvantil slučajne spremenljivke X za verjetnost α, če velja:

$$P(X < q_\alpha) \leq \alpha \qquad P(X \leq q_\alpha) \geq \alpha.$$

Mediano $m = q_{1/2}$ lahko razumemo kot nadomestek za matematično upanje, semiinterkvartilni razmik:

$$s = \frac{q_{3/4} - q_{1/4}}{2}$$

pa za varianco.

6. Slučajna spremenljivka X je porazdeljena zvezno z naslednjo gostoto:

$$p_X(x) = \begin{cases} \dfrac{1}{(1+x)^2} & ;\ x \geq 0 \\ 0 & ;\ \text{sicer} \end{cases}$$

   Izračunajte njeno mediano in semiinterkvartilni razmik.

   Opomba: matematično upanje $E(X)$ ne obstaja!

7. Fakulteta za računalništvo in informatiko ima nad pritličjem 8 različnih nadstropij (M1, 1, M2, ..., M4, 4), ki so bila nekoč označena s številkami od 1 do 8. Denimo, da so tudi sedaj označena tako. V dvigalo vstopi 6 profesorjev, med katerimi si vsak izbere nadstropje naključno, vsakega z isto verjetnostjo in neodvisno od ostalih profesorjev.

   Naj bo H najvišje nadstropje, v katerem se dvigalo ustavi. Poiščite mediano in semiinterkvartilni razmik.