VJEROJATNOST I STATISTIKA

Dodatni zadaci za 10. i 11. tjedan (20. i 21. predavanje)

09 Zakon velikih brojeva i centralni granični teorem

1. Dokažite da za svaku slučajnu varijablu X s očekivanjem a i devijacijom σ vrijedi sljedeća ocjena:

$$\mathbf{P}\left(a-3\sigma<X<a+3\sigma\right)\geq\frac{8}{9}.$$

Koliko iznosi ta vjerojatnost ukoliko se radi o slučajnoj varijabli s normalnom razdiobom?

Rješenje.

Koristimo Čebiševljevu nejednakost koja glasi:

$$\mathbf{P}\left(\left|X-E\left(X\right)\right|\geq\varepsilon\right)\leq\frac{\mathbf{D}\left(X\right)}{\varepsilon^2}.$$

Uvrstimo očekivanje $\mathbf{E}\left(X\right)=a$ i disperziju $\mathbf{D}\left(X\right)=\sigma^2$ u prethodnu formulu:

$$\mathbf{P}\left(\left|X-a\right|\geq\varepsilon\right)\leq\frac{\sigma^2}{\varepsilon^2}.$$

Vidimo da nam treba suprotna nejednakost:

$$\mathbf{P}\left(\left|X-a\right|<\varepsilon\right)=1-\mathbf{P}\left(\left|X-a\right|\geq\varepsilon\right)\geq1-\frac{\sigma^2}{\varepsilon^2},$$

Sada za vrijednost ε odaberemo $\varepsilon=3\sigma$:

$$\mathbf{P}\left(\left|X-a\right|<3\sigma\right)\geq1-\frac{\sigma^2}{9\sigma^2}=\frac{8}{9},$$

što je upravo tražena nejednakost.

Ako slučajna varijabla X ima normalnu razdiobu, tada se radi o poznatom 3σ pravilu te vrijedi:

$$\frac{X-a}{\sigma}\sim\mathcal{N}\left(0,1\right),$$

odnosno

$$\mathbf{P}\left(\left|X-a\right|<3\sigma\right)=\mathbf{P}\left(a-3\sigma<X<a+3\sigma\right)=P\left(-3<\frac{X-a}{\sigma}<3\right)=\Phi^*\left(3\right)=0.9973.$$

2. Broj sunčanih dana u nekom gradu u tijeku jedne godine je slučajna varijabla s matematičkim očekivanjem 75 dana. Pokažite da je vjerojatnost da u tijeku jedne godine u tom gradu ne bude više od 200 sunčanih dana veća od $\frac{5}{8}$.

Rješenje. Označimo našu slučajnu varijablu:

$$X=\text{ broj sunčanih dana u gradu tijekom jedne godine.}$$

Trebamo dokazati da vrijedi sljedeća ocjena

$$\mathbf{P}\left(X\leq200\right)>\frac{5}{8}.$$

Pri rješavanju ovog zadatka koristimo nejednakost Markova koja kaže da za slučajnu varijablu X koja poprima nenegativne vrijednosti i svaki $\varepsilon > 0$ vrijedi:

$$\mathbf{P}\left(X \geq \varepsilon\right) \leq \frac{\mathbf{E}\left(X\right)}{\varepsilon}.$$

Sada je prema Markovljevoj nejednakosti:

$$\mathbf{P}\left(X < \varepsilon\right) = 1 - \mathbf{P}\left(X \geq \varepsilon\right) \geq 1 - \frac{\mathbf{E}\left(X\right)}{\varepsilon} \implies \mathbf{P}\left(X \leq \varepsilon\right) > 1 - \frac{\mathbf{E}\left(X\right)}{\varepsilon}.$$

Uvrstimo li vrijednosti $\varepsilon = 200$ te $\mathbf{E}\left(X\right) = 75$ dobivamo:

$$\mathbf{P}\left(X \leq 200\right) > 1 - \frac{75}{200} = \frac{5}{8}.$$

3. Na ruletskom kolu obilježeno je ukupno 37 polja i to 18 crvenih, 18 crnih i 1 zeleno. Pretpostavimo da se igrač u jednoj igri (jednom okretanju ruleta) kladi u crveno ili crno polje. U slučaju pogotka boje polja dobiva 10 kuna, inače gubi 10 kuna. Kolika je vjerojatnost da nakon 100 klađenja igrač bude na dobitku? Ukoliko umjesto 10 kuna odredimo neki drugi iznos hoće li se ova vjerojatnost promijeniti?

Rješenje.

Označimo s X_i slučajnu varijablu koja modelira dobitak igrača u i-tom okretanju ruleta:

$$X_i \sim \begin{pmatrix} 10 & -10 \\ \frac{18}{37} & \frac{19}{37} \end{pmatrix}, \quad i = 1, \ldots, 100.$$

Ukoliko ulog od 10kn zamijenimo s a dobivamo općenitiju definiciju

$$X_i \sim \begin{pmatrix} a & -a \\ \frac{18}{37} & \frac{19}{37} \end{pmatrix}, \quad i = 1, \ldots, 100.$$

Računamo očekivanje, disperziju i standardnu devijaciju slučajnih varijabli X_i:

$$\begin{aligned} m &= \mathbf{E}\left(X_i\right) = \frac{-a}{37}, \\ \sigma^2 &= \mathbf{D}\left(X_i\right) = a^2\left(1 - \frac{1}{37^2}\right), \\ \sigma &= \sqrt{\mathbf{D}(X_i)} = a\sqrt{\left(1 - \frac{1}{37^2}\right)}. \end{aligned}$$

Slučajna varijabla

$$Y = \sum_{i=1}^{100} X_i$$

predstavlja iznos koji igrač dobiva, odnosno gubi, nakon 100 klađenja. Trebamo odrediti vjerojatnost $\mathbf{P}(Y > 0)$.

Pri rješavanju koristimo centralni granični teorem (CGT) koji kaže da za niz (X_n) jednako distribuiranih nezavisnih slučajnih varijabli s očekivanjem m i disperzijom σ^2 vrijedi:

$$\frac{\sum_{i=1}^{n} X_i - nm}{\sigma\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, 1\right). \tag{1}$$

U našem slučaju je

$$Y^* := \frac{Y + \frac{100a}{37}}{\frac{10\sqrt{1368}}{37}a} \sim \mathcal{N}(0,1),$$

te je tražena vjerojatnost jednaka

$$\mathbf{P}(Y > 0) = \mathbf{P}\left(\frac{Y + \frac{100}{37}a}{\frac{10\sqrt{1368}}{37}a} > \frac{\frac{100}{37}a}{\frac{10\sqrt{1368}}{37}a}\right) = \mathbf{P}\left(Y^* > \frac{10}{\sqrt{1368}}\right) = \frac{1}{2} - \frac{1}{2}\Phi^*(0.27037) = 0.3936.$$

Vidimo da ova vjerojatnost ne ovisi o iznosu a (što ne bismo mogli reći za vjerojatnost $\mathbf{P}(Y > d)$ za neki $d \neq 0$).

4. Koliko puta moramo baciti kocku da bi vjerojatnost da je aritmetička sredina dobivenih brojeva između 3 i 4 bila najmanje 0.95?

Rješenje.

Označimo s X slučajnu varijablu koja modelira jedno bacanje kocke. Ona ima jednoliku razdiobu:

$$X \sim \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix},$$

s očekivanjem

$$m := \mathbf{E}(X) = \frac{21}{6} = \frac{7}{2}$$

i disperzijom

$$\sigma^2 := \mathbf{D}(X) = \mathbf{E}\left(X^2\right) - \mathbf{E}(X)^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}.$$

Neka su za $i = 1, \ldots, n$, X_i nezavisne kopije slučajne varijable X, te neka je njihova aritmetička sredina

$$Y = \frac{\sum_{i=1}^n X_i}{n}.$$

Prema centralnom graničnom teoremu (1) vrijedi

$$\frac{\sum_{i=1}^n X_i - nm}{\sigma\sqrt{n}} \sim \mathcal{N}(0,1).$$

Dijeljenjem brojnika i nazivnika s n dobivamo

$$\frac{\frac{\sum_{i=1}^n X_i}{n} - m}{\sqrt{\frac{\sigma}{n}}} \sim \mathcal{N}(0,1)$$

odnosno

$$Y^* = \frac{Y - \frac{7}{2}}{\sqrt{\frac{35}{12n}}} \sim \mathcal{N}(0,1).$$

Ono što mi želimo jest pronaći najmanji n takav da vjerojatnost da slučajna varijable Y poprimi vrijednosti između 3 i 4 bude najmanje 0.95. Ta vjerojatnost želimo zapisati koristeći slučajnu varijablu Y^* s jediničnom normalnom razdiobom:

$$0.95 \leq \mathbf{P}(3 < Y < 4) = \mathbf{P}\left(\frac{3 - \frac{7}{2}}{\sqrt{\frac{35}{12n}}} < \frac{Y - \frac{7}{2}}{\sqrt{\frac{35}{12n}}} < \frac{4 - \frac{7}{2}}{\sqrt{\frac{35}{12n}}}\right) = \mathbf{P}\left(\left|\frac{Y - \frac{7}{2}}{\sqrt{\frac{35}{12n}}}\right| < \frac{\frac{1}{2}}{\sqrt{\frac{35}{12n}}}\right)$$

$$= \mathbf{P}\left(|Y^*| < \sqrt{\frac{3n}{35}}\right) = \Phi^*\left(\sqrt{\frac{3n}{35}}\right).$$

Odavde slijedi

$$\Phi^*\left(\sqrt{\frac{3n}{35}}\right) \geq 0.95 \implies \sqrt{\frac{3n}{35}} \geq 1.96 \implies n \geq 44.8187.$$

Pritom smo iz tablice (kvantila ili inverznim očitavanjem iz tablice za normalnu razdiobu) dobili $\Phi^*(x) \geq 0.95 \Rightarrow x \geq 1.96$. U konačnici rješenje je:

$$n \geq \lceil 44.81866 \rceil = 45.$$

5. Broj automobila pristiglih na jedan granični prijelaz je Poissonova slučajna varijabla s očekivanjem 72 automobila po satu. Izračunajte vjerojatnost da je tijekom jednog tjedna stiglo više od 12 tisuća automobila na promatrani granični prijelaz.

Rješenje.

Broj dolazaka automobila na granični prijelaz po **satu** modeliramo slučajnom varijablom X koja ima Poissonovu razidobu. Znamo da za Poissonovu slučajnu varijablu $X \sim \mathcal{P}(\lambda)$ vrijedi $\mathbf{E}(X) = \mathbf{D}(X) = \lambda$. Označimo

$$m = \mathbf{E}(X) = 72,\ \sigma = \sqrt{\mathbf{D}(X)} = \sqrt{72}.$$

Ako broj pristiglih automobila unutar i-tog sata opišemo slučajnom varijablom X_i koja je jednako distribuirana kao i slučajna varijable X, onda ukupni broj automobila koji su došli tijekom jednog tjedna iznosi

$$Y = \sum_{i=1}^{168} X_i.$$

Primjenom centralnog graničnog teorema dobivamo:

$$Y^* = \frac{Y - 168 \cdot 72}{\sqrt{168 \cdot 72}} \sim \mathcal{N}(0, 1).$$

Ono što tražimo je vjerojatnost

$$\mathbf{P}(Y > 12000) = \mathbf{P}\left(\frac{Y - 12096}{\sqrt{12096}} > \frac{12000 - 12096}{\sqrt{12096}}\right) = \mathbf{P}(Y^* > -0.8729) = 0.80864.$$

Pimijetimo da slučajna varijabla Y ima Poissonovu razdiobu s parametrom $n\lambda$, gdje je $n = 168$, te bismo bez upotrebe CGT-a trebali izračunati $1 - \sum_{k=0}^{12000} \frac{(n\lambda)^k}{k!} e^{-n\lambda}$.

6. Nezavisne slučajne varijable $X_1, X_2, ..., X_{100}$ su zadane funkcijom gustoće

$$f(x) = \begin{cases} x, & x \in [0, 1], \\ 2 - x, & x \in [1, 2]. \end{cases}$$

(a) Izračunajte $\mathbf{E}(X_i)$ i $\mathbf{D}(X_i)$.

(b) Neka je $\overline{X} = \frac{1}{100} \sum_{i=1}^{100} X_i$. Koristeći centralni granični teorem, odredite vjerojatnost

$$\mathbf{P}\left(0.9 < \overline{X} < 1.02\right).$$

Rješenje.

(a) Računamo očekivanje

$$\mathbf{E}\left(X_i\right) = \int_0^1 x^2 \,\mathrm{d}x + \int_1^2 x\left(2-x\right)\mathrm{d}x = 1,$$

a zatim disperziju

$$\mathbf{D}\left(X_i\right) = \mathbf{E}(X_i^2) - \mathbf{E}\left(X_i\right)^2 = \int_0^1 x^3 \,\mathrm{d}x + \int_1^2 x^2\left(2-x\right)\mathrm{d}x - 1 = \frac{7}{6} - 1 = \frac{1}{6}.$$

(b) Kao zadatku **4**. opet se pojavljuje aritmetička sredina te problemu pristupamo na jednak način. Ovdje koristimo CGT:

$$\overline{X} = \frac{1}{100}\sum_{i=1}^{100} X_i \implies \frac{\overline{X} - \mathbf{E}\left(X_i\right)}{\sqrt{\frac{\mathbf{D}(X_i)}{n}}} \sim \mathcal{N}\left(0,1\right) \implies \frac{\overline{X}-1}{\sqrt{\frac{1}{600}}} \sim \mathcal{N}\left(0,1\right).$$

Neka je

$$Y = \frac{\overline{X}-1}{\sqrt{\frac{1}{600}}} \sim \mathcal{N}\left(0,1\right).$$

Tada je

$$\begin{aligned}\mathbf{P}\left(0.9 < \overline{X} < 1.02\right) &= \mathbf{P}\left(\frac{-0.1}{\frac{1}{10\sqrt{6}}} < Y < \frac{0.02}{\frac{1}{10\sqrt{6}}}\right) = \mathbf{P}\left(-\sqrt{6} < Y < \frac{\sqrt{6}}{5}\right) \\ &= \frac{1}{2}\left(\Phi^*\left(\frac{\sqrt{6}}{5}\right) + \Phi^*\left(\sqrt{6}\right)\right) = 0.68079.\end{aligned}$$