# Matematika s statistiko
# Verjetnost

### Slučajne spremenljivke

(1) V posodi so štiri kroglice s številkami $-2$, $-1$, $0$ in $1$. Iz posode naključno izvlečemo eno kroglico in označimo z $X$ število na njej.

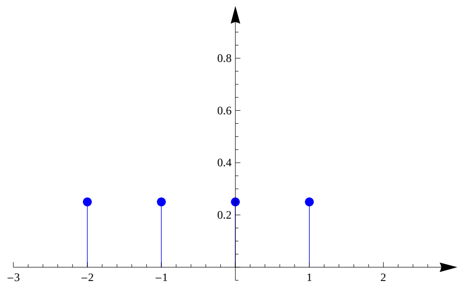
(a) Izračunaj porazdelitvi slučajnih spremenljivk $X$ in $X^2$.

(b) Izračunaj povprečno vrednost slučajne spremenljivke $X$.

*Rešitev:* (a) Slučajna spremenljivka je funkcija, katere vrednost je odvisna od izida poskusa. Zanima nas predvsem, kakšne so verjetnosti, da zavzame neko vrednost. Če iz posode naključno izvlečemo eno kroglico, se lahko zgodi, da bo na njej število $-2$, $-1$, $0$ ali pa $1$. Porazdelitev slučajne spremenljivke $X$ lahko predstavimo s tabelo

$$X : \begin{pmatrix} -2 & -1 & 0 & 1 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}$$

ali pa grafično s histogramom




Slučajna spremenljivka $X^2$ pa lahko zavzame vrednosti $0$, $1$ ali $4$. Njena porazdelitev je

$$X^2 : \begin{pmatrix} 0 & 1 & 4 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{pmatrix}.$$

Možne vrednosti spremenljivk $X$ in $X^2$ v obeh primerih tvorijo neko zaporedje števil. Takšnim slučajnim spremenljivkam rečemo diskretne slučajne spremenljivke.

(b) Povprečna vrednost diskretne slučajne spremenljivke $X$ je definirana s predpisom

$$E(X) = \sum_k x_k \cdot P(X = x_k),$$

kjer vsota teče po vseh možnih vrednostih slučajne spremenljivke $X$. V našem primeru je

$$\begin{aligned} E(X) &= -2 \cdot \frac{1}{4} - 1 \cdot \frac{1}{4} + 0 + 1 \cdot \frac{1}{4}, \\ &= \frac{-2 - 1 + 0 + 1}{4}, \\ &= -\frac{1}{2}. \end{aligned}$$

□

(2) V podjetju je zaposlenih 5 moških in 5 žensk. Za delo na projektu naključno izberejo skupino treh ljudi. Z $X$ označimo število žensk v tej skupini. Izračunaj porazdelitveno shemo in povprečno vrednost slučajne spremenljivke $X$.

*Rešitev:* Možne vrednosti spremenljivke $X$ so 0, 1, 2 in 3, posamezne verjetnosti pa so:

$$\begin{aligned} P(X=0) &= \frac{\binom{5}{0}\binom{5}{3}}{\binom{10}{3}} = \frac{1}{12}, \\ P(X=1) &= \frac{\binom{5}{1}\binom{5}{2}}{\binom{10}{3}} = \frac{5}{12}, \\ P(X=2) &= \frac{\binom{5}{2}\binom{5}{1}}{\binom{10}{3}} = \frac{5}{12}, \\ P(X=3) &= \frac{\binom{5}{3}\binom{5}{0}}{\binom{10}{3}} = \frac{1}{12}. \end{aligned}$$

Porazdelitev slučajne spremenljivke $X$ lahko predstavimo s tabelo

$$X : \begin{pmatrix} 0 & 1 & 2 & 3 \\ \frac{1}{12} & \frac{5}{12} & \frac{5}{12} & \frac{1}{12} \end{pmatrix}.$$

Povprečna vrednost slučajne spremenljivke $X$ je tako enaka

$$E(X) = 0 \cdot \frac{1}{12} + 1 \cdot \frac{5}{12} + 2 \cdot \frac{5}{12} + 3 \cdot \frac{1}{12} = 1.5.$$

To pomeni, da bo v povprečju v skupini enako število žensk kot moških. Seveda pa v nobeni skupini števili moških in žensk ne moreta biti dejansko enaki. $\square$

(3) Igralec igra igro s kocko. Za sodo število pik dobi toliko evrov, kolikor pik je vrgel, za liho število pik pa vedno plača 3 evre. Njegov dobitek je slučajna spremenljivka $X$.

(a) Izračunaj povprečno vrednost slučajne spremenljivke $X$.

(b) Koliko v povprečju igralec zasluži oziroma izgubi, če igra deset iger?

*Rešitev:* (a) Možne vrednosti, ki jih zavzame slučajna spremenljivka $X$, so $-3$, 2, 4 in 6, njena porazdelitev pa je

$$X : \begin{pmatrix} -3 & 2 & 4 & 6 \\ \frac{1}{2} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}.$$

Od tod sklepamo, da je povprečna vrednost slučajne spremenljivke $X$ enaka

$$E(X) = -3 \cdot \frac{1}{2} + 2 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 0.5.$$

Pričakovan dobiček je pozitiven, zato se igralcu to igro splača igrati.

(b) Če igralec to igro ponovi desetkrat, v povprečju zasluži $E(10X) = 5$ evrov. $\square$

(4) Krog s polmerom 1 vržemo v notranjost kroga s polmerom 5, tako da je oddaljenost $R$ med središčema krogov porazdeljena po zakonu

$$p(r) = \tfrac{r}{8},\ r \in [0,4].$$

(a) Izračunaj verjetnost, da manjši krog pokrije središče večjega.

(b) Kolikšna je v povprečju razdalja med središčema krogov?

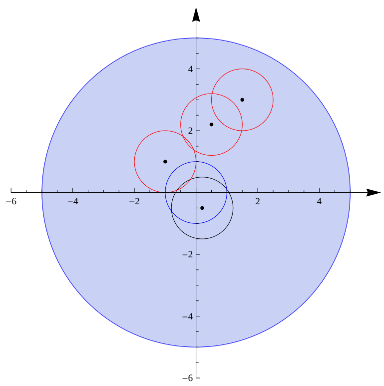
*Rešitev:* (a) Ko vržemo krog s polmerom 1 v notranjost kroga s polmerom 5, pristane njegovo središče nekje v krogu s polmerom 4. Razdalja med središčema $R$ leži na intervalu $[0,4]$, kar pomeni, da imamo opravka z zvezno slučajno spremenljivko.

Ko imamo opravka z zvezno slučajno spremenljivko, moramo za izračun verjetnosti vsote nadomestiti z integrali. To pomeni, da je

$$P(a \le R \le b) = \int_a^b p(r)\, dr.$$

Funkciji $p$ rečemo gostota zvezne slučajne spremenljivke.

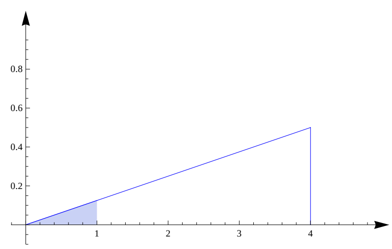
V našem primeru nas zanima verjetnost, da manjši krog pokrije središče večjega.




To se bo zgodilo, če bo $R \in [0,1]$, zato je ta verjetnost enaka

$$P(0 \le R \le 1) = \int_0^1 p(r)\, dr = \int_0^1 \frac{r}{8}\, dr = \frac{r^2}{16}\Big|_0^1 = \frac{1}{16}.$$

Grafično ta verjetnost ustreza ploščini lika pod grafom funkcije $p$ na intervalu $[0,1]$.




Ploščina tega koščka je enaka $\frac{1}{16}$, celoten lik pod grafom funkcije $p$ pa ima ploščino 1.

(b) Povprečno vrednost zvezne slučajne spremenljivke lahko izračunamo s formulo

$$E(R) = \int_{-\infty}^{\infty} r\, p(r)\, dr.$$

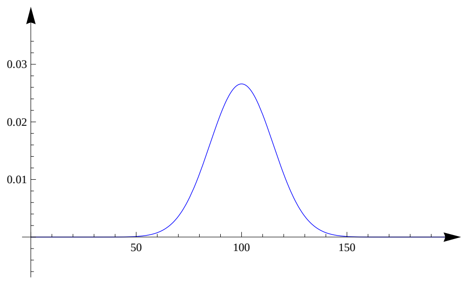
V praksi potem integriramo samo po intervalu, kjer je gostota neničelna. Tako dobimo

$$E(R) = \int_0^4 r\,p(r)\,dr = \int_0^4 \frac{r^2}{8}\,dr = \left.\frac{r^3}{24}\right|_0^4 = \frac{8}{3}.$$

□

(5) Privzemimo, da je IQ prebivalstva slučajna spremenljivka, ki je porazdeljena po zakonu $N(100, 15)$. Oceni približno število ljudi v Sloveniji z IQ-jem: med 85 in 115 ter nad 130.

*Rešitev:* Z $N(\mu, \sigma)$ označimo Gaussovo normalno porazdelitev s povprečno vrednostjo $\mu$ in s standardnim odklonom $\sigma$. V našem primeru je $\mu = 100$ in $\sigma = 15$, graf gostote pa je narisan na spodnji sliki.




Pri računanju ploščin likov pod Gaussovimi krivuljami uporabljamo funkcijo $\Phi$, katere vrednosti so zbrane v tabeli. Limita funkcije $\Phi$ je $\lim_{x\to\infty} \Phi(x) = 0.5$, za negativne vrednosti pa velja

$$\Phi(x) = -\Phi(-x).$$

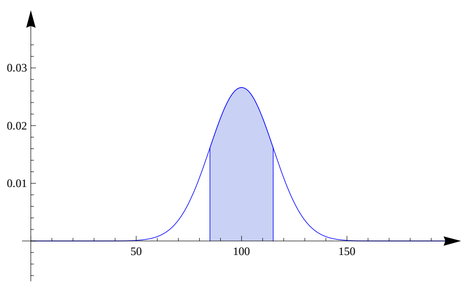
Velja

$$P(a \leq N(\mu, \sigma) \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

Ocena za verjetnost, da ima naključno izbrana oseba $IQ$ med 85 in 115 je tako enaka

$$P(85 \leq N(100, 15) \leq 115) = \Phi\left(\frac{115-100}{15}\right) - \Phi\left(\frac{85-100}{15}\right) = \Phi(1) - \Phi(-1) = 0.68268.$$

Ta verjetnost je enaka ploščini lika na spodnji sliki.



Od tod lahko ocenimo, da je v Sloveniji približno

$$n \approx 2000000 \cdot 0,68268 \approx 1370000$$

ljudi z $IQ$-jem med 85 in 115. Za IQ nad 130 pa velja

$$P(130 \leq N(100, 15)) = \Phi(\infty) - \Phi(2) = 0.0228 \Longrightarrow n \approx 45000.$$

$\square$

(6) V posodi so tri kroglice s številkami $-1$, 0 in 1. Iz posode naključno izvlečemo eno kroglico, jo vrnemo v posodo in nato naključno izvlečemo še eno kroglico. Označimo z $X$ število na prvi izvlečeni kroglici z $Y$ pa vsoto števil na obeh izvlečenih kroglicah. Izračunaj kovarianco in korelacijski koeficient spremenljivk $X$ in $Y$.

*Rešitev:* Kadar imamo opravka z več slučajnimi spremenljivkami hkrati, nas zanima, ali so med sabo povezane. Če sta $X$ in $Y$ slučajni spremenljivki, merimo njuno medsebojno povezanost s kovarianco in pa s korelacijskim koeficientom:

$$\begin{aligned} K(X, Y) &= E(XY) - E(X)E(Y), \\ r(X, Y) &= \frac{K(X, Y)}{\sigma(X)\sigma(Y)}. \end{aligned}$$

Najprej bomo izračunali povprečne vrednosti slučajnih spremenljivk $X$, $Y$ in $XY$, nato pa še standardna odklona spremenljivk $X$ in $Y$.

Če dvakrat zapored izvlečemo po eno kroglico, je skupaj možnih 9 izidov. Vrednost slučajnih spremenljivk $X$, $Y$ in $XY$ pri vsakem izmed izidov bomo predstavili v spodnji tabeli (oklepaj $(-1, 0)$ npr. pomeni, da smo najprej izvlekli $-1$ nato pa 0):

| | $(-1,-1)$ | $(-1,0)$ | $(-1,1)$ | $(0,-1)$ | $(0,0)$ | $(0,1)$ | $(1,-1)$ | $(1,0)$ | $(1,1)$ |
|---|---|---|---|---|---|---|---|---|---|
| $X$ | $-1$ | $-1$ | $-1$ | 0 | 0 | 0 | 1 | 1 | 1 |
| $Y$ | $-2$ | $-1$ | 0 | $-1$ | 0 | 1 | 0 | 1 | 2 |
| $XY$ | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 2 |

Vidimo, da lahko $X$ zavzame vrednosti $-1$, 0 in 1, $Y$ pa vrednosti $-2$, $-1$, 0, 1 in 2. Skupno porazdelitev spremenljivk $X$ in $Y$ predstavimo s tabelo:

| X/Y | $-2$ | $-1$ | 0 | 1 | 2 |
|---|---|---|---|---|---|
| $-1$ | 1/9 | 1/9 | 1/9 | 0 | 0 |
| 0 | 0 | 1/9 | 1/9 | 1/9 | 0 |
| 1 | 0 | 0 | 1/9 | 1/9 | 1/9 |

porazdelitve slučajnih spremenljivk $X$, $Y$ in $XY$ pa s tabelami:

$$X : \begin{pmatrix} -1 & 0 & 1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}, Y : \begin{pmatrix} -2 & -1 & 0 & 1 & 2 \\ \frac{1}{9} & \frac{2}{9} & \frac{3}{9} & \frac{2}{9} & \frac{3}{9} \end{pmatrix}, XY : \begin{pmatrix} 0 & 1 & 2 \\ \frac{5}{9} & \frac{2}{9} & \frac{2}{9} \end{pmatrix}.$$

Od tod sledi $E(X) = E(Y) = 0$ in $E(XY) = \frac{2}{3}$, kar pomeni, da je kovarianca $X$ in $Y$ enaka

$$K(X, Y) = E(XY) - E(X)E(Y) = \frac{2}{3}.$$

Za izračun korelacijskega koeficienta moramo izračunati standardna odklona spremenljivk $X$ in $Y$. Standardni odklon $X$ je definiran s predpisom

$$\sigma(X) = \sqrt{V(X)},$$

kjer je

$$V(X) = \sum_k (x_k - E(X))^2 \cdot P(X = x_k) = E(X^2) - E(X)^2$$

varianca slučajne spremenljivke $X$.
Velja $E(X^2) = \frac{2}{3}$, $E(Y^2) = \frac{4}{3}$, $V(X) = \frac{2}{3}$ in $V(Y) = \frac{4}{3}$. Standardna odklona spremenljivk $X$ in $Y$ sta torej enaka $\sigma(X) = \sqrt{\frac{2}{3}}$ in $\sigma(Y) = \sqrt{\frac{4}{3}}$, zato je korelacijski koeficient teh dveh spremenljivk enak

$$r(X,Y) = \frac{K(X,Y)}{\sigma(X)\sigma(Y)} = \frac{\frac{2}{3}}{\sqrt{\frac{2}{3}}\sqrt{\frac{4}{3}}} = \frac{\sqrt{2}}{2} = 0.71.$$

Korelacijski koeficient dveh slučajnih spremenljivk je zmeraj neko število na intervalu $[-1,1]$. Če je enak 0, rečemo, da sta spremenljivki nekorelirani. Če je vrednost pozitivna, sta spremenljivki povezani, in sicer tako, da obe hkrati ali naraščata ali pa padata. V primeru negativnega koeficienta pa pri padanju ene spremenljivke druga spremenljivka narašča. Če sta vrednosti $\pm 1$, je zveza med spremenljivkama linearna.

V našem primeru je koeficient pozitiven, kar pomeni, da večja vrednost $X$ implicira večjo vrednost $Y$ in obratno. □

(7) Z Laplaceovo aproksimacijo približno izračunaj naslednji verjetnosti:

(a) V 100 metih kovanca pade grb med 40 in 60-krat.
(b) V 4500 metih kocke pade šestica med 750 in 800-krat.

*Rešitev:* Natančne verjetnosti binomske porazdelitve je pri velikem številu poskusov težko računati. Z Laplaceovo aproksimacijo pa lahko binomsko porazdelitev dobro aproksimiramo z ustrezno normalno porazdelitvijo. Pri velikih $n$ uporabljamo aproksimacijo

$$\mathrm{Bin}(n,p) \sim N(np, \sqrt{np(1-p)}).$$

V praksi to pomeni, da za binomsko porazdeljeno slučajno spremenljivko $X$ s parametroma $n$ in $p$ velja

$$P(a \le X \le b) \approx \Phi\left(\frac{b - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - np}{\sqrt{np(1-p)}}\right).$$

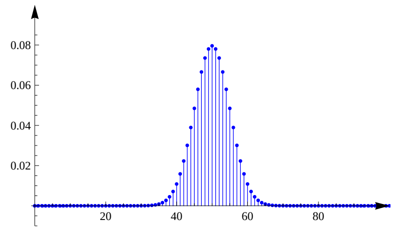
(a) Naj slučajna spremenljivka $X$ označuje, koliko grbov pade v 100 metih kovanca. Potem je $X$ porazdeljena binomsko s parametroma $n = 100$ in $p = 1/2$. Sledi

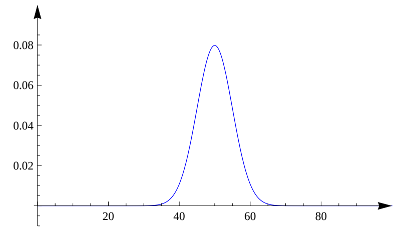
$$P(40 \le X \le 60) \approx \Phi\left(\frac{60-50}{\sqrt{25}}\right) - \Phi\left(\frac{40-50}{\sqrt{25}}\right) = 2\Phi(2) = 0.9544.$$

Če bi šli izračunati natančno vrednost z uporabo Bernoullijeve formule, bi dobili

$$P(40 \le X \le 60) = P(X = 40) + P(X = 41) + \cdots + P(X = 60) = 0.9648.$$

Poglejmo si še histogram binomske porazdelitve $\mathrm{Bin}(100, 1/2)$ in pa gostoto normalne porazdelitve $N(50, 5)$. Vidimo, da se ustrezna Gaussova funkcija zelo prilega histogramu binomske porazdelitve.




(b) Naj sedaj $X$ označuje, koliko šestic je padlo v 4500 metih kovanca. Potem je $X$ porazdeljena binomsko s parametroma $n = 4500$ in $p = 1/6$. Sledi

$$P(750 \leq X \leq 800) \approx \Phi\left(\frac{800-750}{\sqrt{625}}\right) - \Phi\left(\frac{750-750}{\sqrt{625}}\right) = \Phi(2) - \Phi(0) = 0.4722.$$

Natančna vrednost pa je tokrat

$$P(750 \leq X \leq 800) = 0.4838.$$

□

(8) Na izpitu je 50 vprašanj izbirnega tipa, študent pa izpit opravi, če pravilno odgovori na vsaj 25 vprašanj. Denimo, da študent pride na izpit popolnoma nepripravljen in pri vsakem vprašanju naključno izbere enega izmed treh možnih odgovorov. Z Laplaceovo aproksimacijo približno izračunaj verjetnost, da študent opravi izpit.

*Rešitev:* Če študent popolnoma naključno izbira odgovore nalog, je situacija podobna kot pri metu kovanca ali pa kocke. Gre za zaporedje izbir odgovorov, pri čemer študent pri vsaki nalogi z isto verjetnostjo ugane pravilen rezultat.

Naj slučajna spremenljivka $X$ označuje, koliko pravilnih odgovorov študent ugane, če so pri vsaki nalogi možni trije odgovori. Potem je $X$ porazdeljena binomsko s parametroma $n = 50$ in $p = 1/3$. Sledi

$$P(25 \leq X \leq 50) \approx \Phi\left(\frac{50-\frac{50}{3}}{\sqrt{50\cdot\frac{1}{3}\cdot\frac{2}{3}}}\right) - \Phi\left(\frac{50-\frac{50}{3}}{\sqrt{50\cdot\frac{1}{3}\cdot\frac{2}{3}}}\right) = \Phi(10) - \Phi(2.5) = 0.0062.$$

Natančna vrednost pa je

$$P(25 \leq X \leq 50) = 0.0108.$$

Vidimo, da študent nima prav veliko možnosti, da opravi izpit, če ugiba odgovore. □

(9) Igralec igra igro s kocko. Za sodo število pik dobi toliko evrov, kolikor pik je vrgel, za liho število pik pa vedno plača 3 evre. Njegov dobitek je slučajna spremenljivka $X$. Oceni verjetnost, da bo po desetih igrah imel pozitiven izkupiček.

*Rešitev:* Možne vrednosti, ki jih zavzame slučajna spremenljivka $X$, so $-3$, 2, 4 in 6, njena porazdelitev pa je

$$X : \begin{pmatrix} -3 & 2 & 4 & 6 \\ \frac{1}{2} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}.$$

Od tod sklepamo, da je povprečna vrednost slučajne spremenljivke $X$ enaka

$$E(X) = -3 \cdot \frac{1}{2} + 2 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 0.5.$$

Izračunajmo še varianco in standardni odklon. Najprej je

$$\begin{aligned} E(X^2) &= 9 \cdot \frac{1}{2} + 4 \cdot \frac{1}{6} + 16 \cdot \frac{1}{6} + 36 \cdot \frac{1}{6}, \\ &= \frac{83}{6}. \end{aligned}$$

Od tod sledi

$$\begin{aligned} &V(X) = E(X^2) - E(X)^2 = \frac{83}{6} - \frac{1}{4} = \frac{163}{12}, \\ &\sigma(X) = 3.69. \end{aligned}$$

Označimo z $Y$ dobiček po $n$ igrah. Potem imamo oceno

$$P(a \leq Y \leq b) \approx \Phi\left(\frac{b - nE(X)}{\sigma(X)\sqrt{n}}\right) - \Phi\left(\frac{a - nE(X)}{\sigma(X)\sqrt{n}}\right).$$

V našem primeru je $n = 10$, kar nam da

$$P(Y \geq 0) = P(0 \leq Y < \infty) \approx \Phi(\infty) - \Phi\left(\frac{0-5}{3.69\sqrt{10}}\right) = \frac{1}{2} - \Phi(-0.43) = 0.67.$$

Natančna vrednost je $P(Y \geq 0) = 0.68$. $\square$