Poglavlje 2

Slučajne varijable

Primjeri u Excelu vezani za ovu cjelinu nalaze se u dokumentu *Slucajnevarijable.xlsx*.

Slučajna varijabla je funkcija koja svakom mogućem ishodu nekog pokusa pridružuje realni broj. Jednostavnijim riječima, vrijednost slučajne varijable možemo promatrati kao numerički ishod slučajnog eksperimenta. Glavna podjela slučajnih varijabli je na diskretne i kontinuirane. **Diskretna** slučajna varijabla je ona ǩojoj je skup vrijednosti konačan ili prebrojiv, a **kontinuirana** je ona ǩojoj je skup vrijednosti neprebrojiv. Za početak se bavimo diskretnim slučajnim varijablama.

2.1 Diskretna slučajna varijabla

Primjer 2.1
Bacamo tri novčića. (S jedne strane svakog novčića je pismo, s druge glava.) Slučajna varijabla X bilježi broj dobivenih pisama. Pogledajmo koje se sve mogućnosti mogu dogoditi.

Ishod	Broj pisama
GGG	0
GGP	1
GPG	1
PGG	1
GPP	2
PGP	2
PPG	2
PPP	3

Na svakom novčiću imamo dvije mogućnosti, a imamo tri novčića. Stoga je broj svih mogućih ishoda jednak $2 \cdot 2 \cdot 2 = 8$. Skup vrijednosti slučajne varijable X je $\{0, 1, 2, 3\}$. Uočimo da se u tablici neke vrijednosti pojavljuju više puta, tj. nekim je ishodima pridružena ista numerička vrijednost. Svakoj vrijednosti slučajne varijable, x_i, pridružena je njena vjerojatnost, $P(X=x_i)$. Ta je vjerojatnost jednaka

$$P(X=x_i) = \frac{\text{broj ishoda koji daju vrijednost } x_i}{\text{broj svih mogućih ishoda}}.$$

Nula pisama možemo dobiti na samo jedan način (GGG) pa je $P(X = 0) = \frac{1}{8}$. Jedno pismo možemo dobiti na tri načina (GGP, GPG, PGG) pa je $P(X=1) = \frac{3}{8}$. Na isti se način dobiju vjerojatnosti za preostale vrijednosti slučajne varijable.

X	$P(X=x_i)$
0	$\frac{1}{8}$
1	$\frac{3}{8}$
2	$\frac{3}{8}$
3	$\frac{1}{8}$

Vjerojatnosti $P(X=x_i)$ su brojevi između 0 i 1. Možemo ih interpretirati i kao postotke. Ako će se nešto sigurno dogoditi, to ima vjerojatnost 1 (100%), a ako se sigurno neće dogoditi, ima vjerojatnost 0 (0%). Zbroj vjerojatnosti svih vrijednosti slučajne varijable je 1.

Očekivanje diskretne slučajne varijable X s vrijednostima $x_1, x_2, \ldots, x_n$ definira se kao

$$E(X) = \sum_{i=1}^{n} x_i P(x_i) = x_1 P(x_1) + x_2 P(x_2) + \cdots + x_n P(x_n),$$

a njena **varijanca** (disperzija) iznosi

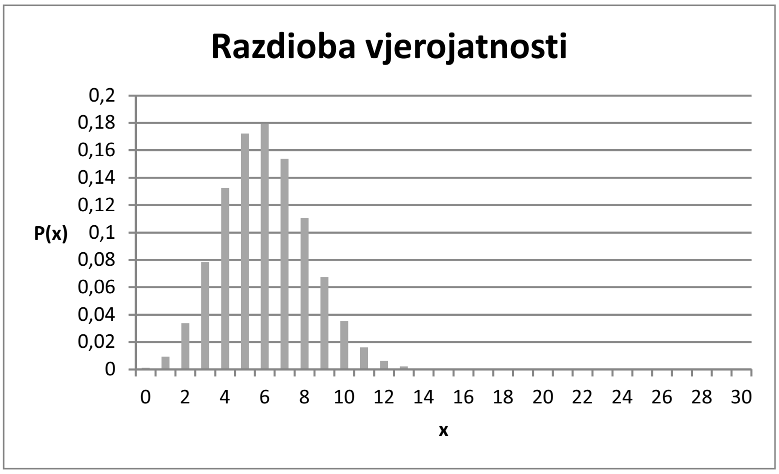
$$V(X) = \sum_{i=1}^{n} (x_i - E(X))^2 P(x_i) = (x_1 - E(X))^2 P(x_1) + \cdots + (x_n - E(X))^2 P(x_n).$$

U prethodnom primjeru očekivanje iznosi

$$E(X) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = 1.5,$$

a varijanca

$$V(X) = (0 - 1.5)^2 \cdot \frac{1}{8} + (1 - 1.5)^2 \cdot \frac{3}{8} + (2 - 1.5)^2 \cdot \frac{3}{8} + (3 - 1.5)^2 \cdot \frac{1}{8} = 0.75.$$



Slika 2.1: Graf razdiobe vjerojatnosti binomne slučajne varijable s parametrima $n = 30$, $p = 0.2$.

2.1.1 Binomna razdioba

Binomna slučajna varijabla je diskretna slučajna varijabla koja se generira tako što se n puta ponavlja isti pokus. Slučajna varijabla registrira koliko je puta pokus uspio, odnosno, koliko se puta dogodio neki fiksirani događaj A koji ima vjerojatnost p. Parametri n i p određuju binomnu slučajnu varijablu X i pišemo

$$X \sim B(n, p).$$

Kako ova slučajna varijaba registrira koliko se puta, od n pokušaja, dogodio A, lako je vidjeti da su njene vrijednosti cijeli brojevi između 0 i n. U najgorem se slučaju događaj A pojavio 0 puta, a u najboljem slučaju svih n puta.

Vjerojatnost svakog pojedinog ishoda i iznosi

$$P(X=i) = \binom{n}{i} p(1-p)^{n-i}, \qquad i = 0, \ldots, n.$$

U Excelu za računanje vjerojatnosti $P(X=i)$ koristimo naredbu

=BINOMDIST(i; n; p; FALSE).

Pomoću iste naredbe možemo računati i vjerojatnost da je slučajna varijabla manja ili jednaka od neke vrijednosti i, $P(X \leq i)$,

=BINOMDIST(i; n; p; TRUE).

Očekivanje binomne slučajne varijable X dano je formulom

$$E(X) = np,$$

a varijanca

$$V(X) = np(1-p).$$

Primjer 2.2
Bacamo kocku 20 puta. (Brojevi na kocki su od 1 do 6.) Slučajna varijabla X bilježi broj dobivenih jedinica. Ovo je binomna slučajna varijabla. Odredimo njene parametre. S obzirom da se pokus ponavlja 20 puta, $n = 20$. Vjerojatnost da dobijemo jedinicu u jednom bacanju je $\frac{1}{6}$ jer je jedinica jedna od ukupno 6 mogućnosti. Stoga je $p = \frac{1}{6}$, tj. $X \sim B(20, \frac{1}{6})$.

Primjer 2.3
Bacamo kockicu 5 puta. Slučajna varijabla Y bilježi broj dobivenih parnih brojeva. Ovo je također binomna slučajna varijabla. Parametri su joj $n = 5$ i $p = \frac{3}{6} = \frac{1}{2}$, jer su tri od šest brojeva na kocki parni.

2.1.2 Poissonova razdioba

Poissonova slučajna varijabla je diskretna slučajna varijabla koja broji koliko se puta pojavio fiksni događaj A. Međutim, za razliku od binomne, Poissonova slučajna varijabla ima prebrojiv (beskonačan) skup vrijednosti

$$\{0, 1, 2, 3, \ldots\}.$$

Poissonova slučajna varijabla određena je parametrom a,

$$X \sim P(a),$$

za koji vrijedi

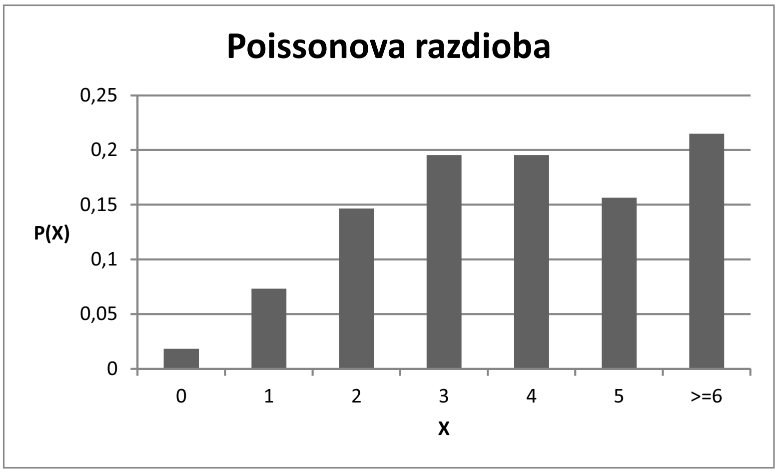
$$P(X=i) = e^{-a}\frac{a^i}{i!}, \qquad i = 0, 1, 2, 3, \ldots$$

Računanje u Excelu je slično kao za binomnu slučajnu varijablu. Da dobijemo vrijednost $P(X=i)$ u Excelu koristimo naredbu

=POISSON(i; a; FALSE),

a za računanje $P(X \leq i)$,

=POISSON(i; a; TRUE).



Slika 2.2: Graf razdiobe vjerojatnosti Poissonove slučajne varijable s parametrom $a = 4$.

Za očekivnje i varijancu Poissonove slučajne varijable X vrijedi

$$E(X) = a, \qquad V(X) = a.$$

Primjer 2.4
Pretpostavimo da Marija prosječno dobije 4 poruke na sat. Slučajna varijabla koja bilježi broj poruka koje je Marija dobila unutar jednog sata je Poissonova s parametrom 4, tj. $X \sim P(4)$.

2.2 Kontinuirana slučajna varijabla

Skup vrijednosti kontinuirane slučajne varijable je neprebrojiv. Ovakva slučajna varijabla poprima sve vrijednosti u nekom intervalu $[a, b]$. **Razdioba vjerojatnosti** na tom intervalu zadana je funkcijom f za koju vrijedi

(i) $f(x) \geq 0, \quad x \in [a, b]$,

(ii) $\int_a^b f(x)\, dx = 1$.

Vjerojatnost da slučajna varijabla poprimi vrijednost iz intervala $[a_0, b_0]$ jednaka je

$$P(a_0 < X < b_0) = \int_{a_0}^{b_0} f(x)\, dx.$$

Funkciju f nazivamo **funkcija gustoće** vjerojatnosti.

Uz funkciju gustoće definira se i **funkcija distribucije** vjerojatnosti normalne slučajne varijable X u x. To je vjerojatnost da je $X < x$. Funkcija distribucije je primitivna funkcija funkcije gustoće vjerojatnosti. Točnije, vrijedi $F'(x) = f(x)$ osim za zanemarivo mnogo vrijednosti x. Računa se po formuli

$$F(x) = \int_{-\infty}^{x} f(t)\ dt,$$

Očekivanje kontinuirane slučajne varijable X je integral

$$E(X) = \int_{-\infty}^{\infty} f(x)\ dx,$$

a njena **varijanca** je

$$V(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x)\ dx.$$

2.2.1 Normalna razdioba

Normalna (Gaussova) slučajna varijabla je kontinuirana slučajna varijabla. Normalna slučajna varijabla X određena je s dva parametra, očekivanjem μ i standardnom devijacijom σ (odnosno varijancom σ^2), te pišemo

$$X \sim N(\mu, \sigma^2).$$

Funkcija gustoće vjerojatnosti normalne slučajne varijable dana je izrazom

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

a u Excelu se računa naredbom

$$\text{=NORMDIST}(x;\ \mu;\ \sigma;\ \text{FALSE}).$$

Funkcija distribucije dobije se naredbom

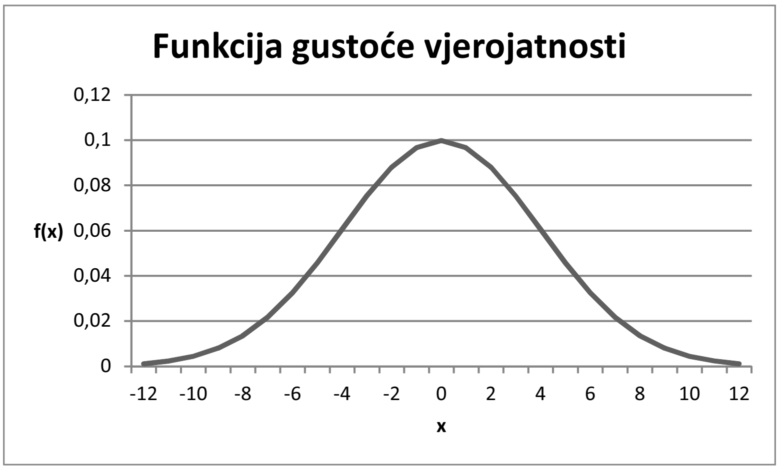
$$\text{=NORMDIST}(x;\ \mu;\ \sigma;\ \text{TRUE}).$$

Interval tri sigme je interval oko očekivanja μ, od 3σ lijevo od očekivanja do 3σ desno od očekivanja, tj.

$$[\mu - 3\sigma, \mu + 3\sigma].$$

Unutar tog intervala nalaze se gotovo sve vrijednosti normalno distribuirane slučajne varijable X, njih 99.7%. Intervali jedna sigma i dvije sigme dobiju se analogno kao $[\mu - \sigma, \mu + \sigma]$, odnosno $[\mu - 2\sigma, \mu + 2\sigma]$.

Uobičajeni primjeri normalne slučajne varijable su varijable koje registriraju rezultate nekog mjerenja ili grešku mjerenja.



Slika 2.3: Graf funkcije gustoće vjerojatnosti normalne slučajne varijable s parametrima $\mu = 0$, $\sigma = 4$.

2.2.2 Eksponencijalna razdioba

Eksponencijalna slučajna varijabla je kontinuirana slučajna varijabla određena parameterom λ,

$$X \sim E(\lambda).$$

Ona poprima sve vrijednosti u iz skupa $\langle 0, \infty \rangle$. Funkcija gustoće vjerojatnosti eksponencijalne slučajne varijable dana je izrazom

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

Iz toga se izračuna da je funkcija distribucije vjerojatnosti

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

U Excelu se vrijednosti ovih funkcija računaju naredbom

=EXPONDIST(x; λ; FALSE),

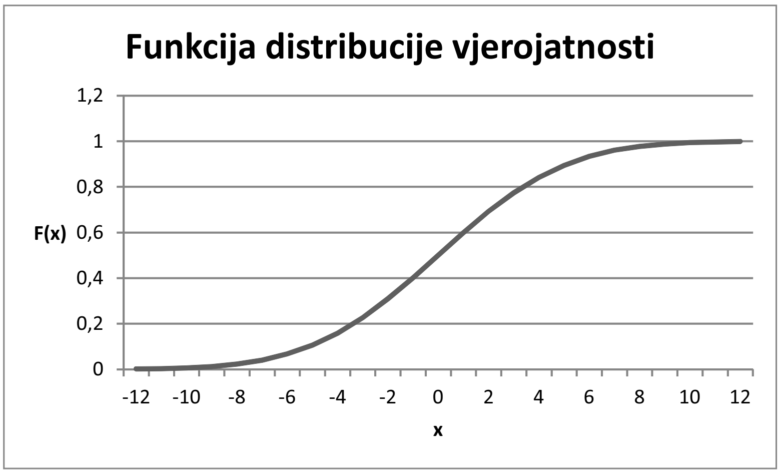
odnosno

=EXPONDIST(x; λ; TRUE).

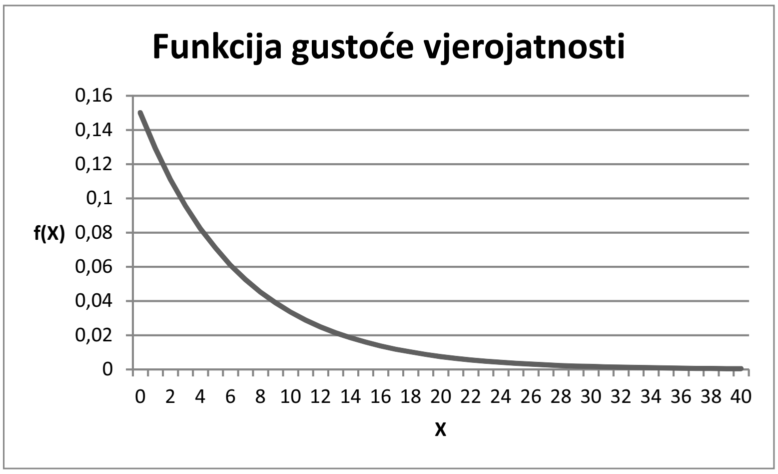
Za očekivnje i varijancu Poissonove slučajne varijable X vrijedi

$$E(X) = \frac{1}{\lambda}, \qquad V(X) = \frac{1}{\lambda^2}.$$

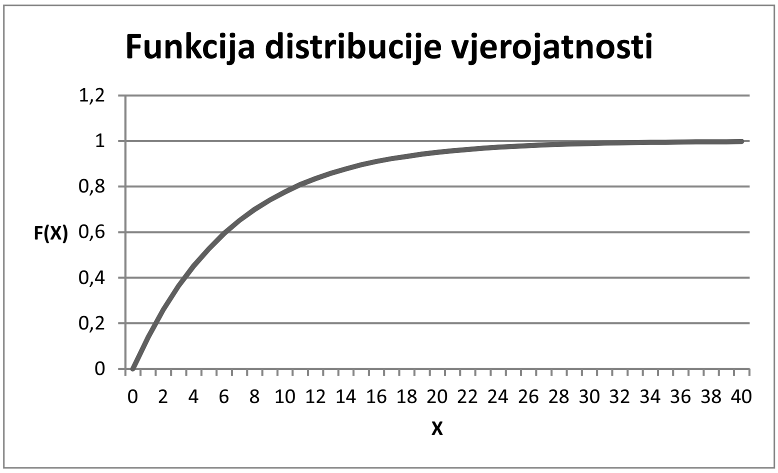
Primjer eksponencijalne slučajne varijale je varijabla koja bilježi vrijeme proteklo do pojave nekog događaja.



Slika 2.4: Graf funkcije distribucije vjerojatnosti normalne slučajne varijable s parametrima $\mu = 0$, $\sigma = 4$.



Slika 2.5: Graf funkcije gustoće vjerojatnosti eksponencijalne slučajne varijable s parametrom $\lambda = 0.15$.



Slika 2.6: Graf funkcije distribucije vjerojatnosti eksponencijalne slučajne varijable s parametrom $\lambda = 0.15$.