ODABRANE PRIMJENE VJEROJATNOSTI I STATISTIKE

primjeri i zadaci

Ante Mimica, Marina Ninčević
Prirodoslovno-matematički fakultet
Matematički odjel

12. svibnja 2010.

Sadržaj

1

Slučajne varijable i vektori. Uvjetno očekivanje

1.1 Funkcije izvodnice

Neka je X nenegativna cjelobrojna slučajna varijabla zadana zakonom razdiobe

$$X \sim \begin{pmatrix} 0 & 1 & 2 & \dots & n & \dots \\ p_0 & p_1 & p_2 & \dots & p_n & \dots \end{pmatrix}.$$

Funkcija izvodnica slučajne varijable X je funkcija

$$G_X \colon [-1,1] \to \mathbb{R},$$

$$G_X(s) = \sum_{n=0}^{\infty} p_n s^n. \tag{1.1}$$

Primijetimo da je G_X dobro definirana, jer je red (1.1) apsolutno konvergentan;

$$\sum_{n=0}^{\infty} |p_n s^n| \leq \sum_{n=0}^{\infty} p_n |s|^n \leq \sum_{n=0}^{\infty} p_n = 1.$$

Također, možemo pisati

$$G_X(s) = \mathbb{E}[s^X] \tag{1.2}$$

Zadatak 1.1 Odredite funkcije izvodnice za sljedeće slučajne varijable

(a) X Bernoullijeva s parametrom $p \in (0,1)$,

(b) $X \sim B(n,p)$, $n \in \mathbb{N}, p \in (0,1)$,

(c) $X \sim P(\lambda)$, $\lambda > 0$,

(d) $X \sim G(p)$, $p \in (0,1)$.

Napomena. Neka su X i Y nenegativne cjelobrojne slučajne varijable.

(a) $G_X(1) = 1$

(b) [1] Ako su X i Y nezavisne, tada je

$$G_{X+Y}(s) = G_X(s)G_Y(s).$$

(c) [2] Ako je

$$G_X(s) = G_Y(s) \text{ na nekoj okolini } 0,$$

onda je $X \stackrel{d}{=} Y$, t.j.

$$\mathbb{P}(X = n) = \mathbb{P}(Y = n), \ \forall n \in \{0, 1, 2, \ldots\}.$$

(d) [3] Ako X ima konačan n-ti moment, t.j.

$$\mathbb{E}[X^n] < \infty,$$

tada je

$$\mathbb{E}[X(X-1)\cdots(X-n+1)] = G_X^{(n)}(1-) := \lim_{s \to 1-} G_X^{(n)}(s).$$

Specijalno je

$$\boxed{\mathbb{E}X = G'_X(1-)} \quad \text{i} \quad \boxed{\operatorname{Var} X = G''_X(1-) + G'_X(1-) - [G'_X(1-)]^2}.$$

Zadatak 1.2 Neka su $X \sim P(\lambda)$ i $Y \sim P(\mu)$ nezavisne slučajne varijable, gdje su $\lambda, \mu > 0$.

(a) Dokažite da je $\mathbb{E}X = \operatorname{Var} X = \lambda$.

(b) Dokažite da je $X + Y \sim P(\lambda + \mu)$.

Zadatak 1.3 Neka je $X_1, X_2, X_3, \ldots$ niz nezavisnih i jednako distribuiranih nedegeneriranih (t.j. nisu g.s. konstantne) nenegativnih cjelobrojnih slučajnih varijabli i neka je N nenegativna cjelobrojna slučajna varijabla nezavisna od gornjih slučajnih varijabli. Definirajmo

$$S_0 = 0, \ S_n = X_1 + \ldots + X_n, \ n \in \mathbb{N}.$$

(a) Dokažite da je

$$G_{S_N} = G_N \circ G_{X_1}.$$

(b) Ako su $\mathbb{E}N < \infty$ i $\mathbb{E}X_1 < \infty$, izračunajte $\mathbb{E}S_N$.

Zadatak 1.4 Za vrijeme praznika Harry se malo više vozio autocestom na kojoj je bilo puno policijskih patrola. U pola situacija kada ga je zaustavila policija, platio je kaznu od \$ 50, dok je u drugoj polovici situacija platio kaznu od \$ 100. Kad je riječ o policiji, Harry zna biti dosta neugodan. Kada ga zaustavi policajac, Harry bez obzira na visinu kazne uvrijedi policajca s vjerojatnošću $p \in (0,1)$ i tada mu policajac još oduzme i vozačku dozvolu. Koliko novca u prosjeku Harry izgubi prije nego mu policajac oduzme vozačku dozvolu?

Zadatak 1.5 Dokažite da je nemoguće napraviti dvije igraće kocke tako da suma brojeva koji su pali bude uniformno distribuirana na skupu $\{2, 3, \ldots, 12\}$.

1.2 Funkcije izvodnice momenata

Neka je X slučajna varijabla takva da je za neki $\delta > 0$

$$\mathbb{E}[e^{tX}] < \infty, \ |t| < \delta. \tag{1.3}$$

Tada definiramo **funkciju izvodnicu momenata** M_X s

$$M_X(t) = \mathbb{E}[e^{tX}], \ |t| < \delta.$$

Napomena. Vrijede slična svojstva kao i za funkcije izvodnice. Za slučajne varijable X i Y takve da je (1.3) vrijedi

(a) $M_X(0) = 1$

(b) [4] Ako je $\mathbb{E}[|X|^k] < \infty$, onda postoji $M_X^{(k)}(0)$ i vrijedi

$$\mathbb{E}[X^k] = M_X^{(k)}(0).$$

(c) Ako su X i Y nezavisne, onda je

$$M_{X+Y}(t) = M_X(t)M_Y(t).$$

Ako je X diskretna slučajna varijabla, onda je

$$M_X(t) = \sum_x e^{tx}\mathbb{P}(X = x),$$

dok je za neprekidnu slučajnu varijablu X s funkcijom gustoće f

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx.$$

Zadatak 1.6 Odredite funkcije izvodnice momenata za sljedeće slučajne varijable

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

(b) $X \sim P(\lambda)$,

(c) $X \sim U(a, b)$,

(d) $X \sim Exp(\lambda)$,

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

Zadatak 1.7 Neka je $X \sim \Gamma(\alpha, \frac{1}{\lambda})$. Odredite funkciju izvodnicu momenata slučajne varijable X i izračunajte joj očekivanje.

Zadatak 1.8 Neka su $X_1, \ldots, X_n$ nezavisne eksponencijalne slučajne varijable s parametrom $\lambda > 0$. Odredite razdiobu slučajne varijable $Y = X_1 + \ldots + X_n$.

Zadatak 1.9 Neka su $X \sim N(\mu_1, \sigma_1^2)$ i $Y \sim N(\mu_2, \sigma_2^2)$ nezavisne slučajne varijable. Odredite razdiobu slučajne varijable $aX + bY$, gdje su $a, b \in \mathbb{R}$.

Zadatak 1.10 Slučajna varijabla ima funkciju izvodnicu momenata

$$M(t) = e^{3(e^t-1)}.$$

Izračunajte $\mathbb{P}(X = 0)$.

1.3 Neprekidni slučajni vektori

(X, Y) je **dvodimenzionalni neprekidni slučajni vektor** ako postoji funkcija

$$f\colon \mathbb{R}^2 \to \mathbb{R}$$

takva da je

(a) $f(x) \geq 0, \ \forall x \in \mathbb{R}$

(b) $\displaystyle\int\int_{\mathbb{R}^2} f(x, y)\, dx\, dy = 1$

i takva da je

$$F_{X,Y}(a, b) := \mathbb{P}(X \leq a, Y \leq b) = \int_{-\infty}^{a} \int_{-\infty}^{b} f(x, y)\, dx\, dy, \ \forall a, b \in \mathbb{R}.$$

Funkciju f zovemo **funkcija gustoće** slučajnog vektora (X, Y).

Napomena.

(a) Za izmjerive[5] skupove $A \subset \mathbb{R}^d$ vrijedi

$$\mathbb{P}((X,Y) \in A) = \int\int_A f(x,y)\,dx\,dy.$$

(b) $\mathbb{P}(X = a, Y = b) = 0$, $\mathbb{P}(X = a, Y \in B) = 0$, za $a, b \in \mathbb{R}$ i $B \subset \mathbb{R}$ izmjeriv.

(c) Ako je $h \colon \mathbb{R}^d \to \mathbb{R}$ izmjeriva[6] takva da je $\mathbb{E}[|h(X,Y)|] < \infty$, onda je

$$\mathbb{E}[h(X,Y)] = \int\int_{\mathbb{R}^2} h(x,y) f(x,y)\,dx\,dy.$$

Zadatak 1.11 Slučajni vektor (X,Y) ima gustoću

$$f(x,y) = C(y-x)e^{-y},\ \ y > 0, |x| \leq y.$$

(a) Odredite C.

(b) Izračunajte $\mathbb{E}[XY]$.

Zadatak 1.12 Neka je X nenegativna neprekidna slučajma varijabla s gustoćom f i neka je $p > 0$ takav da je $\mathbb{E}[X^p] < \infty$. Dokažite da je

$$\mathbb{E}[X^p] = p\int_0^\infty t^{p-1}\mathbb{P}(X > t)\,dt.$$

Napomena. Za $X \sim Exp(\lambda)$ je $\mathbb{P}(X > t) = e^{-\lambda t}$ pa je

$$\mathbb{E}X = \int_0^\infty \mathbb{P}(X > t)\,dt = \frac{1}{\lambda}.$$

Zadatak 1.13 Neka je $X \sim U(0,1)$. Odredite razdiobu slučajne varijable

$$Y = \frac{1}{\lambda}\ln U.$$

Zadatak 1.14 (Buffonov problem) Stol je podijeljen paralelnim linijam koje su međusobno udaljene za $D > 0$. Igla duljine $L > 0$, pri čemu je $L \leq D$ je bačena na stol na slučajan način. Kolika je vjerojatnost da igla presiječe neku od linija?

Zadatak 1.15 Iz segmenta $[0,1]$ slučajno i nezavisno biramo tri točke x, y, z. Izračunajte vjerojatnost da je

$$x > yz.$$

Napomena. Neka js (X_1, X_2) neprekidni slučajni vektor s gustoćom f_{X_1,X_2} i neka su $g_1, g_2 \colon \mathbb{R}^2 \to \mathbb{R}$ derivabilne funkcije takve da sustav

$$\begin{aligned} y_1 &= g_1(x_1, x_2) \\ y_2 &= g_2(x_1, x_2) \end{aligned}$$

ima jedinstveno rješenje $x_1 = h_1(y_1, y_2)$, $x_2 = h_2(y_1, y_2)$. Tada je

$$(Y_1, Y_2) = (g_1(X_1, X_2), g_2(X_1, X_2))$$

neprekidni slučajni vektor s gustoćom

$$f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}(x_1, x_2)|J(x_1, x_2)|^{-1},$$

gdje su $x_1 = h_1(y_1, y_2)$, $x_2 = h_2(y_1, y_2)$ i gdje je

$$J(x_1, x_2) = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \end{vmatrix}$$

Jacobijan takav da je $J(x, y) \neq 0$.

Zadatak 1.16 Neka su $X_1, X_2 \sim N(0, 1)$ nezavisne slučajne varijable.

(a) Odredite razdiobu slučajnog vektora $(X_1 + X_2, X_1 - X_2)$.

(b) Jesu li $X_1 + X_2$ i $X_1 - X_2$ nezavisne slučajne varijable?

Zadatak 1.17 Neka je (X, Y) neprekidni slučajni vektor s gustoćom f. Dokažite da su gustoće slučajnih varijabli $X + Y$ i $X \cdot Y$ dane s

$$\begin{aligned} f_{X+Y}(y) &= \int_{-\infty}^{\infty} f(y - x, x)\, dx \\ f_{X \cdot Y}(y) &= \int_{-\infty}^{\infty} f(\frac{y}{x}, x)|x|\, dx \end{aligned}$$

Napomena. Ako su X i Y nezavisne, onda je

$$\begin{aligned} f_{X+Y}(y) &= \int_{-\infty}^{\infty} f_X(y - x) f_Y(x)\, dx \\ f_{X \cdot Y}(y) &= \int_{-\infty}^{\infty} f_X(\frac{y}{x}) f_Y(x)|x|\, dx \end{aligned}$$

Zadatak 1.18 Iz segmenta $[0,1]$ slučajno i nezavisno biramo dvije točke x i y. Izračunajte vjerojatnost da je

$$\frac{3}{4} \leq x + y \leq \frac{5}{4}.$$

Zadatak 1.19 (Box-Mullerova transformacija) Neka su $X_1, X_2 \sim U(0,1)$ nezavisne slučajne varijable. Dokažite da su

$$\begin{aligned} Y_1 &= \sqrt{-2\ln X_1}\cos(2\pi X_2) \\ Y_2 &= \sqrt{-2\ln X_1}\sin(2\pi X_2) \end{aligned}$$

nezavisne standardne normalne slučajne varijable.

Zadatak 1.20 (Uređene statistike) Neka su $X_1, \ldots, X_n$ nezavisne i jednako distribuirane slučajne varijable s funkcijom distribucije F. Definiramo **uređene statistike** s

$$\begin{aligned} X_{(1)} &= \text{najmanja vrijednost među } X_1, \ldots, X_n = \min\{X_1, \ldots, X_n\} \\ X_{(2)} &= \text{druga po veličini vrijednost među } X_1, \ldots, X_n \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ X_{(n)} &= \text{najveća vrijednost među } X_1, \ldots, X_n = \max\{X_1, \ldots, X_n\}. \end{aligned}$$

Odredite funkciju distribucije slučajnih varijabli X_k, $k = 1, 2, \ldots, n$.

Zadatak 1.21 Neka su $X_1, \ldots, X_n \sim U(0,1)$ nezavisne slučajne varijable. Odredite funkciju gustoće slučajne varijable $X_{(k)}$, za $k = 1, 2, \ldots, n$.

Zadatak 1.22 Čestica se giba u ravnini koracima duljine 1 tako da u svakom koraku ide u nekom slučajnom smjeru, koji je nezavisan od prethodnih koraka. Odredite očekivanu vrijednost kvadrata udaljenosti čestice od početne točke nakon n koraka.

1.4 Uvjetno očekivanje

Neka je (X, Y) slučajni vektor s funkcijom gustoće f. **Uvjetno očekivanje** od X uz uvjet $Y = y$ je definirano za $y \in \mathbb{R}$ takav da je $f_Y(y) > 0$ sa

$$\mathbb{E}[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\, dx, \tag{1.4}$$

gdje je $f_{X|Y}(x|y)$ **uvjetna funkcija gustoće**

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}.$$

Općenitije, vrijedi

$$\mathbb{E}[g(X)|Y = y] = \int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y)\, dx, \tag{1.5}$$

U slučaju diskretnog slučajnog vektora u (1.4) i (1.5) stoji suma umjesto integrala.

Zadatak 1.23 Slučajni vektor ima gustoću

$$f(x,y) = 4y(x-y)e^{-(x+y)},\ x>0,\ 0<x<y.$$

Izračunajte $\mathbb{E}[X|Y=y]$.

Možemo definirati slučajnu varijablu $\mathbb{E}[X|Y]$ na sljedeći način:

$$\mathbb{E}[X|Y](\omega) = \begin{cases} \mathbb{E}[X|Y=y], & \text{za } y = Y(\omega) \text{ takav da je } f_Y(y) > 0 \\ 0, & \text{inače} \end{cases}$$

Nekada je očekivanje slučajne varijable lakše računati uvjetovanjem. Vrijedi

$$\mathbb{E}X = \mathbb{E}[\mathbb{E}[X|Y]],$$

odnosno

$$\mathbb{E}X = \int_{-\infty}^{\infty} \mathbb{E}[X|Y=y] f_Y(y)\, dy. \tag{1.6}$$

Zaista,

$$\begin{aligned}\int_{-\infty}^{\infty} \mathbb{E}[X|Y=y] f_Y(y)\, dy &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X|Y}(x|y) 1_{\{f_Y(y)>0\}} f_Y(y)\, dx\, dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f(x,y)\, dx\, dy = \int_{-\infty}^{\infty} x f_X(x)\, dx = \mathbb{E}X.\end{aligned}$$

Ako je A događaj, onda vrijedi

$$\mathbb{E}[X|A] = \frac{1}{\mathbb{P}(A)}\mathbb{E}[X \cdot 1_A].$$

Zadatak 1.24 Neka je X jedinična Cauchyeva slučajna varijabla, t.j. ima gustoću

$$f(x) = \frac{1}{\pi(1+x^2)}.$$

Izračunajte

$$\mathbb{E}[X|0<X<1].$$

Zadatak 1.25 Rudar je zarobljen u rudniku s 3 vrata. Prva vrata vode do izlaza nakon 2 sata, dok druga i treća vrata rudara dovode do istog mjesta u rudniku nakon 3, odnosno 5 sati. Ako rudar bira vrata s jednakom vjerojatnošću, odredite očekivano vrijeme potrebno rudaru da izađe iz rudnika.

Zadatak 1.26 (Očekivanje i varijanca geometrijske slučajne varijable) Novčić kod kojeg je vjerojatnost da padne pismo jednaka p bacamo sve dok ne padne pismo. Odredite očekivanje i varijancu potrebnog broja bacanja.

Ako je $X = 1_A$ za neki događaj A, onda je

$$\mathbb{E}X = \mathbb{P}(A) \quad \text{i} \quad \mathbb{E}[X|Y=y] = \mathbb{P}(A|Y=y)$$

pa iz (1.6) dobijemo

$$\mathbb{P}(A) = \int_{-\infty}^{\infty} \mathbb{P}(A|Y=y)\, f_Y(y)\, dy.$$

U diskretnom slučaju ovo nije ništa drugo nego formula potpune vjerojatnosti.

Zadatak 1.27 Neka su X i Y nezavisne neprekidne slučajne varijable. Odredite funkciju distribucije slučajne varijable $X + Y$.

Zadatak 1.28 Neka je $U \sim U(0,1)$ i pretpostavimo da je uvjetna razdioba od X uz uvjet $U = p$ binomna razdioba s parametrima n i p. Odredite razdiobu slučajne varijable X.

Zadatak 1.29 Na zabavu za gradonačelnika, koja počinje u trenutku 0 je pozvano n ljudi. Vremena dolazaka uzvanika su nezavisna i eksponencijalno distribuirana s parametrom 1. Vrijeme dolaska gradonačelnika je uniformno distribuirano na intervalu $(0,1)$ i nezavisno od vremena dolazaka uzvanika.

(a) Izračunajte vjerojatnost da točno k uzvanika dođe prije gradonačelnika.

(b) Koji je očekivani broj uzvanika koji je došao prije gradonačelnika?

Definiramo **uvjetnu varijancu** uz uvjet $Y = y$ sa

$$\mathrm{Var}(X|Y=y) = \mathbb{E}[(X - \mathbb{E}[X|Y=y])^2|Y=y]$$

i stavljamo

$$\mathrm{Var}(X|Y)(\omega) = \begin{cases} \mathrm{Var}(X|Y=y), & \text{za } y = Y(\omega) \text{ takav da je } f_Y(y) > 0 \\ 0, & \text{inače} \end{cases}$$

Lako se vidi da je

$$\mathrm{Var}(X|Y=y) = \mathbb{E}[X^2|Y=y] - (\mathbb{E}[X|Y=y])^2.$$

Zadatak 1.30 Dokažite da je

$$\mathrm{Var}\, X = \mathbb{E}[\mathrm{Var}(X|Y)] + \mathrm{Var}(\mathbb{E}[X|Y]).$$

Zadatak 1.31 Neka je $X_1, X_2, \ldots$ niz nezavisnih i jednako distribuiranih slučajnih varijabli s očekivanjem μ i varijancom σ^2 te neka je N nenegativna cjelobrojna slučajna varijabla nezavisna od $X_1, X_2, \ldots$. Definiramo slučajnu varijablu

$$S = \sum_{i=1}^{N} X_i.$$

Odredite $\mathbb{E}S$ i $\mathrm{Var}\, S$.

Zadatak 1.32 Dnevni broj kupaca u nekoj trgovini ima Poissonovu razdiobu s parametrom 10. Svaki kupac nezavisno od drugih kupaca potroši svotu novca koja je uniformno distribuirana na intervalu $(0, 100)$. Odredite očekivanje i varijancu dnevne zarade te trgovine.

Zadatak 1.33 Neka je $U \sim U(0, 1)$. Ponavlja se niz od n pokusa koji su uz uvjet da je $U = u$ nezavisni s Bernoullijevim zakonom s vjerojatnošću uspjeha u. Izračunajte očekivanje i varijancu broja uspjeha u tim pokusima.

1.5 Još neke razdiobe. Intenzitet hazarda.

Znamo da $X \sim Exp(\lambda)$ ima svojstvo **memorijske odsutnosti**:

$$\mathbb{P}(X > t + y | X > s) = \mathbb{P}(X > t)$$

i da je eksponencijalna razdioba jedina neprekidna razdioba s tim svojstvom.

Zadatak 1.34 Pretpostavite da je broj kilometara koje automobil prijeđe prije nego što moramo zamijeniti akumulator eksponencijalno distribuiran s očekivanjem od 10 000 km. Ako želimo prijeći put od 5 000 km, kolika je vjerojatnost da će akumulator izdržati put? Što možete reći o toj vjerojatnosti, ako razdioba nije eksponencijalna?

Zadatak 1.35 Žarulje A i B imaju vijek trajanja koji je modeliran po eksponencijalnoj razdiobi s parametrima λ_A i λ_B. Ako ste ušli u neku prostoriju u kojoj su uključene žarulje A i B, kolika je vjerojatnost da žarulja A pregori prije žarulje B?

Slučajna varijabla ima **Weibullovu** razdiobu s parametrima $c, \alpha > 0$ ako joj je gustoća

$$f(x) = c\alpha x^{\alpha-1} \exp\{-cx^{\alpha}\}, \ x > 0.$$

Napomena. Ako je $\alpha = 1$, onda je $X \sim Exp(c)$.

Zadatak 1.36 Odredite očekivanje Weibullove slučajne varijable s parametrima $c, \alpha > 0$.

Zadatak 1.37 Vrijeme (u satima) do kvara ležaja u nekom motoru je modelirano Weibullovom razdiobom s parametrima $\alpha = 1/2$ i $c = 1/(50\sqrt{2})$.

(a) Izračunajte očekivano vrijeme do kvara.

(b) Kolika je vjerojatnost da će ležaj trajati barem 6 000 sati?

Zadatak 1.38 Neka je $W \sim N(\mu, \sigma^2)$. Odredite razdiobu slučajne varijable $X = e^W$.

Kažemo da slučajna varijabla X ima **lognormalnu** razdiobu s parametrima $\mu \in \mathbb{R}$ i $\sigma^2 > 0$ ako je $X = e^W$, gdje je $W \sim N(\mu, \sigma^2)$.

Zadatak 1.39 Životni vijek poluvodičkog lasera ima lognormalnu razdiobu s parametrima $\mu = 10$ sati i $\sigma = 1.5$ sati.

(a) Kolika je vjerojatnost da laser doživi 10 000 sati?

(b) Koji životni vijek će doživjeti 99 % lasera?

(c) Izračunajte očekivani životni vijek lasera.

Slučajna varijabla X ima **Pareto** razdiobu s parametrima $\alpha, \kappa > 0$, ako joj je gustoća dana s

$$f(x) = \frac{\alpha\kappa^\alpha}{x^{\alpha+1}},\ x \geq \kappa.$$

Zadatak 1.40 Odredite očekivanje slučajne varijable s Pareto razdiobom s parametrima $\alpha\kappa > 0$. Kada postoji?

Neka je X neprekidna pozitivna slučajna variabla s funkcijom distribucije F. **Rep funkcije distribucije** je definiran s

$$\overline{F}(t) = \mathbb{P}(X > t) = 1 - F(t).$$

Intenzitet hazarda je definiran s

$$\lambda(t) = \frac{f(t)}{\overline{F}(t)} = -\frac{d}{dt}(\ln \overline{F}(t)).$$

Napomena.

$$\overline{F}(t) = e^{-\int_0^t \lambda(s)\, ds}$$

Zadatak 1.41 Poznato je da je stopa smrtnosti pušača u svakom trenutku dvostruko veća od stope smrtnosti nepušača. Kako se odnose odgovarajuće vjerojatnosti?

Zadatak 1.42 Odredite za koje parametre intenzitet hazarda Weibullove distribucije raste, a za koje pada.

Zadaci za vježbu

Funkcije izvodnice

1.43 Odredite funkciju izvodnicu, očekivanje i varijancu uniformne slučajne varijable na skupu $\{1, 2, \ldots, n\}$, $n \in \mathbb{N}$.

1.44 Odredite očekivanje i varijancu

(a) Bernoullijeve slučajne varijable s parametrom $p \in (0, 1)$,

(b) geometrijske slučajne varijable s parametrom $p \in (0, 1)$,

(c) binomne slučajne varijable s parametrima $n \in \mathbb{N}$ i $p \in (0, 1)$.

1.45 Neka je G_X funkcija izvodnica nenegativne cjelobrojne slučajne varijable X. Odredite funkciju izvodnicu slučajne varijable $aX + b$, gdje su $a, b \in \mathbb{N}$.

1.46 Pretpostavite da u Zadatku 1.3 vrijedi $E[N^2] < \infty$ i $\mathbb{E}[X_1^2] < \infty$. Dokažite da je

$$\operatorname{Var} S_N = \mathbb{E}N \operatorname{Var} X_1 + (\mathbb{E}X_1)^2 \operatorname{Var} N.$$

1.47 Baca se simetrični novčić. Svaki put kada padne pismo se baca simetrična kocka. Kada padne glava stanemo s igrom. Nađite funkciju izvodnicu sume svih brojeva koji su pali na kocki.

1.48 Bacamo 3 simetrične kocke. Izračunajte da suma brojeva koji su pali bude 9. (Uputa: Nađite funkciju izvodnicu sume brojeva na kockama i promatrajte koeficijent uz s^9.)

1.49 Ploča za pikado je podijeljena na 10 jednakih dijelova označenih brojevima 1, 2, 3,. . ., 10. Robin je početnik u pikadu i s jednakom vjerojatnošću pogađa bilo koji dio na ploči (i ne može promašiti ploču). Izračunajte vjerojatnost da Robin u 3 nezavisna gađanja osvoji točno 24 boda.

1.50 Marko često trči uz rijeku uz koju ima puno komaraca. Svaki komarac koji se približi Marku ga ubode s vjerojatnošću $p \in (0, 1)$. Ako broj komaraca koji se približi Marku za vrijeme jednog treninga ima Poissonovu distribuciju s parametrom $\lambda > 0$, odredite distribuciju broja uboda i prosječni broj uboda.

1.51 Neka je X slučajna varijabla s razdiobom

$$\mathbb{P}(X = k) = \frac{-1}{\ln(1-p)} \frac{p^k}{k}, \ k \in \mathbb{N},$$

gdje je $0 < p < 1$. Odredite funkciju izvodnicu slučajne varijable X i izračunajte $\mathbb{E}X$.

1.52 Neka je $X_1, \ldots, X_n$ niz nezavisnih i jednako distribuiranih slučajnih varijabli takvih da je

$$X_1 + X_2 + \ldots + X_n \sim P(\lambda),$$

gdje je $n \in \mathbb{N}$. Odredite razdiobu slučajne varijable X_1.

1.53 Neka su X i Y nenegativne cjelobrojne slučajne varijable. Za $s, t \in [-1, 1]$ definiramo **zajedničku funkciju izvodnicu** od X i Y sa

$$G_{X,Y}(s, t) = \sum_{m,n=0}^{\infty} s^m t^n \mathbb{P}(X = m, Y = n).$$

Dokažite da su X i Y nezavisne ako i samo ako je

$$G_{X,Y}(s, t) = G_X(s) \cdot G_Y(t), \ \forall s, t \in [-1, 1].$$

1.54 Neka je G_X funkcija izvodnica nenegativne cjelobrojne slučajne varijable X. Pokažite da ze sve $\alpha > 0$ vrijedi

$$\mathbb{E}\left[\frac{1}{X + \alpha}\right] = \int_0^1 s^{\alpha-1} G_X(s) \, \mathrm{d}s.$$

Funkcije izvodnice momenata

1.55 Neka su $X_1, \dots, X_n \sim N(0,1)$ nezavisne slučajne varijable. Odredite funkciju izvodnicu momenata slučajne varijable $X = X_1^2 + \dots + X_n^2$. Koju razdiobu ima X?

1.56 Neka su M_1 i M_2 funkcije izvodnice momenata dvije diskretne slučajne varijable i neka je $0 \le \alpha \le 1$. Koje su od sljedećih funkcija funkcije izvodnice momenata:

$$\text{(a) } M_1(t) \cdot M_2(t) \quad \text{(b) } \alpha M_1(t) + (1-\alpha) M_2(t) \quad \text{(c) } \frac{M_1(\alpha t)}{M_1(\alpha)} \text{ ?}$$

Neprekidni slučajni vektori

1.57 Na dužini duljine L su slučajno i nezavisno odabrane 3 točke x, y, z. Kolika je vjerojatnost da se y nalazi između x i z.

1.58 (a) Slučajni vektor (X, Y) ima gustoću

$$f(x,y) = xe^{-x-y}, \; x > 0, y > 0.$$

Jesu li X i Y nezavisne?

(b) Slučajni vektor (X, Y) ima gustoću

$$f(x,y) = 2, \; 0 < x < y, 0 < y < 1.$$

Jesu li X i Y nezavisne?

1.59 Neka su $X, Y \sim U(0,1)$ nezavisne. Dokažite da za $\alpha > 0$ vrijedi

$$\mathbb{E}[|X-Y|^\alpha] = \frac{2}{(\alpha+1)(\alpha+2)}.$$

1.60 Neka je $X \sim N(0,1)$ i neka je $a > 0$. Definiramo Y s

$$Y = \begin{cases} Z, & X > a \\ 0, & X \le a. \end{cases}$$

Izračunajte $\mathbb{E}Y$.

1.61 Mjesto nesreće na cesti duljine $L > 0$ je uniformno distribuirano. U vrijeme nesreće, položaj vozila hitne pomoći je također uniformno distribuiran na istoj cesti. Ako su mjesto nesreće i položaj vozila hitne pomoći nezavisni, odredite očekivanu udaljenost između mjesta nesreće i vozila hitne pomoći.

1.62 Slučajna varijabla ima jediničnu **Caucyhevu distribuciju** ako ima gustoću

$$f(x) = \frac{1}{\pi(1+x^2)}.$$

Postoji li $\mathbb{E}X$?

1.63 Neka su $X, Y \sim N(0,1)$ nezavisne slučajne varijable. Odredite razdiobu slučajnog vektora $(X, X/Y)$. Koju razdiobu ima X/Y?

1.64 Neka su $X, Y \sim N(0,1)$ nezavisne slučajne varijable. Odredite razdiobu slučajnog vektora (R, Θ), gdje je

$$R = \sqrt{X^2 + Y^2}, \quad \Theta = \operatorname{arctg} \frac{Y}{X}.$$

Jesu li R i Θ nezavisne? Odredite im razdiobu.

1.65 Jesu li R^2 i Θ iz Zadatka 1.64 nezavisne? Odredite gustoću od R^2.

1.66 Neka su $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ uređene statistike n nezavisnih uniformnih slučajnih varijabli na $[0,1]$. Dokažite da za $1 \leq k \leq n+1$ vrijedi

$$\mathbb{P}(X_{(k)} - X_{(k-1)} > t) = (1-t)^n,$$

pri čemu stavljamo $X_{(0)} := 0$ i $X_{(n+1)} := t$.

1.67 Neka su $X \sim \Gamma(\alpha, 1/\lambda)$ i $X \sim \Gamma(\beta, 1/\lambda)$ nezavisne slučajne varijable. Odredite razdiobu slučajnog vektora

$$(X + Y, X/(X+Y)).$$

Odredite marginalne razdiobe ovog slučajnog vektora. Jesu li komponente ovog slučajnog vektora nezavisne?

1.68 Neka su $X_1, \ldots, X_n$ nezavisne i jednako distribuirane neprekidne slučajne varijable s gustoćom f. Dokažite da slučajni vektor uređenih statistika

$$Z = (X_{(1)}, \ldots, X_{(n)})$$

ima gustoću

$$f_Z(x_1, x_2, \ldots, x_n) = n! f(x_1) f(x_2) \cdots f(x_n), \ x_1 < x_2 < \ldots < x_n.$$

Uvjetno očekivanje

1.69 Slučajni vektor ima gustoću

$$f(x,y) = \frac{1}{2} y e^{-xy}, \ x > 0, \ 0 < y < 2.$$

Izračunajte

$$\mathbb{E}[e^{x/2} | Y = 1].$$

1.70 Slučajni vektor ima gustoću

$$f(x,y) = \frac{e^{-x/y} e^{-y}}{y}, \ x, y > 0.$$

Izračunajte

$$\mathbb{E}[X | Y = y].$$

1.71 Slučajni vektor (X, Y) ima gustoću

$$f(x, y) = C(x^2 - y^2)e^{-y},\ y > 0,\ -y \leq x \leq y.$$

Odredite C i izračunajte $\mathbb{E}[X|Y = y]$.

1.72 Neka su $X_1, X_2 \sim G(p)$ nezavisne slučajne varijable. Izračunajte

$$\mathbb{P}(X_1 = i|X_1 + X_2 = n).$$

1.73 Neka je X Poissonova slučajna varijabla s parametrom koji je eksponencijalno distribuiran s parametrom 1. Odredite distribuciju slučajne varijable X.

1.74 Na zabavu je pozvan Marko i još 10 njegovih prijatelja. Vremena dolazaka su nezavisna i uniformno distribuirana na intervalu $(0, 1)$. Odredite očekivani broj i varijancu ljudi koji dođu prije Marka.

1.75 Zatvorenik je zatočen u ćeliji koja ima 3 vrata. Prva i druga vrata vode u zatvorenika tunele, koji ga vrate u ćeliju nakon 2 dana, odnosno 3 dana, dok ga treća vrata odmah dovedu do slobode. Ako zatvorenik bira vrata 1, 2 ili 3 s vjerojatnostima 0.5, 0.3, 0.2, izračunajte očekivano vrijeme koje zatvorenik provede u zatvoru do slobode. Kolika je varijanca?

1.76 Neka je $X \sim Exp(\lambda)$. Izračunajte

$$\mathbb{E}[X|X > 1].$$

1.77 Novčić kod kojeg je vjerojatnost da padne pismo jednaka p se baca dok ne padnu tri pisma zaredom. Odredite očekivani broj bacanja novčića.

1.78 Osiguranje pretpostavlja da je broj prometnih nezgoda koje će klijent imati u jednoj godini distribuiran po Poissonovoj razdiobi. Ako je parametar te razdiobe slučajan i ima eksponencijalnu razdiobu s parametrom $\lambda > 0$, izračunajte vjerojatnost da slučajno odabrani klijent ima točno n prometnih nezgoda u godini.

Još neke razdiobe. Intenzitet hazarda

1.79 Odredite varijancu Weibullove slučajne varijable s parametrima $c, \alpha > 0$.

1.80 Sluăjna varijabla X ima Weibullovu razdiobu s parametrima $c = 1$ i $\alpha = 2$.

(a) Izračunajte $\mathbb{P}(X > 2)$ i $\mathbb{P}(1 < X < 3)$.

(b) Odredite intenzitet hazarda.

1.81 Životni vijek uredaja za magnetsku rezonancu (MR) ima Weibullovu razdiobu s parametrima $\alpha = 2$ i $c = 1/250000$ sati.

(a) Izračunajte očekivano vrijeme života uredaja.

(b) Izračunajte varijancu životnog vijeka uredaja.

(c) Kolika je vjerojatnost da uredaj ne doživi 250 sati?

1.82 Slučajna varijabla X ima lognormalnu razdiobu s parametrima $\mu = 5$ i $\sigma^2 = 9$. Izračunajte:

(a) $\mathbb{P}(X < 13300)$,

(b) vrijednost x za koju je $\mathbb{P}(X \leq x) = 0.95$,

(c) očekivanje i varijancu od X.

1.83 Odredite očekivanje i varijancu lognormalne slučajne varijable s parametrima μ i σ^2.

1.84 Što možete reći o monotonosti intenziteta hazarda Pareto razdiobe?

2

Monte Carlo simulacije

2.1 Uvod R

R je programski sustav za statističko računanje, obradu podataka i grafiku. Ovaj programski sustav je besplatan i možete ga pronaći na `http://www.r-project.org`.

Opis	R	Primjer
Komentar	`#`	`#ovo je primjer komentara`
Operator pridruživanja	`<-`	`x<-1/sqrt(2*pi)*exp(-2)`
Operator konkatenacije	`c`	`c(1,2,2,3)`
Množenje	`*`	`a*b`
Exponent	`^`	`2^3`
Niz od a do b s korakom h	`seq`	`seq(a,b,h)`
Nizovni operator	`:`	`1:10`
x mod y	`x %% y`	`25%%3`

2.1.1 Vektori

Jedan od osnovnih objekata je vektor. Npr.

```
> v<-1:20
> v
 [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
> v^2
 [1]   1   4   9  16  25  36  49  64  81 100 121 144 169 196 225 256 289 324 361
[20] 400
```

Podacima u vektoru možemo pristupiti na razne načine:

```
> v[4]
[1] 4
> v[v%%3==1]
[1]  1  4  7 10 13 16 19
```

Izraz v%%3==1 je primjer logičkog izraza. Npr.

```
> 2^3+3^3==4^3
[1] FALSE
> v%%3==1
 [1]  TRUE FALSE FALSE  TRUE FALSE FALSE  TRUE FALSE FALSE  TRUE FALSE FALSE
[13]  TRUE FALSE FALSE  TRUE FALSE FALSE  TRUE FALSE
```

Elemente u vektore možemo unositi jedan po jedan:

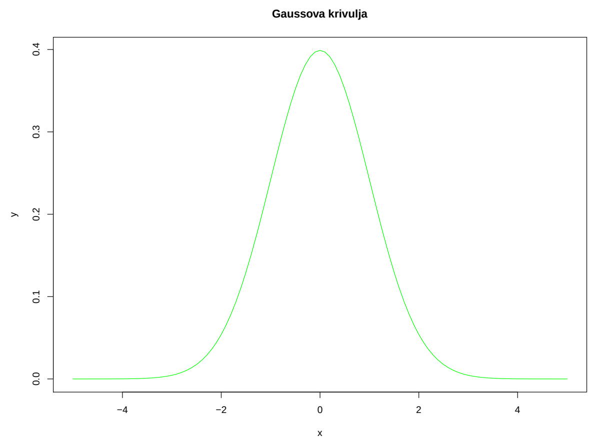
```
> b<scan()
1:
> b<-scan()
1: 101
2: 21
3: 3
4: 45
5:
Read 4 items
> b
[1] 101  21   3  45
```

2.1.2 Funkcije

Funkcija	R	Primjer
Eksponencijalna	exp	exp(-2)
Logaritam	log	log(2)
$\lfloor x \rfloor$	floor	floor(-pi)
$\lceil x \rceil$	ceiling	ceiling(-pi)
$n!$	factorial	factorial(5)
Suma	sum	sum(c(1,2,3))
Kumulativna suma	cumsum	cumsum(c(1,2,3))
Aritmetička sredina	mean	mean(c(1,4,6,3)
Uzoračka varijanca	var	var(c(1,4,6,3))
Standardna devijacija	sd	sd(c(1,4,6,3))
Uzorak	sample	sample(c('pismo','glava'),size=10,replace=T)

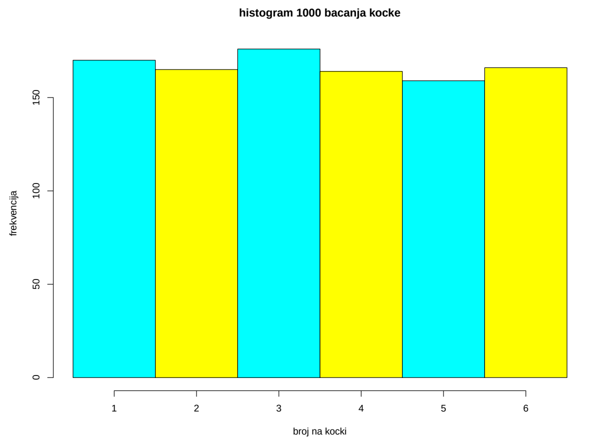
Primjer. (Gaussova krivulja)

```
> gaussian<-function(x,mu,var)
+ exp(-(x-mu)^2/(2*var))/(sqrt(2*pi*var))
> gaussian(0,0,1)
[1] 0.3989423
> f<-function(x)
+ gaussian(x,0,1)
> plot(f,-5,5)
> plot(f,-5,5,xlab='x',ylab='y',
main='Gaussova krivulja',col="green")
```



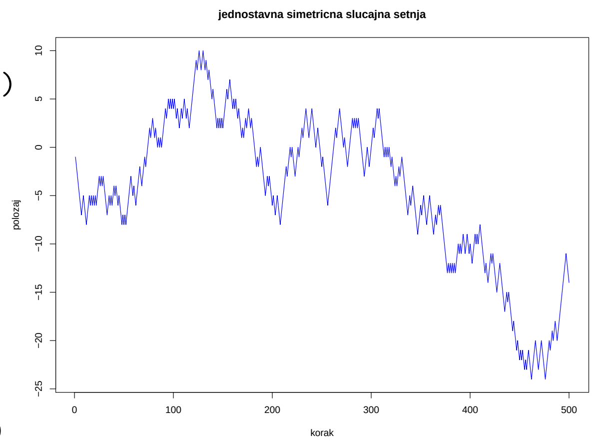
Primjer. (Bacanje kocke)

```
> kocka<-function(n) {
+ k<-sample(1:6,size=n,replace=TRUE)
+ return(k)
+ }
> kocka(5)
[1] 4 2 1 3 6
> hist(kocka(10000))
> hist(kocka(1000),breaks<-seq(0.5,6.5,1),
xlab="broj na kocki", ylab="frekvencija",
main="histogram 1000 bacanja kocke",
col=rep(c("cyan","yellow"),3))
```



Primjer. (Simetrična slučajna šetnja)

```
> srw<-function(n) {
+ steps<-sample(c(-1,1),size=n,replace=TRUE)
+ return(cumsum(steps))
+ }
> srw(10)
 [1]  1  2  1  0 -1  0 -1  0 -1  0
> plot(rw(500),type="l",col="blue",
xlab="korak",ylab="polozaj",
main="jednostavna simetricna
slucajna setnja")
# vremena kada se slucajna setnja vraca u 0
> which(s==0)
 [1]   2   4   6 116 536 540 858 860 862 888 890
```



2.1.3 Matrice

Matrice predstavljamo kao vektore s dimenzijama.

```
> x<-1:12
> dim(x)<-c(3,4)
> x
     [,1] [,2] [,3] [,4]
[1,]    1    4    7   10
[2,]    2    5    8   11
[3,]    3    6    9   12
> x[1,2]
[1] 4
> x[,2]
[1] 4 5 6
> x[1]
[1] 1
> x[1,]
[1]  1  4  7 10
```

Drugi način zadavanja je pomoću funkcije `matrix`.

```
> y<-matrix(1:9,nrow=3,ncol=3)
> y
     [,1] [,2] [,3]
[1,]    1    4    7
[2,]    2    5    8
[3,]    3    6    9
> det(y)
[1] 0
```

Pomoću naredbe `byrow=T` smještamo elemente matrice po retcima.

Pomoću naredbi `rownames` i `colnames` mijenjamo imena redaka i stupaca.

```
> x<-matrix(1:28,ncol=7,byrow=T)
> x
     [,1] [,2] [,3] [,4] [,5] [,6] [,7]
[1,]    1    2    3    4    5    6    7
[2,]    8    9   10   11   12   13   14
[3,]   15   16   17   18   19   20   21
[4,]   22   23   24   25   26   27   28
> colnames(x)<-c('pon','uto','sri','cet','pet','sub','ned')
> x
     pon uto sri cet pet sub ned
```

```
[1,]    1    2    3    4    5    6    7
[2,]    8    9   10   11   12   13   14
[3,]   15   16   17   18   19   20   21
[4,]   22   23   24   25   26   27   28
```

Koristeći naredbu `t` transponiramo matricu

```
> t(x)
    [,1] [,2] [,3] [,4]
pon    1    8   15   22
uto    2    9   16   23
sri    3   10   17   24
cet    4   11   18   25
pet    5   12   19   26
sub    6   13   20   27
ned    7   14   21   28
```

Vektore spajamo u matrice koristeći naredbe `cbind` i `rbind`.

```
> cbind(A=1:5,B=10:14,C=20:24)
     A  B  C
[1,] 1 10 20
[2,] 2 11 21
[3,] 3 12 22
[4,] 4 13 23
[5,] 5 14 24
```

2.1.4 Polja

```
> x<-array(1:12,c(2,2,3))  # primjer 2 x 2 x 3 polja
> x
, , 1

     [,1] [,2]
[1,]    1    3
[2,]    2    4

, , 2

     [,1] [,2]
[1,]    5    7
[2,]    6    8
```

```
, , 3

     [,1] [,2]
[1,]    9   11
[2,]   10   12
```

2.1.5 Faktori

Nekada je korisno kategorijalnim podacima pridruziti neke nivoe ili kategorije.

```
> ocjene<-c(2,1,4,5,4,3,3,2,3,4,2,4)
> focjene<-factor(ocjene,levels=1:5)
> levels(focjene)<-c('nedovoljan','dovoljan','dobar','vrlo dobar','izvrstan')
> focjene
 [1] dovoljan   nedovoljan vrlo dobar izvrstan   vrlo dobar dobar
 [7] dobar      dovoljan   dobar      vrlo dobar dovoljan   vrlo dobar
Levels: nedovoljan dovoljan dobar vrlo dobar izvrstan
> as.numeric(focjene)
 [1] 2 1 4 5 4 3 3 2 3 4 2 4
```

2.1.6 Data frames

U statistici su podaci cesto zadani preko tablica koje sadrže vektore i faktore.

```
> razred<-data.frame(ocjene,focjene)
> razred
   ocjene    focjene
1       2   dovoljan
2       1 nedovoljan
3       4 vrlo dobar
4       5   izvrstan
5       4 vrlo dobar
6       3      dobar
7       3      dobar
8       2   dovoljan
9       3      dobar
10      4 vrlo dobar
11      2   dovoljan
12      4 vrlo dobar
> razred$ocjene[ocjene!=1]
 [1] 2 4 5 4 3 3 2 3 4 2 4
```

2.1.7 Čitanje iz vanjske datoteke

U R možemo učitavati podatke iz vanjskih datoteka. Prvo pomoću naredbi `getwd` i `setwd` dođjemo do direktorija gdje se nalazi datoteka. Npr.

```
> getwd()
[1] "/home/user"
> setwd('R')
> getwd()
[1] "/home/user/R"
> m<-read.table('meteo.data',header=T)
> m
     mjesec srednja_temperatura suncani_dani oborine.mm.
1  sijecanj                -0.1            2        47.0
2   veljaca                 2.0            3        40.6
3    ozujak                 6.2            4        52.2
4   travanj                10.9            3        63.1
5   svibanj                15.7            4        76.4
6    lipanj                19.1            3        96.7
7    srpanj                20.8            7        81.6
8   kolovoz                20.0            8        89.3
9     rujan                16.0            6        85.9
10 listopad                10.8            4        74.2
11  studeni                 5.7            2        81.8
12 prosinac                 1.3            2        64.2
> m$mjesec[m$oborine>80]
[1] lipanj  srpanj  kolovoz rujan   studeni
12 Levels: kolovoz lipanj listopad ozujak prosinac rujan sijecanj ... veljaca
```

Kôd iz datoteke učitavamo naredbom `source`. Npr.

```
> source('srw.R')
> srw(20)
 [1]  1  0 -1 -2 -3 -2 -1  0 -1 -2 -1  0  1  2  3  2  3  2  1  0
```

2.1.8 Naredbe grananja i petlje

Naredba grananja je dana s `if`:

```
if (uvjet) {izraz1}
else {izraz2}
```

gdje je `uvjet` neki logički uvjet. Npr.

```
> bodovi<-80
> if (bodovi>=50) print('ispit polozen') else print('ispit nije polozen')
[1] "ispit polozen"
```

U R-u imamo petlje `for`,`repeat`,`while`.

Primjer. (Newtonova metoda za računanje drugog korijena) Znamo da drugi korijen od $y > 0$ mo zemo računati na način da definiramo rekurzivno niz

$$a_0 = \frac{y}{2}, a_n = \frac{1}{2}\left(a_{n-1} + \frac{y}{a_{n-1}}\right), \ n \geq 1.$$

Tada se lako vidi da je (a_n) konvergentan i $\lim_n a_n = \sqrt{y}$.

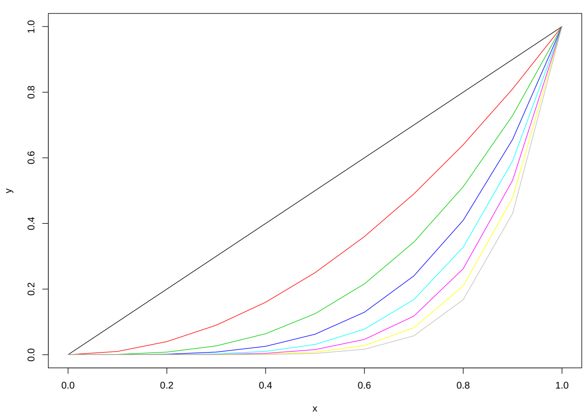
```
> korijen<-function(y,tocnost) {
+ x<-y/2
+ while(abs(x*x-y)>tocnost) x<-(x+y/x)/2
+ return(x)
+ }
> z<-korijen(10323,1e-10)
> z^2
[1] 10323
```

Funkciju `korijen` možemo napisati i pomoću repeat:

```
> korijen<-function(y,tocnost) {
+ x<-y/2
+ repeat {
+ x<-(x+y/x)/2
+ if(abs(x*x-y)<tocnost) break
+ }
+ return(x)
+ }
```

Primjer. (`for` **petlja)**

```
> x<-seq(0,1,0.1)
> plot(x,x,ylab="y",type="l")
> for(j in 2:8) lines(x,x^j,col=j)
```



2.1.9 Spremanje slika

Slike možemo spremiti u `.pdf` ili `.ps` datoteke na sljedeći način:

```
> pdf('slika.pdf')
> plot(x^2)
> dev.off()
X11cairo
       2
```

ili

```
> postscript('slika.ps')
> plot(x^2)
> dev.off()
X11cairo
       2
```

2.2 Simuliranje slučajnih varijabli

2.2.1 Odabir uzorka iz populacije

Pomoću funkcije `sample` biramo uzorak iz neke populacije.

Zadatak 2.1 (Izvlačenje kuglica) U kutiji je 5 bijelih, 3 crvene i 7 zelenih kuglica. Iz kutije izvlačimo 8 kuglica, jednu po jednu. Simulirajte izvlačenje ako

(a) izvučene kuglice vraćamo u kutiju.

(b) izvučene kuglice ne vraćamo u kutiju.

Rješenje.

```
> kutija<-c(rep('b',5),rep('c',3),rep('z',7))
> sample(kutija,size=8,replace=T)
[1] "b" "b" "z" "z" "z" "b" "z" "c"
> sample(kutija,size=8,replace=F)
[1] "z" "b" "b" "b" "c" "z" "z" "b"
```

Zadatak 2.2 (Slučajna permutacija) Simulirajte slučajnu permutaciju skupa slova.

Rješenje.

```
> letters
 [1] "a" "b" "c" "d" "e" "f" "g" "h" "i" "j" "k" "l" "m" "n" "o" "p" "q" "r" "s"
[20] "t" "u" "v" "w" "x" "y" "z"
> sample(letters)
 [1] "l" "o" "h" "n" "b" "k" "e" "f" "d" "r" "v" "s" "y" "p" "c" "q" "g" "z" "x"
[20] "i" "a" "u" "t" "w" "j" "m"
```

Zadatak 2.3 Simulirajte slučajnu šetnju duljine

$$S_0 = 0,\ S_n = X_1 + \ldots + X_n, n \geq 1,$$

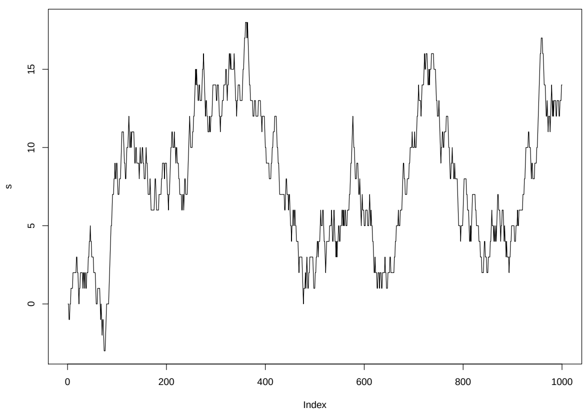
gdje su

$$X_i \sim \begin{pmatrix} -1 & 0 & 1 \\ 0.25 & 0.5 & 0.25 \end{pmatrix}$$

nezavisne slučajne varijable.

Rješenje.

```
> koraci<-sample(c(-1,0,1),
100,replace=T,
prob=c(0.25,0.5,0.25))
> s<-cumsum(koraci)
> plot(s,type="l")
```



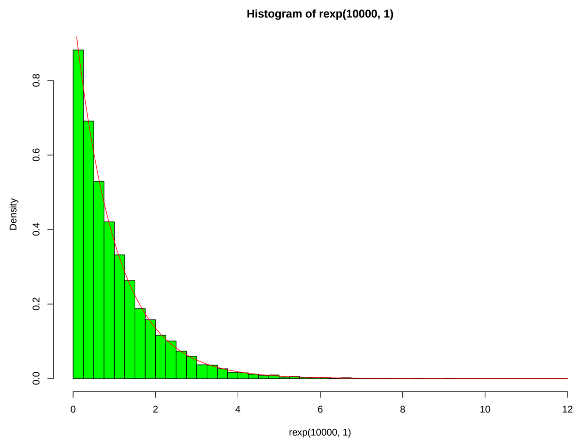
2.2.2 Neke razdiobe implementirane u R-u

Razdioba	Generator	Parametri
uniformna	`runif`	a, b
eksponencijalna	`rexp`	λ
normalna	`rnorm`	μ, σ^2
gama	`rgamma`	$\alpha, 1/\beta$
beta	`rbeta`	α, β
Studentova t	`rt`	n
Fisherova F	`rf`	n_1, n_2
χ^2	`rchisq`	n
Poissonova	`rpois`	λ
geometrijska	`rgeom`	p

Npr. za eksponencijalnu razdiobu imamo

`dexp`	funkcija gustoće	`dexp(x,lambda)`
`pexp`	funkcija distribucije	`pexp(x,lambda)`
`qexp`	$(1 - alpha)$-kvantil	`qexp(alpha,lambda)`
`rexp`	realizacija	`dexp(n,lambda)`

```
> hist(rexp(10000,1),prob=T,
breaks<-c(seq(0,10,0.25),12),
col="green")
> x<-seq(0,12,0.05)
> lines(x,dexp(x,1),col="red")
```



2.2.3 Metoda inverzne transformacije

Neka je X slučajna varijabla s funkcijom distribucije F. Definiramo generalizirani inverz od F sa

$$F^{\leftarrow}(u) = \inf\{x \colon F(x) \geq u\}.$$

U slučaju da je F bijekcija, vrijedi $F^{\leftarrow} = F^{-1}$. S predavanja znamo da za $U \sim U(0,1)$ slučajna varijabla

$$F^{\leftarrow}(U)$$

ima istu razdiobu kao i slučajna varijabla X.

Primjer. (Eksponencijalna razdioba) Simulirajte eksponencijalnu slučajnu varijablu koristeći metodu inverza.

Ako je $X \sim$Exp(λ), onda je

$$F(x) = 1 - e^{-\lambda x}, x > 0.$$

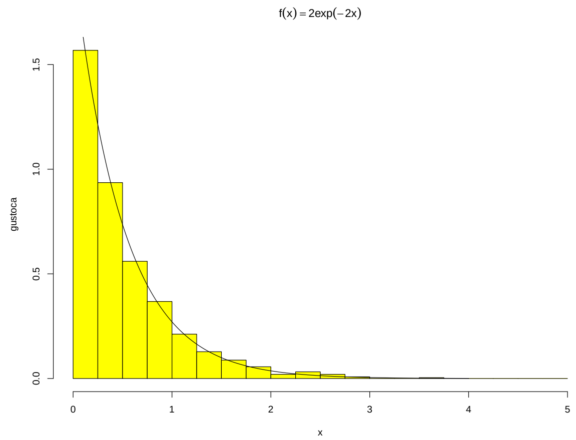
Tada je

$$F^{-1}(u) = -\frac{\log(1-u)}{\lambda}.$$

Za $U \sim U(0,1)$ je tada

$$X \stackrel{d}{=} -\frac{\log(1-U)}{\lambda} \stackrel{d}{=} -\frac{\log U}{\lambda}$$

```
> n<-1000
> lambda<-2
> u<-runif(n)
> r<--log(u)/lambda
> hist(r,prob=T,
main=expression(f(x)==2*exp(-2*x)),
xlab="x",ylab="gustoca",col="yellow",
breaks=seq(0,5,0.25))
> y<-seq(0,4,0.01)
> lines(y,dexp(y,lambda))
```



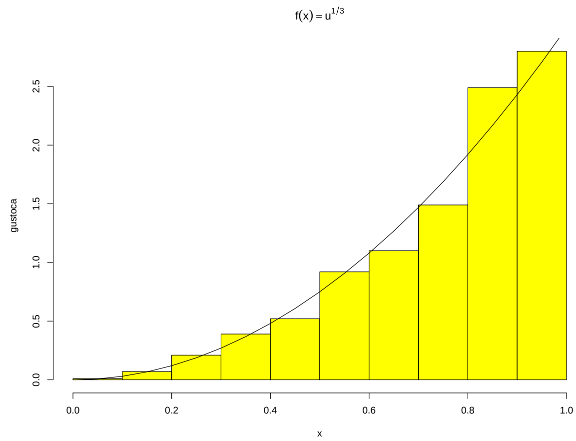
Zadatak 2.4 Simulirajte metodom inverzne transofrmacije slučajnu varijablu s gustoćom

$$f(x) = 3x^2, 0 < x < 1.$$

Rješenje.

$F(x) = x^3,\ 0 < x < 1$ pa je $F^{-1}(x) = x^{1/3}$

```
> n<-1000
> u<-runif(n)
> r<-u^(1/3)
> hist(r,prob=T,
main=expression(f(x)==u^{1/3}),
xlab="x",ylab="gustoca",col="yellow",
breaks=seq(0,1,0.1))
> y<-seq(0,1,0.05)
> lines(y,3*y^3)
```



Zadatak 2.5 (Geometrijska razdioba) Simulirajte geometrijsku slučajnu varijablu koristeći metodu generaliziranog inverza.

Funkcija distribucije F je

$$F(x) = 1 - q^n, \quad \text{za} \quad n-1 < x \leq n, \quad n \geq 1.$$

Pri traženju generaliziranog inverza, trebamo za $0 < u < 1$ pronaći $n \in \mathbb{N}$ takav da je

$$1 - q^{n-1} < u \leq 1 - q^n,$$

odakle dobijemo

$$n - 1 < \frac{\log(1-u)}{\log q} \leq n$$

pa je $n = \lceil \log(1-u)/\log q \rceil$.

```
> n<-1000
> p<-0.25
> u<-runif(n)
> k<-ceiling(log(1-u)/log(1-p))
```

2.2.4 Metoda odbacivanja

Pretpostavimo da možemo simulurati slučajnu varijablu s gustoćom g. Neka je f funkcija gustoće neke slučajne varijable takva da postoji konstanta $c > 0$ takva da je

$$\frac{f(t)}{g(t)} \leq c, \quad \text{za sve } \ t \text{ takve da je } \ f(t) > 0.$$

Tada slučajni varijablu X s gustoćom f možemo simulirati na sljedeći način:

1. Simuliramo Y s gustoćom g i $U \sim U(0,1)$.
2. Ako je
$$U \leq \frac{f(Y)}{g(Y)}$$
onda stavmo $X = Y$. Inače se vratimo na korak 1.

Napomena.

(a) Vrijedi

$$\begin{aligned}
\mathbb{P}(X \leq a) &= \mathbb{P}(Y \leq a | U \leq f(Y)/cg(Y)) \\
&= \frac{\mathbb{P}(Y \leq a, U \leq f(Y)/cg(Y))}{\mathbb{P}(U \leq f(Y)/cg(Y))} \\
&= \frac{\int_{-\infty}^{\infty} \mathbb{P}(Y \leq a, U \leq f(Y)/cg(Y)) | Y = y) g(y)\, dy}{\mathbb{P}(U \leq f(Y)/cg(Y))} \\
&= \frac{\int_{-\infty}^{a} \frac{f(y)}{cg(y)}\, g(y)\, dy}{\mathbb{P}(U \leq f(Y)/cg(Y))} \\
&= \frac{\int_{-\infty}^{a} f(y)\, dy}{c\,\mathbb{P}(U \leq f(Y)/cg(Y))}
\end{aligned}$$

pa za $a \to \infty$ dobijemo

$$1 = \frac{1}{c\,\mathbb{P}(U \leq f(Y)/cg(Y))}, \tag{2.1}$$

odakle je

$$\mathbb{P}(X \leq a) = \int_{-\infty}^{a} f(y)\, dy.$$

(b) Iz algoritma slijedi da će ova metoda generirati traženu slučajnu varijablu s vjerojatnošću

$$\mathbb{P}(U \leq f(Y)/cg(Y)) = \frac{1}{c},$$

odakle slijedi da broj iteracija ima geometrijsku razdiobu s parametrom $\frac{1}{c}$. Stoga je očekivani broj iteracija jednak c.

Zadatak 2.6 Simulirajte Beta$(2,4)$ razdiobu koristeći metodu odbacivanja.

Rješenje. Funkcija gustoće je

$$f(x) = B(2,4)x(1-x)^3 = \frac{5!}{1! \cdot 3!}x(1-x)^3 = 20x(1-x)^3, \ 0 < x < 1.$$

Uzmimo $g(x) = 1, \ 0 < x < 1$. Tada je

$$\frac{f(x)}{g(x)} = 20x(1-x)^3,$$

i deriviranjem dobijemo

$$\left(\frac{f(x)}{g(x)}\right)' = 20(1-x)^2(1-4x),$$

odakle vidimo da se maksimum postiže za $x_0 = \frac{1}{4}$. Stoga je

$$\frac{f(x)}{g(x)} \leq \frac{f(x_0)}{g(x_0)} = \frac{135}{64} =: c.$$

Dakle,

$$\frac{f(x)}{cg(x)} = \frac{256}{27}x(1-x)^3.$$

Simulirajmo uzorak duljine $n = 100$:

```
n<-100
k<-0   # je li prihvaceno
j<-0   # broj iteracija
y<-numeric(n)
while (k<n) {
    u<-runif(1)
    j<-j+1
    x<-runif(1) # simuliramo iz gustoce g
    if (256/27*x*(1-x)^3>u) {
        # prihvacamo x
        k<-k+1
        y[k]<-x
    }
}
```

Broj iteracija potrebnih za simuliranje uzorka duljine 100 je

```
> j
[1] 192
```

Primijetite da je očekivana vrijednost $100c = \frac{135}{64} = 210.9375$.

Zadatak 2.7 Simulirajte apsolutnu vrijednost jedinične normalne razdiobe koristeći metodu odbacivanja.

Rješenje. Stavimo

$$f(x) = \frac{2}{\sqrt{2\pi}} e^{-x^2/2},\ x > 0 \quad \text{i} \quad g(x) = e^{-x},\ x > 0.$$

Tada je

$$\frac{f(x)}{g(x)} = \sqrt{2e/\pi} e^{-(x-1)^2/2} \le \sqrt{2e/\pi} =: c$$

pa je

$$\frac{f(x)}{c\, g(x)} e^{-(x-1)^2/2}.$$

```
n<-100
k<-0   # je li prihvaceno
j<-0   # broj iteracija
y<-numeric(n)
while (k<n) {
    u<-runif(1)
    j<-j+1
    x<--log(runif(1)) # simuliramo eksponencijalnu razdiobu
    if (u<exp(-(x-1)^2/2)) {
        k<-k+1
        y[k]<-x
    }
}
```

2.3 Monte Carlo metoda

Neka js X slučajna varijabla i neka je

$$g\colon \mathbb{R} \to \mathbb{R}$$

funkcija takva da je $\mathbb{E}[g(X)] < \infty$.

Ako su $X_1, X_2, \ldots$ nezavisne kopije slučajne varijable X, onda jaki zakon velikih brojeva kaže da

$$\lim_n \frac{g(X_1) + \ldots + g(X_n)}{n} = \mathbb{E}[g(X)] \quad (g.s.). \tag{2.2}$$

Tako (2.2) daje jedan način za aproksimaciju $\mathbb{E}[g(X)]$ koristeći neku realizaciju slučajnih varijabli $X_1, X_2, \ldots$.

Definiramo **Monte Carlo procjenitelj** za $\mathbb{E}[g(X)]$ s

$$T_n = \frac{1}{n}\sum_{i=1}^{n} g(X_i).$$

Primijetimo da je procjenitelj T_n **nepristran**, t.j.

$$\mathbb{E}T_n = \frac{1}{n}\sum_{i=1}^{N} \mathbb{E}[g(X_i)] = \mathbb{E}[g(x)].$$

Ovakvo računanje ima nekoliko primjena.

(a) **Računanje vjerojatnosti.** Ako je Y neka slučajna varijabla, onda $\mathbb{P}(X \in A)$ možemo računati koristeći procjenitelj

$$\frac{1}{n}\sum_{i=1}^{n} 1_A(Y_i),$$

gdje su $Y_1, Y_2, \ldots$ nezavisne kopije slučajne varijable Y.

(b) **Računanje integrala.** Želimo izračunati

$$\int_a^b g(x)\,dx. \tag{2.3}$$

Primijetimo da je za $U \sim U(a, b)$

$$\mathbb{E}[g(U)] = \frac{1}{b-a}\int_a^b g(x)\,dx$$

pa je za nezavisne slučajne varijable $U_1, U_2, \ldots \sim U(a, b)$ procjenitelj za integral (2.3) dan s

$$\frac{b-a}{n}\sum_{i=1}^{n} g(U_i).$$

(c) **Računanje suma.** Ako želimo računati sume

$$\sum_{i=1}^{k} g(a_i), \tag{2.4}$$

gdje je (a_i) niz brojeva. Tada uzmemo niz nezavisnih slučajnih varijabli $U_1, U_2, \ldots \sim U(\{a_1, a_2, \ldots, a_k\})$ i definiramo procjenitelj za sumu (2.4)

$$\frac{k}{n}\sum_{i=1}^{n} g(U_i).$$

Zadatak 2.8 Izračunajte integral

$$\int_2^4 e^{-x}\,dx.$$

Rješenje.

Primijetimo da je integral jednak

$$\int_2^4 e^{-x}\,dx = e^{-2} - e^{-4}.$$

```
m<-10000
x<-runif(m,min=2,max=4)
integral<-2*mean(exp(-x))
print(integral)
print(exp(-2)-exp(-4))
```

Dobijemo npr.

```
[1] 0.1168570
[1] 0.1170196
```

Zadatak 2.9 Izračunajte

$$\int_0^\infty e^{-x^2}\,dx.$$

Rješenje.

Trebamo imati konačan interval pa prvo primijenimo supsituciju $y = \frac{1}{x+1}$. Tada je

$$\int_0^\infty e^{-x^2}\,dx = \int_0^1 e^{-(\frac{1}{y}-1)^2}\,\frac{dy}{y^2}.$$

```
m<-10000
x<-runif(m)
integral<-mean(exp(-(1/x-1)^2)/x^2)
print(integral)
print(sqrt(pi)/2)
[1] 0.8899181
[1] 0.886227
```

Zadatak 2.10 Odredite π simulacijama.

Rješenje.

Neka je (X,Y) slučajni vektor uniformo distribuiran na kvadratu stranice duljine 2 sa središtem u ishodištu. Primijetimo da je

$$\mathbb{P}(X^2+Y^2 \le 1) = \frac{1^2\pi}{2^2} = \frac{\pi}{4}.$$

Primijetimo da su X,$Y \sim U(-1,1)$ nezavisne slučajne varijable.

```
m<-1000000
x<-2*runif(m)-1 #uniformna razdioba na (-1,1)
y<-2*runif(m)-1 #uniformna razdioba na (-1,1)
procjena<-4*sum((x^2+y^2<1))/m
print(procjena)
print(pi)
[1] 3.142684
[1] 3.141593
```

Zadatak 2.11 Neka su $U_1, U_2, \ldots \sim U(0,1)$ nezavisne slučajne varijable i

$$N = \inf\{n\colon \sum_{i=1}^{n} U_i > 1\}.$$

Simulacijama odredite $\mathbb{E}N$.

Rješenje.

```
n<-10000
k<-1 # brojac
m<-numeric(n)
while (k<n) {
    suma<-0
    while (suma<1) {
        u<-runif(1)
        m[k]<-m[k]+1
        suma<-suma+u
    }
    k<-k+1
}
ocekivanje<-mean(m)
print(ocekivanje)
```

2.4 Uzorkovanje po važnosti

Uzorkovanje po važnosti koristimo da bi smanjili varijancu Monte Carlo procjenitelja za integral

$$\int g(x)\,dx.$$

Neka je X slučajna varijabla s gustoćom f takvom da je

$$f(x) > 0 \quad \text{za} \quad x \in \{x\colon g(x) > 0\}.$$

Označimo

$$Y = \frac{g(X)}{f(X)}.$$

Tada je

$$\int g(x)\,dx = \int \frac{g(x)}{f(x)} f(x)\,dx = \mathbb{E}Y.$$

Tada možemo procijeniti $\mathbb{E}Y$ pomoću Monte Carlo procjenitelja računajući

$$\frac{1}{n}\sum_{i=1}^{n} Y_n = \frac{1}{n}\sum_{i=1}^{n} \frac{g(X_i)}{f(X_i)}, \tag{2.5}$$

gdje su slučajne varijable $X_1, \dots, X_n$ nezavisne s gustoćom f. Gustoću f zovemo **funkcija važnosti**.

Varijanca procjenitelja (2.5) je

$$\operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n} Y_n\right) = \frac{\operatorname{Var} Y}{n}.$$

Ideja je da $\operatorname{Var} Y$ bude dosta mala, što odgovara tome da je Y blizu konstanti. To možemo postići odabirom funkcije f koja je 'blizu' g.

Primjer. Procijenimo integral

$$\int_0^1 \frac{e^{-x}}{1+x^2}\,dx$$

koristeći razne funkcije važnosti:

(a) $f_0(x) = 1$, $0 < x < 1$

(b) $f_1(x) = e^{-x}$, $x > 0$

(c) $f_2(x) = \frac{\pi}{1+x^2}$

(d) $f_3(x) = \frac{e^{-x}}{1-e^{-1}}$, $0 < x < 1$

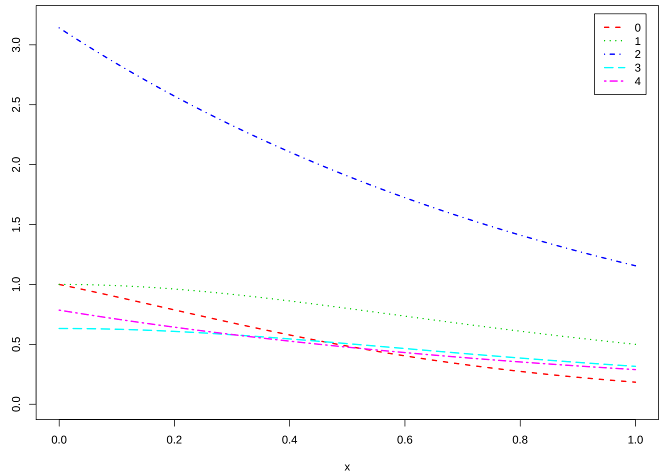
(e) $f_4(x) = \frac{4}{\pi(1+x^2)}$, $0 < x < 1$

```
m<-10000
int<-numeric(5)
dev<-numeric(5)
g<-function(x) {
exp(-x)/(1+x^2)*(x>0)*(x<1)
}

x<-runif(m)       # pomocu f0
fg<-g(x)
int[1]<-mean(fg)
dev[1]<-sd(fg)

x<-rexp(m,1)      # pomocu f1
```

```
fg<-g(x)/exp(-x)
int[2]<-mean(fg)
dev[2]<-sd(fg)

x<-rcauchy(m)      # pomocu f2
i<-c(which(x>1),which(x<0))   # izbacimo sve vrijednosti izvan (0,1)
x[i]<-2
fg<-g(x)/dcauchy(x)
int[3]<-mean(fg)
dev[3]<-sd(fg)

u<-runif(m)      # pomocu f3 i metode generaliziranog inverza
x<--log(1-u*(1-exp(-1)))
fg<-g(x)/(exp(-x)/(1-exp(-1)))
int[4]<-mean(fg)
dev[4]<-sd(fg)

u<-runif(m)      # pomocu f4 i metode generaliziranog inverza
x<-tan(pi*u/4)
fg<-g(x)/(4/((1+x^2)*pi))
int[5]<-mean(fg)
dev[5]<-sd(fg)
```

Promotrimo funkcije g/f_i:

```
x<-seq(0,1,0.01)
w<-2
f1<-exp(-x)
f2<-(1/pi)/(1+x^2)
f3<-exp(-x)/(1-exp(-1))
f4<-4/((1+x^2)*pi)
g<-exp(-x)/(1+x^2)

plot(x,g,type="l",main="",ylab="",
ylim=c(0,3.2),lwd=w,lty=2,col=2)
lines(x,g/f1,lty=3,lwd=w,col=3)
lines(x,g/f2,lty=4,lwd=w,col=4)
lines(x,g/f3,lty=5,lwd=w,col=5)
lines(x,g/f4,lty=6,lwd=w,col=6)
legend("topright",legend=c(0:4),
lty=2:6,lwd=w,inset=0.02,col=2:6)
```



Vidimo da je g/f_3 najbliža konstanti.

3

Markovljevi lanci

3.1 Uvod

Niz slučajnih varijabli $(X_n, n = 0, 1, 2, \ldots)$ zovemo **Markovljev lanac** sa skupom stanja S, ako vrijedi

$$\mathbb{P}(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1) = \mathbb{P}(X_{n+1} = j | X_n = i),$$

za sve $n \in \mathbb{N} \cup \{0\}$, $i, i_1, \ldots, i_n \in S$.

Gornju jednakosit možemo interpretirati na sljedeći način: 'budućnost' ovisi o 'prošlosti' samo kroz 'sadašnjost.

Od sada pretpostavljamo da je S najviše prebrojiv i da je Markovljev lanac **vremenski homogen**, tj.

$$\mathbb{P}(X_{n+1} = j | X_n = i) = \mathbb{P}(X_1 = j | X_0 = i), \quad \text{za sve} \quad n \in \mathbb{N} \cup \{0\}.$$

Možemo definirati vjerojatnosti $\mathbb{P}_i(\cdot)$, $i \in S$ s

$$\mathbb{P}_i(\cdot) := \mathbb{P}(\cdot | X_0 = i).$$

Definiramo vjerojatnost da Markovljev lanac prijeđe iz stanja i u j s

$$p_{ij} := \mathbb{P}(X_1 = j | X_0 = i) = P_i(X_1 = j).$$

Vrijedi

$$p_{ij} \geq 0, \ \sum_{y \in S} p_{ij} = 1.$$

Prijelazne vjerojatnosti zapisujemo u **matricu prijelaznih vjerojatnosti**, koja je npr. za $S = \{1, 2, 3, \ldots\}$ dana s

$$\begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots \\ p_{21} & p_{22} & p_{23} & \cdots \\ p_{31} & p_{32} & p_{33} & \cdots \\ \vdots & \vdots & \vdots & \end{bmatrix}$$

Primjer. (Vremenska prognoza) Pretpostavimo pojednostavljeni model predviđanja vremenea u kojem vremenska prognoza u nekom danu ovisi samo o vrememenskoj prognozi prethodnog dana. Neka je $S = \{0, 1\}$, gdje stanja označavaju

$$0 \ - \ \text{kiša}, \quad 1 \ - \ \text{nije kiša}.$$

Ako kiša pada danas, onda je vjerojatnost da će padati i sutra jednaka α. Ako kiša nije padala danas, onda je vjerojatnost će padati ni sutra jednaka β. Tada je matrica prijelaza

$$\begin{bmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{bmatrix}.$$

△

Primjer. (Slučajna šetnja) Promotrimo slučajnu šetnju na $\mathbb{Z}$:

$$X_0 = 0, \ X_n = \xi_1 + \ldots + \xi_n, \ n \geq 1,$$

gdje su

$$\xi_1, \xi_2, \ldots \sim \begin{pmatrix} -1 & 1 \\ 1-p & p \end{pmatrix}$$

nezavisne slučajne varijable. Tada je $(X_n, n = 0, 1, 2, \ldots)$ Markovljev lanac sa skupom stanja $S = \mathbb{Z}$ i matricom prijelaznih vjerojatnosti

$$p_{i,i+1} = p, \quad p_{i,i-1} = 1-p, \quad i \in \mathbb{Z}.$$

△

Označimo **prijelazne vjerojatnosti u** m **koraka** s $p_{ij}^{(m)}$:

$$p_{ij}^{(m)} = \mathbb{P}(X_{n+k} = j | X_k = i), \ n \geq 0, \ i, j \in S.$$

Zadatak 3.1 Dokažite **Chapman-Kolmogorovljevu jednakost**

$$p_{ij}^{(n+m)} = \sum_{k \in S} p_{ik}^{(n)} p_{kj}^{(m)}.$$

Napomena. Iz prethodnog zadatka za $m = 1$ dobijemo

$$p_{ij}^{(n+1)} = \sum_{k \in S} p_{ik}^{(n)} p_{kj}.$$

Ako s $P^{(n)}$ označimo matricu prijelaznih vjerojatnosti u n koraka, onda gornja jednakost prelazi u

$$P^{(n+1)} = P^{(n)} P.$$

Indukcijom slijedi da je

$$P^{(n)} = P^n.$$

Zadatak 3.2 Promotrite primjer vremenske prognoze s $\alpha = 0.7$ i $\beta = 0.4$. Ako je danas bila kiša, kolika je vjerojatnost da će kiša padati i za četiri dana.

Kažemo da stanja i i j **komuniciraju** ako postoje $m, n \in \mathbb{N}$ takvi da je

$$p_{ij}^m > 0 \quad \text{i} \quad p_{ji}^n > 0.$$

Intuitivno, iz stanja i u j možemo doći u nekoliko koraka, kao i obrnuto. Markovljev lanac je **ireducibilan** ako sva stanja komuniciraju.

Zadatak 3.3 Odredite koje matrice prijelaza odgovaraju ireducibilnom Markovljevom lancu:

$$\text{(a)} \begin{bmatrix} 0.3 & 0.7 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.1 & 0.9 \\ 0 & 0 & 0.8 & 0.2 \end{bmatrix} \qquad \text{(b)} \begin{bmatrix} 0.3 & 0.6 & 0.1 & 0 \\ 0.5 & 0.3 & 0 & 0.2 \\ 0 & 0 & 0.2 & 0.8 \\ 0 & 0.1 & 0.7 & 0.2 \end{bmatrix}?$$

Stanje $i \in S$ je **aperiodično** ako postoji $n \in \mathbb{N}$ takav da je

$$p_{ii}^{(n)} > 0 \quad \text{i} \quad p_{ii}^{(n+1)} > 0.$$

Kažemo da je stanje $i \in S$ **pozitivno povratno** ako je očekivano vrijeme potrebno Markovljevom lancu da iz stanja i opet dođe do i konačno:

$$\mathbb{E}_i[T_i] < \infty, \quad \text{gdje je} \quad T_i = \inf\{n \geq 1 \colon X_n = i\}.$$

Teorem. Ako je Markovljev lanac ireducibilan, aperiodičan i pozitivno povratan, onda postoji $\lim\limits_{n\to\infty} p_{ij}^{(n)}$ i vrijedi

$$\pi_j = \lim_{n\to\infty} p_{ij}^{(n)},$$

gdje je $\pi = (\pi_i \colon i \in S)$ jedinstveno rješenje sustava

$$\pi = \pi P \quad (\text{t.j. } \pi_j = \sum_{i\in S} \pi_i p_{ij})$$

takvo da je

$$\sum_{j\in S} \pi_j = 1.$$

Također vrijedi

$$\mathbb{E}_i[T_i] = \frac{1}{\pi_i},$$

gdje je $T_i = \inf\{n \geq 1 \colon X_n = i\}$ je prvo vrijeme povratka u stanje i. ■

Primijetimo da π definira vjerojatnost na S. π zovemo **stacionarna mjera** Markovljevog lanca.

Ako zadamo razdiobu slučajne varijable X_0, kažemo da smo zadali **početne vjerojatnosti** Markovljevog lanca. U slučaju da je

$$\mathbb{P}(X_0 = i) = a_i,\ i \in S, \quad \text{za}\ \ a = (a_1, a_2, \ldots),\ a_i \geq 0, \sum_{i \in S} a_i = 1,$$

dobijemo

$$\mathbb{P}(X_n = j) = \sum_{i \in S} \mathbb{P}(X_n = j, X_0 = i) = \sum_{i \in S} \mathbb{P}(X_n = j | X_0 = i)\mathbb{P}(X_0 = i) = \sum_{i \in S} a_i p_{ij}^n.$$

Dakle razdioba od X_n je vektorski dana s

$$aP^n.$$

Zadatak 3.4 Neka je $(X_n, n = 0, 1, 2 \ldots)$ Markovljev lanac koji ima stacionarnu mjeru π. Ako X_0 ima razdiobu π, odredite razdiobu od X_n.

Teorem. (Ergodski teorem) Ako je $(X_n, n = 0, 1, 2, \ldots)$ ireducibilan Markovljev lanac sa stacionarnom razdiobom π i ako je $f : S \to \mathbb{R}$ omeđena funkcija, onda je

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} f(X_n) = \sum_{j \in S} f(j)\, \pi_j.$$

■

Zadatak 3.5 Promotrimo primjer vremenske prognoze s $\alpha = 0.7$ i $\beta = 0.4$.

(a) Odredite postotak kišnih dana u nekom dužem periodu.

(b) Ako trgovima kišobranima zarađuje 1000 kn na kišne dane, a 200 kn na dane kada ne pada kiša, odredite prosječnu dnevnu zaradu trgovine u nekom dužem periodu.

(c) Koliki je prosječni razmak između dva kišna dana?

Zadatak 3.6 Muha se kreće po vrhovima trokuta na sljedeći način: ako je u vrhu i, onda se miče u vrh u smjeru kazaljke na satu s vjerojatnošću p_i, odnosno u smjeru suprotnom od kazaljke na satu s vjerojatnošću $q_i = 1 - p_i$, $i = 1, 2, 3$. Ako je $p_1 = 1/2$ i $p_2 = p_3 = 1/4$ odredite

(a) postotak vremena koje muha provede u svakom od vrhova;

(b) koliko često muha napravi pomak u smjeru suprotnom od smjera kazaljke na satu, a onda dva koraka u smjeru kazaljke na satu.

3.2 Simuliranje Markovljevog lanca

Primjer. Promotrimo igru zmije i ljestve. Na ploči označenoj brojevima od 1 do 9 igrač se kreće tako da svaki put baci novčić i ako padne pismo ide za 1 korak, a ako padne glava ide za 2 koraka.

7 - Lj1	8 - Z2	9 - Kraj
6 - Z1	5 - Lj2	4 - Z2
1 - Z1	2 - Lj1	3 - Lj2

Ako dođe npr na polje 2, onda se ljestvama (Lj1) popne na polje 7. S druge strane, ako je npr. na polju 6, onda se po zmiji (Z1) spusti do polja 1. Simulacijama odredite očekivano trajanje igre.

```
p<-rep(0,81)
dim(p)<-c(9,9)
p[1,2]<-0.5
p[1,3]<-0.5
p[2,7]<-1
p[3,5]<-1
p[4,5]<-0.5
p[4,6]<-0.5
p[5,6]<-0.5
p[5,7]<-0.5
p[6,1]<-1
p[7,8]<-0.5
p[7,9]<-0.5
p[8,4]<-1
p[9,9]<-1
igra<-function() {
stanja<-c(1)
     i<-1
     repeat  {
          tmp<-sample(1:9,size=1,replace=T,prob=p[stanja[i],])
          stanja<-c(stanja,tmp)
          if (tmp==9) break
          i<-i+1
      }
      return(stanja)
}
ocek<-function(n) {
        s<-0
        for (i in 1:n) s<-s+length(igra())
        return(s/n)
}
```

Tako npr. dobijemo

```
> ocek(100)
[1] 11.79
> ocek(100)
[1] 13.29
> ocek(100)
[1] 12.98
```

3.3 Markov Chain Monte Carlo

MCMC metoda se koristi kod simuliranja slučajnih varijabli koje nije lako simulirati metodama koje smo da sada vidjeli. Ideja je da se tražena razdioba pokuša dobiti kao stacionarna rardioba π ireducibilnog Markovljevog lanca.

Neka je $\pi = (\pi_i \colon i \in S)$ neka razdioba i neka je Q matrica prijelaza ireducibilnog Markovljevog lanca. Definiramo novi Markovljev lanac $(X_n, n = 0, 1, 2, \ldots)$ na sljedeći način:

(a) Ako je $X_n = i$, onda simuliramo Y iz razdiobe q_i ..

(b) Stavimo $X_{n+1} = Y$ s vjerojatnošću $\alpha(i, Y)$, a inače stavimo $X_{n+1} = X_n$.

Ako definiramo

$$\alpha(i, j) = \min\left(\frac{\pi(j)q_{ji}}{\pi(i)q_{ij}}, 1\right),$$

onda se lako vidi da je π stacionarna razdioba Markovljevog lanca $(X_n, n = 0, 1, 2 \ldots)$.

Sve možemo sažeti u tzv. **Metropolis-Hastings** algoritam:

1. Neka je Q ireducibilna matrica i neka je $k \in S$ neko stanje (koje možemo mijenjati deterministički ili slučajno).

2. Neka je $n = 0$ i $X_0 = k$.

3. Simulirajmo slučajnu varijablu Y iz razdiobe $q_{X_n,\cdot}$. Simulirajmo $U \sim U(0, 1)$.

4. Ako je

$$U < \frac{\pi(Y)q_{Y,X_n}}{\pi(X_n)q_{X_n,Y}}$$

   onda stavimo $X_{n+1} = Y$, a inače $X_{n+1} = X_n$

5. $n = n + 1$

6. Idemo na korak 3.

Primjer. Promatramo mrežu $\{1,2,\ldots,n\} \times \{1,2,\ldots,n\}$ u rešetci $\mathbb{Z}^2$ kao graf. Svaki vrh može biti u stanju 0 ili 1. Kažemo da je konfiguracija dopustiva, ako je nema susjednih vrhova sa stanjem 1. Uz pretpostavku da su sve dopustive konfiguracije jednako vjerojatne trebamo simulirati jednu takvu konfiguraciju. Kombinatorno nije trivijalno odrediti broj svih konfiguracija N pa ćemo se poslužiti MCMC metodom pri simuliranju.

Neka je $V = \{1,2,\ldots,n\} \times \{1,2,\ldots,n\}$. Tada je skup stanja Markovljevog lanca

$$S = \{\xi \in \{0,1\}^V : \xi \text{ je dopustiva}\}.$$

Dakle, trebamo simulirati Markovljev lanac čija je stacionarna distribucija dana s

$$\pi(\xi) = \begin{cases} 1/N & \xi \in \{0,1\}^V \text{ dopustiv} \\ 0 & \text{inače} \end{cases}$$

```
sim<-function(n,m) {
  stanje<-rep(0,(n+2)^2)
  dim(stanje)<-c(n+2,n+2)
  for (l in 1:m) {
    k<-sample(1:(n*n),replace=T,size=1)
    i<-ceiling(k/n)+1
    j<-k-n*(i-2)+1
    ind<-stanje[i,j+1]+stanje[i,j-1]+stanje[i-1,j]+stanje[i+1,j]
    u<-runif(1)
    if ((u<0.5) & (ind==0))
      stanje[i,j]<-1
    }

    plot(expand.grid(1:n,1:n),col=stanje[2:(n+1),2:(n+1)]+3,pch=15,
    xlab="",ylab="",main="")
}
```

Zadaci za vježbu

3.7 Neka je X Markovljev lanac s dva stanja i matricom prijelaza

$$P = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}.$$

Dokažite matematičkom indukcijom da je

$$P^n = \begin{bmatrix} 0.5 + 0.5(2p-1)^n & 0.5 - 0.5(2p-1)^n \\ 0.5 - 0.5(2p-1)^n & 0.5 + 0.5(2p-1)^n \end{bmatrix}.$$

Što možete reći o stacionarnoj distribuciji ovog Markovljevog lanca?

3.8 U nekoj tvrtci je N zaposlenika, gdje je N velik. Svaki zaposlenik radi jedan od 3 posla, pri čemu mijenja vrstu posla po matrici

$$\begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.4 & 0.5 \end{bmatrix}.$$

Koliko posto zaposlenika radi svaki od poslova?

3.9 U nekom gradu dva sunčana dana nikada ne dolaze zaredom. Svaki dan može biti sunčan, oblačan ili kišan. Ako je dan sunčan, onda je jednako vjerojatno da sljedeći bude oblačan ili kišan. Ako je nekog dana oblačno ili kišno, onda u jednom od dva slučaja vrijeme ostaje isto sljedećeg dana, dok u drugom slučaju su preostale mogućnosti za sljedeći dan jednako vjerojatne. Koliki postotak sunčanih dana ima taj grad?

3.10 Čestica se giba po kružnici na kojoj su u smjeru kazaljke na satu točke označene s 0,1,2,3,4. U svakom trentutku čestica ide u smjeru kazaljke na satu s vjerojatnošću p, a u obrnutom smjeru s vjerojatnošću $1 - p$.

(a) Odredite matricu prijelaza.

(b) Koji je očekivani broj koraka koji je potreban čestici da se vrati u početni položaj?

4

Poissonov proces

4.1 Homogeni Poissonov proces

Definicija. Process $(N(t), t \geq 0)$ s vrijednostima u skupu $\{0, 1, 2, \ldots\}$ je **Poissonov proces s intenzitetom** $\lambda > 0$, $\lambda > 0$,ako je

(i) $N_0 = 0$

(ii) N ima **nezavisne priraste**, t.j. za $0 \leq t_1 < t_2 < \ldots < t_n$ su slučajne varijable

$$N_{t_1}, N_{t_2} - N_{t_1}, \ldots, N_{t_n} - N_{t_{n-1}}$$

nezavisne.

(iii) Za $0 \leq s < t$ je

$$N_t - N_s \sim P(\lambda(t - s)).$$

Napomena.

(a) Primijetimo da iz svojstva (iii) slijedi da je za $0 \leq s < t$

$$N_t - N_s \stackrel{d}{=} N_{t-s}.$$

Kažemo da N ima **stacionarne priraste**.

(b) Vrijedi

$$\mathbb{E}N_t = \lambda t.$$

(c) Poissonov proces je primjer procesa koji broji koliko se puta dogodio neki događaj do trenutka t.

Definicija. Dolazna vremena Poissonovog procesa definiramo kao

$$T_n = \inf\{t \geq 0\colon N_t \geq n\},\ n \in \{0, 1, 2 \dots\}.$$

Vrijeme T_n označava vrijeme kada se n-ti put pojavio događaj koji broji Poissonov proces.

Definicija. Međudolazna vremena Poissonovg procesa su definirana kao

$$W_n = T_n - T_{n-1},\ n \in \{1, 2, 3, \dots\}.$$

W_n je vrijeme koje od $(n-1)$-og to n-tog događaja.

Može se pokazati da su $W_1, W_2, \dots \sim Exp(\lambda)$ nezavisne. Tada je

$$T_n = W_1 + \dots + W_n \sim \Gamma(n, \frac{1}{\lambda}).$$

Zadatak 4.1 U neku ordinaciju pacijenti dolaze kao Poisonov proces s intenzitetom $\lambda = \frac{1}{10}$ po minuti. Doktor ima pravilo da ne prima pacijente sve dok u čekaonici ne bude barem 3 pacijenta.

(a) Koje je očekivano vrijeme kada doktor primi prvog pacijenta?

(b) Kolika je vjerojatnost da doktor ne primi ni jednog pacijenta sat vremena?

Zadatak 4.2 Useljenici ulaze na neki teritorij neke države kao Poissonov proces s intenzitetom $\lambda = 1$ dnevno.

(a) Koje je očekivano vrijeme ulaska dvadesetog useljenika?

(b) Kolika je vjerojatnost da između ulaska dvadesetog i dvadesetprvog usljenika prođe više od 10 dana?

Promotrimo sada pojavu poznatu kao stanjivanje Poissonovog procesa. Pretpostavimo da imamo Poissonov proces $(N_t, t \geq 0)$ s intenzitetom $\lambda > 0$. Svaki put kada se dogodi događaj, s vjerojatnošću $p \in \langle 0, 1\rangle$ ga registriramo. Ako broj registriranih događaja do trenutka $t > 0$ označimo s X_t, onda je $(X_t, t \geq 0)$ Poissonov proces s intenzitetom λp.

Zadatak 4.3 U neku zemlju doseljenici dolaze kao Poissonov proces s intenzitetom od 10 doseljenika tjedno. Svaki doseljenik je britanskog podrijetla s vjerojatnošću 1/12. Kolika je vjerojatnost da se u sljedeća 4 tjedna neće u tu zemlju neće useliti nijedan Britanac?

Zadatak 4.4 (Superpozicija Poissonovih procesa) Neka su $(X_t, t \geq 0)$ i $(Y_t, t \geq 0)$ nezavisni Poissonovi procesi s intenzitetima $\lambda > 0$ i $\mu > 0$. Dokažite da je s $N_t = X_t + Y_t$ definiran Poissonov proces s intenzitetom $\lambda + \mu$.

4.2 Nehomogeni Poissonov proces

Definicija. Proces $(N_t, t \geq 0)$ s vrijednostima u skupu $\{0, 1, 2, \ldots\}$ je **nehomogeni Poissonov proces s funkcijom intenziteta** $\lambda(t)$, $t \geq 0$, ako vrijedi

(i) $N_0 = 0$

(ii) N ima nezavisne priraste

(iii) za $0 \leq s < t$ je

$$N_t - N_s \sim P(\int_s^t \lambda(u)\, du).$$

Napomena. Ako je $\lambda(t) = \lambda$, onda dobijemo homogeni Poissonov proces.

Zadatak 4.5 Trgovina se otvara ujutro u 8 sati. Od 8 do 11 kupci dolaze u prosjeku s intenzitetom koji linerano raste od 5 kupaca po satu u 8 ujutro do 20 kupaca po satu u 11 sati. Od 11 do 13 sati intenzitet je konstantan i jednak 20 kupaca po satu, Od 13 do 17 sati intenzitet pada i u 17 sata je jednak 12 kupaca po sau. Pretpostavimo da je broj kupaca u disjunktnim vremenskim periodima nezavisan.

(a) Kolika je vjerojatnost da nema kupaca od 8:30 do 9:30?

(b) Koji je očekivani broj kupaca u periodu od 10 do 12 sati?

4.3 Složeni Poissonov proces

Definicija. Složeni Poissonov proces je proces $(X_t, t \geq 0)$ definiran na sljedeći način:

$$X_t = \sum_{i=1}^{N_t} Y_i,\ t \geq 0,$$

gdje je $(N_t, \geq 0)$ Poissonov proces s intenzitetom $\lambda > 0$, a $\{Y_i, i \geq 1\}$ je familija nezavisnih i jednako distribuiranih slučajnih varijabli nezavisna od Poissonovog procesa N.

Napomena.

(a) Ako stavimo $Y_i = 1$, onda je $X_t = N_t$ pa dobijemo Poissonov proces.

(b) Zbog $N_t \sim P(\lambda t)$ vidimo da je

$$\mathbb{E}X_t = \mathbb{E}N_t \mathbb{E}Y_1 = \lambda t \mathbb{E}Y_1$$

i

$$\operatorname{Var} X_t = \mathbb{E}N_t \operatorname{Var} Y_1 + \operatorname{Var} N(\mathbb{E}Y_1)^2 = \lambda t \mathbb{E}[Y_1^2].$$

Zadatak 4.6 Osiguravajuća kuća isplaćuje police osiguranja kao Poissonov proces s intenzitetom 5 tjedno. Ako je svota novca koja se isplati eksponencijalno distribuirana s očekivanjem od 10 000 kn, odredite očekivanu količinu novca koju osiguravajuća kuća isplati u 4 tjedna.

4.4 Simuliranje Poissonovog procesa

Homogeni Poissonov proces možemo simulirati koristeći činjenicu da su međudolazna vremena nezavisna i eksponencijalno distribuirana.

```
poisson<-function(lambda,t) {
     Tn<-rexp(100*floor(t),lambda)    #medjudolazna vremena
     Sn<-cumsum(Tn)        #dolazna vremena
     return(min(which(Sn>t)))
}
```

Tako je npr.

```
> poisson(3,10)
[1] 36
```

Zadaci za vježbu

4.7 Muhe i ose dolijeću u juhu kao nezavisni Poissonovi procesi s intenzitetima λ i μ. Odredite očekivani broj insekata u juhi u trenutku $t > 0$. Odredite očekivano vrijeme ulijetanja n-rog insekta u juhu. Kolika je vjerojatnost da je prvi insekt koji uleti u juhu osa?

4.8 Prema nekoj znanstvenoj teoriji u dijeljenju stanica se pojavljuju greške po Poissonovom procesu s intenzitetom od 2.5 stanice po godini. Smatra se da neka osoba umre nakon 196 pogrešnih dijeljenja stanica.

(a) Izračunaje prosječni životni vijek osobe.

(b) Kolika je vjerojatnost da osoba doživi starost od 100 godina?

4.9 Vrijeme između dolaska dva vlaka na stanicu je uniformno distribuirano na $(0, 1)$. Putnici dolaze na stanicu po Poissonovom procesu s intenzitetom 7 po satu. Pretpostavite da je vlak upravo napustio stanicu. Neka je X broj putnika koji će stići na sljedeći vlak.

(a) Izračunajte $\mathbb{E}X$.

(b) Izračunajte $\operatorname{Var} X$.

4.10 Klijenti dolaze u banku po Poissonovom procesu s intanzitetom $\lambda > 0$. Ako su u prvom satu u banku došla samo 2 klijenta, kolika je vjerojatnost

(a) da su oba stigla u prvih 20 minuta,

(b) da je barem jedan od njih stigao u prvih 20 min?

4.11 Događaji se pojavljuju po nehomogenom Poissonovom procesu $(X_t, t \geq 0)$ takvom da je

$$\mathbb{E}X_t = t^2 + 2t, \; t \geq 0.$$

Kolika je vjerojatnost da se n događaja pojavi između vremena $t = 4$ i $t = 5$?

4.12 Posjetitelji dolaze u neki restoran po Poissonovom procesu s intenzitetom $\lambda = 10$ po satu. Čim uđu u restoran, prvo pogledaju jelovnik i s vjerojatnošću 0.6 ostanu u restoranu na ručku.

(a) Kolika je vjerojatnost da u prva 3 sata u restoranu ruča barem 3 posjetitelja?

(b) Odredite očekivani broj posjetitelja koji ruča u restoranu u prva 3 sata.

5

Uvod u Bayesovsku statistiku

Često želimo nešto zaključiti o nekom parametru razdiobe ako imamo neke podatke iz te razdiobe. U klasičnom pristupu pretpostavljamo da je parametar nepoznata konstanta, a da podaci predstavljaju realizaciju slučajnog uzorka.

S druge strane, kod Bayesovskog pristupa uzimamo da je parametar koji još ne znamo slučajan, a podaci koje smo dobili konstantni. Kod odabira razdiobe parametra razdiobu možemo odabrati na temelju poznavanja načina kako se provodi slučajni pokus.

5.1 Bayesov teorem

Pretpostavimo da imamo podatke $x = (x_1, \dots, x_n)$ i želimo nešto zaključiti o parametru θ koji je slučajan. Tada za funkciju gustoće vrijedi

$$\begin{aligned} f(\theta|x) &= \frac{f(x,\theta)}{f(x)} = \frac{f(x|\theta)f(\theta)}{f(x)} \\ &\propto f(x|\theta)f(\theta), \end{aligned}$$

jer pretpostavljamo da je $f(x)$ konstanta. Dakle,

$$f(\theta|x) \propto \prod_{i=1}^{n} f(x_i|\theta)f(\theta)$$

ili

$$\text{posterior} \propto \text{likelihood} \times \text{prior}.$$

Ovdje je

(a) $f(\theta)$ **apriorna** gustoća od θ (prior),

(b) $f(x|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$ **vjerodostojnost** (likelihood),

(c) $f(\theta|x)$ **aposteriorna** gustoća od θ (posterior).

Zadatak 5.1 Pretpostavimo da $X_1, \ldots, X_n|\theta \sim B(m, \theta)$ nezavisne i da je $\theta \sim \text{Beta}(\alpha, \beta)$, gdje su $\alpha, \beta > 0$ konstante. Odredite aposteriornu gustoću od θ i izračunajte aposteriorno očekivanje od θ.

Zadatak 5.2 Broj e-mail poruka koji svaki dan primi neka osoba ima Poissonovu razdiobu s očekivanjem λ, gdje je apriorna razdioba od λ eksponencijalna s očekivanjem μ. Pretpostavimo da osoba ima podatke $x_1, \ldots, x_n$, gdje je x_i broj poruka i-tog dana, $i = 1, \ldots, n$.

(a) Odredite aposteriornu razdiobu od λ.

(b) Ako je $\mu = 50$ i ako je osoba primila ukupno 550 poruka u 10 dana, izračunajte Bayesovsku procjenu za λ.

Zadatak 5.3 Zanima nas vjerojatnost θ da će pasti pismo na novčiću ako ga bacimo na određeni način. Pretpostavljamo da je apriorna razdioba beta razdioba s parametrima $\alpha = \beta = 4$. Novčić je bačen 10 puta, a mi znamo samo da je palo manje od 3 pisma. Odredite aposteriornu razdiobu i izračunajte joj konstantu normalizacije.

5.2 Predviđanje

Želimo pronaći buduću vrijednost Z na temelju opaženih podataka. Vrijedi

$$\begin{aligned} f(z|x) &= \frac{f(z,x)}{f(x)} = \frac{\int f(z,x,\theta)\, d\theta}{f(x)} \\ &= \frac{\int f(z|\theta,x) f(\theta,x)\, d\theta}{f(x)} \\ &= \int f(z|\theta,x) f(\theta|x)\, d\theta. \end{aligned}$$

Pretpostavljamo da su X i Z uvjetno nezavisni uz dano θ, t.j.

$$f(x,z|\theta) = f(x|\theta) f(z|\theta).$$

Zato je

$$\begin{aligned} f(z|\theta,x) &= \frac{f(z,x,\theta)}{f(x,\theta)} = \frac{f(z,x|\theta) f(\theta)}{f(x|\theta) f(\theta)} \\ &= \frac{f(z,x|\theta)}{f(x|\theta)} = \frac{f(z|\theta) f(x|\theta)}{f(x|\theta)} = f(z|\theta). \end{aligned}$$

Dakle,

$$f(z|x) = \int f(z|\theta)f(\theta|x)\, d\theta.$$

Za očekivanje buduće vrijednosti Z zbog uvjetne nezavisnosti of X i Z uz uvjet θ dobijemo

$$\mathbb{E}[Z|X] = \mathbb{E}[\mathbb{E}[Z|X,\theta]|X] = \mathbb{E}[\mathbb{E}[Z|\theta]|X].$$

Zadatak 5.4 Odredite očekivanje buduće vrijednosti u Zadatku 5.1.

5.3 MCMC metoda

Ponekad nije jednostavno izračunati konstante kod aposteriorne razdiobe. Unatoč tome, moguće je simulirati tu razdiobo koristeći Markov Chain Monte Carlo metodu.

Ideja je pronaći ireducibilan i aperiodičan Markovljev lanac s vrijednostima u $\mathbb{R}$ kojem je razdioba koju želimo simulirati stacionarna. Primijetite da sada imamo 'prijelazne' vjerojatnosti u obliku funkcije $K(s,r)$.

Metropolis-Hastings algoritam je sljedeći:

(i) Odaberimo neku razdiobu $q(s|r)$.

(ii) Generiramo X_0 iz q.

(iii) Ovaj korak ponavljamo sve dok razdioba od X_n nije dovoljno blizu stacionarnoj razdiobi:

   (a) Generiramo Y iz $q(\cdot|X_t)$.

   (b) Generiramo $U \sim U(0,1)$.

   (c) Ako je
   $$U \leq \frac{f(Y)q(X_n|Y)}{f(X_n)q(Y|X_n)},$$
   onda stavimo $X_{n+1} = Y$, a inače $X_{n+1} = X_n$.

   (d) Povećamo n.

Pokaže se da ako razdioba $\pi(t)$ zadovoljava **uvjet detaljne ravnoteže**:

$$\pi(t)K(t,s) = \pi(s)K(s,t),$$

da je onda stacionarna. Provjerimo da Metropolis-Hastings algoritam zaista generira Markovljev lanac s traženom razdiobom.

U koraku (iii) (c) vidimo da prihvaćamo novu vrijednost s vjerojatnošću

$$\alpha(X_n, Y) = \min\left(1, \frac{f(Y)q(X_n|Y)}{f(X_n)q(Y|X_n)}\right)$$

pa je dovoljno poznavati razdiobu f do na konstantu.

Primjer. Simuliramo Rayleighovu razdiobu, t.j. razdiobu s gustoćom

$$f(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}, \ x > 0.$$

Odaberimo $q(\cdot, r) \sim \chi^2(r)$. Tada je

$$\alpha(X_n, y) = \frac{f(y)q(X_n|y)}{f(X_n)q(y|X_n)} = \frac{ye^{-y^2/2\sigma^2}}{X_n e^{-X_n^2/2\sigma^2}} \cdot \frac{\Gamma(X_n/2)2^{X_n/2}X_n^{y/2-1}e^{-X_n/2}}{\Gamma(y/2)2^{y/2}y^{X_n/2-1}e^{-y/2}}.$$

```
f<-function(x,sigma) return((x/sigma^2)*exp(-x^2/(2*sigma^2)))
m<-10000
sigma<-4
x<-numeric(m)
x[1]<-rchisq(1,df=1)
k<-0
u<-runif(m)

for (i in 2:m) {
     xt<-x[i-1]
     y<-rchisq(1,df=xt)
     alpha<-(f(y,sigma)*dchisq(xt,df=y))/(f(xt,sigma)*dchisq(y,df=xt))
     if (u[i]<=alpha) x[i]<-y else {
          x[i]<-xt
          k<-k+1
     }
}
```

Često se koriste jednostavniji oblici distirbucije q. Npr. $q(r|s) = q(|r-s|)$. Tada je

$$\alpha(r, s) = \min\left(1, \frac{f(s)}{f(r)}\right)$$

5.4 Gibbsov sampler

Ako slučajni vektor (X, Y) ima gustoću $f(x, y)$, onda možemo simulirati (X, Y) pomoću markovljevog lanca (X_n, Y_n) na sljedeći način:

(i) Stavimo $X_0 = x_0$.

(ii) Generiramo Y_{n+1} iz razdiobe $f_{Y|X}(\cdot|X_n)$

(iii) Generiramo Y_{n+1} iz razdiobe $f_{X|Y}(\cdot|Y_{n+1})$

Primjer. Pretpostavimo da imamo razdiobe

$$X|\theta \sim B(n,\theta) \quad \theta \sim \text{Beta}(\alpha,\beta).$$

Tada dobijemo zajedničku razdiobu

$$f(x,\theta) \propto \theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}$$

i posteriorna razdioba je

$$\theta|x \sim \text{Beta}(x+\alpha, n-x+\beta-1)$$

```
m<-10000
n<-10
alpha<-2
beta<-4
x<-numeric(m)
t<-numeric(m)
t[1]<-rbeta(1,alpha,beta)
x[1]<-rbinom(1,n,t[1])
k<-0
u<-runif(m)

for (i in 2:m) {
     x[i]<-rbinom(1,n,t[i-1])
     t[i]<-rbeta(1,alpha+x[i],n-x[i]+beta)
}
```

Zadaci za vježbu

5.5 Za svaku od m nezavisnih polica, vjerojatnost jedne štete u godini je θ, a vjerojatnost niti jedne štete je $1-\theta$. Ukupan broj šteta je slučajna varijabla X. Dostupna su nezavisna opažanja $x_1,\ldots,x_n$ od X. Apriorna distribucija od θ ima gustoću

$$f(\theta) \propto (\theta(1-\theta))^{\beta-1},\ 0<\theta<1,$$

gdje je $\beta > 0$ neka konstanta. Odredite aposteriornu distribuciju od X uz dane $x_1,\ldots,x_n$ i izračunajte Bayesovski procjenitelj od θ ako je

$$n=6,\ m=10,\ x_1=1,\ x_2=4,\ x_3=2,\ x_4=1,\ x_5=1, x_6=3,\ \beta=1.$$

5.6 Neka su $X_1,\ldots,X_n$ uvjetno nezavisne uz dano θ, pri čemu $X_i|\theta$ ima Poissonovu razdiobu s parametrom θ.

(a) Ako je apriorna razdioba od θ gama razdioba s parametrima α i $\beta > 0$, odredite aposteriornu razdiobu od θ.

(b) Odredite očekivanje buduće vrijednosti.

5.7 Pokažite da je aposteriorno očekivanje u Zadatku 5.3 jednako 39/128.

Notes

[1]Ova tvrdnja jednostavno slijedi iz formule (1.2). Naime, zbog nezavisnosti od X i Y je

$$G_{X+Y}(s) = \mathbb{E}[s^X s^Y] = \mathbb{E}[s^X]\mathbb{E}[s^Y] = G_X(s)G_Y(s).$$

[2]Primijetimo da je (1.1) red potencija pa ga možemo derivirati član po član. Tako je ($f^{(n)}$ označava n-tu derivaciju funkcije f)

$$G_X^{(n)}(s) = \sum_{k=n}^{\infty} p_n n(n-1)\cdots(n-k+1)s^{n-k},$$

odakle za $s = 0$ dobijemo

$$\mathbb{P}(X = n) = p_n = \frac{G_X^{(n)}(0)}{n!}.$$

Stoga, ako se G_X i G_Y podudaraju na okolini nule, onda su im sve derivacije u 0 jednake, iz čega slijedi tvrdnja.

[3]Deriviranjem reda potencija (1.1) dobijemo

$$\mathbb{P}_X^{(n)}(s) = \sum_{k=n}^{\infty} p_n n(n-1)\cdots(n-k+1)s^{n-k}.$$

Postojanje n-tog momenta povlači konvergenciju reda $\sum_{k=n}^{\infty} p_n n(n-1)\cdots(n-k+1)$ pa tvrdnja slijedi iz Abelovog teorema.

[4]Za $k = 1$ vrijedi

$$M_X'(0) = \lim_{t\to 0}\frac{M_X(t) - M_X(0)}{t} = \lim_{t\to 0}\mathbb{E}\left[\frac{e^{tX}-1}{t}\right] = \mathbb{E}\left[\lim_{t\to 0}\frac{e^{tX}-1}{t}\right] = \mathbb{E}X,$$

gdje smo upotrijebili Lebesgueov teorem o dominiranoj konvergenciji u trećoj jednakosti.

[5]Na $\mathbb{R}^d$ možemo definirati tzv. Borelovu σ-algebru $\mathcal{B}(\mathbb{R}^d)$, t.j. σ-algebru generiranu svim otvorenim pravokutnicima $(a_1, b_1) \times \cdots \times (a_d, b_d)$. Elemente σ-algebre $\mathcal{B}(\mathbb{R}^d)$ zovemo izmjerivi skupovi.

[6]Ako su $(X, \mathcal{F})$ i $(Y, \mathcal{G})$ dva izmjeriva prostora, onda kažemo da je $f\colon X \to Y$ izmjeriva, ako je $f^{-1}(A) \in \mathcal{F}$, za sve $A \in \mathcal{G}$. Funkcija $f\colon \mathbb{R}^2 \to \mathbb{R}$ je izmjeriva, ako je izmjeriva u odnosu na izmjerive prostore $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ i $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Rješenja

1.43 $G_X(s) = \frac{1-s^n}{ns^n}$, $\mathbb{E}X = \frac{n+1}{2}$, $\operatorname{Var} X = \frac{n^2-1}{12}$

1.44 (a) $G_X(s) = q + ps$, $\mathbb{E}X = p$, $\operatorname{Var} X = pq$ (b) $G_X(s) = \frac{ps}{1-qs}$, $\mathbb{E}X = \frac{1}{p}$, $\operatorname{Var} X = \frac{q}{p^2}$ (c) $G_X(s) = (q + ps)^n$, $\mathbb{E}X = np$, $\operatorname{Var} X = npq$

1.45 $s^b G_X(s^a)$

1.47 $(2 - \frac{s(1-s^6)}{6(1-s)})^{-1}$

1.48 $\frac{25}{216}$

1.49 0.028

1.50 Broj uboda ima Poissonovu razdiobu s parametrom λp (usporedite taj rezultat kasnije s pojmom stanjivanja Poissonovog procesa) pa je prosječni broj uboda jednak λp.

1.51 $G_X(s) = \frac{\ln(1-ps)}{\ln(1-p)}$, $\mathbb{E}X = \frac{-p}{(1-p)\ln(1-p)}$

1.52 $X_1 \sim P(\lambda/n)$

1.55 $M(t) = (1 - 2t)^{-n/2}$ (razdioba je $\chi^2(n) \sim \Gamma(n/2, 2)$)

1.56 (a) Da (b) Da (c) Da

1.57 1/3

1.58 Da. Ne.

1.60 $e^{-a^2/a}/\sqrt{2\pi}$

1.61 $L/3$

1.62 Ne

1.63 X/Y ima jediničnu Cauchyevu razdiobu.

1.63 R i Θ su nezavisne. $\Theta \sim U(0, 2\pi)$, a R ima tzv. Rayleighovu razdiobu, t.j.

$$f_R(r) = re^{-r^2/2},\ r > 0.$$

1.65 Da. $R^2 \sim Exp(1/2) \sim \chi^2(2)$.

1.67 Komponente su nezavisne, pri čemu je $X + Y \sim \Gamma(\alpha + \beta, 1/\lambda)$ i $X/(X + Y) \sim B(\alpha, \beta)$.

1.69 2

1.70 $\mathbb{E}[X|Y=y]=y$

1.71 $C=1/8$, $\mathbb{E}[X|Y=y]=0$

1.72 $1/(n-1)$

1.73 $P(X=n)=2^{-n-1}$, $n=0,1,2,\dots$

1.74 Očekivano broj ljudi koji dođe prije Marka je 5, a varijanca je 10.

1.75 Očekivano vrijeme izlaska je 9.5, a varijanca je 113.5.

1.76 $1+1/\lambda$

1.77 $(2+2q+q^2/p)/(1-q^2)$

1.78 $(n+1)2^{-n-2}$

1.79 $c^{-2/\alpha}\Gamma(1+2/\alpha)-c^{-2/\alpha}\left[\Gamma(1+1/\alpha)\right]^2$

1.80 (a) $\mathbb{P}(X>2)=e^{-4}$, $\mathbb{P}(1<X<3)=e^{-1}-e^{-9}$, (b) $\lambda(t)=2t$

1.81 (a) 443.11 (b) 53650.5 (c) 0.2212

1.82 (a) 0.9332 (b) 20952.2 (c) $\mathbb{E}X=13359.7$, $\operatorname{Var}X=1.45\cdot 10^{12}$

1.83 $\mathbb{E}X=e^{\mu+\sigma^2/2}$, $\operatorname{Var}X=e^{2\mu+\sigma^2}(e^{\sigma^2}-1)$.

1.84 Intenzitet hazarda pada.