VIS - Labosi (prvi ciklus)

Uvod u R na primjerima zadataka iz diskretne vjerojatnosti

Vanessa Keranović, Kristijan Kilassa Kvaternik, Mate Puljiz, Stjepan Šebek, Josip Žubrinić

??.??.2019.

Uvod

Labosi na predmetu "Vjerojatnost i statistika" izvode se u programskom jeziku R, radnoj okolini RStudio, u obliku R Markdown izvještaja koji kombiniraju pisanje teksta s programskim kodom i rezultatima izvođenja koda. Predznanje ovih alata nije nužno za izvedbu jer se kroz labose demonstriraju ključne funkcionalnosti. Kao dodatne materijale preporučamo udžbenik "Programirajmo u R-u" doc. dr. sc. Damira Pintara, dostupan na stranicama vještine "Osnove programskog jezika R" (https://www.fer.unizg.hr/predmet/opjr).

R Markdown

R Markdown dokument sastavljen je od isječaka koda u R-u i teksta oko njih. Trenutnu liniju koda izvodimo kombinacijom tipaka `CTRL+ENTER`, a cijeli isječak kombinacijom `CTRL+SHIFT+ENTER`. Iz R Markdown dokumenta moguće je stvoriti izvještaj u PDF, HTML, DOCX ili drugim formatima (`output` parametar u zaglavlju dokumenta) kombinacijom tipaka `CTRL+SHIFT+K`.

Zadaci

Na ovim labosima riješit ćemo nekoliko zadataka pomoću simulacija u R-u, oslanjajući se pritom na jaki zakon velikih brojeva.

Zadatak 1.

Iz kutije u kojoj se nalazi 5 crnih, 6 bijelih i 7 zelenih kuglica izvlačimo na sreću 4 kuglice. Odredite vjerojatnost da među izvučenim kuglicama: (a) nema crnih, (b) nisu zastupljene sve boje.

Egzaktno rješenje:

(a) $\frac{\binom{13}{4}}{\binom{18}{4}}$,

(b) $1 - \frac{\binom{5}{2}\binom{6}{1}\binom{7}{1}}{\binom{18}{4}} - \frac{\binom{5}{1}\binom{6}{2}\binom{7}{1}}{\binom{18}{4}} - \frac{\binom{5}{1}\binom{6}{1}\binom{7}{2}}{\binom{18}{4}}$.

Simulacije:

```r
set.seed(1518141)
kutija = rep(c("c", "b", "z"), c(5, 6, 7))
broj_ponavljanja = 1e+05
nema_crnih = 0
```

```
nisu_zastupljene_sve_boje = 0

for (i in 1:broj_ponavljanja) {
    uzorak = sample(kutija, size = 4, replace = FALSE)
    nema_crnih = nema_crnih + (!is.element("c", uzorak))
    nisu_zastupljene_sve_boje = nisu_zastupljene_sve_boje + (!(is.element("c",
        uzorak) & is.element("b", uzorak) & is.element("z", uzorak)))
}

a_dio_sim = nema_crnih/broj_ponavljanja
b_dio_sim = nisu_zastupljene_sve_boje/broj_ponavljanja

# Egzaktno rjesenje
a_dio_egz = choose(13, 4)/choose(18, 4)
b_dio_egz = 1 - (choose(5, 2) * choose(6, 1) * choose(7, 1))/choose(18,
    4) - (choose(5, 1) * choose(6, 2) * choose(7, 1))/choose(18,
    4) - (choose(5, 1) * choose(6, 1) * choose(7, 2))/choose(18,
    4)

a_dio_sim
```

```
## [1] 0.23336
```

```
a_dio_egz
```

```
## [1] 0.2336601
```

```
b_dio_sim
```

```
## [1] 0.48432
```

```
b_dio_egz
```

```
## [1] 0.4852941
```

Zadatak 2.

Novčić bacamo dok se dva puta za redom ne pojavi isti znak. Opišite vjerojatnosni prostor i izračunajte vjerojatnost da pokus završi u parnom broju bacanja.

Egzaktno rješenje:

Vjerojatnosni prostor:

$$
\begin{aligned}
\Omega = &\{PP, PGPP, PGPGPP, \dots, PG\cdots PGPP, \dots\} \\
&\cup \{PGG, PGPGG, PGPGPGG, \dots, PG\cdots PGG, \dots\} \\
&\cup \{GG, GPGG, GPGPGG, \dots, GP\cdots GPGG, \dots\} \\
&\cup \{GPP, GPGPP, GPGPGPP, \dots, GP\cdots GPP, \dots\} \\
&\cup \underbrace{\{PGPGPG\dots, GPGPGP\dots\}}_{\mathbf{P}=0} \\
= &A \cup B \cup C \cup D \cup E
\end{aligned}
$$

$$\mathbf{P}(A\cup C) = 2\mathbf{P}(A) = 2\mathbf{P}(\cup_{n=0}^{\infty}\{\underbrace{PG\cdots PG}_{2n}PP\}) = 2\sum_{n=0}^{\infty}\mathbf{P}(\underbrace{PG\cdots PG}_{2n}PP)$$
$$= 2\sum_{n=0}^{\infty}\left(\frac{1}{2}\right)^{2n+2} = \frac{1}{2}\sum_{n=0}^{\infty}\left(\frac{1}{4}\right)^{n} = \frac{1}{2}\frac{1}{1-\frac{1}{4}} = \frac{2}{3}$$

Simulacije:

```
pokus = function() {
    bacanje = rbinom(1, size = 1, prob = 0.5)
    i = 1
    while (1) {
        sljedece_bacanje = rbinom(1, size = 1, prob = 0.5)
        i = i + 1
        if (sljedece_bacanje == bacanje) {
            break
        } else {
            bacanje = sljedece_bacanje
        }
    }
    return(i)
}

set.seed(1956819)

broj_ponavljanja = 10000
zavrsilo_u_parno_bacanja = 0

for (i in 1:broj_ponavljanja) zavrsilo_u_parno_bacanja = zavrsilo_u_parno_bacanja +
    ((pokus()%%2) == 0)

rj_sim = zavrsilo_u_parno_bacanja/broj_ponavljanja
rj_egz = 2/3

rj_sim
```

```
## [1] 0.6634
```

```
rj_egz
```

```
## [1] 0.6666667
```

Zadatak 3.

U poslovnici A nalazi se 100 srećki od kojih je 25 dobitnih, a u poslovnici B 55 srećki od kojih je 5 dobitnih. Marko baca simetričnu kocku - ako na kocki padne broj 1 kupuje dvije srećke u poslovnici A, ako na kocki padne 2 kupuje dvije srećke u poslovnici B, inače kupuje po jednu srećku u svakoj poslovnici. Kolika je vjerojatnost da je točno jedna kupljena srećka dobitna?

Egzaktno rješenje:

- $A = \{\text{točno jedna kupljena srećka je dobitna}\}$
- $H_1 = \{\text{pala je jedinica}\}$
- $H_2 = \{\text{pala je dvojka}\}$
- $H_3 = \{\text{palo je 3, 4, 5 ili 6}\}$
- $\mathbf{P}(H_1) = \frac{1}{6}$, $\mathbf{P}(H_2) = \frac{1}{6}$, $\mathbf{P}(H_3) = \frac{4}{6}$
- $\mathbf{P}(A \mid H_1) = \frac{25 \cdot 75}{\binom{100}{2}} = \frac{25}{66}$, $\mathbf{P}(A \mid H_2) = \frac{5 \cdot 50}{\binom{55}{2}} = \frac{50}{297}$, $\mathbf{P}(A \mid H_3) = \frac{25 \cdot 50 + 75 \cdot 5}{100 \cdot 55} = \frac{13}{44}$,
- $\mathbf{P}(A) = \sum_{i=1}^{3} \mathbf{P}(A \mid H_i)\mathbf{P}(H_i) = \frac{1027}{3564} = 0.288159$

Simulacije:

```
# oznacimo s 0 listice koji nisu dobitni, a s 1 dobitne
# listice
set.seed(5382523)
A = rep(0:1, c(100 - 25, 25))
B = rep(0:1, c(55 - 5, 5))

broj_ponavljanja = 1000
tocno_jedna_dobitna = 0

for (i in 1:broj_ponavljanja) {
    kocka = sample.int(6, size = 1)
    if (kocka == 1) {
        uzorak = sample(A, size = 2, replace = FALSE)
    } else if (kocka == 2) {
        uzorak = sample(B, size = 2, replace = FALSE)
    } else {
        uzorak = c(sample(A, size = 1), sample(B, size = 1))
    }
    tocno_jedna_dobitna = tocno_jedna_dobitna + (sum(uzorak) ==
        1)
}

rj_sim = tocno_jedna_dobitna/broj_ponavljanja
rj_egz = 1027/3564

rj_sim
```

```
## [1] 0.289
```

```
rj_egz
```

```
## [1] 0.2881594
```

Zadatak 4.

Dva igrača naizmjence bacaju simetrični novčić. Pobjednik je onaj kojem prvom padne glava, a za nagradu dobije iznos (u kunama) dvostruko veći od broja bacanja u igri. Koja je vjerojatnost da prvi igrač osvoji više od 100kn?

Egzaktno rješenje:

- $X =$ broj bacanja do pojave glave (uključivo), $X \sim \mathcal{G}(\frac{1}{2})$
- $\mathbf{P}(X = k) = (\frac{1}{2})^{k-1}(\frac{1}{2}) = (\frac{1}{2})^k$

$$\begin{aligned}\mathbf{P}(\{\text{prvi igrač je osvojio više od 100kn}\}) &= \mathbf{P}(\{X = 51\} \cup \{X = 53\} \cup \dots) \\ &= \sum_{k=25}^{\infty} (\frac{1}{2})^{2k+1} = \frac{1}{2} \sum_{k=25}^{\infty} (\frac{1}{4})^k = \frac{1}{2} \frac{(\frac{1}{4})^{25}}{1 - \frac{1}{4}} = \frac{1}{6 \cdot 4^{24}}\end{aligned}$$

Simulacije:

```
pokus = function() {
    i = 1
    while (1) {
        novcic = sample(c("p", "g"), size = 1)
        if (novcic == "g")
            break
        i = i + 1
    }
    return(i)
}

broj_ponavljanja = 1e+05
prvi_igrac_preko_100 = 0

for (i in 1:broj_ponavljanja) {
    broj_bacanja = pokus()
    if (((broj_bacanja%%2) == 1) & (2 * broj_bacanja > 100))
        prvi_igrac_preko_100 = prvi_igrac_preko_100 + 1
}

rj_sim = prvi_igrac_preko_100/broj_ponavljanja
rj_egz = 1/(6 * 4^24)

rj_sim
```

```
## [1] 0
```

```
rj_egz
```

```
## [1] 5.921189e-16
```

Zadatak 5.

Na sreću biramo točku unutar kvadrata $[-1, 1]^2$. Kolika je vjerojatnost da se ta točka nalazi unutar kruga oko ishodišta radijusa 1? Koristeći gornji postupak, aproksimirajte vrijednost broja π.

Egzaktno rješenje:

- $A = \{\text{odabrana točka nalazi se unutra kruga radijusa } 1\}$
- $K(0,1) = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$
- $\mathbf{P}(A) = \frac{m(K(0,1))}{m([-1,1]^2)} = \frac{\pi}{4}$

Simulacije:

```
set.seed(309501)
broj_ponavljanja = 1e+06
pogodjen_krug = 0

for (i in 1:broj_ponavljanja) {
    tocka = runif(2, min = -1, max = 1)
    if (sum(tocka^2) <= 1)
        pogodjen_krug = pogodjen_krug + 1
}

rj_sim = pogodjen_krug/broj_ponavljanja
rj_egz = pi/4

rj_sim
```

```
## [1] 0.785348
```

```
rj_egz
```

```
## [1] 0.7853982
```

```
# Aproksimacija za pi
4 * rj_sim
```

```
## [1] 3.141392
```

Zadatak 6.

Konobar počinje smjenu s 0kn. Od svakog gosta dobije napojnicu i to od 10kn ili 5kn, pri čemu je manja napojnica dva puta vjerojatnija. Izračunajte vjerojatnost da je konobar dobio manje od 30kn od 4 gosta.

Egzaktno rješenje:

- $X = \text{broj većih napojnica}, \ X \sim \mathcal{B}(4, \frac{1}{3})$

$$\begin{aligned}\mathbf{P}(10X + 5(4 - X) < 30) &= \mathbf{P}(5X < 10) = \mathbf{P}(X < 2) \\ &= \mathbf{P}(X = 0) + \mathbf{P}(X = 1) \\ &= \binom{4}{0}\left(\frac{1}{3}\right)^0\left(\frac{2}{3}\right)^4 + \binom{4}{1}\left(\frac{1}{3}\right)^1\left(\frac{2}{3}\right)^3 = \frac{16}{27}\end{aligned}$$

Simulacije:

```
set.seed(105179)
broj_ponavljanja = 10000

broj_vecih_napojnica = rbinom(broj_ponavljanja, size = 4, prob = 1/3)
ukupna_napojnica = 10 * broj_vecih_napojnica + 5 * (4 - broj_vecih_napojnica)

rj_sim = sum(ukupna_napojnica < 30)/broj_ponavljanja
rj_egz = pbinom(1, size = 4, prob = 1/3)

rj_sim
```

```
## [1] 0.5985
```

```
rj_egz
```

```
## [1] 0.5925926
```

Zadatak 7.

Na letu ima mjesta za 300 putnika, pri čemu će putnik koji je kupio kartu zakasniti na njega s vjerojatnošću 0.01. Stoga je aviokompanija prodala više karata, njih 302. Kolika ja vjerojatnost da će na letu biti mjesta za sve putnike s kartom?

Egzaktno rješenje:

- $X =$ broj putnika koji su zakasnili, $X \sim \mathcal{B}(302, 0.01) \approx \mathcal{P}(3.02)$

$$\mathbf{P}(X \geq 2) = 1 - \mathbf{P}(X = 0) - \mathbf{P}(X = 1) = 1 - \binom{302}{0} 0.01^0 0.99^{302} - \binom{302}{1} 0.01^1 0.99^{301} = 0.8053$$

ili

$$\mathbf{P}(X \geq 2) = 1 - \mathbf{P}(X = 0) - \mathbf{P}(X = 1) = 1 - \frac{3.02^0}{0!} e^{-3.02} - \frac{3.02^1}{1!} e^{-3.02} = 0.8038$$

Simulacije:

```
set.seed(63962171)
broj_ponavljanja = 10000

broj_putnika_koji_kasne = rbinom(broj_ponavljanja, size = 302,
    prob = 0.01)

rj_sim = sum(broj_putnika_koji_kasne >= 2)/broj_ponavljanja
rj_egz = 1 - pbinom(1, size = 302, prob = 0.01)
rj_approx = 1 - ppois(1, lambda = 302 * 0.01)

rj_sim
```

```
## [1] 0.8037
```

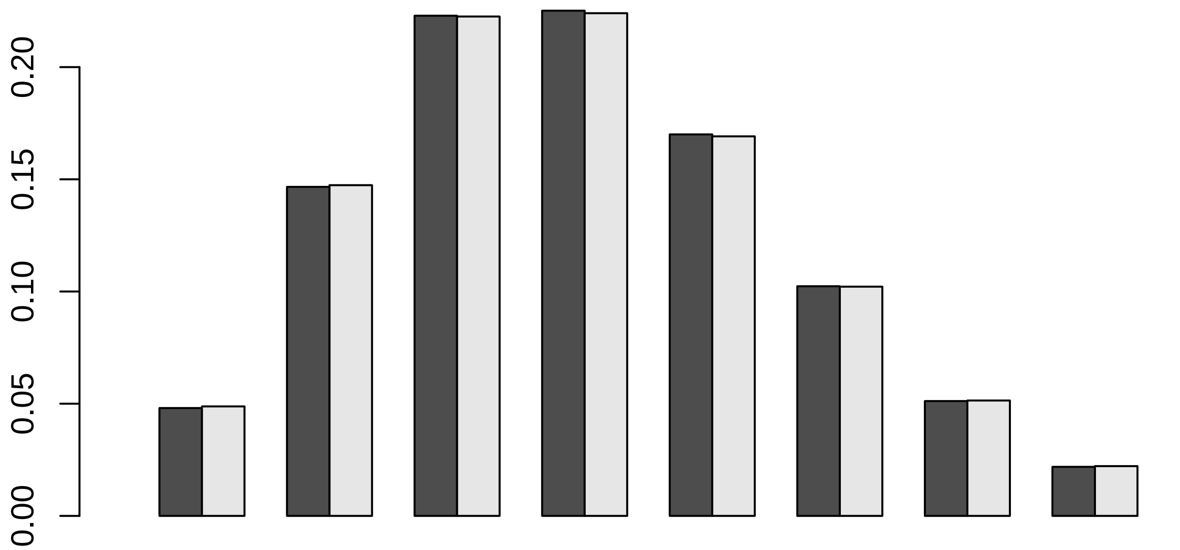
```r
rj_egz
```

```
## [1] 0.8053126
```

```r
rj_approx
```

```
## [1] 0.8038191
```

```r
# Ilustracija aproksimacije binomne slucajne varijable
# Poissonovom
n = 302
p = 0.01
barplot(rbind(dbinom(0:n, size = n, prob = p), dpois(0:n, lambda = n *
    p)), beside = T, xlim = c(0, 22))
```




```r
n = 302
p = 0.1
barplot(rbind(dbinom(0:n, size = n, prob = p), dpois(0:n, lambda = n *
    p)), beside = T, xlim = c(0, 150))
```