cv2.R

Marek

2023-02-23

```
#Diskretne a spojite rozdelenia pravdepodobnosti
#Kazde rozdelenie ma svoj zakladny prikaz a k nemu dane
#pismeno, podla toho, co ratame
# d ratame P(x=k)
# p ratame P(X\leq k)
# q ratame kvantily
# r generujeme nahodne cislo z daneho rozdelenia
# Binomicke rozdelenie
# V statistickej kontrole akosti, n/krat nezavisle na sebe opakujeme
# pokus, sledujeme vyskyt danej udalosti
# Pravdepodobnost danes udalosti v jedinom pokuse je p,
# parametre su n,p, binom

# Pr.1
# Ucinost antibiotika je  80%. Podavame ho 10 pacientom na oddeleni.
# Zratajte tieto pravdepodobnosti:
# Vsetci pacienti sa vyliecia P(x=10)
# Prave 7 sa vylieci P(x=7)
# Najviac 8 sa vylieci P(x<=8).
# Aspon 5 sa vylieci P(x>=5) = P(x>4)  ostra nerovnost je lepsia

n <- 10
p <- 0.8
dbinom(10, n, p) #prva uloha
```

```
## [1] 0.1073742
```

```
dbinom(7, n, p) #druha uloha
```

```
## [1] 0.2013266
```

```
pbinom(8, n, p) #tretia uloha
```

```
## [1] 0.6241904
```

```
pbinom(4, n, p, lower.tail = F) #stvrta uloha - prva moznost
```

```
## [1] 0.9936306
```

```
1 - pbinom(4, n, p) #stvrta uloha - druha moznost
```

```
## [1] 0.9936306
```

```
# DO tej hodnoty je neostra, OD tej hodnoty je ostra
# pbinom ked kumulujem

#Zostrojte tabulku a graf rozdelenia pravdepodobnosti
xB <- 0:10
hustotaB <- dbinom(xB, n, p)
hustotaB
```

```
##  [1] 0.0000001024 0.0000040960 0.0000737280 0.0007864320 0.0055050240
##  [6] 0.0264241152 0.0880803840 0.2013265920 0.3019898880 0.2684354560
## [11] 0.1073741824
```

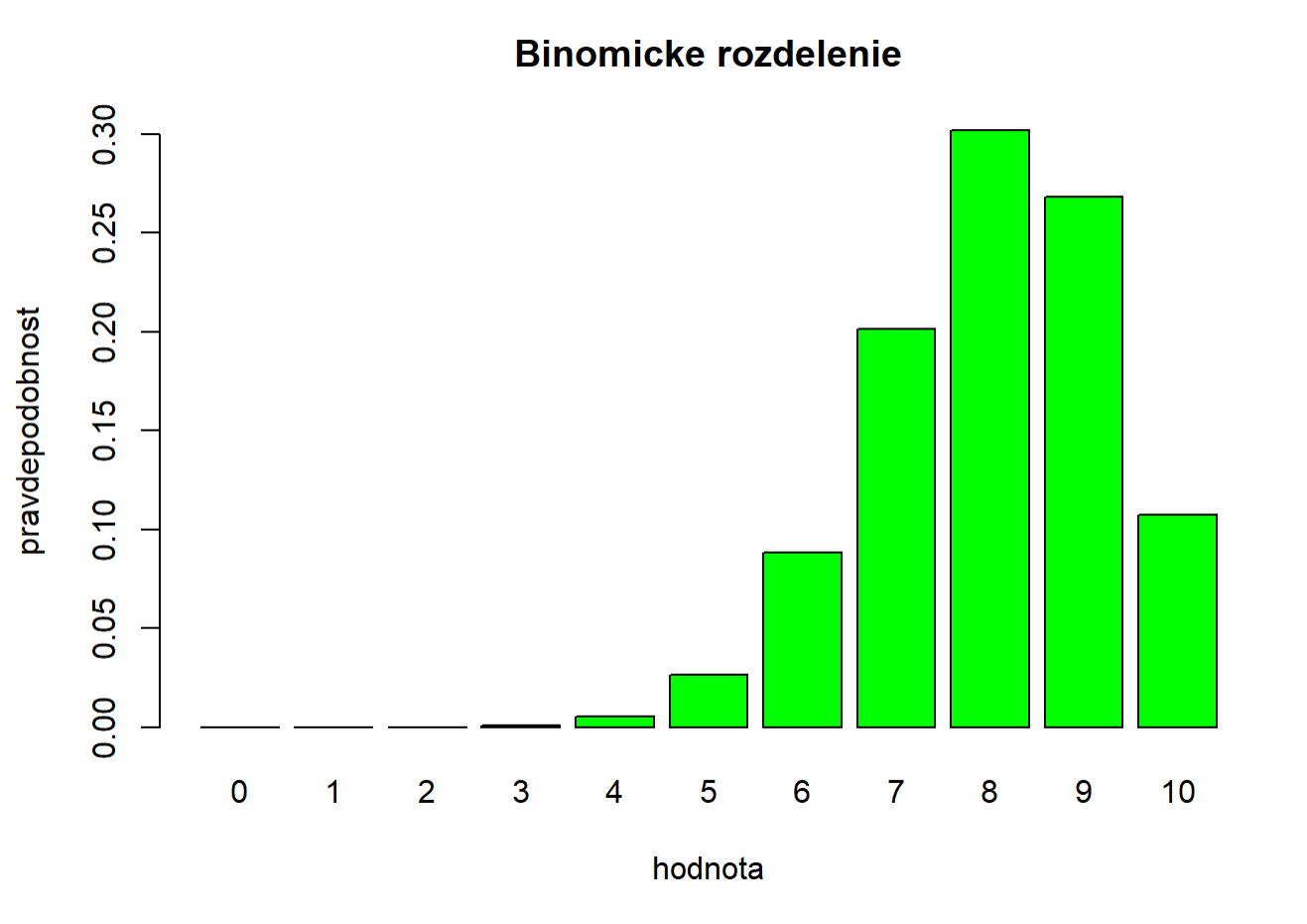
```
(tabulkaB <- data.frame(hodnota = xB, pravdepodobnost = hustotaB))
```

```
##    hodnota pravdepodobnost
## 1        0    0.0000001024
## 2        1    0.0000040960
## 3        2    0.0000737280
## 4        3    0.0007864320
## 5        4    0.0055050240
## 6        5    0.0264241152
## 7        6    0.0880803840
## 8        7    0.2013265920
## 9        8    0.3019898880
## 10       9    0.2684354560
## 11      10    0.1073741824
```

```
tabulkaB
```

```
##    hodnota pravdepodobnost
## 1        0    0.0000001024
## 2        1    0.0000040960
## 3        2    0.0000737280
## 4        3    0.0007864320
## 5        4    0.0055050240
## 6        5    0.0264241152
## 7        6    0.0880803840
## 8        7    0.2013265920
## 9        8    0.3019898880
## 10       9    0.2684354560
## 11      10    0.1073741824
```

```
View(tabulkaB)
barplot(tabulkaB$pravdepodobnost, main = 'Binomicke rozdelenie', names.arg = xB,
        xlab = 'hodnota', ylab = 'pravdepodobnost', col = 'green')
```




```
# hodnoty distribucnej funkcie
(distB <- pbinom(xB, n, p))
```

```
##  [1] 0.0000001024 0.0000041984 0.0000779264 0.0008643584 0.0063693824
##  [6] 0.0327934976 0.1208738816 0.3222004736 0.6241903616 0.8926258176
## [11] 1.0000000000
```

```
##############################################
# Hypergeometricke rozdelenie
# Pouzitie v statistickej kontrole akosti. Mnozina obsahuje
# m-prvkov so sledovanou vlastnostou a
# n-prvkov bez tejto vlastnosti,
# nahodne vyberieme k-prvkov.
# Nahodna premenna je pocet prvkov so sledovanou vlastnostou v nasom vybere.
# Prikaz je hyper, pismenka platia, parametre v poradi m, n, k

# Priklad
# Student sa nauci na skusku 12 z 20 otazok. Test obsahuje 5 otazok.
# Vypocitajte nasledujuce pravdepodobnost:
# Student dostane Acko, zodpovie vsetky otazky P(x=5)
# Student neurobi skusku, zodpovie menej ako 3 P(x<=2)  pre dolnu hranicu neostra
# Student urobi skusku (doplnok k predchadzajucej), zodpovie 3 a viac P(x>2)
# Parametre su m = 12, n = 8, k = 5
m <- 12
n <- 8
k <- 5
# prva uloha
dhyper(5, m, n, k)
```

```
## [1] 0.05108359
```

```
# druha uloha
phyper(2, m, n, k)
```

```
## [1] 0.2961816
```

```
# tretia uloha
1 - phyper(2, m, n, k) # ako doplnok
```

```
## [1] 0.7038184
```

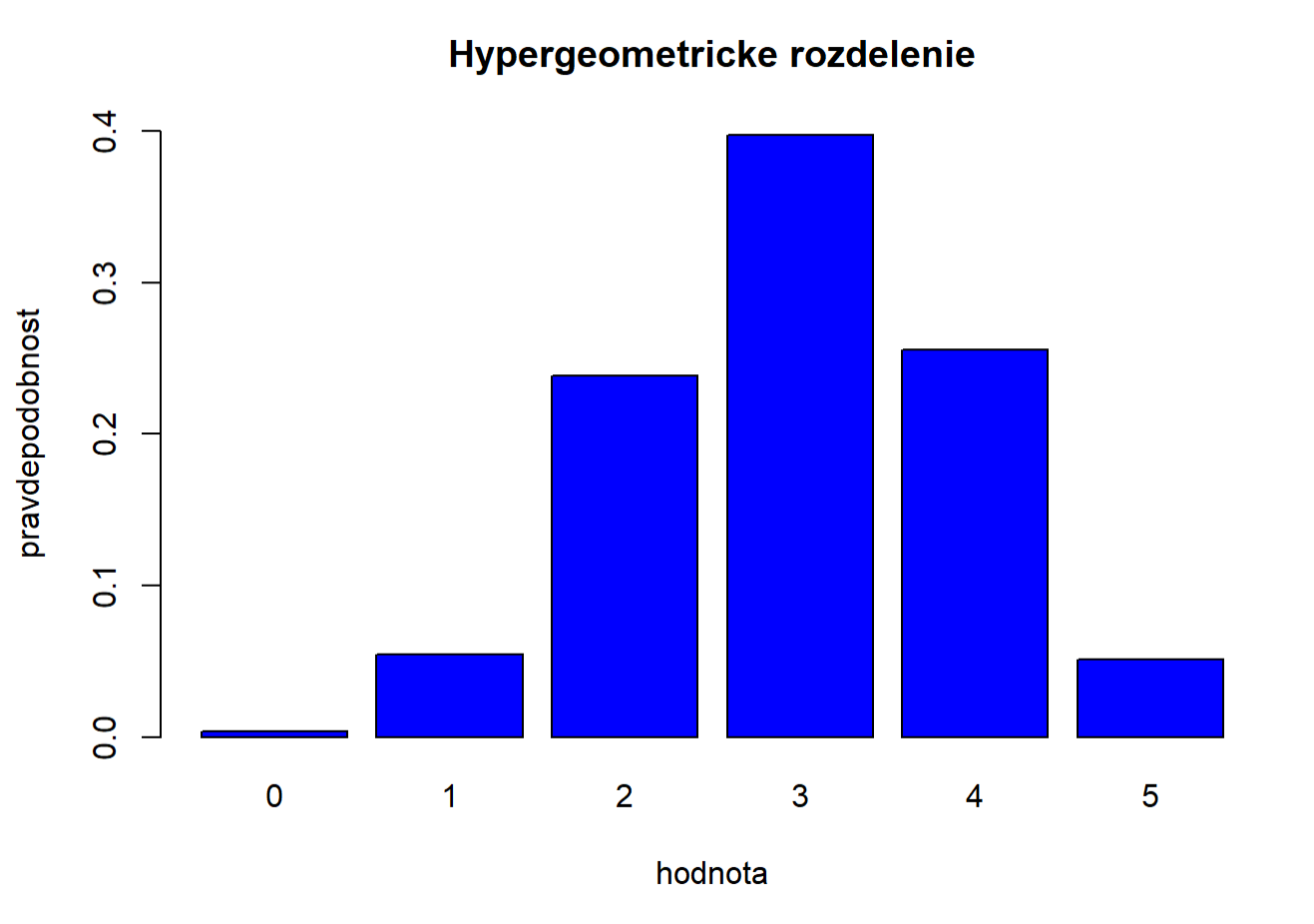
```
phyper(2, m, n, k, lower.tail = F)
```

```
## [1] 0.7038184
```

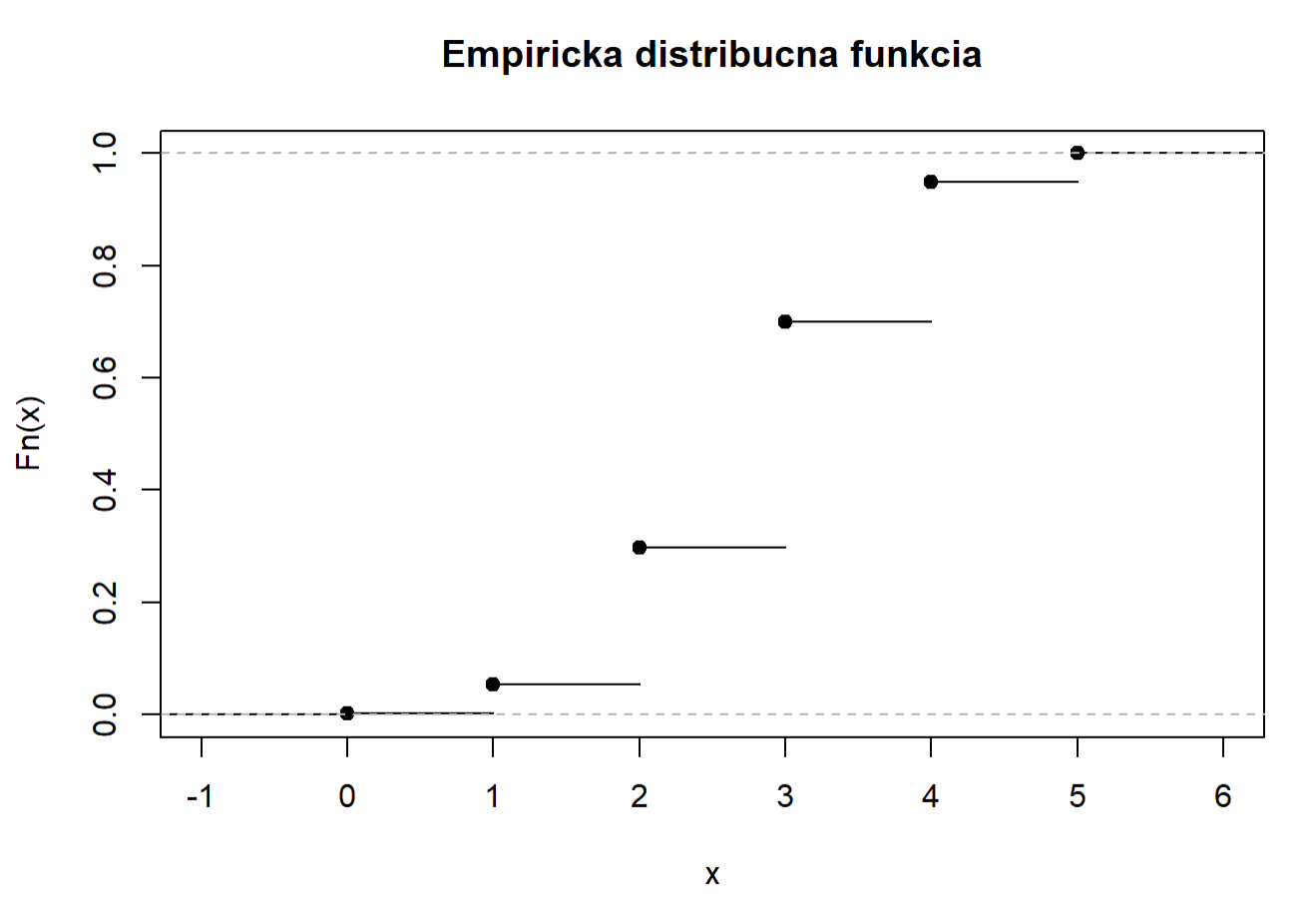
```
# tabulka a graf rozdelenia pravdepodobnosti
xH <- 0:5 # viem zodpovedat ani jednu az vsetkych 5 otazok
hustotaH <- dhyper(xH, m, n, k)
tabulkaH <- data.frame(hodnota = xH, pravdepodobnost = hustotaH)
tabulkaH
```

```
##   hodnota pravdepodobnost
## 1       0     0.003611971
## 2       1     0.054179567
## 3       2     0.238390093
## 4       3     0.397316821
## 5       4     0.255417957
## 6       5     0.051083591
```

```
View(tabulkaH)
barplot(
  tabulkaH$pravdepodobnost,
  main = 'Hypergeometricke rozdelenie',
  names.arg = xH,
  xlab = 'hodnota',
  ylab = 'pravdepodobnost',
  col = 'blue',
  ylim = c(0,0.4)
)
```




```
# nakreslite empiricku distribucnu funkciu pomocou nasimulovanych dat 10 000
data <- rhyper(10000, m, n, k)
plot(ecdf(data), main = 'Empiricka distribucna funkcia')
```




```
##########################################
# Poissonovo rozdelenie
# pouziva sa v teorii hromadnej obsluhy, pravdepodobnosti zriedkavych javov
# v casovom intervale, na nejakom objeme. Ma jediny parameter lambda
# Pri zmene casoveho intervalu treba parameter tiez prepocitat.
# Lambda je ocakavana hodnota v zadani.
# Prikaz je pois
# Na samoobsluznu linku pride 20 ludi za hodinu. Vypocitajte tieto pravdepodobnosti:
# V priebehu 15 min. pride 1 clovek P(x = 1), prepocet lambda = 20/60 je jedna minuta
# 15 = 5
# V priebehu 5 min. nikto nepride P(x = 0), lmbda = 20/60, 5 = 5 / 3
# V priebehu 10 min. pride aspon 10 ludi, P(x >= 10) = P(x > 9),
# lambda = 20/60, 10 = 510/3

#prva uloha
dpois(1, 5)
```

```
## [1] 0.03368973
```

```
#druha uloha
dpois(0, 5/3)
```

```
## [1] 0.1888756
```

```
#tretia uloha
ppois(9, 10/3, lower.tail = F)
```

```
## [1] 0.002356375
```

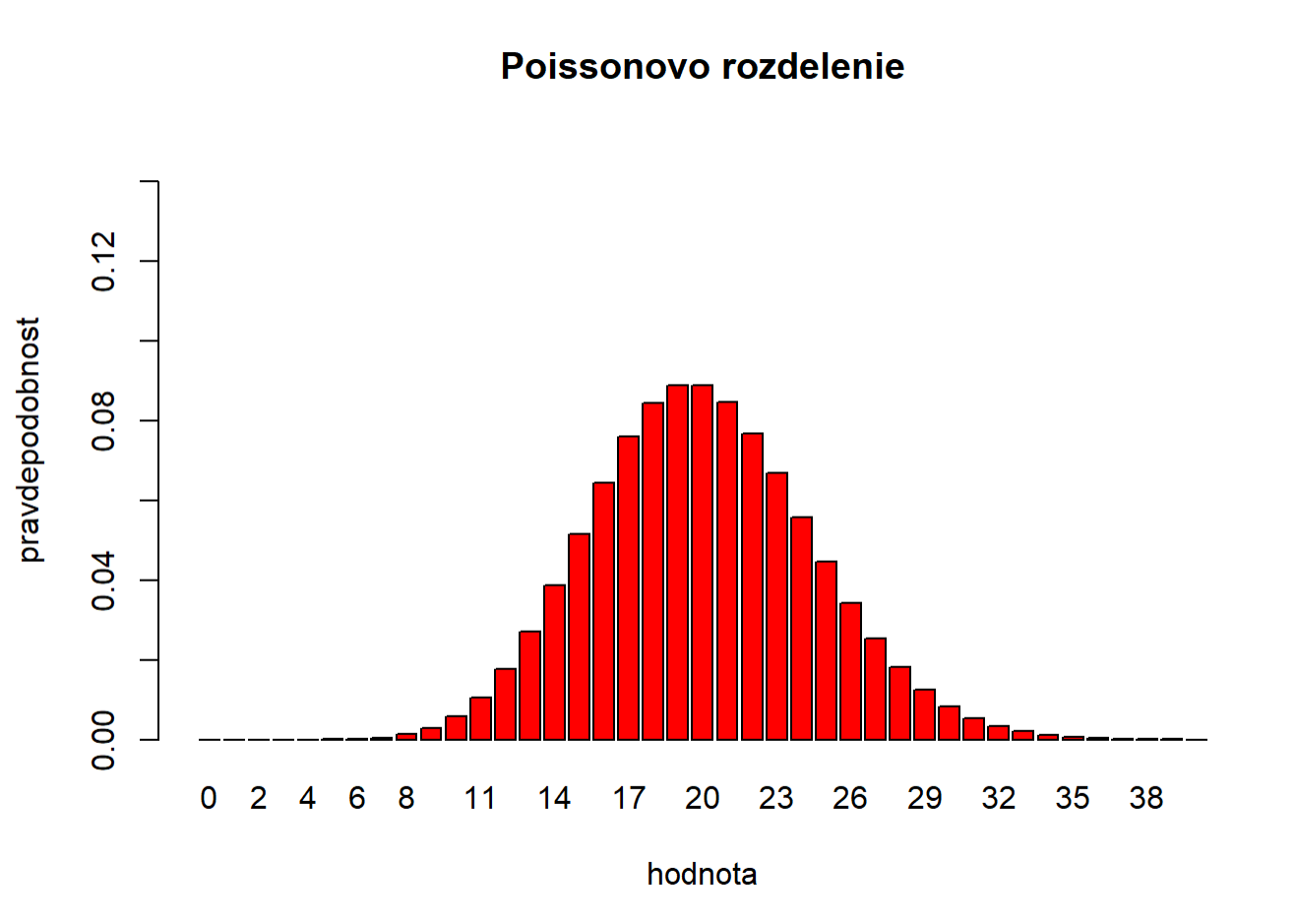
```
# Uvazujme casovy okamih 1 hodinu, urcte maximalny pocet ludi, ktori navstivia
# linku s pravdepodobnostou 90% (na linku pride max. tolko ludi) .. zarucujeme
# sa, ze na 90% tolko pride max
qpois(0.9, 20)
```

```
## [1] 26
```

```
# zostrojte tabulku a graf rozdelenia pravdepodobnosti
# pre interval 1 hodina a prvych 40 hodnot
xP <- 0:40 # viem zodpovedat ani jednu az vsetkych 5 otazok
hustotaP <- dpois(xP, 20)
tabulkaP <- data.frame(hodnota = xP, pravdepodobnost = hustotaP)
tabulkaP
```

```
##    hodnota pravdepodobnost
## 1        0    2.061154e-09
## 2        1    4.122307e-08
## 3        2    4.122307e-07
## 4        3    2.748205e-06
## 5        4    1.374102e-05
## 6        5    5.496410e-05
## 7        6    1.832137e-04
## 8        7    5.234676e-04
## 9        8    1.308669e-03
## 10       9    2.908153e-03
## 11      10    5.816307e-03
## 12      11    1.057510e-02
## 13      12    1.762517e-02
## 14      13    2.711565e-02
## 15      14    3.873664e-02
## 16      15    5.164885e-02
## 17      16    6.456107e-02
## 18      17    7.595420e-02
## 19      18    8.439355e-02
## 20      19    8.883532e-02
## 21      20    8.883532e-02
## 22      21    8.460506e-02
## 23      22    7.691369e-02
## 24      23    6.688147e-02
## 25      24    5.573456e-02
## 26      25    4.458765e-02
## 27      26    3.429819e-02
## 28      27    2.540607e-02
## 29      28    1.814719e-02
## 30      29    1.251530e-02
## 31      30    8.343536e-03
## 32      31    5.382927e-03
## 33      32    3.364329e-03
## 34      33    2.038987e-03
## 35      34    1.199404e-03
## 36      35    6.853739e-04
## 37      36    3.807633e-04
## 38      37    2.058180e-04
## 39      38    1.083253e-04
## 40      39    5.555141e-05
## 41      40    2.777571e-05
```

```
View(tabulkaP)
barplot(
  tabulkaP$pravdepodobnost,
  main = 'Poissonovo rozdelenie',
  names.arg = xP,
  xlab = 'hodnota',
  ylab = 'pravdepodobnost',
  col = 'red',
  ylim = c(0,0.15)
)
```




```
################################
# spojite rozdelenia
# normalne rozdelenie pravdepodobnosti ma dva parametre, strednu hodnotu
# mu (to je asi nejakej pismenko) a smerodajnu odchylku sigma
# prikaz je norm
# Zivotnost bateriek do mobilnych telefonov sa riadi normalnym rozdelenim
# so strednou hodnotou 8 a smerodajnou odchylkou 2
# Ulohy
# Kolko % bateriek treba vymenit do 7.5 roka P(x <= 7.5)
# Kolko % bateriek vydrzi v rozpati 7-9 rokov P(7 <= x <= 9)
# Kolko vydrzi viac ako 10 P(x >= 10)
# Za aku dobu zivotnosi sa mozno zarucit na 90% (tolko a viac)

# prva uloha
pnorm(7.5, mean = 8, sd = 2)
```

```
## [1] 0.4012937
```

```
# druha uloha
# na dva kroky, najskor spocitam po a potom odcitam
pnorm(9, mean = 8, sd = 2) - pnorm(7, mean = 8, sd = 2)
```

```
## [1] 0.3829249
```

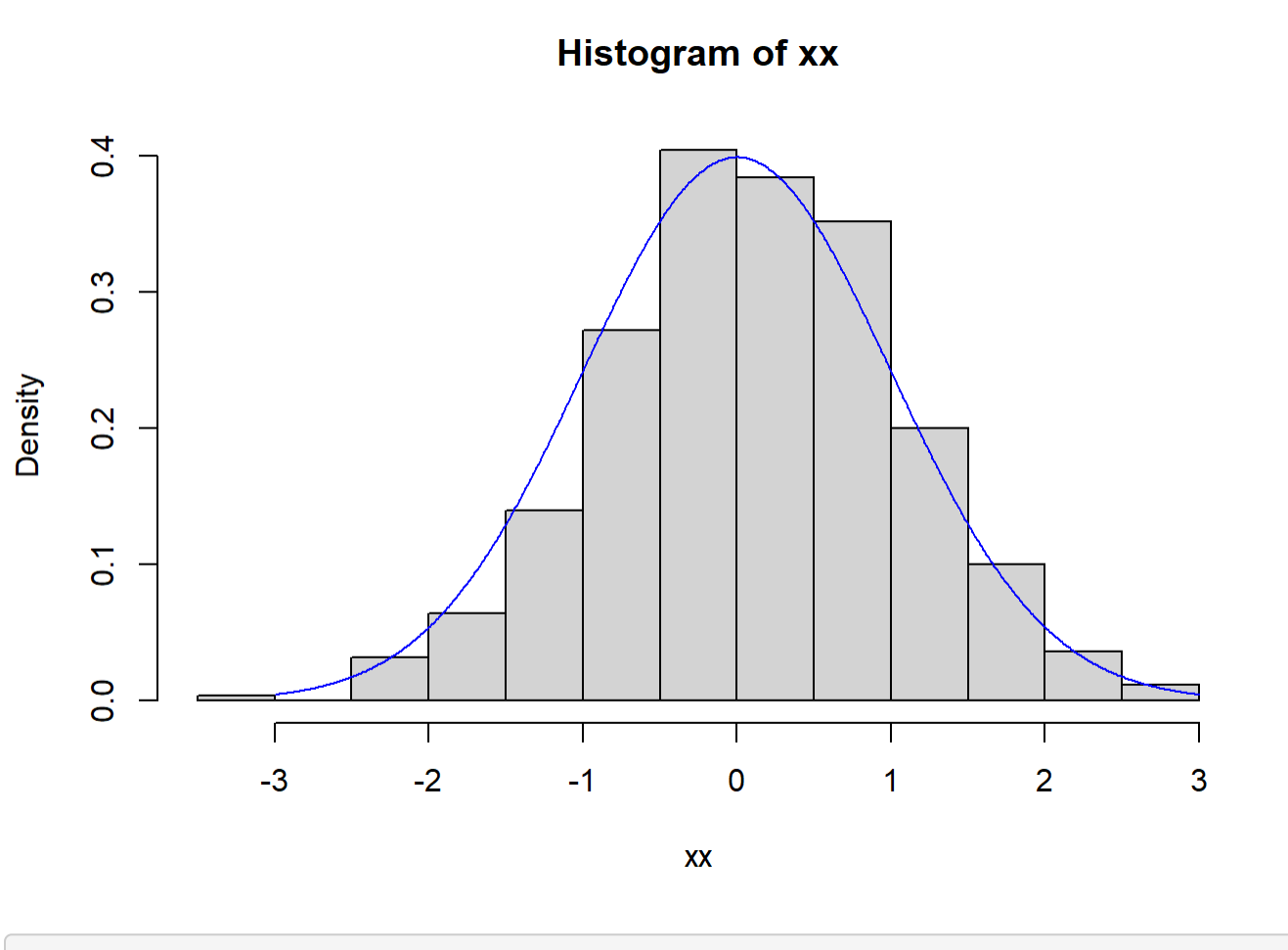
```
# tretia uloha
pnorm(10, mean = 8, sd = 2, lower.tail = F)
```

```
## [1] 0.1586553
```

```
# stvrta uloha
qnorm(0.9, mean = 8, sd = 2, lower.tail = F) #zarucujem sa na tolko a hornu hranicu
```

```
## [1] 5.436897
```

```
# nakreslime histogram nasimulovanych dat, N(0, 1), prelozime hustotu cez histogram
xx <- rnorm(500, mean = 0, sd = 1) # pre histogram
xxx <- seq(-3, 3, 0.01) # pre kreslenie hustoty
hist(xx, freq = F) # freq F na zmenu mierky na pravdepodobnostnu
lines(xxx, dnorm(xxx, 0, 1), col = 'blue')
```




```
#########################################
# Exponencialne rozdelenie
# Zivotnost zariadenia, doba do prvej poruchy, doby medzi poruchami,
# jediny parameter lambda, je to prevratena hodnota strednej hodnoty,
# pozor ako to bude v zadani. E(x) = 1 / lambda
# Priklad
# Dlzka zivotnosti pouzivaneho PC v pocitacovej ucebni je 2 roky
# Vypocitajte PC ma zivotnost aspon 1 rok P(X >= 1)
# PC ma zivotnost najviac % rokov P(X <= 5)
# Za aku zivotnost by ste sa zarucili s pravdepodobnostou 5%, p = 0.05
# parameter je 1/2
# Uloha 1
pexp(1, rate = 1/2, lower.tail = F)
```

```
## [1] 0.6065307
```

```
# Uloha 2
pexp(5, rate = 1/2)
```

```
## [1] 0.917915
```

```
# Uloha 3
qexp(0.05, rate = 1/2, lower.tail = F)
```

```
## [1] 5.991465
```

```
# Nakreslite histogram, prelozte hustotu
xx <- rexp(500, rate = 1/2) # pre histogram
xxx <- seq(0, 5, 0.01) # pre kreslenie hustoty
hist(xx, freq = F) # freq F na zmenu mierky na pravdepodobnostnu
lines(xxx, dexp(xxx, rate = 1/2), col = 'purple')
```