4 Náhodné veličiny

4.1 Diskrétní náhodné veličiny

Binomické rozdělení $\mathrm{Bin}(N, p)$

- Bernoulliho pokusy $X_1, \ldots, X_N$:
  - $X_i = 1 \ldots$ událost nastala; $X_i = 0 \ldots$ událost nenastala; $i = 1, \ldots, N$.
  - $\Pr(X_i = 1) = p$
  - $\Pr(X_i = 0) = 1 - p = q$
- Binomické rozdělení:
  - $X \ldots$ počet událostí v posloupnosti N nezávislých Bernoulliho pokusů, přičemž pravděpodobnost nastání události v každém pokusu je vyjádřena parametrem p.
  - $\sum_{i=1}^{N} X_i = X \sim \mathrm{Bin}(N, p)$.
  - $\boldsymbol{\theta} = (N, p)$
  - pravděpodobnostní funkce:

$$p(x) = \binom{N}{x} p^x (1-p)^{N-x} \qquad x = 0, 1, \ldots, N; \tag{1}$$

  - vlastnosti: $\mathrm{E}[X] = Np$; $\mathrm{Var}[X] = Np(1-p)$
  - `dbinom(x, N, p)`, `pbinom(x, N, p)`

Dataset 5: Počet chlapců v rodinách s 12 dětmi

V rámci studie poměru pohlaví u lidí z roku 1889 bylo na základě záznamů z nemocnic v Sasku zaznamenáno rozdělení počtu chlapců v čtrnáctičlenných rodinách. Mezi $M = 6115$ rodinami s $N = 12$ dětmi byla pozorována početnost narozených chlapců. Údaje ze studie jsou uvedeny v tabulce 1.

Tabulka 1: Počet chlapců v 6 115 rodinách s dvanácti dětmi

n	0	1	2	3	4	5	6	7	8	9	10	11	12	$\sum$
$m_{observed}$	3	24	104	286	670	1033	1343	1112	829	478	181	45	7	6115

Příklad 4.1. Popis reálné situace pomocí binomického modelu
Zaměřte se nyní na dataset 5. Předpokládejme, že náhodná veličina X popisuje počet chlapců v rodinách s dvanácti dětmi. Nalezněte model, který by co nejvhodněji popisoval údaje uvedené v datasetu 5 a odhadněte hodnoty parametrů takového modelu.

Řešení příkladu 4.1
Nejprve se zaměřme na nalezení modelu, který co nejvýstižněji popisuje náhodnou veličinu X. Protože počet chlapců v rodině je vždy celé číslo, budeme jej popisovat pomocí diskrétní náhodné veličiny. V rámci jedné rodiny máme celkem $N = 12$ Bernoulliho pokusů X_i, $i = 1, \dots, 12$, přičemž sledovanou událostí v jednom Bernoulliho pokusu je narození chlapce. Při narození každého z dvanácti dětí tedy buď událost nastala (narodil se chlapec; $X_i = 1$, $i = 1, \dots, 12$) nebo událost nenastala (narodilo se děvče; $X_i = 0$, $i = 1, \dots, 12$). Na základě všech výše uvedených indicií budeme o náhodné veličině X předpokládat, že pochází z binomického rozdělení, tj. $X \sim (N, p)$, kde $N = 12$. Zbývá odhadnout hodnotu parametru p.

Odhad parametru p, tj. odhad pravděpodobnosti narození chlapce v jednom náhodném pokusu, spočítáme jako podíl součtu všech chlapců v rodinách s dvanácti dětmi (viz čitatel vzorce 2) ku celkovému počtu všech dětí v těchto rodinách (viz jmenovatel vzorce 2).

$$\widehat{p} = \frac{\sum_{n=0}^{N} n m_{observed}}{NM} = \frac{0 \times 3 + 1 \times 24 + \cdots + 11 \times 45 + 12 \times 7}{12 \times 6115} = \frac{38\,100}{73\,380} = 0.519215 \doteq 0.5192. \tag{2}$$

```r
M     <- 6115
N     <- 12
n     <- 0 : N
m.obs <- c(3, 24, 104, 286, 670, 1033, 1343, 1112, 829, 478, 181, 45, 7)
p     <- sum(n * m.obs) / (N * M)
round(p, 4)
```

```
[1] 0.5192
```

Interpretace výsledků: Náhodnou veličinu X popisující počet chlapců v rodinách s dvanácti dětmi modelujeme pomocí binomického modelu s parametry N a p, tj. $X \sim \mathrm{Bin}(N, p)$, kde $N = 12$ a $p = 0.5192$. Pravděpodobnost narození chlapce v rodinách s dvanácti dětmi je 51.92%.

Příklad 4.2. Porovnání pozorovaných a očekávaných početností v binomickém modelu
Na základě výše uvedené úvahy popisujeme počet chlapců v rodině s dvanácti dětmi pomocí binomického rozdělení $\mathrm{Bin}(12, 0.5192)$. Nyní ověříme, zda jsme k popisu zvolili vhodné rozdělení. Za předpokladu, že náhodná veličina X popisující počet chlapců v rodinách s dvanácti dětmi pochází z binomického rozdělení $\mathrm{Bin}(12, 0.5192)$, odhadněte očekávané početnosti chlapců v těchto rodinách a porovnejte je s pozorovanými početnostmi.

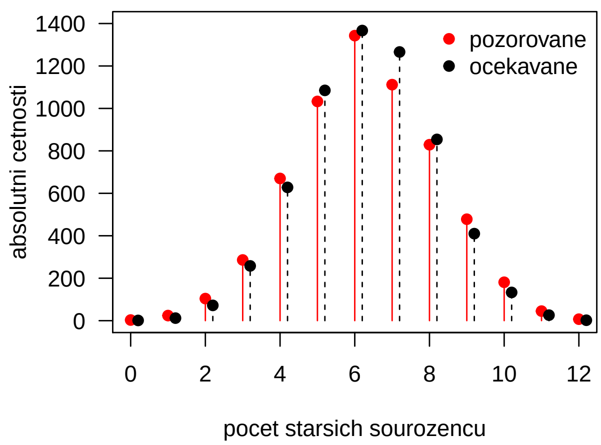
Řešení příkladu 4.2
Za předpokladu, že $X \sim \mathrm{Bin}(12, 0.5192)$ stanovíme pravděpodobnosti, že se v rodině s dvanácti dětmi nenarodí žádný chlapec, narodí právě jeden chlapec, právě dva chlapci, apod. Výsledné pravděpodobnosti vynásobíme počtem rodin, tj. číslem 6115, čímž zjistíme, v kolika rodinách se za přepokladu $X \sim \mathrm{Bin}(12, 0.5192)$ narodí nula chlapců, jeden chlapec, atd. K vypočítání těchto pravděpodobností použijeme pravděpodobnostní funkci $p(x)$, kde $x = 0, 1, \dots, 12$. Hodnoty pravděpodobnostní funkce binomického rozdělení stanovíme příkazem dbinom(), kde prvním argumentem jsou hodnoty $x = 0, 1, \dots, 12$, druhý argument size odpovídá počtu pokusů N a třetí argument prob odpovídá pravděpodobnosti p výskytu události v jednom pokusu. Vektor získaných pravděpodobností vynásobíme počtem rodin $M = 6115$ a zaokrouhlíme na nula desetinných míst (round()). Pomocí příkazů data.frame() a rbind() vytvoříme tabulku pozorovaných a očekávaných četností.

```r
p.exp       <- dbinom(0:12, size = N, prob = p)
m.exp       <- round(p.exp * 6115)
tab         <- data.frame(rbind(pozorovane = m.obs, ocekavane = m.exp))
names(tab) <- 0:12
tab
```

```
           0  1   2   3   4    5    6    7   8   9  10 11 12
pozorovane 3 24 104 286 670 1033 1343 1112 829 478 181 45  7
ocekavane  1 12  72 258 628 1085 1367 1266 854 410 133 26  2
```

Pozorované a očekávané četnosti porovnáme také graficky. Pomocí příkazu plot() s argumentem type = 'h' vykreslíme graf obsahující horizontální čáry červené barvy odpovídající pozorovaným četnostem m.obs. Příkazem points() doplníme do grafu červené body ve výšce pozorovaných četností. Příkazem lines() s argumentem type = 'h' vykreslíme přerušované horizontální čáry černé barvy odpovídající očekávaných četnostem m.exp. Příkazem points() doplníme do grafu černé body ve výšce očekávaných četností. Nakonec do grafu přidáme legendu funkcí legend().

```
plot(0 : 12, m.obs, type = 'h', col = 'red', las = 1, ylim = c(0, 1400),
     ylab = 'absolutni cetnosti', xlab = 'pocet starsich sourozencu')
points(0 : 12, m.obs, pch = 19, col = 'red')
lines (0 : 12 + 0.2, m.exp, type = 'h', lty = 2)
points(0 : 12 + 0.2, m.exp, pch = 19)
legend('topright', pch = 19, col = c('red', 'black'),
        legend = c('pozorovane', 'ocekavane'), bty = 'n')
```



Obrázek 1: Porovnání pozorovaných a očekávaných početností v binomickěm modelu Bin$(12, 0.5192)$

Interpretace výsledků: Z tabulky pozorovaných a očekávaných četností a z grafu 1 vidíme, že zvolené binomické rozdělení Bin$(12, 0.5192)$ je vhodné k popisu počtu chlapců v rodině s dvanácti dětmi.

Příklad 4.3. Výpočet pravděpodobností za předpokladu binomického modelu
Za předpokladu, že náhodná veličina X popisující počet chlapců v rodinách s dvanácti dětmi pochází z binomického rozdělení $\text{Bin}(12, 0.5192)$ vypočítejte pravděpodobnost, že v rodině s dvanácti dětmi bude (a) právě devět chlapců; (b) nejvýše čtyři chlapci; (c) alespoň osm chlapců; (d) čtyři, pět, šest, nebo sedm chlapců.

Řešení příkladu 4.3
Ze vzorce 1 víme, že pravděpodobnostní funkce binomického rozdělení má tvar

$$p(x) = \binom{N}{x} p^x (1-p)^{N-x},$$

kde v našem případě $N = 12$ a $p = 0.5192$. Začneme výpočtem pravděpodobnosti, že v rodině s dvanácti dětmi bude právě devět chlapců. K výpočtu využijeme pravděpodobnostní funkci $p(x)$, kde $x = 9$.

$$\begin{aligned}\Pr(X = 9) &= \binom{12}{9} \times 0.5192^9 \times (1 - 0.5192)^{12-9} = 220 \times 0.5192^9 \times 0.4808^3 \\ &= 0.06703911 \doteq 0.0670.\end{aligned}$$

Kontrolu ručního výpočtu můžeme provést použitím softwaru R. Hodnotu pravděpodobnostní funkce v $x = 9$ získáme příkazem dbinom(), kde prvním argumentem bude hodnota $x = 9$, druhým argumentem počet pokusů N a třetím argumentem pravděpodobnost narození chlapce p. Výsledek zaokrouhlíme na čtyři desetinná místa.

```r
N <- 12
p <- 0.5192
round(dbinom(9, N, p), 4)
```

```
[1] 0.067
```

K výpočtu pravděpodobnosti, že v rodině s dvanácti dětmi budou nejvýše čtyři chlapci využijeme distribuční funkci $F(x)$, kde $x = 4$.

$$\begin{aligned}\Pr(X \le 4) &= \sum_{i=0}^{4} \Pr(X = i) = \Pr(X = 0) + \Pr(X = 1) + \Pr(X = 2) + \Pr(X = 3) + \Pr(X = 4) \\ &= \binom{12}{0} \times 0.5192^0 \times (1 - 0.5192)^{12-0} + \cdots + \binom{12}{4} \times 0.5192^4 \times (1 - 0.5192)^{12-4} \\ &= 1 \times 0.5192^0 \times 0.4808^{12} + \cdots + 495 \times 0.5192^4 \times 0.4808^8 \\ &= 0.1588736 \doteq 0.1589.\end{aligned}$$

Hodnotu distribuční funkce v $x = 4$ získáme pomocí příkazu pbinom(), kde pvním argumentem bude hodnota $x = 4$, druhým argumentem počet pokusů N a třetím argumentem pravděpodobnost narození chlapce p.

```r
round(pbinom(4, N, p), 4)
```

```
[1] 0.1589
```

K výpočtu pravděpodobnosti, že v rodině s dvanácti dětmi bude alespoň osm chlapců využijeme vlastnost komplementarity, tj. $\Pr(X \ge x) = 1 - \Pr(X < x)$ v kombinaci s distribuční funkcí $F(x) = \Pr(X \le x)$.

$$\begin{aligned}\Pr(X \ge 8) &= 1 - \Pr(X < 8) = 1 - \Pr(X \le 7) \\ &= 1 - \left(\sum_{i=0}^{7} \Pr(X = i)\right) = 1 - (\Pr(X = 0) + \cdots + \Pr(X = 7)) \\ &= 1 - \left(\binom{12}{0} \times 0.5192^0 \times (1 - 0.5192)^{12-0} + \cdots + \binom{12}{7} \times 0.5192^7 \times (1 - 0.5192)^{12-7}\right) \\ &= 1 - \left(1 \times 0.5192^0 \times 0.4808^{12} + \cdots + 792 \times 0.5192^7 \times 0.4808^5\right) \\ &= 1 - 0.7669131 = 0.2330869 \doteq 0.2331.\end{aligned}$$

K výpočtu pravděpodobnosti $\Pr(X \ge 8)$ vyžijeme příkaz 1-pbinom().

```r
round(1 - pbinom(7, N, p), 4)
```

```
[1] 0.2331
```

Pravděpodobnost, že v rodině s dvanácti dětmi bude čtyři pět, šest nebo sedm chlapců vypočítáme pomocí pravděpodobnostní funkce $p(x)$.

$$
\begin{aligned}
\Pr(4 \leq X \leq 7) &= \sum_{i=4}^{7} \Pr(X = i) = \Pr(X = 4) + \Pr(X = 5) + \Pr(X = 6) + \Pr(X = 7) \\
&= \binom{12}{4} \times 0.5192^4 \times (1 - 0.5192)^{12-4} + \cdots + \binom{12}{7} \times 0.5192^7 \times (1 - 0.5192)^{12-7} \\
&= 495 \times 0.5192^4 \times 0.4808^8 + \cdots + 792 \times 0.5192^7 \times 0.4808^5 \\
&= 0.7107605 \doteq 0.7108.
\end{aligned} \tag{3}
$$

K výpočtu pravděpodobnosti $\Pr(4 \leq X \leq 7)$ můžeme tentokrát použít buď funkci dbinom(), která vede na výpočet analogický výpočtu 3, tj.

```r
round(sum(dbinom(4 : 7, N, p)), 4)
```

```
[1] 0.7108
```

nebo využít vztahu

$$\Pr(4 \leq X \leq 7) = \Pr(X \leq 7) - \Pr(X < 4) = \Pr(X \leq 7) - \Pr(X \leq 3)$$

a vypočítat $\Pr(4 \leq X \leq 7)$ pomocí příkazu

```r
round(pbinom(7, N, p) - pbinom(3, N, p), 4)
```

```
[1] 0.7108
```

Vidíme, že oba postupy vedou ke stejnému výsledku.

Interpretace výsledků: Pravděpodobnost, že v rodině s dvanácti dětmi bude právě devět chlapců, je 6.70%. Pravděpodobnost, že v rodině s dvanácti dětmi budou nejvýše čtyři chlapci, je 15.89%. Pravděpodobnost, že v rodině s dvanácti dětmi bude alespoň osm chlapců, je 23.31%. Pravděpodobnost, že v rodině s dvanácti dětmi bude čtyři, pět, šest, nebo sedm chlapců, je 71.08%.

Příklad 4.4. Graf pravděpodobnostní a distribuční funkce binomického modelu
Zaměřte se nyní blíže na tvar binomického rozdělení $\mathrm{Bin}(12, 0.5192)$. Nakreslete graf pravděpodobnostní funkce $p(x)$ a graf distribuční funkce $F(x)$ tohoto rozdělení.

Řešení příkladu 4.4
Začneme s vykreslením grafu pravděpodobnostní funkce $p(x)$. Do proměnné px nejprve vložíme hodnoty pravděpodobnostní funkce binomického rozdělení $p(x) = \Pr(X = x)$ pro $x = 0, 1, \ldots, 12$, které vypočítáme příkazem dbinom(). Na prvním místě ve funkci budou hodnoty x, na druhém místě počet pokusů N a na třetím místě pravděpodobnost nastání události p v jednom pokusu. Samotný graf potom vykreslíme příkazem plot() s argumentem type = 'h', který zajistí vykreslení vertikálních čar v hodnotách 0–12 na ose x a s délkou odpovídající hodnotám pravděpodobnostní funkce $p(x)$, $x = 0, \ldots, 12$. Argumentem xlab = '' v rámci příkazu plot() zamezíme vypsání popisku osy x. Dále do grafu doplníme body (points()) ve výšce hodnot funkce $p(x)$. Popisek osy x vykreslíme samostatně příkazem mtext() na pozici pod graf (side = 1) na řádek 2.1 (line). Nakonec do grafu doplníme druhý popisek uvádějící hodnoty parametrů N a p. Text popisku vygenerujeme pomocí kombinace funkcí bquote() a paste(). Uvnitř funkce paste() je vložena syntaxe popisku skládající se ze tří částí oddělených čárkami. První část, tj. N==.(N), vypíše písmeno N, znaménko rovnosti a vyhodnocení proměnné N, tj. 12, druhá část, tj ';', vypíše středník a třetí část p==.(p) vypíše písmeno p, znaménko rovnosti a vyhodnocení proměnné p, tj. 0.5192.

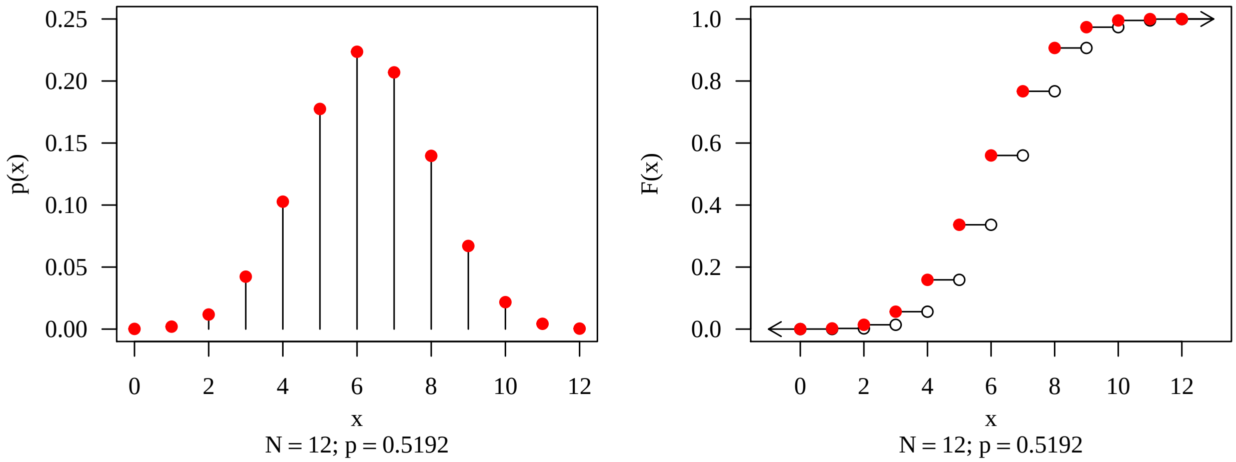
```r
N     <- 12
x     <- 0 : N
p     <- 0.5192

# Graf pravdepodobnostni funkce
px    <- dbinom(x, N, p)
plot(x, px, type = 'h', ylim = c(0, 0.25), ylab = 'p(x)', xlab = '', las = 1)
points(x, px, col = 'red', pch = 19)
mtext('x', side = 1, line = 2.1)
mtext(bquote(paste(N==.(N), '; ', p==.(p))), side = 1, line = 3.2)
```

První krokem pro vykreslení grafu distribuční funkce $F(x)$ je vytvoření vektoru Fx s hodnotami distribuční funkce binomického rozdělení $F(x) = \Pr(X \leq x)$ pro $x = 0, 1, \ldots, 12$, které vypočítáme příkazem pbinom(). Prvním argumentem funkce budou hodnoty x, druhým počet pokusů N a třetím pravděpodobnost nastání události p v jednom pokusu. Konstrukci grafu zahájíme vykreslením prázdného grafu (příkaz plot() s argumentem type = 'n') s rozsahem měřítka osy x od -1 do $N + 1$ (xlim) a rozsahem měřítka osy y od 0 do 1 (ylim). Argumentem xlab = '' potlačíme vypsání popisku osy x. Nyní do grafu dokreslíme horizontální úsečky délky 1 začínající vždy v bodě [x, F(x)] a končící v bodě [x+1, F(x)]. Příkazem arrows() dokreslíme do grafu horozintální šipku umístěnou vlevo dole a směřující doleva, s počátečním bodem [0,0] a koncovým bodem [-1, 0]. Argumentem length zmenšíme velikost zobáčku šipky. Opětovným použitím příkazu arrows() vykreslíme nyní horizotnální šipku umístěnou vpravo nahoře, směřující doprava s počátečním bodem [N, 1] a koncovým bodem [N+1, 1]. Následně do grafu dokreslíme body značící, skok z hodnoty $F(x)$ do hodnoty $F(x + 1)$. Na levém konci každé úsečky vykreslíme bod (points()) se souřadnicí [x, F(x)] s červeným okrajem (bg) a červeným vnitřkem (col) značící, že levý krajní bod každé úsečky má hodnotu distribuční funkce $F(x)$. Oproti tomu, na pravém konci každé úsečky vykreslíme bod (points()) se souřadnicí [x, F(x-1)] s černým okrajem (bg) a bílým vnitřkem (col) značící, že pravý krajní bod každé úsečky nepatří mezi body s hodnotou distribuční funkce $F(x)$. Nakonec dvojnásobným využitím příkazu mtext() doplníme do grafu popisek osy x a popisek s hodnotami parametrů N a p.

```r
# Graf distribucni funkce
Fx    <- pbinom(x, N, p)
plot(x, Fx, type = 'n', xlab = '', ylab = 'F(x)',
     xlim = c(-1, N + 1), ylim = c(0, 1), las = 1)
segments(x, Fx, x + 1, Fx)                       # vodorovne cary
arrows(0, 0,    -1, 0, length = 0.1)             # sipka vlevo dole
arrows(N, 1, N + 1, 1, length = 0.1)             # sipka vpravo dole
points(x, c(0, Fx[1 : N]), pch = 21, bg = 'white', col = 'black') # prazdne body
points(x, Fx, col = 'red', pch = 19)             # cervene body
mtext('x', side = 1, line = 2.1)
mtext(bquote(paste(N==.(N), '; ', p==.(p))), side = 1, line = 3.2)
```

Poissonovo rozdělení $\mathrm{Po}(\lambda)$

- X ... počet událostí, které nastanou v jednotkovém časovém intervalu, přičemž k událostem dochází náhodně, jednotlivě a vzájemně nezávisle. Střední počet těchto událostí je vyjádřen parametrem $\lambda > 0$.
- $X \sim \mathrm{Po}(\lambda)$
- $\theta = \lambda$
- pravděpodobnostní funkce:
$$p(x) = \frac{\lambda^x}{x!} e^{-\lambda} \; x = 0, 1, \ldots;$$
- vlastnosti: $E[X] = \lambda$; $\mathrm{Var}[X] = \lambda$
- dpois(x, lambda), ppois(x, lambda)



Obrázek 2: Pravděpodobnostní funkce (vlevo) a distribuční funkce (vpravo) binomického modelu $\mathrm{Bin}(12, 0.5192)$

Příklad 4.5. Výpočet parametru λ Poissonova modelu
Načtete datový soubor 17-anova-newborns.txt a odstraňte z něj neznámá pozorování. Zaměřte se na znak $X =$*počet starších sourozenců* novorozence. Za předpokladu, že náhodná veličina X popisující počet starších sourozenců novorozence pochází z Poissonova rozdělení parametrem λ odhadněte střední hodnotu počtu starších sourozenců λ.

Řešení příkladu 4.5
Střední hodnotu počtu starších sourozenců odhadneme pomocí vzorce

$$\lambda = \frac{\text{počet starších sourozenců}}{\text{počet novorozenců}} = \frac{\sum_{i=1}^{N} x_i}{N}. \tag{4}$$

```
[1] 0.9428365
```
56

Interpetace výsledků: Střední hodnota počtu starších sourozenců novorozenců v datovém souboru $\lambda = 0.9428$

Příklad 4.6. Porovnání pozorovaných a očekávaných početností v Poissonově modelu
Za předpokladu, že počet starších sourozenců novorozenců pochází z Poissonova rozdělení s parametrem $\lambda = 0.9428$ odhadněte očekávané početnosti starších sourozenců a porovnejte je s pozorovanými početnostmi.

Řešení příkladu 4.6

Příklad 4.7. Výpočet pravděpodobností za předpokladu Poissonova modelu
Vraťem se nyní k příkladu 4.5. Za předpokladu, že data pochází z Poissonova rozdělení s parametrem $\lambda = 0.9428$ určete pravděpodobnost, novorozenec má (a) dva, tři nebo čtyři starší sourozence; (b) alespoň čtyři starší sourozence; (c) nejvýše dva starší sourozence; (d) právě jednoho starší sourozence.

Řešení příkladu 4.7

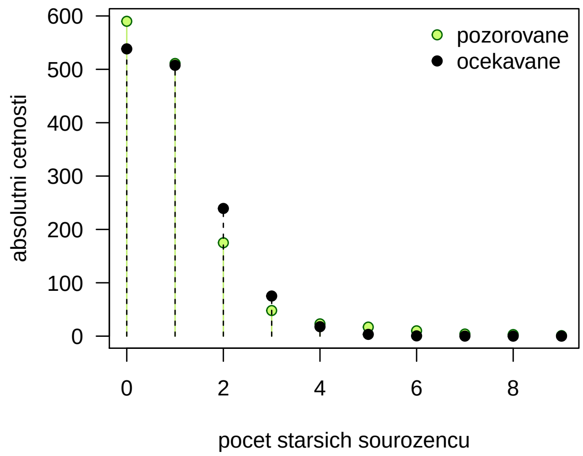
```
[1] 0.2403672
```
57

```
[1] 0.01568161
```
58

```
[1] 0.9299071
```
59

```
[1] 0.367255
```
60

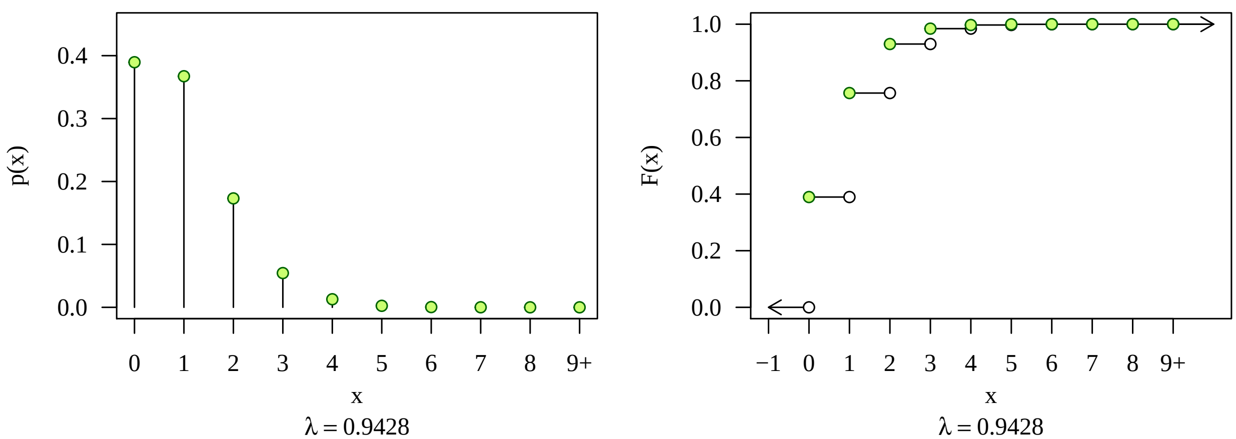
Interpretace výsledů: Pravděpodobnost, že novorozenec bude mít dva, tři nebo čtyři starší sourozence je 24.04%. Pravděpodobnost, že novorozenec bude alespoň čtyři starší sourozence je 1.57%. Pravděpodobnost, novorozenec bude mít nejvýše dva starší sourozence je 92.99%. Pravděpodobnost, že novorozenec bude mít jednoho staršího sourozence je 36.73%.



Obrázek 3: Porovnání pozorovaných a očekávaných početností v Poissonově modelu

Příklad 4.8. Graf pravděpodobnostní a distribuční funkce Poissonova modelu
Nakreslete graf pravděpodobnostní a distribuční funkce Poissonova rozdělení Po(0.9428) v hodnotách $x = 0, 1, 2, 3, 4, 5, 6, 7, 8$, a $x \geq 9$.

Řešení příkladu 4.8



Obrázek 4: Pravděpodobnostní a distribuční funkce Poissonova modelu

Dataset 6: Pruské armádní jednotky

V rámci studie z roku 1898 byly zpracovávány počty smrtelných úrazů v pruských armádních jednotkách způsobené kopnutím koněm. Údaje o smrtelných úrazech po kopnutí koněm by zaznamenávány po dobu dvaceti let u deseti armádních jednotek. Počty úrazů v každé jednotce za jeden rok jsou uvedeny v následující tabulce.

n	0	1	2	3	4	5+	$\sum$
$m_{observed}$	109	65	22	3	1	0	200

Rozsah náhodného výběru je $M = 200$ (10 jednotek $\times$ 20 let).

Příklad 4.9. Výpočet parametru λ Poissonova modelu
Vezměte údaje z datasetu 7. Předpokládejme, že náhodná veličina X popisující počet smrtelných úrazů v pruských armádních jednotkách způsobených kopnutím koněm pochází z Poissonova rozdělení se střední hodnotou λ. Odhadněte střední hodnotu počtu smrtelných úrazů λ.

Řešení příkladu 4.9
Střední hodnotu počtu smrtelných úrazů v pruských armádních jednotkách způsobeným kopnutím koněm odhadneme pomocí vzorce

$$\lambda = \frac{\sum_{n=0}^{N} n m_{observed}}{\sum_{n=0}^{N} m_{observed}}. \tag{5}$$

```
[1] 0.61
```
61

Interpretace výsledků: Střední hodnota výskytu smrtelných úrazů způsobených kopnutím koněm je $\lambda = 0.61$.

Příklad 4.10. Výpočet pravděpodobností za předpokladu Poissonova modelu
Vraťem se nyní k příkladu 4.9. Za předpokladu, že data pochází z Poissonova rozdělení s parametrem $\lambda = 0.61$ určete pravděpodobnost, že v pruských armádních jednotkách dojde k (a) nejvýše dvěma smrtelným úrazům; (b) žádnému smrtelnému úrazu; (c) alespoň jednomu smrtelnému úrazu; (d) právě jednomu smrtelnému úrazu.

Řešení příkladu 4.10

```
[1] 0.9758853
```
62

```
[1] 0.5433509
```
63

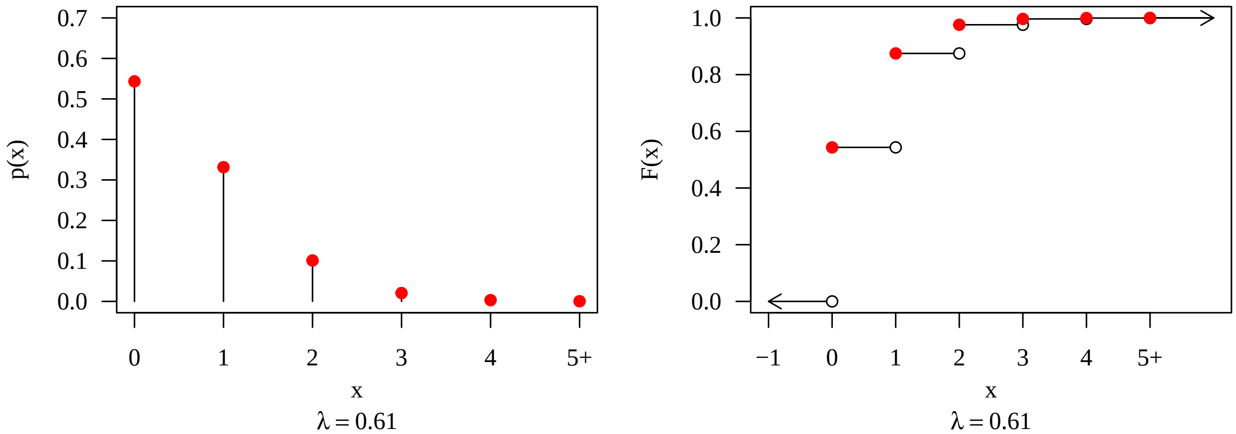
```
[1] 0.1252051
```
64

```
[1] 0.331444
```
65

Interpretace výsledů: Pravděpodobnost, že v pruských armádních jednotkách dojde k nejvýše dvěma smrtelným úrazům způsobených kopnutím koněm je 97.59%. Pravděpodobnost, že v pruských armádních jednotkách nedojde k žádnému smrtelnému úrazu způsobenému kopnutím koněm je 54.34%. Pravděpodobnost, že v pruských armádních jednotkách dojde k alespoň jednomu smrtelnému úrazu způsobenému kopnutím koněm je 12.52%. Pravděpodobnost, že v pruských armádních jednotkách dojde k právě jednomu smrtelnému úrazu způsobenému kopnutím koněm je 33.14%.

Příklad 4.11. Graf pravděpodobnostní a distribuční funkce Poissonova modelu
Nakreslete graf pravděpodobnostní a distribuční funkce Poissonova rozdělení Po(0.61) v hodnotách $x = 0, 1, 2, 3, 4$ a $x \geq 5$.

Řešení příkladu 4.11



Obrázek 5: Pravděpodobnostní a distribuční funkce Poissonova modelu

n	0	1	2	3	4	≥ 5	$\sum$
$m_{observed}$	447	132	42	21	3	2	647

Dataset 7: Dělníci v továrně

V rámci studie počtu úrazů v továrnách byl zaznamenán počet úrazů u každého dělníka v jedné vybrané továrně během roku 1920. Celkový počet dělníků zahrnutých do studie $M = 647$. Údaje ze studie jsou uvedeny v následující tabulce.

Příklad 4.12. Výpočet parametru λ Poissonova modelu
Vezměte údaje z datasetu 7. Předpokládejme, že náhodná veličina X popisující počet úrazů u dělníků v továrně pochází z Poissonova rozdělení se střední hodnotou λ. Odhadněte střední hodnotu počtu úrazů u dělníků v továrně λ.

Řešení příkladu 4.12
Střední hodnotu počtu úrazů dělníků v továrně za jeden rok odhadneme pomocí vzorce

$$\widehat{\lambda} = \frac{\sum_{n=0}^{N} n m_{observed}}{\sum_{n=0}^{N} m_{observed}}. \tag{6}$$

```
[1] 0.4652
```
66

Interpretace výsledků: Střední hodnota počtu úrazů u dělníků v továrně během jednoho roku je $\lambda = 0.4652$.

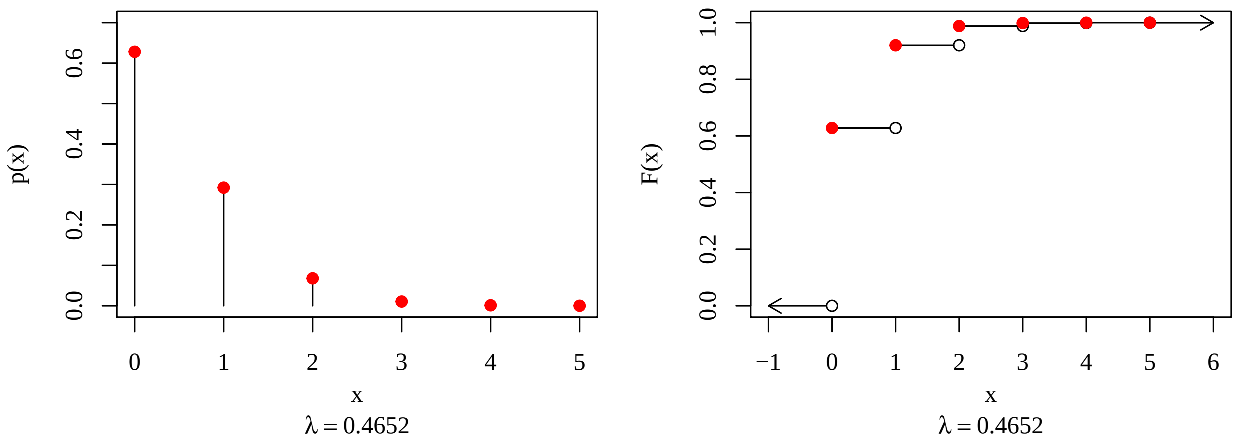
Příklad 4.13. Graf pravděpodobnostní a distribuční funkce Poissonova modelu
Nakreslete graf pravděpodobnostní a distribuční funkce Poissonova rozdělení $\text{Po}(0.4652)$ v hodnotách $x = 0, 1, 2, 3, 4$ a $x \geq 5$.

Řešení příkladu 4.13

Příklad 4.14. Výpočet pravděpodobností na základě Poissonova modelu
Za předpokladu, že náhodná veličina X, udávající počet úrazů u dělníků v továrně, pochází z Poissonova rozdělení s parametrem $\lambda = 0.4652$, tj. $X \sim \text{Po}(\lambda)$ vypočítejte pravděpodobnost, že u náhodně vybraného dělníka dojde během jednoho roku k (a) nula úrazům; (b) třem nebo čtyřem úrazům; (c) nejvýše dvěma úrazům; (d) alespoň jednomu úrazu.

Řešení příkladu 4.14

```
[1] 0.628
```
67



Obrázek 6: Pravděpodobnostní a distribuční funkce Poissonova modelu

```
[1] 0.0118                                                    68
```

```
[1] 0.9881                                                    69
```

```
[1] 0.372                                                     70
```

interpretace výsledků: Pravděpodobnost, že u vybraného dělníka nedojde během roku k žádnému úrazu, je 0.6280. Pravděpodobnost, že u vybraného dělníka dojde během roku k třem nebo čtyřem úrazům, je 0.0118. Pravděpodobnost, že u vybraného dělníka dojde během roku k nejvýše dvěma úrazům, je 0.9881. Pravděpodobnost, že u vybraného dělníka dojde během roku k alespoň jednomu úrazu, je 0.3720.

4.2 Spojité rozdělení

Normální rozdělení $N(\mu, \sigma^2)$

- $X_1, \ldots, X_n$ . . . nezávislé náhodné veličiny
- Normální rozdělení
  - $X \sim N(\mu, \sigma^2)$
  - $\boldsymbol{\theta} = (\mu, \sigma^2)^T$
  - hustota
    $$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\}, \qquad x \in \mathbb{R}.$$
  - vlastnosti $\mathrm{E}[X] = \mu$; $\mathrm{Var}[X] = \sigma^2$
  - dnorm(x, mu, sigma), pnorm(x, mu, sigma), rnorm(M, mu, sigma), qnorm(alpha, mu, sigma)
- Standardizované normální rozdělení
  - $X \sim N(0, 1)$
  - $\boldsymbol{\theta} = (0, 1)^T$
  - hustota
    $$f(x) = \phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}, \qquad x \in \mathbb{R}.$$
  - vlastnosti $\mathrm{E}[X] = 0$; $\mathrm{Var}[X] = 1$

– dnorm(x), pnorm(x), rnorm(M), qnorm(alpha)

- Vlastnosti normálního rozdělení
  - **Věta 1:** Nechť $X_1, \ldots, X_n$ jsou nezávislé náhodné veličiny z normálního rozdělení $N(\mu, \sigma^2)$. Potom náhodná veličina $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.

Příklad 4.15. Výpočet pravděpodobností na základě normálního modelu
Na základě datového souboru obsahujícího osteometrická data klíční kosti (clavicula) angického souboru dokumentovaných skeletů (Parsons, 1916) byla odhadnuta střední hodnota a směrodatná odchylka délky pravé klavikuly u mužů. Střední hodnota $\mu = 151.74$ mm, směrodatná ochylka $s = 11$ mm (viz příklad **??**. Za předpokladu, že data pochází z normálního rozdělení vypočítejte, jaká je pravděpodobnost, že délka pravé klavikuly u mužů bude (a) rovná 150 mm; (b) menší než 140 mm; (c) větší než 160 mm; (d) v rozmezí 140-160 mm.

Řešení příkladu 4.15

```
[1] 0
```
71

```
[1] 0.1429243
```
72

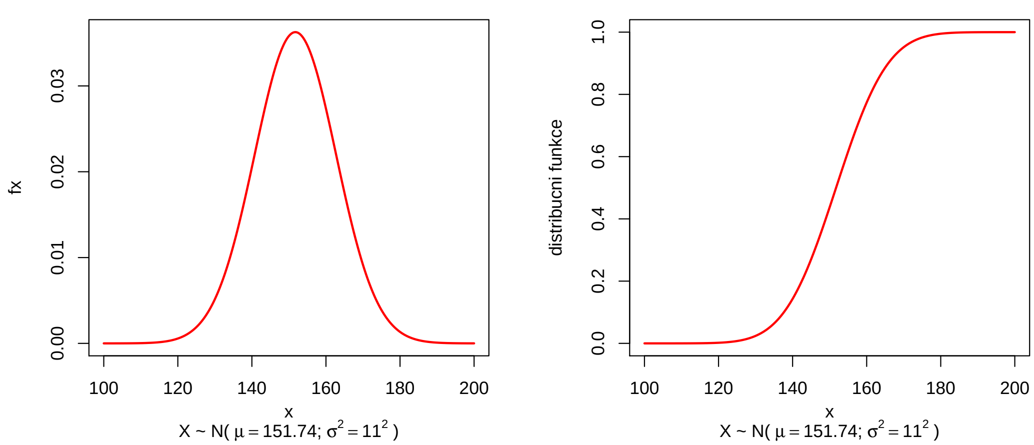
```
[1] 0.2263537
```
73

```
[1] 0.630722
```
74

Interpretace výsledků: Pravděpodobnost, že délka pravé klavikuly u mužů bude rovná 150 mm je 0%, protože data pochází z normálního rozdělení, což je spojitý typ rozdělení a proto $\Pr(X = 150) = 0$. Pravděpodobnost, že délka pravé klavikuly u mužů bude menší než 140 mm je 14.29%. Pravděpodobnost, že délka pravé klavikuly u mužů bude větší než 160 mm je 12.34%. Pravděpodobnost, že délka pravé klavikuly u mužů bude v rozmezí 140–160 mm je 22.64%.

Příklad 4.16. Graf hustoty a distribuční funkce normálního modelu
Vraťme se k příkladu 4.15. Nakreslete graf hustoty a distribuční funkce náhodné veličiny $X \sim N(151.74, 11)$.

Řešení příkladu 4.16



Příklad 4.17. Výpočet pravděpodobností na základě normálního modelu
Na základě datového souboru obsahujícího údaje o porodní hmmotnosti novorozenců v jedné okresní nemocnici za období jednoho roku(Alánová, 2008) byla odhadnuta střední hodnota a směrodatná odchylka porodní hmotnosti novorozenců. Střední hodnota $\mu = 3078.94$ g, směrodatná ochylka $s = 697$ g (viz příklad **??**). Za předpokladu, že data pochází z normálního rozdělení vypočítejte, jaká je pravděpodobnost, že porodní hmotnost novorozence bude

(a) menší než 3800 g; (b) v rozmezí 2500–4200 g; (c) větší než 4000 g; (d) rovná 2100 g.

Řešení příkladu 4.17

```
[1] 0.8495533
```
75

```
[1] 0.743032
```
76

```
[1] 0.09317345
```
77

```
[1] 0
```
78

Interpretace výsledků: Pravděpodobnost, že porodní hmotnost novorozenců bude menší než 3800 g je 84.96%. Pravděpodobnost, že porodní hmotnost novorozenců bude v rozmezí 2500–4200 mm je 74.30%. Pravděpodobnost, že porodní hmotnost novorozenců bude větší než 4000 mm je 9.32%. Pravděpodobnost, že porodní hmotnost novorozenců bude rovná 2100 g je 0%, protože data pochází z normálního rozdělení, což je spojitý typ rozdělení a proto $\Pr(X = 2100) = 0$.

Příklad 4.18. Výpočet pravděpodobností na základě standardizovaného normálního modelu
Vraťme se nyní k předchozímu příkladu 4.17. Za předpokladu, že porodní hmotnost novorozenců pochází z normálního rozdělení $N(3078.94, 697^2)$ vypočítejte pravděpodobnost, že porodní hmotnost novorozence bude (a) menší než 3800 g; (b) v rozmezí 2000–3000 g; (c) větší než 4000 g, (d) rovná 2100 g. Řešení proveďte přes standardizaci náhodné veličiny X.

Řešení příkladu 4.18

```
[1] 0.8495533
```
79

```
[1] 0.743032
```
80

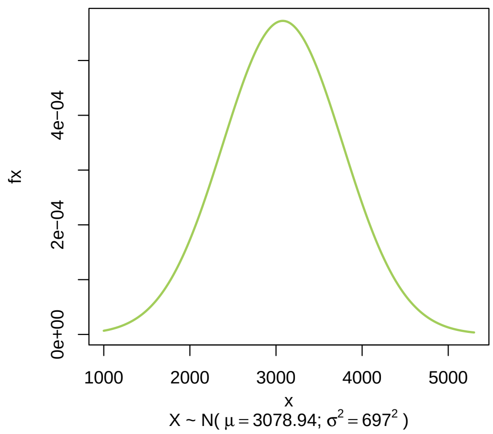
```
[1] 0.09317345
```
81

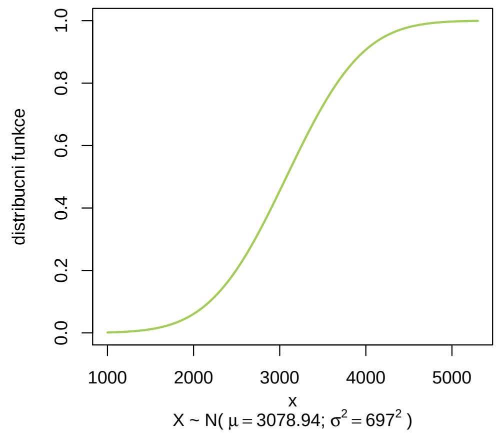
```
[1] 0
```
82

Interpretace výsledků: Pravděpodobnost, že porodní hmotnost novorozenců bude menší než 3800 g je 84.96%. Pravděpodobnost, že porodní hmotnost novorozenců bude v rozmezí 2500–4200 mm je 74.30%. Pravděpodobnost, že porodní hmotnost novorozenců bude větší než 4000 mm je 9.32%. Pravděpodobnost, že porodní hmotnost novorozenců bude rovná 2100 g je 0%, protože data pochází z normálního rozdělení, což je spojitý typ rozdělení a proto $\Pr(X = 2100) = 0$.

Příklad 4.19. Graf hustoty a distribuční funkce normálního modelu
Vraťme se k příkladu 4.17. Nakreslete graf hustoty a distribuční funkce náhodné veličiny $X \sim N(3078.94, 697)$.

Řešení příkladu 4.19


fx
0e+00
2e−04
4e−04
1000
2000
3000
4000
5000
x
X ~ N(μ = 3078.94; σ² = 697²)


distribucni funkce
0.0
0.2
0.4
0.6
0.8
1.0
1000
2000
3000
4000
5000
x
X ~ N(μ = 3078.94; σ² = 697²)

4.3 Aproximace binomického modelu normálním modelem

- Normální rozdělení je limitním rozdělením binomického rozdělení $\mathrm{Bin}(N, p)$, tedy pro $N \to \infty$, $p \to 0.5$:
$$X \sim \mathrm{Bin}(N, p) \to X \sim N(\mu, \sigma^2),$$
kde $\mu = Np$ a $\sigma^2 = Np(1-p)$.
- Haldova podmínka: Nechť $X \sim \mathrm{Bin}(N, p)$ a platí, že $Np > 5$ a $N(1-p) > 5$. Potom rozdělení náhodné proměnné X můžeme aproximovat normálním rozdělením $X \sim N(Np, Np(1-p))$.

Příklad 4.20. Aproximace binomického modelu normálním modelem
Předpokládejme, že pravděpodobnost výskytu dermatoglyfického vzoru *vír* na palci pravé ruky u mužů české populace $p = 0.533$.

1. Jaká je pravděpodobnost, že ve vybraném vzorku 10 mužů bude výskyt dermatoglyfického vzoru *vír* na palci pravé ruky (a) alespoň u sedmi mužů; (b) nejvýše u pěti mužů; (c) u šesti, sedmi nebo osmi mužů.
2. Jaká je pravděpodobnost, že ve vybraném vzorku 100 mužů bude výskyt dermatoglyfického vzoru *vír* na palci pravé ruky (a) alespoň u 56; (b) nejvýše u 53 mužů; (c) u 60–85 mužů.
3. Jaká je pravděpodobnost, že ve vybraném vzorku 300 mužů bude výskyt dermatoglyfického vzoru *vír* na palci pravé ruky (a) alespoň u 164 mužů; (b) nejvýše u 160 mužů; (c) u 170–175.

Požadované pravděpodobnosti vypočítejte exaktně na základě binomického rozdělení a aproximačně na základě normálního rozdělení. Výsledné hodnoty navzájem porovnejte.

Řešení příkladu 4.20

```
          alespon 7 nejvyse 5    8-9
binomicke    0.2313    0.5396 0.0801
normalni     0.3355    0.4172 0.1349
```

```
          alespon 56 nejvyse 53  60-85
binomicke     0.3304     0.5151 0.1067
normalni      0.3666     0.4760 0.1266
```

```
          alespon 164 nejvyse 160 170-175
binomicke      0.3389      0.5272  0.0980
normalni       0.3599      0.5046  0.1059
```

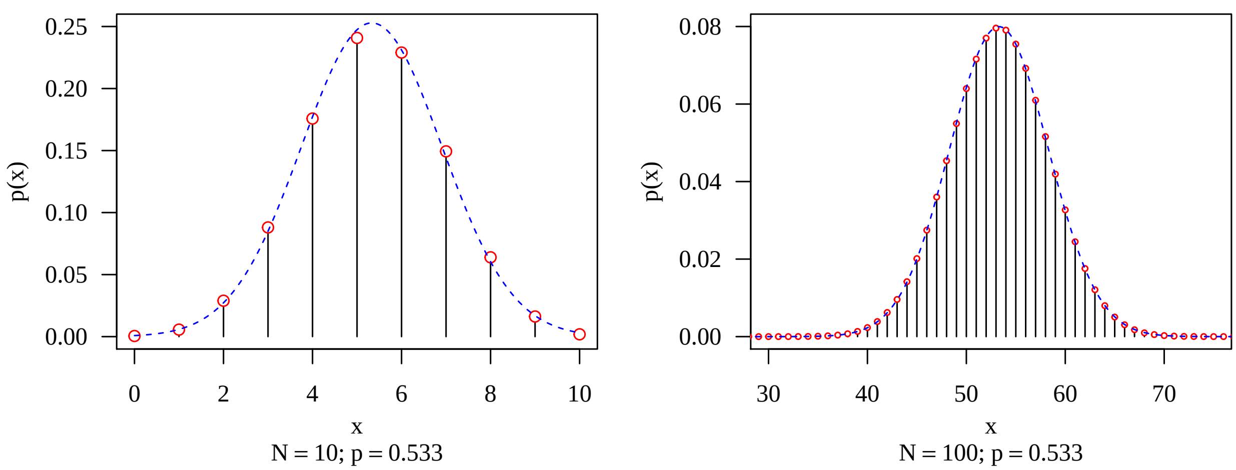
Interpretace výsledků: Pravděpodobnost výskytu dermatoglyfického vzoru *vír* alespoň u sedmi mužů z deseti je 23.13% (resp. 33.55%). Pravděpodobnost, výskytu vzoru *vír* nejvýše u pěti mužů z deseti je 53.96% (resp. 41.72%). Pravděpodobnost, výskytu vzoru *vír* u osmi nebo devíti mužů z deseti je 8.01% (resp. 13.49%).

Pravděpodobnost výskytu dermatoglyfického vzoru *vír* alespoň u 56 mužů ze sta je 33.04% (resp. 36.66%). Pravděpodobnost výskytu vzoru *vír* nejvýše u 53 mužů ze sta je 51.51% (resp. 47.60%). Pravděpodobnost výskytu vzoru *vír* u 60–85 mužů ze sta je 10.67% (resp. 12.66%).

Pravděpodobnost výskytu dermatoglyfického vzoru *vír* alespoň u 164 mužů z 300 je 33.89% (resp. 35.99%). Pravděpodobnost výskytu vzoru *vír* nejvýše u 160 mužů z 300 je 52.72% (resp. 50.46%). Pravděpodobnost výskytu vzoru *vír* u 170–175 mužů z 300 je 9.80% (resp. 10.59%).

Příklad 4.21. Aproximace binomického modelu normálním modelem
Předpokládejme, že pravděpodobnost výskytu dermatoglyfického vzoru *vír* na palci pravé ruky u mužů české populace $p = 0.533$. Pro $N = 10$, $N = 100$ a $N = 1000$ vykreslete graf pravděpodobnostní funkce binomického rozdělení a aproximujte jej křivkou funkce hustoty normálního rozdělení. Hodnoty obou funkcí porovnejte.

Řešení příkladu 4.21



Obrázek 7: Aproximace binomického modelu normálním modelem

4.4 Příklady k samostatnému procvičování

Příklad 4.22. Výpočet pravděpodobností za předpokladu binomického modelu
Předpokládejme, že pravděpodobnost výskytu dermatoglyfického vzoru *vír* na palci pravé ruky u mužů české populace $p = 0.533$. Jaká je pravděpodobnost, že ve vybraném vzorku 10 mužů bude výskyt dermatoglyfického vzoru *vír* na palci pravé ruky (a) právě u pěti mužů; (b) alespoň u 6 mužů; (c) nejvýše u dvou mužů; (d) u šesti, sedmi nebo osmi mužů.

Řešení příkladu 4.22

```
[1] 0.2408
```
92

```
[1] 0.4604
```
93

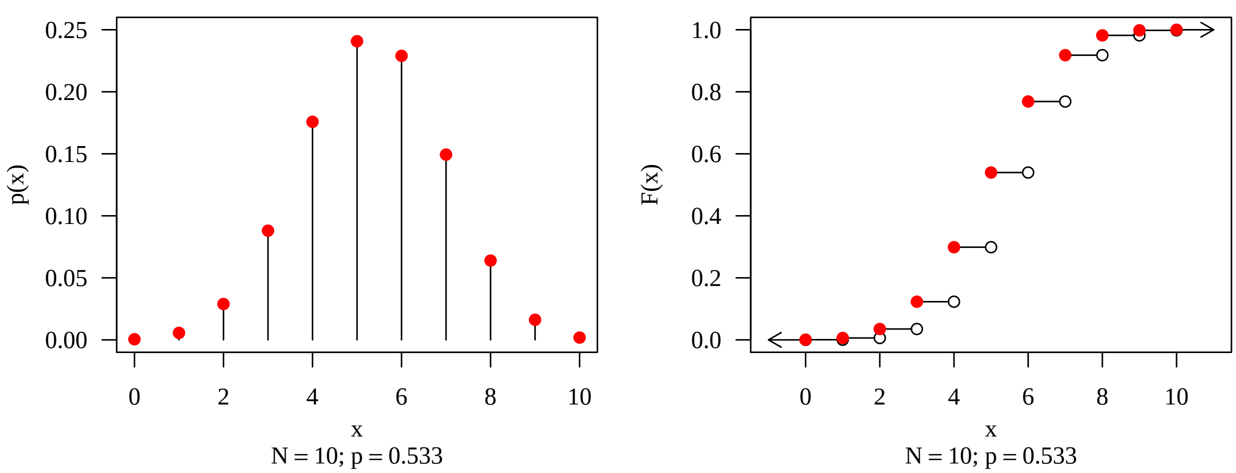
```
[1] 0.035
```
94

```
[1] 0.4423
```
95

Interpretace výsledků: Pravděpodobnost výskytu dermatoglyfického vzoru vír na palci pravé ruky u právě pěti mužů z deseti je 24.08%. Pravděpodobnost výskytu vzoru vír na palci pravé ruky u alespoň šesti mužů z deseti je 46.04%. Pravděpodobnost výskytu vzoru vír na palci pravé ruky u nejvýše dvou mužů z deseti je 3.50%. Pravděpodobnost výskytu vzoru vír na palci pravé ruky u šesti, sedmi nebo osmi mužů z deseti je 44.23%.

Příklad 4.23. Graf pravděpodobnostní a distribuční funkce binomického modelu
Nakreslete graf pravděpodobnostní funkce a distribuční funkce náhodné veličiny $X \sim \mathrm{Bin}(10, 0.533)$.

Řešení příkladu 4.23



Obrázek 8: Pravděpodobnostní funkce (vlevo) a distribuční funkce (vpravo) binomického modelu $\mathrm{Bin}(10, 0.533)$