4 Náhodné veličiny

4.1 Diskrétní náhodné veličiny

Alternativní rozdělení $\mathrm{Alt}(p)$

- Jeden Bernoulliho pokus X:
  - $X = 1 \ldots$ událost nastala; $X = 0 \ldots$ událost nenastala;
  - $\Pr(X = 1) = p$
  - $\Pr(X = 0) = 1 - p = q$
- Alternativní rozdělení:
  - $X \ldots$ výskyt sledované události v jednom Bernoulliho pokusu, přičemž pravděpodobnost nastání události v tomto pokusu je vyjádřena parametrem p.
  - $X \sim \mathrm{Alt}(p)$.
  - $\boldsymbol{\theta} = (p)^T$
  - pravděpodobnostní funkce:

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

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

Příklad 4.1. Popis reálné situace pomocí alternativního rozdělení
Mějme datový soubor 09-one-sample-probability-sutmet.txt obsahující údaje o výskytu epigenetického znaku *sutura metopica* na lebkách Ainů z ostrova Hokaido (více informací viz sekce **??**). Údaje o výskytu epigenetického znaku jsou také k dispozici v tabulce 1.

Tabulka 1: Údaje o výskytu epigenetického znaku *sutura metopica* na lebkách Ainů z ostrova Hokaido

Populace	Výskyt *sutura metopica* (X)		$\sum$
	Ano	Ne	
Ainové	6	178	184

Nechť náhodná veličina X popisuje výskyt epigenetického znaku *sutura metopica* na lebkách Ainů z ostrova Hokaido. Nalezněte rozdělení náhodné veličiny X a odhadněte hodnoty parametrů tohoto rozdělení.

Řešení příkladu 4.1
Nejprve se zamyslíme nad tím, které rozdělení by co nejlépe popisovalo náhodnou veličinu X. V rámci studie jsme zkoumali lebky $M = 184$ Ainů, přičemž u každé lebky jsme si zaznamenali, zda byl na ní přítomen výskyt epigenetického znaku *sutura metopica*. Náhodná veličina X tedy nabývá pouze dvou hodnot, a sice $X = 0$ v případě, že na lebce nebyl epigenetický znak přítomný, nebo $X = 1$ v případě, že na lebce byl epigenetický znak přítomný. Proto náhodná veličina X pochází z alternativního rozdělení, tj. $X \sim \mathrm{Alt}(p)$. Zbývá odhadnout hodnotu parametru p popisující pravděpodobnost výskytu epigenetického znaku *sutura metopica* na lebkách Ainů z ostrova Hokaido.

Odhad parametru p spočítáme jako podíl počtu výskytů epigenetického znaku *sutura metopica* na lebkách Ainů (viz čitatel vzorce 2, kde n zastupuje počet výskytů *sutura metopica* (0 nebo 1) a $m_{observed}$ uvádí, na kolika lebkách byl nulový počet ($n = 0$) výskytů *sutura metopica*, či na kolika lebkách byl jeden ($n = 1$) výskyt *sutura metopica*) ku celkovému počtu zkoumaných lebek (viz jmenovatel vzorce 2).

$$\widehat{p} = \frac{\sum_{n=0}^{1} n m_{observed}}{M} = \frac{0 \times 176 + 1 \times 6}{184} = \frac{6}{184} = 0.0326087 \doteq 0.03261. \tag{2}$$

```
M <- 184
n <- 0:1
m.obs <- c(176, 6)
p <- sum(n * m.obs) / M # 0.0326087
```

Interpretace výsledků: Náhodnou veličinu X popisující výskyt epigenetického znaku *sutura metopica* na lebkách Ainů z ostrova Hokaido modelujeme pomocí alternativního rozdělení s parametrem p, tj. $X \sim \mathrm{Alt}(p)$, kde $p = 0.03261$. Pravděpodobnost výskytu epigenetického znaku *sutura metopica* u populace Ainů z ostrova Hokaido je 3.26%.

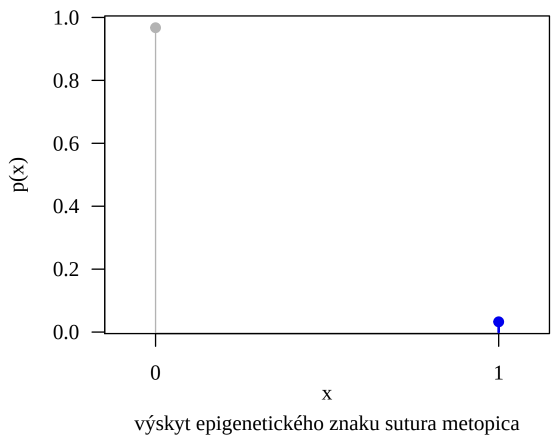
Příklad 4.2. Výpočet pravděpodobností za předpokladu alternativního rozdělení
Za předpokladu, že náhodná veličina X popisující výskyt epigenetického znaku *sutura metopica* u populace Ainů z ostrova Hokaido pochází z alternativního rozdělení a parametrem $p = 0.03261$, tj. $X \sim \text{Alt}(0.03261)$ vypočítejte pravděpodobnost, výskytu epigenetického znaku *sutura metopica* u populace Ainů; (b) pravděpodobnost absence epigenetického znaku *sutura metopica* u populace Ainů; (c) pravděpodobnost nejvýše žádného výskytu epigenetického znaku *sutura metopica*; (d) pravděodobnost nejvýše jednoho výskytu epigenetického znaku *sutura metopica.*

Řešení příkladu 4.2
Ze vzorce 1 víme, že pravděpodobnostní funkce alternativního rozdělení má tvar

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

kde v našem případě $p = 0.03261$.

(a) Začneme výpočtem pravděpodobnosti výskytu epigenetického znaku *sutura metopica* u populace Ainů. Tuto pravděpodobnost si nejprve vizualizujeme (viz obrázek 1).



Obrázek 1: Vizualizace pravděpodobnosti výskytu epigenetického znaku *sutura metopica* u populace Ainů

Z obrázku 1 je krásně viditelné, že naším úkolem je najít hodnotu pravděpodobnostní funkce $p(x)$ v hodnotě $x = 1$.

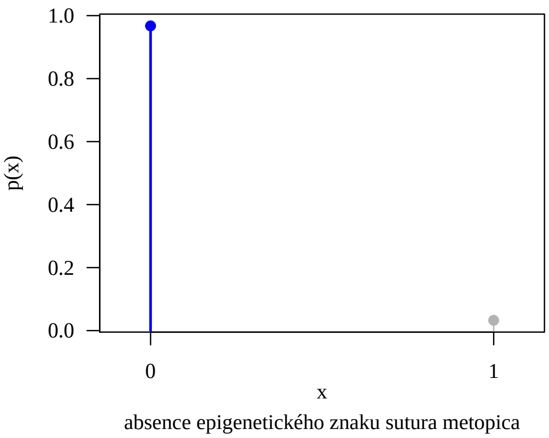
$$\begin{aligned}\Pr(X = 1) &= 0.0326087^1 \times (1 - 0.0326087)^{1-1} = 0.0326087^1 \times 0.9673913^0 = 0.0326087 \times 1 = \\ &= 0.0326087 \doteq 0.03261.\end{aligned}$$

```
x <- 1
p ^ (x) * (1 - p) ^ (1 - x) # 0.0326087
```

Interpretace výsledků: Pravděpodobnost výskytu epigenetického znaku *sutura metopica* v populaci Ainů ostrova Hokaido je 3.26%.

Poznámka: Všimněme si, že uvedený výpočet jsme vůbec provádět nemuseli. Pravděpodobnost výskytu epigenetického znaku *sutura metopica* je stejné jako hodnota parametru p, a to proto, že parametr p je sám pravděpodobností výskytu epigenetického znaku *sutura metopica.* Výpočet jsme prováděli pouze k procvičení výpočtu pravděpodobnostní funkce $p(x)$ alternativního rozdělení.

(b) Nyní se zaměříme na výpočet pravděpodobnosti absence epigenetického znaku *sutura metopica* u populace Ainů. Tuto pravděpodobnost si opět nejprve vizualizujeme (viz obrázek 2).



Obrázek 2: Vizualizace pravděpodobnosti absence epigenetického znaku *sutura metopica* u populace Ainů

Z obrázku 2 je zjevné, že naším úkolem je najít hodnotu pravděpodobnostní funkce $p(x)$ v hodnotě $x = 0$.

$$\begin{aligned}\Pr(X = 0) &= 0.0326087^0 \times (1 - 0.0326087)^{1-0} = 0.0326087^0 \times 0.9673913^1 = 1 \times 0.9673913 = \\ &= 0.9673913 \doteq 0.9674.\end{aligned}$$

```
x <- 0
p ^ (x) * (1 - p) ^ (1 - x) # 0.9673913
```

Interpretace výsledků: Pravděpodobnost absence epigenetického znaku *sutura metopica* v populaci Ainů ostrova Hokaido je 96.74%.

Poznámka: Všimněme si, že uvedený výpočet jsme opět vůbec provádět nemuseli. Pravděpodobnost absence epigenetického znaku *sutura metopica* jsme mohli získat přímo, odečtením pravděpodobnosti výskytu epigenetického znaku *sutura metopica* p od hodnoty 1, tj. $1 - p = 1 - 0.0326087 = 0.9673913$. Výpočet jsme prováděli pouze k procvičení výpočtu pravděpodobnostní funkce $p(x)$ alternativního rozdělení.

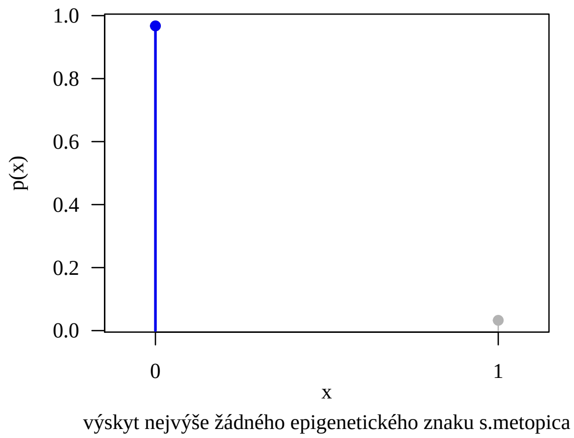
(c) Nyní se zaměříme na výpočet nejvýše žádného výskytu epigenetického znaku *sutura metopica* u populace Ainů. Nejvýše žádný výskyt znamená, že náhodné veličina X nabývá nejvýše hodnoty 0, tj. ze zadání máme za úkol vypočítat $\Pr(X \leq 0)$, tedy hodnotu distribuční funkce $F(x)$ v hodnotě $x = 0$. Tuto pravděpodobnost si nejprve vizualizujeme (viz obrázek 3).

Z obrázku 3 je zjevné, že naším úkolem bude finálně opět nalezení hodnoty pravděpodobnostní funkce $p(x)$ v hodnotě $x = 0$. K pochopení toho, jak jsme se z distribuční funkce $F(x)$ dostali k pravděpodobnostní funkci $p(x)$ stačí, když si uvědomíme, že nejvýše žádný výskyt *sutura metopica* znamená prostě žádný výskyt *sutura metopica*, protože nic menšího než to, že se epigenetický znak na lebce nevyskytuje, neexistuje. Tj. $\Pr(X \leq 0) = \Pr(X = 0)$.

$$\Pr(X \leq 0) = \Pr(X = 0) = 0.9673913 \doteq 0.9674.$$

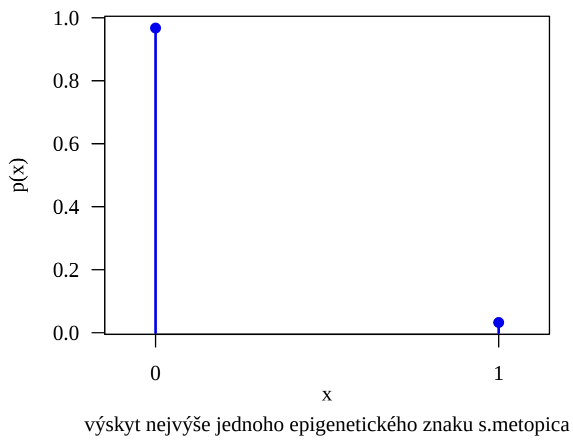
```
x <- 0
p ^ (x) * (1 - p) ^ (1 - x) # 0.9673913
```

Interpretace výsledků: Pravděpodobnost nejvýše žádného výskytu epigenetického znaku *sutura metopica* v populaci Ainů ostrova Hokaido je 96.74%.



Obrázek 3: Vizualizace pravděpodobnosti nejvýše žádného výskytu epigenetického znaku *sutura metopica* u populace Ainů

(d) Nakonec se zaměříme na výpočet nejvýše jednoho výskytu epigenetického znaku *sutura metopica* u populace Ainů. Nejvýše jeden výskyt znamená, že náhodné veličina X nabývá nejvýše hodnoty 1, tj. ze zadání máme za úkol vypočítat $\Pr(X \leq 1)$, tj. hodnotu distribuční funkce $F(x)$ v hodnotě $x = 1$. Tuto pravděpodobnost si nejprve vizualizujeme (viz obrázek 4).



Obrázek 4: Vizualizace pravděpodobnosti nejvýše jednoho výskytu epigenetického znaku *sutura metopica* u populace Ainů

Z obrázku 4 je zjevné, že $\Pr(X \leq 1) = \Pr(X = 0) + \Pr(X = 1)$.

$$\Pr(X \leq 1) = \Pr(X = 0) + \Pr(X = 0) = 0.9673913 + 0.0326087 = 1.$$

```
x <- 0:1
sum(p ^ (x) * (1 - p) ^ (1 - x)) # 1
```

Interpretace výsledků: Pravděpodobnost nejvýše jednoho výskytu epigenetického znaku *sutura metopica* v populaci Ainů ostrova Hokaido je 100.00%.

Poznámka: K pochopení toho, že výsledná pravděpodobnost nejvýše jednoho výskytu epigenetického znaku *sutura metopica* v populaci Ainů vyšla rovna 100 %, stačí, když si uvědomíme, že jiná situace, než výskyt žádného nebo jednoho epigenetického znaku *sutura metopica* na lebce nastat nemůže. Tedy pokud prozkoumáme lebku, je 100 % jisté, že se na ní *sutura metupica* buď vůbec nevyskytne, nebo vyskytne právě jednou. Celkově je tedy jisté, že se na lebce epigenetivcký znak *sutura metopica* vyskytne nejvýše jednou.

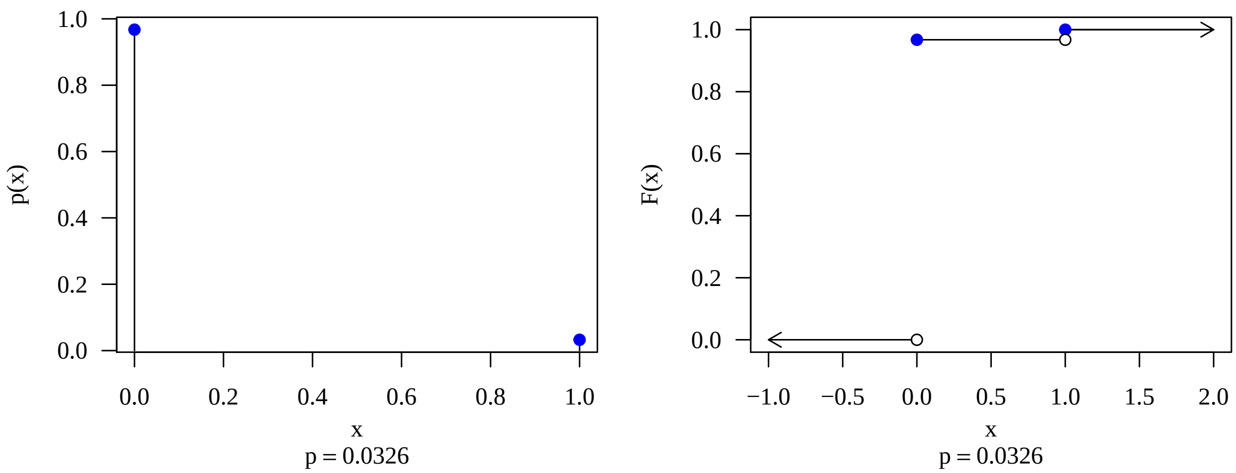
Příklad 4.3. Graf pravděpodobnostní funkce a distribuční funkce alternativního rozdělení
Zaměřte se nyní blíže na tvar alternativního rozdělení Alt(0.03261). Nakreslete graf pravděpodobnostní funkce $p(x)$ a graf distribuční funkce $F(x)$ tohoto rozdělení.

Řešení příkladu 4.3
Začneme s vykreslením grafu pravděpodobnostní funkce $p(x)$. Do proměnné `px` nejprve vložíme hodnoty pravděpodobnostní funkce alternativního rozdělení $p(x) = \Pr(X = x)$ pro $x = 0, 1$, které vypočítáme analogicky, jako v předchozím příkladu 4.2. Samotný graf potom vykreslíme příkazem `plot()` s argumentem `type = 'h'`, který zajistí vykreslení vertikálních čar v hodnotách 0 a 1 na ose x a s délkou odpovídající hodnotám pravděpodobnostní funkce $p(x)$, $x = 0, 1$. 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 hodnotu parametru p. Text popisku vygenerujeme pomocí kombinace funkcí `bquote()` a `paste()`. Uvnitř funkce `paste()` je vložena syntaxe popisku `p == .(round(p, 4))`, která nejprve vypíše písmeno p, následně znaménko rovnosti, a nakonec vyhodnocení proměnné `p` zaokrouhlené na čtyři desetinná místa, tj. 0.03261.

```
x     <- 0 : 1
p     <- 6 / 184
px    <- p ^ x * (1 - p) ^ (1 - x)
plot(x, px, type = 'h', ylab = 'p(x)', xlab = '', las = 1)
points(x, px, col = 'blue2', pch = 19)
mtext('x', side = 1, line = 2.1)
mtext(bquote(paste(p == .(round(p, 4)))), side = 1, line = 3.2)
```

První krokem pro vykreslení grafu distribuční funkce $F(x)$ je vytvoření vektoru `Fx` s hodnotami distribuční funkce alternativního rozdělení $F(x) = \Pr(X \leq x)$ pro $x = 0, 1$. Z předchozího příkladu 4.2 již víme, že $\Pr(X \leq 0) = 0.9673913$ a $\Pr(X \leq 1) = 1$. Konstrukci grafu zahájíme vykreslením prázdného grafu (příkaz `plot()` s argumentem `type = 'n'`) s rozsahem měřítka osy x od -1 do 2 (`xlim`) a rozsahem měřítka osy y od 0 do 1 (`ylim`). Argumentem `xlab = ''` potlačíme vypsání popisku osy x. Dále do grafu dokreslíme horizontální úsečky délky 1 začínající vždy v bodě [x, F(x)] a končící v bodě [x+1, F(x)]. K tomu použijeme funkci `segments()`, u níž povinně specifikujeme čtyři parametry `x0`, `x1`, `y0`, `y1` určující postupně x-ovou souřadnici počátečního bodu, x-ovou souřadnici konečného bodu, y-ovou souřadnici počátečního bodu a y-ovou souřadnici konečného bodu. Pokud za argumenty `x0`, `x1`, `y0`, `y1` dosadíme vektory obsahující všechny x-ové, resp. y-voé souřadnice počátečních, resp. konečných bodů, vykreslí se všechny horizontální úsečky najednou. 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 horizontální šipku umístěnou vpravo nahoře, směřující doprava s počátečním bodem [1, 1] a koncovým bodem [2, 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()`, `pch = 19`) se souřadnicí [x, F(x)] s modrým okrajem a modrým 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()`, `pch = 21`) 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 hodnotou parametru p.

```
Fx <- c(1 - p, 1)
plot(x, Fx, xlab = '', ylab = 'F(x)',
     xlim = c(-1, 2), ylim = c(0, 1), type = 'n', las = 1)
segments(x, Fx, x + 1, Fx)                      # vodorovne cary
arrows(0, 0,    -1, 0, length = 0.1)            # dolni sipka
arrows(N, 1, 2, 1, length = 0.1)                # horni sipka
points(x, Fx, col = 'blue2', pch = 19)          # plne body
points(x, c(0, Fx[1 : N]), pch = 21, bg = 'white', col = 'black') # prazdne body
mtext('x', side = 1, line = 2.1)
mtext(bquote(paste(p == .(round(p, 4)))), side = 1, line = 3.2)
```



Obrázek 5: Pravděpodobnostní funkce (vlevo) a distribuční funkce (vpravo) alternativního rozdělení Alt(0.03261)

Příklad 4.4. Popis reálné situace pomocí alternativního rozdělení
Mějme datový soubor 08-one-sample-probability-sexratio.txt obsahující údaje o pohlaví novorozenců (proměnná sex; m – chlapec, f – děvče) narozených v krajské nemocnici v průběhu jednoho roku (Alánová, 2008), více informací viz sekce ??). Nechť náhodná veličina X popisuje narození děvčete. Nalezněte rozdělení náhodné veličiny X a odhadněte hodnoty parametrů tohoto rozdělení.

Řešení příkladu 4.4
Nejprve načteme datový soubor (read.delim()) a vypíšeme první tři řádky tabulky (head()). Následně z datové tabulky vybereme pomocí dolarové syntaxe sloupec obsahující údaje o pohlaví (data$sex) a odstraníme z něj případná chybějící pozorování (na.omit()). Nakonec zjistíme rozsah náhodného výběru (length()) a počet narozených dívek a chlapců (table()).

```
data <- read.delim('00-Data//08-one-sample-probability-sexratio.txt')
head(data, n = 3)
```

```
  sex
1   m
2   m
3   f
```

```
sex <- na.omit(data$sex)
length(sex) # 1403
```

```
[1] 1403
```

```
table(sex)
```

```
sex
  f   m
674 729
```

Datový soubor obsahuje údaje o pohlaví $M = 1\,403$ novorozenců. Celkem se v krajské nemocnici za jeden rok narodilo 674 děvčat a 729 chlapců.

Za předpokladu, že náhodná veličina X popisuje narození děvčete, nabývá tato veličina pouze dvou hodnot, a sice $X = 1$ v případě, že se narodilo děvče (sex == 'f'), nebo $X = 0$ v případě, že se narodil chlapec (sex == 'm'). Náhodná veličina X tedy pochází z alternativního rozdělení, tj. $X \sim \mathrm{Alt}(p)$. Zbývá odhadnout hodnotu parametru p popisující pravděpodobnost narození děvčete. Odhad parametru p spočítáme jako podíl počtu narozených dívek ku celkovému počtu narozených dětí M.

$$\widehat{p} = \frac{\sum_{n=0}^{1} n m_{observed}}{M} = \frac{0 \times 729 + 1 \times 674}{1403} = \frac{674}{1403} = 0.4803991 \doteq 0.4804. \tag{3}$$

```
n <- 0:1
m.obs <- c(729, 674)
M <- sum(m.obs)
p <- sum(n * m.obs) / M #  0.4803991
```

Interpretace výsledků: Náhodnou veličinu X popisující narození děvčete modelujeme pomocí alternativního rozdělení s parametrem p, tj. $X \sim \mathrm{Alt}(p)$, kde $p = 0.4804$. Pravděpodobnost narození děvčete je 48.04%.

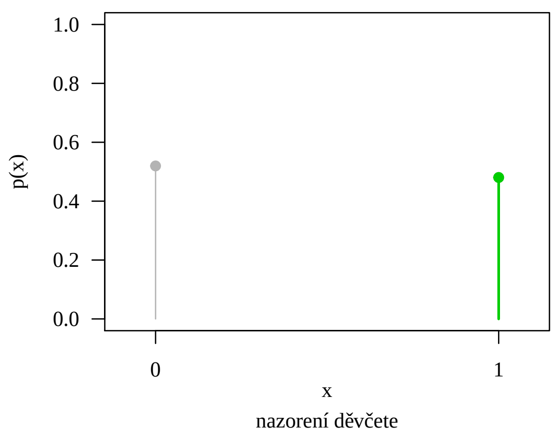
Příklad 4.5. Výpočet pravděpodobností za předpokladu alternativního rozdělení
Za předpokladu, že náhodná veličina X popisující narození děvčete pochází z alternativního rozdělení a parametrem $p = 0.4804$, tj. $X \sim \text{Alt}(0.4804)$ vypočítejte pravděpodobnost narození děvčete; (b) pravděpodobnost narození chlapce; (c) pravděpodobnost narození děvčete nebo chlapce.

Řešení příkladu 4.5
Ze vzorce 1 víme, že pravděpodobnostní funkce alternativního rozdělení má tvar

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

kde v našem případě $p = 0.4804$.

(a) Začneme výpočtem pravděpodobnosti narození děvčete. Tuto pravděpodobnost si nejprve vizualizujme (viz obrázek 6).



Obrázek 6: Vizualizace pravděpodobnosti narození děvčete

Z obrázku 6 je krásně viditelné, že naším úkolem je najít hodnotu pravděpodobnostní funkce $p(x)$ v hodnotě $x = 1$.

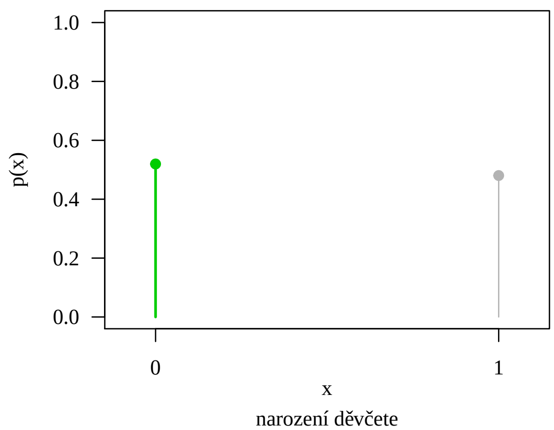
$$\begin{aligned}\Pr(X=1) &= 0.4803991^1 \times (1-0.4803991)^{1-1} = 0.4803991^1 \times 0.5196009^0 = 0.4803991 \times 1 = \\ &= 0.4803991 \doteq 0.4804.\end{aligned}$$

```
x <- 1
p ^ (x) * (1 - p) ^ (1 - x) # 0.4803991
```

Interpretace výsledků: Pravděpodobnost narození děvčete je 48.04%.

Poznámka: Uvedený výpočet jsme vůbec provádět nemuseli. Pravděpodobnost narození děvčete je stejná jako hodnota parametru p, neboť parametr p je sám pravděpodobností narození děvčete. Výpočet jsme prováděli pouze k procvičení výpočtu pravděpodobnostní funkce $p(x)$ alternativního rozdělení.

(b) Nyní se zaměříme na výpočet pravděpodobnosti narození chlapce. Tuto pravděpodobnost si opět nejprve vizualizujeme (viz obrázek 7). Vzhledem k tomu, že náhodná veličina X popisuje narození děvčete, odpovídá narození chlapce situaci, kdy náhodná veličina $X = 0$.



Obrázek 7: Vizualizace pravděpodobnosti narození chlapce

Z obrázku 7 je zjevné, že naším úkolem je najít hodnotu pravděpodobnostní funkce $p(x)$ v hodnotě $x = 0$.

$$\begin{aligned}\Pr(X = 0) &= 0.4803991^0 \times (1 - 0.4803991)^{1-0} = 0.4803991^0 \times 0.5196009^1 = 1 \times 0.5196009 = \\ &= 0.5196009 \doteq 0.5196009.\end{aligned}$$

```
x <- 0
p ^ (x) * (1 - p) ^ (1 - x) # 0.5196009
```

Interpretace výsledků: Pravděpodobnost narození chlace je 51.96%.

Poznámka: Uvedený výpočet jsme opět provádět nemuseli. Z podstaty věci je zjevné, že chlapec se narodil v případě, že se nenarodilo děvče. Proto pravděpodobnost narození chlapce lze vypočítat triviálně odečtením pravděpodobnosti narození děvčete od jedničky, tj. $1 - 0.4803991 = 0.5196009$. Podrobný výpočet jsme prováděli za účelem procvičení výpočtu pravděpodobnosti pomocí pravděpodobnostní funkce $p(x)$.

(c) Nakonec se zaměříme na výpočet pravděpodobnosti narození děvčete nebo chlapce. Narození chlapce odpovídá situaci, kdy náhodná veličina $X = 0$, narození děvčete odpovídá situaci, kdy náhodná veličina $X = 1$. Výpočet pravděpodobnosti narození děvčete nebo chlapce tedy odpovídá pravděpodobnosti $\Pr(X \leq 1)$, což je hodnota distribuční funkce $F(x)$ v hodnotě $x = 1$. Tuto pravděpodobnost si nejprve vizualizujeme (viz obrázek 8).

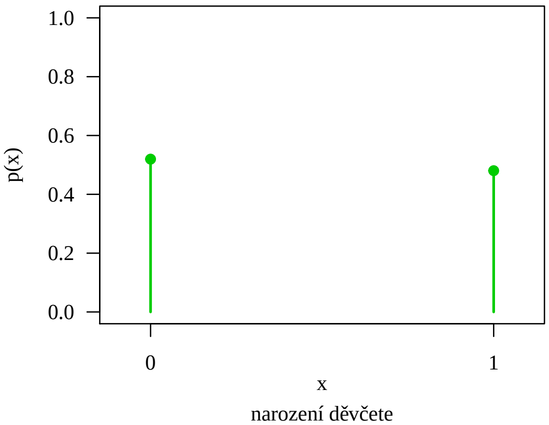
Z obrázku 8 je zjevné, že pravděpodobnost narození děvčete nebo chlapce odpovídá součtu pravděpodobnosti narození chlapce a pravděpodobnosti narození děvčete, tj. $\Pr(X \leq 1) = \Pr(X = 0) + \Pr(X = 1)$.

$$\Pr(X \leq 1) = \Pr(X = 0) + \Pr(X = 1) = 0.5196009 + 0.4803991 = 1.$$

```
x <- 0:1
sum(p ^ (x) * (1 - p) ^ (1 - x)) # 1
```

Interpretace výsledků: Pravděpodobnost narození děvčete nebo chlapce je 100.00%.

Poznámka: Výsledná 100% pravděpodobnost narození děvčete nebo chlapce nás asi příliš nepřekvapí. Je zjevné, že narodí-li se nový jedinec je jisté, že to bude buď chlapec nebo děvče.



Obrázek 8: Vizualizace pravděpodobnosti narození děvčete nebo chlapce

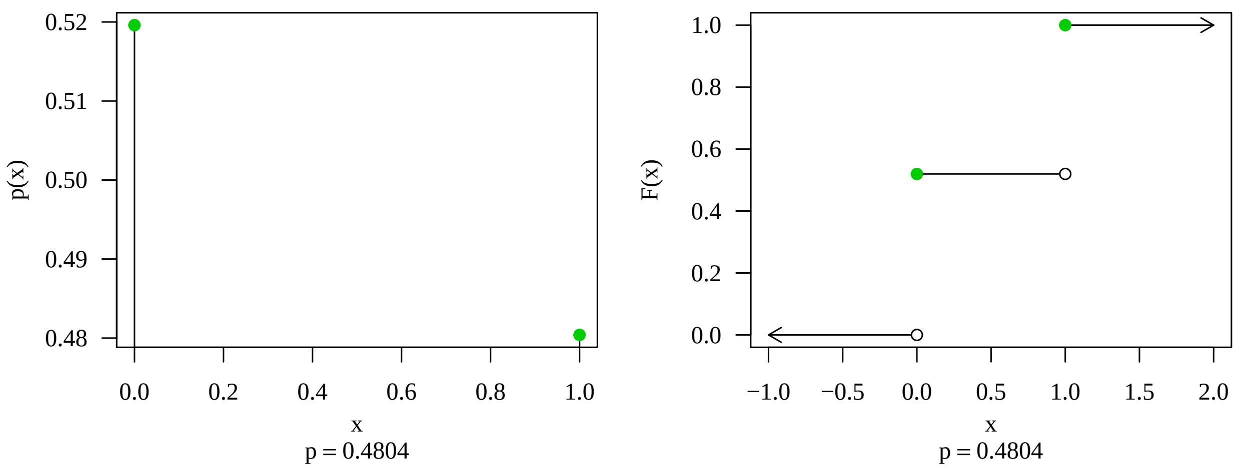
Příklad 4.6. Graf pravděpodobnostní funkce a distribuční funkce alternativního rozdělení
Zaměřte se nyní blíže na tvar alternativního rozdělení Alt(0.4804). Nakreslete graf pravděpodobnostní funkce $p(x)$ a graf distribuční funkce $F(x)$ tohoto rozdělení.

Řešení příkladu 4.6
Začneme s vykreslením grafu pravděpodobnostní funkce $p(x)$. Do proměnné px nejprve vložíme hodnoty pravděpodobnostní funkce alternativního rozdělení $p(x)$, které vypočítáme analogicky, jako v předchozím příkladu 4.5. Základ grafu pravděpodobnostní funkce $p(x)$ v podobě vertikálních čar s výškami odpovídajícím hodnotám $p(x)$ v hodnotách $x = 0, 1$ vykreslíme příkazem plot() s argumentem type = 'h'. V rámci příkazu plot() opět zakážene vypsání popisku osy x (xlab = ' '). Dále do grafu doplníme body (points()) ve výšce hodnot funkce $p(x)$. Nakonec do grafu doplníme popisek osy x a pod něj popisek uvádějící hodnotu parametru p (mtext()). Text popisku s hodnotou parametru p zaokrouhlenou na 4 desetinná místa (round()) vygenerujeme pomocí kombinace příkazů bquote() a paste().

```
x     <- 0 : 1
p     <- 674 / 1403
px    <- p ^ x * (1 - p) ^ (1 - x)
plot(x, px, type = 'h', ylab = 'p(x)', xlab = '', las = 1, ylim = c(0, 1))
points(x, px, col = 'green3', pch = 19)
mtext('x', side = 1, line = 2.1)
mtext(bquote(paste(p == .(round(p, 4)))), side = 1, line = 3.2)
```

Dále se zaměříme na vykreslení grafu distribuční funkce $F(x)$. Do proměnné Fx nejprve vložíme hodnoty distribuční funkce $F(x)$ náhodné veličiny X popisující narození děvčete, tj. $F(x) = \Pr(X \leq x)$ pro $x = 0, 1$. Analogicky jako v příkladu 4.3 platí i zde, že $\Pr(X \leq 0) = \Pr(X = 0)$, tj. pravděpodobnost, že se v jednom pokusu narodí nejvýše nula děvčat odpovídá pravděpodobnosti, že se v jednom pokusu nenarodí žádné děvče, (což odpovídá situaci, že se narodí chlapec). Dále $\Pr(X \leq 1) = 1$ (viz předchozí příklad 4.5, (c)). Konstrukci grafu zahájíme vykreslením prázdného okna bez popisku osy x s rozsahem měřítka osy x od -1 do 2 a rozsahem měřítka osy y od 0 do 1 (příkaz plot() s argumentem type = 'n' a s vhodnou specifikací argumentů xlab, xlim a ylim). Následně do grafu dokreslíme horizontální úsečky délky 1 začínající vždy v bodě [x, F(x)] a končící v bodě [x+1, F(x)] (segments()). Dále do grafu dokreslíme horozintální šipku umístěnou vlevo dole a směřující doleva, s počátečním bodem [0,0] a koncovým bodem [-1, 0] a horizontální šipku umístěnou vpravo nahoře, směřující doprava s počátečním bodem [1, 1] a koncovým bodem [2, 1] (arrows()). Následně na levý konci každé úsečky vykreslíme zelený plný bod se souřadnicí [x, F(x)] (příkaz points() s argumentem pch = 19) značící, že levý krajní bod každé úsečky má hodnotu distribuční funkce $F(x)$, a na pravý konec každé úsečky vykreslíme bod se souřadnicí [x, F(x-1)] s černým okrajem a bílým vnitřkem (příkaz points() s argumentem pch = 21) 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 hodnotou parametru p.

```r
Fx <- c(1 - p, 1)
plot(x, Fx, xlab = '', ylab = 'F(x)',
     xlim = c(-1, 2), ylim = c(0, 1), type = 'n', las = 1)
segments(x, Fx, x + 1, Fx)                        # vodorovne cary
arrows(0, 0,    -1, 0, length = 0.1)              # dolni sipka
arrows(N, 1, 2, 1, length = 0.1)                  # horni sipka
points(x, Fx, col = 'green3', pch = 19)           # plne body
points(x, c(0, Fx[1 : N]), pch = 21, bg = 'white', col = 'black') # prazdne body
mtext('x', side = 1, line = 2.1)
mtext(bquote(paste(p == .(round(p, 4)))), side = 1, line = 3.2)
```



Obrázek 9: Pravděpodobnostní funkce (vlevo) a distribuční funkce (vpravo) alternativního rozdělení Alt(0.4804)

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

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

$$p(x) = \binom{N}{x} p^x (1-p)^{N-x} \qquad x = 0, 1, \ldots, N; \tag{4}$$

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

Příklad 4.7. Popis reálné situace pomocí binomického rozdělení
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 12 dětmi byla pozorována početnost narozených chlapců. Údaje ze studie jsou uvedeny v tabulce 2. Více informací o datovém souboru a studii viz sekce **??**.

Tabulka 2: 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

Nechť náhodná veličina X popisuje počet chlapců v rodinách s dvanácti dětmi. Nalezněte rozdělení náhodné veličiny X a odhadněte hodnoty parametrů tohoto rozdělení.

Řešení příkladu 4.7
Nejprve se zaměřme na nalezení rozdělení, který co nejvýstižněji popisuje náhodnou veličinu X. Protože počet chlapců v rodině je vždy celé číslo, náhodná veličina X popisující počet chlapců v rodině je diskrétní náhodná veličina. Rozdělení náhondé veličiny X musíme tedy hledat mezi diskrétními rozděleními. Dále si uvědomme, že v rámci jedné rodiny máme celkem $N = 12$ Bernoulliho pokusů X_i, $i = 1, \ldots, 12$, přičemž sledovanou událostí v jednom Bernoulliho pokusu je narození chlapce. Při narození každého z dvanácti dětí tedy buď událost nastala (narodil se chlapec; $X_i = 1$, $i = 1, \ldots, 12$) nebo událost nenastala (narodilo se děvče; $X_i = 0$, $i = 1, \ldots, 12$). 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 5) ku celkovému počtu všech dětí v těchto rodinách (viz jmenovatel vzorce 5).

$$\widehat{p} = \frac{\sum_{n=0}^{N} n m_{observed}}{NM} = \frac{0 \times 3 + 1 \times 24 + \cdots + 11 \times 45 + 12 \times 7}{12 \times 6115} = \frac{38\,100}{73\,380} = 0.519215 \doteq 0.5192. \tag{5}$$

```
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 rozdělení 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 4.8. Porovnání pozorovaných a očekávaných početností binomického rozdělení
Na základě výše uvedené úvahy popisujeme počet chlapců v rodině s dvanácti dětmi pomocí binomického modelu $\mathrm{Bin}(12, 0.5192)$. Nyní ověříme, zda jsme k popisu reálné situace 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 4.8
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, atd. Výsledné pravděpodobnosti vynásobíme počtem rodin, tj. číslem 6115, čímž zjistíme, v kolika rodinách se za přepokladu $X \sim \mathrm{Bin}(12, 0.5192)$ nenarodí žádný chlapec, narodí jeden chlapec, narodí dva chlapci, atd. K vypočítání těchto pravděpodobností použijeme pravděpodobnostní funkci $p(x)$, kde $x = 0, 1, \ldots, 12$. Hodnoty pravděpodobnostní funkce binomického rozdělení stanovíme příkazem dbinom(), kde prvním argumentem jsou hodnoty $x = 0, 1, \ldots, 12$, druhý argument size = 12 odpovídá počtu pokusů N a třetí argument prob = 0.5192 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.

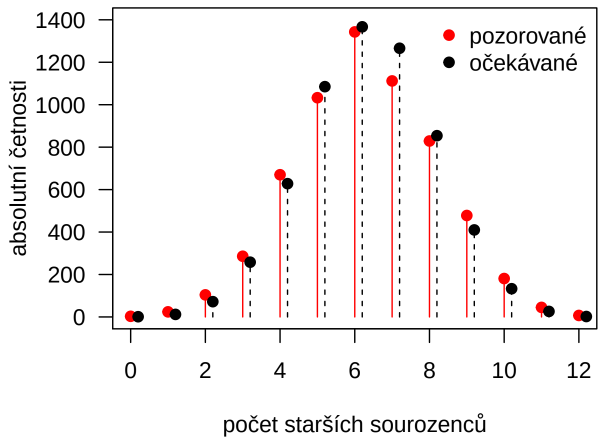
```
p.exp       <- dbinom(0:12, size = N, prob = p)
m.exp       <- round(p.exp * 6115)
tab         <- data.frame(rbind(pozorovane = m.obs, ocekavane = m.exp))
names(tab) <- 0:12
tab
```

```
           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 = 'absolutní četnosti', xlab = 'počet starších sourozenců')
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('pozorované', 'očekávané'), bty = 'n')
```

Z obrázku 10 je zřejmé, že zvolené binomické rozdělení popisuje počet chlapců v rodině s 12 dětmi velmi dobře. Červené body reprezentující pozorované četnosti se drží blízko černých bodů reprezentující očekávané četnosti. Největší rozdíl nastává v počtu rodin se sedmi chlapci, kde pozorujeme o 154 méně rodin, než by teoreticky mělo být. Jde však o jedinou výraznější odchylku, která vzhledem k reálným datům není nikterak závažná.



Obrázek 10: Porovnání pozorovaných a očekávaných početností v binomickěm modelu $\text{Bin}(12, 0.5192)$

Interpretace výsledků: Z tabulky pozorovaných a očekávaných četností a na základě grafické vizualizace soudíme, že zvolené binomické rozdělení $\text{Bin}(12, 0.5192)$ je vhodné k popisu počtu chlapců v rodině s dvanácti dětmi.

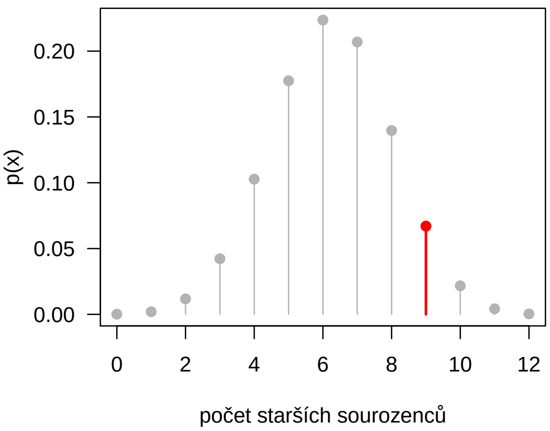
Příklad 4.9. Výpočet pravděpodobností za předpokladu binomického rozdělení
Za předpokladu, že náhodná veličina X popisující počet chlapců v rodinách s dvanácti dětmi pochází z binomického rozdělení $\text{Bin}(12, 0.5192)$ vypočítejte pravděpodobnost, že v rodině s dvanácti dětmi bude (a) právě devět chlapců; (b) nejvýše čtyři chlapci; (c) alespoň osm chlapců; (d) čtyři, pět, šest, nebo sedm chlapců.

Řešení příkladu 4.9
Ze vzorce 4 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 $p = 0.5192$.

(a) Začneme výpočtem pravděpodobnosti, že v rodině s dvanácti dětmi bude právě devět chlapců. Tuto pravděpodobnost si nejprve vizualizujeme (viz obrázek 11).



Obrázek 11: Vizualizace pravděpodobnosti, že v rodině s dvanácti dětmi bude právě devět chlapců

Z obrázku 11 je krásně viditelné, že naším úkolem je najít hodnotu pravděpodobnostní funkce $p(x)$ v hodnotě $x = 9$.

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

Poznámka: Pro zopakování uvádíme, že člen ve tvaru $\binom{n}{k}$ (čteme *n nad k*) je tzv. *kombinační číslo.* Každé kombinační číslo je možné vyjádřit ve tvaru $\binom{n}{k} = \frac{n!}{k!(n-k)!}$, kde pro $n!$ (čteme *n faktorial*) platí, že $n! = n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1$. Zde na ukázku uvádíme konktérně výpočet kombinačního čísla $\binom{12}{9}$. Tj.

$$\binom{12}{9} = \frac{12!}{9!(12-9)!} = \frac{12!}{9! \times 3!} = \frac{12 \times 11 \times \cdots \times 2 \times 1}{(9 \times 8 \times \cdots \times 2 \times 1) \times (3 \times 2 \times 1)} = \frac{12 \times 11 \times 10}{3 \times 2 \times 1} = \frac{1320}{6} = 220.$$

Kombinační číslo lze vypočítat pomocí softwaru R prostřednictvím funkce choose(). Pravděpodobnost, že v rodině s dvanácti dětmi bude právě devět chlapců můžeme tedy přímo vypočítat pomocí softwaru R přepisem vzorce 4.

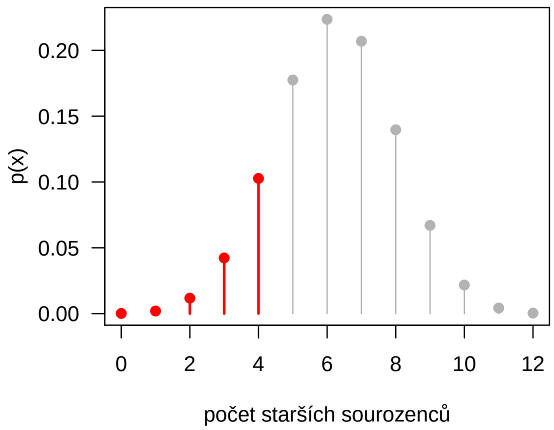
```
N <- 12
p <- 0.5192
x <- 9
choose(N, x) * p ^ x * (1 - p) ^ (N - x) # 0.06703911
```

K přímému výpočtu hodnoty pravděpodobnostní funkce $p(x)$ binomického rozdělení můžeme použít funkci dbinom() implementovanou v softwaru R. Prvním argumentem funkce bude hodnota $x = 9$, druhým argumentem počet pokusů N a třetím argumentem pravděpodobnost narození chlapce p.

```
dbinom(9, N, p) # 0.06703911
```

Interpretace výsledků: Pravděpodobnost, že v rodině s dvanácti dětmi bude právě devět chlapců, je 6.70%.

(b) Nyní se zaměříme na výpočet pravděpodobnosti, že v rodině s dvanácti dětmi budou nejvýše čtyři chlapci. Situaci si opět nejprve vizualizujeme (viz obrázek 12).



Obrázek 12: Vizualizace pravděpodobnosti, že v rodině s dvanácti dětmi budou nejvýše čtyři chlapci

Z obrázku 12 je zřejmé, že tentokrát hledáme hodnotu distribuční funkce $F(x)$ v hodnotě $x = 4$. Zároveň vidíme, že tuto hodnotu můžeme také vypočítat jako součet pravděpodobnostích funkcí $p(x)$ v hodnotách $x = 0, 1, 2, 3$ a 4.

$$\begin{aligned}
\Pr(X \leq 4) &= \sum_{i=0}^{4} \Pr(X = i) = \Pr(X = 0) + \Pr(X = 1) + \Pr(X = 2) + \Pr(X = 3) + \Pr(X = 4) \\
&= \binom{12}{0} \times 0.5192^0 \times (1 - 0.5192)^{12-0} + \binom{12}{1} \times 0.5192^1 \times (1 - 0.5192)^{12-1} + \\
&\quad + \binom{12}{2} \times 0.5192^2 \times (1 - 0.5192)^{12-2} + \binom{12}{3} \times 0.5192^3 \times (1 - 0.5192)^{12-3} + \\
&\quad + \binom{12}{4} \times 0.5192^4 \times (1 - 0.5192)^{12-4} = \\
&= 1 \times 0.5192^0 \times 0.4808^{12} + 12 \times 0.5192^1 \times 0.4808^{11} + 66 \times 0.5192^2 \times 0.4808^{10} + \\
&\quad + 220 \times 0.5192^3 \times 0.4808^9 + 495 \times 0.5192^4 \times 0.4808^8 = \\
&= 0.0001526067 + 0.001977539 + 0.01174513 + 0.04227726 + 0.1027211 = \\
&= 0.1588736 \doteq 0.1589.
\end{aligned}$$

Pravděpodobnost, že v rodině s dvanácti dětmi budou nejvýše čtyři chlaci můžeme vypočítat pomocí softwaru R a to buď ručním přepisem,

```
x <- 0:4
sum(choose(N, x) * p ^ x * (1 - p) ^ (N - x)) # 0.1588736
```

nebo pomocí již známé funkce dbinom(), prostřednictvím které vypočítáme hodnoty pravděpodobnostních funkcí $p(x)$ v hodnotách $x = 0, 1, 2, 3$ a 4, jež následně sečteme příkazem sum(),

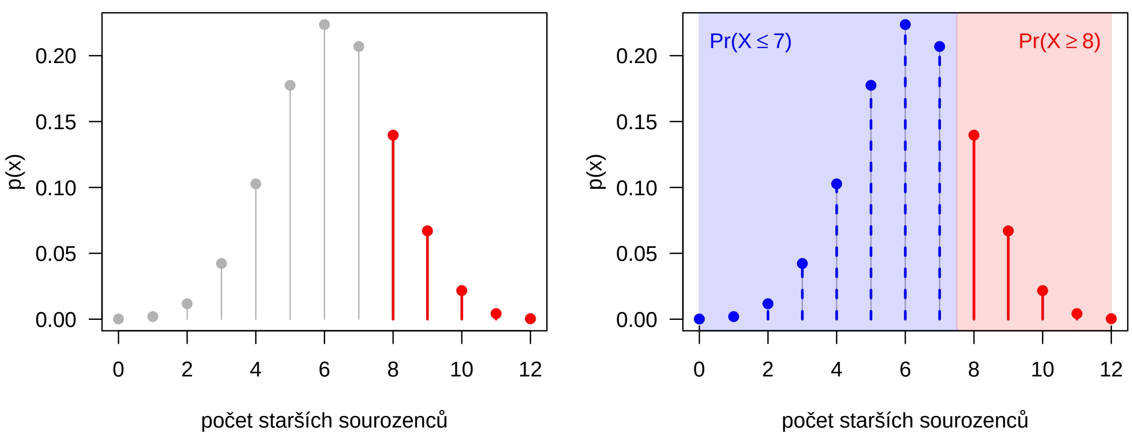
```
sum(dbinom(0:4, N, p)) # 0.1588736
```

nebo pomocí příkazu pbinom(), který vrátí přímo hodnotu distribuční funkce $\Pr(X \leq 4)$. Prvním argumentem funkce bude hodnota $x = 4$, druhým argumentem bude počet pokusů N a třetím argumentem pravděpodobnost narození chlapce p.

```
pbinom(4, N, p) # 0.1588736
```

Interpretace výsledků: Pravděpodobnost, že v rodině s dvanácti dětmi budou nejvýše čtyři chlapci, je 15.89%.

(c) Pravděpodobnost, že v rodině s dvanácti dětmi bude alespoň osm chlapců můžeme vypočítat buď přímo jako součet pravděpodobnostních funkcí $p(x)$ v hodnotách $x = 8, 9, 10, 11$ a 12 (viz obrázek 13 vlevo), nebo můžeme využít vlastnosti komplementarity, tj. $\Pr(X \geq x) = 1 - \Pr(X < x) = 1 - \Pr(X \leq x - 1)$ (viz obrázek 13 vpravo).



Obrázek 13: Vizualizace pravděpodobnosti, že v rodině s dvanácti dětmi bude alespoň osm chlapců (vlevo); vizualizace výpočtu pomocí komplementarity (vpravo)

$$
\begin{aligned}
\Pr(X \geq 8) &= \Pr(X = 8) + \Pr(X = 9) + \Pr(X = 10) + \Pr(X = 11) + \Pr(X = 12) = \\
&= \binom{12}{8} \times 0.5192^8 \times (1 - 0.5192)^{12-8} + \cdots + \binom{12}{12} \times 0.5192^{12} \times (1 - 0.5192)^{12-12} = \\
&= 495 \times 0.5192^8 \times (1 - 0.5192)^4 + \cdots + 1 \times 0.5192^{12} \times (1 - 0.5192)^0 = \\
&= 0.2330869 \doteq 0.2331.
\end{aligned}
$$

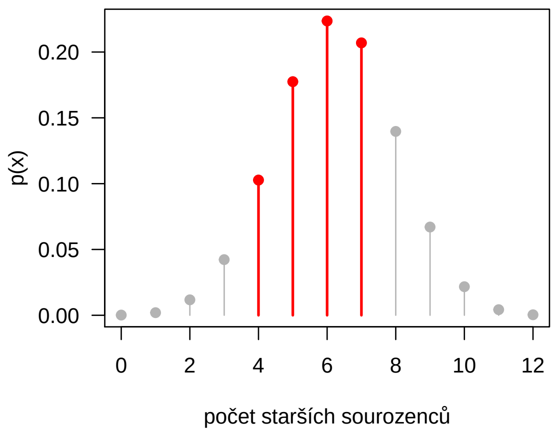
$$
\begin{aligned}
\Pr(X \geq 8) &= 1 - \Pr(X < 8) = 1 - \Pr(X \leq 7) \\
&= 1 - \left(\sum_{i=0}^{7} \Pr(X = i) \right) = 1 - (\Pr(X = 0) + \cdots + \Pr(X = 7)) \\
&= 1 - \left(\binom{12}{0} \times 0.5192^0 \times (1 - 0.5192)^{12-0} + \cdots + \binom{12}{7} \times 0.5192^7 \times (1 - 0.5192)^{12-7}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

Pravděpodobnost, že v rodině s dvanácti dětmi bude alespoň osm chlapců můžeme vypočítat pomocí softwaru , a to buď přímým součtem pravděpodobnostních funkcí $p(x)$ v hodnotách $x = 8, 9, 10, 11$ a 12, nebo odečtením hodnot pravděpodobnostních funkcí $p(x)$ v hodnotách $x = 0, 1, 2, 3, 4, 5, 6$ a 7 od celku, nebo odečtením hodnoty distribuční funkce $\Pr(X \leq 7)$ od celku.

```
sum(dbinom(8:12,  N, p)) # 0.2330869
1 - sum(dbinom(0:7, N, p)) # 0.2330869
1 - pbinom(7, N, p) # 0.2330869
```

Interpretace výsledků: Pravděpodobnost, že v rodině s dvanácti dětmi bude alespoň osm chlapců, je 23.31%.

(d) Nakonec vypočítáme pravděpodobnost, že v rodině s dvanácti dětmi bude čtyři pět, šest nebo sedm chlapců (vizualizace situace viz obrázek 14)



Obrázek 14: Vizualizace pravděpodobnosti, že v rodině s dvanácti dětmi bude čtyři, pět, šest nebo sedm chlapců

Pravděpodobnost, že v rodině s dvanácti dětmi bude čtyři pět, šest nebo sedm chlapců vypočítáme přímým součtem hodnot pravděpodobnostní funkce $p(x)$ v hodnotách $x = 4, 5, 6$ a 7.

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

K výpočtu pravděpodobnosti $\Pr(4 \leq X \leq 7)$ můžeme tentokrát použít buď funkci dbinom(), která vede na výpočet analogický výpočtu 6, tj.

```
sum(dbinom(4 : 7, N, p)) # 0.7107605
```

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

```
pbinom(7, N, p) - pbinom(3, N, p) # 0.7107605
```

Vidíme, že oba postupy vedou ke stejnému výsledku.

Interpretace výsledků: Pravděpodobnost, že v rodině s dvanácti dětmi bude čtyři, pět, šest, nebo sedm chlapců, je 71.08%.

Příklad 4.10. Graf pravděpodobnostní funkce a distribuční funkce binomického rozdělení
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.

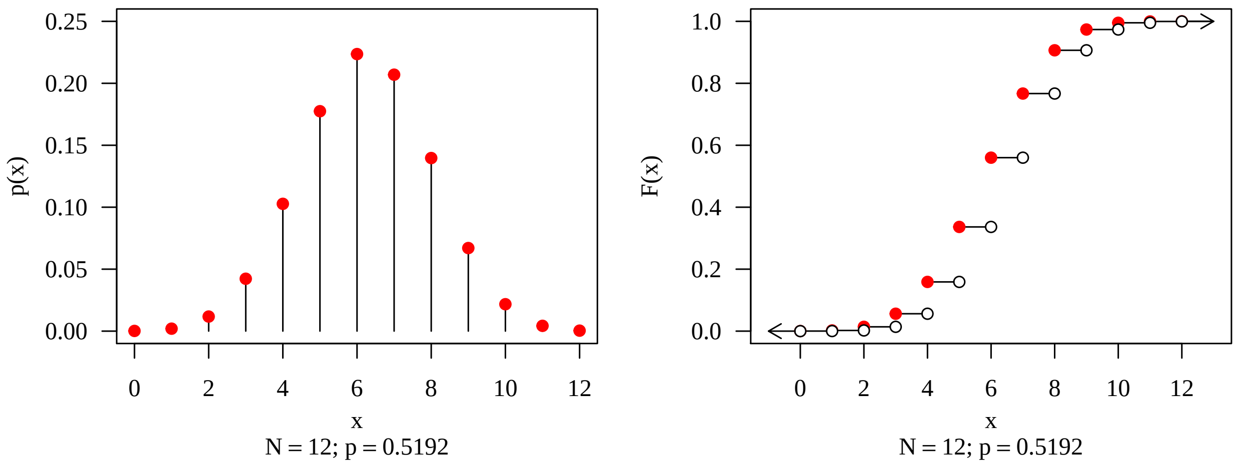
Řešení příkladu 4.10
Začneme s vykreslením grafu pravděpodobnostní funkce $p(x)$. Do proměnné px nejprve vložíme hodnoty pravděpodobnostní funkce binomického rozdělení $p(x) = \Pr(X = x)$ pro $x = 0, 1, \ldots, 12$, které vypočítáme příkazem dbinom(). Na prvním místě ve funkci dbinom() budou hodnoty 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.

```
N     <- 12
x     <- 0 : N
p     <- 0.5192
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 horizontální 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(), pch = 19) se souřadnicí [x, F(x)] s červeným okrajem 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(), pch = 21) 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.

```
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, Fx, col = 'red', pch = 19)          # plne body
points(x, c(0, Fx[1 : N]), pch = 21, bg = 'white', col = 'black') # prazdne body
```

```
mtext('x', side = 1, line = 2.1)
mtext(bquote(paste(N==.(N), '; ', p==.(p))), side = 1, line = 3.2)
```



Obrázek 15: Pravděpodobnostní funkce (vlevo) a distribuční funkce (vpravo) binomického modelu $\mathrm{Bin}(12, 0.5192)$

Příklad 4.11. 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 4.11

```
[1] 0.2408
```

```
[1] 0.4604
```

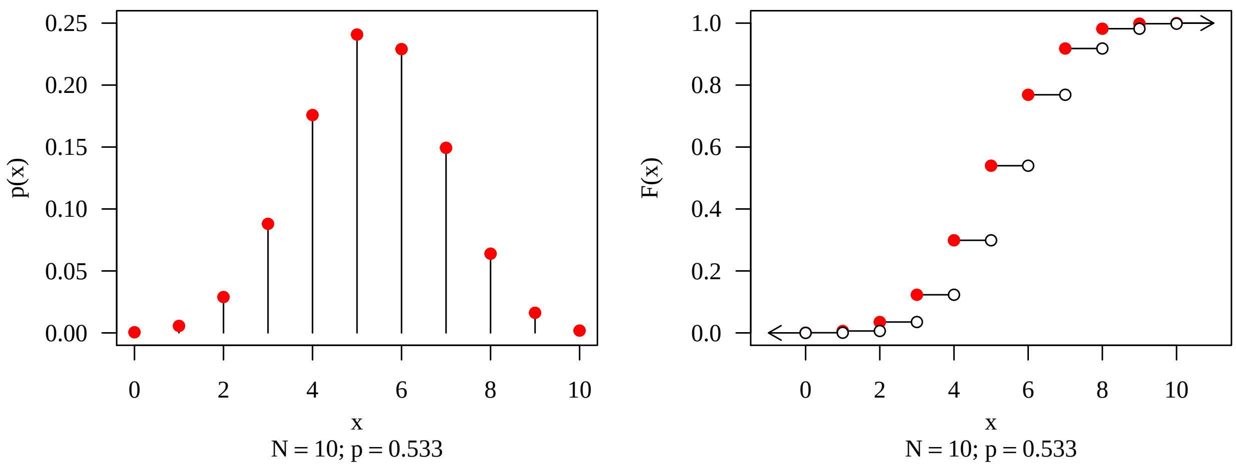
```
[1] 0.035
```

```
[1] 0.4423
```

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 4.12. 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 4.12



Obrázek 16: Pravděpodobnostní funkce (vlevo) a distribuční funkce (vpravo) binomického modelu $\mathrm{Bin}(10, 0.533)$

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)

Příklad 4.13. 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 4.13
Střední hodnotu počtu starších sourozenců odhadneme pomocí vzorce

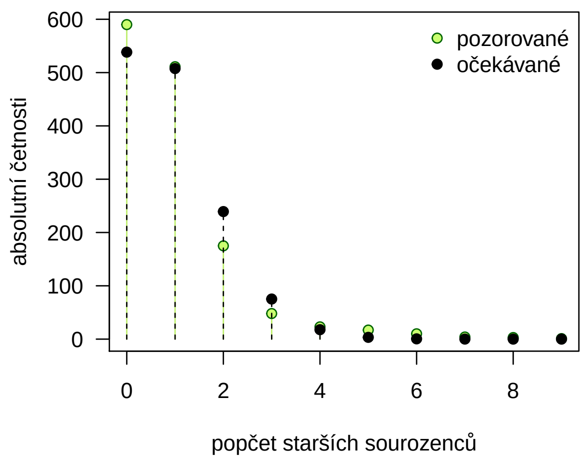
$$\lambda = \frac{\text{počet starších sourozenců}}{\text{počet novorozenců}} = \frac{\sum_{i=1}^{N} x_i}{N}. \tag{7}$$

```
[1] 0.9428365
```
128

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

Příklad 4.14. 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 4.14



Obrázek 17: Porovnání pozorovaných a očekávaných početností v Poissonově modelu

Příklad 4.15. Výpočet pravděpodobností za předpokladu Poissonova modelu
Vraťem se nyní k příkladu 4.13. 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 4.15

```
[1] 0.2403672
```
129

```
[1] 0.01568161
```
130

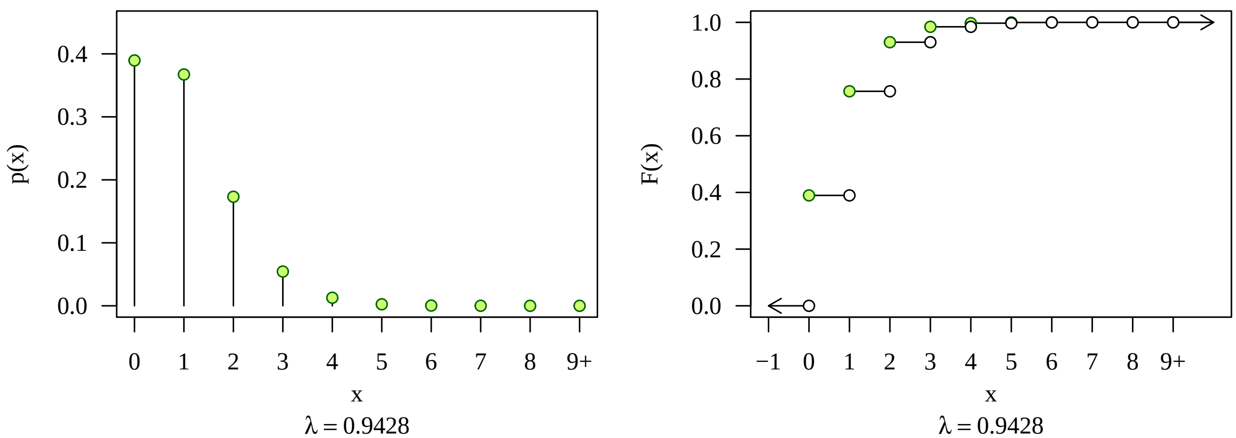
```
[1] 0.9299071
```
131

```
[1] 0.367255
```
132

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%.

Příklad 4.16. 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 4.16



Obrázek 18: 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 4.17. 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 4.17
Střední hodnotu počtu smrtelných úrazů v pruských armádních jednotkách způsobeným kopnutím koněm odhadneme pomocí vzorce

$$\lambda = \frac{\sum_{n=0}^{N} n m_{observed}}{\sum_{n=0}^{N} m_{observed}}. \tag{8}$$

```
[1] 0.61
```
133

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

Příklad 4.18. Výpočet pravděpodobností za předpokladu Poissonova modelu
Vraťem se nyní k příkladu 4.17. 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 4.18

```
[1] 0.9758853
```
134

```
[1] 0.5433509
```
135

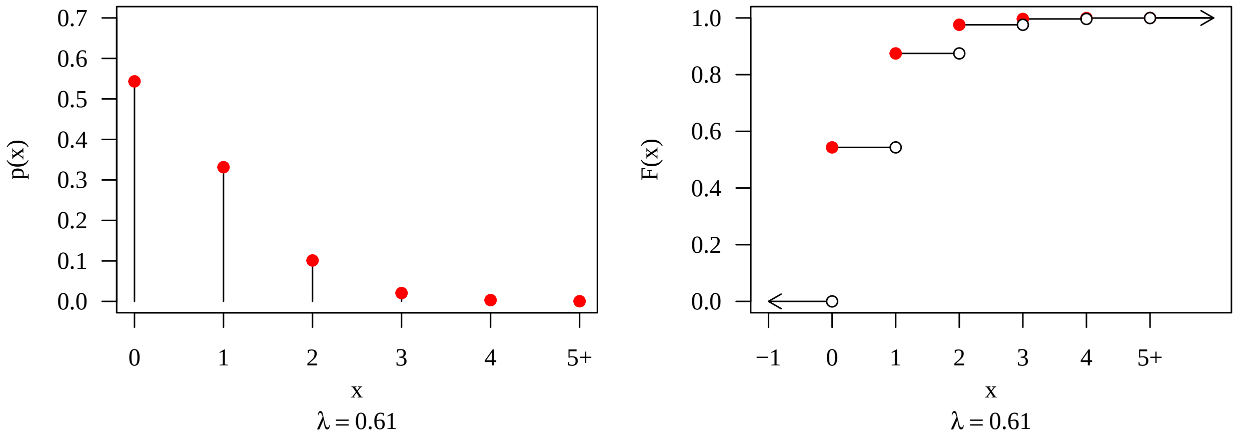
```
[1] 0.1252051
```
136

```
[1] 0.331444
```
137

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 4.19. 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 4.19



Obrázek 19: 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 4.20. 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 4.20
Střední hodnotu počtu úrazů dělníků v továrně za jeden rok odhadneme pomocí vzorce

$$\widehat{\lambda} = \frac{\sum_{n=0}^{N} n m_{observed}}{\sum_{n=0}^{N} m_{observed}}. \tag{9}$$

```
[1] 0.4652                                                                138
```

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

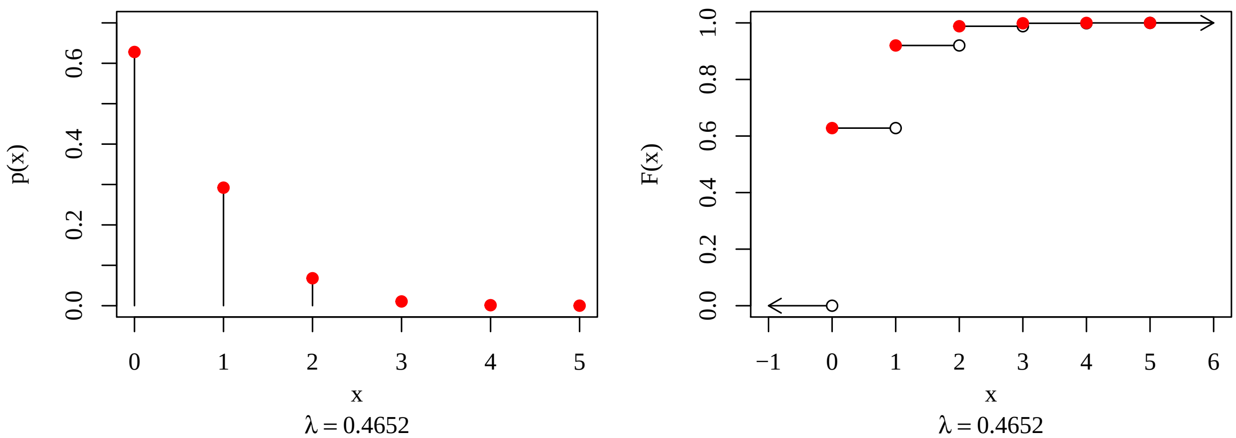
Příklad 4.21. Graf pravděpodobnostní a distribuční funkce Poissonova modelu
Nakreslete graf pravděpodobnostní a distribuční funkce Poissonova rozdělení Po(0.4652) v hodnotách $x = 0, 1, 2, 3, 4$ a $x \geq 5$.

Řešení příkladu 4.21

Příklad 4.22. 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. $X \sim$ Po(λ) 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 4.22

```
[1] 0.628                                                                 139
```



Obrázek 20: Pravděpodobnostní a distribuční funkce Poissonova modelu

```
[1] 0.0118
```
140

```
[1] 0.9881
```
141

```
[1] 0.372
```
142

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 $N(\mu, \sigma^2)$

- $X_1, \dots, X_n$ ... 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 4.23. 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 $\mu = 151.74\,\mathrm{mm}$, směrodatná ochylka $s = 11\,\mathrm{mm}$ (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 4.23

```
[1] 0
```
143

```
[1] 0.1429243
```
144

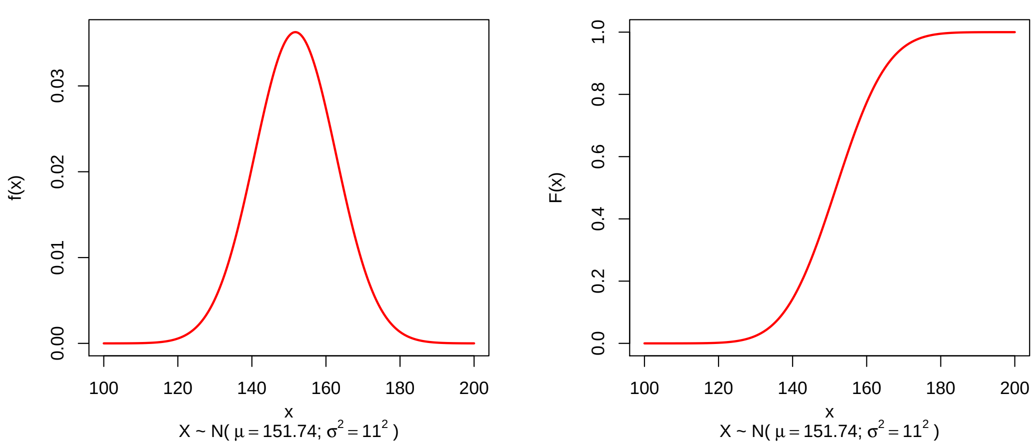
```
[1] 0.2263537
```
145

```
[1] 0.630722
```
146

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 4.24. Graf hustoty a distribuční funkce normálního modelu
Vraťme se k příkladu 4.23. Nakreslete graf hustoty a distribuční funkce náhodné veličiny $X \sim N(151.74, 11)$.

Řešení příkladu 4.24



Příklad 4.25. 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 $\mu = 3078.94\,\mathrm{g}$, směrodatná ochylka $s = 697\,\mathrm{g}$ (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 4.25

```
[1] 0.8495533
```
147

```
[1] 0.743032
```
148

```
[1] 0.09317345
```
149

```
[1] 0
```
150

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 4.26. Výpočet pravděpodobností na základě standardizovaného normálního modelu
Vraťme se nyní k předchozímu příkladu 4.25. 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 4.26

```
[1] 0.8495533
```
151

```
[1] 0.743032
```
152

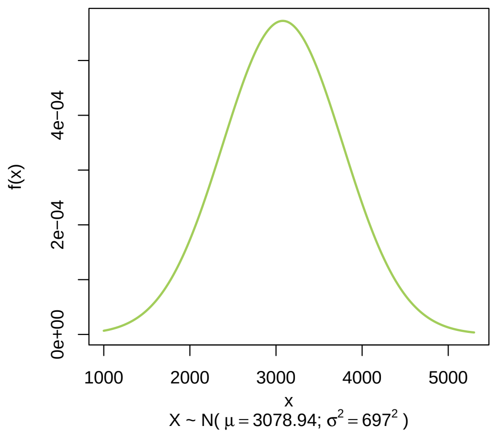
```
[1] 0.09317345
```
153

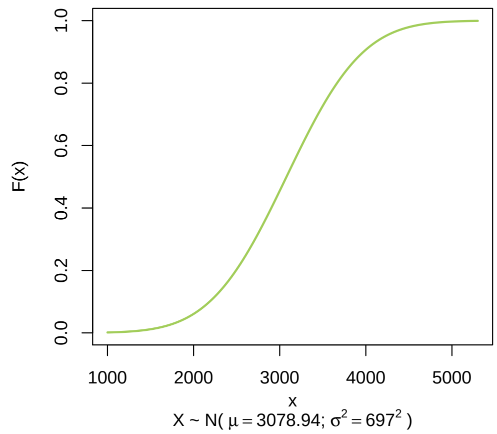
```
[1] 0
```
154

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 4.27. Graf hustoty a distribuční funkce normálního modelu
Vraťme se k příkladu 4.25. Nakreslete graf hustoty a distribuční funkce náhodné veličiny $X \sim N(3078.94, 697)$.

Řešení příkladu 4.27


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


F(x)
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 $\text{Bin}(N, p)$, tedy pro $N \to \infty$, $p \to 0.5$:

$$X \sim \text{Bin}(N, p) \to X \sim N(\mu, \sigma^2),$$

kde $\mu = Np$ a $\sigma^2 = Np(1-p)$.

- Haldova podmínka: Nechť $X \sim \text{Bin}(N, p)$ 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 4.28. 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 4.28

```
          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
```

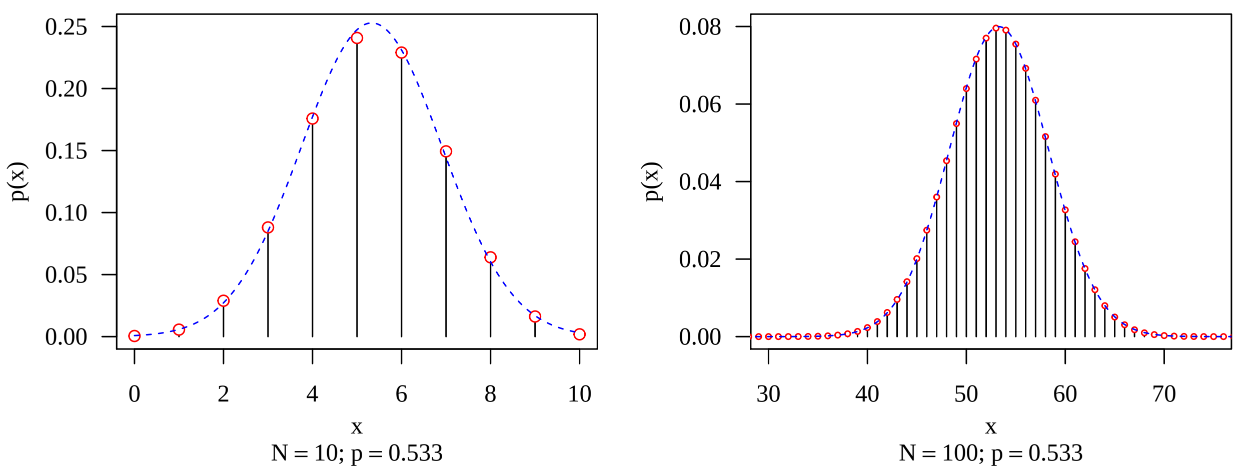
Interpretace výsledků: Pravděpodobnost výskytu dermatoglyfického vzoru *vír* alespoň u sedmi mužů z deseti je 23.13% (resp. 33.55%). Pravděpodobnost, výskytu vzoru *vír* nejvýše u pěti mužů z deseti je 53.96% (resp. 41.72%). Pravděpodobnost, výskytu vzoru *vír* u osmi nebo devíti mužů z deseti je 8.01% (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 4.29. 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 4.29



Obrázek 21: Aproximace binomického modelu normálním modelem