4 Pravděpodobnostní modely pro diskrétní náhodné veličiny

4.1 Základní pojmy počtu pravděpodobnosti

- **Náhodný pokus**: jednorázové uskutečnění konstantně vymezeného souboru definičních podmínek, jehož každé opakování vede k právě jednomu z více možných výsledků.
- **Základní prostor**: neprázdná množina možných výsledků náhodného pokusu. Značí se Ω a nazývá se **jistý jev**. Opakem je **nemožný jev** $\varnothing$. Možné výsledky se značí $\omega_1, \omega_2, \dots$
- **Jev**: vymezená množina výsledků náhodného pokusu. Značí se A, B, C, ...
- **Jev opačný k jevu** A: $\bar{A} = \Omega - A$.
- **Neslučitelné jevy**: $A \cap B = \varnothing$.
- **Jevové pole**: systém všech jevů, značí se $\mathcal{A}$.
- **Měřitelný prostor**: dvojice $(\Omega, \mathcal{A})$.
- **Pravděpodobnost**: množinová funkce $\Pr : \mathcal{A} \to \mathbb{R}$, která splňuje následující tři axiomy:
  - každému jevu přiřazuje nezáporné číslo (axiom nezápornosti),
  - jistému jevu přiřazuje číslo 1 (axiom normovanosti),
  - sjednocení neslučitelných jevů přiřazuje součet pravděpodobností těchto jevů (axiom spočetné aditivity); jde o teoretický protějšek relativní četnosti.
- **Pravděpodobnostní prostor**: trojice $(\Omega, \mathcal{A}, \Pr)$.
- **Klasická pravděpodobnost**: $\Pr(A) = \frac{m(A)}{m(\Omega)}$, kde $m(A)$ je počet možných výsledků příznivých nastoupení jevu A a $m(\Omega)$ je počet všech možných výsledků. Klasická pravděpodobnost se dá použít jen tehdy, když je základní prostor konečný a všechny možné výsledky mají stejnou šanci nastat.
- **Stochasticky nezávislé jevy**: jevy $A_1, \dots, A_n \in \mathcal{A}$ jsou stochasticky nezávislé, když $\forall 1 \leq i < j < \dots < k \leq n$: $\Pr(A_i \cap A_j \cap \dots \cap A_k) = \Pr(A_i)\Pr(A_j)\dots\Pr(A_k)$. Stochastická nezávislost jevů je teoretickým protějškem četnostní nezávislosti množin.
- **Podmíněná pravděpodobnost jevu** A **za podmínky** H: $\Pr(A|H) = \frac{\Pr(A \cap H)}{\Pr(H)}$, $\Pr(H) > 0$; jde o teoretický protějšek podmíněné relativní četnosti.
- **Náhodná veličina**: zobrazení $X : \Omega \to \mathbb{R}$, které možnému výsledku ω přiřazuje číslo $X(\omega)$; jde o teoretický protějšek skalárního znaku. Číselnou realizaci $X(\omega)$ zkráceně značíme x.
- **Dvourozměrný náhodný vektor**: zobrazení $(X, Y)^T$: $\Omega \to \mathbb{R}^2$, které možnému výsledku ω přiřazuje dvojici čísel $(X(\omega), Y(\omega))^T$; jde o teoretický protějšek dvourozměrného znaku. Číselnou realizaci $(X(\omega), Y(\omega))^T$ zkráceně značíme $(x, y)^T$.
- **Transformovaná náhodná veličina**: $Z = g(X)$.
- **Distribuční funkce náhodné veličiny**: $\forall x \in \mathbb{R} : F(x) = \Pr(X \leq x)$; jde o teoretický protějšek empirické distribuční funkce. Známe-li distribuční funkci náhodné veličiny, řekneme, že známe její **pravděpodobnostní rozložení** (používá se též termín pravděpodobnostní model). Toto rozložení zpravidla závisí na nějakém parametru ϑ. Zkráceně píšeme $X \sim L(\vartheta)$ a čteme: náhodná veličina X má rozložení L s parametrem ϑ.

4.2 Diskrétní náhodná veličina

Diskrétní náhodná veličina nabývá pouze konečně nebo spočetně mnoha hodnot. Její pravděpodobnostní chování je popsáno **pravděpodobnostní funkcí**: $\forall x \in \mathbb{R} : p(x) = \Pr(X = x)$; jde o zidealizovaný protějšek četnostní funkce. Pravděpodobnostní funkce je s distribuční funkcí spjata součtovým vztahem: $\forall x \in \mathbb{R} : F(x) = \sum_{t \le x} p(t)$.

4.2.1 Alternativní rozložení

Provedeme pokus, jehož výsledkem je buď nastání sledované události (úspěch) s pravděpodobností ϑ nebo nenastání sledované události (neúspěch) s pravděpodobností $1-\vartheta$. Náhodná veličina X udává počet nastání sledované události v tomto pokusu, nabývá tedy hodnot 0 nebo 1. Píšeme $X \sim A(\vartheta)$.

$$p(x) = \begin{cases} \vartheta^x (1-\vartheta)^{1-x} \text{ pro } x = 0, 1, \\ 0 \text{ jinak.} \end{cases}$$

4.2.2 Binomické rozložení

Provádíme posloupnost n opakovaných nezávislých pokusů, v každém pokusu může nastat sledovaná událost s pravděpodobností ϑ; jde o tzv. bernoulliovskou posloupnost pokusů. Náhodná veličina X udává počet nastání sledované události v této posloupnosti pokusů. Píšeme $X \sim \mathrm{Bi}(n, \vartheta)$; pro $n = 1$ jde o alternativní rozložení.

$$p(x) = \begin{cases} \binom{n}{x} \vartheta^x (1-\vartheta)^{n-x} \text{ pro } x = 0, 1, \ldots, n, \\ 0 \text{ jinak.} \end{cases}$$

Binomické rozložení se pojí s výběry s vracením (s opakováním).

Vzorce pro výpočty pravděpodobností pomocí binomického rozložení

- Pravděpodobnost, že v n pokusech nastane sledovaná událost právě x-krát:
  $\Pr(X = x) = p(x) = \binom{n}{x} \vartheta^x (1-\vartheta)^{n-x}$.
  K výpočtu v softwaru R slouží funkce dbinom(x, n, ϑ).
- Pravděpodobnost, že v n pokusech nastane sledovaná událost nejvýše x_1-krát:
  $\Pr(X \le x_1) = F(x_1) = \sum_{x=0}^{x_1} p(x)$.
  K výpočtu v softwaru R slouží funkce pbinom(x_1, n, ϑ).
- Pravděpodobnost, že v n pokusech nastane sledovaná událost alespoň x_0-krát:
  $\Pr(X \ge x_0) = 1 - \Pr(X \le x_0 - 1) = 1 - F(x_0 - 1) = 1 - \sum_{x=0}^{x_0-1} p(x)$.
  Výpočet lze v softwaru R provést takto: 1 − pbinom(x_0 − 1, n, ϑ).
- Pravděpodobnost, že v prvních n pokusech nastane sledovaná událost alespoň x_0-krát a nejvýše x_1-krát:
  $\Pr(x_0 \le X \le x_1) = F(x_1) - F(x_0 - 1) = \sum_{x=x_0}^{x_1} p(x)$.
  Výpočet lze v softwaru R provést takto: pbinom(x_1, n, ϑ) − pbinom(x_0 − 1, n, ϑ).

Příklad 4.1. Řešený příklad
V rámci studie poměru pohlaví u lidí (Geissler, 1889) bylo na základě záznamů z nemocnic v Sasku zaznamenáno rozložení počtu chlapců v $M = 6115$ rodinách s dvanácti dětmi. Údaje ze studie jsou uvedeny v tabulce 4.1.

Tabulka 4.1: Počet chlapců v rodinách s dvanácti dětmi

x	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

Předpokládejme, že náhodná veličina X popisuje počet chlapců v rodině s dvanácti dětmi. (a) Najděte rozložení, které dostatečně dobře vystihuje pravděpodobnostní chování náhodné veličiny X, a stanovte odhady parametrů tohoto rozložení; (b) ověřte, že nalezené rozložení je vhodné na popis náhodné veličiny X; vypočítejte očekávané absolutní četnosti za předpokladu nalezeného rozložení a graficky je porovnejte s pozorovanými absolutními četnostmi.

Řešení příkladu 4.1

Počet chlapců v rodině s dvanácti dětmi je diskrétní znak, k jeho popisu tedy použijeme diskrétní náhodnou veličinu. Narození chlapce bylo zkoumáno ve dvanácti Bernoulliho pokusech $X_1, \ldots, X_{12}$, přičemž v každému pokusu mohlo dojít k nastání sledované události ($X_i = 1$; narodil se chlapec), nebo k nenastání sledované události ($X_i = 0$; narodilo se děvče). O náhodné veličině X tedy předpokládáme, že pochází z binomického rozložení s parametry $n = 12$ a ϑ, kde ϑ je pravděpodobnost narození chlapce v rodině s dvanácti dětmi. Parametr ϑ odhadneme jako podíl počtu všech narozených chlapců ku celkovému počtu všech narozených dětí v rodinách s dvanácti dětmi.

$$\widehat{\vartheta} = \frac{\text{počet narozených chlapců}}{\text{počet narozených dětí}} = \frac{\sum_{x=0}^{n} x m_{observed}}{nM} = \frac{38\,100}{73\,380} = 0,5192.$$

```r
n <- 12
x <- 0:n
m.obs <- c(3, 24, 104, 286, 670, 1033, 1343, 1112, 829, 478, 181, 45, 7)
M <- sum(m.obs) # 6115
theta <- round(sum(x * m.obs) / (n * M), 4) # 0,5192
```

Odhad parametru $\widehat{\vartheta} = 0,5192$. Pravděpodobnost narození chlapce v rodině s dvanácti dětmi je 51,92 %. O náhodné veličině X předpokládáme, že pochází z binomického rozložení s parametry $n = 12$ a $\vartheta = 0,5192$, tj. $X \sim \mathrm{Bi}(12; 0,5192)$.

Nyní zbývá ověřit, zda binomické rozložení dostatečně dobře popisuje reálná data. Za předpokladu, že $X \sim \mathrm{Bi}(12; 0,5192)$, vypočítáme očekávané absolutní četnosti rodin s 0, 1, ..., 12 chlapci. Pomocí příkazu dbinom() stanovíme nejprve pravděpodobnostní funkci rozložení $\mathrm{Bi}(12; 0,5192)$ v hodnotách 0, 1, ..., 12. Hodnoty pravděpodobnostní funkce následně vynásobíme celkovým počtem rodin ($M = 6\,115$) a zaokrouhlíme na celá čísla. Tím získáme očekávané absolutní četnosti rodin s 0, 1, ..., 12 chlapci. Očekávané absolutní četnosti vložíme společně s pozorovanými absolutními četnostmi do souhrnné tabulky, kterou vytvoříme příkazem data.frame().

```r
m.exp <- round(dbinom(0:12, n, theta) * 6115)
tab <- data.frame(rbind(m.obs, m.exp))
names(tab) <- 0:12
```

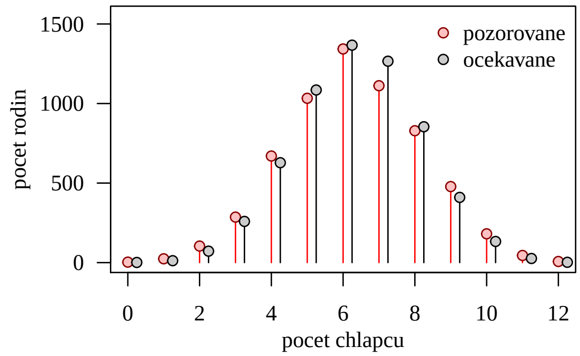
```
      0  1   2   3   4    5    6    7   8   9  10 11 12
m.obs 3 24 104 286 670 1033 1343 1112 829 478 181 45  7
m.exp 1 12  72 259 628 1085 1367 1266 854 410 133 26  2
```

Za předpokladu, že $X \sim \mathrm{Bi}(12; 0,5192)$, se v jedné rodině s dvanácti dětmi nenarodí žádný chlapec, ve dvanácti rodinách se narodí jeden chlapec, v 72 rodinách se narodí dva chlapci, apod.

Pozorované a očekávané absolutní četnosti nyní vzájemně graficky porovnáme (viz obrázek 4.1). Nejprve příkazem plot() s argumentem type = 'h' vykreslíme graf pozorovaných absolutních četností reprezentovaných červenými svislými úsečkami. Příkazem points() zakončíme svislé úsečky červenými body. Následně pomocí příkazu lines() s argumentem type = 'h' dokreslíme do grafu svislé černé úsečky reprezentující očekávané absolutní četnosti, které tentokrát zakončíme šedými body. Pro lepší přehlednost posuneme tyto úsečky a body o vzdálenost 0,25 směrem doprava. Nakonec do grafu doplníme legendu příkazem legend().

```r
plot(x, m.obs, type = 'h', col = 'red', ylim = c(0, 1550),
     xlab = 'pocet chlapcu', ylab = 'pocet rodin', las = 1)
points(x, m.obs, pch = 21, col = 'darkred', bg = 'rosybrown1')
lines (x + 0.25, m.exp, type = 'h', col = 'black')
points(x + 0.25, m.exp, pch = 21, col = 'black', bg = 'grey80')
legend('topright', pch = c(21, 21), col = c('darkred', 'black'), bty = 'n',
       pt.bg = c('rosybrown1', 'grey80'), legend = c('pozorovane', 'ocekavane'))
```

Z obrázku 4.1 vidíme, že pozorované a očekávané absolutní četnosti jsou si hodnotami velmi blízké. Grafická vizualizace tedy podporuje náš závěr, že náhodná veličina X popisující počet chlapců v rodině s dvanácti dětmi pochází



Obrázek 4.1: Porovnání pozorovaných absolutních četností a očekávaných absolutních četností za předpokladu rozložení Bi(12; 0, 5192)

z binomického rozložení s parametry $n = 12$ a $\vartheta = 0,5192$. ★

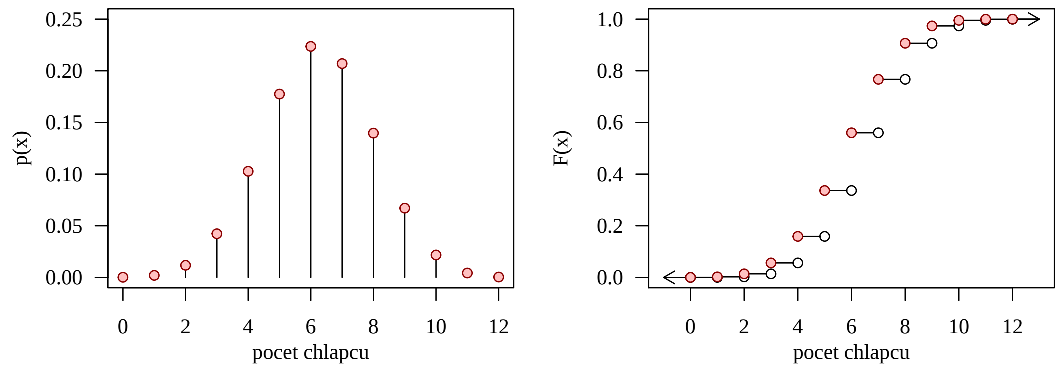
Příklad 4.2. Řešený příklad
Za předpokladu, že náhodná veličina X popisující počet chlapců v rodině s dvanácti dětmi pochází z rozložení Bi(12; 0, 5192), (a) nakreslete graf pravděpodobnostní funkce $p(x)$ a graf distribuční funkce $F(x)$; (b) vypočítejte pravděpodobnost, že v rodině s dvanácti dětmi bude (i) právě sedm chlapců; (ii) nejvýše čtyři chlapci; (iii) alespoň šest chlapců; (iv) tři, čtyři, pět nebo šest chlapců.

Řešení příkladu 4.2
K vykreslení grafu pravděpodobnostní funkce $p(x)$ nejprve příkazem dbinom() vypočítáme hodnoty pravděpodobnostní funkce v bodech $x = 0, 1, \ldots, 12$. Příkazem plot() s argumentem type = 'h' potom vykreslíme hodnoty pravděpodobnostní funkce reprezentované svislými černými úsečkami. Svislé úsečky zakončíme červenými body, které vykreslíme příkazem points(). Výsledný graf je zobrazen na obrázku 4.2 vlevo.

```
n <- 12
x <- 0:n
px <- dbinom(x, n, theta)
plot(x, px, type = 'h', ylim = c(0, 0.25), xlab = 'pocet chlapcu',
     ylab = 'p(x)', las = 1)
points(x, px, pch = 21, col = 'darkred', bg = 'rosybrown1')
```

K vykreslení grafu distribuční funkce $F(x)$ nejprve příkazem pbinom() vypočítáme hodnoty distribuční funkce v bodech $x = 0, 1, \ldots, 12$. Následně si vykresíme prázdný graf příkazem plot() s argumentem type = 'n'. Do prázdného grafu zaneseme nejprve vodorovné úsečky pomocí příkazu segments(x0, y0, x1, y1), kde x0, resp. y0 je vektor x-ových, resp. y-ových souřadnic počátečních bodů úseček a x1, resp. y1 je vektor x-ových, resp. y-ových souřadnic koncových bodů úseček. Dále příkazem arrows() dokreslíme do grafu šipky popisující průběh distribuční funkce na intervalu $(-\infty; 0)$, resp. na intervalu $\langle 12; \infty)$. Syntax příkazu arrows() je analogická syntaxi příkazu segments(); argumentem length specifikujeme délku hrotu šipky. Nakonec příkazem points() dokreslíme každé vodorovné úsečce zprava bílý bod s černým obrysem reprezentující, že úsečka je zprava otevřená, a zleva plný červený bod reprezentující, že úsečka je zleva uzavřená. Výsledný graf je zobrazen na obrázku 4.2 vpravo.

```
Fx <- pbinom(x, n, theta)
plot(x, Fx, xlab = 'pocet chlapcu', type = 'n', xlim = c(-1, 13),
     ylim = c(0, 1), ylab = 'F(x)', las = 1)
segments(x, Fx, x + 1, Fx)
arrows(0, 0, -1, 0, length = 0.1)
arrows(12, 1, 13, 1, length = 0.1)
points(x, c(0, Fx[1:12]), pch = 21, col = 'black', bg = 'white')
points(x, Fx, pch = 21, col = 'darkred', bg = 'rosybrown1')
```



Obrázek 4.2: Graf pravděpodobnostní funkce $p(x)$ (vlevo) a distribuční funkce $F(x)$ (vpravo) rozložení $\mathrm{Bi}(12; 0,5192)$

Nyní se zaměříme na výpočet pravděpodobností za předpokladu binomického rozložení s parametry $n = 12$ a $\vartheta = 0,5192$. Pravděpodobnost, že v rodině s dvanácti dětmi bude právě sedm chlapců, tj. $\Pr(X = 7)$ odpovídá hodnotě pravděpodobnostní funkce $p(x)$ rozložení $\mathrm{Bi}(12; 0,5192)$ v bodě $x = 7$. Tuto hodnotu vypočítáme pomocí příkazu dbinom().

```
n <- 12
theta <- 0.5192
dbinom(7, n, theta) # 0,2069618
```

Pravděpodobnost, že v rodině s dvanácti dětmi budou nejvýše čtyři chlapci, tj. $\Pr(X \leq 4)$ odpovídá hodnotě distribuční funkce $F(x)$ v bodě $x = 4$. Tuto hodnotu vypočítáme příkazem pbinom().

```
pbinom(4, n, theta) # 0,1588736
```

Pravděpodobnost, že v rodině s dvanácti dětmi bude alespoň šest chlapců, tj. $\Pr(X \geq 6)$ vypočítáme tak, že od jedné odečteme pravděpodobnost, že v rodině bude nejvýše pět chlapců. Tuto pravděpodobnost, odpovídající hodnotě distribuční funkce $F(x)$ v bodě $x = 5$, vypočítáme příkazem pbinom().

```
1 - pbinom(5, n, theta) # 0,6636462
```

Pravděpodobnost, že v rodině s dvanácti dětmi budou tři, čtyři, pět nebo šest chlapců, tj. $\Pr(3 \leq X \leq 6)$ můžeme vypočítat dvěma způsoby. Prvním způsobem je, že sečteme hodnoty pravděpodobnostní funkce $p(x)$ v bodech $x = 3$, 4, 5 a 6. Tyto hodnoty vypočítáme pomocí příkazu dbinom() a sečteme příkazem sum(). Druhým způsobem je, že od pravděpodobnosti, že v rodině bude nejvýše šest chlapců, odečteme pravděpodobnost, že v rodině budou nejvýše dva chlapci. Obě pravděpodobnosti jsou hodnotami distribuční funkce $F(x)$ v bodě $x = 6$, resp. v bodě $x = 2$ a vypočítáme je pomocí příkazu pbinom().

```
sum(dbinom(3:6, n, theta)) # 0,546076
pbinom(6, n, theta) - pbinom(2, n, theta) # 0,546076
```

Pravděpodobnost, že v rodině s dvanácti dětmi bude právě sedm chlapců, je 20,70 %. Pravděpodobnost, že v rodině s dvanácti dětmi budou nejvýše čtyři chlapci, je 15,89 %. Pravděpodobnost, že v rodině s dvanácti dětmi bude alespoň šest chlapců, je 66,36 %. Pravděpodobnost, že v rodině s dvanácti dětmi budou tři, čtyři, pět nebo šest chlapců, je 54,61 %.

★

Příklad 4.3. Neřešený příklad
V rámci studie poměru pohlaví u lidí (Geissler, 1889) bylo na základě záznamů z nemocnic v Sasku zaznamenáno rozložení počtu dívek v $M = 6115$ rodinách s dvanácti dětmi. Údaje ze studie jsou uvedeny v tabulce 4.2.

Tabulka 4.2: Počet dívek v rodinách s dvanácti dětmi

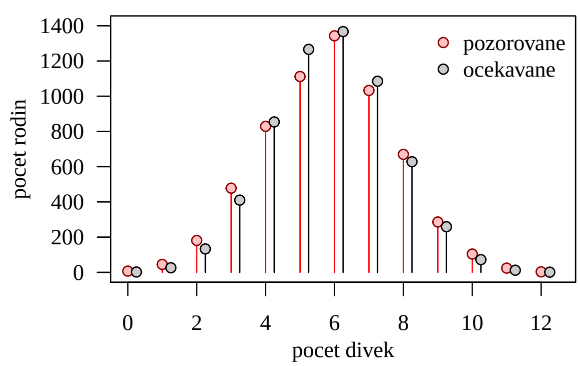
x	0	1	2	3	4	5	6	7	8	9	10	11	12	$\sum$
$m_{observed}$	7	45	181	478	829	1112	1343	1033	670	286	104	24	3	6115

Předpokládejme, že náhodná veličina X popisuje počet dívek v rodině s dvanácti dětmi. (a) Najděte rozložení, které dostatečně dobře vystihuje pravděpodobnostní chování náhodné veličiny X, a stanovte odhady parametrů tohoto rozložení; (b) ověřte, že nalezené rozložení je vhodné na popis náhodné veličiny X; vypočítejte očekávané absolutní četnosti za předpokladu nalezeného rozložení a graficky je porovnejte s pozorovanými absolutními četnostmi.

Výsledky: (a) $X \sim \text{Bi}(12; 0,4808)$; (b) očekávané absolutní četnosti viz tabulka 4.3; grafické porovnání pozorovaných a očekávaných absolutních četností viz obrázek 4.3.

Tabulka 4.3: Očekávaný počet dívek v rodinách s dvanácti dětmi

x	0	1	2	3	4	5	6	7	8	9	10	11	12	$\sum$
$m_{expected}$	2	26	133	410	854	1266	1367	1085	628	259	72	12	1	6115



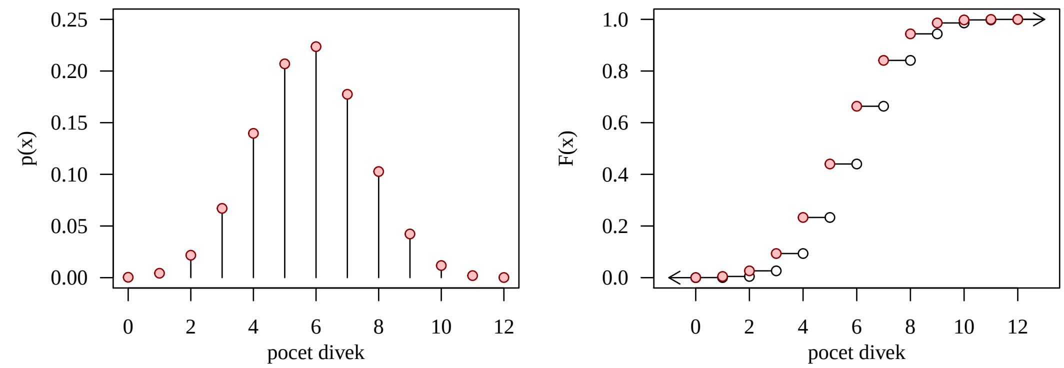
Obrázek 4.3: Porovnání pozorovaných absolutních četností a očekávaných absolutních četností za předpokladu rozložení Bi(12; 0, 4808)

★

Příklad 4.4. Neřešený příklad
Za předpokladu, že náhodná veličina X popisující počet dívek v rodině s dvanácti dětmi pochází z rozložení $\text{Bi}(12; 0,4808)$, (a) nakreslete graf pravděpodobnostní funkce $p(x)$ a graf distribuční funkce $F(x)$; (b) vypočítejte pravděpodobnost, že v rodině s dvanácti dětmi bude (i) šest až jedenáct dívek; (ii) alespoň osm dívek; (iii) právě čtyři dívky; (iv) nejvýše tři dívky.

Výsledky: (a) graf pravděpodobnostní funkce $p(x)$ viz obrázek 4.4 vlevo, graf distribuční funkce $F(x)$ viz obrázek 4.4 vpravo; (b-i) $\Pr(6 \leq X \leq 11) = 0,5598$; (b-ii) $\Pr(X \geq 8) = 0,1589$; (b-iii) $\Pr(X = 4) = 0,1397$; (b-iv) $\Pr(X \leq 3) = 0,0934$.

★



Obrázek 4.4: Graf pravděpodobnostní funkce $p(x)$ (vlevo) a distribuční funkce $F(x)$ (vpravo) rozložení $\mathrm{Bi}(12; 0,4808)$

4.2.3 Hypergeometrické rozložení

V souboru N prvků je M prvků označeno. Náhodně vybereme k prvků bez vracení. Náhodná veličina X udává počet vybraných označených prvků. Píšeme $X \sim \mathrm{Hg}(N, M, k)$.

$$p(x) = \begin{cases} \frac{\binom{M}{x}\binom{N-M}{k-x}}{\binom{N}{k}} \text{ pro } x = \max\{0, M-N+k\}, \ldots, \min\{M, k\}, \\ 0 \text{ jinak.} \end{cases}$$

Hypergeometrické rozložení se pojí s výběry bez vracení (bez opakování).

Vzorce pro výpočty pravděpodobností pomocí hypergeometrického rozložení

- Pravděpodobnost, že ve výběru je právě x označených prvků:
  $\Pr(X = x) = p(x) = \frac{\binom{M}{x}\binom{N-M}{k-x}}{\binom{N}{k}}$.
  K výpočtu v softwaru R slouží funkce `dhyper(x, m, n, k)`, kde
  - x ... počet označených prvků ve výběru,
  - m ... počet označených prvků v souboru ($m = M$),
  - n ... počet neoznačených prvků v souboru ($n = N - M$),
  - k ... počet prvků ve výběru.
- Pravděpodobnost, že ve výběru je nejvýše x_1 označených prvků:
  $\Pr(X \leq x_1) = \sum_{x=\max\{0,M-N+k\}}^{x_1} p(x)$.
  K výpočtu v softwaru R slouží funkce `phyper(x₁, m, n, k)`.
- Pravděpodobnost, že ve výběru je alespoň x_0 označených prvků:
  $\Pr(X \geq x_0) = 1 - \Pr(X \leq x_0 - 1) = 1 - F(x_0 - 1) = 1 - \sum_{x=\max\{0,M-N+k\}}^{x_0-1} p(x)$.
  Výpočet lze v softwaru R provést takto: `1 − phyper(x₀ − 1, m, n, k)`.
- Pravděpodobnost, že ve výběru je alespoň x_0 a nejvýše x_1 označených prvků:
  $\Pr(x_0 \leq X \leq x_1) = F(x_1) - F(x_0 - 1) = \sum_{x=x_0}^{x_1} p(x)$.
  Výpočet lze v softwaru R provést takto: `phyper(x₁, m, n, k) − phyper(x₀ − 1, m, n, k)`.

Příklad 4.5. Řešený příklad
Podle údajů o počtu obyvatelstva v ČR získaných z webových stránek Českého statistického úřadu (www.czso.cz) měl Jihomoravský kraj ke dni 31. 12. 2019 celkem 1 191 989 obyvatel. Rozmístění obyvatel v jednotlivých okresech Jihomoravského kraje je k dispozici v tabulce 4.4.

Tabulka 4.4: Počet obyvatel v okresech Jihomoravského kraje k datu 31. 12. 2019

Okres	Blansko	Brno-město	Brno-venkov	Břeclav	Hodonín	Vyškov	Znojmo	$\sum$
Počet obyvatel	109 136	381 346	224 642	116 291	153 943	92 280	114 351	1 191 989

Předpokládejme, že chceme sestavit reprezentativní vzorek 10-ti obyvatel pocházejících z Jihomoravského kraje a že náhodná veličina X popisuje počet obyvatel z okresu Brno-venkov v reprezentativním vzorku. (a) Najděte rozložení, které dostatečně dobře vystihuje pravděpodobnostní chování náhodné veličiny X, a stanovte odhady parametrů tohoto rozložení; (b) nakreslete graf pravděpodobnostní funkce $p(x)$ a graf distribuční funkce $F(x)$.

Řešení příkladu 4.5

Z datové tabulky 4.4 vidíme, že v celém Jihomoravském kraji bylo k datu 31. 12. 2019 evidováno celkem $N = 1\,191\,989$ obyvatel, z nichž $M = 224\,642$ obyvatel pocházelo z okresu Brno-venkov. Z celého Jihomoravského kraje vybíráme (bez vracení) reprezentativní vzorek $k = 10$ obyvatel. Náhodná veličina X popisující počet obyvatel z okresu Brno-venkov v reprezentativním vzorku 10-ti obyvatel potom pochází z hypergeometrického rozložení s parametry N, M a k, tj. $X \sim \mathrm{Hg}(1\,191\,989, 224\,642, 10)$.

K vykreslení grafu pravděpodobnostní funkce $p(x)$ nejprve příkazem dhyper() vypočítáme hodnoty pravděpodobnostní funkce v bodech $x = 0, 1, \dots, 10$. Hodnoty pravděpodobnostní funkce potom zaznamenáme do grafu příkazem plot() s argumentem type = 'h' jako svislé úsečky, které následně zakončíme fialovými body pomocí příkazu points(). Výsledný graf je zobrazen na obrázku 4.5 vlevo.

```
N <- 1191989
M <- 224642
k <- 10
x <- 0:k
px <- dhyper(x, M, N - M, k)
plot(x, px, type = 'h', xlab = 'pocet obyvatel z okresu Brno-venkov',
     ylab = 'p(x)', las = 1)
points(x, px, pch = 21, col = 'purple4', bg = 'plum2')
```

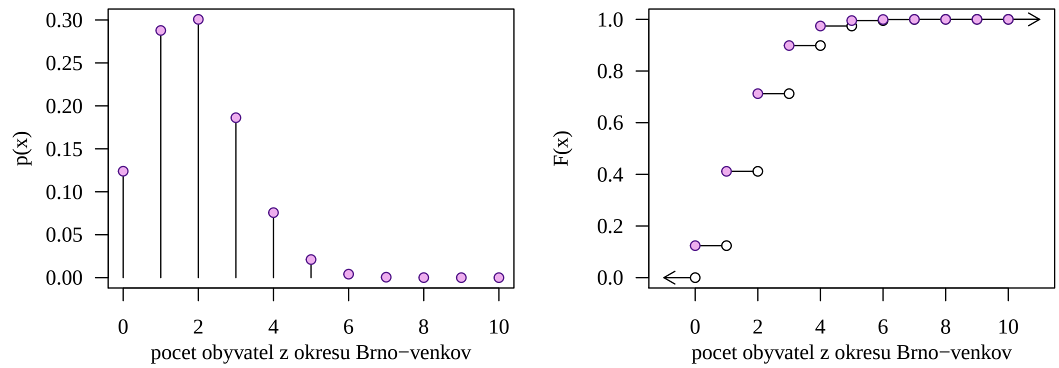
K vykreslení grafu distribuční funkce $F(x)$ nejprve prostřednictvím příkazu phyper() vypočítáme hodnoty distribuční funkce v bodech $x = 0, 1, \dots, 10$. Následně si příkazem plot() s argumentem type = 'n' připravíme prázdný graf, do kterého zaneseme hodnoty distribuční funkce $F(x)$ jako vodorovné úsečky pomocí příkazu segments(). Dále příkazem arrows() dokreslíme do grafu šipky reprezentující průběh distribuční funkce na intervalech $(-\infty; 0)$ a $\langle 10; \infty)$. Nakonec každou úsečku vizuálně zprava otevřeme prostřednictvím bílého bodu s černým obrysem a zleva uzavřeme prostřednictvím plného fialového bodu. Výsledný graf je zobrazen na obrázku 4.5 vpravo.

```
Fx <- phyper(x, M, N - M, k)
plot(x, Fx, type = 'n', xlim = c(-1, 11), ylim = c(0, 1),
     xlab = 'pocet obyvatel z okresu Brno-venkov', ylab = 'F(x)', las = 1)
segments(x, Fx, x + 1, Fx)
arrows(0, 0, -1, 0, length = 0.1)
arrows(10, 1, 11, 1, length = 0.1)
points(x, c(0, Fx[1:10]), pch = 21, col = 'black', bg = 'white')
points(x, Fx, pch = 21, col = 'purple4', bg = 'plum2')
```

★

Příklad 4.6. Řešený příklad

Za předpokladu, že náhodná veličina X popisující počet obyvatel z okresu Brno-venkov v reprezentativním vzorku 10-ti obyvatel z Jihomoravského kraje pochází z rozložení $\mathrm{Hg}(1\,191\,989, 224\,642, 10)$, vypočítejte pravděpodobnost, že v reprezentativním vzorku bude z okresu Brno-venkov (a) právě jeden obyvatel; (b) nejvýše dva obyvatelé; (c) alespoň sedm obyvatel; (d) pět až osm obyvatel.



Obrázek 4.5: Graf pravděpodobnostní funkce $p(x)$ (vlevo) a distribuční funkce $F(x)$ (vpravo) rozložení $\mathrm{Hg}(1\,191\,989, 224\,642, 10)$

Pravděpodobnost, že v reprezentativním vzorku 10-ti obyvatel Jihomoravského kraje bude z okresu Brno-venkov právě jeden obyvatel, tj. $\Pr(X = 1)$ odpovídá hodnotě pravděpodobnostní funkce $p(x)$ rozložení $\mathrm{Hg}(1\,191\,989, 224\,642, 10)$ v bodě $x = 1$. Tuto hodnotu vypočítáme pomocí příkazu dhyper().

```
dhyper(1, M, N - M, k) # 0,287746
```

Pravděpodobnost, že v reprezentativním vzorku 10-ti obyvatel Jihomoravského kraje budou z okresu Brno-venkov nejvýše dva obyvatelé, tj. $\Pr(X \leq 2)$ odpovídá hodnotě distribuční funkce $F(x)$ v bodě $x = 2$. Tuto hodnotu vypočítáme příkazem phyper().

```
phyper(2, M, N - M, k) # 0,7123524
```

Pravděpodobnost, že v reprezentativním vzorku 10-ti obyvatel Jihomoravského kraje bude z okresu Brno-venkov alespoň sedm obyvatel, tj. $\Pr(X \geq 7)$ vypočítáme tak, že od jedné odečteme pravděpodobnost, že v reprezentativním vzorku bude z okresu Brno-venkov nejvýše šest obyvatel. Tuto pravděpodobnost, odpovídající hodnotě distribuční funkce $F(x)$ v bodě $x = 6$, vypočítáme příkazem phyper().

```
1 - phyper(6, M, N - M, k) # 0,0005911696
```

Pravděpodobnost, že v reprezentativním vzorku 10-ti obyvatel Jihomoravského kraje bude z okresu Brno-venkov pět až osm obyvatel, tj. $\Pr(5 \leq X \leq 8)$ můžeme vypočítat dvěma způsoby. Prvním způsobem je, že sečteme hodnoty pravděpodobnostní funkce $p(x)$ v bodech $x = 5$, 6, 7 a 8. Tyto hodnoty vypočítáme pomocí příkazu dhyper(). Druhým způsobem je, že od pravděpodobnosti, že v reprezentativním vzorku bude z okresu Brno-venkov nejvýše osm obyvatel, odečteme pravděpodobnost, že v reprezentativním vzorku budou z okresu Brno-venkov nejvýše čtyři obyvatelé. Obě pravděpodobnosti jsou hodnotami distribuční funkce $F(x)$ v bodě $x = 8$, resp. v bodě $x = 4$ a vypočítáme je pomocí příkazu phyper().

```
sum(dhyper(5:8, M, N - M, k)) # 0,02575766
phyper(8, M, N - M, k) - phyper(4, M, N - M, k) # 0,02575766
```

Pravděpodobnost, že v reprezentativním vzorku 10-ti obyvatel Jihomoravského kraje bude z okresu Brno-venkov právě jeden obyvatel, je 28,77 %. Pravděpodobnost, že v reprezentativním vzorku budou z okresu Brno-venkov nejvýše dva obyvatelé, je 71,24 %. Pravděpodobnost, že v reprezentativním vzorku bude z okresu Brno-venkov alespoň sedm obyvatel, je 0,06 %. Pravděpodobnost, že v reprezentativním vzorku bude z okresu Brno-venkov pět až osm obyvatel, je 2,58 %.

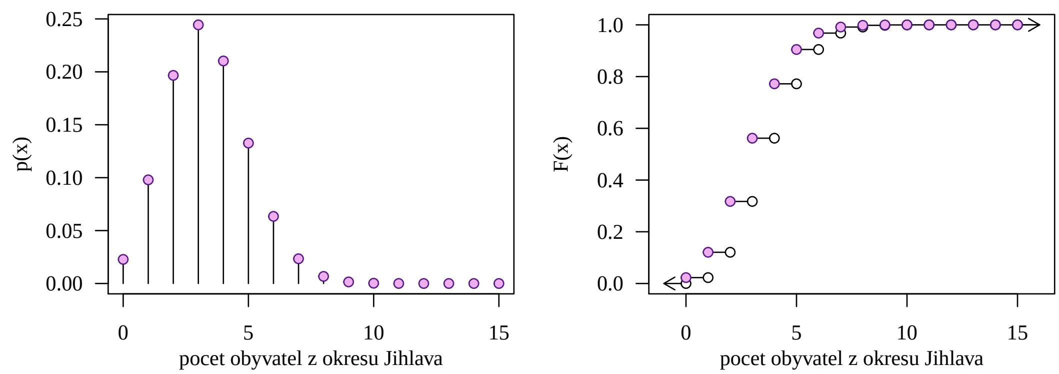
★

Příklad 4.7. Neřešený příklad
Podle údajů o počtu obyvatelstva v ČR získaných z webových stránek Českého statistického úřadu (www.czso.cz) měl kraj Vysočina ke dni 31. 12. 2019 celkem 509 813 obyvatel. Rozmístění obyvatel v jednotlivých okresech kraje Vysočina je k dispozici v tabulce 4.5.

Tabulka 4.5: Počet obyvatel v okresech kraje Vysočina k datu 31. 12. 2019

Okres	Havlíčkův Brod	Jihlava	Pelhřimov	Třebíč	Žďár nad Sázavou	$\sum$
Počet obyvatel	94 915	113 628	72 302	110 810	118 158	509 813

Předpokládejme, že chceme sestavit reprezentativní vzorek 15-ti obyvatel pocházejících z kraje Vysočina a že náhodná veličina X popisuje počet obyvatel z okresu Jihlava v tomto reprezentativním vzorku. (a) Najděte rozložení, které dostatečně dobře vystihuje pravděpodobnostní chování náhodné veličiny X, a stanovte odhady parametrů tohoto rozložení; (b) nakreslete graf pravděpodobnostní funkce $p(x)$ a graf distribuční funkce $F(x)$.
Výsledky: (a) $X \sim \mathrm{Hg}(509\,813, 113\,628, 15)$; (b) graf pravděpodobnostní funkce $p(x)$ viz obrázek 4.6 vlevo, graf distribuční funkce $F(x)$ viz obrázek 4.6 vpravo.



Obrázek 4.6: Graf pravděpodobnostní funkce $p(x)$ (vlevo) a distribuční funkce $F(x)$ (vpravo) rozložení $\mathrm{Hg}(509\,813, 113\,628, 15)$

★

Příklad 4.8. Neřešený příklad
Za předpokladu, že náhodná veličina X popisující počet obyvatel z okresu Jihlava v reprezentativním vzorku 15-ti obyvatel z kraje Vysočina pochází z rozložení $\mathrm{Hg}(509\,813, 113\,628, 15)$, vypočítejte pravděpodobnost, že v reprezentativním vzorku bude z okresu Jihlava (a) nejvýše pět obyvatel; (b) právě tři obyvatelé; (c) čtyři až devět obyvatel; (d) alespoň jeden obyvatel.
Výsledky: (a) $\Pr(X \leq 5) = 0,9047$; (b) $\Pr(X = 3) = 0,2444$; (c) $\Pr(4 \leq X \leq 9) = 0,4380$; (d) $\Pr(X \geq 1) = 0,9772$.

★

4.2.4 Poissonovo rozložení

Náhodná veličina X udává počet sledovaných událostí, které nastanou v jednotkovém časovém intervalu (resp. jednotkové oblasti), přičemž události nastávají náhodně, jednotlivě a vzájemně nezávisle. Parametr $\lambda > 0$ je střední hodnota počtu těchto událostí. Píšeme $X \sim \mathrm{Po}(\lambda)$.

$$p(x) = \begin{cases} \frac{\lambda^x}{x!} e^{-\lambda} \text{ pro } x = 0, 1, \ldots, \\ 0 \text{ jinak.} \end{cases}$$

Vzorce pro výpočty pravděpodobností pomocí Poissonova rozložení

- Pravděpodobnost, že nastane právě x sledovaných událostí:
  $\Pr(X = x) = p(x) = \frac{\lambda^x}{x!} e^{-\lambda}$.
  K výpočtu v softwaru R slouží funkce `dpois(x, λ)`.

- Pravděpodobnost, že nastane nejvýše x_1 sledovaných událostí:
  $\Pr(X \leq x_1) = F(x_1) = \sum_{x=0}^{x_1} p(x)$.
  K výpočtu v softwaru R slouží funkce `ppois(x₁, λ)`.

- Pravděpodobnost, že nastane alespoň x_0 sledovaných událostí:
  $\Pr(X \geq x_0) = 1 - \Pr(X \leq x_0 - 1) = 1 - F(x_0 - 1) = 1 - \sum_{x=0}^{x_0-1} p(x)$.
  Výpočet lze v softwaru R provést takto: `1 - ppois(x₀, λ)`.

- Pravděpodobnost, že nastane alespoň x_0 a nejvýše x_1 sledovaných událostí:
  $\Pr(x_0 \leq X \leq x_1) = F(x_1) - F(x_0 - 1) = \sum_{x=x_0}^{x_1} p(x)$.
  Výpočet lze v softwaru R provést takto: `ppois(x₁, λ) - ppois(x₀ - 1, λ)`.

Upozornění: Je-li počet pokusů n v bernoulliovské posloupnosti pokusů alespoň 30 a pravděpodobnost úspěchu ϑ v jednom pokusu je nejvýše $0,1$, lze pravděpodobnostní funkci rozložení $\mathrm{Bi}(n, \vartheta)$ nahradit pravděpodobnostní funkcí rozložení $\mathrm{Po}(\lambda)$, kde $\lambda = n\vartheta$.

Příklad 4.9. Řešený příklad

V rámci studie (von Bortkiewicz, 1898) byly zpracovány 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 tabulce 4.6.

Tabulka 4.6: Počet smrtelných úrazů v důsledku kopnutí koněm v pruských armádních jednotkách za jeden rok

x	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).

Předpokládejme, že náhodná veličina X popisuje počet smrtelných úrazů v důsledku kopnutí koněm v jedné pruské armádní jednotce za jeden rok. (a) Najděte rozložení, které dostatečně dobře vystihuje pravděpodobnostní chování náhodné veličiny X, a stanovte odhady parametrů tohoto rozložení; (b) ověřte, že nalezené rozložení je vhodné na popis náhodné veličiny X; vypočítejte očekávané absolutní četnosti za předpokladu nalezeného rozložení a graficky je porovnejte s pozorovanými absolutními četnostmi.

Řešení příkladu 4.9

Počet smrtelných úrazů v jedné armádní jednotce za jeden rok je diskrétní znak, k jeho popisu tedy použijeme diskrétní náhodnou veličinu. Všimněme si, že v tabulce 4.6 je uvedeno, v kolika armádních jednotkách nenastal žádný úraz, nastal jeden úraz, ..., nastalo pět a více úrazů. Počet úrazů tedy není shora nijak omezen. O náhodné veličině X tedy předpokládáme, že pochází z Poissonova rozložení s parametrem λ, kde λ je střední počet smrtelných úrazů v jedné armádní jednotce za jeden rok. Hodnotu parametru λ odhadneme jako podíl počtu všech smrtelných úrazů ku celkovému počtu všech armádních jednotek za dvacet let.

$$\widehat{\lambda} = \frac{\text{počet smrtelných úrazů}}{\text{počet armádních jednotek za 20 let}} = \frac{\sum_{x=0}^{n} x m_{observed}}{M} = \frac{122}{200} = 0,61.$$

```
x <- 0:5
m.obs <- c(109, 65, 22, 3, 1, 0)
M <- sum(m.obs) # 200
lambda <- sum(x * m.obs) / M # 0,61
```

Odhad parametru $\widehat{\lambda} = 0,61$. Střední počet smrtelných úrazů v důsledku kopnutí koněm v jedné pruské armádní jednotce za jeden rok je $0,61$. O náhodné veličině X předpokládáme, že pochází z Poissonova rozložení s parametrem $\lambda = 0,61$, tj. $X \sim \text{Po}(0,61)$.

Nyní zbývá ověřit, zda Poissonovo rozložení dostatečně dobře popisuje reálná data, a to tak, že porovnáme očekávané absolutní četnosti s pozorovanými absolutními četnostmi. Za předpokladu, že $X \sim \text{Po}(0,61)$, vypočítáme nejprve příkazem dpois() pravděpodobnosti, že v jedné armádní jednotce došlo během jednoho roku k 0, 1, 2, 3 a 4 úrazům. Dále pomocí příkazu 1 − ppois() vypočítáme pravděpodobnost, že v jedné armádní jednotce došlo během jednoho roku k pěti a více úrazům. Získané pravděpodobnosti vynásobíme počtem armádních jednotek za dvacet let (tj. $M = 200$) a výsledky zaokrouhlíme na celá čísla. Tím získáme očekávané absolutní četnosti armádních jednotek, ve kterých v průběhu jednoho roku došlo k $0, 1, \ldots, 5+$ smrtelným úrazům. Očekávané absolutní četnosti vložíme společně s pozorovanými absolutními četnostmi do souhrnné tabulky.

```
m.exp <- round(c(dpois(0:4, lambda), 1 - sum(dpois(0:4, lambda))) * M)
tab <- data.frame(rbind(m.obs, m.exp))
names(tab) <- c(0:4, '5+')
```

```
        0  1  2 3 4 5+
m.obs 109 65 22 3 1  0
m.exp 109 66 20 4 1  0
```

Za předpokladu, že $X \sim \text{Po}(0,61)$, nedojde ve 109 armádních jednotkách během jednoho roku k žádnému smrtelnému úrazu v důsledku kopnutí koněm, v 66 jednotkách dojde k jednomu smrtelnému úrazu, ve 20 jednotkách dojde ke dvěma smrtelným úrazům, apod.

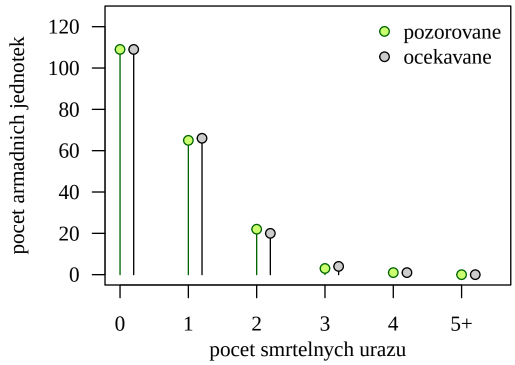
Pozorované a očekávané absolutní četnosti nyní vzájemně graficky porovnáme (viz obrázek 4.7). Příkazem plot() s argumentem type = 'h' vykreslíme graf s pozorovanými absolutními četnostmi reprezentovanými svislými zelenými úsečkami. Argumentem axes = F zakážeme vykreslení měřítka osy x a osy y. Obě měřítka doplníme do grafu zvlášť. Měřítko osy x vykreslíme příkazem axis() s argumentem side = 1. Popisky měřítka osy x, tj. $0, 1, \ldots, 5+$ specifikujeme pomocí argumentu labels. Měřítko osy y vykreslíme příkazem axis() s argumentem side = 2. Svislé zelené úsečky zakončíme zelenými body pomocí příkazu points(). Dále do grafu dokresíme svislé černé úsečky zakončené šedými body reprezentující očekávané absolutní četnosti, a to ve vzdálenosti $0,2$ od úseček a bodů reprezentujících pozorované absolutní četnosti. Nakonec příkazem legend() doplníme do grafu legendu.

```
plot(x, m.obs, type = 'h', xlim = c(0, 5.5), ylim = c(0, 125),
     xlab = 'x', ylab = 'pocet armadnich jednotek', col = 'darkgreen', axes = F)
box(bty = 'o')
axis(side = 1, at = x, labels = c(0:4, '5+'))
axis(side = 2, las = 1)
points(x, m.obs, pch = 21, col = 'darkgreen', bg = 'darkolivegreen1')
lines (x + 0.2, m.exp, type = 'h', col = 'black')
points(x + 0.2, m.exp, pch = 21, col = 'black', bg = 'grey80')
legend('topright', pch = c(21, 21), col = c('darkgreen', 'black'), bty = 'n',
       pt.bg = c('darkolivegreen1', 'grey80'), legend = c('pozorovane', 'ocekavane'))
```

Z obrázku 4.7 vidíme, že pozorované a očekávané absolutní četnosti jsou téměř identické. Grafická vizualizace tedy podporuje náš závěr, že náhodná veličina X popisující počet smrtelných úrazů v důsledku kopnutí koněm v jedné pruské armádní jednotce v průběhu jednoho roku pochází z Poissonova rozložení s parametrem $\lambda = 0,61$. ★

Příklad 4.10. Řešený příklad

Za předpokladu, že náhodná veličina X popisující počet smrtelných úrazů v důsledku kopnutí koněm v jedné pruské armádní jednotce v průběhu jednoho roku pochází z rozložení $\text{Po}(0,61)$, (a) nakreslete graf pravděpodobnostní funkce $p(x)$ a graf distribuční funkce $F(x)$; (b) vypočítejte pravděpodobnost, že během jednoho roku dojde v jedné armádní jednotce (i) k právě dvěma smrtelným úrazům; (ii) k nejvýše jednomu smrtelnému úrazu; (iii) k alespoň třem smrtelným úrazům; (iv) ke dvěma nebo třem smrtelným úrazům.



Obrázek 4.7: Porovnání pozorovaných absolutních četností a očekávaných absolutních četností za předpokladu rozložení $\mathrm{Po}(0,61)$

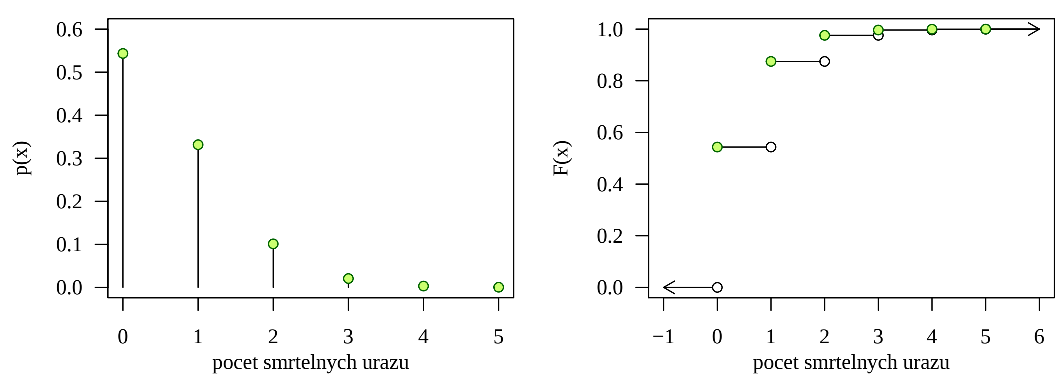
Řešení příkladu 4.10

K vykreslení grafu pravděpodobnostní funkce $p(x)$ nejprve příkazem dpois() vypočítáme hodnoty pravděpodobnostní funkce v bodech $x = 0, 1, \ldots, 5$. Graf pravděpodobnostní funkce (viz obrázek 4.8 vlevo) potom vykreslíme analogicky jako v příkladech 4.2 a 4.5.

```
x <- 0:5
px <- dpois(x, lambda)
plot(x, px, type = 'h', ylim = c(0, 0.6), xlab = 'pocet smrtelnych urazu',
     ylab = 'p(x)', las = 1)
points(x, px, pch = 21, col = 'darkgreen', bg = 'darkolivegreen1')
```

K vykreslení grafu distribuční funkce $F(x)$ nejprve příkazem ppois() vypočítáme hodnoty distribuční funkce v bodech $x = 0, 1, \ldots, 4$. Hodnotu distribuční funkce v bodě $x = 5$ aproximujeme hodnotou 1. Graf distribuční funkce (viz obrázek 4.8 vpravo) vykreslíme analogicky jako v příkladech 4.2 a 4.5.

```
Fx <- c(ppois(0:4, lambda), 1)
plot(x, Fx, type = 'n', xlim = c(-1, 6), ylim = c(0, 1),
     xlab = 'pocet smrtelnych urazu', ylab = 'F(x)', las = 1)
segments(x, Fx, x + 1, Fx)
arrows(0, 0, -1, 0, length = 0.1)
arrows(5, 1, 6, 1, length = 0.1)
points(x, c(0, Fx[1:5]), pch = 21, col = 'black', bg = 'white')
points(x, Fx, pch = 21, col = 'darkgreen', bg = 'darkolivegreen1')
```



Obrázek 4.8: Graf pravděpodobnostní funkce $p(x)$ (vlevo) a distribuční funkce $F(x)$ (vpravo) rozložení $\mathrm{Po}(0,61)$

Nyní se zaměříme na výpočet pravděpodobností za předpokladu Poissonova rozložení s parametrem $\lambda = 0,61$. Pravděpodobnost, že v jedné armádní jednotce dojde během jednoho roku k právě dvěma smrtelným úrazům, tj. $\Pr(X = 2)$ odpovídá hodnotě pravděpodobnostní funkce $p(x)$ rozložení $\mathrm{Po}(0,61)$ v bodě $x = 2$. Tuto hodnotu vypočítáme příkazem dpois().

```
lambda <- 0.61
dpois(2, lambda) # 0,1010904
```

Pravděpodobnost, že v jedné armádní jednotce dojde během jednoho roku k nejvýše jednomu smrtelnému úrazu, tj. $\Pr(X \leq 1)$ odpovídá hodnotě distribuční funkce $F(x)$ v bodě $x = 1$. Tuto hodnotu vypočítáme příkazem ppois().

```
ppois(1, lambda) # 0,8747949
```

Pravděpodobnost, že v jedné armádní jednotce dojde během jednoho roku k alespoň třem smrtelným úrazům, tj. $\Pr(X \geq 3)$ vypočítáme tak, že od jedné odečteme pravděpodobnost, že v jedné armádní jednotce dojde k nejvýše dvěma smrtelným úrazům. Tuto pravděpodobnost, odpovídající hodnotě distribuční funkce $F(x)$ v bodě $x = 2$, vypočítáme příkazem ppois().

```
1 - ppois(2, lambda) # 0,02411467
```

Pravděpodobnost, že v jedné armádní jednotce dojde během jednoho roku ke dvěma nebo třem smrtelným úrazům, můžeme vypočítat dvěma způsoby. Prvním způsobem je, že sečteme hodnoty pravděpodobnostní funkce $p(x)$ v bodech $x = 2$ a $x = 3$. Tyto hodnoty vypočítáme příkazem dpois() a sečteme příkazem sum(). Druhým způsobem je, že od pravděpodobnosti, že v jedné armádní jednotce dojde během jednoho roku k nejvýše třem smrtelným úrazům, odečteme pravděpodobnost, že v jedné armádní jednotce dojde během jednoho roku k nejvýše jednomu smrtelnému úrazu. Obě pravděpodobnosti jsou hodnotami distribuční funkce $F(x)$ v bodě $x = 3$, resp. v bodě $x = 1$ a vypočítáme je pomocí příkazu ppois().

```
sum(dpois(2:3, lambda)) # 0,1216455
ppois(3, lambda) - ppois(1, lambda) # 0,1216455
```

Pravděpodobnost, že v jedné armádní jednotce dojde během jednoho roku k právě dvěma smrtelným úrazům v důsledku kopnutí koněm, je 10,11 %. Pravděpodobnost, že dojde k nejvýše jednomu smrtelnému úrazu, je 87,48 %. Pravděpodobnost, že dojde k alespoň třem smrtelným úrazům, je 2,41 %. Pravděpodobnost, že dojde ke dvěma nebo třem smrtelným úrazům, je 12,16 %.

★

Příklad 4.11. Neřešený příklad

V rámci studie (Greenwood a Yule, 1920) byla publikována data z výstupní zprávy o výskytu průmyslových havárií (v Londýně, 1919), která obsahují počty úrazů u dělnic v továrně při výrobě dělostřeleckých granátů v průběhu pěti týdnů. Celkem bylo do studie zahrnuto $M = 647$ dělnic. Údaje ze studie jsou uvedeny v tabulce 4.7.

Tabulka 4.7: Počet úrazů u dělnic při výrobě dělostřeleckých granátů

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

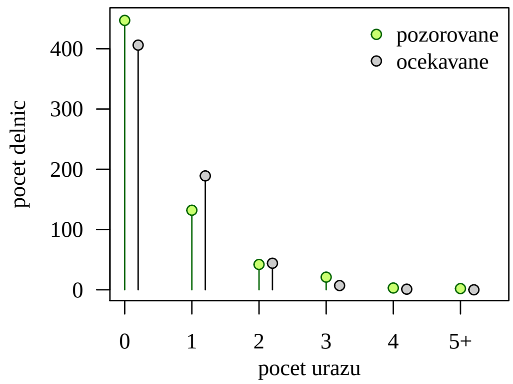
Předpokládejme, že náhodná veličina X popisuje počet úrazů u jedné dělnice při výrobě dělostřeleckých granátů v průběhu pěti týdnů. (a) Najděte rozložení, které dostatečně dobře vystihuje pravděpodobnostní chování náhodné veličiny X, a stanovte odhady parametrů tohoto rozložení; (b) ověřte, že nalezené rozložení je vhodné na popis náhodné veličiny X; vypočítejte očekávané absolutní četnosti za předpokladu nalezeného rozložení a graficky je porovnejte s pozorovanými absolutními četnostmi.

Výsledky: (a) $X \sim \mathrm{Po}(0,4652)$; (b) očekávané absolutní četnosti viz tabulka 4.8; grafické porovnání pozorovaných a očekávaných absolutních četností viz obrázek 4.9.

★

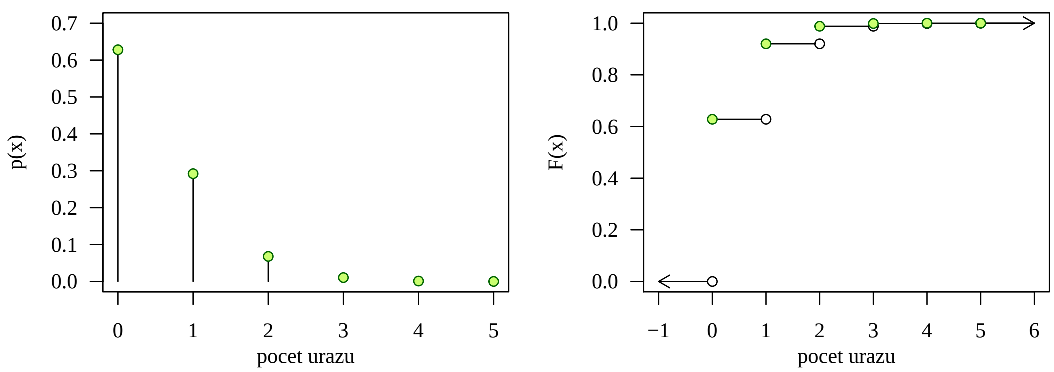
Tabulka 4.8: Očekávaný počet úrazů u dělnic při výrobě dělostřeleckých granátů

x	0	1	2	3	4	5+	$\sum$
$m_{expected}$	406	189	44	7	1	0	647



Obrázek 4.9: Porovnání pozorovaných absolutních četností a očekávaných absolutních četností za předpokladu rozložení $\mathrm{Po}(0,4652)$

Příklad 4.12. Neřešený příklad
Za předpokladu, že náhodná veličina X popisující počet úrazů u jedné dělnice při výrobě dělostřeleckých granátů v průběhu pěti týdnů pochází z rozložení $\mathrm{Po}(0,4652)$, (a) nakreslete graf pravděpodobnostní funkce $p(x)$ a graf distribuční funkce $F(x)$; (b) vypočítejte pravděpodobnost, že v průběhu pěti týdnů dojde u jedné dělnice (i) k alespoň dvěma úrazům; (ii) k jednomu nebo dvěma úrazům; (iii) k nejvýše třem úrazům; (iv) k žádnému úrazu.
Výsledky: (a) graf pravděpodobnostní funkce $p(x)$ viz obrázek 4.10 vlevo, graf distribuční funkce $F(x)$ viz obrázek 4.10 vpravo; (b-i) $\Pr(X \geq 2) = 0,0798$; (b-ii) $\Pr(1 \leq X \leq 2) = 0,3601$; (b-iii) $\Pr(X \leq 3) = 0,9987$; (b-iv) $\Pr(X = 0) = 0,6280$.



Obrázek 4.10: Graf pravděpodobnostní funkce $p(x)$ (vlevo) a distribuční funkce $F(x)$ (vpravo) rozložení $\mathrm{Po}(0,4652)$

★