4 Diskrétní náhodné veličiny

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

- Bernoulliho pokusy $X_1, \ldots, X_N$:
  - $X_i = 1$ ... 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.
  - $\sum_{i=1}^{N} X_i = X \sim \mathrm{Bin}(N, p)$.
  - $\boldsymbol{\theta} = (N, p)$
  - pravděpodobnostní funkce:
$$p(x) = \binom{N}{x} p^x (1-p)^{N-x} \qquad x = 0, 1, \ldots, N;$$
  - vlastnosti: $\mathrm{E}[X] = Np$; $\mathrm{Var}[X] = Np(1-p)$
  - dbinom(x, N, p), pbinom(x, N, p)

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

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

Příklad 4.1. Výpočet parametru p binomického rozdělení
Předpokládejme, že náhodná veličina X popisující počet chlapců v rodinách s dvanácti dětmi pochází z binomického rozdělení s parametrem $N = 12$. Vypočítejte odhad pravděpodobnosti výskytu chlapců v rodinách s dvanácti dětmi.

Řešení příkladu 4.1
Pravděpodobnost p výskytu chlapců v rodinách s dvanácti dětmi odhadneme pomocí vzorce

$$\widehat{p} = \frac{\text{počet narozených chlapců}}{\text{celkový počet narozených dětí}} = \frac{\sum_{n=0}^{N} n m_{observed}}{NM}. \tag{4.1}$$

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

```
[1] 0.5192
```

Interpretace výsledků: Pravděpodobnost výskytu chlapců v rodinách s dvanácti dětmi je .... (............................ %). ★

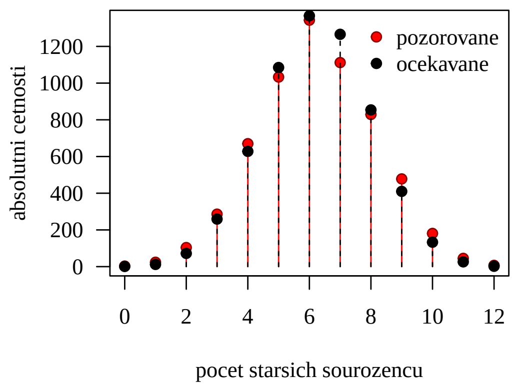
Příklad 4.2. Pozorované a očekávané početnosti v binomickém rozdělení
Za předpokladu, že počet chlapců v rodinách s dvanácti dětmi pochází z binomického rozdělení s parametry $N =$ .... a $p =$ .... odhadněte očekávané početnosti chlapců v rodinách s dvanácti dětmi a porovnejte je s pozorovanými početnostmi.

Řešení příkladu 4.2

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

```
      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
```

```
par(mar = c(4, 4, 1, 1), family = 'Times')
plot(0:12, m.obs, type = 'h', col = 'red',
     xlab = 'pocet starsich sourozencu', ylab = 'absolutni cetnosti', las = 1)
points(0:12, m.obs, pch = 21, col = 'darkred', bg = 'red')
lines (0:12, m.exp, type = 'h', lty = 2, col = 'black')
points(0:12, m.exp, pch = 21, col = 'black', bg = 'black')
legend('topright', pch = c(21, 21), col = c('darkred', 'black'), pt.bg = c('red', 'black'),
       legend = c('pozorovane', 'ocekavane'), bty = 'n')
```



★

Příklad 4.3. Výpočet pravděpodobností za předpokladu binomického rozdělení
Za předpokladu, že náhodná veličina X popisující počet chlapců v rodinách s dvanácti dětmi pochází z binomického rozdělení s parametry $N =$ .... a $p =$ .... vypočítejte pravděpodobnost, že v rodině s dvanácti dětmi bude

a. právě devět chlapců,

b. nejvýše čtyři chlapci,

c. alespoň osm chlapců,

d. čtyři, pět, šest, nebo sedm chlapců.

Řešení příkladu 4.3

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

```
[1] 0.067
```

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

```
[1] 0.1589
```

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

```
[1] 0.2331
```

```
#round(sum(dbinom(8 : 12, N, p)), 4)
round(sum(dbinom(4 : 7, N, p)), 4)
```

```
[1] 0.7108
```

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

Interpretace výsledků: Pravděpodobnost, že v rodině bude právě devět chlapců, je ....%. Pravděpodobnost, že v rodině budou nejvýše čtyři chlapci, je ....%. Pravděpodobnost, že v rodině bude alespoň osm chlapců, je ....%. Pravděpodobnost, že v rodině bude čtyři, pět, šest, nebo sedm chlapců, je ....%. ★

Příklad 4.4. Graf pravděpodobnostní a distribuční funkce binomického rozdělení

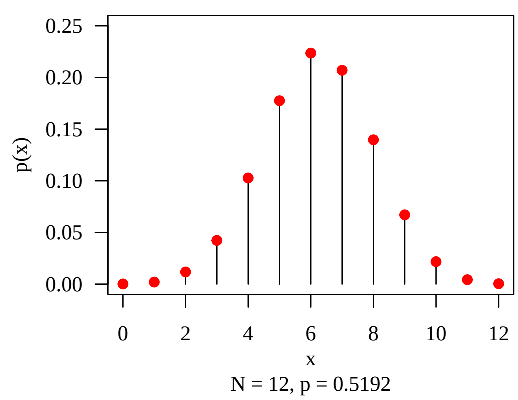
Nakreslete graf pravděpodobnostní funkce a graf distribuční funkce binomického rozdělení $\mathrm{Bin}(N, p)$, kde $N = 12$ a $p = 0.5192$.

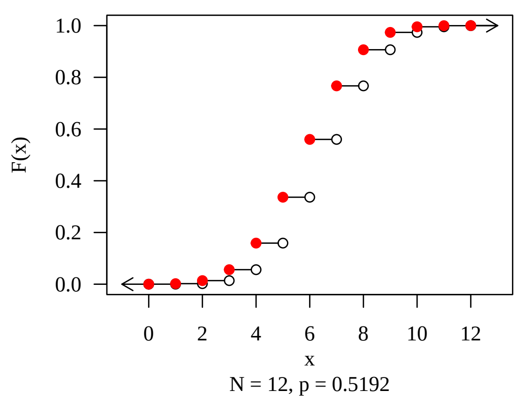
Řešení příkladu 4.4

```
N <- 12
x <- 0 : N
p <- 0.5192
px <- dbinom(x, N, p)
Fx <- pbinom(x, N, p)

par(mar = c(4, 4, 1, 1), family = 'Times')
plot(x, px, type = 'h', ylim = c(0, 0.25), xlab = '', ylab = 'p(x)', las = 1)
mtext('x', side = 1, line = 2)
mtext('N = 12, p = 0.5192', side = 1, line = 3)
points(x, px, col = 'red', pch = 19)

plot(x, Fx, xlab = '', type = 'n', xlim = c(-1, 13), ylim = c(0, 1),
     ylab = 'F(x)', las = 1)
segments(x, Fx, x + 1, Fx)                    # vodorovne cary
arrows(0, 0, -1, 0, length = 0.1)             # dolni sipka
arrows(N, 1, 13, 1, length = 0.1)             # horni sipka
points(x, c(0, Fx[1:N]), pch = 21, col = 'black', bg = 'white') # prazdne body
points(x, Fx, pch = 19, col = 'red')          # plne body
mtext('x', side = 1, line = 2)
mtext('N = 12, p = 0.5192', side = 1, line = 3)
```





★

Alternativní rozdělení Alt(p)

- $X\ldots$ výskyt sledované události v **jednom** Bernoulliho pokusu, přičemž pravděpodobnost nastání události v tomto pokusu je vyjádřena parametrem p.
- speciální případ binomického rozdělení, kde $N = 1$, tj. $X \sim \mathrm{Alt}(p) \simeq X \sim \mathrm{Bin}(1, p)$
- $X \sim \mathrm{Alt}(p)$.
- $\boldsymbol{\theta} = (p)^T$
- pravděpodobnostní funkce
$$p(x) = p^x(1-p)^{1-x} \qquad x = 0, 1; \tag{4.2}$$
- vlastnosti: $\mathrm{E}[X] = p$; $\mathrm{Var}[X] = p(1-p)$
- dbinom(x, 1, p), pbinom(x, 1, p)

Dataset: 09-one-sample-probability-sutmet.txt
Datový soubor 09-one-sample-probability-sutmet.txt obsahuje údaje o výskytu epigenetického znaku *sutura metopica* na lebkách Ainů z ostrova Hokaido. Údaje o výskytu epigenetického znaku jsou 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

Příklad 4.5. Popis reálné situace pomocí alternativního rozdělení
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 4.5
Náhodná veličina X 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. Náhodná veličina X tedy pochází z .... rozdělení, tj. $X \sim$ .... Odhad parametru p spočítáme jako podíl počtu výskytů epigenetického znaku *sutura metopica* na lebkách Ainů ku celkovému počtu zkoumaných lebek.

```
p <- 6 / 184
round(p, 4)
```

```
[1] 0.0326
```

Interpretace výsledků: Výskyt epigenetického znaku *sutura metopica* na lebkách Ainů z ostrova Hokaido pochází z .... rozdělení s parametrem $p =$ .... Pravděpodobnost výskytu epigenetického znaku *sutura metopica* u populace Ainů je ....%. ★

Příklad 4.6. 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 $p =$................, vypočítejte

a. 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. pravděpodobnost nejvýše jednoho výskytu epigenetického znaku *sutura metopica.*

Řešení příkladu 4.6

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

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

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

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

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

Příklad 4.7. 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.

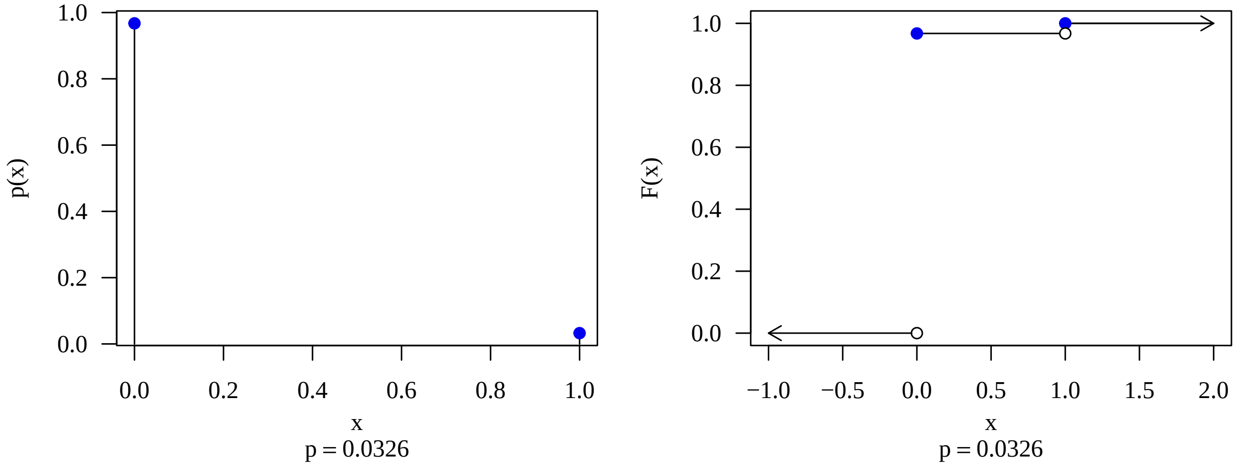
Řešení příkladu 4.7

```
N <- 1
x <- 0:1
px <- dbinom(0:1, 1, p)
Fx <- pbinom(0:1, 1, p)

par(mar = c(4, 4, 1, 1), family = 'Times')
# Graf pravdepodobnostni funkce
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)

# Graf distribucni funkce
plot(x, x, 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 1: Pravděpodobnostní funkce (vlevo) a distribuční funkce (vpravo) alternativního rozdělení Alt(0.03261)

4.2 Poissonovo rozdělení Poiss(λ)

- 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 $\lambda > 0$.
- $X \sim \text{Poiss}(\lambda)$
- $\theta = \lambda$
- pravděpodobnostní funkce:
$$p(x) = \frac{\lambda^x}{x!} e^{-\lambda} \; x = 0, 1, \ldots;$$
- vlastnosti: $E[X] = \lambda$; $\text{Var}[X] = \lambda$
- dpois(x, lambda), ppois(x, lambda)

Příklad 4.8. Výpočet parametru λ Poissonova rozdělení
Načtete datový soubor 17-anova-newborns-2.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 4.8
Střední hodnotu počtu starších sourozenců odhadneme pomocí vzorce

$$\lambda = \frac{\text{počet starších sourozenců}}{\text{počet novorozenců}} = \frac{\sum_{i=1}^{N} x_i}{N}. \tag{4.3}$$

```
data <- read.delim('00-Data//17-anova-newborns-2.txt')
data <- na.omit(data)
prch <- data$prch.N
N <- length(prch) # 1381
(lambda <- sum(prch) / N) # 0.9427951
```

```
[1] 0.9427951
```

Interpetace výsledků: Střední hodnota počtu starších sourozenců novorozenců v datovém souboru λ =.........
....
★

Příklad 4.9. Porovnání pozorovaných a očekávaných početností v Poissonově rozdělení
Za předpokladu, že počet starších sourozenců novorozenců pochází z Poissonova rozdělení s parametrem λ =...........

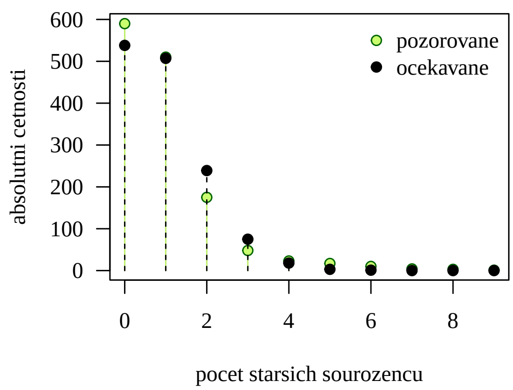
.... odhadněte očekávané početnosti starších sourozenců a porovnejte je s pozorovanými početnostmi.

Řešení příkladu 4.9

```
m.obs <- data.frame(table(prch))$Freq
m.exp <- round(c(dpois(0:8, lambda), 1 - ppois(8, lambda)) * N)
tab <- data.frame(rbind(m.obs, m.exp))
names(tab) <- 0:9
tab
```

```
        0   1   2  3  4  5  6 7 8 9
m.obs 590 510 175 48 23 17 10 4 3 1
m.exp 538 507 239 75 18  3  1 0 0 0
```

```
par(mar = c(4, 4, 1, 1), family = 'Times')
plot(0:9, m.obs, type = 'h', col = 'darkolivegreen2',
     xlab = 'pocet starsich sourozencu', ylab = 'absolutni cetnosti', las = 1)
points(0:9, m.obs, pch = 21, col = 'darkgreen', bg = 'darkolivegreen1')
lines(0:9, m.exp, type = 'h', lty = 2, col = 'black')
points(0:9, m.exp, pch = 21, col = 'black', bg = 'black')
legend('topright', pch = c(21, 21), col = c('darkgreen', 'black'), pt.bg = c('darkolivegreen1', 'black'),
       legend = c('pozorovane', 'ocekavane'), bty = 'n')
```



Obrázek 2: Porovnání pozorovaných a očekávaných početností v Poissonově rozdělení

★

Příklad 4.10. Výpočet pravděpodobností za předpokladu Poissonova rozdělení
Za předpokladu, že data pochází z Poissonova rozdělení s parametrem $\lambda =$ .... určete pravděpodobnost, novorozenec má

a. dva, tři nebo čtyři starší sourozence,

b. alespoň čtyři starší sourozence,

c. nejvýše dva starší sourozence,

d. právě jednoho staršího sourozence.

Řešení příkladu 4.10

```
sum(dpois(2:4, lambda))
```

```
[1] 0.2403526
```

```
1 - ppois(3, lambda)
```

```
[1] 0.01567936
```

```
ppois(2, lambda)
```

```
[1] 0.9299142
```

```
dpois(1, lambda)
```

```
[1] 0.3672541
```

Interpretace výsledů: Pravděpodobnost, že novorozenec má dva, tři nebo čtyři starší sourozence je ....
..............%. Pravděpodobnost, že novorozenec má alespoň čtyři starší sourozence je ....%. Pravděpodobnost, že novorozenec má nejvýše dva starší sourozence je ....%. Pravděpodobnost, že novorozenec má jednoho staršího sourozence je ....%. ★

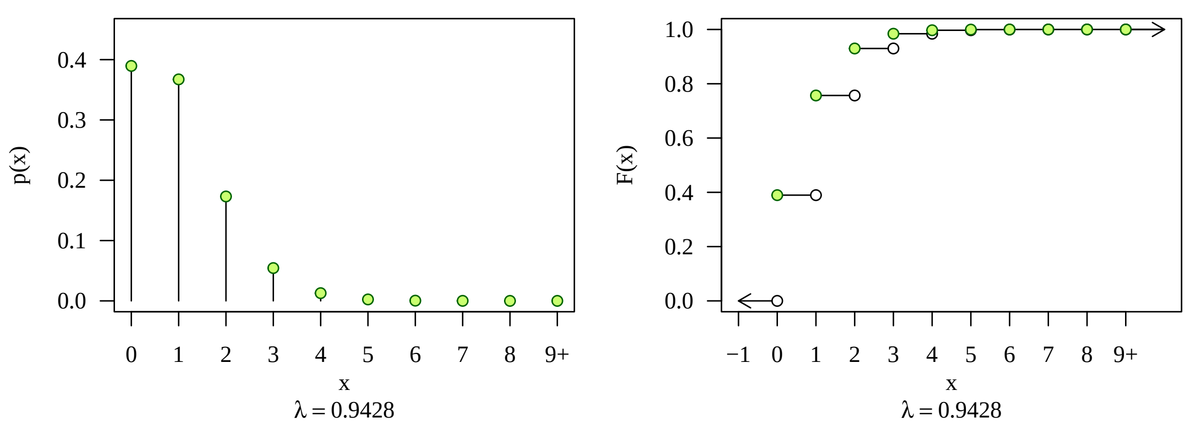
Příklad 4.11. Graf pravděpodobnostní a distribuční funkce Poissonova rozdělení
Nakreslete graf pravděpodobnostní a distribuční funkce Poissonova rozdělení Poiss(0.9428) v hodnotách $x = 0, 1, 2, 3, 4, 5, 6, 7, 8,$ a $x \geq 9$.

Řešení příkladu 4.11

```
N <- 9
x <- 0:N
px <- c(dpois(0:8, lambda), 1 - sum(dpois(0:8, lambda)) )
Fx <- c(ppois(0:8, lambda), 1)

par(mar = c(5, 4, 1, 1), family = 'Times')
plot(x, px, type = 'h', ylim = c(0, 0.45), xlab = '', ylab = 'p(x)', axes = F)
points(x, px, pch = 21, col = 'darkgreen', bg = 'darkolivegreen1')
box(bty = 'o')
axis(1, 0:9, c(0:8, '9+'))
axis(2, las = 1)
mtext('x', side = 1, line = 2)
mtext(bquote(paste(lambda == 0.9428)), side = 1, line = 3)

plot(x, Fx, type = 'n', xlim = c(-1, N + 1), ylim = c(0, 1),
     xlab = '', ylab = 'F(x)', axes = F)
box(bty = 'o')
axis(1, -1:9, c(-1:8, '9+'))
axis(2, las = 1)
segments(x, Fx, x + 1, Fx)                    # vodorovne cary
arrows(0, 0, -1, 0, length = 0.1)           # dolni sipka
arrows(9, 1, 10, 1, length = 0.1)             # horni sipka
points(x, c(0, Fx[1:9]), pch = 21, col = 'black', bg = 'white', cex = 1) # prazdne body
points(x, Fx, pch = 21, col = 'darkgreen', bg = 'darkolivegreen1')  # plne body
mtext('x', side = 1, line = 2)
mtext(bquote(paste(lambda == 0.9428)), side = 1, line = 3)
```



Obrázek 3: Pravděpodobnostní a distribuční funkce Poissonova rozdělení

★