3 Negativně binomické rozdělení $\mathrm{NegBin}(k, p)$

- Bernoulliho pokusy X_1, X_2, ...
  - $X_i = 1$ ... událost nastala (úspěch); $X_i = 0$ ... událost nenastala (neúspěch); $i = 1, 2, \ldots$
  - $\Pr(X_i = 1) = p$
  - $\Pr(X_i = 0) = 1 - p = q$
- X ... počet nastalých úspěchů předcházejících předem stanovenému počtu k neúspěchů
- $X \sim \mathrm{NegBin}(k, p)$
- $\boldsymbol{\theta} = (k, p)$
- pravděpodobnostní funkce

$$p(x) = \binom{x+k-1}{x} p^x (1-p)^k \qquad x = 1, 2, \ldots$$

- vlastnosti: $\mathrm{E}[X] = \dfrac{kp}{1-p}$; $\mathrm{Var}[X] = \dfrac{kp}{(1-p)^2}$
- dnbinom(n, k, 1 - p), pnbinom(n, k, 1 - p), rnbinom(M, k, 1 - p)
- Data:
  - **Dataset 2: 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.

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

Příklad 3.1. Pravděpodobnostní funkce negativně binomického modelu
Naprogramujte v R funkci dNegBinom(x, k, p) počítající hodnoty pravděpodobnostní funkce negativně binomického rozdělení $\mathrm{NegBin}(k, p)$ v hodnotě x. Správnost funkce otestujte na výpočtu $p(x)$, $x = 0, 1, 2$ pro $X \sim \mathrm{NegBin}(k, p)$, kde $k = 5$ a $p = 0.4$. Výsledky ověřte s výsledky funkce dnbinom() dostupné v R. Jaký je rozdíl v syntaxích obou funkcí?

Řešení příkladu 3.1

```
       p0      p1       p2
1 0.07776 0.15552 0.186624
```

$p(0) = 0.7776$; $p(1) = 0.1555$; $p(2) = 0.1866$. ★

Příklad 3.2. Výpočet pravděpodobností na základě negativně binomického modelu
Pravděpodobnost sestřelení terče dosahuje u biatlonisky Koukalové až 95 %. Pravděpodobnost sestřelení terče u biatlonisky Charvátové se pohybuje okolo 65 %. Porovnejte, jaká je pravděpodobnost, že při tréninku sestřelí každá z biatlonistek před prvními dvěma neúspěchy (a) právě pět terčů; (b) nejvýše deset terčů; (c) alespoň sedm terčů; (d) alespoň patnáct terčů; (e) alespoň 25 terčů.

Řešení příkladu 3.2

```
           prave 5 nejvyse 10 alespon 7 alespon 15 alespon 25
Koukalova   0.0116     0.1184    0.9428     0.8108     0.6241
Charvatova  0.0853     0.9576    0.1691     0.0098     0.0002
```

Biatlonistka Koukalová při tréninku před prvními dvěma neúspěchy sestřelí s pravděpodobností 1.16 % právě pět terčů, s pravděpodobností 11.84 % nejvýše 10 terčů, s pravděpodobností 94.28 % alespoň 7 terčů, s pravděpodobností 81.08 % alespoň 15 terčů a s pravděpodobností 62.41 % alespoň 25 terčů.

Biatlonistka Charvátová při tréninku před prvními dvěma neúspěchy sestřelí s pravděpodobností 8.53 % právě pět terčů, s pravděpodobností 95.76 % nejvýše 10 terčů, s pravděpodobností 16.91 % alespoň 7 terčů, s pravděpodobností 0.98 % alespoň 15 terčů a s pravděpodobností 0.02 % alespoň 25 terčů. ★

Příklad 3.3. Výpočet očekávaných početností za předpokladu negativně binomického modelu
V příkladu 2.2 jsme popsali počet úrazů u dělníků v továrně pomocí náhodné veličiny X, která pocházela z Poissonova rozdělení s parametrem $\lambda = 0.4652$ odhadnutým na základě dat. Srovnáním pozorovaných početností m_{obs} s očekávanými početnostmi m_{exp}, za podmínky, že $X \sim \text{Poiss}(0.4652)$ jsme diagnostikovali overdisperzi v datech oproti předpokládanému Poissonovu modelu. Hodnoty pozorovaných a očekávaných početností jsou pro připomenutí uvedeny v následující tabulce.

n	0	1	2	3	4	≥ 5	$\sum$
$m_{observed}$	447	132	42	21	3	2	647
$m_{exp.Poiss}$	406	189	44	7	1	0	647

V případě, že data vykazují po aproximaci Poissonovým rozdělením overdisperzi, bývá vhodné popsat data náhodnou veličinou X pocházející z negativně binomického rozdělení $X \sim \text{NegBin}(k, p)$. Vypočítejte očekávané početnosti výskytu úrazů u dělníků v továrně za předpokladu, že početnosti úrazů pocházejí z negativně binomického rozdělení $\text{NegBin}(k, p)$, kde $\widehat{k} = \frac{\widehat{E[X]}^2}{\widehat{\text{Var}[X]} - \widehat{E[X]}}$ a $\widehat{p} = 1 - \frac{\widehat{E[X]}}{\widehat{\text{Var}[X]}}$. Výsledné početnosti porovnejte s očekávanými početnostmi za předpokladu Poissonova rozdělení.

Řešení příkladu 3.3

```
          k         p
1 0.9548138 0.3276139
```

```
          0   1  2  3 4 5
m.obs   447 132 42 21 3 2
m.exp.p 406 189 44  7 1 0
m.exp.n 443 139 44 14 5 2
```

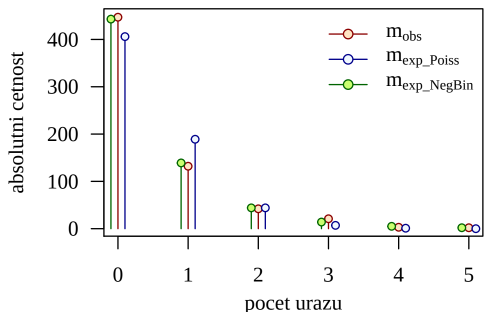
Odhad $\widehat{k} = 0.9548$, odhad $\widehat{p} = 0.3276$.

n	0	1	2	3	4	≥ 5	$\sum$
$m_{observed}$	447	132	42	21	3	2	647
$m_{exp.Poiss}$	406	189	44	7	1	0	647
$m_{exp.NegBin}$	443	139	44	14	5	2	647

Z tabulky vidíme, že očekávané početnosti vypočítané za předpokladu negativně binomického modelu lépe popisují chování pozorovaných početností úrazů u dělníků v továrně než početnosti vypočítané za předpokladu Poissonova modelu. ★

Příklad 3.4. Overdispersion a underdispersion v negativně binomickém modelu
V příkladu 3.3 jsme vypočítali hodnoty očekávaných absolutních početností úrazů dělníků v továrně za podmínky, že data pochází z NegBin(k, p), kde $\widehat{k} = 0.9548$, $\widehat{p} = 0.3276$. Do jednoho grafu zaneste nyní hodnoty pozorovaných početností $m_{observed}$, hodnoty očekávaných početností za předpokladu Poissonova rozdělení $m_{exp.Poiss}$ a hodnoty očekávaných početností za předpokladu negativně binomického rozdělení $m_{exp.NegBin}$. Na základě výsledného grafu stanovte, zda v případě použití negativně binomického modelu dochází k overdisperzi nebo underdisperzi. Závěr podložte srovnáním rozptylu vypočítaného z pozorovaných dat s rozptylem vypočítaným z očekávaných dat.

Řešení příkladu 3.4



Obrázek 1: Pozorované četnosti a očekávané četnosti za předpokladu negativně binomického a Poissonova modelu

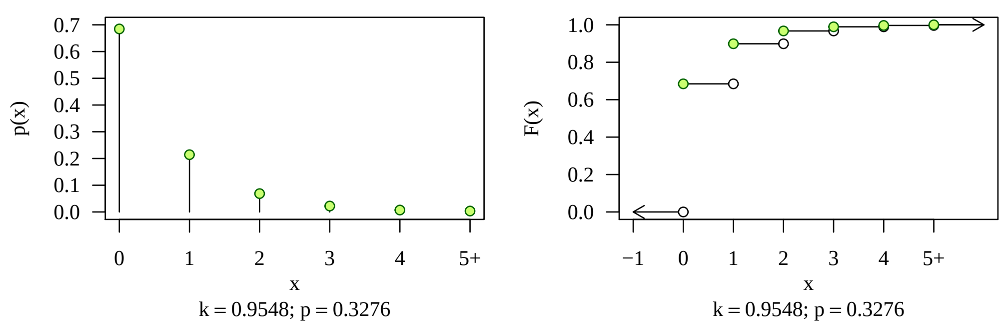
Z vykresleného grafu je zřetelně viditelná blízkost pozorovaných početností a očekávaných početností za předpokladu negativně binomického modelu.

```
    Var.obs Var.exp.n
1 0.6919002 0.6700035
```

Hodnota rozptylu získaného z pozorovaných dat vyšla 0.6919, hodnota rozptylu získaného z očekávaných dat za předpokladu negativně binomického modelu vyšla 0.6700. Vidíme, že hodnoty rozptylů jsou si velmi blízké. V tomto případě tedy nedochází ani k overdisperzi ani k underdisperzi. Pro srovnání připomeňme, že hodnota rozptylu získaného z očekávaných dat za předpokladu Poissonova modelu vyšla 0.4691 (overdisperze; viz příklad 2.2). ★

Příklad 3.5. Graf pravděpodobnostní a distribuční funkce negativně binomického modelu
V příkladu 3.3 jsme odhadli hodnoty parametrů k a p negativně binomického rozdělení NegBin(k, p) jako $\widehat{k} = 0.9548$, $\widehat{p} = 0.3276$. Na základě tohoto rozdělení jsme vypočítali očekávané početnosti výskytu úrazů u dělníků v továrně. Nakreslete graf pravděpodobnostní a distribuční funkce negativně binomického rozdělení NegBin(k, p), kde $k = 0.9548$, $p = 0.3276$, v hodnotách $x = 0, 1, 2, 3, 4$ a $x \geq 5$.

Řešení příkladu 3.5 ★



Obrázek 2: Pravděpodobnostní a distribuční funkce negativně binomického modelu

Příklad 3.6. Výpočet pravděpodobností na základě negativně binomického modelu
Za předpokladu, že náhodná veličina X, udávající počet úrazů u dělníků v továrně, pochází z negativně binomického modelu s parametry $k = 0.9548$ a $p = 0.3276$, tj. $X \sim \text{NegBin}(k, p)$ 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. Výsledky porovnejte s pravděpodobnostmi vypočítanými v příkladu 2.6 za předpokladu Poissonova modelu.

Řešení příkladu 3.6

```
          a          b         c         d
1 0.6845545 0.02929255 0.9672589 0.3154455
```

Pravděpodobnost, že u vybraného dělníka nedojde během roku k žádnému úrazu, je 68.46 %. Pravděpodobnost, že u vybraného dělníka dojde během roku k třem nebo čtyřem úrazům, je 2.93 %. Pravděpodobnost, že u vybraného dělníka dojde během roku k nejvýše dvěma úrazům, je 96.73 %. Pravděpodobnost, že u vybraného dělníka dojde během roku k alespoň jednomu úrazu, je 31.54 %.

počet úrazů	žádný	tři nebo čtyři	nejvýše dva	alespoň jeden
NegBin	0.6846	0.0293	0.9673	0.3154
Poiss	0.6280	0.0118	0.9881	0.3720

★

Příklad 3.7. Střední hodnota a rozptyl náhodné veličiny z negativně binomického modelu
Za předpokladu, že náhodná veličina X, udávající počet úrazů u dělníků v továrně, pochází z negatovně binomického rozdělení, tj. $X \sim \text{NegBin}(k, p)$, s parametry $k = 0.9548$, $p = 0.3276$, vypočítejte střední hodnotu $\text{E}[X]$ a rozptyl $\text{Var}[X]$ náhodné veličiny X. Střední hodnotu a rozptyl porovnejte s jejich odhady vypočítanými na (a) základě očekávaných dat; (b) na základě pozorovaných dat (viz příklad 3.3). Výsledky porovnejte s pravděpodobnostmi vypočítanými za předpokladu Poissonova modelu $\text{Poiss}(\lambda)$, kde $\lambda = 0.4652$.

Řešení příkladu 3.7

```
        E.X     Var.X     E.exp   Var.exp     E.obs   Var.obs
1 0.4652241 0.6919002 0.4621329 0.6700035 0.4652241 0.6919002
```

Střední hodnota počtu úrazů dělníků v továrně je 0.4652 s rozptylem 0.6919, odhad střední hodnoty počtu úrazů dělníků v továrně vypočítaný na základě očekávaných hodnot je 0.4652 s rozptylem 0.6919. Odhad střední hodnoty počtu úrazů dělníků v továrně vypočítaný na základě pozorovaných dat je 0.4621 s rozptylem 0.6700.

Tabulka 1: Porovnání odhadu střední hodnoty $E[X]$ a rozptylu $\text{Var}[X]$ (a) z pozorovaných dat; (b) z očekávaných dat za předpokladu, že $X \sim \text{NegBin}(k, p)$; (c) z očekávaných dat za předpokladu, že $X \sim \text{Poiss}(\lambda)$

	$\widehat{\text{E}[X]}$	$\widehat{\text{Var}[X]}$
pozorovaná data	0.4652	0.6919
očekávaná - NegBin	0.4652	0.6700
očekávaná data - Poiss	0.4668	0.4691

★

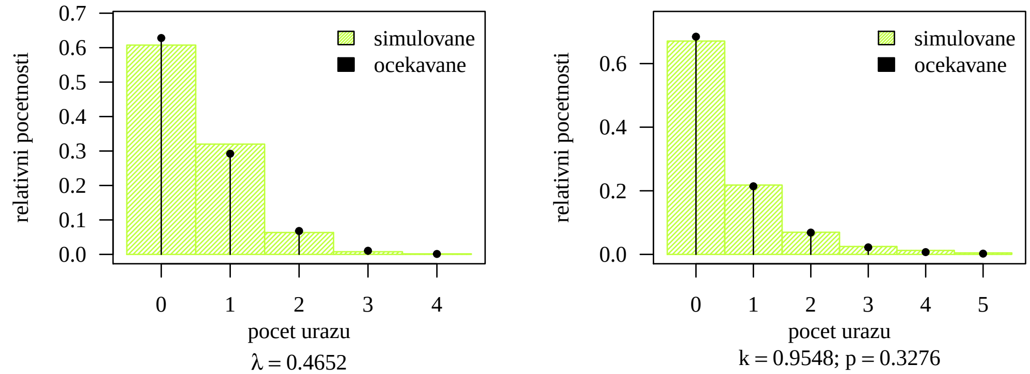
Příklad 3.8. Simulační studie pro negativně binomický model a Poissonův model
Vytvořte simulační studii modelující chování očekávaných početností náhodné veličiny X popisující počet úrazů dělníků v továrně za předpokladu, že

(a) $X \sim \text{Poiss}(\lambda)$, kde $\lambda = 0.4652$,

(b) $X \sim \text{NegBin}(k, p)$, kde $k = 0.9548$ a $p = 0.3276$.

Vygenerujte pseudonáhodná čísla X (početnosti úspěchů) opakovaná M-krát ($M = 647$) pocházející (a) z Poissonova rozělení, tj. $X \sim \text{Poiss}(\lambda)$; (b) z negativně binomického rozdělení, tj. $X \sim \text{NegBin}(k, p)$. Pro každé rozdělení vytvořte histogram vygenerovaných pseudonáhodných čísel a superponujte jej hodnotami očekávaných (teoretických početností).

Řešení příkladu 3.8



Obrázek 3: Porovnání pozorovaných četností a očekávaných četností za předpokladu (a) Poissonova modelu (vlevo); (b) negativně binomického modelu (vpravo)

★