MATEMATICKO-FYZIKÁLNÍ FAKULTA
Univerzita Karlova

BAKALÁŘSKÁ PRÁCE

Albert Zejda

Rozdělení typu $(a, b, 0)$ v neživotním pojištění

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: Mgr. Ing. Pavel Kříž, Ph.D.
Studijní program: Matematika
Studijní obor: Finanční matematika

Praha 2023

Prohlašuji, že jsem tuto bakalářskou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů. Tato práce nebyla využita k získání jiného nebo stejného titulu.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V .... dne ....

Podpis autora

Děkuji panu Mgr. Ing. Pavlovi Křížovi, Ph.D. za ochotu a nespočet věcných připomínek.

Název práce: Rozdělení typu $(a, b, 0)$ v neživotním pojištění

Autor: Albert Zejda

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: Mgr. Ing. Pavel Kříž, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: Nejprve je představena definice rozdělení typu $(a, b, 0)$. Následně je ukázáno, která známá rozdělení definici splňují, které parametry a, b jim přísluší a jsou určeny konkrétní množiny parametrů pro každé z rozdělení. Potom je dokázáno, že žádná další rozdělení tuto definici splňovat nemohou. Je představena metoda maximální věrohodnosti odhadu parametrů a, b přímo z dat. Na závěr je vypracována simulační studie, ve které se porovnávají pravděpodobnosti z odhadnutého rozdělení typu $(a, b, 0)$ z konkrétních dat metodou maximální věrohodnosti s empirickými relativními četnostmi spočítanými z dat.

Klíčová slova: diskrétní rozdělení, neživotní pojištění, třída $(a, b, 0)$, odhad parametrů.

Title: Distributions of $(a, b, 0)$ type in non-life insurance

Author: Albert Zejda

Department: Department of Probability and Mathematical Statistics

Supervisor: Mgr. Ing. Pavel Kříž, Ph.D., Department of Probability and Mathematical Statistics

Abstract: First, a definition of the distribution type $(a, b, 0)$ is introduced. Next, it is shown which known distributions satisfy this definition, the parameters a and b that correspond to them, and specific sets of parameters for each of the distributions are determined. Then, it is proven that no other distributions can satisfy this definition. A maximum likelihood estimation method for estimating the parameters a and b directly from the data is presented. Finally, a simulation study is conducted, in which the probabilities from the estimated distribution type $(a, b, 0)$ from specific data using the maximum likelihood method are compared with the empirical relative frequencies calculated from the data.

Keywords: discrete distributions, non-life insurance, $(a, b, 0)$ class, parameter estimation.

Obsah

Úvod

Poissonovo, binomické, negativně binomické a geometrické rozdělení jsou rozdělení, která se využívají k modelování počtu škod v pojišťovnictví. Jelikož není vždy jasné, které z rozdělení bude pro modelování konkrétního případu ideální, tak vyjádřím všechna tato rozdělení jedním známým vztahem, se kterým bychom mohli dále pracovat a volba rozdělení při modelování počtu škod byla přímočařejší. V první části práce definuji vztah, kterým jsou rozdělení patřící do třídy $(a, b, 0)$ určeny a poté dokazuji, že Poissonovo, binomické, negativně binomické i geometrické rozdělení do třídy $(a, b, 0)$ náleží. Ve druhé části rozděluji parametry a, b do množin přípustných hodnot pro každé rozdělení a dokazuji, že parametry, které nenáleží do množiny přípustných hodnot žádného z rozdělení nemohou určovat rozdělení. V poslední teoretické kapitole se soustřeďuji na odhad parametrů a, b a to konkrétně metodou maximální věrohodnosti. Věrohodnost maximalizuji přes všechny množiny přípustných hodnot parametrů a vybírám optimální řešení. Na konci je ukázána simulační studie odhadů parametrů pomocí mnou zvolené metody maximální věrohodnosti.

Vlastním přínosem v práci je velké množství vlastních výpočtů, odvozování a obrázků uvedených v bakalářské práci a také závěrečná simulační studie. Mezi vlastní odvozování a výpočty patří především: důkaz tvrzení 2, výpočty v podkapitole 1.2, obrázky v podkapitole 2.2 a obrázky v kapitole 4.

1. Třída rozdělení $(a, b, 0)$

Základní vlastnosti rozdělení náležících třídě $(a, b,\ 0)$ včetně definice jsou popsány v knize Klugman a kol. (2019).

1.1 Definice

Definice 1. *Diskrétní rozdělení náhodných veličin s pravděpodobnostní funkcí $\{p_k,\ k = 0,\ 1,\ \ldots\ \}$ náleží třídě $(a, b,\ 0)$ za předpokladu, že existují konstanty a,b takové, že:*

$$p_k = \left(a + \frac{b}{k}\right) p_{k-1}\ ,\ k = 1, 2, \ldots\ , p_0 > 0. \tag{1.1}$$

Pravděpodobnost p_0 se dá získat zpětně z rekurzivní rovnice, protože se pravděpodobnosti musí sčítat na 1.

Pro rozdělení pravděpodobnosti diskrétní náhodné veličiny platí:

$$\sum_{k=0}^{\infty} p_k = 1.$$

Z čehož pro p_0 vyplývá:

$$p_0 = 1 - \sum_{k=1}^{\infty} p_k.$$

1.2 Rozdělení náležící třídě $(a, b, 0)$

Do rozdělení náležících třídě $(a, b,\ 0)$ počítáme Poissonovo, negativně binomické, binomické a speciální případ negativně binomického - geometrické rozdělení.
V této kapitole budu dokazovat, že tato rozdělení do třídy $(a, b,\ 0)$ patří a odvozovat jaké parametry a, b jim přísluší.

1.2.1 Poissonovo rozdělení

Poissonovo rozdělení má pravděpodobnostní funkci zadanou ve tvaru:

$$p_k = \frac{\lambda^k}{k!} e^{-\lambda}\ ,\ k = 1, 2, \ldots\ ,\ \lambda > 0.$$

Z toho vytknutím $\frac{\lambda}{k}$:

$$p_k = \frac{\lambda}{k} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda}.$$

Jelikož chci vyjádřit rekurzivní vztah, napíšu si, jak vypadá p_{k-1}:

$$p_{k-1} = \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda}.$$

Z toho dohromady pro p_k plyne:

$$p_k = \frac{\lambda}{k} p_{k-1}.$$

Takže Poissonovo rozdělení patří podle rovnice (1.1) do třídy rozdělení $(a, b,\ 0)$ a pro parametry platí:

$$a = 0 \ \ , \ \ b = \lambda.$$

Pro Poissonovo rozdělení navíc platí $p_0 = e^{-\lambda}$.

1.2.2 Negativně binomické rozdělení

Pravděpodobnostní funkce negativně binomického rozdělení je dána ve tvaru:

$$p_k = \binom{h+k-1}{k} \left(\frac{1}{\beta+1}\right)^h \left(\frac{\beta}{\beta+1}\right)^k \ , \ k = 1, 2, \dots \ , \ \ \beta > 0 \ , \ h > 0.$$

Jelikož v kombinačním čísle se může vyskytnout i jiné než přirozené číslo, definuji kombinační čísla pro reálná čísla. Vycházet budu z Klugman a kol. (2019, str. 85).

Definice 2. *Nechť $r \in \mathbb{R}$ a $k \in \mathbb{N}_0$, pak*

$$\binom{r}{k} = \frac{r(r-1)\dots(r-k+1)}{k!}$$

nazýváme kombinačním číslem r nad k.

Pravděpodobnostní funkci upravím na rekurzivní tvar:

$$p_k = \frac{h+k-1}{k} \frac{\beta}{1+\beta} \binom{h+k-2}{k-1} \left(\frac{1}{\beta+1}\right)^h \left(\frac{\beta}{\beta+1}\right)^{(k-1)},$$

$$p_k = \left(1 + \frac{h-1}{k}\right) \frac{\beta}{1+\beta} p_{k-1}.$$

Negativně binomické rozdělení také splňuje rovnici (1.1) s parametry:

$$a = \frac{\beta}{1+\beta} \ \ , \ \ b = (h-1)\frac{\beta}{1+\beta}.$$

Doplním $p_0 = (1+\beta)^{-h}$.

1.2.3 Binomické rozdělení

Pro pravděpodobnostní funkci binomického rozdělení platí:

$$p_k = \binom{m}{k} p^k (1-p)^{m-k} \ , \ k = 1, 2, \dots \ , \ 0 < p < 1 \ , \ m \in \mathbb{N}.$$

Znovu upravím na tvar splňující rovnici (1.1):

$$p_k = \frac{m-k+1}{k}\frac{p}{1-p}\binom{m}{k-1}p^{k-1}(1-p)^{m-(k-1)},$$

$$p_k = \left(-1+\frac{m+1}{k}\right)\frac{p}{1-p}p_{k-1}.$$

Z čehož vyplývá, že binomické rozdělení patří do třídy (a,b,0) s parametry:

$$a = -\frac{p}{1-p} \quad , \quad b = (m+1)\frac{p}{1-p}.$$

Také platí, že $p_0 = (1-p)^m$.

1.2.4 Geometrické rozdělení

Pravděpodobnostní funkce geometrického rozdělení má tvar:

$$p_k = \frac{1}{1+\beta}\left(\frac{\beta}{\beta+1}\right)^k \quad , \ k = 1, 2, \dots \ , \ \beta > 0.$$

Upravím na tvar, který splňuje rovnici (1.1):

$$p_k = \frac{\beta}{(1+\beta)^2}\left(\frac{\beta}{\beta+1}\right)^{k-1},$$

$$p_k = \frac{\beta}{(1+\beta)}p_{k-1}.$$

Parametry geometrického rozdělení jsou:

$$a = \frac{\beta}{(1+\beta)} \quad , \quad b = 0.$$

Platí $p_0 = \frac{1}{1+\beta}$.

1.2.5 Shrnutí první kapitoly

Dokázal jsem, že Poissonovo, negativně binomické, binomické a geometrické rozdělení jsou typy rozdělení, jejichž společným znakem je, že splňují rovnici (1.1). V následující tabulce napíšu všechny parametry, které jsem pro tato rozdělení určil.

Rozdělení	a	b	p_0
Poissonovo	0	λ	$e^{-\lambda}$
Negativně binomické	$\frac{\beta}{1+\beta}$	$(h-1)\frac{\beta}{1+\beta}$	$(1+\beta)^{-h}$
Binomické	$-\frac{p}{1-p}$	$(m+1)\frac{p}{1-p}$	$(1-p)^m$
Geometrické	$\frac{\beta}{1+\beta}$	0	$(1+\beta)^{-1}$

Tabulka 1.1: Parametry chakarakterizující rozdělení patřící do třídy $(a, b,\ 0)$

2. Hodnoty parametrů $(a, b, 0)$

2.1 Motivace

V předchozí kapitole jsem ukázal, že Poissonovo, negativně binomické, binomické a geometrické rozdělení splňují rovnici (1.1), což implikuje, že podle definice (1) patří do třídy rozdělení (a, b, 0). V této kapitole bych chtěl ukázat, že tato rozdělení jsou jediná, které splňují rovnici (1.1), neboli jsou jediná, které do třídy rozdělení typu (a, b, 0) patří.
Struktura této kapitoly tedy bude:

- Určení množin parametrů rozdělení.
- Důkaz toho, že pokud parametry a, b nebudou náležet žádné z množin, pak neurčují rozdělení.
- Přechod z množin parametrů (a, b, 0) k původní parametrizaci.

2.2 Množiny parametrů rozdělení

V této kapitole budu Poissonovo, negativně binomické, binomické a geometrické rozdělení vnímat jako rozdělení typu (a, b, 0) s parametry a a b.
Označím M_{dist}, $dist \in \{Poiss, Binom, NegBinom\}$ množiny všech přípustných hodnot parametrů a, b určující dané rozdělení a množinu M jako:

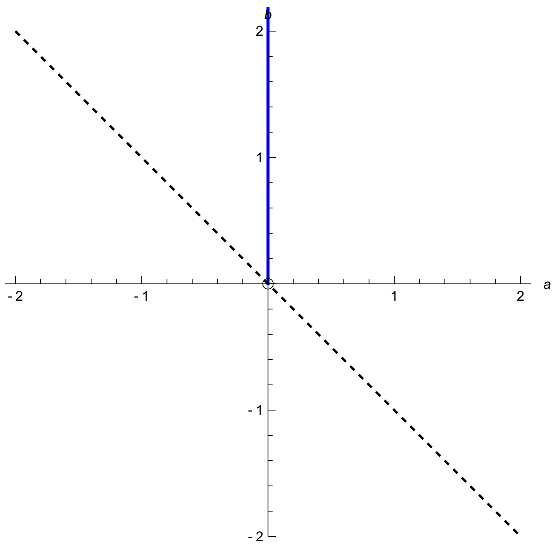
$$M = M_{Poiss} \cup M_{NegBinom} \cup M_{Binom}$$

2.2.1 Poissonovo Rozdělení

Poissonovo rozdělení budu nyní vnímat jako rozdělení typu (a, b, 0) s parametry $a = 0$ a $b = \lambda$. Pro Poissonovo rozdělení platí, že $\lambda > 0$.
Množina M_{Poiss} tedy bude vypadat následovně:

$$M_{Poiss} = \{(a,b) : a = 0,\, b > 0\}.$$

Vykreslím množinu M_{Poiss}.



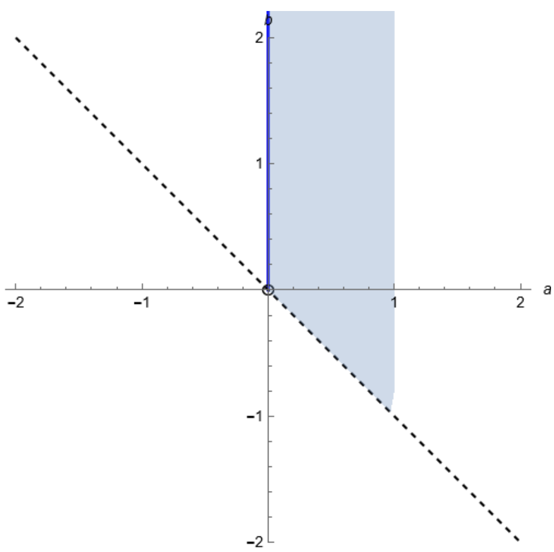
Obrázek 2.1: M_{Poiss}

2.2.2 Negativné binomické rozdělení

Jelikož geometrické rozdělení je speciálním případem negativně binomického rozdělení pro $h = 1$, parametry budu vymezovat pouze obecně pro negativně binomické rozdělení.
Podle tabulky 1.1 pro negativně binomické rozdělení $a = \frac{\beta}{1+\beta}$ a $b = (h-1)\frac{\beta}{1+\beta}$.
Vyjádřím parametr b v závislosti na a. Dostávám $b = (h-1)a$.
Pro výraz $\frac{\beta}{1+\beta}$ platí, že jeho oborem hodnot je interval (0,1).
Také platí, že $h > 0$, a proto $h - 1 > -1$.
Dohromady: $a \in (0,1) \wedge b > -a$.

$$M_{NegBinom} = \{(a,b) : 0 < a < 1, b > -a\}.$$

Obrázek 2.1 rozšířím o množinu všech přípustných hodnot parametrů definující negativně binomické rozdělení.



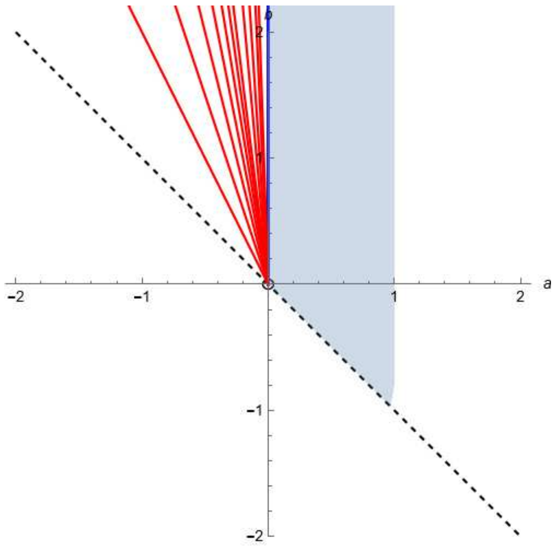
Obrázek 2.2: $M_{Poiss} \cup M_{NegBinom}$

2.2.3 Binomické Rozdělení

Pro binomické rozdělení parametr $a = -\frac{p}{1-p}$ a $b = (m+1)\frac{p}{1-p}$ tedy b v závislosti na a vyjádřím znovu jednodušše: $b = -(m+1)a$.
Pro p je obor hodnot interval (0,1), a proto pro a je obor hodnot $(-\infty,0)$ a zároveň $m \in \mathbb{N}$.

$$M_{Binom} = \{(a,b) : a < 0,\ b = -(m+1)a,\, m \in \mathbb{N}\}$$

Do obrázku 2.2 přidáme množinu M_{Binom}.



Obrázek 2.3: $M_{Poiss} \cup M_{NegBinom} \cup M_{Binom}$

2.3 Pravděpodobnostní funkce na doplňku M

V této kapitole dokážu, že množiny M_{Poiss}, $M_{NegBinom}$ a M_{Binom} obsahují všechny možné dvojice parametrů, které určují pravděpodobnostní funkci. Tedy: Nechť $(a, b) \notin (M_{Poiss} \cup M_{NegBinom} \cup M_{Binom})$, pak p_k není nedegenerovaná (ani jedna z pravděpodoboстí p_k, $k \in 0,\, 1,\, 2,\, \ldots$ nemá hodnotu 1) pravděpodobnostní funkce.
Zprvu musím definovat doplněk k M. Z obrázku 4.3 to bude sjednocení následujících množin:

- $M_1 = \{(a,b) :\ a + b = 0\}$.
- $M_2 = \{(a,b) :\ a + b < 0\}$.
- $M_3 = \{(a,b) :\ a < 0, b = -ar,\ r \in (n, n+1)\,,\ n \in \mathbb{N}\}$.
- $M_4 = \{(a,b) :\ a > 1,\ a + b > 0\}$.

2.3.1 Množina M_1

Tato množina na obrázku 2.3 vyznačuje přímku $b = -a$.

Pro p_1 platí vztah:

$$p_1 = (a + b)p_0.$$

Jelikož $a + b = 0$:

$$p_1 = 0 \implies p_2\ , p_3, \dots = 0.$$

Z toho vyplývá, že $p_0 = 1$, což porušuje předpoklad nedegenerovanosti.

2.3.2 Množina M_2

Množina $M_2 = \{(a,b) : a + b < 0\}$ je na obrázku 2.3 polorovina s hraniční přímkou $a = b$.
Pro p_1 platí:

$$p_1 = (a + b)p_0 \implies p_1 < 0.$$

Což znamená, že p_k není pravděpodobnostní funkce.

2.3.3 Množina M_3

Na obrázku 2.3 je množina $M_3 = \{(a,b) : a < 0, b = -ar, r \in (n, n + 1)\}$ kužel, který má jako strany sousední polopřímky, jež však tato množina neobsahuje.

$$p_k = \left(a - \frac{ar}{k}\right) p_{k-1}.$$

Z toho vyplývá

$$p_k = a\left(1 - \frac{r}{k}\right) p_{k-1}.$$

$a < 0$, aby tedy p_k mohla být pravděpodobnostní funkce musí platit:

$$\left(1 - \frac{r}{k}\right) < 0.$$

Za předpokladu existence pravděpodobnostní funkce dostávám vztah:

$$k \leq r.$$

Jelikož $r \in (n,n + 1)$:

$$k \leq n.$$

Pro p_k, kde $k \leq n$ za předpokladu $p_0 > 0$ platí:

$$p_k = a\left(1 - \frac{r}{k}\right) p_{k-1} > 0.$$

Pro p_{n+1} platí:

$$p_{n+1} = a\left(1 - \frac{r}{n + 1}\right) p_n.$$

Jelikož $n + 1 > r$, tak $p_{n+1} < 0$ a p_k proto nemůže být pravděpodobnostní funkce.

2.3.4 Množina M_4

Na obrázku 2.3 je množina $M_4 = \{(a,b) : a > 1, a + b > 0\}$ posunutý kužel se stranami definované přímkami $b = -a$ a $a = 1$, které však tato množina neobsahuje.

Z definice 1:

$$p_k = \left(a + \frac{b}{k}\right) p_{k-1}.$$

Ten upravím přičtením a odečtením $\frac{a}{k}$:

$$p_k = \left(a + \frac{b}{k} + \frac{a}{k} - \frac{a}{k}\right) p_{k-1}.$$

Z toho vyplývá, že:

$$p_k = \left(\frac{1}{k}(a + b) + a\left(1 - \frac{1}{k}\right)\right) p_{k-1}.$$

Po roznásobení závorky vyjde:

$$p_k = (a + b)\frac{p_{k-1}}{k} + a\frac{k-1}{k}p_{k-1}.$$

Protože první člen je kladný, tak:

$$p_k \geq a\frac{k-1}{k}p_{k-1}.$$

Díky platnosti podmínky, že $a > 1$ platí:

$$p_k \geq \frac{k-1}{k}p_{k-1}.$$

Rozepíšu pravou stranu:

$$p_k \geq \frac{k-1}{k}\frac{k-2}{k-1}\frac{k-3}{k-2}\cdots\frac{1}{2}p_1.$$

Pokrátí se vše až na $\frac{p_1}{k}$, tedy:

$$p_k \geq \frac{p_1}{k}.$$

Pro sumu pravděpodobností platí:

$$\sum_{k=1}^{\infty} p_k \geq p_1 \sum_{k=1}^{\infty} 1/k = \infty.$$

Čímž jsem dokázal, že p_k není pravděpodobnostní funkce.

2.4 Přechod z M_{dist} k původní parametrizaci

V této kapitale dokážu, že libovolné rozdělení typu $(a, b, 0)$, jehož parametry náleží množině M_{dist}, musí být nutně typu *dist*. To sice plyne z úvah v předchozích kapitolách, ale v této kapitole jsem udělal alternativní a především přímý důkaz.

2.4.1 Poissonovo rozdělení

Předpokládám, že $(a,b) \in M_{Poiss} = \{(a,b) : a = 0,\ b > 0\}$.

Pro p_1 a p_2 platí:

$$p_1 = bp_0\ ,\ p_2 = \frac{b^2}{2}p_0.$$

Lehce si uvědomím, že:

$$p_k = \frac{b^k}{k!}p_0.$$

Pro sumu p_k mám:

$$\sum_{k=0}^{\infty} p_k = p_0 \sum_{k=0}^{\infty} \frac{b^k}{k!} = 1.$$

Z toho pro p_0 vyplývá:

$$p_0 = \left(\sum_{k=0}^{\infty} \frac{b^k}{k!}\right)^{-1}.$$

Suma je však rovna Taylorovu rozvoji e^b, tedy:

$$p_0 = e^{-b}.$$

Celkem:

$$p_k = \frac{b^k}{k!}p_0 = \frac{b^k}{k!}e^{-b}.$$

Takže jsem dokázal, že pokud $(a,b) \in M_{Poiss}$, tak p_k je pravděpodobnostní funkce Poissonova rozdělení.

2.4.2 Negativně binomické rozdělení

Nyní předpokládám, že $a,b \in M_{NegBinom} = \{(a,b) : 0 < a < 1, b > -a\}$.

$$p_1 = (a+b)p_0 = a\left(1 + \frac{b}{a}\right)p_0,$$

$$p_2 = \left(a + \frac{b}{2}\right)(a+b)p_0 = a\left(1 + \frac{b}{2a}\right)a\left(1 + \frac{b}{a}\right)p_0,$$

$$\begin{aligned} p_3 &= a\left(1 + \frac{b}{3a}\right)a\left(1 + \frac{b}{2a}\right)a\left(1 + \frac{b}{a}\right)p_0 \\ &= a^3\frac{1}{3}\left(3 + \frac{b}{a}\right)\frac{1}{2}\left(2 + \frac{b}{a}\right)\left(1 + \frac{b}{a}\right)p_0. \end{aligned}$$

Pro p_k platí:

$$p_k = a^k \frac{\left(k + \frac{b}{a}\right)\left(k - 1 + \frac{b}{a}\right)...\left(1 + \frac{b}{a}\right)}{k!}p_0 = a^k \binom{k + \frac{b}{a}}{k} p_0.$$

Ke každému členu přičtu a odečtu 1:

$$p_k = a^k \frac{\left(k + \frac{b}{a} + 1 - 1\right)\left(k - 1 + \frac{b}{a} + 1 - 1\right) ... \left(\frac{b}{a} + 1\right)}{k!} p_0,$$

$$p_k = a^k \frac{\left(-1\right)(-\frac{b}{a} - 1\right) \dots (-1)\left(-k - \frac{b}{a} - 1 + 1\right)}{k!} p_0.$$

Což upravím na:

$$p_k = a^k (-1)^k \binom{-\frac{b}{a} - 1}{k} p_0.$$

Pro sumu pravděpodobostí platí:

$$\sum_{k=0}^{\infty} p_k = p_0 \sum_{k=0}^{\infty} (-a)^k \binom{-\frac{b}{a} - 1}{k} = 1.$$

Při formulaci následujícího tvrzení budu vycházet z Lang (1999)(str.57,58).

Tvrzení 1. *Nechť $r \in \mathbb{R}$ a $k \in \mathbb{N}_0$, potom platí:*

$$(1 + x)^r = \sum_{k=0}^{\infty} \binom{r}{k} x^k \;, \; -1 < x < 1.$$

Z tvrzení 1 vyplývá:

$$p_0 \,(1 - a)^{-\frac{b}{a} - 1} = 1.$$

Pro p_0 tedy:

$$p_0 = (1 - a)^{\frac{b}{a} + 1} . \tag{2.1}$$

Celkem:

$$p_k = \binom{k + \frac{b}{a} - 1}{k} a^k \,(1 - a)^{\frac{b}{a} + 1} .$$

Pro $\beta = \frac{1-a}{a}$ a $h = \frac{b}{a}$ to je negativně binomické rozdělení.

2.4.3 Binomické rozdělení

Nyní uvažuji $(a,b) \in M_{Binom} = \{(a,b) : a < 0, b = -(m + 1)a, m \in \mathbb{N}\}$.

První ukážu, že pro $k > m$ je $p_k = 0$:

$$p_{m+1} = \left(a + \frac{-(m + 1)a}{m + 1}\right) p_m = 0 \implies p_k = 0 \; k > m,$$

$$p_k = \left(a + \frac{-a(m + 1)}{k}\right) p_{k-1} = -a \left(\frac{m - (k - 1)}{k}\right) p_{k-1}.$$

Pro první členy:

$$p_1 = -a\, m\, p_0,$$

$$p_2 = -a\frac{m-1}{2}p_1 = (-a)^2\frac{m(m-1)}{2}p_0,$$

$$p_3 = -a\frac{m-2}{3}p_2 = (-a)^3\frac{m(m-1)(m-2)}{3*2}p_0.$$

Pro p_k tedy platí:

$$p_k = (-a)^k\frac{m(m-1)...(m-k+1)}{k!}p_0.$$

Což je přesně:

$$p_k = (-a)^k\binom{m}{k}p_0.$$

Pro sumu pravděpodobností platí:

$$\sum_{k=0}^{\infty} p_k = \sum_{k=0}^{m} p_k = p_0\sum_{k=0}^{m}(-a)^k\binom{m}{k} = 1.$$

Tedy:

$$p_0 = \left(\sum_{k=0}^{m}(-a)^k\binom{m}{k}\right)^{-1}.$$

Z rovnice (2.1) vím, že pro negativně binomické rozdělení $p_0 = (1-a)^{\frac{b+a}{a}}$. Teď ukážu, že tento vztah platí i na M_{Binom} a dodefinuji ho na M_{Poiss}.

Tvrzení 2. *Nechť existuje nedegenerovaná pravděpodobnostní funkce p_k definovaná dle definice 1, pak*

$$p_0 = (1-a)^{\frac{b+a}{a}}\ ,\ (a,b) \in M\ ,\ a \neq 0,$$

$$p_0 = \lim_{a\to 0}(1-a)^{\frac{b+a}{a}} = e^{-b}\ ,\ (a,b) \in M\ ,\ a = 0.$$

Důkaz. Jelikož vím, že p_k je nedegenerovaná pravědpodobnostní funkce, tak díky poznatkům z kapitoly 2 můžu říct, že p_k je pravděpodobonstní funkce binomického, negativně binomického nebo Poissonova rozdělení. U těch znám a, b a p_0 z tabulky 1.1. Pro Poissonovo rozdělení $a = 0$, $b = \lambda$, $p_0 = e^{-\lambda}$. Podle tvrzení:

$$p_0 = e^{-b} = e^{-\lambda}.$$

Jelikož se p_0 spočítané z rovnice v tvrzení shoduje s p_0 pro Poissonovo rozdělení, tak vidím, že pro Poissonovo rozdělení tvrzení platí.
Pro negativně binomické $a = \frac{\beta}{\beta+1}$, $b = (h-1)\frac{\beta}{\beta+1}$, $p_0 = (1+\beta)^{-h}$. Dosadím do rovnice z tvrzení:

$$p_0 = (1-a)^{\frac{b+a}{a}} = \left(1-\frac{\beta}{\beta+1}\right)^{\frac{\frac{\beta}{\beta+1}(1+(h-1))}{\frac{\beta}{\beta+1}}} = \left(\frac{1}{\beta+1}\right)^{h} = (1+\beta)^{-h}.$$

Znovu docházím k závěru, že p_0 se shodují.
Pro binomické rozdělení $a = -\frac{p}{1-p}$, $b = (m+1)\frac{p}{1-p}$, $p_0 = (1-p)^m$.

$$p_0 = \left(1 + \frac{p}{1-p}\right)^{\frac{\frac{p}{1-p}m}{-\frac{p}{1-p}}} = (1-p)^m.$$

Jelikož p_0 jsou znovu shodné, tak tvrzení platí na celém M a je tedy dokázané.

$\square$

dle tvrzení 2:

$$p_0 = \left((1-a)^{\frac{-a(m+1)+a}{a}}\right)^{-1}.$$

Celkem:

$$p_k = \binom{m}{k}(-a)^k \left((1-a)^{-1}\right)^m.$$

Což je binomické rozdělení s parametrem $p = \frac{a}{a-1}$.

3. Odhady parametrů a, b

3.1 Vyjádření pravděpodobnostní funkce

Z definice vím, že pro rozdělení typu $(a, b, 0)$ platí:

$$p_k = p_{k-1}\left(a + \frac{b}{k}\right).$$

Odvodím vztah pro k=n
pro k=1

$$p_1 = p_0(a + b).$$

pro k=2

$$p_2 = p_1\left(a + \frac{b}{2}\right).$$

kde však vím, že $p_1 = p_0\,(a + b)$, tedy:

$$p_2 = p_0(a + b)\left(a + \frac{b}{2}\right).$$

Vidím, že pro p_n platí:

$$p_n = p_0\prod_{k=1}^{n}\left(a + \frac{b}{k}\right).$$

3.2 Metoda maximální věrohodnosti

Metoda maximální věrohodnosti je standardní metoda hledání odhadu parametrů. V literatuře je popsána například v knize Anděl (2007).

Definice 3. *Nechť $X_1, X_2 \dots X_n$ je náhodný výběr z rozdělení s hustotou $f(x|\theta_0)$ vzhledem k míře μ náležících do parametrického modelu $\mathcal{F}$ = {rozdělení s hustotou $f(x|\theta)$, $\theta \in \Theta \subseteq \mathbb{R}^d$}. Předpokládáme, že různým hodnotám pramaterů $\theta_1 \neq \theta_2$ odpovídají různé hustoty $f(.|\theta_1) \neq f(.|\theta_2)$. Náhodnou funkci*

$$L_n(\theta) = \prod_{i=1}^{n} f(X_i|\theta)$$

nazýváme věrohodnostní funkcí (věrohodností) pro parametr θ v modelu $\mathcal{F}$.

Definice 4. *Maximálně věrohodný odhad parametru θ v modelu $\mathcal{F}$ je definován jako*

$$\hat{\boldsymbol{\theta}}_n = arg \max_{\theta\in\Theta} L_n(\boldsymbol{\theta}).$$

Definice 5. *Náhodnou funkci*

$$l_n(\theta) = \log(L_n(\theta)) = \sum_{i=1}^{n} \log f(X_i|\theta)$$

nazýváme logaritmickou věrohodností.

Definice 6. *Náhodný vektor*

$$U(\theta|X) = \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(X_i|\theta)$$

nazýváme skórovou statistikou.

Pro náhodný výběr $X_1, X_2 ... X_n$ z rozdělení typu $(a, b, 0)$ platí:

$$L_n(a,b) = \prod_{i=1}^{n} (p_{X_i}).$$

Dosadím pozorování z minulého kroku:

$$L_n(a,b) = \prod_{i=1}^{n} \left(p_0 \prod_{k=1}^{X_i} \left(a + \frac{b}{k} \right) \right).$$

Napíšu a upravím logaritmickou věrohodnostní funkci:

$$l_n(a,b) = \sum_{i=1}^{n} \left(\log \left(p_0 \prod_{k=1}^{X_i} \left(a + \frac{b}{k} \right) \right) \right),$$

$$l_n(a,b) = \sum_{i=1}^{n} \left(\log(p_0) + \sum_{k=1}^{X_i} \log \left(a + \frac{b}{k} \right) \right),$$

$$l_n(a,b) = n \log(p_0) + \sum_{i=1}^{n} \sum_{k=1}^{X_i} \log \left(a + \frac{b}{k} \right).$$

Dosadím p_0 z tvrzení 2 s tím, že pro $a = 0$ je rovnice dodefinovaná limitně:

$$l_n(a,b) = n \log(1-a)^{\frac{b+a}{a}} + \sum_{i=1}^{n} \sum_{k=1}^{X_i} \log \left(a + \frac{b}{k} \right),$$

$$l_n(a,b) = \frac{n(b+a)}{a} \log(1-a) + \sum_{i=1}^{n} \sum_{k=1}^{X_i} \log \left(a + \frac{b}{k} \right). \tag{3.1}$$

Odhadování parametrů a, b z (3.1) by bylo velmi obtížné. Jelikož se pokusím najít maximálně věrohodný odhad a, b, tedy pokusím se maximalizovat věrohodnostní funkci na celém M, spočítám maximální logaritmické věrohodnosti pro jednotlivé M_{dist}, spočítám odhady parametrů a vyberu tu, jejíž funkční hodnota v odhadnutých parametrech bude maximální. Protože metoda maximální věrohodnosti nezáleží na parametrizaci, tak budeme pro potřeby maximalizace věrohodnosti používat původní parametrizaci a potom dopočítám odhady pro parametry a, b.

3.2.1 MLE pro Poissonovo rozdělení

Nechť $(a,b) \in M_{Poiss}$, pak:

$$p_k = \frac{\lambda^k}{k!} e^{-\lambda} \; k = 1,2,\dots \; , \; \lambda > 0.$$

Předpokládám, že mám soubor dat obsahující n pozorování $X_1, X_2, \dots, X_n$, o kterých předpokládám, že jsou nezávislé a stejně rozdělené (i.i.d.) náhodné veličiny z Poissonova rozdělení s parametrem λ. Potom:

$$L_n(\lambda) = \prod_{i=1}^{n} \frac{\lambda^{X_i}}{X_i!} e^{-\lambda}.$$

Z toho získám logaritmickou věrohodnostní funkci:

$$l_n(\lambda) = \sum_{i=1}^{n} \left[-\lambda + X_i \log(\lambda) - \log(X_i!)\right].$$

Tedy pro maximálně věrohodný odhad λ:

$$l_{Poiss} = \sum_{i=1}^{n} \left[-\hat{\lambda} + X_i \log(\hat{\lambda}) - \log(X_i!)\right].$$

Abych zjistil maximálně věrohodný odhad λ, musím zderivovat podle λ.

$$\frac{\partial l_{Poiss}}{\partial \lambda} = \sum_{i=1}^{n} \left[-1 + \frac{X_i}{\lambda}\right] = 0.$$

Upravím, abych vyřešil pro λ:

$$\sum_{i=1}^{n} X_i = n\lambda$$

Maximálně věrohodný odhad pro λ je:

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} X_i}{n} = \overline{X_n}.$$

Pokud tedy funkce l_{Poiss} bude mít větší funkční hodnotu než $l_{NegBinom}$ a l_{Binom}, budu odhadovat parametry a, b podle MLE pro Poissonovo rozdělení, tedy $a = 0$ a maximálně věrohodným odhadem b bude:

$$\hat{b} = \hat{\lambda} = \overline{X_n}.$$

3.2.2 MLE pro negativně binomické rozdělení

Nechť $(a,b) \in M_{NegBinom}$, pak:

$$p_k = \binom{h+k-1}{k} \left(\frac{1}{\beta+1}\right)^h \left(\frac{\beta}{\beta+1}\right)^k \; , \; k = 1,2,\dots \; , \quad \beta > 0 \; , \; h > 0.$$

Protože pro negativně binomické rozdělení $E(N) = h\beta$ a $Var(N) = h\beta(1+\beta)$, budu uvažovat negativně binomické rozdělení jako vhodného kandidáta na odhad parametrů pouze tehdy, když výběrový průměr bude menší než výběrový rozptyl.

Předpokládám, že mám soubor dat obsahující n pozorování $X_1, X_2, \dots, X_n$, o kterých předpokládám, že jsou nezávislé a stejně rozdělené (i.i.d.) náhodné veličiny z negativně binomického rozdělení s parametry h a β. Potom:

$$L_n = \prod_{i=1}^{n} p_{X_i} = \prod_{i=1}^{n} \binom{h + X_i - 1}{X_i} \left(\frac{1}{\beta+1}\right)^h \left(\frac{\beta}{\beta+1}\right)^{X_i}.$$

Z toho získám logaritmickou věrohodnostní funkci:

$$l_n(h,\beta) = \sum_{i=1}^{n} \log \binom{h + X_i - 1}{X_i} + h \log\left(\frac{1}{\beta+1}\right) + X_i \log\left(\frac{\beta}{\beta+1}.\right)$$

Tu upravím následovně:

$$l_n(h,\beta) = \sum_{i=1}^{n} \log \binom{h + X_i - 1}{X_i} - h \log(\beta+1) + X_i \log(\beta) - X_i \log(\beta+1).$$

Celkem pro maximálně věrohodné odhady β a h:

$$l_{NegBinom} = \sum_{i=1}^{n} \log \binom{\hat{h} + X_i - 1}{X_i} - \hat{h} \log\left(\hat{\beta}+1\right) + X_i \log\left(\hat{\beta}\right) - X_i \log(\hat{\beta}+1).$$

Spočítám a upravím parciální derivace.

$$\frac{\partial ł_n(h,\beta)}{\partial h} = -\sum_{i=1}^{n} \log(1+\beta) + \sum_{i=1}^{n} \frac{\partial}{\partial h} \log\left(\frac{(h+X_i-1)\dots h}{X_i!},\right)$$

$$\frac{\partial ł_n(h,\beta)}{\partial h} = -n\log(1+\beta) + \sum_{i=1}^{n} \left(\frac{\partial}{\partial h} \log \prod_{m=1}^{X_i} (h + X_i - m) - \frac{\partial}{\partial h} \log X_i!\right),$$

$$\frac{\partial ł_n(h,\beta)}{\partial h} = -n\log(1+\beta) + \sum_{i=1}^{n} \sum_{m=1}^{X_i} \frac{1}{(h+X_i-m)},$$

$$\frac{\partial l_n(h,\beta)}{\partial \beta} = \sum_{i=1}^{n} \left[\frac{X_i}{\beta} - \frac{h+X_i}{\beta+1}\right].$$

Položím rovnici derivovanou podle β rovnou nule a upravím:

$$\frac{1}{\beta}\sum_{i=1}^{n} X_i - \frac{1}{\beta+1}\sum_{i=1}^{n} X_i - \frac{nh}{\beta+1} = 0,$$

$$\frac{\beta+1}{\beta}\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} X_i - nh = 0,$$

$$\sum_{i=1}^{n} X_i \left(\frac{\beta+1}{\beta} - 1\right) = nh,$$

$$\frac{1}{\beta}\sum_{i=1}^{n} X_i = nh,$$

$$\frac{1}{h\beta} = n\left(\sum_{i=1}^{n} X_i\right)^{-1}.$$

Dostávám:

$$\beta h = \overline{X_n}.$$

Z rovnice derivované podle h:

$$n\log(1+\beta) = \sum_{i=1}^{n}\sum_{m=1}^{X_i}\frac{1}{(h+X_i-m)}.$$

Rovnice se dají vyřešit numericky nahrazením $\hat{\beta} = \frac{\overline{X_n}}{\hat{h}}$ ve druhé rovnici:

$$H(\hat{h}) = n\log\left(1+\frac{\overline{X_n}}{\hat{h}}\right) - \sum_{i=1}^{n}\sum_{m=1}^{X_i}\frac{1}{(\hat{h}+X_i-m)} = 0.$$

Budu řešit numericky Newtonovou metodou. Pro k-tou iteraci mám vztah:

$$h_k = h_{k-1} - \frac{H(h_{k-1})}{H'(h_{k-1})}.$$

Pokud tedy funkce $l_{NegBinom}$ bude mít větší funkční hodnotu než l_{Poiss} a l_{Binom}, hodnoty parametrů budou:

$$a = \frac{\overline{X_n}}{\hat{h}+\overline{X_n}}\ ,\ b = \frac{(\hat{h}-1)\overline{X_n}}{\hat{h}+\overline{X_n}}.$$

3.2.3 MLE pro binomické rozdělení

Nechť $(a,b) \in M_{Binom}$ potom:

$$p_k = \binom{m}{k}p^k(1-p)^{m-k}\ ,\ k = 1,2,\dots\ ,\ 0 < p < 1\ ,\ m \in \mathbb{N}.$$

Protože pro binomické rozdělení $E(N) = mp$ a $Var(N) = mp(1-p)$ budu uvažovat binomické rozdělení jako vhodného kandidáta na odhad parametrů pouze tehdy, když bude výběrový průměr větší než výběrový rozptyl.

Pro soubor dat, který obsahuje n pozorování: $X_1, X_2, \dots, X_n$, o kterých předpokládám, že jsou nezávislé a stejně rozdělené (i.i.d.) náhodné veličiny z binomického rozdělení s p a m. Platí:

$$L_n(p,m) = \prod_{i=1}^{n}\left[\binom{m}{X_i}p^{X_i}(1-p)^{m-X_i}\right].$$

Logaritmická věrohodnostní funkce je:

$$l_n(p,m) = \sum_{i=1}^{n}\left[\log\binom{m}{Xi} + X_i\log(p) + (m-X_i)\log(1-p)\right].$$

Pro maximálně věrohodné odhady m a p:

$$l_{Binom} = \sum_{i=1}^{n}\left[\log\binom{\hat{m}}{Xi} + X_i\log(\hat{p}) + (\hat{m}-X_i)\log(1-\hat{p})\right].$$

Z logaritmické věrohodnostní rovnice s parametry p a m vyjádřím maximálně věrohodný odhad p.

$$\frac{\partial l_n(p,m)}{\partial p} = \sum_{i=1}^{n} \left[\frac{X_i}{p} - \frac{m - X_i}{1-p} \right],$$

Položím rovno 0 a vyřeším pro p:

$$\frac{1}{p}\sum_{i=1}^{n} X_i + \frac{1}{1-p}\sum_{i=1}^{n} X_i - \frac{nm}{1-p} = 0,$$

$$\frac{1-p}{p}\sum_{i=1}^{n} X_i + \sum_{i=1}^{n} X_i = nm,$$

$$\frac{1}{p}\sum_{i=1}^{n} X_i = nm,$$

$$\frac{1}{p} = \frac{nm}{\sum_{i=1}^{n} X_i}.$$

Celkem:

$$\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n\ m} = \frac{\overline{X_n}}{\hat{m}}.$$

Kde $\hat{m}$ získáme tímto způsobem:

- Začnu pro $\hat{m}$, jako s největší napozorovanou hodnotou.
- Spočítám $\hat{p}$.
- Spočítám l_{Binom} pro tyto hodnoty.
- Zvětším $\hat{m}$ o 1.
- Opakuju předešlé kroky, dokud nenajdu lokální maximum. Tedy vyberu poslední takové parametry, pro které předchozí výpočet l_{Binom} měl menší hodnotu.

Pokud bude l_{Binom} větší než l_{Poiss} a $l_{NegBinom}$, tak pro $\hat{a}$ platí:

$$\hat{a} = \frac{\overline{X_n}}{-\hat{m} + \overline{X_n}}.$$

a pro $\hat{b}$:

$$\hat{b} = \frac{\overline{X_n}(\hat{m}+1)}{\hat{m} - \overline{X_n}}.$$

4. Simulační studie

V rámci simulační studie vygeneruji data, jejichž rozdělení budu brát jako neznámé, z Poissonova, negativně binomického a binomického rozdělení. Vypočítám si všechny l_{dist}, $dist \in \{Poiss, Binom, NegBinom\}$ (pokud to bude dávat smysl) pro každá data zvlášť. Vyberu pro každý soubor dat l_{dist}, které bude maximální a odhadnu pro něj parametry a a b. Simulační studii budu dělat v programu Wolfram Mathematica 12.3.

4.1 Data z Poissonova rozdělení

Kód pro tuto podkapitolu je v příloze A.1. Jako první jsem určil vztah mezi výběrovým rozptylem a výběrovým průměrem. Zjistil jsem, že pro vygenerované data je větší výběrový rozptyl, takže jsem se snažil najít hodnoty l_{Poiss} a $l_{NegBinom}$. Abych odhadl l_{Poiss}, tak jsem první musel najít maximálně věrohodný odhad λ. Vyšlo:

$$\hat{\lambda} = 9.972.$$

Celkem pro l_{Poiss}:

$$l_{Poiss} = -2574.47.$$

Pro to abych určil hodnotu $l_{NegBinom}$ musím najít maximálně věrohodný odhad h a β. Vyšlo:

$$\hat{h} = 298.951 \quad , \quad \hat{\beta} = 0.0333566.$$

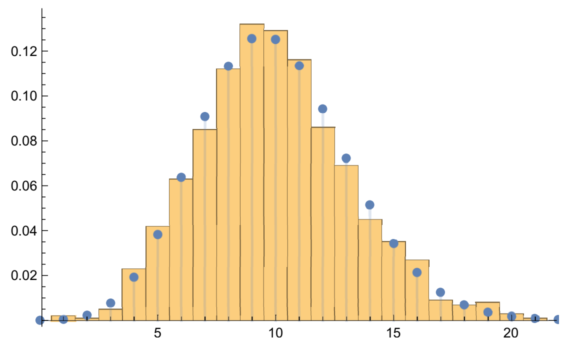
Celkem:

$$l_{NegBinom} = -4522.27.$$

Jelikož l_{Poiss} s odhadem parametru $\hat{\lambda} = 9.972$ má větší hodnotu než $l_{NegBinom}$ spočítám a a b pro M_{Poiss}:

$$\hat{a} = 0 \quad , \quad \hat{b} = 9.972.$$

V grafu porovnám napozorovaná data s odhadnutou pravděpodobnostní funkcí.



Obrázek 4.1: Porovnání odhadu s daty z Poissonova rozdělení

4.2 Data z negativně binomického rozdělení

Kód pro tuto podkapitolu je v příloze A.2. Znovu jsem jako první určil vztah mezi výběrovým rozptylem a výběrovým průměrem. Znovu jsem také zjistil, že pro vygenerované data je větší výběrový rozptyl, takže jsem se snažil najít hodnoty l_{Poiss} a $l_{NegBinom}$. Abych odhadl l_{Poiss}, musel jsem najít maximálně věrohodný odhad λ. Vyšlo:

$$\hat{\lambda} = 8.303.$$

Celkem pro l_{Poiss}:

$$l_{Poiss} = -4321.62.$$

Abych určil hodnotu $l_{NegBinom}$, musím najít maximálně věrohodný odhad h a β. Vyšlo:

$$\hat{h} = 1.95432. \quad , \quad \hat{\beta} = 4.24853.$$

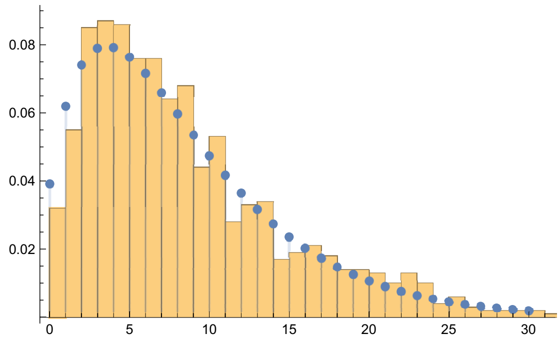
Celkem:

$$l_{NegBinom} = -3102.63.$$

Jelikož $l_{NegBinom}$ s odhadnutými parametry má větší hodnotu než l_{Poiss}, spočítám a a b pro $M_{NegBinom}$:

$$\hat{a} = 0.80947 \quad , \quad \hat{b} = 0.772497.$$

V grafu porovnám napozorovaná data a odhadnutou pravděpodobnostní funkci.



Obrázek 4.2: Porovnání odhadu s daty z negativně binomického rozdělení

4.3 Data z binomického rozdělení

Kód pro tuto podkapitolu je v příloze A.3. Výběrový průměr je tentokrát větší než výběrový rozptyl. Tentokrát jsem tedy hledal hodnoty l_{Poiss} a l_{Binom}. Maximálně věrohodný odhad λ vyšel:

$$\hat{\lambda} = 24.122.$$

Celkem pro l_{Poiss}:

$$l_{Poiss} = -2716.06.$$

Pro maximálně věrohodné odhady m a p vyšlo:

$$\hat{m} = 41 \quad , \quad \hat{p} = 0.588341.$$

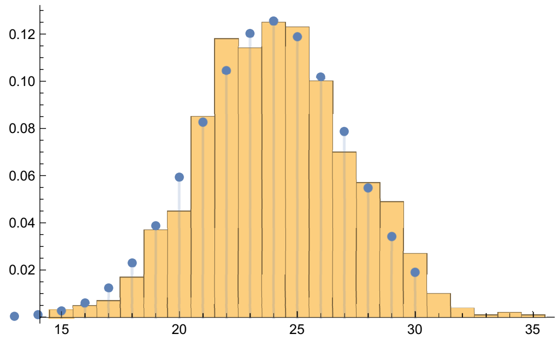
Celkem:

$$l_{Binom} = -2566.74.$$

Jelikož l_{Binom} s odhadnutými parametry má větší hodnotu než l_{Poiss}, spočítám a a b pro M_{Binom}:

$$\hat{a} = -1.4292 \quad , \quad \hat{b} = 60.0263.$$

V grafu porovnám napozorovaná data a odhadnutou pravděpodobnostní funkci.



Obrázek 4.3: Porovnání odhadu s daty z binomického rozdělení

Závěr

V práci jsem zprvu dokázal, že Poissonovo, negativně binomické i binomické rozdělení do třídy $(a, b, 0)$ náleží. Poté jsem provedl úplný rozklad množiny přípustných hodnot parametrů a, b, včetně charakterizace, kterým typům rozdělení jednotlivé podmnožiny náleží. Následně jsem prozkoumal použití metody maximální věrohodnosti pro odhad parametrů a, b. Odvodil jsem logaritmickou věrohodnostní funkci, z níž jsem kvůli složitosti neodhadl parametry a, b a udělal metodu maximální věrohodnosti na množinách přípustných hodnot pro jednotlivá rozdělení. Nakonec jsem udělal simulační studii, ve které jsem využíval předešlých pozorování o odhadech pomocí metody maximální věrohodnosti.

Seznam použité literatury

ANDĚL, J. (2007). *Základy matematické statistiky.* Druhé opravené vydání. Matfyzpress, Praha. ISBN 80-7378-001-1.

KLUGMAN, S. A., PANJER, H. H. a E., W. G. (2019). *Loss Models: From Data to Decisions.* Fifth Edition. John Wiley and Sons, Inc., Hoboken. ISBN 9781119523734.

LANG, S. (1999). *Complex Analysis.* Fourth Edition. Springer Science & Business Media, New York. ISBN 978-1-4757-3083-8.

Seznam obrázků

Seznam tabulek

A. Přílohy

A.1 Kód pro data z Poissonova rozdělení

```
SeedRandom[101]
data = RandomVariate[PoissonDistribution[10], 1000];

Variance[data] > Mean[data]

lPoiss [lambda_] := Sum[-lambda + data[[i]] Log[lambda] -
Log[data[[i]]!], {i, 1, Length[data]}]

lNegBinom[h_, Beta_] := Sum[Log[Binomial[h + data[[i]] - 1, data[[i]]]]
-h Log[Beta + 1] + data[[i]] Log[Beta] - Log[data[[i]] (Beta + 1)],
{i, 1,Length[data]}]

odhadlambda = N[Mean[data]]

m1 = Max[data];

momenthodhadh = N[Mean[data]/( Variance[data]/Mean[data] - 1)]

Newton[H_, h0_, n_] := NestList[# - H[#]/H'[#] &, h0, n]

H[h_] = Length[data] Log[1 + Mean[data]/h] -
   Sum[Sum[1/(h + data[[i]] - m), {m, 1, data[[i]]}], {i, 1,
     Length[data]}];

odhadhseznam = Newton[H, momenthodhadh, 10] // N;

odhadh = odhadhseznam[[10]]

HodnotaPoiss = lPoiss[odhadlambda]

odhadbeta = (1 - (1/(Mean[data]/odhadh + 1)))/(1/(Mean[data]/odhadh +
      1))

HodnotaNegBinom = lNegBinom[odhadh, odhadbeta]

odhadb = N[Mean[data]]

odhada = 0

Show[Histogram[data, Automatic, "Probability"],
 DiscretePlot[
  Table[PDF[PoissonDistribution[\[Lambda]],
     k], {\[Lambda], {9.972}}] // Evaluate, {k, 0, 30},
```

```
  PlotRange -> All, PlotMarkers -> Automatic]]
```

A.2 Kód pro data z negativně binomického rozdělení

```
SeedRandom[101]
data = RandomVariate[NegativeBinomialDistribution[2, 0.2], 1000];

Variance[data] > Mean[data]

lPoiss [lambda_] :=
 Sum[-lambda + data[[i]] Log[lambda] - Log[data[[i]]!], {i, 1,
   Length[data]}]

lNegBinom[h_, Beta_] :=
 Sum[Log[Binomial[h + data[[i]] - 1, data[[i]]]] - h Log[Beta + 1] +
   data[[i]] Log[Beta] - data[[i]] Log[(Beta + 1)], {i, 1,
   Length[data]}]

odhadlambda = N[Mean[data]]

momenthodhadh = N[Mean[data]/( Variance[data]/Mean[data] - 1)]

Newton[H_, h0_, n_] := NestList[# - H[#]/H'[#] &, h0, n]

H[h_] = Length[data] Log[1 + Mean[data]/h] -
   Sum[Sum[1/(h + data[[i]] - m), {m, 1, data[[i]]}], {i, 1,
     Length[data]}];

odhadhseznam = Newton[H, momenthodhadh, 10] // N;

odhadh = odhadhseznam[[10]]

HodnotaPoiss = lPoiss[odhadlambda]

odhadbeta = (1 - (1/(Mean[data]/odhadh + 1)))/(1/(Mean[data]/odhadh +
      1))

HodnotaNegBinom = lNegBinom[2, odhadbeta]

odhada = Mean[data]/(odhadhseznam[[10]] + Mean[data])

odhadb = odhada (odhadhseznam[[10]] - 1)

Show[Histogram[data, Automatic, "Probability"],
```

```
 DiscretePlot[
  Table[PDF[NegativeBinomialDistribution[h, 1/(odhadbeta + 1)],
     k], {h, {1.9543238862436016}}] // Evaluate, {k, 0, 30},
  PlotRange -> All, PlotMarkers -> Automatic]]
```

A.3 Kód pro data z binomického rozdělení

```
SeedRandom[101]
data = RandomVariate[BinomialDistribution[40, 0.6], 1000];

Mean[data] > Variance[data]

lPoiss [lambda_] :=
 Sum[-lambda + data[[i]] Log[lambda] - Log[data[[i]]!], {i, 1,
   Length[data]}]

lBinom[p_, m_] :=
 Sum[Log[Binomial[m, data[[i]]]] +
   data[[i]] Log[p] + (m - data[[i]]) Log[1 - p], {i, 1,
   Length[data]}]

odhadlambda = N[Mean[data]]

m1 = Max[data];

odhadp[m_] := N[Mean[data]/m]

HodnotaPoiss = lPoiss[odhadlambda]

HodnotaBinom =
 N[Maximize[{lBinom[odhadp[odhadm], odhadm], odhadm >= m1,
    odhadm \[Element] Integers}, odhadm]]

odhadm = 41;

odhada = N[Mean[data]/(Mean[data] - odhadm)]

odhadb = N[Mean[data] (odhadm + 1)/(-Mean[data] + odhadm)]

Show[Histogram[data, Automatic, "Probability"],
 DiscretePlot[
  Table[PDF[BinomialDistribution[41, p],
     k], {p, {0.5819513536725999}}] // Evaluate, {k, 0, 30},
  PlotRange -> All, PlotMarkers -> Automatic]]
```