VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA STROJNÍHO INŽENÝRSTVÍ
ÚSTAV MATEMATIKY

FACULTY OF MECHANICAL ENGINEERING
INSTITUTE OF MATHEMATICS

# STATISTICKÁ ANALÝZA SLOŽENÝCH ROZDĚLENÍ

STATISTICAL ANALYSIS OF COMPOUND DISTRIBUTIONS

DIPLOMOVÁ PRÁCE
DIPLOMA THESIS

AUTOR PRÁCE — Bc. ZDENĚK KONEČNÝ
AUTHOR

VEDOUCÍ PRÁCE — doc. RNDr. JAROSLAV MICHÁLEK, CSc.
SUPERVISOR

BRNO 2011

Vysoké učení technické v Brně, Fakulta strojního inženýrství

Ústav matematiky
Akademický rok: 2010/2011

# ZADÁNÍ DIPLOMOVÉ PRÁCE

student(ka): Bc. Zdeněk Konečný

který/která studuje v **magisterském navazujícím studijním programu**

obor: **Matematické inženýrství (3901T021)**

Ředitel ústavu Vám v souladu se zákonem č.111/1998 o vysokých školách a se Studijním a zkušebním řádem VUT v Brně určuje následující téma diplomové práce:

**Statistická analýza složených rozdělení**

v anglickém jazyce:

**Statistical analysis of compound distributions**

Stručná charakteristika problematiky úkolu:

Rozdělení pravděpodobností náhodných veličin, které vzniknou jako součet náhodného počtu nezávislých stejně rozdělených náhodných veličin se nazývají složená rozdělení. V praxi se poměrně často vyskytují při analýze pojišťovacích dat nebo při oceňování ceny poškozených prvků v kontrole jakosti a podobně. V současné době se řada autorů věnuje problematice odhadů parametrů složeného rozdělení, pro speciálně vybraná rozdělení sčítanců a speciální diskrétní rozdělení jejich počtu. Studují se různé metody odhadu parametrů těchto rozdělení a jejich vliv na analýzu složeného rozdělení.

Cíle diplomové práce:

V práci zaveďte složená rozdělení a výpočet jejich charakteristik např. podle [2]. Dále předpokládejte, že jednotliví sčítanci složeného rozdělení mají logaritmicko-normální rozdělení, popište různé přístupy k odhadu jeho parametrů (např. podle [3] a [4]) a dále studujte vliv těchto odhadů na shodu odhadnutého složeného rozdělení s teoretickým složeným rozdělením při různých typech rozdělení počtu sčítanců. Pro rozdělení počtu sčítanců můžete volit např. rozdělení Poissonovo nebo negativně binomické (odhady jejich parametrů můžete zpracovat např. podle [1]). Vliv odhadu parametrů uvažovaného diskrétního rozdělení na shodu odhadnutého a teoretického rozdělení rovněž vyhodnoťte.

Získané výsledky můžete demonstrovat pomocí simulací a případně použít pro analýzu reálných dat.

Seznam odborné literatury:

[1]Lehmann E.L. and Casella G.: Theory of Point Estimation. Springer. New York, 1998
[2]Resnick S.I.: Advantures in Stochastic Processes. Birkhaüser. Boston, 2002
[3] Shen H. and Zhu Z.: Efficient Estimation in Log-normal Linear Models. Journal of Statistical Planning and Inference.138, p.552-567, 2008
[4] Shen H., Brown L.D. and Zhi H..: Efficient Estimation of Log-normal Means with application to Pharmacokinetic Data. Statistics in Medicine 25, p.3023- 3038, 2006

Vedoucí diplomové práce: doc. RNDr. Jaroslav Michálek, CSc.

Termín odevzdání diplomové práce je stanoven časovým plánem akademického roku 2010/2011.

V Brně, dne

L.S.

| | |
|---|---|
| prof. RNDr. Josef Šlapal, CSc.<br>Ředitel ústavu | prof. RNDr. Miroslav Doupovec, CSc.<br>Děkan fakulty |

**Abstrakt**
Složeným rozdělením je nazýváno rozdělení pravděpodobnosti náhodné veličiny, která vznikla jako součet náhodného počtu nezávislých stejně rozdělených náhodných veličin. V této práci je popsáno složené rozdělení spolu s výpočtem jeho charakteristik. Práce se dále zabývá speciálnímu případem složeného rozdělení, jehož jednotlivý sčítanci mají rozdělení logaritmicko-normální (LN) a rozdělení jejich počtu je negativně binomické (NB). Jsou zde popsány i některé přístupy k odhadu parametrů LN a NB rozdělení a dále je studován vliv těchto odhadů na výsledné složené rozdělení.

**Summary**
The probability distribution of a random variable created by summing a random number of the independent and identically distributed random variables is called a compound probability distribution. In this work is described a compound distribution as well as a calculation of its characteristics. Especially, the thesis is focused on studying a special case of compound distribution where each addend has the log-normal distribution and their number has the negative binomial distribution. Here are also described some approaches to estimate the parameters of LN and NB distribution. Further, the impact of these estimates on the final compound distribution is analyzed.

**Klíčová slova**
Složené rozdělení, vytvořující funkce, charakteristická funkce, inverzní věta

**Keywords**
Compound distribution, generating function, characteristic function, inversion theorem

KONEČNÝ, Z.*Statistická analýza složených rozdělení.* Brno: Vysoké učení technické v Brně, Fakulta strojního inženýrství, 2011. 45 s. Vedoucí doc. RNDr. Jaroslav Michálek, CSc.

Prohlašuji, že jsem diplomovou práci *Statistická analýza složených rozdělení* vypracoval samostatně pod vedením doc. RNDr. Jaroslava Michálka, CSc., s použitím materiálů uvedených v seznamu literatury.

Bc. Zdeněk Konečný

Děkuji svému školiteli doc. RNDr. Jaroslavu Michálkovi za odborné vedení, cenné rady a připomínky při zpracování mé diplomové práce.

Bc. Zdeněk Konečný

# Obsah

**1 Úvod** **3**

**2 Základní pojmy a označení** **4**
2.1 Náhodná veličina a její charakteristiky . . . . . 4

**3 Složené rozdělení a jeho charakteristiky** **7**
3.1 Náhodný součet a složené rozdělení . . . . . 7
3.2 Charakteristiky složeného rozdělení . . . . . 7
3.3 Inverzní věta . . . . . 14

**4 Negativně binomické rozdělení** **19**
4.1 Zavedení negativně binomického rozdělení a jeho charakteristiky . . . . . 19
4.2 Způsoby reparametrizace NB rozdělení . . . . . 21
4.3 Odhady parametrů NB rozdělení . . . . . 25
4.3.1 Metoda momentů . . . . . 25
4.3.2 Metoda maximální věrohodnosti . . . . . 26
4.3.3 Bayesovské odhady . . . . . 28

**5 Logaritmicko-normální rozdělení** **29**
5.1 Zavedení logaritmicko-normálního rozdělení a jeho charakteristiky . . . . . 29
5.2 Metody odhadu parametrů LN rozdělení . . . . . 30
5.2.1 Metoda momentů . . . . . 31
5.2.2 Metoda maximální věrohodnosti . . . . . 31

**6 Konkrétní případ složeného rozdělení** **32**
6.1 Určení hustoty složeného rozdělení . . . . . 32
6.2 Vliv odhadu parametrů . . . . . 36
6.2.1 Vliv odhadu parametrů LN rozdělení . . . . . 36
6.2.2 Vliv odhadu parametrů NB rozdělení . . . . . 37

**7 Programy** **40**

**8 Závěr** **42**

**9 Seznam použitých zkratek a symbolů** **44**

# 1 Úvod

Rozdělení pravděpodobnosti náhodných veličin, které vzniknou jako součet náhodného počtu nezávislých stejně rozdělených náhodných veličin se nazývají složená rozdělení. Složená rozdělení se v praxi poměrně často vyskytují při analýze pojišťovacích dat, oceňování ceny poškozených prvků v kontrole jakosti nebo při analýze dešťových srážek. Řada statistiků se v současné době věnuje problematice odhadu parametrů složeného rozdělení, pro speciálně vybraná rozdělení sčítanců a speciální diskrétní rozdělení jejich počtu. Studují se různé metody odhadu těchto rozdělení a jejich vliv na analýzu složeného rozdělení. Například v článcích [12] a [13] se autoři zabývají srážkovým modelem, ve kterém má výskyt přeháněk Poissonovo rozdělení a množství srážek rozdělení exponenciální.

Cílem této práce bylo zavést složené rozdělení a popsat výpočet jeho charakteristik. Dále se práce zabývá speciálním případem složeného rozdělení, jehož sčítanci mají logaritmicko-normální rozdělení a rozdělení jejich počtu je negativně binomické. Práce je rozdělena do osmi kapitol, s jejichž obsahem se nyní podrobněji seznámíme.

Po úvodní kapitole následuje druhá kapitola, ve které je uveden stručný přehled charakteristik náhodné veličiny a zavedeno potřebné značení.

V třetí kapitole je zavedeno složené rozdělení a jsou zde uvedeny definice vytvořující funkce, charakteristické funkce, momentové vytvořující funkce a vytvořující funkce kumulantů a je zde odvozen výpočet těchto charakteristik pro složené rozdělení. Součástí této kapitoly je i inverzní věta a její důkaz.

Čtvrtá kapitola je věnována negativně binomickému rozdělení. Je zde popsáno zavedení negativně binomického rozdělení pomocí Bernoulliovské posloupnosti nezávislých alternativních pokusů. Jsou zde uvedeny různé reparametrizace tohoto rozdělení. Pro všechny reparametrizce byly stanoveny základní chakteristiky (momenty, šikmost, špičatost, charakteristická funkce, ...). Druhá část kapitoly je zaměřena na odhad parametrů NB rozdělení. Nejdříve jsou zde popsány klasické metody odhadu parametrů, tedy metoda momentů a metoda maximální věrohodnosti. Odhady pomocí těchto metod jsou určeny pro všechny reparametrizace. Na závěr je zde uveden bayesovský přístup k odhadu parametrů NB rozdělení.

Logaritmicko-normálnímu rozdělení je věnována celá pátá kapitola. Jsou zde uvedeny jeho základní charakteristiky a určeny momentové a maximálně věrohodné odhady jeho parametrů.

Stěžejní částí této práce je šestá kapitola. Je věnována složenému rozdělení, jehož sčítanci mají logaritmicko-normální rozdělení a rozdělení jejich počtu je negativně binomické. Nejdříve je zde odvozen způsob, jakým lze přibližně určit hustotu zkoumaného složeného rozdělení, známe-li parametry rozdělení sčítanců a jejich počtu. Pomocí simulací náhodného výběru velkého rozsahu z konkrétního složeného rozdělení je posouzena přesnost určené hustoty. Dále je zde pomocí simulací studován vliv odhadu parametrů dílčích rozdělení na shodu odhadnutého složeného rozdělení s teoretickým. Parametry rozdělení sčítanců jsou odhadovány metodou maximální věrohodnosti. Pro odhad parametrů rozdělení počtu sčítanců je volen bayesovský přístup a metoda maximální věrohodnosti.

V sedmé kapitole je uveden přehled a stručný popis funkcí, které byly naprogramovány v softwaru MATLAB. Těchto funkcí bylo užito v šesté kapitole.

Osmá kapitola Závěr obsahuje shrnutí dosažených cílů.

# 2 Základní pojmy a označení

V této kapitole budou připomenuty základní pojmy, kterých bude dále v práci použito. Uvedeme zde základní charakteristiky rozdělení pravděpodobnosti náhodné veličiny. Dále zde budou uvedeny některé důležitě věty, na které se budeme v další části odvolávat. Odstavec je zpracován podle [1],[2],[4],[8] a [10]. Věty jsou v této kapitole uvedeny bez důkazů. Jejich důkazy je možno nalézt v citované literatuře.

V této práci se předpokládá, že je čtenář obeznámen se základy komplexní analýzy a se základními příklady rozdělení diskrétního a spojitého typu.

## 2.1 Náhodná veličina a její charakteristiky

*Pravděpodobnostní prostor* je trojice $(\Omega, \mathcal{A}, \mathrm{P})$, kde $\Omega$ je *prostor elementárních jevů*, $\mathcal{A}$ je nějaká množinová $\sigma$-algebra podmnožin prostoru $\Omega$. Dvojice $(\Omega, \mathcal{A})$ tvoří jevové pole a P je pravděpodobnost na tomto jevovém poli.

Nechť $\mathbb{R}$ značí množinu všech reálných čísel, a $\mathcal{B}$ systém borelovských podmnožin $\mathbb{R}$. Nechť $X(\omega)$ je měřitelná funkce z $(\Omega, \mathcal{A})$ do $(\mathbb{R}, \mathcal{B})$. Pak se $X(\omega)$ nazývá *náhodná veličina* a značí se stručně $X$. Náhodné veličiny budemem značit velkými písmeny z konce abecedy.

Označíme-li dále $[X \in B] = \{\omega \in \Omega : X(\omega) \in B\}$ pro libovolnou množinu $B \subseteq \mathbb{R}$, potom z měřitelnosti funkce $X$ plyne, že $[X \in B] \in \mathcal{A}$ pro každou $B \in \mathcal{B}$ a na $\mathcal{B}$ můžeme zavést indukovanou pravděpodobnost $\mathrm{P}_X$ předpisem $\mathrm{P}_X(B) = \mathrm{P}(X \in B)$, kde pro stručnost píšeme $\mathrm{P}(X \in B)$ místo $\mathrm{P}([X \in B])$. Indukovaná pravděpodobnost $\mathrm{P}_X$ se nazývá *rozdělení pravděpodobnosti náhodné veličiny* $X$. Položíme-li speciálně $B = (-\infty, x\rangle$, dostaneme *distribuční funkci* $F$ náhodné veličiny $X$, tedy

$$F(x) = \mathrm{P}_X\left(-\infty, x\rangle\right) = \mathrm{P}_X(X \leq x), \quad x \in \mathbb{R}.$$

Mezi distribuční funkcí $F$ a rozdělením pravděpodobnosti $\mathrm{P}_X$ je vzájemně jednoznačný vztah, proto můžeme dále při popisu rozdělení pravděpodobnosti náhodné veličiny $X$ pracovat pouze s její distribuční funkcí $F$. Skutečnost, že náhodná veličina $X$ má rozdělení pravděpodobnosti $\mathrm{P}_X$, značíme $X \sim \mathrm{P}_X$ a za $\mathrm{P}_X$ dosadíme podle potřeby symbol příslušného rozdělení.

V matematické statistice jsou významné především dva typy rozdělení pravděpodobnosti:

a) Rozdělení diskrétního typu.
   Náhodná veličina $X$ má *diskrétní rozdělení pravděpodobnosti*, jestliže existuje nejvýše spočetná množina $M \subset \mathbb{R}$ taková, že platí $P_X(M) = 1$. Funkci

$$p(x) = \begin{cases} \mathrm{P}(X = x) & \text{pro } x \in M \\ 0 & \text{pro } x \in \mathbb{R} - M \end{cases}$$

   nazýváme *pravděpodobnostní funkcí* náhodné veličiny $X$ a množinu $M$ *oborem hodnot* náhodné veličiny $X$. Pro diskrétní náhodnou veličinu $X$ s oborem hodnot $M$ a pravděpodobnostní funkcí $p(x)$, budeme v dalším textu používat označení $X \sim (M, p)$.

b) Rozdělení spojitého typu.
Náhodná veličina $X$ má *spojité rozdělení pravděpodobnosti*, existuje-li taková funkce $f(x)$, že platí

$$F(x) = \int_{-\infty}^{x} f(t)dt.$$

V tomto případě je distribuční funkce $F$ absolutně spojitá. Funkce $f$ se nazývá *hustota* rozdělení pravděpodobnosti a je skoro všude (vyhledem k Lebesguově míře) nezáporná. Hustotu příslušnou k dané distribuční funkci budeme značit odpovídajícím stejným malým písmenem.

Informace o rozdělení pravděpodobnosti lze popsat pomocí charakteristik. Jednou z nejvýznamnějších a nejčastěji používaných charakteristik je *střední hodnota* náhodné veličiny $X$, budeme ji značit $EX$. Střední hodnota je definována vztahem

$$EX = \int_{\Omega} X(\omega)d\mathrm{P}(\omega),$$

pokud integrál absolutně konvrguje. Pro výpočetní účely se užívá vzorec

$$EX = \int_{-\infty}^{+\infty} xdF(x),$$

kde integrál napravo je chápán jako Lebesgue-Stieltjesův integrál vzhledem k Lebesgue-Stieltjesově míře $\mu_F$, která je vytvořena distribuční funkcí $F$.

**Věta 2.1 (O střední hodnotě transformované náhodné veličiny)** *Nechť $X$ je náhodná veličina a $F(x)$ její distribuční funkce. Dále nechť $g : \mathbb{R} \to \mathbb{R}$ je borelovská funkce. Pak platí*

$$Eg(X) = \int_{-\infty}^{+\infty} g(x)dF(x),$$

*pokud jeden z integrálů existuje.*

*Má-li náhodná veličina $X$ diskrétní rozdělení $(M, p)$, pak*

$$Eg(X) = \sum_{x \in M} g(x)p(x),$$

*pokud jedna ze stran rovnosti existuje.*

*Má-li náhodná veličina $X$ spojité rozdělení s hustotou $f$, potom*

$$Eg(X) = \int_{-\infty}^{+\infty} g(x)f(x)dx,$$

*pokud jeden z integrálů existuje.*

Další často užívanou charakteristikou rozdělení pravděpodobnosti náhodné veličiny $X$ je rozptyl. Rozptyl budeme značit $DX$ a je definován vztahem

$$DX = E(X - EX)^2,$$

pokud uvedená střední hodnota existuje.

Střední hodnota i rozptyl jsou speciálním případem momentů. Pro libovolné přirozené číslo $r$ definujeme *$r$-tý obecný moment* $\mu'_r$ náhodné veličiny $X$ vztahem

$$\mu'_r = EX^r.$$

Dále *$r$-té centrální momenty* náhodné veličiny $X$ zavádíme vztahem

$$\mu_r = E(X - EX)^r, \qquad \text{pro } r = 0, 1, 2, \ldots.$$

V případě, že některý z výše uvedených integrálů neexistuje, řekneme, že i odpovídající moment neexistuje.

Mezi obecnými a centrálními momenty platí následující vztahy:

$$\begin{aligned} \mu_2 =& \mu'_2 - \mu'^2_1, \\ \mu_3 =& \mu'_3 - 3\mu'_1\mu'_2 + 2\mu'^3_1, \\ \mu_4 =& \mu'_4 - 4\mu'_1\mu'_3 + 6\mu'^2_1\mu'_2 - 3\mu'^4_1. \end{aligned} \tag{2.1}$$

Pomocí centrálních momentů závádíme charakteristiky *šikmost* $\gamma_1$ a *špičatost* $\gamma_2$ vztahy

$$\gamma_1 = \frac{\mu_3}{\mu_2^{\frac{3}{2}}}, \qquad \gamma_2 = \frac{\mu_4}{\mu_2^2}.$$

# 3 Složené rozdělení a jeho charakteristiky

Na začátku této kapitoly zavedeme náhodný součet se složeným rozdělením. Následně se budeme zabývat charakteristikami rozdělení, jež rozdělení jednoznačně určují, což ke konci kapitoly také dokážeme. Uvedeme definice vytvořující funkce, chrakteristické funkce, momentové vytvořující funkce a vytvořující funkce kumulantů. Ukážeme, jak určit tyto charakteristiky náhodného součtu se složeným rozdělením pro případ, kdy známe příslušné charakteristiky jednotlivých sčítanců náhodného součtu.

## 3.1 Náhodný součet a složené rozdělení

**Definice 3.1** Nechť $X_1, X_2, \ldots$ je posloupnost stejně rozdělených náhodných veličin definovaných na pravděpodobnostním prostoru $(\Omega, \mathcal{A}, \mathrm{P})$. Nechť $N \sim (M, p_N)$ je nezáporná celočíselná náhodná veličina a nechť $N$ spolu s $X_1, X_2, \ldots$ tvoří posloupnost nezávislých náhodných veličin. Položme

$$S_N = X_1 + \cdots + X_N \text{ a pro } N = 0 \text{ klademe } S_0 = 0.$$

Potom $S_N$ nazveme *náhodným součtem* a jeho rozdělení se nazývá *složené*.

## 3.2 Charakteristiky složeného rozdělení

**Definice 3.2** Nechť $X \sim (M, p)$ je nezáporná celočíselná náhodná veličina. Pak mocninnou řadu

$$G_X(s) = \sum_{x \in M} p(x) s^x, \qquad s \in \mathbb{C}, |s| \leq 1,$$

nazveme *vytvořující funkcí* náhodné veličiny $X$.

**Tvrzení 3.1** *Nechť $X \sim (M, p)$ je nezáporná celočíselná náhodná veličina. Pak*

$$G_X(s) = E s^X. \tag{3.1}$$

Důkaz: Plyne z věty 2.1.

**Příklad 3.1** *Vytvořující funkce Poissonova rozdělení.* Nechť $X \sim \mathsf{Po}(\lambda)$, potom

$$G_X(s) = \sum_{k=0}^{\infty} p(k) s^k = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} s^k = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda s)^k}{k!} = e^{\lambda(s-1)}.$$

**Příklad 3.2** *Vytvořující funkce geometrického rozdělení.* Nechť $X \sim \mathsf{Ge}(\Theta)$, potom

$$G_X(s) = \sum_{k=0}^{\infty} p(k) s^k = \sum_{k=0}^{\infty} ((1-\Theta)^k \Theta) s^k = \Theta \sum_{k=0}^{\infty} ((1-\Theta)s)^k = \frac{\Theta}{1-(1-\Theta)s}.$$

Z definice vytvořující funkce vyplývá, že vytvořující funkcí je rozdělení náhodné veličiny určeno jednoznačně. Platí totiž

$$p(0) = G_X(0), p(x) = \frac{G_X^{(x)}(0)}{x!} \quad \text{pro} \quad x = 1, 2, \ldots,$$

kde $G_X^{(x)}(s)$ je $x$-tá derivace vytvořující funkce $G_X(s)$. Existence derivací vytvořující funkce $G_X(s)$ plyne z vlastností mocninné řady.

K určení vytvořující funkce náhodného součtu se složeným rozdělením, je zapotřebí nejdříve vědět, jak spočítat vytvořující funkci součtu nezávislých náhodných veličin. To uvádí následující věta.

**Věta 3.1** *Nechť $X_1, \ldots, X_n$ jsou nezávislé nezáporné celočíselné náhodné veličiny, $G_{X_j}(s)$ je vytvořující funkce náhodné veličiny $X_j, j = 1, \ldots, n$. Pak náhodná veličina $S = \sum_{j=1}^n X_j$ má vytvořující funkci*

$$G_S(s) = \prod_{j=1}^n G_{X_j}(s).$$

Důkaz: Přímým výpočtem dostáváme

$$G_S(s) = E(s^S) = E(s^{\sum_{j=1}^n X_j}) = E\Big(\prod_{j=1}^n s^{X_j}\Big) = \prod_{j=1}^n E(s^{X_j}) = \prod_{j=1}^n G_{X_j}(s).$$

**Důsledek 3.1** *Nechť $X_1, \ldots, X_n$ jsou nezávislé stejně rozdělené nezáporné celočíselné náhodné veličiny. Nechť $G_{X_1}$ je vytvořující funkce náhodné veličiny $X_1$. Pak náhodná veličina $S = \sum_{j=1}^n X_j$ má vytvořující funkci*

$$G_S(s) = (G_{X_1}(s))^n.$$

Důkaz: Vytvořující funkce součtu $n$ nezávislých nezáporných celočíselných náhodných veličin je podle věty 3.1 tvaru

$$G_S(s) = \prod_{j=1}^n G_{X_j}(s),$$

kde $G_{X_j}$ je vytvořující funkce náhodné veličiny $X_j, j = 1, \ldots, n$. Jelikož jsou náhodné veličiny $X_1, \ldots, X_n$ navíc stejně rozdělené, jsou si jejich vytvořující funkce rovny. Tedy

$$G_{X_1}(s) = G_{X_2}(s) = \ldots = G_{X_n}(s).$$

Odtud

$$G_S(s) = \prod_{j=1}^n G_{X_1}(s) = (G_{X_1}(s))^n.$$

**Věta 3.2** *Nechť $X_1, X_2, \ldots$ jsou nezávislé stejně rozdělené nezáporné celočíselné náhodné veličiny a nechť $G_X(s)$ je jejich vytvořující funkce. Dále nechť $N \sim (M, p_N)$ je nezáporná celočíselná náhodná veličina, $G_N(s)$ její vytvořující funkce, a nechť $N, X_1, X_2, \ldots$ je posloupnost nezávislých náhodných veličin. Potom náhodný součet $S_N = X_1 + \cdots + X_N$ má vytvořující funkci*

$$G_{S_N}(s) = G_N(G_X(s)).$$

Důkaz: Pro určení vytvořující funkce náhodného součtu vyjdeme ze vztahu (3.1)

$$G_{S_N}(s) = Es^{S_N} = E(Es^{S_N}|N).$$

Užitím důsledku 3.1 obdržíme

$$G_{S_N}(s) = E(G_X(s))^N = G_N(G_X(s)).$$

**Příklad 3.3** *Vytvořující funkce složeného rozdělení.* Nechť $N \sim \mathsf{Po}(\lambda)$ a $X_i \sim \mathsf{Ge}(\Theta), i = 1, 2, \ldots$.

V předchozích příkladech byly určeny vytvořující funkce náhodných veličin s Poissonovým a geometrickým rozdělením. Tedy

$$G_N(s) = e^{\lambda(s-1)}, \quad G_X(s) = \frac{\Theta}{1-(1-\Theta)s},$$

kde $G_X(s)$ je vytvořující funkce náhodných veličin $X_1, X_2, \ldots$. Potom podle věty 3.2 má náhodný součet $S_N = X_1 + \cdots + X_N$ vytvořující funkci

$$G_{S_N}(s) = G_N(G_X(s)) = G_N\left(\frac{\Theta}{1-(1-\Theta)s}\right) = e^{\lambda\left(\frac{\Theta}{1-(1-\Theta)s}-1\right)}.$$

Ve zbylé části odstavce se budeme zabývat dalšími charakteristikami náhodných veličin, jimiž lze složené rozdělení rovněž jednoznačně popsat. Dříve než zavedeme pojem charakteristické funkce, musíme se nejdříve zabývat komplexními náhodnými veličinami, neboť charakteristická funkce náhodné veličiny $X$ je definovaná jako střední hodnota komplexní náhodné veličiny $e^{\mathrm{i}tX}$.

**Definice 3.3** Nechť $X, Y$ jsou reálné náhodné veličiny definované na pravděpodobnostním prostoru $(\Omega, \mathcal{A}, \mathrm{P})$, pak veličinu

$$Z = X + \mathrm{i}Y$$

nazveme *komplexní náhodnou veličinou.*
*Distribuční funkcí* náhodné veličiny $Z$ rozumíme distribuční funkci náhodného vektoru $(X, Y)'$.
*Střední hodnota* náhodné veličiny $Z = X + \mathrm{i}Y$ je definována vztahem

$$EZ = EX + \mathrm{i}EY,$$

pokud střední hodnoty $EX$ a $EY$ existují. Řekneme, že komplexní náhodnné veličiny $Z_1 = X_1 + \mathrm{i}Y_1, Z_2 = X_2 + \mathrm{i}Y_2$ jsou *nezávislé*, jestliže náhodné vektory $(X_1, Y_1)', (X_2, Y_2)'$ jsou nezávislé.

**Poznámka 3.1** Vlastnosti komplexních náhodných veličin jsou analogické vlastnostem (reálných) náhodných veličin.

**Definice 3.4.** *Charakteristickou funkcí* náhodné veličiny $X$ rozumíme funkci

$$\psi_X(t) = E(e^{\mathrm{i}tX}) = E(cos(tX)) + \mathrm{i}E(sin(tX)), \quad t \in \mathbb{R}. \tag{3.2}$$

Má-li náhodná veličina $X$ distribuční funkci $F(x)$, pak podle věty 2.1 pro pro charakteristickou funkci $\psi_X$ dostáváme vztah

$$\psi_X(t) = \int_{-\infty}^{+\infty} e^{\mathrm{i}tx} dF(x).$$

**Věta 3.3 (vlastnosti charakteristické funkce)**

1. *$\psi(0) = 1$ a $|\psi(t)| \leq 1$.*
2. *$\psi(t)$ je stejnoměrně spojitá.*
3. *Pro každé reálné $t$ platí*
$$\psi(-t) = \overline{\psi(t)},$$
*kde $\overline{\psi(t)}$ značí číslo komplexně sdružené k $\psi(t)$.*
4. *Nechť $\psi_X(t)$ je charakteristická funkce náhodné veličiny $X$. Pak náhodná veličina $Y = aX + b$, kde $a, b$ jsou konstanty, má charakteristickou funkci*
$$\psi_Y(t) = e^{\mathrm{i}bt}\psi_X(at).$$

Důkaz: Viz [10] (str. 264).

**Definice 3.5** Existuje-li pro nějaký interval hodnot $z$, obsahující počatek, střední hodnota

$$M_X(z) = E(e^{zX}), \tag{3.3}$$

pak tuto funkci nazveme *momentovou vytvořující funkcí* náhodné veličiny $X$.

**Definice 3.6** *Vytvořující funkci kumulantů* náhodné veličiny $X$ definujeme vztahem

$$K_X(t) = \ln(\psi_X(t)), \quad t \in \mathbb{R}. \tag{3.4}$$

Podobně jako u vytvořující funkce uvedeme i nyní věty, pomocí nichž můžeme určit charakteristickou funkci (resp. momentovou vytvořující funkci, vytvořující funkci kumulantů) náhodného součtu, jestliže známe charakteristickou funkci (resp. momentovou vytvořující funkci, vytvořující funkci kumulantů) jednotlivých sčítanců náhodného součtu.

**Věta 3.4** *Nechť $X_1, \ldots, X_n$ jsou nezávislé náhodné veličiny, $\psi_{X_j}(t)$ je charakteristická funkce náhodné veličiny $X_j, j = 1, \ldots, n$. Pak náhodná veličina $S = \sum_{j=1}^n X_j$ má charakteristickou funkci*

$$\psi_S(t) = \prod_{j=1}^n \psi_{X_j}(t).$$

Důkaz: Přímým výpočtem dostáváme

$$\psi_S(t) = E(e^{\mathrm{i}tS}) = E(e^{\mathrm{i}t\sum_{j=1}^n X_j}) = E(e^{\sum_{j=1}^n \mathrm{i}tX_j}) = E(\prod_{j=1}^n e^{\mathrm{i}tX_j}) = \prod_{j=1}^n E(e^{\mathrm{i}tX_j}) = \prod_{j=1}^n \psi_{X_j}(t).$$

**Důsledek 3.2** *Nechť $X_1, \ldots, X_n$ jsou nezávislé stejně rozdělené náhodné veličiny. Nechť $\psi_{X_1}(t)$ je charakteristická funkce náhodné veličiny $X_1$. Pak náhodná veličina $S = \sum_{j=1}^n X_j$ má charakteristickou funkci*

$$\psi_S(t) = (\psi_{X_1}(t))^n.$$

Důkaz: Charakteristická funkce součtu $n$ nezávislých náhodných veličin je podle věty 3.4 tvaru

$$\psi_S(t) = \prod_{j=1}^n \psi_{X_j}(t),$$

kde $\psi_{X_j}$ je charakteristická funkce náhodné veličiny $X_j, j = 1, \ldots, n$. Jelikož jsou náhodné veličiny $X_1, \ldots, X_n$ navíc stejně rozdělené, jsou si jejich charakteristické funkce rovny. Tedy

$$\psi_{X_1}(t) = \psi_{X_2}(t) = \ldots = \psi_{X_n}(t).$$

Odtud

$$\psi_S(t) = \prod_{j=1}^n \psi_{X_1}(t) = (\psi_{X_1}(t))^n.$$

**Věta 3.5** *Nechť $X_1, X_2, \ldots$ jsou nezávislé stejně rozdělené náhodné veličiny a nechť $\psi_X(t)$ je jejich charakteristická funkce. Dále nechť $N \sim (M, p_N)$ je nezáporná celočíselná náhodná veličina, $G_N(s)$ její vytvořující funkce, a nechť $N, X_1, X_2, \ldots$ je posloupnost nezávislých náhodných veličin. Potom náhodný součet $S_N = X_1 + \cdots + X_N$ má charakteristickou funkci*

$$\psi_{S_N}(t) = G_N(\psi_X(t)).$$

Důkaz: Vyjdeme ze vztahu (3.2)

$$\psi_{S_N}(t) = E(e^{\mathrm{i}tS_N}) = E(Ee^{\mathrm{i}tS_N}|N).$$

Sčítanci náhodného součtu jsou nezávislé stejně rozdělené náhodné veličiny. Pomocí důsledku 3.2 dostáváme

$$\psi_{S_N}(t) = E(\psi_X(t))^N = G_N(\psi_X(t)).$$

**Věta 3.6** *Nechť $X_1, \ldots, X_n$ jsou nezávislé náhodné veličiny, $M_{X_j}(z)$ je momentová vytvořující funkce náhodné veličiny $X_j, j = 1, \ldots, n$. Pak náhodná veličina $S = \sum_{j=1}^n X_j$ má momentovou vytvořující funkci*

$$M_S(z) = \prod_{j=1}^n M_{X_j}(z).$$

Důkaz: Přímým výpočtem dostáváme

$$M_S(z) = E(e^{zS}) = E(e^{z\sum_{j=1}^n X_j}) = E(e^{\sum_{j=1}^n zX_j}) = E(\prod_{j=1}^n e^{zX_j}) = \prod_{j=1}^n E(e^{zX_j}) = \prod_{j=1}^n M_{X_j}(z).$$

**Důsledek 3.3** *Nechť $X_1, \ldots, X_n$ jsou nezávislé stejně rozdělené náhodné veličiny. Nechť $M_{X_1}(z)$ je momentová vytvořující funkce náhodné veličiny $X_1$. Pak náhodná veličina $S = \sum_{j=1}^n X_j$ má momentovou vytvořující funkci*

$$M_S(z) = (M_{X_1}(z))^n.$$

Důkaz: Momentová vytvořující funkce součtu $n$ nezávislých náhodných veličin je podle věty 3.6 tvaru

$$M_S(z) = \prod_{j=1}^{n} M_{X_j}(z),$$

kde $M_{X_j}$ je momentová vytvořující funkce náhodné veličiny $X_j, j = 1, \ldots, n$. Jelikož jsou náhodné veličiny $X_1, \ldots, X_n$ navíc stejně rozdělené, jsou si jejich momentové vytvořující funkce rovny. Tedy

$$M_{X_1}(z) = M_{X_2}(z) = \ldots = M_{X_n}(z).$$

Odtud

$$M_S(z) = \prod_{j=1}^{n} M_{X_1}(z) = (M_{X_1}(z))^n.$$

**Věta 3.7** *Nechť $X_1, X_2, \ldots$ jsou nezávislé stejně rozdělené náhodné veličiny a nechť $M_X(z)$ je jejich momentová vytvořující funkce. Dále nechť $N \sim (M, p_N)$ je nezáporná celočíselná náhodná veličina, $G_N(s)$ její vytvořující funkce, a nechť $N, X_1, X_2, \ldots$ je posloupnost nezávislých náhodných veličin. Potom náhodný součet $S_N = X_1 + \cdots + X_N$ má momentovou vytvořující funkci*

$$M_{S_N}(z) = G_N(M_X(z)).$$

Důkaz: Vyjdeme ze vztahu (3.3)

$$M_{S_N}(z) = E(e^{zS_N}) = E(Ee^{zS_N}|N).$$

Sčítanci náhodného součtu jsou nezávislé stejně rozdělené náhodné veličiny. Pomocí důsledku 3.3 dostáváme

$$M_{S_N}(z) = E(M_X(z))^N = G_N(M_X(z)).$$

**Věta 3.8** *Nechť $X_1, \ldots, X_n$ jsou nezávislé náhodné veličiny, $K_{X_j}(t)$ je vytvořující funkce kumulantů náhodné veličiny $X_j, j = 1, \ldots, n$. Pak náhodná veličina $S = \sum_{j=1}^{n} X_j$ má vytvořující funkci kumulantů*

$$K_S(t) = \sum_{j=1}^{n} K_{X_j}(t).$$

Důkaz: Přímým výpočtem dostáváme

$$K_S(t) = \ln(\psi_S(t)) = \ln(\prod_{j=1}^{n} \psi_{X_j}(t)) = \sum_{j=1}^{n} \ln(\psi_{X_j}(t)) = \sum_{j=1}^{n} K_{X_j}(t).$$

**Důsledek 3.4** *Nechť $X_1, \ldots, X_n$ jsou nezávislé stejně rozdělené náhodné veličiny. Nechť $K_{X_1}(t)$ je vytvořující funkce kumulantů náhodné veličiny $X_1$. Pak náhodná veličina $S = \sum_{j=1}^{n} X_j$ má vytvořující funkci kumulantů*

$$K_S(t) = n \cdot K_{X_1}(t).$$

Důkaz: Vytvořující funkce kumulantů součtu $n$ nezávislých náhodných veličin je podle věty 3.8 rovna

$$K_S(t) = \sum_{j=1}^{n} K_{X_j}(t),$$

kde $K_{X_j}$ je vytvořující funkce kumulantů náhodné veličiny $X_j, j = 1, \ldots, n$. Jelikož jsou náhodné veličiny $X_1, \ldots, X_n$ navíc stejně rozdělené, jsou si jejich vytvořující funkce kumulantů rovny. Tedy

$$K_{X_1}(t) = K_{X_2}(t) = \ldots = K_{X_n}(t).$$

Odtud

$$K_S(t) = \sum_{j=1}^{n} K_{X_1}(t) = n \cdot K_{X_1}(t).$$

**Věta 3.9** *Nechť $X_1, X_2, \ldots$ jsou nezávislé stejně rozdělené náhodné veličiny a nechť $K_X(t)$ je vytvořující funkce kumulantů. Dále nechť $N \sim (M, p_N)$ je nezáporná celočíselná náhodná veličina, $G_N(s)$ její vytvořující funkce, a nechť $N, X_1, X_2, \ldots$ tvoří posloupnost nezávislých náhodných veličin. Potom náhodný součet $S_N = X_1 + \cdots + X_N$ má vytvořující funkci kumulantů.*

$$K_{S_N}(t) = K_N\left(\frac{1}{\mathrm{i}} K_X(t)\right).$$

Důkaz: Označme $\psi_X(t)$ charakteristickou funkci náhodných veličin $X_1, X_2, \ldots$. Vyjdeme ze vztahu (3.4)

$$K_{S_N}(t) = \ln \psi_{S_N}(t) = \ln E(e^{\mathrm{i}tS_N}) = \ln E(Ee^{\mathrm{i}tS_N}|N) = \ln E(\psi_X(t))^N =$$
$$= \ln E(e^{\ln(\psi_X(t))^N}) = \ln E(e^{N \ln \psi_X(t)}) = \ln E(e^{NK_X(t)}) = \ln E(e^{\frac{\mathrm{i}K_X(t)N}{\mathrm{i}}}) = K_N\left(\frac{1}{\mathrm{i}} K_X(t)\right).$$

**Příklad 3.4** Nechť $N \sim \mathsf{Po}(\lambda)$ a $X_i \sim \mathsf{Ex}(\tau), i = 1, 2, \ldots$.

Pro určení charakteristik náhodného součtu $S_N$ je zapotřebí nejprve vypočíst dílčí charakteristiky náhodných veličin $N, X_1, X_2, \ldots$. Charakteristiky příslušící k náhodným veličinám $X_1, X_2, \ldots$ budou značeny indexem $X$.

$$G_N(s) = e^{\lambda(s-1)}, \quad \text{viz příklad 3.1.}$$
$$\psi_N(t) = Ee^{\mathrm{i}tN} = \sum_{k=0}^{\infty} p_N(k) e^{\mathrm{i}tk} = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} e^{\mathrm{i}tk} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{\mathrm{i}t})^k}{k!} =$$
$$= e^{-\lambda} e^{\lambda e^{\mathrm{i}t}} = e^{\lambda(e^{\mathrm{i}t}-1)}$$
$$K_N(t) = \ln \psi_N(t) = \ln e^{\lambda(e^{\mathrm{i}t}-1)} = \lambda(e^{\mathrm{i}t} - 1)$$
$$\psi_X(t) = Ee^{\mathrm{i}tX} = \int_{-\infty}^{+\infty} e^{\mathrm{i}tx} f_X(x) dx = \int_0^{+\infty} \frac{1}{\tau} e^{-\frac{x}{\tau}} e^{\mathrm{i}tx} dx = \frac{1}{\tau} \int_0^{+\infty} e^{x(-\frac{1}{\tau}+\mathrm{i}t)} dx =$$
$$= \frac{1}{\tau} \left[ \frac{1}{\mathrm{i}t - \frac{1}{\tau}} e^{x(-\frac{1}{\tau}+\mathrm{i}t)} \right]_{x=0}^{+\infty} = \frac{1}{\tau} \left( 0 - \frac{1}{\mathrm{i}t - \frac{1}{\tau}} \right) = \frac{1}{1 - \mathrm{i}t\tau}$$
$$K_X(t) = \ln \psi_X(t) = \ln \frac{1}{1 - \mathrm{i}t\tau}$$
$$M_X(z) = Ee^{zX} = \int_{-\infty}^{+\infty} e^{zx} f_X(x) dx = \int_0^{+\infty} \frac{1}{\tau} e^{-\frac{x}{\tau}} e^{zx} dx = \frac{1}{\tau} \int_0^{+\infty} e^{x(-\frac{1}{\tau}+z)} dx =$$

$$= \frac{1}{\tau}\left[\frac{1}{z-\frac{1}{\tau}}e^{x(-\frac{1}{\tau}+z)}\right]_0^{+\infty} = \frac{1}{\tau}\left(0-\frac{1}{z-\frac{1}{\tau}}\right) = \frac{1}{1-z\tau}$$

Potom pro náhodný součet $S_N = X_1 + \ldots + X_N$ platí

$$\begin{aligned}
\psi_{S_N}(t) &= G_N(\psi_X(t)) = G_N\left(\frac{1}{1-\mathrm{i}t\tau}\right) = e^{\lambda(\frac{1}{1-\mathrm{i}t\tau}-1)},\\
M_{S_N}(z) &= G_N(M_X(z)) = G_N\left(\frac{1}{1-z\tau}\right) = e^{\lambda(\frac{1}{1-z\tau}-1)},\\
K_{S_N}(t) &= K_N\left(\frac{K_X(t)}{\mathrm{i}}\right) = K_N\left(\frac{\ln\frac{1}{1-\mathrm{i}t\tau}}{\mathrm{i}}\right) = \lambda(e^{\mathrm{i}\frac{\ln\frac{1}{1-\mathrm{i}t\tau}}{\mathrm{i}}}-1) = \lambda(e^{\ln\frac{1}{1-\mathrm{i}t\tau}}-1) =\\
&= \lambda\left(\frac{1}{1-\mathrm{i}t\tau}-1\right).
\end{aligned}$$

## 3.3 Inverzní věta

V definici 3.4 jsme zavedli pojem charakteristické funkce. Známe-li distribuční funkci $F(x)$ náhodné veličiny $X$, pak její odpovídající charakteristická funkce je určena vztahem

$$\psi_X(t) = \int_{-\infty}^{+\infty} e^{\mathrm{i}tx}dF(x).$$

V tomto odstavci ukážeme, že mezi distribuční a charakteristickou funkcí platí i obrácený vztah. Dokážeme tedy, že každá distribuční funkce je svou charakteristickou funkcí jednoznačně určena. To ovšem znamená, že charakteristickou funkcí je rozdělení pravděpodobnosti náhodné veličiny $X$ jednoznačně zadáno.

V následující větě uvedeme inverzní vzorec, který nám umožňuje při znalosti charakteristické funkce $\psi(t)$ určit přírůstek distribuční funkce $F(x)$ v každém intervalu, jehož koncové body jsou body spojitosti funkce $F(x)$.

**Věta 3.10** *Nechť $\psi(t)$ je charakteristická funkce distribuční funkce $F(x)$ (tj. $\psi(t)$ je charakteristická funkce náhodné veličiny $X$, která má distribuční funkci $F(x)$ a nechť $a, b \in \mathbb{R}, a < b$, jsou body spojitosti distribuční funkce $F(x)$, pak*

$$F(b) - F(a) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\left(\psi(t)\frac{e^{-\mathrm{i}ta}-e^{-\mathrm{i}tb}}{2\mathrm{i}t} - \psi(-t)\frac{e^{\mathrm{i}ta}-e^{\mathrm{i}tb}}{2\mathrm{i}t}\right)dt. \tag{3.5}$$

Důkaz předešlé věty je zpracován podle [10]. K důkazu budeme potřebovat dvě pomocná lemmata.

**Lemma 3.1** *Položme*

$$S(\alpha, T) = \frac{2}{\pi}\int_0^T \frac{\sin(\alpha t)}{t}dt. \tag{3.6}$$

*Pak pro každé reálné $\alpha$ a každé kladné $T$ platí nerovnost*

$$|S(\alpha, T)| \leq 2. \tag{3.7}$$

*Dále platí*

$$\lim_{T\to+\infty} S(\alpha,T) = \frac{2}{\pi}\int_0^{+\infty}\frac{\sin(\alpha t)}{t}dt = \begin{cases} 1 & \text{pro} \quad \alpha>0, \\ 0 & \text{pro} \quad \alpha=0, \\ -1 & \text{pro} \quad \alpha<0, \end{cases} \tag{3.8}$$

*přičemž konvergence je pro libovolné $\delta > 0$ stejnoměrná vzhledem k $|\alpha| \geq \delta$.*

Důkaz: Označme

$$S(x) = \frac{2}{\pi}\int_0^x \frac{\sin u}{u}du.$$

Pro $x = \alpha T$ dostáváme

$$S(\alpha T) = \frac{2}{\pi}\int_0^{\alpha T}\frac{\sin u}{u}du$$

a po substituci $u = \alpha T$ obdržíme

$$S(\alpha T) = \frac{2}{\pi}\int_0^{T}\frac{\sin(\alpha t)}{t}dt.$$

Tedy

$$S(\alpha,T) = S(\alpha T). \tag{3.9}$$

Označme dále

$$c_n = \frac{2}{\pi}\int_{n\pi}^{(n+1)\pi}\frac{\sin v}{v}dv.$$

Substitucí $u = v - n\pi$ dostáváme

$$c_n = \frac{2}{\pi}\int_0^{\pi}\frac{\sin(u+n\pi)}{u+n\pi}du = \frac{2}{\pi}\int_0^{\pi}\frac{\sin u\cos(n\pi)}{u+n\pi}du = (-1)^n\frac{2}{\pi}\int_0^{\pi}\frac{\sin u}{u+n\pi}du,\ n=0,1,2,\dots,$$

takže členy posloupnosti $c_n$ pravidelně střídají znaménko a v absolutní hodnotě klesají. Nyní můžeme $S(x)$ zapsat ve tvaru

$$S(x) = \sum_{k=0}^{n-1} c_k + \frac{2}{\pi}\int_{n\pi}^{x}\frac{\sin u}{u}du \quad \text{pro} \quad n\pi \leq x \leq (n+1)\pi. \tag{3.10}$$

Platí tedy

$$\sum_{k=0}^{n-1} c_k \leq S(x) \leq \sum_{k=0}^{n} c_k \quad \text{pro sudé } n \quad \text{a} \quad n\pi \leq x \leq (n+1)\pi, \tag{3.11}$$

$$\sum_{k=0}^{n} c_k \leq S(x) \leq \sum_{k=0}^{n-1} c_k \quad \text{pro liché } n \quad \text{a} \quad n\pi \leq x \leq (n+1)\pi. \tag{3.12}$$

Z vlastností čísel $c_n$ plyne, že

$$0 \leq S(x) \leq c_0 \quad \text{pro} \quad x \geq 0. \tag{3.13}$$

Dále hodnotu $c_0$ ohraničíme shora. Použijeme zde rozvoje funkce $\frac{\sin u}{u}$ v mocninnou řadu.

$$c_0 = \frac{2}{\pi}\int_0^{\pi}\frac{\sin u}{u}du = \frac{2}{\pi}\int_0^{\pi}\left(1 - \frac{u^2}{3!} + \frac{u^4}{5!} - \cdots\right)du \leq \frac{2}{\pi}\int_0^{\pi}du = 2. \tag{3.14}$$

Dosazením (3.14) do vztahu (3.13) obdržíme

$$0 \leq S(x) \leq 2 \quad \text{pro} \quad x \geq 0, \tag{3.15}$$

a jelikož $S(x)$ je lichá funkce, platí

$$|S(x)| \leq 2 \quad \text{pro} \quad x \in \mathbb{R}. \tag{3.16}$$

Ze vztahů (3.9) a (3.16) dostáváme (3.7). Vztah (3.8) plyne z (3.9) a ze známého vzorce matematické analýzy

$$\lim_{x \to +\infty} S(x) = \frac{2}{\pi} \int_0^{+\infty} \frac{\sin u}{u} du = 1.$$

Stejnoměrná konvergence plyne rovněž z (3.9).

**Lemma 3.2** *Položme pro libovolné $T > 0, z, a, b \in \mathbb{R}$*

$$D(T, z, a, b) = \frac{1}{2\pi} \int_{-T}^{+T} \frac{\sin\left(t(z-a)\right) - \sin\left(t(z-b)\right)}{t} dt \tag{3.17}$$

*a dále*

$$D(z, a, b) = D(+\infty, z, a, b) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{\sin\left(t(z-a)\right) - \sin\left(t(z-b)\right)}{t} dt. \tag{3.18}$$

*Pak pro libovolná $z, a, b \in \mathbb{R}$ a libovolné kladné $T$ platí*

$$|D(T, z, a, b)| \leq 2. \tag{3.19}$$

*Jestliže $a < b$, pak*

$$\lim_{T \to +\infty} D(T, z, a, b) = D(z, a, b) = \begin{cases} 1 & \text{pro} \quad a < z < b, \\ \frac{1}{2} & \text{pro} \quad z = a \quad \text{nebo} \quad z = b, \\ 0 & \text{pro} \quad z < a \quad \text{nebo} \quad z > b, \end{cases} \tag{3.20}$$

*přičemž konvergence je pro libovolné $\delta > 0$ stejnoměrná vzhledem k $|z-a| \geq \delta$ a $|z-b| \geq \delta$.*

Důkaz: Lemma 3.2 je přímým důsledkem lemmatu 3.1. Existuje totiž vztah mezi $D(T, z, a, b)$ a funkcí $S(\alpha, T)$ z lemmatu 3.1. Zřejmě platí, že

$$D(T, z, a, b) = \frac{1}{2}[S(z-a, T) - S(z-b, T)].$$

Nyní se už můžeme zabývat důkazem hlavní věty tohoto odstavce.

Důkaz věty 3.10: Pro charakteristickou funkci $\psi(t)$ platí, že $\psi(-t) = \overline{\psi(t)}$, a podobně pro funkci $e^{-\mathrm{i}ta} - e^{-\mathrm{i}tb}$ máme $e^{-\mathrm{i}ta} - e^{-\mathrm{i}tb} = \overline{e^{\mathrm{i}ta} - e^{\mathrm{i}tb}}$. Pomocí těchto vztahů můžeme vzorec (3.5) upravit na tvar

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} \left( \psi(t) \frac{e^{-\mathrm{i}ta} - e^{-\mathrm{i}tb}}{2\mathrm{i}t} - \psi(-t) \frac{e^{\mathrm{i}ta} - e^{\mathrm{i}tb}}{2\mathrm{i}t} \right) dt =$$

$$= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left( \psi(t) \frac{e^{-\mathrm{i}ta} - e^{-\mathrm{i}tb}}{2\mathrm{i}t} - \overline{\psi(t)} \frac{\overline{e^{-\mathrm{i}ta} - e^{-\mathrm{i}tb}}}{2\mathrm{i}t} \right) dt =$$

$$= \frac{1}{2\pi}\int_{-\infty}^{+\infty} \mathrm{Re}\left\{\psi(t)\frac{e^{-\mathrm{i}ta}-e^{-\mathrm{i}tb}}{\mathrm{i}t}\right\}dt, \tag{3.21}$$

kde symbol $\mathrm{Re}\{z\}$ značí reálnou část komplexního čísla $z$.

Dosazením za

$$\psi(t) = \int_{-\infty}^{+\infty} e^{\mathrm{i}tz}dF(z)$$

do vztahu (3.21) obdržíme

$$\begin{aligned}
&\frac{1}{2\pi}\int_{-\infty}^{+\infty} \mathrm{Re}\left\{\left(\int_{-\infty}^{+\infty} e^{\mathrm{i}tz}dF(z)\right)\frac{e^{-\mathrm{i}ta}-e^{-\mathrm{i}tb}}{\mathrm{i}t}\right\}dt = \\
&= \frac{1}{2\pi}\int_{-\infty}^{+\infty} \mathrm{Re}\left\{\int_{-\infty}^{+\infty}(\cos(zt)+\mathrm{i}\sin(zt))\frac{\cos(-ta)+\mathrm{i}\sin(-ta)-(\cos(-tb)+\mathrm{i}\sin(-tb))}{\mathrm{i}t}\right\}dF(z)dt = \\
&= \frac{1}{2\pi}\int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty} \mathrm{Re}\left\{\frac{\sin(t(z-a))-\sin(t(z-b))}{t}+\frac{\cos(t(z-a))-cos(t(z-b))}{\mathrm{i}t}\right\}dF(z)\right)dt = \\
&= \frac{1}{2\pi}\int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty}\frac{\sin(t(z-a))-\sin(t(z-b))}{t}dF(z)\right)dt.
\end{aligned} \tag{3.22}$$

Pokud by bylo možné v integrálu (3.22) zaměnit pořadí integrace, dostali bychom s pomocí lemmatu 3.2

$$\begin{aligned}
&\frac{1}{2\pi}\int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty}\frac{\sin(t(z-a))-\sin(t(z-b))}{t}dt\right)dF(z) = \\
&\qquad = \int_{-\infty}^{+\infty} D(z,a,b)dF(z) = \int_a^b dF(z) = F(b)-F(a).
\end{aligned} \tag{3.23}$$

Odtud vidíme, že k dokončení důkazu stačí ukázat, že lze v integrálu (3.22) záměnu pořadí integrace provést. Jelikož integrál (3.18) není absolutně konvergentní, není možnost záměny integrace zřejmá. Užitím lemmatu 2.3.2 dostáváme, že rozdíl $D(T,z,a,b)-D(z,a,b)$ pro $T\to\infty$ konverguje k nule stejnoměrně vzhledem ke všem reálným $z$ s výjimkou intervalů $a-\delta < z < a+\delta$ a $b-\delta < z < b+\delta$, kde $\delta$ je libovolné kladné reálné číslo. Současně všude platí nerovnost $|D(T,z,a,b)| \le 2$. Jelikož předpokládáme, že $a$ a $b$ jsou body spojitosti distribuční funkce $F(x)$, platí

$$F(b)-F(a) = \int_{-\infty}^{+\infty} D(z,a,b)dF(z) = \lim_{T\to+\infty}\int_{-\infty}^{+\infty} D(T,z,a,b)dF(z).$$

Dále

$$\begin{aligned}
\int_{-\infty}^{+\infty} D(T,z,a,b)dF(z) &= \int_{-\infty}^{+\infty}\frac{1}{2\pi}\left(\int_{-T}^{+T}\frac{\sin\left(t(z-a)\right)-\sin\left(t(z-b)\right)}{t}dt\right)dF(z) = \\
&= \frac{1}{2\pi}\int_{-T}^{+T}\left(\int_{-\infty}^{+\infty}\frac{\sin\left(t(z-a)\right)-\sin\left(t(z-b)\right)}{t}dF(z)\right)dt.
\end{aligned} \tag{3.24}$$

Záměnu integrace v (3.24) bylo možné provést, jelikož je poslední integrál v oblasti $z\in\mathbb{R}, |t|\le T$ absolutně konvergentní. Shrneme-li tyto výsledky, pak limitním přechodem pro $T\to+\infty$ dostaneme

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty}\frac{\sin(t(z-a))-\sin(t(z-b))}{t}dF(z)\right)dt =$$
$$= \lim_{T\to+\infty}\int_{-\infty}^{+\infty} D(T,z,a,b)dF(z) = \int_{-\infty}^{+\infty} D(z,a,b)dF(z) =$$
$$= \frac{1}{2\pi}\int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty}\frac{\sin(t(z-a))-\sin(t(z-b))}{t}dt\right)dF(z),$$

což zbývalo dokázat.

Následující věta je důsledkem věty 3.10.

**Věta 3.11** *Každá distribuční funkce je svou charakteristickou funkcí jednoznačně určena.*

Důkaz: Ve větě 3.10 je uvedeno, že pomocí charakteristické funkce $\psi(t)$ je určen přírůstek distribuční funkce $F(b)-F(a)$ ve všech bodech spojitosti $a, b$ distribuční funkce $F(x)$, $a < b$. Jelikož každá distribuční funkce má nejvýše spočetně mnoho bodů nespojitosti, lze vybrat posloupnost bodů spojitosti $a_n, n = 1, 2, \ldots$ takovou, že

$$a_n < a_{n-1}, \qquad \lim_{n\to\infty} a_n = -\infty.$$

Vzhledem k tomu, že $F(x)$ je neklesající funkce, platí

$$\lim_{n\to\infty} F(a_n) = \lim_{t\to-\infty} F(t) = 0.$$

Tímto limitním přechodem získáme hodnotu $F(b)$ v libovolném bodě spojitosti $b$ distribuční funkce $F(x)$. Protože každá distribuční funkce je zprava spojitá, můžeme její hodnoty v bodech nespojitosti jednoznačně určit tím způsobem, že ji limitním přechodem doplníme na funkci zprava spojitou.

# 4 Negativně binomické rozdělení

V závěrečné části práce se budeme zabývat složeným rozdělením, jehož sčítanci mají logaritmicko-normální rozdělení a rozdělení jejich počtu je negativně binomické, proto bude tato kapitola věnována právě negativně binomickému rozdělení.

## 4.1 Zavedení negativně binomického rozdělení a jeho charakteristiky

Negativně binomické rozdělení je jedno ze základních rozdělení pravděpodobnosti diskrétního typu. Nejčastější způsob zavedení NB rozdělení vychází z Bernoulliovské posloupnosti nezávislých alternativních pokusů, kdy pravděpodobnost úspěchu v každém pokusu je $\pi \in (0,1)$. Rozdělení náhodné veličiny $X$, která udává počet neúspěšných pokusů předcházejících $\kappa$-tému úspěchu, $\kappa \in \{1, 2, \ldots\}$, se nazývá negativně binomické rozdělení s parametry $\pi$ a $\kappa$. Pro jeho pravděpodobnostní funkci (hustotu vzhledem k čítací míře) tedy platí:

$$\begin{aligned} f(x;\pi,\kappa) &= \binom{x+\kappa-1}{x}(1-\pi)^x\pi^\kappa && \text{pro } x = 0, 1, \ldots \\ &= 0 && \text{jinak.} \end{aligned} \tag{4.1}$$

Výše popsané rozdělení se v některé literatuře nazývá Pascalovo a jako NB rozdělení pak bývá označováno zobecnění tohoto rozdělení pro $\kappa > 0$. V této práci budeme předpokládat, že pro parametry NB rozdělení platí $0 < \pi < 1$ a $\kappa > 0$. Je-li $\kappa = 1$, pak dostáváme geometrické rozdělení.

V dalším textu budeme pro náhodnou veličinu $X$ mající negativně binomické rozdělení tvaru (4.1) používat označení $X \sim \mathsf{NB}(\pi, \kappa)$. Příklady hustot NB rozdělení pro různé hodnoty parametrů $\pi$ a $\kappa$ jsou na obr. 4.1.

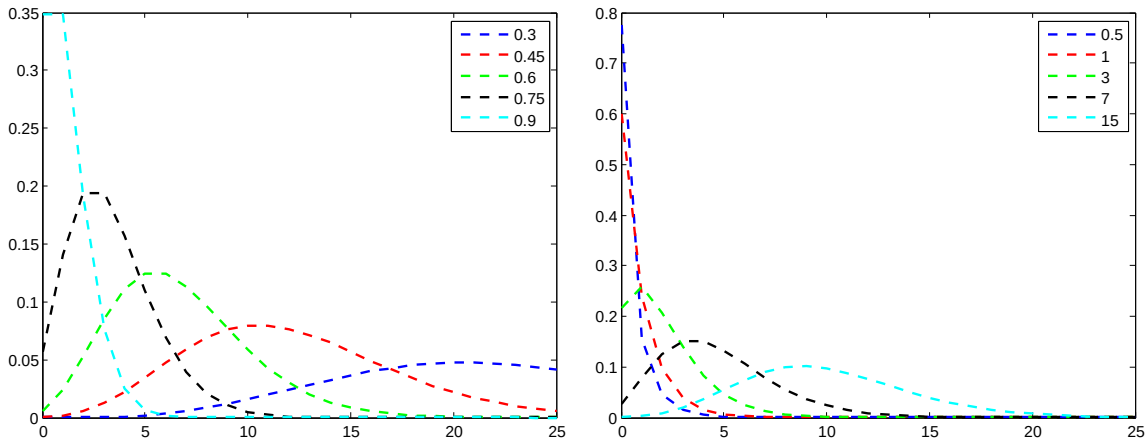
Hustotu (4.1) lze snadno přepsat do tvaru

$$\begin{aligned} f(x;\pi,\kappa) &= \binom{-\kappa}{x}\pi^\kappa(\pi-1)^x && \text{pro } x = 0, 1, \ldots \\ &= 0 && \text{jinak,} \end{aligned}$$

který připomíná pravděpodobnostní funkci binomického rozdělení. Právě z této podobnosti pochází název NB rozdělení.

Pro střední hodnotu $\mu$ a rozptyl $\sigma^2$ platí

$$\mu = EX = \kappa\frac{1-\pi}{\pi} \qquad \text{a} \qquad \sigma^2 = DX = \kappa\frac{1-\pi}{\pi^2}. \tag{4.2}$$




(a) $\kappa = 10$, $\pi$ postupně $0,3; 0,45; 0,6; 0,75; 0,9$

(b) $\pi = 0,6$, $\kappa$ postupně $0,5; 1; 3; 7; 15$

Obrázek 4.1: Pravděpodobnostní funkce NB rozdělení pro různé hodnoty parametrů $\pi$ a $\kappa$. Přestože se jedná o rozdělení diskretního typu, jsou pro větší názornost hustoty zakresleny jako spojité funkce.

Pomocí hustoty (4.1) a vzorců (3.1),(3.2),(3.3) a (3.4) lze snadno určit, že vytvořující funkce $G(s)$, charakteristická funkce $\psi(t)$, momentová vyvořující funkce $M(z)$ a vytvořující funkce kumulantů $K(t)$ negativně binomického rozdělení jsou tvaru

$$\begin{aligned} G(s) &= \left(\frac{\pi}{1-(1-\pi)s}\right)^{\kappa}, \\ \psi(t) &= \left(\frac{\pi}{1-(1-\pi)e^{\mathrm{i}t}}\right)^{\kappa}, \\ M(z) &= \left(\frac{\pi}{1-(1-\pi)e^{z}}\right)^{\kappa}, \\ K(t) &= \kappa \ln \frac{\pi}{1-(1-\pi)e^{\mathrm{i}t}}. \end{aligned} \tag{4.3}$$

Dále uvedeme vzorce pro obecné momenty až do čtvrtého řádu (viz [3]).

Obecné momenty:

$$\begin{aligned} \mu_1' &= EX = \frac{\kappa(1-\pi)}{\pi}, \\ \mu_2' &= EX^2 = \frac{\kappa(1-\pi)}{\pi^2}[\kappa(1-\pi)+1], \\ \mu_3' &= EX^3 = \frac{\kappa(1-\pi)}{\pi^3}[\kappa^2(1-\pi)^2+3\kappa(1-\pi)+2-\pi], \\ \mu_4' &= EX^4 = \frac{\kappa(1-\pi)}{\pi^4}[\kappa^3(1-\pi)^3+6\kappa^2(1-\pi)^2+\kappa(1-\pi)(11-4\pi)+\pi^2-6\pi+6]. \end{aligned}$$

Pomocí vzorců (2.1) lze snadno určit centrální momenty $\mu_2$, $\mu_3$ a $\mu_4$. Známe-li tyto momenty, lze pak nalézt šikmost a špičatost. Jejich výpočty zde nebudeme uvádět, uvedem pouze výsledné vzorce.

Centrální momenty:

$$\begin{aligned}
\mu_2 =& E(X-EX)^2 = DX = \frac{\kappa(1-\pi)}{\pi^2},\\
\mu_3 =& E(X-EX)^3 = \frac{\kappa(1-\pi)(2-\pi)}{\pi^3},\\
\mu_4 =& E(X-EX)^4 = \frac{\kappa(1-\pi)[3\kappa(1-\pi)+\pi^2-6\pi+6]}{\pi^4}.
\end{aligned}$$

Šikmost:

$$\gamma_1 = \frac{\mu_3}{\mu_2^{\frac{3}{2}}} = \frac{2-\pi}{\sqrt{\kappa(1-\pi)}}.$$

Špičatost:

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3 = \frac{\pi^2-6\pi+6}{\kappa(1-\pi)}.$$

Na zápis hustoty NB rozdělení pomocí parametrů $\kappa$ a $\pi$ bude v dalším textu odkazováno jako na parametrizaci 1.

## 4.2 Způsoby reparametrizace NB rozdělení

V analýzach je velmi často zapotřebí testovat hypotézy, jež se týkají střední hodnoty. Proto se velmi často používá reparametrizace s parametry $\mu = EX$ a $\kappa$.

Ze vzorce pro střední hodnotu NB rozdělení o parametrech $\kappa$ a $\pi$, lze vyjádřit parametr $\pi$ ve tvaru

$$\pi = \frac{\kappa}{\kappa+\mu}.$$

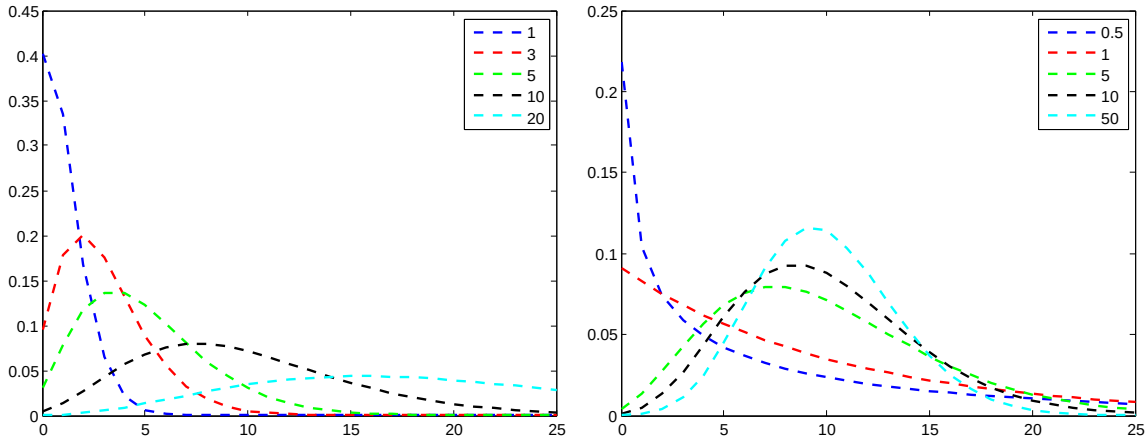
Dosazením do vztahu pro hustotu (4.1) získáme vyjádřní hustoty v nových parametrch $\mu$ a $\kappa$.

$$\begin{aligned}
f(x;\mu,\kappa) &= \binom{x+\kappa-1}{x}\left(\frac{\mu}{\kappa+\mu}\right)^x\left(\frac{\kappa}{\kappa+\mu}\right)^\kappa && \text{pro } x = 0,1,\ldots\\
&= 0 && \text{jinak.}
\end{aligned}$$

Pokud v předchozím vyjádření hustoty navíc zapíšeme binomický koeficient pomocí gama funkce, dostaneme pravděpodobnostní funkci negativně binomického rozdělení ve tvatu

$$\begin{aligned}
f(x;\mu,\kappa) &= \frac{\Gamma(x+\kappa)}{\Gamma(x+1)\Gamma(\kappa)}\left(\frac{\mu}{\kappa+\mu}\right)^x\left(\frac{\kappa}{\kappa+\mu}\right)^\kappa && \text{pro } x = 0,1,\ldots\\
&= 0 && \text{jinak.}
\end{aligned}$$

Pro náhodnou veličinu $X$ mající negativně binomické rozdělení s parametry $\mu$ a $\kappa$ budeme dále použivat označení $X \sim \mathsf{NB}(\mu,\kappa)$.




(a) $\kappa = 5$, $\mu$ postupně $1; 3; 5; 10; 20$

(b) $\mu = 10$, $\kappa$ postupně $0,5; 1; 5; 10; 50$

Obrázek 4.2: Pravděpodobnostní funkce NB rozdělení pro různé hodnoty parametrů $\mu$ a $\kappa$. Přestože se jedná o rozdělení diskretního typu, jsou pro větší názornost hustoty zakresleny jako spojité funkce.

Na obrázku 4.2 jsou příklady hustot NB rozdělení pro různé hodnoty parametrů $\mu$ a $\kappa$. Z obrázků je patrné, že s rostoucím parametrem $\mu$ roste i rozptyl a rozdělení se stává „plošší“, s rostoucím parametrem $\kappa$ se mění tvar rozdělení- rozdělení se stává „symetričtější“.

V této parametrizaci NB rozdělení je $EX = \mu$ a $DX = \mu + \frac{\mu^2}{\kappa}$. Ze vzorce pro rozptyl je patrné, že při daném $\kappa$ je rozptyl kvadratickou funkcí střední hodnoty $\mu$.

Pro tuto parametrizaci dostáváme užitím vzorců (3.1),(3.2),(3.3) a (3.4) vytvořující funkci, charakteristickou funkci, momentovou vyvořující funkci a vytvořující funkci kumulantů ve tvaru

$$
\begin{aligned}
G(s) &= \left(\frac{\kappa}{\kappa + \mu(1-s)}\right)^{\kappa}, \\
\psi(t) &= \left(\frac{\kappa}{\kappa + \mu(1-e^{\mathrm{i}t})}\right)^{\kappa}, \\
M(z) &= \left(\frac{\kappa}{\kappa + \mu(1-e^{z})}\right)^{\kappa}, \\
K(t) &= \kappa \ln \frac{\kappa}{\kappa + \mu(1-e^{\mathrm{i}t})}.
\end{aligned}
$$

Obecné momenty dostaneme ve tvaru

$$\begin{aligned}
\mu_1' =& \mu, \\
\mu_2' =& \mu + \mu^2 + \frac{\mu^2}{\kappa}, \\
\mu_3' =& \mu + 3\mu^2 + 3\frac{\mu^2}{\kappa} + \mu^3 + 3\frac{\mu^3}{\kappa} + 2\frac{\mu^3}{\kappa^2}, \\
\mu_4' =& \mu + 7\mu^2 + 7\frac{\mu^2}{\kappa} + 6\mu^3 + 18\frac{\mu^3}{\kappa} + 12\frac{\mu^3}{\kappa^2} + \mu^4 + 6\frac{\mu^4}{\kappa} + 11\frac{\mu^4}{\kappa^2} + 6\frac{\mu^4}{\kappa^3}
\end{aligned}$$

a pro centrální momenty platí

$$\begin{aligned}
\mu_2 =& \frac{\mu(\kappa+\mu)}{\kappa}, \\
\mu_3 =& \frac{\mu(\kappa+\mu)(\kappa+2\mu)}{\kappa^2}, \\
\mu_4 =& \frac{\mu(\kappa+\mu)(\kappa^2+6\kappa\mu+6\mu^2)}{\kappa^3}.
\end{aligned}$$

Dále pro šikmost a špičatost dostaneme vztahy

$$\begin{aligned}
\gamma_1 &= \frac{\kappa+2\mu}{\sqrt{\kappa\mu(\kappa+\mu)}}, \\
\gamma_2 &= \frac{\kappa^2+6\kappa\mu+6\mu^2}{\kappa\mu(\kappa+\mu)}.
\end{aligned}$$

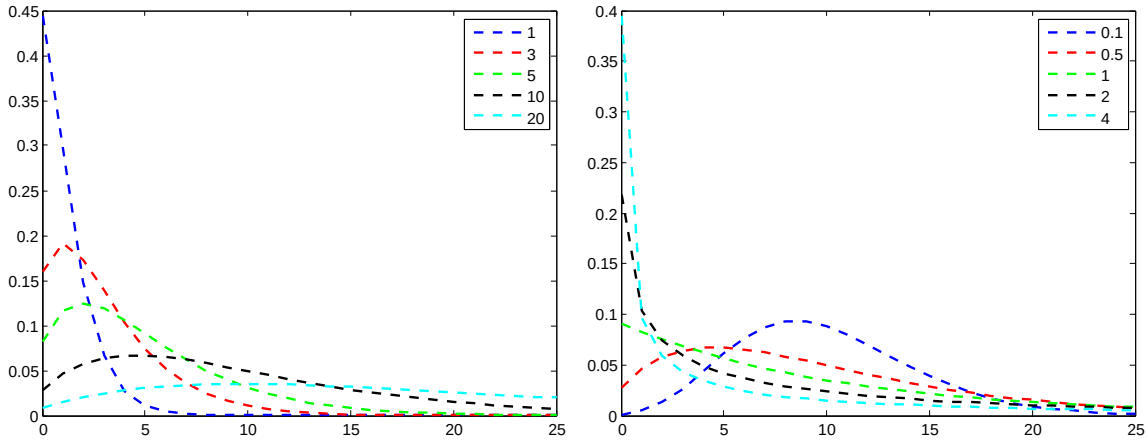
Na zápis hustoty NB rozdělení pomocí parametrů $\mu$ a $\kappa$ bude v dalším textu odkazováno jako na parametrizaci 2.

V některých situacích je obtížné nálézt odhad parametru $\kappa$ (zejména pro velké $\kappa$ a malé $\mu$), proto se zavádí reparametrizace $\mu$, $c$, kde $c = \frac{1}{\kappa}$. Pravděpodobnostní funkce je potom tvaru

$$\begin{aligned}
f(x;\mu,\kappa) &= \frac{\Gamma(x+c^{-1})}{\Gamma(x+1)\Gamma(c^{-1})}\left(\frac{c\mu}{1+c\mu}\right)^x\left(\frac{1}{1+c\mu}\right)^{\frac{1}{c}} && \text{pro } x = 0, 1, \ldots \\
&= 0 && \text{jinak.}
\end{aligned}$$

Pro náhodnou veličinu $X$ mající negativně binomické rozdělení s parametry $\mu$ a $c$ budeme v dalším textu používat označení $X \sim \mathsf{NB}(\mu, c)$.

Příklady hustot NB rozdělení pro různé hodnoty parametrů $\mu$ a $c$ jsou na obr. 4.3.




(a) $c = 0,5$, $\mu$ postupně $1; 3; 5; 10; 20$ (b) $\mu = 10$, $c$ postupně $0,1; 0,5; 1; 2; 4$

Obrázek 4.3: Pravděpodobnostní funkce NB rozdělení pro různé hodnoty parametrů $\mu$ a $c$. Přestože se jedná o rozdělení diskretního typu, jsou pro větší názornost hustoty zakresleny jako spojité funkce.

Vytvořující funkce, charakteristická funkce, momentová vyvořující funkce a vytvořující funkce kumulantů jsou potom tvaru

$$
\begin{aligned}
G(s) &= \left(\frac{1}{1 + c\mu(1-s)}\right)^{\frac{1}{c}}, \\
\psi(t) &= \left(\frac{1}{1 + c\mu(1-e^{\mathrm{i}t})}\right)^{\frac{1}{c}}, \\
M(z) &= \left(\frac{1}{1 + c\mu(1-e^{z})}\right)^{\frac{1}{c}}, \\
K(t) &= \frac{1}{c} \ln \frac{1}{1 + c\mu(1-e^{\mathrm{i}t})}.
\end{aligned}
$$

Pro obecné momenty platí

$$
\begin{aligned}
\mu_1' &= \mu, \\
\mu_2' &= \mu + \mu^2 + \mu^2 c, \\
\mu_3' &= \mu + 3\mu^2 + 3\mu^2 c + \mu^3 + 3\mu^3 c + 2\mu^3 c^2, \\
\mu_4' &= \mu + 7\mu^2 + 7\mu^2 c + 6\mu^3 + 18\mu^3 c + 12\mu^3 c^2 + \mu^4 + 6\mu^4 c + 11\mu^4 c^2 + 6\mu^4 c^3.
\end{aligned}
$$

Pro příslušné centrální momenty dostáváme vztahy

$$
\begin{aligned}
\mu_2 &= \mu(1 + c\mu), \\
\mu_3 &= \mu(1 + c\mu)(1 + 2c\mu), \\
\mu_4 &= \mu(1 + c\mu)(1 + 6c\mu + 6c^2\mu^2).
\end{aligned}
$$

Šikmost a špičatost dostáváme ve tvaru

$$\gamma_1 = \frac{1+2c\mu}{\sqrt{\mu(1+c\mu)}},$$
$$\gamma_2 = \frac{1+6c\mu+6c^2\mu^2}{\mu(1+c\mu)}.$$

Na zápis hustoty NB rozdělení pomocí parametrů $\mu$ a $c$ bude v dalším textu odkazováno jako na parametrizaci 3.

## 4.3 Odhady parametrů NB rozdělení

V tomto odstavci popíšeme různé přístupy k odhadu parametrů negativně binomického rozdělení. Nejdříve budou uvedeny klasické metody odhadu parametrů NB rozdělení- metoda momentů (MM) a metoda maximální věrohodnosti (MMV). Na závěr bude uveden Bayesovský přístup k odhadu parametrů, který je vhodné použít zejména v případech, kdy klasické metody odhadu selhávají.

### 4.3.1 Metoda momentů

Předpokládejme, že je dán náhodný výběr $X_1, \ldots, X_n$ z NB rozdělení. Metoda momentů spočívá v porovnání teoretických a výběrových obecných momentů. Hledané odhady tak snadno získáme řešením momentových rovnic

$$\mu'_k = M'_k, \qquad \text{kde} \quad M'_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k.$$

Nechť $X_i \sim \mathsf{NB}(\pi, \kappa)$. Pro tuto parametrizaci obdržímeme momentové rovnice ve tvaru

$$\frac{\kappa(1-\pi)}{\pi} = \overline{X}$$
$$\frac{\kappa(1-\pi)}{\pi^2}[\kappa(1-\pi)+1] = \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

Jejich řešením dostáváme odhady

$$\hat{\pi} = \frac{\overline{X}}{M_2} \qquad \hat{\kappa} = \frac{\overline{X}^2}{M_2 - \overline{X}},$$

kde $M_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ je druhý centrální výběrový moment.

Nyní uvedeme odhady parametrů NB rozdělení metodou momentů pro ostatní parametrizace.

Nechť $X_i \sim \mathsf{NB}(\mu, \kappa)$, pak

$$\hat{\mu} = \overline{X} \qquad \hat{\kappa} = \frac{\overline{X}^2}{M_2 - \overline{X}}.$$

Nechť $X_i \sim \mathsf{NB}(\mu, c)$, pak

$$\hat{\mu} = \overline{X} \qquad \hat{c} = \frac{M_2 - \overline{X}}{\overline{X}^2}.$$

Odhady získané touto metodou bývají často používány jako počáteční odhady při iterativním řešení nelineárních rovnic, které často vycházejí při hledání maximálně věrohodných odhadů.

Z výrazů pro odhady $\hat{\kappa}$ a $\hat{c}$ je patrné, že pro $M_2 < \overline{X}$ tyto odhady vycházejí záporné, což může působit interpretační a především numerické problémy, například při automatickém použití těchto odhadů jako počátečních odhadů při iterativním hledání řešení věrohodnostních rovnic.

### 4.3.2 Metoda maximální věrohodnosti

Metoda maximální věrohodnosti je pravděpodobně nejpoužívanější metodou určování bodových odhadů. Cílem MMV je odhadnou vektor parametrů $\boldsymbol{\theta}$ tak, aby při každé hodnotě náhodného výběru $X_1 = x_1, \ldots, X_n = x_n$ byla maximalizována věrohodnostní funkce $L(\boldsymbol{\theta}, \mathbf{x})$, kde $\mathbf{x} = (x_1, \ldots, x_n)$, jako funkce $\boldsymbol{\theta}$. Věrohodnostní funkce $L(\boldsymbol{\theta}, \mathbf{x})$ je rovna sdružené hustotě náhodného výběru $X_1, \ldots, X_n$. K hledání maximálně věrohodných odhadů se obvyklé místo věrohodnostní funkce používá logaritmická věrohodnostní funkce $l(\boldsymbol{\theta}, \mathbf{x}) = \ln L(\boldsymbol{\theta}, \mathbf{x})$, která nabývá svého maxima ve stejném bodě jako $L(\boldsymbol{\theta}, \mathbf{x})$.

Nyní stanovíme maximálně věrohodné odhady pro všechny tři parametrizace negativně binomického rozdělení.

- Nechť $X_i \sim \mathsf{NB}(\pi, \kappa)$, pak příslušná věrohodnostní funkce je tvaru

$$L(\pi, \kappa; \mathbf{x}) = \prod_{i=1}^{n} \left[ \frac{\Gamma(x_i + \kappa)}{\Gamma(x_i + 1)\Gamma(\kappa)} (1 - \pi)^{x_i} \pi^{\kappa} \right].$$

  Logaritmováním věrohodnostní funkce dostaneme logaritimickou věrohodnostní funkci

$$l(\pi, \kappa; \mathbf{x}) = \sum_{i=1}^{n} [x_i \ln(1 - \pi) + \kappa \ln \pi + \ln \Gamma(x_i + \kappa) - \ln \Gamma(\kappa) - \ln \Gamma(x_i + 1)].$$

  Derivace logaritmické věrohodnostní funkce podle jednotlivých parametrů položíme rovny nule a dostaneme tak systém věrohodnostních rovnic pro hledané odhady

$$\frac{\partial l}{\partial \pi} = -\frac{\sum_{i=1}^{n} x_i}{1 - \pi} + n\frac{\kappa}{\pi} = 0$$

$$\frac{\partial l}{\partial \kappa} = n \ln \pi + \sum_{i=1}^{n} \Psi(x_i + \kappa) - n\Psi(\kappa) = 0,$$

kde $\Psi(z)$ je digama funkce (derivace logaritmu gama funkce). Z první rovnice snadno získáme odhad

$$\hat{\pi} = \frac{\kappa}{\kappa + \overline{X}},$$

druhá rovnice se řeší iterativně.

- Nechť $X_i \sim \mathsf{NB}(\mu, \kappa)$, pak pro logaritmickou věrohodnostní funkci platí

$$l(\mu, \kappa; \mathbf{x}) = \sum_{i=1}^{n} \left[ x_i \ln \frac{\mu}{\kappa + \mu} + \kappa \ln \frac{\kappa}{\kappa + \mu} + \ln \Gamma(x_i + \kappa) - \ln \Gamma(\kappa) - \ln \Gamma(x_i + 1) \right].$$

Derivováním logaritmické věrohodností funkce obdržíme věrohodnostní rovnice

$$\frac{\partial l}{\partial \mu} = \frac{\kappa}{(\kappa + \mu)} \sum_{i=1}^{n} x_i - n \frac{\kappa}{\kappa + \mu} = 0$$
$$\frac{\partial l}{\partial \kappa} = \sum_{i=1}^{n} \left[ -x_i \frac{1}{\kappa + \mu} + \ln \frac{\kappa}{\kappa + \mu} + \frac{\mu}{\kappa + \mu} + \Psi(x_i + \kappa) - \Psi(\kappa) \right] = 0.$$

Z první rovnice lze jednoduše výjádřit parametr $\mu$, tím dostaneme maximálně věrohodný odhad

$$\hat{\mu} = \overline{X}.$$

Odhad parametru $\kappa$ je nutno hledat numericky.

- Nechť $X_i \sim \mathsf{NB}(\mu, c)$, pak logaritmická věrohodnostní funkce je tvaru

$$l(\mu, c; \mathbf{x}) = \sum_{i=1}^{n} \left[ x_i \ln \frac{c\mu}{1 + c\mu} + \frac{1}{c} \ln \frac{1}{1 + c\mu} + \ln \Gamma(x_i + c^{-1}) - \ln \Gamma(c^{-1}) - \ln \Gamma(x_i + 1) \right].$$

Příslušné věrohodnostní rovnice jsou tvaru

$$\frac{\partial l}{\partial \mu} = \frac{1}{(1 + c\mu)} \sum_{i=1}^{n} x_i - \frac{n}{1 + c\mu} = 0$$
$$\frac{\partial l}{\partial c} = \sum_{i=1}^{n} \left[ \frac{x_i - \mu}{c(1 + c\mu)} + \frac{1}{c^2} \ln(1 + c\mu) - \frac{1}{c^2} \Psi(x_i + c^{-1}) + \frac{1}{c^2} \Psi(c^{-1}) \right] = 0.$$

Z první rovnice opět dostaneme, že maximálně věrohodným odhadem parametru $\mu$ je výběrový průměr. Druhou rovnici je nutno řešit numerickou iterací.

Odhady parametrů $\kappa$ a $c$ získané metodou maximální věrohodnosti bývají většinou rovnocenné. V běžných případech dávají po přepočtu stejný výsledek. MMV ovšem selhává u výběrů, kdy $\kappa$ je velké a $\mu$ současně malé (řešení buď není nalezeno, nebo je nalezen odhad $\hat{\kappa} = \infty$). V takovém případě je vhodné použít jiné metody odhadu- například bayesovský přístup, jež bude popsán v následujícím odstavci.

### 4.3.3 Bayesovské odhady

V tomto odstavci uvedeme bayesovský přístup k odhadu parametrů $\mu$ a $\kappa$ negativně binomického rozdělení. Podrobnější postup určení bayesovský odhadů těchto parametrů je uveden v [3]. Zde uvedeme pouze stěžejní kroky výpočtu odhadů.

Podmíněná apriorní hustota parametru $\mu$ při daném $\sigma^2 = s^2$ je zavedena vztahem $f_{\mu|\sigma^2}(m|s^2) = 1/s^2$ pro $0 \leq m < s^2$ a rovna 0 jinak. Dále pro apriorní hustotu $\sigma^2$ platí $f_{\sigma^2}(s^2) = 1/s^2$ pro $s^2 > 0$. Pak sdružená apriorní hustota $f_{\mu,\sigma^2}(m, s^2) = 1/s^4$ pro $0 \leq$ $\leq m < s^2 < \infty$.

Pro stanovení aposteriorní hustoty $f_{\mu,\sigma^2|\mathbf{X}}(m, s^2|\mathbf{x})$ je použita aproximace hustoty náhodného výběru $X_1, \ldots, X_n$ normálním rozdělením na základě centrální limitní věty.

Nechť $X_i \sim \mathsf{NB}(\pi, \kappa)$, pak pro přirozené $\kappa$ můžeme psát $X_i = \sum_{j=1}^{\kappa} Y_j$, kde $Y_j \sim \mathsf{NB}(\pi, 1)$ jsou nezávislé stejně rozdělené náhodné veličiny. Podle centrální limitní věty má $X_i$ pro velké hodnoty $\kappa$ asymptoticky normální rozdělení $\mathsf{N}(\mu, \sigma^2)$, $0 < \mu < \sigma^2$.

Podle Bayesovy věty lze aposteriorní hustotu zapsat ve tvaru

$$f_{\mu,\sigma^2|\mathbf{X}}(m, s^2|\mathbf{x}) = \frac{f_{\mu,\sigma^2}(m, s^2) f_{\mathbf{X}|\mu,\sigma^2}(\mathbf{x}|m, s^2)}{\int_0^\infty \int_0^{s^2} f_{\mu,\sigma^2}(m, s^2) f_{\mathbf{X}|\mu,\sigma^2}(\mathbf{x}|m, s^2) dm\, ds^2}. \tag{4.4}$$

Hustotu $f_{\mathbf{X}|\mu,\sigma^2}(\mathbf{x}|m, s^2)$ můžeme na základě výše uvedené aproximace zapsat ve tvaru

$$f_{\mathbf{X}|\mu,\sigma^2}(\mathbf{x}|m, s^2) = (2\pi)^{-\frac{n}{2}} (s^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} \frac{(x_i - m)^2}{s^2}\right)$$

a po dosazení do (4.4) a následné úpravě pro aproximovanou aposteriorní hustotu dostaneme

$$f_{\mu,\sigma^2|\mathbf{X}}(m, s^2|\mathbf{x}) = \frac{(s^2)^{-\frac{n}{2}-2} e^{-\frac{1}{2s^2}\sum_{i=1}^{n}(x_i-m)^2}}{\int_0^\infty \int_0^{s^2} (s^2)^{-\frac{n}{2}-2} e^{-\frac{1}{2s^2}\sum_{i=1}^{n}(x_i-m)^2} dm\, ds^2}.$$

Pomocí této hustoty lze stanovit bayesovský ohad $\mu$

$$\hat{\mu} = E(\mu|\mathbf{x}) = \int_0^\infty \int_0^{s^2} m f_{\mu,\sigma^2|\mathbf{X}}(m, s^2|\mathbf{x}) dm\, ds^2,$$

který lze aproximovat výrazem

$$E(\mu|\mathbf{x}) \cong \overline{x} - \frac{E\left[\frac{\sigma}{\sqrt{n}}\left(\phi\left(\frac{(\sigma^2-\overline{x})\sqrt{n}}{\sigma}\right) - \phi\left(-\frac{\overline{x}\sqrt{n}}{\sigma}\right)\right)\right]}{E\left[\Phi\left(\frac{(\sigma^2-\overline{x})\sqrt{n}}{\sigma}\right) - \Phi\left(-\frac{\overline{x}\sqrt{n}}{\sigma}\right)\right]},$$

kde $\phi(z)$ je hustota standardizovaného normálního rozdělení a $\Phi$ je její distribuční funkce.

Podobně pro odhad $\sigma^2$ dostáváme

$$\hat{\sigma}^2 = E(\sigma^2|\mathbf{x}) \cong \frac{E\left[\sigma^2\left(\Phi\left(\frac{(\sigma^2-\overline{x})\sqrt{n}}{\sigma}\right) - \Phi\left(-\frac{\overline{x}\sqrt{n}}{\sigma}\right)\right)\right]}{E\left[\Phi\left(\frac{(\sigma^2-\overline{x})\sqrt{n}}{\sigma}\right) - \Phi\left(-\frac{\overline{x}\sqrt{n}}{\sigma}\right)\right]}$$

a bayesovský odhad parametru $\kappa$ je pak tvaru

$$\hat{\kappa} = \frac{\hat{\mu}^2}{\hat{\sigma}^2 - \hat{\mu}}.$$

Protože $\hat{\mu} \leq \hat{\sigma}^2$, je vždy $\hat{\kappa} > 0$.

Přestože je odvození tohoto odhadu založeno na aproximaci NB rozdělení normálním rozdělením, tedy pro situaci kdy je $\kappa$ velké přirozené, fungují tyto odhady dobře i pro ostatní situace.

# 5 Logaritmicko-normální rozdělení

V kapitole 6 budeme analyzovat složené rozdělení, jehož sčítanci mají logaritmicko-normální rozdělení. Proto bude tato kapitola věnována logaritmicko-narmálnímu rozdělení.

Na začátku této kapitoly zavedeme LN rozdělení a uvedeme jeho základní charakteristiky. Dále popíšeme různé metody odhadu parametrů LN rozdělení.

## 5.1 Zavedení logaritmicko-normálního rozdělení a jeho charakteristiky

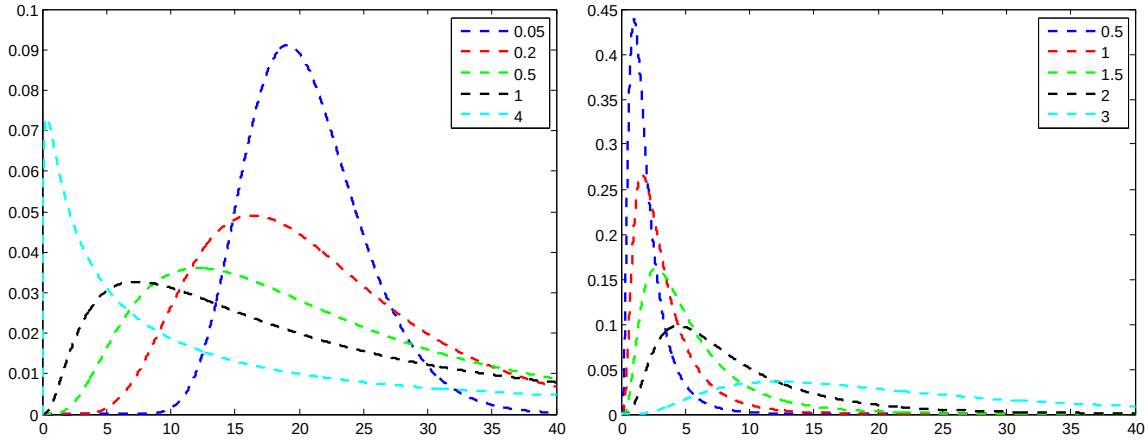
Logaritmicko-normální rozdělení je rozdělení spojitého typu. Náhodná veličina $X$ má logaritmicko-normální rozdělení s parametry $\mu > 0, \sigma^2 > 0$, jestliže transformovaná náhodná veličina $Y = \ln X$ má normální rozdělení $\mathsf{N}(\mu, \sigma^2)$.

Logaritmicko-normální rozdělení má tedy hustotu

$$\begin{aligned} f(x) &= \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} && \text{pro } x > 0 \\ &= 0 && \text{pro } x \leq 0. \end{aligned}$$

Pro náhodnou veličinu $X$ mající logaritmicko-normální rozdělení s parametry $\mu$ a $\sigma^2$ budeme v dalším textu používat označení $X \sim \mathsf{LN}(\mu, \sigma^2)$.

Na obrázku 5.1 jsou příklady hustot logaritmicko-normálního rozdělení pro různé hodnoty parametrů $\mu$ a $\sigma^2$.




(a) $\mu = 3$, $\sigma^2$ postupně $0,05; 0,2; 0,5; 1; 4$

(b) $\sigma^2 = 0,5$, $\mu$ postupně $0,5; 1; 1,5; 2; 3$

Obrázek 5.1: Hustota LN rozdělení pro různé hodnoty parametrů $\mu$ a $\sigma^2$.

Nechť $X \sim LN(\mu, \sigma^2)$, pak pro její střední hodnotu a rorptyl platí

$$EX = e^{\mu + \frac{\sigma^2}{2}} \qquad \text{a} \qquad DX = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1).$$

Výpočet charakteristické funkce LN rozdělení $\psi(t) = Ee^{itX}$ je značně náročný, proto bývá charakteristická funkce často vyjadřována pomocí rozvoje v Taylorovu řadu (viz [14]). Pro charakteristickou funkci pak platí

$$\psi(t) = \sum_{k=0}^{\infty} \frac{(\mathrm{i}t)^k}{k!} e^{k\mu + k^2 \frac{\sigma^2}{2}}. \tag{5.1}$$

Uvedená řada je ovšem divergentní. Nicméně je toto vyjádření postačující pro numerické vyhodnocení charakteristické funkce, pokud horní hranici sumace $K$ zvolíme tak, že

$$\max(|t|, |\mu|) \ll K \ll \frac{2}{\sigma^2} \ln \frac{2}{\sigma^2} \qquad \text{a} \qquad \sigma^2 < 0,1. \tag{5.2}$$

Pro $r$-tý obecný moment platí (viz [7])

$$\mu'_r = e^{r\mu + \frac{r^2}{2}\sigma^2}.$$

Pro první čtyři obecné momenty tedy dostáváme

$$\begin{aligned}
\mu'_1 =& e^{\mu + \frac{\sigma^2}{2}}, \\
\mu'_2 =& e^{2\mu + 2\sigma^2}, \\
\mu'_3 =& e^{3\mu + \frac{9\sigma^2}{2}}, \\
\mu'_4 =& e^{4\mu + 8\sigma^2}.
\end{aligned}$$

Užitím vztahů mezi centrálními a obecnými momenty (2.1) lze po jednoduchém výpočtu určit centrální momenty $\mu_2$, $\mu_3$ a $\mu_4$ ve tvaru

$$\begin{aligned}
\mu_2 =& e^{2\mu + \sigma^2}(e^{\sigma^2} - 1), \\
\mu_3 =& e^{3\mu + \frac{3\sigma^2}{2}}(e^{3\sigma^2} - 3e^{\sigma^2} + 2), \\
\mu_4 =& e^{4\mu + 2\sigma^2}(e^{6\sigma^2} - 4e^{3\sigma^2} + 6e^{\sigma^2} - 3).
\end{aligned}$$

Dále pro šikmost a špičatost platí vztahy

$$\begin{aligned}
\gamma_1 &= \sqrt{e^{\sigma^2} - 1}(e^{\sigma^2} + 2), \\
\gamma_2 &= e^{4\sigma^2} + 2e^{3\sigma^2} + 3e^{2\sigma^2} - 6.
\end{aligned}$$

## 5.2 Metody odhadu parametrů LN rozdělení

V tomto odstavci určíme odhady parametrů NB rozdělení metodou momentů a metodou maximální věrohodnosti.

### 5.2.1 Metoda momentů

Nechť $X_1, \ldots, X_n$ je náhodný výběr z $\mathsf{LN}(\mu, \sigma^2)$. Pak momentové rovnice jsou tvaru

$$e^{\mu+\frac{\sigma^2}{2}} = \overline{X} \tag{5.3}$$

$$e^{2\mu+2\sigma^2} = M_2', \qquad \text{kde} \qquad M_2' = \frac{1}{n}\sum_{i=1}^{n} X_i^2, \tag{5.4}$$

Z rovnice (5.3) vyjádříme parametr $\mu$

$$\mu = \ln \overline{X} - \frac{\sigma^2}{2}. \tag{5.5}$$

Do rovnice (5.4) dosadíme (5.5) a vyjádříme $\sigma^2$, tím získáme odhad parametru $\hat{\sigma^2}$

$$\hat{\sigma}^2 = \ln M_2' - 2\ln \overline{X}. \tag{5.6}$$

Momentový odhad $\hat{\mu}$ získáme dosazením 5.6 do 5.5

$$\hat{\mu} = 2\ln \overline{X} - \frac{\ln M_2'}{2}.$$

Odhady parametrů LN rozdělení získané metodou momentů nejsou funkcí postačující statistiky, proto tyto odhady nejsou příliš kvalitní a je tedy vhodné použít k odhadu parametrů jinou metodu.

### 5.2.2 Metoda maximální věrohodnosti

Věrohodnostní funkce rozdělení $\mathsf{LN}(\mu, \sigma^2)$ je tvaru

$$L(\mu, \sigma^2; \mathbf{x}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma x_i} exp\left(-\frac{(\ln x_i - \mu)^2}{2\sigma^2}\right) = (2\pi)^{-\frac{n}{2}}\sigma^{-n}\prod_{i=1}^{n}\frac{1}{x_i} exp\left(-\sum_{i=1}^{n}\frac{(\ln x_i - \mu)^2}{2\sigma^2}\right).$$

Jejím logaritmováním získáme logaritmickou věrohodnostní funkci

$$l(\mu, \sigma^2; \mathbf{x}) = -\frac{n}{2}\ln(2\pi) - n\ln\sigma - \sum_{i=1}^{n}\ln x_i - \sum_{i=1}^{n}\frac{(\ln x_i - \mu)^2}{2\sigma^2}.$$

Derivace logaritmické věrohodnostní funkce podle jednotlivých parametrů položíme rovny nule a dostaneme tak systém věrohodnostních rovnic pro hledané odhady

$$\frac{\partial l}{\partial \mu} = \sum_{i=1}^{n}\frac{\ln x_i - \mu}{\sigma^2} = 0 \tag{5.7}$$

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{\sigma^2} + \sum_{i=1}^{n}\frac{(\ln x_i - \mu)^2}{\sigma^4} = 0. \tag{5.8}$$

Z rovnice (5.7) vyjádříme parametr $\mu$, tím získáme odhad

$$\hat{\mu} = \frac{\sum_{i=1}^{n}\ln x_i}{n}. \tag{5.9}$$

Do rovnice (5.8) dosadííme (5.9) a vyjádříme $\sigma^2$. Obdržíme odhad

$$\hat{\sigma^2} = \frac{\sum_{i-1}^{n}(\ln x_i - \hat{\mu})^2}{n}.$$

# 6 Konkrétní případ složeného rozdělení

Jak již bylo zmíněno dříve, budeme se nyní zabývat složeným rozdělením, jehož počet sčítanců má negativně binomické rozdělení s parametry $\pi$ a $\kappa$. Rozdělení jednotlivých sčítanců je logaritimicko-normální s parametry $\mu$ a $\sigma^2$. Pro náhodnou veličinu $S$ mající složené rozdělení tohoto typu bude v dalším textu použito označení $S \sim \mathsf{NBLN}(\pi, \kappa, \mu, \sigma^2)$.

Nejdříve ukážeme, jak lze přibližně určit hustotu složeného rozdělení, známe-li parametry rozdělení počtu sčítanců a jednotlivých sčítanců. Dále budeme studovat vliv odhadu parametrů dilčích rozdělení na výsledné složené rozdělení.

Veškeré simulace a další výpočty byly provedeny ve výpočetním systému MATLAB R2009a.

## 6.1 Určení hustoty složeného rozdělení

Předpokládejme, že náhodná veličina $S$ má složené rozdělení takové, že rozdělení počtu sčítanců je NB s parametry $\pi$ a $\kappa$. Jednotlivý sčítanci mají LN rozdělení s parametry $\mu$ a $\sigma^2$, tedy $S \sim \mathsf{NBLN}(\pi, \kappa, \mu, \sigma^2)$.

Jelikož známe rozdělení počtu sčítanců ($N \sim \mathsf{NB}(\pi, \kappa)$) a rozdělení jednotlivých sčítanců ($X \sim \mathsf{LN}(\mu, \sigma^2)$) složeného rozdělení, můžeme podle věty 3.5 určit charakteristickou funkci náhodné veličiny $S$. Pro charakteristickou funkci náhodné veličiny $S$ tedy platí

$$\psi_S(t) = G_N(\psi_X(t)). \tag{6.1}$$

Vytvořující funkce $G_N$ rozdělení počtu sčítanců je tvaru (viz (4.3))

$$G_N(s) = \left(\frac{\pi}{1-(1-\pi)s}\right)^{\kappa}. \tag{6.2}$$

Pro charakteristickou funkci $\psi_X$ použijeme aproximaci (5.1). Tedy

$$\psi_X(t) = \sum_{k=0}^{K} \frac{(\mathrm{i}t)^k}{k!} e^{k\mu + k^2 \frac{\sigma^2}{2}}. \tag{6.3}$$

Dosazením (6.2) a (6.3) do vztahu (6.1) obdržíme aproximaci charakteristické funkce nahodné veličiny $S$.

$$\psi_S(t) = \left(\frac{\pi}{1-(1-\pi)\sum_{k=0}^{K} \frac{(\mathrm{i}t)^k}{k!} e^{k\mu + k^2 \frac{\sigma^2}{2}}}\right)^{\kappa}. \tag{6.4}$$

Výše odvozený vzorec pro aproximaci charakteristické funkce složeného rozdělení náhodné veličiny $S$ lze použít pouze v případě, že jsou splněny podmínky pro aproximaci charakterické funkce LN rozdělení (5.2). Tedy horní hranici sumace $K$ volíme s ohledem na omezení

$$\max(|t|, |\mu|) \ll K \ll \frac{2}{\sigma^2} \ln \frac{2}{\sigma^2}.$$

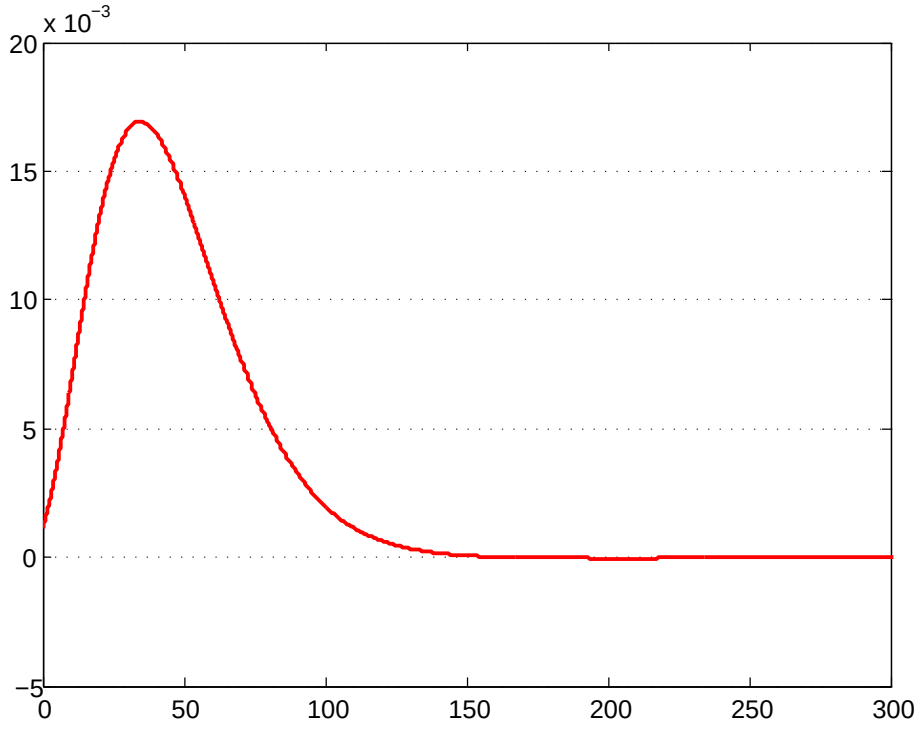
A dále rozptyl $\sigma^2$ LN rozdělení musí splňovat podmínku

$$\sigma^2 < 0,1.$$

Podle věty 3.11 je každá distribuční funkce svou charakteristickou funkcí jednoznačně určena. Pokud tedy známe charakteristickou funkci složeného rozdělení, mělo by být možné určit i hledanou hustotu. Jelikož je charakteristická funkce v podstatě inverzní Fourierovou transformací hustoty náhodné veličiny (až na konstantu $2\pi$), lze pomocí Fourierovy transformace získat z charakteristické funkce původní hustotu.

Pro případ složeného rozdělení $\mathsf{NBLN}(\pi, \kappa, \mu, \sigma^2)$ je určení hustoty z charakteristické funkce pomocí Fourierovy transformace naprogramováno ve funkci `Hustota_NB_LN` (podrobnější informace o funkci viz seznam programů v kapitole 7).

Na obrázku 6.1 je hustota složeného rozdělení $\mathsf{NBLN}(\frac{1}{3}; 5; 1,5; 0,09)$, jež byla určena pomocí funkce `Hustota_NB_LN`.

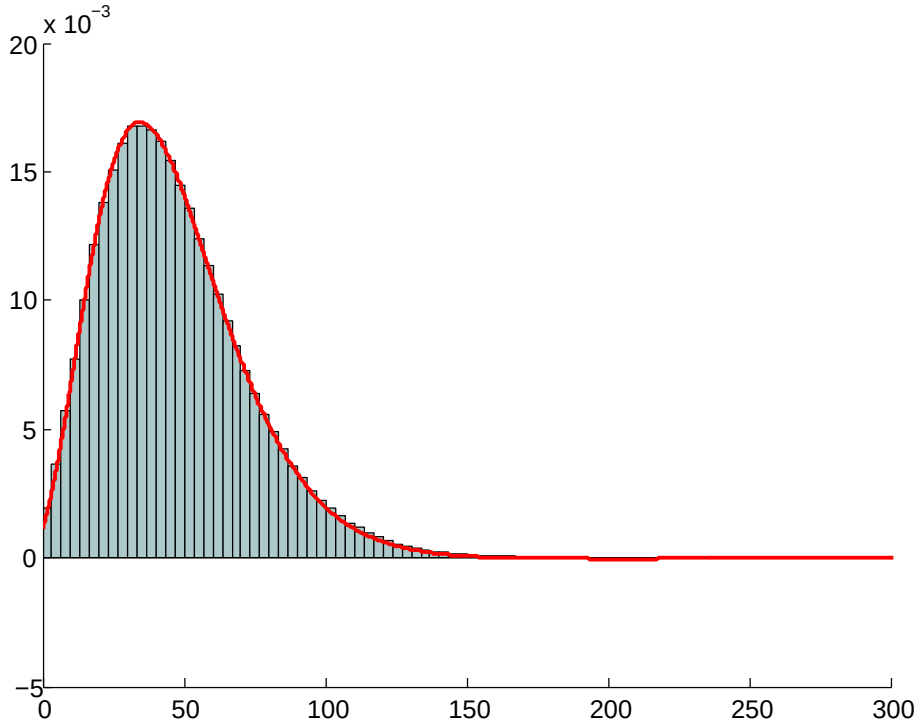



Obrázek 6.1: Hustota náhodné veličiny $S \sim \mathsf{NBLN}(\frac{1}{3}; 5; 1,5; 0,09)$.

Pomocí simulace náhodného výběru velkého rozsahu lze posoudit přesnost, s jakou byla určena hustota studovaného složeného rozdělení. Z histogramu náhodného výběru získáme představu o průběhu hustoty složeného rozdělení. Na obrázku 6.2 je histogram simulovaného náhodného výběru rozsahu $10^6$ ze složeného rozdělení $\mathsf{NBLN}(\frac{1}{3}; 5; 1,5; 0,09)$. Histogram byl upraven tak, aby obsah plochy pod histogramem byl jednotkový. Pro větší názornost byla do stejného obrázku zakreslena i dříve určená hustota. Na obrázku je vidět, že přibližně určená hustota velmi přesně kopíruje průběh histogramu.

Předchozí simulace náhodného výběru ze složeného rozdělení $\mathsf{NBLN}(\frac{1}{3}; 5; 1,5; 0,09)$ a histogram byly získány pomocí funkce `Simulace_NB_LN`, která byla naprogramována v MATLABu a nachází se na přiloženém CD.

Přesnost určené aproximace hustoty složeného rozdělení můžeme lépe posoudit porovnáním odpovídajících kvantilů, které určíme z nasimulovaných dat a z aproximované




Obrázek 6.2: Histogram náhodného výběru rozsahu $10^6$ ze složeného rozdělení $\mathsf{NBLN}(\frac{1}{3}; 5; 1,5; 0,09)$.

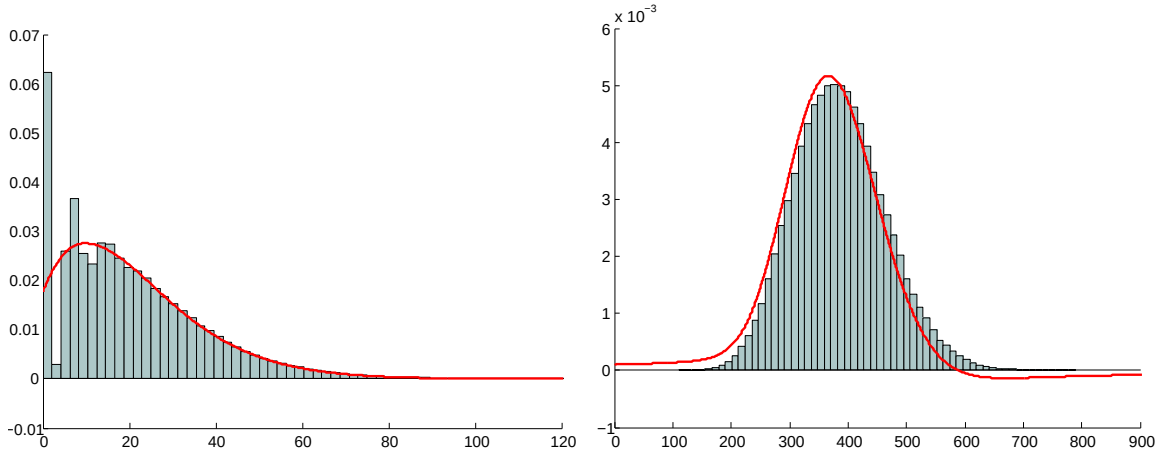
hustoty. Hodnoty kvantilů můžeme s vysokou přesností určit z náhodného výběru velkého rozsahu. Získané hodnoty kvantilů porovnáme z kvantily určenými z aproximované hustoty.

V tabulce 6.1 je přehled kvantilů složeného rozdělení $\mathsf{NBLN}(\frac{1}{3}; 5; 1,5; 0,09)$. Ve druhém sloupci tabulky jsou kvantily určené z již dříve nasimulovaného náhodného výberu rozsahu $10^6$. Tyto kvantily byly určeny pomocí funkce `quantile`, jež je implementována přímo v MATLABu. Ve třetím sloupci jsou kvantily určené z aproximace hustoty, které byly získány pomocí funkce `Kvantil` (viz seznam programů v kapitole 7). V posledním sloupci je rozdíl odpovídajících si kvantilů získaných z nasimulovaných dat a z aproximace hustoty. Rozdíly odpovídajících si kvantilů jsou poměrně malé (pouze rozdíl $0,99$- kvantilů je větší než 1). Funkce získaná pomocí matlabovského příkazu `Hustota_NB_LN` tedy velmi dobře aproximuje hustotu složeného rozdělení $\mathsf{NBLN}(\frac{1}{3}; 5; 1,5; 0,09)$. Pomocí teto funkce můžeme určit $p$-kvantily, pro $p \leq 0,975$, zkoumaného SR s přesností menší než 1.

I když parametry $\mu$ a $\sigma^2$ splňují podmínky pro užití funkce `Hustota_NB_LN` k určení hustoty složeného rozdělení $\mathsf{NBLN}(\pi, \kappa, \mu, \sigma^2)$, není možné tuto funkci použít pro libovolné hodnoty parametrů $\pi$ a $\kappa$. Pomocí simulací bylo zjištěno, že pokud je střední hodnota rozdělení počtu sčítanců menší než 5, nabývá výsledné složené rozdělení velmi často hodnoty 0 (viz obrázek 6.3 (a)). V takovém případě se nejedná o rozdělení spojitého typu a nemá tedy smysl určovat hustotu. Pozorováním bylo dále zjištěno, že funkce `Hustota_NB_LN` nedává dostatečně přesné výsledky pro střední hodnotu rozdělení počtu sčítanců větší než 40 (viz obrázek 6.3 (b)).

| | Kvantily ze simulace | Kvantily z hustoty | Rozdíl kvantilů |
|---|---|---|---|
| $0,10$-kvantil | $16,9829$ | $16,9956$ | $-0,0127$ |
| $0,25$-kvantil | $27,7450$ | $27,6644$ | $0,0806$ |
| $0,50$-kvantil | $42,8468$ | $42,7141$ | $0,1327$ |
| $0,75$-kvantil | $61,6173$ | $61,2919$ | $0,3254$ |
| $0,90$-kvantil | $81,8431$ | $81,3196$ | $0,5235$ |
| $0,95$-kvantil | $95,5811$ | $95,1103$ | $0,4708$ |
| $0,975$-kvantil | $108,4398$ | $108,6050$ | $-0,1652$ |
| $0,99$-kvantil | $124,5661$ | $128,8515$ | $-4,2854$ |

Tabulka 6.1: Porovnání kvantilů rozdělení $\mathsf{NBLN}(\frac{1}{3}; 5; 1,5; 0,09)$ určených z nasimulovaného výběru rozsahu $10^6$ a z aproximace hustoty.




(a) $\mathsf{NBLN}(0,6; 4; 2; 0,09)$ (b) $\mathsf{NBLN}(0,5; 50; 2; 0,09)$

Obrázek 6.3: Příklady složených rozdělení $\mathsf{NBLN}(\pi, \kappa, \mu, \sigma^2)$, jejichž hustotu nelze spolehlivě určit pomocí funkce `Hustota_NB_LN`.

Na obrázku 6.3 (a) je histogram náhodného výběru rozsahu $10^6$ ze složeného rozdělení $\mathsf{NBLN}(0,6; 4; 2; 0,09)$. Odpovídající střední hodnota rozdělení počtu sčítanců je $2,\overline{6}$. V tomto náhodném výběru bylo hodnoty 0 dosaženo 129699krát. Histogram byl opět upraven tak, aby plocha pod ním byla jednotková. Do stejného obrázku byla pro porovnání zakreslena i hustota určená funkcí `Hustota_NB_LN` pro stejné hodnoty parametrů složeného rozdělení.

Na obrázku 6.3 (b) je histogram náhodného výběru rozsahu $2 \cdot 10^5$ ze složeného rozdělení $\mathsf{NBLN}(0,5; 50; 2; 0,09)$. Odpovídající střední hodnota rozdělení počtu sčítanců je 50. Plocha pod histogramem je jednotková. V tomto případě hustota určená funkcí `Hustota_NB_LN` dosahuje značných nepřesností a chvost rozdělení nabývá výrazně záporných hodnot.

## 6.2 Vliv odhadu parametrů

V tomto odstavci bude studován vliv odhadu parametrů rozdělení jednotlivých sčítanců a počtu sčítanců na výsledné složené rozdělení. Vliv odhadu parametrů byl sledován pomocí simulací výběrů ze složeného rozdělení $\mathsf{NBLN}(\frac{1}{3}; 5; 1,5; 0,09)$.

Z nasimulovaných dat malého rozsahu ze složeného rozdělení určíme maximálně věrohodné odhady parametrů jednotlivých sčítanců. Pomocí funkce `Hustota_NB_LN` určíme hustotu složeného rozdělení s těmito odhadnutými parametry. Tuto hustotu porovnáme s hustotou, jež byla určena s přesnými hodnotami parametrů. Dále budeme studovat, jak přibližně určené hodnoty parametrů ovlivnily kvantily složeného rozdělení.

Tento postup zopakujeme i pro maximálně věrohodné a bayesovské odhady parametrů rozdělení počtu sčítanců.

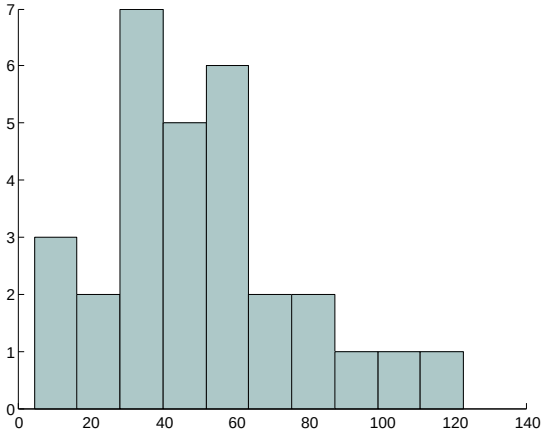
### 6.2.1 Vliv odhadu parametrů LN rozdělení

Na obrázku 6.4 je histogram pro simulovaný náhodný výběr rozsahu 30 ze složeného rozdělení $\mathsf{NBLN}(\frac{1}{3}; 5; 1,5; 0,09)$. Pomocí funkce `lognfit`, jež je implementována přímo v MATLABu, byly určeny maximálně věrohodné odhady parametrů rozdělení jednotlivých sčítanců. Byly získány odhady

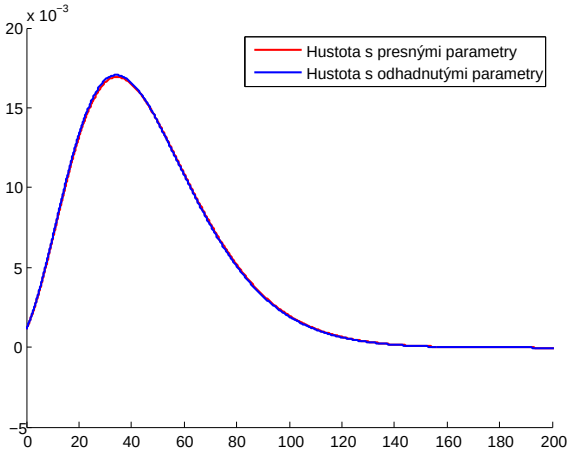
$$\hat{\mu} = 1,4907 \qquad \text{a} \qquad \hat{\sigma}^2 = 0,0931.$$

Tyto odhady jsou velmi přesné, proto i hustota určená s těmito odhadnutými parametry velmi dobře kopíruje průběh hustoty určené s přesnými parametry, viz obrázek 6.5. Červeně je vykreslena hustota složeného rozdělení s přesnými parametry, modře potom hustota rozdělení s odhadnutými parametry rozdělení jednotlivých sčítanců, tj. hustota rozdělení $\mathsf{NBLN}(\frac{1}{3}; 5; 1,4907; 0,0931)$. Maximální rozdíl funkčních hodnot těchto hustot je $9,9486 \cdot 10^{-5}$. Průběhy funkcí jsou tedy takřka totožné.

V tabulce 6.2 je uveden přehled kvantilů určených z těchto hustot. Rozdíly jednotlivých kvantilů jsou menší než jedna. Metoda maximální věrohodnosti pro odhad parametrů LN rozdělení je tedy vhodná již při malém rozsahu výběru ze složeného rozdělení.




Obrázek 6.4: Histogram náhodného výběru rozsahu 30 ze složeného rozdělení $\mathsf{NBLN}(\frac{1}{3}; 5; 1,5; 0,09)$.




Obrázek 6.5: Průběhy hustot

| | Kvantily určené z hustoty s přesnými parametry | Kvantily určené z hustoty s odhadnutými parametry | Rozdíl kvantilů |
|---|---|---|---|
| $0,10$-kvantil | $16,9956$ | $16,8459$ | $0,1497$ |
| $0,25$-kvantil | $27,6644$ | $27,4649$ | $0,1995$ |
| $0,50$-kvantil | $42,7141$ | $42,3646$ | $0,3495$ |
| $0,75$-kvantil | $61,2919$ | $60,8421$ | $0,4498$ |
| $0,90$-kvantil | $81,3196$ | $80,7197$ | $0,5999$ |
| $0,95$-kvantil | $95,1103$ | $94,4103$ | $0,7$ |
| $0,975$-kvantil | $108,6050$ | $107,8051$ | $0,7999$ |
| $0,99$-kvantil | $128,8515$ | $127,9015$ | $0,95$ |

Tabulka 6.2: Porovnání kvantilů rozdělení $\mathsf{NBLN}(\frac{1}{3}; 5; 1,5; 0,09)$ a $\mathsf{NBLN}(\frac{1}{3}; 5; 1,4907; 0,0931)$.

### 6.2.2 Vliv odhadu parametrů NB rozdělení

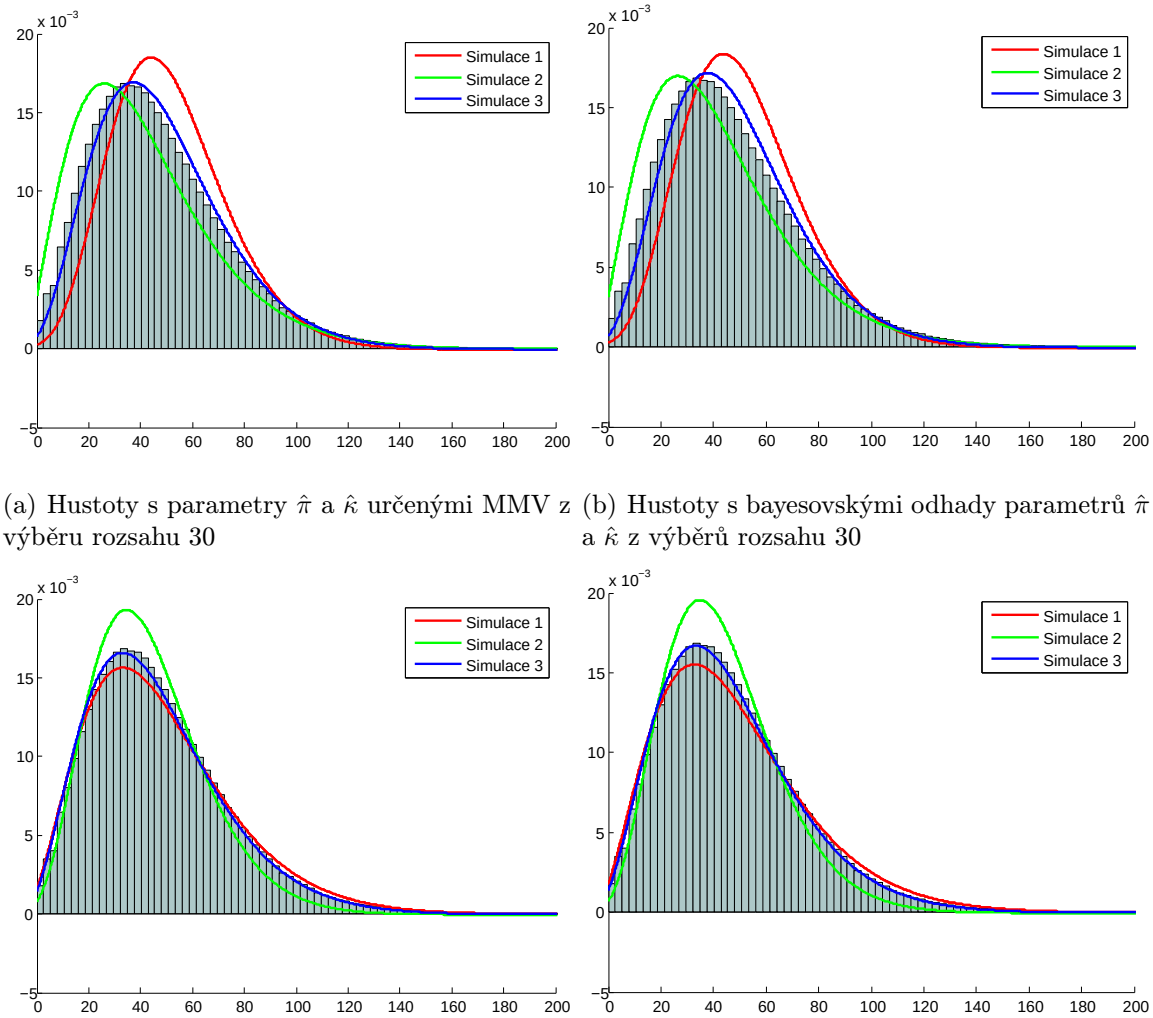
Byly provedeny tři simulace výběru ze složeného rozdělení $\mathsf{NBLN}(\frac{1}{3}; 5; 1,5; 0,09)$ rozsahu 30 a tři simulace rozsahu 300. Rozdělení počtu sčítanců je tedy negativně binomické s hodnotami parametrů $\pi = \frac{1}{3}$ a $\kappa = 5$. Pro všechny tyto výběry byly určeny maximálně věrohodné a bayesovské odhady parametrů rozdělení počtu sčítanců. Výsledné odhady jsou uvedeny v tabulce 6.3. Z hodnot odhadů je patrné, že maximálně věrohodné i bayesovské odhady se od přesných hodnot parametrů často výrazně liší i pro výběry větších rozsahů. Pro některé výběry jsou přesnější maximálně věrohodné odhady (např. simulace číslo 2 rozsahu 300), v jiných případech jsou zase přesnější bayesovské odhady (např. simulace číslo 2 rozsahu 30), avšak odhady určené ze stejných výběrů se od sebe většinou výrazně neliší.

| Číslo simulace | Rozsah simulace 30 | | | | Rozsah simulace 300 | | | |
|---|---|---|---|---|---|---|---|---|
| | MV odhady | | Bayesovské odhady | | MV odhady | | Bayesovské odhady | |
| | $\hat{\pi}$ | $\hat{\kappa}$ | $\hat{\pi}$ | $\hat{\kappa}$ | $\hat{\pi}$ | $\hat{\kappa}$ | $\hat{\pi}$ | $\hat{\kappa}$ |
| 1 | 0,4947 | 10,8351 | 0,4851 | 10,4253 | 0,2276 | 3,9713 | 0,2702 | 3,8279 |
| 2 | 0,2621 | 3,1965 | 0,2698 | 3,3259 | 0,4317 | 7,0763 | 0,4465 | 7,5149 |
| 3 | 0,3558 | 5,7448 | 0,3709 | 6,1308 | 0,3097 | 4,4590 | 0,3154 | 4,5785 |

Tabulka 6.3: Odhady parametrů $\pi$ a $\kappa$ z výběrů ze složeného rozdělení $\mathsf{NBLN}(\frac{1}{3}; 5; 1,5; 0,09)$.

Maximálně věrohodné odhady byly určeny pomocí funkce `nbinfit`, jež je implementována přímo v MATLABu. Bayesovské odhady byly získány pomocí funkce `bayes_k`, která byla autorovi zapůjčena z dizertační práce [3].

Na obrázku 6.6 jsou hustoty složeného rozdělení s hodnotami parametrů rozdělení počtu sčítanců, které byly odhadnuty z nasimulovaných výběrů. Pro větší názornost byl do obrázků zakreslen histogram náhodného výběru rozsahu $10^6$ ze složeného rozdělení s přesnými hodnotami parametrů. Z průběhů hustot je patrné, že hustoty určené s těmito odhady se v mnohých případech výrazně liší od hustoty studovaného rozdělení. Proto i kvantily určené z těchto hustot nejsou příliš přesné, především potom vysoké kvantily. Hodnoty kvantilů SR s bayesovskými odhady parametrů z výběrů rozsahu 30 jsou v tabulce 6.4.




(a) Hustoty s parametry $\hat{\pi}$ a $\hat{\kappa}$ určenými MMV z výběru rozsahu 30

(b) Hustoty s bayesovskými odhady parametrů $\hat{\pi}$ a $\hat{\kappa}$ z výběrů rozsahu 30

(c) Hustoty s parametry $\hat{\pi}$ a $\hat{\kappa}$ určenými MMV z výběru rozsahu 300

(d) Hustoty s bayesovskými odhady parametrů $\hat{\pi}$ a $\hat{\kappa}$ z výběrů rozsahu 300

Obrázek 6.6: Hustoty složeného rozdělení $\mathsf{NBLN}(\hat{\pi}; \hat{\kappa}; 1,5; 0,09)$.

| | Přesné parametry | Simulace 1 | Simulace2 | Simulace 3 |
|---|---|---|---|---|
| $0,10$-kvantil | $16,9956$ | 24.2933 | 12.2011 | 19.2942 |
| $0,25$-kvantil | $27,6644$ | 35.0167 | 22.1664 | 30.1144 |
| $0,50$-kvantil | $42,7141$ | 48.9716 | 37.2621 | 44.9658 |
| $0,75$-kvantil | $61,2919$ | 65.0996 | 57.0384 | 62.8938 |
| $0,90$-kvantil | $81,3196$ | 81.5245 | 79.4670 | 81.8709 |
| $0,95$-kvantil | $95,1103$ | 92.3135 | 95.8583 | 94.6612 |
| $0,975$-kvantil | $108,6050$ | 102.2571 | 113.5036 | 106.9556 |
| $0,99$-kvantil | $128,8515$ | 115.1528 | 160.5002 | 124.2019 |

Tabulka 6.4: Kvantily složeného rozdělení $\mathsf{NBLN}(\frac{1}{3}; 5; 1,5; 0,09)$ určené z hustoty s přesnými parametry a z hustot s bayesovskými odhady parametrů ze simulací rozsahu 30.

K posuzení, zda kvantily určené z hustot s odhadnutými parametry jsou vychýlené od kvantilů určených z hustoty s přesnými parametry, byla provedena následující simulace. 50krát se nasimuloval výběr ze složeného rozdělení $\mathsf{NBLN}(\frac{1}{3}; 5; 1,5; 0,09)$ rozsahu 30. Pro

každý výběr se nalezly MV odhady a bayesovské odhady parametrů $\pi$ a $\kappa$. Pomocí hustot s odhadnutými parametry byly určeny kvantily. Jako výsledný kvantil se vzal průměr z 50 nalezených kvantilů, viz tabulka 6.5.

| | Kvantily určené z hustoty s přesnými parametry | Průměr kvantilů s MV odhady | Průměr kvantilů s bayesovskými odhady |
|---|---|---|---|
| $0,10$-kvantil | $16,9956$ | 18.7919 | 18.5608 |
| $0,25$-kvantil | $27,6644$ | 29.3547 | 29.1762 |
| $0,50$-kvantil | $42,7141$ | 43.9512 | 43.8665 |
| $0,75$-kvantil | $61,2919$ | 61.7146 | 61.7969 |
| $0,90$-kvantil | $81,3196$ | 80.6468 | 80.9414 |
| $0,95$-kvantil | $95,1103$ | 93.5327 | 93.9914 |
| $0,975$-kvantil | $108,6050$ | 106.0929 | 106.7177 |
| $0,99$-kvantil | $128,8515$ | 125.0850 | 125.9159 |

Tabulka 6.5: V tabulce jsou uvedeny kvantily složeného rozdělení $\mathsf{NBLN}(\frac{1}{3}; 5; 1,5; 0,09)$ určené z hustoty s přesnými parametry a dále průměry kvantilů, jež byly získány z hustot s maximálně věrohodnými a bayesovskými odhady parametrů $\pi$ a $\kappa$.

Průměry kvantilů uvedené v tabulce 6.5 se od hodnot kvantilů, které byly určeny z hustoty s přesnými parametry, liší jen v řádu jednotek. Dále průměry kvantilů pro MV a bayesovské odhady si jsou velmi blízké. Použití baesovského přístupu k odhadu parametrů NB rozdělení je však univerzálnější, jelikož lze použít i v případech, kdy MMV selhává.

# 7 Programy

Součástí této práce jsou i tři funkce naprogramované v softwaru MATLAB. V předchozí kapitole bylo těchto funkcí užito k analýze složeného rozdělení $\mathsf{NBLN}(\pi, \kappa, \mu, \sigma^2)$. Funkce se nachází na přiloženém CD. V této kapitole bude uveden seznam těchto funkcí a jejich podrobnější popis.

**funkce** `Hustota_NB_LN`

Tato funkce určí funkční hodnoty hustoty složeného rozděleni v bodech intervalu $\langle 0, 5000 \rangle$. Je zvoleno ekvidistantní dělení, kde vzdálenost sousedních bodů je 0,05. Hustota je získána z charakteristické funkce složeného rozdělení pomocí diskrétní Fourierovy transformace. K výpočtu hodnot charakteristické funkce je použit aproximační vztah (6.4), kde horní mez sumace $K$ klademe rovnu 7. Počet sčítanců složeného rozdělení má negativně binomické rozdělen s parametry `pi_NB` a `kappa_NB`. Jednotlivý sčítanci mají logaritmicko-normální rozdělení s parametry `mu_LN` a `sigma2_LN`. Funkce dále zobrazí průběh hustoty na intervalu $\langle \texttt{x_od}, \texttt{x_do} \rangle$. Funkci lze použít pouze v případě, když $\texttt{sigma2_LN} < 0,1$ a střední hodnota NB rozdělení náleží intervalu $\langle 5; 40 \rangle$.

```
[ x ,f_x ] = Hustota_NB_LN( pi_NB, kappa_NB, mu_LN, sigma2_LN, x_od, x_do )
```

Vstupy:

| | | |
|---|---|---|
| `pi_NB` | – | parametr $\pi$ negativně binomického rozdělení, |
| `kappa_NB` | – | parametr $\kappa$ negativně binomického rozdělení, |
| `mu_LN` | – | parametr $\mu$ logaritmicko-normalního rozdělení, |
| `sigma2_LN` | – | parametr $\sigma^2$ logaritmicko-normalního rozdělení, |
| `x_od` | – | dolní mez pro vykreslení, |
| `x_do` | – | horní mez pro vykreslení – $(0 \leq \texttt{x_od} < \texttt{x_do} \leq 5000)$. |

V přpade, že meze pro vykreslení `x_od` a `x_do` nebudou uvedeny, bude hustota složeného rozdělení vykreslena na maximalním možném intervalu $\langle 0, 5000 \rangle$.

Výstupy:

| | | |
|---|---|---|
| `x` | – | vektor bodů, v nichž je určena hustota složeneho rozdělení, |
| `f_x` | – | vektor hodnot hustoty v `x`. |

**funkce** `Simulace_NB_LN`

Tato funkce nasimuluje náhodný výběr ze složeného rozdělení. Počet sčítanců má negativně binomické rozdělení s parametry `pi_NB` a `kappa_NB`. Jednotlivý sčítanci mají logaritmicko-normalní rozdělení s parametry `mu_LN` a `sigma2_LN`. Funkce dále zobrazí histogram pro nasimulovaný náhodný výběr upravený tak, že plocha pod histogramem je rovna 1.

```
[ SR ,NB ,LN ] = Simulace_NB_LN( rozsahvyberu, pi_NB, kappa_NB, mu_LN,
                 sigma2_LN, poc_sl_hist )
```

Vstupy:

| | | |
|---|---|---|
| `rozsahvyberu` | – | rozsah výběru ze složeného rozdělení, |
| `pi_NB` | – | parametr $\pi$ negativně binomického rozdělení, |
| `kappa_NB` | – | parametr $\kappa$ negativně binomického rozdělení, |
| `mu_LN` | – | parametr $\mu$ logaritmicko-normalního rozdělení, |
| `sigma2_LN` | – | parametr $\sigma^2$ logaritmicko-normalního rozdělení, |
| `poc_sl_hist` | – | počet sloupců histogramu. |

Výstupy:

| | | |
|---|---|---|
| `SR` | – | vektor nasimulovaných dat ze složeného rozdělení, |
| `NB` | – | vektor počtu sčítanců v náhodných součtech, |
| `LN` | – | vektor jednotlivých sčítanců složeného rozdělení. |

**funkce** `Kvantil`

Tato funkce dává `alfa`-kvantil náhodné veličiny spojitého typu, jejíž hustota nabývá v bodech `x` hodntot `f_x`. K výpočtu kvantilu je použita lichoběžníková formule pro numerické integrování.

```
function [ x_alfa ] = Kvantil( x , f_x, alfa )
```

Vstupy:

| | | |
|---|---|---|
| `x` | – | vektor bodů, v nichž je určena hustota rozdělení, |
| `f_x` | – | vektor hodnot hustoty v `x`. |

Výstupy:

| | | |
|---|---|---|
| `x_alfa` | – | alfa-kvantil. |

# 8 Závěr

Na začátku práce je zavedeno složené rozdělení. Rozdělení pravděpodobnosti lze určit mimo jiné pomocí charakteristické funkce, momentové vytvořující funkce a vytvořující funkce kumulantů. Výpočet těchto charakteristik pro složené rozdělení je v práci odvozen a následně i ukázán na příkladu složeného rozdělení, jehož jednotlivý sčítanci mají exponenciální rozdělení a rozdělení jejich počtu je Poissonovo.

Následující dvě kapitoly jsou věnovány logaritmicko-normálnímu a negativně binomickému rozdělení, protože se zbývající část práce zabývá složeným rozdělením, jehož sčítanci mají právě logaritmicko-normální rozdělení a rozdělení jejich počtu je negativně binomické. Tato rozdělení jsou zde podrobně popsána a jsou určeny jejich základní charakteristiky. Značný prostor byl vyhrazen různým metodám odhadu parametrů těchto rozdělení, protože byl dále studován vliv těchto odhadů na výsledné složené rozdělení.

Jak již bylo zmíněno dříve, závěrečná část práce je věnována složenému rozdělení, jehož sčítanci mají LN rozdělení a rozdělení jejich počtu je NB. Nedílnou součástí práce jsou tři funkce naprogramované v MATLABu, s jejichž pomocí bylo studováno toto složené rozdělení. Těmito funkcemi byla velmi přesně určena hustota a kvantily rozdělení $\mathsf{NBLN}(\frac{1}{3}; 5; 1,5; 0,09)$. Dále byl sledován vliv odhadu parametrů na složené rozdělení. Odhady parametrů rozdělení jednotlivých sčítanců, které byly určeny metodou maximální věrohodnosti, jsou velmi přesné již pro výběry malého rozsahu z rozdělení $\mathsf{NBLN}(\frac{1}{3}; 5; 1,5; 0,09)$. Hodnoty kvantilů určených z hustoty s těmito odhadnutými parametry se od kvantilů určených z hustoty složeného rozdělení s přesnými hodnotami parametrů významně nelišily. Mnohem výraznější vliv na shodu odhadnutého a teoretického rozdělení mají odhady parametrů rozdělení počtu sčítanců. Maximálně věrohodné i bayesovské odhady parametrů NB rozdělení jsou často velmi nepřesné i pro simulované výběry většího rozsahu z $\mathsf{NBLN}(\frac{1}{3}; 5; 1,5; 0,09)$. Tedy i kvantily určené s takto odhadnutými parametry mohou být značně nepřesné. Opakováním simulací však bylo zjištěno, že kvantily určené s odhadnutými parametry počtu sčítanců nejsou vychýlené od kvantilů získaných s hustot s přesnými hodnotami parametrů.

# Literatura


[1] ANDĚL, J.: *Základy matematické statistiky.* Praha: MATFYZPRESS, 2007. 358 s. ISBN 80-7378-001-1.

[2] ANDĚL, J.: *Matematická statistika.* Praha: SNTL, 1978. 346 s. ISBN 80-7378-001-1.

[3] DOUDOVÁ, L.: *Statistická analýza populací s negativně binomickým rozdělením.* Brno: Masarykova univerzita, 2009. 109 s.

[4] Dupač, V., Hušková, M.: *Pravděpodobnost a matematická statistika.* Praha: Univerzita Karlova v Praze– Nakladatelství Karolinum, 1999. 162 s. ISBN 80-246-0009-9.

[5] Embrechts, P., Frei , M.: Panjer recursion versus FFT for compound distributions. *Mathematical Methods of Operations Research*, 2009, vol. 69, s. 497-508.

[6] Gil-Pelaez, J.: Note on the inversion theorem. *Biometrika*, 1951, vol. 38, s. 481-482.

[7] Likeš, J.: Inference v lognormálním a Paretově rozdělení. *Robust*, 1980.

[8] MICHÁLEK, J.: *Matematická statistika pro informatiky.* Praha: Státní pedagogické nakladatelství, 1987. 109 s.

[9] RAO, C.R.: *Lineární metody statistické indukce a jejich aplikace.* Praha: ACADEMIA, 1978. 666 s.

[10] RÉNYI, A.: *Teorie pravděpodobnosti.* Praha: ACADEMIA, 1972. 511 s.

[11] RESNICK, S.: *Adventures in stochastic processes.* Boston: Birkhäuser, 1992. ISBN 0-8176-3591-2.

[12] REVFEIM, K.J.A.: An initial model of the relationship between rainfall events and daily rainfalls. *Journal of Hydrology*, 1984, vol. 75, s. 357-363.

[13] Thompson, C.S.: Homogeneity analysis of a rainfall series: an application of the use of a realistic rainfall model. *Journal of Climatology*, 1984, vol. 4, s. 609-619.

[14] *Log-normal distribution.* [online]. Poslední revize 10. 3. 2011.
URL: <`http://en.wikipedia.org/wiki/Lognormal_distribution`>, [cit. 20. 3. 2011].

[15] *Negative binomial distribution.* [online]. Poslední revize 8. 3. 2011.
URL: <`http://en.wikipedia.org/wiki/Negative_binomial_distribution`>, [cit. 12. 3. 2011].

# 9 Seznam použitých zkratek a symbolů

| | |
|---|---|
| NB | negativně binomické rozdělení |
| LN | logaritmicko-normální rozdělení |
| SR | složené rozdělení |
| MM | metoda momentů |
| MMV | metoda maximální věrohodnosti |
| MV odhad | maximálně věrohodný odhad |

| | |
|---|---|
| $(\Omega, \mathcal{A}, \mathrm{P})$ | pravděpodobnostní prostor |
| $\mathbb{R}^n$ | reálný $n$-rozměrný eukleidovský prostor |
| f | hustota rozdělení pravděpodobnosti |
| F | distribuční funkce |
| $EX$ | střední hodnota náhodné veličiny $X$ |
| $DX$ | rozptyl náhodné veličiny $X$ |
| $\overline{X}$ | výběrový průměr |
| $\mu'_r$ | $r$-tý obecný moment |
| $\mu_r$ | $r$-tý centrální moment |
| $M'_r$ | $r$-tý obecný výběrový moment |
| $M_r$ | $r$-tý centrální výběrový moment |
| $\gamma_1$ | šikmost |
| $\gamma_2$ | špičatost |
| $\Gamma$ | gama funkce |
| $L$ | věrohodnostní funkce |
| $l$ | logaritmická věrohodnostní funkce |
| $G_X$ | vytvořující funkce náhodné veličiny $X$ |
| $\psi_X$ | charakteristická funkce náhodné veličiny $X$ |

| | |
|---|---|
| $M_X$ | momentová vytvořující funkce náhodné veličiny $X$ |
| $K_X$ | vytvořující funkce kumulantů náhodné veličiny $X$ |
| $\mathrm{Re}(z)$ | reálná část komplexního čísla $z$ |
| $\mathrm{Im}(z)$ | imaginární část komplexního čísla $z$ |
| $\mathsf{Po}(\lambda)$ | Poissonovo rozdělení s parametrem $\lambda$ |
| $\mathsf{Ge}(p)$ | geometrické rozdělení s parametrem $p$ |
| $\mathsf{NB}(\pi, \kappa)$ | negativně binomické rozdělení s parametry $\kappa$ a $\pi$ |
| $\mathsf{N}(\mu, \sigma^2)$ | normální rozdělení s parametry $\mu$ a $\sigma^2$ |
| $\mathsf{LN}(\mu, \sigma^2)$ | logaritmicko-normální rozdělení s parametry $\mu$ a $\sigma^2$ |
| $\mathsf{NBLN}(\pi, \kappa, \mu, \sigma^2)$ | složené rozdělení, kde počet sčítanců má $\mathsf{NB}(\pi, \kappa)$ a sčítanci májí $\mathsf{LN}(\mu, \sigma^2)$ |
| $\overline{z(t)}$ | číslo komplexně sdružené k $z(t)$ |