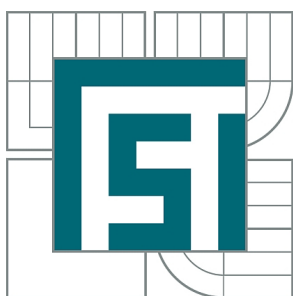


**VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ**

BRNO UNIVERSITY OF TECHNOLOGY



**FAKULTA STROJNÍHO INŽENÝRSTVÍ  
ÚSTAV MATEMATIKY**

FACULTY OF MECHANICAL ENGINEERING  
INSTITUTE OF MATHEMATICS

## **ROZDĚLENÍ EXTRÉMNÍCH HODNOT A JEJICH APLIKACE**

EXTREME VALUE DISTRIBUTIONS WITH APPLICATIONS

**DIZERTAČNÍ PRÁCE**

DOCTORAL THESIS

**AUTOR PRÁCE**

AUTHOR

Ing. MICHAL FUSEK

**VEDOUCÍ PRÁCE**

SUPERVISOR

doc. RNDr. JAROSLAV MICHÁLEK, CSc.

BRNO 2013



## Abstrakt

Práce je zaměřena na rozdělení extrémních hodnot a jejich aplikace. V úvodní části jsou položeny základy teorie extrémních hodnot pro jednorozměrná pozorování. Pomocí limitní věty pro rozdělení maxim jsou zavedeny tři typy extrémálních rozdělení (Gumbelovo, Fréchetovo, Weibullovo) včetně charakterizace jejich oborů atraktivity. Dále jsou popsány dva modely pro odhady parametrických funkcí rozdělení extrémních hodnot vycházející ze zobecněného rozdělení extrémních hodnot (model blokových maxim) a zobecněného Paretova rozdělení (prahový model). Pro tato rozdělení jsou odvozeny odhady parametrů metodou maximální věrohodnosti a metodou pravděpodobnostně vážených momentů. Popsané metody jsou následně použity k analýze srážkových úhrnů v brněnském regionu. Dále je pozornost věnována Gumbelově třídě rozdělení, která se v praxi často vyskytuje. V práci jsou odvozeny metody pro statistickou inferenci mnohonásobně zleva cenzorovaných (cenzorování typu I) výběrů z exponenciálního a Weibullova rozdělení, které jsou následně použity k analýze koncentrací syntetických musk sloučenin. Poslední část práce shrnuje základní poznatky z teorie extrémních hodnot pro dvourozměrná pozorování. Součástí práce je také vytvořený demonstrační software pro rozdělení extrémních hodnot.

## Abstract

The thesis is focused on extreme value distributions and their applications. Firstly, basics of the extreme value theory for one-dimensional observations are summarized. Using the limit theorem for distribution of maximum, three extreme value distributions (Gumbel, Fréchet, Weibull) are introduced and their domains of attraction are described. Two models for parametric functions estimation based on the generalized extreme value distribution (block maxima model) and the generalized Pareto distribution (threshold model) are introduced. Parameters estimates of these distributions are derived using the method of maximum likelihood and the probability weighted moment method. Described methods are used for analysis of the rainfall data in the Brno Region. Further attention is paid to Gumbel class of distributions, which is frequently used in practice. Methods for statistical inference of multiply left-censored samples from exponential and Weibull distribution considering the type I censoring are developed and subsequently used in the analysis of synthetic musk compounds concentrations. The last part of the thesis deals with the extreme value theory for two-dimensional observations. Demonstrational software for the extreme value distributions was developed as a part of this thesis.



## **Klíčová slova**

Rozdělení extrémních hodnot, obor atraktivity, index extrémní hodnoty, bloková maxima, prahový model, maximální věrohodnost, pravděpodobnostně vážené momenty, mnohonásobné cenzorování zleva.

## **Keywords**

Extreme value distribution, domain of attraction, extreme value index, block maxima, threshold model, maximum likelihood, probability weighted moments, multiply left-censored.

FUSEK, M. *Rozdělení extrémních hodnot a jejich aplikace*. Brno: Vysoké učení technické v Brně, Fakulta strojního inženýrství, 2013. 136 s. Vedoucí dizertační práce doc. RNDr. Jaroslav Michálek, CSc.



Prohlašuji, že jsem dizertační práci *Rozdělení extrémních hodnot a jejich aplikace* vypracoval samostatně pod vedením doc. RNDr. Jaroslava Michálka, CSc. s použitím materiálů uvedených v seznamu literatury.

V Brně dne 13. 6. 2013

Ing. Michal Fusek





Touto cestou bych chtěl poděkovat doc. RNDr. Jaroslavu Michálkovi, CSc. za vedení mé dizertační práce, za trpělivost a množství času, které mi věnoval, a za nespočet cenných rad a podnětů nejen k dizertační práci. Děkuji také své rodině za neutuchající podporu během celého studia.



# Obsah

Úvod	13
<b>1 Teorie extrémních hodnot pro jednorozměrná pozorování</b>	<b>17</b>
1.1 Formulace problému . . . . .	17
1.2 Rozdělení extrémního typu . . . . .	19
1.3 Základní typy rozdělení extrémních hodnot a jejich obor atraktivity . . . .	26
1.4 Modely pro odhady parametrických funkcí rozdělení extrémních hodnot . .	32
<b>2 Inference o parametrech rozdělení extrémních hodnot</b>	<b>37</b>
2.1 Metoda maximální věrohodnosti . . . . .	39
2.2 Metoda pravděpodobnostně vážených momentů . . . . .	45
<b>3 Testování hypotéz o parametrech rozdělení extrémních hodnot</b>	<b>49</b>
3.1 Testy dobré shody . . . . .	49
3.2 Asymptotické testy . . . . .	51
<b>4 Statistická analýza dešťových srážek</b>	<b>55</b>
4.1 Vzorkovací techniky . . . . .	55
4.2 Analýza srážek v brněnském regionu . . . . .	56
<b>5 Inference pro cenzorovaná rozdělení z Gumbelova oboru atraktivity</b>	<b>65</b>
5.1 Cenzorované exponenciální rozdělení . . . . .	67
5.2 Cenzorované Weibullovo rozdělení . . . . .	78
5.3 Test přípustnosti nahrazení cenzorovaného Weibullova rozdělení exponenci- álním submodelem . . . . .	90
<b>6 Statistická analýza koncentrací musk sloučenin</b>	<b>91</b>
<b>7 Teorie extrémních hodnot pro dvourozměrná pozorování</b>	<b>97</b>
Závěr	103
Dodatek A Odvození EV indexů	107
Dodatek B Software EVDgraph a EVDsim	121
Dodatek C Přehled programů k analýze cenzorovaných výběrů	125
Literatura	127
Seznam symbolů a zkratk	135



# Úvod

Teorie extrémních hodnot je specifické odvětví matematické statistiky, které se zabývá rozvojem metod a technik pro popis, modelování a predikci neobvyklých a málo častých jevů, které se mohou vyskytnout v mnoha oblastech lidské činnosti. V takových situacích je vždy třeba odhadnout nebo předpovědět úroveň hodnot nějakého reálného procesu, obvykle mimo oblast dosud pozorovaných dat. Přestože jsou takové předpovědi většinou postaveny na důvěře ve správnost konkrétního matematického modelu, jsou v literatuře popsány statistické modely a analýzy, které vedly ke kvalitním predikcím v dané oblasti. Z úspěšných historických aplikací, které dokumentují možnosti teorie extrémních hodnot, lze jako příklad zmínit předpovědi v oblasti radioaktivních emisí provedené Gumbelem v roce 1937 (viz [41]), dále analýzy pevnosti materiálu (viz [94]), analýzy záplav (viz např. [42, 43, 44, 45, 80]), seizmické analýzy (viz [77]) nebo analýzy dešťových srážek (viz [79]).

Prudký rozvoj teorie extrémních hodnot doprovázený jejím použitím v řadě odvětví dokumentuje skutečnost, že do roku 2000 se v literatuře objevilo více než 1000 titulů, které se zabývají jejími úspěšnými aplikacemi (viz [59]). Z moderních aplikací lze zmínit demografické předpovědi délky života provedené de Haanem v roce 2006 na datovém souboru osob narozených v Nizozemí v letech 1877–1881, a které přežily 1. 1. 1971 (viz [22]), dále pak analýzy predikce pevnosti materiálu (viz [92]), modelování velikosti vln oceánu (viz [21]), modelování selhání paměťových buněk (viz [74]), analýzu dat o větru (viz [48]), modelování termodynamiky zemětřesení (viz [63]), analýzu manažerských strategií (viz [19]), zpracování biomedicínských dat (viz [81]), posouzení klimatických změn (viz [89]) nebo aplikace v potravinářství (viz [58]).

Je samozřejmé, že praktické požadavky na předpověď extrémních hodnot datových souborů měly vliv na prudký rozvoj teorie extrémních hodnot, zejména pak podnítily rozvoj související matematické teorie rozdělení extrémních hodnot (zkráceně EV rozdělení z anglického Extrême Value distribution).

Z historického hlediska se traduje jako první úloha v oblasti extrémních hodnot stanovení průměrné největší vzdálenosti  $n$  bodů náhodně vržených na interval  $\langle 0; 1 \rangle$  od levého krajního bodu (bodu 0) tohoto intervalu (Nicolas Bernoulli, 1709). Matematický přístup k rozdělení extrémní (maximální nebo minimální) hodnoty daného statistického souboru se objevuje v pracích autorů Fuller (1914), Griffith (1920) nebo Bortkiewicz (1922). Následovaly práce von Mises (1923) zabývající se stanovením střední hodnoty největšího pozorování, práce Dodda (1923) zabývající se stanovením mediánu největšího pozorování a dále práce Fréchet (1927) zabývající se stanovením možného asymptotického rozdělení největšího pozorování. V roce 1928 Fisher a Tippett ukázali, že limitní rozdělení může být pouze jediné ze tří možných typů. Později v roce 1943 Gnedenko položil rigorózní matematické základy teorie extrémních hodnot a uvedl nutné a postačující podmínky pro slabou konvergenci extrémních pořadových statistik. Na něj navázal de Haan, který v le-

tech 1970 a 1971 pomocí teorie regulárně se měnících funkcí teorii EV rozdělení elegantně rozvinul. V současnosti je velká pozornost věnována konstrukci odhadů parametrů rozdělení extrémního typu v závislosti na vlastnostech a tvaru studovaných datových souborů. Detailní citace uvedených prací i další související literární zdroje lze nalézt v monografiích [8, 15, 22, 59]. Z uvedeného přehledu aplikací teorie EV rozdělení je dobře patrné, že moderní aplikace rozdělení extrémních hodnot se vyskytují v nejrůznějších oborech.

Předložená dizertační práce je rozčleněna na sedm kapitol a tři dodatky. V kapitole 1 jsou shrnuty základní myšlenky teorie extrémních hodnot pro jednorozměrná pozorování. Především je zde uvedena a dokázána limitní věta pro rozdělení maxim a popsány tři základní typy EV rozdělení (Gumbelovo, Fréchetovo, Weibullovo) včetně charakterizace jejich oborů atraktivity. V neposlední řadě jsou zde odvozeny dva modely pro odhady parametrických funkcí rozdělení extrémních hodnot vycházející ze zobecněného rozdělení extrémních hodnot (model blokových maxim) a zobecněného Paretova rozdělení (prahový model), které budou využity v kapitole 4.

Kapitola 2 je věnována metodám odhadu parametrů zobecněného EV rozdělení a zobecněného Paretova rozdělení. Jsou zde shrnuty základy z teorie maximální věrohodnosti a pro uvedená rozdělení jsou odvozeny odhady parametrů metodou maximální věrohodnosti a metodou pravděpodobnostně vážených momentů, které budou využity v kapitole 4.

V kapitole 3 jsou pak stručně popsány vybrané testy dobré shody (Pearsonův  $\chi^2$  test, Kolmogorovův-Smirnovův test, Andersonův-Darlingův test) a asymptotické testy včetně testů s rušivými parametry, které budou použity v kapitolách 4–6.

Metody popsané v kapitolách 1–3 jsou v kapitole 4 použity k analýze reálných hydrologických dat. Konkrétně se bude jednat o statistickou analýzu srážkových úhrnů v brněnském regionu, která vychází z autorových prací [49, 50, 75].

Kapitola 5 je věnována statistické inferenci pro vybraná cenzorovaná rozdělení z oboru atraktivity Gumbelova rozdělení, protože Gumbelova třída se v praxi často vyskytuje. Kapitola byla motivována potřebou analyzovat reálná chemická data, která jsou reprezentována dvojnásobně zleva cenzorovaným výběrem s cenzorováním typu I. Jelikož dosud nebyly dostupné žádné statistické metody, které by bylo možné snadno aplikovat na tento typ dat, bylo nutné vyvinout zcela novou metodiku, která vychází zejména z autorových prací [33, 34, 35]. Konkrétně se jedná o metody pro analýzu mnohonásobně zleva cenzorovaných výběrů z exponenciálního a Weibullova rozdělení. Model pro analýzu jednoho výběru je dále rozšířen na model pro dva nezávislé cenzorované výběry. V neposlední řadě jsou zde v návaznosti na kapitolu 3 odvozeny testovací statistiky pro srovnání dvou nezávislých mnohonásobně zleva cenzorovaných výběrů z exponenciálního a Weibullova rozdělení.

V kapitole 6 jsou metody odvozené v kapitole 5 použity ke zpracování reálných chemických dat. Konkrétně se bude jednat o analýzu koncentrací syntetických musk sloučenin, jejichž přítomnost byla odhalena v rybí tkáni, a posouzení vlivu čistírny odpadních vod na koncentraci musk sloučenin v rybí tkáni. Kapitola vychází zejména z autorových prací

[36, 37].

V kapitole 7 jsou shrnuty základní poznatky z teorie extrémních hodnot pro dvourozměrná pozorování. V dodatku A jsou pak odvozeny indexy extrémních hodnot pro všechna rozdělení z oboru atraktivity EV rozdělení, která jsou uvedena v kapitole 1. V dodatku B je popsán autorem vytvořený demonstrační software pro rozdělení extrémních hodnot, který umožňuje uživateli získat základní představu o tvaru EV rozdělení a o rozděleních z jejich oborů atraktivity. Pomocí tohoto softwaru je také možné posoudit rychlost konvergence daného rozdělení k příslušnému limitnímu EV rozdělení. Všechny metody popsané v kapitole 5 byly pro účely kapitoly 6 softwarově implementovány v prostředí Matlab, přičemž přehled jednotlivých funkcí lze najít v dodatku C.

V celé práci bude veškerá pozornost soustředěna pouze na spojitá rozdělení s absolutně spojitou distribuční funkcí. Předpoklad absolutní spojitosti distribučních funkcí nebude dále připomínán, nicméně důsledně dodržován.





# 1 Teorie extrémních hodnot pro jednorozměrná pozorování

V této kapitole budou uvedeny základní poznatky z teorie extrémních hodnot pro jednorozměrná pozorování. Nejprve bude popsána souvislost mezi vlastnostmi výběrových průměrů a asymptotickou teorií výběrových extrémních hodnot (maxim nebo minim). Dále bude pomocí limitní věty pro rozdělení maxim ukázáno, že EV rozdělení může být pouze jednoho ze tří typů (Gumbelovo, Fréchetovo, Weibullovo). V závěru kapitoly budou popsány dva klasické modely užívané při vyšetřování extrémních hodnot.

## 1.1 Formulace problému

Nechť  $X_1, \dots, X_n$  je náhodný výběr z rozdělení s distribuční funkcí  $F$ , tedy jinými slovy  $X_1, \dots, X_n$  jsou nezávislé náhodné veličiny se stejnou distribuční funkcí  $F$ . Dále  $x^* := \sup\{x : F(x) < 1\}$  bude označovat pravý koncový bod rozdělení s distribuční funkcí  $F$ . To znamená, že  $x^*$  je maximální hodnota, které může náhodná veličina  $X$  s distribuční funkcí  $F$  nabýt. Může nastat i situace, kdy pravý koncový bod  $x^* = \infty$ . Cílem teorie extrémních hodnot je vyšetřit pravděpodobnostní chování maxima  $M_n = \max\{X_1, \dots, X_n\}$  pro velké hodnoty  $n$ , tedy zkoumat limitní chování náhodné veličiny  $M_n$  pro  $n \rightarrow \infty$ . Pozornost bude tedy věnována asymptotické teorii výběrových extrémních hodnot.

Využití asymptotické teorie je v praxi velmi časté. Jde například o situace, kdy rozdělení součtu  $\sum_{i=1}^n X_i$  nebo průměru  $\bar{X}_n = 1/n \sum_{i=1}^n X_i$  náhodného výběru lze aproximovat normálním rozdělením. Ze zákona velkých čísel (viz [5]) je dobře známo, že v případě, kdy rozdělení dané distribuční funkcí  $F(x)$  má střední hodnotu  $\mu$ , platí

$$\bar{X}_n \rightarrow \mu$$

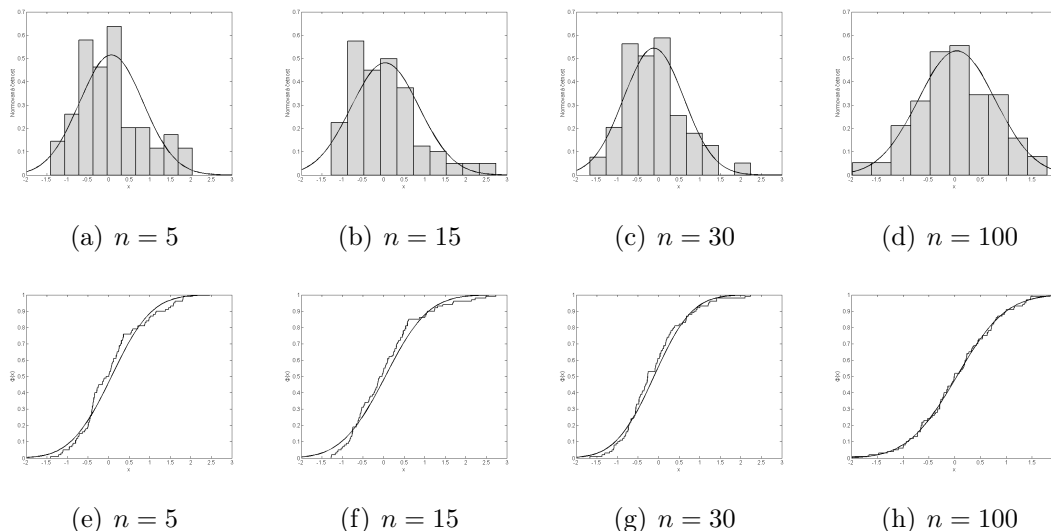
podle pravděpodobnosti pro  $n \rightarrow \infty$ . Existuje-li konečný rozptyl  $\sigma^2$  rozdělení s distribuční funkcí  $F(x)$ , platí podle centrální limitní věty (viz [5]), že standardizovaný výběrový průměr  $\sqrt{n}(\bar{X}_n - \mu)/\sigma$  má asymptoticky normální rozdělení  $N(0; 1)$ . Tedy pro všechna reálná  $x$  platí, že

$$P\left(\frac{\bar{X}_n - b_n}{a_n} \leq x\right) \rightarrow \Phi(x)$$

pro  $n \rightarrow \infty$ , kde  $a_n$  a  $b_n$ ,  $n = 1, 2, \dots$ , jsou posloupnosti vhodných reálných konstant a  $\Phi(x)$  označuje distribuční funkci normovaného normálního rozdělení  $N(0; 1)$ .

V praktických aplikacích je pak zásadní otázkou, kdy lze rozdělení výběrového průměru dostatečně přesně aproximovat normálním rozdělením. V případě, kdy je rozdělení náhodné veličiny  $X$  symetrické, stačí poměrně malý rozsah náhodného výběru  $n$  (často i  $n < 30$ ) a aproximace rozdělení výběrového průměru normálním rozdělením je poměrně kvalitní. Na druhou stranu, je-li rozdělení náhodné veličiny  $X$  silně asymetrické, je třeba k dobré aproximaci pomocí normálního rozdělení poměrně velký rozsah výběru. Na obrázku

1.1 je tato situace graficky demonstrována pro exponenciální rozdělení s parametrem  $\lambda = 2$ , distribuční funkcí  $F(x) = 1 - \exp(-\lambda x)$  a konstantami  $a_n = 1/(\lambda\sqrt{n})$  a  $b_n = 1/\lambda$ . Nejprve byl vygenerován náhodný výběr rozsahu  $n = 5, 15, 30, 100$  z exponenciálního rozdělení. Následně byl spočítán výběrový průměr každého výběru. Pomocí 100 opakovaných simulací byl sestaven histogram rozdělení četností výběrových průměrů a stanovena příslušná výběrová distribuční funkce. Histogram byl porovnán s hustotou a empirická distribuční funkce s distribuční funkcí normovaného normálního rozdělení. Z obrázku 1.1 je dobře patrná rychlost konvergence rozdělení výběrového průměru k normálnímu rozdělení.



Obrázek 1.1: Ilustrace centrální limitní věty pro exponenciální rozdělení s parametrem  $\lambda = 2$  a distribuční funkcí  $F(x) = 1 - \exp(-\lambda x)$  pro rozsah výběru  $n$ .

Problematika studovaná v teorii EV rozdělení je do značné míry paralelní s asymptotickou teorií, která se zabývá asymptotickým chováním výběrových průměrů. Rozdíl je v tom, že výběrový průměr se v teorii EV rozdělení nahrazuje maximem  $M_n = \max\{X_1, \dots, X_n\}$ . Platí následující věta, která je obdobou zákona velkých čísel pro výběrový průměr.

**Věta 1.1.** *Nechť  $x^*$  je pravý koncový bod rozdělení s distribuční funkcí  $F$  a  $M_n = \max\{X_1, \dots, X_n\}$ . Pak platí, že*

$$M_n \rightarrow x^*$$

*podle pravděpodobnosti pro  $n \rightarrow \infty$ .*

*Důkaz.* Nejprve je třeba odvodit distribuční funkci veličiny  $M_n$ , pro kterou platí

$$P(M_n \leq x) = P(\max\{X_1, \dots, X_n\} \leq x) = P\left(\bigcap_{i=1}^n [X_i \leq x]\right) = \prod_{i=1}^n P(X_i \leq x) = F^n(x).$$

Dále se rozliší dva případy:

(1) Jestliže  $x^* < \infty$ , pak pro všechna  $\epsilon > 0$  platí, že

$$\lim_{n \rightarrow \infty} P(|M_n - x^*| > \epsilon) = \lim_{n \rightarrow \infty} [1 - P(|M_n - x^*| \leq \epsilon)]$$

$$\begin{aligned}
&= \lim_{n \rightarrow \infty} [1 - P(x^* - \epsilon \leq M_n \leq x^* + \epsilon)] \\
&= \lim_{n \rightarrow \infty} [1 - F^n(x^* + \epsilon) + F^n(x^* - \epsilon)] = \lim_{n \rightarrow \infty} F^n(x^* - \epsilon) = 0.
\end{aligned}$$

(2) Jestliže  $x^* = \infty$ , pak pro všechna  $x \in \mathbb{R}$  platí, že

$$\lim_{n \rightarrow \infty} P(M_n > x) = \lim_{n \rightarrow \infty} [1 - P(M_n \leq x)] = \lim_{n \rightarrow \infty} [1 - F^n(x)] = 1.$$

□

Dále pro každé reálné  $x$  platí

$$F^n(x) \rightarrow G(x)$$

pro  $n \rightarrow \infty$ , kde limitní funkce  $G(x)$  je degenerovaná distribuční funkce (tj.  $G(x) = 0$  pro  $x < x^*$  a  $G(x) = 1$  pro  $x \geq x^*$ ) v případě, že pravý koncový bod  $x^*$  je konečný. Není-li pravý koncový bod  $x^*$  konečný, je  $G(x)$  identicky rovna nule. Z toho důvodu se podobně jako v centrální limitní větě studuje limitní rozdělení normované náhodné veličiny  $M_n$ , tedy rozdělení náhodné veličiny

$$Z_n = \frac{M_n - b_n}{a_n},$$

kde  $a_n > 0$  a  $b_n$ ,  $n = 1, 2, \dots$ , jsou posloupnosti vhodných reálných konstant. V případě, že takové limitní rozdělení existuje, je nedegenerované a jeho distribuční funkce je  $G(x)$ , lze problém nalezení asymptotického rozdělení přeformulovat pomocí distribučních funkcí. Úkolem je pak zjistit, kdy existuje nedegenerovaná distribuční funkce  $G(x)$  taková, že pro  $n \rightarrow \infty$  existuje limita

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G(x) \quad (1.1)$$

pro každý bod spojitosti limitní funkce  $G(x)$ . Dále je třeba najít nutné a postačující podmínky pro existenci této limity. V případě, že je podmínka (1.1) splněna, rozdělení pravděpodobností reprezentované nedegenerovanou distribuční funkcí  $G$  se nazývá EV rozdělení. Třída všech distribučních funkcí  $F$ , pro které platí rovnice (1.1), se pak nazývá obor atraktivity distribuční funkce  $G$ . Další pozornost bude věnována pouze rozdělení maxim  $M_n = \max \{X_1, \dots, X_n\}$ , viz následující poznámka.

**Poznámka 1.2.** Jelikož platí  $\min \{X_1, \dots, X_n\} = -\max \{-X_1, \dots, -X_n\}$ , lze výsledky získané pro rozdělení maxim  $M_n = \max \{X_1, \dots, X_n\}$  využít i v situaci, kdy je třeba vyšetřit chování výběrových minim.

Další odstavec bude věnován nalezení limitní distribuční funkce  $G$ .

## 1.2 Rozdělení extrémního typu

V tomto odstavci, který je zpracován podle [22], bude stanoveno EV rozdělení reprezentované distribuční funkcí  $G$ . K tomu je nutné určit limitu uvedenou ve vzorci (1.1). Přímý

výpočet této limity však není snadný a k podstatně jednoduššímu výpočtu vede použití kvantilové funkce. Vztah (1.1) se tedy přeformuluje pomocí kvantilové funkce, spočítá se požadovaná limita a potom se zpětně přejde k funkci distribuční. Jde tedy o obdobnou situaci jako v důkazu centrální limitní věty, kdy se od distribuční funkce přejde k funkci charakteristické, s její pomocí se nalezne charakteristická funkce limitního rozdělení a zpětnou transformací se přejde opět k funkci distribuční.

**Definice 1.3.** Nechť  $h$  je neklesající reálná funkce. Pak funkci  $h^{\leftarrow}$  ve tvaru

$$h^{\leftarrow}(y) = \inf\{x : h(x) \geq y\}, \quad y \in \mathbb{R},$$

nazveme zleva spojitou inverzní funkcí k funkci  $h$ .

**Poznámka 1.4.** Je-li funkce  $h$  z definice 1.3 rovna distribuční funkci  $F$ , pak zřejmě funkce  $F^{\leftarrow}$  je dobře známá kvantilová funkce.

Pomocí funkce  $h^{\leftarrow}$  se zavede kvantilová funkce chvostu příslušná distribuční funkci  $F$ .

**Definice 1.5.** Nechť  $F$  je distribuční funkce. Funkci  $U$  určenou předpisem

$$U(t) = \left( \frac{1}{1 - F(t)} \right)^{\leftarrow}, \quad t > 1, \quad (1.2)$$

nazveme kvantilovou funkcí chvostu příslušnou distribuční funkci  $F$ .

**Lemma 1.6.** *Kvantilovou funkcí chvostu z definice 1.5 lze vyjádřit ve tvaru*

$$U(t) = F^{\leftarrow} \left( 1 - \frac{1}{t} \right), \quad t > 1.$$

*Důkaz.* Úpravou vztahu (1.2) dostáváme

$$U^{\leftarrow}(t) = \frac{1}{1 - F(t)}. \quad (1.3)$$

Dosazením substituce

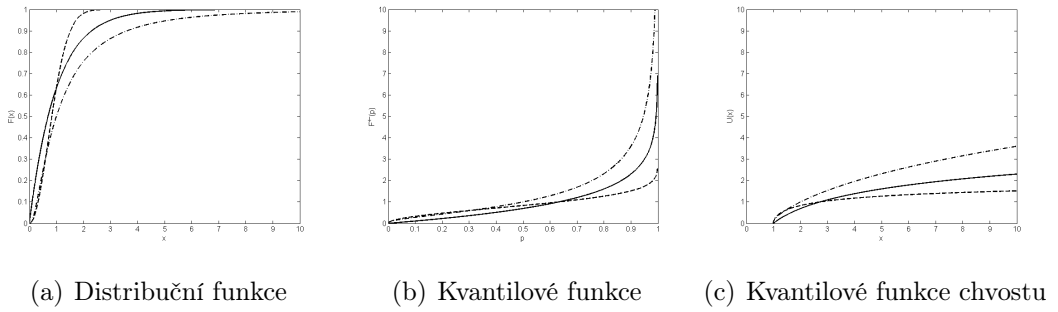
$$\begin{aligned} U^{\leftarrow}(t) &= x, \\ t &= U(x), \end{aligned}$$

do (1.3) dostáváme

$$\begin{aligned} x &= \frac{1}{1 - F(U(x))}, \\ F(U(x)) &= 1 - \frac{1}{x}, \\ U(x) &= F^{\leftarrow} \left( 1 - \frac{1}{x} \right), \quad x > 1, \end{aligned}$$

což je tvrzení lemmatu. □

Kvantilová funkce chvostu má v teorii EV rozdělení podobnou roli jako charakteristická funkce při studiu asymptotického rozdělení součtů nezávislých a stejně rozdělených náhodných veličin. Výhodou této funkce je, že umožňuje získat dobrou představu o chování distribuční funkce  $F$  v oblasti, kde  $F$  nabývá hodnot „blízkých“ jedné. Tato skutečnost je dobře patrná z obrázku 1.2, kde jsou znázorněny distribuční funkce, kvantilová funkce a kvantilová funkce chvostu pro tři vybraná rozdělení. Konkrétně se jedná o exponenciální rozdělení s distribuční funkcí  $F(x) = 1 - \exp(-\lambda x)$ , kvantilovou funkcí  $F^{\leftarrow}(p) = -\ln(1-p)/\lambda$  a kvantilovou funkcí chvostu  $U(x) = \ln(x)/\lambda$  pro  $\lambda = 1$ , Weibullovo rozdělení s distribuční funkcí  $F(x) = 1 - \exp(-\lambda x^\tau)$ , kvantilovou funkcí  $F^{\leftarrow}(p) = [-\ln(1-p)/\lambda]^{1/\tau}$  a kvantilovou funkcí chvostu  $U(x) = [\ln(x)/\lambda]^{1/\tau}$  pro  $\lambda = 1, \tau = 2$ , a lognormální rozdělení s distribuční funkcí  $F(x) = \Phi((\ln(x) - \mu)/\sigma)$  pro  $\mu = 0, \sigma = 1$  (kvantilové funkce nelze explicitně vyjádřit).



Obrázek 1.2: Srovnání distribuční funkce  $F$ , kvantilové funkce  $F^{\leftarrow}$  a kvantilové funkce chvostu  $U$  pro exponenciální (plnou čarou), Weibullovo (čárkovaně) a lognormální (čerchovaně) rozdělení.

Nyní lze výpočet limity ve vztahu (1.1) přeformulovat tak, že distribuční funkce  $F$  se nahradí kvantilovou funkcí chvostu. Logaritmováním a úpravou rovnice (1.1) se dostane

$$\lim_{n \rightarrow \infty} n \ln F(a_n x + b_n) = \ln G(x). \quad (1.4)$$

Dále lze užitím L'Hospitalova pravidla ukázat, že

$$\lim_{n \rightarrow \infty} \frac{-\ln F(a_n x + b_n)}{1 - F(a_n x + b_n)} = 1,$$

a tedy ze vztahu (1.4) postupně plyne, že

$$\begin{aligned} \lim_{n \rightarrow \infty} n(1 - F(a_n x + b_n)) &= -\ln G(x), \\ \lim_{n \rightarrow \infty} \frac{1}{n(1 - F(a_n x + b_n))} &= \frac{1}{-\ln G(x)}. \end{aligned}$$

Odtud lze užitím zleva spojitě inverzní funkce získat ekvivalentní vyjádření rovnice (1.1) pomocí kvantilové funkce chvostu ve tvaru

$$\lim_{n \rightarrow \infty} \frac{U(nx) - b_n}{a_n} = G^{\leftarrow}\left(e^{-\frac{1}{x}}\right) =: D(x), \quad (1.5)$$

přičemž bylo zavedeno označení  $D(x) := G^{\leftarrow}(e^{-\frac{1}{x}})$ .

**Věta 1.7.** *Nechť  $a_n > 0$  a  $b_n$  jsou posloupnosti reálných konstant a  $G$  nedegenerovaná distribuční funkce. Následující tvrzení jsou ekvivalentní:*

1. *Vztah*

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G(x) \quad (1.6)$$

*platí pro každý bod spojitosti  $x$  funkce  $G$ .*

2. *Vztah*

$$\lim_{t \rightarrow \infty} \frac{U(tx) - b(t)}{a(t)} = D(x) \quad (1.7)$$

*platí pro každý bod spojitosti  $x > 0$  funkce  $D(x) = G^-(e^{-1/x})$ , kde  $a(t) := a_{[t]}$ ,  $b(t) := b_{[t]}$  ( $[t]$  označuje celou část  $t$ ).*

*Důkaz.* Ekvivalence (1.5) a (1.6) byla ověřena již dříve. Proto nyní stačí ukázat, že (1.5) implikuje (1.7). Nechť  $x$  je bodem spojitosti funkce  $D(x)$ . Pak pro  $t \geq 1$  platí, že

$$\frac{U([t]x) - b_{[t]}}{a_{[t]}} \leq \frac{U(tx) - b_{[t]}}{a_{[t]}} \leq \frac{U([t]x(1 + 1/[t])) - b_{[t]}}{a_{[t]}}. \quad (1.8)$$

Pro libovolný bod spojitosti  $x_0 > x$  funkce  $D$  s vlastností  $D(x_0) > D(x)$  platí, že pravá strana vztahu (1.8) je menší než  $D(x_0)$ . Jelikož  $x$  je bodem spojitosti funkce  $D(x)$ , dostaneme

$$\lim_{t \rightarrow \infty} \frac{U(tx) - b_{[t]}}{a_{[t]}} = D(x).$$

□

K nalezení limitní distribuční funkce  $G$  tedy stačí výpočtem limity ve vztahu (1.5) najít funkci  $D$  a následně pomocí funkce  $D^-$  stanovit distribuční funkci  $G$ . Třidu EV rozdělení popisuje následující věta, která historicky vychází z prací [31, 38].

**Věta 1.8 (Fisherova a Tippetova (1928), Gnedenkova (1943)).** *Třída EV rozdělení je definována distribuční funkcí  $G_\gamma(\sigma x + \mu)$  pro  $\mu \in \mathbb{R}$ ,  $\sigma > 0$  a  $\gamma \in \mathbb{R}$ , kde*

$$G_\gamma(x) = \begin{cases} \exp \left[ -(1 + \gamma x)^{-\frac{1}{\gamma}} \right] & \text{pro } \gamma \neq 0 \text{ a } 1 + \gamma x > 0, \\ \exp(-e^{-x}) & \text{pro } \gamma = 0. \end{cases} \quad (1.9)$$

*Důkaz.* Uvažujme třídu limitních funkcí  $D(x)$  ve tvaru (1.7) a předpokládejme, že  $x = 1$  je bodem spojitosti funkce  $D(x)$ . Dále položme  $E(x) := D(x) - D(1)$ , přičemž pro všechny body spojitosti  $x > 0$  platí

$$E(x) = D(x) - D(1) = \lim_{t \rightarrow \infty} \frac{U(tx) - b(t)}{a(t)} - \lim_{t \rightarrow \infty} \frac{U(t) - b(t)}{a(t)} = \lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)}. \quad (1.10)$$

Dále pro  $x, y > 0$  užitím úpravy

$$\frac{U(txy) - U(t)}{a(t)} = \frac{U(txy) - U(ty) + U(ty) - U(t)}{a(t)}$$

dostaneme

$$\frac{U(txy) - U(t)}{a(t)} = \frac{U(txy) - U(ty)}{a(ty)} \frac{a(ty)}{a(t)} + \frac{U(ty) - U(t)}{a(t)}. \quad (1.11)$$

Jelikož ze vztahu (1.10) plyne existence limit

$$\lim_{t \rightarrow \infty} \frac{U(txy) - U(t)}{a(t)}, \lim_{t \rightarrow \infty} \frac{U(txy) - U(ty)}{a(ty)}, \lim_{t \rightarrow \infty} \frac{U(ty) - U(t)}{a(t)},$$

existuje pro  $y > 0$  také limita

$$\lim_{t \rightarrow \infty} \frac{a(ty)}{a(t)} =: A(y)$$

a pro  $x, y > 0$  platí

$$E(xy) = E(x)A(y) + E(y). \quad (1.12)$$

Nyní položíme  $s := \ln x$  a  $t := \ln y$  pro  $x, y \neq 1$ , z čehož plyne  $x = e^s$ ,  $y = e^t$ . Dosazením do rovnice (1.12) dostaneme

$$E(e^{s+t}) = E(e^s)A(e^t) + E(e^t). \quad (1.13)$$

Položíme-li  $H(t) := E(e^t)$ , můžeme přepsat (1.13) jako

$$H(s+t) = H(s)A(e^t) + H(t). \quad (1.14)$$

Protože  $H(0) = 0$ , můžeme (1.14) dále přepsat jako

$$\frac{H(s+t) - H(t)}{s} = A(e^t) \frac{H(s) - H(0)}{s}. \quad (1.15)$$

Jelikož je  $H$  monotónní, má derivaci skoro všude a z (1.15) plyne, že má derivaci všude a pro  $s \rightarrow 0$  dostaneme

$$H'(t) = A(e^t)H'(0). \quad (1.16)$$

Poznamenejme, že  $H'(0) \neq 0$ , protože  $H$  nemůže být konstantní, neboť funkce  $G$  je nederivovaná. Dále položíme  $Q(t) := H(t)/H'(0)$ , přičemž  $Q(0) = 0$  a  $Q'(0) = 1$ . Z (1.14) plyne

$$Q(s+t) = A(e^t)Q(s) + Q(t),$$

což můžeme s využitím (1.16) přepsat jako

$$Q(s+t) = Q'(t)Q(s) + Q(t). \quad (1.17)$$

Zaměníme-li nyní  $s$  a  $t$  v (1.17), dostaneme  $Q(s+t) = Q'(s)Q(t) + Q(s)$ , což položíme do rovnosti s (1.17) a postupně dostaneme

$$\begin{aligned} Q'(s)Q(t) + Q(s) &= Q'(t)Q(s) + Q(t), \\ Q(t) \frac{Q'(s) - 1}{s} &= \frac{Q(s)}{s} (Q'(t) - 1), \\ Q(t) \frac{Q'(s) - Q'(0)}{s} &= \frac{Q(s) - Q(0)}{s} (Q'(t) - 1), \end{aligned}$$

z čehož pro  $s \rightarrow 0$  plyne

$$\begin{aligned} Q(t)Q''(0) &= Q'(0)(Q'(t) - 1), \\ Q(t)Q''(0) &= Q'(t) - 1. \end{aligned} \tag{1.18}$$

Jelikož existuje derivace  $Q''(t)$ , derivováním (1.18) dostaneme

$$\begin{aligned} Q'(t)Q''(0) &= Q''(t), \\ Q''(0) &= \frac{Q''(t)}{Q'(t)}. \end{aligned}$$

Položíme-li nyní  $Q''(0) =: \gamma \in \mathbb{R}$ , pak pro všechna  $t$  platí

$$\frac{Q''(t)}{Q'(t)} = (\ln Q'(t))' = \gamma.$$

Odtud vzhledem k počáteční podmínce  $Q(0) = 0$  dostaneme

$$\begin{aligned} Q'(t) &= e^{\gamma t}, \\ Q(t) &= \int_0^t e^{\gamma s} ds. \end{aligned}$$

Protože  $Q(t) = \frac{H(t)}{H'(0)}$ , postupně

$$\begin{aligned} H(t) &= H'(0) \int_0^t e^{\gamma s} ds, \\ H(t) &= H'(0) \frac{e^{\gamma t} - 1}{\gamma}. \end{aligned}$$

Dále  $H(t) = E(e^t)$ ,  $E(x) = D(x) - D(1)$  a  $e^t = x$ , tedy

$$\begin{aligned} E(e^t) &= H'(0) \frac{e^{\gamma t} - 1}{\gamma}, \\ D(e^t) &= D(1) + H'(0) \frac{e^{\gamma t} - 1}{\gamma}, \\ D(x) &= D(1) + H'(0) \frac{x^\gamma - 1}{\gamma}. \end{aligned} \tag{1.19}$$

Položíme-li  $y := D(x)$ , pak  $D^{-1}(y) = x$  a úpravou (1.19) postupně dostaneme

$$\begin{aligned} x^\gamma &= 1 + \gamma \frac{y - D(1)}{H'(0)}, \\ x &= \left( 1 + \gamma \frac{y - D(1)}{H'(0)} \right)^{\frac{1}{\gamma}}, \\ D^{-1}(y) &= \left( 1 + \gamma \frac{y - D(1)}{H'(0)} \right)^{\frac{1}{\gamma}}. \end{aligned} \tag{1.20}$$

Ze vztahu  $y = D(x) = G^{-1}(e^{-\frac{1}{x}})$  plyne, že

$$G(y) = e^{-\frac{1}{x}},$$



$$x = -\frac{1}{\ln G(y)}.$$

Tedy  $D^-(y) = -\frac{1}{\ln G(y)}$ , což po dosazení do (1.20) dává

$$-\frac{1}{\ln G(y)} = \left(1 + \gamma \frac{y - D(1)}{H'(0)}\right)^{\frac{1}{\gamma}},$$

$$G(y) = \exp \left[ - \left(1 + \gamma \frac{y - D(1)}{H'(0)}\right)^{-\frac{1}{\gamma}} \right].$$

Označíme-li nyní  $\mu := D(1)$  a  $\sigma := H'(0)$ , dostaneme

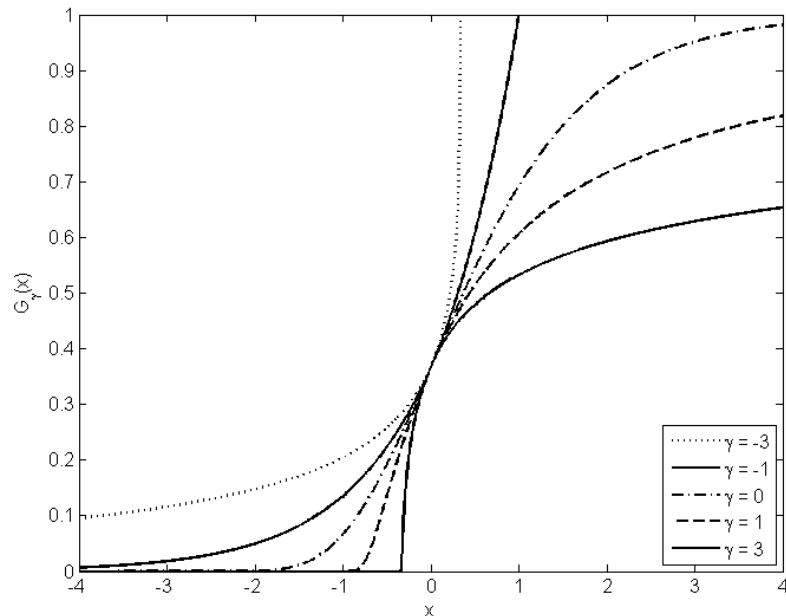
$$G(y) = \exp \left[ - \left(1 + \gamma \frac{y - \mu}{\sigma}\right)^{-\frac{1}{\gamma}} \right],$$

$$G(\sigma x + \mu) = \exp \left[ - (1 + \gamma x)^{-\frac{1}{\gamma}} \right],$$

pro  $1 + \gamma x > 0$ , z čehož plyne tvrzení věty.

Na začátku důkazu jsme pro zjednodušení předpokládali, že  $x = 1$  je bodem spojitosti funkce  $D(x)$ . Pokud by tento předpoklad neplatil, lze v důkazu postupovat podobně s tím, že by se místo funkce  $U(t)$  uvažovala funkce  $U(tx_0)$ , kde  $x_0$  by byl bodem spojitosti funkce  $D$ .

□



Obrázek 1.3: Distribuční funkce EV rozdělení pro různé hodnoty EV indexu  $\gamma$ .

**Definice 1.9.** Parametr  $\gamma$  ve vztahu (1.9) se nazývá index extrémní hodnoty (zkráceně EV index z anglického Extrême Value index).

Ze vztahu (1.9) je patrné, že třída EV rozdělení je až na parametr polohy  $\mu$  a parametr měřítka  $\sigma$  tvořena jednoparametrickou třídou rozdělení s EV indexem  $\gamma$ . Vliv indexu  $\gamma$  na tvar EV rozdělení je graficky znázorněn na obrázku 1.3, z kterého je vidět, že existují tři zásadně odlišné typy EV rozdělení, které budou podrobněji rozebrány v následujícím odstavci. Nejprve lze však s využitím výsledků věty 1.8 přeformulovat tvrzení věty 1.7.

**Věta 1.10.** *Pro  $\gamma \in \mathbb{R}$  jsou následující tvrzení ekvivalentní:*

1. *Existují takové posloupnosti reálných konstant  $a_n > 0$  a  $b_n$ , že*

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G_\gamma(x) = \exp \left[ - (1 + \gamma x)^{-\frac{1}{\gamma}} \right] \quad (1.21)$$

*pro všechna  $x$  a  $1 + \gamma x > 0$ .*

2. *Existuje taková kladná funkce  $a$ , že pro  $x > 0$  platí*

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = D_\gamma(x) = \frac{x^\gamma - 1}{\gamma}, \quad (1.22)$$

*přičemž pro  $\gamma = 0$  je pravá strana vztahu (1.22) rovna  $\ln x$ .*

*Navíc vztah (1.21) zůstává v platnosti pro  $a_n := a(n)$  a  $b_n := U(n)$ .*

*Důkaz.* Ekvivalence (1.21) a (1.22) byla dokázána ve větě 1.7. □

## 1.3 Základní typy rozdělení extrémních hodnot a jejich obor atraktivivity

Jak již bylo zmíněno v předchozím odstavci, rozlišují se tři speciální typy EV rozdělení, které se liší hodnotou EV indexu  $\gamma$ . Konkrétně se jedná o situaci, kdy  $\gamma > 0$ ,  $\gamma = 0$  nebo  $\gamma < 0$ . Jednotlivé případy budou nyní rozebrány. Bez újmy na obecnosti bude dále uvažován nulový parametr polohy  $\mu$  a jednotkový parametr měřítka  $\sigma$ .

### 1.3.1 EV index $\gamma > 0$

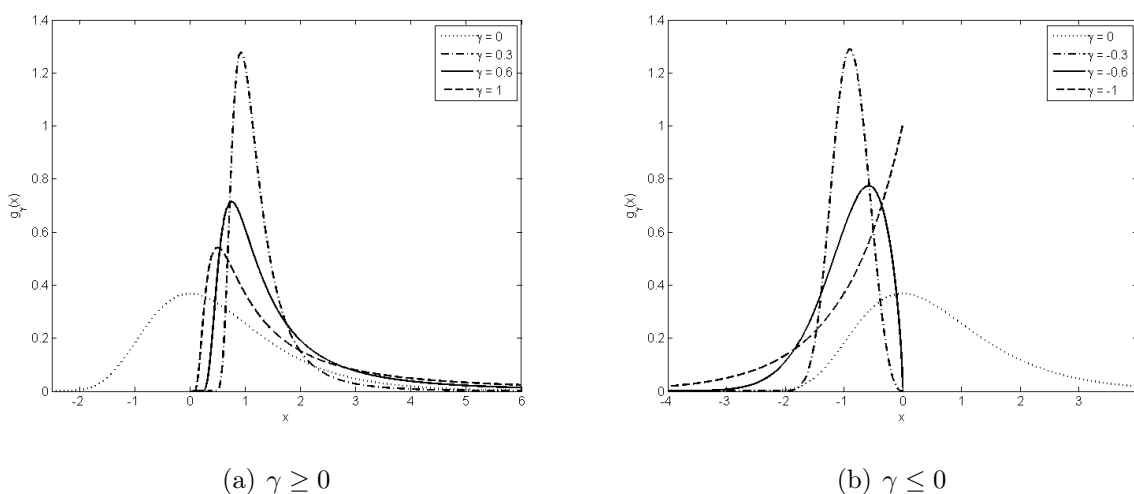
Rozdělení v této třídě mají pravý koncový bod  $x^* = \infty$ , tedy distribuční funkce  $G_\gamma(x) < 1$  pro všechna  $x \in \mathbb{R}$ . Dále platí, že

$$\lim_{x \rightarrow \infty} \frac{1 - G_\gamma(x)}{\gamma^{-\frac{1}{\gamma}} x^{-\frac{1}{\gamma}}} = 1,$$

tedy rozdělení má těžký pravý chvost, což kupříkladu znamená, že neexistují jeho momenty řádu většího nebo rovného  $\frac{1}{\gamma}$ . Zavede-li se reparametrizace a pro  $\alpha = \frac{1}{\gamma} > 0$  se položí  $H_\alpha(x) := G_\gamma((x-1)/\gamma)$ , pak

$$H_\alpha(x) = \begin{cases} \exp(-x^{-\alpha}) & \text{pro } x > 0, \\ 0 & \text{pro } x \leq 0. \end{cases} \quad (1.23)$$

Třída rozdělení s distribuční funkcí (1.23) se nazývá Fréchetova třída rozdělení. Hustoty rozdělení z této třídy jsou pro vybrané hodnoty EV indexu  $\gamma$  graficky znázorněny na obrázku 1.4a.



Obrázek 1.4: Hustota EV rozdělení pro různé hodnoty EV indexu  $\gamma$ .

### 1.3.2 EV index $\gamma = 0$

Rozdělení v této třídě mají pravý koncový bod  $x^* = \infty$  a platí, že

$$\lim_{x \rightarrow \infty} \frac{1 - G_0(x)}{e^{-x}} = 1,$$

tedy rozdělení nemá těžký pravý chvost, což znamená, že všechny jeho momenty existují. Třída rozdělení s distribuční funkcí

$$G_0(x) = \exp(-e^{-x}), \quad x \in \mathbb{R},$$

se nazývá Gumbelova třída rozdělení (někdy též rozdělení dvojité-exponenciálního typu). Hustota tohoto rozdělení ve vztahu ke zbývajícím dvěma typům je graficky znázorněna na obrázcích 1.4a a 1.4b.

### 1.3.3 EV index $\gamma < 0$

Rozdělení v této třídě mají konečný pravý koncový bod a platí, že  $x^* = -\frac{1}{\gamma}$ . Dále platí, že

$$\lim_{x \downarrow 0} \frac{1 - G_\gamma(-\frac{1}{\gamma} - x)}{(-\gamma x)^{-\frac{1}{\gamma}}} = 1,$$

tedy rozdělení má krátký pravý chvost. Zavede-li se reparametrizace a pro  $\alpha = -\frac{1}{\gamma} > 0$  se položí  $H_\alpha(x) := G_\gamma(-\frac{1+x}{\gamma})$ , pak

$$H_\alpha(x) = \begin{cases} \exp(-(-x)^\alpha) & \text{pro } x < 0, \\ 1 & \text{pro } x \geq 0. \end{cases} \quad (1.24)$$

Třída rozdělení s distribuční funkcí (1.24) se nazývá Weibullova třída rozdělení. Hustoty rozdělení z této třídy jsou pro vybrané hodnoty EV indexu  $\gamma$  graficky znázorněny na obrázku 1.4b.

Jak již bylo zmíněno, pokud platí

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G_\gamma(x) \quad (1.25)$$

pro nějaké  $\gamma \in \mathbb{R}$ , pak distribuční funkce  $F$  patří do oboru atraktivity  $G_\gamma$ , což se zapíše jako  $F \in \mathcal{D}(G_\gamma)$ . V závislosti na hodnotě parametru  $\gamma$  lze tedy rozlišit obory atraktivity Gumbelova, Fréchetova a Weibullova rozdělení. Nyní bude odvozena postačující podmínka kladená na distribuční funkci  $F$  garantující existenci posloupností konstant  $a_n > 0$  a  $b_n$  takových, že pro všechna  $x$  a nějaké  $\gamma \in \mathbb{R}$  je splněna podmínka (1.25).

**Věta 1.11 (von Misesova).** *Nechť  $x^*$  je pravý koncový bod rozdělení s distribuční funkcí  $F$ . Dále předpokládejme, že existuje druhá derivace  $F''(x)$  a první derivace  $F'(x) > 0$  pro všechna  $x$  v nějakém levém okolí  $x^*$ . Jestliže platí tzv. von Misesova podmínka*

$$\lim_{t \uparrow x^*} \left( \frac{1 - F(t)}{F'(t)} \right)' = \gamma, \quad (1.26)$$

pak  $F$  patří do oboru atraktivity  $G_\gamma$ .

*Důkaz.* Z definice kvantilové funkce chvostu (viz definice 1.5 a zejména lemmatu 1.6) plyne, že

$$t = \frac{1}{1 - F(U(t))}. \quad (1.27)$$

Derivací (1.27) dostaneme

$$1 = \frac{F'(U(t))U'(t)}{[1 - F(U(t))]^2},$$

což po úpravě dává

$$U'(t) = \frac{[1 - F(U(t))]^2}{F'(U(t))}. \quad (1.28)$$

Derivací (1.28) dostaneme

$$U''(t) = \frac{-2[1 - F(U(t))][F'(U(t))]^2 U'(t) - [1 - F(U(t))]^2 F''(U(t))U'(t)}{[F'(U(t))]^2},$$

což po úpravě dává

$$\begin{aligned} \frac{U''(t)}{U'(t)} &= -2[1 - F(U(t))] - \frac{[1 - F(U(t))]^2 F''(U(t))}{[F'(U(t))]^2}, \\ \frac{tU''(t)}{U'(t)} &= -2 - \frac{[1 - F(U(t))] F''(U(t))}{[F'(U(t))]^2}. \end{aligned} \quad (1.29)$$

Jednoduchou úpravou vztahu (1.26) lze ukázat, že

$$\lim_{t \uparrow x^*} \frac{[1 - F(t)] F''(t)}{[F'(t)]^2} = -\gamma - 1,$$

z čehož lze odvodit, že

$$\lim_{t \rightarrow \infty} \frac{tU''(t)}{U'(t)} = \gamma - 1. \quad (1.30)$$

Nyní je třeba ukázat, že (1.30) implikuje

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma},$$

což je ekvivalentní podmínka pro příslušnost funkce  $F$  do oboru atraktivity  $G_\gamma$ , která plyne z vět 1.7 a 1.10. Jelikož pro  $1 < x_0 < x$  platí, že

$$\ln U'(x) - \ln U'(x_0) = \int_{x_0}^x \frac{U''(s)}{U'(s)} ds,$$

pro  $x > 0$  a  $t, tx > 1$  platí, že

$$\ln U'(tx) - \ln U'(t) = \int_1^x \frac{tsU''(ts)}{sU'(ts)} ds.$$

Položíme-li  $A(t) := \frac{tU''(t)}{U'(t)}$ , pak vzhledem k (1.30) platí

$$\ln U'(tx) - \ln U'(t) = \int_1^x A(ts) \frac{ds}{s} = (\gamma - 1) \ln x = \ln x^{\gamma-1}.$$

Dále pro  $0 < a < b < \infty$  platí, že

$$\lim_{t \rightarrow \infty} \sup_{a \leq x \leq b} \left| \ln \frac{U'(tx)}{U'(t)} - \ln x^{\gamma-1} \right| = 0.$$

Jelikož pro nějaké kladné  $k$  na kompaktním intervalu platí  $|e^s - e^t| < k|s - t|$ , pak

$$\lim_{t \rightarrow \infty} \sup_{a \leq x \leq b} \left| \frac{U'(tx)}{U'(t)} - x^{\gamma-1} \right| = 0,$$

z čehož plyne, že

$$\frac{U(tx) - U(t)}{tU'(t)} - \frac{x^\gamma - 1}{\gamma} = \int_1^x \left( \frac{U'(ts)}{U'(t)} - s^{\gamma-1} \right) ds$$

konverguje k nule. □

**Poznámka 1.12.** Za platnosti (1.26) lze posloupnosti konstant  $a_n, b_n$  ve vztahu (1.25) volit ve tvaru  $a_n = nU'(n) = 1/(nF'(b_n))$  a  $b_n = U(n)$ .

Uvedená podmínka pro příslušnost  $F$  do oboru atraktivity  $G_\gamma$  není jediná. V monografii [22] lze nalézt jejich zevrubný přehled. Nicméně von Misesovu podmínku (1.26) lze v konkrétních praktických situacích většinou poměrně snadno ověřit a je tak možno rozhodnout o limitním chování maxim  $M_n$ , když náhodný výběr pochází z rozdělení s distribuční funkcí  $F$ . Vybraná rozdělení patří do oboru atraktivity EV rozdělení tří výše uvedených typů

Rozdělení	$F(x)$
Benktanderovo II	$1 - x^{\beta-1} \exp\left(\frac{\alpha}{\beta} - \frac{\alpha}{\beta}x^\beta\right),$ $x \geq 1, \alpha > 0, 0 < \beta \leq 1$
Weibullovo	$1 - e^{-\lambda x^\tau},$ $x > 0; \lambda > 0, \tau > 0$
Exponenciální	$1 - e^{-\lambda x},$ $x > 0; \lambda > 0$
Gama	$1 - \int_x^\infty \frac{\lambda^m}{\Gamma(m)} t^{m-1} e^{-\lambda t} dt,$ $x > 0; \lambda > 0, m > 0$
Logistické	$1 - \frac{1}{1+e^x},$ $x \in \mathbb{R}$
Lognormální	$1 - \int_x^\infty \frac{1}{\sqrt{2\pi}\sigma t} \exp\left[-\frac{(\ln t - \mu)^2}{2\sigma^2}\right] dt,$ $x > 0; \mu \in \mathbb{R}, \sigma > 0$

Tabulka 1.1: Rozdělení z oboru atraktivity Gumbelova rozdělení.

Rozdělení	$F(x)$	EV index
Rovnoměrné	$x - x^* + 1,$ $x^* - 1 < x < x^*$	-1
Beta	$1 - \int_{x-x^*+1}^1 \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} t^{p-1} (1-t)^{q-1} dt,$ $x^* - 1 < x < x^*; p > 0, q > 0$	$-\frac{1}{q}$
Reverzní Burrovo	$1 - \left[\frac{\beta}{\beta+(x^*-x)^{-\tau}}\right]^\lambda,$ $x < x^*; \beta > 0, \lambda > 0, \tau > 0$	$-\frac{1}{\lambda\tau}$
Extremální Weibullovo	$\exp(-(x^* - x)^\alpha),$ $x < x^*; \alpha > 0$	$-\frac{1}{\alpha}$

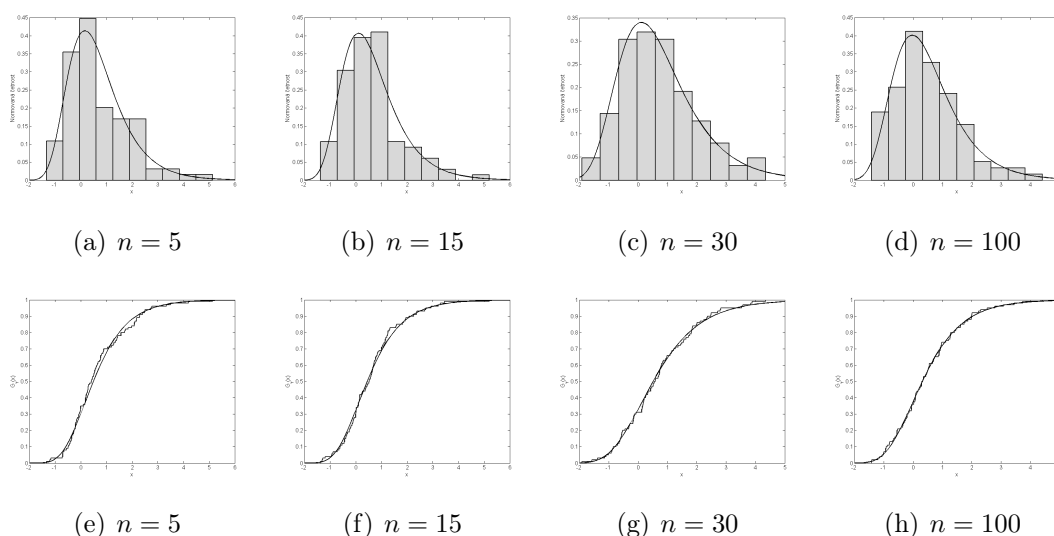
Tabulka 1.2: Rozdělení z oboru atraktivity Weibullova rozdělení.

jsou společně s jejich distribučními funkcemi a EV indexy uvedena v tabulce 1.1 pro rozdělení Gumbelovo (EV index je roven nule), v tabulce 1.2 pro rozdělení Weibullovo a v tabulce 1.3 pro rozdělení Fréchetovo. Platnost von Misesovy podmínky (1.26) byla autorem ověřena pro všechna rozdělení uvedená v tabulkách 1.1–1.3, přičemž kompletní odvození příslušných EV indexů je uvedeno v kapitole Dodatek A.

Představu o tvaru jednotlivých rozdělení uvedených v tabulkách 1.1–1.3 je možné získat pomocí autorem vytvořeného softwarového produktu, který je popsán v kapitole

Rozdělení	$F(x)$	EV index
Paretovo	$1 - x^{-\alpha},$ $x > 0, \alpha > 0$	$\frac{1}{\alpha}$
Zobecněné Paretovo	$1 - \left(1 + \frac{\xi x}{\sigma}\right)^{-\frac{1}{\xi}}$ $x > 0; \sigma > 0, \xi > 0$	$\xi$
Burrovo (Typ XII)	$1 - \left(\frac{\eta}{\eta + x^\tau}\right)^\lambda,$ $x > 0; \eta > 0, \lambda > 0, \tau > 0$	$\frac{1}{\lambda\tau}$
Burrovo (Typ III)	$\left(\frac{\eta}{\eta + x^{-\tau}}\right)^\lambda,$ $x > 0; \eta > 0, \lambda > 0, \tau > 0$	$\frac{1}{\tau}$
$F(m, n)$	$1 - \int_x^\infty \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \left(\frac{m}{n}\right)^{m/2} t^{m/2-1} \left(1 + \frac{m}{n}t\right)^{-(m+n)/2} dt,$ $x > 0; m > 0, n > 0$	$\frac{2}{n}$
Inverzní gama	$1 - \int_x^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} \exp(-\lambda/t)t^{-\alpha-1} dt,$ $x > 0; \alpha > 0, \lambda > 0$	$\frac{1}{\alpha}$
Logaritmické gama	$1 - \int_x^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{-\lambda-1} (\ln t)^{\alpha-1} dt,$ $x > 1; \alpha > 0, \lambda > 0$	$\frac{1}{\lambda}$
Fréchetovo	$\exp(-x^{-\alpha}),$ $x > 0; \alpha > 0$	$\frac{1}{\alpha}$

Tabulka 1.3: Rozdělení z oboru atraktivity Fréchetova rozdělení.



Obrázek 1.5: Ilustrace limitní věty pro rozdělení maxim pro exponenciálního rozdělení.

Dodatek B. S jeho pomocí je rovněž možné posoudit rychlost konvergence daného rozdělení k příslušnému limitnímu EV rozdělení. S využitím tohoto softwaru byly pro ilustraci

vygenerovány náhodné výběry rozsahu  $n = 5, 15, 30, 100$  z exponenciálního rozdělení. Následně bylo spočítáno maximum každého výběru. Pomocí 100 opakovaných simulací byl následně sestaven histogram rozdělení četností výběrových maxim a stanovena příslušná výběrová distribuční funkce. Histogram byl porovnán s hustotou a empirická distribuční funkce s distribuční funkcí Gumbelova rozdělení. Z obrázku 1.5 je dobře patrná rychlost konvergence rozdělení výběrového maxima k limitnímu Gumbelovu rozdělení.

Další odstavec bude zaměřen na modely pro odhady parametrických funkcí EV rozdělení.

## 1.4 Modely pro odhady parametrických funkcí rozdělení extrémních hodnot

Nyní budou popsány dva přístupy, které lze použít při analýze a modelování extrémních hodnot posloupnosti nezávislých pozorování  $X_1, \dots, X_n$ . Nejprve bude zaveden model blokových maxim, který je založený na zobecněném EV rozdělení, a následně bude popsána jeho souvislost s prahovým modelem založeným na zobecněném Paretově rozdělení.

### 1.4.1 Model blokových maxim

V odstavci 1.2 byla zavedena třída EV rozdělení pomocí distribuční funkce  $G_\gamma(\sigma x + \mu)$ . Z funkce  $G_\gamma$ , viz (1.9), lze po jednoduché úpravě získat obecnou distribuční funkci EV rozdělení v závislosti na parametrech  $\mu$ ,  $\sigma$  a  $\gamma$  ve tvaru

$$G(x) = \begin{cases} \exp \left\{ - \left[ 1 + \gamma \left( \frac{x-\mu}{\sigma} \right) \right]^{-\frac{1}{\gamma}} \right\} & \text{pro } \gamma \neq 0 \text{ a } 1 + \gamma \left( \frac{x-\mu}{\sigma} \right) > 0, \\ \exp \left\{ - \exp \left[ - \left( \frac{x-\mu}{\sigma} \right) \right] \right\} & \text{pro } \gamma = 0, \end{cases} \quad (1.31)$$

kde  $\mu \in \mathbb{R}$  je parametr polohy,  $\sigma > 0$  je parametr měřítka a  $\gamma \in \mathbb{R}$  je EV index. Rozdělení s distribuční funkcí (1.31) se nazývá zobecněné EV rozdělení (zkráceně GEV rozdělení z anglického Generalized Extrême Value distribution).

V praktických aplikacích se GEV rozdělení používá k aproximaci rozdělení maxima  $M_n = \max \{X_1, \dots, X_n\}$ , když je dán náhodný výběr  $X_1, \dots, X_n$  z rozdělení s neznámou distribuční funkcí  $F$  dostatečně velkého rozsahu. V takovém případě je datový soubor rozdělen na bloky určité velikosti a z každého bloku se spočítá maximum. Rozdělení členů posloupnosti maxim lze pak aproximovat GEV rozdělením, které je možné využít ke stanovení odhadů vybraných parametrických funkcí. Vzhledem k aplikacím v této práci (viz kapitola 4) bude pozornost věnována odhadu kvantilu  $x_{1-p}$  GEV rozdělení, pomocí kterého lze vyjádřit funkci  $z_p = x_{1-p}$ , která se nazývá úroveň návratu přiřazená periodě návratu  $1/p$  a hojně se používá v hydrologii při modelování srážkového úhrnu. V takovém případě je datový soubor reprezentován časovou řadou srážkových úhrnů.



Jestliže je časová řada rozdělena na bloky o délce jeden rok, pak úroveň návratu  $z_p$  je hodnota měřené veličiny, která je v průměru překročena jednou za  $1/p$  let. Jinými slovy, sledovaná veličina může nabýt hodnoty  $z_p$  nebo větší v libovolném roce s pravděpodobností  $p$ . Úroveň návratu  $z_p$  je tvaru

$$z_p = \begin{cases} \mu - \frac{\sigma}{\gamma} \{1 - [-\ln(1-p)]^{-\gamma}\} & \text{pro } \gamma \neq 0, \\ \mu - \sigma \ln[-\ln(1-p)] & \text{pro } \gamma = 0, \end{cases} \quad (1.32)$$

přičemž její odhad  $\hat{z}_p$  lze získat nahrazením parametrů  $\mu, \sigma, \gamma$  ve vztahu (1.32) jejich odhady  $\hat{\mu}, \hat{\sigma}, \hat{\gamma}$ . Odhady parametrů GEV rozdělení metodou maximální věrohodnosti a metodou pravděpodobnostně vážených momentů jsou odvozeny v kapitole 2. Na základě vlastností maximálně věrohodných odhadů (zkráceně MV odhadů) lze tzv. delta metodou (viz např. [15]) získat odhad rozptylu odhadu úrovně návratu ve tvaru

$$\text{Var}(\hat{z}_p) \approx \nabla z_p^T \mathbf{V} \nabla z_p,$$

kde

$$\mathbf{V} = \mathbf{V}(\hat{\mu}, \hat{\sigma}, \hat{\gamma})$$

je varianční matice, která je odvozena v odstavci 2.1.1, a

$$\nabla z_p^T = \left( \frac{\partial z_p}{\partial \mu}, \frac{\partial z_p}{\partial \sigma}, \frac{\partial z_p}{\partial \gamma} \right) = \left[ 1, -\frac{1 - y_p^{-\gamma}}{\gamma}, \frac{\sigma}{\gamma^2} (1 - y_p^{-\gamma}) - \frac{\sigma}{\gamma} y_p^{-\gamma} \ln y_p \right],$$

kde  $y_p = -\ln(1-p)$  a parametry  $\mu, \sigma, \gamma$  jsou nahrazeny jejich MV odhady  $\hat{\mu}, \hat{\sigma}, \hat{\gamma}$ .

#### 1.4.2 Prahový model

Jak již bylo uvedeno, GEV rozdělení se používá k aproximaci rozdělení maxima  $M_n = \max\{X_1, \dots, X_n\}$ . Formálně zapsáno, vychází se z asymptotického vztahu

$$P(M_n \leq x) \approx G(x) = \exp \left\{ - \left[ 1 + \gamma \left( \frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\gamma}} \right\}.$$

Při hledání odhadu EV indexu  $\gamma$  je také možné vyjít ze vzájemného vztahu mezi GEV rozdělením a zobecněným Paretoovým rozdělením. Nechť tedy  $X_1, \dots, X_n$  je posloupnost nezávislých a stejně rozdělených náhodných veličin s distribuční funkcí  $F$  a nechť

$$M_n = \max\{X_1, \dots, X_n\}$$

má asymptoticky GEV rozdělení s distribuční funkcí

$$G(x) = \exp \left\{ - \left[ 1 + \gamma \left( \frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\gamma}} \right\}$$

pro nějaké  $\mu$ ,  $\sigma > 0$  a  $\gamma$ . Pak lze podmíněné rozdělení náhodné veličiny  $Y = X - u$  za podmínky  $X > u$  pro dostatečně velké  $u$  aproximovat zobecněným Paretovým rozdělením s distribuční funkcí

$$H(y) = 1 - \left(1 + \frac{\gamma y}{\sigma_u}\right)^{-\frac{1}{\gamma}}$$

pro  $\{y : y > 0 \wedge (1 + \gamma y/\sigma_u) > 0\}$ , kde  $\sigma_u = \sigma + \gamma(u - \mu)$ .

**Poznámka 1.13.** Jestliže lze rozdělení blokových maxim sledované veličiny aproximovat GEV rozdělením, pak lze rozdělení sledované veličiny přesahující práh  $u$  aproximovat zobecněným Paretovým rozdělením, přičemž parametr  $\gamma$  zobecněného Paretova rozdělení odpovídá parametru  $\gamma$  GEV rozdělení.

Kritickým bodem uvedeného přístupu je stanovení prahové hodnoty  $u$ , o čemž svědčí celá řada publikací (viz např. [10, 11, 70, 73, 83]). Volba nízkého prahu má za následek špatné asymptotické vlastnosti modelu a velké vychýlení odhadu. Naopak vysoká hodnota prahu vede k malému počtu hodnot nad prahem, což má za následek velký rozptyl odhadu. K výběru prahu se používají různé explorativní techniky (viz [15]):

- Mean Residual Life (MRL) plot vznikne tak, že se do grafu vynesou souřadnice bodů

$$\left(u, \frac{1}{n_u} \sum_{i=1}^{n_u} (x_{(i)} - u)\right) \text{ pro } x > u, \quad (1.33)$$

kde  $x_{(1)}, \dots, x_{(n_u)}$  označuje uspořádané realizace výběru  $x_1, \dots, x_{n_u}$  a  $n_u$  je počet pozorování, která překročí práh  $u$ . Vhodná (v jistém smyslu optimální) prahová hodnota  $u_0$  se zvolí tak, aby byla pro  $u > u_0$  znázorněná závislost v grafu přibližně lineární. Metoda vychází z faktu, že střední hodnota veličiny  $Y = X - u$  za podmínky  $X > u$  je lineární funkcí prahu  $u$  pro  $u > u_0$ , tedy

$$E(X - u | X > u) = \frac{\sigma_u}{1 - \gamma}.$$

Střední hodnota přesahu prahu  $u$  se odhadne výběrovým průměrem, z čehož již plyne podstata vztahu (1.33).

- Techniky využívající stability parametrů. Z grafického znázornění se posoudí průběh hodnot parametrů v závislosti na volbě prahu  $u > u_0$  s tím, že cílem je najít takovou hodnotu  $u_0$ , pro kterou platí:
  - (a) Parametr měřítka  $\sigma_u$  vykazuje lineární průběh pro  $u > u_0$ , protože parametr  $\sigma_u$  zobecněného Paretova rozdělení je lineární funkcí prahové hodnoty  $u$ .
  - (b) Parametr tvaru  $\gamma$  vykazuje konstantní průběh pro  $u > u_0$ , protože parametr  $\gamma$  zobecněného Paretova rozdělení odpovídá parametru  $\gamma$  GEV rozdělení, a tedy musí být konstantní pro  $u > u_0$ .

Jakmile je vybrána vhodná prahová hodnota, lze přistoupit ke statistické inferenci pro prahový model a použít zobecněné Pareto rozdělení ke stanovení odhadu EV indexu  $\gamma$  a dalších parametrických funkcí. Vzhledem k aplikacím v této práci (viz kapitola 4) bude další pozornost opět věnována odhadu úrovně návratu přiřazené periodě návratu  $1/p$ . Úroveň návratu  $x_m$  je hodnota měřené veličiny, která je ve výběru překročena v průměru jednou za  $m = 1/p$  pozorování. V anglicky psané literatuře je zvykem ji nazývat  $m$ -observation return level. Nechť je dále splněn předpoklad, že zobecněné Pareto rozdělení je vhodným modelovým rozdělením pro veličinu  $X$  přesahující práh  $u$ . Tedy pro  $x > u$  platí

$$P(X > x | X > u) = \left[ 1 + \gamma \left( \frac{x - u}{\sigma_u} \right) \right]^{-\frac{1}{\gamma}}.$$

Označuje-li  $\zeta_u = P(X > u)$ , pak

$$P(X > x) = \zeta_u \left[ 1 + \gamma \left( \frac{x - u}{\sigma_u} \right) \right]^{-\frac{1}{\gamma}}.$$

Jelikož  $p = 1/m$ , dále

$$P(X > x_m) = \zeta_u \left[ 1 + \gamma \left( \frac{x_m - u}{\sigma_u} \right) \right]^{-\frac{1}{\gamma}} = \frac{1}{m}.$$

Odtud již lze pro dostatečně velké  $m$  (aby platilo  $x_m > u$ ) snadno stanovit úroveň návratu  $x_m$  ve tvaru

$$x_m = \begin{cases} u + \frac{\sigma_u}{\gamma} [(m\zeta_u)^\gamma - 1] & \text{pro } \gamma \neq 0, \\ u + \sigma_u \ln(m\zeta_u) & \text{pro } \gamma = 0, \end{cases} \quad (1.34)$$

přičemž její odhad  $\hat{x}_m$  lze získat nahrazením parametrů  $\sigma_u$ ,  $\gamma$ ,  $\zeta_u$  ve vztahu (1.34) jejich odhady  $\hat{\sigma}_u$ ,  $\hat{\gamma}$ ,  $\hat{\zeta}_u$ . Odhady parametrů  $\sigma_u$  a  $\gamma$  zobecněného Pareto rozdělení metodou maximální věrohodnosti a metodou pravděpodobnostně vážených momentů jsou odvozeny v kapitole 2. Odhad  $\hat{\zeta}_u$  parametru  $\zeta_u$  se získá jako relativní četnost pozorování, která překročila práh  $u$ . Jelikož počet překročení práhu  $u$  má binomické rozdělení  $\text{Bi}(n, \zeta_u)$ , odhad  $\hat{\zeta}_u$  je MV odhadem parametru  $\zeta_u$ .

Odhad rozptylu odhadu úrovně návratu je opět možné získat delta metodou, přičemž rozptyl odhadu  $\hat{\zeta}_u$  lze určit z vlastností binomického rozdělení a zahrnout do varianční matice. Dále tedy  $\text{Var}(\hat{\zeta}_u) \approx \hat{\zeta}_u(1 - \hat{\zeta}_u)/n$  a varianční matice pro odhady  $\hat{\sigma}_u$ ,  $\hat{\gamma}$ ,  $\hat{\zeta}_u$  je tvaru

$$\mathbf{V}(\hat{\zeta}_u, \hat{\sigma}_u, \hat{\gamma}) = \begin{pmatrix} \hat{\zeta}_u(1 - \hat{\zeta}_u)/n & \mathbf{0} \\ \mathbf{0} & \mathbf{V}(\hat{\sigma}_u, \hat{\gamma}) \end{pmatrix},$$

kde  $\mathbf{V}(\hat{\sigma}_u, \hat{\gamma})$  je varianční matice pro odhady  $\hat{\sigma}_u$ ,  $\hat{\gamma}$ , která je odvozena v odstavci 2.1.2. Pak

$$\text{Var}(\hat{x}_m) \approx \nabla x_m^T \mathbf{V} \nabla x_m,$$

kde

$$\begin{aligned}\nabla x_m^T &= \left( \frac{\partial x_m}{\partial \zeta_u}, \frac{\partial x_m}{\partial \sigma_u}, \frac{\partial x_m}{\partial \gamma} \right) \\ &= \left\{ \sigma_u m^\gamma \zeta_u^{\gamma-1}, \frac{(m\zeta_u)^\gamma - 1}{\gamma}, -\frac{\sigma_u}{\gamma^2} [(m\zeta_u)^\gamma - 1] + \frac{\sigma_u}{\gamma} (m\zeta_u)^\gamma \ln(m\zeta_u) \right\},\end{aligned}$$

přičemž parametry  $\sigma_u$ ,  $\gamma$ ,  $\zeta_u$  jsou nahrazeny jejich MV odhady  $\hat{\sigma}_u$ ,  $\hat{\gamma}$ ,  $\hat{\zeta}_u$ .

V praxi se častěji používá tzv.  $N$ -letá úroveň návratu  $x_N$  (z anglického  $N$ -year return level), což je hodnota měřené veličiny, která je ve výběru překročena v průměru jednou za  $N$  let. Při jejím výpočtu se vychází ze vztahu (1.34) s tím, že  $m = Nn_y$ , kde  $n_y$  je počet pozorování za rok. Úroveň návratu  $x_N$  je tedy tvaru

$$x_N = \begin{cases} u + \frac{\sigma_u}{\gamma} [(Nn_y\zeta_u)^\gamma - 1] & \text{pro } \gamma \neq 0, \\ u + \sigma_u \ln(Nn_y\zeta_u) & \text{pro } \gamma = 0, \end{cases} \quad (1.35)$$

přičemž její odhad  $\hat{x}_N$  lze obdržet nahrazením parametrů  $\sigma_u$ ,  $\gamma$ ,  $\zeta_u$  ve vztahu (1.35) jejich odhady  $\hat{\sigma}_u$ ,  $\hat{\gamma}$ ,  $\hat{\zeta}_u$ . Odhad rozptylu odhadu  $N$ -leté úrovně návratu  $x_N$  se získá podobným způsobem jako pro úroveň návratu  $x_m$ .

## 2 Inference o parametrech rozdělení extrémních hodnot

Jak bylo zmíněno v odstavcích 1.2 a 1.3, v teorii extrémních hodnot hraje klíčovou roli EV index  $\gamma$ , který má zásadní vliv na tvar EV rozdělení. Odhad EV indexu je tedy stěžejní součástí analýzy EV rozdělení a lze jej získat například užitím standardních metod, tedy teorie maximální věrohodnosti, ovšem s jistým omezením. Klasická teorie MV odhadů je totiž založena na regulárním systému hustot (viz definice 2.1), ale hustota EV rozdělení není obecně regulární. Je však možné ukázat (viz [84]), že pro  $\gamma > -0,5$  mají odhady získané metodou maximální věrohodnosti obvyklé asymptotické vlastnosti, tedy rozdělení odhadu  $\hat{\gamma}$  EV indexu  $\gamma$  je asymptoticky normální. V případě, kdy  $-1 < \gamma < -0,5$ , lze MV odhady získat obvyklým způsobem, ale již nemají standardní asymptotické vlastnosti. Konečně pro  $\gamma < -1$  je málo pravděpodobné, že MV odhady EV indexu budou dosažitelné s uspokojivou přesností (rozdělení má velmi krátký ohraničený pravý chvost).

Další možností jak získat odhad EV indexu je užitím neparametrických metod. Výhodou těchto metod je, že při jejich použití odpadá předpoklad na konkrétní rozdělení pravděpodobností. Z neparametrických metod odhadu EV indexu lze zmínit např. Hillův odhad (viz [51]), Pickandsův odhad (viz [78]), momentový odhad (viz např. [22]), odhad metodou pravděpodobnostně vážených momentů (viz [53]), negativní Hillův odhad (viz [29, 85]). Nevýhodou naopak je, že přesnost odhadů EV indexu pomocí většiny výše uvedených metod je silně závislá na určení optimálního počtu horních  $k$  pořadových statistik, které jsou pro konstrukci příslušného odhadu stěžejní. Při „velkém“  $k$  může být odhad značně vychýlený, volba „malého“  $k$  má za následek velký rozptyl odhadu. Určení optimálního  $k$  je tedy kritickým bodem těchto metod. Adaptivní metody pro odhad optimálního  $k$  lze nalézt například v [9, 20, 25, 39]. Výše uvedené neparametrické metody však nejsou dostupné běžnému uživateli, protože se nevyskytují v komerčních softwarech.

V této kapitole bude pozornost upřena nejen na stanovení odhadu EV indexu, ale i na odhad zbývajících parametrů GEV rozdělení a zobecněného Paretova rozdělení, která jsou stěžejní při modelování extrémních hodnot metodou blokových maxim, případně užitím prahového modelu (viz odstavec 1.4). Výsledky z této kapitoly budou dále využity při analýze dešťových srážek v brněnském regionu (viz kapitola 4). K odhadu neznámých parametrů, které jsou klíčové pro popis pravděpodobnostního modelu výše uvedených rozdělení, bude použita metoda maximální věrohodnosti jako zástupce parametrických metod a metoda pravděpodobnostně vážených momentů jako zástupce neparametrických metod. Nejprve však bude podle [5] uvedeno několik vybraných pojmů z teorie maximální věrohodnosti, které budou potřebné dále. Více výsledků lze najít například v [5, 64].

**Definice 2.1.** Nechť  $\mathbf{X} = (X_1, \dots, X_n)^T$  je náhodný vektor s hustotou  $f(\mathbf{x}, \boldsymbol{\theta})$  vzhledem k nějaké  $\sigma$ -konečné míře  $\mu$ , přičemž  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)^T$ . Předpokládejme, že platí:

(1)  $\boldsymbol{\theta} \in \Theta$ , kde  $\Theta$  je neprázdná otevřená množina v  $\mathbb{R}^m$ .

(2) Množina  $M = \{\mathbf{x} : f(\mathbf{x}, \boldsymbol{\theta}) > 0\}$  nezávisí na  $\boldsymbol{\theta}$ .

(3) Pro skoro všechna  $\mathbf{x} \in M$  vzhledem k  $\mu$  a pro všechna  $i = 1, \dots, m$  existují parciální derivace

$$f'_i(\mathbf{x}, \boldsymbol{\theta}) = \frac{\partial f(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_i}.$$

(4) Pro každé  $i$  a pro všechna  $\boldsymbol{\theta} \in \Theta$  platí

$$\int_M f'_i(\mathbf{x}, \boldsymbol{\theta}) d\mu(\mathbf{x}) = 0.$$

(5) Pro každou dvojici  $(i, j)$  existuje konečný integrál

$$J_{ij}(\boldsymbol{\theta}) = \int_M \frac{f'_i(\mathbf{x}, \boldsymbol{\theta}) f'_j(\mathbf{x}, \boldsymbol{\theta})}{f^2(\mathbf{x}, \boldsymbol{\theta})} f(\mathbf{x}, \boldsymbol{\theta}) d\mu(\mathbf{x}).$$

(6) Matice  $\mathbf{J}_n(\boldsymbol{\theta}) = \|J_{ij}(\boldsymbol{\theta})\|_{i,j=1}^m$  je pozitivně definitní pro každé  $\boldsymbol{\theta} \in \Theta$ .

Pak se systém hustot  $\{f(\mathbf{x}, \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$  nazývá regulární a  $\mathbf{J}_n(\boldsymbol{\theta})$  se nazývá Fisherova informační matice.

**Poznámka 2.2.** Bude-li vektor parametrů  $\boldsymbol{\theta}$  jednorozměrný, bude místo termínu Fisherova informační matice (dále jen FIM) užíván termín Fisherova míra informace (dále jen FMI).

**Věta 2.3.** *Nechť systém hustot  $\{f(\mathbf{x}, \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$  je regulární. Předpokládejme, že pro skoro všechna  $\mathbf{x} \in M$  (vzhledem k  $\mu$ ) existují derivace*

$$f''_{ij}(\mathbf{x}, \boldsymbol{\theta}) = \frac{\partial^2 f(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}, \quad i, j = 1, \dots, m,$$

a že pro všechna  $\boldsymbol{\theta} \in \Theta$  platí

$$\int_M f''_{ij}(\mathbf{x}, \boldsymbol{\theta}) d\mu(\mathbf{x}) = 0, \quad i, j = 1, \dots, m.$$

Pak platí

$$J_{ij}(\boldsymbol{\theta}) = - \int_M \frac{\partial^2 \ln f(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} f(\mathbf{x}, \boldsymbol{\theta}) d\mu(\mathbf{x}), \quad i, j = 1, \dots, m. \quad (2.1)$$

*Důkaz.* Důkaz přímo nesouvisí s tématem této práce a lze jej najít například v knihách [5, 64].

□

**Definice 2.4.** FIM  $\mathbf{J}_n(\boldsymbol{\theta})$  s prvky danými vztahem (2.1) se nazývá očekávaná FIM. Matici s prvky

$$\tilde{J}_{ij}(\boldsymbol{\theta}) = - \frac{\partial^2 \ln f(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}, \quad i, j = 1, \dots, m,$$

nazveme výběrovou FIM a bude dále označena jako  $\tilde{\mathbf{J}}_n(\boldsymbol{\theta})$ .

**Poznámka 2.5.** Pro očekávanou a výběrovou FIM platí vztah  $\mathbf{J}_n(\boldsymbol{\theta}) = E\tilde{\mathbf{J}}_n(\boldsymbol{\theta})$ .

**Poznámka 2.6.** Dá se ukázat, že za předpokladů regularity je výběrová FIM nestranným odhadem očekávané FIM a  $\tilde{\mathbf{J}}_n(\boldsymbol{\theta}) \rightarrow \mathbf{J}_n(\boldsymbol{\theta})$  v pravděpodobnosti pro  $n \rightarrow \infty$ .

**Věta 2.7.** Nechť  $X_1, \dots, X_n$  je náhodný výběr z rozdělení s hustotou  $f(x, \boldsymbol{\theta})$  vzhledem k  $\sigma$ -konečné míře  $\nu$ . Nechť systém hustot  $\{f(x, \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$  je regulární a má FIM  $\mathbf{J}(\boldsymbol{\theta})$ . Pak náhodný vektor  $\mathbf{X} = (X_1, \dots, X_n)^\top$  má hustotu

$$f(\mathbf{x}, \boldsymbol{\theta}) = f(x_1, \boldsymbol{\theta}) \dots f(x_n, \boldsymbol{\theta})$$

vzhledem k míře  $\mu = \nu \times \dots \times \nu$ . Systém hustot  $\{f(\mathbf{x}, \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$  je také regulární a pro jeho FIM  $\mathbf{J}_n(\boldsymbol{\theta})$  platí  $\mathbf{J}_n(\boldsymbol{\theta}) = n\mathbf{J}(\boldsymbol{\theta})$ , kde  $\mathbf{J}(\boldsymbol{\theta})$  je informace obsažená v každém jednotlivém  $X_i$ .

*Důkaz.* Důkaz přímo nesouvisí s tématem této práce a lze jej najít například v knihách [5, 64].

□

## 2.1 Metoda maximální věrohodnosti

Metoda maximální věrohodnosti bude zavedena pouze pro regulární hustoty. Není-li systém hustot regulární, lze metodu použít s jistými omezeními. Nastane-li taková situace, bude rozebrána zvlášť.

**Definice 2.8.** Nechť  $\mathbf{X}$  je  $n$ -rozměrný náhodný vektor s hustotou  $f(\mathbf{x}, \boldsymbol{\theta})$ , kde  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)^\top \in \Theta$ . Funkce

$$L(\boldsymbol{\theta}) = f(\mathbf{x}, \boldsymbol{\theta}) = \prod_{i=1}^n f(x_i, \boldsymbol{\theta}) \quad (2.2)$$

se pak nazývá věrohodnostní funkce a funkce

$$l(\boldsymbol{\theta}) = \ln L(\boldsymbol{\theta}) \quad (2.3)$$

logaritmická věrohodnostní funkce.

**Definice 2.9.** Existuje-li hodnota  $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\mathbf{x}) \in \Theta$  maximalizující věrohodnostní funkci (2.2), případně logaritmickou věrohodnostní funkci (2.3), nazývá se maximálně věrohodný odhad (MV odhad) parametru  $\boldsymbol{\theta} \in \Theta$ .

**Definice 2.10.** Systém

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \theta_i} = \mathbf{0}, \quad i = 1, \dots, m, \quad (2.4)$$

se nazývá systém (logaritmických) věrohodnostních rovnic.

**Věta 2.11.** *Nechť systém hustot  $\{f(x, \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$  je regulární a má FIM  $\mathbf{J}(\boldsymbol{\theta})$ . Dále necht jsou splněny následující předpoklady.*

(1) *Nechť  $\Theta \subset \mathbb{R}^m$  je parametrický prostor, který obsahuje takový neprázdný otevřený interval  $\Theta$ , že skutečná hodnota parametru  $\boldsymbol{\theta}_0$  patří do  $\Theta$ .*

(2) *Nechť  $\mathbf{X} = (X_1, \dots, X_n)^\top$ , kde  $X_i$  jsou nezávislé a stejně rozdělené náhodné veličiny s hustotou  $f(x, \boldsymbol{\theta})$  vzhledem k nějaké  $\sigma$ -konečné míře  $\mu$ .*

(3) *Nechť  $M = \{x : f(x, \boldsymbol{\theta}) > 0\}$  nezávisí na  $\boldsymbol{\theta}$ .*

(4) *Nechť  $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$ . Pak  $f(x, \boldsymbol{\theta}_1) = f(x, \boldsymbol{\theta}_2)$  vzhledem k  $\mu$  právě tehdy, je-li  $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ .*

(5) *Derivace*

$$\frac{\partial^3 f(x, \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j \partial \theta_k}$$

*existuje pro skoro všechna  $x$ , pro všechna  $\boldsymbol{\theta} \in \Theta$  a pro všechna  $i, j, k = 1, \dots, m$ .*

(6) *Pro všechna  $\boldsymbol{\theta} \in \Theta$  platí*

$$\int_M f''_{ij}(x, \boldsymbol{\theta}) d\mu(x) = 0, \quad i, j = 1, \dots, m.$$

(7) *Pro všechna  $i, j, k = 1, \dots, m$  existují takové funkce  $M_{ijk}(x) \geq 0$ , že*

$$E_{\boldsymbol{\theta}_0} M_{ijk}(X) < \infty$$

*a*

$$\left| \frac{\partial^3 \ln f(x, \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j \partial \theta_k} \right| \leq M_{ijk}(x)$$

*pro všechna  $\boldsymbol{\theta} \in \Theta$  a skoro všechna  $x \in M$ .*

*Pak platí následující tvrzení.*

(a) *Jestliže  $n \rightarrow \infty$ , pak ke každému  $\epsilon > 0$  existuje s pravděpodobností blížíící se jedné takové řešení  $\hat{\boldsymbol{\theta}}_n$  systému věrohodnostních rovnic, že  $|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0| < \epsilon$ .*

(b) *Položíme-li*

$$\mathbf{U}(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial l(\boldsymbol{\theta})}{\partial \theta_1} \\ \vdots \\ \frac{\partial l(\boldsymbol{\theta})}{\partial \theta_m} \end{pmatrix}, \quad (2.5)$$

*pak pro  $n \rightarrow \infty$  platí*

$$\frac{1}{\sqrt{n}} \mathbf{U}(\boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}; \mathbf{J}(\boldsymbol{\theta}_0)),$$

*kde  $\xrightarrow{d}$  značí konvergenci v distribuci.*



(c) Existuje-li pro každé dostatečně velké  $n$  a pro každou hodnotu  $\mathbf{X}$  takový kořen  $\widehat{\boldsymbol{\theta}}_n$  systému věrohodnostních rovnic, že  $\widehat{\boldsymbol{\theta}}_n$  je konzistentním odhadem parametru  $\boldsymbol{\theta}_0$ , pak

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} \text{N}(\mathbf{0}; \mathbf{J}^{-1}(\boldsymbol{\theta}_0)).$$

*Důkaz.* Důkaz přímo nesouvisí s tématem této práce a lze jej najít například v knihách [5, 65]. □

**Poznámka 2.12.** Vektor  $\mathbf{U}(\boldsymbol{\theta})$  ze vztahu (2.5) se nazývá skórový vektor příslušný hustotě  $f(x, \boldsymbol{\theta})$ .

Dále budou odvozeny MV odhady parametrů GEV rozdělení a zobecněného Paretova rozdělení. Pomocí FIM (viz věta 2.11) pak budou stanoveny rozptyly těchto odhadů.

### 2.1.1 GEV rozdělení

Nechť  $X_1, \dots, X_n$  je náhodný výběr z GEV rozdělení s distribuční funkcí

$$F(x, \mu, \sigma, \gamma) = \exp \left\{ - \left[ 1 + \gamma \left( \frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\gamma}} \right\}$$

a hustotou

$$f(x, \mu, \sigma, \gamma) = \frac{1}{\sigma} \left[ 1 + \gamma \left( \frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\gamma}-1} \exp \left\{ - \left[ 1 + \gamma \left( \frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\gamma}} \right\} \quad (2.6)$$

pro  $x \in \{x : 1 + \gamma \left( \frac{x - \mu}{\sigma} \right) > 0\}$ . Dosazením vztahu (2.6) postupně do vztahů (2.2) a (2.3) lze dostat věrohodnostní funkci ve tvaru

$$\begin{aligned} L(\mu, \sigma, \gamma) &= \prod_{i=1}^n f(x_i, \mu, \sigma, \gamma) \\ &= \frac{1}{\sigma^n} \prod_{i=1}^n \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\gamma}-1} \exp \left\{ - \sum_{i=1}^n \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\gamma}} \right\} \end{aligned}$$

a logaritmickou věrohodnostní funkci ve tvaru

$$\begin{aligned} l(\mu, \sigma, \gamma) &= \ln L(\mu, \sigma, \gamma) = -n \ln \sigma - \left( \frac{1}{\gamma} + 1 \right) \sum_{i=1}^n \ln \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right] \\ &\quad - \sum_{i=1}^n \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\gamma}}. \end{aligned} \quad (2.7)$$

Užitím vztahů (2.4) a (2.7) lze obdržet systém věrohodnostních rovnic

$$\begin{aligned}
\frac{\partial l}{\partial \mu} &= \frac{1+\gamma}{\sigma} \sum_{i=1}^n \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-1} - \frac{1}{\sigma} \sum_{i=1}^n \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\gamma}-1} = 0, \\
\frac{\partial l}{\partial \sigma} &= -\frac{n}{\sigma} + \frac{1+\gamma}{\sigma} \sum_{i=1}^n \left( \frac{x_i - \mu}{\sigma} \right) \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-1} \\
&\quad - \frac{1}{\sigma} \sum_{i=1}^n \left( \frac{x_i - \mu}{\sigma} \right) \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\gamma}-1} = 0, \\
\frac{\partial l}{\partial \gamma} &= \frac{1}{\gamma^2} \sum_{i=1}^n \ln \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right] - \frac{1+\gamma}{\gamma} \sum_{i=1}^n \left( \frac{x_i - \mu}{\sigma} \right) \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-1} \\
&\quad - \frac{1}{\gamma^2} \sum_{i=1}^n \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\gamma}} \\
&\quad \times \left\{ \ln \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right] - \gamma \left( \frac{x_i - \mu}{\sigma} \right) \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-1} \right\} = 0,
\end{aligned} \tag{2.8}$$

jehož vyřešením lze získat MV odhady  $\hat{\mu}$ ,  $\hat{\sigma}$ ,  $\hat{\gamma}$  parametrů  $\mu$ ,  $\sigma$ ,  $\gamma$ . Soustavu je však nutné řešit numericky. Je také možné využít numerickou proceduru založenou na Nelderově-Meadově simplexovém algoritmu (viz [60]), který je implementován v Matlabu (verze 7.12, R2011a), a maximalizovat logaritmickou věrohodnostní funkci (2.7). Při numerickém výpočtu MV odhadů neznámých parametrů je často nutné stanovit počáteční (startovací) hodnoty neznámých parametrů. Jako startovací odhady mohou posloužit například odhady parametrů GEV rozdělení metodou momentů (viz např. [5]).

Dále bude pomocí výběrové FIM stanoven rozptyl odhadů  $\hat{\mu}$ ,  $\hat{\sigma}$ ,  $\hat{\gamma}$  parametrů  $\mu$ ,  $\sigma$ ,  $\gamma$ . Z definice 2.4 plyne, že výběrová FIM je tvaru

$$\tilde{J}_n(\mu, \sigma, \gamma) = \begin{pmatrix} \tilde{J}_{11} & \tilde{J}_{12} & \tilde{J}_{13} \\ \tilde{J}_{21} & \tilde{J}_{22} & \tilde{J}_{23} \\ \tilde{J}_{31} & \tilde{J}_{32} & \tilde{J}_{33} \end{pmatrix} = \begin{pmatrix} -\frac{\partial^2 l}{\partial \mu^2} & -\frac{\partial^2 l}{\partial \mu \partial \sigma} & -\frac{\partial^2 l}{\partial \mu \partial \gamma} \\ -\frac{\partial^2 l}{\partial \sigma \partial \mu} & -\frac{\partial^2 l}{\partial \sigma^2} & -\frac{\partial^2 l}{\partial \sigma \partial \gamma} \\ -\frac{\partial^2 l}{\partial \gamma \partial \mu} & -\frac{\partial^2 l}{\partial \gamma \partial \sigma} & -\frac{\partial^2 l}{\partial \gamma^2} \end{pmatrix},$$

kde

$$\begin{aligned}
\tilde{J}_{11} &= -\frac{\gamma + \gamma^2}{\sigma^2} \sum_{i=1}^n \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-2} + \frac{1+\gamma}{\sigma^2} \sum_{i=1}^n \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\gamma}-2}, \\
\tilde{J}_{12} = \tilde{J}_{21} &= \frac{1+\gamma}{\sigma^2} \sum_{i=1}^n \left\{ \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right] - \gamma \left( \frac{x_i - \mu}{\sigma} \right) \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-2} \right\} \\
&\quad + \frac{1}{\sigma^2} \sum_{i=1}^n \left\{ (1+\gamma) \left( \frac{x_i - \mu}{\sigma} \right) \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\gamma}-2} - \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\gamma}-1} \right\}, \\
\tilde{J}_{13} = \tilde{J}_{31} &= \frac{1}{\gamma \sigma} \sum_{i=1}^n \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-1} - \frac{1+\gamma}{\gamma \sigma} \sum_{i=1}^n \left\{ \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-1} \right.
\end{aligned}$$

$$\begin{aligned}
& - \gamma \left( \frac{x_i - \mu}{\sigma} \right) \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-2} \Big\} + \frac{1}{\gamma^2 \sigma} \sum_{i=1}^n \left\{ \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\gamma}-1} \right. \\
& \times \left\{ \ln \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right] - \gamma \left( \frac{x_i - \mu}{\sigma} \right) \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-1} \right\} \\
& \left. - \gamma^2 \left( \frac{x_i - \mu}{\sigma} \right) \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\gamma}-2} \right\}, \\
\tilde{J}_{22} = & - \frac{n}{\sigma^2} + \frac{1+\gamma}{\sigma^2} \sum_{i=1}^n \left\{ 2 \left( \frac{x_i - \mu}{\sigma} \right) \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-1} \right. \\
& \left. - \gamma \left( \frac{x_i - \mu}{\sigma} \right)^2 \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-2} \right\} + \frac{1}{\sigma^2} \sum_{i=1}^n \left\{ (1+\gamma) \left( \frac{x_i - \mu}{\sigma} \right)^2 \right. \\
& \left. \times \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\gamma}-2} - 2 \left( \frac{x_i - \mu}{\sigma} \right) \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\gamma}-1} \right\}, \\
\tilde{J}_{23} = \tilde{J}_{32} = & \frac{1}{\gamma \sigma} \sum_{i=1}^n \left( \frac{x_i - \mu}{\sigma} \right) \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-1} \\
& - \frac{1+\gamma}{\gamma \sigma} \sum_{i=1}^n \left\{ \left( \frac{x_i - \mu}{\sigma} \right) \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-1} - \gamma \left( \frac{x_i - \mu}{\sigma} \right)^2 \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-2} \right\} \\
& + \frac{1}{\gamma^2 \sigma} \sum_{i=1}^n \left\{ \left( \frac{x_i - \mu}{\sigma} \right) \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\gamma}-1} \left\{ \ln \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right] \right. \right. \\
& \left. \left. - \gamma \left( \frac{x_i - \mu}{\sigma} \right) \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-1} \right\} - \gamma^2 \left( \frac{x_i - \mu}{\sigma} \right)^2 \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\gamma}-2} \right\}, \\
\tilde{J}_{33} = & \frac{2}{\gamma^3} \sum_{i=1}^n \ln \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right] - \frac{2}{\gamma^2} \sum_{i=1}^n \left( \frac{x_i - \mu}{\sigma} \right) \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-1} \\
& - \left( \frac{1}{\gamma} + 1 \right) \sum_{i=1}^n \left( \frac{x_i - \mu}{\sigma} \right)^2 \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-2} + \frac{1}{\gamma^4} \sum_{i=1}^n \left\{ \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\gamma}} \right. \\
& \times \left\{ \ln \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right] - \gamma \left( \frac{x_i - \mu}{\sigma} \right) \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-1} \right\}^2 \\
& - \gamma \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\gamma}} \left\{ 2 \ln \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right] \right. \\
& \left. \left. - 2\gamma \left( \frac{x_i - \mu}{\sigma} \right) \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-1} - \gamma^2 \left( \frac{x_i - \mu}{\sigma} \right)^2 \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-2} \right\} \right\}.
\end{aligned}$$

Z věty 2.11 pak plyne, že odhad rozptylu odhadů parametrů GEV rozdělení lze získat z varianční matice  $\mathbf{V}(\mu, \sigma, \gamma) = \tilde{\mathbf{J}}^{-1}(\mu, \sigma, \gamma)$ , přičemž za  $\mu, \sigma, \gamma$  se dosadí jejich MV odhady  $\hat{\mu}, \hat{\sigma}, \hat{\gamma}$ .

### 2.1.2 Zobecněné Paretovo rozdělení

Nechť  $X_1, \dots, X_n$  je náhodný výběr ze zobecněného Paretova rozdělení s distribuční funkcí

$$F(x) = 1 - \left(1 + \gamma \frac{x}{\sigma}\right)^{-\frac{1}{\gamma}}$$

a hustotou

$$f(x) = \frac{1}{\sigma} \left(1 + \gamma \frac{x}{\sigma}\right)^{-\frac{1}{\gamma}-1} \quad (2.9)$$

pro  $x > 0$ . Dosazením vztahu (2.9) postupně do vztahů (2.2) a (2.3) lze dostat věrohodnostní funkci ve tvaru

$$L(\sigma, \gamma) = \prod_{i=1}^n f(x_i, \sigma, \gamma) = \frac{1}{\sigma^n} \prod_{i=1}^n \left(1 + \gamma \frac{x_i}{\sigma}\right)^{-\frac{1}{\gamma}-1}$$

a logaritmickou věrohodnostní funkci ve tvaru

$$l(\sigma, \gamma) = \ln L(\sigma, \gamma) = -n \ln \sigma - \left(\frac{1}{\gamma} + 1\right) \sum_{i=1}^n \ln \left(1 + \gamma \frac{x_i}{\sigma}\right). \quad (2.10)$$

Užitím vztahů (2.4) a (2.10) lze obdržet systém věrohodnostních rovnic

$$\begin{aligned} \frac{\partial l}{\partial \sigma} &= -\frac{n}{\sigma} + \frac{1+\gamma}{\sigma^2} \sum_{i=1}^n x_i \left(1 + \gamma \frac{x_i}{\sigma}\right)^{-1} = 0, \\ \frac{\partial l}{\partial \gamma} &= \frac{1}{\gamma^2} \sum_{i=1}^n \ln \left(1 + \gamma \frac{x_i}{\sigma}\right) - \frac{1+\gamma}{\gamma \sigma} \sum_{i=1}^n x_i \left(1 + \gamma \frac{x_i}{\sigma}\right)^{-1} = 0, \end{aligned} \quad (2.11)$$

jehož vyřešením lze získat MV odhady  $\hat{\sigma}$ ,  $\hat{\gamma}$  parametrů  $\sigma$ ,  $\gamma$ . Soustavu je však třeba řešit numericky. Opět lze použít numerickou proceduru založenou na Nelderově-Meadově simplexovém algoritmu (viz [60]), který je implementován v Matlabu (verze 7.12, R2011a), a maximalizovat logaritmickou věrohodnostní funkci (2.10). Jako startovací odhady mohou opět posloužit odhady parametrů zobecněného Paretova rozdělení metodou momentů (viz např. [5]).

Dále bude pomocí výběrové FIM stanoven rozptyl odhadů  $\hat{\sigma}$ ,  $\hat{\gamma}$  parametrů  $\sigma$ ,  $\gamma$ . Z definice 2.4 plyne, že výběrová FIM je tvaru

$$\tilde{\mathbf{J}}_n(\sigma, \gamma) = \begin{pmatrix} \tilde{J}_{11} & \tilde{J}_{12} \\ \tilde{J}_{21} & \tilde{J}_{22} \end{pmatrix} = \begin{pmatrix} -\frac{\partial^2 l}{\partial \sigma^2} & -\frac{\partial^2 l}{\partial \sigma \partial \gamma} \\ -\frac{\partial^2 l}{\partial \gamma \partial \sigma} & -\frac{\partial^2 l}{\partial \gamma^2} \end{pmatrix},$$

kde

$$\begin{aligned} \tilde{J}_{11} &= -\frac{n}{\sigma^2} + \frac{1+\gamma}{\sigma^4} \sum_{i=1}^n \left[ 2\sigma x_i \left(1 + \gamma \frac{x_i}{\sigma}\right)^{-1} - \gamma x_i^2 \left(1 + \gamma \frac{x_i}{\sigma}\right)^{-2} \right], \\ \tilde{J}_{12} = \tilde{J}_{21} &= \frac{1}{\gamma \sigma^2} \sum_{i=1}^n x_i \left(1 + \gamma \frac{x_i}{\sigma}\right)^{-1} - \frac{1+\gamma}{\gamma \sigma^3} \sum_{i=1}^n \left[ \sigma x_i \left(1 + \gamma \frac{x_i}{\sigma}\right)^{-1} - \gamma x_i^2 \left(1 + \gamma \frac{x_i}{\sigma}\right)^{-2} \right], \end{aligned}$$

$$\tilde{J}_{22} = \frac{2}{\gamma^3} \sum_{i=1}^n \ln \left( 1 + \gamma \frac{x_i}{\sigma} \right) - \frac{2}{\gamma^2 \sigma} \sum_{i=1}^n x_i \left( 1 + \gamma \frac{x_i}{\sigma} \right)^{-1} - \frac{1 + \gamma}{\gamma \sigma^2} \sum_{i=1}^n x_i^2 \left( 1 + \gamma \frac{x_i}{\sigma} \right)^{-2},$$

Z věty 2.11 pak plyne, že odhad rozptylu odhadů parametrů zobecněného Paretova rozdělení lze získat z varianční matice  $\mathbf{V}(\sigma, \gamma) = \tilde{\mathbf{J}}^{-1}(\sigma, \gamma)$ , přičemž za  $\sigma, \gamma$  se dosadí jejich MV odhady  $\hat{\sigma}, \hat{\gamma}$ .

## 2.2 Metoda pravděpodobnostně vážených momentů

Pravděpodobnostně vážené momenty (dále jen PVM) jako zobecnění běžných momentů rozdělení pravděpodobností byly poprvé představeny v pracích [40, 61]. Hlavní myšlenkou je stejně jako u klasické metody momentů porovnání obecných PVM s jejich empirickými protějšky.

Nechť  $X$  je náhodná veličina s distribuční funkcí  $F(x)$ . Veličiny

$$M_{p,r,s} = \mathbb{E} \{ X^p [F(X)]^r [1 - F(X)]^s \}, \quad (2.12)$$

kde  $p, r, s$  jsou reálná čísla, se nazývají pravděpodobnostně vážené momenty. Pokud je možné vyjádřit inverzní distribuční funkci  $x(F) = F^{\leftarrow}(F)$  k funkci  $F$  v explicitním tvaru, pak je příhodné přepsat (2.12) jako

$$M_{p,r,s} = \int_0^1 (F^{\leftarrow})^p F^r (1 - F)^s dF,$$

což je často nejjednodušší cesta k získání PVM.

PVM ve tvaru  $M_{p,0,0}$  ( $p = 1, 2, \dots$ ) odpovídají klasickým obecným momentům veličiny  $X$ . Často jsou však při odhadování neznámých parametrů daného rozdělení upřednostňovány PVM typu  $M_{1,r,s}$ , neboť obsahují pouze první mocninu  $X$ , a tedy vztah mezi parametry a PVM daného rozdělení jsou jednoduššího tvaru. Jsou-li  $r$  a  $s$  celá čísla, lze  $F^r(1 - F)^s$  vyjádřit jako lineární kombinaci buď mocnin  $F$  nebo mocnin  $(1 - F)$ . Proto je přirozené popsat rozdělení buď momenty typu  $M_{1,r,0}$ ,  $r = 0, 1, 2, \dots$ , nebo typu  $M_{1,0,s}$ ,  $s = 0, 1, 2, \dots$

Výhodou PVM odhadů je skutečnost, že jsou výpočetně nenáročné a hodnoty odhadovaných parametrů jsou vždy přípustné. Pro malý rozsah výběru se často chovají dokonce lépe než odhady získané metodou maximální věrohodnosti (viz [54]). Dnešní popularitu PVM odhadů lze přičíst na vrub jejich jednoduchosti, praktičnosti a dobrým vlastnostem pro malé rozsahy výběru.

V dalších odstavcích bude pozornost věnována odhadu parametrů GEV rozdělení a zobecněného Paretova rozdělení metodou PVM. Dále  $x_{(j)}$  bude značit  $j$ -tý prvek v uspořádané realizaci výběru  $x_1, \dots, x_n$ .

### 2.2.1 GEV rozdělení

Nechť  $X_1, \dots, X_n$  je náhodný výběr z GEV rozdělení s distribuční funkcí

$$F(x, \mu, \sigma, \gamma) = \exp \left\{ - \left[ 1 + \gamma \left( \frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\gamma}} \right\}$$

pro  $x \in \{x : 1 + \gamma \left( \frac{x - \mu}{\sigma} \right) > 0\}$ . Dále budou uvažovány momenty typu  $\beta_r = M_{1,r,0} = E \{X [F(X)]^r\}$ ,  $r = 0, 1, 2, \dots$ . Průběžným užitím substituce  $-\ln(F) = t$  a  $(r+1)t = u$  lze dostat momenty ve tvaru

$$\begin{aligned} \beta_r &= \int_0^1 F^{\leftarrow} F^r dF = \int_0^1 \left\{ \mu + \frac{\sigma}{\gamma} [(-\ln(F))^{-\gamma} - 1] \right\} F^r dF \\ &= \int_0^\infty \left[ \mu + \frac{\sigma}{\gamma} (t^{-\gamma} - 1) \right] e^{-t(r+1)} dt = \int_0^\infty \left[ \mu + \frac{\sigma}{\gamma} t^{-\gamma} - \frac{\sigma}{\gamma} \right] e^{-t(r+1)} dt \\ &= \left( \mu - \frac{\sigma}{\gamma} \right) \int_0^\infty e^{-t(r+1)} dt + \frac{\sigma}{\gamma} \int_0^\infty t^{-\gamma} e^{-t(r+1)} dt \\ &= \left( \mu - \frac{\sigma}{\gamma} \right) (r+1)^{-1} + \frac{\sigma}{\gamma} (r+1)^{\gamma-1} \int_0^\infty u^{-\gamma} e^{-u} du \\ &= \left( \mu - \frac{\sigma}{\gamma} \right) (r+1)^{-1} + \frac{\sigma}{\gamma} (r+1)^{\gamma-1} \Gamma(1-\gamma) \\ &= \frac{1}{r+1} \left\{ \mu - \frac{\sigma}{\gamma} [1 - (r+1)^\gamma \Gamma(1-\gamma)] \right\}, \quad \gamma < 1, \quad \gamma \neq 0, \end{aligned} \tag{2.13}$$

kde  $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$  je dobře známá gama funkce. Dále budou odvozeny rovnice pro stanovení odhadů parametrů  $\mu, \sigma, \gamma$ . Ze vztahu (2.13) plyne pro  $r = 0$

$$\beta_0 = \mu - \frac{\sigma}{\gamma} [1 - \Gamma(1-\gamma)],$$

pro  $r = 1$  lze dostat

$$\begin{aligned} \beta_1 &= \frac{1}{2} \left\{ \mu - \frac{\sigma}{\gamma} [1 - 2^\gamma \Gamma(1-\gamma)] \right\}, \\ 2\beta_1 &= \mu - \frac{\sigma}{\gamma} [1 - \Gamma(1-\gamma) + \Gamma(1-\gamma) - 2^\gamma \Gamma(1-\gamma)], \\ 2\beta_1 &= \mu - \frac{\sigma}{\gamma} [1 - \Gamma(1-\gamma)] + \frac{\sigma}{\gamma} \Gamma(1-\gamma)(2^\gamma - 1), \\ 2\beta_1 - \beta_0 &= \frac{\sigma}{\gamma} \Gamma(1-\gamma)(2^\gamma - 1), \end{aligned}$$

a pro  $r = 2$  lze užitím vztahu  $(\sigma/\gamma)\Gamma(1-\gamma) = (2\beta_1 - \beta_0)/(2^\gamma - 1)$  vyjádřit

$$\begin{aligned} \beta_2 &= \frac{1}{3} \left\{ \mu - \frac{\sigma}{\gamma} [1 - 3^\gamma \Gamma(1-\gamma)] \right\}, \\ 3\beta_2 &= \mu - \frac{\sigma}{\gamma} [1 - \Gamma(1-\gamma) + \Gamma(1-\gamma) - 3^\gamma \Gamma(1-\gamma)], \\ 3\beta_2 &= \mu - \frac{\sigma}{\gamma} [1 - \Gamma(1-\gamma)] + \frac{\sigma}{\gamma} \Gamma(1-\gamma)(3^\gamma - 1), \end{aligned}$$

$$\begin{aligned}
3\beta_2 - \beta_0 &= \frac{\sigma}{\gamma} \Gamma(1 - \gamma)(3^\gamma - 1), \\
3\beta_2 - \beta_0 &= \frac{2\beta_1 - \beta_0}{2^\gamma - 1} (3^\gamma - 1), \\
\frac{3\beta_2 - \beta_0}{2\beta_1 - \beta_0} &= \frac{3^\gamma - 1}{2^\gamma - 1}.
\end{aligned}$$

Celkově tedy lze PVM odhady  $\hat{\mu}$ ,  $\hat{\sigma}$ ,  $\hat{\gamma}$  parametrů  $\mu$ ,  $\sigma$ ,  $\gamma$  získat řešením soustavy rovnic

$$\begin{aligned}
\beta_0 &= \mu - \frac{\sigma}{\gamma} [1 - \Gamma(1 - \gamma)], \\
2\beta_1 - \beta_0 &= \frac{\sigma}{\gamma} \Gamma(1 - \gamma)(2^\gamma - 1), \\
\frac{3\beta_2 - \beta_0}{2\beta_1 - \beta_0} &= \frac{3^\gamma - 1}{2^\gamma - 1},
\end{aligned}$$

přičemž  $\beta_r$  se nahradí odhadem

$$\hat{\beta}_r = \frac{1}{n} \sum_{j=1}^n \left( \prod_{l=1}^r \frac{j-l}{n-l} \right) x_{(j)}$$

který je nestranným odhadem parametru  $\beta_r$  (viz [61]). Odhady  $\hat{\mu}$ ,  $\hat{\sigma}$ ,  $\hat{\gamma}$  splňují podmínky přípustnosti ( $\gamma < 1$ ,  $\sigma > 0$ ) skoro jistě a jsou asymptoticky normální, stejně jako  $\hat{\beta}_0$ ,  $\hat{\beta}_1$ ,  $\hat{\beta}_2$ , pro  $\gamma < 0,5$  (viz [54]).

### 2.2.2 Zobecněné Paretovo rozdělení

Nechť  $X_1, \dots, X_n$  je náhodný výběr ze zobecněného Paretova rozdělení s distribuční funkcí

$$F(x) = 1 - \left( 1 + \gamma \frac{x}{\sigma} \right)^{-\frac{1}{\gamma}}$$

pro  $x > 0$ . V případě zobecněného Paretova rozdělení je příhodné uvažovat momenty typu  $\alpha_s = M_{1,0,s} = E \{ X [1 - F(X)]^s \}$ ,  $s = 0, 1, 2, \dots$ . Užitím substituce  $1 - F = t$  lze dostat

$$\begin{aligned}
\alpha_s &= \int_0^1 F^{\leftarrow}(1 - F)^s dF = \int_0^1 \frac{\sigma}{\gamma} [(1 - F)^{-\gamma} - 1] (1 - F)^s dF \\
&= \frac{\sigma}{\gamma} \int_0^1 (t^{-\gamma} - 1)t^s dt = \frac{\sigma}{\gamma} \int_0^1 (t^{-\gamma+s} - t^s) dt = \frac{\sigma}{\gamma} \left( \frac{1}{s+1-\gamma} - \frac{1}{s+1} \right) \\
&= \frac{\sigma}{(s+1)(s+1-\gamma)}, \quad \gamma < 1.
\end{aligned}$$

Pak PVM odhady  $\hat{\sigma}$ ,  $\hat{\gamma}$  parametrů  $\sigma$ ,  $\gamma$  je možné získat řešením soustavy rovnic

$$\begin{aligned}
\alpha_0 &= \frac{\sigma}{1-\gamma}, \\
\alpha_1 &= \frac{\sigma}{2(2-\gamma)}.
\end{aligned}$$

Po jednoduché úpravě lze dostat

$$\begin{aligned}
\sigma &= \frac{2\alpha_0\alpha_1}{\alpha_0 - 2\alpha_1}, \\
\gamma &= \frac{\alpha_0 - 4\alpha_1}{\alpha_0 - 2\alpha_1},
\end{aligned} \tag{2.14}$$

přičemž  $\alpha_s$  se nahradí odhadem

$$\hat{\alpha}_s = \frac{1}{n} \sum_{j=1}^n \left( \prod_{l=1}^r \frac{n-j-l+1}{n-l} \right) x_{(j)},$$

který je nestranným odhadem parametru  $\alpha_s$  (viz [61]). Veličinu  $\alpha_s$  ve vztahu (2.14) lze nahradit také odhadem

$$\tilde{\alpha}_s = \frac{1}{n} \sum_{j=1}^n \left( 1 - \frac{j}{n} \right)^s x_{(j)},$$

který je sice pouze konzistentní, ale při provedených simulacích vykazoval lepší výsledky. Odhady  $\hat{\sigma}$ ,  $\hat{\gamma}$  a  $\hat{\alpha}_0$ ,  $\hat{\alpha}_1$  jsou pro  $\gamma < 0,5$  asymptoticky normální (viz [53]). Toto omezení však není podstatné při aplikacích, kterými se budeme zabývat v této práci, protože při zpracování hydrologických dat (viz kapitola 4) se hodnota parametru  $\gamma$  obvykle pohybuje v mezích  $-0,5 < \gamma < 0,5$ .



## 3 Testování hypotéz o parametrech rozdělení extrémních hodnot

V této kapitole bude pozornost věnována testování hypotéz o parametrech rozdělení extrémních hodnot. Nejprve bude shrnuto několik základních poznatků o testech dobré shody, které budou využity při posuzování adekvátnosti pravděpodobnostních modelů. Dále budou popsány asymptotické testy vycházející z teorie MV odhadů včetně testů s rušivými parametry, které lze využít například v situaci, kdy je nutné posoudit vhodnost nahrazení navrženého modelu jednodušším submodelem.

### 3.1 Testy dobré shody

Je-li nutné posoudit, jak dobře zvolený statistický model vystihuje určitá data, lze k tomu použít testy dobré shody, které jistým způsobem popisují rozdíl mezi empirickým rozdělením datového souboru a zvoleným teoretickým rozdělením pravděpodobností. Typické využití testů dobré shody je při zjišťování, zda data pochází z předem specifikovaného rozdělení. Stejně tomu bude i v této práci. Níže uvedené testy dobré shody (Pearsonův  $\chi^2$  test, Kolmogorovův-Smirnovův test, Andersonův-Darlingův test) jsou v oblasti matematické statistiky natolik známé, že nebudou probírány dopodrobna a budou shrnuty pouze některé základní poznatky, které budou využity v rámci této práce. Detailní informace o uvedených testech lze najít například v monografii [66]. Všechny zde uvedené testy jsou natolik rozšířené, že je lze najít ve většině běžně používaných statistických softwarů.

Dále  $X_1, \dots, X_n$  bude značit posloupnost nezávislých a stejně rozdělených náhodných veličin z rozdělení s distribuční funkcí  $F$ .

#### 3.1.1 $\chi^2$ test

Pearsonův  $\chi^2$  test je jedním z nejznámějších testů pro posouzení shody mezi pozorovanými daty a navrženým statistickým modelem. Jelikož je test aplikován na data rozdělená do tříd, lze jej použít pro diskrétní i spojitá rozdělení. Pearsonův  $\chi^2$  test je založen na porovnání pozorovaných četností v jednotlivých třídách s očekávanými četnostmi odvozenými z testovaného rozdělení. Nechť je tedy náhodný výběr rozdělen do  $k$  tříd s tím, že pravděpodobnost, že člen výběru padne do třídy  $i$ ,  $i = 1, \dots, k$ , je  $p_i$ . Přitom platí, že  $\sum_{i=1}^k p_i = 1$ . Nechť  $Y_i$ ,  $i = 1, \dots, k$ , jsou četnosti v jednotlivých třídách. Pak sdružené rozdělení  $(Y_1, \dots, Y_k)$  je multinomické rozdělení s pravděpodobnostní funkcí

$$P(Y_1 = y_1, \dots, Y_k = y_k) = \frac{n!}{y_1! \cdots y_k!} p_1^{y_1} \cdots p_k^{y_k},$$

přičemž  $\sum_{i=1}^k y_i = n$ ,  $y_i \in \{0, 1, \dots, n\}$ ,  $i = 1, \dots, k$ . Nulová hypotéza je tvaru  $H_0 : p_i = \pi_i$ ,  $i = 1, \dots, k$ , proti alternativě  $H_1 : p_i \neq \pi_i$  pro nějaké  $i$ .

Testovací statistika  $\chi^2$  testu je tvaru

$$\chi_n^2 = \sum_{i=1}^k \frac{(Y_i - n\pi_i)^2}{n\pi_i} \quad (3.1)$$

a má asymptoticky  $\chi^2$  rozdělení s  $k - 1$  stupni volnosti. Nulová hypotéza se zamítá na hladině významnosti  $\alpha$  v případě, že hodnota testovací statistiky (3.1) přesáhne kritickou hodnotu  $\chi^2$  rozdělení s příslušným stupněm volnosti. Je třeba vzít do úvahy, že  $\chi^2$  test je asymptotický, a proto rozsah výběru  $n$  musí být dostatečný, stejně jako četnosti pozorování v jednotlivých třídách.

Často však nastává situace, kdy  $\pi_i = \pi_i(\boldsymbol{\theta})$ ,  $i = 1, \dots, k$ , závisí na neznámém parametru  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)$ . V takovém případě je třeba tento parametr odhadnout, například metodou  $\chi^2$  minima nebo modifikovanou metodou  $\chi^2$  minima (viz např. [5]). Testovací statistika (3.1) má pak asymptoticky  $\chi^2$  rozdělení s  $k - m - 1$  stupni volnosti.

### 3.1.2 Kolmogorovův-Smirnovův test

Kolmogorovův-Smirnovův (K-S) test patří do skupiny neparametrických testů založených na empirické distribuční funkci. Empirická distribuční funkce je definována vztahem

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq x\}},$$

přičemž  $1_{\{A\}}$  je indikátor jevu  $A$ . Dá se také ukázat (viz [5]), že funkce  $F_n(x)$  je pro všechna  $x \in \mathbb{R}$  nestranným a konzistentním odhadem distribuční funkce  $F(x)$ . Nulová hypotéza je tvaru  $H_0 : F = F_0$  proti alternativě  $H_1 : F \neq F_0$ , kde  $F_0$  je předepsaná distribuční funkce.

Testovací statistika K-S testu je tvaru

$$T_n = \sqrt{n} \sup_{x \in \mathbb{R}} |F_n(x) - F_0(x)| \quad (3.2)$$

a má asymptoticky Kolmogorovo rozdělení s distribuční funkcí

$$K(x) = 1 - 2 \sum_{i=1}^{\infty} (-1)^{i+1} e^{-2i^2 x^2}.$$

Je-li rozsah výběru malý, porovná se testovací statistika (3.2) s kritickými hodnotami K-S testu, které lze najít v tabulkách (viz např. [68]). Pro větší rozsah výběru platí, že je-li  $K(T_n) \geq 1 - \alpha$ , nulová hypotéza se zamítá na hladině významnosti, která se s rostoucím rozsahem výběru blíží číslu  $\alpha$ .

V případě, že funkce  $F$  obsahuje neznáme parametry, je třeba je odhadnout a také modifikovat kritické hodnoty K-S testu. Modifikované kritické hodnoty K-S testu lze získat pomocí simulací, jako tomu bylo například v případě Lillieforsova testu (viz např. [69]), který je modifikací K-S testu pro testování, zda data pochází z normálního rozdělení s neznámými parametry  $\mu$  a  $\sigma^2$ . Většina dostupných softwarů však K-S test obsahuje a řeší situaci modifikace kritických hodnot za uživatele.

### 3.1.3 Andersonův-Darlingův test

Andersonův-Darlingův (A-D) test, stejně jako K-S test, patří do skupiny testů založených na empirické distribuční funkci. A-D test je modifikací K-S testu, přičemž klade větší důraz na chvosty testovaných rozdělení než K-S test. Nulová hypotéza je tvaru  $H_0 : F = F_0$  proti alternativě  $H_1 : F \neq F_0$ , kde  $F_0$  je předepsaná distribuční funkce.

A-D testovací statistika je tvaru

$$A_n^2 = n \int_{-\infty}^{\infty} \frac{[F_n(x) - F_0(x)]^2}{F_0(x)[1 - F_0(x)]} dF_0(x). \quad (3.3)$$

Jak bylo ukázáno v [4], vztah (3.3) lze pro výpočetní účely nahradit vztahem

$$A_n^2 = -n - \frac{1}{n} \sum_{i=1}^n (2i - 1) [\ln F_0(X_{(i)}) + \ln(1 - F_0(X_{(n-i+1)}))], \quad (3.4)$$

kde  $X_{(1)}, \dots, X_{(n)}$  označuje uspořádaný výběr  $X_1, \dots, X_n$ . Nulová hypotéza se zamítá na hladině významnosti  $\alpha$  v případě, že hodnota testovací statistiky (3.4) přesáhne příslušnou kritickou hodnotu. Kritické hodnoty A-D testu na rozdíl od K-S testu závisí na testovaném rozdělení, což na jednu stranu zvyšuje sílu testu, na druhou stranu je však nutné spočítat kritické hodnoty pro každé testované rozdělení.

V případě, že funkce  $F$  obsahuje neznáme parametry, je třeba je odhadnout a také modifikovat kritické hodnoty A-D testu. Kritické hodnoty pro vybraná rozdělení se známými i neznámými parametry lze najít například v [18, 55, 86].

## 3.2 Asymptotické testy

Nyní budou s využitím výsledků z teorie MV odhadů popsány testy založené na věrohodnostní funkci. Nulová hypotéza bude tvaru  $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$  proti alternativě  $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ . V následující větě budou zavedeny tři statistiky, které lze použít při testování hypotézy.

**Věta 3.1.** *Nechť jsou splněny předpoklady věty 2.11. Dále necht*

$$\begin{aligned} LM(\boldsymbol{\theta}_0) &= \frac{1}{n} [\mathbf{U}(\boldsymbol{\theta}_0)]^T [\mathbf{J}(\boldsymbol{\theta}_0)]^{-1} \mathbf{U}(\boldsymbol{\theta}_0), \\ W(\boldsymbol{\theta}_0) &= n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^T \mathbf{J}(\hat{\boldsymbol{\theta}}_n)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0), \\ LR(\boldsymbol{\theta}_0) &= 2 \left[ l(\hat{\boldsymbol{\theta}}_n) - l(\boldsymbol{\theta}_0) \right], \end{aligned}$$

kde  $\mathbf{U}(\boldsymbol{\theta})$  je skórový vektor definován vztahem (2.5). Jsou-li všechny prvky matice  $\mathbf{J}(\boldsymbol{\theta})$  spojité v bodě  $\boldsymbol{\theta}_0$ , pak  $LM(\boldsymbol{\theta}_0)$ ,  $W(\boldsymbol{\theta}_0)$ ,  $LR(\boldsymbol{\theta}_0)$  mají asymptoticky  $\chi_m^2$  rozdělení.

*Důkaz.* Důkaz přímo nesouvisí s tématem této práce a lze jej najít například v knize [5].

□

**Poznámka 3.2.** Za předpokladu spojitosti všech prvků matice  $\mathbf{J}(\boldsymbol{\theta})$  v bodě  $\boldsymbol{\theta}_0$  zůstává tvrzení o asymptotickém rozdělení statistiky  $LM(\boldsymbol{\theta}_0)$  v platnosti, i když se místo matice  $\mathbf{J}(\boldsymbol{\theta}_0)$  použije nějaký její konzistentní odhad. Také tvrzení o asymptotickém rozdělení statistiky  $W(\boldsymbol{\theta}_0)$  zůstává v platnosti, i když se místo matice  $\mathbf{J}(\hat{\boldsymbol{\theta}}_n)$  použije nějaký její konzistentní odhad.

Test založený na statistice  $LM(\boldsymbol{\theta}_0)$  se nazývá skórový test (někdy též Raův test nebo test založený na Lagrangeových multiplikatorech), test založený na statistice  $W(\boldsymbol{\theta}_0)$  se nazývá Waldův test a test užívající statistiku  $LR(\boldsymbol{\theta}_0)$  se nazývá test poměrem věrohodností (likelihood ratio test). Všechny tři testy zamítají nulovou hypotézu na hladině významnosti  $\alpha$ , pokud příslušné testovací statistiky přesáhnou kritickou hodnotu  $\chi_m^2(\alpha)$ .

Výhodou skórového testu je, že ve statistice  $LM(\boldsymbol{\theta}_0)$  nevystupuje maximálně věrohodný odhad  $\hat{\boldsymbol{\theta}}_n$ . Tím se liší od Waldova testu, který znalost maximálně věrohodného odhadu  $\hat{\boldsymbol{\theta}}_n$  vyžaduje. Test založený na věrohodnostním poměru na rozdíl od dvou předchozích testů nevyžaduje znalost FIM  $\mathbf{J}$ , jejíž stanovení může být v některých situacích velmi obtížné.

### 3.2.1 Testy s rušivými parametry

Dále necht' je  $m$ -rozměrný parametr  $\boldsymbol{\theta}$  rozdělen na dvě složky, tedy  $\boldsymbol{\theta} = (\boldsymbol{\tau}^T, \boldsymbol{\psi}^T)^T$ , kde

$$\boldsymbol{\tau} = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_k \end{pmatrix}, \quad \boldsymbol{\psi} = \begin{pmatrix} \theta_{k+1} \\ \vdots \\ \theta_m \end{pmatrix}, \quad 1 \leq k < m.$$

Nulová hypotéza je tvaru  $H_0^* : \boldsymbol{\tau} = \boldsymbol{\tau}_0$  proti alternativě  $H_1^* : \boldsymbol{\tau} \neq \boldsymbol{\tau}_0$ . Pozornost je tedy soustředěna pouze na parametr  $\boldsymbol{\tau}$  a parametr  $\boldsymbol{\psi}$  je sice potřebný k popisu pravděpodobnostního modelu, ale hypotéza  $H_0^*$  se jej netýká. Proto se  $\boldsymbol{\tau}$  nazývá cílový parametr a  $\boldsymbol{\psi}$  rušivý parametr. Dále  $\mathbf{U}(\boldsymbol{\theta})$  je skórový vektor definován vztahem (2.5) a

$$\mathbf{U}_1(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial l(\boldsymbol{\theta})}{\partial \theta_1} \\ \vdots \\ \frac{\partial l(\boldsymbol{\theta})}{\partial \theta_k} \end{pmatrix}, \quad \mathbf{U}_2(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial l(\boldsymbol{\theta})}{\partial \theta_{k+1}} \\ \vdots \\ \frac{\partial l(\boldsymbol{\theta})}{\partial \theta_m} \end{pmatrix}, \quad \mathbf{J}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{J}_{11}(\boldsymbol{\theta}) & \mathbf{J}_{12}(\boldsymbol{\theta}) \\ \mathbf{J}_{21}(\boldsymbol{\theta}) & \mathbf{J}_{22}(\boldsymbol{\theta}) \end{pmatrix},$$

kde  $\mathbf{J}_{11}(\boldsymbol{\theta})$  je matice typu  $k \times k$ . Dále bude třeba užít vzorce pro výpočet inverzní matice rozdělené na bloky, které jsou shrnuty v následujícím lemmatu.

**Lemma 3.3.** *Necht'*

$$\mathbf{J} = \begin{pmatrix} \mathbf{J}_{11} & \mathbf{J}_{12} \\ \mathbf{J}_{21} & \mathbf{J}_{22} \end{pmatrix}$$

je regulární matice, přičemž bloky  $\mathbf{J}_{11}$  a  $\mathbf{J}_{22}$  jsou čtvercové a regulární. Položme

$$\begin{aligned}\mathbf{J}_{11.2} &= \mathbf{J}_{11} - \mathbf{J}_{12}\mathbf{J}_{22}^{-1}\mathbf{J}_{21}, & \mathbf{J}^{11} &= \mathbf{J}_{11.2}^{-1}, & \mathbf{J}^{12} &= -\mathbf{J}_{11.2}^{-1}\mathbf{J}_{12}\mathbf{J}_{22}^{-1}, \\ \mathbf{J}_{22.1} &= \mathbf{J}_{22} - \mathbf{J}_{21}\mathbf{J}_{11}^{-1}\mathbf{J}_{12}, & \mathbf{J}^{22} &= \mathbf{J}_{22.1}^{-1}, & \mathbf{J}^{21} &= -\mathbf{J}_{22.1}^{-1}\mathbf{J}_{21}\mathbf{J}_{11}^{-1}.\end{aligned}$$

Pak

$$\mathbf{J}^{-1} = \begin{pmatrix} \mathbf{J}^{11} & \mathbf{J}^{12} \\ \mathbf{J}^{21} & \mathbf{J}^{22} \end{pmatrix}.$$

*Důkaz.* Lemma lze dokázat ověřením faktu, že součin matice  $\mathbf{J}$  a matice  $\mathbf{J}^{-1}$  dává jednotkovou matici. □

Při určování MV odhadu parametru  $\boldsymbol{\theta}$  je třeba rozlišit dvě situace. V prvním případě se jedná o MV odhad, který není vázán žádnými dalšími podmínkami. Takový odhad bude označen  $\hat{\boldsymbol{\theta}}_n$  a je tvaru

$$\hat{\boldsymbol{\theta}}_n = \begin{pmatrix} \hat{\boldsymbol{\tau}}_n \\ \hat{\boldsymbol{\psi}}_n \end{pmatrix}.$$

V druhém případě je třeba získat MV odhad parametru  $\boldsymbol{\theta}$  za podmínky, že platí  $\mathbf{H}_0^* : \boldsymbol{\tau} = \boldsymbol{\tau}_0$ . Věrohodnostní funkce se tedy maximalizuje pouze vzhledem k parametru  $\boldsymbol{\psi}$  a takový odhad bude označen  $\tilde{\boldsymbol{\psi}}_n$ . Příslušný MV odhad parametru  $\boldsymbol{\theta}$  je pak tvaru

$$\tilde{\boldsymbol{\theta}}_n = \begin{pmatrix} \boldsymbol{\tau}_0 \\ \tilde{\boldsymbol{\psi}}_n \end{pmatrix}.$$

Je-li skutečná hodnota parametru  $\boldsymbol{\theta}$  rovna  $\boldsymbol{\theta}_0 = (\boldsymbol{\tau}_0^T, \boldsymbol{\psi}_0^T)^T$ , pak za platnosti předpokladů věty 2.11 platí pro  $\tilde{\boldsymbol{\psi}}$  a veličiny s tímto odhadem související

$$\begin{aligned}\frac{1}{\sqrt{n}}\mathbf{U}_2(\boldsymbol{\theta}_0) &\xrightarrow{d} \mathbf{N}[\mathbf{0}; \mathbf{J}_{22}(\boldsymbol{\theta}_0)], \\ \sqrt{n}(\tilde{\boldsymbol{\psi}}_n - \boldsymbol{\psi}_0) &\xrightarrow{d} \mathbf{N}[\mathbf{0}; \mathbf{J}_{22}^{-1}(\boldsymbol{\theta}_0)].\end{aligned}$$

Dále nechť

$$\begin{aligned}LM^* &= \frac{1}{n} \left[ \mathbf{U}_1(\tilde{\boldsymbol{\theta}}_n) \right]^T \left[ \mathbf{J}_{11.2}(\tilde{\boldsymbol{\theta}}_n) \right]^{-1} \mathbf{U}_1(\tilde{\boldsymbol{\theta}}_n), \\ W^* &= n(\hat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}_0)^T \mathbf{J}_{11.2}(\hat{\boldsymbol{\theta}}_n)(\hat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}_0), \\ LR^* &= 2 \left[ l(\hat{\boldsymbol{\theta}}_n) - l(\tilde{\boldsymbol{\theta}}_n) \right].\end{aligned}$$

**Věta 3.4.** *Nechť jsou splněny předpoklady věty 2.11 a nechť jsou všechny prvky matice  $\mathbf{J}(\boldsymbol{\theta})$  spojité v bodě  $\boldsymbol{\theta}_0$ . Jestliže  $n \rightarrow \infty$ , pak za platnosti  $\mathbf{H}_0^*$  platí*

$$LM^* \xrightarrow{d} \chi_k^2, \quad W^* \xrightarrow{d} \chi_k^2, \quad LR^* \xrightarrow{d} \chi_k^2.$$

*Důkaz.* Důkaz přímo nesouvisí s tématem této práce a lze jej najít například v knize [5].

□

**Poznámka 3.5.** Za předpokladu spojitosti všech prvků matice  $\mathbf{J}_{11.2}(\boldsymbol{\theta})$  v bodě  $\boldsymbol{\theta}_0$  zůstává tvrzení o asymptotickém rozdělení statistiky  $LM^*$  v platnosti i v případě, že se matice  $\mathbf{J}_{11.2}(\tilde{\boldsymbol{\theta}}_n)$  nahradí nějakou maticí  $\hat{\mathbf{J}}_{11.2}(\tilde{\boldsymbol{\theta}}_n)$ , která je konzistentním odhadem matice  $\mathbf{J}_{11.2}(\boldsymbol{\theta}_0)$ . Stejně tak tvrzení o asymptotickém rozdělení statistiky  $W^*$  zůstává v platnosti i v případě, že se matice  $\mathbf{J}_{11.2}(\hat{\boldsymbol{\theta}}_n)$  nahradí nějakou maticí  $\hat{\mathbf{J}}_{11.2}(\hat{\boldsymbol{\theta}}_n)$ , která je konzistentním odhadem matice  $\mathbf{J}_{11.2}(\boldsymbol{\theta}_0)$ .

## 4 Statistická analýza dešťových srážek

V této kapitole budou teoretické poznatky odvozené v předchozích kapitolách aplikovány na reálných datech. Konkrétně se bude jednat o statistickou analýzu úhrnů dešťových srážek v brněnském regionu, které bylo motivováno potřebou aktualizace hydrologických podkladů pro tvorbu městského odvodnění. Nejprve budou popsány dvě vzorkovací techniky sloužící k výběru dat do statistického souboru, které se v hydrologii nejčastěji používají.

### 4.1 Vzorkovací techniky

Při posuzování hydrologických jevů je často aplikována teorie EV rozdělení, přičemž statistické soubory jsou tvořeny z měřených časových řad vyšetřované náhodné veličiny, v tomto případě srážkového úhrnu. Jak je popsáno například v [56], rozlišují se dva způsoby výběru prvků do statistického souboru hydrologických dat. První je založen na metodě blokových maxim, tedy výběru maximální dosažené hodnoty sledované veličiny v daném měrném období. V případě hydrologických aplikací se často jedná o období jednoho roku. Takové soubory jsou pak nazývány Annual Maximum Series (AMS). Volba délky měrného období (tj. velikosti bloku) jeden rok není náhodná. Pokud je zvolená velikost bloku příliš malá, může dojít k porušení asymptotických vlastností modelu, což vede k vychýlení odhadu. Naopak, je-li velikost bloku příliš velká, počet maxim bude malý, což způsobí značný rozptyl odhadu. Pragmatickou volbou se ukázala být právě velikost bloku rovna jednomu roku. Statistický model pro metodu AMS je založen na GEV rozdělení a je popsán v odstavci 1.4. Druhým způsobem je výběr všech hodnot překračujících určitou prahovou hodnotu. Statistické soubory jsou pak nazývány Peaks Over Threshold (POT). V hydrologii se však častěji užívá termínu Partial Duration Series (PDS). Statistický model pro metodu PDS je založen na zobecněném Paretově rozdělení a je popsán v odstavci 1.4. Řada autorů (viz např. [17, 71, 72, 88]) se také zabývala porovnáváním obou vzorkovacích technik.

V oblasti modelování hydrologických extrémů byla často používána metoda AMS (např. [3, 6, 96]). Výhodou metody AMS je, že jsou jasně stanovena pravidla pro výběr prvků do statistického souboru a je všeobecně považována za metodu, která splňuje podmínku nezávislosti prvků statistického souboru (viz [56]). Nevýhodou však je, že metoda nezohledňuje další extrémní hodnoty v daném roce, které jsou nižší než hodnota maximální. Jelikož lze získat pouze jednu hodnotu za rok, časové řady musí být dostatečně dlouhé.

V současnosti stále více využívanou vzorkovací technikou je metoda PDS, což dokládá řada prací ([10, 11, 13, 70, 71, 82, 95]). Tato metoda stojí na pevných teoretických základech (viz [90]), lépe fituje rozdělení s těžkými chvosty (viz [72]) a do statistického souboru je možné zahrnout více extrémních hodnot za rok, tedy nejen jednu maximální hodnotu. Tím se získá více informací o chování pravého chvostu rozdělení, než při použití pouze ročních maxim, což vede ke zpřesnění odhadů parametrů i kvantilů EV rozdělení. Nevýhodou metody je, že nejsou stanovena jasná pravidla pro výběr prvků do statistického souboru.

Například v [93] byly datové soubory pro metodu PDS sestaveny tak, že jednotlivé extrémy byly od sebe odděleny nižšími srážkovými úhrny nejméně po dobu 8–24 hodin v závislosti na délce trvání deště. V oblasti atmosférických srážek bývá PDS často obsažena v tzv. Event Maxima Series (EMS). Pak se postupuje tak, že v první fázi je dešťová řada rozdělena na jednotlivé dešťové události a následně jsou z těchto událostí vzorkovány jednotlivé délky trvání deště. Není však jednoznačně dáno, jak přesně definovat dešťovou událost. Jednou z možností definice je stanovení minimální délky bezdešťového období mezi jednotlivými událostmi.

Různí autoři k tomuto problému přistupují různým způsobem. Například v [47] stanovili minimální délku bezdešťového období rovnu jedné hodině, což je doba dotoku nejdelší dánské stokové sítě. V [95] oddělili jednotlivé události ustáním deště po dobu rovnou délce uvažovaného deště, minimálně však 12 hodin. V [70] oddělili dešťové události ustáním deště na nejméně jednu hodinu. V [62] vytvořili proměnlivé kritérium ustání deště na 1–6 dnů v závislosti na délce deště. V [11] oddělili dešťové události ustáním deště po dobu 24 hodin. Záleží tedy vždy na tom, za jakým účelem je dané vyhodnocení prováděno.

Z pohledu městského odvodnění se jeví jako vhodné postupovat podle [70]. Pokud se stanoví délka bezdešťového období jako součet doby dotoku stokovou sítí a doby do vyprázdnění retenčních prvků odvodňovacího systému, pak lze předpokládat, že následující dešťová událost nebude ovlivňovat sledované efekty (např. průtoky, přepady z odlehčovací komory) z události předchozí. To je podstatné zejména pro statistické vyhodnocení sledovaných veličin na síti.

Vzhledem k délce dostupných historických dešťových řad bude další odstavec věnován zpracování hydrologických dat pomocí technik založených na PDS.

## 4.2 Analýza srážek v brněnském regionu

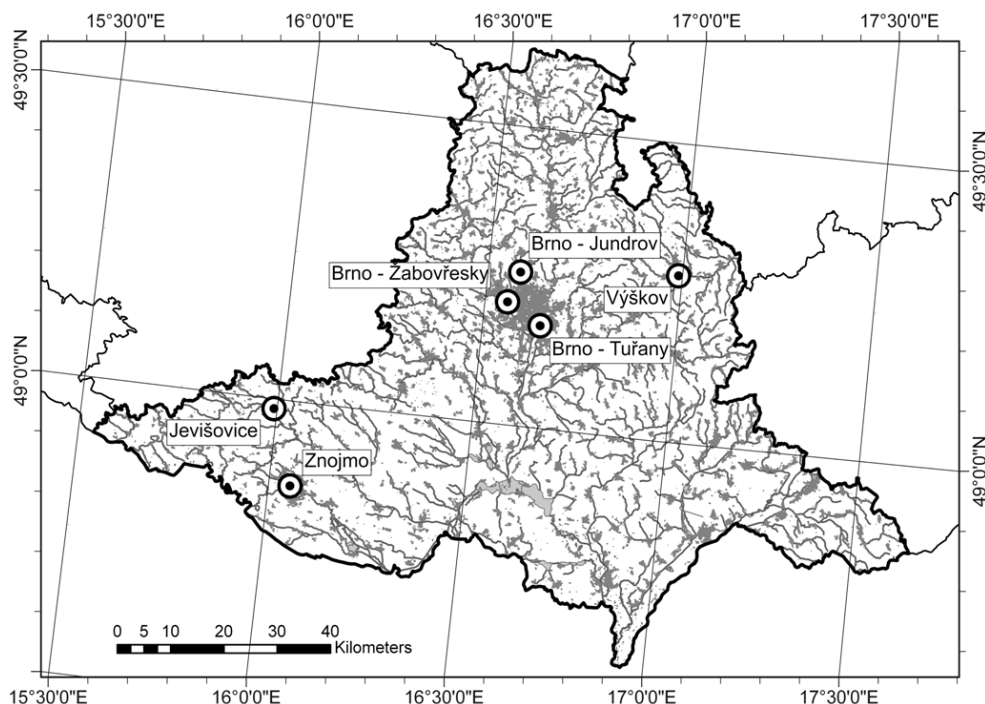
V České republice jsou údaje o srážkových intenzitách získávány z registračních pásek ombrografů (ombrograf = přístroj pro měření srážkových úhrnů) provozovaných Českým hydrometeorologickým ústavem. Pro potřeby moderního zpracování těchto údajů je prováděna digitalizace ombrogramů (ombrogram = grafický záznam údajů ombrografu) odečítáním souřadnic zlomových bodů ze záznamů na digitizéru. K opravě, doplnění a rekonstrukci záznamů se používá celá řada podkladů, zejména měsíční výkaz meteorologických pozorování na dané stanici (denní úhrny srážek ze srážkoměru a záznamy pozorovatele o časovém výskytu a intenzitě srážek), ombrogramy a měsíční výkazy z okolních stanic, radarová pozorování, synoptické mapy a údaje o srážkách ze synoptických a automatických stanic. Výsledný digitalizovaný záznam je ukládán do databáze s časovým rozlišením 1 minuty. V České republice jsou ombrografy běžně osazovány v období květen až září, kdy se objevuje většina přívalových srážek. Proto jsou analýzy extrémních srážek vycházející z těchto měření považovány za spolehlivé. Zaznamenané roky, které neobsahují kompletní měření ve výše uvedeném období, byly ze statistických analýz vyloučeny.



Stanice	Období sledování	Délka řady
Brno–Jundrov	1992–2003	11
Brno–Tuřany	1959–2000	41
Brno–Žabovřesky	1987–2003	16
Jevišovice	1961–2000	37
Vyškov	1961–1992	31
Znojmo–Kuchařovice	1956–2003	27

Tabulka 4.1: Přehled srážkoměrných stanic. Ve sloupci *Délka řady* je uveden počet let s úplnými záznamy.

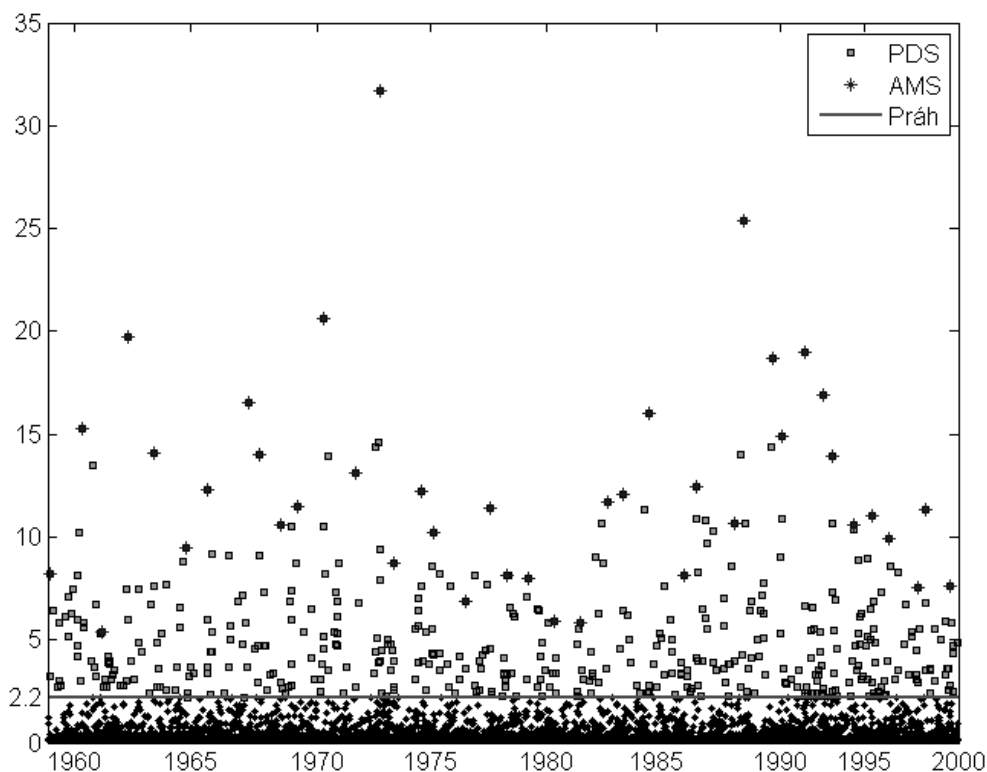
Jako vstupní data pro tuto analýzu slouží kontinuální řady reálných dešťů ve formě srážkových intenzit z 6 srážkoměrných stanic v Jihomoravském kraji (viz tabulka 4.1). Stanice Brno–Žabovřesky, Brno–Tuřany a Brno–Jundrov jsou umístěny na urbanizovaném povodí města Brna, které je centrem kraje a druhým největším městem v České republice. Stanice Vyškov je umístěna v severovýchodní části a stanice Jevišovice a Znojmo–Kuchařovice v jihozápadní části kraje (viz obrázek 4.1).



Obrázek 4.1: Rozmístění srážkoměrných stanic.

Ze sledovaných řad byly podle výše uvedených metodik (AMS a PDS) vyčleněny všechny deště, jejichž doby trvání byly postupně 5, 10, 15, 20, 30, 45, 60, 90, 120, 180, 240 a 360 minut. Tím byly získány soubory dat, kde ke každému roku byl přiřazen blok hodnot intenzit dešťů daného trvání. Příslušná intenzita deště je dána množstvím srážek,

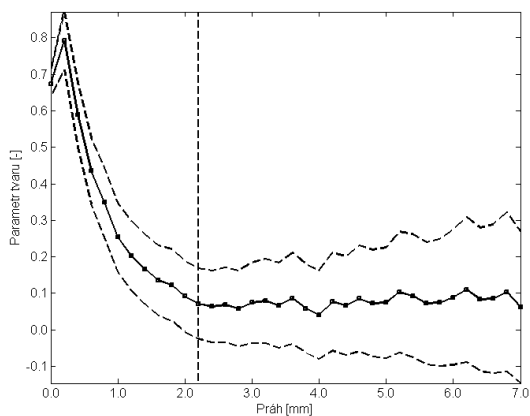
které dopadnou na jednotkovou plochu. Ilustraci rozdílu mezi metodami AMS a PDS co do složení datových souborů lze najít na obrázku 4.2, kde jsou vyneseny srážkové úhrny pro 15minutové deště ze stanice Brno–Tuřany.



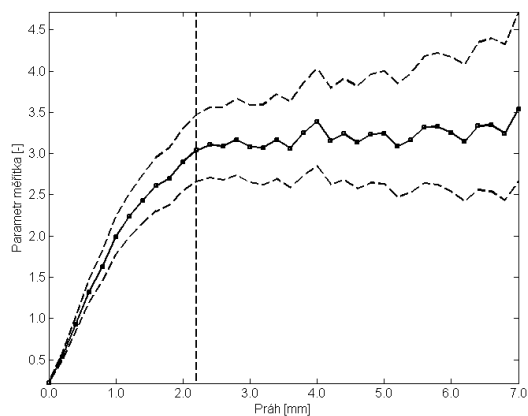
Obrázek 4.2: Složení datových souborů podle metodiky AMS a PDS pro stanici Brno–Tuřany a 15minutový déšť.

Vzhledem k malé délce některých časových řad bude k vyhodnocení srážkových dat použita metodika založená na PDS. Jak již bylo zmíněno (viz odstavec 1.4), nejdůležitějším faktorem u metody PDS je stanovení optimální prahové hodnoty  $u$ . K tomu lze využít metod popsanych v odstavci 1.4. Nejprve je třeba pro všechny uvedené doby trvání deště a různé prahové hodnoty určit MV odhady parametrů zobecněného Paretova rozdělení (viz odstavec 2.1.2). Vhodná prahová hodnota se pak určí pomocí MRL plotu a posouzením stability odhadů parametrů zobecněného Paretova rozdělení v závislosti na prahové hodnotě  $u$ . Výběr prahové hodnoty je demonstrován pro stanici Brno–Tuřany a 15minutový déšť (viz obrázek 4.3). MRL plot vykazuje přibližně lineární trend od hodnoty prahu  $u = 2,2$  mm (viz obrázek 4.3c). Stejně tak závislost parametru tvaru (viz obrázek 4.3a) a měřítka (viz obrázek 4.3b) na velikosti prahu  $u$  je přibližně lineární (respektive konstantní) od hodnoty prahu  $u = 2,2$  mm.

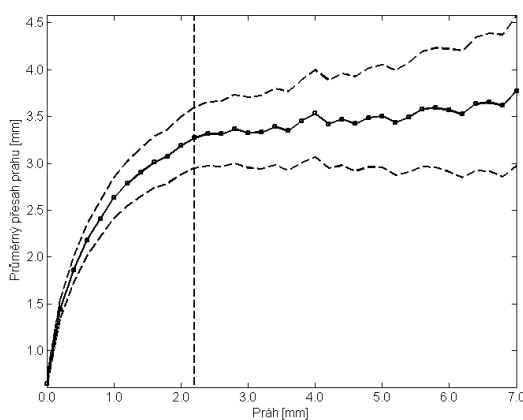
Jakmile jsou stanoveny prahové hodnoty pro všechny stanice a doby trvání deště, je třeba vyšetřit, zda lze hodnoty nad prahem popsat pomocí zobecněného Paretova rozdělení. K tomu se použijí testy dobré shody (viz odstavec 3.1), konkrétně Pearsonův  $\chi^2$  test, K-S test a A-D test, který se v hydrologické praxi zvláště doporučuje. Testy dobré shody lze do-



(a) Parametr tvaru.

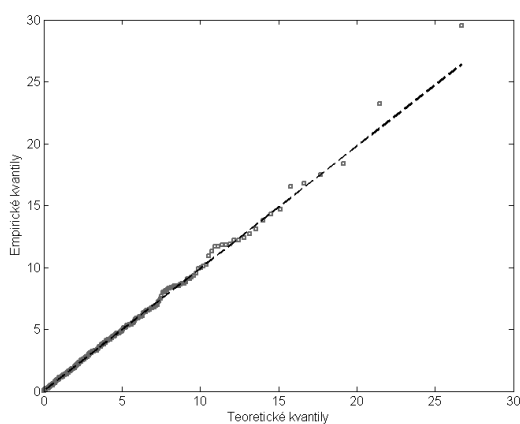


(b) Parametr měřítka.

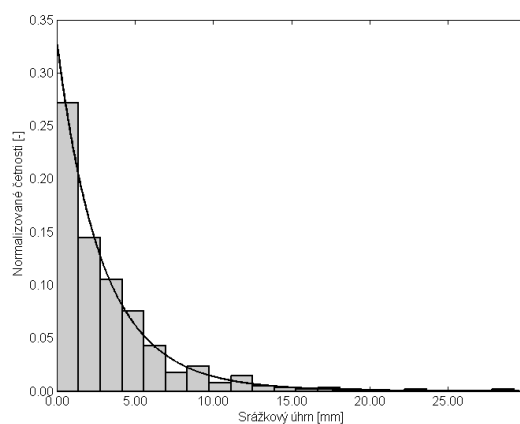


(c) MRL plot.

Obrázek 4.3: MRL plot a posouzení stability parametrů tvaru a měřítka pro stanici Brno–Tuřany, 15minutový déšť a práh  $u_0 = 2,2$  mm. Čárkovaná čára označuje 95% intervaly spolehlivosti pro parametr tvaru, parametr měřítka a pro průměrný přesah prahu.

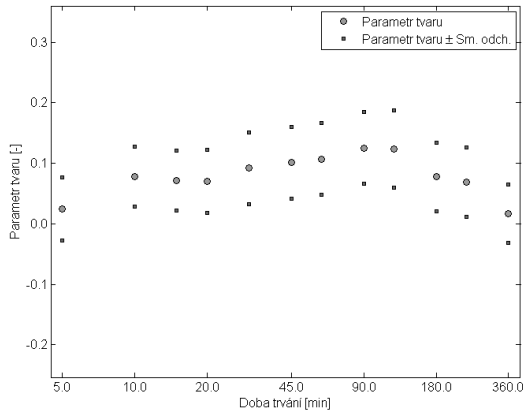


(a) Q-Q plot.

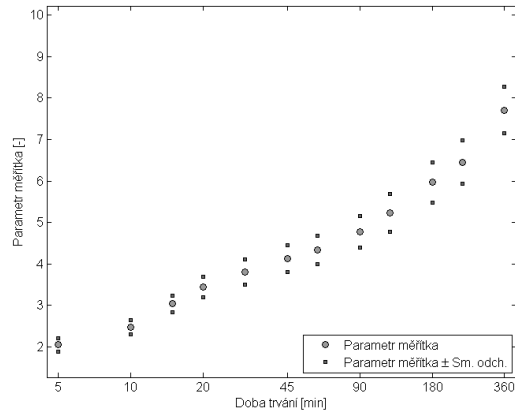


(b) Histogram.

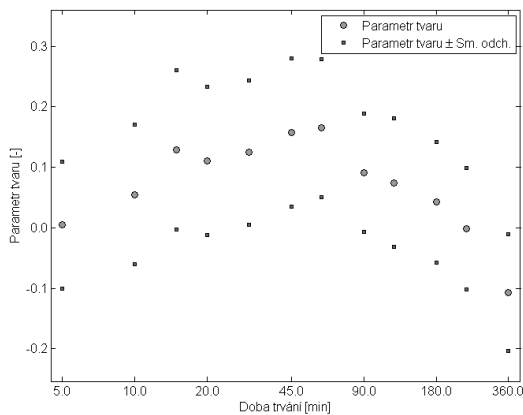
Obrázek 4.4: Q-Q plot a histogram proložený hustotou zobecněného Paretova rozdělení pro stanici Brno–Tuřany a 15minutový déšť.



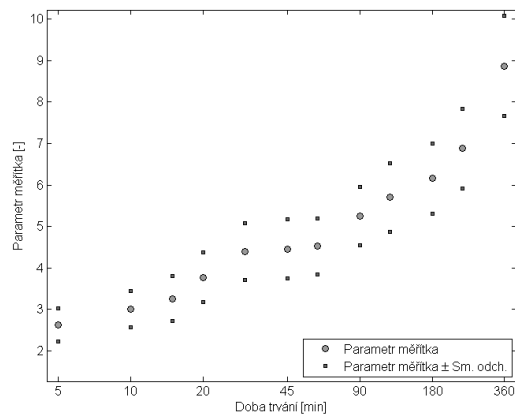
(a) Brno–Tuřany



(b) Brno–Tuřany



(c) Brno–Jundrov

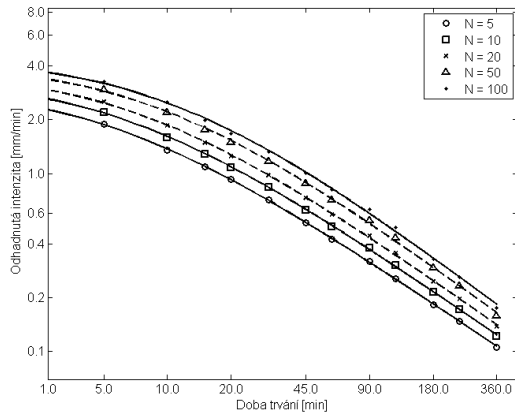


(d) Brno–Jundrov

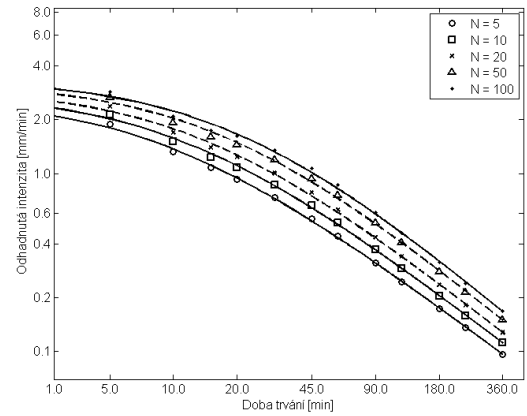
Obrázek 4.5: Odhady parametrů tvaru (vlevo) a měřítka (vpravo) včetně příslušných směrodatných odchylek s logaritmickým měřítkem na ose x.

plnit grafickou analýzou s využitím Q-Q plotů (viz obrázek 4.4a) a porovnáním histogramu datového souboru s hustotou zobecněného Paretova rozdělení (viz obrázek 4.4b). Ukázalo se, že shoda empirického a teoretického zobecněného Paretova rozdělení je velmi dobrá, protože všechny testy dobré shody nezamítají nulovou hypotézu o shodě se zobecněným Paretovým rozdělením na hladině významnosti 0,05. Kompletní tabulku výsledků A-D testu včetně hodnot testovacích statistik a kritických hodnot pro všechny stanice a doby trvání deště lze najít v práci [50]. Další analýza tedy vychází ze zobecněného Paretova rozdělení.

Na obrázku 4.5 jsou znázorněny odhady parametrů zobecněného Paretova rozdělení včetně směrodatných odchylek pro stanici Brno–Tuřany a Brno–Jundrov. Z těchto grafů je vidět, že variabilita parametrů tvaru a měřítka pro krátkou časovou řadu ze stanice Jundrov (11 let) je téměř o řád větší než pro nejdelší časovou řadu ze stanice Tuřany (41 let). Podobná skutečnost byla pozorována i pro ostatní časové řady. Informace o variabilitě jednotlivých parametrů budou dále použity k odhadu úrovně návratu a příslušných intervalů

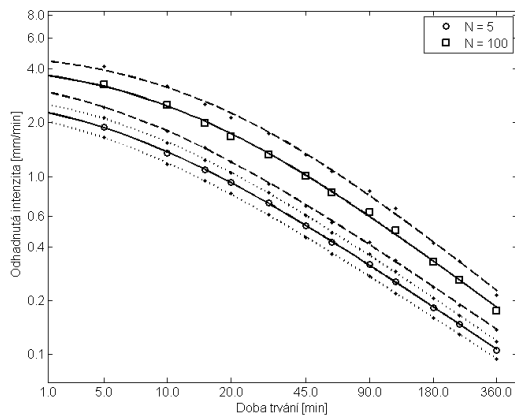


(a) Brno–Tuřany

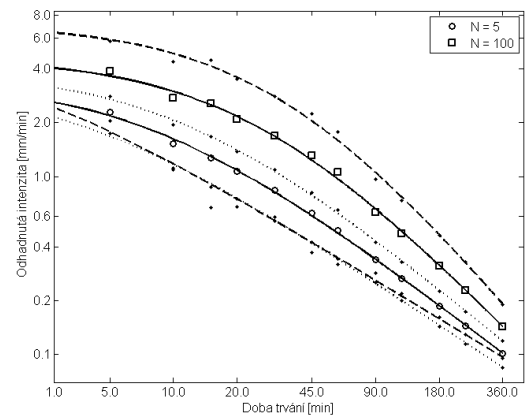


(b) Jevišovice

Obrázek 4.6: Odhad IDF křivek v logaritmických souřadnicích na obou osách.



(a) Brno–Tuřany



(b) Brno–Jundrov

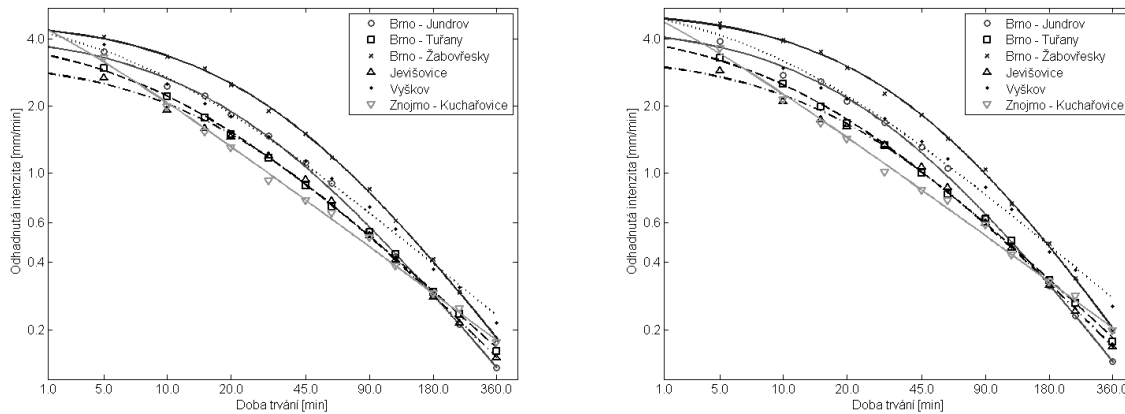
Obrázek 4.7: Odhad IDF křivek v logaritmických souřadnicích na obou osách. 95% intervaly spolehlivosti jsou vyznačeny tečkovaně ( $N = 5$ ) a čárkovaně ( $N = 100$ ).

spolehlivosti.

V hydrologické literatuře je zvykem znázorňovat odhady  $N$ -leté úrovně návratu v podobě tzv. IDF (Intensity-Duration-Frequency) křivek, kde intenzita je definována jako podíl úrovně návratu a doby trvání deště. Na obrázku 4.6 jsou pro stanice s nejdelšími časovými řadami (Tuřany, Jevišovice) znázorněny odhady IDF křivek, které byly získány proložení odhadnutých intenzit křivkou za použití nelineární regrese ve tvaru

$$I = \frac{\beta_1}{(1 + \beta_2 t)^{\beta_3}}$$

kde  $I$  označuje odhadnuté intenzity,  $t$  dobu trvání deště a  $\beta_1$ ,  $\beta_2$ ,  $\beta_3$  regresní parametry. Aby bylo možné detailněji rozeznat příslušné hodnoty intenzit dešťů různé doby trvání, byla v uvedených grafech na obou osách použita logaritmická stupnice.



(a)  $N = 50$  let

(b)  $N = 100$  let

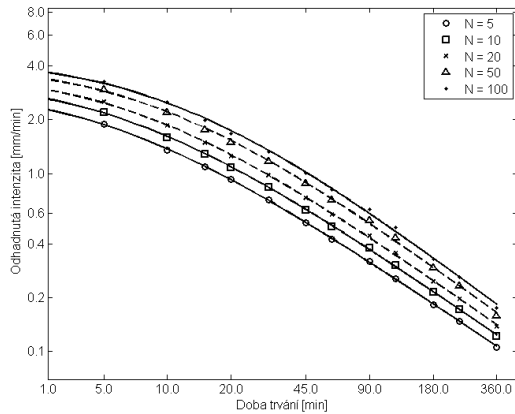
Obrázek 4.8: Srovnání odhadů IDF křivek pro všechny stanice v logaritmických souřadnicích na obou osách.

Na obrázku 4.7 jsou znázorněny IDF křivky pro vybrané stanice včetně 95% intervalů spolehlivosti odvozených delta metodou (viz odstavec 1.4). Z obrázků je patrné, že se snižující se délkou časové řady se snižuje spolehlivost modelu kvůli zvyšující se chybě odhadu.

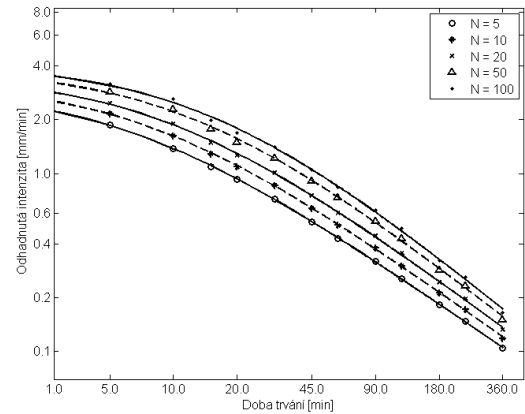
Na obrázku 4.8 jsou za účelem srovnání znázorněny odhady IDF křivek pro všechny stanice a  $N = 50$  let a  $N = 100$  let. Navzdory relativně malé ploše zkoumaného brněnského regionu lze z obrázku posoudit rozdíly mezi jednotlivými stanicemi a to nejen v intenzitě srážek, ale i v průběhu odhadnutých IDF křivek. Předně lze konstatovat, že IDF křivky pro všechny brněnské stanice (Jundrov, Tuřany, Žabovřesky), které jsou relativně blízko u sebe, vykazují přibližně stejné průběhy. Rozdíly v IDF křivkách mezi brněnskými a mimobrněnskými stanicemi jsou s největší pravděpodobností způsobeny odlišnými geografickými a klimatickými faktory v jednotlivých oblastech. Dále lze konstatovat, že IDF křivky pro  $N = 5, 10, 20$  let vykazují podobné průběhy.

Metodologie použitá výše k vyhodnocení IDF křivek byla založena na metodě maximální věrohodnosti pro zobecněné Paretovo rozdělení a asymptotických vlastnostech MV odhadů. Aby bylo možné posoudit asymptotické chování MV odhadů v závislosti na délce řady, provedou se nová vyhodnocení historických dešťových řad. Postupuje se stejným způsobem jako výše s tím, že k odhadu neznámých parametrů zobecněného Paretova rozdělení se použije metoda PVM (viz odstavec 2.2.2).

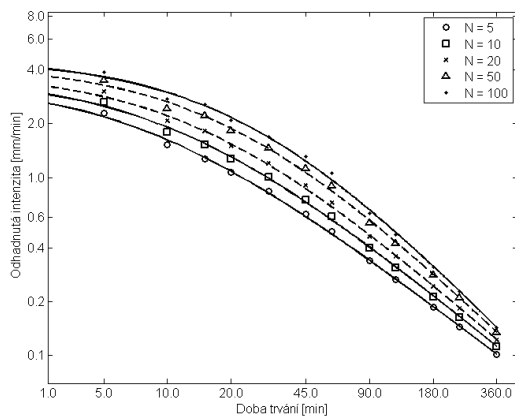
Nové vyhodnocení ukázalo, že takřka ve všech případech jsou odhady IDF křivek téměř identické při použití metody maximální věrohodnosti a metody PVM. Z hlediska praktického využití jsou rozdíly mezi křivkami víceméně zanedbatelné. Výjimku tvoří krátká dešťová řada ze stanice Brno–Jundrov, u které IDF křivky získané použitím metody PVM vykazují vyšší hodnoty intenzit než křivky získané metodou maximální věrohodnosti (viz obrázek 4.9).



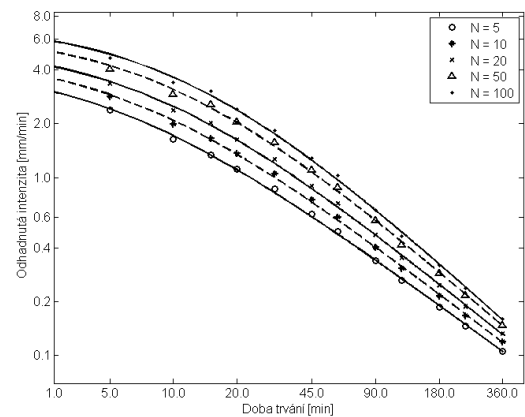
(a) Brno–Tuřany



(b) Brno–Tuřany



(c) Brno–Jundrov



(d) Brno–Jundrov

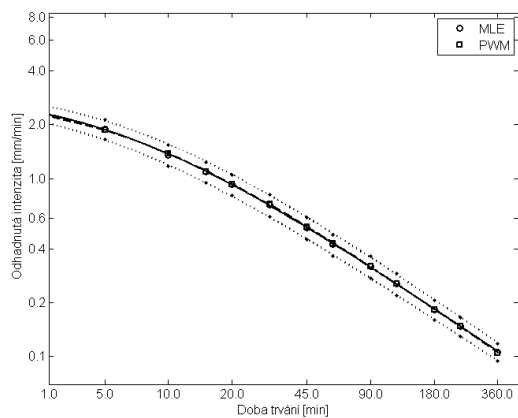
Obrázek 4.9: IDF křivky získané metodou maximální věrohodnosti (vlevo) a metodou PVM (vpravo) v logaritmických souřadnicích na obou osách.

Shodu mezi metodami lze pozorovat i na obrázku 4.10, kde jsou IDF křivky odhadnuté užitím obou metod doplněny 95% intervaly spolehlivosti pro IDF křivky odhadnuté metodou maximální věrohodnosti. Z obrázků je vidět, že IDF křivky získané metodou PVM se nachází uvnitř 95% intervalů spolehlivosti pro IDF křivky získané metodou maximální věrohodnosti. Podobné průběhy lze pozorovat i pro všechny zbývající stanice. Až na mírné rozdíly v odhadu IDF křivek pro stanici Brno–Jundrov je tedy možné konstatovat, že rozdíly v odhadech IDF křivek mezi metodou maximální věrohodnosti a metodou PVM se jeví jako zanedbatelné.

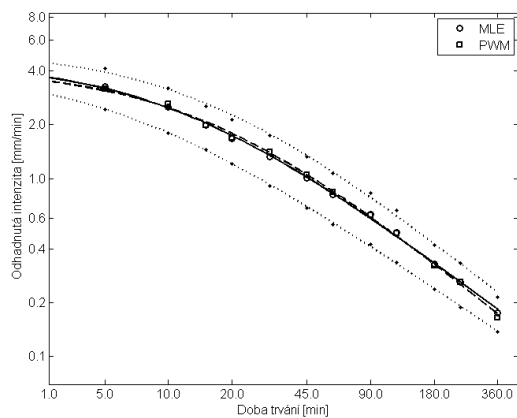
Závěrem je třeba zdůraznit, že navzdory faktu, že odhady IDF křivek pro stanici Brno–Jundrov vykazují podobné průběhy jako křivky u ostatních brněnských stanic, analyzovaná časová řada ze stanice Brno–Jundrov je krátká a není tedy příliš vhodná pro praktické využití z důvodu vyšší variability odhadů.

Výsledky získané použitím teorie extrémních hodnot v rámci této kapitoly byly následně porovnány se stávajícími hydrologickými podklady (viz [91]), přičemž toto srovnání

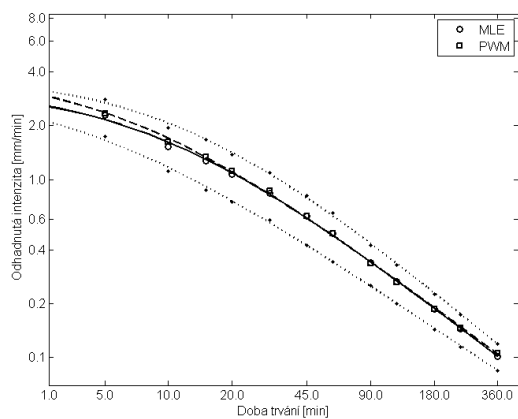
lze najít v práci [50].



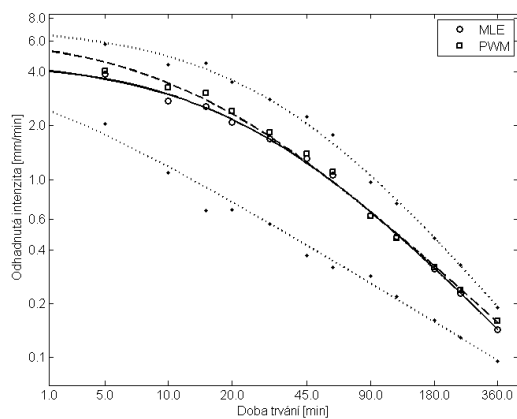
(a) Brno–Tuřany



(b) Brno–Tuřany



(c) Brno–Jundrov



(d) Brno–Jundrov

Obrázek 4.10: IDF křivky získané metodou maximální věrohodnosti (95% intervaly spolehlivosti jsou vyznačeny tečkovaně) a metodou PVM v logaritmických souřadnicích na obou osách.



## 5 Inference pro cenzorovaná rozdělení z Gumbelova oboru atraktivity

Tato kapitola je věnována statistické inferenci pro vybraná cenzorovaná rozdělení z oboru atraktivity Gumbelova rozdělení, protože Gumbelova třída se v praxi často vyskytuje. Inference cenzorovaných rozdělení byla motivována potřebou analyzovat reálná chemická data, která budou blíže popsána v kapitole 6. Konkrétně půjde o data, která jsou reprezentována dvojnásobně zleva cenzorovaným výběrem s cenzorováním typu I. Jelikož dosud nebyly dostupné žádné statistické metody, které by bylo možné snadno aplikovat na tento typ dat, bylo nutné vyvinout zcela novou metodiku, která vychází zejména z autorových prací [33, 34, 35]. Nejprve budou shrnuty základní poznatky o cenzorovaných výběrech a dále bude pozornost věnována odvození modelů pro vybraná cenzorovaná rozdělení z oboru atraktivity Gumbelova rozdělení.

V technické praxi lze často narazit na případy, kdy náhodný výběr není úplný. Například při sledování experimentálních jednotek může nastat situace, kdy rizikový jev (například porouchání součástky) není pozorován u všech jednotek. V takovém případě se mluví o neúplných nebo také cenzorovaných náhodných výběrech.

Obecně se rozlišují dva druhy cenzorování—zleva a zprava. Cenzorování zprava se často užívá v analýze přežití, kdy není možné pozorovat experimentální jednotky po celou dobu jejich provozu až do poruchy. Cenzorování zleva se používá při analýze environmentálních nebo chemických dat, například když se analyzovaná látka vyskytuje v tak nízké koncentraci, že nepřesáhne detekční limit měřicího zařízení. Oba druhy cenzorování se mohou lišit v závislosti na době pozorování experimentálních jednotek nebo detekčním limitu měřicího zařízení. Jestliže je detekční limit fixní, jedná se o cenzorování typu I nebo též o cenzorování časem. V takovém případě je počet cenzorovaných experimentálních jednotek náhodná veličina. Jestliže je fixní počet cenzorovaných jednotek, jedná se o cenzorování typu II nebo též o cenzorování poruchou.

Různé techniky cenzorování a metody statistické inference cenzorovaných dat jsou detailně popsány v mnoha monografiích (viz např. [14, 16]). Většina autorů se zabývá cenzorováním zprava, které je v literatuře dobře rozpracováno pro všechna běžná rozdělení pravděpodobností. V případě cenzorování zleva již nejsou literární prameny tak obsáhlé. Vzhledem k aplikacím cenzorovaných výběrů při analýze chemických dat (viz kapitola 6) bude další pozornost soustředěna na cenzorování typu I zleva.

V mnoha pracích je cenzorování zleva založeno na normálním rozdělení, což je také dobře rozpracováno v literatuře (viz např. [26, 27]). Často je také nutné se zaměřit na mnohonásobně cenzorované výběry (viz např. [1, 2, 14]). V takovém případě je definováno více detekčních limitů  $DL_1 < DL_2 < \dots < DL_k$ ,  $k > 1$ , přičemž k dispozici jsou pouze pozorování nad nejvyšším detekčním limitem  $DL_k$  a počty pozorování pod zbývajících detekčními limity. Mnoho autorů se ve svých pracích zabývá mnohonásobně cenzorovanými

výběry z normálního rozdělení (viz [1, 2]). Toto rozdělení však není příliš vhodné v situacích, kdy měřená veličina (například koncentrace chemické sloučeniny) nabývá pouze kladných hodnot, rozdělení této veličiny je asymetrické a má kladnou šikmost. V takové situaci je vhodnější použít například exponenciální nebo Weibullovo rozdělení, která náleží do oboru atraktivitu Gumbelova rozdělení. Proto bude další pozornost věnována mnohonásobně zleva cenzorovaným výběrům z exponenciálního a Weibullova rozdělení, přičemž cenzorování bude typu I.

Nechť náhodný výběr  $X_1, \dots, X_n$  je mnohonásobně cenzorován zleva s cenzorováním typu I. Dále nechť  $X_{(1)}, \dots, X_{(n)}$  označuje uspořádaný výběr  $X_1, \dots, X_n$ . Pro jednoduchost budou detekční limity označeny  $DL_1 = d_1, DL_2 = d_2, \dots, DL_k = d_k$ , položí se  $d_0 = 0$  a pro zjednodušení zápisu některých vyjádření se položí také  $\ln(d_0) = 0$ . Veličina  $N_i$  bude značit počet pozorování nad detekčním limitem  $d_{i-1}$  a pod nebo rovno detekčnímu limitu  $d_i$ , tedy v intervalu  $(d_{i-1}; d_i]$ , a  $N_0$  bude počet necenzorovaných pozorování. Tedy pro  $N_0 > 0$  jsou pozorování  $X_{(n-N_0+1)}, \dots, X_{(n)}$  necenzorovaná.

Nyní bude uvedeno několik pomocných tvrzení, která budou využita později v důkazech vět 5.7 a 5.17.

**Lemma 5.1.** *Za předpokladu cenzorování typu I má náhodný vektor  $(N_0, N_1, \dots, N_k)$  multinomické rozdělení, tedy*

$$(N_0, N_1, \dots, N_k) \sim \text{Mu}_{k+1}(n, \theta_0, \theta_1, \dots, \theta_k),$$

kde  $\theta_i = F(d_i) - F(d_{i-1})$ ,  $i = 1, \dots, k$ ,  $\theta_0 = 1 - F(d_k)$  a  $n = N_0 + N_1 + \dots + N_k$ .

*Důkaz.* Zřejmé. □

**Lemma 5.2.** *Za předpokladu cenzorování typu I mají marginální četnosti  $N_i$  binomické rozdělení, tedy*

$$N_i \sim \text{Bi}(n, \theta_i), \quad i = 0, 1, \dots, k,$$

kde  $\theta_i = F(d_i) - F(d_{i-1})$ ,  $i = 1, \dots, k$ ,  $\theta_0 = 1 - F(d_k)$  a  $n = N_0 + N_1 + \dots + N_k$ .

*Důkaz.* Plyne z lemmatu 5.1. □

V dalších odstavcích bude popsána nově vyvinutá metodika pro analýzu mnohonásobně zleva cenzorovaných výběrů z exponenciálního a Weibullova rozdělení, přičemž cenzorování bude typu I. Pro odhad neznámých parametrů cenzorovaných rozdělení se často používá věrohodnostní přístup (viz např. [5, 7, 64, 65] nebo kapitola 2 této práce). Stejně bude učiněno i v této práci.

## 5.1 Cenzorované exponenciální rozdělení

Při analýze cenzorovaných dat pocházejících z exponenciálního rozdělení jsou mnohé výsledky, zejména pro cenzorování zprava, dobře známy již od poloviny minulého století (viz např. [23]). Novější práce o statistické analýze náhodných výběrů z exponenciálního rozdělení pro různé typy cenzorování lze najít například v [30, 57, 87], což však nepokrývá výše nastíněnou situaci mnohonásobně zleva cenzorovaných výběrů z exponenciálního rozdělení s cenzorováním typu I. Nejprve bude popsán model mnohonásobně zleva cenzorovaného exponenciálního rozdělení pro jeden výběr, který bude následně rozšířen na případ srovnání dvou nezávislých cenzorovaných výběrů z exponenciálního rozdělení. Z toho důvodu budou v závěru odstavce odvozeny testovací statistiky pro srovnání dvou nezávislých mnohonásobně zleva cenzorovaných výběrů z exponenciálního rozdělení.

### 5.1.1 Model exponenciálního rozdělení pro jeden výběr

Nechť  $X_1, \dots, X_n$  je náhodný výběr z exponenciálního rozdělení s parametrem  $\lambda > 0$ , distribuční funkcí

$$F(x, \lambda) = \begin{cases} 1 - \exp\left(-\frac{x}{\lambda}\right) & \text{pro } x \geq 0, \\ 0 & \text{pro } x < 0, \end{cases} \quad (5.1)$$

a hustotou

$$f(x, \lambda) = \begin{cases} \frac{1}{\lambda} \exp\left(-\frac{x}{\lambda}\right) & \text{pro } x \geq 0, \\ 0 & \text{pro } x < 0. \end{cases} \quad (5.2)$$

Cenzorovaný výběr  $X_1, \dots, X_n$  může být také reprezentován transformovaným náhodným výběrem  $Y_1, \dots, Y_n$  jako

$$Y_i = \begin{cases} X_i & \text{pro } X_i > d_k, \\ d_i & \text{pro } d_{i-1} < X_i \leq d_i, \end{cases}$$

přičemž každá z transformovaných náhodných veličin  $Y_i$  má hustotu

$$g(x, \lambda) = \begin{cases} F(d_i, \lambda) - F(d_{i-1}, \lambda) & \text{pro } x = d_i, \quad i = 1, \dots, k, \\ f(x, \lambda) & \text{pro } x > d_k, \end{cases} \quad (5.3)$$

vzhledem k míře  $\mu$ , což je Lebesgueova míra na  $(d_k; \infty)$ , a čítací míře, která přiřazuje míru jedna každému bodu  $y = d_i$ ,  $i = 1, \dots, k$ . Rozdělení daná hustotou  $g(\cdot, \lambda)$ ,  $\lambda > 0$ , jsou obecně různá, mají společný nosič  $(0; \infty)$ , jednotlivá pozorování  $Y_1, \dots, Y_n$  jsou nezávislá a skutečná hodnota parametru  $\lambda$  je vnitřním bodem intervalu  $(0; \infty)$ .

Užitím rovnice (5.3), viz také [14], lze věrohodnostní funkci cenzorovaného výběru zapsat ve tvaru

$$L(\lambda, N_0, N_1, \dots, N_k, X_{(n-N_0+1)}, \dots, X_{(n)}) = \frac{n!}{N_1! \dots N_k!} \prod_{i=1}^k [F(d_i, \lambda) - F(d_{i-1}, \lambda)]^{N_i} \\ \times \prod_{i=n-N_0+1}^n f(X_{(i)}),$$

přičemž se položí  $\prod_{i=n-N_0+1}^n f(X_{(i)}) = 1$  pro  $N_0 = 0$ . Logaritmická věrohodnostní funkce je pak tvaru

$$l(\lambda, N_0, N_1, \dots, N_k, X_{(n-N_0+1)}, \dots, X_{(n)}) = \ln L = \ln \left( \frac{n!}{N_1! \dots N_k!} \right) \\ + \sum_{i=1}^k N_i \ln [F(d_i, \lambda) - F(d_{i-1}, \lambda)] \quad (5.4) \\ + \sum_{i=n-N_0+1}^n \ln [f(X_{(i)})].$$

Jestliže  $F'_\lambda$  označuje derivaci distribuční funkce  $F$  podle parametru  $\lambda$ ,  $f'_\lambda$  označuje derivaci hustoty  $f$  podle parametru  $\lambda$  a

$$H_i(\lambda) = \frac{\partial}{\partial \lambda} \ln [F(d_i, \lambda) - F(d_{i-1}, \lambda)] \\ = \frac{F'_\lambda(d_i, \lambda) - F'_\lambda(d_{i-1}, \lambda)}{F(d_i, \lambda) - F(d_{i-1}, \lambda)} = \frac{d_{i-1} \exp\left(-\frac{d_{i-1}}{\lambda}\right) - d_i \exp\left(-\frac{d_i}{\lambda}\right)}{\lambda^2 \left[ \exp\left(-\frac{d_{i-1}}{\lambda}\right) - \exp\left(-\frac{d_i}{\lambda}\right) \right]},$$

pak věrohodnostní rovnice pro odhad parametru  $\lambda$  je tvaru

$$\frac{\partial l}{\partial \lambda} = \sum_{i=1}^k N_i H_i(\lambda) + \sum_{i=n-N_0+1}^n \frac{f'_\lambda(X_{(i)})}{f(X_{(i)})} = 0. \quad (5.5)$$

Dosazením hustoty (5.2) do vztahu (5.5) lze věrohodnostní rovnici přepsat jako

$$\frac{\partial l}{\partial \lambda} = \sum_{i=1}^k N_i H_i(\lambda) - \frac{N_0}{\lambda} + \frac{1}{\lambda^2} \sum_{i=n-N_0+1}^n X_{(i)} = 0. \quad (5.6)$$

MV odhad  $\hat{\lambda}$  parametru  $\lambda$  lze obdržet jako řešení rovnice (5.6). Rovnici je však nutné řešit numericky. Lze také využít numerickou proceduru založenou na Nelderově-Meadově simplexovém algoritmu (viz [60]), který je implementován v Matlabu (verze 7.12, R2011a), a maximalizovat logaritmickou věrohodnostní funkci (5.4). Numerická studie (viz také [46, 76]) prokázala, že řešení je jediné. Tedy podle [65], užitím vlastností hustoty  $g$  zmíněné výše, je řešení rovnice (5.6) konzistentním odhadem parametru  $\lambda$ .

Dále bude pomocí FMI (viz kapitola 2) odvozen rozptyl odhadu parametru  $\lambda$ . Nejprve bude odvozena výběrová FMI, která je nestranným odhadem očekávané FMI, a která je v mnoha aplikacích upřednostňována (viz např. [2, 28]), zvláště pak v situacích, kdy je stanovení očekávané FMI komplikované.

**Věta 5.3.** Výběrová FMI mnohonásobně zleva cenzorovaného (cenzorování typu I) náhodného výběru z exponenciálního rozdělení s distribuční funkcí (5.1) je tvaru

$$\tilde{J}_n(\lambda) = - \sum_{i=1}^k N_i H'_i(\lambda) - \frac{N_0}{\lambda^2} + \frac{2}{\lambda^3} \sum_{i=n-N_0+1}^n X_{(i)}, \quad (5.7)$$

kde

$$H'_i(\lambda) = \frac{\partial H_i(\lambda)}{\partial \lambda} = \frac{(d_{i-1}^2 - 2d_{i-1}\lambda) \exp\left(-\frac{d_{i-1}}{\lambda}\right) - (d_i^2 - 2d_i\lambda) \exp\left(-\frac{d_i}{\lambda}\right)}{\lambda^4 \left[ \exp\left(-\frac{d_{i-1}}{\lambda}\right) - \exp\left(-\frac{d_i}{\lambda}\right) \right]} - \frac{\left[ d_{i-1} \exp\left(-\frac{d_{i-1}}{\lambda}\right) - d_i \exp\left(-\frac{d_i}{\lambda}\right) \right]^2}{\lambda^4 \left[ \exp\left(-\frac{d_{i-1}}{\lambda}\right) - \exp\left(-\frac{d_i}{\lambda}\right) \right]^2}.$$

*Důkaz.* Jak bylo uvedeno v kapitole 2, výběrová FMI je definována jako

$$\tilde{J}_n(\lambda) = - \frac{\partial^2 l}{\partial \lambda^2}.$$

Užitím vztahu (5.6) lze pak snadno stanovit druhou derivaci logaritmické věrohodnostní funkce podle parametru  $\lambda$ , což dává tvrzení věty. □

Další pozornost bude věnována stanovení očekávané FMI. K tomu bude potřeba několik pomocných tvrzení.

**Lemma 5.4.** Necht  $X_1, \dots, X_n$  je náhodný výběr z exponenciálního rozdělení. Pak hustota pořadové veličiny  $X_{(i)}$  je tvaru

$$f_{(i)}(x) = n \binom{n-1}{i-1} \frac{1}{\lambda} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} \exp\left(-\frac{(n-i+j+1)x}{\lambda}\right), \quad i = 1, \dots, n.$$

*Důkaz.* Hustota veličiny  $X_{(i)}$  je obecně tvaru (viz [52])

$$f_{(i)}(x) = n \binom{n-1}{i-1} f(x) [F(x)]^{i-1} [1 - F(x)]^{n-i}, \quad i = 1, 2, \dots, n.$$

Místo  $F$  se dosadí distribuční funkce (5.1) a místo  $f$  hustota (5.2) necenzorovaného exponenciálního rozdělení a postupnými úpravami lze obdržet hustotu  $f_{(i)}(x)$  ve tvaru

$$\begin{aligned} f_{(i)}(x) &= n \binom{n-1}{i-1} \frac{1}{\lambda} \exp\left(-\frac{x}{\lambda}\right) \left[1 - \exp\left(-\frac{x}{\lambda}\right)\right]^{i-1} \exp\left(-\frac{n-i}{\lambda}x\right) \\ &= n \binom{n-1}{i-1} \frac{1}{\lambda} \exp\left(-\frac{(n-i+1)x}{\lambda}\right) \sum_{j=0}^{i-1} \binom{i-1}{j} (-1)^j \exp\left(-\frac{jx}{\lambda}\right) \\ &= n \binom{n-1}{i-1} \frac{1}{\lambda} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} \exp\left(-\frac{(n-i+j+1)x}{\lambda}\right). \end{aligned}$$

□

**Lemma 5.5.** *Nechť  $X_1, \dots, X_n$  je náhodný výběr z exponenciálního rozdělení. Pak střední hodnota  $E(X_{(i)})$  náhodné veličiny  $X_{(i)}$  je tvaru*

$$E(X_{(i)}) = n \binom{n-1}{i-1} \lambda \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} (n-i+j+1)^{-2}, \quad i = 1, \dots, n.$$

*Důkaz.* Postupnými úpravami s využitím lemmatu 5.4 a faktu, že  $\int_0^\infty x e^{-x} dx = \Gamma(2) = 1$ , lze odvodit střední hodnotu  $E(X_{(i)})$  ve tvaru

$$\begin{aligned} E(X_{(i)}) &= \int_0^\infty x f_{(i)}(x) dx \\ &= n \binom{n-1}{i-1} \frac{1}{\lambda} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} \int_0^\infty x \exp\left(-\frac{(n-i+j+1)x}{\lambda}\right) dx \\ &= n \binom{n-1}{i-1} \frac{1}{\lambda} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} (n-i+j+1)^{-2} \int_0^\infty t \exp\left(-\frac{t}{\lambda}\right) dt \\ &= n \binom{n-1}{i-1} \lambda \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} (n-i+j+1)^{-2} \int_0^\infty z \exp(-z) dz \\ &= n \binom{n-1}{i-1} \lambda \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} (n-i+j+1)^{-2}. \end{aligned}$$

□

**Lemma 5.6.** *Nechť  $X_1, \dots, X_n$  je náhodný výběr z exponenciálního rozdělení. Pak střední hodnota součtu náhodných veličin  $E\left(\sum_{i=n-N_0+1}^n X_{(i)}\right)$  je tvaru*

$$\begin{aligned} E\left(\sum_{i=n-N_0+1}^n X_{(i)}\right) &= \sum_{n_0=0}^n \left[ \sum_{i=n-n_0+1}^n E(X_{(i)}) \right] \binom{n}{n_0} \exp\left(-\frac{n_0 d_k}{\lambda}\right) \\ &\quad \times \left[ 1 - \exp\left(-\frac{d_k}{\lambda}\right) \right]^{n-n_0}. \end{aligned}$$

*Důkaz.* Z vlastností podmíněné pravděpodobnosti plyne, že  $E(E(X|Y)) = E(X)$ , kde  $X$ ,  $Y$  jsou náhodné veličiny. Vzhledem k tomu, že z lemmatu 5.2 plyne, že  $N_0 \sim \text{Bi}(n, \theta_0)$ , postupnými úpravami lze odvodit

$$\begin{aligned} E\left(\sum_{i=n-N_0+1}^n X_{(i)}\right) &= E\left[E\left(\sum_{i=n-N_0+1}^n X_{(i)} \mid N_0\right)\right] \\ &= \sum_{n_0=0}^n E\left(\sum_{i=n-n_0+1}^n X_{(i)} \mid N_0 = n_0\right) P(N_0 = n_0) \\ &= \sum_{n_0=0}^n \left[ \sum_{i=n-n_0+1}^n E(X_{(i)}) \right] \binom{n}{n_0} \theta_0^{n_0} (1-\theta_0)^{n-n_0} \\ &= \sum_{n_0=0}^n \left[ \sum_{i=n-n_0+1}^n E(X_{(i)}) \right] \binom{n}{n_0} \exp\left(-\frac{n_0 d_k}{\lambda}\right) \left[ 1 - \exp\left(-\frac{d_k}{\lambda}\right) \right]^{n-n_0}. \end{aligned}$$

□

**Věta 5.7.** Očekávaná FMI mnohonásobně zleva cenzorovaného (cenzorování typu I) náhodného výběru z exponenciálního rozdělení s distribuční funkcí (5.1) je tvaru

$$J_n(\lambda) = nJ(\lambda),$$

kde

$$\begin{aligned} J(\lambda) = & \sum_{i=1}^k \frac{\left[ d_{i-1} \exp\left(-\frac{d_{i-1}}{\lambda}\right) - d_i \exp\left(-\frac{d_i}{\lambda}\right) \right]^2}{\lambda^4 \left[ \exp\left(-\frac{d_{i-1}}{\lambda}\right) - \exp\left(-\frac{d_i}{\lambda}\right) \right]} + \frac{(d_k^2 - 2d_k\lambda) \exp\left(-\frac{d_k}{\lambda}\right)}{\lambda^4} \\ & - \frac{1}{\lambda^2} \exp\left(-\frac{d_k}{\lambda}\right) + \frac{2}{\lambda^2} \sum_{n_0=0}^n \left\{ \binom{n}{n_0} \exp\left(-\frac{n_0 d_k}{\lambda}\right) \left[ 1 - \exp\left(-\frac{d_k}{\lambda}\right) \right]^{n-n_0} \right. \\ & \left. \times \sum_{i=n-n_0+1}^n \binom{n-1}{i-1} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} (n-i+j+1)^{-2} \right\}. \end{aligned}$$

*Důkaz.* Z poznámky 2.5 plyne, že

$$J_n(\lambda) = \mathbb{E} \tilde{J}_n(\lambda). \quad (5.8)$$

Dosazením  $\tilde{J}_n$  ze vztahu (5.7) do (5.8) dostáváme

$$J_n(\lambda) = - \sum_{i=1}^k H'_i(\lambda) \mathbb{E}(N_i) - \frac{\mathbb{E}(N_0)}{\lambda^2} + \frac{2}{\lambda^3} \mathbb{E} \left( \sum_{i=n-N_0+1}^n X_{(i)} \right). \quad (5.9)$$

Z lemmatu 5.2 plyne, že  $N_i \sim \text{Bi}(n, \theta_i)$ ,  $i = 0, \dots, k$ , tedy očekávané četnosti jsou tvaru

$$\mathbb{E}(N_i) = n\theta_i = \begin{cases} n \left[ \exp\left(-\frac{d_{i-1}}{\lambda}\right) - \exp\left(-\frac{d_i}{\lambda}\right) \right] & \text{pro } i = 1, \dots, k, \\ n \exp\left(-\frac{d_k}{\lambda}\right) & \text{pro } i = 0. \end{cases}$$

Kombinací výsledků lemmat 5.5 a 5.6 dostáváme

$$\begin{aligned} \mathbb{E} \left( \sum_{i=n-N_0+1}^n X_{(i)} \right) = & \sum_{n_0=0}^n \sum_{i=n-n_0+1}^n n \binom{n-1}{i-1} \lambda \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} \\ & \times (n-i+j+1)^{-2} \binom{n}{n_0} \exp\left(-\frac{n_0 d_k}{\lambda}\right) \left[ 1 - \exp\left(-\frac{d_k}{\lambda}\right) \right]^{n-n_0}. \end{aligned}$$

Dosazením výše uvedených výsledků do vztahu (5.9) dostaneme tvrzení věty. □

Vzhledem k asymptotickým vlastnostem MV odhadu  $\hat{\lambda}$  má (viz věta 2.11)  $\sqrt{n}(\hat{\lambda} - \lambda)$  asymptoticky normální rozdělení  $N(0; \sigma^2(\lambda))$ , kde  $\sigma^2(\lambda) = J^{-1}(\lambda)$ . Asymptotické vlastnosti odhadu  $\hat{\lambda}$  a odhadu jeho rozptylu (vycházejícího z výběrové a očekávané FMI) pro různé rozsahy výběru a různý počet detekčních limitů  $k$  budou vyšetřeny v simulační studii v následujícím odstavci.

### 5.1.2 Posouzení vlastností odhadů na základě simulací

Odhady parametrů a parametrických funkcí odvozených v předchozím odstavci budou dále analyzovány a porovnány užitím simulací. Nejprve bylo vygenerováno 10 000 jednonásobně, dvojnásobně a trojnásobně zleva cenzorovaných výběrů rozsahu  $n = 10, 20, 30, 50, 100, 500$  z exponenciálního rozdělení. Jelikož  $\lambda$  je parametr měřítka a MV odhady jsou invariantní vůči měřítku, lze bez ztráty na obecnosti položit  $\lambda = 1$ . Detekční limity  $d_i, i = 1, \dots, k, k = 1, 2, 3$ , byly zvoleny jako kvantily exponenciálního rozdělení užitím rovnice  $q_i = F(d_i, \lambda)$ , kde  $q_i$  jsou uvedeny v tabulce 5.1.

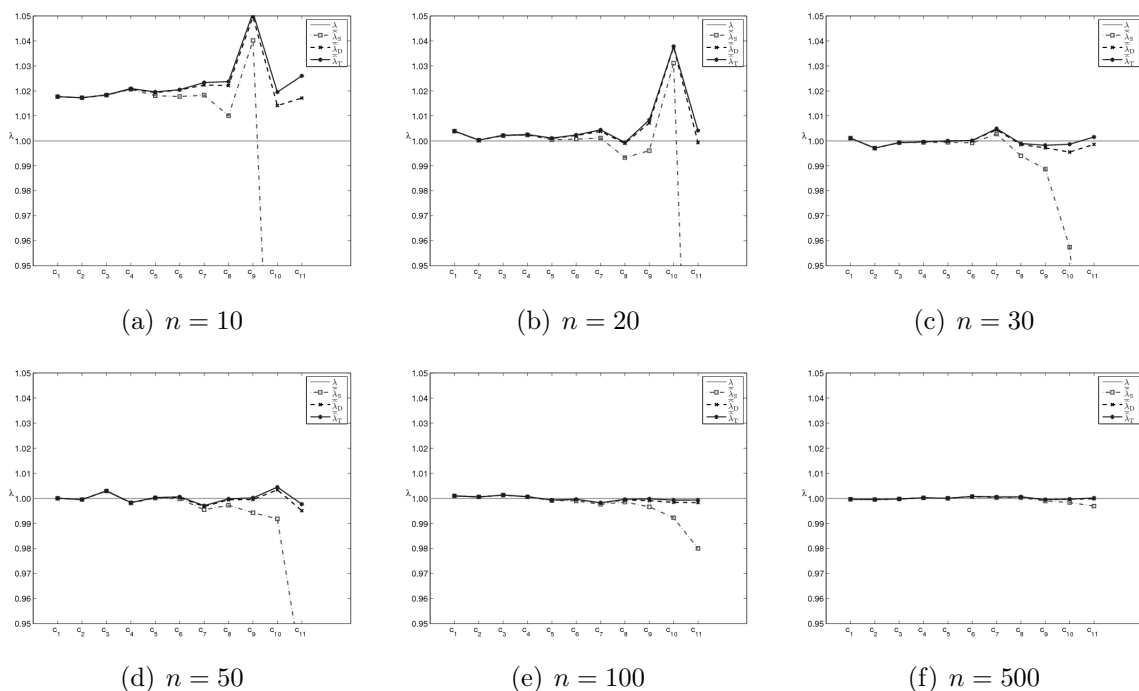
Cenzorování	Jednonásobné	Dvojnásobné		Trojnásobné		
	$q_1$	$q_1$	$q_2$	$q_1$	$q_2$	$q_3$
$c_1$	0,05	0,03	0,05	0,02	0,03	0,05
$c_2$	0,10	0,05	0,10	0,03	0,07	0,10
$c_3$	0,20	0,10	0,20	0,07	0,13	0,20
$c_4$	0,30	0,15	0,30	0,10	0,20	0,30
$c_5$	0,40	0,20	0,40	0,13	0,27	0,40
$c_6$	0,50	0,25	0,50	0,17	0,33	0,50
$c_7$	0,60	0,30	0,60	0,20	0,40	0,60
$c_8$	0,70	0,35	0,70	0,23	0,47	0,70
$c_9$	0,80	0,40	0,80	0,27	0,53	0,80
$c_{10}$	0,90	0,45	0,90	0,30	0,60	0,90
$c_{11}$	0,95	0,48	0,95	0,32	0,63	0,95

Tabulka 5.1: Kvantily pro stanovení detekčních limitů pro jednonásobné, dvojnásobné a trojnásobné cenzorování a různá schémata cenzorování  $c_i, i = 1, \dots, 11$ .

Příslušná schémata cenzorování jsou označena jako  $c_1, \dots, c_{11}$  a odpovídají kvantilům, které určují detekční limity pro  $k = 1, 2, 3$ . Tedy například  $q_1$  ve sloupci *Jednonásobné* v tabulce 5.1 značí podíl jednonásobně cenzorovaných pozorování a tím popisuje dané schéma cenzorování. Schéma cenzorování  $c_1$  reprezentuje nejmenší podíl (5 %) cenzorovaných dat a schéma  $c_{11}$  reprezentuje největší podíl (95 %) cenzorovaných dat pro jednonásobně, dvojnásobně a trojnásobně cenzorované výběry. Velký podíl cenzorovaných pozorování odpovídá reálným datům, která budou analyzována v kapitole 6.

Dále byly užitím rovnice (5.6) spočteny MV odhady parametru  $\lambda$  ze všech 10 000 výběrů a jejich průměrné hodnoty  $\widehat{\lambda}$  jsou na obrázku 5.1 pro různá schémata cenzorování z tabulky 5.1. Z obrázku 5.1 a tabulky 5.2, kde jsou uvedeny průměrné MV odhady  $\widehat{\lambda}$  a jejich průměrné střední kvadratické chyby, je vidět, že pro malé rozsahy výběru ( $n < 20$ ) jsou MV odhady parametru  $\lambda$  značně vychýlené i v případě, kdy počet cenzorovaných pozorování je nízký. Pro  $n \geq 20$  je vychýlení MV odhadů velmi malé a z praktického pohledu zanedbatelné pro schémata cenzorování  $c_6$  (viz obrázek 5.1) a nižší. Dopad násobnosti cenzorování na odhad parametru  $\lambda$  je patrný pouze pro vyšší detekční limity v závislosti na





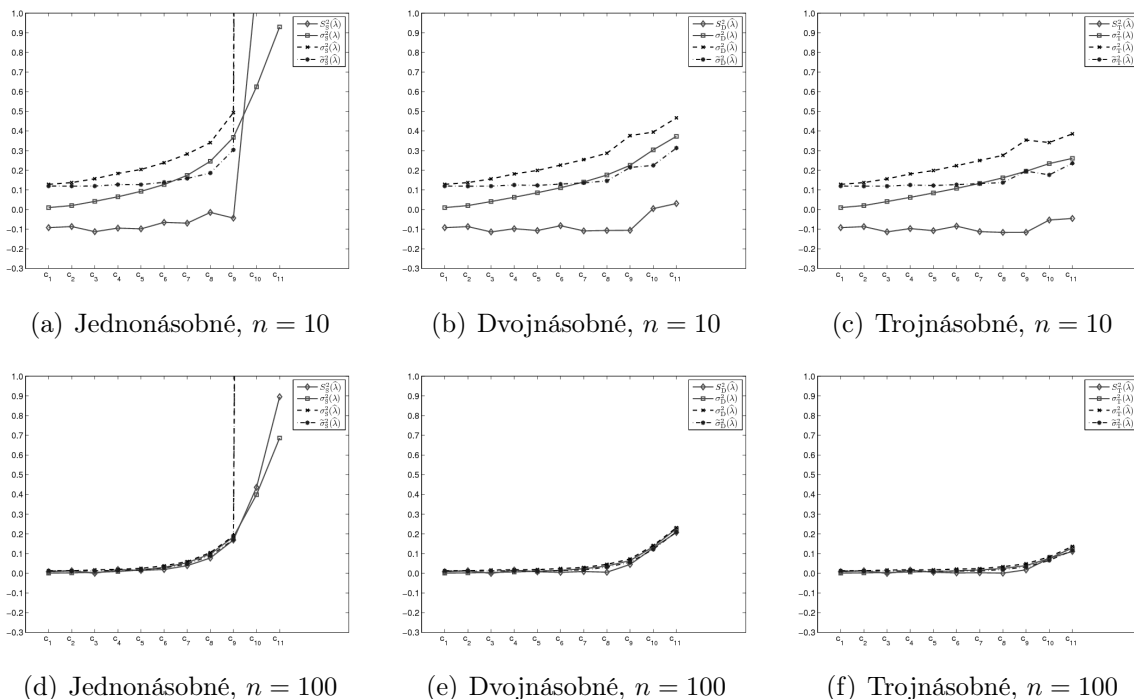
Obrázek 5.1: Průměrné MV odhady parametru  $\lambda$  pro jednonásobné (index S), dvojnásobné (index D), trojnásobné (index T) cenzorování a různé rozsahy výběru  $n$ .

Cenzorování	$n = 10$			$n = 100$		
	S	D	T	S	D	T
$c_1$	1,02 (0,092)	1,02 (0,092)	1,02 (0,092)	1,00 (0,010)	1,00 (0,010)	1,00 (0,010)
$c_2$	1,02 (0,092)	1,02 (0,092)	1,02 (0,092)	1,00 (0,010)	1,00 (0,010)	1,00 (0,010)
$c_3$	1,02 (0,090)	1,02 (0,090)	1,02 (0,090)	1,00 (0,010)	1,00 (0,010)	1,00 (0,010)
$c_4$	1,02 (0,091)	1,02 (0,091)	1,02 (0,091)	1,00 (0,010)	1,00 (0,010)	1,00 (0,010)
$c_5$	1,02 (0,091)	1,02 (0,090)	1,02 (0,090)	1,00 (0,010)	1,00 (0,010)	1,00 (0,010)
$c_6$	1,02 (0,094)	1,02 (0,093)	1,02 (0,092)	1,00 (0,010)	1,00 (0,010)	1,00 (0,010)
$c_7$	1,02 (0,094)	1,02 (0,090)	1,02 (0,090)	1,00 (0,010)	1,00 (0,010)	1,00 (0,010)
$c_8$	1,01 (0,099)	1,02 (0,090)	1,02 (0,090)	1,00 (0,011)	1,00 (0,010)	1,00 (0,010)
$c_9$	1,04 (0,097)	1,05 (0,092)	1,05 (0,092)	1,00 (0,012)	1,00 (0,010)	1,00 (0,010)
$c_{10}$	0,82 (0,351)	1,01 (0,101)	1,02 (0,095)	0,99 (0,016)	1,00 (0,011)	1,00 (0,011)
$c_{11}$	0,61 (0,568)	1,02 (0,103)	1,03 (0,096)	0,98 (0,025)	1,00 (0,012)	1,00 (0,011)

Tabulka 5.2: Průměrné MV odhady  $\bar{\lambda}$  a jejich průměrné střední kvadratické chyby (v závorce) pro jednonásobné (S), dvojnásobné (D), trojnásobné (T) cenzorování a různé rozsahy výběru  $n$ .

rozsahu výběru. Pro  $n \geq 30$  je rozdíl mezi jednonásobným, dvojnásobným a trojnásobným cenzorováním téměř zanedbatelný až do schématu  $c_7$ , při kterém je v souladu s očekáváním nejvyšší vychýlení odhadu v případě jednonásobného cenzorování.

Rozptyl MV odhadu  $\hat{\lambda}$  parametru  $\lambda$  je možné odhadnout výběrovým rozptylem  $s^2(\hat{\lambda})$  ze simulovaných hodnot. Dále bude pomocí simulací porovnán asymptotický rozptyl  $\sigma^2(\lambda) = J^{-1}(\lambda)$  s odhadem  $\sigma^2(\hat{\lambda}) = J^{-1}(\hat{\lambda})$  založeným na očekávané FMI (viz věta 5.7) a odhadem  $\tilde{\sigma}^2(\hat{\lambda}) = \tilde{J}^{-1}(\hat{\lambda})$  založeným na výběrové FMI (viz věta 5.3). Odhady  $\sigma^2(\hat{\lambda})$  a  $\tilde{\sigma}^2(\hat{\lambda})$  byly zprůměrovány (10 000 výběrů) a od teď, když se bude mluvit o odhadech  $\sigma^2(\hat{\lambda})$ ,  $\tilde{\sigma}^2(\hat{\lambda})$  asymptotického rozptylu  $\sigma^2(\lambda)$ , budou tím myšleny průměrné odhady z 10 000 výběrů (tedy  $\sigma^2(\hat{\lambda}) = \overline{J^{-1}(\hat{\lambda})}$ ,  $\tilde{\sigma}^2(\hat{\lambda}) = \overline{\tilde{J}^{-1}(\hat{\lambda})}$ ). Tyto dva odhady společně s příslušným empirickým výběrovým rozptylem  $S^2 = ns^2(\hat{\lambda})$  budou dále srovnány s asymptotickým rozptylem v závislosti na násobnosti cenzorování (jednonásobné, dvojnásobné, trojnásobné) a rozsahu výběru  $n$ . Vzhledem k relativně velkému počtu výběrů nám odhad  $S^2(\hat{\lambda})$  umožňuje posoudit vychýlení odhadů  $\sigma^2(\hat{\lambda})$ ,  $\tilde{\sigma}^2(\hat{\lambda})$  a asymptotický rozptyl  $\sigma^2(\lambda)$ .



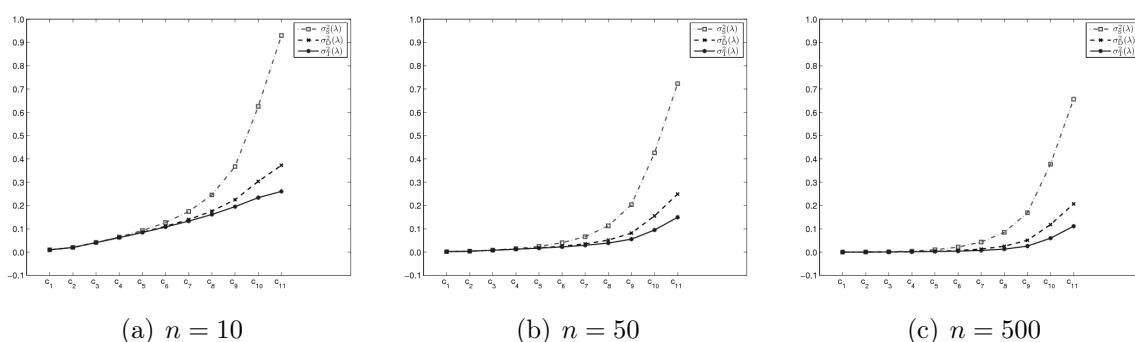
Obrázek 5.2: Srovnání odhadů rozptylu  $S^2(\hat{\lambda}) = ns^2(\hat{\lambda})$ ,  $\sigma^2(\hat{\lambda}) = \overline{J^{-1}(\hat{\lambda})}$ ,  $\tilde{\sigma}^2(\hat{\lambda}) = \overline{\tilde{J}^{-1}(\hat{\lambda})}$  a asymptotického rozptylu  $\sigma^2(\lambda) = J^{-1}(\lambda)$  pro jednonásobné, dvojnásobné, trojnásobné cenzorování a různé rozsahy výběru  $n$  s logaritickým měřítkem na ose  $y$ .

Ze srovnání výše uvedených charakteristik rozptylu (viz obrázek 5.2 a tabulka 5.3 pro dvojnásobné cenzorování) je vidět, že očekávaná výchylka odhadu  $\sigma^2(\hat{\lambda})$  je značná pro malý rozsah výběru, například hodnota  $\sigma^2(\lambda)$  (respektive  $\sigma^2(\hat{\lambda})$ ) je rovna 1,01 (1,14) pro  $n = 10$ , 1,00 (1,04) pro  $n = 30$  a 1,00 (1,01) pro  $n = 100$  uvažujeme-li schéma  $c_1$  a dvojnásobné cenzorování. Dále odhad  $S^2(\hat{\lambda})$  nabývá nižších hodnot než asymptotický rozptyl  $\sigma^2(\lambda)$  pro

Cenzorování	$n = 10$				$n = 100$			
	$S^2(\hat{\lambda})$	$\sigma^2(\lambda)$	$\sigma^2(\hat{\lambda})$	$\tilde{\sigma}^2(\hat{\lambda})$	$S^2(\hat{\lambda})$	$\sigma^2(\lambda)$	$\sigma^2(\hat{\lambda})$	$\tilde{\sigma}^2(\hat{\lambda})$
$c_1$	0,9122	1,0101	1,1372	1,1270	1,0076	1,0010	1,0131	1,0121
$c_2$	0,9170	1,0204	1,1473	1,1266	1,0130	1,0020	1,0133	1,0114
$c_3$	0,8925	1,0419	1,1691	1,1266	0,9994	1,0042	1,0169	1,0129
$c_4$	0,9071	1,0647	1,1990	1,1335	1,0164	1,0068	1,0184	1,0123
$c_5$	0,8987	1,0896	1,2206	1,1310	1,0078	1,0103	1,0189	1,0108
$c_6$	0,9212	1,1174	1,2540	1,1384	1,0049	1,0153	1,0245	1,0143
$c_7$	0,8972	1,1504	1,2898	1,1457	1,0090	1,0234	1,0299	1,0175
$c_8$	0,8990	1,1925	1,3322	1,1573	1,0055	1,0377	1,0465	1,0316
$c_9$	0,8998	1,2524	1,4571	1,2392	1,0460	1,0667	1,0752	1,0575
$c_{10}$	1,0049	1,3550	1,4837	1,2519	1,1370	1,1441	1,1515	1,1293
$c_{11}$	1,0313	1,4517	1,5948	1,3682	1,2308	1,2537	1,2601	1,2336

Tabulka 5.3: Srovnání odhadů rozptylu  $S^2(\hat{\lambda}) = ns^2(\hat{\lambda})$ ,  $\sigma^2(\hat{\lambda}) = \overline{J^{-1}(\hat{\lambda})}$ ,  $\tilde{\sigma}^2(\hat{\lambda}) = \overline{\tilde{J}^{-1}(\hat{\lambda})}$  a asymptotického rozptylu  $\sigma^2(\lambda) = J^{-1}(\lambda)$  pro dvojnásobné cenzorování a různé rozsahy výběru  $n$ .

malé rozsahy výběru, ale selhává pro vysokou úroveň (schéma) jednonásobného cenzorování. Všechny odhady rozptylu jsou téměř identické pro rozsah výběru  $n > 100$ . Variabilita odhadů je podle očekávání nejmenší pro trojnásobné cenzorování a různé rozsahy výběru a schémata cenzorování. Například pro schéma cenzorování  $c_{11}$  a rozsah výběru  $n = 10$  je asymptotický rozptyl  $\sigma^2(\lambda)$  roven 2, 53 v případě jednonásobného, 1, 45 v případě dvojnásobného a 1, 30 v případě trojnásobného cenzorování. Podobně pro rozsah výběru  $n = 100$ .



Obrázek 5.3: Asymptotický rozptyl  $\sigma^2(\lambda) = J^{-1}(\lambda)$  pro jednonásobné (index S), dvojnásobné (index D), trojnásobné (index T) cenzorování a různé rozsahy výběru  $n$  s logaritmickým měřítkem na ose  $x$ .

Asymptotický rozptyl  $\sigma^2(\lambda)$  (získaný z očekávané FMI) byl dále analyzován pro různá schémata cenzorování a rozsahy výběru. Z obrázku 5.3 a tabulky 5.4 je vidět, že rozdíly

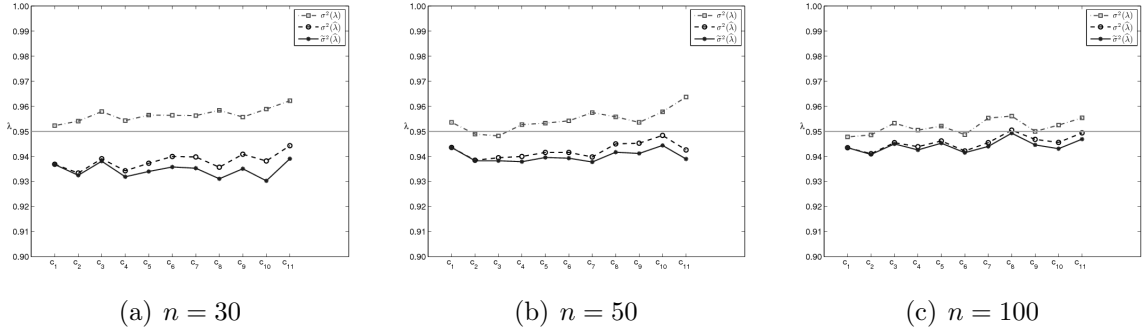
Cenzorování	$n = 10$			$n = 50$			$n = 500$		
	S	D	T	S	D	T	S	D	T
$c_1$	1,010	1,010	1,010	1,002	1,002	1,002	1,000	1,000	1,000
$c_2$	1,021	1,020	1,020	1,004	1,004	1,004	1,000	1,000	1,000
$c_3$	1,043	1,042	1,042	1,009	1,008	1,008	1,002	1,001	1,001
$c_4$	1,067	1,065	1,064	1,015	1,013	1,013	1,004	1,002	1,002
$c_5$	1,097	1,090	1,088	1,025	1,019	1,017	1,010	1,004	1,003
$c_6$	1,136	1,117	1,114	1,041	1,026	1,023	1,022	1,007	1,004
$c_7$	1,191	1,150	1,143	1,069	1,036	1,030	1,045	1,013	1,007
$c_8$	1,279	1,192	1,175	1,120	1,053	1,040	1,089	1,026	1,013
$c_9$	1,444	1,252	1,215	1,226	1,085	1,057	1,184	1,052	1,026
$c_{10}$	1,869	1,355	1,264	1,531	1,168	1,100	1,459	1,126	1,062
$c_{11}$	2,534	1,452	1,298	2,062	1,283	1,161	1,929	1,230	1,118

Tabulka 5.4: Asymptotický rozptyl  $\sigma^2(\lambda) = J^{-1}(\lambda)$  pro jednonásobné (S), dvojnásobné (D), trojnásobné (T) cenzorování a různé rozsahy výběru  $n$ .

mezi asymptotickými rozptyly pro jednonásobné, dvojnásobné a trojnásobné cenzorování jsou zanedbatelné po schéma  $c_6$  pro libovolný rozsah výběru.

Užitím asymptotického rozptylu  $\sigma^2(\lambda)$  lze určit interval spolehlivosti pro parametr  $\lambda$ . Pravděpodobnost pokrytí 95% intervalu spolehlivosti je pak dána procentem případů (z 10 000 opakování), kdy odhadnutý 95% interval spolehlivosti obsahuje skutečnou hodnotu parametru. Výsledky simulací ukázaly, že pravděpodobnosti pokrytí jsou velmi podobné pro všechny rozsahy výběru a schémata  $c_1-c_8$  v případě jednonásobného, dvojnásobného i trojnásobného cenzorování. Z toho důvodu byly spočteny také pravděpodobnosti pokrytí 95% intervalu spolehlivosti pro různé odhady rozptylu, konkrétně  $\sigma^2(\hat{\lambda})$  a  $\tilde{\sigma}^2(\hat{\lambda})$ . Vliv typu odhadu asymptotického rozptylu na pravděpodobnost pokrytí je ilustrován na obrázku 5.4 pro případ dvojnásobného cenzorování. Z obrázku je vidět, že odhad založený na očekávané FMI (viz věta 5.7) vykazuje lepší výsledky než odhad založený na výběrové FMI (viz věta 5.3) pro všechna schémata cenzorování. Podobné výsledky lze pozorovat i v případě jednonásobného a trojnásobného cenzorování.

Z provedených simulací lze usoudit, jaké vychýlení odhadů  $\hat{\lambda}$  a  $\sigma^2(\hat{\lambda})$  lze očekávat pro různé rozsahy výběru  $n = 10, 20, 30, 50, 100, 500$ , schémata cenzorování  $c_1-c_{11}$  (viz tabulka 5.1) a násobnosti cenzorování (jednonásobné, dvojnásobné, trojnásobné). Ukázalo se, že užitím očekávané FMI lze provést přijatelnou statistickou inferenci o parametru  $\lambda$  pro rozsahy výběru  $n \geq 30$  a schémata cenzorování  $c_1-c_6$ , kde  $c_6$  odpovídá 50% kvantilu exponenciálního rozdělení (tedy 50 % cenzorovaných hodnot). Dále v případě vyšší úrovně cenzorování (schémata  $c_6-c_{11}$ ) a malého rozsahu výběru ( $n = 30$ ) je pravděpodobnost pokrytí spočtená s využitím  $\sigma^2(\hat{\lambda})$  (tedy užitím očekávané FMI) o asi 1 % nižší než předepsaná 95% mez spolehlivosti. Použijeme-li odhad  $\tilde{\sigma}^2(\hat{\lambda})$  (tedy užitím výběrové FMI) je pravděpodobnost



Obrázek 5.4: Pravděpodobnosti pokrytí pro parametr  $\hat{\lambda}$  a různé odhady rozptylu při uvažování dvojnásobného cenzorování.

pokrytí o asi 2 % nižší než předepsaná 95% mez spolehlivosti.

Aby bylo možné srovnat dva nezávislé cenzorované výběry z exponenciálního rozdělení, bude dále odvozen model cenzorovaného exponenciálního rozdělení pro dva výběry.

### 5.1.3 FIM pro dva výběry z cenzorovaného exponenciálního rozdělení

Nechť  $X_{j,1}, \dots, X_{j,n}$ ,  $j = 1, 2$ , jsou dva nezávislé cenzorované výběry z exponenciálního rozdělení s distribuční funkcí (5.1), hustotou (5.2) a parametry  $\lambda_1 = \lambda$  pro první výběr a  $\lambda_2 = \lambda + \alpha$  pro druhý výběr. Parametr  $\alpha$  charakterizuje rozdíl mezi rozdělením prvního výběru a rozdělením druhého výběru. Je-li  $\alpha = 0$ , rozdělení obou výběrů jsou identická. Dále  $X_{j,(1)}, \dots, X_{j,(n)}$ ,  $j = 1, 2$ , označuje uspořádaný výběr  $X_{j,1}, \dots, X_{j,n}$  a veličiny  $N_{j,i}$  jsou četnosti odpovídající četnostem  $N_i$ ,  $i = 0, 1, \dots, k$ , kde  $j$  značí číslo výběru ( $j = 1, 2$ ). Logaritmická věrohodnostní funkce sdruženého výběru  $X_{1,1}, \dots, X_{1,n}, X_{2,1}, \dots, X_{2,n}$  je pak tvaru

$$l_R(\alpha, \lambda) = l(\lambda, N_{1,0}, \dots, N_{1,k}, X_{1,(n-N_{1,0}+1)}, \dots, X_{1,(n)}) + l(\lambda + \alpha, N_{2,0}, \dots, N_{2,k}, X_{2,(n-N_{2,0}+1)}, \dots, X_{2,(n)}) \quad (5.10)$$

a MV odhady  $\hat{\alpha}, \hat{\lambda}$  parametrů  $\alpha, \lambda$  se získají maximalizací logaritmické věrohodnostní funkce (5.10).

Dále lze užitím FMI pro jednovýběrový model exponenciálního rozdělení stanovit očekávanou FIM pro dvouvýběrový model.

**Věta 5.8.** *Očekávaná FIM sdruženého mnohonásobně zleva cenzorovaného (cenzorování typu I) náhodného výběru z exponenciálního rozdělení je tvaru*

$$J_n^R(\alpha, \lambda) = \begin{pmatrix} J_{11}^R & J_{12}^R \\ J_{21}^R & J_{22}^R \end{pmatrix} = \begin{pmatrix} -E \frac{\partial^2 l_R}{\partial \alpha^2} & -E \frac{\partial^2 l_R}{\partial \alpha \partial \lambda} \\ -E \frac{\partial^2 l_R}{\partial \lambda \partial \alpha} & -E \frac{\partial^2 l_R}{\partial \lambda^2} \end{pmatrix},$$

kde

$$J_{11}^R = J_{12}^R = J_{21}^R = J_n(\lambda + \alpha, N_{2,0}, \dots, N_{2,k}, X_{2,(n-N_{2,0}+1)}, \dots, X_{2,(n)}),$$

$$J_{22}^R = J_n(\lambda, N_{1,0}, \dots, N_{1,k}, X_{1,(n-N_{1,0}+1)}, \dots, X_{1,(n)}) \\ + J_n(\lambda + \alpha, N_{2,0}, \dots, N_{2,k}, X_{2,(n-N_{2,0}+1)}, \dots, X_{2,(n)}),$$

a  $J_n$  je uvedeno ve větě 5.7

Důkaz. Zřejmé. □

### 5.1.4 Srovnání dvou nezávislých cenzorovaných výběrů z exponenciálního rozdělení

Je-li potřeba porovnat dva nezávislé zleva cenzorované výběry z exponenciálního rozdělení, lze k tomu použít asymptotické testy s rušivými parametry, které jsou popsány v odstavci 3.2. Nulová hypotéza je tvaru  $H_0 : \alpha = 0$  proti alternativě  $H_1 : \alpha \neq 0$ , tedy  $\lambda$  je rušivý parametr. Testovací statistiky mají asymptoticky  $\chi^2$  rozdělení s jedním stupněm volnosti a jsou tvaru

$$LM = \frac{U_1^2(0, \tilde{\lambda})}{J_{n,11.2}^R(0, \tilde{\lambda})}, \\ W = \hat{\alpha}^2 J_{n,11.2}^R(\hat{\alpha}, \hat{\lambda}), \\ LR = 2 \left[ l_R(\hat{\alpha}, \hat{\lambda}) - l_R(0, \tilde{\lambda}) \right], \quad (5.11)$$

kde  $J_{n,11.2}^R(\alpha, \lambda) = J_{11}^R - J_{12}^R (J_{22}^R)^{-1} J_{21}^R$  je transformace FIM  $\mathbf{J}_n^R(\alpha, \lambda)$  s dosazenými parametry odhadnutými za platnosti  $H_0$  (označeno vlnkou), případně bez dalších omezení (označeno stříškou). Funkce  $U_1(\alpha, \lambda) = \partial l_R / \partial \alpha$  je skórová funkce s dosazenými parametry odhadnutými za platnosti  $H_0$  (označeno vlnkou). Nulová hypotéza se zamítne na dané hladině významnosti v případě, že hodnota zvolené testovací statistiky (5.11) přesáhne příslušnou kritickou hodnotu  $\chi^2$  rozdělení.

## 5.2 Cenzorované Weibullovo rozdělení

Nyní bude popsán model mnohonásobně zleva cenzorovaného Weibullova rozdělení pro jeden výběr, který bude následně rozšířen na případ srovnání dvou nezávislých cenzorovaných výběrů z Weibullova rozdělení. Z toho důvodu budou v závěru odstavce odvozeny testovací statistiky pro srovnání dvou nezávislých mnohonásobně zleva cenzorovaných výběrů z Weibullova rozdělení.

### 5.2.1 Model Weibullova rozdělení pro jeden výběr

Nechť  $X_1, \dots, X_n$  je náhodný výběr z Weibullova rozdělení s parametry  $\lambda > 0$ ,  $\tau > 0$ , distribuční funkcí

$$F(x, \lambda, \tau) = \begin{cases} 1 - \exp \left[ - \left( \frac{x}{\lambda} \right)^\tau \right] & \text{pro } x \geq 0, \\ 0 & \text{pro } x < 0, \end{cases} \quad (5.12)$$

a hustotou

$$f(x, \lambda, \tau) = \begin{cases} \frac{\tau}{\lambda^\tau} x^{\tau-1} \exp \left[ - \left( \frac{x}{\lambda} \right)^\tau \right] & \text{pro } x \geq 0, \\ 0 & \text{pro } x < 0. \end{cases} \quad (5.13)$$

Věrohodnostní funkci cenzorovaného výběru lze zapsat (viz [14]) ve tvaru

$$L(\lambda, \tau, N_0, N_1, \dots, N_k, X_{(n-N_0+1)}, \dots, X_{(n)}) = \frac{n!}{N_1! \dots N_k!} \prod_{i=1}^k [F(d_i, \lambda, \tau) - F(d_{i-1}, \lambda, \tau)]^{N_i} \\ \times \prod_{i=n-N_0+1}^n f(X_{(i)}),$$

přičemž se položí  $\prod_{i=n-N_0+1}^n f(X_{(i)}) = 1$  pro  $N_0 = 0$ . Logaritmická věrohodnostní funkce je pak tvaru

$$l(\lambda, \tau, N_0, N_1, \dots, N_k, X_{(n-N_0+1)}, \dots, X_{(n)}) = \ln L = \ln \left( \frac{n!}{N_1! \dots N_k!} \right) \\ + \sum_{i=1}^k N_i \ln [F(d_i, \lambda, \tau) - F(d_{i-1}, \lambda, \tau)] \\ + \sum_{i=n-N_0+1}^n \ln [f(X_{(i)})]. \quad (5.14)$$

Jestliže  $F'_\lambda$  (respektive  $F'_\tau$ ) označuje derivaci distribuční funkce podle parametru  $\lambda$  (respektive  $\tau$ ),  $f'_\lambda$  (respektive  $f'_\tau$ ) označuje derivaci hustoty podle parametru  $\lambda$  (respektive  $\tau$ )

a

$$H_i^\lambda(\lambda, \tau) = \frac{\partial}{\partial \lambda} \ln [F(d_i, \lambda, \tau) - F(d_{i-1}, \lambda, \tau)] = \frac{F'_\lambda(d_i, \lambda, \tau) - F'_\lambda(d_{i-1}, \lambda, \tau)}{F(d_i, \lambda, \tau) - F(d_{i-1}, \lambda, \tau)} \\ = \frac{\tau \left\{ d_{i-1}^\tau \exp \left[ - \left( \frac{d_{i-1}}{\lambda} \right)^\tau \right] - d_i^\tau \exp \left[ - \left( \frac{d_i}{\lambda} \right)^\tau \right] \right\}}{\lambda^{\tau+1} \left\{ \exp \left[ - \left( \frac{d_{i-1}}{\lambda} \right)^\tau \right] - \exp \left[ - \left( \frac{d_i}{\lambda} \right)^\tau \right] \right\}}, \\ H_i^\tau(\lambda, \tau) = \frac{\partial}{\partial \tau} \ln [F(d_i, \lambda, \tau) - F(d_{i-1}, \lambda, \tau)] = \frac{F'_\tau(d_i, \lambda, \tau) - F'_\tau(d_{i-1}, \lambda, \tau)}{F(d_i, \lambda, \tau) - F(d_{i-1}, \lambda, \tau)} \\ = \frac{d_i^\tau \ln \left( \frac{d_i}{\lambda} \right) \exp \left[ - \left( \frac{d_i}{\lambda} \right)^\tau \right] - d_{i-1}^\tau \ln \left( \frac{d_{i-1}}{\lambda} \right) \exp \left[ - \left( \frac{d_{i-1}}{\lambda} \right)^\tau \right]}{\lambda^\tau \left\{ \exp \left[ - \left( \frac{d_{i-1}}{\lambda} \right)^\tau \right] - \exp \left[ - \left( \frac{d_i}{\lambda} \right)^\tau \right] \right\}}, \quad i = 1, \dots, k,$$

pak věrohodnostní rovnice pro odhad parametrů  $\lambda$  a  $\tau$  jsou tvaru

$$\frac{\partial l}{\partial \lambda} = \sum_{i=1}^k N_i H_i^\lambda(\lambda, \tau) + \sum_{i=n-N_0+1}^n \frac{f'_\lambda(X_{(i)})}{f(X_{(i)})} = 0, \\ \frac{\partial l}{\partial \tau} = \sum_{i=1}^k N_i H_i^\tau(\lambda, \tau) + \sum_{i=n-N_0+1}^n \frac{f'_\tau(X_{(i)})}{f(X_{(i)})} = 0. \quad (5.15)$$

Dosazením hustoty (5.13) do vztahu (5.15) lze pak věrohodnostní rovnice přepsat jako

$$\begin{aligned}\frac{\partial l}{\partial \lambda} &= \sum_{i=1}^k N_i H_i^\lambda(\lambda, \tau) - N_0 \frac{\tau}{\lambda} + \frac{\tau}{\lambda^{\tau+1}} \sum_{i=n-N_0+1}^n X_{(i)}^\tau = 0, \\ \frac{\partial l}{\partial \tau} &= \sum_{i=1}^k N_i H_i^\tau(\lambda, \tau) + N_0 \frac{(1 - \tau \ln \lambda)}{\tau} + \sum_{i=n-N_0+1}^n \ln X_{(i)} + \frac{\ln \lambda}{\lambda^\tau} \sum_{i=n-N_0+1}^n X_{(i)}^\tau \\ &\quad - \frac{1}{\lambda^\tau} \sum_{i=n-N_0+1}^n X_{(i)}^\tau \ln X_{(i)} = 0.\end{aligned}\quad (5.16)$$

MV odhady  $\hat{\lambda}$  a  $\hat{\tau}$  parametrů  $\lambda$  a  $\tau$  lze získat numerickým řešením rovnic (5.16). Je také možné použít Nelderův-Meadův simplexový algoritmus, který je implementován v Matlabu (verze 7.12, R2011a), a maximalizovat logaritmicou věrohodnostní funkci (5.14).

Dále bude pomocí FIM (viz kapitola 2) stanoven rozptyl odhadu parametrů  $\lambda$  a  $\tau$ . Nejprve bude odvozena výběrová FIM, která je nestranným odhadem očekávané FIM, a která bývá často upřednostňována, zejména v situacích, kdy je stanovení očekávané FIM komplikované.

**Věta 5.9.** *Výběrová FIM mnohonásobně zleva cenzorovaného (cenzorování typu I) náhodného výběru z Weibullova rozdělení s distribuční funkcí (5.12) je tvaru*

$$\tilde{\mathbf{J}}_n(\lambda, \tau) = \begin{pmatrix} \tilde{J}_{11} & \tilde{J}_{12} \\ \tilde{J}_{21} & \tilde{J}_{22} \end{pmatrix}, \quad (5.17)$$

kde

$$\begin{aligned}\tilde{J}_{11} &= - \sum_{i=1}^k N_i H_i^{\lambda\lambda}(\lambda, \tau) - \frac{N_0 \tau}{\lambda^2} + \frac{\tau^2 + \tau}{\lambda^{\tau+2}} \sum_{i=n-N_0+1}^n X_{(i)}^\tau, \\ \tilde{J}_{22} &= - \sum_{i=1}^k N_i H_i^{\tau\tau}(\lambda, \tau) + \frac{N_0}{\tau^2} + \frac{(\ln \lambda)^2}{\lambda^\tau} \sum_{i=n-N_0+1}^n X_{(i)}^\tau - \frac{2 \ln \lambda}{\lambda^\tau} \sum_{i=n-N_0+1}^n X_{(i)}^\tau \ln X_{(i)} \\ &\quad + \frac{1}{\lambda^\tau} \sum_{i=n-N_0+1}^n X_{(i)}^\tau (\ln X_{(i)})^2, \\ \tilde{J}_{12} = \tilde{J}_{21} &= - \sum_{i=1}^k N_i H_i^{\lambda\tau}(\lambda, \tau) + \frac{N_0}{\lambda} + \frac{\tau \ln \lambda - 1}{\lambda^{\tau+1}} \sum_{i=n-N_0+1}^n X_{(i)}^\tau - \frac{\tau}{\lambda^{\tau+1}} \sum_{i=n-N_0+1}^n X_{(i)}^\tau \ln X_{(i)},\end{aligned}$$

přičemž

$$\begin{aligned}H_i^{\lambda\lambda}(\lambda, \tau) &= \frac{\partial H_i^\lambda(\lambda, \tau)}{\partial \lambda} = \frac{(d_i^\tau \lambda^\tau \tau^2 + d_i^\tau \lambda^\tau \tau - d_i^{2\tau} \tau^2) \exp \left[ - \left( \frac{d_i}{\lambda} \right)^\tau \right]}{\lambda^{2\tau+2} \left\{ \exp \left[ - \left( \frac{d_{i-1}}{\lambda} \right)^\tau \right] - \exp \left[ - \left( \frac{d_i}{\lambda} \right)^\tau \right] \right\}} \\ &\quad - \frac{(d_{i-1}^\tau \lambda^\tau \tau^2 + d_{i-1}^\tau \lambda^\tau \tau - d_{i-1}^{2\tau} \tau^2) \exp \left[ - \left( \frac{d_{i-1}}{\lambda} \right)^\tau \right]}{\lambda^{2\tau+2} \left\{ \exp \left[ - \left( \frac{d_{i-1}}{\lambda} \right)^\tau \right] - \exp \left[ - \left( \frac{d_i}{\lambda} \right)^\tau \right] \right\}}\end{aligned}$$



$$\begin{aligned}
& - \frac{\tau^2 \left\{ d_{i-1}^\tau \exp \left[ - \left( \frac{d_{i-1}}{\lambda} \right)^\tau \right] - d_i^\tau \exp \left[ - \left( \frac{d_i}{\lambda} \right)^\tau \right] \right\}^2}{\lambda^{2\tau+2} \left\{ \exp \left[ - \left( \frac{d_{i-1}}{\lambda} \right)^\tau \right] - \exp \left[ - \left( \frac{d_i}{\lambda} \right)^\tau \right] \right\}^2}, \\
H_i^{\tau\tau}(\lambda, \tau) &= \frac{\partial H_i^\tau(\lambda, \tau)}{\partial \tau} = \frac{(d_i^\tau \lambda^\tau - d_i^{2\tau}) \left[ \ln \left( \frac{d_i}{\lambda} \right) \right]^2 \exp \left[ - \left( \frac{d_i}{\lambda} \right)^\tau \right]}{\lambda^{2\tau} \left\{ \exp \left[ - \left( \frac{d_{i-1}}{\lambda} \right)^\tau \right] - \exp \left[ - \left( \frac{d_i}{\lambda} \right)^\tau \right] \right\}} \\
& - \frac{(d_{i-1}^\tau \lambda^\tau - d_{i-1}^{2\tau}) \left[ \ln \left( \frac{d_{i-1}}{\lambda} \right) \right]^2 \exp \left[ - \left( \frac{d_{i-1}}{\lambda} \right)^\tau \right]}{\lambda^{2\tau} \left\{ \exp \left[ - \left( \frac{d_{i-1}}{\lambda} \right)^\tau \right] - \exp \left[ - \left( \frac{d_i}{\lambda} \right)^\tau \right] \right\}} \\
& - \frac{\left\{ d_{i-1}^\tau \ln \left( \frac{d_{i-1}}{\lambda} \right) \exp \left[ - \left( \frac{d_{i-1}}{\lambda} \right)^\tau \right] - d_i^\tau \ln \left( \frac{d_i}{\lambda} \right) \exp \left[ - \left( \frac{d_i}{\lambda} \right)^\tau \right] \right\}^2}{\lambda^{2\tau} \left\{ \exp \left[ - \left( \frac{d_{i-1}}{\lambda} \right)^\tau \right] - \exp \left[ - \left( \frac{d_i}{\lambda} \right)^\tau \right] \right\}^2}, \\
H_i^{\lambda\tau}(\lambda, \tau) &= \frac{\partial H_i^\lambda(\lambda, \tau)}{\partial \tau} = \frac{\left[ d_i^{2\tau} \tau \ln \left( \frac{d_i}{\lambda} \right) - d_i^\tau \lambda^\tau \tau \ln \left( \frac{d_i}{\lambda} \right) - d_i^\tau \lambda^\tau \right] \exp \left[ - \left( \frac{d_i}{\lambda} \right)^\tau \right]}{\lambda^{2\tau+1} \left\{ \exp \left[ - \left( \frac{d_{i-1}}{\lambda} \right)^\tau \right] - \exp \left[ - \left( \frac{d_i}{\lambda} \right)^\tau \right] \right\}} \\
& - \frac{\left[ d_{i-1}^{2\tau} \tau \ln \left( \frac{d_{i-1}}{\lambda} \right) - d_{i-1}^\tau \lambda^\tau \tau \ln \left( \frac{d_{i-1}}{\lambda} \right) - d_{i-1}^\tau \lambda^\tau \right] \exp \left[ - \left( \frac{d_{i-1}}{\lambda} \right)^\tau \right]}{\lambda^{2\tau+1} \left\{ \exp \left[ - \left( \frac{d_{i-1}}{\lambda} \right)^\tau \right] - \exp \left[ - \left( \frac{d_i}{\lambda} \right)^\tau \right] \right\}} \\
& + \frac{\tau \left\{ d_{i-1}^\tau \exp \left[ - \left( \frac{d_{i-1}}{\lambda} \right)^\tau \right] - d_i^\tau \exp \left[ - \left( \frac{d_i}{\lambda} \right)^\tau \right] \right\}}{\lambda^{2\tau+1} \left\{ \exp \left[ - \left( \frac{d_{i-1}}{\lambda} \right)^\tau \right] - \exp \left[ - \left( \frac{d_i}{\lambda} \right)^\tau \right] \right\}^2} \\
& \times \frac{\left\{ d_{i-1}^\tau \ln \left( \frac{d_{i-1}}{\lambda} \right) \exp \left[ - \left( \frac{d_{i-1}}{\lambda} \right)^\tau \right] - d_i^\tau \ln \left( \frac{d_i}{\lambda} \right) \exp \left[ - \left( \frac{d_i}{\lambda} \right)^\tau \right] \right\}}{\lambda^{2\tau+1} \left\{ \exp \left[ - \left( \frac{d_{i-1}}{\lambda} \right)^\tau \right] - \exp \left[ - \left( \frac{d_i}{\lambda} \right)^\tau \right] \right\}^2}, \quad i = 1, \dots, k.
\end{aligned}$$

*Důkaz.* Jak bylo uvedeno v kapitole 2, výběrová FIM je definována jako

$$\tilde{\mathbf{J}}_n(\lambda, \tau) = \begin{pmatrix} -\frac{\partial^2 l}{\partial \lambda^2} & -\frac{\partial^2 l}{\partial \lambda \partial \tau} \\ -\frac{\partial^2 l}{\partial \tau \partial \lambda} & -\frac{\partial^2 l}{\partial \tau^2} \end{pmatrix}.$$

Užitím vztahu (5.16) lze pak snadno stanovit druhé derivace logaritmičké věrohodnostní funkce podle parametrů  $\lambda$  a  $\tau$ , což dává tvrzení věty. □

Další pozornost bude věnována stanovení očekávané FIM. K tomu bude potřeba několik pomocných tvrzení.

**Lemma 5.10.** *Nechť  $X_1, \dots, X_n$  je náhodný výběr z Weibullova rozdělení. Pak hustota pořadové veličiny  $X_{(i)}$  je tvaru*

$$f_{(i)}(x) = n \binom{n-1}{i-1} \frac{\tau}{\lambda^\tau} x^{\tau-1} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} \exp \left[ - \left( \frac{x}{\lambda} \right)^\tau (n-i+j+1) \right], \quad i = 1, \dots, n.$$

*Důkaz.* Hustota veličiny  $X_{(i)}$  je obecně tvaru (viz [52])

$$f_{(i)}(x) = n \binom{n-1}{i-1} f(x) [F(x)]^{i-1} [1-F(x)]^{n-i}, \quad i = 1, 2, \dots, n.$$

Místo  $F$  se dosadí distribuční funkce (5.12) a místo  $f$  hustota (5.13) necenzorovaného Weibullova rozdělení a postupnými úpravami lze dostat hustotu  $f_{(i)}(x)$  ve tvaru

$$\begin{aligned} f_{(i)}(x) &= n \binom{n-1}{i-1} \frac{\tau}{\lambda^\tau} x^{\tau-1} \exp \left[ - \left( \frac{x}{\lambda} \right)^\tau \right] \left\{ 1 - \exp \left[ - \left( \frac{x}{\lambda} \right)^\tau \right] \right\}^{i-1} \exp \left[ - \left( \frac{x}{\lambda} \right)^\tau (n-i) \right] \\ &= n \binom{n-1}{i-1} \frac{\tau}{\lambda^\tau} x^{\tau-1} \exp \left[ - \left( \frac{x}{\lambda} \right)^\tau (n-i+1) \right] \sum_{j=0}^{i-1} \binom{i-1}{j} (-1)^j \exp \left[ -j \left( \frac{x}{\lambda} \right)^\tau \right] \\ &= n \binom{n-1}{i-1} \frac{\tau}{\lambda^\tau} x^{\tau-1} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} \exp \left[ - \left( \frac{x}{\lambda} \right)^\tau (n-i+j+1) \right]. \end{aligned}$$

□

**Lemma 5.11.** *Nechť  $X_1, \dots, X_n$  je náhodný výběr z Weibullova rozdělení. Pak střední hodnota  $E(X_{(i)}^\tau)$  transformované náhodné veličiny  $X_{(i)}^\tau$  je tvaru*

$$E(X_{(i)}^\tau) = n \binom{n-1}{i-1} \lambda^\tau \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} (n-i+j+1)^{-2}, \quad i = 1, \dots, n.$$

*Důkaz.* Postupnými úpravami s využitím lemmatu 5.10 a faktu, že  $\int_0^\infty te^{-t} dt = \Gamma(2) = 1$ , lze odvodit střední hodnotu  $E(X_{(i)}^\tau)$  ve tvaru

$$\begin{aligned} E(X_{(i)}^\tau) &= \int_0^\infty x^\tau f_{(i)}(x) dx \\ &= n \binom{n-1}{i-1} \frac{\tau}{\lambda^\tau} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} \int_0^\infty x^{2\tau-1} \exp \left[ -x^\tau \frac{(n-i+j+1)}{\lambda^\tau} \right] dx \\ &= n \binom{n-1}{i-1} \lambda^\tau \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} (n-i+j+1)^{-2} \int_0^\infty te^{-t} dt \\ &= n \binom{n-1}{i-1} \lambda^\tau \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} (n-i+j+1)^{-2}. \end{aligned}$$

□

**Lemma 5.12.** *Nechť  $X_1, \dots, X_n$  je náhodný výběr z Weibullova rozdělení. Pak střední hodnota  $E(X_{(i)}^\tau \ln X_{(i)})$  transformované náhodné veličiny  $X_{(i)}^\tau \ln X_{(i)}$  je tvaru*

$$\begin{aligned} E(X_{(i)}^\tau \ln X_{(i)}) &= n \binom{n-1}{i-1} \frac{\lambda^\tau}{\tau} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} (n-i+j+1)^{-2} \\ &\quad \times \left[ \ln \left( \frac{\lambda^\tau}{n-i+j+1} \right) + 1 - \gamma_e \right], \quad i = 1, \dots, n, \end{aligned}$$

kde  $\gamma_e \doteq 0,57722$  je Eulerova konstanta.

*Důkaz.* Postupnými úpravami s využitím lemmatu 5.10 a faktů, že  $\int_0^\infty te^{-t}dt = \Gamma(2) = 1$  a  $\int_0^\infty t \ln t e^{-t}dt = \Gamma'(2) = 1 - \gamma_e$ , lze odvodit střední hodnotu  $E\left(X_{(i)}^\tau \ln X_{(i)}\right)$  ve tvaru

$$\begin{aligned}
E(X_{(i)}^\tau \ln X_{(i)}) &= \int_0^\infty x^\tau \ln x f_{(i)}(x) dx = n \binom{n-1}{i-1} \frac{\tau}{\lambda^\tau} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} \\
&\quad \times \int_0^\infty x^{2\tau-1} \ln x \exp\left[-x^\tau \frac{(n-i+j+1)}{\lambda^\tau}\right] dx \\
&= n \binom{n-1}{i-1} \frac{\tau}{\lambda^\tau} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} \frac{1}{\tau} \\
&\quad \times \int_0^\infty x^{2\tau-1} \ln x^\tau \exp\left[-x^\tau \frac{(n-i+j+1)}{\lambda^\tau}\right] dx \\
&= n \binom{n-1}{i-1} \frac{\lambda^\tau}{\tau} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} (n-i+j+1)^{-2} \\
&\quad \times \int_0^\infty t \ln\left(\frac{\lambda^\tau}{n-i+j+1} t\right) e^{-t} dt \\
&= n \binom{n-1}{i-1} \frac{\lambda^\tau}{\tau} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} (n-i+j+1)^{-2} \\
&\quad \times \left[ \ln\left(\frac{\lambda^\tau}{n-i+j+1}\right) \int_0^\infty te^{-t} dt + \int_0^\infty t \ln t e^{-t} dt \right] \\
&= n \binom{n-1}{i-1} \frac{\lambda^\tau}{\tau} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} (n-i+j+1)^{-2} \\
&\quad \times \left[ \ln\left(\frac{\lambda^\tau}{n-i+j+1}\right) + 1 - \gamma_e \right].
\end{aligned}$$

□

**Lemma 5.13.** *Nechť  $X_1, \dots, X_n$  je náhodný výběr z Weibullova rozdělení. Pak střední hodnota  $E\left[X_{(i)}^\tau (\ln X_{(i)})^2\right]$  transformované náhodné veličiny  $X_{(i)}^\tau (\ln X_{(i)})^2$  je tvaru*

$$\begin{aligned}
E\left[X_{(i)}^\tau (\ln X_{(i)})^2\right] &= n \binom{n-1}{i-1} \frac{\lambda^\tau}{\tau^2} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} (n-i+j+1)^{-2} \\
&\quad \times \left\{ \left[ \ln\left(\frac{\lambda^\tau}{n-i+j+1}\right) \right]^2 + 2 \ln\left(\frac{\lambda^\tau}{n-i+j+1}\right) (1 - \gamma_e) \right. \\
&\quad \left. + \frac{\pi^2}{6} - 2\gamma_e + \gamma_e^2 \right\}, \quad i = 1, \dots, n,
\end{aligned}$$

kde  $\gamma_e \doteq 0,57722$  je Eulerova konstanta.

*Důkaz.* Postupnými úpravami s využitím lemmatu 5.10 a faktů, že  $\int_0^\infty te^{-t}dt = \Gamma(2) = 1$ ,  $\int_0^\infty t \ln t e^{-t}dt = \Gamma'(2) = 1 - \gamma_e$  a  $\int_0^\infty t (\ln t)^2 e^{-t}dt = \Gamma''(2) = \frac{\pi^2}{6} - 2\gamma_e + \gamma_e^2$ , lze odvodit

střední hodnotu  $E \left[ X_{(i)}^\tau (\ln X_{(i)})^2 \right]$  ve tvaru

$$\begin{aligned}
E \left[ X_{(i)}^\tau (\ln X_{(i)})^2 \right] &= \int_0^\infty x^\tau (\ln x)^2 f_{(i)}(x) dx = n \binom{n-1}{i-1} \frac{\tau}{\lambda^\tau} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} \\
&\quad \times \int_0^\infty x^{2\tau-1} (\ln x)^2 \exp \left[ -x^\tau \frac{(n-i+j+1)}{\lambda^\tau} \right] dx \\
&= n \binom{n-1}{i-1} \frac{\tau}{\lambda^\tau} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} \frac{1}{\tau^2} \\
&\quad \times \int_0^\infty x^{2\tau-1} (\ln x^\tau)^2 \exp \left[ -x^\tau \frac{(n-i+j+1)}{\lambda^\tau} \right] dx \\
&= n \binom{n-1}{i-1} \frac{\lambda^\tau}{\tau^2} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} (n-i+j+1)^{-2} \\
&\quad \times \int_0^\infty t \left[ \ln \left( \frac{\lambda^\tau}{n-i+j+1} t \right) \right]^2 e^{-t} dt \\
&= n \binom{n-1}{i-1} \frac{\lambda^\tau}{\tau^2} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} (n-i+j+1)^{-2} \\
&\quad \times \int_0^\infty t \left[ \ln \left( \frac{\lambda^\tau}{n-i+j+1} \right) + \ln t \right]^2 e^{-t} dt \\
&= n \binom{n-1}{i-1} \frac{\lambda^\tau}{\tau^2} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} (n-i+j+1)^{-2} \\
&\quad \times \left\{ \left[ \ln \left( \frac{\lambda^\tau}{n-i+j+1} \right) \right]^2 \int_0^\infty t e^{-t} dt \right. \\
&\quad \left. + 2 \ln \left( \frac{\lambda^\tau}{n-i+j+1} \right) \int_0^\infty t \ln t e^{-t} dt + \int_0^\infty t (\ln t)^2 e^{-t} dt \right\} \\
&= n \binom{n-1}{i-1} \frac{\lambda^\tau}{\tau^2} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} (n-i+j+1)^{-2} \\
&\quad \times \left\{ \left[ \ln \left( \frac{\lambda^\tau}{n-i+j+1} \right) \right]^2 + 2 \ln \left( \frac{\lambda^\tau}{n-i+j+1} \right) (1 - \gamma_e) \right. \\
&\quad \left. + \frac{\pi^2}{6} - 2\gamma_e + \gamma_e^2 \right\}.
\end{aligned}$$

□

**Lemma 5.14.** *Nechť  $X_1, \dots, X_n$  je náhodný výběr z Weibullova rozdělení. Pak střední hodnota součtu transformovaných náhodných veličin  $E \left( \sum_{i=n-N_0+1}^n X_{(i)}^\tau \right)$  je tvaru*

$$\begin{aligned}
E \left( \sum_{i=n-N_0+1}^n X_{(i)}^\tau \right) &= \sum_{n_0=0}^n \left[ \sum_{i=n-n_0+1}^n E \left( X_{(i)}^\tau \right) \right] \binom{n}{n_0} \exp \left[ -n_0 \left( \frac{d_k}{\lambda} \right)^\tau \right] \\
&\quad \times \left\{ 1 - \exp \left[ - \left( \frac{d_k}{\lambda} \right)^\tau \right] \right\}^{n-n_0}.
\end{aligned}$$

*Důkaz.* Podobným způsobem jako v důkazu lemmatu 5.6. □

**Lemma 5.15.** *Nechť  $X_1, \dots, X_n$  je náhodný výběr z Weibullova rozdělení. Pak střední hodnota součtu transformovaných náhodných veličin  $E \left( \sum_{i=n-n_0+1}^n X_{(i)}^\tau \ln X_{(i)} \right)$  je tvaru*

$$E \left( \sum_{i=n-n_0+1}^n X_{(i)}^\tau \ln X_{(i)} \right) = \sum_{n_0=0}^n \left[ \sum_{i=n-n_0+1}^n E \left( X_{(i)}^\tau \ln X_{(i)} \right) \right] \\ \times \binom{n}{n_0} \exp \left[ -n_0 \left( \frac{d_k}{\lambda} \right)^\tau \right] \left\{ 1 - \exp \left[ - \left( \frac{d_k}{\lambda} \right)^\tau \right] \right\}^{n-n_0}.$$

*Důkaz.* Podobným způsobem jako v důkazu lemmatu 5.6. □

**Lemma 5.16.** *Nechť  $X_1, \dots, X_n$  je náhodný výběr z Weibullova rozdělení. Pak střední hodnota součtu transformovaných náhodných veličin  $E \left[ \sum_{i=n-n_0+1}^n X_{(i)}^\tau (\ln X_{(i)})^2 \right]$  je tvaru*

$$E \left[ \sum_{i=n-n_0+1}^n X_{(i)}^\tau (\ln X_{(i)})^2 \right] = \sum_{n_0=0}^n \left\{ \sum_{i=n-n_0+1}^n E \left[ X_{(i)}^\tau (\ln X_{(i)})^2 \right] \right\} \\ \times \binom{n}{n_0} \exp \left[ -n_0 \left( \frac{d_k}{\lambda} \right)^\tau \right] \left\{ 1 - \exp \left[ - \left( \frac{d_k}{\lambda} \right)^\tau \right] \right\}^{n-n_0}.$$

*Důkaz.* Podobným způsobem jako v důkazu lemmatu 5.6. □

**Věta 5.17.** *Očekávaná FIM mnohonásobně zleva cenzorovaného (cenzorování typu I) náhodného výběru z Weibullova rozdělení s distribuční funkcí (5.12) je tvaru*

$$\mathbf{J}_n(\lambda, \tau) = n\mathbf{J}(\lambda, \tau) = n \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix},$$

kde

$$J_{11} = \sum_{i=1}^k \frac{\tau^2 \left\{ d_{i-1}^\tau \exp \left[ - \left( \frac{d_{i-1}}{\lambda} \right)^\tau \right] - d_i^\tau \exp \left[ - \left( \frac{d_i}{\lambda} \right)^\tau \right] \right\}^2}{\lambda^{2\tau+2} \left\{ \exp \left[ - \left( \frac{d_{i-1}}{\lambda} \right)^\tau \right] - \exp \left[ - \left( \frac{d_i}{\lambda} \right)^\tau \right] \right\}} \\ - \frac{(d_k^\tau \lambda^\tau \tau^2 + d_k^\tau \lambda^\tau \tau - d_k^{2\tau} \tau^2) \exp \left[ - \left( \frac{d_k}{\lambda} \right)^\tau \right] - \frac{\tau}{\lambda^2} \exp \left[ - \left( \frac{d_k}{\lambda} \right)^\tau \right]}{\lambda^{2\tau+2}} \\ + \frac{\tau^2 + \tau}{\lambda^2} \sum_{n_0=0}^n \sum_{i=n-n_0+1}^n \binom{n-1}{i-1} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} (n-i+j+1)^{-2} \\ \times \binom{n}{n_0} \exp \left[ -n_0 \left( \frac{d_k}{\lambda} \right)^\tau \right] \left\{ 1 - \exp \left[ - \left( \frac{d_k}{\lambda} \right)^\tau \right] \right\}^{n-n_0}, \\ J_{22} = \sum_{i=1}^k \frac{\left\{ d_{i-1}^\tau \ln \left( \frac{d_{i-1}}{\lambda} \right) \exp \left[ - \left( \frac{d_{i-1}}{\lambda} \right)^\tau \right] - d_{i-1}^\tau \ln \left( \frac{d_i}{\lambda} \right) \exp \left[ - \left( \frac{d_i}{\lambda} \right)^\tau \right] \right\}^2}{\lambda^{2\tau} \left\{ \exp \left[ - \left( \frac{d_{i-1}}{\lambda} \right)^\tau \right] - \exp \left[ - \left( \frac{d_i}{\lambda} \right)^\tau \right] \right\}}$$

$$\begin{aligned}
& - \frac{(d_k^\tau \lambda^\tau - d_k^{2\tau}) \left[ \ln \left( \frac{d_k}{\lambda} \right) \right]^2 \exp \left[ - \left( \frac{d_k}{\lambda} \right)^\tau \right] + \frac{1}{\tau^2} \exp \left[ - \left( \frac{d_k}{\lambda} \right)^\tau \right]}{\lambda^{2\tau}} \\
& + (\ln \lambda)^2 \sum_{n_0=0}^n \sum_{i=n-n_0+1}^n \binom{n-1}{i-1} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} (n-i+j+1)^{-2} \\
& \times \binom{n}{n_0} \exp \left[ -n_0 \left( \frac{d_k}{\lambda} \right)^\tau \right] \left\{ 1 - \exp \left[ - \left( \frac{d_k}{\lambda} \right)^\tau \right] \right\}^{n-n_0} \\
& - \frac{2 \ln \lambda}{\tau} \sum_{n_0=0}^n \sum_{i=n-n_0+1}^n \binom{n-1}{i-1} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} (n-i+j+1)^{-2} \\
& \times \left[ \ln \left( \frac{\lambda^\tau}{n-i+j+1} \right) + 1 - \gamma_e \right] \\
& \times \binom{n}{n_0} \exp \left[ -n_0 \left( \frac{d_k}{\lambda} \right)^\tau \right] \left\{ 1 - \exp \left[ - \left( \frac{d_k}{\lambda} \right)^\tau \right] \right\}^{n-n_0} \\
& + \frac{1}{\tau^2} \sum_{n_0=0}^n \sum_{i=n-n_0+1}^n \binom{n-1}{i-1} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} (n-i+j+1)^{-2} \\
& \times \left\{ \left[ \ln \left( \frac{\lambda^\tau}{n-i+j+1} \right) \right]^2 + 2 \ln \left( \frac{\lambda^\tau}{n-i+j+1} \right) (1 - \gamma_e) \right. \\
& \left. + \frac{\pi^2}{6} - 2\gamma_e + \gamma_e^2 \right\} \binom{n}{n_0} \exp \left[ -n_0 \left( \frac{d_k}{\lambda} \right)^\tau \right] \left\{ 1 - \exp \left[ - \left( \frac{d_k}{\lambda} \right)^\tau \right] \right\}^{n-n_0}, \\
J_{12} = J_{21} = & - \sum_{i=1}^k \frac{\tau \left\{ d_{i-1}^\tau \exp \left[ - \left( \frac{d_{i-1}}{\lambda} \right)^\tau \right] - d_i^\tau \exp \left[ - \left( \frac{d_i}{\lambda} \right)^\tau \right] \right\}}{\lambda^{2\tau+1} \left\{ \exp \left[ - \left( \frac{d_{i-1}}{\lambda} \right)^\tau \right] - \exp \left[ - \left( \frac{d_i}{\lambda} \right)^\tau \right] \right\}} \\
& \times \left\{ d_{i-1}^\tau \ln \left( \frac{d_{i-1}}{\lambda} \right) \exp \left[ - \left( \frac{d_{i-1}}{\lambda} \right)^\tau \right] - d_i^\tau \ln \left( \frac{d_i}{\lambda} \right) \exp \left[ - \left( \frac{d_i}{\lambda} \right)^\tau \right] \right\} \\
& + \frac{[d_k^\tau \tau \lambda^\tau \ln \left( \frac{d_k}{\lambda} \right) + d_k^\tau \lambda^\tau - d_k^{2\tau} \tau \ln \left( \frac{d_k}{\lambda} \right)] \exp \left[ - \left( \frac{d_k}{\lambda} \right)^\tau \right] + \frac{1}{\lambda} \exp \left[ - \left( \frac{d_k}{\lambda} \right)^\tau \right]}{\lambda^{2\tau+1}} \\
& + \frac{\tau \ln \lambda - 1}{\lambda} \sum_{n_0=0}^n \sum_{i=n-n_0+1}^n \binom{n-1}{i-1} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} (n-i+j+1)^{-2} \\
& \times \binom{n}{n_0} \exp \left[ -n_0 \left( \frac{d_k}{\lambda} \right)^\tau \right] \left\{ 1 - \exp \left[ - \left( \frac{d_k}{\lambda} \right)^\tau \right] \right\}^{n-n_0} \\
& - \frac{1}{\lambda} \sum_{n_0=0}^n \sum_{i=n-n_0+1}^n \binom{n-1}{i-1} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} (n-i+j+1)^{-2} \\
& \times \left[ \ln \left( \frac{\lambda^\tau}{n-i+j+1} \right) + 1 - \gamma_e \right] \\
& \times \binom{n}{n_0} \exp \left[ -n_0 \left( \frac{d_k}{\lambda} \right)^\tau \right] \left\{ 1 - \exp \left[ - \left( \frac{d_k}{\lambda} \right)^\tau \right] \right\}^{n-n_0}.
\end{aligned}$$

*Důkaz.* Z poznámky 2.5 plyne, že

$$\mathbf{J}_n(\lambda, \tau) = \mathbf{E} \tilde{\mathbf{J}}_n(\lambda, \tau). \tag{5.18}$$

Dosazením  $\tilde{\mathbf{J}}_n$  ze vztahu (5.17) do (5.18) dostáváme

$$\mathbf{J}_n(\lambda, \tau) = n\mathbf{J}(\lambda, \tau) = n \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix},$$

kde

$$\begin{aligned} nJ_{11} &= - \sum_{i=1}^k H_i^{\lambda\lambda}(\lambda, \tau) \mathbb{E}(N_i) - \frac{\mathbb{E}(N_0)\tau}{\lambda^2} + \frac{\tau^2 + \tau}{\lambda^{\tau+2}} \mathbb{E} \left( \sum_{i=n-N_0+1}^n X_{(i)}^\tau \right), \\ nJ_{22} &= - \sum_{i=1}^k H_i^{\tau\tau}(\lambda, \tau) \mathbb{E}(N_i) + \frac{\mathbb{E}(N_0)}{\tau^2} + \frac{(\ln \lambda)^2}{\lambda^\tau} \mathbb{E} \left( \sum_{i=n-N_0+1}^n X_{(i)}^\tau \right) \\ &\quad - \frac{2 \ln \lambda}{\lambda^\tau} \mathbb{E} \left( \sum_{i=n-N_0+1}^n X_{(i)}^\tau \ln X_{(i)} \right) + \frac{1}{\lambda^\tau} \mathbb{E} \left[ \sum_{i=n-N_0+1}^n X_{(i)}^\tau (\ln X_{(i)})^2 \right], \quad (5.19) \\ nJ_{12} = nJ_{21} &= - \sum_{i=1}^k H_i^{\lambda\tau}(\lambda, \tau) \mathbb{E}(N_i) + \frac{\mathbb{E}(N_0)}{\lambda} + \frac{\tau \ln \lambda - 1}{\lambda^{\tau+1}} \mathbb{E} \left( \sum_{i=n-N_0+1}^n X_{(i)}^\tau \right) \\ &\quad - \frac{\tau}{\lambda^{\tau+1}} \mathbb{E} \left( \sum_{i=n-N_0+1}^n X_{(i)}^\tau \ln X_{(i)} \right), \end{aligned}$$

přičemž přesné vyjádření  $H_i^{\lambda\lambda}$ ,  $H_i^{\tau\tau}$  a  $H_i^{\lambda\tau}$  lze najít ve větě 5.9. Z lemmatu 5.2 plyne, že  $N_i \sim \text{Bi}(n, \theta_i)$ ,  $i = 0, \dots, k$ , tedy očekávané četnosti jsou tvaru

$$\mathbb{E}(N_i) = n\theta_i = \begin{cases} n \left\{ \exp \left[ - \left( \frac{d_{i-1}}{\lambda} \right)^\tau \right] - \exp \left[ - \left( \frac{d_i}{\lambda} \right)^\tau \right] \right\} & \text{pro } i = 1, \dots, k, \\ n \exp \left[ - \left( \frac{d_k}{\lambda} \right)^\tau \right] & \text{pro } i = 0. \end{cases}$$

Dosadíme-li výsledky z lemmat 5.11–5.13 do lemmat 5.14–5.16, dostaneme

$$\begin{aligned} \mathbb{E} \left( \sum_{i=n-N_0+1}^n X_{(i)}^\tau \right) &= n\lambda^\tau \sum_{n_0=0}^n \sum_{i=n-n_0+1}^n \binom{n-1}{i-1} \sum_{j=0}^{i-1} (-1)^j \\ &\quad \times \binom{i-1}{j} (n-i+j+1)^{-2} \\ &\quad \times \binom{n}{n_0} \exp \left[ -n_0 \left( \frac{d_k}{\lambda} \right)^\tau \right] \left\{ 1 - \exp \left[ - \left( \frac{d_k}{\lambda} \right)^\tau \right] \right\}^{n-n_0}, \\ \mathbb{E} \left( \sum_{i=n-n_0+1}^n X_{(i)}^\tau \ln X_{(i)} \right) &= n \frac{\lambda^\tau}{\tau} \sum_{n_0=0}^n \sum_{i=n-n_0+1}^n \binom{n-1}{i-1} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} \\ &\quad \times (n-i+j+1)^{-2} \left[ \ln \left( \frac{\lambda^\tau}{n-i+j+1} \right) + 1 - \gamma_e \right] \\ &\quad \times \binom{n}{n_0} \exp \left[ -n_0 \left( \frac{d_k}{\lambda} \right)^\tau \right] \left\{ 1 - \exp \left[ - \left( \frac{d_k}{\lambda} \right)^\tau \right] \right\}^{n-n_0}, \\ \mathbb{E} \left\{ \sum_{i=n-n_0+1}^n X_{(i)}^\tau [\ln X_{(i)}]^2 \right\} &= n \frac{\lambda^\tau}{\tau^2} \sum_{n_0=0}^n \sum_{i=n-n_0+1}^n \binom{n-1}{i-1} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} \\ &\quad \times (n-i+j+1)^{-2} \left\{ \left[ \ln \left( \frac{\lambda^\tau}{n-i+j+1} \right) \right]^2 \right. \end{aligned}$$

$$\begin{aligned}
& + 2 \ln \left( \frac{\lambda^\tau}{n - i + j + 1} \right) \left( 1 - \gamma_e \right) + \frac{\pi^2}{6} - 2\gamma_e + \gamma_e^2 \Big\} \\
& \times \binom{n}{n_0} \exp \left[ -n_0 \left( \frac{d_k}{\lambda} \right)^\tau \right] \left\{ 1 - \exp \left[ - \left( \frac{d_k}{\lambda} \right)^\tau \right] \right\}^{n-n_0},
\end{aligned}$$

což po dosazení do vztahů (5.19) dává tvrzení věty.  $\square$

Vzhledem k asymptotickým vlastnostem MV odhadu  $\hat{\lambda}$  (respektive  $\hat{\tau}$ ) má (viz věta 2.11)  $\sqrt{n}(\hat{\lambda} - \lambda)$  (respektive  $\sqrt{n}(\hat{\tau} - \tau)$ ) asymptoticky normální rozdělení  $N(0; \sigma^2(\lambda))$  (respektive  $N(0; \sigma^2(\tau))$ ), kde  $\sigma^2(\lambda)$ ,  $\sigma^2(\tau)$  odpovídají diagonálním prvkům varianční matice  $\mathbf{V}(\lambda, \tau) = \mathbf{J}^{-1}(\lambda, \tau)$ .

Aby bylo možné srovnat dva nezávislé cenzorované výběry z Weibullova rozdělení bude dále odvozen model cenzorovaného Weibullova rozdělení pro dva výběry.

### 5.2.2 FIM pro dva výběry z cenzorovaného Weibullova rozdělení

Nechť  $X_{j,1}, \dots, X_{j,n}$ ,  $j = 1, 2$ , jsou dva nezávislé cenzorované výběry z Weibullova rozdělení s distribuční funkcí (5.12), hustotou (5.13) a parametry  $\lambda_1 = \lambda$ ,  $\tau_1 = \tau$  pro první výběr a  $\lambda_2 = \lambda + \alpha$ ,  $\tau_2 = \tau + \beta$  pro druhý výběr. Parametry  $\alpha$  a  $\beta$  charakterizují rozdíl mezi rozděleními prvního výběru a rozděleními druhého výběru. Je-li  $\alpha = 0$  a  $\beta = 0$ , rozdělení obou výběrů jsou identická. Dále  $X_{j,(1)}, \dots, X_{j,(n)}$ ,  $j = 1, 2$ , označuje uspořádaný výběr  $X_{j,1}, \dots, X_{j,n}$  a veličiny  $N_{j,i}$  jsou četnosti odpovídající četnostem  $N_i$ ,  $i = 0, 1, \dots, k$ , kde  $j$  značí číslo výběru ( $j = 1, 2$ ). Logaritmická věrohodnostní funkce sdruženého výběru  $X_{1,1}, \dots, X_{1,n}, X_{2,1}, \dots, X_{2,n}$  je pak tvaru

$$\begin{aligned}
l_{\mathbf{R}}(\alpha, \beta, \lambda, \tau) &= l(\lambda, \tau, N_{1,0}, \dots, N_{1,k}, X_{1,(n-N_{1,0}+1)}, \dots, X_{1,(n)}) \\
&+ l(\lambda + \alpha, \tau + \beta, N_{2,0}, \dots, N_{2,k}, X_{2,(n-N_{2,0}+1)}, \dots, X_{2,(n)})
\end{aligned} \tag{5.20}$$

a MV odhady  $\hat{\alpha}, \hat{\beta}, \hat{\lambda}, \hat{\tau}$  parametrů  $\alpha, \beta, \lambda, \tau$  se získají maximalizací logaritmické věrohodnostní funkce (5.20).

Dále lze užitím FIM pro jednovýběrový model Weibullova rozdělení stanovit očekávanou FIM pro dvouvýběrový model.

**Věta 5.18.** *Očekávaná FIM sdruženého mnohonásobně zleva cenzorovaného (cenzorování typu I) náhodného výběru z Weibullova rozdělení je tvaru*

$$\mathbf{J}_n^{\mathbf{R}}(\alpha, \beta, \lambda, \tau) = \begin{pmatrix} J_{11}^{\mathbf{R}} & J_{12}^{\mathbf{R}} & J_{13}^{\mathbf{R}} & J_{14}^{\mathbf{R}} \\ J_{21}^{\mathbf{R}} & J_{22}^{\mathbf{R}} & J_{23}^{\mathbf{R}} & J_{24}^{\mathbf{R}} \\ J_{31}^{\mathbf{R}} & J_{32}^{\mathbf{R}} & J_{33}^{\mathbf{R}} & J_{34}^{\mathbf{R}} \\ J_{41}^{\mathbf{R}} & J_{42}^{\mathbf{R}} & J_{43}^{\mathbf{R}} & J_{44}^{\mathbf{R}} \end{pmatrix} = \begin{pmatrix} -\mathbf{E} \frac{\partial^2 l_{\mathbf{R}}}{\partial \alpha^2} & -\mathbf{E} \frac{\partial^2 l_{\mathbf{R}}}{\partial \alpha \partial \beta} & -\mathbf{E} \frac{\partial^2 l_{\mathbf{R}}}{\partial \alpha \partial \lambda} & -\mathbf{E} \frac{\partial^2 l_{\mathbf{R}}}{\partial \alpha \partial \tau} \\ -\mathbf{E} \frac{\partial^2 l_{\mathbf{R}}}{\partial \beta \partial \alpha} & -\mathbf{E} \frac{\partial^2 l_{\mathbf{R}}}{\partial \beta^2} & -\mathbf{E} \frac{\partial^2 l_{\mathbf{R}}}{\partial \beta \partial \lambda} & -\mathbf{E} \frac{\partial^2 l_{\mathbf{R}}}{\partial \beta \partial \tau} \\ -\mathbf{E} \frac{\partial^2 l_{\mathbf{R}}}{\partial \lambda \partial \alpha} & -\mathbf{E} \frac{\partial^2 l_{\mathbf{R}}}{\partial \lambda \partial \beta} & -\mathbf{E} \frac{\partial^2 l_{\mathbf{R}}}{\partial \lambda^2} & -\mathbf{E} \frac{\partial^2 l_{\mathbf{R}}}{\partial \lambda \partial \tau} \\ -\mathbf{E} \frac{\partial^2 l_{\mathbf{R}}}{\partial \tau \partial \alpha} & -\mathbf{E} \frac{\partial^2 l_{\mathbf{R}}}{\partial \tau \partial \beta} & -\mathbf{E} \frac{\partial^2 l_{\mathbf{R}}}{\partial \tau \partial \lambda} & -\mathbf{E} \frac{\partial^2 l_{\mathbf{R}}}{\partial \tau^2} \end{pmatrix},$$

kde

$$J_{11}^{\mathbf{R}} = J_{13}^{\mathbf{R}} = J_{31}^{\mathbf{R}} = n J_{11}(\lambda + \alpha, \tau + \beta, N_{2,0}, \dots, N_{2,k}, X_{2,(n-N_{2,0}+1)}, \dots, X_{2,(n)}),$$



$$\begin{aligned}
J_{12}^R &= J_{21}^R = J_{14}^R = J_{41}^R = J_{23}^R = J_{32}^R \\
&= nJ_{12}(\lambda + \alpha, \tau + \beta, N_{2,0}, \dots, N_{2,k}, X_{2,(n-N_{2,0}+1)}, \dots, X_{2,(n)}), \\
J_{22}^R &= J_{24}^R = J_{42}^R = nJ_{22}(\lambda + \alpha, \tau + \beta, N_{2,0}, \dots, N_{2,k}, X_{2,(n-N_{2,0}+1)}, \dots, X_{2,(n)}), \\
J_{33}^R &= nJ_{11}(\lambda, \tau, N_{1,0}, \dots, N_{1,k}, X_{1,(n-N_{1,0}+1)}, \dots, X_{1,(n)}) \\
&\quad + nJ_{11}(\lambda + \alpha, \tau + \beta, N_{2,0}, \dots, N_{2,k}, X_{2,(n-N_{2,0}+1)}, \dots, X_{2,(n)}), \\
J_{44}^R &= nJ_{22}(\lambda, \tau, N_{1,0}, \dots, N_{1,k}, X_{1,(n-N_{1,0}+1)}, \dots, X_{1,(n)}) \\
&\quad + nJ_{22}(\lambda + \alpha, \tau + \beta, N_{2,0}, \dots, N_{2,k}, X_{2,(n-N_{2,0}+1)}, \dots, X_{2,(n)}), \\
J_{34}^R &= J_{43}^R = nJ_{12}(\lambda, \tau, N_{1,0}, \dots, N_{1,k}, X_{1,(n-N_{1,0}+1)}, \dots, X_{1,(n)}) \\
&\quad + nJ_{12}(\lambda + \alpha, \tau + \beta, N_{2,0}, \dots, N_{2,k}, X_{2,(n-N_{2,0}+1)}, \dots, X_{2,(n)}),
\end{aligned}$$

a  $J_{11}, J_{12}, J_{22}$  lze najít ve větě 5.17

*Důkaz.* Zřejmé. □

### 5.2.3 Srovnání dvou nezávislých cenzorovaných výběrů z Weibullova rozdělení

Je-li potřeba porovnat dva nezávislé zleva cenzorované výběry z Weibullova rozdělení, lze k tomu použít asymptotické testy s rušivými parametry, které jsou popsány v odstavci 3.2. Nulová hypotéza je tvaru  $H_0 : (\alpha, \beta)^T = (0, 0)^T$  proti alternativě  $H_1 : (\alpha, \beta)^T \neq (0, 0)^T$ , tedy  $\lambda, \tau$  jsou rušivé parametry. Testovací statistiky mají asymptoticky  $\chi^2$  rozdělení s dvěma stupni volnosti a jsou tvaru

$$\begin{aligned}
LM &= \mathbf{U}_1^T(0, 0, \tilde{\lambda}, \tilde{\tau}) \left[ \mathbf{J}_{n,11.2}^R(0, 0, \tilde{\lambda}, \tilde{\tau}) \right]^{-1} \mathbf{U}_1(0, 0, \tilde{\lambda}, \tilde{\tau}), \\
W &= (\hat{\alpha}, \hat{\beta})^T \left[ \mathbf{J}_{n,11.2}^R(\hat{\alpha}, \hat{\beta}, \hat{\lambda}, \hat{\tau}) \right] (\hat{\alpha}, \hat{\beta}), \\
LR &= 2 \left[ l_R(\hat{\alpha}, \hat{\beta}, \hat{\lambda}, \hat{\tau}) - l_R(0, 0, \tilde{\lambda}, \tilde{\tau}) \right],
\end{aligned} \tag{5.21}$$

kde

$$\mathbf{J}_{n,11.2}^R(\alpha, \beta, \lambda, \tau) = \begin{pmatrix} J_{11}^R & J_{12}^R \\ J_{21}^R & J_{22}^R \end{pmatrix} - \begin{pmatrix} J_{13}^R & J_{14}^R \\ J_{23}^R & J_{24}^R \end{pmatrix} \begin{pmatrix} J_{33}^R & J_{34}^R \\ J_{43}^R & J_{44}^R \end{pmatrix}^{-1} \begin{pmatrix} J_{31}^R & J_{32}^R \\ J_{41}^R & J_{42}^R \end{pmatrix}$$

je transformace FIM  $\mathbf{J}_n^R(\alpha, \beta, \lambda, \tau)$  s dosazenými parametry odhadnutými za platnosti  $H_0$  (označeno vlnkou), případně bez dalších omezení (označeno stříškou). Funkce  $\mathbf{U}_1(\alpha, \beta, \lambda, \tau) = (\partial l_R / \partial \alpha, \partial l_R / \partial \beta)^T$  je skórová funkce s dosazenými parametry odhadnutými za platnosti  $H_0$  (označeno vlnkou). Nulová hypotéza se zamítne na dané hladině významnosti v případě, že hodnota zvolené testovací statistiky (5.21) přesáhne příslušnou kritickou hodnotu  $\chi^2$  rozdělení.

### 5.3 Test přípustnosti nahrazení cenzorovaného Weibullova rozdělení exponenciálním submodelem

Často může nastat situace, kdy je model cenzorovaného Weibullova rozdělení pro popis zvoleného datového souboru příliš složitý a model cenzorovaného exponenciálního rozdělení by byl dostačující. Proto bude dále popsán test pro posouzení přípustnosti nahrazení cenzorovaného Weibullova rozdělení exponenciálním submodelem.

Je-li třeba posoudit vhodnost nahrazení Weibullova rozdělení exponenciálním rozdělením, lze použít asymptotické testy s rušivými parametry, které jsou popsány v odstavci 3.2. Nulová hypotéza je vyjádřena jako omezení na hodnotu parametru  $\tau$  Weibullova rozdělení, tedy při nezamítnutí  $H_0 : \tau = 1$  proti alternativě  $H_1 : \tau \neq 1$  na zvolené hladině významnosti lze použít exponenciální rozdělení místo Weibullova rozdělení. V tomto případě je  $\tau$  cílový parametr a  $\lambda$  rušivý parametr. Testovací statistiky jsou pro necenzorované exponenciální a Weibullovo rozdělení odvozeny v knize [5].

V případě mnohonásobně zleva cenzorovaného Weibullova a exponenciálního rozdělení mají testovací statistiky asymptoticky  $\chi^2$  rozdělení s jedním stupněm volnosti a jsou tvaru

$$\begin{aligned} LM &= \frac{U_1^2(\tilde{\lambda}, 1)}{J_{n,22.1}(\tilde{\lambda}, 1)}, \\ W &= (\hat{\tau} - 1)^2 J_{n,22.1}(\hat{\lambda}, \hat{\tau}), \\ LR &= 2 \left[ l(\hat{\lambda}, \hat{\tau}) - l(\tilde{\lambda}, 1) \right], \end{aligned} \tag{5.22}$$

kde logaritmická věrohodnostní funkce  $l(\lambda, \tau)$  je dána vztahem (5.14). Dále  $J_{n,22.1}(\lambda, \tau) = n(J_{22} - J_{21}J_{11}^{-1}J_{12})$  je transformace FIM  $\mathbf{J}_n(\lambda, \tau)$ , která je uvedena ve větě 5.17, s dosazenými parametry odhadnutými za platnosti  $H_0$  (označeno vlnkou), případně bez dalších omezení (označeno stříškou). Funkce  $U_1(\lambda, \tau) = \partial l(\lambda, \tau) / \partial \tau$  je skórová funkce s dosazenými parametry odhadnutými za platnosti  $H_0$  (označeno vlnkou). Nulová hypotéza se zamítne na dané hladině významnosti v případě, že hodnota zvolené testovací statistiky (5.22) přesáhne příslušnou kritickou hodnotu  $\chi^2$  rozdělení.

## 6 Statistická analýza koncentrací musk sloučenin

V této kapitole budou metody odvozené v kapitole 5 použity ke zpracování reálných chemických dat. Konkrétně se bude jednat o analýzu koncentrací syntetických musk sloučenin, jejichž přítomnost byla odhalena v rybí tkáni. Závěrem bude posouzen vliv čistírny odpadních vod (ČOV) na koncentraci musk sloučenin v rybí tkáni.

Syntetické musk sloučeniny neboli syntetické vonné látky představují skupinu organických environmentálních kontaminantů, a to především kvůli své perzistenci, bioakumulacímu potenciálu a toxicitě (viz např. [12, 24]). Běžně se užívají jako náhrada přírodního pižma v parfémtech, detergentech (prací a čisticí prostředky), kosmetických produktech a produktech osobní péče (mýdla, šampony, tělová mýdla, atd.). K průniku těchto látek do životního prostředí dochází především prostřednictvím odpadních vod kvůli jejich nedostačitému odstranění v ČOV. Akumulace musk sloučenin v životním prostředí (povrchové vody, sedimenty) má za následek jejich výskyt v potravním řetězci, zvláště pak ve vodním ekosystému, neboť díky lipofilním vlastnostem mají schopnost biokoncentrace v různých typech vodních organismů. Tyto látky lze rovněž najít v lidském těle, například v tkáni nebo tělních tekutinách jako je krev nebo mateřské mléko (viz např. [67]), jako důsledek konzumace ryb. Je důležité sledovat koncentraci musk sloučenin v rybí tkáni, protože přítomnost musk sloučenin může mít negativní vliv na lidské tělo. Například musk ambrette je považován za silný fotoalergen, který může mít při vysokých dávkách neurotoxické a mutagenní účinky. Musk tibeten může mít karcinogenní účinky a některé musk sloučeniny jsou považovány za endokrinní disruptory.

Reálná data tvoří 60 ryb z čeledi kaprovitých (jelec tloušť), které byly chyceny v řece Svratce poblíž ČOV Brno–Modřice. Polovina ryb byla chycena před ČOV (skupina 1) a polovina za ČOV (skupina 2). Při analýze vzorků rybí tkáně, konkrétně svaloviny, byla odhalena přítomnost několika musk sloučenin.

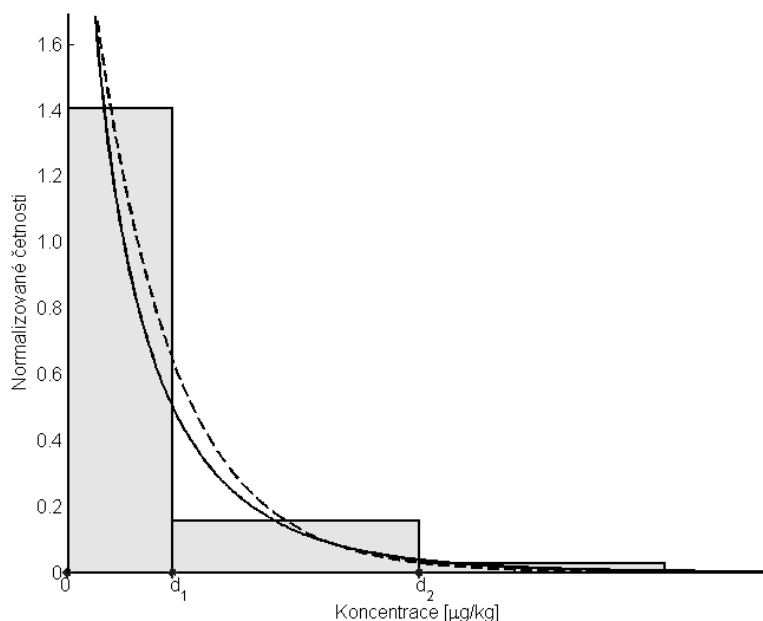
Při analýze musk sloučenin (a chemických látek obecně) lze často narazit na situaci, kdy sledovaná látka ve vzorku buď chybí, anebo je přítomna v tak nízké koncentraci, že nedosahuje detekčního limitu měřicího zařízení. Provedené chemické analýzy totiž nedovolují přesné stanovení příslušných koncentrací, pokud se výsledné hodnoty nachází pod mezí detekce nebo mezí stanovitelnosti (kvantifikace) použité metody. Mez detekce (LOD z anglického Limit Of Detection) je nejnižší koncentrace látky, kterou lze ještě rozlišit od nepřítomnosti této látky ve vzorku a mez stanovitelnosti (LOQ z anglického Limit Of Quantification) je nejnižší koncentrace, při které je možné od sebe odlišit dvě různé hodnoty koncentrace látky. Jelikož jsou předepsány dvě fixní meze ( $d_1 = \text{LOD}$  a  $d_2 = \text{LOQ}$ ), bude další analýza založena na dvojnásobně zleva cenzorovaných výběrech, přičemž cenzorování bude typu I.

Jak již bylo řečeno, v situacích, kdy měřená veličina nabývá pouze kladných hodnot, rozdělení této veličiny je asymetrické a má kladnou šikmost, je vhodné použít například

Sloučenina	$N_{1,0}$	$N_{1,1}$	$N_{1,2}$	$d_1$	$d_2$
	$N_{2,0}$	$N_{2,1}$	$N_{2,2}$		
Phantolid	0	18	12	0,5454	1,8180
	1	23	6		
Traseolid	0	24	6	1,1053	3,6844
	0	22	8		
Galaxolid	4	3	23	8,9488	29,8294
	8	0	22		
Tonalid	7	6	17	1,9846	6,6154
	18	8	4		

Tabulka 6.1: Analyzované musk sloučeniny.

exponenciální nebo Weibullovo rozdělení. Proto budou k analýze koncentrací využity výsledky odvozené v kapitole 5. Použití těchto metod bude ilustrována pro čtyři polycyklické musk sloučeniny (phantolid, traseolid, galaxolid, tonalid), viz tabulka 6.1.



Obrázek 6.1: Histogram koncentrace phantolidu s proloženou hustotou exponenciálního (čárkovaně) a Weibullova rozdělení (plnou čarou).

Nejprve bude modelována koncentrace musk sloučenin pomocí Weibullova rozdělení. Užitím metod odvozených v odstavci 5.2 se získají MV odhady parametrů  $\lambda$  a  $\tau$  dvojnásobně zleva cenzorovaného Weibullova rozdělení. Při pohledu na histogramy koncentrací musk sloučenin však vyzvala myšlenka, zda by nebylo možné nahradit rozdělení Weibullovo

Sloučenina	Skupina 1			Skupina 2		
	H	Stat	p-hodnota	H	Stat	p-hodnota
Phantolid	0	2,69	0,10	0	1,27	0,26
Traseolid	0	2,19	0,14	0	1,78	0,18
Galaxolid	1	9,49	0,00	1	19,40	0,00
Tonalid	1	6,19	0,01	1	4,21	0,04

Tabulka 6.2: Test poměrem věrohodností pro posouzení vhodnosti nahrazení Weibullova rozdělení exponenciálním rozdělením. Sloupec *Stat* obsahuje hodnoty testovací statistiky  $LR$  a  $H = 0$  (respektive  $H = 1$ ) značí, že se nulová hypotéza nezamítá (respektive zamítá) na hladině významnosti 0,05.

rozdělením exponenciálním. Proto byl užitím metod odvozených v odstavci 5.1 získán MV odhad parametru  $\lambda$  dvojnásobně zleva cenzorovaného exponenciálního rozdělení. Na obrázku 6.1 je pro ilustraci znázorněn histogram koncentrace phantolidu s proloženou hustotou exponenciálního a Weibullova rozdělení, ze kterého je patrné, že by takové nahrazení mohlo být přípustné.

K potvrzení vhodnosti nahrazení Weibullova rozdělení exponenciálním rozdělením lze použít asymptotický test s rušivými parametry, který je popsán v odstavci 5.3. Bude se tedy testovat nulová hypotéza  $H_0 : \tau = 1$  proti alternativě  $H_1 : \tau \neq 1$ . Při nezamítnutí  $H_0$  na zvolené hladině významnosti lze použít exponenciální rozdělení místo Weibullova rozdělení. V tabulce 6.2 lze najít výsledky testu poměrem věrohodností, ze kterých je patrné, že exponenciální rozdělení lze použít v případě sloučenin phantolid a traseolid a nelze použít v případě sloučenin galaxolid a tonalid. Zbývající asymptotické testy (Waldův test, skórový test) ukázaly obdobné výsledky. Další pozornost tedy bude věnována modelu exponenciálního rozdělení a Weibullovo rozdělení přijde na řadu později.

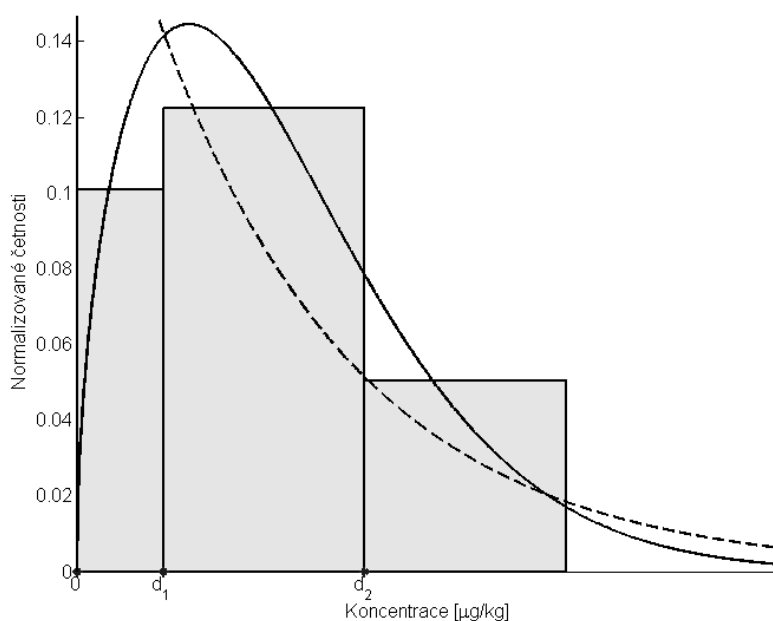
Aby bylo možné posoudit vliv ČOV na koncentraci phantolidu a traseolidu v rybí tkáni, bude třeba testovat, zda jsou rozdělení koncentrací musk sloučenin ve skupině 1 a 2 identická. K tomu lze využít model popsany v odstavci 5.1.3 a testy popsané v odstavci 5.1.4. Bude se tedy testovat nulová hypotéza  $H_0 : \alpha = 0$  proti alternativě  $H_1 : \alpha \neq 0$ . V tabulce 6.3 jsou výsledky skórového testu, testu poměrem věrohodností a Waldova testu, ze kterých je patrné, že ČOV nemá významný vliv na rozdělení koncentrací phantolidu a traseolidu, protože všechny testy nezamítají nulovou hypotézu na hladině významnosti 0,05.

Síly skórového testu, testu poměrem věrohodností a Waldova testu pro srovnání dvou nezávislých cenzorovaných výběrů z exponenciálního rozdělení byly porovnány užitím simulací, ze kterých vyplynulo, že při analýze koncentrací musk sloučenin je vhodné upřednostnit skórový test a test poměrem věrohodností. Podrobnosti lze najít v [36].

Sloučenina	LM			LR			W		
	H	Stat	p-hodnota	H	Stat	p-hodnota	H	Stat	p-hodnota
Phantolid	0	0,60	0,44	0	0,57	0,45	0	0,53	0,47
Traseolid	0	0,32	0,57	0	0,31	0,58	0	0,30	0,58

Tabulka 6.3: Srovnání rozdělení koncentrací musk sloučenin phantolid a traseolid mezi skupinou 1 a 2 užitím skórového testu (LM), testu poměrem věrohodností (LR) a Waldova testu (W). Sloupec *Stat* obsahuje hodnoty příslušné testovací statistiky a  $H = 0$  značí, že se nezamítá nulová hypotéza na hladině významnosti 0,05.

Nyní bude pozornost upřena na sloučeniny galaxolid a tonalid, u kterých nebylo možné nahradit Weibullovo rozdělení exponenciálním rozdělením. Na obrázku 6.2 je pro ilustraci znázorněn histogram koncentrace tonalidu s proloženou hustotou exponenciálního a Weibullova rozdělení, ze kterého je vidět, že v tomto případě není proložení exponenciálním rozdělením nejlepší volba, což potvrdil i test poměrem věrohodností (viz tabulka 6.2).



Obrázek 6.2: Histogram koncentrace tonalidu s proloženou hustotou exponenciálního (čárkovaně) a Weibullova rozdělení (plnou čarou).

Aby bylo možné posoudit vliv ČOV na koncentraci galaxolidu a tonalidu v rybí tkáni, bude se opět testovat, zda jsou rozdělení koncentrací musk sloučenin ve skupině 1 a 2 identická. K tomu se využije model popsáný v odstavci 5.2.2 a testy popsané v odstavci 5.2.3. Bude se tedy testovat nulová hypotéza  $H_0 : (\alpha, \beta)^T = (0, 0)^T$  proti alternativě  $H_1 : (\alpha, \beta)^T \neq (0, 0)^T$ . V tabulce 6.4 jsou výsledky skórového testu, testu poměrem věrohodností a Waldova testu, ze kterých je patrné, že ČOV nemá významný vliv na rozdělení

Sloučenina	LM			LR			W		
	H	Stat	p-hodnota	H	Stat	p-hodnota	H	Stat	p-hodnota
Galaxolid	0	2,19	0,33	0	2,06	0,36	0	1,96	0,38
Tonalid	0	2,52	0,28	0	2,52	0,28	0	2,37	0,31

Tabulka 6.4: Srovnání rozdělení koncentrací musk sloučenin galaxolid a tonalid mezi skupinou 1 a 2 užitím skórového testu (LM), testu poměrem věrohodností (LR) a Waldova testu (W). Sloupec *Stat* obsahuje hodnoty příslušné testovací statistiky a  $H = 0$  značí, že se nezamítá nulová hypotéza na hladině významnosti 0,05.

koncentrací galaxolidu a tonalidu, protože všechny testy nezamítají nulovou hypotézu na hladině významnosti 0,05.





## 7 Teorie extrémních hodnot pro dvourozměrná pozorování

Teorie extrémních hodnot se neomezuje pouze na případ jednorozměrných pozorování, ale existuje dostatečně bohatá teorie i o vícerozměrných extrémech. Aby bylo docíleno větší přehlednosti v probírané problematice, bude pozornost soustředěna pouze na dvourozměrný případ. V této kapitole budou tedy stručně shrnuty základy z teorie extrémních hodnot pro dvourozměrná pozorování. Vzhledem k tomu, že tato práce je zaměřena především na teorii extrémních hodnot pro jednorozměrná pozorování a její aplikace, budou některé věty v této kapitole uvedeny bez důkazů s tím, že příslušné důkazy lze najít v monografii [22].

Nechť  $(X_1, Y_1), \dots, (X_n, Y_n)$  je posloupnost nezávislých a stejně rozdělených náhodných vektorů s distribuční funkcí  $F$ . Vektor  $(X, Y)$  může reprezentovat například výšku mořské hladiny ve dvou různých lokacích nebo dvojici výška vln a výška klidné mořské hladiny. Při analýze dvourozměrných (a obecně vícerozměrných) extrémů vyvstává důležitá otázka, a to jakým způsobem definovat maximum množiny vektorů  $(X_1, Y_1), \dots, (X_n, Y_n)$ . Ukázalo se, že stačí určit maxima z veličin  $X_1, \dots, X_n$  a  $Y_1, \dots, Y_n$  odděleně a následně z jednotlivých maxim sestavit nový vektor. Výsledný vektor maxim tedy většinou nebude roven některému z původních vektorů  $(X_i, Y_i)$ ,  $i = 1, \dots, n$ . Dále předpokládejme, že existují takové posloupnosti reálných konstant  $a_n > 0$ ,  $b_n$ ,  $c_n > 0$ ,  $d_n$ , a simultánní distribuční funkce  $G(x, y)$  s nedegenerovanými marginálními distribučními funkcemi  $G_1(x) = G(x, \infty)$ ,  $G_2(y) = G(\infty, y)$ , že pro všechny body spojitosti  $(x, y) \in \mathbb{R}^2$  funkce  $G(x, y)$  platí

$$\lim_{n \rightarrow \infty} P \left( \frac{\max \{X_1, \dots, X_n\} - b_n}{a_n} \leq x, \frac{\max \{Y_1, \dots, Y_n\} - d_n}{c_n} \leq y \right) = G(x, y). \quad (7.1)$$

Rozdělení s distribuční funkcí  $G$  definovanou vztahem (7.1) s nedegenerovanými marginálními distribučními funkcemi  $G_1$ ,  $G_2$  se pak nazývá dvourozměrné EV rozdělení.

Dále bude stanovena třída všech možných limitních rozdělení s distribuční funkcí  $G$ . K tomu se využijí teoretické poznatky z kapitoly 1. Jelikož ze vztahu (7.1) plyne konvergence jednorozměrných marginálních rozdělení, platí

$$\lim_{n \rightarrow \infty} P \left( \frac{\max \{X_1, \dots, X_n\} - b_n}{a_n} \leq x \right) = G_1(x) \quad (7.2)$$

a

$$\lim_{n \rightarrow \infty} P \left( \frac{\max \{Y_1, \dots, Y_n\} - d_n}{c_n} \leq y \right) = G_2(y), \quad (7.3)$$

přičemž tato marginální rozdělení jsou spojitého typu. Nyní se zvolí takové posloupnosti konstant  $a_n$ ,  $b_n$ ,  $c_n$ ,  $d_n$ , že pro nějaká  $\gamma_1, \gamma_2 \in \mathbb{R}$  platí (viz věta 1.8)

$$G_1(x) = \exp \left[ -(1 + \gamma_1 x)^{-\frac{1}{\gamma_1}} \right] \quad (7.4)$$

a

$$G_2(y) = \exp \left[ -(1 + \gamma_2 y)^{-\frac{1}{\gamma_2}} \right]. \quad (7.5)$$

Jelikož  $F(x, y)$  představuje spojitě rozdělení a marginální distribuční funkce  $G_1(x)$  a  $G_2(y)$  jsou nedegenerované, je spojitá i simultánní distribuční funkce  $G(x, y)$ . Dále nechť  $F_i$ ,  $i = 1, 2$ , jsou marginální distribuční funkce k funkci  $F$  a  $U_i(t) = F_i^{\leftarrow}(1 - 1/t)$ ,  $t > 1$ ,  $i = 1, 2$ , jsou příslušné kvantilové funkce chvostu. Pak z věty 1.10 plyne, že existuje taková kladná funkce  $a_i(t)$ ,  $i = 1, 2$ , že pro všechna  $x > 0$  platí

$$\lim_{t \rightarrow \infty} \frac{U_i(tx) - U_i(t)}{a_i(t)} = \frac{x^{\gamma_i} - 1}{\gamma_i}, \quad (7.6)$$

a

$$\lim_{t \rightarrow \infty} \frac{a_i(tx)}{a_i(t)} = x^{\gamma_i}.$$

Dále vztahy (7.2)–(7.5) platí (viz věta 1.7) také pro

$$\begin{aligned} a_n &:= a_1(\lfloor n \rfloor), \\ b_n &:= U_1(\lfloor n \rfloor), \\ c_n &:= a_2(\lfloor n \rfloor), \\ d_n &:= U_2(\lfloor n \rfloor), \end{aligned}$$

kde  $\lfloor n \rfloor$  označuje celou část  $n$ . Z (7.6) plyne, že pro  $x, y > 0$  platí

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{U_1(nx) - b_n}{a_n} &= \frac{x^{\gamma_1} - 1}{\gamma_1}, \\ \lim_{n \rightarrow \infty} \frac{U_2(ny) - d_n}{c_n} &= \frac{y^{\gamma_2} - 1}{\gamma_2}. \end{aligned} \quad (7.7)$$

Vztah (7.1) lze dále přepsat jako

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n, c_n y + d_n) = G(x, y).$$

Jestliže  $x_n \rightarrow u$ ,  $y_n \rightarrow v$ , pak z monotonie funkce  $F$  a spojitosti funkce  $G$  plyne

$$\lim_{n \rightarrow \infty} F^n(a_n x_n + b_n, c_n y_n + d_n) = G(u, v). \quad (7.8)$$

Nyní se položí

$$\begin{aligned} x_n &:= \frac{U_1(nx) - b_n}{a_n}, \\ y_n &:= \frac{U_2(ny) - d_n}{c_n}, \end{aligned}$$

pro  $x, y > 0$ . Pak, vezme-li se do úvahy vztah (7.7), lze po dosazení do (7.8) dostat

$$\lim_{n \rightarrow \infty} F^n(U_1(nx), U_2(ny)) = G\left(\frac{x^{\gamma_1} - 1}{\gamma_1}, \frac{y^{\gamma_2} - 1}{\gamma_2}\right),$$

a tím i tvrzení následující věty.

**Věta 7.1.** *Nechť existují takové reálné konstanty  $a_n > 0$ ,  $b_n$ ,  $c_n > 0$ ,  $d_n$ , že*

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n, c_n y + d_n) = G(x, y)$$

*platí pro všechna  $(x, y)$  funkce  $G$  a marginální distribuční funkce  $G_1(x)$ ,  $G_2(y)$  jsou tvaru*

$$G_1(x) = \exp \left[ -(1 + \gamma_1 x)^{-\frac{1}{\gamma_1}} \right]$$

*a*

$$G_2(y) = \exp \left[ -(1 + \gamma_2 y)^{-\frac{1}{\gamma_2}} \right].$$

*Pak pro  $F_1(x) := F(x, \infty)$ ,  $F_2(y) := F(\infty, y)$ ,  $U_i(x) := F_i^{\leftarrow}(1 - 1/x)$ ,  $i = 1, 2$ , a všechna  $x, y > 0$  platí, že*

$$\lim_{n \rightarrow \infty} F^n(U_1(nx), U_2(ny)) = G_0(x, y),$$

*kde*

$$G_0(x, y) := G \left( \frac{x^{\gamma_1} - 1}{\gamma_1}, \frac{y^{\gamma_2} - 1}{\gamma_2} \right)$$

*a  $\gamma_1, \gamma_2$  jsou marginální EV indexy ze vztahů (7.4) a (7.5).*

**Věta 7.2.** *K libovolné distribuční funkci EV rozdělení ze vztahu (7.1) existuje za platnosti (7.4) a (7.5) konečná míra  $\nu$  definovaná na  $\langle 0; \pi/2 \rangle$ , která je reprezentována distribuční funkcí  $\Psi$  pomocí vztahu*

$$G \left( \frac{x^{\gamma_1} - 1}{\gamma_1}, \frac{y^{\gamma_2} - 1}{\gamma_2} \right) = G_0(x, y) = \exp \left[ - \int_0^{\frac{\pi}{2}} \max \left\{ \frac{\cos \theta}{x}, \frac{\sin \theta}{y} \right\} \Psi(d\theta) \right], \quad x, y > 0. \quad (7.9)$$

*Parametry  $\gamma_1$  a  $\gamma_2$  jsou EV indexy marginálních rozdělení příslušných rozdělení s distribuční funkcí  $G$ . Navíc platí podmínka*

$$\int_0^{\frac{\pi}{2}} \cos \theta \Psi(d\theta) = \int_0^{\frac{\pi}{2}} \sin \theta \Psi(d\theta) = 1. \quad (7.10)$$

*Naopak každá konečná míra reprezentovaná distribuční funkcí  $\Psi$  definuje limitní distribuční funkci  $G$  ve vztahu (7.1) prostřednictvím vztahu (7.9) za předpokladu, že je splněna podmínka (7.10).*

*Důkaz.* Důkaz lze najít v monografii [22].

□

**Definice 7.3.** Konečná míra  $\nu$  zavedená ve větě 7.2 se nazývá spektrální míra.

**Věta 7.4.** *Ke každému limitnímu rozdělení s distribuční funkcí  $G$  danou vztahem (7.1) s příslušnými marginálními distribučními funkcemi (7.4) a (7.5) existuje:*

(1) Konečná míra (reprezentovaná distribuční funkcí  $\Psi$ ) na intervalu  $\langle 0; \pi/2 \rangle$  taková, že pro  $x, y > 0$  platí

$$G\left(\frac{x^{\gamma_1} - 1}{\gamma_1}, \frac{y^{\gamma_2} - 1}{\gamma_2}\right) = G_0(x, y) = \exp\left[-\int_0^{\pi/2} \max\left\{\frac{\cos \theta}{x}, \frac{\sin \theta}{y}\right\} \Psi(d\theta)\right] \quad (7.11)$$

s podmínkou

$$\int_0^{\pi/2} \cos \theta \Psi(d\theta) = \int_0^{\pi/2} \sin \theta \Psi(d\theta) = 1.$$

(2) Rozdělení (reprezentované distribuční funkcí  $H$ ) soustředěné na intervalu  $\langle 0; 1 \rangle$  se střední hodnotou 0,5 takové, že pro  $x, y > 0$  platí

$$G\left(\frac{x^{\gamma_1} - 1}{\gamma_1}, \frac{y^{\gamma_2} - 1}{\gamma_2}\right) = G_0(x, y) = \exp\left[-2 \int_0^1 \max\left\{\frac{w}{x}, \frac{1-w}{y}\right\} H(dw)\right]. \quad (7.12)$$

(3) Konečná míra (reprezentovaná distribuční funkcí  $\Lambda$ ) na intervalu  $\langle 0; \pi/2 \rangle$  taková, že pro  $x, y > 0$  platí

$$\begin{aligned} G\left(\frac{x^{\gamma_1} - 1}{\gamma_1}, \frac{y^{\gamma_2} - 1}{\gamma_2}\right) &= G_0(x, y) \\ &= \exp\left[-\int_0^{\pi/2} \max\left\{\frac{\min\{1, \operatorname{tg} \theta\}}{x}, \frac{\min\{1, \operatorname{cotg} \theta\}}{y}\right\} \Lambda(d\theta)\right] \end{aligned} \quad (7.13)$$

s podmínkou

$$\int_0^{\pi/2} \min\{1, \operatorname{tg} \theta\} \Lambda(d\theta) = \int_0^{\pi/2} \min\{1, \operatorname{cotg} \theta\} \Lambda(d\theta) = 1.$$

Parametry  $\gamma_1$  a  $\gamma_2$  jsou EV indexy marginálních rozdělení příslušných rozdělení s distribuční funkcí  $G$ .

Naopak libovolná konečná míra reprezentovaná distribuční funkcí  $\Psi$ ,  $H$  nebo  $\Lambda$  zapříčiní existenci limitní distribuční funkce  $G$  v (7.1) prostřednictvím vztahů (7.11), (7.12) nebo (7.13) s tím, že musí být splněny přidružené podmínky.

**Poznámka 7.5.** Z vyjádření (7.11)–(7.13) plyne, že limitní rozdělení ve vztahu (7.1) jsou charakterizována pouze spektrální mírou a EV indexy příslušných marginálních rozdělení.

Spektrální míry  $\Psi$ ,  $H$  a  $\Lambda$  nejsou jediné, ale jsou nejčastěji používané a jednu na druhou lze přetransformovat. Záleží tedy na konkrétní situaci, kterou míru je vhodnější použít.

Jak bylo zmíněno v poznámce 7.5, třída limitních rozdělení ve vztahu (7.1) je charakterizována pouze EV indexy  $\gamma_1$  a  $\gamma_2$  marginálních rozdělení příslušných distribuční funkce  $G$  a spektrální mírou, která určuje závislostní strukturu mezi jednotlivými veličinami. To se odráží v podmínce oboru atraktivity, která kromě marginálních rozdělení zahrnuje i závislostní strukturu.

Nechť  $G : \mathbb{R}^2 \rightarrow \mathbb{R}_+$  je distribuční funkce EV rozdělení. Jestliže pro posloupnost reálných konstant  $a_n > 0$ ,  $b_n, c_n > 0$ ,  $d_n$  platí

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n, c_n y + d_n) = G(x, y)$$

pro všechna  $x, y \in \mathbb{R}$ , pak říkáme, že distribuční funkce  $F$  patří do oboru atraktivity distribuční funkce  $G$ .

**Věta 7.6.** *Nechť  $G$  je distribuční funkce EV rozdělení a  $\exp[-(1 + \gamma_i x)^{-1/\gamma_i}]$ ,  $i = 1, 2$ , jsou marginální distribuční funkce příslušné funkci  $G$ . Dále necht'  $\Psi$  nebo  $H$  nebo  $\Lambda$  je spektrální míra, která byla zavedena ve větě 7.4. Platí následující tvrzení.*

(1) *Jestliže distribuční funkce  $F$  náhodného vektoru  $(X, Y)$  se spojitými marginálními distribučními funkcemi  $F_1, F_2$  patří do oboru atraktivity funkce  $G$ , pak platí následující ekvivalentní podmínky:*

(a) *Pro  $U_i = (1/(1 - F_i))^\leftarrow$ ,  $i = 1, 2$  a  $x, y > 0$  platí*

$$\lim_{t \rightarrow \infty} \frac{1 - F(U_1(tx), U_2(ty))}{1 - F(U_1(t), U_2(t))} = S(x, y), \quad (7.14)$$

*kde  $S(x, y) := \ln G((x^{\gamma_1} - 1)/\gamma_1, (y^{\gamma_2} - 1)/\gamma_2) / \ln G(0, 0)$ .*

(b) *Pro všechna  $r > 1$  a všechny body spojitosti  $\theta \in \langle 0; \pi/2 \rangle$  funkce  $\Psi$  platí*

$$\lim_{t \rightarrow \infty} P \left( V^2 + W^2 > t^2 r^2 \wedge \frac{W}{V} \leq \operatorname{tg} \theta \mid V^2 + W^2 > t^2 \right) = r^{-1} \frac{\Psi(\theta)}{\Psi(\frac{\pi}{2})}, \quad (7.15)$$

*kde  $V := 1/(1 - F_1(X))$  a  $W := 1/(1 - F_2(Y))$ .*

(c) *Pro všechna  $r > 1$  a všechny body spojitosti  $s \in \langle 0; 1 \rangle$  funkce  $H$  platí*

$$\lim_{t \rightarrow \infty} P \left( V + W > tr \wedge \frac{V}{V + W} \leq s \mid V + W > t \right) = r^{-1} H(s), \quad (7.16)$$

*kde  $V := 1/(1 - F_1(X))$  a  $W := 1/(1 - F_2(Y))$ .*

(d) *Pro všechna  $r > 1$  a všechny body spojitosti  $\theta \in \langle 0; \pi/2 \rangle$  funkce  $\Lambda$  platí*

$$\lim_{t \rightarrow \infty} P \left( \max \{V, W\} > tr \wedge \frac{V}{W} \leq \operatorname{tg} \theta \mid \max \{V, W\} > t \right) = r^{-1} \frac{\Lambda(\theta)}{\Lambda(\frac{\pi}{2})}, \quad (7.17)$$

*kde  $V := 1/(1 - F_1(X))$  a  $W := 1/(1 - F_2(Y))$ .*

(2) *Naopak jestliže spojitě marginální distribuční funkce  $F_i$  patří do oboru atraktivity funkce  $\exp[-(1 + \gamma_i x)^{-1/\gamma_i}]$ ,  $i = 1, 2$ , a libovolný ze vztahů (7.14)–(7.17) platí pro nějakou kladnou funkci  $S$  nebo nějakou ohraničenou distribuční funkci  $\Psi$ ,  $H$  nebo  $\Lambda$ , pak  $F$  patří do oboru atraktivity  $G$ .*

*Důkaz.* Důkaz lze najít v monografii [22].

□

Dosud bylo ukázáno, že mnohorozměrné EV rozdělení je charakterizováno EV indexy marginálních EV rozdělení a spektrální mírou charakterizující závislostní strukturu mezi veličinami. Neexistuje tedy exaktní parametrizace mnohorozměrného EV rozdělení. Jako vhodný přístup k odhadu závislostní struktury se jeví použití neparametrických metod. Rozsáhlejší analýza mnohorozměrných extrémů však již přesahuje rámec této práce. Další výsledky z teorie mnohorozměrných extrémů lze najít například v [8, 15, 22, 59].

## Závěr

Dizertační práce je zaměřena na rozdělení extrémních hodnot a jejich aplikace. Jelikož není v současné době dostupná žádná česky psaná monografie s touto tematikou, jsou v práci shrnuty základní poznatky z teorie extrémních hodnot pro jednorozměrná a dvourozměrná pozorování. Vzhledem k aplikacím EV rozdělení v této práci a dostupnosti reálných dat je největší pozornost věnována jednorozměrnému případu.

Nejprve jsou v práci shrnuty základní myšlenky teorie extrémních hodnot pro jednorozměrná pozorování. Především je zde uvedena a dokázána limitní věta pro rozdělení maxim, což je analogie centrální limitní věty pro rozdělení výběrových průměrů. Ukázalo se, že limitní rozdělení může být jednoho ze tří typů (Gumbelovo, Fréchetovo, Weibullovo). Dále bylo pro 18 vybraných rozdělení pravděpodobností dokázáno, že patří do oboru atraktivity některého ze tří uvedených limitních rozdělení. Následně byly představeny a odvozeny dva modely pro odhady parametrických funkcí rozdělení extrémních hodnot. První je model blokových maxim vycházející z GEV rozdělení, které je zobecněním tří výše uvedených limitních rozdělení. Dále bylo ukázáno, že za jistých předpokladů je možné přejít od modelu blokových maxim k tzv. prahovému modelu, který je založen na zobecněném Paretově rozdělení.

Aby bylo možné výše uvedené modely využít pro analýzu reálných dat, byla další pozornost věnována metodám odhadu parametrů GEV rozdělení a zobecněného Paretova rozdělení. Pro uvedená rozdělení byly odvozeny odhady parametrů metodou maximální věrohodnosti, včetně odvození rozptylů těchto odhadů vycházejících z výběrové FIM, a metodou pravděpodobnostně vážených momentů. Za účelem testování adekvátnosti použití výše uvedených modelů byly stručně popsány vybrané testy dobré shody (Pearsonův  $\chi^2$  test, Kolmogorovův-Smirnovův test, Andersonův-Darlingův test) a asymptotické testy vycházející z teorie maximální věrohodnosti včetně testů s rušivými parametry.

Popsané metody byly dále použity k analýze reálných hydrologických dat. Nejprve byly popsány dvě vzorkovací techniky (AMS, PDS) sloužící k sestavení statistického souboru hydrologických dat, které se nejčastěji používají v hydrologii. Následně byly užitím metody PDS a zobecněného Paretova rozdělení modelovány úhrny dešťových srážek v brněnském regionu. Tato aplikace EV rozdělení byla motivována potřebou aktualizace hydrologických podkladů pro tvorbu městského odvodnění a vychází z autorových prací [49, 50, 75].

Další pozornost byla věnována statistické inferenci pro vybraná cenzorovaná rozdělení z oboru atraktivity Gumbelova rozdělení, protože Gumbelova třída se v praxi často vyskytuje. Tato část práce vychází zejména z autorových prací [33, 34, 35] a byla motivována potřebou analyzovat reálná chemická data, která jsou reprezentována dvojnásobně zleva cenzorovaným výběrem s cenzorováním typu I. Jelikož dosud nebyly dostupné žádné statistické metody, které by bylo možné snadno aplikovat na tento typ dat, byly vyvinuty nové metody pro analýzu mnohonásobně zleva cenzorovaných výběrů z exponenciálního

a Weibullova rozdělení. Pro obě cenzorovaná rozdělení byly odvozeny odhady parametrů metodou maximální věrohodnosti, včetně odvození rozptylů těchto odhadů vycházejících nejen z výběrové FIM, ale i z očekávané FIM, která má, jak se ukázalo, lepší asymptotické vlastnosti. Odvozená metodika pro analýzu jednoho výběru byla dále rozšířena pro dva nezávislé cenzorované výběry. Aby bylo možné mezi sebou porovnat dva nezávislé mnohonásobně zleva cenzorované výběry z exponenciálního a Weibullova rozdělení, byly odvozeny příslušné testovací statistiky, které vychází z asymptotických testů s rušivými parametry. Testy s rušivými parametry jsou v případě necenzorovaných výběrů poměrně známou statistickou technikou (viz např. [5]). Zde však bylo jejich použití rozšířeno pro mnohonásobně zleva cenzorované výběry z exponenciálního a Weibullova rozdělení.

Vyvinuté metody byly počítačově implementovány v prostředí Matlab a následně použity ke zpracování reálných chemických dat. Konkrétně se jednalo o analýzu koncentrací syntetických musk sloučenin, které vycházelo zejména z autorových prací [36, 37]. Použitím metod na porovnání dvou nezávislých cenzorovaných výběrů bylo ukázáno, že vliv ČOV na koncentraci musk sloučenin v rybí tkáni je zanedbatelný. Celkově se tedy ukázalo, že odvozené metody jsou dobře použitelné pro zpracování chemických či environmentálních dat, která jsou reprezentována mnohonásobně zleva cenzorovaným výběrem z exponenciálního a Weibullova rozdělení.

V rámci této práce byl také vytvořen demonstrační software pro rozdělení extrémních hodnot, který umožňuje uživateli získat základní představu o tvaru EV rozdělení a o rozděleních z jejich oborů atraktivitu. Pomocí tohoto softwaru je také možné posoudit rychlost konvergence daného rozdělení k příslušnému limitnímu EV rozdělení.



# Dodatky



## A Odvození EV indexů

Tato kapitola je věnována ověření von Misesovy podmínky (1.26) a také s tím spojenému odvození EV indexů pro rozdělení z tabulek 1.1–1.3. Připomeňme, že je nutné určit limitu

$$\lim_{x \uparrow x^*} \left( \frac{1 - F(x)}{F'(x)} \right)' = \gamma,$$

kde  $\gamma$  je EV index, který je roven nule pro rozdělení patřící do oboru atraktivity Gumbelova rozdělení.

### Rozdělení z oboru atraktivity Gumbelova rozdělení

**Věta A.1.** *Uvažujme Benktanderovo (typ II) rozdělení s parametry  $\alpha > 0$ ,  $0 < \beta \leq 1$ , distribuční funkcí*

$$F(x) = \begin{cases} 1 - x^{\beta-1} \exp\left(\frac{\alpha}{\beta} - \frac{\alpha}{\beta}x^\beta\right) & \text{pro } x \geq 1, \\ 0 & \text{pro } x < 1, \end{cases}$$

a hustotou

$$f(x) = \begin{cases} (1 - \beta + \alpha x^\beta)x^{\beta-2} \exp\left(\frac{\alpha}{\beta} - \frac{\alpha}{\beta}x^\beta\right) & \text{pro } x \geq 1, \\ 0 & \text{pro } x < 1. \end{cases}$$

Pak EV index tohoto rozdělení je  $\gamma = 0$ .

*Důkaz.* Dosazením hustoty a distribuční funkce do vztahu (1.26) a postupnými úpravami lze odvodit

$$\begin{aligned} \gamma &= \lim_{x \uparrow \infty} \left[ \frac{x^{\beta-1} \exp\left(\frac{\alpha}{\beta} - \frac{\alpha}{\beta}x^\beta\right)}{(1 - \beta + \alpha x^\beta)x^{\beta-2} \exp\left(\frac{\alpha}{\beta} - \frac{\alpha}{\beta}x^\beta\right)} \right]' = \lim_{x \uparrow \infty} \left[ \frac{x}{1 - \beta + \alpha x^\beta} \right]' \\ &= \lim_{x \uparrow \infty} \frac{1 - \beta + \alpha x^\beta - \alpha \beta x^\beta}{(1 - \beta + \alpha x^\beta)^2} = \lim_{x \uparrow \infty} \frac{(1 - \beta + \alpha - \alpha \beta)x^\beta}{(1 - \beta + \alpha x^\beta)^2} = 0. \end{aligned}$$

□

**Věta A.2.** *Uvažujme Weibullovo rozdělení s parametry  $\lambda > 0$ ,  $\tau > 0$ , distribuční funkcí*

$$F(x) = \begin{cases} 1 - e^{-\lambda x^\tau} & \text{pro } x \geq 0, \\ 0 & \text{pro } x < 0, \end{cases}$$

a hustotou

$$f(x) = \begin{cases} \lambda \tau x^{\tau-1} e^{-\lambda x^\tau} & \text{pro } x \geq 0, \\ 0 & \text{pro } x < 0. \end{cases}$$

Pak EV index tohoto rozdělení je  $\gamma = 0$ .

*Důkaz.* Dosazením hustoty a distribuční funkce do vztahu (1.26) a postupnými úpravami lze odvodit

$$\gamma = \lim_{x \uparrow \infty} \left[ \frac{e^{-\lambda x^\tau}}{\lambda \tau x^{\tau-1} e^{-\lambda x^\tau}} \right]' = \lim_{x \uparrow \infty} \left( \frac{x^{1-\tau}}{\lambda \tau} \right)' = \lim_{x \uparrow \infty} \frac{(1-\tau)x^{-\tau}}{\lambda \tau} = 0.$$

□

**Věta A.3.** Uvažujme exponenciální rozdělení s parametrem  $\lambda > 0$ , distribuční funkcí

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{pro } x \geq 0, \\ 0 & \text{pro } x < 0, \end{cases}$$

a hustotou

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{pro } x \geq 0, \\ 0 & \text{pro } x < 0. \end{cases}$$

Pak EV index tohoto rozdělení je  $\gamma = 0$ .

*Důkaz.* Dosazením hustoty a distribuční funkce do vztahu (1.26) a postupnými úpravami lze odvodit

$$\gamma = \lim_{x \uparrow \infty} \left[ \frac{e^{-\lambda x}}{\lambda e^{-\lambda x}} \right]' = \lim_{x \uparrow \infty} \left( \frac{1}{\lambda} \right)' = 0.$$

□

**Věta A.4.** Uvažujme gama rozdělení s parametry  $\lambda > 0$ ,  $m > 0$ , distribuční funkcí

$$F(x) = \begin{cases} 1 - \int_x^\infty \frac{\lambda^m}{\Gamma(m)} t^{m-1} e^{-\lambda t} dt & \text{pro } x > 0, \\ 0 & \text{pro } x \leq 0, \end{cases}$$

a hustotou

$$f(x) = \begin{cases} \frac{\lambda^m}{\Gamma(m)} x^{m-1} e^{-\lambda x} & \text{pro } x > 0, \\ 0 & \text{pro } x \leq 0, \end{cases}$$

Pak EV index tohoto rozdělení je  $\gamma = 0$ .

*Důkaz.* Dosazením hustoty a distribuční funkce do vztahu (1.26) a postupnými úpravami lze odvodit

$$\begin{aligned} \gamma &= \lim_{x \uparrow \infty} \left[ \frac{\int_x^\infty \frac{\lambda^m}{\Gamma(m)} t^{m-1} e^{-\lambda t} dt}{\frac{\lambda^m}{\Gamma(m)} x^{m-1} e^{-\lambda x}} \right]' = \lim_{x \uparrow \infty} \frac{1}{\left[ \frac{\lambda^m}{\Gamma(m)} x^{m-1} e^{-\lambda x} \right]^2} \left\{ - \left[ \frac{\lambda^m}{\Gamma(m)} x^{m-1} e^{-\lambda x} \right]^2 \right. \\ &\quad \left. - \left[ \int_x^\infty \frac{\lambda^m}{\Gamma(m)} t^{m-1} e^{-\lambda t} dt \right] \left[ \frac{\lambda^m}{\Gamma(m)} (m-1) x^{m-2} e^{-\lambda x} - \frac{\lambda^{m+1}}{\Gamma(m)} x^{m-1} e^{-\lambda x} \right] \right\} \\ &= \lim_{x \uparrow \infty} \left\{ -1 - (m-1) \lambda^{-m} \Gamma(m) x^{-m} e^{\lambda x} \left[ \int_x^\infty \frac{\lambda^m}{\Gamma(m)} t^{m-1} e^{-\lambda t} dt \right] \right\} \end{aligned}$$

$$\begin{aligned}
& + \lambda^{1-m} \Gamma(m) x^{1-m} e^{\lambda x} \left[ \int_x^\infty \frac{\lambda^m}{\Gamma(m)} t^{m-1} e^{-\lambda t} dt \right] \Big\} \\
= & -1 - (m-1) \lambda^{-m} \Gamma(m) \lim_{x \uparrow \infty} \left[ \frac{\int_x^\infty \frac{\lambda^m}{\Gamma(m)} t^{m-1} e^{-\lambda t} dt}{\frac{x^m}{e^{\lambda x}}} \right] + \lambda^{1-m} \Gamma(m) \\
& \times \lim_{x \uparrow \infty} \left[ \frac{\int_x^\infty \frac{\lambda^m}{\Gamma(m)} t^{m-1} e^{-\lambda t} dt}{\frac{x^m}{x e^{\lambda x}}} \right] \\
= & -1 + (m-1) \lim_{x \uparrow \infty} \left[ \frac{x^{m-1} e^{-\lambda x}}{e^{-\lambda x} (m x^{m-1} - \lambda x^m)} \right] - \lim_{x \uparrow \infty} \left[ \frac{\lambda x^{m-1} e^{-\lambda x}}{e^{-\lambda x} (m x^{m-2} - x^{m-2} - \lambda x^{m-1})} \right] \\
= & -1 + (m-1) \lim_{x \uparrow \infty} \left( \frac{1}{m - \lambda x} \right) - \lim_{x \uparrow \infty} \left( \frac{\lambda}{m x^{-1} - x^{-1} - \lambda} \right) = 0.
\end{aligned}$$

□

**Věta A.5.** Uvažujme logistické rozdělení s distribuční funkcí

$$F(x) = 1 - \frac{1}{1 + e^x} \quad \text{pro } x \in \mathbb{R},$$

a hustotou

$$f(x) = \frac{e^x}{(1 + e^x)^2} \quad \text{pro } x \in \mathbb{R}.$$

Pak EV index tohoto rozdělení je  $\gamma = 0$ .

*Důkaz.* Dosazením hustoty a distribuční funkce do vztahu (1.26) a postupnými úpravami lze odvodit

$$\gamma = \lim_{x \uparrow \infty} \left[ \frac{\frac{1}{1+e^x}}{\frac{e^x}{(1+e^x)^2}} \right]' = \lim_{x \uparrow \infty} \left( \frac{1+e^x}{e^x} \right)' = \lim_{x \uparrow \infty} \left( -\frac{1}{e^x} \right) = 0.$$

□

**Věta A.6.** Uvažujme lognormální rozdělení s parametry  $\mu \in \mathbb{R}$ ,  $\sigma > 0$ , distribuční funkcí

$$F(x) = \begin{cases} 1 - \int_x^\infty \frac{1}{\sqrt{2\pi\sigma t}} \exp \left[ -\frac{(\ln t - \mu)^2}{2\sigma^2} \right] dt & \text{pro } x > 0, \\ 0 & \text{pro } x \leq 0, \end{cases}$$

a hustotou

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma x}} \exp \left[ -\frac{(\ln x - \mu)^2}{2\sigma^2} \right] & \text{pro } x > 0, \\ 0 & \text{pro } x \leq 0. \end{cases}$$

Pak EV index tohoto rozdělení je  $\gamma = 0$ .

*Důkaz.* Dosazením hustoty a distribuční funkce do vztahu (1.26) a postupnými úpravami lze odvodit

$$\gamma = \lim_{x \uparrow \infty} \left\{ \frac{\int_x^\infty \frac{1}{\sqrt{2\pi\sigma t}} \exp \left[ -\frac{(\ln t - \mu)^2}{2\sigma^2} \right] dt}{\frac{1}{\sqrt{2\pi\sigma x}} \exp \left[ -\frac{(\ln x - \mu)^2}{2\sigma^2} \right]} \right\}' = \lim_{x \uparrow \infty} \frac{1}{\left\{ \frac{1}{\sqrt{2\pi\sigma x}} \exp \left[ -\frac{(\ln x - \mu)^2}{2\sigma^2} \right] \right\}^2}$$

$$\begin{aligned}
& \times \left\{ - \left\{ \frac{1}{\sqrt{2\pi}\sigma x} \exp \left[ -\frac{(\ln x - \mu)^2}{2\sigma^2} \right] \right\}^2 - \int_x^\infty \frac{1}{\sqrt{2\pi}\sigma t} \exp \left[ -\frac{(\ln t - \mu)^2}{2\sigma^2} \right] dt \right. \\
& \times \left. \left\{ -\frac{1}{\sqrt{2\pi}\sigma x^2} \exp \left[ -\frac{(\ln x - \mu)^2}{2\sigma^2} \right] - \frac{(\ln x - \mu)}{\sqrt{2\pi}\sigma^3 x^2} \exp \left[ -\frac{(\ln x - \mu)^2}{2\sigma^2} \right] \right\} \right\} \\
& = \lim_{x \uparrow \infty} \left\{ -1 + \sqrt{(2\pi)}\sigma \exp \left[ \frac{(\ln x - \mu)^2}{2\sigma^2} \right] \int_x^\infty \frac{1}{\sqrt{2\pi}\sigma t} \exp \left[ -\frac{(\ln t - \mu)^2}{2\sigma^2} \right] dt \right. \\
& \quad \left. + \frac{\sqrt{(2\pi)}}{\sigma} (\ln x - \mu) \exp \left[ \frac{(\ln x - \mu)^2}{2\sigma^2} \right] \int_x^\infty \frac{1}{\sqrt{2\pi}\sigma t} \exp \left[ -\frac{(\ln t - \mu)^2}{2\sigma^2} \right] dt \right\} \\
& = -1 + \sqrt{(2\pi)}\sigma \lim_{x \uparrow \infty} \left\{ \frac{\int_x^\infty \frac{1}{\sqrt{2\pi}\sigma t} \exp \left[ -\frac{(\ln t - \mu)^2}{2\sigma^2} \right] dt}{\exp \left[ -\frac{(\ln x - \mu)^2}{2\sigma^2} \right]} \right\} \\
& \quad + \frac{\sqrt{(2\pi)}}{\sigma} \lim_{x \uparrow \infty} \left\{ \frac{\int_x^\infty \frac{1}{\sqrt{2\pi}\sigma t} \exp \left[ -\frac{(\ln t - \mu)^2}{2\sigma^2} \right] dt}{\frac{\exp \left[ -\frac{(\ln x - \mu)^2}{2\sigma^2} \right]}{\ln x - \mu}} \right\} \\
& = -1 + \lim_{x \uparrow \infty} \left\{ \frac{-\frac{1}{x} \exp \left[ -\frac{(\ln x - \mu)^2}{2\sigma^2} \right]}{-\frac{1}{\sigma^2 x} (\ln x - \mu) \exp \left[ -\frac{(\ln x - \mu)^2}{2\sigma^2} \right]} \right\} \\
& \quad + \frac{1}{\sigma^2} \lim_{x \uparrow \infty} \left\{ \frac{-\frac{1}{x} \exp \left[ -\frac{(\ln x - \mu)^2}{2\sigma^2} \right]}{-\frac{1}{x} \exp \left[ -\frac{(\ln x - \mu)^2}{2\sigma^2} \right] \left[ \frac{1}{\sigma^2} - \frac{1}{(\ln x - \mu)^2} \right]} \right\} \\
& = -1 + \sigma^2 \lim_{x \uparrow \infty} \left( \frac{1}{\ln x - \mu} \right) + \frac{1}{\sigma^2} \lim_{x \uparrow \infty} \left\{ \frac{1}{\left[ \frac{1}{\sigma^2} - \frac{1}{(\ln x - \mu)^2} \right]} \right\} = -1 + \frac{1}{\sigma^2} \sigma^2 = 0.
\end{aligned}$$

□

## Rozdělení z oboru atraktivity Fréchetova rozdělení

**Věta A.7.** *Uvažujme Paretovo rozdělení s parametrem  $\alpha > 0$ , distribuční funkci*

$$F(x) = \begin{cases} 1 - x^{-\alpha} & \text{pro } x > 1, \\ 0 & \text{pro } x \leq 1, \end{cases}$$

a hustotou

$$f(x) = \begin{cases} \alpha x^{-\alpha-1} & \text{pro } x > 1, \\ 0 & \text{pro } x \leq 1. \end{cases}$$

Pak EV index tohoto rozdělení je  $\gamma = \frac{1}{\alpha}$ .

*Důkaz.* Dosazením hustoty a distribuční funkce do vztahu (1.26) a postupnými úpravami lze odvodit

$$\gamma = \lim_{x \uparrow \infty} \left( \frac{x^{-\alpha}}{\alpha x^{-\alpha-1}} \right)' = \frac{1}{\alpha}.$$

□

**Věta A.8.** Uvažujme zobecněné Paretovo rozdělení s parametry  $\xi > 0$ ,  $\sigma > 0$ , distribuční funkcí

$$F(x) = \begin{cases} 1 - \left(1 + \frac{\xi x}{\sigma}\right)^{-\frac{1}{\xi}} & \text{pro } x > 0, \\ 0 & \text{pro } x \leq 0, \end{cases}$$

a hustotou

$$f(x) = \begin{cases} \frac{1}{\sigma} \left(1 + \frac{\xi x}{\sigma}\right)^{-\frac{1}{\xi}-1} & \text{pro } x > 0, \\ 0 & \text{pro } x \leq 0. \end{cases}$$

Pak EV index tohoto rozdělení je  $\gamma = \xi$ .

*Důkaz.* Dosazením hustoty a distribuční funkce do vztahu (1.26) a postupnými úpravami lze odvodit

$$\gamma = \lim_{x \uparrow \infty} \left[ \frac{\left(1 + \frac{\xi x}{\sigma}\right)^{-\frac{1}{\xi}}}{\frac{1}{\sigma} \left(1 + \frac{\xi x}{\sigma}\right)^{-\frac{1}{\xi}-1}} \right]' = \lim_{x \uparrow \infty} \left( \frac{1 + \frac{\xi x}{\sigma}}{\frac{1}{\sigma}} \right)' = \xi.$$

□

**Věta A.9.** Uvažujme Burrovo (typ XII) rozdělení s parametry  $\eta > 0$ ,  $\lambda > 0$ ,  $\tau > 0$ , distribuční funkcí

$$F(x) = \begin{cases} 1 - \left(\frac{\eta}{\eta + x^\tau}\right)^\lambda & \text{pro } x > 0, \\ 0 & \text{pro } x \leq 0, \end{cases}$$

a hustotou

$$f(x) = \begin{cases} \frac{\lambda \tau x^{\tau-1}}{\eta + x^\tau} \left(\frac{\eta}{\eta + x^\tau}\right)^\lambda & \text{pro } x > 0, \\ 0 & \text{pro } x \leq 0. \end{cases}$$

Pak EV index tohoto rozdělení je  $\gamma = \frac{1}{\lambda \tau}$ .

*Důkaz.* Dosazením hustoty a distribuční funkce do vztahu (1.26) a postupnými úpravami lze odvodit

$$\begin{aligned} \gamma &= \lim_{x \uparrow \infty} \left[ \frac{\left(\frac{\eta}{\eta + x^\tau}\right)^\lambda}{\frac{\lambda \tau x^{\tau-1}}{\eta + x^\tau} \left(\frac{\eta}{\eta + x^\tau}\right)^\lambda} \right]' = \lim_{x \uparrow \infty} \left( \frac{\eta + x^\tau}{\lambda \tau x^{\tau-1}} \right)' = \lim_{x \uparrow \infty} \left( \frac{\eta}{\lambda \tau x^{\tau-1}} + \frac{x}{\lambda \tau} \right)' \\ &= \lim_{x \uparrow \infty} \left[ \frac{\eta(1 - \tau)}{\lambda \tau x^\tau} + \frac{1}{\lambda \tau} \right] = \frac{1}{\lambda \tau}. \end{aligned}$$

□

**Věta A.10.** Uvažujme Burrovo (typ III) rozdělení s parametry  $\eta > 0$ ,  $\lambda > 0$ ,  $\tau > 0$ , distribuční funkcí

$$F(x) = \begin{cases} \left(\frac{\eta}{\eta + x^{-\tau}}\right)^\lambda & \text{pro } x > 0, \\ 0 & \text{pro } x \leq 0, \end{cases}$$

a hustotou

$$f(x) = \begin{cases} \frac{\lambda\tau\left(\frac{\eta}{\eta+x^{-\tau}}\right)^\lambda x^{-\tau-1}}{\eta+x^{-\tau}} & \text{pro } x > 0, \\ 0 & \text{pro } x \leq 0. \end{cases}$$

Pak EV index tohoto rozdělení je  $\gamma = \frac{1}{\tau}$ .

Důkaz. Dosazením hustoty a distribuční funkce do vztahu (1.26) a postupnými úpravami lze odvodit

$$\begin{aligned} \gamma &= \lim_{x \uparrow \infty} \left[ \frac{1 - \left(\frac{\eta}{\eta+x^{-\tau}}\right)^\lambda}{\frac{\lambda\tau\left(\frac{\eta}{\eta+x^{-\tau}}\right)^\lambda x^{-\tau-1}}{\eta+x^{-\tau}}} \right]' = \lim_{x \uparrow \infty} \frac{1}{\left[ \frac{\lambda\tau\left(\frac{\eta}{\eta+x^{-\tau}}\right)^\lambda x^{-\tau-1}}{\eta+x^{-\tau}} \right]^2} \\ &\times \left\{ - \left[ \frac{\lambda\tau\left(\frac{\eta}{\eta+x^{-\tau}}\right)^\lambda x^{-\tau-1}}{\eta+x^{-\tau}} \right]^2 - \left[ 1 - \left(\frac{\eta}{\eta+x^{-\tau}}\right)^\lambda \right] \left[ \frac{\lambda^2\tau^2\left(\frac{\eta}{\eta+x^{-\tau}}\right)^\lambda x^{-2\tau-2}}{(\eta+x^{-\tau})^2} \right. \right. \\ &\left. \left. + \frac{\lambda\tau(-\tau-1)\left(\frac{\eta}{\eta+x^{-\tau}}\right)^\lambda x^{-\tau-2}(\eta+x^{-\tau}) + \lambda\tau^2\left(\frac{\eta}{\eta+x^{-\tau}}\right)^\lambda x^{-2\tau-2}}{(\eta+x^{-\tau})^2} \right] \right\} \\ &= \lim_{x \uparrow \infty} \left\{ -1 - \frac{1 - \left(\frac{\eta}{\eta+x^{-\tau}}\right)^\lambda}{\left(\frac{\eta}{\eta+x^{-\tau}}\right)^\lambda} + \frac{\left[ 1 - \left(\frac{\eta}{\eta+x^{-\tau}}\right)^\lambda \right] (\eta+x^{-\tau})}{\lambda x^{-\tau} \left(\frac{\eta}{\eta+x^{-\tau}}\right)^\lambda} \right. \\ &\left. + \frac{\left[ 1 - \left(\frac{\eta}{\eta+x^{-\tau}}\right)^\lambda \right] (\eta+x^{-\tau})}{\lambda\tau x^{-\tau} \left(\frac{\eta}{\eta+x^{-\tau}}\right)^\lambda} - \frac{1 - \left(\frac{\eta}{\eta+x^{-\tau}}\right)^\lambda}{\lambda \left(\frac{\eta}{\eta+x^{-\tau}}\right)^\lambda} \right\} \\ &= \lim_{x \uparrow \infty} \left[ - \left(\frac{\eta}{\eta+x^{-\tau}}\right)^{-\lambda} + \frac{(\eta+x^{-\tau})x^\tau \left(\frac{\eta}{\eta+x^{-\tau}}\right)^{-\lambda}}{\lambda} - \frac{(\eta+x^{-\tau})x^\tau}{\lambda} \right. \\ &\left. + \frac{(\eta+x^{-\tau})x^\tau \left(\frac{\eta}{\eta+x^{-\tau}}\right)^{-\lambda}}{\lambda\tau} - \frac{(\eta+x^{-\tau})x^\tau}{\lambda\tau} - \frac{\left(\frac{\eta}{\eta+x^{-\tau}}\right)^{-\lambda}}{\lambda} + \frac{1}{\lambda} \right] \\ &= \lim_{x \uparrow \infty} \frac{1}{\lambda\tau} \left[ -\lambda\tau \left(\frac{\eta}{\eta+x^{-\tau}}\right)^{-\lambda} + \eta\tau x^\tau \left(\frac{\eta}{\eta+x^{-\tau}}\right)^{-\lambda} + \tau \left(\frac{\eta}{\eta+x^{-\tau}}\right)^{-\lambda} - \eta\tau x^\tau \right. \\ &\left. - \tau + \eta x^\tau \left(\frac{\eta}{\eta+x^{-\tau}}\right)^{-\lambda} + \left(\frac{\eta}{\eta+x^{-\tau}}\right)^{-\lambda} - \eta x^\tau - 1 - \tau \left(\frac{\eta}{\eta+x^{-\tau}}\right)^{-\lambda} + \tau \right] \\ &= \frac{1}{\lambda\tau} \left\{ \lim_{x \uparrow \infty} \left[ \eta x^\tau \left(\frac{\eta}{\eta+x^{-\tau}}\right)^{-\lambda} - \eta x^\tau \right] + \lim_{x \uparrow \infty} \left(\frac{\eta}{\eta+x^{-\tau}}\right)^{-\lambda} - 1 \right. \\ &\left. - \lim_{x \uparrow \infty} \lambda\tau \left(\frac{\eta}{\eta+x^{-\tau}}\right)^{-\lambda} + \lim_{x \uparrow \infty} \left[ \eta\tau x^\tau \left(\frac{\eta}{\eta+x^{-\tau}}\right)^{-\lambda} - \eta\tau x^\tau \right] \right\} \end{aligned}$$



$$\begin{aligned}
&= \frac{1}{\lambda\tau} \left\{ \lim_{x \uparrow \infty} \left[ \eta x^\tau \left( \left( \frac{\eta}{\eta + x^{-\tau}} \right)^{-\lambda} - 1 \right) \right] - \lambda\tau + \lim_{x \uparrow \infty} \left[ \eta^\tau x^\tau \left( \left( \frac{\eta}{\eta + x^{-\tau}} \right)^{-\lambda} - 1 \right) \right] \right\} \\
&= \frac{1}{\lambda\tau} \left\{ (\eta + \eta^\tau) \lim_{x \uparrow \infty} \left[ \frac{\left( \frac{\eta}{\eta + x^{-\tau}} \right)^{-\lambda} - 1}{\frac{1}{x^\tau}} \right] - \lambda\tau \right\} \\
&= \frac{1}{\lambda\tau} \left\{ (\eta + \eta^\tau) \lim_{x \uparrow \infty} \left[ \frac{\lambda \left( \frac{\eta}{\eta + x^{-\tau}} \right)^{-\lambda}}{\eta + x^{-\tau}} \right] - \lambda\tau \right\} = \frac{1}{\lambda\tau} \left[ (\eta + \eta^\tau) \frac{\lambda}{\eta} - \lambda\tau \right] = \frac{1}{\tau}.
\end{aligned}$$

□

**Věta A.11.** *Uvažujme  $F$  rozdělení s parametry  $m > 0$ ,  $n > 0$ , distribuční funkcí*

$$F(x) = \begin{cases} 1 - \int_x^\infty \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \left(\frac{m}{n}\right)^{\frac{m}{2}} t^{\frac{m}{2}-1} \left(1 + \frac{m}{n}t\right)^{-\frac{m+n}{2}} dt & \text{pro } x > 0, \\ 0 & \text{pro } x \leq 0, \end{cases}$$

a hustotou

$$f(x) = \begin{cases} \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \left(\frac{m}{n}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1} \left(1 + \frac{m}{n}x\right)^{-\frac{m+n}{2}} & \text{pro } x > 0, \\ 0 & \text{pro } x \leq 0. \end{cases}$$

Pak EV index tohoto rozdělení je  $\gamma = \frac{2}{n}$ .

*Důkaz.* Dosazením hustoty a distribuční funkce do vztahu (1.26) a postupnými úpravami lze odvodit

$$\begin{aligned}
\gamma &= \lim_{x \uparrow \infty} \left[ \frac{\int_x^\infty \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \left(\frac{m}{n}\right)^{\frac{m}{2}} t^{\frac{m}{2}-1} \left(1 + \frac{m}{n}t\right)^{-\frac{m+n}{2}} dt}{\frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \left(\frac{m}{n}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1} \left(1 + \frac{m}{n}x\right)^{-\frac{m+n}{2}}} \right]' \\
&= \lim_{x \uparrow \infty} \frac{1}{\left[ \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \left(\frac{m}{n}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1} \left(1 + \frac{m}{n}x\right)^{-\frac{m+n}{2}} \right]^2} \\
&\quad \times \left\{ - \left[ \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \left(\frac{m}{n}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1} \left(1 + \frac{m}{n}x\right)^{-\frac{m+n}{2}} \right]^2 \right. \\
&\quad - \left[ \int_x^\infty \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \left(\frac{m}{n}\right)^{\frac{m}{2}} t^{\frac{m}{2}-1} \left(1 + \frac{m}{n}t\right)^{-\frac{m+n}{2}} dt \right] \\
&\quad \times \left[ \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \left(\frac{m}{n}\right)^{\frac{m}{2}} \left(\frac{m}{2} - 1\right) x^{\frac{m}{2}-2} \left(1 + \frac{m}{n}x\right)^{-\frac{m+n}{2}} \right. \\
&\quad \left. \left. + \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \left(\frac{m}{n}\right)^{\frac{m}{2}+1} \left(-\frac{m+n}{2}\right) x^{\frac{m}{2}-1} \left(1 + \frac{m}{n}x\right)^{-\frac{m+n}{2}-1} \right] \right\} \\
&= \lim_{x \uparrow \infty} \left\{ -1 - \frac{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})}{\Gamma(\frac{m+n}{2})} \left(\frac{m}{n}\right)^{-\frac{m}{2}} \left(\frac{m}{2} - 1\right) x^{-\frac{m}{2}} \left(1 + \frac{m}{n}x\right)^{\frac{m+n}{2}} \right\}
\end{aligned}$$

$$\begin{aligned}
& \times \left[ \int_x^\infty \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right)^{\frac{m}{2}} t^{\frac{m}{2}-1} \left(1 + \frac{m}{n}t\right)^{-\frac{m+n}{2}} dt \right] \\
& - \frac{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{m+n}{2}\right)} \left(\frac{m}{n}\right)^{-\frac{m}{2}+1} \left(\frac{m+n}{2}\right) x^{-\frac{m}{2}+1} \left(1 + \frac{m}{n}x\right)^{\frac{m+n}{2}-1} \\
& \times \left[ \int_x^\infty \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right)^{\frac{m}{2}} t^{\frac{m}{2}-1} \left(1 + \frac{m}{n}t\right)^{-\frac{m+n}{2}} dt \right] \Big\} \\
= & -1 - \frac{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{m+n}{2}\right)} \left(\frac{m}{n}\right)^{-\frac{m}{2}} \left(\frac{m}{2} - 1\right) \\
& \times \lim_{x \uparrow \infty} \left[ \frac{\int_x^\infty \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right)^{\frac{m}{2}} t^{\frac{m}{2}-1} \left(1 + \frac{m}{n}t\right)^{-\frac{m+n}{2}} dt}{\frac{x^{\frac{m}{2}}}{\left(1 + \frac{m}{n}x\right)^{\frac{m+n}{2}}}} \right] \\
& - \frac{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{m+n}{2}\right)} \left(\frac{m}{n}\right)^{-\frac{m}{2}+1} \left(\frac{m+n}{2}\right) \\
& \times \lim_{x \uparrow \infty} \left[ \frac{\int_x^\infty \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right)^{\frac{m}{2}} t^{\frac{m}{2}-1} \left(1 + \frac{m}{n}t\right)^{-\frac{m+n}{2}} dt}{\frac{x^{\frac{m}{2}}\left(1 + \frac{m}{n}x\right)}{x\left(1 + \frac{m}{n}x\right)^{\frac{m+n}{2}}}} \right] \\
= & -1 - \left(\frac{m}{2} - 1\right) \lim_{x \uparrow \infty} \left\{ \frac{-x^{\frac{m}{2}-1} \left(1 + \frac{m}{n}x\right)^{-\frac{m+n}{2}}}{\left(1 + \frac{m}{n}x\right)^{-\frac{m+n}{2}} \left[ \frac{m}{2}x^{\frac{m}{2}-1} - \frac{m}{n} \left(\frac{m+n}{2}\right) x^{\frac{m}{2}} \left(1 + \frac{m}{n}x\right)^{-1} \right]} \right\} \\
& - \frac{m}{n} \left(\frac{m+n}{2}\right) \\
& \times \lim_{x \uparrow \infty} \left\{ \frac{-x^{\frac{m}{2}-1} \left(1 + \frac{m}{n}x\right)^{-\frac{m+n}{2}}}{\left(1 + \frac{m}{n}x\right)^{-\frac{m+n}{2}} \left[ \left(\frac{m}{2} - 1\right) x^{\frac{m}{2}-2} \left(1 + \frac{m}{n}x\right) + \frac{m}{n} x^{\frac{m}{2}-1} \left(1 - \frac{m+n}{2}\right) \right]} \right\} \\
= & -1 + \left(\frac{m}{2} - 1\right) \lim_{x \uparrow \infty} \left( \frac{x^{-1} + \frac{m}{n}}{\frac{m}{2}x^{-1} - \frac{m}{2}} \right) + \frac{m}{n} \left(\frac{m+n}{2}\right) \lim_{x \uparrow \infty} \left[ \frac{1}{\left(\frac{m}{2} - 1\right) x^{-1} - \frac{m}{2}} \right] \\
= & -1 - \left(\frac{m}{2} - 1\right) \frac{2}{n} - \left(\frac{m}{n}\right) \left(\frac{m+n}{2}\right) \frac{2}{m} = \frac{2}{n}.
\end{aligned}$$

□

**Věta A.12.** Uvažujme inverzní gama rozdělení s parametry  $\alpha > 0$ ,  $\lambda > 0$ , distribuční funkcí

$$F(x) = \begin{cases} 1 - \int_x^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{-\alpha-1} \exp\left(-\frac{\lambda}{t}\right) dt & \text{pro } x > 0, \\ 0 & \text{pro } x \leq 0, \end{cases}$$

a hustotou

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(-\frac{\lambda}{x}\right) & \text{pro } x > 0, \\ 0 & \text{pro } x \leq 0. \end{cases}$$

Pak EV index tohoto rozdělení je  $\gamma = \frac{1}{\alpha}$ .

*Důkaz.* Dosazením hustoty a distribuční funkce do vztahu (1.26) a postupnými úpravami lze odvodit

$$\begin{aligned}
\gamma &= \lim_{x \uparrow \infty} \left[ \frac{\int_x^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{-\alpha-1} \exp\left(-\frac{\lambda}{x}\right) dt}{\frac{\lambda^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(-\frac{\lambda}{x}\right)} \right]' = \lim_{x \uparrow \infty} \frac{1}{\left[ \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(-\frac{\lambda}{x}\right) \right]^2} \\
&\times \left\{ - \left[ \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(-\frac{\lambda}{x}\right) \right]^2 - \left[ \int_x^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{-\alpha-1} \exp\left(-\frac{\lambda}{x}\right) dt \right] \right. \\
&\times \left. \left[ \frac{(-\alpha-1) \lambda^\alpha x^{-\alpha-2} \exp\left(-\frac{\lambda}{x}\right)}{\Gamma(\alpha)} + \frac{\lambda^{\alpha+1} x^{-\alpha-3} \exp\left(-\frac{\lambda}{x}\right)}{\Gamma(\alpha)} \right] \right\} \\
&= \lim_{x \uparrow \infty} \left\{ -1 + (\alpha+1) \lambda^{-\alpha} \Gamma(\alpha) x^\alpha \exp\left(\frac{\lambda}{x}\right) \left[ \int_x^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{-\alpha-1} \exp\left(-\frac{\lambda}{x}\right) dt \right] \right. \\
&\quad \left. - \lambda^{-\alpha+1} \Gamma(\alpha) x^{\alpha-1} \exp\left(\frac{\lambda}{x}\right) \left[ \int_x^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{-\alpha-1} \exp\left(-\frac{\lambda}{x}\right) dt \right] \right\} \\
&= -1 + (\alpha+1) \lim_{x \uparrow \infty} \left[ \frac{\int_x^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{-\alpha-1} \exp\left(-\frac{\lambda}{x}\right) dt}{\frac{1}{\lambda^{-\alpha} \Gamma(\alpha) x^\alpha \exp\left(\frac{\lambda}{x}\right)}} \right] - \lim_{x \uparrow \infty} \left[ \frac{\int_x^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{-\alpha-1} \exp\left(-\frac{\lambda}{x}\right) dt}{\frac{1}{\lambda^{-\alpha+1} \Gamma(\alpha) x^{\alpha-1} \exp\left(\frac{\lambda}{x}\right)}} \right] \\
&= -1 + (\alpha+1) \lim_{x \uparrow \infty} \left( \frac{1}{\alpha - \frac{\lambda}{x}} \right) - \lim_{x \uparrow \infty} \left( \frac{\frac{\lambda}{x}}{\alpha - 1 - \frac{\lambda}{x}} \right) = -1 + (\alpha+1) \frac{1}{\alpha} = \frac{1}{\alpha}.
\end{aligned}$$

□

**Věta A.13.** Uvažujme logaritmické gama rozdělení s parametry  $\alpha > 0$ ,  $\lambda > 0$ , distribuční funkcí

$$F(x) = \begin{cases} 1 - \int_x^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} (\ln t)^{\alpha-1} t^{-\lambda-1} dt & \text{pro } x > 1, \\ 0 & \text{pro } x \leq 1, \end{cases}$$

a hustotou

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} (\ln x)^{\alpha-1} x^{-\lambda-1} & \text{pro } x > 1, \\ 0 & \text{pro } x \leq 1. \end{cases}$$

Pak EV index tohoto rozdělení je  $\gamma = \frac{1}{\lambda}$ .

*Důkaz.* Dosazením hustoty a distribuční funkce do vztahu (1.26) a postupnými úpravami lze odvodit

$$\begin{aligned}
\gamma &= \lim_{x \uparrow \infty} \left[ \frac{\int_x^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} (\ln t)^{\alpha-1} t^{-\lambda-1} dt}{\frac{\lambda^\alpha}{\Gamma(\alpha)} (\ln x)^{\alpha-1} x^{-\lambda-1}} \right]' = \lim_{x \uparrow \infty} \frac{1}{\left[ \frac{\lambda^\alpha}{\Gamma(\alpha)} (\ln x)^{\alpha-1} x^{-\lambda-1} \right]^2} \\
&\times \left\{ - \left[ \frac{\lambda^\alpha}{\Gamma(\alpha)} (\ln x)^{\alpha-1} x^{-\lambda-1} \right]^2 - \left[ \int_x^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} (\ln t)^{\alpha-1} t^{-\lambda-1} dt \right] \right. \\
&\times \left. \left[ \frac{(\alpha-1) \lambda^\alpha (\ln x)^{\alpha-2} x^{-\lambda-2}}{\Gamma(\alpha)} + \frac{(-\lambda-1) \lambda^\alpha (\ln x)^{\alpha-1} x^{-\lambda-2}}{\Gamma(\alpha)} \right] \right\}
\end{aligned}$$

$$\begin{aligned}
&= \lim_{x \uparrow \infty} \left\{ -1 - (\alpha - 1) \lambda^{-\alpha} \Gamma(\alpha) (\ln x)^{-\alpha} x^\lambda \left[ \int_x^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} (\ln t)^{\alpha-1} t^{-\lambda-1} dt \right] \right. \\
&\quad \left. + (\lambda + 1) \lambda^{-\alpha} \Gamma(\alpha) (\ln x)^{-\alpha+1} x^\lambda \left[ \int_x^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} (\ln t)^{\alpha-1} t^{-\lambda-1} dt \right] \right\} \\
&= -1 - (\alpha - 1) \lim_{x \uparrow \infty} \left[ \frac{\int_x^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} (\ln t)^{\alpha-1} t^{-\lambda-1} dt}{\frac{1}{\lambda^{-\alpha} \Gamma(\alpha) (\ln x)^{-\alpha} x^\lambda}} \right] \\
&\quad + (\lambda + 1) \lim_{x \uparrow \infty} \left[ \frac{\int_x^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} (\ln t)^{\alpha-1} t^{-\lambda-1} dt}{\frac{1}{\lambda^{-\alpha} \Gamma(\alpha) (\ln x)^{-\alpha+1} x^\lambda}} \right] \\
&= -1 - (\alpha - 1) \lim_{x \uparrow \infty} \left( \frac{\frac{1}{\ln x}}{\lambda - \frac{\alpha}{\ln x}} \right) + (\lambda + 1) \lim_{x \uparrow \infty} \left( \frac{1}{\frac{1-\alpha}{\ln x} + \lambda} \right) = -1 + (\lambda + 1) \frac{1}{\lambda} = \frac{1}{\lambda}.
\end{aligned}$$

□

**Věta A.14.** *Uvažujme Fréchetovo rozdělení s parametrem  $\alpha > 0$ , distribuční funkcí*

$$F(x) = \begin{cases} \exp(-x^{-\alpha}) & \text{pro } x > 0, \\ 0 & \text{pro } x \leq 0, \end{cases}$$

a hustotou

$$f(x) = \begin{cases} \alpha x^{-\alpha-1} \exp(-x^{-\alpha}) & \text{pro } x > 0, \\ 0 & \text{pro } x \leq 0. \end{cases}$$

Pak EV index tohoto rozdělení je  $\gamma = \frac{1}{\alpha}$ .

*Důkaz.* Dosazením hustoty a distribuční funkce do vztahu (1.26) a postupnými úpravami lze odvodit

$$\begin{aligned}
\gamma &= \lim_{x \uparrow \infty} \left[ \frac{1 - \exp(-x^{-\alpha})}{\alpha x^{-\alpha-1} \exp(-x^{-\alpha})} \right]' = \lim_{x \uparrow \infty} \frac{1}{[\alpha x^{-\alpha-1} \exp(-x^{-\alpha})]^2} \\
&\quad \times \left\{ -[\alpha x^{-\alpha-1} \exp(-x^{-\alpha})]^2 - [1 - \exp(-x^{-\alpha})] \right. \\
&\quad \left. \times \left[ \alpha(-\alpha - 1)x^{-\alpha-2} \exp(-x^{-\alpha}) + \alpha^2 x^{-2\alpha-2} \exp(-x^{-\alpha}) \right] \right\} \\
&= \lim_{x \uparrow \infty} \left\{ -1 + \frac{\alpha + 1}{\alpha} x^\alpha \exp(x^{-\alpha}) [1 - \exp(-x^{-\alpha})] - \exp(x^{-\alpha}) [1 - \exp(-x^{-\alpha})] \right\} \\
&= -1 + \frac{\alpha + 1}{\alpha} \lim_{x \uparrow \infty} \left[ \frac{1 - \exp(-x^{-\alpha})}{\frac{1}{x^\alpha \exp(x^{-\alpha})}} \right] - \lim_{x \uparrow \infty} \{ \exp(x^{-\alpha}) [1 - \exp(-x^{-\alpha})] \} \\
&= -1 + \frac{\alpha + 1}{\alpha} \lim_{x \uparrow \infty} \left( \frac{1}{1 - x^{-\alpha}} \right) = -1 + \frac{\alpha + 1}{\alpha} = \frac{1}{\alpha}.
\end{aligned}$$

□

## Rozdělení z oboru atraktivity Weibullova rozdělení

**Věta A.15.** *Uvažujme rovnoměrné rozdělení s distribuční funkcí*

$$F(x) = \begin{cases} 1 & \text{pro } x \geq x^*, \\ x - x^* + 1 & \text{pro } x^* - 1 < x < x^*, \\ 0 & \text{pro } x \leq x^* - 1, \end{cases}$$

a hustotou

$$f(x) = \begin{cases} 1 & \text{pro } x^* - 1 < x < x^*, \\ 0 & \text{jinde.} \end{cases}$$

Pak EV index tohoto rozdělení je  $\gamma = -1$ .

*Důkaz.* Dosazením hustoty a distribuční funkce do vztahu (1.26) a postupnými úpravami lze odvodit

$$\gamma = \lim_{x \uparrow x^*} \left( \frac{x^* - x}{1} \right)' = -1.$$

□

**Věta A.16.** *Uvažujme beta rozdělení s parametry  $p > 0$ ,  $q > 0$ , distribuční funkcí*

$$F(x) = \begin{cases} 1 & \text{pro } x \geq x^*, \\ 1 - \int_{1-(x^*-x)}^1 \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} t^{p-1} (1-t)^{q-1} dt & \text{pro } x^* - 1 < x < x^*, \\ 0 & \text{pro } x \leq x^* - 1, \end{cases}$$

a hustotou

$$f(x) = \begin{cases} \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} [1 - (x^* - x)]^{p-1} (x^* - x)^{q-1} & \text{pro } x^* - 1 < x < x^*, \\ 0 & \text{jinde.} \end{cases}$$

Pak EV index tohoto rozdělení je  $\gamma = -\frac{1}{q}$ .

*Důkaz.* Dosazením hustoty a distribuční funkce do vztahu (1.26) a postupnými úpravami lze odvodit

$$\begin{aligned} \gamma &= \lim_{x \uparrow x^*} \left\{ \frac{\int_{1-(x^*-x)}^1 \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} t^{p-1} (1-t)^{q-1} dt}{\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} [1 - (x^* - x)]^{p-1} (x^* - x)^{q-1}} \right\}' \\ &= \lim_{x \uparrow x^*} \frac{1}{\left\{ \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} [1 - (x^* - x)]^{p-1} (x^* - x)^{q-1} \right\}^2} \\ &\quad \times \left\{ - \left\{ \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} [1 - (x^* - x)]^{p-1} (x^* - x)^{q-1} \right\}^2 \right. \\ &\quad \left. - \int_{1-(x^*-x)}^1 \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} t^{p-1} (1-t)^{q-1} dt \left\{ \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} (p-1) [1 - (x^* - x)]^{p-2} (x^* - x)^{q-1} \right. \right. \end{aligned}$$

$$\begin{aligned}
& - \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} (q-1) [1 - (x^* - x)]^{p-1} (x^* - x)^{q-2} \Big\} \Big\} \\
= & \lim_{x \uparrow x^*} \left\{ -1 - \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} (p-1) [1 - (x^* - x)]^{-p} (x^* - x)^{-q+1} \right. \\
& \times \int_{1-(x^*-x)}^1 \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} t^{p-1} (1-t)^{q-1} dt + \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} (q-1) [1 - (x^* - x)]^{-p+1} (x^* - x)^{-q} \\
& \times \left. \int_{1-(x^*-x)}^1 \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} t^{p-1} (1-t)^{q-1} dt \right\} \\
= & -1 - \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} (p-1) \lim_{x \uparrow x^*} \left\{ \frac{(x^* - x) \int_{1-(x^*-x)}^1 \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} t^{p-1} (1-t)^{q-1} dt}{(x^* - x)^q} \right\} \\
& + \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} (q-1) \lim_{x \uparrow x^*} \left\{ \frac{\int_{1-(x^*-x)}^1 \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} t^{p-1} (1-t)^{q-1} dt}{(x^* - x)^q} \right\} \\
= & -1 - \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} (p-1) \lim_{x \uparrow x^*} \left\{ \frac{-\int_{1-(x^*-x)}^1 \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} t^{p-1} (1-t)^{q-1} dt}{-q (x^* - x)^{q-1}} \right. \\
& \left. + \frac{-\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} (x^* - x)^q [1 - (x^* - x)]^{p-1}}{-q (x^* - x)^{q-1}} \right\} \\
& + (q-1) \lim_{x \uparrow x^*} \left\{ \frac{-[1 - (x^* - x)]^{p-1} (x^* - x)^{q-1}}{-q (x^* - x)^{q-1}} \right\} \\
= & -1 - (p-1) \lim_{x \uparrow x^*} \left\{ \frac{-[1 - (x^* - x)]^{p-1} (x^* - x)^{q-1}}{-q(q-1) (x^* - x)^{q-2}} \right\} + (q-1) \frac{1}{q} \\
= & -1 + (q-1) \frac{1}{q} = -\frac{1}{q}.
\end{aligned}$$

□

**Věta A.17.** Uvažujme reverzní Burrovo rozdělení s parametry  $\beta > 0$ ,  $\lambda > 0$ ,  $\tau > 0$ , distribuční funkcí

$$F(x) = \begin{cases} 1 - \left[ \frac{\beta}{\beta + (x^* - x)^{-\tau}} \right]^\lambda & \text{pro } x < x^*, \\ 1 & \text{pro } x \geq x^*, \end{cases}$$

a hustotou

$$f(x) = \begin{cases} \frac{\lambda \tau (x^* - x)^{-\tau-1}}{\beta + (x^* - x)^{-\tau}} \left[ \frac{\beta}{\beta + (x^* - x)^{-\tau}} \right]^\lambda & \text{pro } x < x^*, \\ 0 & \text{pro } x \geq x^*. \end{cases}$$

Pak EV index tohoto rozdělení je  $\gamma = -\frac{1}{\lambda\tau}$ .

*Důkaz.* Dosazením hustoty a distribuční funkce do vztahu (1.26) a postupnými úpravami lze odvodit

$$\gamma = \lim_{x \uparrow x^*} \left\{ \frac{\left[ \frac{\beta}{\beta + (x^* - x)^{-\tau}} \right]^\lambda}{\frac{\lambda \tau (x^* - x)^{-\tau-1}}{\beta + (x^* - x)^{-\tau}} \left[ \frac{\beta}{\beta + (x^* - x)^{-\tau}} \right]^\lambda} \right\}' = \lim_{x \uparrow x^*} \left[ \frac{\beta + (x^* - x)^{-\tau}}{\lambda \tau (x^* - x)^{-\tau-1}} \right]'$$

$$= \lim_{x \uparrow x^*} \left[ \frac{\beta}{\lambda \tau (x^* - x)^{-\tau-1}} + \frac{(x^* - x)}{\lambda \tau} \right]' = \lim_{x \uparrow x^*} \left[ -\frac{\beta(\tau + 1)(x^* - x)^\tau}{\lambda \tau} - \frac{1}{\lambda \tau} \right] = -\frac{1}{\lambda \tau}.$$

□

**Věta A.18.** *Uvažujme extrémální Weibullovo rozdělení s parametrem  $\alpha > 0$ , distribuční funkci*

$$F(x) = \begin{cases} \exp[-(x^* - x)^\alpha] & \text{pro } x < x^*, \\ 1 & \text{pro } x \geq x^*, \end{cases}$$

a hustotou

$$f(x) = \begin{cases} \alpha (x^* - x)^{\alpha-1} \exp[-(x^* - x)^\alpha] & \text{pro } x < x^*, \\ 0 & \text{pro } x \geq x^*. \end{cases}$$

Pak EV index tohoto rozdělení je  $\gamma = -\frac{1}{\alpha}$ .

*Důkaz.* Dosazením hustoty a distribuční funkce do vztahu (1.26) a postupnými úpravami lze odvodit

$$\begin{aligned} \gamma &= \lim_{x \uparrow x^*} \left\{ \frac{1 - \exp[-(x^* - x)^\alpha]}{\alpha (x^* - x)^{\alpha-1} \exp[-(x^* - x)^\alpha]} \right\}' = \lim_{x \uparrow x^*} \frac{1}{\{\alpha (x^* - x)^{\alpha-1} \exp[-(x^* - x)^\alpha]\}^2} \\ &\quad \times \left\{ -\{\alpha (x^* - x)^{\alpha-1} \exp[-(x^* - x)^\alpha]\}^2 - \{1 - \exp[-(x^* - x)^\alpha]\} \right. \\ &\quad \left. \times \left\{ -\alpha(\alpha - 1)(x^* - x)^{\alpha-2} \exp[-(x^* - x)^\alpha] + \alpha^2 (x^* - x)^{2\alpha-2} \exp[-(x^* - x)^\alpha] \right\} \right\} \\ &= \lim_{x \uparrow x^*} \left\{ -1 + \frac{\alpha - 1}{\alpha} (x^* - x)^{-\alpha} \exp[(x^* - x)^\alpha] \{1 - \exp[-(x^* - x)^\alpha]\} \right. \\ &\quad \left. - \exp[(x^* - x)^\alpha] \{1 - \exp[-(x^* - x)^\alpha]\} \right\} \\ &= -1 + \frac{\alpha - 1}{\alpha} \lim_{x \uparrow x^*} \left\{ \frac{\exp[(x^* - x)^\alpha] \{1 - \exp[-(x^* - x)^\alpha]\}}{(x^* - x)^\alpha} \right\} \\ &\quad - \lim_{x \uparrow x^*} \left\{ \exp[(x^* - x)^\alpha] \{1 - \exp[-(x^* - x)^\alpha]\} \right\} \\ &= -1 + \frac{\alpha - 1}{\alpha} \lim_{x \uparrow x^*} \exp[(x^* - x)^\alpha] = -1 + \frac{\alpha - 1}{\alpha} = -\frac{1}{\alpha}. \end{aligned}$$

□





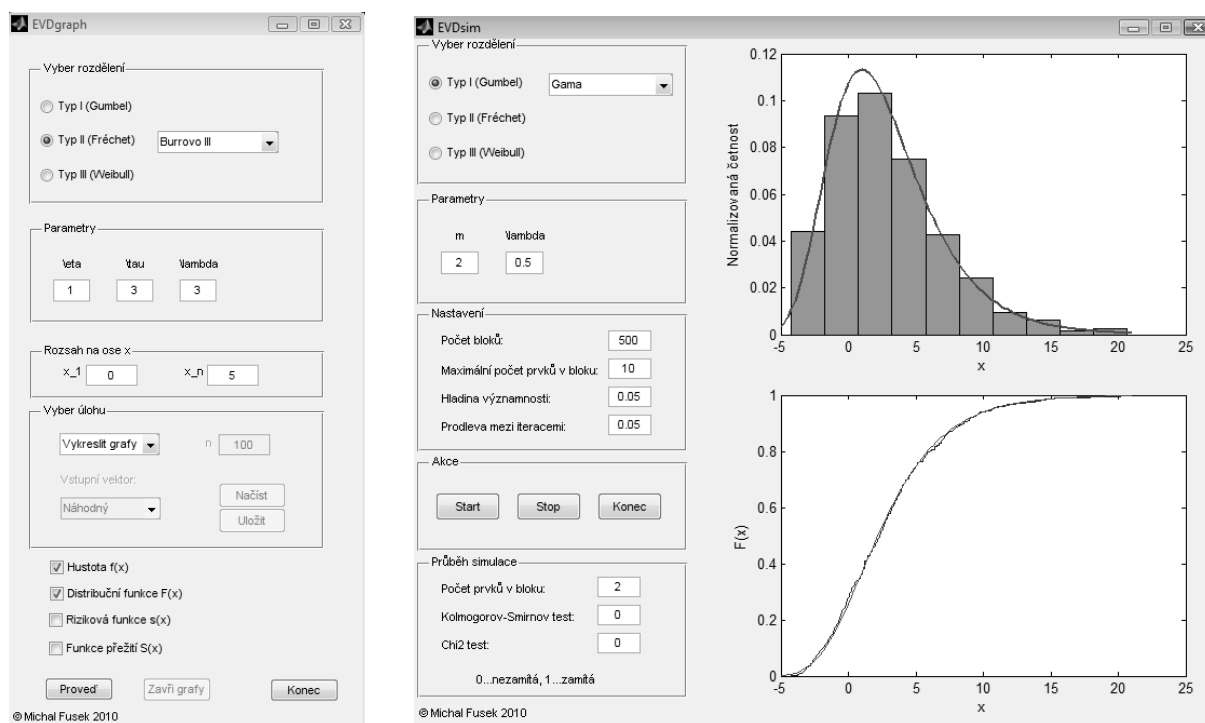
## B Software EVDgraph a EVDsim

Tato kapitola vychází z autorovy práce [32] a je věnována popisu vytvořených demonstračních softwarů pro rozdělení extrémních hodnot. Software *EVDgraph* umožňuje uživateli získat základní představu o tvaru EV rozdělení a o rozděleních z jejich oborů atraktivity. Pomocí softwaru *EVDsim* je možné posoudit rychlost konvergence daného rozdělení k příslušnému limitnímu EV rozdělení. Dále je možné získat základní představu, jak velké musí být bloky, z nichž se počítají bloková maxima, a jaký musí být jejich počet, aby konvergence k limitnímu EV rozdělení byla dostatečně kvalitní.

Ke správné funkčnosti programů je nutný Matlab se statistickým toolboxem (testováno na verzi R2011a). Oba produkty jsou dostupné na adrese:

<http://dl.dropbox.com/u/32398211/EVDgraph.zip>,

<http://dl.dropbox.com/u/32398211/EVDsim.zip>.



(a) EVDgraph

(b) EVDsim

Obrázek B.1: Demonstrační programy.

### Program EVDgraph

Program EVDgraph (viz obrázek B.1(a)) slouží ke grafickému znázornění rozdělení z oboru atraktivity EV rozdělení, která jsou uvedena v tabulkách 1.1–1.3.

## Spuštění

Po spuštění Matlabu se do příkazového řádku napíše `EVDgraph`. Je však nutné se nacházet v adresáři s programem. Dále budou popsány jednotlivé funkce.

## Popis programu

Program je rozčleněn na několik částí. V sekci *Vyber rozdělení* lze zvolit jeden ze tří typů EV rozdělení (Gumbelovo, Fréchetovo, Weibullovo). V závislosti na volbě se otevře podnabídka rozdělení patřících do oboru atraktivity zvoleného EV rozdělení. V sekci *Parametry* se zobrazí parametry příslušné danému rozdělení, jejichž hodnoty lze podle potřeby měnit. V sekci *Rozsah na ose x* lze zvolit rozsah na ose  $x$  při vykreslování grafů. Konečně v sekci *Vyber úlohu* lze pro daná rozdělení zvolit následující úlohy:

### (a) Vykreslit grafy

Vykreslí zvolené funkce, tedy hustotu  $f(x)$ , distribuční funkci  $F(x)$ , rizikovou funkci  $s(x) = f(x)/(1 - F(x))$ , případně funkci přežití  $S(x) = 1 - F(x)$ , vybraného rozdělení s příslušnými parametry.

### (b) Histogram

Vykreslí histogram ze vstupního datového vektoru a proloží jej hustotou zvoleného rozdělení. Neznámé parametry rozdělení jsou odhadnuty metodou maximální věrohodnosti ze vstupního souboru a jejich hodnoty se zobrazí v sekci *Parametry*. Ve většině případů je výpočet MV odhadů numericky stabilní. Nicméně se může stát, že při některých hodnotách parametrů, jejichž odhad vyžaduje speciální numerické postupy, se mohou kvůli nízké podmíněnosti úlohy objevit numerické obtíže.

Jako vstupní vektor je možné vybrat buď náhodný výběr rozsahu  $n$  ze zvoleného rozdělení, nebo načíst data z externího `txt` souboru. Data v souboru musí být ve sloupci a desetinná místa musí být oddělena tečkou místo čárky. Pokud se zvolí vstupní vektor náhodný, program vygeneruje náhodný výběr ze zvoleného rozdělení, který je možné po vykreslení histogramu uložit do externího `txt` souboru a uchovat jej pro pozdější využití. Program tedy může sloužit i jako generátor náhodných čísel z vybraného rozdělení s příslušnými parametry. Náhodná, přesněji pseudonáhodná, čísla jsou generována s využitím kvantilové funkce a pseudonáhodných čísel z rovnoměrného rozdělení (metoda inverse transform sampling).

### (c) Q-Q plot

Vykreslí Q-Q plot, který porovnává závislost teoretických kvantilů zvoleného rozdělení a empirických kvantilů vstupního vektoru. Neznámé parametry rozdělení jsou opět odhadovány metodou maximální věrohodnosti ze vstupního vektoru a jejich hodnoty se zobrazí v sekci *Parametry*.

Samozřejmostí je možnost uložení veškerých grafických výstupů.

## Program EVDsim

Program EVDsim (viz obrázek B.1(b)) slouží ke grafickému znázornění rychlosti konvergence rozdělení z oboru atraktivity EV rozdělení k danému rozdělení limitnímu (Gumbellovo, Fréchetovo, Weibullovo).

### Spuštění

Po spuštění Matlabu se do příkazového řádku napíše `EVDsim`. Je však nutné se nacházet v adresáři s programem. Dále budou popsány jednotlivé funkce.

### Popis programu

Program je opět rozčleněn na několik částí, přičemž sekce shodné s programem EVDgraph již nebudou uvedeny. V sekci *Nastavení* lze zvolit počet bloků, maximální velikost bloku (počet prvků, ze kterých se počítá maximum), hladinu významnosti pro testování hypotéz a prodlevu mezi jednotlivými iteracemi. Sekce *Průběh simulace* informuje o velikosti bloku v dané iteraci a také o výsledcích testů dobré shody (Pearsonův  $\chi^2$  test a K-S test).

Simulace probíhá buď do doby, než dojde ke konvergenci rozdělení výběrových maxim k limitnímu rozdělení (což v tomto případě znamená, že oba testy dobré shody nezamítají hypotézu, že výběr pochází z daného limitního rozdělení, na dané hladině významnosti), anebo než je dosaženo zvolené maximální velikosti bloku. Na grafech je pak vidět případná shoda mezi histogramem a hustotou limitního rozdělení a mezi empirickou distribuční funkcí náhodného výběru a distribuční funkcí limitního rozdělení.



## C Přehled programů k analýze cenzorovaných výběrů

<code>expfit_oneS.m</code>	výpočet MV odhadu parametru $\lambda$ dvojnásobně zleva cenzorovaného výběru z exponenciálního rozdělení
<code>expfit_twoS.m</code>	výpočet MV odhadů parametrů $\alpha, \lambda$ sdruženého dvojnásobně zleva cenzorovaného výběru z exponenciálního rozdělení
<code>wblfit_oneS.m</code>	výpočet MV odhadů parametrů $\lambda, \tau$ dvojnásobně zleva cenzorovaného výběru z Weibullova rozdělení
<code>wblfit_twoS.m</code>	výpočet MV odhadů parametrů $\alpha, \beta, \lambda, \tau$ sdruženého dvojnásobně zleva cenzorovaného výběru z Weibullova rozdělení
<code>histExpWbl.m</code>	vykreslení histogramu pro dvojnásobně zleva cenzorovaný výběr a proložení histogramu hustotou exponenciálního a Weibullova rozdělení
<code>wblT0exp.m</code>	testování hypotézy o přípustnosti nahrazení dvojnásobně zleva cenzorovaného Weibullova rozdělení exponenciálním rozdělením
<code>compExp.m</code>	testování hypotézy o shodnosti rozdělení dvou nezávislých dvojnásobně zleva cenzorovaných výběrů z exponenciálního rozdělení
<code>compWbl.m</code>	testování hypotézy o shodnosti rozdělení dvou nezávislých dvojnásobně zleva cenzorovaných výběrů z Weibullova rozdělení



## Literatura

- [1] ABOUEISSA, A. E.-M. A. Maximum likelihood estimators of population parameters from multiply censored samples. *Environmetrics*, 2009, vol. **20**, pp. 312–330.
- [2] ABOUEISSA, A. E.-M. A.; STOLINE, M. R. Maximum likelihood estimators of population parameters from doubly left-censored samples. *Environmetrics*, 2006, vol. **17**, pp. 811–826.
- [3] ALEXANDER, G. N.; KAROLY, A.; SUSTS, A. B. Equivalent distributions with applications to rainfall as an upper bound to flood distributions. *Journal of Hydrology*, 1969, vol. **9**, no. 3, pp. 322–344.
- [4] ANDERSON, T. W.; DARLING, D. A. A test of goodness-of-fit. *J. Am. Stat. Assoc.*, 1954, vol. **49**, pp. 765–769.
- [5] ANDĚL, J. *Základy matematické statistiky*. Praha: Matfyzpress, 2005. 358 s. ISBN 80-86732-40-1.
- [6] ASHKAR, F.; BOBÉE, B.; RASMUSSEN, P.; ROSBJERG, D. A perspective on the annual maximum flood approach to flood frequency analysis. In *K. W. Hipel (ed.), Stochastic and Statistical Methods in Hydrology and Environmental Engineering, Extreme Values: Floods and Droughts*, 1994, vol. **1**, Kluwer, pp. 3-14.
- [7] BARNDORFF-NIELSEN, O. E.; COX, D. R. *Inference and Asymptotics*. London: Chapman and Hall/CRC, 1994. 360 p. ISBN 978-0412494406.
- [8] BEIRLANT, J.; GOEGEBEUR, Y.; SEGERS, J.; TEUGELS, J. *Statistics of Extremes: Theory and Applications*. Hoboken, N.J.: Wiley, 2004. 522 p. ISBN 978-0-471-97647-9.
- [9] BEIRLANT, J.; VYNCKIER, P.; TEUGELS, J. L. Tail index estimation, Pareto quantile plots and regression diagnostics. *J. Amer Statist. Association*, 1996, vol. **91**, pp. 1659–1667.
- [10] BEN-ZVI, A. Fit of probability distributions to upper sub-samples of partial duration series. In *K. W. Hipel (ed.), Stochastic and Statistical Methods in Hydrology and Environmental Engineering, Extreme Values: Floods and Droughts*, 1994, vol. **1**, Kluwer, pp. 95–107.
- [11] BEN-ZVI, A. Rainfall intensity-duration-frequency relationships derived from large partial duration series. *Journal of Hydrology*, 2009, vol. **367**, pp. 104–114.

- [12] BESTER, K. Analysis of musk fragrances in environmental samples. *Journal of Chromatography A*, 2009, vol. **1216**, no. 3, pp. 470–480.
- [13] BUIHAND, T. A. The partial duration series method with a fixed number of peaks. *Journal of Hydrology*, 1989, vol. **109**, no. 1–2, pp. 1–9.
- [14] COHEN, A. C. *Truncated and Censored Samples*. New York: Marcel Dekker, 1991. 328 p. ISBN 978-0824784478.
- [15] COLES, S. *An Introduction to Statistical Modeling of Extreme Values*. London: Springer, 2001. 208 p. ISBN 1-85233-459-2.
- [16] COX, D. R.; OAKES, D. *Analysis of Survival Data*. New York: Chapman and Hall/CRC, 1984. 208 p. ISBN 978-0412244902.
- [17] CUNNANE, C. A particular comparison of annual maxima and partial duration series methods of flood frequency predictions. *Journal of Hydrology*, 1973, vol. **18**, no. 3–4, pp. 257–271.
- [18] D'AGOSTINO, R. B.; STEPHENS, M. A. *Goodness-of-fit-techniques*. New York: M. Dekker, 1986. 576 p. ISBN 978-0824774875.
- [19] DAHAN, E.; MENDELSON, H. An extreme value model of concept testing. *Manage. Sci.*, 2001, vol. **47**, pp. 102–116.
- [20] DANIELSSON, J.; DE HAAN, L.; PENG, L.; DE VRIES, C. G. Using a bootstrap method to choose the sample fraction in tail index estimation. *J. Multivariate Analysis*, 2001, vol. **76**, pp. 226–248.
- [21] DAWSON, T. H. Maximum wave crests in heavy seas. *Journal of Offshore Mechanics and Arctic Engineering*, 2000, vol. **122**, no. 3, pp. 222–224.
- [22] DE HAAN, L.; FERREIRA, A. *Extreme Value Theory. An Introduction*. London: Springer, 2006. 421 p. ISBN 0387239464.
- [23] DEEMER, W. L.; VOTAW, D. F. Estimation of parameters of truncated or censored exponential distributions. *Annals of Mathematical Statistics*, 1955, vol. **26**, pp. 498–504.
- [24] DRAISCI, R.; MARCHIAFAVA, C.; FERRETTI, E.; PALLESCHI, L.; CATELLANI, G.; ANASTASIO, A. Evaluation of musk contamination of freshwater fish in Italy by accelerated solvent extraction and gas chromatography with mass spectrometric detection. *Journal of Chromatography A*, 1998, vol. **814**, no. 1–2, pp. 187–197.
- [25] DREES, H.; KAUFMANN, E. Selecting the optimal sample fraction in univariate extreme value estimation. *Stoch. Proc. Appl.*, 1998, vol. **75**, pp. 149–172.



- [26] EL-SHAARAWI, A. H.; DOLAN, D. M. Maximum likelihood estimation of water concentrations from censored data. *Canadian Journal of Fisheries and Aquatic Sciences*, 1989, vol. **46**, pp. 1033–1039.
- [27] EL-SHAARAWI, A. H.; ESTERBY, S. R. Replacement of censored observations by a constant: An evaluation. *Water Research*, 1992, vol. **26**, pp. 835–844.
- [28] FAHRMEIER, L.; TUTZ, G. *Multivariate Statistical Modelling Based on Generalized Linear Models*. 2nd ed. New York: Springer-Verlag, 2001. 574 p. ISBN 978-0387951874.
- [29] FALK, M. Some best estimators for distributions with finite endpoint. *Statistics*, 1995, vol. **27**, pp. 115–125.
- [30] FERNÁNDEZ, A. J. Estimation and hypothesis testing for exponential lifetime models with double censoring and prior information. *Journal of Economic and Social Research*, 2000, vol. **2**, pp. 1–17.
- [31] FISHER, R. A.; TIPPETT, L. H. C. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Cambridge Philos. Soc.*, 1928, vol. **24**, pp. 180–190.
- [32] FUSEK, M. Demonstrační software pro rozdělení extrémních hodnot. *Inf. Bull. České stat. spol.*, 2011, vol. **22**, s. 47–51.
- [33] FUSEK, M.; MICHÁLEK, J. Statistical analysis of doubly left-censored samples from exponential distribution. In *18th International Conference on Soft Computing MENDEL 2012*. Brno, Czech Republic, 2012, pp. 564–569.
- [34] FUSEK, M.; MICHÁLEK, J. Statistical inference of doubly left-censored samples from Weibull distribution. In *XX International Conference PDMU-2012 Problems of Decision Making under Uncertainties Proceedings - Applied Papers*. Brno, Czech Republic, 2012, pp. 31–40.
- [35] FUSEK, M.; MICHÁLEK, J. Statistical analysis of type I multiply left-censored samples from exponential distribution. 2013, odesláno k publikaci.
- [36] FUSEK, M.; MICHÁLEK, J. Statistical methods for analyzing musk compounds concentration based on doubly left-censored samples. *International Journal of Mathematical Models and Methods in Applied Sciences*, 2013, přijato k publikaci.
- [37] FUSEK, M.; MICHÁLEK, J.; ZOUHAR, L.; VÁVROVÁ, M. Statistical analysis of musk compounds concentrations in fish tissue based on doubly left-censored samples. In *Recent Advances in Energy, Environment and Economic Development - Proceedings of the 3rd International Conference on Development, Energy, Environment, Economics*. Paris, France, 2012, pp. 247–250.

- [38] GNEDENKO, B. V. Sur la distribution limite du terme maximum d' une série aléatoire. *Ann. Math.*, 1943, vol. **44**, pp. 423–453.
- [39] GOMES, M. I.; OLIVEIRA, O. The bootstrap methodology in statistics of extremes - choice of the optimal sample fraction. In *Extremes*, 2001, vol. **4**, no. 4, pp. 331–358.
- [40] GREENWOOD, J. A.; LANDWEHR, J. M.; MATALAS, N. C.; WALLIS, J. R. Probability-weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Water Resour. Res.*, 1979, vol. **15**, pp. 1049–1054.
- [41] GUMBEL, E. J. Les intervalles extremes entre les émissions radio actives. *J. Phys. Radium*, 1937, vol. **8**, pp. 446–452.
- [42] GUMBEL, E. J. The return period of flood flows. *Ann. Math. Statist.*, 1941, vol. **12**, pp. 163–190.
- [43] GUMBEL, E. J. On the plotting of flood discharges. *Trans. Amer. Geophys. Union*, 1944, vol. **25**, pp. 699–719.
- [44] GUMBEL, E. J. Floods estimated by probability methods. *Engrg. News-Record*, 1945, vol. **134**, pp. 97–101.
- [45] GUMBEL, E. J. The statistical forecast of floods. *Ohio Water Resources Board*, 1949, Bulletin No. **15**, pp. 1–21.
- [46] GUPTA, R. D.; KUNDU, D. Generalized exponential distribution. *Austral. & New Zealand J. Statist.*, 1999, vol. **41**, pp. 173–188.
- [47] HARREMOËS, P.; MIKKELSEN, P. S. Properties of extreme point rainfall I: Results from a rain gauge system in Denmark. *Atmospheric Research*, 1995, vol. **37**, no. 4, pp. 277–286.
- [48] HARRIS, R. I. The accuracy of design values predicted from extreme value analysis. *Journal of Wind Engineering and Industrial Aerodynamics*, 2001, vol. **89**, no. 2, pp. 153–164.
- [49] HELLEBRAND, R.; MICHÁLEK, J.; FUSEK, M.; PRAX, P. Aktualizace hydrologických podkladů pro potřeby městské hydrologie. In *Hydrologické podklady pro potřeby městského odvodnění*. Brno: FAST VUT Brno, 2010. s. 10–23. ISBN: 978-80-214-4096-8.
- [50] HELLEBRAND, R.; MICHÁLEK, J.; FUSEK, M.; PRAX, P. Modelling precipitation extremes for the purpose of urban drainage in the Czech Republic. 2012, odesláno k publikaci.

- [51] HILL, B. M. A simple general approach to inference about the tail of a distribution. *Ann. Statist.*, 1975, vol. **3**, pp. 1163–1174.
- [52] HOGG, R. V.; MCKEAN, J. W.; CRAIG, A. *Introduction to Mathematical Statistics*. 6th ed. Upper Saddle River, N.J.: Pearson Education, 2005. 692 p. ISBN 978-0130085078.
- [53] HOSKING, J. R. M.; WALLIS, J. R. Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, 1987, vol. **29**, pp. 339–349.
- [54] HOSKING, J. R. M.; WALLIS, J. R.; WOOD, E. F. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, 1985, vol. **27**, pp. 251–261.
- [55] CHOULAKIAN, V.; STEPHENS, M. A. Goodness-of-fit tests for the generalized Pareto distribution. *Technometrics*, 2001, vol. **43**, no. 4, pp. 478–484.
- [56] CHOW, V. T.; MAIDMENT, D. R.; MAYS, L. W. *Applied Hydrology*. New York: McGraw-Hill, 1988. 572 p. ISBN 978-0070108103.
- [57] IYER, S. K.; JAMMALAMADAKA, S. R.; KUNDU, D. Analysis of middle censored data with exponential lifetime distribution. *J. Statist. Plann. Inference*, 2008, vol. **138**, pp. 3550–3560.
- [58] KAWAS, M. L.; MOREIRA, R. G. Characterization of product quality attributes of tortilla chips during the frying process. *Journal of Food Engineering*, 2001, vol. **47**, no. 2, pp. 97–107.
- [59] KOTZ, S.; NADARAJAH, S. *Extreme Value Distributions. Theory and Applications*. London: Imperial College Press, 2000. 187 p. ISBN 1860942245.
- [60] LAGARIAS, J. C.; REEDS, J. A.; WRIGHT, M. H.; WRIGHT, P. E. Convergence properties of the Nelder–Mead simplex method in low dimensions. *SIAM J. Optim.*, 1998, vol. **9**, pp. 112–147.
- [61] LANDWEHR, J.; MATALAS, N.; WALLIS, J. Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. *Water Resour. Res.*, 1979, vol. **15**, pp. 1055–1064.
- [62] LANGOUSIS, A.; VENEZIANO, D. Intensity-duration-frequency curves from scaling representations of rainfall. *Water Resources Research*, 2007, vol. **43**, no. 2, DOI: 10.1029/2006WR005245.
- [63] LAVENDA, B.; CIPOLLONE, E. Extreme value statistics and thermodynamics of earthquakes: aftershock sequences. *Annals of Geophysics*, 2000, vol. **43**, no. 5, pp. 967–982.

- [64] LEHMANN, E. L. *Elements of Large-Sample Theory*. New York: Springer, 1999. 631 p. ISBN 0-387-98595-6.
- [65] LEHMANN, E. L.; CASELLA, G. *Theory of Point Estimation*. 2nd ed. New York: Springer-Verlag, 1998. 616 p. ISBN 978-0387985022.
- [66] LEHMANN, E. L.; ROMANO, J. P. *Testing Statistical Hypotheses*. 3rd ed. New York: Springer, 2005. 800 p. ISBN 978-0387988641.
- [67] LIGNELL, S.; DARNERUD, P. O.; AUNE, M.; CNATTINGIUS, S.; HAJŠLOVA, J.; SETKOVA, L.; GLYNN, A. Temporal trends of synthetic musk compounds in mother's milk and associations with personal use of perfumed products. *Environ. Sci. Technol.*, 2008, vol. **42**, no. 17, pp. 6743–6748.
- [68] LIKEŠ, J.; LAGA, J. *Základní statistické tabulky*. Praha: SNTL, 1978. 488 s.
- [69] LILLIEFORS, H. On the Kolmogorov–Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, 1967, vol. **62**, pp. 399–402.
- [70] MADSEN, H.; MIKKELSEN, P. S.; ROSBJERG, D.; HARREMOES, P. Regional estimation of rainfall intensity-duration-frequency curves using generalized least squares regressions of partial duration series. *Water Resources Research*, 2002, vol. **38**, no. 11, pp. 21-1–21-11.
- [71] MADSEN, H.; PEARSON, C. P.; ROSBJERG, D. Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events, 2. Regional modeling. *Water Resources Research*, 1997, vol. **33**, no. 4, pp. 759–769.
- [72] MADSEN, H.; RASMUSSEN, P. F.; ROSBJERG, D. Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events, 1. At site modeling, *Water Resources Research*, 1997, vol. **33**, no. 4, pp. 747–757.
- [73] MCCUEN, R. M.; JOHNSON, P. A.; HROMADKA, T. V. Regionalized partial-duration balanced-hydrograph model. *Journal Irrigation and Drainage Engineering ASCE*, 1993, vol. **119**, pp. 1036–1051.
- [74] MCNULTY, P. J.; SCHEICK, L. Z.; ROTH, D. R.; DAVIS, M. G.; TORTORA, M. R. S. First failure predictions for EPROMs of the type flown on the MPTB satellite. *IEEE Transactions on Nuclear Science*, 2000, vol. **47**, no. 6, pp. 2237–2243.
- [75] MICHÁLEK, J.; FUSEK, M.; HELLEBRAND, R. Srovnání statistických metod odhadů IDF křivek. In *Workshop Adopfa Patery 2012 Extrémní hydrologické jevy v povodích*. Brno, 2012, s. 163–170. ISBN: 978-80-02-02423-1.

- [76] MITRA, S.; KUNDU, D. Analysis of left censored data from the generalized exponential distribution. *J. Stat. Comput. Simul.*, 2008, vol. **78**, pp. 669–679.
- [77] NORDQUIST, J. M. Theory of largest values applied to earthquake magnitudes. *Trans. Amer. Geophys. Union*, 1945, vol. **26**, pp. 29–31.
- [78] PICKANDS III, J. Statistical inference using extreme order statistics. *Ann. Statist.*, 1975, vol. **3**, pp. 119–131.
- [79] POTTER, W. D. Normalcy tests of precipitation and frequency studies of runoff on small watersheds. *Technical bulletin/United States Department of Agriculture*, 1949, vol. **985**, 24 p.
- [80] RANTZ, S. F.; RIGGS, H. C. Magnitude and frequency of floods in the Columbia river basin. *U.S. Geological Survey, Water Supply Paper*, 1949, vol. **1080**, pp. 317–476.
- [81] ROBERTS, S. J. Extreme value statistics for novelty detection in biomedical data processing. *IEE Proceedings—Science, Measurement and Technology*, 2000, vol. **147**, no. 6, pp. 363–367.
- [82] ROSBJERG, D.; MADSEN, H.; RASMUSSEN, P. F. Prediction in partial duration series with generalized Pareto distribution exceedances. *Water Resources Research*, 1992, vol. **28**, pp. 3001–3010.
- [83] SMITH, J. A. Estimating the upper tail of flood frequency distributions. *Water Resources Research*, 1987, vol. **23**, pp. 1657–1666.
- [84] SMITH, R. L. Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 1985, vol. **72**, no. 1, pp. 67–90.
- [85] SMITH, R. L.; WEISSMAN, I. Maximum likelihood estimation of the lower tail of a probability distribution. *J. Royal Statist. Soc. Ser. B*, 1985, vol. **47**, pp. 285–298.
- [86] STEPHENS, M. A. Asymptotic results for goodness-of-fit statistics with unknown parameters. *Ann. Statist.*, 1976, vol. **4**, pp. 357–369.
- [87] SUNDBERG, R. Comparison of confidence procedures for type I censored exponential lifetimes. *Lifetime Data Anal.*, 2001, vol. **7**, pp. 393–413.
- [88] TAKEUCHI, K. Annual maximum series and partial duration series - evaluation of Langbein's formula and Chow's discussion. *Journal of Hydrology*, 1984, vol. **68**, no. 1–4, Global Water: Science and Engineering The Ven Te Chow Memorial Volume, pp. 275–284.

- [89] THOMPSON, M. L.; REYNOLDS, J.; COX, L. H.; GUTTORP, P.; SAMPSON, P. D. A review of statistical methods for the meteorological adjustment of tropospheric ozone. *Atmospheric Environment*, 2001, vol. **35**, no. 3, pp. 617–630.
- [90] TODOROVIC, P. Stochastic models of floods. *Water Resources Research*, 1978, vol. **14**, pp. 345–356.
- [91] TRUPL, J. *Intenzity krátkodobých dešťů v povodích Labe, Odry a Moravy*. VÚV Praha, Práce a studie 97, 1958. 34 s.
- [92] TRYON, R. G.; CRUSE, T. A. Probabilistic mesomechanics for high cycle fatigue life prediction. *Journal of Engineering Materials and Technology—Transactions of the AMSE*, 2000, vol. **122**, pp. 209–214.
- [93] VAN MONTFORT, M. A. J.; WITTER, J. V. The generalized Pareto distribution applied to rainfall depths. *Hydrological Science Journal*, 1986, vol. **31**, pp. 151–162.
- [94] WEIBULL, W. A statistical theory of the strength of materials. *Ing. Vet. Akad. Handlingar*, 1939, vol. **151**, pp. 5–45.
- [95] WILLEMS, P. Compound intensity/duration/frequency-relationships of extreme precipitation for two seasons and two storm types. *Journal of Hydrology*, 2000, vol. **233**, no. 1–4, pp. 189–205.
- [96] YEJVJEVICH, V. *Probability and Statistics in Hydrology*. Fort Collins, CO: Water Resources Publications, 1972. 302 p. ISBN 978-0918334008.

# Seznam symbolů a zkratek

## Symbols

$F, G$	distribuční funkce
$f, g$	hustota
$\Phi(x)$	distribuční funkce normovaného normálního rozdělení $N(0; 1)$
$X \sim N(0; 1)$	$X$ je náhodná veličina s rozdělením $N(0; 1)$
$x^*$	pravý koncový bod rozdělení
$h^{\leftarrow}$	zleva spojitá inverzní funkce k funkci $h$
$F^{\leftarrow}$	kvantilová funkce
$U(t)$	kvantilová funkce chvostu
$[t]$	celá část $t$
$\gamma$	EV index
$\mathcal{D}(G)$	obor atraktivity distribuční funkce $G$
$\approx$	aproximace
$\mathbf{V}$	varianční (variančně-kovarianční) matice
$\mathbf{J}_n$	očekávaná Fisherova informační matice
$\tilde{\mathbf{J}}_n$	výběrová Fisherova informační matice
$x_{(j)}$	$j$ -tý prvek v uspořádaném výběru $x_1, \dots, x_n$
$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$	gama funkce
$\xrightarrow{d}$	konvergence v distribuci
$\gamma_e \doteq 0,57722$	Eulerova konstanta
$\vee$	nebo
$\wedge$	a zároveň
$\mathbb{R}_+$	množina kladných reálných čísel včetně nuly
$\mathbb{R}_+^*$	množina kladných reálných čísel
$\Psi, H, \Lambda$	spektrální míra

## Zkratky

A-D test	Andersonův-Darlingův test
AMS	Annual maximum series
ČOV	čistírna odpadních vod
EV rozdělení	rozdělení extrémních hodnot
EV index	index extrémní hodnoty
FIM	Fisherova informační matice
FMI	Fisherova míra informace
GEV rozdělení	zobecněné rozdělení extrémních hodnot
K-S test	Kolmogorovův-Smirnovův test
LOD	mez detekce
LOQ	mez stanovitelnosti
MV odhad	odhad metodou maximální věrohodnosti
PDS	Partial duration series
PVM odhad	odhad metodou pravděpodobnostně vážených momentů