



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA STROJNÍHO INŽENÝRSTVÍ

FACULTY OF MECHANICAL ENGINEERING

ÚSTAV MATEMATIKY

INSTITUTE OF MATHEMATICS

METODY ODHADU PARAMETRŮ ROZDĚLENÍ EXTRÉMNÍHO
TYPU S APLIKACEMI

EXTREME VALUE DISTRIBUTION PARAMETER ESTIMATION AND ITS APPLICATION

DIZERTAČNÍ PRÁCE

DOCTORAL THESIS

AUTOR PRÁCE

AUTHOR

Ing. JAN HOLEŠOVSKÝ

VEDOUCÍ PRÁCE

SUPERVISOR

doc. RNDr. JAROSLAV MICHÁLEK, CSc.

Abstrakt

Předložená práce je zaměřena na teorii extrémních hodnot a její užití v aplikačních úlohách. V první části je zavedeno rozdělení extrémních hodnot a popsány jeho vlastnosti. Na základě předložených tvrzení jsou diskutovány dva přístupy k analýze extrémních hodnot, a sice model blokových maxim a prahový model postavený na zobecněném Paretově rozdělení. Ačkoliv je první jmenovaný v mnoha ohledech chápán jako robustnější, patří prahový model ke stále častěji užívaným přístupům. Samotná volba prahu, která má zásadní vliv na kvalitu odhadu, však pořád patří k nedořešeným problémům tohoto přístupu. Především na techniky určení vhodné prahové hodnoty je tato práce zaměřena. Z aplikačního hlediska jsou pak nejzajímavější adaptivní přístupy určení prahu, které danou volbu vhodně automatizují. Pro porovnání vybraných adaptivních technik byla provedena simulační studie a tyto byly dále použity pro analýzu srážkových úhrnů v jiho-moravském regionu. Dále se práce věnuje v poslední době rozvíjeným metodám odhadu extrémních hodnot stacionárních řad. V praxi je často nutné z měřené časové řady vzorkovat přibližně nezávislá pozorování. Použití teorie pro stacionární řady přitom tento problém redukce dat zcela eliminuje. Jak je ukázáno, běžně používané metody vzorkování se v tomto kontextu ukazují jako nevhodné a užití pokročilých technik pro stacionární řady vede k lepším odhadům extrémních hodnot.

Summary

The thesis is focused on extreme value theory and its applications. Initially, extreme value distribution is introduced and its properties are discussed. At this basis are described two models mostly used for an extreme value analysis, i.e. the block maxima model and the Pareto-distribution threshold model. The first one takes advantage in its robustness, however recently the threshold model is mostly preferred. Although the threshold choice strongly affects estimation quality of the model, an optimal threshold selection still belongs to unsolved issues of this approach. Therefore, the thesis is focused on techniques for proper threshold identification, mainly on adaptive methods suitable for the use in practice. For this purpose a simulation study was performed and acquired knowledge was applied for analysis of precipitation records from South-Moravian region. Further on, the thesis also deals with extreme value estimation within a stationary series framework. Usually, an observed time series needs to be separated to obtain approximately independent observations. The use of the advanced theory for stationary series allows to avoid the entire separation procedure. In this context the commonly applied separation techniques turn out to be quite inappropriate in most cases and the estimates based on theory of stationary series are obtained with better precision.

Klíčová slova

rozdělení extrémních hodnot, index extrémní hodnoty, zobecněné Paretovo rozdělení, extrémální index, neparametrické odhady

Keywords

extreme value distribution, extreme value index, generalized Pareto distribution, extremal index, nonparametric estimation

HOLEŠOVSKÝ, J. *Metody odhadu parametrů rozdělení extrémního typu s aplikacemi*. Brno: Vysoké učení technické v Brně, Fakulta strojního inženýrství, 2016. 144 s. Vedoucí dizertační práce doc. RNDr. Jaroslav Michálek, CSc.

Prohlašuji, že jsem dizertační práci *Metody odhadu parametrů rozdělení extrémního typu s aplikacemi* vypracoval samostatně pod vedením doc. RNDr. Jaroslava Michálka, CSc., s použitím materiálů uvedených v seznamu literatury.

V Brně dne 25. dubna, 2016

Ing. Jan Holešovský

Rád bych poděkoval doc. RNDr. Jaroslavu Michálkovi, CSc., za jeho vedení mé dizertační práce, veškerou ochotu a trpělivost, kterou mi věnoval, a v neposlední řadě za množství cenných rad nejen v odborných oblastech či k samotné dizertační práci. Mé poděkování patří také mým nejbližším za jejich podporu během celého studia.

| | |
|--|------------|
| Úvod | 13 |
| 1 Základy teorie extrémních hodnot | 15 |
| 1.1 Problematika extrémních hodnot | 15 |
| 1.2 Rozdělení extrémních hodnot | 17 |
| 1.3 Základní typy EV rozdělení | 24 |
| 1.4 Podmínka druhého řádu | 31 |
| 2 Odhady parametrů rozdělení extrémních hodnot | 37 |
| 2.1 Základní modely pro odhady parametrů a parametrických funkcí EV rozdělení | 37 |
| 2.2 Odhady metodou maximální věrohodnosti | 40 |
| 2.3 Semiparametrické přístupy pro odhady parametrů GEV rozdělení | 47 |
| 2.4 Metoda bootstrap | 53 |
| 3 Metody výběru prahových hodnot | 57 |
| 3.1 Grafické metody | 57 |
| 3.2 Adaptivní volba prahu v semiparametrickém modelu | 59 |
| 3.3 Adaptivní volba prahu založená na věrohodnostní funkci | 63 |
| 4 Statistická analýza extrémních dešťových srážek | 67 |
| 4.1 Srážková data a techniky vzorkování | 67 |
| 4.2 Simulační studie pro volbu prahu adaptivními technikami | 71 |
| 4.3 Analýza srážek pomocí jednorozměrné teorie nezávislých veličin | 77 |
| 5 Teorie extrémních hodnot pro stacionární procesy | 89 |
| 5.1 Rozdělení extrémních hodnot | 89 |
| 5.2 Metody odhadu extrémálního indexu | 93 |
| 6 Statistická analýza pomocí teorie stacionárních řad | 99 |
| 6.1 Simulační studie pro porovnání vlastností odhadů extrémálního indexu | 99 |
| 6.2 Analýza srážkových časových řad | 101 |
| 6.3 Validace měření ve formě časových řad | 106 |
| Závěr | 111 |
| A Odvození vybraných pomocných funkcí podmínek atraktivity a druhého řádu | 113 |
| B Odvození skórového vektoru a Fisherovy informační matice MT-GP modelu adaptivní volby prahové hodnoty | 125 |
| C Software EVDest | 129 |
| D Popis programů k analýze extrémních hodnot | 131 |

OBSAH

Literatura **133**

Seznam symbolů a zkratek **143**

Teorie extrémních hodnot (zkráceně EV teorie z angl. Extreme Value) je oblast matematické statistiky zaměřená na rozvoj metod určených pro popis a modelování řídicí se vyskytujících, extrémních jevů. Potřeba užití adekvátních technik predikce málo častých událostí se objevuje v mnoha oblastech lidské činnosti. Zejména se jedná o nutnost odhadu vlastností chvostů daných rozdělení pravděpodobnosti, neboť tyto mají zásadní vliv na selhání či poruchy sledovaného procesu. Obvykle je pro takový proces vyžadována předpověď dosažení nějaké vysoké, předem stanovené úrovně, přičemž tato se často nachází mimo oblast pozorovaných dat. Ačkoliv jsou tyto predikce vždy postaveny na důvěře ve správnost konkrétního modelu, EV teorie dnes představuje pravděpodobně nejadekvátnější prostředek pro popis vlastností extrémů náhodných veličin.

Historicky se jako první úloha EV teorie uvádí problém nalezení pravého koncového bodu $a > 0$ spojitého rovnoměrného rozdělení na intervalu $\langle 0, a \rangle$ (Bernoulli, 1709). V moderním pojetí se problém související s extrémí náhodných veličin objevuje v pracích Fullera (1914) či Griffitha (1920), první systematický popis lze nalézt v publikaci Bortkiewiczze (1922), který se zabýval extrémními hodnotami výběrů pocházejících z normálního rozdělení. Pro rozvoj asymptotické teorie je pak klíčový článek Fishera a Tippetta [52], kteří odvodili jediné tři možné typy asymptotických rozdělení největších pozorování a zobecnili tak předešlé práce Frécheta (1927). Gnedenko [59] dále tyto poznatky rozšířil a uvedl nutné a postačující podmínky pro slabou konvergenci horních pořadových statistik. De Haan (1970, 1971) pak celou EV teorii rozvinul pomocí teorie regulárně se měnících funkcí. Detailní citace na výše uvedené práce uvádí publikace [86], kde lze nalézt i některé doplňující literární zdroje.

V současnosti lze největší část rozvoje EV teorie přisoudit praktickým potřebám reálných úloh zejména v oblasti hydrologie a ekonomie. V poslední době se přitom v literatuře stále více objevují aplikace z nejrůznějších vědních oblastí i mimo tyto tradiční obory. Z moderních řešených problematik lze zmínit například předpověď srážkových extrémů [56, 69], modelování výšky přílivu [50] či rychlosti větru [46, 50], odhady frekvence výskytu velmi nízkých teplot [124, 92], analýzu rozsáhlých požárů [81] či znečištění životního prostředí [110], posouzení návratnosti pojištění hmotného majetku [32] a dalších finančních dopadů [85], redukci šumu obrazových dat [113], nebo posouzení kapacity telekomunikačních sítí [127].

Přes celou řadu aktuálních článků a monografií však patří EV teorie ke stále se rozvíjejícím oborům matematické statistiky s mnoha dosud nedorozřešenými tématy. Vybranými oblastmi, které zůstávají nadále otevřeny dalšímu výzkumu, se zabývá tato práce. Členění kapitol předkládané dizertační práce je následující. První kapitola obsahuje stručnou charakteristiku problematiky teorie extrémních hodnot, zpracovanou převážně podle monografií [38, 33, 18], přičemž je formulováno a také dokázáno základní tvrzení vedoucí k zavedení rozdělení extrémních hodnot (EV rozdělení). Dále jsou v krátkosti shrnuty vybrané vlastnosti tohoto rozdělení, mimo jiné je také stanovena tzv. podmínka druhého řádu. Splnění této podmínky je zásadním požadavkem pro existenci asymptotických rozdělení dále studovaných odhadů parametrů EV rozdělení.

V druhé kapitole je podán popis dvou běžně užívaných přístupů k analýze extrémních hodnot, a sice modelu blokových maxim a prahového modelu. Následně jsou v krátkosti

představeny metody odhadu, které budou dále použity k řešení praktických úloh. Popis metody maximální věrohodnosti byl zpracován podle knih [14, 31, 94], přičemž jsou diskutovány konkrétní vlastnosti odhadů v blokovém i prahovém modelu. Stávající poznatky z oblasti semiparametrických odhadů byly přebrány z monografie [38], kde lze také nalézt důkazy předkládaných tvrzení. Stručně jsou podle [48] také popsány základy metody bootstrap, jíž bude v kontextu EV teorie použito pro vybrané adaptivní techniky určení prahové hodnoty.

Třetí kapitola je soustředěna na metody výběru prahových hodnot v prahovém modelu, které stále patří k nejasným součástem EV teorie. Jsou zmíněny jak klasické explorativní postupy, tak dvě v poslední době rozvinuté techniky adaptivní volby prahu.

Ve čtvrté kapitole jsou uvedené metody aplikovány k odhadům frekvencí výskytů extrémních srážkových intenzit. Jedná se o autorovu původní reálnou aplikační úlohu (viz [1]) založenou na potřebě aktualizace hydrologických podkladů pro vyhodnocování extrémních srážkových událostí. Zde je navázáno na dřívější studie v této oblasti, přičemž cílem této dizertační práce je především srovnání parametrických a semiparametrických technik k odhadům extrémních srážek se zvláštním zřetelem na adaptivní techniky volby prahové hodnoty. Pro porovnání těchto adaptivních přístupů byla provedena simulační studie a získané poznatky byly uplatněny i při řešení zmíněné aplikační úlohy.

V páté kapitole je popsáno rozšíření EV teorie pro stacionární procesy. Dochází tak k zavedení tzv. parametru extrémálního indexu, který vyjadřuje míru korelovanosti pozorování na extrémních úrovních. Při analýze pomocí jednorozměrné teorie pro nezávislá pozorování je pro časové řady nutné použít dodatečné vzorkovací techniky, které jsou založeny na expertních znalostech v dané aplikační oblasti. Aplikací metod pro extrémní hodnoty stacionárních procesů je tato potřeba zcela eliminována, čímž je možné docílit zpřesnění odhadů parametrů a parametrických funkcí.

Šestá kapitola nejprve popisuje autorem provedenou simulační studii ke srovnání vybraných současných přístupů k odhadům extrémálního indexu. Dříve provedená analýza srážkových dat je v tomto smyslu aktualizována, přičemž jsou učiněna některá doporučení k použitým technikám. Dále je představena původní autorova aplikace EV teorie stacionární procesů pro identifikaci odlehlých pozorování v časových řadách.

V dodatku A jsou pro vybraná rozdělení pravděpodobnosti autorem odvozené pomocné funkce podmínky prvního a druhého řádu. V dodatku B je pak provedeno detailní odvození testovacích statistik adaptivní metody volby prahu, která je popsána v třetí kapitole. Všechny výše diskutované metody byly také autorem práce počítačově implementovány v prostředí Matlab. V dodatku C je představen autorem vytvořený software pro analýzu extrémních hodnot založenou na teorii pro nezávislá pozorování i časové řady. Některé další implementace jsou shrnuty v dodatku D, kde je možné nalézt přehled jednotlivých funkcí.

Vzhledem k řešeným aplikačním problémům budou v celé práci uvažovány jen případy absolutně spojitých rozdělení. Tato skutečnost bude striktně dodržována, přičemž předpoklad absolutní spojitosti nebude dále připomínán.

Základy teorie extrémních hodnot

V následujících odstavcích budou shrnuty základní poznatky z teorie extrémních hodnot v rozsahu nutném pro další kapitoly této práce. Nejdříve je pozornost věnována popisu elementárního řešeného problému, načež je vzápětí formulováno a dokázáno základní tvrzení teorie extrémních hodnot. V posledních dvou odstavcích jsou předloženy některé doplňující poznatky a tvrzení potřebná pro studium metod odhadů. Jsou zde popsány třídy možných limitních rozdělání extrémních hodnot včetně jejich vlastností a formulována tzv. podmínka druhého řádu, jejíž splnění je nutným požadavkem pro existenci asymptotických rozdělání semiparametrických odhadů parametrů rozdělání extrémních hodnot. Tato kapitola je zpracována podle publikací [38, 18, 49], základní přehled teorie extrémních hodnot je možné nalézt také v [33].

1.1. Problematika extrémních hodnot

Teorie extrémních hodnot (zkráceně EV teorie z angl. Extreme Value) je zaměřena na asymptotické chování extrémních hodnot, tedy hodnot, které jsou pozorovány s malou pravděpodobností. Uvažujme pravděpodobnostní prostor (Ω, \mathcal{A}, P) a na něm náhodný výběr X_1, \dots, X_n rozsahu $n \in \mathbb{N}$ z rozdělání s distribuční funkcí $F(x)$, tj. X_1, \dots, X_n jsou nezávislé náhodné veličiny se stejnou distribuční funkcí $F(x)$. Cílem EV teorie je tak vyšetřit pravděpodobnostní chování výběrového maxima $M_n = \max\{X_1, \dots, X_n\}$, resp. výběrového minima $m_n = \min\{X_1, \dots, X_n\}$, pro případ kdy $n \rightarrow \infty$. Zřejmě se lze omezit jen na popis výběrového maxima M_n , které se z aplikačních hledisek jeví jako zajímavější, neboť případ výběrového minima lze převést na předchozí pomocí vztahu

$$\max\{X_1, \dots, X_n\} = -\min\{-X_1, \dots, -X_n\}.$$

EV teorie je tak jakousi obdobou asymptotické teorie výběrových průměrů, která je popsána pomocí centrální limitní věty. Asymptotické vlastnosti náhodných veličin jsou v praktických situacích velmi často využívány. Běžně se lze setkat s případy, kdy lze rozdělání výběrového průměru $\bar{X}_n = \sum_{i=1}^n X_i/n$ aproximovat normálním rozděláním. Toto nahrazení bývá aplikováno buď z důvodu, že pro n konečné není možné rozdělání \bar{X}_n exaktně odvodit, nebo, ačkoli distribuční funkce $F(x)$ není známá, je k dispozici výběr dostatečného rozsahu.

Má-li rozdělání dané distribuční funkcí $F(x)$ střední hodnotu EX , pak ze zákona velkých čísel [14] vyplývá, že pro $n \rightarrow \infty$ platí

$$\bar{X}_n \xrightarrow{P} EX,$$

kde \xrightarrow{P} označuje konvergenci podle pravděpodobnosti, tj. pro každé $\varepsilon > 0$ platí

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - EX| > \varepsilon) = 0.$$

Existuje-li navíc pro dané rozdělení konečný rozptyl $\text{var}(X)$, potom podle centrální limitní věty [14] platí, že asymptotickým rozdělením normovaného výběrového průměru $\sqrt{n}(\bar{X}_n - EX)/\sqrt{\text{var}(X)}$ je standardizované normální rozdělení $N(0,1)$. Tedy

$$\frac{\bar{X}_n - EX}{\sqrt{\text{var}(X)}} \sqrt{n} \xrightarrow{d} U \sim N(0,1),$$

kde $U \sim N(0,1)$ označuje, že náhodná veličina U má rozdělení $N(0,1)$, a \xrightarrow{d} značí konvergenci v distribuci, tj. pro všechna $u \in \mathbb{R}$ je splněno

$$\lim_{n \rightarrow \infty} P\left(\frac{\bar{X}_n - EX}{\sqrt{\text{var}(X)}} \sqrt{n} \leq u\right) = \Phi(u),$$

kde $\Phi(u)$ je distribuční funkce standardizovaného normálního rozdělení.

Buď $F_{M_n}(x) := P(M_n \leq x)$ distribuční funkce výběrového maxima M_n . S využitím předpokladu nezávislosti náhodných veličin X_1, \dots, X_n je možné odvodit funkci $F_{M_n}(x)$ v následujícím tvaru

$$F_{M_n}(x) = P(\max\{X_1, \dots, X_n\} \leq x) = P\left(\bigcap_{i=1}^n [X_i \leq x]\right) = F^n(x), \quad (1.1)$$

kde $F(x)$ je distribuční funkce veličin $X_i, i = 1, \dots, n$. V dalším textu označme $x^* := \sup\{x : F(x) < 1\}$ pravý koncový bod rozdělení s distribuční funkcí $F(x)$. Tedy x^* je největší možná hodnota, které může náhodná veličina X nabývat. Může však nastat i situace, kdy $x^* = \infty^1$. Obdobou zákona velkých čísel pro výběrový průměr je následující tvrzení.

VĚTA 1.1. Nechtě X_1, \dots, X_n je náhodný výběr z rozdělení s distribuční funkcí $F(x)$ a pravým koncovým bodem x^* . Je-li $M_n = \max\{X_1, \dots, X_n\}$, pak pro $n \rightarrow \infty$ platí

$$M_n \xrightarrow{P} x^*.$$

Důkaz. Pro ověření předchozího tvrzení musíme rozlišit dva případy:

(i) Pokud $x^* < \infty$, potom pro libovolné $\varepsilon > 0$ platí

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

neboť $F(x^* - \varepsilon) < 1$.

¹Přirozeně v případě $x^* = \infty$ nemůže náhodná veličina X nabývat hodnoty x^* . V tomto případě máme na mysli, že realizace veličiny X může být libovolné reálné číslo.

(ii) Pokud $x^* = \infty$, potom pro všechna $x \in \mathbb{R}$ platí

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

□

Nyní se zaměříme na limitní rozdělení výběrového maxima. Ze vztahu (1.1) je patrné, že pro $n \rightarrow \infty$

$$F_{M_n}(x) = F^n(x) \rightarrow G_{\text{deg}}(x),$$

kde $G_{\text{deg}}(x)$ je degenerovaná funkce v x^* , tj. $G_{\text{deg}}(x) = 0$ pro $x < x^*$ a $G_{\text{deg}}(x) = 1$ pro $x \geq x^*$. Pro $x^* < \infty$ je pak $G_{\text{deg}}(x)$ distribuční funkcí. V případě, že pravý koncový bod není konečný, je limitní funkce $G_{\text{deg}}(x)$ identicky rovna nule. Poslední jmenovaná situace nastává v celé řadě běžně používaných rozdělení (mimo jiné v případě normálního rozdělení) a přitom o vlastnostech extrémních hodnot nepodává příliš dobrou představu. Z tohoto důvodu je stejně jako v případě centrální limitní věty potřeba výběrová maxima řádně normalizovat. Předpokládejme, že existují posloupnosti reálných konstant $a_n > 0$, b_n ($n \in \mathbb{N}$) tak, že veličina $(M_n - b_n)/a_n$ má nedegenerované limitní rozdělení, tj.

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G(x), \quad (1.2)$$

pro každý bod spojitosti funkce $G(x)$, kde $G(x)$ je nějaká nedegenerovaná distribuční funkce. Rozdělení s distribuční funkcí $G(x)$ z podmínky (1.2) se nazývá *rozdělením extrémních hodnot* (zkráceně EV rozdělení). Podmínka (1.2) se nazývá *podmínkou atraktivity* a třída všech distribučních funkcí $F(x)$ (resp. příslušných rozdělení) vyhovující podmínce (1.2) se pak nazývá *obor atraktivity distribuční funkce $G(x)$* .

Na možné tvary limitních rozdělení bude zaměřena následující část. Budou také stanoveny nutné a postačující podmínky pro to, aby dané rozdělení patřilo do oboru atraktivity nějakého EV rozdělení.

1.2. Rozdělení extrémních hodnot

Podobně jako hraje při dokazování centrální limitní věty roli charakteristická funkce [91], má v EV teorii obdobný význam tzv. kvantilová funkce chvostu. Nejprve proto tento pojem zavedeme a představíme alternativní formulaci podmínky atraktivity (1.2). Následně bude stanovena třída možných limitních rozdělení v (1.2), tedy třída EV rozdělení.

Uvažujme nějakou neklesající funkci f a zaveďme k ní funkci f^{\leftarrow} danou následovně

$$f^{\leftarrow}(x) = \inf\{y \in \mathbb{R} : f(y) \geq x\}. \quad (1.3)$$

Protože pro každé $x_0 \in \mathbb{R}$ platí

$$\lim_{x \uparrow x_0} f^{\leftarrow}(x) = \lim_{x \uparrow x_0} \inf\{y \in \mathbb{R} : f(y) \geq x\} = \inf\{y \in \mathbb{R} : f(y) \geq x_0\} = f^{\leftarrow}(x_0),$$

je vidět, že funkce $f^{\leftarrow}(x)$ je zleva spojitá. Lze také ukázat, že $(f^{\leftarrow})^{\leftarrow}(x) \leq f(x)$ [112]. V případě, že f je rostoucí, pak $(f^{\leftarrow})^{\leftarrow}(x) = f(x)$ a f^{\leftarrow} je identická s f^{-1} . Funkce f^{\leftarrow} zavedená vztahem (1.3) je tudíž zobecněním pojmu inverzní funkce.

DEFINICE 1.2. Necht' f je neklesající reálná funkce. Pak se funkce f^{\leftarrow} daná vztahem (1.3) nazývá *zleva spojitou inverzní funkcí k funkci $f(x)$* .

Zřejmě, je-li F distribuční funkcí nějaké náhodné veličiny, pak F^{\leftarrow} je běžně užívanou kvantilovou funkcí. Dále pomocí pojmu zleva spojitě inverze zavedeme kvantilovou funkci chvostu. Podrobné vlastnosti zleva spojitých inverzních funkcí včetně jejich důkazů lze nalézt např. v monografiích [112] či [49].

DEFINICE 1.3. Necht' F je distribuční funkce náhodné veličiny X . Pak funkci U danou vztahem

$$U(t) = \left(\frac{1}{1 - F(t)} \right)^{\leftarrow} \quad (1.4)$$

definovanou pro $t > 1$, nazýváme *kvantilovou funkcí chvostu náhodné veličiny X* .

Snadno lze ukázat [49], že kvantilovou funkcí chvostu z definice 1.3 můžeme vyjádřit v následujícím tvaru, tedy

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

LEMMA 1.4. Necht' f_n je posloupnost neklesajících funkcí a g je neklesající funkce. Necht' je dán otevřený interval $I := (a, b)$ a pro každý bod spojitosti $x \in I$ funkce g necht' platí

$$\lim_{n \rightarrow \infty} f_n(x) = g(x).$$

Bud'te $f_n^{\leftarrow}, g^{\leftarrow}$ zleva spojitě inverzní funkce k funkcím f_n a g . Pak pro všechna $x \in (g(a), g(b))$, která jsou body spojitosti funkce g^{\leftarrow} , platí

$$\lim_{n \rightarrow \infty} f_n^{\leftarrow}(x) = g^{\leftarrow}(x). \quad (1.5)$$

Důkaz. Předěšlé lemma je dokázáno např. v práci [38]. □

VĚTA 1.5. Necht' $a_n > 0, b_n, n \in \mathbb{N}$, jsou reálné konstanty a $G(x)$ nedegenerovaná distribuční funkce. Pak následující tvrzení jsou ekvivalentní:

1. Pro každý bod spojitosti $x \in \mathbb{R}$ funkce $G(x)$ platí

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

2. Pro každý bod spojitosti $x > 0$ funkce $D(x) := G^{\leftarrow}(e^{-1/x})$ platí

$$\lim_{t \rightarrow \infty} \frac{U(tx) - b(t)}{a(t)} = D(x), \quad (1.6)$$

přičemž $a(t) := a_{[t]}, b(t) := b_{[t]}$, kde $[t]$ označuje celočíselnou² část t .

²Přesněji $[t] = \max\{m \in \mathbb{Z} : m \leq t\}$, protože ovšem uvažujeme jen hodnoty $t > 0$, jedná se tak o celočíselnou část t .

Důkaz. Vyjdeme z platnosti podmínky atraktivnosti (1.2), tedy $\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G(x)$. Pro každý bod spojitosti $x \in \mathbb{R}$ funkce $G(x)$ takový, že $0 < G(x) < 1$, lze logaritmováním obou stran rovnice dostat (1.2) ve tvaru

$$\lim_{n \rightarrow \infty} n \ln F(a_n x + b_n) = \ln G(x). \quad (1.7)$$

Snadno se ukáže, že platí

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

a proto je tedy (1.7) ekvivalentní výrazu

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

nebo také výrazu

$$\lim_{n \rightarrow \infty} \frac{1}{n [1 - F(a_n x + b_n)]} = \frac{1}{-\ln G(x)}. \quad (1.8)$$

Nyní, dle lemmatu 1.4, přeformulujeme rovnost (1.8) pomocí zleva spojitých inverzních funkcí. Nechť tedy $U(t) = (1/(1 - F(t)))^{\leftarrow}$, jak bylo zavedeno v definici 1.3. Pak pro $x > 0$ je (1.8) ekvivalentní rovnosti

$$\lim_{n \rightarrow \infty} \frac{U(nx) - b_n}{a_n} = D(x), \quad (1.9)$$

přičemž bylo zavedeno označení $D(x) := G^{\leftarrow}(e^{-1/x})$. Dále nechť $x > 0$ je bodem spojitosti funkce $D(x)$. Pak pro $t \geq 1$ máme

$$\frac{U(\lfloor t \rfloor x) - b_{\lfloor t \rfloor}}{a_{\lfloor t \rfloor}} \leq \frac{U(tx) - b_{\lfloor t \rfloor}}{a_{\lfloor t \rfloor}} \leq \frac{U(\lfloor t \rfloor x(1 + 1/\lfloor t \rfloor)) - b_{\lfloor t \rfloor}}{a_{\lfloor t \rfloor}}. \quad (1.10)$$

Pro libovolný bod spojitosti $x_0 > x$ funkce D s vlastností $D(x_0) > D(x)$ platí, že pravá strana vztahu (1.10) je menší než $D(x_0)$. Protože x je bodem spojitosti funkce $D(x)$, dostaneme

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

□

Články Fishera a Tippetta (1928, [52]) a dále Gnedenka (1943, [59]) byly jedny z prvních, které se zabývaly rozdělením výběrových maxim. Z historického hlediska byly postupně odvozeny tři třídy možných nedegenerovaných limitních rozdělení, které splňují rovnost (1.2). Tyto třídy jsou známé jako Gumbelova, Fréchetova a Weibullova třída, také označované po řadě jako EV rozdělení typu I, II a III. Konkrétní vlastnosti rozdělení náležících do těchto tříd budou později detailně popsány v odstavci 1.3. Následující tvrzení, také označované jako *Základní věta teorie extrémních hodnot*, popisuje třídu EV rozdělení obecněji pomocí parametrizace odvozené nezávisle von Misesem (1954, [130]) a Jenkinsonem (1955, [82]).

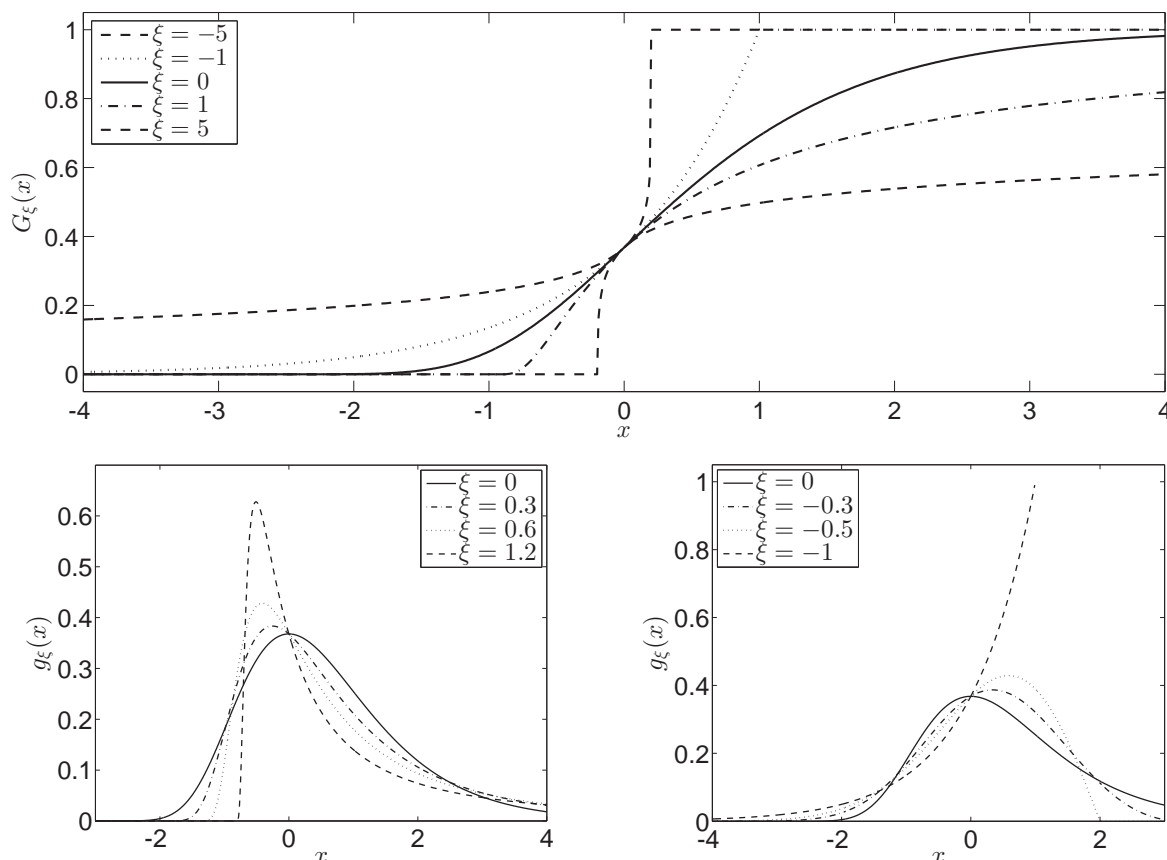
VĚTA 1.6. FISHER A TIPPETT (1928), GNEDENKO (1943) NechtĚ je splněna podmínka (1.2) pro nějaké reálné konstanty $a_n > 0, b_n$ ($n \in \mathbb{N}$). Pak distribuční funkce EV rozdělení je ve tvaru $G_\xi(\sigma x + \mu)$ pro $\mu \in \mathbb{R}, \sigma > 0$ a $\xi \in \mathbb{R}$, kde

$$G_\xi(x) = \begin{cases} \exp \left[- (1 + \xi x)^{-1/\xi} \right], & \xi \neq 0, \\ \exp [-e^{-x}], & \xi = 0, \end{cases} \quad (1.11)$$

definovaná na $\{x : 1 + \xi x > 0\}$.

Vzhledem k von Misesově parametrizaci, která shrnuje všechny třídy EV rozdělení (viz odstavec 1.3), je rozdělení s distribuční funkcí (1.11) označováno jako *zobecněné EV rozdělení* (zkráceně GEV rozdělení z angl. Generalized Extreme Value). Vyjma parametru polohy μ a parametru měřítka $\sigma > 0$ tvoří tedy GEV rozdělení jednoparametrickou třídu rozdělení s parametrem tvaru ξ .

POZNÁMKA. Parametr ξ má vzhledem k (1.11) zásadní vliv na chvosty EV rozdělení, jak je naznačeno níže na obrázku 1.1, a je tak často označován jako tzv. *index extrémní hodnoty* (zkráceně EV index). Dále budeme označením $F \in \mathcal{D}(G_\xi)$ rozumět, že distribuční funkce F patří do oboru atraktivit EV rozdělení s EV indexem ξ . Snadno se ukáže, že obor atraktivit je dán jednoznačně, tj. pokud $F \in \mathcal{D}(G_{\xi_1})$ a $F \in \mathcal{D}(G_{\xi_2})$, pak nutně $\xi_1 = \xi_2$.



Obrázek 1.1: Distribuční funkce EV rozdělení (nahore) a hustoty EV rozdělení (dole) pro různé hodnoty EV indexu.

Důkaz věty 1.6. Předpokládejme, že $x = 1$ je bodem spojitosti funkce $D(x)$. Pokud by tomu tak nebylo, provedly by se následující kroky důkazu pro jiný bod spojitosti x_0 , přičemž bychom $U(t)$ nahradili za $U(x_0t)$. Pro $x > 0$ tedy zavedme funkci

$$E(x) := D(x) - D(1) = \lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)}. \quad (1.12)$$

Pro $y > 0$ můžeme rozepsat

$$\frac{U(txy) - U(t)}{a(t)} = \frac{U(txy) - U(ty)}{a(ty)} \frac{a(ty)}{a(t)} + \frac{U(ty) - U(t)}{a(t)}. \quad (1.13)$$

Snadno se ověří, že obě limity $\lim_{t \rightarrow \infty} a(ty)/a(t)$ i $\lim_{t \rightarrow \infty} (U(ty) - U(t))/a(t)$ existují pro $y > 0$. Označme tedy

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

Pak pro $x, y > 0$ můžeme s využitím vztahu (1.13) psát

$$E(xy) = \lim_{t \rightarrow \infty} \frac{U(txy) - U(t)}{a(t)} = E(x)A(y) + E(y).$$

Nyní zavedme označení $H(x) = E(e^x)$ a pro $x, y \neq 1$ buď $s := \ln x, t := \ln y$. Předchozí rovnost tak můžeme přepsat a postupnými úpravami získáme

$$\begin{aligned} H(t+s) &= H(s)A(e^t) + H(t), \\ \frac{H(t+s) - H(t)}{s} &= \frac{H(s) - H(0)}{s} A(e^t), \end{aligned} \quad (1.14)$$

kde $H(0) = 0$. Protože funkce H je monotónní, zcela jistě existuje nějaké t , v němž je H diferencovatelná. Navíc, dle předchozího vyjádření, je H diferencovatelná ve všech bodech a pro $s \rightarrow 0$ tak dostaneme

$$H'(t) = H'(0)A(e^t). \quad (1.15)$$

Funkce H nemůže být konstantní, protože funkce G_ξ je nedegenerovaná, tedy $H'(0) \neq 0$. Položme $Q(t) := H(t)/H'(0)$, přičemž $Q(0) = 0, Q'(0) = 1$. Ze vztahu (1.14) plyne

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

což můžeme s využitím (1.15) přepsat jako

$$Q(t+s) = Q(s)Q'(t) + Q(t). \quad (1.16)$$

Zaměníme-li nyní v (1.16) t a s , pak vzhledem ke komutativitě sčítání plyne z (1.16)

$$Q(s)Q'(t) + Q(t) = Q(t)Q'(s) + Q(s).$$

Odtud úpravou a dále pro $s \rightarrow 0$ dostáváme po řadě

$$\begin{aligned} Q(t) \frac{Q'(s) - 1}{s} &= \frac{Q(s)}{s} (Q'(t) - 1), \\ Q(t)Q''(0) &= Q'(t) - 1. \end{aligned} \quad (1.17)$$

Vztah (1.17) derivujeme vzhledem k t a máme

$$Q''(0)Q'(t) = Q''(t),$$

neboli

$$(\ln Q'(t))' = Q''(0) =: \xi \in \mathbb{R}.$$

Odtud vzhledem k počáteční podmínce $Q(0) = 0$ plynou následující rovnosti

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

Nyní postupně zpětnými kroky dostáváme následující vyjádření pro funkce H , E a D , tudíž

$$\begin{aligned} H(t) &= H'(0) \frac{e^{\xi t} - 1}{\xi}, \\ E(t) &= H'(0) \frac{t^\xi - 1}{\xi}, \\ D(t) &= D(1) + H'(0) \frac{t^\xi - 1}{\xi}. \end{aligned} \tag{1.18}$$

Odtud plyne, že zleva spojitá inverzní funkce je ve tvaru

$$D^{\leftarrow}(t) = \left(1 + \xi \frac{t - D(1)}{H'(0)} \right)^{1/\xi}. \tag{1.19}$$

Protože $D(t)$ byla zavedena jako $D(t) = G^{\leftarrow}(e^{-1/t})$, lze vyjádřit

$$D^{\leftarrow}(t) = \frac{-1}{\ln G(t)}. \tag{1.20}$$

Porovnáním rovnic (1.19) a (1.20) obdržíme konečně

$$G(t) = \exp \left[- \left(1 + \xi \frac{t - D(1)}{H'(0)} \right)^{-1/\xi} \right].$$

Označíme-li nyní $\mu := D(1)$, $\sigma := H'(0)$ a $x = (t - \mu)/\sigma$, dostáváme funkci $G_\xi(x)$ ve tvaru (1.11) a tvrzení je tak dokázáno. \square

POZNÁMKA. Stojí za zmínku, že třída EV rozdělení je totožná s třídou tzv. *max-stabilních rozdělení*, tj. rozdělení, pro která existují takové reálné konstanty $a_n > 0$, b_n , že pro všechna $x \in \mathbb{R}$ platí

$$G_\xi^n(a_n x + b_n) = G_\xi(x). \tag{1.21}$$

Tato vlastnost hraje také významnou roli v kapitole 5, zejména pak fakt, že EV index je totožný pro funkce na obou stranách předešlého vztahu.

Závěrem tohoto odstavce ještě formulujeme ekvivalentní tvary podmínky atraktivity (1.2), které budeme často využívat v následujícím textu.

VĚTA 1.7. Následující tvrzení jsou ekvivalentní:

1. Pro každé $x \in \mathbb{R}$ existují reálné konstanty $a_n > 0, b_n (n \in \mathbb{N})$ tak, že

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G_\xi(x). \quad (1.22)$$

2. Existuje kladná funkce $a(t)$ taková, že pro $x > 0$ platí

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = D_\xi(x) := \frac{x^\xi - 1}{\xi}, \quad (1.23)$$

přičemž pro $\xi = 0$ chápeme pravou stranu jako $D_\xi(x) = \ln x$.

3. Existuje kladná funkce $\varphi(t)$ tak, že

$$\lim_{t \uparrow x^*} \frac{1 - F(t + x\varphi(t))}{1 - F(t)} = (1 + \xi x)^{-1/\xi} \quad (1.24)$$

pro všechna $x \in \mathbb{R}$ splňující $1 + \xi x > 0$.

Důkaz. Ekvivalence (1.22) a (1.23) byla ověřena již v důkaze věty 1.6, kde z (1.18) plyne $D_\xi(x) = E(x)/H'(0)$. Zbývá tedy dokázat, že (1.23) je ekvivalentní s (1.24).

Bud' f neklesající funkce. Z vlastností zleva spojitých inverzních funkcí (viz např. [112]) plyne pro $x \in \mathbb{R}$ $f^{\leftarrow}(f(x)) \leq x$. Zřejmě lze pak pro libovolné $\varepsilon > 0$ psát

$$f^{\leftarrow}(f(x) - \varepsilon) \leq x \leq f^{\leftarrow}(f(x) + \varepsilon).$$

Odtud se pro $t \uparrow x^*$ dostane

$$\begin{aligned} \frac{(1 - \varepsilon)^\xi - 1}{\xi} &\leftarrow \frac{U\left(\frac{1 - \varepsilon}{1 - F(t)}\right) - U\left(\frac{1}{1 - F(t)}\right)}{a\left(\frac{1}{1 - F(t)}\right)} < \frac{t - U\left(\frac{1}{1 - F(t)}\right)}{a\left(\frac{1}{1 - F(t)}\right)} < \\ &< \frac{U\left(\frac{1 + \varepsilon}{1 - F(t)}\right) - U\left(\frac{1}{1 - F(t)}\right)}{a\left(\frac{1}{1 - F(t)}\right)} \rightarrow \frac{(1 + \varepsilon)^\xi - 1}{\xi}, \end{aligned} \quad (1.25)$$

a protože tato nerovnost platí pro každé $\varepsilon > 0$, vyplývá, že

$$\lim_{t \uparrow x^*} \frac{t - U\left(\frac{1}{1 - F(t)}\right)}{a\left(\frac{1}{1 - F(t)}\right)} = 0.$$

Za předpokladu (1.23) tedy pro všechna $x > 0$ dostáváme

$$\lim_{t \uparrow x^*} \frac{U\left(\frac{x}{1 - F(t)}\right) - t}{a\left(\frac{1}{1 - F(t)}\right)} = \frac{x^\xi - 1}{\xi}.$$

Nyní využijeme lemmatu 1.4 a předchozí vztah přepíšeme pomocí zleva spojitých inverzních funkcí. Obdrží se tudíž

$$\lim_{t \uparrow x^*} \frac{1 - F(t)}{1 - F\left(t + xa\left(\frac{1}{1 - F(t)}\right)\right)} = (1 + \xi x)^{1/\xi},$$

čímž je dokázáno, že (1.23) implikuje podmínku (1.24). Opačná implikace se pak dokáže obdobně. \square

POZNÁMKA. Rovnost (1.22) je splněna pro $a_n = a(n)$ a $b_n = U(n)$ ze vztahu (1.23). Podmínka (1.24) je splněna pro $\varphi(t) = a(1/(1 - F(t)))$. Funkce $a(t)$ z podmínky (1.23) je označována jako *pomocná funkce prvního řádu*. Určení této funkce pro známou kvantilovou funkci chvostu $U(t)$ dovoluje ověřit, zda příslušné rozdělení pravděpodobnosti splňuje (1.23), resp. (1.22), a rozhodnout tak o oboru atraktivit extrémních hodnot. Odvození konkrétních tvarů pomocných funkcí $a(t)$ pro vybraná rozdělení pravděpodobnosti je možné nalézt v příloze A.

1.3. Základní typy EV rozdělení

Obrázek 1.1 ilustruje třídu EV rozdělení pro různé hodnoty EV indexu ξ a ukazuje také, že tato třída zahrnuje rozdělení s celou řadou různých vlastností. Tradičně se rozlišují tři podtřídy EV rozdělení tak, jak byly historicky odvozeny, a sice Fréchetova třída pro $\xi > 0$, Gumbelova třída pro $\xi = 0$ a Weibullova třída pro $\xi < 0$. Jednotlivé případy budou podrobněji rozebrány níže. Funkční charakteristiky jednotlivých tříd rozdělení je možné komplexně popsat pomocí *regulárně se měnících funkcí*, a proto bude nejdříve tento pojem zaveden v následující definici.

DEFINICE 1.8. Kladná měřitelná funkce $f : \mathbb{R}^+ \rightarrow \mathbb{R}$ se nazývá *regulárně se měnící funkcí (v nekonečnu)*, pokud pro všechna $x > 0$ a nějaké $\alpha \in \mathbb{R}$ platí

$$\lim_{t \rightarrow \infty} \frac{f(tx)}{f(t)} = x^\alpha. \quad (1.26)$$

Píšeme pak $f \in \mathcal{R}_\alpha$. Číslo α se pak nazývá *index regulární variace* (zkráceně RV index). Je-li navíc $\alpha = 0$, nazývá se pak funkce f *pomalou se měnící funkcí*.

Fréchetova třída EV rozdělení ($\xi > 0$)

V literatuře je tato třída označována také jako EV rozdělení typu II. Z historického hlediska je distribuční funkce rozdělení Fréchetova typu často uváděna v následujícím tvaru při parametrizaci $\Phi_\alpha(x) = G_\xi((x - 1)/\xi)$, tedy

$$\Phi_\alpha(x) = \begin{cases} 0, & x \leq 0, \\ \exp(-x^{-\alpha}), & x > 0, \end{cases}$$

kde $\alpha = 1/\xi > 0$.

Pro rozdělení v této třídě platí $G_\xi(x) < 1$ pro všechna $x \in \mathbb{R}$, tj. pravý koncový bod $x^* = \infty$. Navíc, levý koncový bod $x_* := \inf\{x : F(x) > 0\}$ je konečný a je roven $-1/\xi$. Pro $x \rightarrow \infty$ je $1 - G_\xi(x) \simeq (\xi x)^{-1/\xi}$, kde symbol \simeq značí asymptotickou ekvivalenci, tedy

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

a tudíž rozdělení z této třídy mají těžké pravé chvosty. Neexistují tak momenty řádu většího nebo rovného $1/\xi$, což kupříkladu pro $\xi \geq 1/2$ znamená, že rozptyl není konečný, a pro $\xi \geq 1$ dokonce neexistuje ani střední hodnota.

Významná rozdělení patřící do oboru atraktivit Fréchetova rozdělení shrnuje tabulka 1.1. Patří mezi ně např. Studentovo, Fisherovo-Snedecorovo či Fréchetovo rozdělení. Navíc do této třídy patří všechna rozdělení Paretova typu, tj. rozdělení, jejichž distribuční funkce a kvantilová funkce chvostu jsou pro $\xi > 0$ ve tvaru

$$\begin{aligned} 1 - F(x) &= x^{-1/\xi} \ell_F(x), \\ U(t) &= t^\xi \ell_U(t), \end{aligned}$$

kde $\ell_F(x), \ell_U(t) \in \mathcal{R}_0$. Tato rozdělení s těžkými chvosty hrají důležitou roli v mnoha modelech pro odhady parametrů rozdělení extrémních hodnot, jak bude diskutováno v kapitole 3. Detailní popis Fréchetovy třídy rozdělení ve smyslu regulárně se měnících funkcí nabízí následující věta.

VĚTA 1.9. Následující tvrzení jsou ekvivalentní:

- (i) $F \in \mathcal{D}(G_\xi)$ pro $\xi > 0$,
- (ii) $x^* = \infty$ a platí $1 - F(t) \in \mathcal{R}_{-1/\xi}$,
- (iii) $U(t) \in \mathcal{R}_\xi$.

Důkaz. Podrobně provedeme důkaz pro jednostranné implikace (i) \Rightarrow (ii) a (i) \Rightarrow (iii), opačné implikace lze dokázat obdobně. Nejprve se budeme zabývat důkazem implikace mezi tvrzeními (i) a (iii).

Nechť $F \in \mathcal{D}(G_\xi)$ pro $\xi > 0$. Pak je splněna podmínka (1.23) a tato implikuje

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{a(tx)}{a(t)} &= \lim_{t \rightarrow \infty} \left(\frac{U(txy) - U(t)}{a(t)} - \frac{U(tx) - U(t)}{a(t)} \right) / \left(\frac{U(txy) - U(tx)}{a(tx)} \right) = \\ &= \left(\frac{(xy)^\xi - 1}{\xi} - \frac{x^\xi - 1}{\xi} \right) / \left(\frac{y^\xi - 1}{\xi} \right) = x^\xi. \end{aligned}$$

Vzhledem k předcházející rovnosti pro $Z > 1$ platí

$$\lim_{k \rightarrow \infty} \frac{U(Z^{k+1}) - U(Z^k)}{U(Z^k) - U(Z^{k-1})} = \lim_{k \rightarrow \infty} \left(\frac{U(Z^{k+1}) - U(Z^k)}{a(Z^k)} \right) / \left(\frac{U(Z^k) - U(Z^{k-1})}{Z^\xi a(Z^{k-1})} \right) = Z^\xi,$$

a tudíž pro $0 < \varepsilon < 1 - Z^{-\xi}$, $k \geq n_0(\varepsilon)$, kde $n_0(\varepsilon)$ je dostatečně velké, dostáváme

$$\begin{aligned} (U(Z^k) - U(Z^{k-1})) Z^\xi (1 - \varepsilon) &\leq U(Z^{k+1}) - U(Z^k) \leq \\ &\leq (U(Z^k) - U(Z^{k-1})) Z^\xi (1 + \varepsilon). \end{aligned} \quad (1.27)$$

Nyní nerovnosti (1.27) sečteme pro $k = n_0, \dots, n$, výsledek vydělíme výrazem $U(Z^n)$ a dostaneme tak nerovnost

$$Z^\xi (1 - \varepsilon) \frac{U(Z^n) - U(Z^{n_0-1})}{U(Z^n)} \leq \frac{U(Z^{n+1}) - U(Z^{n_0})}{U(Z^n)} \leq Z^\xi (1 + \varepsilon) \frac{U(Z^n) - U(Z^{n_0-1})}{U(Z^n)}.$$

Pro $n \rightarrow \infty$ se obdrží

$$Z^\xi (1 - \varepsilon) \leq \lim_{n \rightarrow \infty} \frac{U(Z^{n+1}) - U(Z^{n_0})}{U(Z^n)} = \lim_{n \rightarrow \infty} \frac{U(Z^{n+1})}{U(Z^n)} \leq Z^\xi (1 + \varepsilon),$$

a poněvadž tato nerovnost platí pro libovolné $\varepsilon > 0$, plyne odtud

$$\lim_{n \rightarrow \infty} \frac{U(Z^{n+1})}{U(Z^n)} = Z^\xi \quad (1.28)$$

pro všechna $Z > 1$. Dále pro každé $x > 1$ definujeme číslo $n(x) \in \mathbb{N}$ tak, aby

$$Z^{n(x)} \leq x < Z^{n(x)+1}. \quad (1.29)$$

Pro $t, x > 1$ potom platí nerovnosti

$$\frac{U(Z^{n(t)} Z^{n(x)})}{U(Z^{n(t)+1})} \leq \frac{U(tx)}{U(t)} \leq \frac{U(Z^{n(t)+1} Z^{n(x)+1})}{U(Z^{n(t)})}. \quad (1.30)$$

Levou stranu (1.30) je možné rozepsat jako

$$\frac{U(Z^{n(t)} Z^{n(x)})}{U(Z^{n(t)})} \frac{U(Z^{n(t)})}{U(Z^{n(t)+1})}.$$

Dle vztahu (1.28) tento výraz pro $t \rightarrow \infty$ konverguje k výrazu $Z^{\xi n(x)} Z^{-\xi}$, který je dle (1.29) zdola ohraničen $(x/Z^2)^\xi$. Podobně pravá strana výrazu (1.30) je shora ohraničena $(xZ^2)^\xi$. Pro $Z \rightarrow 1$ tak dostáváme přímo tvrzení (iii).

Nyní dokážeme, že z tvrzení (iii) plyne (ii). Z definice kvantilové funkce chvostu lze pro $\varepsilon > 0$ psát

$$U\left(\frac{1-\varepsilon}{1-F(t)}\right) \leq t \leq U\left(\frac{1+\varepsilon}{1-F(t)}\right),$$

tedy

$$\frac{U\left(\frac{x}{1-F(t)}\right)}{U\left(\frac{1+\varepsilon}{1-F(t)}\right)} \leq \frac{1}{t} U\left(\frac{x}{1-F(t)}\right) \leq \frac{U\left(\frac{x}{1-F(t)}\right)}{U\left(\frac{1-\varepsilon}{1-F(t)}\right)}. \quad (1.31)$$

Nechť $U(t) \in \mathcal{R}_\xi$. Pak levá a pravá strana výrazu (1.31) konvergují k výrazům $(x/(1+\varepsilon))^\xi$ a $(x/(1-\varepsilon))^\xi$. Protože ale nerovnost (1.31) platí pro všechna $\varepsilon > 0$, dostane se rovnost

$$\lim_{t \rightarrow \infty} \frac{1}{t} U\left(\frac{x}{1-F(t)}\right) = x^\xi.$$

Dále se použije lemma 1.4 a předchozí rovnice se vyjádří pomocí zleva spojitých inverzních funkcí. Obdrží se tudíž

$$\lim_{t \rightarrow \infty} \frac{1-F(t)}{1-F(tx)} = x^{1/\xi}$$

a implikace je tak dokázána. □

Gumbelova třída EV rozdělení ($\xi = 0$)

V případě Gumbelovy třídy (též EV rozdělení typu I) je pravý koncový bod $x^* = \infty$ a levý koncový bod $x_* = -\infty$, tedy nosič distribuční funkce pokrývá celou reálnou osu a pro všechna $x \in \mathbb{R}$ je $0 < G_0(x) < 1$. Distribuční funkce Gumbelova oboru atraktivity je přímo ve tvaru (1.11), tj.

$$G_0(x) = \exp(-e^{-x}).$$

Oproti Fréchetově třídě zde pro $x \rightarrow \infty$ dostáváme $1 - G_0(x) \simeq e^{-x}$ a momenty všech řádů existují. Pro rozdělení patřící do Gumbelovy třídy lze ukázat (viz [38]), že $a(t) \in \mathcal{R}_0$ a také $U(t) \in \mathcal{R}_0$. Některá významná rozdělení patřící do Gumbelova oboru atraktivity jsou shrnuta v tabulce 1.2, řadí se mezi ně např. exponenciální, normální či Weibullovo rozdělení.

Weibullova třída EV rozdělení ($\xi < 0$)

Weibullova třída, také označována jako EV rozdělení typu III, zahrnuje všechna rozdělení $G_\xi(x)$ s EV indexem $\xi < 0$. Její distribuční funkce se často vyskytuje v následujícím tvaru při reparametrizaci $\Psi_\alpha(x) = G_\xi(-(1+x)/\xi)$ pro $\alpha = -1/\xi > 0$, tedy

$$\Psi_\alpha(x) = \begin{cases} \exp(-(-x)^\alpha), & x < 0, \\ 1, & x \geq 0. \end{cases}$$

Pravý koncový bod pro rozdělení z Weibullové třídy je konečný a $x^* = -1/\xi$. Naproti tomu je levý koncový bod $x_* = -\infty$. Protože pro $x \rightarrow \infty$ dostáváme $1 - G_\xi(x^* - x) \simeq (-\xi x)^{-1/\xi}$, mají rozdělení omezené pravé chvosty. Pro rozdělení patřící do Weibullové třídy lze ukázat (viz [38]), že $x^* - U(t) \in \mathcal{R}_\xi$ a $1 - F(x^* - t) \in \mathcal{R}_{-1/\xi}$. Důkaz zde lze vést obdobně jako u tvrzení 1.9. Tabulka 1.3 popisuje vybraná rozdělení pravděpodobnosti z Weibullova oboru atraktivity, řadí se sem mimo jiné např. rovnoměrné či beta rozdělení.

Vzhledem k tomu, že třída EV rozdělení je totožná s třídou max-stabilních rozdělení splňujících (1.21), zřejmě vždy platí $G_\xi \in \mathcal{D}(G_\xi)$, tedy kupříkladu Gumbelovo rozdělení patří do Gumbelova oboru atraktivity apod.

V závěru tohoto odstavce ještě stanovíme von Misesovu postačující podmínku pro to, aby daná distribuční funkce patřila do oboru atraktivity nějakého EV rozdělení. Tato podmínka je často nejvhodnějším způsobem, jak tuto skutečnost ověřit.

VĚTA 1.10. VON MISESOVA PODMÍNKA ATRAKTIVITY Buď $F(x)$ distribuční funkce s pravým koncovým bodem x^* . Nechť existuje $F''(x)$ a nechť $F'(x)$ je kladná pro všechna x v nějakém levém okolí bodu x^* . Pak $F \in \mathcal{D}(G_\xi)$, pokud

$$\lim_{t \uparrow x^*} \left(\frac{1 - F(t)}{F'(t)} \right)' = \xi. \quad (1.32)$$

Důkaz. Úpravou vztahu (1.32) lze ukázat, že platí rovnost

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

která je ekvivalentní výrazu

$$\lim_{t \rightarrow \infty} \frac{tU''(t)}{U'(t)} = \xi - 1. \quad (1.33)$$

Ukážeme, že (1.33) implikuje podmínku (1.23), a tedy i podmínku atraktivitý (1.22). Vyděme z následující triviální rovnosti

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

Derivací předešlého vztahu dostaneme nejprve

$$U'(t) = \frac{[1 - F(U(t))]^2}{F'(U(t))}, \quad (1.34)$$

derivací (1.34) po úpravě postupně obdržíme

$$\begin{aligned} \frac{U''(t)}{U'(t)} &= -2[1 - F(U(t))] - \frac{[1 - F(U(t))]^2 F''(U(t))}{[F'(U(t))]^2}, \\ \frac{tU''(t)}{U'(t)} &= -2 - \frac{[1 - F(U(t))] F''(U(t))}{[F'(U(t))]^2}. \end{aligned} \quad (1.35)$$

Pro $x > 0$ a $t, tx > 1$, platí

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

přičemž s využitím (1.35) lze dále psát

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

Jelikož pro $0 < a < b < \infty$ platí

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

a pro nějaké kladné k na kompaktním intervalu platí $|e^s - e^t| < k|s - t|$, dostane se tak

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

Odtud už vyplývá následující konvergence

$$\frac{U(tx) - U(t)}{tU'(t)} - \frac{x^\xi - 1}{\xi} = \int_1^x \left(\frac{U'(ts)}{U'(t)} - s^{\xi-1} \right) ds \rightarrow 0$$

a platnost tvrzení je tak dokázána. \square

Některé další tvary předešlé postačující podmínky pro konkrétní případy $\xi > 0$, $\xi = 0$ a $\xi < 0$ jsou odvozeny např. v [38].

Tabulka 1.1: Rozdělení patřící do Fréchetova oboru atraktivity.

| Rozdělení | Distribuční funkce $F(t)$ a kvantilová funkce chvostu $U(t)$ | EV index |
|------------------------|--|-------------------------|
| Paretovo | $F(t) = 1 - \left(\frac{\alpha}{x}\right)^{1/\xi}, \quad x \geq \alpha; \xi > 0, \alpha \in \mathbb{R}$ $U(t) = \alpha t^\xi, \quad t > 1$ | ξ |
| Zobecněné Paretovo | $F(x) = 1 - (1 + \xi x)^{-1/\xi}, \quad x > 0; \xi > 0$ $U(t) = (t^\xi - 1)/\xi, \quad t > 1$ | ξ |
| Fréchetovo | $F(x) = \exp(-x^{-1/\xi}), \quad x \geq 0; \xi > 0$ $U(t) = \left(-\ln\left(1 - \frac{1}{t}\right)\right)^{-\xi}, \quad t > 1$ | ξ |
| $F(m, n)$ | $\frac{dF(x)}{dx} = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1} \left(1 + \frac{m}{n}x\right)^{-\frac{m+n}{2}},$ $x > 0; m, n > 0$ předpis pro $U(t)$ není znám | $\frac{2}{n}$ |
| Studentovo | $\frac{dF(x)}{dx} = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad x \in \mathbb{R}; n > 0$ předpis pro $U(t)$ není znám | $\frac{1}{n}$ |
| Burrovo (typ XII) | $F(x) = 1 - \left(\frac{\eta}{\eta+x^\tau}\right)^\lambda, \quad x > 0; \eta, \tau, \lambda > 0$ $U(t) = (\eta [t^{1/\lambda} - 1])^{1/\tau}, \quad t > 1$ | $\frac{1}{\lambda\tau}$ |
| Burrovo (typ III) | $F(x) = \left(\frac{\eta}{\eta+x^{-\tau}}\right)^\lambda, \quad x > 0; \eta, \tau, \lambda > 0$ $U(t) = \left(\frac{\eta}{(1-t^{-1})^{1/\lambda}} - \eta\right)^{-1/\tau}, \quad t > 1$ | $\frac{1}{\tau}$ |
| Inverzní gama | $\frac{dF(x)}{dx} = \frac{\lambda^\alpha}{\Gamma(\alpha)} e^{-\lambda/x} x^{-\alpha-1}, \quad x > 0; \lambda, \alpha > 0$ předpis pro $U(t)$ není znám | $\frac{1}{\alpha}$ |
| Logaritmičké gama | $\frac{dF(x)}{dx} = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{-\lambda-1} (\ln x)^{\alpha-1}, \quad x > 1; \lambda, \alpha > 0$ předpis pro $U(t)$ není znám | $\frac{1}{\lambda}$ |
| Loglogistické | $F(x) = \left(1 + \left(\frac{x}{\alpha}\right)^{-\beta}\right)^{-1}, \quad x > 0; \alpha, \beta > 0$ $U(t) = \alpha(t-1)^{1/\beta}, \quad t > 1$ | $\frac{1}{\beta}$ |
| Halova třída rozdělení | $U(t) = Ct^\xi (1 + Dt^\rho + o(t^\rho))$ pro $t > 1$; $C > 0, \xi > 0, \rho < 0$ | ξ |

Tabulka 1.2: Rozdělení patřící do Gumbelova oboru atraktivity.

| Rozdělení | Distribuční funkce $F(t)$ a kvantilová funkce chvostu $U(t)$ |
|------------------------|---|
| Exponenciální | $F(x) = 1 - \exp(-\lambda x), \quad x > 0; \lambda > 0$ $U(t) = \frac{1}{\lambda} \ln t, \quad t > 1$ |
| Weibullovo | $F(x) = 1 - \exp(-(x/\lambda)^\alpha), \quad x > 0; \lambda, \alpha > 0$ $U(t) = \lambda (\ln t)^{1/\alpha}, \quad t > 1$ |
| Normální | $\frac{dF(x)}{dx} = (\sigma\sqrt{2\pi})^{-1} \exp\left(-\frac{(x-\mu)^2}{\sigma^2}\right), \quad x \in \mathbb{R};$ $\mu \in \mathbb{R}, \sigma > 0$ předpis pro $U(t)$ není znám |
| Logaritmicko normální | $\frac{dF(x)}{dx} = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2} (\ln x - \mu)^2\right\}, \quad x > 0;$ $\mu \in \mathbb{R}, \sigma > 0$ předpis pro $U(t)$ není znám |
| Gama | $\frac{dF(x)}{dx} = \frac{\lambda^m}{\Gamma(m)} \exp(-\lambda x)x^{m-1}, \quad x > 0; \lambda, m > 0$ předpis pro $U(t)$ není znám |
| Logistické | $F(x) = (1 + \exp(-\frac{x-\mu}{\sigma}))^{-1}, \quad x \in \mathbb{R}; \mu \in \mathbb{R},$ $\sigma > 0$ $U(t) = \mu + \sigma \ln(t-1), \quad t > 1$ |
| Benktanderovo (typ II) | $F(x) = 1 - x^{-(1-\beta)} \exp\left\{-\frac{\alpha}{\beta} (x^\beta - 1)\right\},$ $x > 0; \alpha, \beta > 0$ předpis pro $U(t)$ není znám |
| Gumbelovo | $F(x) = \exp(-e^{-x}), \quad x \in \mathbb{R}$ $U(t) = -\ln\left(\ln \frac{t}{t-1}\right), \quad t > 1$ |

Tabulka 1.3: Rozdělení patřící do Weibullova oboru atraktivity.

| Rozdělení | Distribuční funkce $F(t)$ a kvantilová funkce chvostu $U(t)$ | EV index |
|-----------------------|---|--------------------------|
| Rovnoměrné | $F(x) = \frac{x-a}{b-a}, \quad a \leq x \leq b; a, b \in \mathbb{R}, a < b$ $U(t) = \left(1 - \frac{1}{t}\right)(b-a) + a, \quad t > 1$ | -1 |
| Beta | $\frac{dF(x)}{dx} = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} x^{p-1}(1-x)^{q-1}, 0 \leq x \leq 1;$ $p, q > 0$ předpis pro $U(t)$ není znám | $-\frac{1}{q}$ |
| Reverzní Burrovo | $F(x) = 1 - \left(\frac{\eta}{\eta+(-x)^{-\tau}}\right)^\lambda, \quad x \leq 0; \eta, \tau, \lambda > 0$ $U(t) = -(\eta[t^{1/\lambda} - 1])^{-1/\tau}, \quad t > 1$ | $-\frac{1}{\lambda\tau}$ |
| Extremální Weibullovo | $F(x) = \exp\{-(-x)^\alpha\}, \quad x < 0; \alpha > 0$ $U(t) = -\left(\ln \frac{t}{t-1}\right)^{1/\alpha}, \quad t > 1$ | $-\frac{1}{\alpha}$ |

1.4. Podmínka druhého řádu

Tento odstavec bude zaměřen na charakterizaci konvergence ve vztahu (1.22), resp. (1.23) a bude zavedena tzv. podmínka druhého řádu. Splnění této podmínky bude později nutným požadavkem pro existenci asymptotických rozdělení semiparametrických odhadů parametrů EV rozdělení popsanych v odstavci 2.3. V tomto kontextu hraje podmínka druhého řádu důležitou roli také při použití prahového modelu (viz odstavec 2.1), zejména v kombinaci s některými adaptivními technikami výběru prahové hodnoty, na které je tato práce dále zaměřena.

DEFINICE 1.11. Řekneme, že kvantilová funkce chvostu $U(t)$ (resp. příslušné rozdělení pravděpodobnosti) splňuje podmínku druhého řádu, pokud existuje nějaká kladná funkce $a(t)$ a kladná či záporná funkce $A(t)$ taková, že $\lim_{t \rightarrow \infty} A(t) = 0$ a pro $x > 0$ platí

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx) - U(t)}{a(t)} - D_\xi(x)}{A(t)} = H_{\xi, \rho}(x) := \frac{1}{\rho} \left(\frac{x^{\xi+\rho} - 1}{\xi + \rho} - \frac{x^\xi - 1}{\xi} \right). \quad (1.36)$$

Parametr $\rho \leq 0$ se nazývá *parametr druhého řádu* a funkce $A(t)$ se pak označuje jako tzv. *pomocná funkce druhého řádu*.

V případě, že $\xi = 0$ nebo $\rho = 0$, chápeme funkci $H_{\xi, \rho}(x)$ ve tvaru, který se obdrží jejím limitním přechodem pro $\xi \rightarrow 0$, resp. $\rho \rightarrow 0$, tj. ve tvaru

$$H_{\xi, \rho}(x) = \begin{cases} \frac{1}{\rho} \left(\frac{x^{\xi+\rho} - 1}{\xi + \rho} - \frac{x^\xi - 1}{\xi} \right), & \xi \neq 0 \neq \rho, \\ \frac{1}{\xi} \left(x^\xi \ln x - \frac{x^\xi - 1}{\xi} \right), & \xi \neq 0 = \rho, \\ \frac{1}{\rho} \left(\frac{x^\rho - 1}{\rho} - \ln x \right), & \xi = 0 \neq \rho, \\ \frac{1}{2} \ln^2 x, & \xi = 0 = \rho. \end{cases} \quad (1.37)$$

Zejména je třeba upozornit na skutečnost, že funkce $H_{\xi,\rho}(x)$ není násobkem funkce $D_\xi(x)$. Uvědomme si, že funkce $a(t)$ z podmínky (1.36) může (ale také nemusí) být totožná s funkcí $a(t)$ z podmínky atraktivity (1.23). Lze navíc ukázat (viz např. [38]), že

$$|A(t)| \in \mathcal{R}_\rho.$$

Označí-li se $a_1(t) := a(t)A(t)$, pak se vztah (1.36) může přepsat do ekvivalentního tvaru

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t) - a(t)D_\xi(x)}{a_1(t)} = H_{\xi,\rho}(x). \quad (1.38)$$

Nahradíme nyní funkci $a(t)$ ve vztahu (1.38) za funkci $a(t) + ca_1(t)$ pro nějakou konstantu $c \in \mathbb{R}$. Dostaneme tak

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{U((tx) - U(t) - a(t)(1 + cA(t))D_\xi(x)}{a_1(t)(1 + cA(t))} &= \\ &= \lim_{t \rightarrow \infty} \left[\frac{1}{1 + cA(t)} \frac{U(tx) - U(t) - a(t)D_\xi(x)}{a_1(t)} - \frac{cD_\xi(x)}{1 + cA(t)} \right] = \\ &= H_{\xi,\rho}(x) - cD_\xi(x). \end{aligned} \quad (1.39)$$

Protože je podmínka druhého řádu splněna a $a(t) \simeq a(t) + ca_1(t)$, plyne odtud, že funkce $H_{\xi,\rho}(x)$ je vždy předchozího typu jako v (1.39). Obecně zůstává podmínka druhého řádu v platnosti, nahradíme-li ve vztahu (1.36) funkce $a(t)$, $A(t)$ po řadě funkcemi $a_*(t)$, $A_*(t)$ s vlastností $A_*(t) \simeq A(t)$ a $a_*(t)/a(t) - 1 = o(A(t))$ [38].

DŮSLEDEK 1.12. Jestliže je splněna podmínka (1.36) pro nějaké funkce $a(t)$, $A(t)$ a $\rho \leq 0$, pak existuje kladná funkce $a_0(t)$ a kladná či záporná funkce $A_0(t)$ tak, že platí

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx) - U(t)}{a_0(t)} - D_\xi(x)}{A_0(t)} = \Psi_{\xi,\rho}(x), \quad (1.40)$$

kde

$$\Psi_{\xi,\rho}(x) := \begin{cases} \frac{x^{\xi+\rho}-1}{\xi+\rho}, & \xi + \rho \neq 0, \rho < 0, \\ \ln x, & \xi + \rho = 0, \rho < 0, \\ \frac{x^\xi}{\xi} \ln x, & \rho = 0 \neq \xi, \\ \frac{1}{2} \ln^2 x, & \rho = 0 = \xi, \end{cases}$$

a kde dále

$$a_0(t) := \begin{cases} a(t) \left(1 - \frac{1}{\rho} A(t)\right), & \rho < 0, \\ a(t) \left(1 - \frac{1}{\xi} A(t)\right), & \rho = 0 \neq \xi, \\ a(t), & \rho = 0 = \xi, \end{cases}$$

$$A_0(t) := \begin{cases} \frac{1}{\rho} A(t), & \rho < 0, \\ A(t), & \rho = 0. \end{cases}$$

Hodnoty parametru druhého řádu ρ a tvary pomocných funkcí druhého řádu $A(t)$ pro rozdělení z Fréchetova, Gumbelova a Weibullova oboru atraktivty uvádějí tabulky 1.4, 1.5 a 1.6. V dodatku A je možné nalézt detailní odvození velké části z nich. Platnost podmínky druhého řádu je možné ověřit například pomocí následující postačující podmínky.

VĚTA 1.13. VON MISESOVA PODMÍNKA DRUHÉHO ŘÁDU Necht' $U(t)$ je kvantilová funkce chvostu, $U''(t)$ existuje a

$$A(t) := \frac{tU''(t)}{U'(t)} - \xi + 1. \quad (1.41)$$

Jestliže $A(t)$ nemění znaménko pro dostatečně velké t , $\lim_{t \rightarrow \infty} A(t) = 0$ a $|A(t)| \in \mathcal{R}_\rho$, pro nějaké $\rho \leq 0$, pak pro $x > 0$ platí

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx) - U(t)}{a(t)} - \frac{x^\xi - 1}{\xi}}{A(t)} = H_{\xi, \rho}(x).$$

Důkaz. Bud' $U(t)$ kvantilová funkce chvostu. Protože $U'(t) \in \mathcal{R}_{\xi-1}$, pak zřejmě

$$\ln U'(tx) - \ln U'(t) = \int_1^x \frac{tuU''(tu)}{U'(tu)} \frac{du}{u} \rightarrow \int_1^x (\xi - 1) \frac{du}{u}.$$

Odtud tedy dostáváme

$$\begin{aligned} \frac{\frac{U(tx) - U(t)}{a(t)} - \frac{x^\xi - 1}{\xi}}{A(t)} &= \frac{\int_1^x \left(\frac{U'(ts)}{U'(t)} - s^{\xi-1} \right) ds}{A(t)} = \\ &= \frac{\int_1^x s^{\xi-1} t^{\xi-1} \int_1^s \frac{U''(tu)(tu)^{1-\xi} + (1-\xi)(tu)^{-\xi} U'(tu)}{U'(t)} t du ds}{A(t)} = \\ &= \frac{\int_1^x s^{\xi-1} \int_1^s \left(\frac{tuU''(tu)}{U'(t)} - \xi + 1 \right) \frac{U'(tu)}{U'(t)} u^{-\xi} du ds}{A(t)} = \\ &= \int_1^x s^{\xi-1} \int_1^s \frac{A(tu)}{A(t)} \frac{U'(tu)}{U'(t)} u^{-\xi} du ds. \end{aligned}$$

Protože pro $u > 0$ konvergují členy $A(tu)/A(t)$ a $U'(tu)/U'(t)$ stejnoměrně ke u^ρ a $u^{\xi-1}$ (viz [38], věta B.1.4), lze předchozí rovnost dále zapsat ve tvaru integrálu $\int_1^x s^{\xi-1} \int_1^s u^{\rho-1} du ds$, jehož integrací se zřejmě získá funkce $H_{\xi, \rho}(x)$. Tvrzení je tak dokázáno. \square

Tabulka 1.4: Rozdělení patřící do Fréchetova oboru atraktivitv.

| Rozdělení | Vhodný tvar funkce $a(t)$ | Vhodný tvar funkce $A(t)$ | Parametr druhého řádu |
|--------------------|---|--|---|
| Paretovo | $a(t) = \xi \alpha t^\xi$ | nesplňuje podmínku druhého řádu | $-\infty$ |
| Zobecněné Paretovo | $a(t) = t^\xi$ | $A(t) = [\xi(t^\xi - 1)]^{-1}$ | $-\xi$ |
| Fréchetovo | $a(t) = \begin{cases} \frac{2\xi t^{\xi+1}}{2t+\xi-1}, & \xi \neq 1, \\ \frac{12t^3}{12t^2-1}, & \xi = 1 \end{cases}$ | $A(t) = \begin{cases} \frac{\xi-1}{2t}, & \xi \neq 1, \\ -\frac{1}{6t^2}, & \xi = 1 \end{cases}$ | -1 ($\xi \neq 1$) -2 ($\xi = 1$) |
| Cauchyho | $a(t) = t/\pi$ | $A(t) = \pi^2/(3t^2)$ | -2 |
| Studentovo [66] | tvar funkce není známý | tvar funkce není známý | $-\frac{2}{n}$ |
| Burrovo (typ XII) | $a(t) = \frac{\eta^{1/\tau}}{\lambda^\tau} \frac{t^{1/(\lambda^\tau)+1/\lambda}}{t^{1/\lambda} + \frac{1}{\lambda}}$ | $A(t) = (\lambda t^{1/\lambda})^{-1}(\tau^{-1} - 1)$ | $-\frac{1}{\lambda}$ |
| Burrovo (typ III) | $a(t) = \frac{\lambda^{1/\tau} \eta^{-1/\tau}}{\tau} \frac{t^{1/\tau+1}}{t - \frac{1}{\lambda}(\frac{1}{\tau} - 1)}$ | $A(t) = (\lambda t)^{-1}(\tau^{-1} - 1)$ | -1 |
| Loglogistické | $a(t) = \frac{\alpha}{\beta} t^{1/\beta}$ | $A(t) = \frac{1-\beta}{\beta t}$ | -1 |
| Hallová třída | $a(t) = C\xi t^\xi$ | $A(t) = \frac{\xi+1}{\xi} D t^\rho$ | ρ |

Tabulka 1.5: Rozdělení patřící do Gumbelova oboru atraktivity.

| Rozdělení | Vhodný tvar funkce $a(t)$ | Vhodný tvar funkce $A(t)$ | Parametr druhého řádu |
|----------------|---|--|-----------------------|
| Exponenciální | $a(t) = \lambda^{-1}$ | nesplňuje podmínku druhého řádu | $-\infty$ |
| Weibullovo | $a(t) = \frac{\lambda}{\alpha} (\ln t)^{(1-\alpha)/\alpha}$ | $A(t) = \frac{1-\alpha}{\alpha \ln t}$ | 0 |
| Normální [38] | tvar funkce není známý | tvar funkce není známý | 0 |
| Gama [38] | tvar funkce není známý | tvar funkce není známý | 0 |
| Logistické | $a(t) = \frac{\sigma t}{t-1}$ nebo $a(t) = \sigma t \ln \left(\frac{t}{t-1} \right)$ | $A(t) = -\frac{1}{t-1}$ | -1 |
| Gumbelovo [55] | $a(t) = 1$ | $A(t) = \frac{1}{2t}$ | -1 |

Tabulka 1.6: Rozdělení patřící do Weibullova oboru atraktivit.

| Rozdělení | Vhodný tvar funkce $a(t)$ | Vhodný tvar funkce $A(t)$ | Parametr druhého řádu |
|-----------------------|---|---|-----------------------|
| Rovnoměrné | $a(t) = \frac{b-a}{t}$ | nesplňuje podmínku druhého řádu | $-\infty$ |
| Reverzní Burrovo | $a(t) = \frac{\eta^{1/\tau}}{\lambda^\tau} t^{-1}/(\lambda^\tau)$ | $A(t) = \frac{\tau}{1+\tau} t^{-1/\lambda}$ | $-\frac{1}{\lambda}$ |
| Extremální Weibullovo | $a(t) = \frac{1}{\alpha} t^{-1/\alpha}$ | $A(t) = \frac{1+\alpha}{2\alpha t}$ | -1 |

Odhady parametrů rozdělení extrémních hodnot

V této kapitole budou popsány vybrané metody odhadů parametrů EV rozdělení. V první části nejdříve představíme základní přístupy vyhodnocení dat, které vedou na použití GEV rozdělení, a sice tzv. model blokových maxim a prahový model. Následně se zaměříme přímo na techniky odhadu. Ve druhém odstavci budou stručně shrnuty základy často užívané parametrické metody odhadu, metody maximální věrohodnosti. Zde bude diskutováno její konkrétní použití pro odhady parametrických funkcí za předpokladu modelu blokových maxim i prahového modelu. V poslední době se pozornost zaměřila především na neparametrické přístupy odhadu parametrů EV rozdělení. Řada autorů se zabývala odhady parametrů metodou pravděpodobnostně vážených momentů [29, 30, 44, 135] nebo metodou L-momentů [12, 74]. V této kapitole budou rozebrány semiparametrické přístupy odhadu EV indexu, a to Hillův a momentový odhad v prahovém modelu. Tyto semiparametrické odhady jsou závislé na vhodné volbě prahové hodnoty a vzhledem k tomuto bude také provedena jejich diskuze. Technikám adaptivní volby prahové hodnoty v kombinaci s uvedenými odhady pak bude později věnována kapitola 3. Významná třída těchto adaptivních přístupů je založena na metodě bootstrap, kterou v závěru následující kapitoly v krátkosti představíme.

2.1. Základní modely pro odhady parametrů a parametrických funkcí EV rozdělení

V tomto odstavci budou popsány dva základní přístupy pro analýzu a modelování extrémních hodnot nezávislých pozorování. V aplikačních úlohách, kdy je potřeba učinit závěry o rozdělení extrémních hodnot, je nutné z dané řady pozorování nejprve vybrat dostatečně velké pozorované hodnoty, které mohou pocházet z nějakého EV rozdělení. Řada autorů vyzdvihuje fakt [34, 22], že v praktických situacích jsou často extrémní výchyly při měření náhodných veličin způsobeny jinými mechanismy než „běžná“ pozorování. Jako jedno z možných užití EV teorie se tak nabízí nejenom často vyžadovaný odhad frekvence výskytu extrémních jevů, ale také například možnost automatické validace dat, kdy je potřeba očistit pozorování sledovaných veličin od vnějších vlivů. Právě ve způsobu výběru významných pozorování se liší modely popsané níže. Nejprve bude popsán tzv. *model blokových maxim*, který je založen na GEV rozdělení. Tento model má některá úskalí, která se snaží překonat *prahový model* založený na zobecněném Paretově rozdělení.

Model blokových maxim

V první kapitole bylo ukázáno, že třída EV rozdělení je dána distribuční funkcí G_ξ GEV rozdělení ve tvaru (1.11). Doposud byla pozornost soustředěna na popis vlastností EV rozdělení vzhledem k EV indexu ξ a GEV rozdělení bylo chápáno jako jednoparametrická třída rozdělení s parametrem tvaru ξ . Nyní uvažujme obecněji GEV rozdělení včetně parametru polohy $\mu \in \mathbb{R}$ a parametru měřítka $\sigma > 0$ s distribuční funkcí $G(x)$ v následujícím tvaru

$$G(x) = \begin{cases} \exp \left\{ - \left[1 + \xi \left(\frac{x-\mu}{\sigma} \right) \right]^{-1/\xi} \right\}, & \text{pro } \xi \neq 0 \text{ a } 1 + \xi \left(\frac{x-\mu}{\sigma} \right) > 0, \\ \exp \left\{ - \exp \left[- \left(\frac{x-\mu}{\sigma} \right) \right] \right\}, & \text{pro } \xi = 0, \end{cases} \quad (2.1)$$

kde EV index $\xi \in \mathbb{R}$ je parametr tvaru. Příležitostně budeme psát $\text{GEV}(\mu, \sigma, \xi)$, abychom explicitně vyjádřili, že máme na mysli GEV rozdělení s uvedenými parametry polohy, měřítka a tvaru.

Metoda blokových maxim byla navržena Gumbem a poprvé byla použita v práci [67]. Uvažujme, že je dán náhodný výběr X_1, \dots, X_k rozsahu $k = mn$ s neznámou distribuční funkcí $F \in \mathcal{D}(G_\xi)$. Princip metody je následující. Daný náhodný výběr se nejprve rozdělí na m bloků délky n , z nichž se posléze vyberou vždy jen ta nejvyšší pozorování. Označme $M_{n,i}, i = 1, \dots, m$ bloková maxima, tedy

$$M_{n,i} = \max\{X_{(i-1)n+1}, \dots, X_{in}\}, \quad i = 1, \dots, m.$$

Jak udává věta 1.6, pro každé $i = 1, \dots, m$ platí

$$\lim_{n \rightarrow \infty} P \left(\frac{M_{n,i} - b_n}{a_n} \leq x \right) = G(x), \quad (2.2)$$

kde $G(x)$ je distribuční funkce nějakého $\text{GEV}(\mu, \sigma, \xi)$ rozdělení. Pro konečné a dostatečně velké n můžeme platnost vztahu (2.2) chápat asymptoticky, tedy pro $i = 1, \dots, m$ lze použít aproximaci

$$P(M_{n,i} \leq x) \approx G((x - b_n)/a_n) = G^*(x),$$

kde G^* je distribuční funkce $\text{GEV}(\mu^*, \sigma^*, \xi^*)$ rozdělení s parametry $\mu^* = a_n \mu + b_n$, $\sigma^* = a_n \sigma$ a $\xi^* = \xi$. Je tak vidět, že se funkce G a G^* liší pouze volbou parametru polohy a měřítka. Tento fakt nepředstavuje pro analýzu extrémních hodnot žádné omezení, protože všechny parametry musí být odhadnuty v každém případě.

Nyní se zaměříme na odhady vybraných parametrických funkcí. Vzhledem k aplikačním úlohám, kterým se tato práce věnuje v kapitole 4, bude věnována pozornost zejména odhadům vybraných vysokých kvantilů. Typicky se jedná o kvantily odpovídající nějaké hodnotě $r > 1$ tak, aby pravděpodobnost překročení byla $1/r$. V praktických situacích bývá obvyklá volba $r \gg k$, kde k je rozsah daného výběru, a zřejmě zde EV teorie představuje vhodný nástroj k odhadům těchto kvantilů. V aplikačních oblastech se hodnota r často nazývá *perioda návratu* a zmíněný kvantil z_r je pak označován jako tzv. *návratová úroveň*. Návratová úroveň z_r je tedy taková hodnota sledované náhodné veličiny, která je průměrně překročena jednou za r pozorování. Jedná se vlastně o $(1 - r^{-1})$ kvantil modelového rozdělení, v případě metody blokových maxim tedy GEV rozdělení.

Návratová úroveň je v následujícím tvaru

$$z_r = \begin{cases} \mu - \frac{\sigma}{\xi} \{1 - [-\ln(1 - r^{-1})]\}, & \text{pro } \xi \neq 0, \\ \mu - \sigma \ln[-\ln(1 - r^{-1})], & \text{pro } \xi = 0, \end{cases} \quad (2.3)$$

přičemž její odhad lze získat nahrazením parametrů μ, σ, ξ ve vztahu (2.3) vhodnými odhady $\hat{\mu}, \hat{\sigma}, \hat{\xi}$. Odhady parametrů a parametrických funkcí pomocí metody maximální věrohodnosti budou diskutovány v odstavci 2.2. Někdy se v praxi vyskytuje požadavek určení N -leté návratové úrovně, tedy hodnoty z_N , která je průměrně překročena jednou každých N let. Je-li pozorovaná řada rozdělena na bloky o délce odpovídající trvání jednoho roku, pak N -letou úroveň obdržíme jednoduše položením $r = N$.

Nesnáze metody blokových maxim jsou zřejmé: při konstantním rozsahu $k = mn$ původního náhodného výběru způsobuje velká volba délky bloku n malý počet bloků m a tím i malý rozsah náhodného výběru $M_{n,1}, \dots, M_{n,m}$. Odhady parametrů GEV rozdělení tak bývají zatíženy značnou variabilitou. Naopak při velkém počtu bloků je jejich délka malá, což může vést k vychýlení odhadů vlivem nedostatečné aproximace asymptotickým rozdělením. Metodě blokových maxim je také často vytýkáno (viz např. [95, 97]), že zanedbává případná jiná velká pozorování, která ovšem nedosahují hodnot příslušných blokových maxim. Původní náhodný výběr je tak značně redukován, což vede k významné ztrátě obsažené informace.

Je však obecně přijímáno [80], že na rozdíl od prahového modelu diskutovaného níže, je tento přístup robustnější vzhledem k porušení např. předpokladu nezávislosti původní řady X_1, \dots, X_{mn} . Modely odhadu parametrů pro stacionární řady budou detailně popsány v kapitole 5 a zde bude nadále uvažován předpoklad nezávislosti.

Prahový model

Nechť X je náhodná veličina s distribuční funkcí $F \in \mathcal{D}(G_\xi)$. Z (1.24) pro $x > 0$ splňující $1 + \xi x > 0$ plyne rovnost

$$\lim_{u \uparrow x^*} \frac{1 - F(u + x\varphi(u))}{1 - F(u)} = \lim_{u \uparrow x^*} P\left(X - u > x\varphi(u) \mid X > u\right) = (1 + \xi x)^{-1/\xi}, \quad (2.4)$$

kde φ je nějaká kladná funkce pro $x > 0$. Označme nyní $\sigma_u := \varphi(u)$ parametr závislý na $u \in \mathbb{R}$ a zaveďme náhodnou veličinu $Y = X - u$. Hodnotu u budeme dále nazývat *prahovou hodnotou* nebo také zkráceně *prahem*. Ze vztahu (2.4) vyplývá

$$\lim_{u \uparrow x^*} P(Y > x \mid X > u) = \left(1 + \xi \frac{x}{\sigma_u}\right)^{-1/\xi} =: 1 - H(x), \quad (2.5)$$

kde $H(x)$ je distribuční funkce zobecněného Paretova rozdělení (zkráceně GP rozdělení z angl. Generalized Pareto) s parametrem tvaru $\xi \in \mathbb{R}$ a parametrem měřítka $\sigma_u > 0$ definovaná pro $\{x : x > 0, 1 + \xi x/\sigma_u > 0\}$. Pro případ $\xi = 0$ se opět $H(x)$ dostane limitním přechodem pro $\xi \rightarrow 0$ ve tvaru $1 - \exp(-x/\sigma_u)$.

Tedy pro dostatečně velkou prahovou hodnotu u lze podmíněné rozdělení náhodné veličiny $Y = X - u$ za podmínky $X > u$ aproximovat GP rozdělením, přičemž parametr tvaru $\xi \in \mathbb{R}$ GP rozdělení odpovídá parametru tvaru $\xi \in \mathbb{R}$ příslušného limitního GEV rozdělení výběrových maxim. Pro parametr měřítka je možné odvodit (viz [18]) rovnost

$$\sigma_u = \sigma + \xi(u - \mu), \quad (2.6)$$

kde μ, σ jsou parametry polohy a měřítka příslušného GEV rozdělení výběrových maxim.

Zabývejme se opět odhady návratových úrovní z_r . Při pevném, dostatečně velkém u dostáváme úpravou vztahu (2.5) pro $x > u$ aproximaci

$$P(X > x) \approx P(X > u) (1 - H(x - u)). \quad (2.7)$$

Pro jednoduchost označme $\lambda_u := P(X > u)$. Návratovou úroveň můžeme vyjádřit jako kvantil rozdělení (2.7) ve tvaru

$$z_r = \begin{cases} u + \frac{\sigma_u}{\xi} \left[(\lambda_u r)^\xi - 1 \right], & \text{pro } \xi \neq 0, \\ u + \sigma_u \ln(\lambda_u r), & \text{pro } \xi = 0. \end{cases} \quad (2.8)$$

V případě, že hledaná návratová úroveň je N -letá úroveň, položíme $r = Nn_y$, kde n_y je průměrný počet pozorování za rok. Odhad návratové úrovně získáme nahrazením parametrů ξ, σ_u, λ_u jejich vhodnými odhady $\hat{\xi}, \hat{\sigma}_u, \hat{\lambda}_u$. Odhady parametrů prahového modelu pomocí metody maximální věrohodnosti budou odvozeny v odstavci 2.2. V poslední době jsou rozvíjeny semiparametrické metody odhadu, které jsou vhodné při použití s některými adaptivními metodami výběru prahu. Tyto budou probrány v části 2.3, adaptivní volby prahové hodnoty pak v odstavcích 3.2 a 3.3.

Prahový model byl odvozen Pickandsem (1975, [109]) a v současnosti je stále více používaným modelem pro odhady parametrů GEV rozdělení (viz práce [12, 19, 25, 50, 89, 96, 114]). V literatuře je tento přístup označován nejběžněji jako Peaks-Over-Threshold (zkráceně POT). Oproti metodě blokových maxim neopomíjí ostatní extrémní pozorování, která nedosahují velikosti blokových maxim, a zpravidla tak dovoluje zahrnout do analýzy soubor většího rozsahu. Na druhou stranu vhodná volba prahu je kritickou částí celého modelu. Je-li práh u zvolen příliš nízko, může dojít ke špatné aproximaci limitního vztahu (2.5) a hrozí, že parametry modelu budou odhadnuty se značným vychýlením. Zvolíme-li práh u příliš vysoko, dojde jen k malému počtu překročení prahu a odhady parametrů budou zatíženy velkou variabilitou. Podobně je volba prahu kritická vzhledem k odhadům návratových úrovní (2.8) a tyto mohou být v některých situacích značně vychýlené. Proto tato skutečnost bude také předmětem pozdějších úvah, zejména dále v odstavcích věnovaných konkrétním aplikacím.

V praktických situacích je třeba volit práh co nejnižší tak, aby byla zaručena dostatečná aproximace pomocí GP rozdělení. Volba optimálního prahu, vyvažujícího vychýlení a variabilitu, však stále patří k nedorozřešeným otázkám popisované problematiky. Výběru prahové hodnoty je věnována celá kapitola 3, která shrnuje jak běžně používané metody, tak i některé inovativní přístupy rozvinuté v poslední době. Představené metody jsou následně aplikovány pro analýzu extrémních srážkových úhrnů v odstavci 4, přičemž pro metody výběru prahu je zde provedena detailní simulační studie.

2.2. Odhady metodou maximální věrohodnosti

V následujícím odstavci připomeneme některé základní pojmy a vlastnosti odhadů získaných metodou maximální věrohodnosti, přičemž dále budou rozebrány odhady parametrů v modelu blokových maxim i prahovém modelu. Odstavec je zpracován podle publikací [14, 31, 94]. Vzhledem k rozsahu předkládané práce neuvádíme důkazy následujících tvrzení a čtenáře odkazujeme na citovanou literaturu.

Uvažujme náhodný výběr X_1, \dots, X_n z rozdělení s hustotou $f(x, \boldsymbol{\theta})$ vzhledem k nějaké σ -konečné míře μ , kde $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)^T$ je vektor parametrů daného rozdělení. Označíme-li $\mathbf{X} = (X_1, \dots, X_n)^T$ náhodný vektor s hustotou $f(\mathbf{x}, \boldsymbol{\theta})$, pak zřejmě tato hustota lze vyjádřit ve tvaru

$$f(\mathbf{x}, \boldsymbol{\theta}) = \prod_{i=1}^n f(x_i, \boldsymbol{\theta}). \quad (2.9)$$

Metoda maximální věrohodnosti (MV metoda) je založena na regulárním systému hustot, který nejprve zavedeme v následující definici.

DEFINICE 2.1. Nechť náhodný vektor $\mathbf{X} = (X_1, \dots, X_n)^T$ má hustotu $f(\mathbf{x}, \boldsymbol{\theta})$ vzhledem k nějaké σ -konečné míře μ . Předpokládejme, že platí:

- (i) $\boldsymbol{\theta} \in \Theta$, kde Θ je neprázdná otevřená množina v \mathbb{R}^m .
- (ii) Množina $M = \{\mathbf{x} : f(\mathbf{x}, \boldsymbol{\theta}) > 0\}$ nezávisí na $\boldsymbol{\theta}$.
- (iii) Pro skoro všechna $\mathbf{x} \in M$ vzhledem k μ a pro všechna $i = 1, \dots, m$ existují parciální derivace $f'_i(\mathbf{x}, \boldsymbol{\theta}) := \frac{\partial f(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_i}$.
- (iv) Pro každé $i = 1, \dots, m$ a pro všechna $\boldsymbol{\theta} \in \Theta$ platí $\int_M f'_i(\mathbf{x}, \boldsymbol{\theta}) d\mu(\mathbf{x}) = 0$.
- (v) Pro každou dvojici (i, j) existuje konečný integrál

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

- (vi) 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* (zkráceně FIM). Uvažujeme-li vektor $\boldsymbol{\theta}$ jednorozměrný, nazývá se FIM *Fisherova míra informace* (zkráceně FMI). V případě, že budeme uvažovat jednorozměrný vektor $\mathbf{X} = X$, budeme FIM značit jednoduše $\mathbf{J}(\boldsymbol{\theta})$.

Lze navíc ukázat (viz např. [14]), že je-li X_1, \dots, X_n náhodný výběr z rozdělení s hustotou $f(x, \boldsymbol{\theta})$ a systém $\{f(x, \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$ je regulární s FIM $\mathbf{J}(\boldsymbol{\theta})$, pak hustota $f(\mathbf{x}, \boldsymbol{\theta})$ náhodného vektoru $\mathbf{X} = (X_1, \dots, X_n)^T$ také tvoří regulární systém hustot s FIM $\mathbf{J}_n(\boldsymbol{\theta})$, přičemž

$$\mathbf{J}_n(\boldsymbol{\theta}) = n\mathbf{J}(\boldsymbol{\theta}).$$

Pokud dále existují druhé derivace $f''_{ij}(\mathbf{x}, \boldsymbol{\theta})$ a $\int_M f''_{ij}(\mathbf{x}, \boldsymbol{\theta}) d\mu(\mathbf{x}) = 0$, pak pro každé $i, j = 1, \dots, m$ platí

$$J_{ij}(\boldsymbol{\theta}) = - \int_M \frac{\partial^2 \ln f(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} f(\mathbf{x}, \boldsymbol{\theta}) d\mu(\mathbf{x}). \quad (2.10)$$

FIM $\mathbf{J}_n(\boldsymbol{\theta})$ s prvky (2.10) se nazývá *očekávaná FIM*. Matice $\tilde{\mathbf{J}}_n(\boldsymbol{\theta}) = \left(\tilde{J}_{ij}(\boldsymbol{\theta}) \right)_{i,j=1}^m$ s prvky

$$\tilde{J}_{ij}(\boldsymbol{\theta}) := - \frac{\partial^2 \ln f(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \quad (2.11)$$

se označuje jako *výběrová FIM*.

DEFINICE 2.2. Mějme náhodný výběr X_1, \dots, X_n z rozdělení s hustotou $f(x, \boldsymbol{\theta})$. Funkce

$$L(\boldsymbol{\theta}) := f(\mathbf{x}, \boldsymbol{\theta}) = \prod_{i=1}^n f(x_i, \boldsymbol{\theta}) \quad (2.12)$$

pro pevně zvolené $\mathbf{x} \in \mathbb{R}^n$ se nazývá *věrohodnostní funkce* a funkce

$$l(\boldsymbol{\theta}) := \ln L(\boldsymbol{\theta}) = \sum_{i=1}^n \ln f(x_i, \boldsymbol{\theta}) \quad (2.13)$$

je *logaritmická věrohodnostní funkce*. Statistika $\widehat{\boldsymbol{\theta}}$, pro kterou platí

$$\widehat{\boldsymbol{\theta}} \in \operatorname{argmax}_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta}),$$

se nazývá *maximálně věrohodný odhad* (zkráceně MV odhad) parametru $\boldsymbol{\theta}$.

Zřejmě lze MV odhad získat i maximalizací výrazu (2.13), neboť logaritmus jako rostoucí funkce zachovává uspořádání. Metodami matematické analýzy lze MV odhad často najít jako řešení tzv. *systému (logaritmických) věrohodnostních rovnic* ve tvaru

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

Ve většině případů je však nutné hledat MV odhady pomocí numerických metod, neboť předešlá soustava nebývá zpravidla řešitelná analyticky. V situacích, kdy namísto parametru uvažujeme nějakou parametrickou funkci $g(\boldsymbol{\theta})$, je možné ukázat, že je-li $\widehat{\boldsymbol{\theta}}$ MV odhad parametru $\boldsymbol{\theta}$, pak $g(\widehat{\boldsymbol{\theta}})$ je MV odhadem parametrické funkce $g(\boldsymbol{\theta})$ [31].

Ze vztahů (2.10) a (2.11) je vidět, že očekávanou FIM $\mathbf{J}_n(\boldsymbol{\theta})$ je možné získat pomocí logaritmické věrohodnostní funkce následujícím způsobem

$$J_{ij}(\boldsymbol{\theta}) = \mathbb{E} \left(\widetilde{J}_{ij}(\boldsymbol{\theta}) \right) = -\mathbb{E} \left(\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right). \quad (2.14)$$

VĚTA 2.3. [94] Nechť systém hustot $\{f(x, \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$ je regulární a má FIM $\mathbf{J}(\boldsymbol{\theta})$. Nechť platí následující předpoklady:

- (a) Nechť $\Theta \subset \mathbb{R}^m$ je parametrický prostor, který obsahuje takový neprázdný otevřený interval ω , že skutečná hodnota parametru $\boldsymbol{\theta}_0$ patří do ω .
- (b) Nechť $\mathbf{X} = (X_1, \dots, X_n)^T$, kde X_i jsou nezávislé, stejně rozdělené náhodné veličiny s hustotou $f(x, \boldsymbol{\theta})$ vzhledem k nějaké σ -konečné míře μ .
- (c) Nechť $M = \{x : f(x, \boldsymbol{\theta}) > 0\}$ nezávisí na $\boldsymbol{\theta}$.
- (d) Nechť $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$. Pak $f(x, \boldsymbol{\theta}_1) = f(x, \boldsymbol{\theta}_2)$ vzhledem k μ právě tehdy, je-li $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$.
- (e) Derivace $\frac{\partial^3 f(x, \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j \partial \theta_k}$ existuje pro skoro všechna x , pro všechna $\boldsymbol{\theta} \in \omega$ a pro všechna $i, j, k = 1, \dots, m$.

(f) Pro všechna $\boldsymbol{\theta} \in \omega$ platí

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

(g) Pro všechna $i, j, k, = 1, \dots, m$ existují funkce $M_{ijk}(x) \geq 0$ tak, že $EM_{ijk}(X) < \infty$ 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 $\boldsymbol{\theta} \in \omega$ a skoro všechna $x \in M$.

Pak platí následující tvrzení.

(i) Jestliže $n \rightarrow \infty$, pak ke každému $\varepsilon > 0$ existuje s pravděpodobností blížící se jedné takové řešení $\widehat{\boldsymbol{\theta}}_n$ systému věrohodnostních rovnic, že $|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0| < \varepsilon$.

(ii) Označme

$$\mathbf{U}(\boldsymbol{\theta}) = \left(\frac{\partial l(\boldsymbol{\theta})}{\partial \theta_1}, \dots, \frac{\partial l(\boldsymbol{\theta})}{\partial \theta_m} \right)^T. \quad (2.15)$$

Pak pro $n \rightarrow \infty$ platí

$$\frac{1}{\sqrt{n}} \mathbf{U}(\boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{J}(\boldsymbol{\theta}_0)). \quad (2.16)$$

(iii) 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

$$\sqrt{n} (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, [\mathbf{J}(\boldsymbol{\theta}_0)]^{-1}). \quad (2.17)$$

Veličina $\mathbf{U}(\boldsymbol{\theta})$ definovaná v předchozím tvrzení se nazývá *skórový vektor příslušný hustotě $f(x, \boldsymbol{\theta})$* .

Dále se zaměříme na MV odhady parametrů GEV a GP rozdělení. V konkrétních situacích je často analyticky nemožné určit rozdělení daného odhadu pro konečný rozsah výběru n a obvykle se užívají známé asymptotické vlastnosti odhadů. Předchozí tvrzení tak poskytuje prostředek pro odhad vlastností MV odhadů parametrů či parametrických funkcí na základě asymptotické normality. Výhodou MV metody pak je, že nejsou kladeny žádné další požadavky na splnění podmínky druhého řádu.

MV odhady parametrů GEV rozdělení (resp. GP rozdělení) lze však získat jen s jistými omezeními, protože třída GEV rozdělení není obecně regulární a jsou tak porušeny základní předpoklady věty 2.3. Jak je ukázáno v práci [120], konvergence MV odhadů závisí na hodnotě EV indexu. Můžeme tak v zásadě rozlišit čtyři případy $\xi > -1/2$, $\xi = -1/2$, $-1 < \xi < -1/2$ a $\xi \leq -1$. Pokud $\xi > -1/2$, mají odhady získané metodou maximální věrohodnosti obvyklé asymptotické vlastnosti, tj. asymptotické rozdělení odhadů je normální. Pro $\xi = -1/2$ byly MV odhady odvozeny v [119] pro semiparametrický model. V případě, že $-1 < \xi < -1/2$, lze odhady získat běžným způsobem, ovšem tyto již nemají obvyklé asymptotické vlastnosti [107]; přesto však, jak je ukázáno v [137], pro $\xi > -1$ představují odhady získané metodou maximální věrohodnosti konzistentní odhady. Nakonec pro $\xi \leq -1$, kdy mají rozdělení velmi krátké pravé chvosty, je málo pravděpodobné, že budou MV odhady dosaženy s dostatečnou přesností.

MV odhady parametrů v modelu blokových maxim

Metoda blokových maxim je založena na tříparametrickém GEV rozdělení (2.1). Dle (2.12) lze tak věrohodnostní funkci odvodit ve tvaru

$$L(\mu, \sigma, \xi) = \begin{cases} \sigma^{-n} \exp \left\{ - \sum_{i=1}^n \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right]^{-1/\xi} \right\} \times \\ \quad \times \prod_{i=1}^n \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right]^{-1/\xi}, & \text{pro } \xi \neq 0, \\ \exp \left\{ - \sum_{i=1}^n \exp \left[- \left(\frac{x_i - \mu}{\sigma} \right) \right] \right\}, & \text{pro } \xi = 0, \end{cases} \quad (2.18)$$

za předpokladu, že $1 + \xi(x_i - \mu)/\sigma > 0$ pro každé $i = 1, \dots, n$. Obdobně, za stejných podmínek, je logaritmická věrohodnostní funkce ve tvaru

$$l(\mu, \sigma, \xi) = \begin{cases} -n \ln \sigma - \left(1 + \frac{1}{\xi} \right) \sum_{i=1}^n \ln \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right] - \\ \quad - \sum_{i=1}^n \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right]^{-1/\xi}, & \text{pro } \xi \neq 0, \\ -n \ln \sigma - \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right) - \sum_{i=1}^n \exp \left[- \left(\frac{x_i - \mu}{\sigma} \right) \right], & \text{pro } \xi = 0. \end{cases} \quad (2.19)$$

MV odhady $\hat{\mu}, \hat{\sigma}, \hat{\xi}$ parametrů μ, σ, ξ se dostanou maximalizací výrazu (2.19), resp. jako řešení příslušné soustavy věrohodnostních rovnic. Tuto soustavu je nutné řešit numericky, např. lze využít optimalizační proceduru založenou na Nelderově-Meadově simplexovém algoritmu [90, 106], který je implementován v řadě softwarů. Pro nalezení MV odhadů pomocí numerických metod bývá nutné stanovit vhodné počáteční hodnoty - tyto lze pro parametry GEV rozdělení získat např. metodou momentů [14].

Označme $\mathbf{V} = \mathbf{V}(\hat{\mu}, \hat{\sigma}, \hat{\xi})$ varianční matici MV odhadů $\hat{\mu}, \hat{\sigma}, \hat{\xi}$. Ze vztahu (2.17) plyne, že odhad varianční matice lze získat pomocí výběrové FIM jako $\mathbf{V}(\hat{\mu}, \hat{\sigma}, \hat{\xi}) \approx \tilde{\mathbf{J}}_n^{-1}(\hat{\mu}, \hat{\sigma}, \hat{\xi})$. Detailní odvození prvků matice $\tilde{\mathbf{J}}_n^{-1}(\mu, \sigma, \xi)$ je provedeno v [57].

Dále můžeme přistoupit k MV odhadům parametrických funkcí. Zde se opět zaměříme na odhad návratové úrovně \hat{z}_r (2.3), jejíž MV odhad získáme nahrazením parametrů μ, σ, ξ jejich MV odhady. Na základě vlastností MV odhadů je možné získat odhad rozptylu statistiky \hat{z}_r pomocí tzv. *delta metody* [31, 33] ve tvaru

$$\text{var}(\hat{z}_r) \approx \nabla z_r^T \mathbf{V} \nabla z_r,$$

kde pro $\xi \neq 0$

$$\nabla z_r^T = \left(\frac{\partial z_r}{\partial \mu}, \frac{\partial z_r}{\partial \sigma}, \frac{\partial z_r}{\partial \xi} \right) = \left(1, -\frac{1 - y_r^{-\xi}}{\xi}, \frac{\sigma}{\xi^2} (1 - y_r^{-\xi}) - \frac{\sigma}{\xi} y_r^{-\xi} \ln y_r \right),$$

a kde bylo zavedeno označení $y_r := -\ln(1 - r^{-1})$, přičemž parametry μ, σ, ξ jsou nahrazeny jejich MV odhady $\hat{\mu}, \hat{\sigma}, \hat{\xi}$. Na základě asymptotické normality MV odhadů plynoucí ze vztahu (2.17) je možné určit $100(1 - \alpha)\%$ asymptotický (oboustranný) interval spolehlivosti pro z_r jako dvojici statistik

$$\langle \hat{z}_r - u_{1-\alpha/2} \sqrt{\text{var}(\hat{z}_r)}, \hat{z}_r + u_{1-\alpha/2} \sqrt{\text{var}(\hat{z}_r)} \rangle, \quad (2.20)$$

kde $u_{1-\alpha/2}$ je $(1 - \alpha/2)$ kvantil standardizovaného normálního rozdělení.

MV odhady parametrů v prahovém modelu

Zaměříme pozornost na MV odhady GP rozdělení, na kterém je prahový model postaven. Věrohodnostní funkce GP rozdělení je ve tvaru

$$L(\sigma_u, \xi) = \begin{cases} \sigma_u^{-n} \prod_{i=1}^n \left(1 + \xi \frac{x_i}{\sigma_u}\right)^{-1/\xi-1}, & \text{pro } \xi \neq 0, \\ \sigma_u^{-n} \exp \left\{ -\sigma_u^{-1} \sum_{i=1}^n x_i \right\}, & \text{pro } \xi = 0. \end{cases} \quad (2.21)$$

Logaritmická věrohodnostní funkce GP rozdělení je dána jako

$$l(\sigma_u, \xi) = \begin{cases} -n \ln \sigma_u - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^n \ln \left(1 + \xi \frac{x_i}{\sigma_u}\right), & \text{pro } \xi \neq 0, \\ -n \ln \sigma_u - \frac{1}{\sigma_u} \sum_{i=1}^n x_i, & \text{pro } \xi = 0. \end{cases} \quad (2.22)$$

Maximalizací výrazu (2.22) obdržíme MV odhady příslušných parametrů. Protože systém věrohodnostních rovnic opět nelze řešit analyticky, je nutné použít vhodné numerické metody.

Hodnota $\lambda_u := P(X > u)$ se pro pevně stanovený práh u odhadne jako relativní četnost pozorování nad prahem u . Protože počet překročení prahu je náhodná veličina s binomickým rozdělením $\text{Bi}(n, \lambda_u)$, je takto definovaný odhad $\hat{\lambda}_u$ také MV odhadem. Z vlastností binomického rozdělení lze tedy odhad rozptylu $\hat{\lambda}_u$ určit jako $\text{var}(\hat{\lambda}_u) \approx \hat{\lambda}_u(1 - \hat{\lambda}_u)/n$.

Bud' $\mathbf{V}(\hat{\sigma}_u, \hat{\xi})$ varianční matice odhadů $\hat{\sigma}_u, \hat{\xi}$. Pak podle (2.17) je možné získat odhad varianční matice pomocí výběrové FIM, $\mathbf{V}(\hat{\sigma}_u, \hat{\xi}) \approx \tilde{\mathbf{J}}_n^{-1}(\hat{\sigma}_u, \hat{\xi})$. Odvození varianční matice je možné nalézt např. v práci [57]. Za předpokladu, že doba překročení a míra překročení $y = x - u$ jsou nezávislé (tj. λ_u nezávisí na parametrech σ_u, ξ), lze odhadnout varianční matici odhadů $\hat{\lambda}_u, \hat{\sigma}_u, \hat{\xi}$ pomocí výše uvedeného jako

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

Odhad rozptylu statistiky z_r získáme opět pomocí delta metody, tudíž ve tvaru

$$\text{var}(\hat{z}_r) \approx \nabla z_r^T \mathbf{V} \nabla z_r,$$

kde pro $\xi \neq 0$

$$\begin{aligned} \nabla z_r^T &= \left(\frac{\partial z_r}{\partial \lambda_u}, \frac{\partial z_r}{\partial \sigma_u}, \frac{\partial z_r}{\partial \xi} \right) = \\ &= \left(\sigma_u r^\xi \lambda_u^{\xi-1}, \frac{1}{\xi} [(r\lambda_u)^\xi - 1], -\frac{\sigma_u}{\xi^2} [(r\lambda_u)^\xi - 1] + \frac{\sigma_u}{\xi} (r\lambda_u)^\xi \ln(r\lambda_u) \right), \end{aligned}$$

přičemž parametry λ_u, σ_u, ξ jsou nahrazeny jejich MV odhady $\hat{\lambda}_u, \hat{\sigma}_u, \hat{\xi}$. Asymptotický $100(1 - \alpha)\%$ interval spolehlivosti obdržíme obdobně jako ve vztahu (2.20).

Asymptotické testy s rušivými parametry

Na závěr odstavce 2.2 ještě doplníme jedno důležité tvrzení pro testování hypotéz s rušivými parametry, které bude použito v dalším textu. Nechť je parametr $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)^T$ rozdělený na dvě složky $\boldsymbol{\theta} = (\boldsymbol{\tau}^T, \boldsymbol{\psi}^T)^T$, kde $\boldsymbol{\tau} = (\theta_1, \dots, \theta_k)^T$ a $\boldsymbol{\psi} = (\theta_{k+1}, \dots, \theta_m)^T$ pro nějaké k ($1 \leq k < m$). Cílem je stanovení vhodného kritéria pro testování nulové hypotézy ve tvaru $H : \boldsymbol{\tau} = \boldsymbol{\tau}_0$ proti alternativě $A : \boldsymbol{\tau} \neq \boldsymbol{\tau}_0$. Parametry obsažené v $\boldsymbol{\psi}$ jsou tak sice potřebné k popisu pravděpodobnostního modelu, ale testovaná hypotéza se týká jen parametru $\boldsymbol{\tau}$. Běžně se proto $\boldsymbol{\psi}$ nazývá *rušivý parametr* a $\boldsymbol{\tau}$ *cílový parametr*.

Nechť je dán skórový vektor $\mathbf{U}(\boldsymbol{\theta}) = (U_1, \dots, U_m)^T$ definovaný vzorcem (2.15). Dále označme

$$\begin{aligned} \mathbf{U}_1(\boldsymbol{\theta}) &= (U_1, \dots, U_k)^T, \\ \mathbf{U}_2(\boldsymbol{\theta}) &= (U_{k+1}, \dots, U_m)^T, \end{aligned}$$

a Fisherovu informační matici $\mathbf{J} = \mathbf{J}(\boldsymbol{\theta})$ danou definicí 2.1 rozdělme podle složek parametru $\boldsymbol{\theta}$ následovně

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_{11} & \mathbf{J}_{12} \\ \mathbf{J}_{21} & \mathbf{J}_{22} \end{bmatrix}, \quad (2.23)$$

kde \mathbf{J}_{11} je matice typu $k \times k$.

Dále je potřeba rozlišit dvě situace. V prvním případě se určí MV odhad parametru $\boldsymbol{\theta}$, který není vázán žádnými dodatečnými podmínkami. Takto získaný odhad bude označen $\hat{\boldsymbol{\theta}}$. Protože zde uvažujeme parametr $\boldsymbol{\theta}$ rozdělený na bloky, budeme psát

$$\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\tau}}^T, \hat{\boldsymbol{\psi}}^T)^T,$$

kde $\hat{\boldsymbol{\tau}}$ a $\hat{\boldsymbol{\psi}}$ jsou jednotlivé složky MV odhadu $\hat{\boldsymbol{\theta}}$. Dále se určí odhad parametru $\boldsymbol{\theta}$ za platnosti testované nulové hypotézy, tj. za podmínky, že $\boldsymbol{\tau} = \boldsymbol{\tau}_0$. Příslušná věrohodnostní funkce se tedy maximalizuje jen vzhledem k parametru $\boldsymbol{\psi}$, tento odhad označme $\hat{\boldsymbol{\psi}}_0$. Odpovídající odhad parametru $\boldsymbol{\theta}$ bude pak označen $\hat{\boldsymbol{\theta}}_0$ a je ve tvaru

$$\hat{\boldsymbol{\theta}}_0 = (\boldsymbol{\tau}_0^T, \hat{\boldsymbol{\psi}}_0^T)^T.$$

Pro další účely je potřeba určit inverzní matici \mathbf{J}^{-1} k FIM \mathbf{J} , kterou popisuje následující lemma.

LEMMA 2.4. Nechť matice \mathbf{J} daná vztahem (2.23) je regulární matice, přičemž bloky $\mathbf{J}_{11}, \mathbf{J}_{22}$ jsou čtvercové a regulární. Pak inverzní matice \mathbf{J}^{-1} je ve tvaru

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

kde

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

Důkaz. Předchozí tvrzení se snadno dokáže ověřením, že součin obou matic \mathbf{J} , \mathbf{J}^{-1} dává jednotkovou matici. \square

VĚTA 2.5. [14] Necht' jsou splněny předpoklady věty 2.3 a prvky matice $\mathbf{J}(\boldsymbol{\theta})$ jsou spojitě v bodě $\boldsymbol{\theta}_0$. Necht' $\mathbf{U}(\boldsymbol{\theta})$ je skórový vektor daný vztahem (2.15) a $\mathbf{J}_{11.2}(\boldsymbol{\theta})$ je matice daná vztahem (2.4). Pak náhodné veličiny

$$LM(\boldsymbol{\theta}_0) = \frac{1}{n} \left[\mathbf{U}_1(\hat{\boldsymbol{\theta}}_0) \right]^T \left[\mathbf{J}_{11.2}(\hat{\boldsymbol{\theta}}_0) \right]^{-1} \mathbf{U}_1(\hat{\boldsymbol{\theta}}_0), \quad (2.24)$$

$$W(\boldsymbol{\theta}_0) = n (\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0)^T \mathbf{J}_{11.2}(\hat{\boldsymbol{\theta}}) (\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0), \quad (2.25)$$

$$LR(\boldsymbol{\theta}_0) = 2 \left[l(\hat{\boldsymbol{\theta}}) - l(\hat{\boldsymbol{\theta}}_0) \right] \quad (2.26)$$

mají za platnosti nulové hypotézy $H : \boldsymbol{\tau} = \boldsymbol{\tau}_0$ asymptoticky χ_k^2 rozdělení.

Výsledky této důležité věty budou dále použity v odstavci 3.3, kde bude pozornost zaměřena na odhady parametrů prahového modelu, zejména pak na adaptivní techniky určení vhodné prahové hodnoty.

2.3. Semiparametrické přístupy pro odhady parametrů GEV rozdělení

Semiparametrické odhady GEV rozdělení jsou založeny na prahovém modelu, který byl představen v odstavci 2.1. Jedná se o neparametrické odhady využívající pořadové statistiky, které však předpokládají jistá omezení parametrického prostoru, nejčastěji pak omezení týkající se oboru atraktivity. Mějme náhodný výběr X_1, \dots, X_n a označme $X_{(1)}, \dots, X_{(n)}$ uspořádaný náhodný výběr, kde $X_{(1)} \leq \dots \leq X_{(n)}$ a $X_{(i)}$ označuje i -tou pořadovou statistiku.

V předchozích odstavcích byla pozornost omezena na asymptotické chování výběrového maxima $M_n = X_{(n)}$ pro $n \rightarrow \infty$. Obecně lze ukázat (viz [38, 93]), že pro pevné $k = 0, 1, \dots, n-1$ a konstanty $a_n > 0, b_n$ ze vztahu (1.2) platí

$$\frac{X_{(n-k)} - b_n}{a_n} \xrightarrow{d} \frac{\left(\sum_{i=1}^{k+1} E_i \right)^\xi - 1}{\xi},$$

kde E_i jsou nezávislé náhodné veličiny se standardizovaným exponenciálním rozdělením s distribuční funkcí $F(x) = 1 - e^{-x}$ pro $x \geq 0$. Lze však také uvažovat pořadové statistiky při volbě $k = k(n) \rightarrow \infty$ pro $n \rightarrow \infty$. Přidáme-li navíc podmínku $k(n)/n \rightarrow 0$, pak při takové volbě k lze řádnou normalizací obdržet asymptoticky normálně rozdělené statistiky. Pomocí následujícího tvrzení nejprve zavedeme Hillův odhad EV indexu, který patří k základním semiparametrickým odhadům. Vzhledem ke komplexnosti předkládaných tvrzení nebudou v tomto odstavci provedeny důkazy, neboť tyto by vyžadovaly obsáhlý výklad teorie regulárně se měnících funkcí, který je již mimo rámec této práce.

VĚTA 2.6. [38] Distribuční funkce F patří do oboru atraktivity EV rozdělení G_ξ pro $\xi > 0$ právě tehdy, když $F(x) < 1$ pro všechna $x \in \mathbb{R}$, $\int_1^\infty (1 - F(x)) \frac{dx}{x} < \infty$ a je splněna rovnost

$$\lim_{t \rightarrow \infty} \frac{\int_t^\infty (1 - F(x)) \frac{dx}{x}}{1 - F(t)} = \xi. \quad (2.27)$$

Hillův odhad $\widehat{\xi}_H$ EV indexu ξ vychází z uvedeného vztahu (2.27). Protože

$$\int_t^\infty (1 - F(x)) \frac{dx}{x} = \int_t^\infty (\ln x - \ln t) dF(x),$$

odpovídá vztah (2.27) rovnosti

$$\lim_{t \rightarrow \infty} E(\ln X - \ln t \mid X > t) = \xi. \quad (2.28)$$

Pro náhodný výběr X_1, \dots, X_n označme dále $F_n(x)$ empirickou distribuční funkci, tj. $F_n(x) := n^{-1} \sum_{i=1}^n \mathbf{1}_{[X_i \leq x]}$, kde $\mathbf{1}_{[X_i \leq x]}$ je charakteristická funkce jevu $[X_i \leq x]$. Nahradíme-li ve vztahu (2.28) parametr t pořadovou statistikou $X_{(n-k)}$ a distribuční funkci $F(x)$ jejím empirickým protějškem $F_n(x)$, dostaneme tak Hillův odhad

$$\widehat{\xi}_H(k) := \frac{\int_{X_{(n-k)}}^\infty (\ln s - \ln X_{(n-k)}) dF_n(s)}{1 - F_n(X_{(n-k)})},$$

neboli

$$\widehat{\xi}_H(k) := \frac{1}{k} \sum_{i=0}^{k-1} \ln X_{(n-i)} - \ln X_{(n-k)}. \quad (2.29)$$

Vzhledem k omezení platnosti rovnosti (2.27) je však Hillův odhad použitelný jen pro případ $\xi > 0$.

VĚTA 2.7. [41] Nechť X_1, \dots, X_n je náhodný výběr z rozdělení s distribuční funkcí $F \in \mathcal{D}(G_\xi)$, $\xi \in \mathbb{R}$ a $x^* > 0$. Označme

$$M_n^{(j)}(k) := \frac{1}{k} \sum_{i=0}^{k-1} (\ln X_{(n-i)} - \ln X_{(n-k)})^j. \quad (2.30)$$

Pokud pro $n \rightarrow \infty$ platí $k = k(n) \rightarrow \infty$, $k/n \rightarrow 0$, potom

$$\frac{M_n^{(j)}(k)}{\left(a \left(\frac{n}{k}\right) / U \left(\frac{n}{k}\right)\right)^j} \xrightarrow{P} \prod_{i=1}^j \frac{i}{1 - i\xi_-}, \quad (2.31)$$

kde $\xi_- := \min(0, \xi)$.

Pomocí předchozího označení můžeme tedy psát $\widehat{\xi}_H(k) = M_n^{(1)}(k)$. Zaměříme-li se na případ, kdy $\xi \leq 0$, pak lze ukázat $\widehat{\xi}_H \xrightarrow{P} 0$ [18]. Tento fakt motivuje nový odhad ve tvaru $\widehat{\xi} = \widehat{\xi}_H + \widehat{\xi}_-$, který by byl vhodný pro $\xi \in \mathbb{R}$. Pomocí věty 2.7 lze odvodit

$$\frac{\left(M_n^{(1)}(k)\right)^2}{M_n^{(2)}(k)} \xrightarrow{P} \frac{1 - 2\xi_-}{2(1 - \xi_-)},$$

a toto vede k definici momentového odhadu ve tvaru

$$\widehat{\xi}_M(k) := M_n^{(1)}(k) + 1 - \frac{1}{2} \left(1 - \frac{\left(M_n^{(1)}(k)\right)^2}{M_n^{(2)}(k)} \right)^{-1}. \quad (2.32)$$

Vzhledem k platnosti (2.31) je použitelnost momentového odhadu podmíněna kladným koncovým bodem rozdělení náhodného výběru. V případě, že tomu tak není, nabízí se možnost posunout dané rozdělení, aby $x^* > 0$. Zde je ovšem potřeba zmínit, že momentový odhad (stejně jako Hillův odhad) není invariantní vůči posunutí, ačkoli je invariantní vůči změně měřítka. Pro $x^* \leq 0$ je pak zapotřebí použít jiné odhady (např. Pickandsův [109], pravděpodobnostně vážený momentový [75], de Vriesův [39], zobecněný Hillův [53]), velká část je jich shrnuta např. v pracích [118, 63]. Přesto však představuje $\widehat{\xi}_M$ často používaný odhad, zejména díky svým asymptotickým vlastnostem v porovnání s ostatními odhady (dále viz obrázky 2.1).

VĚTA 2.8. [40, 39] Nechť X_1, \dots, X_n je náhodný výběr z rozdělení s distribuční funkcí $F(x)$ a nechť $F \in \mathcal{D}(G_\xi)$ pro $\xi > 0$. Pokud pro $n \rightarrow \infty$ platí $k = k(n) \rightarrow \infty, k/n \rightarrow 0$, pak

$$\widehat{\xi}_H(k) \xrightarrow{P} \xi. \quad (2.33)$$

Pokud navíc $F(x)$ splňuje podmínku druhého řádu (1.36) s $\rho \leq 0$ a existuje konečná limita

$$\lim_{n \rightarrow \infty} \sqrt{k} A\left(\frac{n}{k}\right) = \lambda,$$

potom

$$\sqrt{k} \left(\widehat{\xi}_H(k) - \xi \right) \xrightarrow{d} \xi N + \frac{\lambda}{1 - \rho}, \quad (2.34)$$

kde $N \sim N(0, 1)$.

Obdobné tvrzení lze formulovat i pro momentový odhad EV indexu, zde je však potřeba uvažovat podmínku druhého řádu v jiném tvaru. Předpokládejme, že pro $x > 0$ je splněna podmínka druhého řádu (1.36), tj.

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx) - U(t)}{a(t)} - D_\xi(x)}{A(t)} = H_{\xi, \rho}(x).$$

Je-li $\rho < 0$ v případě, že $\xi > 0$, a $\xi \neq \rho$, pak platí (viz [38], věta B.3.16)

$$\lim_{t \rightarrow \infty} \frac{\frac{\ln U(tx) - \ln U(t)}{a(t)/U(t)} - D_{\xi_-}(x)}{Q(t)} = H_{\xi_-, \rho'}(x), \quad (2.35)$$

přičemž $\rho' \leq 0$ a $Q(t)$ je vhodná funkce neměnicí znaménko taková, že pro $t \rightarrow \infty$ je $Q(t) \rightarrow 0$. Konkrétně

$$\rho' = \begin{cases} \rho, & \text{pro } \xi < \rho \leq 0, \\ \xi, & \text{pro } \rho < \xi \leq 0, \\ -\xi, & \text{pro } 0 < \xi < -\rho \text{ a } l \neq 0, \\ \rho, & \text{pro } 0 < \xi < -\rho \text{ a } l = 0, \text{ nebo pro } \xi \geq -\rho, \end{cases}$$

kde $l := \lim_{t \rightarrow \infty} (U(t) - a(t)/\xi)$. Ve vztahu (2.35) tedy $Q(t)$ hraje obdobnou roli jako funkce $A(t)$ v podmínce (1.36), jen ve smyslu funkce $\ln U(t)$ namísto $U(t)$. Možná volba $Q(t)$ je následující [38]

$$Q(t) = \begin{cases} A(t), & \xi < \rho \leq 0, \\ \xi_+ - \frac{a(t)}{U(t)}, & \rho < \xi \leq 0 \text{ nebo } (0 < \xi < -\rho \text{ a } l \neq 0) \text{ nebo } \xi = -\rho, \\ \frac{\rho}{\xi + \rho} A(t), & (0 < \xi < -\rho \text{ a } l = 0) \text{ nebo } \xi > -\rho > 0, \\ A(t), & \xi > \rho = 0, \end{cases}$$

kde $\xi_+ := \max(\xi, 0)$ a l je limita uvedená výše.

VĚTA 2.9. [41] Nechť X_1, \dots, X_n je náhodný výběr z rozdělení s distribuční funkcí $F(x) \in \mathcal{D}(G_\xi)$ a $x^* > 0$. Pokud pro $n \rightarrow \infty$ platí $k = k(n) \rightarrow \infty, k/n \rightarrow 0$, pak

$$\widehat{\xi}_M(k) \xrightarrow{P} \xi. \quad (2.36)$$

Pokud je navíc splněna podmínka druhého řádu (1.36) pro $\xi \neq \rho$ a pro Q ze vztahu (2.35) platí, že existuje konečná limita

$$\lim_{n \rightarrow \infty} \sqrt{k} Q\left(\frac{n}{k}\right) = \lambda,$$

potom

$$\sqrt{k} \left(\widehat{\xi}_M(k) - \xi \right) \xrightarrow{d} \sqrt{\text{var}_\xi} N + \lambda b_{\xi, \rho}, \quad (2.37)$$

kde $N \sim N(0, 1)$ a kde dále

$$b_{\xi, \rho} := \begin{cases} \frac{(1-\xi)(1-2\xi)}{(1-\xi-\rho)(1-2\xi-\rho)}, & \xi < \rho \leq 0, \\ \frac{\xi(1+\xi)}{(1-\xi)(1-3\xi)}, & \rho < \xi \leq 0, \\ \frac{-\xi}{(1+\xi)^2}, & 0 < \xi < -\rho, l \neq 0, \\ \frac{\xi - \xi\rho + \rho}{\rho(1-\rho)^2}, & (0 < \xi < -\rho, l = 0) \text{ nebo } \xi \geq -\rho > 0, \\ 1, & \xi > \rho = 0, \end{cases} \quad (2.38)$$

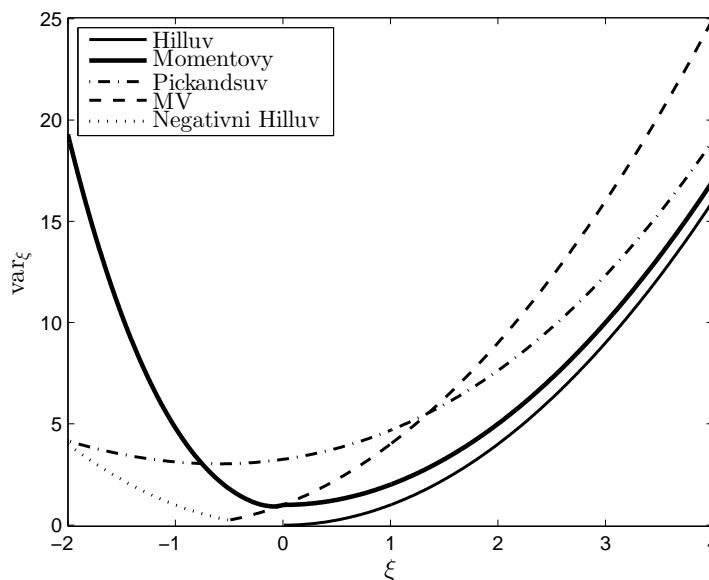
přičemž $l := \lim_{t \rightarrow \infty} (U(t) - a(t)/\xi)$,

$$\text{var}_\xi := \begin{cases} \xi^2 + 1, & \xi \geq 0, \\ \frac{(1-\xi)^2(1-2\xi)(1-\xi+6\xi^2)}{(1-3\xi)(1-4\xi)}, & \xi < 0. \end{cases} \quad (2.39)$$

Obě předešlá tvrzení udávají v podstatě totéž: limitním rozdělením $\widehat{\xi}_H, \widehat{\xi}_M$ je normální rozdělení s možným nenulovým vychýlením, které závisí na původním rozdělení pozorovaných hodnot skrze parametry ξ a ρ . Při aproximaci vztahů (2.34) a (2.37) pro nějaké konečné n je však důležité vhodně určit hodnotu k a tím i podíl n/k . Situaci výběru vhodného počtu horních k pořadových statistik demonstrujeme na jednodušším Hillově odhadu. Vzhledem k (2.34) lze psát

$$\widehat{\xi}_H(k) - \xi \approx \frac{\xi N}{\sqrt{k}} + \frac{\lambda}{\sqrt{k}(1-\rho)} \approx \frac{\xi N}{\sqrt{k}} + \frac{A\left(\frac{n}{k}\right)}{1-\rho}, \quad (2.40)$$

kde \approx^d označuje přibližnou rovnost v distribuci. Mějme rozsah výběru n pevný a dostatečně velký a uvažujme Hillův odhad ve dvou různých bodech $k_1 < k_2$. Varianční složka ξ/\sqrt{k} je zřejmě klesající při rostoucím k , tj. $\xi/\sqrt{k_1} > \xi/\sqrt{k_2}$, zatímco $|A(n/k_1)| > |A(n/k_2)|$, a tedy složka vychýlení roste s rostoucím k . Obdobná situace je u všech semiparametrických odhadů EV indexu a volba vhodného k odpovídá volbě vhodného prahu u v (parametrickém) prahovém modelu s tím rozdílem, že nyní je jako práh uvažována hodnota pořadové statistiky $X_{(n-k)}$. Hledáme tedy optimální počet horních pořadových statistik tak, aby bylo dosaženo nějakého kompromisu mezi možným vychýlením odhadu pro velké k a možnou velkou variabilitou při k malém. Více bude na metody volby vhodného počtu horních pořadových statistik zaměřen odstavec 3.2.



Obrázek 2.1: Asymptotické rozptyly běžně užívaných semiparametrických odhadů EV indexu.

Na obrázku 2.1 jsou v závislosti na hodnotě EV indexu zobrazeny asymptotické rozptyly vybraných významných semiparametrických odhadů: Hillova a momentového odhadu, dále Pickandsova ($\xi \in \mathbb{R}$), MV odhadu ($\xi > -1/2$) a negativního Hillova odhadu ($\xi < -1/2$). Z obrázku je vidět, že pro $\xi > 0$ má $\hat{\xi}_H$ z uvedených jednoznačně nejmenší rozptyl, který se příliš neliší od rozptylu $\hat{\xi}_M$. Navíc má momentový odhad relativně malý rozptyl pro hodnoty $-1/2 < \xi < 0$. Z tohoto důvodu se i v dalších odstavcích soustředíme výhradně na tyto dva odhady.

V první kapitole bylo ukázáno, že podmínka atraktivita je ekvivalentní podmínce

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\xi - 1}{\xi}. \quad (2.41)$$

Vztah (2.41) vede k aproximaci

$$U(x) \approx U(t) + a(t) \frac{\left(\frac{x}{t}\right)^\xi - 1}{\xi}, \quad \text{pro } x > t.$$

Připomeňme, že návratová úroveň z_r , která nás bude v aplikační části zejména zajímat, je $(1 - r^{-1})$ kvantil modelového rozdělení, tj. $z_r = F^{\leftarrow}(1 - r^{-1}) = U(r)$. Dosadíme-li tedy v předcházející rovnosti $t = n/k$ a $x = r$, dostáváme výraz

$$z_r = U(r) \approx U\left(\frac{n}{k}\right) + \frac{a\left(\frac{n}{k}\right)}{\xi} \left[\left(\frac{kr}{n}\right)^\xi - 1 \right],$$

který je zjevně obdobný vztahu (2.8). Jak jsme již naznačili, v semiparametrickém modelu prahová hodnota u odpovídá vhodné pořadové statistice $X_{(n-k)}$, tedy $\widehat{U}(n/k) = X_{(n-k)}$ je odhad $U(n/k)$. Pravděpodobnost překročení prahu $\lambda_u := P(X > U(n/k))$ byla odhadnuta jako $\widehat{\lambda}_u := k/n$, což odpovídá odhadu $\widehat{\lambda}_u$ ze vztahu (2.8) jakožto relativní četnosti pozorování nad prahem. Zřejmě člen $a(n/k)$ bude hrát roli parametru měřítka GP rozdělení.

Uvažujme momentový odhad EV indexu $\widehat{\xi}_M(k)$. Potom definujme příslušný momentový odhad parametru měřítka

$$\widehat{\sigma}_M(k) := X_{(n-k)} M_n^{(1)}(k) \left(1 - \widehat{\xi}_-(k)\right), \quad (2.42)$$

kde $\widehat{\xi}_-(k) := \widehat{\xi}_M(k) - M_n^{(1)}(k)$.

VĚTA 2.10. [38] Nechť je dán náhodný výběr X_1, \dots, X_n z rozdělení s distribuční funkcí $F \in \mathcal{D}(G_\xi)$, $\xi \in \mathbb{R}$ a necht' $x^* > 0$. Předpokládejme, že pro $n \rightarrow \infty$ je $k = k(n) \rightarrow \infty$, $k/n \rightarrow 0$. Pak

$$\frac{\widehat{\sigma}_M(k)}{a\left(\frac{n}{k}\right)} \xrightarrow{P} 1. \quad (2.43)$$

Pokud dále platí, že

- (i) podmínka druhého řádu (1.36) je splněna pro $\xi \neq \rho$,
- (ii) existuje konečná limita

$$\lim_{n \rightarrow \infty} \sqrt{k} Q\left(\frac{n}{k}\right) = \lambda,$$

kde $Q := A$ z (1.36) pro $\xi > 0$ a $\rho = 0$, případně s Q z (2.35) jinak,

potom

$$\sqrt{k} \left(\frac{\widehat{\sigma}_M(k)}{a\left(\frac{n}{k}\right)} - 1 \right) \xrightarrow{d} \sqrt{\text{var}_\xi} N + \lambda b_{\xi, \rho}, \quad (2.44)$$

kde $N \sim N(0, 1)$ a kde dále

$$b_{\xi, \rho} := \begin{cases} -\frac{\rho}{(1-\xi-\rho)(1-2\xi-\rho)}, & \xi < \rho \leq 0, \\ -\frac{\xi}{(1-2\xi)(1-3\xi)}, & \rho < \xi \leq 0, \\ \frac{\xi}{(1+\xi)^2}, & l = 0, 0 < \xi < -\rho, \\ -\frac{\rho}{(1-\rho)^2}, & (l = 0, 0 < \xi < -\rho) \text{ nebo } \xi \geq -\rho > 0, \\ 0, & \xi > \rho = 0, \end{cases} \quad (2.45)$$

přičemž $l := \lim_{t \rightarrow \infty} (U(t) - a(t)/\xi)$,

$$\text{var}_\xi := \begin{cases} \xi^2 + 2, & \xi \geq 0, \\ \frac{2-16\xi+51\xi^2-69\xi^3+50\xi^4-24\xi^5}{(1-2\xi)(1-3\xi)(1-4\xi)}, & \xi < 0. \end{cases} \quad (2.46)$$

Odhad návratové úrovně příslušné periodě návratu r je tedy při použití semiparametrických momentových odhadů ve tvaru

$$\widehat{z}_r = X_{(n-k)} + \frac{\widehat{\sigma}_M(k)}{\widehat{\xi}_M(k)} \left[\left(\frac{kr}{n} \right)^{\widehat{\xi}_M(k)} - 1 \right], \quad (2.47)$$

kde k je nějaká vhodná pevně zvolená hodnota. Jak je odvozeno např. v [38]

$$\sqrt{k} \left(\widehat{\xi}_M(k) - \xi, \frac{\widehat{\sigma}_M(k)}{a(n/k)} - 1 \right) \xrightarrow{d} (R, S),$$

kde (R, S) jsou normálně rozdělené náhodné veličiny se středními hodnotami $\lambda(b_{\xi, \rho}^{\widehat{\xi}_M}, b_{\xi, \rho}^{\widehat{\sigma}_M})$ danými vztahy (2.38) a (2.45), rozptyly $(\text{var}_{\widehat{\xi}_M}^{\xi}, \text{var}_{\widehat{\sigma}_M}^{\sigma})$ odvozenými ve (2.39) a (2.46) a kovariancí

$$\text{cov}(R, S) = \begin{cases} \xi - 1, & \xi \geq 0, \\ \frac{(1-\xi)^2(-1+4\xi-12\xi^2)}{(1-3\xi)(1-4\xi)}, & \xi < 0. \end{cases}$$

Tedy

$$\text{cov} \left(\widehat{\xi}_M(k), \widehat{\sigma}_M(k) \right) = \frac{a \left(\frac{n}{k} \right)}{k} \text{cov}(R, S) \approx \begin{cases} \frac{\widehat{\sigma}_M(k)}{k} \left(\widehat{\xi}_M(k) - 1 \right), & \xi \geq 0, \\ \frac{\widehat{\sigma}_M(k)}{k} \frac{(1-\widehat{\xi}_M(k))^2(-1+4\widehat{\xi}_M(k)-12\widehat{\xi}_M(k)^2)}{(1-3\widehat{\xi}_M(k))(1-4\widehat{\xi}_M(k))}, & \xi < 0. \end{cases}$$

Protože počet překročení prahu λ_u je náhodná veličina s binomickým rozdělením $\text{Bi}(n, \lambda_u)$, lze odhad rozptylu statistiky $\widehat{\lambda}_u = k/n$ určit jako $\text{var}(\widehat{\lambda}_u) \approx \widehat{\lambda}_u(1 - \widehat{\lambda}_u)/n = k(n - k)/n^2$.

Odhad rozptylu odhadu návratové úrovně se následně získá pomocí delta metody podobně jako v odstavci 2.2. Označme $\mathbf{V} = \mathbf{V}(\widehat{\lambda}_u, \widehat{\sigma}_M, \widehat{\xi}_M)$ varianční matici odhadů $\widehat{\lambda}_u, \widehat{\sigma}_M, \widehat{\xi}_M$. Na základě asymptotických vlastností daných odhadů lze varianční matici \mathbf{V} odhadnout jako

$$\mathbf{V} = \begin{pmatrix} k(n-k)/n^2 & 0 & 0 \\ 0 & \text{var}_{\widehat{\xi}_M}^{\widehat{\sigma}_M} & \text{cov} \left(\widehat{\xi}_M(k), \widehat{\sigma}_M(k) \right) \\ 0 & \text{cov} \left(\widehat{\xi}_M(k), \widehat{\sigma}_M(k) \right) & \text{var}_{\widehat{\xi}_M}^{\widehat{\xi}_M} \end{pmatrix}.$$

Užitím delta metody se dostane

$$\text{var}(\widehat{z}_r) \approx \nabla z_r^T \mathbf{V} \nabla z_r, \quad (2.48)$$

kde ∇z_r^T je vektor prvních parciálních derivací z_r daný v odstavci 2.2 (str. 45), přičemž jednotlivé parametry jsou nahrazeny jejich semiparametrickými protějšky. Dle vztahu (2.20) lze stanovit asymptotický $100(1 - \alpha)\%$ interval spolehlivosti pro odhad návratové úrovně.

2.4. Metoda bootstrap

Metoda bootstrap je metodou odhadu patřící do skupiny tzv. výpočetně intenzivních statistických přístupů. Poprvé byla představena v Efronově práci [47]. Historicky vychází

z jiné výpočetně intenzivní metody jackknife. V souvislosti s odhady v EV teorii je možné bootstrap použít například k určení variability odhadů parametrů modelového rozdělení či parametrických funkcí [21, 28, 88, 105] nebo k redukci vychýlení odhadů [63, 64, 62]. Nabízí tak alternativní přístup k běžně užívaným technikám založeným na asymptotických výsledcích.

Nechť je dán náhodný výběr X_1, \dots, X_n z rozdělení o distribuční funkci $F(x) = F(x, \boldsymbol{\theta})$, kde $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)^T$ je vektor parametrů daného rozdělení. Buď $\tau = \tau(F)$ nějaká charakteristika daného rozdělení o distribuční funkci F ; typicky se jedná o parametr daného rozdělení nebo parametrickou funkci. Jako odhad hodnoty τ uvažujme nějakou vhodnou statistiku $T = T(X_1, \dots, X_n)$. Explicitní odvození rozdělení statistiky T může být mnohdy velmi obtížné či dokonce neproveditelné, a to i v případech, kdy je distribuční funkce F známa. Často se dá setkat s případy, kdy je známo jen asymptotické rozdělení statistiky T pro $n \rightarrow \infty$, avšak pro konečný rozsah výběru n je toto rozdělení neznámé. Pro malé rozsahy n pak obvykle vede použití asymptotické teorie k nepřesným výsledkům. Zejména v těchto situacích je možné použít metodu bootstrap, která kombinuje tzv. *substituční princip* a metodu Monte Carlo [115].

Buď \hat{F} nějaký odhad distribuční funkce F . Běžně se lze setkat s případem, kdy $\hat{F}(x) = F_n(x)$, tedy neznámá funkce F je odhadnuta svou empirickou distribuční funkcí. Obecně je však možné, je-li dána nějaká dodatečná informace o typu rozdělení, uvažovat i odhady F typu $\hat{F}(x) = F(x, \hat{\boldsymbol{\theta}})$, přičemž $\hat{\boldsymbol{\theta}}$ je nějaký odhad parametru $\boldsymbol{\theta}$ (získaný např. MV metodou, momentovou metodou apod.)¹. Substituční princip spočívá v nahrazení funkce F tímto jejím odhadem \hat{F} . Vzhledem k předchozímu pak lze charakteristiku $\tau(F)$ odhadnout pomocí hodnoty $\tau(\hat{F})$.

Je-li kupříkladu hledanou charakteristikou střední hodnota daného rozdělení, tj. $\tau(F) = E_F(x)$, kde E_F explicitně vyjadřuje střední hodnotu určenou vzhledem k rozdělení F , můžeme tuto střední hodnotu odhadnout statistikou $T = \tau(\hat{F}) = E_{\hat{F}}(X_1, \dots, X_n)$. Pokud navíc $\hat{F} = F_n$, lze pro tuto střední hodnotu

$$\tau(F) = \int_{-\infty}^{\infty} X \, dF(X)$$

získat její odhad pomocí substitučního principu ve tvaru

$$\tau(\hat{F}) = \int_{-\infty}^{\infty} X \, dF_n(X) = \frac{1}{n} \sum_{i=1}^n X_i.$$

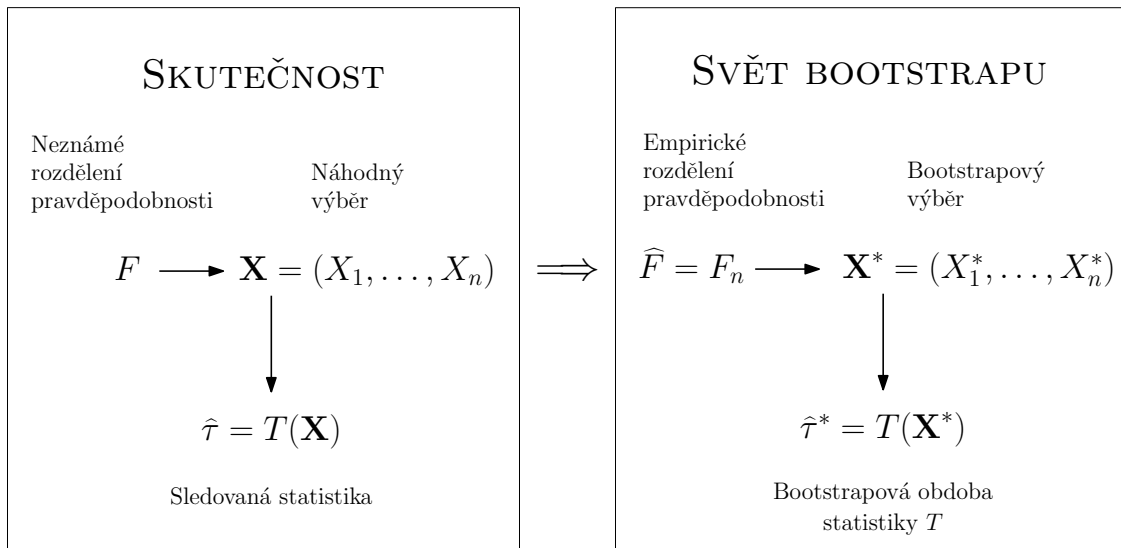
Obecně ale ne všechny statistiky můžeme obdržet pomocí tohoto substitučního principu.

Dále bude pozornost zaměřena na neparametrickou variantu metody bootstrap pro $\hat{F}(x) = F_n(x)$, která bývá v literatuře uvažována nejčastěji (viz např. [29, 28, 27, 69]). Není tak nutné požadovat žádné významně omezující vlastnosti původního rozdělení, resp. distribuční funkce F . Jako *bootstrapový výběr* rozsahu n (přidružený výběru X_1, \dots, X_n) se označuje náhodný výběr X_1^*, \dots, X_n^* z rozdělení o distribuční funkci F_n . Tudíž každá z veličin X_i^* , $i = 1, \dots, n$, nabývá hodnoty X_i , $i = 1, \dots, n$, s pravděpodobností $\frac{1}{n}$ a jedná

¹V literatuře jsou různé typy odhadů \hat{F} pak rozlišovány jako tzv. parametrický a neparametrický bootstrap.

se tak o náhodný výběr s vracením z množiny $\{X_1, \dots, X_n\}$. V dalších úvahách nahradíme původní distribuční funkci F funkcí F_n a náhodný výběr X_1, \dots, X_n bootstrapovým výběrem. Namísto hledané charakteristiky původního rozdělení $\tau(F)$ se tak budeme zajímat o její bootstrapovou obdobu $\tau^* = \tau(F_n)$, k jejímu odhadu pak bude sloužit statistika $T(X_1^*, \dots, X_n^*)$.

Celou situaci shrnuje schéma obrázku 2.2. Pozorováním náhodného výběru X_1, \dots, X_n , resp. statistiky $T(X_1, \dots, X_n)$, se snažíme učinit závěry o charakteristikách neznámé distribuční funkce F . Podobně, sledujeme-li náhodnou veličinu X^* s oborem hodnot $\{X_1, \dots, X_n\}$ a distribuční funkcí F_n , závěry o charakteristikách rozdělení F_n lze učinit na základě pozorování náhodného výběru X_1^*, \dots, X_n^* . Vzhledem k tomu, že funkce F_n je odhadem funkce F , potažmo rozdělení reprezentovaného funkcí F , lze závěry o charakteristikách rozdělení F učinit i na základě bootstrapového výběru. Výhoda metody bootstrap pak spočívá především v možnosti libovolného opakování pozorování výběru X_1^*, \dots, X_n^* . Jak je diskutováno např. v [48], je možné bootstrap použít v nejrůznějších situacích bez ohledu na složitost statistiky T .



Obrázek 2.2: Schéma relací metody bootstrap. (Přejato a upraveno z knihy [48].)

Opakováním, řekněme B krát, pozorování bootstrapového výběru se získají odhady $T^{*(b)}$, $b = 1, \dots, B$, charakteristiky $\tau(F)$. Odhad rozdělení statistiky T se pak dostane pomocí empirického rozdělení pozorování $T^{*(b)}$, $b = 1, \dots, B$. Při rozsahu původního náhodného výběru n je však počet všech různých bootstrapových výběrů $B = \binom{2n-1}{n}$. Ani při současném výkonu výpočetní techniky proto není pro velké n možné uvažovat všechny různé varianty bootstrapových výběrů. Nejčastěji se proto pro získání bootstrapových výběrů aplikuje metoda Monte Carlo [115], kdy se generuje velký počet (řádově tisíce) náhodných a vzájemně nezávislých bootstrapových výběrů z rozdělení F_n . Některá doporučení ohledně dostačujících hodnot B jsou uvedena např. v knize [48]. Na základě teorie výběrových průměrů lze uvažovat statistiku

$$T^* = \frac{1}{B} \sum_{b=1}^B T^{*(b)}, \quad (2.49)$$

kteřá slouží jako odhad odhadu hledané charakteristiky $\tau(F)$. Statistiku (2.49) pak označujeme jako *bootstrapový odhad* charakteristiky $\tau(F)$. Dále je možné získat odhad rozptylu $\text{var} T$ pomocí bootstrapového odhadu ve tvaru

$$\text{var}^*T = \frac{1}{B-1} \sum_{b=1}^B (T^{*(b)} - T^*)^2.$$

Techniku bootstrapování lze aplikovat na širokou škálu statistických problémů: k určení variability odhadů, určení intervalů spolehlivosti, redukci vychýlení odhadů či testování hypotéz [117]. Bootstrap je možné dokonce snadno zobecnit i pro složitější struktury, které obsahují nějakou formu závislosti (viz [48]). Často se tak bootstrapování objevuje ve spojitosti s odhady charakteristik časových řad (např. pro predikce či odhady variability parametrů řad) nebo v regresní analýze (intervalu spolehlivosti u nelineární regrese).

Teoretické výsledky vyplývající z centrálních limitních vět ukazují, že bootstrapové odhady jsou za určitých podmínek konzistentními odhady. Jak uvádí [117], bývá nekonzistentnost bootstrapových odhadů zpravidla způsobena nedostatečnou hladkostí používané statistiky T nebo citlivostí vyplývající z vlastností chvostů rozdělení F . Vlastnosti bootstrapových odhadů přitom obvykle také závisí na způsobu, jakým jsou bootstrapové výběry získány (různé techniky bootstrapování pro různé problematiky jsou shrnuty v [48]). Metoda bootstrapování pak je zcela nevhodná, nejsou-li splněny předpoklady centrální limitní věty pro asymptotické rozdělení, tj. nemá-li sledovaná náhodná veličina konečný rozptyl apod.

V současnosti představují významnou problematiku EV teorie metody adaptivní volby prahové hodnoty [37, 70, 45, 65, 24]. Tyto metody jsou postaveny na prahovém Paretově modelu představeném v odstavci 2.1. Volba vhodné prahové hodnoty je přitom základním požadavkem pro správné určení odhadů sledovaných parametrů GP, resp. GEV rozdělení a dalších parametrických funkcí. Adaptivní přístupy k určení prahu určují optimální prahovou hodnotu pomocí různých statistických kritérií, přičemž se tak snaží eliminovat vliv často subjektivní volby prahu. Některé, v poslední době rozvinuté, metody adaptivní volby prahu jsou dále popsány v odstavcích 3.2 a 3.3. Právě první z nich je na principu bootstrapování založena.

3

Metody výběru prahových hodnot

V předchozí kapitole byly představeny dva základní přístupy k analýze extrémních hodnot, a sice model blokových maxim a prahový model. V této části se zaměříme na druhý zmiňovaný. Jak již bylo diskutováno v části 2.1, prahový model zpravidla dovoluje zahrnout do analýzy výběry většího rozsahu. Na druhou stranu je zde velmi kritická volba prahu, která vyžaduje citlivé vyvážení variability a vychýlení. Volbě vhodné prahové hodnoty, resp. vhodného počtu horních pořadových statistik, bude věnována následující kapitola. Nejprve představíme běžně užívané grafické metody a dále se zaměříme na adaptivní techniky výběru prahu, které nabízí možnost v jistém smyslu objektivně rozhodnout o vhodnosti prahu a také pohodlnou automatizaci výběru prahových hodnot.

3.1. Grafické metody

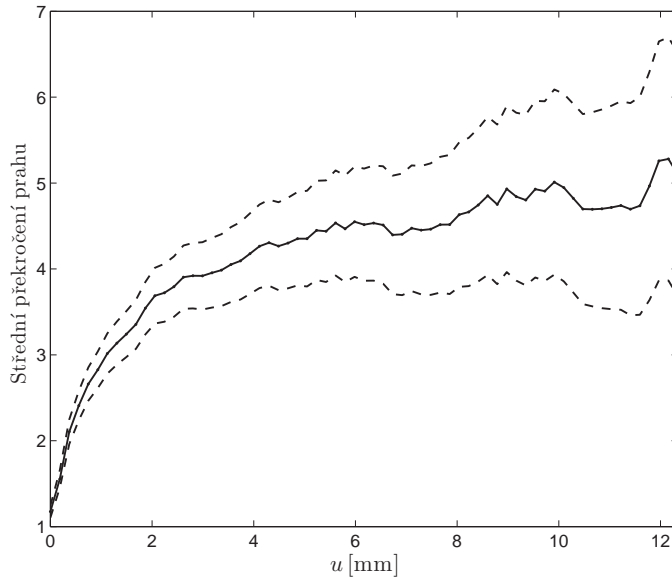
Grafické metody určení prahové hodnoty jsou založeny na vlastnostech GP rozdělení jakožto limitního rozdělení hodnot překračujících práh u . První metoda, známá jako *Mean Residual Life plot* (MRL plot), vychází ze vztahu pro střední hodnotu GP rozdělení, která existuje pro $\xi < 1$. Z rovnic (2.5) a (2.6) plyne následující relace

$$E(X - u \mid X > u) = \frac{\sigma_u}{1 - \xi} = \frac{\sigma + \xi(u - \mu)}{1 - \xi}, \quad (3.1)$$

kde μ, σ, ξ jsou parametry nějakého GEV rozdělení, kterému odpovídá příslušné $GP(\sigma_u, \xi)$ rozdělení. Tedy pro μ, σ, ξ pevné je střední hodnota $E(X - u \mid X > u)$ lineárně závislá na prahové hodnotě u . Je-li dán náhodný výběr X_1, \dots, X_n a jeho realizace $X_1 = x_1, \dots, X_n = x_n$, MRL plot spočívá ve vynesení závislosti empirického protějšku uvedené střední hodnoty na prahové hodnotě, tj. ve vynesení bodů

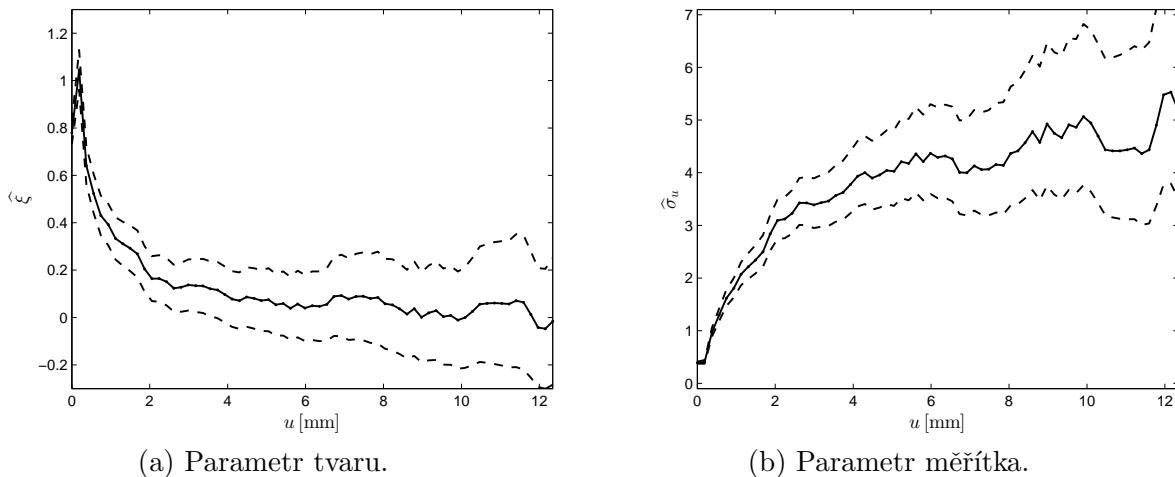
$$\bar{x}_u(u) = \frac{1}{n_u} \sum_{i=1}^n (x_{(i)} - u)_+,$$

kde $a_+ = \max(a, 0)$, $u < x_{(n)}$ a $n_u := \max\{i : x_{(n-i+1)} > u\}$ je počet pozorování nad prahem u . Jako optimální práh $u_0 = u_0(n)$ je potom zvolena taková hodnota, pro kterou je vynesená závislost přibližně lineární pro $u > u_0$. Na obrázku 3.1 je zobrazen MRL plot měřených 30minutových dešťových srážek, jejichž podrobné analýze bude věnována následující kapitola. Spolu s ním je zde také vynesena 95% interval spolehlivosti určený na základě asymptotické normality výběrového průměru.



Obrázek 3.1: Mean residual life plot - průměrná hodnota překročení v závislosti na prahové hodnotě u spolu s 95 % asymptotickým intervalem spolehlivosti.

Další běžně používaná metoda využívá vztahu mezi GP rozdělením a příslušným GEV rozdělením. Je přirozené požadovat stabilitu odhadů parametrů $GEV(\mu, \sigma, \xi)$ rozdělení jakožto limitního zákona rozdělení výběrových maxim. Jsou-li parametry μ, σ, ξ pevné, pak dle (2.6) parametr měřítka σ_u GP rozdělení lineárně závisí na prahové hodnotě u a parametr tvaru ξ přímo odpovídá EV indexu. Grafické určení vhodného prahu tedy spočívá ve vynesení odhadů $\hat{\xi}(u), \hat{\sigma}_u(u)$ parametrů GP rozdělení v závislosti na u . Optimální prahová hodnota u_0 se pak vybere tak, aby pro $u > u_0$ byl odhad $\hat{\xi}(u)$ přibližně konstantní a $\hat{\sigma}_u(u)$ byl přibližně lineární, resp. konstantní je-li ξ blízké nule. Grafické znázornění stability odhadů pro dešťové srážky analyzované v kapitole 4 jsou zobrazeny na obrázku 3.2. Odhady parametrů zde byly stanoveny pomocí metody maximální věrohodnosti a jejich 95% intervaly spolehlivosti na základě asymptotické normality.



(a) Parametr tvaru.

(b) Parametr měřítka.

Obrázek 3.2: Závislost MV odhadů parametrů na prahové hodnotě u spolu s 95% asymptotickým intervalem spolehlivosti.

Obě představené grafické metody byly implementovány do softwaru EVDest (viz dodatky C), který jako nadstavba nad prostředím Matlab nabízí uživatelsky příjemnou analýzu

extrémních hodnot. Grafické volby prahu zde slouží jako základní nástroj pro určení prahové hodnoty. Pro jiné použití mimo software EVDest je k této práci také přiložena funkce `mrlp_par_plot.m` (viz dodatek D), kde jsou opět obě grafické techniky zpracovány. Detailní popis použití funkce lze nalézt také přímo ve zdrojovém kódu.

Shrnutí některých dalších, dnes již méně často využívaných, grafických metod lze nalézt například v práci [116]. Grafické metody volby prahu jsou oblíbené zejména díky své jednoduchosti. Na druhou stranu jsou výše představené přístupy zatíženy značnou mírou subjektivního rozhodování, díky čemuž jsou častým cílem kritiky (např. viz [116]). V praktických situacích tak bývají k určení vhodné prahové hodnoty ve značné míře vyžadovány zkušenosti z dané aplikační oblasti. V současnosti proto roste zájem o metody adaptivního určování prahu, které jednak nabízí možnost pohodlné automatizace výběru prahu. Převážně se však snaží odstranit lidský faktor a pomocí prostředků matematické statistiky rozhodnout o vhodnosti prahu. Tyto metody jsou však zpravidla výpočetně náročnější a jejich podstata méně intuitivní. V následujícím textu se na tyto metody adaptivní volby prahu dále soustředíme, zejména pak na významnou třídu metod pro semiparametrické odhady, která je založena na technice bootstrapování.

3.2. Adaptivní volba prahu v semiparametrickém modelu

Mějme dán náhodný výběr X_1, \dots, X_n z rozdělení o distribuční funkcí $F(x)$. Připomeňme, že semiparametrické odhady jsou obecně postaveny na uspořádaném náhodném výběru $X_{(1)} \leq \dots, X_{(n)}$ za případného omezení oboru atraktivitu příslušného GEV rozdělení. Obdobou volby prahové hodnoty u v parametrickém modelu (viz odstavec 3.1) je zde volba vhodného počtu $k = k(n)$ horních pořadových statistik $X_{(n)}, \dots, X_{(n-k)}$, které jsou v prahovém modelu brány v potaz, přičemž pro $n \rightarrow \infty$ je $k \rightarrow \infty, k/n \rightarrow 0$.

Vzhledem ke složitosti dalších úvah je vhodné se omezit na speciální případy rozdělení. Předpokládejme, že dané rozdělení s distribuční funkcí $F(x)$ patří do takové třídy rozdělení, pro kterou je pomocná funkce druhého řádu ve tvaru

$$A(t) = ct^\rho, \quad c \neq 0, \rho < 0. \quad (3.2)$$

Nejčastěji bývá pro semiparametrický model uvažována tzv. Hallova třída rozdělení [45, 65] s těžkými chvosty a kvantilovou funkcí chvostu ve tvaru

$$U(t) = Ct^\xi (1 + Dt^\rho + o(t^\rho)),$$

kde $C > 0, \xi > 0$ je EV index rozdělení, $\rho < 0$ je parametr druhého řádu (viz tabulka 1.1 a 1.4) a $o(t^\rho)$ je funkce „malé o“, tedy funkce rostoucí pro $t \rightarrow \infty$ pomaleji než t^ρ (viz str. 143). Hallova třída přitom představuje značně obecnou třídu rozdělení, mezi něž patří např. Fréchetovo, Burrovo (typ XII), GP či $|t|$ rozdělení. Detailní odvození tvarů pomocných funkcí prvního a druhého řádu Hallovy třídy rozdělení lze nalézt v dodatku A.

Za určitých předpokladů na funkci druhého řádu lze u semiparametrických odhadů $\hat{\xi}(k)$ EV indexu ukázat, že tyto odhady mají asymptoticky normální rozdělení (viz odstavec 2.3), tj. obecně můžeme psát

$$\sqrt{k} \left(\hat{\xi}(k) - \xi \right) \xrightarrow{d} \sqrt{\text{var}_\xi} N + \lambda b_{\xi, \rho},$$

kde $N \sim N(0, 1)$, var_ξ je složka variability závislé na ξ a $\lambda b_{\xi, \rho}$ je složka vychýlení závislé na ξ a parametru druhého řádu $\rho \leq 0$. Přitom limita $\lim_{n \rightarrow \infty} \sqrt{k} A\left(\frac{n}{k}\right) = \lambda$ je konečná a funkce A je pomocná funkce z podmínky druhého řádu (1.36), resp. pro momentový odhad $\widehat{\xi}_M(k)$ pomocná funkce druhého řádu z podmínky (2.35).

Optimální počet $k_0 = k_0(n)$ horních pořadových statistik je třeba volit s ohledem na značné vychýlení odhadu pro k velké a velkou variabilitu odhadu pro k malé. Obvykle je v semiparametrickém modelu k_0 voleno tak, aby byla minimalizována *střední kvadratická chyba* (zkráceně MSE z angl. Mean Square Error) odhadu EV indexu ξ , tedy

$$k_0 \in \underset{k}{\operatorname{argmin}} \operatorname{MSE} \left(\widehat{\xi}(k) \right) = \underset{k}{\operatorname{argmin}} \operatorname{E} \left(\widehat{\xi}(k) - \xi \right)^2. \quad (3.3)$$

Podobně jako ve vztahu (2.40) můžeme aproximovat rozdělení odhadu $\widehat{\xi}(k)$ následovně

$$\widehat{\xi}(k) - \xi \stackrel{d}{\approx} \frac{\sqrt{\operatorname{var}_\xi}}{\sqrt{k}} N + A\left(\frac{n}{k}\right) b_{\xi, \rho}. \quad (3.4)$$

Protože střední hodnota (3.3) nemusí existovat (typicky pro rozdělení z Fréchetova oboru atraktivity, odstavec 1.3), bývá namísto výrazu (3.3) minimalizována asymptotická střední kvadratická chyba (AMSE), tj. kombinací (3.4) a (3.3) máme

$$k_0 \in \underset{k}{\operatorname{argmin}} \left(\frac{\operatorname{var}_\xi}{k} + A^2\left(\frac{n}{k}\right) b_{\xi, \rho}^2 \right). \quad (3.5)$$

Za předpokladu Hallovy třídy rozdělení můžeme pro $A(t) = ct^\rho$ předchozí přepsat jako

$$\underset{k}{\operatorname{argmin}} \left(\frac{\operatorname{var}_\xi}{k} + c^2 b_{\xi, \rho}^2 \left(\frac{n}{k}\right)^{2\rho} \right)$$

a pro $r := n/k$ pak dostáváme

$$\underset{(n/r)=1,2,\dots}{\operatorname{argmin}} \left(\frac{r \operatorname{var}_\xi}{n} + c^2 b_{\xi, \rho}^2 r^{2\rho} \right).$$

Pro jednoduchost hledáme řešení náležití

$$\underset{t>0}{\operatorname{argmin}} \left(\frac{t \operatorname{var}_\xi}{n} + c^2 b_{\xi, \rho}^2 t^{2\rho} \right).$$

Klasickými přístupy matematické analýzy, když se vyjde z první a druhé derivace, se snadno získá řešení předchozího problému. Zpětnou substitucí $t = r = n/k$ se dostane, za předpokladu (3.2), vhodný počet k_0 horních pořadových statistik ve tvaru

$$k_0(n) = \left[\left(\frac{\operatorname{var}_\xi}{-2\rho c^2 b_{\xi, \rho}^2} \right)^{1/(1-2\rho)} n^{-2\rho/(1-2\rho)} \right]. \quad (3.6)$$

Přikročme nyní k odhadu hodnoty k_0 pomocí adaptivní techniky založené na metodě bootstrap. Tento přístup byl poprvé navržen Hallem [70] a později rozvinut například v práci [45] pro momentový odhad, resp. v práci [65] pro Hillův odhad. Ze vztahu (3.6) je patrné, že odhad hodnoty k_0 vyžaduje, mimo jiné, také odhad parametru druhého řádu.

Problematika odhadu ρ spolu s obtížemi, které jej doprovází, je v literatuře často diskutována. Autoři práce [63] dokonce uvažují pevně stanovený odhad $\hat{\rho} = -1$, přičemž argumentují tím, že existující odhady nejsou příliš kvalitní. Použití techniky bootstrapování pro odhad k_0 zvyšuje výpočetní náročnost odhadu, na druhou stranu vhodnou kombinací více stávajících odhadů umožňuje vyhnout se odhadu parametru ρ . Od roku 2000 bylo sice navrženo několik pokročilejších odhadů parametru druhého řádu [42, 54, 61, 66, 135], jakýkoli odhad ρ však vnáší do odhadu k_0 další nadbytečnou variabilitu. Navíc je často předpokládáno splnění podmínky třetího řádu (viz [54]). Princip adaptivní bootstrapové metody je následující.

1. Při minimalizaci výrazu (3.3) je nejprve neznámý EV index ξ nahrazen nějakým jiným vhodným odhadem $\hat{\xi}_{\text{aux}}(k)$. Podobně je nahrazena také distribuční funkce $F(x)$, vzhledem ke které je určena daná střední hodnota, empirickou distribuční funkcí $F_n(x)$.
2. Následně je z rozdělení $F_n(x)$ bootstrapován náhodný výběr $X_1^*, \dots, X_{n_1}^*$ o rozsahu $n_1 < n$ a pro $k = 2, \dots, n_1$ jsou určeny bootstrapové odhady $\hat{\xi}^*(k)$, $\hat{\xi}_{\text{aux}}^*(k)$. Označme dále $q_{n_1}^*(k) = \left(\hat{\xi}^*(k) - \hat{\xi}_{\text{aux}}^*(k)\right)^2$.

3. Krok 2 je opakován B krát nezávisle, čímž se obdrží bootstrapová řada $q_{n_1,b}^*(k)$, $k = 2, \dots, n_1$, pro každé opakování $b = 1, \dots, B$.

4. Bootstrapový odhad AMSE (3.3)

$$\widehat{\text{AMSE}}^*(n_1, k) = \frac{1}{B} \sum_{b=1}^B q_{n_1,b}^*(k)$$

je minimalizován vzhledem ke k . Necht' $k_0^*(n_1)$ je bod, v němž je minima dosaženo.

5. Rozsah n_1 je nahrazen rozsahem $n_2 = \lceil (n_1)^2/n \rceil$ a kroky 2 až 4 se opakují, čímž se obdobně dle kroku 4 získá hodnota $k_0^*(n_2)$.

6. Bootstrapový odhad \hat{k}_0^* optimálního k minimalizujícího výraz (3.3) je získán ve tvaru

$$\hat{k}_0^* = \frac{(k_0^*(n_1))^2}{k_0^*(n_2)}. \quad (3.7)$$

- (7.) Z podstaty techniky bootstrap může odhad \hat{k}_0^* významněji kolísat. Je tak vhodné kroky 2 až 6 N krát nezávisle opakovat, přičemž dle (3.7) získáme odhady $\hat{k}_{0,i}^*$, $i = 1, \dots, N$. Konečný odhad k_0 je pak získán kombinací hodnot $\hat{k}_{0,1}^*, \dots, \hat{k}_{0,N}^*$, typicky je možné použít průměr či medián.

Kritickým bodem předešlého postupu je volba vhodného pomocného odhadu $\hat{\xi}_{\text{aux}}$. Draisma a kol. [45] se ve své práci zaměřili na studium momentového odhadu $\hat{\xi}_M(k)$,

přičemž jako pomocný odhad byl zvolen odhad $\widehat{\xi}_{M3}(k)$ využívající třetích momentů, tedy odhad ve tvaru

$$\widehat{\xi}_{M3}(k) := \sqrt{\frac{M_n^{(2)}(k)}{2} + 1} - \frac{2}{3} \left(1 - \frac{M_n^{(1)}(k)M_n^{(2)}(k)}{M_n^{(3)}(k)} \right)^{-1}, \quad (3.8)$$

který má podobné asymptotické vlastnosti jako momentový odhad. Označme

$$\widehat{\xi}_{M4}^*(k) := \left(\widehat{\xi}_M^*(k) - \widehat{\xi}_{M3}^*(k) \right) \mathbf{1}_{\left[\left| \widehat{\xi}_M^*(k) - \widehat{\xi}_{M3}^*(k) \right| \leq k^{\delta-1/2} \right]},$$

kde $\delta > 0$ a odhady $\widehat{\xi}_M^*(k)$, $\widehat{\xi}_{M3}^*(k)$ jsou získány z nějakého vhodného bootstrapového výběru. Jak je dokázáno v [45], je-li $\bar{k}_0^*(n)$ řešením $\operatorname{argmin}_k \mathbb{E} \left(\widehat{\xi}_{M4}^*(k) \right)^2$, kde n je rozsah bootstrapového výběru, pak

$$\widehat{k}_0^* / \frac{\left(\bar{k}_0^*(n_1) \right)^2}{\bar{k}_0^*(n_2)} \xrightarrow{P} 1.$$

Protože uvedená konvergence platí pro libovolné $\delta > 0$, plyne odtud odhad (3.7).

Gomes a Oliveira [65] studovali adaptivní volbu k_0 při použití Hillova odhadu $\widehat{\xi}_H(k)$. Neznámý EV index ve své práci navrhli nahradit za odhad $\widehat{\xi}_H(k_{\text{aux}})$, kde k_{aux} je nějaká vhodná počáteční hodnota zvolená pevně tak, aby $\widehat{\xi}_H(k_{\text{aux}})$ byl konzistentním odhadem ξ . Optimální počet $k_0 = k_0(n, k_{\text{aux}})$ horních pořadových statistik je určen řešením výrazu

$$k_0 \in \operatorname{argmin}_k \mathbb{E} \left(\widehat{\xi}_H(k) - \widehat{\xi}_H(k_{\text{aux}}) \right)^2.$$

Simulační studie v [65] ukazuje, že ačkoliv je získaný odhad \widehat{k}_0^* téměř nezávislý na volbě rozsahu n_1 , uvedená metoda je značně citlivá na volbu pomocného parametru k_{aux} , který má zásadní vliv na odhad hodnoty k_0 . Především postup Draismy a kol. se tak ukazuje jako vhodnější.

Přístup založený na myšlenkách Draismy a kol. byl rozvinut v pracích [37, 65]. Neznámý parametr ξ zde byl nahrazen jinými pomocnými statistikami explicitně závislejšími na k . Ačkoli navržené postupy dovolují relaxovat často nejasnou volbu optimálního prahu pomocí více prahových hodnot, celkově situaci značně komplikují, neboť tyto již vyžadují odhad parametru druhého řádu ρ .

Oba přístupy [45, 65] byly softwarově implementovány v prostředí Matlab. Pro adaptivní volbu prahové hodnoty je tak možné použít funkce `db_moe.m` a `db_hill.m`, které jsou přiloženy k této práci. Jejich popis je uveden v dodatku D, detailní popis vstupních a výstupních parametrů je možné nalézt přímo ve zdrojovém kódu. Obě funkce jsou optimalizovány vzhledem k výpočetní náročnosti celého algoritmu (díky maticovému zápisu jsou všechny bootstrapové odhady určeny zároveň) a jsou pro uvedené účely daleko vhodnější než kombinace funkcí `moe.m`, `moe3.m`, `hill.m` s momenty (2.7) počítanými pomocí funkce `jth_moment.m`. Poslední uvedené funkce lze v příloze také nalézt; tyto jsou svou jednoduchostí často vhodnější pro běžné užití mimo algoritmus dvojitého bootstrapování.

3.3. Adaptivní volba prahu založená na věrohodnostní funkci

V poslední době se na adaptivní techniky určení prahové hodnoty zaměřila řada autorů (viz např. [56, 63, 68, 101, 124, 128, 131]). Zde se soustředíme na postup navržený v práci [102], který svou podstatou vychází z grafických metod určení prahu, konkrétně z techniky studující stabilitu parametrů. Omezme se na určení prahové hodnoty vzhledem ke stabilitě EV indexu ξ (viz obrázek 3.2.a), jakožto klíčového parametru popisujícího vlastnosti chvostů EV rozdělení. Přístupy uvedenými v části 3.1 je vhodná prahová hodnota u_0 z grafu závislosti určena tak, aby pro $u \geq u_0$ přibližně platilo $\xi(u) = \xi(u_0)$. Jedná se tedy o testování nulové hypotézy

$$H : \xi(u) = \xi(u_0), \quad \text{pro } u \geq u_0,$$

přičemž hodnota u_0 je zvolena pevně. Autoři publikace [102] navrhli metodu, která je založena na věrohodnostní funkci a jež je postavena na přístupu navrženém ve studii [131]. Uvažujme dvojici prahů (v, u) , $v < u$, a označme ξ_{vu} , resp. ξ_u , parametr tvaru GP rozdělení odhadnutý pro pozorování na intervalu (v, u) , resp. (u, ∞) . Metoda odvozená ve [131] je postavena na testování nulové hypotézy $H : \xi_{vu} = \xi_u$ pro nějaké pevné hodnoty v, u . Jak je pojednáno ve [102], předešlý přístup je zcela nevhodný pro praktické použití, neboť vyžaduje testování uvedené hypotézy pro všechna v, u z nějakého intervalu možných prahových hodnot, čímž výrazně narůstá také výpočetní náročnost uvedeného postupu. Uvedený model navíc předpokládá konstantní parametr tvaru ξ_u na značně rozsáhlém intervalu.

Northrop a Coleman [102] proto uvedený postup modifikovali a ve své práci uvažují diskretizaci pomocí m prahových hodnot u_1, \dots, u_m , $u_1 < \dots < u_m$, tudíž

$$\xi(x) = \begin{cases} \xi_i, & \text{pro } u_i < x < u_{i+1}, \quad i = 1, \dots, m-1, \\ \xi_m, & \text{pro } x > u_m. \end{cases} \quad (3.9)$$

Parametr tvaru ξ je tak modelován jako po částech konstantní funkce $\xi(x)$ se změnami v bodech $u_i, i = 2, \dots, m$. Uvedený přístup je označován jako *Multiple-Threshold Generalized Pareto model* (zkráceně jej budeme označovat jako MT-GP model).

Nechť je dána náhodná veličina X z rozdělení s distribuční funkcí $F(x)$. Bud' $v_i = u_i - u_1$ pro $i = 1, \dots, m$ a $w_i = v_{i+1} - v_i$ pro $i = 1, \dots, m-1$. Označme Y překročení prahu u_1 , tj. $Y = X - u_1$. Předpokládejme dále, že náhodná veličina $Y - v_i = X - u_i$ má za podmínky $v_i < Y < v_{i+1}$ (tj. za podmínky $u_i < X < u_{i+1}$) GP(σ_i, ξ_i) rozdělení, přičemž $v_{m+1} := v_m - \sigma_m / \xi_m$, pokud $\xi_m < 0$, a $v_{m+1} := \infty$ jinak. Distribuční funkce $H_i(x)$ a hustota $h_i(x)$ GP(σ_i, ξ_i) rozdělení jsou ve tvarech daných vztahem (2.5), tudíž

$$H_i(x) = 1 - \left(1 + \xi_i \frac{x}{\sigma_i}\right)_+^{-1/\xi_i}, \quad (3.10)$$

$$h_i(x) = \frac{1}{\sigma_i} \left(1 + \xi_i \frac{x}{\sigma_i}\right)_+^{-(1+1/\xi_i)}. \quad (3.11)$$

Označme $f_Y(y)$ hustotu náhodné veličiny Y . Pak má náhodná veličina $Y - v_i$ za podmínky $v_i < Y < v_{i+1}$ podmíněnou hustotu $f_Y(y | v_i < y < v_{i+1})$ ve tvaru

$$f_Y(y | v_i < y < v_{i+1}) = \frac{h_i(y)}{H_i(w_i)}, \quad v_i < y < v_{i+1}. \quad (3.12)$$

Nechť $p_i := P(Y > v_i)$. Předpokládejme dále, že prahová hodnota u_1 je zvolena tak, aby bylo $X > u_1$, tj. $p_1 = P(Y > v_1) = 1$. Potom platí

$$p_i = \prod_{j=1}^{i-1} (1 - H_j(w_j)).$$

Odtud plyne relace

$$P(v_i < Y < v_{i+1}) = p_i - p_{i+1} = p_i H_i(w_i), \quad (3.13)$$

a tedy

$$f_Y(y) = \prod_{i=1}^m [p_i h_i(y)] \mathbf{1}_{[v_i < y < v_{i+1}]}. \quad (3.14)$$

Aby nedocházelo k nespojitostem v hustotě $f_Y(y)$, je vhodné uvažovat parametry měřítka ve tvaru $\sigma_{i+1} = \sigma_i + \xi_i w_i = \sigma_1 + \sum_{j=1}^i \xi_j w_j$, kde $\sigma_1 > 0$. Původní model s $2m$ parametry je tak možné parametrizovat pomocí menšího počtu $m + 1$ parametrů $\boldsymbol{\theta} = (\sigma_1, \xi_1, \dots, \xi_m)$. Nechť je dán náhodný výběr $\mathbf{Y} = (Y_1, \dots, Y_n)$ překročení prahu u_1 z hustoty (3.14). Užitím (2.13) je možné zapsat logaritmickou věrohodnostní funkci v následujícím tvaru

$$\begin{aligned} l(\sigma_1, \xi_1, \dots, \xi_m) &= \prod_{i=1}^n f(y_i) = \\ &= \sum_{i=1}^n \sum_{j=1}^m \mathbf{1}_{[v_j < y_i < v_{j+1}]} \left\{ \ln p_j - \ln \sigma_j - \left(1 + \frac{1}{\xi_j}\right) \ln \left[1 + \xi_j \frac{y_i - v_j}{\sigma_j}\right] \right\}, \end{aligned} \quad (3.15)$$

za podmínek, že $1 + \xi_j w_j / \sigma_j > 0$ a $1 + \xi_j (y_i - v_j) / \sigma_j > 0$ pro všechna $j = 1, \dots, m$, $i = 1, \dots, n$.

Uvažujme nyní situaci, kdy chceme ověřit, zda je parametr tvaru ξ identický na všech intervalech (v_i, v_{i+1}) , $i = 1, \dots, m$, tj. testujeme nulovou hypotézu $H : \xi_1 = \dots = \xi_m$. Zamítnutí této hypotézy by pak ukazovalo, že pro $u \geq u_1$ nebylo dosaženo uspokojivé stability EV indexu, a bylo by tak nutné uvažovat jako vhodný práh nějakou hodnotu větší než u_1 . Označme dále $\widehat{\boldsymbol{\theta}}$ MV odhad parametru $\boldsymbol{\theta} = (\sigma_1, \xi_1, \dots, \xi_m)$ získaný z (3.15) a $\widehat{\boldsymbol{\theta}}_0$ MV odhad získaný za platnosti nulové hypotézy, tj. za platnosti $\xi_1 = \dots = \xi_m$. Nulovou hypotézu lze testovat na základě statistik (2.26) a (2.24), tedy

$$LM(\widehat{\boldsymbol{\theta}}_0) = \frac{1}{n} \mathbf{U}^T(\widehat{\boldsymbol{\theta}}_0) \mathbf{J}^{-1}(\widehat{\boldsymbol{\theta}}_0) \mathbf{U}(\widehat{\boldsymbol{\theta}}_0), \quad (3.16)$$

$$LR(\widehat{\boldsymbol{\theta}}_0) = 2 \left(l(\widehat{\boldsymbol{\theta}}) - l(\widehat{\boldsymbol{\theta}}_0) \right), \quad (3.17)$$

kde $U(\widehat{\boldsymbol{\theta}}_0)$ je skórový vektor příslušný hustotě (3.14) (viz věta 2.3) a $\mathbf{J}(\widehat{\boldsymbol{\theta}}_0)$ je očekávaná FIM. Jedná se asymptotické testy s rušivými parametry. Za předpokladu, že $\xi_m > -1/2$

a tedy MV odhady mají obvyklé asymptotické vlastnosti [120] (viz odstavec 2.2 pro diskuzi regularity GP rozdělení), mají podle věty 2.5 uvedené statistiky asymptoticky χ_{m-1}^2 rozdělení. Odvození tvarů statistik $U(\hat{\boldsymbol{\theta}}_0)$ a $\mathbf{J}(\hat{\boldsymbol{\theta}}_0)$ lze nalézt v dodatku B, tyto byly autorem práce ověřeny a částečně zjednodušeny.

V případě zamítnutí nulové hypotézy $H : \xi_1 = \dots, \xi_m$ se přistupuje k postupnému testování stability EV indexu pro prahové hodnoty (u_i, \dots, u_m) , dokud použité kritérium nevede pro nějaké $i = 2, \dots, m - 1$ k jejímu přijetí. Při nulové hypotéze $H : \xi_i = \dots = \xi_m$ je tak asymptotické rozdělení statistik (3.16) a (3.17) χ_{m-i}^2 .

Výhodou statistiky (3.16) je, že model je nutné odhadnout jen pro sledovanou prahovou hodnotu u_i . Naproti tomu je výpočet statistiky (3.17) více výpočetně náročný, neboť je vždy nutné odhadnout všech m parametrů tvaru. Ukazuje se, že zásadní částí v MT-GP modelu je vhodná volba počtu m prahových hodnot. Do jisté míry je tak úloha hledání vhodného prahu u_0 přenesena právě na tento problém optimálního stupně diskretizace. Problematika volby m byla částečně studována v [102], ovšem pro detailnější popis situace by bylo vhodné provést rozsáhlou simulační studii. Na jednu stranu je vhodné volit m co největší tak, aby bylo dosaženo co možná nejhustšího pokrytí intervalu možných prahových hodnot. Na druhou stranu při velkých hodnotách m jsou parametry modelu odhadnuty se značnou variabilitou a rozdíly mezi odhady parametrů jsou tak detekovány s menší pravděpodobností. Navíc pro velké m mohou také nastat problémy spojené s konvergencí MV metody. Odstavec 4.2 bude zaměřen na odhady prahu pomocí uvedené adaptivní techniky pro případ srážkových měření, přičemž volba m bude diskutována pro některé speciální případy rozdělení. Cílem však bude převážně porovnání MT-GP metody a semiparametrického přístupu založeného na dvojitým bootstrappingu.

Statistická analýza extrémních dešťových srážek

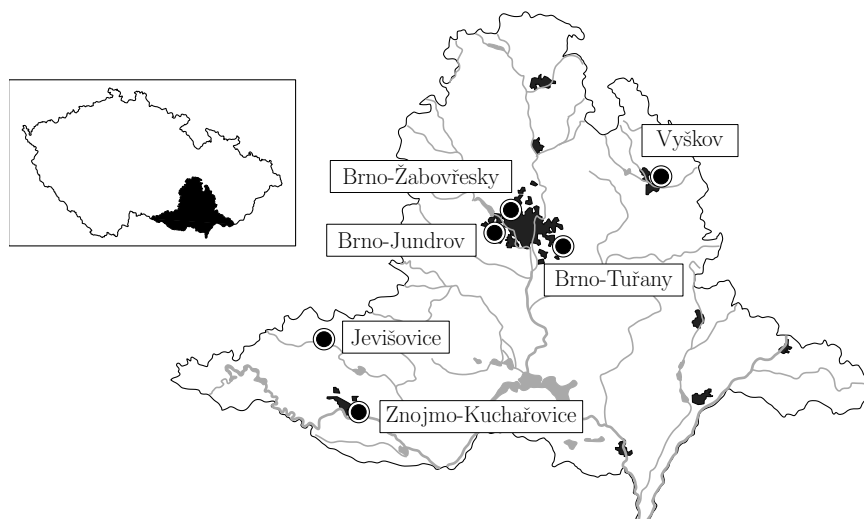
Poznatky z předešlých kapitol budou aplikovány k odhadu extrémních srážkových úhrnů v jihomoravském regionu. Tato analýza byla motivována nutností aktualizace hydrologických podkladů pro potřeby plánování a údržby městského odvodňovacího systému. Zejména pak bude pozornost zaměřena na odhady tzv. *IDF křivek* (zkráceně z angl. *Intensity-Duration-Frequency*), které představují základní hydrologický nástroj pro vyhodnocení frekvence výskytu intenzivních dešťových srážek. Dodnes u nás používané odhady IDF křivek (viz [126]) se jeví jako zastaralé a nově vyvinuté statistické metody by tak mohly přispět k jejich zpřesnění.

Nejprve budou popsána uvažovaná srážková data a v souladu s odstavcem 2.1 bude diskutováno použití modelu blokových maxim a prahového modelu, které se v hydrologii používají nejčastěji. Budou také shrnuty užívané vzorkovací techniky, které slouží k výběru dat do statistického souboru. Vzhledem k prahovému modelu se zaměříme na jeho kritickou část, a sice volbu prahové hodnoty. Diskutovány budou představené adaptivní přístupy výběru prahu a jejich vlastnosti budou studovány pomocí simulační studie. Následně také porovnáme adaptivní přístupy s klasickými grafickými technikami.

4.1. Srážková data a techniky vzorkování

Údaje o srážkových úhrnech jsou v České republice získávány z registračních pásek ombrografů (přístrojů pro měření dešťových srážek) provozovaných Českým hydrometeorologickým ústavem (ČHMÚ). Ombrografy, jakožto plovákové srážkoměry, z podstaty své konstrukce neumožňují měření pevných částí a bývají tak běžně osazovány jen v období od května do září, kdy se však v České republice vyskytuje většina přívalových srážek. Proto bývají analýzy extrémních srážek vycházející z těchto měření považovány za spolehlivé. Na plovák ombrografu je napojeno registrační zařízení, které zapisuje údaje na otáčející se papír, čímž vzniká tzv. ombrogram, neboli grafický záznam údajů ombrografu. Ombrograf umožňuje přesné kontinuální zaznamenávání srážkových úhrnů. Od roku 1997 jsou postupně mechanické přístroje nahrazovány automaty s digitálním zápisem. Jak uvádí například autoři práce [87], v roce 2003 měřilo v klimatologické síti ČHMÚ 165 mechanických ombrografů a 77 automatických srážkoměrů. Historické údaje z mechanických ombrografů ale tvoří většinu dostupných dat a pro potřeby moderního zpracování je nutné je digitalizovat. Digitalizace se provádí pomocí digitizéru odečítáním souřadnic zlomových bodů a výsledný záznam je ukládán do databáze s časovým rozlišením 1 minuty. Pro ko-

rekcí dat, jejich doplnění či rekonstrukci se používá celá řada podkladů, zejména výkazy měsíčních meteorologických pozorování na dané stanici, ale také měsíční výkazy a ombrografické záznamy z okolních stanic či údaje o srážkách ze synoptických a automatických stanic [87].



Obrázek 4.1: Poloha srážkoměrných stanic.

Jako vstupní data pro analýzu srážek, jíž je věnována tato kapitola, slouží digitalizované záznamy srážkových intenzit reálných dešťových srážek ze 6 srážkoměrných stanic situovaných na území Jihomoravského kraje (viz obrázek 4.1). Jedná se o stanice Brno-Tuřany, Brno-Žabovřesky a Brno-Jundrov, které se nacházejí v urbanizované oblasti města Brna, druhého největšího města České republiky a centra kraje. Dále jsou uvažovány stanice Znojmo-Kuchařovice a Jevišovice, které jsou umístěny v jihozápadní části kraje a stanice Vyškov v severozápadní části Jihomoravského kraje. Z důvodů neúplnosti historických měření byly u vybraných stanic vynechány některé roky záznamu, které zejména kvůli nedostatku doprovodných dat nebylo možné rekonstruovat. Sledovaná období spolu s délkami úplných záznamů jsou shrnuta v tabulce 4.1.

Tabulka 4.1: Přehled srážkoměrných stanic, jejich sledovaného období a délky řad s úplnými záznamy.

| Stanice | Sledované období | Délka dostupné řady [roky] |
|--------------------|------------------|----------------------------|
| Brno-Tuřany | 1959-2000 | 41 |
| Brno-Žabovřesky | 1987-2003 | 16 |
| Brno-Jundrov | 1992-2003 | 11 |
| Znojmo-Kuchařovice | 1956-2003 | 27 |
| Jevišovice | 1961-2000 | 37 |
| Vyškov | 1961-1992 | 31 |

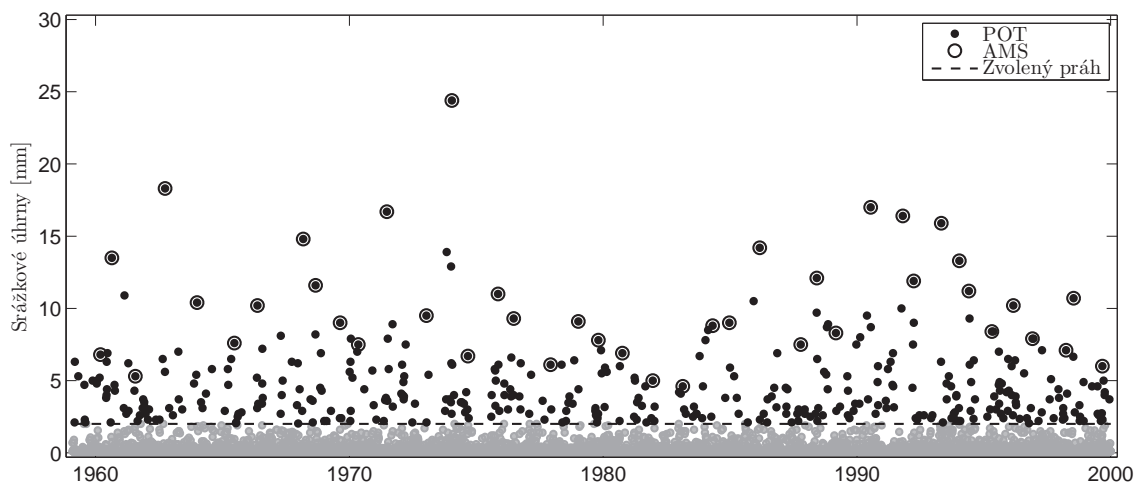
Teorie extrémních hodnot je při posuzování hydrologických jevů často aplikována. Zejména se jedná o situace, kdy se zajímáme o odhad frekvence vzácných či extrémních jevů jako je např. výskyt tzv. stoleté vody, padesátileté vody apod. Jak uvádí například práce [80], v zásadě se rozlišují dva základní způsoby výběru prvků pro statistické zpracování. První je postaven na metodě blokových maxim (viz odstavec 2.1), kdy jsou do analyzovaného statistického souboru zahrnuty jen ty největší prvky v daném bloku pozorování. V hydrologii bývají většinou uvažovány bloky o délce jednoho roku a metoda blokových maxim je zde pak označována jako *Annual Maxima Series* (AMS) metoda. Volba délky bloku odpovídající jednomu roku má své opodstatnění. Délku bloku je potřeba zvolit dostatečně velkou, aby byla zaručena dobrá aproximace pomocí asymptotického modelu a odhadované parametry a parametrické funkce nebyly zatíženy velkým vychýlením. Naopak je-li délka bloku zvolena příliš velká, dochází k nárůstu variability odhadů. Zde se bloky o délce jednoho roku ukazují jako velmi vhodné. Výběrem jednoročního maxima jsou také eliminovány nestacionarity dešťových srážek v průběhu roku. Navíc měření dešťových srážek s rozlišením 1 minuty zřejmě nelze pokládat za nezávislá pozorování a tedy pro analýzu extrémních hodnot pomocí EV teorie založené na nezávislých veličinách je nutné z dostupných měření vybrat taková pozorování, která můžeme pokládat za nezávislá. V případě AMS modelu jsou vybraná roční maxima obecně považována za přibližně nezávislá měření [80] a odpadá tak nutnost užití dalších vzorkovacích technik. Teorii extrémních hodnot pro závislé výběry bude více věnována kapitola 5, kde představíme příslušné techniky odhadu dané závislosti. Metoda AMS byla v oblasti hydrologických extrémů často používána, o čemž svědčí například publikace [17, 136, 105]. Významnou nevýhodou AMS je ovšem fakt, že jsou zanedbávány případné další extrémní hodnoty, které ovšem nedosahují příslušných blokových maxim. Protože je tak značně redukována obsažená informace, vyžaduje použití AMS dostupnost dostatečně dlouhých časových řad.

Jiný způsob výběru prvků je založen na prahovém POT modelu (viz odstavec 2.1), v hydrologii označovaným také často jako *Partial Duration Series*. Do statistického souboru jsou zahrnuty hodnoty překračující nějakou dostatečně vysokou prahovou hodnotu. Oproti metodě AMS dovoluje POT model zpravidla zahrnout do statistického souboru více pozorování během roku a v současnosti tak představuje stále více používanou techniku, což dokládá řada prací ([20, 19, 25, 134, 84, 98, 108, 114]). Ukazuje se také, že POT lépe odpovídá rozdělením s těžkými chvosty [97]. Řada autorů se také zabývala srovnáním obou metodik AMS a POT, můžeme zmínit například práce [35, 95, 97, 125, 122]. Nevýhodou prahového modelu ovšem je, že výběr prvků do statistického souboru není jasně stanoven a v případě, že datový soubor nepředstavuje pozorování nezávislých veličin, je nutné použít dodatečné vzorkovací techniky. V kapitole 5 bude ukázáno, že v praktických aplikačních problémech se porušení dosud uvažovaných předpokladů nejčastěji vyskytuje ve formě jakési krátkodobé závislosti, tj. pozorování můžeme uvažovat za přibližně nezávislá jen je-li mezi nimi dostatečně dlouhý časový rozestup. Tato vlastnost se pak v POT modelu promítá ve formě shlukování extrémních hodnot.

Při analýze extrémů pomocí přístupu POT se v oblasti atmosférických srážek v první fázi tedy postupuje tak, že se z měřených časových řad separují tzv. *dešťové události*. Tyto jsou mezi sebou uvažovány jako přibližně nezávislé. Následně se z každé události pro předem stanovené trvání deště vyberou největší pozorované srážkové úhrny, tj. největší klouzavé součty úhrnů s oknem odpovídajícím danému trvání deště [96]. Pro tyto je dále možné postupovat pomocí teorie nezávislých veličin. Není však jednoznačně dáno, jak dešťovou událost definovat. Různí autoři ke vzorkování srážkového POT modelu přistupují

různě. Například v [19] byly dešťové události odděleny ustáním deště po dobu alespoň 24 hodin. V [71] separovali dvě dešťové události pomocí alespoň hodinové bezsrážkové periody odpovídající době, která je nutná k dotoku nejdlejší dánské stokové sítě. Další autoři uvažují proměnlivá kritéria v závislosti na délce trvání deště. Autoři práce [129] separovali dvě dešťové události pomocí snížení srážkových úhrnů po dobu alespoň 8–24 hodin v závislosti na délce trvání deště. Podobný přístup byl aplikován v [96], resp. ve [134], kde byly dešťové události odděleny bezsrážkových obdobím o délce alespoň jedné hodiny, resp. 12 hodin, pro déle trvající deště však byla použita separační perioda odpovídající době trvání deště.

Aplikace různých metodik vzorkování nezávislých pozorování závisí na studované problematice a konkrétní aplikační oblasti. Jak naznačuje práce [99], právě přístup separace uvedený v [96] se ukazuje jako nejvhodnější pro potřeby návrhu městského odvodnění v brněnském regionu. Vzhledem k délce dostupných časových řad lze metodu AMS aplikovat jen se značnými omezeními, neboť dochází k významné redukci datového souboru. Pozorované řady ze stanic Brno-Tuřany či Jevišovice lze považovat za vhodné pro použití AMS, avšak měření z některých dalších stanic nejsou dostatečně rozsáhlá. Jedná se zejména o stanice Brno-Jundrov a Brno-Žabovřesky, kde by AMS metoda pro dostupné dešťové řady byla zcela nevhodná (viz délky dostupných řad v tabulce 4.1). Z tohoto důvodu se dále soustředíme výhradně na použití prahového POT modelu, který zejména v případě krátkých časových řad dovoluje učinit přesnější závěry o charakteristikách extrémních srážek. Obrázek 4.2 ilustruje rozdíly složení výběrů uvažovaných při použití AMS a POT modelů pro stanici Brno-Tuřany a srážkové úhrny pro 10minutové deště, přičemž tyto byly nejprve vzorkovány podle metodiky [96].



Obrázek 4.2: Demonstrace rozdílů složení datových souborů srážkových úhrnů při použití metod AMS a POT; zde pro stanici Brno-Tuřany při délce trvání deště 10 minut. Vyznačena je prahová hodnota $u = 2$ mm, horizontální osa vyznačuje sledované období.

Ve spojitosti s analýzou srážkových měření byla provedena řada počítačových implementací. Výše představenou techniku vzorkování podle [96] lze provést pomocí funkce `ems.m`. Tato je navržena pro získání přibližně nezávislých agregovaných pozorování (např. zde diskutovaných srážkových úhrnů). Vzorkování pomocí blokových maxim i vzorkování pomocí prahového POT modelu je obecněji věnován odstavce 5.1. Pro tyto účely je pak možné využít funkci `decluster.m`, která je vhodná pro získání přibližně nezávislých, ne-

agregovaných (tj. původních) měření. Více detailů lze nalézt v odstavci D, resp. přímo ve zdrojových kódech funkcí.

4.2. Simulační studie pro volbu prahu adaptivními technikami

Zásadním krokem při použití POT modelu je vhodná volba prahové hodnoty u . V kapitole 3 byly zmíněny dva adaptivní přístupy volby prahu, dvojitě bootstrapová semiparametrická metoda a MT-GP metoda založená na stabilitě EV indexu. Tyto lze s úspěchem použít i pro analýzu dešťových srážek. Použití obou metodik pro odhad frekvence srážkových extrémů byla diskutována v [10], kde byly oba přístupy aplikovány k odhadům IDF křivek (viz odstavec 4.3). V této části se především zaměříme na výsledky simulační studie [4], jež porovnává vlastnosti zmiňovaných technik.

Pro simulační studii byla vybrána modelová rozdělení pravděpodobnosti, která svou povahou nejlépe odpovídají uvažovaným pozorováním srážek z oblasti brněnského regionu. Vzhledem k analýze dat v následujícím odstavci 4.3, je zejména zajímavé studovat vlastnosti adaptivních metod výběru prahu v závislosti na délce dostupných pozorovaných řad. Jako referenční stanice byla vybrána stanice Brno-Jundrov a generovány náhodné výběry z rozdělení, které svou povahou odpovídají právě situaci na této stanici. Vhodná modelová rozdělení byla vybrána z široké řady spojitých rozdělení s kladným nosičem, přičemž jako hledisko při posuzování vhodnosti modelu bylo zvoleno Akaikeho informační kritérium AIC a bayesovské informační kritérium BIC. Obě tyto statistiky vyhodnocují jak kvalitu proložení, tak složitost modelového rozdělení. Uvažujme modelové rozdělení s hustotou $f_0(x, \boldsymbol{\theta})$, kde $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)$ je vektor parametrů. Pro náhodný výběr X_1, \dots, X_n z modelového rozdělení označme $L_0(\boldsymbol{\theta})$ příslušnou věrohodnostní funkci a $\hat{\boldsymbol{\theta}}$ MV odhad parametru $\boldsymbol{\theta}$. Daná kritéria jsou vyjádřena vztahy

$$\begin{aligned} \text{AIC} &= 2m - 2 \ln L_0(\hat{\boldsymbol{\theta}}), \\ \text{BIC} &= m \ln n - 2 \ln L_0(\hat{\boldsymbol{\theta}}). \end{aligned}$$

Při porovnání dvou modelových rozdělení pomocí výše uvedených kritérií je tak preferováno to, pro které je hodnota statistiky AIC, resp. BIC, nižší. Mimo uvedených kritérií lze také uvažovat jen maximum věrohodnostní funkce L_0 , udávané nejčastěji ve tvaru záporné logaritmické věrohodnostní funkce $-\ln L_0(\hat{\boldsymbol{\theta}})$, přičemž jako vhodnější se pak opět ukazuje to rozdělení, které má tuto hodnotu nižší. Je třeba mít na paměti, že všechna uvedená kritéria udávají pouze relativní vhodnost modelu a absolutní vhodnost modelu tak není zaručena.

V tabulce 4.2 jsou shrnuty hodnoty AIC a BIC kritérií pro řadu modelových rozdělení pro referenční stanici Brno-Jundrov při trvání deště 180 minut. Vzhledem k uvedeným výsledkům se z testovaných rozdělení zaměříme pouze na vybrané modely rozdělení. Konkrétně budeme uvažovat rozdělení gama, Fréchetovo, GP, lognormální, loglogistické a Weibullovo (pro tvary parametrizací odkazujeme na tabulky 1.1, 1.2), které jsou uvedenými kritérii označeny jako nejvhodnější. Vhodnost uvažovaných rozdělení byla ověřena testy dobré shody, konkrétně Pearsonovým χ^2 testem [14], Kolmogorovovým-Smirnovým testem (K-S testem) [16] a Andersonovým-Darlingovým testem (A-D testem) [36].

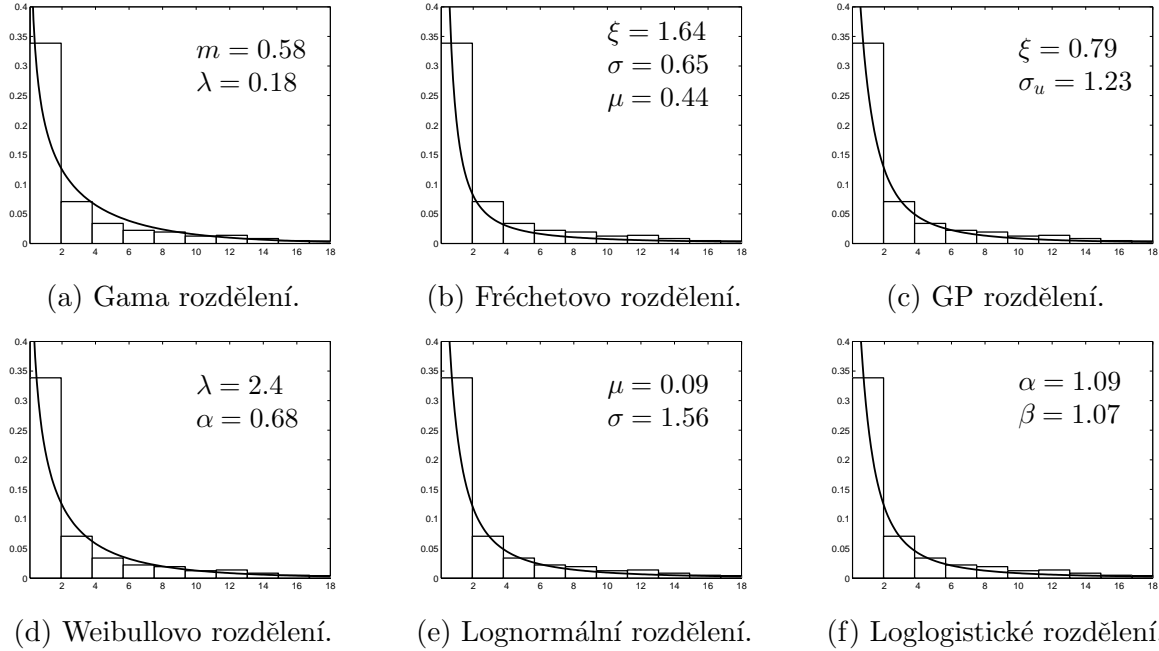
Tabulka 4.2: Hodnoty kritérií AIC, BIC a $-\ln L(\hat{\theta})$ pro modelová rozdělení srážkových dat ze stanice Brno-Jundrov s trváním deště 180 minut.

| Rozdělení | AIC | BIC | $-\ln L_0(\hat{\theta})$ |
|------------------|--------|---------|--------------------------|
| Zobecněné Pareto | 2956,9 | 2970,8 | 1475,4 |
| Lognormální | 3053,8 | 3063,1 | 1524,9 |
| GEV | 3089,3 | 3103,3 | 1541,6 |
| Loglogistické | 3111,9 | 3121,2 | 1553,9 |
| Weibullovo | 3137,1 | 3146,4 | 1566,5 |
| Gama | 3189,3 | 3198,6 | 1592,7 |
| Nakagamiho | 3324,0 | 3333,3 | 1660,0 |
| Exponenciální | 3387,2 | 3391,8 | 1692,6 |
| Logistické | 4471,6 | 4481,0 | 2233,8 |
| Rayleighovo | 5980,3 | 5984,9 | 2989,1 |
| Riceovo | 5982,3 | 55991,6 | 2989,1 |

Na obrázku 4.3 je pak ukázáno proložení histogramů odhady hustot pravděpodobností pro uvažovanou stanici. Fréchetovo, GP a loglogistické rozdělení patří do Fréchetovy třídy EV rozdělení a všechna tato rozdělení splňují předpoklady semiparametrického dvojité bootstrapového modelu, konkrétně patří do Hallovy třídy rozdělení (viz odvození pomocných funkcí v dodatku A). Další uvažovaná rozdělení patří do Gumbelova oboru, přičemž navíc parametr druhého řádu je pro gama a Weibullovo rozdělení roven nule. Jsou tak porušeny základní předpoklady semiparametrického modelu výběru prahu (3.2) na tvar pomocné funkce druhého řádu. V těchto případech je přesto možné sledovat chování dvojité bootstrapové metody volby prahu a porovnat její ne-/vhodnost ve srovnání s MT-GP přístupem, neboť v praktických situacích lze předpoklady semiparametrického modelu ověřit jen se značnými obtížemi.

V dalších odstavcích se soustředíme především na popis situace a odhadů pro dva konkrétní datové soubory, a sice pro případ stanice Brno-Tuřany a Brno-Jundrov, které představují v naší analýze nejdelší a nejkratší dostupné pozorované časové řady. Zejména pak bude vhodné se zajímat o odlišnosti různých přístupů k odhadům parametrických funkcí. Protože rozsahy vzorkovaných nezávislých pozorování se pro uvažované délky trvání dešťů (viz analýza extrémních srážek dále v odstavci 4.3) pohybují od 1836 do 10791 pro stanici Brno-Tuřany a v rozmezí od 618 do 3228 pro stanici Brno-Jundrov, byly simulovány náhodné výběry z výše uvedených rozdělení o rozsazích řádově odpovídajícím těmto hodnotám. Konkrétně byly generovány náhodné výběry o rozsazích $n = 500, 1000, 2000, 3000, 5000$ a 8000 .

Pro metodu MT-GP se omezíme jen na použití statistiky (3.16). Jak bylo zmíněno v odstavci 3.3, použití statistiky (3.17) vyžaduje odhady parametrů tvaru na všech prahových podintervalech, což pro simulační studii vede k neúnosným výpočetním časům. Z tohoto důvodu budou testy založené na statistice (3.17) použity jen pro analýzu konkrétních datových souborů srážkových úhrnů v následujícím odstavci. Pro případ testo-



Obrázek 4.3: Histogramy srážkových úhrnů pro 180minutové deště ze stanice Brno-Jundrov proložené odhady hustot vybraných rozdělení včetně odhadů jejich parametrů při parametrizacích uvedených v tabulkách 1.1, 1.2.

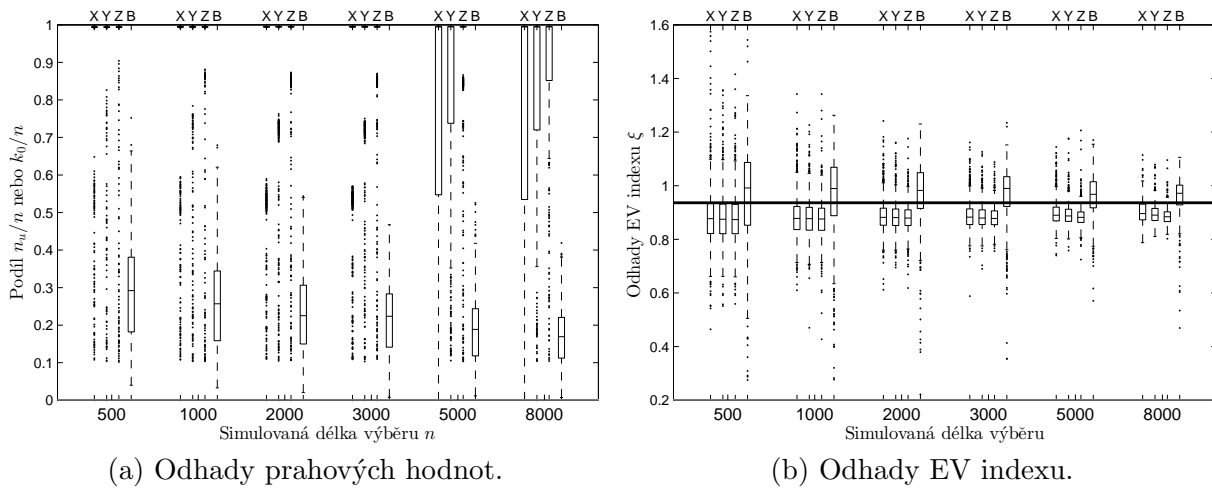
vacího kritéria (3.16) bylo zvoleno následující nastavení. Pro každý uvažovaný rozsah n bylo generováno $M = 1500$ náhodných výběrů z modelových rozdělení. Optimální prahová hodnota u_0 je hledána v intervalu mezi 0,5% a 90% empirickým kvantilem, tj. například pro $n = 1000$ mezi hodnotami $x_{(5)}$ a $x_{(900)}$. A priori tak předpokládáme, že nižší prahové hodnoty nejsou optimální ve smyslu stability EV indexu a pro práh větší než $x_{(900)}$ již je dostatečné stability dosaženo. V odstavci 3.3 věnovanému MT-GP modelu byla nastíněna problematika volby počtu m prahových hodnot a tím stupně diskretizace EV indexu. Prozatím provedené simulace v práci [102] však na tuto otázku nedávají jednoznačnou odpověď. Vysoká hodnota m dovoluje detailní studium stability EV indexu, má však za následek velkou variabilitu odhadů. V návaznosti na publikaci [102] tak v našem případě volíme tři různé úrovně diskretizace $m = 10, 20$ a 40 . Prahová hodnota u_0 je určena na základě asymptotického rozdělení statistiky (3.16), přičemž jako u_0 je zvolena nejmenší hodnota prahu u , pro kterou není testovaná nulová hypotéza zamítnuta na hladině významnosti $\alpha = 0,05$.

Odhady získané pomocí MT-GP modelu byly porovnány s dvojitě bootstrapovou metodou volby prahu. Pro účely této simulační studie byl zvolen dvojitě bootstrapový přístup studovaný Draismou a kol. [45], tedy při použití momentového odhadu EV indexu (2.32) v kombinaci se statistikou (3.8) jakožto pomocným odhadem $\hat{\xi}_{\text{aux}}$. Uvedená adaptivní technika odhadu k_0 horních pořadových statistik je, na rozdíl od dvojitě bootstrapové metody studované např. v [65], pro uvažovaná modelová rozdělení vhodnější. Technika popsaná v [65] vyžaduje stanovení počáteční hodnoty k_{aux} , která zásadně ovlivňuje odhad k_0 . Navíc použití Hillova odhadu v [65] klade pro hodnoty EV indexu značná omezení. Z důvodu potřeby značných výpočetních kapacit byl při použití adaptivní bootstrapové metody volby prahu počet generovaných výběrů omezen na $M = 500$, přičemž počet bootstrapových výběrů byl v každém opakování nastaven na $B = 250$. Ze studie [65] vyplývá, že rozsah bootstrapového výběru n_1 je vhodné volit co největší tak, aby $n_1 < n$. V našem

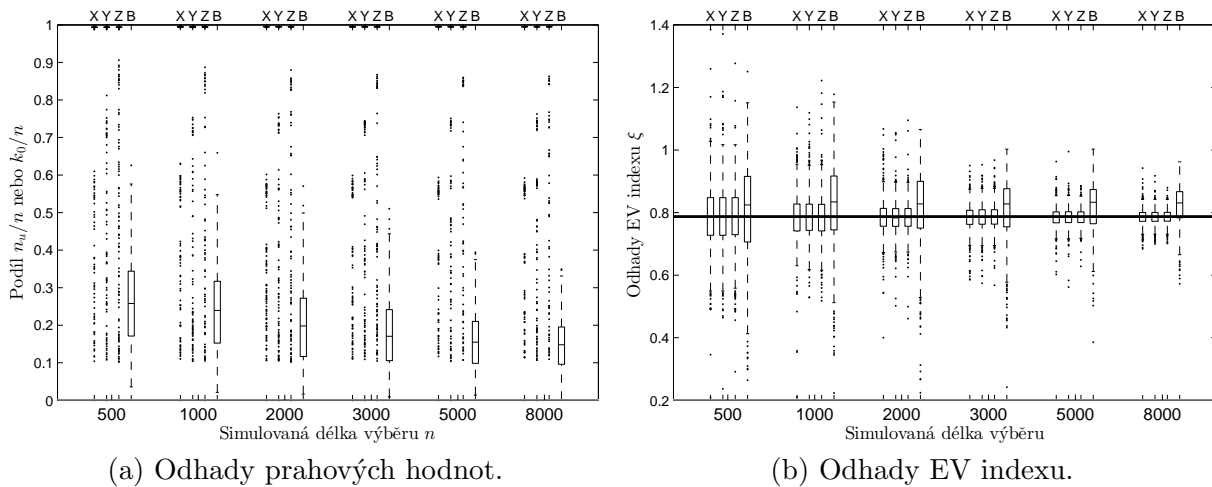
případě tak bylo určeno $n_1 = \lfloor n^{0,995} \rfloor$. Z podstaty techniky bootstrap může odhad \widehat{k}_0^* hodnoty k_0 při různých opakováních kolísat. Dle bodu (7) algoritmu na str. 61 je tak celý postup $N = 100$ krát nezávisle opakován pro každý generovaný výběr z modelového rozdělení, přičemž dle vztahu (3.7) tak byly získány odhady $\widehat{k}_{0,i}^*$, $i = 1, \dots, N$. Konečný odhad \widehat{k}_0 je pak stanoven jako

$$\widehat{k}_0 = \operatorname{median}_{i=1, \dots, N} \widehat{k}_{0,i}^*.$$

Na následujících obrázcích jsou formou box-plotů vizualizovány výsledky popsané simulační studie. Obrázky 4.4.a až 4.9.a ukazují vždy podíly nadprahových hodnot n_u/n (sloupce X, Y, Z), resp. podíly optimálního počtu horních pořadových statistik \widehat{k}_0/n (sloupec B), pro různá modelová rozdělení a uvažované délky simulovaných výběrů.



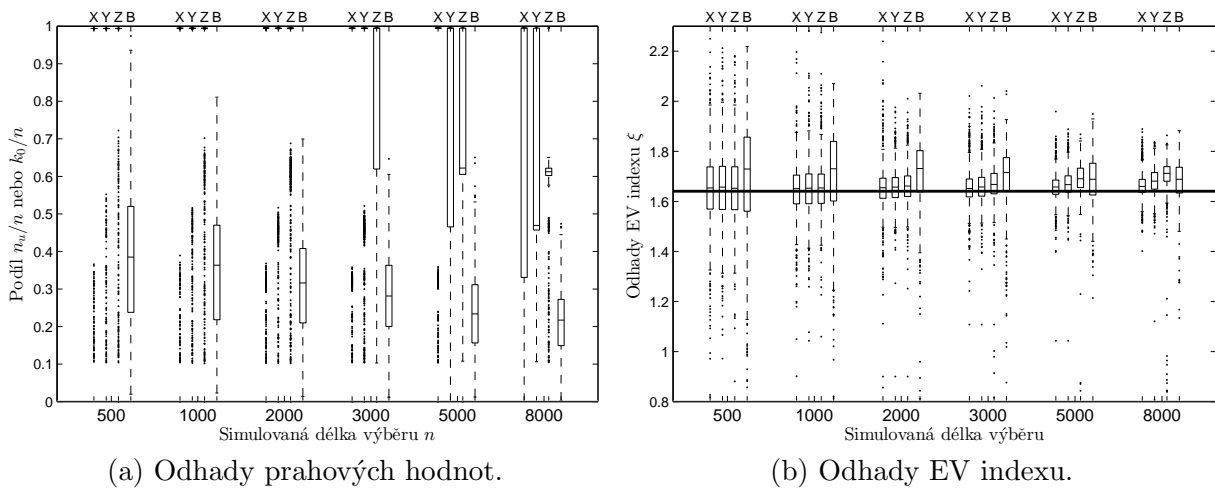
Obrázek 4.4: Loglogistické rozdělení: box-ploty podílů nadprahových hodnot n_u/n resp. \widehat{k}_0/n (vlevo) a box-ploty příslušných odhadů EV indexu (vpravo) pro různé rozsahy náhodných výběrů a volbu prahu získanou MT-GP metodou při $m = 10, 20, 40$ (sloupce X, Y, Z) a semiparametrickou dvojitě bootstrapovou metodou (sloupec B).



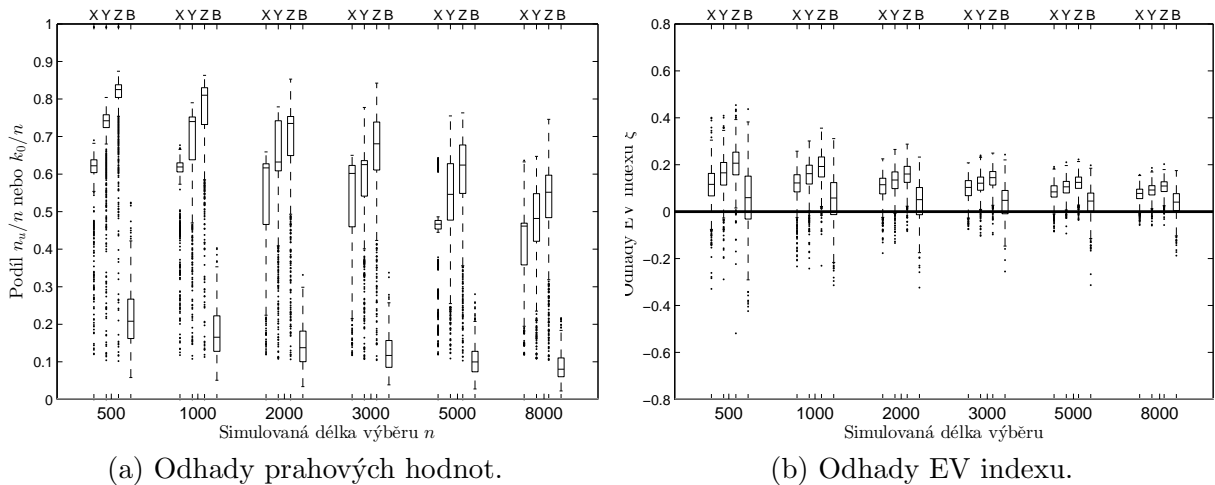
Obrázek 4.5: GP rozdělení: box-ploty podílů nadprahových hodnot n_u/n resp. \widehat{k}_0/n (vlevo) a box-ploty příslušných odhadů EV indexu (vpravo) pro různé rozsahy náhodných výběrů a volbu prahu získanou MT-GP metodou při $m = 10, 20, 40$ (sloupce X, Y, Z) a semiparametrickou dvojitě bootstrapovou metodou (sloupec B).

Na obrázcích 4.4.b až 4.9.b jsou pak zobrazeny odpovídající odhady EV indexu pomocí MV metody (pro práh u určený pomocí MT-GP modelu - sloupce X, Y, Z) a pomocí momentového odhadu (pro semiparametrický model - sloupec B). Vyznačená vodorovná úsečka vyznačuje skutečnou hodnotu EV indexu.

Je vidět, že mediány hodnot \hat{k}_0 , tj. optimálních hodnot horních pořadových statistik, jsou ve všech studovaných případech menší než mediány hodnot n_u pro libovolné m . Jistá nestabilita MT-GP metodiky v odhadu prahu je pozorovatelná pro střední rozsahy výběrů z lognormálního a Fréchetova rozdělení (obrázky 4.8.a, 4.6.a) a pro velké rozsahy výběrů z loglogistického rozdělení (obrázek 4.4.a). V těchto případech kolísají p-hodnoty testovacího kritéria (3.16) blízko hladiny významnosti $\alpha = 0,05$. Autoři práce [102] zmiňují v podobných situacích možnou modifikaci volby vhodného prahu tak, aby bylo dosaženo určité stability odpovídajících p-hodnot. Lze například požadovat, aby pro prahy $u \geq u_0$

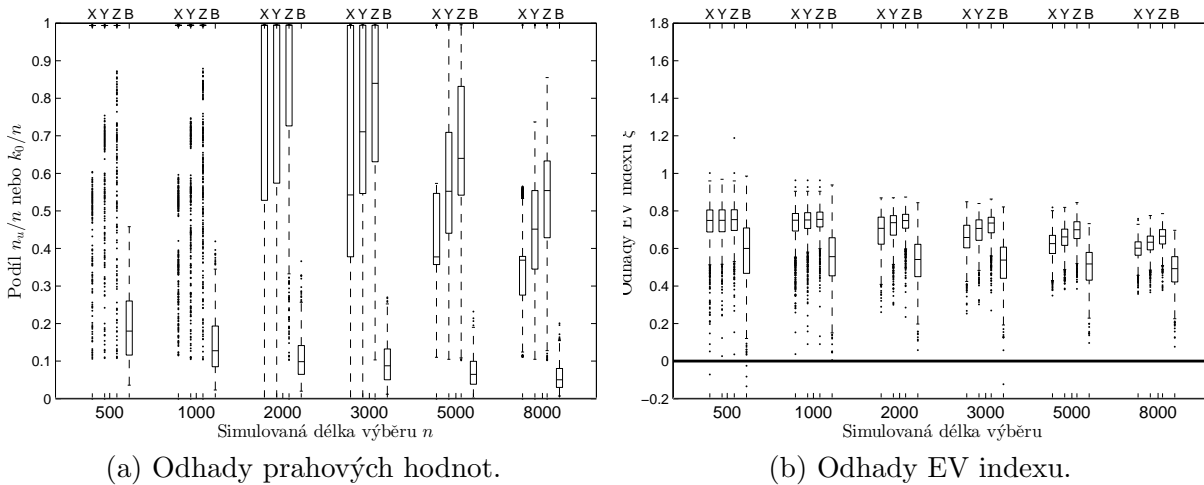


Obrázek 4.6: Fréchetovo rozdělení: box-ploty podílů nadprahových hodnot n_u/n resp. \hat{k}_0/n (vlevo) a box-ploty příslušných odhadů EV indexu (vpravo) pro různé rozsahy náhodných výběrů a volbu prahu získanou MT-GP metodou při $m = 10, 20, 40$ (sloupce X, Y, Z) a semiparametrickou dvojitě bootstrapovou metodou (sloupec B).

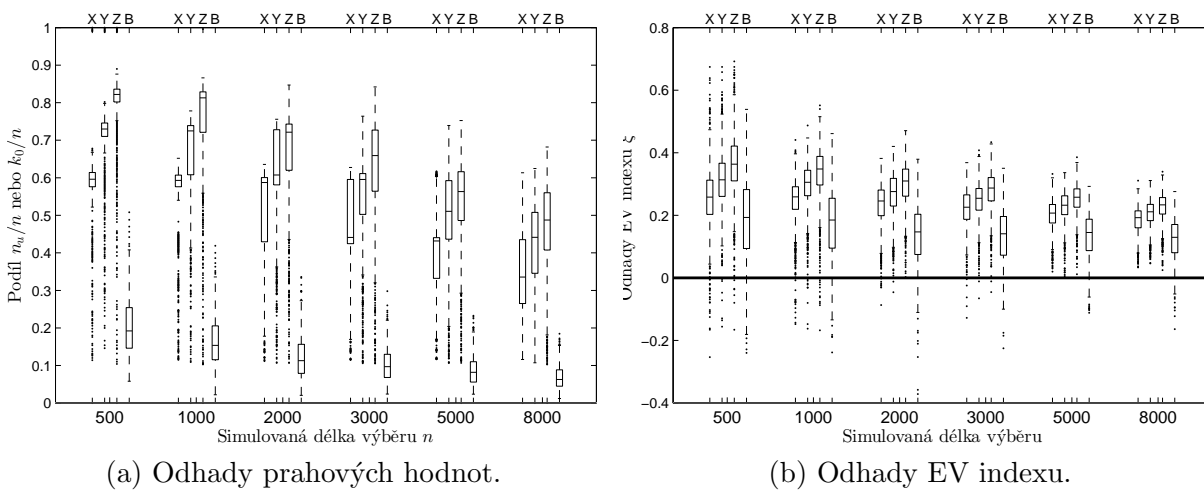


Obrázek 4.7: Gama rozdělení: box-ploty podílů nadprahových hodnot n_u/n resp. \hat{k}_0/n (vlevo) a box-ploty příslušných odhadů EV indexu (vpravo) pro různé rozsahy náhodných výběrů a volbu prahu získanou MT-GP metodou při $m = 10, 20, 40$ (sloupce X, Y, Z) a semiparametrickou dvojitě bootstrapovou metodou (sloupec B).

tyto hodnoty neklesly pod nějakou stanovenou úroveň, řekněme kupříkladu 2α . Naneštěstí tento přístup MT-GP techniku značně komplikuje, neboť mimo nejasné úrovně diskretizace m je tak do modelu vnášen další pomocný parametr ovlivňující volbu prahu u_0 . Z obrázku 4.5.a je patrné, že adaptivní volba prahu při použití MT-GP modelu správně identifikuje optimální prahové hodnoty pro výběry z GP rozdělení a hodnoty n_u se pohybují blízko n . Podobná situace nastává i pro loglogistické rozdělení (obrázek 4.4.a), kde parametr tvaru α tohoto rozdělení byl odhadnut blízko jedné (obrázek 4.3), přičemž v takovém případě loglogistické rozdělení splývá s GP rozdělením. Závislost odhadu vhodné prahové hodnoty u_0 na úrovni diskretizace m je patrná zejména na obrázku 4.7.a či 4.9.a pro gama a Weibullovo rozdělení. Situace tak odpovídá výsledkům práce [102], kdy při velkém m jsou rozdíly mezi jednotlivými prahy detekovány s menší pravděpodobností.



Obrázek 4.8: Lognormální rozdělení: box-ploty podílů nadprahových hodnot n_u/n resp. \hat{k}_0/n (vlevo) a box-ploty příslušných odhadů EV indexu (vpravo) pro různé rozsahy náhodných výběrů a volbu prahu získanou MT-GP metodou při $m = 10, 20, 40$ (sloupce X, Y, Z) a semiparametrickou dvojitě bootstrapovou metodou (sloupec B).



Obrázek 4.9: Weibullovo rozdělení: box-ploty podílů nadprahových hodnot n_u/n resp. \hat{k}_0/n (vlevo) a box-ploty příslušných odhadů EV indexu (vpravo) pro různé rozsahy náhodných výběrů a volbu prahu získanou MT-GP metodou při $m = 10, 20, 40$ (sloupce X, Y, Z) a semiparametrickou dvojitě bootstrapovou metodou (sloupec B).

Oba studované adaptivní přístupy volby prahu využívají zcela jiné principy odhadu, a jejich srovnání proto bude vhodnější provést skrze odhady EV indexů. Kvalitní odhady prahu pomocí metodiky MT-GP v případě GP rozdělení vedou k odhadům EV indexu, které lze z praktického hlediska pokládat za nevychýlené. Naproti tomu odhady získané metodou dvojitého bootstrapování vykazují v tomto směru horší výsledky. Navíc jsou odhady EV indexu zatíženy větší variabilitou. V případě dalších rozdělení náležících do Hallovy třídy jsou odhady získané oběma přístupy srovnatelné co se týká vychýlení i variability, vyjma Fréchetova rozdělení při malých až středně velkých rozsazích ($n \leq 3000$), kde jsou odhady získané metodou MT-GP méně vychýlené. Pro rozdělení nepatřící do Hallovy třídy, tj. rozdělení gama, lognormální a Weibullovo, je variabilita srovnatelná, zejména při větších rozsazích simulovaných výběrů. Naopak, ačkoli zde dvojité bootstrapová metoda nemá teoretickou oporu, odhady EV indexu získané tímto přístupem se ukazují jako méně vychýlené.

4.3. Analýza srážek pomocí jednorozměrné teorie nezávislých veličin

Odhady parametrů prahového POT modelu

Nyní přikročíme k analýze konkrétních datových souborů srážkových úhrnů v jihomoravském regionu pomocí přístupu POT. Dostupné časové řady byly nejprve dle metodiky [96] rozděleny na přibližně nezávislé dešťové události, přičemž z těchto byly dále vyčleněny největší srážkové úhrny pro trvání deště odpovídající době 5, 10, 15, 20, 30, 45, 60, 90, 120, 180, 240 a 360 minut. Takto bylo získáno celkově 12 datových souborů, pro které je možné dále použít přístup POT bez dalších omezení, tj. jednotlivé soubory představují přibližně nezávislá pozorování, přičemž lze přímo aplikovat metodiku představenou v odstavci 2.1. Z vyvzorkovaných srážkových úhrnů je možné určit intenzity dešťů jako množství srážek, které za danou dobu trvání deště dopadne na jednotkovou plochu. Tyto budou dále použity při odhadech IDF křivek níže.

Vhodná prahová hodnota POT modelu bude určena technikami diskutovanými v kapitole 3. Při odhadech limitního GP rozdělení překročení prahu parametrickou MV metodou bude prahová hodnota zjištěna nejprve grafickými technikami. Následně bude aplikována adaptivní technika MT-GP představená v odstavci 3.3. Nakonec budeme uvažovat semi-parametrický POT model, přičemž optimální počet horních pořadových statistik bude pro momentový odhad určen technikou dvojitého bootstrapování diskutovanou v odstavci 3.2.

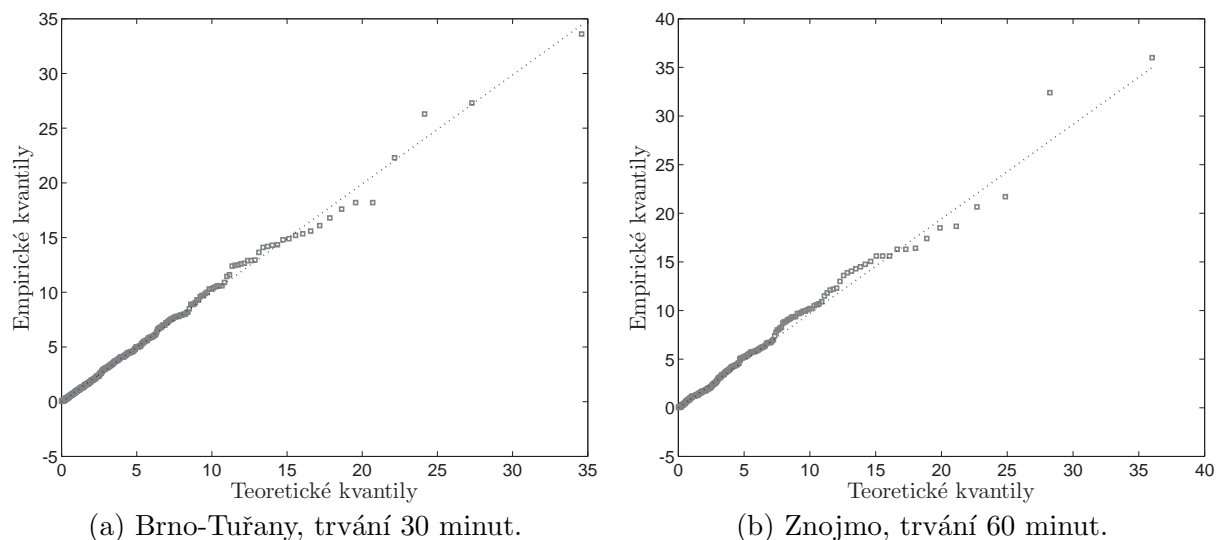
V tabulce 4.3 jsou uvedeny optimální prahové hodnoty, které byly stanoveny grafickou analýzou MRL plotu a posouzením stability MV odhadů. Prahové hodnoty bylo možné z průběhů grafů odečíst v naprosté většině případů. V některých jednotlivých situacích, kdy použité grafické metody nevedly k jednoznačné identifikaci prahu, bylo přihlédnuto k výsledkům u datových souborů se srovnatelným trváním deště. Vhodnost vybraných prahů lze pozorovat např. pro stanici Brno-Tuřany na obrázcích 3.1 a 3.2, kde při době trvání deště 10 minut byl zvolen práh $u_0 = 3,9$. Tato hodnota u_0 odpovídá teoretickým požadavkům, tedy MRL plot a graf závislosti parametru měřítka σ_u vykazují přibližně lineární trend a parametr tvaru ξ je přibližně konstantní pro hodnoty prahu $u \geq u_0$.

Pro všechny uvažované stanice a doby trvání deště je dále třeba vyšetřit vhodnost stanoveného POT modelu, tedy zda lze pozorované hodnoty nad prahem popsat pomocí

Tabulka 4.3: Optimální prahové hodnoty u_0 [mm] získané grafickými technikami.

| Stanice \ Trvání deště [min] | 5 | 10 | 15 | 20 | 30 | 45 | 60 | 90 | 120 | 180 | 240 | 360 |
|------------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Brno-Tuřany | 1,4 | 1,5 | 2,2 | 2,8 | 3,9 | 4,5 | 5,1 | 6,4 | 7,9 | 8,1 | 8,5 | 8,8 |
| Brno-Žabovřesky | 1,8 | 2,7 | 3,5 | 4,8 | 5,4 | 6,0 | 6,5 | 6,9 | 7,1 | 7,4 | 7,8 | 9,4 |
| Brno-Jundrov | 1,9 | 2,0 | 3,0 | 3,3 | 4,0 | 4,9 | 5,0 | 5,3 | 6,9 | 7,2 | 8,1 | 8,5 |
| Znojmo | 1,6 | 2,2 | 2,6 | 2,7 | 3,4 | 4,2 | 4,5 | 5,2 | 6,8 | 7,2 | 8,0 | 8,8 |
| Jevišovice | 2,0 | 2,8 | 3,3 | 3,7 | 4,5 | 5,7 | 6,1 | 6,4 | 6,9 | 7,1 | 7,4 | 9,1 |
| Vyškov | 1,3 | 1,3 | 1,8 | 2,2 | 3,0 | 3,2 | 4,1 | 5,0 | 5,5 | 6,2 | 7,0 | 8,2 |

GP rozdělení. Správnost zvolených prahových hodnot byla ověřena testy dobré shody, konkrétně byl použit Pearsonův χ^2 test, K-S test a A-D test, který je v hydrologii obzvláště doporučován. Žádný z uvedených testů nezamítl vhodnost GP rozdělení jakožto modelového rozdělení na hladině významnosti $\alpha = 0,05$. Kritické hodnoty A-D testu byly převzaty z publikace [79], detailní popis výsledků lze pak nalézt v práci [1]. Shoda empirického rozdělení s GP rozdělením byla také ověřena vizuálně s využitím Q-Q plotů (viz obrázek 4.10) a porovnáním histogramů datových souborů s hustotami příslušných GP rozdělení. Ukázalo se tak, že GP rozdělení velmi dobře odpovídá empirickému rozdělení nadprahových hodnot ve všech studovaných případech a další analýza tak bude vycházet právě z tohoto modelového rozdělení.

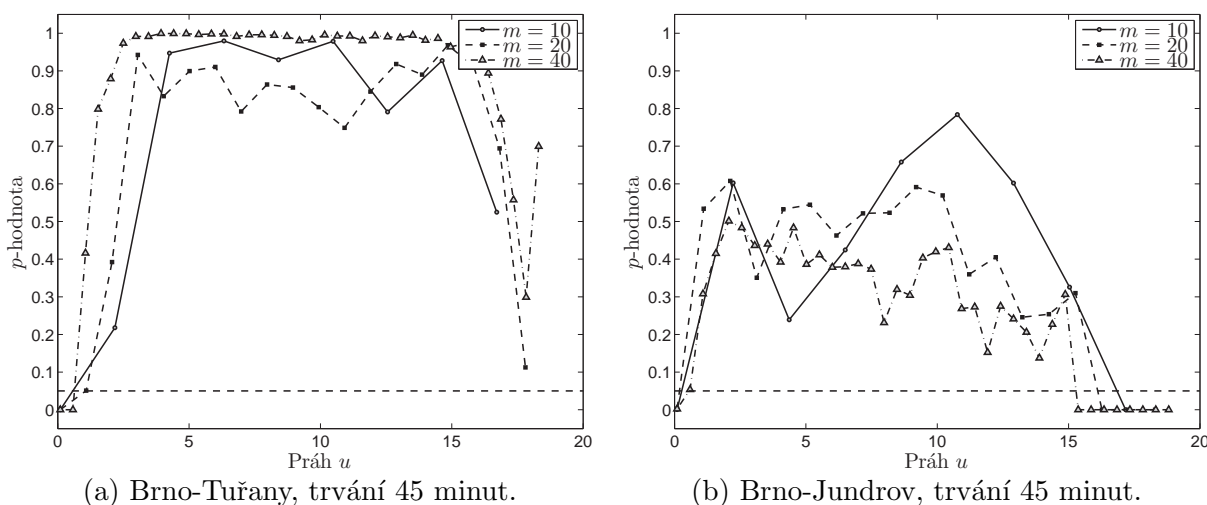
**Obrázek 4.10:** Q-Q plot empirických kvantilů nadprahových hodnot a teoretických GP kvantilů pro stanice Brno-Tuřany a Znojmo-Kuchařovice při trvání deště 30 a 60 minut.

Dále byly vhodné prahové hodnoty stanoveny pomocí adaptivního přístupu MT-GP [1]. Ze simulační studie z odstavce 4.2 vyplývá, že při použití testovacího kritéria založeného na skórovém vektoru (3.16) má sice volba stupně diskretizace m vliv na identifikaci optimální prahové hodnoty u_0 , avšak příslušné odhady EV indexu jsou pro uvažované hodnoty m ve většině případů srovnatelné. Přesto byl vzhledem k diskusi v práci [102] zvolen následující „defenzivní“ přístup. Pro každou stanici a dobu trvání deště bylo vždy $u_0(m)$ určeno pro diskretizaci pomocí $m = 10, 20, 40$ prahových hodnot mezi 10% a 99,5%

empirickými kvantily, přičemž pro uvedené hodnoty m bylo $u_0(m)$ určeno jako nejnižší práh, pro který testovaná hypotéza nebyla zamítnuta na hladině významnosti $\alpha = 0,05$. Na závěr byla optimální prahová hodnota u_0 určena jako nejvyšší z těchto hodnot, tj. $u_0 = \max_m \{u_0(m)\}$. Tyto prahy jsou pro vybrané stanice Brno-Tuřany a Brno-Jundrov shrnuty v tabulce 4.4.

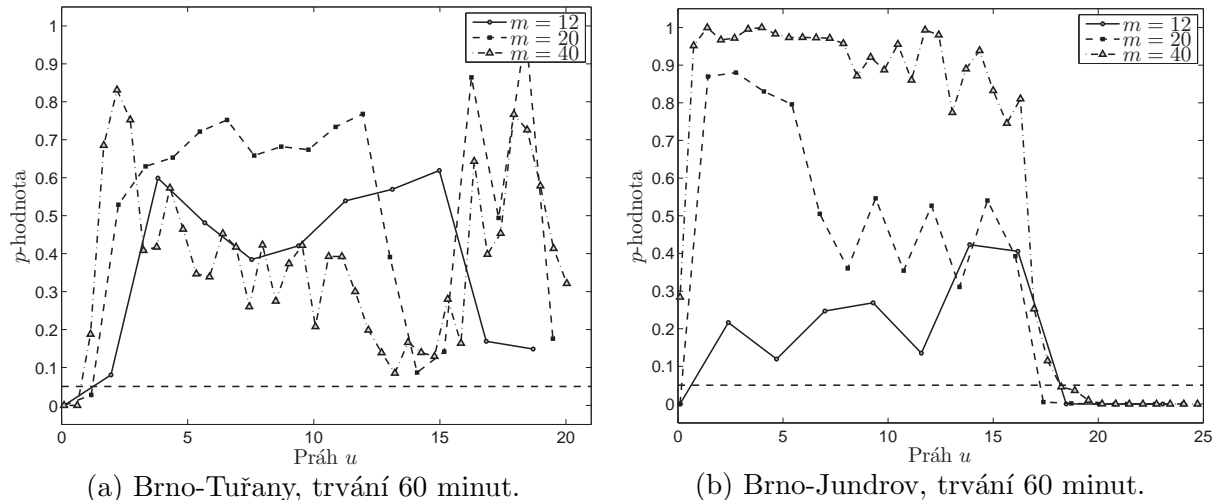
Tabulka 4.4: Optimální prahové hodnoty u_0 [mm] určené metodou MT-GP pro stanice Brno-Tuřany a Brno-Jundrov při použití skórové statistiky (LM,(3.16)) a poměru věrohodností (LR,(3.17)).

| Stanice \ Trvání deště [min] | | 5 | 10 | 15 | 20 | 30 | 45 | 60 | 90 | 120 | 180 | 240 | 360 |
|------------------------------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| Tuřany | (LM) | 1,19 | 1,40 | 2,48 | 2,10 | 1,91 | 2,18 | 2,37 | 2,70 | 3,04 | 3,54 | 3,88 | 4,30 |
| | (LR) | 0,99 | 1,60 | 2,04 | 2,40 | 3,07 | 3,50 | 3,82 | 4,36 | 4,91 | 5,60 | 6,10 | 6,71 |
| Jundrov | (LM) | 0,73 | 1,12 | 1,39 | 1,57 | 1,98 | 2,23 | 2,91 | 3,31 | 3,36 | 3,43 | 3,57 | 3,82 |
| | (LR) | 1,17 | 1,77 | 2,21 | 2,51 | 3,17 | 3,59 | 4,69 | 5,35 | 5,43 | 5,55 | 5,78 | 6,13 |



Obrázek 4.11: Vybrané průběhy p-hodnot při testování vhodnosti prahu pomocí skórové statistiky (3.16) při různých úrovních diskretizace $m = 10, 20$ a 40 . Horizontální přerušovaná čára vyznačuje hladinu významnosti $\alpha = 0,05$.

Na obrázku 4.11 je pro dva konkrétní případy ukázán typický výsledek při testování nulové hypotézy pomocí skórové statistiky (3.16), konkrétně je zde vizualizován průběh p-hodnot určených na základě limitního χ^2 rozdělení. Obdobně byla optimální prahová hodnota u_0 určena za použití testovacího kritéria založeného na poměru věrohodností (3.17), přičemž zde bylo zvoleno $m = 12, 20$ a 40 . Při volbě $m = 10$ byly zaznamenány problémy s konvergencí MV metody. Konkrétní výsledky lze nalézt v tabulce 4.4. Průběhy p-hodnot pro stanice Brno-Tuřany a Brno-Jundrov jsou zobrazeny na obrázku 4.12. Vhodnost prahových hodnot určených metodou MT-GP byla opět ověřena skrze shodu rozdělení nadprahových hodnot a modelového GP rozdělení pomocí testů dobré shody. Pro prahy určené skórovou statistikou (3.16) se shoda empirického a modelového rozdělení ukázala jako dobrá, přičemž žádný z použitých testů dobré shody vhodnost GP rozdělení nezamítl na hladině významnosti $\alpha = 0,05$. V případě prahů určených použitím statistiky



Obrázek 4.12: Vybrané průběhy p-hodnot při testování vhodnosti prahu pomocí poměru věrohodností (3.17) při různých úrovních diskretizace $m = 12, 20$ a 40 . Horizontální přerušovaná čára vyznačuje hladinu významnosti $\alpha = 0,05$.

(3.17) se ukázaly některé prahové hodnoty jako méně vhodné. Z popsaných výsledků se toto zejména týká krátkých dob trvání deště do 20 minut pro stanici Brno-Jundrov a dob trvání deště v rozmezí 30–60 minut pro stanici Brno-Tuřany. Výhodou adaptivního přístupu MT-GP je však především schopnost automatizovaného výběru prahu. Z tohoto důvodu budeme namísto úpravy vybraných prahových hodnot předpokládat jejich vhodnost ve všech případech a soustředíme se dále na vlastnosti odhadů parametrů modelového GP rozdělení.

V posledním případě byl uvažován semiparametrický POT model a optimální počet k_0 horních pořadových statistik byl určen adaptivní technikou dvojitého bootstrapování. Pro účely této analýzy byl vybrán přístup studovaný Draismou a kol. [45] založený na momentovém odhadu EV indexu (2.32). Na rozdíl od postupu navrženého např. v [65] je zřejmě splněn nutný předpoklad $x^* > 0$ a přitom postup dle [45] neklade další požadavky na příslušný parametr tvaru ξ . Bylo uvažováno následující nastavení výpočetního schématu. Počet bootstrapových výběrů byl stanoven $B = 250$, přičemž tato hodnota se dle literatury ukazuje jako dostačující [48]. Délka bootstrapových výběrů přitom byla na doporučení studie [65] stanovena $n_1 = \lfloor n^{0,995} \rfloor$, kde n je délka uvažované řady přibližně nezávislých srážkových úhrnů. Rozsah n_2 je určen bodem (5) dvojité bootstrapového algoritmu (str. 61). Dle bodu (7) byl postup algoritmu opakován $N = 1000$ krát, přičemž takto byly získány odhady $\hat{k}_{0,i}^*, i = 1, \dots, N$, hodnoty k_0 (viz obrázek 4.13). Konečný odhad \hat{k}_0 byl stanoven obdobně jako v odstavci 4.2, tedy

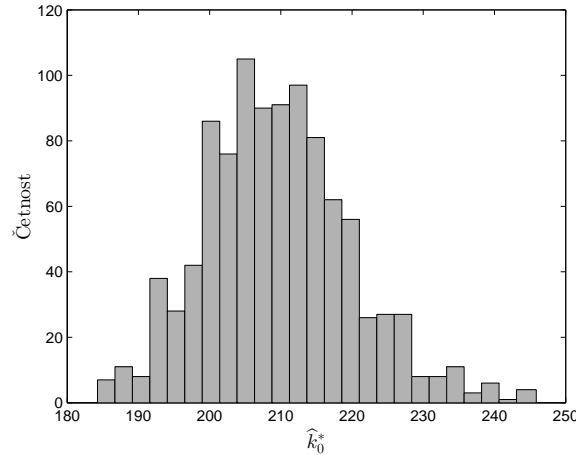
$$\hat{k}_0 = \text{median}_{i=1, \dots, N} \hat{k}_{0,i}^*.$$

Optimální počet horních pořadových statistik spolu s počty nadprahových hodnot určených předešlými přístupy lze pro vybrané stanice nalézt v tabulce 4.5. Nabízí se tak srovnání délek uvažovaných výběrů, které budou použity ke stanovení odhadů parametrů POT modelu. Lze pozorovat, že použití semiparametrického modelu vede k menšímu počtu nadprahových hodnot pro dlouhou časovou řadu ze stanice Tuřany. Při srovnání s metodou MT-GP jsou tyto rozdíly nejmarkantnější. V případě krátké časové řady ze stanice Jundrov jsou rozdíly méně výrazné. Při srovnání s grafickou volbou prahu je vidět,

Tabulka 4.5: Srovnání počtu nadprahových hodnot určených adaptivní dvojitě bootstrapovou metodou (\hat{k}_0), grafickými technikami (n_u^G) a adaptivní metodou MT-GP s použitím statistik LM (n_u^{LM}) a LR (n_u^{LR}).

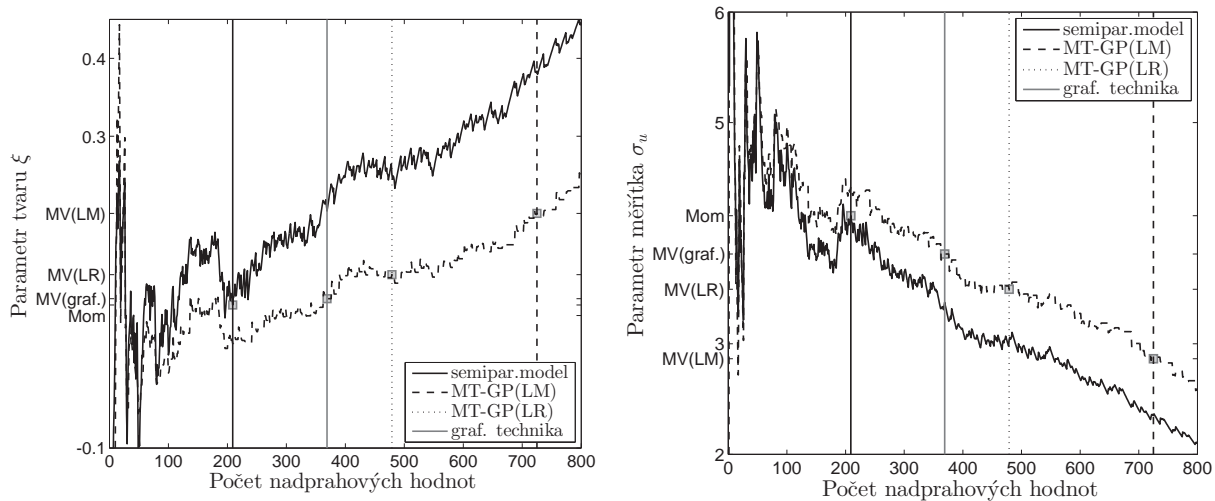
| Počet hodnot deště [min] | Trvání | 5 | 10 | 15 | 20 | 30 | 45 | 60 | 90 | 120 | 180 | 240 | 360 |
|-----------------------------|----------------------|------|------|------|------|------|------|------|------|------|------|------|------|
| Tuřany | \hat{k}_0 | 174 | 186 | 243 | 215 | 209 | 111 | 111 | 225 | 245 | 186 | 221 | 278 |
| | n_u^G | 352 | 506 | 458 | 423 | 369 | 376 | 369 | 329 | 271 | 309 | 318 | 334 |
| | n_u^{LM} | 423 | 540 | 418 | 556 | 725 | 776 | 804 | 807 | 777 | 740 | 725 | 689 |
| | n_u^{LR} | 487 | 471 | 499 | 498 | 479 | 496 | 507 | 535 | 516 | 500 | 481 | 482 |
| Jundrov | \hat{k}_0 | 56 | 41 | 44 | 50 | 68 | 70 | 83 | 105 | 90 | 103 | 115 | 120 |
| | n_u^G | 81 | 116 | 101 | 102 | 96 | 96 | 110 | 119 | 99 | 112 | 107 | 107 |
| | n_u^{LM} | 192 | 205 | 210 | 211 | 207 | 216 | 191 | 194 | 197 | 205 | 206 | 196 |
| | n_u^{LR} | 157 | 147 | 150 | 150 | 145 | 149 | 147 | 146 | 140 | 141 | 134 | 131 |
| Podíly počtu hodnot | | | | | | | | | | | | | |
| Tuřany | \hat{k}_0/n_u^G | 0,49 | 0,37 | 0,53 | 0,51 | 0,57 | 0,30 | 0,30 | 0,68 | 0,90 | 0,60 | 0,69 | 0,83 |
| | \hat{k}_0/n_u^{LM} | 0,41 | 0,34 | 0,58 | 0,39 | 0,29 | 0,14 | 0,14 | 0,29 | 0,32 | 0,25 | 0,30 | 0,40 |
| | \hat{k}_0/n_u^{LR} | 0,36 | 0,39 | 0,49 | 0,43 | 0,44 | 0,22 | 0,22 | 0,42 | 0,47 | 0,37 | 0,46 | 0,58 |
| Jundrov | \hat{k}_0/n_u^G | 0,69 | 0,35 | 0,44 | 0,49 | 0,71 | 0,73 | 0,75 | 0,88 | 0,91 | 0,92 | 1,07 | 1,12 |
| | \hat{k}_0/n_u^{LM} | 0,29 | 0,20 | 0,21 | 0,24 | 0,33 | 0,32 | 0,43 | 0,54 | 0,46 | 0,50 | 0,56 | 0,61 |
| | \hat{k}_0/n_u^{LR} | 0,36 | 0,28 | 0,29 | 0,33 | 0,48 | 0,47 | 0,56 | 0,72 | 0,64 | 0,73 | 0,86 | 0,92 |

že dvojitě bootstrapová technika volí podobnou prahovou hodnotu pro rostoucí trvání deště.



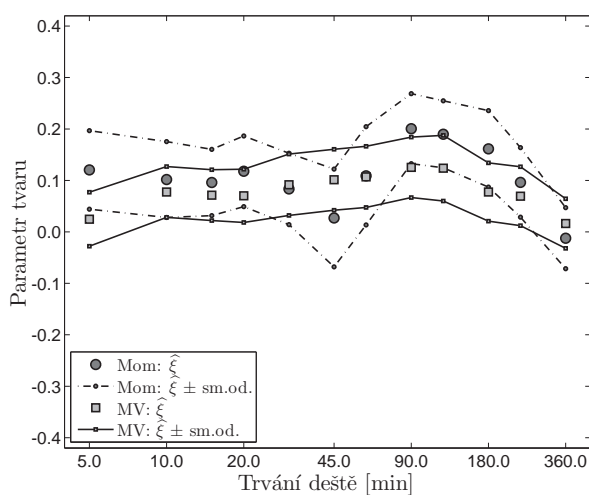
Obrázek 4.13: Histogram $N = 1000$ bootstrapových odhadů \hat{k}_0^* pro stanici Brno-Tuřany při trvání deště 30 minut.

Poté, co byly pro všechny datové soubory určeny optimální části vhodné k POT analýze, byly odhadnuty parametry příslušných POT modelů. Pro hodnoty prahu u_0 určené grafickými technikami a metodou MT-GP byly parametry limitního GP rozdělení odhadnuty MV metodou, zatímco v případě semiparametrického modelu byly pro dané \hat{k}_0 parametry odhadnuty pomocí momentových odhadů (2.32) a (2.42). Na obrázku 4.14 jsou ukázány průběhy MV a momentových odhadů parametrů GP rozdělení pro stanici Tuřany při různých volbách hodnot u , resp. k , včetně vybraných optimálních hodnot. Lze zde velmi dobře pozorovat vlastnosti odhadů parametrů, konkrétně jejich rostoucí vychýlení pro rostoucí počet nadprahových hodnot (tj. klesající práh) a značnou variabilitu pro malý počet nadprahových hodnot.

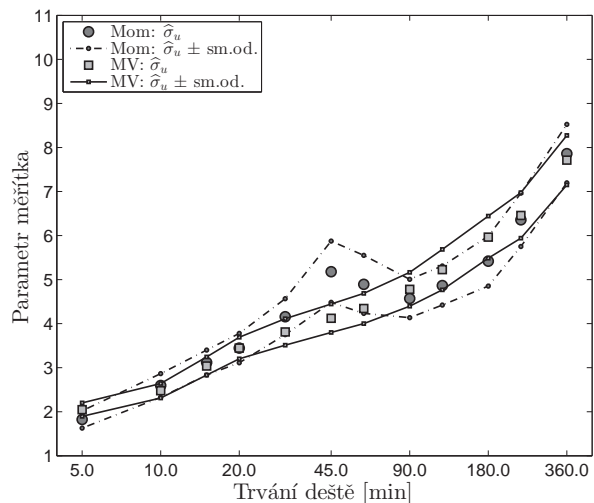


Obrázek 4.14: Průběh MV (čárkovaně) a momentového odhadu (Mom , tučně) parametru GP rozdělení s vyznačenými optimálními hodnotami získanými grafickou analýzou, metodou MT-GP a semiparametrickou dvojitě bootstrapovou technikou pro stanici Brno-Tuřany, doba trvání deště 30 minut.

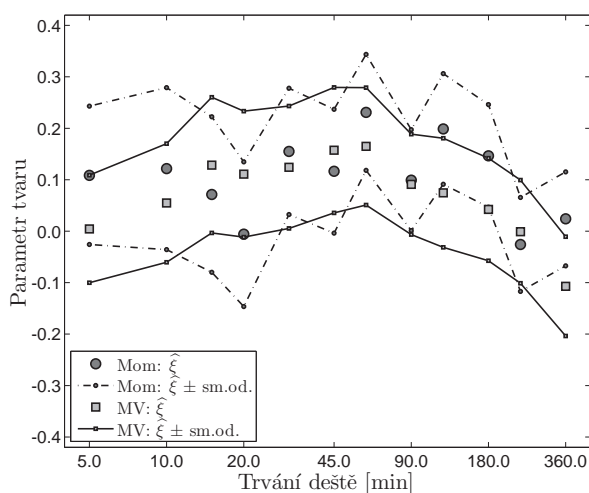
Na obrázku 4.15 jsou pak znázorněny odhady parametru tvaru ξ a měřítka σ_u pro všechna uvažovaná trvání deště pro stanice Brno-Tuřany a Brno-Jundrov. Parametry zde byly určeny jednak momentovými odhady při volbě \hat{k}_0 dle tabulky 4.5 a dále MV metodou při volbách prahových hodnot určených grafickými metodami, přičemž jsou zobrazeny také odhady směrodatných odchylek získané na základě asymptotické normality odhadů. V první řadě lze pozorovat, jak bylo ostatně očekáváno, že v případě krátké pozorované řady ze stanice Brno-Jundrov jsou odhady parametrů zatíženy větší variabilitou než v případě stanice Brno-Tuřany. Šírka příslušných asymptotických 95% intervalů spolehlivosti se pak liší téměř o jeden řád. Podobné chování bylo pozorováno i pro časové řady z ostatních stanic v regionu. Obdobně jsou na obrázku 4.16 porovnány momentové odhady s MV odhady při volbách prahu určených adaptivním přístupem MT-GP - viz tabulka 4.4. Protože prahové hodnoty určené metodou MT-GP jsou ve většině případů nižší než prahy odečtené grafickými technikami, jsou MV odhady parametrů na obrázku 4.16



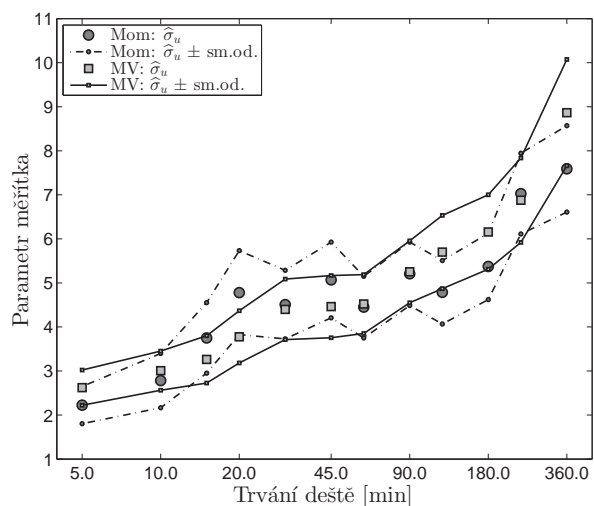
(a) Brno-Tuřany, parametr tvaru.



(b) Brno-Tuřany, parametr měřítka.



(c) Brno-Jundrov, parametr tvaru.

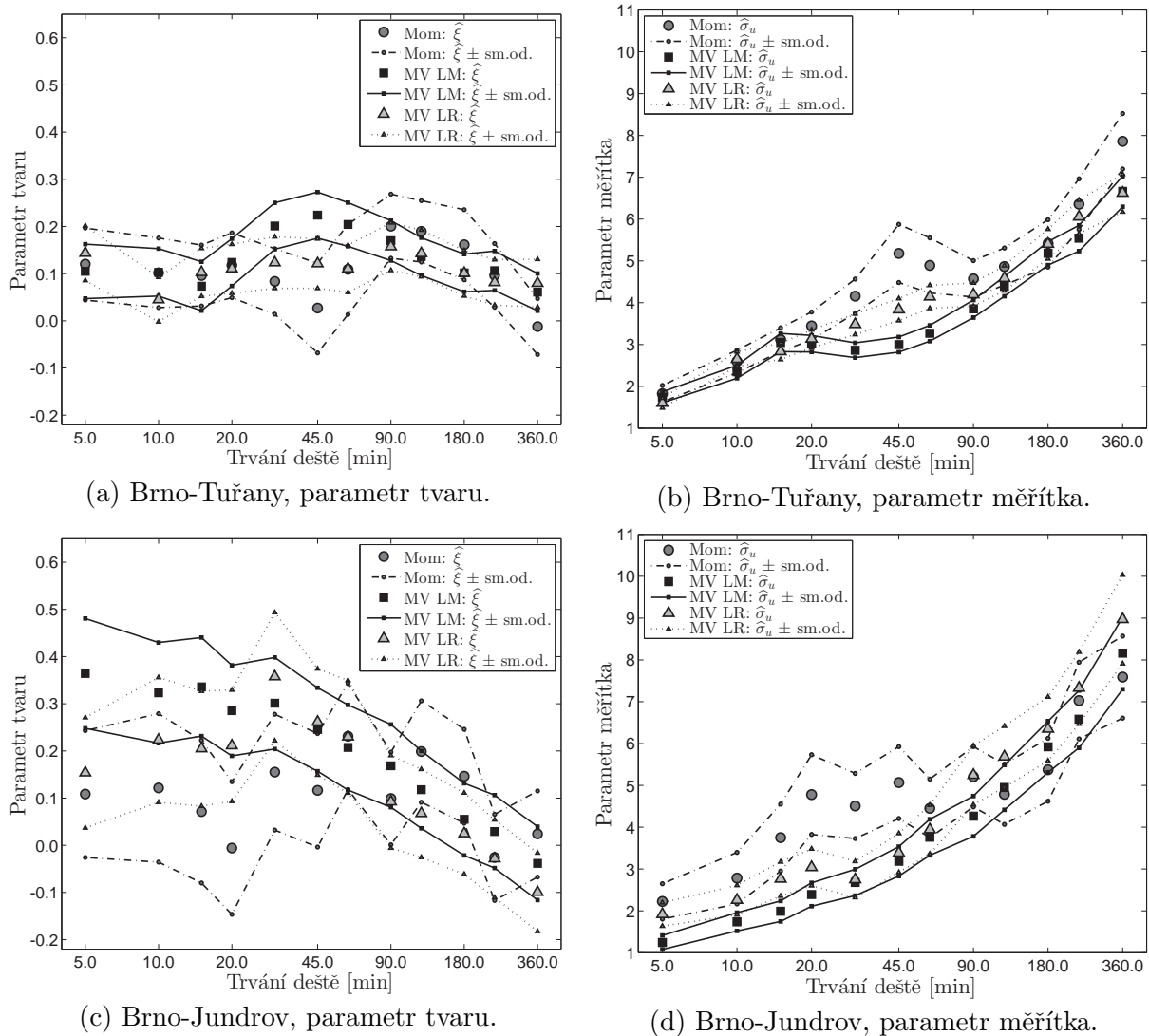


(d) Brno-Jundrov, parametr měřítka.

Obrázek 4.15: Parametry tvaru a měřítka získané semiparametrickými momentovými odhady (*Mom*) a MV metodou při volbě prahu pomocí grafických přístupů pro stanice Brno-Tuřany a Brno-Jundrov. Směrodatné odchylky byly získány na základě asymptotické normality. Horizontální osa je zobrazena v logaritmickém měřítku.

zatíženy menší variabilitou. Na druhé straně je vidět zjevná rozdílnost bodových odhadů sledovaných parametrů v porovnání s MV odhady s prahem určeným grafickými techniky. Toto se týká především případů, kdy byla k určení prahu použita skórová testovací statistika. Teoretické aspekty by měly vést k závěrům, že při vyšší hodnotě prahu dochází k menšímu vychýlení bodových odhadů, avšak vzhledem ke značné variabilitě odhadů (především pro krátké časové řady) není možné takové závěry jasně stanovit.

Při porovnání s momentovými odhady jsou MV odhady s graficky určeným prahem zatíženy menší variabilitou pro delší časové řady. Pro krátkou řadu Brno-Jundrov pak dochází u MV odhadů k nárůstu variability (viz tabulka 4.6), a proto se v těchto případech jeví semiparametrické odhady jako vhodnější. Na druhou stranu vyžaduje dvojitě bootstrapová metoda odhadu k_0 značné výpočetní kapacity. Obdobné výsledky se ukazují i pro MV odhady při prazích určených metodou MT-GP.

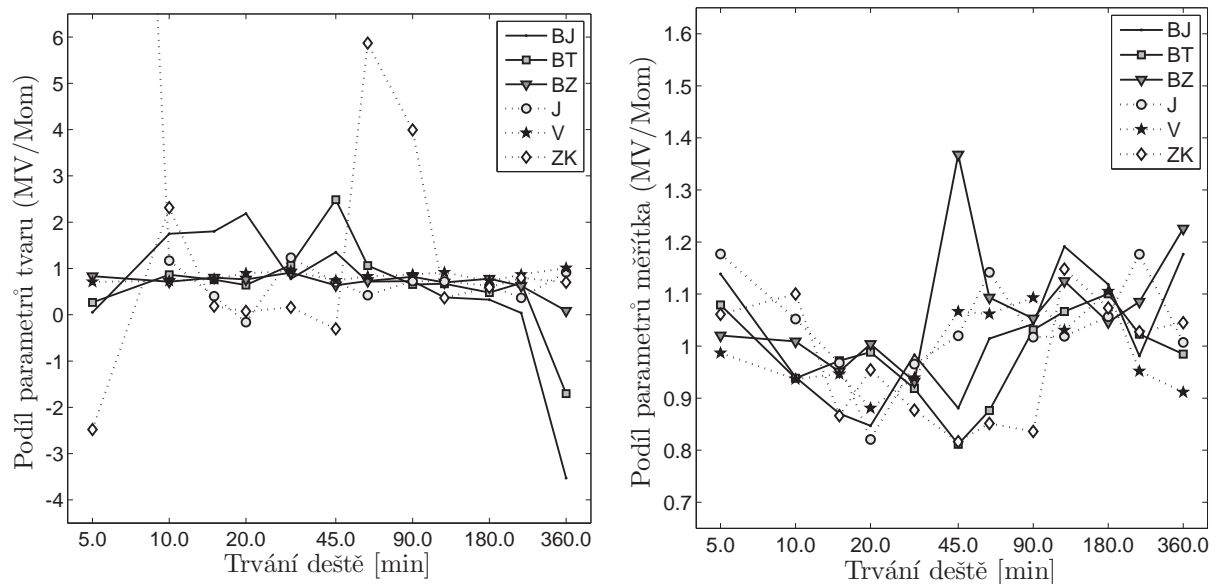


Obrázek 4.16: Parametry tvaru a měřítka získané semiparametrickými momentovými odhady (*Mom*) a MV metodou při volbě prahu pomocí přístupu MT-GP při použití testovacích statistik LM a LR pro stanice Brno-Tuřany a Brno-Jundrov. Směrodatné odchylky byly získány na základě asymptotické normality. Horizontální osa je zobrazena v logaritmickém měřítku.

Tabulka 4.6: Poměr asymptotických směrodatných odchylek (Mom/MV) odhadů parametru tvaru a měřítka získaných MV metodou při volbě prahu grafickými technikami a semiparametrickými momentovými odhady (Mom). Uvedeny jsou výsledky pro stanice Brno-Tuřany (BT) a Brno-Jundrov (BJ).

| Poměr sm. odchylek (Mom/MV) | | 5 | 10 | 15 | 20 | 30 | 45 | 60 | 90 | 120 | 180 | 240 | 360 |
|-----------------------------|----|------|------|------|------|------|------|------|------|------|------|------|------|
| param. tvaru | BT | 1,45 | 1,49 | 1,31 | 1,33 | 1,16 | 1,60 | 1,61 | 1,16 | 1,02 | 1,31 | 1,18 | 1,24 |
| | BJ | 1,28 | 1,36 | 1,15 | 1,16 | 1,03 | 0,99 | 0,99 | 1,01 | 1,01 | 1,00 | 0,91 | 0,95 |
| param. měřítka | BT | 1,33 | 1,67 | 1,38 | 1,38 | 1,35 | 2,11 | 1,95 | 1,15 | 0,97 | 1,18 | 1,17 | 1,18 |
| | BJ | 1,09 | 1,59 | 1,49 | 1,50 | 1,12 | 1,22 | 1,05 | 0,99 | 0,87 | 0,91 | 0,96 | 0,81 |

Na obrázku 4.17 jsou ukázány poměry odhadnutých parametrů tvaru a měřítka pro všechny studované stanice. Zde se jedná opět o momentové odhady a MV odhady obdržené na základě volby prahu dle grafických přístupů. Pro stanice lokalizované v jihozápadní části regionu (Jevišovice a Znojmo-Kuchařovice) lze pozorovat význačné odchylky obou odhadů. Tento fakt lze však především přisuzovat skutečnosti, že příslušné hodnoty parametrů se pohybují blízko nule, což zřejmě odpovídá povaze tohoto subregionu.



Obrázek 4.17: Podíly odhadů (MV/Mom) parametrů tvaru (vlevo) a měřítka (vpravo) získaných MV metodou při volbě prahu grafickými technikami a semiparametrickými momentovými odhady. Zobrazeny jsou stanice Brno-Tuřany (BT), Brno-Jundrov (BJ), Brno-Žabovřesky (BZ), Jevišovice (J), Vyškov (V) a Znojmo-Kuchařovice (ZK).

V poslední době se řada autorů zaměřovala na některé neparametrické metody odhadu, například na metodu pravděpodobnostně vážených momentů (PVM,[29, 30, 30, 75]) či metodu L-momentů [12, 74]. Srovnání předešlé MV metody s metodou PVM bylo detailně provedeno v práci [1], která také ukázala, že pro studované datové soubory vedou oba tyto přístupy k velmi podobným odhadům parametrů modelového GP rozdělení.

Odhady IDF křivek

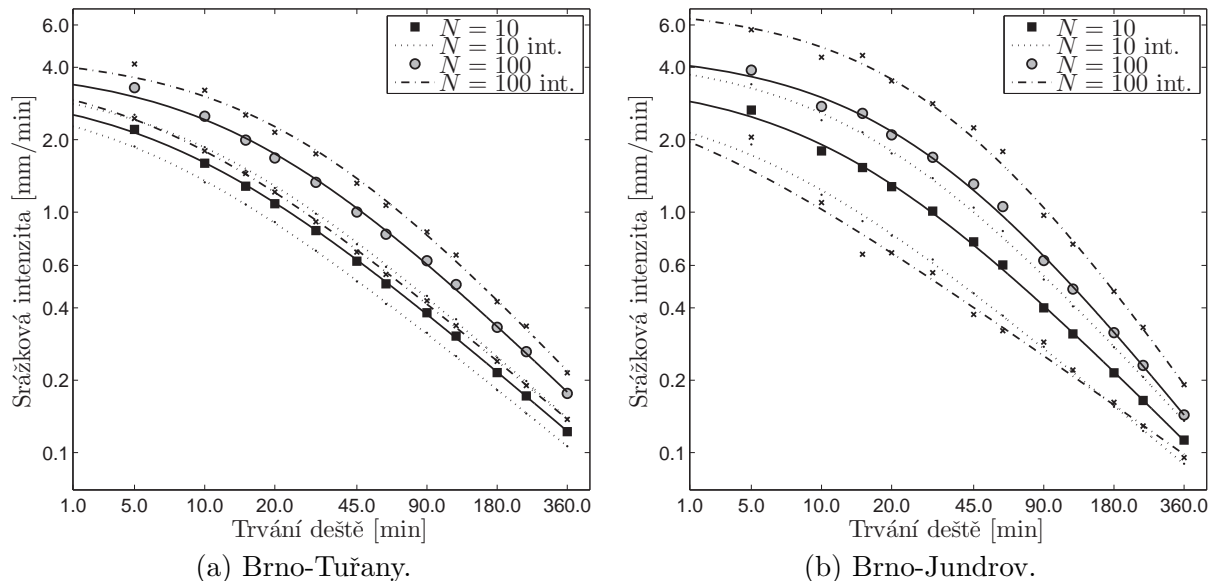
Dále přikročíme k odhadům návratových úrovní, konkrétně k odhadům IDF křivek. Nejprve byly ze vyvzorkovaných nezávislých srážkových úhrnů určeny srážkové intenzity jako podíl srážkových úhrnů a příslušné doby trvání deště. IDF křivky zobrazují pro požadovanou frekvenci výskytu příslušné periodě návratu r závislost intenzity dešťových srážek na trvání deště. Pro praktické použití bývá zvykem určit IDF křivky pro N -letou frekvenci, přičemž sledovaná intenzita se při daném trvání deště objeví v průměru jednou za N let. V kontextu POT modelu je tedy perioda návratu r určena jako $r = Nn_y$, kde n_y je průměrný počet pozorování za jeden rok (viz odstavec 2.1).

Pro snadnější orientaci bývají získané srážkové intenzity proloženy křivkou za použití nelineární regrese. Běžně se volí regresní křivka v následujícím tvaru

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

jejíž použití dokládají např. publikace [80, 108]. $I(t)$ zde označuje odhadnutou intenzitu, t dobu trvání deště a $\beta_1, \beta_2, \beta_3$ jsou regresní koeficienty. Meziregionální srovnání IDF křivek pro celou studovanou oblast Jihomoravského kraje již bylo provedeno v práci [57]. Zde se zaměříme především na srovnání různých metodik odhadů pro jednotlivé časové řady a na rozdíly mezi jednotlivými stanicemi tak v našem případě nebude nutné brát zřetel.

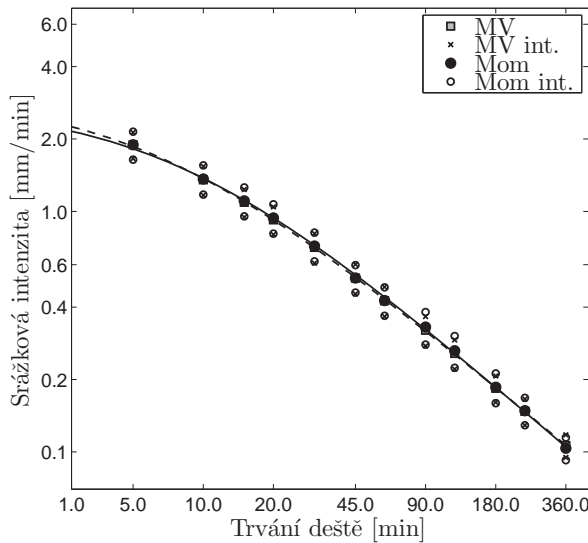
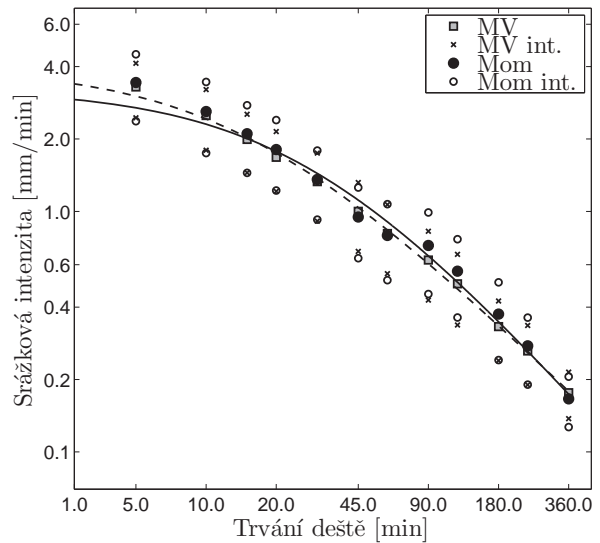
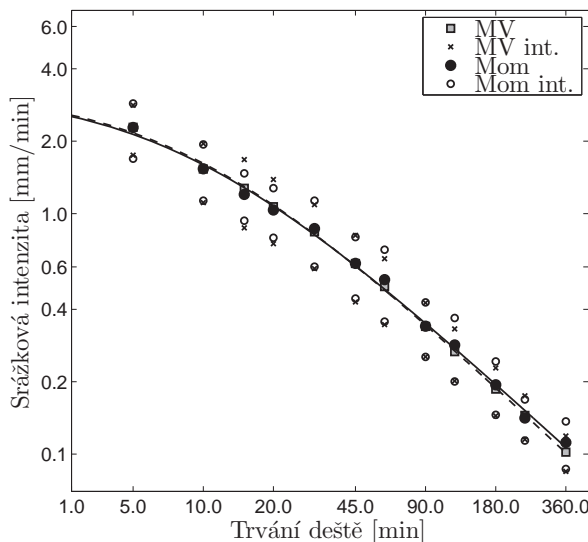
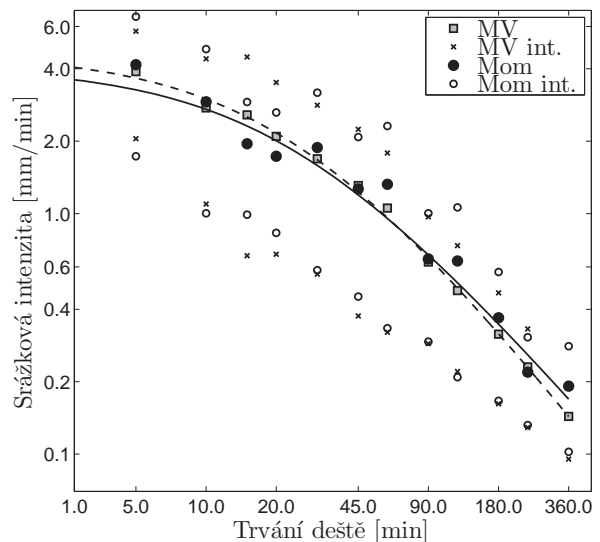
Na obrázku 4.18 jsou zobrazeny odhady IDF křivek pro vybrané stanice Tuřany a Jundrov získané MV metodou při grafické volbě prahové hodnoty. Intervaly spolehlivosti byly určeny na základě asymptotických vlastností MV odhadů pomocí delta metody. Přirozeně lze pozorovat rostoucí šířku intervalů pro rostoucí hodnotu N . Viditelná je i větší variabilita pro krátkou časovou řadu ze stanice Jundrov.



Obrázek 4.18: N -leté návratové úrovně získané MV metodou při grafické volbě prahové hodnoty. IDF křivky byly získány proložení pomocí regresní křivky (plná čára), 95% intervaly spolehlivosti jsou vyznačeny přerušovanými čarami. Osy jsou v logaritmickém měřítku.

Na obrázku 4.19 jsou vizualizovány odhady IDF křivek získané MV metodou při grafické volbě prahu a semiparametrickými momentovými odhady, přičemž 95% intervaly spolehlivosti byly opět získány delta metodou na základě asymptotického rozdělení. Pro

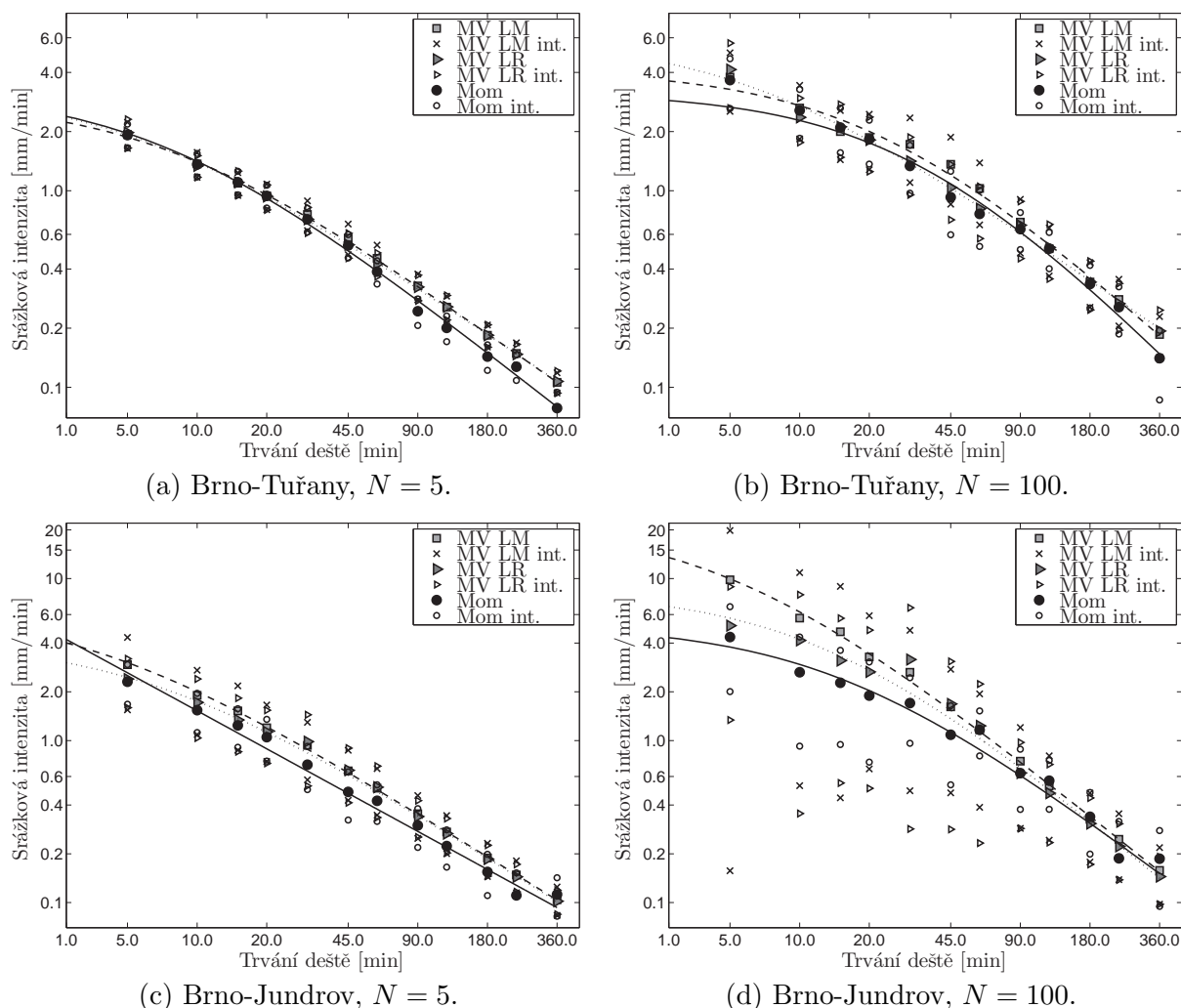
stanici Tuřany lze odhady z praktického hlediska považovat za totožné. Rozdíly mezi oběma metodami jsou patrné pro krátkou časovou řadu ze stanice Jundrov, zejména při odhadech dlouhodobých návratových úrovní. Křivky získané MV metodou vykazují menší variabilitu pro kratší doby trvání deště (do 60 minut), pro delší trvání je pak variabilita menší pro momentové odhady, jak je doloženo v tabulce 4.6. Podobná situace byla pozorována také u obdobně dlouhé časové řady ze stanice Žabovřesky. V těchto situacích je tak možné semiparametrické metody doporučit. Pro dostatečně dlouhé pozorované řady je však použití semiparametrického POT modelu nevhodné, neboť ani značná výpočetní náročnost adaptivního bootstrapového algoritmu nevede k výraznému zpřesnění odhadů IDF křivek.

(a) Brno-Tuřany, $N = 5$.(b) Brno-Tuřany, $N = 100$.(c) Brno-Jundrov, $N = 5$.(d) Brno-Jundrov, $N = 100$.

Obrázek 4.19: MV odhady při grafické volbě prahu (čárkovaně) a momentové odhady (plně) IDF křivek pro stanice Brno-Tuřany a Brno-Jundrov a N -leté návratové úrovně. 95% intervaly spolehlivosti nebyly regresní křivkou proloženy. Osy jsou v logaritmickém měřítku.

Dále je zobrazeno srovnání odhadů IDF křivek získaných semiparametrickou metodou a MV metodou při volbě prahu pomocí přístupu MT-GP. Při pohledu na obrázek 4.20 je

patrné, že jednotlivé odhady se značně liší bez ohledu na délku časové řady a hodnotu N . Pro časovou řadu ze stanice Tuřany se návratové úrovně liší především pro delší doby trvání deště, variabilita je zde vesměs podobná. Pro stanici Jundrov je vidět extrémní nárůst variability MV odhadů IDF křivek pro 100leté návratové úrovně. Volba prahu pomocí metody MT-GP má za následek velké vychýlení křivek oproti semiparametrickému odhadu a MV odhady se zde ukazují jako prakticky nepoužitelné kvůli extrémní šířce příslušných intervalů spolehlivosti. Na druhou stranu pro dlouhé doby trvání deště (240 a 360 minut) se MV odhady pro stanici Jundrov ukazují jako přesnější.



Obrázek 4.20: MV odhady při volbě prahu pomocí MT-GP metody (LM čárkovaně, LR tečkovaně) a momentové odhady (plně) IDF křivek pro stanice Brno-Tuřany a Brno-Jundrov a N -leté návratové úrovně. 95% intervaly spolehlivosti nebyly regresní křivkou proloženy. Osy jsou v logaritmickém měřítku.

Další metody k odhadům IDF křivek lze nalézt např. v práci [1]. Zde byly křivky odhadnuty pomocí neparametrické PVM metody, která byla studována v kombinaci s volbou prahové hodnoty grafickými technikami. Jak již bylo zmíněno v části věnované odhadům parametrů, PVM metoda pro diskutovaná srážková data vede k výsledkům, které jsou srovnatelné s odhady pomocí MV metody při obdobné volbě prahu. Z tohoto důvodu zde předeslané výsledky podrobněji neuvádíme.

Teorie extrémních hodnot pro stacionární procesy

5.1. Rozdělení extrémních hodnot

Dosud představené metody statistického vyhodnocení extrémních hodnot byly založeny na předpokladu nezávislosti jednotlivých náhodných veličin. Vzájemná nezávislost však představuje v mnoha praktických situacích zcela nerealistický požadavek a bývá tak nutné použít dodatečné techniky vzorkování, jak byly představeny např. v odstavci 4.1. Pro zobecnění EV teorie pro případ závislých náhodných veličin je třeba se omezit na některé speciální případy přítomné závislosti. Vzhledem k povaze typických aplikačních problémů, kdy pozorování vystupují většinou ve formě časových řad, je výhodné se soustředit na případ nahrazení uvažovaného náhodného výběru nějakou (striktně) stacionární řadou. Jedná se tak o přirozené zobecnění dosud představené EV teorie. Předpoklad striktní stacionarity, který je více realistickým předpokladem pro mnoho fyzikálních procesů, zaručuje pro libovolné indexy i, j ($i \leq j$) a číslo $m \in \mathbb{N}$, že sdružené rozdělení náhodných vektorů (X_i, \dots, X_j) a $(X_{i+m}, \dots, X_{j+m})$ je totožné. Tato situace pak odpovídá náhodným veličinám, jež mohou být vzájemně závislé, avšak jejichž vlastnosti jsou v čase homogenní. Protože jinými než striktně stacionárními řadami se níže zabývat nebudeme, budeme tyto označovat označovat jednoduše jako *stacionární*¹ (více např. viz [15]).

Omezení ve formě stacionarity se ukazuje jako příliš obecné k tomu, aby umožnilo smysluplné rozvinutí EV teorie. Je nutné přidat požadavek na omezení přítomné závislosti tak, aby závislost mezi veličinami X_i, X_j byla pro rostoucí vzdálenost $|i - j|$ nějakým způsobem limitována. Obecně je podmínka tohoto typu v literatuře označována jako *podmínka silného mixing* (zkráceně SM podmínka z angl. strong mixing; více viz např. [93], řada podmínek na omezení závislosti je shrnuta v [23]). Nejjednodušším případem je tzv. *m-závislost*, kdy X_i a X_j jsou pro $|i - j| > m$ skutečně nezávislé. Vzhledem k řešeným aplikačním problémům však obecně SM podmínka nepředstavuje žádná významná omezení. Ukázkou mohou být např. analyzovaná srážková data z kapitoly 4, kdy časově vzdálená pozorování jsou buď skutečně nezávislá, nebo je závislost přítomna v tak malé míře, že z praktického hlediska na ni není třeba brát ohled. V literatuře je však nejběžněji uvažována následující Leadbetterova $D(u_n)$ podmínka.

¹V anglicky psané literatuře bývá někdy pojmem *stationary series* označována právě striktně stacionární řada.

DEFINICE 5.1. Řekneme, že stacionární řada X_1, \dots, X_n splňuje $D(u_n)$ podmínku, jestliže pro všechny indexy $1 \leq i_1 < \dots < i_p < j_1 < \dots < j_q \leq n$ takové, že $j_1 - i_p > l$, platí

$$\left| P(X_{i_1} \leq u_n, \dots, X_{i_p} \leq u_n, X_{j_1} \leq u_n, \dots, X_{j_q} \leq u_n) - \right. \\ \left. - P(X_{i_1} \leq u_n, \dots, X_{i_p} \leq u_n) P(X_{j_1} \leq u_n, \dots, X_{j_q} \leq u_n) \right| \leq \alpha(n, l), \quad (5.1)$$

kde $\alpha(n, l_n) \rightarrow 0$ pro nějakou posloupnost l_n , pro kterou $l_n/n \rightarrow 0$ při $n \rightarrow \infty$.

Pro nezávislé náhodné veličiny je $\alpha(n, l)$ identicky nulová funkce. Na rozdíl od obecné SM podmínky, $D(u_n)$ podmínka musí být splněna jen pro určité hodnoty prahu u_n a je tak poněkud slabším požadavkem. Typicky budeme vyžadovat, aby práh u_n nějakým způsobem rostl spolu s rostoucím rozsahem n . Uvedená podmínka zaručuje, že náhodné jevy $[X_i > u_n]$ a $[X_j > u_n]$ jsou přibližně nezávislé za předpokladu, že práh u_n je vhodně vysoký a vzdálenost mezi indexy i a j je dostatečná, neboli při velkém časovém rozmezí jsou extrémní pozorování blízká nezávislým realizacím. Dále stanovíme tvrzení obdobné Základní větě teorie extrémních hodnot, která byla představena v kapitole 1. Tvrzení uvedená v následujícím odstavci budou uvedena bez důkazů, přičemž pro jejich provedení bude odkázáno na citovanou literaturu.

VĚTA 5.2. [93] Nechtě X_1, \dots, X_n je stacionární řada a $M_n = \max\{X_1, \dots, X_n\}$. Pokud existují takové konstanty $a_n > 0, b_n$ tak, že

$$\lim_{n \rightarrow \infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = G(x),$$

kde $G(x)$ je nedegenerovaná distribuční funkce, a pokud je dále splněna $D(u_n)$ podmínka pro $u_n = a_n x + b_n$ pro každé $x \in \mathbb{R}$ takové, že $G(x) > 0$, pak $G(x)$ je distribuční funkcí GEV rozdělení.

Předchozí tvrzení udává, že výběrová maxima stacionární řady se řídí stejnými limitními zákony jako tomu bylo pro nezávislé náhodné veličiny, ačkoliv limitní GEV rozdělení se může lišit. V následujícím odstavci budeme mimo stacionární řady X_1, X_2, \dots, X_n často uvažovat tzv. *přidružený náhodný výběr* $X_1^*, X_2^*, \dots, X_n^*$, tj. náhodný výběr se stejným (jednorozměrným) marginálním rozdělením jako má původní stacionární řada. Následující věta popisuje vztah mezi limitními rozděleními výběrových maxim stacionární řady a k ní přidruženého náhodného výběru.

VĚTA 5.3. [93] Nechtě X_1, \dots, X_n je stacionární řada a X_1^*, \dots, X_n^* je k ní přidružený náhodný výběr. Označme příslušná výběrová maxima $M_n = \max\{X_1, \dots, X_n\}$, $M_n^* = \max\{X_1^*, \dots, X_n^*\}$. Pokud existují takové konstanty $a_n > 0, b_n$, že

$$\lim_{n \rightarrow \infty} P\left(\frac{M_n^* - b_n}{a_n} \leq x\right) = G(x), \quad (5.2)$$

kde $G(x)$ je nedegenerovaná distribuční funkce, a pokud je dále splněna $D(u_n)$ podmínka pro $u_n = a_n x + b_n$ pro každé x takové, že $G(x) > 0$, potom

$$\lim_{n \rightarrow \infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = G_\theta(x), \quad (5.3)$$

kde $G_\theta(x)$ je opět nedegenerovaná distribuční funkce. Navíc platí $G_\theta(x) = [G(x)]^\theta$, kde $\theta \in (0, 1)$.

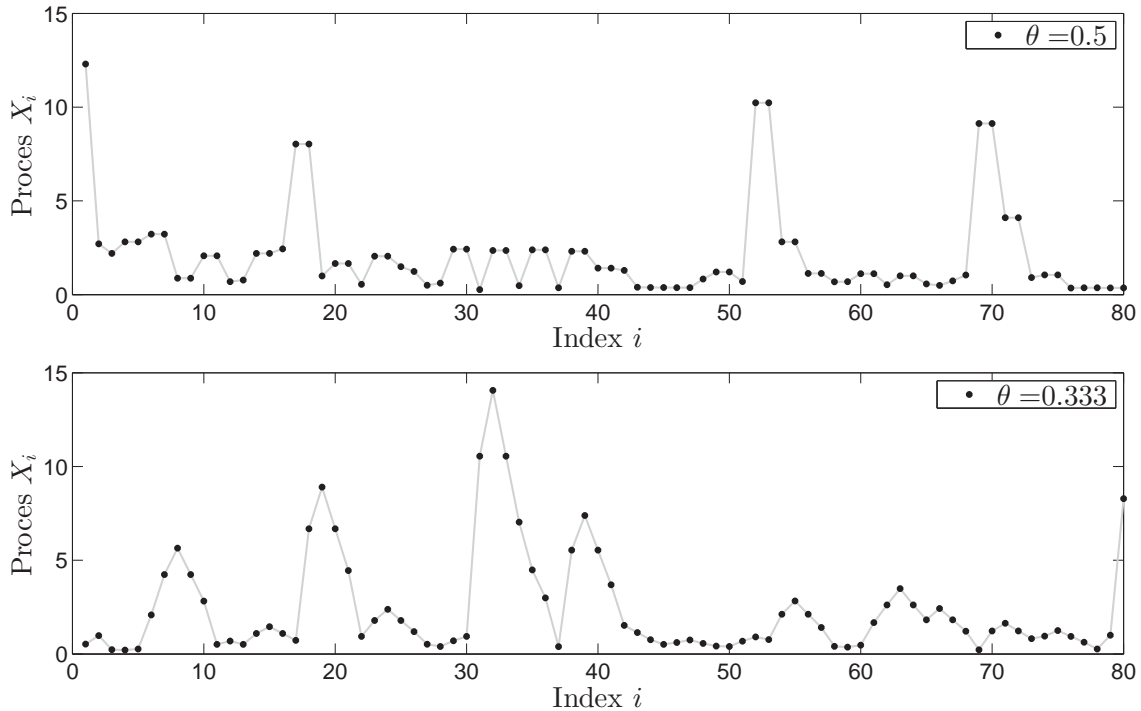
Z předcházejícího tvrzení tedy plyne, že zatímco limitním rozdělením výběrových maxim nezávislých veličin je $\text{GEV}(\mu, \sigma, \xi)$ rozdělení s distribuční funkcí $G(x)$, výběrová maxima stacionární řady konvergují v distribuci k $\text{GEV}(\mu_\theta, \sigma_\theta, \xi_\theta)$ rozdělení s distribuční funkcí $G_\theta(x) = [G(x)]^\theta$. Pro $\xi \neq 0$ mezi parametry obou rozdělení platí následující vztahy

$$\mu_\theta = \mu - \frac{\sigma}{\xi} (1 - \theta^\xi), \quad \sigma_\theta = \sigma \theta^\xi, \quad \xi_\theta = \xi, \quad (5.4)$$

pro $\xi = 0$ je pak $\mu_\theta = \mu + \sigma \ln \theta$, $\sigma_\theta = \sigma$.

Parametr θ se nazývá *extremální index* a vyjadřuje míru krátkodobé závislosti na extrémních úrovních. Pro posloupnost nezávislých veličin je zřejmě $\theta = 1$, opačná implikace ovšem neplatí [93]. Příklad $\theta = 0$ nebudeme dále uvažovat, protože je do jisté míry patologický. Jak ovšem dokládá práce [43], není tento případ nemožný.

Vliv nějaké krátkodobé závislosti (např. splňující $D(u_n)$ podmínku) lze v realizacích stacionárních řad pozorovat ve formě shlukování měření na extrémních úrovních. Čím větší je míra této závislosti, tím větší shluky se vytváří. Na obrázku 5.1 jsou ilustrativně znázorněny realizace procesu klouzavých maxim [18] popsaného níže v odstavci 6.1. Protože přítomnost krátkodobé závislosti vede k zavedení parametru θ (viz (5.3)), zřejmě tak existuje vazba mezi extrémálním indexem a shlukováním extrémních hodnot.



Obrázek 5.1: Realizace procesu klouzavých maxim pro různá nastavení parametrů (viz odstavce 6.1). Nižší hodnota θ vede k větší velikosti shluků extrémních hodnot.

Extremální index je možné interpretovat několika způsoby, nejčastěji je uvažován následující výklad dle [93]. Mějme stacionární řadu X_1, \dots, X_n , kterou rozdělíme do bloků délky r_n , $r_n = o(n)$. *Shlukem extrémních hodnot* se rozumí množina všech překročení nějakého vysokého prahu u_n v daném bloku za předpokladu, že takové překročení nastalo. Označme pro $j = 1, \dots, r_n$

$$\pi_n(j) = P \left(\sum_{i=1}^{r_n} \mathbf{1}_{[X_i > u_n]} = j \mid M_{r_n} > u_n \right) \quad (5.5)$$

pravděpodobnost, že v jednom bloku nastane j překročení prahu u_n . Pak, za určitých podmínek konvergence rozdělení π_n [77], lze ukázat, že

$$\theta^{-1} = \lim_{n \rightarrow \infty} \sum_{j=1}^{r_n} j \pi_n(j) = \lim_{n \rightarrow \infty} \mathbb{E} \left[\sum_{i=1}^{r_n} \mathbf{1}_{[X_i > u_n]} \middle| M_{r_n} > u_n \right]. \quad (5.6)$$

Podle (5.6) tedy θ^{-1} představuje limitní střední hodnotu velikosti shluku. Jiná interpretace byla odvozena O'Brienem [104] v následujícím tvaru

$$\theta = \lim_{n \rightarrow \infty} P \left(X_2 \leq u_n, \dots, X_{r_n} \leq u_n \middle| X_1 > u_n \right), \quad (5.7)$$

tudíž θ představuje pravděpodobnost, že překročení prahu u_n bude následováno pozorováními pod tímto prahem.

V praktických situacích, kdy jsou k dispozici pozorování nějaké stacionární řady, není možné přímé užití teorie pro nezávislé veličiny. Obecně je možno postupovat dvěma způsoby, a sice (i) extrémální index nějak vhodně odhadnout nebo (ii) použít nějaké filtrační schéma, které by vliv θ eliminovalo.

Druhé zmiňované je v literatuře označováno jako *declusterování* (z angl. declustering), neboli „odshlukování“. Cílem declusterovacích technik je z původní stacionární řady vybrat pozorování, která můžeme považovat za přibližně nezávislá, přičemž pro tyto je možné použít standardní metody z kapitoly 2. V praxi jsou používány dvě metody declusterování, které byly diskutovány v odstavci 4.1. V literatuře jsou tyto techniky označovány jako *blocks declustering* [121, 33] a *runs declustering* [50, 18]. V první z nich je daná řada rozdělena na bloky, přičemž všechna překročení daného prahu ve stejném bloku jsou chápána jako jeden shluk (dle (5.6)). Předpokládá se pak vzájemná nezávislost různých shluků. Při použití runs declusterování se nejprve stanoví hodnota *runs* parametru $l \in \mathbb{N}$ a pro danou prahovou hodnotu u se dva shluky chápou jako nezávislé, pokud jsou odděleny alespoň l po sobě následujícími podprahovými hodnotami. Typicky pak z každého shluku vybíráme jen nejvyšší pozorovanou hodnotu. Použití declusterovacích technik má ovšem za následek (často významnou) redukci původní pozorované řady. Navíc v obou případech je volba pomocných parametrů prováděna pomocí různých explorativních technik a je tak zatížena značnou mírou subjektivity.

Jako lepší se tak jeví odhadnout extrémální index pomocí nějakého vhodného odhadu $\hat{\theta}$. Protože limitním rozdělením výběrových maxim stacionární řady je GEV rozdělení (5.3), lze k analýze použít standardní metody bez ohledu na závislost. Odhady parametrů či parametrických funkcí se poté korigují dle získaného odhadu θ . Výhodou toho postupu je možnost využití všech dostupných pozorování, což může významně snížit variabilitu odhadů parametrů modelového rozdělení.

Obě zmiňované techniky declusterování byly softwarově implementovány do prostředí Matlab a pro blocks i runs declusterování lze tak použít přiloženou funkci `decluster.m`. Popis uvedené funkce je uveden v dodatku D, detaily ohledně vstupních či výstupních parametrů lze nalézt přímo ve zdrojovém kódu. Nastavením parametru funkce je uživateli dána možnost zvolit mezi přístupy k declusterování; dále je nutné zadat buď velikost bloku, nebo velikost runs parametru a zvolenou prahovou hodnotu. Částečně byly techniky declusterování také implementovány i do softwaru EVDest (zde jen formou runs declusterování, viz dodatek C). EVDest umožňuje analýzu extrémních hodnot stacionárních řad v rozsahu, který je představen v této kapitole.

5.2. Metody odhadu extrémálního indexu

Jednoduchý odhad extrémálního indexu lze zavést jako průměrnou velikost shluku, tj. ve tvaru

$$\hat{\theta} = \frac{C(u)}{N(u)}, \quad (5.8)$$

kde $C(u)$ označuje počet shluků a $N(u)$ značí počet pozorování nad prahem u_n . V literatuře se objevují různé přístupy k identifikaci shluků. Dle (5.6) se daná stacionární řada X_1, \dots, X_n rozdělí na m bloků délky r ($mr \approx n$), přičemž počet shluků je odhadnut jako počet bloků, v nichž alespoň jedno pozorování překročí daný práh u . Takto zavedený odhad extrémálního indexu je označován jako *blokový odhad*. Ze vztahu (5.7) vychází jiný odhad počtu shluků. Při dané hodnotě runs parametru $l \in \mathbb{N}$ a prahu u uvažujeme dva shluky jako různé, pokud jsou odděleny alespoň l po sobě jdoucími podprahovými pozorováními, tj.

$$C(u) = \sum_{i=1}^{n-l} \mathbf{1}_{[X_i > u]} (1 - \mathbf{1}_{[X_{i+1} > u]}) \dots (1 - \mathbf{1}_{[X_{i+l} > u]}).$$

Tento odhad θ je označován jako *runs odhad*. Obě hodnoty $C(u), N(u)$ jsou však silně závislé na hodnotě u , což spolu s nejasnou volbou prahu vede často k velkému vychýlení odhadu θ . Detailní srovnání blokového a runs odhadu je provedeno např. v práci [121], diskuzi optimálních hodnot parametrů l, m lze nalézt v [103]. V poslední době se objevují nové variace těchto odhadů, např. dvojprahový runs odhad [92], který se snaží redukovat důležitost vhodné volby prahu. Tento je však poměrně výpočetně náročný a stále málo robustní oproti jiným odhadům diskutovaným níže.

Odhad podle Gomes

Jiný přístup navrhla ve své práci Gomes [60], jejíž odhad $\hat{\theta}_G$ parametru θ vychází ze vztahů (5.2) a (5.3). Protože přítomnost krátkodobé závislosti (5.1) se v limitním GEV rozdělení výběrových maxim projevuje pouze zavedením extrémálního indexu, lze θ získat porovnáním dané stacionární řady X_1, \dots, X_n a jejího přidruženého náhodného výběru X_1^*, \dots, X_n^* .

Vyjdeme z metody blokových maxim představené v odstavci 2.1. Stacionární řadu X_1, \dots, X_n rozdělíme do m bloků délky k tak, aby $mk \approx n$. Označme $M_{n,1}, \dots, M_{n,m}$, resp. $M_{n,1}^*, \dots, M_{n,m}^*$ bloková maxima dané stacionární řady, resp. nějakého k ní přidruženého náhodného výběru. Dle metody blokových maxim lze přímo získat MV odhady $\hat{\mu}_\theta, \hat{\sigma}_\theta, \hat{\xi}_\theta$ parametrů příslušného $\text{GEV}(\mu_\theta, \sigma_\theta, \xi_\theta)$ rozdělení, jakožto modelového rozdělení veličin M_n . Dále se zabýváme odhady parametrů $\text{GEV}(\mu, \sigma, \xi)$ rozdělení přidruženého náhodného výběru. Gomes uvažovala takový přidružený náhodný výběr X_1^*, \dots, X_n^* , který vznikne náhodným přerovnáním původní stacionární řady. Takto získaný náhodný výběr slouží jako aproximace řady, která má stejné marginální rozdělení jako X_1, \dots, X_n , přičemž metodou blokových maxim lze pomocí příslušných blokových maxim $M_{n,1}^*, \dots, M_{n,m}^*$ určit MV odhady $\hat{\mu}, \hat{\sigma}, \hat{\xi}$ parametrů μ, σ, ξ . Simulační studie [5] ukazuje, že počet přerovnání původní stacionární řady má jen nepatrný vliv na odhad extrémálního indexu za předpokladu, že rozsah řady n je dostatečný a délka bloku k je uvažována v rozumných rozpětích (dále viz odstavce 6.1).

S využitím vztahů (5.4) zavedla Gomes [60] odhad $\widehat{\theta}_G$ extrémálního indexu ve tvaru

$$\widehat{\theta}_G = \left(\frac{\widehat{\sigma}}{\widehat{\sigma}_\theta} \right)^{-1/\bar{\xi}}, \quad (5.9)$$

kde

$$\bar{\xi} = \frac{\widehat{\sigma} - \widehat{\sigma}_\theta}{\widehat{\mu} - \widehat{\mu}_\theta} \quad (5.10)$$

označuje sdružený odhad EV indexu ξ .

Odhad rozptylu odhadu extrémálního indexu $\widehat{\theta}_G$ určíme pomocí delta metody [31]. Pro zjednodušení označme dále $\boldsymbol{\omega} = (\mu_\theta, \sigma_\theta, \xi_\theta, \mu, \sigma, \xi)$. Parametry GEV rozdělení byly odhadnuty pomocí metody blokových maxim a bude proto vhodné využít příslušných variančních matic. Buďte \mathbf{V}_θ^* a \mathbf{V}^* varianční matice MV odhadů $(\widehat{\mu}_\theta, \widehat{\sigma}_\theta, \widehat{\xi}_\theta)$ a $(\widehat{\mu}, \widehat{\sigma}, \widehat{\xi})$ (viz odstavec 2.2) a zaveďme sdruženou varianční matici $\overline{\mathbf{V}} = \overline{\mathbf{V}}(\widehat{\boldsymbol{\omega}})$ MV odhadu $\widehat{\boldsymbol{\omega}}$, tj. matici ve tvaru

$$\overline{\mathbf{V}} = \begin{bmatrix} \mathbf{V}_\theta^* & \mathbf{0} \\ \mathbf{0} & \mathbf{V}^* \end{bmatrix}.$$

Rozptyl $\text{var}(\widehat{\theta}_G)$ je možné odhadnout jako

$$\text{var}(\widehat{\theta}_G) \approx \nabla \widehat{\theta}_G^T \overline{\mathbf{V}} \nabla \widehat{\theta}_G, \quad (5.11)$$

kde z (5.9) máme

$$\begin{aligned} \nabla \widehat{\theta}_G^T &= \left(\frac{\partial \widehat{\theta}_G}{\partial \mu_\theta}, \frac{\partial \widehat{\theta}_G}{\partial \sigma_\theta}, \frac{\partial \widehat{\theta}_G}{\partial \xi_\theta}, \frac{\partial \widehat{\theta}_G}{\partial \mu}, \frac{\partial \widehat{\theta}_G}{\partial \sigma}, \frac{\partial \widehat{\theta}_G}{\partial \xi} \right) = \\ &= \left(\widehat{\theta}_G (\sigma - \sigma_\theta)^{-1} \ln \frac{\sigma}{\sigma_\theta}, \widehat{\theta}_G \bar{\xi}^{-1} \left[\sigma_\theta^{-1} - (\sigma - \sigma_\theta) \ln \frac{\sigma}{\sigma_\theta} \right], 0, \right. \\ &\quad \left. -\widehat{\theta}_G (\sigma - \sigma_\theta)^{-1} \ln \frac{\sigma}{\sigma_\theta}, \widehat{\theta}_G \bar{\xi}^{-1} \left[(\sigma - \sigma_\theta)^{-1} \ln \frac{\sigma}{\sigma_\theta} - \sigma^{-1} \right], 0 \right), \end{aligned}$$

přičemž parametr $\boldsymbol{\omega}$ je nahrazen svým MV odhadem $\widehat{\boldsymbol{\omega}}$.

Zabývejme se nyní opět odhady vybraných parametrických funkcí, a sice odhadem návratové úrovně z_r příslušné periodě návratu r (tj. $(1 - r^{-1})$ kvantilem modelového rozdělení). V návaznosti k declusterování diskutovanému výše se jeví odhadnutí extrémálního indexu výhodné především pro možnost zahrnutí do analýzy všech pozorování nad zvoleným prahem. Z tohoto důvodu je smysluplné se soustředit výhradně na použití prahového POT modelu. Připomeňme, že všechny techniky volby vhodné prahové hodnoty lze použít bez rozdílu i v případě stacionární řady, jak vyplývá z tvrzení 5.2.

Uvažujme stacionární řadu X_1, \dots, X_n splňující podmínky věty 5.2. Pro konečné a dostatečně velké n můžeme platnost vztahu (5.3) chápat asymptoticky, tj. můžeme aproximovat

$$P(M_n \leq x) \approx G^\theta(x) \approx F^{n\theta}(x), \quad (5.12)$$

kde $F(x)$ je marginální distribuční funkce náhodných veličin X_1, \dots, X_n . Parametry měřítka a polohy není nutné uvažovat, neboť jak bylo dříve ukázáno v odstavci věnovaném

metodě blokových maxim, lze vztah (5.3) převést na předchozí jednoduchou reparametrizací. Současně, jak vyplývá z (2.5), při použití POT modelu pro nezávislé veličiny můžeme aproximovat

$$P(X \leq x) \approx 1 - \lambda_u(1 - H(x - u)), \quad (5.13)$$

kde bylo v souladu s předchozím označeno $\lambda_u := P(X > u)$ a $H(x)$ je distribuční funkce GP rozdělení. Pro dostatečně vysokou prahovou hodnotu u pravá strana předchozího vztahu zastupuje člen $F^n(x)$ v (5.12). Kombinací předchozích lze návratovou úroveň z_r obdržet řešením rovnosti

$$[1 - \lambda_u(1 - H(z_r - u))]^\theta = 1 - r^{-1},$$

odtud tedy

$$z_r = \begin{cases} u + \frac{\sigma_u}{\xi} \left[\left\{ \lambda_u^{-1} \left[1 - (1 - r^{-1})^{\theta-1} \right] \right\}^{-\xi} - 1 \right], & \text{pro } \xi \neq 0, \\ u + \sigma_u \ln \left\{ \lambda_u \left[1 - (1 - r^{-1})^{\theta-1} \right]^{-1} \right\}, & \text{pro } \xi = 0. \end{cases} \quad (5.14)$$

Pro N -letou návratovou úroveň opět položíme $r = Nn_y$, kde n_y je průměrný počet pozorování za jeden rok. Odhad návratové úrovně je možné obdržet nahrazením parametrů POT modelu jejich odhady $\hat{\xi}, \hat{\sigma}_u, \hat{\lambda}_u$, tj. odhady získanými pomocí *všech* nadprahových hodnot.

Soustředme se dále výhradně na odhady získané MV metodou. Označme $\mathbf{V}_{\text{GP}}(\hat{\sigma}_u, \hat{\xi})$ varianční matici odhadů $\hat{\sigma}_u, \hat{\xi}$ (např. viz [57]). Podle výsledků odstavce 2.2 můžeme přibližně určit rozptyl $\text{var}(\hat{\lambda}_u) \approx \hat{\lambda}_u(1 - \hat{\lambda}_u)/n$. Nechť \mathbf{V} je varianční matice odhadů všech parametrů $\hat{\lambda}_u, \hat{\sigma}_u, \hat{\xi}, \hat{\theta}_G$. Pro praktické účely je vhodné učinit předpoklad (viz [50]), že $\hat{\theta}_G$ není korelovaný s žádným se zbývajících odhadů $\hat{\xi}, \hat{\sigma}_u, \hat{\lambda}_u$. V tom případě pak můžeme varianční matici \mathbf{V} vyjádřit ve tvaru

$$\mathbf{V} = \begin{pmatrix} \hat{\lambda}_u(1 - \hat{\lambda}_u)/n & \mathbf{0} & 0 \\ \mathbf{0} & \mathbf{V}_{\text{GP}}(\hat{\sigma}_u, \hat{\xi}) & 0 \\ 0 & 0 & \text{var}(\hat{\theta}_G) \end{pmatrix}.$$

Odhad rozptylu statistiky \hat{z}_r získaný delta metodou je pak možné ve tvaru

$$\text{var}(\hat{z}_r) \approx \nabla z_r^T \mathbf{V} \nabla z_r,$$

kde pro $\xi \neq 0$

$$\begin{aligned} \nabla z_r^T &= \left(\frac{\partial z_r}{\partial \lambda_u}, \frac{\partial z_r}{\partial \sigma_u}, \frac{\partial z_r}{\partial \xi}, \frac{\partial z_r}{\partial \theta} \right) = \\ &= \left(\sigma_u \left(\frac{\beta}{\lambda_u} \right)^{-\xi} \lambda_u^{-1}, \xi^{-1} \left[\left(\frac{\beta}{\lambda_u} \right)^{-\xi} - 1 \right], -\frac{\sigma_u}{\xi} \left\{ \left(\frac{\beta}{\lambda_u} \right)^{-\xi} \left[1 + \xi \ln \left(\frac{\beta}{\lambda_u} \right) \right] - 1 \right\}, \right. \\ &\quad \left. -\frac{\sigma_u}{\theta^2} \left(\frac{\beta}{\lambda_u} \right)^{-\xi} \beta^{-1} (1 - \beta) \ln(1 - r^{-1}) \right), \end{aligned}$$

přičemž $\beta := 1 - (1 - r^{-1})^{\theta-1}$. Uvedené parametry ξ, σ_u, λ_u jsou v předchozím vztahu nahrazeny jejich MV odhady a θ je nahrazen odhadem $\hat{\theta}_G$.

Odhad Ancony-Navarreteho a Tawna

Modifikace předchozího odhadu podle Gomes byla navržena Anconou-Navarreteho a Tawnem [13]. Namísto odhadu parametrů $(\mu_\theta, \sigma_\theta, \xi_\theta)$ a (μ, σ, ξ) zvlášt' navrhli autoři práce odhadnout všechny parametry současně pomocí sdruženého rozdělení veličin $M_{n,i}$ a $M_{n,i}^*$, $i = 1, \dots, m$. Blokovaná maxima $M_{n,1}^*, \dots, M_{n,m}^*$ přidruženého náhodného výběru jsou přitom určena dle postupu navrženého Gomes [60], tedy jedná se o maxima řady vzniklé náhodným přerovnáním původní stacionární řady.

Dle vztahů (5.4) lze parametry $\text{GEV}(\mu_\theta, \sigma_\theta, \xi_\theta)$ rozdělení popsat pomocí čtveřice $(\mu, \sigma, \xi, \theta)$ a tudíž je nutné odhadnout jen tyto parametry. Pro odhad parametrů např. pomocí MV metody je především nutné, aby byl splněn požadavek vzájemné nezávislosti veličin

$$(M_{n,1}, \dots, M_{n,m}, M_{n,1}^*, \dots, M_{n,m}^*), \quad (5.15)$$

který zde zřejmě splněn není. Autoři práce [13] však argumentují asymptotickou nevýznamností přítomné závislosti a pro praktické účely ji považují za zanedbatelnou.

MV odhady parametrů $(\mu, \sigma, \xi, \theta)$ sdruženého rozdělení vektoru $(M_{n,1}, \dots, M_{n,m}, M_{n,1}^*, \dots, M_{n,m}^*)$ jsou pak určeny maximalizací logaritmické věrohodnostní funkce (pro $\xi \neq 0$)

$$\begin{aligned} l(\mu, \sigma, \xi, \theta) = & -m \ln \sigma - \sum_{i=1}^m \left[1 + \xi \frac{x_i^* - \mu}{\sigma} \right]^{-1/\xi} - \left(1 + \frac{1}{\xi} \right) \sum_{i=1}^m \ln \left[1 + \xi \frac{x_i^* - \mu}{\sigma} \right] - \\ & - m [\ln \sigma + \xi \ln \theta] - \sum_{i=1}^m \left[\frac{\sigma + \xi(x_i - \mu)}{\sigma \theta^\xi} \right]^{-1/\xi} - \left(1 + \frac{1}{\xi} \right) \sum_{i=1}^m \ln \left[\frac{\sigma + \xi(x_i - \mu)}{\sigma \theta^\xi} \right], \end{aligned} \quad (5.16)$$

kde x_i , resp. x_i^* , pro $i = 1, \dots, m$ označují pozorované hodnoty blokových maxim $M_{n,i}$, resp. $M_{n,i}^*$, pro $i = 1, \dots, m$. Uvedený MV odhad extrémálního indexu budeme dále označovat $\hat{\theta}_{AT}$.

Rozptyl odhadu parametru θ (ať už odhadu $\hat{\theta}_G$ či $\hat{\theta}_{AT}$) je možné určit na základě asymptotických vlastností MV odhadů. Tyto zde kvůli rozsahu předkládané práce neuvádíme, ale spolu s uvedenými odhady extrémálního indexu jsou implementovány v softwaru Matlab ve funkci `theta_GAT.m`, která je přiložena v dodatku D této práce. Odvození rozptylu odhadu $\hat{\theta}_{AT}$ se provede přímočaře pomocí výběrové FIM jako $\tilde{\mathbf{J}}^{-1}(\hat{\mu}, \hat{\sigma}, \hat{\xi}, \hat{\theta}_{AT})$ podle (2.11). Odhad rozptylu odhadu $\hat{\theta}_{AT}$ se obdrží delta metodou podle vztahu (5.11). Funkce `theta_GAT.m` nabízí možnost určit odhad $\hat{\theta}_G$ či $\hat{\theta}_{AT}$ pro zvolenou délku bloku (kvůli výpočetní stabilitě je $\hat{\theta}_G$ určen vždy a použit pro inicializaci maximalizace věrohodnostní funkce (5.16)). Detailnější informace lze nalézt v dodatku D nebo přímo ve zdrojovém kódu. Odhad podle Gomes byl navíc také implementován do softwaru `EVDest` (viz dodatek C), kde slouží jako základní odhad shlukování hodnot.

Další odhady

V krátkosti shrňme ještě některé, v poslední době rozvinuté, metody odhadu extrémálního indexu θ . Ferro a Segers ve své práci [51] navrhli odhad založený na asymptotických vlastnostech časů mezi překročeními dané prahové hodnoty. Uvažujme stacionární řadu X_1, \dots, X_n s marginální distribuční funkcí $F(x)$ a dostatečně vysoký práh u . Označme

$T(u)$ náhodnou veličinu, která v distribuci odpovídá veličině $\min\{j \geq 1 : X_{j+1} > u\}$ za podmínky $X_1 > u$. Tedy pro $j \geq 1$ máme

$$P(T(u) = j) = P(X_2 \leq u, \dots, X_j \leq u, X_{j+1} > u \mid X_1 > u).$$

Dle výsledku v práci [51] lze ukázat, že za určitých podmínek regularity, platí při $u \rightarrow x^*$

$$\overline{F}(u)T(u) \xrightarrow{d} T_\theta, \quad (5.17)$$

kde $\overline{F} := 1 - F$ a kde T_θ je náhodná veličina odpovídající směsi rozdělení

$$T_\theta \stackrel{d}{=} (1 - \theta)\varepsilon_0 + \theta E_\theta,$$

přičemž ε_0 označuje veličinu s degenerovaným rozdělením² v $x = 0$ a E_θ je veličina s exponenciálním rozdělením se střední hodnotou θ^{-1} .

Tedy při změně měřítka časů $T(u)$ mezi překročeními prahu u pomocí $\overline{F}(u)$ můžeme časy mezi veličinami ve stejném shluku extrémních hodnot modelovat pomocí degenerovaného rozdělení, zatímco časy mezi jednotlivými (nezávislými) shluky modelovat pomocí exponenciálního rozdělení. Vyjdeme-li, pro nějaký dostatečně velký rozsah n , z aproximace rozdělení veličiny $T(u)$ pomocí vztahu (5.17), lze hodnotu $\overline{F}(u)$ odhadnout z empirické distribuční funkce jako podíl počtu nadprahových hodnot, tj. v souladu s předchozím $\widehat{\overline{F}}(u) = \widehat{\lambda}_u$.

Autoři práce [51] navrhli momentový odhad extrémálního indexu vycházející z aproximace vztahu (5.17). Uvedený odhad je však silně závislý na prahové hodnotě, jejíž volba stále patří mezi nedořešené problémy modelování extrémních hodnot. Navíc, jak upozorňuje práce [123], aproximace veličiny $T(u)$ pomocí vztahu (5.17) není příliš vhodná, neboť tato předpokládá, že se časy mezi překročeními prahu v rámci shluků (při změně měřítka dle $\overline{F}(u)$) limitně blíží nule. Použití subasymptotické aproximace pomocí (5.17) tak vede k tomu, že časy uvnitř shluků jsou modelovány jako malé, avšak z exponenciálního rozdělení získané realizace. Práce [123] navrhuje získání odhadu θ pomocí asymptotického rozdělení veličiny $S(u) = T(u) - 1$ namísto $T(u)$, avšak tento postup vynucuje splnění celé řady dalších podmínek, které je v praxi možné ověřit jen s velkými obtížemi.

²Degenerovaným rozdělením v bodě $x = 0$ rozumíme rozdělení takové, pro které je $F(x) = 1$ pro $x \geq 0$ a $F(x) = 0$ jinak.

6

Statistická analýza pomocí teorie stacionárních řad

V této kapitole budou představené metody odhadu extrémálního indexu aplikovány na reálné datové soubory. Nejprve budou shrnuty výsledky autorem provedené simulační studie [5] a dále, v návaznosti na odstavec 4.3, budou nabyté poznatky aplikovány na měření srážkových úhrnů z jihomoravského regionu. Zde bude pozornost opět zaměřena zejména na odhady IDF křivek, přičemž důraz bude kladen na vhodnost použití vzorkovacích technik diskutovaných v kapitole 4. Závěrem bude představena autorova původní aplikace EV teorie stacionárních řad pro identifikaci odlehlých pozorování časových řad.

6.1. Simulační studie pro porovnání vlastností odhadů extrémálního indexu

Za účelem porovnání výše představených odhadů extrémálního indexu $\hat{\theta}_G, \hat{\theta}_{AT}$ byla provedena simulační studie [5]. Zde byly uvažovány dva typy stacionárních procesů (splňující $D(u_n)$ podmínku) se známými charakteristikami shluků extrémních hodnot, pomocí nichž byly generovány nezávislé realizace. Pro takto získané stacionární řady byl následně extrémální index odhadnut výše uvedenými odhady $\hat{\theta}_G$ a $\hat{\theta}_{AT}$.

Mějme posloupnost nezávislých náhodných veličin $Z_i, i = 1, 2, \dots$, se standardizovaným Fréchetovým rozdělením, tj. $F(z) = \exp(-1/z)$ pro $z > 0$ (viz tabulka 1.1). Nejprve byl uvažován max-autoregresní (maxAR) proces X_1, X_2, \dots , tedy proces ve tvaru

$$X_i = \max\{\alpha X_{i-1}, (1 - \alpha)Z_i\}, \quad i = 2, 3, \dots, \quad (6.1)$$

kde $0 \leq \alpha < 1$ a $X_1 = Z_1$. Jednoduše se odvodí, že marginální rozdělení veličin $X_i, i = 1, 2, \dots$, je opět standardizované Fréchetovo, ačkoliv se již nejedná o vzájemně nezávislé náhodné veličiny. Lze ukázat (např. viz [18]), že extrémální index maxAR procesu je $\theta = 1 - \alpha$. Jako zobecnění maxAR procesu byl uvažován i proces klouzavých maxim (KM proces), který lze zapsat ve tvaru

$$X_i = \max_{j=0, \dots, p} \{\alpha_j Z_{i+j}\}, \quad i = 1, 2, \dots, \quad (6.2)$$

kde koeficienty α_j splňují $\alpha_0 > 0, \alpha_p > 0$ a $\alpha_j \geq 0$ pro $j = 1, \dots, p-1$, a navíc $\sum_{j=0}^p \alpha_j = 1$. Veličiny X_1, X_2, \dots , pak mají opět Fréchetovo rozdělení, přičemž extrémální index KM procesu je $\theta = \max_{j=0, \dots, p} \{\alpha_j\}$ [18].

Oba diskutované odhady θ , odhad podle Gomes i odhad Ancony-Navarreteho a Tawna, jsou založeny na odhadech parametrů rozdělení blokových maxim. Uvažovaná délka bloku k tak hraje zásadní roli při použití těchto odhadů. Práce [60, 13] předpokládaly vhodnou délku bloku $k = \sqrt{n}$, kde n je rozsah stacionární řady. Daná řada je tak rozdělena na $m = \sqrt{n}$ bloků, čímž je prakticky délce bloku a počtu bloků přidělena stejná váha při odhadech θ . I přes vliv hodnoty k na odhad extrémálního indexu se ukazuje [5], že je-li délka bloku zvolena v rozumném rozpětí (např. mezi $\sqrt{n}/2$ a $2\sqrt{n}$), hraje počet přerovnáání původní řady pro získání přidruženého náhodného výběru jen malou roli. Z tohoto důvodu je možné vliv počtu přerovnáání zanedbat.

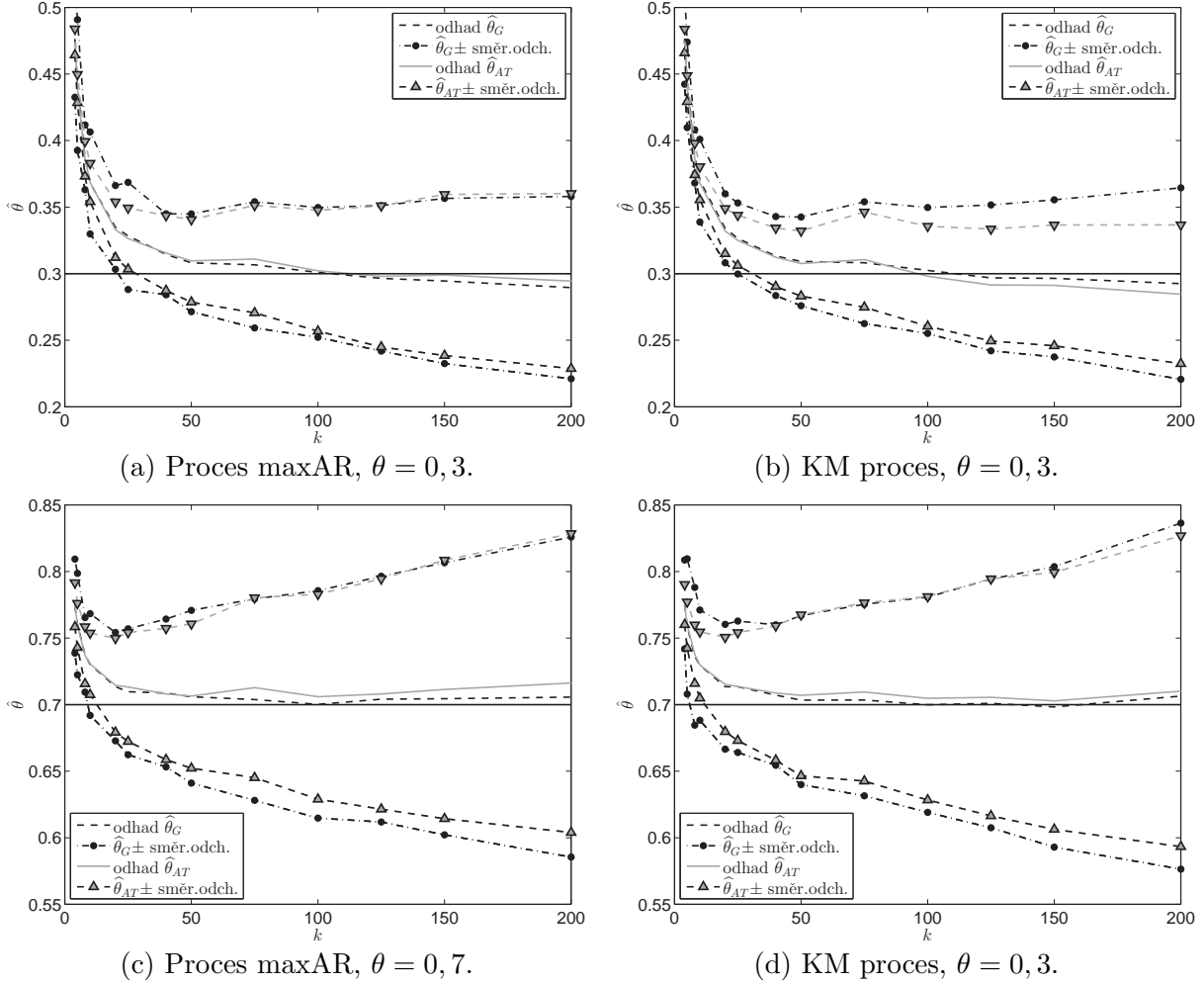
Pro účely simulační studie bylo generováno vždy $N = 1000$ realizací stacionární řady (6.1), resp. (6.2), rozsahu $n = 10000$ při hodnotách $\theta = 0, 1; 0, 2; \dots; 0, 9$. Odhady $\hat{\theta}_G, \hat{\theta}_{AT}$ byly pak získány pro různé hodnoty délky bloku k . Na obrázku 6.1 jsou zobrazeny výsledky simulační studie - bodové odhady extrémálního indexu byly získány průměrováním N nezávislých opakování a z těchto byly odhadnuty i směrodatné odchylky.

Jak lze pozorovat na obrázku 6.1, odhad $\hat{\theta}_G$ je méně vychýlený v naprosté většině případů. Na druhou stranu, pro $\theta \rightarrow 0$ či $\theta \rightarrow 1$ vychýlení $\hat{\theta}_G$ roste a $\hat{\theta}_{AT}$ je zde více robustním odhadem. Detailní výsledky pro délku bloku $k = \sqrt{n}$ jsou shrnuty v tabulce 6.1. Pro rostoucí hodnotu k (při konstantním rozsahu n) roste variabilita obou odhadů, neboť kvůli klesajícímu počtu bloků roste také variabilita odhadů parametrů GEV rozdělení. Porovnáním výsledků v tabulce 6.1 a na obrázku 6.1 je zřejmé, že variabilita odhadu $\hat{\theta}_{AT}$ je všeobecně menší než variabilita odhadu $\hat{\theta}_G$. Vzhledem k menší citlivosti na vhodnou délku bloku k lze však preferovat odhad podle Gomes.

Zároveň se z představené simulační studie ukazuje, že volba k , kterou ve své práci [60] Gomes apriori uvažovala a která byla převzata i v práci [13], je často nejvhodnějším kompromisem mezi vychýlením a variabilitou. Toto platí pro oba simulované procesy při různých hodnotách θ . Podobné výsledky byly získány i při simulacích řad menšího rozsahu, jejichž výsledky podrobně neukazujeme.

Tabulka 6.1: Srovnání odhadů $\hat{\theta}_G, \hat{\theta}_{AT}$ pro délku bloku $k = \sqrt{n} = 100$. Odhady byly získány jako průměr $N = 1000$ nezávislých opakování, směrodatné odchylky jsou udány v závorkách.

| Extremální index θ | | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|---------------------------|---------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| maxAR | $\hat{\theta}_G$ | 0.094 (0.031) | 0.197 (0.041) | 0.301 (0.049) | 0.402 (0.055) | 0.503 (0.069) | 0.608 (0.073) | 0.700 (0.085) | 0.806 (0.090) | 0.902 (0.096) |
| | $\hat{\theta}_{AT}$ | 0.095 (0.024) | 0.197 (0.037) | 0.302 (0.045) | 0.408 (0.055) | 0.506 (0.064) | 0.613 (0.069) | 0.706 (0.077) | 0.811 (0.081) | 0.900 (0.077) |
| KM | $\hat{\theta}_G$ | 0.103 (0.037) | 0.201 (0.043) | 0.302 (0.047) | 0.404 (0.053) | 0.503 (0.065) | 0.606 (0.075) | 0.700 (0.081) | 0.807 (0.095) | 0.907 (0.100) |
| | $\hat{\theta}_{AT}$ | 0.091 (0.019) | 0.192 (0.029) | 0.298 (0.038) | 0.405 (0.046) | 0.504 (0.057) | 0.610 (0.070) | 0.705 (0.076) | 0.809 (0.085) | 0.902 (0.079) |



Obrázek 6.1: Závislost odhadů extrémálního indexu $\hat{\theta}_G, \hat{\theta}_{AT}$ na délce bloku k pro stacionární řady rozsahu $n = 10000$ generované pomocí (6.1) a (6.2). Skutečná hodnota θ je vyznačena horizontální čarou.

6.2. Analýza srážkových časových řad

Přikročíme nyní k analýze datových souborů měřených dešťových srážek v jihomoravském regionu, které byly představeny v kapitole 4. Protože pozornost bude opět zaměřena především na odhady IDF křivek, byly z dostupných časových řad vyčleněny srážkové úhrny pro trvání deště odpovídající době 5, 10, 15, 20, 30, 45, 60, 90, 120, 180, 240 a 360 minut, pro které je dále možné určit intenzity dešťů jako množství srážek za danou dobu trvání. V odstavci 4.3, kde byl k analýze použit přístup POT založený na nezávislých pozorováních, byla s použitím dodatečných technik ze srážkových úhrnů vyčleněna měření, která bylo možné považovat za nezávislá. Namísto toho je možné pro danou řadu odhadnout její extrémální index, přičemž následně lze použít přístup POT pro všechna dostupná pozorování. Cílem tohoto odstavce je porovnání odhadů parametrů GP rozdělení a IDF křivek získaných metodologií POT pro přibližně nezávislá pozorování (získaná užitím metodologie [96]) a pro všechna dostupná pozorování ve formě časové řady.

Pro účely následující analýzy srážkových měření lze použít autorem implementované funkce `ems.m` a `intensity.m`. První z nich realizuje vzorkování přibližně nezávislých agregovaných pozorování podle metodiky [96] a jejím užitím lze získat řadu nezávislých

srážkových úhrnů pro různé doby trvání deště. Druhá uvedená funkce pak umožňuje z dané pozorované srážkové řady získat časovou řadu disjunktně agregovaných srážek, tj. časovou řadu srážkových úhrnů za danou dobu trvání. Obě funkce jsou detailněji popsány v dodatku D.

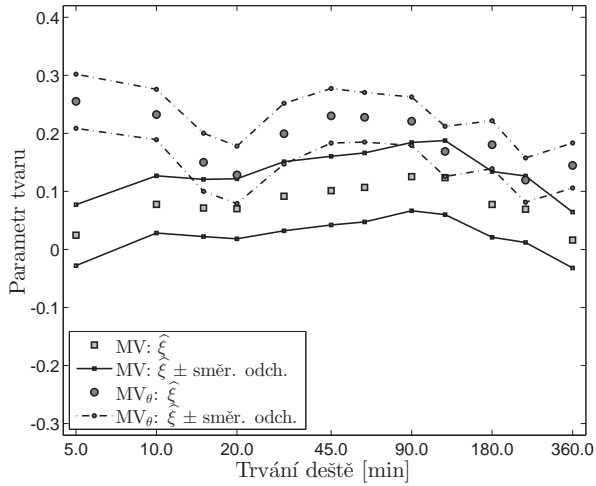
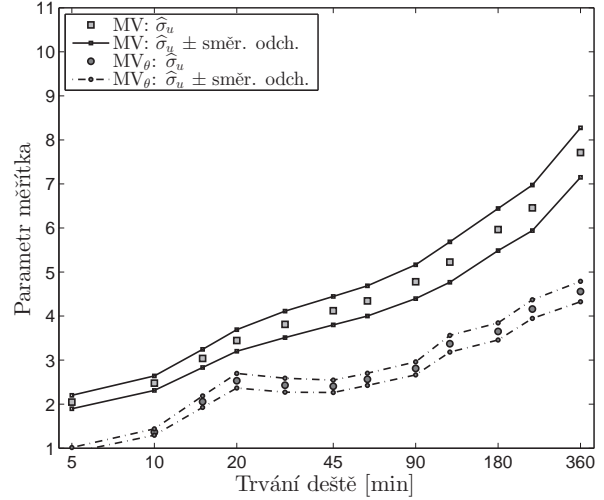
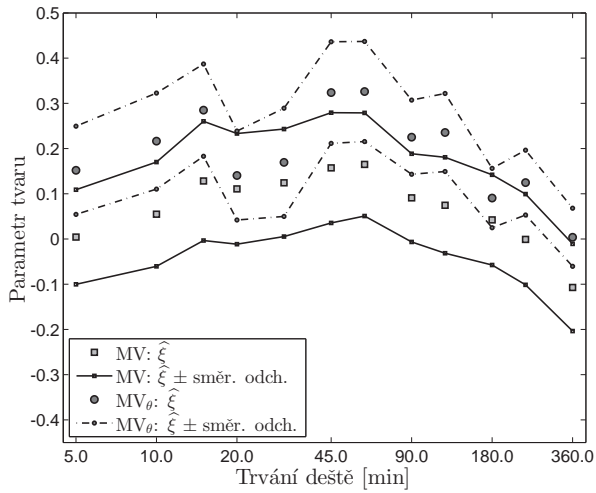
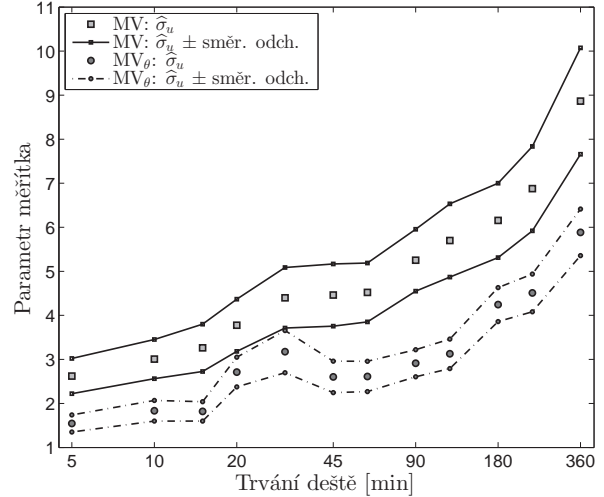
Odhady parametrů

K odhadu extrémálního indexu byl použit jak představený odhad $\hat{\theta}_G$, tak odhad $\hat{\theta}_{AT}$. Pro oba tyto odhady byla uvažována délka bloku odpovídající \sqrt{n} , kde n označuje délku dostupné časové řady. Tato volba se dle výše představené simulační studie [5] často ukazuje jako nejvýhodnější kompromis mezi vychýlením a variabilitou odhadu. Pro další účely bude nutné rozlišovat mezi parametry limitního GP rozdělení nadprahových hodnot řady nezávislých pozorování (ozn. ξ, σ_u) a původní časové řady (ozn. $\xi_\theta, \sigma_{u,\theta}$). Protože na rozdíl od analýzy z odstavce 4.3 byla pro odhady parametrů GP rozdělení použita všechna dostupná pozorování, bylo nutné pro všechny datové soubory znovu určit vhodné prahové hodnoty. Zde byly pro jednoduchost použity grafické techniky určení prahu popsané v 3.1. Odhady všech parametrů GP rozdělení byly následně určeny MV metodou.

Na obrázku 6.2 jsou zobrazeny odhady parametrů $\xi, \sigma_u, \xi_\theta, \sigma_{u,\theta}$ pro vybrané stanice Brno-Tuřany a Brno-Jundrov včetně jejich směrodatných odchylek získaných na základě asymptotické normality. Rozdíly v parametrech měřítka $\sigma_u, \sigma_{u,\theta}$ plynou z relací (2.6) a (5.4) a nelze tak očekávat jejich shodu. Na druhou stranu dle teoretických poznatků má být $\xi = \xi_\theta$. Vzhledem k velké variabilitě odhadů pro řadu ze stanice Brno-Jundrov je vzájemné vychýlení odhadů viditelné spíše pro delší dostupnou řadu ze stanice Brno-Tuřany. Nejlépe lze situaci zhodnotit pro krátké trvání deště v době 5 a 10 minut. Tyto odlišnosti mohou být způsobeny buď jinou volbou prahových hodnot nebo metodologií vzorkování přibližně nezávislých pozorování [96]. Podrobnou analýzou bylo zjištěno, že původní volba prahové hodnoty vede jen k zanedbatelnému zlepšení shody odhadů a majorantní část vzájemného vychýlení tak lze přisuzovat přístupu vzorkování. Nabízí se tak diskuze, zda je uvedená technika vzorkování [96] opravdu vhodná a zda jsou dešťové události, především pro krátké doby trvání deště, separovány dostatečně dlouhými intervaly bezsrážkového období.

Obrázek 6.3 detailně ukazuje odhady rozptylů odhadů parametrů tvaru ξ, ξ_θ a měřítka $\sigma_u, \sigma_{u,\theta}$ pro diskutované stanice Brno-Tuřany a Brno-Jundrov. Variabilita parametrů tvaru je vesměs srovnatelná bez ohledu na trvání deště a délku pozorované řady. Naopak variabilita parametrů měřítka $\sigma_{u,\theta}$ odhadnutých na základě všech dostupných pozorování je výrazně nižší než variabilita σ_u .

Pro odhady extrémálního indexu a odhady jejich variability byla použita autorem vytvořená funkce `theta_GAT.m` implementovaná v prostředí Matlab. Jak již bylo uvedeno, délka bloku k byla pro odhady θ určena jako $k = \sqrt{n}$, kde n je rozsah příslušného datového souboru. Pro stanice s nejkratší měřenou časovou řadou, tj. Brno-Jundrov a Brno-Žabovřesky, se však tato volba ukázala jako nevhodná. Pro krátkou dobu trvání deště vede malá délka bloku k problémům s konvergencí MV metody. Navíc lze očekávat, že extrémální index se při změně doby trvání deště bude měnit dostatečně hladce. Zde se proto volba bloku $k = \sqrt{n}$ ukázala jako nevyhovující a v těchto případech musela být hodnota k adaptována tak, aby pro krátké doby trvání deště bylo k vyšší. Vzhledem k požadavku hladkosti průběhu extrémálního indexu byla pro jednotlivé doby trvání deště 5; 10; ...; 360 minut zvolena hodnota $k = 2\sqrt{n}; 1, 8\sqrt{n}; 1, 6\sqrt{n}; \dots; \sqrt{n}; \sqrt{n}; \sqrt{n}$. Průběh

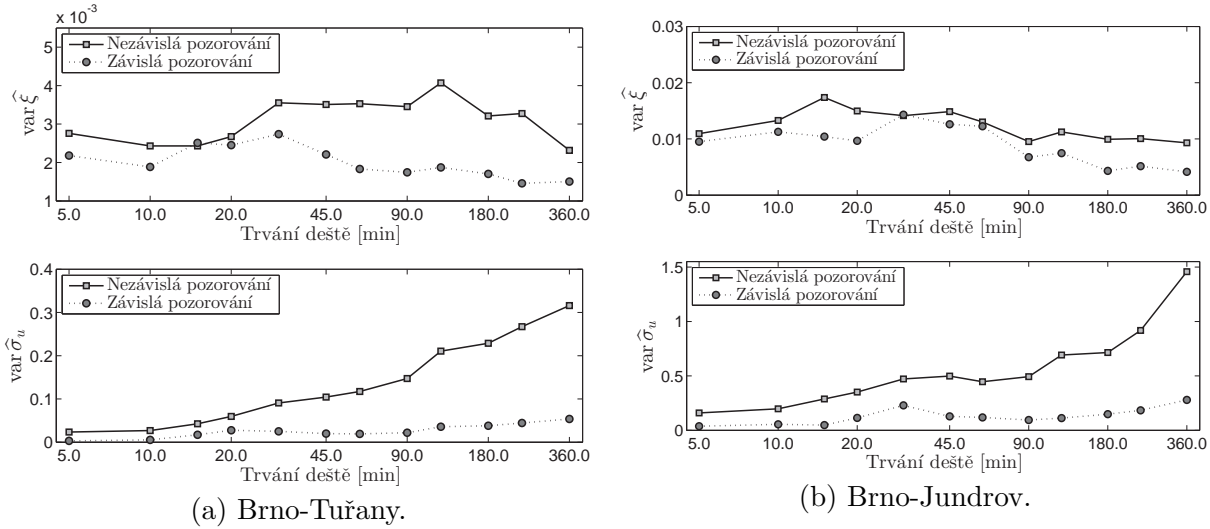
(a) Brno-Tuřany, parametr tvaru ξ .(b) Brno-Tuřany, parametr měřítka σ_u .(c) Brno-Jundrov, parametr tvaru ξ .(d) Brno-Jundrov, parametr měřítka σ_u .

Obrázek 6.2: Parametry tvaru a měřítka získané MV metodou pro nezávislá vzorkovaná pozorování a všechna (závislá) pozorování. Směrodatné odchylky byly získány na základě asymptotické normality. Horizontální osa je zobrazena v logaritmickém měřítku.

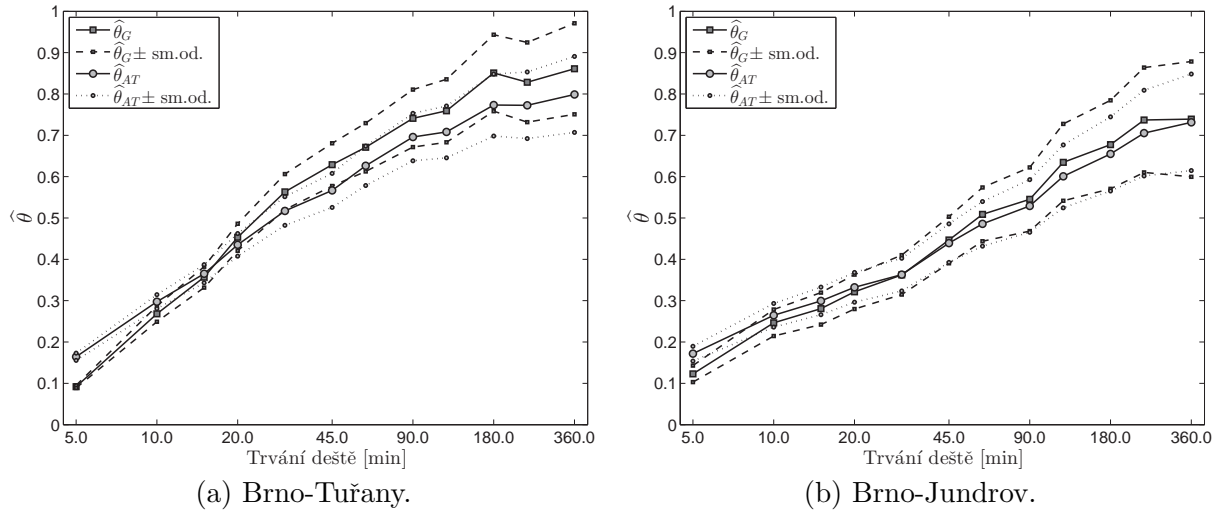
odhadů extrémálního indexu $\hat{\theta}_G$ a $\hat{\theta}_{AT}$ je zobrazen na obrázku 6.4. Jak bylo očekáváno, roste hodnota θ spolu s dobou trvání deště. Dochází tedy k menšímu shlukování extrémních hodnot a závislost mezi jednotlivými nadprahovými pozorováními stále více klesá. Na obrázku 6.4 jsou také zobrazeny odhady směrodatných odchylek odhadů extrémálního indexu, přičemž tyto byly získány na základě asymptotické normality dle odstavce 5.2.

Odhady IDF křivek

Odhady parametrů GP rozdělení a extrémálního indexu byly dále použity pro odhady IDF křivek. Obrázek 6.5 ukazuje odhady 5- a 100letých návratových úrovní pro jednotlivá trvání deště pro stanice Brno-Tuřany a Brno-Jundrov. Jsou zde ukázány návratové úrovně odhadnuté pomocí závislých (tj. všech) pozorování časové řady (ozn. ZP) a pomocí přibližně nezávislých pozorování vzorkovaných metodikou [96] (ozn. NP). V případě časové řady zde byl použit odhad extrémálního indexu podle Gomes. IDF křivky byly následně získány proložení těchto návratových úrovní vhodnou regresní křivkou (viz



Obrázek 6.3: Odhady rozptylů MV odhadů parametrů tvaru a měřítka GP rozdělení získané analýzou nezávislých vzorkovaných pozorování a všech (závislých) pozorování.



Obrázek 6.4: Odhady extrémálního indexu $\hat{\theta}_G, \hat{\theta}_{AT}$ a jejich směrodatné odchylky určené pro časové řady srážkových úhrnů podle trvání deště.

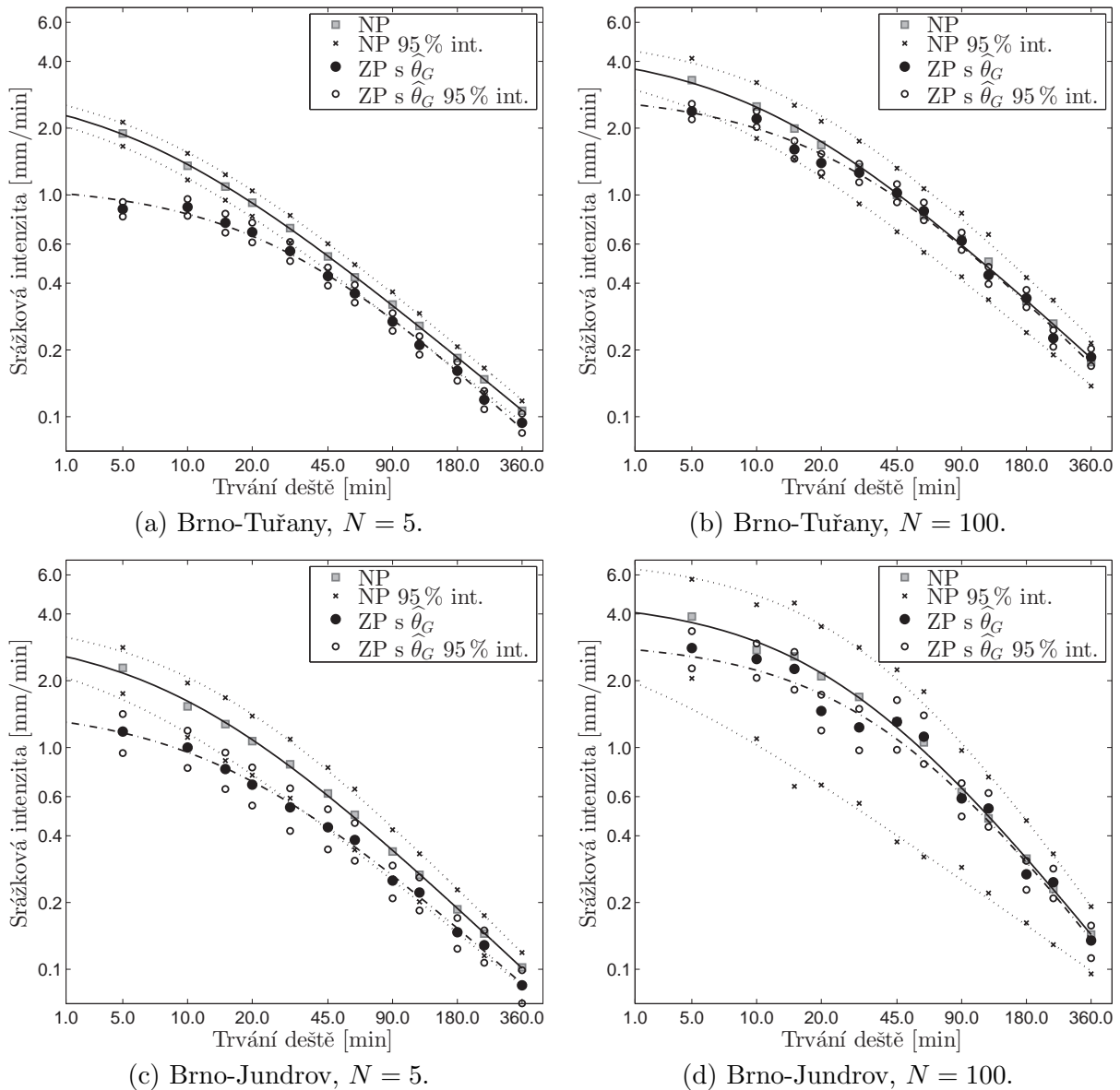
odstavec 4.3). Stejně tak byly proloženy 95% intervaly spolehlivosti návratových úrovní získaných z nezávislých pozorování. Intervaly spolehlivosti obdržené v případě použití všech pozorování nebyly proloženy z důvodu jednodušší orientace.

Při odhadu extrémálního indexu pomocí odhadu Ancony-Navarreteho a Tawna $\hat{\theta}_{AT}$ byly získány odhady návratových úrovní a jejich intervalů spolehlivosti, které jsou velmi podobné těm, kdy je použit odhad $\hat{\theta}_G$. Shoda je zřejmá z průběhu obou odhadů na obrázku 6.4 a z tohoto důvodu zde výsledky podrobně nepředstavujeme.

Na první pohled je zřejmé, že použití všech dostupných pozorování vede k významné redukci variability odhadů návratových úrovní bez ohledu na délku měřené řady. Na rozdíl od nezávislých pozorování v případě analýzy původních časových řad vstupuje do odhadu také variabilita odhadu extrémálního indexu. Vhodnější odhady θ by tak dále mohly přispět ke snížení variability odhadů návratových úrovní.

Na druhou stranu je jasné vidět vzájemné vychýlení odhadů návratových úrovní. Typicky se jedná o případy odpovídající době trvání deště mezi 5 a 45 minutami. Tyto

neshody mohou být opět v principu způsobeny buď nepřesností odhadu θ či chybou některého modelu, zejména pak nevhodností výběru přibližně nezávislých pozorování metodikou [96]. Obrázek 6.4 však ukazuje, že zejména pro krátká trvání deště je variabilita odhadů extrémálního indexu velmi malá. Proto se jako pravděpodobnější jeví nevhodnost metodiky [96]. Připomeňme, že pomocí tohoto přístupu je časová řada rozdělena na dešťové události. Jako nezávislé události jsou pak vybrány ty, které jsou odděleny bezdešťovým obdobím o délce alespoň $\min(60, t)$ minut, kde t je sledované trvání deště. Jedno z možných doporučení při použití techniky [96] by tak mohla být úprava separačního kritéria, aby byly dešťové události odděleny delší periodou (jak vesměs na obrázku 6.5 ukazují situace pro delší trvání deště).



Obrázek 6.5: MV odhady IDF křivek pro různé N -leté návratové úrovně získané analýzou nezávislých pozorování (NP; plná čára) a závislých pozorování s odhadem $\hat{\theta}_G$ (ZP; čerchovaná čára). IDF křivky byly proloženy regresní křivkou, 95% intervaly spolehlivosti byly proloženy jen pro případ NP. Osy jsou v logaritmickém měřítku.

Pro 5leté návratové úrovně je také patrné, že na základě analýzy celé časové řady dochází k celkovému podhodnocení odhadovaných IDF křivek. Na druhou stranu jsou tyto odhady, vyjma diskutovaných krátkých dob trvání deště, obsažené v 95% intervalech spolehlivosti pro odhady získané pomocí nezávislých pozorování a naopak. Zde je otevřena cesta dalšímu výzkumu, neboť vhodnější odhady extrémálního indexu by mohly dále přispět ke zvýšení přesnosti odhadů IDF křivek.

6.3. Validace měření ve formě časových řad

V tomto odstavci popíšeme jiné praktické použití EV teorie stacionárních řad. V environmentálních vědách, ale také v jiných technických a ekonomických oblastech, se často vyskytuje požadavek validace získaných měření časových řad. Typicky se jedná o nutnost identifikace odlehlých pozorování, jež nebyla způsobena vlastní sledovanou veličinou. Tato pozorování mohou být zapříčiněna např. chybami měřících přístrojů či jinými vnějšími vlivy. Analýza původní časové řady přitom může vést k nesprávným odhadům vlastností původní sledované veličiny. Běžně sice bývají získaná pozorování doprovázena také metadaty, která popisují vybrané podmínky měření (zejména funkčnost jednotlivých přístrojů a dále pozorování některých doprovodných veličin), přesto je ale v praktických situacích často nutné provést dodatečnou validaci na základě empirických zkušeností a konkrétních znalostí v daném oboru. Někdy pak metadata nejsou dostupná vůbec. V tomto odstavci bude proto pozornost zaměřena na identifikaci odlehlých pozorování v časových řadách, která chápeme jako řídké se vyskytující extrémní jevy. Představená metoda validace pak může sloužit buď jako plně samostatná automatická technika, případně jako technika pomocná, jejímž účelem je upozornit na „podezřelá“ měření, o kterých je potřeba rozhodnout dodatečným posouzením.

Validaci dat za použití EV teorie není v literatuře dosud věnována velká pozornost. Základní shrnutí metod identifikace odlehlých hodnot publikovaných do roku 2004 lze nalézt v práci [73], užití EV teorie se pak věnují například publikace [113, 110, 26]. Téměř vždy se však autoři omezují na případy Gaussovských procesů a předpokládají obor atraktivity v rámci Gumbelovy třídy EV rozdělení.

Níže navržený postup je založen na odhadech parametrů GEV rozdělení stacionární řady, které byly představeny v kapitole 5. Při rozsáhlých a dlouhotrvajících měřeních sledované náhodné veličiny většinou není splněn předpoklad stacionarity příslušné časové řady. Zde se omezíme na případy, kdy je stacionarita porušena jen z důvodu nestacionární střední hodnoty řady, tj. pro X_1, \dots, X_n máme $m = m_t, 1 \leq t \leq n$, kde $m_t := EX_t$. Je-li k dispozici nějaká dodatečná informace o této střední hodnotě, je možné odhadnout její průběh pomocí vybraného parametrického modelu např. užitím regresní analýzy. Ve většině případů je však tento průběh neznámý nebo je jeho popis natolik komplikovaný, že lze parametrické modely užít jen se značnými omezeními. Jako vhodnější se pak jeví aplikace neparametrických technik, které budou uvažovány i v případě níže.

Popis dat

K analýze byly vybrány časové řady měření polétavého prachu s označením PM_{10} , tedy pevných částic menších než $10 \mu m$. Vysoké koncentrace PM_{10} mají významný vliv na zdraví člověka, protože jsou bez větších potíží schopny proniknout až do dolních cest

dýchacích. Zde pak způsobují podráždění dýchacích cest, akutní respirační infekce, případně mohou přispět k rozvoji vážných onemocnění jako je např. rakovina či chronická obstrukční plicní nemoc [83]. Z těchto důvodů se jeví sledování PM_{10} jako důležité a hodnoty koncentrace PM_{10} tak patří k významným ukazatelům znečištění ovzduší. Podle směrnice Rady Evropské unie [111] je pak vyžadováno, aby denní průměr PM_{10} nepřekračoval hranici $50 \mu\text{g}/\text{m}^3$ po více než 35 dní v kalendářním roce.

Měření PM_{10} byla automaticky prováděna na pěti lokacích v městě Brně umístěných v oblastech se zvýšenou zátěží na kvalitu ovzduší, přičemž časové rozlišení měřicího přístroje bylo nastaveno na 1 hodinu. Konkrétně se jedná o stanice Arboretum, Lány, Svatoplukova, Výstaviště a Zvonařka. Dostupné byly časové řady za období od 11/2007 až 11/2015 pro poslední tři uvedené stanice, 11/2006 až 11/2015 pro stanici Arboretum a 9/2002 až 11/2015 pro stanici Lány. V tomto odstavci budou ukázány jen výsledky pro první uvedenou stanici Arboretum, bližší informace lze dále nalézt v práci [8]. Některé postupy k analýze koncentrací PM_{10} měřených v brněnském regionu lze nalézt kupříkladu v práci [76], kde se autoři zabývali identifikací faktorů přispívajícím ke znečištění v městě Brně. V [78] lze pak nalézt modely k predikci hodnot PM_{10} .

Jádrové vyhlazování

Uvažujme řadu náhodných veličin X'_1, \dots, X'_n s nekonstantní střední hodnotou m_t , $1 \leq t \leq n$. Následující postup identifikace odlehlých pozorování lze shrnout do několika základních bodů. Nejprve se vybranými neparametrickými přístupy určí odhad \hat{m}_t střední hodnoty m_t . Je-li stacionarita původní řady porušena jen nekonstantností střední hodnoty, pak lze nově vzniklou řadu

$$X_t = X'_t - \hat{m}_t, \quad 1 \leq t \leq n,$$

považovat již za přibližně stacionární řadu. Zde je pak možné metodami odstavce 5 odhadnout parametry příslušného rozdělení extrémních hodnot výběrových maxim včetně odhadu extrémálního indexu θ . Na základě těchto odhadů je možné dále určit odhady zvolené parametrické funkce, zejména pak návratové úrovně z_r příslušné periodě návratu r pozorování. Kombinací hodnot $\hat{m}_t + \hat{z}_r$ pro dané r a $1 \leq t \leq n$ obdržíme v jistém smyslu intervalový odhad pro extrémní hodnoty původní časové řady. Podobně lze postupovat v případě identifikace odlehlých nízkých pozorování a určit tak intervalový odhad řady pro minimální hodnoty. Z teoretického hlediska se uvedený postup jeví jako vhodnější než běžně užívané intervaly spolehlivosti založené na teorii nezávislých veličin, neboť zavedením extrémálního indexu je brána v potaz také závislost mezi jednotlivými pozorováními.

Pro účely neparametrického odhadu průběhu střední hodnoty byla pozornost zaměřena na techniku jádrového vyhlazování. Vzhledem k zaměření této práce budou shrnuty jen základní pojmy tohoto přístupu, více o jádrových odhadech je možné nalézt např. v knize [133]. *Jádrovou funkcí* se rozumí symetrická funkce $K(x)$, pro kterou platí $\int_{-\infty}^{\infty} K(x) dx = 1$. Pro naše účely se dále omezíme na jádrové funkce s nosičem $\text{supp}(K) = \langle -1, 1 \rangle$. Pro $h > 0$ označme $K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right)$. Parametr h je označován jako *šířka vyhlazovacího okna*. Nosičem funkce $K_h(x)$ je tedy interval $\langle -h, h \rangle$. Jádrový odhad funkce m_t v bodě t_0 lze obecně zapsat ve tvaru

$$\hat{m}_{t_0}(h) = \sum_{t=1}^n W_t(t_0, h) X'_t,$$

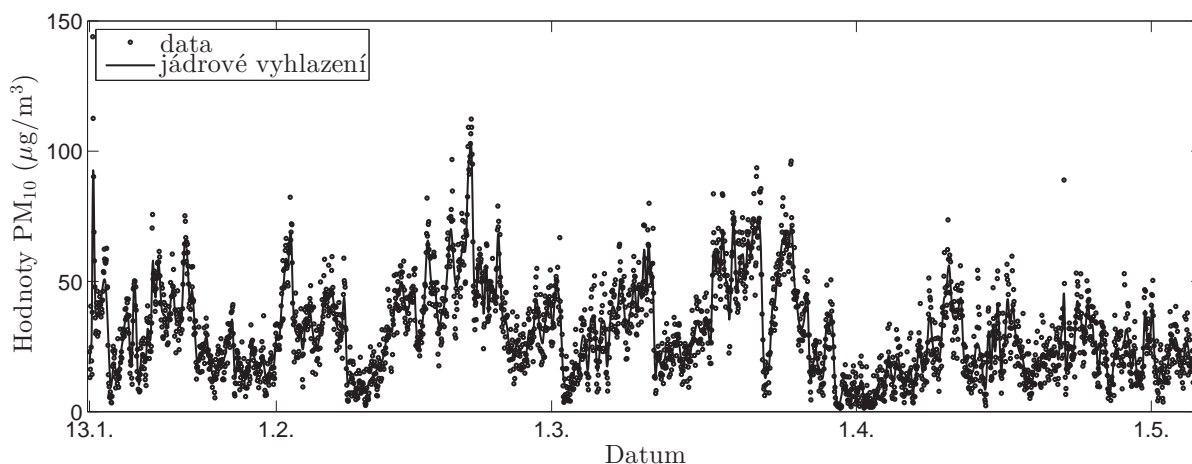
kde váhy W_t pro $1 \leq t \leq n$ závisejí na šířce vyhlazovacího okna. Jádrový odhad hledané střední hodnoty m_t v bodě t_0 se tedy dostane jako vážený průměr pozorování ležících v okolí t_0 . Různým nastavením vah W_t se dostanou různé jádrové odhady funkce m_t . Mezi nejvýznamnější patří především Nadarayův-Watsonův odhad [100], lokálně lineární odhad či Gasserův-Müllerův odhad [58]. Zde bude použit poslední uvedený, přičemž bude zvolena Epanečnikova jádrová funkce (viz [133]), tedy funkce

$$K(x) = \frac{3}{4}(1 - x^2)\mathbf{1}_{[|x| \leq 1]}.$$

Případnou diskuzi o vhodnosti tohoto jádra vynecháme a odkazujeme na [8]. Z pohledu mnoha statistických vlastností je však Epanečnikovo jádro optimální (viz [133]) a patří mezi nejčastěji užívané jádrové funkce.

Volba parametru h ovlivňuje hladkost odhadu \hat{m}_t a jeho volba je zásadním problémem jádrových vyhlazovacích metod. Velká hodnota h vede k přehlazení (a to k průměru dat), h malé pak způsobuje malé vyhlazení (vedoucí až k reprodukci dat). Na základě teoretických úvah je pro optimální hodnotu h^* často kladen požadavek, aby byla minimalizována střední integrální kvadratická chyba odhadu \hat{m}_t . V praxi se běžně postupuje tak, že se h volí metodou křížového ověřování [133], přičemž tento odhad bývá zpravidla blízký h^* . Některé pokročilé postupy spočívají v adaptivní volbě lokální šířky okna v závislosti na průběhu konkrétní časové řady [72]. Tento postup byl použit také v publikaci [8], kde je možné nalézt doplňující informace k technice zde provedeného jádrového vyhlazení.

Odhad funkce m_t pomocí jádrové vyhlazení podle [8] je zobrazen na obrázku 6.6. Konkrétně se jedná o stanici Arboretum a časový úsek od 13. ledna 2015 do 6. května 2015. Na odhady návratových úrovní extrémních hodnot bude pozornost zaměřena níže.



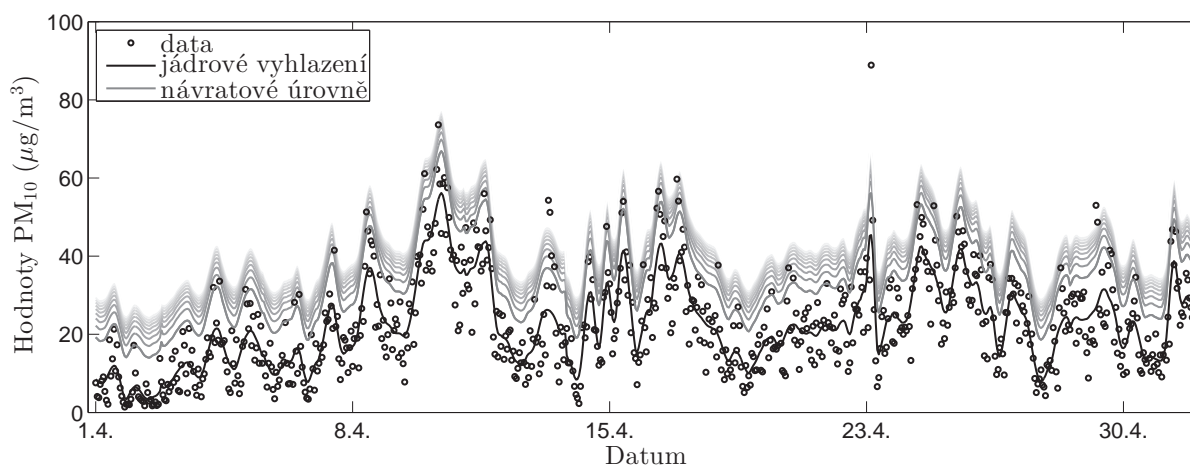
Obrázek 6.6: Odhad střední hodnoty řady PM_{10} pomocí jádrového vyhlazení pro stanici Brno-Arboretum v období od 13. ledna 2015 do 6. května 2015 (více viz [8]).

Identifikace odlehlých hodnot

Odečtením odhadu střední hodnoty \hat{m}_t byla získána řada X_1, \dots, X_n . Pro odhady extrémních hodnot byl opět zvolen prahový POT model v kombinaci s MV metodou pro odhady parametrů GP rozdělení. Nejprve byla vhodná volba prahové hodnoty u stanovena pevně pomocí výběrového 90% kvantilu, dále byla použita adaptivní technika MT-GP v kombinaci se skórovou statistikou (3.16) popsaná v odstavci 3.3. Uvedené techniky výběru prahu

byly zvoleny zejména pro možnost snadné automatizace odhadu a tudíž i validace dat. Shoda rozdělení nadprahových hodnot s GP rozdělením byla vždy ověřena jednak vizuálně (histogramy, Q-Q ploty) a také testy dobré shody (χ^2 , K-S a A-D testy). Pro žádný z uvedených testů přitom nebyla shoda rozdělení na hladině významnosti 0,05 zamítnuta. K odhadu extrémálního indexu θ byl uvažován odhad podle Gomes $\hat{\theta}_G$, přičemž délka bloku k byla zvolena na základě simulační studie [5] $k = \sqrt{n}$.

Pro získané odhady všech parametrů byla stanovena návratová úroveň \hat{z}_r pro periody návratu $r = 24, 48, \dots, 240$ hodin. Tyto návratové úrovně tedy odpovídají hladině koncentrací PM₁₀, které jsou v průměru překročeny jednou za 1, 2, \dots , 10 dní. Výsledky analýzy měření PM₁₀ pro stanici Brno-Arboretum a období měsíce dubna jsou ukázány na obrázku 6.7. Při volbě prahu pomocí 90% výběrového kvantilu ($u = 8,14 \mu\text{g}/\text{m}^3$)



Obrázek 6.7: Odhady návratových úrovní extrémních hodnot pro řadu PM₁₀ při volbě prahové hodnoty jako 90% výběrového kvantilu reziduí jádrového vyhlazení. Zobrazena je stanice Brno-Arboretum v průběhu měsíce dubna 2015. Uvažovány byly periody návratu příslušné $r = 24, 48, \dots, 240$ pozorováním.

byl počet nadprahových hodnot v měsíci dubnu pro jednotlivá $r = 24, 48, \dots, 240$ stanoven 43, 20, 11, 10, 7, 7, 6, 5, 4 a 4. Poslední uvedená překročení 10denní návratové úrovně byla identifikována pro pozorování ve dnech 13.4. v 8:00 a 9:00, 22.4. ve 3:00 a 28.4. v 6:00 a na obrázku 6.7 je možné je jasně identifikovat. Použití adaptivní techniky MT-GP vedlo k volbě prahové hodnoty $u = 11,18 \mu\text{g}/\text{m}^3$ a za stejné období byly jednotlivé počty překročení určeny 41, 21, 15, 11, 10, 8, 7, 6, 6 a 5. Na všech úrovních tedy dochází oproti předešlému postupu ke zvýšení počtu pozorování, která překračují danou návratovou úroveň. Pro 10denní návratovou úroveň tak došlo navíc k označení měření ze dne 28.4. v 7:00 hodin, které bylo v dříve uvedeném případě blízko 10denní hranici.

Stanovení konkrétní periody r návratové úrovně z_r je nutné založit na znalostech fyzikálních a dalších procesů a její volba tak převážně závisí na expertních znalostech v daném oboru. Celkově však obě metody určení prahu vedou k porovnatelným závěrům. Analýzou delších časových úseků byla také pozorována velmi dobrá shoda mezi očekávanými a skutečnými počty překročení všech odhadů návratových úrovní, což svědčí o vhodnosti uvedeného postupu k označení extrémních hodnot pozorovaných se stanovenou frekvencí.

Představená práce shrnuje soudobé poznatky teorie extrémních hodnot a podtrhuje tak jejich aktuálnost při řešení praktických inženýrských úloh. Převážně jsou diskutována dvě zásadní témata spojená s odhady frekvence výskytů extrémních jevů. V první řadě se jedná o problematiku spojenou s volbou vhodné prahové hodnoty, která stále patří k nedořešeným otázkám prahového modelu, zároveň je však kritickou částí celého modelu. V současnosti roste zájem o vývoj adaptivních přístupů volby prahu, jejichž výhoda spočívá jednak v automatizaci celého procesu a především pak v absenci subjektivního rozhodování, které je součástí běžných praktických postupů. Významná třída adaptivních metod, dvojitě bootstrapová technika poprvé navržená Hallem [70], je spojena se semi-parametrickými odhady parametrů rozdělení extrémních hodnot. Jiný přístup k určení vhodného prahu (označovaný MT-GP) publikovali Northrop a Coleman [102] a tento je založen na asymptotických, maximálně věrohodných testech s rušivými parametry. Obě tyto metody jsou postaveny na pevných teoretických základech, na druhou stranu vyžadují značné výpočetní kapacity.

Za účelem srovnání těchto přístupů byla provedena simulační studie. Skrze odhady EV indexu se ukazuje, že obě techniky volby prahu vedou k porovnatelným odhadům v rámci Hallovy třídy rozdělení. Mimo Hallovu třídu se pak ukazuje, že dvojitě bootstrapová metoda přispívá k menšímu vychýlení odhadů, ačkoliv toto zde nemá jasné teoretické opodstatnění. Následně byly získané poznatky použity k analýze reálných srážkových úhrnů měřených v jihomoravském regionu. Cílem bylo navázat na předchozí studie zhotovené za účelem aktualizace IDF křivek, důležitého hydrologického nástroje pro ohodnocení extrémních dešťů. V tomto ohledu se ukazuje, že použití adaptivních technik určení prahu může vést ke snížení variability odhadů tam, kde jsou k dispozici jen krátké časové řady. Zejména lze tyto přístupy doporučit pro odhady dlouhodobých extrémních událostí. V případě dostatečně velkého počtu pozorování jsou však odhady získané na základě adaptivních metod srovnatelné s běžnými grafickými technikami určení prahu, přičemž tyto zpravidla nabízí jednodušší použití a nevyžadují takové výpočetní kapacity.

V druhé části se práce zabývá zobecněním teorie extrémních hodnot pro stacionární řady. Ačkoliv základy této oblasti byly položeny již před několika desetiletími, k rozvoji dochází až nyní s nově se objevujícími odhady extrémního indexu. Zde bylo autorem práce provedeno srovnání dvou nejvýznamnějších odhadů rozvinutých v poslední době a následně byla těmito pokročilými technikami opětovně analyzována předchozí srážková data. Zásadní zjištění bylo odhaleno v nedostacích vzorkovacích technik, které bývají používány při vyhodnocení pozorování pomocí představené teorie nezávislých veličin. S ohledem k odhadům IDF křivek dochází aplikací těchto technik ke značnému vychýlení odhadů, navíc použitím teorie pro stacionární řady lze významně redukovat variabilitu odhadů. V tomto směru jsou v práci učiněna některá praktická doporučení.

V závěru se práce zabývá použitím teorie extrémních hodnot stacionárních řad pro účely validace časových řad. Kombinací s neparametrickými technikami vyhlazování je možné pro nestacionární časové řady označit odlehlá pozorování, která lze v tomto kontextu chápat jako extrémní jevy. Uvedený postup má pak daleko větší teoretické opodstatnění než běžně užívané metody, např. založené na intervalech spolehlivosti určených pomocí jednorozměrné teorie nezávislých veličin.



Odvození vybraných pomocných funkcí podmínek atraktivity a druhého řádu

V této části budou odvozeny pomocné funkce $a(t)$, resp. $A(t)$, které vystupují v podmínce atraktivity, resp. podmínce druhého řádu. Podmínka atraktivity je nutnou a postačující podmínkou pro dané rozdělení s kvantilovou funkcí chvostu $U(t)$ tak, aby patřilo do oboru atraktivity $\mathcal{D}(G_\xi)$ některého rozdělení extrémních hodnot. Podmínka atraktivity v jednom z jejích ekvivalentních tvarů je dána následovně

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = D_\xi(x) := \frac{x^\xi - 1}{\xi}, \quad (\text{A.1})$$

kde $a(t)$ je nějaká kladná funkce a daná rovnost platí pro všechna $x > 0$. V případě $\xi = 0$ chápeme pravou stranu rovnosti jako $\ln x$, kterou dostaneme limitním přechodem.

Jak je odvozeno např. v [38], univerzální volbu funkce $a(t)$ pro příslušné obory atraktivity je možno provést následujícím způsobem:

$$a(t) = \begin{cases} \xi U(t), & \text{pro } \xi > 0, \\ g(U(t)), & \text{pro } \xi = 0, \\ -\xi(x^* - U(t)), & \text{pro } \xi < 0, \end{cases} \quad (\text{A.2})$$

přičemž x^* označuje pravý koncový bod rozdělení s distribuční funkcí $F(x)$ a funkce $g(t)$ je definována jako

$$g(t) = \frac{\int_t^{x^*} (1 - F(s)) ds}{1 - F(t)}.$$

Autoři publikace [132] ukazují možnost volit

$$a(t) = h(U(t)) \quad \text{pro } \xi \in \mathbb{R}, \quad (\text{A.3})$$

při

$$h(t) = \frac{1 - F(t)}{F'(t)}.$$

V konkrétních případech je však volba pomocné funkce $a(t)$ dle výše uvedených vztahů často velmi výpočetně náročná, mnohdy nemožná či nevede k jednoduchému ověření podmínky atraktivity. Zde je pak nutné použít některé dodatečné techniky, které budou diskutovány v příslušných případech.

Pro připomenutí uvedme také podmínku druhého řádu. Zde je navíc vyžadováno, aby existovala nějaká funkce $A(t)$, která nemění znaménko a navíc $\lim_{t \rightarrow \infty} A(t) = 0$. Pak $U(t)$ splňuje podmínku druhého řádu, jestliže pro všechna $x > 0$ platí

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx) - U(t)}{a(t)} - \frac{x^\xi - 1}{\xi}}{A(t)} = H_{\xi, \rho}(x) := \frac{1}{\rho} \left(\frac{x^{\xi + \rho} - 1}{\xi + \rho} - \frac{x^\xi - 1}{\xi} \right), \quad (\text{A.4})$$

kde $\rho \leq 0$ je parametr druhého řádu. Pro případy $\xi = 0$ nebo $\rho = 0$ lze pravou stranu rovnosti opět obdržet limitním přechodem ve tvaru (1.37) (str. 31).

Jak ukazuje důsledek 1.12, lze nalézt takové funkce $a_0(t)$, $A_0(t)$, že

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx) - U(t)}{a_0(t)} - D_\xi(x)}{A_0(t)} = \Psi_{\xi, \rho}(x), \quad (\text{A.5})$$

kde $\Psi_{\xi, \rho}(x) = \frac{x^{\xi + \rho} - 1}{\xi + \rho}$, resp. některý z jejích limitních tvarů pro $\xi = 0$ či $\rho = 0$. Práce [38] ukazuje, že volbu pomocné funkce a_0 lze provést následovně:

$$a_0(t) := \begin{cases} ct^\xi, & \rho < 0, \\ -\xi(U(\infty) - U(t)), & \xi < \rho = 0, \\ \xi U(t), & \xi > \rho = 0, \\ \tilde{U}(t) + \bar{U}(t), & \xi = \rho = 0, \end{cases} \quad (\text{A.6})$$

kde $c := \lim_{t \rightarrow \infty} t^{-\xi} a(t) > 0$, $U(\infty) := \lim_{t \rightarrow \infty} U(t)$, funkce $\bar{U}(t)$ je definována vztahem

$$\bar{U}(t) := \begin{cases} U(t) - c \frac{t^\xi - 1}{\xi}, & \rho < 0, \\ t^{-\xi}(U(\infty) - U(t)), & \xi < \rho = 0, \\ t^{-\xi} U(t), & \xi > \rho = 0, \\ \tilde{U}(t), & \xi = \rho = 0, \end{cases} \quad (\text{A.7})$$

přičemž pro nějakou integrovatelnou funkci g bylo zavedeno značení

$$\tilde{g}(t) := g(t) - \frac{1}{t} \int_0^t g(s) ds.$$

Funkci A_0 můžeme pak zvolit v následující tvaru

$$A_0(t) := \begin{cases} -(\xi + \rho) \frac{\bar{U}(\infty) - \bar{U}(t)}{a_0(t)}, & \xi + \rho < 0, \rho < 0, \\ (\xi + \rho) \frac{\bar{U}(t)}{a_0(t)}, & \xi + \rho > 0, \rho < 0, \\ \frac{\tilde{U}(t)}{a_0(t)}, & \xi + \rho = 0, \rho < 0, \\ \frac{\bar{U}(t)}{a_0(t)}, & \xi \neq \rho = 0, \\ \frac{\tilde{U}(t)}{a_0(t)}, & \xi = \rho = 0. \end{cases} \quad (\text{A.8})$$

Rozdělení z Gumbelova oboru atraktivity

Exponenciální rozdělení. Uvažujme exponenciální rozdělení s parametrem $\lambda > 0$ a s distribuční funkcí

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

Kvantilová funkce chvostu je ve tvaru $U(t) = \frac{1}{\lambda} \ln t$ pro $t > 1$. Protože zřejmě

$$U(tx) - U(t) = \frac{1}{\lambda} \ln x,$$

nabízí se jednoduchá volba funkce $a(t)$ tak, aby byla splněna rovnost (A.1) pro pravou stranu $\ln x$, a sice konstantní funkce $a(t) = \lambda^{-1}$. V podmínce (A.1) obdržíme pravou stranu ve tvaru $\ln x$ a zřejmě tedy exponenciální rozdělení patří do Gumbelova oboru atraktivity (tj. $\xi = 0$).

Nyní hledíme tvar funkce $a(t)$ pomocí výše uvedených vztahů (A.2) a (A.3). V prvním případě lze tak jednoduchým výpočtem získat $g(t) = \lambda^{-1}$ a tedy volba $a(t)$ dle (A.2) je shodná s předešlou volbou $a(t) = \lambda^{-1}$. Ve druhém případě dle (A.3) obdržíme opět $h(t) = \lambda^{-1}$ a volba je identická.

Protože pro exponenciální rozdělení nastává v podmínce (A.1) rovnost přímo, je čitatel v podmínce (A.4) roven nule. Není tak možné získat pravou stranu v požadovaném tvaru funkce $H_{\xi, \rho}(x)$ a tedy exponenciální rozdělení nesplňuje podmínku druhého řádu. V tomto případě píšeme $\rho = -\infty$.

Logistické rozdělení. Uvažujme logistické rozdělení s parametry $\mu \in \mathbb{R}$, $\sigma > 0$ a s distribuční funkcí

$$F(x) = \frac{1}{1 + e^{-\frac{x-\mu}{\sigma}}}.$$

Kvantilová funkce chvostu je dána jako $U(t) = \mu + \sigma \ln(t-1)$ pro $t > 1$. Dle (A.2) je tak možné provést volbu $a(t) = g(U(t))$, kde

$$g(t) = \sigma \left(1 + e^{\frac{t-\mu}{\sigma}} \right) \ln \left(1 + e^{-\frac{t-\mu}{\sigma}} \right),$$

a tedy

$$a(t) = \sigma t \ln \left(\frac{t}{t-1} \right). \quad (\text{A.9})$$

Následně ověříme podmínku atraktivity. Protože $U(tx) - U(t) = \sigma \ln \left(\frac{tx-1}{t-1} \right)$, můžeme psát

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \lim_{t \rightarrow \infty} \frac{\ln \left(\frac{tx-1}{t-1} \right)}{t \ln \left(\frac{t}{t-1} \right)} = \ln x,$$

neboť $t \ln \left(\frac{t}{t-1} \right) \rightarrow 1$ pro $t \rightarrow \infty$. Tudíž $a(t)$ daná vztahem (A.9) je vhodnou funkcí a zřejmě logistické rozdělení patří do Gumbelova oboru atraktivity. Jednodušší tvar $a(t)$ lze získat pomocí vztahu (A.3). Zde lze ukázat, že

$$h(t) = \sigma \left(1 + e^{-\frac{t-\mu}{\sigma}} \right)$$

a následně tedy $a(t) = \sigma t/(t-1)$. Opět ověříme platnost podmínky (A.1). Protože pro $t \rightarrow \infty$ je $t/(t-1) \rightarrow 1$, můžeme psát

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \lim_{t \rightarrow \infty} \frac{\ln\left(\frac{tx-1}{t-1}\right)}{\frac{t}{t-1}} = \ln x.$$

Při odvození vhodného tvaru pomocné funkce $A(t)$ lze například vyjít z von Misesovy postačující podmínky druhého řádu (věta 1.13). Pro logistické rozdělení máme

$$A(t) = \frac{-1}{t-1}.$$

Zřejmě jsou splněny všechny předpoklady tvrzení 1.13, zejména $A(t)$ nemění znaménko pro $t \rightarrow \infty$ a $\lim_{t \rightarrow \infty} A(t) = 0$. Snadno lze ověřit, že $|A(t)| \in \mathcal{R}_{-1}$. Ověříme tak platnost podmínky (A.4) při volbě pomocné funkce prvního řádu $a(t) = t/(t-1)$. Postupnými úpravami a použitím l'Hospitalova pravidla obdržíme

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{\frac{U(tx)-U(t)}{a(t)} - D_\xi(x)}{A(t)} &= \lim_{t \rightarrow \infty} \frac{\frac{\ln\left(\frac{tx-1}{t-1}\right)}{\frac{t}{t-1}} - \ln x}{\frac{-1}{t-1}} = \lim_{t \rightarrow \infty} \frac{\ln\left(\frac{tx-1}{t-1}\right) - \frac{t}{t-1} \ln x}{\frac{-t}{(t-1)^2}} = \\ &= \lim_{t \rightarrow \infty} \frac{\frac{(1-x)(t-1)}{tx-1} + \ln x}{\frac{t+1}{t-1}} = \frac{1-x}{x} + \ln x. \end{aligned}$$

Tedy logistické rozdělení splňuje podmínku druhého řádu s parametrem $\rho = -1$.

Weibullovo rozdělení. Uvažujme Weibullovo rozdělení s parametry $\lambda, \alpha > 0$ a s distribuční funkcí

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

Kvantilová funkce chvostu je ve tvaru $U(t) = \lambda (\ln t)^{1/\alpha}$ pro $t > 1$. Volba pomocné funkce $a(t)$ dle kritéria (A.2) vede ke značně obtížnému výpočtu. Namísto toho můžeme použitím binomického rozvoje $(x+y)^c = x^c + cx^{c-1}y + c(c-1)/2x^{c-2}y^2 + o(x^{c-2})$ psát následující

$$\begin{aligned} U(tx) - U(t) &= \lambda [(\ln t + \ln x)^{1/\alpha} - (\ln t)^{1/\alpha}] = \\ &= \lambda \left[\frac{1}{\alpha} (\ln t)^{(1-\alpha)/\alpha} \ln x + \frac{1-\alpha}{2\alpha^2} (\ln t)^{(1-2\alpha)/\alpha} (\ln x)^2 + o((\ln t)^{(1-2\alpha)/\alpha}) \right]. \quad (\text{A.10}) \end{aligned}$$

Snadno se tak nahlédne, že pomocná funkce $a(t) = \frac{\lambda}{\alpha} (\ln t)^{(1-\alpha)/\alpha}$, tedy

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \lim_{t \rightarrow \infty} \left(\ln x + \frac{1-\alpha}{2\alpha} \frac{(\ln x)^2}{\ln t} + o((\ln t)^{-1}) \right) = \ln x.$$

Dle kritéria (A.3) získáme funkci $a(t)$ v identickém tvaru, přičemž zde je $h(t) = \lambda^\alpha t^{1-\alpha}/\alpha$.

Při volbě pomocné funkce druhého řádu lze vyjít z předchozího vyjádření a ze tvaru podmínky (A.5). Zřejmě můžeme volit $a_0(t) = a(t)$ a

$$A_0(t) = \frac{1-\alpha}{\alpha \ln t}.$$

Podmínka druhého řádu (A.5) je pak splněna, neboť

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx) - U(t)}{a_0(t)} - D_0(x)}{A_0(t)} = \lim_{t \rightarrow \infty} \frac{\frac{1-\alpha}{2\alpha} \frac{(\ln x)^2}{\ln t} + o((\ln t)^{-1})}{\frac{1-\alpha}{\alpha \ln t}} = \lim_{t \rightarrow \infty} \left(\frac{(\ln x)^2}{2} + o(1) \right) = \frac{(\ln x)^2}{2}.$$

Tedy v případě Weibullova rozdělení je $\xi = \rho = 0$. Dle důsledku 1.12 tak dostáváme pomocné funkce podmínky druhého řádu (A.4) ve tvaru

$$a(t) = a_0(t) = \frac{\lambda}{\alpha} (\ln t)^{1/\alpha-1}, \quad A(t) = A_0(t) = \frac{1-\alpha}{\alpha \ln t}.$$

Rozdělení z Fréchetova oboru atraktivity

Pro Fréchetův obor atraktivity je při volbě $a(t) = \xi U(t)$ (A.2) podmínka atraktivity zřejmě triviálně splněna, neboť lze uvedenou podmínku zjednodušit na tvar

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \lim_{t \rightarrow \infty} \frac{1}{\xi} \left(\frac{U(tx)}{U(t)} - 1 \right). \quad (\text{A.11})$$

Jak bylo ukázáno v odstavci 1.3, je pro $\xi > 0$ funkce $U(t) \in \mathcal{R}_\xi$, a tedy pravá strana je rovna funkci $D_\xi(x)$. Přesto však v níže studovaných případech platnost této podmínky alespoň naznačíme.

Paretovo rozdělení. Uvažujme Paretovo rozdělení s parametry $\xi > 0$, $\alpha \in \mathbb{R}$ a s distribuční funkcí

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

Kvantilová funkce chvostu je $U(t) = \alpha t^\xi$ pro $t > 1$ a tedy $U(tx) - U(t) = \alpha t^\xi (x^\xi - 1)$. Snadno se ověří (např. pomocí von Misesovy podmínky 1.10), že index extrémní hodnoty pro Paretovo rozdělení je roven parametru ξ . Dle (A.2) získáme funkci $a(t)$ ve tvaru

$$a(t) = \xi \alpha t^\xi,$$

pomocí vztahu (A.3) je volba zcela identická při $h(t) = \xi t$. Ověříme tak platnost podmínky atraktivity. Dostáváme

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \lim_{t \rightarrow \infty} \frac{\alpha t^\xi (x^\xi - 1)}{\xi \alpha t^\xi} = \frac{x^\xi - 1}{\xi}.$$

Protože v podmínce (A.1) dostáváme přímo rovnost, není možné nalézt funkci $A(t)$ tak, aby byla splněna podmínka (A.4), a tedy $\rho = -\infty$.

Burrovo rozdělení (typ XII). Uvažujme Burrovo rozdělení typu XII s parametry $\lambda, \tau, \eta > 0$ a s distribuční funkcí

$$F(x) = \begin{cases} 1 - \left(\frac{\eta}{\eta + x^\tau}\right)^\lambda, & \text{pro } x \geq 0, \\ 0, & \text{pro } x < 0. \end{cases}$$

Kvantilová funkce chvostu je ve tvaru $U(t) = (\eta [t^{1/\lambda} - 1])^{1/\tau}$ pro $t > 1$. Dle (A.2) můžeme zvolit pomocnou funkci $a(t)$ ve tvaru $a(t) = \frac{1}{\lambda\tau} U(t)$. Dostáváme tak podmínku atraktivity ve tvaru

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \lim_{t \rightarrow \infty} \frac{U(tx)}{\frac{1}{\lambda\tau} \eta^{1/\tau} (t^{1/\lambda} - 1)^{1/\tau}} - \lambda\tau = \lambda\tau x^{1/(\lambda\tau)} - \lambda\tau.$$

Volbou pomocné funkce $a(t)$ dle kritéria (A.3) obdržíme poněkud složitější tvar $a(t) = \frac{1}{\lambda\tau} \frac{t^{1/\lambda}}{t^{1/\lambda} - 1} U(t)$ při $h(t) = \frac{1}{\lambda\tau} (\eta + t^\tau) t^{1-\tau}$, který je asymptoticky ekvivalentní předchozímu vyjádření.

Pro nalezení vhodného tvaru funkce $A(t)$ budeme vycházet z tvaru podmínky druhého řádu (A.5). Nejprve přepišme kvantilovou funkci chvostu $U(t)$ užitím binomického rozvoje $(t-1)^c = t^c - ct^{c-1} + o(t^{c-1})$, tudíž

$$U(t) = \eta^\alpha (t^\beta - 1)^\alpha = \eta^\alpha (t^{\alpha\beta} - \alpha t^{\beta(\alpha-1)} + o(t^{\beta(\alpha-1)})),$$

přičemž byla zavedena parametrizace $\alpha = 1/\tau, \beta = 1/\lambda$. Můžeme pak psát

$$U(tx) - U(t) = \eta^\alpha t^{\alpha\beta} (x^{\alpha\beta} - 1 - \alpha t^{-\beta} [x^{\beta(\alpha-1)} - 1] + o(t^{-\beta})).$$

Protože EV index Burrova rozdělení je $\xi = 1/(\tau\lambda) = \alpha\beta$, platí $D_\xi(x) = (x^{\alpha\beta} - 1)/(\alpha\beta)$. Snadno se nahlédne, že podmínka (A.5) je splněna při volbě

$$a_0(t) = \alpha\beta\eta^\alpha t^{\alpha\beta}, \quad A_0(t) = -\frac{\alpha - 1}{t^\beta},$$

pro $\rho = -\beta = -1/\lambda$, čímž podle důsledku 1.12 je podmínka druhého řádu (A.4) splněna pro

$$a(t) = \frac{a_0(t)}{1 - A_0(t)} = \alpha\beta\eta^\alpha \frac{t^{\beta(\alpha+1)}}{t^\beta + \alpha - 1}, \quad A(t) = \rho A_0(t) = \frac{\beta(\alpha - 1)}{t^\beta}.$$

Fréchetovo rozdělení. Uvažujme Fréchetovo rozdělení s parametrem $\gamma > 0$ a s distribuční funkcí

$$F(x) = \begin{cases} \exp(-x^{-1/\xi}), & \text{pro } x \geq 0, \\ 0, & \text{pro } x < 0. \end{cases}$$

Kvantilová funkce chvostu je ve tvaru $U(t) = (-\ln [1 - \frac{1}{t}])^{-\xi}$ pro $t > 1$. Pro $|t| \leq 1$ můžeme použít přepis $\ln(1+t) = t - \frac{t^2}{2} + \frac{t^3}{3} + o(t^3)$ a neboť $(1+t)^\alpha = 1 + \alpha t + \frac{\alpha(\alpha-1)}{2} t^2 + o(t^2)$, můžeme kvantilovou funkci chvostu postupně přepsat ve tvaru

$$\begin{aligned} U(t) &= \left(-\ln \left[1 - \frac{1}{t} \right] \right)^{-\xi} = t^\xi \left(1 + \frac{1}{2t} + \frac{1}{3t^2} + o(t^{-2}) \right)^{-\xi} = \\ &= t^\xi \left(1 - \frac{\xi}{2t} + \frac{\xi(3\xi - 5)}{24t^2} + o(t^{-2}) \right). \end{aligned}$$

Funkce $a(t)$ je podle kritéria (A.2) ve tvaru $a(t) = \xi U(t)$. Asymptoticky ekvivalentní volba je $a(t) = \xi t^\xi(1 + o(1))$, $a(t) = \xi t^\xi$ je však ještě jednodušší. Ověříme platnost podmínky atraktivitu. Nejprve můžeme zjednodušit výraz

$$\begin{aligned} U(tx) - U(t) &= \xi t^\xi \left(\frac{x^\xi - 1}{\xi} - \frac{x^{\xi-1} - 1}{2t} + (3\xi - 5) \frac{x^{\xi-2} - 1}{24t^2} + o(t^{-2}) \right) = \\ &= \begin{cases} \xi t^\xi \left(\frac{x^\xi - 1}{\xi} - \frac{x^{\xi-1} - 1}{2t} + o(t^{-1}) \right), & \text{pro } \xi \neq 1, \\ t \left(x - 1 - \frac{1}{12t^2} (x^{-1} - 1) + o(t^{-2}) \right), & \text{pro } \xi = 1. \end{cases} \end{aligned}$$

nyň je zřejmé, že při volbě $a(t) = \xi t^\xi(1 + o(1))$ je podmínka atraktivitu splněna.

Hledejme dále pomocnou funkci druhého řádu $A(t)$. Pomocí kritérií (A.6) a (A.8) lze obdržet vyjádření (A.5) při volbě

$$a_0(t) = \xi t^\xi, \quad A_0(t) = \begin{cases} -\frac{\xi-1}{2t}, & \xi \neq 1, \\ \frac{1}{12t^2}, & \xi = 1. \end{cases}$$

Platnost (A.5) ověříme. Pro $\xi = 1$ dosazením a_0, A_0 a úpravou dostáváme

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx) - U(t)}{a_0(t)} - D_\xi(x)}{A_0(t)} = \lim_{t \rightarrow \infty} \left(-(x^{-1} - 1) + o(1) \right) = 1 - x^{-1}.$$

V případě $\xi \neq 1$ obdržíme úpravou vztahu (A.5)

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx) - U(t)}{a_0(t)} - D_\xi(x)}{A_0(t)} = \lim_{t \rightarrow \infty} \left(\frac{x^{\xi-1} - 1}{\xi - 1} + o(1) \right) = \frac{x^{\xi-1} - 1}{\xi - 1}.$$

Tedy parametr druhého řádu $\rho = -2$ pro $\xi = 1$, $\rho = -1$ pro $\xi \neq 1$ a dle důsledku 1.12 máme

$$a(t) = \begin{cases} \frac{2\xi t^{\xi+1}}{2t + \xi - 1}, & \xi \neq 1, \\ \frac{12t^3}{12t^2 - 1}, & \xi = 1, \end{cases} \quad A(t) = \begin{cases} \frac{\xi-1}{2t}, & \xi \neq 1, \\ -\frac{1}{6t^2}, & \xi = 1. \end{cases}$$

Hallová třída rozdělení. Nyní uvažujme obecnější případ tzv. Hallovy třídy rozdělení s těžkými chvosty, tj. rozdělení s kvantilovou funkcí chvostu ve tvaru

$$U(t) = Ct^\xi (1 + Dt^\rho + o(t^\rho)),$$

kde $C > 0, D \in \mathbb{R}, \xi > 0$ a $\rho < 0$. Zřejmě můžeme funkci $a(t)$ zvolit dle (A.2) ve tvaru $a(t) = \xi U(t)$, tj. máme

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \lim_{t \rightarrow \infty} \frac{1}{\xi} \left(\frac{U(tx)}{U(t)} - 1 \right) = \lim_{t \rightarrow \infty} \frac{1}{\xi} \left(\frac{x^\xi (1 + Dt^\rho x^\rho + o(t^\rho))}{(1 + Dt^\rho + o(t^\rho))} - 1 \right) = \frac{x^\xi - 1}{\xi}.$$

EV index Hallovy třídy rozdělení je tak roven parametru ξ .

Dále se zabýváme podmínkou druhého řádu. Dle vztahu (A.6) můžeme brát $a_0(t) = ct^\xi$, kde $c = C\xi$. Dostaneme tak rovnost

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx) - U(t)}{a_0(t)} - D_\xi(x)}{A_0(t)} = \lim_{t \rightarrow \infty} \frac{1}{A_0(t)} \left[\frac{Dt^\rho}{\xi} (x^{\xi+\rho} - 1) + o(t^\rho) \right] = \Psi_{\xi, \rho}(x).$$

Zřejmě, lze pravou stranu $\Psi_{\xi, \rho}(x)$ z (A.5) pro $\xi \neq 0$ obdržet volbou $A_0(t) = \frac{\xi+\rho}{\xi} Dt^\rho$, přičemž parametr druhého řádu je roven parametru $\rho < 0$.

Zobecněné Paretovo (GP) rozdělení. Uvažujme GP rozdělení s parametrem $\xi > 0$ a s distribuční funkcí

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

Kvantilová funkce chvostu je ve tvaru $U(t) = \frac{1}{\xi}(t^\xi - 1)$ pro $t > 1$. Protože GP rozdělení je speciálním případem rozdělení patřící do Ĥallovoy třídy rozdělení s parametry $\rho = -\xi$, $C = 1/\xi$, $D = -1$, lze z předchozího dostat $a_0(t) = t^\xi$. Odvození funkce $A(t)$ je však nutné provést jiným způsobem, neboť dle výše uvedeného dostaneme $A(t)$ jako identicky nulovou funkci.

Dle [38], tvrzení 2.3.9, pro $\xi > 0$ a nějakou kladnou či zápornou funkci $A(t)$ musí být splněno

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx)}{U(t)} - x^\xi}{A(t)} = x^\xi \frac{x^\rho - 1}{\rho}. \quad (\text{A.12})$$

Rozepsáním předchozí limity máme

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx)}{U(t)} - x^\xi}{A(t)} = \lim_{t \rightarrow \infty} \frac{1}{A(t)} \frac{x^\xi - 1}{t^\xi - 1}$$

a tedy jediná možná volba pro získání pravé strany (A.12) je $A(t) = (\xi(t^\xi - 1))^{-1}$, přičemž pravá strana je pak ve tvaru $(x^\xi - 1)/\xi$. Tudíž parametr druhého řádu $\rho = -\xi$.

Cauchyho rozdělení. Uvažujme Cauchyho rozdělení s distribuční funkcí

$$F(x) = \frac{1}{2} + \frac{1}{\pi} \operatorname{arctg} x, \quad \text{pro } x \in \mathbb{R}.$$

Kvantilová funkce chvostu je ve tvaru $U(t) = \operatorname{tg} \left(\pi \left[\frac{1}{2} - \frac{1}{t} \right] \right)$ pro $t > 1$. Dle (A.2) lze volit $a(t) = U(t)$, avšak pro Cauchyho rozdělení se nabízí jednodušší varianta $a(t) = t/\pi$. Protože $\operatorname{tg} \left(x + \frac{\pi}{2} \right) = -\operatorname{cotg} x$, můžeme psát $U(t) = \operatorname{cotg} \frac{\pi}{t} = \frac{t}{\pi} - \frac{\pi}{3t} + o(t^{-1})$. Dosazením do podmínky atraktivitv dostáváme

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \lim_{t \rightarrow \infty} \frac{t(x-1) - \frac{1}{t} \left(\frac{\pi^2}{3x} - \frac{\pi^2}{3} \right) + o(t^{-1})}{t} = x - 1,$$

čímž je podmínka (A.1) splněna pro $\xi = 1$.

Rozepišme dále čitatele podmínky druhého řádu (A.5), tj. máme

$$\begin{aligned} \frac{U(tx) - U(t)}{a_0(t)} - D_1(x) &= \\ &= \frac{\pi \left[\left(\frac{tx}{\pi} - \frac{\pi}{3tx} + o(t^{-1}) \right) - \left(\frac{t}{\pi} - \frac{\pi}{3t} + o(t^{-1}) \right) \right]}{t} - (x-1) = \frac{\pi^2}{3t^2} \left(1 - \frac{1}{x} \right) + o(t^{-2}). \end{aligned}$$

Zřejmě tak lze funkci $A_0(t)$ zvolit jako $A_0(t) = \frac{\pi^2}{3t^2}$, čímž dostaneme výsledek limity (A.5) ve tvaru $1 - x^{-1}$ a tedy $\rho = -2$.

Burrovo rozdělení (typ III). Uvažujme Burrovo rozdělení typu III s parametry $\lambda, \tau, \eta > 0$ a s distribuční funkcí

$$F(x) = \begin{cases} \left(\frac{\eta}{\eta+x^{-\tau}}\right)^\lambda, & \text{pro } x > 0, \\ 0, & \text{pro } x \leq 0. \end{cases}$$

Kvantilová funkce chvostu je ve tvaru $U(t) = \left(\frac{\eta}{(1-t^{-1})^{1/\lambda}} - \eta\right)^{-1/\tau}$. Označme dále $c = 1/\tau, b = 1/\lambda$. Pomocí binomického rozvoje můžeme funkci U vyjádřit ve tvaru

$$\begin{aligned} U(t) &= \eta^{-c} \left(\frac{t^b - (t-1)^b}{(t-1)^b}\right)^{-c} = \eta^{-c} \left(\frac{t^b - bt^{b-1} + o(t^{b-1})}{bt^{b-1}(1+o(1))}\right)^c = \\ &= (b\eta)^{-c} (t-b+o(1))^c = (b\eta)^{-c} (t^c - cbt^{c-1} + o(t^{c-1})). \end{aligned}$$

Odtud

$$U(tx) - U(t) = (b\eta)^{-c} t^c (x^c - 1 - cbt^{-1}(x^{c-1} - 1) + o(t^{-1})).$$

Zřejmě je pak podmínka atraktivity (A.1) splněna pro $a(t) = c(b\eta)^{-c} t^c$, tj.

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \lim_{t \rightarrow \infty} \frac{1}{c} (x^c - 1 - cbt^{-1}(x^{c-1} - 1) + o(t^{-1})) = \frac{x^c - 1}{c}.$$

Soustředme se nyní na podmínku druhého řádu. Dosazením funkce $A_0(t)$ ve tvaru

$$A_0(t) = -\frac{b(c-1)}{t}$$

do podmínky (A.5) dostáváme

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx) - U(t)}{a_0(t)} - D_\xi(x)}{A_0(t)} = \lim_{t \rightarrow \infty} \frac{\left(\frac{x^c - 1}{c} - \frac{b}{t}(x^{c-1} - 1) + o(t^{-1}) - D_\xi(x)\right)}{-\frac{b(c-1)}{t}} = \frac{x^c - 1}{c - 1}$$

a tedy $\rho = -1$. Dle důsledků 1.12 tak dostáváme

$$a(t) = \frac{a_0(t)}{1 - A_0(t)} = \frac{c(b\eta)^{-c} t^{c+1}}{t - b(c-1)}, \quad A(t) = \rho A_0(t) = \frac{b(c-1)}{t}.$$

Loglogistické rozdělení. Uvažujme loglogistické rozdělení s parametry $\alpha, \beta > 0$ a s distribuční funkcí

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

Snadno se nahlédne, že loglogistické rozdělení je speciální typ Burrova rozdělení typu III s parametry $(\lambda, \tau, \eta) = (1, \beta, \alpha^{-\beta})$. Pomocné funkce jsou tedy ve tvaru $a(t) = \frac{\alpha}{\beta} t^{1/\beta}$, $A(t) = (1 - \beta)/(\beta t)$ a parametr druhého řádu $\rho = -1$.

Rozdělení z Weibullova oboru atraktivity

Rovnoměrné rozdělení. Uvažujme rovnoměrné rozdělení s parametry $a, b, a < b$ a s distribuční funkcí

$$F(x) = \begin{cases} \frac{x-a}{b-a}, & \text{pro } a \leq x \leq b, \\ 0, & \text{jinak.} \end{cases}$$

Pro rovnoměrné rozdělení máme $U(t) = \frac{t-1}{t}(b-a) + a$ pro $t > 1$ a $U(tx) - U(t) = \frac{x-1}{x} \frac{b-a}{t}$. Při volbě $a(t)$ dle (A.3) dostáváme

$$a(t) = \frac{b-a}{t},$$

přičemž $h(t) = b - t$. Použití vztahu (A.2) vede ke shodnému tvaru funkce $a(t)$, neboť pravý koncový bod $x^* = b$. Ověříme tedy platnost podmínky atraktivity:

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

Tedy rovnoměrné rozdělení splňuje podmínku atraktivity s indexem extrémní hodnoty $\xi = -1$.

Při ověření podmínky druhého řádu ve vztahu (A.1) nastává rovnost přímo a žádnou volbou $A(t)$ není možné získat rovnost v rovnici (A.4). Rovnoměrné rozdělení tak nesplňuje podmínku druhého řádu a $\rho = -\infty$.

Extremální Weibullovo rozdělení. Uvažujme extremální Weibullovo rozdělení s parametrem $\alpha > 0$ a s distribuční funkcí

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

Kvantilová funkce chvostu je ve tvaru $U(t) = -\left[-\left(\ln\left(1 - \frac{1}{t}\right)\right)\right]^{1/\alpha}$ pro $t > 1$. Funkci $U(t)$ lze zapsat ve tvaru

$$U(t) = -t^{-1/\alpha} \left[1 + \frac{1}{2\alpha t} + o(t^{-1}) \right].$$

Odvození pomocných funkcí se provede stejně jako v případě Fréchetova rozdělení pro $\xi = -1/\alpha$, a tedy tyto jsou ve tvaru

$$a_0(t) = \frac{1}{\alpha} t^{-\frac{1}{\alpha}}, \quad A_0(t) = \frac{1 + \alpha}{2\alpha t},$$

přičemž EV index tohoto rozdělení je $\xi = -1/\alpha$ a parametr druhého řádu je $\rho = -1$.

Reverzní Burrovo rozdělení. Uvažujme reverzní Burrovo rozdělení s parametry $\eta, \tau, \lambda > 0$ a s distribuční funkcí

$$F(x) = \begin{cases} 1 - \left(\frac{\eta}{\eta + (-x)^{-\tau}} \right)^\lambda, & \text{pro } x \leq 0, \\ 1, & \text{pro } x > 0. \end{cases}$$

Kvantilová funkce chvostu je ve tvaru $U(t) = -[\eta(t^{1/\lambda} - 1)]^{-1/\tau}$ pro $t > 1$. Pravý koncový bod tohoto rozdělení je $x^* = 0$ a tedy funkci $a(t)$ můžeme dle (A.2) volit ve tvaru $a(t) = \xi U(t)$, kde EV index $\xi = \frac{-1}{\lambda\tau}$. Dostáváme tak

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} &= \lim_{t \rightarrow \infty} \lambda\tau \left(1 - \frac{U(tx)}{U(t)} \right) = \lim_{t \rightarrow \infty} \lambda\tau \left(1 - \frac{[(tx)^{1/\lambda}]^{-1/\tau}}{[t^{1/\lambda} - 1]^{-1/\tau}} \right) = \\ &= \lambda\tau (1 - x^{1/(\lambda\tau)}) = \frac{x^{-1/(\lambda\tau)} - 1}{\frac{-1}{\lambda\tau}}. \end{aligned}$$

Dle (A.6) lze dále volit $a_0(t)$ ve tvaru $a_0(t) = ct^{-1/(\lambda\tau)}$, kde $c = \frac{1}{\lambda\tau}\eta^{1/\tau}$, tedy $a_0(t) = \frac{\eta^{1/\tau}}{\lambda\tau}t^{-1/(\lambda\tau)}$. Protože

$$U(tx) - U(t) = \eta^{\frac{1}{\tau}} \left[t^{\frac{-1}{\lambda\tau}} \left(1 - x^{\frac{-1}{\lambda\tau}} \right) + \frac{1}{\tau} t^{\frac{1}{\lambda}(\frac{-1}{\tau}-1)} \left(1 - x^{\frac{1}{\lambda}(\frac{-1}{\tau}-1)} \right) + o\left(t^{\frac{1}{\lambda}(\frac{-1}{\tau}-1)}\right) \right],$$

máme

$$\lim_{t \rightarrow \infty} \frac{1}{A_0(t)} \left[\frac{U(tx) - U(t)}{a_0(t)} - D_\xi(x) \right] = \lim_{t \rightarrow \infty} \frac{1}{A_0(t)} \left[\lambda t^{\frac{-1}{\lambda}} \left(1 - x^{\frac{1}{\lambda}(\frac{-1}{\tau}-1)} \right) + o\left(t^{\frac{-1}{\lambda}}\right) \right],$$

přičemž (A.5) získáme pro $A_0(t) = \frac{\tau}{1+\tau}t^{-1/\lambda}$. Odtud lze pak obdržet parametr druhého řádu $\rho = \frac{-1}{\lambda}$.

B

Odvození skórového vektoru a Fisherovy informační matice MT-GP modelu adaptivní volby prahové hodnoty

V této části bude, v návaznosti na odstavce 3.3, odvozen tvar skórového vektoru a FIM MT-GP modelu adaptivní volby prahu. Základní odvození je možné také nalézt přímo v publikaci [102], zde překládané bylo autorem práce detailně ověřeno a částečně zjednodušeno.

MT-GP model je postaven na myšlenkách grafických technik určení optimálního prahu, kdy je vyžadována stabilita parametru tvaru ξ . Zavede se tak diskretizace ξ pomocí m prahových hodnot u_1, \dots, u_m , čímž se dostane m hodnot ξ_1, \dots, ξ_m . Nechť X je náhodná veličina z rozdělení s distribuční funkcí $F(x)$. Buď $v_i = u_i - u_1$ pro $i = 1, \dots, m$ a $w_i = v_{i+1} - v_i$ pro $i = 1, \dots, m - 1$. Označme Y překročení prahu u_1 , tj. $Y = X - u_1$. Předpokládejme dále, že náhodná veličina $Y - v_i = X - u_i$ má za podmínky $v_i < Y < v_{i+1}$ GP(σ_i, ξ_i) rozdělení, přičemž $v_{m+1} := v_m - \sigma_m / \xi_m$ pro $\xi_m < 0$ a $v_{m+1} := \infty$ jinak. Přitom je vhodné uvažovat parametry měřítka ve tvaru $\sigma_i = \sigma_1 + \sum_{j=1}^i \xi_j w_j$. Pak je navíc model možné popsat pomocí menšího počtu $m + 1$ parametrů $\boldsymbol{\theta} = (\sigma_1, \xi_1, \dots, \xi_m)$.

Pro ověření platnosti nulové hypotézy $H : \xi_1 = \dots = \xi_m$, resp. dále $H : \xi_i = \dots = \xi_m$, lze využít statistiku (3.16), tedy statistiku ve tvaru

$$LM(\hat{\boldsymbol{\theta}}_0) = \frac{1}{n} \mathbf{U}^T(\hat{\boldsymbol{\theta}}_0) \mathbf{J}^{-1}(\hat{\boldsymbol{\theta}}_0) \mathbf{U}(\hat{\boldsymbol{\theta}}_0).$$

Zde bude odvozen tvar skórového vektoru $U(\boldsymbol{\theta})$ a FIM $\mathbf{J}(\boldsymbol{\theta})$.

Výhodnější je parametrizovat model pomocí parametru $\boldsymbol{\theta}^* = (\sigma_1, \phi_1, \dots, \phi_m)$, kde $\phi_i = \xi_i / \sigma_i$. V případě, že $n = 1$, je pak logaritmická věrohodnostní funkce (3.15) ve tvaru

$$l(\boldsymbol{\theta}^*) = \sum_{i=1}^m \mathbf{1}_{[v_i < y < v_{i+1}]} \left\{ \ln p_i - \ln \sigma_i - \left(1 + \frac{1}{\sigma_i \phi_i} \right) \ln [1 + \phi_i (y - v_i)] \right\}. \quad (\text{B.1})$$

Snadno se ukáže, že první členy pravé strany lze vyjádřit následovně, tudíž

$$\ln p_i = \begin{cases} 0, & i = 1, \\ -\sum_{k=1}^{i-1} \frac{\ln(1 + \phi_k w_k)}{\sigma_k \phi_k}, & i = 2, \dots, m, \end{cases}$$

$$\ln \sigma_i = \begin{cases} \ln \sigma_1, & i = 1, \\ \ln \sigma_1 + \sum_{k=1}^{i-1} \ln(1 + \phi_k w_k), & i = 2, \dots, m. \end{cases} \quad (\text{B.2})$$

Ze vztahu (B.2) lze tedy postupně psát

$$\sigma_i = \sigma_1 \prod_{k=1}^{i-1} (1 + \phi_k w_k), \quad (\text{B.3})$$

$$\frac{\partial \sigma_i}{\partial \sigma_1} = \prod_{k=1}^{i-1} (1 + \phi_k w_k) = \frac{\sigma_i}{\sigma_1}. \quad (\text{B.4})$$

Označme $T_i := 1 + \phi_i(y - v_i)$, $R_i := (\sigma_i \phi_i)^{-1} \ln T_i$, $\gamma_i := 1 + \phi_i w_i$. Derivováním věrohodnostní funkce (B.1) dostáváme

$$\begin{aligned} \frac{\partial l(\boldsymbol{\theta}^*)}{\partial \sigma_1} &= -\frac{1}{\sigma_1} \sum_{i=1}^m \mathbf{1}_{[v_i < y < v_{i+1}]} \{\ln p_i + 1 - R_i\}, \\ \frac{\partial l(\boldsymbol{\theta}^*)}{\partial \phi_k} &= \sum_{i=1}^m \mathbf{1}_{[v_i < y < v_{i+1}]} \left\{ \frac{\partial \ln p_i}{\partial \phi_k} - \frac{\partial \ln \sigma_i}{\partial \phi_k} - \frac{\partial \ln T_i}{\partial \phi_k} - \frac{\partial R_i}{\partial \phi_k} \right\}, \end{aligned}$$

kde

$$\begin{aligned} \frac{\partial \ln p_i}{\partial \phi_k} &= \mathbf{1}_{[k \leq i-1]} \frac{1}{\sigma_k \phi_k} \left\{ \frac{\ln \gamma_k}{\phi_k} - \frac{w_k}{\gamma_k} \right\} + \frac{w_k}{\gamma_k} \sum_{j=1}^{i-1} \mathbf{1}_{[k \leq j-1]} \frac{\ln(\gamma_j)}{\sigma_j \phi_j}, \\ \frac{\partial \ln \sigma_i}{\partial \phi_k} &= \mathbf{1}_{[k \leq i-1]} \frac{w_k}{\gamma_k}, \\ \frac{\partial \ln T_i}{\partial \phi_k} &= \mathbf{1}_{[k=i]} \frac{y - v_i}{T_i}, \\ \frac{\partial R_i}{\partial \phi_k} &= \mathbf{1}_{[k=i]} \frac{1}{\sigma_i \phi_i} \left\{ \frac{y - v_i}{T_i} - \sigma_i R_i \right\} - \mathbf{1}_{[k \leq i-1]} \frac{w_k R_i}{\gamma_k}. \end{aligned}$$

Odtud lze již přímo získat skórový vektor $\mathbf{U}_y(\hat{\boldsymbol{\theta}}_0^*)$ pro případ $n = 1$ a pozorování hodnoty y . Pro obecnější případ n -rozměrného náhodného výběru $\mathbf{Y} = (Y_1, \dots, Y_n)$ se skórový vektor $\mathbf{U}(\hat{\boldsymbol{\theta}}_0^*)$ jednoduše dostane součtem všech skórových vektorů pro jednotlivá pozorování odvozených na základě věrohodnostní funkce (B.1), tedy

$$\mathbf{U}(\hat{\boldsymbol{\theta}}_0^*) = \sum_{i=1}^n \mathbf{U}_{y_i}(\hat{\boldsymbol{\theta}}_0^*).$$

Prvky výběrové FIM $\tilde{\mathbf{J}}(\boldsymbol{\theta}^*) = \left(\tilde{J}_{ij}(\boldsymbol{\theta}^*) \right)_{i,j=1}^{m+1} = \left(-\frac{\partial^2 l(\boldsymbol{\theta}^*)}{\partial \theta_i^* \partial \theta_j^*} \right)_{i,j=1}^{m+1}$ obdržíme dle (2.11) ve tvaru

$$\begin{aligned} -\frac{\partial^2 l(\boldsymbol{\theta}^*)}{\partial \sigma_1^2} &= -\frac{1}{\sigma_1^2} \sum_{i=1}^m \mathbf{1}_{[v_i < y < v_{i+1}]} \{2 \ln p_i + 1 - 2R_i\}, \\ -\frac{\partial^2 l(\boldsymbol{\theta}^*)}{\partial \sigma_1 \partial \phi_k} &= \frac{1}{\sigma_1} \sum_{i=1}^m \mathbf{1}_{[v_i < y < v_{i+1}]} \left\{ \frac{\partial \ln p_i}{\partial \phi_k} - \frac{\partial R_i}{\partial \phi_k} \right\}, \\ -\frac{\partial^2 l(\boldsymbol{\theta}^*)}{\partial \phi_k^2} &= -\sum_{i=1}^m \mathbf{1}_{[v_i < y < v_{i+1}]} \left\{ \frac{\partial^2 \ln p_i}{\partial \phi_k^2} - \frac{\partial^2 \ln \sigma_i}{\partial \phi_k^2} - \frac{\partial^2 \ln T_i}{\partial \phi_k^2} - \frac{\partial^2 R_i}{\partial \phi_k^2} \right\}, \end{aligned}$$

a pro $k > l$

$$-\frac{\partial^2 l(\boldsymbol{\theta}^*)}{\partial \phi_k \partial \phi_l} = -\sum_{i=1}^m \mathbf{1}_{[v_i < y < v_{i+1}]} \left\{ \frac{\partial^2 \ln p_i}{\partial \phi_k \partial \phi_l} - \frac{\partial^2 R_i}{\partial \phi_k \partial \phi_l} \right\},$$

přičemž uvedené parciální derivace jsou dány následovně:

$$\frac{\partial^2 \ln p_i}{\partial \phi_k^2} = \mathbf{1}_{[k \leq i-1]} \frac{1}{\sigma_k \phi_k^2} \left\{ \frac{2w_k}{\gamma_k} + \frac{w_k^2 \phi_k}{\gamma_k^2} - \frac{2 \ln \gamma_k}{\phi_k} \right\} - \frac{w_k^2}{\gamma_k^2} \sum_{j=1}^{i-1} \mathbf{1}_{[k \leq j-1]} \frac{2 \ln \gamma_j}{\sigma_j \phi_j},$$

$$\frac{\partial^2 \ln \sigma_i}{\partial \phi_k^2} = \mathbf{1}_{[k \leq i-1]} \frac{-w_k^2}{\gamma_k^2},$$

$$\frac{\partial^2 \ln T_i}{\partial \phi_k^2} = \mathbf{1}_{[k=i]} \frac{-(y-v_i)^2}{T_i^2},$$

$$\begin{aligned} \frac{\partial^2 R_i}{\partial \phi_k^2} &= \mathbf{1}_{[k=i]} \frac{1}{\sigma_i \phi_i} \left\{ -\frac{(y-v_i)^2}{T_i^2} + \frac{2}{\phi_i^2} \ln T_i - \frac{2}{\phi_i} \frac{y-v_i}{T_i} \right\} + \\ &+ \mathbf{1}_{[k \leq i-1]} \frac{2}{\sigma_i \phi_i} \frac{w_k^2 \ln T_i}{\gamma_k^2}, \end{aligned}$$

$$\frac{\partial^2 \ln p_i}{\partial \phi_k \partial \phi_l} = \mathbf{1}_{[k \leq i-1]} \frac{w_l}{\sigma_k \phi_k \gamma_l} \left\{ \frac{w_k}{\gamma_k} - \frac{\ln \gamma_k}{\phi_k} \right\} - \frac{w_k w_l}{\gamma_k \gamma_l} \sum_{j=1}^{i-1} \mathbf{1}_{[k \leq j-1]} \frac{\ln \gamma_j}{\sigma_j \phi_j},$$

$$\frac{\partial^2 R_i}{\partial \phi_k \partial \phi_l} = \mathbf{1}_{[k=i]} \frac{w_l}{\sigma_i \phi_i \gamma_l} \left\{ \frac{\ln T_i}{\phi_i} - \frac{y-v_i}{T_i} \right\} + \mathbf{1}_{[k \leq i-1]} \frac{w_k w_l \ln T_i}{\sigma_i \phi_i \gamma_k \gamma_l}.$$

Prvky očekávané FIM $\mathbf{J}(\boldsymbol{\theta}^*) = (J_{ij}(\boldsymbol{\theta}^*))_{i,j=1}^{m+1}$ lze dále získat integrací dle (2.14). Přitom je vhodné u parciálních derivací věrohodnostní logaritmické funkce $l(\boldsymbol{\theta}^*)$ rozlišit členy, které neobsahují a které obsahují proměnnou y . Dle předchozích vyjádření tak lze každý prvek $\tilde{J}_{rs}(\boldsymbol{\theta}^*)$ výběrové FIM vyjádřit jako součet funkcí A_i^{rs} , $B_i^{rs}(y)$, přičemž prvky očekávané FIM obdržíme ve tvaru

$$\begin{aligned} J_{rs}(\boldsymbol{\theta}^*) &= -E \left(\frac{\partial^2 l(\boldsymbol{\theta}^*)}{\partial \theta_r \partial \theta_s} \right) = \sum_{i=1}^m \int_{v_i}^{v_{i+1}} \{A_i^{rs} + B_i^{rs}(y)\} p_i h_i(y) dy = \\ &= \sum_{i=1}^m A_i^{rs} p_i \left(1 - \gamma_i^{-1/\xi_i}\right) + \sum_{i=1}^m \int_{v_i}^{v_{i+1}} B_i^{rs}(y) p_i h_i(y) dy. \end{aligned} \quad (\text{B.5})$$

První člen vyjádření (B.5) byl získán z (3.13), odkud plyne $P(v_i < Y < v_{i+1}) = p_i \left(1 - \gamma_i^{-1/\xi_i}\right)$. V závislosti na hodnotách indexů (r, s) zahrnuje druhý člen vztahu (B.5) vybrané následující členy [102]:

$$\begin{aligned} \int_{v_i}^{v_{i+1}} \frac{y-v_i}{T_i} h_i(y) dy &= \sigma_i (1 + \xi_i)^{-1} + \mathbf{1}_{[i < m]} \phi_i^{-1} \left\{ \gamma_i^{-(1+1/\xi_i)} (1 + \xi_i)^{-1} - \gamma_i^{1/\xi_i} \right\}, \\ \int_{v_i}^{v_{i+1}} \frac{(y-v_i)^2}{T_i^2} h_i(y) dy &= 2\sigma_i^2 (1 + \xi_i)^{-1} (1 + 2\xi_i)^{-1} - \mathbf{1}_{[i < m]} \phi_i^{-2} \times \\ &\times \left\{ \gamma_i^{-1/\xi_i} - 2\gamma_i^{-(1+1/\xi_i)} (1 + \xi_i)^{-1} + \gamma_i^{-(2+1/\xi_i)} (1 + 2\xi_i)^{-1} \right\}, \\ \int_{v_i}^{v_{i+1}} \ln T_i h_i(y) dy &= \xi_i - \mathbf{1}_{[i < m]} \gamma_i^{-1/\xi_i} (\xi_i \ln \gamma_i). \end{aligned}$$

C

Software EVDest

Tato část je věnována popisu softwaru EVDest [7], který byl vytvořen primárně za účelem určení odhadů návratových úrovní extrémních událostí a jejich intervalů spolehlivosti. Software využívá jak blokového, tak prahového modelu, které byly diskutovány v odstavci 2.1, na jejichž základě jsou určeny odhady parametrů modelového GEV, resp. GP rozdělení. Software je možné použít jednak pro případ nezávislých pozorování, ale také pro stacionární řady, které byly popsány v kapitole 5. Pro účely nakládání se závislými pozorováními nabízí software možnost odhadu extrémálního indexu podle Gomes (viz odstavec 5.2), případně je také vybaven základní možností declusterování (runs declusterování). Software EVDest je dostupný na adrese:

<http://www.umat.feec.vutbr.cz/software/evdest.html>

Spuštění

Ke správné funkčnosti je potřeba Matlab se statistickým toolboxem (software byl testován na verzi Matlab 8.2 R2013b). Pro správný chod softwaru je nutné, aby byl nainstalován MATLAB Compiler Runtime; tento je přiložen k softwaru EVDest, případně je ke stažení na internetových stránkách Mathworks. Spuštění softwaru se provede otevřením souboru `EVDest.exe`.

Popis programu

Volba modelu se vždy provádí v záložce `File`. Na výběr má uživatel následující možnosti:

Nezávislá pozorování (Independent Series)

V této části jsou zahrnuty modely pro analýzu nezávislých pozorování. V záložce je dále možné zvolit mezi modelem blokových maxim a přístupem POT (prahový model). V obou případech je vstupem datový soubor nezávislých měření. K softwaru jsou přiloženy testovací datové soubory vycházející z analyzovaných srážkových dat, které je možné použít; data vhodná pro vyhodnocení jednotlivými metodami jsou příslušně označena v názvu.

V prahovém modelu je možné volně nastavit vhodné prahové hodnoty - pro snadnější volbu jsou v této sekci implementovány grafické metody volby prahu představené v odstavci 3.1 (viz obrázek C.1). Při použití modelu blokových maxim je nutné, aby vstupní datový soubor obsahoval již hodnoty blokových maxim.

V obou případech je navíc uživateli dána možnost testování shody modelového rozdělení pomocí testů dobré shody (χ^2 test, K-S test, A-D test) či pomocí grafických metod (histogram, Q-Q plot). Při použití metody blokových maxim je možné odhadnout

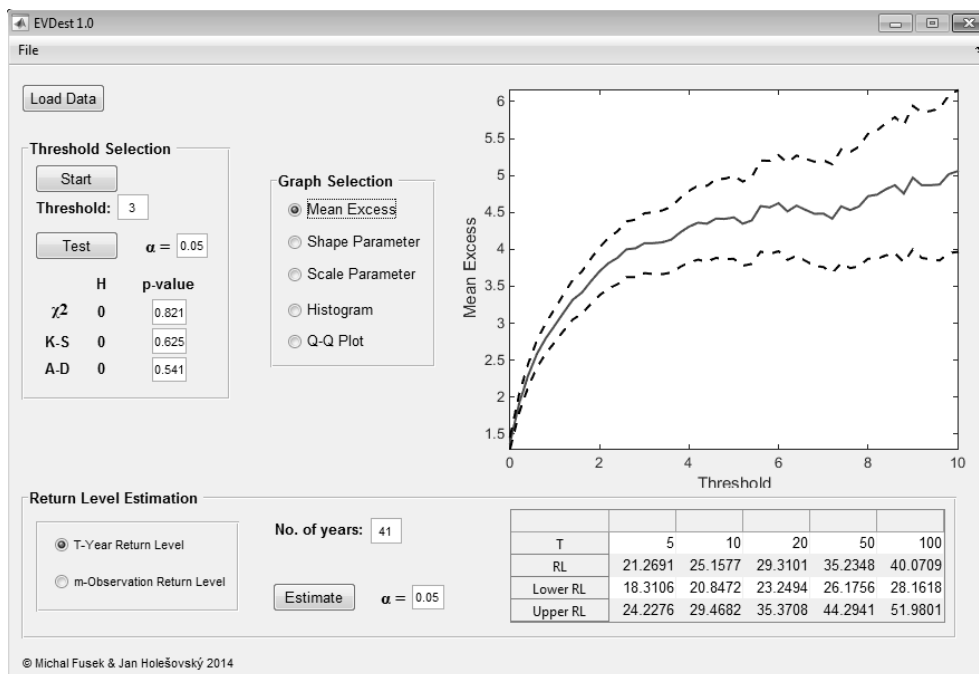
jen návratové úrovně (RT – Return Levels) T -letých událostí, zatímco při použití prahového modelu lze nastavit návratovou úroveň pro libovolný počet m pozorování. Při použití prahového modelu je však nutné zadat počet let měření (No. of years). Pro danou hladinu významnosti α software také určuje asymptotický interval spolehlivosti pro návratovou úroveň (Lower RT, Upper RT), přičemž všechny odhady jsou zobrazovány v grafu.

Závislá pozorování (Dependent Series)

Při volbě analýzy závislých pozorování nabízí software možnost použití prahového modelu s využitím runs declusterování (odstavec 5.1) nebo s využitím odhadu extrémálního indexu. Pro oba účely slouží testovací data v souboru `Dependent.txt`. Volba prahu se opět provede na základě grafických metod.

Při declusterování je nutné ve spodní části nastavit hodnotu declusterovacího runs parametru. Testovací data představují minutová měření, tj. hodnota tohoto parametru pak určuje, kolika minutami podprahových hodnot jsou odděleny různé shluky. Pro různá nastavení declusterovacího parametru lze pozorovat změny v odhadech návratových úrovní a jejich intervalů spolehlivosti.

Prahový model postavený na odhadu extrémálního indexu využívá odhad Gomes (viz odstavec 5.2), který se ukazuje jako dostatečně robustní. Je také odhadnuta variabilita odhadu $\hat{\theta}_G$, přičemž tato je využita při konstrukci intervalových odhadů návratových úrovní. Nastavení pomocných parametrů odhadu $\hat{\theta}_G$ není uživateli dáno k dispozici, optimální hodnoty parametrů však vychází ze studií [60, 13, 5].



Obrázek C.1: Prostředí softwaru EVDest.

D

Popis programů k analýze extrémních hodnot

K analýze extrémních hodnot byly v softwaru Matlab vytvořeny následující skripty a funkce. Tyto využívají statistický a optimalizační toolbox (testováno pro verzi Matlab 2014a), který je nutné mít nainstalován. Některé z dále představených funkcí jsou také součástí softwaru EVDest (viz dodatek C).

| | |
|--------------------------|--|
| <code>db_hill.m</code> | Výpočet asymptotických kvadratických odchylek $q_{m,b}^*(k)$ Hillova odhadu EV indexu pro použití dvojité bootstrapového algoritmu odhadu vhodného počtu k horních pořadových statistik (viz algoritmus na str. 61). Funkce je vzhledem k časové náročnosti opakovaných výpočtů vhodnější než funkce <code>hill.m</code> , neboť počítá hodnoty $q_{m,b}^*(k)$ pro všechny hodnoty $b = 1, \dots, B$ a $k = 2, \dots, m$ současně. |
| <code>db_moe.m</code> | Výpočet odhadů asymptotických kvadratických odchylek EV indexu $q_{m,b}^*(k)$ pro použití dvojité bootstrapového algoritmu odhadu vhodného počtu k horních pořadových statistik (viz algoritmus na str. 61). Funkce je vzhledem k časové náročnosti opakovaných výpočtů vhodnější než funkce <code>moe.m</code> , neboť počítá hodnoty $q_{m,b}^*(k)$ pro všechny hodnoty $b = 1, \dots, B$ a $k = 2, \dots, m$ současně. |
| <code>decluster.m</code> | Vzorkování přibližně nezávislých pozorování ze souvislé časové řady pomocí technik <code>blocks</code> a <code>runs declusteringu</code> (odstavec 5.1). |
| <code>ems.m</code> | Vzorkování přibližně nezávislých agregovaných pozorování (míněno pro srážkové úhrny) ze souvislé časové řady na základě nastavitelné separační periody dle metodiky [96]. |
| <code>hill.m</code> | Hillův odhad EV indexu. |
| <code>intensity.m</code> | Vzorkování závislých agregovaných pozorování (míněno pro srážkové úhrny) ze souvislé časové řady. Agregace probíhá na základě disjunktních bloků dle sledované doby trvání (např. deště). |

| | |
|------------------------------|---|
| <code>jth_moment.m</code> | Pomocná funkce pro určení j -tých momentů pro momentové odhady parametrů ξ a σ_u (str. 48). |
| <code>moe.m</code> | Momentový odhad EV indexu a parametru měřítka GP rozdělení (str. 50, 52). |
| <code>moe3.m</code> | Pomocný momentový odhad $\hat{\xi}_{M3}(k)$ EV indexu založený na momentech 3. řádu (str. 62). |
| <code>mrlp_par_plot.m</code> | MRL plot, grafy stability parametrů GP rozdělení a testy dobré shody nadprahových hodnot. |
| <code>theta_GAT.m</code> | Odhad extrémálního indexu a jeho variability pomocí metody Gomes a Ancony-Navarreteho a Tawna (odstavec 5.2). |

Vlastní publikace ve vědeckém časopisu s IF

- [1] HOLEŠOVSKÝ, J., FUSEK, M., BLACHUT, V., MICHÁLEK, J. Comparison of precipitation extremes estimation using parametric and nonparametric methods. *Hydrological Sciences Journal*, 2015. Přijato k publikaci. DOI 10.1080/02626667.2015.1111517.
- [2] OLEJNÍČKOVÁ, Z., HOLEŠOVSKÝ, J., VÁVROVÁ, M., KRÁLOVÁ, Z., MICHÁLEK, J. Methylmercury in tissues of fish from the Svatka River, the Czech Republic. *Fresenius Environmental Bulletin*, 2014, vol. **23**, no. 12b, pp. 3319-3324.
- [3] Škarková, P., ZLÁMALOVÁ GARGOŠOVÁ, H., HOLEŠOVSKÝ, J., VÁVROVÁ, M., MICHÁLEK, J., OLEJNÍČKOVÁ, Z. Application of statistical methods for ecotoxicological data evaluation. *Fresenius Environmental Bulletin*, 2015, vol. **24**, no. 5, pp. 1692-1698.

Vlastní publikace indexované ve Scopus

- [4] HOLEŠOVSKÝ, J., FUSEK, M., MICHÁLEK, J. Modelling of precipitation extremes using parametric and nonparametric methods with automated threshold selection. *International Journal of Mathematics and Computers in Simulation*, 2015, vol. **9**, pp. 94-102.
- [5] HOLEŠOVSKÝ, J., FUSEK, M., MICHÁLEK, J. Extreme value estimation for correlated observations. In *Proceedings of 20th International Conference on Soft Computing MENDEL 2014*. Brno, Czech Republic, 2014, pp. 359-364. ISBN 978-80-214-4984-8.
- [6] HOLEŠOVSKÝ, J., POPELA, P., ROUPEC, J. Disruption in Congested Networks. In *Proceedings of 19th International Conference on Soft Computing MENDEL 2013*. Brno, Czech Republic, 2013, pp. 191-196. ISBN 978-80-214-4755-4.

Další vlastní publikace

- [7] FUSEK, M., HOLEŠOVSKÝ, J. *EVDest* [software]. Ver. 1.0, 2014 [citováno 9.1.2016]. Dostupné z <http://www.umat.feec.vutbr.cz/software/evdest.html>.
- [8] HOLEŠOVSKÝ, J., ČAMPULOVÁ, M., MICHÁLEK, J. Time-series outlier identification based on extreme value theory for stationary processes. Připraveno k odeslání, 2016.

- [9] HOLEŠOVSKÝ, J., FUSEK, M. Metody analýzy extrémních hodnot a jejich softwarová implementace. Odesláno k publikaci, 2015.
- [10] HOLEŠOVSKÝ, J., FUSEK, M., MICHÁLEK, J. Automated threshold selection for parametric and non-parametric estimates of intensity-duration-frequency curves. In *Proceedings of the 1st International Conference on Mathematical Methods & Computational Techniques in Science & Engineering, MMCTSE 2014*. Athens, Greece, 2014, pp. 114-118. ISBN 987-1-61804-256-9.
- [11] HOLEŠOVSKÝ, J., POPELA, P. Stochastic Extensions of a Traffic Assignment Problem. In *XX International Conference PDMU-2012: Problems of Decision Making under Uncertainties, Proceeding - Applied Papers*. Brno, Czech Republic, 2012, pp. 61-70. ISBN 978-80-7231-897-1.

Citovaná literatura

- [12] ADAMOWSKI, K. Regional analysis of annual maximum and partial duration flood data by nonparametric and L-moment methods. *Journal of Hydrology*, 2000, vol. **229**, no. 3-4, pp. 219-231.
- [13] ANCONA-NAVARRETE, M.A., TAWN, J.A. A Comparison of Methods for Estimating the Extremal Index. *Extremes*, 2000, vol. **3**, no. 1, pp. 5-38.
- [14] ANDĚL, J. *Základy matematické statistiky*. 3. vyd. Praha: Matfyzpress, 2005. 358 s. ISBN 80-86732-40-1.
- [15] ANDĚL, J. *Statistická analýza časových řad*. 1. vyd. Praha: Státní nakladatelství technické literatury, 1976. 271 s.
- [16] ANDĚL, J. *Statistické metody*. 3. vyd. Praha: Matfyzpress, 2007. 299 s. ISBN 80-86732-08-8.
- [17] ASHKAR, F., BOBÉE, B., RASMUSSEN, P., ROSBJERG, D. A perspective on the annual maximum flood approach to flood frequency analysis. *Stochastic and Statistical Models in Hydrology and Environmental Engineering, Extreme Values: Flood and Droughts*, 1994, vol. **1**, pp. 3-14.
- [18] BEIRLANT, J., GOEGEBEUR, Y., SEGERS, J., TEUGELS, J., DE WALL, D., FERRO, C. *Statistics of Extremes: Theory and Applications*. 1st edition. Chichester: Wiley, 2004. 514 p. ISBN: 978-0-471-97647-9.
- [19] BEN ZVI, A. Rainfall intensity-duration-frequency relationships derived from large partial duration series. *Journal of Hydrology*, 2009, vol. **367**, pp. 104-114.
- [20] BEN ZVI, A. Fit of probability distributions to upper sub-samples of partial duration series. *Stochastic and Statistical Methods in Hydrology and Environmental Engineering, Extreme Values: Flood and Droughts*, 1994, vol. **1**, pp. 95-107.

- [21] BOLÍVAR-CIMÉ, A., DÍAZ-FRANCÉS, E., ORTEGA, J. Optimality of profile likelihood intervals for quantiles of extreme value distributions: application to environmental disasters. *Hydrologic Sciences Journal*, 2015, vol. **60**, no. 4, pp. 651-670.
- [22] BONI, G., PARODI, A., RUDARI, R. Extreme rainfall events: Learning from rain-gauge time series. *Journal of Hydrology*, 2006, vol. **327**, pp. 304-314.
- [23] BRADLEY, R.C. Basic Properties of Strong Mixing Coonditions. A Survey and Some Open Questions. *Probability Surveys*, 2005, vol. **2**, pp. 107-144.
- [24] BRILHANTE, M.F., GOMES, M.I., PESTANA, D. A simple generalisation of the Hill estimator. *Computational Statistics and Data Analysis*, 2013, vol. **57**, pp. 518-535.
- [25] BUISSHAND, 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.
- [26] BURRIDGE, P., ROBERT TAYLOR, A.M. Additive outlier detection via extreme value theory. *Journal of Time Series Analysis*, 2006, vol. **27**, no. 5, pp. 685-701.
- [27] BURN, D.H. A framework for regional estimation of intenisty-duration-frequency (IDF) cruves. *Hydrological Processes*, 2014, vol. **28**, pp. 4209-4218.
- [28] BURN, D.H. The use of resampling for estimating confidence intervals for single site and pooled frequency analysis. *Hydrologic Sciences Journal*, 2003, vol. **48**, no. 1, pp. 25-38.
- [29] CAEIRO, F., GOMES, M.I., VANDEWALLE, B. Semi-parametric probability-weighted-moments estimation revisited. *Methodology and Computing in Applied Probability*, 2014, vol. **16**, no. 1, pp. 1-29.
- [30] CAEIRO, F., GOMES, M.I. Semi-parametric tail inference through probability-weighted-moments. *Journal of Statistical Planning and Inference*, 2010, vol. **141**, pp. 937-950.
- [31] CASELLA, G., BERGER, R.L. *Statistical Inference*. 2nd edition. Pacific Grove: Thomson Learning, 2002. 660 p. ISBN-13 978-0-495-39187-6.
- [32] CERCHIARA, R.R. *Mathematical and Statistical Methods in Insurance and Finance*. 1st edition. Milan: Springer, 2008. 208 p. ISBN 978-88-470-0703-1. Chapter 3, FFT, extreme value theory and simulation to model non-life insurance claims dependences, pp. 61-65.
- [33] COLES, S. *An Introduction to Statistical Modeling of Extreme Values*. 3rd edition. London: Springer, 2001. ISBN 1-85233-459-2.
- [34] COLES, S., PERICCHI, L.R., SISSON, S. A fully probabilistic approach to extreme rainfall modeling. *Journal of Hydrology*, 2003, vol. **273**, pp. 35-50.
- [35] CUNNANE, C. A particular comparison of annual and partial duration series methods of flood frequency predictions. *Journal of Hydrology*, 1973, vol. **18**, no. 3-4, pp. 257-271.

- [36] D'AGOSTINO, R.B., STEPHENS, M.A. *Goodness-of-fit techniques*. 1st edition. New York: Dekker, 1986. 576 p. ISBN 978-0824774875.
- [37] DANIELSSON, J., DE HAAN, L., PENG, L., DE VRIES, C.G. Using a bootstrap method to choose the sample fraction in the tail index estimation. *Journal of Multivariate Analysis*, 2001, vol. **76**, pp. 226-248.
- [38] DE HAAN, L., FERREIRA, A. *Extreme Value Theory: An Introduction*. 1st edition. New York: Springer, 2006. ISBN-10 0-387-23946-4.
- [39] DE HAAN, L., PENG, L. Comparison of tail index estimators. *Statistica Neerlandica*, 1998, vol. **52**, pp. 60-70.
- [40] DEHEUVELS, P., HÄUSLER, E., MASON, D.M. Almost sure convergence of the Hill estimator. *Mathematical Proceedings of the Cambridge Philosophical Society*, 1988, vol. **104**, pp. 371-381.
- [41] DEKKERS, A.L.M., EINMAHL, J.H.J., DE HAAN, L. A moment estimator for the index of an extreme-value distribution. *Annals of Statistics*, 1989, vol. **17**, no. 4, pp. 1833-1855.
- [42] DEME, E.H., GARDES, L., GIRARD, S. On the estimation of the second order parameter for heavy-tailed distribution. *REVSTAT*, 2013, vol. **11**, no. 3, pp. 277-299.
- [43] DENZEL, G.E., O'BRIEN, G.L. Limit theorems for extreme values of chain-dependent processes. *Annals of Probability*, 1975, vol. **3**, pp. 773-779.
- [44] DIEBOLT, J., GUILLOU, A., RACHED, I. Approximation of the distribution of excesses through a generalized probability-weighted moments method. *Journal of Statistical Planning and Inference*, 2007, vol. **137**, no. 3, pp. 841-857.
- [45] DRAISMA, G., DE HAAN, L., PENG, L., PEREIRA, T.T. A Bootstrap-based Method to Achieve Optimality in Estimating the Extreme-Value Index. *Extremes*, 1999, vol. **2**, no. 4, pp. 367-404.
- [46] DUPUIS, D.J., FIELD, C.A. Large Wind Speeds: Modeling and Outlier Detection. *Journal of Agricultural, Biological, and Environmental Statistics*, 2004, vol. **9**, no. 1, pp. 105-121.
- [47] EFRON, B. Bootstrap methods: another look at the jackknife. *Annals of Statistics*, 1979, vol. **7**, pp. 1-26.
- [48] EFRON, B., TIBSHIRANI, R.J. *An Introduction to the Bootstrap*. 1st edition. Boca Raton: Chapman & Hall, 1994. 436 p. ISBN 0-412-04231-2.
- [49] EMBRECHTS, P., KLÜPPELBERG, C., MIKOSCH, T. *Modelling Extremal Events: for Insurance and Finance*. 1st edition. Berlin: Springer, 1997. 645 p. ISBN 978-3-540-60931-5.
- [50] FAWCETT, L., WALSHAW, D. Estimating return levels from serially dependent extremes. *Environmentrics*, 2012, vol. **23**, pp. 272-283.

- [51] FERRO, C.A.T., SEGERS, J. Inference for clusters of extreme values. *Journal of Royal Statistical Society: Series B*, 2003, vol. **65**, no. 2, pp. 545-556.
- [52] FISHER, R.A., TIPPETT, L.H.C. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceeding of Cambridge Philosophical Society*, 1928, vol. **24**, pp. 180-190.
- [53] FRAGA ALVES, M.I. A Location Invariant Hill-Type Estimator. *Extremes*, 2001, vol. **4**, no. 3, pp. 199-217.
- [54] FRAGA ALVES, M.I., GOMES, M.I., DE HAAN, L. A new class of semi-parametric estimators of the second order parameter. *Portugaliae Mathematica*, 2003, vol. **60**, no. 1, pp. 193-213.
- [55] FRAGA ALVES, M.I., GOMES, M.I., DE HAAN, L., NEVES, C. A Note on Second Order Conditions in Extreme Value Theory: Linking General and Heavy Tail Conditions. *REVSTAT*, 2007, vol. **5**, no. 3, pp. 285-304.
- [56] FUKUTOME, S., LINIGER, M.A., SÜVEGES, M. Automatic threshold and run parameter selection: a climatology for extreme hourly precipitation in Switzerland. *Theoretical and Applied Climatology*, 2014, vol. **120**, no. 3, pp. 403-416.
- [57] FUSEK, M. Rozdělení extrémních hodnot a jejich aplikace. Vysoké učení technické v Brně, Brno, 2013. Dizertační práce, 136 s.
- [58] GASSER, T., MÜLLER, H.G., MAMMITZSCH, V. Kernels for Nonparametric Curve Estimation. *Journal of the Royal Statistical Society: Series B*, 1985, vol. **47**, no. 2, pp. 238-252.
- [59] GNEDENKO, B.V. Sur la distribution limite du terme maximum d' une série aléatoire. *Annals of Mathematics*, 1943, vol. **44**, pp. 423-453.
- [60] GOMES, M.I. On the estimation of parameters of rare events in environmental time series. In *Statistics for the Environment 2: Water Related Issues*. Chichester: Wiley, 1993. pp. 225-241.
- [61] GOMES, M.I., DE HAAN, L., PENG, L. Semi-parametric estimation of the second order parameter - asymptotic and finite sample behavior. *Extremes*, 2002, vol. **5**, no. 4, pp. 387-414.
- [62] GOMES, M.I., FIGUEIREDO, F., MARTINS, M.J., NEVES, M. Resampling methodologies and reliable tail estimation. *South African Statistical Journal*, 2015, vol. **49**, no. 1, pp. 1-20.
- [63] GOMES, M.I., MARTINS, M.J., NEVES, M. Alternatives to a Semi-Parametric Estimator of Parameters of rare Events—The Jackknife Methodology. *Extremes*, 2000, vol. **3**, no. 3, pp. 207-229.
- [64] GOMES, M.I., MENDONÇA, S., PESTANA, D. Adaptive reduces-bias tail index and VaR estimation via the bootstrap methodology. *Communications in Statistics - Theory and Methods*, 2011, vol. **40**, no. 16, pp. 2946-2968.

- [65] GOMES, M.I., OLIVEIRA, O. The Bootstrap Methodology in Statistics of Extremes - Choice of the Optimal Sample Fraction. *Extremes*, 2001, vol. **4**, no. 4, pp. 331-358.
- [66] GOMES, M.I., RODRIGUES, L.H., PEREIRA, H., PESTANA, D. Tail index and second-order parameters' semi-parametric estimation based on the log-excesses. *Journal of Statistical Computation and Simulation*, 2010, vol. **80**, no. 6, pp. 653-666.
- [67] GUMBEL, E.J. *Statistics of Extremes*. 1st edition. New York: Columbia University Press, 1958. 396 p. ISBN-13 978-1626549876
- [68] GONZALES, J., RODRIGUEZ, D., SUED, M. Threshold selection for extremes under a semiparametric model. *Statistical Methods and Applications*, 2013, vol. **22**, no. 4, pp. 481-500.
- [69] HAILEGEORGIS, T.T., THOROLFSSON, S.T., ALFREDSEN, K. Regional frequency analysis of extreme precipitation with consideration on uncertainties to update IDF curves for the city of Trondheim. *Journal of Hydrology*, 2013, vol. **498**, pp. 305-318.
- [70] HALL, P. Using bootstrap to estimate mean squared error and selecting parameter in nonparametric problems. *Journal of Multivariate Analysis*, 1990, vol. **32**, pp. 177-203.
- [71] 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.
- [72] HERRMANN, E. Local Bandwidth Choice in Kernel Regression Estimation. *Journal of Computational and Graphical Statistics*, 1997, vol. **6**, no. 1, pp. 35-54.
- [73] HODGE, V.J., AUSTIN, J. A Survey of Outlier Detection Methodologies. *Artificial Intelligence Review*, 2004, vol. **22**, pp. 85-126.
- [74] HOSKING, J.R.M., WALLIS, J.R. *Regional frequency analysis: An approach based on L-moments*. 1st edition. Cambridge: Cambridge University Press, 1997. 224 s. ISBN 05-214-3045-3.
- [75] HOSKING, J.R.M., WALLIS, J.R. Parameter and Quantile Estimation for the Generalized Pareto Distribution. *Technometrics*, 1987, vol. **29**, no. 3, pp. 339-349.
- [76] HRDLIČKOVÁ, Z., MICHÁLEK, J., KOLÁŘ, M., VESELÝ, V. Identification of factors affecting air pollution by dust aerosol PM10 in Brno City, Czech Republic. *Atmospheric Environment*, 2008, vol. **42**, pp. 8661-8673.
- [77] HSING, T., HUSLER, J., LEADBETTER, M.R. On the exceedance point process for a stationary sequence. *Probability Theory and Related Fields*, 1988, vol. **78**, pp. 97-112.
- [78] HÜBNEROVÁ, Z., MICHÁLEK, J. Analysis of daily average PM10 predictions by generalized linear models in Brno, Czech Republic. *Atmospheric Pollution Research*, 2014, vol. **5**, pp. 471-476.

- [79] CHOULAKIAN, V., STEPHENS, M.A. Goodness-of-fit tests for the generalized Pareto distribution. *Technometrics*, 2001, vol. **43**, no. 4, pp. 478-484.
- [80] CHOW, V.T., MAIDMENT, D.R., MAYS, L.W. *Applied Hydrology*. 1st edition. New-York: McGraw-Hill, 1988. 572 s. ISBN 978-0070108103.
- [81] JIANG, Y., ZHUANG, Q. Extreme value analysis of wildfires in Canadian boreal forest ecosystems. *Canadian Journal of Forest Research*, 2011, vol. **41**, no. 9, pp. 1836-1851.
- [82] JENKINSON, A.F. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 1955, vol. **81**, pp. 158-171.
- [83] KAMPA, M., CASTANAS, E. Human health effects of air pollution. *Environmental Pollution*, 2008, vol. **151**, no. 2, pp. 362-367.
- [84] KHALIQ, M.N., OUARDA, T.B.M.J., ONDO, J.-C., GACHON, P., BOBÉE, B. Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observation: A review. *Journal of Hydrology*, 2006, vol. **329**, pp. 534-552.
- [85] KLÜPPELBERG, C., MIKOSCH, T. Large deviations of heavy-tailed random sums with applications in insurance and finance. *Journal of Applied Probability*, 1997, vol. **34**, no. 2, pp. 293-308.
- [86] KOTZ, S., NADARAJAH, S. *Extreme Value Distributions: Theory and Applications*. 1st edition. London: Imperial College Press, 2000. 185 p. ISBN 1860942245.
- [87] KVĚTON, V., ZAHRADNÍČEK, J., ŽÁK, M. Kontrola kvality a digitalizace ombrogramů v ČHMÚ. *Meteorologické zprávy*, 2004, roč. **57**, č. 2, s. 47-52.
- [88] KYSELÝ, J. Coverage probability of bootstrap confidence intervals in heavy-tailed frequency models, with application to precipitation data. *Theoretical and Applied Climatology*, 2010, vol. **101**, no. 3, pp. 345-361.
- [89] KYSELÝ, J., PICEK, J., BERANOVÁ, R. Estimating extremes in climate change simulations using the peaks-over-threshold method with a non-stationary threshold. *Global and Planetary Change*, 2010, vol. **72**, no. 1-2, pp. 55-68.
- [90] LAGARIAS, J.C., REEDS, J.A., WRIGHT, M.H., WRIGHT, P.E. Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM Journal on Optimization*, 1998, vol. **9**, pp. 112-147.
- [91] LARSEN, R.J., MARX, M.L. *An Introduction to Mathematical Statistics and Its Applications*. 5th edition. Boston: Pearson Education, 2012. ISBN-13 978-0-312-69394-5.
- [92] LAURINI, F., TAWN, J.A. New Estimators for the Extremal Index and Other Cluster Characteristics. *Extremes*, 2003, vol. **6**, pp. 189-211.

- [93] LEADBETTER, M.R., LINDGREN, G., ROOTZÉN, H. *Extremes and Related Properties of Random Sequences and Series*. 1st edition. New York: Springer, 1983. 336 p. ISBN 978-1-4612-5451-5.
- [94] LEHMANN, E.L., CASELLA, G. *Theory of Point Estimation*. 2nd edition. New York: Springer, 1998. 616 p. ISBN 0-387-98502-6.
- [95] 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. 795-769.
- [96] MADSEN, H., MIKKELSEN, P.S., ROSBJERG, D., HARREMÖES, P. Regional estimation of rainfall intensity-duration-frequency curves using generalized least squares regression of partial duration series. *Water Resources Research*, 2002, vol. **38**, no. 11, pp. 21-1–21-11.
- [97] 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.
- [98] MAILHOT, A., LACHANCE-CLOUTIER, S., TALBOT, G., FAVRE, A.-C. Regional estimates of intense rainfall based on the Peak-Over-Threshold (POT) approach. *Journal of Hydrology*, 2013, vol. **476**, pp. 188-199.
- [99] MICHÁLEK, J., FUSEK, M., HELLEBRAND, R. Srovnání statistických metod odhadů IDF křivek. In *Workshop Adolfa Patery 2012 Extrémní hydrologické jevy v povodích*. Brno, 2012, s. 163-170. ISBN 978-80-02-02423-1.
- [100] NADARAYA, E.A. On estimating regression. *Theory Probability and its Applications*, 1964, vol. **10**, pp. 186-190.
- [101] NEVES, C., FRAGA ALVES, M.I. Reiss and Thomas' automatic selection of the number of extremes. *Computational Statistics & Data Analysis*, 2004, vol. **47**, pp. 689-704.
- [102] NORTHROP, P.J., COLEMAN, C.L. Improved threshold diagnostic plots for extreme value analyses. *Extremes*, 2014, vol. **17**, pp. 289-303.
- [103] NOVAK, S.Y., WEISSMAN, I. On blocks and runs estimator of the extremal index. *Journal of Statistical Planning and Inference*, 1998, vol. **66**, no. 2, pp. 281-288.
- [104] O'BRIEN, G.L. Extreme values for stationary and Markov sequences. *Annals of Probability*, 1987, vol. **15**, pp. 281-291.
- [105] OVEREEM, A., BUISHAND, A., HOLLEMAN, I. Rainfall depth-duration-frequency curves and their uncertainties. *Journal of Hydrology*, 2008, vol. **348**, pp. 124-134.
- [106] PANJER, H.H. *Operational Risk: Modeling Analytics*. 1st edition. Hoboken: Wiley Interscience, 2006. 431 p. ISBN 978-0471760894

- [107] PENG, L., QI, Y. Maximum likelihood estimation of extreme value index for irregular cases. *Journal of Statistical Planning and Inference*, 2009, vol. **139**, pp. 3361-3376.
- [108] PÉREZ-ZANÓN, N., CASAS-CASTILLO, M.C., RODRÍGUEZ-SOLA, R., PENA, J.C., RIUS, A., SOLÉ, J.G., REDANO, A. (2015). Analysis of extreme rainfall in the Ebre Observatory (Spain). *Theoretical and Applied Climatology*, Springer, 1-10. DOI 10.1007/s00704-015-1476-0.
- [109] PICKANDS, J. Statistical inference using extreme order statistics. *Annals of Statistics*, 1975, vol. **3**, no. 1, pp. 119-131.
- [110] QUINTELA-DEL-RÍO, A., FRANCISCO-FERNÁNDEZ, M. Nonparametric functional data estimation applied to ozone data: Prediction and extreme value analysis. *Chemosphere*, 2011, vol. **82**, no. 6, pp. 800-808.
- [111] Rada Evropské unie. Směrnice Rady 1999/30/EC ze dne 22. dubna 1999 o mezních hodnotách pro oxid siřičitý, oxid dusičitý a oxidy dusíku, částice a olovo ve vnějším ovzduší. *Úřední věstník Evropské unie*, 1999, L 163 0041-0060.
- [112] RESNICK, S.I. *Extreme Values, Regular Variation, and Point Processes*. 1st edition. New York: Springer, 1987. 320 p. ISBN 978-0-387-75952-4.
- [113] ROBERTS, S.J. Novelty detection using extreme value statistics. *IEE Proceedings: Vision, Image and Signal Processing*, 1999, vol. **146**, no. 3, pp. 124-129.
- [114] ROSBJERG, D., MADSEN, H., RASMUSSEN, P.F: Prediction in partial duration series with generalized Pareto distribution exceedances. *Water Resources Research*, 1992, vol. **14**, pp. 345-356.
- [115] RUBINSTEIN, R.Y., KROESE, D.P. *Simulation and the Monte Carlo Method*. 2nd edition. Hoboken: Wiley, 2008. 372 p. ISBN: 978-0-470-17794-5.
- [116] SCARROTT, C., MACDONALD, A. A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT*, 2012, vol. **10**, no. 1, pp. 33-60.
- [117] SHAO, J., TU, D. *The Jackknife and Bootstrap*. 1st edition. New York: Springer, 1995. 517 p. ISBN: 978-1-4612-0795-5.
- [118] SHINYIE, W.L., ISMAIL, N., JEMAIN, A.A. Semi-parametric Estimation fo Selecting Optimal Threshold of Extreme Rainfall Events. *Water Resources Management*, 2013, vol. **27**, no. 7, pp. 2325-2352.
- [119] SMITH, R.L. Estimating tails of probability distribution. *Annals of Statistics*, 1987, vol. **15**, no. 3, pp. 1174-1207.
- [120] SMITH, R.L. Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 1985, vol. **72**, no. 1, pp. 67-90.
- [121] SMITH, R.L., WEISSMAN, I. Estimating the Extremal Index. *Journal of the Royal Statistical Society, Series B*, 1994, vol. **56**, pp. 515-528.

LITERATURA

- [122] SRIKANTHAN, S. Annual maximum or partial duration series: A review. In *Proceedings of the 34th Hydrology and Water Resources Symposium, HWRS 2012*. Sydney, Australia, 2012, pp. 407-414.
- [123] SÜVEGES, M. Likelihood estimation of the extremal index. *Extremes*, 2007, vol. **10**, pp. 41-55.
- [124] SÜBEGES, M., DAVISON, A.C. Model misspecification in peaks over threshold analysis. *The Annals of Applied Statistics*, 2010, vol. **4**, no. 1, pp. 203-221.
- [125] 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, pp. 275-284.
- [126] TRUPL, J. *Intenzity krátkodobých dešťů v povodích Labe, Odry a Moravy*. Praha: Výzkumná ústav vodohospodářský, 1958. Práce a studie, 34 s.
- [127] UCHIDA, M. Traffic data analysis using extreme value theory. *NTT R and D*, 2003, vol. **52**, pp. 239-244.
- [128] UM, M.-J., CHO, W., HEO, J.-H. A comparative study of the adaptive choice of thresholds in extreme hydrologic events. *Stochastic Environmental Research and Risk Assessment*, 2010, vol. **24**, no. 5, pp. 611-623.
- [129] VAN MONTFORD, M.A.J., WITTER, J.V. The generalized Pareto distribution applied to rainfall depths. *Hydrological Science Journal*, 1986, vol. **31**, pp. 151-162.
- [130] VON MISES, R. La distribution de la plus grande de n valeurs. *American Mathematical Society, Selected Papers*, Providence, 1954.
- [131] WADSWORTH, J.L., TAWN, J.A. Likelihood-based procedures for threshold diagnostics and uncertainty in extreme value modelling. *Journal of Royal Statistical Society: Series B*, 2012, vol. **74**, no. 3, pp. 543-567.
- [132] WADSWORTH, J.L., TAWN, J.A., JONATHAN, P. Accounting for choice of measurement scale in extreme value modeling. *The Annals of Applied Statistics*, 2010, vol. **4**, no. 3, pp. 1558-1578.
- [133] WAND, M.P., JONES, M.C. *Kernel Smoothing*. 1st edition. Boca Raton: CRC Press, 1994. 224 p. ISBN-13: 978-0412552700.
- [134] 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
- [135] WORMS, J., WORMS, R. Estimation of second order parameters using probability weighted moments. *ESAIM - Probability and Statistics*, 2012, vol. **16**, pp. 97-113.
- [136] YEVJEVICH, V. *Probability and statistics in Hydrology*. 1st edition. Fort Collins: Water Resources Publications, 1972. 302 p. ISBN 978-0918334008.
- [137] ZHOU, C. Existence and consistency of the maximum likelihood estimator for the extreme value index. *Journal of Multivariate Analysis*, 2009, vol. **100**, pp. 794-815.

Seznam symbolů a zkratek

| | |
|-------------------------------------|---|
| $\rightarrow, \uparrow, \downarrow$ | Bodová konvergence, konvergence zleva, konvergence zprava |
| \xrightarrow{P} | Konvergence podle pravděpodobnosti |
| \xrightarrow{d} | Konvergence v distribuci |
| $\stackrel{d}{=}$ | Rovnost v distribuci |
| $\stackrel{d}{\approx}$ | Přibližná rovnost v distribuci, tj. $X \stackrel{d}{\approx} Y \Rightarrow \forall x : P(X \leq x) \approx P(Y \leq x)$ |
| \simeq | Asymptotická ekvivalence, tj. $a(t) \simeq b(t) \Rightarrow \lim_{t \rightarrow \infty} a(t)/b(t) = 1$ |
| c_+, c_- | $\max(c, 0), \min(c, 0)$ |
| $\lfloor c \rfloor$ | Největší celé číslo menší než nebo rovné c |
| $o(g)$ | Landauova notace, $f(x) \in o(g(x)) \Leftrightarrow \lim_{x \rightarrow \infty} f(x)/g(x) = 0$ |
| $\mathbf{1}_{[A]}$ | Charakteristická funkce jevu A |
| P | Pravděpodobnost |
| $\text{supp}(f)$ | Nosič funkce f |
| \mathbb{R} | Množina reálných čísel |
| Γ | Gama funkce |
| $N(\mu, \sigma^2)$ | Normální rozdělení s parametrem polohy μ a parametrem měřítka σ^2 |
| X, Y | Náhodné veličiny |
| \mathbf{X} | Náhodný vektor |
| $X \sim N(\mu, \sigma^2)$ | Náhodná veličina X má dané rozdělení |
| F | Distribuční funkce |
| F_n | Empirická distribuční funkce |
| f^{\leftarrow} | Zleva spojitá inverzní funkce k f , str. 17 |
| U | Kvantilová funkce chvostu, $U = (1/(1 - F))^{\leftarrow}$, str. 18 |
| ξ | Index extrémní hodnoty |
| $\rho \leq 0$ | Parametr druhého řádu |
| \mathcal{R}_α | Množina regulárně se měnících funkcí s indexem α |
| θ | Extremální index |
| σ_u | Parametr měřítka GP rozdělení |
| $\hat{\lambda}$ | Odhad parametru λ |
| x^*, x_* | Pravý a levý koncový bod rozdělení, str. 16 |
| G_ξ, G | Distribuční funkce GEV rozdělení s EV indexem ξ , str. 20 |

SEZNAM ZKRATEK

| | |
|----------------------|--|
| H | Distribuční funkce GP rozdělení, str. 39 |
| D_ξ | Limitní funkce podmínky atraktivity, str. 23 |
| $H_{\xi,\rho}$ | Limitní funkce podmínky druhého řádu, str. 31 |
| $\mathcal{D}(G_\xi)$ | Obor atraktivity GEV rozdělení s EV indexem ξ |
| a, A | Pomocná funkce prvního a druhého řádu, str. 23 a 31 |
| Φ | Distribuční funkce standardizovaného normálního rozdělení |
| χ_n^2 | χ^2 rozdělení s n stupni volnosti |
| EV index | Index extrémní hodnoty ξ |
| GEV rozdělení | Zobecněné rozdělení extrémních hodnot (Generalized Extreme Value) |
| GP rozdělení | Zobecněné Paretovo rozdělení (Generalized Pareto) |
| KM proces | Proces klouzavých maxim |
| maxAR proces | Max-autoregresní proces |
| MT-GP | Multiple-Threshold Generalized Pareto; metoda určení prahové hodnoty |
| MV metoda | Metoda maximální věrohodnosti |
| MV odhad | Maximálně věrohodný odhad |
| RV index | Index regulární variace |