Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

BAKALÁŘSKÁ PRÁCE

Karatun Ksenia

Kombinovaná rozdělení výší škod

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce:	RNDr. Lucie Mazurová, Ph.D.
Studijní program:	Matematika
Studijní obor:	Finanční matematika

Praha 2018

Prohlašuji, že jsem tuto bakalářskou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

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

V .... dne .... Podpis autora

Na tomto místě bych chtěla poděkovat vedoucí práce RNDr. Lucii Mazurové, Ph.D., za trpělivost, věnovaný čas a cenné připomínky poskytnuté při vypracování této práce. Dále děkuji své rodině za podporu během psaní práce a celého studia.

Název práce: Kombinovaná rozdělení výší škod

Autor: Karatun Ksenia

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Lucie Mazurová, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: V této práce se věnujeme kombinovaným rozdělením, která lze použít pro modelovaní výše škod v určitých odvětvích neživotního pojištění. V první části je definovan obecný kombinovaný model a jsou uvedené základní vlastnosti. V druhé části jsou popsány modely, založené na Weibullovo rozdělení a rozděleni transformované rodiny beta se svými hustotami a základními charakteristikami. Ve třeti části je popsán algoritmus odhadu parametrů kombinovaného modelu a metody hodnocení vhodností modelu. Poslední část je věnována praktickému modelování nad dvěma soubory realných dat.

Klíčová slova: kombinované rozdělení, výše škody, metoda maximální věrohodnosti.

Title: Composite distributions of loss sizes

Author: Karatun Ksenia

Department: Department of Probability and Mathematical Statistics

Supervisor: RNDr. Lucie Mazurová, Ph.D., Department of Probability and Mathematical Statistics

Abstract: In this work, we deal with composite distributions that can be used to model loss sizes in some specific classes of non-life insurance. The first part contains definition of the general composite model and its special features. The second part describes models that are made up by piecing together Weibull distribution and distributions belonging to a family of transformed beta distributions. The third part describes algorithm that computes the maximum likelihood estimators for parameters of composite distribution and criteria of the relative quality of statistical models. In the last part we apply composite models to two real data sets.

Keywords: composite distribution, loss size, MLE method.

Obsah

Úvod

Stanovení předpokládaného počtu škod a jejich výši je jedním z nejdůležitějších problémů neživotního pojištění. V neživotním pojištění výskyt škody se považuje za náhodilou událost a její výše za náhodnou veličinu, která se dá popsat vhodným pravděpodobnostním rozdělením. Cílem této práci je seznamit čtenáře s metodou modelování výše škod rozdělením, které vznikne kombinací dvou parametrických rozdělení.

Práce je členěna do čtyř kapitol. V první kapitole definujeme kombinovaný model a uvedeme základní vlastností modelu. V druhé kapitole budou představeny tři kombinovaná rozděleni, která jsou kombinací Weibullovo rozdělení a rozděleni transformované rodiny beta, jejich vlastnosti a základní charakteristiky. V této kapitole budou shrnuty modely, které popsali ve svých článcích Scollnik a Sun (2012) a Abu Bakar, Hamzah, Maghsoudi a Nadarajah (2015). Třetí kapitola je věnována popisu algoritmu odhadu parametrů kombinovaných rozdělení metodou maximální věrohodnosti na příkladu kombinovaného Weibull-Paretova rozdělení. Potom uvedeme metody hodnocení uvažovaných modelů. V poslední kapitole ukážeme praktickou aplikaci kombinovaných rozdělení na reálných datech.

Kapitola 1

Kombinované rozdělení

1.1 Zavedení modelu

Typickou vlastností výše škod v určitých odvětvích pojištění je kladná šikmost, která je způsobená vyšším počtem výskytu menších škod. Základní modely pro popis vyše škod jsou například exponenciální, logaritmicko-normální a Weibullovo rozdělení. Pokud navíc data svědčí o významné pravděpodobnosti velmi vysokých škod, výše uvedená rozdělení se nehodí k modelování celého souboru dat, jelikož jejich hustoty pro $x \to \infty$ konvergují k nule poměrně rychle. Vyšší škody jsou obvykle modelované zvlášť rozdělením s těžkým chvostem, jako například je Paretovo rozdělení. Cooray a Ananda (2005) poprvé navrhli způsob, jak popsat data obou typů jedním rozdělením - model, založený na kombinovaném rozdělení.

Abychom definovali kombinované rozdělení, nejdřív musíme zavést pojem useknutého rozdělení. Dále uvedeme odvození jeho hustoty.

Definice 1 (Useknuté rozdělení). *Nechť $(a,b] \subset (0,\infty)$ je polouzavřený interval. Podmíněné rozdělení náhodné veličiny $X \in (0,\infty)$, za podmínky $X \in (a,b]$, s distribuční funkci*

$$F_X^*(x) = P[X \leq x | a < X \leq b] \tag{1.1}$$

nazveme useknuté rozdělení náhodné veličiny X.

Hustota useknutého rozdělení. Nechť X je spojitá náhodná veličina s hustotou rozdělení $f_X(x)$, $x \in [0,\infty)$. Hustota useknutého rozdělení náhodné veličiny X, za podmínky $a < X \leq b$, je určena vztahem:

$$f_X^*(x) = \begin{cases} \frac{f_X(x)}{F(b)-F(a)}, & \text{je-li } a < x \leq b \\ 0, & \text{jinak} \end{cases} \tag{1.2}$$

Vztah (1.2) odvodíme pomocí podmíněné pravděpodobnosti :

$$F_X^*(x) = P[X \leq x | a < X \leq b] = \frac{P[a < X \leq x]}{P[a < X \leq b]} = \frac{\int_a^x f_X(y)dy}{\int_a^b f_X(y)dy}$$

Derivovaním distribuční funkce dojdeme k hustotě useknutého rozdělení:

$$f_X^*(x) = [F_X^*(x)]' = \frac{f_X(x)}{\int_a^b f_X(y)dy} = \frac{f_X(x)}{F(b)-F(a)}, x \in (a,b].$$

Dále uvedeme definici kombinovaného modelu, sestrojeného kombinací n pravděpodobnostních rozdělení. Definujeme model pomoci hustoty.

Definice 2 (Hustota kombinovaného rozdělení)**.** *Uvažujeme n spojitých pravděpodobnostních rozdělení s nosičem $(0,\infty)$ a hustotami $f_1(x), \dots f_n(x)$. Spojitá náhodná veličina $X \in (0,\infty)$ má kombinované rozdělení právě tehdy, jestliže hustota její je nezáporná funkce, která je definovana vztahem:*

$$f_X(x) = \begin{cases} a_1 f_1^*(x), & \text{je-li } c_0 < x \leq c_1 \\ a_2 f_2^*(x), & \text{je-li } c_1 < x \leq c_2 \\ \vdots & \vdots \\ a_n f_n^*(x), & \text{je-li } c_{n-1} < x \leq c_n \end{cases},$$

kde

$$f_i^* = \frac{f_i(x)}{\int_{c_{i-1}}^{c_i} f_i(x)dx}$$

je hustota useknutého rozdělení z definici 1, kde X má hustotu $f_i(x)$, za podmínky $c_{i-1} < X \leq c_i$, pro $i = 1, \dots ,n$. Navic platí :

$$\sum_{i=1}^{n} a_i = 1,$$

kde $a_i \in (0,1)$ jsou váhy pro $i = 1, \dots ,n$

V dálším textu budeme uvažovat speciální případ definice 2 - kombinaci dvou rozdělení s koncovým bodem θ, který bude parametrem modelu:

$$f_X(x) = \begin{cases} a_1 f_1^*(x), & \text{je-li } 0 < x \leq \theta \\ a_2 f_2^*(x), & \text{je-li } \theta < x < \infty \end{cases} \tag{1.3}$$

1.2 Vlastnosti modelu

V této podkapitole zavedeme další vlastnosti modelu - spojitost a diferencovatelnost v koncovém bodě. Tyto vlastnosti omezují široký model, definovaný vztahem (1.3).

Spojitost. Vztah (1.3) nedává žádnou informaci o spojitosti hustoty kombinovaného rozdělení v koncovém bodě θ. Aby levá část rozdělení byla spojená s pravou v bodě θ, přidáme podmínku :

$$\lim_{x \to \theta-} f_X(x) = \lim_{x \to \theta+} f_X(x)$$

Diferencovatelnost. Funkce hustoty má byt dostatečně hladká v koncovém bodě θ, druhou podmínkou bude:

$$\lim_{x \to \theta-} f'(x) = \lim_{x \to \theta+} f'(x)$$

Váhy. V zavedeném modelu součet vah se má rovnat 1. Sestrojíme-li model ze dvou rozdělení, váhy a_1,a_2 lze napsat ve tvaru $a_1 = \lambda$, $a_2 = 1 - \lambda$, kde $0 < \lambda < 1$. Z definice 2 a podmínky spojitosti odvodíme váhy:

$$\lambda * f_1^*(\theta) = (1 - \lambda) * f_2^*(\theta)$$

Po dosazení:

$$f_1^*(\theta) = \frac{f_1(\theta)}{F_1(\theta) - F_1(-\infty)} = \frac{f_1(\theta)}{F_1(\theta)} \quad f_2^*(\theta) = \frac{f_2(\theta)}{F_2(\infty) - F_2(\theta)} = \frac{f_2(\theta)}{1 - F_2(\theta)} \tag{1.4}$$

Dostaváme:

$$\lambda = \frac{f_2(\theta)F_1(\theta)}{f_2(\theta)F_1(\theta) + f_1(\theta)[1 - F_2(\theta)]} \tag{1.5}$$

Výše uvedený vzorec ukazuje, že za platnosti podmínky spojitosti se dají váhy vyjádřit jako funkce zbyvajících parametrů modelu.

Kapitola 2

Kombinovaná rozdělení, založená na Weibullovo rozdělení a rozdělení transformované rodiny beta

Weibullovo rozdělení, jak jsme naznačili na začátku kapitoly 1, se velmi často používá k modelování výše škod. Dvouparametrické Weibullovo rozdělení má dva parametry - parametr tvaru a parametr měřítka. Je zobecněním exponenciálního rozdělení, ale může aproximovat lognormální a normální rozdělení, v závislosti na hodnotě parametru tvaru. V důsledku této flexibility se často uplatňuje v praxi.

2.1 Weibullovo rozdělení

Weibullovo rozdělení je spojité dvouparametrické rozdělení s parametrem tvaru $\beta > 0$ a parametrem měřítka $\gamma > 0$. Distribuční funkce je daná vztahem:

$$F(x) = 1 - \exp\left\{-\left(\frac{x}{\gamma}\right)^{\beta}\right\}, \quad x \geq 0 \tag{2.1}$$

a hustota:

$$f(x) = \frac{\beta}{x}\left(\frac{x}{\gamma}\right)^{\beta} exp\left\{-\left(\frac{x}{\gamma}\right)^{\beta}\right\}, \quad x \geq 0. \tag{2.2}$$

K výpočtu střední hodnoty a rozptylu použijme nasledující vztahy:

$$\begin{aligned} EX &= \int_{-\infty}^{\infty} x * f(x)dx \\ varX &= EX^2 - (EX)^2. \end{aligned} \tag{2.3}$$

Střední hodnota Weibullova rozdělení je pak rovna:

$$\begin{aligned} E(X) &= \int_0^{\infty} x * \frac{\beta}{x}\left(\frac{x}{\gamma}\right)^{\beta} exp\left\{-\left(\frac{x}{\gamma}\right)^{\beta}\right\} dx \overset{s}{=} \left[\begin{array}{l} y = \left(\frac{x}{\gamma}\right)^{\beta} \\ dx = \frac{\beta}{\gamma}\left(\frac{x}{\gamma}\right)^{\beta-1} dy \end{array}\right] = \\ &= \gamma \int_0^{\infty} y^{\frac{1}{\beta}} e^{-y} dy = \gamma\Gamma(1 + \frac{1}{\beta}), \end{aligned} \tag{2.4}$$

kde $\Gamma(1+\frac{1}{\beta})$ je Eulerova Γ funkce, která se zavádí vztahem

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt, \quad x > 0 \tag{2.5}$$

Pro výpočet rozptylu použijme stejnou substituci jako je výše a dostaneme:

$$var(X) = \gamma^2 \Gamma(1+\frac{2}{\beta}) - \gamma^2 \Gamma^2(1+\frac{1}{\beta})$$

Vztahy pro střední hodnotu a rozptyl jsou platné za předpokladu, že $x \geq 0$, $\beta > 0$ a $\gamma > 0$.

Dále budeme definovat kombinovaná rozdělení, která budou založená na Weibullovem rozdělení a rozdělení z transformované rodiny beta rozdělení. K tomu definujeme transformované beta rozdělení a ukažeme, že rozdělení, která patří do této rodiny skutečně mají těžší chvosty, než Weibullovo rozdělení.

2.2 Transormovaná rodina beta

Nejobecnějším zástupcem transformované rodiny beta je transformované beta rozdělení, které má čtyři parametry - $\alpha > 0$, $\tau > 0$, $\theta > 0$, $\xi > 0$. Hustota je definovaná jako

$$f(x) = \frac{\Gamma(\alpha+\tau)}{\Gamma(\alpha)\Gamma(\tau)} \frac{\xi \, (x/\theta)^{\xi\tau}}{x \left[1 + (x/\theta)^\xi\right]^{\alpha+\tau}}, \quad x > 0 \tag{2.6}$$

Chování chvostů. Funkce $S(x) = 1 - F(x) = P(X > x)$, známá jako funkce přežití, vyjadřuje chování chvostu (viz Klugman a kol. (2004), str. 50). Porovnat chování chvostu dvou rozdělení lze limitou podílu jejich funkci přežití při $x \to \infty$. Pokud limita je konstantní nenulová, chvosty se chovají podobně. Pomocí L'Hopitalova pravidla ukážeme, že tento vztah platí i pro podíl hustot:

$$\lim_{x\to\infty} \frac{S_1(x)}{S_2(x)} \stackrel{L'H}{=} \lim_{x\to\infty} \frac{S_1'(x)}{S_2'(x)} = \lim_{x\to\infty} \frac{-f_1(x)}{-f_2(x)}$$

Chování chvostů transformovaného beta rozdělení ovlivňují paramtry ξ a α:

$$\lim_{x\to\infty} \frac{1-F(x)}{x^{-\alpha\xi}} \stackrel{L'H}{=} \lim_{x\to\infty} \frac{-f(x)}{-\alpha\xi x^{-\alpha\xi-1}} = \frac{\Gamma(\alpha+\tau)\theta^{\xi\alpha}}{\Gamma(\alpha)\Gamma(\tau)\alpha} \lim_{x\to\infty} \frac{x^{\xi(\alpha+\tau)}}{[\theta^\xi + x^\xi]^{\alpha+\tau}} = c,$$

kde $c = \frac{\Gamma(\alpha+\tau)\theta^{\xi\alpha}}{\Gamma(\alpha)\Gamma(\tau)\alpha}$ je konstanta. Tak při $x \to \infty$ platí:

$$1 - F(x) \sim c * x^{-\alpha\xi}$$

$$-f(x) \sim -c * \alpha\xi x^{-\alpha\xi-1}$$

Limita podílu funkcí přežití Weibullovo rozdělení a transformavaného beta se pak rovna:

$$\lim_{x\to\infty} \frac{\frac{\beta}{\gamma}\left(\frac{x}{\gamma}\right)^{\beta-1} exp\left\{-\left(\frac{x}{\gamma}\right)^{\beta}\right\}}{c*\alpha\xi x^{-\alpha\xi-1}} =$$

$$= \frac{\beta}{c*\gamma^{\beta}\alpha\xi} \lim_{x\to\infty} exp\left\{-(\frac{x^{\beta}}{\gamma}) + (\beta-1)\ln(x) + (\alpha\xi+1)\ln(x)\right\} = 0.$$

Tím se ukázalo, že transformované beta rozdělení má těžší chvosty, než rozdělení Weibulla.

2.3 Kombinované Weibull-Paretovo rozdělení

Paretovo rozdělení je speciálním případem transformovaného beta rozdělení, kde $\tau = \xi = 1$. V této práci budeme uvažovat takzvánou evropskou verzi Paretova rozdělení, která se využívá pro rozdělení větších škod. Její počatek tedy není v nule, ale v bodě θ, který zároveň bude koncovým bodem kombinovaného rozdělení. Kombinované Weibull-Paretovo rozdělení získáme dosazením do vztahu (1.3) hustot rozděleni Weibullova a Paretova rozdělení:

$$f_X(x) = \begin{cases} \lambda\frac{f_1(x)}{F_1(\theta)}, & \text{je-li } 0 < x \le \theta \\ (1-\lambda)\frac{\alpha\theta^{\alpha}}{x^{\alpha+1}}, & \text{je-li } \theta < x < \infty \end{cases}$$

kde $F_1(x)$ je distribuční funkce a $f_1(x)$ je hustota Weibullova rozdělení z (2.1) a (2.2). Po dosazení do (1.5) dostaváme λ :

$$\lambda = \frac{\frac{\alpha}{\beta}}{\frac{\alpha}{\beta} + \frac{\left(\frac{\theta}{\gamma}\right)^{\beta}}{exp\left\{\left(\frac{\theta}{\gamma}\right)^{\beta}\right\}-1}}.$$

Vezmeme v úvahu druhou vlastnost modelu $f'(\theta-) = f'(\theta+)$, dostaneme se k rovnosti :

$$\left(\frac{\theta}{\gamma}\right)^{\beta} = \frac{\alpha}{\beta} + 1.$$

Tato rovnost snižuje počet neznámých parametrů modelu - např. γ se dá vyjádřit pomocí ostatních parametrů. Navíc se zjednosuší rovnice pro λ:

$$\lambda = \frac{\alpha * exp\left\{\frac{\alpha}{\beta}+1\right\} - \alpha}{\alpha * exp\left\{\frac{\alpha}{\beta}+1\right\} + \beta}.$$

Dále uvedeme další charakteristiku rozdělelení - *k-tý* obecný moment.

Definice 3 (k-tý obecný moment). *Nechť X je náhodná veličina s hustotou rozdělení f(x) a nosičem [0,+∞).*

$$\mu'_k = E(X^k) = \int_0^{\infty} x^k * f(x)dx$$

je k-tým obecným momentem náhodné veličiny X, pokud integrál na pravé straně existuje.

Vypočet *k-tého* obecného momenta Weibullova rozdělení vede k Γ-funkci, ale v našem případě klasická integrální definice (2.5) nebude postačující, potřebujeme integrovat na intervalu $(0,\theta]$. K tomu si uvedeme definici neúplné Γ-funkce.

Definice 4 (Neúplná Γ-funkce)**.** *Rozdělením integrálu* $\int_0^\infty t^{x-1}e^{-t}dt$ *v bodě* $\theta \geq 0$, *získáme dvě neúplné* Γ*-funkce :*

$$\begin{aligned}\gamma(x,\theta) &= \int_0^\theta t^{x-1}e^{-t}dt \\ \Gamma(x,\theta) &= \int_\theta^\infty t^{x-1}e^{-t}dt,\end{aligned} \tag{2.7}$$

kde $\gamma(x,\theta) + \Gamma(x,\theta) = \Gamma(x)$ *pro* $x > 0$.

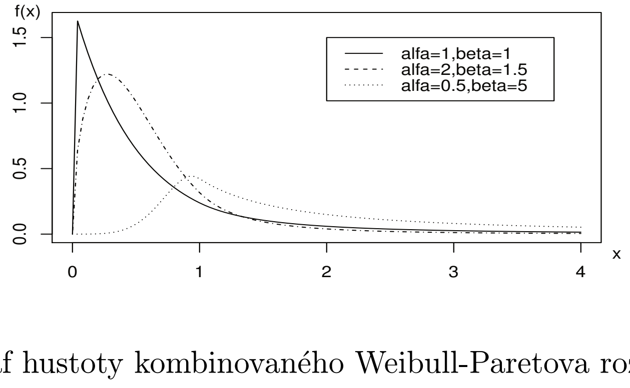
Potom *k-tý* obecný moment nahodné veliciny X, která má kombinované Weibull-Paretovo rozdělení je:

$$\begin{aligned}E(X^k) &= \int_0^\infty x^k f_X(x)dx = \frac{\lambda}{F_1(\theta)}\int_0^\theta f_1(x)dx + (1-\lambda)\int_\theta^\infty f_2(x)dx = \\ &= \frac{\lambda}{F_1(\theta)}\int_0^\theta x^k \frac{\beta}{x}\left(\frac{x}{\gamma}\right)^\beta exp\left\{-\left(\frac{x}{\gamma}\right)^\beta\right\}dx + (1-\lambda)\int_\theta^\infty x^k\frac{\alpha\theta^\alpha}{x^{\alpha+1}}dx \Rightarrow\end{aligned}$$

Pro vypočet prvního integrálu použijme stejnou substituci jako v (2.4) a dostáváme:

$$\begin{aligned}\Rightarrow= \frac{\lambda\gamma^k}{F_1(\theta)}&\int_0^\theta y^{\frac{k}{\beta}}exp^{-y}dy + (1-\lambda)\frac{\alpha\theta^k}{\alpha-k} = \\ &= \frac{\lambda\gamma^k}{F_1(\theta)}\left(\int_0^\infty y^{\frac{k}{\beta}}exp^{-y}dy - \int_\theta^\infty y^{\frac{k}{\beta}}exp^{-y}dy\right) + (1-\lambda)\frac{\alpha\theta^k}{\alpha-k} = \\ &\overset{(2.7)}{=} \frac{\lambda\gamma^k}{1-exp\{-(\theta/\gamma)^\beta\}}\left[\Gamma(1+\frac{k}{\beta}) - \Gamma(1+\frac{k}{\beta},\theta)\right] + (1-\lambda)\frac{\alpha\theta^k}{\alpha-k}\end{aligned} \tag{2.8}$$

pro $\alpha > k$. Pomocí *k-tého* obecného momenta se dají spočítat i další charakteristiky rozdělení - střední hodnota je první obecný moment, rozptyl lze spočitat ze vztahu (2.3).



Obrázek 2.1: Graf hustoty kombinovaného Weibull-Paretova rozělení pro $\theta = 1$

2.4 Kombinované Weibull-Pareto II rozdělení

Paretovo rozdělení II typu je spojité pravděpodobnostní rozdělení s distribuční funkcí

$$G(x) = 1 - \left(1 + \frac{x-\theta}{\alpha\delta}\right)^{-\alpha}, \quad x > \theta \tag{2.9}$$

kde $\theta > 0$ - parametr polohy, $\delta > 0$ - parametr měřítka, $\alpha > 0$ - parametr tvaru. Odvozuje se za pomocí Paretova rozdělení.

Nechť Y je náhodná veličina, která má transformované beta rozdělení, kde $\tau = \xi = 1$ a $\theta = \alpha$, tj. $Y \sim Pa(\alpha,\alpha)$. Potom hustota rozdělení má tvar:

$$f_Y(y) = \frac{\alpha^{\alpha+1}}{(\alpha+y)^{\alpha+1}} = \left(1 + \frac{y}{\alpha}\right)^{-(\alpha+1)}, \quad y \geq 0$$

Distribuční funkce Y je rovná

$$F_Y(y) = \int_0^y f_Y(z)dz = \int_0^y \frac{\alpha^{\alpha+1}}{(\alpha+z)^{\alpha+1}}dz = \frac{\alpha^{\alpha+1}}{-(\alpha+1)+1}\left[(\alpha+z)^{-(\alpha+1)+1}\right]_0^y$$

tedy

$$F_Y(y) = 1 - \left(1 + \frac{y}{\alpha}\right)^{-\alpha}$$

Definujeme náhodnou veličinu X vztahem

$$X = \theta + \delta Y$$

kde θ, $\delta > 0$. Pak platí:

$$P[X > x] = P[\theta + \delta Y > x] = 1 - P[Y \leq \frac{x-\theta}{\delta}] = \left(1 + \frac{x-\theta}{\alpha\delta}\right)^{-\alpha}$$

Potom X má Paretovo rozdělení typu II s parametry $\alpha > 0$, $\theta > 0$, $\delta > 0$ s distribuční funkcí, která je daná vztahem (2.9). Použitím vztahu $G'(x) = g(x)$ dostaváme, že hustota rozdělení je rovna

$$g(x) = \frac{\alpha(\alpha\delta)^\alpha}{(\alpha\delta + x - \theta)^{\alpha+1}}, \quad x > \theta$$

Konečný tvar hustoty, který lze dohledat ve světové literatuře, získáme substitucí $\eta = \alpha\delta - \theta$:

$$g(x) = \frac{\alpha(\eta+\theta)^\alpha}{(\eta+x)^{\alpha+1}}, \quad x > \theta \tag{2.10}$$

kde $\alpha > 0$, $\theta > 0$, $\eta > -\theta$.
Hustotu kombinovaného Weibull-Pareto II rozdělení definujeme vztahem:

$$f_X(x) = \begin{cases} \lambda\frac{f_1(x)}{F_1(\theta))}, & \text{je-li } 0 < x \leq \theta \\ (1-\lambda)g(x), & \text{je-li } \theta < x < \infty \end{cases} \tag{2.11}$$

s parametry $\alpha > 0$, $\beta > 0$, $\theta > 0$, $\eta > -\theta$, $\gamma > 0$ a $0 \leq \lambda \leq 1$, kde $f_1(x)$ a $F_1(x)$ jsou hustota a distribuční funkce Weibullova rozdělení a $g(x)$ je hustota Paretova

rozdělení typu II, která je daná vztahem (2.10).
Po dosazení do (1.5) dostaváme λ:

$$\lambda = \frac{\frac{\alpha}{\beta}}{\frac{\alpha}{\beta} + \frac{\eta+\theta}{\theta} \frac{\left(\frac{\theta}{\gamma}\right)^\beta}{exp\left\{\left(\frac{\theta}{\gamma}\right)^\beta\right\}-1}}.$$

Z podmínky differencovatelnosti plyne:

$$\left(\frac{\theta}{\gamma}\right)^\beta = \frac{\alpha\theta - \eta}{\beta(\eta+\theta)} + 1$$

Touto rovností se počet neznámych parametrů snižil, tedy kombinované rozdělení Weibull-Pareto II má čtyři neznáme parametry.
Pro k-tý obecný moment platí:

$$E(X^k) = \int_0^\infty x^k f_x(x)dx = \frac{\lambda}{F_1(\theta)} \int_0^\theta f_1(x)dx + (1-\lambda)\int_\theta^\infty f_2(x)dx = (*)$$

První integrál je již vypočten v (2.8), s využitím vztahu $(a+b)^k = \sum_{n=1}^k \binom{k}{n} a^{k-n}b^n$ spočteme druhý integrál:

$$\int_\theta^\infty f_2(x)dx = \int_\theta^\infty \alpha x^k \frac{(\gamma+\theta)^\alpha}{(\gamma+x)^{\alpha+1}}dx \overset{s}{=} \begin{bmatrix} y = \frac{\gamma+\theta}{\gamma+x} \\ dx = \frac{-(\gamma+x)^2}{\gamma+\theta}dy \end{bmatrix} =$$

$$= \int_0^1 \frac{\alpha((\gamma+\theta)-\gamma y)^k}{y^{k-\alpha+1}}dy = \int_0^1 \frac{\alpha \sum_{n=1}^k \binom{k}{n}(\gamma+\theta)^n(-\gamma)^{k-n}y^{k-n}}{y^{k-\alpha+1}}dy =$$

$$= \alpha \sum_{n=0}^k \binom{k}{n} (\gamma+\theta)^n(-\gamma)^{k-n} \int_0^1 y^{\alpha-k-1}dy = \alpha \sum_{n=0}^k \frac{\binom{k}{n}(\gamma+\theta)^n(-\gamma)^{k-n}}{\alpha-n}$$

Tedy k-tý obecný moment náhodné veličiny X, která má Weibull-Paretovo typu II rozdělení je roven:

$$(*) = \frac{\lambda\gamma^k\left[\Gamma(1+\frac{k}{\beta}) - \Gamma(1+\frac{k}{\beta},\theta)\right]}{1-exp\left\{-(\theta/\gamma)^\beta\right\}} + (1-\lambda)\alpha\sum_{n=0}^k \frac{\binom{k}{n}(\gamma+\theta)^n(-\gamma)^{k-n}}{\alpha-n},$$

pro $\alpha > k$.

2.5 Kombinované Weibull-Burrovo rozdělení

Burrovo rozdělení, známé také jako Singh-Maddalovo rozdělení nebo rozdělení Pareto typu IV, je speciálním případem transormovaného beta rozdělení, kde $\tau = 1$. Hustota Burrova rozdělení je daná vztahem:

$$f(x) = \frac{\alpha\xi\left(\frac{x}{\theta}\right)^\xi}{x\left[1+\left(\frac{x}{\theta}\right)^\xi\right]^{\alpha+1}}, \quad x > 0 \tag{2.12}$$

Použitím vztahu $F(x) = \int_0^x f(t)dt$ dostáváme, že distribuční funkce Burrova rozdělení se rovná

$$F(x) = 1 - \left(1 + \left(\frac{x}{\theta}\right)^{\xi}\right)^{-\alpha}, \quad x > 0 \tag{2.13}$$

Na rozdíl od rozdělení Pareto a Pareto typu II, náhodná veličina X, která má Burrovo rozdělení, může nabývat hodnot od nuly do nekonečna, tedy není omezena bodem θ, který jsme považovali za koncový bod v předchozích kombinovaných rozděleních. Proto si musíme zvolit dělící bod odlišný od θ, který bude parametrem modelu. Kombinované Weibull-Burrovo rozdělení pak definujeme takto:

$$f_X(x) = \begin{cases} \lambda\frac{f_1(x)}{F_1(\sigma)}, & \text{je-li } 0 < x \leq \sigma \\ (1-\lambda)\frac{f_2(x)}{1-F_2(\sigma)}, & \text{je-li } \sigma < x < \infty \end{cases} \tag{2.14}$$

kde $f_1(x)$ a $F_1(x)$ je hustota a distribuční funkce Weibullova rozdělení, s neznámým parametry $\beta > 0$, $\gamma > 0$, $f_2(x)$ a $F_2(x)$ je hustota a distribuční funkce Burrova rozdělení definované vztahem (2.12) a (2.13), s neznámými parametry $\alpha > 0$, $\theta > 0$, $\xi > 0$.
Z rovnosti (1.5) dostaneme λ:

$$\lambda = \frac{\alpha\xi\sigma^{\xi}e^{\left(\frac{\sigma}{\gamma}\right)^{\beta}} + \sigma\theta^{\xi}\left(1 + \left(\frac{\sigma}{\theta}\right)^{\xi}\right)}{\alpha\xi\sigma^{\xi}e^{\left(\frac{\sigma}{\gamma}\right)^{\beta}} + \beta\theta^{\xi}\left(\frac{\sigma}{\gamma}\right)^{\beta}\left(1 + \left(\frac{\sigma}{\theta}\right)^{\xi}\right)}$$

Koncový bod σ získáme z rovnice, kterou dává podmínka diferencovatelnosti. V tomoto případě σ nelze explicitně vyjádřit, ale lze použít k výpočtu numerické metody. Potom Weibull-Burrovo rozdělení má pět parametrů.

Výpočet k-tého obecného momentu Burrova rozdělení vede k hypergeometrické funkci; uvedeme si jeji definici.

Definice 5 (Hypergeometrická funkce)**.** *Hypergeometrická funkce* ${}_2F_1(a,b,c;x)$ *definovaná řadou*

$$\sum_{n=0}^{\infty} \frac{(a)_n(b)_n}{(c)_n n!} z^n$$

kde a,b,c jsou libovolné reálné nebo komplexní parametry, mimo $c = 0, -1, -2..$, *z - komplexní proměnná a* $(x)_n = x(x+1)..(x+n-1)$.

Věta 1. *Pokud* $Re(c) > 0$, $Re(b) > 0$ *a* $|x| < 1$ *platí:*

$${}_2F_1(a,b,c;x) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 t^{b-1}(1-t)^{c-b-1}(1-xt)^{-a}dt \tag{2.15}$$

Důkaz. Viz [Andrews a kol. (1999), Kapitola 2.2].

❑

Potom k-tý obecný moment náhodné veličiny, která má Weibull-Burrovo rozdělení je rovný:

$$E(X^k) = \frac{\lambda}{F_1(\sigma)} \int_0^{\sigma} x^k f_1(x)dx + \frac{(1-\lambda)}{F_2(\sigma)} \int_{\sigma}^{\infty} x^k f_2(x)dx$$

První integrál byl spočitan v (2.8), spočteme druhý integrál:

$$\int_{\sigma}^{\infty} x^k f_2(x)dx = \int_{\sigma}^{\infty} x^k \frac{\alpha\xi\left(\frac{x}{\theta}\right)^{\xi}}{x\left[1+\left(\frac{x}{\theta}\right)^{\xi}\right]^{\alpha+1}} \stackrel{s}{=} \left[\begin{array}{l} y = \left(\frac{x}{\theta}\right)^{\xi} \\ dx = \frac{\theta}{\xi}\left(\frac{x}{\theta}\right)^{1-\xi} dy \end{array}\right] =$$

$$= \alpha\theta^k \int_{\left(\frac{\sigma}{\theta}\right)^{\xi}}^{\infty} y^{\frac{k}{\xi}} \left[1+y\right]^{-(\alpha+1)} dy \Rightarrow$$

Označme $\left(\frac{\sigma}{\theta}\right)^{\xi} = a$:

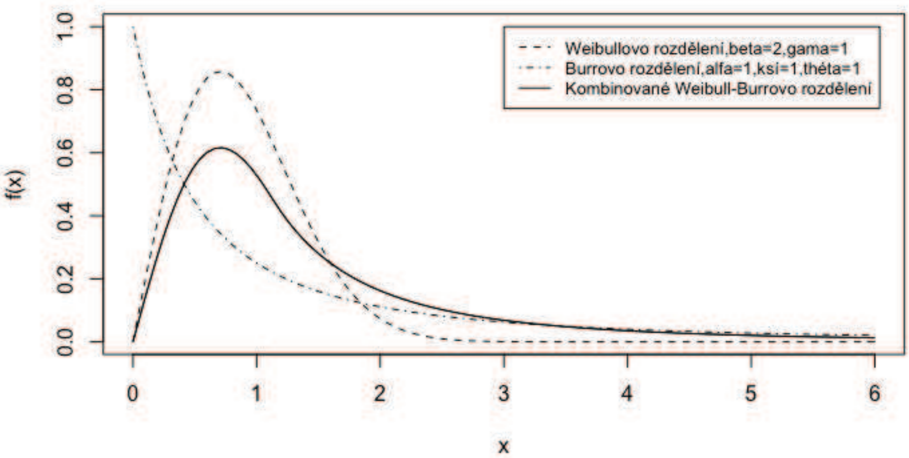
$$\Rightarrow \stackrel{s}{=} \left[\begin{array}{l} z = \frac{a}{y} \\ dy = -\frac{y^2}{a}dy \end{array}\right] = \alpha\theta^k \int_0^1 a^{\frac{k}{\xi}-\alpha} z^{\alpha-\frac{k}{\xi}-1} \left[1+z\frac{1}{a}\right]^{-(\alpha+1)} dz \Rightarrow$$

Tento integrál je speciálním případem integrálu z věty 1, kde $c = b+1$ a $x = -\frac{1}{a}$ a pokud platí $|\left(\frac{\theta}{\sigma}\right)^{\xi}| < 1$, $\alpha - \frac{k}{\xi} > 0$. Dosadíme $a = \left(\frac{\sigma}{\theta}\right)^{\xi}$ a dostáváme:

$$\Rightarrow = \frac{\alpha\theta^k \left[\left(\frac{\theta}{\sigma}\right)^{\xi}\right]^{\alpha-\frac{k}{\xi}}}{\alpha - \frac{k}{\xi}} {}_2F_1(\alpha+1, \alpha-\frac{k}{\xi}, \alpha-\frac{k}{\xi}+1, -\left(\frac{\theta}{\sigma}\right)^{\xi}) \stackrel{ozn.}{=} J_k(\sigma)$$

Tedy k-tý centrální moment nahodné veličiny X, která má kombinované Weibull-Burrovo rozdělení je roven:

$$E(X^k) = \frac{\lambda}{1-exp\left\{-(\sigma/\gamma)^{\beta}\right\}}\left[\Gamma(1+\frac{k}{\beta}) - \Gamma(1+\frac{k}{\beta},\sigma)\right] + \frac{1-\lambda}{(1+(x/\theta)^{\xi})^{-\alpha}} J_k(\sigma).$$



Obrázek 2.2: Graf hustoty Weibullova rozdělení, Burrova rozdělení a kombinovaného Weibull-Burrova rozdělení s paramtry $\alpha = 1$, $\theta = 1$, $\xi = 1$, $\beta = 2$, $\gamma = 1$

Kapitola 3

Metoda maximální věrohodnosti a metody srovnání vhodnosti modelů

3.1 Metoda maximální věrohodnosti

V této části bude popsán algoritmus odhadu parametrů kombinovaných rozdělení. Jednou z nejpoužívanějších metod ve statistice pro hledaní odhadu parametra (resp. parametrů) je metoda maximální věrohodnosti. Princip této metody je založen na maximalizaci věrohodnostní funkce. Dále uvedeme základní pojmy a principy metody, a uvedeme postup na příkladě Weibull-Paretova rozdělení.

Nechť $\boldsymbol{X} = (X_1, \ldots, X_n)$ je náhodný vektor, jehož složky jsou nezávislé, se sdruženou hustotou $f(x_1, \ldots, x_n|\boldsymbol{\theta})$, kde $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$ je vektor neznámých parametrů z parametrického prostoru $\Theta \in \mathbb{R}^m$. Předpokladáme, že $f(x_1, \ldots, x_n|\boldsymbol{\theta_1}) \neq f(x_1, \ldots, x_n|\boldsymbol{\theta_2})$, pro kterákoliv $\boldsymbol{\theta_1} \neq \boldsymbol{\theta_2}$.

Věrohodnostní funkce. Sdružená hustota rozdělení náhodného vektoru $\boldsymbol{X}$, s tím, že jeho složky jsou nezávislé a stejně rozdělené, je rovna $f(x_1, \ldots, x_n|\boldsymbol{\theta}) = f(x_1|\boldsymbol{\theta}) \cdot \ldots \cdot f(x_n|\boldsymbol{\theta})$. Definujeme dále funkce věrohodností jako sdruženou hustotu $f(x_1, \ldots, x_n|\boldsymbol{\theta})$ závislou na parametru $\boldsymbol{\theta}$, spočitanou v pozorovaných hodnotách $X_1, \ldots, X_n$:

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

Takto definovaná funkce $L_n(\boldsymbol{\theta})$ určuje pravděpodobnost toho, zda pozorování $X_1, \ldots, X_n$ pochází z předpokladaného rozdělení. Hlavní myšlenkou tedy najit takovou hodnotu parametru $\boldsymbol{\theta}$, aby hodnota věrohodnostní funkce byla maximální.

Maximálně věrohodný odhad parametru $\boldsymbol{\theta}$ je maximum věrohodnostní funkce:

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

Maximum funkce najdeme řešením soustavy rovnic $\frac{\partial L_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0$. V praxi se nicméně často používá logaritmus funkce věrohodnosti, vlastnosti logaritmu usnadňují výpočet.

Poznámka. V důsledku toho, že je logaritmus monotónní, rostoucí na celém definičním oboru funkce, věrohodnostní funkce $L_n(\boldsymbol{\theta})$ a její logaritmus $\log L_n(\boldsymbol{\theta})$ nabývají maxima ve stejném bodě.

Tedy v bodech nenulové hustoty $f(X_1, \ldots, X_n|\theta)$ definujeme funkci

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

kterou nazýváme **logaritmická věrohodnost** a hledáme věrohodnostní odhad jako řešení sytému logaritmických věrohodnostních rovnic:

$$\frac{\partial \sum_{i=1}^{n} \ln f(X_i|\boldsymbol{\theta})}{\partial \theta_j} = 0, \quad j = 1, \ldots, m. \tag{3.1}$$

Odhady získané metodou maximalní věrohodností nejsou vždy nestranné a konzistentní. Nicméně při určitém počtu pozorování je možné dokázat, že tato metoda poskytuje odhady, které jsou asymptoticky normální.

Algoritmus odhadu parametrů Weibull-Paretova rozdělení. Nechť $X_1, \ldots, X_n$ je náhodný vyběr z kombinovaného Weibull-Paretova rozděleni. Bez újmy na obecnosti, předpokladáme, že $x_1 \leq x_2 \leq \cdots \leq x_n$. Neznámé parametry Weibull-Paretova rozdělení jsou $\alpha > 0$, $\beta > 0$, $\gamma > 0$, $\theta > 0$, ale jak jsme ukázali v (2.3), jeden z parametrů je funkcí zbyvajících parametrů. Vyjádříme γ:

$$\gamma = \sqrt[\beta]{\frac{\beta \theta^{\beta}}{\alpha + \beta}}$$

a hledáme odhad parametru α, β a θ.

Předpokladejme, že neznámý parametr θ leží mezi m-tým a $(m+1)$-ým pozorováním náhodného výběru nebo $x_m \leq \theta \leq x_{m+1}$. Potom logaritmická věrohodnost vypadá následovně:

$$\begin{aligned} l_n(\boldsymbol{\theta}) &= \lambda \sum_{i=1}^{m} \ln(f_1^*(x_i)) + (1-\lambda) \sum_{i=m+1}^{n} \ln(f_2^*(x_i)) = \\ &= n\ln(\alpha) + n\ln(\alpha+\beta) - n\ln(\alpha\, e^{\frac{\alpha}{\beta}+1} + \beta) + (\alpha m - \beta M)\ln(\theta) + \\ &+ (\frac{\alpha}{\beta}+1)\left(M - \sum_{x_i \leq \theta} \left(\frac{x_i}{\theta}\right)^{\beta}\right) + (\beta - 1)\sum_{x_i \leq \theta} \ln(x_i) - (\alpha+1)\sum_{x_i \geq \theta} \ln(x_i), \end{aligned}$$

kde $m = \sum_{i=1}^{n} I\{x_i \leq \theta\}$ a $M = \sum_{i=1}^{n} I\{x_i \geq \theta\}$. Tedy je vidět, že odhady parametrů jsou závislé na m, jelikož se hodnota logaritmické věrohodnosti mění v závislostí na m. Nicméně hodnota m není známa, proto dále budeme postupovat podle nasledujícího algoritmu (Cooray a Ananda (2005)):

Krok 1: Pro každé $m = 1, \ldots, n-1$ spočteme maximalní věrohodný odhad parametrů kombinovaného rozdělení $\widehat{\alpha}$, $\widehat{\beta}$, $\widehat{\theta}$ jako řešení soustavy logaritmických věrohodnostních rovnic:

$$\begin{cases} \frac{\partial \ln(\boldsymbol{\theta})}{\partial \alpha} = 0 \quad \frac{\partial \ln(\boldsymbol{\theta})}{\partial \beta} = 0 \quad \frac{\partial \ln(\boldsymbol{\theta})}{\partial \theta} = 0 \end{cases}$$

když $x_m \leq \widehat{\theta} \leq x_{m+1}$, pak maximální věrohodné odhady jsou rovné

$$\widehat{\alpha}^{MLE} = \widehat{\alpha}, \quad \widehat{\beta}^{MLE} = \widehat{\beta}, \quad \widehat{\theta}^{MLE} = \widehat{\theta}$$

a algoritmus se zastaví.

Krok 2: Pokud jsme nenašli řešení pro θ v prvním kroku, nastávají dvě možnosti: buďto $m = n$ a věrohodnostní funkci $L_n(\boldsymbol{\theta})$ vypočteme jako součin pouze hustot f_1, anebo $m = 0$ a $L_n(\boldsymbol{\theta})$ bude součinem pouze hustot f_2.

V případě kombinovaného Weibull-Paretova rozdělení odhady parametrů nelze analyticky vyjádřit, systém logaritmických věrohodnostních rovnic je třeba řešit numericky.

3.2 Hodnocení vhodnosti modelů

Existuje řada postupů jak posoudit vhodnost modelu. Cooray a Ananda (2005) uvádí jako jeden ze základních kritérií **hodnotu funkci logaritmické věrohodnosti** po dosazení odhadnutých parametrů. Jak jsme již výše uvedli, čím větší je hodnota věrohodnostní funkce, tím je větší i pravděpodobost, že data pocházejí z předpokladaného rozdělení. Tento postup má však smysl pouze pokud rozhodujeme mezi modely se stejným počtem prarametrů. V případě, kdy je potřeba porovnávát modely s různým počtem parametrů, je vhodné použit kritéria, která penalizují počet parametrů modelu. Uvedeme nyní některá z těchto kritérií.

Akaikeho informační kritérium (AIC) bylo představeno Hirotugu Akaikem v roce 1974. Toto kritérium je dáno vzorcem (Abu Bakar a kol. (2015)):

$$AIC = 2k - 2l_n(\boldsymbol{\theta}),$$

kde k je počet parametrů a $l_n(\boldsymbol{\theta})$ je hodnota logarotmické věrohodnostní funkce. Dostáváme výsledek, který je kompromisem mezi složitostí modelu a jeho přesností. Porovnáváme-li několik parametrických modelů, budeme preferovat model, který má nejmenší hodnotu AIC.

Bayesovo informační kritérium (BIC) bylo představeno Schwarzem v roce 1978, proto se někdy označuje jako Schwarzovo kritérium (SBC). Kritérium je dáno vzorcem (Abu Bakar a kol. (2015)):

$$BIC = k\ln(n) - 2l_n(\boldsymbol{\theta}),$$

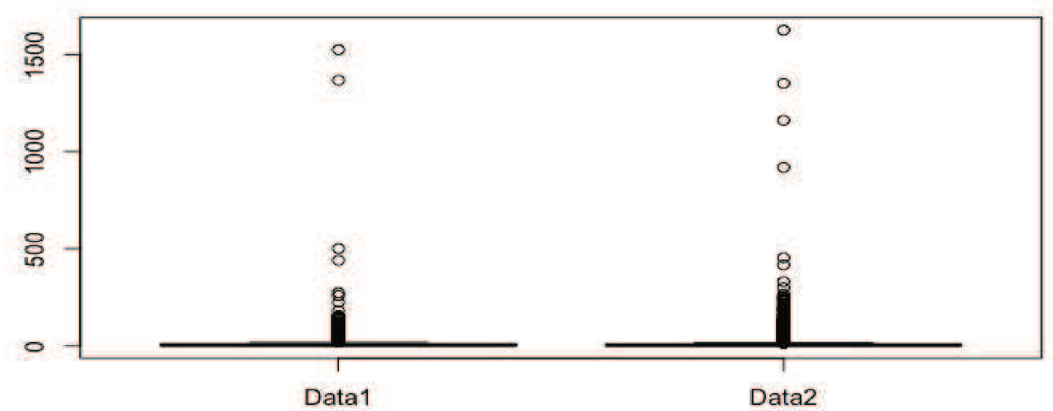
kde n je počet pozorování. Jak je vidět, penalizace parametrů u BIC je větší, než u AIC.

Kapitola 4

Modelování nad reálnými daty

V praktické části této práce budeme provadět modelování nad dvěma reálnými soubory dat. První soubor obsahuje 9613 záznamů o výších škod vzniklých v důsledku požáru od roku 1982 do 1996 na území Francie (dále jen Data1). Výše škod se pohybuje v rozmezí od 1.034747 do 1624.743 (v milionech francouzských franků, FRF). Střední hodnota se rovná 9.039918, dolní kvartil je roven 1.908664 a horní 7.815564. Datový soubor lze nalezt v balíčku "CASdatasets"softwarového prostředí R pod nazvem "frecomfire". Druhý soubor obsahuje 2387 záznamů o škodach způsobených přerušením podnikatelské činnosti, stejně jako první soubor na území Francie (dále jen Data2), v rozmezí od 1.00289 do 1524.49017 (ve ste tisicích francouzských franků, FRF). Střední hodnota souboru dat se rovná 10.96223, dolní kvartil je roven 1.99262 a horní 8.90563. V balíčku R "CASdatasets"je datový soubor pod nazvem "frebiloss".

Výpočtene hodnoty kvartilů ukazují, že oba soubory dat silně zešikmené. Dolní a horní kvartily jsou velmi vzdálené od maximálních hodnot. Znázorníme data graficky pomocí krabicových diagramů:



Obrázek 4.1: Krabicové diagramy souborů zkoumaných dat

Vidíme, že oba soubory vzkazují stejné vlastností - velký počet malých škod a malý počet velkých. Také lze pozorovat, že jsou oba soubory silně kladně zešikmené.

Pro výpočet odhadu parametrů a nalezení hodnot AIC a BIC využijme již výše zmíněný software R. Jak jsme již ukázali v minulé kapitole, algoritmus maximální věrohodnosti vede k řešení soustavy nelineárních rovnic. Pro řešení této soustavy použijme funkci "nlm", která hledá řešení pomocí Newtonovy metody. Výsledky odhadů parametrů zkoumaných rozdělení pro oba soubory dat v tabulce 4.1 a 4.2.

Rozdělení	**Odhad parametrů**	**NLL***	**AIC**	**BIC**
Weibullovo	$\widehat{\beta} = 0{,}7935222$ $\widehat{\gamma} = 7{,}3953103$	29991,76	59987,51	60001,85
Paretovo	$\widehat{\alpha} = 0{,}7033735$ $\widehat{\theta} = 1{,}0347472$	26991,28	53986,55	54000,9
Pareto II	$\widehat{\theta} = 1{,}0347449$ $\widehat{\alpha} = 1{,}056684$ $\widehat{\eta} = 0{,}008856295$	27880,45	55766,9	55788,41
Burrovo	$\widehat{\alpha} = 0{,}06760312$ $\widehat{\xi} = 11{,}90192226$ $\widehat{\theta} = 0{,}80140886$	26650,89	53307,78	53329,29
CWP	$\widehat{\beta} = 14{,}9530580$ $\widehat{\alpha} = 0{,}7664223$ $\widehat{\theta} = 1{,}305622$	26712,63	53431,26	53452,77
CWPII	$\widehat{\beta} = 39{,}9892138$ $\widehat{\alpha} = 1{,}378473$ $\widehat{\eta} = 2{,}248496$ $\widehat{\theta} = 1{,}126465$	26248,69	52505,38	52534,06
CWB	$\widehat{\beta} = 38{,}3504103$ $\widehat{\gamma} = 1{,}1328153$ $\widehat{\alpha} = 2{,}5311051$ $\widehat{\xi} = 0{,}7132602$ $\widehat{\theta} = 5{,}5653363$	26240,52	52491,04	52526,89

Pozn: *NLL - negative log-likelihood

Tabulka 4.1: Maximálně věrohodné odhady pro Data1

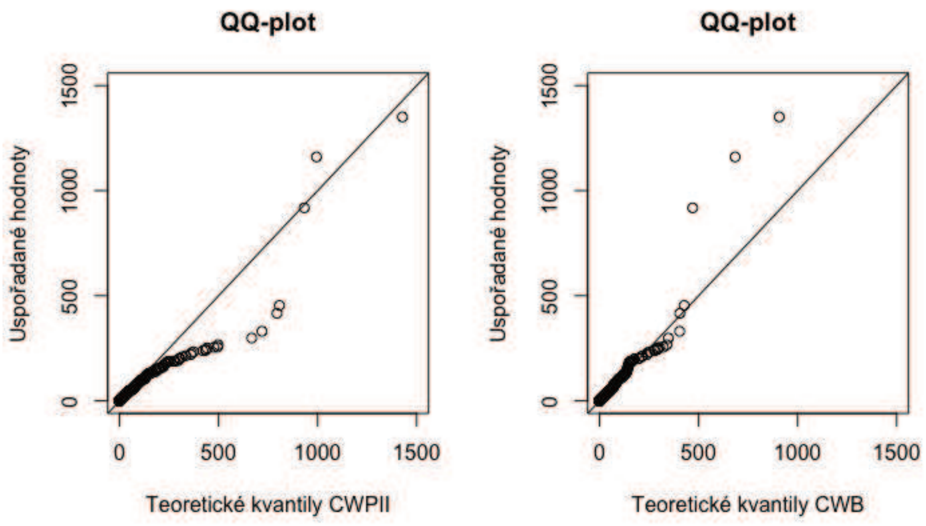
Jak je vidět z tabulky 4.1 kombinované Weibull-Burrovo rozdělení má nejmenší hodnotu logarotmické věrohodnostní funkce vynásobenou -1, tedy pravděpodobnost toho, že data pochazejí z tohoto rozdělení je největší a to v obou případech. Zároveň má Weibull-Burrovo rozdělení nejmenší hodnoty AIC a BIC, i když má největší počet parametrů. Druhý nejlepší výsledek dává kombinované Weibull-Paretovo II rozdělení jak pro soubor Data1, tak i pro soubor Data2. Nejhorším kandidátem se pro soubor Data1 stalo Weibullovo rozdělení, pro soubor Data2 nejhůře dopadlo rozdělení Pareto typu II.

Rozdělení	**Odhad parametrů**	**NLL***	**AIC**	**BIC**
Weibullovo	$\widehat{\beta} = 0{,}7473416$ $\widehat{\gamma} = 8{,}2265517$	7766,155	15536,31	15547,87
Paretovo	$\widehat{\alpha} = 0{,}6537811$ $\widehat{\theta} = 1{,}0028900$	7058,573	14121,15	14132,7
Pareto II	$\widehat{\theta} = 1{,}0028872$ $\widehat{\alpha} = 1{,}662072$ $\widehat{\eta} = 0{,}4988266$	7840,204	15686,41	15703,74
Burrovo	$\widehat{\alpha} = 0{,}09729753$ $\widehat{\xi} = 7{,}96648083$ $\widehat{\theta} = 0{,}77114025$	6990,397	13986,79	14004,13
CWP	$\widehat{\beta} = 7{,}1237442$ $\widehat{\alpha} = 0{,}7528249$ $\widehat{\theta} = 1{,}56524$	6999,147	14004,29	14021,63
CWPII	$\widehat{\beta} = 23{,}1924097$ $\widehat{\alpha} = 1{,}378473$ $\widehat{\eta} = 3{,}376381$ $\widehat{\theta} = 1{,}079977$	6880,982	13769,96	13793,08
CWB	$\widehat{\beta} = 182{,}7261326$ $\widehat{\gamma} = 1{,}0116970$ $\widehat{\alpha} = 1{,}6274685$ $\widehat{\xi} = 0{,}9102909$ $\widehat{\theta} = 3{,}9901364$	6868,439	13746,88	13775,77

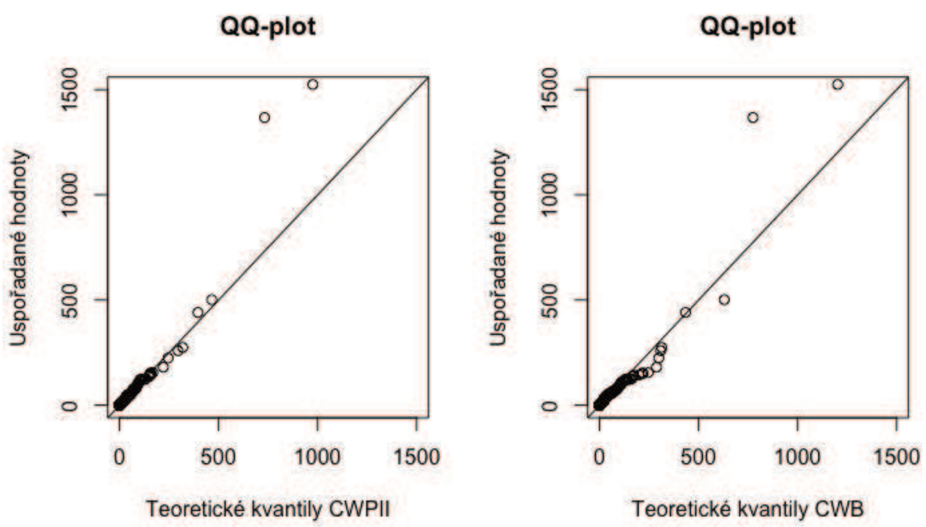
Pozn: *NLL - negative log-likelihood

Tabulka 4.2: Maximálně věrohodné odhady pro Data2

Dosažené výsledky je vhodné znázornit také graficky. Jedna z nejpoužívanějších grafickych metod je kvantil-kvantilový graf (Q-Q plot). Tato technika je založená na porovnání teoretických kvantilů předpokládaného rozdělení s empirickými kvantily uspořádaného souboru zkoumaných dat. Pochází-li data z uvažovaného rozdělení, měly by body kvantil-kvantilového grafu ležet přibližně na přímce. Obrázek 4.2 ukazuje, že kombinované Weibull-Burrovo rozdělení v porovnání s kombinovaným Weibull-Pareto II dává lepší výsledky - body na grafu jsou skoro na přímce. Na obrázku 4.3 je vidět, že kombinované Weibull-Pareto II rozdělení lepé vystihuje charakter dat - body na grafu jsou blížší k přímce, nicméně numericky jsme dostali lepší výsledek pro kombinované Weibull-Burrovo rozdělení.

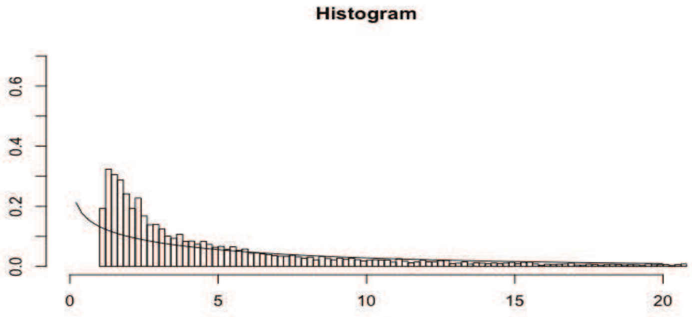


Obrázek 4.2: Kvantil-kvantilový graf (Data1)

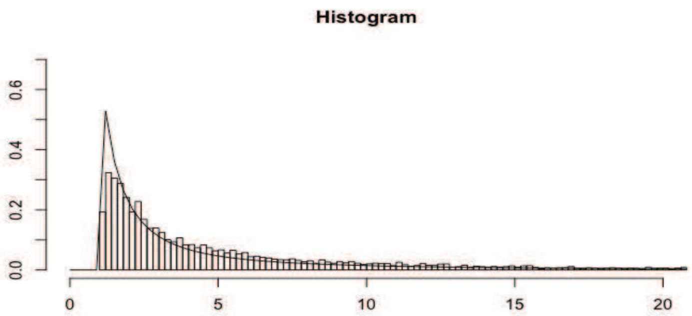


Obrázek 4.3: Kvantil-kvantilový graf (Data2)

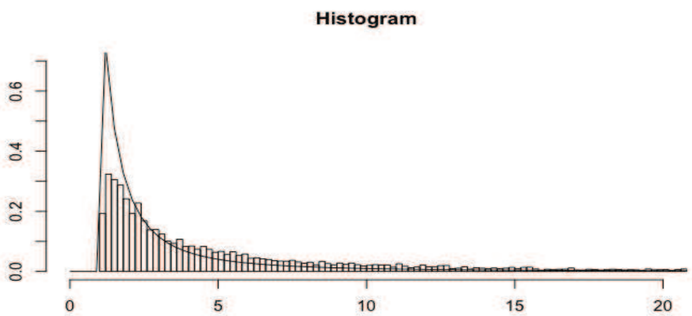
Týto výledky lze pozorovat i z histogramu. Histogram je slpupcový graf, který znazorňuje distribuce dat. Níže uvedeme srovnání histogramu s hustotami základních rozdělení uvažovaých v práci a hustotami kombinovaných rozdělení pro soubor Data1.



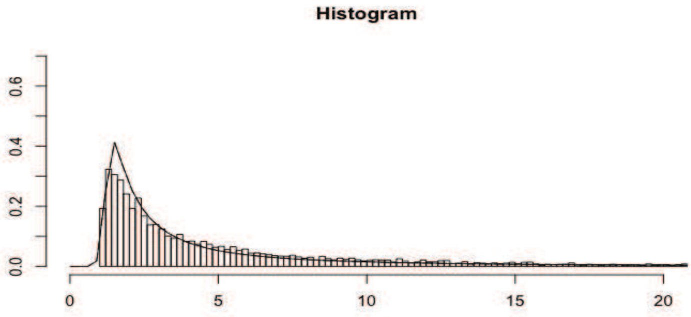
Obrázek 4.4: Histogram souboru Data1 a hustota Weibullova rozdělení s odhadnutými parametry



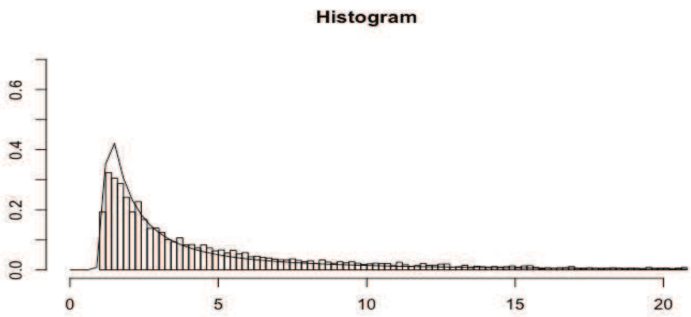
Obrázek 4.5: Histogram souboru Data1 a hustota Paretova rozdělení s odhadnutými parametry



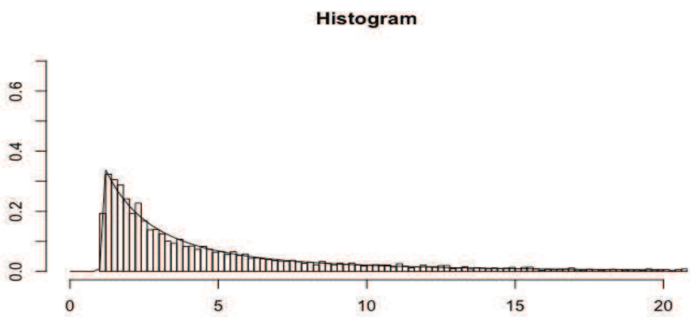
Obrázek 4.6: Histogram souboru Data1 a hustota Paretova rozdělení typu II s odhadnutými parametry



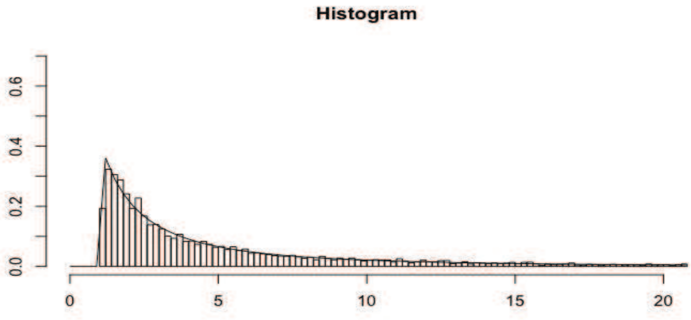
Obrázek 4.7: Histogram souboru Data1 a hustota Burrova rozdělení s odhadnutými parametry



Obrázek 4.8: Histogram souboru Data1 a hustota kombinovaného Weibull-Paretova rozdělení s odhadnutými parametry



Obrázek 4.9: Histogram souboru Data1 a hustota kombinovaného Weibull-Paretova typu II rozdělení s odhadnutými parametry



Obrázek 4.10: Histogram souboru Data1 a hustota kombinovaného Weibull-Burrova rozdělení s odhadnutými parametry

Závěr

Pro modelování výše škod někdy potřebujeme modely, která jsou složitější než základní spojitá rozdělení. V této práci jsme se seznamili s některými kombinovanými rozdělení, která vznikly kombinací dvou základních rozdělení a uvedli odvození jejich základnćh charakteristik. Dále jsme popsali algoritmus odhadu parametrů kombinovaných rozdělení, který v praktické čaśti aplikovali na dva soubory reálných dat.

Složitější modely, jako jsou kombinovaná rozdělení, mají i své nevýhody. Odhad parametrů metodou maximalní věrohodností je početně náročný. Jak jsme ukázali na příkladu kombinovaného Weibull-Paretova rozdělení, pro výpočet odhadu parametrů musíme vyřešit soustavu věrohodnostních rovnic, kterou se nedá vyřešit analytickým přistupem. K hledání odhadů parametrů tedy používáme numerické metody, kde jako vstup musíme zadat počáteční odhad parametrů, které ovlivňují výsledné odhady.

Kombinovaná rozdělení obtížněji použitelné na reální data, než modely základaní. Nicméně v praktické části jsme došli k závěru, že lepé vystihují charakter dat a mohou sloužit užitečným nástrojem pro modelování výše škod. Týto výsledky jsme znazornili i graficky.

Literatura

ABU BAKAR, S. S. A., HAMZAH, N., MAGHSOUDI, M. a NADARAJAH, S. (2015). Modeling loss data using composite models. *Insurance: Mathematics and Economics*, **61**, 146–154.

ANDREWS, G., ASKEY, R. a ROY, R. (1999). *Special Functions*. Cambridge University Press, Cambridge.

COORAY, K. a ANANDA, M. M. A. (2005). Modeling actuarial data with a composite lognormal-pareto model. *JScandinavian Actuarial Journal*, **5**, 321–334.

KLUGMAN, S. A., PANJER, H. H. a WILLMOT, G. E. (2004). *Loss Models: From Data to Decisions*. Second Edition. John Wiley & Sons, Inc., Hoboken. ISBN ISBN 0-471-21577-5.

SCOLLNIK, D. a SUN, C. (2012). Modeling withweibull-pareto models. *North American Actuarial Journal*, **16**(2), 260–272.

Seznam obrázků

Seznam tabulek