



**MATEMATICKO-FYZIKÁLNÍ
FAKULTA**
Univerzita Karlova

DIPLOMOVÁ PRÁCE

Petra Čápová

**Modely úhrnů škod se závislou frekvencí
a severitou**

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: RNDr. Lucie Mazurová, Ph.D.

Studijní program: Matematika

Studijní obor: Finanční a pojistná matematika

Praha 2017

Prohlašuji, že jsem tuto diplomovou práci vypracovala 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 má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V dne

Podpis autora

Název práce: Modely úhrnů škod se závislou frekvencí a severitou

Autor: Petra Čápková

Katedra: Katedra pravděpodobnosti a matematické statistiky

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

Abstrakt: V neživotním pojištění se obvykle předpokládá nezávislost mezi počtem a výší škod. Tato práce však ukazuje, že může být předpoklad nezávislosti vynechán. Zabýváme se tedy modelováním závislosti mezi frekvencí a severitou škod. Pro zahrnutí závislosti do modelu úhrnu škod uvažujeme dvě metody. První metoda využívá zobecněné lineární modely a druhá metoda uvedená v této práci je založena na modelování závislosti pomocí kopul. Uvádíme i model s nezávislou frekvencí a severitou škod. Tento model je porovnáván s popsányými metodami v simulační části práce. Do všech uvedených modelů zahrnujeme také závislost na vysvětlujících (tarifních) proměnných.

Klíčová slova: modelování závislosti, počet škod, velikost škody, zobecněný lineární model, kopula

Title: Aggregate loss models with dependent frequency and severity

Author: Petra Čápková

Department: Department of Probability and Mathematical Statistics

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

Abstract: In non-life insurance, the independence between the number and size of claims is usually assumed. However, this thesis shows that the assumption of independence can be omitted. We deal with the dependency modeling between frequency and severity of claims. For including the dependence to the total claims model, we consider two methods. The first method uses generalized linear models and the second method used in the thesis is based on dependence modeling by copulas. We also perform a model with independent frequency and severity of claims. This model is compared with the described methods in the simulation part of the thesis. We include dependency on explanatory (rating) variables in all of these models.

Keywords: dependence modeling, number of claims, claims size, generalized linear model, copula

Ráda bych na tomto místě poděkovala RNDr. Lucii Mazurové, Ph.D. za odborné vedení, poskytnutou literaturu a cenné rady, které mi pomohly při zpracování této diplomové práce. Velké poděkování patří také mé rodině, která mi studium umožnila a vždy mě plně podporovala.

Obsah

Úvod	2
1 Základní teoretické poznatky	3
1.1 Zobecněný lineární model	3
1.1.1 Lineární regresní model	3
1.1.2 Zobecněný lineární model (GLM)	3
1.1.3 Rodina exponenciálních rozdělení (EDF)	5
1.2 Kopuly	5
1.2.1 Definice kopul a Sklarova věta	6
1.2.2 Fréchetovy-Hoeffdingovy meze	7
1.2.3 Gaussova kopula	7
1.2.4 Archimédovské kopuly	8
1.2.5 Kendallovo tau	10
2 Model úhrnu škod s nezávislou frekvencí a severitou	11
2.1 Střední hodnota úhrnu škod	11
2.2 Rozptyl úhrnu škod	12
2.3 Specifikace modelu pro portfolio pojistných smluv	13
3 Model úhrnu škod se závislou frekvencí a severitou	14
3.1 Zahrnutí závislosti pomocí zobecněných lineárních modelů	14
3.1.1 Střední hodnota úhrnu škod	15
3.1.2 Rozptyl úhrnu škod	16
3.1.3 Sdružené pravděpodobnostní rozdělení	17
3.2 Zahrnutí závislosti pomocí kopul	18
3.2.1 Marginální rozdělení	20
3.2.2 Rozdělení úhrnu škod	21
3.2.3 Specifikace modelu v případě daných tarifních proměnných	22
4 Simulační část	24
4.1 Příklad 1	25
4.2 Příklad 2	32
Závěr	41
Seznam použité literatury	42
Seznam obrázků	43
Seznam tabulek	44

Úvod

V diplomové práci se věnujeme nejnovějším přístupům k modelování úhrnu škod, kdy uvažujeme možnou závislost mezi frekvencí a severitou škod. V praxi se obvykle pracuje s tradičním kolektivním modelem rizika, ve kterém se předpokládá nezávislost počtu a výši škod, avšak reálná data takový předpoklad splňovat nemusí. Uvažujme například havarijní pojištění vozidla. Řidiči, kteří havarují vícekrát za rok, mají často menší škody, a tak může v datech vzniknout zmiňovaná závislost. Použijeme-li na data se závislou frekvencí a severitou škod model, který předpokládá jejich nezávislost, nemusíme tak dostat kvalitní odhady budoucího úhrnu škod.

V práci shrneme dvě navrhované metody pro zahrnutí možné závislosti počtů a výši škod do modelu. První metoda využívá zobecněné lineární modely (GLM, generalized linear models) a zahrnuje požadovanou závislost tak, že uvažuje počet škod jako vysvětlující proměnnou pro průměrnou výši škody. Touto metodou se zabývá Garrido a kol. (2016) a Shi a kol. (2015). Druhý přístup k zahrnutí závislosti je již obecnější a používá dvourozměrné kopuly pro sdružené rozdělení frekvence a severity. Tento model popisuje Krämer a kol. (2013). Do obou metod také zakomponujeme závislost na hodnotách tarifních proměnných.

První kapitola práce shrnuje základní teoretické poznatky, které jsou důležité pro uvedené metody zahrnutí závislosti do modelu úhrnu škod. Jedna podkapitola se věnuje zobecněným lineárním modelům včetně vysvětlení klasického lineárního modelu. V druhé podkapitole pak najdeme definici kopul, některé jejich vlastnosti a vybrané druhy kopul používané dle Krämer a kol. (2013) k modelování závislosti mezi frekvencí a severitou škod.

V další kapitole popisujeme tradiční model úhrnů škod podle Garrido a kol. (2016), který předpokládá nezávislost počtů a výši škod. Tento model uvádíme hlavně z důvodu porovnání s modely, kde takový předpoklad neuvažujeme.

Ve třetí kapitole se dostáváme k hlavní části práce, tedy k metodám zahrnutí závislosti frekvence a severity do modelu úhrnu škod. V rámci metody využívající zobecněné lineární modely jsme odvodili střední hodnotu a rozptyl úhrnu škod a sdružené rozdělení průměrné výše škody a počtu škod. V podkapitole zabývající se zahrnutím závislosti pomocí kopul uvádíme také sdružené rozdělení průměrné výše škody a počtu škod včetně hustoty úhrnu škod (Krämer a kol. (2013)).

V simulační části (Kapitola 4) ukazujeme použití obou popsaných metod na generovaných datech se závislou frekvencí a severitou a porovnáváme výsledky s případem, kdy na taková data použijeme model předpokládající nezávislost počtu a výši škod.

1. Základní teoretické poznatky

1.1 Zobecněný lineární model

1.1.1 Lineární regresní model

Nejprve si vysvětlíme normální lineární model (viz Dobson (2002)), jehož zobecněním se poté dostaneme k zobecněnému lineárnímu modelu. Lineární regresní model nám umožňuje vyjádřit vztah mezi vysvětlovanou proměnnou (odezvou) Y a množinou vysvětlujících proměnných (prediktorů, regresorů) pomocí lineární funkce neznámých parametrů modelu. Předpokládejme, že máme k dispozici m pozorování. Označme i -tou hodnotu odezvy jako Y_i a předpokládejme, že Y_i , $i = 1, \dots, m$ jsou nezávislé, dále označme hodnoty vysvětlujících proměnných x_j , $j = 1, \dots, p$ pro i -té pozorování jako x_{ij} a neznámé odhadované parametry modelu β_j , $j = 0, \dots, p$. Lineární regresní model zapisujeme ve tvaru

$$Y = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \epsilon,$$

kde β_0 je absolutní člen a ϵ je náhodná chyba. Pokud ϵ_i označuje chybovou složku i -tého pozorování, pak model pro i -té pozorování má následující tvar

$$Y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \epsilon_i,$$

kdy předpokládáme, že ϵ_i jsou nezávislé, stejně rozdělené náhodné veličiny, pro které platí

$$\epsilon_i \sim N(0, \sigma^2), \quad i = 1, \dots, m. \quad (1.1)$$

V normálním lineárním modelu je střední hodnota odezvy v závislosti na vysvětlujících proměnných vyjádřena lineární funkcí odhadovaných parametrů modelu, což nelze říci o zobecněném lineárním modelu, který bude popsán níže. V lineárním regresním modelu tedy platí

$$E(Y|x_1, \dots, x_p) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p.$$

1.1.2 Zobecněný lineární model (GLM)

Dále v práci budeme využívat zobecněný lineární model dle Dobson (2002). V takovém modelu platí, že střední hodnota vysvětlované proměnné v závislosti na regresorech je určitou funkcí lineární kombinace těchto regresorů. Pro lineární kombinaci vysvětlujících proměnných x_1, x_2, \dots, x_p si zavedeme funkci η , kterou lze vyjádřit jako

$$\eta = \eta(x_1, x_2, \dots, x_p) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p = \sum_{j=0}^p \beta_j x_j, \quad (1.2)$$

kdy je nutné položit $x_0 \equiv 1$. Funkce η se někdy nazývá lineární prediktor. Pro hodnoty vysvětlujících proměnných $x_1 = x_{i1}, \dots, x_p = x_{ip}$ máme lineární prediktor η_i vyjádřený vztahem

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}.$$

Nyní zavedeme linkovací (spojovací) funkci, která nám umožní vyjádřit vztah mezi střední hodnotou vysvětlované proměnné a lineární kombinací prediktorů. Nejprve uvádíme značení, které budeme dále používat

$$\begin{aligned} \mu &= E(Y|x_1, \dots, x_p), \\ \mu_i &= E(Y_i|x_{i1}, \dots, x_{ip}). \end{aligned}$$

Spojovací funkci budeme značit g . Tato funkce g musí být monotónní a diferencovatelná. V zobecněném lineárním modelu pak platí vztah

$$g(\mu) = \eta,$$

kde lineární prediktor η je definován vztahem (1.2). To znamená, že lze odvodit vztah pro střední hodnotu odezvy μ v závislosti na vysvětlujících proměnných jako

$$\mu = g^{-1}(\eta),$$

kde g^{-1} je inverzní funkce ke spojovací funkci g . Tato inverzní funkce existuje díky předpokladu monotónnosti linkovací funkce g .

Pro hodnoty regresorů $x_1 = x_{i1}, \dots, x_p = x_{ip}$ lze zobecněný lineární model vyjádřit následujícím způsobem

$$g(\mu_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}. \quad (1.3)$$

Pomocí vzorce (1.3) již lze odvodit vztah pro střední hodnotu vysvětlované proměnné v závislosti na daných hodnotách vysvětlujících proměnných, a tedy platí

$$\mu_i = E(Y_i|x_{i1}, \dots, x_{ip}) = g^{-1}(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}),$$

kde g^{-1} je opět inverzní funkce k linkovací funkci g .

Příklad (Logaritmická linkovací funkce):

Jednou z nejčastěji používaných linkovacích funkcí je funkce logaritmická. Je zřejmé, že tato funkce splňuje požadované vlastnosti - monotónnost a diferencovatelnost. Uvažujme-li tedy logaritmickou spojovací funkci $g(\mu) = \log(\mu)$, zobecněný lineární model pro střední hodnotu μ vysvětlované proměnné Y v závislosti na regresorech x_1, \dots, x_p je pak ve tvaru

$$\log(\mu) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p.$$

Jelikož víme, že inverzní funkce k přirozenému logaritmu je funkce exponenciální, můžeme výše uvedený model vyjádřit jako

$$\mu = e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}.$$

Poznámka: Funkce $\log(t)$ je v celé práci chápána jako přirozený logaritmus.

1.1.3 Rodina exponenciálních rozdělení (EDF)

V rámci zobecněného lineárního modelu předpokládáme, že vysvětlovaná proměnná Y má pravděpodobnostní rozdělení z takzvané rodiny exponenciálních rozdělení. Tento předpoklad je určitým zobecněním předpokladu u lineárního regresního modelu, kde se uvažuje normální rozdělení chybové složky (viz (1.1)), a tedy i odezvy.

Pravděpodobnostní rozdělení náhodné veličiny X je součástí rodiny exponenciálních rozdělení dle Ohlsson a Johansson (2010), pokud jeho hustotu f_X můžeme vyjádřit v následujícím tvaru

$$f_X(x, \theta, \phi) = \exp \left\{ \frac{x\theta - b(\theta)}{\phi/w} + c(x, \phi, w) \right\}, \quad (1.4)$$

kde $w > 0$ je daná váha, $\phi > 0$ disperzní parametr, θ je kanonický parametr splňující $\theta \in \Theta \subset \mathbb{R}$ pro otevřenou množinu Θ , dále $b(\theta)$ je kumulantová funkce a $c(x, \phi, w)$ normalizační funkce.

Pokud je kumulantová funkce $b(\theta)$ dvakrát spojitě diferencovatelná na množině Θ , pak pro střední hodnotu a rozptyl náhodné veličiny X s rozdělením z EDF platí

$$\begin{aligned} E(X) &= b'(\theta), \\ \text{var}(X) &= \frac{\phi}{w} \cdot b''(\theta), \end{aligned}$$

kde $b'(\theta)$ značí první derivaci a $b''(\theta)$ pak druhou derivaci kumulantové funkce $b(\theta)$.

Existuje-li inverzní funkce k první derivaci kumulantové funkce a označíme-li si střední hodnotu náhodné veličiny X jako μ (tedy $E(X) = \mu$), lze rozptyl náhodné veličiny X vyjádřit následujícím způsobem

$$\text{var}(X) = \frac{\phi}{w} \cdot V(\mu).$$

Funkce $V(\mu)$ v uvedené rovnici značí rozptylovou funkci, pro kterou platí vztah

$$V(\mu) = b''[(b')^{-1}(\mu)],$$

kde $(b')^{-1}(\mu)$ je inverzní funkce k první derivaci kumulantové funkce.

Do rodiny exponenciálních rozdělení patří nejen normální rozdělení, ale například také gama, Poissonovo a binomické rozdělení.

1.2 Kopuly

V této podkapitole popíšeme, co jsou to kopuly, shrneme některé jejich vlastnosti a uvedeme několik druhů kopul. Vycházíme hlavně z Nelsen (2007).

1.2.1 Definice kopul a Sklarova věta

Kopuly jsou nejčastěji používány k modelování závislosti mezi náhodnými veličinami. Jiný pohled může říkat, že kopuly jsou v podstatě funkce, které propojují vícerozměrné distribuční funkce s jejich jednorozměrnými marginálními funkcemi. Tento význam vyplývá ze Sklarovy věty, kterou zmíníme později.

Nejprve si uvedeme dvě ekvivalentní definice pro kopulu (Nelsen (2007)).

Definice 1 (Kopula). d -rozměrná kopula je distribuční funkce d -rozměrného náhodného vektoru, jehož všechna jednorozměrná marginální rozdělení jsou rovnoměrná na intervalu $(0,1)$.

Definice 2 (Kopula - ekvivalentní definice). Funkce $C(u_1, \dots, u_d) : [0,1]^d \rightarrow [0,1]$ je kopula právě tehdy, když:

1. $C(u_1, \dots, u_d)$ je neklesající v každé proměnné u_i , $i = 1, \dots, d$;
2. $C(1, \dots, 1, u_i, 1, \dots, 1) = u_i$ pro každé $i = 1, \dots, d$, $u_i \in [0,1]$;
3. pro každou dvojici vektorů $(u_1^{(1)}, \dots, u_d^{(1)})$, $(u_1^{(2)}, \dots, u_d^{(2)}) \in [0,1]^d$ takovou, že $u_i^{(1)} \leq u_i^{(2)}$, $i = 1, \dots, d$, platí následující

$$\sum_{i_1=1}^2 \dots \sum_{i_d=1}^2 (-1)^{i_1+\dots+i_d} \cdot C(u_1^{(i_1)}, \dots, u_d^{(i_d)}) \geq 0.$$

Na tomto místě uvádíme známou Sklarovu větu, díky které lze mimo jiné vidět, jakým způsobem kopula spojuje sdruženou distribuční funkci náhodného vektoru s jejími marginálními distribučními funkcemi.

Věta 1 (Sklarova věta). Nechť F je sdružená distribuční funkce d -rozměrného rozdělení s marginálními distribučními funkcemi F_1, \dots, F_d . Potom existuje kopula $C : [0,1]^d \rightarrow [0,1]$ taková, že pro každé $x_1, \dots, x_d \in [-\infty, +\infty]$ platí

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)).$$

Pokud jsou navíc marginální distribuční funkce F_1, \dots, F_d spojitě, potom je kopula C určena jednoznačně. Jinak je určena jednoznačně na $\text{Ran}(F_1) \times \dots \times \text{Ran}(F_d)$, kde $\text{Ran}(F_i)$ je obor hodnot distribuční funkce F_i .

Naopak, pokud C je kopula a F_1, \dots, F_d jsou jednorozměrné distribuční funkce, potom funkce F definovaná jako $F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d))$ je sdružená distribuční funkce d -rozměrného rozdělení s marginálními distribučními funkcemi F_1, \dots, F_d .

Důkaz. Jednotlivé kroky důkazu jsou podrobně popsány v Nelsen (2007) na straně 41.

□

1.2.2 Fréchetovy-Hoeffdingovy meze

Každá kopula $C(u_1, \dots, u_d)$ je omezena zdola i shora určitými hranicemi, které nazýváme Fréchetovy-Hoeffdingovy meze (Nelsen (2007)). Pro každou kopulu platí

$$\max\left(\sum_{i=1}^d u_i + 1 - d, 0\right) \leq C(u_1, \dots, u_d) \leq \min(u_1, \dots, u_d)$$

pro $u_i \in [0,1]$ pro každé $i = 1, \dots, d$.

Představme si nyní tři základní kopuly. **Horní Fréchetova-Hoeffdingova mez** (kopula komonotónie), kterou značíme $M(u_1, \dots, u_d)$, je definována jako

$$M(u_1, \dots, u_d) = \min(u_1, \dots, u_d), \quad u_i \in [0,1], i = 1, \dots, d.$$

Kopula, která je ve tvaru **dolní Fréchetovy-Hoeffdingovy meze** ($d=2$), se nazývá kopula kontramonotónie a je definována pouze pro dvourozměrný případ. Pro tuto kopulu $W(u_1, u_2)$ platí

$$W(u_1, u_2) = \max(u_1 + u_2 - 1, 0), \quad u_1, u_2 \in [0,1].$$

Třetí základní kopulou je takzvaná **nezávislá kopula**, kterou značíme $\Pi(u_1, \dots, u_d)$ a je dána následujícím vztahem

$$\Pi(u_1, \dots, u_d) = u_1 \cdots u_d, \quad u_i \in [0,1], i = 1, \dots, d. \quad (1.5)$$

Jelikož pro naše účely zahrnutí závislosti frekvence a severity do modelu úhrnu škod si vystačíme pouze s dvourozměrnými kopulami, následující kopuly již uvedeme pouze pro dvourozměrný případ. Tedy dále předpokládáme, že platí $d = 2$.

1.2.3 Gaussova kopula

Další důležitou kopulou je Gaussova kopula, kterou budeme dále využívat při simulační studii.

Gaussova kopula pro případ $d = 2$ je kopula dvourozměrného normálního rozdělení. To znamená, že tuto kopulu můžeme napsat ve tvaru

$$C_\rho^{Ga}(u_1, u_2) = \Phi_2\left(\Phi^{-1}(u_1), \Phi^{-1}(u_2)\right), \quad (1.6)$$

kde Φ_2 je sdružená distribuční funkce dvourozměrného normálního rozdělení s vektorem středních hodnot $\mu = (0,0)^T$ a korelační maticí

$$R = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad (1.7)$$

kde ρ je korelační koeficient. Aby byla korelační matice regulární, musí pro korelační koeficient platit $|\rho| < 1$.

Dále se ve vzorci (1.6) vyskytují symboly $\Phi^{-1}(u_1)$, respektive $\Phi^{-1}(u_2)$, kterými označujeme kvantilovou funkci jednorozměrného normovaného normálního rozdělení $N(0,1)$ v bodě u_1 , resp. u_2 .

V následující rovnici uvádíme dvourozměrnou Gaussovou kopulu v podrobnějším tvaru

$$C_\rho^{Ga}(u_1, u_2) = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho^2}} \cdot \exp \left\{ -\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{2(1-\rho^2)} \right\} dx_1 dx_2.$$

1.2.4 Archimédovské kopuly

Významnou skupinou kopul jsou takzvané archimédovské kopuly (Nelsen (2007), str. 89 – 124), mezi které patří například pro naši práci důležitá Claytonova, Gumbelova či Frankova rodina kopul. K definici archimédovských kopul je nutné nejprve definovat pseudoinverzní funkci.

Definice 3 (Pseudoinverzní funkce). *Nechť $\varphi : [0,1] \rightarrow [0,\infty]$ je spojitá, ryze klesající funkce taková, že platí $\varphi(1) = 0$. Pseudoinverzní funkcí k φ je funkce $\varphi^{[-1]} : [0,\infty] \rightarrow [0,1]$, která je definována následujícím způsobem*

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t), & 0 \leq t \leq \varphi(0), \\ 0, & \varphi(0) \leq t \leq \infty. \end{cases}$$

Z definice pseudoinverzní funkce vyplývá, že $\varphi^{[-1]}$ je spojitá, ryze klesající funkce na $[0, \varphi(0)]$ a nerostoucí na $[0, \infty]$. Dále poznamenejme, že pokud platí $\varphi(0) = \infty$, pak je pseudoinverzní funkce shodná s inverzní funkcí ($\varphi^{[-1]} = \varphi^{-1}$).

Nyní již můžeme uvést definici archimédovské kopuly.

Definice 4 (Archimédovská kopula). *Nechť $\varphi : [0,1] \rightarrow [0,\infty]$ je spojitá, ryze klesající a konvexní funkce taková, že platí $\varphi(1) = 0$. Nechť dále $\varphi^{[-1]}$ je pseudoinverzní funkcí k funkci φ . Pak je dvourozměrná archimédovská kopula C definována jako*

$$C(u_1, u_2) = \varphi^{[-1]}(\varphi(u_1) + \varphi(u_2)).$$

Funkci φ z předchozí definice říkáme generátor archimédovské kopuly. Platí-li navíc $\varphi(0) = \infty$, pak je funkce φ nazývána přesným generátorem a kopula $C(u_1, u_2) = \varphi^{-1}(\varphi(u_1) + \varphi(u_2))$ přesnou archimédovskou kopulou.

Představme si nejznámější rodiny archimédovských kopul, které se navzájem liší tvarem generátoru. Začneme s **Claytonovou rodinou**, jejíž generátor je vyjádřen následujícím způsobem

$$\varphi_{\theta}^{Cl}(t) = \frac{1}{\theta} (t^{-\theta} - 1),$$

kde $\theta \in [-1, \infty) \setminus \{0\}$ je parametr.

Na základě uvedeného generátoru je Claytonova kopula definována vztahem

$$C_{\theta}^{Cl}(u_1, u_2) = \left[\max(u_1^{-\theta} + u_2^{-\theta} - 1, 0) \right]^{-\frac{1}{\theta}}$$

pro $u_1, u_2 \in [0, 1]$ a $\theta \in [-1, \infty) \setminus \{0\}$. Poznamenejme ještě, že v případě $\theta > 0$ je $\varphi_{\theta}^{Cl}(t)$ přesným generátorem a kopula $C_{\theta}^{Cl}(u_1, u_2)$ přesnou archimédovskou kopulou.

Další významnou rodinou archimédovských kopul je **Gumbelova rodina**. Generátorem pro Gumbelovu rodinu kopul je

$$\varphi_{\theta}^{Gu}(t) = (-\log t)^{\theta},$$

kde $\theta \in [1, \infty)$ je opět parametr. Právě definovaný generátor φ_{θ}^{Gu} je přesným generátorem na celém intervalu $[1, \infty)$.

Dvourozměrná Gumbelova kopula je tedy přesnou archimédovskou kopulou a lze ji zapsat v následujícím tvaru

$$C_{\theta}^{Gu}(u_1, u_2) = \exp \left\{ - \left[(-\log u_1)^{\theta} + (-\log u_2)^{\theta} \right]^{\frac{1}{\theta}} \right\},$$

kde opět $\theta \in [1, \infty)$ a $u_1, u_2 \in [0, 1]$.

Poslední archimédovskou rodinou kopul, kterou si představíme, je **Frankova rodina**. Generátor pro tuto rodinu kopul φ_{θ}^{Fr} je definovaný následujícím vztahem

$$\varphi_{\theta}^{Fr}(t) = -\log \frac{e^{-\theta t} - 1}{e^{-\theta} - 1},$$

kde $\theta \in \mathbb{R} \setminus \{0\}$ je parametr. Generátor Frankovy rodiny kopul je, stejně jako v případě Gumbelovy rodiny, přesný.

Frankova kopula C_{θ}^{Fr} je přesná archimédovská kopula, která je dána vzorcem

$$C_{\theta}^{Fr}(u_1, u_2) = -\frac{1}{\theta} \log \left(1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1} \right)$$

pro parametr $\theta \in \mathbb{R} \setminus \{0\}$ a pro $u_1, u_2 \in [0, 1]$.

1.2.5 Kendallovo tau

V rámci základních teoretických poznatků ještě zbývá zmínit se o mírách závislosti. V této práci budeme využívat pořádkové korelace. Častou volbou je Kendallovo tau, proto se zaměříme právě na tuto míru závislosti (viz Nelsen (2007), strana 126), kterou použijeme i v simulační části práce.

Mějme dvourozměrný náhodný vektor (X_1, X_2) . Kendallovo tau je definováno jako

$$\tau = P[(X_1 - \widetilde{X}_1)(X_2 - \widetilde{X}_2) > 0] - P[(X_1 - \widetilde{X}_1)(X_2 - \widetilde{X}_2) < 0],$$

kde vektor $(\widetilde{X}_1, \widetilde{X}_2)$ je nezávislou kopií vektoru (X_1, X_2) .

Kendalovo tau nabývá hodnot z intervalu $[-1, 1]$. Pokud jsou náhodné veličiny X_1 a X_2 nezávislé, pak $\tau = 0$. Jestliže mají náhodné veličiny X_1, X_2 spojitá rozdělení a C je jejich kopula, potom lze Kendallovo tau vyjádřit jako

$$\tau = 4 \int_0^1 \int_0^1 C(u_1, u_2) dC(u_1, u_2) - 1.$$

2. Model úhrnu škod s nezávislou frekvencí a severitou

Nejprve si pro účely srovnání výsledků v rámci simulační části práce popíšeme jednodušší model úhrnu škod, který předpokládá nezávislost mezi počtem a výší škod. Tento model je popsán například v Garrido a kol. (2016). Uvažujme určitou pojistnou smlouvu (resp. pojistníka). Úhrn škod za uvažované období (nejčastěji jeden rok) pro tuto smlouvu lze vyjádřit jako

$$S = \sum_{j=1}^N Y_j,$$

kde N je diskrétní náhodná veličina vyjadřující počet škod uvažovaného pojistníka v daném období a Y_j , $j \in \{1, 2, \dots\}$ jsou spojité náhodné veličiny udávající výše jednotlivých škod pojistníka. Dle úmluvy uvažujeme $S \equiv 0$, pokud platí $N = 0$.

V případě tohoto modelu předpokládáme, že jednotlivé výše škod Y_j , $j \in \{1, 2, \dots\}$ jsou navzájem nezávislé, stejně rozdělené a nezávisí na počtu škod N . Označme obecně Y náhodnou veličinu se stejným rozdělením, jaké mají jednotlivé výše škod. Pak dle Klugman a kol. (2004) pro střední hodnotu a rozptyl úhrnu škod daného pojistníka za výše uvedených předpokladů platí

$$E(S) = E(N)E(Y), \quad (2.1)$$

$$\text{var}(S) = E(N)\text{var}(Y) + \text{var}(N) [E(Y)]^2. \quad (2.2)$$

2.1 Střední hodnota úhrnu škod

Máme-li k dispozici vektor tarifních proměnných $x = (x_1, \dots, x_p)^T$, mohou být podle Garrido a kol. (2016) tyto proměnné začleněny do našeho modelu pomocí zobecněných lineárních modelů pro náhodné veličiny N a Y . Zobecněné lineární modely byly zavedeny v první kapitole. Pokud označíme střední hodnotu náhodné veličiny N v závislosti na vektoru prediktorů x jako ν a střední hodnotu náhodné veličiny Y v závislosti na stejném vektoru prediktorů x jako μ , pak pro tyto střední hodnoty platí

$$\nu = E(N|x) = g_N^{-1}(x^T \alpha), \quad (2.3)$$

$$\mu = E(Y|x) = g_Y^{-1}(x^T \beta), \quad (2.4)$$

kde g_N^{-1} a g_Y^{-1} jsou inverzní funkce k předem daným linkovacím funkcím g_N a g_Y

(často například logaritmické linkovací funkce), α a β jsou vektory regresních koeficientů o rozměru $p \times 1$. Spojovací funkce g_N a g_Y náhodných veličin N a Y musí být monotónní a diferencovatelné (viz Kapitola 1, část 1.1.2).

Nyní můžeme za pomoci středních hodnot (2.3) a (2.4) a vztahu (2.1) vyjádřit střední hodnotu úhrnu škod při daném vektoru tarifních proměnných x následujícím způsobem

$$E(S|x) = \nu\mu = g_N^{-1}(x^T\alpha) \cdot g_Y^{-1}(x^T\beta).$$

Pokud navíc zvolíme obě linkové funkce logaritmické, tedy pokud platí rovnosti $\log(\nu) = x^T\alpha$ a $\log(\mu) = x^T\beta$, pak lze střední hodnotu úhrnu škod S v závislosti na vektoru tarifních proměnných x zapsat jako

$$E(S|x) = \nu\mu = e^{x^T\alpha + x^T\beta}.$$

2.2 Rozptyl úhrnu škod

Dále budeme předpokládat, že rozdělení počtu škod i výší škod jsou z exponenciální rodiny (Kapitola 1, část 1.1.3) s vahou $w = 1$. Hustota zmíněných náhodných veličin tak bude podle vzorce (1.4) ve tvaru

$$f_Z(z; \theta, \phi) = \exp \left\{ \frac{z\theta - b(\theta)}{\phi} + c(z, \phi) \right\}, \quad (2.5)$$

kde $b(\theta)$ a $c(z, \phi)$ jsou známé funkce, θ je kanonický parametr a ϕ disperzní parametr. Rozptyl náhodné veličiny Z s hustotou f_Z splňuje

$$\text{var}(Z) = \phi V(\mu),$$

kde μ je střední hodnota náhodné veličiny Z a $V(\mu)$ je její rozptylová funkce.

Vycházíme-li ze vzorce (2.2), platí obecně pro rozptyl úhrnu škod S za daného vektoru tarifních proměnných x v našem modelu s nezávislou frekvencí a severitou škod následující rovnost

$$\text{var}(S|x) = E(N|x)\text{var}(Y|x) + \text{var}(N|x) [E(Y|x)]^2. \quad (2.6)$$

Jsou-li navíc obě výše zmíněná rozdělení (počtu škod a výší škod) z exponenciální rodiny s vahou $w = 1$, potom pro rozptyl úhrnu škod S platí

$$\text{var}(S|x) = \nu\phi V_Y(\mu) + \psi V_N(\nu)\mu^2, \quad (2.7)$$

kde V_N značí rozptylovou funkci náhodné veličiny N (počet škod) s disperzním parametrem ψ a V_Y rozptylovou funkci náhodné veličiny Y (výše škody) s disperzním parametrem ϕ . V souladu s předchozím značením, ν je střední hodnota počtu škod v závislosti na vektoru vysvětlujících proměnných x a μ střední hodnota výší škod při daném vektoru tarifních proměnných x (vzorce (2.3) a (2.4)).

2.3 Specifikace modelu pro portfolio pojistných smluv

Uvažujme portfolio m pojistných smluv (pojistníků). Často máme k dispozici hodnoty tarifních proměnných pro každou pojistnou smlouvu. Pro $i \in \{1, \dots, m\}$ pak náhodná veličina N_i značí počet nastalých škod i -tého pojistníka a vektor $x_i = (x_{i1}, \dots, x_{ip})^T$ obsahuje jeho hodnoty tarifních proměnných. Jednotlivé výše škod na i -té pojistné smlouvě označíme Y_{ij} , $j \in \{1, 2, \dots\}$.

Úhrn škod i -tého pojistníka, který značíme S_i , lze tak vyjádřit vzorcem

$$S_i = \sum_{j=1}^{N_i} Y_{ij}$$

a celkový úhrn škod S uvažovaného portfolia je součtem úhrnů škod všech pojistníků v portfoliu. To znamená, že platí

$$S = \sum_{i=1}^m \sum_{j=1}^{N_i} Y_{ij}.$$

Pokud uvažujeme zobecněné lineární modely pro počet i výše škod, lze odhadnout parametry pro oba modely odděleně. V případě logaritmických linkovacích funkcí máme

$$\nu_i = e^{x_i^T \alpha}, \tag{2.8}$$

$$\mu_i = e^{x_i^T \beta} \tag{2.9}$$

pro všechny pojistné smlouvy $i \in \{1, \dots, m\}$, kde $\alpha = (\alpha_1, \dots, \alpha_p)^T$ je vektorový parametr počtu škod a $\beta = (\beta_1, \dots, \beta_p)^T$ je vektorový parametr výší škod.

3. Model úhrnu škod se závislou frekvencí a severitou

V této kapitole se zaměříme na model úhrnu škod, kde nepředpokládáme nezávislost mezi počtem a výší škod, jako tomu bylo v předchozím případě. Ukážeme si různé postupy zahrnutí závislosti mezi těmito náhodnými veličinami do modelu.

3.1 Zahrnutí závislosti pomocí zobecněných lineárních modelů

Uvažujeme stejný model pro úhrn škod na pojistné smlouvě za určité období jako v předchozí kapitole, máme tedy

$$S = \sum_{j=1}^N Y_j,$$

kde předpokládáme, že při daném počtu škod N jsou výše jednotlivých škod Y_1, \dots, Y_N vzájemně nezávislé a stejně rozdělené. V tomto modelu budeme pracovat s průměrnou výší škody pojistníka. Předpokládejme, že tato průměrná výše škody je závislá na počtu škod daného pojistníka.

Pro kladný počet škod ($N > 0$) náhodná veličina \bar{Y} značí průměrnou výši škody a lze ji tak vyjádřit následujícím způsobem

$$\bar{Y} = \frac{\sum_{j=1}^N Y_j}{N}. \quad (3.1)$$

Opět dle úmluvy máme $S \equiv 0$, pokud platí $N = 0$. Úhrn škod S pro $N > 0$ můžeme za použití vzorce (3.1) vyjádřit ve tvaru

$$S = \sum_{j=1}^N Y_j = N \cdot \frac{\sum_{j=1}^N Y_j}{N} = N\bar{Y}. \quad (3.2)$$

Tento model zahrnuje podle Garrido a kol. (2016) požadovanou závislost tak, že počet škod N je vysvětlující proměnnou v zobecněném lineárním modelu pro náhodnou veličinu \bar{Y} (průměrná výše škody).

Předpokládáme-li dále, že výše škod Y_j , $j \in \{1, \dots, N\}$ při daném N mají rozdělení z EDF se střední hodnotou μ a disperzním parametrem ϕ (značíme $Y_j \sim \text{EDF}(\mu, \phi)$), pak pro průměrnou výši škody platí

$$\bar{Y} \sim \text{EDF}(\mu, \phi/N).$$

3.1.1 Střední hodnota úhrnu škod

Jako v předchozím modelu máme vektor vysvětlujících (tarifních) proměnných uvažované pojistné smlouvy, který značíme $x = (x_1, \dots, x_p)^T$. Zobecněný lineární model pro počet škod pojistníka lze vyjádřit stejným vzorcem (2.3) jako v modelu s nezávislou frekvencí a severitou škod

$$\nu = E(N|x) = g_N^{-1}(x^T \alpha),$$

kde g_N^{-1} je inverzní funkce k linkovací funkci g_N náhodné veličiny N (počet škod) a $\alpha = (\alpha_1, \dots, \alpha_p)^T$ je p -rozměrný vektor regresních koeficientů. Zápis $E(N|x)$ značí střední hodnotu počtu škod N v závislosti na vektoru tarifních proměnných x .

Změna oproti modelu popsanému v předchozí kapitole nastává ve vyjádření druhého zobecněného lineárního modelu, tentokrát pro průměrnou výši škody \bar{Y} . V tomto případě musíme zahrnout závislost na proměnné počet škod N . V zobecněném lineárním modelu pro \bar{Y} uvažujeme, že počet škod N je další vysvětlující proměnnou. Dostáváme tak model

$$\mu_\theta = E(\bar{Y}|N, x) = g_Y^{-1}(x^T \beta + \theta N),$$

kde g_Y^{-1} je inverzní funkce k dané linkovací funkci g_Y , vektor $\beta = (\beta_1, \dots, \beta_p)^T$ je opět p -rozměrný vektor regresních koeficientů a θ je reálné číslo indikující stupeň závislosti průměrné výše škody \bar{Y} na počtu škod N . Zápis $E(\bar{Y}|N, x)$ označuje střední hodnotu průměrné výše škody za podmínky počtu škod v závislosti na vektoru tarifních proměnných.

Ze vzorce pro μ_θ je nyní zřejmé, že v případě nulového stupně závislosti ($\theta = 0$) přecházíme k modelu úhrnu škod s nezávislou frekvencí a severitou, který byl popsán v předchozí kapitole, neboť platí

$$\mu_\theta = \mu_0 = g_Y^{-1}(x^T \beta) = \mu,$$

kde μ je zavedeno vzorcem (2.4). Naopak pokud $\theta \neq 0$, pak následující odvození za použití (3.2) ukazuje, že se střední hodnota úhrnu škod v závislosti na vektoru tarifních proměnných v tomto případě liší od této střední hodnoty v modelu s nezávislou frekvencí a severitou škod:

$$\begin{aligned} E(S|x) &= E[E(S|N, x)|x] = E[E(N\bar{Y}|N, x)|x] \\ &= E[N E(\bar{Y}|N, x)|x] \neq E(N|x)E(Y|x). \end{aligned} \quad (3.3)$$

Navíc dále uvažujme, že g_Y je logaritmická linkovací funkce. Potom pro podmíněnou střední hodnotu průměrné výše škody dostáváme

$$\mu_\theta = e^{x^T \beta + \theta N} = \mu e^{\theta N}, \quad (3.4)$$

kde μ je střední hodnota výší škod Y v závislosti na tarifních proměnných x zavedená vzorcem (2.4) v rámci modelu s nezávislou frekvencí a severitou.

Pomocí (3.4) a s využitím odvození použitých ve vzorci (3.3) můžeme nyní střední hodnotu úhrnu škod při daném vektoru tarifních proměnných zapsat následujícím způsobem

$$E(S|x) = E(N\mu_\theta|x) = E(N\mu e^{\theta N}|x) = \mu M'_N(\theta|x), \quad (3.5)$$

kde M_N je momentová vytvořující funkce náhodné veličiny N a funkce M'_N pak značí derivaci této momentové vytvořující funkce podle proměnné θ , vše v závislosti na tarifních proměnných x (viz Garrido a kol. (2016) nebo Frees a kol. (2011)). To znamená, že platí následující vztahy

$$\begin{aligned} M_N(\theta|x) &= E(e^{\theta N}|x), \\ M'_N(\theta|x) &= E(Ne^{\theta N}|x). \end{aligned} \quad (3.6)$$

3.1.2 Rozptyl úhrnu škod

Než odvodíme rozptyl úhrnu škod v závislosti na vektoru tarifních proměnných, vyjádříme si nejprve podmíněný rozptyl úhrnu škod vzhledem k N . Pro lepší přehlednost vynecháme v následujících vzorcích vektor tarifních proměnných x ze značení.

Podmíněná střední hodnota úhrnu škod S při daném počtu škod N splňuje dle (3.2)

$$E(S|N) = NE(\bar{Y}|N) = N\mu_\theta, \quad (3.7)$$

kde μ_θ je dáno rovností (3.4). Dále uvádíme vyjádření podmíněné střední hodnoty druhé mocniny úhrnu škod S^2 při daném N

$$E(S^2|N) = N^2E(\bar{Y}^2|N) = N^2 [\text{var}(\bar{Y}|N) + \mu_\theta^2]. \quad (3.8)$$

Jelikož předpokládáme, že rozdělení výší škod je z exponenciální rodiny, pak pro podmíněný rozptyl úhrnu škod S při daném N s využitím (3.7) a (3.8) platí

$$\begin{aligned} \text{var}(S|N) &= E(S^2|N) - [E(S|N)]^2 = N^2\text{var}(\bar{Y}|N) \\ &= N^2 [(\phi_\theta/N)V_Y(\mu_\theta)] = N\phi_\theta V_Y(\mu_\theta), \end{aligned} \quad (3.9)$$

kde ϕ_θ označuje disperzní parametr rozdělení výší škod a V_Y je rozptylová funkce výší škod.

Využijeme-li vztahů (3.7) a (3.9), můžeme vyjádřit rozptyl úhrnu škod S v rámci modelu se závislou frekvencí a severitou jako

$$\begin{aligned} \text{var}(S) &= E[\text{var}(S|N)] + \text{var}[E(S|N)] \\ &= E [N\phi_\theta V_Y(\mu e^{\theta N})] + \text{var}(N\mu_\theta) \\ &= E [N\phi_\theta V_Y(\mu e^{\theta N})] + \text{var}(N\mu e^{\theta N}) \\ &= \phi_\theta E [NV_Y(\mu e^{\theta N})] + \mu^2 \left\{ E(N^2 e^{2\theta N}) - [E(Ne^{\theta N})]^2 \right\} \\ &= \phi_\theta E [NV_Y(\mu e^{\theta N})] + \mu^2 \left\{ M''_N(2\theta) - [M'_N(\theta)]^2 \right\}. \end{aligned}$$

Zahrneme-li nyní do značení závislost na vektoru tarifních proměnných x , dostaneme (viz Garrido a kol. (2016))

$$\text{var}(S|x) = \phi_\theta \text{E} \left[NV_Y(\mu e^{\theta N})|x \right] + \mu^2 \left\{ M_N''(2\theta|x) - [M_N'(\theta|x)]^2 \right\}, \quad (3.10)$$

kde M_N' je první derivace momentové vytvořující funkce náhodné veličiny N podle proměnné θ (vzorec (3.6)) a M_N'' vyjadřuje druhou derivaci momentové vytvořující funkce M_N podle proměnné θ . Tedy platí

$$M_N''(\theta|x) = \text{E}(N^2 e^{\theta N} |x).$$

Pokud zvolíme v našem modelu $\theta = 0$, což znamená nulový stupeň závislosti, máme $\phi_\theta = \phi_0 = \phi$, neboli disperzní parametr je shodný s disperzním parametrem pro výši škod v případě modelu úhrnu škod s nezávislou frekvencí a severitou popsaného v Kapitole 2. Jelikož víme, že platí $\text{var}(N|x) = \text{E}(N^2|x) - [\text{E}(N|x)]^2$, můžeme vyjádřit rozptyl úhrnu škod při daném vektoru tarifních proměnných pro nulový stupeň závislosti následujícím vzorcem

$$\begin{aligned} \text{var}(S|x) &= \phi \text{E} \{ NV_Y(\mu)|x \} + \mu^2 \text{var}(N|x) \\ &= \nu \phi V_Y(\mu) + \psi V_N(\nu) \mu^2, \end{aligned} \quad (3.11)$$

kde ψ je disperzní parametr a V_N rozptylová funkce rozdělení počtu škod N . Značení ν a μ je použito ve shodě s předchozím textem dle (2.3) a (2.4).

Vidíme, že rozptyl úhrnu škod S při daném vektoru tarifních proměnných x v případě nulového stupně závislosti (3.11) se shoduje s rozptylem úhrnu škod při daném x v rámci modelu s nezávislou frekvencí a severitou škod (Kapitola 2, vzorec (2.7)).

3.1.3 Sdružené pravděpodobnostní rozdělení

V této části práce budeme opět uvažovat určitou pojistnou smlouvu (pojistníka). Předpokládejme, že máme pro takovou pojistnou smlouvu k dispozici vektor vysvětlujících (tarifních) proměnných $x = (x_1, \dots, x_p)^T$.

Nyní budeme vycházet z Shi a kol. (2015) a odvodíme sdružené pravděpodobnostní rozdělení náhodných veličin \bar{Y} a N pro uvažovanou pojistnou smlouvu v závislosti na vektoru tarifních proměnných. Předpokládejme tentokrát pouze kladný počet škod, tedy $N > 0$. Diskrétní náhodná veličina N udává počet škod daného pojistníka. Náhodná veličina \bar{Y} je spojitá a udává průměrnou výši škody uvažovaného pojistníka.

Sdružené rozdělení náhodných veličin \bar{Y} a N v závislosti na vektoru x můžeme na základě podmíněné pravděpodobnosti zapsat jako

$$\text{P}(\bar{Y} \leq y, N = n|x) = \text{P}(N = n|x) \cdot \text{P}(\bar{Y} \leq y|N = n, x) \quad (3.12)$$

pro $y \in (0, \infty)$ a $n \in \mathbb{N}$.

Na základě (3.12) lze odvodit sdruženou hustotu náhodných veličin \bar{Y} a N v závislosti na vektoru tarifních proměnných x jako

$$\begin{aligned} f_{\bar{Y},N}(y,n|x) &= \frac{\partial}{\partial y} \mathbb{P}(\bar{Y} \leq y, N = n|x) \\ &= \mathbb{P}(N = n|x) \cdot \frac{\partial}{\partial y} \mathbb{P}(\bar{Y} \leq y|N = n,x) \end{aligned} \quad (3.13)$$

kde $\mathbb{P}(N = n|x)$ je pravděpodobnostní funkce náhodné veličiny N v závislosti na vektoru x , kterou značíme $f_N(n|x)$, a $\frac{\partial}{\partial y} [\mathbb{P}(\bar{Y} \leq y|N = n,x)]$ je podmíněná hustota průměrné výše škody \bar{Y} při daném N . Pro sdruženou hustotu (3.13) tedy platí

$$f_{\bar{Y},N}(y,n|x) = f_N(n|x) \cdot f_{\bar{Y}|N}(y|n,x).$$

3.2 Zahrnutí závislosti pomocí kopul

V této podkapitole si ukážeme, jakým způsobem lze využít kopuly v modelu úhrnu škod se závislou frekvencí a severitou.

Stejně jako v předchozí podkapitole uvažujeme dvě náhodné veličiny. Diskrétní náhodná veličina N označuje počet škod daného pojistníka. Spojitá náhodná veličina \bar{Y} reprezentuje průměrnou výši škody a je definována stejně jako v předchozí části práce vzorcem (3.1). Úhrn škod je poté dán opět vztahem (3.2).

Předpokládáme, že sdružené rozdělení průměrné výše škody \bar{Y} a počtu škod N je definováno kopulou C , která závisí na parametru θ (viz Krämer a kol. (2013)). Takovou kopulu budeme dále značit jako $C(\cdot, \cdot | \theta)$. Distribuční funkce sdruženého rozdělení $F_{\bar{Y},N|\theta}$ zmíněných náhodných veličin splňuje

$$F_{\bar{Y},N|\theta}(y,n) = C(F_{\bar{Y}}(y), F_N(n) | \theta), \quad (3.14)$$

kde $F_{\bar{Y}}$ je distribuční funkce průměrné výše škody a F_N je distribuční funkce počtu škod.

Ve vzorci (3.14) budeme dále uvažovat kopulu C z jedné ze čtyř rodin kopul popsanych v první kapitole. Jedná se o Gaussovu, Claytonovu, Gumbelovu a Frankovu rodinu kopul. Každá tato rodina kopul je závislá na parametru θ . V případě Gaussovy kopuly je tímto parametrem korelační koeficient, který je v první kapitole ve vzorci (1.7) označen jako ρ . U Claytonovy kopuly uvažujeme pouze přesnou Claytonovu kopulu, tedy pro parametr θ musí platit $\theta \in (0, \infty)$.

Kendalovo tau τ , které jsme definovali v první kapitole (část 1.2.5), může být v rámci každé zmíněné rodiny kopul vyjádřeno pomocí parametru θ . V Tabulce 3.1 je uveden tento vztah společně s rozsahem parametru θ pro Gaussovu, Claytonovu, Gumbelovu a Frankovu rodinu kopul.

rodina	rozsah θ	vztah τ a θ
Gauss	$(-1,1)$	$\tau = \frac{2}{\pi} \arcsin(\theta) \in \mathbb{R}$
Clayton	$(0,\infty)$	$\tau = \frac{\theta}{\theta+2} \in (0,\infty)$
Gumbel	$[1,\infty)$	$\tau = \frac{\theta-1}{\theta} \in [0,\infty)$
Frank	$\mathbb{R} \setminus \{0\}$	$\tau = 1 - \frac{4}{\theta} [1 - D_1(\theta)] \in \mathbb{R} \setminus \{0\}$

Tabulka 3.1: Rozsah parametru θ a vztah Kendallova tau τ a parametru θ pro vybrané rodiny kopul (zdroj: Krämer a kol. (2013)).

V Tabulce 3.1 se objevila Debyeova funkce $D_k(x)$, která je definována následujícím způsobem

$$D_k(x) = \frac{k}{x^k} \int_0^x \frac{t^k}{e^t - 1} dt,$$

kde k je přirozené číslo ($k \in \mathbb{N}$).

Pro další účely potřebujeme znát sdruženou hustotu náhodných veličin \bar{Y} (průměrná výše škody) a N (počet škod). Tato sdružená hustota $f_{\bar{Y},N}$ je definována vztahem

$$f_{\bar{Y},N}(y,n) = \frac{\partial}{\partial y} \mathbb{P}(\bar{Y} \leq y, N = n). \quad (3.15)$$

K odvození sdružené hustoty $f_{\bar{Y},N}$ za použití kopuly $C(\cdot, \cdot | \theta)$ budeme využívat první parciální derivaci kopuly podle první proměnné, kterou označíme $D_1(\cdot, \cdot | \theta)$. To znamená, že platí

$$D_1(u_1, u_2 | \theta) = \frac{\partial}{\partial u_1} C(u_1, u_2 | \theta)$$

pro $u_1, u_2 \in (0,1)$.

Tabulka 3.2 ukazuje první parciální derivace $D_1(\cdot, \cdot | \theta)$ v bodech u_1 a u_2 pro námi uvažované čtyři rodiny kopul.

rodina	první parciální derivace $D_1(u_1, u_2 \theta)$
Gauss	$\Phi \left(\frac{\Phi^{-1}(u_2) - \theta \Phi^{-1}(u_1)}{\sqrt{1 - \theta^2}} \right)$
Clayton	$(u_1^{-\theta} + u_2^{-\theta} - 1)^{-(1/\theta)-1} u_1^{-\theta-1}$
Gumbel	$u_1^{-1} \exp \left\{ - \left[(-\log u_1)^\theta + (-\log u_2)^\theta \right]^{1/\theta} \right\}$
Frank	$\frac{e^\theta (e^{\theta u_2} - 1)}{e^{\theta(u_1+1)} + e^{\theta(u_2+1)} - e^\theta - e^{\theta(u_1+u_2)}}$

Tabulka 3.2: První parciální derivace $D_1(u_1, u_2 | \theta)$ pro vybrané rodiny kopul (zdroj: Krämer a kol. (2013)).

Na základě vzorce (3.15) lze sdruženou hustotu spojité náhodné veličiny \bar{Y} a diskrétní náhodné veličiny N vyjádřit jako

$$\begin{aligned} f_{\bar{Y},N}(y,n) &= \frac{\partial}{\partial y} P(\bar{Y} \leq y, N \leq n) - \frac{\partial}{\partial y} P(\bar{Y} \leq y, N \leq n-1) \\ &= \frac{\partial}{\partial y} C(F_{\bar{Y}}(y), F_N(n)|\theta) - \frac{\partial}{\partial y} C(F_{\bar{Y}}(y), F_N(n-1)|\theta) \\ &= f_{\bar{Y}}(y) [D_1(F_{\bar{Y}}(y), F_N(n)|\theta) - D_1(F_{\bar{Y}}(y), F_N(n-1)|\theta)], \end{aligned} \quad (3.16)$$

neboť hustota průměrné výše škody $f_{\bar{Y}}(y)$ je derivací její distribuční funkce, kterou značíme $F_{\bar{Y}}(y)$.

3.2.1 Marginální rozdělení

Prozatím byla sdružená hustota odvozena obecně pro dvojici tvořenou diskrétní a spojitou náhodnou veličinou. V této části se zaměříme pouze na kladné náhodné veličiny průměrné výše škody \bar{Y} a počtu škod N . Připomínáme, že dle úmluvy pro úhrn škod S platí $S \equiv 0$, pokud $N = 0$.

Průměrnou výši škody budeme modelovat pomocí gama rozdělení, jehož hustotu lze vyjádřit ve tvaru

$$f_{\bar{Y}}(y|\mu,\delta) = \frac{1}{y \cdot \Gamma(\frac{1}{\delta})} \left(\frac{y}{\mu\delta}\right)^{\frac{1}{\delta}} \exp\left(-\frac{y}{\mu\delta}\right), \quad y > 0, \quad (3.17)$$

kde parametr $\mu > 0$ je zároveň střední hodnotou rozdělení a $\delta > 0$ je disperzní parametr (McCullagh a Nelder (1989)).

Dále uvažujme, že počet škod N za předpokladu, že platí $N > 0$, má Poissonovo rozdělení useknuté v nule (zero-truncated Poisson, ZTP). Hustotu náhodné veličiny N můžeme tedy zapsat jako

$$f_N(n|\lambda) = \frac{\lambda^n}{n![1 - \exp(-\lambda)]} \cdot \exp(-\lambda), \quad n = 1, 2, \dots,$$

kde $\lambda > 0$ je parametr rozdělení (Klugman a kol. (2004)).

V Tabulce 3.3 jsou uvedeny základní momenty marginálních rozdělení, která uvažujeme pro náhodné veličiny - průměrnou výši škody \bar{Y} a počet škod N .

	průměrná výše škody \bar{Y}	počet škod N
rozdělení	gama	ZTP
parametry	$\mu > 0, \delta > 0$	$\lambda > 0$
střední hodnota	$E(\bar{Y}) = \mu$	$E(N) = \frac{\lambda}{1 - e^{-\lambda}}$
rozptyl	$\text{var}(\bar{Y}) = \mu^2 \delta$	$\text{var}(N) = \frac{\lambda[1 - e^{-\lambda}(\lambda + 1)]}{(1 - e^{-\lambda})^2}$

Tabulka 3.3: Parametry, střední hodnota a rozptyl dvou uvažovaných marginálních rozdělení pro náhodné veličiny \bar{Y} a N (zdroj: Krämer a kol. (2013)).

3.2.2 Rozdělení úhrnu škod

Úhrn škod S , který je definován vzorcem (3.2), je kladná (pokud pro počet škod platí $N > 0$) spojitá náhodná veličina, která má čtyři parametry μ, δ, λ a θ na základě předpokládaných marginálních rozdělení a předpokládané kopuly.

Nejprve si zapíšeme sdruženou hustotu průměrné výše škody a počtu škod v závislosti na parametrech μ, δ, λ a θ . Ze vzorce (3.16) tedy pro tuto sdruženou hustotu dostáváme rovnost

$$\begin{aligned} f_{\bar{Y}, N}(y, n | \mu, \delta, \lambda, \theta) \\ = f_{\bar{Y}}(y | \mu, \delta) [D_1(F_{\bar{Y}}(y | \mu, \delta), F_N(n | \lambda) | \theta) - D_1(F_{\bar{Y}}(y | \mu, \delta), F_N(n - 1 | \lambda) | \theta)] \end{aligned}$$

pro $y > 0$ a $n = 1, 2, \dots$, kde $F_{\bar{Y}}$ je nyní distribuční funkce gama rozdělení a F_N distribuční funkce Poissonova rozdělení useknutého v nule.

Dále nás zajímá rozdělení úhrnu škod S . Následující věta z Krämer a kol. (2013) uvádí vyjádření jeho hustoty.

Věta 2. *Rozdělení úhrnu škod S je dáno hustotou $f_S(s | \mu, \delta, \lambda, \theta)$, pro kterou platí*

$$\begin{aligned} f_S(s | \mu, \delta, \lambda, \theta) \\ = \sum_{n=1}^{\infty} \left[D_1 \left(F_{\bar{Y}} \left(\frac{s}{n} | \mu, \delta \right), F_N(n | \lambda) | \theta \right) - D_1 \left(F_{\bar{Y}} \left(\frac{s}{n} | \mu, \delta \right), F_N(n - 1 | \lambda) | \theta \right) \right] \\ \cdot \frac{1}{n} f_{\bar{Y}} \left(\frac{s}{n} | \mu, \delta \right) \end{aligned}$$

pro $s > 0$.

Důkaz. Abychom si zjednodušili značení, vynecháme v tomto důkazu parametry ze vzorců. Uvažujme dvourozměrný náhodný vektor

$$(S, N)^T \in \mathbb{R}^+ \times \{1, 2, \dots\}.$$

Nyní odvodíme sdruženou hustotu náhodného vektoru $(S, N)^T$ (podobně jako v (3.15))

$$\begin{aligned} f_{S, N}(s, n) &= \frac{\partial}{\partial s} \mathbb{P}(S \leq s, N = n) \\ &= \frac{\partial}{\partial s} \mathbb{P}\left(\bar{Y} \leq \frac{s}{n}, N = n\right), \end{aligned}$$

neboť $\bar{Y} = \frac{S}{N}$. Zavedeme-li substituci $y = \frac{s}{n}$, dostaneme za použití vzorce (3.15)

$$\begin{aligned} f_{S, N}(s, n) &= \frac{\partial}{\partial y} \mathbb{P}\left(\bar{Y} \leq y, N = n\right) \cdot \frac{\partial y}{\partial s} \\ &= f_{\bar{Y}, N}\left(\frac{s}{n}, n\right) \cdot \frac{1}{n}. \end{aligned}$$

Zbývá dosadit sdruženou hustotu náhodných veličin \bar{Y} a N podle vzorce (3.16). Věta je pak dokázána vyjádřením marginální hustoty náhodné veličiny S (sumace přes diskrétní náhodnou veličinu N), tedy platí

$$f_S(s) = \sum_{n=1}^{\infty} f_{S, N}(s, n).$$

□

3.2.3 Specifikace modelu v případě daných tarifních proměnných

Doposud byly v podkapitole 3.2 průměrná výše škody a počet škod modelovány nezávisle na tarifních (vysvětlujících) proměnných. Dále budeme uvažovat, že máme tyto vysvětlující proměnné k dispozici.

Předpokládejme, že máme opět portfolio m pojistných smluv (pojistníků). Nechtě $\bar{Y}_i \in \mathbb{R}^+$, $i = 1, 2, \dots, m$ jsou nezávislé spojité náhodné veličiny, kde \bar{Y}_i značí průměrnou výši škody i -tého pojistníka. Dále mějme $N_i \in \mathbb{N}$, $i = 1, 2, \dots, m$ nezávislé diskrétní náhodné veličiny, kde N_i označuje počet škod i -tého pojistníka. Uvažujme vektor tarifních proměnných $x_i \in \mathbb{R}^p$ pro i -tou smlouvu, tedy $x_i = (x_{i1}, \dots, x_{ip})^T$.

Pro marginální rozdělení se využívají zobecněné lineární modely s logaritmickou linkovací funkcí

$$\begin{aligned} N_i &\sim \text{ZTP}(\lambda_i), & \log(\lambda_i) &= x_i^T \alpha, \\ \bar{Y}_i &\sim \text{gama}(\mu_i, \delta), & \log(\mu_i) &= x_i^T \beta, \end{aligned} \tag{3.18}$$

kde $\alpha \in \mathbb{R}^p$ je vektor regresních koeficientů pro počet škod a $\beta \in \mathbb{R}^p$ vektor regresních koeficientů pro průměrnou výši škody (Krämer a kol. (2013)).

Ve specifikaci marginálních modelů (3.18) si lze všimnout, že druhý parametr gama rozdělení δ je shodný pro všechny pojistné smlouvy.

Obecně lze uvažovat dva různé vektory tarifních proměnných, jeden pro počet škod a druhý pro průměrnou výši škody. Tyto vektory mohou obsahovat jiné proměnné i jiný počet těchto proměnných. V naší práci však předpokládáme shodný vektor pro obě popsané náhodné veličiny.

Označíme-li si úhrn škod i -tého pojistníka jako S_i , platí

$$S_i = \bar{Y}_i \cdot N_i.$$

Celkový úhrn škod portfolia, který značíme S , je pak součtem úhrnů škod všech pojistníků v portfoliu

$$S = \sum_{i=1}^m S_i = \sum_{i=1}^m \bar{Y}_i \cdot N_i.$$

4. Simulační část

V této kapitole si pomocí dvou příkladů ukážeme použití obou popsaných metod, které lze uvažovat pro zahrnutí závislosti frekvence a severity do modelu úhrnu škod. V prvním příkladě porovnáme použití modelu s nezávislou frekvencí a severitou škod (Kapitola 2) a modelu, který předpokládá závislost mezi počtem a výší škod za pomoci zobecněných lineárních modelů (Kapitola 3, část 3.1). V druhém příkladě pak srovnáme použití modelu popsaného v rámci druhé kapitoly s modelem předpokládajícím závislost frekvence a severity škod s využitím kopul (Kapitola 3, část 3.2).

V simulační části práce uvažujeme havarijní pojištění vozidla. Pro vysvětlení počtu a výší škod v obou příkladech budeme uvažovat tři stejné tarifní proměnné. První proměnnou je pohlaví pojistníka. Tato tarifní proměnná obsahuje logicky dvě kategorie (muž a žena), každou s pravděpodobností $1/2$. Další tarifní proměnnou je věk pojistníka. V rámci tohoto regresoru uvažujeme celočíselné hodnoty z intervalu $[18, 70]$, každou hodnotu se stejnou pravděpodobností (tedy s pravděpodobností $1/53$). Posledním uvažovaným regresorem je typ pojištěného vozidla, kde předpokládáme tři kategorie (motocykl, osobní automobil a nákladní automobil), opět každý typ se stejnou pravděpodobností (tedy $1/3$). Referenční kategorií u proměnné pohlaví jsme zvolili muže a u proměnné typ vozidla motocykl.

Dále jsme si pro účely simulace dat určili hodnoty regresních koeficientů $\alpha = (\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4)^T$ pro náhodnou veličinu počet škod N a $\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)^T$ pro průměrnou výši škody \bar{Y} . Koeficienty α_0 a β_0 jsou absolutními členy (intercepty). Dále α_1 a β_1 odpovídají kategorii žena u tarifní proměnné pohlaví, α_2 a β_2 jsou regresními koeficienty proměnné věk. U tarifní proměnné typ vozidla koeficienty α_3 a β_3 přísluší kategorii osobní automobil (OA) a α_4 a β_4 kategorii nákladní automobil (NA).

Zvolené hodnoty popsaných regresních koeficientů pro počet škod a průměrnou výši škody můžeme vidět v Tabulce 4.1. V rámci naší simulační části práce tak mají ženy kladný efekt na počet škod (lze chápat jako počet nastalých nehod) a naopak záporný efekt na průměrnou výši škody. Nejvyššího kladného efektu na průměrnou výši škody si lze všimnout u nákladního automobilu. V našich simulacích tak předpokládáme, že nákladní automobil v průměru způsobí vyšší škody než motocykl.

	interc.	žena	věk	OA	NA
počet škod N	-1,00	0,40	0,04	0,50	0,30
průměrná výše škody \bar{Y}	-0,30	-0,80	-0,04	1,10	1,50

Tabulka 4.1: Regresní koeficienty pro počet škod a průměrnou výši škody.

4.1 Příklad 1

V prvním příkladě uvažujeme portfolio obsahující 700 pojistných smluv (pojistníků). Pro všechny pojistné smlouvy jsme vygenerovali pomocí softwaru R hodnoty tarifních proměnných popsaných výše a sestavili tak regresní matici $X \in \mathbb{R}^{700 \times 5}$. Pro představu uvádíme prvních pět řádků této matice X (Tabulka 4.2).

Vidíme, že prvním generovaným pojistníkem je muž ve věku 62 let s pojištěným motocyklem (referenční kategorie typu vozidla). Poté například čtvrtý pojistník je žena ve věku 63 let s pojištěným osobním automobilem.

interc.	žena	věk	OA	NA
1	0	62	0	0
1	1	31	0	0
1	0	19	0	0
1	1	63	1	0
1	1	35	0	0
...

Tabulka 4.2: Prvních pět řádků regresní matice X (Příklad 1).

S využitím matice X a koeficientů α, β generujeme počty škod N_i a výše škod $Y_{i,j}$, $j = 1, \dots, N_i$ pro každou pojistnou smlouvu $i = 1, \dots, 700$. Při generování těchto hodnot předpokládáme určitou závislost mezi frekvencí a severitou škod.

Bylo provedeno 100 simulací, kdy v každé takové simulaci byly generovány pro každého pojistníka $i = 1, \dots, 700$ nejprve počty škod N_i z Poissonova rozdělení se střední hodnotou ν_i , pro kterou na základě (2.8) platí

$$\nu_i = \exp(-1 + 0,4x_{i1} + 0,04x_{i2} + 0,5x_{i3} + 0,3x_{i4}),$$

kde $x_i = (x_{i0}, x_{i1}, x_{i2}, x_{i3}, x_{i4})$ je i -tý řádek matice X a $x_{i0} = 1$ pro všechna $i = 1, \dots, 700$.

Pro pojistné smlouvy, kde byl počet škod N_i generován jako 0, jsme hodnotu úhrnu škod S_i nastavili také na hodnotu 0. Ve zbylých případech jsme generovali výše škod Y_{i1}, \dots, Y_{iN_i} (již pro známé vygenerované hodnoty N_i) z gama rozdělení s parametry μ_{θ_i} (střední hodnota) a $\delta = 0,25$, jehož hustotu uvažujeme ve tvaru (3.17). Pro střední hodnotu μ_{θ_i} navíc dle (3.4) platí

$$\mu_{\theta_i} = \exp(-0,3 - 0,8x_{i1} - 0,04x_{i2} + 1,1x_{i3} + 1,5x_{i4} + \theta N_i),$$

kde $x_i = (x_{i0}, x_{i1}, x_{i2}, x_{i3}, x_{i4})$ je opět i -tý řádek matice X a θ vyjadřuje předem specifikovaný stupeň závislosti mezi frekvencí a severitou škod. V tomto příkladě

jsme uvažovali pět různých hodnot stupně závislosti $\theta \in \{0,10; 0,15; 0,20; 0,25; 0,30\}$ a simulace jsme provedli pro všechny tyto hodnoty.

Z generovaných dat v každé simulaci byly odhadnuty regresní koeficienty pro počet škod a pro průměrnou výši škody na základě daných hodnot tarifních proměnných (matice X). Odhady regresních koeficientů $\alpha = (\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4)^T$ pro počet škod v modelu se závislou i v modelu s nezávislou frekvencí a severitou škod jsou shodné, neboť jsou odhadovány pomocí stejného zobecněného lineárního modelu. Odhady regresních koeficientů $\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)^T$ pro průměrnou výši škody se však ve zmíněných modelech liší. V rámci modelu se závislou frekvencí a severitou škod přibude odhad stupně závislosti θ . Všechny regresní koeficienty byly odhadnuty metodou nejmenších čtverců pomocí softwaru R. Odhad regresních koeficientů α budeme značit $\hat{\alpha}$ a koeficientů β pak $\hat{\beta}$ (model s nezávislou frekvencí a severitou škod) a $\check{\beta}$ (model se závislou frekvencí a severitou škod). Odhad stupně závislosti θ dále značíme $\check{\theta}$.

Poté jsme vypočítali průměr odhadnutých koeficientů $\hat{\alpha}$, $\hat{\beta}$, $\check{\beta}$ a $\check{\theta}$ ze 100 proběhnutých simulací. Popsaný postup jsme aplikovali na všechny zvolené hodnoty stupně závislosti $\theta \in \{0,10; 0,15; 0,20; 0,25; 0,30\}$.

Výsledné průměry odhadnutých koeficientů z našich simulací můžeme vidět v Tabulkách 4.3, 4.4 a 4.5. Jelikož data byla generována pomocí vzorců pro model úhrnu škod se závislou frekvencí a severitou, je zřejmé, že pro tento model vycházejí odhady koeficientů přibližně stejné pro všechny stupně závislosti θ a zároveň odpovídají zvoleným regresním koeficientům (viz Tabulka 4.1), které jsme použili pro výpočet středních hodnot ν_i a μ_{θ_i} pro $i = 1, \dots, 700$.

Naopak, použijeme-li pro odhad regresních koeficientů model s nezávislou frekvencí a severitou škod, hodnoty odhadů $\hat{\beta}$ se pro různé stupně závislosti θ značně liší.

θ	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$
0,10	-1,005	0,403	0,040	0,503	0,312
0,15	-0,986	0,398	0,040	0,500	0,304
0,20	-1,002	0,401	0,040	0,496	0,304
0,25	-1,015	0,402	0,040	0,501	0,301
0,30	-1,005	0,403	0,040	0,499	0,302

Tabulka 4.3: Zprůměrované hodnoty odhadnutých regresních koeficientů α pro počet škod N .

θ	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
0,10	-0,745	-0,626	-0,024	1,329	1,619
0,15	-0,993	-0,537	-0,015	1,454	1,681
0,20	-1,248	-0,426	-0,006	1,580	1,740
0,25	-1,558	-0,317	0,005	1,744	1,813
0,30	-1,833	-0,197	0,015	1,891	1,873

Tabulka 4.4: Zprůměrované hodnoty odhadnutých regresních koeficientů β pro průměrnou výši škody \bar{Y} při použití modelu úhrnu škod s nezávislou frekvencí a severitou.

θ	$\check{\beta}_0$	$\check{\beta}_1$	$\check{\beta}_2$	$\check{\beta}_3$	$\check{\beta}_4$	$\check{\theta}$
0,10	-0,292	-0,800	-0,040	1,098	1,499	0,101
0,15	-0,300	-0,804	-0,040	1,101	1,500	0,150
0,20	-0,293	-0,798	-0,040	1,095	1,496	0,200
0,25	-0,303	-0,800	-0,040	1,097	1,500	0,251
0,30	-0,300	-0,799	-0,040	1,101	1,505	0,300

Tabulka 4.5: Zprůměrované hodnoty odhadnutých regresních koeficientů β pro průměrnou výši škody \bar{Y} při použití modelu úhrnu škod se závislou frekvencí a severitou.

Pro účely porovnání středních hodnot celkového úhrnu škod S si vyjádříme teoretické hodnoty této střední hodnoty pro všechny uvažované hodnoty θ . Pro střední hodnotu celkového úhrnu škod S_i i -tého pojistníka v závislosti na vektoru tarifních proměnných $x_i = (x_{i0}, x_{i1}, x_{i2}, x_{i3}, x_{i4})^T$ v modelu se závislou frekvencí a severitou škod podle (3.5) platí

$$E(S_i|x_i) = \mu_i M'_{N_i}(\theta|x_i),$$

kde M'_{N_i} je derivace momentové vytvořující funkce počtu škod N_i a μ_i je střední hodnota výši škod (definována vzorcem (2.9)), předpokládáme-li jejich nezávislost na počtu škod.

V popsané simulaci generujeme pro každého pojistníka počet škod N_i z Poissonova rozdělení se střední hodnotou ν_i . Odvodíme tedy nejprve momentovou vytvořující funkci pro takové Poissonovo rozdělení

$$\begin{aligned} M_{N_i}(\theta|x_i) &= \mathbb{E}\left(e^{\theta N_i}|x_i\right) = \sum_{k=0}^{\infty} \frac{\nu_i^k}{k!} e^{-\nu_i} e^{\theta k} \\ &= e^{-\nu_i} \cdot \sum_{k=0}^{\infty} \frac{(\nu_i e^{\theta})^k}{k!} = e^{-\nu_i} \cdot e^{\nu_i e^{\theta}} \\ &= e^{\nu_i(e^{\theta}-1)}. \end{aligned}$$

(Výpočet lze nalézt také v Klugman a kol. (2004) na straně 37.)

Zderivujeme-li nyní momentovou vytvořující funkci M_{N_i} podle θ , dostaneme

$$M'_{N_i}(\theta|x_i) = e^{\nu_i(e^{\theta}-1)} \cdot \nu_i e^{\theta} = \nu_i \cdot \exp[\nu_i(e^{\theta}-1) + \theta].$$

Střední hodnotu úhrnu škod i -tého pojistníka v závislosti na hodnotách tarifních proměnných x_i můžeme pro předem zvolený stupeň závislosti θ vyjádřit vzorcem

$$\mathbb{E}(S_i|x_i) = \mu_i \nu_i \cdot \exp[\nu_i(e^{\theta}-1) + \theta], \quad (4.1)$$

kde $\nu_i = \exp(x_i^T \alpha)$ a $\mu_i = \exp(x_i^T \beta)$ pro zvolené hodnoty regresních koeficientů α a β (viz Tabulka 4.1).

Sečteme-li střední hodnoty úhrnu škod všech pojistníků v portfoliu, dostaneme střední hodnotu celkového úhrnu škod portfolia S v závislosti na hodnotách tarifních proměnných (matice X)

$$\mathbb{E}(S|X) = \sum_{i=1}^{700} \mathbb{E}(S_i|x_i). \quad (4.2)$$

Právě popsané střední hodnoty jsme spočítali pro všechny uvažované stupně závislosti θ . Jelikož data v tomto příkladu byla generována pomocí modelu se závislou frekvencí a severitou škod, je vypočtená střední hodnota teoretickou očekávanou hodnotou celkového součtu všech vygenerovaných škod pojistníků v portfoliu. Takové hodnoty můžeme porovnat pro uvažované stupně závislosti θ v Tabulce 4.6. Součty vygenerovaných škod jsou spočteny jako průměry těchto součtů ze 100 provedených simulací. Vidíme, že součty škod odpovídají vypočteným středním hodnotám pomocí vzorců (4.1) a (4.2), jak jsme očekávali.

stupeň závislosti θ	0,10	0,15	0,20	0,25	0,30
střední hodnota úhrnu škod	1174,9	1694,9	2598,1	4289,3	7731,9
součet vygenerovaných škod	1174,0	1686,2	2566,7	4242,5	7527,1

Tabulka 4.6: Střední hodnoty úhrnu škod pro různé stupně závislosti θ při předpokládané závislosti mezi počtem a výší škod v porovnání s průměrnými součty vygenerovaných škod z provedených simulací.

Abychom mohli ukázat, co způsobí na našich datech s předpokládanou závislostí mezi frekvencí a severitou škod nekorektní volba modelu, vypočítali jsme si odhady úhrnu škod pro oba modely (se závislou i nezávislou frekvencí a severitou).

Pro model se závislou frekvencí a severitou škod (dále jen závislý model) postupujeme obdobně jako při výpočtu střední hodnoty, ale s tím rozdílem, že místo zvolených hodnot regresních koeficientů α a β a stupně závislosti θ použijeme vypočtené průměrné odhady těchto koeficientů $\hat{\alpha}$ a $\hat{\beta}$ a průměrný odhad stupně závislosti $\hat{\theta}$ (viz Tabulky 4.3 a 4.5). Odhad úhrnu škod i -tého pojistníka v závislosti na hodnotách tarifních proměnných x_i pro závislý model \check{S}_i spočteme jako

$$\check{S}_i = \check{\mu}_i \hat{\nu}_i \cdot \exp[\hat{\nu}_i (e^{\hat{\theta}} - 1) + \hat{\theta}],$$

kde $\hat{\nu}_i = \exp(x_i^T \hat{\alpha})$ a $\check{\mu}_i = \exp(x_i^T \hat{\beta})$. Odhad \check{S} celkového úhrnu škod portfolia je pak součtem odhadů úhrnů škod pro všechny pojistné smlouvy

$$\check{S} = \sum_{i=1}^{700} \check{S}_i.$$

V případě modelu s nezávislou frekvencí a severitou (dále nezávislý model) používáme při výpočtu odhadu úhrnu škod odpovídající průměrné odhady regresních koeficientů $\hat{\alpha}$ a $\hat{\beta}$ (viz Tabulky 4.3 a 4.4). Odhad úhrnu škod i -tého pojistníka \hat{S}_i v závislosti na hodnotách tarifních proměnných je poté určen vzorcem

$$\hat{S}_i = \hat{\mu}_i \hat{\nu}_i,$$

kde platí $\hat{\nu}_i = \exp(x_i^T \hat{\alpha})$ a $\hat{\mu}_i = \exp(x_i^T \hat{\beta})$. Opět jsme vypočítali odhad celkového úhrnu škod portfolia

$$\hat{S} = \sum_{i=1}^{700} \hat{S}_i.$$

Odhady celkových úhrnů pro závislý a nezávislý model jsme vyčíslili pro uvažované stupně závislosti θ . Porovnat tyto odhady společně se střední hodnotou vypočtenou pomocí zvolených regresních koeficientů a zvolenými stupni závislosti můžeme v Tabulce 4.7. Pro lepší představu jsme tyto hodnoty vynesli do společného grafu (viz Obrázek 4.1).

Použijeme-li pro výpočet odhadu celkového úhrnu škod model, ve kterém předpokládáme nezávislost mezi počtem a výší škod, lze si všimnout pro naše zvolené stupně závislosti θ , že v případě dat generovaných pomocí závislého modelu podhodnocuje nezávislý model tyto odhady. Pro vyšší uvažovaný stupeň závislosti při generování dat je toto podhodnocení výraznější. Předpokládáme-li θ rovno hodnotě 0,3 střední hodnota úhrnu škod vychází přibližně 7731,9, zatímco odhad tohoto úhrnu škod pomocí nezávislého modelu je přibližně 4691,9. Naopak odhady celkových úhrnů škod pomocí závislého modelu podle očekávání odpovídají střední hodnotě těchto úhrnů, neboť data byla generována právě pomocí závislého modelu.

stupeň závislosti θ	0,10	0,15	0,20	0,25	0,30
střední hodnota úhrnu škod	1174,9	1694,9	2598,1	4289,3	7731,9
odhad úhrnu škod (závislý m.)	1170,4	1688,9	2563,5	4275,3	7667,5
odhad úhrnu škod (nezávislý m.)	1064,3	1444,2	2029,5	3018,9	4691,9

Tabulka 4.7: Střední hodnoty úhrnu škod, odhady úhrnu škod s použitím odhadnutých regresních koeficientů pomocí závislého a nezávislého modelu pro zvolené hodnoty stupně závislosti θ .

Abychom neporovnávali pouze odhady celkových úhrnů škod, vypočítali jsme pro uvažované hodnoty stupně závislosti θ průměrný procentuální rozdíl (APD, average percent difference) odhadů úhrnu škod na pojistné smlouvě mezi závislým a nezávislým modelem (Garrido a kol. (2016)). Procentuální rozdíl odhadů úhrnu škod i -tého pojistníka lze vyjádřit jako

$$APD_i = \frac{\check{\mu}_i \hat{\nu}_i \cdot \exp[\hat{\nu}_i(e^\theta - 1) + \check{\theta}]}{\hat{\mu}_i \hat{\nu}_i} - 1,$$

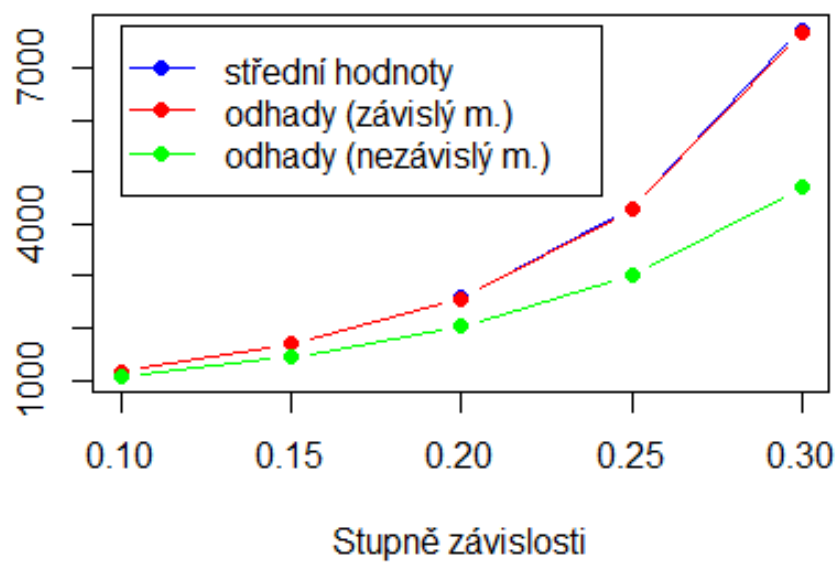
kde opět platí $\hat{\nu}_i = \exp(x_i^T \hat{\alpha})$, $\check{\mu}_i = \exp(x_i^T \check{\beta})$ a $\hat{\mu}_i = \exp(x_i^T \hat{\beta})$ pro hodnoty tarifních proměnných i -tého pojistníka x_i . Průměrný procentuální rozdíl odhadů úhrnu škod na pojistné smlouvě mezi závislým a nezávislým modelem je pak určen vzorcem

$$APD = 100 \cdot \frac{1}{700} \sum_{i=1}^{700} APD_i.$$

Popsaný průměrný procentuální rozdíl odhadů úhrnu škod na pojistné smlouvě jsme vypočítali pro všechny uvažované hodnoty stupně závislosti θ na základě našich simulací a výsledky lze vidět v Tabulce 4.8. Pokud jsme pro simulace uvažovali stupeň závislosti roven hodnotě 0,3, pak průměrný procentuální rozdíl odhadů úhrnu škod na pojistné smlouvě je velmi významný (přibližně 32,6 %).

stupeň závislosti θ	0,10	0,15	0,20	0,25	0,30
průměrný procentuální rozdíl [%]	9,31	14,52	19,66	26,20	32,59

Tabulka 4.8: Průměrný procentuální rozdíl odhadů úhrnů škod na pojistné smlouvě mezi závislým a nezávislým modelem pro uvažované hodnoty stupně závislosti θ .



Obrázek 4.1: Střední hodnoty úhrnu škod, odhady úhrnu škod s použitím odhadnutých regresních koeficientů pomocí závislého a nezávislého modelu pro zvolené hodnoty stupně závislosti θ v grafu.

4.2 Příklad 2

V druhém příkladě budeme uvažovat portfolio s nižším počtem pojistníků, konkrétně 400, ale pouze kladný počet škod pro každou pojistnou smlouvu. Pro každého pojistníka jsme opět vygenerovali hodnoty tarifních proměnných.

Tyto vysvětlující proměnné uvažujeme stejné jako v předchozím příkladě se shodnými předpokládanými pravděpodobnostmi při jejich generování. Pro i -tou smlouvu tak máme vektor tarifních proměnných $x_i = (x_{i0}, x_{i1}, x_{i2}, x_{i3}, x_{i4})^T$, kde $x_{i0} = 1$ pro každé $i = 1, \dots, 400$. Dále x_{i1} nabývá hodnoty 0 pro muže a 1 pro ženy, x_{i2} udává věk i -tého pojistníka. Pokud je pojištěným vozidlem osobní automobil, platí $x_{i3} = 1$, nebo může být pojištěným vozidlem nákladní automobil, pak $x_{i4} = 1$. Jsou-li obě hodnoty x_{i3} a x_{i4} nulové, pak je pojištěným vozidlem motocykl.

Regresní matice $X \in \mathbb{R}^{400 \times 5}$ obsahuje všechny generované hodnoty tarifních proměnných, kdy i -tý řádek odpovídá i -té pojistné smlouvě. V Tabulce 4.9 můžeme vidět prvních pět řádků sestavené regresní matice X . Podle námi provedené simulace je tak prvním pojistníkem v portfoliu muž ve věku 40 let s pojištěným nákladním automobilem. A například třetí pojistník je 28letá žena s pojištěným motocyklem.

interc.	žena	věk	OA	NA
1	0	40	0	1
1	0	61	1	0
1	1	28	0	0
1	0	63	0	0
1	0	59	1	0
...

Tabulka 4.9: Prvních pět řádků regresní matice X (Příklad 2).

Abychom mohli generovat pro každou pojistnou smlouvu počet škod a průměrnou výši škody, spočítali jsme nejprve parametry λ_i a μ_i pro všechny smlouvy $i = 1, \dots, 400$ podle (3.18), kde za vektory koeficientů $\alpha = (\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4)^T$ pro počet škod N_i a $\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)^T$ pro průměrnou výši škody \bar{Y}_i jsou použity hodnoty zvolené na začátku simulační části práce (viz Tabulka 4.1).

Pro simulaci našich marginálních modelů tedy pro každou pojistnou smlouvu i používáme

$$\begin{aligned} N_i &\sim \text{ZTP}(\lambda_i), & \lambda_i &= e^{x_i^T \alpha}, \\ \bar{Y}_i &\sim \text{gama}(\mu_i, \delta), & \mu_i &= e^{x_i^T \beta}, \end{aligned}$$

kde x_i , $i = 1, \dots, 400$ jsou řádky matice X a α a β jsou vektory zvolených regresních koeficientů pro počet škod a průměrnou výši škody (viz Tabulka 4.1).

Podle uvedených vzorců jsme pro každou pojistnou smlouvu pomocí generované matice X a koeficientů α a β nejprve vypočítali hodnoty parametrů λ_i a μ_i (parametr δ byl zvolen $\delta = 2$), které jsme dále použili pro generování počtu škod N_i a průměrné výše škody \bar{Y}_i pomocí Claytonovy kopuly se zvolenou hodnotou Kendallova tau τ a uvažovaných marginálních rozdělání (ZTP(λ_i) pro počet škod a gama(μ_i, δ) pro průměrnou výši škody). Dvojici (\bar{Y}_i, N_i) , kterou jsme získali popsáním postupem, jsme vygenerovali stokrát pro každého pojistníka v portfoliu a vyčíslili průměry.

Všechny výpočty i popsaná simulace pomocí kopuly byly provedeny v softwaru R (pomocí Yan (2007)). Z generovaných počtů škod a průměrných výší škod jsme poté odhadli metodou maximální věrohodnosti regresní koeficienty pomocí čtyř kopul, kterým jsme se věnovali v rámci teoretické části práce (Kapitola 3, část 3.2). Dále jsme odhadli koeficienty pro případ, kdy předpokládáme nezávislost mezi frekvencí a severitou škod ($\tau = 0$). Pro tento účel jsme použili nezávislou kopulu definovanou vztahem (1.5). Pro sdruženou hustotu náhodných veličin \bar{Y} a N v případě nezávislosti platí

$$f_{\bar{Y}, N}(y, n | \mu, \delta, \lambda) = f_{\bar{Y}}(y | \mu, \delta) \cdot f_N(n | \lambda),$$

kde $f_{\bar{Y}}$ je v našem případě hustota gama rozdělení a f_N hustota Poissonova rozdělení useknutého v nule.

Sdružené hustoty $f_{\bar{Y}, N}$ pro Claytonovu, Frankovu, Gaussovou a Gumbelovu kopulu jsou dány vzorcem (3.16) a parciálními derivacemi uvedenými v Tabulce 3.2.

Maximálně věrohodné odhady (MLE) koeficientů jsme získali maximalizací logaritmicke věrohodnostní funkce dle Krämer a kol. (2013)

$$\ell(v | y^*, n^*) = \sum_{i=1}^{400} \log(f_{\bar{Y}, N}(y_i, n_i | v)),$$

kde $y^* = (y_1, \dots, y_{400})^T$ a $n^* = (n_1, \dots, n_{400})^T$ jsou námi generované hodnoty průměrné výše škody a počtu škod pro všechny pojistné smlouvy a $v = (\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \delta, \theta)^T$ je vektor neznámých parametrů. MLE jsou pak

$$\hat{v} = \arg \max_v \ell(v | y^*, n^*).$$

Výpočet maximálně věrohodných odhadů byl proveden pomocí softwaru R optimalizační metodou BFGS (quasi-Newtonova metoda). Odhadnuté koeficienty pro počet škod $\alpha = (\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4)^T$ a pro průměrnou výši škody $\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)^T$ můžeme vidět v uvedených tabulkách pro různé hodnoty Kendallova tau vždy v porovnání se zvolenými regresními koeficienty. Data byla postupně generována Claytonovou kopulou s Kendallovým tau $\tau = 0,1$ (Tabulky 4.10 a 4.11), $\tau = 0,3$ (Tabulky 4.12 a 4.13) a $\tau = 0,5$ (Tabulky 4.14 a 4.15).

kopula	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$
Clayton	-0,934	0,375	0,039	0,504	0,289
Frank	-0,943	0,380	0,039	0,499	0,288
Gauss	-0,941	0,379	0,039	0,499	0,287
Gumbel	-0,253	0,316	0,033	0,432	0,236
nezávislá	-0,944	0,380	0,039	0,499	0,289
zvolené koef.	-1,000	0,400	0,040	0,500	0,300

Tabulka 4.10: Odhady koeficientů α při použití vybraných kopul pro Kendallovo tau $\tau = 0,1$ v porovnání se zvolenými hodnotami těchto koeficientů.

kopula	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
Clayton	-0,296	-0,830	-0,040	1,090	1,500
Frank	-0,285	-0,830	-0,040	1,090	1,500
Gauss	-0,285	-0,830	-0,040	1,090	1,500
Gumbel	-0,303	-0,827	-0,040	1,094	1,503
nezávislá	-0,283	-0,831	-0,040	1,090	1,500
zvolené koef.	-0,300	-0,800	-0,040	1,100	1,500

Tabulka 4.11: Odhady koeficientů β při použití vybraných kopul pro Kendallovo tau $\tau = 0,1$ v porovnání se zvolenými hodnotami těchto koeficientů.

kopula	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$
Clayton	-0,935	0,389	0,039	0,484	0,286
Frank	-0,950	0,392	0,040	0,478	0,281
Gauss	-0,955	0,393	0,040	0,481	0,278
Gumbel	-0,253	0,330	0,033	0,412	0,241
nezávislá	-0,935	0,387	0,039	0,484	0,289
zvolené koef.	-1,000	0,400	0,040	0,500	0,300

Tabulka 4.12: Odhady koeficientů α při použití vybraných kopul pro Kendallovo tau $\tau = 0,3$ v porovnání se zvolenými hodnotami těchto koeficientů.

kopula	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
Clayton	-0,280	-0,818	-0,041	1,132	1,509
Frank	-0,286	-0,816	-0,040	1,132	1,511
Gauss	-0,287	-0,815	-0,040	1,136	1,510
Gumbel	-0,282	-0,815	-0,040	1,136	1,510
nezávislá	-0,264	-0,819	-0,041	1,134	1,510
zvolené koef.	-0,300	-0,800	-0,040	1,100	1,500

Tabulka 4.13: Odhady koeficientů β při použití vybraných kopul pro Kendallovo tau $\tau = 0,3$ v porovnání se zvolenými hodnotami těchto koeficientů.

kopula	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$
Clayton	-0,897	0,379	0,039	0,469	0,254
Frank	-0,875	0,386	0,038	0,472	0,255
Gauss	-0,890	0,387	0,039	0,480	0,258
Gumbel	-0,104	0,321	0,031	0,390	0,208
nezávislá	-0,920	0,390	0,039	0,490	0,281
zvolené koef.	-1,000	0,400	0,040	0,500	0,300

Tabulka 4.14: Odhady koeficientů α při použití vybraných kopul pro Kendallovo tau $\tau = 0,5$ v porovnání se zvolenými hodnotami těchto koeficientů.

kopula	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
Clayton	-0,387	-0,781	-0,039	1,077	1,488
Frank	-0,361	-0,775	-0,039	1,079	1,486
Gauss	-0,362	-0,775	-0,039	1,078	1,486
Gumbel	-0,362	-0,776	-0,039	1,082	1,491
nezávislá	-0,338	-0,780	-0,039	1,075	1,490
zvolené koef.	-0,300	-0,800	-0,040	1,100	1,500

Tabulka 4.15: Odhady koeficientů β při použití vybraných kopul pro Kendallovo tau $\tau = 0,5$ v porovnání se zvolenými hodnotami těchto koeficientů.

Jak lze vidět, maximálně věrohodné odhady regresních koeficientů α a β většinou odpovídají zvoleným koeficientům. Nejvíce se liší pro všechny uvažované hodnoty τ odhad koeficientu α_0 v případě Gumbelovy kopuly, tedy intercept pro počet škod.

Z generovaných dat pomocí Claytonovy kopuly a tří různých hodnot Kendallova tau $\tau \in \{0,1; 0,3; 0,5\}$ jsme vypočítali maximálně věrohodné odhady regresních koeficientů pomocí Claytonovy, Frankovy, Gaussovy, Gumbelovy a nezávislé kopuly. Abychom však mohli porovnat, který model je pro generovaná data nejvhodnější, vyčíslili jsme pro všechny uvedené kopuly a hodnoty Kendallova tau Akaikeho informační kritérium (AIC, Akaike information criterion).

Akaikeho informační kritérium se počítá pro každý model samostatně a slouží pro porovnání více modelů. Preferujeme model s nejnižší hodnotou AIC. Akaikeho informační kritérium zohledňuje nejen hodnotu věrohodnostní funkce, ale také počet odhadovaných parametrů. Toto kritérium lze v našem případě vypočítat podle vzorce

$$AIC = 2k - 2\ell(\hat{v}|y^*, n^*),$$

kde k je počet odhadovaných parametrů modelu, $\ell(\hat{v}|y^*, n^*)$ je hodnota logaritmické věrohodnostní funkce v odhadnutých parametrech (tedy maximální hodnota logaritmické věrohodnostní funkce), \hat{v} je odhad vektorového parametru $v \in \mathbb{R}^k$ a $y^* \in \mathbb{R}^{400}$ a $n^* \in \mathbb{R}^{400}$ jsou generované hodnoty průměrné výše škody a počtu škod dříve popsáním způsobem.

V našem případě je počet parametrů k u Claytonovy, Frankovy, Gaussovy a Gumbelovy kopuly roven hodnotě 12 a u nezávislého modelu hodnotě 11, neboť zde nemáme parametr τ (Kendalovo tau) vyjadřující stupeň závislosti mezi frekvencí a severitou škod.

Akaikeho informační kritérium jsme spočítali pro modely využívající postupně všechny zmiňované kopuly včetně nezávislého modelu a pro tři hodnoty Kendallova tau $\tau \in \{0,1; 0,3; 0,5\}$. Výsledné hodnoty můžeme vidět v Tabulce 4.16.

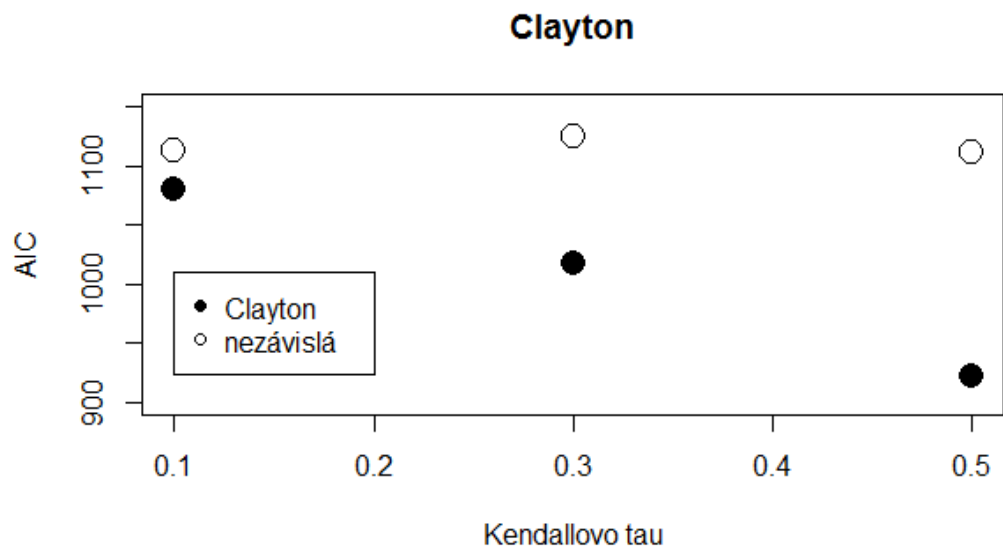
kopula	$\tau = 0,1$	$\tau = 0,3$	$\tau = 0,5$
Clayton	1080,3	1017,8	921,8
Frank	1097,8	1030,5	899,4
Gauss	1102,0	1042,3	900,6
Gumbel	1208,4	1141,1	1085,9
nezávislá	1113,2	1124,5	1111,1

Tabulka 4.16: Vypočítané Akaikeho informační kritérium pro uvažované kopuly a hodnoty Kendallova tau.

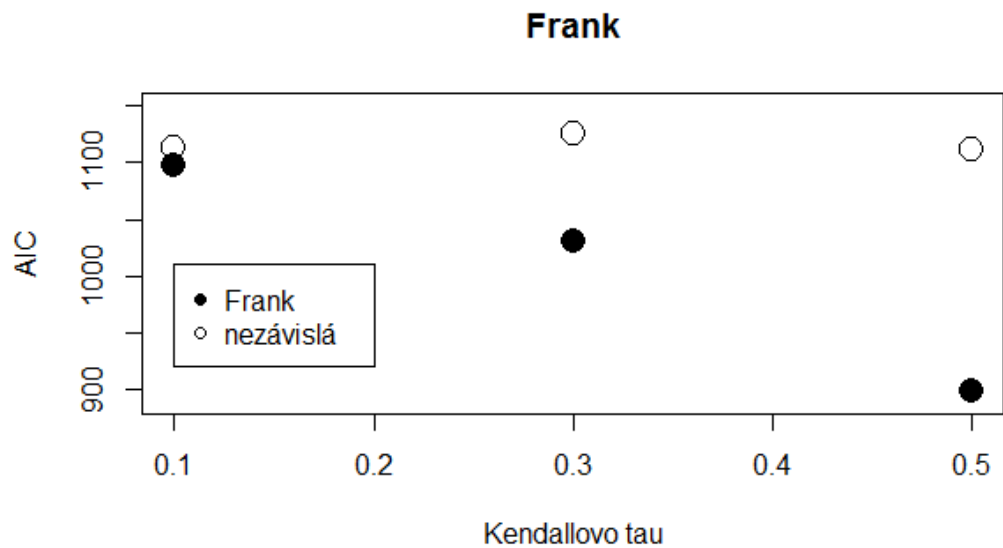
V případě generovaných dat s nižší uvažovanou závislostí ($\tau = 0,1$) je podle očekávání nejvhodnějším modelem ten, který využívá Claytonovu kopulu. Naopak nejméně vhodným modelem podle AIC je model s Gumbelovou kopulou, který je dokonce i méně vhodný než nezávislý model. Stejně krajní modely (nejlepší i nejhorší) byly zjištěny pomocí Akaikeho informačního kritéria pro případ $\tau = 0,3$. Pokud jsme při generování dat uvažovali Kendallovo tau rovno hodnotě 0,5, nejméně vhodným modelem se již dle očekávání stal model předpokládající nezávislost mezi frekvencí a severitou škod. Naopak nejvhodnějším modelem byl tentokrát model s Frankovou kopulou i přesto, že data byla generována pomocí Claytonovy kopuly.

Ve všech případech byla data generována pomocí modelu předpokládajícího závislost mezi počtem a výší škod s různým stupněm této závislosti $\tau \in \{0,1; 0,3; 0,5\}$. Použití nezávislého modelu na taková data je nekorektní volbou, a nemusí tedy správně odhadnout celkový úhrn škod. Abychom mohli porovnat použití nezávislého modelu oproti závislým modelům využívajícím kopuly, zobrazili jsme si grafy Akaikeho informačního kritéria pro tyto modely, jednotlivě pro uvažované kopuly (viz Obrázky 4.2, 4.3, 4.4, 4.5).

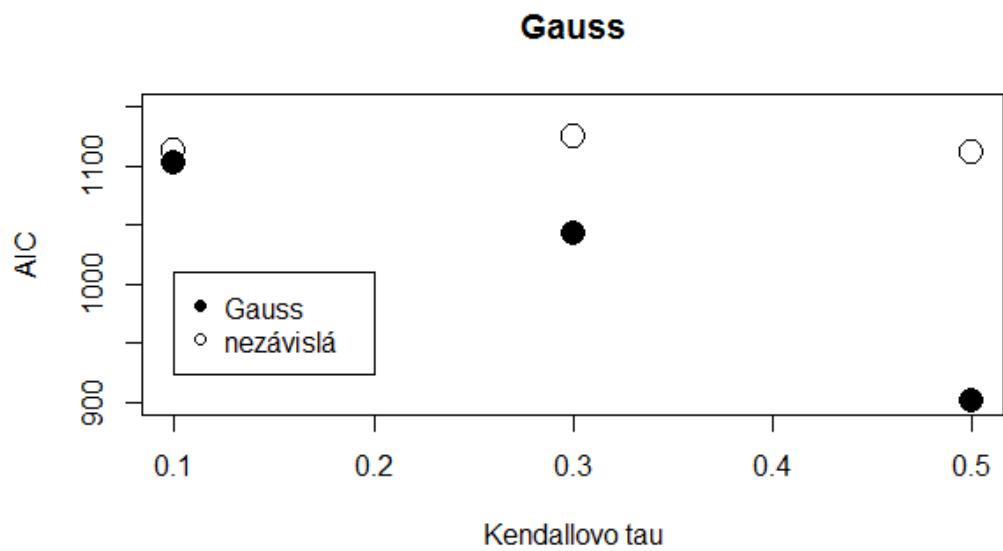
Pro Claytonovu, Frankovu a Gaussovu kopulu je zřejmé, že modely předpokládající závislost s využitím těchto kopul jsou korektnější volbou pro závislá data než model nezávislý. Také platí, že čím větší závislost frekvence a severity v datech uvažujeme, tím je nezávislý model méně vhodný oproti modelu předpokládajícímu možnou závislost. Gumbelova kopula se pro modelování závislosti na našich datech jeví jako nevhodná, neboť pro případy $\tau = 0,1$ a $\tau = 0,3$ dosahuje model s Gumbelovou kopulou nejvyšší hodnoty Akaikeho informačního kritéria a pouze pokud platí $\tau = 0,5$, má tento model nižší hodnotu AIC než model předpokládající nezávislost počtu a výší škod.



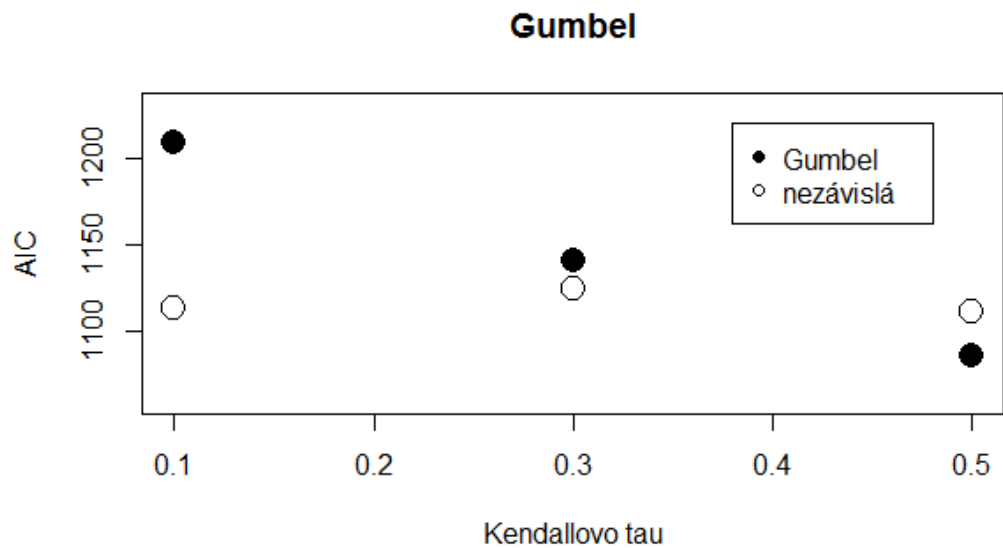
Obrázek 4.2: Akaikeho informační kritérium pro Claytonovu a nezávislou kopulu pro uvažované hodnoty Kendallova tau.



Obrázek 4.3: Akaikeho informační kritérium pro Frankovu a nezávislou kopulu pro uvažované hodnoty Kendallova tau.



Obrázek 4.4: Akaikeho informační kritérium pro Gaussovu a nezávislou kopulu pro uvažované hodnoty Kendallova tau.



Obrázek 4.5: Akaikeho informační kritérium pro Gumbelovu a nezávislou kopulu pro uvažované hodnoty Kendallova tau.

Závěr

Při modelování úhrnu škod v praxi je tradičně předpokládáno, že počet škod a výše škod jsou navzájem nezávislé. Tento předpoklad pak vede k jednoduchému vyjádření odhadu úhrnu škod. Avšak reálná data tuto nezávislost mohou porušovat. V práci jsme ukázali, že předpoklad nezávislosti může být vynechán.

Byly popsány dvě metody zahrnující závislost mezi frekvencí a severitou do modelu úhrnu škod a obě tyto metody jsme v rámci simulační části práce porovnali s klasickým modelem úhrnu škod předpokládajícím nezávislost počtu a výši škod.

V obou příkladech jsme simulovali hodnoty tarifních proměnných pro předem daný počet pojistníků v uvažovaném portfoliu a určili jsme si hodnoty regresních koeficientů potřebných pro vygenerování dat. V prvním příkladě jsme generovali počet a výše škod pro uvažované pojistné smlouvy pomocí závislého modelu využívajícího zobecněné lineární modely pro pět různých stupňů závislosti. Na tato data jsme poté použili modely se závislou i nezávislou frekvencí a severitou a odhadli hodnoty regresních koeficientů pro počet škod a průměrnou výši škody. Dále jsme ukázali, že při použití nekorektního nezávislého modelu dochází pro dané stupně závislosti k podhodnocení odhadů úhrnu škod a toto podhodnocení roste s rostoucím stupněm závislosti. Průměrný procentální rozdíl těchto odhadů na pojistné smlouvě byl pro stupeň závislosti 0,3 vyšší než 32%.

Metodu zahrnutí závislosti počtu a výši škod pomocí kopul jsme ukázali ve druhém příkladě. Pro generování dat jsme si vybrali Claytonovu kopulu a vygenerovali jsme počet škod a průměrnou výši škody pro 400 pojistníků pro tři různé hodnoty Kendallova tau τ . Poté jsme metodou maximální věrohodnosti odhadli regresní koeficienty počtu škod a průměrné výše škody pomocí modelů s Claytonovou, Frankovou, Gaussovou, Gumbelovou a také s nezávislou kopulou představující model, který předpokládá nezávislost mezi frekvencí a severitou škod. Pro porovnání uvedených modelů jsme spočítali Akaikeho informační kritérium pro všechny modely s využitím generovaných dat a porovnali modely uvažující závislost frekvence a severity škod s nezávislým modelem. Pro naše data se jeví Gumbelova kopula jako nevhodná, neboť pro dvě hodnoty Kendallova tau dosáhl model s touto kopulou vyššího Akaikeho informačního kritéria než model nezávislý. Zbylé tři kopuly jsou již dle očekávání vhodnější volbou než model předpokládající nezávislost frekvence a severity škod pro všechny hodnoty Kendallova tau. Navíc s vyšší závislostí dat (vyšší τ) roste rozdíl AIC mezi nezávislým a závislým modelem.

Pokud máme k dispozici data s určitou závislostí mezi počtem a výší škod, doporučuji volit pro tato data jeden z představených modelů v této práci, neboť použití modelu předpokládajícího nezávislost frekvence a severity škod není korektní volbou a odhady úhrnu škod nemusí být správné.

Seznam použité literatury

- DOBSON, A. J. (2002). *An Introduction to Generalized Linear Models*. Second Edition. CRC Press. ISBN 1-58488-165-8.
- FREES, E. W., GAO, J. a ROSENBERG, M. A. (2011). Predicting the Frequency and Amount of Health Care Expenditures. *North American Actuarial Journal*, **15**(3), 377–392.
- GARRIDO, J., GENEST, C. a SCHULZ, J. (2016). Generalized linear models for dependent frequency and severity of insurance claims. *Insurance: Mathematics and Economics*, **70**, 205–215.
- KLUGMAN, S. A., PANJER, H. H. a WILLMOT, G. E. (2004). *Loss models: from data to decisions*. Second Edition. John Wiley & Sons, New Jersey. ISBN 0-471-21577-5.
- KRÄMER, N., BRECHMANN, E. C., SILVESTRINI, D. a CZADO, C. (2013). Total loss estimation using copula-based regression models. *Insurance: Mathematics and Economics*, **53**(3), 829–839.
- MCCULLAGH, P. a NELDER, J. A. (1989). *Generalized Linear Models*. Second Edition. Springer, London. ISBN 978-0-412-3760-6.
- NELSEN, R. B. (2007). *An Introduction to Copulas*. Springer, New York. ISBN 0-387-98623-5.
- OHLSSON, E. a JOHANSSON, B. (2010). *Non-Life Insurance Pricing with Generalized Linear Models*. EAA Lecture Notes. Springer, Berlin. ISBN 978-3-642-10790-0.
- SHI, P., FENG, X. a IVANTSOVA, A. (2015). Dependent frequency-severity modeling of insurance claims. *Insurance: Mathematics and Economics*, **64**, 417–428.
- YAN, J. (2007). Enjoy the Joy of Copulas: With a Package copula. *Journal of Statistical Software*, **21**(4), 1–21.

Seznam obrázků

4.1	Střední hodnoty úhrnu škod, odhady úhrnu škod s použitím odhadnutých regresních koeficientů pomocí závislého a nezávislého modelu pro zvolené hodnoty stupně závislosti θ v grafu.	31
4.2	Akaikeho informační kritérium pro Claytonovu a nezávislou kopulu pro uvažované hodnoty Kendallova tau.	39
4.3	Akaikeho informační kritérium pro Frankovu a nezávislou kopulu pro uvažované hodnoty Kendallova tau.	39
4.4	Akaikeho informační kritérium pro Gaussovu a nezávislou kopulu pro uvažované hodnoty Kendallova tau.	40
4.5	Akaikeho informační kritérium pro Gumbelovu a nezávislou kopulu pro uvažované hodnoty Kendallova tau.	40

Seznam tabulek

3.1	Rozsah parametru θ a vztah Kendallova tau τ a parametru θ pro vybrané rodiny kopul (zdroj: Krämer a kol. (2013)).	19
3.2	První parciální derivace $D_1(u_1, u_2 \theta)$ pro vybrané rodiny kopul (zdroj: Krämer a kol. (2013)).	19
3.3	Parametry, střední hodnota a rozptyl dvou uvažovaných marginálních rozdělání pro náhodné veličiny \bar{Y} a N (zdroj: Krämer a kol. (2013)).	21
4.1	Regresní koeficienty pro počet škod a průměrnou výši škody.	24
4.2	Prvních pět řádků regresní matice X (Příklad 1).	25
4.3	Zprůměrované hodnoty odhadnutých regresních koeficientů α pro počet škod N	26
4.4	Zprůměrované hodnoty odhadnutých regresních koeficientů β pro průměrnou výši škody \bar{Y} při použití modelu úhrnu škod s nezávislou frekvencí a severitou.	27
4.5	Zprůměrované hodnoty odhadnutých regresních koeficientů β pro průměrnou výši škody \bar{Y} při použití modelu úhrnu škod se závislou frekvencí a severitou.	27
4.6	Střední hodnoty úhrnu škod pro různé stupně závislosti θ při předpokládané závislosti mezi počtem a výší škod v porovnání s průměrnými součty vygenerovaných škod z provedených simulací.	28
4.7	Střední hodnoty úhrnu škod, odhady úhrnu škod s použitím odhadnutých regresních koeficientů pomocí závislého a nezávislého modelu pro zvolené hodnoty stupně závislosti θ	30
4.8	Průměrný procentuální rozdíl odhadů úhrnů škod na pojistné smlouvě mezi závislým a nezávislým modelem pro uvažované hodnoty stupně závislosti θ	30
4.9	Prvních pět řádků regresní matice X (Příklad 2).	32
4.10	Odhady koeficientů α při použití vybraných kopul pro Kendallovo tau $\tau = 0,1$ v porovnání se zvolenými hodnotami těchto koeficientů.	34
4.11	Odhady koeficientů β při použití vybraných kopul pro Kendallovo tau $\tau = 0,1$ v porovnání se zvolenými hodnotami těchto koeficientů.	34
4.12	Odhady koeficientů α při použití vybraných kopul pro Kendallovo tau $\tau = 0,3$ v porovnání se zvolenými hodnotami těchto koeficientů.	35
4.13	Odhady koeficientů β při použití vybraných kopul pro Kendallovo tau $\tau = 0,3$ v porovnání se zvolenými hodnotami těchto koeficientů.	35
4.14	Odhady koeficientů α při použití vybraných kopul pro Kendallovo tau $\tau = 0,5$ v porovnání se zvolenými hodnotami těchto koeficientů.	36
4.15	Odhady koeficientů β při použití vybraných kopul pro Kendallovo tau $\tau = 0,5$ v porovnání se zvolenými hodnotami těchto koeficientů.	36
4.16	Vypočítané Akaikeho informační kritérium pro uvažované kopuly a hodnoty Kendallova tau.	37