

**MATEMATICKO-FYZIKÁLNÍ
FAKULTA**
Univerzita Karlova

DIPLOMOVÁ PRÁCE

Lucie Zemánková

Vícerozměrné modely počtů škod

Katedra pravděpodobnosti a matematické statistiky

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

Studijní program: Matematika

Studijní obor: Finanční a pojistná matematika

Praha 2019

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 Praze dne 17. 7. 2019

Na tomto místě bych chtěla poděkovat vedoucí mé diplomové práce, RNDr. Lucii Mazurové, Ph.D., za pomoc při vypracovávání práce a ochotu poskytnout mi cenné rady při konzultacích. Dále bych chtěla poděkovat rodičům za finanční i morální podporu během mého celého studia a v neposlední řadě můj dík patří mému životnímu partneru Tomášovi, který mi byl velkou oporou při studiu.

Název práce: Vícerozměrné modely počtů škod

Autor: Lucie Zemánková

Katedra: Katedra pravděpodobnosti a matematické statistiky

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

Abstrakt: Vícerozměrné modely počtů škod je možné využít při modelování počtů škod z různých odvětví, které mohou být navzájem provázány závislostní strukturou. Stejně jako v případě jednorozměrných počtů škod se k modelování převážně využívá Poissonova a negativně binomického rozdělení, které je rozšířeno do dalších dimenzí. Zobecnění rozdělení pro více rozměrů se často provádí pomocí tzv. šokových proměnných, kdy je jedna náhodná veličina obsažena ve všech rozměrech náhodného vektoru modelujícího počty škod. Komplexnějším přístupem k modelování závislostí je modelování pomocí kopulí. Porovnání těchto modelů je provedeno na simulovaném příkladu počtů škod ze dvou různých garancí autopojištění.

Klíčová slova: dvojrozměrné Poissonovo rozdělení, počty škod, negativně binomické rozdělení, kopule.

Title: Multivariate claim numbers models

Author: Lucie Zemánková

Department: Department of Probability and Mathematical Statistics

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

Abstract: Multidimensional frequency models can be used for modeling number of claims from different branches which are somehow dependent on each other. As in the one-dimensional case Poisson distribution and negative binomial distribution are primarily used for modeling multidimensional claim counts data, only they are extended to higher dimensions. The generalization of multi-dimensional distributions is often done using so-called shock variables, where one random variable is included in all dimensions of a random vector which models claim counts. The more comprehensive approach to modeling dependence uses copulas. Comparison of these models is done on a simulated data of number of claims from two different car insurance guarantees.

Keywords: bivariate Poisson distribution, claim frequency, negative binomial distribution, copulas.

Obsah

Úvod	3
1 Poissonovské modely	5
1.1 Poissonovo rozdělení	5
1.2 V nule modifikované Poissonovo rozdělení	5
1.3 Odhady parametrů	7
1.4 Dvourozměrné poissonovské modely	8
1.4.1 V nule modifikované dvourozměrné Poissonovo rozdělení	10
1.4.2 Na diagonále modifikované dvourozměrné Poissonovo rozdělení	11
1.5 Trojrozměrné poissonovské modely	11
1.5.1 Model se společnou kovariancí	12
1.5.2 Model s úplnou kovariancí	13
1.5.3 V nule modifikované trojrozměrné modely	16
2 Negativně binomické modely	17
2.1 Negativně binomické rozdělení	17
2.2 Trojrozměrné negativně binomické modely	17
2.2.1 Model NB-I	18
2.2.2 Model NB-II	18
2.2.3 Trojrozměrné negativně binomické rozdělení se společnou kovariancí	19
2.2.4 Trojrozměrné negativně binomické rozdělení s úplnou kovariancí	20
3 Modely s kopulemi	23
3.1 Základní kopule	23
3.1.1 Gaussovské kopule	24
3.1.2 Archimédovské kopule	25
3.2 Míry závislosti	27
3.2.1 Pearsonův korelační koeficient	27
3.2.2 Pořadové korelace	28
3.2.3 Koeficienty závislosti chvostů	29
3.3 Odhad parametrů kopulí	30
3.3.1 Pořadové korelace	30
3.3.2 Maximálně věrohodný odhad	30
4 Porovnání modelů na simulovaných datech	31
4.1 Generování dat	31
4.2 Porovnání poissonovských modelů na datech	34
4.3 Modelování pomocí kopulí	37
Závěr	39
Seznam použité literatury	40

Seznam obrázků	41
Seznam tabulek	42
Seznam použitých zkratk	43

Úvod

Modelování počtů škod a stanovení výše pojistného za pojistnou smlouvu spolu velmi úzce souvisí. Pojišťovna pro určení ceny za pojištění musí dokázat odhadnout, kolik bude muset klientovi vyplatit na škodách souvisejících s danou pojistnou smlouvou. Proto se modelují počty škod pro parametry dané pojistné smlouvy. Pro různé druhy pojištění se liší požadované údaje o pojištěných osobách nebo věcech. Při pojištění smrti v rámci životního pojištění potřebuje pojišťovna znát například věk, trvalé bydliště, výšku, váhu, zaměstnání pojištěného nebo zda pojištěná osoba kouří. Při pojištění domu nebo bytu je zase potřeba znát přesnou adresu jednotky pro určení nebezpečí plynoucích z povodní nebo velkého větru. V této práci se budeme ale především zabývat autopojištěním. V tomto odvětví se zjišťují jednak údaje o pojištěném vozidle – např. značka automobilu, rok výroby, zabezpečení vozu, VIN kód, a jednak údaje o osobě, která vozidlo vlastní.

S pomocí zjištěných údajů je pojistná smlouva zařazena do základní třídy, ve které se vyskytují smlouvy se stejnou rizikovostí (předpokládá se pro tyto smlouvy stejné rozdělení počtů škod). Pro tyto třídy je pomocí matematických metod odhadnut očekávaný počet škod a z toho určena výše základního pojistného, kterému se říká *apriorní pojistné*. Apriorní pojistné je určeno z údajů známých při sjednání smlouvy a je stejné pro všechny pojistky ve stejné třídě.

Ne všechny údaje jsou ale při sjednání smlouvy známe, případně je obtížné je kvantifikovat, jako například schopnosti řidiče včas zareagovat, styl jeho jízdy nebo znalost dopravních předpisů. Jako zastoupení těchto ukazatelů se využívá informace, která je známá pro všechny řidiče, a to počet předchozích škod, které eviduje Česká kancelář pojistitelů. Na základě těchto dodatečných informací je pomocí systému Bonus-Malus stanoveno finální pojistné, kterému se říká *aposteriorní pojistné*.

Důležitá část při tvorbě pojistného je odhad počtů škod. Ten nejčastěji probíhá pomocí zobecněných lineárních modelů, kde jsou smlouvy z portfolia rozděleny do homogenních tříd, pro které se modeluje očekávaný počet škod jako funkce popisujících faktorů. Pro modely počtů škod se často používá Poissonův regresní model, kde počty škod mají Poissonovo rozdělení.

V rámci autopojištění jsou zahrnuty různé typy garancí – povinné ručení, havarijní pojištění, a to se dále může dělit například na pojištění skel, střet se zvěří, pojištění proti přírodním událostem a mnoho dalších. Pro každou z těchto garancí je modelován počet škod zvlášť, ale ne vždy je při tarifování zohledněn vzájemný závislostní vztah mezi jednotlivými garancemi. Střední hodnota celkového počtu škod není ovlivněna případnou závislostí mezi různými garancemi. Pro zachycení rizikovosti tarifované skupiny smluv se může použít bezpečnostní přírážka, o jejíž násobek je navýšen očekávaný počet škod. Bezpečnostní přírážka je často popsána pomocí rozptylu, směrodatné odchylky nebo nějaké jejich kombinace. Rozptyl součtu počtů škod se ale již liší pro případ, kdy jsou složky vektorů závislé a nezávislé, a je tedy potřeba vzájemnou závislost mezi složkami brát v potaz.

Ne všechny faktory, které jsou použity v zobecněných lineárních modelech mohou být měřitelné. Je tedy možné, že riziko škody nebude dostatečně podchyceno a třídy, o kterých předpokládáme, že jsou homogenní, budou ve skutečnosti

heterogenní. Což je možné napravit například zavedením Bonus-Malus systému.

Dalším problémem, který není možné zcela podchytit pomocí Poissonova regresního modelu, je problém overdispersione – rozptyl počtů škod je větší než jejich střední hodnota. Vzhledem k vlastnostem Poissonova rozdělení – jak je uvedeno v oddílu 1.1, není možné podchytit větší rozptyl než střední hodnotu pomocí Poissonova rozdělení. Proto se využívá například v nule modifikované Poissonova rozdělení, viz oddíl 1.2.

Poslední problém, který je třeba zmínit, je předpoklad vzájemné nezávislosti garancí. Je pádný důvod očekávat, že předpoklad nezávislosti splněn nebude. Stane-li se dopravní nehoda mezi dvěma vozidly, tak má-li řidič, který dopravní nehodu zavinil, pojištění navíc k povinnému ručení i havarijní pojištění, tak se bude plnit z obou těchto garancí. Z povinného ručení se bude platit škoda způsobená druhému vozidlu a z havarijního pojištění se bude platit oprava vlastního vozu. Vzniknou takto v podstatě dvě škody na dvou různých garancích, které se ovšem vážou k jedné dopravní nehodě. Pro zachycení reality je tedy předpoklad nezávislosti potřeba rozvolnit.

V této práci se budeme zabývat právě rozvolněním požadavku nezávislosti mezi různými garancemi, neboli mezi složkami vícerozměrného náhodného vektoru. Nejprve si zavedeme v oddílu 1.4 model dvourozměrného Poissonova rozdělení, u kterého je kladná závislost mezi složkami dvourozměrného náhodného vektoru vyjádřena pomocí jednoho z parametrů rozdělení. Toto rozdělení budeme dále modifikovat v odstavcích 1.4.1 a 1.4.2, abychom byli schopni zachytit overdispersioni. Rozšíření pro trojrozměrné náhodné vektory zavedeme v oddílu 1.5. Dále v kapitole 2 budeme prezentovat negativně binomické modely rozšířené do tří dimenzí. V kapitole 3 zachytíme závislostní strukturu mezi jednotlivými garancemi pomocí kopulí. Nakonec některé teoretické výsledky budeme v kapitole 4 demonstrovat na simulovaném vzorku dat.

1. Poissonovské modely

Počty škod z nějakého pojištění jsou náhodné veličiny, protože nikdy není dopředu jasné, kolik škod nastane v daném portfoliu. Pojišťovny však musí, mimo jiné kvůli určení výše pojistného nebo nutnosti výpočtu rezerv, modelovat počty škod v daném portfoliu dopředu v čase. K modelování počtů škod se používají diskrétní rozdělení. Za předpokladu nezávislosti škod se nejčastěji využívá Poissonovo nebo negativně-binomické rozdělení. Základní charakteristiky jednorozměrného Poissonova rozdělení jsou shrnuty v oddílu 1.1. Jelikož Poissonovo rozdělení má stejnou střední hodnotu jako rozptyl, není možné zachytit případ, kdy je rozptyl větší než střední hodnota (overdisperte). Je možné využít rozšíření Poissonova rozdělení na v nule modifikované, kdy hodnotě 0 je přiřazena větší pravděpodobnost, než by měla být dle Poissonova rozdělení.

Předpoklad nezávislosti škod je často nesplněn, protože při jedné škodě dochází k plnění z různých druhů garancí. Například na škodách z autopojištění je v mnoha případech pozitivní závislost mezi škodami z povinného ručení a ostatním krytím (havarijní pojištění, krytí skel, atd.)

Budeme chtít rozvolnit předpoklad nezávislosti, proto v oddílu 1.4 zavedeme model dvourozměrného Poissonova rozdělení. Díky tomuto rozdělení budeme moci zachytit pozitivní závislost mezi počty škod ze dvou různých garancí na jedné pojistce. Dále stejně jako v jednorozměrném případě zavedeme modifikace dvourozměrného Poissonova rozdělení na v nule modifikované rozdělení a na diagonále modifikované rozdělení.

V oddílu 1.5 dvourozměrný případ rozšíříme i pro trojrozměrné náhodné veličiny, kde uvedeme dva typy modelů – v části 1.5.1 model se společnou kovariancí a v části 1.5.2 model s úplnou kovariancí. Pro model s úplnou kovariancí si v části 1.5.3 uvedeme jeho rozšíření modifikací v nule.

1.1 Poissonovo rozdělení

Jednorozměrné Poissonovo rozdělení je používáno pro modelování četnosti jevů, které nastaly v daném časovém intervalu. Jedná se o diskrétní rozdělení.

Definice 1. *Náhodná veličina N má Poissonovo rozdělení, pokud pro všechna $k = 0, 1, 2, \dots$ a $\lambda > 0$ platí*

$$P(N = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Pravděpodobnost $P(N = k)$ pro Poissonovo rozdělení budeme označovat $P_{PD}(k)$, kde písmena PD jsou zkratkou anglického názvu *Poisson distribution*.

Základní charakteristiky Poissonova rozdělení jsou shrnuty v tabulce 1.1, kde je vidět, že střední hodnota a rozptyl Poissonova rozdělení jsou shodné.

1.2 V nule modifikované Poissonovo rozdělení

Jak je uvedeno v Klugman (1998), v nule modifikované rozdělení je možné zavést dvěma způsoby, buď pomocí tzv. třídy rozdělení $(a, b, 1)$, nebo jako směs

Základní charakteristika	Hodnota
střední hodnota	λ ,
rozptyl	λ ,
šikmost	$1/\sqrt{\lambda}$,
špičatost	$1/\lambda$.

Tabulka 1.1: Základní charakteristiky Poissonova rozdělení.

rozdělení z třídy $(a, b, 0)$ a degenerovaného rozdělení soustředěného v nule.

Definice 2. Diskrétní rozdělení s pravděpodobnostní funkcí $P(k)$ patří do třídy $(a, b, 0)$, pokud existují reálné konstanty a, b takové, že pro $k = 1, 2, 3, \dots$ platí

$$\frac{P(k)}{P(k-1)} = a + \frac{b}{k}. \quad (1.1)$$

Hodnota pravděpodobnosti $P(0)$ je jednoznačně určena z podmínky

$$\sum_{k=0}^{\infty} P(k) = 1.$$

Definice 3. Diskrétní rozdělení s pravděpodobnostní funkcí $P(k)$ patří do třídy $(a, b, 1)$, pokud existují reálné konstanty a, b takové, že pro $k = 2, 3, 4, \dots$ platí

$$\frac{P(k)}{P(k-1)} = a + \frac{b}{k}. \quad (1.2)$$

Hodnota pravděpodobnosti $P(0)$ může být určena libovolně z intervalu $[0, 1)$, hodnota $P(1)$ je následně jednoznačně určena z podmínky

$$\sum_{k=1}^{\infty} P(k) = 1 - P(0).$$

V případě, že hodnota $P(0)$ je rovná nule, vznikne v nule useknuté rozdělení. Takové rozdělení se používá, pokud neočekáváme, že se v datech budou objevovat nulové hodnoty, což většinou není případ dat týkajících se počtů škod v portfoliu. Pokud je pravděpodobnost $P(0)$ větší než nula, potom se jedná o v nule modifikované rozdělení.

Poissonovo rozdělení patří do třídy $(a, b, 0)$, protože splňuje vztah (1.1) s konstantami $a = 0, b = \lambda$:

$$\frac{P_{PD}(k)}{P_{PD}(k-1)} = \frac{\frac{\lambda^k}{k!} e^{-\lambda}}{\frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda}} = \frac{\lambda}{k}.$$

V nule modifikované Poissonovo rozdělení je možné zavést také pomocí Poissonova rozdělení a degenerovaného rozdělení soustředěného v nule. Pravděpodobnostní funkce v nule modifikovaného Poissonova rozdělení má následující tvar

$$\begin{aligned} P(0) &= p + (1-p)P_{PD}(0) = p + (1-p)e^{-\lambda}, \\ P(k) &= (1-p)P_{PD}(k) = (1-p)\frac{\lambda^k}{k!}e^{-\lambda}, \quad k = 1, 2, \dots, p \in (0, 1). \end{aligned}$$

Základní charakteristiky v nule modifikovaného Poissonova rozdělení jsou shrnuty v tabulce 1.2. Rozptyl v nule modifikovaného Poissonova rozdělení je možné vyjádřit pomocí jeho střední hodnoty, označíme-li ji μ , jako $\mu + \frac{p}{1-p}\mu^2$. Odtud je vidět, že rozptyl je vždy v případě v nule modifikovaného rozdělení větší než střední hodnota, protože $\frac{p}{1-p}\mu^2$ je vždy větší než nula.

Základní charakteristika	Hodnota
střední hodnota	$(1 - p) \lambda$,
rozptyl	$(1 - p) \lambda + p(1 - p) \lambda^2$.

Tabulka 1.2: Základní charakteristiky v nule modifikovaného Poissonova rozdělení.

1.3 Odhady parametrů

V této části si ukážeme postup odhadování parametrů, jak je uveden například v AlMuhayfith a kol. (2016).

Pro odhad parametrů rozdělení náhodné veličiny, která popisuje počty, je možné použít Poissonovu regresi. Obecně je Poissonovu regresi možné použít na vzájemně nezávislé náhodné vektory \mathbf{N}_i , pro jejichž složky N_{ki} platí, že mají Poissonovo rozdělení s parametry λ_{ki} , tedy jsou charakterizovány pravděpodobností

$$P[N_{ki} = n_{ki} | \mathbf{x}_i^k] = \frac{e^{-\lambda_{ki}} \lambda_{ki}^{n_{ki}}}{n_{ki}!},$$

kde \mathbf{x}_i^k jsou vektory nezávisle proměnných. Index i označuje i -té pozorování, nabývá hodnot $1, \dots, I$, kde I je celkový počet pozorování. Index k označuje k -té složky ve vektorech \mathbf{N}_i .

Dále předpokládáme, že parametry rozdělení jsou modelovány pomocí nezávislých proměnných \mathbf{x}_i^k a vektoru regresních koeficientů $\boldsymbol{\beta}_k$. Použijeme logaritmickou linkovou funkci a dostaneme vyjádření

$$\log \lambda_{ki} = \mathbf{x}_i^k \top \boldsymbol{\beta}_k, \quad i = 1, \dots, I; \quad k = 1, \dots, K.$$

Pro modelování různých parametrů λ_{ki} mohou být použity odlišné vysvětlující proměnné, proto jsou jednotlivé vysvětlující proměnné označeny navíc horním indexem k , aby byla zdůrazněna příslušnost ke k -tému parametru rozdělení.

Odhad parametru $\widehat{\lambda}_{ki}$ získáme jako $\exp\{\mathbf{x}_i^k \top \widehat{\boldsymbol{\beta}}_k\}$, kde $\widehat{\boldsymbol{\beta}}_k$ jsou odhady parametrů $\boldsymbol{\beta}_k$ pomocí metody maximální věrohodnosti. Při této metodě se využívá hledání argumentu maxima věrohodnostní funkce. Věrohodnostní funkce $L(\boldsymbol{\theta})$ má za argument vektor neznámých parametrů daného rozdělení $\boldsymbol{\theta}$ a je definována jako sdružená distribuční funkce pozorovaných veličin.

Častěji se využívá logaritmus věrohodnostní funkce označovaný jako $l(\boldsymbol{\theta})$. Argument maxima logaritmu věrohodnostní funkce je možné nalézt položením derivací podle neznámých parametrů rovných nule.

1.4 Dvourozměrné poissonovovské modely

Dvourozměrné Poissonovo rozdělení je rozšířením Poissonova rozdělení pro dvourozměrné náhodné vektory, jejichž složky nemusí být vzájemně nezávislé. Toto rozdělení můžeme definovat, jak je uvedeno v Bermudez (2009), jednak pomocí sdružené pravděpodobnostní funkce, jako v definici 4, nebo pomocí tří nezávislých veličin s Poissonovým rozdělením.

Definice 4. *Náhodný vektor (N_1, N_2) má dvourozměrné Poissonovo rozdělení, pokud pro všechna $n_1, n_2 = 0, 1, 2, \dots$ a $\lambda_0, \lambda_1, \lambda_2 > 0$ je sdružená pravděpodobnostní funkce vektoru (N_1, N_2) dána vztahem*

$$P(N_1 = n_1, N_2 = n_2) = e^{-(\lambda_0 + \lambda_1 + \lambda_2)} \frac{\lambda_1^{n_1}}{n_1!} \frac{\lambda_2^{n_2}}{n_2!} \sum_{k=0}^{\min(n_1, n_2)} \binom{n_1}{k} \binom{n_2}{k} k! \left(\frac{\lambda_0}{\lambda_1 \lambda_2} \right)^k. \quad (1.3)$$

Pravděpodobnost $P(N_1 = n_1, N_2 = n_2)$ označíme jako $P_{BPD}(n_1, n_2)$, kde písmena v dolním indexu jsou zkratkou anglického názvu *Bivariate Poisson distribution*.

Dvourozměrné Poissonovo rozdělení má náhodný vektor $\mathbf{N} = (N_1, N_2)^\top$, který je po složkách generovaný jako součet náhodných veličin s Poissonovým rozdělením Y_0, Y_1, Y_2 , které jsou vzájemně nezávislé,

$$\begin{aligned} N_1 &= Y_1 + Y_0, \\ N_2 &= Y_2 + Y_0. \end{aligned} \quad (1.4)$$

Náhodné veličiny Y_0, Y_1, Y_2 mají Poissonovo rozdělení postupně s parametry $\lambda_0, \lambda_1, \lambda_2$. Marginální distribuce N_1 , resp. N_2 , jsou Poissonovy s parametrem $\lambda_1 + \lambda_0$, resp. $\lambda_2 + \lambda_0$. Náhodná veličina Y_0 je také někdy označována jako společný šok, protože je o ni navýšena hodnota náhodné veličiny s Poissonovým rozdělením v obou dimenzích vektoru \mathbf{N} .

Ukážeme, že náhodný vektor \mathbf{N} má sdruženou pravděpodobnostní funkci (1.3).

$$\begin{aligned} P(N_1 = n_1, N_2 = n_2) &= P(Y_1 + Y_0 = n_1, Y_2 + Y_0 = n_2) = \\ &= \sum_{k=0}^{\min(n_1, n_2)} P(Y_1 + Y_0 = n_1, Y_2 + Y_0 = n_2 | Y_0 = k) P(Y_0 = k) = \\ &= e^{-\lambda_0} \sum_{k=0}^{\min(n_1, n_2)} \frac{\lambda_0^k}{k!} P(Y_1 = n_1 - k, Y_2 = n_2 - k) = \\ &= e^{-\lambda_0} \sum_{k=0}^{\min(n_1, n_2)} \frac{\lambda_0^k}{k!} P(Y_1 = n_1 - k) P(Y_2 = n_2 - k) = \\ &= e^{-(\lambda_0 + \lambda_1 + \lambda_2)} \sum_{k=0}^{\min(n_1, n_2)} \frac{\lambda_0^k}{k!} \frac{\lambda_1^{n_1 - k}}{(n_1 - k)!} \frac{\lambda_2^{n_2 - k}}{(n_2 - k)!} = \\ &= e^{-(\lambda_0 + \lambda_1 + \lambda_2)} \lambda_1^{n_1} \lambda_2^{n_2} \sum_{k=0}^{\min(n_1, n_2)} \left(\frac{\lambda_0}{\lambda_1 \lambda_2} \right)^k \frac{1}{(n_1 - k)! (n_2 - k)! k!} = \\ &= e^{-(\lambda_0 + \lambda_1 + \lambda_2)} \frac{\lambda_1^{n_1}}{n_1!} \frac{\lambda_2^{n_2}}{n_2!} \sum_{k=0}^{\min(n_1, n_2)} \binom{n_1}{k} \binom{n_2}{k} k! \left(\frac{\lambda_0}{\lambda_1 \lambda_2} \right)^k, \end{aligned}$$

kde jsme využili vzájemnou nezávislost náhodných veličin Y_0, Y_1, Y_2 .

Definice dvourozměrného Poissonova rozdělení pomocí vztahu (1.4) je vhodnější pro generování dvourozměrných dat se vzájemnou pozitivní závislostí. Kovarianci mezi složkami dvourozměrného vektoru $\begin{pmatrix} N_1 \\ N_2 \end{pmatrix}$ je možné vyjádřit díky definici veličin N_1, N_2 jako

$$\text{Cov}(N_1, N_2) = \text{Cov}(Y_1 - Y_0, Y_2 - Y_0) = \text{Cov}(Y_0, Y_0) = \text{Var}(Y_0) = \lambda_0,$$

z čehož je patrné, že parametr λ_0 z rozdělení náhodné veličiny Y_0 reprezentuje míru závislosti náhodných veličin N_1 a N_2 . Pokud je parametr λ_0 roven 0, náhodné veličiny N_1 a N_2 jsou nekorelované. Navíc má náhodná veličina Y_0 nulový rozptyl, čili je skoro jistě rovna 0, a tím pádem jsou náhodné veličiny N_1 a N_2 nezávislé, protože náhodné veličiny Y_1 a Y_2 jsou nezávislé.

V tabulce 1.3 jsou uvedeny základní charakteristiky dvourozměrného Poissonova rozdělení.

Základní charakteristika	Hodnota
vektor středních hodnot	$\begin{pmatrix} \lambda_1 + \lambda_0 \\ \lambda_2 + \lambda_0 \end{pmatrix},$
rozptylová matice	$\begin{pmatrix} \lambda_1 + \lambda_0 & \lambda_0 \\ \lambda_0 & \lambda_2 + \lambda_0 \end{pmatrix}.$

Tabulka 1.3: Základní charakteristiky dvourozměrného Poissonova rozdělení.

Odhad parametrů modelu

Pro náhodné veličiny s dvourozměrným Poissonovým rozdělením můžeme pro odhad parametrů rozdělení použít Poissonovu regresi. Díky vyjádření (1.4) víme, že náhodné veličiny N_1 a N_2 mají Poissonovo rozdělení postupně s parametry $\lambda_1 + \lambda_0$ a $\lambda_2 + \lambda_0$.

Vezmeme náhodný výběr o velikosti I pro náhodnou veličinu N_1 ,

$$\mathbf{N}_1 = (N_{11}, \dots, N_{1I})^\top$$

a náhodný výběr se stejnou velikostí pro náhodnou veličinu N_2 ,

$$\mathbf{N}_2 = (N_{21}, \dots, N_{2I})^\top.$$

Předpoklady modelu pro odhad parametrů můžeme schematicky zapsat následovně

$$\begin{aligned} (N_{1i}, N_{2i}) &\sim \text{BP}(\lambda_{0i}, \lambda_{1i}, \lambda_{2i}), \\ \log \lambda_{0i} &= \mathbf{x}_i^{0\top} \boldsymbol{\beta}_0, \\ \log \lambda_{1i} &= \mathbf{x}_i^{1\top} \boldsymbol{\beta}_1, \\ \log \lambda_{2i} &= \mathbf{x}_i^{2\top} \boldsymbol{\beta}_2, \end{aligned} \tag{1.5}$$

kde $i = 1, \dots, I$ a I je počet pozorování (počet pojistek, u kterých sledujeme počty škod). \mathbf{x}_i^k jsou vektory vysvětlujících proměnných a $\boldsymbol{\beta}_k$ jsou příslušné vektory regresních koeficientů – v našem případě $k = 0, 1, 2$.

Ve vztahu (1.5) je sice uvedena i rovnice, ze které je možné odhadnout parametr λ_{0i} , ale aby byla zachována struktura modelu, tak není vhodné odhadovat parametr λ_{0i} pomocí vysvětlujících proměnných, protože tento parametr zachycuje vzájemnou korelaci náhodných veličin (N_{1i}, N_{2i}) .

V dalších odstavcích si uvedeme dvě rozšíření dvourozměrného Poissonova rozdělení pro případy, kdy chceme podchytit hodnoty v datech, které se mohou častěji vyskytovat.

1.4.1 V nule modifikované dvourozměrné Poissonovo rozdělení

V nule modifikované rozdělení se používá pro data, která mají zvýšený počet nulových hodnot. Což je častý případ v portfoliu smluv, kdy je větší počet pojistek, na kterých nebyla hlášena žádná škoda oproti očekávanému počtu smluv s žádnou škodou.

V případě dvourozměrného rozdělení se jedná o vyšší počet hodnot $(0, 0)$. V nule modifikované dvourozměrné Poissonovo rozdělení můžeme zkonstruovat, obdobně jako v jednorozměrném případě, z dvourozměrného Poissonova rozdělení navýšením pravděpodobnosti výskytu události $[0, 0]$. V nule modifikované dvourozměrné Poissonovo rozdělení je popsáno pravděpodobnostmi

$$\begin{aligned} P(N_1 = 0, N_2 = 0) &= p + (1 - p) P_{BPD}(0, 0), \\ P(N_1 = n_1, N_2 = n_2) &= (1 - p) P_{BPD}(n_1, n_2), \quad (n_1, n_2) \neq (0, 0), \end{aligned}$$

kde $p \in (0, 1)$. Pravděpodobnost $P_{BPD}(n_1, n_2)$ je definována v (1.3). V tomto případě mají marginální distribuční funkce v nule modifikované jednorozměrné Poissonovo rozdělení. Pravděpodobnostní funkci v nule modifikovaného dvourozměrného rozdělení označíme jako P_{ZIBPD} z anglického *zero inflated Poisson distribution*.

Kovarianci mezi složkami N_1 a N_2 je možné vyjádřit jako

$$\begin{aligned} \text{cov}(N_1, N_2) &= \mathbf{E} N_1 N_2 - \mathbf{E} N_1 \mathbf{E} N_2 = \\ &= \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} n_1 n_2 P_{ZIBPD}(n_1, n_2) - (1 - p)^2 (\lambda_1 + \lambda_0) (\lambda_2 + \lambda_0) = \\ &= \sum_{n_1=1}^{\infty} \sum_{n_2=1}^{\infty} (1 - p) n_1 n_2 P_{BPD}(n_1, n_2) - (1 - p)^2 (\lambda_1 + \lambda_0) (\lambda_2 + \lambda_0) = \\ &= (1 - p) \mathbf{E}_{BPD} [N_1 N_2] - (1 - p)^2 (\lambda_1 + \lambda_0) (\lambda_2 + \lambda_0) = \\ &= (1 - p) [\text{cov}_{BPD}(N_1, N_2) + \mathbf{E}_{BPD} N_1 \mathbf{E}_{BPD} N_2] - (1 - p)^2 (\lambda_1 + \lambda_0) (\lambda_2 + \lambda_0) = \\ &= (1 - p) [\lambda_0 + (\lambda_1 + \lambda_0) (\lambda_2 + \lambda_0)] - (1 - p)^2 (\lambda_1 + \lambda_0) (\lambda_2 + \lambda_0), \quad (1.6) \end{aligned}$$

kde \mathbf{E}_{BPD} označuje střední hodnotu počítanou vzhledem k pravděpodobnostní funkci dvourozměrného Poissonova rozdělení.

Ostatní charakteristiky v nule modifikovaného dvourozměrného Poissonova rozdělení jsou shrnuty v tabulce 1.4.

Toto rozdělení je možné dále rozšířit pro případy, kdy je počet škod v obou složkách vektoru \mathbf{N} stejný, jak je ukázáno v další části.

Základní charakteristika	Hodnota
vektor středních hodnot	$\begin{pmatrix} (1-p)(\lambda_1 + \lambda_0) \\ (1-p)(\lambda_2 + \lambda_0) \end{pmatrix},$
rozptylová matice	$\begin{pmatrix} (1-p)(\lambda_1 + \lambda_0) + p(1-p)(\lambda_1 + \lambda_0)^2 & \\ & (1-p)(\lambda_2 + \lambda_0) + p(1-p)(\lambda_2 + \lambda_0)^2 \end{pmatrix} \quad (1.6)$

Tabulka 1.4: Základní charakteristiky v nule modifikovaného dvourozměrného Poissonova rozdělení.

1.4.2 Na diagonále modifikované dvourozměrné Poissonovo rozdělení

Diagonálou jsou myšleny ty hodnoty, kde $n_1 = n_2$, tedy nejen událost $[0, 0]$, ale všechny události $[n, n]$, kde $n \in \mathbb{N} \cup \{0\}$.

Jelikož se často stává, že při jedné autonehodě nastává škoda jak z povinného ručení, tak z havarijního pojištění, je pravděpodobné, že pojistky se stejnými počty škod z obou pojištění mají vyšší pravděpodobnost výskytu, než je $PBD(n, n)$ pro $n \in \mathbb{N} \cup \{0\}$.

Pravděpodobnosti na diagonále modifikovaného dvourozměrného Poissonova rozdělení je možné vyjádřit pro $p \in (0, 1)$ jako

$$\begin{aligned} P(N_1 = n, N_2 = n) &= p P_D(n) + (1-p) P_{BPD}(n, n), \\ P(N_1 = n_1, N_2 = n_2) &= (1-p) P_{BPD}(n_1, n_2), \quad n_1 \neq n_2. \end{aligned}$$

Pravděpodobnost $P_D(n)$ je pravděpodobnostní funkce libovolného diskrétního rozdělení s vektorem parametrů θ .

Speciální případy na diagonále modifikovaného dvourozměrného Poissonova rozdělení jsou pro $p = 0$, kdy vznikne dvourozměrné Poissonovo rozdělení, a případ, kdy P_D je degenerované rozdělení v nule, kdy vznikne v nule modifikované dvourozměrné Poissonovo rozdělení.

1.5 Trojrozměrné poissonovské modely

Doposud jsme se zabývali maximálně dvěma různými druhy garancí, u kterých jsme zkoumali počty škod. Nyní přidáme ještě další druh garance. Můžeme tak například sledovat počty škod z povinného ručení, počty škod z havarijního pojištění a počty škod z ostatního krytí. Stejně jako v případě dvou garancí, budeme chtít podchytit vztah mezi jednotlivými garancemi, nebudeme tak předpokládat nezávislost počtů škod v jednotlivých garancích.

V této kapitole si zavedeme, stejně jako v Bermudez a Karlis (2011), modely pro trojrozměrné počty škod modelované pomocí Poissonova trojrozměrného rozdělení. Nejprve si uvedeme model se společnou kovariancí, a potom model s úplnou kovariancí.

1.5.1 Model se společnou kovariancí

Model se společnou kovariancí popisuje trojrozměrný náhodný vektor počtů škod, kde jednotlivé složky vektoru mají stejnou vzájemnou kovarianci. Jedná se o podobný model jako v oddíle 1.4, pouze je rozšířený do dalšího rozměru.

Definice 5. *Náhodný vektor (N_1, N_2, N_3) má trojrozměrné Poissonovo rozdělení se společnou kovariancí, pokud pro všechna $n_1, n_2, n_3 = 0, 1, 2, \dots$ a $\lambda_0, \lambda_1, \lambda_2, \lambda_3 > 0$ je sdružená pravděpodobnostní funkce vektoru (N_1, N_2, N_3) dána vztahem*

$$P(N_1 = n_1, N_2 = n_2, N_3 = n_3) = e^{-\lambda} \sum_{k=0}^n \frac{\lambda_0^k}{k!} \frac{\lambda_1^{n_1-k}}{(n_1-k)!} \frac{\lambda_2^{n_2-k}}{(n_2-k)!} \frac{\lambda_3^{n_3-k}}{(n_3-k)!}, \quad (1.7)$$

kde $\lambda = \lambda_0 + \lambda_1 + \lambda_2 + \lambda_3$ a $n = \min(n_1, n_2, n_3)$.

Pravděpodobnost (1.7) označíme jako P_{TPDCC} , kde písmena v dolním indexu jsou zkratkou anglického názvu *Trivariate Poisson distribution – common covariance*.

Stejně jako v oddílu 1.4 je možné definovat model trojrozměrného Poissonova rozdělení se společnou kovariancí také pomocí trojrozměrného náhodného vektoru $\mathbf{N} = (N_1, N_2, N_3)^\top$, který splňuje

$$\begin{aligned} N_1 &= Y_1 + Y_0, \\ N_2 &= Y_2 + Y_0, \\ N_3 &= Y_3 + Y_0. \end{aligned}$$

Náhodné veličiny Y_0, Y_1, Y_2, Y_3 jsou nezávislé a mají jednorozměrné Poissonovo rozdělení s parametry $\lambda_0, \lambda_1, \lambda_2, \lambda_3 > 0$. Parametr λ_0 vyjadřuje vzájemnou kovarianci mezi jednotlivými složkami vektoru, která je pro všechny složky stejná. Marginální jednorozměrné rozdělení N_i je Poissonovo s parametrem $\lambda_i + \lambda_0$, kde $i \in \{1, 2, 3\}$.

Sdružené rozdělení náhodných veličin N_1, N_2, N_3 lze snadno převést do tvaru (1.7). Použijeme zavedené značení z definice 5, a to n a λ .

$$\begin{aligned} P(N_1 = n_1, N_2 = n_2, N_3 = n_3) &= P(Y_1 + Y_0 = n_1, Y_2 + Y_0 = n_2, Y_3 + Y_0 = n_3) = \\ &= \sum_{k=0}^n P(Y_1 + Y_0 = n_1, Y_2 + Y_0 = n_2, Y_3 + Y_0 = n_3 | Y_0 = k) P(Y_0 = k) = \\ &= e^{-\lambda_0} \sum_{k=0}^n \frac{\lambda_0^k}{k!} P(Y_1 = n_1 - k, Y_2 = n_2 - k, Y_3 = n_3 - k) = \\ &= e^{-\lambda_0} \sum_{k=0}^n \frac{\lambda_0^k}{k!} P(Y_1 = n_1 - k) P(Y_2 = n_2 - k) P(Y_3 = n_3 - k) = \\ &= e^{-\lambda} \sum_{k=0}^n \frac{\lambda_0^k}{k!} \frac{\lambda_1^{n_1-k}}{(n_1-k)!} \frac{\lambda_2^{n_2-k}}{(n_2-k)!} \frac{\lambda_3^{n_3-k}}{(n_3-k)!}. \end{aligned}$$

V tabulce 1.5 jsou shrnuty základní charakteristiky trojrozměrného Poissonova rozdělení se společnou kovariancí. Je vidět, že střední hodnota a rozptyl jsou v jednotlivých dimenzích stejné, jelikož se jedná o po složkách Poissonovo rozdělení.

Základní charakteristika	Hodnota
vektor středních hodnot	$\begin{pmatrix} \lambda_1 + \lambda_0 \\ \lambda_2 + \lambda_0 \\ \lambda_3 + \lambda_0 \end{pmatrix},$
rozptylová matice	$\begin{pmatrix} \lambda_1 + \lambda_0 & \lambda_0 & \lambda_0 \\ \lambda_0 & \lambda_2 + \lambda_0 & \lambda_0 \\ \lambda_0 & \lambda_0 & \lambda_3 + \lambda_0 \end{pmatrix}.$

Tabulka 1.5: Základní charakteristiky trojrozměrného Poissonova rozdělení se společnou kovariancí.

Odhad parametrů modelu

Pro odhad parametrů $\lambda_0, \dots, \lambda_3$ trojrozměrného Poissonova modelu použijeme vícerozměrnou Poissonovu regresi.

Využijeme pozorování počtů škod pro každou garanci, které značíme N_{ki} , kde k přísluší k -té garanci $k \in (1, 2, 3)$ a i označuje i -tou pojistnou smlouvu z celkového počtu I smluv v portfoliu.

Stejně jako v části 1.3 využijeme logaritmickou linkovou funkci a budeme tak modelovat logaritmus parametru λ_{ki} pro $k \in (1, 2, 3)$ a $i \in (1, \dots, I)$ pomocí vysvětlujících proměnných \mathbf{x}_i^k .

Předpoklady modelu pro odhad parametrů můžeme schematicky zapsat následovně

$$\begin{aligned}
(N_{1i}, N_{2i}, N_{3i}) &\sim \text{TPDCC}(\lambda_{0i}, \lambda_{1i}, \lambda_{2i}, \lambda_{3i}), \\
\log \lambda_{1i} &= \mathbf{x}_i^{1\top} \boldsymbol{\beta}_1, \\
\log \lambda_{2i} &= \mathbf{x}_i^{2\top} \boldsymbol{\beta}_2, \\
\log \lambda_{3i} &= \mathbf{x}_i^{3\top} \boldsymbol{\beta}_3,
\end{aligned} \tag{1.8}$$

kde \mathbf{x}_i^k jsou vektory vysvětlujících proměnných a $\boldsymbol{\beta}_k$ jsou vektory příslušných regresních koeficientů – v našem případě $k = 1, 2, 3$.

Parametr λ_0 většinou není modelován pomocí vektoru vysvětlujících proměnných, protože vyjadřuje vzájemnou kovarianci mezi složkami vektoru, a použití vysvětlujících proměnných by bylo obtížně interpretovatelné.

Model se společnou kovariancí je možné použít pouze pro shodné kovariance mezi složkami vektoru a není možné pomocí něj modelovat overdispersi. Proto tento model v následující kapitole rozšíříme na model s úplnou kovariancí.

1.5.2 Model s úplnou kovariancí

Model s úplnou kovariancí je rozšířením modelu se společnou kovariancí pro různé párové kovariance vektoru \mathbf{N} .

Definice 6. *Náhodný vektor (N_1, N_2, N_3) má trojrozměrné Poissonovo rozdělení s úplnou kovariancí, pokud pro všechna $n_1, n_2, n_3 = 0, 1, 2, \dots$ a $\mu_1, \mu_2, \mu_3, \lambda_{12}, \lambda_{13}$,*

$\lambda_{23} > 0$ je sdružená pravděpodobnostní funkce vektoru (N_1, N_2, N_3) dána vztahem

$$P(N_1 = n_1, N_2 = n_2, N_3 = n_3) = e^{-\lambda} \sum_{k_1=0}^{s_1} \sum_{k_2=0}^{s_2} \sum_{k_3=0}^{s_3} \frac{\lambda_{12}^{k_1}}{k_1!} \frac{\lambda_{13}^{k_2}}{k_2!} \frac{\lambda_{23}^{k_3}}{k_3!} \cdot \frac{\mu_1^{n_1-k_1-k_2}}{(n_1-k_1-k_2)!} \frac{\mu_2^{n_2-k_1-k_3}}{(n_2-k_1-k_3)!} \frac{\mu_3^{n_3-k_2-k_3}}{(n_3-k_2-k_3)!}, \quad (1.9)$$

kde $\lambda = \lambda_{12} + \lambda_{23} + \lambda_{13} + \mu_1 + \mu_2 + \mu_3$, $s_1 = \min(n_1, n_2)$, $s_2 = \min(n_2 - s_1, n_3)$ a $s_3 = \min(n_2 - s_1, n_3 - s_2)$.

Pravděpodobnost (1.9) označíme jako P_{TPDFC} , kde písmena v dolním indexu jsou zkratkou anglického názvu *Trivariate Poisson distribution – full covariance*.

Trojrozměrný náhodný vektor s trojrozměrným Poissonovým rozdělením s úplnou kovariancí může být po složkách modelován pomocí devíti náhodných veličin, které mají Poissonovo rozdělení a jsou vzájemně nezávislé

$$\begin{aligned} N_1 &= Y_1 + Y_{12} + Y_{13}, \\ N_2 &= Y_2 + Y_{12} + Y_{23}, \\ N_3 &= Y_3 + Y_{13} + Y_{23}. \end{aligned}$$

Náhodné veličiny Y_1, Y_2, Y_3 mají Poissonovo rozdělení postupně s parametry $\mu_1, \mu_2, \mu_3 > 0$. Náhodné veličiny Y_{12}, Y_{13}, Y_{23} mají také Poissonovo rozdělení, ale postupně s parametry $\lambda_{12}, \lambda_{13}, \lambda_{23} > 0$. Kovarianční matici náhodného vektoru \mathbf{N} je možné vyjádřit jako matici

$$\text{cov}(\mathbf{N}) = \begin{pmatrix} \mu_1 + \lambda_{12} + \lambda_{13} & \lambda_{12} & \lambda_{13} \\ \lambda_{12} & \mu_2 + \lambda_{12} + \lambda_{23} & \lambda_{23} \\ \lambda_{13} & \lambda_{23} & \mu_3 + \lambda_{13} + \lambda_{23} \end{pmatrix}. \quad (1.10)$$

Základní charakteristiky trojrozměrného Poissonova rozdělení s úplnou kovariancí jsou shrnuty v tabulce 1.6.

Základní charakteristika	Hodnota
vektor středních hodnot	$\begin{pmatrix} \mu_1 + \lambda_{12} + \lambda_{13} \\ \mu_2 + \lambda_{12} + \lambda_{23} \\ \mu_3 + \lambda_{13} + \lambda_{23} \end{pmatrix},$
rozptylová matice	(1.10).

Tabulka 1.6: Základní charakteristiky trojrozměrného Poissonova rozdělení s úplnou kovariancí.

Ukážeme, že sdružené rozdělení náhodných veličin N_1, N_2, N_3 odpovídá tvaru (1.9). Použijeme zavedené značení z definice 6, a to s_1, s_2, s_3 a λ .

$$\begin{aligned} P(N_1 = n_1, N_2 = n_2, N_3 = n_3) &= \\ &= \sum_{k_1=0}^{s_1} \sum_{k_2=0}^{s_2} \sum_{k_3=0}^{s_3} P(Y_1 + Y_{12} + Y_{13} = n_1, Y_2 + Y_{12} + Y_{23} = n_2, \end{aligned}$$

$$\begin{aligned}
& Y_3 + Y_{13} + Y_{23} = n_3 \mid Y_{12} = k_1, Y_{13} = k_2, Y_{23} = k_3) P(Y_{12} = k_1 \mid Y_{13} = k_2, Y_{23} = k_3) \cdot \\
& \quad \cdot P(Y_{13} = k_2 \mid Y_{23} = k_3) P(Y_{23} = k_3) = \\
& = \sum_{k_1=0}^{s_1} \sum_{k_2=0}^{s_2} \sum_{k_3=0}^{s_3} P(Y_1 = n_1 - k_1 - k_2, Y_2 = n_2 - k_1 - k_3, Y_3 = n_3 - k_2 - k_3) \cdot \\
& \quad \cdot P(Y_{12} = k_1) P(Y_{13} = k_2) P(Y_{23} = k_3) = \\
& = \sum_{k_1=0}^{s_1} \sum_{k_2=0}^{s_2} \sum_{k_3=0}^{s_3} P(Y_1 = n_1 - k_1 - k_2) P(Y_2 = n_2 - k_1 - k_3) P(Y_3 = n_3 - k_2 - k_3) \cdot \\
& \quad \cdot P(Y_{12} = k_1) P(Y_{13} = k_2) P(Y_{23} = k_3) = \\
& = \sum_{k_1=0}^{s_1} \sum_{k_2=0}^{s_2} \sum_{k_3=0}^{s_3} e^{-\lambda} \frac{\mu_1^{n_1 - k_1 - k_2}}{(n_1 - k_1 - k_2)!} \frac{\mu_2^{n_2 - k_1 - k_3}}{(n_2 - k_1 - k_3)!} \frac{\mu_3^{n_3 - k_2 - k_3}}{(n_3 - k_2 - k_3)!} \frac{\lambda_{12}^{k_1}}{k_1!} \frac{\lambda_{13}^{k_2}}{k_2!} \frac{\lambda_{23}^{k_3}}{k_3!}.
\end{aligned}$$

Při úpravách jsme využili vzájemnou nezávislost všech náhodných veličin Y_i , kde $i = 1, 2, 3, 12, 13, 23$.

Odhad parametrů modelu

Pro odhad parametrů trojrozměrného Poissonova modelu s úplnou kovariancí je možné opět použít Poissonovu regresi, a to buď pro všechny parametry modelu, nebo pouze pro parametry střední hodnoty μ_1, μ_2, μ_3 .

Pro vyjádření parametrů máme k dispozici hodnoty vysvětlujících proměnných \mathbf{x}_i^k a vektor regresních koeficientů $\boldsymbol{\beta}_k$, kde $k \in (1, \dots, K)$ a $i \in (1, \dots, I)$, K je počet odhadovaných parametrů pomocí Poissonovy regrese, I je celkový počet pozorování.

Předpoklady modelu pro odhad parametrů můžeme schematicky zapsat následovně

$$\begin{aligned}
(N_{1i}, N_{2i}, N_{3i}) & \sim \text{TPDFC}(\mu_{1i}, \mu_{2i}, \mu_{3i}, \lambda_{12i}, \lambda_{13i}, \lambda_{23i}), \\
\log \mu_{1i} & = \mathbf{x}_i^1 \boldsymbol{\beta}_1, \\
\log \mu_{2i} & = \mathbf{x}_i^2 \boldsymbol{\beta}_2, \\
\log \mu_{3i} & = \mathbf{x}_i^3 \boldsymbol{\beta}_3, \\
\log \lambda_{12i} & = \mathbf{x}_i^4 \boldsymbol{\beta}_4, \\
\log \lambda_{13i} & = \mathbf{x}_i^5 \boldsymbol{\beta}_5, \\
\log \lambda_{23i} & = \mathbf{x}_i^6 \boldsymbol{\beta}_6,
\end{aligned} \tag{1.11}$$

kde $i = 1, \dots, I$, \mathbf{x}_i^k jsou vektory vysvětlujících proměnných a $\boldsymbol{\beta}_k$ jsou vektory příslušných regresních koeficientů – v našem případě $k = 1, \dots, 6$.

Model zahrnující odhad všech parametrů je velmi komplikovaný pro interpretaci, proto je vhodnější odhadovat pomocí Poissonovy regrese pouze parametry popisující střední hodnoty rozdělení, tedy μ_1, μ_2, μ_3 . Schematicky zapsáno, dostáváme model

$$\begin{aligned}
(N_{1i}, N_{2i}, N_{3i}) & \sim \text{TPDFC}(\mu_{1i}, \mu_{2i}, \mu_{3i}, \lambda_{12i}, \lambda_{13i}, \lambda_{23i}), \\
\log \mu_{1i} & = \mathbf{x}_i^1 \boldsymbol{\beta}_1, \\
\log \mu_{2i} & = \mathbf{x}_i^2 \boldsymbol{\beta}_2, \\
\log \mu_{3i} & = \mathbf{x}_i^3 \boldsymbol{\beta}_3,
\end{aligned}$$

kde \mathbf{x}_i^k jsou vektory vysvětlujících proměnných a β_k jsou vektory regresních parametrů. V tomto případě indexy k nabývají hodnot 1, 2, 3.

1.5.3 V nule modifikované trojrozměrné modely

Rozšířením trojrozměrného Poissonova modelu s úplnou kovariancí je trojrozměrný Poissonův model, který je v nule modifikovaný. Modifikaci je možné provést po složkách, kdy jednorozměrné Poissonovo rozdělení je modifikováno v nule. Touto modifikací vznikne trojrozměrné v nule modifikované Poissonovo rozdělení.

Hlavní důvod modifikace v nule je možnost zachycení většího výskytu případů, které mají ve všech dimenzích nulový počet škod, a modelování overdisperze, stejně jako v částech 1.2 a 1.4.1.

Uvedeme pouze základní případ, kdy je modifikace pravděpodobnosti prováděna ve všech dimenzích pomocí stejné hodnoty p . Pro další rozšíření by bylo možné uvažovat modifikaci nulových hodnot v každé dimenzi s jinou hodnotou např. postupně s p_1, p_2, p_3 .

Budeme vycházet z pravděpodobnosti P_{TPDFC} trojrozměrného Poissonova modelu s úplnou kovariancí, kterou upravíme a dostaneme pravděpodobnost

$$\begin{aligned} P_{TPDZI}(0, 0, 0) &= p + (1 - p) P_{TPDFC}(0, 0, 0), \\ P_{TPDZI}(n_1, n_2, n_3) &= (1 - p) P_{TPDFC}(n_1, n_2, n_3), \quad (n_1, n_2, n_3) \neq (0, 0, 0), \end{aligned}$$

kde $p \in (0, 1)$.

2. Negativně binomické modely

Pro modelování počtů škod se nejčastěji využívá dvou rozdělení – Poissonova a negativně binomického. Modifikaci Poissonova rozdělení pro vícerozměrné případy jsme si ukázali v kapitole 1. V této kapitole si uvedeme definici vícerozměrného negativně binomického rozdělení podle Shi a Valdez (2014), konkrétně pro případ trojrozměrného rozdělení. Výhodou negativně binomického rozdělení oproti Poissonovu rozdělení je to, že pomocí něho je možné zachytit overdispersi i bez toho, aby se musela používat v nule modifikovaná varianta rozdělení.

2.1 Negativně binomické rozdělení

Negativně binomické rozdělení je možné definovat různými způsoby. Uvedeme definici podle Shi a Valdez (2014) s využitím gama funkce.

Definice 7. *Náhodná veličina N má negativně binomické rozdělení s parametry ψ a η , zkráceně $N \sim \text{NB}(\psi, \eta)$, pokud pro všechna $n = 0, 1, 2, \dots$, $\psi > 0$ a $\eta > 0$ platí*

$$P(N = n) = \frac{\Gamma(\eta + n)}{\Gamma(\eta)\Gamma(n + 1)} \left(\frac{1}{1 + \psi} \right)^\eta \left(\frac{\psi}{1 + \psi} \right)^n, \quad (2.1)$$

kde $\Gamma(n) = \int_0^\infty t^{n-1} e^{-t} dt$.

Pravděpodobnost $P(N = n)$ pro negativně binomické rozdělení budeme označovat $P_{NBD}(n)$, kde písmena *NBD* jsou zkratkou anglického názvu *negative binomial distribution*.

V tabulce 2.1 jsou shrnuty základní charakteristiky negativně binomického rozdělení. Jak je ihned patrné z vyjádření rozptylu, ten je oproti střední hodnotě větší o kladný člen $\eta\psi^2$.

Základní charakteristika	Hodnota
střední hodnota	$\eta\psi$,
rozptyl	$\eta(\psi + \psi^2)$.

Tabulka 2.1: Základní charakteristiky negativně binomického rozdělení.

Jednorozměrné negativně binomické rozdělení nyní rozšíříme do tří dimenzí.

2.2 Trojrozměrné negativně binomické modely

Uvedeme si definici trojrozměrného negativně binomického rozdělení podle Shi a Valdez (2014). Pro definici využijeme dva typy modelů negativně binomického rozdělení – model NB-I popsáný v části 2.2.1 a model NB-II popsáný v části 2.2.2.

Stejně jako v případě trojrozměrného Poissonova rozdělení budeme pro negativně binomické rozdělení uvažovat dvě různé struktury závislostí mezi jednotlivými členy trojrozměrného vektoru. Nejprve si definujeme trojrozměrné negativně binomické rozdělení se společnou kovariancí, kde jsou jednotlivé členy vektoru

provázány stejnou „šokovou“ náhodnou veličinou. Dále si uvedeme zobecnění tohoto rozdělení, a to negativně binomické rozdělení s úplnou kovariancí, kde je kovarianční struktura členů vektoru libovolná.

Pro definici trojrozměrného negativně binomického rozdělení využijeme dvě různé reparametrizace tvaru pravděpodobnosti (2.1). Pro střední hodnotu budeme předpokládat, že je zachycena pomocí vysvětlujících proměnných následujícím způsobem

$$\lambda := \eta \cdot \psi = \exp\{\mathbf{x}^\top \boldsymbol{\beta}\}, \quad (2.2)$$

kde \mathbf{x} je vektor vysvětlujících proměnných a $\boldsymbol{\beta}$ je vektor regresních parametrů.

Při daném parametru η získáme z rovnice (2.2) vyjádření pro parametr ψ , následně dosadíme tyto parametry do pravděpodobnostní funkce negativně binomického rozdělení (2.1). Různou definicí parametru η dostaneme dva různé modely, které dále využijeme pro definici trojrozměrného negativně binomického rozdělení.

2.2.1 Model NB-I

První model – model NB-I, dostaneme za předpokladu vyjádření parametru η jako

$$\eta = \sigma^{-2} \exp\{\mathbf{x}^\top \boldsymbol{\beta}\}.$$

Z rovnice (2.2) vyjádříme parametr ψ jako

$$\psi = \sigma^2.$$

Pravděpodobnost (2.1) potom přechází do tvaru

$$f^{\text{NB-I}}(n | \mathbf{x}, \boldsymbol{\beta}, \sigma^2) = \frac{\Gamma(\sigma^{-2} \exp\{\mathbf{x}^\top \boldsymbol{\beta}\} + n)}{\Gamma(\sigma^{-2} \exp\{\mathbf{x}^\top \boldsymbol{\beta}\}) \Gamma(n + 1)} \cdot \left(\frac{1}{1 + \sigma^2}\right)^{\sigma^{-2} \exp\{\mathbf{x}^\top \boldsymbol{\beta}\}} \left(\frac{\sigma^2}{1 + \sigma^2}\right)^n. \quad (2.3)$$

2.2.2 Model NB-II

Druhý model – NB-II, dostaneme za předpokladu vyjádření parametru η jako

$$\eta = \sigma^{-2}.$$

Z rovnice (2.2) vyjádříme parametr ψ

$$\psi = \sigma^2 \exp\{\mathbf{x}^\top \boldsymbol{\beta}\}.$$

Pravděpodobnost (2.1) potom přechází do tvaru

$$f^{\text{NB-II}}(n | \mathbf{x}, \boldsymbol{\beta}, \sigma^2) = \frac{\Gamma(\sigma^{-2} + n)}{\Gamma(\sigma^{-2}) \Gamma(n + 1)} \cdot \left(\frac{1}{1 + \sigma^2 \exp\{\mathbf{x}^\top \boldsymbol{\beta}\}}\right)^{\sigma^{-2}} \left(\frac{\sigma^2 \exp\{\mathbf{x}^\top \boldsymbol{\beta}\}}{1 + \sigma^2 \exp\{\mathbf{x}^\top \boldsymbol{\beta}\}}\right)^n. \quad (2.4)$$

2.2.3 Trojrozměrné negativně binomické rozdělení se společnou kovariancí

Trojrozměrné negativně binomické rozdělení je postaveno na stejném principu jako trojrozměrné Poissonovo rozdělení se společnou kovariancí uvedené v části 1.5.1. Kovariance mezi jednorozměrnými náhodnými veličinami, ze kterých se skládá trojrozměrný náhodný vektor, je pro všechny dvojice stejná.

Definice 8. *Náhodný vektor $\mathbf{N} = (N_1, N_2, N_3)^\top$ má trojrozměrné negativně binomické rozdělení se společnou kovariancí, pokud pro všechna $n_1, n_2, n_3 = 0, 1, \dots$ je sdružená pravděpodobnostní funkce vektoru \mathbf{N} dána vztahem*

$$\begin{aligned} P(N_1 = n_1, N_2 = n_2, N_3 = n_3 | \mathbf{x}) &= \\ &= \sum_{s=0}^n f_0(s) f_1(n_1 - s | \mathbf{x}^1) f_2(n_2 - s | \mathbf{x}^2) f_3(n_3 - s | \mathbf{x}^3), \end{aligned} \quad (2.5)$$

kde $n = \min(n_1, n_2, n_3)$, \mathbf{x} je regresní matice složená z vektorů vysvětlujících proměnných $\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3$,

$$f_0(\cdot) = f^{NB-I}(\cdot | 1, \log \lambda_0, \sigma^2)$$

a

$$f_j(\cdot | \mathbf{x}^j) = f^{NB-I}(\cdot | \mathbf{x}^j, \beta_j, \sigma^2),$$

kde $j = 1, 2, 3$.

Pravděpodobnost (2.5) označíme jako P_{TBNCC} , kde písmena v dolním indexu jsou zkratkou anglického názvu *trivariate negative binomial distribution – common covariance*.

Trojrozměrné negativně binomické rozdělení se společnou kovariancí má náhodný vektor $\mathbf{N} = (N_1, N_2, N_3)^\top$, který je generovaný náhodnými veličinami U_0, U_1, U_2, U_3 ,

$$\begin{aligned} N_1 &= U_1 + U_0, \\ N_2 &= U_2 + U_0, \\ N_3 &= U_3 + U_0. \end{aligned}$$

Náhodné veličiny U_0, U_1, U_2, U_3 jsou vzájemně nezávislé a mají jednorozměrné negativně binomické rozdělení s parametry (ψ, η_j) – parametr ψ je stejný u všech náhodných veličin U_j , kde $j \in \{0, 1, 2, 3\}$, zatímco parametry η_j mohou být odlišné pro každou z náhodných veličin.

Využijeme vztah (2.2) a pro parametr ψ budeme předpokládat

$$\psi = \sigma^2.$$

Parametry η_j následně nabývají hodnot

$$\eta_j = \frac{\lambda_j}{\sigma^2}.$$

Po této reparametrizaci dostaneme základní charakteristiky trojrozměrného negativně binomického rozdělení se společnou kovariancí, které jsou uvedeny v tabulce 2.2.

Základní charakteristika	Hodnota
vektor středních hodnot	$\begin{pmatrix} \lambda_1 + \lambda_0 \\ \lambda_2 + \lambda_0 \\ \lambda_3 + \lambda_0 \end{pmatrix},$
rozptylová matice	$(1 + \sigma^2) \begin{pmatrix} \lambda_1 + \lambda_0 & \lambda_0 & \lambda_0 \\ \lambda_0 & \lambda_2 + \lambda_0 & \lambda_0 \\ \lambda_0 & \lambda_0 & \lambda_3 + \lambda_0 \end{pmatrix}.$

Tabulka 2.2: Základní charakteristiky trojrozměrného negativně binomického rozdělení se společnou kovariancí, kde $\sigma^2 = \psi$ a $\lambda_j = \eta_j \psi$.

Náhodné veličiny N_j , $j = 1, 2, 3$, mají jednorozměrné negativně binomické rozdělení s parametry $(\psi, \eta_j + \eta_0)$. Proto stejně jako v jednorozměrném případě je rozptyl složek N_j větší než jejich střední hodnota, konkrétně pro každou složku platí

$$\mathbb{E} N_j = (1 + \sigma^2) \text{var} N_j.$$

Ukážeme, že náhodný vektor \mathbf{N} s trojrozměrným negativně binomickým rozdělením se společnou kovariancí má za sdruženou pravděpodobnostní funkci 2.5.

$$\begin{aligned} P(N_1 = n_1, N_2 = n_2, N_3 = n_3 | \mathbf{x}) &= \\ &= P(U_1 + U_0 = n_1, U_2 + U_0 = n_2, U_3 + U_0 = n_3 | \mathbf{x}) = \\ &= \sum_{s=0}^n P(U_1 = n_1 - s, U_2 = n_2 - s, U_3 = n_3 - s, | \mathbf{x}, U_0 = s) P(U_0 = s | \mathbf{x}) = \\ &= \sum_{s=0}^n P_{\mathbf{x},s}(U_1 = n_1 - s) P_{\mathbf{x},s}(U_2 = n_2 - s) P_{\mathbf{x},s}(U_3 = n_3 - s) P_{\mathbf{x}}(U_0 = s) = \\ &= \sum_{s=0}^n f^{NB-I}(n_1 - s | \mathbf{x}^1, \boldsymbol{\beta}_1, \sigma^2) f^{NB-I}(n_2 - s | \mathbf{x}^2, \boldsymbol{\beta}_2, \sigma^2) \cdot \\ &\quad \cdot f^{NB-I}(n_3 - s | \mathbf{x}^3, \boldsymbol{\beta}_3, \sigma^2) f^{NB-I}(s | 1, \log \lambda_0, \sigma^2), \end{aligned}$$

kde $P_{\mathbf{x},s}$ značí pravděpodobnost za podmínky znalosti regresní matice \mathbf{x} a za podmínky $[U_0 = s]$.

Předpoklad společné kovariance je možné rozvolnit a dostaneme tak obecnější kovarianční strukturu.

2.2.4 Trojrozměrné negativně binomické rozdělení s úplnou kovariancí

Trojrozměrné negativně binomické rozdělení se společnou kovariancí zobecníme na trojrozměrné negativně binomické rozdělení s úplnou kovariancí, kde je kovarianční struktura určena pomocí obecné kovarianční matice a kovariance mezi dvojicemi náhodných veličin z trojrozměrného náhodného vektoru je různá pro různé dvojice.

Definice 9. Náhodný vektor $\mathbf{N} = (N_1, N_2, N_3)^\top$ má trojrozměrné negativně binomické rozdělení s úplnou kovariancí, pokud pro všechna $n_1, n_2, n_3 = 0, 1, \dots$ je

sdížená pravděpodobnostní funkce vektoru \mathbf{N} dána vztahem

$$P(N_1 = n_1, N_2 = n_2, N_3 = n_3 | \mathbf{x}) = \sum_{s_1=0}^{r_1} \sum_{s_2=0}^{r_2} \sum_{s_3=0}^{r_3} f_{12}(s_1) f_{13}(s_2) f_{23}(s_3) \cdot f_1(n_1 - s_1 - s_2 | \mathbf{x}^1) f_2(n_2 - s_1 - s_3 | \mathbf{x}^2) f_3(n_3 - s_2 - s_3 | \mathbf{x}^3), \quad (2.6)$$

kde $r_1 = \min(n_1, n_2)$, $r_2 = \min(n_1 - s_1, n_3)$, $r_3 = \min(n_2 - s_1, n_3 - s_2)$, \mathbf{x} je regresní matice složená z vektorů vysvětlujících proměnných $\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3$,

$$f_{jk}(\cdot) = f^{NB-I}(\cdot | 1, \log \lambda_{jk}, \sigma^2)$$

a

$$f_j(\cdot | \mathbf{x}^j) = f^{NB-I}(\cdot | \mathbf{x}^j, \beta_j, \sigma^2),$$

kde $j, k = 1, 2, 3$.

Pravděpodobnost (2.6) označíme jako P_{TBNFC} , kde písmena v dolním indexu jsou zkratkou anglického názvu *trivariate negative binomial distribution – full covariance*.

Trojrozměrné negativně binomické rozdělení s úplnou kovariancí má náhodný vektor $\mathbf{N} = (N_1, N_2, N_3)^\top$, který je generovaný náhodnými veličinami $U_1, U_2, U_3, U_{12}, U_{13}, U_{23}$, kde platí

$$\begin{aligned} N_1 &= U_1 + U_{12} + U_{13}, \\ N_2 &= U_2 + U_{12} + U_{23}, \\ N_3 &= U_3 + U_{13} + U_{23}. \end{aligned}$$

Náhodné veličiny $U_1, U_2, U_3, U_{12}, U_{13}, U_{23}$ jsou vzájemně nezávislé a mají jednorozměrné negativně binomické rozdělení s parametry (ψ, η_j) – parametr ψ je shodný pro všechny náhodné veličiny U_j , kde $j \in \{1, 2, 3, 12, 13, 23\}$, zatímco parametry η_j mohou být odlišné pro každou z náhodných veličin.

Parametry ψ a η opět přeparametrizujeme na model typu NB-I, dostaneme tak vztahy

$$\begin{aligned} \psi &= \sigma^2, \\ \eta_j &= \frac{\lambda_j}{\sigma^2}. \end{aligned}$$

Náhodné veličiny N_j mají opět jednorozměrné negativně binomické rozdělení, proto kovarianční matice náhodného vektoru \mathbf{N} má tvar

$$\text{cov}(\mathbf{N}) = (1 + \sigma^2) \begin{pmatrix} \lambda_1 + \lambda_{12} + \lambda_{13} & \lambda_{12} & \lambda_{13} \\ \lambda_{12} & \lambda_2 + \lambda_{12} + \lambda_{23} & \lambda_{23} \\ \lambda_{13} & \lambda_{23} & \lambda_3 + \lambda_{13} + \lambda_{23} \end{pmatrix}. \quad (2.7)$$

V tabulce 2.3 jsou shrnuty základní charakteristiky trojrozměrného negativně binomického rozdělení s úplnou kovariancí.

Základní charakteristika	Hodnota
vektor středních hodnot	$\begin{pmatrix} \lambda_1 + \lambda_{12} + \lambda_{13} \\ \lambda_2 + \lambda_{12} + \lambda_{23} \\ \lambda_3 + \lambda_{13} + \lambda_{23} \end{pmatrix},$
rozptylová matice	(2.7).

Tabulka 2.3: Základní charakteristiky trojrozměrného Poissonova rozdělení s úplnou kovariancí.

Ukážeme, že náhodný vektor \mathbf{N} s trojrozměrným negativně binomickým rozdělením s úplnou kovariancí má za sdruženou pravděpodobnostní funkci 2.6.

$$\begin{aligned}
& P(N_1 = n_1, N_2 = n_2, N_3 = n_3 | \mathbf{x}) = \\
& = P(U_1 + U_{12} + U_{13} = n_1, U_2 + U_{12} + U_{23} = n_2, U_3 + U_{13} + U_{23} = n_3 | \mathbf{x}) = \\
& = \sum_{s_1=0}^{r_1} \sum_{s_2=0}^{r_2} \sum_{s_3=0}^{r_3} P(U_1 = n_1 - s_1 - s_2, U_2 = n_2 - s_1 - s_3, U_3 = n_3 - s_2 - s_3 | \\
& | \mathbf{x}, U_{12} = s_1, U_{13} = s_2, U_{23} = s_3) P(U_{12} = s_1 | \mathbf{x}) P(U_{13} = s_2 | \mathbf{x}) P(U_{23} = s_3 | \mathbf{x}) = \\
& = \sum_{s_1=0}^{r_1} \sum_{s_2=0}^{r_2} \sum_{s_3=0}^{r_3} P_{\mathbf{x},s}(U_1 = n_1 - s_1 - s_2) P_{\mathbf{x},s}(U_2 = n_2 - s_1 - s_3) P_{\mathbf{x},s}(U_3 = n_3 - s_2 - s_3) \cdot \\
& \quad \cdot P_{\mathbf{x},s}(U_{12} = s_1) P_{\mathbf{x},s}(U_{13} = s_2) P_{\mathbf{x},s}(U_{23} = s_3) = \\
& = \sum_{s_1=0}^{r_1} \sum_{s_2=0}^{r_2} \sum_{s_3=0}^{r_3} f^{NB-I}(n_1 - s_1 - s_2 | \mathbf{x}^1, \boldsymbol{\beta}_1, \sigma^2) f^{NB-I}(n_2 - s_1 - s_3 | \mathbf{x}^2, \boldsymbol{\beta}_2, \sigma^2) \cdot \\
& \quad \cdot f^{NB-I}(n_3 - s_2 - s_3 | \mathbf{x}^3, \boldsymbol{\beta}_3, \sigma^2) f^{NB-I}(s_1 | 1, \log \lambda_{12}, \sigma^2) \cdot \\
& \quad \cdot f^{NB-I}(s_2 | 1, \log \lambda_{13}, \sigma^2) f^{NB-I}(s_3 | 1, \log \lambda_{23}, \sigma^2),
\end{aligned}$$

kde $P_{\mathbf{x},s}$ značí pravděpodobnost za podmínky znalosti regresní matice \mathbf{x} a za podmínky $[U_{12} = s_1, U_{13} = s_2, U_{23} = s_3]$.

Další rozšíření trojrozměrných negativně binomických modelů již uvádět nebudeme. Nyní přejdeme od modelování závislosti mezi složkami náhodných veličin pomocí šokových proměnných na modelování závislosti pomocí kopulí.

3. Modely s kopulemi

V této kapitole si zavedeme pojem kopule a společně s ním uvedeme Sklarovu větu, která se následně využívá i při definování rodiny gaussovských kopulí. Jako další příklad rodiny kopulí uvedeme rodinu archimédovských kopulí, které jsou generovány pomocí tzv. generátoru kopulí. Dále si uvedeme míry závislostí a základní odhady parametrů kopulí. Pro tuto kapitolu budeme čerpat z McNeil (2005) a Nelsen (2006).

Kopule jsou vícerozměrné sdružené distribuční funkce, které popisují závislostní strukturu mezi náhodnými veličinami. Jejich marginální distribuční funkce jsou rovnoměrné na intervalu $[0, 1]$. Přesná definice je uvedena v definici 10.

Definice 10. *Funkce $C(u_1, u_2, \dots, u_d) : [0, 1]^d \rightarrow [0, 1]$ je kopule, pokud je to sdružená distribuční funkce nějakého d -rozměrného rozdělení a všechna marginální rozdělení jsou rovnoměrná na intervalu $[0, 1]$.*

Základní vlastnost propojující kopuli a sdruženou distribuční funkci d -rozměrného rozdělení je popsána ve Sklarově větě.

Věta 1 (Sklarova věta). *Nechť $F(x_1, \dots, x_d)$ je sdružená distribuční funkce d -rozměrného rozdělení s marginálními distribučními funkcemi $F_1(x_1), \dots, F_d(x_d)$. Potom existuje kopule C taková, že pro každý vektor $(x_1, \dots, x_d)^T \in [-\infty, \infty]^d$ platí*

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)) \quad (3.1)$$

Pokud jsou F_1, \dots, F_d distribuční funkce spojitého rozdělení, potom je kopule C definována jednoznačně. Pokud nejsou distribuční funkce F_1, \dots, F_d absolutně spojitě, tak je kopule C definována jednoznačně pouze na množině

$$\text{Ran}(F_1) \times \dots \times \text{Ran}(F_d),$$

kde $\text{Ran}(F_i), i = 1, \dots, d$, je obor hodnot distribuční funkce F_i .

Důkaz. viz. McNeil (2005) – str. 187.

3.1 Základní kopule

V této části si uvedeme nejprve vlastnosti kopulí a dále příklady základních kopulí, které budeme dále využívat v kapitole 4.

Marginální rozdělení $F_1(\cdot), \dots, F_d(\cdot)$ mohou být spojitá i diskrétní. Hlavním rozdílem je jednoznačnost kopule definované ze Sklarovy věty. Zatímco pro spojitě rozdělení náhodných veličin je kopule definovaná ze sdružené distribuční funkce jednoznačně, pro diskrétní náhodné veličiny je jednoznačnost zaručena pouze na určité množině hodnot.

Pro spojitě náhodné veličiny X a Y platí, že nezávislost těchto dvou náhodných veličin je jednoznačně popsána, pokud jejich příslušná kopule definovaná pomocí Sklarovy věty je kopulí nezávislosti. Tedy pro $u, v \in (0, 1)$ platí

$$X \perp Y \Leftrightarrow C(u, v) = \Pi(u, v) = uv$$

Obdobně X a Y jsou komonotónní (resp. kontramonotónní) náhodné veličiny právě tehdy, když jejich kopule dle Sklarovy věty je kopulí komonotonie – kopule horní Fréchetovy meze M (resp. kontramonotonie – kopule dolní Fréchetovy meze W):

$$\begin{aligned} M(u_1, \dots, u_d) &= \min(u_1, \dots, u_d), \\ W(u_1, u_2) &= [u_1 + u_2 - 1]_+, \end{aligned}$$

kde $[\cdot]_+$ značí kladnou část, tedy $\max(\cdot, 0)$. Kopule dolní Fréchetovy meze je kopulí pouze pro dvourozměrnou kopuli.

Poznámka 1. Náhodné veličiny X_1, \dots, X_d jsou *komonotónní* právě tehdy, když náhodný vektor (X_1, \dots, X_d) má stejné rozdělení jako vektor $(\nu_1(Z), \dots, \nu_d(Z))$, kde Z je libovolná náhodná veličina a $\nu_i, i = 1, \dots, d$ jsou neklesající funkce.

Poznámka 2. Náhodné veličiny X_1, X_2 jsou *kontramonotónní* právě tehdy, když sdružené rozdělení náhodných veličin X_1 a X_2 je stejné jako rozdělení náhodného vektoru $(\nu_1(Z), \nu_2(Z))$, kde Z je libovolná náhodná veličina, ν_1 je neklesající funkce a ν_2 je nerostoucí funkce, nebo naopak.

Zatímco pro spojitě náhodné veličiny je ve výše zmíněných případech ekvivalence, tak pro diskrétní náhodné veličiny tyto ekvivalence neplatí. Platí pouze následující implikace

$$\begin{aligned} C(u, v) = \Pi(u, v) &\Rightarrow X \perp Y, \\ C(u, v) = M(u, v) &\Rightarrow X \text{ a } Y \text{ jsou komonotónní,} \\ C(u, v) = W(u, v) &\Rightarrow X \text{ a } Y \text{ jsou kontramonotónní.} \end{aligned}$$

V případě nespojitých náhodných veličin nejsou kopule popsány jednoznačně, náhodné veličiny mohou být nezávislé, ale jejich kopule nemusí nutně být kopulí nezávislosti.

Dále si uvedeme dvě rodiny kopulí – gaussovskou a archimédovskou rodinu.

3.1.1 Gaussovské kopule

Jedná se o rodinu implicitních kopulí, které jsou definovány na základě Sklarovy věty z d -rozměrného normálního rozdělení.

Náhodný vektor $\mathbf{X} = (X_1, \dots, X_d)^\top$ má d -rozměrné normální rozdělení s hustotou

$$f(\mathbf{x}) = \frac{1}{(2\pi)^d |\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\},$$

kde $\mathbf{x} = (x_1, \dots, x_d)^\top$, $\boldsymbol{\mu}$ je vektor středních hodnot a Σ je kovarianční matice vektoru \mathbf{X} , $|\Sigma|$ značí její determinant. Zkráceně píšeme $\mathbf{X} \sim N_d(\boldsymbol{\mu}, \Sigma)$.

Vektor \mathbf{X} normujeme na vektor $\mathbf{Y} = (Y_1, \dots, Y_d)$ transformací

$$Y_i = \frac{X_i - \mu_i}{\sqrt{\Sigma_{ii}}}, \quad i = 1, \dots, d,$$

provedenou na jednotlivé složky vektoru \mathbf{X} . Získáme vektor \mathbf{Y} , jehož složky mají $N(0, 1)$ rozdělení. Protože jsme použili monotónní transformace, vektor \mathbf{Y} má stejnou kopuli jako vektor \mathbf{X} .

Gaussovská kopule je kopule náhodného vektoru \mathbf{Y} , který má d-rozměrné normální rozdělení s nulovým vektorem středních hodnot a jednotkovým rozptylem jednotlivých složek vektoru \mathbf{Y} , definovaná jako

$$\begin{aligned} C_R^{Ga}(u_1, \dots, u_d) &= P \left[Y_1 \leq \Phi^{-1}(u_1), \dots, Y_d \leq \Phi^{-1}(u_d) \right] = \\ &= \Phi_R(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d)), \end{aligned}$$

kde $\Phi(\cdot)$ je distribuční funkce jednorozměrného normálního rozdělení $N(0, 1)$, R je korelační matice vektoru \mathbf{Y} , $\Phi_R(\cdot)$ je sdružená distribuční funkce vektoru \mathbf{Y} , $(u_1, \dots, u_d) \in [0, 1]^d$.

Kopule nezávislosti

Speciálním případem gaussovské kopule je kopule nezávislosti pro $\Sigma = I_d$

$$C_{I_d}^{Ga}(u_1, \dots, u_d) = u_1 \cdot \dots \cdot u_d.$$

3.1.2 Archimédovské kopule

Rodinou archimédovských kopulí jsou nazývány kopule, které vzniknou pomocí funkce $\varphi : [0, 1] \rightarrow \mathbb{R}$, která se nazývá *generátor kopule*, následujícím způsobem

$$C(u_1, \dots, u_d) = \varphi^{-1}(\varphi(u_1) + \dots + \varphi(u_d)). \quad (3.2)$$

V případě, kdy inverzní funkce φ^{-1} není jednoznačně definovaná, tak se používá tzv. pseudoinverzní funkce, která je definována v definici 11.

Definice 11. Nelsen (2006) *Nechť φ je spojitá klesající funkce z $[0, 1]$ do $[0, \infty]$ taková, že $\varphi(1) = 0$. Pseudoinverzní funkce k funkci φ je funkce $\varphi^{[-1]} : [0, \infty] \rightarrow [0, 1]$ definovaná jako*

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

V případě, že je $\varphi(0) = \infty$, je $\varphi^{[-1]} = \varphi^{-1}$.

Archimédovské kopule jsou symetrické, tedy platí pro ně vztah

$$C(u, v) = C(v, u),$$

tuto vlastnost je třeba brát v potaz při používání archimédovských kopulí při modelování závislosti.

Nyní si uvedeme příklady základních archimédovských kopulí.

Claytonova kopule

Generátorem kopule je pro Claytonovu kopuli funkce

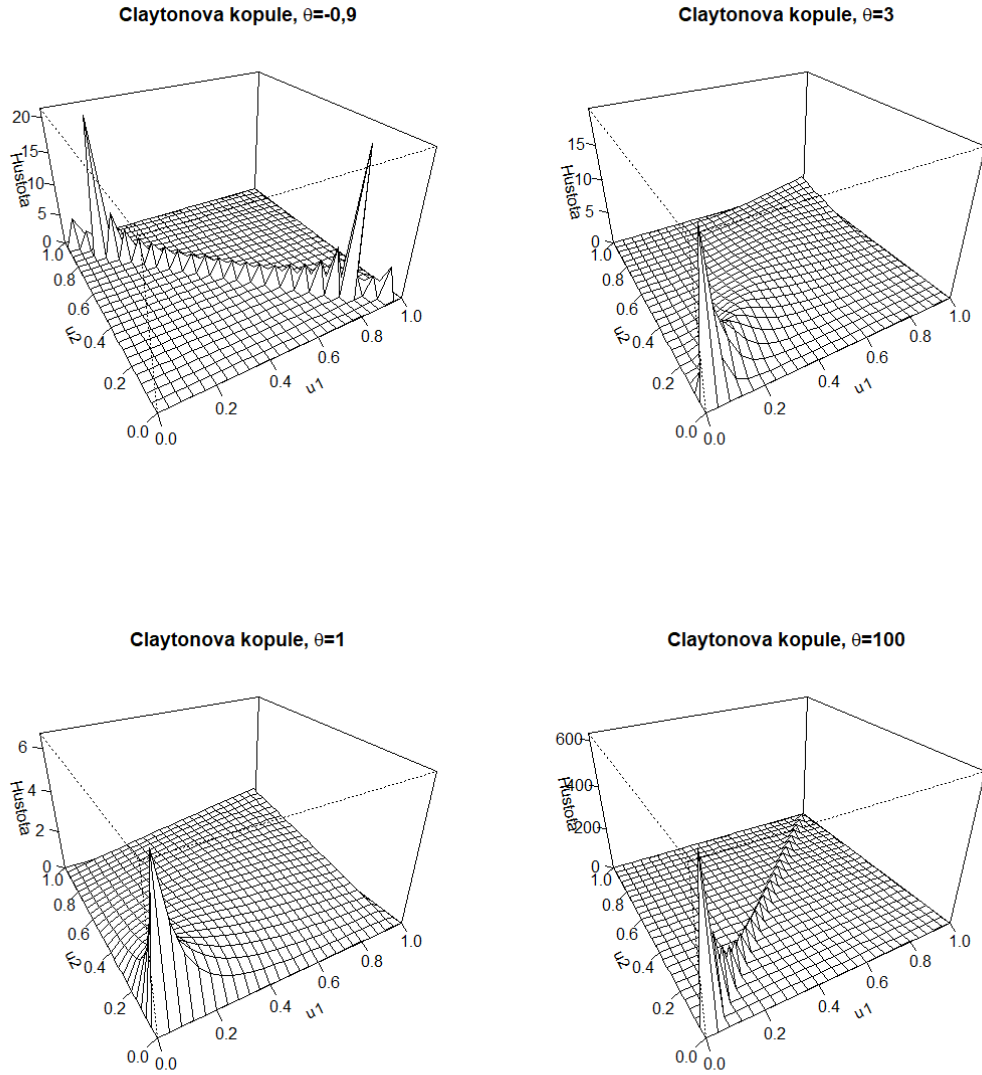
$$\varphi(u) = \frac{1}{\theta} (u^{-\theta} - 1),$$

pomocí které následně vznikne v dvourozměrném případě kopule

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

kde $\theta \in [-1, \infty] \setminus \{0\}$.

Pro hodnoty parametru θ blíží se k nule získáme kopuli nezávislosti, pro $\theta = -1$ získáme kopuli dolní Fréchetovy meze a pro θ blíží se nekonečnu získáme kopuli horní Fréchetovy meze. Na obrázku 3.1 jsou zobrazeny hustoty kopule pro různé hodnoty parametru θ .



Obrázek 3.1: Znáornění hustoty Claytonovy kopule pro různé parametry θ .

Gumbelova kopule

Generátor pro Gumbelovu kopuli je funkce

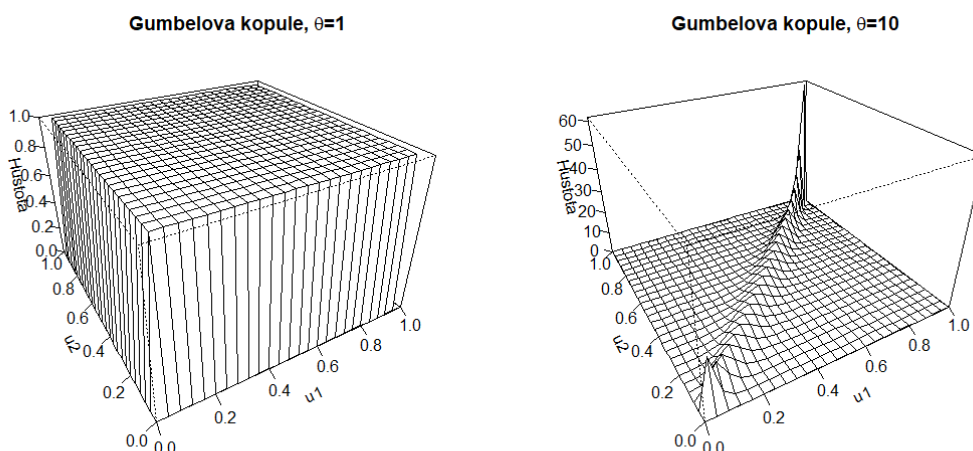
$$\varphi(u) = (-\log u)^\theta.$$

V dvourozměrném případě vznikne kopule

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

kde $1 \leq \theta < \infty$.

Pro hodnotu parametru $\theta = 1$ přechází Gumbelova kopule do tvaru kopule nezávislosti – viz obrázek 3.2.



Obrázek 3.2: Znárodnění hustoty Gumbelovy kopule pro různé parametry θ .

3.2 Míry závislosti

Míru závislosti mezi dvěma náhodnými veličinami je možné vyjádřit různými způsoby. Jako základní koeficient, který popisuje lineární závislost, si uvedeme Pearsonův korelační koeficient. Dále definujeme pořadové korelace a koeficienty závislosti chvostů, což jsou hodnoty odvozené na základě kopulí, proto postihují i jiné než lineární závislosti. Vzhledem k tomu, že kopule zachycuje závislost mezi náhodnými veličinami, je úzce spjata s pojmem korelace, a proto zde uvádíme dvě míry závislosti, které jsou odvozeny pomocí kopule daných náhodných veličin.

3.2.1 Pearsonův korelační koeficient

Základní korelační koeficient pro zachycení lineární závislosti je definován jako hodnota korelace mezi náhodnými veličinami X_1 a X_2

$$\rho(X_1, X_2) = \frac{\text{cov}(X_1, X_2)}{\sqrt{\text{var}(X_1) \text{var}(X_2)}},$$

kde rozptyly náhodných veličin X_1 a X_2 jsou konečné. Často tento korelační koeficient také bývá označován jako lineární korelační koeficient. Korelační koeficient nabývá hodnot $[-1, 1]$, kde hodnota -1 značí perfektní negativní korelaci a hodnota 1 perfektní pozitivní korelaci. Pro vzájemně nezávislé náhodné veličiny X_1 a X_2 je korelace nulová. Pokud je však korelace mezi náhodnými veličinami nulová, nemusí to nutně znamenat, že jsou vzájemně nezávislé.

Lineární korelace závisí jednak na kopuli dvourozměrného rozdělení, protože se v definici korelačního koeficientu vyskytuje vzájemná kovariance náhodných

veličin, a jednak na marginálním rozdělení náhodných veličin X_1 a X_2 , protože využíváme rozptyl těchto náhodných veličin.

3.2.2 Pořadové korelace

Na rozdíl od lineární korelace pořadové korelace závisí pouze na hodnotách dvourozměrné kopule a nejsou závislé na marginálních rozděleních náhodných veličin X_1 a X_2 . Jak již název napovídá, pro vyjádření pořadových korelací budeme využívat pořadí v seřazeném náhodném výběru. Uvedeme si dvě pořadové korelace, Kendallovo tau (někdy též označováno jako ρ_τ) a Spearmanovo rhó (označováno ρ_S).

Abychom mohli vyjádřit význam Kendallova tau, zavedeme si nejprve dva pojmy – *konkordance* a *diskordance*.

Definice 12. *Body (x_1, x_2) a $(\tilde{x}_1, \tilde{x}_2)$ z \mathbb{R}^2 jsou konkordantní, pokud*

$$(x_1 - \tilde{x}_1)(x_2 - \tilde{x}_2) > 0.$$

Pokud platí opačná nerovnost, tak řekneme, že tyto dva body jsou diskordantní.

Kendallovo tau

Kendallovo tau je mírou konkordance náhodných vektorů (X_1, X_2) a $(\tilde{X}_1, \tilde{X}_2)$, kde náhodný vektor $(\tilde{X}_1, \tilde{X}_2)$ je nezávislou kopií náhodného vektoru (X_1, X_2) – jejich rozdělení je stejné, ale jsou navzájem nezávislé.

Čím více jsou hodnoty složek náhodných vektorů provázány ve stejném směru (se zvětšující se hodnotou náhodné veličiny X_1 se zvětšuje i hodnota náhodné veličiny X_2), tím větší jsou hodnoty Kendallova tau.

Definice 13. *Pro náhodné veličiny X_1, X_2 je hodnota Kendallova tau definována jako*

$$\rho_\tau(X_1, X_2) = E \left[\text{sign} \left[(X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2) \right] \right],$$

kde $(\tilde{X}_1, \tilde{X}_2)$ je nezávislá kopie náhodného vektoru (X_1, X_2)

Kendallovo tau je možné vyjádřit jako rozdíl pravděpodobností konkordance náhodných vektorů (X_1, X_2) a $(\tilde{X}_1, \tilde{X}_2)$ a diskordance těchto vektorů

$$\rho_\tau(X_1, X_2) = P \left[(X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2) > 0 \right] - P \left[(X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2) < 0 \right].$$

Spearmanovo rhó

Spearmanovo rhó je možné také definovat pomocí pojmů konkordance a diskordance, ale vhodnější je definice uvedená v definici 14.

Definice 14. *Pro náhodné veličiny X_1 a X_2 s marginálními distribučními funkcemi F_1 a F_2 je Spearmanovo rhó definováno jako*

$$\rho_S(X_1, X_2) = \rho(F_1(X_1), F_2(X_2)),$$

kde $\rho(\cdot, \cdot)$ je lineární korelační koeficient.

Vlastnosti pořadových korelací

V obou případech hodnoty korelací jsou z intervalu $[-1, 1]$. Hodnoty 1 nabývají pro komonotónní náhodné veličiny a hodnoty -1 pro kontramonotónní veličiny. Pokud jsou dvě náhodné veličiny nezávislé, tak hodnota korelací je 0, zatímco opačná implikace neplatí, tedy nulovost korelací neimplikuje nezávislost.

Pro obě korelace také platí, že pro spojitá marginální rozdělení lze pořadové korelace vyjádřit pomocí jednoznačně určené kopule. Pro spojitě náhodné veličiny s kopulí C platí

$$\rho_\tau(X_1, X_2) = 4 \int_0^1 \int_0^1 C(u_1, u_2) dC(u_1, u_2) - 1, \quad (3.3)$$

$$\rho_S(X_1, X_2) = 12 \int_0^1 \int_0^1 (C(u_1, u_2) - u_1 u_2) du_1 du_2. \quad (3.4)$$

Důkaz těchto tvrzení je možné najít například v McNeil (2005) na straně 207.

3.2.3 Koeficienty závislosti chvostů

Koeficienty závislosti chvostů měří závislost mezi dvěma náhodnými veličinami pouze pomocí kopule. Tyto koeficienty měří závislost na chvostech dvourozměrného rozdělení.

Definice 15. (McNeil (2005)) *Nechť X_1 a X_2 jsou náhodné veličiny s marginálními distribučními funkcemi F_1 a F_2 . Koeficient horní závislosti chvostu je definován jako*

$$\lambda_U(X_1, X_2) = \lim_{q \rightarrow 1^-} P \left[X_2 > F_2^{-1}(q) | X_1 > F_1^{-1}(q) \right],$$

pokud limita $\lambda_U \in [0, 1]$ existuje.

Koeficient dolní závislosti chvostu je definován jako

$$\lambda_L(X_1, X_2) = \lim_{q \rightarrow 0^+} P \left[X_2 \leq F_2^{-1}(q) | X_1 \leq F_1^{-1}(q) \right],$$

pokud limita $\lambda_L \in [0, 1]$ existuje.

Pokud jsou F_1 a F_2 spojitě distribuční funkce, je možné vyjádřit koeficienty λ_U a λ_L pouze pomocí jednoznačně určené kopule dvourozměrného rozdělení

$$\lambda_L = \lim_{q \rightarrow 0^+} \frac{C(q, q)}{q},$$

$$\lambda_U = \lim_{q \rightarrow 0^+} \frac{\widehat{C}(q, q)}{q},$$

kde $\widehat{C}(q, q)$ je kopule přežití příslušná kopuli $C(q, q)$.

3.3 Odhad parametrů kopulí

Máme k dispozici d-rozměrné vektory dat $\mathbf{X}_1, \dots, \mathbf{X}_n$, kde $\mathbf{X}_t = (X_{t,1}, \dots, X_{t,d})^\top$, $t = 1, \dots, n$. Předpokládáme, že všechny vektory mají stejnou distribuční funkci $F(x_1, \dots, x_d)$.

Odhad parametrů kopule je možné provést například pomocí maximálně věrohodného odhadu nebo pomocí pořadových korelací.

3.3.1 Pořadové korelace

Odhad Spearmanova korelačního koeficientu je možné vyjádřit následujícím způsobem

$$\widehat{\rho}_S(X_{t,i}, X_{t,j}) = \frac{12}{n(n^2-1)} \sum_{t=1}^n \left[\text{rank}(X_{t,i}) - \frac{1}{2}(n+1) \right] \left[\text{rank}(X_{t,j}) - \frac{1}{2}(n+1) \right],$$

kde jsme využili faktu, že pořadí $\text{rank}(X_{t,i})$ má diskrétní rovnoměrné rozdělení na množině $\{1, \dots, n\}$ se střední hodnotou $\frac{n+1}{2}$ a rozptylem $\frac{(n^2-1)}{12}$.

Odhad Kendallova tau lze pomocí pozorování vyjádřit jako

$$\widehat{\rho}_\tau(X_{t,i}, X_{t,j}) = \binom{n}{2}^{-1} \sum_{1 \leq t < s \leq n} \text{sign}[(X_{t,i} - X_{s,i})(X_{t,j} - X_{s,j})].$$

Následně je možné odhadnuté pořadové korelace dosadit do rovnice 4.2 nebo 3.4, odkud pak získáme odhad parametrů kopule.

3.3.2 Maximálně věrohodný odhad

Máme kopuli C_θ s vektorem neznámých parametrů θ . Maximálně věrohodný odhad parametrů θ je získán jako argument maxima logaritmu věrohodnostní funkce, která má v tomto případě tvar

$$\log L(\theta, \widehat{U}_1, \dots, \widehat{U}_n) = \sum_{t=1}^n \log c_\theta(\widehat{U}_t), \quad (3.5)$$

kde c_θ je hustota kopule C_θ a \widehat{U}_t jsou pseudo-pozorování.

Odhady pomocí MLE se doporučuje provádět ve dvou krocích, kdy nejprve jsou odhadnuty jednorozměrné marginální distribuční funkce, případně se použije empirická distribuční funkce $\widetilde{F}_i(x) = \frac{1}{n+1} \sum_{t=1}^n I[X_{t,i} \leq x]$. A až v druhém kroku jsou odhadnuty parametry kopule C_θ pomocí maximalizace 3.5.

4. Porovnání modelů na simulovaných datech

V této kapitole si ukážeme porovnání některých modelů z předchozích kapitol na námi nasimulovaných datech. Nejprve si vygenerujeme řádky regresní matice \mathbb{X} značící jednotlivé pojistky v portfoliu, ze kterých budeme dále simulovat počty škod pro dvě různé garance. Celkový počet škod N pro jednu pojistku je součtem počtů škod z dvou dílčích garancí – N^1 a N^2 .

Porovnáme směrodatné odchylky z nasimulovaných počtů škod pro případ, kdy budeme uvažovat nezávislost mezi jednotlivými garancemi, a pro vzájemně závislé garance. Očekáváme, že výsledná hodnota pro závislé garance by měla být vyšší než v případě nezávislosti.

Označíme π střední hodnotu celkového počtu škod z obou garancí, π_1 střední hodnotu počtu škod z první garance, π_2 střední hodnotu počtu škod z druhé garance, θ_1 , θ_2 směrodatné odchylky pro dané garance. Máme k dispozici následující vztahy

$$\begin{aligned} N &= N^1 + N^2, & \theta_1 &= \sqrt{\text{var } N^1}, \\ \pi_1 &= \mathbb{E}[N^1], & \theta_2 &= \sqrt{\text{var } N^2}, \\ \pi_2 &= \mathbb{E}[N^2], & \theta &= \sqrt{\text{var } N}. \\ \pi &= \pi_1 + \pi_2, \end{aligned} \tag{4.1}$$

4.1 Generování dat

Pro generování počtů škod použijeme v nule modifikované dvourozměrné Poissonovo rozdělení. Následné porovnání vypočteného pojistného provedeme pomocí poissonovských modelů a za pomoci kopulí.

Nejprve budeme generovat regresní matici. Zvolíme následující popisné veličiny: pohlaví, věk a typ vozidla. Pro pohlaví uvažujeme hodnoty muž a žena, pro typ vozidla uvažujeme tři typy vozidel – osobní automobil označovaný jako OA, nákladní automobil označovaný jako NA a motorku označovanou jako MO. Pro popis pohlaví použijeme jednu tarifní faktorovou proměnnou v_1 , jejíž hodnota 1 označuje ženu a hodnota 0 muže. Věk nabývá hodnot od 18 do 75 let a je zaznamenán v proměnné v_2 , nebudeme uvažovat možné řidiče motorek od 15 do 18 let. Pro popis typu vozidla použijeme dvě tarifní faktorové proměnné v_3, v_4 , pokud je v_3 rovna 1, jedná se o OA, pokud je v_4 rovna 1, jedná se o NA a v ostatních případech se jedná o motorku. Tarifní proměnná v_0 je intercept, který zachycuje muže ve věku 18 let s typem vozidla motorka.

Budeme předpokládat, že pravděpodobnost výskytu jednotlivých hodnot popisných veličin je stejná. Pravděpodobnost výskytu ženy je tedy stejná jako pravděpodobnost výskytu muže, a to 1/2, pravděpodobnost pro každý typ vozidla je 1/3 a pravděpodobnost jednotlivých věků je 1/58.

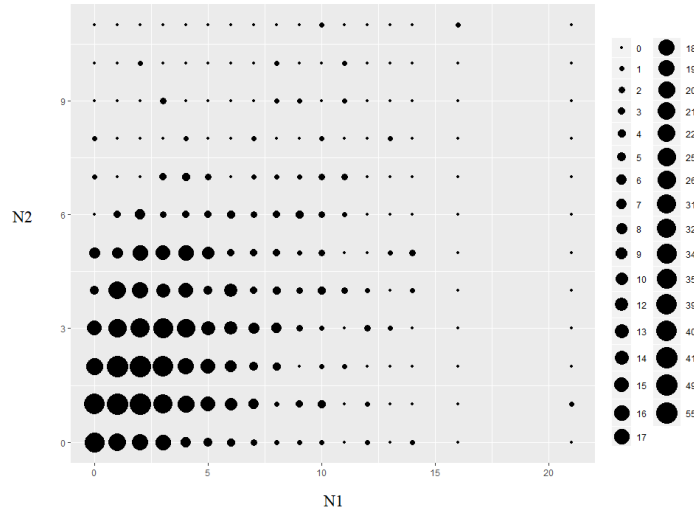
Velikost datového souboru označíme n a budeme předpokládat, že máme k dispozici 1000 pozorování, tedy $n = 1000$. Regresní matice \mathbb{X} má rozměry 1000×5 . V tabulce 4.1 je vidět prvních 8 řádků regresní matice \mathbb{X} . První řádek matice popisuje pojistku, kde je pojistníkem žena ve věku 19 let, která si pojišťuje motorku. V druhém pozorování, máme k dispozici taktéž ženu ve věku 59 let, ale

pojištěné vozidlo je nákladní automobil. Takto máme celkem 1000 řádků matice.

Řádek	v_0	v_1	v_2	v_3	v_4
1	1	1	19	0	0
2	1	1	59	0	1
3	1	0	25	0	1
4	1	0	74	0	0
5	1	1	33	0	1
6	1	0	62	1	0
7	1	0	68	1	0
8	1	0	50	0	1
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots

Tabulka 4.1: Prvních 8 řádků regresní matice pozorování \mathbb{X} .

K zavedeným tarifním proměnným zafixujeme vektory hodnot koeficientů, podle kterých následně sestrojíme počty škod. Budeme uvažovat dva typy garancí, náhodný výběr $\mathbf{N}^1 = (N_1^1, \dots, N_{1000}^1)$ o velikosti 1000 bude zaznamenávat počty škod z povinného ručení a náhodný výběr $\mathbf{N}^2 = (N_1^2, \dots, N_{1000}^2)$ o velikosti 1000 bude zaznamenávat počty škod z ostatních garancí.



Obrázek 4.1: Počty škod pro dané dvojice hodnot (N^1, N^2) s rozlišením zastoupení jednotlivých hodnot dle velikosti puntíků.

Budeme generovat data z dvourozměrného Poissonova rozdělení pomocí společných šoků, jak bylo uvedeno v (1.4). Pouze s tím rozdílem, že jako tři nezávislé vektory Y_0, Y_1, Y_2 využijeme náhodné veličiny s Poissonovým rozdělením, které je v nule modifikované. Zvolíme následující vektory regresních parametrů

$$\begin{aligned} \beta^0 &= (\beta_0^0, \beta_1^0, \beta_2^0, \beta_3^0, \beta_4^0) = (-1,00; 0,30; 0,015; 0,25; 0,33), \\ \beta^1 &= (\beta_0^1, \beta_1^1, \beta_2^1, \beta_3^1, \beta_4^1) = (-1,20; 0,50; 0,03; 0,70; 0,40), \\ \beta^2 &= (\beta_0^2, \beta_1^2, \beta_2^2, \beta_3^2, \beta_4^2) = (-0,80; 0,40; 0,02; 0,40; 0,20). \end{aligned}$$

Vektory parametrů $\boldsymbol{\lambda}^0, \boldsymbol{\lambda}^1, \boldsymbol{\lambda}^2$ dostaneme pro $j = 0, 1, 2$ pomocí vzorce

$$\boldsymbol{\lambda}^j = \exp\{\mathbf{X} \boldsymbol{\beta}^j\},$$

kde $\exp\{\mathbf{X} \boldsymbol{\beta}^j\}$ označuje exponenciálu provedenou na složky vektoru $\mathbf{X} \boldsymbol{\beta}^j$.

První čtyři řádky hodnot parametrů $\lambda_i^0, \lambda_i^1, \lambda_i^2, i = 1, \dots, n$ jsou zobrazeny v tabulce 4.2.

i	λ_i^0	λ_i^1	λ_i^2
1	0,66	0,87	0,98
2	1,67	4,35	2,66
3	0,74	0,95	0,90
4	1,12	2,77	1,97
\vdots	\vdots	\vdots	\vdots
výběrový průměr	1,11	2,79	1,88

Tabulka 4.2: První čtyři hodnoty parametrů $\boldsymbol{\lambda}^j, j = 0, 1, 2$ a jejich výběrové průměry.

Pomocí parametrů $\lambda_i^0, \lambda_i^1, \lambda_i^2$ budeme generovat počty škod náhodných vektorů $\mathbf{Y}_0, \mathbf{Y}_1, \mathbf{Y}_2$ z v nule modifikovaného Poissonova rozdělení s parametrem modifikace v nule $p = 0,34$. Následné počty škod z uvažovaných garancí \mathbf{N}^1 a \mathbf{N}^2 dostaneme ze vztahu (1.4).

Charakteristika	\mathbf{N}^1	\mathbf{N}^2
Minimum	0	0
Maximum	17	15
Medián	2	2
Výběrový průměr	3,09	2,60
Směrodatná odchylka	2,69	2,07

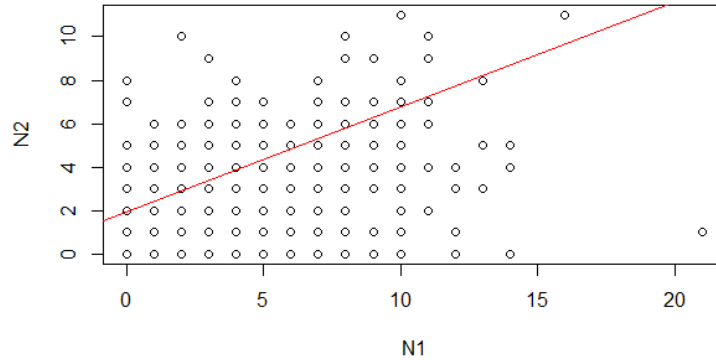
Tabulka 4.3: Základní charakteristiky náhodných výběrů \mathbf{N}^1 a \mathbf{N}^2 .

Základní charakteristiky náhodných výběrů \mathbf{N}^1 a \mathbf{N}^2 jsou uvedeny v tabulce 4.3. Na obrázku 4.1 jsou zobrazeny počty škod pro dvojice počtů škod N^1 a N^2 .

Mezi náhodnými veličinami \mathbf{N}^1 a \mathbf{N}^2 jsme pomocí náhodné veličiny Y_0 vytvořili závislost. Na obrázku 4.2 je vyobrazen vztah mezi \mathbf{N}^1 a \mathbf{N}^2 společně s regresní přímkou, která popisuje lineární vztah mezi \mathbf{N}^1 a \mathbf{N}^2 . V tabulce 4.4 jsou uvedeny průměrné výběrové korelační koeficienty pro dvojice N^1 a N^2 .

Název	Označení	Hodnota
Pearsonův korelační koeficient	ρ	0,3273
Spearmanův korelační koeficient	ρ_S	0,2953
Kendallové tau	ρ_τ	0,2322

Tabulka 4.4: Různé korelační koeficienty pro vzájemnou korelaci mezi N^1 a N^2 .



Obrázek 4.2: Závislost N^1 na N^2 společně s regresní přímkou zachycující lineární závislost mezi náhodnými výběry.

4.2 Porovnání poissonovských modelů na datech

Nasimulované počty škod N^1 a N^2 zároveň s regresní maticí \mathbb{X} použijeme na odhady parametrů. Porovnáme modely, které uvažují vztah mezi N^1 a N^2 – dvourozměrný Poissonův model BPD a na diagonále modifikovaný dvourozměrný Poissonův model DIBPD, a modely, které tento vztah neuvažují, resp. předpokládají nezávislost mezi N^1 a N^2 – dvojitý Poissonův model DPD a dvojitý na diagonále modifikovaný Poissonův model DIDPD. Očekáváme, že nejlepší odhad dostaneme pro model dvourozměrného na diagonále modifikovaného rozdělení, protože jsme při vytváření dat využili modifikaci v nule.

Odhady provedeme pomocí funkcí *lm.bp* a *lm.dibp* z balíčku *bivpois* v softwaru R. Obě funkce pro odhad parametrů využívají EM algoritmus, který je popsán v poznámce 3. Omezíme běh algoritmu na maximální počet 300 iterací nebo přesnost 10^{-6} podle toho, co nastane dříve.

Poznámka 3 (EM algoritmus). Pro tuto část je čerpáno ze zápisků k přednášce NMST434 - Moderní statistické metody, Omelka (2019). EM algoritmus je zkratkou pro *Expectation-maximization algorithm*.

EM algoritmus se používá pro odhad neznámých parametrů v případě, kdy nejsou všechny náhodné veličiny pozorovány. Pozorovaná data jsou brána jako realizace náhodných vektorů $(X_1, \mathbf{Z}_1^\top)^\top, \dots, (X_n, \mathbf{Z}_n^\top)^\top$, kde $\mathbb{X} = (X_1, \dots, X_n)$ jsou pozorované hodnoty a vektory $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ nejsou pozorovány a označují se za chybějící pozorování. Předpokládá se, že rozdělení nepozorovaných náhodných vektorů nezávisí na odhadovaných parametrech pozorovaných náhodných veličin.

Věrohodnostní funkce je sdružená distribuční funkce X a \mathbf{Z} , jelikož však nejsou známy hodnoty \mathbf{Z} , není možné použít obyčejnou metodu maximální věrohodnosti. Pro hledání argumentu maxima se využívá logaritmus věrohodnostní funkce. Postupuje se iteračně, kdy v každé iteraci proběhnou dva kroky.

V $(k + 1)$ -ní iteraci v prvním kroku je spočítána střední hodnota vzhledem k nepozorovaným náhodným vektorům z logaritmické věrohodnostní funkce sdruženého rozdělení za podmínky, že jsou známy pozorované hodnoty \mathbb{X} , a ta je označena Q ,

$$Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k)}) = \mathbb{E}_{\hat{\boldsymbol{\theta}}^{(k)}} [l(\boldsymbol{\theta}) | \mathbb{X}],$$

kde θ je vektor neznámých parametrů, $\hat{\theta}^{(k)}$ je hodnota odhadnutých parametrů v k -tém iteračním kroku, se kterou se pracuje ve střední hodnotě.

Druhý krok je maximalizační, kdy je spočítána nová lepší hodnota odhadu parametrů a je označena jako $\hat{\theta}^{(k+1)}$,

$$\hat{\theta}^{(k+1)} = \operatorname{argmax}_{\theta \in \Theta} Q(\theta, \hat{\theta}^{(k)}).$$

V iteracích se pokračuje, dokud není splněna jedna z ukončovacích podmínek, a to buď je překročen maximální počet iterací, nebo je dosažena požadovaná přesnost. Iterační algoritmus dokonverguje k maximálně věrohodnému odhadu parametrů.

Odhady parametrů uvažovaných modelů jsou zobrazeny v tabulce 4.5. Pro modely, které jsou na diagonále modifikovány, budeme uvažovat za modifikující rozdělení postupně diskrétní rozdělení s jedním parametrem (v nule modifikované rozdělení), diskrétní rozdělení s dvěma parametry (rozdělení na dvou bodech – (0, 0) a (1, 1)) a Poissonovo rozdělení. Pro každý z těchto modelů dostáváme navíc odhad parametrů modifikujícího rozdělení a odhad parametru p . Parametr jedno-parametrického diskrétního rozdělení označíme θ_0 , parametry dvou-parametrického diskrétního rozdělení označíme θ_0 a θ_1 , parametr Poissonova rozdělení označíme θ_0 . V případě, že daný parametr není součástí rozdělení, tak jeho hodnota není uvedena.

Koef.	DPD	DIDPD-D2	BPD	DIBPD-D1	DIBPD-D2	DIBPD-PO
$\widehat{\lambda}_0$			0,73	0,68	0,68	0,68
$\widehat{\lambda}_1$	3,16	3,33	2,43	2,62	2,63	2,63
$\widehat{\lambda}_2$	2,53	2,65	1,80	1,94	1,95	1,95
$\widehat{\theta}_0$				0,41	0,40	0,59
$\widehat{\theta}_1$					0,11	
\widehat{p}		0,06		0,04	0,06	0,04

Tabulka 4.5: Porovnání odhadů parametrů pro různé modely.

	DPD	DIDPD-D2	BPD	DIBPD-D1	DIBPD-D2	DIBPD-PO
AIC	8296,17	8162,64	8180,94	8089,89	8091,80	8090,53
BIC	8352,18	8235,45	8242,55	8162,70	8170,22	8163,34

Tabulka 4.6: Informační kritéria AIC a BIC pro uvažované modely – nejnižší hodnoty jsou vyznačeny tučně.

V tabulce 4.6 jsou shrnuta informační kritéria AIC a BIC pro všechny uvažované modely. Jak jsme očekávali, tak jako nejlepší model bylo zvoleno na diagonále modifikované dvourozměrné Poissonovo rozdělení s jedním parametrem. Tento model zachycuje jak modifikaci v nule, kterou jsme při tvorbě dat použili, tak závislostní vztah mezi složkami vektorů \mathbf{N}^1 a \mathbf{N}^2 .

Nyní se zaměříme na porovnání součtu očekávaného počtu škod π a směrodatné odchylky θ pro odhadnuté modely. Využijeme vzorců uvedených v 4.1.

Pokud uvažujeme nezávislost mezi složkami náhodných vektorů \mathbf{N}^1 a \mathbf{N}^2 , tak dostáváme vzorce

$$\begin{aligned}\pi &= \pi_1 + \pi_2 = \mathbf{E} N^1 + \mathbf{E} N^2, \\ \theta &= \sqrt{\text{var}(N^1 + N^2)} = \sqrt{\text{var} N^1 + \text{var} N^2}.\end{aligned}$$

Pro odhady střední hodnoty a rozptylu použijeme vyjádření základních charakteristik příslušného rozdělení. Pro dvojité Poissonovo rozdělení dostáváme z tabulky 1.3 a použitím příslušných odhadů parametrů vyjádření

$$\begin{aligned}\hat{\pi}^{DIPD} &= \widehat{\lambda}_1 + \widehat{\lambda}_2, \\ \hat{\theta}^{DIPD} &= \sqrt{\widehat{\lambda}_1 + \widehat{\lambda}_2}.\end{aligned}$$

Pro rozdělení DIDPD, tedy na diagonále modifikované dvojité Poissonovo rozdělení má očekávaný počet škod a směrodatná odchylka tvar

$$\begin{aligned}\hat{\pi}^{DIDPD} &= (1 - \hat{p}) \widehat{\lambda}_1 + (1 - \hat{p}) \widehat{\lambda}_2, \\ \hat{\theta}^{DIDPD} &= \sqrt{(1 - \hat{p}) \widehat{\lambda}_1 + \hat{p}(1 - \hat{p}) \widehat{\lambda}_1 + (1 - \hat{p}) \widehat{\lambda}_2 + \hat{p}(1 - \hat{p}) \widehat{\lambda}_2}.\end{aligned}$$

Budeme-li uvažovat dvourozměrné rozdělení, které bere v potaz závislost mezi jednotlivými složkami, tak máme vyjádření

$$\begin{aligned}\pi &= \pi_1 + \pi_2 = \mathbf{E} N^1 + \mathbf{E} N^2, \\ \theta &= \sqrt{\text{var}(N^1 + N^2)} = \sqrt{\text{var} N^1 + \text{var} N^2 + 2 \text{cov}(N_1, N_2)}.\end{aligned}$$

Pro rozdělení dvourozměrné Poissonovo rozdělení BPD dostáváme z tabulky 1.3 vyjádření π a θ ve tvaru

$$\begin{aligned}\hat{\pi}^{BPD} &= (\widehat{\lambda}_1 + \widehat{\lambda}_0) + (\widehat{\lambda}_2 + \widehat{\lambda}_0), \\ \hat{\theta}^{BPD} &= \sqrt{(\widehat{\lambda}_1 + \widehat{\lambda}_0) + (\widehat{\lambda}_2 + \widehat{\lambda}_0) + 2 \widehat{\lambda}_0}.\end{aligned}$$

Pro v nule modifikované rozdělení DIBPD-D1 dostáváme z tabulky 1.4 vyjádření

$$\begin{aligned}\hat{\pi}^{DIBPD-D1} &= (1 - \hat{p}) (\widehat{\lambda}_1 + \widehat{\lambda}_0) + (1 - \hat{p}) (\widehat{\lambda}_2 + \widehat{\lambda}_0), \\ \hat{\theta}^{DIBPD-D1} &= \sqrt{(1 - \hat{p}) (\widehat{\lambda}_1 + \widehat{\lambda}_0) + \hat{p}(1 - \hat{p}) (\widehat{\lambda}_1 + \widehat{\lambda}_0) + (1 - \hat{p}) (\widehat{\lambda}_2 + \widehat{\lambda}_0) + \\ &\quad + \hat{p}(1 - \hat{p}) (\widehat{\lambda}_2 + \widehat{\lambda}_0) + 2(1 - \hat{p}) [\widehat{\lambda}_0 + (\widehat{\lambda}_1 + \widehat{\lambda}_0) (\widehat{\lambda}_2 + \widehat{\lambda}_0)] - \\ &\quad - 2(1 - \hat{p})^2 (\widehat{\lambda}_1 + \widehat{\lambda}_0) (\widehat{\lambda}_2 + \widehat{\lambda}_0)}.\end{aligned}$$

Hodnoty středních hodnot celkových počtů škod a směrodatných odchylek jsou zobrazeny v tabulce 4.7. Jak je vidět v tabulce 4.7, tak směrodatná odchylka pro modely, které uvažují nezávislost jednotlivých složek počtů škod, je nižší než u modelů, které závislost zachycují. Což odpovídá přesně tomu, co jsme očekávali, protože v případě, kdy nebudeme při tarifování uvažovat kladnou závislost mezi počty škod z různých garancí, tak je riziko, že se při jedné nehodě bude plnit z obou garancí, nižší.

V další části budeme modelovat závislosti mezi garancemi pomocí kopulí a porovnáme, která kopule je k tomuto účelu na našich datech nejvhodnější.

	DPD	DIDPD-D2	BPD	DIBPD-D1
π	5,692	5,632	5,692	5,673
θ	2,386	2,441	2,675	2,816

Tabulka 4.7: Porovnání střední hodnoty počtů škod a směrodatné odchylky pro různé modely.

4.3 Modelování pomocí kopulí

V této části použijeme na vygenerovaná data počtů škod modely kopulí. Využijeme k tomu balíček *Copula* implementovaný v softwaru R. Nejprve z pozorovaných počtů škod vytvoříme takzvaná pseudo-pozorování popsaná v Schepsmeier (2018) jako

$$u_i^j = \frac{r_i^j}{n+1}, \quad i = 1, \dots, n; j = 1, \dots, d,$$

kde r_i^j je pořadí hodnoty N_i^j mezi všemi hodnotami náhodného vektoru $\mathbf{N}^j = (N_1^j, \dots, N_n^j)$, n je celkový počet pozorování, který máme k dispozici – v našem případě $n = 1000$, d je počet vektorů pozorování, tedy $d = 2$, protože máme škody ze dvou typů garancí.

Charakteristika	pseudo- \mathbf{N}^1	pseudo- \mathbf{N}^2
Minimum	0,062	0,057
Maximum	0,999	0,998
Medián	0,575	0,452
Výběrový průměr	0,500	0,500
Směrodatná odchylka	0,286	0,284

Tabulka 4.8: Základní popisné charakteristiky pseudo-pozorování vytvořených z počtů škod.

Hodnoty pseudo-pozorování, které jsme nyní dostali v podstatě odpovídají použití empirické distribuční funkce na hodnoty počtů škod, které máme k dispozici a následnému přeškálování faktorem $\frac{n}{n+1}$. Základní charakteristiky pseudo-pozorování jsou shrnuty v tabulce 4.8.

Pomocí metody maximální věrohodnosti pro pseudo-pozorování odhadneme parametry kopulí na našich vygenerovaných datech. Budeme uvažovat následující kopule: gaussovskou kopuli, Claytonovu kopuli a Gumbelovu kopuli. Kopule popisuje pouze jeden parametr, odhady parametrů těchto kopulí jsou uvedeny v tabulce 4.9.

Kopule	Parametr
gaussovská	0,348
Claytonova	0,605
Gumbelova	1,259

Tabulka 4.9: Odhady parametrů různých kopulí pro simulovaná data počtů škod.

Následně jsou spočítány hodnoty Akaikeho kritéria pro všechny odhadnuté

kopule. Pro kopuli c s vektorem parametrů θ hodnota Akaikeho kritéria odpovídá

$$AIC = -2 \sum_{i=1}^N \log \left(c \left(u_i^1, u_i^2 | \theta \right) \right) + 2k,$$

kde k je počet parametrů kopule.

Kopule	AIC
gaussovská	3 296,49
Claytonova	3 154,58
Gumbelova	3 404,04

Tabulka 4.10: Hodnot AIC pro různé kopule popisující závislostní strukturu počtů škod z povinného ručení a z ostatních garancí.

Pro porovnání hodnot směrodatné odchylky pro různé kopule si uvedeme tvrzení z Pfeifer (2004).

Tvrzení 2. *Nechť $(X, Y)^\top$ je dvourozměrný náhodný vektor s kopulí C a marginálními distribučními funkcemi F a G takovými, že $E|X| < \infty$, $E|Y| < \infty$, $E|XY| < \infty$. Potom kovariance mezi X a Y může být vyjádřena pomocí kopule jako*

$$\text{cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [C(F(x), G(y)) - \Pi(F(x), G(y))] dx dy. \quad (4.2)$$

Vzhledem k tomu, že známe marginální rozdělení, ze kterých jsme data generovali, tak použijeme jako distribuční funkce F a G distribuční funkce v nule modifikovaného dvourozměrného Poissonova rozdělení.

Pro diskrétní Poissonovo rozdělení přejdou dvojně integrály ve dvojně sumy. Směrodatnou odchylku součtu škod θ následně získáme pomocí kopulí jako

$$\theta^{\text{kopule}} = \sqrt{\text{var } N^1 + \text{var } N^2 + 2 \text{cov}(N^1, N^2)},$$

kde se hodnoty rozptylu a kovariance liší pro různé kopule vždy podle použití dané kopule ve vzorci (4.2).

	gaussovská	Claytonova	Gumbelova
θ	1,892	1,931	1,727

Tabulka 4.11: Hodnoty směrodatné odchylky součtu počtů škod při použití různých kopulí.

V tabulce 4.11 jsou uvedeny hodnoty směrodatných odchylek pro gaussovskou, Claytonovu a Gumbelovu kopuli. Nejvyšší hodnota směrodatné odchylky je při modelování Claytonovou kopulí. Claytonova kopule tedy zachycuje největší rizikovost.

Závěr

V této práci jsme se věnovali popisu vícerozměrných modelů počtů škod. Nejprve jsme zobecnili jednorozměrné Poissonovo rozdělení na vícerozměrná Poissonova rozdělení tak, aby byla zachycena vzájemná závislost mezi jednotlivými složkami náhodného vektoru. Uvedli jsme definice pro dvou- a trojrozměrná Poissonova rozdělení, základní charakteristiky daných rozdělení společně s odhady parametrů a různé modifikace těchto rozdělení. Dále jsme zobecnění do více rozměrů provedli pro negativně binomické rozdělení, kde jsme se věnovali převážně trojrozměrnému případu. Nejobecnější model zachycující závislosti mezi jednotlivými složkami je model kopulí, kdy závislostní struktura je popsána kopulí d-
rozměrného rozdělení. Uvedli jsme základní kopule společně s jejich vlastnostmi, mírami závislosti a odhady parametrů kopulí.

V poslední kapitole jsme na simulovaná data vyzkoušeli použít modely uvedené v teoretické části práce. Ukázali jsme na hodnotách směrodatné odchylky součtu počtů škod ze dvou různých garancí, že při existenci závislosti mezi složkami je důležité používat modely, které tuto závislost zohledňují. Směrodatná odchylka byla u modelů, které nezachycují vzájemnou závislost, výrazně menší než u vícerozměrných modelů, které vzájemnou závislost zachycovaly. Pokud se rizikovost klienta popisuje pomocí nějakého násobku směrodatné odchylky nebo rozptylu, tak je potom potřeba používat vícerozměrné modely zachycující závislosti mezi jednotlivými složkami.

Seznam použité literatury

- ALMUHAYFITH, F. E., ALZAID, A. A. a OMAIR, M. A. (2016). On bivariate poisson regression models. *Journal of King Saud University - Science*, **28**(2), 178 – 189. ISSN 1018-3647.
- BERMUDEZ, L. a KARLIS, D. (2011). Bayesian multivariate poisson models for insurance ratemaking. *Insurance Mathematics and Economics*, **48**, 226 – 236. ISSN 0167-6687.
- BERMUDEZ, L. (2009). A priori ratemaking using bivariate poisson regression models. *Insurance Mathematics and Economics*, **44**, 135 – 141. ISSN 0167-6687.
- KLUGMAN, S. A. (1998). *Loss models : from data to decisions*. Wiley series in probability and mathematical statistics. Applied probability and statistics section. Wiley. ISBN 0-471-23884-8.
- MCNEIL, A. J. (2005). *Quantitative risk management : concepts, techniques and tools*. Princeton series in finance. Princeton University Press. ISBN 0-691-12255-5.
- NELSEN, R. B. (2006). *An introduction to copulas*. Springer series in statistics. Springer. ISBN 0-387-28659-4.
- OMELKA, M. (2019). Modern statistical methods. URL https://www.karlin.mff.cuni.cz/~omelka/Soubory/nmst434/nmst434_course-notes.pdf. datum pristupu 7. 7. 2019.
- PFEIFER, D. NESLEHOVA, J. (2004). Modeling and generating dependent risk processes for IRM and DFA. *ASTIN Bulletin*, **34**(2), 333 – 360.
- SCHEPSMEIER, U. STOEBE, J. (2018). Statistical inference of vine copulas. URL <https://cran.r-project.org/web/packages/VineCopula/VineCopula.pdf>. datum pristupu 7. 7. 2019.
- SHI, P. a VALDEZ, E. A. (2014). Multivariate negative binomial models for insurance claim counts. *Insurance Mathematics and Economics*, **55**, 18 – 29. ISSN 0167-6687.

Seznam obrázků

3.1	Znázornění hustoty Claytonovy kopule pro různé parametry θ . . .	26
3.2	Znázornění hustoty Gumbelovy kopule pro různé parametry θ . . .	27
4.1	Počty škod pro dané dvojice hodnot (N^1, N^2) s rozlišením zastoupení jednotlivých hodnot dle velikosti puntíků.	32
4.2	Závislost N^1 na N^2 společně s regresní přímkou zachycující lineární závislost mezi náhodnými výběry.	34

Seznam tabulek

1.1	Základní charakteristiky Poissonova rozdělení.	6
1.2	Základní charakteristiky v nule modifikovaného Poissonova rozdělení.	7
1.3	Základní charakteristiky dvourozměrného Poissonova rozdělení.	9
1.4	Základní charakteristiky v nule modifikovaného dvourozměrného Poissonova rozdělení.	11
1.5	Základní charakteristiky trojrozměrného Poissonova rozdělení se společnou kovariancí.	13
1.6	Základní charakteristiky trojrozměrného Poissonova rozdělení s úplnou kovariancí.	14
2.1	Základní charakteristiky negativně binomického rozdělení.	17
2.2	Základní charakteristiky trojrozměrného negativně binomického rozdělení se společnou kovariancí, kde $\sigma^2 = \psi$ a $\lambda_j = \eta_j \psi$	20
2.3	Základní charakteristiky trojrozměrného Poissonova rozdělení s úplnou kovariancí.	22
4.1	Prvních 8 řádků regresní matice pozorování \mathbb{X}	32
4.2	První čtyři hodnoty parametrů λ^j , $j = 0, 1, 2$ a jejich výběrové průměry.	33
4.3	Základní charakteristiky náhodných výběrů \mathbf{N}^1 a \mathbf{N}^2	33
4.4	Různé korelační koeficienty pro vzájemnou korelaci mezi N^1 a N^2	33
4.5	Porovnání odhadů parametrů pro různé modely.	35
4.6	Informační kritéria AIC a BIC pro uvažované modely – nejnižší hodnoty jsou vyznačeny tučně.	35
4.7	Porovnání střední hodnoty počtů škod a směrodatné odchylky pro různé modely.	37
4.8	Základní popisné charakteristiky pseudo-pozorování vytvořených z počtů škod.	37
4.9	Odhady parametrů různých kopulí pro simulovaná data počtů škod.	37
4.10	Hodnot AIC pro různé kopule popisující závislostní strukturu počtů škod z povinného ručení a z ostatních garancí.	38
4.11	Hodnoty směrodatné odchylky součtu počtů škod při použití různých kopulí.	38

Seznam použitých zkratek

MLE ... maximálně věrohodný odhad z anglického Maximal Likelihood Estimate

PD ... Poissonovo rozdělení z anglického Poisson distribution

BPD ... dvourozměrné Poissonovo rozdělení z anglického Bivariate Poisson Distribution

DPD ... dvojité Poissonovo rozdělení z anglického Double Poisson distribution

DIBPD ... na diagonále modifikované dvourozměrné Poissonovo rozdělení z anglického diagonal inflated bivariate Poisson distribution

ZIBPD ... v nule modifikované dvourozměrné Poissonovo rozdělení z anglického zero inflated bivariate Poisson distribution

DIDPD ... na diagonále modifikované dvojité Poissonovo rozdělení z anglického diagonal inflated double Poisson distribution

TPDCC ... trojrozměrné Poissonovo rozdělení se společnou kovariancí z anglického Trivariate Poisson Distribution with Common Covariance

TPDFC ... trojrozměrné Poissonovo rozdělení s úplnou kovariancí z anglického Trivariate Poisson Distribution with Full Covariance

NBD nebo NB ... negativně binomické rozdělení z anglického Negative Binomial Distribution

TBNCC ... negativně binomické rozdělení se společnou kovariancí z anglického Negative Binomial Distribution with Common Covariance

TBNFC ... negativně binomické rozdělení s úplnou kovariancí z anglického Negative Binomial Distribution with Full Covariance

$Ran(\cdot)$... obor hodnot funkce z argumentu

AIC ... Akaikeho informační kritérium

BIC ... Bayesovo informační kritérium