

UNIVERZITA PARDUBICE

FAKULTA EKONOMICKO-SPRÁVNÍ

Simulační metody modelování kolektivního rizika

Pavel Berka

Diplomová práce

2012

UNIVERZITA PARDUBICE

FACULTY OF ECONOMY AND ADMINISTRATION

Simulation modeling methods of collective risk

Pavel Berka

Thesis

2012

ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: Bc. Pavel Berka
Osobní číslo: E09845
Studijní program: N6209 Systémové inženýrství a informatika
Studijní obor: Pojistné inženýrství
Název tématu: Simulační metody modelování kolektivního rizika
Zadávací katedra: Ústav matematiky

Z á s a d y p r o v y p r a c o v á n í :

Modelování počtu a výše individuálních škod
Kolektivní model rizika
Monte Carlo simulace hodnot kolektivního rizika
Testy dobré shody při využití simulovaných dat
Využití výpočetní techniky a vhodného programovacího jazyku


Rozsah grafických prací: –
Rozsah pracovní zprávy: cca 50 stran
Forma zpracování diplomové práce: tištěná/elektronická
Seznam odborné literatury:

- [1] BEARD, R. E.: Risk Theory, the Stochastic Basis of Insurance, Chapman and Hall, London 1994.
- [2] BOLAND, P. J.: Statistical and probabilistic methods in actuarial science, Chapman 2007, Boca Raton, FL.
- [3] CIPRA, T.: Teorie rizika v pojistné matematice, MFF UK a Česká pojišťovna, Praha 1991.
- [4] DAYKIN, C. D., PENTIKÄINEN, T., PESONEN, M.: Practical Risk Theory for Actuaries, Chapman and Hall, London 1996.
- [5] FABIAN, F., KLUIBER, Z.: Metoda Monte Carlo a možnosti jejího uplatnění. Praha: Nakladatelství PROSPEKTRUM, 1998.
- [6] HORÁKOVÁ, G., MUCHA, V.: Teória rizika v poistení. Bratislava: Vydavateľstvo EKONÓM, 2006.
- [7] PACÁKOVÁ, V.: Aplikovaná poistná štatistika - 3. prepracované a doplnené vydanie. Bratislava: IURA EDITION, 2004.
- [8] RUBINSTEIN, REUVEN Y.: Simulation and the Monte Carlo Method. New York: John Wiley & Sons, Inc., 1981.

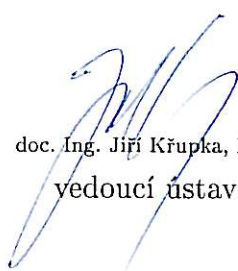
Vedoucí diplomové práce: prof. RNDr. Viera Pacáková, CSc.
UK Praha

Datum zadání diplomové práce: 30. června 2011

Termín odevzdání diplomové práce: 30. dubna 2012


doc. Ing. Renáta Myšková, Ph.D.
děkanka

L.S.


doc. Ing. Jiří Krupka, Ph.D.
vedoucí ústavu

V Pardubicích dne 4. srpna 2010

Prohlašuji:

Tuto práci jsem vypracoval samostatně. Veškeré literární prameny a informace, které jsem v práci využil, jsou uvedeny v seznamu použité literatury.

Byl jsem seznámen s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorský zákon, zejména se skutečností, že Univerzita Pardubice má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona, a s tím, že pokud dojde k užití této práce mnou nebo bude poskytnuta licence o užití jinému subjektu, je Univerzita Pardubice oprávněna ode mne požadovat přiměřený příspěvek na úhradu nákladů, které na vytvoření díla vynaložila, a to podle okolností až do jejich skutečné výše.

Souhlasím s prezenčním zpřístupněním své práce v Univerzitní knihovně Univerzity Pardubice.

V Pardubicích dne 26. 04. 2012

Pavel Berka

Děkuji prof. RNDr. Věře Pacákové, Ph.D. za profesionální vedení, cenné rady a připomínky, které napomohly zkvalitnění a dokončení diplomové práce.

Souhrn

Diplomová práce se zabývá metodami modelování kolektivního rizika pomocí simulační metody Monte Carlo a její součástí je i praktická ukázka řešení ve statistickém programovacím jazyku R.

Řešení je doplněno výpočty a jejich ověření ve statistickém programu Statgraphics Centurion XIV, spolu s výpočty v tabulkovém procesoru Excel.

Klíčová slova

kolektivní riziko, Monte Carlo, simulační metody, programovací statistický jazyk R

Abstract

This thesis deals with the collective risk modelling using Monte Carlo simulation methods and includes a practical demonstration of the solution in the statistical programming language R. The solution is complemented by calculations and their verification in the statistical program Statgraphics Centurion XIV, together with calculations in the spreadsheet program Excel.

Keywords

collective risk, Monte Carlo, simulation methods, statistical programming language R

Obsah

Úvod	11
1 Pojem rizika	13
1.1 Pojištění	15
1.2 Risk management	16
2 Modelování pojistných rizik	18
2.1 Individuální model rizika	21
2.2 Kolektivní model rizika	21
2.2.1 Obecná charakteristika kolektivního rizika	22
2.2.2 Základní charakteristiky kolektivního rizika S	22
2.2.3 Modely počtu pojistných nároků	23
2.2.4 Modely výše škod	25
2.3 Ověření vhodnosti zvoleného rozdělení	27
2.3.1 Pearsonův χ^2 - test dobré shody	27
2.3.2 Kolmogorov-Smirnovův test	28
2.4 Aproximace kolektivního modelu rizika	30
2.4.1 Aproximace normálním rozdělením	30
2.4.2 Aproximace posunutým gama rozdělením	31
3 Využití simulace pro aproximaci rozdělení kolektivního rizika	32
3.1 Metoda Monte Carlo	32
3.1.1 Přesnost odhadu	33
3.1.2 Postup realizace metody Monte Carlo	33
4 Programovací jazyk R	35
4.1 Prostředí R	35
4.1.1 Analýza dat	35
4.1.2 R jako programovací jazyk	36

4.1.3	Více než statistický software	36
5	Aplikace metod aproximace kolektivního rizika S	38
5.1	Rozdělení počtu škod	38
5.1.1	Výpočet v R.....	38
5.1.2	Výpočet v programech Excel a Statgraphics	41
5.2	Rozdělení výše škod.....	42
5.2.1	Výpočet v R.....	42
5.2.2	Výpočet v programech Excel a Statgraphics	47
5.3	Simulace Monte Carlo v R	49
5.3.1	Simulace v R.....	49
5.4	Pravděpodobnostní rozdělení S pomocí systému Statgraphics	53
5.4.1	Posunuté lognormální rozdělení	53
5.4.2	Posunuté gamma rozdělení.....	55
5.4.3	Normální rozdělení	57
6	Závěr	58
7	Seznam použité literatury.....	59

Seznam tabulek

Tab. 1: Subjektivní a objektivní rizika, zdroj [10.].....	14
Tab. 2: Standardní rozdělení pravděpodobnosti v základním prostředí R, zdroj [21.]	37
Tab. 3: Test shody pro rozdělení počtu škod Excel, zdroj vlastní	41
Tab. 4: Test shody pro rozdělení počtu škod Statgraphics, zdroj vlastní	42
Tab. 5: Výpočet parametrů metodou maximální věrohodnosti, zdroj vlastní	47
Tab. 6: Výsledné hodnoty metody maximální věrohodnosti, zdroj vlastní.....	48
Tab. 7: Test shody pro rozdělení výšky škod, zdroj vlastní	49
Tab. 8: Výsledné parametry posunutého lognormálního rozdělení, zdroj vlastní	53
Tab. 9: Výsledek testu dobré shody posunutého lognormálního rozdělení, zdroj vlastní.....	54
Tab. 10: Výsledek Kolmogorov-Smirnovova testu posunutým lognormálním rozdělením, zdroj vlastní	54
Tab. 11: Výsledné parametry posunutého gamma rozdělení, zdroj vlastní	55
Tab. 12: Výsledek testu dobré shody posunutého gamma rozdělení, zdroj vlastní.....	56
Tab. 13: Výsledek Kolmogorov-Smirnovova testu posunutého gama rozdělení, zdroj vlastní .	56

Seznam obrázků

Obr. 1: Kolektivní rezerva, zdroj [10.].....	16
Obr. 2: Algoritmus generování náhodných čísel metodou Monte Carlo, zdroj vlastní	34
Obr. 3: Výsledné grafické zobrazení z programu R, zdroj vlastní.....	40
Obr. 4: Histogram výsledných hodnot získaných metodou Monte Carlo, zdroj vlastní	52
Obr. 5: Grafické porovnání shody kolektivního rizika s posunutým Lognormálním rozdělením, zdroj vlastní	55
Obr. 6: Grafické porovnání shody kolektivního rizika s posunutým gamma rozdělením, zdroj vlastní	57

Seznam příloh

Příloha A: Ověření Poissonova rozdělení.....	61
Příloha B: Ověření Paretova rozdělení.....	64
Příloha C: Simulace metodou Monte Carlo.....	67
Příloha D: Výsledky metody Monte Carlo v programu Statgraphics	69

Úvod

Současné trendy kolem rámcové směrnice Solvency II (SII) kladou mimořádný důraz na řízení rizika a konkrétní mechanismy (interní i externí) s použitím sofistikovaných nástrojů.

Při přípravě na budoucí systém je v pojišťovnách nutné důsledně připravit způsob analýzy a modelování jednotlivých rizik a jejich vzájemných vztahů pomocí aktuárských metod, založených na pokročilých metodách matematiky, pravděpodobnosti, statistiky a operačního výzkumu s využitím vhodných datových souborů pojišťovny. Vývoj požadovaných souborů je mimořádně náročný a bude klást vysoké nároky na lidské i technické zdroje, hlavně na oblast informačních technologií. [19.]

Český pojišťovací trh se nejen proto věnuje stále více analýze vlastních dat z pohledu organizace, zpracování a zejména jejich praktickému využití. Důraz je zde kladen především na podporu obchodu a lepší odhady míry rizika, které na sebe pojišťovny v rámci zvyšujícího se zájmu o komerční pojištění přebírají od svých klientů ve stále větší míře.

Díky rychle se měnícím přírodním, ekonomickým, ale i společenským trendům, které vládou moderní, někdy též nazývané turbulentní, době je analýza rizikového kapitálu nutností, kterou musí moderní pojišťovna bezpodmínečně, pakliže chce být konkurence schopná, zajišťovat.

Rizikový kapitál je v pojišťovnictví kvantifikován pomocí teorie rizika založené především na pokročilých metodách teorie pravděpodobnosti a matematické statistiky. Moderní přístup teorie rizika je představován zejména modely kolektivního rizika, jejichž základ je tvořen rozdělením počtu a výšky individuálních pojistných plnění a slouží jako ukázka široké aplikace těchto metod. [16.]

Předkládaná diplomová práce se snaží nastínit možnosti moderní statistiky s využitím IT. Statistické programy, jako v této práci použitý Statgraphics Centurion XIV. a tabulkový procesor Excel, slouží jako odrazový můstek při řešení netriviálních praktických problémů pojistné praxe.

V některých, značně pokročilejších, situacích, kdy je potřeba vytvořit komplexní aplikaci umožňující např. personalizovaný výběr statistických funkcí, specifické GUI¹, vstupy či výstupy, které se budou dále zpracovávat v jiném programu nebo předávat v některém specifickém formátu apod., je vhodné sáhnout po některém z množství programovacích jazyků.

¹ GUI – grafické uživatelské prostředí

Pro praktickou ukázkou řešení úloh, se kterými se v pojišťovně setkáme a ke kterým je možný zařadit také modelování kolektivního rizika pomocí metody Monte Carlo je vybrán programovací jazyk R, který v sobě kloubí dostupnost a relativní jednoduchost s programátorskou volností.

R, jakožto statistický software, obsahuje řadu důležitých statistických funkcí již v základní instalaci. Další je možné jednoduše dodat pomocí některé z množství podpůrných knihoven.

Nespornou výhodou je možnost propojení s dalšími programovacími jazyky jako C++, Java apod., s jejichž pomocí je pro profesionála možné uspokojit ty nejspecifičtější potřeby.

Cílem diplomové práce je položit teoretický základ pro aplikaci metod kolektivního rizika, výpočet úloh pro zjištění rozdělení pravděpodobnosti modelu počtu a výšky pojistných škod a následná aplikace získaných výsledků pro simulaci kolektivního rizika pomocí metody Monte Carlo. Výsledek simulace bude taktéž podroben zkoumání na zjištění rozdělení náhodné proměnné, kterou je celkové pojistné plnění pojišťovny za rok, spolu se zjištěním základních charakteristik rozdělení.

Diplomová práce ukáže možnosti využití programovacího jazyku R při řešení statistických úloh pojišťoven, včetně již zmíněné aplikace algoritmů pro simulaci metodou Monte Carlo. Řešení v programovacím jazyku je doplněno a konfrontováno použitím statistického software Statgraphics Centurion a tabulkového procesoru Excel.

1 Pojem rizika

Základním pojmem v pojišťovnictví je riziko, které představuje možnost vzniku pojistné události. Pojistitel, který na sebe profesionálně přebírá velké množství rizik různého charakteru, musí k hodnocení rizika zaujmout velmi zodpovědný postoj. Riziko pojistitele spočívá v nebezpečí, zda přijaté pojistné bude dostačující na vyplacení všech oprávněných pojistných plnění. [16.]

Teorie rizika je první oblastí pojistných věd, ve které se vzhledem k praktickému rozvoji zejména neživotního pojištění začali využívat pokročilé metody teorie pravděpodobnosti a matematické statistiky. [16.]

Pojem rizika nemá v literatuře jednoznačnou definici. Obecně proto můžeme riziko definovat jako:

- pravděpodobnost či možnost vzniku ztráty, obecně nezdaru,
- odchýlení skutečných a očekávaných výsledků,
- nebezpečí chybného rozhodnutí,
- variabilita možných výsledků nebo nejistota jejich dosažení. [20.]

V ekonomii je pojem rizika užíván v souvislosti s nejednoznačností průběhu určitých skutečných ekonomických procesů spolu s nejednoznačností jejich výsledků. [23.]

Ducháčková [10.], str. 15, ve svých Principech pojištění a pojišťovnictví definuje pojem rizika jako „Možnost vzniku události s výsledkem odchylným od cíle s určitou objektivní pravděpodobností (statistickou či matematickou).“ Riziko se tedy, na rozdíl od pravé nejistoty, dá měřit. V případě pojistného rizika je tudíž známo rozdělení pravděpodobnosti. [10.]

V rámci pojišťovnictví je riziko vnímáno jako potenciální možnost vzniku pojistné události s určitou známou objektivní pravděpodobností. Zároveň se v pojišťovnách jako riziko často označuje každá uzavřená pojistná smlouva, popř. každá z jejích složek, pokud se jedná o sdružená pojištění.

Riziko, jak bylo výše definováno, se může vyskytovat v různých souvislostech. V závislosti na povaze příslušného jevu či procesu mohou realizací příslušného rizika vzniknout: [10.]

- Výhradně negativní (záporné) odchylky od cíle, kdy se mluví o tzv. čistém riziku (nebezpečí ztrát).

U čistého rizika platí dále skutečnost, že toto není záměrně lidmi podstupováno.

- Záporné i kladné odchylky od cíle, kdy jde o tzv. spekulativní (záměrné) riziko. Jsou to situace spojené s hraním hazardních her, sázením, spekulacemi na burze, investováním, podnikáním apod. Subjektem je toto riziko dobrovolně podstupováno.

Z výše uvedeného je jasné, že předmětem pojištění se mohou stát jen tzv. čistá rizika. Rizika spekulativní (tj. uměle vytvořená) pojistit nelze. [25.]

Uvedené rozdělení však není v praktickém použití stejně jasné a jednoznačné jako v teorii. V praxi je u čistého rizika nutno sledovat i jeho objektivní a subjektivní stránku. Objektivní riziko je dáno objektivními skutečnostmi, které nezávisí na lidech. Subjektivní riziko potom existuje v závislosti na činnosti člověka, bez ohledu na to, zda na vědomé či nevědomé. Takové rizikové momenty závisí na duševních a charakterových vlastnostech daného subjektu. Mezi objektivním a subjektivním rizikem je často složité najít přesnou hranici.

U rizik ohrožujících osoby se pod objektivním rizikem rozumí souhrn rizikových momentů u osob s průměrnými duševními a charakterovými vlastnostmi. Subjektivní riziko představují potom odchylky od tohoto průměru. Stručný přehled je nastíněn v Tab. 1.

Tab. 1: Subjektivní a objektivní rizika, zdroj [10.]

Subjektivní rizika	Objektivní rizika
na základě konání a jednání lidí neopatrnost, schopnosti a charakterové vlastnosti, morální riziko, <i>například: žhářství, dovednost při manuální práci, riskantní jízda řidiče</i>	na základě objektivně daných skutečností <i>například: blesk, přírodní katastrofa</i>

Ekonomické subjekty se snaží realizaci rizika a jeho možných důsledků předcházet pomocí preventivních opatření. Všeobecně však riziko není možné vyloučit.[17.]

Pro rizika, jejichž realizaci nelze odvrátit nebo omezit, nebo jen v určité míře prostřednictvím specifických nástrojů (různých smluvních dohod, konkrétních technických nástrojů jako ochranná zařízení apod.), přichází v úvahu finanční krytí. Tedy zabezpečení finanční náhrady škod vzniklých realizací rizik. Finanční krytí se uskutečňuje v různých formách, závislých na nositeli a charakteru rizika. [10.]

Jde o: [10.]

- krytí prostřednictvím státu,
- individuální krytí rizika,
- pojištění.

Pojišťovnu samozřejmě zajímá třetí bod.

1.1 Pojištění

Podstatou pojištění je ochrana proti negativním důsledkům nahodilosti. [10.] Pojišťovna pak z titulu placení pojistného převádí veškerá přijatá rizika do finanční roviny. Případné škody jsou v běžné praxi taktéž v penězích vypláceny.

Pojištění tedy znamená ochranu proti pojistným rizikům pomocí přenesení onoho rizika na specializovanou instituci - pojistitele. Jedná se o tvorbu rezerv na krytí rizik prostřednictvím příspěvků na pojištění od jednotlivých pojistníků. Jde o tvorbu kolektivní rezervy a rozdělení rizika mezi více zúčastněných, kdy krytí rizik není shora omezeno naspořenými prostředky jednotlivého pojistníka. [10.]

Rizika takového druhu, která jsou pro pojistníka jako jednotlivce neúnosná, převezme pojišťovna a vytvoří z nich soubor převzatých rizik – tzv. pojistný kmen. Pokud je tento pojistný kmen dostatečně velký a homogenní, je pojišťovna schopna pomocí správně vypočteného pojistného riziku nejen čelit, ale z přijatého pojistného také vytvářet zisk.[25.][10.]

Pojišťovna při tomto postupu vychází z principu solidarity mezi pojištěnými, který umožňuje rozvrhnout případnou vzniklou škodu mezi všechny její klienty vystavenými stejnému riziku, i když škodu utrpěla jen část z nich, viz Obr. 1. [25.]

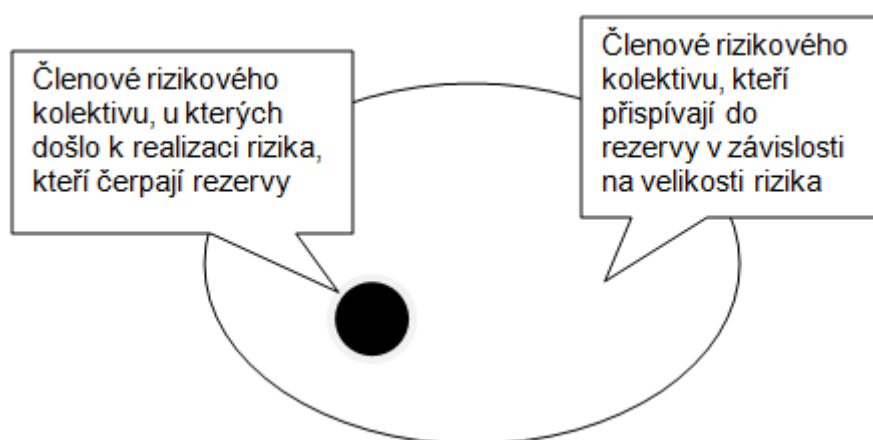
Pro pojišťovnu se tak rizika převzatá od klientů mění na tzv. pojistně technické riziko, tj. riziko, že přijaté pojistné nebude dostatečné pro krytí pojistného plnění. Toto riziko lze vyjádřit jako výši variability očekávaného stavu, který se odrazí ve vypláceném pojistném plnění. Přitom platí pravidlo, že čím větší je pojistný kmen, tím více klesá pojistně technické riziko. [25.]

Realizaci pojistného rizika pak nazýváme pojistnou událostí.

Pojistně technické riziko pojistitele ovlivňují hlavně dvě charakteristiky:

- počet výskytů pojistných událostí,
- velikost škody pojistných událostí.

Počet pojistných událostí a výška pojistných plnění jsou v terminologii teorie pravděpodobnosti náhodné proměnné. Znalost rozdělení pravděpodobnosti těchto náhodných proměnných je základem řešení všech dalších zásadních problémů pojišťovny souvisejících s řízením pojistně technického rizika tak, aby se snížil jeho negativní dopad na hospodaření pojišťovny. [19.]



Obr. 1: Kolektivní rezerva, zdroj [10.]

To vše klade velký důraz na pojišťovnu, aby přesně vypočetla pravděpodobnosti vzniku pojistných událostí, jejich výše a tím i výše placeného pojistného. Konkrétním výpočtům a jejich simulacím se budeme věnovat v samostatných kapitolách.

1.2 Risk management

Snaha lidí o zvládnutí rizika pomocí určitých vědeckých přístupů vedla ke vzniku speciálního oboru risk management (řízení rizika).

Řízení rizika v pojišťovnách využívá teorii rizika, jako důležité součásti moderních pojistných věd. Teorie rizika se zabývá analýzou pravděpodobnostních vlastností počtu a výšky pojistných škod a stochastických vlastností pojistných procesů. [19.]

Řízení rizika napomáhá racionálnímu jednání v rizikových situacích tak, aby se chránila stávající a budoucí aktiva podniku. K tomu se využívá znalostí z oblasti managementu, marketingu, financí, oceňování, statistiky, pojišťovnictví, právní úpravy apod. Mimo jiné se

počítá s využitím moderních počítačových systémů, které na bázi simulací odhadují možné rizikové scénáře. [23.]

Proces řízení rizika sestává ze tří fází: [10.]

- identifikace rizika,
- ocenění a kvantifikace rizik,
- kontrola a financování těchto rizik.

Se znalostí této klasifikace bychom výpočty v diplomové práci zařadily právě do činnosti managementu rizik pojišťovny.

- Pod identifikací rizika je možné si představit pojistnou smlouvu, či soubor pojistných smluv uzavřených pojišťovnou.
- Oceněním a kvantifikací rizika se budou zabývat další kapitoly, kdy je nejprve zjištěno rozdělení pravděpodobností počtu a výše škod a poté na základě zjištěných údajů vytvořena simulace podobných škodních průběhů pro zjištění kolektivního rizika, se kterým pojišťovna pracuje např. při výpočtu pojistného, rozhodování o vhodném zajištění apod.
- Do kontroly a financování rizik je možné zařadit nutnost pojišťovny držet dostatečnou míru likvidního kapitálu pro výplaty pojistných plnění a neustálou kontrolu správnosti a změn vypočtených údajů.

V praxi je samozřejmě celý systém rozsáhlejší, komplexnější a více sofistikovaný.

2 Modelování pojistných rizik

Model rizika slouží v pojišťovnictví pro účely modelování celkové výše škod Z , které nastaly během daného časového období délky T (nejčastěji během jednoho roku) v daném pojistném kmeni. [16.]

Pojistná rizika mají náhodný charakter a v činnosti pojišťovny se jako náhodné výkyvy projevují hodnoty těchto náhodných proměnných:

- N – počet pojistných škod v určitém období
- X – výška individuálních pojistných plnění, které vyplývají pro pojišťovnu z jednotlivých pojistných smluv v určitém portfoliu pojistných smluv za určité období
- S – výška celkových pojistných plnění pojišťovny za určité časové období

Náhodná proměnná N je diskrétní s možnými hodnotami $k = 0, 1, 2, \dots$. Náhodné proměnné X a S jsou pak spojité, i když někdy bude výhodné považovat náhodnou proměnnou S také za diskrétní. Pro možné hodnoty x proměnné X platí $x \in (0, +\infty)$, případně při jiných typech pojištění je interval možných hodnot proměnné X zdola ohraničený pojistnou sumou v životních pojištěních a pojistnou částkou v majetkových a odpovědnostních pojištěních. Pro náhodnou proměnnou S platí $S \geq 0$. [17.]

V modelech kolektivního rizika se uplatňují zejména centrální limitní věta, vlastnosti momentové vytvářející funkce, touto funkcí odvozené počáteční a centrální momenty, konvoluce distribučních funkcí a vlastnosti podmíněných středních hodnot. [18.]

o Centrální limitní věta

Pokud mají prvky X_1, X_2, \dots, X_n posloupnosti nezávislých náhodných proměnných stejné rozdělení se střední hodnotou μ a směrodatnou odchylkou σ , pak rozdělení náhodné proměnné

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

se s rostoucím n blíží k normálnímu rozdělení se střední hodnotou μ a směrodatnou odchylkou $\frac{\sigma}{\sqrt{n}}$.

Z této věty plyne, že aritmetický průměr jako náhodná proměnná je za velmi malých omezení asymptoticky normálně rozdělen. Jeho náhodné chování můžeme aproximovat pomocí normálního rozdělení.

Zapsané jako: [12.]

$$\bar{X} \approx N\left(\mu; \frac{\sigma^2}{n}\right) \text{ pro } n \rightarrow \infty$$

o Momentová vytvářející funkce

Momentová vytvářející funkce má v pojistných vědách mimořádný význam a časté použití. Je definovaná vztahem: [18.]

$$M_x(z) = E(e^{zX})$$

Momentová vytvářející funkce má několik výhodných vlastností, z nichž první dvě velmi ulehčují výpočet počátečních a centrálních momentů: [18.]

1. Nechť X je libovolná náhodná proměnná s momentovou vytvářející funkcí $M_x(z)$. Potom pro libovolné celé kladné číslo k platí:

$$\frac{d^k}{dz^k} M_x(z)|_{z=0} = E(X^k) = m_k$$

2. Pro $k = 2$ a $k = 3$ platí vztah:

$$\frac{d^k}{dz^k} \ln M_x(z)|_{z=0} = E\{[X - E(X)]^k\} = \mu_k$$

o Konvoluce distribučních funkcí:

Předpokládá se, že $\{X_i\}_{i=1}^n$ jsou nezávislé a identicky rozdělené náhodné proměnné se společnou distribuční funkcí $F(x)$. Distribuční funkci $F^{*n}(x)$ jejich součtu $\sum_{i=1}^n X_i$ se nazývá n -násobnou konvolucí distribučních funkcí: [18.]

$$F^{*n}(x) = P(X_1 + X_2 + \dots + X_n \leq x)$$

○ **Vlastnosti podmíněných středních hodnot**

Pro modely kolektivního rizika se uplatňují zejména následující dva vztahy, které platí pro libovolné náhodné proměnné X, Y : [13.][18.]

$$E(X) = E[E(X/Y)]$$

$$D(X) = E[D(X/Y)] + D[E(X/Y)]$$

2.1 Individuální model rizika

Pravděpodobnostní modely individuálního rizika, stručně nazývané individuální modely rizika, byly v historickém vývoji teorie rizika prvním stupněm jeho rozvoje. Historicky tedy předcházeli teorii kolektivních modelů rizika. Kolektivní modely rizika jsou všeobecnějšími modely celkových pojistných plnění, resp. Individuální modely rizika můžeme považovat za speciální případ kolektivních modelů rizika. [13.][18.]

Individuální model rizika pracuje s riziky příslušejícími jednotlivým (konkrétním) pojistným smlouvám.

Označme S_n výšku pojistného plnění z portfolia n individuálních rizik. Potom můžeme vyjádřit

$$S_n = Y_1 + Y_2 + \dots + Y_n$$

přičemž $Y_j, j = 1, 2, \dots, n$ je výška pojistného plnění pro j -tou pojistku z portfolia n pojistek.

Výše škod v individuálním modelu rizika Y_1, \dots, Y_n je tedy reprezentována posloupností škodních nároků odpovídajících jednotlivým pojistným smlouvám z pojistného kmene tvořeného n smlouvami. Většinou jsou Y_i navzájem nezávislé náhodné veličiny a pro některé j může platit $Y_j = 0$, jelikož ne pro každou pojistku nastane ve sledovaném období pojistná událost. [18.]

2.2 Kolektivní model rizika

Vychází z předpokladu, že v homogenním pojistném kmenu (příp. v dané tarifní skupině) lze výše škodních nároků z jednotlivých pojistných událostí považovat za stejně rozdělené (a většinou navzájem nezávislé) náhodné veličiny

$$S = \sum_{i=1}^N X_i$$

Celková výše škod v kolektivním modelu rizika X_1, \dots, X_N je posloupnost škodních nároků (uspořádaná na rozdíl od individuálního modelu rizika zcela libovolně bez ohledu na to, jakým pojistným smlouvám tyto škody příslušejí) a N je počet pojistných událostí v uvažovaném období. Většinou jsou X_i navzájem nezávislé a stejně rozdělené náhodné veličiny, které jsou nezávislé s náhodnou veličinou N . [13.][16.]

2.2.1 Obecná charakteristika kolektivního rizika

Označíme písmenem N počet pojistných plnění za celé portfolio pojišťovny v daném období, písmenem X_i velikost i -tého pojistného plnění pro $i = 1, 2, \dots, N$. Pak

$$S = X_1 + X_2 + \dots + X_N$$

představuje celkovou (agregovanou) sumu pojistných plnění, tzv. kolektivní riziko pojišťovny pro uvažované období, nejčastěji kalendářní rok. Přitom platí:

$X_1 + X_2 + \dots + X_N$ jsou nezávislé a identicky rozdělené náhodné veličiny,

náhodné veličiny N, X_1, X_2, \dots, X_N jsou vzájemně nezávislé,

pokud $N = 0$, pak i $S = 0$. [18.]

2.2.2 Základní charakteristiky kolektivního rizika S

Poměrně snadno lze dokázat, že charakteristiky kolektivního rizika můžeme vyjádřit pomocí charakteristik náhodných veličin N a X_i . [16.]

Společnou distribuční funkci identicky rozdělených náhodných proměnných X_i označme $F(x)$.

Dále ať symbol $m_k, k = 1, 2, 3, \dots$ značí k -tý počáteční moment identicky rozdělených individuálních pojistných plnění X_i , tj. $m_k = E(X_i^k)$. Potom platí: [13.] [18.]

$$E(S) = E(N)m_1$$

$$D(S) = E(N)(m_2 - m_1^2) + D(N)m_1^2$$

Speciální případy složeného rozdělení S dostáváme v závislosti od rozdělení počtu pojistných plnění N . Pro praktické použití je nejdůležitější případ, kde N má Poissonovo rozdělení.

- Složené Poissonovo rozdělení

Nechť je S celkové pojistné plnění, přičemž N má Poissonovo rozdělení s parametrem λ , symbolicky $N \sim Po(\lambda)$ a $F(x)$ je společná distribuční funkce identicky rozdělených individuálních pojistných plnění X_i .

Pak S má složené Poissonovo rozdělení s parametry λ a $F(x)$ a označuje se $CoPo(\lambda, F(x))$. [18.]

V tomto případě platí: [18.]

$$E(S) = \lambda m_1$$

$$D(S) = \lambda m_2$$

$$\gamma_1 = \frac{\lambda m_3}{(\lambda m_2)^2} > 0$$

Protože $\lambda > 0$ a $X_i \geq 0$, platí i $\gamma_1 > 0$. Složené Poissonovo rozdělení je tedy vždy pozitivně zešikmené, dokonce i tehdy, když je individuální pojistné plnění X negativně zešikmené. Pro velké hodnoty parametru λ se rozdělení $CoPo(\lambda, F(x))$ stává symetričtější, protože platí, že $\gamma_1 \rightarrow 0$, když $\lambda \rightarrow \infty$.

2.2.3 Modely počtu pojistných nároků

Vycházejí z pravděpodobnostního rozdělení náhodné veličiny N (počet pojistných událostí v uvažovaném období) nabývající většinou hodnot $0, 1, 2, \dots$ [16.]

V praktické části je využito Poissonovo rozdělení $Po(\lambda)$.

○ Poissonovo rozdělení $N \sim Po(\lambda)$

Poissonovo rozdělení $Po(\lambda)$, $\lambda > 0$ vzniká jako limitní případ binomického rozdělení $Bi(n; \pi)$, když současně platí $n \rightarrow \infty$ a $\pi \rightarrow 0$, přitom střední hodnota $n\pi$ se rovná konstantě λ .

Náhodná proměnná N s možnými hodnotami $k = 0, 1, 2, \dots$ má Poissonovo rozdělení s parametrem λ právě tehdy, když jeho pravděpodobnostní funkce má tvar: [18.]

$$P(k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}, \quad k = 0, 1, 2, \dots$$

Je jedno-parametrické rozdělení s $\lambda > 0$. N má význam počtu pojistných událostí ve velkém počtu nezávislých homogenních pojistných smluv s malou pravděpodobností pojistné události (tzv. „rozdělení řídkých jevů“).

$$E(N) = D(N) = \lambda$$

K odhadu parametrů pravděpodobnostního rozdělení se nejčastěji používají metoda momentů a metoda maximální věrohodnosti. [18.]

Princip metody momentů spočívá v tom, že momenty základního souboru odhadujeme odpovídajícími výběrovými momenty. Nevýhodou této metody je, že získané výsledky nemají všeobecně vlastnosti dobrých odhadů. [14.]

$$P(X = x_i) = \frac{\lambda^{x_i}}{x_i!} \cdot e^{-\lambda}$$

Vybereme n prvků x_1, \dots, x_n

$$m_1 = \frac{1}{n} \sum_{i=1}^n x_i$$

pak tedy

$$\text{est } \lambda = \tilde{\lambda} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$$

Metoda maximální věrohodnosti je metoda, kterou je možné aplikovat ve velmi rozmanitých situacích a odhady pomocí této metody získané mají velmi dobré vlastnosti v porovnání s jinými metodami, např. metodou momentů.

Vynikající asymptotické vlastnosti maximálně věrohodných odhadů zaručují kvalitní odhady zejména pomocí výběrových souborů velkého rozsahu. [14.][18.]

$$P(X = x_i) = \frac{\lambda^{x_i}}{x_i!} \cdot e^{-\lambda}$$

Funkce maximální věrohodnosti

$$L(x_1, x_2, \dots, x_n, \lambda) = \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!} \cdot e^{-\lambda}$$

$$\ln L(x_1, x_2, \dots, x_n, \lambda) = \sum_{i=1}^n [-\lambda + x_i \ln \lambda - \ln(x_i!)]$$

$$\frac{d \ln L(x_1, x_2, \dots, x_n, \lambda)}{d \lambda} = \sum_{i=1}^n \left(-1 + \frac{x_i}{\lambda}\right)$$

Hledáme takové $\hat{\lambda}$, aby platilo $\sum_{i=1}^n \left(-1 + \frac{x_i}{\lambda}\right) = 0$. Pro maximálně věrohodný odhad parametru λ dostáváme stejný výsledek jako pro odhadu metodou momentů:

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$$

2.2.4 Modely výše škod

Vycházejí z pravděpodobnostního rozdělení náhodné veličiny X (výše škody obvykle na jednu pojistnou událost v uvažovaném období) nabývající nezáporných hodnot. Nejčastěji využívaná pravděpodobnostní rozdělení jsou: [16.][18.]

- Exponenciální rozdělení,
- Paretovo rozdělení,
- Weibullovo rozdělení,
- Gamma rozdělení,
- Lognormální rozdělení.

V praktické části je použito Paretovo rozdělení.

○ Paretovo rozdělení

Náhodná proměnná X má Paretovo rozdělení $Pa(\alpha; \lambda)$ právě tehdy, když funkční vyjádření jeho hustoty pravděpodobnosti má tvar: [18.]

$$f(x) = \begin{cases} \frac{\alpha \lambda^\alpha}{(\lambda + x)^{\alpha+1}}, & x > 0, \alpha > 0, \lambda > 0 \\ 0, & x \leq 0 \end{cases}$$

distribuční funkce

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

Je dvouparametrické rozdělení s $a > 0$ a $b > 0$. Vzhledem k „těžkým koncům“ se používá v situacích s odlehlými extrémními hodnotami škod, např. v pojištění nemocenském, požárním (dřevěných budov) aj.

$$E(X) = \frac{\lambda}{\alpha - 1} \text{ pro } \alpha > 1$$

Níže jsou uvedeny vzorce pro odhad parametrů α a λ pomocí metody momentů. Právě u Paretova rozdělení se používá metoda momentů zejména proto, jelikož odhady parametrů pomocí metody maximální věrohodnosti jsou numericky náročnější.

Bez využití statistického softwaru bychom se spokojili s nimi, i když odhady získané touto metodou nemají takové vlastnosti. [18.]

$$\tilde{\alpha} = \frac{2s^2}{s^2 - \bar{x}^2}$$

$$\tilde{\lambda} = (\tilde{\alpha} - 1)\bar{x}$$

Určení maximálně věrohodných odhadů α , λ Paretova rozdělení je numericky značně náročné. Proto uvádím výsledky bez dokazování. [18.]

$$\hat{\alpha} = \frac{n}{\sum_{i=1}^n \ln\left(1 + \frac{x_i}{\lambda}\right)} = A$$

$$\hat{\alpha} = \frac{\sum_{i=1}^n \frac{1}{\lambda + x_i}}{\sum_{i=1}^n \frac{x_i}{\lambda(\lambda + x_i)}} = B$$

Nyní máme dvě různé vyjádření pro maximálně věrohodný odhad $\hat{\alpha}$. Z jejich rovnosti dostaneme funkci druhého parametru λ .

$$f(\lambda) = \frac{\sum_{i=1}^n \frac{1}{\lambda + x_i}}{\sum_{i=1}^n \frac{x_i}{\lambda(\lambda + x_i)}} - \frac{n}{\sum_{i=1}^n \ln\left(1 + \frac{x_i}{\lambda}\right)}$$

Řešením nelineární rovnice $f(\lambda) = 0$ získáme maximálně věrohodný odhad $\hat{\lambda}$. Dosazením řešení odhadu $\hat{\lambda}$ zpět vztahů pro A a B dostaneme maximálně věrohodný odhad $\hat{\alpha}$.

2.3 Ověření vhodnosti zvoleného rozdělení

Pro ověření vhodnosti zvoleného rozdělení se používají zejména

- Pearsonův χ^2 -test dobré shody,
- Kolmogorov-Smirnovův test dobré shody.

2.3.1 Pearsonův χ^2 - test dobré shody

χ^2 -test dobré shody slouží k ověření shody empirického rozdělení (rozdělení četností výběrových údajů) s předpokládaným teoretickým rozdělením s hustotou pravděpodobnosti $f(x; \Theta)$, kde Θ je vektor parametrů nejčastěji odhadnutých z výběrového souboru. Na hladině významnosti α poté dochází k rozhodnutí, zda přijmout, či zamítnout nulovou hypotézu: [19.]

H_0 : náhodná proměnná X má rozdělení s hustotou $f(x; \Theta)$.

H_1 : náhodná proměnná X nemá rozdělení s hustotou $f(x; \Theta)$.

Východiskem pro test hypotézy H_0 jsou empirické údaje x_1, x_2, \dots, x_n náhodné proměnné X , zjištěné z výběrového souboru a rozříděné do k skupin s četnostmi O_1, O_2, \dots, O_k . Pro test se využívá testovací charakteristika χ^2 vyjádřená vztahem: [17.]

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

kde χ^2 -rozdělení má počet stupňů volnosti $k - 1 - p$, přičemž p je počet odhadnutých parametrů předpokládaného teoretického rozdělení s hustotou $f(x; \Theta)$, O_i jsou empirické, skutečně zjištěné četnosti hodnot x_i diskrétní proměnné nebo intervalu (x_{i-1}, x_i) hodnot x spojitě proměnné X a E_i jsou příslušné teoretické, očekávané četnosti vyjádřené vztahem: [17.]

$$E_i = np_i$$

kde n je rozsah výběrového souboru a p_i je pravděpodobnost hodnoty x_i diskrétní proměnné s předpokládaným rozdělením $p_i = P(x_i) = P(X = x_i)$, respektive pravděpodobnost intervalu hodnot $x \in (x_{i-1}, x_i)$ spojitě proměnné $p_i = P(x_{i-1} < X \leq x_i) = F(x_i) - F(x_{i-1})$, kde $F(x)$ je distribuční funkce předpokládaného teoretického rozdělení.

Čím větší je neshoda mezi předpokládaným a skutečným rozdělením pravděpodobností základního souboru, tím budou větší rozdíly mezi skutečnými a teoretickými četnostmi. Z tohoto důvodu budeme nulovou hypotézu zamítat při velkých hodnotách těchto rozdílů

a tudíž i velké hodnotě testovacího kritéria χ . Proto kritickou oblast volíme ve tvaru $\chi > \chi_{krit}$. [14.]

Za předpokladu platnosti H_0 má náhodná veličina χ pro $n \rightarrow \infty$ asymptoticky χ^2 rozdělení pravděpodobností.

Kritická oblast je definována vztahem:

$$W = \{\chi; \chi > \chi_{\alpha, k-1-p}\} \text{ kde } \chi_{\alpha, k-1-p} = F_{\chi^2, k-1-p}^{-1}(1 - \alpha)$$

Kritickou hranici $\chi_{\alpha, k-1-p}$ hledáme v tabulkách χ^2 rozdělení pravděpodobností.

Nejčastěji se při testu používá hladina významnosti $1 - \alpha = 0,95$. Potom $\chi_{1-\alpha}^2 = \chi_{0,95}^2$ je 95. percentil χ^2 rozdělení s $k - 1 - p$ stupni volnosti. [14.][17.]

2.3.2 Kolmogorov-Smirnovův test

V případě, že náhodný výběr pochází ze spojitého rozdělení specifikovaného distribuční funkcí $F(x)$, a rozsah n výběrového souboru není postačující na použití χ^2 -testu, potom lze použít Kolmogorov-Smirnovův test dobré shody pro testování nulové hypotézy. [17.]

H_0 : náhodná proměnná X má rozdělení s distribuční funkcí $F(x)$.

H_1 : náhodná proměnná X nemá rozdělení s distribuční funkcí $F(x)$.

Tento test na rozdíl od χ^2 -testu vychází z netříděných, vzestupně uspořádaných výběrových údajů $x_1 \leq x_2 \leq \dots \leq x_n$, nedochází tedy ke ztrátě informací.

Nechť x_1, x_2, \dots, x_n je náhodný výběr z rozdělení pravděpodobností se spojitou distribuční funkcí F .

Výběrová (empirická) distribuční funkce F_n je definována vztahem: [14.]

$$F_n(x) = \begin{cases} 0 & x \leq x_{(1)} \\ \frac{i}{n} & x_{(i)} < x \leq x_{(i+1)}, i = 1, 2, \dots, n-1 \\ 1 & x > x_{(n)} \end{cases}$$

Lze dokázat, že při rostoucím n konverguje empirická distribuční funkce F_n k distribuční funkci F .

Testovací charakteristika je definována vztahem

$$d_n = \sup_x |F_n(x) - F(x)|$$

jako maximální absolutní odchylka výběrové (empirické) distribuční funkce $F_n(x)$ od spojitě distribuční funkce $F(x)$, kterou předpokládá nulová hypotéza.[14.]

V případě, že je podle předpokladu funkce $F(x)$ spojitá, je možné najít rozdělení statistiky d_n , které nezávisí na $F(x)$. Ve statistických tabulkách pro Kolmogorov-Smirnovův test jsou uvedené kvantily $d_{n;0,95}$ a $d_{n;0,99}$ tohoto rozdělení pro rozsah výběru $n = 1, 2, \dots, 100$, které při testování hypotézy slouží jako kritické hodnoty. [14.][19.]

2.4 Aproximace kolektivního modelu rizika

Aproximativní metody pro stanovení rozdělení pravděpodobnosti kolektivního rizika S se v praxi používají místo přesných analytických vzorců pro výpočet distribuční funkce $F_S(x)$, protože ty jsou často příliš složité a pracné. Pozornost je nejčastěji zaměřena na aproximaci rozdělení S pomocí normálního rozdělení a posunutého gama rozdělení. [16.]

2.4.1 Aproximace normálním rozdělením

Předpokládejme, že pro kolektivní riziko S známe základní charakteristiky: střední hodnotu $E(S) = \mu$ a rozptyl $D(S) = \sigma^2$. Protože $S = \sum_{i=1}^N X_i$ je součtem nezávislých a identicky rozdělených náhodných proměnných, nabízí se podle centrální limitní věty aproximace rozdělení S normálním rozdělením $N(\mu = E(S), \sigma^2 = D(S))$. [16.][17.]

Předpokládejme, že pro kolektivní riziko S známe základní charakteristiky: střední hodnotu $E(S) = \mu$ a rozptyl $D(S) = \sigma^2$. Protože $S = \sum_{i=1}^N X_i$ je součtem nezávislých a identicky rozdělených náhodných proměnných, nabízí se podle centrální limitní věty aproximace rozdělení S normálním rozdělením $N(\mu = E(S), \sigma^2 = D(S))$.

Pak pro libovolné s platí: [17.]

$$G(s) = P(S \leq s) = P\left(\frac{S - \mu}{\sigma} \leq \frac{s - \mu}{\sigma}\right) \sim \Phi\left(\frac{s - \mu}{\sigma}\right)$$

kde $\Phi(z)$ je distribuční funkce normovaného normálního rozdělení.

Čím větší je počet N pojistných plnění, tím je aproximace $G(s)$ pomocí distribuční funkce normovaného normálního rozdělení lepší. Předností je jednoduchost, má však i své nedostatky. [19.]

- Bez ohledu na hodnoty μ a σ^2 , při normální aproximaci platí, že $P(S < 0) > 0$. Navzdory podmínce, že všechna pojistná plnění jsou nezáporná, je zřejmé, že ve skutečnosti $P(S < 0) = 0$.
- Hustota normálního rozdělení je symetrická, což znamená, že pravý konec konverguje velmi rychle k nule. Při mnohých typech pojištění je však rozdělení S pravostranné, s pomalým klesáním pravé strany hustoty k ose x , tedy s dost vysokou pravděpodobností i extrémních pojistných plnění. Pro tyto typy pojištění má normální aproximace tendenci podhodnocovat hodnoty $P(S > x)$ pro velké hodnoty x .

Z pohledu pojišťovatele je to velmi nežádoucí skutečnost, protože vysoká pojistné plnění pro něho mají závažný finanční důsledek. [16.]

2.4.2 Aproximace posunutým gama rozdělením

Nechť jsou známé či spolehlivě odhadnutelné tři základní charakteristiky celkového pojistného plnění S . A to střední hodnota $E(S) = \mu$, rozptyl $D(S) = \sigma^2$ a koeficient šikmosti γ . V takovém případě můžeme použít posunuté gama rozdělení pro aproximaci rozdělení S . Posunuté gama rozdělení je rozdělení náhodné veličiny $k + Y$, kde k je konstanta, náhodná veličina Y má gama rozdělení s parametry α a β , symbolicky $Y \sim G(\alpha; \beta)$, a charakteristiky $E(Y) = \frac{\alpha}{\beta}$, $D(Y) = \frac{\alpha}{\beta^2}$ a koeficientem šikmosti $\frac{2}{\sqrt{\alpha}} > 0$. [16.]

Parametry k , α a β se určí za předpokladu, že náhodná proměnná $k + Y$ má první tři momenty shodné s adekvátními momenty S , které jsou známé. Získává se tím systém tří rovnic s neznámými k , α a β : [17.][18.]

$$\gamma = \frac{2}{\sqrt{\alpha}}$$

$$\sigma^2 = \frac{\alpha}{\beta^2}$$

$$\mu = k + \frac{\alpha}{\beta}$$

Protože hodnoty μ , σ^2 a γ jsou podle předpokladu známé, postupně se vypočítají hodnoty neznámých parametrů

$$\alpha = \frac{4}{\gamma^2}$$

$$\beta = \sqrt{\frac{\alpha}{\sigma^2}}$$

$$k = \mu - \frac{\alpha}{\beta}$$

Výhodou této aproximace je to, že je jednodušší určit pravděpodobnost $P(a < k + Y < b)$, než pravděpodobnost $P(a < S < b)$. Takové pravděpodobnosti v případě gama rozdělení umožňuje určit většina statistických programů, včetně tabulkového procesoru Excel.

3 Využití simulace pro aproximaci rozdělení kolektivního rizika

Simulace provádíme s cílem prozkoumat chování složitějších modelů, v našem případě pro nalezení přibližného tvaru složeného rozdělení kolektivního rizika S . Vycházíme při tom ze znalosti pravděpodobností základních jevů, které jsou pro celý proces směrodatné. Pokud určíme tyto pravděpodobnosti a mechanismus, jak nové jevy vznikají z jevů základních, můžeme postupovat dvěma cestami. Buď nové pravděpodobnosti vypočítáme postupy k tomu určenými, anebo proces simulujeme - to znamená, že ho mnohokrát opakujeme a jednotlivé pravděpodobnosti odhadujeme pomocí relativních četností jevů, které nás zajímají. Těmto metodám se také říká metody Monte Carlo. Při modelování nám obvykle pomáhá počítač, ale někdy je lze realizovat i manuálně. [12.]

Simulace a počítání relativních četností představují často jednodušší cestu jak odhadnout neznámé pravděpodobnosti. Naše simulace povede k validním výsledkům, pokud jsme model dobře sestavili. Simulace v kontextu statistiky lze definovat jako využívání tabulek náhodných čísel nebo generování náhodných čísel počítačem s cílem napodobit reálné procesy. [12.]

3.1 Metoda Monte Carlo

Samotná metoda Monte Carlo byla poprvé formulována v 18. století na problému tzv. Buffonovi jehly Georges-Louis Leclercem, Comte de Bufonem. Později ji zdokonalila a prakticky použila dvojice J. von Neumann a S. Ulam při vývoji atomové bomby během 2. světové války. Při výzkumu chování neutronů bylo třeba vyřešit problém, jaké procento neutronů v určité spršce pronikne nějakou překážkou, např. nádrží vody určitých rozměrů. Při řešení tohoto problému předpovědi života neutronu byla použita technika kola rulety, odtud plyne i název metody. [11.]

Při řešení metody Monte Carlo je použita třída algoritmů pro simulaci systémů, která je založena na provádění náhodných experimentů s modelem systému a jejich vyhodnocení. Výsledkem provedení velkého množství experimentů je obvykle pravděpodobnost určitého jevu, se kterou můžeme dále pracovat. Jistou výhodou metody je její snadná implementace, která je bohužel vyvážena relativně malou přesností. [12.] [22.]

3.1.1 Přesnost odhadu

Přesnost a efektivnost celého výpočtu metodou Monte Carlo na počítači je určen třemi základními faktory: [15.]

1. kvalitou generátoru náhodných čísel, resp. pseudonáhodných čísel
2. výběrem racionálního algoritmu výpočtu
3. kontrolou přesnosti získaného výsledku.

Přesnost metody Monte Carlo obecně roste s počtem opakování n . Přesnost dále závisí na konkrétní aplikaci, zda lze odvodit nějaké další asymptotické vlastnosti. K odhadu chyby výsledku získaného metodou Monte Carlo se většinou používá střední kvadratická chyba aritmetického průměru. Chyba výsledku získaného pomocí n historií je úměrná $\frac{1}{\sqrt{n}}$. Aby se tudíž zlepšil výsledek o jeden řád, musí se počet historií zvýšit alespoň o dva řády. Abychom tedy získali výsledek s přesností na 6 desetinných míst, což odpovídá přesnosti jiných metod, musíme získat 1012 historií.² [15.]

3.1.2 Postup realizace metody Monte Carlo

Simulovaným náhodným procesem jsou hodnoty s_1, s_2, \dots, s_n celkových škod S . Realizací simulace získáme nejprve hodnoty počtu pojistných škod N a následně výšky pojistných plnění X , přičemž předpokládáme znalost rozdělení pravděpodobnosti a jeho parametrů nebo alespoň odhadů parametrů rozdělení pravděpodobnosti náhodných proměnných N a X . Uvedený postup se realizuje podle následujícího algoritmu. [24.]

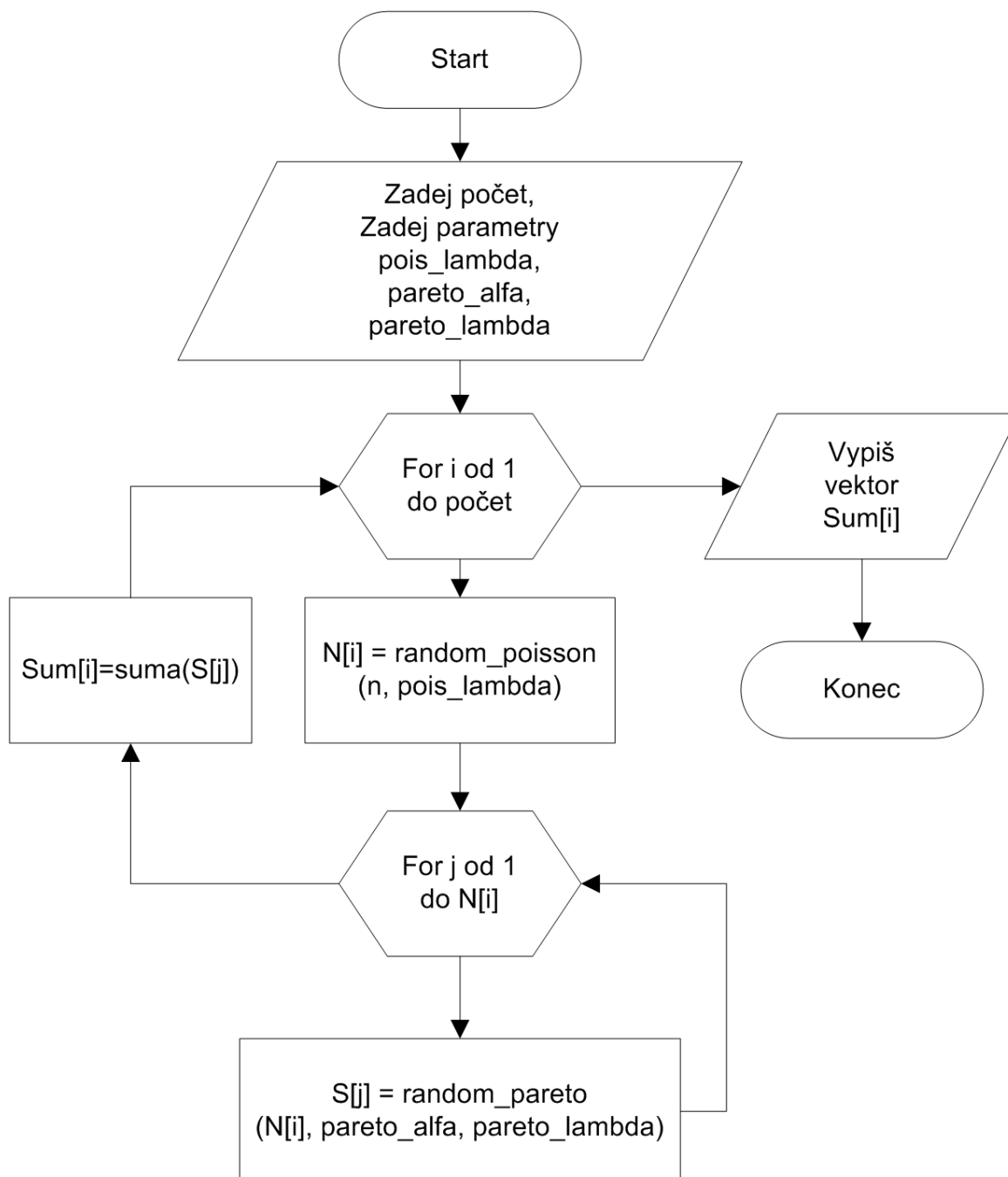
Vygeneruje se náhodné číslo n_1 z rozdělení počtu škod N pomocí generátoru náhodných čísel z rozdělení N . Pokud programovací jazyk přímo dovoluje vygenerovat náhodné číslo z daného rozdělení, pak je toto číslo konečné. Takto aplikujeme v programovacím jazyku R. V opačném případě, kdy nejsme schopni vygenerovat přímo náhodné číslo z daného rozdělení pravděpodobnosti, jsme nuceni vygenerovat náhodné číslo r_1 z rovnoměrného rozdělení na intervalu $< 0; 1 >$ a poté na takto získané číslo využít vztahu $n_1 = Q(r_1)$, kde $Q(r_i)$ je kvantilová funkce rozdělení N , n_1 je náhodné číslo z daného rozdělení N .

Analogicky vygenerujeme právě n_i hodnot individuálních škod x_1, x_2, \dots, x_{n_1} z rozdělení výšky škod X . Součet těchto hodnot bude první vygenerovanou hodnotou celkových škod s_1 .

² Předpokladem takového odhadu přesnosti je, že aritmetický průměr výstupů má normální distribuci. Nejsme-li si jisti distribucí výsledků (což většinou nejsme, proto používáme tuto metodu), je vhodné počet opakování raději nadstřelit, dovolí-li to hardware.

První dva kroky zopakujeme n -krát, přičemž n musí být dostatečně velké číslo, nejlépe několik stovek až tisíců. Dostaneme tak hodnoty s_1, s_2, \dots, s_n z neznámého rozdělení celkových škod S , představující jeden náhodný výběr z tohoto rozdělení.

Vzhledem k dostatečnému počtu vygenerovaných hodnot můžeme rozdělení pravděpodobnosti kolektivního rizika S modelovat metodami statistiky, například použitím testů dobré shody.



Obr. 2: Algoritmus generování náhodných čísel metodou Monte Carlo, zdroj vlastní

4 Programovací jazyk R

V praktické části byl využit k řešení programovací jazyk R. Tento volně šiřitelný software pro statistické výpočty je vhodný jak díky své distribuci, tak i širokou základnou uživatelů, kteří vytvářejí vhodné prostředí pro učení a konzultaci problémů.

R je jazykem a prostředím pro statistické výpočty a grafiku. Jedná se o GNU projekt podobný jazyku a prostředí S vyvinutému v Bell Laboratories (dříve AT&T, nyní Lucent Technologies) Johnem Chambersem a kolegy. R lze považovat za odlišnou implementaci jazyka S. Existují některé významné rozdíly mezi R a S, nicméně většina kódu napsaného pro S bude beze změn fungovat též v R. [26.]

- **Široká škála metod a snadná rozšiřitelnost**

R poskytuje širokou škálu statistických metod (lineární a nelineární modely, klasické testy, analýza časových řad, klasifikace, klastrování, ...) a grafických technik. R je dále snadno rozšiřitelné o další metody. Jazyk S implementovaný v R je často volen jako nástroj pro výzkum nových statistických metod, přičemž R samotné poskytuje cestu k účasti na těchto aktivitách. [21.]

- **Grafy na profesionální úrovni**

Jednou z nejsilnějších částí R je snadnost, s kterou lze vytvářet dobře navrhnuté obrázky a grafy v profesionální kvalitě. Do grafů lze snadno v případě potřeby vkládat matematické symboly a vzorce. Standardní nastavení pro kresbu grafů bylo voleno s maximální pečlivostí, nicméně uživateli je ponechána plná kontrola nad výsledným vzhledem grafu. [26.]

4.1 Prostředí R

Nejprve je nutné si stáhnout instalační balíček programu R³. Tento je určen přímo pro Windows jako tzv. samoinstalační balíček. Ten zabezpečí instalaci všech důležitých součástí.

4.1.1 Analýza dat

R poskytuje softwarové prostředky pro manipulaci s daty, výpočty a grafická zobrazování. R zahrnuje prostředky pro efektivní manipulaci a ukládání dat, sadu operátorů pro výpočty na polích a maticích, rozsáhlé, konzistentní a integrované prostředky pro analýzu dat, grafické

³ Samoinstalační balíček pro Windows / Linux / Unix je ke stažení na oficiálních stránkách R - <http://cran.r-project.org/bin/windows/base/>.

prostředky pro analýzu a zobrazování dat, ať již na obrazovce nebo v tištěné podobě, dobře navržený, jednoduchý a efektivní programovací jazyk obsahující podmínky, cykly, uživatelem definované rekurzivní funkce, prostředky pro vstup a výstup. [26.][21.]

4.1.2 R jako programovací jazyk

R, podobně jako S je navrženo jako skutečný programovací jazyk a umožňuje tedy uživateli přidávat dovednosti definování nových funkcí. Většina samotného systému je napsána v R dialektu jazyka S, což umožňuje uživateli snadno sledovat volené algoritmy. Výpočetně náročné postupy mohou být naprogramovány v C, C++ nebo Fortranu, poté připojeny k R a volány za běhu. Pokročilí uživatelé mohou psát kódy v C, jež přímo manipulují s R objekty. [21.]

4.1.3 Více než statistický software

Mnozí uživatelé si pod R představují statistický software. Je lepší si pod R představit prostředí, v němž jsou kromě jiného implementovány statistické metody. Pomocí balíčků (packages) lze R snadno rozšiřovat. Přibližně osm balíčků je součástí oficiální distribuce R. Mnohé další balíčky pokrývající velmi rozsáhlou oblast moderní statistiky jsou dostupné přes CRAN (Comprehensive R Archive Network). [26.]

V základním prostředí je například vytvořena podpora velkého množství rozdělení pravděpodobností, jak je možné vidět v následující tabulce.

Tab. 2: Standardní rozdělení pravděpodobnosti v základním prostředí R, zdroj [21.]

Rodění	Příkaz v R	Parametry	Výchozí hodnota
Beta	beta	shape1, shape2	
Binomial	binom	size, prob	
Cauchy	cauchy	location, scale	0, 1
Chi-square	chisq	df	
Exponential	exp	1/mean	1
F	f	df1, df2	
Gamma	gamma	shape, 1/scale	NA. 1
Geometric	geom	prob	
Hypergeometric	hyper	m, n, k	
Log-normal	lnorm	mean, sd	0, 1
Logistic	logis	location, scale	0, 1
Normal	norm	mean, sd	0, 1
Poisson	pois	lambda	
Student	t	df	
Uniform	unif	min, max	0, 1
Weibull	weibull	shape	

o **Dokumentace**

R má svůj vlastní, LaTeXu podobný formát pro tvorbu dokumentace, jenž je používán při poskytování jak on-line dokumentace v několika formátech, tak tištěné dokumentace.

5 Aplikace metod aproximace kolektivního rizika S

Nyní přejdeme k vyřešení příkladu, na kterém si ukážeme postup modelování metodou Monte Carlo. Samotné simulaci bude předcházet vyšetření nutných přípravních úkolů.

5.1 Rozdělení počtu škod

Z nejmenované pojišťovny známe údaje o počtu pojistných plnění pro 80000 uzavřených pojistných smluv v minulém roce. [2.]

5.1.1 Výpočet v R

Nejprve načteme dodatečné knihovny, které budeme používat. Dodateční knihovny načteme příkazem:

```
> library()
```

Do tohoto příkazu vložíme buďto fyzickou cestu k `package` nebo, pakliže jej stáhneme do adresáře `/library/` ve složce R, pouze název požadovaného `package`.

Data jsou uložena v sešitu tabulkového procesoru Excel. Pro práci s nimi je nutné uložit tato data do souboru `csv`. Poté je můžeme načíst do prostředí programu R.

```
tab.poissonData = read.csv(  
+ file='c:\\Users\\Berka\\Desktop\\  
+ diplomka-final\\prakticka\\poisson-r.csv',  
+ sep=';',header=F)
```

Zde určíme fyzickou cestu k souboru, který načítáme, separátor, kterým jsou data oddělena a zda data obsahují první řádek hlavičky.

Takto načtená data si můžeme vypsát

```
> csvtab.poissonData
```

Nyní je nutné určit parametr λ , díky kterému budeme moci pokračovat v rozboru daného souboru.

```
lambda.ocekavana = vector()  
for(i in 1: length(csvtab.poissonData)) {
```



```

      lambda.ocekavana           [i]           =
csvtab.poissonData[1,i]*csvtab.poissonData[2,i]
}
sum.poissonData = sum(csvtab.poissonData[2,])
lambda.odhad = sum(lambda.ocekavana)/sum.poissonData
> lambda.odhad

```

Pakliže bychom neměli již tabulku četností, ale zdrojová neseříděná data, mohli bychom pro zjištění parametru použít příkaz

```
lambda.est = mean(poisson)
```

V obou případech vyjde λ stejně, ve druhém případě je však odhad o poznání jednodušší.

- o **χ^2 test dobré shody v R**

Nyní využijeme knihovny `vcd`, která obsahuje kód pro zjištění shody s Poissonovým rozdělením.

```

gof = goodfit(poisson, type = c("poisson"),par =
+ list(lambda = lambda.est))
summary(gof)

```

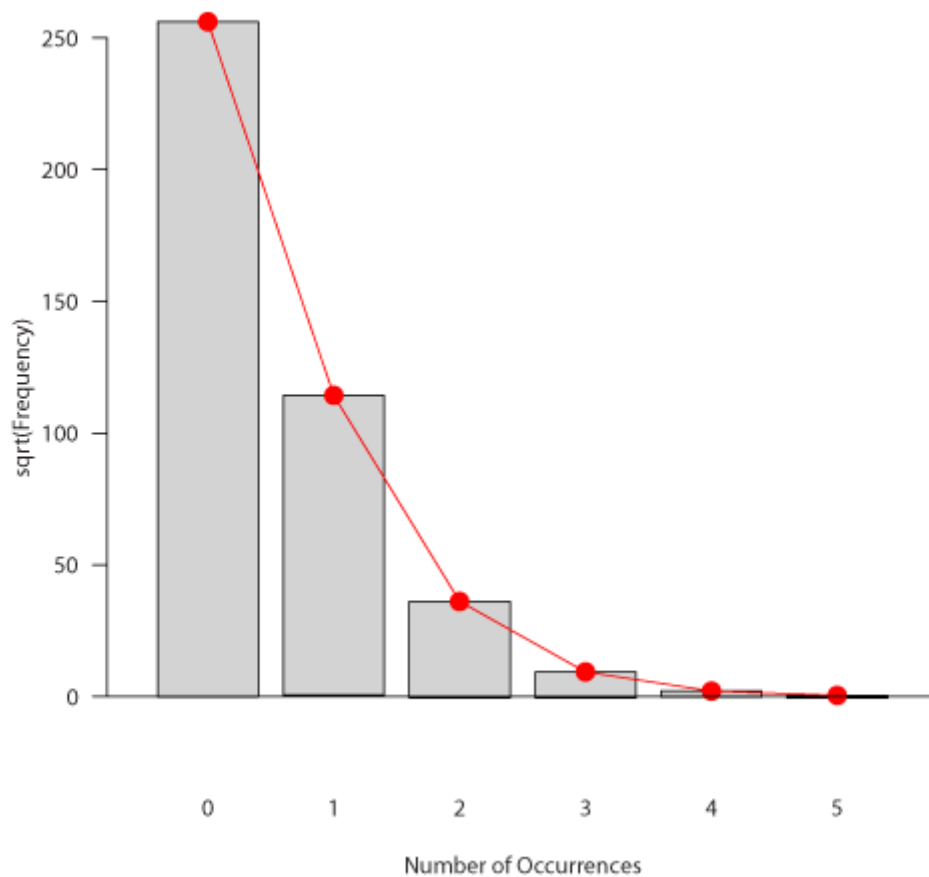
```

      Goodness-of-fit test for poisson distribution

```

	X ²	df	P(> X ²)
Pearson	8.307104	5	0.1401030

```
plot(gof)
```



Obr. 3: Výsledné grafické zobrazení z programu R, zdroj vlastní

Výsledek v programu R jednoznačně neprokazuje, že Poissonovo rozdělení je vhodným. Pro kontrolu vytvoříme χ^2 test shody také v tabulkovém procesoru Excel.

Nezkrácený optimalizovaný zdrojový kód naleznete v Příloha A: Ověření Poissonova rozdělení.

5.1.2 Výpočet v programech Excel a Statgraphics

Jelikož parametr λ vypočtený metodou momentů je pro Poissonovo rozdělení stejný jako parametr vypočtený metodou maximální věrohodnosti, nebudu zde uvádět názornější výpočet pro tabulkový procesor Excel.

Ověříme předpoklad, že počet pojistných plnění pro jednu pojistnou smlouvu má Poissonovo rozdělení pravděpodobnosti s parametrem λ , odhadnutého pomocí aritmetického průměru.

H_0 : náhodná proměnná X má Poissonovo rozdělení pravděpodobnosti

H_1 : náhodná proměnná X nemá Poissonovo rozdělení pravděpodobnosti

Tab. 3: Test shody pro rozdělení počtu škod Excel, zdroj vlastní

i	claims in general insurance portfolio		nc*f	pi	npi	χ
	number of claims	frequency				
1	0	65623	0	0,81939	65550,88	0,079347
2	1	12929	12929	0,16322	13057,74	1,269192
3	2	1344	2688	0,01626	1300,55	1,451589
4	3	98	294	0,00108	86,3565	1,569886
5	4	5	20	0,00005	4,300556	0,113758
6	5	1	5	2,14E-06	0,17133	4,007882
7	6	0	0	7,11E-08	0,005688	0,005688
		80000	16360	1	χ	8,497343
		\bar{x}	0,2045		$\chi_{\alpha,n}^2$	11,0705

Jelikož na hladině významnosti $\alpha = 0,05$ je $\chi < \chi_{\alpha,n}^2$, na dané hladině významnosti nulovou hypotézu nezamítáme. Můžeme přijmout předpoklad, že počet pojistných plnění má Poissonovo rozdělení pravděpodobnosti.

Tab. 4: Test shody pro rozdělení počtu škod Statgraphics, zdroj vlastní

	Lower	Upper	Observed	Expected	
	Limit	Limit	Frequency	Frequency	Chi-Squared
at or below		0,0	65623	65550,88	0,08
	1,0	1,0	12929	13057,74	1,27
	2,0	2,0	1344	1300,55	1,45
	3,0	3,0	98	86,36	1,57
	4,0		6	4,48	0,52

Chi-Squared = 4,88752 with 3 d.f. P-Value = 0,180218

Jak je možné vidět, také podle statistického programového balíku Statgraphics, ač vytvořil malinko jiné intervaly, můžeme na hladině $\alpha = 0,05$ hypotézu o shodě s Poissonovým rozdělením přijmout.

5.2 Rozdělení výše škod

Následující soubor dat se 40 hodnotami (v 1000 Kč) pochází z portfolia domácího pojištění. Budeme zjišťovat, zda hodnoty odpovídají Paretovu rozdělení pravděpodobnosti, tedy $X \sim Pareto(\alpha, \lambda)$. [2.]

10 11 15 22 28 30 32 36 38 48 51
 55 56 68 68 85 87 94 103 104 105 106
 109 119 121 137 178 181 226 287 310 321 354
 393 438 591 1045 1210 1212 2423

5.2.1 Výpočet v R

Tyto hodnoty zavedeme do programu R jako vektor pomocí příkazu

```
> x = c(10, 11, 15, 22, 28, 30, 32, 36, 38, 48, 51,
+ 55, 56, 68, 68, 85, 87, 94, 103, 104, 105, 106,
+ 109, 119, 121, 137, 178, 181, 226, 287, 310, 321, 354,
+ 393, 438, 591, 1045, 1210, 1212, 2423)
```

Takto v programu vytvoříme vektor x s požadovanými hodnotami.

Jeho průměr spolu se směrodatnou odchylkou zjistíme pomocí příkazů:

```
> prumer_x = mean(x)
> s = sd(x)
> s_2 = sd(x)
```

Nyní máme v proměnné `prumer_x` uložen průměr, v proměnné `s` výběrovou směrodatnou odchylku a konečně v proměnné `s_2` pak výběrovou charakteristiku s^2 . Tyto hodnoty budeme potřebovat pro zjištění parametrů pomocí metody momentů a dále pak metody maximální věrohodnosti.

- **Vypočtení parametrů pomocí metody momentů v R**

Nyní pomocí metody momentů vypočteme parametry pro Paretovo rozložení pravděpodobností α a λ .

Do programu R tyto vzorečky opět zaneseme pomocí kódu

```
> alfa = (2 * s^2) / (s^2 - prum_x^2)
> lambda = (alfa - 1) * prum_x
```

nebo

```
> alfa = (2 * s_2) / (s_2 - prum_x^2)
```

Nyní máme v proměnné `alfa` uložen odhad parametru α a v proměnné `lambda` odhad parametru λ .

Maximálně věrohodné odhady $\hat{\alpha}$, $\hat{\lambda}$ jsou ale kvalitnější než odhady $\tilde{\alpha}$ a $\tilde{\lambda}$ provedené metodou momentů. Výběrová charakteristika s^2 , z které odhady získáváme má pro Paretovo rozdělení vysokou variabilitu, jelikož i vysoké hodnoty x jsou v tomto rozdělení značně pravděpodobné.

V programu R bychom mohli tento výpočet získat pomocí následujícího kódu.

- **χ^2 test dobré shody v R**

Nejprve je nutné zjistit reálné výskyty v jednotlivých intervalech. Pro tento úkol si vytvoříme funkci s názvem `interval`:

```
interval <- function(x, dolni_mez, horni_mez) {
  a = 0;
  for (i in 1:length(x)) {
    if( (x[i]>dolni_mez) & (x[i]<horni_mez) ) {
      a = a+1
    }
  }
}
```

```

    }
    return(a)
}

```

Jejím voláním s jednotlivými mezemi zjistíme počet výskytů.

```

> interval(x, 0, 42.594)
[1] 9
> interval(x, 42.594, 102.270)
[1] 9
...

```

Vložíme tyto hodnoty do vektoru O_i a do vektoru HH pravé strany intervalů.

```

> O_i = c(interval(x, 0, 42.594),
+ interval(x, 42.594, 102.270),
+ interval(x, 102.270, 196.444),
+ interval(x, 196.444, 322.336),
+ interval(x, 322.336, max(x)+1))
> HH = c(42.594,102.27,196.444,322.336)

```

Nyní vytvoříme $F(HH)$:

```

> F_HH <- function(meze, lambda, alfa) {
+   HH_vektor = vector()
+   for (i in 1:length(meze)) {
+     HH_vektor[i] = 1 - ( lambda /
+       + (lambda + meze[i]) )^alfa
+   }
+   return(HH_vektor)
+ }

```

Požadovaná funkce bude vracet vektor s výslednými hodnotami, jak je možné vidět dále.

```

> F_HH_cisla = F_HH(HH, lambda, alfa)
[1] 0.2000307 0.4000465 0.6000472 0.7500381

```

Nyní vypočítáme p_i

```

> p_i = function(F_HH) {
+   suma = 0

```

```

+   vektor = vector()
+   for (i in 1:length(F_HH)) {
+       vektor[i] = F_HH[i] - suma
+       suma = suma + vektor[i]
+   }
+   vektor[5] = 1 - suma
+   return(vektor)
+ }
> p_i_cisla = p_i(F_HH_cisla)

```

V dalším kroku vypočteme E_i . Mohli bychom vzít jednoduše vektor `p_i_cisla` a vynásobit každé číslo v něm `sum_x`.

```

> E_i_cisla = p_i_cisla * 40

```

Pak ale dostaneme v našem případě čísla typu `double`. Tedy neceločíselné hodnoty. Ve snaze převést tato čísla na celá použijeme nejspíše příkaz `as.integer()`, který převede desetinná čísla na celá. Bohužel ale bere pouze celou část a část za desetinou čárkou ztratíme. Abychom předešli této nepříjemnosti, vytvoříme opět funkci, která správně převede čísla z desetinných na celá.

```

> E_i = function(p_i, sum_x) {
+   E_i_x = p_i
+   x_i = 0
+   for (i in 1:length(p_i)) {
+       p_i[i] = p_i[i] * 40
+       E_i_x = as.integer(p_i[i])
+       x_i = p_i[i] - E_i_x
+       if (x_i < 0.5) {
+           p_i[i] = E_i_x
+       } else {
+           p_i[i] = E_i_x + 1
+       }
+   }
+   return(p_i)
+ }
> E_i_cisla = E_i(p_i_cisla, sum(x))

```

Nyní přistoupíme k poslednímu kroku, vypočtení χ^2 testu.

```
> chi_square = function(Oi, Ei) {  
+ chi_sq = vector()  
+ for(i in 1:length(Oi)) {  
+ chi_sq = (Oi - Ei)^2 / Ei  
+ }  
+ return(chi_sq)  
+ }  
> chi = chi_square(Oi, E_i_cisla)  
> sum(chi)  
[1] 1.816667
```

Nezkrácený optimalizovaný zdrojový kód je uveden v Příloha B: Ověření Paretova rozdělení.

5.2.2 Výpočet v programech Excel a Statgraphics

Nyní vypočítáme odhady $\hat{\alpha}$ a $\hat{\lambda}$ pomocí metody maximální věrohodnosti.

- **Vypočtení parametrů pomocí metody maximální věrohodnosti**

Pro tento úkol využijeme funkce Řešitel v tabulkovém procesoru Excel.

Tab. 5: Výpočet parametrů metodou maximální věrohodnosti, zdroj vlastní

		A		B	
xi	xi - průměr x	1+xi/lambda	ln(1+xi/lambda)	1/(lambda+xi)	xi/(lambda(lambda+xi))
10	68998,15563	1,0160731	0,01594525	0,00158188	2,79111E-05
11	68473,80563	1,0176804	0,017525879	0,001579381	3,06489E-05
15	66396,40563	1,0241096	0,023823533	0,001569466	4,15062E-05
22	62837,95563	1,0353607	0,034749885	0,001552411	6,01509E-05
28	59865,85563	1,0450045	0,044021239	0,001538085	7,57822E-05
30	58891,15563	1,0482192	0,047092686	0,001533368	8,09227E-05
32	57924,45563	1,0514338	0,050154729	0,00152868	8,60288E-05
36	56015,05563	1,057863	0,056250828	0,001519389	9,61392E-05
38	55072,35563	1,0610776	0,059284998	0,001514786	0,000101144
48	50478,85563	1,0771507	0,074319274	0,001492182	0,00012568
51	49139,80563	1,0819726	0,078785831	0,001485532	0,000132885
55	47382,40563	1,0884018	0,084710376	0,001476757	0,000142384
56	46948,05563	1,0900091	0,086186044	0,00147458	0,00014474
68	41891,85563	1,1092968	0,103726268	0,001448941	0,000172429
68	41891,85563	1,1092968	0,103726268	0,001448941	0,000172429
85	35221,90563	1,136621	0,128059786	0,001414108	0,000209906
87	34475,20563	1,1398356	0,130884011	0,00141012	0,000214187
94	31924,75563	1,1510867	0,140706455	0,001396337	0,000228965
103	28789,60563	1,1655525	0,153195181	0,001379007	0,000247512

...

1212	882331,4556	2,9480541	1,081145317	0,000545209	0,001095725
2423	4623897,606	4,8945008	1,588112298	0,00032839	0,001303002

Do vzorečku pro zjištění odhadu $\hat{\alpha}$ vkládáme odhady získané metodou momentů. Poté si v Excelu vytvoříme tabulku s výpočtem jednotlivých kroků spolu s výslednými hodnotami.

Tab. 6: Výsledné hodnoty metody maximální věrohodnosti, zdroj vlastní

maximální věrohodnost		původní hodnoty	
A	3,4763	A	3,2413
B	3,4763	B	3,42113
$\hat{\lambda}$	622,159	$\tilde{\lambda}$	565,867
$f(\lambda)$	1,1E-07		

Výsledkem je odhad A a B spolu s odhadem $\tilde{\lambda}$ získaným metodou momentů. Výsledky můžeme vidět v pravé části tabulky nadepsané jako původní hodnoty.

Víme, že rovnice $A - B$ musí být rovna 0 ($f(\lambda) = 0$). Využijeme tudíž Řešitele, kde buňce $f(\lambda)$ nastavíme požadovanou hodnotu 0 a budeme měnit hodnotu $\hat{\lambda}$.

Tím, že vše je uvnitř sešitu Excel provázáno vzorci, tudíž je zaručeno, že se změní i A a B tak, aby $f(\lambda) = 0$ (nebo číslo limitně se blíží k nule).

- χ^2 test dobré shody

Pro zjištění, zda se v našem případě jedná opravdu o Paretovo rozložení pravděpodobnosti a pro doplnění tabulky získaných a očekávaných hodnot využijeme χ^2 test dobré shody.

H_0 : náhodná proměnná X má Paretovo rozdělení pravděpodobnosti

H_1 : náhodná proměnná X nemá Paretovo rozdělení pravděpodobnosti

Pro tento test jsou zadány intervaly, které spolu s výsledkem testu jsou vidět v tabulce.

Pro výpočet použijeme parametr λ získaný metodou momentů.

Tab. 7: Test shody pro rozdělení výšky škod, zdroj vlastní

i	$< x_i - 1; x_i)$	O_i	HH	$F(HH)$	p_i	E_i	E_i	$\frac{(O_i - E_i)^2}{E_i}$
1	$< 0; 42,594)$	9	42,594	0,20003	0,20003	8,00123	8	0,125
2	$< 42,594; 102,270)$	9	102,27	0,40005	0,20002	8,00063	8	0,125
3	$< 102,270; 196,444)$	10	196,444	0,60005	0,2	8,00003	8	0,5
4	$< 196,444; 322,336)$	4	322,336	0,75004	0,14999	5,99963	6	0,6666667
5	$< 322,366; +\infty)$	8			0,24996	9,99848	10	0,4
	Σ	40			1	χ		1,8166667
						stup.volnosti		3
						α		0,05
						$P(\chi > \chi_{\alpha,n}^2) = \alpha$		5,991465

Kritickou hodnotu χ^2 získáme buďto ze statistických tabulek nebo přímo z Excelu příkazem =CHIINV(α , STUPNĚ_VOLNOSTI).

Protože $\chi < \chi_{\alpha,n}^2$ nezamítáme nulovou hypotézu, že výšky škod mají Paretovo rozdělení pravděpodobnosti, na hladině významnosti $\alpha = 0,05$.

5.3 Simulace Monte Carlo v R

Dle výsledků v předchozích dvou podkapitolách jsme zjistili, že počet škod má Poissonovo rozdělení pravděpodobnosti $N \sim Po(\lambda = 1992)$ a výšky škod mají Paretovo rozdělení pravděpodobnosti $Pa(\alpha = 3,4763, \lambda = 622.159)$.

S využitím těchto výsledků můžeme přistoupit k samotné simulaci.

5.3.1 Simulace v R

V R vygenerujeme 10000 hodnot z Poissonova rozdělení pravděpodobnosti s parametrem $\lambda = 1992$.

```
> x_pois = rpois(10000,1992)
```

V dalších řádcích načteme knihovnu `actuar`, která mimo jiné obsahuje již hotovou funkci pro generování náhodných čísel z Paretova rozdělení pravděpodobnosti. Jako počet hodnot uvedeme vždy i -tou vygenerovanou hodnotu z rozdělení množství škod. Tento postup opakujeme až do vygenerování všech výšek individuálních škod.

```
> library(actuar)
> x_pareto = rpareto(x_pois[1], 1.7394, 37277.8)
```

Nyní všechny generované náhodné veličiny daného i -tého opakování sečteme a dostaneme i -tou hodnotu kolektivního rizika S .

```
> sum(x_pareto)
```

Pro zjednodušení si opět vytvoříme funkci, která bude celou záležitost řešit za nás a jejíž výsledkem bude vektor kolektivního rizika.

```
> x = function(n, pois_lambda, pareto_alfa, pareto_lambda) {
+   m = list()
+   m_sum = array(0, dim = n)
+   x_pois = rpois(n, pois_lambda)
+
+   for (i in 1:length(x_pois)) {
+     m[[i]] = rpareto(x_pois[i], pareto_alfa,
+       + pareto_lambda)
+     m_sum[i] = sum(m[[i]])
+   }
+   return(m_sum)
+ }
```

Potom stačí pouze zavolat vytvořenou funkci pomocí níže uvedeného příkazu. Pod proměnnou `monte_carlo` se skrývá vektor všech vygenerovaných hodnot kolektivního rizika S , kterých je právě 10000.

```
> monte_carlo = x(1000, 2000, 1.7394, 37277.8)
```

Výsledné hodnoty si můžeme zobrazit jednoduše voláním příkazu

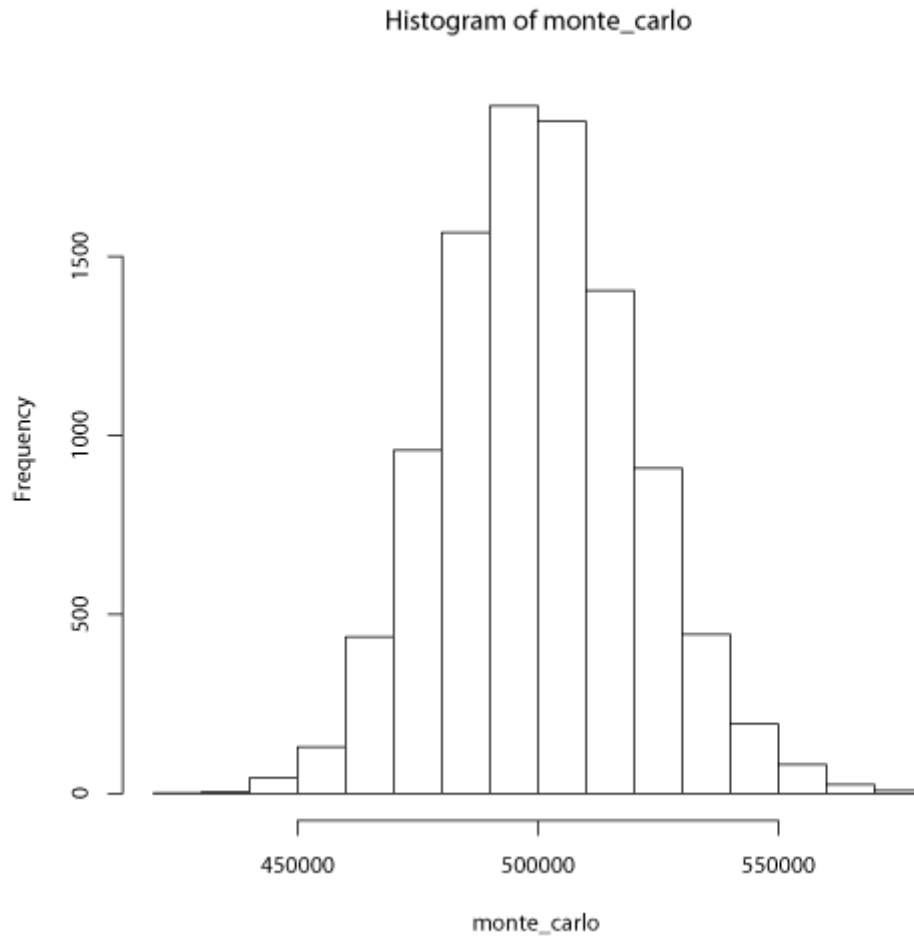
```
> monte_carlo  
[1] 492673,2114 528286,9292 497193,1973 492521,1897 488510,0079  
[5] 485589,3477 514243,1717 509208,3092 507731,9773 500341,8403  
[10] 520069,7322 484191,3625 488641,6521 523933,5354 486297,7953  
[15] 502459,2714 488176,0003 482816,1853 487603,862 512157,273  
[20] 101571980 127444461 95673600 202458366 103371045 91800898  
...  
[10000] 512704,476 548048,463 492534,8404 429290,1495 509751,953
```

Vygenerovali jsme si 10000 hodnot, které odpovídají, v našem případě, simulaci 10000 let, které by pojišťovna musela sbírat data. Nyní vyšetříme hodnoty a porovnáme je s některými nejčastějšími rozděleními kolektivního rizika.

Abychom mohli hodnoty vyšetřit v některém ze statistických programů, je nutné je vyexportovat z prostředí R.

K tomu slouží např. `package foreign`, který umí exportovat hodnoty různých formátů.

```
> library(foreign)
```



Obr. 4: Histogram výsledných hodnot získaných metodou Monte Carlo, zdroj vlastní

Typ, ve kterém jsou data uložena je nutné z vektoru převést na frame.

```
> monte_carlo_frame = data.frame(simulace=monte_carlo)
> class(monte_carlo_frame)
> codefile<-tempfile()
> write.foreign(monte_carlo_frame, "monte_carlo.xls",
+ codefile, package="SPSS")
```

Nyní jsme vyexportovali data monte_carlo_frame do souboru monte_carlo.xls, který je určen pro statistický program SPSS Statistica.

Nezkrácený optimalizovaný zdrojový kód naleznete v Příloha C: Simulace metodou Monte Carlo.

5.4 Pravděpodobnostní rozdělení S pomocí systému Statgraphics

Data vyexportovaná z programu R jsme transformovali do systému Statgraphics, kde vyšetříme shodu s některým z předpokládaných rozdělení pravděpodobnosti.

V našem případě jsme zvolili:

- Posunuté lognormální rozdělení
- Posunuté gamma rozdělení
- Normální rozdělení

5.4.1 Posunuté lognormální rozdělení

Jako první jsou uvedeny výsledky ověření 3-parametrického lognormálního rozdělení jako vhodného modelu škoda na základě výběru pořízeného z 10000 hodnot od 429290 do 598434, získaných simulací Monte Carlo v R.

- Maximálně věrohodné odhady parametrů

Tab. 8: Výsledné parametry posunutého lognormálního rozdělení, zdroj vlastní

<i>Lognormal (3-Parameter)</i>
mean = 500226,
standard deviation = 20635,4
lower threshold = 77515,8

- χ^2 -test dobré shody

Tab. 9: Výsledek testu dobré shody posunutého lognormálního rozdělení, zdroj vlastní

	<i>Lower</i>	<i>Upper</i>	<i>Observed</i>	<i>Expected</i>	
	<i>Limit</i>	<i>Limit</i>	<i>Frequency</i>	<i>Frequency</i>	<i>Chi-Squared</i>
at or below		435000,	3	3,24	0,02
	435000,	450000,	52	47,86	0,36
	450000,	465000,	340	341,72	0,01
	465000,	480000,	1213	1241,15	0,64
	480000,	495000,	2497	2454,18	0,75
	495000,	510000,	2762	2801,57	0,56
	510000,	525000,	1973	1943,56	0,45
	525000,	540000,	845	857,63	0,19
	540000,	555000,	256	250,72	0,11
	555000,	570000,	48	50,37	0,11
above	570000,		11	8,00	1,12

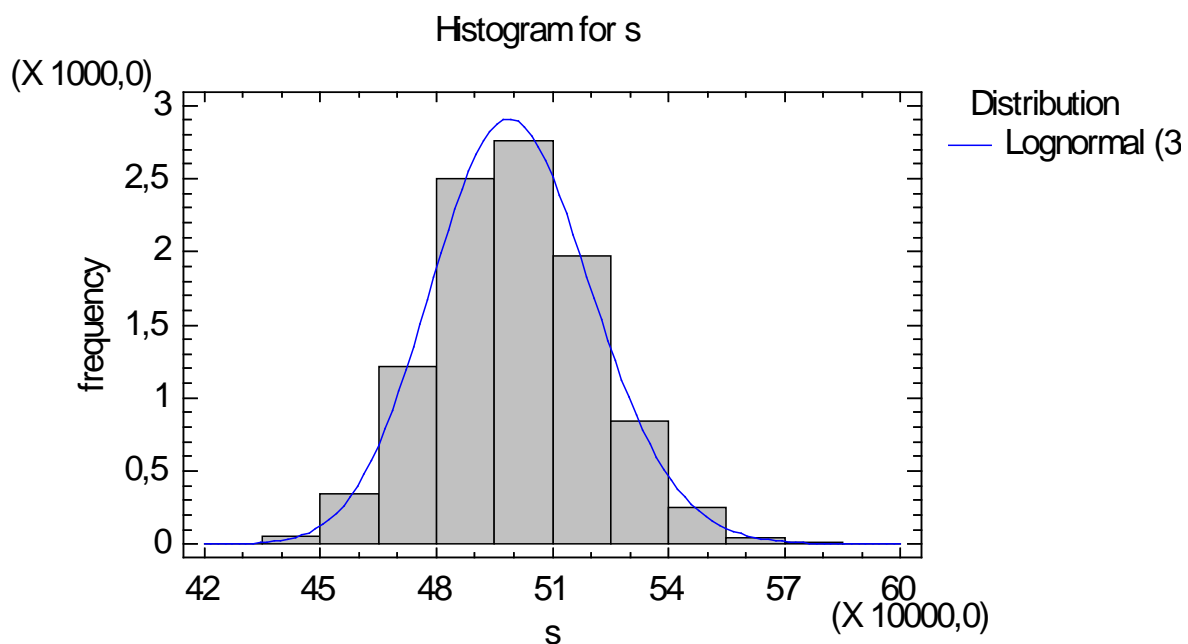
Chi-Squared = 4,30621 with 7 d.f. P-Value = 0,743909

- Kolmogorov-Smirnovův text Test

Tab. 10: Výsledek Kolmogorov-Smirnovova testu posunutým lognormálním rozdělením, zdroj vlastní

	<i>Lognormal (3-Parameter)</i>
DPLUS	0,00513365
DMINUS	0,00551626
DN	0,00551626
P-Value	0,921174

Postup a výsledky testů dobré shody pro 3-parametrické Lognormální rozdělení ukázali, že dobře modeluje data získaná simulací, jak je ostatně vidět i na Obr. 5: Grafické porovnání shody kolektivního rizika s posunutým Lognormálním rozdělením, zdroj vlastní.



Obr. 5: Grafické porovnání shody kolektivního rizika s posunutým Lognormálním rozdělením, zdroj vlastní

5.4.2 Posunuté gamma rozdělení

Jako druhý v pořadí byl proveden test s posunutým gamma rozdělením.

- Maximálně věrohodné odhady parametrů

Tab. 11: Výsledné parametry posunutého gamma rozdělení, zdroj vlastní

<i>Gamma (3-Parameter)</i>
shape = 189,896
scale = 0,000667809
lower threshold = 215869,

- χ^2 -test dobré shody

Tab. 12: Výsledek testu dobré shody posunutého gamma rozdělení, zdroj vlastní

	<i>Lower</i>	<i>Upper</i>	<i>Observed</i>	<i>Expected</i>	
	<i>Limit</i>	<i>Limit</i>	<i>Frequency</i>	<i>Frequency</i>	<i>Chi-Squared</i>
at or below		435000,	3	3,17	0,01
	435000,	450000,	52	47,73	0,38
	450000,	465000,	340	342,22	0,01
	465000,	480000,	1213	1242,07	0,68
	480000,	495000,	2497	2452,93	0,79
	495000,	510000,	2762	2799,90	0,51
	510000,	525000,	1973	1944,49	0,42
	525000,	540000,	845	858,83	0,22
	540000,	555000,	256	250,75	0,11
	555000,	570000,	48	50,07	0,09
above	570000,		11	7,82	1,29

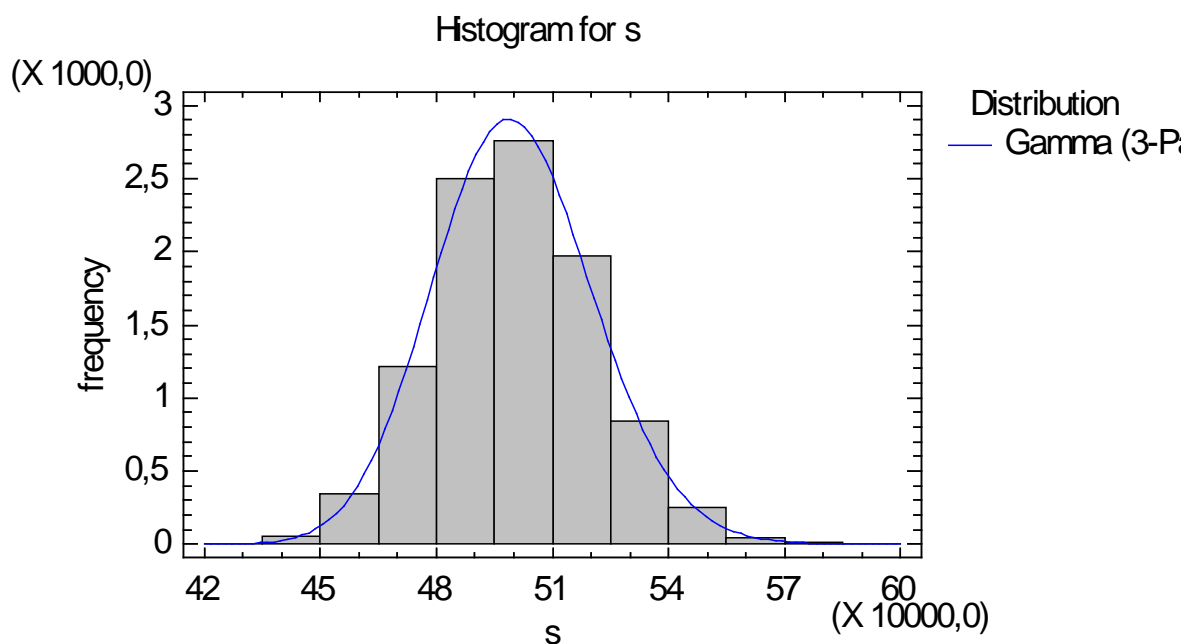
Chi-Squared = 4,51539 with 7 d.f. P-Value = 0,718863

- Kolmogorov-Smirnovův text Test

Tab. 13: Výsledek Kolmogorov-Smirnovova testu posunutého gama rozdělení, zdroj vlastní

	<i>Gamma (3-Parameter)</i>
DPLUS	0,00528812
DMINUS	0,00563245
DN	0,00563245
P-Value	0,908902

Výsledky ukazují, že i posunuté gamma rozdělení lze také přijmout jako vhodný pravděpodobnostní model kolektivního rizika S na základě hodnot získaných simulací.



Obr. 6: Grafické porovnání shody kolektivního rizika s posunutým gamma rozdělením, zdroj vlastní

5.4.3 Normální rozdělení

Normální rozdělení aproximuje zadané hodnoty sic nejhůře, avšak na hladině $\alpha = 0,05$ bychom jej také mohli přijmout jako vyhovující. Z důvodu malé přesnosti jsou výsledky tohoto rozdělení ke shlednutí v Příloha D: Výsledky metody Monte Carlo v programu Statgraphics.

6 Závěr

V úvodní kapitole byly uvedeny základní pojmy problematiky pojišťovnictví, které jsou využity v dalších kapitolách diplomové práce. Byly vysvětleny pojmy rizika, základů pojišťovnictví, pojištění, teorie rizika a managementu rizik, do kterého bychom mohli začlenit praktickou část diplomové práce.

Navazující kapitola se zaměřila na modelování pojistných rizik, jejich obecné charakteristiky, ale hlavně modely počtu a výšky pojistných plnění spolu s ověřením vhodnosti některých rozdělení pravděpodobnosti. V závěru byla pak vyložena teorie aproximace kolektivního rizika. Tato kapitola je důležitá, k pochopení praktické části diplomové práce, ve které se teoreticky popsané metody používají.

V následujících dvou kapitolách je uveden postup využití simulační metody Monte Carlo, která je v diplomové práci použita pro simulaci hodnot kolektivního rizika, spolu s představením programovacího jazyku R, ve kterém byly kroky simulace naprogramovány.

Praktická část diplomové práce je konkrétní ukázkou komplexního řešení problému na reálných datech o počtu a výši pojistných plnění použitím programovacího statistického jazyku. Výsledky simulace hodnot kolektivního rizika S v jazyku R jsou doplněny výpočty ve statistickém programu Statgraphics a tabulkovém procesoru Excel.

Znalost pravděpodobnostního rozdělení kolektivního rizika je pro pojišťovnu mimořádně důležitá jako základ k řešení mnohých závažných problémů, jako jsou stanovení rizikového pojistného, výpočet hodnot v riziku, řešení problému zajištění apod.

Diplomová práce slouží jako ukáзка řešení problému modelování S využitím moderních simulačních technik s podporou nástrojů informačních technologií.

Diplomová práce popisuje také možnosti využití programovacího jazyku R v pojišťovnictví a prakticky prezentuje postupy, kterými se dané problémy, včetně simulace metodou Monte Carlo, dají řešit. Ukazuje také možnosti skloubení programovacího jazyka R s výpočtovými možnostmi programu Statgraphics či tabulkového procesoru Excel, případně porovnává získané výsledky řešení.

7 Seznam použité literatury

- [1.] BEARD, R, Teivo PENTIKÄINEN a E PESONEN. *Risk theory: the stochastic basis of insurance*. 3rd ed. New York: Chapman and Hall, 1984, 408 s. ISBN 04-122-5980-X.
- [2.] BOLAND, Philip J. *Statistical and probabilistic methods in actuarial science*. Boca Raton, FL: Chapman, c2007, 351 s. Interdisciplinary statistics. ISBN 15-848-8695-1.
- [3.] CIPRA, Tomáš. *Finanční a pojistné vzorce*. 1. vyd. Praha: Grada, 2006, 374 s. ISBN 80-247-1633-X.
- [4.] CIPRA, Tomáš. *Kapitálová přiměřenost ve financích a solventnost v pojišťovnictví*. Vyd. 1. Praha: Ekopress, 2002, 271 s. ISBN 80-861-1954-8.
- [5.] CIPRA, Tomáš. *Teorie rizika v pojistné matematice*. Praha: MFF UK a Česká pojišťovna, 1991. ISBN 990002154X.
- [6.] DAŇHEL, Jaroslav. *Kapitoly z pojistné teorie*. Vyd. 1. Praha: Oeconomica, 2002, 139 s. ISBN 80-245-0306-9.
- [7.] DAYKIN, C, Teivo PENTIKÄINEN a M PESONEN. *Practical risk theory for actuaries*. 1st ed. New York: Chapman, 1994, 546 s. ISBN 04-124-2850-4.
- [8.] Definice a vyřešení problému pomocí Řešitele. In: *Nápověda Excel* [online]. [cit. 2012-04-20]. Dostupné z: <http://office.microsoft.com/cs-cz/excel-help/definice-a-vyreseni-problemu-pomoci-resitele-HP005199671.aspx>
- [9.] DORDA, Michal. Úvod do metody Monte Carlo. In: *Vysoká škola báňská - Technická univerzita Ostrava* [online]. 2010 [cit. 2012-04-20]. Dostupné z: http://homel.vsb.cz/~dor028/Aplikace_5.pdf
- [10.] DUCHÁČKOVÁ, Eva. *Principy pojištění a pojišťovnictví: studijní text pro kombinovanou formu studia*. 3. vyd. - přeprac. Praha: Ekopress, c2009, 224 s. ISBN 978-80-86929-51-4.
- [11.] FABIAN, František. *Metoda Monte Carlo a možnosti jejího uplatnění*. 1.vyd. Praha: Prospektrum, 1998, 148 s. ISBN 80-717-5058-1.
- [12.] HENDL, Jan. *Přehled statistických metod: analýza a metaanalýza dat*. 3., přeprac. vyd. Praha: Portál, 2009, 695 s. ISBN 978-80-7367-482-3.
- [13.] HORÁKOVÁ, Galina a Vladimír MUCHA. *Teória rizika v poistení*. Bratislava: EKONÓM, 2006. ISBN 80-225-2141-8.
- [14.] KUBANOVÁ, Jana. *Statistické metody pro ekonomickou a technickou praxi*. třetí. Bratislava: STATIS, 2008. ISBN 978-80-85659-47-4.

- [15.] KUPČÍK, Jakub. *Statistická metoda Monte Carlo*. Zlín, 2009. Dostupné z: <http://dspace.k.utb.cz/handle/10563/10535>. Diplomová práce. Univerzita Tomáše Bati ve Zlíně. Vedoucí práce Ing. Bronislav Chramcov, Ph.D.
- [16.] PACÁKOVÁ, Viera a Veronika BALCÁRKOVÁ. *Možnosti aproximace rozdělení kolektivního rizika: Scientific papers of the University of Pardubice. Series D. Faculty of Economics and Administration*. Pardubice: Univerzita Pardubice, 2010. ISSN 1211–555X. Dostupné z: <http://dspace.upce.cz/handle/10195/38521>
- [17.] PACÁKOVÁ, Viera, Jana KUBANOVÁ, Bohdan LINDA a Erik ŠOLTĚS. *Modelování a simulace pojistných rizik*. první. Pardubice: Univerzita Pardubice, 2012. ISBN 978-80-7395-457-4.
- [18.] PACÁKOVÁ, Viera. *Aplikovaná poistná štatistika*. 3., preprac. a dopl. vyd. Bratislava: Elita, 2004, 248 s. ISBN 80-807-8004-8.
- [19.] Pojistná rizika: Inaugurační přednáška. PACÁKOVÁ, Viera. *Ekonomická univerzita v Bratislavě* [online]. [cit. 2012-04-20]. Dostupné z: http://www.fhi.sk/sk/katedry/ks/nastiahnutie/doc.-rnr.-viera-pacakova_-phd.-inauguracna-prednaska.html
- [20.] RAIS, Karel. *Risk management: studijní text pro kombinovanou formu studia*. Vyd. 1. Brno: Akademické nakladatelství CERM, 2007, 270 s. ISBN 978-80-214-3510-0.
- [21.] ROBERT, Christian P a George CASELLA. *Introducing Monte Carlo methods with R*. New York: Springer, c2010, 283 s. Use R!. ISBN 978-144-1915-764.
- [22.] RUBINSTEIN, Reuven Y. *Simulation and the Monte Carlo method*. New York: Wiley, c1981, 278 s. ISBN 04-710-8917-6.
- [23.] SMEJKAL, Vladimír. *Řízení rizik*. 1. vyd. Praha: Grada, 2003, 270 s. ISBN 80-247-0198-7.
- [24.] ŠVEC, Petr. *Metody kvantifikace rizikového kapitálu v neživotní pojišťovně*. Pardubice, 2009. Dostupné z: <http://hdl.handle.net/10195/34061>. Diplomová práce. Univerzita Pardubice. Vedoucí práce prof. RNDr. Viera Pacáková, Ph.D.
- [25.] VLASATÍKOVÁ, Michala. *Kalkulace pojistného v neživotním pojištění*. Brno, 2006. Dostupné z: <http://is.muni.cz/th/50406/>. Diplomová práce. Masarykova universita. Vedoucí práce RNDr. František Čámský.
- [26.] *Webová stránka programovacího jazyku R* [online]. [cit. 2012-04-20]. Dostupné z: <http://www.r-project.cz/about.html> .

Příloha A: Ověření Poissonova rozdělení

```
#####  
## Ověření Poissonova rozdělení  
## u rozdělení počtu škod  
## s využitím knihoven R  
#####  
## pro práci využíváme následující dodatečné knihovny (packages)  
library(stats)  
library(vcd)  
library(foreign)  
## načtení dat z externího csv souboru  
csvtab.poissonData = read.csv(  
file='c:\\Users\\Berka\\Desktop\\  
+ diplomka-final\\prakticka\\poisson-r.csv',  
+ sep=';',header=F)  
## odhad lambdy z načtených dat  
lambda.ocekavana = vector()  
for(i in 1: length(csvtab.poissonData)) {  
  lambda.ocekavana [i] =  
+  csvtab.poissonData[1,i]*csvtab.poissonData[2,i]  
}  
sum.poissonData = sum(csvtab.poissonData[2,])  
lambda.odhad = sum(lambda.ocekavana)/sum.poissonData  
##  
## funkce pro převod z tabulky četností do čistých dat  
##  
fce = function(x, n) {  
  if(n>0){
```

```

vektor = vector()
  for(i in 1:n){
    vektor[i] = x
  }
  return(vektor)
} else {
  return(F)
}
}

poissonData = list()
for(i in 1:length(csvtab.poissonData)) {
  x = csvtab.poissonData[1,i]
  n = csvtab.poissonData[2,i]
  poissonData[[i]] = fce(x, n)
}

## v listu poissonData jsou vloženy jednotlivé počty,
## potřebujeme je pro jednodušší práci přenést do vektoru
poisson = c(poissonData[[1]],poissonData[[2]],poissonData[[3]],
+ poissonData[[4]],poissonData[[5]],poissonData[[6]],
+ poissonData[[7]])

## vytvoří tabulku četností z čistých dat
tab.os = table(poisson)

tab.os

## odhad parametru lambda z čistých dat
lambda.est = mean(poisson)

lambda.est

## empirické četnosti hodnot
freq.os = vector()

```



```
for(i in 1: length(tab.os)) freq.os[i]<-tab.os[[i]]
freq.os
## očekávané četnosti hodnot
freq.ex = (dpois(0:max(poisson),lambda=lambda.odhad)*80000)
freq.ex
## absolute goodness of fit index
acc = mean(abs(freq.os-trunc(freq.ex)))
## relative (percent) goodness of fit index
acc/mean(freq.os)*100
## Fits a discrete (count data) distribution for goodness-of-fit
## tests.
gof = goodfit(poisson, type = c("poisson"),par = lambda.est)
summary(gof)
plot(gof)
## pro kontrolu data vygenerujeme do externího souboru
poisson_frame = data.frame(simulace=poisson)
codefile<-tempfile()
write.foreign(poisson_frame,"poisson_frame.xls",codefile,
+ package="SPSS")
```

Příloha B: Ověření Paretova rozdělení

```
#####  
## Ověření Paretova rozdělení  
## u rozdělení výšky škod  
#####  
## vložení vektoru výšek škod  
x = c(10,11,15,22,28,30,32,36,38,48,51,55,56,68,68,85,87,94,103,  
+ 104,105,106,109,119,121,137,178,181,226,287,310,321,354,393,  
+ 438,591,1045,1210,1212,2423)  
## základní charakteristiky  
prumer_x = mean(x)  
s = sd(x)  
s_2 = var(x)  
## odhady parametrů metodou momentů  
alfa = (2 * s^2) / (s^2 - prum_x^2)  
alfa_2 = (2 * s_2) / (s_2 - prum_x^2)  
lambda = (alfa - 1) * prum_x  
## intervaly chi-kvadrat testu dobré shody  
## a toho vypočtené reálné četnosti  
interval <- function(x, dolni_mez, horni_mez) {  
  a = 0;  
  for (i in 1:length(x)) {  
    if( (x[i]>dolni_mez) & (x[i]<horni_mez) ) {  
      a = a+1  
    }  
  }  
  return(a)  
}
```

```

Oi = c(interval(x, 0, 42.594),interval(x, 42.594, 102.270),
+ interval(x, 102.270, 196.444),interval(x, 196.444, 322.336),
+ interval(x, 322.336, max(x)+1))

## horní hranice
HH = c(42.594,102.27,196.444,322.336)
F_HH <- function(meze, lambda, alfa) {
  HH_vektor = vector()
  for (i in 1:length(meze)) {
    HH_vektor[i] = 1 - ( lambda / (lambda + meze[i]) )^alfa
  }
  return(HH_vektor)
}

F_HH_cisla = F_HH(HH, lambda, alfa)

## pravděpodobnosti výskytu
p_i = function(F_HH) {
  suma = 0
  vektor = vector()
  for (i in 1:length(F_HH)) {
    vektor[i] = F_HH[i] - suma
    suma = suma + vektor[i]
  }
  vektor[5] = 1 - suma
  return(vektor)
}

p_i_cisla = p_i(F_HH_cisla)

## očekávané četnosti
E_i = function(p_i, sum_x) {
  E_i_x = vector()

```

```

x_i = 0
for (i in 1:length(p_i)) {
  p_i[i] = p_i[i] * 40
  E_i_x = as.integer(p_i[i])
  x_i = p_i[i] - E_i_x
  if (x_i > 0.5) {
    p_i[i] = E_i_x + 1
  } else {
    p_i[i] = E_i_x
  }
}
return(p_i)
}
E_i_cisla = E_i(p_i_cisla, sum(x))
## výpočet chí
chi_square = function(Oi, Ei) {
  chi_sq = vector()
  for(i in 1:length(Oi)) {
    chi_sq = (Oi - Ei)^2 / Ei
  }
  return(chi_sq)
}
chi = chi_square(Oi, E_i_cisla)
sum(chi)

```

Příloha C: Simulace metodou Monte Carlo

```
#####  
## Simulace pomocí metody  
## Monte Carlo  
#####  
## načtení knihovny, kterou budeme  
## používat pro generování náhodných  
## čísel z paretova rozdělení  
library(actuar)  
library(foreign)  
## funkce metody monte carlo  
x = function(n, pois_lambda, pareto_alfa, pareto_lambda) {  
  m = list()  
  m_sum = array(0, dim = n)  
  x_pois = rpois(n, pois_lambda)  
  for (i in 1:length(x_pois)) {  
    m[[i]] = rpareto(x_pois[i], pareto_alfa, pareto_lambda)  
    m_sum[i] = sum(m[[i]])  
  }  
  return(m_sum)  
}  
## voláme funkci pro metodu MC s parametry, které jsme vypočetli  
## dříve  
monte_carlo = x(10000, 1992, 3.4763, 622.159)  
## histogram metody  
hist(monte_carlo)  
## pro export dat musíme data upravit z vektoru na frame typ  
monte_carlo_frame = data.frame(simulace=monte_carlo)
```

```
class(monte_carlo_frame)
## exportujeme data do zadaného souboru
codefile<-tempfile()
write.foreign(monte_carlo_frame,"monte_carlo.xls",codefile,
+ package="SPSS")
```

Příloha D: Výsledky metody Monte Carlo v programu Statgraphics

Uncensored Data - s

Data variable: s

10000 values ranging from 429290, to 598434,

Fitted Distributions

<i>Normal</i>
mean = 500226,
standard deviation = 20637,5

Goodness-of-Fit Tests for s

Kolmogorov-Smirnov Test

	<i>Normal</i>
DPLUS	0,0128603
DMINUS	0,00719794
DN	0,0128603
P-Value	0,0732006

