



**MATEMATICKO-FYZIKÁLNÍ  
FAKULTA**  
Univerzita Karlova

**BAKALÁŘSKÁ PRÁCE**

Karla Váchová

**Operační riziko a značkový  
Poissonův proces**

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: doc. RNDr. Michal Pešta, Ph.D.

Studijní program: Finanční matematika

Praha 2022

Prohlašuji, že jsem tuto bakalářskou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů. Tato práce nebyla využita k získání jiného nebo stejného titulu.

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 11. května 2022

Karla Váchová

Tuto formou bych ráda na tomto místě poděkovala vedoucímu bakalářské práce doc. RNDr. Michalovi Peštovi, Ph.D. za věnovaný čas, cenné rady i za návrh tématu v oblasti, která mě zajímá, a námět k tvorbě praktické části.

Název práce: Operační riziko a značkový Poissonův proces

Autor: Karla Váchová

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: doc. RNDr. Michal Pešta, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: Předmětem této bakalářské práce s názvem „Operační riziko a značkový Poissonův proces“ je modelování operačního rizika pomocí Poissonova značkového procesu. Poissonův proces je typ bodového procesu modelující náhodně rozmístěné body v nějakém nosném prostoru. Díky jeho matematickým vlastnostem je poměrně často užívaným modelem například v biologii, astronomii, ekologii nebo ekonomii. Tato bakalářská práce popisuje jeho základní vlastnosti a využívá značkového Poissonova procesu k modelování výší a počtu škod spadajících pod operační riziko banky.

Klíčová slova: operační riziko, Poissonův proces, značkový proces

Title: Operational risk and marked Poisson process

Author: Karla Váchová

Department: Department of Probability and Mathematical Statistics

Supervisor: doc. RNDr. Michal Pešta, Ph.D., Department of Probability and Mathematical Statistics

Abstract: The subject of this bachelor thesis entitled „Operational risk and marked Poisson process“ is the modelling of operational risk using marked Poisson process. The Poisson process is a type of a point process that models randomly distributed points on some underlying space. Because of its mathematical properties, it is a quite frequently used model in biology, astronomy, ecology or economics, for example. This bachelor thesis describes its basic properties and uses the marked Poisson process to model loss frequency and severity belonging to bank's operational risk.

Keywords: operational risk, Poisson process, marked process

# Obsah

Úvod	2
<b>1 Operační riziko</b>	<b>3</b>
1.1 Definice . . . . .	3
1.2 Přístupy k modelování . . . . .	3
1.3 Parametrický přístup . . . . .	5
1.3.1 Rozdělení četnosti škod . . . . .	5
1.3.2 Rozdělení výší škod . . . . .	6
1.3.3 Celkový model . . . . .	6
1.3.4 Vhodnost modelu . . . . .	7
1.3.5 Posouzení přesnosti modelu . . . . .	7
<b>2 Poissonův proces</b>	<b>8</b>
2.1 Základní pojmy . . . . .	8
2.2 Neznačkový Poissonův proces . . . . .	9
2.2.1 Definice . . . . .	10
2.2.2 Vlastnosti . . . . .	11
2.3 Značkový Poissonův proces . . . . .	12
2.3.1 Vlastnosti . . . . .	12
2.3.2 Typy modelů . . . . .	13
2.3.3 Definice . . . . .	13
2.4 Poznámky ke značení . . . . .	16
<b>3 Aplikace</b>	<b>17</b>
3.1 Popis dat . . . . .	17
3.2 Rozdělení počtu škod . . . . .	19
3.3 Rozdělení výší škod . . . . .	20
3.4 Simulace . . . . .	23
<b>Závěr</b>	<b>25</b>
<b>Seznam použité literatury</b>	<b>27</b>

# Úvod

K pochopení jevu nebo popisu neznámého systému lze využít modelování pomocí stochastických (náhodných) procesů. Tyto teoretické modely nám dávají přibližnou představu o tom, jak by se daný jev mohl chovat. Vytvoření takového modelu vyžaduje pochopení jak oblasti, ze které data pocházejí, tak i principu stochastických procesů.

Poissonův proces je náhodný proces, kterým lze popsat výskyt náhodných bodů v dané množině. Tento náhodný proces je poměrně často užívaným modelem díky jeho vlastnostem a limitnímu chování. Jednodimenzionální Poissonův proces, kterým se budeme v této práci povětšinou zabývat, může být modelem pro nastání náhodných událostí v daném časovém intervalu. V této práci využijeme Poissonův značkový proces k modelování výší a počtů škod, které spadají pod operační riziko banky.

Schopnost účinně měřit a řídit rizika je velmi důležitou součástí činnosti každé finanční instituce. Cílem této práce je nalezení vhodného modelu pro výši celkové ztráty plynoucí z operačního rizika na základě sepsaného aparátu s následnou aplikací na simulovaná data. Model zmíněné celkové ztráty vychází z modelů pro frekvenci a severitu individuálních ztrát. Práce je rozdělena do tří kapitol.

V první kapitole je čtenář krátce seznámen s pojmem operačního rizika, které je následně uvedeno do souvislosti s procesem modelování. Možné ztráty plynoucí z tohoto rizika jsou totiž ovlivněny nastáním náhodných událostí. Banka proto nemůže přesně určit, jak velká ztráta skutečně nastane, ale pokud se zaměří na sestavení vhodného modelu, může ztráty alespoň přibližně predikovat. V této části je motivována volba Poissonova rozdělení pro modelování počtu ztrát. Dále jsou zmíněny nejběžnější přístupy pro sestavení vhodného modelu, které banky mohou využít, a jejich teoretické porovnání, jakož i problémy, jež při modelování rizika mohou nastat.

Ve druhé kapitole je zaveden Poissonův proces jako druh čítacího procesu. Jsou také popsány některé jeho důležité charakteristiky. Definice a věty jsou též doplněny o komentář napomáhající pochopení v kontextu modelování frekvence nastání náhodných událostí v čase. Díky tomu si čtenář osvojí znalosti potřebné k porozumění závěrečné části této práce.

Třetí kapitola aplikuje ekonomický i matematický aparát z předchozích částí na datový soubor obsahující ztráty plynoucí z operačního rizika, k čemuž je využito Poissonova procesu pro modelování počtu škod. Pro modelování jejich výše se využije rozšíření Poissonova procesu o značkování. Data jsou obsažena v balíčku `OpVaR` dostupném ve statistickém softwaru R, který bude využit také ke zpracování a grafické reprezentaci výsledků i dílčích kroků. Pro datový soubor bude vybráno rozdělení na základě jeho charakteristik a posouzena vhodnost této volby. Závěrečná část obsahuje predikci maximální možné ztráty pro další období a porovnání teoretického odhadu se skutečnou výší ztráty.

# 1. Operační riziko

Detaily k teorii shrnuté v této kapitole týkající se risk managementu popisuje například Wernz (2014). Obecné normy v této oblasti vychází ze standardů Basel (současně Basel III), což jsou dokumenty představující doporučení pro bankovní právo a regulaci vydávané Basilejským výborem pro bankovní dohled.

## 1.1 Definice

Operační riziko patří mezi základní typy rizik v bankovním sektoru. Lze ho definovat jako riziko ztráty plynoucí z nedostatků či selhání vnitřních procesů, osob a systémů nebo z vnějších událostí. Při modelování frekvencí operačního rizika je Poissonův proces poměrně často užívanou metodou.

Na analýzu operačního rizika a jeho řízení se subjekty začaly více zaměřovat zejména po jeho implementaci do konceptu výpočtu kapitálového požadavku dle Basel II, tedy po roce 2006.

Samotný pojem je poměrně těžko definovatelný, protože se tento typ rizika vyskytuje téměř ve všech činnostech banky a je provázaný i s jinými typy rizik. Proto je také relativně těžké určit ztráty, které z operačního rizika plynou.

Basilejský výbor pro bankovní dohled definuje operační riziko následovně:

*Operační riziko je riziko ztráty, které vyplývá z nedostatků či selhání vnitřních procesů, osob a systémů nebo z vnějších událostí.*

*Tato definice zahrnuje právní riziko, ale vylučuje riziko reputační a strategické.*

Údaje o událostech vztahujících se k operačnímu riziku jsou ve standardech Basel rozděleny na 7 typů událostí a 8 linií podnikání. V této práci ale budeme pracovat s daty jako jedním celkem a nebudeme tyto kategorie rozlišovat.

Pod operační riziko spadá například vnitřní a vnější nekalé jednání, selhání IT systémů nebo lidského faktoru či ztráty způsobené škodou na fyzickém majetku.

## 1.2 Přístupy k modelování

Aby byla banka schopna na zmíněné situace nastávající náhodně reagovat, je vhodné riziko modelovat a následně predikovat ztráty. Z důvodu rychlého vývoje finančních technologií i zvyšující se komplexity operací naráží instituce na nové typy rizikových událostí, proto je také důležité modely neustále přizpůsobovat.

Pro výpočet kapitálového požadavku byly v Basel II navrženy tři přístupy:

- BIA (*Basic Indicator Approach* – přístup základního ukazatele),
- SA (*Standardized Approach* – standardizovaný přístup),
- AMA (*Advanced Measurement Approach* – pokročilý přístup).

První dva výše zmíněné přístupy využívají indikátor na bázi podílu z hrubých výnosů instituce. Nejsou tedy založené na výších škod plynoucích z rizikových událostí jako takových, v této práci se jimi proto dále zabývat nebudeme.

Přístupy BIA a SA zohledňují některé charakteristiky instituce, díky AMA lze ale uplatnit i vlastní „know-how“ a zvolit metodu modelování. Tato samostatnost je výhodou, jelikož potenciálně umožňuje dosáhnout i vyšší citlivosti na riziko. Složitost výsledných AMA modelů a obtížná srovnatelnost zapříčiněná širokou škálou různých interních modelů však přináší v některých směrech komplikace. Naopak nemodelové metody dle některých dostatečně nezohledňují komplexitu problému a charakteristiky instituce. Proto Basel III definuje SMA (*Standardised Measurement Approach*), která má kombinovat jednoduchost SA a citlivost AMA.

Nejčastější metodou v rámci AMA je LDA (*Loss Distribution Approach*). Princip spočívá v modelování dvou složek – frekvence a severity. Předpokladem aktuárského přístupu je, že ztráty jsou iid veličiny a že rozdělení frekvence nezávisí na rozdělení severity. Aplikace této metody umožňuje také odhadnout maximální možnou ztrátu na hladině spolehlivosti  $\alpha$  pro dané období, tj. hodnotu v riziku VaR (*Value at Risk*), což je tedy taková hodnota, která nebude s předepsanou (vysokou) pravděpodobností v požadovaném časovém horizontu překročena. Ze statistického pohledu jde o  $\alpha$ -kvantil rozdělení ztrát. Označíme-li tedy  $X$  ztrátu, definujeme  $P[X \leq VaR] = \alpha$ , z čehož odvodíme

$$VaR_\alpha = \inf_{x \in \mathbb{R}} \{P[X > x] \leq 1 - \alpha\}.$$

Tuto metodu i její modifikace detailněji popisuje například Cipra (2015).

Úkolem je nalézt vyhovující rozdělení, případně jejich vhodnou kombinaci, popisující počet a výši ztrát z nastalých událostí. Používají se 3 techniky:

- modelování pomocí parametrického rozdělení,
- teorie extrémních hodnot (EVT – *Extreme Value Theory*),
- neparametrický odhad na základě empirického rozdělení škod.

První přístup předpokládá, že data pocházejí z konkrétního parametrického rozdělení. Parametry lze odhadnout například metodou maximální věrohodnosti, momentovou metodou atd. Nelze však otestovat všechna možná rozdělení a vybrat to nejvhodnější, proto se často využívá zobecněných rozdělení s mnoha parametry (typicky 4 a více). Parametry popisují třeba charakteristiky polohy a variability, šikmost a špičatost. Jednodušší rozdělení získáme odvozením ze zobecněného.

Rozdělení ztrát lze rozložit na tělo (hodnoty pod určitým prahem) a chvost. Při modelování operačního rizika nás zajímá především chvost, protože extrémní ztráty (přestože nastávají méně často) mají na kapitál obvykle dominantní vliv. Přístup EVT se zaměřuje na popis extrémních hodnot, tj. událostí s nízkou četností a vysokou závažností, modelovaných právě chvostem rozdělení. K tomu využívá jednu ze dvou metod. Metoda blokového maxima se zakládá na rozčlenění časového intervalu na stejně dlouhé disjunktní intervaly (bloky). V každém z nich se pak zkoumá nejvyšší hodnota (blokové maximum), která se použije k sestavení modelu. Nevýhoda je, že se velká část dat nevyužije. Metoda prahové hodnoty naopak zkoumá pouze takové události, které překročily určitou mez. Používá se hlavně v situacích, kdy nepozorujeme výraznou sezónnost dat.

Neparametrický odhad je založen na simulaci z historických dat. Nevýhoda je, že odhady mohou být vychýlené. Tato metoda může riziko podhodnocovat, jelikož předpokládá, že budoucí ztráta nebude vyšší než dosud nejvyšší zaznamenaná. Naopak riziko bude nadhodnocené, pokud je v zaznamenaných datech neobvykle vysoká ztráta.



## 1.3 Parametrický přístup

V této části popíšeme postup volby vhodného rozdělení k modelování četnosti a výše ztrát v parametrickém přístupu. Předpokladem je, že náhodné události jsou nezávislé, stejně rozdělené s předem daným rozdělením s neznámými parametry.

Ještě před samotným modelováním je dobré si uvědomit strukturu dat. Dále se budeme zabývat modely využívajícími rozdělení s těžkými chvosty, jelikož mnoho událostí souvisejících s operačním rizikem nastává poměrně často, ale závažnost způsobených ztrát je nízká (HFLI – *High Frequency Low Impact*). Tyto události nejsou pro banku tak podstatné, jelikož existují způsoby, jak dopady zmírnit nebo jim předcházet. Jiným typem jsou události mající extrémní dopad, které by mohly vést až ke kolapsu instituce. Ztráty z nich plynoucí se naopak vyskytují s malou četností, ale jsou velice vysoké (LFHI – *Low Frequency High Impact*). Právě tyto ztráty, modelované chvostem nějakého rozdělení, jsou pro banku důležité.

### 1.3.1 Rozdělení četnosti škod

Frekvence nebo také četnost ztrát vyjadřuje jejich počet za dané období. Podle studie BCBS (2009) nejvíce bank při modelování frekvence ztrát využívá Poissonovo rozdělení, celkem 93 % dotázaných AMA bank (viz tabulka 1.1).

Rozdělení	Zastoupení
Poissonovo rozdělení	93 %
Negativně binomické rozdělení	19 %
Ostatní	7 %

*Poznámka:* Banky mohly zvolit více než jednu odpověď.

Tabulka 1.1: Procentuální zastoupení rozdělení využitých k modelování frekvencí.

V této práci budeme při popisu počtu ztrát uvažovat Poissonovo rozdělení  $Pois(\lambda)$ . Použití Poissonova rozdělení je přirozené z následujícího důvodu.

Rozdělme interval  $(0, t]$  na  $n$  krátkých časových intervalů a předpokládejme, že ve všech je stejně pravděpodobný výskyt jedné události a pravděpodobnost výskytu více událostí v jednom intervalu je zanedbatelná. Necht' je očekávaný počet událostí  $\lambda t$ . V každém intervalu tedy musí být pravděpodobnost výskytu jedné události rovna  $\lambda t/n$  a počet událostí v intervalu má binomické rozdělení  $Bi(n, \lambda t/n)$ . Pro  $n \rightarrow \infty$  se zvyšuje počet intervalů a zkracuje se jejich délka. Rozdělení počtu událostí se přibližuje Poissonovu rozdělení s parametrem  $\lambda t$ .

Poissonovo rozdělení je vhodné i díky aditivitě. Výsledkem by měl být model pro počet ročních ztrát. V praxi ale nelze provést roční součty ztrát, jelikož banka obvykle nemá dostatečnou historii dat (buď vede záznam kratší dobu nebo tak velké množství dat z praktických důvodů neuchovává). Díky aditivitě je možné z odhadu parametru měsíčních ztrát získat odhad parametru pro roční ztráty. Další důvod k volbě Poissonova procesu je potřeba odhadu jediného parametru. Negativně binomické rozdělení, podobné Poissonovu, sice umožňuje modelovat rozdílnou střední hodnotu a rozptyl, ale bylo by nutné odhadovat dva parametry, což vyžaduje i větší množství dat. Předpoklad Poissonova rozdělení pro frekvenci událostí také říká, že nezáleží na čase uplynulém od poslední události.

### 1.3.2 Rozdělení výší škod

Severita je výše neboli závažnost jednotlivých ztrát. K modelování severit lze také zvolit libovolné vhodné rozdělení nebo jejich kombinaci, preference už ale nejsou tak jednoznačné. Nejčastěji bylo podle dříve zmíněné studie BCBS (2009) používáno log-normální rozdělení (viz tabulka 1.2).

Použité rozdělení	Zastoupení
Log-normální rozdělení	33 %
Weibullovo rozdělení	17 %
Empirická rozdělení	14 %
Gamma rozdělení	7 %
Směs log-normálních rozdělení	7 %
Směs log-normálního a gamma rozdělení	5 %
Zobecněné Paretovo rozdělení	5 %
Ostatní	16 %

*Poznámka:* Banky mohly zvolit více než jednu odpověď.

Tabulka 1.2: Procentuální zastoupení rozdělení využitých k modelování severit.

Z důvodů vysvětlených dříve v této kapitole se obvykle volí takové rozdělení, které má poměrně vysokou špičatost, znatelné zešikmením vpravo a těžký pravý chvost tvořený odlehlými pozorováními (Rippel a Teplý (2008)). Dále také ztráty nemohou nabývat záporných hodnot (záleží na zvoleném značení, každopádně nemění znaménko), proto je též přirozené uvažovat pouze ta rozdělení, která mají nezáporný nosič. Množství parametrů zvoleného rozdělení ovlivňuje následnou flexibilitu modelu, zároveň se ale s vyšším počtem parametrů zvyšuje složitost odhadů i modelu, a tedy také následná aplikace a interpretovatelnost. Některé studie (např. Dutta a Perry (2006)) popsaly poměrně vhodné vlastnosti a chování zobecněných rozdělení se čtyřmi parametry (zobecněné beta rozdělení druhého typu GB2 nebo g-and-h rozdělení).

### 1.3.3 Celkový model

V rámci modelování metodou LDA je pak třeba rozdělení frekvencí a severit zkombinovat, abychom získali celkovou akumulovanou ztrátu. K tomu využijeme tzv. složené rozdělení, což je rozdělení náhodné veličiny  $S(t) = \sum_{i=1}^{N(t)} m_i$ , kde  $N(t)$  značí počet událostí nastalých za časové období  $t$  a  $m_i, i = 1, \dots, N(t)$  jsou výše příslušných škod, které považujeme za nezávislé (vzájemně i na  $N$ ) stejně rozdělené. Veličina  $S(t)$  pak značí celkovou ztrátu pro dané období  $t$ . Kapitálový požadavek v kontextu operačního rizika je definován jako 99,9% hodnota v riziku (VaR), což je odpovídající kvantil rozdělení pro příští období (na hladině 0,999).

### 1.3.4 Vhodnost modelu

Test zvoleného modelu je důležitou součástí risk managementu. Nevhodně zvolený model, který by riziko podhodnocoval, by z dlouhodobého hlediska mohl způsobit, že banka nebude mít dostatečnou kapitálovou rezervu ke krytí ztrát. Naopak pokud by model riziko nadhodnocoval, přijde instituce o disponibilní prostředky, které by jinak mohla investovat.

Vizuálně lze porovnat empirické rozdělení reálných dat a zvolené teoretické rozdělení (označme  $F$ ) tak, že je zobrazíme do jednoho grafu. K tomu lze využít kvantilový diagram ( $Q-Q$  plot), který porovnává kvantily příslušných rozdělení. Přestože takové vizuální porovnání sice není formálním testem, poskytuje dobrou představu o struktuře – symetrie, odlehle hodnoty atd. Navíc, pokud máme menší počet dat, lze posoudit přesnost modelu spíše na základě kvantilového diagramu než formálního testu. To se hodí například pro modelování chvostů.

K formálnímu posouzení slouží statistické testy. Testujeme nulovou hypotézu

$$H_0 : \text{data mají rozdělení } F$$

proti alternativě

$$H_1 : \text{data nemají rozdělení } F.$$

Lze použít různé statistické testy, například jde o Kolmogorov-Smirnovův test, Cramér von-Mises test nebo Anderson-Darlingův test, který je nejvhodnější, jelikož klade větší důraz na chvosty rozdělení, což je z dříve zmíněných důvodů pro testování ztrát z operačního rizika žádoucí (Dutta a Perry (2006)). Tyto testy ve své standardní podobě předpokládají, že žádný z parametrů nebyl odhadován ze stejných dat, ze kterých počítáme testovou statistiku. Jelikož v našem případě bude třeba parametry odhadovat, je třeba vzít tento fakt v potaz.

Kromě tohoto statistického posouzení přesnosti by nás dále mohlo zajímat, zda je výsledný odhad kapitálového požadavku realistický, jestli jsou charakteristiky zaznamenaných a odhadnutých dat logicky konzistentní.

V souvislosti s modelem můžeme také zkoumat jeho flexibilitu a jednoduchost, přičemž tyto charakteristiky jdou proti sobě – čím více parametrů model má, tím by potenciálně mohl být přesnější (při vhodně zvoleném rozdělení), ale také komplexnější a tím pádem i složitější.

### 1.3.5 Posouzení přesnosti modelu

Přesnost výsledného modelu je možné posoudit pomocí tzv. backtestingu. Prakticky se jedná o porovnání empirického odhadu (kvantilu) VaR s reálnou, skutečně pozorovanou, ztrátou  $S(t)$  v nějakém časovém intervalu  $t$ . Bude nás pak zajímat, kolikrát se stane, že bude skutečná ztráta vyšší než ztráta odhadnutá ze simulovaných průběhů. Definujme tedy indikátorovou funkci

$$I_\alpha(t) := \begin{cases} 1 & S(t) \leq \widehat{VaR}_\alpha, \\ 0 & \text{jinak.} \end{cases}$$

Pokud je model vhodně zvolený, měl by přibližně splňovat  $P[I_\alpha(t) = 1] = \alpha$ . Pokud by  $P[I_\alpha(t) = 1] > \alpha$ , je riziko podhodnocené. Naopak model, pro který vyjde, že  $P[I_\alpha(t) = 1] < \alpha$ , riziko nadhodnocuje.

Formulace statistických testů podrobněji popisuje například Campbell (2005). Můžeme zmínit třeba Kupiecův test z roku 1995 založený na věrohodnostním poměru, který je známý také jako POF test (*Proportion of Failure*).

## 2. Poissonův proces

Definice a značení v této kapitole vychází především z knihy Ross (1996).

### 2.1 Základní pojmy

Při snaze popsat nějaký neznámý systém nebo jev je často vhodné dívat se na soubor příslušných zkoumaných náhodných veličin jako na náhodný proces, který je definovaný následovně.

**Definice 1.** *Nechť  $T \subseteq \mathbb{R}$  je neprázdná množina. Systém náhodných veličin  $\{X(t), t \in T\}$  definovaných na témže pravděpodobnostním prostoru nazýváme náhodným procesem.*

Pokud nás při popisu systému zajímá počet výskytů nějakých bodů v určité množině, využijeme k tomu čítací proces. Čítací proces je takový náhodný proces, který má zprava spojitě neklesající trajektorie, nabývá pouze celočíselných hodnot a startuje z nezáporné hodnoty.

*Poznámka.* Trajektorie je realizace náhodného procesu, tedy nenáhodná funkce.

V kontextu této práce budou zkoumanými body nastalé události pozorované na množině, kterou budeme chápat jako časový interval. Čítací proces nejčastěji popisuje právě počet výskytů událostí v čase, tj. na  $[0, \infty)$ . Proces ale má širší využití a lze díky němu popsat například výskyt objektů v časoprostoru. Používá se třeba k popisu populace na geografickém území, výskytu rostlin na nějaké ploše nebo k modelování výskytu přírodních katastrof. Dále se v této kapitole budeme zabývat jednodimenzionálními procesy, teorii lze ale rozšířit do více dimenzí.

Nyní už přejdeme k formální definici čítacího procesu.

**Definice 2.** *Řekneme, že náhodný proces  $\{N(t), t \geq 0\}$  je čítací proces, pokud splňuje následující podmínky:*

1.  $N(t) \geq 0$ ,
2.  $N(t)$  nabývá pouze celočíselných hodnot,
3. pro  $s < t$  platí  $N(s) \leq N(t)$ ,
4. pro  $s < t$  je rozdíl  $N(t) - N(s)$  roven počtu bodů v intervalu  $(s, t]$ .

Někdy bývá čítací proces definovaný následovně, pomocí indikátorové funkce. Takto definovaný čítací proces je ekvivalentní s procesem zavedeným v definici 2.

**Věta 1.** *Řekneme, že náhodný proces  $\{N(t), t \geq 0\}$  je čítací proces, pokud existuje neklesající posloupnost nezávislých náhodných veličin  $\xi_i, i \in \mathbb{N}$  takových, že*

$$N(t) = \sum_{i=1}^{\infty} I\{\xi_i \leq t\}.$$

Veličiny  $\xi_i, i \in \mathbb{N}$  v této definici lze chápat jako časy výskytů jednotlivých událostí a  $N(t)$  jako počet událostí, které nastaly do času  $t$ .

Čítací proces si lze představit jako proces, který přičte 1 v každém bodě nastání události. Na obrázku 2.1 je zobrazena trajektorie jedné realizace čítacího procesu, která je typická právě těmito „skoky“. Konkrétně se v tomto případě jedná o Poissonův proces, kterým se budeme zabývat později v této kapitole.

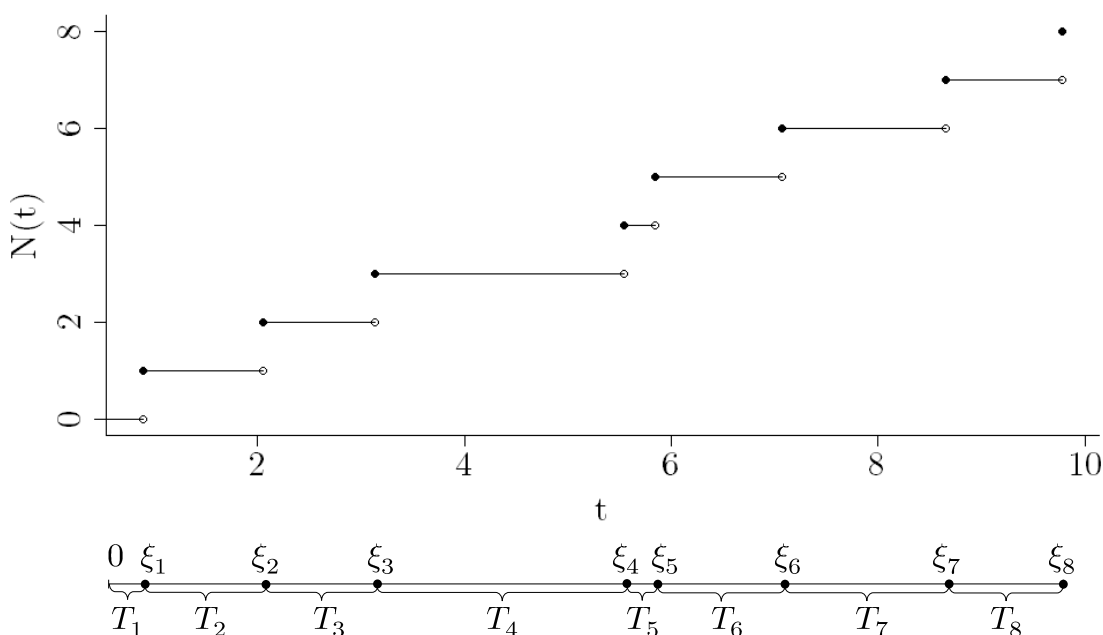
K zavedení Poissonova procesu je třeba definovat ještě dvě vlastnosti přírůstků náhodného procesu. První z těchto vlastností je nezávislost.

**Definice 3.** Řekneme, že náhodný proces (se spojitým časem)  $\{X(t), t \geq 0\}$  má nezávislé přírůstky, jsou-li přírůstky  $X(t_2) - X(t_1), \dots, X(t_n) - X(t_{n-1})$  procesu  $\{X(t), t \geq 0\}$  nezávislé náhodné veličiny pro všechna  $0 \leq t_1 < t_2 < \dots < t_n$ .

Pro čítací proces je počet událostí v intervalu  $(t_{i-1}, t_i]$  roven  $N(t_i) - N(t_{i-1})$ . Čítací proces má tedy nezávislé přírůstky, pokud jsou nezávislé počty událostí nastalých v disjunktních časových intervalech  $(t_1, t_2], \dots, (t_{n-1}, t_n]$ . To lze využít při modelování vzájemně nezávislých událostí nastávajících náhodně v čase.

Druhou důležitou vlastností přírůstků je stacionarita, která říká, že počet událostí v libovolném časovém intervalu závisí pouze na jeho délce.

**Definice 4.** Řekneme, že náhodný proces se spojitým časem  $\{X(t), t \geq 0\}$  má stacionární přírůstky, pokud mají přírůstky  $X(t_2 + s) - X(t_1 + s)$  stejné rozdělení jako  $X(t_2) - X(t_1)$  pro libovolná  $0 \leq t_1 < t_2$  a  $s > 0$ .



Obrázek 2.1: Realizace a příslušná trajektorie čítacího (Poissonova) procesu.

## 2.2 Neznačkovaný Poissonův proces

Pro úplnost ještě připomeňme pravděpodobnostní funkci Poissonova rozdělení s parametrem  $\lambda > 0$ , dále  $\text{Pois}(\lambda)$ , tvaru

$$P[N = n] = \frac{\lambda^n}{n!} e^{-\lambda}, \quad n = 0, 1, 2, \dots,$$

příslušnou střední hodnotu  $E N = \lambda$  a rozptyl  $\text{var } N = \lambda$ .

## 2.2.1 Definice

Poissonův proces je často interpretován jako náhodný bodový proces na reálné přímce (obvykle jde o polopřímku zachycující diskrétní množinu stavů se spojitým časem na kladné poloose, tj.  $t \geq 0$ ).

Dále v textu uvádíme věty, které jsou v některé literatuře formulovány jako definice Poissonova procesu. Tyto formulace jsou ale ekvivalentní s definicí 1, důkaz uvádí například Ross (1996).

**Definice 5.** Čítací proces  $\{N(t), t \geq 0\}$  se nazývá (homogenní) Poissonův proces s intenzitou  $\lambda > 0$ , pokud splňuje:

1.  $N(0) = 0$ ,
2. proces má nezávislé přírůstky,
3. přírůstky  $N(t+h) - N(t)$  mají Poissonovo rozdělení s parametrem  $\lambda h$ .

Bod 3 znamená, že počet událostí v libovolném časovém intervalu délky  $h$  má Poissonovo rozdělení se střední hodnotou  $\lambda h$ .

V souladu s definicí čítacího procesu je tedy parametr rozdělení úměrný délce časového intervalu s konstantou úměrnosti  $\lambda$ , která představuje střední hodnotu počtu událostí za jednotkový časový interval. Pro všechna  $h, t \geq 0$  pak platí

$$P[N(t+h) - N(t) = n] = \frac{(\lambda h)^n}{n!} e^{-\lambda h}, \quad n \in \mathbb{N}_0.$$

Dále také platí, že  $E[N(t)] = \lambda t$ , jelikož hodnotu v čase  $t$  lze díky vlastnosti 1 vyjádřit jako  $N(t) = N(t) - N(0) + N(0) = N(t) - N(0)$ , což je náhodná veličina s Poissonovým rozdělením s parametrem  $\lambda t$  (opět dle bodu 3 předchozí definice). Z této úvahy tedy dostáváme

$$P[N(t) = n] = e^{-\lambda t} \frac{(\lambda t)^n}{n!}, \quad n \in \mathbb{N}_0.$$

V situaci, ve které chceme rozhodnout, zda se skutečně jedná o Poissonův proces, je výhodné využít následující větu, která se často používá jako ekvivalentní definice Poissonova procesu.

**Věta 2.** Čítací proces  $\{N(t), t \geq 0\}$  se nazývá (homogenní) Poissonův proces s intenzitou  $\lambda > 0$ , pokud splňuje:

1.  $N(0) = 0$ ,
2. proces má stacionární a nezávislé přírůstky,
3.  $P[N(h) = 1] = \lambda h + o(h)$ ,
4.  $P[N(h) \geq 2] = o(h)$ ,

kde  $o(h)$  značí libovolnou funkci takovou, že  $\lim_{h \rightarrow 0^+} \frac{o(h)}{h} = 0$ .

Bod 3 definice 5 by totiž mohlo být v praxi obtížné ověřit. Ve větě 2 je tento bod přeformulovaný tak, že pravděpodobnost výskytu jediné události v krátkém časovém intervalu je úměrná délce tohoto intervalu s konstantou úměrnosti rovnou intenzitě procesu, tedy  $\lambda$ . Zároveň musí pravděpodobnost výskytu dvou nebo více událostí s klesající délkou intervalu klesat k 0 rychleji než délka intervalu.

## 2.2.2 Vlastnosti

**Definice 6.** Definujeme posloupnost náhodných veličin značících dobu nastání  $k$ -té události  $\{\xi_k, k \in \mathbb{N}\}$  předpisem  $\xi_k := \inf_{t \geq 0} \{N(t) = k\}$ ,  $k = 1, 2, \dots$  a  $\xi_0 := 0$ . Dále definujeme dobu mezi dvěma po sobě jdoucími událostmi, tj. dobu mezi  $k$ -tou a  $(k-1)$ -ní událostí, jako náhodné veličiny  $T_k = \xi_k - \xi_{k-1}$ .

Zavedené veličiny jsou ilustrovány na obrázku 2.1.

**Věta 3.** Necht  $\{N(t), t \geq 0\}$  je Poissonův proces s intenzitou  $\lambda > 0$ .

- Veličiny  $T_k, k = 1, 2, \dots$  mají exponenciální rozdělení s parametrem  $\lambda$ .
- Veličiny  $\xi_k = \sum_{i=1}^k T_i$  mají Erlangovo rozdělení s parametry  $\lambda$  a  $k$ .

Rozdělení náhodných veličin  $T_k \sim \text{Exp}(\lambda)$  je ve shodě s bodem 2 ve větě 2 požadujícím nezávislé stacionární přírůstky. Ten říká, že nastání jedné události nemá vliv na nastání jiných událostí. Exponenciální rozdělení „nemá paměť“, tj. pro  $X \sim \text{Exp}(\lambda)$  platí  $P[X > t + h | X > t] = P[X > h]$ .

Díky těmto vztahům lze Poissonův proces chápat jako součet posloupnosti nezávislých náhodných veličin s exponenciálním rozdělením. Tato vlastnost se hodí pro vytváření modelů založených na Poissonově rozdělení.

**Věta 4.** Necht je  $\{T_k\}_{k=1}^{\infty}$  posloupnost dob mezi dvěma po sobě jdoucími událostmi,  $T_k \sim \text{Exp}(\lambda), \lambda > 0$  a  $\{\xi_k\}_{k=1}^{\infty} = \sum_{k=1}^{\infty} T_k$ . Potom  $N(t) = \sum_{k=1}^{\infty} I\{\xi_k \leq t\}$  definuje Poissonův proces s intenzitou  $\lambda$ .

V předchozích částech jsme uvažovali proces s konstantní intenzitou  $\lambda > 0$ . Nyní Poissonův proces zobecníme tak, že budeme uvažovat intenzitu jako funkci času  $t$ . Tato vlastnost je v praxi užitečná, jelikož díky ní lze modelovat situace, kdy je nastání události v určitých časech pravděpodobnější.

Nehomogenní Poissonův proces zavedeme obdobně jako homogenní. Hlavní rozdíl spočívá v tom, že již nepředpokládáme stacionární přírůstky.

**Definice 7.** Necht  $\lambda(t) : [0, \infty) \rightarrow [0, \infty)$  je integrovatelná funkce. Čítací proces  $\{N(t), t \geq 0\}$  se nazývá nehomogenní Poissonův proces s intenzitou  $\lambda(t)$  splňuje:

1.  $N(0) = 0$ ,
2. proces má nezávislé přírůstky,
3.  $P[N(t+h) - N(t) = 1] = \lambda(t)h + o(h)$ ,
4.  $P[N(t+h) - N(t) \geq 2] = o(h)$ ,

kde  $o(h)$  značí libovolnou funkci takovou, že  $\lim_{h \rightarrow 0^+} \frac{o(h)}{h} = 0$ .

Stejně tak lze zavést nehomogenní Poissonův proces v analogii s definicí 1. Počet událostí je náhodná veličina s Poissonovým rozdělením, nyní však parametr může záviset na čase nastání události, tj.  $N(t+h) - N(t) \sim \text{Pois}\left(\int_t^{t+h} \lambda(\xi) d\xi\right)$ . Pokud je  $\{N(t), t \geq 0\}$  homogenní Poissonův proces s intenzitou  $\lambda = 1$ , potom definujeme nehomogenní proces jako  $N(\Lambda(t))$ , tj. homogenní proces s nelineární časovou osou. Toto  $\Lambda$  definované jako  $\Lambda(t) = \int_0^t \lambda(\xi)(d)\xi$  nazýváme také míra intenzity procesu.

Všimněme si, že pro konstantní  $\lambda(t) = \lambda$  mají přírůstky  $N(t+h) - N(t)$  Poissonovo rozdělení s parametrem  $\int_t^{t+h} \lambda d\xi = \lambda h$  a jde o homogenní proces.

## 2.3 Značkový Poissonův proces

Tato část popisující značkový Poissonův proces vychází především z knih Kingman (1993) a Daley a Vere-Jones (2003). Připomeňme, že  $N(t)$  označuje počet událostí v intervalu  $[0, t]$ , tj.  $N(t) = \max\{n : \xi_n \leq t\}$ . Zjednodušený zápis  $\xi \in N(t)$  budeme chápat tak, že událost  $\xi$  nastala do času  $t$ .

### 2.3.1 Vlastnosti

Před definicí značkového procesu uvedeme několik důležitých vlastností. Následující věta bývá v literatuře označována jako *Superposition Theorem*, viz například Kingman (1993).

**Věta 5.** *Je-li dáno  $k$  nezávislých Poissonových procesů  $\{N_k(t), t \geq 0\}$ ,  $k \in \mathbb{N}$  s parametry  $\lambda_k$  a  $N(t) = \sum_{k=1}^{\infty} N_k(t)$ . Pak  $\{N(t), t \geq 0\}$  tvoří Poissonův proces s parametrem  $\sum_{k=1}^{\infty} \lambda_k$ .*

*Poznámka.* Vybereme-li bod  $\xi$  z konečného součtu  $k$  Poissonových procesů, pak pravděpodobnost, že bod  $\xi$  patří do  $N_j$  je  $P[\xi \in N_j] := \lambda_j / \sum_{i=1}^k \lambda_i$ .

Opačnou operací je ředění (ztenčení, angl. *thinning*), jenž spočívá v odebrání bodů procesu podle nějakého pravidla. Zbývající body vytvoří nový proces, stejně tak body odstraněné. Pro jednoduchost začněme nejdříve s procesem, kde se vyskytují pouze dva typy událostí.

**Věta 6.** *Nechť  $\{N(t), t \geq 0\}$  je Poissonův proces s intenzitou  $\lambda$ . Uvažujme, že každá událost (nezávisle na ostatních) je buď typu 1 s pravděpodobností  $p$  nebo typu 2 s pravděpodobností  $1 - p$ , kde  $p \in (0, 1)$ . Označme  $N_1(t)$  počet událostí typu 1 a  $N_2(t)$  počet událostí typu 2. Pak*

- $\{N_1(t), t \geq 0\}$  je Poissonův proces s intenzitou  $\lambda p$ ,
- $\{N_2(t), t \geq 0\}$  je Poissonův proces s intenzitou  $\lambda(1 - p)$ ,
- procesy  $\{N_1(t), t \geq 0\}$  a  $\{N_2(t), t \geq 0\}$  jsou nezávislé.

Větu lze zobecnit pro  $k$  typů událostí s analogickým značením následovně. Jde o tzv. *Colouring Theorem*, jelikož si lze představit, že máme  $k$  různých barev a každý z bodů Poissonova procesu nabarvíme  $i$ -tou barvou s pravděpodobností  $p_i$  nezávisle na obarveních ostatních bodů.

**Věta 7.** *Mějme Poissonův proces  $\{N(t), t \geq 0\}$  s intenzitou  $\lambda$ . Každý bod zařadíme nezávisle na ostatních do některé z  $k$  kategorií s pravděpodobností  $p_i$ ,  $\sum_{i=1}^k p_i = 1$ ,  $i \in \{1, 2, \dots, k\}$ . Označme  $N_i(t)$  počet bodů typu  $i$ . Pak jsou  $N_i(t)$  nezávislé náhodné veličiny s Poissonovým rozdělením s příslušnými intenzitami  $\lambda_i = p_i \lambda$ .*

*Poznámka.* Pokud závisí pravděpodobnosti  $p_i$  na čase nastání, potom platí

$$N_i(t) \sim \text{Pois} \left( \lambda \int_0^t p_i(\xi) d\xi \right).$$

I díky těmto operacím aplikovaným na bodové procesy lze vytvářet vhodné matematické modely. Jedním z důvodů, proč je Poissonův proces často používaný, je ten, že za určitých podmínek vznikají po spojení nebo rozdělení Poissonových procesů opět Poissonovy procesy.



### 2.3.2 Typy modelů

Existuje velké množství tříd modelů, které založeny na bodových procesech. Podstatou jejich fungování je vytvoření výsledného modelu pomocí dvou kroků: nejdříve generují proces a následně na generované body aplikují nějakou operaci, která je podmíněná realizací původního procesu. Jedná se například o Coxovy nebo shlukové procesy, které v podstatě rozšiřují jednoduchý Poissonův proces stejným způsobem jako smíšené a složené rozdělení rozšiřuje Poissonovo. Velkou roli zde hraje nezávislost, díky které lze některé vztahy zjednodušit. Tyto procesy jsou popsány ve většině odbornější literatury o bodových procesech (viz například Kingman (1993), Daley a Vere-Jones (2003)).

Dalším příkladem jsou pak právě značkové procesy, které popisují události na základě jejich polohy a nějaké vlastnosti. Jak později uvidíme, jde o speciální typ bodového procesu na kartézsky násobených prostorech (angl. *product space*).

Předpokládáme, že se body (události) vyskytují v kompletních separabilních metrických prostorech  $\mathcal{X}$ . Pro účely této bakalářské práce dále bude  $\mathcal{X} = \mathbb{R}^+$ . Poznamenejme ale, že lze uvažovat i prostory více dimenzí případně jiné obecnější prostory. Například modely založené na značkových procesech s událostmi v  $\mathbb{R}^2$  jsou poměrně často užívané v ekologii, demografii nebo seismologii.

Dále budeme značit  $A \in \mathcal{B}(\mathcal{X})$ , tedy libovolnou borelovskou množinu. Zápis  $N(A)$  pak značí počet bodů v množině  $A$ , tedy v souvislosti se zvolenou definicí Poissonova procesu zavádíme  $N(A)$  jako přírůstek na množině  $A$ . Poznamenejme, že na omezených množinách  $A$  musí platit  $N(A) < \infty$ .

### 2.3.3 Definice

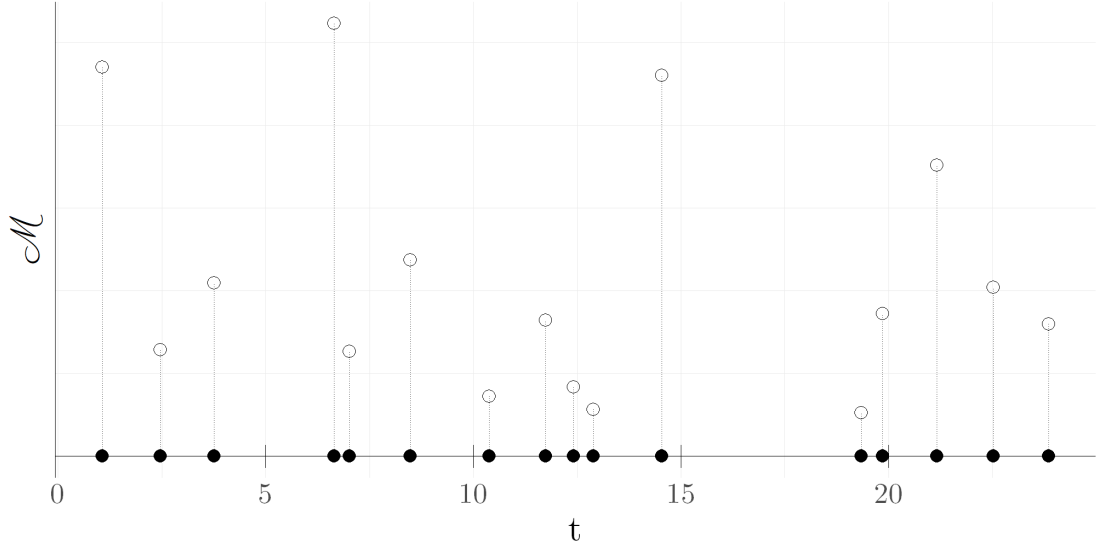
V mnoha modelech nás kromě popisu bodového procesu zajímají i nějaké další informace, které bod přináší. Zavádíme proto pojem značkový (kótovaný) bodový proces definovaný jako posloupnost bodů se značkami (kótami), což jsou náhodné veličiny, které zpravidla říkají, co se v daném bodě stalo. Ve značkovém Poissonově procesu tyto značky často reprezentují váhy spojené s daným bodem, například peněžní hodnotu ve financích, sílu v seismologii atd.

**Definice 8.** *Značkový Poissonův proces na úplném separabilním metrickém prostoru  $\mathcal{X}$  se značkami v úplném separabilním metrickém prostoru  $\mathcal{M}$  je bodový proces  $N^* = \{(\xi_i, m_i), \xi_i \in \mathcal{X}\}$  na  $\mathcal{X} \times \mathcal{M}$ , kde  $N$  je Poissonův proces na  $\mathcal{X}$  a  $\xi_i$  jsou příslušné body tohoto procesu.*

Značkový bodový proces lze chápat jako proces na prostoru  $\mathcal{X} \times \mathcal{M}$  (viz obrázek 2.2) a příslušný neznačkový proces  $N$  je pak marginálním procesem umístění bodů.

Jelikož  $N$  je neznačkový bodový proces (konečná míra na omezené množině), musí platit  $N(A) = N^*(A \times \mathcal{M}) < \infty$ . Z tohoto důvodu ne každý bodový proces na  $\mathcal{X} \times \mathcal{M}$  je značkovým procesem. Pokud je ale prostor značek  $\mathcal{M}$  konečný, tj. nějaká množina  $\{1, \dots, m\}$  pro libovolné  $m$  konečné, je konečný i každý prvek  $N_i^*(\cdot) = N^*(\cdot \times \{i\})$  a lze psát  $N(\cdot) = N^*(\cdot \times \{1, \dots, m\}) = \sum_{i=1}^m N_i(\cdot)$ , kde  $N$  je příslušný neznačkový proces.

Na obrázku 2.2 je simulace značkového Poissonova procesu, kde příslušný neznačkový proces je definovaný na nezáporné reálné přímce značící čas,  $t \geq 0$ . Náhodné značky nabývají hodnot v prostoru značek  $\mathcal{M}$ .



Obrázek 2.2: Ilustrace značkováného Poissonova procesu.

Intenzita procesu  $\lambda_{N^*}(A \times M)$  udává střední hodnotu počtu událostí (bodů) v množině  $A \subset \mathcal{X}$ , které mají značku v množině  $M \in \mathcal{B}(\mathcal{M})$ .

**Definice 9.** *Intenzita značkováného bodového procesu  $N^*$  je definována jako  $\lambda_{N^*}(A \times M) = \mathbb{E}[N(A \times M)]$ , kde  $t \geq 0$  a  $M \in \mathcal{B}(\mathcal{M})$  je libovolná množina v  $\sigma$ -algebře prostoru značek  $\mathcal{B}(\mathcal{M})$ .*

Dále definujeme dva pojmy související se značkami procesu. První z těchto vlastností je nezávislost, která je pro značkováný Poissonův proces definována následovně.

**Definice 10.** *Nechť  $N^* = \{(\xi_i, m_i), \xi_i \in N\}$  je značkováný Poissonův proces. Řekneme, že proces  $N^*$  má nezávislé značky, pokud jsou značky  $\{m_i\}$  vzájemně nezávislé náhodné veličiny.*

Druhým pojmem jsou nepředvídatelné značky, což jsou takové, které nezávisí na předchozích hodnotách příslušného značkováného procesu.

**Definice 11.** *Nechť  $N^* = \{(\xi_i, m_i), \xi_i \in N\}$  je značkováný Poissonův proces a  $\mathcal{X} = \mathbb{R}$ . Pak říkáme, že  $N^*$  má nepředvídatelné značky, pokud rozdělení značky v bodě  $\xi_i$  nezávislé na bodech a značkách  $\{(\xi_j, m_j), \xi_j \in N\}$  pro  $\xi_j < \xi_i$ .*

Tedy jsou-li značky  $m_i$  nezávislé stejně rozdělené náhodné veličiny, říkáme, že má značkováný proces nezávislé značky. Naopak pokud jsou značky podmíněně iid (pro daný předchozí průběh procesu), říkáme, že jsou nepředvídatelné.

Je-li rozdělení bodového procesu invariantní vůči posunutí bodů se zachováním značek, jde o stacionární značkováný proces. Definice stacionarity pro případ značkováného procesu je analogií definice 4 pro neznačkováný proces.

**Definice 12.** *Řekneme, že značkováný bodový proces  $N^*$  je stacionární, jestliže rozdělení procesu  $N^* + y = \{(\xi_i + y, m_i), (\xi_i, m_i) \in N^*\}$  je stejné jako rozdělení procesu  $N^*$  pro libovolné  $y \in \mathbb{R}$ .*

Nyní nás bude zajímat struktura procesů s nezávislými značkami. V mnoha aplikacích je  $\mathcal{M} = \mathbb{R}^+$  a definujeme náhodnou míru

$$\lambda(A) = \int_{A \times \mathcal{M}} mN(d\xi \times dm) = \sum_{\xi_i \in A} m_i.$$

Tento vzorec je zobecněním vztahu zavedeného v definici 9. Všimněme si také, že pokud  $m_i = m$  skoro jistě pro všechna  $i$ , pak  $\lambda(A) = mN(A)$ .

Pokud se na značky díváme jako na posloupnost hodnot reprezentující váhy, jak bylo zmíněno na začátku části 2.3.3, měří  $\lambda$  celkovou hodnotu přes určitý časový interval nebo oblast pro body s příslušnými vahami.

Při modelování procesu a značek nás bude zajímat inference intenzity procesu v čase a příslušné podmíněné rozdělení značek.

Následující věta je velmi důležitá, jelikož díky ní lze za určitých předpokladů modelovat značky a body zvlášť. Zmíněným předpokladem je nezávislost těchto dvou náhodných veličin.

**Věta 8.** *Značkovaný Poissonův proces s podmíněným rozdělením značek  $F(\cdot|\cdot)$  s příslušným neznačkovaným procesem s parametrem  $\lambda$  je ekvivalentní Poissonovu procesu na prostoru  $\mathcal{X} \times \mathcal{M}$  s mírou intenzity  $\Lambda(d\xi \times dm) = \lambda(d\xi)F(dm|\xi)$ .*

Jde o zobecnění „barvení“, které bylo zmíněno ve větě 7. Díky tomu lze proces rozdělit na podprocesy podle hodnot značek a interpretovat ho jako  $k$ -tici procesů.

Z dříve uvedených vět a definice značkovaného procesu, lze odvodit následující. Pokud proces rozdělíme na podprocesy  $N_i$  podle hodnot značek  $i = 1, \dots, k$ , jsou tyto procesy Poissonovými s intenzitou  $\lambda(A)p_i$ , kde  $p_i$  je pravděpodobnost, že bude mít daný bod  $\xi_i$  značku  $m_i$ . Značky se obvykle považují za iid veličiny, mohou ale také záviset na čase. V tomto případě by pak výsledné podprocesy měly intenzity  $\lambda(A)p_i(A)$ . Navíc, pokud předpokládáme, že jednomu bodu  $\xi_i$  přiřadíme pouze jednu značku  $m_i$  z množiny  $\mathcal{M}$ , jsou podprocesy nezávislé náhodné veličiny.

Předpoklad nezávislých značek je poněkud restriktivní a v praxi je možné, že budou značky nějakým způsobem záviset na poloze (času nastání události). Příkladem takových značek jsou značky vytvořené určitým postupem založeným na neznačkovaných bodech (například jde o geometrické rozmístění – vzdálenost k nejbližšímu bodu procesu, počet následujících bodů v dané vzdálenosti atd.).

Značkování může také záviset na intenzitě příslušného bodového procesu. Pro ilustraci situací, kdy rozdělení značek závisí na intenzitě procesu, se často uvádějí příklady z oblasti biologie rostlin (rostliny mohou být systematicky menší v oblastech s vyšší hustotou porostu, například z důvodu konkurence a tím pádem nedostatku živin) nebo fyziky materiálů (velikost částic může záviset na hustotě jejich výskytu v nějakém materiálu).

V teorii rizika se často využívá tzv. složené rozdělení. Definujme tedy závěrem složený Poissonův proces.

**Definice 13.** *Nechť  $S(t) = \sum_{i=1}^{N(t)} Y_i, t \geq 0$ , kde  $\{N(t), t \geq 0\}$  je Poissonův proces a  $\{Y_i, i \geq 1\}$  je posloupnost nezávislých stejně rozdělených náhodných veličin, které jsou navíc nezávislé na  $\{N(t), t \geq 0\}$ . Pak náhodný proces  $\{X(t), t \geq 0\}$  nazýváme složený Poissonův proces.*

Jde o speciální případ značkovaného Poissonova procesu. Funkce  $S(t)$  je funkcí značkovaného procesu definovanou pro každé  $t \geq 0$  jako součet všech značek v takových bodech procesu, které nastaly do času  $t$ .

## 2.4 Poznámky ke značení

Slovo *bodový* se v označení často vynechává. Existují však i jiné Poissonovy procesy, které místo bodů popisují složitější matematické objekty jako například přímky. Těmto procesům se věnuje i kniha Kingman (1993) na stranách 73–78. Takové procesy lze také interpretovat jako bodové s tím rozdílem, že tentokrát jsou definované na prostoru přímek. V celé této práci mluvíme vždy o bodových procesech, i pokud to není explicitně zmíněno.

Definice Poissonova procesu se také liší, zejména v závislosti na oblasti použití. Na Poissonův proces  $N$  se lze dívat jako na náhodnou množinu. Pak píšeme  $\xi_i \in N$ , což značí, že  $\xi_i$  je náhodný bod, který náleží do Poissonova bodového procesu  $N$ .

Obecněji je Poissonův proces náhodná čítací míra a  $N(B)$  chápeme jako počet bodů Poissonova bodového procesu  $N$  v nějaké borelovské množině  $B$ , což je náhodná veličina.

Na reálné přímce, což je většinou případ uvažovaný v této práci, lze Poissonův bodový proces definovat jako čítací proces se značením  $\{N(t), t \geq 0\}$ . Znalost  $N(t)$  pro všechna  $t \geq 0$  postačuje k tomu, abychom určili  $N(A)$  pro borelovské množiny  $A \subset [0, \infty)$ .

Lze ukázat, že pro Poissonův bodový proces  $N$  definovaný na eukleidovském prostoru  $\mathbb{R}^d$  a měřitelnou funkci  $f$  na  $\mathbb{R}^d$  jsou zápisy  $\int_{\mathbb{R}^d} f(t)N(dt)$  a  $\sum_{\xi_i \in N} f(\xi_i)$  ekvivalentní. Záleží na tom, zda se na problém díváme z pohledu teorie míry nebo teorie množin.

## 3. Aplikace

V této části bude teorie sepsaná v předchozích částech aplikována na konkrétní data. Ta budou zpracována ve statistickém softwaru R a jsou obsažena v balíčku OpVaR, který je v tomto softwaru dostupný k instalaci.

V dalších výpočtech budeme v souladu s metodou LDA uvažovat, že četnost je nezávislá na výši škod. Budeme také výše ztrát považovat za nezávislé stejně rozdělené náhodné veličiny.

### 3.1 Popis dat

V datovém souboru `lossdat` jsou obsažena data o ztrátě z rizikových událostí. Tento soubor se skládá ze 4 buněk, které by měly odpovídat 2 liniím podnikání a 2 typům událostí. Jak jsme uvedli již dříve, v této práci se nebudeme zabývat jednotlivými typy událostí a liniemi podnikání, proto byla zvolena pouze jedna buňka (tedy jeden typ události a jedna linie podnikání). Zvolili jsme první buňku z `lossdat`, modelování výší a počtu škod by u ostatních bylo založeno na stejném principu, pravděpodobně by se lišila volba rozdělení a odhady parametrů.

Buňka obsahuje údaje o výši ztráty v tisících amerických dolarů, časovou značku (dobu nastání události) a údaj o časovém období (40 tříměsíčních úseků). V této práci ale délku období budeme volit vlastní.

Dokumentace k tomuto balíčku neobsahuje informace o tom, jakým způsobem byla data vygenerována. Můžeme se pouze domnívat, že soubor vznikl například simulací pomocí metody Monte Carlo s přihlédnutím buď na nějaký známý vzorek reálných dat, nebo na charakteristické vlastnosti tohoto typu rizika.

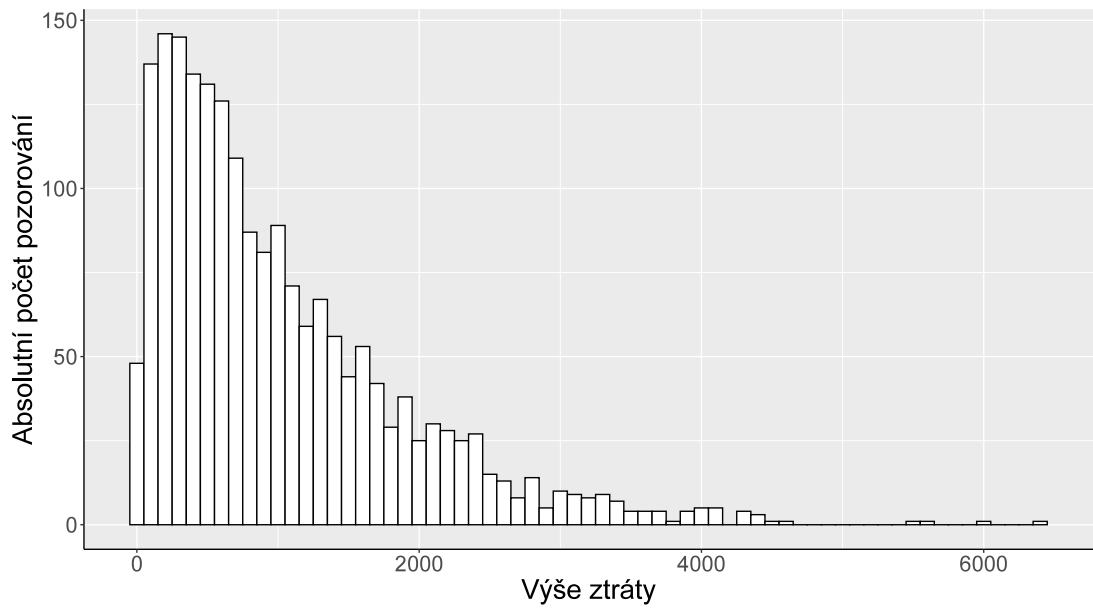
Simulace Monte Carlo se často používá, pokud není k dispozici dostatek dat. Postup je takový, že je nejdříve odhadnuto parametrické rozdělení ztrát a poté je provedena simulace metodou Monte Carlo spočívající v generování velkého počtu průběhů s využitím daného pravděpodobnostního rozdělení v tom smyslu, že náhodně vybíráme vzorky z tohoto rozdělení. Tím se získá větší množství dat, se kterými můžeme dále pracovat. Zásadní je ale kvalitně sestavený model.

Celkem máme k dispozici 1965 dat nasimulovaných pro roky 2007 až 2016. Popisné charakteristiky výší ztrát jsou zaznamenané v tabulce 3.1.

Minimum	Medián	Maximum	Průměr	Směrodat. odchylka
5	760	6382	1017	882

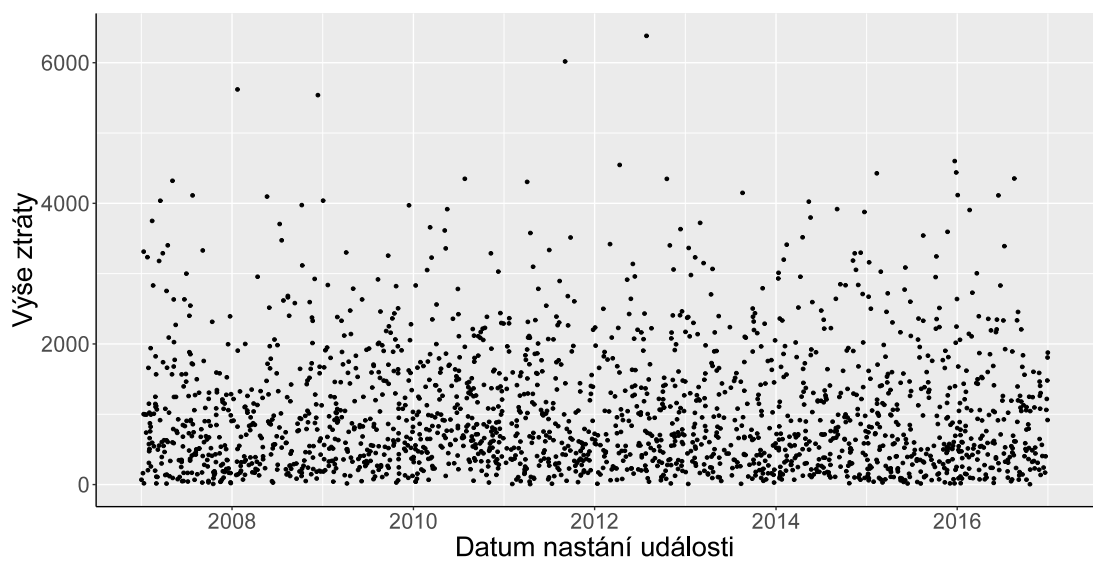
Tabulka 3.1: Popisné charakteristiky výší ztrát.

V tabulce 3.1 si všimněme značného rozdílu mezi minimální a maximální výší ztráty. Je znatelný rozdíl mezi mediánem a průměrem, což naznačuje, že empirické rozdělení výší nebude symetrické. Průměr je větší než medián, šikmost bude tedy kladná (konkrétně  $\approx 1,55$ ) a zešikmení vpravo, což indikuje odlehlá pozorování napravo od průměru. V souladu s dříve sepsanou teorií jde o rozdělení s těžkým pravým chvostem. Koeficient špičatosti je přibližně 6,14, jde tedy o leptokurtické rozdělení. Pro ilustraci je na obrázku 3.1 vykreslen histogram.



Obrázek 3.1: Histogram ztrát.

Data byla uměle nasimulovaná a mají vlastnosti, které jsou pro tento druh rizika typické. Především jde o nízkou četnost vyšších ztrát, jak je kromě výše zobrazeného histogramu na obrázku 3.1 vidět i ze scatterplotu na obrázku 3.2.



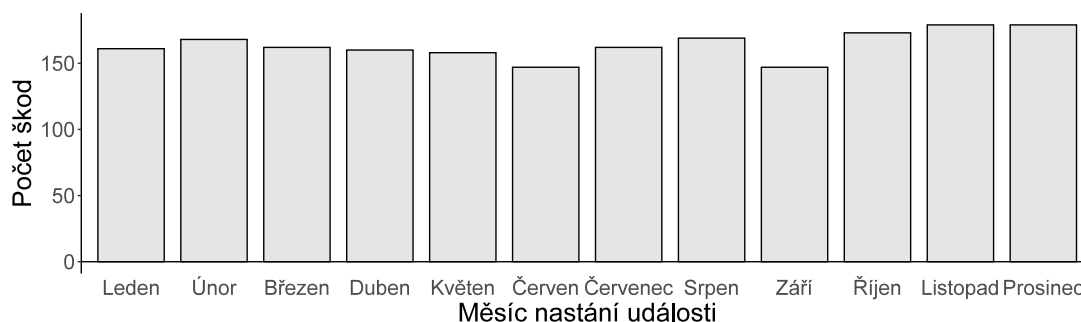
Obrázek 3.2: Ztráty v letech 2007–2016.

Nyní se budeme zabývat sestavením parametrického modelu pro popis počtu a výši škod. K tomu využijeme předpokladu, že jsou frekvence a závažnost ztráty nezávislé veličiny (tedy výše ztráty nezávisí na okamžiku, kdy k ní došlo, což se zdá jako oprávněný předpoklad).

## 3.2 Rozdělení počtu škod

Jak bylo zmíněno dříve, Poissonovo rozdělení je častou volbou pro modelování výskytů událostí. V této práci se jinými volbami nebudeme zabývat.

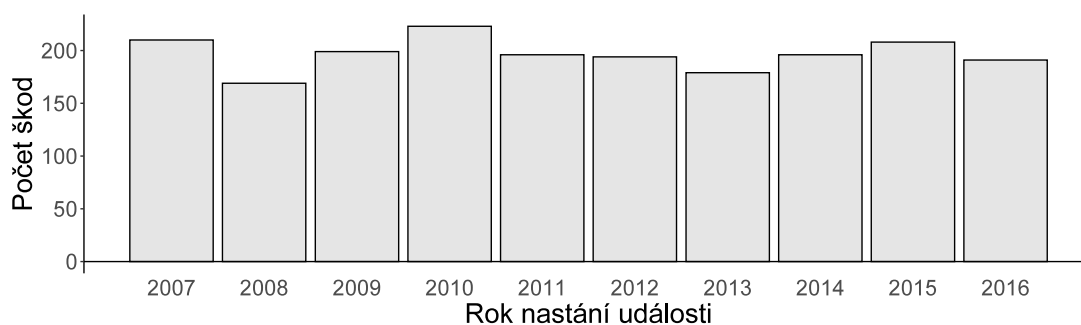
V kapitole 2 byly zmíněny dva druhy Poissonova procesu v závislosti na tom, zda je proces stacionární či nikoliv. Zavedli jsme homogenní případ, pokud je parametr  $\lambda$  konstantní, a jako zobecnění nehomogenní proces, který má parametr proměnný v čase. Abychom určili, o jaký případ se může jednat, podíváme se, zda je znatelný výraznější rozdíl v průběhu roku v jednotlivých měsících.



Obrázek 3.3: Počet ztrát v jednotlivých měsících.

Data zřejmě dle obrázku 3.3 nevykazují výrazné známky sezónnosti, tj. nelze říci, že by v určitých měsících nastávalo výrazně více škod než v jiných. Třeba v pojištnictví ale poměrně často počty nahlášených škod například v souvislosti s dopravními nehodami závisí na ročním období, zde by byla sezónnost znatelná.

Na obrázku 3.4 zobrazujícím roční počty nastalých událostí už data vykazují vyšší výkyvy, nicméně pro odhad parametru nehomogenního procesu bychom pravděpodobně potřebovali více informací. Data se sice pro jednotlivé roky lehce liší, nicméně změny nejsou nijak drastické a z toho důvodu pro následující části volíme homogenní Poissonův proces s konstantní intenzitou  $\lambda$ .



Obrázek 3.4: Počet ztrát v jednotlivých měsících.

Odhadem parametru  $\lambda$  Poissonova rozdělení je počet událostí za jednotkový časový interval. Zvolíme-li měsíční interval, získáme odhad  $\hat{\lambda} = \frac{1965}{10 \cdot 12} = 16,375$ . Očekáváme tedy asi 16 rizikových událostí měsíčně. Volba délky intervalu závisí na množství dat, která jsou k dispozici. Náš datový soubor obsahuje dostatek záznamů, může se ale stát, že banka takovým množstvím dat nebude disponovat. Řešení je několik. Díky aditivitě Poissonova rozdělení lze zvolit jiný jednotkový interval, dále lze využít například různé simulace nebo i externí databáze.

### 3.3 Rozdělení výší škod

V souladu s poznatky z kapitoly 1, kde jsme zkoumali statistické vlastnosti rozdělení výší škod, i s výsledky studie BCBS (2009), která ukázala, že jsou nejčastěji z parametrických rozdělení používaná log-normální, Weibullovo a gama, provedeme odhad parametrů právě pro tato rozdělení. K tomu využijeme funkci `fitdist()` z balíčku `fitdistrplus`.

Konkrétně byla funkce implementována následujícím způsobem.

```
> fitdist(x, "dist", method = "mge", gof = "ADR")
```

V tomto zápise značí `x` vektor ztrát a `"dist"` jedno ze zvolených rozdělení, tedy `"plnorm"`, `"pweibull"` a `"pgamma"`. Kombinace posledních dvou vstupních parametrů `method = "mge"` a `gof = "ADR"` nám umožňuje zvolit metodu MDE (*Minimum Distance Estimation*, zkratka MGE pochází z alternativního názvu *Maximum Goodness-of-fit Estimation*) s použitím Anderson-Darlingova kritéria s vyšší vahou pro pravé chvosty rozdělení, což je právě ta část rozdělení, která je pro modelování ztrát z operačního rizika důležitá. Princip je analogický jako u metody nejmenších čtverců, která je také jednou z metod MDE. Při této volbě `method = "mge"`, `gof = "ADR"` chceme minimalizovat integrál vážených kvadrátů rozdílů mezi empirickými a odhadnutými funkcemi. Tedy matematické kritérium vychází z integrálu tvaru

$$\int_{-\infty}^{\infty} \frac{(F_n(x) - F(x))^2}{1 - F(x)} dx.$$

Parametrizace všech tří dříve vybraných rozdělení je v programu R v tomto kontextu následující.

- Log-normální  $\text{LN}(\mu, \sigma)$  má hustotu  $f(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left\{-\frac{(\log(x)-\mu)^2}{2\sigma^2}\right\}$ .
- Weibullovo  $\text{Weibull}(k, \gamma)$  má hustotu  $f(x) = \frac{k}{\gamma} \left(\frac{x}{\gamma}\right)^{k-1} \exp\left\{-\left(\frac{x}{\gamma}\right)^k\right\}$ .
- Gama  $\Gamma(\alpha, \beta)$  má hustotu  $f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} \exp\left\{-\frac{x}{\beta}\right\}$ .

Odhadnuté parametry pro jednotlivá rozdělení výší škod jsou v tabulce 3.3.

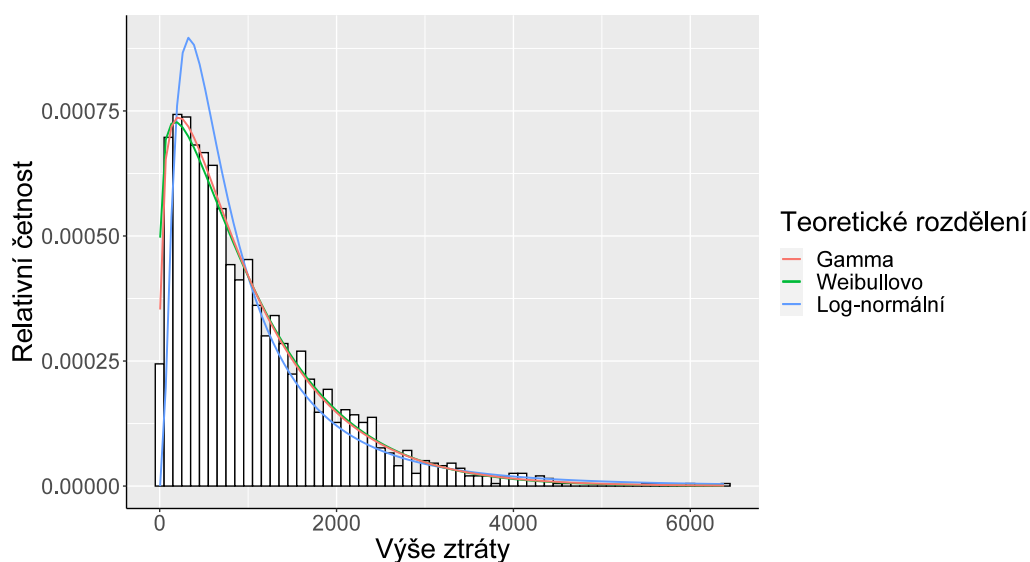
$\text{LN}(\hat{\mu}, \hat{\sigma})$	$\text{Weibull}(\hat{k}, \hat{\gamma})$	$\Gamma(\hat{\alpha}, \hat{\beta})$
$\hat{\mu} = 6,608$	$\hat{k} = 1,143$	$\hat{\alpha} = 1,265$
$\hat{\sigma} = 0,901$	$\hat{\gamma} = 1067,761$	$\hat{\beta} = 809,812$

Tabulka 3.2: Odhady parametrů zvolených rozdělení.

*Poznámka.* K odhadu parametrů pro gama rozdělení bylo třeba zvolit počáteční hodnoty  $\alpha_0, \beta_0$ . Používáme vzorce  $\alpha_0 = \bar{m}_n^2/s_n^2$ ,  $\beta_0 = s_n^2/\bar{m}_n$ , kde  $\bar{m}_n$  je výběrový průměr a  $s_n^2$  výběrový rozptyl. Výběrem máme na mysli pozorované výše škod.

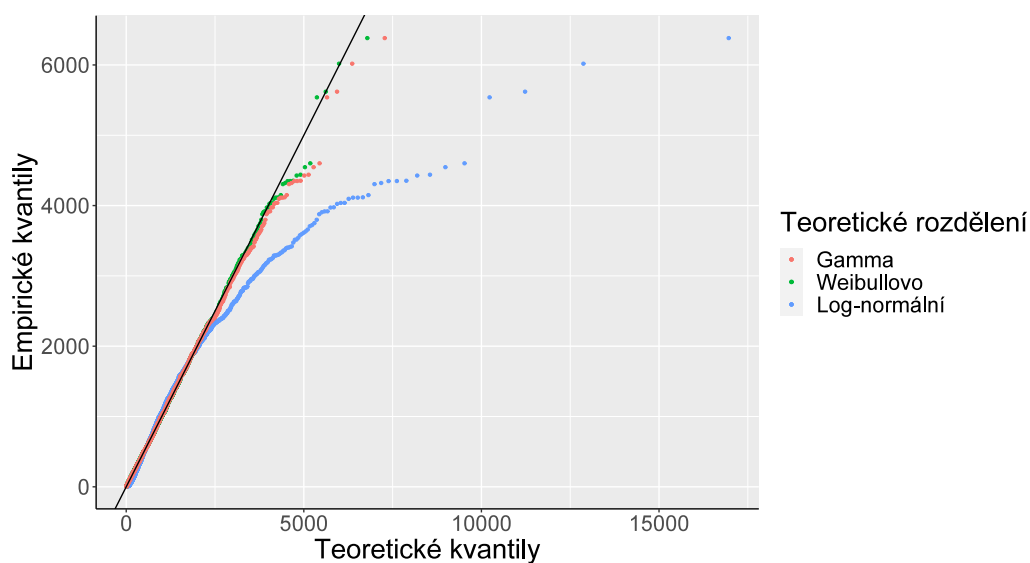


Nyní vizuálně porovnáme odhadnutá rozdělení a pozorovaná data. Porovnání histogramu a hustot zvolených rozdělení s odhadnutými parametry je ilustrováno obrázkem 3.5. Zdá se, že log-normální rozdělení data dobře nevystihuje, nicméně histogram není ideální volbou pro rozhodování, jelikož podstatně záleží na volbě šířky sloupců. Navíc v tomto konkrétním případě jsou hustoty Weibullova a gama rozdělení tak blízko, že lze jen těžko posoudit, které je lepší. Díky histogramu však získáváme prvotní informaci o tom, zda odhadnuté rozdělení není zásadně odlišné. Na základě tohoto vizuálního porovnání bychom vyřadili log-normální rozdělení.



Obrázek 3.5: Porovnání histogramu a teoretických hustot.

Dále by nás mohl zajímat spíše Q-Q plot, který je zachycený na obrázku 3.6. Z kvantilového diagramu již získáme lepší představu o vhodnosti jednotlivých rozdělení. I nyní vidíme, že log-normální rozdělení zřejmě není vhodnou volbou. Dle kvantilového diagramu se zdá, že nejbližší je Weibullovo rozdělení.



Obrázek 3.6: Kvantilový diagram.

Podívejme se tedy na formální statický test, konkrétně Anderson-Darlingův. Tento test přiřazuje vyšší váhy chvostům rozdělení. Testujeme nulovou hypotézu, že data pochází postupně z každého ze tří zvolených rozdělení (s odhadnutými parametry) proti alternativě, že pochází z libovolného jiného rozdělení.

Konkrétně byla použita funkce `ad.test()` z balíčku `gofest` se vstupem:

- `x` značící vektor zaznamenaných ztrát,
- `"dist"` je testované rozdělení, tj. `"plnorm"`, `"pweibull"` a `"pgamma"`,
- `params` specifikuje příslušné (odhadnuté) parametry a
- `estimated = TRUE` znamená, že parametry byly odhadnuty.

```
> ad.test(x, "dist", params, estimated = TRUE)
```

Jak jsme zmínili na samém konci kapitoly 1, v testu je nutné zohlednit to, že příslušné parametry byly odhadnuty z dat, na kterých chceme testovat, jinak by byly výsledky testů značně zkreslené (získali bychom poměrně vysoké  $p$ -hodnoty z důvodu provádění testu na stejných datech, ze kterých byly dříve odhadnuty parametry).

Jedna z metod poskytujících řešení je podrobněji popsána v Braun (1980) na stranách 53–63 včetně odvození jejích statistických vlastností. Tuto metodu využijeme i v této práci.

Parametry testovaných teoretických rozdělení jsme dříve odhadli z původního datasetu obsahujícího 1965 záznamů, viz tabulka 3.3. K testování nyní náhodně rozdělíme data na více přibližně stejně velkých skupin (žádná skupina by neměla obsahovat více dat než 10 až 15 procent dat původního vzorku). Zde jsme využili dělení do 44 skupin, tedy každá skupina obsahuje 44/1965 původních dat, což je lehce přes 2 %. S užitím těchto skupin vypočteme příslušnou testovou statistiku a  $p$ -hodnotu pro rozdělení odhadnuté z celého vzorku.

Jelikož je rozdělení dat do skupin provedené příkazem `estimated = "TRUE"` náhodné, mohou se po opakovaném výpočtu výsledné hodnoty lišit. Nulovou hypotézu bychom na hladině  $\alpha = 0,05$  zamítali pro log-normální rozdělení. Z tabulky 3.3 je také možné vyčíst, že pro Weibullovo rozdělení vyšla nejnižší hodnota testové statistiky, nicméně rozdíl oproti gama rozdělení opět není nijak výrazný, pokud bereme v potaz to, že výsledky jednotlivých testů se mohou lišit v závislosti na dělení do skupin, které je náhodné.

Rozdělení	LN	Weibull	$\Gamma$
Parametry	(6,608; 0,901)	(1,143; 1 067,761)	(1,265; 809,812)
Testová statistika	9,0048	3,2771	3,4190
$p$ -hodnota	0,0020	0,5891	0,5299

Tabulka 3.3: Výsledky Anderson-Darlingova testu s randomizací.

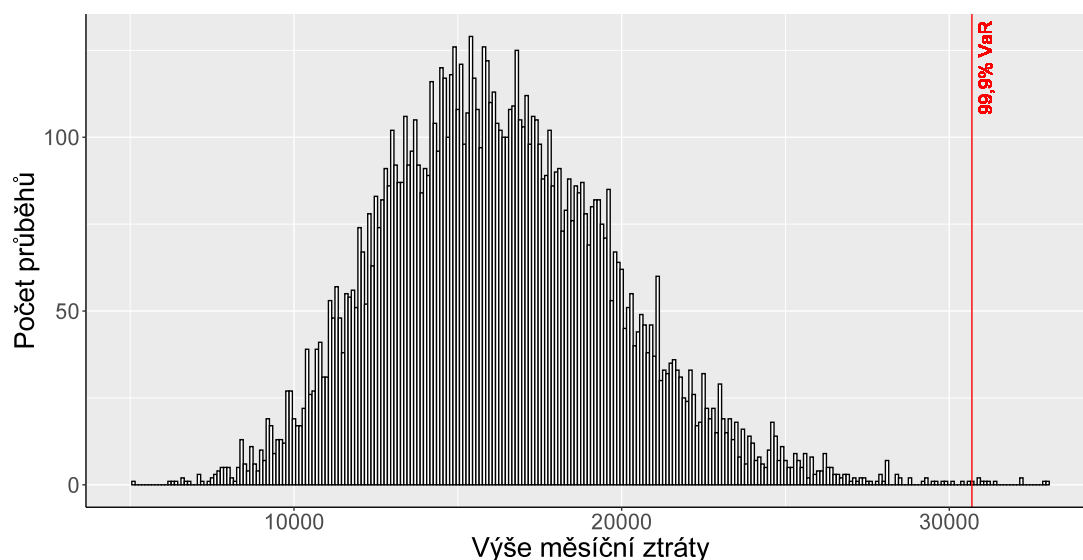
## 3.4 Simulace

Nyní máme potřebné údaje pro generování simulací. Díky těmto simulovaným průběhům pak můžeme vypočítat maximální možnou ztrátu pro zvolenou hladinu spolehlivosti. Z předpokladu nezávislosti četnosti nastalých událostí a výší ztrát (tedy neznačkovaného Poissonova procesu a značek) lze modelovat tyto dvě složky zvlášť.

K simulacím využijeme první vlastnosti z věty 3 a budeme simulovat doby mezi událostmi, z nichž pomocí kumulativních součtů získáme doby nastání událostí. K simulaci výší škod využijeme Weibullovo rozdělení s odhadnutými parametry, tj.  $\text{Weibull}(1,143; 1067,761)$ .

Zaměříme na predikci možných ztrát v následujícím měsíci, tedy v lednu 2017. Provedeme 10 000 simulací scénářů, tj. predikcí vývoje procesu do budoucna pro tento časový interval o délce jednoho měsíce. Predikce vychází z vybraných rozdělení s odhadnutými parametry, tedy  $\text{Pois}(16,375)$  a  $\text{Weibull}(1,143; 1067,761)$ . Následně spočítáme celkovou ztrátu pro tento úsek jako součet simulovaných výší ztrát jednotlivých scénářů.

Díky těmto simulacím můžeme vypočítat například hodnotu v riziku (VaR, viz podkapitola 1.2). Histogram výší škod ze simulovaných událostí je na obrázku 3.7 a zobrazuje, kolik průběhů má výslednou celkovou měsíční ztrátu reprezentovanou příslušným sloupcem. Zároveň je vyznačena hodnota VaR pro  $\alpha = 0,999$ . V našem případě vychází  $VaR_{0,999} = 30690$ . Tento výsledek interpretujeme tak, že nejvyšší možná ztráta pro měsíc leden na 99,9% hladině spolehlivosti je 30 690 000 USD.



Obrázek 3.7: Histogram simulovaných měsíčních výší ztrát.

Odhad hodnoty v riziku je často prvním cílem analýzy takovýchto dat. Díky VaR finanční instituce získá odhad maximální možné ztráty, na kterou by se měla v nadcházejícím období připravit.

V praxi by byl postup pravděpodobně opačný, tj. nejdříve bychom provedli backtesting a následně, pokud by byl sestavený model vhodný, by se vypočítal odhad maximální možné ztráty a kapitálového požadavku. Přesto se ale na závěr zaměříme ještě na to, jak námi sestavené modely obstojí v porovnání se skutečně nastalou měsíční ztrátou. V kapitole 3.3 jsme mohli vidět, že Weibullovo i gama rozdělení s příslušnými parametry vystihují data poměrně dobře.

K porovnání využijeme dříve představená data s tím rozdílem, že při odhadech nebudeme zahrnovat poslední měsíc, tedy prosinec roku 2016. Odhad parametrů pro Poissonovo rozdělení i pro rozdělení výše ztrát provedeme naprosto stejným způsobem jako v podkapitole 3.3, tentokrát ale z dat bez posledního měsíce. Výsledky odhadů jsou následující.

- Pro Poissonův proces je odhad intenzity  $\hat{\lambda} = 16,403$ .
- Pro Weibull( $k, \gamma$ ) jsou odhady parametrů  $\hat{k} = 1,141$  a  $\hat{\gamma} = 1068,175$ .
- Pro  $\Gamma(\alpha, \beta)$  jsou odhadnuté parametry  $\hat{\alpha} = 1,262$  a  $\hat{\beta} = 812,067$ .

Můžeme si všimnout porovnáním s tabulkou 3.3, že odhady se příliš nezměnily. To potvrzuje, že výše škod zřejmě nezávisí na čase nastání.

Opět provedeme 10 000 simulací průběhů, tentokrát pro Weibullovo i gama rozdělení. Pokud hodnoty seřadíme vzestupně, empirický odhad  $\widehat{VaR}_{99,5}$  získáme jako 9 950. hodnotu simulované měsíční ztráty ze všech 10 000. Odhad hodnoty v riziku pro měsíc prosinec je:

- $\widehat{VaR}_{99,5}^{(W)} = 26986,66$  v případě Weibullova rozdělení,
- $\widehat{VaR}_{99,5}^{(g)} = 27500,98$  s využitím gama rozdělení.

Skutečná celková ztráta za měsíc prosinec roku 2016 činila 11 586 000 USD. Odhady 26 986 660 USD ze simulací s Weibullovo rozdělením i 27 500 980 USD při využití gama rozdělení by tak zajistily tvorbu dostatečné kapitálové rezervy k pokrytí operační ztráty pro tento měsíc. V této konkrétní situaci můžeme také říci, že lepší volbou by pro banku bylo s drobným rozdílem Weibullovo rozdělení, jelikož riziko výsledný model riziko méně nadhodnotil a banka by tak měla více disponibilních prostředků, které by mohla investovat. Rozdíl mezi těmito dvěma odhady ale není nijak výrazný v porovnání s řádem výše ztráty.

# Závěr

S určitou mírou rizika se ve finančním sektoru setká každá instituce. Operační riziko je specifické svou provázaností s ostatními typy rizik a se všemi činnostmi banky, nelze ho tedy jednotně charakterizovat. Ztráty vyplývající z tohoto typu rizika mohou být obrovské, a i z toho důvodu bylo implementováno do výpočtu kapitálového požadavku v Basilejských dohodách. Mimo to se oblasti, ve kterých rizikové události nastávají, neustále vyvíjí vlivem rozvoje technologií i zvyšující se komplexity všech operací probíhajících v instituci. Z těchto důvodů finanční instituce více zaměřily svou pozornost na modelování ztrát v této oblasti, díky čemuž pak lze potenciální ztráty předpovídat, připravit se na ně a případně jim předcházet. Přístup tvorby vlastních modelů je bankami často preferovaný, jelikož každá instituce má své specifické charakteristiky, a právě díky interním modelům je může zohlednit při výpočtech.

V této bakalářské práci byl nejdříve shrnut potřebný související ekonomický a matematický aparát. Konkrétně jsme se v teoretické části věnovali základním pojmům z oblasti teorie operačního rizika. Uvedli jsme výhody, které zdůvodňují použití Poissonova rozdělení pro modelování počtů škod. Byla také diskutována volba vhodných rozdělení pro výše nastalých škod.

V následující části jsme se věnovali náhodným bodovým procesům, zejména Poissonovu bodovému procesu a jeho vlastnostem. Především jsme se soustředili na Poissonův značkovaný proces a jeho charakteristiky.

V aplikační části byla zpracována a analyzována data zahrnutá v balíčku `OpVar` ve statistickém softwaru R. Tato data byla použita k vytvoření modelu pro počet a výše škod spadající pod operační riziko banky. Poté bylo provedeno 10 000 simulací možných budoucích scénářů a z nich následně vypočtena hodnota v riziku pro následující měsíc. Závěrem bylo ověřeno, že sestavený model by stačil k pokrytí skutečné ztráty pro tuto banku. Samozřejmě volba modelu se pro jiné finanční instituce bude lišit, jelikož každá má jiné charakteristiky (specializace, velikost, místo působnosti atd.).

V průběhu tvorby modelu jsme se nesetkali s žádnými výraznějšími problémy, jelikož šlo o simulovaná data a situace tak byla oproti praxi značně jednodušší. V reálné situaci bychom ale museli zohledňovat více faktorů, jako například vliv inflace, snížení ztráty z rizikové události díky pojištění atd.

Těž je běžné, že instituce nemá dostatek interních dat. Tento problém se řeší kromě využití externích databází také různými simulacemi a predikcemi.

Navíc jsou také často zaznamenávány pouze ztráty překračující danou mez (tzv. `threshold`), ať už třeba z důvodu nákladů na uchování dat či prevence chyb při zaznamenávání velkého množství nízkých ztrát. Z důvodu nedostatku dat banka často sahá po datech z externích databází. V souvislosti s tím je třeba mít na paměti, že se může lišit i zvolený `threshold` nejen mezi různými institucemi, ale i mezi jednotlivými odděleními v bance. Pokud bychom toto useknutí dat ignorovali a modelovali bychom nepodmíněné rozdělení těchto neúplných dat, bude odhad parametrů rozdělení frekvence i severity a následně i celkový model pravděpodobně vychýlený. Při zanedbání dat pod zmíněnou mezí bude odhad VaR podhodnocený. V praxi ale většina ztrát plyne z dat, která danou mez překročí, a malé ztráty by tedy neměly výrazný vliv na VaR určenou horními kvantily.

Dalším problémem může být, že LFHI data budou vyžadovat jiný model (jinou volbu rozdělení) než HFLI data. Často se tak k modelování používá vícero různých rozdělení a modelujeme tělo a chvost zvlášť.

Jak uvádí Dutta a Perry (2006) některé tyto problémy může řešit například volba g-and-h nebo GB2 rozdělení, což jsou složitější rozdělení se 4 parametry.

Cílem bylo sepsat matematickou teorii související s Poissonovými procesy v jednotném značení a následně ji použít na data k modelování počtů a výší škod spadajících pod operační riziko banky. Díky sestrojenému modelu jsme byli schopni predikovat maximální možnou ztrátu pro nadcházející období a porovnat odhad se skutečnou ztrátou.

Závěrem stojí za zmínku, že ať už jde o vytváření modelů v jakékoli oblasti zájmu, je třeba mít na paměti, že vytvořené modely jsou vždy zjednodušením reálné situace.

# Seznam použité literatury

- BASEL COMMITTEE ON BANKING SUPERVISION (2009). *Observed range of practice in key elements of Advanced Measurement Approaches (AMA)*. [online]. Basel: Bank for International Settlements. ISBN 92-9131-799-3. Poslední změna 28.7.2009, [cit. 24.4.2022]. Dostupné z <https://www.bis.org/publ/bcbs160b.pdf>.
- BRAUN, HENRY (1980). A Simple Method for Testing Goodness of Fit in the Presence of Nuisance Parameters. *Journal of the Royal Statistical Society. Series B (Methodological)*. London: Royal Statistical Society, **42**(1). ISSN 0035-9246.
- CAMPBELL, SEAN D. (2005). A Review of Backtesting and Backtesting Procedures. *Finance and Economics Discussion Series*. Washington: Board of Governors of the Federal Reserve System (U.S.), **21** [cit. 8.5.2022]. Dostupné z <https://EconPapers.repec.org/RePEc:fip:fedgfe:2005-21>.
- CIPRA, TOMÁŠ (2015). *Riziko ve financích a pojišťovnictví: Basel III a Solvency II*. Praha: Ekopress. ISBN 978-80-87865-24-8.
- DALEY, D. J. a VERE-JONES, D. (2003). *An Introduction to the Theory of Point Processes*. New York: Springer. ISBN 978-0-387-21564-8.
- DUTTA, KABIR a PERRY, JASON (2006). A Tale of Tails: An Empirical Analysis of Loss Distribution Models for Estimating Operational Risk Capital. *Federal Reserve Bank of Boston Working Papers* [online]. Boston: FRB of Boston, **13**(06) [cit. 24.4.2022]. Dostupné z DOI:10.2139/ssrn.918880.
- KINGMAN, J. F. C. (1993). *Poisson processes*. Oxford: Clarendon Press. ISBN 0-19-853693-3.
- RIPPEL, MILAN a TEPLÝ, PETR (2008). Operational Risk – Scenario Analysis. *IES Working Papers* [online]. Prague: IES FSV, **15** [cit. 24.4.2022]. Dostupné z <http://ies.fsv.cuni.cz/sci/publication/show/id/3471/lang/cs%3E>.
- ROSS, SHELDON M. (1996). *Stochastic Processes*. 2. vyd. New York: Wiley. ISBN 0-471-12062-6.
- WERNZ, JOHANNES (2014). *Bank Management and Control*. Berlin: Springer. ISBN 978-3-642-40373-6.