

MASARYKOVA UNIVERZITA
PŘÍRODOVĚDECKÁ FAKULTA
ÚSTAV MATEMATIKY A STATISTIKY

Diplomová práce

BRNO 2017

BC. MARTIN BÍZA



MASARYKOVA UNIVERZITA
PŘÍRODOVĚDECKÁ FAKULTA
ÚSTAV MATEMATIKY A STATISTIKY



Modely výskytu událostí, jejich ověření a aplikace

Diplomová práce

Bc. Martin Bíza

Vedoucí práce: doc. RNDr. Petr Lánský, CSc.

Brno 2017

Bibliografický záznam

Autor:	Bc. Martin Bíza Přírodovědecká fakulta, Masarykova univerzita Ústav matematiky a statistiky
Název práce:	Modely výskytu událostí, jejich ověření a aplikace
Studijní program:	Matematika
Studijní obor:	Finanční matematika
Vedoucí práce:	doc. RNDr. Petr Lánský, CSc.
Akademický rok:	2016/2017
Počet stran:	88 + viii
Klíčová slova:	Stochastické procesy, Poissonův proces, procesy obnovy, intervaly mezi událostmi, neživotní pojištění

Bibliographic Entry

Author: Bc. Martin Bíza
Faculty of Science, Masaryk University
Department of Mathematics and Statistics

Title of Thesis: Models of series of events, statistical inference and applications

Degree Programme: Mathematics

Field of Study: Financial Mathematics

Supervisor: doc. RNDr. Petr Lánský, CSc.

Academic Year: 2016/2017

Number of Pages: 88 + viii

Keywords: Stochastic processes, Poisson process, renewal processes, interarrival times, non-life insurance

Abstrakt

V této diplomové práci se věnujeme studiu procesů, které popisují výskyt událostí. V první kapitole se seznamujeme s pojmy z oblasti stochastických procesů. Postupně na ně klademe omezující podmínky, které stále více konkretizují jejich zprvu zcela obecnou podobu. Druhá kapitola se zabývá procesy čistého zrodu s důrazem na Poissonův proces. Dále se věnujeme procesům obnovy, které jsou spojovacím můstkem Cramérova-Lundbergova a obecnějšího Sparre Andersenova modelu v teorii ruinování. V poslední kapitole aplikujeme teoretické poznatky na empirická data z oblasti pojišřovnictví.

Abstract

The diploma thesis is dedicated to the study of processes describing series of events. In the first chapter we introduce the basic terms related to stochastic processes. These are specified with the aid of restrictive conditions that further define their initially general nature. Further we study pure birth processes with a view to the Poisson process. In the next chapter renewal processes are mentioned that form a bridge between Cramér-Lundberg model to the more general Sparre Andersen one in the ruin theory. Eventually, we apply the theory to the empiric dataset taken from the insurance area.



ZADÁNÍ DIPLOMOVÉ PRÁCE

Akademický rok: 2016/2017

Ústav: Ústav matematiky a statistiky

Student: Bc. Martin Bíza

Program: Matematika

Obor: Finanční matematika

Ředitel *Ústavu matematiky a statistiky* PŘF MU Vám ve smyslu Studijního a zkušebního řádu MU určuje diplomovou práci s názvem:

Název práce: Modely výskytu událostí, jejich ověření a aplikace

Název práce anglicky: Models of series of events, statistical inference and applications

Oficiální zadání:

Pro modelování časového výskytu nerozlišitelných událostí se používá teorie stochastických bodových procesů. Cílem diplomové práce je seznámit se s touto teorií a popsat vlastnosti odhadů neznámých parametrů těchto modelů. Na simulovaných datech demonstrovat vyložené metody. Očekávanou částí diplomové práce je i aplikace na realná data podle volby studenta i rozšíření interpretace teorie v této oblasti.

Literatura: D.R. Cox, P.A.W. Lewis: *The Statistical Analysis of Series of Events*, Methuen, London, 1966. P.A.W. Lewis (Editor). *Stochastic Point Processes; Statistical Analysis, Theory and Applications*. Wiley-Interscience, New York and London, 1972.

Jazyk závěrečné práce:

Vedoucí práce: doc. RNDr. Petr Lánský, CSc.

Datum zadání práce: 4. 8. 2014

V Brně dne: 6. 11. 2016

Souhlasím se zadáním (podpis, datum):

.....
Bc. Martin Bíza
student

.....
doc. RNDr. Petr Lánský, CSc.
vedoucí práce

.....
prof. RNDr. Jan Slovák, DrSc.
ředitel Ústavu matematiky a
statistiky

Poděkování

Na tomto místě bych chtěl poděkovat především doc. RNDr. Petru Lánskému, CSc. za jeho odborné vedení a věcnou kritiku. Dále také doc. RNDr. Martinu Kolářovi, Ph.D. za odbornou pomoc při studiu související literatury a přínosnou debatu nad zpracováním dat z pojišťovací praxe. Doc. RNDr. Michalu Veselému, Ph.D. patří velký dík za doporučení a korekturu v oblasti stylizace matematického textu. V neposlední řadě děkuji také společnosti Assicurazioni Generali S.p.A. za svolení k použití interních dat pro aplikační část mé práce a svým kolegům za pomoc při jejich výběru a získání z databáze.

Na tomto místě nesmím a nechci opomenout ani všechny vyučující Masarykovy univerzity a řecké Univerzity v Patře, kteří svými přednáškami přispěli k rozvoji mých znalostí vedoucích až k sepsání této práce. Zároveň děkuji své nejbližší rodině za podporu v průběhu celého mého studia.

Prohlášení

Prohlašuji, že jsem svoji diplomovou práci vypracoval samostatně s využitím informačních zdrojů, které jsou v práci citovány.

Brno 26. dubna 2017

.....
Bc. Martin Bíza

Obsah

Úvod	1
Přehled použitého značení	2
Kapitola 1. Stochastické procesy	4
1.1 Definice a charakteristiky	4
1.2 Markovské řetězce	7
1.2.1 Matice intenzit přechodu	9
1.3 Procesy zrodu a zániku	10
1.3.1 Pravděpodobnosti a intenzity přechodu	11
1.3.2 Limitní pravděpodobnosti	12
1.3.3 Procesy čistého zrodu	13
Kapitola 2. Procesy čistého zrodu	15
2.1 Pravděpodobnosti přechodu	15
2.2 Poissonův proces	18
2.2.1 Nehomogenní Poissonův proces	18
2.2.2 Homogenní Poissonův proces	20
2.3 Procesy s nákazou	22
2.3.1 Pozitivní nákaza	23
2.3.2 Negativní nákaza	24
2.4 Obecná intenzita přechodu	25
2.4.1 Obecný tvar nákazy	26
Kapitola 3. Intervaly mezi událostmi	28
3.1 Procesy obnovy	29
3.1.1 Asymptotické vlastnosti	30
3.1.2 Rovnice obnovy	31
3.2 Rozdělení časů mezi událostmi	32
3.2.1 Číselné charakteristiky rozdělení	33
3.2.2 Exponenciální rozdělení	35
3.2.3 Weibullovo rozdělení	36
3.2.4 Gama rozdělení	38
3.2.5 Lognormální rozdělení	39
3.2.6 Paretovo rozdělení	41

3.3 Testy o shodě rozdělení	43
3.3.1 χ^2 test dobré shody	43
3.3.2 Kolmogorovův-Smirnovův test	44
Kapitola 4. Aplikace	46
4.1 Data	46
4.1.1 Typy pojištění	47
4.2 Příprava dat	48
4.2.1 Analýza sezónnosti	49
4.2.2 Nulové škody	51
4.2.3 Expozice	52
4.3 Zpracování a výsledky	54
4.3.1 Metoda maximální věrohodnosti	56
4.3.2 Metoda minimálního χ^2	58
4.3.3 Srovnání metod	58
Závěr	65
Příloha	66
A Datové podklady pro vybrané grafy čtvrté kapitoly	66
B Modifikovaná metoda maximální věrohodnosti	70
C Zdrojový kód vybraných grafů v softwaru R	71
Seznam použité literatury	83
Seznam obrázků	85
Seznam tabulek	86
Rejstřík	87

Úvod

Rozvoj moderních vědních odvětví a výzkum v nich si už zcela jistě nelze představit bez rozsáhlého matematického aparátu. Teoretické modely nám dávají jednoduchou a intuitivní představu o tom, podle jakého vzoru mohou věci fungovat a situace se v čase vyvíjet. Tyto pokusy o popis a predikci jsou ale vždy – ať už více či méně – pouhým zjednodušením reálného prostředí a v běžném životě tak narážíme na fakt, že rozličné a v modelech nepostižené jevy způsobují nesoulad mezi empirickým pozorováním a výsledky, které nám poskytují výpočty. Tento fakt vede stále více ke snaze popsat pozorované jevy co možná nejdůkladněji a nejvěrněji. Této problematice se věnují stochastické procesy, které budou v diplomové práci zkoumány a které se uplatňují prakticky ve všech přírodovědných oborech, ekonomii, pojišťovnictví, demografii i jinde. Od počátku 20. století se na vystavění teorie stochastických procesů podíleli přední matematici. Za všechny jmenujme Markova, Craméra či Kolmogorova.

Ústředním tématem této práce je modelování výskytů událostí. V matematickém modelu nastání nějakého jevu v budoucnosti může vystupovat mnoho proměnných. V zásadě však můžeme hovořit o dvou typech v závislosti na jejich vztahu k historickým hodnotám zkoumaného procesu. Nejprve uvažme faktory, které explicitně závisí na známé historii náhodného procesu. Určité stavy minulosti a přítomnosti v tomto případě přímo ovlivňují jeho budoucí hodnoty. Kvalitativně odlišným typem jsou faktory, které na historických stavech zkoumaného náhodného procesu nezávisí. Jedná se o externí efekty, které mohou být i samy stochastickými procesy. Budeme-li například v čase sledovat počet aut, která přijedou na čerpací stanici k benzínovému stojanu, bude zřejmě budoucí vývoj tohoto procesu záviset na ceně pohonných hmot. Při zdražení benzínu budou *ceteris paribus* mezi dvěma událostmi pravděpodobně delší intervaly, protože část řidičů s vozem na benzín zvolí alternativní způsob dopravy. Je to příklad externího faktoru, který v modelu explicitně neuvažujeme. Protože jednou z determinantů ceny pohonných hmot je i cena souvisejících komodit obchodovaných na burze, je tento faktor navíc sám stochastickým procesem. Ambicí této práce ale není zkoumání těchto situací, a proto od nich abstrahujeme. Nadto mohou být tyto jevy částečně nebo dokonce zcela zahrnuty v historii daného procesu, a implicitně tak jeho hodnoty přesto ovlivňovat.

V aplikační části této diplomové práce prověříme teoretické poznatky na reálné situaci z pojišťovnictví. Na tuto oblast finančních trhů mají dopad tzv. pojistné události, při jejichž nastání je pojišťovna povinna vyplatit pojištěnému předem smlouvenou výplatu. Výskyt pojistných událostí je náhodný, a proto je zde vhodné využít aparátu stochastických procesů.

Přehled použitého značení

\mathbb{R}	množina reálných čísel
\mathbb{R}^+	množina kladných reálných čísel
\mathbb{R}_0^+	množina nezáporných reálných čísel
\mathbb{R}^n	n -rozměrný prostor reálných čísel
\mathbb{N}	množina přirozených čísel
\mathbb{N}_0	množina přirozených čísel a nuly
(Ω, \mathcal{A})	měřitelný prostor na $\Omega \neq \emptyset$ se systémem podmnožin \mathcal{A}
(Ω, \mathcal{A}, P)	pravděpodobnostní prostor s mírou pravděpodobnosti P
$P(X = x)$	pravděpodobnost, že náhodná veličina X nabude hodnoty x
$P(X = x Y = y)$	pravděpodobnost, že $X = x$ za podmínky $Y = y$
$\{X_t(\omega)\}, \{X_t\}$	stochastický proces
$\{N_t(\omega)\}, \{N_t\}$	čítací proces
(S, \mathcal{S})	stavový prostor na $S \neq \emptyset$ se systémem podmnožin \mathcal{S}
$\gamma(s, t), \tilde{\gamma}(t)$	autokovarianční funkce
$\rho(s, t), \tilde{\rho}(t)$	autokorelační funkce
EX	střední hodnota náhodné veličiny X
EX^k	k -tý obecný moment náhodné veličiny X
DX	rozptyl náhodné veličiny X
$\psi_X(t)$	charakteristická funkce náhodné veličiny X
$Exp(\lambda)$	exponenciální rozdělení s parametrem λ
$Po(\lambda)$	Poissonovo rozdělení s parametrem λ
$Wb(k, \lambda)$	Weibullovo rozdělení s parametry k a λ
$\Gamma(k, \lambda)$	gama rozdělení s parametry k a λ
$LN(\mu, \sigma^2)$	lognormální rozdělení s parametry μ a σ^2
$Pa(\sigma, \alpha)$	Paretovo rozdělení druhého řádu s parametry σ a α
$MN(n; p_1, \dots, p_k)$	multinomické rozdělení s parametry n a p_1, \dots, p_k
$N(0, 1)$	standardizované normální rozdělení
$\chi^2(v)$	χ^2 rozdělení s v stupni volnosti
$D(n)$	rozdělení jednovýběrové Kolmogorovovy-Smirnovovy statistiky s n stupni volnosti

$D(m, n)$	rozdělení dvouvýběrové Kolmogorovovy-Smirnovovy statistiky s m, n stupni volnosti
$\mathcal{F}_\alpha(\mathbf{p})$	α -kvantil rozdělení \mathcal{F} s parametry \mathbf{p}
$\Phi(x)$	hodnota distribuční funkce $N(0, 1)$ v bodě x
$F * G$	konvoluce funkcí F a G
F^{*n}	n -tá konvoluce funkce F
\hat{p}	maximálně věrohodný odhad parametru p
$\Gamma(k)$	gama funkce
$\gamma(k, x)$	dolní neúplná gama funkce
$\psi(k)$	digama funkce
$\binom{n}{k}$	kombinační číslo n nad k
i	imaginární jednotka
\mathbf{p}	sloupcový vektor
\mathbf{p}^T	řádkový vektor
\mathbf{P}	dvourozměrná matice

Kapitola 1

Stochastické procesy

Stochastické procesy jsou dynamickou součástí teorie pravděpodobnosti a zobecněním pojmu náhodné veličiny. Setkáváme se s nimi všude tam, kde je vývoj nějakého procesu v čase ovlivněn zákony pravděpodobnosti a není zcela deterministický. Nachází uplatnění v různorodých vědních disciplínách. Jsou důležitým aparátem pro popis procesů ve fyzice, demografii, ekonomii i v dalších oborech. Ekvivalentně můžeme použít označení náhodné procesy.

V této kapitole zdefinujeme stochastický proces, popíšeme jeho základní charakteristiky a zaměříme se na speciální případy, které využijeme v dalším textu. Čerpat budeme především ze zdrojů [14] a [15]. Ke studiu náhodných veličin, ze kterých stochastické procesy vychází, lze využít např. [1, kapitoly 1 a 3] nebo [16, kapitola 2]. Jejich základní číselné charakteristiky budou popsány v odstavci 3.2.

1.1 Definice a charakteristiky

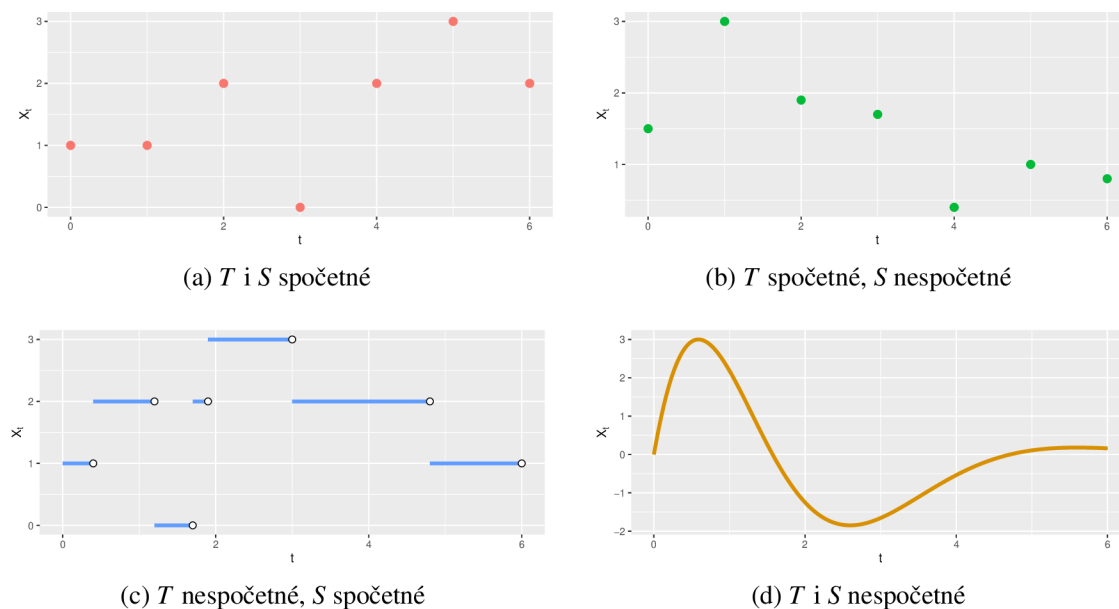
Definice 1.1. Mějme pravděpodobnostní prostor (Ω, \mathcal{A}, P) a neprázdnou indexovou množinu $T \subseteq \mathbb{R}$. Pak *Stochastický proces* je systém náhodných veličin $\{X_t(\omega); \omega \in \Omega, t \in T\}$ definovaných na (Ω, \mathcal{A}, P) .

Pokud bude z textu jasné, co se množinami Ω a T myslí, budeme pro jednoduchost náhodný proces značit jen $\{X_t\}$. V závislosti na mohutnosti T mluvíme o procesu s diskrétním časem, pokud je T spočetné, nebo se spojitým časem, pokud je T nespočetné. Na množinu T lze nahlížet jako na plynutí času. V takovém případě mluvíme o X_t jako o hodnotě náhodného procesu $\{X_t\}$ v čase t .

Definice 1.2. Dvojice (S, \mathcal{S}) , kde S je množina hodnot náhodných veličin X_t a \mathcal{S} je σ -algebra na S , se nazývá *stavový prostor* procesu $\{X_t\}$.

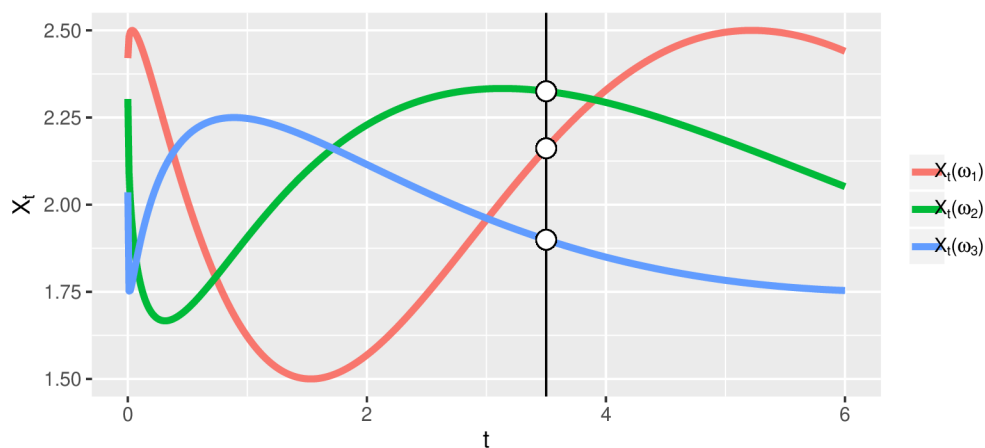
Pokud náhodné veličiny X_t nabývají pouze diskrétních hodnot, hovoříme o procesu s diskrétními stavy, nabývají-li libovolných hodnot z nedegenerovaného intervalu, mluvíme

o procesu se spojitými stavy. Obrázek 1.1 ukazuje, jak může vypadat stochastický proces pro různé kombinace mohutností množin T a S .



Obrázek 1.1: Příklady trajektorií stochastických procesů

Náhodný proces lze chápat jako funkci dvou proměnných ω a t . Pro pevné $\omega \in \Omega$ je $X_t(\cdot)$ reálnou funkcí proměnné t , kterou označujeme jako *trajektorii* procesu $\{X_t\}$. Pro pevné $t \in T$ je $X_{t,\cdot}(\omega)$ náhodná veličina. Obrázek 1.2 ilustruje souvislost mezi těmito pojmy.



Obrázek 1.2: Trajektorie a náhodná veličina

Stejně jako náhodná veličina, tak i náhodný proces je vybaven distribuční funkcí. V tomto případě však mluvíme o celém systému distribučních funkcí.

Definice 1.3. Mějme $n \in \mathbb{N}$, $(t_1, \dots, t_n) \in T^n$ a $(x_1, \dots, x_n) \in \mathbb{R}^n$. Distribuční funkcí náhodného vektoru $(X_{t_1}, \dots, X_{t_n})$ rozumíme funkci

$$F_{t_1, \dots, t_n}(x_1, \dots, x_n) = P(X_{t_1} \leq x_1, \dots, X_{t_n} \leq x_n).$$

Pro libovolné $n \in \mathbb{N}$ a $(t_1, \dots, t_n) \in T^n$ dostáváme *systém distribučních funkcí* náhodného procesu $\{X_t\}$. Uvažujme dále dvě podmínky:

a) pro libovolnou permutaci i_1, \dots, i_n čísel $1, \dots, n$ platí

$$F_{t_{i_1}, \dots, t_{i_n}}(x_{i_1}, \dots, x_{i_n}) = F_{t_1, \dots, t_n}(x_1, \dots, x_n);$$

b) pro $t_{n+1} \in T$ a $x_{n+1} \rightarrow \infty$ platí

$$F_{t_1, \dots, t_n, t_{n+1}}(x_1, \dots, x_n, x_{n+1}) = F_{t_1, \dots, t_n}(x_1, \dots, x_n).$$

Platí-li tyto podmínky, pak o systému distribučních funkcí $\{F_{t_1, \dots, t_n}(x_1, \dots, x_n)\}$ mluvíme jako o *konzistentním*. Podle [15, strana 8] umíme ke každému náhodnému procesu najít konzistentní systém distribučních funkcí. Zároveň podle Kolmogorovovy věty [tamtéž] platí i opačné tvrzení.

Definice 1.4. Nechť $\{X_t\}$ je takový proces, že pro každé $t \in T$ existuje střední hodnota EX_t . Potom funkce $\mu_t = EX_t$ definovaná na T se nazývá *střední hodnota* procesu $\{X_t\}$.

O procesu, pro který platí $\mu_t = \mu$, říkáme, že je *stacionární* ve střední hodnotě. Pokud navíc $\mu_t \equiv 0$ pak je tento proces *centrovaný*.

Definice 1.5. Nechť pro náhodný proces $\{X_t\}$ platí $E|X_t|^2$ pro všechna $t \in T$. Potom funkce dvou proměnných definovaná na $T \times T$ předpisem

$$\gamma(s, t) = E(X_s - \mu_s)(X_t - \mu_t)$$

se nazývá *autokovarianční funkce* procesu $\{X_t\}$. Hodnota $\sigma_t^2 = \gamma(t, t)$ je *rozptyl* procesu $\{X_t\}$ v čase t .

Pokud pro všechna $s, t \in T$ platí, že $\gamma(s, t) = \gamma(0, |s - t|)$, pak říkáme, že je proces $\{X_t\}$ *stacionární* v autokovarianci. Autokovarianční funkci potom zjednodušeně zapisujeme jako $\gamma(s, t) = \tilde{\gamma}(s - t)$. *Autokorelační funkci* náhodného procesu definujeme obdobně jako korelaci dvou náhodných veličin. Je dána vztahem

$$\rho(s, t) = \frac{\gamma(s, t)}{\sqrt{\gamma(s, s)\gamma(t, t)}}.$$

Pro náhodný proces stacionární v autokovarianci pak z předchozího máme

$$\rho(s, t) = \tilde{\rho}(s - t) = \frac{\tilde{\gamma}(s - t)}{\tilde{\gamma}(0)}.$$

Důležitým pojmem při studiu stochastických procesů je stacionarita.

Definice 1.6. Řekneme, že náhodný proces $\{X_t\}$ je *striktně stacionární*, jestliže pro libovolné $n \in \mathbb{N}$, libovolná $(x_1, \dots, x_n) \in \mathbb{R}^n$ a libovolná $(t_1, \dots, t_n) \in T^n$ a h tak, že $t_i, t_i + h \in T$ platí

$$F_{t_1, \dots, t_n}(x_1, \dots, x_n) = F_{t_1 + h, \dots, t_n + h}(x_1, \dots, x_n).$$

Striktní stacionarita je velmi silná podmínka. Proto se často setkáváme s pojmem *slabé stacionarity*, který je definován pro procesy s konečnými druhými momenty (též procesy druhého řádu). Náhodný proces je slabě stacionární právě tehdy, když je stacionární ve střední hodnotě a v autokovarianci.

Definice stochastických procesů jen velmi málo omezuje škálu vlastností, které tyto mohou mít. Některé z nich usnadňují práci při odvozování statistických vlastností a dává smysl je předpokládat u procesů v mnohých praktických úlohách. Dobrým příkladem je *stacionarita přírůstků* náhodného procesu. Přírůstek $X_t - X_s$ je pro $s, t \in T$ a $t > s$ stacionární, pokud jeho výše závisí pouze na délce intervalu $t - s$. Pravděpodobnostní rozdělení přírůstku o dané délce tedy nezávisí na časovém okamžiku, do kterého jej umístíme. Náhodný proces má navíc přírůstky *nezávislé*, pokud jsou na sobě nezávislé přírůstky v libovolných dvou disjunktních intervalech. V tomto případě tedy velikost přírůstku v jednom časovém úseku nemá vliv na přírůstek v časovém úseku jiném.

Nebude-li explicitně uvedeno jinak, budeme v dalším textu předpokládat, že množina T symbolizuje čas. Zároveň chceme, aby měl počátek, proto $t \geq 0$.

1.2 Markovské řetězce

Ze stochastických procesů vydělujeme část, do které spadají procesy, jejichž budoucí stavy závisí pouze na přítomnosti a znalost minulosti neposkytuje žádnou dodatečnou informaci.

Definice 1.7. Řekneme, že náhodný proces $\{X_t, t \geq 0\}$ definovaný na pravděpodobnostním prostoru (Ω, \mathcal{A}, P) je *Markovův řetězec* se spočetnou množinou stavů S a spojitým časem, jestliže

$$P(X_t = k + n \mid X_s = k, X_{s_1} = k_1, \dots, X_{s_p} = k_p) = P(X_t = k + n \mid X_s = k) \quad (1.1)$$

pro všechna $k + n, k, k_1, \dots, k_p \in S$ a $s, s_1, \dots, s_p \leq t$, pro která $P(X_s = k, X_{s_1} = k_1, \dots, X_{s_p} = k_p) > 0$.

Identitu (1.1) označujeme jako *markovskou vlastnost* (též Markovovu) a o procesech, které ji splňují hovoříme jako o *markovských*. Lze tedy říci, že v případě těchto procesů nezáleží na tom, jestli máme k dispozici přehled všech stavů od počátku až do současnosti nebo pouze nejaktuálnější stav. Poznamenejme, že markovská vlastnost je slabší než nezávislost přírůstků, protože dovoluje historickému vývoji ovlivňovat budoucí stavy procesu prostřednictvím přítomnosti. Současně ji ale připouští jako jedinou formu vlivu.

V této práci budeme zkoumat především výskyt událostí, které budeme počítat a které mohou nastat v libovolném okamžiku. Z tohoto důvodu se omezujeme na definici markovského řetězce pouze pro spočetnou množinu stavů a spojitý čas. Definici pro diskrétní případ najdeme např. v [15, strana 15].

U Markovských procesů nás zajímají především *přechodové pravděpodobnosti*. Říkají, s jakou pravděpodobností nabude přírůstek dané hodnoty za jistý časový interval. Označme $p_{k,k+n}(s,t)$ jako pravděpodobnost, že proces mezi časy s a t přejde ze stavu k do stavu $k+n$. Díky markovské vlastnosti procesu definujeme

$$p_{k,k+n}(s,t) = P(X_t - X_s = n \mid X_s = k), \quad (1.2)$$

přičemž $0 \leq s \leq t < \infty$ a $k, k+n \in S$. Z definice 1.7 musí tato přechodová pravděpodobnost existovat pro každé $k+n \in S$. Obecně může záviset jak na s , tak na t . V případě, že závisí pouze na jejich rozdílu $t-s$, je příslušný proces *homogenní*. Za těchto okolností budeme pro jednoduchost značit $p_{k,k+n}(s,t) = p_{k,k+n}(t-s)$. Vlastnost homogenity se podobá stacionaritě přírůstků, ale není s ní ekvivalentní. Stacionarita se týká nepodmíněné pravděpodobnosti přírůstků, zatímco homogenita pravděpodobnosti podmíněné počátečním stavem procesu. Jako příklad uveďme Polyův proces definovaný v [13, strana 158], který má stacionární přírůstky, ale přesto je nehomogenní [13, strana 160].

Bez újmy na obecnosti zvolme množinu stavů S tak, aby $X_0 = 0$. Označme

$$p_n(t) = p_{0,n}(0,t) = P(X_t = n) \quad (1.3)$$

pro $n \in S$ jako pravděpodobnostní rozdělení procesu $\{X_t\}$ v čase t . Z věty o celkové pravděpodobnosti máme pro rozdělení přírůstku vztah

$$P(X_t - X_s = n) = \sum_{k \in S} P(X_t - X_s = n \mid X_s = k) P(X_s = k) = \sum_{k \in S} p_{k,k+n}(s,t) p_k(s). \quad (1.4)$$

Pokud má proces nezávislé přírůstky, pak musí být nezávislé zejména $X_t - X_s$ a $X_s - X_0 = X_s$. Potom

$$P(X_t - X_s = n) = P(X_t - X_s = n \mid X_s = 0) = p_{0,n}(s,t).$$

Toto je nutná podmínka nezávislosti přírůstků procesu. Není to ale podmínka postačující, jelikož jsme uvažovali pouze dva konkrétní přírůstky.

Pro $s = 0$ dostáváme z rovnice (1.4) vztah

$$P(X_t = n) = \sum_{k \in S} p_{k,k+n}(0,t) p_k(0),$$

což v případě homogenity $\{X_t\}$ můžeme přepsat do tvaru

$$p_n(t) = \sum_{k \in S} p_{k,k+n}(t) p_k(0).$$

Sestavme čtvercovou matici $\mathbf{P}(t) = \{p_{k,k+n}(t); k, k+n \in S\}$ a vektor $\mathbf{p}(t) = \{p_k(t); k \in S\}$. Maticově lze nyní předchozí vztah vyjádřit jako

$$\mathbf{p}(t)^T = \mathbf{p}(0)^T \mathbf{P}(t),$$

kde $\mathbf{p}(0)$ nazýváme vektorem počátečních pravděpodobností a $\mathbf{P}(t)$ maticí intenzit přechodu homogenního markovského řetězce. Navíc např. v [15, strana 72] nalezneme, že platí

$$\mathbf{P}(s+t) = \mathbf{P}(s)\mathbf{P}(t).$$

Z poslední rovnice je zřejmě vhodné dodefinovat $\mathbf{P}(0) = \mathbf{I}$.

1.2.1 Matice intenzit přechodu

Pro homogenní Markovovy řetězce zavádíme také matici intenzit přechodu. Její prvky jsou úzce spjaty s průměrnou pravděpodobností přechodu mezi dvěma stavy. Číslo $p_{k,k+n}(h)/h$ pro $k, k+n \in S$ označuje průměrnou pravděpodobnost přechodu ze stavu k do stavu $k+n$ za čas h a číslo $(1 - p_{k,k}(h))/h$ odpovídá průměrné pravděpodobnosti výstupu ze stavu k za čas h . Limitním přechodem pro $h \rightarrow 0^+$ dostáváme *intenzitu přechodu* z k do $k+n$

$$q_{k,k+n} = \lim_{h \rightarrow 0^+} \frac{p_{k,k+n}(h)}{h} \quad (1.5)$$

a *intenzitu výstupu* z k

$$q_k = \lim_{h \rightarrow 0^+} \frac{1 - p_{k,k}(h)}{h} \quad (1.6)$$

Intenzitě výstupu se někdy říká také celková intenzita. Podle [12, kap. 14, věty 1.1 a 1.2] obě tyto limity existují.

Definice 1.8. Mějme nezáporná čísla $q_{k,k+n}$ resp. q_k definovaná vztahy (1.5) resp. (1.6). Pak matice $\mathbf{Q} = \{q_{k,k+n}; k, k+n \in S\}$, kde $q_{k,k} = -q_k$, se nazývá *matice intenzit přechodu* homogenního Markovova řetězce.

Protože pro libovolné $k \in S$ a $n \neq 0$ platí $q_k = \sum_{k+n \in S} q_{k,k+n}$, je matice intenzit přechodu kvazistochastická.

Definice 1.9. Nechť $\{X_t\}$ je homogenní Markovův řetězec se spojitým časem, množinou stavů S a maticemi pravděpodobností přechodu $\mathbf{P}(t)$. Potom vektor $\pi = \{0 \leq \pi_k; k \in S, \sum_{k \in S} \pi_k = 1\}$ takový, že

$$\pi^T = \pi^T \mathbf{P}(t)$$

pro libovolné $t \geq 0$, se nazývá *stacionární rozdělení* řetězce $\{X_t\}$.

Nalezení stacionárního vektoru přímo z definice 1.9 s sebou může nést výpočetní komplikace. Proto se většinou volí kalkulace pomocí matice intenzit přechodu. Za splnění dosti obecných podmínek uvedených např. v [15, strana 92] lze stacionární rozdělení nalézt řešením rovnice

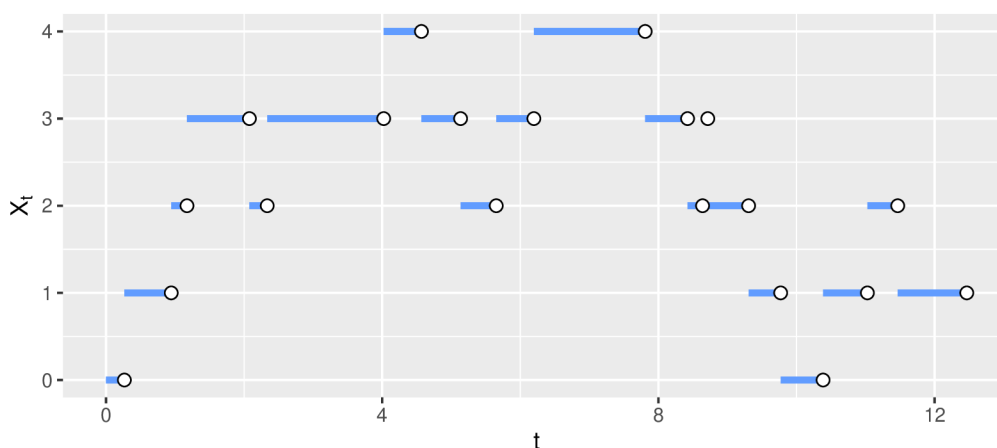
$$\mathbf{0}^T = \pi^T \mathbf{Q}.$$

Další výklad k Markovovým řetězcům a jejich vlastnostem najdeme např. v [7, kapitola 6] nebo v [11, kapitoly 2–4].

1.3 Procesy zrodu a zániku

Do této chvíle jsme podle definice 1.7 pro množinu stavů S požadovali pouze to, aby byla spočetná. Už samotný název této podkapitoly napovídá, že jde o terminologii z biologie, kde se procesy zrodu a zániku používají jako matematické modely vývoje počtu jedinců v populaci. Tento počet přirozeně nemůže být záporný. Proto se budeme dále věnovat pouze takovým procesům, pro které platí $S = \mathbb{N}_0$. Všimněme si, že tento požadavek nic neubírá na obecnosti procesu, protože každou spočetnou množinu lze přeznačit na \mathbb{N}_0 .

U těchto procesů rozlišujeme dva jevy. Při události odpovídající zrodu se posouváme ze stavu k do stavu $k+1$ a analogicky při události odpovídající zániku z k do $k-1$. V případě $k=0$ můžeme v tomto stavu pouze setrvat nebo se posunout do stavu $k=1$. Na obrázku 1.3 vidíme možnou trajektorii procesu zrodu a zániku.



Obrázek 1.3: Trajektorie procesu zrodu a zániku

1.3.1 Pravděpodobnosti a intenzity přechodu

V procesech zrodu a zániku klademe na pravděpodobnosti přechodu (1.2) další požadavky. Pro velmi malý časový interval délky $h > 0$ předpokládáme, že

1. pravděpodobnost výskytu zrodu nebo zániku je přibližně proporcionální k h a
2. pravděpodobnost nastání libovolných dvou nerozlišitelných jevů za h je téměř nulová.

Použití slovních spojení „přibližná proporcionalita“ a „téměř nulový“ je však pro další práci příliš vágní a musíme je pečlivěji vysvětlit.

Definice 1.10. Řekneme, že libovolnou funkci f jedné reálné proměnné h můžeme označit symbolem $o(h)$, pokud platí

$$\lim_{h \rightarrow 0^+} \frac{f(h)}{h} = 0.$$

Znamená to tedy, že funkce f jde pro $h \rightarrow 0^+$ k nule rychleji než funkce lineární. Poznamenejme, že $o(h)$ je spíše označení funkce s výše uvedenou vlastností než funkce jako taková. V důsledku toho je výraz $co(h) + do(h) = o(h)$ korektní pro libovolná $c, d \in \mathbb{R}$, přestože za normálních okolností by platil pouze pro vybrané uspořádané dvojice $[c, d]$. Jinými slovy tento vztah říká, že lineární kombinace dvou funkcí s vlastností $o(h)$ má opět tuto vlastnost.

Předpokládejme, že se nacházíme v čase t a ve stavu k , tj. $X_t = k$. Chceme pomocí výše uvedených předpokladů a s využitím notace $o(h)$ zapsat pravděpodobnost, že do času $t+h$ pro velmi malou hodnotu h nastane právě jedna událost zrodu resp. zániku. Vzhledem k zažitému značení můžeme psát

$$p_{k,k+1}(t, t+h) = \lambda_k(t)h + o(h)$$

pro $k \in \mathbb{N}_0$ resp.

$$p_{k,k-1}(t, t+h) = \mu_k(t)h + o(h)$$

pro $k \in \mathbb{N}$. Dodefinujme $\mu_0(t) = 0$. Protože koeficient proporcionality¹ závisí pouze na přítomnosti je tento proces Markovský. Nemusí však obecně být homogenní. Abychom zajistili, že výskyt více než jednoho zrodu resp. zániku za čas h je velmi nepravděpodobný², doplňme

$$p_{k,k+n}(t, t+h) = o(h)$$

pro $k \in \mathbb{N}_0$ a $|n| \geq 2$. Z vlastnosti notace $o(h)$ a z vlastností pravděpodobnosti zůstává

$$p_{k,k}(t, t+h) = 1 - (\lambda_k(t) + \mu_k(t))h + o(h)$$

pro $k \in \mathbb{N}_0$.

Závěrem tohoto odstavce ještě doplňme matici intenzit přechodu pro homogenní procesy zrodu a zániku.

$$\mathbf{Q} = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \dots \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & 0 & \dots \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & \ddots \\ \vdots & \vdots & \ddots & \ddots & \ddots \end{pmatrix} \quad (1.7)$$

Všimněme si, že funkce $\lambda_0, \lambda_1, \dots$ a μ_1, μ_2, \dots jsou intenzitami přechodu z (1.5).

1.3.2 Limitní pravděpodobnosti

Podle [15, strana 111] můžeme limitní pravděpodobnosti $\pi_k = \lim_{t \rightarrow \infty} p_k(t)$ homogenních řetězců hledat jako řešení soustavy $\pi^T \mathbf{Q} = \mathbf{0}^T$, které vyhovuje podmínkám $\pi_k \geq 0$ a $\sum_{k \in \mathbb{N}_0} \pi_k = 1$.

Uvažujme $k \geq 1$. Po rozepsání dostáváme

$$\begin{aligned} -\lambda_0 \pi_0 + \mu_1 \pi_1 &= 0, \\ \lambda_{k-1} \pi_{k-1} - (\lambda_k + \mu_k) \pi_k + \mu_{k+1} \pi_{k+1} &= 0. \end{aligned}$$

¹Vzhledem k přítomnosti členu $o(h)$ můžeme přesněji hovořit o přibližné proporcionalitě

²Zde ekvivalentní k použitému pojmu "pravděpodobnost je téměř nulová"

Položíme-li

$$K_k = \mu_k \pi_k - \lambda_{k-1} \pi_{k-1},$$

můžeme soustavu přepsat do tvaru

$$\begin{aligned} K_1 &= 0, \\ K_{k+1} - K_k &= 0, \end{aligned}$$

z čehož zřejmě plyne, že $K_k = 0$. Díky tomu pišme

$$\pi_k = \frac{\lambda_{k-1}}{\mu_k} \pi_{k-1} = \frac{\lambda_{k-1} \lambda_{k-2}}{\mu_k \mu_{k-1}} \pi_{k-2} = \dots = \frac{\lambda_{k-1} \lambda_{k-2} \dots \lambda_0}{\mu_k \mu_{k-1} \dots \mu_1} \pi_0 = \rho_k \pi_0,$$

přičemž jsme použili označení

$$\rho_k = \frac{\lambda_{k-1} \lambda_{k-2} \dots \lambda_0}{\mu_k \mu_{k-1} \dots \mu_1}.$$

Dodefinujme ještě $\rho_0 = 1$. Aby $\sum_{k \in \mathbb{N}_0} \pi_k = 1$ dostáváme konečně

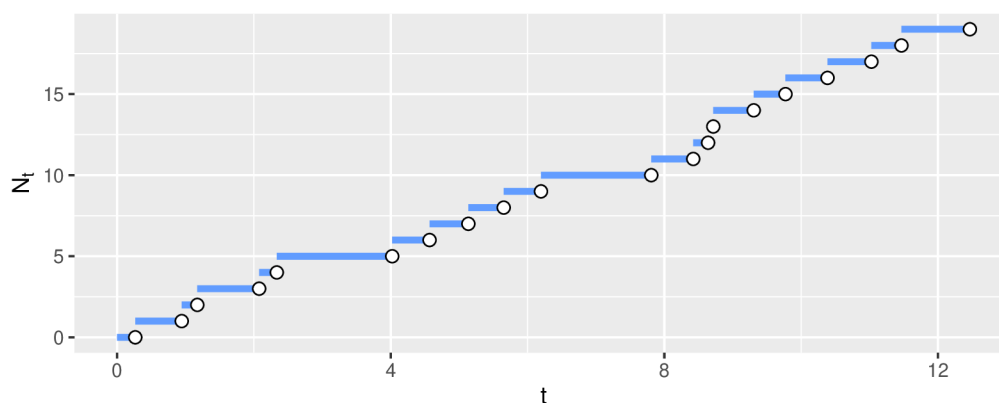
$$\pi_k = \rho_k \left(\sum_{j=0}^{\infty} \rho_j \right)^{-1}$$

pro $k \geq 0$.

1.3.3 Procesy čistého zrodu

Procesy čistého zrodu jsou speciálním případem procesů popsaných v předchozím odstavci pro intenzitu zániku $\mu_k(t) \equiv 0$ pro všechna $k \in \mathbb{N}$. Opustíme-li biologickou terminologii, můžeme říct, že vnímáme pouze jednu rozlišitelnou událost, která nastává s intenzitou $\lambda_k(t)$, kde t je aktuální čas a k aktuální stav systému. Na obrázku 1.4 vidíme možnou trajektorii procesu čistého zrodu.

Procesy, jejichž stav udává počet výskytů nějaké události za zvolený časový interval, označujeme jako *čítací*. Proces čistého zrodu tuto podmínku splňuje. Jelikož je počet výskytů událostí stěžejním tématem této práce, budeme dále pro odlišení od ostatních náhodných procesů čítací proces v čase t označovat symbolem $\{N_t\}$. Stejně jako dosud $N_0 = 0$. Uvažme dva časy s a t , kde $s < t$. Snadno lze nahlédnout, že pro čítací procesy



Obrázek 1.4: Trajektorie procesu čistého zrodu

platí za této podmínky nerovnost $N_s \leq N_t$.

Přechodové pravděpodobnosti procesu čistého zrodu jsou pro $h \rightarrow 0$ tvaru

$$p_{k,k+n}(t,t+h) = \begin{cases} 0 & n \leq -1; \\ 1 - \lambda_k(t)h + o(h) & n = 0; \\ \lambda_k(t)h + o(h) & n = 1; \\ o(h) & n \geq 2. \end{cases} \quad (1.8)$$

Z podmínky $\mu_k(t) \equiv 0$ a matice (1.7) vyplývá, že matice intenzit přechodu procesu čistého zrodu je tvaru

$$\mathbf{Q} = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \cdots \\ 0 & -\lambda_1 & \lambda_1 & 0 & \cdots \\ 0 & 0 & -\lambda_2 & \lambda_2 & \ddots \\ \vdots & \vdots & \ddots & \ddots & \ddots \end{pmatrix}.$$

Kapitola 2

Procesy čistého zrodu

V rámci této kapitoly se zaměříme na odvození pravděpodobností přechodu procesů čistého zrodu. Důkladněji rozebereme speciální případy, kdy intenzita $\lambda_k(t)$ závisí pouze na stavu k , nebo čase t , nebo je pro libovolnou dvojici $[k, t]$ konstantní. V této části vycházíme především ze zdrojů [13] a [16].

2.1 Pravděpodobnosti přechodu

Z odstavce 1.3.3 víme, jak se přechodové pravděpodobnosti procesu čistého zrodu chovají ve velmi krátkých časových intervalech. V aplikacích však nejčastěji potřebujeme jejich popis na intervalech libovolné délky. Musíme tedy odvodit pravděpodobnosti přechodu $p_{k,k+n}(s, t)$ z (1.2).

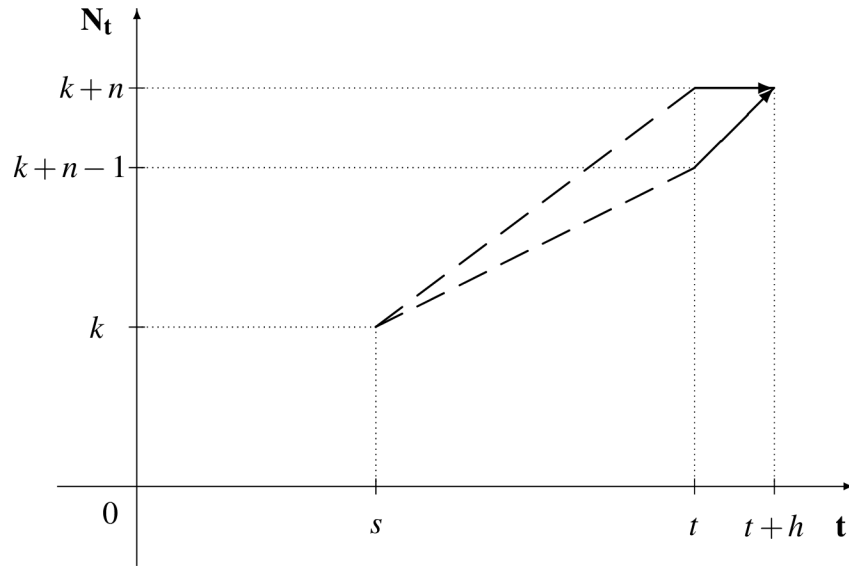
Mějme proces čistého zrodu $\{N_t\}$, pro který víme, že $N_s = k$ a $N_{t+h} = k + n$. Máme tedy přírůstek o velikosti n na intervalu $(s, t + h)$. Rozdělme jej v bodě t na dva disjunktní podintervaly (s, t) a $(t, t + h)$. Díváme se nejdříve na výsledek procesu a jeho pravděpodobnost v čase t a poté pro fixovaný stav N_t můžeme díky znalosti N_s a N_{t+h} dopočítat i přírůstek na $(t, t + h)$. Pro tento přírůstek navíc známe přechodové pravděpodobnosti (1.8). Celá situace je schematicky znázorněna na obrázku 2.1.

Díky nezávislosti přírůstků a z věty o celkové pravděpodobnosti platí *Chapmanova-Kolmogorova rovnice*

$$p_{k,k+n}(s, t+h) = \sum_{j=0}^n p_{k,k+j}(s, t) p_{k+j,k+n}(t, t+h).$$

Protože pro $j \leq n - 2$ je z (1.8) suma na pravé straně rovna $o(h)$, můžeme psát

$$p_{k,k+n}(s, t+h) = p_{k,k+n-1}(s, t) \lambda_{k+n-1}(t) h + p_{k,k+n}(s, t) [1 - \lambda_{k+n}(t) h] + o(h).$$

Obrázek 2.1: Rozdělení intervalu $(s, t+h)$ v bodě t

Jednoduchou úpravou dostáváme rovnici

$$\frac{p_{k,k+n}(s, t+h) - p_{k,k+n}(s, t)}{h} = p_{k,k+n-1}(s, t)\lambda_{k+n-1}(t) - p_{k,k+n}(s, t)\lambda_{k+n}(t) + \frac{o(h)}{h}$$

a limitním přechodem pro $h \rightarrow 0^+$ máme konečně

$$\frac{d}{dt}p_{k,k+n}(s, t) = p_{k,k+n-1}(s, t)\lambda_{k+n-1}(t) - p_{k,k+n}(s, t)\lambda_{k+n}(t). \quad (2.1)$$

Určení přechodových pravděpodobností odpovídá vyřešení nehomogenní lineární diferenciální rovnice prvního řádu (2.1). Pro $n = 0$ máme

$$\begin{aligned} \frac{d}{dt}p_{k,k}(s, t) &= \underbrace{p_{k,k-1}(s, t)}_0 \lambda_{k-1}(t) - p_{k,k}(s, t)\lambda_k(t), \\ \frac{d}{dt}p_{k,k}(s, t) &= -p_{k,k}(s, t)\lambda_k(t). \end{aligned}$$

Přeznačíme $y = t$ a řešíme získanou homogenní rovnici metodou integračního faktoru. Postupně píšme

$$\begin{aligned} \frac{d}{dy}p_{k,k}(s, y) \exp \left\{ \int_s^y \lambda_k(x) dx \right\} + p_{k,k}(s, y) \lambda_k(y) \exp \left\{ \int_s^y \lambda_k(x) dx \right\} &= 0, \\ \frac{d}{dy} \left(p_{k,k}(s, y) \exp \left\{ \int_s^y \lambda_k(x) dx \right\} \right) &= 0. \end{aligned}$$

Po integraci podle y v mezích od s do t dostáváme

$$p_{k,k}(s,t) \exp \left\{ \int_s^t \lambda_k(x) dx \right\} - p_{k,k}(s,s) \exp \left\{ \int_s^s \lambda_k(x) dx \right\} = c$$

pro libovolné reálné c . Protože pro počáteční podmínky zřejmě platí

$$\begin{aligned} p_{k,k}(s,s) &= 1 \quad \text{a} \\ p_{k,k+n}(s,s) &= 0 \quad \text{pro } n \in \mathbb{N}, \end{aligned} \tag{2.2}$$

můžeme dále psát

$$p_{k,k}(s,t) \exp \left\{ \int_s^t \lambda_k(x) dx \right\} - 1 = c.$$

Pro $t = s$ dostáváme z (2.2) integrační konstantu $c = 0$, a proto konečně

$$p_{k,k}(s,t) = \exp \left\{ - \int_s^t \lambda_k(x) dx \right\}. \tag{2.3}$$

Řešme nyní rovnici (2.1) pro $n \in \mathbb{N}$. Opět provedeme přeznačení y za t a k řešení použijeme metodu integračního faktoru ve tvaru $\exp \left\{ \int_s^y \lambda_{k+n}(x) dx \right\}$. Dostáváme rovnici

$$\begin{aligned} \frac{d}{dy} \left(p_{k,k+n}(s,y) \exp \left\{ \int_s^y \lambda_{k+n}(x) dx \right\} \right) &= \\ &= p_{k,k+n-1}(s,y) \lambda_{k+n-1}(y) \exp \left\{ \int_s^y \lambda_{k+n}(x) dx \right\}, \end{aligned}$$

kteřá oproti předchozímu případu již není homogenní. Označme nehomogenní část jako $I(y)$ a integrujme obě strany rovnice podle y od s do t . Obdržíme

$$p_{k,k+n}(s,t) \exp \left\{ \int_s^t \lambda_{k+n}(x) dx \right\} - \underbrace{p_{k,k+n}(s,s)}_0 \exp \left\{ \underbrace{\int_s^s \lambda_{k+n}(x) dx}_0 \right\} = \int_s^t I(y) dy + c,$$

přičemž $c \in \mathbb{R}$, a tedy

$$p_{k,k+n}(s,t) \exp \left\{ \int_s^t \lambda_{k+n}(x) dx \right\} = \int_s^t I(y) dy + c.$$

Pro $t = s$ dostáváme z počátečních podmínek (2.2) opět $c = 0$. Navíc, schématicky zapsáno, máme $\int_s^t = - \int_t^s$. Výsledná rekurentní rovnice je proto tvaru

$$p_{k,k+n}(s,t) = \int_s^t \lambda_{k+n-1}(y) \exp \left\{ - \int_y^t \lambda_{k+n}(x) dx \right\} p_{k,k+n-1}(s,y) dy. \tag{2.4}$$

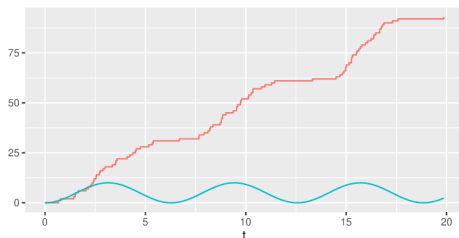
Ukazuje se, že rovnici (2.4) není možné pro $p_{k,k+n}(s,t)$ explicitně vyřešit pro libovolnou posloupnost intenzit přechodu, viz [13, strana 123]. Je také nutno poznamenat, že ne každá posloupnost intenzit přechodu vede k přípustnému výsledku. Předně musí být $\sum_{n=0}^{\infty} p_{k,k+n}(s,t) = 1$. Platnost této podmínky je třeba zajistit vhodnou volbou. Podle [15, strana 60] je tato rovnost splněna, pokud

$$\sum_{k=0}^{\infty} \left[\max_{t \geq 0} \lambda_k(t) \right]^{-1} = \infty. \quad (2.5)$$

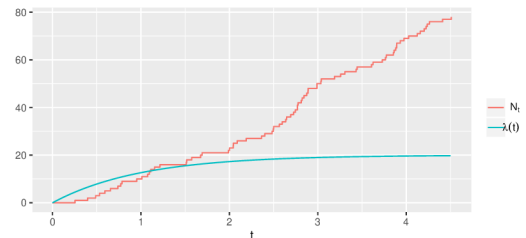
2.2 Poissonův proces

2.2.1 Nehomogenní Poissonův proces

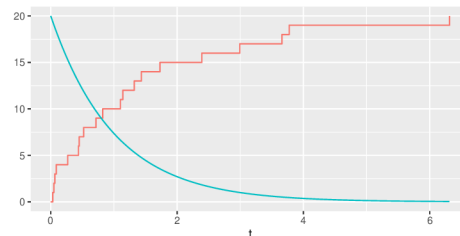
Naštěstí pro důležité a poměrně obecné případy lze explicitní řešení rovnic (2.3) a (2.4) nalézt. Uvažujme situaci, kdy přechodové pravděpodobnosti nezávisí na aktuálním stavu systému. Pro funkce intenzit přechodu to znamená, že pro libovolné nezáporné celé k platí $\lambda_k(t) = \lambda(t)$. Takovému procesu říkáme *nehomogenní Poissonův proces*. Na obrázku 2.2 vidíme jeho možné trajektorie pro různé volby $\lambda(t)$.



(a) $\lambda(t) = 5(1 - \cos(t))$



(b) $\lambda(t) = 20(1 - e^{-x})$



(c) $\lambda(t) = 20e^{-x}$

Obrázek 2.2: Trajektorie nehomogenního Poissonova procesu

Pro jednodušší zápis bez újmy na obecnosti nejprve označme $p_n(s, t) := p_{k, k+n}(s, t)$. Pro $n = 0$ máme snadno z rovnice (2.3) vztah

$$p_0(s, t) = \exp \left\{ - \int_s^t \lambda(x) dx \right\}. \quad (2.6)$$

Pro $n \in \mathbb{N}$ je situace komplikovanější. Vyjděme z tvrzení (viz například [13, strana 124]), že platí

$$p_n(s, t) = \exp \left\{ - \int_s^t \lambda(x) dx \right\} \frac{[\int_s^t \lambda(x) dx]^n}{n!}. \quad (2.7)$$

Vztah (2.7) zřejmě platí pro $n = 0$. Dále předpokládejme jeho platnost pro $n - 1$. Potom po dosazení do (2.4) postupně píšme

$$\begin{aligned} p_n(s, t) &= \int_s^t \lambda(y) \exp \left\{ - \int_y^t \lambda(x) dx \right\} p_{n-1}(s, y) dy = \\ &= \int_s^t \lambda(y) \exp \left\{ - \int_y^t \lambda(x) dx \right\} \exp \left\{ - \int_s^y \lambda(x) dx \right\} \frac{[\int_s^y \lambda(x) dx]^{n-1}}{(n-1)!} dy = \\ &= \frac{1}{(n-1)!} \exp \left\{ - \int_s^t \lambda(x) dx \right\} \underbrace{\int_s^t \lambda(y) \left[\int_s^y \lambda(x) dx \right]^{n-1} dy}_{I(y)}. \end{aligned} \quad (2.8)$$

Proveďme substituci

$$f(y) = \int_s^y \lambda(x) dx \quad (2.9)$$

a dále nechť je Λ primitivní funkce k λ . Pro derivaci funkce f platí

$$\frac{df}{dy} = \frac{d}{dy} (\Lambda(y) - \Lambda(s)) = \lambda(y).$$

Pro integrál $I(y)$ v (2.8) tedy máme

$$I(y) = \int_s^t \frac{df(y)}{dy} f^{n-1}(y) dy$$

a odtud z věty o derivaci součinu funkcí můžeme psát

$$I(y) = \frac{1}{n} \int_s^t \frac{d}{dy} (f^{n-1}(y)) dy = \frac{1}{n} (f^n(t) - f^n(s)).$$

Po návratu k původní proměnné ze substituce (2.9) už snadno vidíme, že

$$I(y) = \frac{1}{n} \left[\int_s^t \lambda(x) dx \right]^n.$$

Po dosazení $I(y)$ do rovnice (2.8) ihned dostaneme přechodové pravděpodobnosti (2.7). Jelikož jsme na začátku předpokládali libovolnou hodnotu přírůstku $n - 1$, tak jsme indukcí ověřili platnost výsledku pro všechna $n = 0, 1, \dots$

Poznamenejme, že přechodové pravděpodobnosti nezávisí na aktuálním stavu k , a přírůstky tohoto procesu jsou tedy nezávislé. Pro marginální rozdělení N_t ve shodě se značením v (1.3) máme

$$p_n(t) = P(N_t = n) = \exp \left\{ - \int_0^t \lambda(x) dx \right\} \frac{\left[\int_0^t \lambda(x) dx \right]^n}{n!}. \quad (2.10)$$

Odtud vidíme, že stav procesu v libovolném čase t má Poissonovo rozdělení s parametrem $\int_0^t \lambda(x) dx$.

2.2.2 Homogenní Poissonův proces

Přechodové pravděpodobnosti *homogenního Poissonova procesu* nezávisí jednak na jeho stavu (jak tomu bylo v předchozím případě), ale ani na čase, ve kterém se právě nachází. Celá množina funkcí intenzit přechodu se proto redukuje na jedinou konstantní funkci a platí $\lambda_k(t) = \lambda$. Matice intenzit přechodu je tvaru

$$\mathbf{Q} = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \dots \\ 0 & -\lambda & \lambda & 0 & \dots \\ 0 & 0 & -\lambda & \lambda & \ddots \\ \vdots & \vdots & \ddots & \ddots & \ddots \end{pmatrix}.$$

Na obrázku 2.3 jsou vykresleny trajektorie homogenního Poissonova procesu pro různé volby λ .

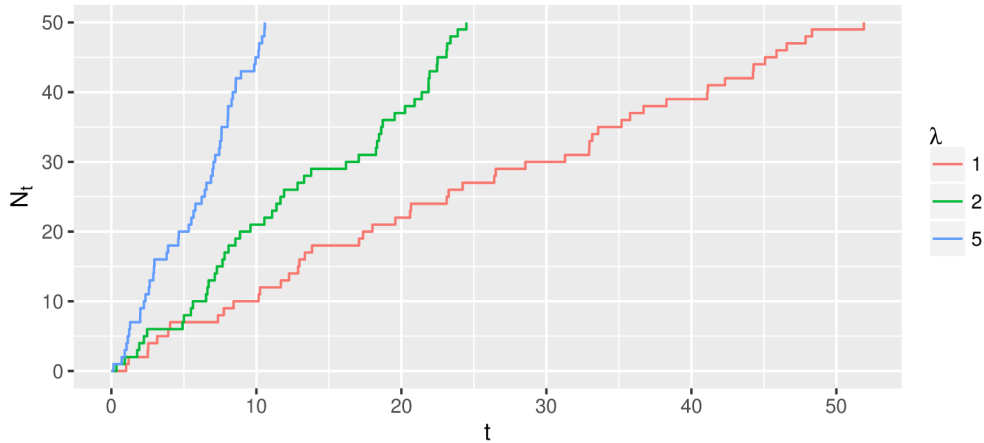
Přechodové pravděpodobnosti jsou pro $n = 0$ z (2.6) tvaru

$$p_0(s, t) = e^{-\lambda(t-s)} \quad (2.11)$$

a pro $n \in \mathbb{N}$ z (2.7) podobně

$$p_n(s, t) = e^{-\lambda(t-s)} \frac{[\lambda(t-s)]^n}{n!}. \quad (2.12)$$

Marginální rozdělení homogenního Poissonova procesu je z předchozího



Obrázek 2.3: Trajektorie homogenního Poissonova procesu

$$p_n(t) = P(N_t = n) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}, \quad (2.13)$$

a opět má tedy Poissonovo rozdělení, avšak nyní s parametrem λt .

Poznamenejme, že v rovnicích (2.11) a (2.12) zápis $\lambda(t-s)$ představuje součin intenzity přechodu a rozdílu časů t a s , nikoli funkci λ se svým argumentem. Protože

$$P(N_t - N_s = n) = \sum_{k=0}^{\infty} p_{k, k+n}(s, t) p_k(s) = p_n(s, t) \underbrace{\sum_{k=0}^{\infty} p_k(s)}_1 = e^{-\lambda(t-s)} \frac{[\lambda(t-s)]^n}{n!} \quad (2.14)$$

nezávisí na k a navíc na s a t pouze prostřednictvím jejich rozdílu, tak homogenní Poissonův proces má nezávislé a stacionární přírůstky. Lze ukázat, že jsou navíc i stejně rozdělené. Označme W_k za čas mezi k minus první a k -tou událostí pro $k \in \mathbb{N}$, kde W_1 je čas první události. Potom

$$P(W_1 > t) = P(N_t = 0) = e^{-\lambda t},$$

a tedy podle 3.2.2 je $W_1 \sim Exp(1/\lambda)$. Dále pišme

$$\begin{aligned} P(W_2 > t \mid W_1 = s) &= P(W_1 + W_2 > s + t \mid W_1 = s) = P(N_{t+s} = 1 \mid N_s = 1) = \\ &= P(N_{t+s} - N_s = 0 \mid N_s = 1) = P(N_{t+s} - N_s = 0), \end{aligned}$$

kde poslední rovnost plyne z nezávislosti přírůstků v Poissonově procesu. Díky (2.14) dostáváme

$$P(W_2 > t \mid W_1 = s) = e^{-\lambda t}.$$

Protože tato rovnost platí pro libovolné $s \geq 0$, tak $P(W_2 > t) = e^{-\lambda t}$ a W_2 je nezávislé na W_1 . Analogickou úvahou zjistíme, že W_1, W_2, W_3, \dots jsou nezávislé a stejně rozdělené náhodné veličiny z exponenciálního rozdělení s parametrem $1/\lambda$.

Poissonův proces má široké uplatnění pro jeho elegantní vlastnosti. Jednou z nich je jeho chování v procesu ředění, který se hojně využívá v pojišťovnictví. Uvažujme kladnou náhodnou veličinu X s distribuční funkcí F_X , která se v čase realizuje tak, že její výskyty splňují homogenní Poissonův proces $\{N_t : t \geq 0\}$ s intenzitou λ . Realizace veličiny X jsou vzájemně nezávislé a jsou nezávislé i na procesu $\{N_t\}$. Mějme také pevně dané kladné reálné číslo d ¹. Sestrojme nyní nový proces $\{N_t^d\}$ tak, že z původního procesu odstraníme všechny události, jejichž příslušné realizace veličiny X nepřevýšily hodnotu d . Lze ukázat, že $\{N_t^d\}$ je opět Poissonovým procesem s intenzitou $[1 - F_X(d)]\lambda$, viz [16, strana 310]. Tuto praktickou vlastnost lze ukázat i pro nehomogenní Poissonův proces.

2.3 Procesy s nákazou

V předchozí podkapitole jsme se zabývali případem, kdy intenzita přechodu závisela na čase. *Procesy s nákazou* (angl. *processes with contagion*) popisují situaci, kdy tato veličina závisí na stavu daného procesu. V rámci tohoto odstavce budeme navíc předpokládat, že intenzitu přechodu ovlivňuje pouze tento stav, a odtud tedy $\lambda_k(t) = \lambda_k$. Za nejjednodušší případ můžeme považovat *lineární závislost intenzity na k* . Definujme tedy

$$\lambda_k = \alpha + \beta k, \quad k = 0, 1, \dots \quad (2.15)$$

pro $\alpha > 0$ a $\beta \neq 0$. Pokud $\lambda_k < 0$, pak klademe $\lambda_k = 0$. Tento proces je v literatuře označován jako *Yuleův proces s driftem* či *lineární proces růstu*.

Intenzita (2.15) nezávisí na t , a proto je tento proces homogenní. Zároveň ale závisí na jeho aktuálním stavu, tedy přírůstky nejsou nezávislé. Je-li $\beta > 0$, pak mluvíme o *pozitivní nákaze*. V tomto případě lze říci, že každý další výskyt zkoumaného jevu ještě zvětšuje jeho frekvenci v budoucnosti. Naopak pro $\beta < 0$ označujeme (2.15) za *negativní nákazu*. Při této situaci více výskytů v minulosti snižuje pravděpodobnost výskytů dalších. Pro úplnost doplníme, že pro $\beta = 0$ se dostáváme zpět k homogennímu Poissonovu procesu s intenzitou α .

Podle [13, strana 126] jsou přechodové pravděpodobnosti procesů s nákazou pro $n \in \mathbb{N}_0$ tvaru

$$p_{k,k+n}(s,t) = \binom{\frac{\alpha}{\beta} + k + n - 1}{n} e^{-(\alpha + \beta k)(t-s)} \left[1 - e^{-\beta(t-s)}\right]^n. \quad (2.16)$$

¹V aplikacích z pojišťovnictví označuje spoluúčast (angl. *deductible*)

Toto tvrzení ověříme indukcí. Pro $n = 0$ máme z rovnice (2.3) výsledek

$$p_{k,k}(s,t) = e^{-(\alpha+\beta k)(t-s)},$$

z čehož ihned vidíme, že (2.16) pro $n = 0$ platí. Dále předpokládejme, že tato rovnost platí pro $n - 1$. Počítejme

$$\begin{aligned} p_{k,k+n}(s,t) &= \int_s^t [\alpha + \beta(k+n-1)] \exp \left\{ - \int_y^t [\alpha + \beta(k+n)] dx \right\} p_{k,k+n-1}(s,y) dy = \\ &= \int_s^t [\alpha + \beta(k+n-1)] \exp \left\{ - \int_y^t [\alpha + \beta(k+n)] dx \right\} \\ &\quad \binom{\frac{\alpha}{\beta} + k + n - 2}{n-1} e^{-(\alpha+\beta k)(t-s)} [1 - e^{-\beta(y-s)}]^{n-1} dy = \\ &= \binom{\frac{\alpha}{\beta} + k + n - 1}{n} e^{-(\alpha+\beta k)(t-s)} \int_s^t \beta n [1 - e^{-\beta(y-s)}]^{n-1} e^{-\beta n(t-y)} dy = \\ &= \binom{\frac{\alpha}{\beta} + k + n - 1}{n} e^{-(\alpha+\beta k)(t-s)} \left\{ [e^{-\beta(t-y)} - e^{-\beta(t-s)}]^n \Big|_{y=s}^{y=t} \right\}, \end{aligned}$$

z čehož po dosazení za y ihned dostáváme (2.16) pro libovolné $n \in \mathbb{N}$.

Doplňme ještě matici intenzit přechodu

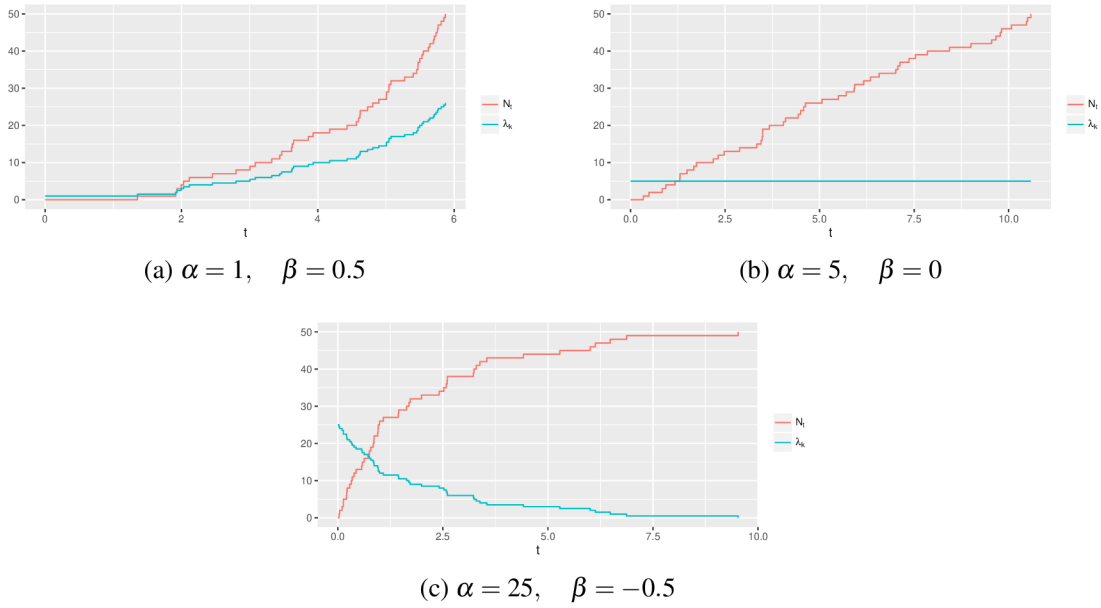
$$\mathbf{Q} = \begin{pmatrix} -\alpha & \alpha & 0 & 0 & \cdots \\ 0 & -(\alpha + \beta) & \alpha + \beta & 0 & \cdots \\ 0 & 0 & -(\alpha + 2\beta) & \alpha + 2\beta & \ddots \\ \vdots & \vdots & \ddots & \ddots & \ddots \end{pmatrix}.$$

Obrázek (2.4) ukazuje možné trajektorie procesu s nákazou pro různé volby α a β .

2.3.1 Pozitivní nákaza

Nyní chceme znát rozdělení přechodových pravděpodobností (2.16). Ukazuje se, že toto rozdělení závisí na znaménku koeficientu β . Připomeňme, že intenzita λ_k musí být vždy nezáporná. V případě pozitivní nákazy je tato podmínka vždy splněna. Navíc pokud platí (2.5), máme zaručeno, že součet pravděpodobností přechodu přes všechny možné přírůstky je roven jedné. Protože $k \in \mathbb{N}_0$, tak

$$\sum_{k=0}^{\infty} \left[\max_{t \geq 0} \lambda_k(t) \right]^{-1} = \sum_{k=0}^{\infty} \frac{1}{\alpha + \beta k} = \infty,$$



Obrázek 2.4: Trajektorie procesů s nákazou

a (2.5) tedy skutečně platí. Dále rovnici (2.16) můžeme jednoduše upravit do tvaru

$$p_{k,k+n}(s,t) = \binom{\frac{\alpha}{\beta} + k + n - 1}{n} \left[e^{-\beta(t-s)} \right]^{\frac{\alpha}{\beta} + k} \left[1 - e^{-\beta(t-s)} \right]^n,$$

což je podle [1, strana 21] pravděpodobnostní funkce negativně binomického rozdělení s parametry

$$r = \frac{\alpha}{\beta} + k,$$

$$p = e^{-\beta(t-s)}.$$

2.3.2 Negativní nákaza

V případě negativní nákazy nemáme zaručenu nezápornost intenzity λ_k z (2.15), protože lineární funkce se zápornou směrnicí klesá pod všechny meze. Pro k proto musí platit nerovnost $\alpha + \beta k \geq 0$, z čehož vyplývá, že $k \leq M$ pro nějaké kladné reálné M . Zároveň musí opět platit podmínka (2.5), kterou můžeme nyní přepsat do tvaru

$$\sum_{k=0}^M \frac{1}{\alpha + \beta k} = \infty.$$

Tato podmínka je splněna pouze tehdy, pokud $M = -\alpha/\beta$. Z toho dále vyplývá, že takové M je nezáporné celé číslo. Po této úvaze již můžeme rovnici (2.16) postupně upravit do tvaru

$$\begin{aligned} p_{k,k+n}(s,t) &= \binom{-M+k+n-1}{n} e^{-(\alpha+\beta k)(t-s)} \left[1 - e^{-\beta(t-s)}\right]^n = \\ &= \frac{1}{n!} \prod_{j=0}^{n-1} (-M+k+j) e^{-[\alpha+\beta(n+k)](t-s)} \left[e^{-\beta(t-s)} - 1\right]^n = \\ &= \frac{(-1)^n}{n!} \prod_{j=0}^{n-1} (M-k-j) e^{\beta(t-s)(-\frac{\alpha}{\beta}-n-k)} (-1)^n \left[e^{-\beta(t-s)} - 1\right]^n = \\ &= \binom{M-k}{n} \left[1 - e^{\beta(t-s)}\right]^n \left[e^{\beta(t-s)}\right]^{M-k-n} \end{aligned}$$

Výraz na posledním řádku je zřejmě podle [1, strana 20] pravděpodobnostní funkce binomického rozdělení s parametry

$$\begin{aligned} n &= M - k, \\ p &= 1 - e^{\beta(t-s)}. \end{aligned}$$

Analýza procesu s negativní nákazou implicitně předpokládá, že $k \in \{0, 1, \dots, M-1\}$. To dává dobrý smysl, uvědomíme-li si, že s M událostmi dospěje $\lambda_k = 0$ a další výskyty se již nemohou objevit. Dosadíme-li tuto nulovou intenzitu do rovnice (2.4), dostáváme $p_{M,M+n}(s,t) = 0$ pro $n \in \mathbb{N}_0$.

2.4 Obecná intenzita přechodu

Věnujme se nyní krátce obecnějším tvarům intenzity přechodu. Přírozeně můžeme uvažovat kombinaci intenzity v nehomogenním Poissonově procesu a v procesu s lineární nákazou. V tomto případě máme

$$\lambda_k(t) = (\alpha + \beta k) \lambda(t), \quad k = 0, 1, \dots$$

Pro $\beta \neq 0$ však nedostáváme žádnou kvalitativní odlišnost od procesu s lineární nákazou, protože podle [13, strana 128] opět obdržíme negativně binomické resp. binomické rozdělení přechodových pravděpodobností pro kladnou resp. zápornou nákazu. Obecný tvar těchto pravděpodobností je podle stejného zdroje

$$p_{k,k+n}(s,t) = \binom{\frac{\alpha}{\beta} + k + n - 1}{n} e^{-(\alpha+\beta k) \int_s^t \lambda(x) dx} \left[1 - e^{-\beta \int_s^t \lambda(x) dx}\right]^n.$$

Samozřejmě pro $\beta = 0$ můžeme bez újmy na obecnosti položit $\alpha = 1$ a ihned dostáváme intenzitu pro nehomogenní Poissonův proces. Ačkoliv jsou přechodové pravděpodobnosti přítomností činitele $\lambda(t)$ v zásadě neovlivněny, charakter výsledného procesu může být odlišný, protože do něj vnáší nehomogenitu přírůstků, jejichž výše už nezávisí pouze na délce časového intervalu.

2.4.1 Obecný tvar nákazy

V případě lineární nákazy diskutované v odstavci 2.3 intenzita přechodu buď roste nade všechny meze nebo naopak klesá tak, že v reálném čase dosáhne nuly, což znamená, že už nemůžeme pozorovat žádný další výskyt. Jistě existují aplikace, ve kterých je tento koncept dobrým popisem skutečnosti, avšak v některých případech může být nevhodný. Uvažme například situaci, kdy každá událost snižuje pravděpodobnost jejího dalšího výskytu, ale v reálném čase tato pravděpodobnost nikdy nedosáhne nuly. Mohl by to být dobrý příklad chování mladého řidiče, který po několika prvních nehodách s přibývajícím věkem a zkušenostmi snižuje svou rizikovost. Nikdy však nelze vyloučit, že působením náhodných vlivů nezpůsobí nehodu další, a proto intenzita nikde nebude nulová.²

Definujme intenzitu přechodu ve tvaru

$$\lambda_k(t) = c_k \lambda(t), \quad k = 0, 1, \dots \quad (2.17)$$

Opět tedy předpokládáme tvar součinu, z něhož jeden činitel popisuje nehomogenitu závislou na čase t a druhý je funkcí aktuálního stavu procesu k . Po dosazení do rovnice (2.3) máme ihned

$$p_{k,k}(s,t) = \exp \left\{ -c_k \int_s^t \lambda(x) dx \right\}$$

a podle [13, strana 128] jsou pro $n \in \mathbb{N}$ pravděpodobnosti přechodu dány rovnicí

$$p_{k,k+n}(s,t) = \sum_{j=0}^n m_j(k,n) \exp \left\{ -c_{k+j} \int_s^t \lambda(x) dx \right\},$$

kde

$$m_j(k,n) = \frac{\prod_{i=0}^{n-1} c_{k+i}}{\prod_{\substack{i=0 \\ i \neq j}}^n (c_{k+i} - c_{k+j})}.$$

²V případě negativní lineární nákazy by výsledkem byla skutečnost, že se po několika nehodách řidič rozhodne úplně přestat řídit

Obecně je proces čistého zrodu s intenzitou přechodu definovanou v (2.17) nehomogenní. Naproti tomu právě tato součinná definice implikuje existenci tzv. *operačního času*, který je transformací času původního.³ Díváme-li se na proces z pohledu této veličiny, pak se jeví jako homogenní. Zavedeme-li operační čas τ ve tvaru

$$\tau(t) = \int_0^t \lambda(x) dx,$$

pak snadno ukážeme, že proces s intenzitou z (2.17) závisí na τ pouze prostřednictvím rozdílu $\tau(t) - \tau(s)$.

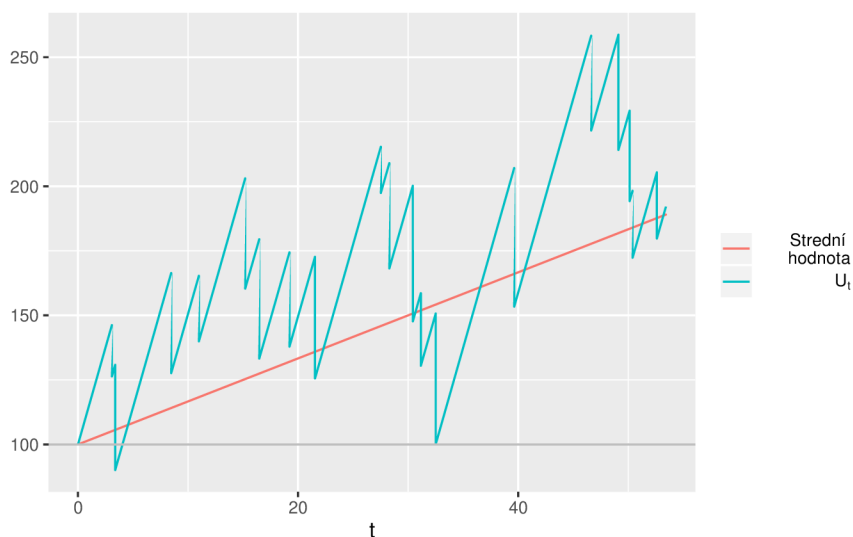
Detailnější diskuzi věnující se procesům čistého zrodu s obecnou intenzitou přechodu lze nalézt například v [4, kapitola 2], [14, kapitola 7] nebo [17, kapitola 3].

³Více o této problematice například v [4, strana 49]

Kapitola 3

Intervaly mezi událostmi

Procesy popsané v předešlé kapitole nachází uplatnění v mnoha vědních disciplínách. Jednou z nich je nepochybně i matematika neživotního pojištění, kde jsou ústředním objektem zájmu v teorii ruinování. Základním modelem, který dává do souvislosti proces náhodného výskytu pojistných událostí a výši příslušných plnění, je model Cramérův-Lundbergův popsaný např. v [13, kapitola 11]. Předpokládá, že události tvoří homogenní Poissonův proces, a tedy že intervaly mezi událostmi jsou nezávislé náhodné veličiny s exponenciálním rozdělením. Na obrázku 3.1 je znázorněn možný vývoj stavu kapitálu pojišťovny U_t v čase t při platnosti tohoto modelu. Pro úplnost doplňme, že navíc předpokládáme i deterministický lineární výběr pojistného a vzájemně nezávislá a stejně rozdělená pojistná plnění.



Obrázek 3.1: Vývoj kapitálu podle Cramérova-Lundbergova modelu

Erik Sparre Andersen byl dánský matematik dvacátého století, který se věnoval teorii pravděpodobnosti. Vystudoval na univerzitě v Kodani, stal se profesorem na univerzitě

v Aarhusu a byl také členem Královské dánské akademie věd. V roce 1957 navrhl ve svém příspěvku k Mezinárodnímu kongresu aktuárů v New Yorku zobecnění modelu Craméra a Lundberga. Viz [18, strana 104]. Předpokládal, že doby mezi dvěma událostmi jsou opět nezávislé a stejně rozdělené náhodné veličiny, avšak opustil předpoklad exponenciálního rozdělení.

V této kapitole vycházíme především z [1], [8], [10] a [15].

3.1 Procesy obnovy

Definice 3.1. Mějme posloupnost $\{X_n, n \in \mathbb{N}_0\}$ nezávislých náhodných veličin nabývajících pouze nezáporných hodnot. Dále necht X_1, X_2, \dots mají stejné rozdělení s distribuční funkcí F a konečnou střední hodnotou. Položme

$$S_n = \sum_{k=0}^n X_k$$

pro $n \in \mathbb{N}_0$. Potom proces náhodných veličin $\{N_t, t \geq 0\}$ takových, že

$$N_t = \sup\{n : S_n \leq t\},$$

se nazývá *proces obnovy*.

Náhodné veličiny X_1, X_2, \dots jsou doby mezi událostmi (obnovami). Jestliže je v předchozí definici $P(X_0 > 0) > 0$, říkáme, že N_t je *proces obnovy se zpožděním*. Naopak, pokud $X_0 = S_0 = 0$, pak považujeme počátek za čas nastání první události a mluvíme o *čistém procesu obnovy*. Příkladem může být počet návratů do stavu k ve vhodně definovaném Markovově řetězci. V případě, že se řetězec na počátku nacházel ve stavu k , pak jde o čistý proces obnovy ve všech ostatních případech dostáváme proces zpožděný.

Náhodná veličina S_n značí čas, kdy dojde k výskytu $(n + 1)$ -ní události. N_t je počet obnov do času t včetně. Rozdělení veličiny N_t získáme snadno z ekvivalence jevů

$$[N_t \leq n] \Leftrightarrow [S_n > t]$$

pro $n \in \mathbb{N}_0$. V tomto případě máme

$$P(N_t = n) = P(S_{n-1} \leq t < S_n), \quad (3.1)$$

kde $n \in \mathbb{N}$. Doplňme, že zřejmě $P(N_t = 0) = P(t < X_0)$. Odtud

$$P(N_t > n) = P(S_n \leq t) = F_n(t),$$

kde F_n je distribuční funkce náhodné veličiny S_n . Protože S_n je součtem X_0, X_1, \dots, X_n , můžeme F_n vyjádřit jako konvoluci distribučních funkcí těchto náhodných veličin. Jestliže $X_0 = 0$, potom

$$F_n(t) = F^{*n}(t).$$

V případě, že $X_0 > 0$ a má distribuční funkci G , pak dostáváme

$$F_n(t) = G * F^{*n}(t).$$

V procesech obnovy zavádíme *funkci obnovy* definovanou jako $m(t) = EN_t$. Pro tuto funkci podle [15, strana 125] platí

$$m(t) = \sum_{n=0}^{\infty} P(S_n \leq t) = \sum_{n=0}^{\infty} F^{*n}(t) < \infty$$

pro libovolné $t \in \mathbb{R}_0^+$.

3.1.1 Asymptotické vlastnosti

Podívejme se nejprve na asymptotické vlastnosti počtu obnov N_t . Předpokládejme, že $P(X_0 < \infty) = 1$ a $EX_1 = \mu$. Podle rovnice (3.1) můžeme psát

$$S_{N_t-1} \leq t < S_{N_t},$$

$$\frac{S_{N_t-1}}{N_t-1} \leq \frac{t}{N_t} < \frac{S_{N_t}}{N_t}.$$

Podle silného zákona velkých čísel s pravděpodobností jedna platí

$$\lim_{t \rightarrow \infty} \frac{S_{N_t}}{N_t} = \lim_{t \rightarrow \infty} \left(\frac{1}{N_t} \sum_{k=0}^{N_t} X_k \right) = \mu.$$

Navíc zřejmě

$$\lim_{t \rightarrow \infty} \frac{N_t - 1}{N_t} = 1,$$

a proto

$$\mu \leq \lim_{t \rightarrow \infty} \frac{t}{N_t} < \mu.$$

Odtud již snadno vidíme, že

$$\lim_{t \rightarrow \infty} \frac{N_t}{t} = \frac{1}{\mu}.$$

Hodnotě $1/\mu$ říkáme *intenzita procesu obnovy*. Obdržení výsledek je do značné intuitivní, protože události nastávají v průměru po μ časových jednotkách. Lze tedy očekávat, že intenzita jejich výskytu odpovídá převrácené hodnotě μ . Podle [16, strana 425] navíc platí i rovnost

$$\lim_{t \rightarrow \infty} \frac{EN_t}{t} = \lim_{t \rightarrow \infty} \frac{m(t)}{t} = \frac{1}{\mu}. \quad (3.2)$$

O (3.2) mluvíme jako o *elementární větě obnovy*.

Známe-li i rozptyl časů mezi obnovami, přičemž $DX_1 = \sigma^2 < \infty$, pak podle [15, strana 127] platí

$$\lim_{t \rightarrow \infty} P \left(\frac{N_t - \frac{t}{\mu}}{\sigma \sqrt{\frac{t}{\mu^3}}} \leq x \right) = \Phi(x),$$

kde Φ je distribuční funkce standardizovaného normálního rozdělení.

Uvažme nyní, že obnova v čase S_n je ohodnocena jistým výnosem R_n . Předpokládejme, že $\{R_n, n \in \mathbb{N}\}$ tvoří posloupnost nezávislých a stejně rozdělených náhodných veličin. Tyto jsou navíc nezávislé na R_0 , které ale může mít obecně jiné rozdělení. Nechť

$$R(t) = \sum_{n=0}^{N_t-1} R_n$$

pro $t \geq 0$ je celkový výnos do času t . Jsou-li náhodné veličiny X_0, R_0 konečné a zároveň je-li $E|R_1| < \infty$, potom podle [15, strana 130] platí

$$\lim_{t \rightarrow \infty} \frac{R(t)}{t} = \lim_{t \rightarrow \infty} \frac{ER(t)}{t} = \frac{\rho}{\mu}, \quad (3.3)$$

kde $\rho = ER_1$.

3.1.2 Rovnice obnovy

Konvoluční rovnice tvaru

$$Z(t) = z(t) + (Z * F)(t) = z(t) + (F * Z)(t), \quad (3.4)$$

kde jsou všechny funkce definovány na množině \mathbb{R}_0^+ a pro $t < 0$ jsou identicky nulové, se nazývá *rovnice obnovy*. Funkce Z je neznámá, z známá a F je distribuční funkce nezáporné náhodné veličiny. Připomeňme, že z vlastností konvoluce máme

$$(Z * F)(t) = \int_0^t Z(t-y) dF(y).$$

Dosazujeme-li rekurentně za Z do rovnice (3.4), dostáváme

$$Z = z + Z * F = z + z * F + Z * F^{*2} = \dots = z + \sum_{n=1}^{\infty} z * F^{*n},$$

z čehož již snadno vidíme řešení pro Z ve tvaru

$$Z = z * \sum_{n=0}^{\infty} F^{*n}.$$

Na závěr tohoto odstavce zjistíme, zda funkce obnovy $m(t)$ splňuje rovnici (3.4). Položme $Z(t) = m(t)$. Z vlastností konvoluce plyne

$$m(t) = F^{0*}(t) + \sum_{n=1}^{\infty} (F * F^{*(n-1)})(t) = F^{0*}(t) + \left(F * \sum_{n=1}^{\infty} F^{*(n-1)} \right)(t) = 1 + (F * m)(t).$$

Funkce obnovy tedy vyhovuje rovnici obnovy pro $z(t) \equiv 1$. Obdobným postupem dostáváme pro proces obnovy se zpožděním rovnost

$$m(t) = G(t) + (F * m)(t)$$

a i v tomto případě vyhovuje příslušná funkce obnovy rovnici (3.4) pro $z(t) = G(t)$.

Další výklad k procesům obnovy je možné nalézt v [16, kapitola 7].

3.2 Rozdělení časů mezi událostmi

V této podkapitole se seznámíme s rozděleními, které jsou charakteristické pro modelování časů mezi dvěma událostmi. Nebude-li explicitně uvedeno jinak, budeme dále uvažovat pouze spojitě nezáporné náhodné veličiny. U zkoumaných rozdělení uvedeme i maximálně věrohodné odhady jejich parametrů, abychom se na ně mohli odkázat a využít je v poslední kapitole. Podrobný popis metody maximální věrohodnosti pro odhad parametrů rozdělení najdeme v [1, podkapitola 7.6].

3.2.1 Číselné charakteristiky rozdělení

U každého rozdělení pravděpodobnosti můžeme zkoumat rozličné charakteristiky. V rámci tohoto odstavce zadefinujeme ty z nich, které budeme v dalším textu používat u konkrétních rozdělení.

- Hustota pravděpodobnosti

Definice 3.2. Řekneme, že náhodná veličina X definovaná na pravděpodobnostním prostoru (Ω, \mathcal{A}, P) je *absolutně spojitého typu*, jestliže existuje nezáporná integrovatelná funkce f taková, že

$$P(X \in A) = \int_A f(x) dx$$

pro každé $A \in \mathcal{A}$. Funkci f nazýváme *hustotou pravděpodobnosti* náhodné veličiny X .

Předpokládáme čas mezi událostmi libovolný kladný. Proto definujeme pouze náhodnou veličinu absolutně spojitého typu.

- Distribuční funkce

Definice 3.3. Buď X náhodná veličina na pravděpodobnostním prostoru (Ω, \mathcal{A}, P) . Potom funkci $F(x) = P(X \leq x)$ pro $x \in \mathbb{R}$ nazýváme *distribuční funkcí* náhodné veličiny X .

Doplňme, že s distribuční funkcí je úzce spjata *funkce přežití*, kterou zavádíme vztahem

$$S(x) = 1 - F(x).$$

V některé literatuře se můžeme setkat se značením $\bar{F}(x)$.

- Funkce rizika

Riziková funkce h je definována jako hustota pravděpodobnosti toho, že událost nastane v intervalu $(x, x + \Delta x)$ za podmínky, že do času x nenastala. Pro náhodnou veličinu X a $\Delta x \rightarrow 0^+$ máme

$$h(x) = \lim_{\Delta x \rightarrow 0^+} \frac{P(x < X \leq x + \Delta x \mid X > x)}{\Delta x} = \frac{1}{P(X > x)} \lim_{\Delta x \rightarrow 0^+} \frac{P(x < X \leq x + \Delta x)}{\Delta x},$$

odkud

$$h(x) = \frac{f(x)}{1 - F(x)}.$$

S využitím funkce přežití můžeme psát

$$h(x) = \frac{f(x)}{S(x)} = -\frac{d}{dx} (\ln S(x)).$$

- Střední hodnota

Definice 3.4. Mějme náhodnou veličinu X na (Ω, \mathcal{A}, P) . Potom číslo

$$EX = \int_{\Omega} X(\omega) dP(\omega),$$

pokud existuje, nazýváme *střední hodnotou* veličiny X .

Z věty o přenosu integrace uvedené např. v [1, strana 14] a z [6, strana 72] pro nezápornou absolutně spojitou náhodnou veličinu X platí

$$EX = \int_0^{\infty} xf(x) dx$$

- Rozptyl

Definice 3.5. Mějme náhodnou veličinu X na (Ω, \mathcal{A}, P) . Potom pro $k \in \mathbb{N}$ nazýváme čísla

$$EX^k$$

k -tým *obecným momentem* a čísla

$$E(X - EX)^k$$

k -tým *centrovaným momentem* náhodné veličiny X .

Druhý centrovaný moment označujeme jako *rozptyl* a značíme jej DX . O druhé odmocnině z tohoto čísla hovoříme jako o *směrodatné odchylce*. Pro nezápornou absolutně spojitou náhodnou veličinu X obdobně jako u střední hodnoty platí

$$DX = \int_0^{\infty} (x - EX)^2 f(x) dx.$$

- Charakteristická funkce

Definice 3.6. Nechť X je náhodná veličina na (Ω, \mathcal{A}, P) . Potom se funkce

$$\psi_X(t) = Ee^{tX}$$

pro $t \in \mathbb{R}$ nazývá *charakteristickou funkcí* náhodné veličiny X .

Charakteristická funkce existuje vždy, protože

$$e^{tX} = \cos(tX) + i \sin(tX)$$

a $|\cos(tX)| \leq 1$, $|\sin(tX)| \leq 1$. Pro nezápornou absolutně spojitou náhodnou veličinu počítáme charakteristickou funkci podle [6, strana 84] ze vztahu

$$\int_{\mathbb{R}} e^{tx} f(x) dx.$$

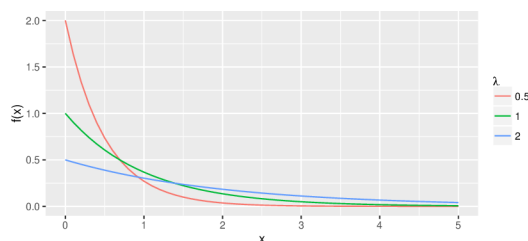
Další používané číselné charakteristiky rozdělení pravděpodobnosti a odvození jejich vlastností najdeme například v [6, kapitola 4].

3.2.2 Exponenciální rozdělení

Jak jsme již ukázali v odstavci 2.2.2, exponenciální rozdělení je rozdělením intervalů mezi příchody dvou po sobě jdoucích událostí v Poissonově procesu. Je to jediné rozdělení, které má rizikovou funkci konstantní. Pro tuto vlastnost o něm hovoříme jako o rozdělení „bez paměti“. Má jeden parametr $\lambda > 0$ a značíme jej $Exp(\lambda)$.

- Hustota pravděpodobnosti

$$f(x) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}}$$



Obrázek 3.2: Hustota pravděpodobnosti exponenciálního rozdělení

- Distribuční funkce

$$F(x) = 1 - e^{-\frac{x}{\lambda}}$$

- Funkce rizika

$$h(x) = \frac{1}{\lambda}$$

- Charakteristická funkce

$$\psi(t) = (1 - i\lambda t)^{-1}$$

- Střední hodnota

$$EX = \lambda$$

- Rozptyl

$$DX = \lambda^2$$

- Maximálně věrohodný odhad parametru

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

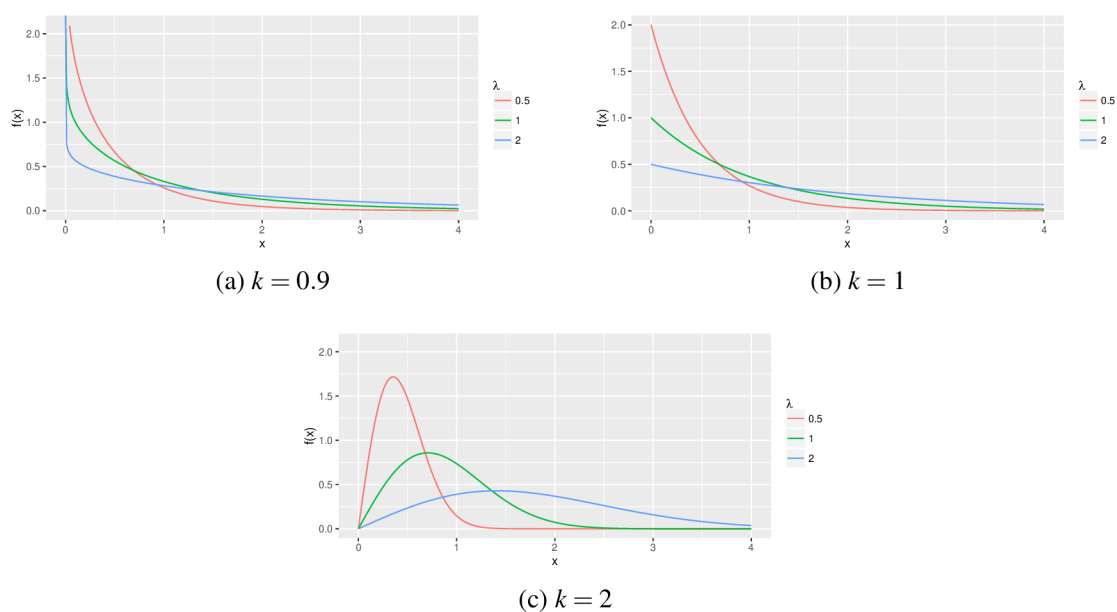
Uvedený vztah je platný pro náhodný výběr o rozsahu n . Pro další rozdělení budeme předpokládat tutéž velikost.

3.2.3 Weibullovo rozdělení

Weibullovo rozdělení splňuje náhodná veličina, která představuje například životnost technických součástí. Disponuje dvěma parametry $\lambda > 0$ a $k > 0$. Druhý jmenovaný parametr charakterizuje mechanické namáhání, způsob použití, stáří zařízení a jiné. Značíme $Wb(k, \lambda)$. Exponenciální rozdělení je speciálním případem Weibullova pro $k = 1$.

- Hustota pravděpodobnosti

$$f(x) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k}$$



Obrázek 3.3: Hustota pravděpodobnosti Weibullova rozdělení

- Distribuční funkce

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

- Funkce rizika

$$h(x) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1}$$

- Charakteristická funkce

$$\psi(t) = \sum_{i=1}^{\infty} \frac{(i\lambda t)^i}{i!} \Gamma\left(1 + \frac{i}{k}\right)$$

Funkce $\Gamma(k) = \int_0^{\infty} t^{k-1} e^{-t} dt$ je gama funkce.

- Střední hodnota

$$EX = \lambda \Gamma\left(1 + \frac{1}{k}\right)$$

- Rozptyl

$$DX = \lambda^2 \left[\Gamma\left(1 + \frac{2}{k}\right) - \Gamma^2\left(1 + \frac{1}{k}\right) \right]$$

- Maximálně věrohodné odhady parametrů

$$\hat{\lambda} = \left(\frac{1}{n} \sum_{i=1}^n x_i^{\hat{k}} \right)^{\hat{k}^{-1}}$$

$$\frac{1}{\hat{k}} = \frac{\sum_{i=1}^n x_i^{\hat{k}} \ln x_i}{\sum_{i=1}^n x_i^{\hat{k}}} - \frac{1}{n} \sum_{i=1}^n \ln x_i$$

Nejprve se numericky řeší druhá rovnice, ze které se získá odhad parametru k . Následně se tento odhad použije při analytickém odhadu parametru λ z první rovnice.

3.2.4 Gama rozdělení

Náhodná veličina, která je součtem $k \in \mathbb{N}$ náhodných veličin s exponenciálním rozdělením se stejným parametrem $\lambda > 0$, má Erlangovo rozdělení pravděpodobnosti. Připustíme-li, aby $k \in \mathbb{R}^+$, pak má gama rozdělení. Značíme $\Gamma(k, \lambda)$.

- Hustota pravděpodobnosti

$$f(x) = \frac{1}{\Gamma(k)\lambda^k} x^{k-1} e^{-\frac{x}{\lambda}}$$

- Distribuční funkce

$$F(x) = \frac{1}{\Gamma(k)} \gamma\left(k, \frac{x}{\lambda}\right)$$

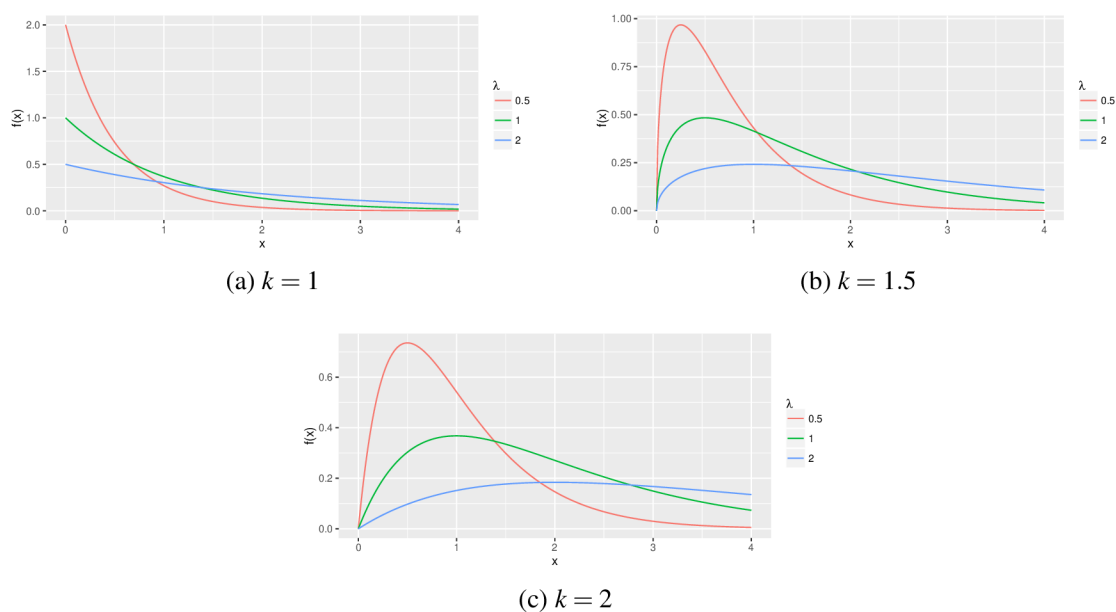
Funkce $\gamma(k, x) = \int_0^x t^{k-1} e^{-t} dt$ je tzv. dolní neúplná gama funkce.

- Funkce rizika

Tvar funkce rizika nelze pro gama rozdělení vhodně zjednodušit. Proto při jejím výpočtu vycházíme přímo z definiční formule.

- Charakteristická funkce

$$\psi(t) = (1 - i\lambda t)^{-k}$$



Obrázek 3.4: Hustota pravděpodobnosti gama rozdělení

- Střední hodnota

$$EX = k\lambda$$

- Rozptyl

$$DX = k\lambda^2$$

- Maximálně věrohodné odhady parametrů

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

$$\ln \hat{k} = \psi(\hat{k}) + \ln \left(\frac{1}{n} \sum_{i=1}^n x_i \right) - \frac{1}{n} \sum_{i=1}^n \ln x_i$$

Funkce $\psi(k) = \frac{d}{dk} \ln \Gamma(k)$ je tzv. digama funkce. Druhou rovnici lze pro \hat{k} opět řešit numericky.

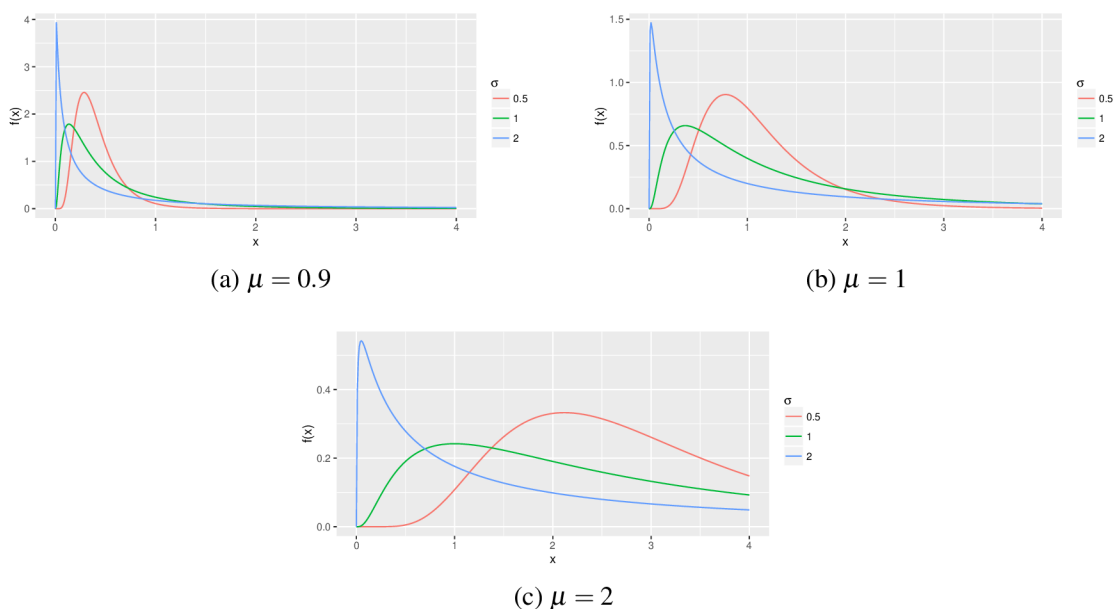
3.2.5 Lognormální rozdělení

Uvažme náhodnou veličinu Y , která se řídí normálním rozdělením. Potom transformovaná náhodná veličina $X = e^Y$ splňuje lognormální (též logaritnicko-normální) rozdělení

pravděpodobnosti. Značíme $LN(\mu, \sigma^2)$, kde $\mu \in \mathbb{R}$ a $\sigma^2 > 0$ jsou parametry příslušného normálního rozdělení veličiny Y . Například v Blackově-Scholesově modelu pro oceňování finančních derivátů definovaném v [9, kapitola 14] se předpokládá lognormální rozdělení ceny akcie.

- Hustota pravděpodobnosti

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$$



Obrázek 3.5: Hustota pravděpodobnosti lognormálního rozdělení

- Distribuční funkce

$$F(x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right)$$

Funkce Φ je distribuční funkce standardizovaného normálního rozdělení.

- Funkce rizika

Opět jako v případě gama rozdělení neumíme tvar funkce rizika zjednodušit.

- Charakteristická funkce

Lognormální rozdělení nemá podle [2, strana 2] charakteristickou funkci analytickou v bodě 0, a nemůže tedy být vyjádřena jako konvergentní nekonečná mocninná řada. V citované publikaci lze nalézt její částečnou analýzu.

- Střední hodnota

$$EX = e^{\mu + \frac{\sigma^2}{2}}$$

- Rozptyl

$$DX = (e^{\sigma^2} - 1) e^{2\mu + \frac{\sigma^2}{2}}$$

- Maximálně věrohodné odhady parametrů

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n \ln x_i$$

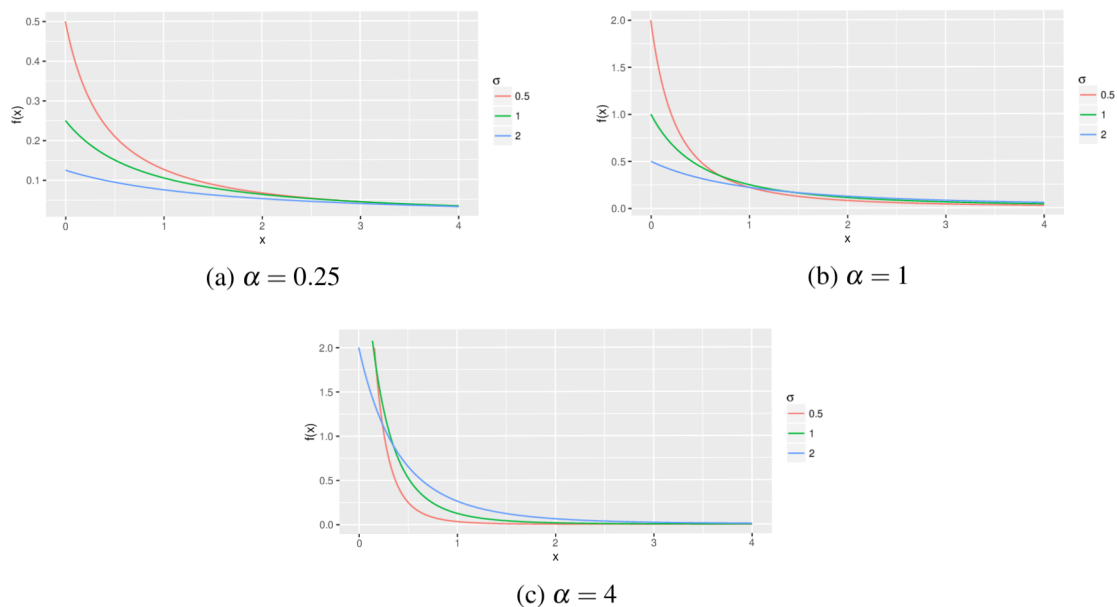
$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (\ln x_i - \hat{\mu})^2$$

3.2.6 Paretovo rozdělení

V literatuře se můžeme setkat s různými definicemi Paretova rozdělení. V této práci budeme používat tzv. Paretovo rozdělení druhého řádu, které je zadefinováno např. v [10, strana 575]. Tomuto rozdělení se také někdy říká Lomaxovo. Disponuje dvěma parametry $\alpha > 0$ a $\sigma > 0$. Oproti výše popsaným rozdělením pravděpodobnosti neklesá jeho hustota pro x jdoucí k nekonečnu exponenciálně ale polynomicky. O takovýchto rozděleních říkáme, že mají těžký pravý konec. Pro jednoduchost při zápisu vynecháme informaci o řádu rozdělení a budeme značit symbolem $Pa(\sigma, \alpha)$.

- Hustota pravděpodobnosti

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



Obrázek 3.6: Hustota pravděpodobnosti Paretova rozdělení

- Distribuční funkce

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

- Funkce rizika

$$h(x) = \frac{\alpha}{x + \sigma}$$

- Charakteristická funkce
Charakteristickou funkci Paretova rozdělení nebudeme pro její složitost uvádět.

- Střední hodnota

$$EX = \frac{\sigma}{\alpha - 1}$$

Střední hodnota existuje pouze pro $\alpha > 1$.

- Rozptyl

$$DX = \frac{\alpha\sigma^2}{(\alpha - 1)^2(\alpha - 2)}$$

Rozptyl je definován pouze pro $\alpha > 2$. Analogicky jsou momenty k -tého řádu podle [10, strana 577] definovány pro $\alpha > k$.

- Maximálně věrohodné odhady parametrů

$$\hat{\alpha} = n \left[\sum_{i=1}^n \ln \left(1 + \frac{x}{\hat{\sigma}} \right) \right]^{-1}$$

$$\sum_{i=1}^n \left(1 + \frac{\hat{\sigma}}{x} \right)^{-1} \left[1 + \frac{n}{\sum_{i=1}^n \ln \left(1 + \frac{x}{\hat{\sigma}} \right)} \right] = n$$

Druhou rovnici řešíme pro $\hat{\sigma}$ numericky.

3.3 Testy o shodě rozdělení

V aplikační části této práce budeme zkoumat, jakým rozdělením se řídí časy mezi událostmi. Ukážeme si proto nyní statistické nástroje, kterými lze rozhodnout, zda analyzovaná data podporují vybrané rozdělení.

3.3.1 χ^2 test dobré shody

Tento test je konstruován pro náhodné veličiny z multinomického rozdělení.

Definice 3.7. Mějme n nezávislých pokusů, které mohou vyústit vždy právě jedním z $k \geq 2$ možných jevů A_1, \dots, A_k . Dále nechť pro $i \in \{1, \dots, k\}$ je pravděpodobnost $P(A_i) = p_i$ stejná pro všechny pokusy a $\sum_{i=1}^k p_i = 1$. Označme symbolem X_i náhodnou veličinu, která udává počet nastání i -tého jevu. Řekneme, že náhodné veličiny X_1, \dots, X_k mají *multinomické* rozdělení pravděpodobnosti, jestliže je jejich sdružená pravděpodobnostní funkce dána vzorcem

$$P(X_1 = x_1, \dots, X_k = x_k) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k},$$

kde $\sum_{i=1}^k x_i = n$. Tuto skutečnost značíme $X_1, \dots, X_k \sim MN(n; p_1, \dots, p_k)$.

Poznamenejme, že marginální rozdělení mají opět multinomická rozdělení. Navíc pro $k = 2$ dostáváme rozdělení binomické.

Přidržíme-li se značení z definice 3.7, pak má statistika

$$\chi^2 = \sum_{i=1}^k \frac{(X_i - np_i)^2}{np_i} \tag{3.5}$$

podle [1, strana 270] pro $n \rightarrow \infty$ asymptoticky χ^2 rozdělení s $k - 1$ stupni volnosti. Tato statistika je označována jako *Pearsonova*. Jakmile $\chi^2 \geq \chi_\alpha^2(k - 1)$, zamítáme na hladině významnosti α hypotézu, že náhodný výběr pochází z multinomického rozdělení $MN(n; p_1, \dots, p_k)$.

Náhodná veličina řídicí se spojitým rozdělením pravděpodobnosti se může realizovat nespočetně mnoha výsledky. Chceme-li proto využít χ^2 test pro rozhodnutí, zda náhodný výběr pochází ze spojitého rozdělení, musíme jej nejprve vhodně rozdělit do třídících intervalů (neboli tříd). Veličina X_i potom značí absolutní četnost výsledků v i -té třídě. Je důležité si uvědomit, že test dobré shody je asymptotický, a proto je možné jej využít jen při dostatečně velkém rozsahu výběru n . V literatuře se obvykle uvádí, že teoretická absolutní četnost každé třídy musí být větší než pět, tj. $np_i \geq 5$. Podle [1, strana 271] se užívá i slabší *Yarnoldovo kritérium*, které požaduje pouze $np_i \geq 5q$, kde q je podíl tříd, pro které platí $np_i < 5$.

Doposud jsme předpokládali, že známe všechny parametry testovaného rozdělení. Když neznáme právě p parametrů, pak podle [3, strana 205] platí $\chi^2 \sim \chi^2(k - p - 1)$, kde k je opět počet třídících intervalů. Nabízí se také otázka, jakou metodou odhadovat vektor neznámých parametrů $\mathbf{a} = (a_1, \dots, a_p)^T$. Při testování, zda náhodný výběr pochází z konkrétního rozdělení, většinou nechceme tuto hypotézu zamítnout. Proto je zřejmě vhodné hledat odhad $\hat{\mathbf{a}}$ vektoru \mathbf{a} takový, pro který je statistika (3.5) s $p_i = p_i(\mathbf{a})$ minimální. O tomto postupu se hovoří jako o *metodě minimálního χ^2* . Celou proceduru nalezneme v [1, podkapitola 12.3]. Doplňme, že podle stejného zdroje nemůžeme v tomto případě použít Yarnoldovo kritérium k rozhodnutí o dostatečnosti rozsahu výběru.

3.3.2 Kolmogorovův-Smirnovův test

Kolmogorovův-Smirnovův test je neparametrický. Zkoumáme pomocí něj průběh empirických distribučních funkcí $F_1(x)$ resp. $F_2(y)$ dvou nezávislých náhodných výběrů X_1, \dots, X_n resp. Y_1, \dots, Y_m . Testujeme hypotézu, že všech $n + m$ náhodných veličin pochází z téhož rozdělení proti alternativě, že jsou jejich distribuční funkce rozdílné.

Kolmogorovův-Smirnovův test využívá statistiku

$$D = \max_{x \in \mathbb{R}} |F_1(x) - F_2(x)|, \quad (3.6)$$

kteřá nemá žádné známé rozdělení pravděpodobnosti a její kritické hodnoty jsou tabulovány. Hypotézu o shodě rozdělení výběrů X_1, \dots, X_n a Y_1, \dots, Y_m zamítáme na hladině významnosti α , pokud $D \geq D_\alpha(m, n)$. Pro velké rozsahy výběrů lze podle [3, strana 200] kritickou hodnotu aproximovat vzorcem

$$D_\alpha(m, n) \approx \sqrt{\frac{n+m}{2nm} \ln \frac{2}{\alpha}}.$$

Porovnáváme-li jeden náhodný výběr X_1, \dots, X_n s teoretickým rozdělením, idea testu se nemění a opět bereme za testovou statistiku největší odchylku distribučních funkcí. Uvědomíme-li si, že v tomto případě je $m \rightarrow \infty$, můžeme navíc pro velké n aproximovat kritickou hodnotu testu vztahem

$$D_\alpha(n) \approx \sqrt{\frac{1}{2n} \ln \frac{2}{\alpha}}.$$

Kapitola 4

Aplikace

V této kapitole se budeme věnovat aplikaci teoretických poznatků zkoumaných v předchozím textu na reálnou situaci v pojišťovnictví. Jak již bylo řečeno v úvodu, tuto oblast ovlivňují pojistné události a pojišťovna, která v takovém případě musí vyplatit klientovi pojistné plnění, nemůže s jistotou určit, kdy taková událost nastane. Předmětem této analýzy bude prozkoumání intervalů mezi dvěma událostmi a určení rozdělení pravděpodobnosti, které je nejlíp popisuje. Toto rozdělení může být následně využito ve Sparre Andersenově modelu a spolu s rozdělením velikostí nároků vytvořit komplexní představu o pravděpodobnosti, že pojišťovna nebude schopna dostát svým závazkům.

V následujícím textu se u čtenáře předpokládá elementární znalost pojmů z pojišťovnictví. Tyto lze nalézt například v [5].

4.1 Data

Data, která budeme analyzovat, pochází z české pobočky pojišťovacího koncernu Assicurazioni Generali S.p.A., který byl založen roku 1831 v italském Terstu. Vzhledem k důvěrnému charakteru těchto dat, byla tato obdržena bez nároku na jejich úplnost a správnost. Veškeré obsažené nesrovnalosti ale nemají dopad na výsledky testů, které budeme provádět. Data byla zpracována v softwaru Microsoft Office Excel 2007. Statistické testování bylo provedeno a grafy vykresleny statistickým programem **R**. Poznamenejme, že v celé kapitole uijeme k oddělení desetinné části čísel tečku.

V tabulce 4.1 vidíme základní strukturu dat. V prvních třech sloupcích je jednoznačná identifikace škody. Linkou se rozumí část smluv z portfolia pojišťovny, které jsou co do rizikovosti a struktury homogenní. V tomto případě jsou linkami konkrétní produkty, které budou zevrubněji popsány dále. Den vzniku škody je uveden ve formátu RRRRMMDD. Zbylé tři sloupce se týkají informací o finančních pohybech na dané škodě. Po řadě označují již přiznanou výplatu, výši rezervy na případné další plnění a celkem evidované finanční prostředky spojené s danou škodou. V datech jsou zaznamenány i nulové škody, tj. škody, na kterých byla historicky vytvořena nenulová rezerva, ale na základě likvidace nedošlo k jejich uznání jako pojistné události. Částky jsou uvedeny v českých korunách. Časové rozpětí dnů vzniku událostí je od 1. 1. 2003 do 31. 12. 2016.

Identifikace škody			Informace o plnění		
Číslo	Linka	Den vzniku	Paid	RBNS	Incurred
300****711	18	20131101	6 828	0	6 828
300****716	32	20150206	29 257	0	29 257
12/****986	54	20120705	57 000	0	57 000
14/****466	32	20140127	0	0	0
05/****189	54	20050605	394 160	0	394 160
300****035	05	20150307	4 094	0	4 094
300****576	55	20160708	0	250 000	250 000
10/****991	54	20100803	341 614	0	341 614
300****557	32	20100429	0	180 000	180 000
300****086	05	20160723	6 022	10 000	16 022
⋮	⋮	⋮	⋮	⋮	⋮

Tabulka 4.1: Struktura dat

4.1.1 Typy pojištění

Podívejme se nyní na jednotlivé linky a ukažme si, jaká pojištění jsou v každé z nich obsažena.

Pojištění dopravy – majetek (03)

Pojištění je určeno zejména pro vlastníky převáženého zboží. Vztahuje se na veškeré nahodilé škody na přepravovaném zboží na území celého světa, jak z důvodu dopravní nehody nebo odcizení, tak z důvodu přírodních katastrof. Kryje jeho skutečnou hodnotu, a to i nad odpovědností limit dopravce. Pojištění chrání zboží i na obchodních a uměleckých výstavách.

Pojištění dopravy – odpovědnost (34)

Toto pojištění se na rizika spojená s transportem zboží dívá optikou dopravce, který převoz zajišťuje. Rizika jsou tedy velmi podobná těm z předchozího případu, ale chrání přepravce, který je v případě poškození svěřeného majetku povinen nahradit vzniklou škodu majiteli zboží.

Občanská odpovědnost (05)

Pojištění chrání klienty v situaci, kdy v běžném občanském životě způsobí někomu škodu, kterou jsou povinni nahradit. Vztahuje se na škody vzniklé ve spojení s vedením domácnosti, provozováním rekreační sportovní činnosti nebo opatrováním domácích zvířat.

Pojištění se nevztahuje na hospodářská zvířata nebo škody způsobené zaměstnavateli.

Pofesní odpovědnost (32)

Tento produkt je určen kvalifikovaným profesionálům s náročnou prací, jejichž chyby mohou mít závažné důsledky. Pojištění pokrývá ublížení na zdraví ale i finanční škody. Nejběžnějšími profesemi, které toto pojištění pokrývá, jsou např. lékaři, advokáti, soudní znalci, daňoví poradci, architekti a tlumočníci.

Pojištění úvěrů (18)

Klienty tohoto typu pojištění jsou subjekty poskytující finanční úvěry. Za pojistnou událost se považuje případ, kdy dlužník nedostojí svým závazkům vůči věřiteli. Vyloučeno je zde životní riziko, tj. nesplacení dluhu z důvodu úmrtí.

Pojištění plodin (54)

Tento produkt je určen pěstitelům plodin na území České republiky. Pojištění se vztahuje mimo jiné na obilniny, krmné pícniny, okopaniny, léčivé a kořeninové rostliny, ale i na trvalé travní porosty a trvalé kultury jakými jsou např. vinná réva nebo ovocné sady. Mezi pojištěná rizika patří krupobití, požár, vichřice, povodeň, záplava, sesuv půdy nebo vymrznutí.

Pojištění hospodářských zvířat (55)

Pojištění pokrývá rizika spojená s chovem hospodářských zvířat. Vztahuje se na velká zvířata (prasata, ovce, skot, kozy), drůběž i farmovou zvěř. Mimo podobných rizik jako u pojištění plodin jsou pokryty i škody způsobené zasažením elektrickým proudem, úraz, akutní onemocnění nebo škody způsobené neohlášeným výpadkem elektrického proudu.

Graf 4.1 (tab. A.1 přílohy) ukazuje počty škod po jednotlivých linkách a letech vzniku.

4.2 Příprava dat

Základní datový soubor, který máme k dispozici obsahuje všechny potřebné informace k provedení analýzy intervalů mezi dvěma po sobě následujícími událostmi. Než se do tohoto cvičení budeme moci pustit, musíme ale data nejprve očistit o případy, se kterými z nějakého důvodu nechceme pracovat. V rámci této sekce provedeme postupnými kroky toto očištění a připravíme data do potřebného stavu.



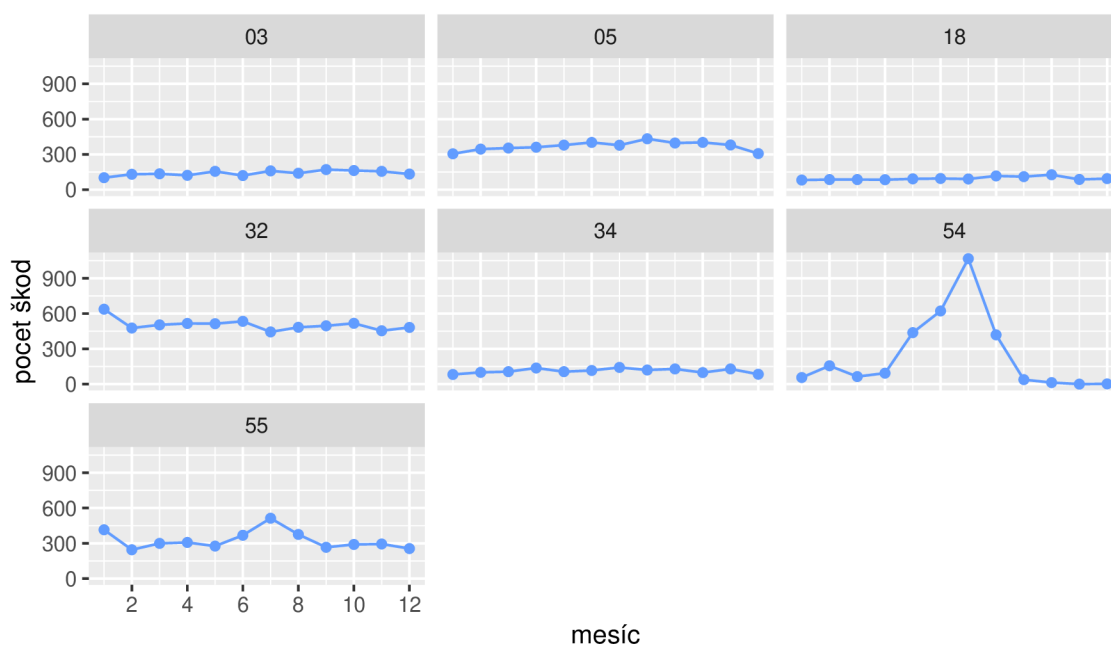
Obrázek 4.1: Počet škod podle linek a let vzniku

4.2.1 Analýza sezónnosti

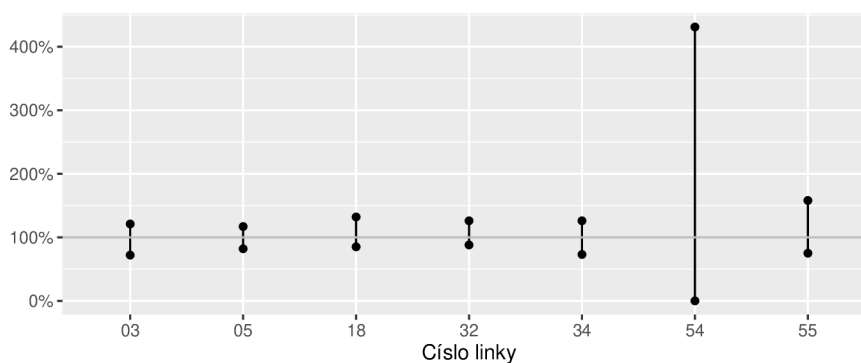
Proces obnovy předpokládá, že časy mezi dvěma událostmi jsou nezávislé a stejně rozdělené náhodné veličiny. Nezávislost můžeme u dostatečně velkého portfolia s mnoha vzájemně neovlivňujícími se klienty snadno předpokládat. U sezónních dat se ale dostáváme do situace, kdy druhý předpoklad zřejmě splněn není. V čase se střídají úseky, kdy je pravděpodobnost nastání události větší, a naopak období, kdy je tato pravděpodobnost menší. Intervaly mezi událostmi mohou mít stejný typ rozdělení (exponenciální, Weibullovo apod.). Jistě však nebudou mít shodné alespoň parametry, které v něm vystupují.

V rámci této analýzy budeme zkoumat sezónnost s roční periodou. Jistě můžeme v datech pozorovat i např. měsíční nebo naopak delší než roční periodicitu. Stručně uvedme, proč se těmito případy nebudeme zabývat. Lze si jen obtížně představit, jaký faktor by mohl způsobit delší než roční periodicitu počtu pojistných událostí. Zvláštním druhem takové příčiny by mohly být pravidelně kolísající počty pojistných smluv. Tuto situaci budeme řešit v odstavci 4.2.3. Navíc z obrázku 4.1 je patrné, že k periodickému chování s více než roční periodou v žádné z linek nedochází. Področní periodu lze uvažovat například denní. Vzhledem k větší aktivitě lidí a s tím spojeného rizika nastání pojistné události je pravděpodobnost škody větší přes den než v noci. Data, která máme k dispozici, však nedisponují časovým údajem, a proto můžeme od této skutečnosti abstrahovat.

Z grafu 4.2 (tab. A.2 přílohy) je vidět, že kromě pojištění plodin (linka 54) a pojištění hospodářských zvířat (linka 55) se s roční periodicitou pojistných událostí nesetkáváme. To dokládá i obrázek 4.3 (tab. A.3 přílohy) ukazující rozpětí poměrů počtu škod v jednotlivých měsících vzhledem k jejich průměru.



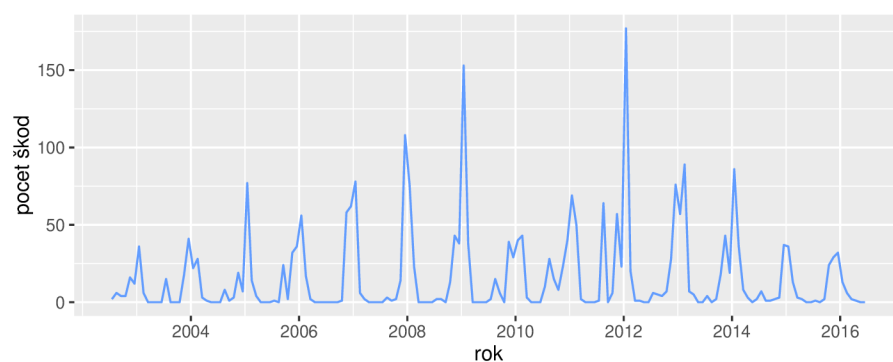
Obrázek 4.2: Počet škod podle linek a měsíců vzniku



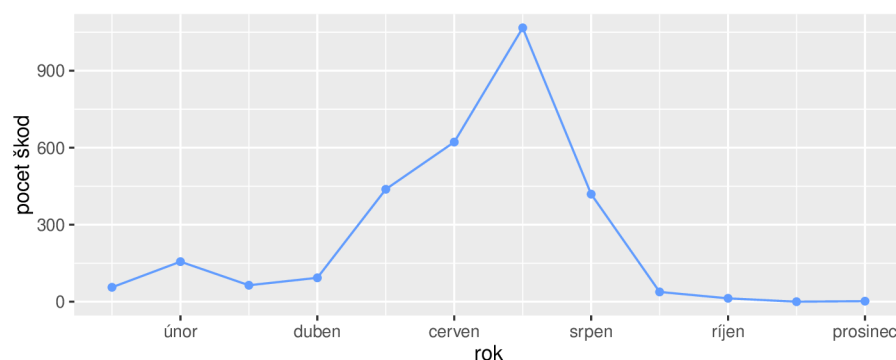
Obrázek 4.3: Max. a min. odchylka dat od průměru

Naopak je evidentní, že zejména v lince 54 se setkáváme s roční sezónností. Na grafech 4.4 je detailnější obraz pojištění plodin. Tento produkt je bezpochyby silně ovlivněn ročním cyklem počasí. Konkrétně na obrázku 4.4b si můžeme všimnout zvýšených hodnot v únoru a dále pak od května do srpna. Na začátku roku se zemědělci potýkají s mrazy, které tradičně přichází právě v únoru, a v letních měsících sužují pole kroupy a povodně z bleskových bouřek.

Identifikovali jsme tedy pojištění plodin a pojištění hospodářských zvířat jako ty, které vykazují sezónní charakter s roční periodou. Do dalších analýz nezahrneme pouze linku 54. U linky 55 budeme mít tuto skutečnost na paměti, ale budeme s tímto pojištěním dále pracovat.



(a) Měsíční úhrny škod



(b) Průměry úhrnů škod podle měsíců

Obrázek 4.4: Sezónnost pojištění plodin (54)

4.2.2 Nulové škody

Po vybrání vhodných linek můžeme přistoupit k samotné úpravě dat. V těch jsou obsaženy i nulové škody, tj. škody, na kterých byla historicky vytvořena nenulová rezerva, ale na základě likvidace nedošlo k jejich uznání jako pojistné události. O tyto škody musíme data očistit, protože ve skutečnosti nejsou v této analýze pro pojišťovnu relevantní. Nulové škody mohou být ale předmětem jiného testování. Při velkém podílu nulových škod by pojišťovna mohla začít uvažovat o jasnější komunikaci směrem ke klientům, jaké škody jsou v rámci sjednaného produktu skutečně kryty. Mohla by tím snížit náklady na likvidaci a případně si i zvýšit renomé, které při častém odmítání plnění obvykle trpí i v případě, že je pojišťovna v právu.

Odstranění nulových škod ale nestačí. V datech jsou i škody, které jsou otevřené, tzn. mají vytvořenu nenulovou rezervu, ale nebylo na nich prozatím nic vyplaceno. V takovém případě může dojít buď k výplatě pojistného plnění, nebo se škoda uzavře jako nulová. Část těchto škod by s odstupem času tedy z analýzy vypadla.¹ K vyřešení tohoto problému jsme spočetli poměr nulových škod ke všem zavřeným škodám a právě tento podíl otevřených škod s prozatím nulovou výplatou byl náhodně odstraněn.

¹Procesu zpřesňování informací z pohledu výplaty a výše rezervy se říká *run-off*

V tabulce 4.2 jsou shrnuty výsledky procesu odstranění nulových škod. Podle očekávání má na redukci dat díky promptní likvidaci mnohem zásadnější dopad vypuštění zavřených nulových škod. Run-off otevřených s aktuálně nulovou výplatou přináší s výjimkou linky 32 pouze drobnou korekci. Přesto je tento krok důležitý pro správnou a především kompletní přípravu dat.

Položka	Číslo linky					
	03	05	18	32	34	55
počet škod	1 616	3 396	1 153	5 612	1 321	3 527
nulové	362	1 047	607	1 936	512	282
nulové/zavřené	24.51 %	30.12 %	52.69 %	46.86 %	41.06 %	9.07 %
run-off	15	30	1	591	32	3
bez nulových	1 239	2 319	545	3 085	777	3 242

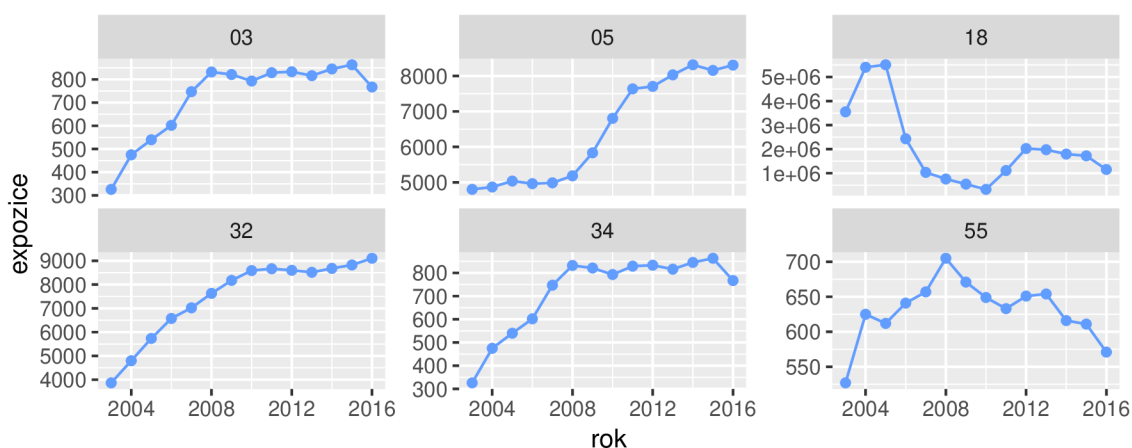
Tabulka 4.2: Odstranění nulových škod

4.2.3 Expozice

V odstavci 4.2.1 jsme hovořili mimo jiné i o periodickém chování dat s větší než roční periodou. Z grafu 4.1 je patrné, že žádný produkt takovou tendenci nevykazuje. Naproti tomu bychom při shodě rozdělení intervalů mezi dvěma událostmi očekávali, že počty škod budou pro jednotlivé roky náhodně oscilovat kolem svého průměru. K tomu ovšem u většiny linek nedochází a naopak je u některých z nich patrný trend. Toto chování je způsobeno výkyvy velikosti rizika, kterému je pojišťovna vystavena (exponována). Předpokládáme, že toto kolísání je způsobeno výhradně změnou *expozice*, kterou budeme u linky 18 měřit zaslouženým pojistným a ve všech ostatních případech počtem pojistek.² Je zřejmé, že při větší expozici bude za jinak stejných podmínek docházet ke škodám častěji. Poznamenáme, že obecně to nemusí být jediný důvod změny rizika, které pojišťovna podstupuje. Např. kvalitní segmentací klientů, výpovědí smluv s velkou škodovostí a dalšími nástroji může efektivně snížit riziko svého portfolia při zachování jeho velikosti. Od této skutečnosti ale pro zjednodušení problému v dalším textu odhlédneme.

Na obrázku 4.5 (tab. A.4 přílohy) vidíme výkyvy expozice v jednotlivých letech související s rozvojem pojišťovny. Očištění dat o vliv expozice provedeme obdobně jako v případě run-offu nulových škod. Vhodně zvolíme základní hladinu expozice pro každou linku zvlášť. Následně pro každý rok vypočteme poměr mezi zvolenou základní expozicí a expozicí příslušnou danému roku. Tento podíl škod ponecháme v datovém souboru. Škody, které z dat odstraníme, vybereme zcela náhodně. Tímto procesem dostaneme pro jednotlivé linky počty škod, ke kterým by došlo, kdyby ve všech letech byla stejná expozice, tedy stejné riziko.

²Expozici můžeme měřit i jinak, např. u krátkodobých smluv je vhodné zvolit počet dnů trvání pojistky



Obrázek 4.5: Expozice

Poznamenejme, že výběr základní expoziční musí být proveden s rozvahou. Její hodnota by neměla být příliš malá, abychom nemuseli odstraňovat mnoho škod. Naopak při velké hodnotě musíme zřejmě zcela vyloučit všechny roky, pro které je reálná expoziční menší než zvolená základní. Tabulka 4.3 shrnuje výběr základní expoziční a roky, které byly ponechány pro další analýzu.

Položka	Číslo linky					
	03	05	18	32	34	55
základní expoziční	747	4 803	327 042	8 175	747	527
data nově od roku	2007	2003	2006	2009	2007	2003
data nově do roku	2016	2016	2015	2016	2016	2016

Tabulka 4.3: Nastavení základní expoziční

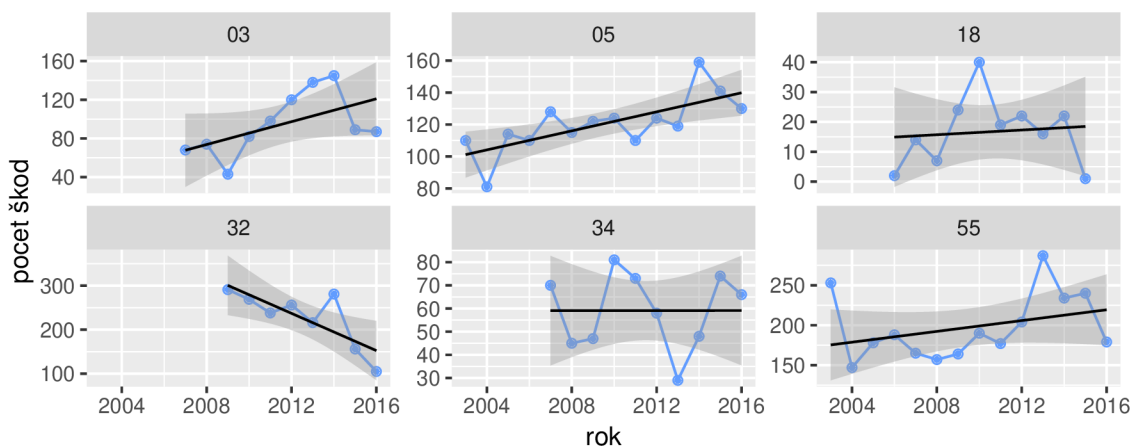
Celkem dostáváme počty škod sepsané v tabulce 4.4. Pro každou linku navíc máme k dispozici i údaj o tom, jak jednotlivé úpravy ovlivnily stavy dat.

Položka	Číslo linky					
	03	05	18	32	34	55
počet škod	1 616	3 396	1 153	5 612	1 321	3 527
nulové + run-off	377	1 077	608	2 527	544	285
expoziční	295	632	378	1 273	186	479
finální počet škod	944	1 687	167	1 812	591	2 763

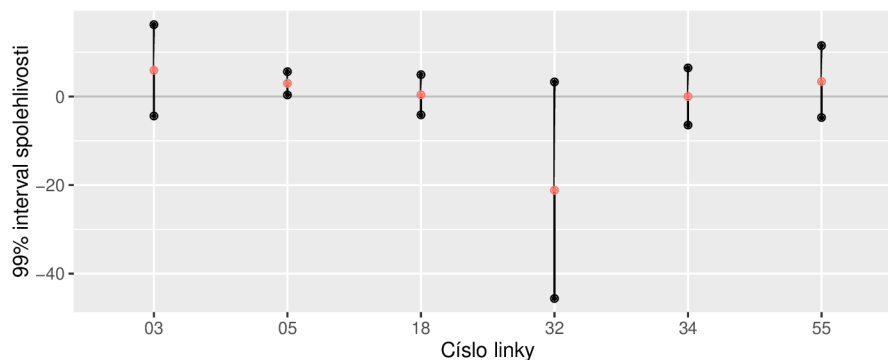
Tabulka 4.4: Finální počet škod

Na závěr tohoto odstavce je vhodné zhodnotit, zda korekce dat podle expoziční přispěla k odstranění trendů v počtech škod. Konečný počet škod po všech úpravách rozdělený podle linek a škodních let vidíme na obrázku 4.6 (tab. A.5 přílohy). Do grafu je zanesena i regresní přímka a 95% konfidenční intervaly. Již z tohoto obrázku je patrné, že škody nevykazují takovou trendovost jako v 4.1. Graf 4.7 (tab. A.6 přílohy) shrnuje výsledky

t-testu o významnosti směrnice regresní přímky na 99% hladině spolehlivosti. Ukazuje se, že kromě linky 05 nemůžeme na této hladině zamítnout nulovost směrnice regresní přímky a tedy nepřítomnosti trendu. U občanské odpovědnosti je p-hodnota t-testu rovna 0.004728, což se od testované 1% hladiny významnosti řádově neliší, a proto i tuto linku ponecháme v dalších analýzách.



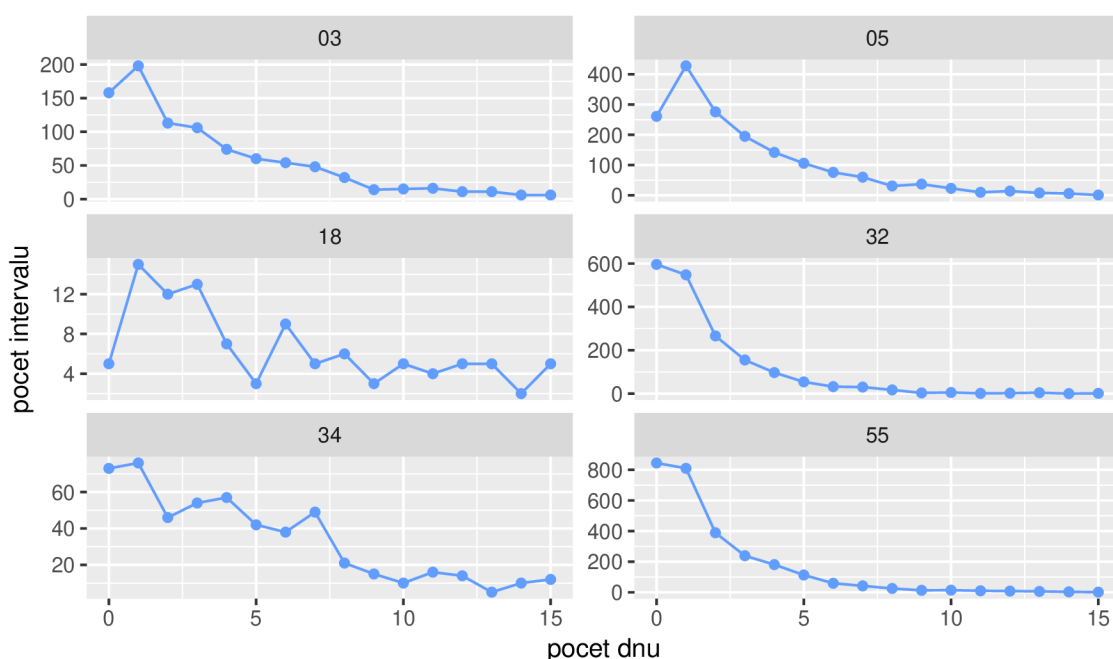
Obrázek 4.6: Finální počet škod po letech po očištění



Obrázek 4.7: Analýza trendu finálního počtu škod

4.3 Zpracování a výsledky

Nyní máme data připravená k testování. Na grafu 4.8 (tab. A.7 přílohy) vidíme absolutní četnosti intervalů mezi dvěma po sobě následující událostmi podle počtu dnů mezi nimi. Neuvažujeme přitom interval od počátku měření do první události ani interval od poslední události do konce měření. U všech linek je patrný klesající trend. U těch z nich, ve kterých máme k dispozici více dat, si můžeme od intervalu jedna povšimnout poklesu monotónního.



Obrázek 4.8: Absolutní četnosti intervalů mezi škodami

Na následujících stranách odhadneme parametry rozdělení popsanych v odstavci 3.2 metodou maximální věrohodnosti a metodou minimálního χ^2 . Pomocí χ^2 testu dobré shody následně prověříme, jak dobře jednotlivá rozdělení fitují empirická data. Diskretizaci teoretických rozdělení provedeme metodou centrální diference, tzn. relativní četnosti intervalu nula budeme porovnávat s integrálem hustoty rozdělení od nuly do 0.5, relativní četnosti intervalu jedna s integrálem hustoty rozdělení od 0.5 do 1.5 atd.

Je vhodné poznamenat, že stanovení dělicích bodů pro diskretizaci není v případě dat, která máme k dispozici, jednoduchým úkolem. Demonstrujme tuto skutečnost na příkladu. Představme si, že u dvou po sobě následujících škod známe i čas jejich nastání. Předpokládejme, že první z nich se stala v pondělí v 6:15 a druhá v úterý ve 23:39. Obě škody dělí přesně 1.725 dne. Uvážíme-li rozdělení pomocí centrální diference, pak tato doba mezi událostmi spadá do intervalu (1.5, 2.5). Pokud bychom vzali pouze den nastání – tedy informaci v dostupných datech skutečně viditelnou – pak je doba mezi událostmi rovna jednomu dni, a proto ji zařadíme do intervalu (0.5, 1.5). Přestože diskretizací přes centrální diferenci situaci zjednodušíme, u této metody zůstaneme. Navíc může zřejmě dojít i k opačné záměně, tj. pozorujeme dvoudenní interval mezi událostmi, který by při znalosti přesného času škod, které jej ohraničují, spadl do diskretizačního intervalu (0.5, 1.5).³

³Pro představu uvažme například interval mezi škodami, které nastaly v pondělí ve 21:00 a ve středu ve 3:00 (tj. 1.25 dne)

4.3.1 Metoda maximální věrohodnosti

Metoda maximální věrohodnosti je velmi často užívána pro odhad parametrů. V odstavci 3.2 jsme ukázali vztahy, kterými je možno tyto odhady pro jednotlivá rozdělení získat. Metodu maximální věrohodnosti bychom mohli beze zbytku využít, pokud bychom měli u každé škody uveden i přesný čas a doby mezi událostmi by tak byly přibližně spojitě.⁴ Bohužel data, která máme k dispozici, specifikují pouze den nastání škody, což s sebou nese dodatečný problém, který je nutno řešit.

Vezměme si např. jednodenní interval mezi škodami. Počet dat, která do něj spadají, se rovná počtu dvojic událostí, které nastaly ve dvou po sobě následujících dnech. Každé toto jednotlivé pozorování vstupuje do odhadu parametrů metodou maximální věrohodnosti hodnotou jedna. Ve skutečnosti však může být doba mezi popsányými událostmi i menší (pondělí 21:00 a úterý 3:00, tj. 0.25 dne) či větší (pondělí 3:00 a úterý 21:00, tj. 1.75 dne) než jedna. Limitně mohou být hodnoty pozorování v tomto případě v intervalu od nuly do dvou.

Obdobná logika platí i pro vícedenní intervaly. Máme-li v datech dvoudenní odstup škod, pak původně od sebe mohly být tyto škody vzdáleny minimálně jeden a maximálně tři dny atp. Hodnota, kterou disponujeme v rámci dat, je tedy středem intervalu, ve kterém se příslušná skutečná doba mezi škodami mohla realizovat. Pro intervaly 1, 2, 3, ... proto v metodě maximální věrohodnosti využijeme právě tyto hodnoty.

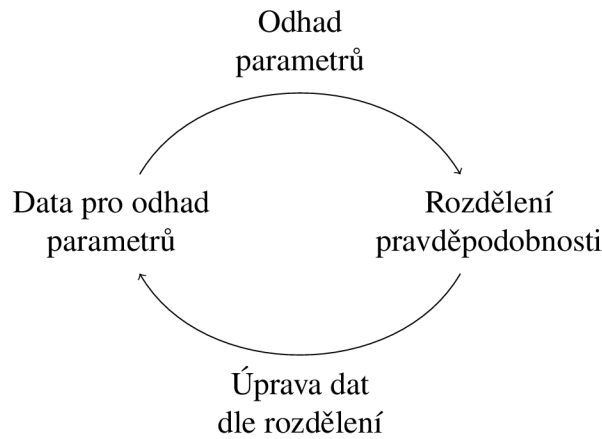
Jiná situace je u škod, které nastaly během jediného dne. Skutečný odstup takových škod může být pouze od nuly do jedné, jelikož čas nemůže být záporný. Pro tato pozorování budeme opět uvažovat střed intervalu, ve kterém se mohla ve skutečnosti realizovat, a proto v datech nahradíme nulové body hodnotou 0.5.

Modifikovaná metoda maximální věrohodnosti

Závěrem tohoto odstavce se zamysleme nad tím, jak obecně pro metodu maximální věrohodnosti upravit data podobná těm, se kterými nyní pracujeme. Nabízí se myšlenka nahradit každé pozorování střední hodnotou odhadnutého rozdělení „useknutého“⁵ na intervalu, ve kterém se toto pozorování mohlo ve skutečnosti realizovat. Bohužel parametry rozdělení můžeme odhadnout až ve chvíli, kdy přesně známe hodnoty, které do věrohodnostní funkce vstupují. Tento problém je schématicky naznačen v diagramu na následující straně. Nic nás ovšem neomezuje oba kroky udělat současně. To znamená, že do věrohodnostní funkce nebudou vstupovat konkrétní data, ale funkce, která bude také závislá na odhadovaném parametru.

⁴V reálném světě nejsme ani nejdokonalějšími přístroji schopni měřit čas spojitě. Pokud bychom ale měli k dispozici čas nastání události s přesností na minuty, pak by se pro účely analýz prováděných v této práci, dalo hovořit o tom, že intervaly mezi škodami jsou přibližně výběrem ze spojitého rozdělení pravděpodobnosti.

⁵Viz angl. „truncated distributions“



Jelikož máme diskrétní data s realizacemi z nezáporných celých čísel, můžeme věrohodnostní funkci pro rozdělení s hustotou f psát jako

$$L = \prod_{i=1}^n f(y_i) = \prod_{k=0}^{\infty} [f(k)]^{n_k},$$

kde y_i je i -té pozorování, n jejich celkový počet a n_k značí počet těch z nich, která se realizovala hodnotou k . Mějme nezápornou spojitou náhodnou veličinu X z rozdělení \mathcal{F} s jedním parametrem λ . Nechť má tato veličina hustotu pravděpodobnosti f a distribuční funkci F . Dále definujme náhodnou veličinu X_k pro $k \in \mathbb{N}$, která vznikne „useknutím“ rozdělení \mathcal{F} na intervalu $(k-1, k+1)$. Označme $X_k \sim Tr\mathcal{F}(\lambda; k-1, k+1)$. Pro úplnost dodefinujme $X_0 \sim Tr\mathcal{F}(\lambda; 0, 1)$. Nyní můžeme podle předchozích úvah zapsat věrohodnostní funkci ve tvaru

$$L = \prod_{k=0}^{\infty} f(EX_k)^{n_k} = \prod_{k=0}^{\infty} L_k^{n_k},$$

kde L_k je pouze označení pro snazší zápis. Snadno určíme i logaritmus věrohodnostní funkce jako

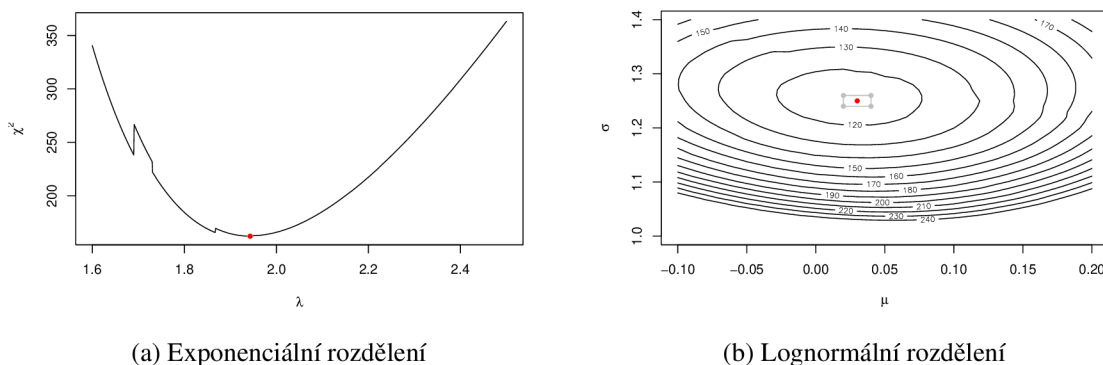
$$l = \ln L = \sum_{k=0}^{\infty} n_k \ln L_k = \sum_{k=0}^{\infty} n_k \ln l_k.$$

Nyní již standardním postupem pomocí minimalizace funkce l určíme maximálně věrohodný odhad parametru λ rozdělení \mathcal{F} . V příloze B lze nalézt konkrétní výpočet pro exponenciální rozdělení.

Máme-li konkretizováno rozdělení, pak můžeme řešit i problém s rozřazením dat do diskretizačních intervalů. Nasimulujeme pro $k \in \mathbb{N}_0$ vždy n_k hodnot z příslušného „useknutého“ rozdělení. Teprve tato čísla rozřadíme do zvolených třídících intervalů. Tímto postupem se přiblížíme k datům, která bychom získali, pokud bychom dokázali okamžik vzniku události lépe specifikovat. Je to proto, že tento proces reflektuje proměnnou intenzitu nastání události v rámci jednotlivých diskretizačních intervalů.

4.3.2 Metoda minimálního χ^2

Odhady parametrů rozdělení metodou minimálního χ^2 budeme oproti předchozímu případu získávat vždy numericky. Výhodou této metody je skutečnost, že nepotřebujeme znát konkrétní časy mezi událostmi a zcela nám postačí pouze diskrétní data, která máme k dispozici. Oproti [1, podkapitola 12.3], kde je tato metoda popsána, budeme v průběhu numerického výpočtu pro testované vektory parametrů⁶ měnit počet tříd intervalů. Připomeňme, že podle odstavce 3.3.1 je nutné, aby každá třída měla teoretickou četnost větší než pět. Z tohoto důvodu nemusí být minimalizované funkce spojité, jak je vidět na grafu 4.9. Pro lognormální rozdělení byl pro přehlednost zvolen vrstevnicový graf. Poznamenejme, že obdélník vymezený šedou čarou reprezentuje přesnost odhadu. Jinými slovy, skutečné minimum je ve vnitřku této oblasti.



Obrázek 4.9: Pearsonova testová statistika vybraných rozdělení (linka 55)

Jak uvidíme při srovnání metody minimálního χ^2 s metodou maximální věrohodnosti, dává prvně jmenovaná lepší výsledky. Tuto skutečnost lze intuitivně očekávat, jelikož v metodě minimálního χ^2 minimalizujeme testovací statistiku, kterou využíváme pro rozhodování o tom, jestli dané rozdělení dobře fituje data. Je vhodné poznamenat, že porovnání obou zmíněných metod provádíme na základě p-hodnoty. Z tohoto pohledu ale nemusí být metoda minimálního χ^2 vždy lepší. V některých případech je možné, že odhad parametru resp. parametrů s Pearsonovou statistikou vyšší než minimální má i vyšší p-hodnotu kvůli většímu počtu stupňů volnosti příslušného χ^2 rozdělení. Tato situace se ovšem objevuje jen zřídka.

4.3.3 Srovnání metod

V poslední části této kapitoly představíme výsledky obou metod popsanych v předchozím textu. Prozkoumáme, zda navržená rozdělení mohou být těmi, ze kterých empirická data

⁶Resp. testované hodnoty parametru v případě jednoparametrického exponenciálního rozdělení

pochází. Testy budeme provádět na 1% hladině významnosti. V případě, že identifikujeme více vhodných rozdělení, navrhneme na základě p-hodnoty to z nich, které popisuje data nejlépe.

Pojištění dopravy – majetek (03)

V případě majetkového pojištění dopravy nemůžeme zamítnout, že se intervaly mezi událostmi řídí Weibullovým, gama nebo Paretovým rozdělením s parametry specifikovanými v tabulce 4.5. Tato rozdělení nezamítá metoda minimálního χ^2 . Metoda maximální věrohodnosti připouští pouze Pareto rozdělení pravděpodobnosti. Uvážíme-li pouze rozdělení, která testujeme, pak podle velikosti p-hodnoty empirická data nejlépe popisuje gama rozdělení s parametry odhadnutými metodou minimálního χ^2 . Na grafu 4.10c jsou sloučeny třídy s teoretickou četností menší než pět.

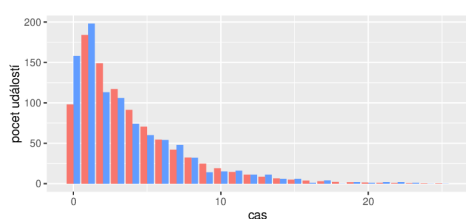
Param.	Rozdělení pravděpodobnosti				
	Exp	Wb	Γ	LN	Pa II
k		1.002	1.087		
λ	3.629	3.946	3.628		
μ				0.846	
σ				1.052	35.388
α					9.976
p-hod.	1.76E-04	1.24E-05	2.25E-09	6.66E-31	0.017

(a) MLE odhady

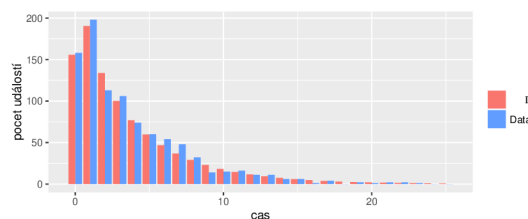
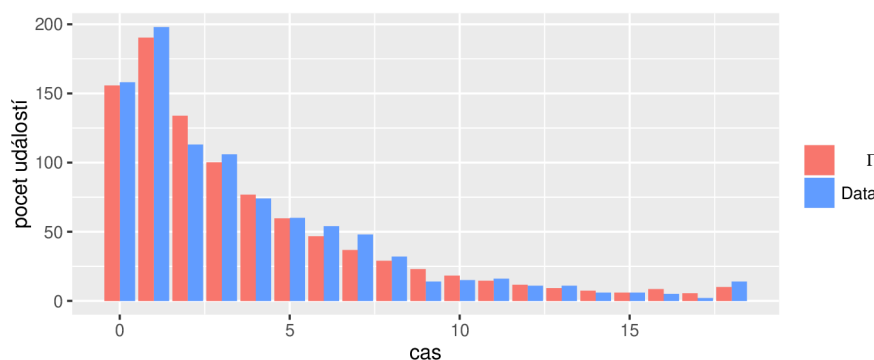
Param.	Rozdělení pravděpodobnosti				
	Exp	Wb	Γ	LN	Pa II
k		0.874	0.809		
λ	3.629	3.67	4.761		
μ				0.716	
σ				1.27	23.513
α					7.234
p-hod.	1.76E-04	0.258	0.264	1.38E-11	0.021

(b) Min. χ^2 odhady

Tabulka 4.5: Odhady parametrů (03)



(a) MLE fit

(b) Min. χ^2 fit(c) Min. χ^2 fit – sloučení tříd

Obrázek 4.10: Srovnání teoretického rozdělení s daty (03)

Občanská odpovědnost (05)

U pojištění občanské odpovědnosti se plně projevují rozdíly mezi metodami pro odhad parametrů. Nejlepší volbou pro výběr rozdělení je Weibullovo a Paretovo rozdělení s parametry odhadnutými metodou minimálního χ^2 . Druhé jmenované má při zvýšení přesnosti větší p-hodnotu.

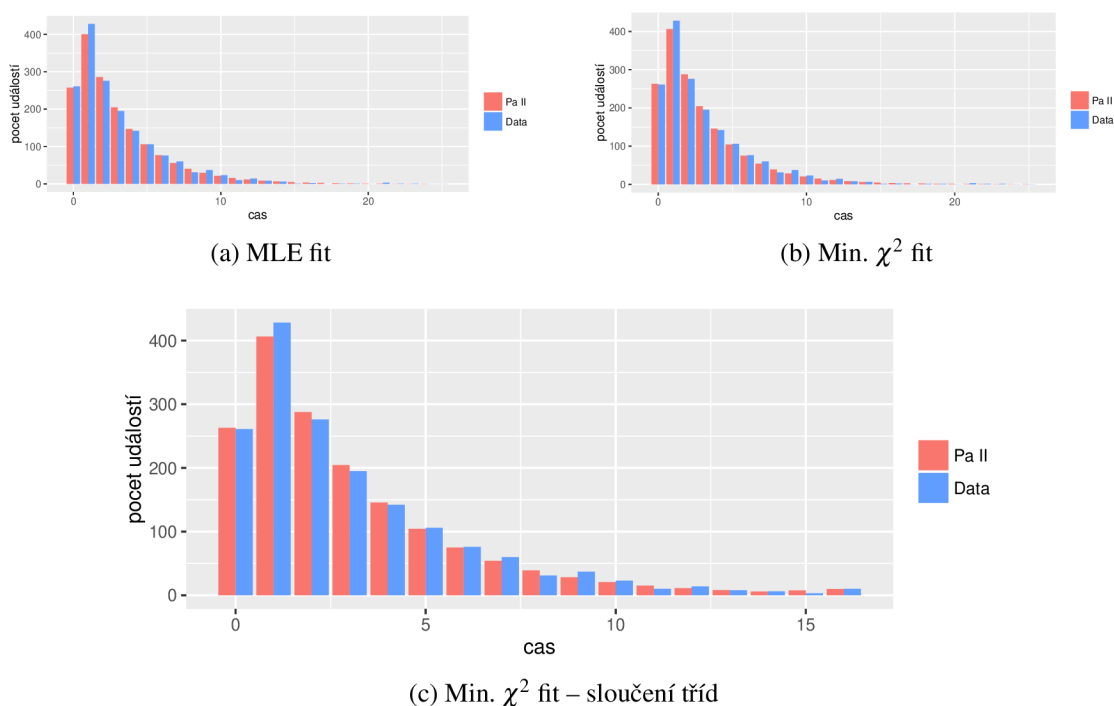
Param.	Rozdělení pravděpodobnosti				
	Exp	Wb	Γ	LN	Pa II
k		1.096	1.292		
λ	3.106	3.23	2.404		
μ				0.699	
σ				0.938	99.934
α					33.176
p-hod.	0.440	1.27E-05	1.99E-14	9.23E-46	0.222
	✓				✓

(a) MLE odhady

Param.	Rozdělení pravděpodobnosti				
	Exp	Wb	Γ	LN	Pa II
k		0.975	0.963		
λ	2.951	3.004	3.154		
μ				0.614	
σ				1.155	88.739
α					30.177
p-hod.	0.491	0.541	0.511	5.16E-12	0.541
	✓	✓	✓		⊗

(b) Min. χ^2 odhady

Tabulka 4.6: Odhady parametrů (05)



Obrázek 4.11: Srovnání teoretického rozdělení s daty (05)

Pojištění úvěrů (18)

V pojištění úvěrů máme ze všech linek nejméně dat a intervaly mezi událostmi jsou největší. Testy jsou proto značně benevolentnější než v ostatních případech. Tato skutečnost má dopad i na jejich výsledky, kdy zamítáme pouze lognormální rozdělení s parametry

odhadnutými metodou maximální věrohodnosti. Nejvyšší p-hodnoty dosahuje Weibullovo rozdělení s parametry vypočtenými právě touto metodou. Je to příklad situace, kdy metoda maximální věrohodnosti může poskytnout lepší odhad parametrů co do velikosti p-hodnoty χ^2 testu, než metoda minimálního χ^2 . Připomeňme, že tato skutečnost je dána rozdílnými počty tříd intervalů.

Navzdory těmto faktům by bylo vhodné vybrat Paretovo rozdělení s min. χ^2 parametry, které má druhou nejvyšší p-hodnotu. Tento výběr můžeme podložit argumentem, že intervaly mezi škodami v pojištění úvěrů mohou být dlouhé řádově až několik měsíců. Paretovo rozdělení má oproti ostatním kandidátům nejtěžší pravý konec, jelikož klesá polynomiálně. Po zvážení tohoto faktu bychom vybrali právě toto rozdělení.

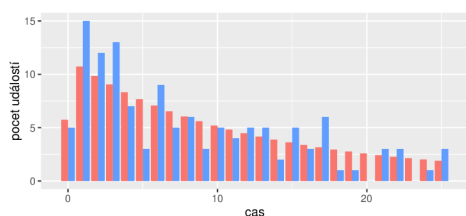
Param.	Rozdělení pravděpodobnosti				
	Exp	Wb	Γ	LN	Pa II
k		0.854	0.806		
λ	18.72	17.19	23.224		
μ				2.194	
σ				1.342	55.161
α					3.89
p-hod.	0.033	0.286	0.151	0.003	0.037
	✓	⊗	✓		✓

(a) MLE odhady

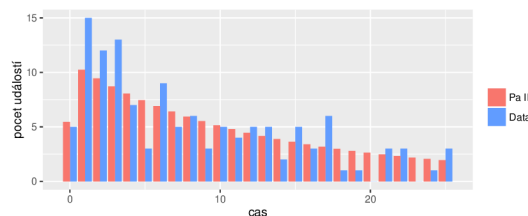
Param.	Rozdělení pravděpodobnosti				
	Exp	Wb	Γ	LN	Pa II
k		0.756	0.733		
λ	17.495	14.467	24.325		
μ				2.065	
σ				1.409	61.025
α					4.091
p-hod.	0.095	0.240	0.229	0.097	0.272
	✓	✓	✓	✓	✓

(b) Min. χ^2 odhady

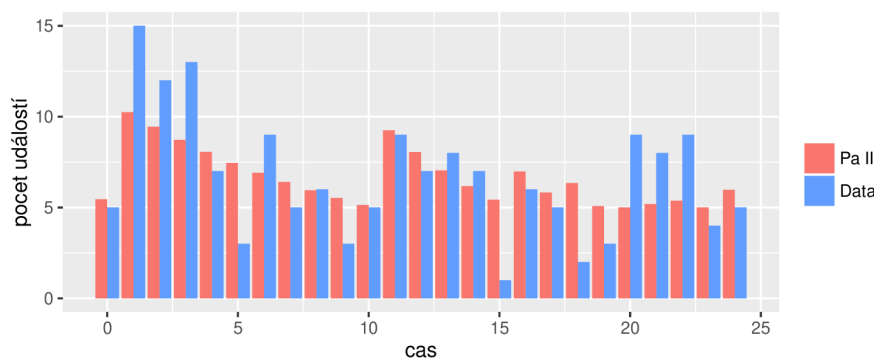
Tabulka 4.7: Odhady parametrů (18)



(a) MLE fit



(b) Min. χ^2 fit



(c) Min. χ^2 fit – sloučení tříd

Obrázek 4.12: Srovnání teoretického rozdělení s daty (18)

Profesní odpovědnost (32)

U pojištění profesní odpovědnosti je jasně nejvhodnější volbou rozdělení gama s parametry odhadnutými minimálním χ^2 . Na grafu 4.13, a to zejména na prvních třech intervalech, vidíme, že v rámci tohoto rozdělení dostáváme oproti metodě maximální věrohodnosti výrazně lepší odhad.

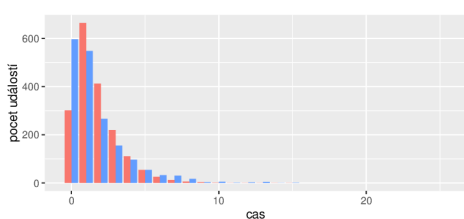
Param.	Rozdělení pravděpodobnosti				
	Exp	Wb	Γ	LN	Pa II
k		1.14	1.459		
λ	1.777	1.877	1.218		
μ				0.195	
σ				0.828	64.499
α					37.301
p-hod.	1.69E-15	3.64E-52	4.18E-94	1.2E-127	3.26E-13

(a) MLE odhady

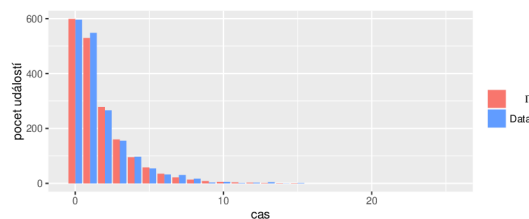
Param.	Rozdělení pravděpodobnosti				
	Exp	Wb	Γ	LN	Pa II
k		0.849	0.731		
λ	1.634	1.506	2.268		
μ				-0.079	
σ				1.226	6.594
α					4.89
p-hod.	1.08E-13	0.289	0.393	3.65E-09	0.001

(b) Min. χ^2 odhady

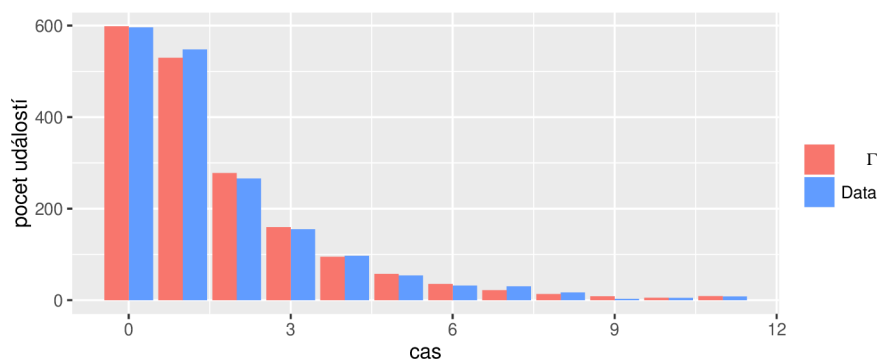
Tabulka 4.8: Odhady parametrů (32)



(a) MLE fit



(b) Min. χ^2 fit



(c) Min. χ^2 fit – sloučení tříd

Obrázek 4.13: Srovnání teoretického rozdělení s daty (32)

Pojištění dopravy – odpovědnost (34)

V odpovědnostním pojištění dopravy se – jako v jediné lince – nepodařilo prokázat dobrou fitaci dat některým z uvažovaných rozdělení. Na druhou stranu nejvyšší dosažené p-hodnoty u Weibullova a gama rozdělení se k jednocentní hladině významnosti velmi blíží. Tato skutečnost může být spojena s charakterem analyzovaného pojištění. Jak již bylo řečeno, pojištění dopravy kryje i předměty na výstavách, veletrzích apod. Tyto akce se konají v konkrétním termínu a je na ně dopravováno a na nich vystavováno mnoho exponátů. Z tohoto důvodu se můžeme domnívat, že intervaly mezi škodami nejsou zcela nezávislé, což procesy obnovy nepřipouští.

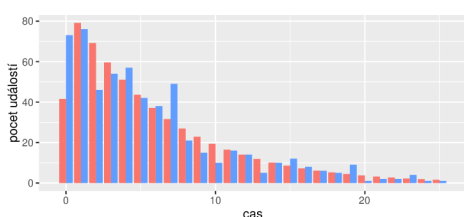
Param.	Rozdělení pravděpodobnosti				
	Exp	Wb	Γ	LN	Pa II
k		1.002	1.046		
λ	6.221	6.227	5.948		
μ				1.279	
σ				1.136	112.183
α					19.032
p-hod.	2.03E-04	1.06E-04	8.98E-06	7.15E-27	4.29E-04

(a) MLE odhady

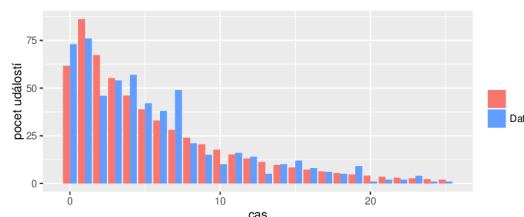
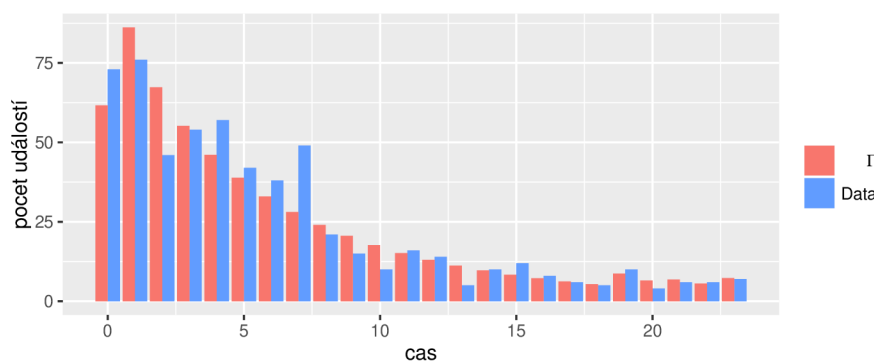
Param.	Rozdělení pravděpodobnosti				
	Exp	Wb	Γ	LN	Pa II
k		0.916	0.852		
λ	6.159	5.959	7.284		
μ				1.185	
σ				1.324	103.889
α					17.816
p-hod.	2.08E-04	0.003	0.006	3.17E-16	4.52E-04

(b) Min. χ^2 odhady

Tabulka 4.9: Odhady parametrů (34)



(a) MLE fit

(b) Min. χ^2 fit(c) Min. χ^2 fit – sloučení tříd

Obrázek 4.14: Srovnání teoretického rozdělení s daty (34)

Pojištění hospodářských zvířat (55)

V pojištění hospodářských zvířat můžeme na obrázku 4.15 opět srovnat rozdíl mezi metodami pro odhad parametrů. Jako statisticky významné se ukazuje být Weibullovo a gama rozdělení s parametry odhadnutými metodou minimálního χ^2 , přičemž první jmenované má větší p-hodnotu.

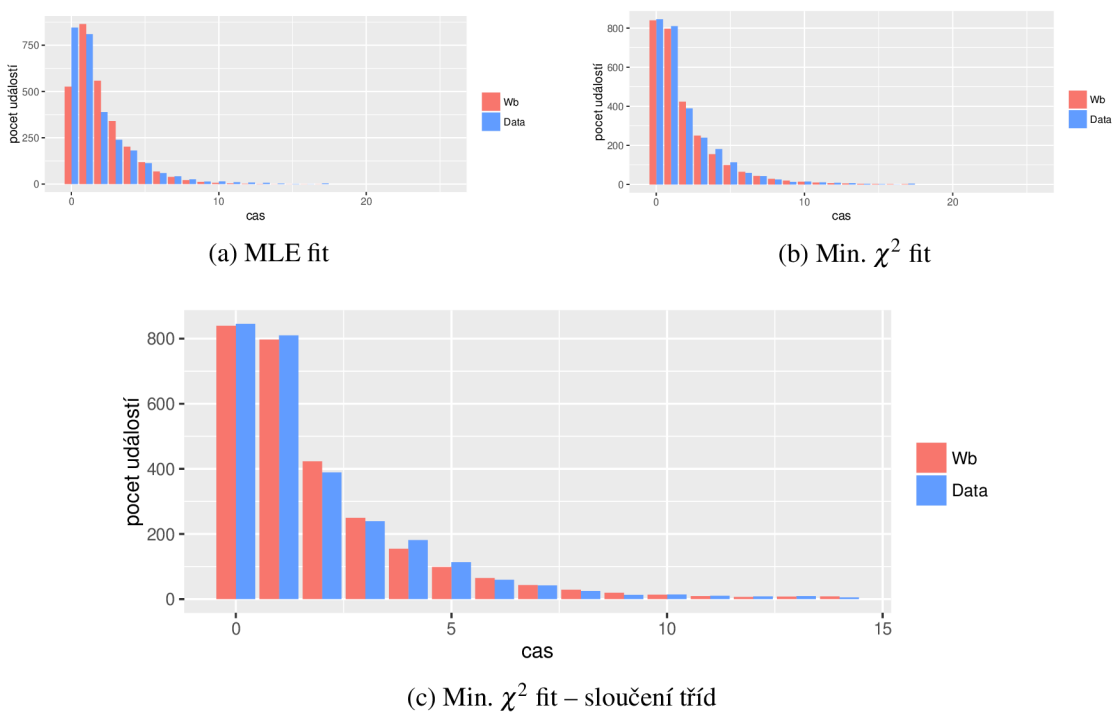
Param.	Rozdělení pravděpodobnosti				
	Exp	Wb	Γ	LN	Pa II
k		1.091	1.336		
λ	2.002	2.081	1.499		
μ				0.275	
σ				0.875	28.031
α					15.003
p-hod.	7.60E-30	6.34E-73	9.9E-132	3.0E-170	7.16E-18

(a) MLE odhady

Param.	Rozdělení pravděpodobnosti				
	Exp	Wb	Γ	LN	Pa II
k		0.826	0.715		
λ	1.943	1.709	2.642		
μ				0.025	
σ				1.254	6.288
α					4.225
p-hod.	4.80E-29	0.245	0.198	8.32E-16	2.86E-06

(b) Min. χ^2 odhady

Tabulka 4.10: Odhady parametrů (55)



Obrázek 4.15: Srovnání teoretického rozdělení s daty (55)

Závěr

V teoretické části jsme prodiskutovali základní témata, která jsou spojena se stochastickými procesy, přičemž jsme věnovali zvýšenou pozornost Markovovým řetězcům a zejména potom Poissonovu procesu. Dalším bodem našeho zájmu pak byly procesy obnovy a s nimi úzce spjatá rozdělení intervalů mezi událostmi. V rámci aplikační části jsme se zevrubně zabývali zpracováním empirických dat a jejich analýzou. U většiny zkoumaných případů se nám podařilo prokázat, že se statisticky významnou pravděpodobností splňují charakter procesů obnovy.

V průběhu zpracování reálných dat jsme se setkali s několika problémy. Byly jimi například jejich hrubost, která vedla k limitovanému využití statistických metod, nebo nutnost odstranění některých pozorování při jejich přípravě, jejímž logickým následkem je ztráta informace. Rozšířením této práce by mohl být podrobný průzkum právě těchto fenoménů. Je nutno poznamenat, že velká hrubost dat nebyla způsobena jejich nevhodným výběrem, ale plošně zavedenou praxí, kterou pojišťovny standardně využívají při likvidaci. Právě z tohoto důvodu je zde velký prostor pro rozvoj specifických metod, které by se dokázaly vypořádat s tímto typem datové struktury.

Logickou nadstavbou teoretické části by jistě byla zevrubná analýza obou zmíněných modelů z oblasti teorie ruinování, tedy Cramérova-Lunbergova a Sparre Andersenova.

Příloha

A Datové podklady pro vybrané grafy čtvrté kapitoly

Rok vzniku	Číslo linky						
	03	05	18	32	34	54	55
2003	42	134	1	288	51	86	265
2004	44	123	0	323	55	128	191
2005	77	151	1	341	55	133	219
2006	81	145	26	399	56	170	235
2007	87	179	54	376	107	207	218
2008	88	169	22	493	81	227	238
2009	63	188	59	515	111	289	214
2010	114	246	87	518	130	177	236
2011	146	269	195	437	161	245	229
2012	171	293	247	464	110	350	271
2013	204	311	284	391	103	284	394
2014	222	450	168	478	85	221	295
2015	145	377	9	348	115	105	303
2016	132	361	0	241	101	110	219
celkem	1 616	3 396	1 153	5 612	1 321	2 732	3 527

Tabulka A.1: Počet škod podle linek a let vzniku

Měsíc vzniku	Číslo linky						
	03	05	18	32	34	54	55
1	102	305	82	637	82	56	415
2	131	345	86	477	100	156	245
3	135	354	86	504	106	64	299
4	122	361	85	516	137	93	307
5	156	379	92	514	106	438	276
6	120	402	95	534	116	622	368
7	160	378	91	444	142	1 067	513
8	140	433	117	483	120	419	375
9	171	397	111	496	129	38	266
10	163	402	127	517	98	13	290
11	156	380	87	454	129	0	294
12	133	307	94	482	84	2	255

Tabulka A.2: Počet škod podle linek a měsíců vzniku

	Číslo linky						
	03	05	18	32	34	54	55
min	121 %	117 %	132 %	126 %	126 %	431 %	158 %
max	72 %	82 %	85 %	88 %	73 %	0 %	75 %

Tabulka A.3: Max. a min. odchylka dat od průměru

Rok	Číslo linky						
	03	05	18	32	34	55	
2003	326	4 803	3 550 373	3 866	326	527	
2004	475	4 870	5 405 501	4 798	475	625	
2005	540	5 037	5 507 679	5 735	540	612	
2006	602	4 964	2 430 991	6 575	602	641	
2007	747	4 988	1 033 104	7 023	747	657	
2008	832	5 185	758 693	7 630	832	705	
2009	821	5 834	545 359	8 175	821	671	
2010	793	6 805	327 042	8 590	793	649	
2011	829	7 637	1 109 982	8 666	829	633	
2012	833	7 705	2 025 600	8 600	833	651	
2013	816	8 031	1 977 021	8 517	816	654	
2014	845	8 315	1 795 213	8 680	845	616	
2015	863	8 156	1 721 307	8 825	863	611	
2016	767	8 306	1 154 210	9 106	767	571	

Tabulka A.4: Expozice

Rok vzniku	Číslo linky					
	03	05	18	32	34	55
2003	-	110	-	-	-	253
2004	-	81	-	-	-	147
2005	-	114	-	-	-	178
2006	-	110	2	-	-	188
2007	68	128	14	-	70	165
2008	74	115	7	-	45	157
2009	43	122	24	291	47	164
2010	82	124	40	269	81	190
2011	98	110	19	238	73	177
2012	120	124	22	256	58	204
2013	138	119	16	216	29	287
2014	145	159	22	281	48	234
2015	89	141	1	156	74	240
2016	87	130	-	105	66	179
celkem	944	1 687	167	1 812	591	2 763

Tabulka A.5: Finální počet škod po letech po očištění

Položka	Číslo linky					
	03	05	18	32	34	55
směrnice přímky	5.915	2.978	0.394	- 21.167	0.006	3.382
dolní mez	- 4.399	0.348	- 4.148	- 45.622	- 6.441	- 4.745
horní mez	16.229	5.608	4.936	3.289	6.454	11.51
p-hodnota	0.091	0.005	0.778	0.018	0.998	0.228

Tabulka A.6: Analýza trendu finálního počtu škod

Počet dní	Číslo linky					
	03	05	18	32	34	55
0	158	261	5	596	73	845
1	198	428	15	548	76	810
2	113	276	12	266	46	389
3	106	195	13	155	54	239
4	74	142	7	97	57	181
5	60	106	3	54	42	113
6	54	76	9	32	38	59
7	48	60	5	30	49	42
8	32	31	6	17	21	25
9	14	37	3	3	15	13
10	15	23	5	5	10	14
11	16	10	4	1	16	10
12	11	14	5	2	14	8
13	11	8	5	4	5	6
14	6	6	2	0	10	3
15	6	1	5	1	12	1
16	1	2	3	0	8	0
17	4	0	6	0	6	4
18	0	1	1	0	5	0
19	2	1	1	0	9	0
20	1	0	0	0	1	0
21	2	3	3	0	2	0
22	2	1	3	0	2	0
23	1	1	0	0	4	0
24	0	0	1	0	1	0
25	0	0	3	0	1	0
26	0	0	1	0	2	0
27	2	1	1	0	2	0
28	3	0	0	0	0	0
29	1	0	1	0	2	0
30	1	0	0	0	0	0
31	0	0	0	0	2	0
32	0	1	0	0	0	0
33	0	0	2	0	1	0
34	0	0	1	0	1	0
35	0	0	2	0	0	0
36	1	0	1	0	0	0
37	0	0	2	0	0	0
38	0	0	2	0	0	0
39	0	0	2	0	0	0
40	0	1	1	0	1	0
41	0	0	1	0	0	0
42	0	0	1	0	1	0
43	0	0	2	0	0	0
44	0	0	2	0	0	0
45	0	0	1	0	0	0
46	0	0	0	0	0	0
47	0	0	2	0	0	0
48	0	0	3	0	0	0
49	0	0	1	0	0	0
50	0	0	2	0	1	0
...
55	0	0	1	0	0	0
63	0	0	1	0	0	0
73	0	0	1	0	0	0
74	0	0	1	0	0	0
76	0	0	1	0	0	0
83	0	0	1	0	0	0
92	0	0	1	0	0	0
95	0	0	1	0	0	0
104	0	0	1	0	0	0
123	0	0	1	0	0	0

Tabulka A.7: Počet dní mezi škodami

B Modifikovaná metoda maximální věrohodnosti

Tato část přílohy pojednává o odhadu parametru λ exponenciálního rozdělení modifikovanou metodou maximální věrohodnosti, která byla nastíněna v odstavci 4.3.1. Značení zde použité odpovídá konvenci zavedené v téže části textu.

- Základní rozdělení

$$X \sim \text{Exp}\left(\frac{1}{\lambda}\right)$$

$$f(x) = \lambda e^{-\lambda x}$$

$$F(x) = 1 - e^{-\lambda x}$$

- Useknutá rozdělení

$$X_0 \sim \text{TrExp}\left(\frac{1}{\lambda}; 0, 1\right)$$

$$f_{X_0}(x) = \frac{\lambda e^{-\lambda x}}{F(1) - F(0)} = \frac{\lambda e^{-\lambda x}}{1 - e^{-\lambda}}$$

$$X_k \sim \text{TrExp}\left(\frac{1}{\lambda}; k-1, k+1\right), \quad k \in \mathbb{N}$$

$$f_{X_k}(x) = \frac{\lambda e^{-\lambda x}}{F(k+1) - F(k-1)} = \frac{\lambda e^{-\lambda x}}{e^{-\lambda(k-1)} - e^{-\lambda(k+1)}}$$

- Výpočet střední hodnoty useknutých rozdělení

$$EX_0 = \int_0^1 \frac{\lambda x e^{-\lambda x}}{1 - e^{-\lambda}} dx = \dots = \frac{1}{\lambda} + \frac{1}{1 - e^{-\lambda}}$$

$$EX_k = \int_{k-1}^{k+1} \frac{\lambda x e^{-\lambda x}}{e^{-\lambda(k-1)} - e^{-\lambda(k+1)}} dx = \dots = \frac{1}{\lambda} + (k-1) + \frac{2}{1 - e^{2\lambda}}$$

- Výpočet derivace logaritmické věrohodnostní funkce

$$\frac{\partial l_0}{\partial \lambda} = \frac{\partial}{\partial \lambda} \ln(\lambda e^{-\lambda EX_0}) = \frac{\partial}{\partial \lambda} \left(\ln \lambda - 1 + \frac{\lambda}{e^\lambda - 1} \right) = \frac{1}{\lambda} + \frac{e^\lambda(1-\lambda) - 1}{(e^\lambda - 1)^2}$$

$$\begin{aligned} \frac{\partial l_k}{\partial \lambda} &= \frac{\partial}{\partial \lambda} \ln(\lambda e^{-\lambda EX_k}) = \frac{\partial}{\partial \lambda} \left(\ln \lambda - 1 - \lambda(k-1) + \frac{2\lambda}{e^{2\lambda} - 1} \right) = \\ &= \frac{1}{\lambda} - (k-1) + 2 \frac{e^{2\lambda}(1-2\lambda) - 1}{(e^{2\lambda} - 1)^2} \end{aligned}$$

- Shrnutí

Odhad parametru λ dostaneme vyřešením rovnice

$$\frac{\partial l}{\partial \lambda} = \sum_0^{\infty} n_k \frac{\partial l_k}{\partial \lambda} = 0.$$

Snadno lze nahlédnout, že i v případě exponenciálního rozdělení je nutné tuto rovnici řešit numericky. U složitějších typů rozdělení s více parametry by se obtížnost úlohy ještě zvýšila.

C Zdrojový kód vybraných grafů v softwaru R

```
library(ggplot2) #grafika
library(grid) # grafika
library(Renext) #Lomax rozdeleni

#####
##### KAPITOLA 1 #####
#####

# stochastické procesy
# sp1
df1.1<-data.frame(x=seq(0,6),y=c(1,1,2,0,2,3,2))
ggplot(df1.1,aes(x,y)) + labs(x="t",y=expression(X[t])) +
  geom_point(size=3,col="#F8766D")

# sp2
df1.2<-data.frame(x=seq(0,6),y=c(1.5,3,1.9,1.7,0.4,1,0.8))
ggplot(df1.2,aes(x,y)) + labs(x="t",y=expression(X[t])) +
  geom_point(size=3,col="#00BA38")

# sp3
df1.3<-data.frame(x=c(0,0.4,1.2,1.7,1.9,3,4.8),y=c(1,2,0,2,3,2,1),
  vx=c(0.4,1.2,1.7,1.9,3,4.8,6))
ggplot(df1.3,aes(x,y)) + labs(x="t",y=expression(X[t])) +
  geom_segment(aes(xend=vx, yend=y),size=1.5,col="#619CFF") +
  geom_point(aes(x=vx),size=2.5,shape=21,fill="white")

# sp4
t<-seq(0,6,0.01)
```

```
df1.4<-data.frame(x=t, y=3/0.26694*sin(t*exp(-t))*cos(t))
ggplot(df1.4,aes(x,y)) + labs(x="t",y=expression(X[t])) +
  geom_line(size=1.5,col="#D89000")

# trajektorie a NV
t<-seq(0,6,0.01)
y1<-2+sin(3*t^(1/2)+1)/2
y2<-2+sin(4*t^(1/3)+2)/3
y3<-2+sin(5*t^(1/4)+3)/4
df1.5<-data.frame(id=factor(rep(c(1,2,3),each=601)),x=rep(t,3),
  y=c(y1,y2,y3))
ggplot(df1.5,aes(x,y)) + labs(x="t",y=expression(X[t])) +
  geom_line(aes(colour=id),size=1.5) +
  geom_vline(xintercept=3.5) +
  geom_point(aes(x=3.5,y=2.161691),size=4,shape=21,fill="white") +
  geom_point(aes(x=3.5,y=2.325357),size=4,shape=21,fill="white") +
  geom_point(aes(x=3.5,y=1.899401),size=4,shape=21,fill="white") +
  theme(legend.title=element_blank()) +
  scale_colour_discrete(
    labels=c(expression(paste(X[t], '(', omega[1], ')')),
      expression(paste(X[t], '(', omega[2], ')')),
      expression(paste(X[t], '(', omega[3], ')'))))

# proces zrodu a zaniku
x<-c(0.2658633,0.9448095,1.1727146,2.0767420,2.3345838,4.0216092,
  4.5678973,5.1353035,5.6516231,6.1963257,7.8087089,8.4225299,
  8.6396434,8.7152521,9.3065322,9.7709044,10.3852231,11.0268804,
  11.4661720)
#x<-cumsum(rexp(19))
df1.6<-data.frame(x=c(0,x),vx=c(x,x[19]+1),
  y=c(0,1,2,3,2,3,4,3,2,3,4,3,2,3,2,1,0,1,2,1))
ggplot(df1.6,aes(x,y)) + labs(x="t",y=expression(X[t])) +
  geom_segment(aes(xend=vx, yend=y),size=1.5,col="#619CFF") +
  geom_point(aes(x=vx),size=2.5,shape=21,fill="white")

# proces cisteho zrodu
df1.7<-df1.6
df1.7$y<-seq(0,19)
ggplot(df1.7,aes(x,y)) + labs(x="t",y=expression(N[t])) +
  geom_segment(aes(xend=vx, yend=y),size=1.5,col="#619CFF") +
  geom_point(aes(x=vx),size=2.5,shape=21,fill="white")
```

```
#####  
##### KAPITOLA 2 #####  
#####  
  
# Nehomogenni PP  
# nhpp1  
n<-200  
max.lambda<-10  
x<-cumsum(rexp(n,max.lambda))  
lambda<-5*(1-cos(x))  
p<-lambda/max.lambda  
z<-rbinom(n,1,p)  
  
t0<-c(0,x[which(z==1)])  
t<-c(0,rep(x[which(z==1)],each=2))  
Nt<-c(rep(seq(0,length(t0)-2),each=2),length(t0)-1)  
y<-seq(0,max(t),0.01)  
fy<-5*(1-cos(y))  
id<-factor(c(rep(1,length(t)),rep(2,length(y))))  
df2.1<-data.frame(id,t=c(t,y),Nt=c(Nt,fy))  
  
ggplot(df2.1,aes(t,Nt)) + geom_line(aes(colour=id)) +  
  scale_colour_discrete(name="",  
                        labels=c(expression(N[t]),  
                                expression(paste(lambda (t))))) +  
  labs(x="t", y="")  
  
# nhpp2  
n<-100  
max.lambda<-20  
x<-cumsum(rexp(n,max.lambda))  
lambda<-20*(1-exp(-x))  
p<-lambda/max.lambda  
z<-rbinom(n,1,p)  
  
t0<-c(0,x[which(z==1)])  
t<-c(0,rep(x[which(z==1)],each=2))  
Nt<-c(rep(seq(0,length(t0)-2),each=2),length(t0)-1)  
y<-seq(0,max(t),0.01)  
fy<-20*(1-exp(-y))  
id<-factor(c(rep(1,length(t)),rep(2,length(y))))  
df2.2<-data.frame(id,t=c(t,y),Nt=c(Nt,fy))  
  
ggplot(df2.2,aes(t,Nt)) + geom_line(aes(colour=id)) +  
  scale_colour_discrete(name="")
```

```
                                labels=c(expression(N[t]),
                                expression(paste(lambda (t)))) +
labs(x="t",y="")

# nhpp3
n<-1000
max.lambda<-20
x<-cumsum(rexp(n,max.lambda))
lambda<-20*exp(-x)
p<-lambda/max.lambda
z<-rbinom(n,1,p)

t0<-c(0,x[which(z==1)])
t<-c(0,rep(x[which(z==1)],each=2))
Nt<-c(rep(seq(0,length(t0)-2),each=2),length(t0)-1)
y<-seq(0,max(t),0.01)
fy<-20*exp(-y)
id<-factor(c(rep(1,length(t)),rep(2,length(y))))
df2.3<-data.frame(id,t=c(t,y),Nt=c(Nt,fy))

ggplot(df2.3,aes(t,Nt)) + geom_line(aes(colour=id)) +
  scale_colour_discrete(name="",
                        labels=c(expression(N[t]),
                        expression(paste(lambda (t)))) +
  labs(x="t",y="")

# Homogenni PP
# hpp
n<-50
x1<-cumsum(rexp(n,1))
x2<-cumsum(rexp(n,2))
x3<-cumsum(rexp(n,5))

t1<-c(0,rep(x1,each=2))
t2<-c(0,rep(x2,each=2))
t3<-c(0,rep(x3,each=2))
Nt<-c(rep(seq(0,length(x1)-1),each=2),length(x1))
id<-factor(rep(c(1,2,3),each=length(Nt)))
df2.4<-data.frame(id,t=c(t1,t2,t3),Nt=rep(Nt,3))

ggplot(df2.4,aes(t,Nt)) + geom_line(aes(colour=id)) +
  scale_colour_discrete(name=expression(paste(lambda)),
                        labels=c(1,2,5)) +
  labs(x="t",y=expression(N[t]))
```

```
# Procesy s nakazou
# cont1
n<-50
a<-1
b<-0.5
x<-numeric()
y<-numeric()
for (i in 0:(n-1)) {
  x<-c(x, rexp(1, a+b*i))
  y<-c(y, a+b*i)
}
x<-cumsum(x)
t<-c(0, rep(x, each=2))
Nt<-c(rep(seq(0, length(x)-1), each=2), length(x))
fy<-c(rep(y, each=2), y[length(y)]+b)
id<-factor(rep(c(1,2), each=length(t)))
df2.5<-data.frame(id, t=rep(t, 2), Nt=c(Nt, fy))

ggplot(df2.5, aes(t, Nt)) + geom_line(aes(colour=id)) +
  scale_colour_discrete(name="",
                        labels=c(expression(N[t]),
                                   expression(paste(lambda[k])))) +
  labs(x="t", y="")

# cont2
n<-50
a<-5
b<-0
x<-numeric()
y<-numeric()
for (i in 0:(n-1)) {
  x<-c(x, rexp(1, a+b*i))
  y<-c(y, a+b*i)
}
x<-cumsum(x)
t<-c(0, rep(x, each=2))
Nt<-c(rep(seq(0, length(x)-1), each=2), length(x))
fy<-c(rep(y, each=2), y[length(y)]+b)
id<-factor(rep(c(1,2), each=length(t)))
df2.6<-data.frame(id, t=rep(t, 2), Nt=c(Nt, fy))

ggplot(df2.6, aes(t, Nt)) + geom_line(aes(colour=id)) +
  scale_colour_discrete(name="")
```



```
                                labels=c(expression(N[t]),
                                expression(paste(lambda[k]))) +
labs(x="t",y="")

# cont3
n<-50
a<-25
b<--0.5
x<-numeric()
y<-numeric()
for (i in 0:(n-1)) {
  x<-c(x,exp(1,a+b*i))
  y<-c(y,a+b*i)
}
x<-cumsum(x)
t<-c(0,rep(x,each=2))
Nt<-c(rep(seq(0,length(x)-1),each=2),length(x))
fy<-c(rep(y,each=2),y[length(y)]+b)
id<-factor(rep(c(1,2),each=length(t)))
df2.7<-data.frame(id,t=rep(t,2),Nt=c(Nt,fy))

ggplot(df2.7,aes(t,Nt)) + geom_line(aes(colour=id)) +
  scale_colour_discrete(name="",
                        labels=c(expression(N[t]),
                        expression(paste(lambda[k]))) +
  labs(x="t",y="")

#####
##### KAPITOLA 3 #####
#####

#CL model
u0<-100
c<-15
w<-round(1000*exp(20,1/3))
x<-c(0,cumsum(rgamma(19,20,0.5)))
xt<-numeric()
for (i in 1:20) {
  xt<-c(xt,rep(x[i],w[i]))
}
t<-seq(1,length(xt))/1000
ut<-u0+c*t-xt
```

```
df3.1<-data.frame(id=factor(rep(c(1,2),each=length(t))),
                  x=rep(t,2),y=c(mean=u0+(c-20*2/3)*t,ut))
ggplot(df3.1,aes(x,y)) + labs(x="t",y="") +
  geom_line(aes(colour=id)) + geom_hline(yintercept=u0,col="gray") +
  scale_colour_discrete("",labels=c("Střední \n hodnota",
                                    expression(U[t])))
```

```
id<-factor(rep(c(1,2,3),each=401))
df4.1<-data.frame(id)
x<-seq(0,4,0.01)
df4.1$x<-rep(x,3)
```

```
# Exponential
```

```
df4.1$f<-c(dexp(x,2),dexp(x,1),dexp(x,0.5))
ggplot(df4.1,aes(x,f)) + geom_line(aes(colour=id)) +
  scale_colour_discrete(name=expression(paste(lambda)),
                        labels=c(0.5,1,2)) +
  labs(x="x",y="f(x)")
```

```
# Weibull
```

```
k<-0.9
df4.1$f<-c(dweibull(x,k,0.5),dweibull(x,k,1),dweibull(x,k,2))
ggplot(df4.1,aes(x,f)) + geom_line(aes(colour=id)) +
  scale_colour_discrete(name=expression(paste(lambda)),
                        labels=c(0.5,1,2)) +
  labs(x="x",y="f(x)") + ylim(0,2.1)
```

```
k<-1
```

```
df4.1$f<-c(dweibull(x,k,0.5),dweibull(x,k,1),dweibull(x,k,2))
ggplot(df4.1,aes(x,f)) + geom_line(aes(colour=id)) +
  scale_colour_discrete(name=expression(paste(lambda)),
                        labels=c(0.5,1,2)) +
  labs(x="x",y="f(x)")
```

```
k<-3
```

```
df4.1$f<-c(dweibull(x,k,0.5),dweibull(x,k,1),dweibull(x,k,2))
ggplot(df4.1,aes(x,f)) + geom_line(aes(colour=id)) +
  scale_colour_discrete(name=expression(paste(lambda)),
                        labels=c(0.5,1,2)) +
  labs(x="x",y="f(x)")
```

```
# Gama
```

```
k<-1
df4.1$f<-c(dgamma(x,k,0.5),dgamma(x,k,1),dgamma(x,k,2))
```

```
ggplot(df4.1,aes(x,f)) + geom_line(aes(colour=id)) +
  scale_colour_discrete(name=expression(paste(lambda)),
                        labels=c(0.5,1,2)) +
  labs(x="x",y="f(x)")

k<-1.5
df4.1$f<-c(dgamma(x,k,0.5),dgamma(x,k,1),dgamma(x,k,2))
ggplot(df4.1,aes(x,f)) + geom_line(aes(colour=id)) +
  scale_colour_discrete(name=expression(paste(lambda)),
                        labels=c(0.5,1,2)) +
  labs(x="x",y="f(x)")

k<-2
df4.1$f<-c(dgamma(x,k,0.5),dgamma(x,k,1),dgamma(x,k,2))
ggplot(df4.1,aes(x,f)) + geom_line(aes(colour=id)) +
  scale_colour_discrete(name=expression(paste(lambda)),
                        labels=c(0.5,1,2)) +
  labs(x="x",y="f(x)")

# Lognormal
m<--1
df4.1$f<-c(dlnorm(x,m,0.5),dlnorm(x,m,1),dlnorm(x,m,2))
ggplot(df4.1,aes(x,f)) + geom_line(aes(colour=id)) +
  scale_colour_discrete(name=expression(paste(sigma)),
                        labels=c(0.5,1,2)) +
  labs(x="x",y="f(x)")

m<-0
df4.1$f<-c(dlnorm(x,m,0.5),dlnorm(x,m,1),dlnorm(x,m,2))
ggplot(df4.1,aes(x,f)) + geom_line(aes(colour=id)) +
  scale_colour_discrete(name=expression(paste(sigma)),
                        labels=c(0.5,1,2)) +
  labs(x="x",y="f(x)")

m<-1
df4.1$f<-c(dlnorm(x,m,0.5),dlnorm(x,m,1),dlnorm(x,m,2))
ggplot(df4.1,aes(x,f)) + geom_line(aes(colour=id)) +
  scale_colour_discrete(name=expression(paste(sigma)),
                        labels=c(0.5,1,2)) +
  labs(x="x",y="f(x)")

# Pareto (Lomax)
a<-0.25
df4.1$f<-c(dlomax(x,0.5,a),dlomax(x,1,a),dlomax(x,2,a))
ggplot(df4.1,aes(x,f)) + geom_line(aes(colour=id)) +
```

```
scale_colour_discrete(name=expression(paste(sigma)),
                      labels=c(0.5,1,2)) +
labs(x="x",y="f(x)")

a<-1
df4.1$f<-c(dlomax(x,0.5,a),dlomax(x,1,a),dlomax(x,2,a))
ggplot(df4.1,aes(x,f)) + geom_line(aes(colour=id)) +
  scale_colour_discrete(name=expression(paste(sigma)),
                      labels=c(0.5,1,2)) +
  labs(x="x",y="f(x)")

a<-4
df4.1$f<-c(dlomax(x,0.5,a),dlomax(x,1,a),dlomax(x,2,a))
ggplot(df4.1,aes(x,f)) + geom_line(aes(colour=id)) +
  scale_colour_discrete(name=expression(paste(sigma)),
                      labels=c(0.5,1,2)) +
  labs(x="x",y="f(x)") + ylim(0,2.1)

#####
##### KAPITOLA 4 #####
#####

# claims_year
d5.1<-c(42,44,77,81,87,88,63,114,146,171,204,222,145,132,134,123,
        151,145,179,169,188,246,269,293,311,450,377,361,1,0,1,26,
        54,22,59,87,195,247,284,168,9,0,288,323,341,399,376,493,
        515,518,437,464,391,478,348,241,51,55,55,56,107,81,111,
        130,161,110,103,85,115,101,86,128,133,170,207,227,289,177,
        245,350,284,221,105,110,265,191,219,235,218,238,214,236,
        229,271,394,295,303,219)
df5.1<-data.frame(id1=rep(c("03","05","18","32","34","54","55"),
                        each=14),
                 year=rep(seq(2003,2016),7),claims=d5.1)
ggplot(df5.1,aes(year,claims)) + geom_point(col="#619CFF") +
  geom_line(col="#619CFF") + facet_wrap(~id1) +
  labs(x="rok",y="počet škod")

# claims_month
d5.2<-c(102,131,135,122,156,120,160,140,171,163,156,133,305,345,354,
        361,379,402,378,433,397,402,380,307,82,86,86,85,92,95,91,
        117,111,127,87,94,637,477,504,516,514,534,444,483,496,517,
        454,482,82,100,106,137,106,116,142,120,129,98,129,84,56,156,
        64,93,438,622,1067,419,38,13,0,2,415,245,299,307,276,368,
```

```
513,375,266,290,294,255)
df5.2<-data.frame(idl=rep(c("03","05","18","32","34","54","55"),
                        each=12),
                 year=rep(seq(1,12),7),claims=d5.2)
ggplot(df5.2,aes(year,claims)) + geom_point(col="#619CFF") +
  geom_line(col="#619CFF") + facet_wrap(~idl) +
  labs(x="měsíc",y="počet škod") +
  scale_x_continuous(breaks=seq(2,12,2))

# season
d5.3<-c(1.21,0.72,1.17,0.82,1.32,0.85,1.26,0.88,1.26,0.73,4.31,0,
        1.58,0.75)
df5.3<-data.frame(idl=rep(c("03","05","18","32","34","54","55"),
                        each=2),
                 Mm=rep(c("max","min"),7),value=d5.3)
ggplot(df5.3,aes(idl,value)) + geom_line() + geom_point() +
  geom_hline(yintercept=1,col="gray") +
  scale_y_continuous(labels=percent) +
  labs(x="Číslo linky",y="")

# season_55_year
d5.4<-c(2,6,4,4,16,12,36,6,0,0,0,0,15,0,0,0,18,41,22,28,3,1,0,0,0,
        8,1,3,19,7,77,14,4,0,0,0,1,0,24,2,32,36,56,17,2,0,0,0,0,0,
        0,1,58,62,78,6,2,0,0,0,0,3,1,2,14,108,76,23,0,0,0,0,2,2,0,
        13,43,38,153,38,0,0,0,0,2,15,6,0,39,29,40,43,3,0,0,0,10,28,
        15,8,23,40,69,50,2,0,0,0,1,64,0,6,57,23,177,20,1,1,0,0,6,5,
        4,7,28,76,57,89,7,5,0,0,4,0,2,18,43,19,86,36,8,3,0,2,7,1,1,
        2,3,37,36,13,3,2,0,0,1,0,2,24,29,32,13,6,2,1,0,0)
df5.4<-data.frame(month=seq(1,168),claims=d5.4)
ggplot(df5.4,aes(month,claims)) + geom_line(col="#619CFF") +
  scale_x_continuous(breaks=seq(18.5,162.5,24),
                    labels=seq(2004,2016,2)) +
  labs(x="rok",y="počet škod")

# season_55_month
d5.5<-c(56,156,64,93,438,622,1067,419,38,13,0,2)
df5.5<-data.frame(month=seq(1,12),claims=d5.5)
ggplot(df5.5,aes(month,claims)) + geom_point(col="#619CFF") +
  geom_line(col="#619CFF") + labs(x="rok",y="počet škod") +
  scale_x_continuous(breaks=seq(2,12,2),
                    labels=c("únor","duben","červen","srpen",
                              "říjen","prosinec"))

# exposure
d5.6<-c(326,475,540,602,747,832,821,793,829,833,816,845,863,767,
```

```
4803,4870,5037,4964,4988,5185,5834,6805,7637,7705,8031,
8315,8156,8306,3550373,5405501,5507679,2430991,1033104,
758693,545359,327042,1109982,2025600,1977021,1795213,
1721307,1154210,3866,4798,5735,6575,7023,7630,8175,8590,
8666,8600,8517,8680,8825,9106,326,475,540,602,747,832,821,
793,829,833,816,845,863,767,527,625,612,641,657,705,671,
649,633,651,654,616,611,571)
df5.6<-data.frame(idl=rep(c("03","05","18","32","34","55"),each=14),
                  year=rep(seq(2003,2016),6),claims=d5.6)
ggplot(df5.6,aes(year,claims)) + geom_point(col="#619CFF") +
  geom_line(col="#619CFF") + facet_wrap(~idl,scales="free_y") +
  labs(x="rok",y="expozice")

# claims_year_final
d5.7<-c(NA,NA,NA,NA,68,74,43,82,98,120,138,145,89,87,110,81,114,110,
128,115,122,124,110,124,119,159,141,130,NA,NA,NA,2,14,7,24,
40,19,22,16,22,1,NA,NA,NA,NA,NA,NA,NA,291,269,238,256,216,
281,156,105,NA,NA,NA,NA,70,45,47,81,73,58,29,48,74,66,253,
147,178,188,165,157,164,190,177,204,287,234,240,179)
df5.7<-data.frame(idl=rep(c("03","05","18","32","34","55"),each=14),
                  year=rep(seq(2003,2016),6),claims=d5.7)
ggplot(df5.7,aes(year,claims)) + geom_point(col="#619CFF") +
  geom_line(col="#619CFF") + facet_wrap(~idl,scales="free_y") +
  labs(x="rok",y="počet škod") +
  geom_smooth(method="lm",col="black",size=0.5)

# trend_analysis
d5.8<-c(5.915,-4.399,16.229,2.978,0.348,5.608,0.394,-4.148,4.936,
-21.167,-45.622,3.289,0.006,-6.441,6.454,3.382,-4.745,11.51)
df5.8<-data.frame(idl=rep(c("03","05","18","32","34","55"),each=3),
                  item=rep(c("slope","int","int"),6),value=d5.8)
ggplot(df5.8,aes(idl,value,col=item)) +
  geom_hline(yintercept=0,col="gray") + geom_line(col="black") +
  geom_point(show.legend=F) +
  scale_colour_manual(values=c("black","#F8766D")) +
  labs(x="Číslo linky",y="99% interval spolehlivosti")

# interarrival_times (IAT)
d5.9<-c(158,198,113,106,74,60,54,48,32,14,15,16,11,11,6,6,261,428,
276,195,142,106,76,60,31,37,23,10,14,8,6,1,5,15,12,13,7,3,
9,5,6,3,5,4,5,5,2,5,596,548,266,155,97,54,32,30,17,3,5,1,2,
4,0,1,73,76,46,54,57,42,38,49,21,15,10,16,14,5,10,12,845,
810,389,239,181,113,59,42,25,13,14,10,8,6,3,1)
df5.9<-data.frame(idl=rep(c("03","05","18","32","34","55"),each=16),
                  days=rep(seq(0,15),6),IAT=d5.9)
```

```
ggplot(df5.9,aes(days,IAT)) + geom_point(col="#619CFF") +  
  geom_line(col="#619CFF") +  
  facet_wrap(~idl,scales="free_y",ncol=2) +  
  labs(x="počet dnů",y="počet intervalů")
```

Seznam použité literatury

- [1] ANDĚL, Jiří. *Základy matematické statistiky*. Praha: Matfyzpress, 2005. ISBN 80-867-3240-1.
- [2] BAROUCH, Eytan, a Gordon M. KAUFMAN, *On Sums of Lognormal Random Variables*. Massachusetts: M.I.T., 1976. WP 831-76.
- [3] BUDÍKOVÁ, Marie, Maria KRÁLOVÁ a Bohumil MAROŠ. *Průvodce základními statistickými metodami*. Praha: Grada, 2010. ISBN 978-80-247-3243-5.
- [4] BUHLMANN, Hans. *Mathematical methods in risk theory*. New York: Springer, 2005. ISBN 978-3-540-05117-6.
- [5] CIPRA, Tomáš. *Pojistná matematika: teorie a praxe*. 2. aktualiz. vyd. Praha: Ekopress, 2006. ISBN 80-869-2911-6.
- [6] FORBELSKÁ, Marie, a Jan KOLÁČEK, *Pravděpodobnost a statistika I*. Studijní text k předmětu M3121. Brno: Masarykova univerzita, Přírodovědecká fakulta.
- [7] GRIMMETT, Geoffrey, a David STIRZAKER. *Probability and random processes*. 3rd ed. New York: Oxford University Press, 2001. ISBN 978-0-198-57222-0.
- [8] HENDL, Jan. *Přehled statistických metod: analýza a metaanalýza dat*. 4. rozš. vyd. Praha: Portál, 2012. ISBN 978-80-262-0200-4.
- [9] HULL, John. *Options, Futures and Other Derivatives*. 5th ed. New Jersey: Prentice Hall, 2003. ISBN 0-13-046592-5.
- [10] JOHNSON, Norman L., Samuel KOTZ a Narayanaswamy BALAKRISHNAN. *Continuous univariate distributions*. Volume 1, 2nd ed. New York: Wiley, 1995. ISBN 04-715-8495-9.
- [11] KARLIN, Samuel, a Howard M. TAYLOR. *A first course in stochastic processes*. 2nd ed. New York: Academic Press, 1975. ISBN 978-0-12-398552-1.
- [12] KARLIN, Samuel, a Howard M. TAYLOR. *A second course in stochastic processes*. New York: Academic Press, 1981. ISBN 978-0-12-398650-4.
- [13] KLUGMAN, Stuart A., Harry H. PANJER a Gordon E. WILLMOT. *Loss models: from data to decisions*. 3rd ed. Hoboken: John Wiley, 2008. ISBN 978-047-0187-814.

- [14] PARZEN, Emanuel. *Stochastic processes*. Philadelphia: Society for Industrial and Applied Mathematics, 1999. ISBN 978-0-898714-41-8.
- [15] PRÁŠKOVÁ, Zuzana, a Petr LACHOUT. *Základy náhodných procesů I*. 2. vyd. Praha: Matfyzpress, 2012. ISBN 978-80-7378-210-8.
- [16] ROSS, Sheldon M. *Introduction to probability models*. 9th ed. Boston: Academic Press, 2007. ISBN 978-0-12-598062-3.
- [17] SCHMIDT, Klaus D. *Lectures on risk theory*. Stuttgart: Teubner, 1996. ISBN 978-351-9027-355.
- [18] THORIN, Olof. *The Journal of the International Actuarial Association: Some Comments on the Sparre Andersen Model in the Risk Theory*. ASTIN Bulletin [online]. 1974, 17(September), 104-125 [cit. 2017-03-05]. Dostupné z: <http://www.actuaries.org/LIBRARY/ASTIN/vol8no1/104.pdf>

Seznam obrázků

1.1	Příklady trajektorií stochastických procesů	5
1.2	Trajektorie a náhodná veličina	5
1.3	Trajektorie procesu zrodu a zániku	11
1.4	Trajektorie procesu čistého zrodu	14
2.1	Rozdělení intervalu $(s, t + h)$ v bodě t	16
2.2	Trajektorie nehomogenního Poissonova procesu	18
2.3	Trajektorie homogenního Poissonova procesu	21
2.4	Trajektorie procesů s nákazou	24
3.1	Vývoj kapitálu podle Cramérova-Lundbergova modelu	28
3.2	Hustota pravděpodobnosti exponenciálního rozdělení	35
3.3	Hustota pravděpodobnosti Weibullova rozdělení	37
3.4	Hustota pravděpodobnosti gama rozdělení	39
3.5	Hustota pravděpodobnosti lognormálního rozdělení	40
3.6	Hustota pravděpodobnosti Paretova rozdělení	42
4.1	Počet škod podle linek a let vzniku	49
4.2	Počet škod podle linek a měsíců vzniku	50
4.3	Max. a min. odchylka dat od průměru	50
4.4	Sezónnost pojištění plodin (54)	51
4.5	Expozice	53
4.6	Finální počet škod po letech po očištění	54
4.7	Analýza trendu finálního počtu škod	54
4.8	Absolutní četnosti intervalů mezi škodami	55
4.9	Pearsonova testová statistika vybraných rozdělení (linka 55)	58
4.10	Srovnání teoretického rozdělení s daty (03)	59
4.11	Srovnání teoretického rozdělení s daty (05)	60
4.12	Srovnání teoretického rozdělení s daty (18)	61
4.13	Srovnání teoretického rozdělení s daty (32)	62
4.14	Srovnání teoretického rozdělení s daty (34)	63
4.15	Srovnání teoretického rozdělení s daty (55)	64

Seznam tabulek

4.1	Struktura dat	47
4.2	Odstranění nulových škod	52
4.3	Nastavení základní expozice	53
4.4	Finální počet škod	53
4.5	Odhady parametrů (03)	59
4.6	Odhady parametrů (05)	60
4.7	Odhady parametrů (18)	61
4.8	Odhady parametrů (32)	62
4.9	Odhady parametrů (34)	63
4.10	Odhady parametrů (55)	64
A.1	Počet škod podle linek a let vzniku	66
A.2	Počet škod podle linek a měsíců vzniku	67
A.3	Max. a min. odchylka dat od průměru	67
A.4	Expozice	67
A.5	Finální počet škod po letech po očištění	68
A.6	Analýza trendu finálního počtu škod	68
A.7	Počet dní mezi škodami	69

Rejstřík

E

elementární věta obnovy, 31
expozice, 52

F

funkce

autokorelační, 6
autokovarianční, 6
digama, 39
distribuční, 6, 33
gama, 37
dolní neúplná, 38
charakteristická, 35
obnovy, 30
přežití, 33
rizika, 33

H

hustota pravděpodobnosti, 33

I

intenzita

celková, 9
procesu obnovy, 31
přechodu, 9
výstupu, 9

K

kritérium

Yarnoldovo, 44

M

matice

intenzit přechodu, 10

metoda

maximální věrohodnosti, 56
modifikovaná, 56
minimálního χ^2 , 44, 58

moment

centrovaný, 34
obecný, 34

N

náhodná veličina

absolutně spojitá, 33

nákaza

lineární, 22
negativní, 22
pozitivní, 22

O

operační čas, 27

P

pravděpodobnost

přechodová, 8

proces

čistého zrodu, 13
čítací, 13
homogenní, 8
lineárního růstu, 22
markovský, 8
obnovy, 29
čistý, 29
se zpožděním, 29

Poissonův

homogenní, 20
nehomogenní, 18

ředění, 22

s nákazou, 22

stochastický, 4

Yuleův s driftem, 22

zrodu a zániku, 10

přírůstek

nezávislý, 7
stacionární, 7

R

rovnice

- Chapman-Kolmogorova, 15
- obnovy, 32
- rozdělení
 - Erlangovo, 38
 - exponenciální, 35
 - gama, 38
 - lognormální, 39
 - Lomaxovo, 41
 - multinomické, 43
 - Paretovo, 41
 - s těžkým koncem, 41
 - stacionární, 10
 - Weibullovo, 36
- rozptyl, 6, 34
- run-off, 51
- Ř**
- řetězec
 - Markovův, 7
- S**
- směrodatná odchylka, 34
- stacionarita
 - slabá, 7
 - striktní, 7
- statistika
 - Pearsonova, 44
- stavový prostor, 4
- střední hodnota, 6, 34
- T**
- test
 - dobré shody, 43
 - Kolmogorovův-Smirnovův, 44
- trajektorie, 5
- V**
- vlastnost
 - markovská, 8

