MATEMATICKO-FYZIKÁLNÍ FAKULTA
Univerzita Karlova

BAKALÁŘSKÁ PRÁCE

Ondřej Přítel

Useknuté čítací procesy

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: doc. RNDr. Michal Pešta, Ph.D.
Studijní program: Finanční matematika
Studijní obor: Finanční matematika

Praha 2023

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

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V .... dne ....

Podpis autora

Chtěl bych zde poděkovat doc. RNDr. Michal Pešta, Ph.D. za vedení mé bakalářské práce. Jeho rady a poskytnutá péče mi byly velkým přínosem. Dále bych si dovolil poděkovat mé rodině, která mě vždy dokázala povzbudit a motivovat.

Název práce: Useknuté čítací procesy

Autor: Ondřej Přítel

Katedra: Katedra pravděpodobnosti a matematické statistiky

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

Abstrakt: Tato práce se zabývá predikcí počtu pojistných událostí, v situaci kdy data o vznicích událostí jsou useknutá. Podstata useknutí dat spočívá v tom, že v současnosti pozorujeme pouze události, které již byly nahlášeny pojišťovnám. Vzniky a hlášení událostí budeme reprezentovat dvourozměrným nehomogenním Poissonovým procesem. Intenzita procesu vzniků je odvozena pomocí Kingmanova Displacement theorem, který ji počítá konvolucí intenzity procesu hlášení a hustoty prodlení mezi vznikem a nahlášením. Odhady parametrických funkcí intenzity hlášení a hustoty jsou odhadnuty pomocí metody maximální věrohodnosti. V práci je navíc uvedena obecná teorie týkající se čítacích procesů s primárním zaměřením na homogenní a nehomogenní Poissonův proces.

Klíčová slova: useknutá data | čítací proces | Poissonův proces | opožděné hlášení

Title: Truncated counting processes

Author: Ondřej Přítel

Department: Department of Probability and Mathematical Statistics

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

Abstract: The aim of this thesis is the prediction of insurance events under the condition that the data related to the occurrence of the events is truncated. The nature of the truncation lies in the fact that in the present we observe only those events that were already reported to the insurance company. Occurrences and reporting are modeled by a two-dimensional non-homogeneous Poisson process. The intensity of occurrences is derived from Kingman's Displacement theorem and is computed as a convolution of the intensity of reporting and the density of the delay in between occurrences and reporting. The estimations of the parametric function of the intensity of reporting and the distribution are preformed using the maximum likelihood method. In addition, theoretical background concerning counting processes primarily directed to the Poison processes is discussed in this thesis.

Keywords: truncated data | counting process | Poisson process | reporting delay

Obsah

Úvod

Pojišťovna je finanční instituce, která na základě pojistné smlouvy finančně kryje budoucí pojištěná rizika pojištěného subjektu. Budoucí pojištěná rizika pojištěného subjektu jsou zároveň i rizikem pro pojišťovny. Správná stochastická predikce pojistných událostí je pro pojišťovny klíčová, aby mohla dostát svým budoucím závazkům vůči pojištěnci.

V této práci předkládáme matematickou teorii spojenou s predikcí pojistných událostí, která bere v potaz všechna data pro pojišťovny dostupná. Současné modely používané v pojišťovnách se zabývají pouze agregovanými daty za delší časové období. Náš přístup vnímá posloupnost vzniků pojistných událostí jako stochastický proces a umožňuje nám vystavět teorii pro predikci předpokládaného počtu budoucích pojistných událostí, která vychází z jednotlivý časů vzniku pojistných událostí.

Problematika vytvoření takovéto predikce tkví v charakteru dat. Data o vzniklých událostech, které máme v současnosti k dispozici, jsou useknutá. Dnes víme pouze o událostech, které již byli nahlášeny pojišťovnám. Nezaznamenáváme události, které již vznikly, ale k dnešnímu datu stále nebyli pojišťovnám nahlášeny. Tato charakteristika dat se nazývá useknutí a motivuje nás k zavedení Kigmanova Displacement theorem (Kingman, 1993), který v této práci nazýváme větou o posunutí. Předpokládejme, že máme k dispozici dva procesy, které jsou vzájemně provázány. Věta o posunutí říká, že intenzitu jednoho procesu můžeme odvodit z intenzity druhého procesu a z jejich vzájemného vztahu.

V praktické části této práce tuto větu použijeme k určení intenzity procesu vzniku pojistných událostí. Při její aplikaci budeme uvažovat intenzitu neuseknutého procesu hlášení a hustotu času mezi vznikem události a jejím nahlášením.

1. Čítací a Poissonův proces

V této kapitole se zabýváme čítacími procesy v jejich jednorozměrné i dvojrozměrné formě. Mezi nejpoužívanější čítací procesy patří Poissonův proces. Zavedeme homogenního a nehomogenní Poissonův proces a popíšeme vlastnosti obou variant. Dále představíme pojem useknutého procesu a naznačíme jeho využití v praktické části.

1.1 Čítací proces

Primárním zdrojem pro tuto i následující podkapitoly je monografie od Rosse (Ross, 1996), dále jsme čerpali z prací Tijms (2003) a Andersen a kol. (1992).

Nejprve připomeňme pojem náhodného procesu.

Definice 1 (Náhodný proces).
Náhodným procesem $\{X(t),\ t \in T\}$ *rozumíme soubor náhodných veličin.*

Proměnou t budeme pro účely této bakalářské práce chápat jako čas. Pokud je množina T spočetná, pak se jedná o proces s diskrétním časem neboli časovou řadu. V opačném případě nazýváme soubor náhodných veličin $\{X(t),\ t \in T\}$ pouze procesem nebo procesem se spojitým časem. V této práci se budeme zabývat pouze procesy spojitého typu.

Čítací proces je náhodný proces, kde každému kladnému t přiřadíme náhodnou veličinu s určitými vlastnostmi.

Definice 2 (Čítací proces).
Náhodný proces $\{N(t),\ t \geq 0\}$ *nazýváme čítací proces, pokud splňuje následující předpoklady:*

(i) $N(t) \geq 0$*,* $N(0) = 0$*,*

(ii) $N(t) \in \mathbb{N}$*,*

(iii) pokud $t_1 \leq t_2$*, pak* $N(t_1) \leq N(t_2)$*.*

Náhodná veličina $N(t)$ udává celkový počet událostí, které nastaly do času t. Náhodná veličina $N(t_2) - N(t_1)$ vyjadřuje počet událostí, které nastaly v časovém intervalu $(t_1, t_2]$.

Nyní zavedeme několik pojmů: nezávislé přírůstky, stacionární přírůstky a čas výskytu.

- Říkáme, že proces má nezávislé přírůstky, pokud je počet událostí, které se odehrály v disjunktních časových intervalech, nezávislý. Pro $t_0 < t_1 < \cdots < t_n$ jsou náhodné veličiny $N(t_1) - N(t_0), N(t_2) - N(t_1), \ldots, N(t_{n-1}) - N(t_n)$ nezávislé.

- O procesu řekneme, že je stacionární, pokud pravděpodobnostní rozdělení náhodné veličiny $N(t_2) - N(t_1),\ t_1 < t_2$ závisí pouze na délce intervalu $(t_1, t_2]$.

Poznámka. Za platnosti výše uvedených vlastností, proces stále nebude mít nezávislé a stejně rozdělené přírůstky. Abychom získali stejně rozdělené přírůstky, musely by intervaly mezi událostmi být stejně velké.

- Čas výskytu n-té události značíme Z_n a $Z_n = \sum_{i=1}^{n} U_i$, kde veličina U_i představuje čas mezi $(i-1)$-ní a i-tou událostí. $U_1 = Z_1$ je čas do první události.

Praktická část této práce bude zkoumat aplikace dvourozměrného čítacího procesu.

Definice 3 (Dvojrozměrný čítací proces)**.**
Dvojrozměrný čítací proces $[\{M(s), N(t)\},\ s \geq 0, t \geq 0]$ *chápeme jako náhodný proces o dvou složkách, které reprezentují počet výskytů dvou diskrétních událostí.*

Proměnné s a t mohou a nemusí náležet stejnému prostoru.

Právě uvedená podkapitola popisuje čítací procesy, ve kterých proměnnou t vnímáme jako čas. Odkloňme se o tohoto předpokladu a předložme jiný pohled na čítací procesy. Nyní nebudeme vnímat t jako čas, ale jako polohu. Proměnná t bude náležet přímce (cyklostezce). Náhodný proces $\{N\}$ bude určovat posloupnost bodů na této přímce (cyklistů na cyklostezce). Body $Z_1 < \cdots < Z_n$ reprezentují polohu cyklistů. Na tuto posloupnost nahlížíme v neměnném čase.

Dvourozměrný čítací proces si pak můžeme představit jako posloupnost chodců $\{M(t)\}$ a cyklistů $\{N(t)\}$ na jedné cyklostezce v daném čase. Zde můžeme předpokládat, že proměnné obou procesů jsou totožné. Jiný přístup by mohl znázorňovat posloupnost cyklistů v dvou různých časech. $\{M(s)\}$ by byla posloupnost cyklistů v počátečním čase a $\{N(t)\}$ by byla posloupnost cyklistů v nějakém čase budoucím. Je zřejmé, že poloha cyklisty v budoucím čase závisí na jeho startovní pozici, proto proměnné s a t na sobě v určitém smyslu závisí.

1.2 Useknutí

Useknutí dat obecně znamená omezení našeho pozorování na data, která splňují určitá kritéria. V kontextu této práce bude určující podmínkou čas.

Definice 4 (Useknutí)**.**
Nechť $\{N(t), t \geq 0\}$ *je náhodný proces a* c *je nezáporná konstanta. Řekneme, že* $\{N_L(t),\ t \in [c, \infty)\}$ *je zleva useknutý, když proces pozorujeme pouze pro* $t \geq c$*.*

Obdobně pro useknutí zprava. Proces $\{N_R(t),\ t \in [0, c]\}$ *je zprava useknutý, když proces pozorujeme pouze pro* $t \leq c$*.*

Dvojrozměrný čítací proces $[\{M(s), N(t)\},\ s,t \geq 0]$, který má zprava useknutou první složkou, bude předmětem této práce. Čas useknutí bude záviset na druhé složce.

1.3 Homogenní Poissonův proces

V této podkapitole prostudujeme Poissonův proces se stacionárními přírůstky. Proces, který splňuje předpoklad stacionarity se nazývá homogenní. Na nehomogenní proces, tedy proces, který předpoklad stacionarity nesplňuje, se zaměříme v příští podkapitole.

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

(i) $N(0) = 0$,

(ii) proces má nezávislé a stacionární přírůstky,

(iii) $\mathsf{P}[N(t+s) - N(t) = 1] = \lambda s + o(s)$,

(iv) $\mathsf{P}[N(t+s) - N(t) \geq 2] = o(s)$.

Z definice vyplývají dva triviální, ale důležité poznatky. Očekávaná hodnota Poissonova procesu v čase t je λt a pravděpodobnost nastání dvou událostí ve stejný čas se limitně blíží nule.

Proces, který je zadefinován právě uvedeným způsobem, se nazývá Poissonův, protože náhodné veličiny $N(t)$ mají Poissonovo rozdělení. Toto tvrzení nyní dokážeme. Principem důkazu je Poissonova aproximace binomického rozdělení.

Rozdělme interval $[0,t]$ na k shodných podintervalů. Jak již bylo naznačeno v poznámce za definicí, pro $k \to \infty$ se pravděpodobnost výskytu dvou událostí v jednom podintervalu blíží k 0. Formálněji:

$$\begin{aligned}
&\mathsf{P}[\text{2 nebo více událostí se uskuteční v nějakém podintervalu}] \\
&\leq \sum_{i=1}^{k} \mathsf{P}[\text{2 nebo více událostí se uskuteční v } i\text{-tém podintervalu}] \\
&= k \cdot o\left(\frac{t}{k}\right) \\
&= t \cdot \frac{o(t/k)}{t/k} \\
&\to 0, \quad k \to \infty.
\end{aligned}$$

$N(t)$ se bude s pravděpodobností blížící se k jedné rovnat počtu podintervalů, ve kterých nastala událost. Zároveň, díky stacionaritě a nezávislosti přírůstků, bude náhodná veličina $N(t)$ mít binomické rozdělení s parametry k a $p = \lambda\frac{t}{k} + o\left(\frac{t}{k}\right)$. Z Poissonovy aproximace vyplývá, že pro $k \to \infty$ bude $N(t)$ mít Poissonovo rozdělení se střední hodnotou rovnou

$$\lim_{k\to\infty} kp = \lim_{k\to\infty} k\left[\lambda\frac{t}{k} + o\left(\frac{t}{k}\right)\right] = \lambda t + \lim_{k\to\infty} k\left[t\frac{o(t/k)}{t/k}\right] = \lambda t.$$

Navíc díky stacionaritě přírůstků můžeme toto tvrzení rozšířit i pro náhodnou veličinu $N(t+s) - N(t)$.

Tvrzení 1. *Náhodná veličina* $\{N(t+s) - N(t),\ t,s \geq 0\}$*, reprezentující počet událostí nastalých v intervalu délky t, má Poissonovo rozdělení s parametrem* λs*. Neboli* $\forall s,t \geq 0$ *platí*

$$\mathsf{P}[N(t+s) - N(t) = n] = e^{-\lambda s}\frac{(\lambda s)^n}{n!}, \quad n = 0,1,\dots\ .$$

U čítacích procesů jsme zavedli dvě náhodné veličiny:

- U_n - interval mezi $(n-1)$ a n-tou událostí,
- Z_n - čas výskytu i-té události.

V případě homogenních Poissonových procesů můžeme obě veličiny charakterizovat distribuční funkcí.

Tvrzení 2. *Časy mezi dvěma událostmi* $\{U_n,\ n = 1{,}2,\dots\}$ *jsou nezávislé náhodné veličiny s exponenciálním rozdělením.*

Časy výskytů $\{Z_n,\ n = 1{,}2,\dots\}$ *jsou nezávislé náhodné veličiny s Erlangovým rozdělením.*

Důkaz. Tvrzení dokážeme rozdílně pro $n = 1$ a pro $n > 1$. Jev $\{U_1 > t\}$ nastane pouze, když do času t nenastanou žádné události.

$$\mathsf{P}[U_1 > t] = \mathsf{P}[N(t) = 0] = \frac{(\lambda t)^0}{0!} e^{-\lambda t} = e^{-\lambda t}$$

T_1 má zřejmě exponenciální rozdělení s parametrem λ. Pro $n > 1$ jev $\{U_n > s\}$ za podmínky $\{\sum_{i=1}^{n-1} U_i = t\}$ nastane pouze tehdy pokud, během $(t, t+s)$ nenastanou žádné události.

$$\mathsf{P}\left[U_n > s \,\middle|\, \sum_{i=1}^{n-1} U_i = t\right] = \mathsf{P}[U_n > s] = e^{-\lambda s}$$

První rovnost plyne z nezávislosti přírůstků, druhá z jejich stacionarity.

Erlangovo rozdělení je rozdělení sumy n nezávislých exponenciálních rozdělení.

□

Náhodná veličina U_n je nezávislá a stejně rozdělená. Proces, který má nezávislé stejně rozdělené časy mezi událostmi, nazveme procesem bez paměti.

Uveďme další ekvivalentní způsob jak definovat Poissonův proces. Mějme posloupnost $\{Z_n, n \geq 1\}$ definovanou jako v tvrzení 2. Pak počet událostí v Poissonově procesu $\{N(t), t \geq 0\}$ můžeme definovat jako :

$$N(t) = \sum_{i=0}^{\infty} \mathbb{1}\{Z_i <\leq t\} \quad \text{nebo také } N(t) = \max_{i\geq 1}\{Z_i \leq t\}$$

Tato definice dobře vyjadřuje vztah mezi počtem událostí a časy výskytů.

1.4 Nehomogenní Poissonův proces

Nehomogenní Poissonův proces je zobecněním homogenní varianty, ve kterém intenzita procesu $\lambda(t)$ je funkcí času.

Definice 6 (Nehomogenní Poissonův proces).
Čítací proces $\{N(t),\ t \geq 0\}$ *nazýváme nehomogenní Poissonův proces s intenzitou* $\lambda(t),\ t \geq 0$, *kde* $\lambda(t)$ *je nezáporná, měřitelná funkce, pokud:*

(i) $N(0) = 0$,

(ii) *proces má nezávislé přírůstky,*

(iii) $\mathsf{P}[N(t+s) - N(t) = 1] = \lambda(t)s + o(s)$,

(iv) $\mathsf{P}[N(t+s) - N(t) \geq 2] = o(s)$.

Předpis očekávané hodnoty nehomogenního Poissonova procesu přejímáme od Rosse (Ross, 2010) a to ve tvaru

$$\mathsf{E}\left[N(t)\right] = \int_0^t \lambda(y)\mathrm{d}y = \mu(t).$$

Funkci $\mu(t)$ nazveme funkcí střední hodnoty nehomogenního Poissonova procesu. Uvědomme si, že při konstantní intenzitě $\lambda(t) = \lambda$ platí $\mu(t) = \lambda t$.

Věta 3. *Nechť $\{N(t), t \geq 0\}$ je nehomogenní Poissonův proces s intenzitou $\lambda(t)$ a funkcí střední hodnoty $\mu(t)$. Pak počet událostí z intervalu $(t, t+s]$, neboli náhodná veličina $N(t+s) - N(t)$, $t,s > 0$ má Poissonovo rozdělení s parametrem $\mu(t+s) - \mu(t)$.*

Zapsáno formálně:

$$\mathsf{P}[N(t+s) - N(t) = n] = e^{-[\mu(t+s)-\mu(t)]}\frac{[\mu(t+s) - \mu(t)]^n}{n!}, \quad t,s > 0 \tag{1.1}$$

Důkaz. Zavedeme značení

$$p_n(s) = \mathsf{P}[N(t+s) - N(t) = n], \quad n = 0,1,\ldots.$$

Veličina $p_n(s + \Delta s)$ značí pravděpodobnost, že n událostí nastane mezi časy t a $t + s + \Delta s$.

Z definice 6 bodu (iv) víme, že pro $\Delta s \to 0$, stav procesu v čase $t + s + \Delta s$ může být n, pouze pokud stav v čase $t+s$ je n nebo $n-1$. Zapsáno v zavedeném značení

$$p_n(s + \Delta s) = p_{n-1}(s)[\lambda(t+s)\Delta s + o(\Delta s)] + p_n(s)[1 - \lambda(t+s)\Delta s + o(\Delta s)],$$

$$\Delta s \to 0.$$

Tuto rovnost převedeme do tvaru diferenciální rovnice následujícími úpravami:

$$p_n(s + \Delta s) - p_n(s) = p_{n-1}(s)\lambda(t+s)\Delta s - p_n(s)\lambda(t+s)\Delta s$$

$$+ o(\Delta s)[p_n(s) + p_{n-1}(s)];$$

$$\frac{p_n(s + \Delta s) - p_n(s)}{\Delta s} = p_{n-1}(s)\lambda(t+s) - p_n(s)\lambda(t+s)$$

$$+ \frac{o(\Delta s)[p_n(s) + p_{n-1}(s)]}{\Delta s};$$

$$p_n'(s) + \lambda(t+s)p_n(s) = \lambda(t+s)p_{n-1}(s). \tag{1.2}$$

Třetí ekvivalence platí díky $\Delta s \to 0$.

Nyní řešíme obyčejnou lineární diferenciální rovnici 1. řádu. Řešením této rovnice bude pravděpodobnostní rozdělení veličiny $N(t+s) - N(t)$. Diferenciální rovnici nebudeme řešit běžným způsobem, kterým bychom hledali řešení homogenní rovnice a partikulární řešení. Náš důkaz založíme na indukci. Vyřešíme homogenní rovnici a zavedeme indukční předpoklad, že rovnice (1.1) je platná pro $n-1$.

Pro $n = 0$ získáme homogenní rovnici. Definiční obor funkce $p_0(s)$ je $[t,\infty)$ a intenzitu uvažujeme na intervalu $[0,t+s]$. Pak

$$p_0'(s) + \lambda(t+s)p_0(s) = 0;$$

$$\frac{p_0'(s)}{p_0(s)} = -\lambda(t+s);$$

$$\int \frac{p_0'(s)}{p_0(s)} = -\int_0^{t+s} \lambda(x)\mathrm{d}(x);$$

$$\ln p_0(s) = -\mu(t+s) + C;$$

$$p_0(s) = Ke^{-\mu(t+s)}.$$

Pro $s = 0$ získáme krajní bod $p_0(0) = 1$. Využijeme tohoto poznatku a dopočítáme konstantu K pro $s = 0$:

$$p_0(0) = 1 = Ke^{-\mu(t+0)};$$

$$K = e^{\mu(t)}.$$

Řešením homogenní rovnice je

$$p_0(s) = e^{\mu(t)}e^{-\mu(t+s)} = e^{-[\mu(t+s)+\mu(t)]}.$$

Tímto jsme dokázali rovnost (1.1) pro $n = 0$. Uveďme indukční předpoklad

$$p_{n-1}(s) = e^{-[\mu(t+s)-\mu(t)]}\frac{[\mu(t+s)-\mu(t)]^{n-1}}{(n-1)!}. \tag{1.3}$$

Nejprve upravíme diferenciální rovnici (1.2). Rovnici přenásobíme $e^{\mu(t+s)}$:

$$e^{\mu(t+s)}[p_n'(s) + \lambda(t+s)p_n(s)] = e^{\mu(t+s)}\lambda(t+s)p_{n-1}(s);$$

$$\frac{\mathrm{d}}{\mathrm{d}s}\left[e^{\mu(t+s)}p_n(s)\right] = e^{\mu(t+s)}\lambda(t+s)p_{n-1}(s). \tag{1.4}$$

Ekvivalence využívá derivaci součinu (1.5) a derivaci integrálu podle horní meze (1.6).

$$\frac{\mathrm{d}}{\mathrm{d}s}\left[e^{\mu(t+s)}p_n(s)\right] = e^{\mu(t+s)}\frac{\mathrm{d}}{\mathrm{d}s}\left[\int_0^{t+s}\lambda(x)\mathrm{d}x\right]p_n(s) + e^{\mu(t+s)}p_n'(s) \tag{1.5}$$

$$\frac{\mathrm{d}}{\mathrm{d}s}\left[\int_0^{t+s}\lambda(x)\mathrm{d}x\right] = \frac{\mathrm{d}}{\mathrm{d}s}[\mu(t+s)] = \lambda(t+s) \tag{1.6}$$

Nyní do rovnice (1.4) dosadíme indukční předpoklad (1.3) a rovnici zintegrujeme:

$$\frac{\mathrm{d}}{\mathrm{d}s}\left[e^{\mu(t+s)}p_n(s)\right] = \lambda(t+s)e^{\mu(t+s)}e^{-[\mu(t+s)-\mu(t)]}\frac{[\mu(t+s)-\mu(t)]^{n-1}}{(n-1)!};$$

$$\int \frac{\mathrm{d}}{\mathrm{d}s}[e^{\mu(t+s)}p_n(s)]\mathrm{d}s = \int \lambda(t+s)e^{\mu(t)}\frac{[\mu(t+s)-\mu(t)]^{n-1}}{(n-1)!}\mathrm{d}s;$$

$$e^{\mu(t+s)}p_n(s) = e^{\mu(t)}\frac{[\mu(t+s)-\mu(t)]^{n}}{n!} + C. \tag{1.7}$$

Výsledek levé strany rovnice je zřejmý. Pravou stranu rovnice osvětlíme následující derivací:

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}s}e^{\mu(t)}\frac{[\mu(t+s)-\mu(t)]^{n}}{n!} + C \\
&= \frac{e^{\mu(t)}}{n!}\frac{\mathrm{d}}{\mathrm{d}s}\left(\mu(t+s)-\mu(t)\right)^n \\
&= \frac{e^{\mu(t)}}{n!}n\left(\mu(t+s)-\mu(t)\right)^{n-1}\lambda(t+s) \\
&= \lambda(t+s)e^{\mu(t)}\frac{[\mu(t+s)-\mu(t)]^{n-1}}{(n-1)!}.
\end{aligned}$$

Pro $n \geq 1$ a $s = 0$ víme, že $p_n(0) = 0$. Tento poznatek aplikujeme v rovnici (1.7) a dopočítáme konstantu C:

$$e^{\mu(t)}p_n(0) = e^{\mu(t)}\frac{[\mu(t)-\mu(t)]^{n}}{(n)!} + C;$$

$$0 = 0 + C.$$

Dosazením do (1.7) získáme:

$$p_n(s) = e^{-[\mu(t+s)-\mu(t)]}\frac{[\mu(t+s)-\mu(t)]^{n}}{(n)!}.$$

$\square$

2. Odhad intenzity u Poissonova procesu

V této práci budeme předpokládat, že intenzita je parametrickou funkcí tvaru

$$\lambda(t) = \lambda(t, \boldsymbol{\theta}),$$

kde $\lambda(t,\boldsymbol{\theta}) > 0$, pro $t > 0$ a $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$. Speciálním příkladem je konstantní intenzita $\lambda(t) = \lambda$, ta mám definuje homogenní Poissonův proces.

Abychom mohli modelovat Poissonův proces, tak musíme najít odhad intenzity. Metoda maximální věrohodnosti nám dobře poslouží k získání vhodných odhadů. Výpočet maximálně věrohodného odhadu blízce sleduje práci Maciak M. (2021) s tím rozdílem, že zmíněná práce se zabývá odhady pro Hawkesův proces. Dále se teorie opírá o články Zhao a Xie (1996) a Lawless (1987).

Uvažujme Poissonův proces $\{N(t),t > 0\}$ s intenzitou $\lambda(t,\boldsymbol{\theta})$, který pozorujeme na časovém intervalu $[0,A]$. A nechť posloupnost časů $Z_1 < \cdots < Z_{N(A)}$ reprezentuje časy nastání událostí. Pak dle Thompsona (Thompson, 2012) věrohodnostní funkce je tvaru

$$L(t,\boldsymbol{\theta}) = \exp\{-\mu(A,\boldsymbol{\theta})\} \prod_{i=1}^{N(A)} \lambda(Z_i, \boldsymbol{\theta}).$$

Zlogaritmování získáme logistickou věrohodnost:

$$l(t,\boldsymbol{\theta}) = -\int_0^A \lambda(x,\boldsymbol{\theta})\mathrm{d}x + \sum_{i=1}^{N(A)} \log(\lambda(Z_i,\boldsymbol{\theta})).$$

Maximálně věrohodný odhad $\hat{\boldsymbol{\theta}}$ získáme maximalizací logaritmické věrohodnosti $l(t,\boldsymbol{\theta})$ dle parametru $\boldsymbol{\theta}$

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\Theta} l(t,\hat{\boldsymbol{\theta}}).$$

Připomeňme zde metodu výpočtu maximálně věrohodného odhadu.

1. Nalezneme první derivaci logaritmické věrohodnosti:

$$U(t,\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} l(t,\boldsymbol{\theta}).$$

2. Adepti pro maximálně věrohodný odhad $\hat{\boldsymbol{\theta}}$ řeší rovnici $U(t,\hat{\boldsymbol{\theta}}) = \mathbf{0}$, neboli

$$\frac{\partial}{\partial \boldsymbol{\theta}} l(t,\hat{\boldsymbol{\theta}}) = \mathbf{0}.$$

3. Dopočteme druhou derivaci logaritmické věrohodnosti a upravíme ji do tvaru:

$$I(t,\boldsymbol{\theta}) = -\frac{1}{N(A)} \frac{\partial^2}{\partial \boldsymbol{\theta}^\top \partial \boldsymbol{\theta}} l(t,\boldsymbol{\theta}).$$

4. Logaritmická věrohodnost je konkávní a odhad $\hat{\boldsymbol{\theta}}$ musí být globálním maximem, pokud platí

$$I(t,\boldsymbol{\theta}) > 0, \qquad \text{pro } \forall \boldsymbol{\theta} \in \Theta.$$

Konkrétní tvar maximálně věrohodného odhadu $\hat{\boldsymbol{\theta}}$, získáme až pro konkrétní parametrickou intenzitu $\lambda(t,\boldsymbol{\theta})$

3. Věta o posunutí

Věta, kterou v této kapitole nazýváme věta o posunutí, je jednorozměrnou variantou Kingmanova Displacement theorem (Kingman, 1993). Velmi zjednodušeně řečeno nám umožní z jednoho Poissonova procesu vytvořit jiný.

Pro názorné vysvětlení se vraťme k příkladu z první kapitoly, kde jsme polohu cyklistů na cyklostezce prezentovali pomocí čítacího procesu $\{N(t),t \geq 0\}$. Nyní budeme procesem $\{N(t),t \geq 0\}$ rozumět Poissonův proces s intenzitou $\lambda(t)$. Opět se odkloníme od pohledu na t jako na čas a pro účely následujícího odstavce budeme t chápat jako polohu na $\mathbb{R}$ (cyklostezce). Naopak budeme předpokládat, že celý proces $\{N\}$ se odehrává v čase a a polohu i-tého cyklisty na cyklostezce v čase a označíme N_i.

Nyní si představme, že uběhne nějaký čas a naším přáním je vyjádřit polohu cyklistů v pozdějším čase b. Věta o posunutí nám říká, že polohu cyklistů v pozdějším čase lze vyjádřit pomocí nového Poissonova procesu $\{M(s)\}$. Proměnná s také náleží prostoru $\mathbb{R}$, ale protože se jedná o polohu cyklisty v jiném čase zavádíme novou proměnou. Polohu i-tého cyklisty v čase b budeme značit T_i. Předpokládáme, že všichni cyklisti se pohybují jedním směrem a že posunutí jednoho cyklisty nezávisí na posunutí druhého, neboli T_i je nezávislé na T_j pro $i \neq j$. Zároveň je zřejmé, že poloha T_i cyklisty v čase b závisí na jeho poloze Z_i v čase a. Nyní zbývá pouze vyjádřit o kolik se daný cyklista posunul. Jak již bylo naznačeno, vzdálenosti ujeté různými cyklisty na sobě vzájemně nezávisí, ale závisí na počáteční poloze cyklisty. Tuto vzdálenost vyjádříme pomocí podmíněné hustoty $\gamma(s|Z_i = t)$.

Uveďme větu o posunutí. Využijeme právě zavedené značení pro intenzity a hustotu. V definici věty se opět vrátíme k pohledu na veličiny t a s jako na čas a veličinu Z_i budeme vnímat jako čas nastání i-té události z procesu $\{N(t),t \geq 0\}$.

Věta 4 (O posunutí)**.** *Nechť* $\{N(t), t \geq 0\}$ *je Poissonův proces na reálné polopřímce* $[0,\infty)$ *s intenzitou* $\lambda_N(t)$ *a nechť* $\{Z_i\}$ *je posloupnost časů výskytů událostí z procesu* $\{N(t)\}$*. Předpokládejme, že body* Z_i *jsou náhodně transformovány takovým způsobem, že pozice transformovaných bodů jsou vzájemně nezávislé. A nechť rozdělení transformovaných bodů má podmíněnou hustotu* $\gamma(s|Z_i = t)$*, kde body* t *náleží prostoru procesu* $\{N\}$ *a body* s *náleží transformovanému prostoru. Pak transformované body tvoří Poissonův proces* $\{M(s),s \geq 0\}$ *s intenzitou*

$$\lambda_M(s) = \int_0^\infty \lambda_N(t)\gamma(s|Z_i = t)\mathrm{d}t.$$

Důkaz. Důkaz Displacement theorem, ze kterého tato věta vychází, je uveden v Kingmanovi (Kingman, 1993).

□

Nyní uveďme několik numerických příkladů na použití věty o posunutí. Budeme uvažovat situaci, kdy proces $\{N\}$ se odehrává před procesem $\{M\}$. Hustota γ transformuje body Z_i dále od počátku. Transformované body budeme značit T_i a rozdíl $T_i - Z_i = W_i$ vyjadřuje rozsah posunutí.

V těchto příkladech budeme volit různé parametrické tvary intenzity λ_N a hustoty γ. Za intenzitu budeme volit buď exponenciálu nebo mocninu. Posunuté body budeme reprezentovat pomocí rozdělení s těžkými chvosty, abychom zajistili, že čím dále od původního bodu Z_i budeme, tím méně pravděpodobný bude výskyt transformovaných bodů T_i.

Příklad 1.
V tomto prvním ilustračním příkladu budeme intenzitu procesu $\{N\}$ reprezentovat pomocí exponenciály. Jako hustotu γ zvolíme hustotu podmíněného exponenciálního rozdělení. Exponenciální hustotu musíme posunout do budu Z_i.

Uveďme parametrické tvary intenzity $\lambda_N(t)$ a hustoty $\gamma(s|Z_i = t)$:

$$\lambda_N(t) = e^{a+bt} \qquad \text{pro } t > 0;$$

$$\gamma(s|Z_i = t) = ce^{-c(s-t)} \qquad \text{pro } t < s < \infty.$$

Je potřeba pamatovat také na definiční obory parametrů: $a \in \mathbb{R}, b \in \mathbb{R}, c > 0$. Do věty o posunutí dosadíme zvolené parametrické funkce. Poznamenejme, že výsledná intenzita $\lambda_M(s)$ je vlastně konvolucí exponenciály a hustoty exponenciálního rozdělení. Integrál bude tvaru:

$$\begin{aligned}
&\int_0^s e^{a+bt} ce^{-c(s-t)} \mathrm{d}t \\
&= ce^a e^{-cs} \int_0^s e^{bt+ct} \mathrm{d}t \\
&= ce^{a-cs} \left[\frac{1}{b+c} e^{(b+c)t} \right]_0^s \\
&= ce^{a-cs} \left[\frac{e^{(b+c)s}}{b+c} - \frac{1}{b+c} \right] \\
&= \frac{c}{b+c} (e^{a+bs} - e^{a-cs}).
\end{aligned}$$

Intenzita Poissonova procesu je nezáporná funkce. Aby dopočtený integrál mohl být intenzitou, musí být kladný. Tento integrál bude kladný pro všechna $a,b \in \mathbb{R}, c > 0$

Příklad 2.
V druhém příkladu intenzita procesu $\{N\}$ bude opět exponenciální funkce a pro podmíněnou hustotu γ zvolíme posunuté Gamma rozdělení:

$$\lambda_N(t) = e^{a+bt} \qquad \text{pro } t > 0;$$

$$\gamma(s|Z_i = t) = \frac{\beta^\alpha}{\Gamma(\alpha)} (s-t)^{\alpha-1} e^{\beta(s-t)} \qquad \text{pro } t < s < \infty.$$

O parametrech víme: $a,b \in \mathbb{R}, \alpha > 0,\beta > 0$. Použitím věty o posunutí dopočítáme intenzitu procesu $\{M\}$:

$$\lambda_M(s) = \int_0^s e^{a+bt} \frac{\beta^\alpha}{\Gamma(\alpha)} (s-t)^{\alpha-1} e^{\beta(s-t)} \mathrm{d}t$$

$$= \frac{\beta^\alpha}{\Gamma(\alpha)} e^{a+\beta s} \int_0^s (s-t)^{\alpha-1} e^{t(b-\beta)} \mathrm{d}t.$$

Řešení tohoto integrálu vede k zavedení neúplné Gamma funkce, což je nad rámec této práce. Je nutné poznamenat, že kdyby parametry a, b a α,β byly známé, pak integrál by stále nebylo možné dopočítat. Proto nebudeme kombinaci exponenciály a hustoty Gamma rozdělení volit k výpočtu intenzity $\lambda_M(s)$.

Příklad 3.
Uveďme poslední příklad, kde intenzitu λ_N modelujeme pomocí exponenciální funkce. Za hustotu γ nyní dosadíme hustotu logaritmicko-normálního rozdělení.

$$\lambda_N(t) = e^{a+bt} \qquad \text{pro } t > 0;$$

$$\gamma(s|Z_i = t) = \frac{1}{(s-t)\sqrt{2\pi}\sigma} \exp\left\{-\frac{[\ln(s-t)-\mu]^2}{2\sigma^2}\right\} \qquad \text{pro} t < s < \infty.$$

Pro parametry platí : $a,b,\mu \in \mathbb{R}$ a $\sigma > 0$. Hustotu λ_M získáme konvolucí intenzity λ_N a hustoty γ:

$$\lambda_M(s) = \int_0^s e^{a+bt} \frac{1}{(s-t)\sqrt{2\pi}\sigma} \exp\left\{-\frac{[\ln(s-t)-\mu]^2}{2\sigma^2}\right\} \mathrm{d}t$$

$$= \frac{e^a}{\sqrt{2\pi}\sigma} \int_0^s \frac{e^{bt}}{s-t} \exp\left\{-\frac{\ln^2(s-t) - 2\ln(s-t)\mu + \mu^2}{2\sigma^2}\right\} \mathrm{d}t$$

$$= \frac{e^a}{\sqrt{2\pi}\sigma} e^{-\frac{\mu^2}{2\sigma^2}} \int_0^s \frac{e^{bt}}{s-t} e^{-\frac{2}{2\sigma^2}\ln(s-t)} e^{\frac{2\mu}{2\sigma^2}\ln(s-t)} \mathrm{d}t$$

$$= \frac{e^{a-\frac{\mu^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} \int_0^s \frac{e^{bt}}{s-t} (s-t)^{-\frac{1}{\sigma^2}} (s-t)^{\frac{\mu}{\sigma^2}} \mathrm{d}t$$

$$= \frac{e^{a-\frac{\mu^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} \int_0^s e^{bt} (s-t)^{\frac{\mu-1}{\sigma^2}-1} \mathrm{d}t.$$

Tento integrál opět vede k neúplné beta funkci a proto ho nedopočítáváme. Tuto konvoluci nebudeme používat k výpočtu intenzity $\lambda_M(s)$.

Příklad 4.
V posledním příkladě zadefinujeme intenzitu jinak než dosud. Použijeme mocninou funkci. Hustota γ bude stejná jako v prvním příkladě.

$$\lambda_N(t) = a + bt + ct^2 \qquad \text{pro } t > 0;$$

$$\gamma(s|Z_i = t) = de^{-d(s-t)} \qquad \text{pro } t < s < \infty.$$

Parametry u mocninné funkce zvolíme všechny kladné, pak platí $a,b,c,d > 0$. Nyní do integrálu z věty o posunutí dosadíme zvolené funkce:

$$
\begin{aligned}
&\int_0^s \left(a + bt + ct^2\right) de^{-d(s-t)} \mathrm{d}t \\
&\quad = \int_0^s ade^{-d(s-t)} + bdte^{-d(s-t)} + cdt^2 e^{-d(s-t)} \mathrm{d}t \\
&\quad = ade^{-ds} \int_0^s e^{dt} \mathrm{d}t + bde^{ds} \int_0^s te^{dt} \mathrm{d}t + cde^{ds} \int_0^s t^2 e^{dt} \mathrm{d}t. \qquad (3.1)
\end{aligned}
$$

Integrály vypočítáme každý zvlášť a pak je do rovnice dosadíme:

$$
\int_0^s e^{dt} \mathrm{d}t = \frac{1}{d}(e^{ds} - 1);
$$

$$
\int_0^s te^{dt} \mathrm{d}t \overset{PP}{=} \left| \begin{matrix} t & e^{dt} \\ 1 & \frac{1}{d}e^{dt} \\ 0 & \frac{1}{d^2}e^{dt} \end{matrix} \right| = \left[\frac{t}{d}e^{dt} - \frac{1}{d^2}e^{dt} \right]_0^s = e^{ds}\left(\frac{s}{d} - \frac{1}{d^2} \right) + \frac{1}{d^2};
$$

$$
\int_0^s t^2 e^{dt} \mathrm{d}t \overset{PP}{=} \left| \begin{matrix} t^2 & e^{dt} \\ 2t & \frac{1}{d}e^{dt} \\ 2 & \frac{1}{d^2}e^{dt} \\ 0 & \frac{1}{d^3}e^{dt} \end{matrix} \right| = \left[\frac{t^2}{d}e^{dt} - \frac{2t}{d^2}e^{dt} + \frac{2}{d^3}e^{dt} \right]_0^s = e^{ds}\left(\frac{s^2}{d} - \frac{2s}{d^2} + \frac{2}{d^3} \right) - \frac{2}{d^3}.
$$

Vypočtené integrály dosadíme do (3.1) a získáme hustotu procesu $\{M\}$:

$$
\begin{aligned}
\lambda_M(s) = ade^{-ds}\frac{1}{d}(e^{ds} - 1) + bde^{-ds}\left[e^{ds}\left(\frac{s}{d} - \frac{1}{d^2} \right) + \frac{1}{d^2} \right] \\
+ cde^{-ds}\left[e^{ds}\left(\frac{s^2}{d} - \frac{2s}{d^2} + \frac{2}{d^3} \right) - \frac{2}{d^3} \right] \\
= a\left(1 - e^{-ds}\right) - b\left(s + \frac{1}{d} - \frac{e^{-ds}}{d} \right) + c\left(s^2 + \frac{2s}{d} + \frac{2}{d^2} - \frac{2e^{-ds}}{d^2} \right).
\end{aligned}
$$

4. Modelování pomocí dvourozměrného Poissonova procesu

Uveďme úlohu z pojišťovnictví. Jednou z hlavních povinností pojišťoven je odhadování počtu pojistných událostí. Pojišťovny se o pojistných událostech nedozví při jejich vzniku, ale až při jejich nahlášení. Mezi vznikem události a jejím nahlášení v praxi existuje prodleva, jejíž délka není předem známá. Tato skutečnost způsobuje, že v současnosti již některé události nastaly, ale pojišťovna je zatím nezaznamenává. Tyto useknutá data jsou jednou z hlavních složek rizika pro pojišťovny. Zpětnou predikcí počtu pojistných událostí se v pojišťovnách zabývá oddělení pojistné matematiky.

Nyní sestavíme model, kterým danou situaci popíšeme. Označme časy výskytů pojistných událostí jako posloupnost časů $\{T_i\}$ a označme časy nahlášení posloupností $\{Z_i\}$. Pro každou událost musí platit $T_i \leq Z_i$, protože událost nemůže být nahlášena před svým vznikem, kdyby tomu tak bylo, jednalo by se o pojistný podvod. Zároveň můžeme soudit, že událost nemůže být nahlášena přesně v okamžiku vzniku. Z tohoto důvodu se omezíme na ostrou nerovnost $T_i < Z_i$. Délka prodlení mezi vznikem a nahlášením závisí na mnoha vnějších faktorech a pro i-tou událost ji budeme značit jako W_i.

Uvažujme situaci, kdy tyto proměnné pozorujeme od času 0 do současnosti A. Některé události, které nastaly do času A, nepozorujeme, protože jejich nahlášení nastane až v čase budoucím $(A,B]$, $B > A$. Data zaznamenávající vzniky událostí jsou v důsledku toho useknutá a pozorujeme pouze takové události pro které platí $Z_i < A$. Vztah zmíněných veličin můžeme vyjádřit jednoduchou rovnicí $Z_i - T_i = W_i$.

Posloupnosti časů vzniku a nahlášení jsou náhodné a budeme je reprezentovat čítacími procesy. Oba případy popisujeme Poissonovým procesem. Je třeba si uvědomit, že použití Poissonova procesu nekoresponduje s reálnou situací. Poissonův proces předpokládá nezávislost přírůstků a obzvlášť v kontextu vzniku pojistných událostí, tento předpoklad není vždy platný. Nicméně použití procesu, který bere v potaz provázané pojistné události, by situaci výrazně komplikovalo a překročilo prostor této práci vyměřený. Poissonův proces reprezentující časy vzniku pojistných událostí značíme $\{M\}$ a jeho intenzitu λ_M. Poissonův proces $\{N\}$ s intenzitou λ_N odpovídá časům nahlášení. Tyto dva procesy spolu tvoří dvourozměrný Poissonův proces, odtud název této kapitoly. Čas prodlení W_i budeme vnímat jako náhodnou veličinu s podmíněnou hustotou $\gamma(s|Z_i = t)$.

Počet nahlášených a tedy i pozorovaných událostí je roven hodnotě procesu $\{N\}$ v čase A. Předpokládejme, že máme k dispozici data

$$\{T_i,Z_i\}_{i=1,\ldots,N(A)}$$

a naším úkolem predikovat počet událostí, které vzniknou od času A do času B.

Odhad intenzity vzniků pojistných událostí $\lambda_M(s)$ získáme následujícím způsobem:

1. Mějme parametrickou formu intenzity hlášení událostí $\lambda_N(t,\boldsymbol{\theta})$. Pomocí metody maximální věrohodnosti získáme odhady parametrů $\hat{\boldsymbol{\theta}}$.

2. Podmíněná hustota časů prodlení $\gamma(s|Z_i = t,\boldsymbol{\tau})$ je parametrická funkce s parametrem $\boldsymbol{\tau}$. Vypočteme maximálně věrohodný odhad parametru $\hat{\boldsymbol{\tau}}$.

3. Získané odhady parametrů dosadíme zpět do parametrických funkcí a získáme odhad intenzity $\lambda_N(t,\hat{\boldsymbol{\theta}})$ a odhad hustoty $\gamma(s|Z_i = t,\hat{\boldsymbol{\tau}})$.

4. Aplikací těchto odhadů ve větě o posunutí obdržíme odhad intenzity vzniků událostí $\lambda_M(s,\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\tau}})$.

Predikci procesu nyní provedeme na reálných datech. Veškeré výpočtu jsou prováděny s pomocí programovacím jazyku R (R Core Team, 2021) v prostředí RStudio (RStudio Team, 2015).

4.1 Data

Dataset, se kterým budeme pracovat, pochází od České kanceláře pojistitelů. Obsahuje záznamy o pojistných událostech způsobených automobily od roku 2000 do roku 2019. Informace dostupné ke každému záznamu jsou rozděleny do 6 kategorií:

- ID číslo přiřazené k pojistné události;
- Type povaha škodné události (například zda se jednalo pouze o poškození vozidla nebo zda byl i někdo zraněn);
- Accident čas vzniku pojistné události (počítaný ve dnech od roku 1900);
- Reporting čas nahlášení pojistné události;
- Payment čas výplaty pojistného plnění pojistníkovi;
- Amount výše pojistného plnění v českých korunách.

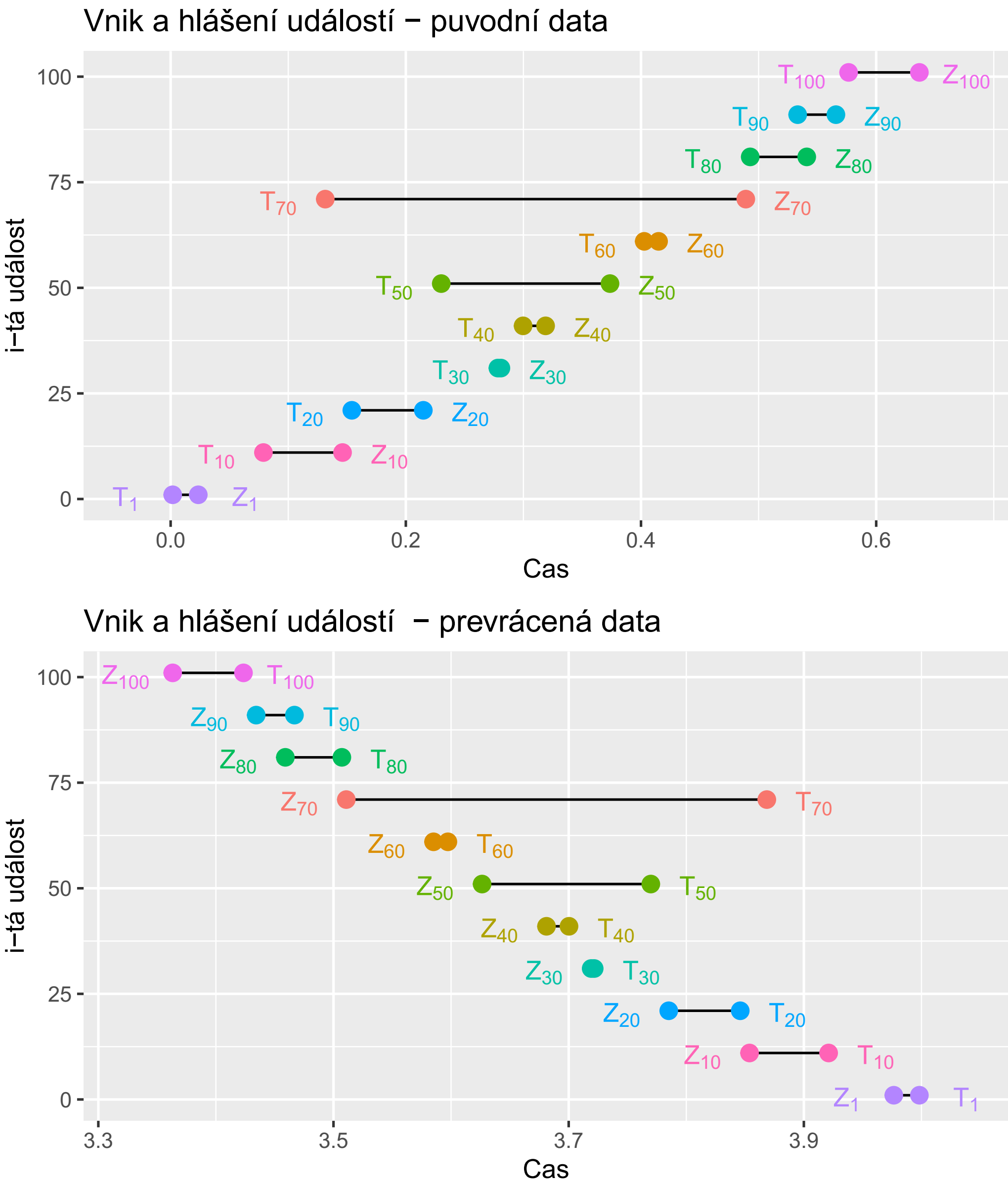
Vzhledem k charakteru našeho cíle se budeme zabývat pouze časy vzniku a nahlášení pojistných událostí. Navíc rozdělíme data podle povahy škodné události. Budeme se věnovat pojistným událostem, při kterých došlo k ublížení na zdraví. Zaměříme se na časový úsek od 1.1.2013 do 31.12.2016. Transformujeme časy v našem datasetu tak, že je budeme počítat v letech. Tato transformace má své numerické výhody při hledání intenzity vzniků událostí. Navíc označíme datum 1.1.2013 jako 0 a tedy datum 31.12.2017 budeme značit 4.

Abychom mohli ověřit správnost našich výsledků, rozdělíme data na trénovací a testovací. Za trénovací budeme považovat data událostí, které vnikly po 1.1.2013 a byly nahlášeny do 30.9.2016. Testovací data se budou týkat událostí z posledního čtvrtletí roku 2016.

4.2 Otočení časů událostí

Aby naše data byla přizpůsobena teorii a numerickým příkladům z kapitoly 3 je potřeba je transformovat. Vztah časů vzniků a nahlášení se dá vyjádřit následujícím způsobem $T_i < Z_i$ a platí $Z_i - T_i = W_i$, kde W_i je čas prodlení mezi vnikem a nahlášením. V kapitole 3 však pracujeme s náhodnými veličinami pro které platí $Z_i < T_i$ a $T_i - Z_i = W_i$

Tento problém nás vede k transformaci našich dat. Značení časů pomocí této transformace obrátíme. Tím je myšleno, že datum 1.1.2013 budeme značit 4 a datum 31.12.2016 budeme značit 0. Například událost, která původně nastala v čase 0,75, nyní nastane v čase 3,25. Oba procesy $\{N\}$ a $\{M\}$ plynou od času 4 do času 0. Je třeba si uvědomit, že za platnosti této transformace, pro všechna i platí $Z_i < T_i$ a tedy vztah mezi vznikem a nahlášením odpovídá $T_i - Z_i = W_i$, kde $\forall i : W_i > 0$.



Obrázek 4.1: Převrácení časů událostí.

4.3 Odhady pro intenzitu $\lambda_N(t)$

Budeme uvažovat parametrické intenzity $\lambda_N(t)$ z kapitoly 3. Pro odhady parametrů intenzit využijeme teoretické poznatky z kapitoly 2 a funkci mle2() z balíčku "bbmle". Odhady parametrů jsou uvedeny v tabulce 4.1.

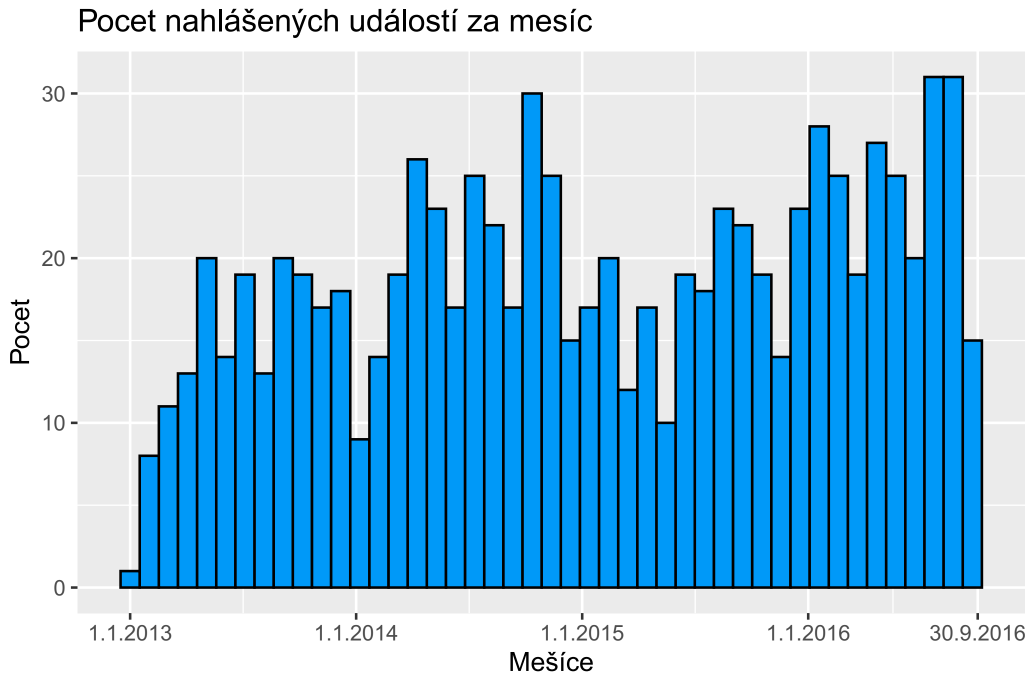
Intenzita	**Odhady parametrů**	
e^{a+bt}	$\hat{a} =$	5,4431
	$\hat{b} =$	$-0{,}0427$
$a + bt + ct^2$	$\hat{a} =$	121,8706
	$\hat{b} =$	147,4432
	$\hat{c} =$	$-38{,}3026$

Tabulka 4.1: Maximálně věrohodné odhady parametrů intenzity procesu $\{M\}$.

Odhady parametrů exponenciální funkce můžeme interpretovat následovně. V čase 0, který vzhledem k otočení dat odpovídá 31.12.2016, je intenzita hlášení událostí nejvyšší. Kdyby byla intenzita konstantní, pak by odpovídala přibližně $e^{5{,}44} \approx 231$ událostem za rok. S vzrůstajícím t, tedy díky transformaci s nižším datem nahlášení události, se intenzita hlášení snižuje. Za platnosti těchto odhadů by 1.1.2013 intenzita hlášení odpovídala $e^{5{,}44-0{,}04\cdot 4} = 195$ událostem za rok.

Obdobně lze interpretovat i odhady parametrů pro mocninou funkci.

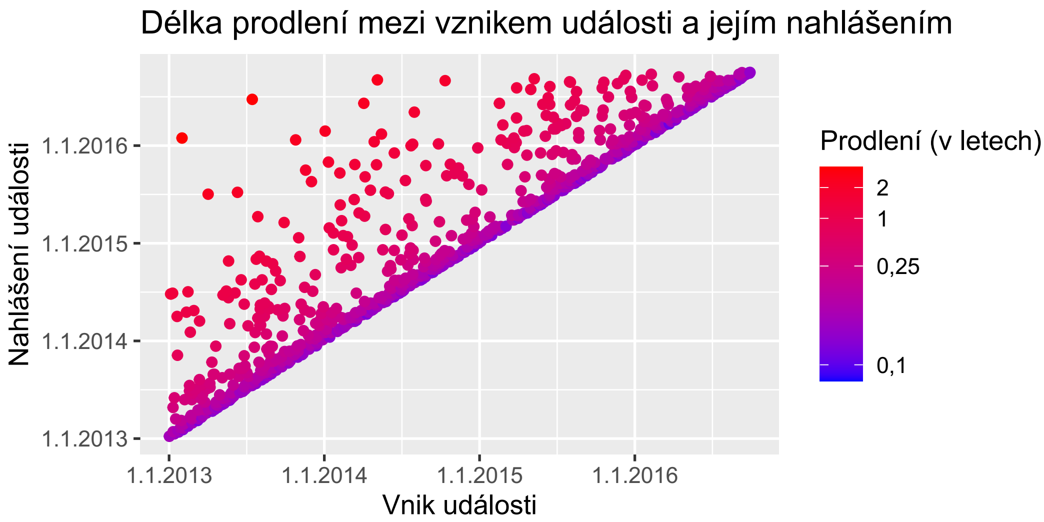
Na obrázku 4.2 je uveden histogram, který znázorňuje počty škodných událostí v jednotlivých měsících. Se zvyšujícím se datem počet nahlášených událostí mírně stoupá. Toto empirické pozorování odpovídá našim maximálně věrohodným odhadům.



Obrázek 4.2: Histogram počtu nahlášených událostí za měsíc.

4.4 Odhady pro hustotu $\gamma(s|Z_i = t)$

Hustota $\gamma(s|Z_i = t)$ je podmíněná hustota veličiny W_i, která určuje délku prodlení mezi nahlášením a vznikem události. Na obrázku 4.3 jsou znázorněny časy prodlení. Odpovídají vertikální vzdálenosti bodů od diagonály. Z vizualizace lze odvodit, že pro hustotu $\gamma(s|Z_i = t)$ musíme volit rozdělení s těžkými chvosty, protože výskyty extrémních hodnot jsou vzácné. Budeme volit posunuté exponenciální rozdělení.



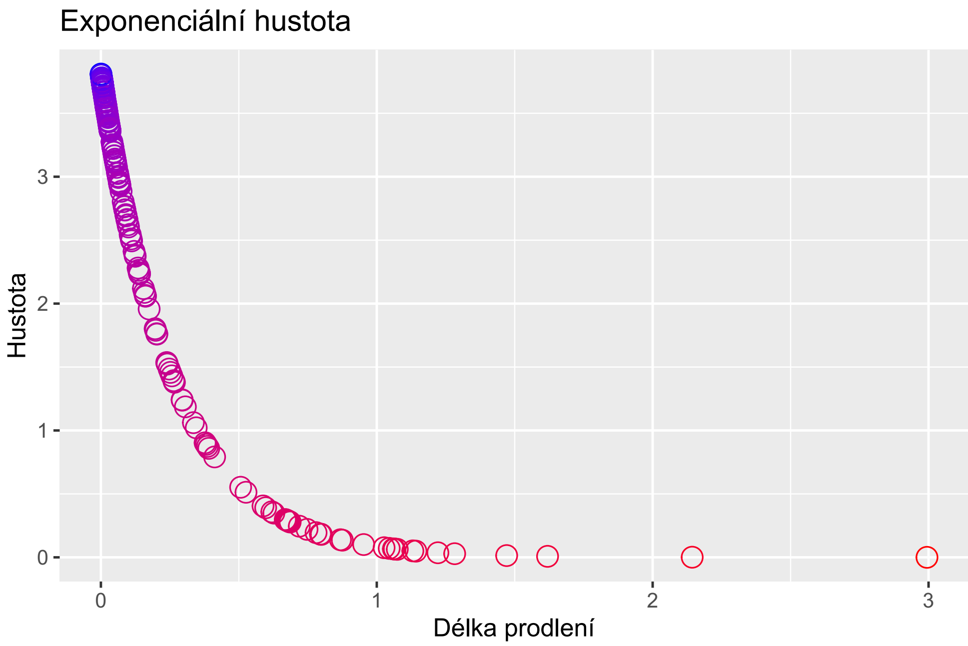
Obrázek 4.3: Reálné délky časů prodlení.

Parametry hustot odhadneme pomocí metody maximální věrohodnosti. Teorie vztahující se k maximálně věrohodným odhadům parametrů hustot je dobře známá a proto ji zde neuvádíme. Odhady parametrů shrneme v následující tabulce.

Hustota	**Odhady parametrů**	
$ce^{-c(s-t)}$	$\hat{c} =$	3,8150

Tabulka 4.2: Maximálně věrohodné odhady parametrů intenzity procesu $\{M\}$.

Obrázek 4.4 zobrazuje hustotu veličiny W_i vykreslené za pomoci odhadnutých parametrů. Graf vykreslujeme bodově a pro větší přehlednost využíváme pouze 200 náhodně vybraných časů prodlení z našeho datasetu. Úseky, kde se body slévají do jednolité přímky naznačují, že v těchto časech dochází k většímu množství událostí. Z této grafické reprezentace je možné usoudit, že exponenciální hustota dobře reprezentuje situaci prodlení. Je však nutné poznamenat, že se stále jedná pouze o grafické znázornění a neprovádíme žádné testy, kterými bychom parametry ověřovali. Testy ověřující spolehlivost maximálně věrohodných odhadů jsou známé, ale jejich aplikace není předmětem této práce.



Obrázek 4.4: Hustota časů prodlení.

4.5 Predikce počtu vzniklých událostí

Získané odhady parametrů intenzity hlášení škod a hustoty prodlení dosadíme do příkladů z kapitoly 3. Tímto dosazením získáme konkrétní tvary intenzity λ_M. Pro predikce počtu vzniklých událostí v posledním čtvrtletí využijeme funkci střední hodnoty zmíněnou v kapitole 1.

Z kapitoly 3 se budeme konkrétně zabývat příklady 1 a 4.

Příklad 1. Parametrický předpis intenzity procesu $\{M\}$ je následujícího tvaru :

$$\lambda_M(s) = \frac{c}{b+c}(e^{a+bs} - e^{a-cs}).$$

Pomocí získané intenzity a funkce střední hodnoty dopočítáme očekávaný počet vzniků v období od 1.10.2016 do 31.12.2016. Připomeňme, že funkce střední hodnoty odpovídá integrálu intenzity přes uvažované období. Stále pracujeme s otočenými daty a proto budeme integrovat od 0 do 0,25:

$$\mathsf{E}\left[M(0{,}25)\right] = \int_0^{0{,}25} \lambda_M(s,\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\tau}})\mathrm{d}s$$

$$= 21{,}09.$$

Příklad 4.
V kapitole 3 jsem v příkladu 4 získali parametrickou formu intenzity ve tvaru:

$$\lambda_M(s) = a\left(1 - e^{-ds}\right) - b\left(s + \frac{1}{d} - \frac{e^{-ds}}{d}\right) + c\left(s^2 + \frac{2s}{d} + \frac{2}{d^2} - \frac{2e^{-ds}}{d^2}\right).$$

Odhad střední hodnoty procesu vniku škod v posledním čtvrtletí je

$$E[M(0{,}25)] = 17{,}58.$$

Pokud výsledky porovnáme s reálnými daty není predikce příliš úspěšná. Reálný počet vzniklých událostí v posledním čtvrtletí je 44 událostí. Tuto nekonzistenci predikce s realitou můžeme odůvodnit nesprávnou volbou intenzit. Lze totiž předpokládat, že intenzita vzniků škodných událostí bude mít sezónní charakter. Například v zimních obdobích lidé častěji havarují, protože se zvyšují rizika spojená se zasněženými silnicemi.

Závěr

Vzhledem k odchylce našich výsledků s reálnými daty je zřejmé, že námi zvolené intenzity a hustoty nevhodně reprezentují skutečnost. Ale navzdory nepříznivým výsledkům praktické části věříme, že modelování počtu budoucích událostí pomocí predikce useknutého procesu vzniku událostí je krok správným směrem. Tento pohled nám umožňuje využít všechna data, která mají pojišťovny k dispozici. Tím máme na mysli, že pro každou událost se zabýváme časem vzniku a časem jejího nahlášení. Výhoda tohoto přístupu oproti agregovaným datům za delší časová období spočívá v možnosti zaznamenat i malé změny ve frekvenci dat způsobené například sezónním chováním nebo aktuální strategií pojišťoven.

To, že jsme této vlastnosti nevyužili, vedlo k nekonzistenci mezi našimi odhady a realitou v praktické části. Při budoucí aplikaci tohoto typu predikce bychom doporučili zvolit jiné tvary intenzity hlášení $\lambda_N(t)$. Využití goniometrických funkcí v této intenzitě by nám umožnilo uvažovat sezónnost našich dat. Jako rozdělení časů prodlení by bylo vhodné volit spíše logaritmicko-normální nebo gamma rozdělení, protože lépe modelují skutečnost než exponenciální rozdělení. Většina událostí je nahlášena v několika následujících dnech po pojistné události. Bylo by zajímavé předpokládat, že parametry těchto hustot mohou záviset na čase.

Praktické využití v pojišťovnách by se navíc dalo ještě zlepšit tím, že ke každé události bychom přiřadili náhodnou veličinu, která by reprezentovala výši škody způsobené pojistnou událostí. Pak bychom mohli, kromě budoucího počtu pojistných událostí, predikovat i výši závazků pojišťoven vůči pojištěným.

Seznam použité literatury

Andersen, P. K., Borgan, O., Gill, R. D. a Keiding, N. (1992). *Statistical models based on counting processes.* II. Series. Springer Science & Business Media. ISBN 0-387-94519-9.

Kingman, J. F. C. (1993). *Poisson processes*, volume 3 of *Oxford Studies in Probability.* The Clarendon Press Oxford University Press, New York. ISBN 0-19-853693-3. Oxford Science Publications.

Lawless, J. F. (1987). Regression methods for poisson process data. *Journal of the American Statistical Association*, **82**(399), 808–815. ISSN 01621459.

Maciak M., Okhrin O., P. M. (2021). Infinitely stochastic micro reserving. *Insurance, mathematics economics*, **100**, 31.

R Core Team (2021). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. URL `https://www.R-project.org/`.

Ross, S. M. (1996). *Stochastic processes.* 2nd ed. John Wiley & Sons, New York. ISBN 04-7112-062-6.

Ross, S. M. (2010). Chapter 5 - the exponential distribution and the poisson process. In Ross, S. M., editor, *Introduction to Probability Models (Tenth Edition)*, pages 291–370. Academic Press, Boston, tenth edition edition. ISBN 978-0-12-375686-2.

RStudio Team (2015). *RStudio: Integrated Development Environment for R.* RStudio, Inc., Boston, MA. URL `http://www.rstudio.com/`.

Thompson, W. A. (2012). *Point Process Models with Applications to Safety and Reliability*, volume 1. Springer New York, NY, New York. ISBN 978-1-4613-1067-9. Oxford Science Publications.

Tijms, H. C. (2003). *A first course in stochastic models.* John Wiley and Sons. ISBN 0-471-49881-5.

Zhao, M. a Xie, M. (1996). On maximum likelihood estimation for a general non-homogeneous Poisson process. *Scandinavian Journal of Statistics*, **23**, 597–607.