

## MODELY 2D BODOVÝCH PROCESŮ A JEJICH VYUŽITÍ V ANALÝZE NÁHODNÝCH SOUČTŮ

PETR VOLF

ABSTRACT. The contribution deals with the compound (cumulative) random process, i. e. the sequence of random increments at random moments. The process is modeled as a 2D point process (or 2D random counting measure), via the intensities of both components. The study starts from the case of compound Poisson process and generalizes to the cases when the components depend on each other, and also on a set of covariates.

Резюме. В статье изучаются случайные кумулятивные процессы. Они представлены как 2-мерные точечные процессы (или точечные меры), их модель построена на основе 2D интенсивности, которая одновременно кажется компенсатором процесса. Разные варианты модели обсуждены (включая модель регрессионную) и представлены консистентные оценки интенсивностей. В качестве примера исследованы процессы (случайные суммы) финансовых транзакций.

### 1. KUMULATIVNÍ PROCES A JEHO MODEL

Procesy náhodných přírůstků v náhodných okamžicích (kumulativní procesy, složené – compound – procesy) mají trajektorie v rovině, kde 1. osa je čas ( $t$ ) a 2. osa je hodnota součtu ( $y$ ). Tradiční model popisuje zvláště bodový proces časů a rozdělení přírůstků (hustotou, distribuční funkcí, momenty – viz i Volf, Robust 2000). Obě složky mohou vzájemně záviset, resp. mohou záviset na společné historii a kovariátách. Zkusíme tuto situaci modelovat jako 2D bodový proces, tj. i složku  $y$  popsat intenzitou. Zřejmý je takový popis v případě, kdy okamžiky přírůstků jsou dány Poissonovým procesem a přírůstky jsou i.i.d. s exponenciálním rozdělením - pak jde i ve směru  $y$  o Poissonův proces. Budeme se zabývat možnostmi tohoto modelu pro obecnější situace, především z hlediska formulace modelu, odhadu jeho komponent a predikce průběhu procesu. Zmíníme i možné aplikace, konkrétně posloupnosti finančních transakcí či pojistných plnění, kde otázka predikce úzce souvisí s otázkou pravděpodobnosti bankrotu (ruin probability, risk process).

Uvažujme tedy proces náhodných přírůstků v náhodných časech, tzv. kumulativní či složený (compound) bodový proces

$$C(t) = \int_0^t X(s) dN(s) = \sum_{T_i \leq t} X(T_i),$$

kde  $N(t)$  je čítací proces ( $T_i$  jsou jeho náhodné body, tj. bodový proces  $0 < T_1 < T_2 < \dots$ ,  $N(t) = \sum_{j=1}^{\infty} 1[T_j \leq t]$ ) a  $X(t)$  je náhodná veličina – přírůstek v čase  $t$ . I zde tedy vystupuje zvláště proces časových bodů a zvláště přírůstky. Jak jsme řekli,

---

2000 *Mathematics Subject Classification.* 62N05, 62P30, 60K05.

*Klíčová slova.* Random point process, compound process, intensity.

Práce na této tématice je podporována grantem GA ČR č. 402/01/0539.

jako alternativu budeme modelovat i proces přírůstků pomocí rizikové funkce, případně intenzity. Rizikovou funkcí přitom myslíme nezápornou funkci, charakteristiku rozdělení pravděpodobnosti, zatímco intenzitu budeme chápat jako charakteristiku z rizikové funkce odvozenou, která udává aktuální pravděpodobnost výskytu náhodného bodu. Jaké výhody má model založený na rizikové funkci?

- (1) Riziková funkce je rovnocenná charakteristika distribuce náhodné veličiny spojitého typu, definovaná jako  $h(t) = f(t)/(1 - F(t))$ . Statisticky je výhodnější analyzovat kumulativní rizikovou funkci  $H(t) = \int_0^t h(s) ds$ , jejíž běžný odhad (Nelsonův–Aalenův) má stejně dobré asymptotické vlastnosti jako empirická distribuční funkce.
- (2) Obecněji, intenzita charakterizuje ‘lokální pravděpodobnost’ výskytu náhodného bodu (zpravidla podmíněnou, například ‘historií’), aniž by musela být spojována s nějakou náhodnou veličinou – tak je tomu u čítacích procesů. (opět viz literatura o čítacích procesech, včetně teorie martingalů a kompenzátorů, také Volf, 1992.)
- (3) Nejjednodušší situace, popsaná Poissonovým procesem časových bodů a i.i.d. exponenciálními přírůstky (nezávislémi i na čase), tzv. složený Poissonův proces, si o takovýto popis přímo říká. V takovémto případě je parametr exponenciálního rozdělení zároveň rizikovou funkcí tohoto rozdělení, a také výsledného Poissonova procesu.

Obecně tedy půjde o model 2D bodového procesu (či 2D čítací náhodné míry). Procesy tohoto typu se zabývali např. Greenwood a Wefelmeyer (1994), základní monografií je třeba Jacod a Shirjajev (1987).

## 2. SLOŽENÝ POISSONŮV PROCES

Uvažujme nejdříve homogenní Poissonův proces okamžiků přírůstků, s konstantní rizikovou funkcí  $h$ , a i.i.d. exponenciální přírůstky s parametrem  $g$  (tj.  $g$  je riziková funkce,  $EX = 1/g$ ). Příslušný složený Poissonův proces  $C(t)$  (předpokládejme, že je pozorovaný na  $[0, T]$ ) se tedy skládá ze 2 nezávislých homogenních Poissonových procesů, které vlastně odpovídají projekci  $C(t)$  do osy  $x = t$  (coby vodorovné osy) a osy  $y = c$  (svislé). V ose  $t$  pozorujeme bodový (či čítací) proces  $N(t)$  s náhodnými body  $0 < T_1 < T_2 < \dots < T_K$ ,  $K = N(T)$ . Ve směru osy  $y$  pozorujeme náhodné body  $0 < Y_1 < Y_2 < \dots < Y_K$ , kde  $Y_j = C(T_j)$ ,  $Y_j - Y_{j-1} = X_j$  jsou ony exponenciální přírůstky ( $Y_0 = 0$ ). Maximálně věrohodné odhady parametrů jsou  $\hat{h} = N(T)/T$ ,  $\hat{g} = S(C(T))/C(T)$ , kde jsme označili  $S(y) = \sum_{j=1}^K 1[Y_j \leq y]$  čítací proces ‘počítající’ náhodné body ve směru  $y$ . Samozřejmě  $S(C(T)) = N(T)$ .

**2.1. Proces rizika - odbočka do finanční matematiky.** Představme si velice zjednodušený případ, kdy výše uvedený proces  $C(t)$  je proces pohledávek, které jsme nuceni zaplatit, a k dispozici na to máme spojitě vybírané příspěvky, které dosahují velikosti  $u + vt$  v čase  $t$ . Potom proces  $R(t) = u + vt - C(t)$  je proces rizika („risk process“) a jakmile je  $R(t) < 0$ , můžeme začít uvažovat o bankrotu. Základní řešenou otázkou tedy je, jak nastavit parametry  $u$  a  $v$  (tj. počáteční rezervu a parametr růstu rezervy) tak, aby pravděpodobnost ‘zruinování’ (tj. jevu  $\inf_t R(t) < 0$  v  $[0, \infty)$ , případně v  $[0, T]$ ) byla dostatečně malá. Nechci zde zdržovat přetlumočením výsledků, které jednak sám dobře neznám, jednak lze nalézt v literatuře, např. Rolski et al (1999), Asmussen (2000), Embrechts et al (1997). Pro případ uvedený zde a nekonečný časový horizont je úloha řešena přesně

$$P\left(\inf_t R(t) < 0\right) = \frac{h}{gv} \exp\left(-\frac{(v-h/g) \cdot u}{v/g}\right) = (1-p) e^{-pug},$$

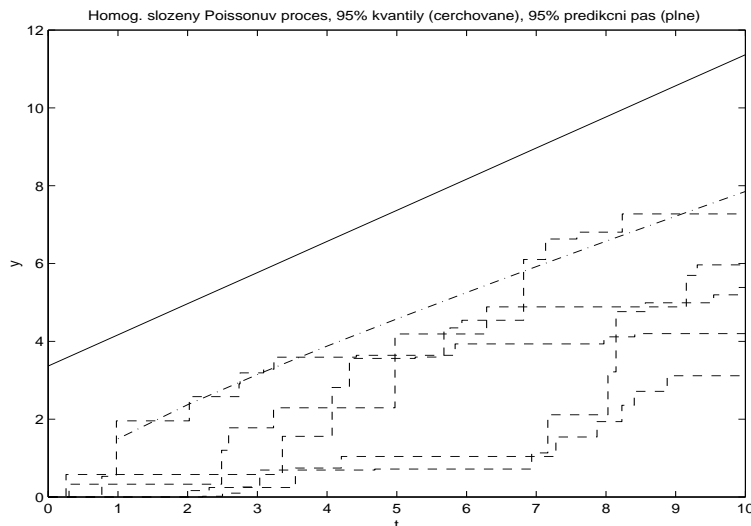
kde  $p = (v - h/g)/v$  vlastně porovnává rychlost růstu rezervy  $v$  se střední hodnotou požadavků (za jednotku času)  $h \cdot EX$  (a mělo by být  $p > 0$ ).

Samozřejmě si můžeme představit situace podobného typu z jiných oblastí, např. problém překročení kritické hladiny v přehradě, nebo zahlcení obslužné linky. Modely náhodných procesů pro popis těchto situací se pak neomezují jen na čítací procesy, ale používají se i jiné třídy, např. difusní procesy, nebo tzv. Lévyho procesy – to není ovšem téma této práce.

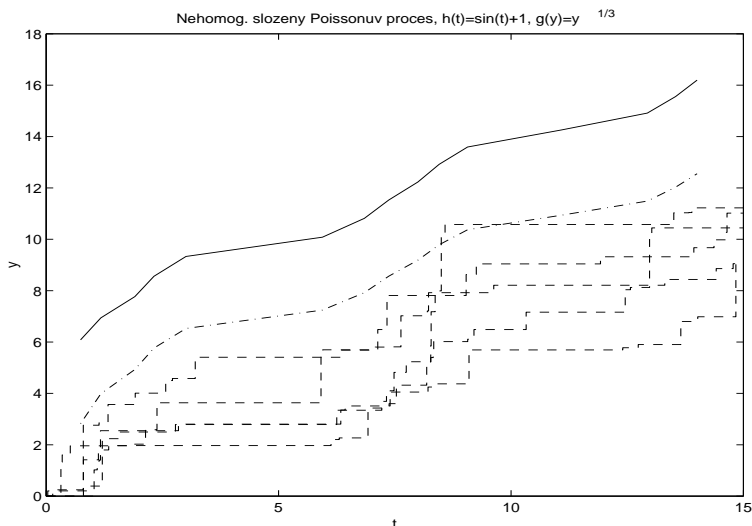
Případy s jinými distribucemi přírůstků a případy s konečným horizontem vedou na složitější řešení (viz zmíněná literatura). Uvědomme si také, že řešení – nalezení vhodné přímky  $u + vt$  – také řeší úlohu predikčního pásu, tj. tato přímka udává horní hranici, přes kterou se proces dostane jen se zvolenou pravěpodobností.

Na obrázku 1 byly nejprve zkonstruovány v každém bodě  $t = 1, 2, \dots, K$  95% kvantily rozdělení veličiny  $Y_t = \sum_{j=1}^t X_j$ , tj. gamma rozdělení s hustotou  $f(y) = e^{-gy} y^{t-1} g^t / \Gamma(t)$ , čerchovaná linka je spojuje, a pak byla spočtena horní 95% predikční hranice  $u + v(t)$  se směrnici  $v = 0.8$  a s  $u = 0.368$  vypočteným k danému  $v$  a  $\alpha = 5\%$  (plná čára).

**2.2. Nehomogenní Poissonovy procesy.** Uvažujme teď trochu složitější situaci, a to že obě složky budou nehomogenní Poissonovy procesy, s rizikovou funkcí  $h(t)$  pro proces časových okamžiků a s rizikovou funkcí  $g(y)$  procesu postupného nárůstu  $C(t)$ , tj. jeho projekce do osy  $y$ . Nezabýváme se nyní tedy rozdělením velikosti jednotlivých přírůstků, i když model pro přírůstky bychom mohli formulovat obdobně. Vztah mezi součtem a jednotlivými přírůstky je dán konvolucí, což není vztah zrovna jednoduchý. Je řešitelný pomocí charakteristických funkcí, pokud bychom vyšli z nějakého konkrétního rozdělení přírůstků.



**Obr. 1** Několik trajektorií složeného Poissonova procesu s  $h = 1$ ,  $g = 2$ .



**Obr. 2** Trajektorie, 95% kvantily a predikční křivka pro nehomogenní složený Poissonův proces.

Ve směru  $t$  tedy opět pozorujeme čítací proces  $N(t)$  a jeho body  $0 < T_1 < T_2 < \dots < T_K < T$ , kde  $K = N(T)$ , ve směru  $y$  čítací proces  $S(y)$  s body  $0 < Y_1 < Y_2 < \dots < Y_K = C(T)$ . Pozorovaný proces  $C(t)$  v daném  $t$  můžeme vyjádřit i jako  $C(t) = \max_{s \leq t} \int_0^\infty y dS(y) dN(s)$ . Zatím předpokládáme, že proces  $S(y)$  nezávisí na  $t$  (což samozřejmě není příliš realistické, zato je realistické, že jeho riziková funkce závisí na jeho stavu, výši  $y$ ).

Představme si nyní, že pozorujeme data  $n$  (vzájemně nezávislých) realizací tohoto 2D bodového procesu. Věrohodnostní funkci pro  $h(t)$  můžeme napsat takto:

$$V_i = \prod_{i=1}^n \left\{ \prod_{t>0} h(t)^{dN_i(t)} \cdot \exp \left( - \int_0^T h(t) dt \right) \right\},$$

kde  $i = 1, 2, \dots, n$  jsou indexy jednotlivých realizací. Po logaritmování a řešení věrohodnostních rovnic dostaneme odhad Nelsonova–Aalenova typu pro kumulativní rizikovou funkci  $H(t) = \int_0^t h(s) ds$  (znovu, viz literatura, např. Volf, 1992):

$$\hat{H}(t) = \sum_{i=1}^n \int_0^t \frac{dN_i(s)}{n} = \frac{1}{n} \sum_{i=1}^n N_i(t^-) = \frac{1}{n} \sum_{i=1}^n \sum_j 1[T_{ij} < t],$$

kde jsme  $T_{ij}$  označili  $j$ -tý bod  $i$ -tého procesu.

Obdobně, věrohodnostní funkci pro  $g(y)$  bychom mohli psát jako

$$V_y = \prod_{i=1}^n \left\{ \prod_{y>0} g(y)^{dS_i(y)} \cdot \exp \left( - \int_0^\infty g(y) J_i(y) dy \right) \right\},$$

kde novým prvkem je indikátor  $J_i(y) = 1$  pro  $y \leq C_i(T)$ ,  $= 0$  jinak. V daném  $y$  tedy  $\sum_{i=1}^n J_i(y)$  říká, kolik realizací dosáhlo celkového součtu  $C_i(T) \geq y$ . Opět můžeme získat odhad pro  $G(y) = \int_0^y g(x) dx$ :

$$\hat{G}(y) = \sum_{i=1}^n \int_0^y \frac{dS_i(x)}{\sum_{k=1}^n J_k(x)} = \sum_{i=1}^n \sum_j \frac{1[Y_{ij} < y]}{\sum_{k=1}^n J_k(Y_{ij})}.$$

Tyto odhady jsou konzistentní (stejněměrně), asymptoticky normální, jak plyne z dnes již standardní teorie čítacích procesů (viz Andersen et al, 1993 a také Volf, 1992).

Odhady rizikových funkcí se získají jádrovým vyrovnáním, tj. (označme  $\mathcal{K}$  zvolené jádrové funkce):

$$\hat{h}(t) = \frac{1}{d_1} \sum_{i=1}^n \sum_j \mathcal{K}_1 \left( \frac{t - T_{ij}}{d_1} \right) \Delta \hat{H}(T_{ij}), \quad \hat{g}(y) = \frac{1}{d_2} \sum_{i=1}^n \sum_j \mathcal{K}_2 \left( \frac{y - Y_{ij}}{d_2} \right) \Delta \hat{G}(Y_{ij}).$$

Pokud jde o predikci, snadno ji odvodíme na základě analogie se situací pro homogenní procesy. Riziková funkce  $h(t)$  resp.  $g(y)$  vlastně představují změnu stupnice na ose  $t$  resp.  $y$ . Konkrétně, jsou-li body  $T_j$  (resp.  $Y_j$ ) realizovány procesem s rizikovou funkcí  $h(t)$  (resp.  $g(y)$ ), pak body  $\tau_j = H(T_j)$  ( $w_j = G(Y_j)$ ) jsou reprezentací složeného Poissonova procesu s rizikovou funkcí  $h^* \equiv 1$  v ose  $t$  a se standardními exponenciálními přírůstky, tj. také s  $g^* \equiv 1$ . Takže, pokud pro tento ‘jednotkový’ proces má v bodě  $t$  predikční (95 % např.) křivka hodnotu  $q(t)$  (například může jít o přímkou  $u + v \cdot t$  spočtenou dle úvah v minulém odstavci), pak pro náš nehomogenní proces dostaneme v bodě  $H^{-1}(t)$  hodnotu  $G^{-1}(q(t))$ .

Funkce  $H(t)$ ,  $G(t)$  jsou spojitě, zpravidla ryze rostoucí, takže mají inverzní funkce. Horší je to pokud máme k dispozici jen odhady, které jsou skokovité, po částech konstantní, ale velikost skoků je řádu  $n^{-1}$  a jak jsme řekli, odhady jsou stejněměrně konzistentní (na ohraničených intervalech  $[0, T] \times [0, Y]$ ), takže alespoň v limitě dostaneme přesnou predikční hranici.

Na obr. 2 je několik trajektorií procesu s  $h(t) = \sin(t) + 1$ ,  $g(y) = y^{1/3}$ , dále opět čerchované spojené 95% kvantily v jednotlivých  $t$  a 95% predikční hranice, obojí vzniklé uvedenou transformací měřítek. V případě predikční hranice jsme takto transformovali přímkou  $u_1 + v_1 \cdot t$  spočtenou pro standardní proces, konkrétně jsme vzali  $v_1 = 2v$ ,  $u_1 = 2u$ ,  $u, v$  z případu na obr. 1, kde bylo  $EX = 1/2$ .

### 3. OBECNÝ MODEL, KOMPENZÁTOR, FILTRACE, ODHADY

Budeme se nyní věnovat dosti obecnému popisu kumulativního procesu. Budeme jej modelovat jako 2D bodový proces, jehož komponenty  $t$  a  $y$  nejsou ‘v rovnováze’, tj. realizace nějakého dalšího bodu  $Y_j$  nastane právě jen když dojde k realizaci bodu  $T_j$ . Nyní již nevystačíme se zcela odděleným popisem složky  $t$  a  $y$ , neboť obě mohou být závislé vzájemně i na nějaké společné filtraci, kovariátách apod.

V průmětu do osy  $t$  pozorujeme standardní čítací proces, který spolugeneruje filtraci  $\mathcal{F}(t)$ , má predikovatelnou (náhodnou) intenzitu  $\lambda(t)$  (tj. adaptovanou na  $\mathcal{F}(t^-)$ ). Filtrace může být mezi jinými spolugenerována i skupinou (náhodných) kovariát  $z(t)$  – opět predikovatelných, mezi nimi i aktuálním stavem procesu  $C(t^-)$ , a případně indikátorem pozorovatelnosti  $I(t)$ . Dále,  $M(t) = N(t) - L(t)$  je martingal, kde  $L(t) = \int_0^t \lambda(s) ds$ , a zpravidla se předpokládá, že existuje riziková funkce  $h(t, z) \geq 0$ , měřitelná, omezená, taková že  $\lambda(t) = h(t, z(t)) \cdot I(t)$ . Zde tedy vidíme názorně vztah mezi rizikovou funkcí a intenzitou.

Pokud pozorujeme zároveň  $i = 1, 2, \dots, n$  procesů, předpokládá se, že jejich inovace  $dM_i(t)$  jsou podmíněně (při dané  $\mathcal{F}(t^-)$ ) nezávislé.

Jak nyní vypadá čítací proces  $S(y)$  pozorovaný v průmětu do osy  $y$ ? Opět budeme uvažovat jeho (náhodnou) intenzitu  $\mu(y)$  a příslušnou filtraci  $\sigma(y)$  resp. její zleva spojitou verzi  $\sigma(y^-)$ , na kterou je  $\mu(y)$  adaptovaná. Otázka je, jak popsat souvislost  $\sigma(y)$  s  $\mathcal{F}(t)$ . Jedna možnost je následující:

S využitím trajektorií procesu  $C(t)$  (uvažujeme je spojité zprava), je možné definovat k danému  $y$  příslušné  $t_y = C^{-1}(y) = \sup\{t : C(t) < y\}$ . Dále,  $\sigma(y^-)$  bude generována funkcemi měřitelnými vzhledem k  $\mathcal{F}(t_y) \otimes \mathcal{B}[0, y]$  a spojitými zleva v  $y$ . Nakonec,  $\sigma(y) = \sigma\{\sigma(y^-) \cup d\mathcal{M}(y)\}$ , kde  $d\mathcal{M}(y) = dS(y) - d\mu(y)$  je opět martingalový přírůstek, inovace procesu, nezávislá na  $\sigma(y^-)$ . Opět budeme předpokládat, že intenzita se dá napsat jako  $\mu(y) = g(y, z(t_y)) \cdot J(y)$ , kde  $z(t)$  jsou (náhodné) kovariáty, jedna z nich je samotný čas  $z(t) = t$ ,  $J(y)$  je indikátor (tentýž jako v části 2.2, náhodný, ale pozorovaný) říkající, zda je proces v daném  $y$  (těsně před  $y$ ) pozorován,  $g(y, z) \geq 0$  je riziková funkce (měřitelná, ohraničená).

Protože se proces  $C(t)$  vlastně pohybuje po schodovité jednorozměrné dráze, dá se celá situace vyjádřit i následujícím způsobem: Uvažujme kumulativní proces jako standardní 1D bodový proces vzhledem k parametru  $v$ , který běží střídavě ve směru  $t$  a ve směru  $y$ . Dalo by se i psát  $v = t \cdot I(t) + y \cdot J(y)$ , kde indikátor  $I(t) = 1$  vždy mezi  $(T_{j-1}, Y_{j-1})$  a  $(T_j, Y_{j-1})$  (při  $T_0 = Y_0 = 0$ ), zatímco indikátor  $J(y) = 1$  mezi  $(T_j, Y_{j-1})$  a  $(T_j, Y_j)$ . Takže takto popsany čítací proces  $N^*(v)$  je adaptován na filtraci  $\mathcal{S}(v) = \mathcal{F}(t) \otimes \sigma(y)$ , jeho intenzita se dá napsat jako

$$h(t, Z) \cdot I(t) + g(y, Z) \cdot J(y),$$

kde  $Z$  jsou procesy kovariát a dalších vstupů, které mohou být obecně funkcí  $t$  i  $y$ , adaptované na  $\mathcal{S}(v)$  a predikovatelné. Z toho plyne i standardní tvar věrohodnostní funkce a odhadů. Konkrétním tvarem odhadů má však smysl se zabývat až pro konkrétní specifikace modelu, například pro případ studovaný v příštím odstavci.

**Poznámka k predikci:** V takto pojatém případě je změna měřítka zpravidla různá (a náhodná) pro každý jednotlivý proces, není již možné odvodit predikční hranici transformací z homogenního případu. Východiskem pro získání takové hranice může být například simulace (vyzkoušená také již v Volf, Robust 2000).

**Poznámka k modelu:** Zdá se možná jednodušší modelovat rozdělení přírůstků, v různých  $t$  bychom pak pozorovali jen jednu náhodnou veličinu resp. čítací proces do jeho první události. Model intenzity by byl zcela obdobný.

**3.1. Případ bez regrese, neparametrická inference.** Pro specifikaci závislosti rizikových funkcí na kovariátách máme k dispozici řadu modelů (nejpopulárnější je Coxův), nyní ale ještě setrvejme u situace bez regrese. Uvažujme tedy rizikové funkce  $h(t)$ ,  $g(y, t)$  – tj. předpokládáme alespoň závislost přírůstku složky  $y$  na čase, kdy k němu dochází. Dále uvažujme  $n$  procesů  $C_i(t)$ , pozorovaných v  $t \in [0, T] \times [0, Y]$ , tj.  $I_i(t) = 1$  pro  $t \in [0, T]$ ,  $J_i(y) = 1$  pro  $y \in [0, \min(Y, C(T))]$ . Logaritmus věrohodnostní funkce lze napsat jako:

$$\begin{aligned} \mathcal{L} = & \sum_{i=1}^n \left\{ \int_0^T \log h(t) dN_i(t) + \int_0^Y \log g(y, t) dS_i(y) - \right. \\ & \left. - \int_0^T h(t) I_i(t) dt - \int_0^Y g(y, t) J_i(y) dy \right\}. \end{aligned}$$

Odhad pro  $h(t)$ , resp. pro  $H(t) = \int_0^t h(s) ds$  je standardní, viz kapitola 2. Pozorujeme čítací procesy  $S_i(y)$ ,  $i = 1, \dots, n$ , které však pro každé  $i$  mají přírůstky v jiných časech.

Chceme odhadnout funkci  $g(y, t)$ , do které nám vlastně  $t$  vstupuje jako na  $y$  závislá kovariáta  $t_i(y)$ , pro  $i$ -tý proces. Tj. pro každé pevné  $(y, t)$  máme nanejvýš 1 bod v datech, ‘triviálním’ přímým odhadem by byla 1 v oněch bodech, 0 jinde – tj. realizace našich  $n$  2D čítacích měř.

**3.2. Odhad kumulované rizikové funkce přírůstků.** Budeme funkci  $g(y, t)$  odhadovat jako 2-rozměrnou funkci, tj. jako intenzitu (nejlépe v kumulativní verzi) vzhledem k  $y$ , a s parametrem  $t$ . Nejjednodušší postup je tento: Zvolme disjunktní dělení intervalu  $[0, T]$  na podintervaly  $\mathcal{T}_r$ ,  $r = 1, 2, \dots, m_n$  (stejně délky  $T/m_n$ ) a definujme čítací značené (marked) procesy  $S_i(y, r)$  registrující ty skoky, při nichž  $t_i(y)$  je v  $\mathcal{T}_r$ . Označme  $S^{(n)}(y, r) = \sum_{i=1}^n S_i(y, r)$ , dále  $J^{(n)}(y, r) = \sum_{i=1}^n 1[t_i(y) \in \mathcal{T}_r] \cdot J_i(y)$  počet individuálních procesů, které přispívají onomu procesu  $S^{(n)}(y, r)$ . Označme pro dané  $t$  ještě  $r_t$  takový index, že  $t \in \mathcal{T}_{r_t}$ . Takže pro každé  $r$  uděláme Nelson–Aalenův odhad

$$\hat{G}(y, r) = \int_0^y \frac{dS^{(n)}(x, r)}{J^{(n)}(x, r)}, \tag{1}$$

coby odhad funkce  $G(y, t) = \int_0^y g(x, t) dx$  pro každé  $t \in \mathcal{T}_r$ . Takovýto odhad (na základě pevných disjunktních oken v  $[0, T]$ ) můžeme samozřejmě vylepšit užitím posuvného okna a jádrového vyrovnání vzhledem k  $t$ , se stejnými limitními vlastnostmi. Pokud jde o teorii limitních vlastností, odkážeme zde na výsledek, který odvodili McKeague a Utikal (1990) pro případ kdy na místě  $t$  vystupoval nějaký obecný pozorovaný proces kovariát. Ti také ukázali, že má smysl odhadovat i dvojítě kumulativní intenzitu  $\mathcal{G}(y, t) = \int_0^y \int_0^t g(x, s) dx ds$ . Připomeneme zde metodu, při které jako odhad  $\mathcal{G}(y, t)$  navrhli

$$\hat{\mathcal{G}}(y, t) = \int_0^t \hat{G}(y, r_s) ds,$$

což je vlastně součet přes jednotlivé intervaly  $\mathcal{T}_r$  do  $t$ . V každém případě, tento odhad je dobře definován. Je nasnadě, že nyní, pro limitní výsledky, budeme předpokládat, že  $n \rightarrow \infty$ , také  $m_n \rightarrow \infty$  a  $n/m_n \rightarrow \infty$ . Dále ještě předpokládejme, že počet procesů, z kterých se odhady počítají, roste stejnoměrně, tj. že

$$\frac{J^{(n)}(y, r) m_n}{n}$$

konverguje v pravděpodobnosti k nějaké (měřitelné) funkci  $a(y, t)$ , stejnoměrně na  $[0, Y] \times [0, T]$ , která je kladná, odražená od nuly. McKeague a Utikal (1990) pak dokazují následující tvrzení:

- (1)  $\hat{G}(y, r_t) \xrightarrow{P} G(y, t)$ , pro každé  $t$ , stejnoměrně v  $y \leq Y$ .
- (2) Pokud ještě navíc  $n/m_n^2 \rightarrow 0$ , tak pro každé  $t$  platí  $\sqrt{n/m_n}(\hat{G}(y, r_t) - G(y, t)) \xrightarrow{\mathcal{D}} \mathcal{U}(y, t)$ , tj. slabá (v Skorochodově  $D[0, Y]$ ) konvergence, kde  $\mathcal{U}(y, t)$  je vzhledem k  $y$  Gaussovský martingal s varianční funkcí

$$u(y, t) = \int_0^y \frac{g(x, t)}{a(x, t)} dx.$$

Jinými slovy, limitní proces lze psát jako

$$U(y, t) = \int_0^y \sqrt{\frac{g(x, t)}{a(x, t)}} dW(x),$$

kde  $W(x)$  je standardní Wienerův proces.

Dále:

- (3) Odhad  $\hat{\mathcal{G}}(y, t)$  konverguje k  $\mathcal{G}(y, t)$  v pravděpodobnosti, stejnoměrně v  $[0, Y] \times [0, T]$ .

(4)  $\sqrt{n}(\hat{G}(y, t) - \mathcal{G}(y, t)) \xrightarrow{D} U^{(2)}(y, t)$ , kde

$$U^{(2)}(y, t) = \int_0^y \int_0^t \sqrt{\frac{g(x, s)}{a(x, s)}} dW(x, s),$$

$W(x, s)$  je standardní Wienerovo pole.

McKeague a Utikal (1990) dále ukazují, že jádrovým vyrovnáním (vzhledem k  $y$  i  $t$ ) lze obdržet i odhady funkce  $g(y, t)$ , pokud je Lipschitzovsky spojitá. Konkrétně, odhad

$$\hat{g}(y, t) = \frac{1}{d_1} \int_0^y \mathcal{K}_1 \left( \frac{y-x}{d_1} \right) d\hat{G}(y, r_t)$$

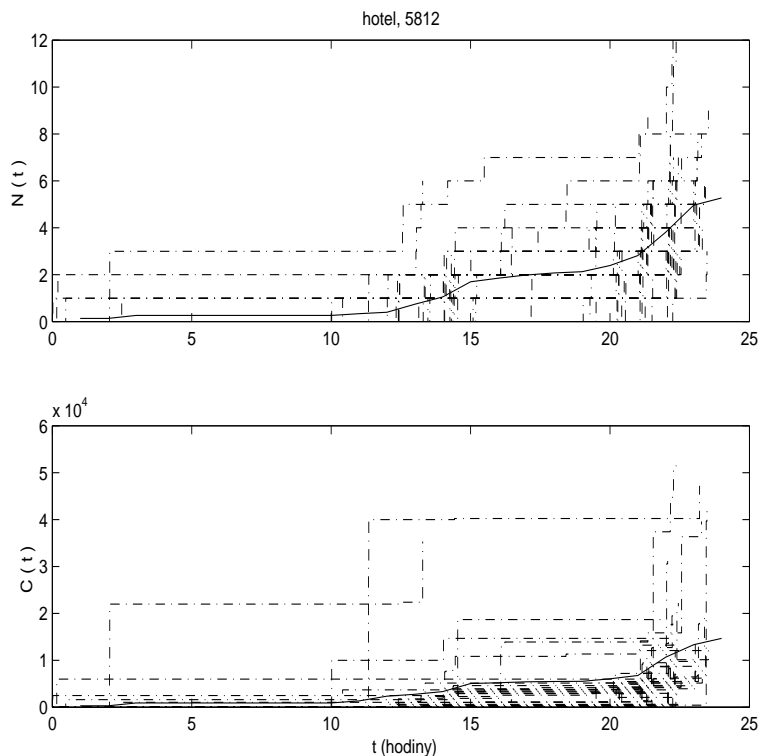
je stejnoměrně konzistentní v  $y$ , pro každé  $t$ , a dále

$$\tilde{g}(y, t) = \frac{1}{d_2} \int_0^t \mathcal{K}_2 \left( \frac{t-s}{d_2} \right) \hat{g}(y, s) ds$$

je stejnoměrně konzistentní i vzhledem k  $t$  v  $[0, T]$ .

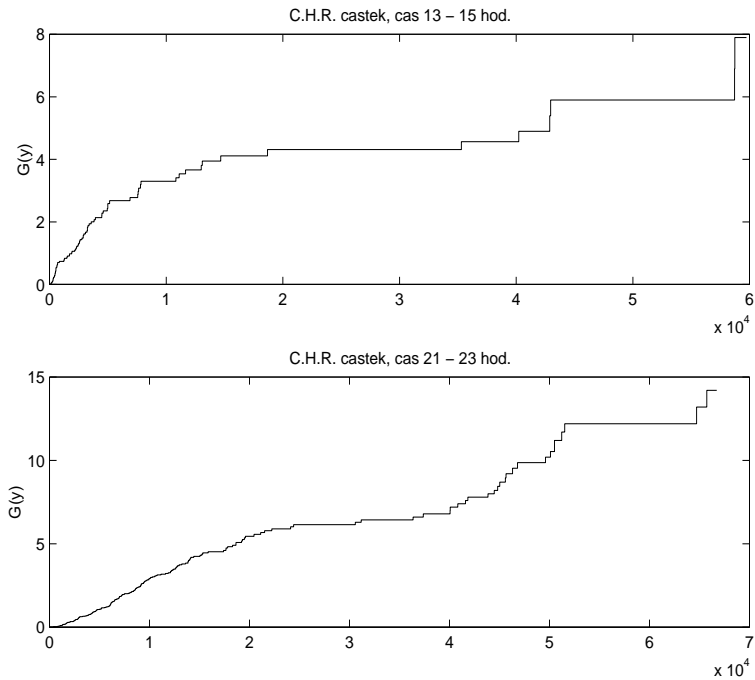
#### 4. PŘÍKLAD

Jako data uvažujeme procesy plateb kreditními kartami v určitém restauračním zařízení. Jeden proces odpovídá platbám za jeden den, data obsahují údaje za 97 dní. Ukázka několika procesů je na obr. 3, spolu s průměry, které vlastně představují Nelson–Aalenovy odhady kumulativních intenzit (srovnej Volf, 2001). Na obr. 4 jsou pak odhady kumulativních intenzit  $G(y, t)$ , a to odhadnuté v „hrubém“ časovém rozpětí dvou hodin, jednak mezi 13. a 15. hodinou a mezi 21. a 23. hodinou. Pro srovnání obr. 5 ukazuje histogram hodnot přírůstků – plateb – v oněch dvou intervalech.

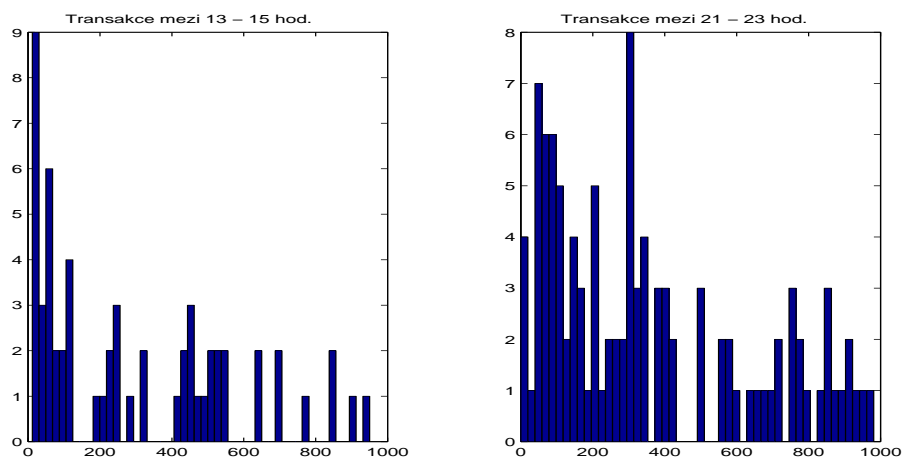


**Obr. 3** Ukázka dat, čítacího a kumulativního procesu karetých transakcí.





Obr. 4 Odhady kumulativních rizikových funkcí  $G(y)$  pro různé časové intervaly.



Obr. 5 Histogramy hodnot transakcí v různých časových intervalech.

#### LITERATURA

- [1] Andersen P.K., Borgan O., Gill R.D., and Keiding N. (1993): Statistical Models Based on Counting Processes. Springer, New York.
- [2] Asmussen S. (2000): Ruin Probabilities. World Scientific, Singapore.
- [3] Embrechts P., Klüppelberg K., and Mikosch T. (1997): Modelling Extremal Events. Springer, Berlin.
- [4] Greenwood P. E., Wefelmeyer W. (1994): Optimality properties of empirical estimators for multivariate point processes. Journal of Multiv. Analysis 19, 202-217.

- [5] ougaard P. (2000): Analysis of Multivariate Survival Data. Springer, New York.
- [6] acod J., Shirjajev A. N. (1987): Limit Theorems for Stochastic Processes. Springer, Berlin.
- [7] cKeague I. W. and Utikal K. J. (1990): Identifying nonlinear covariate effects in semimartingale regression models. Probability Theory and Related Fields 87, 1–25.
- [8] olski T., Schmidli H., Schmidt V., and Teugels J. (1999): Stochastic Processes for Insurance and Finance. Wiley.
- [9] olf P. (1992): Regresní modely v analýze přežívání. Sborník Robust 1992, JČMF Praha, 220–235 .
- [10] olf P. (2001): Statistical models and analysis of cumulated damage processes. Sborník Robust 2000, JČMF Praha, 357–367.

INSTITUTE OF INFORMATION THEORY AND AUTOMATION, ACADEMY OF SCIENCES OF THE CZECH  
REPUBLIC, POD VODÁRENSKOU VĚŽÍ 4, 182 08 PRAGUE 8, CZECH REPUBLIC  
E-MAIL: [volf@utia.cas.cz](mailto:volf@utia.cas.cz)