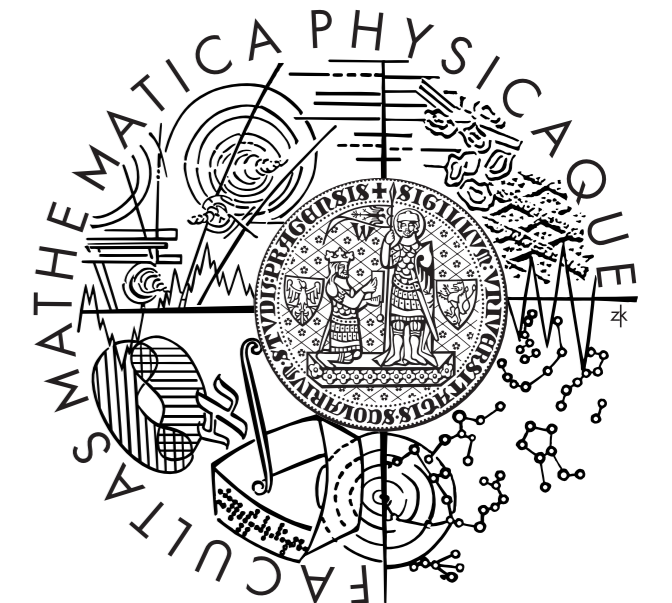




Časoprostorové Coxovy bodové procesy s Lévyho bází

Jiří Dvořák, Michaela Prokešová

dvorak@karlin.mff.cuni.cz, prokesov@karlin.mff.cuni.cz
Katedra pravděpodobnosti a matematické statistiky, MFF UK, Praha



ABSTRAKT

Coxův bodový proces, neboli dvojně stochastický proces, nabízí flexibilní rámec pro modelování shlukových bodových procesů. Bude představen model nestacionárního časoprostorového Coxova procesu využívající jádrové zhlazení Lévyho báze jako řídicí funkce intenzity a budou diskutovány možnosti odhadu parametrů modelu v případě časoprostorové separability použitého vyhlazovacího jádra.

Přestože tento proces není separabilní, umožňuje splnění této podmínky založit odhady parametrů na marginálním časovém a prostorovém procesu, tedy projekci procesu do časové, resp. prostorové domény.

COXOVY PROCESY S LÉVYHO BÁZÍ

Umožňují modelovat shlukové procesy s variabilní vahou shluků (středním počtem bodů ve shluku).

Λ nezáporná náhodná míra na \mathbb{R}^d s hustotou $\lambda(u)$ vzhledem k Lebesgueově míře

Coxův bodový proces – tzv. dvojně stochastický proces – podmíněně při daném $\Lambda = \Lambda_0$ má rozdělení Poissonova procesu s mírou intenzity Λ_0 .

Coxův proces s Lévyho bází (LCP) – náhodná míra Λ má hustotu

$$\lambda(u) = \int_{\mathbb{R}^d} k(u, v) L(dv), \quad u \in \mathbb{R}^d,$$

kde L je nezáporná Lévyho báze na \mathbb{R}^d (nekonečně dělitelná, nezávisle rozptýlená náhodná míra), k je (správně integrovatelné) vyhlazovací jádro. Předpokládáme, že pro fixovanou jednu souřadnici se přes druhou vyintegruje na 1. Pro detaily viz [3].

Shot-noise reprezentace LCP

Pro bázi L bez gaussovské části je k dispozici shot-noise reprezentace LCP ve tvaru

$$\lambda(u) = \sum_{(v,r) \in \Phi} r k(u, v),$$

kde Φ je poissonovská míra na $\mathbb{R}^d \times (0, \infty)$ s mírou intenzity $U(dv, dr)$, viz [3].

Stacionární LCP:

- $k(u, v) = k(v - u)$,
- faktorizace $U(dv, dr) = \mu V(dr) dv$.

Parametr μ popisuje počet shluků, míra V určuje typ báze a rozdělení vah shluků.

Příklad: gama báze s $V(dr) = I(r > 0) r^{-1} e^{-\theta r} dr, \theta > 0$.

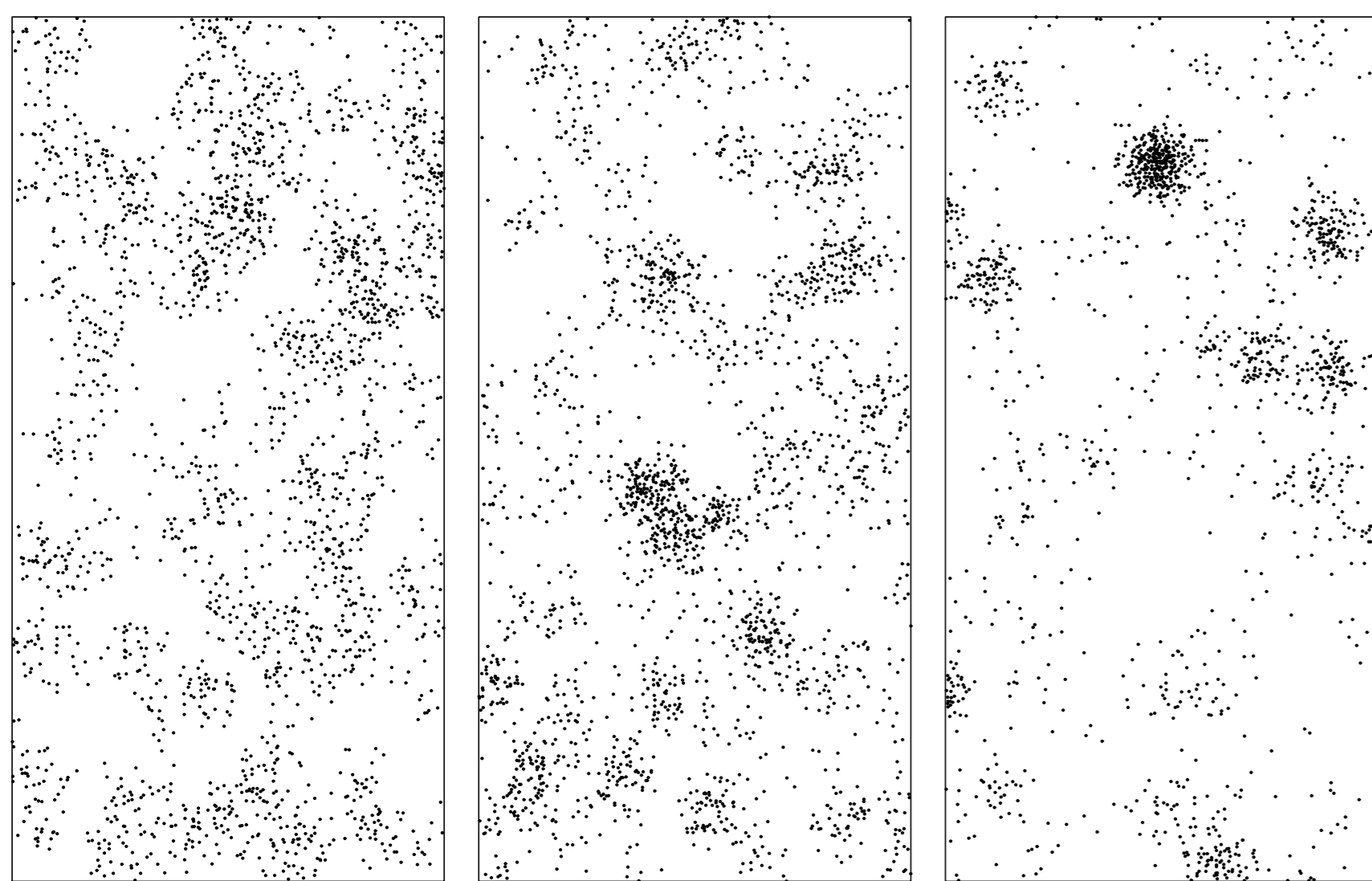


Figure 1: Ukázka realizací stacionárních LCP s různým typem báze (zleva Poissonova, gama a inverzní gaussovská báze).

STACIONÁRNÍ MODEL – PROSTOROVÝ

Momentové charakteristiky stacionárního LCP X dostupné v uzavřeném tvaru, pro přehlednost dále uvažujeme stacionární LCP s gama bází:

$$\rho = \frac{\mu}{\theta}, \quad g(u, v) = 1 + \frac{1}{\mu} I_k(v - u), \quad K(r) = \int_{B(0,r)} g(u) du,$$

- ρ – intenzita procesu, počet bodů procesu na jednotku objemu,
- g – párová korelační funkce, $g(u, v) du dv$ odpovídá pravděpodobnosti pozorování dvojice bodů procesu v místech u a v ,
- $I_k(v - u) = \int_{\mathbb{R}^d} k(v - z) k(u - z) dz$,
- K – tzv. K -funkce, $K(r)$ odpovídá střednímu počtu bodů ve vzdálenosti menší než r od typického bodu procesu, normovanému intenzitou.

Odhad parametrů

X – uvažovaný proces,
 W – pozorovací okno

Odhad parametru gama báze μ a (vektorového) parametru ω jádra k metodou minimálního kontrastu – minimalizace míry diskrepance $D(\mu, \omega)$ vzhledem k μ, ω :

$$D(\mu, \omega) = \int_{R_1}^{R_2} [\widehat{K}^q(r) - K^q(r; \mu, \omega)]^2 dr,$$

- $K(r; \mu, \omega)$ – teoretický tvar K -funkce,
- $\widehat{K}(r)$ – neparametrický odhad empirické K -funkce,
- q – exponent stabilizující rozptyl, typicky $q = 1/2$ nebo $1/4$,
- R_1, R_2 – vhodně zvolené parametry.

Odhad parametru θ pak ze znalosti $\widehat{\mu}$ (viz výše) a přímočarého odhadu $\widehat{\rho} = |X \cap W|/|W|$.

NESTACIONÁRNÍ MODEL – PROSTOROVÝ

Nehomogenita vnesena do modelu pomocí ztenčení závislého na poloze – funkce nehomogenity $f(u)$ splňující $\max f = 1$, každý bod stacionárního procesu $x \in X$ ponechán s pstí $f(x)$, jinak odstraněn. Zbylé body – nehomogenní proces X_f .

Ztenčení se projevuje v momentových charakteristikách 1. řádu, 2. řádu ne $\Rightarrow X_f$ má vlastnost SOIRS (převážena stacionarita 2. řádu, viz [1]) \Rightarrow možnost definovat párovou korelační funkci a K -funkci i v nehomogenním případě.

Opět model s gama bází:

$$\rho(u) = \frac{\mu}{\theta} f(u), \quad g(u, v) = 1 + \frac{1}{\mu} I_k(v - u), \quad K(r) = \int_{B(0,r)} g(u) du.$$

Dvoukrokový odhad parametrů

X_f – ztenčený proces,
 W – pozorovací okno,
 $f(u)$ – funkce nehomogenity splňující $\max_W f = 1$

Postup odhadu:

1. odhad $\rho(u)$ jádrově nebo parametricky (viz [5]),
2. odhad μ a ω minimálním kontrastem jako ve stacionárním případě.

$\widehat{\rho}(u)$ nutné k neparametrickému odhadu K -funkce \Rightarrow vyšší variabilita odhadu $\widehat{K} \Rightarrow$ vyšší variabilita odhadu parametrů než ve stacionárním případě

Odhad parametru θ pak ze znalosti $\widehat{\mu}$, skutečnosti, že $\max_W f = 1$, a rovnice

$$\int_W \widehat{\rho}(u) du = |X_f \cap W|.$$

ČASOPROSTOROVÝ MODEL

Časoprostorový LCP X v $\mathbb{R}^d \times \mathbb{R}$. Shot-noise reprezentace:

$$\lambda(u, t) = \sum_{(r,v,s) \in \Phi} r k((u, t), (v, s)), \quad (u, t) \in \mathbb{R}^d \times \mathbb{R},$$

Φ poissonovská míra na $\mathbb{R}^+ \times \mathbb{R}^d \times \mathbb{R}$ s mírou intenzity U, k vyhlazovací jádro.

Stacionární časoprostorový LCP X :

- $k((u, t), (v, s)) = k(v - u, s - t)$,
- faktorizace $U(dr, d(u, t)) = \mu V(dr) d(u, t), \quad \mu > 0$.

Charakteristiky stacionárního časoprostorového LCP s gama bází:

$$\rho(u, t) = \frac{\mu}{\theta}, \quad g((u, t), (v, s)) = 1 + \frac{1}{\mu} I_k(v - u, s - t),$$

kde $I_k(v - u, s - t) = \int_{\mathbb{R}^d} \int_{\mathbb{R}} k(v - z, s - \tau) k(u - z, t - \tau) d\tau dz$.

Ztenčený proces X_f s funkcí nehomogenity $f(u, t)$:

$$\rho(u, t) = \frac{\mu}{\theta} f(u, t), \quad g((u, t), (v, s)) = 1 + \frac{1}{\mu} I_k(v - u, s - t).$$

Nehomogenní verze K -funkce definována jako

$$K(r, t) = \int_{B(0,r)} \int_{-t}^t g(u, s) ds du.$$

Dvoukrokový odhad parametrů

Analogicky jako v prostorovém případě. V neparametrickém odhadu $K(r, t)$ používány odhady časoprostorové funkce intenzity $\rho(u, t)$ – typicky odhadována z malého množství dat, nestabilní, vede k nestabilním odhadům parametrů modelu.

PROJEKTOVANÉ PROCESY

Při splnění určitých předpokladů je možné založit odhady na projekci procesu do prostorové a časové domény (inspirováno článkem [4]).

Pro pevné prostorové pozorovací okno $W, |W| > 0$, a časový interval $[0, T], T > 0$, definujeme projektované procesy

$$X_{space} = \{u : (u, t) \in X_f, t \in [0, T]\}, \\ X_{time} = \{t : (u, t) \in X_f, u \in W\}.$$

Dodatečné předpoklady – separabilita:

- $f(u, t) = f_1(u) f_2(t), u \in W, t \in [0, T]$,
- $k((u, t), (v, s)) = k_1(v - u) k_2(s - t), u \in W, t \in [0, T], v \in \mathbb{R}^d, s \in \mathbb{R}$.

Pak $I_k(v - u, s - t) = I_{k_1}(v - u) I_{k_2}(s - t)$, kde

$$I_{k_1}(v - u) = \int_{\mathbb{R}^d} k_1(v - z) k_1(u - z) dz \\ I_{k_2}(s - t) = \int_{\mathbb{R}} k_2(s - \tau) k_2(t - \tau) d\tau.$$

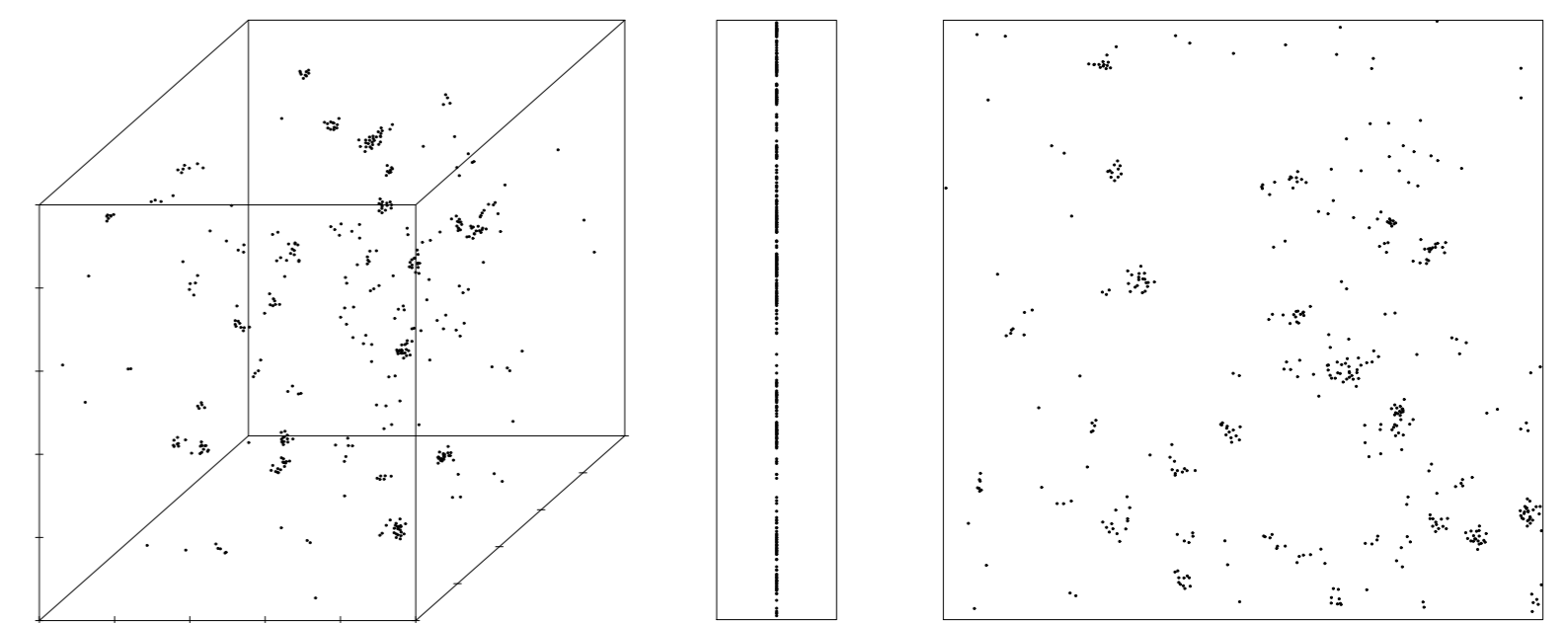


Figure 2: Ukázka realizace nestacionárního časoprostorového LCP s gama bází (vlevo) a jeho časová a prostorová projekce (uprostřed a vpravo).

Prostorová projekce

$$\rho_{space}(u) = \frac{\mu}{\theta} f_1(u) \int_0^T f_2(s) ds, \\ g_{space}(u, v) = g_{space}(v - u) = 1 + C_{space} \frac{I_{k_1}(v - u)}{\mu}, \\ K_{space}(r) = \pi r^2 + \frac{C_{space}}{\mu} \int_{\|u\| < r} I_{k_1}(u) du,$$

$$C_{space} = \frac{1}{(\int_0^T f_2(\tau) d\tau)^2} \int_0^T \int_0^T \int_{\mathbb{R}^d} f_2(s) f_2(t) k_2(s - \tau) k_2(t - \tau) d\tau ds dt.$$

Časová projekce

$$\rho_{time}(t) = \frac{\mu}{\theta} f_2(t) \int_W f_1(z) dz, \\ g_{time}(s, t) = g_{time}(t - s) = 1 + C_{time} \frac{I_{k_2}(t - s)}{\mu}, \\ K_{time}(t) = 2t + \frac{C_{time}}{\mu} \int_{-t}^t I_{k_2}(s) ds,$$

$$C_{time} = \frac{1}{(\int_W f_1(z) dz)^2} \int_W \int_W \int_{\mathbb{R}^d} f_1(u) f_1(v) k_1(u - z) k_1(v - z) dz du dv.$$

Odhady parametrů

Jako výše dvoukrokově, zvlášť pro X_{space} (odhady f_1 a parametrů k_1) a X_{time} (odhady f_2 a parametrů k_2).

Pro vhodně zvolené modely jsou konstanty C_{space}, C_{time} k dispozici v uzavřeném tvaru \Rightarrow odhady μ, θ .

Výhoda: stabilnější odhady $\rho_{space}, \rho_{time}$ než časoprostorové $\rho \Rightarrow$ stabilnější odhady $K_{space}, K_{time} \Rightarrow$ stabilnější odhady parametrů.

Simulační studie

Pro zjištění přesnosti odhadů v konkrétním případě byla provedena simulační studie, viz článek [2]. Klíčová zjištění:

- odhady parametrů jader k_1, k_2 – rozumné,
- odhady parametrů báze μ, θ – těžké, stabilnější s dopočítáním pomocí C_{space} než s C_{time} ,
- odhady z X_{time} – hrozí problémy s překrýváním shluků.

Poděkování. Práce na tomto projektu byla podpořena granty GAČR P201/10/0472 a SVV 265 315.

Literatura.

- [1] Baddeley A.J., Møller J. a Waagepetersen R. (2000). *Non- and semiparametric estimation of interaction in inhomogeneous point patterns*. Stat. Neerl. **54**, 329–50.
- [2] Dvořák J. a Prokešová M. (2012). *Statistics for inhomogeneous space-time shot-noise Cox processes*. Podáno.
- [3] Hellmund G., Prokešová M. a Vedel Jensen E.B. (2008). *Lévy-based Cox point processes*. Adv. Appl. Probab. **40**, 603–629.
- [4] Møller J. a Ghorbani M. (2012). *Aspects of second-order analysis of structured inhomogeneous spatio-temporal point processes*. Stat. Neerlandica, přijato.
- [5] Waagepetersen R.P. a Guan Y. (2009). *Two-Step Estimation for Inhomogeneous Spatial Point Processes*. J. R. Stat. Soc. B **71**, 685–702.