

# Časoprostorové Coxovy bodové procesy s Lévyho bází

Jiří Dvořák<sup>1,2</sup>, Michaela Prokešová<sup>1</sup>

<sup>1</sup>Katedra pravděpodobnosti a matematické statistiky, MFF UK, Praha,

<sup>2</sup>Ústav teorie informace a automatizace AV ČR, v.v.i., Praha,

ROBUST 2012  
Němčičky, 11. 9. 2012

Shot-noise reprezentace:

$$\Lambda(u, t) = \sum_{(r, v, s) \in \Phi} r k((u, t), (v, s)), \quad (u, t) \in \mathbb{R}^2 \times \mathbb{R},$$

$\Phi$  je poissonovská míra na  $\mathbb{R}^+ \times \mathbb{R}^2 \times \mathbb{R}$  s mírou intenzity  $U$ ,  
 $k$  vyhlazovací jádro.

Shot-noise reprezentace:

$$\Lambda(u, t) = \sum_{(r, v, s) \in \Phi} r k((u, t), (v, s)), \quad (u, t) \in \mathbb{R}^2 \times \mathbb{R},$$

$\Phi$  je poissonovská míra na  $\mathbb{R}^+ \times \mathbb{R}^2 \times \mathbb{R}$  s mírou intenzity  $U$ ,  $k$  vyhlazovací jádro.

Stacionarita:

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

Shot-noise reprezentace:

$$\Lambda(u, t) = \sum_{(r, v, s) \in \Phi} r k((u, t), (v, s)), \quad (u, t) \in \mathbb{R}^2 \times \mathbb{R},$$

$\Phi$  je poissonovská míra na  $\mathbb{R}^+ \times \mathbb{R}^2 \times \mathbb{R}$  s mírou intenzity  $U$ ,  $k$  vyhlazovací jádro.

Stacionarita:

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

Pro názornost volíme konkrétní bázi (gama):

$$V(dr) = I(r > 0) r^{-1} e^{-\theta r} dr, \quad \theta > 0.$$

Stacionární proces:

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

$$I_k(v - u, s - t) = \int_{\mathbb{R}^2} \int_{\mathbb{R}} k(v - z, s - \tau) k(u - z, t - \tau) d\tau dz.$$

Stacionární proces:

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

$$l_k(v - u, s - t) = \int_{\mathbb{R}^2} \int_{\mathbb{R}} k(v - z, s - \tau) k(u - z, t - \tau) d\tau dz.$$

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

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

Možnost definovat nehomogenní  $K$ -funkci  $K(r, t)$ .

# Dvukrokový odhad parametrů modelu

- $X$  ... realizace z uvažovaného modelu,
- $W \times [0, T]$  ... pozorovací okno,  $|W| > 0$ ,  $T > 0$ ,
- $k$  ... pravděpodobnostní jádro,
- $\max_{W \times [0, T]} f = 1$ .

# Dvoukrokový odhad parametrů modelu

- $X$  . . . realizace z uvažovaného modelu,
  - $W \times [0, T]$  . . . pozorovací okno,  $|W| > 0$ ,  $T > 0$ ,
  - $k$  . . . pravděpodobnostní jádro,
  - $\max_{W \times [0, T]} f = 1$ .
1. odhad funkce intenzity  $\frac{\mu}{\theta} f(u, t)$  (parametrický vs. jádrový), tedy odhad  $f$  a  $\frac{\mu}{\theta}$  s využitím

$$|X \cap (W \times [0, T])| = \frac{\mu}{\theta} \int_W \int_0^T \hat{f}(u, t) dt du,$$



# Dvukrokový odhad parametrů modelu

- $X$  ... realizace z uvažovaného modelu,
  - $W \times [0, T]$  ... pozorovací okno,  $|W| > 0$ ,  $T > 0$ ,
  - $k$  ... pravděpodobnostní jádro,
  - $\max_{W \times [0, T]} f = 1$ .
1. odhad funkce intenzity  $\frac{\mu}{\theta} f(u, t)$  (parametrický vs. jádrový), tedy odhad  $f$  a  $\frac{\mu}{\theta}$  s využitím

$$|X \cap (W \times [0, T])| = \frac{\mu}{\theta} \int_W \int_0^T \hat{f}(u, t) dt du,$$

2. odhad párové korelační funkce nebo  $K$ -funkce, pak minimální kontrast, tedy odhad jádra  $k$  a  $\frac{1}{\mu}$ .

Metoda minimálního kontrastu, prostorový případ:

$$\int_{r_{min}}^{r_{max}} (\widehat{K}(r)^q - K(r; \omega)^q)^2 dr$$

Časoprostorový případ:

$$\int_{r_{min}}^{r_{max}} \int_{t_{min}}^{t_{max}} (\widehat{K}(r, t)^q - K(r, t; \omega)^q)^2 dt dr$$

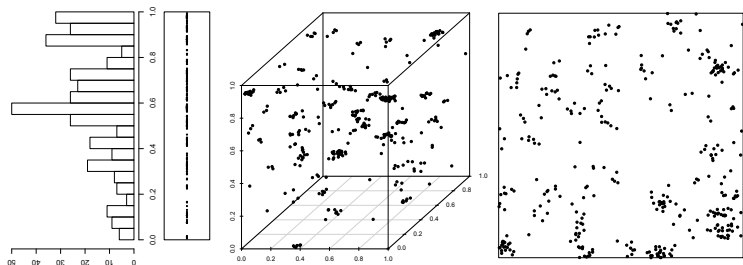
$q \dots$  exponent stabilizující rozptyl, typicky  $q = 1/2$  nebo  $1/4$

# Projekce do prostorové a časové domény

Inspirováno článkem Møller & Ghorbani (2012).

Prostorová projekce:  $X_{space} = \{u : (u, t) \in X, t \in [0, T]\}$

Časová projekce:  $X_{time} = \{t : (u, t) \in X, u \in W\}$



- $f(u, t) = f_1(u)f_2(t), u \in W, t \in [0, T],$
- $k((u, t), (v, s)) = k_1(v - u)k_2(s - t), u \in W, t \in [0, T],$   
 $v \in \mathbb{R}^2, s \in \mathbb{R}.$

- $f(u, t) = f_1(u)f_2(t)$ ,  $u \in W$ ,  $t \in [0, T]$ ,
- $k((u, t), (v, s)) = k_1(v - u)k_2(s - t)$ ,  $u \in W$ ,  $t \in [0, T]$ ,  
 $v \in \mathbb{R}^2$ ,  $s \in \mathbb{R}$ .

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

$$I_{k_1}(v - u) = \int_{\mathbb{R}^2} k_1(v - z)k_1(u - z) dz$$

$$I_{k_2}(s - t) = \int_{\mathbb{R}} k_2(s - \tau)k_2(t - \tau) d\tau.$$

$$X_{space} = \{u : (u, t) \in X, t \in [0, T]\}$$

$$\rho_{space}(u) = \frac{\mu}{\theta} f_1(u) \int_0^T f_2(s) ds,$$

$$g_{space}(u, v) = g_{space}(v - u) = 1 + C_{space} \frac{I_{k1}(v - u)}{\mu},$$

$$K_{space}(r) = \pi r^2 + \frac{C_{space}}{\mu} \int_{\|u\| < r} I_{k1}(u) du,$$

kde

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

$$X_{time} = \{t : (u, t) \in X, u \in W\}$$

$$\rho_{time}(t) = \frac{\mu}{\theta} f_2(t) \int_W f_1(z) dz,$$

$$g_{time}(s, t) = g_{time}(t - s) = 1 + C_{time} \frac{l_{k2}(t - s)}{\mu},$$

$$K_{time}(t) = 2t + \frac{C_{time}}{\mu} \int_{-t}^t l_{k2}(s) ds,$$

kde

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

- $$\hat{K}_{space}(r) = \frac{1}{|W|} \sum_{i \neq j} \frac{I(\|u_i - u_j\| \leq r)}{w_1(u_i, u_j) \hat{\rho}_{space}(u_i) \hat{\rho}_{space}(u_j)}$$



## Druhý krok – konkrétně

$$\bullet \hat{K}_{space}(r) = \frac{1}{|W|} \sum_{i \neq j} \frac{I(\|u_i - u_j\| \leq r)}{w_1(u_i, u_j) \hat{\rho}_{space}(u_i) \hat{\rho}_{space}(u_j)}$$

$$K_{space}(r) = \pi r^2 + \frac{C_{space}}{\mu} \int_{\|u\| < r} I_{k_1}(u) du$$

min. kontrast  $\Rightarrow$  odhad  $k_1, \frac{1}{\mu} C_{space}(k_2, f_2)$

- $$\hat{K}_{space}(r) = \frac{1}{|W|} \sum_{i \neq j} \frac{I(\|u_i - u_j\| \leq r)}{w_1(u_i, u_j) \hat{\rho}_{space}(u_i) \hat{\rho}_{space}(u_j)}$$

$$K_{space}(r) = \pi r^2 + \frac{C_{space}}{\mu} \int_{\|u\| < r} I_{k_1}(u) du$$

min. kontrast  $\Rightarrow$  odhad  $k_1, \frac{1}{\mu} C_{space}(k_2, f_2)$

- $$\hat{K}_{time}(t) = \frac{1}{T} \sum_{i \neq j} \frac{I(|t_i - t_j| \leq t)}{w_2(t_i, t_j) \hat{\rho}_{time}(t_i) \hat{\rho}_{time}(t_j)}$$

$$K_{time}(t) = 2t + \frac{C_{time}}{\mu} \int_{-t}^t I_{k_2}(s) ds$$

min. kontrast  $\Rightarrow$  odhad  $k_2, \frac{1}{\mu} C_{time}(k_1, f_1)$

- Model – gama báze,  $\mu, \theta$ ,  $W \times [0, T] = [0, 1]^3$ ,
- $k_1$  – hustota izotropního centrovaného dvourozměrného normálního rozdělení,  $\sigma$ ,
- $k_2$  – hustota rovnoměrného rozdělení na intervalu  $[0, t_0]$ ,  $t_0$ ,
- $f(x, y, t) \propto \exp\{\beta_1 x + \beta_2 y + \beta_3 t\}$ ,  $\beta_1, \beta_2, \beta_3$ ,
- různé kombinace parametrů, střední počet bodů v okně  $W \times [0, T]$  cca 350 – 1430.

- Model – gama báze,  $\mu, \theta$ ,  $W \times [0, T] = [0, 1]^3$ ,
  - $k_1$  – hustota izotropního centrovaného dvourozměrného normálního rozdělení,  $\sigma$ ,
  - $k_2$  – hustota rovnoměrného rozdělení na intervalu  $[0, t_0]$ ,  $t_0$ ,
  - $f(x, y, t) \propto \exp\{\beta_1 x + \beta_2 y + \beta_3 t\}$ ,  $\beta_1, \beta_2, \beta_3$ ,
  - různé kombinace parametrů, střední počet bodů v okně  $W \times [0, T]$  cca 350 – 1430.
- 
- Parametry jader  $\sigma, t_0$  – rozumné ( $\sigma$  z  $g$ -funkce  $g_{space}$ ,  $t_0$  z  $K$ -funkce  $K_{time}$ ),
  - parametry báze  $\mu, \theta$  – těžké, stabilnější s použitím  $C_{space}$ ,
  - odhady z  $X_{time}$  – možné problémy s překrýváním shluků.

Møller & Ghorbani (2012):

$$g_1(u) = \frac{1}{T^2} \int_T \int_T g(u, s - t) ds dt,$$

$$g_2(t) = \frac{1}{|W|^2} \int_W \int_W g(u - v, t) du dv,$$

Møller & Ghorbani (2012):

$$g_1(u) = \frac{1}{T^2} \int_T \int_T g(u, s - t) ds dt,$$

$$g_2(t) = \frac{1}{|W|^2} \int_W \int_W g(u - v, t) du dv,$$

$$K_1(r) = \int_{\|u\| < r} g_1(u) du, \quad K_2(t) = \int_{-t}^t g_2(s) ds$$

Møller & Ghorbani (2012):

$$g_1(u) = \frac{1}{T^2} \int_T \int_T g(u, s - t) ds dt,$$

$$g_2(t) = \frac{1}{|W|^2} \int_W \int_W g(u - v, t) du dv,$$

$$K_1(r) = \int_{\|u\| < r} g_1(u) du, \quad K_2(t) = \int_{-t}^t g_2(s) ds$$

$$\int \int h_1(u, v) g_1(u - v) du dv = \frac{1}{T^2} \mathbb{E} \sum_{(u,s), (v,t) \in X: s, t \in [0, T]}^{\neq} \frac{h_1(u, v)}{\rho(u, s) \rho(v, t)},$$

$h_1$  nezáporná borelovská funkce na  $\mathbb{R}^2 \times \mathbb{R}^2$

$$\hat{K}_1(r) = \frac{1}{|W|T^2} \sum_{i \neq j} \frac{I(\|u_i - u_j\| \leq r)}{w_1(u_i, u_j) \hat{\rho}(u_i, t_i) \hat{\rho}(u_j, t_j)},$$

$$\hat{K}_2(t) = \frac{1}{|W|^2 T} \sum_{i \neq j} \frac{I(|t_i - t_j| \leq t)}{w_2(t_i, t_j) \hat{\rho}(u_i, t_i) \hat{\rho}(u_j, t_j)}$$



$$\hat{K}_1(r) = \frac{1}{|W|T^2} \sum_{i \neq j} \frac{I(\|u_i - u_j\| \leq r)}{w_1(u_i, u_j) \hat{\rho}(u_i, t_i) \hat{\rho}(u_j, t_j)},$$

$$\hat{K}_2(t) = \frac{1}{|W|^2 T} \sum_{i \neq j} \frac{I(|t_i - t_j| \leq t)}{w_2(t_i, t_j) \hat{\rho}(u_i, t_i) \hat{\rho}(u_j, t_j)}$$

$$\hat{K}_{space}(r) = \frac{1}{|W|} \sum_{i \neq j} \frac{I(\|u_i - u_j\| \leq r)}{w_1(u_i, u_j) \hat{\rho}_{space}(u_i) \hat{\rho}_{space}(u_j)},$$

$$\hat{K}_{time}(t) = \frac{1}{T} \sum_{i \neq j} \frac{I(|t_i - t_j| \leq t)}{w_2(t_i, t_j) \hat{\rho}_{time}(t_i) \hat{\rho}_{time}(t_j)}$$

- Jednodušší tvar charakteristik 2. řádu – žádné konstanty  $C_{space}$ ,  $C_{time}$ ,
- komplikovanější odhady charakteristik 2. řádu,
- odhad parametrů modelu jako výše, využití  $K_1$ ,  $K_2$ ,

- Jednodušší tvar charakteristik 2. řádu – žádné konstanty  $C_{space}$ ,  $C_{time}$ ,
- komplikovanější odhady charakteristik 2. řádu,
- odhad parametrů modelu jako výše, využití  $K_1$ ,  $K_2$ ,
  
- simulační studie – stejné realizace:
  - odhad  $\sigma$  pro nízkou intenzitu dobrý, pro vysokou selhává,
  - odhady  $\mu$ ,  $\theta$  těžké, lepší založit na  $K_1$ ,
  - většinou nižší rel. vychýlení odhadů,
  - většinou vyšší rel. MSE, zejména pro  $t_0$ ,

- Jednodušší tvar charakteristik 2. řádu – žádné konstanty  $C_{space}$ ,  $C_{time}$ ,
- komplikovanější odhady charakteristik 2. řádu,
- odhad parametrů modelu jako výše, využití  $K_1$ ,  $K_2$ ,
  
- simulační studie – stejné realizace:
  - odhad  $\sigma$  pro nízkou intenzitu dobrý, pro vysokou selhává,
  - odhady  $\mu$ ,  $\theta$  těžké, lepší založit na  $K_1$ ,
  - většinou nižší rel. vychýlení odhadů,
  - většinou vyšší rel. MSE, zejména pro  $t_0$ ,
  - rozdíly ale malé.

- Jednodušší tvar charakteristik 2. řádu – žádné konstanty  $C_{space}$ ,  $C_{time}$ ,
- komplikovanější odhady charakteristik 2. řádu,
- odhad parametrů modelu jako výše, využití  $K_1$ ,  $K_2$ ,
  
- simulační studie – stejné realizace:
  - odhad  $\sigma$  pro nízkou intenzitu dobrý, pro vysokou selhává,
  - odhady  $\mu$ ,  $\theta$  těžké, lepší založit na  $K_1$ ,
  - většinou nižší rel. vychýlení odhadů,
  - většinou vyšší rel. MSE, zejména pro  $t_0$ ,
  - rozdíly ale malé.
  
- Závěrem: preferujeme naši variantu odhadu.

- Problém identifikace báze – charakteristikami 2. řádu není báze určena, nutno ji zvolit předem; odhad pouze dvou parametrů.

- Problém identifikace báze – charakteristikami 2. řádu není báze určena, nutno ji zvolit předem; odhad pouze dvou parametrů.
- Jiné charakteristiky (3. řádu, . . . ) mohou umožnit odhadnout více – pro získání stabilních odhadů nutné velké množství dat.

- Problém identifikace báze – charakteristikami 2. řádu není báze určena, nutno ji zvolit předem; odhad pouze dvou parametrů.
- Jiné charakteristiky (3. řádu, . . . ) mohou umožnit odhadnout více – pro získání stabilních odhadů nutné velké množství dat.
- Co použít pro odhady nebo ověření modelu?  
Možní kandidáti:
  - $T$ -funkce (Schladitz & Baddeley (2000)),
  - rezidua pro Neyman-Scott procesy (Jalilian & Vahidi-Asl (2011)),
  - další?



- 1 Baddeley A.J., Møller J., Waagepetersen R. (2000). Non- and semiparametric estimation of interaction in inhomogeneous point patterns. *Stat. Neerlandica* 54, 329–350.
- 2 Hellmund G., Prokešová M., Jensen E.B.V. (2008) Lévy-based Cox point processes. *Adv. Appl. Probab.* 40, 603–629.
- 3 Jalilian A.H., Vahidi-Asl, M.Q. (2011) Residual analysis for inhomogeneous Neyman-Scott processes. *Scand. J. Statist.* 38, 617–630.
- 4 Møller J. (2003) Shot noise Cox processes. *Adv. Appl. Probab.* 35, 614–640.
- 5 Møller J., Ghorbani M. (2012) Aspects of second-order analysis of structured inhomogeneous spatio-temporal point processes. *Stat. Neerlandica*, přijato.
- 6 Schladitz K., Baddeley A.J. (2000) A third-order point process characteristic. *Scand. J. Statist.* 27, 657–671.

# Numerické výsledky – naše metoda

$\mu$	true values			rel. bias				rel. MSE			
	$\theta$	$t_0$	$\sigma$	$\hat{\mu}$	$\hat{\theta}$	$\hat{t}_0$	$\hat{\sigma}$	$\hat{\mu}$	$\hat{\theta}$	$\hat{t}_0$	$\hat{\sigma}$
50	1/20	0.015	0.01	0.169	0.104	-0.028	0.099	0.096	0.247	0.022	0.089
			0.02	0.181	0.192	-0.033	-0.047	0.116	0.326	0.022	0.010
	0.030	0.01	0.180	0.110	-0.047	0.094	0.103	0.285	0.036	0.080	
		0.02	0.155	0.215	-0.062	-0.042	0.100	0.365	0.042	0.011	
50	1/40	0.015	0.01	0.150	0.119	-0.032	0.127	0.080	0.265	0.012	0.094
			0.02	0.144	0.170	-0.031	-0.047	0.087	0.275	0.014	0.007
	0.030	0.01	0.131	0.152	-0.062	0.103	0.065	0.281	0.028	0.079	
		0.02	0.154	0.159	-0.055	-0.042	0.094	0.281	0.025	0.007	
100	1/20	0.015	0.01	0.089	0.107	-0.028	0.016	0.041	0.151	0.019	0.039
			0.02	0.121	0.154	-0.035	-0.041	0.071	0.210	0.018	0.013
	0.030	0.01	0.065	0.089	-0.063	0.023	0.038	0.147	0.038	0.042	
		0.02	0.138	0.205	-0.067	-0.052	0.078	0.240	0.038	0.013	
100	1/40	0.015	0.01	0.081	0.087	-0.031	0.017	0.032	0.144	0.013	0.032
			0.02	0.089	0.117	-0.024	-0.031	0.057	0.164	0.012	0.010
	0.030	0.01	0.070	0.105	-0.046	0.036	0.029	0.128	0.029	0.040	
		0.02	0.124	0.126	-0.054	-0.041	0.060	0.154	0.027	0.010	

# Numerické výsledky – alternativní metoda

$\mu$	true values			rel. bias				rel. MSE			
	$\theta$	$t_0$	$\sigma$	$\hat{\mu}$	$\hat{\theta}$	$\hat{t}_0$	$\hat{\sigma}$	$\hat{\mu}$	$\hat{\theta}$	$\hat{t}_0$	$\hat{\sigma}$
50	1/20	0.015	0.01	0.078	0.012	-0.019	-0.002	0.066	0.237	0.029	0.007
			0.02	0.122	0.083	-0.023	-0.001	0.134	0.346	0.029	0.012
	0.030	0.01	0.093	0.049	-0.014	0.005	0.067	0.254	0.060	0.006	
		0.02	0.109	0.108	-0.031	-0.005	0.120	0.369	0.055	0.013	
50	1/40	0.015	0.01	0.085	0.078	-0.024	-0.005	0.055	0.245	0.015	0.002
			0.02	0.096	0.064	-0.024	0.020	0.110	0.308	0.017	0.184
	0.030	0.01	0.090	0.084	-0.043	-0.004	0.053	0.273	0.035	0.003	
		0.02	0.111	0.046	-0.023	0.014	0.118	0.311	0.035	0.183	
100	1/20	0.015	0.01	0.055	0.023	-0.021	0.159	0.054	0.179	0.025	1.66
			0.02	0.082	0.045	-0.017	0.086	0.100	0.251	0.023	0.957
	0.030	0.01	0.036	-0.001	-0.037	0.114	0.049	0.185	0.056	1.12	
		0.02	0.098	0.102	-0.047	0.155	0.119	0.288	0.050	2.03	
100	1/40	0.015	0.01	0.054	0.023	-0.022	0.081	0.043	0.184	0.016	0.741
			0.02	0.093	0.048	-0.026	0.036	0.085	0.216	0.016	0.458
	0.030	0.01	0.053	0.039	-0.021	0.019	0.035	0.172	0.038	0.110	
		0.02	0.113	0.045	-0.031	0.030	0.087	0.202	0.032	0.499	