Odhad rozdělení dob mezi událostmi z krátkých časových oken

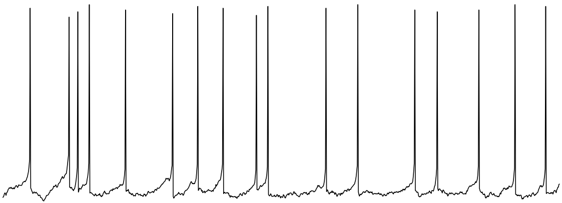
Zbyněk Pawlas

Katedra pravděpodobnosti a matematické statistiky
Matematicko-fyzikální fakulta
Univerzita Karlova v Praze

16. zimní škola JČMF ROBUST 2010
1. února 2010, Králíky

Motivace z neurofyziologie

- informace v nervovém systému je přenášena posloupností akčních potenciálů (tzv. spiků) generovaných jednotlivými neurony
- pro přenos signálu je daleko důležitější doba než tvar nebo amplituda spiku → časové kódování
- dobu spiku lze považovat za jeden bod na časové ose ⇒ časové uspořádání spiků můžeme modelovat bodovými procesy



Informace z mnoha neuronů

často je studována situace, kdy máme jeden dlouhý záznam z jednoho neuronu

ovšem v praxi cílový neuron dostává velké množství záznamů od propojených okolních neuronů, které musí ve velmi krátkém čase vyhodnotit a zareagovat

Cíl: zkoumat scénář, kdy populace neuronů přenáší informaci v krátkém časovém intervalu

Z. Pawlas, L. B. Klebanov, M. Prokop, P. Lánský (2008): parametrické odhady momentů intervalů mezi dvěma spiky

Bodové procesy

diskrétní události vyskytující se v náhodných časech

$$\cdots < X_{-1} < X_0 \leq 0 < X_1 < X_2 < \cdots$$

(jednoduchý) bodový proces $\Phi = \{X_i, i \in \mathbb{Z}\}$ – náhodná lokálně konečná posloupnost (navzájem různých) bodů v $\mathbb{R}$

$\Phi(t) = \#\{i : X_i \in [0, t]\}, \qquad \Phi(t) < k \iff X_k > t$

stacionární bodový proces – rozdělení invariantní vzhledem k posunutí
intenzita $\lambda = \mathrm{E}\Phi(1)$ – střední počet bodů v jednotkovém intervalu

čas první události (spiku) po čase 0: X_1
doby mezi událostmi: $T_i = X_{i+1} - X_i$, $i \in \mathbb{Z}$

Modely bodových procesů

používají se různé modely pro spontánní aktivitu neuronů

dvě různá zobecnění Poissonova procesu:

- proces obnovy
- dvojně stochastický Poissonův proces (Coxův proces)

v našem případě uvažujeme stacionární bodové procesy (předpoklad stacionarity není příliš restriktivní, protože procesy jsou pozorovány v krátkém časovém okně)

Modely stacionárních bodových procesů

- stacionární Poissonův bodový proces:

  $\{T_i\}$ iid $\sim \operatorname{Exp}(\lambda)$

- stacionární proces obnovy:

  $\{T_i\}$ iid s distribuční funkcí F a střední hodnotou μ

  X_1 má hustotu $f_1(x) = \frac{1-F(x)}{\mu}$

- smíšený Poissonův proces:

  náhodná řídící intenzita Λ

  podmíněně při $\Lambda = \lambda$ je Φ stacionární Poissonův proces s intenzitou λ

Redukované Palmovo rozdělení

Φ stacionární bodový proces na $\mathbb{R}$ s intenzitou λ
existuje markovské jádro $x \mapsto P_x^!$ takové, že

$$\mathrm{E} \sum_{X \in \Phi} g(X, \Phi \setminus \{X\}) = \lambda \int \int g(x, \mu)\, P_x^!(\mathrm{d}\mu)\, \mathrm{d}x$$

pro libovolnou nezápornou měřitelnou funkci g
přitom lze volit $P_x^!(U) = P_0^!(U - x)$ a

$$P_0^!(U) = \frac{1}{\lambda} \mathrm{E} \sum_{X \in \Phi \cap [0,1]} \mathbf{1}\{(\Phi \setminus \{X\}) - X \in U\}$$

Interpretace: $P_0^!$ je rozdělení $\Phi \setminus \{0\}$ za podmínky, že $0 \in \Phi$

Distribuční funkce dob mezi událostmi

$$F(t) = P_0^!(\{X_1 \leq t\}), \quad t \geq 0$$

stacionární Poissonův proces s intenzitou λ:

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

stacionární proces obnovy s distribuční funkcí F:

$$F(t) = F(t)$$

smíšený Poissonův proces s řídicí intenzitou Λ:

$$F(t) = \mathrm{E}(\mathrm{E}[\mathrm{P}(T \leq t) \mid \Lambda]) = 1 - \mathrm{E}e^{-\Lambda t}$$

Naše úloha

uvažujme nezávislé stejně rozdělené stacionární bodové procesy $\Phi_1, \ldots, \Phi_n$ (homogenní populace nezávislých neuronů)

bodové procesy jsou pozorovány v časovém intervalu délky $\Delta > 0$ (srovnatelná se střední dobou mezi událostmi)

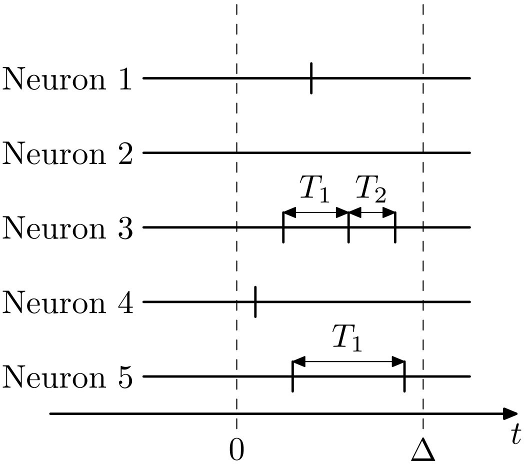
počátek pozorování je volen bez ohledu na probíhající neuronální aktivitu

Cíl: neparametricky odhadnout distribuční funkci dob mezi událostmi (spiky)

$$F(t) = P_0^!(\{X_1 \leq t\})$$

případně $G(t) = \frac{F(t)}{F(\Delta)}$ distribuční funkci dob mezi událostmi za podmínky, že jsou menší než Δ

Schematické znázornění dat



Odhady v případě jednoho okna pozorování I

Empirická distribuční funkce:

$$\tilde{F}(t) = \frac{1}{\Phi(\Delta) - 1} \sum_{i=1}^{\Phi(\Delta)-1} \mathbf{1}\{T_i \leq t\}, \quad t \leq \Delta$$

Problémy: 1. náhodný počet pozorování, závislý s $\{T_i\}$
2. výběrové vychýlení – nemůžeme pozorovat doby mezi událostmi delší než Δ
3. nevyužívá všechnu informaci obsaženou v datech – víme, že $T_{\Phi(\Delta)} = X_{\Phi(\Delta)+1} - X_{\Phi(\Delta)} > B = \Delta - X_{\Phi(\Delta)}$

Modifikovaná empirická distribuční funkce:

$$\hat{F}(t) = \begin{cases} \frac{\Phi(\Delta)-1}{\Phi(\Delta)} \tilde{F}(t), & t \leq B \\ \tilde{F}(t), & t > B \end{cases}$$

Odhady v případě jednoho okna pozorování II

Kaplanův-Meierův odhad

$$1 - \hat{F}^{KM}(t) = \prod_{s \leq t} \left(1 - \frac{\#\{i = 1, \ldots, \Phi(\Delta) - 1 : T_i = s\}}{\#\{i = 1, \ldots, \Phi(\Delta) - 1 : T_i \geq s\} + \mathbf{1}\{B \geq s\}} \right)$$

Odhad založený na zmenšeném výběru

$$\hat{F}^{RS}(t) = \frac{1}{\Phi(\Delta - t)} \sum_{i=1}^{\Phi(\Delta - t)} \mathbf{1}\{T_i \leq t\}$$

Využití odhadů z jednoho okna na případ více oken

dva přístupy:

- zprůměrování jednotlivých odhadů
- shromáždění informace ze všech oken

Zprůměrování jednotlivých odhadů

$$\bar{F}_n(t) = \frac{1}{n} \sum_{k=1}^{n} \hat{F}_{\Phi_k}(t)$$

$$\bar{F}_n^{KM}(t) = \frac{1}{n} \sum_{k=1}^{n} \hat{F}_{\Phi_k}^{KM}(t)$$

$$\bar{F}_n^{RS}(t) = \frac{1}{n} \sum_{k=1}^{n} \hat{F}_{\Phi_k}^{RS}(t)$$

Spojení údajů z jednotlivých oken

$$1 - \hat{F}_n^{KM}(t) =$$

$$\prod_{s \le t} \left(1 - \frac{\sum\limits_{k:\Phi_k(\Delta)>1} \sum\limits_{i=1}^{\Phi_k(\Delta)-1} \mathbf{1}\{T_i^{(k)} = s\}}{\sum\limits_{k:\Phi_k(\Delta)>1} \sum\limits_{i=1}^{\Phi_k(\Delta)-1} \mathbf{1}\{T_i^{(k)} \ge s\} + \sum\limits_{k:\Phi_k(\Delta)\ge 1} \mathbf{1}\{B^{(k)} \ge s\})} \right)$$

$$\hat{F}_n^{RS}(t) = \frac{\sum\limits_{k=1}^{n} \sum\limits_{i=1}^{\Phi_k(\Delta-t)} \mathbf{1}\{T_i^{(k)} \le t\}}{\sum\limits_{k=1}^{n} \Phi_k(\Delta - t)}$$

Odhad pro smíšený Poissonův proces I

Laplaceův funkcionál pro Poissonův proces

$$L_\Phi(f) = \exp\left\{-\lambda \int_{\mathbb{R}} (1 - \mathrm{e}^{-f(x)})\, \mathrm{d}x\right\}$$

Laplaceův funkcionál pro smíšený Poissonův proces

$$L_\Phi(f) = \mathrm{E}\exp\left\{-\int f(x)\,\Phi(\mathrm{d}x)\right\} = \mathrm{E}\exp\left\{-\Lambda \int (1 - \mathrm{e}^{-f(x)})\, \mathrm{d}x\right\}$$

speciálně pro $f(x) = -\log(1 - t/\Delta)$, $0 \le x \le \Delta$,

$$L_\Phi(f) = \mathrm{E}\mathrm{e}^{-\Lambda t}, \quad t < \Delta,$$

$$F(t) = 1 - \mathrm{E}\mathrm{e}^{-\Lambda t} = 1 - L_\Phi(f)$$

Odhad pro smíšený Poissonův proces II

nestranný odhad Laplaceova funkcionálu

$$\widehat{L_\Phi(f)} = \frac{1}{n}\sum_{k=1}^{n} \exp\left\{-\sum_{i=1}^{\Phi_k(\Delta)} f(X_i^{(k)})\right\}$$

nestranný odhad F:

$$\begin{aligned}\hat{F}_n^{MP}(t) &= 1 - \frac{1}{n}\sum_{k=1}^{n} \exp\left\{\sum_{i=1}^{\Phi_k(\Delta)} \log\left(1 - \frac{t}{\Delta}\right)\right\} \\ &= 1 - \frac{1}{n}\sum_{k=1}^{n} \left(1 - \frac{t}{\Delta}\right)^{\Phi_k(\Delta)}\end{aligned}$$

nepotřebujeme přesné časy události, stačí počty v jednotlivých oknech

Konzistence a asymptotická normalita

Odhad založený na zmenšeném výběru:

$$Z_k(t) = \sum_{i=1}^{\Phi_k(\Delta - t)} \mathbf{1}\{T_i^{(k)} \le t\}, \quad \hat{\lambda}_n^{RS} = \frac{1}{n(\Delta - t)} \sum_{k=1}^{n} \Phi_k(\Delta - t)$$

$$\mathrm{E} Z_k(t) = \lambda(\Delta - t) F(t), \quad \mathrm{E}\Phi_k(\Delta - t) = \lambda(\Delta - t)$$

$$\Rightarrow \hat{F}_n^{RS}(t) = \frac{\sum_{k=1}^{n} Z_k(t)}{n(\Delta - t)\hat{\lambda}_n^{RS}} \xrightarrow[n\to\infty]{} \frac{\lambda F(t)}{\lambda} = F(t) \quad \text{s.j.}$$

když Φ je proces obnovy nebo smíšený Poissonův proces, tak

$$Y_n^{RS}(t) = \sqrt{n}\left(\hat{F}_n^{RS}(t) - F(t)\right) \xrightarrow[n\to\infty]{} N(0, \frac{F(t)(1 - F(t))}{\lambda(\Delta - t)})$$

Slabá konvergence

Φ smíšený Poissonův proces, $\mathrm{E}\Lambda^2 < \infty$, $\Delta_0 < \Delta$

$\{Y_n^{RS}(t), t \in [0, \Delta_0]\}$ konverguje slabě v $D([0, \Delta_0])$
k centrovanému gaussovskému procesu $\{Y^{RS}(t), t \in [0, \Delta_0]\}$
s kovarianční funkcí
$\mathrm{E}\, Y^{RS}(t)\, Y^{RS}(s) = (F(s \wedge t) - F(s)F(t))/(\lambda(\Delta - (s \wedge t))$

$$Y_n^{MP}(t) = \sqrt{n}\left(\hat{F}_n^{MP}(t) - F(t)\right), \quad t \in [0, \Delta_0]$$

konverguje slabě v $C([0, \Delta_0])$ k centrovanému gaussovskému
procesu $\{Y^{MP}(t), t \in [0, \Delta_0]\}$ s kovarianční funkcí
$\mathrm{E}\, Y^{MP}(t)\, Y^{MP}(s) = \mathrm{E}e^{-(s+t)\Lambda + st\Lambda/\Delta} - \mathrm{E}e^{-\Lambda t}\mathrm{E}e^{-\Lambda s}$

Simulace

$\Delta = 1,\ n = 400$

proces obnovy: $T_i \sim \Gamma(a, b)$, $\mathrm{E}\,T = 1$, $\mathrm{CV}\ T = \frac{\sqrt{\mathrm{Var}\ T}}{\mathrm{E}\,T} = 1{,}5$

$$F(t) = \int_0^t \frac{b^a}{\Gamma(a)} \mathrm{e}^{-bs} s^{a-1}\,\mathrm{d}s$$

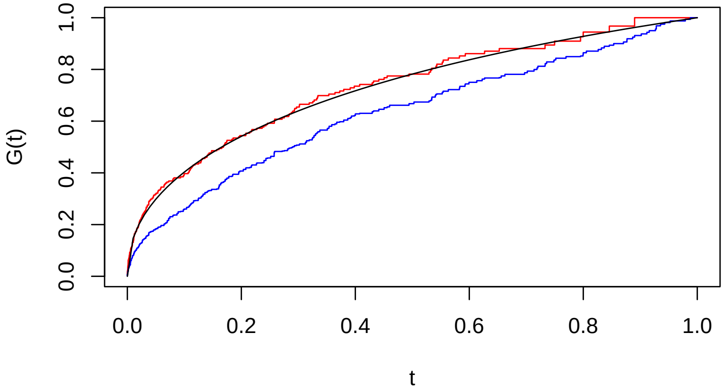
smíšený Poissonův proces: $\Lambda \sim \Gamma(a, b)$, $\mathrm{E}\,T = 1$, $\mathrm{CV}\ T = 1{,}5$

$$F(t) = 1 - \left(\frac{b}{b+t}\right)^a$$

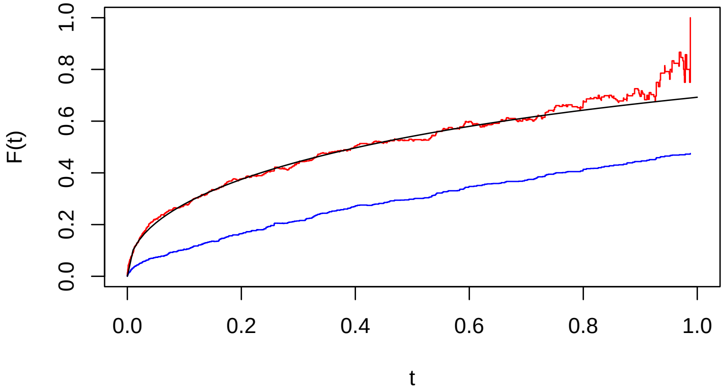
odhad $G(t)$:

$$\hat{G}(t) = \frac{\hat{F}(t)}{\hat{F}(\Delta)}$$

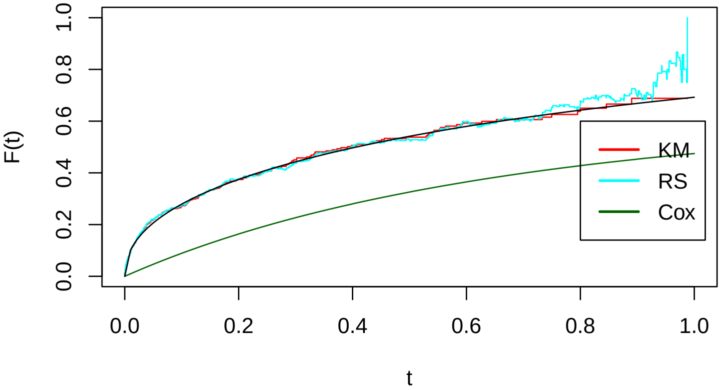
Proces obnovy – Kaplanův-Meierův odhad



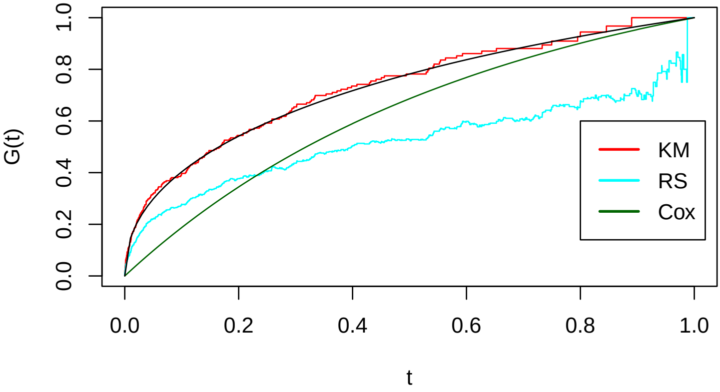
Proces obnovy – odhad založený na zmenšeném výběru



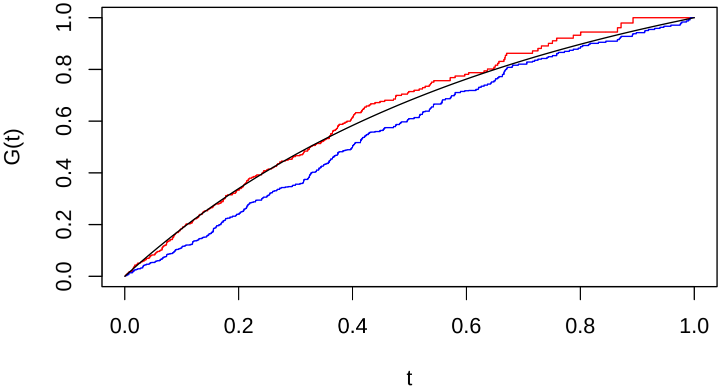
Proces obnovy – porovnání odhadů



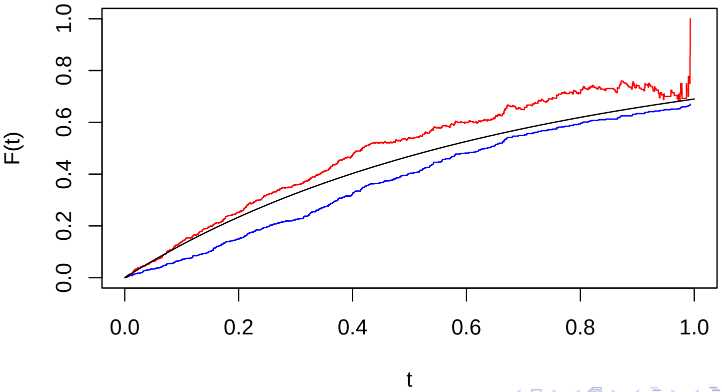
Proces obnovy – porovnání odhadů



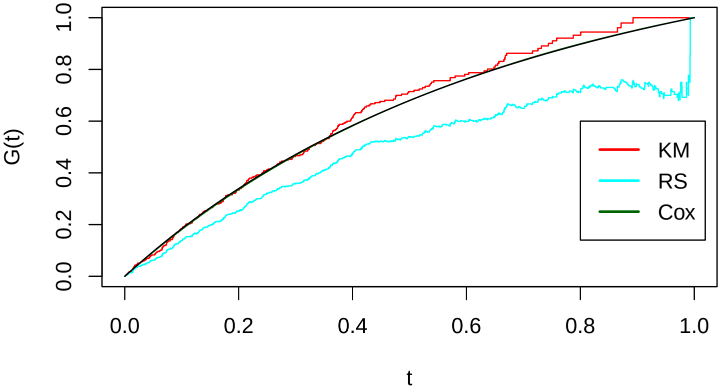
Smíšený Poissonův proces – Kaplanův-Meierův odhad



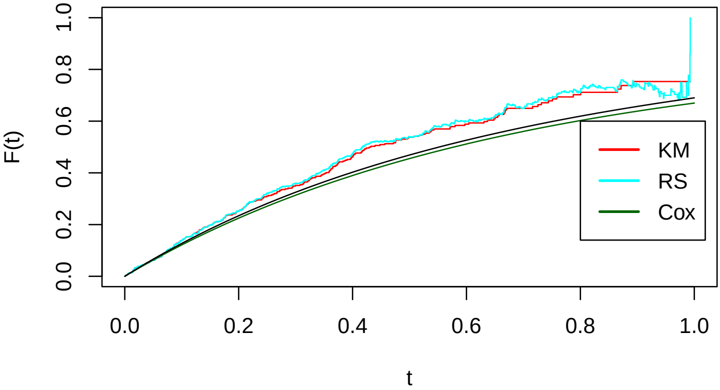
Smíšený Poissonův proces – odhad založený na zmenšeném výběru



Smíšený Poissonův proces – porovnání odhadů



Smíšený Poissonův proces – porovnání odhadů



Závěrečné poznámky

- odhady získané sloučením informace ze všech oken jsou přesnější než odhady obdržené zprůměrováním
- v případě procesu obnovy má nejmenší odchylku od skutečné distrib. funkce odhad Kaplanova-Meierova typu
- odhad založený na zmenšeném výběru dává dobré výsledky pro malá t, pro velká t se hodně informace z dat ignoruje; výsledný odhad nemusí být neklesající funkce – je možné uvažovat horní obálku $\hat{F}_n^{RS+}(t) = \sup_{s \leq t} \hat{F}_n^{RS}(s)$
- místo času od poslední události $(B = \Delta - X_{\Phi(\Delta)})$ bychom v odhadech mohli uvažovat časy do první události (X_1)
- v případě procesu obnovy existuje vztah mezi rozdělením B a rozdělením dob mezi událostmi T, z odhadu rozdělení B tak můžeme dostat odhad rozdělení T