Neparametrické odhady podmíněné rizikové funkce

Iveta Selingerová

Ústav matematiky a statistiky
Přírodovědecká fakulta
Masarykova univerzita

Finanční matematika v praxi III
a Matematické modely a aplikace
3.-6. září 2013

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Obsah

Analýza přežití

- Čas přežití T (čas od počáteční události do koncové události)
- Cenzorování (levostranné, intervalové, **pravostranné**)
- $T_1, T_2, \ldots, T_n$ nezávislé a stejně rozdělené časy přežití s distribuční funkcí F
- $C_1, C_2, \ldots, C_n$ nezávislé a stejně rozdělené časy cenzorování s distribuční funkcí G
- (X_i, δ_i), kde $X_i = min(T_i, C_i)$ a $\delta_i = I_{T_i \leq C_i}$
- $X_1, X_2, \ldots, X_n$ nezávislé a stejně rozdělené časy sledování s distribuční funkcí L splňující $L(x) = 1 - (1 - F(x))(1 - G(x))$

Otázky v analýze přežití

1. Jaká je pravděpodobnost přežití např. po 5 letech? Kdy je největší riziko úmrtí?
2. Je rozdíl v přežití např. mezi ženami a muži?
3. Jak závisí přežití např. na věku?

Funkce přežití a riziková funkce

- Funkce přežití $\overline{F}(x) = P(T \geq x) = 1 - F(x)$
  - Kaplan-Meierův odhad funkce přežití

$$\widehat{\overline{F}}(x) = \prod_{i:X_{(i)}<x} \left(\frac{n-i}{n-i+1} \right)^{\delta_{(i)}}$$

kde $X_{(i)}$ značí i-té pozorování seřazených $X_1, X_2, \ldots, X_n$ a $\delta_{(i)}$ je odpovídající indikátor cenzorování

- Riziková funkce $\lambda(x)\Delta x = P(x \leq T < x + \Delta x | T \geq x)$

$$\lambda(x) = \frac{f(x)}{\overline{F}(x)}$$

Funkce přežití a riziková funkce

- Kumulativní riziková funkce $\Lambda(x) = \int_0^x \lambda(u)\mathrm{d}u$
  - Nelson-Aalenův odhad

$$\hat{\Lambda}_n(x) = \sum_{x_{(i)} \leq x} \frac{\delta_{(i)}}{n - i + 1}$$

Předpoklady

1. $[0, T]$ interval, pro který $L(T) < 1$
2. $\lambda \in C^2[0, T]$
3. K reálná funkce splňující podmínky
   1. $K \in Lip[-1, 1]$
   2. $\mathrm{supp}(K) = [-1, 1], K(-1) = K(1) = 0$
   3. $$\int_{-1}^{1} x^j K(x)\mathrm{d}x = \begin{cases} 0 & j = 1, \\ 1 & j = 0, \\ \beta_2 \neq 0 & j = 2. \end{cases}$$

   Taková funkce se nazývá jádro řádu 2
4. {h(n)} nenáhodná posloupnost kladných čísel splňující

$$\lim_{n\to\infty} h(n) = 0, \lim_{n\to\infty} h(n)n = \infty$$

Jádrové odhady rizikové funkce

- $\lambda(t) = \frac{f(t)}{\overline{F}(t)}$ and $\overline{L}(x) = \overline{F}(x)\overline{G}(x)$

$$\lambda(x) = \frac{\overline{G}(x)f(x)}{\overline{L}(x)} = \frac{r(x)}{\overline{L}(x)}$$

$r(x)$ hustota úmrtí, $\overline{L}(x)$ funkce přežití pozorovaných časů

- $\hat{\lambda}(x) = \frac{\hat{r}(x)}{\hat{\overline{L}}(x)}$
- $\hat{\lambda_1}(x) = \frac{\frac{1}{nh}\sum_{i=1}^{n}\delta_i K\left(\frac{x-X_i}{h}\right)}{1-L_n(x)}$
- $\hat{\lambda_2}(x) = \frac{\frac{1}{nh}\sum_{i=1}^{n}\delta_i K\left(\frac{x-X_i}{h}\right)}{1-\frac{1}{n}\sum_{i=1}^{n} W\left(\frac{x-X_i}{h}\right)}$, where $W(x) = \int_{-\infty}^{x} K(t)\mathrm{d}t$

Jádrové odhady rizikové funkce

- Konvoluce jádra K s Nelson-Aalenovým odhadem

$$\hat{\lambda}_3(x) = \frac{1}{h} \int K\left(\frac{x-u}{h}\right) \mathrm{d}\hat{\Lambda}_n(u) =$$

$$= \frac{1}{h} \sum_{i=1}^{n} K\left(\frac{x - X_{(i)}}{h}\right) \frac{\delta_{(i)}}{n-i+1}$$

Volba vyhlazovacího parametru

- Metoda křížového ověřování $\hat{h}_{CV} = \arg\min_{h \in H_n} CV(h)$

$$CV(h) = \int_0^\infty \hat{\lambda}^2(x)\mathrm{d}x - 2\frac{1}{n}\sum_{i=1}^{n} \frac{\hat{\lambda}_{-i}(X_i)}{(1 - L_n(X_i))}\delta_i$$

- Metoda maximální věrohodnosti $\hat{h}_{ML} = \arg\max_{h \in H_n} ML(h)$

$$ML(h) = \prod_{i=1}^{n} \hat{\lambda}_{-i}(X_i)^{\delta_i} \overline{F}_{-i}(X_i)$$

- iterační metoda, plug-in metoda, ...

Podmíněná riziková funkce

- $(X_i, \delta_i, \mathbf{Y}_i)$, kde $X_i = min(T_i, C_i)$, $\delta_i = I_{T_i \leq C_i}$ a $\mathbf{Y}_i$ je vektor kovariát
- $\lambda(x|y)\Delta x = P(x \leq T < x + \Delta x | T \geq x, Y = y)$

$$\lambda(x|y) = \frac{f(x|y)}{\overline{F}(x|y)} = \frac{\overline{G}(x|y)f(x|y)}{\overline{L}(x|y)} = \frac{r(x|y)}{\overline{L}(x|y)}$$

- $\Lambda(x|y) = \int_0^x \lambda(u|y)\mathrm{d}u$

Coxův model proporcionálního rizika

- $\lambda(x|y) = \lambda_0(x)\exp(\beta y)$, kde $\lambda_0(x)$ je základní riziko a β je parametr
- Poměr riziko:

$$\frac{\lambda(x|y)}{\lambda(x|y^*)} = \frac{\lambda_0(x)\exp(\beta y)}{\lambda_0(x)\exp(\beta y^*)} = \exp(\beta(y - y^*))$$

- $\widehat{\beta}$ maximalizace věrohodnostní funkce

Jádrové odhady podmíněné rizikové funkce

- $$\hat{\lambda_2}(x|y) = \frac{\frac{1}{h_x}\sum_{i=1}^{n} w_i(y)\delta_i K\left(\frac{x-X_i}{h_x}\right)}{\sum_{i=1}^{n} w_i(y) W\left(\frac{X_i-x}{h_x}\right)}$$

  kde $W(x) = \int_{-\infty}^{x} K(t)\mathrm{d}t$ a $w_i(y)$ jsou váhy
- Nadaraya – Watsonovy váhy

$$w_i(y) = \frac{K\left(\frac{y-Y_i}{h_y}\right)}{\sum_{j=1}^{n} K\left(\frac{y-Y_j}{h_y}\right)}$$

Jádrové odhady podmíněné rizikové funkce

- $$\hat{\lambda_3}(x|y) = \frac{1}{h_x}\int K\left(\frac{x-u}{h_x}\right)\mathrm{d}\hat{\Lambda}_n(u|y) =$$
$$= \frac{1}{h_x}\sum_{i=1}^{n} K\left(\frac{x-X_{(i)}}{h_x}\right)\frac{\delta_{(i)}w_{(i)}(y)}{1-\sum_{j=1}^{i-1}w_{(j)}(y)}$$

kde $X_{(i)}$ jsou seřazené pozorované časy, $\delta_{(i)}$ je odpovídající indikátor cenzorování and $w_{(i)}(y)$ je odpovídající váha

Metoda křížového ověřování

$$ISE = \int\int (\hat{\lambda}(x|y) - \lambda(x|y))^2 f(y)\mathrm{d}x\mathrm{d}y =$$
$$= \int\int \hat{\lambda}^2(x|y) f(y)\mathrm{d}x\mathrm{d}y - 2\int\int \hat{\lambda}(x|y)\lambda(x|y) f(y)\mathrm{d}x\mathrm{d}y + \cdots =$$
$$= \int\int \hat{\lambda}^2(x|y) f(y)\mathrm{d}x\mathrm{d}y - 2\int\int \frac{\hat{\lambda}(x|y)}{\overline{L}(x|y)} r(x|y) f(y)\mathrm{d}x\mathrm{d}y + \cdots$$

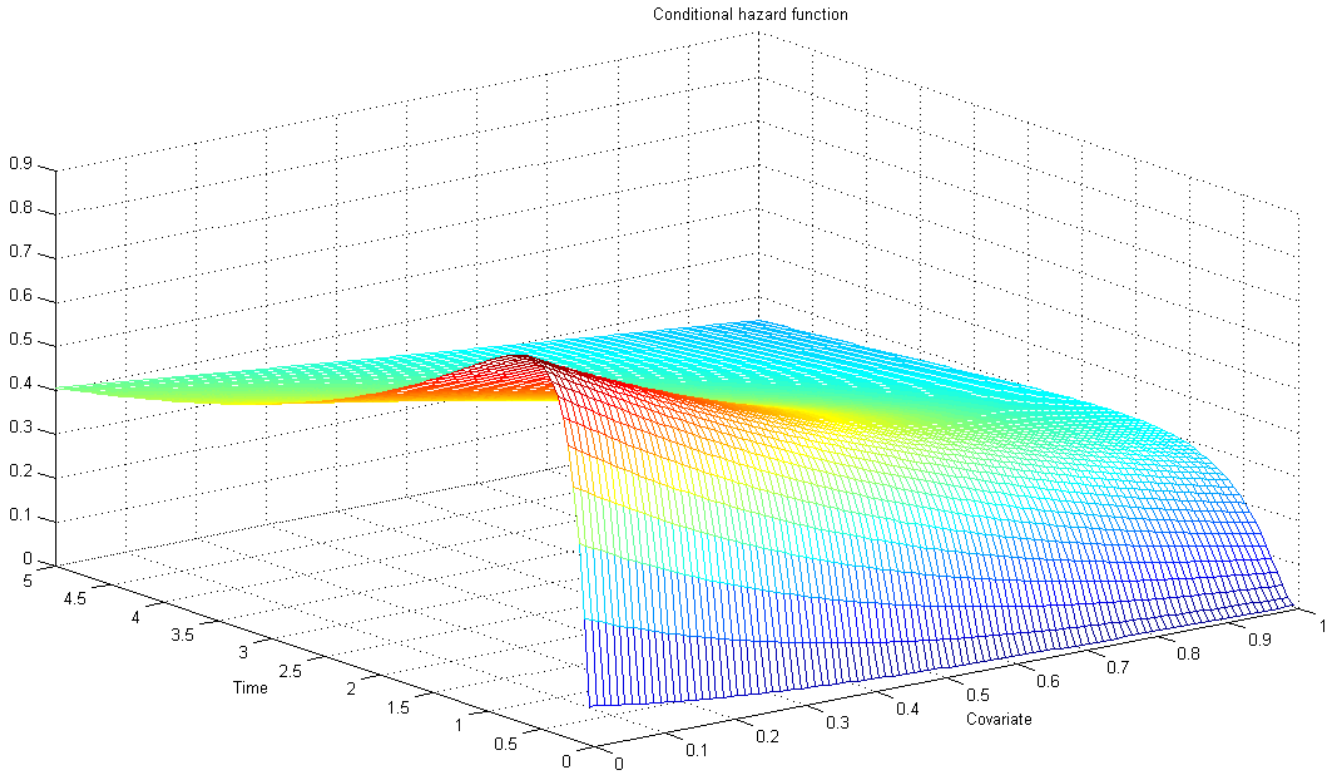
$$CV(h_x, h_y) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\infty} \hat{\lambda}^2(x|Y_i)\mathrm{d}x - 2\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{\lambda}_{-i}(X_i|Y_i)}{\hat{\overline{L}}(X_i|Y_i)}\delta_i$$

$$(\hat{h}_x, \hat{h}_y) = \arg\min CV(h_x, h_y)$$

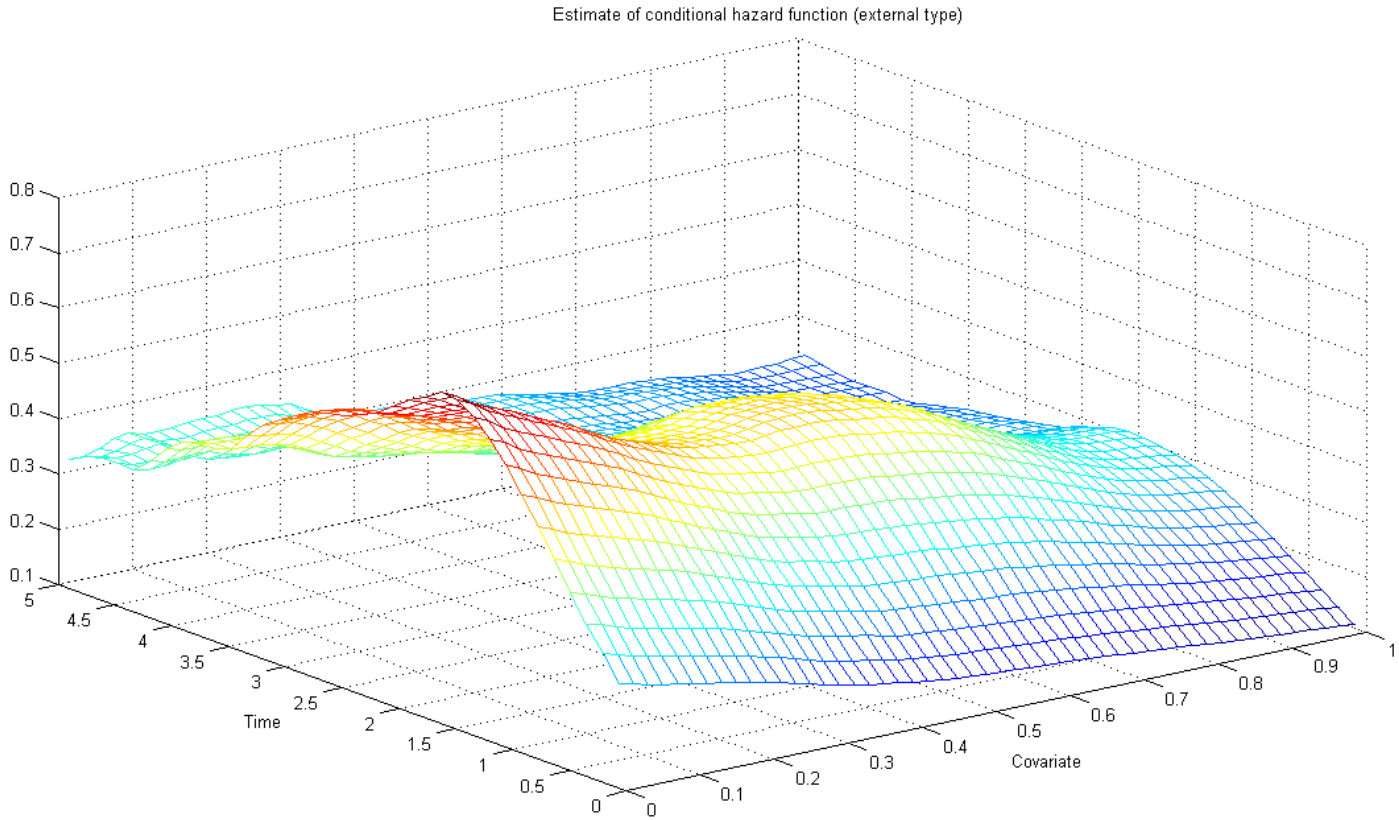
Simulace

Simulace 500 pozorování. Cenzorování 15 %.

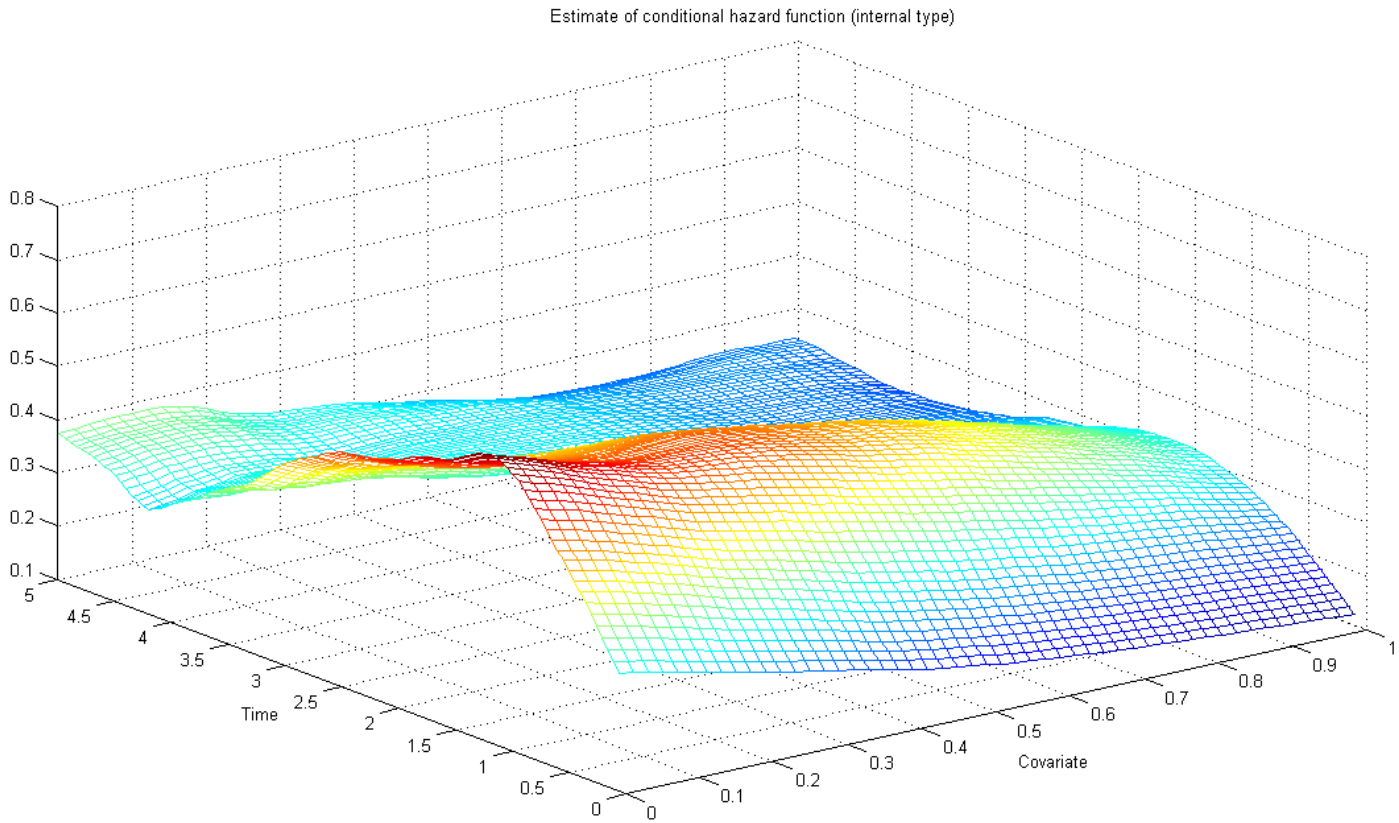
$\log T_i = Y_i + \varepsilon_i,\ Y_i \sim U(0,1),\ \varepsilon_i \sim N(0,1),\ C_i \sim \log N(1.9, 0.6^2)$



Vnější typ odhadu



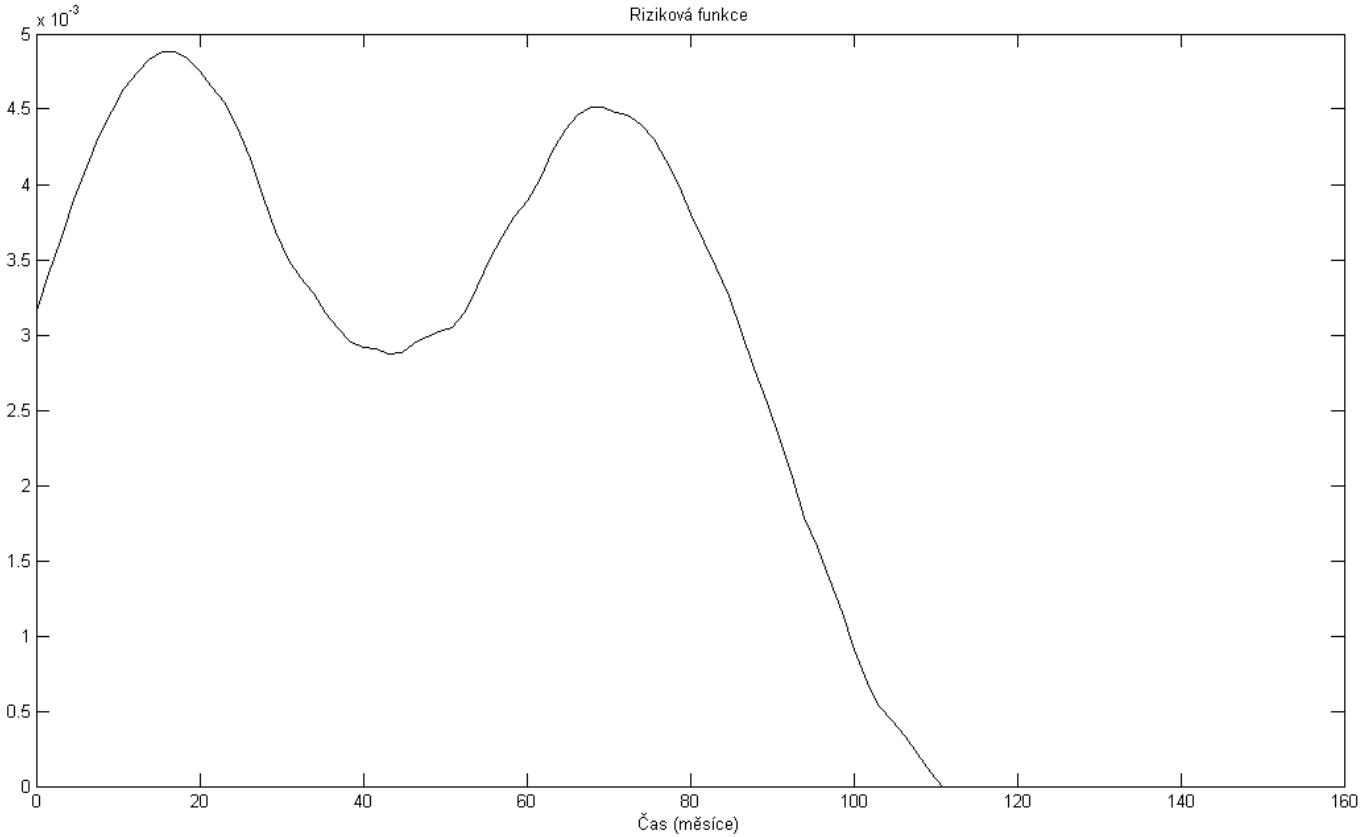
Vnitřní typ odhadu



Pacienti s rakovinou ledvin

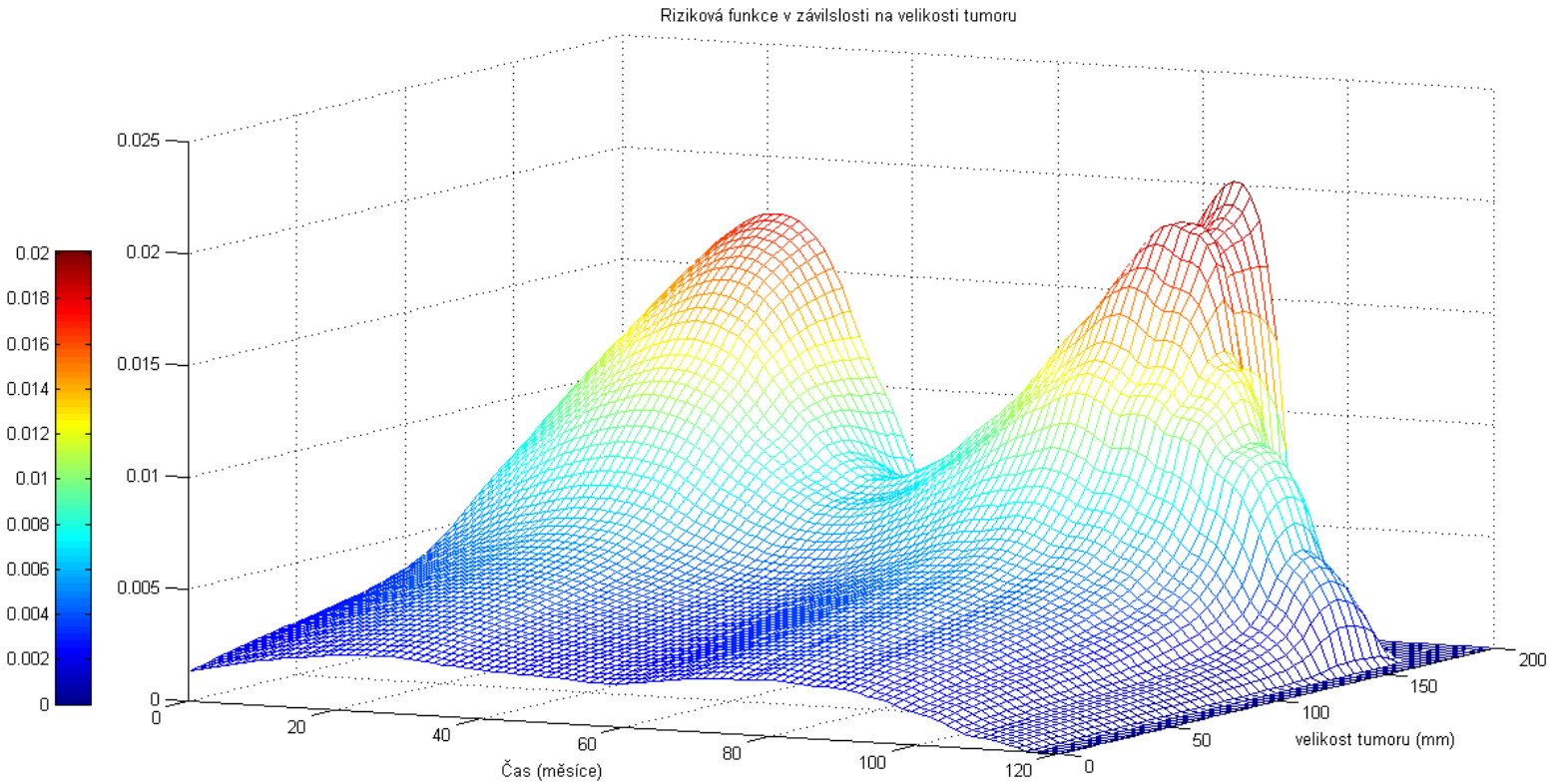
- 391 pacientů, kteří byli v letech 2003-2011 léčeni na MOÚ v Brně
- 75 úmrtí (cenzorování 80 %)
- čas sledování 9 dní až 154 měsíců (medián 38 měsíců)
- velikost tumoru 0 až 200 mm (průměr 54 mm), hranice pro nepříznivou předpověď 70 mm
- věk 21 až 93 let (průměr 61 let)

Pacienti s rakovinou ledvin



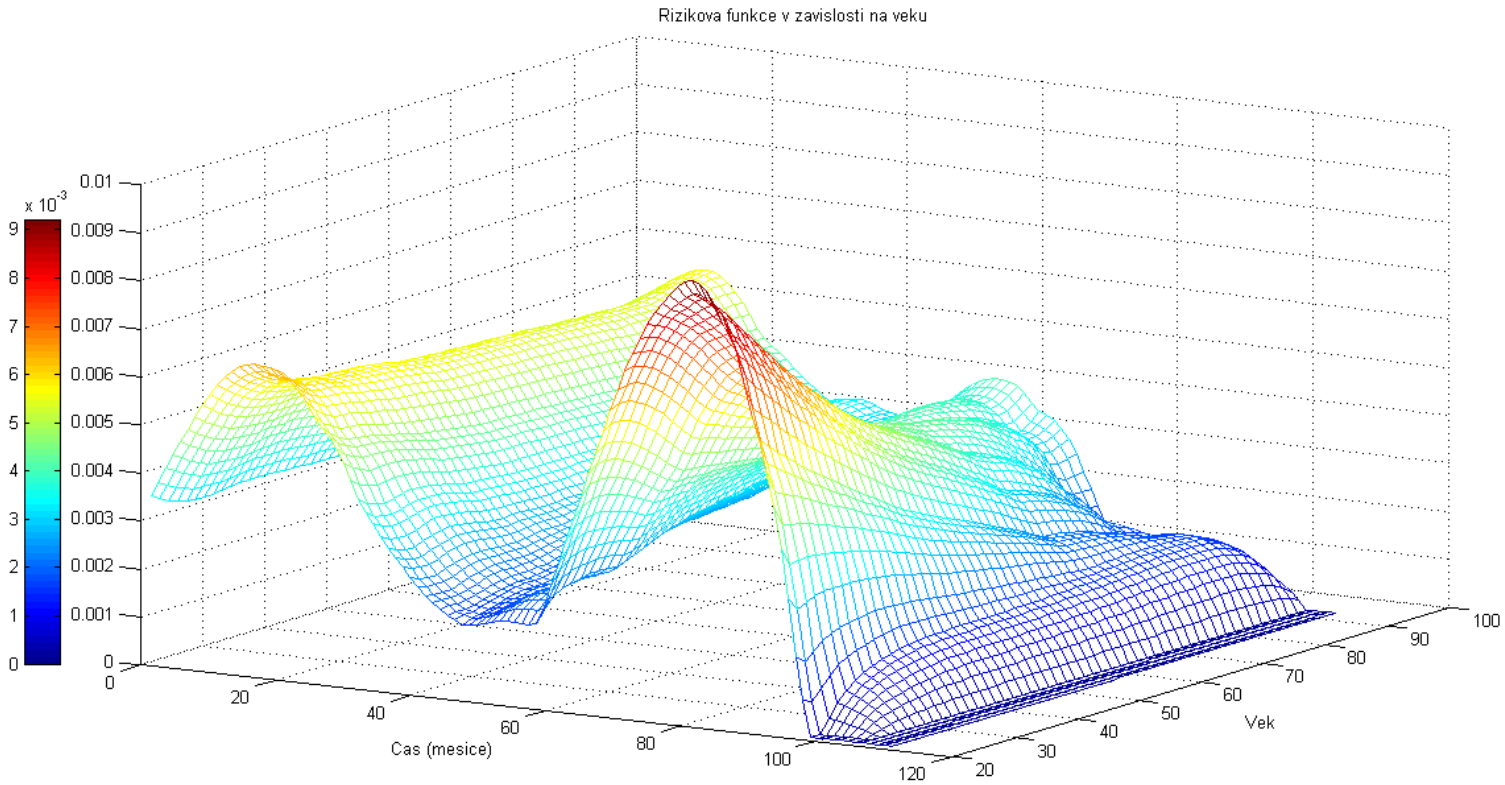
Pacienti s rakovinou ledvin

kovariáta = velikost tumoru



Pacienti s rakovinou ledvin

kovariáta = věk



Děkuji za pozornost.