Populační modely v analýze přežití

Ondřej Černý, Zdeněk Pospíšil

Ústav matematiky a statistiky
Přírodovědecká fakulta
Masarykova univerzita

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

evropský sociální fond v ČR
EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost
AMathNet síť pro transfer znalostí v aplikované matematice
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Obsah

- Úvod
- Stochastický vs. deterministický přístup
- Populační modely a analýza přežití
    - Odhad parametrů
    - Příklad na odhad CLFS

- Klasické odhady
  - Kaplan-Meier(1958)
  - Nelson-Aalen
- Problémy
  - Malý výskyt událostí
  - Jen některé typy cenzorování

Možné pohledy
Stochastický přístup

- Analýza přežití se zabývá analýzou času přežití, který je popsán náhodnou veličinou T
  - $P(T < t) = F(t) = \int_0^t f(u)\mathrm{d}u$
- Funkce přežití
  - $S(t) = P(T \geq t) = 1 - F(t)$
- Riziková funkce - pravděpodobnost, že jedinec nepřežije Δt, dožil-li se t
  - $\lambda(t) = \frac{P(t \leq T < t + \Delta t | T \geq t))}{\Delta t} = \frac{f(t)}{S(t)} = -\frac{\mathrm{d}}{\mathrm{d}t} \ln(S(t))$

Možné pohledy

Deterministický přístup

- Analýza přežití se zabývá analýzou přežití jedinců v kohortě o velikosti $N(t)$
- Funkce přežití - podíl počtu jedinců živých v čase t a na počátku
  - $S(t) = \frac{N(t)}{N(0)}$
  - $N(t + \Delta t) = N(t) - d(t)N(t)\Delta t$
  - $d(t) = -\frac{\mathrm{d}}{\mathrm{d}t} \ln(N(t))$

Možné pohledy
Shrnutí

	stochasticky T	deterministicky N
funkce přežití	$S(t) = P(T \geq t)$	$\frac{N(t)}{N(0)}$
distribuční funkce	$F(t) = 1 - S(t)$	$1 - \frac{N(t)}{N(0)}$
hustota	$f(t) = \frac{\mathrm{d}}{\mathrm{d}t}(F(t))$	$-\frac{\frac{\mathrm{d}}{\mathrm{d}t}N(t)}{N(0)}$
riziková funkce	$\frac{f(t)}{S(t)} = -\frac{\mathrm{d}}{\mathrm{d}t}\ln(S(t))$	$d(t) = -\frac{\mathrm{d}}{\mathrm{d}t}\ln(N(t))$

- Uvažme maticový populační model s k třídami
  $\mathbf{n}(t+1) = A\mathbf{n}(t)$
- Dvourozměrný systém
- Potřebujeme popis kohorty v jednotlivých časech
- A má neznámé parametry- některé v závislosti na modelu však můžeme pokládat za nulové

Odhady parametrů

Regrese

- známe stavy kohorty $n(0), n(1), \ldots, n(T)$
- $n_i(t+1) = \sum_{j=1}^{k} a_{ij} n_j(t) + \epsilon_i(t)$
- $\mathrm{E}(\epsilon_i) = 0, \mathrm{D}(\epsilon_i) = \sigma^2$
- $\mathbf{Y}_i = \mathbf{X}\beta + \epsilon_i$
- $\beta = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_i$

Odhady parametrů

Metoda maximální věrohodnosti

- známe stavy kohorty $n(0), n(1), \dots, n(T)$
- $n_i(t+1) = \mathrm{e}^{d_i(t)} \sum_{j=1}^{k} a_{ij} n_j(t)$
- $d(t) \sim \mathrm{N}(0, \Sigma)$
- $\ln(n_i(t+1)) = d_i(t) + \ln \sum_{j=1}^{k} a_{ij} n_j(t)$
- $m(t) := \ln(n_i(t))$
- věrohodnostní rovnice (vzhledem k nezávislosti)
  $P(m(1), m(2), \dots, m(T) | A, \Sigma, m(0)) =$
  $\prod_{t=1}^{T} P(m(t) | A, \Sigma, m(t-1)) =$
  $\prod_{t=1}^{T} \frac{1}{\sqrt{(2\pi)^k \det\Sigma}} e^{-\frac{1}{2}(m(t)-\hat{m}(t))^T \Sigma^{-1} (m(t)-\hat{m}(t))} \to \max$
- Σ neznáme, $\hat{m}$ lze určit z matice
- problém s nezávislostí, nekorelovaností

- Uvažme maticový populační model s k třídami

$$\mathbf{n}(t+1) = A\mathbf{n}(t)$$

$$\mathbf{n}(\mathrm{t+1}) = \begin{pmatrix} n_1(t) & \cdots & 0 & n_2(t) & \cdots & 0 & \cdots & n_k(t) & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & n_1(t) & 0 & \cdots & n_2(t) & \cdots & 0 & \cdots & n_k(t) \end{pmatrix} \begin{pmatrix} a_{11} \\ \vdots \\ a_{k1} \\ \vdots \\ a_{1k} \\ \vdots \\ a_{kk} \end{pmatrix}$$

$$\mathbf{n}(t+1) = N(t) vecA$$

Vypustíme sloupce z $N(t)$ odpovídající nulovým parametrům ve vektoru $vecA$

$$N(t) \to M(t)$$

$$vecA \to p$$

Dostáváme tak model

$$\mathbf{n}(t+1) = M(t)p$$

Předpokládejme, že známe stav populace v časech $0, 1, 2, \ldots, T$

$$\begin{pmatrix} \mathbf{n}(1) \\ \mathbf{n}(2) \\ \vdots \\ \mathbf{n}(T) \end{pmatrix} = \begin{pmatrix} M(0) \\ M(1) \\ \vdots \\ M(T-1) \end{pmatrix} p$$

$$\mathbf{z} = \mathbf{M}p$$

Pro populaci bez náhodných vlivů by platilo

$$\mathbf{z} - \mathbf{M}p = \mathbf{o}$$

Řešíme tedy úlohu kvadratického programování

$$\frac{1}{2}p^T Cp - d^T p \to min,$$

kde

$$C = \mathbf{M}^T\mathbf{M}$$

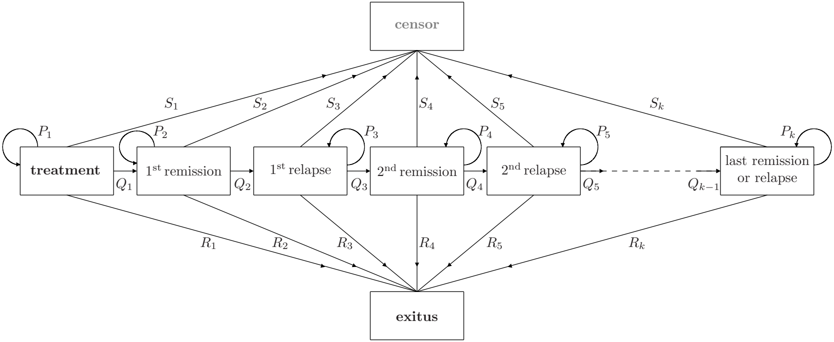
$$d = \mathbf{M}^T\mathbf{z}$$

Souhrn

- Máme parametry
- Výhoda - práce s málo daty, snadné cenzorování, možnost snadného podchycení složitějších funkcí
- DFS-Pravděpodobnost být na živu a v remisi v nějakém intervalu po zákroku
- CDFS-Pravděpodobnost být na živu a v původní nebo následné remisi

Kvadratické programování a CLFS

Možné stavy imatinib



- Studie 20 pacientů (49-2366 dní)
- Model $\mathbf{n}(t+1) = A\mathbf{n}(t)$

$$A = \begin{pmatrix} P_1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ Q_1 & P_2 & 0 & \cdots & 0 & 0 & 0 \\ 0 & Q_2 & P_3 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & P_{k-2} & 0 & 0 \\ R_1 & R_2 & R_3 & \cdots & R_{k-2} & P_{k-1} & 0 \\ S_1 & S_2 & S_3 & \cdots & S_{k-2} & 0 & P_k \end{pmatrix}$$

- Funkcionální omezení z matice

- Předpokládejme, že cenzorování je nezávislé na dalších procesech (smrt, relaps, remise)
- Nové pravděpodobnosti budou úměrné pravděpodobnostem v původním modelu

$$p_i = \frac{P_i}{P_i + Q_i + R_i}, r_i = \frac{R_i}{P_i + Q_i + R_i} \quad i = 1, 2, \cdots, k-2$$

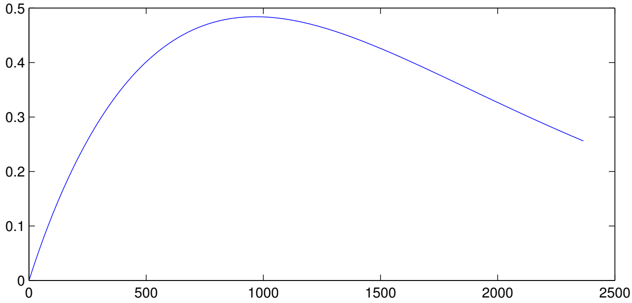
$$q_i = \frac{Q_i}{P_i + Q_i + R_i} \quad i = 1, 2, \cdots, k-3$$

- Máme nový model $\mathbf{n}(t+1) = B\mathbf{n}(t)$

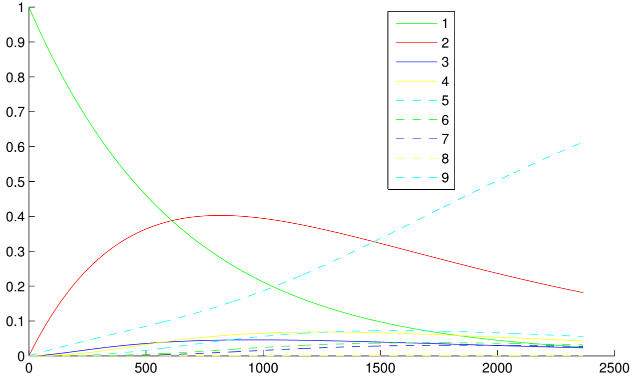
$$B = \begin{pmatrix} p_1 & 0 & 0 & \cdots & 0 & 0 \\ q_1 & p_2 & 0 & \cdots & 0 & 0 \\ 0 & q_2 & p_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & p_{k-2} & 0 \\ r_1 & r_2 & r_3 & \cdots & r_{k-2} & 1 \end{pmatrix}$$

- CLFS pak určíme takto

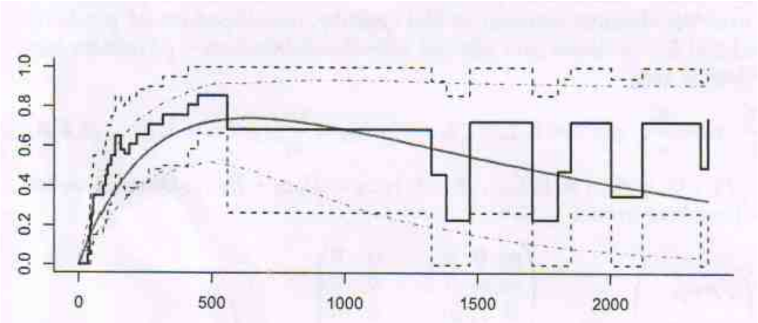
$$CLFS(t) = \frac{1}{N}(n_2(t) + n_4(t) + \cdots + n_{2[(k-2)/2]}(t))$$


0.5
0.4
0.3
0.2
0.1
0
0
500
1000
1500
2000
2500

Kvadratické programování a CLFS
Složky



Kvadratické programování a CLFS
Odhad konfindenčního intervalu



Děkuji za pozornost a přeji pěkný den