

Groupe de Travail des Doctorants

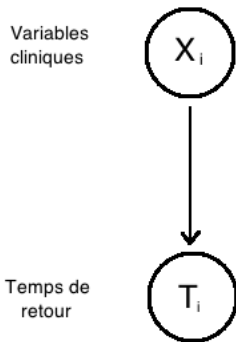
Modèles de survie en grande dimension

Un modèle de mélange et ses applications

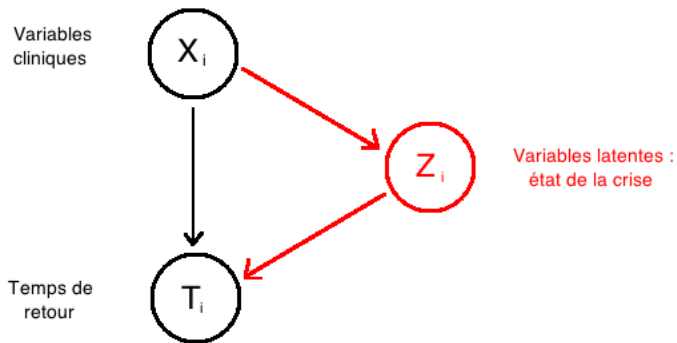
Simon Bussy

Le 11 Avril 2017

Modélisation



Modélisation



Modèle de mélange

- $Z_i \in \{0, 1\}$ v.a. latente t.q. $\forall i \in \llbracket 1, n \rrbracket, \forall k \in \{0, 1\}, T_i | Z_i = k \sim \ell_k$

Modèle de mélange

- $Z_i \in \{0, 1\}$ v.a. latente t.q. $\forall i \in \llbracket 1, n \rrbracket, \forall k \in \{0, 1\}, T_i | Z_i = k \sim \ell_k$
- Quantité primordiale : $\pi_0(X_i) = \mathbb{P}(Z_i = 0 | X_i) = \frac{1}{1 + e^{-X_i^\top \beta}}$

Modèle de mélange

- $Z_i \in \{0, 1\}$ v.a. latente t.q. $\forall i \in \llbracket 1, n \rrbracket, \forall k \in \{0, 1\}, T_i | Z_i = k \sim \ell_k$
- Quantité primordiale : $\pi_0(X_i) = \mathbb{P}(Z_i = 0 | X_i) = \frac{1}{1 + e^{-X_i^\top \beta}}$
- $\forall i \in \llbracket 1, n \rrbracket, Z_i \sim \mathcal{B}(1 - \pi_0(X_i))$

Modèle de mélange

- $Z_i \in \{0, 1\}$ v.a. latente t.q. $\forall i \in \llbracket 1, n \rrbracket, \forall k \in \{0, 1\}, T_i | Z_i = k \sim \ell_k$
- Quantité primordiale : $\pi_0(X_i) = \mathbb{P}(Z_i = 0 | X_i) = \frac{1}{1 + e^{-X_i^\top \beta}}$
- $\forall i \in \llbracket 1, n \rrbracket, Z_i \sim \mathcal{B}(1 - \pi_0(X_i))$
- $\forall i \in \llbracket 1, n \rrbracket, \forall t \in \mathbb{N}^*, f_{T_i | X_i}(t) = \pi_0(X_i) f_{T_i | X_i, 0}(t) + (1 - \pi_0(X_i)) f_{T_i | X_i, 1}(t)$

Mélange censuré

- Censure : $T^c = T \wedge C$ et $\delta = \mathbb{1}_{\{T \leq C\}}$

Mélange censuré

- Censure : $T^c = T \wedge C$ et $\delta = \mathbb{1}_{\{T \leq C\}}$
- Echantillon de n patients : $(X_1, T_1^c, \delta_1), \dots, (X_n, T_n^c, \delta_n) \in \mathbb{R}^d \times \mathbb{N}^* \times \{0; 1\}$

Mélange censuré

- Censure : $T^c = T \wedge C$ et $\delta = \mathbb{1}_{\{T \leq C\}}$
- Echantillon de n patients : $(X_1, T_1^c, \delta_1), \dots, (X_n, T_n^c, \delta_n) \in \mathbb{R}^d \times \mathbb{N}^* \times \{0; 1\}$
- Alors en notant $\mathbf{T}^c = [T_1^c, \dots, T_n^c]$, $\mathbf{\Delta} = [\delta_1, \dots, \delta_n]$ et $\bar{F} = 1 - F$, la log-vraisemblance du modèle s'écrit

$$\begin{aligned} \ell_n(\theta; \mathbf{T}^c, \mathbf{\Delta}) &= \frac{1}{n} \sum_{i=1}^n \log \left[\{ \pi_{0,\beta}(X_i) f_{T,0}(T_i^c; \alpha_0) + (1 - \pi_{0,\beta}(X_i)) f_{T,1}(T_i^c; \alpha_1) \} \bar{G}(T_i^{c-}) \right]^{\delta_i} \\ &\quad \times \left[\{ \pi_{0,\beta}(X_i) \bar{F}_{T,0}(T_i^{c-}; \alpha_0) + (1 - \pi_{0,\beta}(X_i)) \bar{F}_{T,1}(T_i^{c-}; \alpha_1) \} g(T_i^c) \right]^{1-\delta_i} \end{aligned}$$

- But : estimer $(\alpha_0, \alpha_1, \beta)$

Inférence pour le modèle

- Minimiser $-\ell_n(\theta) + \gamma((1 - \eta)\|\beta\|_1 + \frac{\eta}{2}\|\beta\|_2^2)$

Inférence pour le modèle

- Minimiser $-\ell_n(\theta) + \gamma((1 - \eta)\|\beta\|_1 + \frac{\eta}{2}\|\beta\|_2^2)$
- Algorithme qui s'inspire d'un EM.

Inférence pour le modèle

- Minimiser $-\ell_n(\theta) + \gamma((1 - \eta)\|\beta\|_1 + \frac{\eta}{2}\|\beta\|_2^2)$
- Algorithme qui s'inspire d'un EM.
- Au pas l , on calcule $Q_n(\theta, \theta^{(l)}) = \mathbb{E}_{\theta^{(l)}}[\ell_n^{comp}(\theta; \mathbf{T}^c, \mathbf{\Delta}, \mathbf{Z}) | \mathbf{T}^c, \mathbf{\Delta}]$

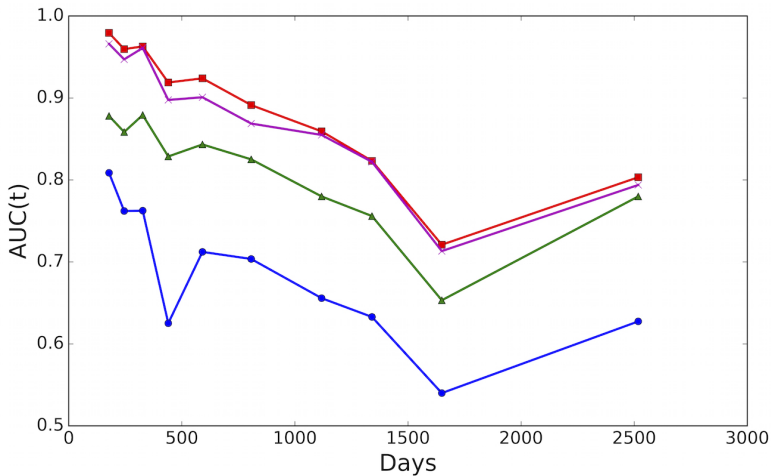
Inférence pour le modèle

- Minimiser $-\ell_n(\theta) + \gamma((1 - \eta)\|\beta\|_1 + \frac{\eta}{2}\|\beta\|_2^2)$
- Algorithme qui s'inspire d'un EM.
- Au pas l , on calcule $Q_n(\theta, \theta^{(l)}) = \mathbb{E}_{\theta^{(l)}}[\ell_n^{comp}(\theta; \mathbf{T}^c, \mathbf{\Delta}, \mathbf{Z}) | \mathbf{T}^c, \mathbf{\Delta}]$
- On maximise la quantité précédente pour mettre à jour les paramètres.

Inférence pour le modèle

- Minimiser $-\ell_n(\theta) + \gamma((1 - \eta)\|\beta\|_1 + \frac{\eta}{2}\|\beta\|_2^2)$
- Algorithme qui s'inspire d'un EM.
- Au pas l , on calcule $Q_n(\theta, \theta^{(l)}) = \mathbb{E}_{\theta^{(l)}}[\ell_n^{comp}(\theta; \mathbf{T}^c, \mathbf{\Delta}, \mathbf{Z}) | \mathbf{T}^c, \mathbf{\Delta}]$
- On maximise la quantité précédente pour mettre à jour les paramètres.
- On utilise l'algorithme L-BGFS-B pour la mise à jour de β .

TCGA Data



Données Red-CVO

Notebook !