



Méthodes de survie et Machine Learning

M2 Données Massives en Santé

le 23/01/2020

Simon Bussy

<https://simonbussy.com>

Postdoc @INSERM, UMRS 1138, Centre de Recherche des Cordeliers
Cofounder, CTO & Head of Research @Califrais

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Programme de la journée

M2 – DMS
Machine learning
2/11

Intitulé du cours :

Méthodes de survie et Machine Learning

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Intitulé du cours :

Méthodes de survie et Machine Learning

Cadre

- ▶ Apprentissage supervisé

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Intitulé du cours :

Méthodes de survie et Machine Learning

Cadre

- ▶ Apprentissage supervisé
- ▶ Analyse de survie

Programme

Rappel

Généralités
La survie

Quelques modèles

Cox PH
C-mix
Binacox

TP python

PBCseq
Consignes

Références

Intitulé du cours :

Méthodes de survie et Machine Learning

Cadre

- ▶ Apprentissage supervisé
- ▶ Analyse de survie
- ▶ Grande dimension

Programme

Rappel

Généralités
La survie

Quelques modèles

Cox PH
C-mix
Binacox

TP python

PBCseq
Consignes

Références

Intitulé du cours :

Méthodes de survie et Machine Learning

Cadre

- ▶ Apprentissage supervisé
- ▶ Analyse de survie
- ▶ Grande dimension

Programme

- I. Petit rappel théorique sur la survie

Programme

Rappel

Généralités
La survie

Quelques modèles

Cox PH
C-mix
Binacox

TP python

PBCseq
Consignes

Références

Intitulé du cours :

Méthodes de survie et Machine Learning

Cadre

- ▶ Apprentissage supervisé
- ▶ Analyse de survie
- ▶ Grande dimension

Programme

- I. Petit rappel théorique sur la survie
- II. Introduction de quelques modèles

Programme

Rappel

Généralités
La survie

Quelques modèles

Cox PH
C-mix
Binacox

TP python

PBCseq
Consignes

Références

Intitulé du cours :

Méthodes de survie et Machine Learning

Cadre

- ▶ Apprentissage supervisé
- ▶ Analyse de survie
- ▶ Grande dimension

Programme

- I. Petit rappel théorique sur la survie
- II. Introduction de quelques modèles
- III. TP sous python sur les données PBCseq...

Programme

Rappel

Généralités
La survie

Quelques modèles

Cox PH
C-mix
Binacox

TP python

PBCseq
Consignes

Références

Intitulé du cours :

Méthodes de survie et Machine Learning

Cadre

- ▶ Apprentissage supervisé
- ▶ Analyse de survie
- ▶ Grande dimension

Programme

- I. Petit rappel théorique sur la survie
- II. Introduction de quelques modèles
- III. TP sous python sur les données PBCseq... en autonomie l'après-midi → questions à envoyer à 17h30 max à simon.bussy@gmail.com (hors bug de code!)

Programme

Rappel

Généralités
La survie

Quelques modèles

Cox PH
C-mix
Binacox

TP python

PBCseq
Consignes

Références

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

I. Petit rappel théorique sur la survie

- ▶ Identifier les facteurs prédictifs

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Généralités

- ▶ Identifier les facteurs prédictifs
- ▶ Problèmes récurrents dans les études cliniques :

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Généralités

- ▶ Identifier les facteurs prédictifs
- ▶ Problèmes récurrents dans les études cliniques :
 - ▶ prédiction du risque

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Généralités

- ▶ Identifier les facteurs prédictifs
- ▶ Problèmes récurrents dans les études cliniques :
 - ▶ prédiction du risque
 - ▶ identification des covariables impliquées

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Généralités

- ▶ Identifier les facteurs prédictifs
- ▶ Problèmes récurrents dans les études cliniques :
 - ▶ prédiction du risque
 - ▶ identification des covariables impliquées

Deux cadres

- ▶ *Binary outcome setting*

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Généralités

- ▶ Identifier les facteurs prédictifs
- ▶ Problèmes récurrents dans les études cliniques :
 - ▶ prédiction du risque
 - ▶ identification des covariables impliquées

Deux cadres

- ▶ *Binary outcome setting*
 - ▶ En se basant sur un seuil ϵ pré-défini

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Généralités

- ▶ Identifier les facteurs prédictifs
- ▶ Problèmes récurrents dans les études cliniques :
 - ▶ prédiction du risque
 - ▶ identification des covariables impliquées

Deux cadres

- ▶ *Binary outcome setting*
 - ▶ En se basant sur un seuil ϵ pré-défini
 - ▶ Résultats très dépendants du choix de ϵ [4]

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Généralités

- ▶ Identifier les facteurs prédictifs
- ▶ Problèmes récurrents dans les études cliniques :
 - ▶ prédiction du risque
 - ▶ identification des covariables impliquées

Deux cadres

- ▶ *Binary outcome setting*
 - ▶ En se basant sur un seuil ϵ pré-défini
 - ▶ Résultats très dépendants du choix de ϵ [4]
- ▶ *Survival analysis setting* : pas de seuil *a priori*

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Généralités

- ▶ Identifier les facteurs prédictifs
- ▶ Problèmes récurrents dans les études cliniques :
 - ▶ prédiction du risque
 - ▶ identification des covariables impliquées

Deux cadres

- ▶ *Binary outcome setting*
 - ▶ En se basant sur un seuil ϵ pré-défini
 - ▶ Résultats très dépendants du choix de ϵ [4]
- ▶ *Survival analysis setting* : pas de seuil *a priori*

Objectif

- ▶ Modèles d'analyse de survie \rightarrow fonctions de survie estimées \rightarrow prédictions binaires pour un ϵ donné

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Généralités

- ▶ Identifier les facteurs prédictifs
- ▶ Problèmes récurrents dans les études cliniques :
 - ▶ prédiction du risque
 - ▶ identification des covariables impliquées

Deux cadres

- ▶ *Binary outcome setting*
 - ▶ En se basant sur un seuil ϵ pré-défini
 - ▶ Résultats très dépendants du choix de ϵ [4]
- ▶ *Survival analysis setting* : pas de seuil *a priori*

Objectif

- ▶ Modèles d'analyse de survie \rightarrow fonctions de survie estimées \rightarrow prédictions binaires pour un ϵ donné
- ▶ Méthodologie pour la création de features pertinentes

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Généralités

- ▶ Identifier les facteurs prédictifs
- ▶ Problèmes récurrents dans les études cliniques :
 - ▶ prédiction du risque
 - ▶ identification des covariables impliquées

Deux cadres

- ▶ *Binary outcome setting*
 - ▶ En se basant sur un seuil ϵ pré-défini
 - ▶ Résultats très dépendants du choix de ϵ [4]
- ▶ *Survival analysis setting* : pas de seuil *a priori*

Objectif

- ▶ Modèles d'analyse de survie \rightarrow fonctions de survie estimées \rightarrow prédictions binaires pour un ϵ donné
- ▶ Méthodologie pour la création de features pertinentes
- ▶ Comparer des méthodes d'apprentissage issues des deux cadres

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

La survie

La fonction de survie

- ▶ $T \geq 0$: v.a. durée de survie ; densité f , f.d.r. F

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

La survie

La fonction de survie

- ▶ $T \geq 0$: v.a. durée de survie ; densité f , f.d.r. F
- ▶ Survie $S(t) = \mathbb{P}[T > t] = 1 - F(t) = \int_t^{+\infty} f(u)du$

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

La survie

La fonction de survie

- ▶ $T \geq 0$: v.a. durée de survie ; densité f , f.d.r. F
- ▶ Survie $S(t) = \mathbb{P}[T > t] = 1 - F(t) = \int_t^{+\infty} f(u)du$

Le risque instantané

- ▶ $\lambda(t) = \lim_{dt \rightarrow 0} \frac{\mathbb{P}[t \leq T < t + dt | T \geq t]}{dt}$

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

La survie

La fonction de survie

- ▶ $T \geq 0$: v.a. durée de survie ; densité f , f.d.r. F
- ▶ Survie $S(t) = \mathbb{P}[T > t] = 1 - F(t) = \int_t^{+\infty} f(u)du$

Le risque instantané

- ▶ $\lambda(t) = \lim_{dt \rightarrow 0} \frac{\mathbb{P}[t \leq T < t + dt | T \geq t]}{dt}$
- ▶ $\lambda(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} \log S(t)$

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

La survie

La fonction de survie

- ▶ $T \geq 0$: v.a. durée de survie ; densité f , f.d.r. F
- ▶ Survie $S(t) = \mathbb{P}[T > t] = 1 - F(t) = \int_t^{+\infty} f(u)du$

Le risque instantané

- ▶ $\lambda(t) = \lim_{dt \rightarrow 0} \frac{\mathbb{P}[t \leq T < t + dt | T \geq t]}{dt}$
- ▶ $\lambda(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} \log S(t)$
- ▶ $S(t) = \exp\{-\int_0^t \lambda(u)du\} = \exp\{-\Lambda(t)\}$ avec $S(0) = 1$ et $\Lambda(t)$ le risque cumulé entre l'instant initial et t

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

La survie

La fonction de survie

- ▶ $T \geq 0$: v.a. durée de survie ; densité f , f.d.r. F
- ▶ Survie $S(t) = \mathbb{P}[T > t] = 1 - F(t) = \int_t^{+\infty} f(u)du$

Le risque instantané

- ▶ $\lambda(t) = \lim_{dt \rightarrow 0} \frac{\mathbb{P}[t \leq T < t + dt | T \geq t]}{dt}$
- ▶ $\lambda(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} \log S(t)$
- ▶ $S(t) = \exp\{-\int_0^t \lambda(u)du\} = \exp\{-\Lambda(t)\}$ avec $S(0) = 1$ et $\Lambda(t)$ le risque cumulé entre l'instant initial et t

La notion de censure

- ▶ Labels: $Y_i = \min\{T_i, C_i\}$ et $\Delta_i = \mathbb{1}_{\{T_i \leq C_i\}}$

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

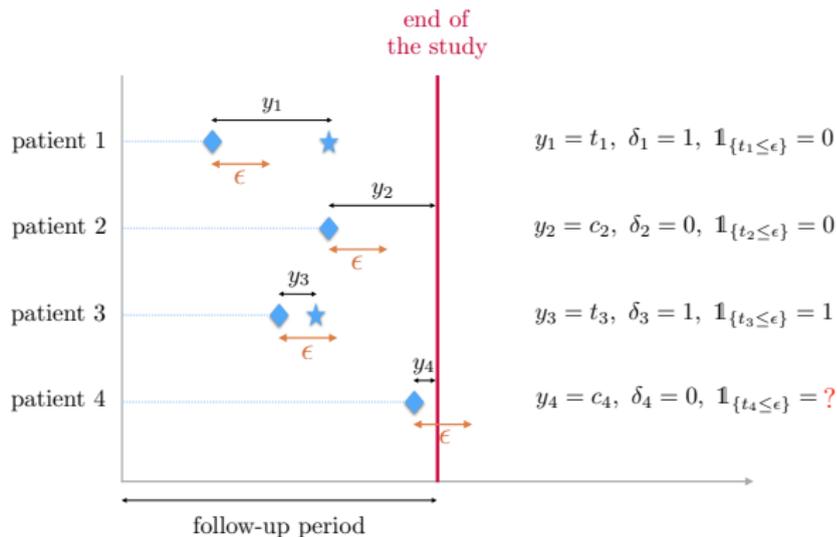
Consignes

Références

La survie

La notion de censure

- Labels: $Y_i = \min\{T_i, C_i\}$ et $\Delta_i = \mathbb{1}_{\{T_i \leq C_i\}}$



Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

La survie

La fonction de survie

- ▶ $T \geq 0$: v.a. durée de survie ; densité f , f.d.r. F
- ▶ Survie $S(t) = \mathbb{P}[T > t] = 1 - F(t) = \int_t^{+\infty} f(u)du$

Le risque instantané

- ▶ $\lambda(t) = \lim_{dt \rightarrow 0} \frac{\mathbb{P}[t \leq T < t + dt | T \geq t]}{dt}$
- ▶ $\lambda(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} \log S(t)$
- ▶ $S(t) = \exp\{-\int_0^t \lambda(u)du\} = \exp\{-\Lambda(t)\}$ avec $S(0) = 1$ et $\Lambda(t)$ le risque cumulé entre l'instant initial et t

La notion de censure

- ▶ Labels: $Y_i = \min\{T_i, C_i\}$ et $\Delta_i = \mathbb{1}_{\{T_i \leq C_i\}}$
- ▶ Ecrire la vraisemblance dans ce contexte! voir Bussy [1]

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

II. Introduction de quelques modèles

► $\lambda_i(t|X_i = x_i) = \lambda_0^*(t) \exp(f^*(x_i))$

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

- ▶ $\lambda_i(t|X_i = x_i) = \lambda_0^*(t) \exp(f^*(x_i))$
- ▶ Fonctions de risque proportionnelles :

$$\frac{\lambda_i(t|X_i = x_i)}{\lambda_j(t|X_j = x_j)} = \exp(f^*(x_i) - f^*(x_j))$$

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

- ▶ $\lambda_i(t|X_i = x_i) = \lambda_0^*(t) \exp(f^*(x_i))$
- ▶ Fonctions de risque proportionnelles :

$$\frac{\lambda_i(t|X_i = x_i)}{\lambda_j(t|X_j = x_j)} = \exp(f^*(x_i) - f^*(x_j))$$

- ▶ En général, $f^*(x_i) = x_i^\top \beta^*$ avec $\beta^* \in \mathbb{R}^d$

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

- ▶ $\lambda_i(t|X_i = x_i) = \lambda_0^*(t) \exp(f^*(x_i))$
- ▶ Fonctions de risque proportionnelles :

$$\frac{\lambda_i(t|X_i = x_i)}{\lambda_j(t|X_j = x_j)} = \exp(f^*(x_i) - f^*(x_j))$$

- ▶ En général, $f^*(x_i) = x_i^\top \beta^*$ avec $\beta^* \in \mathbb{R}^d$
- ▶ Log-vraisemblance partielle :

$$\ell_n(\beta) = \frac{1}{n} \sum_{i=1}^n \delta_i (x_i^\top \beta - \log \sum_{i': z_{i'} \geq z_i} \exp(x_{i'}^\top \beta))$$

- ▶ $\lambda_i(t|X_i = x_i) = \lambda_0^*(t) \exp(f^*(x_i))$
- ▶ Fonctions de risque proportionnelles :

$$\frac{\lambda_i(t|X_i = x_i)}{\lambda_j(t|X_j = x_j)} = \exp(f^*(x_i) - f^*(x_j))$$

- ▶ En général, $f^*(x_i) = x_i^\top \beta^*$ avec $\beta^* \in \mathbb{R}^d$
- ▶ Log-vraisemblance partielle :

$$\ell_n(\beta) = \frac{1}{n} \sum_{i=1}^n \delta_i (x_i^\top \beta - \log \sum_{i': z_{i'} \geq z_i} \exp(x_{i'}^\top \beta))$$

- ▶ $\hat{\beta} \in \operatorname{argmin}_{\beta \in \mathbb{R}^d} -\ell_n(\beta)$

Le modèle du C-mix (voir Bussy et al. [2])

- Contexte de l'analyse de survie

$$Y = \min(T, C) \quad \text{et} \quad \Delta = \mathbb{1}_{\{T \leq C\}}$$

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Le modèle du C-mix (voir Bussy et al. [2])

- ▶ Contexte de l'analyse de survie

$$Y = \min(T, C) \quad \text{et} \quad \Delta = \mathbb{1}_{\{T \leq C\}}$$

- ▶ Variable latente $Z \in \{0, \dots, K - 1\}$

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Le modèle du C-mix (voir Bussy et al. [2])

- ▶ Contexte de l'analyse de survie

$$Y = \min(T, C) \quad \text{et} \quad \Delta = \mathbb{1}_{\{T \leq C\}}$$

- ▶ Variable latente $Z \in \{0, \dots, K-1\}$
- ▶ Modèle de mélange $f(t|X = x) = \sum_{k=0}^{K-1} \pi_{\beta_k}(x) f_k(t; \alpha_k)$

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Le modèle du C-mix (voir Bussy et al. [2])

- ▶ Contexte de l'analyse de survie

$$Y = \min(T, C) \quad \text{et} \quad \Delta = \mathbb{1}_{\{T \leq C\}}$$

- ▶ Variable latente $Z \in \{0, \dots, K-1\}$

- ▶ Modèle de mélange $f(t|X = x) = \sum_{k=0}^{K-1} \pi_{\beta_k}(x) f_k(t; \alpha_k)$

- ▶ $\pi_{\beta_k}(x) = \mathbb{P}[Z = k|X = x] = \frac{e^{x^\top \beta_k}}{\sum_{k=0}^{K-1} e^{x^\top \beta_k}}$ (softmax)

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Le modèle du C-mix (voir Bussy et al. [2])

- ▶ Contexte de l'analyse de survie

$$Y = \min(T, C) \quad \text{et} \quad \Delta = \mathbb{1}_{\{T \leq C\}}$$

- ▶ Variable latente $Z \in \{0, \dots, K-1\}$
- ▶ Modèle de mélange $f(t|X = x) = \sum_{k=0}^{K-1} \pi_{\beta_k}(x) f_k(t; \alpha_k)$
- ▶ $\pi_{\beta_k}(x) = \mathbb{P}[Z = k|X = x] = \frac{e^{x^\top \beta_k}}{\sum_{k=0}^{K-1} e^{x^\top \beta_k}}$ (softmax)
- ▶ $Z|X = x \sim \mathcal{M}(\pi_{\beta_0}(x), \dots, \pi_{\beta_{K-1}}(x))$

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Le modèle du C-mix (voir Bussy et al. [2])

- ▶ Contexte de l'analyse de survie

$$Y = \min(T, C) \quad \text{et} \quad \Delta = \mathbb{1}_{\{T \leq C\}}$$

- ▶ Variable latente $Z \in \{0, \dots, K-1\}$
- ▶ Modèle de mélange $f(t|X = x) = \sum_{k=0}^{K-1} \pi_{\beta_k}(x) f_k(t; \alpha_k)$
- ▶ $\pi_{\beta_k}(x) = \mathbb{P}[Z = k|X = x] = \frac{e^{x^\top \beta_k}}{\sum_{k=0}^{K-1} e^{x^\top \beta_k}}$ (softmax)
- ▶ $Z|X = x \sim \mathcal{M}(\pi_{\beta_0}(x), \dots, \pi_{\beta_{K-1}}(x))$
- ▶ Hypothèses : (i) $T|Z, X \perp\!\!\!\perp C|Z, X$ [5] (ii) $C \perp\!\!\!\perp Z$ [6]

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Le modèle du C-mix (voir Bussy et al. [2])

- ▶ Contexte de l'analyse de survie

$$Y = \min(T, C) \quad \text{et} \quad \Delta = \mathbb{1}_{\{T \leq C\}}$$

- ▶ Variable latente $Z \in \{0, \dots, K-1\}$
- ▶ Modèle de mélange $f(t|X = x) = \sum_{k=0}^{K-1} \pi_{\beta_k}(x) f_k(t; \alpha_k)$
- ▶ $\pi_{\beta_k}(x) = \mathbb{P}[Z = k|X = x] = \frac{e^{x^\top \beta_k}}{\sum_{k=0}^{K-1} e^{x^\top \beta_k}}$ (softmax)
- ▶ $Z|X = x \sim \mathcal{M}(\pi_{\beta_0}(x), \dots, \pi_{\beta_{K-1}}(x))$
- ▶ Hypothèses : (i) $T|Z, X \perp\!\!\!\perp C|Z, X$ [5] (ii) $C \perp\!\!\!\perp Z$ [6]
- ▶ Échantillon *i.i.d.* $(x_1, y_1, \delta_1), \dots, (x_n, y_n, \delta_n) \in \mathbb{R}^d \times \mathbb{R}_+ \times \{0, 1\}$

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Le modèle du C-mix (voir Bussy et al. [2])

- ▶ Contexte de l'analyse de survie

$$Y = \min(T, C) \quad \text{et} \quad \Delta = \mathbb{1}_{\{T \leq C\}}$$

- ▶ Variable latente $Z \in \{0, \dots, K-1\}$
- ▶ Modèle de mélange $f(t|X = x) = \sum_{k=0}^{K-1} \pi_{\beta_k}(x) f_k(t; \alpha_k)$
- ▶ $\pi_{\beta_k}(x) = \mathbb{P}[Z = k|X = x] = \frac{e^{x^\top \beta_k}}{\sum_{k=0}^{K-1} e^{x^\top \beta_k}}$ (softmax)
- ▶ $Z|X = x \sim \mathcal{M}(\pi_{\beta_0}(x), \dots, \pi_{\beta_{K-1}}(x))$
- ▶ Hypothèses : (i) $T|Z, X \perp\!\!\!\perp C|Z, X$ [5] (ii) $C \perp\!\!\!\perp Z$ [6]
- ▶ Échantillon *i.i.d.* $(x_1, y_1, \delta_1), \dots, (x_n, y_n, \delta_n) \in \mathbb{R}^d \times \mathbb{R}_+ \times \{0, 1\}$
- ▶ $\theta = (\alpha_0, \dots, \alpha_K, \beta_0, \dots, \beta_K)^\top$, log-vraisemblance du C-mix

$$\begin{aligned} \ell_n(\theta) = n^{-1} \sum_{i=1}^n \left\{ \delta_i \log \left[\bar{G}(y_i^-) \sum_{k=0}^{K-1} \pi_{\beta_k}(x_i) f_k(y_i; \alpha_k) \right] \right. \\ \left. + (1 - \delta_i) \log \left[g(y_i) \sum_{k=0}^{K-1} \pi_{\beta_k}(x_i) \bar{F}_k(y_i^-; \alpha_k) \right] \right\} \end{aligned}$$

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

- ▶ Un algorithme QNEM pour l'inférence

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

- ▶ Un algorithme QNEM pour l'inférence
- ▶ Meilleures performances que les modèles CURE et Cox PH : en prédiction, sélection de variable, robustesse, temps de calcul, interprétabilité

Programme

Rappel

Généralités
La survie

Quelques modèles

Cox PH
C-mix
Binacox

TP python

PBCseq
Consignes

Références

Programme

Rappel

Généralités
La survie

Quelques modèles

Cox PH
C-mix
Binacox

TP python

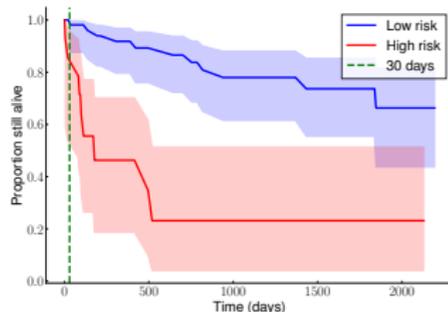
PBCseq
Consignes

Références

- ▶ Un algorithme QNEM pour l'inférence
- ▶ Meilleures performances que les modèles CURE et Cox PH : en prédiction, sélection de variable, robustesse, temps de calcul, interprétabilité
- ▶ Détection de sous-groupes de patients relativement à leurs risques

Le C-mix en pratique

- ▶ Un algorithme QNEM pour l'inférence
- ▶ Meilleures performances que les modèles CURE et Cox PH : en prédiction, sélection de variable, robustesse, temps de calcul, interprétabilité
- ▶ Détection de sous-groupes de patients relativement à leurs risques



(a) Courbes de survie

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

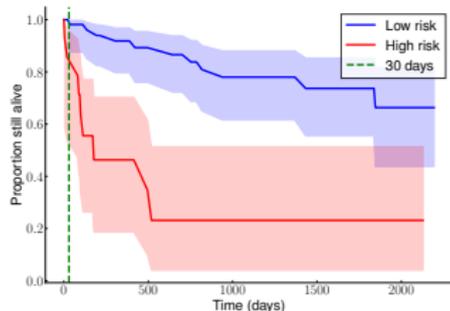
PBCseq

Consignes

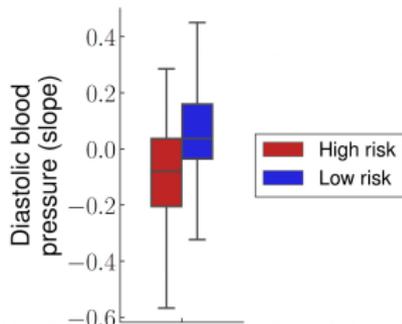
Références

Le C-mix en pratique

- ▶ Un algorithme QNEM pour l'inférence
- ▶ Meilleures performances que les modèles CURE et Cox PH : en prédiction, sélection de variable, robustesse, temps de calcul, interprétabilité
- ▶ Détection de sous-groupes de patients relativement à leurs risques



(a) Courbes de survie



(b) Boxplot

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Le C-mix en pratique

- ▶ Un algorithme QNEM pour l'inférence
- ▶ Meilleures performances que les modèles CURE et Cox PH : en prédiction, sélection de variable, robustesse, temps de calcul, interprétabilité
- ▶ Détection de sous-groupes de patients relativement à leurs risques

Code Python

- ▶ Disponible à <https://github.com/SimonBussy/C-mix>

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Le C-mix en pratique

- ▶ Un algorithme QNEM pour l'inférence
- ▶ Meilleures performances que les modèles CURE et Cox PH : en prédiction, sélection de variable, robustesse, temps de calcul, interprétabilité
- ▶ Détection de sous-groupes de patients relativement à leurs risques

Code Python

- ▶ Disponible à <https://github.com/SimonBussy/C-mix>
- ▶ Programmes annotés, notebooks et tutoriels

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Code Python

- ▶ Disponible à <https://github.com/SimonBussy/C-mix>
- ▶ Programmes annotés, notebooks et tutoriels

```
1 from QNEM.inference import QNEM
2 from QNEM.simulation import CensoredGeomMixtureRegression
3
4 # Generate data
5 simu = CensoredGeomMixtureRegression(n_samples=1000, n_features=100,
6                                     n_active_features=30)
7 X, Y, delta = simu.simulate()
8
9 # Choose between C-mix or CURE model
10 model = 'C-mix'
11
12 # Fit the model with a penalty strength equal to 10
13 learner = QNEM(model=model, l_elastic_net=10, eta=.1, max_iter=100,
14               tol=1e-6, warm_start=True, fit_intercept=True)
15 learner.fit(X, Y, delta)
16
17 # Obtain the estimated marker
18 coeffs = learner.coefs
19 marker = QNEM.predict_proba(X, fit_intercept, coeffs)
```

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Binacox : le contexte

- ▶ Biomarqueurs → décision clinique : choix de seuils

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Binacox : le contexte

- ▶ Biomarqueurs → décision clinique : choix de seuils
- ▶ Méthodes actuelles basées sur les tests multiples (MT) [7]

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Binacox : le contexte

- ▶ Biomarqueurs → décision clinique : choix de seuils
- ▶ Méthodes actuelles basées sur les tests multiples (MT) [7]
- ▶ Binacox : méthode pronostique pour

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Binacox : le contexte

- ▶ Biomarqueurs → décision clinique : choix de seuils
- ▶ Méthodes actuelles basées sur les tests multiples (MT) [7]
- ▶ Binacox : méthode pronostique pour
 - ▶ détecter de **multiples** seuils par covariable continue,

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Binacox : le contexte

- ▶ Biomarqueurs → décision clinique : choix de seuils
- ▶ Méthodes actuelles basées sur les tests multiples (MT) [7]
- ▶ Binacox : méthode pronostique pour
 - ▶ détecter de **multiples** seuils par covariable continue,
 - ▶ de façon **multivariée**,

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Binacox : le contexte

- ▶ Biomarqueurs → décision clinique : choix de seuils
- ▶ Méthodes actuelles basées sur les tests multiples (MT) [7]
- ▶ Binacox : méthode pronostique pour
 - ▶ détecter de **multiples** seuils par covariable continue,
 - ▶ de façon **multivariée**,
 - ▶ dans un contexte de **grande dimension**

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Binacox : le contexte

- ▶ Biomarqueurs → décision clinique : choix de seuils
- ▶ Méthodes actuelles basées sur les tests multiples (MT) [7]
- ▶ Binacox : méthode pronostique pour
 - ▶ détecter de **multiples** seuils par covariable continue,
 - ▶ de façon **multivariée**,
 - ▶ dans un contexte de **grande dimension**
- ▶ $Z = T \wedge C$, $\Delta = \mathbb{1}(T \leq C)$ et $X \in \mathbb{R}^p$

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Binacox : le contexte

- ▶ Biomarqueurs → décision clinique : choix de seuils
- ▶ Méthodes actuelles basées sur les tests multiples (MT) [7]
- ▶ Binacox : méthode pronostique pour
 - ▶ détecter de **multiples** seuils par covariable continue,
 - ▶ de façon **multivariée**,
 - ▶ dans un contexte de **grande dimension**
- ▶ $Z = T \wedge C$, $\Delta = \mathbb{1}(T \leq C)$ et $X \in \mathbb{R}^p$
- ▶ Risque instantané pour un patient i donné par

$$\lambda^*(t|X_i = x_i) = \lambda_0^*(t) \exp \left\{ \underbrace{\sum_{j=1}^p \sum_{k=1}^{K_j^*+1} \beta_{j,k}^* \mathbb{1}(x_{i,j} \in I_{j,k}^*)}_{f^*(x_i)} \right\}$$

où $I_{j,k}^* = (\mu_{j,k-1}^*, \mu_{j,k}^*]$ pour $k \in \{1, \dots, K_j^* + 1\}$

[Programme](#)[Rappel](#)[Généralités](#)[La survie](#)[Quelques modèles](#)[Cox PH](#)[C-mix](#)[Binacox](#)[TP python](#)[PBCseq](#)[Consignes](#)[Références](#)

Binacox : le contexte

- ▶ Biomarqueurs → décision clinique : choix de seuils
- ▶ Méthodes actuelles basées sur les tests multiples (MT) [7]
- ▶ Binacox : méthode pronostique pour
 - ▶ détecter de **multiples** seuils par covariable continue,
 - ▶ de façon **multivariée**,
 - ▶ dans un contexte de **grande dimension**
- ▶ $Z = T \wedge C$, $\Delta = \mathbb{1}(T \leq C)$ et $X \in \mathbb{R}^p$
- ▶ Risque instantané pour un patient i donné par

$$\lambda^*(t|X_i = x_i) = \lambda_0^*(t) \exp \left\{ \underbrace{\sum_{j=1}^p \sum_{k=1}^{K_j^*+1} \beta_{j,k}^* \mathbb{1}(x_{i,j} \in I_{j,k}^*)}_{f^*(x_i)} \right\}$$

où $I_{j,k}^* = (\mu_{j,k-1}^*, \mu_{j,k}^*]$ pour $k \in \{1, \dots, K_j^* + 1\}$

- ▶ But : estimer simultanément

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Binacox : le contexte

- ▶ Biomarqueurs → décision clinique : choix de seuils
- ▶ Méthodes actuelles basées sur les tests multiples (MT) [7]
- ▶ Binacox : méthode pronostique pour
 - ▶ détecter de **multiples** seuils par covariable continue,
 - ▶ de façon **multivariée**,
 - ▶ dans un contexte de **grande dimension**
- ▶ $Z = T \wedge C$, $\Delta = \mathbb{1}(T \leq C)$ et $X \in \mathbb{R}^p$
- ▶ Risque instantané pour un patient i donné par

$$\lambda^*(t|X_i = x_i) = \lambda_0^*(t) \exp \left\{ \underbrace{\sum_{j=1}^p \sum_{k=1}^{K_j^*+1} \beta_{j,k}^* \mathbb{1}(x_{i,j} \in I_{j,k}^*)}_{f^*(x_i)} \right\}$$

où $I_{j,k}^* = (\mu_{j,k-1}^*, \mu_{j,k}^*]$ pour $k \in \{1, \dots, K_j^* + 1\}$

- ▶ But : estimer simultanément
 - ▶ $\mu^* = (\mu_{1,1}^*, \dots, \mu_{1,K_1^*}^*, \dots, \mu_{p,1}^*, \dots, \mu_{p,K_p^*}^*)^\top \in \mathbb{R}^{K^*}$

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Binacox : le contexte

- ▶ Biomarqueurs → décision clinique : choix de seuils
- ▶ Méthodes actuelles basées sur les tests multiples (MT) [7]
- ▶ Binacox : méthode pronostique pour
 - ▶ détecter de **multiples** seuils par covariable continue,
 - ▶ de façon **multivariée**,
 - ▶ dans un contexte de **grande dimension**
- ▶ $Z = T \wedge C$, $\Delta = \mathbb{1}(T \leq C)$ et $X \in \mathbb{R}^p$
- ▶ Risque instantané pour un patient i donné par

$$\lambda^*(t|X_i = x_i) = \lambda_0^*(t) \exp \left\{ \underbrace{\sum_{j=1}^p \sum_{k=1}^{K_j^*+1} \beta_{j,k}^* \mathbb{1}(x_{i,j} \in I_{j,k}^*)}_{f^*(x_i)} \right\}$$

où $I_{j,k}^* = (\mu_{j,k-1}^*, \mu_{j,k}^*]$ pour $k \in \{1, \dots, K_j^* + 1\}$

- ▶ But : estimer simultanément
 - ▶ $\mu^* = (\mu_{1,1}^*, \dots, \mu_{1,K_1^*}^*, \dots, \mu_{p,1}^*, \dots, \mu_{p,K_p^*}^*)^\top \in \mathbb{R}^{K^*}$
 - ▶ $\beta^* = (\beta_{1,1}^*, \dots, \beta_{1,K_1^*+1}^*, \dots, \beta_{p,1}^*, \dots, \beta_{p,K_p^*+1}^*)^\top \in \mathbb{R}^{K^*+p}$

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Binacox : le contexte

- ▶ Biomarqueurs → décision clinique : choix de seuils
- ▶ Méthodes actuelles basées sur les tests multiples (MT) [7]
- ▶ Binacox : méthode pronostique pour
 - ▶ détecter de **multiples** seuils par covariable continue,
 - ▶ de façon **multivariée**,
 - ▶ dans un contexte de **grande dimension**
- ▶ $Z = T \wedge C$, $\Delta = \mathbb{1}(T \leq C)$ et $X \in \mathbb{R}^p$
- ▶ Risque instantané pour un patient i donné par

$$\lambda^*(t|X_i = x_i) = \lambda_0^*(t) \exp \left\{ \underbrace{\sum_{j=1}^p \sum_{k=1}^{K_j^*+1} \beta_{j,k}^* \mathbb{1}(x_{i,j} \in I_{j,k}^*)}_{f^*(x_i)} \right\}$$

où $I_{j,k}^* = (\mu_{j,k-1}^*, \mu_{j,k}^*]$ pour $k \in \{1, \dots, K_j^* + 1\}$

- ▶ But : estimer simultanément
 - ▶ $\mu^* = (\mu_{1,1}^*, \dots, \mu_{1,K_1^*}^*, \dots, \mu_{p,1}^*, \dots, \mu_{p,K_p^*}^*)^\top \in \mathbb{R}^{K^*}$
 - ▶ $\beta^* = (\beta_{1,1}^*, \dots, \beta_{1,K_1^*+1}^*, \dots, \beta_{p,1}^*, \dots, \beta_{p,K_p^*+1}^*)^\top \in \mathbb{R}^{K^*+p}$
 - ▶ $K^* = \sum_{j=1}^p K_j^*$

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

- ▶ Binarisation par encodage “one-hot” [8] $x_i^B \in \mathbb{R}^{p+d}$

- Binarisation par encodage “one-hot” [8] $x_i^B \in \mathbb{R}^{p+d}$

$x_i = (x_{i,1}, \dots, x_{i,p})^\top \in \mathbb{R}^p$ transformé en

$x_i^B = (x_{i,1,1}^B, \dots, x_{i,1,d_1+1}^B, x_{i,2,1}^B, \dots, x_{i,2,d_2+1}^B, \dots, x_{i,p,1}^B, \dots, x_{i,p,d_p+1}^B)^\top \in \mathbb{R}^{p+d}$,

t.q. pour $i = 1, \dots, n$ et $l = 1, \dots, d_j + 1$, on a

$$x_{i,j,l}^B = \begin{cases} 1 & \text{si } x_{i,j} \in I_{j,l} = (\mu_{j,l-1}, \mu_{j,l}] \text{ avec } \mu_{j,l} = q_j(l/d_j + 1) \\ 0 & \text{sinon} \end{cases}$$

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Binacox : la méthode

- Binarisation par encodage “one-hot” [8] $x_i^B \in \mathbb{R}^{p+d}$

$x_i = (x_{i,1}, \dots, x_{i,p})^\top \in \mathbb{R}^p$ transformé en

$$x_i^B = (x_{i,1,1}^B, \dots, x_{i,1,d_1+1}^B, x_{i,2,1}^B, \dots, x_{i,2,d_2+1}^B, \dots, x_{i,p,1}^B, \dots, x_{i,p,d_p+1}^B)^\top \in \mathbb{R}^{p+d},$$

t.q. pour $i = 1, \dots, n$ et $l = 1, \dots, d_j + 1$, on a

$$x_{i,j,l}^B = \begin{cases} 1 & \text{si } x_{i,j} \in I_{j,l} = (\mu_{j,l-1}, \mu_{j,l}] \text{ avec } \mu_{j,l} = q_j(l/d_j + 1) \\ 0 & \text{sinon} \end{cases}$$

$X_{\bullet,1}$

⋮

12,5

9,2

3,1

8,7

⋮

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Binacox : la méthode

- Binarisation par encodage “one-hot” [8] $x_i^B \in \mathbb{R}^{p+d}$

$x_i = (x_{i,1}, \dots, x_{i,p})^\top \in \mathbb{R}^p$ transformé en

$$x_i^B = (x_{i,1,1}^B, \dots, x_{i,1,d_1+1}^B, x_{i,2,1}^B, \dots, x_{i,2,d_2+1}^B, \dots, x_{i,p,1}^B, \dots, x_{i,p,d_p+1}^B)^\top \in \mathbb{R}^{p+d},$$

t.q. pour $i = 1, \dots, n$ et $l = 1, \dots, d_j + 1$, on a

$$x_{i,j,l}^B = \begin{cases} 1 & \text{si } x_{i,j} \in I_{j,l} = (\mu_{j,l-1}, \mu_{j,l}] \text{ avec } \mu_{j,l} = q_j(l/d_j + 1) \\ 0 & \text{sinon} \end{cases}$$

$$\begin{array}{l} \xrightarrow{d_1 = 4} \\ X_{\bullet,1} \quad q_1(0.25) = 7,9 \quad q_1(0.5) = 9,6 \quad q_1(0.75) = 11,7 \\ \vdots \\ 12,5 \\ 9,2 \\ 3,1 \\ 8,7 \\ \vdots \end{array}$$

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Binacox : la méthode

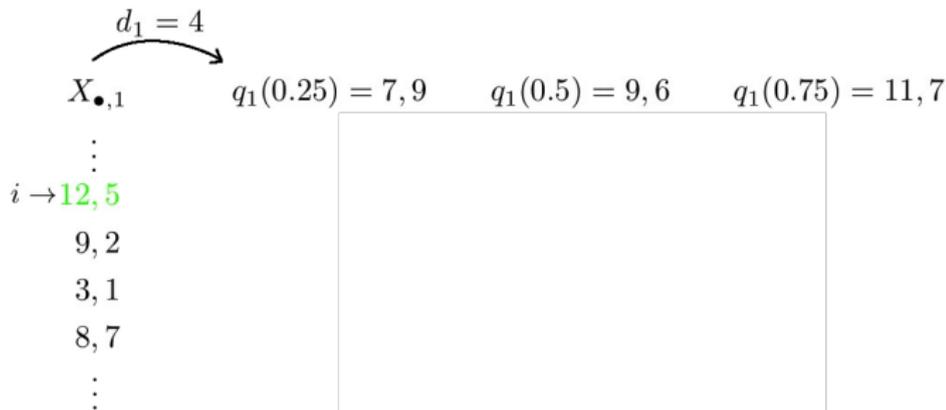
- Binarisation par encodage “one-hot” [8] $x_i^B \in \mathbb{R}^{p+d}$

$x_i = (x_{i,1}, \dots, x_{i,p})^\top \in \mathbb{R}^p$ transformé en

$$x_i^B = (x_{i,1,1}^B, \dots, x_{i,1,d_1+1}^B, x_{i,2,1}^B, \dots, x_{i,2,d_2+1}^B, \dots, x_{i,p,1}^B, \dots, x_{i,p,d_p+1}^B)^\top \in \mathbb{R}^{p+d},$$

t.q. pour $i = 1, \dots, n$ et $l = 1, \dots, d_j + 1$, on a

$$x_{i,j,l}^B = \begin{cases} 1 & \text{si } x_{i,j} \in I_{j,l} = (\mu_{j,l-1}, \mu_{j,l}] \text{ avec } \mu_{j,l} = q_j(l/d_j + 1) \\ 0 & \text{sinon} \end{cases}$$



Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Binacox : la méthode

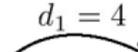
- Binarisation par encodage “one-hot” [8] $x_i^B \in \mathbb{R}^{p+d}$

$x_i = (x_{i,1}, \dots, x_{i,p})^\top \in \mathbb{R}^p$ transformé en

$$x_i^B = (x_{i,1,1}^B, \dots, x_{i,1,d_1+1}^B, x_{i,2,1}^B, \dots, x_{i,2,d_2+1}^B, \dots, x_{i,p,1}^B, \dots, x_{i,p,d_p+1}^B)^\top \in \mathbb{R}^{p+d},$$

t.q. pour $i = 1, \dots, n$ et $l = 1, \dots, d_j + 1$, on a

$$x_{i,j,l}^B = \begin{cases} 1 & \text{si } x_{i,j} \in I_{j,l} = (\mu_{j,l-1}, \mu_{j,l}] \text{ avec } \mu_{j,l} = q_j(l/d_j + 1) \\ 0 & \text{sinon} \end{cases}$$

	$d_1 = 4$				
					
	$X_{\bullet,1}$	$q_1(0.25) = 7,9$	$q_1(0.5) = 9,6$	$q_1(0.75) = 11,7$	
$i \rightarrow$	$12,5$	0	0	0	1
	9,2				
	3,1				
	8,7				
	⋮				

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Binacox : la méthode

- Binarisation par encodage “one-hot” [8] $x_i^B \in \mathbb{R}^{p+d}$

$x_i = (x_{i,1}, \dots, x_{i,p})^\top \in \mathbb{R}^p$ transformé en

$x_i^B = (x_{i,1,1}^B, \dots, x_{i,1,d_1+1}^B, x_{i,2,1}^B, \dots, x_{i,2,d_2+1}^B, \dots, x_{i,p,1}^B, \dots, x_{i,p,d_p+1}^B)^\top \in \mathbb{R}^{p+d}$,

t.q. pour $i = 1, \dots, n$ et $l = 1, \dots, d_j + 1$, on a

$$x_{i,j,l}^B = \begin{cases} 1 & \text{si } x_{i,j} \in I_{j,l} = (\mu_{j,l-1}, \mu_{j,l}] \text{ avec } \mu_{j,l} = q_j(l/d_j + 1) \\ 0 & \text{sinon} \end{cases}$$

	$d_1 = 4$				
					
	$X_{\bullet,1}$	$q_1(0.25) = 7,9$	$q_1(0.5) = 9,6$	$q_1(0.75) = 11,7$	
$i \rightarrow$	\vdots	\vdots	\vdots	\vdots	\vdots
$12,5$	\rightarrow	0	0	0	1
$9,2$	\rightarrow	0	1	0	0
$3,1$	\rightarrow	1	0	0	0
$8,7$	\rightarrow	0	1	0	0
\vdots		\vdots	\vdots	\vdots	\vdots

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

- ▶ Binarisation par encodage “one-hot” [8] $x_i^B \in \mathbb{R}^{p+d}$
- ▶ $f_\beta : x_i \mapsto \beta^\top x_i^B = \sum_{j=1}^p f_{\beta_{j,\bullet}}(x_i)$ pour estimer f^*

$$\text{où } f_{\beta_{j,\bullet}}(x_i) = \sum_{l=1}^{d_j+1} \beta_{j,l} \mathbb{1}(x_{i,j} \in I_{j,l})$$

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

- ▶ Binarisation par encodage “one-hot” [8] $x_i^B \in \mathbb{R}^{p+d}$
- ▶ $f_\beta : x_i \mapsto \beta^\top x_i^B = \sum_{j=1}^p f_{\beta_j, \bullet}(x_i)$ pour estimer f^*
- ▶ log-vraisemblance négative partielle $\ell_n(f_\beta)$

$$\ell_n(f_\beta) = -\frac{1}{n} \sum_{i=1}^n \delta_i \left\{ f_\beta(x_i) - \log \sum_{i': z_{i'} \geq z_i} e^{f_\beta(x_{i'})} \right\}$$

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Binacox : la méthode

- ▶ Binarisation par encodage “one-hot” [8] $x_i^B \in \mathbb{R}^{p+d}$
- ▶ $f_\beta : x_i \mapsto \beta^\top x_i^B = \sum_{j=1}^p f_{\beta_j, \bullet}(x_i)$ pour estimer f^*
- ▶ log-vraisemblance négative partielle $\ell_n(f_\beta)$
- ▶ $\hat{\beta} \in \operatorname{argmin}_{\beta \in \mathcal{B}_{p+d}(\mathbb{R})} \{ \ell_n(f_\beta) + \operatorname{bina}(\beta) \}$, avec

$$\operatorname{bina}(\beta) = \sum_{j=1}^p \left(\sum_{l=2}^{d_j+1} \omega_{j,l} |\beta_{j,l} - \beta_{j,l-1}| + \delta_1(\beta_{j,\bullet}) \right)$$

et

$$\delta_1(u) = \begin{cases} 0 & \text{si } \mathbf{1}^\top u = 0, \\ \infty & \text{sinon,} \end{cases}$$

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Programme

Rappel

Généralités
La survie

Quelques modèles

Cox PH
C-mix
Binacox

TP python

PBCseq
Consignes

Références

- ▶ Binarisation par encodage “one-hot” [8] $x_i^B \in \mathbb{R}^{p+d}$
- ▶ $f_\beta : x_i \mapsto \beta^\top x_i^B = \sum_{j=1}^p f_{\beta_j, \bullet}(x_i)$ pour estimer f^*
- ▶ log-vraisemblance négative partielle $\ell_n(f_\beta)$
- ▶ $\hat{\beta} \in \operatorname{argmin}_{\beta \in \mathcal{B}_{p+d}(\mathbb{R})} \{ \ell_n(f_\beta) + \operatorname{bina}(\beta) \}$
- ▶ $J_j(\hat{\beta}) = \{ l : \beta_{j,l} \neq \beta_{j,l-1}, \text{ pour } l = 2, \dots, d_j + 1 \} = \{ \hat{l}_{j,1}, \dots, \hat{l}_{j,s_j} \}$

- ▶ Binarisation par encodage “one-hot” [8] $x_i^B \in \mathbb{R}^{p+d}$
- ▶ $f_\beta : x_i \mapsto \beta^\top x_i^B = \sum_{j=1}^p f_{\beta_{j,\bullet}}(x_i)$ pour estimer f^\star
- ▶ log-vraisemblance négative partielle $\ell_n(f_\beta)$
- ▶ $\hat{\beta} \in \operatorname{argmin}_{\beta \in \mathcal{B}_{p+d}(\mathbb{R})} \{ \ell_n(f_\beta) + \operatorname{bina}(\beta) \}$
- ▶ $J_j(\hat{\beta}) = \{ l : \beta_{j,l} \neq \beta_{j,l-1}, \text{ pour } l = 2, \dots, d_j + 1 \} = \{ \hat{l}_{j,1}, \dots, \hat{l}_{j,s_j} \}$
- ▶ $\hat{\mu}_{j,\bullet} = (\mu_{j,\hat{l}_{j,1}}, \dots, \mu_{j,\hat{l}_{j,s_j}})^\top$ avec $s_j = |J_j(\hat{\beta})| = \hat{K}_j$

```
1 from tick.simulation import SimuCoxRegWithCutPoints
2 from tick.preprocessing.features_binarizer import FeaturesBinarizer
3 from tick.inference import CoxRegression
4
5 # Generate data
6 simu = SimuCoxRegWithCutPoints(n_samples=1000, n_features=20)
7 X, Y, delta = simu.simulate()
8
9 # Binarize features
10 binarizer = FeaturesBinarizer(n_cuts=50)
11 X_bin = binarizer.fit_transform(X)
12
13 # Fit the model with a penalty strength equal to `C`
14 learner = CoxRegression(penalty='binarsity',
15                          blocks_start=binarizer.blocks_start,
16                          blocks_length=binarizer.blocks_length,
17                          C=10)
18 learner.fit(X_bin, Y, delta)
19
20 # Obtain the estimated vector
21 beta = learner.coeefs
```

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

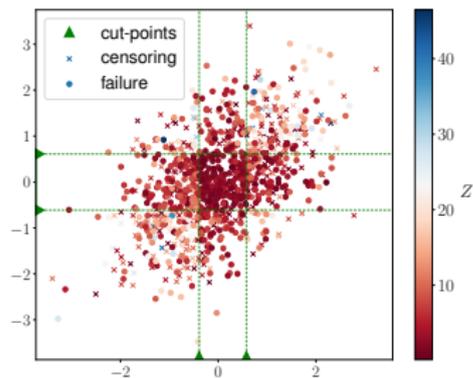
Binacox

TP python

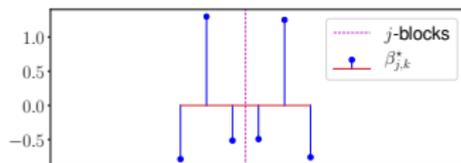
PBCseq

Consignes

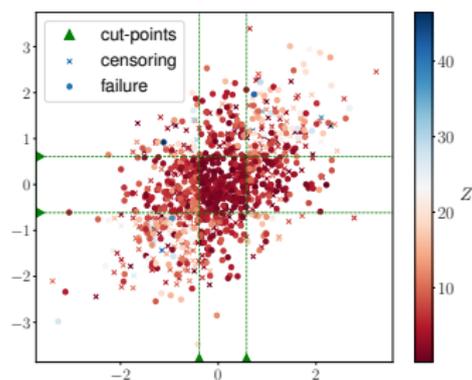
Références



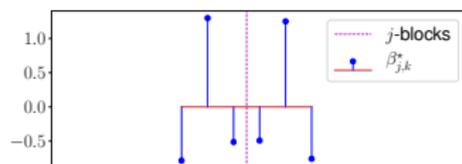
(a) Données générées



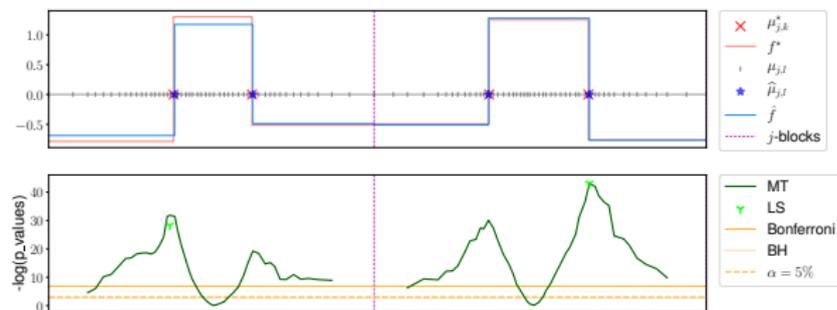
(b) β^* correspondant



(a) Données générées



(b) β^* correspondant



(c) Résultats des différentes méthodes

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

IV. TP sous python

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

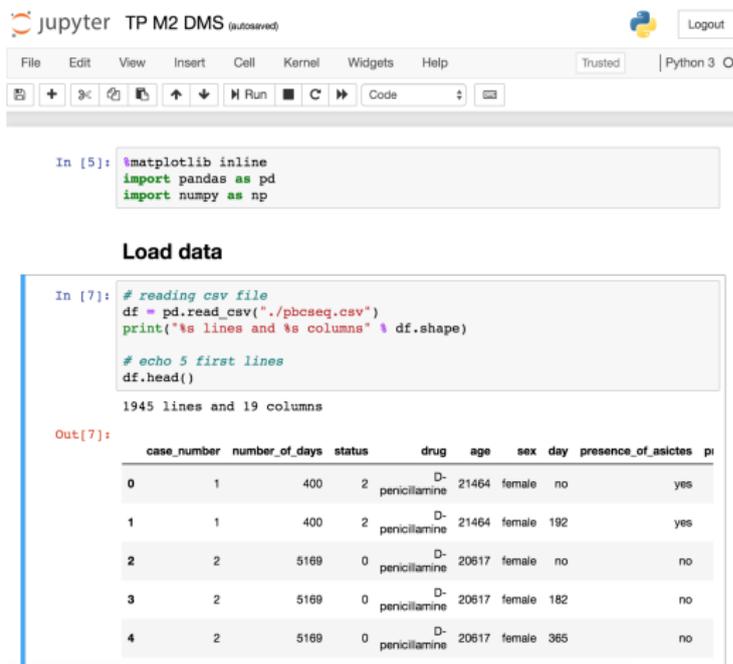
Consignes

Références

▶ <https://openml.org/d/516> : télécharger le .csv

Les données PBCseq

- ▶ <https://openml.org/d/516> : télécharger le .csv
- ▶ Charger les données sur votre notebook jupyter



The screenshot shows a Jupyter Notebook titled "TP M2 DMS (autosaved)". The interface includes a menu bar (File, Edit, View, Insert, Cell, Kernel, Widgets, Help), a toolbar with icons for file operations and execution, and a "Python 3" environment selector. The notebook contains two code cells. The first cell, labeled "In [5]:", imports matplotlib, pandas, and numpy. The second cell, labeled "In [7]:", reads a CSV file, prints its dimensions, and displays the first five rows. The output of the second cell shows the dimensions and a table of data.

```
In [5]: %matplotlib inline
import pandas as pd
import numpy as np
```

Load data

```
In [7]: # reading csv file
df = pd.read_csv("./pbcseq.csv")
print("%s lines and %s columns" % df.shape)

# echo 5 first lines
df.head()
```

1945 lines and 19 columns

Out[7]:

	case_number	number_of_days	status	drug	age	sex	day	presence_of_asictes	p
0	1	400	2	D-penicillamine	21464	female	no	yes	
1	1	400	2	D-penicillamine	21464	female	192	yes	
2	2	5169	0	D-penicillamine	20617	female	no	no	
3	2	5169	0	D-penicillamine	20617	female	182	no	
4	2	5169	0	D-penicillamine	20617	female	365	no	

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Consignes du TP

Avant de commencer

- ▶ Description des données

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Avant de commencer

- ▶ Description des données
- ▶ Statistiques descriptives et visualisation

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Consignes du TP

Avant de commencer

- ▶ Description des données
- ▶ Statistiques descriptives et visualisation

Extraction de features

- ▶ Utiliser par exemple `tsfresh`

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Consignes du TP

Avant de commencer

- ▶ Description des données
- ▶ Statistiques descriptives et visualisation

Extraction de features

- ▶ Utiliser par exemple `tsfresh`
- ▶ Bien d'autres méthodes sophistiquées existent.
Exemple: GP [3]

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Consignes du TP

Avant de commencer

- ▶ Description des données
- ▶ Statistiques descriptives et visualisation

Extraction de features

- ▶ Utiliser par exemple `tsfresh`
- ▶ Bien d'autres méthodes sophistiquées existent.
Exemple: GP [3]
- ▶ On se place dans un contexte de grande dimension!

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Consignes du TP

Avant de commencer

- ▶ Description des données
- ▶ Statistiques descriptives et visualisation

Extraction de features

- ▶ Utiliser par exemple `tsfresh`
- ▶ Bien d'autres méthodes sophistiquées existent.
Exemple: GP [3]
- ▶ On se place dans un contexte de grande dimension!

Comparaison de modèles

- ▶ Considérer (at least) Cox, AFT, RSF, C-mix, Binacox

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Consignes du TP

Avant de commencer

- ▶ Description des données
- ▶ Statistiques descriptives et visualisation

Extraction de features

- ▶ Utiliser par exemple `tsfresh`
- ▶ Bien d'autres méthodes sophistiquées existent.
Exemple: GP [3]
- ▶ On se place dans un contexte de grande dimension!

Comparaison de modèles

- ▶ Considérer (at least) Cox, AFT, RSF, C-mix, Binacox
- ▶ Exemple de comparaison : <https://github.com/SimonBussy/early-readmission-prediction>

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Consignes du TP

Avant de commencer

- ▶ Description des données
- ▶ Statistiques descriptives et visualisation

Extraction de features

- ▶ Utiliser par exemple `tsfresh`
- ▶ Bien d'autres méthodes sophistiquées existent.
Exemple: GP [3]
- ▶ On se place dans un contexte de grande dimension!

Comparaison de modèles

- ▶ Considérer (at least) Cox, AFT, RSF, C-mix, Binacox
- ▶ Exemple de comparaison : <https://github.com/SimonBussy/early-readmission-prediction>
- ▶ Visualiser les courbes d'apprentissage en CV

Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

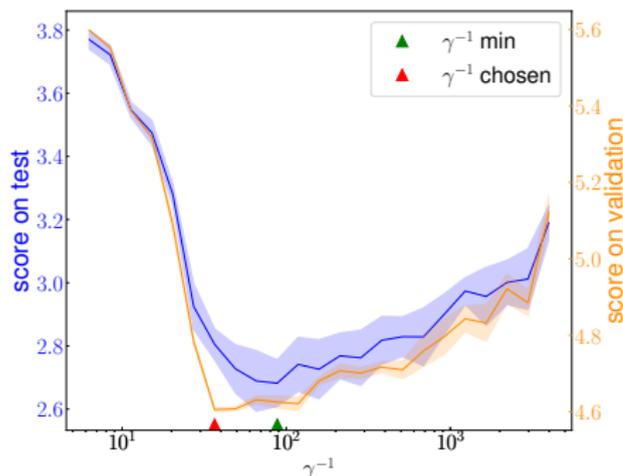
PBCseq

Consignes

Références

Comparaison de modèles

- ▶ Considérer (at least) Cox, AFT, RSF, C-mix, Binacox
- ▶ Exemple de comparaison : <https://github.com/SimonBussy/early-readmission-prediction>
- ▶ Visualiser les courbes d'apprentissage en CV



Programme

Rappel

Généralités

La survie

Quelques modèles

Cox PH

C-mix

Binacox

TP python

PBCseq

Consignes

Références

Programme

Rappel

Généralités
La survie

Quelques modèles

Cox PH
C-mix
Binacox

TP python

PBCseq
Consignes

Références

- [1] Simon Bussy. *Introduction of high-dimensional interpretable machine learning models and their applications*. PhD thesis, SORBONNE UNIVERSITE, 2019.
- [2] Simon Bussy, Agathe Guilloux, Stéphane Gaïffas, and Anne-Sophie Jannot. C-mix: A high-dimensional mixture model for censored durations, with applications to genetic data. *Statistical Methods in Medical Research*, 0(0):0962280218766389, 2018.
- [3] Simon Bussy, Raphaël Veil, Vincent Looten, Anita Burgun, Stéphane Gaïffas, Agathe Guilloux, Brigitte Ranque, and Anne-Sophie Jannot. Comparison of methods for early-readmission prediction in a high-dimensional heterogeneous covariates and time-to-event outcome framework. *BMC medical research methodology*, 19(1):50, 2019.
- [4] Hung-Chia Chen, Ralph L Kodell, Kuang Fu Cheng, and James J Chen. Assessment of performance of survival prediction models for cancer prognosis. *BMC medical research methodology*, 12(1):102, 2012.
- [5] J. P. Klein and M. L. Moeschberger. *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media, 2005.
- [6] Lynn Kuo and Fengchun Peng. A mixture-model approach to the analysis of survival data. *Biostatistics-Basel-*, 5:255–272, 2000.
- [7] B. Lausen and M. Schumacher. Maximally selected rank statistics. *Biometrics*, pages 73–85, 1992.
- [8] J. Wu and S. Coggeshall. *Foundations of Predictive Analytics (Chapman & Hall/CRC Data Mining and Knowledge Discovery Series)*. Chapman & Hall/CRC, 1st edition, 2012.