

Introduction

Le modèle de
Cox lasso

Courbes ROC
dépendantes
du temps

Discussions

Estimation des capacités pronostiques

Equipe d'Accueil 4275 "Biostatistique, recherche clinique et mesures subjectives en santé", Université de Nantes

Vendredi 25 Mai 2012



UNIVERSITÉ DE NANTES



CENTRE HOSPITALIER
UNIVERSITAIRE DE NANTES

itun

institut
transplantation
urologie
néphrologie
INSERM - UMR 643

Introduction

Le modèle de
Cox lasso

Courbes ROC
dépendantes
du temps

Discussions

1. Introduction

2. Le modèle de Cox lasso

3. Courbes ROC dépendantes du temps

4. Discussions

Introduction

Le modèle de
Cox lassoCourbes ROC
dépendantes
du temps

Discussions

- Identifier un ensemble de gènes pronostiques de l'évolution d'une maladie pour :
 - Améliorer le suivi et la prise en charge en fonction du pronostic.
 - Identifier des cibles thérapeutiques permettant de modifier ce pronostic.

Introduction

Le modèle de
Cox lasso

Courbes ROC
dépendantes
du temps

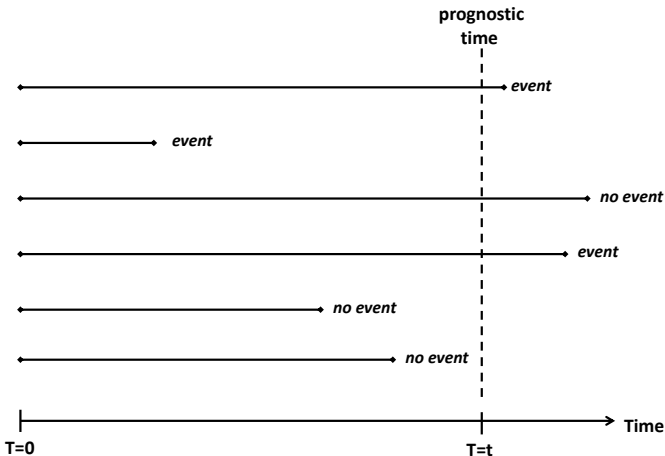
Discussions

- Développement des techniques dites de haut-débit :
 - Nombre (P) très important de variables mesurées simultanément pour un sujet.
- Méthodes utilisées en recherche fondamentale :
 - Taille de l'échantillon (N) faible.

Problème

- $P \gg N$ alors que toute la théorie statistique est construite pour l'inverse.
- Littérature assez importante en médecine diagnostique mais pas en pronostic (censure).

- Variable à expliquer = Temps d'apparition d'un événement



Introduction

Le modèle de
Cox lasso

Courbes ROC
dépendantes
du temps

Discussions

- Estimateur de Kaplan-Meier
- Statistique du LogRank → p-value
- Modèle à risque proportionnel de Cox → HR + p-value

Introduction

Le modèle de
Cox lassoCourbes ROC
dépendantes
du temps

Discussions

- Sélection des variables à partir de facteurs de risque cliniques mesurés au même moment, puis corrélations avec le temps d'événement. * †

Problème

- Comment construire une signature *sous-clinique* ?

*. Alizadeh et al. Distinct types of diffuse large b-cell lymphoma identified by gene expression profiling. *Nature*, 403 (6769) :503-11, 2000.

†. Rosenwald et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-b cell lymphoma. *New England Journal of Medicine*, 346(25) :1937-47, 2002.

Introduction

Le modèle de
Cox lassoCourbes ROC
dépendantes
du temps

Discussions

- Sélection des variables directement sur l'analyse du temps d'événement.
 - Statistique du LogRank pour 2 à 3 classes pour chaque gène pour toutes les valeurs seuils possibles.
 - Modèle de Cox (univarié, puis multivarié).

Problèmes

- La p-value n'a pas de sens pour une multitude de tests.
- Corrélation \neq Prédiction

Introduction

Le modèle de
Cox lassoCourbes ROC
dépendantes
du temps

Discussions

- Sélection des variables directement sur l'analyse du temps d'événement.
 - Modèle de Cox (univarié, puis multivarié) sur échantillon d'apprentissage.
 - Validation sur un échantillon indépendant.

Problèmes

- Réduction du nombre de sujets (déjà très faible).
- Répétition de la procédure jusqu'à l'obtention de résultats sympathiques...

Introduction

Le modèle de
Cox lassoCourbes ROC
dépendantes
du temps

Discussions

- 1 Modèle de Cox avec pénalisation de lasso
 - Sélection des variables + Construction de la signature
- 2 Correction de l'aire sous la courbe ROCT
 - Evaluation de la capacité pronostique

Introduction

Le modèle de
Cox lasso

Courbes ROC
dépendantes
du temps

Discussions

1. Introduction

2. Le modèle de Cox lasso

3. Courbes ROC dépendantes du temps

4. Discussions

Introduction

Le modèle de
Cox lasso

Courbes ROC
dépendantes
du temps

Discussions

- X_j est le vecteur des P variables avec $X_j = (X_{j1}, \dots, X_{jP})$.
- T_j est le temps d'apparition de l'événement.
- C_j est le temps du dernier suivi (censure à droite).
- Si $C_j \geq T_j$, alors T_j est observé et $\Delta_j = 1$.

La fonction de risque de connaître l'événement au temps t sachant $X_j = x_j$ s'écrit :

$$h(t|x_j) = h_0(t) \exp(\eta_j)$$

- $h_0(t)$ est la fonction de risque de base (non estimées).
- η_j est le score pronostique pour le sujet j :

$$\eta_j = x_j' \beta = \beta_1 x_{j1} + \dots + \beta_P x_{jP}$$

- Interprétation : $HR_{X_j=1/X_j=0} = \exp(\beta_j) \quad \forall j = 1, \dots, P$

Introduction

Le modèle de
Cox lassoCourbes ROC
dépendantes
du temps

Discussions

$$\eta_j = \mathbf{x}_j' \boldsymbol{\beta} = \beta_1 x_{j1} + \dots + \beta_P x_{jP}$$

- Si X_j facteur de risque : $HR_{X_j=1/X_j=0} > 1$ et $\beta_j > 0$.
- Si X_j facteur protecteur : $HR_{X_j=1/X_j=0} < 1$ et $\beta_j < 0$.
- Si X_j facteur indépendant : $HR_{X_j=1/X_j=0} = 1$ et $\beta_j = 0$.

⇒ Quand $\eta_j \nearrow$, le risque d'apparition de l'événement \nearrow .

- Soit $t_1 < t_2 < \dots < t_r$ les différents temps des r événements observés.
- La probabilité que le sujet k déclare l'événement en t_k parmi les sujets à risque au même temps est égale à :

$$l_k(\beta) = \frac{h(t_k | x_k)}{\sum_{i: t_i \geq t_k} h(t_i | x_i)}$$

$$l_k(\beta) = \frac{h_0(t_k) \exp(x'_k \beta)}{\sum_{i: t_i \geq t_k} h_0(t_i) \exp(x'_i \beta)}$$

$$l_k(\beta) = \frac{\exp(x'_k \beta)}{\sum_{i: t_i \geq t_j} \exp(x'_i \beta)}$$

- Ne dépend pas de $h_0(t)$

Si les individus **indépendants** et en passant au log :

$$\ell(\beta) = \sum_{k=1}^r \left\{ x'_k \beta - \log \left(\sum_{i:t_i \geq t_j} \exp(x'_i \beta) \right) \right\}$$

La maximisation de cette fonction permet d'estimer l'effet des variables sur le temps de survie :

$$\hat{\beta} = \arg \max_{\beta} \{ \ell(\beta) \}$$

Problèmes

- Trop de paramètres à estimer.
- Sélection sur la pvalue inadéquate.
- Colinéarité importante de certaines variables.

Introduction

Le modèle de
Cox lassoCourbes ROC
dépendantes
du temps

Discussions

$$\hat{\beta} = \arg \max_{\beta} \left\{ \ell(\beta) - \lambda \sum_{p=1}^P |\beta_p| \right\}$$

- λ est le paramètre de régularisation (*tuning parameter*).
- Permet d'obtenir un modèle parcimonieux :
 - Restriction de la valeur de certains coefficients.
 - Annulation de la valeur des coefficients associés aux variables les moins informatives.

‡. R Tibshirani. Regression shrinkage and selection via the Lasso.
Journal of the Royal Statistical Society : Series B, 58 :267-288, 1996

Illustration de l'approche Lasso à partir des données de Rosenwald

Introduction

Le modèle de
Cox lasso

Courbes ROC
dépendantes
du temps

Discussions

- 240 patients avec un lymphome B diffus à grandes cellules (DLBCL).
- Patients traités par anthracycline.
- 7399 gènes mesurés par puces au début du traitement.
- Objectif : prédire la mortalité à partir d'une signature basée sur quelques dizaines de gènes.

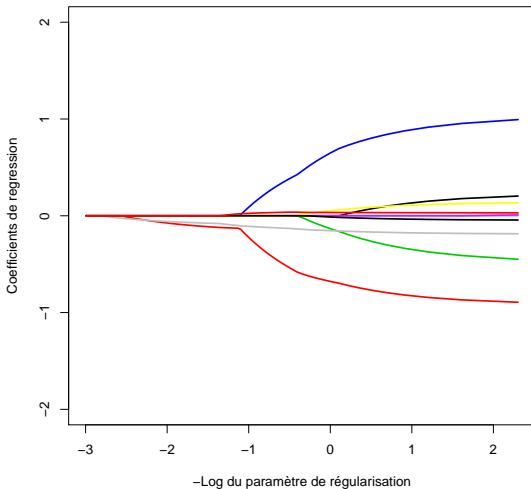
Illustration de l'approche Lasso (les 500 premiers gènes)

Introduction

Le modèle de
Cox lasso

Courbes ROC
dépendantes
du temps

Discussions



- Division de l'échantillon en K groupes ($k = 1, \dots, K$).
- Soit $\hat{\beta}_{(-k)}(\lambda)$ l'estimation de β sans le k ième groupe.
- Soit $\ell_{(-k)}(\hat{\beta}_{(-k)}(\lambda))$ la log vraisemblance partielle.
- Soit $\ell(\hat{\beta}_{(-k)}(\lambda))$ la log vraisemblance partielle avec tous les patients mais avec les paramètres précédents.

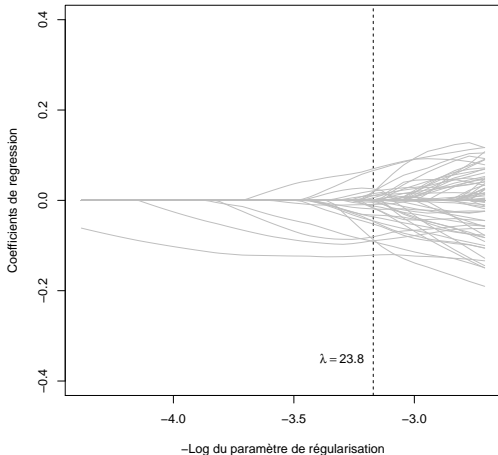
$$CV(\lambda) = \sum_{k=1}^K \{ \ell(\hat{\beta}_{(-k)}(\lambda)) - \ell_{(-k)}(\hat{\beta}_{(-k)}(\lambda)) \}$$

$$\hat{\lambda} = \arg \max_{\lambda} \{ CV(\lambda) \}$$

- Remarque : éviter minimisation AIC ou BIC.

§. Bovelstad et al. Predicting survival from microarray data - a comparative study. *Bioinformatics*, 23 :2080-7, 2007.

- Partitionnement en 5 groupes / $\hat{\lambda} = 23.8$ / 27 gènes retenus



→ Quelles sont les capacités pronostiques de $\hat{\eta}_j$?

Introduction

Le modèle de
Cox lasso

Courbes ROC
dépendantes
du temps

Discussions

1. Introduction

2. Le modèle de Cox lasso

3. Courbes ROC dépendantes du temps

4. Discussions

- Supposons que si $\hat{\eta}_j > c$, le patient est à risque de subir l'événement avant le temps τ .
- Le taux de faux négatifs est égal à :

$$FNR_{\tau}(c) = P(\hat{\eta} \leq c | T \leq \tau)$$

- Le taux de faux positifs est égal à :

$$FPR_{\tau}(c) = P(\hat{\eta} > c | T > \tau)$$

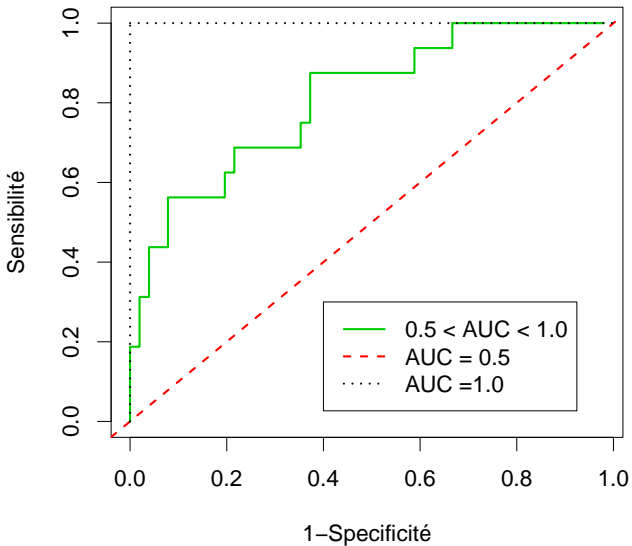
- Courbe ROC dépendante du temps (ROct) :
 - $1 - FNR_{\tau}(c)$ en fonction de $FPR_{\tau}(c)$ pour toutes les valeurs de c .

¶. Heagerty et al. Time-Dependent ROC Curves for Censored Survival Data and a Diagnostic Marker. Biometrics, 56 :337-44, 2000.

Introduction

Le modèle de
Cox lassoCourbes ROC
dépendantes
du temps

Discussions



Introduction

Le modèle de
Cox lassoCourbes ROC
dépendantes
du temps

Discussions

$$FNR_{\tau}(c) = \frac{\sum_{j=1}^N I(\hat{\eta}_j \leq c, T_j \leq \tau, \Delta_j = 1)}{\sum_{j=1}^N I(T_j \leq \tau, \Delta_j = 1)}$$

$$FPR_{\tau}(c) = \frac{\sum_{j=1}^N I(\hat{\eta}_j > c, T_j > \tau)}{\sum_{j=1}^N I(T_j > \tau)}$$

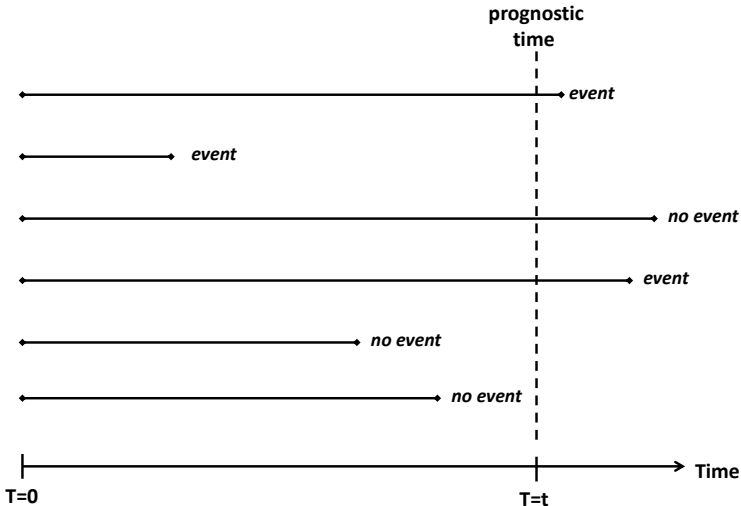
Problème

- Les données sont souvent censurées pour un pronostic à long terme.

Introduction

Le modèle de
Cox lassoCourbes ROC
dépendantes
du temps

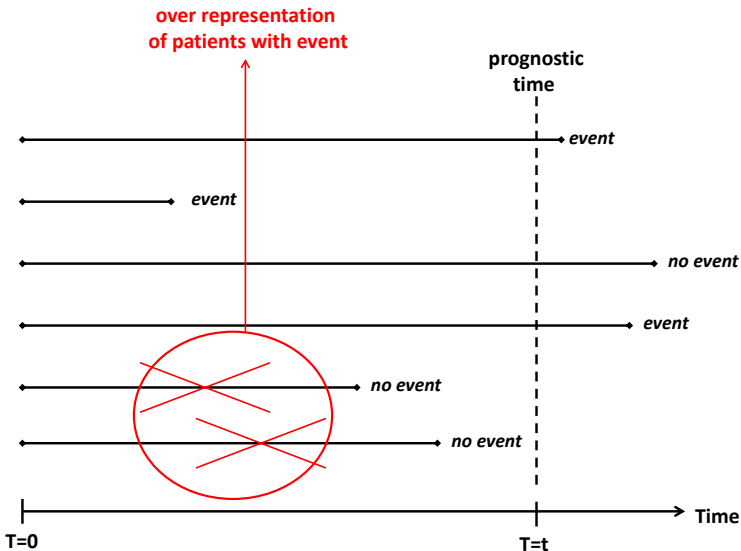
Discussions



Introduction

Le modèle de
Cox lassoCourbes ROC
dépendantes
du temps

Discussions



Introduction

Le modèle de
Cox lassoCourbes ROC
dépendantes
du temps

Discussions

$$S(\tau|\hat{\eta} \leq c) = 1 - P(T \leq \tau|\hat{\eta} \leq c) \text{ et } G(c) = P(\hat{\eta} \leq c)$$

$$\Downarrow$$

Théorème de Bayes

$$\Downarrow$$

$$FNR_{\tau}(c) = P(\hat{\eta} \leq c|T \leq \tau)$$

$$+$$

Estimateur des plus proches voisins ||

||. Akritas. Nearest neighbor estimation of a bivariate distribution under random censoring. *Annals of Statistics*, 22 :1299-327, 1994

**.. Heagerty et al. Time-Dependent ROC Curves for Censored Survival Data and a Diagnostic Marker. *Biometrics*, 56 :337-44, 2000. 28 / 43

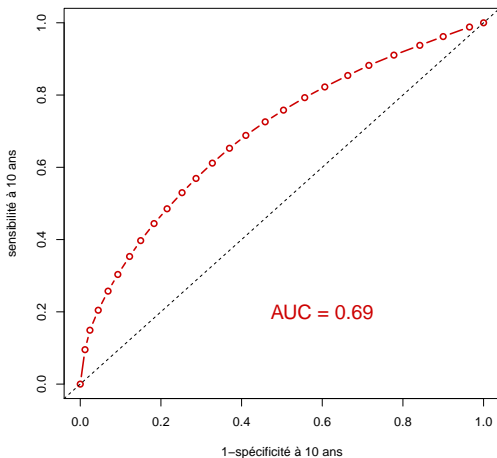
Application à la signature de Rosenwald ($\hat{\lambda} = 23.8$, 27 gènes)

Introduction

Le modèle de Cox lasso

Courbes ROC dépendantes du temps

Discussions



→ Le sur-ajustement semble maîtrisé grâce à Lasso.

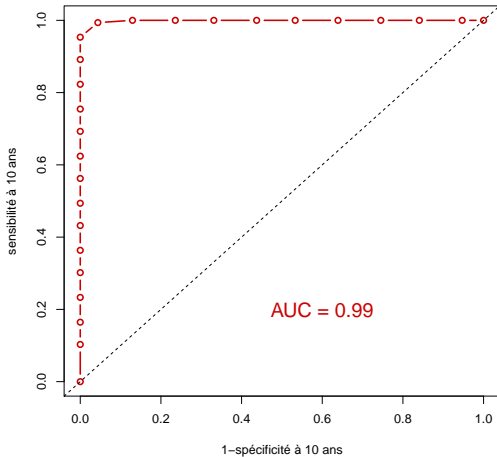
Application à la signature de Rosenwald ($\lambda = 3.0$, 197 gènes)

Introduction

Le modèle de
Cox lasso

Courbes ROC
dépendantes
du temps

Discussions



Introduction

Le modèle de
Cox lasso

Courbes ROC
dépendantes
du temps

Discussions

- Méthode classique : séparation apprentissage et validation.
- Problème : aggravation de la faible puissance.
- Ex Rosenwald :
 - Apprentissage - $N = 160 - \hat{\lambda} = 20.8 - 24$ gènes.
 - Validation - $N = 80$.

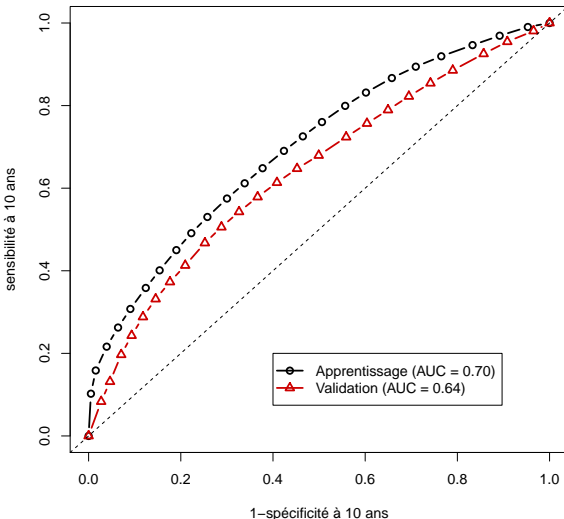
††. Molinaro et al. Prediction error estimation : a comparison of resampling methods. Bioinformatics, 21(15) :3301-3307, 2005.

Introduction

Le modèle de
Cox lasso

Courbes ROC
dépendantes
du temps

Discussions



Introduction

Le modèle de
Cox lassoCourbes ROC
dépendantes
du temps

Discussions

- B échantillons de taille N avec remise (Bootstrap).
- Pour les B échantillons des patients inclus ($b = 1, \dots, B$) :
 - Détermination de λ .
 - Sélection des variables.
 - Estimation des paramètres β .
- Pour les B échantillons des patients non-inclus :
 - Calcul des quantités $\widehat{FNR}_\tau^b(c)$ et $\widehat{FPR}_\tau^b(c)$

Introduction

Le modèle de
Cox lassoCourbes ROC
dépendantes
du temps

Discussions

$$\widehat{FNR}_{\tau}^{BCV}(c) = B^{-1} \sum_{b=1}^B \widehat{FNR}_{\tau}^b(c)$$

$$\widehat{FPR}_{\tau}^{BCV}(c) = B^{-1} \sum_{b=1}^B \widehat{FPR}_{\tau}^b(c)$$

- Courbe ROCt par BCV :

→ $1 - \widehat{FNR}_{\tau}^{BCV}(c)$ en fonction de $\widehat{FPR}_{\tau}^{BCV}(c)$ pour toutes les valeurs de c .

- Sous-estimation des capacités pronostiques.

$$\overline{FNR}_\tau(c) = B^{-1} \sum_{b=1}^B \widehat{FNR}_\tau^{b+}(c)$$

$$\overline{FNR}_\tau(c) = B^{-1} \sum_{b=1}^B \widehat{FPR}_\tau^{b+}(c)$$

- $\widehat{FNR}_\tau^{b+}(c)$ et $\widehat{FPR}_\tau^{b+}(c)$ sont obtenus avec le modèle de l'itération b mais en utilisant tous les patients, qu'ils soient inclus ou non dans l'échantillon d'apprentissage.
- Courbe ROCt apparente :
 - $1 - \overline{FNR}_\tau(c)$ en fonction de $\overline{FPR}_\tau(c)$ pour toutes les valeurs de c .
- Surestimation des capacités pronostiques.

Introduction

Le modèle de
Cox lassoCourbes ROC
dépendantes
du temps

Discussions

$$\widehat{FNR}_\tau^{632}(c) = 0.368 \overline{FNR}_\tau(c) + 0.632 \widehat{FNR}_\tau^{BCV}(c)$$

$$\widehat{FPR}_\tau^{632}(c) = 0.368 \overline{FPR}_\tau(c) + 0.632 \widehat{FPR}_\tau^{BCV}(c)$$

- Courbe ROCt bootstrap 0.632 :

→ $1 - \widehat{FNR}_\tau^{632}(c)$ en fonction de $\widehat{FPR}_\tau^{632}(c)$ pour toutes les valeurs de c .

- Petite sous-estimation des capacités pronostiques. ‡‡

‡‡. Efron and Tibshirani. Improvements on cross-validation : The .632+ bootstrap method. Journal of the American Statistical Association, 92(438) :548-560, 1997.

Introduction

Le modèle de
Cox lassoCourbes ROC
dépendantes
du temps

Discussions

$$\widehat{FNR}_{\tau}^{632+}(c) = \{1 - \psi(\hat{r}_{FNR,\tau}(c))\} \overline{FNR}_{\tau}(c) + \psi(\hat{r}_{FNR,\tau}(c)) \widehat{FNR}_{\tau}^{BCV}(c)$$

$$\widehat{FPR}_{\tau}^{632+}(c) = \{1 - \psi(\hat{r}_{FPR,\tau}(c))\} \overline{FPR}_{\tau}(c) + \psi(\hat{r}_{FPR,\tau}(c)) \widehat{FPR}_{\tau}^{BCV}(c)$$

- $\psi(x) = 0.632/(1 - 0.368x)$
- $\hat{r}_{FNR,\tau}(c) = \{\widehat{FNR}_{\tau}^{BCV}(c) - \overline{FNR}_{\tau}(c)\} / \{\hat{\gamma}_{FNR,\tau}(c) - \overline{FNR}_{\tau}(c)\}$
- $\hat{r}_{FPR,\tau}(c) = \{\widehat{FPR}_{\tau}^{BCV}(c) - \overline{FPR}_{\tau}(c)\} / \{\hat{\gamma}_{FPR,\tau}(c) - \overline{FPR}_{\tau}(c)\}$
- Courbe ROCt bootstrap 0.632+ :

→ $1 - \widehat{FNR}_{\tau}^{632+}(c)$ en fonction de $\widehat{FPR}_{\tau}^{632+}(c)$ pour toutes les valeurs de c .

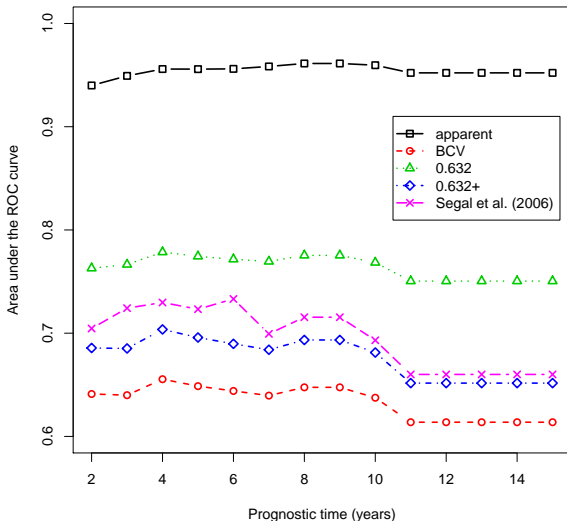
Simulations : moyennes (écart-types)

Level	N	cens.	Apparent	BCV	0.632	0.632+		
Introduction	0/750	60	0.3	0.743(0.103)	0.518(0.071)	0.608(0.059)	0.528(0.080)	
			0.5	0.781(0.097)	0.504(0.058)	0.614(0.053)	0.513(0.065)	
			0.7	0.745(0.108)	0.500(0.058)	0.596(0.058)	0.510(0.063)	
	Le modèle de Cox lasso	125	0.3	0.782(0.071)	0.514(0.050)	0.623(0.040)	0.527(0.057)	
			0.5	0.741(0.082)	0.498(0.054)	0.595(0.047)	0.508(0.062)	
			0.7	0.701(0.118)	0.502(0.058)	0.581(0.059)	0.514(0.066)	
		250	0.3	0.799(0.035)	0.508(0.039)	0.625(0.027)	0.521(0.046)	
			0.5	0.777(0.062)	0.505(0.046)	0.612(0.039)	0.516(0.054)	
			0.7	0.693(0.121)	0.505(0.059)	0.579(0.061)	0.519(0.067)	
Courbes ROC dépendantes du temps	2/3	60	0.3	0.846(0.050)	0.789(0.063)	0.810(0.057)	0.808(0.058)	
			0.5	0.818(0.082)	0.742(0.092)	0.770(0.086)	0.765(0.089)	
			0.7	0.749(0.119)	0.656(0.110)	0.690(0.110)	0.677(0.116)	
	125	0.3	0.842(0.041)	0.818(0.044)	0.827(0.042)	0.826(0.042)		
		0.5	0.840(0.049)	0.803(0.058)	0.816(0.054)	0.815(0.055)		
		0.7	0.798(0.076)	0.732(0.086)	0.756(0.081)	0.751(0.084)		
	250	0.3	0.838(0.032)	0.825(0.033)	0.830(0.033)	0.830(0.033)		
		0.5	0.836(0.033)	0.817(0.037)	0.824(0.035)	0.824(0.035)		
		0.7	0.828(0.047)	0.795(0.054)	0.807(0.051)	0.806(0.051)		
	Discussions	2/750	60	0.3	0.902(0.039)	0.646(0.083)	0.751(0.058)	0.677(0.090)
				0.5	0.869(0.071)	0.591(0.091)	0.702(0.071)	0.614(0.102)
				0.7	0.827(0.086)	0.534(0.072)	0.649(0.061)	0.549(0.083)
		125	0.3	0.905(0.032)	0.783(0.051)	0.831(0.036)	0.819(0.048)	
			0.5	0.890(0.041)	0.729(0.073)	0.791(0.053)	0.765(0.075)	
			0.7	0.834(0.087)	0.619(0.086)	0.702(0.070)	0.647(0.097)	
250		0.3	0.898(0.023)	0.811(0.027)	0.845(0.022)	0.843(0.025)		
		0.5	0.903(0.025)	0.789(0.037)	0.833(0.027)	0.825(0.034)		
		0.7	0.881(0.052)	0.729(0.060)	0.785(0.047)	0.766(0.059)		

Introduction

Le modèle de
Cox lassoCourbes ROC
dépendantes
du temps

Discussions



Gain de temps de calcul

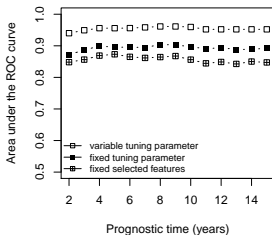
Introduction

Le modèle de Cox lasso

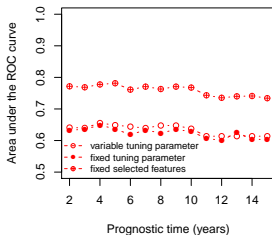
Courbes ROC dépendantes du temps

Discussions

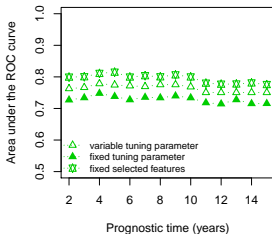
apparent



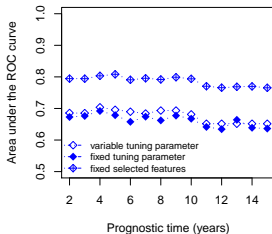
BCV



0.632



0.632+



Introduction

Le modèle de
Cox lasso

Courbes ROC
dépendantes
du temps

Discussions

1. Introduction

2. Le modèle de Cox lasso

3. Courbes ROC dépendantes du temps

4. Discussions

Introduction

Le modèle de
Cox lasso

Courbes ROC
dépendantes
du temps

Discussions

- Hypothèses peu vérifiables :
 - Hypothèses de log linéarité.
 - Hypothèses de proportionnalité des risques.
- Cette méthode ne remplace pas "apprentissage/validation" si $N > P$.
- Moins intuitif que la méthode "apprentissage/validation" :
- Esprit conservateur du reviewing médical.

Introduction

Le modèle de
Cox lassoCourbes ROC
dépendantes
du temps

Discussions

- Méthodologie applicable hors technique haut débit. Ex :
 - N=500 patients avec 100 événements.
 - 40 variables explicatives possibles binaires.
 - 20 variables continues avec 4 possibilités de codage.
 - 5 interactions pertinentes.
 - $P = 20 + 20 \times 4 + 5 = 105$.
- Analyses à venir :
 - AT1R et le risque rejet aigu du greffon rénal.
 - Pronostic de la mortalité relative en transplantation rénal.
 - Pronostic du délai de reprise au démarrage du greffon rénal (pas de censure).
 - Cytokines et risque de GVHD chronique en greffe de moelle.