Available online at www.sciencedirect.com
ScienceDirect
C. R. Acad. Sci. Paris, Ser. I 344 (2007) 641–644
COMPTES RENDUS MATHEMATIQUE
http://france.elsevier.com/direct/CRASS1/

Statistique

Modélisation de risques concurrents par un modèle de mélange semi-paramétrique

Jean-François Dupuy [a], Gabriel Escarela [b]

[a] *Laboratoire de Statistique et Probabilités, Université Paul Sabatier, 118, route de Narbonne, 31062 Toulouse, France*
[b] *Departamento de Matemáticas, Universidad Autónoma Metropolitana, Unidad Iztapalapa, Av. San Rafael Atlixco No. 186, Col. Vicentina, C.P. 09340, D.F., Mexique*

Reçu le 11 juin 2006 ; accepté après révision le 23 mars 2007

Disponible sur Internet le 7 mai 2007

Présenté par Paul Deheuvels

Résumé

Nous nous intéressons à un modèle de mélange semi-paramétrique proposé dans la littérature pour modéliser des risques concurrents dans un contexte de durées de vie censurées à droite et observées avec des variables explicatives. Nous montrons la convergence forte des estimateurs qui ont été proposés pour estimer les paramètres de ce modèle. ***Pour citer cet article : J.-F. Dupuy, G. Escarela, C. R. Acad. Sci. Paris, Ser. I 344 (2007).***

Abstract

Modeling competing risks with a semi-parametric mixture model. We consider a semi-parametric mixture model for competing risks in right-censored survival data with explanatory variables. We prove strong consistency of estimators of the parameters in this model. ***To cite this article: J.-F. Dupuy, G. Escarela, C. R. Acad. Sci. Paris, Ser. I 344 (2007).***

1. Introduction

Dans les modèles à risques concurrents, chaque sujet est soumis à plusieurs risques de décès en même temps, et l'observation s'interrompt au premier type de décès observé. On observe pour chaque sujet une durée de vie aléatoire T^0. On suppose que T^0 peut être censurée aléatoirement à droite par une variable aléatoire positive C. On note $T = \min(T^0, C)$ et $\Delta = 1\{T^0 \leqslant C\}$, où $1\{\cdot\}$ désigne la fonction indicatrice. Si le décès se produit, on observe un indice H à valeurs dans $\mathcal{J} = \{1, \ldots, J\}$, qui indique la cause du décès. H n'est pas observé pour un sujet censuré.

Une modélisation possible de ce type de phénomène prend la forme d'un modèle de mélange. Ce modèle est déterminé par les fonctions de risque instantané $\lambda_j(t|Z) = \lim_{h\downarrow 0} h^{-1}\mathbb{P}[t \leqslant T^0 < t + h|T^0 \geqslant t, Z, H = j]$ et les probabilités conditionnelles $p^{j,X} = \mathbb{P}[H = j|X]$ $(j \in \mathcal{J})$, où Z et X sont respectivement des p et q vecteurs de variables explicatives (qui peuvent avoir des composantes communes). Le modèle le plus couramment retenu pour

Adresses e-mail : dupuy@math.ups-tlse.fr, dupuy@cict.fr (J.-F. Dupuy), ge@xanum.uam.mx (G. Escarela).

1631-073X/$ – see front matter © 2007 Académie des sciences. Publié par Elsevier Masson SAS. Tous droits réservés.
doi:10.1016/j.crma.2007.03.031

les $p^{j,X}$ suppose que $p^{j,X} = \exp(\gamma_j' X)/\sum_{k=1}^{J} \exp(\gamma_k' X)$ ($\gamma_j \in \mathbb{R}^q$, $j \in \mathcal{J}$, $\gamma_J = 0$). Ce modèle de mélange présente l'avantage de ne pas nécessiter d'hypothèses sur la loi jointe des durées de vie sous les différents risques.

Les aspects algorithmiques de l'estimation dans ce type de modèle ont récemment fait l'objet de plusieurs articles [4,5], lorsque les $\lambda_j(t|Z)$ sont modélisés semi-paramétriquement par un modèle de Cox supposant $\lambda_j(t|Z) = \lambda_j(t)\exp(\beta_j' Z)$, où $\{\lambda_j(t)\colon\ t \geqslant 0, j \in \mathcal{J}\}$ sont J fonctions de risque de base inconnues et $\{\beta_j, j \in \mathcal{J}\}$ sont des paramètres de régression inconnus de $\mathbb{R}^p$. L'asymptotique des estimateurs obtenus n'a en revanche pas été étudiée (elle l'a été pour une formulation paramétrique des $\lambda_j(t|Z)$, [2]).

Dans le cadre du modèle de mélange semi-paramétrique, nous montrons la consistance d'estimateurs du maximum de vraisemblance de $\theta = \{(\beta_j, \gamma_j, \Lambda_j), j \in \mathcal{J}\}$, où $\Lambda_j(\cdot) = \int_0^{\cdot} \lambda_j(s)\,\mathrm{d}s$.

2. Notations et estimation du maximum de vraisemblance non-paramétrique

On note $[0, \tau]$ ($\tau < \infty$) l'intervalle d'observation et $(T_i, \Delta_i, H_i, Z_i, X_i)$ ($i = 1, \dots, n$) n répliques indépendantes du quintuplet (T, Δ, H, Z, X). Pour le sujet i, on observe $(T_i, \Delta_i, H_i, Z_i, X_i)$ si $\Delta_i = 1$ et $(T_i, \Delta_i, Z_i, X_i)$ si $\Delta_i = 0$. On suppose que la censure C est indépendante de T^0 conditionnellement à Z et non-informative.

On note $\theta_0 = \{(\beta_{j,0}, \gamma_{j,0}, \Lambda_{j,0}), j \in \mathcal{J}\}$ la « vraie » valeur du paramètre. On suppose que $\forall j \in \mathcal{J}$, $\beta_{j,0} \in \mathcal{B} \subset \mathbb{R}^p$, $\gamma_{j,0} \in \mathcal{G} \subset \mathbb{R}^q$, où $\mathcal{B}$ et $\mathcal{G}$ sont des compacts connus, et que $\Lambda_{j,0}$ est croissante et dérivable sur $[0, \tau]$, de dérivée $\lambda_{j,0}$, et telle que $\Lambda_{j,0}(0) = 0$ et $\Lambda_{j,0}(\tau) < \infty$.

La vraisemblance de θ au vu de l'échantillon des n sujets s'écrit : $L_n(\theta) =$

$$\prod_{i=1}^{n} \left\{ \prod_{j=1}^{J} \left(\lambda_j(T_i|Z_i) \exp\left(-\int_0^{T_i} \lambda_j(s|Z_i)\,\mathrm{d}s \right) p_i^{j,X_i} \right)^{\Delta_i^j} \cdot \left(\sum_{j=1}^{J} \exp\left(-\int_0^{T_i} \lambda_j(s|Z_i)\,\mathrm{d}s \right) p_i^{j,X_i} \right)^{1-\Delta_i} \right\},$$

où $\Delta_i^j = \Delta_i 1\{H_i = j\}$. Cette vraisemblance faisant intervenir les paramètres fonctionnels λ_j ($j \in \mathcal{J}$), [4] et [5] proposent d'estimer θ_0 en adaptant à L_n la méthode dite du maximum de vraisemblance non-paramétrique (MVNP), décrite dans [1] pour le modèle de Cox usuel.

Appliquée au modèle de mélange semi-paramétrique décrit dans l'introduction, la méthode du MVNP consiste à chercher le maximum d'une version modifiée $\tilde{L}_n$ de L_n, obtenue en remplaçant dans L_n la fonction Λ_j, pour tout $j \in \mathcal{J}$, par une fonction $\tilde{\Lambda}_j$ en escalier croissante, dont les sauts se produisent aux instants de décès observés pour la cause j. La fonction $\tilde{L}_n$ est appelée vraisemblance non-paramétrique, par analogie avec la terminologie employée dans le cas du modèle de Cox usuel [1]. Son maximum est cherché sur l'espace Θ_n des $\{(\beta_j, \gamma_j, \tilde{\Lambda}_j), j \in \mathcal{J}\}$. Si ce maximum existe et est atteint en $\hat{\theta}_n = \{(\hat{\beta}_{j,n}, \hat{\gamma}_{j,n}, \hat{\Lambda}_{j,n}), j \in \mathcal{J}\}$, $\hat{\theta}_n$ est appelé estimateur du MVNP de θ_0.

Le résultat suivant assure l'existence d'un estimateur du MVNP de θ_0. Il n'est pas mentionné dans [4] et [5]. Sa démonstration est analogue à celle de la Proposition 2 de [3], nous l'omettrons.

Lemme 2.1. *Il existe un estimateur du MVNP* $\hat{\theta}_n = \arg\max_{\theta \in \Theta_n} \tilde{L}_n(\theta)$ *de* θ_0 *dans* Θ_n.

Nous renvoyons à [4] et [5] pour les aspects algorithmiques du calcul de $\hat{\theta}_n$. Pour tout $i = 1, \dots, n$ et tout $j \in \mathcal{J}$, soit $N_i^j(t) = 1\{T_i \leqslant t\}\Delta_i^j$, $Y_i(t) = 1\{t \leqslant T_i\}$, et soit Γ_i^j la variable aléatoire définie par $\Gamma_i^j = 1\{H_i = j\}$. Γ_i^j n'est observée que si $\Delta_i = 1$. Notons $E_\theta[\Gamma_i^j]$ l'espérance de Γ_i^j conditionnelle aux données observées et calculée avec la valeur θ du paramètre. [4] et [5] montrent que

$$E_\theta\left[\Gamma_i^j\right] = \Delta_i^j + \frac{(1-\Delta_i)\exp(-\Lambda_j(T_i)\exp(\beta_j' Z_i) + \gamma_j' X_i)}{\sum_{k=1}^{J} \exp(-\Lambda_k(T_i)\exp(\beta_k' Z_i) + \gamma_k' X_i)}.$$

Notons $h^j(u, \theta) = Y(u)\mathrm{e}^{\beta_j' Z} E_\theta[\Gamma^j]$, $h_i^j(u, \theta) = Y_i(u)\mathrm{e}^{\beta_j' Z_i} E_\theta[\Gamma_i^j]$, et

$$H_{j,n}(u, \theta) = n^{-1} \sum_{i=1}^{n} h_i^j(u, \theta) := \mathbb{P}_n h^j(u, \theta).$$

[4] et [5] montrent que l'estimateur du MVNP $\hat{\Lambda}_{j,n}$ de $\Lambda_{j,0}$ vérifie l'équation :

$$\hat{\Lambda}_{j,n}(t) = n^{-1} \sum_{i=1}^{n} \int_0^t \frac{\mathrm{d}N_i^j(u)}{H_{j,n}(u, \hat{\theta}_n)}, \quad j \in \mathcal{J}. \tag{1}$$

Dans [4] et [5], les auteurs montrent aussi que $\forall j \in \mathcal{J}$,

$$\Lambda_{j,0}(t) = \int_0^t \frac{P_0 \, \mathrm{d}N^j(u)}{H_j(u, \theta_0)},$$

où $H_j(u, \theta) = P_0 h^j(u, \theta)$ et $P_0 V$ désigne l'espérance d'une variable aléatoire V sous θ_0. L'expression (1) se révèle utile dans l'implémentation de l'algorithme itératif proposé par [4] et [5] pour atteindre l'estimation du MVNP de θ_0. Cette expression va aussi nous servir pour montrer la consistance de $\hat{\theta}_n$.

3. Consistance des estimateurs

Toutes les variables aléatoires sont définies sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$. Nous supposons que Z est bornée et pour chacune des J causes, on peut observer un décès à tout instant : $\inf_{j \in \mathcal{J}} P_0[Y(\tau)\Delta^j] > 0$. Nous supposons aussi les conditions d'identifiabilité suivantes : les matrices de variance–covariance de X et Z sont définies positives et $\forall j \in \mathcal{J}$, $P_0[\Delta^j | T, Z, X] > 0$ presque sûrement (p.s.).

Nous énonçons notre résultat principal ainsi qu'un lemme utile pour sa démonstration.

Lemme 3.1. *Pour tout* $j \in \mathcal{J}$, $\limsup_n \hat{\Lambda}_{j,n}(\tau) < \infty$ *p.s.*

Preuve. On a $H_{j,n}(u, \hat{\theta}_n) \geqslant m \mathbb{P}_n[Y(\tau)\Delta^j]$, où $m = \min_{\beta_j, Z} \mathrm{e}^{\beta_j Z}$, d'où p.s., $H_{j,n}(u, \hat{\theta}_n) \geqslant m P_0[Y(\tau)\Delta^j] + \mathrm{o}(1)$. Ainsi, p.s., $1/H_{j,n}(u, \hat{\theta}_n) \leqslant \mathrm{O}(1)$ et $\hat{\Lambda}_{j,n}(\tau) \leqslant \mathrm{O}(1)[P_0 N^j(\tau) + \mathrm{o}(1)]$, d'où le résultat. $\square$

Théorème 3.2. *On a, pour tout* $j \in \mathcal{J}$, $\hat{\beta}_{j,n} \xrightarrow{p.s.} \beta_{j,0}$, $\hat{\gamma}_{j,n} \xrightarrow{p.s.} \gamma_{j,0}$ *et* $\sup_{t \in [0,\tau]} |\hat{\Lambda}_{j,n}(t) - \Lambda_{j,0}(t)| \xrightarrow{p.s.} 0$.

Preuve. Nous montrons tout d'abord que de toute sous-suite de $(\hat{\theta}_n)$, on peut extraire une sous-suite $(\hat{\theta}_{n_k})$ qui converge vers un $\theta^* = \{(\beta_j^*, \gamma_j^*, \Lambda_j^*), j \in \mathcal{J}\}$. Notons $\hat{\xi}_{j,n} = (\hat{\beta}_{j,n}, \hat{\gamma}_{j,n})$.

La suite $(\hat{\xi}_{1,n}, \ldots, \hat{\xi}_{J,n})$ est bornée dans $\mathbb{R}^{J(p+q)}$, donc de toute sous-suite de $(\hat{\xi}_{1,n}, \ldots, \hat{\xi}_{J,n})$, on peut extraire une sous-suite $(\hat{\xi}_{1,m_k}, \ldots, \hat{\xi}_{J,m_k})$ qui converge vers une limite $(\xi_1^*, \ldots, \xi_J^*) \in (\mathcal{B} \times \mathcal{G})^J$.

D'après le Lemme 3.1 et le lemme de Helly, pour tout $j \in \mathcal{J}$, il existe p.s. une suite extraite $(\hat{\Lambda}_{j,l_k})$ de $(\hat{\Lambda}_{j,m_k})$ et une fonction croissante Λ_j^* telle que $\hat{\Lambda}_{j,l_k}(t) \to \Lambda_j^*(t)$ en tout $t \in [0, \tau]$ où Λ_j^* est continue. On peut identifier Λ_j^* en notant que $u \mapsto H_{j,l_k}(u, \hat{\theta}_{l_k})$ étant décroissante et bornée, il existe une suite extraite $(H_{j,n_k}(u, \hat{\theta}_{n_k}))$ de $(H_{j,l_k}(u, \hat{\theta}_{l_k}))$ et une fonction $H_j^*(u)$ telle que $H_{j,n_k}(u, \hat{\theta}_{n_k}) \to H_j^*(u)$ en tout point de continuité de H_j^*. Par le théorème de Lebesgue,

$$\hat{\Lambda}_{j,n_k}(t) = \int_0^t \frac{n_k^{-1} \sum_{i=1}^{n_k} \mathrm{d}N_i^j(u)}{H_{j,n_k}(u, \hat{\theta}_{n_k})} \longrightarrow \int_0^t \frac{P_0 \, \mathrm{d}N^j(u)}{H_j^*(u)} := \Lambda_j^*(t),$$

et ce, en tout $t \in [0, \tau]$ où Λ_j^* est continue. Montrons que Λ_j^* est continue sur $[0, \tau]$. On a p.s.

$$\left|\hat{\Lambda}_{j,n_k}(t) - \hat{\Lambda}_{j,n_k}(s)\right| \leqslant \mathrm{O}(1)\mathbb{P}_{n_k}\big(N^j(t) - N^j(s)\big) = \mathrm{O}(1) P_0\big(N^j(t) - N^j(s)\big) + \mathrm{o}(1), \quad \forall s, t \in [0, \tau].$$

Sous nos conditions de régularité, $P_0 N^j$ est continue, d'où p.s., $\Delta\hat{\Lambda}_{j,n_k}(s) = \mathrm{o}(1)$ $\forall s \in [0, \tau]$, où $\Delta\hat{\Lambda}_{j,n_k}(s)$ désigne le saut de $\hat{\Lambda}_{j,n_k}$ en s. Or $\forall s \in [0, \tau]$, $\Delta\hat{\Lambda}_{j,n_k}(s) \to \Delta\Lambda_j^*(s)$ d'où $\Delta\Lambda_j^*(s) = 0$ et Λ_j^* est continue sur $[0, \tau]$. Par le théorème de Dini, $\hat{\Lambda}_{j,n_k}$ converge uniformément vers Λ_j^* sur $[0, \tau]$, p.s.

Ainsi, de toute sous-suite de $(\hat{\theta}_n)$, on peut extraire une sous-suite $(\hat{\theta}_{n_k})$ qui converge vers un θ^*. Nous montrons maintenant que $\theta^* = \theta_0$. Posons

$$\bar{\Lambda}_{j,n}(t) = n^{-1} \sum_{i=1}^{n} \int_0^t \frac{\mathrm{d}N_i^j(u)}{H_{j,n}(u,\theta_0)}$$

et montrons que $\bar{\Lambda}_{j,n}$ converge p.s. uniformément vers $\Lambda_{j,0}$. On montre :

$$\sup_{t\in[0,\tau]} \left| \bar{\Lambda}_{j,n}(t) - \Lambda_{j,0}(t) \right| \leqslant \sup_{u\in[0,\tau]} \left| \frac{1}{H_{j,n}(u,\theta_0)} - \frac{1}{H_j(u,\theta_0)} \right| + \sup_{u\in[0,\tau]} \left| (\mathbb{P}_n - P_0)\left(\frac{\Delta^j 1\{T \leqslant u\}}{H_j(T,\theta_0)} \right) \right|.$$

On montre que la classe de fonctions $\{u \mapsto h^j(u,\theta_0)\colon\ u \in [0,\tau]\}$ est une classe de Glivenko–Cantelli (voir [6]), d'où $\sup_{u\in[0,\tau]} |H_{j,n}(u,\theta_0) - H_j(u,\theta_0)| \xrightarrow{\text{p.s.}} 0$. De plus, $H_j(u,\theta_0) \geqslant m P_0[Y(u)\Delta^j] > 0$, d'où

$$\sup_{u\in[0,\tau]} \left| \frac{1}{H_{j,n}(u,\theta_0)} - \frac{1}{H_j(u,\theta_0)} \right| \xrightarrow{\text{p.s.}} 0.$$

On montre aussi que

$$\left\{ u \mapsto \frac{\Delta^j 1\{T \leqslant u\}}{H_j(T,\theta_0)}\colon\ u \in [0,\tau] \right\}$$

est Glivenko–Cantelli. Donc finalement,

$$\sup_{t\in[0,\tau]} \left| \bar{\Lambda}_{j,n}(t) - \Lambda_{j,0}(t) \right| \xrightarrow{\text{p.s.}} 0.$$

Maintenant, notons $\bar{\theta}_n = \{(\beta_{j,0}, \gamma_{j,0}, \bar{\Lambda}_{j,n}),\ j \in \mathcal{J}\}$ et $p_*^{j,X}$ (respectivement $p_0^{j,X}$) la quantité $p^{j,X}$ calculée sous θ^* (respectivement θ_0). Alors on montre :

$$\begin{aligned}
0 &\leqslant n^{-1}\left[\ln \tilde{L}_{n_k}(\hat{\theta}_{n_k}) - \ln \tilde{L}_{n_k}(\bar{\theta}_{n_k})\right] \\
&\to P_0 \Bigg\{ \sum_{j=1}^{J} \Bigg[\int_0^\tau \ln \frac{H_j(u,\theta_0)}{H_j^*(u)} + (\beta_j^* - \beta_{j,0})' Z + \ln \frac{p_*^{j,X}}{p_0^{j,X}} \,\mathrm{d}N^j(u) - \left[\Lambda_j^*(T) \mathrm{e}^{\beta_j^{*\prime} Z} - \Lambda_{j,0}(T) \mathrm{e}^{\beta_{j,0}' Z} \right] \Delta^j \Bigg] \\
&\quad + (1-\Delta) \ln \left(\frac{\sum_{j=1}^J p_*^{j,X} \exp(-\Lambda_j^*(T) \mathrm{e}^{\beta_j^{*\prime} Z})}{\sum_{j=1}^J p_0^{j,X} \exp(-\Lambda_{j,0}(T) \mathrm{e}^{\beta_{j,0}' Z})} \right) \Bigg\} \quad \text{p.s.} \\
&= -P_0\left[\ln\left(L_1(\theta_0)/L_1(\theta^*)\right)\right] \leqslant 0.
\end{aligned}$$

On conclut en montrant que $P_0[\ln(L_1(\theta_0)/L_1(\theta^*))] = 0$ p.s. implique $\theta^* = \theta_0$ (identifiabilité). L'identifiabilité des paramètres se montre en adoptant le même principe de démonstration que [3] et [7]. Ainsi, pour presque tout $\omega \in \Omega$, de toute sous-suite de $(\hat{\theta}_n)$, on peut extraire une sous-suite $(\hat{\theta}_{n_k})$ qui converge vers θ_0, d'où $\hat{\theta}_n \xrightarrow{\text{p.s.}} \theta_0$. $\square$

Références

[1] P.K. Andersen, O. Borgan, R.D. Gill, N. Keiding, Statistical Models Based on Counting Processes, Springer, 1993.

[2] K.C. Choi, X. Zhou, Large sample properties of mixture models with covariates for competing risks, J. Multivariate Anal. 82 (2002) 331–366.

[3] J.-F. Dupuy, I. Grama, M. Mesbah, Asymptotic theory for the Cox model with missing time-dependent covariate, Ann. Statist. 34 (2006) 903–924.

[4] G. Escarela, R. Bowater, Fitting a semi-parametric mixture model for competing risks in survival data, Comm. Statist. Theory Methods, in press.

[5] S.K. Ng, G.J. McLachlan, An EM-based semi-parametric mixture model approach to the regression analysis of competing risks data, Statist. Med. 22 (2003) 1097–1111.

[6] A.W. Van der Vaart, J.A. Wellner, Weak Convergence and Empirical Processes, Springer, Berlin, 1996.

[7] D. Zeng, J. Cai, Asymptotic results for maximum likelihood estimators in joint analysis of repeated measurements and survival time, Ann. Statist. 33 (2005) 2132–2163.