Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
C. R. Acad. Sci. Paris, Ser. I 336 (2003) 81–84
COMPTES RENDUS MATHEMATIQUE

Statistique/Probabilités

Normalité asymptotique des estimateurs semiparamétriques dans le modèle de Cox avec covariable manquante nonignorable

Asymptotic normality of semiparametric estimators in the Cox model with nonignorable missing covariate

Jean-François Dupuy [a], Ion Grama [b], Mounir Mesbah [b]

[a] *CEREMADE, Université Paris Dauphine, 75775 Paris cedex 16, France*
[b] *Laboratoire de statistique, Université de Bretagne-Sud, 56000 Vannes, France*

Reçu le 9 août 2002 ; accepté après révision le 3 décembre 2002

Présenté par Paul Deheuvels

Résumé

Soit T une variable aléatoire représentant une durée jusqu'à un événement. Dans le modèle de Cox $\lambda_T(t)\,\mathrm{e}^{\beta Z(t)}$, si la valeur de Z à l'instant d'événement n'est pas observée, Dupuy et Mesbah (Lifetime Data Analysis 8 (2002) 99–115) ont proposé d'estimer les paramètres β et $\Lambda_T(t)=\int_0^t \lambda_T(s)\,\mathrm{d}s$ en maximisant une vraisemblance obtenue en modélisant conjointement les durées et la covariable $(Z(t))_{t\geqslant 0}$. Nous montrons que les estimateurs obtenus ont une distribution asymptotique normale. ***Pour citer cet article : J.-F. Dupuy et al., C. R. Acad. Sci. Paris, Ser. I 336 (2003).***

Abstract

Let T denote a random duration until some event of interest. In the Cox model $\lambda_T(t)\,\mathrm{e}^{\beta Z(t)}$, if the value of Z at event time is unobserved, Dupuy and Mesbah (Lifetime Data Analysis 8 (2002) 99–115) have proposed to estimate the parameters β and $\Lambda_T(t)=\int_0^t \lambda_T(s)\,\mathrm{d}s$ by maximizing a likelihood obtained from a joint model for survival and the longitudinal covariate data. We show that the estimators derived from this joint likelihood are asymptotically normally distributed. ***To cite this article: J.-F. Dupuy et al., C. R. Acad. Sci. Paris, Ser. I 336 (2003).***

1. Introduction

Le modèle de Cox définit la loi d'un instant d'événement T conditionnelle à une covariable dépendant du temps $(Z(t))_{t\geqslant 0}$ par la fonction de risque instantané

Adresses e-mail : dupuy@ceremade.dauphine.fr (J.-F. Dupuy), ion.grama@univ-ubs.fr (I. Grama), mounir.mesbah@univ-ubs.fr (M. Mesbah).

1631-073X/03/$ – see front matter © 2003 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS. Tous droits réservés.
doi:10.1016/S1631-073X(03)00003-7

$$\lambda_{T|Z}(t) = \lambda_T(t)\,\mathrm{e}^{\beta Z(t)}, \tag{1}$$

où β est un paramètre de régression et λ_T une fonction de risque de base considérée comme un paramètre de nuisance. Cox [4] a proposé d'estimer β au vu d'un n-échantillon de données censurées $(X_i, \delta_i, \{Z_i(s), 0 \leqslant s \leqslant X_i\})_{1\leqslant i\leqslant n}$, où $X_i = T_i \wedge C_i$, $\delta_i = 1_{\{T_i \leqslant C_i\}}$, et C_i représente le temps de censure, en maximisant une vraisemblance partielle qui ne fait pas intervenir $\lambda_T(\cdot)$. Breslow [3] a proposé l'estimateur $\hat{\Lambda}_T(t) = \sum_{X_i \leqslant t} \delta_i / [\sum_{j=1}^n \mathrm{e}^{\hat{\beta} Z_j(X_i)} 1_{\{X_i \leqslant X_j\}}]$ de la fonction de risque cumulé $\Lambda_T(t) = \int_0^t \lambda_T(s)\,\mathrm{d}s$. Andersen et Gill [1] ont montré la consistance et la normalité asymptotique de l'estimateur du maximum de vraisemblance partielle de β et de l'estimateur de Breslow de $\Lambda_T(\cdot)$.

Si les valeurs $\{Z_i(X_i)\colon\ i = 1, \ldots, n\}$ aux instants d'événement ne sont pas observées, Dupuy et Mesbah [6] proposent de modéliser conjointement (X, δ) et $Z(\cdot)$ pour en déduire une vraisemblance permettant d'estimer β et $\Lambda_T(\cdot)$ au vu des données incomplètes. Une première justification de cette méthode est apportée par Dupuy, Grama et Mesbah [5], qui montrent la consistance des estimateurs de β et $\Lambda_T(\cdot)$ obtenus en maximisant la vraisemblance proposée dans [6].

Nous poursuivons ici la justification de l'approche développée dans [6], en établissant la normalité asymptotique de ces estimateurs. Une autre solution au problème de l'estimation dans le modèle de Cox avec une covariable manquante nonignorable est proposée par Scharfstein, Rotnitzky et Robins [8]. Elle repose sur une analyse de sensibilité qui nécessite la connaissance d'un intervalle $[-b, b]$ contenant la vraie valeur de β. Notons que la méthode étudiée ici n'impose pas cette restriction.

2. Estimation semiparamétrique dans le modèle de Cox avec covariable manquante nonignorable

1. Supposons que l'on observe une covariable $Z(\cdot)$ sur n sujets dans l'intervalle de temps $[0, \tau]$. Supposons que $Z(\cdot)$ soit une fonction en escalier de valeur constante $Z(t) = Z_j$ sur les intervalles $(t_{j-1}, t_j]$, où t_j est l'instant d'observation de Z_j $(j = 1, \ldots, K,\ t_0 = 0,\ t_K = \tau)$. Z_0 est une valeur initiale mesurée en t_0. La valeur Z_j n'est pas mesurée si $X \leqslant t_j$.

Soit $a_t = \max(k\colon\ t_k < t)$ $(t > 0)$ l'indice de la dernière observation de $Z(\cdot)$ avant t. L'ensemble $\{Z(s),\ 0 \leqslant s \leqslant X\}$ se ramène à $\{Z_0, \ldots, Z_{a_X}, Z(X)\}$ mais les observations portent sur le vecteur aléatoire $Y = (X, \delta, Z_0, \ldots, Z_{a_X})$. L'observation de $Z(\cdot)$ à l'instant X est manquante et la loi de X en dépend : on dit que $Z(X)$ est une valeur manquante nonignorable (cette définition est rappelée dans [8]).

Le problème statistique est d'estimer β et $\Lambda_T(\cdot)$ dans le modèle (1) au vu des observations incomplètes fournies par n répliques indépendantes $(Y_i)_{1\leqslant i\leqslant n}$.

2. Supposons que pour chaque $j \geqslant 1$, le vecteur aléatoire $(Z_0, \ldots, Z_j)$ à valeurs dans $\mathbb{R}^{j+1}$ admette une densité $f(Z_0, \ldots, Z_j; \alpha)$ sur $\mathbb{R}^{j+1}$, paramétrée par $\alpha \in \mathbb{R}^p$ $(p \geqslant 1)$. Des conditions de régularité sur $Z(\cdot)$ et f sont données dans [5] et [6]. Dupuy et Mesbah [6] construisent un modèle conjoint pour $(X, \delta, Z(\cdot))$ à partir du modèle (1). Ce modèle conjoint est paramétré par $\theta = (\beta, \Lambda_T(\cdot), \alpha)$, où β et $\Lambda_T(\cdot)$ sont les paramètres d'intérêt du modèle.

Dupuy et Mesbah [6] obtiennent la vraisemblance de θ au vu d'une observation incomplète Y en intégrant la densité du vecteur complet $(Y, Z(X))$ sur l'ensemble des valeurs de la variable inobservée $Z(X)$. Nous supposons que la distribution de C ne dépend pas de θ (censure non informative), que T et C sont indépendants conditionnellement à $Z(\cdot)$ et que C ne dépend pas des valeurs inobservées de $Z(\cdot)$.

3. Contrairement à la vraisemblance partielle de Cox, la vraisemblance obtenue dans [6] (nous la notons L_n par la suite) dépend du paramètre fonctionnel $\Lambda_T(\cdot)$. Pour trouver un maximum de L_n, l'espace de recherche d'un estimateur de $\Lambda_T(\cdot)$ est restreint à l'ensemble des fonctions en escalier $\Lambda_{T,n}(\cdot)$ positives, croissantes sur $[0, \infty)$, et dont les sauts $\Delta\Lambda_{T,n}(X_k)$ se produisent aux $p(n)$ $(p(n) \leqslant n)$ instants d'événements non censurés X_k $(k = 1, \ldots, p(n))$. L'estimateur de θ obtenu sera dit estimateur du maximum de vraisemblance semiparamétrique. Notons $\hat{\theta}_n =$

$(\hat\beta_n, \hat\Lambda_{T,n}(\cdot), \hat\alpha_n)$ cet estimateur, obtenu en maximisant sur l'espace $\{(\beta, \Delta\Lambda_{T,n}(X_1), \ldots, \Delta\Lambda_{T,n}(X_{p(n)}), \alpha)$: $\beta \in \mathbb{R},\ \Delta\Lambda_{T,n}(X_k) \in \mathbb{R},\ k = 1, \ldots, p(n),\ \alpha \in \mathbb{R}^p\}$ la vraisemblance

$$L_n(\theta) = \prod_{i=1}^{n} \int \Delta\Lambda_{T,n}^{\delta_i}(X_i) \exp\left[\delta_i \beta Z(X) - \sum_{k=1}^{p(n)} \Delta\Lambda_{T,n}(X_k)\, \mathrm{e}^{\beta Z_i(X_k)} 1_{\{X_k \leqslant X_i\}}\right] \times f(Z_{i0}, \ldots, Z_{ia_{X_i}}, Z(X); \alpha)\, \mathrm{d}Z(X).$$

4. Soit $VB[0,\tau]$ l'espace des applications à variation bornée de $[0,\tau]$ dans $\mathbb{R}$, et $H = \{h = (h_1, h_2, h_3)$: $h_1 \in \mathbb{R},\ h_2 \in VB[0,\tau],\ h_3 \in \mathbb{R}^p\}$. Si X_1 et X_2 sont deux variables aléatoires, nous notons $E_\theta[g(X_1)|X_2]$ l'espérance conditionnelle de $g(X_1)$ sachant X_2, paramétrée par θ.

Dupuy et Mesbah [6] montrent que $\hat\theta_n$ peut être obtenu en résolvant l'équation d'estimation $S_{n,\theta}(\theta)(h) = 0$, où

$$S_{n,\tilde\theta}(\theta)(h) = \frac{1}{n} \sum_{i=1}^{n} E_{\tilde\theta}\left[h_1 \delta Z(X) - \int_0^X h_1 Z(u)\, \mathrm{e}^{\beta Z(u)}\, \mathrm{d}\Lambda_T(u) - \int_0^X h_2(u)\, \mathrm{e}^{\beta Z(u)}\, \mathrm{d}\Lambda_T(u) + \delta h_2(X) + h_3' \frac{\partial}{\partial\alpha} \ln f\big(Z_0, \ldots, Z_{a_X}, Z(X); \alpha\big) \Big| Y_i\right].$$

Soit

$$S_{\tilde\theta}(\theta)(h) = E_{\tilde\theta}\left[h_1 \delta Z(X) - \int_0^X h_1 Z(u)\, \mathrm{e}^{\beta Z(u)}\, \mathrm{d}\Lambda_T(u) - \int_0^X h_2(u)\, \mathrm{e}^{\beta Z(u)}\, \mathrm{d}\Lambda_T(u) + \delta h_2(X) + h_3' \frac{\partial}{\partial\alpha} \ln f\big(Z_0, \ldots, Z_{a_X}, Z(X); \alpha\big)\right].$$

Dans le cas particulier où la covariable $Z(\cdot)$ est complètement observée, le maximum de vraisemblance semiparamétrique revient à estimer β par le maximum de vraisemblance partielle (en pratique, un algorithme de type Newton est utilisé pour calculer cette estimation) et $\Lambda_T(\cdot)$ par l'estimateur de Breslow [2].

En revanche, pour calculer les estimations de β, $\Lambda_T(\cdot)$ et α à partir de la vraisemblance $L_n(\theta)$, pour laquelle il n'existe pas de forme explicite, Dupuy et Mesbah [6] utilisent l'algorithme itératif EM. Les auteurs donnent les formules nécessaires à sa mise en oeuvre (formules explicites de calcul des valeurs itérées des estimations de $\Lambda_T(\cdot)$ et α, algorithme de Newton–Raphson pour estimer β à chaque itération de l'EM, formules de quadrature pour l'approximation des intégrales). Notons que pour estimer les paramètres d'intérêt β et $\Lambda_T(\cdot)$, il est nécessaire d'estimer α, qui intervient dans le calcul des espérances E_θ.

3. Distribution asymptotique des estimateurs semiparamétriques

Soit H_{12} l'ensemble des $h \in H$ tels que $h = (h_1, h_2, 0)$. Soit $l^\infty(H_{12})$ l'ensemble des applications bornées de H_{12} dans $\mathbb{R}$, muni de la norme $\|U\| = \sup_{h \in H_{12}} |U(h)|$. Notons $\theta_0 = (\beta_0, \Lambda_{T,0}(\cdot), \alpha_0)$ la vraie valeur de θ. Murphy [7] a montré la consistance et la normalité asymptotique des estimateurs des paramètres du modèle de fragilité, en considérant $\Lambda_T(\cdot)$ comme un élément de $l^\infty(VB[0,\tau])$ défini par $h \mapsto \int_0^\tau h(u)\, \mathrm{d}\Lambda_T(u)$. De même, nous allons utiliser cette approche ici pour gérer le problème de la dimension de l'estimateur $\hat\theta_n$ et du paramètre θ. Nous considérons le paramètre θ comme un élément de $l^\infty(H)$ défini par $h \mapsto \theta(h) = h_1\beta + \int_0^\tau h_2(u)\, \mathrm{d}\Lambda_T(u) + h_3'\alpha$. L'espace des paramètres θ est maintenant un sous-espace Θ de $l^\infty(H)$ et $S_{n,\theta}$ et S_θ sont deux applications de Θ dans $l^\infty(H)$.

Notre résultat est le suivant :

Théorème 3.1. *La suite* $(\sqrt{n}(\hat{\beta}_n - \beta_0), \sqrt{n}(\hat{\Lambda}_{T,n} - \Lambda_{T,0}))$ *converge en loi dans* $l^\infty(H_{12})$ *vers un processus gaussien centré* G*, de fonction de covariance*

$$\operatorname{cov}\big[G(h), G(h^*)\big] = \sigma_{1,\theta_0}^{-1}(h^*)h_1 + \int_0^\tau h_2(u)\sigma_{2,\theta_0}^{-1}(h^*)(u)\,\mathrm{d}\Lambda_{T,0}(u),$$

où $\sigma_{\theta_0}^{-1} = (\sigma_{1,\theta_0}^{-1}, \sigma_{2,\theta_0}^{-1})$ *est l'inverse de l'opérateur* $\sigma_{\theta_0} = (\sigma_{1,\theta_0}, \sigma_{2,\theta_0})$ *de* H_{12} *dans* H_{12} *défini par* $\sigma_{1,\theta_0}(h) = E_{\theta_0}[\int_0^X Z(u)\,\mathrm{e}^{\beta_0 Z(u)}[h_1 Z(u) + h_2(u)]\,\mathrm{d}\Lambda_{T,0}(u)]$, $\sigma_{2,\theta_0}(h)(u) = E_{\theta_0}[[h_1 Z(u) + h_2(u)]\,\mathrm{e}^{\beta_0 Z(u)} 1_{\{u \leqslant X\}}]$.

Démonstration. Nous nous intéressons ici à la distribution asymptotique de $(\sqrt{n}(\hat{\beta}_n - \beta_0), \sqrt{n}(\hat{\Lambda}_{T,n} - \Lambda_{T,0}))$. Nos résultats sont donc énoncés sur la restriction $l^\infty(H_{12})$ de l'espace $l^\infty(H)$. Dupuy, Grama et Mesbah [5] présentent un résultat de convergence en loi de $(\sqrt{n}(\hat{\beta}_n - \beta_0), \sqrt{n}(\hat{\Lambda}_{T,n} - \Lambda_{T,0}), \sqrt{n}(\hat{\alpha}_n - \alpha_0))$ dans l'espace $l^\infty(H)$.

La démonstration repose sur le Théorème 3.3.1 de Van der Vaart et Wellner [9], qui énonce des conditions suffisantes pour la convergence en loi d'estimateurs obtenus comme solutions d'une équation d'estimation $S_n(\phi) = 0$. Nous montrons dans les propositions suivantes que les conditions de ce théorème sont vérifiées pour nos estimateurs (rappelons que la consistance de $\hat{\theta}_n$ a été montrée dans [5]).

Nous montrons tout d'abord le résultat suivant sur la distribution asymptotique du score $\sqrt{n}(S_{n,\hat{\theta}_n}(\theta_0) - S_{\theta_0}(\theta_0))$.

Proposition 3.2. *La suite* $\sqrt{n}(S_{n,\hat{\theta}_n}(\theta_0) - S_{\theta_0}(\theta_0))$ *converge en loi dans* $l^\infty(H_{12})$ *vers un processus gaussien centré* $\mathcal{G}$ *de fonction de covariance* $\operatorname{cov}(\mathcal{G}(h), \mathcal{G}(h^*)) = h_1\sigma_{1,\theta_0}(h^*) + \int_0^\tau h_2(u)\sigma_{2,\theta_0}(h^*)(u)\,\mathrm{d}\Lambda_{T,0}(u)$.

Nous montrons ensuite que S_{θ_0} est différentiable au sens de Fréchet. Sa différentielle $\dot{S}_{\theta_0}$ joue un rôle analogue à celui de la matrice d'information de Fisher dans le cas paramétrique.

Proposition 3.3. *L'opérateur* $S_{\theta_0} : \theta \mapsto S_{\theta_0}(\theta)$ *de* Θ *dans* $l^\infty(H_{12})$ *est différentiable au sens de Fréchet en* θ_0*, et sa différentielle* $\dot{S}_{\theta_0}(\theta_0)$ *au point* θ_0 *est donnée par :* $\dot{S}_{\theta_0}(\theta_0)(\theta)(h) = -\beta\sigma_{1,\theta_0}(h) - \int_0^\tau \sigma_{2,\theta_0}(h)(u)\,\mathrm{d}\Lambda_T(u)$.

Nous établissons ensuite la propriété suivante de l'opérateur σ_{θ_0}.

Lemme 3.4. *L'opérateur* σ_{θ_0} *de* H_{12} *dans* H_{12} *est inversible et son inverse* $\sigma_{\theta_0}^{-1}$ *est continue.*

Ce lemme nous permet finalement de montrer la dernière condition du Théorème 3.3.1 de Van der Vaart et Wellner [9] :

Proposition 3.5. *L'opérateur* $\dot{S}_{\theta_0}(\theta_0) : \Theta \to l^\infty(H_{12})$ *est inversible et son inverse est continue sur l'image de* $\dot{S}_{\theta_0}(\theta_0)$.

Références

[1] P.K. Andersen, R.D. Gill, Cox's regression model for counting processes: a large sample study, Ann. Statist. 10 (1982) 1100–1120.
[2] V. Bagdonavičius, M.S. Nikulin, Accelerated Life Models: Modeling and Statistical Analysis, Chapman & Hall, 2001.
[3] N.E. Breslow, Contribution to the discussion on the paper by D.R. Cox, J. Roy. Statist. Soc. Ser. B 34 (1972) 216–217.
[4] D.R. Cox, Regression models and life-tables (with discussion), J. Roy. Statist. Soc. Ser. B 34 (1972) 187–220.
[5] J.-F. Dupuy, I. Grama, M. Mesbah, Asymptotic theory for the Cox model with missing time-dependent covariate, 2002, soumis.
[6] J.-F. Dupuy, M. Mesbah, Joint modeling of event time and nonignorable missing longitudinal data, Lifetime Data Analysis 8 (2002) 99–115.
[7] S.A. Murphy, Asymptotic theory for the frailty model, Ann. Statist. 23 (1995) 182–198.
[8] D.O. Scharfstein, A. Rotnitzky, J.M. Robins, Adjusting for nonignorable drop-out using semiparametric nonresponse models, J. Amer. Statist. Assoc. 94 (1999) 1096–1120.
[9] A.W. Van der Vaart, J.A. Wellner, Weak Convergence and Empirical Processes, Springer, Berlin, 1996.