Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
C. R. Acad. Sci. Paris, Ser. I 346 (2008) 567–570
Comptes Rendus Mathematique
http://france.elsevier.com/direct/CRASS1/

Statistique

Test du log-rank et statistique du score de vraisemblance partielle pour une durée de survie dépendant d'un taux de dégradation

Véronique Boisson, Mounir Mesbah

L.S.T.A., Université Pierre et Marie Curie – Paris 6, 175, rue du Chevaleret, boîte 158, 75013 Paris, France

Reçu le 20 décembre 2007 ; accepté le 26 mars 2008

Présenté par Paul Deheuvels

Résumé

Nous considérons une population pour laquelle la durée de survie, $T(\cdot)$, dépend d'un taux de dégradation de qualité de vie x préalablement fixé. Ainsi, nous obtenons la durée de survie $T(x)$. Nous nous intéressons à la généralisation de la statistique du score de vraisemblance partielle pour la durée de survie $T(x)$ puis nous établissons les propriétés asymptotiques des estimateurs. Nous obtenons également la statistique de test du log-rank pour un taux de dégradation de qualité de vie x fixé. ***Pour citer cet article : V. Boisson, M. Mesbah, C. R. Acad. Sci. Paris, Ser. I 346 (2008).***

Abstract

Log-rank test and partial likelihood score statistic for failure time dependent on a threshold of degradation. We consider a population where the survival time, $T(\cdot)$, is depending of a fixed critical rate x of degradation of quality of life. So we obtain a survival time $T(x)$. We are interested in extending the partial likelihood score statistic to survival time $T(x)$ and then we establish properties of estimators. We establish also the log-rank statistical test for a fixed critical rate of degradation of quality of life x. ***To cite this article: V. Boisson, M. Mesbah, C. R. Acad. Sci. Paris, Ser. I 346 (2008).***

1. Introduction

La qualité de vie occupe une place de plus en plus importante lors d'études cliniques et plus particulièrement pour celles concernant les maladies chroniques. Celle-ci est mesurée à l'aide de questionnaires spécifiques à des dates t_j ($j = 0, \dots, T$) préalablement déterminées. L'évolution de la qualité de vie q_j à la date t_j par rapport à la date initiale t_0 est reflétée par la notion de dégradation de qualité de vie à la date t_j définie par $d_j = (q_0 - q_j)q_0^{-1}$. Un nouveau problème est de connaître la date où le score de dégradation de qualité de vie d'un patient dépasse un taux critique x préalablement fixé.

Adresses e-mail : veronique.boisson@upmc.fr (V. Boisson), mounir.mesbah@upmc.fr (M. Mesbah).

1631-073X/$ – see front matter © 2008 Académie des sciences. Publié par Elsevier Masson SAS. Tous droits réservés.
doi:10.1016/j.crma.2008.03.028

La théorie des distributions asymptotiques pour la statistique de test basée sur le score de la vraisemblance partielle [4] a été donnée par Tsiatis [8]. Les propriétés asymptotiques des estimateurs $\hat{\beta}$ et $\hat{\Lambda}$ ont été obtenues par Tsiatis [7] et étendues aux processus de comptage multivariés par Andersen et Gill [1].

Dans cette Note, nous étendons la statistique de test du score de vraisemblance partielle, ainsi que les propriétés asymptotiques établies par Andersen et Gill [1], au cas d'une durée de survie dépendant d'un taux de dégradation x préalablement fixé. Nous obtenons ensuite la statistique de test du log-rank pour un taux de dégradation x fixé.

2. Statistique du score de vraisemblance partielle pour un taux de dégradation x fixé et propriétés asymptotiques

Soit x le taux critique de dégradation de qualité de vie dont la valeur est préalablement fixée et soit $T(x) = \inf\{t; q_0^{-1}(q_0 - q_t) > x\}$ la durée de survie correspondante. Notons par τ la réalisation de $T(x)$. Soit C la variable de censure à droite indépendante de la durée de survie $T(x)$ et $\mathbf{Z}$ le vecteur des covariables. Les observations consistent en le triplet $(X(x), \delta(x), \mathbf{Z})$ où $X(x) = T(x) \wedge C$ et $\delta(x) = \mathbf{1}_{(T(x) \leqslant C)}$. Nous nous plaçons dans le cadre classique du modèle de Cox [3] i.e. en notant β le vecteur de dimension p des paramètres de régression, la fonction de risque instantané λ est définie par $\lambda(t, x/\mathbf{Z}(t)) = \lambda_0(t, x)\exp\{\beta'\mathbf{Z}(t)\}$.

Nous considérons une population de n patients indépendants. Les informations contenues dans la paire $(X_i(x), \delta_i(x))$ étant équivalentes à la connaissance des informations fournies par le processus de comptage $N_i(t, x) = \mathbf{1}_{(X_i(x) \leqslant t, \delta_i(x)=1)}$ et le processus de censure $Y_i(t, x) = \mathbf{1}_{(X_i(x) \geqslant t)}$, les observations consistent en les n triplets $\{N_i, Y_i, \mathbf{Z}_i\}$.

Pour tout $k = 0, 1, 2$ nous introduisons les notations,

$$\mathbf{S}^{(k)}(\beta, t, x) = \frac{1}{n}\sum_{i=1}^{n}\bigl\{\mathbf{Z}_i(t)\bigr\}^{\otimes k} Y_i(t, x)\exp\bigl\{\beta'\mathbf{Z}_i(t)\bigr\}, \tag{1}$$

$$\mathbf{E}(\beta, t, x) = \frac{\mathbf{S}^{(1)}(\beta, t, x)}{S^{(0)}(\beta, t, x)} \quad \text{et} \quad \mathbf{V}(\beta, t, x) = \frac{\mathbf{S}^{(2)}(\beta, t, x)}{S^{(0)}(\beta, t,)} - \bigl\{\mathbf{E}(\beta, t, x)\bigr\}^{\otimes 2}. \tag{2}$$

avec pour tout vecteur $\mathbf{X}$, $\mathbf{X}^{\otimes 0} = 1$, $\mathbf{X}^{\otimes 1} = \mathbf{X}$ et $\mathbf{X}^{\otimes 2} = \mathbf{X}\mathbf{X}'$. Nous notons également par β_0 la vraie valeur de β, par $D[0, \tau]$ l'espace des fonctions continues à droite et limitées à gauche et par $\mathbb{P}$ la convergence en probabilité.

Dans cette Note nous supposons que les conditions suivantes sont vérifiées :

(C1) Le temps τ est tel que, $\int_0^\tau \lambda_0(u, x)\,\mathrm{d}u < \infty$.

(C2) Pour $\mathbf{S}^{(j)}$, $j = 0, 1, 2$, définie en (1) il existe un voisinage $\mathcal{B}$ de β_0, un voisinage $\mathcal{X}$ de x et respectivement un scalaire, un vecteur et une fonction matricielle $s^{(0)}$, $\mathbf{s}^{(1)}$ et $\mathbf{s}^{(2)}$ définis sur $\mathcal{B} \times [0, \tau] \times \mathcal{X}$ tels que pour $j = 0, 1, 2$, $\sup_{u\in[0,\tau],\, \beta\in\mathcal{B},\, x\in\mathcal{X}} \|\mathbf{S}^{(j)}(\beta, u, x) - \mathbf{s}^{(j)}(\beta, u, x)\|_\infty \xrightarrow[n\to\infty]{\mathbb{P}} 0$.

(C3) Il existe $\delta > 0$ tel que, $n^{-1/2}\sup_{1\leqslant i\leqslant n,\, 0\leqslant u\leqslant\tau,\, x\in\mathcal{X}} |\mathbf{Z}_i(u)| Y_i(u, x)\mathbf{1}_{\{\beta_0'\mathbf{Z}_i(u) > -\delta|\mathbf{Z}_i(u)|\}} \xrightarrow[n\to\infty]{\mathbb{P}} 0$.

(C4) Soient $\mathcal{B}$, $\mathcal{X}$ et $\mathbf{s}^{(j)}$, $j = 0, 1, 2$, définie en (C2), posons $\mathbf{e} = \mathbf{s}^{(1)}/s^{(0)}$ et $\mathbf{v} = \mathbf{s}^{(2)}/s^{(0)} - \mathbf{e}^{\otimes 2}$ alors, pour tout $\beta \in \mathcal{B}$, $x \in \mathcal{X}$ et $0 \leqslant u \leqslant \tau$, $\frac{\partial}{\partial\beta}s^{(0)}(\beta, u, x) = \mathbf{s}^{(1)}(\beta, u, x)$ et $\frac{\partial^2}{\partial\beta^2}s^{(0)}(\beta, u, x) = \mathbf{s}^{(2)}(\beta, u, x)$.

(C5) Les fonctions $\mathbf{s}^{(j)}$, $j = 0, 1, 2$, sont bornées sur $\mathcal{B} \times [0, \tau] \times \mathcal{X}$ et de plus, $s^{(0)}$ ne s'annule pas sur $\mathcal{B} \times [0, \tau] \times \mathcal{X}$; pour $j = 0, 1, 2$, la famille de fonctions $\mathbf{s}^{(j)}(\cdot, u, x)$, $0 \leqslant u \leqslant \tau$ et $x \in \mathcal{X}$ est équicontinue au point β_0.

(C6) La matrice $\boldsymbol{\Sigma}(\beta_0, \tau, x) = \int_0^\tau \mathbf{v}(\beta_0, u, x)s^{(0)}(\beta_0, u, x)\lambda_0(u, x)\,\mathrm{d}u$ est définie positive.

Nous sommes en mesure d'établir le résultat principal de cette section :

Théorème 2.1. (i) *Nous supposons que les conditions sont vérifiées alors, le processus*

$$\frac{1}{\sqrt{n}}\mathbf{U}(\beta_0, t, x) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^t \bigl\{\mathbf{Z}_i(u) - \mathbf{E}(\beta_0, u, x)\bigr\}N_i(\mathrm{d}u, x),$$

converge faiblement dans $D[0,\tau]^p$ *vers un processus gaussien p-varié de moyenne nulle tel que, chaque processus est à accroissement indépendant. La matrice de covariance vaut*

$$\{\boldsymbol{\Sigma}(\beta_0,t,x)\}_{ll'} = \int_0^t \{\mathbf{v}(\beta_0,u,x)\}_{ll'} s^{(0)}(\beta_0,s)\lambda_0(u,x)\,\mathrm{d}u.$$

(ii) *Si de plus,* $\hat{\beta}$ *est un estimateur consistant de* β_0 *alors,*

$$\sup_{0\leqslant t\leqslant\tau}\left\|\frac{1}{n}\int_0^t \sum_{i=1}^n \mathbf{V}(\hat{\beta},u,x)N_i(\mathrm{d}u,x) - \boldsymbol{\Sigma}(\beta_0,t,x)\right\|_\infty \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

Élément de preuve du Théorème 2.1. La partie (i) est une application du Théorème 5.3.5 de Fleming et Harrington [5] tandis que la partie (ii) utilise l'Inégalité de Lenglart [6]. □

Remarque. Comme $\mathcal{I}(\beta,t,x) = \int_0^t \sum_{i=1}^n \mathbf{V}(\beta,u,x)N_i(\mathrm{d}u,x)$, où $\mathcal{I}$ est la matrice d'information de Fisher, la partie (ii) du Théorème 2.1 montre que $n^{-1}\mathcal{I}(\beta_0,t,x)$ et $n^{-1}\mathcal{I}(\hat{\beta},t,x)$ sont deux estimateurs uniformément consistants sur $[0,\tau]$ et de variance $\boldsymbol{\Sigma}(\beta_0,\tau,x)$ pour le processus de score en $\beta=\beta_0$.

Nous obtenons la consistance et la normalité asymptotique de $\hat{\beta}$ l'estimateur de β :

Théorème 2.2 *(Consistance). Soient* $\hat{\beta}$ *l'estimateur du maximum de vraisemblance partielle de* β *et* β_0 *la vraie valeur du paramètre* β *pour le modèle d'intensité multiplicative alors,* $\hat{\beta} \xrightarrow[n\to\infty]{\mathbb{P}} \beta_0$.

Élément de preuve du Théorème 2.2. Le théorème repose sur la convergence du log de la vraisemblance partielle pour une fonction maximisée en $\beta=\beta_0$ et sur un lemme de convexité se trouvant dans l'annexe de Andersen et Gill [1]. □

Théorème 2.3 *(Normalité asymptotique). Soit* $\boldsymbol{\Sigma}(\beta_0,t,x)$ *définie comme dans la Condition* (C6) *alors,* $\sqrt{n}\,(\hat{\beta}-\beta_0)$ *converge en distribution lorsque n tend vers l'infini, vers une variable gaussienne p-variée centrée de matrice de covariance* $\{\boldsymbol{\Sigma}(\beta_0,\tau,x)\}^{-1}$.

Un estimateur $\hat{\Lambda}_0$ de la fonction de risque cumulée Λ_0 a été donné par Breslow [2] ; une généralisation de cet estimateur prenant en compte le score x est donnée par,

$$\hat{\Lambda}_0(t,x) = \int_0^t \left\{\sum_{i=1}^n Y_i(u,x)\exp\bigl(\hat{\beta}'\mathbf{Z}_i(u)\bigr)\right\}^{-1}\left\{\sum_{i=1}^n N_i(\mathrm{d}u,x)\right\}.$$

Nous établissons la convergence faible de $\sqrt{n}\,(\hat{\Lambda}_0-\Lambda_0)$.

Théorème 2.4 *(Convergence faible). Nous supposons que les conditions sont vérifiées alors,* $\sqrt{n}\,(\hat{\Lambda}_0-\Lambda_0)$ *converge faiblement sur* $D[0,\tau]$ *vers un processus gaussien de moyenne nulle, à accroissement indépendant et de matrice de covariance,*

$$\int_0^t \{s_0(\beta_0,u,x)\}^{-1}\lambda_0(u,x)\,\mathrm{d}u + \mathbf{Q}'(\beta_0,t,x)\boldsymbol{\Sigma}^{-1}(\beta_0,\tau,x)\mathbf{Q}(\beta_0,t,x),$$

où $\mathbf{Q}(\beta_0,t,x) = \int_0^t \mathbf{e}(\beta_0,u,x)\lambda_0(u,x)\,\mathrm{d}u$.

3. Statistique du test du log-rank pour un taux de dégradation x fixé

Dans cette section, pour un taux de dégradation x préalablement fixé, nous cherchons à tester l'égalité des fonctions de survie de deux groupes A et B de patients. Par conséquent, l'hypothèse nulle à tester est $H_0 : S_A(t,x) = S_B(t,x)$ contre l'hypothèse alternative $H_1 : S_A(t,x) \neq S_B(t,x)$. Nous supposons que les covariables sont réduites à $Z_i = \mathbf{1}_{(i\in B)}$. Nous obtenons la statistique de test du log-rank :

Théorème 3.1. *Pour un taux x préalablement fixé, sous l'hypothèse H_0,*

$$\mathbf{U}(\beta_0,\tau,x)^t \Sigma(\beta_0,\tau,x)^{-1}\mathbf{U}(\beta_0,\tau,x) \rightsquigarrow \chi_1^2.$$

Élément de preuve du Théorème 3.1. C'est une conséquence directe du Théorème 2.1. □

Remarque 1. Notons $n_G(t,x) = \sum_{i\in G}\mathbf{1}_{(T_i(x)\geqslant t)}$, $(G = A, B)$ le nombre de patients du groupe G n'ayant pas subi de dégradation à l'instant t et $n(t,x) = n_A(t,x) + n_B(t,x)$ le nombre total de patients non dégradés à l'instant t. Notons $m_G(t,x)$ le nombre de dégradations observées à l'instant t dans le groupe G et $m(t,x) = m_A(t,x) + m_B(t,x)$ le nombre total de dégradations ayant lieu à l'instant t. Pour un patient i nous notons τ_i la réalisation de l'événement $T_i(x)$. Alors, sous l'hypothèse H_0, le Théorème 3.1 peut se réécrire

$$\frac{[U_n(x)]^2}{\widehat{\mathrm{Var}}(U_n(x))} = \frac{\left[\sum_{i=1}^n\left\{m_B(\tau_i,x) - m(\tau_i,x)\frac{n_B(\tau_i,x)}{n(\tau_i,x)}\right\}\right]^2}{\sum_{i=1}^n m(\tau_i,x)\frac{n(\tau_i,x)-m(\tau_i,x)}{n(\tau_i,x)-1}\,\frac{n_A(\tau_i,x)n_B(\tau_i,x)}{n^2(\tau_i,x)}} \rightsquigarrow \chi_1^2.$$

Remarque 2. Dans le cas de la comparaison de l'égalité des fonctions de survie de k groupes de patients, le Théorème 3.1 devient $\mathbf{U}(\beta_0,\tau,x)^t \Sigma(\beta_0,\tau,x)^{-1}\mathbf{U}(\beta_0,\tau,x) \rightsquigarrow \chi_{k-1}^2$.

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] N.E. Breslow, Covariance analysis of censored survival data, Biometrics 30 (1974) 89–99.
[3] D.R. Cox, Regression models and life tables (with discussion), J. Roy. Statist. Soc. Ser. B 34 (1972) 187–220.
[4] D.R. Cox, Partial likelihood, Biometrika 62 (2) (1975) 269–276.
[5] T.R. Fleming, D.P. Harrington, Counting Processes and Survival Analysis, in: Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons Inc., New York, 1991.
[6] E. Lenglart, Relation de domination entre deux processus, Ann. Inst. H. Poincaré Ser. B 13 (1977) 171–179.
[7] A. Tsiatis, A large sample study of Cox's regression model, Ann. Statist. 9 (1981) 93–108.
[8] A. Tsiatis, The asymptotic distribution of the efficient scores test for the proportional hazards model calculated over time, Biometrika 68 (1981) 311–315.