Available online at www.sciencedirect.com
SCIENCE DIRECT

C. R. Acad. Sci. Paris, Ser. I 340 (2005) 377–382

COMPTES RENDUS MATHEMATIQUE

http://france.elsevier.com/direct/CRASS1/

Statistics/Probability Theory

Statistical analysis of survival and reliability data with multiple crossings of survival functions

Vilijandas Bagdonavičius [a], Mikhail Nikulin [b,1]

[a] University of Vilnius, 24, Naugarduko, Vilnius, Lithuania
[b] Statistique mathématique, université Victor Segalen – Bordeaux 2, 146, rue Léo-Saignat, 33076 Bordeaux, France

Received 24 August 2004; accepted after revision 18 January 2005

Presented by Paul Deheuvels

Abstract

We present a statistical model and semiparametric estimation procedure for analysis of survival data with multiple cross-effects (MCE) of survival functions. A goodness-of-fit test for the proportional hazards model against the MCE model is proposed. ***To cite this article: V. Bagdonavičius, M. Nikulin, C. R. Acad. Sci. Paris, Ser. I 340 (2005).***

Résumé

Analyse statistique des données de survie et de fiabilité avec multiples effets de croisement des fonctions de survie. On propose un modèle et une procédure semiparamétrique d'estimation pour analyser les données de survie avec multiples effets de croisement (MCE) de fonctions de survie. Un test d'ajustement pour le modèle des risques proportionnels contre le modèle MCE est proposé. ***Pour citer cet article : V. Bagdonavičius, M. Nikulin, C. R. Acad. Sci. Paris, Ser. I 340 (2005).***

Version française abrégée

Soient $S_x(t)$ et $\lambda_x(t)$ la fonction de survie et la fonction de hasard sous la covariable $x = (x_1, \dots, x_m)^{\mathrm{T}}$. Notons par $\Lambda_x(t) = -\log\{S_x(t)\}$ la fonction de hasard cumulé sous x.

Le modèle MCE est donné par (1), où $\lambda(t)$ est la fonction de hasard de base, $\delta = (\delta_1, \dots, \delta_m)^{\mathrm{T}}$, $\gamma = (\gamma_1, \dots, \gamma_m)^{\mathrm{T}}$, $\beta = (\beta_1, \dots, \beta_m)^{\mathrm{T}}$.

Sous le modèle MCE les fonctions de survie S_x et S_y se croisent une ou deux fois ou ne se croisent pas.

E-mail address: nikou@sm.u-bordeaux2.fr (M. Nikulin).

[1] This work is done with the support of the Commission Permanente du Conseil Régional d'Aquitane: Dossier 20000204009.

1631-073X/$ – see front matter © 2005 Académie des sciences. Published by Elsevier SAS. All rights reserved.
doi:10.1016/j.crma.2005.01.019

On considère des données censurées à droite : (X_i, δ_i, x_i), où X_i et δ_i sont donnés par (3), $i = 1, \ldots, n$.

En utilisant les notations (3) et (4), les fonctions score modifiées pour le paramètre $\theta = (\beta^{\mathrm{T}}, \gamma^{\mathrm{T}}, \delta)^{\mathrm{T}}$ sont données par (5)–(7).

Notons par $T_1^* < \cdots < T_r^*$ les moments de décès distincts. Les valeurs de la fonction $\tilde{\Lambda}(t, \theta)$ (voir (8)) peuvent être trouvées recursivement (voir (9)).

Les estimateurs des fonctions de survie et de hasard cumulé sont donnés par (10).

On a construit un test d'ajustement pour le modèle des hasards proportionnels (PH) contre le modèle MCE. Le test est basé sur la statistique $\widehat{U} = (\widehat{U}_\gamma^{\mathrm{T}}, \widehat{U}_\delta^{\mathrm{T}})^{\mathrm{T}}$ (voir (11), (12). On a démontré que la statistique du test Y^2, donnée par (13), converge en loi vers la loi du chi deux à $2m$ degrés de liberté.

Les conditions pour que ce résultat soit vrai sont les conditions standard pour la normalité asymptotique de l'estimateur $\hat{\beta}$ sous le modèle PH. Le résultat est obtenu de la façon suivante : on développe les fonctions (11) normalisées autour de la vrai valeur β_0 de β, on utilise l'expression de $n^{1/2}(\hat{\beta} - \beta_0)$ en termes des integrales par rapport au martingales $M_i = N_i - \int Y_i \mathrm{e}^{\beta_0^{\mathrm{T}} x_i}\, \mathrm{d}\Lambda$ (sous le modèle PH), on trouve les limites des variations et covariations prévisibles de $\widehat{U}_\gamma$ et $\widehat{U}_\delta$, on vérifie la condition de Lindeberg du théorème limite centrale pour des martingales (voir Fleming and Harrington [6], Ch. 5, Theorems 5.3.4., 5.3.5.) et on trouve la loi limite de $n^{-1/2}\widehat{U}$. Un estimateur consistant de la matrice de covariance limite est utilisé pour construire la statistique du test.

1. Introduction

Piantadosi [8] (Chapter 19, pages 483–488) gives the data concerning the survival times of lung cancer patients. There were 164 patients divided in two groups who received radiotherapy (sample size of 86) or radiotherapy plus 'CAP' (sample size of 78). The Kaplan–Meier estimators tend to cross twice: at a time around 7 months and around 33 months.

We give a model and estimation procedures for the data with two or one crossing. Generalization to the case of more than two crossings is evident. These types of models arise very often in reliability and survival analysis, see, for example, Klein and Moeschenberger [7], Wu [10,9]. Analysis of the data with single crossing is given in Hsieh [5], Bagdonavičius, Hafdi and Nikulin [3], Wu [10].

2. Modeling

Let $S_x(t)$ and $\lambda_x(t)$ be the survival and the hazard rate functions under a m-dimensional possibly time dependent explanatory variable $x = (x_1, \ldots, x_m)^{\mathrm{T}}$. Denote by $\Lambda_x(t) = -\log\{S_x(t)\}$ the cumulative hazard under x.

Let us consider the following model:

Multiple cross-effects (*MCE*) *model*: for all $x \in E \subset R^m$ from a set of explanatory variables E the hazard rate $\lambda_x(t)$ has the form

$$\lambda_x(t) = \mathrm{e}^{\beta^{\mathrm{T}} x(t)}\bigl(1 + \gamma^{\mathrm{T}} x(t)\Lambda(t) + \delta^{\mathrm{T}} x(t)\Lambda^2(t)\bigr)\lambda(t), \qquad \Lambda(t) = \int_0^t \lambda(u)\,\mathrm{d}u, \tag{1}$$

where $\lambda(t)$ and $\Lambda(t)$ are the baseline hazard rate and the baseline cumulative hazard respectively. Here $\delta = (\delta_1, \ldots, \delta_m)^{\mathrm{T}}, \gamma = (\gamma_1, \ldots, \gamma_m)^{\mathrm{T}}, \beta = (\beta_1, \ldots, \beta_m)^{\mathrm{T}}$.

The cumulative hazard under the constant covariate x is

$$\Lambda_x(t) = \mathrm{e}^{\beta^{\mathrm{T}} x}\Lambda(t)\left(1 + \frac{1}{2}\gamma^{\mathrm{T}} x\Lambda(t) + \frac{1}{3}\delta^{\mathrm{T}} x\Lambda^2(t)\right). \tag{2}$$

It is supposed that Λ is continuous and increasing on $(0,\infty)$, $\Lambda(0)=0$, $\Lambda(\infty)=\infty$. The function $\Lambda_x(t)$ also has these properties when $\delta^{\mathrm T}x>0$ or when $\delta^{\mathrm T}x=0$, $\gamma^{\mathrm T}x\geqslant 0$. These conditions are supposed to be satisfied in what follows.

In the case $\gamma=\delta=0$ the MCE model coincides with the proportional hazards (PH) model.

Proposition 2.1. *If the MCE model is true then for any constant covariates $x, y\in E$ the survival functions S_x and S_y do not cross, cross once or twice on $(0,\infty)$ in dependence on the values of the parameter $\theta=(\beta,\gamma,\delta)$.*

Sketch of the proof. Fix x, y. The equality (2) implies that

$$\frac{\Lambda_y(t)}{\Lambda_x(t)}=g(z):=a\,\frac{1+b_2z+c_2z^2}{1+b_1z+c_1z^2},$$

where

$$z=\Lambda(t),\quad a=\mathrm e^{\beta^{\mathrm T}(y-x)},\quad b_1=\frac12\gamma^{\mathrm T}x,\quad b_2=\frac12\gamma^{\mathrm T}y,\quad c_1=\frac13\delta^{\mathrm T}x>0,\quad c_2=\frac13\delta^{\mathrm T}y>0.$$

Suppose that $b_1c_2-b_2c_1>0$. Otherwise, we could investigate the ratio Λ_x/Λ_y. The number of roots of the derivative g' is lesser or equal to 2. In the case of two roots the smaller root is the point of maximum and the bigger root is the point of minimum of the function g. If one of the roots is non-positive then the function $y=g(x)$ cannot intersect the straight line $y=1$ more than twice on $(0,\infty)$. It is so even when both roots are positive because then $c=c_2/c_1<1$, $g(0)=a>g(\infty)=ac$.

None, one or two intersections (which is equivalent to crossings of survival functions) are possible. For example, if $b_1=1.5, b_2=1, c_1=1, c_2=2$, then we have two, one, none intersections taking $a=1.5, 0.9, 3$, respectively. If $b_1=-1, b_2=-1.5, c_1=2, c_2=1.5$ then take $a=1.7, 1.2, 0.9$, respectively. □

3. Semiparametric estimation

Suppose that n objects are observed. The ith of them is observed under the explanatory variable x_i. Denote by T_i and C_i the failure and censoring times for the ith object and set

$$X_i=\min(T_i,C_i),\quad \delta_i=\mathbf 1_{\{T_i\leqslant C_i\}},\quad N_i(t)=\mathbf 1_{\{T_i\leqslant t,\delta_i=1\}},\quad Y_i(t)=\mathbf 1_{\{X_i\geqslant t\}},\tag{3}$$

where $\mathbf 1_A$ denotes the indicator of the event A. Set

$$N(t)=\sum_{i=1}^n N_i(t),\quad Y(t)=\sum_{i=1}^n Y_i(t),\quad \tau=\sup\bigl\{t\colon Y(t)>0\bigr\}.\tag{4}$$

The modified score functions are

$$\widetilde U_\beta(\theta)=U_\beta(\theta,\tilde\Lambda)=\sum_{i=1}^n\int_0^\tau\bigl\{x_i-E_\beta(u,\tilde\Lambda,\theta)\bigr\}\,\mathrm dN_i(u),\tag{5}$$

$$\widetilde U_\gamma(\theta)=U_\gamma(\theta,\tilde\Lambda)=\sum_{i=1}^n\int_0^\tau\left(\frac{x_i\tilde\Lambda(u,\theta)}{1+\gamma^{\mathrm T}x_i\tilde\Lambda(u,\theta)+\delta^{\mathrm T}x_i\tilde\Lambda^2(u,\theta)}-E_\gamma(u,\tilde\Lambda,\theta)\right)\mathrm dN_i(u),\tag{6}$$

$$\widetilde U_\delta(\theta)=U_\delta(\theta,\tilde\Lambda)=\sum_{i=1}^n\int_0^\tau\left(\frac{x_i\tilde\Lambda^2(u,\theta)}{1+\gamma^{\mathrm T}x_i\tilde\Lambda(u,\theta)+\delta^{\mathrm T}x_i\tilde\Lambda^2(u,\theta)}-E_\delta(u,\tilde\Lambda,\theta)\right)\mathrm dN_i(u),\tag{7}$$

where

$$
\begin{aligned}
&\tilde{\Lambda}(t,\theta)=\int_0^t \frac{\mathrm{d}N(u)}{S^{(0)}(u-,\tilde{\Lambda},\theta)},\\
&S^{(0)}(u,\Lambda,\theta)=\sum_{j=1}^{n} Y_j(u)\,\mathrm{e}^{\beta^{\mathrm{T}}x_j}\big[1+\gamma^{\mathrm{T}}x_j\Lambda(u)+\delta^{\mathrm{T}}x_j\Lambda^2(u)\big],\\
&E_\beta(u,\Lambda,\theta)=\frac{S^{(1)}_\beta(u,\Lambda,\theta)}{S^{(0)}(u,\Lambda,\theta)},\quad E_\gamma(u,\Lambda,\theta)=\frac{S^{(1)}_\gamma(u,\Lambda,\theta)}{S^{(0)}(u,\Lambda,\theta)},\quad E_\delta(u,\Lambda,\theta)=\frac{S^{(1)}_\delta(u,\Lambda,\theta)}{S^{(0)}(u,\Lambda,\theta)},\\
&S^{(1)}_\beta(u,\Lambda,\theta)=\sum_{j=1}^{n} x_jY_j(u)\,\mathrm{e}^{\beta^{\mathrm{T}}x_j}\big(1+\gamma^{\mathrm{T}}x_j\Lambda(u)+\delta^{\mathrm{T}}x_j\Lambda^2(u)\big),\\
&S^{(1)}_\gamma(u,\Lambda,\theta)=\sum_{j=1}^{n} x_jY_j(u)\,\mathrm{e}^{\beta^{\mathrm{T}}x_j}\Lambda(u),\qquad S^{(1)}_\delta(u,\Lambda,\theta)=\sum_{j=1}^{n} x_jY_j(u)\,\mathrm{e}^{\beta^{\mathrm{T}}x_j}\Lambda^2(u).
\end{aligned}
\tag{8}
$$

The modified maximum likelihood estimator $\hat{\theta}=(\hat{\beta},\hat{\gamma},\hat{\delta})$ is the solution of the system of equations $\widetilde{U}_\beta(\theta)=0$, $\widetilde{U}_\gamma(\theta)=0$, $\widetilde{U}_\delta(\theta)=0$.

For fixed θ the 'estimator' $\tilde{\Lambda}$ can be found recurrently. Really, let $T_1^*<\cdots<T_r^*$ be observed and ordered distinct failure times, $r\leqslant n$. Note by d_i the number of failures at the moment T_i. Then $\tilde{\Lambda}(0;\theta)=0$,

$$
\tilde{\Lambda}(T_1^*;\theta)=\frac{\mathrm{d}_1}{S^{(0)}(0,\tilde{\Lambda},\theta)},\quad\text{and}\quad \tilde{\Lambda}(T_{j+1}^*;\theta)=\tilde{\Lambda}(T_j^*;\theta)+\frac{d_{j+1}}{S^{(0)}(T_j^*,\tilde{\Lambda}_0,\theta)}
\tag{9}
$$

for $j=1,\dots,r-1$. Given the consistency of $\tilde{\Lambda}$, the asymptotic covariance matrix of $\sqrt{n}(\hat{\theta}-\theta)$ is obtained by standard methods using the functional delta method and the central limit theorem for martingales, see Andersen et al. [1], Bagdonavičius and Nikulin [2]. For consistency proofs of estimators, given by the equations of the type (8), see Ceci and Mazliak [4].

The baseline cumulative hazard Λ and the survival function S_x under any $x\in E$ of the explanatory variable is estimated by $\hat{\Lambda}(t)=\tilde{\Lambda}(t,\hat{\theta})$ and $\hat{S}_x(t)=\mathrm{e}^{-\hat{\Lambda}_x(t)}$, where

$$
\hat{\Lambda}_x(t)=\mathrm{e}^{\hat{\beta}^{\mathrm{T}}x}\hat{\Lambda}(t)\left(1+\frac{1}{2}\mathrm{e}^{\hat{\gamma}^{\mathrm{T}}x}\hat{\Lambda}(t)+\frac{1}{3}\mathrm{e}^{\hat{\delta}^{\mathrm{T}}x}\hat{\Lambda}^2(t)\right).
\tag{10}
$$

4. Goodness-of-fit for the PH model against the MCE model

Let us construct a test for testing the hypothesis H_0: $\lambda_x(t)=\mathrm{e}^{\beta^{\mathrm{T}}x}\lambda(t)$ of the adequacy of the PH model versus the alternative H_1 given by (1) with $(\gamma,\delta)\neq(0,0)$.

Let

$$
\widehat{U}_\rho=U_\rho\big((\hat{\beta}^{\mathrm{T}},0,0)^{\mathrm{T}},\hat{\Lambda}\big),\quad(\rho=\beta,\gamma,\delta),
$$

where $\hat{\beta}$ is the partial likelihood estimator of the regression parameter β and $\hat{\Lambda}$ is the Breslow estimator of Λ under PH model.

Note that under the PH model $\widehat{U}_\beta=0$, and

$$
\widehat{U}_\gamma=\sum_{i=1}^{n}\int_0^\tau \hat{\Lambda}(u-)\big(x_i-E(u,\hat{\beta})\big)\,\mathrm{d}N_i(u),\qquad \widehat{U}_\delta=\sum_{i=1}^{n}\int_0^\tau \hat{\Lambda}^2(u-)\big(x_i-E(u,\hat{\beta})\big)\,\mathrm{d}N_i(u),
\tag{11}
$$

where

$$E(u,\beta)=\frac{S^{(1)}(u,\beta)}{S^{(0)}(u,\beta)},\quad S^{(0)}(u,\beta)=\sum_{j=1}^{n}Y_j(u)\,\mathrm{e}^{\beta^{\mathrm{T}}x_j},\quad S^{(1)}(u,\beta)=\sum_{j=1}^{n}x_jY_j(u)\,\mathrm{e}^{\beta^{\mathrm{T}}x_j}.$$

The test for H_0 is based on the statistic

$$\widehat{U}=\bigl(\widehat{U}_\gamma^{\mathrm{T}},\widehat{U}_\delta^{\mathrm{T}}\bigr)^{\mathrm{T}}. \tag{12}$$

Since the stochastic process $n^{-1/2}\widehat{U}$ converges in distribution to a zero mean Gaussian process, we can use the asymptotic distribution of $\widehat{U}$ under the PH model to construct a test.

Theorem 4.1. *Under standard conditions of regularity and the PH model*

$$Y^2=n^{-1}\widehat{U}^{\mathrm{T}}\widehat{D}^{-1}\widehat{U}\xrightarrow{\mathcal{D}}\chi^2(2m), \tag{13}$$

where $\widehat{D}$ is a consistent estimator of the limit covariance matrix of the random vector $n^{-1/2}\widehat{U}$:

$$\widehat{D}=\begin{pmatrix}\widehat{D}_\gamma & \widehat{D}_{\gamma\delta}\\ \widehat{D}_{\gamma\delta} & \widehat{D}_\delta\end{pmatrix},$$

$$\widehat{D}_\gamma=\widehat{\Sigma}_\gamma^{**}(\tau)-\widehat{\Sigma}_\gamma^{*}(\tau)\widehat{\Sigma}^{-1}(\tau)\bigl(\widehat{\Sigma}_\gamma^{*}(\tau)\bigr)^{\mathrm{T}},\qquad \widehat{D}_\delta=\widehat{\Sigma}_\delta^{**}(\tau)-\widehat{\Sigma}_\delta^{*}(\tau)\widehat{\Sigma}^{-1}(\tau)\bigl(\widehat{\Sigma}_\delta^{*}(\tau)\bigr)^{\mathrm{T}},$$

$$\widehat{D}_{\gamma\delta}=\widehat{\Sigma}_{\gamma\delta}^{**}(\tau)-\widehat{\Sigma}_\gamma^{*}(\tau)\widehat{\Sigma}^{-1}(\tau)\bigl(\widehat{\Sigma}_\delta^{*}(t)\bigr)^{\mathrm{T}},$$

$$\widehat{\Sigma}(t)=n^{-1}\int_0^t V(u,\hat\beta)\,\mathrm{d}N(u),\qquad \widehat{\Sigma}_\gamma^{*}(t)=n^{-1}\int_0^t V(u,\hat\beta)\hat\Lambda(u)\,\mathrm{d}N(u),$$

$$\widehat{\Sigma}_\gamma^{**}(t)=n^{-1}\int_0^t V(u,\hat\beta)\hat\Lambda^2(u)\,\mathrm{d}N(u),\qquad \widehat{\Sigma}_\delta^{*}(t)=\widehat{\Sigma}_\gamma^{**}(t),$$

$$\widehat{\Sigma}_{\gamma\delta}^{**}(t)=n^{-1}\int_0^t V(u,\hat\beta)\hat\Lambda^3(u)\,\mathrm{d}N(u),\qquad \widehat{\Sigma}_\delta^{**}(t)=n^{-1}\int_0^t V(u,\hat\beta)\hat\Lambda^4(u)\,\mathrm{d}N(u),$$

$$V(u,\hat\beta)=\frac{S^{(2)}(u,\hat\beta)}{S^{(0)}(u,\hat\beta)}-E(u,\hat\beta)\bigl(E(u,\hat\beta)\bigr)^{\mathrm{T}},\qquad S^{(2)}(t,\hat\beta)=\sum_{j=1}^{n}x_j(t)(x_j(t))^{\mathrm{T}}Y_j(t)\,\mathrm{e}^{\hat\beta^{\mathrm{T}}x_j}.$$

The conditions needed for the result of the theorem are the standard conditions for the asymptotic normality of the estimator $\hat\beta$ under the PH model (see Andersen et al. [1], pp. 496–497). The result is obtained developing the normed functions (11) around the true value β_0 of β, using the expression of $n^{1/2}(\hat\beta-\beta_0)$ in terms of integrals with respect to the counting process martingales $M_i=N_i-\int Y_i\mathrm{e}^{\beta_0^{\mathrm{T}}x_i}\,\mathrm{d}\Lambda$ under the PH model, calculating the limits of predictable variations and covariations of $\widehat{U}_\gamma$ and $\widehat{U}_\delta$, and verifying the Lindeberg condition of the central limit theorem for martingales (see Fleming and Harrington [6], Ch. 5, Theorems 5.3.4., 5.3.5.), and so finding the limit distribution of $n^{-1/2}\widehat{U}$. The consistent estimator of the limit covariation matrix is used to construct the test statistic T.

The critical region of the chi-square type test with approximate signification level α is

$$Y^2>\chi^2_{1-\alpha}(2m),$$

where $\chi^2_{1-\alpha}(2m)$ is the $(1-\alpha)$-quantile of the chi-square distribution with $2m$ degrees of freedom.

It would be interesting to obtain a test for the hypothesis of one crossing against two crossing. Note that if $\delta = 0$ then one or none crossings are possible and when $\delta \neq 0$ then two, one or zero crossings can take place. So we cannot obtain nested models here and the idea used testing the PH model does not work in such situation.

References

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

[2] V. Bagdonavičius, M. Nikulin, Accelerated Life Models: Modeling and Statistical Analysis, Chapman and Hall/CRC, Boca Raton, 2002.

[3] V. Bagdonavičius, M. Hafdi, M. Nikulin, Analysis of survival data with cross-effects of survival functions, Biostatistics 5 (3) (2004) 415–425.

[4] C. Ceci, L. Mazliak, Optimal design in nonparametric life testing, Preprint January 2002, Laboratoire de probabilités et modèles aléatoirs, Universités Paris VI et VII, 2002.

[5] F. Hsieh, On heteroscedastic hazards regression models: theory and application, J. Roy. Statist. Soc. Ser. B 63 (2001) 63–79.

[6] T.R. Fleming, D.P. Harrington, Counting Processes and Survival Analysis, Wiley, New York, 1991.

[7] J.P. Klein, M.L. Moeschenberger, Survival Analysis. Statistics for Biology and Health, Springer, New York, 1997.

[8] S. Piantadosi, Clinical Trials: A Methodologic Perspective, Wiley, New York, 1997.

[9] H.-D.I. Wu, Effect of ignoring heterogeneity in hazards regression, in: Parametric and Semiparametric Models with Applications to Reliability, Survival Analysis, and Quality of Life, Birkhäuser, Boston, 2004, pp. 239–252.

[10] H.-D.I. Wu, A partial score test for difference among heteroscedastic populations, Preprint of The School of Public Health, China Medical College, Taichung, Taiwan, 21 October, 2002.