

A MATRIX-VALUED COUNTING PROCESS WITH FIRST-ORDER INTERACTIVE INTENSITIES

BY A. C. PEDROSO DE LIMA¹ AND PRANAB K. SEN

Universidade de São Paulo and University of North Carolina, Chapel Hill

A matrix-valued counting process that allows for modeling of multivariate failure-time data is presented. The inclusion of covariates in a Cox-type regression model is considered, and asymptotic properties of the estimates of regression parameters appearing in the model are studied.

1. Introduction. The modern theory of counting process and martingales as treated by Brémaud (1981) and others provides the necessary theoretical background for the development of rigorous and general theory of regression models adapted to censored data. The seed of such an approach seems to reside in the work of Aalen (1975). The more recent books by Fleming and Harrington (1991) and Andersen, Borgan, Gill and Keiding (1993) take into account a very broad spectrum of applications for the methodology. In the multivariate setting, several approaches have been proposed to deal with the estimation of the multivariate time scale survival function such as, for example, the Dabrowska estimator [Dabrowska (1988)] involving conditional and marginal hazards and the Prentice and Cai estimator [Prentice and Cai (1992)] considering a Volterra-type equation to relate the marginal failure times to their covariance function. In both works, the main issue is not related to the effect that covariates may have on the survival experience, but rather to the estimation of the survival function. In this work we propose a model to handle multivariate failure-time data when our main interest is to consider situations where one is concerned with the effect of covariates on the multiple endpoint regression, and, hence, our goal is to draw statistical conclusions for such regression parameters in the model. In this context, Cai and Prentice (1995) consider a marginal model approach with adjustment for covariates of the marginal estimates. Our proposed model makes use of interactions in the joint distribution instead of adjustments by the covariance estimation suggested by them.

2. The matrix-valued counting process model. In order to develop the model and asymptotic properties, we will consider a bivariate model. The extension of the results to the k -variate situation is discussed later. Let (T_1, T_2) be nonnegative random vector defined in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. In principle we assume that the elements of such a vector are not necessarily inde-

Received August 1994; revised November 1996.

¹Research supported in part by CAPES-Brazil Grant 1175/90-2.

AMS 1991 subject classifications. Primary 62M99; secondary 60G44, 60G55, 90B25.

Key words and phrases. Multivariate failure-time, martingale, predictable process, Cox-type regression model, concomitant variables, censoring, maximum partial likelihood estimators.

pendent and have a joint survival function $\mathbf{S}_{12}(t_1, t_2)$. The marginal survival functions are represented by $\mathbf{S}_1(t_1)$ and $\mathbf{S}_2(t_2)$. We also assume the presence of censoring, represented by a nonnegative random variable C , independent of (T_1, T_2) , defined in the same probability space. Typically, in real life one does not observe necessarily T_h or C ; instead, the observed quantities are $Z_h = T_h \wedge C_h$ and $\delta_h = \mathbb{I}\{Z_h = T_h\}$, where $\mathbb{I}\{A\}$ represents the indicator or characteristic function. We define the counting processes

$$(2.1) \quad N_h(t) = \mathbb{I}\{Z_h \leq t; \delta_h = 1\}, \quad t \geq 0, \quad h = 1, 2,$$

representing a right-continuous function that assumes value zero, jumping to one when the particular event associated with T_h occurs. Since the quantities in (2.1) are defined on dependent random variables, it makes sense to consider the process

$$(2.2) \quad \mathbf{N} = \{\mathbf{N}(t) = (N_1(t), N_2(t))', \quad t \geq 0\}.$$

We assume that the quantities above are measurable with respect to $\mathcal{N}_t^\# = \sigma\{\mathbf{N}^\#(s), 0 \leq s \leq t\}$, the self-exciting filtration defined by the vector-valued counting process $\mathbf{N}^\#(t)$ with elements $N_h^\#(t) = \mathbb{I}\{Z_h \leq t\}$. Note that $\mathcal{N}_t^\#$ contains information on the processes $N_h(t)$ as well as on their dependence.

In order to characterize the counting processes above, we define the $\mathcal{N}_t^\#$ -predictable processes

$$Y_h(t) = \mathbb{I}\{Z_h \geq t\}, \quad t \geq 0, \quad h = 1, 2.$$

If we pretend for a moment that the components of \mathbf{N} are independent, then the multiplicative intensity model of Aalen (1978) would apply; that is, the associated intensity process of N_h would be given by

$$(2.3) \quad \lambda_h(t) = \alpha_h(t)Y_h(t), \quad h = 1, 2,$$

where $\alpha_h(t)$ is the marginal hazard function, defined by

$$\alpha_h(t) = \lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}\{T_h \in (t, t + \Delta t] \mid T_h > t\}}{\Delta t}.$$

Under the assumption of independence, collecting the intensity processes defined in (2.3) in a vector $\boldsymbol{\lambda}$, we can write

$$(2.4) \quad \boldsymbol{\lambda}(t) = \begin{pmatrix} \lambda_1(t) \\ \lambda_2(t) \end{pmatrix} = \begin{pmatrix} \alpha_1(t) & 0 \\ 0 & \alpha_2(t) \end{pmatrix} \begin{pmatrix} Y_1(t) \\ Y_2(t) \end{pmatrix} = \boldsymbol{\alpha}(t)\mathbf{Y}(t),$$

and this fully specifies the counting process defined in (2.2) when independence holds. It is our goal now to modify (2.4) in order to get a model for a more general situation where the independence may not be tenable. In such a case it is expected that the interpretation for the unknown deterministic functions $\alpha_h(t)$ should change and, also, the off-diagonal elements may be different from zero. Let us approach this situation considering a generalization

of the heuristic approach given for (2.3) [see, e.g., Andersen, Borgan, Gill and Keiding (1993) for the univariate case.] In this case one may write

$$(2.5) \quad \lambda_h(t) = E \{ dN_h(t) \mid \mathcal{N}_t^\# \},$$

and note that $\mathcal{N}_t^\#$ contains information whether or not one (or both) component(s) have failed just before t . If the component h has failed before t , then expression (2.5) equals zero. In other words, we need to consider the following situations.

1. No component has failed at instant t ; that is, $Y_1(t) = Y_2(t) = 1$.
2. The first component has failed before t but the second has not; that is, $Y_1(t) = 0$ and $Y_2(t) = 1$.
3. Only the second component has failed before t ; that is, $Y_1(t) = 1$ and $Y_2(t) = 0$.
4. Both components failed before t , in which case $Y_1(t) = Y_2(t) = 0$.

If we want to consider the intensity process for the first component, then we only consider cases where $Y_1(t) = 1$. This, together with expression (2.5), allows us to write

$$\lambda_1(t) = E \{ dN_1(t) \mid \mathcal{N}_t^\# \} = p_1^{(1)}(t)Y_1(t)[1 - Y_2(t)] + p_2^{(1)}(t)Y_1(t)Y_2(t),$$

where $p_1^{(1)}(t) = \lim_{\Delta t \rightarrow 0} (\Delta t)^{-1} \mathbb{P}\{T_1 \in (t, t + \Delta t] \mid T_1 > t; T_2 \leq t\}$ and $p_2^{(1)}(t) = \lim_{\Delta t \rightarrow 0} (\Delta t)^{-1} \mathbb{P}\{T_1 \in (t, t + \Delta t] \mid T_1 > t; T_2 > t\}$ are conditional hazard functions. Note that these functions take into consideration the status of the component 2. Using a similar approach, we can write the intensity $\lambda_2(t)$ for the other component.

Based on those quantities we can represent the intensity process by the product of matrices

$$(2.6) \quad \boldsymbol{\lambda}(t) = \begin{pmatrix} \lambda_1(t) \\ \lambda_2(t) \end{pmatrix} = \begin{pmatrix} Y_1(t) & 0 \\ 0 & Y_2(t) \end{pmatrix} \begin{pmatrix} \alpha_{11}(t) & \alpha_{12}(t) \\ \alpha_{21}(t) & \alpha_{22}(t) \end{pmatrix} \begin{pmatrix} Y_1(t) \\ Y_2(t) \end{pmatrix} \\ = \text{Diag}(\mathbf{Y}(t))\boldsymbol{\alpha}(t)\mathbf{Y}(t),$$

where the elements of $\boldsymbol{\alpha}(t)$ are given by $\alpha_{11}(t) = p_1^{(1)}(t)$, $\alpha_{12}(t) = p_2^{(1)}(t) - p_1^{(1)}(t)$, $\alpha_{21}(t) = p_1^{(2)}(t) - p_2^{(2)}(t)$ and $\alpha_{22}(t) = p_2^{(2)}(t)$.

The *matrix-valued counting process model* is defined in the following way. Suppose that $\mathbf{N}_1, \dots, \mathbf{N}_n$ are n copies of the process \mathbf{N} defined in (2.2). Then the matrix-valued counting process is given by

$$(2.7) \quad \mathbf{N}(t) = (\mathbf{N}_1(t), \dots, \mathbf{N}_n(t)),$$

with an associated intensity process given by (2.6). The columns of \mathbf{N} are independent but each column, in this case, consists of two possibly dependent elements.

As an illustration, let us consider the following parametric model, due to Sarkar (1987).

EXAMPLE. Consider an absolutely continuous bivariate exponential distribution where the joint survival function for the vector (T_1, T_2) is given by

$$\begin{aligned} & \mathbb{P}\{T_1 \geq t_1; T_2 \geq t_2\} \\ &= \begin{cases} \exp(-(\beta_2 + \beta_{12})t_2)\{1 - A(\beta_1 t_2)\}^{-\gamma} A(\beta_1 t_1)^{1+\gamma}, & 0 < t_1 \leq t_2, \\ \exp(-(\beta_1 + \beta_{12})t_1)\{1 - A(\beta_2 t_1)\}^{-\gamma} A(\beta_2 t_2)^{1+\gamma}, & 0 < t_2 \leq t_1, \end{cases} \end{aligned}$$

where $\beta_1 > 0, \beta_2 > 0, \beta_{12} > 0, \gamma = \beta_{12}/(\beta_1 + \beta_2)$ and $A(z) = 1 - e^{-z}, z > 0$. Note that, if $\beta_{12} = 0$, the joint distribution factorizes in two exponential distributions and then T_1 and T_2 are independent.

When no censoring is present, define the bivariate counting process with elements $N_1(t) = \mathbb{I}\{T_1 \leq t\}, N_2(t) = \mathbb{I}\{T_2 \leq t\}$ and the predictable processes $Y_1(t) = \mathbb{I}\{T_1 \geq t\}$ and $Y_2(t) = \mathbb{I}\{T_2 \geq t\}$. Then, after some long algebraic manipulations, we obtain the elements of $\alpha(t)$ that may be obtained by taking linear combinations of the quantities $p_i^{(h)}(t)$ as shown after expression (2.6) and, hence, the intensity process can be expressed as

$$\begin{aligned} \begin{pmatrix} \lambda_1(t) \\ \lambda_2(t) \end{pmatrix} &= \begin{pmatrix} \beta_1 & 0 \\ 0 & \beta_2 \end{pmatrix} \begin{pmatrix} Y_1(t) \\ Y_2(t) \end{pmatrix} \\ &+ \gamma \begin{pmatrix} \beta_1 + \frac{\beta_2}{A(\beta_2 t)} & -\frac{\beta_2}{A(\beta_2 t)} Y_1(t) \\ -\frac{\beta_1}{A(\beta_1 t)} Y_2(t) & \beta_2 + \frac{\beta_1}{A(\beta_1 t)} \end{pmatrix} \begin{pmatrix} Y_1(t) \\ Y_2(t) \end{pmatrix}. \end{aligned}$$

The dependence structure becomes explicit in this expression in the sense that if $\gamma = 0$ the resulting expression for the intensity process vector will be the same one obtained when working with two independent exponential random variables, with parameters β_1 and β_2 .

The Sarkar model is an important member of the family of bivariate exponential distributions, and there are others, too. Pedroso de Lima and Sen (1997) have examined these models incorporating the dual intensity functions based on the current matrix-valued counting process approach, providing a coherent picture of these bivariate exponential distributions and their placing in multiple endpoint survival analysis.

3. The bivariate model with covariates. Here we consider the bivariate model specified in (2.6) incorporating covariates as in a Cox-type regression model. Throughout this section we will assume that $t \in [0, \tau], \tau > 0$ and, in addition to the quantities defined earlier, we also have a set of time-dependent covariates $X_1(t), \dots, X_q(t)$. The covariates are assumed to be observed for all individuals. For simplicity of presentation we consider only the case $q = 1$.

For n independent individuals, let $\mathbf{N}_1, \dots, \mathbf{N}_n$ be copies of \mathbf{N} defined in (2.2) with corresponding intensity processes given by $\lambda_1, \dots, \lambda_n$, where

$$\lambda_i(t) = \begin{pmatrix} \lambda_{1i}(t) \\ \lambda_{2i}(t) \end{pmatrix} = \begin{pmatrix} \alpha_{11}^{(i)}(t)Y_{1i}(t) + \alpha_{12}^{(i)}(t)Y_{1i}(t)Y_{2i}(t) \\ \alpha_{22}^{(i)}(t)Y_{2i}(t) + \alpha_{21}^{(i)}(t)Y_{1i}(t)Y_{2i}(t) \end{pmatrix}.$$

Since for each individual i , one observes the covariate $X_i(t)$, $i = 1, \dots, n$, we assume that each element of α can be expressed through a multiplicative form such that

$$\begin{aligned}
 p_{1i}^{(1)}(t) &= \gamma_{11}(t) \exp(\beta_1 X_i(t)), & \Rightarrow & \left\{ \begin{aligned} \alpha_{11}^{(i)}(t) &= \gamma_{11}(t) \exp(\beta_1 X_i(t)) \\ &= \alpha_{11}^o(t) \exp(\beta_1 X_i(t)), \end{aligned} \right. \\
 p_{2i}^{(1)}(t) &= \gamma_{12}(t) \exp(\beta_1 X_i(t)), & & \left\{ \begin{aligned} \alpha_{12}^{(i)}(t) &= (\gamma_{12}(t) - \gamma_{11}(t)) \exp(\beta_1 X_i(t)) \\ &= \alpha_{12}^o(t) \exp(\beta_1 X_i(t)), \end{aligned} \right. \\
 p_{1i}^{(2)}(t) &= \gamma_{21}(t) \exp(\beta_2 X_i(t)), & & \left\{ \begin{aligned} \alpha_{21}^{(i)}(t) &= (\gamma_{21}(t) - \gamma_{22}(t)) \exp(\beta_2 X_i(t)) \\ &= \alpha_{21}^o(t) \exp(\beta_2 X_i(t)), \end{aligned} \right. \\
 p_{2i}^{(2)}(t) &= \gamma_{22}(t) \exp(\beta_2 X_i(t)) & & \left\{ \begin{aligned} \alpha_{22}^{(i)}(t) &= \gamma_{22}(t) \exp(\beta_2 X_i(t)) \\ &= \alpha_{22}^o(t) \exp(\beta_2 X_i(t)). \end{aligned} \right.
 \end{aligned}$$

In addition, we simplify further the model with the assumption that $\alpha_{11}^o(t) = \theta_1^{-1} \alpha_{12}^o(t)$ and $\alpha_{22}^o(t) = \theta_2^{-1} \alpha_{21}^o(t)$, for $\theta_h > -1$, $h = 1, 2$. Then the intensity process vector can be written as

$$(3.8) \quad \lambda_i(t) = \begin{pmatrix} \alpha_{11}^o(t)(Y_{1i}(t) + \theta_1 Y_{1i}(t)Y_{2i}(t)) \exp(\beta_1 X_i(t)) \\ \alpha_{22}^o(t)(Y_{2i}(t) + \theta_2 Y_{1i}(t)Y_{2i}(t)) \exp(\beta_2 X_i(t)) \end{pmatrix}$$

Based on that, the problem at hand consists in finding estimates for β_j and θ_j , $j = 1, 2$. Since both failure times (for the two components) are assumed to be observed at the exact instant they occur, we have that no two components can jump at the same instant t for the same subject and, hence, when estimating β we will consider a likelihood whose contribution of individual i at time t , if any, will be restricted to $\lambda_{1i}(t) / \sum_{j=1}^n \lambda_{1j}(t)$ when N_{1i} jumps or $\lambda_{2i}(t) / \sum_{j=1}^n \lambda_{2j}(t)$ when N_{2i} jumps. The likelihood can then be written as a product of ratios involving those quantities and, considering the proportionality assumption for the off-diagonal terms in $\alpha(t)$, we are able to cancel the unknown baseline functions $\alpha_{ij}^o(t)$. Cai and Prentice's (1995) marginal model approach relates, in our notation, to $\lambda_{hi}(t) = Y_{hi}(t) \alpha_{hh}^o(t) \exp(\beta X_{hi}(t))$, $h = 1, 2$, where both components depend on the same regression parameter β . In our case, in (3.8) the parameters θ_h incorporate dependence among the components without imposing the homogeneity of the regression parameters β_1, β_2 .

Let $\delta = (\delta_1, \delta_2)$, with $\delta_k = (\beta_k, \theta_k)$ and suppose the true parameter value is represented by δ^o . Then the log-likelihood can be expressed as

$$\begin{aligned}
 \log L(\delta) &= \int_0^\tau \left(\sum_{i=0}^n \sum_{h=1}^2 \beta_h X_i(t) + \log\{Y_{hi}(t) + \theta_h Y_{1i}(t)Y_{2i}(t)\} \right. \\
 &\quad \left. - \log \left\{ \sum_{j=1}^n \exp(\beta_h X_j(t))(Y_{hj}(t) + \theta_h Y_{1j}(t)Y_{2j}(t)) \right\} \right) dN_{hi}(t).
 \end{aligned}$$

Computation of the score vector takes place for a pair of parameters for each component. Therefore it is convenient to consider a partitioned vector where the first element is a two-vector containing the derivatives of the log-likelihood with respect to the parameters related to the first component and the same quantities for the second element with information related to the second component. It should be noted that the score vector is also a stochastic process in $[0, \tau]$. For $t = \tau$, we write

$$\mathbf{U}(\tau; \delta) = \begin{pmatrix} \mathbf{U}^{(1)}(\tau; \delta) \\ \mathbf{U}^{(2)}(\tau; \delta) \end{pmatrix},$$

where the first element of $\mathbf{U}^{(h)} = (U_1^{(h)}, U_2^{(h)})'$, $h = 1, 2$, is given by

$$\begin{aligned} (3.9) \quad U_1^{(h)}(\tau; \delta) &= \frac{\partial \log L(\delta)}{\partial \beta_h} \\ &= \int_0^\tau \sum_{i=1}^n \left(X_i(t) - \frac{\sum_{j=1}^n X_j(t) w_j(\delta_h, t)}{\sum_{j=1}^n w_j(\delta_h, t)} \right) dN_{hi}(t), \end{aligned}$$

and where $w_j(\delta_h, t) = \exp(\beta_h X_j(t))(Y_{hj}(t) + \theta_h Y_{1j}(t)Y_{2j}(t))$. Note that (3.9) is similar to the expression one obtains when considering the univariate case. The basic difference relies on the possibly time-dependent *weights* w_j that are taking into account the predictable processes associated with both components. The second element for the score vector is given by

$$\begin{aligned} (3.10) \quad U_2^{(h)}(\tau; \delta) &= \frac{\partial \log L(\delta)}{\partial \theta_h} \\ &= \int_0^\tau \sum_{i=1}^n \left(\frac{Y_{1i}(t)Y_{2i}(t)}{Y_{hi}(t) + \theta_h Y_{1i}(t)Y_{2i}(t)} - \frac{\sum_{j=1}^n \exp(\beta_h X_j(t))Y_{1j}(t)Y_{2j}(t)}{\sum_{j=1}^n w_j(\delta_h, t)} \right) dN_{hi}(t). \end{aligned}$$

Maximum partial likelihood estimators (MPLE) can be obtained by solving the equations $\mathbf{U}(\tau; \delta) = \mathbf{0}$, which need to be computed iteratively. Let us denote the MPLE by $\hat{\delta}$. Asymptotic properties of such estimators are studied using the standard martingale theory. Hence, we first note that expressions (3.9) and (3.10) can be written as martingales when $\delta = \delta^o$. This is done in a manner similar to the univariate case, developed in Andersen and Gill (1982). First we note that $dN_{hi}(t) = N_{hi}(t) - N_{hi}(t^-) = dM_{hi}(t) - \lambda_{hi}(t) dt$, where $M_{hi}(t)$ is a local square integrable martingale. Plugging this quantity into expressions (3.9) and (3.10), we obtain a difference of two integrals, one involving the martingale M_{hi} and the other involving the intensity process. It turns out that the latter equals zero. Therefore, it follows that the first element of the score function can be written as

$$(3.11) \quad U_1^{(h)}(\tau; \delta^o) = \int_0^\tau \sum_{i=1}^n \left(X_i(t) - \frac{\sum_{j=1}^n X_j(t) w_j(\delta_h^o, t)}{\sum_{j=1}^n w_j(\delta_h^o, t)} \right) dM_{hi}(t),$$

and, hence, (3.11) is also a square integrable martingale. A similar result follows for the second element of the score vectors; that is,

$$U_2^{(h)}(\tau; \delta^o) = \int_0^\tau \sum_{i=1}^n \left(\frac{Y_{1i}(t)Y_{2i}(t)}{Y_{hi}(t) + \theta_h^o Y_{1i}(t)Y_{2i}(t)} - \frac{\sum_{j=1}^n \exp(\beta_h^o X_j(t)) Y_{1j}(t) Y_{2j}(t)}{\sum_{j=1}^n w_j(\delta_h^o, t)} \right) dM_{hi}(t),$$

also a square integrable martingale. These facts are used to derive the asymptotic distribution for the score function, incorporating Rebolledo's central limit theorem for martingales. Such a theorem assumes that the *predictable variation processes* satisfy certain conditions. In order to compute such processes, denoted by $\langle \cdot, \cdot \rangle$, the following processes are defined:

$$S_h^{(j)}(\delta, t) = (1/n) \sum_{i=1}^n X_i^j(t) (Y_{hi}(t) + \theta_h Y_{1i}(t) Y_{2i}(t)) \exp(\beta_h X_i(t)), \quad j=0, 1, 2,$$

$$S_h^{(3)}(\delta, t) = (1/n) \sum_{i=1}^n \exp(\beta_h X_i(t)) Y_{1i}(t) Y_{2i}(t),$$

$$S_h^{(4)}(\delta, t) = (1/n) \sum_{i=1}^n \frac{Y_{1i}(t) Y_{2i}(t)}{Y_{hi}(t) + \theta_h Y_{1i}(t) Y_{2i}(t)} \exp(\beta_h X_i(t)),$$

$$S_h^{(5)}(\delta, t) = (1/n) \sum_{i=1}^n X_i(t) Y_{1i}(t) Y_{2i}(t) \exp(\beta_h X_i(t)).$$

Defining also $U_j^{(h,n)} = n^{-1/2} U_j^{(h)}$, $j = 1, 2$, and using well-known properties of the predictable processes involved, the predictable variation process of $U_1^{(h,n)}$ will be given by

$$\langle U_1^{(h,n)}, U_1^{(h,n)} \rangle(t) = \int_0^t \left(S_h^{(2)}(\delta^o, s) - \frac{(S_h^{(1)}(\delta^o, s))^2}{S_h^{(0)}(\delta^o, s)} \right) \alpha_{hh}^o(s) ds.$$

Similarly,

$$\langle U_2^{(h,n)}, U_2^{(h,n)} \rangle(t) = \int_0^t \left(S_h^{(4)}(\delta^o, s) - \frac{(S_h^{(3)}(\delta^o, s))^2}{S_h^{(0)}(\delta^o, s)} \right) \alpha_{hh}^o(s) ds$$

and

$$\langle U_1^{(h,n)}, U_2^{(h,n)} \rangle(t) = \int_0^t \left(S_h^{(5)}(\delta^o, s) - \frac{S_h^{(1)}(\delta^o, s) S_h^{(3)}(\delta^o, s)}{S_h^{(0)}(\delta^o, s)} \right) \alpha_{hh}^o(s) ds.$$

Now, parallel to the univariate case, we enlist a set of conditions.

C.1. $\int_0^\tau \alpha_{hh}(s) ds < \infty$.

C.2. We assume that $\sup_{\{t \in [0, \tau]; \delta \in \mathcal{D}\}} \|S_h^{(j)}(\delta, t) - s_h^{(j)}(\delta, t)\| \rightarrow_P 0$, with $s_h^{(0)}$ bounded away from zero on $\mathcal{D} \times [0, \tau]$ and $s_h^{(j)}$ continuous functions of δ on \mathcal{D} , $j = 1, \dots, 5$, where \mathcal{D} is a neighborhood of the true parameter vector.

C.3. There exists $\gamma > 0$, such that $n^{-1/2} \sup_{\{1 \leq i \leq n; t \in [0, \tau]\}} |X_i(t)| Y_{hi}(t) \times \mathbb{I}\{\beta_h^o X_i(t) > -\gamma |X_i(t)|\} \rightarrow_P 0$.

C.4. There exists a positive definite matrix Σ , with elements involving the limiting functions $s_h^{(j)}(\cdot, \cdot)$, as defined below.

The following theorem relates to the asymptotic distribution for the score function.

THEOREM 1. *For a bivariate counting process with intensity process defined by (3.8) under the regularity conditions C.1–C.3, the stochastic process $n^{-1/2}\mathbf{U}$, with \mathbf{U} defined in (3.9) and (3.10), converges in distribution to a continuous Gaussian martingale \mathbf{W} with the cross-covariance function $E\{\mathbf{W}(s)[\mathbf{W}(t)]'\} = \Sigma(s \wedge t)$ given by*

$$\Sigma(t) = \begin{pmatrix} \Sigma_1(t) & \mathbf{0} \\ \mathbf{0} & \Sigma_2(t) \end{pmatrix},$$

where for each $h(= 1, 2)$, $\Sigma_h(t)$ is a matrix with elements

$$\begin{aligned} (\Sigma_h(t))_{11} &= \int_0^t \left(s_h^{(2)}(\delta^o, s) - \frac{(s_h^{(1)}(\delta^o, s))^2}{s_h^{(0)}(\delta^o, s)} \right) \alpha_{hh}^o(s) ds, \\ (\Sigma_h(t))_{22} &= \int_0^t \left(s_h^{(4)}(\delta^o, s) - \frac{(s_h^{(3)}(\delta^o, s))^2}{s_h^{(0)}(\delta^o, s)} \right) \alpha_{hh}^o(s) ds, \\ (\Sigma_h(t))_{12} &= \int_0^t \left(s_h^{(5)}(\delta^o, s) - \frac{s_h^{(1)}(\delta^o, s) s_h^{(3)}(\delta^o, s)}{s_h^{(0)}(\delta^o, s)} \right) \alpha_{hh}^o(s) ds. \end{aligned}$$

PROOF. The proof exploits a technique by Andersen and Gill (1982), with some modifications, to include our contemplated more general setup. We refer to Pedroso de Lima (1995) for details.

The asymptotic distribution of the maximum partial likelihood is given by the following theorem. Our derivation is based on the classical Le Cam (1956) approach for the usual maximum likelihood estimator, and here also we refer to Pedroso de Lima (1995) for the details.

THEOREM 2. *Let $\widehat{\delta}$ be a value that maximizes the partial likelihood and suppose that conditions C.1–C.4 hold. Then, if δ^o is the true value for the parameter δ ,*

$$\sqrt{n}(\widehat{\delta} - \delta^o) \rightarrow_D \mathcal{N}(\mathbf{0}, \Sigma^{-1}).$$

We include the following corollaries to conclude our discussion.

COROLLARY 1. *The estimator $\widehat{\boldsymbol{\delta}}$ is a consistent estimator for $\boldsymbol{\delta}^o$.*

This corollary is a direct consequence of the theorem; that is, one can immediately verify that $\|\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}^o\| = o_p(1)$

COROLLARY 2. *The covariance matrix $\boldsymbol{\Sigma}$ can be consistently estimated by $n^{-1}\mathcal{J}(\widehat{\boldsymbol{\delta}})$.*

Corollary 2 follows from the assumptions made in Theorem 2 and from Corollary 1. In fact, since $\widehat{\boldsymbol{\delta}}$ is a consistent estimator, there exists a value n_0 such that, for all $n \geq n_0$ it will be in the neighborhood \mathcal{D} , such that condition C.4 will be true, and, hence, each element of the matrix $n^{-1}\mathcal{J}(\widehat{\boldsymbol{\delta}})$ converges to the respective element of $\boldsymbol{\Sigma}$.

4. The model with $K > 2$ components. In the case of more than two components in the model, its intensity process is somewhat more complex due to the multiplicity of higher order combinations of the predictable processes. To illustrate this point, let us consider the case with three components, involving the nonnegative random vector $\mathbf{T} = (T_1, T_2, T_3)'$. As in the bivariate case, define the matrix-valued counting process (2.7) where each column now is given by a 3×1 vector of counting processes $\mathbf{N}_i = (N_{1i}, N_{2i}, N_{3i})'$, based on n copies \mathbf{T}_i of \mathbf{T} . Also, the predictable vector is given by $\mathbf{Y}_i = (Y_{1i}, Y_{2i}, Y_{3i})'$, where $Y_{hi}(t) = \mathbb{I}\{Z_{hi} \geq t\}$.

In order to compute the intensity processes, we need to consider the $2^3 = 8$ possibilities represented by the combinations of 0's and 1's of the elements of the vector $\mathbf{Y}_i(t)$ but only four (for which the corresponding predictable process is not zero at time t) are meaningful for each component. For example, for the first component we consider only the cases where $Y_{1i}(t) = 1$, as otherwise the component will have already failed and the conditional hazard function will be zero. The following notations are then introduced for the conditional hazard functions (dropping out the subscript i to simplify the notation).

When $Y_1(t) = Y_2(t) = Y_3(t) = 1$, the conditional hazard is given by

$$p_{123}^{(1)}(t) = \lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}\{T_1 \in (t, t + \Delta t] \mid T_1 > t, T_2 > t, T_3 > t\}}{\Delta t};$$

when $Y_1(t) = Y_2(t) = 1$ and $Y_3(t) = 0$,

$$p_{12}^{(1)}(t) = \lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}\{T_1 \in (t, t + \Delta t] \mid T_1 > t, T_2 > t, T_3 \leq t\}}{\Delta t};$$

when $Y_1(t) = Y_3(t) = 1$ and $Y_2(t) = 0$,

$$p_{13}^{(1)}(t) = \lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}\{T_1 \in (t, t + \Delta t] \mid T_1 > t, T_2 \leq t, T_3 > t\}}{\Delta t};$$

when $Y_1(t) = 1$ and $Y_2(t) = Y_3(t) = 0$,

$$p_1^{(1)}(t) = \lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}\{T_1 \in (t, t + \Delta t] \mid T_1 > t, T_2 \leq t, T_3 \leq t\}}{\Delta t},$$

so that the intensity process will be given by one of the four expressions above, that, depending on the value of \mathbf{Y} , can be written as

$$(4.12) \quad \begin{aligned} \lambda_1(t) &= \alpha_1^{(1)}(t)Y_1(t) + \alpha_{13}^{(1)}(t)Y_1(t)Y_3(t) + \alpha_{12}^{(1)}(t)Y_1(t)Y_2(t) \\ &\quad + \alpha_{123}^{(1)}(t)Y_1(t)Y_2(t)Y_3(t). \end{aligned}$$

The same scheme applies to the second and third components, with only changes in the indices:

$$(4.13) \quad \begin{aligned} \lambda_2(t) &= \alpha_2^{(2)}(t)Y_2(t) + \alpha_{23}^{(2)}(t)Y_2(t)Y_3(t) + \alpha_{12}^{(2)}(t)Y_1(t)Y_2(t) \\ &\quad + \alpha_{123}^{(2)}(t)Y_1(t)Y_2(t)Y_3(t); \end{aligned}$$

$$(4.14) \quad \begin{aligned} \lambda_3(t) &= \alpha_3^{(3)}(t)Y_3(t) + \alpha_{23}^{(3)}(t)Y_2(t)Y_3(t) + \alpha_{13}^{(3)}(t)Y_1(t)Y_3(t) \\ &\quad + \alpha_{123}^{(3)}(t)Y_1(t)Y_2(t)Y_3(t). \end{aligned}$$

Based on expressions (4.12)–(4.14), we may note that each expression has a term involving the predictable process for the corresponding component, $\binom{3}{2}$ terms involving the product of two predictable processes and one term involving the product of the three processes $Y_h(t)$, $h = 1, 2, 3$. This structure resembles the usual multifactor analysis of variance or categorical data models, where typically one considers models involving the main effects and interactions of various orders. When collecting all three quantities defined above in a vector of intensity processes, to emphasize these interactions, one may rewrite the model as

$$(4.15) \quad \begin{aligned} \boldsymbol{\lambda}(t) = \begin{pmatrix} \lambda_1(t) \\ \lambda_2(t) \\ \lambda_3(t) \end{pmatrix} &= \begin{pmatrix} \alpha_1^{(1)}(t) \\ 0 \\ 0 \end{pmatrix} Y_1(t) + \begin{pmatrix} 0 \\ \alpha_2^{(2)}(t) \\ 0 \end{pmatrix} Y_2(t) \\ &\quad + \begin{pmatrix} 0 \\ 0 \\ \alpha_3^{(3)}(t) \end{pmatrix} Y_3(t) + \begin{pmatrix} \alpha_{12}^{(1)}(t) \\ \alpha_{12}^{(2)}(t) \\ 0 \end{pmatrix} Y_1(t)Y_2(t) \\ &\quad + \begin{pmatrix} \alpha_{13}^{(1)}(t) \\ 0 \\ \alpha_{13}^{(3)}(t) \end{pmatrix} Y_1(t)Y_3(t) + \begin{pmatrix} 0 \\ \alpha_{23}^{(2)}(t) \\ \alpha_{23}^{(3)}(t) \end{pmatrix} Y_2(t)Y_3(t) \\ &\quad + \begin{pmatrix} \alpha_{123}^{(1)}(t) \\ \alpha_{123}^{(2)}(t) \\ \alpha_{123}^{(3)}(t) \end{pmatrix} Y_1(t)Y_2(t)Y_3(t), \end{aligned}$$

where the first three terms on the r.h.s. of expression (4.15) represent the main effects, the following three terms the first-order interaction and the last term the second-order interaction.

Whenever the second-order interaction is negligible, the above expression turns out to be

$$(4.16) \quad \boldsymbol{\lambda}(t) = \text{Diag}(\mathbf{Y}(t))\boldsymbol{\alpha}(t)\mathbf{Y}(t),$$

and that resembles (2.6).

The same reasoning can be offered for higher dimension problems, taking into account the additional complications due to higher-order interactions. For example, for a K component problem, the intensity process will involve up to the $(K - 1)$ th order interaction. In fact, in this situation, each component will have intensity process that can be written as

$$\lambda_j(t) = \beta_j(t)Y_j(t) \left[1 + \sum_{\{l \in \{1, \dots, K\} \setminus j\}} \gamma_{jl}(t)Y_l(t) + \sum_{\{l \neq l' \in \{1, \dots, K\} \setminus j\}} \delta_{jll'}(t)Y_l(t)Y_{l'}(t) + \dots \right],$$

where $\beta_j = \alpha_{jj}^{(j)}$, $\gamma_{jl} = \alpha_{jl}^{(j)}/\alpha_{jj}^{(j)}$, $\delta_{jll'} = \alpha_{jll'}^{(j)}/\alpha_{jj}^{(j)}$ depend on the conditional hazard functions as in the case $K = 3$. In this representation the first term inside parenthesis is related to the independent situation, the second term with the first-order interaction, and so on. If it is reasonable to assume that the second and higher-order interactions are null, then this model can be rewritten as in (2.6) and (4.16); that is, in general,

$$(4.17) \quad \boldsymbol{\lambda}(t) = \begin{pmatrix} Y_1(t) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & Y_K(t) \end{pmatrix} \begin{pmatrix} \alpha_{11}(t) & \dots & \alpha_{1K}(t) \\ \vdots & \ddots & \vdots \\ \alpha_{K1}(t) & \dots & \alpha_{KK}(t) \end{pmatrix} \begin{pmatrix} Y_1(t) \\ \vdots \\ Y_K(t) \end{pmatrix}.$$

We conclude this section with a brief remark on the assumption that the second or higher-order interactions are null. It may be noted that the model represented by (4.17) has K^2 infinite-dimensional parameters represented by the functions $\alpha_{ij}(t)$. In order to reduce the dimensionality of the parameter space and to enhance estimability, additional assumptions are usually imposed on such models. This also enables us to study asymptotic properties of the estimators in a reduced model with greater flexibility. If higher-order interactions are allowed in the model, then the problem becomes much more complex. Assuming that interactions are null is a common practice in some fields of statistics (such as factorial experiments, multivariate paired comparisons) and we adopt this approach since we believe the simplifications are considerable. However, further investigation on the implications of such assumption is needed. A more careful examination shows that this assumption entails assuming that the failure times are *conditionally independent*. For example, when $K = 3$, assuming that there is no second-order interaction is

equivalent to saying that, given one of the failure times, the other two are conditionally independent. Finally we note that in the case where $K = 2$, no assumption is needed since the model will involve only first-order interactions that are being taken into account in model (4.17).

5. Application to the two-sample problem. Suppose that in a comparative clinical trial one is interested in studying the occurrence of two events, possibly censored, where each patient is randomized in one of two groups, *placebo* and *treatment*. Our main interest resides in studying the efficacy of treatment with respect to prolonging the time for the occurrence of one (or both) events. Here, $K = 2$ and we define a time-independent covariate X_i assuming values zero or one, depending on whether a particular individual i is assigned to the placebo or treatment groups. The variable $N_{ki}(t)$ indicates if the k th event has occurred for individual i at time t .

Asymptotic properties for the MPLE are given by Theorems 1 and 2, and we proceed to appraise the assumptions described there. Note that for a fixed t , each one of the $S_k^{(j)}(\cdot, t)$ may be thought of as an average of independent and identically distributed random variables. Therefore, one may apply the Khintchine law of large numbers and verify the pointwise (for each t) convergence (in probability) to appropriate deterministic functions. If this function is monotone, then considering the stochastic version of the Rockafellar lemma presented in Heiller and Willers (1988), we obtain that the pointwise convergence is equivalent to the uniform convergence in assumption C.2.

Thus, for $S_h^{(0)}(\delta_h^o, t) = (1/n) \sum_{i=1}^n w_i(\delta_h^o, t)$ we note that

$$E \{w_i(\delta_h^o, t)\} = \pi_0(\mathbf{S}_h^{\text{Pl}}(t) + \theta_h^o \mathbf{S}_{12}^{\text{Pl}}(t)) + \pi_1 \exp(\beta_h^o)(\mathbf{S}_h^{\text{Tr}}(t) + \theta_h^o \mathbf{S}_{12}^{\text{Tr}}(t)),$$

where $\pi_0 = 1 - \pi_1 = \mathbb{P}(X_i = 0)$ is the probability that a particular individual will be assigned to placebo, $\mathbf{S}_h^{\text{Pl}}(t) = \mathbb{P}(T_{hi} \geq t \mid X_i = 0)$ is the marginal survival function related to component h and $\mathbf{S}_{12}^{\text{Pl}}(t) = \mathbb{P}(T_{1i} \geq t; T_{2i} \geq t \mid X_i = 0)$ is the joint survival function. The functions $\mathbf{S}_h^{\text{Tr}}(t)$ and $\mathbf{S}_{12}^{\text{Tr}}(t)$ are similarly defined for the treatment group. Since both survival functions are nonincreasing functions, the uniform convergence in C.2 follows.

Similarly, $E[X_i^j(Y_{hi}(t) + \theta_h^o Y_{1i}(t)Y_{2i}(t)) \exp(\beta_h^o X_i)] = \pi_1 \exp(\beta_h^o)(\mathbf{S}_h^{\text{Tr}}(t) + \theta_h^o \mathbf{S}_{12}^{\text{Tr}}(t))$, for $j = 1, 2$, so that the monotonicity of $\mathbf{S}_h^{\text{Tr}}(t)$ and $\mathbf{S}_{12}^{\text{Tr}}(t)$ takes care of $S_h^{(1)}(t)$ and $S_h^{(2)}(t)$.

Note that $E[Y_{1i}(t)Y_{2i}(t) \exp(\beta_h^o X_i)] = (\pi_0 \mathbf{S}_{12}^{\text{Pl}}(t) + \pi_1 \exp(\beta_h^o)) \mathbf{S}_{12}^{\text{Tr}}(t)$, is also a monotone function, and, hence, the uniform convergence for $S_h^{(3)}(t)$ follows from the pointwise convergence.

We also have the same result for $S_h^{(4)}(t)$ since

$$E \left[\frac{Y_{1i}(t)Y_{2i}(t)}{Y_{hi}(t) + \theta_h^o Y_{1i}(t)Y_{2i}(t)} \exp(\beta_h^o X_i) \right] = \frac{1}{1 + \theta_h^o} (\pi_0 \mathbf{S}_{12}^{\text{Pl}}(t) + \pi_1 \exp(\beta_h^o)) \mathbf{S}_{12}^{\text{Tr}}(t)$$

is also a monotone function of t .

For $S_h^{(5)}$, note that $E[X_i Y_{1i}(t) Y_{2i}(t) \exp(\beta_h^\circ X_i)] = \pi_1 \exp(\beta_h^\circ) \mathbf{S}_{12}^{\text{Tr}}(t)$ is a monotone function of t . It follows then, for $j = 1, \dots, 5$ the uniform condition C.2 is true. Moreover, as $n \rightarrow \infty$,

$$\begin{aligned} \langle U_1^{(h,n)}, U_1^{(h,n)} \rangle(t) &\rightarrow_P \pi_0 \pi_1 \exp(\beta_h^\circ) \int_0^t [\mathbf{S}_h^{\text{Tr}}(u) + \theta_h^\circ \mathbf{S}_{12}^{\text{Tr}}(u)] [\mathbf{S}_h^{\text{Pl}}(u) + \theta_h^\circ \mathbf{S}_{12}^{\text{Pl}}(u)] \\ &\quad \times \left\{ \pi_0 [\mathbf{S}_h^{\text{Pl}}(u) + \theta_h^\circ \mathbf{S}_{12}^{\text{Pl}}(u)] \right. \\ &\quad \left. + \pi_1 \exp(\beta_h^\circ) [\mathbf{S}_h^{\text{Tr}}(u) + \theta_h^\circ \mathbf{S}_{12}^{\text{Tr}}(u)] \right\}^{-1} \alpha_h^\circ(u) du, \\ \langle U_2^{(h,n)}, U_2^{(h,n)} \rangle(t) &\rightarrow_P \int_0^t [\pi_0 \mathbf{S}_{12}^{\text{Pl}}(u) + \exp(\beta_h^\circ) \pi_1 \mathbf{S}_{12}^{\text{Tr}}(u)] \\ &\quad \times [\pi_0 (\mathbf{S}_h^{\text{Pl}}(u) - \mathbf{S}_{12}^{\text{Pl}}(u)) + \pi_1 \exp(\beta_h^\circ) (\mathbf{S}_h^{\text{Tr}}(u) - \mathbf{S}_{12}^{\text{Tr}}(u))] \\ &\quad \times [\pi_0 (\mathbf{S}_h^{\text{Pl}}(u) + \theta_h^\circ \mathbf{S}_{12}^{\text{Pl}}(u)) \\ &\quad + \pi_1 \exp(\beta_h^\circ) (\mathbf{S}_h^{\text{Tr}}(u) + \theta_h^\circ \mathbf{S}_{12}^{\text{Tr}}(u))]^{-1} \frac{\alpha_h^\circ(u)}{(1 + \theta_h^\circ)} du, \\ \langle U_1^{(h,n)}, U_2^{(h,n)} \rangle(t) &\rightarrow_P \pi_0 \pi_1 \exp(\beta_h^\circ) \int_0^t [\mathbf{S}_{12}^{\text{Tr}}(u) \mathbf{S}_h^{\text{Pl}}(u) - \mathbf{S}_h^{\text{Tr}}(u) \mathbf{S}_{12}^{\text{Pl}}(u)] \\ &\quad \times \left\{ \pi_0 [\mathbf{S}_h^{\text{Pl}}(u) + \theta_h^\circ \mathbf{S}_{12}^{\text{Pl}}(u)] \right. \\ &\quad \left. + \pi_1 \exp(\beta_h^\circ) [\mathbf{S}_h^{\text{Tr}}(u) + \theta_h^\circ \mathbf{S}_{12}^{\text{Tr}}(u)] \right\}^{-1} \alpha_h^\circ(u) du. \end{aligned}$$

Since in this case the continuity in conditions C.2 and C.3 are trivial, the asymptotic convergence of $n^{1/2}(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}^\circ)$ is given by a normal distribution with mean zero and covariance matrix $\boldsymbol{\Sigma}$ whose elements can be estimated by

$$\begin{aligned} (\widehat{\boldsymbol{\Sigma}}_h)_{11} &= \int_0^\tau \left\{ \frac{(Y_h^{\text{Tr}}(t) + \widehat{\theta}_h Y_{12}^{\text{Tr}}(t)) \exp(\widehat{\beta}_h)}{(Y_h^{\text{Pl}}(t) + \widehat{\theta}_h Y_{12}^{\text{Pl}}(t)) + (Y_h^{\text{Tr}}(t) + \widehat{\theta}_h Y_{12}^{\text{Tr}}(t)) \exp(\widehat{\beta}_h)} \right. \\ &\quad \left. - \left(\frac{(Y_h^{\text{Tr}}(t) + \widehat{\theta}_h Y_{12}^{\text{Tr}}(t)) \exp(\widehat{\beta}_h)}{(Y_h^{\text{Pl}}(t) + \widehat{\theta}_h Y_{12}^{\text{Pl}}(t)) + (Y_h^{\text{Tr}}(t) + \widehat{\theta}_h Y_{12}^{\text{Tr}}(t)) \exp(\widehat{\beta}_h)} \right)^2 \right\} d\bar{N}_h(t), \\ (\widehat{\boldsymbol{\Sigma}}_h)_{22} &= \frac{1}{(1 + \widehat{\theta}_h)^2} \int_0^\tau d\bar{N}_h^{11}(t) \\ &\quad - \int_0^\tau \left(\frac{Y_{12}^{\text{Pl}}(t) + \exp(\widehat{\beta}_h) Y_{12}^{\text{Tr}}(t)}{(Y_h^{\text{Pl}}(t) + \widehat{\theta}_h Y_{12}^{\text{Pl}}(t)) + (Y_h^{\text{Tr}}(t) + \widehat{\theta}_h Y_{12}^{\text{Tr}}(t)) \exp(\widehat{\beta}_h)} \right)^2 d\bar{N}_h(t), \\ (\widehat{\boldsymbol{\Sigma}}_h)_{12} &= (\widehat{\boldsymbol{\Sigma}}_h)_{21} \\ &= \int_0^\tau \left(\frac{Y_{12}^{\text{Tr}}(t) \exp(\widehat{\beta}_h)}{(Y_h^{\text{Pl}}(t) + \widehat{\theta}_h Y_{12}^{\text{Pl}}(t)) + (Y_h^{\text{Tr}}(t) + \widehat{\theta}_h Y_{12}^{\text{Tr}}(t)) \exp(\widehat{\beta}_h)} \right. \\ &\quad \left. - \frac{(Y_{12}^{\text{Pl}}(t) + Y_{12}^{\text{Tr}}(t) \exp(\widehat{\beta}_h)) (Y_h^{\text{Tr}}(t) + \widehat{\theta}_h Y_{12}^{\text{Tr}}(t)) \exp(\widehat{\beta}_h)}{[(Y_h^{\text{Pl}}(t) + \widehat{\theta}_h Y_{12}^{\text{Pl}}(t)) + (Y_h^{\text{Tr}}(t) + \widehat{\theta}_h Y_{12}^{\text{Tr}}(t)) \exp(\widehat{\beta}_h)]^2} \right) d\bar{N}_h(t), \end{aligned}$$

where at time t , $Y_h^{\text{Tr}}(t)$ and $Y_h^{\text{Pl}}(t)$ represent the number of individuals with the component h at risk for treatment and placebo groups respectively, $Y_{12}^{\text{Tr}}(t)$ and $Y_{12}^{\text{Pl}}(t)$ represent the number of individuals with both components at risk for treatment and placebo groups, and $\bar{N}_{h^*}^{11}(t)$ represents the number of failures divided by n for those individuals with no failure in any component.

Acknowledgments. The authors are grateful to the Editor and the referees for their valuable comments on the manuscript.

REFERENCES

- AALEN, O. O. (1975). Statistical inference for a family of counting processes. Ph.D. dissertation, Dept. Statistics, Univ. California, Berkeley.
- AALEN, O. O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6** 701–726.
- ANDERSEN, P. K. and GILL, R. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10** 1100–1120.
- ANDERSEN, P. K., BORGAN, Ø., GILL, R. D. and KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York.
- BRÉMAUD, P. (1981). *Point Processes and Queues*. Springer, New York.
- CAI, J. and PRENTICE, R. L. (1995). Estimating equations for hazard ratio parameters on correlated failure time data. *Biometrika* **82** 151–164.
- DABROWSKA, D. (1988). Kaplan–Meier estimate on the plane. *Ann. Statist.* **16** 1475–1489.
- FLEMING, T. R. and HARRINGTON, D. P. (1991). *Counting Processes and Survival Analysis*. Wiley, New York.
- HEILLER, S. and WILLERS, R. (1988). Asymptotic normality of R -estimates in the linear model. *Statistics* **19** 173–184.
- LE CAM, L. (1956). On the asymptotic theory of estimation and testing hypotheses. *Proc. Third Berkeley Symp. Math. Statist. Probability* **1** 129–156. Univ. California Press, Berkeley.
- PEDROSO DE LIMA, A. C. (1995). A Matrix-valued counting process model with applications to survival analysis. Ph.D. dissertation, Dept. Biostatistics, Univ. North Carolina, Chapel Hill.
- PEDROSO DE LIMA, A. C. and SEN, P. K. (1997). Bivariate exponentials, dual counting processes and their intensities. *J. Combin. Inform. System Sci.* To appear.
- PRENTICE, R. L. and CAI, J. (1992). Covariance and survivor function estimation using censored multivariate failure time data. *Biometrika* **79** 495–512.
- SARKAR, S. K. (1987). A continuous bivariate exponential distribution. *J. Amer. Statist. Assoc.* **82** 667–675.

DEPARTAMENTO DE ESTATÍSTICA
UNIVERSIDADE DE SÃO PAULO
CAIXA POSTAL 66281
05315-970 SÃO PAULO, S.P.
BRAZIL
E-MAIL: acarlos@ime.usp.br

DEPARTMENT OF BIostatISTICS
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NORTH CAROLINA 27599-7400