STOCHASTIC CALCULUS AS A TOOL IN SURVIVAL ANALYSIS: A REVIEW

by

Ian W. McKeague[1] and Klaus J. Utikal

FSU Technical Report No. M-790
USARO Technical Report No. D-102

June, 1988
Revised: March, 1989

Department of Statistics
The Florida State University
Tallahassee, FL 32306-3033

This is an invited paper for a special issue of Applied Mathematics and Computation (1990, vol. 38, pp. 23–49) devoted to "Theory and Applications of Stochastic Differential Equations (Ordinary and Partial)".

[1] Research supported by the Army Research Office under Grant DAAL03-86-K-0094.

AMS 1980 subject classifications. 60G44, 62G05, 62M99.

Key words and phrases. Censored survival data, counting processes, martingale methods.

1. Introduction

Beginning with Aalen's 1975 thesis from Berkeley, there has been a rapid increase in the use of stochastic calculus as a tool in the study of survival analysis. Aalen realized that multivariate counting processes provide a natural framework for the study of censored survival data and that a central role is played by martingales, predictable processes and stochastic integrals. The counting process approach has been successfully applied in the study of Nelson's cumulative hazard estimator (Aalen, 1978), the Kaplan–Meier product-limit estimator (Aalen and Johansen, 1978; Gill, 1980, 1983), nonparametric k-sample tests (Andersen et al., 1982) and Cox's proportional hazards regression model (Andersen and Gill, 1982), to name just a few examples.

The list of such applications is now quite long, and as new estimators, models and data structures are introduced, they too can often be treated under the umbrella of the counting process approach. Some of the benefits of this are: (1) we can bring to bear powerful results from the theory of stochastic processes; (2) more general censoring patterns can be allowed; (3) i.i.d. assumptions no longer play a central role; (4) straightforward, yet rigorous, proofs can be given. Although some background in stochastic processes is needed to appreciate these advantages, it turns out (as remarked by Arjas, 1985) that the σ-fields and martingales involved are far more concrete and practical than the traditional appoach via elementary mathematics.

The purpose of the present paper is to outline these developments with special emphasis on some of the recent applications of Rebolledo's martingale central limit theorem to the study of asymptotic distributions of estimators and test statistics. We shall make no attempt to review the stochastic calculus relevant to survival analysis, but rather refer to the books of Liptser and Shiryayev (1978) and Kopp (1984) as necessary. Earlier survey articles and books containing material on the use of counting process theory in survival analysis have been written by Gill (1980), Jacobsen (1982), Davis (1983), Andersen and Borgan (1985), Shorack and Wellner (1986), Karr (1986), Prakasa Rao (1987) and Andersen et al. (1988). We also mention the forthcoming book of Andersen et al. (1989).

2. The Nelson–Aalen and Kaplan–Meier estimators

2.1. Cumulative hazard and survival functions

Let T, representing the survival time of an individual, be a positive random variable with distribution function F having a density f. The hazard function (or failure rate function) of T is defined by $\lambda(t) = f(t)/S(t)$ for t such that $S(t) = 1 - F(t) > 0$. Here S is called the survival function. Since

$$\lambda(t) = \lim_{\Delta t \downarrow 0} \frac{P(t < T \leq t + \Delta t | T > t)}{\Delta t}$$

we may interpret $\lambda(t)\Delta t$ as the probability of failure in the time interval t to $t + \Delta t$ given survival up to time t. It is easily seen that the hazard function determines the distribution. For

$$\lambda(t) = -\frac{d}{dt} \log S(t),$$

so that, in terms of the cumulative hazard function $\Lambda(t) = \int_0^t \lambda(s)\,ds$, we have $S(t) = e^{-\Lambda(t)}$.

When F does not have a density, the hazard function cannot be defined, but the cumulative hazard function is defined by

$$\Lambda(t) = \int_{(0,t]} \frac{dF(u)}{S(u-)}$$

for t such that $S(t-) = \lim_{u \uparrow t} S(u) > 0$. Throughout the paper we shall consider Λ and S only on a fixed interval $[0, \tau]$ where $\tau > 0$ is a time point such that $S(\tau-) > 0$. The survival function can be represented in terms of the cumulative hazard function by $S = \mathcal{E}(-\Lambda)$, where $\mathcal{E}$ is the Doléans-Dade exponential defined by the following result.

Theorem 2.1. (A special case of Doléans-Dade (1970)) Let X be a right-continuous function of bounded variation with $X_0 = 0$. Then the equation

$$Z_t = 1 + \int_{(0,t]} Z_{s-}\, dX_s$$

has a unique solution which is bounded on finite intervals. The solution, denoted $\mathcal{E}(X)_t$, is given by Z_t equal to

$$\mathcal{E}(X)_t = e^{X_t^c} \prod_{s \leq t} (1 + \Delta X_s),$$

where $X_t^c = X_t - \sum_{s \leq t} \Delta X_s$; $\Delta X_s = X_s - X_{s-}$.

A proof of this result can be found in Liptser and Shiryayev (1978, pp.255–256). As an immediate consequence (see Wellner, 1985) we obtain the formula

$$S = \mathcal{E}(-\Lambda), \tag{2.1}$$

since S satisfies the equation

$$S(t) = 1 + \int_{(0,t]} S(u-)\, d(-\Lambda(u)).$$

An alternative way of representing S in terms of Λ is to use product integration. Let X be a finite Borel measure on the positive real line and denote its distribution function by the same symbol: $X(t) = X((0,t])$. The product integral of X is defined by

$$\prod_{(0,t]} (1 + dX) = \lim_{\max |t_i - t_{i-1}| \to 0} \prod_i (1 + X((t_{i-1}, t_i])),$$

where $0 = t_0 < t_1 < \ldots < t_n = t$ is a partition of $(0, t]$. It can be shown (see Gill and Johansen, 1988) that

$$S(t) = \prod_{(0,t]} (1 - d\Lambda). \tag{2.2}$$

2.2. Random censorship models and the counting process formulation

Censoring is the presence of incomplete or missing observations in survival data. It typically takes the form of random right or left censorship. For example, consider a study for determining the distribution of the age T at which a certain chronic disease or other permanent condition appears in an individual. Right censoring occurs if the individual dies or leaves the study before the disease appears. However, the disease may have already appeared before the individual entered the study, and this results in left censoring. In each case, the exact value of T cannot be determined, but some useful information is still available. One of the most elegant features of the counting process formulation is that it unifies these and more general censoring schemes under the notion of "predictable censoring". Estimators of the survival function which derive from such schemes can then be studied in a unified way.

Define the basic counting process $N_t^* = I(T \le t)$. In the absence of censoring, N_t^* is the only counting process we need to consider. To introduce censoring into the picture, let $(C_t,\ t \ge 0)$ be a $\{0,1\}$-valued stochastic process, called the censoring process, which is understood to indicate censorship at time t if $C_t = 0$. The observed counting process is given by

$$N_t = \int_0^t C_s dN_s^* = \begin{cases} 1 & \text{if } T \le t \text{ and } C_T = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Next, define $\mathcal{F}_t^T = \sigma(N_s^*,\ s \le t)$, the σ-field generated by N^* up to time t. Let $(\mathcal{G}_t)$ be a right-continuous filtration such that $\mathcal{G}_t$ is independent of T for all t and write $\mathcal{F}_t = \mathcal{F}_t^T \vee \mathcal{G}_t$, the σ-field generated by $\mathcal{F}_t^T$ and $\mathcal{G}_t$. Also define the processes $R_t = I(T \ge t)$ and $Y_t = C_t R_t$, so that Y_t is the indicator that the individual is observed to be at risk at time t. The following result is crucial.

Theorem 2.2. Suppose that the censoring process (C_t) is predictable with respect to the filtration $(\mathcal{F}_t)$. Then the process

$$M_t = N_t - \int_0^t Y_s \, d\Lambda_s \tag{2.3}$$

$(0 \le t \le \tau)$ is an $\mathcal{F}_t$-martingale.

In the language of counting processes this result is saying that the counting process N_t has compensator $A_t = \int_0^t Y_s \, d\Lambda_s$ with respect to the filtration $(\mathcal{F}_t)$, see Liptser and Shiryayev (1978, p. 239). If T has a hazard function $\lambda(t)$, which we assume in the sequel, then the result implies that N_t has intensity process $Y_t \lambda(t)$ and we may write (2.3) in the form of a stochastic differential equation:

$$dN_t = Y_t \lambda(t) \, dt + dM_t.$$

The result is intuitively reasonable in view of the interpretation of $Y_t \lambda(t) \, dt$ as the probability of observing a failure in the time interval t to $t + dt$ given survival up to time t.

The censoring process (C_t) is left-continuous in most applications, so in order to show that it is predictable, it suffices to show that it is $(\mathcal{F}_t)$-adapted. In particular, random right and left censorship schemes are obtained by taking $C_t = I(L < t \leq C)$, where L and C are the left and right "censoring times" respectively, assuming that (L, C) is independent of T, and setting $\mathcal{G}_t = \mathcal{G}_0 = \sigma(L, C)$. For this type of censoring T is observed only when it lies in the interval $[L, C]$, in which case L is observed as well. If T is not in $[L, C]$, it is only known whether $T < L$ or $T > C$, with L or C observed correspondingly.

The counting process approach to the analysis of censored survival data can be applied only when the processes (N_t) and (Y_t) are observable. This is the case for the left and right censoring schemes above. However, if L is not observable when T lies in the interval $[L, C]$, as for the double censoring scheme considered by Turnbull (1974), then (Y_t) is not observable. Then alternative approaches need to be used; an iterative method (based on the EM algorithm) for the calculation of maximum likelihood estimators of the survival function from doubly censored data was developed by Turnbull (1974), see also Chang and Yang (1987).

In fact the counting process approach is *fully* successful only when $N = (N_t)$ and $Y = (Y_t)$ contain all the information that is present in the data concerning the survival function (i.e. (N, Y) is a sufficient statistic). Otherwise, estimators based on N and Y will not be efficient. That happens with left censoring, since if $T < L$ we observe L *as well as* knowing that $T < L$, yet the processes N and Y are identically zero. The difficulty here is that the censoring process (C_t) carries information about the survival function not available from N and Y when $T < L$; see Andersen et al. (1988, p.51).

Proof of Theorem 2.2. The main step in the proof is to show that $M_t^* = N_t^* - \int_0^t R_s d\Lambda_s$ is an $\mathcal{F}_t^T$-martingale. Then, since $\mathcal{G}_t$ is independent of $\mathcal{F}_t^T$, it follows from a standard result on conditional expectation (see Chung, 1974, p.308) that M_t^* is an $\mathcal{F}_t$-martingale. Consequently, since $M_t = \int_0^t C_s dM_s^*$ can be interpreted as a stochastic integral (see Kopp, 1984, p.149), M_t is an $\mathcal{F}_t$-martingale. Now to show that M_t^* is a martingale, let $u < t \leq \tau$ and note that

$$E\left(M_t^* - M_u^* | \mathcal{F}_u^T\right) = E\left(I(T > u)\left(M_t^* - M_u^*\right) | \mathcal{F}_u^T\right) = I(T > u)\, E\left(M_t^* - M_u^* \,| T > u\right)$$

since, by a simple argument using the π-λ theorem, $A \cap \{T > u\}$ is either the empty set or $\{T > u\}$ for any $A \in \mathcal{F}_u^T$. Also

$$E\left(N_t^* - N_u^* \,| T > u\right) = P(u < T \leq t | T > u) = \frac{F(t) - F(u)}{S(u)}$$

and with $A_t = \int_0^t R_s \, d\Lambda_s = \Lambda_T$,

$$\begin{aligned} E\left(A_t - A_u | T > u\right) &= E\left(A_t - A_u | T > t\right) P(T > t | T > u) \\ &\quad + E(A_t - A_u | u < T \leq t)\, P(T \leq t | T > u) \\ &= (\Lambda(t) - \Lambda(u)) \frac{S(t)}{S(u)} + \int_{(u,t]} \frac{(\Lambda(v) - \Lambda(u))}{F(t) - F(u)} dF(v)\, \frac{(F(t) - F(u))}{S(u)} \\ &= \frac{1}{S(u)} \int_{(u,t]} S(v-)\, d\Lambda(v) \\ &= \frac{F(t) - F(u)}{S(u)} = E\left(N_t^* - N_u^* | T > u\right), \end{aligned}$$

where we have used the integration by parts formula for Stieltjes integrals (see Liptser and Shiryayev, 1978, p.253). This completes the proof.

The basic idea in the above proof is that the martingale property of $(M_t^*, \mathcal{F}_t^T)$ is preserved when independent events are added to $\mathcal{F}_t^T$. It is natural to ask "How much can we enlarge $\mathcal{F}_t^T$, while preserving the martingale property of M_t^* ?" We know of no general answer to this question, although some recent work of Jacobsen (1986, 1988) dealing with right-censoring is of interest in this regard.

To extend the counting process formulation to n individuals with corresponding survival times $T_1, ..., T_n$, introduce processes N_i, Y_i, M_i, $i = 1, ..., n$ and filtrations $(\mathcal{F}_{it})$, $i = 1, ..., n$ having the same structure as N, Y, M and $(\mathcal{F}_t)$. It is convenient to view the martingales M_i, $i = 1, ..., n$ with respect to the same filtration. If $\mathcal{F}_{1t}, ..., \mathcal{F}_{nt}$ are independent σ-fields for all t, then $M_1, ..., M_n$ are martingales (in fact orthogonal martingales) with respect to the filtration $\mathcal{F}_t^{(n)} = \mathcal{F}_{1t} \vee \cdots \vee \mathcal{F}_{nt}$. For some applications, however, it may be too restrictive to assume that the $\mathcal{F}_{it}$ are independent. The general counting process formulation does not require any such assumption; it only requires that $M_1, ..., M_n$ are martingales with respect to *some* filtration $(\mathcal{F}_t^{(n)})$. There is no loss of generality in setting $\tau = 1$, so in the sequel all processes are defined on the interval $[0, 1]$.

Aalen's (1978) *multiplicative intensity model* is now formulated as follows. Let $N(t) = (N_1(t), ..., N_n(t))'$, $t \in [0, 1]$ be a multivariate counting process with respect to a right-continuous filtration $(\mathcal{F}_t^{(n)})$ with $\mathcal{F}_0^{(n)}$ containing all subsets of P-null sets of $\mathcal{F}_1^{(n)}$. This means that N is adapted to the filtration and has components N_i which are right-continuous step functions, zero at time zero, with jumps of size $+1$ such that no two components jump simultaneously. The counting process N_i is assumed to have intensity

$$\lambda_i(t) = Y_i(t)\, \lambda(t), \tag{2.4}$$

where $Y_i(t)$ is a predictable $\{0, 1\}$-valued process and λ is a fixed hazard function. By Davis (1983, Proposition 4.2)

$$M_i(t) = N_i(t) - \int_0^t \lambda_i(s)\, ds, \quad i = 1, ..., n \tag{2.5}$$

are orthogonal local square integrable martingales on $[0, 1]$ with predictable variation process

$$\langle M_i, M_i \rangle_t = \int_0^t \lambda_i(s)\, ds. \qquad (2.6).$$

If $E\, N_i(1) < \infty$ then M_i is in fact a square integrable martingale on $[0, 1]$. As Aalen does, we shall assume throughout that M_i, $i = 1, ..., n$ are square integrable martingales.

2.3. Left truncation

Left truncation, or delayed entry, is a type of selection bias (cf. Vardi, 1985) found in survival data. It arises when the observed survival time, denoted T', is distributed according to the conditional law of the survival time of interest T given that T is larger than some random time of left truncation L. For instance, in our previous example of the age T at which a certain chronic disease or other permanent condition appears in an individual, left truncation will be present if we only sample from among individuals who are free of the disease at the start of the study. Then the distribution of the observed time T' that the disease appears is the distribution of T truncated by L = the age of the individual at the start of the study. Estimation of a survival function based on left truncated data has been studied by Woodroofe (1985) and Wang, Jewell and Tsai (1986). It has only recently been realized that estimators based on such data can be studied within the counting process framework, see Andersen et al. (1988) and Keiding and Gill (1988).

Left truncated data consists of n i.i.d. replications $(L'_i, T'_i), i = 1, \ldots, n$ of (L', T'), where (L', T') are random variables with the conditional distribution of (L, T) given $L < T$. Andersen et al. (1988) prove a "left truncated" version of Theorem 2.2, which we now briefly describe in our notation. Assume that T and L are independent and $P(L < T) > 0$. Replace T by T' in the processes N^*, R_t and the σ-field $\mathcal{F}_t^T$ of Section 2.2, and set $C_t = I(L' < t)$, $\mathcal{G}_t = \sigma(L')$. Then the "left truncated" counting process $N_t = \int_0^t C_s \, dN_s^*$ satisfies the conclusion of Theorem 2.2, where Λ is the cumulative hazard function of T as before. Thus, if T has a hazard function λ, the i.i.d. replications $N_1, \ldots, N_n$ of N satisfy Aalen's multiplicative intensity model described above, with $Y_i(t) = I(L'_i < t \le T'_i)$. Also, $(N_i, Y_i), i = 1, \ldots, n$ are observable and contain all the information in the data.

2.4. The Nelson–Aalen estimator

An estimator for the cumulative hazard function Λ was introduced by Nelson (1969) in the case of right-censored survival data, and extended by Aalen (1975, 1978) to his multiplicative intensity model. The so called Nelson–Aalen estimator is defined by

$$\hat{\Lambda}(t) = \int_0^t \frac{dN^{(n)}(s)}{Y^{(n)}(s)},$$

where $N^{(n)} = \sum_{i=1}^n N_i$, $Y^{(n)} = \sum_{i=1}^n Y_i$ and $1/0 \equiv 0$. A motivation for this estimator is provided by formally solving the stochastic differential equation

$$dN_t^{(n)} = Y_t^{(n)} d\Lambda_t + dM_t^{(n)} \tag{2.7}$$

(where $M^{(n)} = \sum_{i=1}^n M_i$) for Λ_t and ignoring the "noise" term $\int_0^t \frac{dM^{(n)}(s)}{Y^{(n)}(s)}$ (which is a martingale).

The asymptotic distribution of $\hat{\Lambda}$ can be derived under the following asymptotic stability condition:

(AS) There exists a function ρ, bounded away from zero on $[0,1]$, such that $\bar{Y}^{(n)}(t) \equiv \frac{1}{n} Y^{(n)}(t)$ satisfies

$$\sup_{t \in [0,1]} |\bar{Y}^{(n)}(t) - \rho(t)| \xrightarrow{P} 0 \quad \text{as } n \to \infty.$$

Let $D[0,1]$ denote the space of functions on $[0,1]$ which are right-continuous on $[0,1)$ with left limits on $(0,1]$, and equip it with the Skorohod topology (see Billingsley, 1968, p.111). Convergence in distribution will be denoted $\xrightarrow{\mathcal{D}}$.

Theorem 2.3. (Aalen, 1978) Suppose that the asymptotic stability condition (AS) holds. Then, under Aalen's multiplicative intensity model, $\sqrt{n}(\hat{\Lambda}-\Lambda)\xrightarrow{\mathcal{D}}m$ in $D[0,1]$ as $n\to\infty$, where m is a continuous Gaussian martingale with covariance function

$$\mathrm{Cov}(m_s, m_t) = \int_0^{s\wedge t} \frac{\lambda(u)}{\rho(u)}\, du.$$

Proof. Using (2.7) we have

$$\sqrt{n}(\hat{\Lambda}(t)-\Lambda(t)) = \bar{M}_t^{(n)} - R_t^{(n)}, \tag{2.8}$$

where

$$\bar{M}_t^{(n)} = \frac{1}{\sqrt{n}}\int_0^t \frac{dM^{(n)}(s)}{\bar{Y}^{(n)}(s)}, \quad R_t^{(n)} = \sqrt{n}\int_0^t I(\bar{Y}^{(n)}(s)=0)\, d\Lambda_s.$$

Since the condition (AS) implies that

$$\sup_{t\in[0,1]} |R_t^{(n)}| \xrightarrow{P} 0, \tag{2.9}$$

to complete the proof it suffices to show that $\bar{M}^{(n)}\xrightarrow{\mathcal{D}}m$ in $D[0,1]$. Note that $\bar{M}_t^{(n)}$ is a square integrable $\mathcal{F}_t^{(n)}$-martingale. Now apply the version of Rebolledo's (1980) martingale central limit theorem stated in Andersen and Gill (1982) with $p=1$ and $H_{1l}^{(n)}(t) = n^{-\frac{1}{2}}(\bar{Y}^{(n)}(t))^{-1}$. By (2.5), (2.6) and (AS) we have

$$\langle \bar{M}^{(n)}, \bar{M}^{(n)}\rangle_t = \int_0^t \frac{\lambda(s)}{\bar{Y}^{(n)}(s)}\, ds \xrightarrow{P} \int_0^t \frac{\lambda(s)}{\rho(s)}\, ds. \tag{2.10}$$

The Lindeberg condition, here given by

$$\frac{1}{n}\sum_{i=1}^n \int_0^1 \frac{1}{(\bar{Y}^{(n)}(s))^2}\lambda_i(s)\, I\Big(\Big|\frac{1}{\bar{Y}^{(n)}(s)}\Big| > \epsilon\sqrt{n}\Big)\, ds \xrightarrow{P} 0$$

for all $\epsilon > 0$, follows from (AS). This completes the proof.

It is possible to use Theorem 2.3 to construct confidence bands for Λ, see Andersen and Borgan (1985, p.114). In many applications it is of interest to estimate the hazard function λ itself. It is possible to develop an asymptotic distribution theory for pointwise estimators of $\lambda(t)$, $\hat{\lambda}(t)$ say, using Rebolledo's martingale central limit theorem, much as in the proof of Theorem 2.3. This has been done for kernel estimators (Ramlau-Hansen, 1983), spline sieve estimators (Karr, 1987), grouped data based estimators (Borgan and Ramlau-Hansen, 1985; McKeague, 1988b) and penalized maximum likelihood estimators (Antoniadis, 1987). Integrating any one of these estimators provides another estimator $\int_0^\cdot \hat{\lambda}(s)\,ds$ of Λ, which (not surprisingly) turns out to have the same asymptotic distribution as the Nelson–Aalen estimator.

Note that in the i.i.d. case, in which $Y_i,\ i = 1, ..., n$ are i.i.d. replicates of one another and Y_i has left-continuous paths with right hand limits, the asymptotic stability condition (AS) can be checked using Ranga Rao's (1963) law of large numbers. For the right and left censoring schemes we have $\rho(t) = P(L < t \le T \wedge C)$; and under left truncation $\rho(t) = P(L' < t \le T')$. The condition that the function ρ is bounded away from zero will not be satisfied close to $t = 0$ when there is left censoring or left truncation. This problem can be dealt with by first only estimating $\Lambda^\epsilon = \Lambda - \Lambda(\epsilon)$ on an interval $[\epsilon, 1]$ over which (AS) is satisfied, and then showing that $\sqrt{n}\sup_{t\in[0,\epsilon]} |\hat{\Lambda}(t) - \Lambda(t)|$ becomes negligible, uniformly in n, as $\epsilon \to 0$. This has been done for left truncation under the assumption that $\int_0^\epsilon [G(t)]^{-1}\,dF(t) < \infty$ for some (and so for all) $\epsilon > 0$, where G is the distribution function of L; see Keiding and Gill (1988).

2.5. The Kaplan–Meier estimator

In view of (2.1) and (2.2) it is reasonable to estimate the survival function S by the Doléans-Dade exponential or product integral of $-\hat{\Lambda}$, where $\hat{\Lambda}$ is the Nelson–Aalen estimator. Define

$$\widehat{S}(t) = \mathcal{E}(-\hat{\Lambda})_t = \prod_{(0,t]} (1 - d\hat{\Lambda}).$$

Since the continuous part of $\hat{\Lambda}$ is zero, $\widehat{S}$ reduces to the so called "product-limit" estimator

$$\widehat{S}(t) = \prod_{s\le t}\Big(1 - \frac{\Delta N_s^{(n)}}{Y_s^{(n)}}\Big),$$

which was originally introduced by Kaplan and Meier (1958) in the case of right-censored survival data.

Breslow and Crowley (1974) gave the first proof of weak convergence of the Kaplan–Meier estimator; Gill (1980, 1983) gave a proof based on martingale methods. We shall present Gill's proof in the context of the multiplicative intensity model. In the classical i.i.d. random censorship model other proofs are possible. Gill and Johansen (1988) recently gave a proof using Hadamard differentiability and a functional version of the delta method. That approach works for general (not necessarily continuous) survival functions and can be used to study the asymptotic behavior of the bootstrapped Kaplan–Meier estimator; see Gill (1987). David Pollard (1988) has informed us that a proof via the theory of empirical processes is also possible. We refer the reader to Akritas (1986), Horváth and Yandell (1987) and Lo and Singh (1986) for results on the bootstrapped Kaplan–Meier estimator.

Theorem 2.4. Suppose that the asymptotic stability condition (AS) holds. Then, under Aalen's multiplicative intensity model, $\sqrt{n}(\widehat{S} - S) \xrightarrow{\mathcal{D}} S(\cdot)\, m(\cdot)$ in $D[0, 1]$ as $n \to \infty$, where m is the Gaussian martingale of Theorem 2.3.

Proof. First note that $\mathcal{E}(\Lambda - \hat{\Lambda})_t = \mathcal{E}(\Lambda)_t \mathcal{E}(-\hat{\Lambda})_t = \widehat{S}_t / S_t$, so, by Theorem 2.1,

$$\frac{\widehat{S}_t}{S_t} = 1 + \int_0^t \frac{\widehat{S}_{u-}}{S_u} d(\Lambda - \hat{\Lambda})(u).$$

Thus, using (2.8), we obtain

$$\sqrt{n}(\widehat{S}_t - S_t) = -S_t \int_0^t \frac{\widehat{S}_{u-}}{S_u} d\bar{M}_u^{(n)} + R_t^{(n)}, \tag{2.11}$$

where $R_t^{(n)}$ is a remainder term (different from the original $R_t^{(n)}$) satisfying (2.9). To complete the proof, it suffices to show that $\bar{m}^{(n)} \xrightarrow{\mathcal{D}} m$ in $D[0,1]$, where

$$\bar{m}_t^{(n)} = \int_0^t \frac{\widehat{S}_{u-}}{S_u} \, d\bar{M}_u^{(n)}.$$

Now $\bar{m}^{(n)}$ is a square integrable martingale with predictable variation process

$$\langle \bar{m}^{(n)}, \bar{m}^{(n)} \rangle_t = \int_0^t \left[\frac{\widehat{S}_{u-}}{S_u} \right]^2 \frac{\lambda(u)}{\bar{Y}^{(n)}(u)} \, du.$$

Thus, by (AS), we have $\langle \bar{m}^{(n)}, \bar{m}^{(n)} \rangle_1 = O_P(1)$. Using Lenglart's (1977) inequality and (2.11) it follows that $\sup_{t \in [0,1]} |\widehat{S}_t - S_t| \xrightarrow{P} 0$. Hence

$$\langle \bar{m}^{(n)}, \bar{m}^{(n)} \rangle_t \xrightarrow{P} \int_0^t \frac{\lambda(u)}{\rho(u)} \, du = \langle m, m \rangle_t.$$

The Lindeberg condition for $\bar{m}^{(n)}$ is checked in the same way it was checked for $\bar{M}^{(n)}$. The result follows by Rebolledo's martingale central limit theorem.

It is natural to ask whether the above results have any extension to two-dimensional survival times, $T = (T_1, T_2)$ say. Data of that kind can arise, for example, in a study of the ages T_1, T_2 at which two different chronic diseases appear in an individual. Unfortunately, many of the techniques that are useful in the univariate case are no longer applicable in the bivariate case. In particular, a two-parameter martingale central limit theorem is not available. The best results that are currently available are all for the i.i.d. case with right censoring and rely on classical methods; see Tsai, Leurgans and Crowley (1986) and Dabrowska (1988) for instance.

3. Regression models for survival data

In most applications of survival analysis it is important to consider the effects that covariates may have upon the survival times of individuals in the study. This can be done by using a regression model for the conditional hazard function $\lambda(t, z) = \lambda(t|z)$ of the survival time of an individual who has a covariate vector $z = (z_1, ..., z_p)'$, say, at time t. The well known proportional hazards model of Cox (1972) has been the most popular model, but in recent years other models have begun to be considered. We list Cox's model and various other alternative nonparametric and semiparametric models with which we are familiar as follows.

(1) Cox's (1972) proportional hazards model:

$$\lambda(t,z) = \lambda_0(t)\, e^{\beta_0' z},$$

where λ_0 is an unknown baseline hazard function and β_0 is a vector of p unknown parameters.

(2) Aalen's (1980) additive risk model:

$$\lambda(t,z) = \sum_{j=1}^{p} \alpha_j(t)\, z_j,$$

where $\alpha_1, ..., \alpha_p$ are unknown functions.

(3) The general nonparametric model. Beran (1981) considered

$$\lambda(t,z) \text{ is arbitrary.}$$

(4) Variations on Cox's model. The general proportional hazards model

$$\lambda(t,z) = \lambda_0(t)\, r(z),$$

where λ_0 is an unknown baseline hazard function and r is an unknown "relative risk" function was proposed by Thomas (1983). Hastie and Tibshirani (1987) suggested the generalized additive model $r(z) = \sum_{j=1}^{p} r_j(z_j)$, where $r_1, ..., r_p$ are unknown functions. Prentice and Self (1983) take $r(z) = r_0(\beta_0' z)$, where r_0 is known and β_0 is a vector of p unknown parameters. Zucker (1986) and Zucker and Karr (1987) generalized Cox's model by allowing β_0 to be time-dependent.

Since the papers of Andersen and Gill (1982) and Aalen (1980), which developed asymptotic theory for the models (1) and (2), martingale methods have been used to obtain asymptotic theory for most of these models. In this section we review some of that work. Throughout we shall use the following counting process framework, extending the multiplicative intensity model to allow for covariates. Suppose that $N(t) = (N_1(t), ..., N_n(t))'$ is a multivariate counting process with respect to a right-continuous filtration $(\mathcal{F}_t^{(n)})$. The counting process N_i, which records events in the life of the ith individual, is assumed to have intensity process $\lambda_i(t) = Y_i(t)\, \lambda(t, Z_i(t))$, where $Y_i(t)$ is a predictable $\{0,1\}$-valued process as before, and $Z_i(t) = (Z_{i1}(t), ..., Z_{ip}(t))'$ is a p-vector of predictable covariate processes. The martingales $M_1, ..., M_n$ are again defined by (2.5). For simplicity, we shall only consider the i.i.d. case in which (N_i, Y_i, Z_i), $i = 1, ..., n$ are i.i.d. replicates of (N, Y, Z). Also, assume that Y and Z are left-continuous with right hand limits and the covariate processes are bounded.

3.1. Cox's proportional hazards model

In this section we shall briefly sketch the main results of Andersen and Gill (1982). We refer to the review paper of Davis (1983) for an informal discussion of these results and the motivation behind the estimators. For simplicity of presentation, we shall assume that the covariates are scalar valued ($p = 1$).

Cox (1972, 1975) proposed that inference for β_0 in (1) be based on the partial likelihood function

$$L(\beta) = \prod_{i=1}^{n} \Big\{ \frac{e^{\beta Z_i(T_i)}}{\sum_{j \in \mathcal{R}_i} e^{\beta Z_j(T_i)}} \Big\}^{\delta_i},$$

where δ_i and T_i are the indicator of noncensorship and the survival time for the ith individual respectively, and $\mathcal{R}_i = \{j : T_j \wedge C_j > T_i\}$ is the "risk set" consisting of all individuals which are observed to be at risk at time T_i. Let $\hat{\beta}$ be the value that maximizes $L(\beta)$. In terms of the underlying counting processes, the estimate $\hat{\beta}$ is the unique solution to $\frac{\partial}{\partial \beta} \log L(\beta) = U(\beta, 1) = 0$, where

$$U(\beta, t) = \sum_{i=1}^{n} \int_0^t \Big\{ Z_i(u) - \frac{S^{(1)}(\beta, u)}{S^{(0)}(\beta, u)} \Big\} dN_i(u), \tag{3.1}$$

$$S^{(j)}(\beta, t) = \frac{1}{n} \sum_{i=1}^{n} Z_i(t)^j Y_i(t)\, e^{\beta Z_i(t)}, \tag{3.2}$$

for $j = 0, 1, 2$, where $0^0 \equiv 1$. The following theorem gives the asymptotic distribution of $\hat{\beta}$. Define

$$s^{(j)}(\beta, t) = E\, S^{(j)}(\beta, t), \quad e = s^{(1)}/s^{(0)}, \quad v = s^{(2)}/s^{(0)} - e^2 \text{ and } \Sigma = \int_0^1 v(\beta_0, t)\, s^{(0)}(\beta_0, t)\, \lambda_0(t)\, dt.$$

<u>Theorem 3.1</u>. Suppose that λ_0 is integrable over $[0, 1]$, $s^{(0)}(\cdot, \cdot)$ is bounded away from 0 in a neighborhood of β_0, and $\Sigma > 0$. Then $\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{\mathcal{D}} N\,(0, \Sigma^{-1})$.

<u>Proof</u>. (Sketch) By the mean value theorem

$$U(\hat{\beta}, 1) - U(\beta_0, 1) = -\mathcal{I}(\beta^*, 1)\,(\hat{\beta} - \beta_0), \tag{3.3}$$

where β^* lies between β_0 and $\hat{\beta}$, and

$$\mathcal{I}(\beta, t) = -\frac{\partial}{\partial \beta} U(\beta, t) = \int_0^t \Big\{ \frac{S^{(2)}(\beta, u)}{S^{(0)}(\beta, u)} - \Big(\frac{S^{(1)}(\beta, u)}{S^{(0)}(\beta, u)} \Big)^2 \Big\} dN^{(n)}(u),$$

where $N^{(n)} = \sum_{i=1}^{n} N_i$. But $U(\hat{\beta}, 1) = 0$, so from (3.3) we obtain

$$\sqrt{n}(\hat{\beta} - \beta_0) = \frac{n^{-\frac{1}{2}} U(\beta_0, 1)}{n^{-1} \mathcal{I}(\beta^*, 1)}. \tag{3.4}$$

The key step in the proof is to see that $U(\beta_0, \cdot)$ is a martingale:

$$U(\beta_0,t)=\sum_{i=1}^{n}\int_0^t\Big\{Z_i(u)-\frac{S^{(1)}(\beta_0,u)}{S^{(0)}(\beta_0,u)}\Big\}dM_i(u).$$

Let m_1 be a continuous Gaussian martingale with variation process

$$\langle m_1\rangle_t=\int_0^t v(\beta_0,u)\,s^{(0)}(\beta_0,u)\lambda_0(u)\,du. \tag{3.5}$$

Then

$$\langle n^{-\frac{1}{2}}U(\beta_0,\cdot)\rangle_t=\int_0^t\Big\{S^{(2)}(\beta_0,u)-\frac{(S^{(1)}(\beta_0,u))^2}{S^{(0}(\beta_0,u)}\Big\}\lambda_0(u)\,du\xrightarrow{P}\langle m_1\rangle_t,$$

so by Rebolledo's martingale central limit theorem $n^{-\frac{1}{2}}U(\beta_0,\cdot)\xrightarrow{\mathcal{D}}m_1$ in $D[0,1]$. Consequently, $n^{-\frac{1}{2}}U(\beta_0,1)\xrightarrow{\mathcal{D}}N\,(0,\Sigma)$ and from (3.4), to complete the proof it suffices to show that $n^{-1}\mathcal{I}(\beta^*,1)\xrightarrow{P}\Sigma$. This is done in Andersen and Gill (1982, p.1108), but it is to be expected because they show that $\hat\beta\xrightarrow{P}\beta_0$, so $\beta^*\xrightarrow{P}\beta_0$, and we can write $n^{-1}\mathcal{I}(\beta_0,1)$ in the form

$$\int_0^1\Big\{S^{(2)}(\beta_0,u)-\frac{(S^{(1)}(\beta_0,u))^2}{S^{(0)}(\beta_0,u)}\Big\}\lambda_0(u)\,du+\frac{1}{n}\int_0^1\Big\{\frac{S^{(2)}(\beta_0,u)}{S^{(0)}(\beta_0,u)}-\Big(\frac{(S^{(1)}(\beta_0,u))^2}{S^{(0)}(\beta_0,u)}\Big)^2\Big\}dM^{(n)}(u),$$

where $M^{(n)}=\sum_{i=1}^n M_i$. The first term above tends in probabilty to Σ and the second term tends in probabilty to zero, by Lenglart's inequality.

Along the lines of Breslow (1972, 1974), the cumulative baseline hazard function $\Lambda_0(t)=\int_0^t\lambda_0(s)\,ds$ can be estimated by the Nelson–Aalen type estimator

$$\hat\Lambda_0(t)=\int_0^t\frac{dN^{(n)}(u)}{nS^{(0)}(\hat\beta,u)}.$$

The following result, implicit in Andersen and Gill (1982), gives the asymptotic distribution of $\hat\Lambda$.

Theorem 3.2. Under the conditions of Theorem 3.1, $\sqrt{n}(\hat\Lambda_0-\Lambda_0)\xrightarrow{\mathcal{D}}m_0(\cdot)+\psi(\cdot)\,m_1(1)$ in $D[0,1]$, where m_0 and m_1 are independent zero mean Gaussian martingales, $\langle m_1\rangle_t$ is defined by (3.5),

$$\langle m_0\rangle_t=\int_0^t\frac{\lambda_0(u)}{s^{(0)}(\beta_0,u)}\,du\qquad\text{and}\qquad\psi(t)=\Sigma^{-1}\int_0^t e(\beta_0,u)\,\lambda_0(u)\,du.$$

3.2. Aalen's additive risk model

In some applications, additive risk models are more appropriate than proportional hazards models. However, although parametric additive risk models have been used in survival analysis (especially in epidemiology) for many years (see the references in Breslow, 1986; Muirhead and

Darby, 1987), the nonparametric additive risk model (2) has only been studied recently (Aalen, 1980, 1988; McKeague, 1986, 1988a, 1988b; Huffer and McKeague, 1988; Mau, 1986, 1988).

Let $\alpha = (\alpha_1, ..., \alpha_p)'$ and denote $Y_{ij}(t) = Y_i(t)\, Z_{ij}(t)$, $A(t) = \int_{t_0}^t \alpha(s)\, ds$ for fixed t_0, $0 \le t_0 \le 1$. Aalen (1980) proposed estimators $\hat{A}$ of A of the form $\hat{A}(t) = \int_{t_0}^t Y^-(s)\, dN(s)$, where $Y^-(s)$ is a predictable generalized inverse of the $n \times p$ matrix $Y(s) = (Y_{ij}(s))$. In the case $p = 1$, with $(Y^-(s))_{1j} = (\sum_{k=1}^n Y_{k1}(s))^{-1}$, $i = 1, ..., n$, $\hat{A}$ is the Nelson–Aalen estimator. For $p > 1$, Aalen suggested using $Y^-(s) = (Y'(s)\, Y(s))^{-1} Y'(s)$, where here and in the sequel, for any square matrix (or scalar) D, D^{-1} denotes the inverse of D if D is invertible, the zero matrix otherwise. Aalen observed that this choice of Y^- can be motivated by a formal least squares principle and that the resulting estimator $\tilde{A}(t) = \int_{t_0}^t (Y'(s)\, Y(s))^{-1} Y'(s)\, dN(s)$, referred to as Aalen's least squares estimator, probably gives reasonable but not optimal estimates of A. Huffer and McKeague (1988) proposed using the following generalized inverse of $Y(s)$:

$$Y^-(s) = (Y'(s)\, \hat{W}(s)\, Y(s))^{-1} Y'(s)\, \hat{W}(s), \tag{3.6}$$

where $\hat{W}(t)$ is the $n \times n$ diagonal matrix with ith diagonal entry $\hat{W}_i(t) = (\hat{\lambda}_i(t))^{-1}$, and

$$\hat{\lambda}_i(t) = \sum_{j=1}^p \hat{\alpha}_j(t)\, Y_{ij}(t), \tag{3.7}$$

where $\hat{\alpha}_j$ is a predictable estimator of α_j. The estimator $\hat{\alpha}_j$ is taken to be the jth component of the smoothed least squares estimator

$$\hat{\alpha}(t) = \frac{1}{b_n} \int_0^1 K\Big(\frac{t-s}{b_n}\Big)\, d\tilde{A}(s),$$

where K is a left-continuous bounded kernel function having integral 1, support $[0, 1]$ and $b_n > 0$ is a bandwidth parameter. The choice of generalized inverse (3.6) defines the so called weighted least squares estimator

$$\hat{A}(t) = \int_{t_0}^t (Y'(s)\, \hat{W}(s)\, Y(s))^{-1} Y'(s) \hat{W}(s)\, dN(s). \tag{3.8}$$

In the case of a single covariate the weighted least squares estimator coincides with the Nelson–Aalen estimator.

A heuristic explanation for the choice of weight matrix $\hat{W}(t)$ is as follows. By conditioning on the past $\mathcal{F}_t^{(n)}$ we may interpret the stochastic differential equation $dN(t) = Y(t)\alpha(t)\, dt + dM(t)$, increment by increment, as standard linear regression model with heteroscedastic errors $dM(t)$, where $M = (M_1, ..., M_n)'$. We should choose the weight matrix to be proportional to the inverse of the error covariance matrix $\mathrm{Cov}(dM(t)|\mathcal{F}_t^{(n)})$ = the $n \times n$ diagonal matrix with ith diagonal entry $\lambda_i(t)\, dt$. However, $\lambda_i(t)$ depends on the unknown $\alpha(t)$. Estimating $\lambda_i(t)$ by (3.7), where the estimator $\hat{\alpha}(t)$ only depends on the past (since the kernel K has support $[0, 1]$), leads to $\hat{W}(t)$.

The following result, which gives the asymptotic distribution of $\hat{A}$, is a special case of Theorem 3.2 of McKeague (1988a). Let $L(t)$ and $V(t)$ denote the $p \times p$ matrices with entries $L_{jk}(t) = E\, Y_{1j}(t)\, Y_{1k}(t)$ and $V_{jk}(t) = E\, Y_{1j}(t)\, Y_{1k}(t)\, \lambda_1^{-1}(t)$, respectively, and let $D[t_0, 1]^p$ denote the product of p copies of the Skorohod space $D[t_0, 1]$.

Theorem 3.3. Suppose that $\alpha_1, ..., \alpha_p$, $L(\cdot)$, $V(\cdot)$ are continuous, $L(t)$ and $V(t)$ are nonsingular for all $t \in [0, 1]$, $\lambda(t, Z_t)$ is bounded away from zero, $b_n \to 0$, $nb_n^2 \to \infty$, and the kernel function K has bounded variation. Let $0 < t_0 < 1$. Then, under Aalen's additive risk model, $\sqrt{n}(\hat{A} - A) \xrightarrow{\mathcal{D}} m$ in $D[t_0, 1]^p$, where m is a p-variate continuous Gaussian martingale with mean zero and covariance function

$$\mathrm{Cov}(m_j(t), m_k(t)) = \int_{t_0}^{t} (V^{-1}(s))_{jk}\, ds.$$

3.3. The general nonparametric model

The fully nonparametric model (3) was first studied by Beran (1981). It can be applied successfully only when the sample size is very large and there are a small number of covariates. Inference for this model has been studied further by Doksum and Yandell (1982), Dabrowska (1987a, 1987b), McKeague and Utikal (1987, 1988a) and Cheng (1987). In this section we discuss the main result of McKeague and Utikal (1988a) who introduced an estimator for the "doubly" cumulative hazard function $\mathcal{A}(t, z) = \int_0^z \int_0^t \lambda(s, x)\, ds\, dx$, $(t, z) \in [0, 1]^2$. This estimator turns out to be important in the development of goodness-of-fit tests for specific regression models. For simplicity we shall only consider the case of a single covariate ($p = 1$).

Let $\mathcal{I}_r$, $r = 1, ..., d_n$ be a partition of the unit interval, where $\mathcal{I}_r = [z_{r-1}, z_r)$, $z_r = r/d_n$ and d_n is an increasing sequence of positive integers. Let $N_{ir}(t)$ be the counting process which registers the jumps of $N_i(t)$ when $Z_i(t) \in \mathcal{I}_r$, so that $N_{ir}(t) = \int_0^t I\{Z_i(s) \in \mathcal{I}_r\} dN_i(s)$. Beran (1981) suggested that the cumulative conditional hazard function $\Lambda(t, z) = \int_0^t \lambda(s, z)\, ds$ could be estimated by the Nelson–Aalen type estimator

$$\tilde{\Lambda}(t, z) = \int_0^t \frac{dN_r^{(n)}(s)}{Y_r^{(n)}(s)}, \quad \text{for } z \in \mathcal{I}_r,$$

and that the conditional survival function $S(t|z) = e^{-\Lambda(t,z)}$ could be estimated by the product-limit estimator

$$\widetilde{S}(t|z) = \prod_{s \le t} (1 - \Delta\tilde{\Lambda}(s, z)),$$

where $Y_r^{(n)}(s) = \sum_{i=1}^n I\{Z_i(s) \in \mathcal{I}_r\} Y_i(s)$ and $N_r^{(n)} = \sum_{i=1}^n N_{ir}$. Here d_n should tend to infinity at a suitable rate as $n \to \infty$. Dabrowska (1987a, 1987b) obtained weak convergence results for such estimators in the case of right-censoring and non-time-dependent covariate, using the classical approach of Breslow and Crowley (1974). McKeague and Utikal (1987), using the martingale approach, obtained asymptotic results for $\tilde{\Lambda}$ under general predictable censoring and time-dependent covariates.

McKeague and Utikal (1988a) proposed to estimate $\mathcal{A}$ by

$$\tilde{\mathcal{A}}(t,z) = \int_0^z \tilde{\Lambda}(t,x)\,dx, \tag{3.9}$$

and obtained the following weak convergence result for $\tilde{\mathcal{A}}$. Let $\int_0^t \int_0^z \phi(s,x)\,dW(s,x)$ denote a continuous version of the Wiener integral of a function $\phi \in L^2([0,1]^2,\, ds\,dx)$ with respect to a Brownian sheet W; see Wong and Zakai (1974). Suppose that for each $t \in [0,1]$, the random vector (Z_t, Y_t) is absolutely continuous with respect to the product of Lebesgue measure on [0,1] and counting measure, and denote the corresponding density by $f_{Z(t)\,Y(t)}(z,y)$. Also, assume that $f_{Z(t)\,Y(t)}(z,1)$ is a positive, continuous function of $(t,z) \in [0,1]^2$. Let D_2 denote the extension of Skorohod space $D[0,1]$ to functions on $[0,1]^2$, as defined in Neuhaus (1971).

Theorem 3.4. Suppose that λ is Lipschitz, $d_n^2/n \to \infty$ and $d_n = o(n^\delta)$ for some $\delta \in (\frac{1}{2}, 1)$. Then $\sqrt{n}(\tilde{\mathcal{A}} - \mathcal{A}) \xrightarrow{\mathcal{D}} m$ in D_2 as $n \to \infty$, where $m = (m(t,z),\ (t,z) \in [0,1]^2)$ is given by

$$m(t,z) = \int_0^t \int_0^z \sqrt{h(s,x)}\,dW(s,x),$$
$$h(s,x) = \frac{\lambda(s,x)}{f_{Z(s)\,Y(s)}(x,1)}.$$

Proof. (Sketch) It can be shown easily that $\sqrt{n}(\tilde{\mathcal{A}} - \mathcal{A})$ is asymptotically equivalent in distribution to $\widetilde{M}^{(n)}$, where

$$\widetilde{M}^{(n)}(t,z) = \frac{\sqrt{n}}{d_n} \sum_{r=1}^{[zd_n]} \int_0^t \frac{dM_r^{(n)}(s)}{Y_r^{(n)}(s)}, \tag{3.10}$$
$$M_r^{(n)}(t) = \sum_{i=1}^n \int_0^t I\{Z_i(s) \in \mathcal{I}_r\} Y_i(s)\,dM_i(s), \quad r = 1, ..., d_n.$$

Since $\widetilde{M}^{(n)}(\cdot, z)$ is a martingale for each fixed z, Rebolledo's martingale central limit theorem can be used to show that the finite dimensional distributions of $\widetilde{M}^{(n)}$ converge to those of m (cf. the proof of Theorem 2.3). Finally, $\{\widetilde{M}^{(n)},\ n \geq 1\}$ is shown to be tight in D_2 by checking the moment conditions of Bickel and Wichura (1971).

3.4. The general proportional hazards model

Tibshirani (1984) and Hastie and Tibshirani (1986) considered a local partial likelihood technique for estimating the log relative risk function $\eta(z) = \log r(z)$ in the general proportional hazards model (4) with $p = 1$. O'Sullivan (1986a, 1986b) studied a penalized partial likelihood estimator for η and established consistency of that estimator.

McKeague and Utikal (1988a) considered estimating the cumulative relative risk function

$R(z) = \int_0^z r(x)\,dx$ by $\hat{R}(z) = \tilde{\mathcal{A}}(1, z)$, where $\tilde{\mathcal{A}}$ is defined by (3.9). By Theorem 3.4 and the continuous mapping theorem (Billingsley, 1968) we obtain that $\sqrt{n}(\hat{R} - R) \xrightarrow{\mathcal{D}} m^R$, where m^R is a continuous Gaussian martingale with covariance function

$$\mathrm{Cov}(m^R(z_1), m^R(z_2)) = \int_0^{z_1 \wedge z_2} \int_0^1 h(s, x)\,ds\,dx,$$

provided that Λ_0 is constrained to satisfy $\Lambda_0(1) = 1$ (to ensure identifiability). Similarly, the cumulative baseline hazard function Λ_0 can be estimated by $\hat{\Lambda}(t) = \tilde{\mathcal{A}}(t, 1)$. If R is constrained to satisfy $R(1) = 1$, then $\sqrt{n}(\hat{\Lambda} - \Lambda_0) \xrightarrow{\mathcal{D}} m^0$, where m^0 is a continuous Gaussian martingale with covariance function

$$\mathrm{Cov}(m^0(t_1), m^0(t_2)) = \int_0^{t_1 \wedge t_2} \int_0^1 h(s, x)\,dx\,ds.$$

3.5. Goodness-of-fit tests

There is an extensive literature on goodness-of-fit tests for Cox's proportional hazards regression model, see the references in Arjas (1988). Recently, McKeague and Utikal (1988a, 1988b) have developed consistent goodness-of-fit tests for Cox's model, Aalen's additive risk model and the general proportional hazards model against the alternative of the general nonparametric model (3).

Consider testing the null hypothesis H_0: Cox's proportional hazards model (1) holds over the region $(t, z) \in [0, 1]^2$. Under H_0, the natural estimator of $\mathcal{A}$ is

$$\hat{\mathcal{A}}(t, z) = \hat{\Lambda}(t) \int_0^z e^{\hat{\beta} x} dx,$$

where $\hat{\beta}$ and $\hat{\Lambda}$ are defined in Section 3.1 and, if $(T_i, Z_i(T_i))$ falls outside $[0, 1]^2$, the survival time T_i is regarded as being censored (i.e. δ_i is set to 0). Define $S^{(j)}(\beta, t) = n^{-1} \sum_{i=1}^n Z_i(t)^j Y_j(t)\, I(0 \leq Z_i(t) \leq 1)\, e^{\beta Z_i(t)}$, and define the quantities $s^{(j)}$, Σ etc. of Section 3.1 in terms of this $S^{(j)}$.

<u>Theorem 3.5</u>. (McKeague and Utikal, 1988b) Suppose that the conditions of Theorems 3.1 and 3.4 hold. Then, under H_0, $\sqrt{n}(\tilde{\mathcal{A}} - \hat{\mathcal{A}}) \xrightarrow{\mathcal{D}} m'$ in D_2, where

$$\begin{aligned}
m'(t, z) &= \int_0^t \int_0^z \sqrt{h(u, x)}\,dW(u, x) - b(z) \int_0^t \int_0^1 \frac{\sqrt{g(u, x)}}{s^{(0)}(\beta_0, u)}\,dW(u, x) \\
&\quad - c(t, z) \int_0^1 \int_0^1 \Big\{x - \frac{s^{(1)}(\beta_0, u)}{s^{(0)}(\beta_0, u)}\Big\} \sqrt{g(u, x)}\,dW(u, x), \\
h(u, x) &= \frac{\lambda_0(u)\, e^{\beta_0 x}}{f_{Z(u)\,Y(u)}(x, 1)}, \\
g(u, x) &= \lambda_0(u)\, e^{\beta_0 x} f_{Z(u)\,Y(u)}(x, 1), \\
b(z) &= \int_0^z e^{\beta_0 x} dx, \\
c(t, z) &= \Sigma^{-1} \big(\Lambda_0(t) \int_0^z x e^{\beta_0 x} dx - b(z) \int_0^t e(\beta_0, u)\, \lambda_0(u)\, du\big).
\end{aligned}$$

Proof. (Sketch) Using a Taylor series expansion of $e^{\beta z}$ about $\beta = \beta_0$ and the representation of $\sqrt{n}(\hat{\Lambda} - \Lambda_0)$ given by Andersen and Gill (1982, (2.8)), it can be shown easily that $\sqrt{n}(\tilde{\mathcal{A}} - \hat{\mathcal{A}})$ is asymptotically equivalent in distribution to the process $\widetilde{M}^{(n)}(t,z) - b(z)\,\widehat{M}_0(t) - c(t,z)\,\widehat{M}_1(1)$, where $\widetilde{M}^{(n)}$ is defined by (3.10), $\widehat{M}_1(t) = n^{-\frac{1}{2}}U(\beta_0, t)$ and $\widehat{M}_0(t)$ is the martingale part of $\sqrt{n}(\hat{\Lambda} - \Lambda_0)$ given by

$$\widehat{M}_0(t) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^t \frac{1}{S^{(0)}(u)} I(0 \le Z_i(u) \le 1)\, dM_i(u).$$

Using Rebolledo's martingale central limit theorem (see the proof of Theorem 3.4 of Andersen and Gill, 1982) it can be shown that $(\widehat{M}_0, \widehat{M}_1) \xrightarrow{\mathcal{D}} (m_0, m_1)$ jointly in $D[0,1]^2$, where m_0 and m_1 are the independent Gaussian martingales defined in Theorem 3.2. The key step in the proof is to see that (m_0, m_1) can be represented in terms of a single Brownian sheet process W:

$$m_0(t) = \int_0^t \int_0^1 \frac{\sqrt{g(u,x)}}{s^{(0)}(\beta_0,u)}\, dW(u,x),$$

$$m_1(t) = \int_0^t \int_0^1 \left\{ x - \frac{s^{(1)}(\beta_0,u)}{s^{(0)}(\beta_0,u)} \right\} \sqrt{g(u,x)}\, dW(u,x).$$

Then, using Rebolledo's martingale central limit theorem again, and also using Theorem 3.4, it can be shown that $(\widetilde{M}, \widehat{M}_0, \widehat{M}_1) \xrightarrow{\mathcal{D}} (m, m_0, m_1)$ jointly in $D_2 \times D[0,1]^2$. Applying the continuous mapping theorem, we obtain $\sqrt{n}(\tilde{\mathcal{A}} - \hat{\mathcal{A}}) \xrightarrow{\mathcal{D}} m - bm_0 - cm_1(1) = m'$. This completes the proof.

In order to test H_0 against the alternative that λ has the general form of Section 3.3 we might consider using statistics of Kolmogorov–Smirnov type or Cramér–von Mises type:

$$\sqrt{n} \sup_{(t,z)\in[0,1]^2} |\tilde{\mathcal{A}}(t,z) - \hat{\mathcal{A}}(t,z)| \qquad \text{or} \qquad \sqrt{n} \int_0^1 \int_0^1 (\tilde{\mathcal{A}}(t,z) - \hat{\mathcal{A}}(t,z))^2\, dt\, dz$$

which have asymptotic distributions $\sup_{(t,z)\in[0,1]^2} |m'(t,z)|$ and $\int_0^1 \int_0^1 (m'(t,z))^2\, dt\, dz$ respectively. However, general tables for these distributions are not available. The distribution of m' would need to be transformed a standard form (such as that of a Brownian sheet) in order for Theorem 3.5 to be useful. Instead, McKeague and Utikal (1988b) (following Schoenfeld, 1980) have used Theorem 3.5 to derive the following chi-squared test based on a partition of the product of the time and covariate state spaces into cells.

Let $0 = t_0 < \cdots < t_R = 1$ and $0 = z_0 < \cdots < z_L = 1$ and denote $\mathcal{T}_r = (t_{r-1}, t_r]$ and $\mathcal{Z}_l = (z_{l-1}, z_l]$ so that the cells $\mathcal{J}_{rl} = \mathcal{T}_r \times \mathcal{Z}_l$ partition $[0,1]^2$. The increment of $X \equiv \sqrt{n}(\tilde{\mathcal{A}} - \hat{\mathcal{A}})$ over $\mathcal{J}_{rl}$ is given by $Q_{rl}^{(n)} = X(\mathcal{J}_{rl}) = X(t_r, z_l) - X(t_r, z_{l-1}) - X(t_{r-1}, z_l) + X(t_{r-1}, z_{l-1})$. Under H_0 and the conditions of Theorem 3.1 we have that $Q^{(n)} = (Q_{rl}^{(n)},\ r = 1, \ldots, R;\ l = 1, \ldots, L)$ converges in distribution to the Gaussian random array $Q = (Q_{rl},\ r = 1, \ldots, R;\quad l = 1, \ldots, L)$ with mean zero and covariance

$$\mathrm{Cov}(Q_{rl}, Q_{r'l'}) = H(\mathcal{J}_{rl} \cap \mathcal{J}_{r'l'}) - b(\mathcal{Z}_l)\, b(\mathcal{Z}_{l'}) \int_{\mathcal{T}_r \cap \mathcal{T}_{r'}} \frac{d\Lambda_0(u)}{s^{(0)}(\beta_0,u)} - c(\mathcal{J}_{rl})\, c(\mathcal{J}_{r'l'})\Sigma,$$

where $b(\mathcal{Z}_l)$ and $c(\mathcal{J}_{rl})$ denote increments of b and c. A consistent estimator for this covariance can be obtained by inserting the usual estimates of β_0, Λ_0, $s^{(0)}$, Σ and $e(\beta_0, \cdot)$ in the last two terms above and estimating the first term by $\hat{H}(\mathcal{J}_{rl})$, where

$$\hat{H}(t, z) = \frac{n}{d_n^2} \sum_{r=1}^{[zd_n]} e^{\hat{\beta} x_r} \int_0^t \frac{d\hat{\Lambda}_0(s)}{Y_r^{(n)}(s)}.$$

If we write $Q^{(n)}$ and Q in the form of column vectors $U^{(n)}$ and U, respectively, (by stacking columns one on top of each other, say) and let $\hat{C}^{(n)}$ denote the corresponding estimate of the covariance matrix C of U, then our test statistic is given by

$$\hat{\Gamma}^{(n)} = U^{(n)\prime} \hat{C}^{(n)-1} U^{(n)}.$$

Under H_0 and the conditions of Theorem 3.5, provided C is of full rank, we obtain that $\hat{\Gamma}^{(n)}$ has a limiting χ_q^2 distribution, where $q = RL$.

References

Aalen, O. O. (1975). Statistical inference for a family of counting processes, Ph.D. thesis, Berkeley; reprinted (1976) at Institute of Mathematical Statistics, University of Copenhagen.

Aalen, O. O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6** 701–726.

Aalen, O. O. (1980). A model for nonparametric regression analysis of counting processes. *Lecture Notes in Statistics* **2**. Springer-Verlag, New York.

Aalen O. O. (1988). A linear regression model for the analysis of life times. Tech. Report, Section of Medical Statistics, University of Oslo.

Aalen, O. O. and Johansen, S. (1978). An empirical transition matrix for non-homogen -eous Markov chains based on censored observations. *Scand. J. Statist.* **5** 141–150.

Akritas, M. G. (1986). Bootstrapping the Kaplan-Meier estimator. *J. Amer. Stat. Assoc.* **81** 1032–1038.

Andersen, P. K., Borgan, Ø. (1985). Counting process models for life history data: a review. *Scand. J. Statist.* **12** 97–158.

Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1982). Linear nonparametric tests for comparison of counting processes, with applications to censored survival data (with discussion). *Int. Statist. Rev.* **50** 219–258. Correction **52** p.225.

Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1988). Censoring, truncation and filtering in statistical models based on counting processes. *Contemporary Math.* **80** 19–59.

Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1989). *Statistical Models for Counting Processes*, (in preparation).

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10** 1100–1120.

Antoniadis, A. (1987). A penalty method for nonparametric estimation of the intensity function of a counting process. Tech. Report, Dept. of Mathematics, University of California, Irvine.

Arjas, E. (1985). Contribution to the discussion on the paper by Andersen and Borgan. *Scand. J. Statist.* **12** 150–153.

Arjas, E. (1988). A graphical method for assessing goodness-of-fit in Cox's proportional hazards model. *J. Amer. Statist. Assoc.* **83** 204–212.

Beran, R. (1981). Nonparametric regression with randomly censored survival data. Tech. Report, Dept. of Statistics, University of California, Berkeley.

Bickel, P. J. and Wichura, M. J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Statist.* **42** 1656–1670.

Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

Borgan, Ø. and Ramlau-Hansen, H. (1985). Demographic incidence rates and estimation of intensities with incomplete information. *Ann. Statist.* **13** 564–582.

Breslow, N. (1972). Contribution to the discussion of the paper by D. R. Cox. *J. Roy. Statist. Soc. B.* **34** 187–220.

Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics* **30** 89–99.

Breslow, N. E. (1986). Cohort analysis in epidemiology, in A. C. Atkinson and S. E. Fienberg, eds., *A Celebration of Statistics: the ISI Centenary Volume*. Springer-Verlag, New York, 109–143.

Breslow, N. and Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *Ann. Statist.* **2** 437–453.

Chang, M. N. and Yang, G. L. (1987). Strong consistency of a nonparametric estimator of the survival function with doubly censored data. *Ann. Statist.* **15** 1536–1547.

Cheng, P. E. (1987). Nonparametric estimation of survival curve under dependent censorship. Tech. Report, Institute of Statistics, Academia Sinica, Taipei.

Chung, K. L. (1974). *A Course in Probabilty Theory.* Academic Press, New York.

Cox, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. B.* **34** 187–220.

Cox, D. R. (1975). Partial likelihood. *Biometrika* **62** 269–276.

Dabrowska, D. M. (1987a). Nonparametric regression with censored survival time data. *Scand. J. Statist.* **14** 181–197.

Dabrowska, D. M. (1987b). Uniform consistency of nearest neighbour and kernel conditional Kaplan-Meier estimates. Preprint, University of California, Berkeley.

Dabrowska, D. M. (1988). Kaplan–Meier estimate on the plane. *Ann. Statist.* **16** 1475–1489.

Davis, M. H. A. (1983). The martingale theory of point processes and its application to the analysis of failure-time data. Electronic Systems Effectiveness and Life Cycle Costing (ed. by J. K. Skwirzinsky) *Nato ASI series* **F3** 135–155.

Doksum, K. A. and Yandell, B. S. (1982). Properties of regression estimates based on censored survival data. *A Festschrift for E. L. Lehmann*, ed. by P. J. Bickel, K. A. Doksum, and J. L. Hodges, Jr.

Doléans-Dade, C. (1970). Quelques applications de la formule de changement de variables pour les semimartingales. *Z. Wahrsch. Verw. Gebiete* **16** 181–194.

Gill, R. D. (1980). *Censoring and stochastic integrals.* Mathematical Centre Tracts 124. Mathematisch Centrum Amsterdam.

Gill, R. D. (1983). Large sample behavior of the product-limit estimator on the whole line. *Ann. Statist.* **11** 49–58.

Gill, R. D. (1987). Non- and semi-parametric maximum likelihood estimators and the von Mises method (part 1). CWI Report MS-R8709, Centrum voor Wiskunde en Informatica, Amsterdam. To appear in *Scand. J. Statist.*.

Gill, R. D. and Johansen, S. (1987). Product integrals and counting processes. CWI Report MS-R8707, Centrum voor Wiskunde en Informatica, Amsterdam. To appear in *Ann. Statist.*.

Hastie, T. J. and Tibshirani, R. J. (1986). Generalized additive models (with discussion). *Stat. Sci.* **1** 297–310.

Horváth, L. and Yandell, B. S. (1987). Convergence rates for the bootstrapped product-limit process. *Ann. Statist.* **15** 1155–1173.

Huffer, F. W. and McKeague, I. W. (1988). Survival analysis using additive risk models. Tech. Report, Dept. of Statistics, Florida State University, Tallahassee.

Jacobsen, M. (1982). *Statistical Analysis of Counting Processes. Lecture Notes in Statistics* **12**. Springer-Verlag, New York.

Jacobsen, M. (1986). Right censoring and the Kaplan-Meier and Nelson-Aalen estimators. Preprint, Institute of Mathematical Statistics, University of Copenhagen.

Jacobsen, M. (1988). Right censoring and the Kaplan-Meier and Nelson-Aalen estimators. Summary of results. *Contemporary Mathematics* **80** 61–65.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations, *J. Amer. Statist. Assoc.* **53** 457–481.

Karr, A. F. (1986). *Point Processes and their Statistical Inference*, Dekker, New York.

Karr, A. F. (1987). Maximum likelihood estimation in the multiplicative intensity model via sieves. *Ann. Statist.* **15** 473–490.

Keiding, N. and Gill, R. (1988). Random truncation models and Markov processes. Preprint, Statistical Research Unit, University of Copenhagen.

Kopp, P. E. (1984). *Martingales and Stochastic Integrals.* Cambridge University Press, Cambridge.

Lenglart, E. (1977). Relation de domination entre deux processus. *Ann. Inst. H. Poincare* **13** 171–179.

Liptser, R. S. and Shiryayev, A. N. (1978). *Statistics of Random Processes II: Applications.* Springer-Verlag, New York.

Lo, S.-H. and Singh, K. (1986). The product-limit estimator and the bootstrap: Some asymptotic representations. *Probab. Theory Related Fields* **71** 455–465.

Mau, J. (1986). On a graphical method for the detection of time-dependent effects of covariates in survival data. *Appl. Statist.* **35** 245–255.

Mau. J. (1988). A comparison of counting process models for complicated life histories. *Appl. Stoch. Models Data Anal.* **4**.

McKeague, I. W. (1986). Estimation for a semimartingale regression model using the method of sieves. *Ann. Statist.* **14** 579–589.

McKeague, I. W. (1988a). Asymptotic theory for weighted least squares estimators in Aalen's additive risk model. *Contemporary Math.* **80** 139–152.

McKeague, I. W. (1988b). A counting process approach to the regression analysis of grouped survival data. *Stoch. Process. Appl.* **28** 221–239.

McKeague, I. W. and Utikal, K. J. (1987). Inference for a nonlinear semimartingale regression model. Tech. Report, Dept. of Statistics, Florida State University, Tallahassee.

McKeague, I. W. and Utikal, K. J. (1988a). Identifying nonlinear covariate effects in semimartingale regression models. Tech. Report, Dept. of Statistics, Florida State University, Tallahassee.

McKeague, I. W. and Utikal, K. J. (1988b). Goodness-of-fit tests for additive hazards and proportional hazards models. Tech. Report, Dept. of Statistics, Florida State University, Tallahassee.

Muirhead, C. R. and Darby S. C. (1987). Modelling the relative and absolute risks of radiation-induced cancers. *J. R. Statist. Soc. A* **150** 83–118.

Nelson, W. (1969). Hazard plotting for incomplete failure data, *J. Qual. Tech.* **1** 27–52.

Neuhaus, G. (1971). On weak convergence of stochastic processes with multidimensional time parameter. *Ann. Math. Statist.* **42** 1285–1295.

O'Sullivan, F. (1986a). Nonparametric estimation in the Cox proportional hazards model. Tech. Report, Dept. of Statistics, University of California, Berkeley.

O'Sullivan, F. (1986b). Relative risk estimation. Tech. Report, Dept. of Statistics, University of California, Berkeley.

Pollard, D. (1988). Personal communication.

Prentice, R. L. and Self, S. G. (1983). Asymptotic distribution theory for Cox-type regression models with general relative risk form. *Ann. Statist.* **11** 804–813.

Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions. *Ann. Statist.* **11** 453–466.

Rao, B. L. S. Prakasa (1987). *Asymptotic Theory of Statistical Inference.* Wiley.

Rao, R. Ranga (1963). The law of large numbers for $D[0,1]$-valued random variables. *Theor. Probab. Appl.* **8** 70–74.

Rebolledo, R. (1980). Central limit theorems for local martingales. *Z. Wahrsch. verw. Gebiete* **51** 269–286.

Schoenfeld, D. (1980). Chi-squared goodness-of-fit tests for the proportional hazards regression model. *Biometrika* **67** 145–53.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

Thomas, D. C. (1983). Non-parametric estimation and tests of fit for dose response relations. *Biometrics* **39** 263–268.

Tibshirani, R. J. (1984). Local likelihood estimation. Tech. Report and unpublished Ph.D. dissertation, Dept. of Statistics, Stanford University.

Tsai, W.-Y., Leurgans, S. and Crowley, J. (1986). Nonparametric estimation of a bivariate survival function in the presence of censoring. *Ann. Statist.* **14** 1351–1365.

Turnbull, B. W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *J. Amer. Statist. Assoc.* **69** 169–173.

Vardi, Y. (1985). Empirical distributions in selection bias models. *Ann. Statist.* **13** 178–203.

Wellner, J. A. (1985). A heavy censoring limit theorem for the product limit estimator. *Ann. Statist.* **13** 150–162.

Woodroofe, M. (1985). Estimating a distribution function with truncated data. *Ann. Statist.* **13** 163–177.

Wong, E. and Zakai, M. (1974). Martingales and stochastic integrals for processes with a multidimensional parameter. *Z. Wahrsch. verw. Gebiete* **51** 109–122.

Zucker, D. M. (1986). Survival data regression analysis with time-dependent covariate effects. Unpublished Ph.D. Dissertation, Johns Hopkins University.

Zucker, D. M. and Karr, A. F. (1987). Survival data regression analysis with time-dependent covariate effects: a penalized partial likelihood approach. Preprint, Dept. of Mathematical Sciences, Johns Hopkins University. To appear in *Ann. Statist.*.