Sankhyā : The Indian Journal of Statistics
1999, Volume 61, Series A, Pt. 2, pp. 174-188

PARAMETRIC ESTIMATION OF HAZARD FUNCTIONS WITH STOCHASTIC COVARIATE PROCESSES

By SIMEON M. BERMAN
Courant Institute of Mathematical Sciences, New York
and
HALINA FRYDMAN
Stern School of Business, New York

SUMMARY. Let $X(t),\ t \geq 0$, be a real or vector valued stochastic process and T a random killing-time of the process which generally depends on the sample function. In the context of survival analysis, T represents the time to a prescribed event (e.g. system failure, time of disease symptom, etc.) and $X(t)$ is a stochastic covariate process, observed up to time T. The conditional distribution of T, given $X(t),\ t \geq 0$, is assumed to be of a known functional form with an unknown vector parameter θ; however, the distributions of $X(\cdot)$ are not specified. For an arbitrary fixed $\alpha > 0$ the observable data from a single realization of T and $X(\cdot)$ is $\min(T, \alpha)$, $X(t),\ 0 \leq t \leq \min(T, \alpha)$. For $n \geq 1$ the maximum likelihood estimator of θ is based on n independent copies of the observable data. It is shown that solutions of the likelihood equation are consistent and asymptotically normal and efficient under specified regularity conditions on the hazard function associated with the conditional distribution of T. The Fisher information matrix is represented in terms of the hazard function. The form of the hazard function is very general, and is not restricted to the commonly considered cases where it depends on $X(\cdot)$ only through the present point $X(t)$. Furthermore, the process $X(\cdot)$ is a general, not necessarily Markovian process.

1. **Introduction**

Let (Ω, P) be a probability space, and let the following family of random variables be defined on it: a real or vector valued stochastic process $X(t),\ t \geq 0$, and a nonnegative real random variable T with the property

$$P(T > t \big| X(s),\, s \geq 0) = P(T > t \big| X(s),\, 0 \leq s \leq t) \qquad \ldots (1.1)$$

for all $t > 0$. Such a random variable is called a "killing-time" of the process. The basic condition assumed here on the probability measure P is:

Paper received. March 1998; revised October 1998.
AMS (1991) *subject classifications.* 62F12, 62M09, 62P10.
Key words and phrases. Asymptotic distribution of estimators, hazard function, killing-time, likelihood function, stochastic marker process.

For some integer $m \geq 1$, there is an unknown parameter $\theta \in R^m$ such that the conditional probability measure in (1.1) depends on θ: $P_\theta(T > t \big| X(s),\, s \geq 0)$; and the restriction of P to the sigma-field generated by $X(s),\ s \geq 0$, does not depend on θ.

The theme of this work is the estimation of the unknown parameter θ on the basis of observations of T and the values of $X(s)$, for $s \leq T$. The latter restriction on the values of $X(\cdot)$ is a consequence of the role of T as a time at which the process is killed, so that it is not observable after that time. More precisely, we prescribe a constant $\alpha,\ 0 < \alpha \leq \infty$, such that a single observation of the process is defined as

$$\min(T,\, \alpha), \quad X(s), \quad 0 \leq s \leq \min(T,\, \alpha). \qquad \ldots (1.2)$$

Let $(T_i,\, X_i(s),\, s \geq 0),\ i = 1, \ldots, n$, be n independent copies of $(T,\, X(s),\, s \geq 0)$; then, the set of independent copies of the subprocess (1.2) is the observation set. The introduction of the constant α allows the possibility of a bounded sampling interval.

The object of this investigation is to determine the conditions on the family of measures P such that the maximum likelihood estimator $\hat{\theta}$ of θ is consistent and asymptotically normal and efficient. The results are guided by the fact that the likelihood function of a single observation (1.2) factors into a product of the conditional likelihood of T, given the X-values, and the marginal likelihood of the X-values, where the latter does not depend on θ. Thus the likelihood function is essentially based on the conditional density obtained from (1.1).

We will consider the case where the conditional distribution of T has a smooth density function. We define the conditional hazard function of T as

$$q(t,\, X(\cdot),\, \theta) = -\frac{\frac{d}{dt} P_\theta(T > t \big| X(s),\, s \geq 0)}{P_\theta(T > t \big| X(s),\, s \geq 0)}\ . \qquad \ldots (1.3)$$

Under the assumption (1.1), $q(t,\, X(\cdot),\, \theta)$ depends on $X(\cdot)$ only for $X(s),\ 0 \leq s \leq t$. It follows from the definition of q that the conditional density of T, which is the essential factor of the single-observation likelihood function, is representable as

$$q(t,\, X(\cdot),\, \theta)\, \exp\left(-\int_0^t q(s,\, X(\cdot),\, \theta)\, ds\right). \qquad \ldots (1.4)$$

Therefore, the conditions for the properties of the maximum likelihood estimator are stated in terms of conditions on q.

This work was motivated by the estimation issues arising in the area of survival analysis concerned with the modelling of the hazard of failure as a function of a "marker", a time-varying covariate represented by a stochastic process. Here the "killing-time" is identified as the "survival-time". Knowledge of the hazard function q and the current and past values of the marker process can, in principle, yield survival probabilities for the future.

Marker models have been applied in medicine to study the relationship of survival to the clinically observable variables. Some recent papers on the estimation and other aspects of marker models in the context of studying the progression of AIDS are Fusaro et al (1993), Jewell and Kalbfleisch (1992), Self and Pawitan (1992), and Tsiatis et al (1995). Fusaro et al (1993) consider nonparametric estimation of marker-dependent hazard functions. Jewell and Kalbfleisch (1992) discuss a number of methodological issues associated with marker models in medicine. Self and Pawitan (1992), Tsiatis et al (1995), and Yashin and Manton (1997), consider estimation in marker models when the marker process is observed periodically.

Marker-dependent hazard functions have also emerged recently as an important ingredient of a new class of models in finance for pricing financial instruments subject to the risk of default (see Lando (1996) and the references therein). One of the major issues in this class of models is the specification and estimation of the hazard-rate of default as a function of time-varying state variables, represented by stochastic processes. Here the "killing-time" is identified as the time of default and the stochastic process $X(t)$, $t > 0$ describes the evolution of a state variable known to influence the likelihood of default. For example in Jarrow et al (1997), the hazard-rate of default of a defaultable bond is modelled as a function of its credit rating which is represented by a continuous time Markov chain.

The main result, Theorem 4.1, specifies the conditions on the hazard functions that are sufficient for the asymptotic properties of $\hat{\theta}$, the maximum likelihood estimator of θ. We show that if the hazard function $q(t, x(\cdot), \theta)$, which is, of course, a stochastic process, satisfies specified regularity conditions, then the conditions in the hypothesis of the classical theorem of Cramer (1946, p. 500) are valid, and so, by the latter theorem, $\hat{\theta}$ is consistent and asymptotically normal and efficient. The limiting covariance matrix is also expressed as a functional of q. Our result is related to the general result in Andersen, Borgan, Gill and Keiding (1992, p. 420–426) on maximum likelihood estimation of parametrically specified intensity processes. While it may be possible to establish the asymptotic properties of $\hat{\theta}$ in our case by showing that our set of sufficient conditions implies the set of conditions given in Andersen et al (1992, Condition VI.1.1, p. 420), we chose instead the simpler and more natural route of directly verifying Cramer's conditions. Our results employ the classical central limit theorem, the law of large numbers, and a simple stochastic integral with respect to a process with conditionally orthogonal increments.

Section 2 contains an analysis of the likelihood function of a single realization (1.2). The logarithmic derivatives with respect to the parameters (θ_j) are represented as stochastic integrals with respect to a process with conditionally orthogonal increments.

Section 3 restates Cramer's theorem in a form that is more directly applicable to our framework. Section 4 contains the main result, specifying the conditions on the hazard function that are sufficient for the asymptotic properties of $\hat{\theta}$.

Section 5 includes several new examples illustrating the application of Theorem 4.1. In the examples it is assumed that the hazard function depends on $X(s),\ 0 \le s \le t$ only through $X(t)$. Under additional assumption that $X(t)$ is a Markov process the limiting covariance matrix is explicitly calculated in several examples using the results on the joint distribution of $(X(T),\, T)$ obtained in Berman and Frydman (1996).

Statistical inference for $\hat{\theta}$ when a single observation of the process consists of a subset of (1.2), including the case when a marker process is observed only periodically, will be considered in a forthcoming paper.

2. **Likelihood Function**

Let T be a killing-time for a process X such that the conditional distribution of T given $X(\cdot)$ depends on an unknown parameter θ. Suppose that the conditional density,

$$p(t,\, X(\cdot),\, \theta) = -\frac{d}{dt} P_\theta(T > t \big| X(s),\, s \ge 0) \qquad \ldots (2.1)$$

exists for each $t > 0$, almost surely. If the finite-dimensional distributions of $X(\cdot)$ have densities, then, for any finite set $t_0, t_1, \ldots t_k$ of positive numbers, the joint density of $(T,\, X(t_j),\, j = 0, 1, \ldots, k)$ at a point $(t,\, x_j,\, j = 0, 1, \ldots, k)$ is of the form

$$f(t, x_0, x_1, \ldots, x_k; \theta) = E\left[p(t, X(\cdot), \theta) \big| X(t_j) = x_j, j = 0, 1, \ldots, k\right] h(x_0, \ldots, x_k), \qquad \ldots (2.2)$$

where $h(x_0, \ldots, x_k)$ is the joint density of $X(t_j)$ at $x_j,\ j = 0, 1, \ldots k$, and does not depend on θ. The function $f(T, X(t_0), \ldots, X(t_k); \theta)$ is the likelihood function of a single observed realization $(T,\, X(t_0), \ldots, X(t_k))$. The likelihood function for a sample of n independent copies is a product of the corresponding single-realization likelihood functions (2.2). In the method of maximum likelihood, the function h in (2.2) may be ignored because it does not depend on θ. Therefore, it suffices to represent the single-realization likelihood as

$$E\left[p(t,\, X(\cdot),\, \theta) \big| X(t_j), j = 0, 1, \ldots, k\right]. \qquad \ldots (2.3)$$

at $t = T$.

Under the assumption (1.1) and the definition (2.1), it follows that $p(t,\, X(\cdot),\, \theta)$ is measurable with respect to the sigma-field generated by $X(s),\ 0 \le s \le t$; therefore,

$$E(p(t,\, X(\cdot),\, \theta) \big| X(s),\, 0 \le s \le t) = p(t,\, X(\cdot),\, \theta). \qquad \ldots (2.4)$$

For arbitrary $t > 0$, the conditional hazard function (1.3) is

$$q(t, X(\cdot), \theta) = -\frac{d}{dt} \log \int_t^\infty p(s, X(\cdot), \theta)\, ds; \qquad \ldots (2.5)$$

then, by (1.4),

$$\log p(t, X(\cdot), \theta) = \log q(t, X(\cdot), \theta) - \int_0^t q(s, X(\cdot), \theta)\, ds. \qquad \ldots (2.6)$$

LEMMA 2.1. *Define the random functions*

$$N(t) = 1(T \le t), \qquad Y(t) = 1(T > t), \qquad \ldots (2.7)$$

and

$$M(t) = N(t) - \int_0^t q(s, X(\cdot), \theta)\, Y(s)\, ds. \qquad \ldots (2.8)$$

If

$$\int_0^t E\left[q(s, X(\cdot), \theta)\right]^2 ds < \infty, \quad \textit{for all} \quad t > 0, \qquad \ldots (2.9)$$

then $M(t)$, $t \ge 0$, has, under the conditional probability measure $P(\cdot \,|X(s),$ $s \ge 0)$, orthogonal increments, and, with probability 1,

$$E\left[(M(t) - M(s))^2 \,\middle|X(u),\, u \ge 0\right] = \int_s^t p(u, X(\cdot), \theta)\, du \qquad \ldots (2.10)$$

for $0 \le s < t$.

PROOF. By the definition (2.7) of N and Y:

$$E\left[(N(t) - N(s))^2 \,\middle|X(\cdot)\right] = E\left[N(t) - N(s)\middle|X(\cdot)\right]$$
$$= P\left(s < T \le t\middle|X(\cdot)\right) = \int_s^t p(u, X(\cdot), \theta)\, du;$$

and

$$E\left\{N(t) \int_s^t q(u, X(\cdot), \theta)\, Y(u)\, du\middle|X(\cdot)\right\}$$
$$= E\left\{\int_s^t q(u, X(\cdot), \theta)\, 1(u < T \le t)\, du\middle|X(\cdot)\right\}$$
$$= \int_s^t q(u, X(\cdot), \theta) \int_u^t p(v, X(\cdot), \theta)\, dv\, du;$$

and

$$
\begin{aligned}
&E\left[\left(\int_s^t q(u,\, X(\cdot),\, \theta)\, Y(u)\, du\right)^2 \Big| X(\cdot)\right] \\
&= 2E\left\{\int_s^t \left(\int_u^t q(v,\, X(\cdot),\, \theta)\, Y(v)\, dv\right) q(u,\, X(\cdot),\, \theta)\, du \Big| X(\cdot)\right\} \\
&= 2\int_s^t q(u,\, X(\cdot),\, \theta) \int_u^t p(v,\, X(\cdot),\, \theta)\, dv\, du.
\end{aligned}
$$

The expectations are defined by virtue of (2.9). This confirms (2.10). The result $E\,M(t) \equiv 0$ and the orthogonality of the increments follow by similar calculations. □

Suppose, for almost all sample functions $X(\cdot)$, that the last term in (2.6) has partial derivatives with respect to the components θ_j of θ, and that the differentiation may be done under the integral sign; then

$$
\begin{aligned}
\frac{\partial}{\partial\theta_j} \log p(t,\, X(\cdot),\, \theta) \;&=\; \frac{\partial}{\partial\theta_j} \log q(t,\, X(\cdot),\, \theta) \\
&\;-\; \int_0^t \frac{\partial}{\partial\theta_j} \log q(s,\, X(\cdot),\, \theta)\, q(s,\, X(\cdot),\, \theta)\, ds.
\end{aligned}
$$

It follows that for almost all $X(\cdot)$, the score function $(\partial/\partial\theta_j)\, \log p(t,\, X(\cdot),\, \theta)$ at $t = T$ is representable as

$$
\int_0^\infty \frac{\partial}{\partial\theta_j} \log q(s,\, X(\cdot),\, \theta)\, d\,M(s), \qquad \ldots(2.11)
$$

where the stochastic integral with respect to $M(s)$ is well defined, by virtue of Lemma 2.1, under the assumption

$$
\int_0^\infty \left(\frac{\partial}{\partial\theta_j} \log q(s,\, X(\cdot),\, \theta)\right)^2 p(s,\, X(\cdot),\, \theta)\, ds < \infty. \qquad \ldots(2.12)
$$

3. Cramer's Theorem on the Solutions of the Likelihood Equation

Let $\mathbf{Y}$ be a random vector or stochastic process, and let $f(\mathbf{Y},\, \theta)$ be the likelihood function of a single realization of $\mathbf{Y}$. Here f is a known functional of $(\mathbf{Y},\, \theta)$, and $\theta = (\theta_1, \ldots, \theta_m)$ is an unknown real-vector parameter. As is well-known in the theory of maximum likelihood, f is unique except for factors that do not depend on θ. Let θ^0 be the true value of θ, and let B be an arbitrary neighborhood of θ^0. Let $\mathbf{Y}_i,\ i = 1, 2, \ldots$ be independent copies of

$\mathbf{Y}$ and put $L(\theta) = \prod_{i=1}^{n} f(\mathbf{Y}_i, \theta)$, the likelihood function for the sample of n independent realizations. The following result is equivalent to the vector version of the classical theorem of Cramer (1946), page 500, on the consistency and asymptotic normality of maximum likelihood estimators.

LEMMA 3.1. *Suppose that the following conditions hold:* L.3.1. *The partial derivatives of* $\log f$, $D^{(1)} \log f = (\partial/\partial\theta_j) \log f$, $D^{(2)} \log f = (\partial^2/\partial\theta_j\, \partial\theta_h) \log f$, *and* $D^{(3)} \log f = (\partial^3/\partial\theta_j\, \partial\theta_h\, \partial\theta_\ell) \log f$ *exist and are continuous for* $\theta \in B$, *almost surely for* $j, h, \ell = 1, \ldots m$. L.3.2. *We have*

$$E\left(\frac{\partial}{\partial\theta_j} \log f\right)_{\theta=\theta^0} = 0, \quad E\left(\frac{\partial}{\partial\theta_j} \log f\right)^2_{\theta=\theta^0} < \infty, \quad \textit{for } j = 1, \ldots, m\ ; \qquad \ldots (3.1)$$

the matrix $(\sigma_{jh}) = \sum$, *where*

$$\sigma_{jh} = E\left[\frac{\partial}{\partial\theta_j} \log f \frac{\partial}{\partial\theta_h} \log f\right]_{\theta=\theta^0}, \quad j, h = 1, \ldots, m, \qquad \ldots (3.2)$$

is strictly positive definite; and σ_{jh} *is also equal to*

$$-E\left(\frac{\partial^2 \log f}{\partial\theta_j\, \partial\theta_h}\right)_{\theta=\theta^0}, \qquad \ldots (3.3)$$

which is assumed to be finite.

L.3.3. *There exists a real-valued functional* $H(\mathbf{Y})$ *such that* $\sup_{\theta\in B} \left|D^{(3)} \log f\right| \leq H(\mathbf{Y})$, *for all third-order partial derivatives and almost all* $\mathbf{Y}$, *and* $E\, H(\mathbf{Y}) < \infty$.

Then the system of likelihood equations $(\partial/\partial\theta_j) \log L(\theta) = 0$, $1 \leq j \leq m$, *has a solution* $\hat{\theta}$, *depending on* n, *such that* $\hat{\theta} \to \theta^0$ *in probability, for* $n \to \infty$, *and* $\sqrt{n}\left(\hat{\theta} - \theta^0\right) \xrightarrow{d} N\left(0, \sum^{-1}\right)$, *where* $\sum$ *is defined in L.3.2.*

The subject of this paper is the maximum likelihood estimation of the unknown parameter θ on the basis of the set of independent realizations of the subprocess (1.2). As shown in Section 2, the unknown parameter θ enters the likelihood function only through the conditional distribution of T, given the observations on $X(\cdot)$. Hence, by the uniqueness of the likelihood function up to factors that do not depend on θ, the single-realization likelihood function f in the statement of Lemma 3.1 may be taken to be the function (2.3). The logarithmic derivatives are representable in terms of the associated hazard functions q by means of the stochastic integral (2.11). The conditions L.3.1–L.3.3 in the hypothesis of Lemma 3.1 will be shown to be fulfilled under corresponding conditions on the hazard functions.

As in most applications of maximum likelihood, the matrix $\sum = \sum(\theta)$ is generally unknown and has to be estimated from the same data used to estimate θ. If the underlying process $X(t)$ is specified so that $\sum(\theta)$ can be determined by mathematical analysis, then the natural estimator of $\sum$ is $\sum(\hat{\theta})$, where $\hat{\theta}$ is the estimator of θ. $\sum(\hat{\theta})$ is called the parametric estimator. If $X(t)$ is itself not specified or if $\sum$ cannot be sufficiently identified as a function of θ, then it can often be estimated by applying the law of large numbers as will be shown at the end of Section 4. Such an estimator is called nonparametric.

4. **Main Result**

The sample data consists of a set of n independent copies of (1.2) for some $\alpha > 0$. The likelihood function of a single realization (1.2) is obtained by a modification of (1.4) to account for the length α of the observation interval. In view of the two possibilities, $T \leq \alpha$ and $T > \alpha$, we obtain the likelihood function of one realization,

$$\begin{aligned} f(T,\, X(\cdot),\, \theta) \quad &= 1(T \leq \alpha)\, q(T,\, X(\cdot),\, \theta)\, \exp\left(- \int_0^T q(s,\, X(\cdot),\, \theta)\, ds \right) \\ &\quad + 1(T > \alpha)\, \exp\left(- \int_0^\alpha q(s,\, X(\cdot),\, \theta)\, ds \right) \\ &= (q(T,\, X(\cdot),\, \theta))^{1(T \leq \alpha)}\, \exp\left(- \int_0^\alpha 1(T > s)\, q(s,\, X(\cdot),\, \theta)\, ds \right); \end{aligned}$$

hence

$$\log f = 1(T \leq \alpha)\, \log q(T,\, X(\cdot),\, \theta) - \int_0^\alpha 1(T > s)\, q(s,\, X(\cdot),\, \theta)\, ds. \qquad \ldots (4.1)$$

If the derivative of $\int_0^\alpha 1(T > s)\, q(s,\, X(\cdot),\, \theta)\, ds$ with respect to θ_j can be taken inside the integral, then

$$\begin{aligned} \tfrac{\partial}{\partial \theta_j} \log f &= 1(T \leq \alpha)\, \tfrac{\partial}{\partial \theta_j} \log q(T,\, X(\cdot),\, \theta) \\ &\quad - \textstyle\int_0^\alpha 1(T > s)\, q(s,\, X(\cdot),\, \theta)\, \tfrac{\partial}{\partial \theta_j} \log q(s,\, X(\cdot),\, \theta)\, ds. \end{aligned} \qquad \ldots (4.2)$$

The following theorem furnishes conditions on q which are sufficient for the conditions on the function f assumed in Lemma 3.1.

THEOREM 4.1. *Assume the following three sets of conditions:* L.4.1. *Let* $D^{(i)}$, $i = 1, 2, 3$ *be the differential operators defined in L.3.1. Then for almost all realizations* $(T,\, X(\cdot))$ *the functions* $D^{(i)} \log q(T,\, X(\cdot),\, \theta)$

and $D^{(i)} \int_0^\alpha 1(T > s)\, q(s,\, X(\cdot),\, \theta)\, ds,\ i = 1, 2, 3,$ *exist and are continuous for* $\theta \in B$, *and furthermore,*

$$D^{(i)} \int_0^\alpha 1(T > s)\, q(s,\, X(\cdot),\, \theta)\, ds = \int_0^\alpha 1(T > s)\, D^{(i)}\, q(s,\, X(\cdot),\, \theta)\, ds.$$

L.4.2. *For every* $j,\, h = 1, \ldots, m,$ *and* $\theta \in B$.

$$\int_0^\alpha E\left\{\left|\frac{\partial}{\partial\theta_j}\, \log q(s,\, X(\cdot),\, \theta)\right|^2 p(s,\, X(\cdot),\, \theta)\right\} ds < \infty,$$

and

$$\int_0^\alpha E\left\{\left|\frac{\partial^2}{\partial\theta_j\, \partial\theta_h}\, \log q(s,\, X(\cdot),\, \theta)\right|^2 p(s,\, X(\cdot),\, \theta)\right\} ds < \infty;$$

and the following (nonnegative definite) matrix is strictly positive definite at $\theta = \theta^0$:

$$\int_0^\alpha E\left\{\tfrac{\partial}{\partial\theta_j}\, \log q(s,\, X(\cdot),\, \theta) \quad \tfrac{\partial}{\partial\theta_h} \log q(s,\, X(\cdot),\, \theta) \cdot p(s,\, X(\cdot),\, \theta)\right\} ds, \quad j,\, h = 1, \ldots, m. \qquad \ldots (4.3)$$

L.4.3. *There are nonnegative Borel functions* $H_1(t,\, X(\cdot))$ *and* $H_2(t,\, X(\cdot))$ *such that for all third-order derivatives* $D^{(3)}$, *and all* $\theta \in B$, *and all* $0 \le t \le \alpha$, *and almost all sample functions* $X(\cdot)$, *we have the inequalities*

$$\left|D^{(3)} \log q(t,\, X(\cdot),\, \theta)\right| \le H_1(t,\, X(\cdot)),$$
$$\left|D^{(3)}\, q(t,\, X(\cdot),\, \theta)\right| \le H_2(t,\, X(\cdot)),$$

$$\int_0^\alpha E\left\{H_1(t,\, X(\cdot))\, p(t,\, X(\cdot),\, \theta^0)\right\} dt < \infty,$$
$$\int_0^\alpha E\left\{H_2(t,\, X(\cdot)) \int_t^\infty p(s,\, X(\cdot),\, \theta^0)\right\} dt < \infty.$$

Then the conditions L.3.1, L.3.2, and L.3.3 of Lemma 3.1 hold for $f = f(T,\, X(\cdot),\, \theta)$, *and with* $\sum$ *as the matrix with entries (4.3).*

PROOF. It is obvious that L.4.1 implies L.3.1 and that L.4.3 implies L.3.3 in the case where $\log f$ is given by (4.2).

There remains only to prove L.3.2. Under L.4.1, $\frac{\partial}{\partial\theta_j} \log f$ in (4.2) is representable as

$$\int_0^\alpha \frac{\partial}{\partial\theta_j}\, \log q(t,\, X(\cdot),\, \theta)\, dM(t), \qquad \ldots (4.4)$$

where $M(t)$ is given by (2.8), which is analogous to (2.11) except that the upper limit α is inserted in the place of ∞. By the conditional orthogonality of the increments of M and formula (2.10), the expected value of (4.4) is equal to 0, and the variance is equal to the (finite) integral (4.3) with $j = h$. This confirms (3.1).

For the confirmation of the positive definiteness of $\sum$, observe that, by the representation (4.4) of $(\partial/\partial\theta_j)\log f$, the covariance σ_{jh} in (3.2) is equal to

$$E\left\{\int_0^\alpha \frac{\partial}{\partial\theta_j}\log q(t,\, X(\cdot),\, \theta)\, dM(t) \int_0^\alpha \frac{\partial}{\partial\theta_h}\log q(t,\, X(\cdot),\, \theta)\, dM(t)\right\}$$

at $\theta = \theta^0$, which is equal to the integral (4.3). The resulting matrix is, by hypothesis, strictly positive definite.

Finally, we verify (3.3). Recall the general result from elementary calculus for an arbitrary smooth function f of several variables:

$$\frac{\partial^2 f}{\partial\theta_j\,\partial\theta_h} = f\left\{\frac{\partial^2 \log f}{\partial\theta_j\,\partial\theta_h} + \frac{\partial \log f}{\partial\theta_j}\,\frac{\partial \log f}{\partial\theta_h}\right\} \qquad \ldots (4.5)$$

From (4.2) it follows under condition L.4.1 that

$$\frac{\partial^2 \log f}{\partial\theta_j\,\partial\theta_h} = 1(T \le \alpha)\,\frac{\partial^2 \log q(T,\, X(\cdot),\, \theta)}{\partial\theta_j\,\partial\theta_h} - \int_0^\alpha 1(T > s)\,\frac{\partial^2\, q(s,\, X(\cdot),\, \theta)}{\partial\theta_j\,\partial\theta_h}\, ds\ ,$$

which, by (4.5) with q in the place of f, is equal to

$$1(T \le \alpha)\,\frac{\partial^2 \log q(T,\, X(\cdot),\, \theta)}{\partial\theta_j\,\partial\theta_h} - \int_0^\alpha 1(T > s)\, q(s,\, X(\cdot),\, \theta)$$
$$\left\{\frac{\partial^2 \log q(s,\, X(\cdot),\, \theta)}{\partial\theta_j\,\partial\theta_h} + \frac{\partial \log q(s,\, X(\cdot),\, \theta)}{\partial\theta_j}\,\frac{\partial \log q(s,\, X(\cdot),\, \theta}{\partial\theta_h}\right\} ds\ .$$

By the definition (2.8) of $M(t)$, the previous sum is representable as

$$-\quad 1(T \le \alpha)\,\frac{\partial \log q(T,\, X(\cdot),\, \theta)}{\partial\theta_j}\cdot\frac{\partial \log q(T,\, X(\cdot),\, \theta)}{\partial\theta_h}$$
$$+\quad \int_0^\alpha \left\{\frac{\partial^2 \log q(s,\, X(\cdot),\, \theta)}{\partial\theta_j\,\partial\theta_h} + \frac{\partial \log q(s,\, X(\cdot),\, \theta)}{\partial\theta_j}\cdot\frac{\partial \log q(s,\, X(\cdot),\, \theta)}{\partial\theta_h}\right\} dM(s).$$

By Lemma 2.1, it follows that

$$E\left\{\frac{\partial^2 \log f}{\partial\theta_j\,\partial\theta_h}\,\middle|\, X(\cdot)\right\} = -\,E\left\{1(T \le \alpha)\,\frac{\partial \log q(T,\, X(\cdot),\, \theta)}{\partial\theta_j}\,\frac{\partial \log q(T,\, X(\cdot),\, \theta)}{\partial\theta_h}\,\middle|\, X(\cdot)\right\}.$$

By taking expectations over $X(\cdot)$, we see that the preceding expression becomes the entry $(j,\, h)$ of the matrix defined by (4.3). This operation is justified under L.4.2.

□

The element (4.3) of $\sum$ is representable as

$$\sigma_{jh} = E\left\{1(T \le \alpha)\, \frac{\partial}{\partial\theta_j}\, \log q(T,\, X(\cdot),\, \theta)\, \frac{\partial}{\partial\theta_h}\, \log q(T,\, X(\cdot),\, \theta)\right\}. \quad \ldots (4.6)$$

If $(T_i,\, X_i(\cdot),\, i = 1, \ldots, n)$ are the independent copies of $(T,\, X(\cdot))$ used in the estimation procedure, then, by the Law of Large Numbers,

$$\frac{1}{n}\sum_{i=1}^{n} 1(T_i \le \alpha)\, \frac{\partial}{\partial\theta_j}\, \log q(T_i,\, X_i(\cdot),\, \theta)\, \frac{\partial}{\partial\theta_h}\, \log q(T_i,\, X_i(\cdot),\, \theta)$$

is a consistent nonparametric estimator of σ_{jh}. If the distributions of $X(\cdot)$ are known and σ_{jh} can be explicitly calculated as a function $\sigma_{jh}(\theta)$, then, by Theorem 4.1, $\sigma_{jh}(\hat\theta)$ is a consistent parametric estimator of $\sigma_{jh}(\theta)$ if the latter function is continuous.

5. **Examples**

We consider the class of hazard functions q where the dependence on $X(s)$, $0 \le s \le t$, is limited to $X(t)$ alone: There is a real-valued function $q(t,\, x,\, \theta)$ such that $q(t,\, X(\cdot),\, \theta) = q(t,\, X(t),\, \theta)$. Other models, where the dependence of $q(t,\, X(\cdot),\, \theta)$ on $X(\cdot)$ is through some functional of $X(s)$, $0 \le s \le t$, such as $X(t - a)$ $(a > 0)$ or $\int_0^t X(s)\, ds$, are certainly plausible. However, in most applications the dependence of q on $X(s)$, $0 \le s \le t$, is limited to $X(t)$ alone. In particular we first take q to be of the form

$$q(t,\, x,\, \theta) = \sum_{j=1}^{m} \theta_j\, q_j(t,\, x)\, 1(x \in J_j), \quad \ldots (5.1)$$

where $J_j,\ j = 1, \ldots, m$, are disjoint Borel sets forming a decomposition of the real line, (q_j) are positive functions, and $1(\cdot)$ is the indicator function. From the relations

$$\frac{\partial q}{\partial\theta_j} = 1(x \in J_j)\, q_j(t,\, x), \qquad \frac{\partial \log q}{\partial\theta_j} = 1(x \in J_j)/\theta_j\ ,$$

and from (4.2) it follows that the likelihood equation for θ_j has the solution

$$\hat\theta_j = \frac{\sum_{i=1}^{n} 1(X_i(T_i) \in J_j,\ T_i \le \alpha)}{\sum_{i=1}^{n} \int_0^\alpha 1(T_i > s,\, X_i(s) \in J_j)\, q_j(s,\, X_i(s))\, ds}\ , \quad \ldots (5.2)$$

where $(T_i,\, X_i(\cdot)),\ i = 1, \ldots, n$ are the independent observed realizations. From the elementary form of $\partial \log q/\partial\theta_j$ it is simple to formulate general conditions under which Theorem 4.1 can be applied.

The information matrix (σ_{jh}) in (4.6) takes the form

$$\begin{aligned}\sigma_{jh} &= \tfrac{1}{\theta_j^2}\, P(T \le \alpha,\, X(T) \in J_j), \quad \text{for} \quad j = h,\\ &= 0, \quad \text{for} \quad j \ne h.\end{aligned}$$

According to the discussion following (4.6), the nonparametric estimator of σ_{jj} is

$$\frac{1}{n\,\hat{\theta}_j^2} \sum_{i=1}^{n} 1\left(T_i \le \alpha,\, X_i(T_i) \in J_j\right) ,$$

where $\hat{\theta}_j$ is given by (5.2). The parametric estimator is obtained, under the assumption that $P\left(T \le \alpha,\, X(T) \in J_j\right) = P_\theta\left(T \le \alpha,\, X(T) \in J_j\right)$ can be calculated as a function of θ, by using the estimator $\hat{\theta}$ in the place of θ^0 :

$$\hat{\theta}_j^{-2}\, P_{\hat{\theta}}\left(T \le \alpha,\, X(T) \in J_j\right). \qquad \ldots (5.4)$$

If $m = 1$ in (5.1), then $q(t,\, x,\, \theta) = \theta q(t,\, x)$, and (5.2) and (5.3) assume the simpler forms:

$$\hat{\theta} = \frac{\sum_{i=1}^{n} 1(T_i \le \alpha)}{\sum_{i=1}^{n} \int_0^\alpha 1\left(T_i > s\right) q(s,\, X_i(s))\, ds} , \qquad \ldots (5.5)$$

and

$$\sigma_{11} = \frac{1}{\theta^2}\, P(T \le \alpha) . \qquad \ldots (5.6)$$

Next we discuss the calculation of (5.3) and (5.6) in specific cases. In each case it is easy to verify the conditions of Theorem 4.1. In Examples 5.1 and 5.2 we use the results on the joint distribution of $(X(T),\, T)$ obtained in Berman and Frydman (1996). Example 5.3 employs the results from Yashin (1985) and Yashin and Manton (1997) on the formula for the marginal survival function of T. In the final example q has a form different from (5.1).

Example 5.1. Let $X(t),\ t \ge 0$, be a Poisson process with parameter λ. Take $q(t,\, x,\, \theta) = \theta x$; then, as shown by Berman and Frydman (1996),

$$P\left(T \le \alpha \middle| X(0) = i\right) = 1 - \exp\left[-\lambda\alpha - \theta\alpha i + \frac{\lambda}{\theta}\left(1 - e^{-\theta\alpha}\right)\right]$$

and the parametric estimator of $\sum$ is obtained from this by (5.6).

Example 5.2. Suppose that $X(t),\ t \ge 0$, is a continuous-time Markov chain with state space $S = (1, \ldots, m)$, matrix generator A, and the killing-rate function $q(t,\, x,\, \theta) = \sum_{j=1}^{m} \theta_j\, 1(x = j)$. Then, by (5.2),

$$\hat{\theta}_j = \frac{\sum_{i=1}^{n} 1\left(T_i \le \alpha,\, X_i(T_i) = j\right)}{\sum_{i=1}^{n} \int_0^\alpha 1\left(T_i > s,\, X_i(s) = j\right) ds} ,$$

and $\sigma_{jj}(\theta^\theta)$ is given by (5.3). Put $D = \text{diag}\,(\theta_1, \ldots, \theta_m)$; then, by the results of Berman and Frydman (1996), for any $h,\ j \in S$,

$$P\left(T < \alpha,\ X(T) = j \big| X(0) = h\right) = \theta_j \left[\left(I - e^{\alpha(A-D)}\right)(D-A)^{-1}\right]_{hj}$$

so that, for any initial distribution (π_h) on S,

$$P\left(T < \alpha,\ X(T) = j\right) = \theta_j \sum_{h=1}^{m} \pi_h \left[\left(I - e^{\alpha(A-D)}\right)(D-A)^{-1}\right]_{hj} .$$

This model is related to that of Jarrow, Lando and Turnbull (1997).

Example 5.3. Let $X(t),\ t \geq 0$, be a Gaussian diffusion defined by

$$dX(t) = [a_0(t) + a_1(t)\, X(t)]\, dt + \sigma(t)\, dW(t), \qquad \ldots (5.7)$$

where $X(0)$ is a constant, and $a_0(t),\ a_1(t),\ b(t)$ are known functions of time, and define

$$q(t,\, x,\, \theta) = \theta x^2. \qquad \ldots (5.8)$$

Yashin (1985) showed that in this case the distribution of $X(t)$ conditioned on $T > t$ is also Gaussian with the mean and variance functions satisfying a set of ordinary differential equations. This result combined with the general formula in Yashin and Manton (1997) for the marginal survival function:

$$P(T > t) = \exp\left(-\int_0^t E[q(u,\, X(u),\, \theta) | T > u]\, du\right)$$

makes the calculation of $P(T > t)$ tractable in the model given by (5.7) and (5.8). In particular, suppose that $X(t)$ is a Brownian motion with variance parameter σ^2 (i.e. $a_0(t) = a_1(t) \equiv 0,\ \sigma(t) = \sigma$). Then the calculations give

$$P(T > t) = \Big(\cosh 2\sqrt{\theta}\,\sigma\alpha\Big)^{-1/2}$$

so that (5.6) is of the form

$$\theta^{-2}\left[1 - \Big(\cosh 2\sqrt{\theta}\,\sigma\alpha\Big)^{-1/2}\right].$$

This model is related to that of Yashin and Manton (1997).

Example 5.4. Consider the hazard function

$$q(t,\, x,\, \theta) = a + (x - \theta)^2 ,$$

where $a > 0$ is known and θ is a real-valued unknown parameter. (This is not in the special class (5.1).) The likelihood equation based on the independent realizations $(T_i,\, X_i(\cdot))$, $i = 1, \ldots, n$, is, by (4.2), equivalent to

$$\theta = \frac{\sum_{i=1}^{n} \int_0^{T_i \wedge \alpha} X_i(s)\, ds - \sum_{i=1}^{n} \left\{ \frac{X_i(T_i)\, 1(T_i \le \alpha)}{(X_i(T_i)-\theta)^2 + a} \right\}}{\sum_{i=1}^{n} (T_i \wedge \alpha) - \sum_{i=1}^{n} \frac{1(T_i \le \alpha)}{(X_i(T_i)-\theta)^2 + a}},$$

where $T \wedge \alpha = \min(T,\, \alpha)$. The latter is solvable by successive approximation. By (4.6),

$$\sigma_{11} = 4E \left[\frac{(X(T)-\theta)\, 1(T \le \alpha)}{(X(T)-\theta)^2 + a} \right]^2 .$$

This model is related to that of Yashin and Manton (1997).

Remark 5.5. The hazard functions $h_0(t) + \beta X(t)$ and $h_0(t)(1 + \beta X(t))$ were considered by Jewell and Kalbfleisch (1992) and Self and Pawitan (1992), respectively. In their models β is an unknown parameter and $h_0(t)$ an unknown function; thus, these are semi-parametric models. By contrast, our model is purely parametric since the functional form of q is assumed to be given.

Acknowledgements. The work of Halina Frydman was partially supported by summer research grant from the NYU Salomon Center. She also thanks Biostatistics Department at the University of Copenhagen for providing hospitable environment during her work on this paper.

References

Andersen P.K., Borgan O., Gill R. and Keiding N. (1993). *Statistical Models Based on Counting Processes.* Springer, New York.

Berman, S. and Frydman H. (1996). Distributions associated with Markov processes with killing. *Communications in Statistics: Stochastic Models*, **12** 367–388.

Cramer H. (1946). *Mathematical Methods of Statistics.* Princeton University Press, Princeton, New Jersey.

Fusaro E.F., Nielsen J.P, and Sheike T.H. (1993). Marker-dependent hazard estimation: An application to Aids, *Statistics in Medicine*, **12**, 843–865.

Jarrow R., Lando D. and Turnbull S. (1997). A Markov model for the term structure of credit risk spreads. *The Review of Financial Studies*, **10** 481–523.

Jewell N.P. and Kalbfleisch J.D. (1992). Marker models in survival analysis and applications to issues associated with Aids. in *Aids Epidemiology: Methodological Issues*, N. Jewell, K. Dietz, V. Farewell, Editors, 211–230. Birkhäuser, Boston.

Lando D. (1996). Modeling bonds and derivatives with default risk. Working paper, Institute of Mathematical Statistics, University of Copenhagen.

Tsiatis A.A., De Gruttola V. and Wulfsohn M.S. (1995). Modelling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with Aids. *Jour. Amer. Statist. Assoc.*, **90**, 27–37.

Self S. and Pawitan Y. (1992). Modelling a marker of disease progression and onset of disease. In *AIDS Epidemiology: Methodological Issues*, N. Jewell, K. Dietz, V. Farewell, Editors. 231–255, Birkhäuser, Boston.

Yashin A.I. (1985). Dynamics in survival analysis: conditional Gaussian property versus Cameron-Martin formula. In *Statistics and Control of Stochastic Processes*, N. V. Krylov, R. S. Lipster and A. A. Novikov, Editors, 466–475, Springer.

Yashin A.I., and Manton K.G (1997). Effects of unobserved and partially observed covariate processes on system failure: a review of models and estimation strategies. *Statistical Science*, **12**, 20–34.

Simeon M. Berman
Courant Institute of
Mathematical Sciences
New York University
251 Mercer Street
New York, NY 10012.
e-mail : sberman@cims.nyu.edu

Halina Frydman
Stern School of Business
New York University
44 West 4th Street
New York, NY 10012.
e-mail: hfrydman@stern.nyu.edu