OPTIMAL NONPARAMETRIC ESTIMATION FOR SEMIMARTINGALES

by

A. Thavaneswaran and M.E. Thompson
Department of Statistics and Actuarial Science

University of Waterloo,
Waterloo,
Ontario,
Canada N2L 3G1

Institute of Statistics Mimeo Series No. 1825

May 1987

OPTIMAL NONPARAMETRIC ESTIMATION FOR SEMIMARTINGALES

A.Thavaneswaran and M.E.Thompson

Department of Statistics and Actuarial Science,
University of Waterloo,
Waterloo,
Ontario,
Canada N2L 3G1 .

ABSTRACT

This paper extends a result of Godambe on parametric estimation for discrete time stochastic processes to nonparametric estimation for the continuous time case. Following Hasminskii and Ibragimov (1980), the nonparametric problem is formulated as a parametric one but with infinite dimensional parameter. Let $\{ P \}$ be a family of probability measures such that (Ω,F,P) is complete , $(F_t,t\geq0)$ is a standard filtration, and $X=(X_t,F_t,t\geq0)$ is a semimartingale for every $P\epsilon \{ P \}$. For a parameter $\alpha(t)$ suppose $X_t=V_{t,\alpha}+M_{t,\alpha}$ where the V_α process is predictable and locally of bounded variation and the H_α process is a local martingale. Consider estimating equations for $\alpha(t)$ of the form $\int_0^t a_{u,\alpha}dM_{u,\alpha}=0$ where the a_α process is predictable. Under regularity conditions, an optimal form for a_α in the sense of Godambe (Ann. Math. Statist. 31 (1960), 1208-11) is determined. The method is applied to cases where M is linear in α. It is shown that Nelson-Aalen estimate for the cumulative hazard function is optimal in Godambe's sense. A new estimate is obtained for an extended gamma process model. Semimartingale theory is used to indicate proofs of asymptotic normality of test statistics under the null hypotheses considered.

Keywords: Estimating function; Optimality criterion; Semimartingales; Test statistics.

AMS (1980) subject classifications: Primary 62G05, 62G10; secondary 60G35, 60B11, 60H05.

March 27, 1987

OPTIMAL NONPARAMETRIC ESTIMATION FOR SEMIMARTINGALES

1. INTRODUCTION.

Generally, in the literature on stochastic processes estimation is investigated in terms of asymptotic properties. The theory of nonparametric estimation and hypothesis testing for continuous time processes has attracted quite a large number of contributions. However, most of these contributions have been concerned with inference for one particular process or another; for example, considerable literature on nonparametric estimation for counting processes has appeared in papers such as those of Aalen [1] and Gill [9], and Hasminski and Ibragimov [15] have investigated diffusion processes. In this paper we attempt to provide a theory of optimal estimation and testing (treated in the literature only for the counting process case, as the number of realizations tends to infinity) for a semimartingale model which includes point processes, diffusion processes, Ito processes, semi-Markov processes, gamma processes , Liptser's model etc..

This paper extends the result of Godambe on parametric estimation for discrete time stochastic processes to nonparametric estimation for the continuous time case. Following Hasminskii and Ibragimov [14], the nonparametric problem is formulated as a parametric one but with infinite dimensional parameter. Under regularity conditions, an optimal estimating equation for the unknown parameter is obtained in a class of estimating equations. Application of the method sometimes leads to new estimates and test procedures for familiar problems, or to a new motivation for an already well-studied technique. It is shown that this approach yields estimates consistent with other proposals: for Aalen's multiplicative intensity model for counting processes the nonparametric estimate for the cumulative hazard function is optimal in Godambe's sense , as is the maximum likelihood estimate for the signal in the white noise model. New nonparametric estimates are obtained for Liptser's model and for an extended gamma process model. Semimartingale theory is used to indicate proofs of asymptotic normality of test statistics under the null hypotheses considered.

Now our main object is to study the simplest nonparametric estimation for semimartingales through the optimal estimating equation. Following Hasminskii and Ibragimov [15] and Grenander [12] we formulate the nonparametric problem as a parametric one. Certainly in a sense any nonparametric statistical problem can be parametrized by introducing a properly chosen parameter. In fact, an important feature of parametric problems consists in the possibility of imbedding the parameter set in finite dimensional Euclidean space and using the structure of this space. Suppose we are given a nonparametric problem where the parameter set is a subset of some infinite dimensional normed linear space. Then the basic idea is to use the structure of this normed linear space and treat the problem as a parametric one but with an infinite dimensional parameter.

In section 2 we formulate Godambe's optimality criterion in an abstract parameter space (Hilbert space). Viewing the nonparametric problem as a parametric one with the parameter in this infinite dimensional space we obtain an optimal estimating equation for a semimartingale model in section 3.

In the parametric setup the likelihood ratio statistic $R(\theta_0)=2[MaxL_n(\theta)-L_n(\theta_0)] =2(L_n(\hat{\theta})-L_n(\theta_0))$, where $L_n(\theta)$ is the log likelihood , is used for testing a hypothesis of the form $H_0{:}\theta=\theta_0$. Analogously we propose a test statistic based on the optimal estimating function to test the hypothesis $H_0{:}\alpha(t)=\alpha_0(t)$ where $\alpha(t)$ is the parameter of interest, and we derive its asymptotic properties both when the time interval becomes infinite and when the number of copies/realizations becomes infinite using a central limit theorem for semimartingales.

If we consider an infinite dimensional parameter space then our estimating function must also take values in an infinite dimensional space. In order to formulate the optimality criterion for an infinite dimensional parameter space, which we assume to be a real separable Hilbert space H , we consider the notions of mean, dispersion operator, correlation operator and covariance operator on H , following the treatments of A.V.Skorohod [26], A.V.Balakrishnan [3] .

Let x , y , z be elements of H . Let (x,y) denote the scalar product of elements of H and $||x||^2=(x,x)$ the norm of an element x .

Definition : A functional on H is a function defined on H and taking values in the real scalar field R .

Definition : A bounded linear functional f on H is a functional such that

(i)(linearity) $f(\alpha x+\beta y)=\alpha f(x)+\beta f(y)$ for all x , y in H and α,β in R ;

(ii)(boundedness) there exists a finite M such that $|f(x)|<M||x||$ for x in H .

Definition : A bounded (satisfying $||Tx||\leq M||x||$) linear (satisfying (i)) map from a Hilbert space H to a Hilbert space Y is called a bounded linear operator.

Definition : For each bounded linear operator T there exists a unique operator T^* such that $(Tx,y)=(x,T^*y)$ for all x , y in H . This operator T^* is called the conjugate (adjoint) of T .

Definition : An operator T from H to H is self-adjoint if $T^*=T$, and positive (we write $T>0$) if $(Tx,x)>0$ for all x in H , and non-negative (we write $T>0$) if $(Tx,x)\geq 0$ for all $x \epsilon H$.

When H is Euclidean n-space, a linear operator T is representable by an nxn matrix which is symmetric if T is self-adjoint and positive [semi] definite if T is positive [non-negative].

Let $B(H,H)$ be the class of bounded linear operators from H to H .

Definition : A bounded linear operator T in $B(H,H)$ is said to be invertible (or regular) if there exists a bounded linear operator S from H to H (the inverse of T) s.t. $ST=TS=I$, the identity operator.

Definition : Let H_1 and H_2 both be separable(guaranteeing the existence of a countable base). A bounded linear operator A mapping H_1 into H_2 is said to be a nuclear operator (or trace class operator) if for any orthonormal sequence { e_n } in H_1 and a similar sequence { g_n } in H_2 , we have

$$\sum_{n=1}^{\infty} |(Ae_n,g_n)|<\infty .$$

For the special case where $H_1=H_2$, we can define the trace of a nuclear operator by

$$TrA=\sum_{n=0}^{\infty}(Ae_n,e_n) .$$

Let $X_1,X_2,\ldots,X_N$ be H -valued random variables on a probability space (Ω,F,P) . That is, every set of form { $\omega : X_i(\omega)\epsilon B$ } for B a Borel subset of H is in F . Then (z,X_i) corresponds to a real random variable for $i=1,2,3,\cdots,N$.

Definition :- Assume that for all $z_1,z_2,\ldots,z_N$ in H the integral

$E(z_1,X_1)\ldots(z_N,X_N)=\sigma_X(z_1,\cdots,z_N)$ exists. Then the functional $\sigma_X(z_1,\ldots,z_N)$ may be called the moment functional of order N .

Theorem :- If the N -th moment functional is defined for the random variables $X_1,X_2,\ldots,X_N$, then it is a bounded N -linear form. (c.f. A.V.Skorohod [26] Theorem 1, p. 13).

Definition :- The characteristic functional for an H -valued random variable X is defined as $\theta(z)=E\exp[i(z,X)]$ for all z in H .

E.g:- By a Gaussian random variate taking values in a Hilbert space H we understand one whose characteristic functional has the form $\theta(z)=\exp[i(a,z)-1/2(Az,z)]$ where a in H and A is a symmetric bounded nonnegative operator on H .

Remark:- If such an operator A exists it is the dispersion operator (to be defined below), which is the infinite dimensional analogue of the variance-covariance matrix in the finite dimensional case.

Definition:- If $\sigma_X(z)$ is defined it is a linear functional on H and consequently by the Riesz representation theorem there exists an element a in H s.t. $\sigma_X(z)=(a,z)$. This element a is called the mean of the random variable X .

i.e. we have for all z in H , $(a,z)=E(X,z)$.

The second moment function $\sigma_X(z_1,z_2)$ if defined is a bilinear functional. This means that there exists a bounded symmetric(self-adjoint) linear operator D_{12} on H s.t. $\sigma_X(z_1,z_2)=(D_{12}z_1,z_2)$. The existence of this operator follows from Balakrishnan [3] p.151, and it is defined by the relation $(D_{12}z_1,z_2)$ $=E[(X_1,z_1)(X_2,z_2)]$.

Definition :- A covariance operator may be defined by the relation $(C_{12}z_1,z_2)=\sigma_X(z_1,z_2)-\sigma_{X_1}(z_1)\sigma_{X_2}(z_2)$ $=E[(X_1-a_1,z_1)(X_2-a_2,z_2)]$ where a_1,a_2 are the respective means of X_1,X_2 . When $X_1=X_2=X,A=C_{12}$ may be called the dispersion operator of X .

Note: The above is not Skorohod's use of the term covariance operator.

Lemma :-(Skorohod) If $E\|X\|^2<\infty$, then the dispersion operator A exists, is symmetric and nonnegative and has finite trace;

$$Tr A=E\|X\|^2 \text{ if } EX=0 .$$

The following theorem plays a key role in finding a sufficient condition for the optimality to hold.

Theorem:- If U , U_1,U_2 are bounded linear operators from H into H then

(i) the dispersion operator A_U of UX is UAU^* where U^* is the conjugate operator of U and A is the dispersion operator corresponding to X .

(ii) the dispersion operator of $U_1X_1+U_2X_2$ is

$$B=U_1A_{11}U_1^*+U_2A_{22}U_2^*+U_1C_{12}U_2^*+U_2C_{21}U_1^* ,$$

where A_{11} and A_{22} are the dispersion operators of X_1,X_2 respectively , and C_{12} is the corresponding covariance operator.

Without loss of generality we assume that the mean of X_1,X_2 is zero.

Proof:- (c.f. Skorohod [26] p.151, Balakrishnan [3] p.309).

(i) $(A_Uz_1,z_2)=E(Ux,z_1)(Ux,z_2)=E(x,U^*z_1)(x,U^*z_2)=(AU^*z_1,U^*z_2)=(UAU^*z_1,z_2)$.

(ii)
$$\begin{aligned}(Bz_1,z_2)&=E[(U_1X_1+U_2X_2,z_1)(U_1X_1+U_2X_2,z_2)]\\&=E[(U_1X_1,z_1)(U_1X_1,z_2)]+E[(U_2X_2,z_1)(U_2X_2,z_2)]\\&+E[(U_1X_1,z_1)(U_2X_2,z_2)]+E[(U_2X_2,z_1)(U_1X_1,z_2)]\end{aligned}$$

$$=(A_{11}U_1^* z_1, U_1^* z_2)+(A_{22}U_2^* z_1, U_2^* z_2)+(C_{21}U_1^* z_1, U_2^* z_2)+(C_{12}U_2^* z_1, U_1^* z_2) \text{ for all } z_1, z_2 \epsilon H .$$

Hence the result.

Note: $C_{12}=C_{21}^*$.

2. OPTIMALITY CRITERION.

In this section we prove that Godambe's ([10] ,[11]) optimality criterion developed in the framework of parametric estimation for a finite sample carries over in a natural way to many diverse nonparametric problems. In a number of papers dealing with the accuracy of non-parametric estimation, i.e., in the situation where the set F of admissible distributions is not finite dimensional, Levit ([18], [19]) used an approach involving very general informational inequalities to prove the asymptotic optimality of a wide class of non-parametric estimates in the case of real functionals. Our aim of this section is to broaden the class of non-parametric estimators for which finite sample optimality can be proved. This general approach allows us to consider both types of estimation, parametric and nonparametric, from a naturally unified point of view.

We state the basic probabilistic assumptions. First, we assume that (Ω,F,P), is a complete probability space for each P in a family { P } of probability measures. We assume also a family $\mathcal{F}=(F_t, t\geq 0)$ of σ -algebras satisfying the standard conditions for a filtration. We denote by D the space of right-continuous functions $x=(x_t, t\geq 0)$, having limits on the left. We use $X=(X_t,F_t,t\geq 0)$ to denote an F_t -adapted random process with trajectories in the space D. For simplicity we assume that $X_0=0$. We shall denote by M ($\mathcal{F}$, P), M_{loc} ($\mathcal{F}$, P), M_{loc}^2 ($\mathcal{F}$, P) the classes of uniformly integrable, local and locally square-integrable martingales $X=(X_t,F_t)$ respectively. Next we denote by V_{loc} ($\mathcal{F}$, P) the class of random processes $V=(V_t,F_t)$ which have locally (that is, on each finite time interval) bounded variation P a.s. .

Assume that the process $X=(X_t,F_t)$ is a semimartingale for each P ; that is , for each P it can be represented in the form

(1) $$X_t=V_t+M_t ,$$

where $V=(V_t,F_t)\epsilon V_{loc}$ ($\mathcal{F}$, P) and $M=(M_t,F_t)\epsilon M_{loc}$ ($\mathcal{F}$, P). The representation (1) if V is predictable is called the Doob-Meyer decomposition and is unique for each $P\epsilon$ { P }. When we allow V and M to depend on $P\epsilon$ { P } only through $\alpha(P)$ for a parameter α the model (1) can be rewritten as

(2) $$X_t=V_{t,\alpha}+M_{t,\alpha} .$$

If $M\epsilon M_{loc}^2$ ($\mathcal{F}$, P) then there exists a unique predictable process $<M>=(<M>_t,F_t)$ called the quadratic characteristic of M , such that $M^2-<M>\epsilon M_{loc}$ ($\mathcal{F}$, P) (Liptser and Shiryayev [23]).

For example, eventually we want to consider the semimartingale model of the form

(3) $$dX_t=\alpha(t)dR_t+dM_{t,\alpha} \text{ for } t \text{ in } [0,T] ,$$

where R is a real increasing right continuous process with $R_0=0$, and $M\epsilon M_{loc}^2(F,P)$ has variance process given by $<M>_t=\int_0^t C_{s,\alpha}dR_s$, with C being predictable. The function $\alpha=\alpha(t)$, the parameter to be estimated, belongs to a space of functions on $[0,T]$.

First we consider the following general set up. Let $\alpha \epsilon H$, a real separable Hilbert space. Define the mappings $G(X,\alpha)$ such that for each fixed X value , G maps H to H_1 , a real separable Hilbert space , possibly the same as H .

For each fixed $\alpha_0 \epsilon H$ and X value define the derivative $\frac{\partial G}{\partial \alpha}|$ to be the linear operator

$L_{\alpha_0}:H \longrightarrow H_1$ such that

$$L_{\alpha_0}(u)=\frac{d}{d\epsilon}G(X,\alpha_0+\epsilon u)| \ ,$$

assuming the right hand side exists in H_1 .

Remark: Skorohod [26] calls this a weak derivative; Hasminskii and Ibragimov [15] define a less generally applicable version of this, the Frechet derivative , in which case the operator L_{α_0} is bounded (continuous).

For each u in H , $L_{\alpha_0}(u)$ as a function of X is a random element of H_1 , and if its expectation exists as defined in section 1, $EL_{\alpha_0}(u)$ also defines a linear operator $EL_{\alpha_0}:H \text{ --> } H_1$. It may be written $E_\alpha L_{\alpha_0}$ where E_α is the expectation operator corresponding to $\alpha(P)$. Let V_0 be the inverse(assuming it exists) of $E_{\alpha_0}L_{\alpha_0}$, so that V_0 is a linear operator $H_1 \text{ --> } H$.

Let the composition $V_0 o G(X,\alpha_0)$ be denoted by $h(X):H \text{ --> } H$, suppressing for the time being the dependence on α_0 .

Let us assume $H=H_1$.

Following Godambe [11], consider a parameter α to be a function of $P\epsilon$ { P } . Let $G(X,\alpha)=(G_t(X,\alpha)),F_t)$ represent a family of processes indexed by α such that $E_P G_t(X,\alpha)=0$ for each t , for each P and $\alpha=\alpha(P)$. This corresponds to the unbiasedness property of Godambe's [10] optimality criterion which, adapted to this situation, says that G^0 is optimal in $\mathcal{G}$- the class of unbiased estimating functions -if

$$Q=A_h-A_{h^0}$$

is a non-negative operator for all $G\epsilon\mathcal{G}$ and for all P where $h(X)=V_0 o G(X,\alpha_0)$ and $h^0(X)=V_0^0 o G^0(X,\alpha_0)$ and A_h is the dispersion operator for h under α_0 .

Note that from the unbiasedness $A_h=D_h$, the bilinear operator defined in section 1.

Now

$$Q=A_h-A_{h^0}=D_h-D_{h^0}$$
$$=V_0 D_G V_0^*-V_0^0 D_{G^0} V_0^{0*} \text{ (by theorem (i)).}$$

It is easy to see that

$$D_{V_0 G-V_0^0 G^0}=V_0 D_G V_0^*+V_0^0 D_{G^0} V_0^{0*}-V_0 D_{G,G^0} V_0^{0*}-V_0^0 D_{G^0,G} V_0^*=Q$$

if $E_{\alpha_0}L_{\alpha_0}=D_{G,G^0}K$ (equivalently $V_0 D_{G,G^0}=V_0^0 D_{G^0}$ and $D_{G^0}V_0^{0*}=D_{G^0,G}V_0^*$) for all $G\epsilon\mathcal{G}$ and K a constant invertible operator : $H \text{ --> } H$.

Thus we have the result that G^0 is optimal in if $E_{\alpha_0}L_{\alpha_0}=D_{G,G^0}K$ for all G in $\mathcal{G}$ and $\alpha_0\epsilon H$ and $P\epsilon$ { P }

.

3 EXAMPLES .

Example 3.0: For the semimartingale model (3)

$$dX_t=\alpha(t)dR_t+dM_{t,\alpha}$$

Following Godambe [11] we consider estimating equations of the form $G_t(X,\alpha)=0$ where $G_t(X,\alpha)=\int_0^t a_{s,\alpha}dM_{s,\alpha}$, $a_{s,\alpha}$ is predictable w.r.t. $F_t=\sigma[X_s,s\leq t]$ and realizations of G_t belong to $L_2[0,T]$.

Assume that for each P ,

(a) $M=(M_{t,\alpha},F_t)\epsilon M^2_{loc}(\mathcal{A},P)$, and $M_{t,\alpha}$ is Frechet differentiable w.r.t. α , with the corresponding operator $I_{s,\alpha}$.

(b) $(a_{s,\alpha},F_s)$ is predictable and Frechet differentiable w.r.t. α ,

(c) the Ito stochastic integrals $\int_0^t \frac{\partial a_{s,\alpha}}{\partial \alpha}dM_{s,\alpha}$, $\int_0^t a_{s,\alpha}dM_{s,\alpha}$ exist ,

(d) the operator $\frac{\partial G_{t,\alpha}}{\partial \alpha}$ has nonzero expectation and can be expressed as $\int_0^t \frac{\partial a_{s,\alpha}}{\partial \theta}dM_{s,\alpha}+\int_0^t a_{s,\alpha}dI_{s,\alpha}$.

The optimal estimating function if it exists will take the form

$$G_t^0(X,\alpha)=\int_0^t a^0_{s,\alpha}dM_{s,\alpha} .$$

For it is easy to show that

$$E_{\alpha_0}L_{\alpha_0}(u)=-E_{\alpha_0}\int a_{s,\alpha_0}u(s)dR_s .$$

and the corresponding covariance operator(c.f. Balakrishnan p. 317) is given by

$E_{\alpha_0}[G_tG_t^0]u(s)=E_{\alpha_0}[\int a^0_{s,\alpha_0}a_{s,\alpha_0}u(s)d\langle M\rangle_{s,\alpha}]$ where $\langle M\rangle_{s,\alpha}$ is the variance process of $M=(M_t,F_t)$.

Hence the optimal $a^0_{s,\alpha}$ may be written formally as $=\frac{dR_s}{d\langle M\rangle_{s,\alpha}}=C^{-1}_{\alpha,s}$.

This gives the optimal estimating equation as

$$G_t^0(X,\alpha)=\int_0^t \frac{dR_s}{d\langle M\rangle_{s,\alpha}}dM_s=\int_0^t \frac{dR_s}{d\langle M\rangle_{s,\alpha}}[dX_s-\alpha(s)dR_s] .$$

In the examples considered below the optimal estimating equation takes the form $\int_0^t (dX_s-\alpha(s)dR_s)=0$.

Note: As a consequence it can be argued that for any bounded linear functional of α the optimal estimate is obtained by substituting the optimal estimate of α for α .

Example 3.1: Aalen's (1978) model.

Aalen was the first to introduce the following semimartingale to study the nonparametric estimate for censored data.

Assume that we are observing on the interval [0,1] a random function X_t where

(4) $$dX_t = \alpha(t)Y(t)dt + dM_{t,\alpha}$$

where the unknown parameter α and the observable Y belong to a convex subset Σ of the Hilbert space $L_2(0,1)$, and M is a zero mean square integrable martingale. In the life testing setup X_t is the number of deaths up to time t and $Y(t)$ is the number at risk at time $t-$, and both are observables.

We next consider the problem of estimating the value of a given bounded linear functional

$F : \Sigma \dashrightarrow R$ at an unknown point α ; the problem of estimating $\beta_t = \int_0^t \alpha(s)J(s)ds$, where $J(s) = I\{Y(s)>0\}$, is a special case of this.

Let $F(\alpha) = \int_0^1 f(t)\alpha(t)dt$ be a bounded linear functional on Σ .

Let Σ be the ball in $L_2(0,1)$ of radius 1 and center 0.

Using the result of Example 3.0 for the model of the form (4) the optimal estimating equation can be written formally as

$$G_t^0(X,\alpha) = \int_0^t Y(s)\frac{ds}{d\langle M\rangle_{s,\alpha}}dM_{s,\alpha} .$$

In the case of a counting process the variance process $\langle M\rangle_{t,\alpha} = \int_0^t \alpha(s)Y(s)ds$, and we have

$$G_t^0(X,\alpha) = \int_0^t \alpha^{-1}(s)J(s) \quad dM_{s,\alpha} .$$

This is true for all $t\in [0,1]$. Hence $G_t^0(X,\alpha)=0$ is equivalent to $\int_0^t J(s)dM_{s,\alpha}=0$, and for the functional $\beta_t = \int_0^t \alpha(s)J(s)ds$, the estimate is given by

$$\beta_t^0 = \int_0^t J(s)\frac{dX_s}{Y(s)} .$$

This gives the optimality for the Nelson-Aalen estimate for the cumulative hazard function β_t which is widely used in survival analysis. Jacobsen [17] also proved that Aalen estimator is asymptotically equivalent to the maximum likelihood estimator in an extended model. Instead of considering the extended model we have shown that the estimate is optimal for Aalen's model in Godambe's sense by considering a class of unbiased estimating functions.

Now it is of interest to show that the optimal estimate obtained above attains the Cramer-Rao lower bound.

Recently Grenander [12] has extended the Cramer-Rao inequality to an abstract parameter space and considered estimation of an infinite dimensional parameter for a Poisson process model. Following his work(example 3, p.488) we show that our estimate attains the Cramer-Rao lower bound for the counting process model.

For the counting process model the likelihood may be written (see Liptser and Shiryayev [23])

$$f_t(X,\alpha)=\exp(\int_0^t \ln(Y(s)\alpha(s))dX_s-\int_0^t [1-Y(s)\alpha(s)]ds)$$

where t is such that $\int_0^t (1+\alpha(s)Y(s))ds<\infty$.

Grenander's theorem says that under some regularity conditions for an unbiased estimate α^* of α and for a bounded linear functional l the Cramer-Rao inequality is given by

$$Var_\alpha[l(\alpha^*)]\geq\frac{l^2(u)}{E_\alpha[\lambda_u^2(X,\alpha)/f^2(X,\alpha)]}$$

where $\lambda_u(X,\alpha)$ is the weak derivative of $f(X,\alpha)$ at α in the direction of u .

He suggests that u can be chosen in such a way that the R.H.S. of the above inequality is as large as possible.

When $l_t(\alpha)=\beta_t=\int_0^t \alpha(s)ds$ it is easy to show that the R.H.S. of the Cramer-Rao equality equals

$$\frac{[\int_0^t u(s)ds]^2}{[\int_0^t \frac{u(s)^2Y(s)}{\alpha(s)}ds]},$$

which is maximized for u such that $u(s)Y(s)=\alpha(s)$.

Hence the optimal estimate β^0 attains the Cramer-Rao lower bound .

i.e. for a counting process model the optimal estimate for the cumulative hazard function is the unbiased estimate which attains the lower bound.

Example 3.2:

The diffusion process model given by Hasminskii and Ibragimov [14] is of the form

$$dX_t=\alpha(t)dt+\epsilon dW_t \text{ for t in } [0,T]$$

with $\alpha(t)$ the unknown parameter and ϵ known.

Using the result of example 3.0 the optimal estimating equation is given by

$$G_t^0(X,\alpha)=\int_0^t dW_s$$

which gives the formal estimate $\alpha^0(t)=\frac{dX_t}{dt}$ Therefore for any bounded linear functional $F(\alpha)=\int_0^T f(t)\alpha(t)dt$ of α the estimate F^0 would be given by $F^0=\int_0^T f(t)dX_t$.

An argument similar to that of example 3.1 can be given to show that F^0 attains the Cramer-Rao lower bound.

We now show that the optimal estimate for a linear functional embraces the minimax estimate for the diffusion process model. Hasminskii and Ibragimov [14] studied the nonparametric estimation problem for the above model. There it is assumed that α is in $\Sigma \subseteq L_2[0,1]=L_2$. For the value of a linear functional $F(\alpha)$ on Σ at the point α , the class of linear estimates $\hat{F}=\int_0^1 m(t)dX(t)$ where m in L_2 is denoted by M . The minimax risk is

$$\Delta^2(\epsilon,M)=Inf\,Sup E_\alpha|F(\alpha)-\hat{F}|^2 .$$

The following theorem was established.

Theorem 3.3: Let Σ be a convex, closed and symmetric subset of $L_2[0,1]$.

Then

$$\Delta^2(\epsilon,M)=Sup\,\epsilon^2F^2(\alpha)(\epsilon^2+||\alpha||^2)^{-1} ,$$

where $||.||=||.||_{L_2[0,1]}$.

This problem corresponds to ours when Σ coincides with all of L_2 . In this case the functional F is bounded and representable in the form $F(\alpha)=\int_0^1 l(t)\alpha(t)dt$.

Then the optimal estimate $F^0=\int_0^1 l(t)dX(t)$ has risk

$E_\alpha|F^0-F(\alpha)|^2=\epsilon^2E|\int_0^1 l(t)dW(t)|^2=\epsilon^2||F||^2$ which coincides with the bound in the above theorem, since the Cauchy- Schwarz inequality implies that

$$Sup\,\epsilon^2\frac{F^2(\alpha)}{\epsilon^2+||\alpha||^2}=\epsilon^2||F||^2 .$$

i.e. in this case the optimal estimate is the minimax estimate.

Example 3.4:

Liptser [21] considered the semimartingale model of the form

$$dX_t=\alpha(t)d\langle N\rangle_t+dN_t \text{ for } t \text{ in } [0,T]$$

where N_t is a zero mean square integrable martingale with right continuous trajectories having left hand limits and $\alpha(t)$ is independent of t . Here we allow α to be a function of t and consider the corresponding nonparametric problem.

The optimal estimating equation for α in this case is $G_t^0(X,\alpha)=\int_0^t dN_s$, and the corresponding estimate is

$$\alpha^0=\frac{dX_t}{d\langle N\rangle_t} .$$

Thus we have obtained a new optimal nonparametric estimate for Liptser's model.

Example 3.5

In this example we consider a new semimartingale model for the extended gamma process defined by Dykstra and Laud [5] and obtain a new estimate.

Let $G(\gamma,\beta)$ denote the gamma distribution with γ the shape parameter and β the scale parameter for $\gamma,\beta>0$. Let $\gamma(t)$, $t\geq 0$, be a nondecreasing continuous real valued function such that $\gamma(0)=0$, and let $\beta(t)$, $t\geq 0$, be a positive right continuous real valued function, bounded away from zero, with left hand limits. Let $Z(t)$, $t\geq 0$, defined on an appropriate probability space (Ω,F,P) , denote a gamma process with independent increments corresponding to $\gamma(t)$. That is , $Z(0)=0$, $Z(t)$ has independent increments, and for $t>s$, $Z(t)-Z(s)$ is $G(\gamma(t)-\gamma(s),1)$. Then for the extended gamma process defined by $X(t)=\int_0^t \beta(s)dZ(s)$ we propose a semimartingale model given by

$$dX_t=\beta(t)d\gamma(t)+dM_t$$

where M is a locally square integrable martingale with mean zero and variance process $<M>_t=\int_0^t \beta^2(s)d\gamma(s)$. Then the optimal estimating equation for $F(\beta)=\int_0^t \beta(s)d\gamma(s)$ is given by $G^0(X,\beta)=\int_0^t \frac{1}{\beta(s)}dM_s$ and the optimal estimate by $F^0=\int_0^t dX_s$. Thus we have obtained a new estimate for a new semimartingale model.

4. ASYMPTOTIC EQUIVALENCE OF PARAMETER AND ESTIMATE.

In this section we consider the asymptotic equivalence of parameter and optimal estimate in the linear case as $T \rightarrow \infty$.

For a model of the form

$$dX_t=\alpha(t)dR_t+dM_{t,\alpha} ,$$

where R is of bounded variation, the optimal estimating equation is of the form $G_t^0(X,\alpha)=\int_0^t (dX_s-\alpha(s)dR_s)=0$.

Hence the optimal estimate of $\beta_t=\int_0^t \alpha(s)dR_s$ can be written as

$$\beta_t^0=\beta_t+M_{t,\alpha} .$$

Theorem:- If (i) $\frac{\beta_t}{<M>_{t,\alpha}} \rightarrow \sigma^2>0$ a.s.

and (ii) $<M>_{t,\alpha} \rightarrow \infty$ a.s.

then

$$\frac{\beta_t^0}{\beta_t} \rightarrow 1 \text{ a.s. as } t \rightarrow \infty .$$

Proof:

$$\frac{\beta_t^0}{\beta_t}=1+\frac{M_{t,\alpha}}{\beta_t}=1+\frac{\left(\frac{M_{t,\alpha}}{<M>_{t,\alpha}}\right)}{\left(\frac{\beta_t}{<M>_{t,\alpha}}\right)} .$$

Now by the strong law of large numbers for martingales (Liptser [21]) we have $\frac{M_{t,\alpha}}{<M>_{t,\alpha}}$ --> 0 a.s. under assumption (ii) of the theorem.

Using condition(i) and taking the limit as t --> ∞ we have the result.

Note: For Aalen's model $\beta_t=<M>_{t,\alpha}$ and hence $\sigma^2=1$. For the diffusion process model of Example 3.2 the assumption(i) is the same as $\frac{\beta_t}{t}$ --> σ^2 a.s..

5. TEST STATISTICS FROM THE ESTIMATING EQUATION.

In this section we propose a test statistic based on the optimal estimating equation to test the hypothesis $H_0{:}\alpha(t)=\alpha_0(t)$ where $\alpha_0(t)$ is known , in the collection of semimartingale models

$$dX_n(t)=\alpha(t)dR_n(t)+dM_n(t)$$

for n =1,2,3,....

Then the statistic,in analogy with the likelihood ratio statistic in the parametric set up regarding the likelihood equation as the optimal one, may be defined under H_0 by

$$SS_n(\alpha_0(t))=G_{n,t}^0(\alpha^0)-G_{n,t}^0(\alpha_0)=\int_0^t dM_n(s) ,$$

where $G_{n,t}(\alpha)=\int_0^t b_{s,\alpha}dM_n(s)$ and $G_{n,t}^0(\alpha)$ is the corresponding optimal estimating function as in Example 3.0 .

i.e. $SS_n(\alpha_0(t))\epsilon M^2(\mathcal{J},P)$.

Now we describe two cases in which the asymptotics may be studied.

Case (a): If we consider a sequence of semimartingales ($X_n(t){:}t\epsilon[0,T]$ for fixed T) as before then we may use the asymptotically standard normally distributed statistic $U_n(\alpha_0,T)=\frac{SS_n(\alpha_0,T)}{<SS_n(T)>^{1/2}}$ for testing H_0 .

Note:- The proof of asymptotic normality follows from a martingale central limit theorem (see for example Liptser and Shiryayev [13]) applied directly to the standardized martingale $U_n(\alpha_0,T)$.

This is the case which is extensively studied in the literature; see for example Anderson et. al. [2] Example 4.2.

Case (b): If we consider a single realization of the semimartingale $X=(X_t,F_t)$ in the time interval $[0,T]$ then we may use the asymptotically (T --> ∞) standard normally distributed statistic $U(T,\alpha_0)=\frac{SS_1(\alpha_0,T)}{<SS_1(\alpha_0,T)>^{1/2}}$ for testing H_0 .

Note: The proof of asymptotic normality follows from a central limit theorem (Feigin [7] or Linkov [illegible]) applied directly to the standardized martingale $U(T,\alpha_0)$.

In the literature this case is considered only for the parametric set up . Hence the above is a generalization in some sense.

Most of the recently suggested test statistics for the one-sample situation are special cases of our proposed statistic and hence their asymptotic distribution can be found from martingale central limit theorems. This generalization provides a new motivation for an already well-studied technique in the counting process case and also provides new test statistics together with their asymptotic distribution for a general semimartingale model. Moreover the theory of optimal estimating equations suggests a method to motivate test statistics to test values of infinite dimensional parameters and is hence applicable in the nonparametric set up.

1. Bibliography

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

2 Anderson P.K.,Borgan O., Gill R. and Keiding N.(1982) Linear nonparametric tests for comparison of counting processes with application to censored survival data. Int. Stat. Review 50 , 219-258.

3 Balakrishnan A.V. (1981) Applied Functional Analysis, Application of Mathematics 3. Springer-Verlag; New York.

4 Billingsley P. (1961) Statistical Inference for Markov Processes. The University of Chicago Press; Chicago and London.

5 Dykstra R.L. and Laud P. (1981) A Bayesian nonparametric approach to reliability. Ann. Statist. 9 , 356-367.

6 Elliot R.J.(1982) Stochastic Calculus and Applications. Springer-Verlag; New York.

7 Feigin P.D.(1976) Maximum likelihood estimation for continuous-time stochastic processes. Adv. in Prob. 8, 712-736.

8 Geman S. and Hwang C.R. (1982) Nonparametric maximum likelihood estimation by the method of sieves. Ann. Statist. 10 , 401-414.

9 Gill R.D. (1980) Censoring and Stochastic Integrals. Mathematisch Centrum Amsterdam.

10 Godambe V.P. (1960) An optimum property of regular maximum likelihood equation . Ann. Math. Statist. 31, 1208-11.

11 Godambe V.P. (1985) The foundations of finite sample estimation in stochastic processes. Biometrika 72, 419-428.

12 Grenander Ulf(1981) Abstract Inference . Wiley; New York.

13 Harrington D.P. and Fleming T.R. (1982) A class of rank test procedures for censored survival data. Biometrika 69, 533-546.

14 Hasminskii R.Z. and Ibragimov I.A. (1985) On nonparametric estimation of the value of a linear functional in a Gaussian white noise. Theory Prob. Appl. 29, 18-32.

15 Hasminskii R.Z. and Ibragimov I.A.(1980) Some estimation problems for stochastic differential equations, Lecture Notes in Control and Information Sciences, 25, 1-12.

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

17 Jacobsen M. (1984) Maximum likelihood estimation in the multiplicative intensity model: A survey. Int. Stat. Review 52, 193-208.

18 Levit B. (1973) On optimality of some statistical estimates. Proc. Prague Symp. Asymptotic Statistics 2, 215-238.

19 Levit B. (1975) On efficiency of a class of nonparametric estimates. Theory Prob. and Appl. 20, 723-740.

20 Lin'kov Yu.N. (1982) Asymptotic properties of statistical estimators and tests for Markov processes. Theor. Probability and Math. Statist. No. 25, 83-98.

21 Liptser R.S.(1980) A strong law of large numbers for local martingales. Stochastics 3, 217-228.

22 Liptser R.S. and Shiryayev A.N. (1977). Statistics of Random Processes. 2. Springer-Verlag;New York.

23 Liptser R.S. and Shiryayev A.N. (1980) A functional central limit theorem for semimartingales. Theory Prob. Appl. 25,667-688.

24 Metivier M.(1982) Semimartingales. Walter de Gruyter; New York.

25 Nelson W. (1972) Theory and application of hazard plotting for censored failure data. Technometrics 14, 945-965.

26 Skorohod A.V. (1974) Integration in Hilbert Spaces. Springer-Verlag; New York.

27 Thavaneswaran A. and Thompson M.E. (1984) Optimal estimation for semimartingales.(to be published in Journal of Appl. Prob. 1986).