Stochastic Processes and their Applications 28 (1988) 259-265
North-Holland

A CRITERION FOR FILTERING IN SEMIMARTINGALE MODELS

A. THAVANESWARAN
Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina 27599-3260, USA

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

Received 15 December 1986
Revised 26 August 1987

Recently there has been a growing interest in the problems of inference for stochastic processes when the underlying distribution is not specified in terms of a parametric family. Godambe's (1985) approach is here employed to obtain estimates for random signals for a continuous semimartingale model. The method, which avoids specification of the underlying distribution, leads to estimation for nonconjugate prior situations which is computationally attractive as well as optimal in a restricted sense. A number of techniques in the recent literature are special cases.

estimating functions * Ito's differential rule

1. Introduction

The problem of filtering of random standard processes consists of the following. On some probability space (Ω, F, P) with filtration $(F_t, t \geq 0)$ there is given a partially observable two-dimensional process (X, Z) where $X = (X_t, t \geq 0)$ is the observable component and $Z = (Z_t, t \geq 0)$ is the unobservable component. It is required to find a "convenient" representation for an appropriate estimate of the current state $\theta_t = h_t(Z)$ of an unobservable process $\Theta = (\theta_t, t \geq 0)$, based on observations for the process X up to the moment t.

In this paper we introduce an optimality criterion for the process of estimating functions for θ_t, $t \geq 0$. It is based on a criterion introduced by Ferreira (1982a), and is analogous to the criterion used for stochastic process parameter estimation by Godambe (1984), Thavaneswaran and Thompson (1986) and Hutton and Nelson (1986).

In Section 2 we introduce the optimality criterion and class of estimating functions for treating the discrete time case, and in Section 3 the continuous time case is considered.

0304-4149/88/$3.50

2. Optimal estimation in discrete time case

We begin not with stochastic processes but with the case of jointly distributed random variables. Let X, θ be jointly distributed random variables with finite second moments, with X but not θ being observed. An estimating equation for the realized value of θ is an equation $g(X, \theta)=0$ which is solved for θ in terms of the data X. The function $g(X, \theta)$ may be called an estimating function. Within a class G of unbiased estimating functions satisfying $Eg=0$ and appropriate regularity conditions, a function can be said to be optimal if it minimizes the quotient $E[g/E(\partial g/\partial\theta)]^2$ for all $g\in G$ (Ferreira, 1982a). Briefly, an estimating function which is optimal in this sense has minimal variance, subject to having a fixed "sensitivity" $E[\partial g/\partial\theta)$; in large sample situations the quotient to be minimized approximates the variance of the resulting estimate of θ. A sufficient condition for g^0 to be optimal in G is that $E[gg^0]=KE[\partial g/\partial\theta]$ for all g in G where K is a constant. This condition has been shown also to be necessary by H. Mantel. See Godambe and Thompson (1985).

Ferreira (1982a) showed that in a wide class of unbiased estimating functions when the joint distribution of X, θ is known the optimal estimating function is $\partial \log f_{\theta|X}(\theta|x)/\partial\theta$, yielding as estimate of θ the "posterior" mode, given $X=x$. It is also easily shown that if we restrict to the class of estimating functions of the form

$$g(X, \theta)=t(X)-\theta$$

where $E(t(X)-\theta)=0$, the optimal estimating function is $g(X, \theta)=E(\theta|X)-\theta$, and the estimate is the "posterior" mean of θ, given X. Note that this coincides with the posterior mode when the posterior distribution is symmetric and unimodal, and that its form will be optimal for the class of underlying joint distributions with conditional expectations $E(\theta|X)$ the same as for the given joint distribution.

Suppose we restrict the class of estimating functions further to functions having the form

$$g(X, \theta)=\theta-a-CX,$$

with $Eg=0$, or equivalently,

$$g(X, \theta)=\theta-E(\theta)-C(X-E(X)).$$

Here, as in the previous case, since $\partial g/\partial\theta=1$ the optimality criterion will lead to minimal mean squared error; and since here g is linear in X we may refer to the estimate as "least squares" or the "least squares approximation" to the optimal estimate. The optimal g has the form given in the following theorem.

Theorem 2.1. *Let θ, X be random variables having finite moments $m_\theta=E(\theta)$, $m_X=E(X)$, $D_{\theta\theta}=\mathrm{Cov}(\theta,\theta)$, $D_{\theta X}=\mathrm{Cov}(X,\theta)$, $D_{XX}=\mathrm{Cov}(X,X)$. Let*

$$G=[g|g=(\theta-m_\theta)-C(X-m_X) \text{ for some constant } C];$$

then the optimal estimating function for θ is given by $g^*=\theta-m_\theta-D_{\theta X}D_{XX}^{-1}(X-m_X)$ and the optimal estimate is $\theta^*=m_\theta+D_{\theta X}D_{XX}^{-1}(X-m_X)$.

Remark. If (θ, X) are jointly normal then $\theta^* = E(\theta|X)$, the conditional expectation, which is the optimal estimate of θ in the mean square sense.

Proof. g^* is optimal in G if $E(gg^*) = KE(\partial g/\partial\theta)$ for every g in G, or equivalently since for $g \in G$ the expectation $E(\partial g/\partial\theta) = 1$, if $E[(gg^* - g^{*2})] = 0$. With g^* as given, $E^*_{gg} = D_{\theta\theta} - D_{\theta X}D^{-1}_{XX}D_{\theta X}$, which is independent of C, proving the result. □

If the class of underlying joint distributions of X, θ is sufficiently broad that only the means m_θ and m_X are known, the functions g in G are the only unbiased estimating functions for θ. It is true that computing the optimal g involves knowledge (possibly of $D_{\theta X}$, D^{-1}_{XX}) that may make other unbiased estimating functions available. However, if the product $D_{\theta X}D^{-1}_{XX}$ is only approximately known, an approximately optimal estimating function from G can still be computed, while G remains the source of those estimating functions which are certainly unbiased. In the filtering context, Krichagina et al. (1984) suggest a related justification for restriction to linear functions of θ, namely "simplicity in computations and lower sensitivity with respect to data inaccuracy".

We now turn to the case of estimating the current state of a signal process Θ when the process X is observed and both processes have a discrete time parameter.

Theorem 2.2. *Let $X = (X_0, X_1, \ldots)$ and $\Theta = (\theta_0, \theta_1, \ldots)$ be square integrable stochastic processes on (Ω, F, P) and for $t = 0, 1, 2, \ldots$, let F^X_t be the σ-field generated by $X_0, \ldots, X_t$. Then in the class $G = [g|g = \theta_{t+1} - E(\theta_{t+1}|F^X_t) - C(t, X)(X_{t+1} - E(X_{t+1}|F^X_t))]$ of estimating functions with $C(t, X)$ measurable with respect to F^X_t the optimal one is given by*

$$g^* = [\theta_{t+1} - E(\theta_{t+1}|F^X_t) - d_{12}d^{-1}_{22}(X_{t+1} - E(X_{t+1}|F^X_t))],$$

and the optimal estimate is

$$\theta^*_{t+1} = E(\theta_{t+1}|F^X_t) + d_{12}d^{-1}_{22}(X_{t+1} - E(X_{t+1}|F^X_t))$$

where $d_{22} = \mathrm{Cov}(X_{t+1}, X_{t+1}|F^X_t)$ and $d_{12} = \mathrm{Cov}(X_{t+1}, \theta_{t+1}|F^X_t)$.

Proof. Since $E(g|F^X_t) = 0$ a.s. for each $g \in G$, g^* is optimal in G if $E(gg^*|F^X_t) = KE(\partial g/\partial\theta)|F^X_t)$ a.s. for every g in G, or equivalently, since $\partial g/\partial\theta = 1$, if $E[g^*(g^* - g)|F^X_t] = 0$ a.s., for every g in G. Then the optimal value for C is given by $C^* = d_{12}d^{-1}_{22}$, as in the proof of Theorem 2.1. □

Remark. West, Harrison and Migon (1985) studied the related dynamic forecasting problem, in the discrete time case, for dynamic generalized linear models using the same state evolution (same θ sequence) but they specified the conjugate prior form for the natural parameter in the general exponential family model (observation sequence). The key feature of their analysis is the use of conjugate prior and posterior distributions for the exponential family parameter. In our analysis we do not make any distributional assumptions but we restrict ourselves to a certain class of estimating functions. Neither approach is full Bayesian since a full Bayesian analysis requires also the posterior for $\theta_t|F_t^X$; this is not available because the prior is only partially specified (only the first two moments) and the model does not provide a likelihood for θ_t.

3. Optimal estimation in continuous time case

3.1. The Theorem

In this section we derive the analogous optimal estimating equation for continuous semimartingale models regarding parameters as random signals. Let $(X_t, t\geq 0)$ be the $\mathbb{R}^k$ valued observation process governed by the semimartingale model of the form

$$X_t = X_0 + \left(\int_0^t Z_u^{\mathrm{T}} dR_u\right)^{\mathrm{T}} + H_t,$$

and let the $\mathbb{R}^n$ valued state (cumulative signal) process $(\theta_t, t\geq 0)$ have a semimartingale representation

$$\theta_t = \theta_0 + \left(\int_0^t \mu_u^{\mathrm{T}} dr_u\right)^{\mathrm{T}} + m_t$$

where H, m are continuous independent vector-valued martingales of dimensions k and n respectively. The integrators R and r are observable continuous matrix-valued processes of dimensions $k\times k$ and $n\times n$ respectively, increasing in the sense that for $s<t$, $\int_s^t \mathrm{d}R_u$ and $\int_s^t \mathrm{d}r_u$ are positive definite. The Z and μ processes are unobservable; typically, Z is a functional of the cumulative signal process $(\theta_t, t\geq 0)$. The processes $(X_t, t\geq 0)$ and $(\theta_t, t\geq 0)$ are assumed continuous and square integrable.

In the following lemmas, $\hat{\eta}_t$ denotes $E(\eta_t|F_t^X)$ for any random process $(\eta_t, t\geq 0)$, where now F_t^X is generated by $(X_s, r_s, R_s, 0\leq s\leq t)$.

Lemma 3.1. *If $(\eta_t, F_t, t\geq 0)$ is a [square integrable] [continuous] martingale, then so is $(\hat{\eta}_t, F_t^X, t\geq 0)$.*

Proof of martingale property:

$$E(\hat{\eta}_t|F_s^X) = E(E(\eta_t|F_t^X)|F_s^X) = E(E(\eta_t|F_s)|F_s^X) = E(\eta_s|F_s^X) = \hat{\eta}_s \text{ a.s.,} \quad s<t.$$

Lemma 3.2. *If $(\gamma_u, u\geq 0)$ is adapted to $(F_u, u\geq 0)$ and R is a matrix integrator adapted to $(F_t^X, t\geq 0)$, then $E(\int_0^t \gamma_u^{\mathrm{T}}\,\mathrm{d}R_u|F_t^X) - \int_0^t \hat{\gamma}_u^{\mathrm{T}}\,\mathrm{d}R_u$ if it exists and is integrable is a martingale with respect to $(F_t^X, t\geq 0)$.*

Proof. This follows since

$$E\left(\int_s^t \gamma_u^{\mathrm{T}}\,\mathrm{d}R_u - \int_s^t \hat{\gamma}_u^{\mathrm{T}}\,\mathrm{d}R_u | F_s^X\right) = 0 \quad \text{a.s.}$$

Now from Lemmas 3.1, 3.2 we note that by taking expectations conditional on F_t^X we can represent the observation process as

$$X_t = X_0 + \left(\int_0^t \hat{Z}_u^{\mathrm{T}}\,\mathrm{d}R_u\right)^{\mathrm{T}} + \tilde{H}_t$$

where $(\tilde{H}_t, t \geqslant 0)$ is a martingale with respect to $(F_t^X, t \geqslant 0)$. Moreover,

$$\hat{\theta}_t = \hat{\theta}_0 + \left(\int_0^t \hat{\mu}_u^{\mathrm{T}}\,\mathrm{d}r_u\right)^{\mathrm{T}} + \tilde{m}_t$$

where $(\tilde{m}_t, t \geqslant 0)$ is also a martingale with respect to $(F_t^X, t \geqslant 0)$.

Let G be the class of continuous square integrable estimating functions g of the form

$$g_t = \theta_t - \hat{\theta}_0 - \left(\int_0^t \hat{\mu}_u^{\mathrm{T}}\,\mathrm{d}r_u\right)^{\mathrm{T}} - \int_0^t C_u\,\mathrm{d}\tilde{H}_u$$

$$= M_t - \int_0^t C_u\,\mathrm{d}\tilde{H}_u$$

where $(C_t, t \geqslant 0)$ is an $n \times k$ dimensional $(F_t^X, t \geqslant 0)$-predictable process, integrable with respect to $\mathrm{d}\tilde{H}_u$. Note that $E(g_t | F_t^X)$ is a mean 0 martingale with respect to $(F_t^X, t \geqslant 0)$, and in this sense g_t in G may be said to be unbiased. Its form suggests that it is the increments of the θ process which are being optimally estimated, rather than the θ values themselves. Adapting the optimality criterion of previous sections, we say g^* in G is optimal if

$$E(g_t - g_t^*)g_t^* = 0 \quad \text{for each } t. \quad \square \tag{3.1}$$

Theorem 3.3. *The function $g^* \in G$ is optimal if*

$$g_t^* = (\theta_t - \hat{\theta}_0) - \left(\int_0^t \hat{\mu}_u^{\mathrm{T}}\,\mathrm{d}r_u\right)^{\mathrm{T}} - \int_0^t C_u^*\,\mathrm{d}\tilde{H}_u$$

where $C_u^{*\mathrm{T}}$ can be represented as

$$E(d\langle \tilde{H}, \tilde{H}^{\mathrm{T}}\rangle_u^+ d\langle \tilde{H}, M^{\mathrm{T}}\rangle_u | F_u^X),$$

the symbol + denotes inverse and $M_t = \theta_t - \theta_0 - (\int_0^t \hat{\mu}_u^{\mathrm{T}}\,\mathrm{d}r_u)^{\mathrm{T}}$.

Proof. For simplicity we give the proof for $k = 1$, $n = 1$. Using Ito's differential rule for the product UV of two semimartingales U and V, namely

$$\mathrm{d}(UV)_t = U_t\,\mathrm{d}V_t + V_t\,\mathrm{d}U_t + \mathrm{d}\langle U, V\rangle_t$$

where $\langle U, V\rangle$ is the corresponding quadratic variation process associated with U, V, we have

$$d((g_t-g_t^*)g_t^*)=(g_t-g_t^*)\,dg_t^*+g_t^*\,d(g_t-g_t^*)+d\langle g-g^*, g^*\rangle_t.$$

We note the following:

(a) $g_t-g_t^*=\int_0^t(C_u-C_u^*)\,d\tilde{H}_u$, $t\geq 0$ is a martingale with respect to $(F_t^X, t\geq 0)$, and hence $E(\int_0^t g_u^*\,d(g_u-g_u^*)|F_t^X)$, $t\geq 0$, is a zero mean martingale with respect to $(F_t^X, t\geq 0)$;

(b) similarly $E(\int_0^t(g_u-g_u^*)\,dg_u^*|F_t^X)$, $t\geq 0$ is the same kind of object.

Applying Lemmas 3.1, 3.2 we have

$$E[(g_t-g_t^*)g_t^*|F_t^X]=E(\langle g-g^*, g^*\rangle_t|F_t^X)+N_t \tag{3.2}$$

where $(N_t, t\geq 0)$ is a zero-mean $(F_t^X, t\geq 0)$ martingale.

Using Lemma 3.2, (3.2) implies that (3.1) would be true if

$$\int_0^t (C_u-C_u^*)C_u^*\,d\langle\tilde{H},\tilde{H}\rangle_u=\int_0^t (C_u-C_u^*)E(d\langle\tilde{H}, M\rangle_u|F_u^X)$$

which is true if $C_u^*=E(d\langle\tilde{H}, M\rangle_u/d\langle\tilde{H},\tilde{H}\rangle_u|F_u^X)$, and the optimal estimate is given by

$$\theta_t^*=\hat{\theta}_0+\int_0^t\hat{\mu}_u\,du+\int_0^t C_u^*\,d\tilde{H}_u.$$

Remark. When $H=W$, a Wiener process, this corresponds to the model treated in Elliott (1982, p. 280), and the solution is easily seen to reduce to the one he derives via the innovation method. Thus θ_t^* is the "posterior" or conditional mean of θ_t, given F_t^X. In the case where H, m are R^k, R^n valued processes respectively, the model of this section corresponds to the setup in Curtain and Pritchard (1978, p. 4).

Remark. The *criterion* for optimality in (3.1) is of course applicable to more general semimartingale models than the one considered above.

References

R.F. Curtain and A.J. Pritchard, Infinite Dimensional Linear Systems Theory (Springer-Verlag, New York, 1978).

P.E. Ferreira, Estimating equations in the presence of prior knowledge, Biometrika 69 (1982a) 667–669.

P.E. Ferreira, Multiparametric estimating equations, Ann. Inst. Statist. Math. 34 Part A (1982b) 423–431.

R.J. Elliott, Stochastic Calculus and Applications (Springer-Verlag, New York, 1982).

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

V.P. Godambe and M.E. Thompson, Logic of least squares-revisited (1985, preprint).

J.E. Hutton and P.I. Nelson, Quasi-likelihood estimation for semimartingales, Stochastic Processes and their Applications 22 (1986) 245–257.

N.V. Krichagina, R.S. Liptser and E.Y. Rubinovich, Kalman filter for Markov processes, in: N.V. Krylov, R.S. Liptser and A.A. Novikov, eds., Statistics and Control of Stochastic Processes. Steklov Seminar, 1984 (Optimization Software, Inc., New York, 1984) 197-213.

A. Thavaneswaran and M.E. Thompson, Optimal estimation for semimartingales, J. Appl. Prob. 23 (1986) 409-417.

M. West, P. Harrison and S. Migon, Dynamic generalized linear models and Bayesian forecasting, J. Amer. Statist. Assoc. 80 (1985) 73-97.