

Optimal estimation for semimartingale neuronal models

M.K. Habib

Center of Computational Statistics, George Mason University, Fairfax, VA, USA

A. Thavaneswaran

Department of Statistics, University of Manitoba, Winnipeg, Man., Canada

Received March 1988; accepted 20 November 1990

Abstract: The subthreshold behavior of the somal-membrane potential of a single neuron is modeled as a semimartingale. This model extends all Itô-type stochastic neuronal models which have been treated in the literature. The theory of optimal estimating functions is applied to estimate the parameters of the proposed model. This is a likelihood-free method which does not require imposing any distributional assumptions on the noise or driving process. This method includes conditional least squares, quasi-least squares as special cases. Under regularity conditions together with distributional assumptions, it leads to the maximum likelihood method.

AMS Subject Classification: 60G55, 60K99, 62F99.

Key words and phrases: Asymptotic normality; consistency; neuron; neuronal models; neurophysiology; parameter estimation; optimal estimation; semimartingales.

1. Introduction

This paper is concerned with developing semimartingale models of the subthreshold activity of the somal-membrane potential of nerve cells or neurons. These models contain parameters which reflect neuronal properties such as the effective somal-membrane time constant, the potency or amplitudes of post-synaptic potentials (PSPs), and measures of variability of synaptic input. Estimation of these parameters under various experimental conditions is crucial to the study of the ways neurons integrate temporal patterns of synaptic input. The semimartingale models considered here make no distributional assumptions on the noise process and

Correspondence to: Prof. A. Thavaneswaran, Dept. of Statistics, University of Manitoba, Winnipeg, Man., Canada R3T 2N2.

Research supported by the Office of Naval Research under contract numbers N00014-83-K-0387 and N00014-89-J-1502.

extend the Itô-type stochastic neuronal models studied in the literature (see e.g. Johannesma, 1968; Capocelli and Ricciardi, 1971; Ricciardi and Sacerdote, 1979; Tuckwell and Cope, 1980; Habib, 1985). In Lánský (1983) the membrane potential of a single neuron was modeled as a diffusion process observed over an interval $[0, T]$, maximum likelihood estimators of the model's parameters were presented, and the asymptotic properties of these estimators were discussed as $T \rightarrow \infty$. In practice, though, the subthreshold behavior of the membrane potential is observed over a random interval $[0, \tau]$ (when τ is the first passage of the membrane potential of the neuron's threshold). It is experimentally feasible to observe several copies or trajectories of this randomly stopped process (Ferster, 1987). Habib (1985) modeled the membrane potential of a neuron as a randomly stopped diffusion process and studied the asymptotic properties of the maximum likelihood estimators as the number of observed independent copies of the process increases to infinity. The theory of maximum likelihood estimation (m.l.e.) 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 model or another; for example, the considerable literature on estimation for birth-death processes. Feigin (1976) investigated the asymptotic theory of m.l.e. for diffusion processes. Reviews of inference for stochastic processes can be found in Basawa and Prakasa Rao (1980) and Bellach (1980; 1983).

In the present paper we are concerned with a more general semimartingale neuronal model and the study of the estimates of the model's parameter using a likelihood free method. The semimartingale models proposed in this study include Poisson driven models, diffusion models and Itô-Markov (or mixed) models. Furthermore, in order to avoid imposing restrictive as well as distributional assumptions on the driving (or noise) process that are difficult to verify, the theory of optimal estimating functions is employed. This method has been considered by Godambe (1960, 1985), Thavaneswaran and Thompson (1986) and Hutten and Nelson (1986). Heyde (1992) gives an elegant and unified treatment of the theories of optimal estimating functions, least-squares and maximum likelihood. See also Godambe and Heyde (1987).

The objective then of the methods presented in this paper is to provide neuroscientists with quantitative means in order to estimate reliably physiologically meaningful parameters of appropriate stochastic neuronal models. For instance, a careful analysis of the estimated parameters obtained under different experimental conditions should enable the experimental neuroscientist to draw inference concerning the ways these parameters change in response to experience. This may shed light on some aspects of the mechanisms involved in neuronal plasticity in response to natural and experimentally controlled experience.

In Section 2, several stochastic differential equation models that describe the behavior of the somal-membrane potential of a neuron are considered. It is shown that all the stochastic neuronal models treated in the literature are special cases of the semimartingale neuronal model considered in this paper. In Section 3, the

method of optimal estimating functions is used to obtain the estimates of the parameters of interest. In Section 4, the asymptotic properties such as strong consistency and normality of these estimates are derived.

2. Temporal stochastic neuronal models

For a brief description of the morphology of nerve cells (or neurons) see Wegman and Habib (1990). Now assume that the state of the neuron is characterized by the difference in potential across its membrane near a spatially restricted area of the soma called the initial segment (trigger zone or spike initiation region). The membrane potential $V(t)$, at any point in time t , is modeled as a stochastic process defined on a probability space (Ω, F, P) . The membrane potential is subject to instantaneous changes due to the occurrence of post-synaptic potentials (PSP) of two different types: excitatory post-synaptic potentials (EPSP) which occur according to mutually independent Poisson processes $P(\lambda_k^e; t)$ with rates λ_k^e ($k = 1, 2, \dots, n_1$), each accompanied by an instantaneous displacement of $V(t)$ by a constant amount $\alpha_k^e > 0$ ($k = 1, 2, \dots, n_1$); and inhibitory post-synaptic potentials (IPSP) which occur according to independent Poisson processes $P(\lambda_k^i; t)$ with potential displacement $\alpha_k^i > 0$ ($k = 1, 2, \dots, n_2$). Between PSPs, $V(t)$ decays exponentially to a resting potential. The PSPs sum linearly at the trigger zone, and when $V(t)$ reaches a certain (constant) level S , called the neuron's threshold, an action (spike) takes place. Following the action potential, $V(t)$ is reset to the resting potential with a time constant ϱ^{-1} . Based on this physical model and considering n_1 excitatory synapses and n_2 inhibitory ones, the membrane potential $V(t)$ is modeled as a solution of the stochastic differential equation

$$dV(t) = -\varrho V(t) dt + \sum_{k=1}^{n_1} \alpha_k^e dP(\lambda_k^e; t) - \sum_{k=1}^{n_2} \alpha_k^i dP(\lambda_k^i; t), \quad (2.1)$$

$V(0) = V_0$. Under certain conditions the solution of (2.1) is a homogeneous Markov process with discontinuous sample paths. This model is known as Stein's model (Stein, 1965) and is a special case of the well known Poisson driven Markov process models (see Snyder, 1975).

Diffusion models in which the discontinuities of $V(t)$ are smoothed out have been sought as approximations to the discontinuous model (2.1) (see e.g. Ricciardi, 1976). The types of approximation are justified on the grounds that for many types of neurons in the central nervous system, synapses are densely packed along the dendritic tree. For example, there exists on the order of 20 000 synapses on the surface of a typical motor neuron. As a result of this extensive input, the jumps of $V(t)$ are small and the rates of occurrence of the post-synaptic potentials are very large. In this case the approximation of the Poisson driven Markov model by a diffusion model is appropriate and is accomplished by allowing the amplitudes α^e, α^i to tend to zero and the frequencies λ^e, λ^i to become large in a certain manner. The ac-

curacy of the diffusion approximation (and its use in studies of interspike interval calculations) is discussed by Tuckwell and Cope (1980). From a practical point of view, the diffusion model is of greater utility to experimentalists who are not interested in the individual characteristics and the functional impact of every individual synapse of the neuron, but rather in the collective behavior of these synapses. It should be noted, though, that diffusion models are inadequate for describing the membrane potential which exhibits frequent action potentials with large amplitudes. This requires more general models, namely, Itô–Markov models which are driven by Wiener as well as point processes. Semimartingale models include Itô–Markov models as a special case.

Kallianpur (1983) established a diffusion approximation for model (2.1) using the functional central limit theorem for semimartingales of Liptser and Shirayev (1980, 1981). Under some regularity conditions it was shown that model (2.1) can be approximated by the diffusion model

$$dV(t) = (-\varrho V(t) + \mu) dt + \sigma dW(t), \quad 0 \leq t \leq T, \quad (2.2)$$

$V(0) = V_0$, where W is the standard Wiener process (or Brownian motion), i.e., $W(0) = 0$, the sample paths of W are continuous, and for $0 < t_1 < t_2 < \dots < t_{n-1} < t_n$, the increments

$$W(t_1), W(t_2) - W(t_1), \dots, W(t_n) - W(t_{n-1})$$

are independent and normally distributed, with mean zero and variance $t_1, t_2 - t_1, \dots, t_n - t_{n-1}$, respectively. Given n independent trajectories $\{V_k(t), 0 \leq t \leq \tau_k\}$, $k = 1, 2, \dots, n$, $\tau_1, \tau_2, \dots, \tau_n$ are independent random variables with $P(\tau_k < \infty) = 1$, $k = 1, 2, \dots, n$. Habib (1985) derived maximum likelihood estimators of the parameters ϱ and μ and established these large sample properties such as strong consistency and asymptotic normality assuming σ is known. This assumption about σ is not a serious restriction since by using Levy's characterization it can be determined nonparametrically. It should be noted that large sample properties of maximum likelihood estimators of drift parameters of diffusion processes have been extensively discussed in the literature for the case of observing one trajectory over time interval $0 \leq t \leq T$ as $T \rightarrow \infty$. (See e.g. Brown and Hewitt, 1975, Feigin, 1976, Basawa and Prakasa Rao, 1980.) For an extensive review of such problems see Bellach (1980).

2.1. An Itô–Markov neuronal model

The diffusion model (2.2) describes the subthreshold behavior of the membrane potential of neurons with extensive synaptic input and post-synaptic potentials (PSPs) with relatively small amplitudes. It is also assumed that there are no PSPs with large dominating amplitudes. The diffusion model is thus appropriate for describing the subthreshold activity of the membrane potential of the neuron under study when it is experiencing spontaneous activity (see e.g. Favella et al., 1982, and

Lánský, 1983). These models are inadequate for describing the membrane potential while a neuron is driven by an external stimulus, since in this case the neuron receives effective synaptic input through a limited number of synapses (Tanaka, 1983) in addition to the diffusion input. We, therefore, consider a stochastic neuronal model which takes into account the influence of extensive low amplitude synaptic input as well as PSPs with large amplitudes, which may be reflecting the influence of a number of dominating synapses. These synapses may be electronically close to the initial segment. The activity of these synapses will be modeled by a linear combination of independent point processes. That is, the model is driven by diffusion as well as point processes. This mixed model is a special case of a well known class of stochastic processes called Itô–Markov processes (see Ikeda and Watanabe, 1981). Now assume that in addition to the extensive synaptic input leading to the diffusion model (2.2), there are n_1 EPSPs arriving according to independent point processes $N(\lambda_k^e(t), t)$ with random intensities $\lambda_k^e(t)$, and EPSP amplitudes α_k^e , $k = 1, 2, \dots, n_1$. In addition, IPSPs are arriving according to the independent processes $N(\lambda_k^i(t), t)$, with the corresponding parameters $\lambda_k^i(t)$ and α_k^i , $k = 1, 2, \dots, n_2$. Now consider the following extended mixed model to describe the membrane potential of a stimulus driven neuron:

$$\begin{aligned} dV(t) = & (-\varrho V(t) + \mu) dt + \sigma dW(t) \\ & + \sum_{k=1}^{n_1} \alpha_k^e dN(\lambda_k^e(t), t) - \sum_{k=1}^{n_2} \alpha_k^i dN(\lambda_k^i(t), t). \end{aligned} \quad (2.3)$$

A possible physiological interpretation of this model may be as follows. A relatively small number of pre-synaptic neurons are activated as a result of the presentation of a certain stimulus to the receptive field of the post-synaptic neuron. In the presence of an effective stimulus, a limited number of pre-synaptic neurons will fire in response to the stimulus, while the rest of the pre-synaptic neurons are firing spontaneously. Assume n_1 (stimulus activated) excitatory synapses and n_2 (stimulus activated) inhibitory synapses. The input at the excitatory (inhibitory) synapses arrives according to independent Poisson processes with amplitudes α^e (α^i) and rates λ^e (λ^i). The subthreshold potential, $V(t)$, of the post-synaptic neuron is modeled in this case by the stochastic differential equation (2.3). In the absence of an effective stimulus, the rates of the Poisson processes will be small, and hence the terms representing the Poisson input will drop from the model. In this case, model (2.3) reduces to (2.2).

In this paper, the membrane potential of a single neuron is modeled as a semimartingale driven by a locally square-integrable martingale having right-continuous trajectories with finite left hand limits. Because of the absence of distributional assumptions on the driving process, likelihood-free estimation methods need to be considered. In this paper the method of optimal estimation is applied (Godambe, 1960; 1985). The following notation will be needed in the sequel. Let $(\Omega, \mathcal{F}, P, \{F_t\}_{t \geq 0}, P)$ be a complete filtered probability space; i.e. (Ω, \mathcal{F}, P) is a prob-

ability space, $\{F_t, t \geq 0\}$ is a non-decreasing family of sub- σ -fields of F , and F_t is augmented by the sets of F having zero P -measure for all t . Also assume that $\{F_t\}$ is right-continuous, i.e. $F_t = F_{t+}$ for all t when $F_t = F_{t+} = \bigcap_{s \geq t} F_s$. Let D be the measurable space of all right-continuous functions defined on the real line, $x = \{x(t), t \geq 0\}$, having left-hand limits endowed with the Skorokhod topology. A stochastic process $X = \{X(t), t \geq 0\}$ is said to be adapted to a filtration $F = \{F_t, t \geq 0\}$, if $X(t)$ is F_t -measurable for each $t \geq 0$. The notation $(X(t), F_t)$ will be used to indicate that $X(t)$ is F_t -measurable for each $t \geq 0$.

A stochastic process $\{M(t), F_t, t \geq 0\}$ is called a *martingale* if the following two conditions hold:

- (i) $E|M(t)| < \infty, \quad t \geq 0$;
- (ii) $E(M(t) | F_s) = M(s), \quad s < t$.

A local martingale $M = (M(t), F_t)$ is an adapted process for which one can find a sequence of stopping times $\{\tau_n, n \geq 1\}$ increasing to infinity such that the stopped processes $M^n = (M(t \wedge \tau_n), F_t), n \geq 1$, are martingales. Let $P(F)$ be the smallest σ -field generated by F -adapted processes with continuous trajectories. An F -adapted process $\{X(t), t \geq 0\}$ is said to be predictable if the mapping $(t, \omega) \rightarrow X(t, \omega)$ is measurable with respect to the σ -field $P(F)$. (Processes with left (hand) continuous trajectories are predictable processes.)

Let $M(f), M_{\text{loc}}(F, P), M_{\text{loc}}^c(F, P), M_{\text{loc}}^d(F, P), M_{\text{loc}}^2(F, P)$ denote the classes of uniformly integrable, local, continuously local, purely discontinuous local, and locally square-integrable martingales $X = \{X(t), F_t, t \geq 0\}$, respectively. Next, denote by $B_{\text{loc}}(F, P)$ the class of stochastic processes $B = (B(t), F_t)$ which have bounded variation on each finite time interval. Let $E_{\text{loc}}(E, P)$ be a class of locally integrable processes $E = (E(t), F_t)$. Note that each predictable process of class $B_{\text{loc}}(F, P)$ belongs to $E_{\text{loc}}(F, P)$. If $E \in E_{\text{loc}}(F, P)$, there exists a predictable process $\bar{E} = (\bar{E}(t), F_t)$ such that $M_{\text{loc}}(F, P)$. The process \bar{E} is called the *compensator* of the process E .

Each local martingale M admits the decomposition $M = M^c + M^d$, where M^c is a continuous local martingale and M^d is a purely discontinuous local martingale. If M is a square-integrable martingale, then M^2 is a submartingale, and according to the Doob-Myer decomposition there exists a predictable increasing process, denoted by $\langle M \rangle$ (or $\rangle M; M \rangle$), which is called the quadratic characteristic or the variance process of the martingale M . $\langle M \rangle$ may be calculated from the formula

$$\langle M \rangle_t = \lim_{n \rightarrow \infty} \sum_{k=0}^{n-1} E[(M_{t(k+1)/n} - M_{tk/n})^2 | F_{tk/n}].$$

A stochastic process $X = (X(t), F_t)$ is said to be a semimartingale if it can be represented in the form

$$X(t) = V(t) + M(t), \quad t \geq 0, \quad (2.4)$$

where $V = (V(t), F_t) \in B_{\text{loc}}(F, P)$ and $M = (M(t), F_t) \in M_{\text{loc}}(f, P)$. A semimartingale $X = \{X(t), F_t, t \geq 0\}$ also has the canonical representation

$$X(t) = E(t) + M_t^c + \int_0^t \int_{0 < |x| < 1} x \mu(ds, dx) + \int_0^t \int_{0 < |x| \leq 1} x(\mu - \nu)(ds, dx), \quad (2.5)$$

where $E = (E(t), F_t) \in E_{\text{loc}}(P, F)$, $M^c = (M^c(t), F_t) \in M_{\text{loc}}^c(P, F)$, $\mu(dt, dx)$ is the integral random measure of jumps of the semimartingale X ,

$$\mu\left(0, t, \Gamma = \sum_{0 < s \leq t} I(\Delta X_s \in \Gamma)\right), \quad \Gamma \in \mathcal{B}(\mathbb{R} \setminus \{0\}),$$

$\Delta X(s) = X(s) - X(s-)$ and $\nu(dt, dx)$ is the compensator or the dual predictable projection of the μ . See Shirayayev (1981) for a more detailed discussion of the canonical representation (2.5).

It is clear that the stochastic neuronal models (2.1)–(2.3) are special cases of the semimartingale models (2.4) and (2.5). This demonstrates the relevance of problems of parameter estimation for certain semimartingale models to stochastic neuronal modeling and analysis.

3. Optimal estimation

In this section the theory of optimal estimating functions is applied to estimate parameters of a certain class of semimartingale models which contains the neuronal models (2.1)–(2.3) as special cases. The method of optimal estimation is a likelihood-free method which does not require specific distributional assumptions of the driving or noise process. In this sense it is a more robust method than that of the maximum likelihood procedure. Furthermore, the method of optimal estimation is applicable in certain situations where the method of maximum likelihood estimation may fail. For example, if the diffusion coefficient α in model (2.2) depends on the unknown parameter ϱ or μ , then it is easy to argue that the probability measures P_V and P_W induced (on the appropriate measurable space) by the processes V and W , respectively, are singular and hence the corresponding Radon–Nykodým derivative or the likelihood process (function) does not exist. In this case the usual method of maximum likelihood estimation is unapplicable (see Hutten and Nelson, 1986). It will be shown below that the optimal estimator exists in this case and that it is strongly consistent.

In order to motivate the method of optimal estimation consider the following simple example: let θ be a parameter, X is a random or (vector valued random) variable with probability density function $f(X; \theta)$. Consider the likelihood function $l(\theta; X) = \ln f(X; \theta)$. If f is assumed to be continuously differentiable with respect to θ , the score function S is defined by

$$S(\theta; X) = \frac{\partial l}{\partial \theta}(\theta; X), \quad (3.1)$$

and under the appropriate regularity conditions, the maximum likelihood estimate (MLE) of θ is given by the solution of equation

$$S(\theta; X) = 0. \quad (3.2)$$

An estimating function, $G(\theta, X)$, is a function of the parameter θ and observations X . A typical example is the score function $S(\theta; X)$ in (3.1). An estimating equation for the parameter θ is obtained by setting the estimating function $G(x, \theta)$ equal to zero:

$$G(\theta, x) = 0, \quad (3.3)$$

which is solved for θ in terms of the data x to obtain the estimate. An estimating function G which satisfies the condition $E[G(\theta, x)] = 0$ for all θ is called an unbiased estimating function. The unbiasedness of the estimating function leads to consistent estimator. The class of such functions is denoted by L . An estimating function is said to be optimum in L if it minimizes the quotient

$$E \left[G(\theta, x) / E \left(\frac{\partial G}{\partial \theta}(\theta, x) \right) \right]^2, \quad (3.4)$$

for all $G \in L$ and θ (Godambe, 1960). This optimality criterion was motivated by interpreting the numerator $E_{x/\theta}(G^2)$ as the variable of G , and the denominator $\{E_{x/\theta}(\partial G/\partial \theta)\}^2$ as the sensitivity of the estimating function. Minimizing (3.4), therefore, implies a small bias for the estimate. Godambe (1960) proved that the score function is the optimal estimating function within the class of all unbiased estimating functions in the above sense. For further motivation of the optimality based on efficiency consideration see Lindsey (1985). Finally, to motivate the specific choice of estimating functions which will be considered in the sequel in this paper, consider the following example of diffusion model:

$$dx_t = f(t, x_t, \theta) dt + dW_t, \quad 0 \leq t \leq T.$$

The score function in this case is given by

$$S_T(\theta, x) = \int_0^T \frac{\partial}{\partial \theta} f(t, x_t, \theta) dW_t, \quad (3.5)$$

under the assumption that f is differentiable with respect to θ . If f is assumed to be predictable, then it follows, from the well known fact that the integral of a predictable process with respect to a martingale (namely $W(t)$) is itself a martingale, that $S_T(\theta, x)$ is a martingale. A similar argument can also be given in the counting process case. This motivates the choice of an estimating function as an integral of predictable process with respect to martingale, in addition to the consideration that this class is a cts time version of Godambe (1985). It is also easy to show that the above class of estimating functions contains the estimating function obtained by quasi-least squares as a special case (cf. Christopeit, 1986).

Godambe's optimality criterion. Following Godambe (1985), consider parameter θ to be a function of $P \in \mathcal{P}$ (a family of probability measures). Let $G(V, \theta) = (G_t(V, \theta), F_t; t \geq 0)$ represent a family of processes indexed by θ such that

$E_P G_t(V, \theta) = 0$ for exact $t \geq 0$, for each P and $\theta = \theta(P)$. This corresponds to the unbiasedness property of Godambe (1960). Godambe's optimality criteria, adapted to this situation, reads G^0 is optimal in L , the class of unbiased estimating functions, if $Q = A_h - A_h^0$ is non-negative definite for all $G \in \mathcal{L}$ and for all $P \in \mathcal{P}$, where

$$h(V) = E \left[\frac{\partial G}{\partial \theta} \right] G(V, \theta),$$

and

$$h^0(v) = E \left[\frac{\partial G^0}{\partial \theta} \right]^{-1} G^0(V, \theta),$$

and A is the variance-covariance matrix for h under θ_0 (the true value of θ). The following sufficient condition for optimality to hold is due to M.E. Thompson.

Lemma 3.1. G^0 is optimal in L if

$$E \left[\frac{\partial G}{\partial \theta} \right] = K E [G G^0]$$

for all $G \in \mathcal{L}$.

Proof (cf. Thavaneswaran (1985), p. 57). Let, for each P_θ^τ , the observed process $\{V(t), 0 \leq t \leq \tau\}$, where τ is in $f\{t: V(t) \geq S\}$ (for a constant S representing the neuron's threshold), be a semimartingale of the form

$$V_t = \int_0^t f_{s,\theta} d\lambda_s + H_{t,\theta}, \quad 0 \leq t \leq \tau, \quad \theta \in \mathbb{R}^p, \quad (3.6)$$

where

- (i) $\{\lambda_t\}$ is a real monotonic-nondecreasing, right-continuous process with $\lambda(0) = 0$,
- (ii) $\{H_{t,\theta}, F_t^V\}$ is a càdlàg, locally square-integrable martingale with a predictable variation (process)

$$\langle H \rangle_{t,\theta} = \int_0^t b_{s,\theta} d\lambda_s,$$

and the processes $\{\lambda_t\}$, $\{b_{t,\theta}\}$ and $\{f_{t,\theta}\}$ are predictable, where $\{F_t^V\}$ denotes the natural filtration, i.e. $F_t^V = \sigma\{V_s, s \leq t\}$. Furthermore, assume that the elements of $f_{t,\theta}$ are continuously differentiable. The objective here is to estimate θ based on a realization $\{V_t, 0 < t \leq \tau\}$ where τ is a random variable. Because of the use of the natural filtration $\{F_t^V\}$, the processes $\{f_{t,\theta}\}$ and $\{b_{t,\theta}\}$ are observable up to knowing the true value of θ .

Now based on the discussion following (3.5), it is natural to look for p -dimensional estimating functions of the form

$$G_{t,\theta} = \int_0^t a_{s,\theta} dH_{s,\theta}, \quad (3.7)$$

generated by a predictable process $\{a_{s,\theta}\}$. Then, using the properties of Itô integrals,

$$E[G^{0'}G] = E \int_0^t a_{s,\theta} d\langle H \rangle_{s,\theta} a_{s,\theta}^{0'},$$

where a' denotes the transpose of a , and

$$G_\theta^0 = \int_0^t a_{s,\theta}^0 dH_{s,\theta}.$$

Hence,

$$E[G_0 G_\theta^{0'}] = E \int_0^t a_{s,\theta} b_{s,\theta} \alpha'_{s,\theta} d\lambda_s.$$

Moreover,

$$E \left[\frac{\partial G_\theta}{\partial \theta} \right] = E \int_0^\tau a_{s,\theta} \frac{\partial}{\partial \theta} (dH_{s,\theta}) = -E \int_0^\tau a_{s,\theta} f'_{s,\theta} d\lambda_s, \quad (3.8)$$

where $f'_{s,\theta} = (\partial/\partial\theta)f_{s,\theta}$. It follows that the optimal estimating function is given by

$$G_\theta^0 = \int_0^\tau f'_{s,\theta} b_{s,\theta}^+ dH_{s,\theta}, \quad (3.9)$$

provided that $b_{s,\theta}^+$, the inverse of $b_{s,\theta}$, exists.

Example 3.1. In this example, model (2.2) is extended by replacing the Wiener process $\{W_t\}$ in (2.2) by a square-integrable martingale $\{M_t\}$:

$$\begin{aligned} dV_t &= (-\varrho V_t + \mu) dt + \sigma dM_t \\ &= (-\varrho, \mu) \begin{bmatrix} V_t \\ 1 \end{bmatrix} dt + \sigma dM_t = \theta' f(V_t) dt + \sigma dM_t, \end{aligned} \quad (3.10)$$

where

$$\theta = \begin{bmatrix} -\varrho \\ \mu \end{bmatrix}, \quad f(V_t) = \begin{bmatrix} V_t \\ 1 \end{bmatrix}, \quad \langle M \rangle_t = t, \quad \text{and} \quad \sigma > 0$$

(assumed to be known). The optimal estimating function is given by

$$G_{n,\theta}^0 = \sum_{k=1}^n \int_0^{\tau_k} \begin{bmatrix} V^{(k)} \\ 1 \end{bmatrix} dM_s^{(k)} \quad \text{a.s.} \quad (3.11)$$

Note the following:

(i) (3.11) is the same as the m.l.e. equation in Habib (1985), where it is assumed that $M_t = W_t$ (the standard Wiener process). Therefore, it is clear that without imposing any distributional assumption on the noise (driving process) M , the method of optimal estimation allows for the estimation of the unknown parameter under the assumption that $\langle M \rangle_t = t$;

(ii) M_t may be a purely discontinuous square-integrable martingale with $\langle M \rangle_t = t$, e.g. $M_t = N_t - t$ (where N_t is the standard Poisson process). This shows that model (3.7) also extends model (2.1).

4. Asymptotic properties of the optimal estimators

In this section asymptotic properties such as strong consistency and asymptotic normality of the optimal estimator θ_n^0 of θ (c.f. model (3.6)) are based on observing n trajectories $\{V_t^{(k)}, \tau_{k-1} \leq t \leq \tau_k\}$, where τ_k are stopping times, $k = 1, 2, \dots, n$, and $\tau_0 = 0$ a.s. Assume throughout this section that $f_{t,\theta} = f(V_t)\theta$. In this case model (3.6) takes the form

$$dV_t^{(k)} = f(V_t^{(k)})\theta d\lambda_s + dH_{s,\theta}^{(k)}, \quad \tau_{k-1} \leq t \leq \tau_k, \quad (4.1)$$

where $k = 1, 2, \dots, n$, and $H^{(k)}$ are n square-integrable martingales.

Now, from (3.9), it follows that the optimal estimation function is given by

$$G_{n,\theta}^0 = \sum_{k=1}^n \int_{\tau_{k-1}}^{\tau_k} f^{0'}(V_s^{(k)}) b_{s,\theta}^+ dH_{s,\theta}^{(k)},$$

and the optimal estimator θ_n^0 satisfies the equation

$$\sum_{k=1}^n \int_{\tau_{k-1}}^{\tau_k} f'(V_s^{(k)}) b_{s,\theta}^+ dV_s^{(k)} = \sum_{k=1}^n \int_{\tau_{k-1}}^{\tau_k} f'(V_s^{(k)}) b_{s,\theta}^+ f(V_s^{(k)}) \theta d\lambda_s.$$

Hence θ_n^0 is given by

$$\theta_n^0 = \left(\sum_{k=1}^n \int_{\tau_{k-1}}^{\tau_k} f'(V_s^{(k)}) b_{s,\theta}^+ f(V_s^{(k)}) d\lambda_s \right)^{-1} \left(\sum_{k=1}^n \int_{\tau_{k-1}}^{\tau_k} f'(V_s^{(k)}) b_{s,\theta}^+ dV_s^{(k)} \right). \quad (4.2)$$

Theorem 4.1 (Strong consistency). *Let*

$$m_{\tau_k} = \int_{\tau_{k-1}}^{\tau_k} f'(V_s^{(k)}) b_{s,\theta}^+ dH_{s,\theta}^{(k)}. \quad (4.3)$$

Assume that

$$M_n = \left(\sum_{k=1}^n \langle m \rangle_{\tau_k} \right) \rightarrow \infty \quad \text{as } n \rightarrow \infty,$$

where

$$\langle m \rangle_{\tau_k} = \int_{\tau_{k-1}}^{\tau_k} f'(V_s^{(k)}) b_{s,\theta}^+ f(V_s^{(k)}) d\lambda_s.$$

Then

$$\theta_n^0 \rightarrow \theta \quad \text{a.s. as } n \rightarrow \infty.$$

Proof. From (4.1) and (4.2) it follows that

$$\begin{aligned} \theta_n^0 &= \theta + \left(\sum_{k=1}^n \int_{\tau_{k-1}}^{\tau_k} f'(V_s^{(k)}) b_{s,\theta}^+ f(V_s^{(k)}) d\lambda_s \right)^{-1} \left(\sum_{k=1}^n \int_{\tau_{k-1}}^{\tau_k} f'(V_s^{(k)}) b_{s,\theta}^+ dH_{s,\theta}^{(k)} \right) \\ &= \theta + \left(\sum_{k=1}^n \langle m \rangle_{\tau_k} \right)^{-1} \left(\sum_{k=1}^n m_{\tau_k} \right), \end{aligned}$$

and $\langle M \rangle_{\tau_k}$ is given by (4.3). Now, let $F_{\tau_k} = \sigma\{A \in F_\infty : A \cap [\tau_k \leq t] \in F_t\}$ for all $t \geq 0$, where $F_\infty = \sigma(\cup_{t \geq 0} F_t)$. From the optional stopping theorem (Elliot, 1982, p. 22), it follows that $\{m_{\tau_k}, F_{\tau_k}\}$, $k = 1, 2, \dots, n$, is a martingale.

Since $\sum_{k=1}^n \langle m \rangle_{\tau_k} \rightarrow \infty$ as $n \rightarrow \infty$,

$$\left(\sum_{k=1}^n \langle m \rangle_{\tau_k} \right)^{-1} \left(\sum_{k=1}^n m_{\tau_k} \right) \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

by using the strong law of large numbers for martingales (see e.g. Shiryaev, 1984, p. 487). This concludes the proof of the theorem. \square

Next, the asymptotic normality of the optimal estimator θ_n^0 of θ will be established. Let

$$R_n = \text{var}(M_n) = E \langle m \rangle_n.$$

A Cramér-Wald argument will be used to prove $R_n^{-1/2} M_n \xrightarrow{d} N(0, I)$, i.e. for a p -dimensional vector λ of real number, it is sufficient to prove that

$$\lambda' R_n^{-1/2} M_n \xrightarrow{d} N(0, \lambda' \lambda).$$

Consider

$$\lambda' R_n^{-1/2} M_n = \sum_{k=1}^n \xi_{\tau_k n},$$

where

$$\xi_{\tau_k n} = \sum_{k=1}^n \lambda' R_n^{1/2} m_{\tau_k n}.$$

Further assume that the following three conditions are satisfied:

$$(A.1) \quad \max_{1 \leq k \leq n} |m_{\tau_k n}| \xrightarrow{p} 0,$$

$$(A.2) \quad \sum_{k=1}^n m_{\tau_k n}^2 \xrightarrow{p} \lambda' \lambda,$$

and

$$(A.3) \quad E \left(\max_{1 \leq k \leq n} m_{\tau_k n}^2 \right) \text{ is bounded in } n.$$

Then, using the martingale central limit theorem of Hall and Heyde (1980, Theorem 3.2) along with the Cramér-Wald device, the following result on the asymptotic normality of θ_n^0 follows:

Theorem 4.2 (Asymptotic normality). *Under conditions (A.1)–(A.3), and assuming that $R_n^{-1} \langle m \rangle_n \xrightarrow{p} N(0, I)$, it follows that*

$$(a) \quad R_n^{-1/2} M_n \xrightarrow{d} N(0, I_p),$$

$$(b) \quad R_n^{+1/2} (\theta_n^0 - \theta) \xrightarrow{d} N(0, I).$$

Note: a set of easily verifiable necessary conditions for (A.1)–(A.3) together with an application to nonlinear time series models can be found in Tjøstheim (1986).

Acknowledgements

We are grateful to Dr. Charles Smith for reading the manuscript and offering several useful remarks.

References

- Basawa, I.V. and B.L.S. Prakasa Rao (1980). *Statistical Inference for Stochastic Processes*. Academic Press, New York.
- Bellach, B. (1980). Consistency, asymptotic normality and asymptotic efficiency of the maximum-likelihood estimator in linear stochastic differential equations. *Math. Operationsforsch. Statist. Ser. Statistics* **11**, 227–266.
- Bellach, B. (1983). Parameter estimators in linear stochastic differential equations and their asymptotic properties. *Math. Operationsforsch. Statist. Ser. Statistics* **14**, 141–191.
- Brown, B.M. and J.I. Hewitt (1975). Asymptotic likelihood theory for diffusion processes. *J. Appl. Probab.* **12**, 228–238.
- Christopeit, N. (1986). Quasi-least-squares estimation in semimartingale regression models. *Stochastic* **16**, 255–278.
- Capocelli, R.M. and L.M. Ricciardi (1971). Diffusion approximation and first passage time problem for a model neuron. *Kybernetik* **8**, 214–223.
- Elliott, R.J. (1982). *Stochastic Calculus and Applications*. Springer, New York.
- Favella, L. et al. (1982). First passage time problems and related computational models. *Cybernet. Syst. Ser.* **13**, 95–178.
- Feigin, P.D. (1976). Maximum likelihood estimation for continuous-time stochastic processes. *Adv. in Appl. Probab.* **8**, 712–736.
- Ferster, D. (1987). Origin of orientation-selective EPSPs in neurons of cat visual cortex. *J. Neurosci.* **7**, 1780–1791.
- Godambe, V.P. (1960). An optimum property of regular maximum likelihood equation. *Ann. Math. Statist.* **31**, 1208–1211.
- Godambe, V.P. (1985). The foundations of finite sample estimation in stochastic processes. *Biometrika* **12**, 419–428.
- Godambe, V.P. and C.C. Heyde (1987). Quasi-likelihood and optimal estimation. *Internat. Statist. Rev.* **55**, 231–244.
- Habib, M.K. (1985). Parameter estimation for randomly stopped processes and neuronal modeling. UNC Institute of Statistics, Mimeo Series No. 1492.
- Hall, P. and C.C. Heyde (1980). *Martingale Limit Theory and its Application*. Academic Press, New York.
- Heyde, C.C. (1992). New developments in inference for temporal stochastic processes. *J. Statist. Plann. Inference* **33**, 121–129 (this issue).
- Hutten, J.E. and P.I. Nelson (1986). Quasi-likelihood estimation for semimartingales. *Stochastic Process. Appl.* **22**, 245–257.
- Ikeda, N. and S. Watanabe (1981). *Stochastic Differential Equations and Diffusion Processes*. North-Holland, Amsterdam.

- Johannesma, P.I.M. (1968). Diffusion models for the stochastic activity of neurons. In: E.R. Caianello, Ed., *Neuronal Networks*. Springer, New York.
- Kallianpur, G. (1983). On the diffusion approximation to a discontinuous model for a single neuron. In: P.K. Sen, Ed., *Contributions to Statistics: Essays in Honor of Norman L. Johnson*. North-Holland, Amsterdam.
- Kandel, E.R. and J.H. Schwartz (1981). *Principles of Neural Science*. Elsevier/North-Holland, New York.
- Lánský, P. (1983). Inference for the diffusion models of neuronal activity. *Math. Biosci.* **67**, 247–260.
- Lindsey, B. (1985). Using empirical partially Bayes inference for increased efficiency. *Ann. Statist.* **13**, 914–931.
- Liptser, R.S. and A.N. Shiryaev (1980). A functional central limit theorem for semi-martingales. *Theory Probab. Appl.* **25** (4), 683–703.
- Liptser, R.S. and A.N. Shiryaev (1981). On necessary and sufficient conditions on a central limit theorem for semi-martingales. *Theory Prob. Appl.* **25** (1), 123–137.
- Ricciardi, L.M. (1976). Diffusion approximation for a multi-input model neuron. *Biol. Cybernet.* **24**, 237–240.
- Ricciardi, L.M. and L. Sacerdote (1979). The Örnstein–Uhlenbeck process as a model for neuronal activity. *Biol. Cybernet.* **35**, 1–9.
- Shiryaev, A.N. (1981). Martingales: Recent developments, results and applications. *Internat. Statist. Rev.* **49**, 199–233.
- Shiryaev, A.N. (1984). *Probability*. Springer, New York.
- Snyder, D.L. (1975). *Random Point Processes*. Wiley, New York.
- Stein, R.B. (1965). Some models of neuronal variability. *Biophys. J.* **5**, 173–195.
- Tanaka, K. (1983). Cross-correlation analysis of geniculostriate neuronal relationships in cats. *J. Neurophysiol.* **49**, 1303–1318.
- Tuckwell, H.C. and D.K. Cope (1980). Accuracy of neuronal interspike times calculated from a diffusion approximation. *J. Theoret. Biol.* **83**, 377–387.
- Thavaneswaran, A. (1985). Estimation for Semimartingales. Ph.D. Thesis. University of Waterloo, Canada.
- Thavaneswaran, A. and M.E. Thompson (1986). Optimal estimation for semimartingales. *J. Appl. Probab.* **23**, 409–417.
- Tjøstheim, D. (1986). Estimation in nonlinear time series models. *Stochastic Process. Appl.* **21**, 251–273.
- Wegman, E.J. and M.K. Habib (1990). Stochastic methods of neural systems. *J. Statist. Plann. Inference* **33**, 5–25 (this issue).