

ON BAYESIAN NONPARAMETRIC ESTIMATION FOR STOCHASTIC PROCESSES

by

M.E. Thompson and A. Thavaneswaran

University of Waterloo

and

University of North Carolina, Chapel Hill

Institute of Statistics Mimeo Series No. 1836

August 1987

**ON BAYESIAN NONPARAMETRIC ESTIMATION  
FOR STOCHASTIC PROCESSES**

M.E. Thompson and A.Thavaneswaran

University of Waterloo and University of North Carolina, Chapel Hill

*ABSTRACT*

The problem of Bayesian estimation of a continuous time stochastic 'signal' process, on which the observation process depends linearly, is discussed. It is noted that the situation treated by Ammann (Stochastic Processes and Applications, 1985) is a special case. Considerations involved in the generalization of his Laplace transform functional approach are described.

## 1. Introduction

The most general setting of the problem considered here is that of a stochastic differential equation model

$$dX_t = \eta_t dR_t + dM_{\eta,t}, t \geq 0 \quad (1.1)$$

where  $M_\eta$  is a local martingale,  $X$  and  $R$  are observable processes, and  $\eta$  as a function of  $t$  is the object of inference, the model "parameter". Since this "parameter" is a function and hence infinite dimensional, estimating  $\eta$  is usually thought of as "non-parametric" estimation. Indicating a dependence of  $M$  on  $\eta$  is a way of saying that when the "true" measure of  $X$  is one having a particular  $\eta$  as "parameter" value then  $X_t - \int_0^t \eta_s dR_s$  is a local martingale.

If we view  $(\eta_t, t \geq 0)$  as itself a realization of a stochastic process, then given observations on  $X, R$  on the time interval  $[0, T]$  we can in principle make inference about  $\eta$  via Bayes' theorem.

In this paper we shall discuss the determination of a posterior measure for  $\eta$  via the examination of two special cases. The first is the discrete time analogue of (1.1) when the  $\eta$  sequence or its partial sums are Markovian. The second, based on the work of Ammann (1985), is the case where  $\eta$  has independent increments in continuous time, and the observation process is a point counting process.

In the final section we discuss the considerations which would be involved in extending the Laplace transform functional technique to the more general problem. It should be noted that the purpose of this paper is the survey of the main ingredients of the problem, and that therefore an attempt is made to keep mathematical details to a

minimum, at least until Section 5.

## 2. A Discrete Time Analogue

The discrete time analogue of (1.1) is

$$X_n = X_0 + \sum_{j=1}^n (\Delta R)_j \eta_j + M_n, \quad n = 1, 2, \dots \quad (2.1)$$

where  $M$  is a zero mean martingale, the processes  $X$  and  $\Delta R$  are observable, and the object of inference is the sequence  $\eta$ .

We consider for illustrative purposes two special and relatively tractable models for  $\eta$ , imposing a Markovian dependence first on the 'integral' of  $\eta$  with respect to  $R$ , and second upon  $\eta$  itself.

In the first case, using the 'parameter transformation'

$$\beta_n = \sum_{j=1}^n \eta_j \quad (2.2)$$

we have the model

$$X_n - X_{n-1} = (\Delta R)_n (\beta_n - \beta_{n-1}) + M_n - M_{n-1}$$

and if we assume that the sequence  $\beta$  is a realization of a non-homogeneous Markov chain, and that given  $\eta$  and the past up to  $n-1$  the increment  $M_n - M_{n-1}$  depends at most on  $\beta_{n-1}, \beta_n$ , we have a theoretically tractable inference problem. For the posterior probability function

$$P(\beta_1 = i_1, \dots, \beta_m = i_m \mid X_1 = x_1, \dots, X_N = x_N), \quad m > N$$

is a normalizing constant times

$$\begin{aligned}
 & p_{i_0 i_1}^0 p_{i_1 i_2}^1 \cdots p_{i_{m-1} i_m}^{m-1} \\
 & \times f(x_0) f_{0 i_1}^0(x_1 - i_1 - x_0) f_{i_1 i_2}^1(x_2 - i_2 - x_1 + i_1) \cdots \\
 & f_{i_{N-1} i_N}^{N-1}(x_N - i_N - x_{N-1} + i_{N-1})
 \end{aligned} \tag{2.3}$$

where  $p_{i_k}^j$  denotes a one-step transition probability at time  $j$  and the "likelihood component"  $f_{i_j}$  is the frequency function for the  $j$ th increment of  $M$ . Since therefore in form

$$P(\beta_1 = i_1, \dots, \beta_m = i_m \mid X_1 = x_1, \dots, X_N = x_N) = \tilde{p}_{i_0 i_1}^0 \tilde{p}_{i_1 i_2}^1 \cdots \tilde{p}_{i_{m-1} i_m}^{m-1} \tilde{P} \tag{2.4}$$

where  $\tilde{P}$  is a normalizing constant, it is easily seen that the sequence  $\beta$  is a posteriori a Markov chain (non homogeneous) with modified or updated transition probabilities.

The normalizing constant  $\tilde{P}$  is the  $i_0$ th row sum of a product of  $m$  matrices. If  $m$  or the cardinality of the state space is large the computation of  $\tilde{P}$  and summarization of the posterior measure will be difficult. However,  $\tilde{P}$  need not be computed if we are interested only in comparing the posterior probabilities of sequences  $(i_1, \dots, i_m)$ , given a certain observation  $(x_1, \dots, x_N)$ . Some inroads can be made into the problem of summarization even in large scale problems, where it is possible for example to find the modal sequence  $(i_1, \dots, i_m)$  using 'stochastic relaxation' methods. See Geman and Geman (1984).

In the second case, if we assume that  $\eta$  is *a priori* a realization of a non-homogeneous Markov chain, and that given  $\eta$  and the past up to time  $n - 1$  the increment  $M_n - M_{n-1}$  depends **at most** on  $\eta_{n-1}, \eta_n$ , the form for the posterior distribution of  $\eta$  is analogous to (2.3):

$$\begin{aligned}
 &P(\eta_1 = i_1, \dots, \eta_m = i_m \mid X_1 = x_1, \dots, X_N = x_N; \Delta R_1 = r_1, \dots, \Delta R_N = r_N) \\
 &\quad \propto p_{i_0 i_1}^0 p_{i_1 i_2}^1 \dots p_{i_{m-1} i_m}^{m-1} f(x_0) f_{\alpha_1}(x_1 - x_0 - r_1 i_1) f_{i_1 i_2}(x_2 - x_1 - r_2 i_2) \\
 &\quad \dots f_{i_{N-1} i_N}^{N-1}(x_N - x_{N-1} - r_N i_N). \tag{2.5}
 \end{aligned}$$

The discussion of the practical use of the posterior distribution in the first case clearly applies here also.

In general it seems that the prior and posterior dependence structures for  $\beta$  or  $\eta$  will be of the same order provided the increments of the martingale  $M$  have distributions depending only on sufficiently recent  $\eta$  values.

### 3. Independent increment priors on the process $\eta$ : the Laplace transform functional

In continuous time, corresponding to the two situations discussed in the preceding section, we could consider here two sorts of prior process models for  $\eta$ , corresponding roughly to the cases

(i)  $\int_0^t \eta_s ds, t \geq 0$  is modelled

(ii)  $\eta$  itself is modelled.

In the survival analysis and reliability literature there arise point process models of the form

$$dN_t = \eta_t Z_t dt + dM_t$$

where  $\eta_t Z_t$  is the predictable intensity process,  $Z_t$  being the size of a 'risk set' and  $(\eta_t, t \geq 0)$  the hazard function of a survival distribution.

Some authors have used process priors with independent increments (gamma and Dirichlet processes) to model the cumulative hazard  $\int_0^t \eta_s ds$ , which is the negative of the natural logarithm of the survivor function. See Wild and Kalbfleisch (1981) and the references contained therein. Others such as Dykstra and Laud (1981) and Ammann (1985) have used such process priors to model the hazard function itself.

From the discrete time results we would expect that the two types of models can be dealt with in similar ways, and this is indeed the case. We will here focus mainly on situation (ii), where  $\eta$  itself is modelled, partly to avoid questions about the sense in which a process with discontinuous paths can model the cumulative hazard function or intensity for a point process.

Particularly when the stochastic process  $\eta$  is non-negative, one may try to characterize its associated prior measure via the Laplace transform functional, which is defined for a certain class of functions  $\xi = (\xi(t), t \geq 0)$  by

$$\Phi(\xi) = E \exp \left[ - \int_0^{\infty} \xi(t) d\eta_t \right]. \quad (3.1)$$

From this it is possible to derive as special cases marginal and joint Laplace transforms for the variates  $\eta_t$ , and to use these in turn to derive moments by differentiation. For example, the ordinary Laplace transform for  $\eta_T$  is  $\Phi(\xi)$  where  $\xi(t) = \xi_0$  for  $t \in [0, T]$ ,  $\xi(t) = 0$  for  $t > T$ ,  $\Phi(\xi)$  now being considered as a function of  $\xi_0$ .

As indicated above, Ammann (1985) considers variants of the special case of situation (ii) in which  $\eta = (\eta_t, t \geq 0)$  is a non-negative stochastic process with independent increments. In such a case the Laplace transform functional takes the form

$$\Phi(\xi) = \exp \left\{ -\int_0^{\infty} \xi(t) d\gamma(t) + \int_0^{\infty} \int_{R^+} (e^{-v\xi(t)} - 1) Q(dv, dt) \right\} \quad (3.2)$$

where  $R^+ = (0, \infty)$  and  $Q$  is the Lévy measure of  $\eta$ ; the set function  $Q(B, A)$  can be interpreted as the expected number of jumps of  $\eta$  paths in time set  $A$  having values in  $B$ . For example, if  $\eta$  is a compound Poisson process,

$$Q(dv, dt) = \lambda F(dv) dt \quad (3.3)$$

where  $\lambda$  is the intensity of jumps and  $F$  the distribution function of their values.

Another example, used by Ammann (1985) and Dykstra and Laud (1981), is that of the gamma process, for which

$$Q(dv, dt) = v^{-1} e^{-vA(t)} dv dA(t)$$

where  $b(t) > 0$  is a deterministic "scale function" and  $dA(t)$  is a deterministic locally finite "shape measure". This process has a representation

$$\eta_t = \int_0^t b(s) dZ_s$$

where  $(Z_s, s \geq 0)$  has independent increments and  $Z_t - Z_s$  is distributed as  $\Gamma(A(t) - A(s), 1)$ .

The inference problem in this framework can be stated as follows. Give the "prior" Laplace transform or characteristic functional for  $\eta$ , and given the observation of  $R, X$  on  $[0, T]$ , find the "posterior" or "conditional" Laplace transform or characteristic functional for  $\eta$ , namely

$$\Phi_1(\xi) = E \left[ \exp \left\{ -\int_0^{\infty} \xi(t) d\eta_t \mid X_t, R_t, 0 \leq t \leq T \right\} \right]$$



The filtering problem of "on line" estimation of the current  $\eta$  value can be regarded as the problem of computing  $\Phi_1(\xi)$  where  $\xi(t)$  is a constant  $\xi_0$  on  $[0, T]$  and 0 otherwise. Computation of these or the analogous and more generally applicable characteristic functions is the aim of what Brémaud (1981) calls the "probability reference method" of filtering.

#### 4. Inference from point process data when $\eta$ has independent increments

For a point process, the model of (1.1) can often be specialized to

$$dN_t = \eta_t Z_t dt + dM_{\eta, t} \quad (4.1)$$

where  $N = (N_t, t \geq 0)$  is the point counting process, and  $(\eta_t Z_t, t \geq 0)$  is the (predictable) intensity process. Again, in reliability applications, the observable  $Z_t$  is the size of the "risk set" at time  $t$  and  $\eta$  is the "hazard function" of a survival distribution under investigation.

The form for the "likelihood process" for a point process observed on  $[0, T]$  is well known (see Section 5.2). Let  $(\lambda_t, t \geq 0)$  be a predictable version of the intensity process. If  $N_T = N_0 = 0$ , the likelihood is

$$L_T \propto \exp \left\{ -\int_0^T \lambda_t dt \right\} = \exp \left\{ -\int_0^\infty (T-t)^+ d\lambda_t \right\} \quad (4.2)$$

If  $N_T > 0$ , we have

$$L_T \propto \left[ \prod_{i=1}^{N_T} \lambda_{t_i} \right] \exp \left\{ -\int_0^T \lambda_t dt \right\}$$

where  $t_1, \dots, t_{N_T}$  are the times of increase of  $N$  on  $[0, T]$ . Thus if  $N_T > 0$ ,

$$L_T^\alpha \frac{\partial^{N_T}}{\partial \theta_1 \dots \partial \theta_{N_T}} \Big|_{\theta=0} \exp \left\{ \sum_{i=1}^{N_T} \theta_i \int_0^\alpha I(t \leq t_i) d\lambda_t - \int_0^\infty (T-t)^+ d\lambda_t \right\} \quad (4.3)$$

where  $I(t \leq t_i)$  denotes the indicator of its argument as a function of  $t$ . If  $(Z_t, t \geq 0)$  of (4.1) is a (predictable) pure jump process, like the size of the risk set or a process which is identically 1, we can set

$$d\lambda_t = Z_t d\eta_t$$

above. We assume that this is so.

A form of Bayes' theorem (Section 5.1) applied to (3.1) and (4.2) implies that the conditional or posterior Laplace transform function is, if  $N_T = 0$ ,

$$\begin{aligned} \Phi_1(\xi) &= E \left[ \exp \left\{ - \int_0^\infty \xi(t) d\eta_t \mid N_T, Z_t, 0 \leq t \leq T \right\} \right] \\ &= \Phi(\xi + \xi_L) \mid \Phi(\xi_L) \end{aligned}$$

where  $\xi_L(t) = (T-t)^+ Z_t$ . If  $N_T > 0$ , we similarly use (3.1) and (4.3) to obtain

$$\Phi_1(\xi) = \left\{ \frac{\partial^{N_T}}{\partial \theta_1 \dots \partial \theta_{N_T}} \Big|_{\theta=0} \Phi(\xi + \xi_L) \right\} \div \left\{ \frac{\partial^{N_T}}{\partial \theta_1 \dots \partial \theta_{N_T}} \Big|_{\theta=0} \Phi(\xi_L) \right\}$$

where  $\xi_L(t) = \left[ - \sum_{i=1}^{N_T} \theta_i I(t \leq t_i) + (T-t)^+ \right] Z_t$ .

From these formulas the expressions obtained by Ammann (1985) can be derived. In particular, if  $\eta$  is *a priori* non-negative with independent increments and  $\gamma(t)$  of (3.2) is identically 0, then for  $N_T = 0$ ,

$$\Phi_1(\xi) = \exp \left\{ \int_0^\infty \int_{R^+} (e^{-v\xi(t)} - 1) e^{-v\xi_L(t)} Q(dv, dt) \right\}.$$

That is, conditionally,  $(\eta_t, t \geq 0)$  is an independent increment process with Levy

measure

$$Q_1(d\nu, dt) = e^{-\nu\xi_L(t)} Q(d\nu, dt)$$

where  $\xi_L(t) = (T-t)^+ Z_t$ . If, further,  $\eta$  is *a priori* a gamma process with scale function  $b(t)$ , shape measure  $dA(t)$ , it is *a posteriori* a gamma process with scale function  $[b^{-1}(t) + \xi_L(t)]^{-1}$ , and the same shape measure  $dA(t)$ . Alternatively, if  $\eta$  is compound Poisson *a priori* (see (3.3)) and  $N_T = 0$  then

$$Q_1(d\nu, dt) = e^{-(T-t)^+ Z_t} \lambda F(d\nu) dt,$$

and  $\eta$  is still compound Poisson with the same jump size distribution but with new intensity function  $\lambda_1(t) = \lambda \exp\{-(T-t)^+ Z_t\}$ , which takes value  $\lambda$  for  $t > T$ . If  $N_T > 0$ , it can be shown that

$$\Phi_1(\xi) = \exp \left\{ \int_0^\infty \int_{R^+} (e^{-\nu\xi(t)} - 1) Q_1(d\nu, dt) \right\} \times q(\xi)/q(0)$$

where

$$Q_1(d\nu, dt) = e^{-\nu(T-t)^+ Z_t} Q(d\nu, dt);$$

$$q(\xi) = \sum_{\sigma \in \Gamma} \prod_{r \in \sigma} \left\{ \int \int I(t \leq \min_{i \in r} t_i) (\nu Z_i)^j e^{-\nu\xi(t)} Q_1(d\nu, dt) \right\};$$

$\Gamma$  denotes the set of all distinct partitions of the integers  $1, \dots, N_T$ ;  $\sigma$  is an element of  $\Gamma$ ;  $r$  is a group in  $\sigma$ ; and  $j$  is the size of  $r$ .

*A posteriori* the process  $\eta$  no longer has independent increments, but is likely to be Markov by analogy with the discrete time example of Section 2.

In the next and final section we discuss the ingredients for a generalization of the type of calculation presented here.

## 5. Ingredients for a generalization

### 5.1 Bayes' theorem

In parametric statistical inference, Bayes' theorem tells us how the observation transforms a prior distribution on the parameter space to a posterior distribution. In terms of densities when these are defined, the theorem is usually written as

$$\Pi(\theta | x) = C f(x; \theta) \Pi(\theta)$$

where  $C^{-1} = \int f(x; \theta) \Pi(\theta) d\theta$ ,  $f$  as a function of  $x$  is the density of the observation, and  $\Pi(\theta)$ ,  $\Pi(\theta | x)$  represent respectively the prior and posterior densities of the parameter relative to some basic  $\sigma$ -finite measure  $\mu_0(\theta)$  on the parameter space.

However we can also write the theorem as a transformation of expectation operators. If  $g$  is an arbitrary function of  $\theta$ ,

$$\int g(\theta) \Pi(\theta | x) d\mu_0(\theta) = \int g(\theta) h(x; \theta) \Pi(\theta) d\mu_0(\theta)$$

where  $h(x; \theta) = C f(x; \theta)$ . Thus, taking a posterior expectation given  $x$  is equivalent to multiplying  $g$  by the prior-normalized likelihood  $h$  and then taking the prior expectation. It is this expression of Bayes' theorem which is most easily generalized to the stochastic process context. The following 'abstract version' is taken from Liptser and Shiriyayev (1977).

Let  $(\Omega, \mathcal{F}, P)$  be a probability space, and let  $\theta = \theta(\omega)$  and  $X = X(\omega)$  be random variables with values in the measurable spaces  $(\Theta, \beta_\Theta)$ ,  $(\mathcal{X}, \beta_X)$  respectively. Further,

let  $\mathcal{F}_\theta$  and  $\mathcal{F}_x$  be the  $\sigma$ -fields generated respectively by  $\theta, X$ ; and let  $Q$  be the restriction of the measure  $P$  to  $(\Omega, \mathcal{F}_x)$ . Define  $Q(A; \omega) = E[I_A | \mathcal{F}_\theta](\omega)$ , where  $I_A$  is the indicator of an event  $A$  in  $\mathcal{F}_x$ . Clearly  $Q(A) = E[Q(A; \omega)] = \int_{\Omega} Q(A; \omega) dP(\omega)$ . Now let  $g$  be a measurable real-valued function on  $(\Theta, \beta_\theta)$  such that  $E|g(\theta(\omega))| = 0$ , and for  $A \in \mathcal{F}_x$  define

$$\begin{aligned} G(A) &= \int_{\Omega} g(\theta(\omega)) Q(A; \omega) dP(\omega) \\ &= E[E(g(\theta) I_A | \mathcal{F}_\theta)] . \end{aligned}$$

Then the signed measure  $G$  is absolutely continuous with respect to  $Q$ , and we have as a preliminary form of Bayes' theorem

$$E[g(\theta) | \mathcal{F}_x] = \frac{dG}{dQ} \quad (\text{P.a.s.}) ,$$

the right hand side being a Radon-Nikodym derivative.

Further suppose that there is a representation

$$Q(A; \omega) = \int_A f(\omega, \tilde{\omega}) dP_0(\tilde{\omega}) \quad (\text{P.a.s.})$$

where  $f(\omega, \tilde{\omega})$  is a nonnegative measurable function on  $(\Omega \times \Omega, \mathcal{F}_x \times \mathcal{F}_\theta)$  and  $P_0$  is a probability measure. (See Liptser and Shirayayev (1977), p.287, for sufficient conditions.) Then it can be shown that

$$E[g(\theta) | \mathcal{F}_x](\omega) = \int_{\Omega} g(\theta(\tilde{\omega})) p(\omega, \tilde{\omega}) dP(\tilde{\omega}) \quad (\text{P.a.s.}) \quad (5.1)$$

where  $p(\omega, \tilde{\omega}) = f(\omega, \tilde{\omega}) / \int_{\Omega} f(\omega, \tilde{\omega}) dP(\tilde{\omega})$ . Clearly  $p(\omega, \tilde{\omega})$  is analogous to the normal-

ized likelihood  $h$  above.

In application of this formula to the situation we consider, both  $\theta$  and  $X$  are realizations of jointly distributed stochastic processes,  $\theta$  corresponding to  $(\eta_t, t \geq 0)$  and  $X$  to  $(X_t, R_t, t \in [0, T])$ . In cases where it is useful to characterize the process  $\eta$  by a Laplace transform functional, we consider  $g(\theta)$  to be  $\exp[-\int_0^\infty \xi(t) d\eta_t]$ . The computation implied by (5.1) is straightforward if the 'prior' Laplace transform functional of  $\eta$  is tractable and if the 'likelihood'  $f(\omega, \bar{\omega})$  is expressible as  $\exp[-\int_0^\infty \xi_L(t) d\eta_t]$  (or a derivative thereof).

## 5.2 The likelihood ratio process

A useful reference for the terms and results used here is the paper of Shiriyayev (1981).

We suppose that there is a standard filtration  $(\mathcal{F}_t, t \geq 0)$  on the measurable space  $(\Omega, \mathcal{F})$ , with  $\mathcal{F} = \mathcal{F}_\infty$ . Suppose that  $P_\theta$  and  $P_0$  are two probability measures on  $\mathcal{F}$  (each  $\mathcal{F}_t$  completed with respect to both), and that  $P_{\theta_t}$  and  $P_{0t}$  are their respective restrictions to  $\mathcal{F}_t$ . We say that  $P_\theta$  is locally absolutely continuous with respect to  $P_0$  if for each  $t$  we have  $P_{\theta_t} \ll P_{0t}$ .

When this local absolute continuity holds, we may denote by  $Z_t$  the Radon-Nikodym density of  $P_{\theta_t}$  with respect to  $P_{0t}$ , choosing a version so that  $Z = (Z_t, \mathcal{F}_t)$  is a martingale with respect to  $P$ , having right-continuous sample paths with left hand limits. We may then define  $Z_\infty = \lim Z_t$  as  $t \rightarrow \infty$ . It can be shown that  $P_\theta \ll P_0$  iff  $E_0 Z_\infty = 1$ , a condition which is equivalent to uniform integrability of  $Z$

with respect to  $P_0$ .

To express the likelihood ratio process  $Z$  in 'exponential' form we introduce a process  $\xi = (\xi_t, \mathcal{F}_t, t \geq 0)$  where

$$\xi_t = \int_0^t Z_{s-}^* dZ_s$$

and  $a^*$  is  $a^{-1}$  if  $a \neq 0$  and 0 if  $a = 0$ . It can be shown that  $Z_t$  has the Doléans-Dade (1970) representation

$$Z_t = \exp \left( \xi_t - \frac{1}{2} \langle \xi^c \rangle_t + \sum_{0 < s \leq t} [\ell n(1 + \Delta \xi_s) - \Delta \xi_s] \right) \quad (5.2)$$

where  $\langle \xi^c \rangle$  is the 'quadratic variation' process of the 'continuous martingale part' of the (local martingale)  $\xi$  and  $\Delta \xi_s$  represents  $\xi_s - \xi_{s-}$ . Let us denote the right hand side of (5.2) by  $\mathcal{E}(\xi)_t$ .

A result related to (5.2) which facilitates the matching of likelihood ratio processes and process models is provided by a generalized version of Girsanov's theorem (Elliott, 1982). That is, suppose  $Z = \mathcal{E}(\xi)$  is a uniformly integrable positive martingale and that a new probability measure  $P_\theta$  is defined on  $(\Omega, \mathcal{F}, P_0)$  by putting

$$\frac{dP_\theta}{dP_0} = Z_\infty.$$

If  $M$  is a local martingale under measure  $P_0$ , and the process  $\langle M, \xi \rangle$  exists under  $P_0$ , then

$$M_{\theta, t} = M_t - \langle M, \xi \rangle_t$$

defines a local martingale  $M_\theta$  under  $P_\theta$ .

For example suppose  $X = (X_t, t \geq 0)$  is standard Brownian motion under  $P_0$  and  $\zeta_t = \int_0^t \eta_s dX_s$  where  $(\eta_s, s \geq 0)$  is a 'predictable' process. Then

$$Z_t = \exp\left[\int_0^t \eta_s dX_s - \frac{1}{2} \int_0^t |\eta_s|^2 ds\right] . \quad (5.3)$$

If we take  $M_t = X_t$ , Girsanov's theorem tells us that under the measure  $P_\eta$  with likelihood ratio process  $Z$ ,

$$M_{\eta,t} = X_t - \int_0^t \eta_s ds$$

is a local martingale; in fact, it can then easily be shown that  $M_\eta$  is standard Brownian motion under  $P_\eta$ , and hence that (5.3) represents a likelihood for  $\eta$  in the "model"

$$dX_t = \eta_t dt + \text{Gaussian white noise} .$$

Similarly, suppose  $N = (N_t, t \geq 0)$  is a Poisson counting process with mean intensity 1 under  $P_0$  and  $\zeta_t = \int_0^t (\lambda_s - 1)d(N_s - s)$  where  $(\lambda_s, s \geq 0)$  is a positive predictable process. Then

$$Z_t = \exp\left\{\int_0^t (\lambda_s - 1)ds\right\} \prod_{i=1}^{N_t} \lambda_{\tau_i} , \quad (5.4)$$

where  $\tau_i$  are the jump times of  $N$ . If we take  $M_t = N_t - t$ , the theorem tells us that under the measure  $P_\lambda$  with likelihood ratio process  $Z$ ,  $N_t - \int_0^t \lambda_s ds$  is a local martingale; this defines (5.4) as the likelihood ratio process for a point process with intensity process  $(\lambda_s, s \geq 0)$ , as used in Section 4.



In the context<sup>of</sup> more complicated models, such as for example when the observation process is thought to come from a continuous time Markov chain, it seems most convenient to start by conjecturing a form for  $Z_t$ , calculating  $\zeta_t$  as  $\int_0^t Z_s^- dZ_s$ , and proceeding from there to check the consistency of the Girsanov theorem with a semimartingale representation of the model derived from first principles.

Note that the uniform integrability requirement for  $Z$  presents no problem if the observation is restricted to interval  $[0, T]$  where  $T$  is finite.

### 5.3 The analogue of the Laplace transform functional for processes with dependent increments

Although the prior Laplace transform functional would always characterize the prior measure of the 'parameter' process  $\eta$ , its use is most convenient when  $\eta$  has independent increments, because it is calculable as a single dimensional integral with respect to local characteristics of the process. The same is true for the more generally applicable characteristic functional.

Moreover, if the likelihood process is the "exponential" of a process  $\zeta$  which has a continuous martingale part, then even if  $\zeta$  itself is linear in  $\eta$  the variate  $Z_T$  will not be expressible as  $\exp\{-\int_0^\infty \xi_L(t) d\eta_t\}$ , as is required for the most satisfactory exploitation of the Laplace transform functional.

At the same time, the exponential nature of the likelihood process suggests that the use of some sort of exponential transform of the prior measure would be useful in applications.

We can take heart from the fact that if  $\eta$  has non-negative independent increments,  $\exp\{-\int_0^t \xi(s)d\eta_s\}/E\exp\{-\int_0^t \xi(s)d\eta_s\}$  is a martingale. For general local martingale  $\eta$ , there is an analogous multiplicative decomposition of

$$Y_t = \exp\{i\int_0^t \xi(s)d\eta_s\} = \exp\{i\int_0^t dN_s\},$$

namely

$$Y_t = \tilde{Y}_t A_t^{-1}$$

where  $\tilde{Y}$  is a positive local martingale,  $A_t = \mathcal{E}(a_t)$ ,

$$a_t(\omega) = -\frac{1}{2}\langle N^c \rangle_t + \int_0^t \int_{R \setminus \{0\}} (e^{i\nu} - 1 - i\nu) Q(\omega, ds, d\nu)$$

and  $Q$  is the compensating measure of the jump measure of the process  $N$ . See Liptser and Shiriyayev (1980), Doléans-Dade (1970). Thus it would seem that a practically useful way to characterize the prior and posterior measures of  $\eta$  is to use appropriate multiplicative decompositions of the process  $Y$ , and that these will indicate how the 'local characteristics' of  $\eta$  are transformed, conditionally on the observations.

## REFERENCES

1. Ammann, L.P. (1985) Conditional Laplace transforms for Bayesian non-parametric inference in reliability theory. *Stochastic Processes and Applications* 20, 197-212.
2. Brémaud, P. (1981) *Point processes and queues, martingale dynamics*. Springer-Verlag, New York.
3. Doléans-Dade, C. (1970) Quelques applications de la formule de changement de variables pour les semimartingales. *Zeitschrift für Wahrscheinlichkeits-theorie verw. Geb.* 16, 181-194.
4. Dykstra, R.L. and Laud, P. (1981) A Bayesian nonparametric approach to reliability. *Annals of Statistics* 9, 356-367.
5. Elliott, R.J. (1982) *Stochastic Calculus and Applications*. Springer-Verlag, New York.
6. Geman, S. and Geman, D. (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE-PAMI* 6, 721-741.
7. Liptser, R.S. and Shiriyayev, A.N. (1977) *Statistics of Random Processes I. General Theory*. Springer-Verlag, New York.
8. Liptser, R. and Shiriyayev, A.N. (1980). A functional central limit theorem for semimartingales. *Theory Prob. Applic.* 26, 132-137.
9. Shiriyayev, A.N. (1981) Martingales: recent developments, results and applications. *International Statistical Review* 49, 199-233.
10. Wild, C.J. and Klabfleisch, J.D. (1981) A note on a paper by Ferguson and Phadia. *Annals of Statistics* 9,1061-1065.