On time-reversibility and estimating functions for Markov processes.

Mathieu Kessler
Departamento de Matemática Aplicada
Universidad Politécnica de Cartagena
Spain

Michael Sørensen
Department of Statistics and Operations Research
University of Copenhagen
Denmark

Abstract

The time-reversibility of a Markov process implies a particular structure of the score function. It is explored which martingale estimating functions and other unbiased estimating functions have a similar structure. This leads to an estimating function with a semiparametric efficiency property. Also relations to martingale estimating functions based on eigenfunctions of the infinitesimal generator are found.

Key words: diffusion processes, discretely sampled continuous time Markov models, martingale estimating functions, the Poisson equation, quasi-likelihood, semiparametric models.

1 Introduction

It is often useful or necessary to model time series data sampled at discrete time points by a continuous time process. This is for instance the case, when the dynamics of the phenomenon under study are given by a stochastic differential equation. It is also the case in finance, where models for pricing derivative assets are usually formulated as continuous time models. Unfortunately, the likelihood function is in many cases not explicitly available for discretely sampled continuous time models, but estimating functions have turned out to be very useful in obtaining estimators and drawing inference for models where the likelihood function is not explicitly known; see e.g. Bibby and Sørensen (1995, 2001), Sørensen (1997, 2000), Genon-Catalot, Jeantheau, and Larédo (1999), Kessler and Sørensen (1999), Kessler (2000), Jacobsen (2001), H. Sørensen (2001), and Kessler and Paredes (2002). An application to financial data was given in Bibby and Sørensen (1997), while Pedersen (2000) used the method to estimate the nitrous oxide emission rate from the soil surface.

For Markov models martingale estimating functions are often a sum of terms of the following form: A function of an observation minus its conditional expectation given the previous observation. It is, however, not always obvious how to choose the function of the data on which to base the estimating function. The results in this paper are a contribution to the investigation of this question. We study when martingale estimating functions and other unbiased estimating functions have the same structure as the score function for a time-reversible Markov process. In Section 2 we introduce some necessary regularity conditions on the models and review three basic properties of the score function for a time-reversible Markov process. In particular, we derive the time-reversibility condition for estimating functions. In Section 3 we study when martingale estimating functions satisfy this condition. In this way we arrive at an estimating function that is efficient in the sense of semiparametric models, and we find relations to martingale estimating functions based on eigenfunctions of the infinitesimal generator. Finally, in Section 3 we consider estimating functions that satisfy the time-reversibility condition and study when they are martingale estimating functions or, more generally, when they are unbiased.

2 Conditions on the model and on the estimating functions

Consider a statistical model, parametrized by $\Theta \subseteq \mathbb{R}^p$, for a continuous-time stochastic process X with state space $E \subseteq \mathbb{R}$. It is assumed that X is a Markov process for each θ, and that the transition distribution has a strictly positive density $y \mapsto p(t,x,y;\theta)$ with respect to the Lebesgue measure on E. Specifically, $y \mapsto p(t,x,y;\theta)$ is the density of X_t given $X_0 = x$ ($t > 0$). We denote the class of infinitesimal generators by $\{A_\theta : \theta \in \Theta\}$.

Suppose the data are observations at discrete time points $X_{t_1}, \ldots, X_{t_n}$, $0 < t_1 < \cdots < t_n$, and that $X_0 = x_0$ is non-random. Then the likelihood function is given by

$$L_n(\theta) = \prod_{i=1}^{n} p(\Delta_i, X_{t_{i-1}}, X_{t_i}; \theta),$$

where $\Delta_i = t_i - t_{i-1}$ with $t_0 = 0$, and the score function by

$$U_n(\theta) = \sum_{i=1}^{n} \partial_\theta \log p(\Delta_i, X_{t_{i-1}}, X_{t_i}; \theta). \tag{2.1}$$

In many continuous-time models there is no tractable expression for the transition density p, but inference about the parameter θ can often usefully be made using another estimating function

$$G_n(\theta) = \sum_{i=1}^{n} g(\Delta_i, X_{t_{i-1}}, X_{t_i}; \theta), \tag{2.2}$$

where g is some function with values in $\mathbb{R}^p$ that is more tractable than $\partial_\theta \log p$, and which should be thought of as an approximation to $\partial_\theta \log p$. An estimator of θ is obtained by solving the equation $G_n(\theta) = 0$. In order to get as close to the efficient likelihood inference as possible, it seems preferable that g should have as many properties in common with $\partial_\theta \log p$ as possible. Let us therefore review three important properties of $\partial_\theta \log p$, and formulate these as conditions on estimating functions. To simplify the exposition, we shall from now on assume that the observation times are equidistant, i.e. that, for some $\Delta > 0$, $t_i = \Delta i$ for all i.

Under weak conditions the score function is a martingale for quite general stochastic process models, see e.g. Barndorff-Nielsen and Sørensen (1994). Indeed, under conditions allowing the interchange of integration and differentiation,

$$\begin{aligned} \int \partial_\theta \log p(t,x,y;\theta) p(t,x,y;\theta) dy &= \int \frac{\partial_\theta p(t,x,y;\theta)}{p(t,x,y;\theta)} p(t,x,y;\theta) dy \\ &= \int \partial_\theta p(t,x,y;\theta) dy = \partial_\theta \int p(t,x,y;\theta) dy = 0. \end{aligned}$$

Thus a reasonable condition on an estimating function is that it is a martingale, i.e. that g satisfies the following condition.

Condition P

$$\int_E g(\Delta, x, y; \theta) p(\Delta, x, y; \theta) dy = 0 \tag{2.3}$$

for all $\theta \in \Theta$ and all $x \in E$.

In many cases simpler and more explicit estimating functions can be obtained by replacing Condition P by the weaker Condition Q given below. To do this we need to assume that X is ergodic for all $\theta \in \Theta$. We also assume that the invariant probability measure has a density with respect to the Lebesgue measure on E and denote the density by μ_θ. Define the two-dimensional stationary distribution Q_θ^t by

$$Q_\theta^t(x,y) = p(t,x,y;\theta)\mu_\theta(x). \tag{2.4}$$

For p-dimensional functions f and h defined on E and E^2, respectively, we will use the notation

$$\mu_\theta(f) = \int_E f(x)\mu_\theta(x) dx \quad \text{and} \quad Q_\theta^t(h) = \int_E h(x,y) Q_\theta^t(x,y) dx dy.$$

Then we can formulate the following condition.

Condition Q

$$Q_\theta^\Delta(g(\Delta;\theta)) = 0 \quad \textit{for all } \theta \in \Theta. \tag{2.5}$$

Obviously, Condition P implies Condition Q, so the score function $U_n(\theta)$ satisfies Condition Q under weak conditions.

Condition Q is an asymptotic unbiasedness condition on the estimating function G, which (under some regularity conditions) is enough to ensure that there exists a consistent estimator of

θ which solves the estimating equation $G_n(\theta) = 0$ (at least for n large enough). The asymptotic theory of estimating functions is reviewed in Sørensen (1999). In order to ensure asymptotic normality of estimators obtained from estimating functions satisfying Condition Q, but not Condition P, the following condition is needed on the model.

Condition 2.1 *For every $\theta \in \Theta$, zero is an isolated point in the set of eigenvalues for the infinitesimal generator A_θ.*

Under Condition Q and Condition 2.1 we have that

$$\frac{1}{\sqrt{n}} G_n(\theta) \xrightarrow{D} N_p(0, V_\theta)$$

as $n \to \infty$, where the expression for the asymptotic variance matrix is rather complicated and involves the potential of X; for details see e.g. Kessler (2000). Properties implying Condition 2.1 in the case of diffusion models were studied in Hansen and Scheinkman (1995), Genon-Catalot, Jeantheau and Larédo (2000), and Kessler (2000). In particular, for one-dimensional diffusion models simple conditions can be given.

Now suppose X is *time-reversible*. Then $Q_\theta^t(x,y) = Q_\theta^t(y,x)$, or more explicitly,

$$p(t,x,y;\theta)\mu_\theta(x) = p(t,y,x;\theta)\mu_\theta(y) \tag{2.6}$$

for all $(x,y) \in E^2$, all $t > 0$ and all $\theta \in \Theta$. Equation (2.6) implies that

$$\partial_\theta \log p(t,x,y;\theta) + \ell_\theta(x) = \partial_\theta \log p(t,y,x;\theta) + \ell_\theta(y),$$

where

$$\ell_\theta(x) = \partial_\theta \log \mu_\theta(x) = \frac{\partial_\theta \mu_\theta(x)}{\mu_\theta(x)}. \tag{2.7}$$

Time-reversibility is not an uncommon property for a Markov process. For instance, all one-dimensional diffusions are time-reversible, see Kent (1978). For time-reversible processes it is of interest to study the consequences of imposing the following condition, which is satisfied by the score function.

Condition R

$$g(\Delta,x,y;\theta) + \ell_\theta(x) = g(\Delta,y,x;\theta) + \ell_\theta(y) \tag{2.8}$$

for all $\theta \in \Theta$ and all $(x,y) \in E^2$.

Note that, even when the transition density is intractable, the density of the invariant density may be known, as is the case for one-dimensional diffusions. Therefore, it may well be easy to check Condition R and thus possible to find a tractable g satisfying (2.8). Examples will be given in the following sections.

In the rest of the paper we will assume that the invariant measure satisfies the following Condition 2.2. By $L_p^2(\mu_\theta)$ we denote the set of p-dimensional functions defined on E for which every coordinate is square integrable with respect to μ_θ, and by $L_{p,0}^2(\mu_\theta)$ we denote the subspace of $L_p^2(\mu_\theta)$ of functions f for which $\mu_\theta(f) = 0$.

Condition 2.2

$$\ell_\theta \in L_{p,0}^2(\mu_\theta) \quad \textit{for all} \quad \theta \in \Theta.$$

A sufficient condition ensuring that $\mu_\theta(\ell_\theta) = 0$ is that the family of functions $\{\partial_\theta \mu_\theta(x) | \, \theta \in \Theta\}$ is locally dominated integrable with respect to the Lebesgue measure on E.

3 Martingale estimating functions

In this section we will study which martingale estimating functions satisfy Condition R. We first consider estimating functions of the form

$$g(t, x, y; \theta) = f_\theta(y) - \pi_{\theta,t} f_\theta(x), \tag{3.1}$$

where f_θ is a p-dimensional function defined in E, and $\pi_{\theta,t}$ denotes the transition operator defined by $\pi_{\theta,t} f(x) = E_\theta(f(X_t)|X_0 = x)$. This choice is a natural way to obtain a function g satisfying Condition P.

Obviously, g satisfies Condition P, but for which choices of f_θ is Condition R satisfied too? We can assume that $t = 1$ and that $\mu_\theta(f_\theta) = 0$ without loss of generality, because we can change the time-scale and we can subtract $\mu_\theta(f_\theta)$ from f_θ without changing g. In the following we will suppress t in the notation. Under Condition 2.2, it is not difficult to see that for functions f_θ satisfying $\mu_\theta(f_\theta) = 0$, Condition R is satisfied if and only if f_θ satisfies the equation

$$\pi_\theta(f_\theta) = \ell_\theta - f_\theta. \tag{3.2}$$

Indeed, equation (2.8) is equivalent to

$$f_\theta(y) + \pi_\theta f_\theta(y) - \ell_\theta(y) = f_\theta(x) + \pi_\theta f_\theta(x) - \ell_\theta(x)$$

for all $(x, y) \in E^2$. Hence the function $q_\theta(y) = f_\theta(y) + \pi_\theta f(y) - \ell_\theta(y)$ must be constant (as a function of y), and its constant value is $\mu_\theta(q_\theta) = \mu_\theta(f_\theta) + \mu_\theta(\pi_\theta f_\theta) - \mu_\theta(\ell_\theta) = 0$. Note that (3.2) is the Poisson equation for the operator $-\pi_\theta$. To solve this equation, we suppose that Condition 2.1 is satisfied. Under this condition, since it implies that π_θ is a contraction in $L^2(\mu_\theta)$, the operator $\pi_\theta + I$, where I denotes the identity operator, has a bounded inverse, V_θ, on $L^2_{p,0}(\mu_\theta)$. Hence the solution to the equation (3.2) is

$$f_\theta = V_\theta \ell_\theta. \tag{3.3}$$

Specifically,

$$V_\theta = \sum_{i=0}^{\infty} (-1)^i \pi_\theta^i, \tag{3.4}$$

where the sum converges in $L^2_{p,0}(\mu_\theta)$ because π_θ is a strong contraction under Condition 2.1.

Using that $\pi_\theta V_\theta = 1 - V_\theta$ and (3.4), we see that the function g given by (3.1) and (3.3) can also be written as

$$\begin{aligned} g(x, y; \theta) &= V_\theta \ell_\theta(y) - \pi_\theta V_\theta \ell_\theta(x) = V_\theta \ell_\theta(y) + V_\theta \ell_\theta(x) - \ell_\theta(x) \\ &= \ell_\theta(y) - [\pi_\theta V_\theta \ell_\theta(y) + \pi_\theta V_\theta \ell_\theta(x)] = \ell_\theta(y) + \sum_{i=1}^{\infty} (-1)^i [\pi_\theta^i \ell_\theta(y) + \pi_\theta^i \ell_\theta(x)]. \end{aligned} \tag{3.5}$$

Interestingly, martingale estimating functions of this form were derived via a completely different route in Kessler, Schick and Wefelmeyer (2001). They considered semiparametric reversible Markovian models for X, where only the family of invariant measures $\{\mu_\theta | \, \theta \in \Theta\}$ was specified parametrically. It was proved in the paper that the estimator obtained from (2.2) with g given by (3.5) is efficient in the sense of semiparametric models, i.e. it yields an estimator whose asymptotic variance attains the semiparametric lower bound. To obtain a semiparametric estimator, it is obviously necessary to estimate $\pi_\theta V_\theta$ non-parametrically from the data. Kessler, Schick and Wefelmeyer (2001) also proved that estimation based on this efficient martingale estimating function is asymptotically equivalent to estimation based on the simple estimating function

$$\sum_{i=1}^{n} \tilde{f}(X_{i\Delta};\theta),$$

where

$$\tilde{f}(x;\theta) = 2V_\theta \ell_\theta(x) - \ell_\theta(x) = (V_\theta - \pi_\theta V_\theta)\ell_\theta(x).$$

This simple estimating function is of the form considered in Kessler (2000).

Obviously, parameters on which the invariant probability measure μ_θ does not depend cannot be estimated by the estimating function given by (3.5). If μ_θ does not depend on the ith coordinate of θ_i, then the ith coordinate of ℓ_θ will be equal to zero, and hence so will the ith coordinate of g. Thus g maps into a sub-space of a dimension equal to the number of parameters needed to parametrize μ_θ.

Let us study the operator V_θ a little more in the case where the set of eigenvalues for the infinitesimal generator A_θ is discrete. Denote the eigenvalues and the corresponding eigenfunctions by $\{\lambda_i^\theta : i = 0,1,\ldots\}$ ($\lambda_0^\theta \leq \lambda_1^\theta \leq \cdots$) and $\{\varphi_i^\theta : i = 0,1,\ldots\}$. We assume that the eigenfunctions are normalized such that $\|\varphi_i^\theta\|_\theta = 1$ for all i and that the set of eigenfunctions constitute a complete orthonormal system in $L_1^2(\mu_\theta)$. Here $\|\cdot\|_\theta$ denotes the norm in $L_1^2(\mu_\theta)$. We can assume that $\varphi_0^\theta = 1$. The space $L_{1,0}^2(\mu_\theta)$ is the orthogonal complement to the subspace spanned by φ_0^θ, so any $h \in L_{p,0}^2(\mu_\theta)$ has a representation $\sum_{i=1}^\infty a_i \varphi_i^\theta$, where the jth coordinate of a_i equals $\langle h_j, \varphi_i^\theta\rangle_\theta$. Here h_j is the jth coordinate of h, and $\langle\cdot,\cdot\rangle_\theta$ denotes the inner product in $L_1^2(\mu_\theta)$. It is not difficult to see that

$$V_\theta h = \sum_{i=1}^{\infty} \frac{a_i}{1+e^{-\lambda_i^\theta}} \varphi_i^\theta = h - \sum_{i=1}^{\infty} \frac{a_i}{1+e^{\lambda_i^\theta}} \varphi_i^\theta. \tag{3.6}$$

Now consider $V_\theta \ell_\theta$. A standard calculation shows that $\langle \ell_\theta, \varphi_i^\theta\rangle_\theta = -\mu_\theta(\partial_\theta \varphi_i^\theta)$, provided that the interchange of integration and differentiation is allowed. A standard condition ensuring this is that the family of functions $\{\partial_\theta(\mu_\theta \varphi_i^\theta)\}$ is locally dominated integrable with respect to the Lebesgue measure. Thus we obtain that

$$f_\theta = V_\theta \ell_\theta = \ell_\theta + \sum_{i=1}^{\infty} \frac{\mu_\theta(\partial_\theta \varphi_i^\theta)}{1+e^{\lambda_i^\theta}} \varphi_i^\theta. \tag{3.7}$$

In Kessler and Sørensen (1999) it was proposed to construct martingale estimating functions from eigenfunctions of the infinitesimal generator. From (3.5) we would get an estimating function of the type considered in that paper by truncating the series in (3.7) and by not including ℓ_θ. Equation (3.7) is a further argument for using this type of estimating function, but indicates that it might be worthwhile to study the case where also the function ℓ_θ is included.

Using (3.5) we see that

$$g(x,y;\theta) = \ell_\theta(y) + \sum_{i=1}^{\infty} \frac{\mu_\theta(\partial_\theta \varphi_i^\theta)}{1+e^{\lambda_i^\theta}} [\varphi_i^\theta(x) + \varphi_i^\theta(y)].$$

Example 3.1 Consider the diffusion model

$$dX_t = -\beta(v_t - \alpha)dt + \sigma\sqrt{X_t}dW_t,$$

which is known in the finance literature as the CIR model for short term interest rates, see Cox, Ingersoll and Ross (1985). This process is ergodic and its stationary distribution is the gamma distribution with shape parameter $2\beta\alpha\sigma^{-2}$ and scale parameter $2\beta\sigma^{-2}$ provided that $\beta > 0$, $\alpha > 0$, $\sigma > 0$, and $2\theta\alpha \geq \sigma^2$. Hence we can estimate the parameters $\theta_1 = 2\beta\alpha\sigma^{-2}$ and $\theta_2 = 2\beta\sigma^{-2}$ by means of the estimating function given by (3.7). We see that

$$\ell_\theta(x) = \begin{pmatrix} \log(\theta_2 x) - \Psi(\theta_1) \\ \\ -x + \theta_1/\theta_2 \end{pmatrix},$$

where Ψ denotes the di-gamma function.

The eigenfunctions of the generator are Laguerre polynomials evaluated at $\theta_2 x$, see e.g. Karlin and Taylor (1981). Specifically, the i-th eigenfunction is the i-th order polynomial

$$\varphi_i^\theta(x) = \sum_{j=0}^{i} \rho_{ij}(\theta_1, \theta_2) x^j$$

with coefficients

$$\rho_{ij}(\theta_1, \theta_2) = \frac{(-\theta_2)^j}{j!} \begin{pmatrix} i + \theta_1 - 1 \\ i - j \end{pmatrix},$$

$j = 0, \ldots, i$, $i = 1, 2, \ldots$. The corresponding eigenvalue is $i\theta$. To calculate (3.7) we need to calculate $\mu_\theta(\partial_\theta \varphi_i^\theta)$, which is given by

$$\mu_\theta(\partial_{\theta_k} \varphi_i^\theta) = \sum_{j=0}^{i} [\partial_{\theta_k} \rho_{ij}(\theta_1, \theta_2)](\theta_1 + j - 1)^{(j)} \theta_2^j.$$

where $a^{(j)} = a(a-1)\cdots(a-j+1)$. With this specification, (3.7) does not seem to correspond to any known function, and in practice one would have to truncate the series, say at $i = N$. A more efficient estimating function would be obtained by using the optimal martingale estimating function based on ℓ_θ, $\varphi_1^\theta, \ldots, \varphi_N^\theta$. The results in this section can therefore be interpreted as an argument for using this estimating function for inference.

$\square$

The more general weighted martingale estimating functions with g of the form

$$g(t, x, y; \theta) = \alpha(t, x; \theta)[f_\theta(y) - \pi_{\theta,t} f_\theta(x)] \tag{3.8}$$

are less tractable and probably do not often satisfy Condition R. To investigate this type of estimating function, we can again assume that $t = 1$ and suppress t in the notation. A function g of this form satisfies Condition P. Condition R is equivalent to

$$\begin{aligned} &\alpha(x;\theta) f_\theta(y) - \alpha(y;\theta) f_\theta(x) = \\ &\qquad \ell_\theta(y) - \ell_\theta(x) + \alpha(x;\theta)\pi_\theta f_\theta(x) - \alpha(y;\theta)\pi_\theta f_\theta(y), \end{aligned} \tag{3.9}$$

which implies that

$$\alpha'(x;\theta) f'_\theta(y) - \alpha'(y;\theta) f'_\theta(x) = 0, \tag{3.10}$$

where a prime indicates differentiation with respect to x (or y). To simplify the notation let us assume here that θ, α, and f are one-dimensional. Equation (3.10) states that the function $f'_\theta(x)/\alpha'(x;\theta)$ is constant, so a necessary condition that Condition R is satisfied is that

$$\alpha(x;\theta) = c_1(\theta) f_\theta(x) + c_2(\theta). \tag{3.11}$$

The case $c_1 = 0$ was treated above. By inserting (3.11) into (3.9), we find that the function $\ell_\theta(x) - [c_1(\theta) f_\theta(x) - c_2(\theta)]\pi_\theta f_\theta(x)$ is equal to a constant, which must be $-c_1(\theta)\langle f_\theta, \pi_\theta f_\theta\rangle_\theta$. Usually $\langle f_\theta, \pi_\theta f_\theta\rangle_\theta \neq 0$. Thus we have obtained the non-linear equation

$$\ell_\theta(x) = c_1(\theta)\left[f_\theta(x)\pi_\theta f_\theta(x) - \langle f_\theta, \pi_\theta f_\theta\rangle_\theta\right] + c_2(\theta)\pi_\theta f_\theta(x), \tag{3.12}$$

which does not seem tractable in general. In the following simple example it can, however, be solved.

Example 3.2 Consider the *Ornstein-Uhlenbeck process,* which is the solution of the stochastic differential equation

$$dX_t = -\theta X_t dt + dW_t.$$

Its invariant measure is the normal distribution with mean zero and variance $1/(2\theta)$, so

$$\ell_\theta(x) = -(x^2 - (2\theta)^{-1}).$$

Consider the function $f_\theta(x) = \kappa_\theta x$. Since $\pi_\theta f_\theta(x) = \kappa_\theta e^{-\theta} x$, it follows that

$$f_\theta(x)\pi_\theta f_\theta(x) - \langle f_\theta, \pi_\theta f_\theta\rangle_\theta = \kappa_\theta^2 e^{-\theta}(x^2 - (2\theta)^{-1}).$$

Thus equation (3.12) is satisfied with $c_1 = -1$ if we choose $\kappa_\theta = e^{\theta/2}$. The corresponding g equals

$$g(x, y; \theta) = x[y - xe^{-\theta}].$$

The martingale estimating function with this g is optimal in the class of estimating functions with g of the form $g(x, y; \theta) = \alpha(x; \theta)[y - xe^{-\theta}]$. The corresponding estimator equals, in fact, the maximum likelihood estimator of θ in the model $dX_t = -\theta X_t dt + \sigma dW_t$.

□

4 Estimating functions satisfying Condition R

In this section we will attack the problem from a different angle. Here we will assume that g satisfies Condition R, and then study what is further needed for Condition P or Condition Q to hold. Again we will simplify matters by assuming that the observations are equidistant with $\Delta = 1$. Obviously, an estimating function (2.2) satisfies Condition R if and only if

$$g(x, y; \theta) = \ell_\theta(y) + \tilde{f}_\theta(x, y) \tag{4.1}$$

for all x, y and θ, where $\tilde{f}_\theta$ is symmetric in x and y. Hence we need only discuss how to choose the function $\tilde{f}_\theta$.

First we consider the case

$$\tilde{f}_\theta(x, y) = h_\theta(x) + h_\theta(y) \tag{4.2}$$

for some function h_θ. In this case Condition P is satisfied precisely when

$$\pi_\theta h_\theta = -\pi_\theta \ell_\theta - h_\theta.$$

This is again the Poisson equation (3.2), only with ℓ_θ replaced by $-\pi_\theta \ell_\theta$. Thus

$$h_\theta = -\pi_\theta V_\theta \ell_\theta,$$

and

$$g(x, y; \theta) = \ell_\theta(y) - [\pi_\theta V_\theta \ell_\theta(y) + \pi_\theta V_\theta \ell_\theta(x)] = V_\theta \ell_\theta(y) - \pi_\theta V_\theta \ell_\theta(x).$$

Comparison to (3.5) shows that this is the same estimating function as the one derived in Section 3.

We conclude our investigation by considering the more general case

$$\tilde{f}_\theta(x, y) = h_\theta(x) k_\theta(y) + h_\theta(y) k_\theta(x). \tag{4.3}$$

To simplify matters a bit we assume that θ, h_θ and k_θ are one-dimensional. Here Condition P is satisfied precisely when

$$\pi_\theta \ell_\theta + k_\theta \pi_\theta h_\theta + h_\theta \pi_\theta k_\theta = 0.$$

Suppose we can choose a k_θ such that $\pi_\theta k_\theta(x) \neq 0$ for all $x \in E$. Then

$$\frac{k_\theta}{\pi_\theta k_\theta} \pi_\theta h_\theta = -\frac{\pi_\theta \ell_\theta}{\pi_\theta k_\theta} - h_\theta, \tag{4.4}$$

which is the Poisson equation for the operator $K_\theta = (k_\theta / \pi_\theta k_\theta) \pi_\theta$. Suppose the operator $K_\theta + I$ has a bounded inverse, U_θ, defined on some subset $\mathcal{U}_\theta$ of $L_0^2(\mu_\theta)$. Then the solution to equation (4.4)

$$h_\theta = U_\theta m_\theta,$$

provided that the function

$$m_\theta = -\frac{\pi_\theta \ell_\theta}{\pi_\theta k_\theta}$$

belongs to the domain $\mathcal{U}_\theta$ of U_θ. The operator U_θ is given by

$$U_\theta h = \sum_{i=0}^{\infty} (-1)^i K_\theta^i h, \tag{4.5}$$

for any $h \in L_0^2(\mu_\theta)$ for which the sum converges (in $L^2(\mu_\theta)$). Thus $\mathcal{U}_\theta$ is contained in the set of functions f for which the sum (4.5) converges. In general, $\mathcal{U}_\theta$ is a proper subset of $L_0^2(\mu_\theta)$ depending on the choice of the function k_θ.

A bit more can be said if we choose k_θ equal to an eigenfunction φ_θ of the generator with eigenvalue λ_θ, because under weak regularity conditions $\pi_\theta \varphi_\theta = e^{-\lambda_\theta} \varphi_\theta$. Simple sufficient conditions for diffusion processes are given in Kessler and Sørensen (1999). Above we required that $\pi_\theta \varphi_\theta(x) \neq 0$ for all $x \in E$, so the considerations here only apply to models with an eigenfunction without zero points in E. In this situation $K_\theta = e^{\lambda_\theta} \pi_\theta$, so $\mathcal{U}_\theta$ contains the functions $f \in L_0^2(\mu_\theta)$ for which $\pi_\theta f \leq e^{-\lambda_1} f$ for some $\lambda_1 > \lambda_\theta$, i.e. $\mathcal{U}_\theta$ contains the span of all eigenfunctions with eigenvalue strictly larger than λ_θ.

Finally, to obtain simpler results, we turn to Condition Q, which is satisfied when

$$\langle k_\theta, \pi_\theta h_\theta \rangle_\theta + \langle \pi_\theta k_\theta, h_\theta \rangle_\theta = 0.$$

Since we have assumed that the observed Markov process is time-reversible, π_θ is self-adjoint, so this equation simplifies to

$$\langle k_\theta, \pi_\theta h_\theta \rangle_\theta = 0. \tag{4.6}$$

This problem has several solutions. Define

$$h_\theta = \sum_{i \in I_1} a_i(\theta) \varphi_i^\theta$$

$$k_\theta = \sum_{i \in I_2} b_i(\theta) \varphi_i^\theta,$$

where $I_j \subseteq \mathbb{N}_0$, $j = 1,2$ with $I_1 \cap I_2 = \emptyset$, where $a_i(\theta)$ and $b_i(\theta)$ are real numbers, and where the φ_i^θ-s are eigenfunctions of the infinitesimal generator A_θ. Under the regularity conditions mentioned above ensuring that $\pi_\theta \varphi_j^\theta = e^{-\lambda_j^\theta}\varphi_j^\theta$, where λ_j^θ is the eigenvalue corresponding to φ_j^θ, we have for $i \neq j$ that $\langle \varphi_i^\theta, \pi_\theta \varphi_j^\theta \rangle_\theta = e^{-\lambda_j^\theta} \langle \varphi_i^\theta, \varphi_j^\theta \rangle_\theta = 0$. Hence equation (4.6) is satisfied.

Example 4.1 Consider the class of diffusion processes given as the solutions of

$$dX_t = -\theta \tan(X_t)dt + dW_t, \quad \theta \geq \tfrac{1}{2}.$$

The solutions are ergodic diffusions on the interval $(-\frac{\pi}{2}, \frac{\pi}{2})$ that can be thought of as an Ornstein-Uhlenbeck process on a finite interval. The invariant measure is given by

$$\mu_\theta(x) = \frac{\Gamma(2\theta+1)\cos^{2\theta}(x)}{4^\theta \Gamma(\theta + \frac{1}{2})^2}, \quad x \in (-\tfrac{\pi}{2}, \tfrac{\pi}{2}),$$

where Γ denotes the gamma-function, so

$$\ell_\theta(x) = 2\log(\cos(x)) + 2\left(\Psi(2\theta+1) - \Psi(\theta + \tfrac{1}{2})\right) - \log(4),$$

where Ψ denotes the di-gamma function. The eigenfunctions are $\varphi_i^\theta(x) = C_i^\theta(\sin(x))$, where C_i^θ is a Gegenbauer polynomial of order i, see Kessler and Sørensen (1999). The first two non-trivial eigenfunctions are $\sin(x)$ and $2(\theta+1)\sin^2(x) - 1$, so

$$\begin{aligned} g(x,y;\theta) &= 2\log(\cos(x)) + 2\left(\Psi(2\theta+1) - \Psi(\theta+\tfrac{1}{2})\right) - \log(4) \\ &\quad + \sin(x)\left(2(\theta+1)\sin^2(y) - 1\right) + \sin(y)\left(2(\theta+1)\sin^2(x) - 1\right) \end{aligned}$$

satisfies Condition R. That the necessary regularity conditions are satisfied for this model is demonstrated in Kessler and Sørensen (1999).

□

Acknowledgement.

Most of this research was done while Mathieu Kessler stayed at the Department of Theoretical Statistics, University of Copenhagen. This stay was jointly financed by MaPhySto - Centre for Mathematical Physics and Stochastics, funded by a grant from The Danish National Research Foundation, and the University of Copenhagen. Our collaboration was also supported by the European Commission through the Research Training Network DYNSTOCH under the Human Potential Programme.

References

Barndorff-Nielsen, O.E. and Sørensen, M. (1994): A review of some aspects of asymptotic likelihood theory for stochastic processes. *Int. Statist. Rev.* **62**, 133–165.

Bibby, B.M. and Sørensen, M. (1995): Martingale estimating functions for discretely observed diffusion processes. *Bernoulli* **1**, 17–39.

Bibby, B.M. and Sørensen, M. (1997): A hyperbolic diffusion model for stock prices. *Finance and Stochastics* **1**, 25 – 41.

Bibby, B.M. and Sørensen, M. (2001): Simplified estimating functions for diffusion models with a high-dimensional parameter. *Scand. J. Statist.* **28**, 99–112.

Cox, J.C., Ingersoll, J.E. and Ross, S.A. (1985): A theory of the term structure of interest rates. *Econometrica* **53**, 385–407.

Genon-Catalot, V., Jeantheau, T. and Larédo, C. (1999): Parameter estimation for discretely observed stochastic volatility models. *Bernoulli* **5**, 855 – 872.

Genon-Catalot, V., Jeantheau, T. and Larédo, C. (2000): Stochastic volatility models as hidden Markov models and statistical applications. *Bernoulli* **6**, 1051 – 1079.

Hansen, L.P. and Scheinkman, J.A. (1995): Back to the future: Generating moment implications for continuous-time Markov processes. *Econometrica* **63**, 767–804.

Jacobsen, M. (2001): Discretely observed diffusions: classes of estimating functions and small Δ-optimality. *Scand. J. Statist.* **28**, 123 – 150.

Karlin, S. and Taylor, H.M. (1981). *A Second Course in Stochastic Processes.* Orlando: Academic Press.

Kent, J. (1978): Time-reversible diffusions. *Adv. Appl. Prob.* **10**, 819 – 835.

Kessler, M. (2000): Simple and explicit estimating functions for a discretely observed diffusion process. *Scand. J. Statist.* **27**, 65 – 82.

Kessler, M. and Paredes, S. (2002): Computational aspects related to martingale estimating functions for a discretely observed diffusion. To appear in *Scand. J. Statist.*

Kessler, M., Schick, A., and Wefelmeyer, W. (2001): The information contained in the marginal law of a Markov chain. *Bernoulli* **7**, 243 – 266.

Kessler, M. and Sørensen, M. (1999): Estimating equations based on eigenfunctions for a discretely observed diffusion process. *Bernoulli* **5**, 299 – 314.

Pedersen, A.R. (2000): Estimating the nitrous oxide emission rate from the soil surface by means of a diffusion mod. *Scand. J. Statist.* **27**, 385 – 403.

Sørensen, H. (2001): Discretely observed diffusions: Approximation of the continuous-time score function. *Scand. J. Statist.* **28**, 113 – 121.

Sørensen, M. (1997): Estimating functions for discretely observed diffusions: A review. In Basawa, I.V., Godambe, V.P. and Taylor, R.L. (eds.): *Selected Proceedings of the Symposium on Estimating Functions.* IMS Lecture Notes - Monograph Series, Vol. 32, 305–325.

Sørensen, M. (1999): On asymptotics of estimating functions. *Brazilian Journal of Probability and Statistics* **13**, 111–136.

Sørensen, M. (2000): Prediction-based Estimating Functions. *Econometrics Journal* **3**, 123–147.