Exponential family inference for diffusion models

Michael Sørensen
University of Aarhus
Denmark

Abstract

We consider ergodic diffusion processes for which the class of invariant measures is an exponential family, and study inference based on the class of invariant probability measures when the diffusion has been observed at discrete time points. When the drift depends linearly on the parameters, the invariant measures form an exponential family. It is investigated how the usual exponential family inference, which can be done by means of standard statistical computer packages, works when the observations are from a diffusion process. In particular, the limit distributions of estimators and test statistics are derived. As an example, we consider classes of diffusions with generalized inverse Gaussian marginals. A particular instance is the well-known Cox-Ingersoll-Ross model from mathematical finance.
Key Words: Asymptotic Normality; Consistency; Cox-Ingersoll-Ross model; Discrete time observation; Inference for Diffusion Processes; Estimating Functions; Exponential families; Generalized inverse Gaussian diffusions.

1 Introduction

Diffusion processes often provide a useful alternative to the discrete time stochastic processes traditionally used in time series analysis as models for observations at discrete time points of a phenomenon that develops dynamically in time. In many fields of application it is natural to model the dynamics in continuous time, whereas dynamic modelling in discrete time contains an element of arbitrariness. This is particularly so when the time between observations is not equidistant. An example is financial data, where the models used to price derivative assets such as options are based on continuous time models, usually diffusion models, which must be fitted to time series of stock prices, interest rates or currency exchange rates.

Statistical inference for diffusion processes based on discrete time observations can only rarely be based on the likelihood function as this is usually not explicitly available. The likelihood function is a product of transition densities, as follows easily from the fact that diffusions are Markov processes, but explicit expressions for the transition densities are only known in some special cases. Alternatives are simulated likelihood inference (Pedersen 1995a, 1995b) or martingale estimating functions (Bibby and Sørensen, 1995, 1996, Kessler, 1995, Kessler and Sørensen, 1995, Sørensen, 1996). Both approaches are, however, rather computer-intensive, and it is desirable to have a simple alternative that can, at least, be used for a first analysis of data.

Kessler (1996) studied a type of estimating functions for diffusion models which include the pseudo score function obtained by pretending that the data are independent observations from a distribution that belongs to the class of invariant measures. Here we study in detail the particular case, where the class of invariant measures is an exponential family. This is a particularly interesting case because the inference is simple and the analysis of data can be done using standard statistical computer packages containing procedures for analysing generalized linear models. When the drift depends linearly on the parameters, the invariant measures form an exponential family. The estimator obtained in this paper is obviously not efficient. Questions of relative efficiency within the class of estimating functions that are a sum of functions dependent on the parameters and only one of the observations were studied by Kessler (1996).

It should be noted that the diffusion models considered in this paper are not exponential families of stochastic processes in the sense of Küchler and Sørensen (1997). They do also not belong to the type of models studied by Ycard (1992), who excludes the case of stationary processes.

In section 2, diffusion models where the invariant measures form an exponential family are studied and some classical exponential family results are reviewed. In Section 3, we find the exponential family estimator, $\hat{\theta}_n$, and prove that it is consistent and asymptotically normal. The asymptotic variance is different from that in the case of independent observations, so we study an estimator for the asymptotic variance of $\hat{\theta}_n$. Finally, results about the limit distribution of the pseudo likelihood ratio test statistic are given. In the final Section 4, a flexible class of diffusions with invariant measures in the exponential family of generalized inverse Gaussian distributions is introduced and investigated.

2 Diffusion models where the invariant measures form an exponential family

Consider a one-dimensional system that we would like to model by means of the ordinary differential equation

$$\frac{dx_t}{dt} = \sum_{i=1}^{p} \beta_i b_i(x_t), \tag{2.1}$$

where $b_1, \ldots, b_p$ are known continuous functions, and $\beta_1, \ldots, \beta_p$ are unknown parameters about which we would like to draw inference.

Suppose we have observed the system at the time points $0 < t_1 < \cdots < t_n$, so that we have the data $X_{t_1}, \ldots, X_{t_n}$. If the data do not follow (2.1) exactly, we might want to try to use the diffusion model given by the class of stochastic differential equations

$$dX_t = \sum_{i=1}^{p} \beta_i b_i(X_t)dt + \lambda v(X_t)dW_t, \tag{2.2}$$

where W is a standard Wiener process, $v > 0$ is a known function and $\lambda > 0$ is an real parameter. We consider only functions $b_1, \ldots, b_p$ and $v > 0$ that satisfy the usual conditions ensuring the existence of a unique weak solution. An idea about how to choose v could be obtained by studying suitable plots of the data. The parameter λ is an index of the level of noise in the system. Denote the interior of the state space of a solution X of (2.2) by (l, r), where $-\infty \leq l < r \leq \infty$. The state space is assumed to be the same for all parameter values.

With the reparametrization $\theta_i = \beta_i / \lambda^2$, $i = 1, \ldots, p$, and the definition

$$T_i(x) = \int_{x_0}^{x} \frac{2b_i(y)}{v^2(y)} dy, \quad i = 1, \ldots, p \tag{2.3}$$

for some fixed $x_0 \in (l, r)$, we have the following representation of the scale measure

$$s(x; \theta) = \exp\left(- \sum_{i=1}^{p} \theta_i T_i(x) \right) \tag{2.4}$$

and of the speed measure

$$m(x; \theta) = v(x)^{-2} \exp\left(\sum_{i=1}^{p} \theta_i T_i(x) \right). \tag{2.5}$$

We use the notation $\theta = (\theta_1, \ldots, \theta_p)^T$. Transposition of a vector or matrix a is denoted by a^T. Consider the parameter sets

$$\Theta = \left\{ \theta \in \mathbb{R}^p : \int_l^r m(x; \theta) dx < \infty \right\} \tag{2.6}$$

and

$$\Theta_1 = \left\{ \theta \in \mathrm{IR}^p : \int_l^{x_0} s(x;\theta)dx = \int_{x_0}^r s(x;\theta)dx = \infty \right\}. \tag{2.7}$$

For $\theta \in \Theta$, let μ_θ denote the probability measure on (l,r) with density

$$h(x;\theta) = \exp\left(\sum_{i=1}^p \theta_i T_i(x) - 2\log v(x) - \phi(\theta) \right), \tag{2.8}$$

where

$$\phi(\theta) = \log \int_l^r m(x;\theta)dx. \tag{2.9}$$

Then the class of probability measures $\{\mu_\theta : \theta \in \Theta\}$ is an exponential family. If $\theta \in \Theta_1$, the diffusion process X is ergodic with invariant measure μ_θ. It is quite possible that Θ_1 is a proper subset of Θ. For instance for the Cox-Ingersoll-Ross model, which is obtained for $p = 2$, $b_1(x) = 1$, $b_2(x) = x$, and $v(x) = \sqrt{x}$, we find that $\Theta = (0,\infty) \times (-\infty, 0)$ and $\Theta_1 = [1,\infty) \times (-\infty, 0)$. In this case, the invariant measures are gamma distributions. It is easy to see that we have the following result.

Lemma 2.1 *Suppose that $(l,r) = \mathrm{IR}$ and that v is bounded. Then $\Theta = \Theta_1$.*

We assume that the functions $b_1/v^2, \ldots, b_p/v^2$ are linearly independent. This implies that also the functions $T_1, \ldots, T_p$ are linearly independent, and that, hence, the exponential representation (2.8) is minimal and the parameter θ is identifiable.

The following are standard results for exponential families; see Barndorff-Nielsen (1978) or Brown (1986). For $\theta \in \operatorname{int}\Theta$, the mapping $\phi(\theta)$ is infinitely often differentiable, and $T(X_t)$ has moments of any order if $X_t \sim \mu_\theta$. In particular,

$$\mathrm{E}_{\mu_\theta}(T(X_t)) = \partial_\theta \phi(\theta), \tag{2.10}$$

$$\mathrm{V}_{\mu_\theta}(T(X_t)) = \partial_\theta^2 \phi(\theta). \tag{2.11}$$

We use the notation $\partial_\theta \phi(\theta)$ for the vector of partial derivatives of ϕ with respect to the coordinates of θ, and $\partial_\theta^2 \phi(\theta)$ for the $p \times p$ matrix of second order partial derivatives.

Condition 2.2 *The cumulant transform ϕ is steep.*

Let $\mathcal{T}$ denote the closure of the convex hull of the image of (l,r) by T. The following result is also classical; Barndorff-Nielsen (1978) or Brown (1986).

Lemma 2.3 *Under Condition 2.2 the mapping*

$$\partial_\theta \phi : \ \operatorname{int}\Theta \mapsto \operatorname{int}\mathcal{T}$$

is a homeomorphism.

In the next section, we shall consider inference based on the exponential family of invariant measures. By this procedure we can, obviously, only draw inference about the parameter $\theta = (\theta_1, \ldots, \theta_p)^T$. If λ is known, this is not a problem. When λ is not known, we can always estimate it by means of a quadratic martingale estimating function, see Bibby and Sørensen (1996). Properties of estimators of multi-dimensional parameters, where some components of the parameter are estimated using the class of marginal distributions, while the other components are estimated by means of a martingale estimating function, are studied in Bibby and Sørensen (1997). Even without estimating λ, we can draw inference about the ratio between the value of the β's in (2.1) based on the invariant measures.

If one of the β's is known, we can also estimate λ using only the invariant measures. If, specifically, the dynamic system under study is ideally modelled by the deterministic equation

$$\frac{dx_t}{dt} = b_1(x_t) + \sum_{i=2}^{p} \beta_i b_i(x_t), \tag{2.12}$$

where $b_1, \ldots b_p$ are known continuous functions and $\beta_2, \ldots, \beta_p$ are unknown parameters. Then we might try the stochastic model defined by

$$dX_t = \left[b_1(X_t) + \sum_{i=2}^{p} \beta_i b_i(X_t) \right] dt + \lambda v(X_t) dW_t, \tag{2.13}$$

where W is a standard Wiener process, and $v > 0$ is a known function, while $\lambda > 0$ is unknown. After the reparametrization $\theta_1 = 1/\lambda^2$ and $\theta_i = \beta_i/\lambda^2$, $i = 2, \ldots, p$, the scale and speed measures are given by (2.4) and (2.5). As above, X is ergodic if $\theta \in \Theta_1$, and the invariant measure has the density (2.8).

Diffusion models with exponential families of invariant measures can be obtained in other ways than the one described above. For instance, the class of stochastic differential equations

$$dX_t = \lambda \exp\left(-\tfrac{1}{2} \sum_{i=1}^{p} \theta_i T_i(X_t) \right) dW_t,$$

where $T_1, \ldots, T_p$ are known functions, has speed measures given by

$$m(x; \theta) = \exp\left(\sum_{i=1}^{p} \theta_i T_i(x) \right).$$

In this case the scale measure is $s(x; \theta) = 1$, so $\Theta_1 = \Theta$ and X is ergodic when $\theta \in \Theta$. The invariant measure is given by (2.8) with $v = 1$. Models of this type, with a few non-exponential parameters added, were used in Bibby and Sørensen (1997) to model stock prices.

3 Exponential family inference for diffusion models

In this section we consider inference about the parameters in the invariant measure. We base the inference on the exponential pseudo likelihood function obtained by pretending that the data $X_{t_1}, \ldots, X_{t_n}$ are independent observations with a common distribution belonging to the class of invariant measures $\{\mu_\theta : \theta \in \Theta_1\}$. The corresponding pseudo score function is

$$G_n(\theta) = \sum_{j=1}^{n} \left[T(X_{t_j}) - m(\theta) \right], \tag{3.1}$$

where

$$m(\theta) = \partial_\theta \phi(\theta). \tag{3.2}$$

This is an estimating function of the type studied in Kessler (1996).

Lemma 2.3 implies that under Condition 2.2 the estimating equation $G_n(\theta) = 0$ has the unique solution

$$\hat{\theta}_n = m^{-1} \left(\frac{1}{n} \sum_{j=1}^{n} T(X_{t_j}) \right) \in \operatorname{int} \Theta, \tag{3.3}$$

provided that $\frac{1}{n} \sum_{j=1}^{n} T(X_{t_j}) \in \operatorname{int} \mathcal{T}$.

We will now study the properties of this estimator when the number of observations tends to infinity. We simplify matters by assuming that the observations are made at equidistant time points, i.e. $t_j = j\Delta$, $j = 1, \ldots, n$, for some $\Delta > 0$.

Proposition 3.1 *Suppose $\theta \in \Theta_1 \cap \operatorname{int} \Theta$, and that Condition 2.2 holds. Then the estimator $\hat{\theta}_n$ exists with a probability tending to one as $n \to \infty$, and*

$$\hat{\theta}_n \to \theta$$

in probability as $n \to \infty$ when θ is the true parameter value.

Proof: Since X is ergodic ($\theta \in \Theta_1$), it follows by the ergodic theorem that

$$\frac{1}{n} \sum_{j=1}^{n} T(X_{j\Delta}) \to E_{\mu_\theta}(T(X_t)) \in \operatorname{int} \mathcal{T}$$

in probability as $n \to \infty$, and using (2.10), the proposition follows because $\theta \in \operatorname{int} \Theta$.

□

Define

$$f(x; \theta) = T(x) - \partial_\theta \phi(\theta). \tag{3.4}$$

Then by (2.10) and (2.11) we have $f_i(\theta) \in L_0^2(\mu_\theta)$ when $\theta \in \operatorname{int}\Theta$. Here f_i denotes the ith coordinate of f, and $L_0^2(\mu_\theta)$ is the set of real functions on (l, r) that are square integrable with expectation zero under μ_θ.

The estimating function $G_n(\theta)$, given by (3.1), is not a martingale when θ is the true parameter value, so we cannot directly use the central limit theorem for martingales to obtain results about asymptotic normality of $G_n(\theta)$ and $\hat{\theta}_n$. Therefore, we need further assumptions. Let $a(x;\theta)$ denote the drift of the diffusion, and define

$$
\begin{aligned}
u(x;\theta) = {} & \tfrac{1}{2}[a(x;\theta)^2\lambda^{-2}v(x)^{-2} + a'(x;\theta)] \\
& - v'(x)a(x;\theta)/v(x) + \tfrac{1}{8}\lambda^2 v'(x)^2 - \tfrac{1}{4}\lambda^2 v(x)v''(x).
\end{aligned}
$$

A prime denotes differentiation with respect to x.

Condition 3.2

$$
\min\{\lim_{x\to l} u(x;\theta), \lim_{x\to r} u(x;\theta)\} > 0
$$

for all $\theta \in \Theta_1$.

Under Condition 3.2, there exists, for every $\theta \in \Theta$, a constant $\lambda_\theta > 0$ such that for $g \in L_0^2(\mu_\theta)$

$$
\|\Pi_{\Delta,\theta} g\|_\theta \leq e^{-\lambda_\theta \Delta}\|g\|_\theta \tag{3.5}
$$

for all $\Delta > 0$. Here $\|\cdot\|_\theta$ is the norm in $L^2(\mu_\theta)$, i.e. $\|g\|_\theta^2 = \int g^2 d\mu_\theta$, and $\Pi_{\Delta,\theta}$ denotes the operator defined by

$$
\Pi_{\Delta,\theta} g(x) = E_\theta(g(X_\Delta)|X_0 = x) = \int g(y)p(\Delta, x, y;\theta)dy,
$$

where $p(\Delta, x, y;\theta)$ is the transition density for X, i.e. the conditional density of X_Δ given $X_0 = x$ when θ is the true parameter value. The contraction property (3.5) of $\Pi_{\Delta,\theta}$ implies that the operator

$$
U_\theta g(x) = \sum_{i=0}^{\infty} E_\theta(g(X_{i\Delta})|X_0 = x) = \sum_{i=0}^{\infty} \Pi_{\Delta,\theta}^i g(x)
$$

is well-defined for $g \in L_0^2(\mu_\theta)$. The sum converges in $L_0^2(\mu_\theta)$.

Theorem 3.3 *Suppose that* $\theta \in \Theta_1 \cap \operatorname{int}\Theta$ *and that Condition 3.2 holds. Then*

$$
\frac{1}{\sqrt{n}} G_n(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} f(X_{i\Delta};\theta) \xrightarrow{\mathcal{D}} N(0, A_\theta) \tag{3.6}
$$

when θ *is the true parameter value. Here*

$$
A_\theta = \int h(x, y;\theta)h(x, y;\theta)^T Q_\Delta^\theta(dx, dy) \tag{3.7}
$$

with

$$h(x,y;\theta) = U_\theta f(y;\theta) - U_\theta f(x;\theta) + f(x;\theta) \tag{3.8}$$

and

$$Q^\theta_\Delta(dx,dy) = p(\Delta,x,y;\theta)dy\mu_\theta(dx). \tag{3.9}$$

This is an application of the multivariate version of the central limit theorem in Florens-Zmirou (1989), which follows easily by the Cramér-Wold device; see also Kessler (1996). It follows from the central limit theorem for martingales because $G_n(\theta) = \sum_{i=1}^n h(X_{(j-1)\Delta}, X_{j\Delta};\theta) - U_\theta f(X_{n\Delta};\theta) + U_\theta f(X_0;\theta)$. Here the sum is a martingale when θ is the true parameter value, since $\Pi_{\Delta,\theta}(U_\theta f) = U_\theta f - f$. According to Florens-Zmirou (1989), the following condition also implies that Theorem 3.3 holds.

Condition 3.4 *For every $\theta \in \Theta_1$ there exist constants c_θ and M_θ such that*

$$[a(x;\theta)/v(x) - \tfrac{1}{2}\lambda^2 v'(x)]\mathrm{sign}(x) \le -c_\theta$$

for all x satisfying $|x| > M_\theta$.

This follows from Florens-Zmirou's condition by applying the standard transformation $t(x) = \int_0^x v(y)^{-1}dy$ to X to obtain a diffusion of the type considered by her. It is easy to see that

$$\begin{aligned}(A_\theta)_{ij} &= \int U_\theta f_i(y;\theta)U_\theta f_j(y;\theta)\mu_\theta(y)dy \qquad (3.10)\\ &\quad - \int [U_\theta f_i(x;\theta) - f_i(x;\theta)][U_\theta f_j(x;\theta) - f_j(x;\theta)]\mu_\theta(x)dx\\ &= \int \sum_{k=1}^\infty E_\theta[f_j(X_{k\Delta};\theta)|X_0 = x]f_i(x;\theta)\mu_\theta(x)dx\\ &\quad + \int \sum_{k=1}^\infty E_\theta[f_i(X_{k\Delta};\theta)|X_0 = x]f_j(x;\theta)\mu_\theta(x)dx\\ &\quad + \int f_i(x;\theta)f_j(x;\theta)\mu_\theta(x)dx\\ &= \sum_{k=1}^\infty \{E_{\mu_\theta}[f_i(X_0;\theta)f_j(X_{k\Delta};\theta)]\\ &\quad + E_{\mu_\theta}[f_j(X_0;\theta)f_i(X_{k\Delta};\theta)]\} + \int f_i(x;\theta)f_j(x;\theta)\mu_\theta(x)dx\\ &= H(\theta)_{ij} + \partial_\theta^2\phi(\theta)_{ij}\end{aligned}$$

where

$$H(\theta) = 2\sum_{k=1}^\infty c_k(\theta) \tag{3.11}$$

with

$$c_k(\theta) = \{\mathrm{Cov}_{\mu_\theta}(f_i(X_0;\theta), f_j(X_{k\Delta};\theta))\}_{i,j=1,\ldots,p} \tag{3.12}$$
$$= \left\{\int f_i(x;\theta) f_j(y;\theta) Q^\theta_{k\Delta}(dx,dy)\right\}_{i,j=1,\ldots,p}.$$

We have used that X is time reversible for $\theta \in \Theta_1$; see Kent (1978). Without any further conditions we have the following result about the distribution of our estimator.

Theorem 3.5 *Suppose that $\theta \in \Theta_1 \cap \mathrm{int}\,\Theta$ and that Condition 2.2 and Condition 3.2 hold. Then*

$$\sqrt{n}(\hat\theta_n - \theta) \xrightarrow{\mathcal{D}} N(0, V(\theta)^{-1} A_\theta V(\theta)^{-1}), \tag{3.13}$$

where

$$V(\theta) = \partial_\theta^2 \phi(\theta), \tag{3.14}$$

when θ is the true parameter value. Moreover,

$$V(\theta)^{-1} A_\theta V(\theta)^{-1} = V(\theta)^{-1} + V(\theta)^{-1} H(\theta) V(\theta)^{-1} \tag{3.15}$$

and

$$(V(\theta)^{-1} H(\theta) V(\theta)^{-1})_{ij} = O(e^{-\lambda_\theta \Delta}) \tag{3.16}$$

as $\Delta \to \infty$.

Note that $V(\theta)$ is the covariance matrix of $T(X_t)$ when $X_t \sim \mu_\theta$; cf. (2.11).

Proof: To prove (3.13) consider the standard expansion

$$0 = G_n(\hat\theta_n) = G_n(\theta) - nV(\tilde\theta_n^{(1)}, \ldots, \tilde\theta_n^{(p)})(\hat\theta_n - \theta),$$

where each of $\tilde\theta_n^{(i)}$, $i = 1, \ldots, p$ is a convex combination of $\hat\theta_n$ and θ, and where $V(\theta_1, \ldots, \theta_p)$ is the matrix the ith row of which equals the ith row of V evaluated at the argument θ_i. Since $V(\theta)$ is continuous and invertible, the result follows by applying Proposition 3.1 and Theorem 3.3. Formula (3.15) follows immediately from (3.10). Finally, (3.16) follows by the Cauchy-Schwarz inequality:

$$|c_k(\theta)_{ij}| = |E_{\mu_\theta}[f_i(X_0;\theta) f_j(X_{k\Delta};\theta)]| \tag{3.17}$$
$$\le \|f_i(\theta)\|_\theta \, \|\Pi_{k\Delta,\theta} f_j(\theta)\|_\theta \le e^{-\lambda_\theta \Delta k} \|f_i(\theta)\|_\theta \, \|f_j(\theta)\|_\theta,$$

where we have used (3.5). Thus, $|H(\theta)_{ij}| \le 2e^{-\lambda_\theta \Delta}\sqrt{V(\theta)_{ii} V(\theta)_{jj}}/(1 - e^{-\lambda_\theta \Delta})$.

□

In order to use Theorem 3.5 in practice, we need an estimate of the covariance matrix A_θ. This is not straightforward because the observations are not independent. In the econometric literature there are estimators for A_θ that are robust towards correlation between the observations, see e.g. Gallant (1987). However, we prefer to exploit the expression $A_\theta = H(\theta) + V(\theta)$; cf. (3.10). Let $\hat{c}_k^{(n)}(\theta)$ denote the obvious estimator for the matrix of covariances $c_k(\theta)$:

$$\hat{c}_k^{(n)}(\theta)_{ij} = \frac{1}{n-k} \sum_{l=1}^{n-k} f_i(X_{l\Delta};\theta) f_j(X_{(l+k)\Delta};\theta), \tag{3.18}$$

where $n > k$. Then an estimator of $H(\theta)$ is given by

$$\hat{H}_n(\theta) = 2 \sum_{k=1}^{\nu(n)} \hat{c}_k^{(n)}(\theta), \tag{3.19}$$

where $\nu(n) < n$ and $\nu(n) \to \infty$ as $n \to \infty$.

Proposition 3.6 *Suppose that the conditions of Theorem 3.5 are satisfied and that $\nu(n) = o(\sqrt{n})$. Then*

$$\hat{H}_n(\theta) \to H(\theta) \quad and \quad \hat{H}_n(\hat{\theta}_n) \to H(\theta) \tag{3.20}$$

in probability as $n \to \infty$ provided that θ is the true parameter value. If $\nu(n) = n^\alpha$ with $0 < \alpha < \frac{1}{2}$, then $|\hat{H}_n(\hat{\theta}_n)_{ij} - H(\theta)_{ij}| = O_p(n^{\alpha - \frac{1}{2}})$. If $\nu(n)/\log(n) = O(1)$, then $|\hat{H}_n(\hat{\theta}_n)_{ij} - H(\theta)_{ij}| = o_p(n^{-\frac{1}{2}+\epsilon})$ for every $\epsilon \in (0, \frac{1}{2})$. The same is true of $|\hat{H}_n(\theta)_{ij} - H(\theta)_{ij}|$.

Proof: Since

$$|\hat{H}_n(\hat{\theta}_n)_{ij} - H(\theta)_{ij}| \leq 2 \sum_{k=1}^{\nu(n)} |\hat{c}_k^{(n)}(\hat{\theta}_n)_{ij} - c_k(\theta)_{ij}| + 2 \sum_{k=\nu(n)+1}^{\infty} |c_k(\theta)_{ij}|,$$

where the last sum on the right-hand side is dominated by a constant (dependent on θ) times $e^{-\lambda_\theta \Delta \nu(n)}$, cf. (3.17), the proposition follows if we can prove that each of the terms in the first sum on the right-hand side is of order $O_p(n^{-\frac{1}{2}})$. By straightforward calculations

$$\begin{aligned} \sqrt{n}\,[\hat{c}_k^{(n)}(\hat{\theta}_n)_{ij} - c_k(\theta)_{ij}] \;\; = \;\; & \frac{\sqrt{n}}{n-k} \sum_{l=1}^{n-k} [T_i(X_{l\Delta}) T_j(X_{(l+k)\Delta}) - \tau(\theta)_{ij}] \\ & - m(\hat{\theta}_n)_i \frac{\sqrt{n}}{n-k} \sum_{l=1}^{n-k} [T_j(X_{(l+k)\Delta}) - m(\theta)_j] \\ & - m(\hat{\theta}_n)_j \frac{\sqrt{n}}{n-k} \sum_{l=1}^{n-k} [T_i(X_{l\Delta}) - m(\theta)_i] \end{aligned}$$

$$- m(\theta)_j \sqrt{n} \, [m(\hat{\theta}_n)_i - m(\theta)_i]$$

$$- m(\theta)_i \sqrt{n} \, [m(\hat{\theta}_n)_j - m(\theta)_j]$$

$$+ \sqrt{n} \, [m(\hat{\theta}_n)_i m(\hat{\theta}_n)_j - m(\theta)_i m(\theta)_j],$$

where $\tau(\theta)_{ij} = \int T_i(x) T_j(y) \, Q^{\theta}_{k\Delta}(dx, dy)$. That the three last terms on the right-hand side converge in distribution follows from Theorem 3.5 by the δ-method, because m is continuously differentiable. To see that the first term on the right-hand side converges in distribution, note that for $n = km$ this term equals

$$\sqrt{\frac{m}{(m-1)k}} \sum_{r=1}^{k} \left[\frac{1}{\sqrt{m-1}} \sum_{l=1}^{m-1} \{ T_i(X_{((l-1)k+r)\Delta}) T_j(X_{(lk+r)\Delta}) - \tau(\theta)_{ij} \} \right]. \quad (3.21)$$

The difference between the first term on the right-hand side and (3.21) with m equal to the integer part of n/k goes to zero in probability as $n \to \infty$. Since $T(X_t)$ has moments of any order when $X_t \sim \mu_\theta$ and $\theta \in \operatorname{int} \Theta$, it follows by the Cauchy-Schwarz inequality that $\int T_i(x)^2 T_j(y)^2 Q^{\theta}_{k\Delta}(dx, dy) < \infty$. Moreover, the terms in the sum in square brackets in (3.21) have zero expectation under $Q^{\theta}_{k\Delta}$, so we can apply the central limit theorem in Florens-Zmirou (1989) to prove that the normalized sum in the square brackets converges in distribution as $m \to \infty$. That the second and third terms on the right-hand side of the expression for the difference between $\hat{c}$ and c converge in distribution is the same result as Theorem 3.3. To prove this theorem, the central limit theorem was used too. This completes the proof of the last part of (3.20). The first part of (3.20) follows by similar, but slightly simpler, arguments. The results about the order of convergence are obvious from the proof of (3.20).

□

Finally, we consider tests of the hypothesis

$$H : \theta = \theta_0.$$

Obvious test statistics are the Wald test statistics

$$n(\hat{\theta}_n - \theta_0)^T V(\theta) \hat{A}_n^{-1} V(\theta)(\hat{\theta}_n - \theta_0)$$

and the pseudo score statistic

$$n^{-1} G_n(\theta_0)^T \hat{A}_n^{-1} G_n(\theta_0),$$

where

$$\hat{A}_n = V(\hat{\theta}_n) + \hat{H}_n(\hat{\theta}_n). \quad (3.22)$$

Both of these test statistics are asymptotically χ^2-distributed with p degrees of freedom as $n \to \infty$ when the hypothesis is true. When we use $\hat{H}_n$, we always assume that $\nu(n) = o(\sqrt{n}\,)$.

It is of some interest to study the limiting distribution of the log-likelihood ratio test statistic, which is computed by most computer packages.

Proposition 3.7 *Suppose that the hypotesis H is true. Then*

$$-2\log Q \xrightarrow{\mathcal{D}} \lambda_1(\theta_0)Z_1 + \cdots + \lambda_p(\theta_0)Z_p \tag{3.23}$$

as $n \to \infty$, where $\lambda_1(\theta_0), \ldots, \lambda_p(\theta_0)$ are the eigenvalues of $A_{\theta_0}^{\frac{1}{2}} V(\theta_0)^{-1} A_{\theta_0}^{\frac{1}{2}}$, and where $Z_1, \ldots, Z_p$ are independent $\chi^2(1)$-distributed random variables.

A very accurate saddlepoint approximation to the limit distribution in (3.23) was derived in Jensen (1995).

Proof: The result follows from Theorem 3.5 because

$$\begin{aligned}-2\log Q &= n(\hat{\theta}_n - \theta_0)^T V(\theta_0)(\hat{\theta}_n - \theta_0) + o_p(1)\\ &= [A_{\theta_0}^{-\frac{1}{2}} V(\theta_0)\sqrt{n}(\hat{\theta}_n - \theta_0)]^T A_{\theta_0}^{\frac{1}{2}} V(\theta_0)^{-1} A_{\theta_0}^{\frac{1}{2}} [A_{\theta_0}^{-\frac{1}{2}} V(\theta_0)\sqrt{n}(\hat{\theta}_n - \theta_0)] + o_p(1).\end{aligned}$$

□

In the one-dimensional case $p = 1$, we see that

$$\frac{-2\log Q}{1 + \hat{H}_n(\theta_0)/V(\theta_0)} \xrightarrow{\mathcal{D}} \chi^2(1), \tag{3.24}$$

as $n \to \infty$ when H is true.

Similar results can easily be derived when testing more general hypotheses about θ. When the reduction in the degrees of freedom is one, a simple normalization of the test statistic resulting in a $\chi^2(1)$ limit distribution can be found in analogy with (3.24).

4 Generalized inverse Gaussian diffusions

In this section we introduce a flexible class of positive diffusions with invariant measures belonging to a particular exponential family, namely the class of generalized inverse Gaussian distributions. These distributions were introduced by Good (1953), and their statistical properties were studied in detail by Jørgensen (1982). A generalized inverse Gaussian distribution has density

$$\tfrac{1}{2}(\psi/\chi)^{\gamma/2} K_\gamma\left(\sqrt{\chi\psi}\right)^{-1} x^{\gamma-1} \exp\left[-\tfrac{1}{2}\left(\chi x^{-1} + \psi x\right)\right], \quad x > 0, \tag{4.1}$$

with respect to the Lebesgue measure on $(0, \infty)$. Here K_γ is the modified Bessel function of the third kind with index γ. The parameter domain is as follows. For $\gamma > 0$, $\psi > 0$ and $\chi \geq 0$; for $\gamma = 0$, $\psi > 0$ and $\chi > 0$; for $\gamma < 0$, $\psi \geq 0$ and $\chi > 0$. For $\gamma = -\frac{1}{2}$, (4.1) is a standard inverse Gaussian distribution; for $\chi = 0$, we get the gamma-distributions; and for $\psi = 0$ the upper tail is of the Pareto type. The diffusions discussed in this section include the Cox-Ingersoll-Ross model from mathematical finance.

We consider diffusions given as solutions of

$$dX_t = \left(\beta_1 X_t^{2\alpha-1} - \beta_2 X_t^{2\alpha} + \beta_3 X_t^{2(\alpha-1)}\right) dt + \lambda X_t^\alpha dW_t, \quad X_0 > 0. \tag{4.2}$$

We define new parameters by $\theta_i = 2\beta_i\lambda^{-2}$, $i = 1, 2, 3$. Inference about θ_1, θ_2, and θ_3 can be drawn by the exponential family methods discussed in this paper, when the diffusion is ergodic. We assume that $\alpha \geq 0$ and $\lambda > 0$ are known. If α and λ are not known, inference about these parameters must be drawn by other methods such as martingale estimating functions.

The scale measure has the density

$$s(x) = x^{-\theta_1} \exp[\theta_2 x + \theta_3 x^{-1}], \quad x > 0.$$

Since $x^\alpha > 0$ for $x > 0$ and the function $x^{-2\alpha}$ is integrable on any compact sub-interval of $(0, \infty)$, it follows by a corollary to results in Engelbert and Schmidt (1985) that (4.2) has a unique weak solution, provided that the scale function $S(x) = \int_1^x s(y)dy$ satisfies

$$-S(0) = S(\infty) = \infty. \tag{4.3}$$

The condition (4.3) clearly holds when $\beta_2 > 0$ and $\beta_3 > 0$. If $\theta_1 \geq 1$, the condition is also satisfied for $\beta_3 = 0$, and if $\theta_1 \leq 1$, it holds for $\beta_2 = 0$ too.

The speed measure has the density

$$x^{\theta_1 - 2\alpha} \exp[-\theta_2 x - \theta_3 x^{-1}], \quad x > 0, \tag{4.4}$$

which is a finite measure whenever $\gamma = 1 + \theta_1 - 2\alpha$, $\psi = 2\theta_2$ and $\chi = 2\theta_3$ belong to the parameter space of the generalized inverse Gaussian distribution. In that case, the diffusion X is ergodic with an invariant measure that has the density (4.1), provided that also the condition (4.3) is satisfied. Thus, the parameter values for which the diffusion is ergodic are as follows. For $\theta_1 \geq 1$, $\theta_2 > 0$ and $\theta_3 \geq 0$; for $1 - 2\alpha \leq \theta_1 < 1$, $\theta_2 > 0$ and $\theta_3 > 0$; for $\theta_1 < 1 - 2\alpha$, $\theta_2 \geq 0$ and $\theta_3 > 0$. This defines the set Θ_1 introduced in Section 2. Note that when $\beta_2 = 0$, the upper tail of the invariant measure is of the Pareto type.

An interesting property of a solution of (4.2) is its *revertion* around the level $[\beta_1 + \sqrt{\beta_1^2 + 4\beta_2\beta_3}]/(2\beta_2)$, when $\beta_2 > 0$. If $\beta_2 = 0$, it reverts around

$-\beta_3/\beta_1$, provided that $\beta_1 < 0$. This cannot be called mean-reversion, a term often used in econometrics, because of the non-linearity of the drift.

Another remarkable property of an inverse Gaussian diffusion X solving (4.2) is that, by Ito's formula, its *inverse* X^{-1} is also an inverse Gaussian diffusion with diffusion coefficient $\lambda x^{2-\alpha}$. Thus, if $\alpha = 1$, the diffusion coefficient is the same for X and X^{-1}. If, moreover, $\theta_1 = 1$ and $\theta_2 = \theta_3 = 0$, the diffusions X and X^{-1} are identical.

For an ergodic process on $(0, \infty)$ the boundary 0 is either an entrance boundary or a natural boundary. For $\beta_3 > 0$, it is an entrance boundary for X when $\alpha < \frac{3}{2}$, while it is natural when $\alpha \geq \frac{3}{2}$. If $\beta_3 = 0$, it is entrance for $1 \leq \theta_1 < 2$ and natural when $\theta_1 \geq 2$.

Let us finally consider three particular cases. When $\alpha = 0$, the drift is $a(x) = \beta_1 x^{-1} - \beta_2 + \beta_3 x^{-2}$. For $\beta_1 > 0$ and $\beta_2 = \beta_3 = 0$, the solution of (4.2) is the Bessel process, which is not ergodic. However, if β_1 or β_2 are chosen such that there is sufficient pull away from infinity (and of course still enough pull away from zero), an ergodic diffusion is obtained. Next consider $\alpha = \frac{1}{2}$, where the drift is $a(x) = \beta_1 - \beta_2 x + \beta_3 x^{-1}$. For $\beta_3 = 0$, we obtain the Cox-Ingersoll-Ross model for interest rates, while the radial Ornstein-Uhlenbeck process (Karlin and Taylor, 1981, p. 333) is obtained for $\beta_1 = 0$. When $\alpha = 1$, the drift is $a(x) = \beta_1 x - \beta_2 x^2 + \beta_3$.

A solution of the stochastic differential equation

$$dX_t = \left\{ v(X_t)v'(X_t) + \tfrac{1}{2}v(X_t)^2 \left[(\gamma - 1)X_t^{-1} - \tfrac{1}{2}\psi + \tfrac{1}{2}\chi X_t^{-2} \right] \right\} dt + v(X_t)dW_t$$

will also, under suitable regularity conditions, be ergodic with invariant measure given by (4.1). We have, however, in this paper decided to restrict the attension to solutions of (4.2).

References

Barndorff-Nielsen, O.E. (1978): *Information and Exponential Families.* Wiley, Chichester.

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. (1996): On estimation for discretely observed diffusions: A review. *Theory of Stochastic Processes* **2 (18)**, 49–56.

Bibby, B.M. and Sørensen, M. (1997): Simplified estimating functions for diffusion models with a high-dimensional parameter. In preparation.

Brown, L.D. (1986): *Fundamentals of Statistical Exponential Families.* Institute of Mathematical Statistics, Hayward.

Florens-Zmirou, D. (1989): Approximate discrete-time schemes for statistics of diffusion processes. *Statistics*, **20**, 547–557.

Engelbert, H.J. and Schmidt, W. (1985): On solutions of one-dimensional stochastic differential equations without drift. *Z. Wahrscheinlichkeitstheorie verw. Gebiete.* **68**, 287–314.

Gallant, A.R. (1987): *Nonlinear Statistical Models.* Wiley, New York.

Good, I.J. (1953): The population frequencies of species and the estimation of population parameters. *Biometrika* **40**, 237–260.

Jensen, J.L. (1995): *Saddlepoint Approximations.* Oxford University Press, Oxford.

Jørgensen, B. (1982): *Statistical Properties of the Generalized Inverse Gaussian Distribution.* Lecture Notes in Statistics 9. Springer-Verlag, New York.

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

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

Kessler, M. (1995): Martingale estimating functions for a Markov chain. Preprint, Laboratoire de Probabilités, Université Paris VI.

Kessler, M. (1996): Simple and explicit estimating functions for a discretely observed diffusion process. Research Report No. 336, Department of Theoretical Statistics, University of Aarhus.

Kessler, M. and Sørensen, M. (1995): Estimating equations based on eigenfunctions for a discretely observed diffusion process. Research Report No. 332, Department of Theoretical Statistics, University of Aarhus.

Küchler, U. and Sørensen, M. (1997): *Exponential Families of Stochastic Processes.* Springer-Verlag, New York.

Pedersen, A.R. (1995a): A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scand. J. Statist.* **22**, 55–71.

Pedersen, A.R. (1995b): Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes. *Bernoulli* **1**, 257–279.

Ycard, B. (1992): Markov processes and exponential families. *Stoch. Proc. Appl.* **41**, 203–214.