Hindawi Publishing Corporation
International Journal of Mathematics and Mathematical Sciences
Volume 2011, Article ID 906846, 26 pages
doi:10.1155/2011/906846

Research Article

On Local Linear Approximations to Diffusion Processes

X. L. Duan,[1] Z. M. Qian,[1] and W. A. Zheng[2]

[1] *Department of Mathematics, Mathematical Institute, University of Oxford, 24-29 St Giles', Oxford OX1 3LB, UK*
[2] *Department of Statistics, SFS, ITCS, East China Normal University, Shanghai 200062, China*

Correspondence should be addressed to W. A. Zheng, wazheng@stat.ecnu.edu.cn

Received 27 December 2010; Revised 29 April 2011; Accepted 27 June 2011

Academic Editor: Shyam Kalla

Diffusion models have been used extensively in many applications. These models, such as those used in the financial engineering, usually contain unknown parameters which we wish to determine. One way is to use the maximum likelihood method with discrete samplings to devise statistics for unknown parameters. In general, the maximum likelihood functions for diffusion models are not available, hence it is difficult to derive the exact maximum likelihood estimator (MLE). There are many different approaches proposed by various authors over the past years, see, for example, the excellent books and Kutoyants (2004), Liptser and Shiryayev (1977), Kushner and Dupuis (2002), and Prakasa Rao (1999), and also the recent works by Aït-Sahalia (1999), (2004), (2002), and so forth. Shoji and Ozaki (1998; see also Shoji and Ozaki (1995) and Shoji and Ozaki (1997)) proposed a simple local linear approximation. In this paper, among other things, we show that Shoji's local linear Gaussian approximation indeed yields a good MLE.

1. Introduction

Diffusion processes are used as theoretical models in analyzing random phenomena evolved in continuous time. These models may be described in terms of Itô's type stochastic differential equations

$$\mathrm{d}X_t = A(X_t,\theta)\mathrm{d}t + \sigma(X_t,\theta)\mathrm{d}W_t, \tag{1.1}$$

where $(W_t)_{t\geq 0}$ is a Brownian motion, with some unknown parameters θ to be determined in rational ways.

It is, however, difficulty to derive the maximum likelihood estimator for θ if the diffusion coefficient (i.e., the volatility) σ is unknown. On the other hand, in practice, the volatility is determined first by using the fact that

$$\sigma^2 T = \lim \sum_{j=1}^{n} \left(X_{(j-1)T/n} - X_{jT/n}\right)^2 \quad \text{in prob.} \tag{1.2}$$

when σ is a constant. Therefore we will limit ourselves on diffusion models with constant volatility:

$$\mathrm{d}X_t = A(X_t, \theta)\mathrm{d}t + \mathrm{d}W_t. \tag{1.3}$$

Since there is no much difference at technical level, we will consider one-dimensional models only. That is, we will assume throughout the paper that W is a one-dimensional Brownian motion, and X is real valued.

The distribution μ_T^X of $(X_t)_{t\geq 0}$ over a finite time interval $[0, T]$ has a density with respect to the Wiener measure μ_T^W (the law of the Brownian motion W), given by the Cameron-Martin formula:

$$\frac{\mathrm{d}\mu_T^X}{\mathrm{d}\mu_T^W} = \exp\left[\int_0^T A(X_t, \theta)\mathrm{d}X_t - \frac{1}{2}\int_0^T A(X_t, \theta)^2 \mathrm{d}t\right], \tag{1.4}$$

which is in turn the likelihood function with continuous observation. In practice, only discrete values $X_{t_0}, \ldots, X_{t_n}$ may be observed over the duration $[0, T]$, where $0 = t_0 < t_1 < \cdots < t_n = T$ and $t_i - t_{i-1} = \delta$. The corresponding likelihood function $\widetilde{L}(\theta)$ is the conditional expectation under Wiener measure:

$$\mathbb{E}\left\{\widetilde{L}(\theta) \mid X_{t_0}, \ldots, X_{t_n}\right\} = \frac{\prod_{j=1}^{n} p_\theta\left(\delta, X_{t_{j-1}}, X_{t_j}\right)}{\prod_{j=1}^{n} G\left(\delta, X_{t_{j-1}}, X_{t_j}\right)}, \tag{1.5}$$

where $p_\theta(t, x, y)$ is the conditional probability density function of X_t given $X_0 = x$, and $G(t, x, y)$ is the Gaussian density $1/\sqrt{2\pi t}\exp\{-|x-y|^2/2t\}$ (see [1]). Since the denominator of (1.5) does not depend on θ, we may simply consider the numerator

$$L(X_{t_0}, \ldots, X_{t_n}) \equiv \prod_{j=1}^{n} p_\theta\left(\delta, X_{t_{j-1}}, X_{t_j}\right), \tag{1.6}$$

as a likelihood function. Therefore, the MLE for θ under a discrete observation may be found by solving either explicitly if possible or numerically the likelihood equation

$$\nabla L(X_{t_0}, \ldots, X_{t_n}) = 0. \tag{1.7}$$

The difficulty with this approach is that, unless for a very special drift vector field A, an explicit formula for $p_\theta(t, x, y)$ is not known. To overcome this difficulty, many approximation methods have been proposed in the literature by various authors. The idea is to replace the diffusion model (1.3) by an approximation model for which an explicit formula for the likelihood function is available. One possible candidate is of course the Euler-Maruyama approximation

$$\widetilde{X}_{t_j} - \widetilde{X}_{t_{j-1}} = A\left(\widetilde{X}_{t_{j-1}}, \theta\right)\delta + \xi_j\sqrt{\delta}, \tag{1.8}$$

where $\{\xi_j\}$ is an i.i.d. sequence with standard normal distribution $N(0,1)$ and $\widetilde{X}_0 = X_0$. However, the likelihood function $L_1(X_0,\ldots,X_n)$ for this model is not, in general, close enough to that of the diffusion model if measured in terms of the ratio of their corresponding likelihood functions

$$\frac{L(X_{t_0},\ldots,X_{t_n})}{L_1(X_{t_0},\ldots,X_{t_n})}. \tag{1.9}$$

The second approach is to discretize the likelihood function $\mathrm{d}\mu_T^X/\mathrm{d}\mu_T^W$ for continuous observations. In order to utilize this likelihood function, we need to handle the Itô integral $\int_0^T A(X_t,\theta)\mathrm{d}X_t$ which is defined only in probability sense. If $A = \nabla f$ (where f is a C^1-function) is a gradient field, then, according to Itô's formula,

$$\frac{\mathrm{d}\mu_T^X}{\mathrm{d}\mu_T^W} = \exp\left[f(X_T,\theta) - \int_0^T\left(\frac{1}{2}(\Delta f)(X_T,\theta) + \frac{1}{2}|A(X_T,\theta)|^2\right)\mathrm{d}t\right], \tag{1.10}$$

here the right-hand side involves only the sample X. This idea to get rid of Itô's integral and replace it by an ordinary one has far-reaching consequences, see the interesting paper [2] for some applications.

One can also use approximations to the probability density function $p_\theta(t,x,y)$ and construct functions which are close to the maximum likelihood function. There are a great number of articles devoted to this approach, such as [3–5], for example. The difficulty, however, is that even $f(t,x,y)$ is a uniform approximation of $p_\theta(t,x,y)$, there is no guarantee that the approximate likelihood function $\prod_j f(t,x_{j-1},x_j)$ would tend to $\prod_j p_\theta(t,x_{j-1},x_j)$ when $n \to \infty$.

In this paper we consider the linear diffusion approximation proposed by Shoji and Ozaki [6] to the diffusion model (1.3), which leads to the following approximation of the likelihood function $L(X_{t_0},\ldots,X_{t_n})$:

$$L_2(X_{t_0},\ldots,X_{t_n}) = \prod_{j=1}^{n} h_j\left(\delta, X_{t_{j-1}}, X_{t_j}\right), \tag{1.11}$$

where $t_j = jT/n$ so that X_{t_j} is a sample with fixed duration $\delta = t_j - t_{j-1}$ over $[0,T]$, and $h_j(t,x,y)$ is the probability transition density of the following linear diffusion model

$$d\widehat{X}_t = \left(A\left(X_{t_{j-1}},\theta\right) + A'\left(X_{t_{j-1}},\theta\right)\left(\widehat{X}_t - X_{t_{j-1}}\right)\right)\mathrm{d}t + \mathrm{d}W_t, \tag{1.12}$$

when $t_{j-1} \le t < t_j$ and $\widehat{X}_{t_{j-1}} = X_{t_{j-1}}$.

The approximation (1.12) is called the local linearization of the diffusion model (1.3), which has been studied in Shoji and Ozaki [6]. Shoji has showed numerically that the local linearizations do yield better estimates. Shoji's approximation was revisited in Prakasa Rao [7], without a definite conclusion.

The main goal of the paper is to prove Theorem 3.1 which implies that the local linear approximations (1.12) is efficient for the propose of deriving MLE with discrete samples.

The paper is organized as follows. In Section 2, we present the MLE for linear models such as (1.12). In Section 3, we state our main result for Shoji's local linear approximation, and give some comments about the conditions on the sampling data. Our main theorem provides a deterministic convergence rate for the likelihood functions. In Section 4, we prove that the likelihood function for the local linear approximation converges to the Cameron-Martin density but only in probability sense. Sections 5, 6, and 7 are devoted to the proof of our main result. In Section 5, we state the main tool, a representation formula for diffusions, established by Qian and Zheng [8]. In Section 6, we develop the main technical estimates in order to prove Theorem 3.1, whose proof is completed in Section 7. Section 8 contains a discussion about the Euler-Maruyama approximation which concludes the paper.

2. Linear Diffusions

Let us begin with the MLE of parameters a, b, and $\sigma > 0$ for the linear diffusion model (Mishra and Bishwal [9] discussed a similar model):

$$dX_t = (b - aX_t)\mathrm{d}t + \sigma \mathrm{d}W_t, \tag{2.1}$$

whose finite-dimensional distributions are Gaussian, determined through the probability transition function $h(t,x,y)$. Fortunately we have an explicit formula for h. Indeed the linear equation (2.1) may be solved explicitly and its solution is given by the formula

$$X_t = e^{-at}X_0 + \frac{b}{a}\left(1 - e^{-at}\right) + \sigma \int_0^t e^{-a(t-s)}\mathrm{d}W_s, \tag{2.2}$$

(formula (6.8) of Karatzas and Shreve [10], page 354), and therefore

$$h(t,x,y) = \frac{1}{\sqrt{\pi}\sigma}\sqrt{\frac{a}{1-e^{-2at}}}\exp\left[-\frac{a}{1-e^{-2at}}\frac{\left|y - e^{-at}x - (b/a)\left(1-e^{-at}\right)\right|^2}{\sigma^2}\right]. \tag{2.3}$$

Suppose we have a discrete sample observed over the equal time scale during the period $[0,T]$, $X_{iT/n}$, $i = 0,\ldots,n$. According to the Markov property, their joint distribution, or the maximum likelihood function

$$L(a,b,\sigma;x_0,\ldots,x_n) = \mu(x_0)\prod_{j=1}^{n} h\left(\delta, x_{j-1}, x_j\right), \tag{2.4}$$

where $\delta = T/n$, and $\mu(x)$ is the probability density function of the initial distribution. Therefore the logarithmic of the maximum likelihood function

$$\begin{aligned} l(a,b,\sigma;\{x_i\}) &= \log \mu(x_0) + \sum_{j=1}^{n} \log h\left(\frac{1}{n}, x_{j-1}, x_j\right) \\ &= \log \mu(x_0) - \frac{n}{2}\log \pi - n \log \sigma + \frac{n}{2}\log a - \frac{n}{2}\log\left(1 - e^{-2aT/n}\right) \\ &\quad - \frac{na}{1 - e^{-2aT/n}} \frac{1}{\sigma^2 T} \sum_{j=1}^{n} \left(x_j - \frac{b}{a} - \left(x_{j-1} - \frac{b}{a}\right)e^{-aT/n}\right)^2. \end{aligned} \tag{2.5}$$

The maximum likelihood estimates for a, b, and σ are the stationary points of l, that is solutions to the equation $\nabla l = 0$. Set $\rho = e^{-aT/n}$. Then $a = -(n/T)\log \rho$ and $\beta = b/a$.

Proposition 2.1. *The maximum likelihood estimates for the linear diffusion model* (2.1) *with discrete observations are given by*

$$\begin{aligned} \widehat{a} &= \frac{n}{T} \log \frac{\sum_{j=1}^{n} x_{j-1}^2 - (1/n)\left(\sum_{j=1}^{n} x_{j-1}\right)^2}{\sum_{j=1}^{n} x_{j-1}x_j - (1/n)\sum_{j=1}^{n} x_{j-1} \sum_{j=1}^{n} x_j}, \\ \widehat{\beta} &= \frac{1}{n}\sum_{j=1}^{n} x_j + \frac{1}{n}\frac{\widehat{\rho}}{1-\widehat{\rho}}(x_n - x_0), \\ \widehat{\sigma}^2 &= \frac{1}{T}\frac{2\widehat{a}}{1-\widehat{\rho}^2}\sum_{j=1}^{n}\left(x_j - \widehat{\rho}x_{j-1} - (1-\widehat{\rho})\widehat{\beta}\right)^2. \end{aligned} \tag{2.6}$$

As an interesting consequence we have the following.

Corollary 2.2. *The maximum likelihood estimators* $(\widehat{a}, \widehat{b}, \widehat{\sigma})$ *to the linear diffusion model* (2.1) *are not sufficient statistics while* $(\widehat{a}, \widehat{b}, \widehat{\sigma}, X_0, X_n)$ *are sufficient.*

3. Diffusion Models

We consider the diffusion model (1.3). Our approach and our conclusions are applicable to multidimensional cases as long as the diffusion coefficients are constant. For simplicity, we only consider one-dimensional case. The question is to estimate θ under a discrete observation $\{x_0, \ldots, x_n\}$ over the time scale δ in the time interval $[0, T]$. Then, up to a constant factor, its maximum likelihood function

$$L(x_0, \ldots, x_n) = \prod_{j=1}^{n} p(\delta, x_{j-1}, x_j), \tag{3.1}$$

where $p(t,z,y)$ is the transition probability density of (X_t) (we have dropped the subscript θ for simplicity). The approximation maximum likelihood function, proposed in [6], is given by

$$L_2(x_0,\dots,x_n)=\prod_{j=1}^{n}h_j(\delta,x_{j-1},x_j), \tag{3.2}$$

where $h_j(t,x,y)$ is the transition density function to the linear diffusion model

$$dX_t=\left(A(x_{j-1},\theta)+A'(x_{j-1},\theta)(X_t-x_{j-1})\right)dt+dW_t, \tag{3.3}$$

which is the first-order approximation to (1.3).

In what follows we assume that A has bounded first and second derivatives and

$$\left|A'(x,\theta)\right|,\qquad \left|A''(x,\theta)\right|\leq 2C_0, \tag{3.4}$$

for some constant $C_0>0$ independent of parameters θ.

The main result of the paper is follows.

Theorem 3.1. *Assume that $A'(\cdot,\theta)$ and $A''(\cdot,\theta)$ are bounded uniformly in θ. Let $T>0$ be a fixed time and $C>0$ be a constant. Suppose $\{x_j^n\}_{j\leq n}$ $(n=1,2,\dots)$ is a family of discrete samples such that*

$$\left|x_j^n-x_{j-1}^n\right|^2\leq C\delta(n),\qquad \left|x_j^n\right|\leq C, \tag{3.5}$$

for all pair (j,n) such that $j\leq n$, $n=1,2,\dots$, where $\delta(n)=T/n$. Then

$$\lim_{n\to\infty}\frac{L(x_0^n,\dots,x_n^n)}{L_2(x_0^n,\dots,x_n^n)}=1, \tag{3.6}$$

where L and L_2 are defined in (3.1) *and* (3.2) *with $\delta=\delta(n)=T/n$.*

The convergence in (3.6) happens in a deterministic sense, and therefore conditions such as $|x_j^n-x_{j-1}^n|^2\leq C\delta(n)$ and $|x_j^n|\leq C$ are reasonable. The first condition, that is $|x_j^n-x_{j-1}^n|^2\leq C\delta(n)$, just says the "variance" of the sample cannot be too big. Since

$$\sum_j\left|X_{jT/n}-X_{(j-1)T/n}\right|^2\longrightarrow T\quad\text{in probability}, \tag{3.7}$$

so that on average we should have $|x_j^n-x_{j-1}^n|^2\leq C\delta(n)$. Since (X_t) has continuous sample paths, so that $\{X(\omega)_t : t\in[0,T]\}$ for a fixed sample point ω is bounded. Since x_j^n are sampled from the fixed duration $[0,T]$, thus we can assume that $\{x_j^n\}$ is bounded, though here we have a countable many samples. It is possible to relax this constraint, for example, we may impose that $|x_j^n|\leq Cn^{\alpha}$ with $\alpha<1/2$, but for simplicity we only consider the bounded case. This condition is placed as a kind of "integrability" condition on the samples.

From the asymptotic of the transition density function $p(t,x,y)$, it is easy to see that

$$\lim_{\delta\to 0}\frac{p(\delta,x_{j-1},x_j)}{h_j(\delta,x_{j-1},x_j)}=1, \tag{3.8}$$

for each j, while, as our observation $\{x_0,\dots,x_n\}$ happens over a fixed time interval $[0,T]$, the ratio (3.6) as $n\to\infty$ is really an infinite product, its behavior thus depends on the global behavior of $p(t,x,y)$. Although there are many results about bounds of $p(t,x,y)$ in the literature (see [2, 11] e.g.), the best we could find are those which yield (3.8) uniformly in x_j, none of them yields the precise limit (3.6). In fact, the proof of (3.6) depends on careful estimates on $p(t,x,y)$ through a representation formula established in [8].

4. Linear Diffusion Approximations

Without losing generality, we may assume that $T=1$. Let $X_{j/n}$ be a discrete observation of the diffusion model (1.3) at $t_j=j/n$ $(j=0,\dots,n)$. For simplicity, write $X_{j/n}$ as X_j if no confusion may arise. Consider the family of linear diffusions

$$\mathrm{d}X_t^j=\left(A(X_{j-1},\theta)+A'(X_{j-1},\theta)\left(X_t^j-X_{j-1}\right)\right)\mathrm{d}t+\mathrm{d}W_t, \tag{4.1}$$

with $X^j_{(j-1)/n}=X_{j-1}$. Let

$$\begin{aligned} b_j&=A(X_{j-1},\theta)-A'(X_{j-1},\theta)X_{j-1},\\ a_j&=A'(X_{j-1},\theta). \end{aligned} \tag{4.2}$$

Then

$$X_t^j=\mathrm{e}^{a_j(t-t_{j-1})}X_{j-1}+\frac{\mathrm{e}^{a_j(t-t_{j-1})}-1}{a_j}b_j+\int_{t_{j-1}}^t\mathrm{e}^{a_j(t-s)}\mathrm{d}W_s, \tag{4.3}$$

so that

$$\begin{aligned} h_j(\delta,X_{j-1},X_j)&=\mathbb{P}\left(X^j_{t_j}=X_j\mid X^j_{t_{j-1}}=X_{j-1}\right)\\ &=\sqrt{\frac{2a_j}{e^{2a_j\delta}-1}}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-(a_j/(\mathrm{e}^{2a_j\delta}-1))(X_j-X_{j-1}-((\mathrm{e}^{a_j\delta}-1)/a_j)(b_j+a_jX_{j-1}))^2}, \end{aligned} \tag{4.4}$$

where $\delta=1/n$. The approximating likelihood function is

$$L_2(\delta)=\prod_{j=1}^n h_j(\delta,X_{j-1},X_j). \tag{4.5}$$

We need to compare this function to the likelihood function with continuous observation—the Cameron-Martin density, which, however, should be discounted with respect to the Wiener measure. Thus we have to renormalize $L_2(\delta)$ against the discrete version of Brownian motion, which is given by

$$\widehat{L}_2(\delta) = \prod_{j=1}^{n} \frac{h_j(\delta, X_{j-1}, X_j)}{G(\delta, X_{j-1}, X_j)}, \tag{4.6}$$

where

$$G(\delta, x, y) = \frac{1}{\sqrt{2\pi\delta}} \exp\left(-\frac{|x-y|^2}{2\delta}\right). \tag{4.7}$$

Hence its logarithmic

$$\begin{aligned} \widehat{l}_2(\delta) = &\frac{1}{2}\sum_j \log \frac{2\delta a_j}{e^{2a_j\delta}-1} + \frac{1}{2\delta}\sum_j |X_j - X_{j-1}|^2 \\ &- \frac{1}{2}\sum_j \frac{2a_j}{e^{2a_j\delta}-1}\left(X_j - X_{j-1} - \frac{e^{a_j\delta}-1}{a_j}(b_j + a_j X_{j-1})\right)^2. \end{aligned} \tag{4.8}$$

Proposition 4.1. *One has*

$$\lim_{\delta\downarrow 0} \widehat{l}_2(\delta) = l \tag{4.9}$$

uniformly in θ, in probability with respect to the Wiener measure, where l is the log of the Cameron-Martin density (1.4).

Proof. Let $D_j = X_j - X_{j-1}$. Then

$$\begin{aligned} \widehat{l}_2(\delta) = &\frac{1}{2}\sum_j \log \frac{2\delta a_j}{e^{2a_j\delta}-1} + \frac{1}{2\delta}\sum_j \left(1 - \frac{2\delta a_j}{e^{2a_j\delta}-1}\right) D_j^2 \\ &- \sum_j \frac{1}{e^{a_j\delta}+1}(b_j + a_j X_{j-1})^2 \frac{e^{a_j\delta}-1}{a_j} \\ &+ 2\sum_j \frac{1}{e^{a_j\delta}+1}(b_j + a_j X_{j-1}) D_j. \end{aligned} \tag{4.10}$$

Since $b_j = A(X_{j-1},\theta) - a_j X_{j-1}$ and $a_j = A'(X_{j-1},\theta)$, so that

$$\begin{aligned}\widehat{l}_2(\delta) &= \frac{1}{2}\sum_j \log\frac{2\delta a_j}{e^{2a_j\delta}-1} + \frac{1}{2\delta}\sum_j\left(1-\frac{2\delta a_j}{e^{2a_j\delta}-1}\right)D_j^2 \\ &\quad - \sum_j \frac{1}{e^{a_j\delta}+1}A\left(X_{j-1},\theta\right)^2\frac{e^{a_j\delta}-1}{a_j} \\ &\quad + 2\sum_j\frac{1}{e^{a_j\delta}+1}A\left(X_{j-1},\theta\right)D_j.\end{aligned} \tag{4.11}$$

However,

$$\begin{aligned}\lim_{\delta\downarrow 0} 2\sum_j \frac{1}{e^{a_j\delta}+1}A\left(X_{j-1},\theta\right)D_j &= \int_0^1 A(X_t,\theta)\mathrm{d}X_t, \\ \lim_{\delta\downarrow 0}\sum_j\frac{1}{e^{a_j\delta}+1}A\left(X_{j-1},\theta\right)^2\frac{e^{a_j\delta}-1}{a_j} &= \frac{1}{2}\int_0^1 |A|^2(X_t,\theta)\mathrm{d}t, \\ \lim_{\delta\downarrow 0}\frac{1}{2}\sum_j\log\frac{2\delta a_j}{e^{2a_j\delta}-1} &= -\frac{1}{2}\int_0^1 A'(X_t,\theta)\mathrm{d}t, \\ \lim_{\delta\downarrow 0}\frac{1}{2\delta}\sum_j\left(1-\frac{2\delta a_j}{e^{2a_j\delta}-1}\right)D_j^2 &= \frac{1}{2}\int_0^1 A'(X_t,\theta)\mathrm{d}t\end{aligned} \tag{4.12}$$

in probability. The claim thus follows immediately. □

5. A Representation Formula

From this section, we develop necessary estimates in order to prove Theorem 3.1. In this section, we recall the main tool in our proof, a representation formula proved by Qian and Zheng [8]. Based on this formula, we prove the main estimate (6.65), which has independent interest, in the next section. We conclude the proof of Theorem 3.1 in Section 7.

Let $x \in \mathbb{R}$. Consider the linear diffusion

$$dX_t = \left(A(x,\theta) + A'(x,\theta)(X_t - x)\right)\mathrm{d}t + \mathrm{d}W_t, \tag{5.1}$$

whose probability transition function is also denoted by $h(t,z,y)$. Recall that $p(t,z,y)$ is the probability transition function of the diffusion defined by (1.3). The strong solution of (5.1) is given by

$$X_t = X_0 + \sigma(a,t)^2(aX_0 + b) + \int_0^t e^{a(t-s)}\mathrm{d}W_s, \tag{5.2}$$

so that

$$h(t,z,y) = \frac{1}{\sqrt{2\pi}\sigma(2a,t)} \exp\left\{-\frac{1}{2\sigma(2a,t)^2}\left(y - z - \sigma(a,t)^2(b+az)\right)^2\right\}, \tag{5.3}$$

where $b = A(x,\theta) - xA'(x,\theta)$, $a = A'(x,\theta)$ and $\sigma(a,t) = \sqrt{(\mathrm{e}^{at}-1)/a}$.

Observe that for any $a \in \mathbb{R}$, $t \to \sigma(a,t)$ is increasing, and

$$\lim_{t\downarrow 0} \frac{\sigma(a,t)}{\sqrt{t}} = 1 \quad \text{for any } a. \tag{5.4}$$

We will also use the fact that

$$\sigma(2a,t) = \sqrt{\frac{\mathrm{e}^{2at}-1}{2a}} = \sqrt{\frac{\mathrm{e}^{at}+1}{2}}\sigma(a,t). \tag{5.5}$$

Lemma 5.1. *For $x \in \mathbb{R}$, and $C(z) = A(z,\theta) - A(x,\theta) - A'(x,\theta)(z-x)$. Then*

$$|C(z)| \le C_0 \min\left\{2|z-x|, \frac{1}{2}|z-x|^2\right\}. \tag{5.6}$$

Our main tool is a representation formula (5.7) discovered in [8]. Let $(X_t, \mathbb{P}^x)$ be the solution to the linear stochastic differential equation (5.1).

Proposition 5.2. *For $x, y \in \mathbb{R}$ and $T > 0$ one has*

$$\frac{p(T,x,y)}{h(T,x,y)} = 1 + \int_0^T \mathbb{P}^x\left\{U_t C(X_t) \frac{\nabla h(T-t, X_t, y)}{h(T,x,y)}\right\} dt, \tag{5.7}$$

where

$$U_t = \exp\left\{\int_0^t C(X_s)dW_s - \frac{1}{2}\int_0^t |C|^2(X_s)ds\right\}, \tag{5.8}$$

which is a martingale under the probability $\mathbb{P}^x$.

To prove (3.6), we need to estimate the double integral appearing on the right-hand side of (5.7), which requires a precise estimate for

$$I_t = \mathbb{P}^x\left\{U_t C(X_t) \frac{\nabla h(T-t, X_t, y)}{h(T,x,y)}\right\}, \tag{5.9}$$

which can be achieved since we know the precise form $h(T,x,y)$. Of course, if we knew the joint distribution of (U_t, X_t), our task would be easy, but unfortunately it is rarely the case.

Our arguments are based on the fact that (U_t) is a martingale under $\mathbb{P}^x$, together with some delicate estimates for the functional integral

$$\mathbb{P}^x\left\{\left|\frac{\nabla h(T-t,X_t,y)}{h(T,x,y)}\right|^p\right\}, \tag{5.10}$$

which will be done in the next section.

6. Main Estimates

We use the notations established in the previous section. Let $T>0$, $x,y\in\mathbb{R}$ and $d=y-x$. Then

$$\frac{\nabla_z h(T-t,z,y)}{h(T-t,z,y)}=\frac{2a\mathrm{e}^{a(T-t)}}{e^{2a(T-t)}-1}\left(y-\mathrm{e}^{a(T-t)}z-\sigma(a,T-t)^2b\right), \tag{6.1}$$

and therefore

$$\begin{aligned}\frac{\nabla_z h(T-t,X_t,y)}{h(T,x,y)}&=\frac{2a\mathrm{e}^{a(T-t)}}{\mathrm{e}^{2a(T-t)}-1}\frac{h(T-t,X_t,y)}{h(T,x,y)}\\&\quad\times\left(y-\mathrm{e}^{a(T-t)}X_t-\sigma(a,T-t)^2b\right)\\&=\frac{\sigma(2a,T)\mathrm{e}^{a(T-t)}}{\sigma(2a,T-t)^3}\mathrm{e}^{(1/2\sigma(2a,T)^2)S(T)^2}\\&\quad\times\left(y-\mathrm{e}^{a(T-t)}X_t-\sigma(a,T-t)^2b\right)\\&\quad\times\mathrm{e}^{-(1/2\sigma(2a,T-t)^2)|y-\mathrm{e}^{a(T-t)}X_t-\sigma(a,T-t)^2b|^2},\end{aligned} \tag{6.2}$$

where

$$S(T)=\left|y-x-\sigma(a,T)^2(ax+b)\right|. \tag{6.3}$$

For $t\in(0,T)$ and $p>1$ we set

$$\begin{aligned}\beta_p(t)&=\sqrt{\frac{a}{p(\mathrm{e}^{2aT}-1)-(p-1)(\mathrm{e}^{2a(T-t)}-1)}}\left|\sigma(a,T)^2(b+ax)-d\right|,\\\alpha_p(t)&=\sqrt{\frac{\mathrm{e}^{2a(T-t)}-1}{p(\mathrm{e}^{2aT}-1)-(p-1)(\mathrm{e}^{2a(T-t)}-1)}},\end{aligned} \tag{6.4}$$

for simplicity.

Lemma 6.1. *For any $p > 1$ one has*

$$\sqrt{\frac{1}{p}}\left|\sigma(a,T)^2(b+ax)-d\right| \leq \sqrt{\frac{e^{2aT}-1}{a}}\beta_p(t) \leq \left|\sigma(a,T)^2(b+ax)-d\right|,$$
$$\sqrt{\frac{1}{p}}\sqrt{\frac{e^{2a(T-t)}-1}{e^{2aT}-1}} \leq \alpha_p(t) \leq \sqrt{\frac{e^{2a(T-t)}-1}{e^{2aT}-1}}, \tag{6.5}$$

for all $t \in [0,T]$.

Proof. The two inequalities follow from the fact that

$$\sqrt{\frac{\mathrm{e}^{2aT}-1}{p(\mathrm{e}^{2aT}-1)-(p-1)(\mathrm{e}^{2a(T-t)}-1)}} \tag{6.6}$$

assumes its maximum 1 and minimum $\sqrt{1/p}$. □

Since

$$\frac{\beta_p(t)}{\alpha_p(t)} = \sqrt{\frac{a}{\mathrm{e}^{2a(T-t)}-1}}\left|\sigma(a,T)^2(b+ax)-d\right|,$$
$$\frac{\mathrm{e}^{2a(T-t)}-1}{\mathrm{e}^{2aT}-1}\frac{\beta_{2p}(t)^2}{\alpha_{2p}(t)^2} = \frac{a}{\mathrm{e}^{2aT}-1}\left|\sigma(a,T)^2(b+ax)-d\right|^2, \tag{6.7}$$

so that

$$\frac{\nabla_z h(T-t,X_t,y)}{h(T,x,y)} = \frac{\sigma(2a,T)}{\sigma(2a,T-t)^3}\mathrm{e}^{((\mathrm{e}^{2a(T-t)}-1)/(\mathrm{e}^{2aT}-1))(\beta_{2p}(t)^2/\alpha_{2p}(t)^2)+a(T-t)}$$
$$\times\left(y-\mathrm{e}^{a(T-t)}X_t-\sigma(a,T-t)^2 b\right)$$
$$\times\,\mathrm{e}^{-(1/2\sigma(2a,T-t)^2)|y-\mathrm{e}^{a(T-t)}X_t-\sigma(a,T-t)^2 b|^2}, \tag{6.8}$$

which yield, together with (5.7), the following.

Lemma 6.2. *One has*

$$\frac{p(T,x,y)}{h(T,x,y)} = 1+\int_0^T \frac{2ae^{a(T-t)}}{e^{2a(T-t)}-1}\sqrt{\frac{e^{2aT}-1}{e^{2a(T-t)}-1}}e^{((e^{2a(T-t)}-1)/(e^{aT}-1))(\beta_{2p}(t)^2/\alpha_{2p}(t)^2)}$$
$$\times\,\mathbb{P}^x\{U_t C(X_t)K(t)\}dt, \tag{6.9}$$

where

$$K(t) = \left(y - e^{a(T-t)} X_t - \frac{e^{a(T-t)} - 1}{a} b \right) \times e^{-(a/(e^{2a(T-t)}-1))|y - e^{a(T-t)} X_t - ((e^{a(T-t)}-1)/a) b|^2}. \tag{6.10}$$

Let

$$J(p)_t = \sqrt[p]{\mathbb{P}^x\left(|K(t)|^p\right)}$$
$$D(p)_t = \sqrt[2p]{\mathbb{P}^x\left(\mathrm{e}^{(2p/(2p-1)) \int_0^t |C|^2(X_s) \mathrm{d}s} |C(X_t)|^{2p}\right)}. \tag{6.11}$$

Lemma 6.3. *Choose $\zeta > 0$ such that*

$$\mathbb{P} \exp\left(\zeta \sup_{[0,1]} |W_t|^2 \right) = c < +\infty. \tag{6.12}$$

Then for any $T > 0$ and $\lambda > 0$, such that

$$\frac{4\lambda C_0}{e^{aT}} \frac{e^{2aT} - 1}{a} \leq \zeta, \tag{6.13}$$

one has

$$\mathbb{P}^x\left(e^{\lambda \int_0^t |C|^2(X_s) ds} \right) \leq c \quad \forall t \leq T. \tag{6.14}$$

Proof. Let

$$Y_t = X_t - x - \sigma(a,t)^2 (ax + b)$$
$$= \int_0^t \mathrm{e}^{a(t-s)} \mathrm{d}W_s. \tag{6.15}$$

Then

$$|C|^2(X_s) \leq 4C_0 |X_s - x|^2$$
$$\leq 4C_0 \left| Y_s + \sigma(a,s)^2 (ax + b) \right|^2 \tag{6.16}$$
$$\leq 4C_0 \sigma(a,s)^4 |(ax+b)|^2 + 4C_0 C_2 |Y_s|^2,$$

and therefore

$$\mathbb{P}^x\left(\mathrm{e}^{\lambda \int_0^t |C|^2(X_s) \mathrm{d}s} \right) \leq \mathrm{e}^{4\lambda C_0 \sigma(a,t)^4 |(ax+b)|^2 t} \mathbb{P}^x\left(\mathrm{e}^{4\lambda C_0 \int_0^t |Y_s|^2 \mathrm{d}s} \right). \tag{6.17}$$

On the other hand

$$Y_s = e^{as}\int_0^s \mathrm{e}^{-ar}\mathrm{d}W_r, \tag{6.18}$$

so that $M_s \equiv e^{-as}Y_s$ is a martingale with

$$\langle M\rangle_s = \int_0^s \mathrm{e}^{-2ar}\mathrm{d}r = \frac{\mathrm{e}^{2as}-1}{2a}. \tag{6.19}$$

Thus M_s is a time change of a standard Brownian motion, and

$$M_s = B_{(\mathrm{e}^{2as}-1/2a)}, \tag{6.20}$$

for some standard Brownian motion $(B_t)_{t\geq 0}$. Since

$$2\lambda C_0\frac{1}{\mathrm{e}^{at}}\frac{\mathrm{e}^{2at}-1}{2a} \leq \zeta, \tag{6.21}$$

so we have

$$\begin{aligned}
\mathbb{P}^x\left\{\mathrm{e}^{4\lambda C_0\int_0^t \mathrm{e}^{as}|M_s|^2\mathrm{d}s}\right\} &= \mathbb{P}\exp\left(4\lambda C_0\int_0^t e^{as}\left|B_{(\mathrm{e}^{2as}-1)/2a}\right|^2\mathrm{d}s\right)\\
&= \mathbb{P}\exp\left(4\lambda C_0\int_0^{(\mathrm{e}^{2at}-1)/2a}\frac{s}{\sqrt{2as+1}}|B_s|^2\mathrm{d}s\right)\\
&\leq \mathbb{P}\exp\left(4\lambda C_0\frac{1}{\mathrm{e}^{at}}\left(\frac{\mathrm{e}^{2at}-1}{2a}\right)^2\sup_{[0,((\mathrm{e}^{2at}-1)/2a)]}|B_s|^2\right)\\
&= \mathbb{P}\exp\left(4\lambda C_0\frac{1}{\mathrm{e}^{at}}\frac{\mathrm{e}^{2at}-1}{2a}\sup_{[0,1]}|B_s|^2\right)\\
&\leq c.
\end{aligned} \tag{6.22}$$

□

Corollary 6.4. *For $p > 1$ and $T > 0$ to be such that*

$$\frac{16pC_0}{2p-1}\frac{1}{e^{aT}}\frac{e^{2aT}-1}{2a} \leq \zeta, \tag{6.23}$$

one has

$$\mathbb{P}^x\left\{e^{(4p/(2p-1))\int_0^t |\mathrm{C}|^2(X_s)ds}\right\} \leq c \quad \forall t \leq T. \tag{6.24}$$

In what follows, we always assume that $T > 0$ is chosen such that the condition (6.23) is satisfied. Next we estimate $D(p)_t$, which is provided in the following.

Lemma 6.5. *Let* $p > 1$. *Then*

$$D(p)_t \le C_1\left(\sigma(2a,t)^2 + \sigma(a,t)^2(ax+b)^2\right), \tag{6.25}$$

where the positive constant C_1 *depends only on* p, ζ, *and* C_0.

Proof. Let

$$F(t) = \mathbb{P}^x\left(e^{(2p/(2p-1))\int_0^t |C|^2(X_s)ds}|C(X_t)|^{2p}\right). \tag{6.26}$$

Then, by the Hölder inequality

$$\begin{aligned} F(t) &= \sqrt{\mathbb{P}^x\left(e^{(4p/(2p-1))\int_0^t |C|^2(X_s)ds}\right)}\sqrt{\mathbb{P}^x\left(|C(X_t)|^{4p}\right)} \\ &\le \sqrt{C_0}\sqrt{\mathbb{P}^x\left(|C(X_t)|^{4p}\right)}. \end{aligned} \tag{6.27}$$

Next we estimate the expectation $\mathbb{P}^x|C(X_t)|^{4p}$. Since

$$\begin{aligned} |C(X_t)| &\le \frac{C_0}{2}|X_t - x|^2 \\ &\le C_0|Y_t|^2 + C_0\sigma(a,T)^2(ax+b)^2, \end{aligned} \tag{6.28}$$

so that

$$|C(X_t)|^{4p} \le 2^{4p-1}C_0^{4p}|Y_t|^{8p} + 2^{4p-1}C_0^{4p}\sigma(a,T)^{8p}(ax+b)^{8p}. \tag{6.29}$$

On the other hand

$$\begin{aligned} \mathbb{P}^x|Y_t|^{8p} &= \frac{1}{\sqrt{2\pi}\sigma(2a,t)}\int |z|^{8p}\exp\left(-\frac{|z|^2}{\sigma(2a,t)^2}\right)dz \\ &= \sigma(2a,t)^{8p}, \end{aligned} \tag{6.30}$$

so that

$$\sqrt[4p]{\mathbb{P}^x|C(X_t)|^{4p}} \le 2^{(4p-1)/4p}C_0\sigma(2a,t)^2 + 2^{(4p-1)/4p}C_0\sigma(a,t)^2(ax+b)^2. \tag{6.31}$$

□

Lemma 6.6. *Let $T > 0$ satisfy condition* (6.23), $x,\ y \in \mathbb{R}$, *and* $p > 1$ *and* $q > 1$ *such that* $(1/p) + (1/q) = 1$. *Then*

$$\frac{p(T,x,y)}{h(T,x,y)} \leq 1 + \exp\left(\frac{e^{2a(T-t)}-1}{e^{aT}-1}\frac{\beta_{2p}(t)^2}{\alpha_{2p}(t)^2}\right) \times \int_0^T \frac{2ae^{a(T-t)}}{e^{2a(T-t)}-1}\sqrt{\frac{e^{2aT}-1}{e^{2a(T-t)}-1}}D(p)_t J(2p)_t dt. \tag{6.32}$$

Proof. Since

$$\left(U_t \mathrm{e}^{-((q-1)/2)\int_0^t |\mathrm{C}|^2(\mathrm{X}_s)\mathrm{d}s}\right)^q = \exp\left\{\int_0^t qC(\mathrm{X}_s)\mathrm{d}W_s - \frac{1}{2}q^2\int_0^t |\mathrm{C}|^2(\mathrm{X}_s)\mathrm{d}s\right\} \tag{6.33}$$

is a martingale under $\mathbb{P}^x$, so that

$$\mathbb{P}^x\left(U_t \mathrm{e}^{-((q-1)/2)\int_0^t |\mathrm{C}|^2(\mathrm{X}_s)\mathrm{d}s}\right)^q = 1. \tag{6.34}$$

By the Hölder inequality we deduce that

$$|\mathbb{P}^x(U_t C(\mathrm{X}_t)K(t))| \leq D(p)_t J(2p)_t. \tag{6.35}$$

Equation (6.32), follows from the representation (6.9). □

Lemma 6.7. *Let $p > 1$. Then*

$$J(p)_t \leq \sqrt[p]{2^{p-1}}\alpha_p(t)^{1+(1/p)}e^{-\beta_p^2} \times \left\{\sqrt[p]{\varepsilon_p}\sqrt{\frac{e^{2aT}-e^{2a(T-t)}}{2a}} + \sqrt{\frac{e^{2a(T-t)}-1}{a}}\beta_p(t)\right\}. \tag{6.36}$$

Proof. We have

$$\mathbb{P}^x\{|K(t)|^p\} \leq \mathbb{P}^x\left\{\left|y - e^{a(T-t)}\mathrm{X}_t - \frac{e^{a(T-t)}-1}{a}b\right|^p \times \mathrm{e}^{-(pa/(e^{2a(T-t)}-1))|y-e^{a(T-t)}\mathrm{X}_t-((e^{a(T-t)}-1)/a)b|^2}\right\}. \tag{6.37}$$

Under the probability $\mathbb{P}^x$,

$$\begin{aligned} Z_t &\equiv e^{a(T-t)}\left(X_t - e^{at}x - \frac{e^{at}-1}{a}b\right) \\ &= \int_0^t e^{a(T-s)}\mathrm{d}W_s \end{aligned} \tag{6.38}$$

is a central normal distribution with variance

$$\int_0^t e^{2a(T-s)}\mathrm{d}s = \frac{\mathrm{e}^{2aT} - \mathrm{e}^{2a(T-t)}}{2a}. \tag{6.39}$$

In terms of Z_t and $d = y - x$

$$\begin{aligned} \mathbb{P}^x\left\{|K(t)|^p\right\} \le \mathbb{P}^x\Bigg\{&\left|Z_t + \frac{e^{aT}-1}{a}(b+ax) - d\right|^p \\ &\times \mathrm{e}^{-(pa/(e^{2a(T-t)}-1))|Z_t+((e^{aT}-1)/a)(b+ax)-d|^2}\Bigg\}. \end{aligned} \tag{6.40}$$

Making change of variable

$$N = \sqrt{\frac{2a}{\mathrm{e}^{2aT} - \mathrm{e}^{2a(T-t)}}}Z_t. \tag{6.41}$$

Then, under $\mathbb{P}^x$, N has the standard normal distribution $N(0,1)$, so that

$$\begin{aligned} \mathbb{P}^x|K(t)|^p \le \mathbb{P}\Bigg\{&\left|\sqrt{\frac{\mathrm{e}^{2aT}-\mathrm{e}^{2a(T-t)}}{2a}}N + \frac{e^{aT}-1}{a}(b+ax) - d\right|^p \\ &\times \mathrm{e}^{-(pa/(e^{2a(T-t)}-1))|\sqrt{((\mathrm{e}^{2aT}-\mathrm{e}^{2a(T-t)})/2a)}N+((e^{aT}-1)/a)(b+ax)-d|^2}\Bigg\} \\ = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}\Bigg\{&\left|\sqrt{\frac{\mathrm{e}^{2aT}-\mathrm{e}^{2a(T-t)}}{2a}}z + \frac{e^{aT}-1}{a}(b+ax) - d\right|^p \\ &\times \mathrm{e}^{-(pa/(e^{2a(T-t)}-1))|\sqrt{((\mathrm{e}^{2aT}-\mathrm{e}^{2a(T-t)})/2a)}z+((e^{aT}-1)/a)(b+ax)-d|^2-(z^2/2)}\Bigg\}dz. \end{aligned} \tag{6.42}$$

Let us simplify the last integral. Indeed, set

$$\eta = \sqrt{\frac{p\mathrm{e}^{2aT} - (p-1)\mathrm{e}^{2a(T-t)} - 1}{e^{2a(T-t)} - 1}}\, z,$$
$$E_t = 2p\sqrt{\frac{\mathrm{e}^{2aT} - \mathrm{e}^{2a(T-t)}}{2\left(e^{2a(T-t)} - 1\right)}}\left(\frac{e^{aT}-1}{a}(b+ax) - d\right) \times \sqrt{\frac{a}{p\mathrm{e}^{2aT} - (p-1)\mathrm{e}^{2a(T-t)} - 1}}. \tag{6.43}$$

Then we rewrite the term appearing in the exponential in the last line of (6.42)

$$\begin{aligned} &-\frac{pa}{e^{2a(T-t)}-1}\left|\sqrt{\frac{\mathrm{e}^{2aT} - \mathrm{e}^{2a(T-t)}}{2a}}z + \frac{e^{aT}-1}{a}(b+ax) - d\right|^2 \\ &= -\frac{1}{2}\left(\eta^2 + 2E_t\eta\right) - \frac{pa}{e^{2a(T-t)}-1}\left(\frac{e^{aT}-1}{a}(b+ax) - d\right)^2 \\ &= -\frac{1}{2}(\eta + E_t)^2 + \frac{1}{2}E_t^2 - \frac{pa}{e^{2a(T-t)}-1}\left(\frac{e^{aT}-1}{a}(b+ax) - d\right)^2, \end{aligned} \tag{6.44}$$

together with

$$\begin{aligned} &\frac{1}{2}E_t^2 - \frac{pa}{e^{2a(T-t)}-1}\left(\frac{e^{aT}-1}{a}(b+ax) - d\right)^2 \\ &\quad = -\frac{pa}{p\mathrm{e}^{2aT} - (p-1)\mathrm{e}^{2a(T-t)} - 1}\left(\frac{e^{aT}-1}{a}(b+ax) - d\right)^2 \end{aligned} \tag{6.45}$$

the inequality (6.42) may be rewritten as follows:

$$\begin{aligned} \mathbb{P}^x|K(t)|^p \le{}& \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-(pa/(p\mathrm{e}^{2aT}-(p-1)\mathrm{e}^{2a(T-t)}-1))(((e^{aT}-1)/a)(b+ax)-d)^2} \\ &\times \int_{\mathbb{R}}\left|\sqrt{\frac{\mathrm{e}^{2aT} - \mathrm{e}^{2a(T-t)}}{2a}}z + \frac{\mathrm{e}^{aT}-1}{a}(b+ax) - d\right|^p \\ &\times \mathrm{e}^{-(1/2)(\eta+E_t)^2}\mathrm{d}z. \end{aligned} \tag{6.46}$$

Making change of variable in the last integral

$$v = \eta + E_t = \sqrt{\frac{p\mathrm{e}^{2aT} - (p-1)\mathrm{e}^{2a(T-t)} - 1}{e^{2a(T-t)} - 1}} z + E_t, \tag{6.47}$$

so that

$$\mathrm{d}z = \sqrt{\frac{e^{2a(T-t)} - 1}{p\mathrm{e}^{2aT} - (p-1)\mathrm{e}^{2a(T-t)} - 1}} \mathrm{d}v,$$

$$\begin{aligned}\sqrt{\frac{\mathrm{e}^{2aT} - \mathrm{e}^{2a(T-t)}}{2a}} z + \frac{e^{aT} - 1}{a}(b + ax) - d &= \sqrt{\frac{\left(e^{2a(T-t)} - 1\right)\left(\mathrm{e}^{2aT} - \mathrm{e}^{2a(T-t)}\right)}{2a\left(p\mathrm{e}^{2aT} - (p-1)\mathrm{e}^{2a(T-t)} - 1\right)}} v \\ &+ \frac{\mathrm{e}^{2a(T-t)} - 1}{p\mathrm{e}^{2aT} - (p-1)\mathrm{e}^{2a(T-t)} - 1}\left(\frac{e^{aT} - 1}{a}(b + ax) - d\right).\end{aligned} \tag{6.48}$$

Thus (6.46) yields that

$$\begin{aligned}\mathbb{P}^x|K(t)|^p \le{}& \left(\sqrt{\frac{\mathrm{e}^{2a(T-t)} - 1}{2a}}\right)^{p+1} \sqrt{\frac{2a}{p\mathrm{e}^{2aT} - (p-1)\mathrm{e}^{2a(T-t)} - 1}} \\ &\times \mathrm{e}^{-(pa/(p\mathrm{e}^{2aT} - (p-1)\ \mathrm{e}^{2a(T-t)} - 1))(((e^{aT}-1)/a)(b+ax)-d)^2} Q_t,\end{aligned} \tag{6.49}$$

where

$$\begin{aligned}Q_t ={}& \int_{\mathbb{R}} \left| \sqrt{\frac{\mathrm{e}^{2aT} - \mathrm{e}^{2a(T-t)}}{p\mathrm{e}^{2aT} - (p-1)\mathrm{e}^{2a(T-t)} - 1}} z \right. \\ &\left. + \frac{\sqrt{2a\left(\mathrm{e}^{2a(T-t)} - 1\right)}\left(\left(\left(e^{aT} - 1\right)/a\right)(b + ax) - d\right)}{p\mathrm{e}^{2aT} - (p-1)\mathrm{e}^{2a(T-t)} - 1} \right|^p \frac{\mathrm{e}^{-(1/2)z^2}}{\sqrt{2\pi}} \mathrm{d}z \\ \le{}& 2^{p-1}\varepsilon_p \left| \sqrt{\frac{\mathrm{e}^{2aT} - \mathrm{e}^{2a(T-t)}}{p\mathrm{e}^{2aT} - (p-1)\mathrm{e}^{2a(T-t)} - 1}} \right|^p \\ &+ 2^{p-1} \left| \frac{\sqrt{2a\left(\mathrm{e}^{2a(T-t)} - 1\right)}\left(\left(\left(e^{aT} - 1\right)/a\right)(b + ax) - d\right)}{p\mathrm{e}^{2aT} - (p-1)\mathrm{e}^{2a(T-\mathrm{t})} - 1} \right|^p\end{aligned} \tag{6.50}$$

$$\varepsilon_p = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} |z|^p \mathrm{e}^{-(1/2)z^2} \mathrm{d}z.$$

Therefore

$$
\begin{aligned}
J(p)_t \leq{} & \left(\sqrt{\frac{\mathrm{e}^{2a(T-t)}-1}{2a}}\right)^{1+(1/p)}\left(\frac{2a}{p\mathrm{e}^{2aT}-(p-1)\mathrm{e}^{2a(T-t)}-1}\right)^{1/(2p)} \\
& \times \mathrm{e}^{-(a/(p\mathrm{e}^{2aT}-(p-1)\mathrm{e}^{2a(T-t)}-1))(((e^{aT}-1)/a)(b+ax)-d)^2} \\
& \times \left\{\sqrt[p]{2^{p-1}\varepsilon_p}\left|\sqrt{\frac{\mathrm{e}^{2aT}-\mathrm{e}^{2a(T-t)}}{p\mathrm{e}^{2aT}-(p-1)\mathrm{e}^{2a(T-t)}-1}}\right|\right. \\
& \left.+\sqrt[p]{2^{p-1}}\left|\sqrt{\frac{2(\mathrm{e}^{2a(T-t)}-1)}{p\mathrm{e}^{2aT}-(p-1)\mathrm{e}^{2a(T-t)}-1}}\right|\sqrt{\frac{a\left|((e^{aT}-1)/a)(b+ax)-d\right|^2}{p\mathrm{e}^{2aT}-(p-1)\mathrm{e}^{2a(T-t)}-1}}\right\},
\end{aligned}
\tag{6.51}
$$

which is equivalent to the required inequality. □

Lemma 6.8. *Let*

$$
H(p)_t \equiv \exp\left(\frac{e^{2a(T-t)}-1}{e^{aT}-1}\frac{\beta_{2p}(t)^2}{\alpha_{2p}(t)^2}\right)J(2p)_t. \tag{6.52}
$$

Then

$$
\frac{p(T,x,y)}{h(T,x,y)} \leq 1+\int_0^T \frac{2ae^{a(T-t)}}{e^{2a(T-t)}-1}\sqrt{\frac{e^{2aT}-1}{e^{2a(T-t)}-1}}D(p)_t H(p)_t dt, \tag{6.53}
$$

$$
\begin{aligned}
H(p)_t \leq{} & \sqrt[2p]{2^{2p-1}}\alpha_{2p}(t)^{1+(1/2p)}e^{-(2p-1)((e^{2aT}-e^{2a(T-t)})^2/(e^{2a(T-t)}-1)(e^{2aT}-1))\beta_{2p}(t)^2} \\
& \times\left\{\sqrt[p]{\varepsilon_p}\sqrt{\frac{e^{2aT}-e^{2a(T-t)}}{2a}}+\sqrt{\frac{e^{2a(T-t)}-1}{a}}\beta_{2p}(t)\right\}.
\end{aligned}
\tag{6.54}
$$

Proof. Indeed, by Lemma 6.7 we have

$$
\begin{aligned}
H(p)_t \leq{} & \sqrt[2p]{2^{2p-1}}\alpha_{2p}(t)^{1+(1/2p)}\mathrm{e}^{-\beta_{2p}^2+((\mathrm{e}^{2a(T-t)}-1)/(\mathrm{e}^{2aT}-1))(\beta_{2p}(t)^2/\alpha_{2p}(t)^2)} \\
& \times\left\{\sqrt[p]{\varepsilon_p}\sqrt{\frac{\mathrm{e}^{2aT}-\mathrm{e}^{2a(T-t)}}{2a}}+\sqrt{\frac{\mathrm{e}^{2a(T-t)}-1}{a}}\beta_p(t)\right\}.
\end{aligned}
\tag{6.55}
$$

On the other hand

$$
-\beta_{2p}^2+\frac{\mathrm{e}^{2a(T-t)}-1}{\mathrm{e}^{2aT}-1}\frac{\beta_{2p}(t)^2}{\alpha_{2p}(t)^2} = (p-1)\beta_{2p}(t)^2\frac{\mathrm{e}^{2\alpha T}-\mathrm{e}^{2\alpha(T-t)}}{\mathrm{e}^{2aT}-1}, \tag{6.56}
$$

thus

$$H(p)_t \leq \sqrt[2p]{2^{2p-1}}\alpha_{2p}(t)^{1+(1/2p)} \mathrm{e}^{(p-1)\beta_{2p}(t)^2((\mathrm{e}^{2\alpha T}-\mathrm{e}^{2\alpha(T-t)})/(\mathrm{e}^{2\alpha T}-1))} \times \left\{ \sqrt[p]{\varepsilon_p}\sqrt{\frac{\mathrm{e}^{2aT}-\mathrm{e}^{2a(T-t)}}{2a}} + \sqrt{\frac{\mathrm{e}^{2a(T-t)}-1}{a}}\beta_{2p}(t) \right\}. \tag{6.57}$$

□

Let

$$B(p)_t \equiv \mathrm{e}^{(p-1)\beta_{2p}(t)^2((\mathrm{e}^{2\alpha T}-\mathrm{e}^{2\alpha(T-t)})/(\mathrm{e}^{2\alpha T}-1))} \times \left\{ \sqrt[p]{\varepsilon_p}\sqrt{\frac{\mathrm{e}^{2aT}-\mathrm{e}^{2a(T-t)}}{2a}} + \sqrt{\frac{\mathrm{e}^{2a(T-t)}-1}{a}}\beta_{2p}(t) \right\}. \tag{6.58}$$

Then (6.53) and (6.54) imply that

$$\frac{p(T,x,y)}{h(T,x,y)} \leq 1 + \frac{\sqrt[2p]{2^{2p-1}}}{(\mathrm{e}^{2aT}-1)^{1/4p}} \int_0^T \frac{2a\mathrm{e}^{a(T-t)}}{(\mathrm{e}^{2a(T-t)}-1)^{1-(1/4p)}} D(p)_t B(p)_t \mathrm{d}t. \tag{6.59}$$

Lemma 6.9. *One has*

$$B(p)_t \leq e^{(p-1)\beta_{2p}(t)^2((e^{2\alpha T}-e^{2\alpha(T-t)})/(e^{2\alpha T}-1))} \times \left\{ \sqrt[p]{\varepsilon_p}\sqrt{\frac{e^{2aT}-e^{2a(T-t)}}{2a}} + \sqrt{\frac{e^{2a(T-t)}-1}{e^{2aT}-1}}S(T) \right\}. \tag{6.60}$$

In particular

$$B(p)_t \leq \sqrt[p]{\varepsilon_p}\sqrt{\frac{e^{2aT}-1}{2a}} + S(T),$$

$$\frac{p(T,x,y)}{h(T,x,y)} \leq 1 + \frac{\sqrt[2p]{2^{2p-1}}}{(e^{2aT}-1)^{1/4p}} \left\{ \sqrt[p]{\varepsilon_p}\sqrt{\frac{e^{2aT}-1}{2a}} + S(T) \right\} \times \int_0^T \frac{2ae^{a(T-t)}e^{(p-1)\beta_{2p}(t)^2((e^{2\alpha T}-e^{2\alpha(T-t)})/(e^{2\alpha T}-1))}}{(e^{2a(T-t)}-1)^{1-(1/4p)}} D(p)_t dt. \tag{6.61}$$

Proof. Let

$$G(p)_t = \frac{2ae^{a(T-t)}}{e^{2a(T-t)}-1}\sqrt{\frac{e^{2aT}-1}{e^{2a(T-t)}-1}} \sqrt[2p]{2^{2p-1}}\alpha_{2p}(t)^{1+(1/2p)}. \tag{6.62}$$

Then

$$\frac{p(T,x,y)}{h(T,x,y)} \le 1 + \int_0^T G(p)_t D(p)_t B(p)_t \mathrm{d}t. \tag{6.63}$$

Since

$$\begin{aligned} G(p)_t &= \frac{2ae^{a(T-t)}}{\left(e^{2a(T-t)}-1\right)^{1-(1/4p)}\left(e^{2aT}-1\right)^{1/4p}} \sqrt[2p]{2^{2p-1}} \\ &\quad\times\left(\sqrt{\frac{e^{2aT}-1}{2p\left(\mathrm{e}^{2aT}-1\right)-(2p-1)\left(\mathrm{e}^{2a(T-t)}-1\right)}}\right)^{1+(1/2p)} \\ &\le \frac{2ae^{a(T-t)}\sqrt[2p]{2^{2p-1}}}{\left(e^{2a(T-t)}-1\right)^{1-(1/4p)}\left(e^{2aT}-1\right)^{1/4p}}, \end{aligned} \tag{6.64}$$

which implies the required estimate. □

By collecting all estimates we have established, we may obtain the following.

Proposition 6.10. *There is a constant $C_2 > 0$ depending only on ζ and C_d such that*

$$\begin{aligned} \frac{p(T,x,y)}{h(T,x,y)} &\le 1 + C_2 e^{|a|T} e^{(p-1)(a/(e^{2aT}-1))S(T)^2}\left(\frac{e^{aT}+1}{2}+(ax+b)^2\right) \\ &\quad\times\frac{e^{aT}-1}{a}\left(\sqrt[p]{\varepsilon_p}\sqrt{\frac{e^{2aT}-1}{2a}}+S(T)\right), \end{aligned} \tag{6.65}$$

where

$$S(T) = \left|\frac{e^{aT}-1}{a}(b+ax)-(y-x)\right|. \tag{6.66}$$

Proof. Indeed

$$\begin{aligned} \frac{p(T,x,y)}{h(T,x,y)} &\le 1 + C_1 \mathrm{e}^{(p-1)(a/(e^{2aT}-1))S(T)^2}\left(\frac{\mathrm{e}^{2aT}-1}{a}\right)^{-1/4p}\left(\sqrt[p]{\varepsilon_p}\sqrt{\frac{\mathrm{e}^{2aT}-1}{2a}}+S(T)\right) \\ &\quad\times\int_0^T \mathrm{e}^{a(T-t)}\left(\frac{\mathrm{e}^{2a(T-t)}-1}{a}\right)^{(1/4p)-1}\left(\sigma(2a,t)^2+\sigma(a,t)^2(ax+b)^2\right)\mathrm{d}t \end{aligned}$$

$$
\begin{aligned}
&\leq 1 + C e^{(p-1)(a/(e^{2aT}-1))S(T)^2}\left(\frac{e^{2aT}-1}{a}\right)^{-1/4p}\left(\sqrt[p]{\varepsilon_p}\sqrt{\frac{e^{2aT}-1}{2a}}+S(T)\right)\\
&\quad\times\left(\frac{e^{aT}+1}{2}+(ax+b)^2\right)\times\frac{e^{aT}-1}{a}\int_0^T e^{a(T-t)}\left(\frac{e^{2a(T-t)}-1}{a}\right)^{(1/4p)-1}dt.
\end{aligned}
\tag{6.67}
$$

While,

$$
\begin{aligned}
\int_0^T \frac{e^{a(T-t)}}{\left((e^{2a(T-t)}-1)/2a\right)^{1-(1/4p)}}dt &= \int_0^{(e^{2aT}-1)/2a}\frac{1}{s^{1-(1/4p)}}\frac{1}{\sqrt{1+2as}}ds\\
&\leq 4pe^{|a|T}\left(\frac{e^{2aT}-1}{2a}\right)^{1/4p},
\end{aligned}
\tag{6.68}
$$

and it thus yields our key estimate (6.65). □

Similarly we have a lower bound

$$
\begin{aligned}
\frac{p(T,x,y)}{h(T,x,y)} &\geq 1 - C_3 e^{|a|T}e^{(p-1)(a/(e^{2aT}-1))S(T)^2}\left(\frac{e^{aT}+1}{2}+(ax+b)^2\right)\\
&\quad\times\frac{e^{aT}-1}{a}\left(\sqrt[p]{\varepsilon_p}\sqrt{\frac{e^{2aT}-1}{2a}}+S(T)\right),
\end{aligned}
\tag{6.69}
$$

where C_3 depends only on ζ and C_0.

7. Proof of Theorem 3.1

We are now in a position to prove Theorem 3.1. We may assume that $T = 1$, so that $\delta(n) = 1/n$. Let x_j^n $(j = 0, 1, \ldots, n)$ be discrete samplings with time scale $\delta = \delta(n) = 1/n$ on $[0, 1]$. By our assumptions, $|x_j^n - x_{j-1}^n|^2 \leq C\delta(n)$, and $|x_j^n| \leq C$ for all pair (j, n) such that $0 \leq j \leq n$ and $n \geq 1$. For simplicity we write x_j for x_j^n if no confusion may arise.

In the proof below, we will use C_i to denote nonnegative constants which may depend on C, $T(= 1)$ and the bounds of A' and A'' appearing in our diffusion model (1.3), but independent of n.

Recall that $h_j(t, x, y)$ is the probability transition density function of the diffusion (3.3), that is,

$$
dX_t = \left(b_j + a_j X_t\right)dt + dW_t, \tag{7.1}
$$

where

$$
b_j = A(x_{j-1}, \theta) - x_{j-1}A'(x_{j-1}, \theta), \qquad a_j = A'(x_{j-1}, \theta). \tag{7.2}
$$

According to (6.65) we have

$$\frac{p(\delta,x_{j-1},x_j)}{h_j(\delta,x_{j-1},x_j)} \le 1 + C_8 e^{|a_j|\delta} e^{((p-1)a_j/(e^{2a_j\delta}-1))\, S_j^2}\left(\frac{e^{a_j\delta}+1}{2} + \left(a_j x_{j-1}+b_j\right)^2\right) \times \frac{e^{a_{j-1}\delta}-1}{a_{j-1}}\left(\sqrt[p]{\varepsilon_p}\sqrt{\frac{e^{2a_{j-1}\delta}-1}{2a_{j-1}}} + S_j\right), \tag{7.3}$$

where

$$S_j = \left|\frac{e^{a_j\delta}-1}{a_j}\left(b_j + a_j x_{j-1}\right) - \left(x_j - x_{j-1}\right)\right|. \tag{7.4}$$

Since a_j and x_j are bounded,

$$\begin{aligned} \left|a_j x_{j-1} + b_j\right| &= \left|A(x_{j-1},\theta)\right| \le C_4\left(1+\left|x_{j-1}\right|\right) \\ &\le C_9, \end{aligned} \tag{7.5}$$

so that

$$S_j \le \frac{e^{a_j\delta}-1}{a_j\delta} C_9\delta + C\sqrt{\delta} \le C_{10}\sqrt{\delta}. \tag{7.6}$$

Thus

$$\begin{aligned} \frac{(p-1)a_j}{e^{2a_j\delta}-1} S_j^2 &= \frac{(p-1)a_j\delta}{e^{2a_j\delta}-1}\frac{S_j^2}{\delta} \le C_{11}, \\ \frac{e^{a_j\delta}+1}{2} &+ \left(a_j x_{j-1}+b_j\right)^2 \le C_{12}. \end{aligned} \tag{7.7}$$

Therefore

$$\begin{aligned} \frac{p(\delta,x_{j-1},x_j)}{h_j(\delta,x_{j-1},x_j)} &\le 1 + C_{13}\frac{e^{a_{j-1}\delta}-1}{a_{j-1}\delta}\left(\sqrt[p]{\varepsilon_p}\sqrt{\frac{e^{2a_{j-1}\delta}-1}{2a_{j-1}\delta}}\sqrt{\delta} + S_j\right)\delta \\ &\le 1 + C_{14}\sqrt{\delta}\delta = 1 + C_{14}\frac{1}{n^{3/2}}, \end{aligned} \tag{7.8}$$

where we have used (7.6). It follows that

$$\begin{aligned} \lim_{\delta\to 0}\prod_{j=1}^{n}\frac{p(\delta,x_{j-1},x_j)}{h_j(\delta,x_{j-1},x_j)} &\le \lim_{n\to\infty}\left(1 + C_{14}\frac{1}{n^{3/2}}\right)^n \\ &= 1. \end{aligned} \tag{7.9}$$

Similarly we have

$$\lim_{n\to\infty}\prod_{j=1}^{n}\frac{p(\delta, x_{j-1}, x_j)}{h_j(\delta, x_{j-1}, x_j)} \geq 1. \tag{7.10}$$

Therefore

$$\lim_{n\to\infty}\frac{L(x_0^n,\dots,x_n^n)}{L_2(x_0^n,\dots,x_n^n)} = 1, \tag{7.11}$$

and the proof of Theorem 3.1 is complete.

8. The Euler-Maruyama Approximation

Recall that the Euler-Maruyama approximation to (1.3) is a Markov chain given by

$$X_j = X_{j-1} + A(\theta, X_{j-1})\delta + \xi_j\sqrt{\delta}, \tag{8.1}$$

where $\{\xi_j\}$ is an i.i.d. random sequence, with standard normal $N(0,1)$. The conditional distribution of X_j given $X_{j-1} = x_{j-1}$ is Gaussian with mean $x_{j-1} + A(\theta, x_{j-1})\delta$ and variance δ so that the likelihood function is given as

$$L_1(x_0,\dots,x_n) = \frac{1}{(2\pi\delta)^{n/2}}\prod_{j=1}^{n}\exp\left\{-\frac{\left|x_j - x_{j-1} - A(\theta, x_{j-1})\delta\right|^2}{2\pi\delta}\right\}. \tag{8.2}$$

Applying the representation formula (5.7) we have the following.

Proposition 8.1. *It holds that*

$$\frac{L(x_0,\dots,x_n)}{L_1(x_0,\dots,x_n)} = \prod_j\left(1 - \sqrt{\delta}e^{\alpha_j^2/2\delta}\int_0^\delta \mathbb{P}\left\{U_j(s)c_j\left(X_s^j,\theta\right)\frac{W_s+\alpha_j}{(\delta-s)^{3/2}}e^{-(W_s+\alpha_j)^2/2(\delta-s)}\right\}ds\right), \tag{8.3}$$

where $(W_t)_{t\geq 0}$ *is the standard Brownian motion,* $X_s^j = x_{j-1} + \lambda_{j-1}s + W_s$,

$$U_j(t) = \exp\left[\int_0^t c_j\left(X_s^j,\theta\right)dW_s - \frac{1}{2}\int_0^t \left|c_j\right|^2\left(X_s^j,\theta\right)ds\right], \tag{8.4}$$

$\lambda_j = A(x_{j-1},\theta)$, $\alpha_j = \lambda_{j-1}\delta - x_j + x_{j-1}$, *and*

$$c_j(z,\theta) = A(z,\theta) - A\left(x_{j-1},\theta\right). \tag{8.5}$$

From which we may deduce the following estimate.

Proposition 8.2. *If $A(x,\theta)$ is bounded and Lipschitz continuous, uniformly in θ, then the maximum likelihood function with discrete sampling is stable, in the sense that*

$$\prod_j \left(1 - C_{01}\delta e^{\alpha_j^2/2\delta}\right) \leq \frac{L(x_0,\ldots,x_n)}{L_1(x_0,\ldots,x_n)} \leq \prod_j \left(1 + C_{02}\delta e^{\alpha_j^2/2\delta}\right), \tag{8.6}$$

for some constant C_{01} and C_{02}, where $\alpha_j = A(x_{j-1},\theta)\delta - (x_j - x_{j-1})$.

However, this estimate does not lead to the same result as for the local linear approximation. It is not known (to our best knowledge) whether

$$\lim_{n\to\infty} \frac{L\left(x_0^n,\ldots,x_n^n\right)}{L_1\left(x_0^n,\ldots,x_n^n\right)} = 1 \tag{8.7}$$

holds or not under similar conditions in Theorem 3.1.

Acknowledgment

The authors thank Dr. Zeng for providing a list of errors contained in the first version of the paper.

References

[1] T. J. Lyons and W. A. Zheng, "On conditional diffusion processes," *Proceedings of the Royal Society of Edinburgh. Section A*, vol. 115, no. 3-4, pp. 243–255, 1990.
[2] L. C. G. Rogers, "Smooth transition densities for one-dimensional diffusions," *The Bulletin of the London Mathematical Society*, vol. 17, no. 2, pp. 157–161, 1985.
[3] Y. Aït-Sahalia, "Transition densities for interest rate and other nonlinear diffusions," *The Journal of Finance*, vol. 54, no. 4, pp. 1361–1395, 1999.
[4] Y. Aït-Sahalia, "Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach," *Econometrica*, vol. 70, no. 1, pp. 223–262, 2002.
[5] Y. Aït-Sahalia and P. A. Mykland, "Estimators of diffusions with randomly spaced discrete observations: a general theory," *The Annals of Statistics*, vol. 32, no. 5, pp. 2186–2222, 2004.
[6] I. Shoji and T. Ozaki, "Estimation for nonlinear stochastic differential equations by a local linearization method," *Stochastic Analysis and Applications*, vol. 16, no. 4, pp. 733–752, 1998.
[7] B. L. S. Prakasa Rao, *Statistical Inference for Diffusion Type Processes*, vol. 8 of *Kendall's Library of Statistics*, Edward Arnold, London, UK, 1999.
[8] Z. Qian and W. Zheng, "A representation formula for transition probability densities of diffusions and applications," *Stochastic Processes and Their Applications*, vol. 111, no. 1, pp. 57–76, 2004.
[9] M. N. Mishra and J. P. N. Bishwal, "Approximate maximum likelihood estimation for diffusion processes from discrete observations," *Stochastics and Stochastics Reports*, vol. 52, no. 1-2, pp. 1–13, 1995.
[10] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, vol. 113 of *Graduate Texts in Mathematics*, Springer, New York, NY, USA, 1988.
[11] J. R. Norris, "Long-time behaviour of heat flow: global estimates and exact asymptotics," *Archive for Rational Mechanics and Analysis*, vol. 140, no. 2, pp. 161–195, 1997.