Statistics & Probability Letters 43 (1999) 207–215

Large deviations inequalities for the maximum likelihood estimator and the Bayes estimators in nonlinear stochastic differential equations

J.P.N. Bishwal *

Stat-Math Unit, Indian Statistical Institute, 203 B.T. Road, Calcutta 700 035, India

Received June 1998; received in revised form September 1998

Abstract

Exponential bounds on the large deviation probability of the maximum likelihood estimator and the Bayes estimators of the parameter appearing nonlinearly in the drift coefficient of homogeneous Itô's stochastic differential equations are obtained under some regularity conditions. © 1999 Elsevier Science B.V. All rights reserved

MSC: primary 62M05; 62F12; secondary 60H10; 60J60

Keywords: Nonlinear Itô's stochastic differential equations; Diffusion processes; Drift parameter; Maximum likelihood estimator; Bayes estimator; Large deviations bounds

1. Introduction

In the estimation of the parameter appearing nonlinearly in the drift coefficient of Itô's stochastic differential equations having a stationary solution, the weak consistency, asymptotic normality and asymptotic efficiency of the maximum likelihood estimator (MLE) and the Bayes estimator were obtained by Kutoyants (1977, 1984) via studying the local asymptotic normality (LAN) property of the model. Lanksa (1979) obtained the strong consistency and asymptotic normality of the more general minimum contrast estimator which includes the MLE. Prakasa Rao and Rubin (1981) obtained the strong consistency and asymptotic normality of the MLE by studying the weak convergence of the least-squares random field where the families of stochastic integrals were studied by Fourier analytic methods. Prakasa Rao (1982) studied the weak consistency, asymptotic normality and asymptotic efficiency of the maximum probability estimator. All the above authors assumed the parameter to be a scalar. For the multidimentional drift parameter, Bose (1983a, b) obtained the strong consistency and asymptotic normality of the MLE and the Bayes estimators, respectively. In this paper we obtain bounds on the large deviation probabilities for the maximum likelihood and the Bayes estimator under some regularity conditions. We follow the method in Ibragimov and Has'minskii (1981). This method was

* E-mail: tsmv9705@isical.ac.in.

PII: S0167-7152(98)00261-2

used by Prakasa Rao (1984) to obtain the large deviation probability bound for the least-squares estimator in the nonlinear regression model with Gaussian errors.

The paper is organised as follows: In Section 2 we prepare notations, assumptions and preliminaries. Section 3 contains the large deviation inequality for the MLE and Section 4 contains the large deviation results for the Bayes estimator.

2. Notations, assumptions and preliminaries

Let $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geqslant 0} P)$ be a stochastic basis satisfying the usual hypotheses on which we define a stationary ergodic diffusion process $\{X_t, t\geqslant 0\}$ satisfying the Itô SDE

$$\begin{aligned} \mathrm{d}X_t &= f(\theta, X_t)\,\mathrm{d}t + \mathrm{d}W_t, \quad t\geqslant 0, \\ X_0 &= \xi, \end{aligned} \tag{2.1}$$

where $\{W_t, t\geqslant 0\}$ is a standard Wiener process, $E[\xi^2] < \infty$, $f(\theta,x)$ is a known real-valued function continuous on $\Theta \times \mathbb{R}$ where Θ is a closed interval of the real line and the parameter θ is unknown, which is to be estimated on the basis of the observation of the process $\{X_t, 0\leqslant t\leqslant T\} \equiv X_0^T$. Let θ_0 be the true value of the parameter which lies inside the parameter space Θ.

Let P_θ^T be the measure generated by the process X_0^T on the space (C_T, B_T) of continuous functions on $[0,T]$ with the associated Borel σ-algebra B_T under the supremum norm when θ is the true value of the parameter. Let E_θ^T be the expectation with respect to the measure P_θ^T. Suppose P_θ^T is absolutely continuous with respect to $P_{\theta_0}^T$, then it is well known that (see Liptser and Shiryayev, 1977)

$$L_T(\theta) = \frac{\mathrm{d}P_\theta^T}{\mathrm{d}P_{\theta_0}^T}(X_0^T) = \exp\left\{ \int_0^T [f(\theta,X_s) - f(\theta_0,X_s)]\,\mathrm{d}W_s - \frac{1}{2}\int_0^T [f(\theta,X_s) - f(\theta_0,X_s)]^2\,\mathrm{d}s \right\} \tag{2.2}$$

is the Radon–Nikodym derivative of P_θ^T w.r.t. $P_{\theta_0}^T$. The MLE θ_T of θ based on X_0^T is defined as

$$\theta_T := \operatorname*{argmax}_{\theta\in\Theta} L_T(\theta). \tag{2.3}$$

If $L_T(\theta)$ is continuous in θ, it can be shown that there exists a measurable MLE by using Lemma 3.3 in Schmetterer (1974). Hereafter, we assume the existence of such a measurable MLE. We will also assume that the following regularity conditions on $f(\theta,x)$ are satisfied. C denotes a generic constant throughout the paper.

(A1) (i) $f(\theta,x)$ is continuous on $\Theta\times\mathbb{R}$.

(ii) $|f(\theta,x)|\leqslant M(\theta)(1+|x|)\ \forall\theta\in\Theta,\ x\in\mathbb{R},\ \sup\{M(\theta),\ \theta\in\Theta\}<\infty$.

(iii) $|f(\theta,x)-f(\theta,y)|\leqslant M(\theta)|x-y|\ \forall\theta\in\Theta,\ \forall x,y\in\mathbb{R}$.

(iv) $|f(\theta,x)-f(\phi,x)|\leqslant J(x)|\theta-\phi|\ \forall\theta,\ \phi\in\Theta,\ \forall x\in\mathbb{R}$ where $J(\cdot)$ is continuous and $E[J^2(\xi)]<\infty$.

(v) $I(\theta)=E|f(\theta,\xi)-f(\theta_0,\xi)|^2>0\ \forall\theta\neq\theta_0$.

(A2) (i) The first partial derivative of f w.r.t. θ exists and is denoted by $f_\theta^{(1)}(\theta,x)$. The derivative evaluated at θ_0 is denoted by $f_\theta^{(1)}(\theta_0,x)$.

(ii) $\beta=E[f^{(1)}(\theta,x)]^2<\infty$.

(iii) There exists $\alpha>0$ s.t.

$$|f_\theta^{(1)}(\theta,x)-f_\theta^{(1)}(\theta,x)\leqslant J(x)|\theta-\phi|^\alpha \quad \forall x\in\mathbb{R}$$

$\forall\theta,\ \phi\in\Theta$ and J is as in (A1)(iv).

(iv) $|f_\theta^{(1)}(\theta,x)|\leqslant N(\theta)(1+|x|)\ \forall\theta\in\Theta,\ \forall x\in\mathbb{R},\ \sup\{N(\theta),\ \theta\in\Theta\}<\infty$.

(A3) There exists a positive constant C such that

$$E_\theta\left[\exp\left\{-u^2(3T)^{-1}\int_0^T \inf_{\phi\in\Theta}(f_\theta^{(1)}(\phi,X_t))^2\,dt\right\}\right]\leqslant C\exp(-u^2C)\quad \text{for all } u.$$

Under assumptions (A1) and (A2), Prakasa Rao and Rubin (1981) proved the strong consistency and asymptotic normality of θ_T as $T\to\infty$. Assumption (A3) is used to prove our large deviation result. This assumption is satisfied for the linear case $f(\theta,x)=\theta x$, i.e., for the Ornstein–Uhlenbeck process.

3. Large deviation bounds for the MLE

Before we obtain bounds on the probabilities of large deviation for the MLE θ_T we shall give a more general result.

Theorem 3.1. *Under assumptions* (A1)–(A3), *for* $\rho>0$, *we have*

$$\sup_{\theta\in\Theta} P_\theta^T\{\sqrt{T}|\theta_T-\theta|\geqslant\rho\}\leqslant B\exp(-b\rho^2)$$

for some positive constants b and B independent of ρ and T.

By substituting $\rho=\sqrt{T}\varepsilon$ in Theorem 3.1, the following corollary is obtained.

Corollary 3.1. *Under the conditions of Theorem* 3.1, *for arbitrary* $\varepsilon>0$ *and all* $T>0$, *we have*

$$\sup_{\theta\in\Theta} P_\theta^T\{|\theta_T-\theta|>\varepsilon\}\leqslant B\exp(-CT),$$

where B and C are positive constants independent of ε and T.

To prove Theorem 3.1 we shall use the following revised version of Theorem 19 of Ibragimov and Has'minskii (1981, p. 372), (see Kallianpur and Selukar, 1993, p. 330).

Lemma 3.2. *Let $\zeta(t)$ be a real-valued random function defined on a closed subset F of the Euclidean space $\mathbb{R}^k$. We shall assume that the random process $\zeta(t)$ is measurable and separable. Assume that the following condition is fulfilled: there exist numbers $m\geqslant r>k$ and a function $H(x):\mathbb{R}^k\to\mathbb{R}^1$ bounded on compact sets such that for all $x,h\in F$, $x+h\in F$,*

$$E|\zeta(x)|^m\leqslant H(x),$$

$$E|\zeta(x+h)-\zeta(x)|^m\leqslant H(x)|h|^r.$$

Then with probability one the realizations of $\zeta(t)$ are continuous functions on F. Moreover, set

$$w(\delta,\zeta,L)=\sup_{x,y\in F,\,|x|,\,|y|\leqslant L,\,|x-y|\leqslant\delta}|\zeta(x)-\zeta(y)|,$$

then

$$E(w(h;\zeta,L))\leqslant B_0\left(\sup_{|x|<L}H(x)\right)^{1/m}L^k h^{(r-k)/m}\log(h^{-1}),$$

where $B_0=64^k(1-2^{-(r-k)m})^{-1}+(2^{(m-r)/m}-1)^{-1}$.

Let us consider the likelihood ratio process

$$Z_T(u) = \frac{\mathrm{d}P^T_{\theta+uT^{-1/2}}}{\mathrm{d}P^T_\theta}(X_0^T).$$

By (2.2) with $g_t(u) = f(\theta + uT^{-1/2}, X_t) - f(\theta, X_t)$, we have

$$Z_T(u) = \exp\left\{ \int_0^T [f(\theta + uT^{-1/2}, X_t) - f(\theta, X_t)]\,\mathrm{d}W_t - \frac{1}{2}\int_0^T [f(\theta + uT^{-1/2}, X_t) - f(\theta, X_t)]^2\,\mathrm{d}t \right\}$$

$$= \exp\left\{ \int_0^T g_t(u)\,\mathrm{d}W_t - \frac{1}{2}\int_0^T g_t^2(u)\,\mathrm{d}t \right\}.$$

Lemma 3.3. *Under the assumptions* (A1)–(A3), *we have*
1. $E_\theta^T[Z_T^{1/2}(u_1) - Z_T^{1/2}(u_2)]^2 \leqslant C(u_2 - u_1)^2$,
2. $E_\theta^T[Z_T^{1/2}(u)] \leqslant C\exp(-Cu^2)$.

Proof. Note that

$$\begin{aligned} E_\theta^T[Z_T^{1/2}(u_1) - Z_T^{1/2}(u_2)]^2 &= E_\theta^T[Z_T(u_1)] + E_\theta^T[Z_T(u_2)] - 2E_\theta^T[Z_T^{1/2}(u_1)Z_T^{1/2}(u_2)] \\ &\leqslant 2 - 2E_\theta^T[Z_T^{1/2}(u_1)Z_T^{1/2}(u_2)]. \end{aligned} \tag{3.1}$$

From Gikhman and Skorohod (1972), for all u we have

$$E_\theta^T[Z_T(u)] = E_\theta^T\left[\exp\left\{\int_0^T g_t(u)\,\mathrm{d}W_t - \frac{1}{2}\int_0^T g_t^2(u)\,\mathrm{d}t\right\}\right] \leqslant 1. \tag{3.2}$$

Let

$$\begin{aligned} &\theta_1 = \theta + u_1 T^{-1/2}, \qquad \theta_2 = \theta + u_2 T^{-1/2}, \\ &\delta_t = f(\theta_2, X_t) - f(\theta_1, X_t) \end{aligned} \tag{3.3}$$

and

$$V_T = \exp\left\{\frac{1}{2}\int_0^T \delta_t\,\mathrm{d}W_t - \frac{1}{4}\int_0^T \delta_t^2\,\mathrm{d}t\right\} = \left(\frac{\mathrm{d}P^T_{\theta_2}}{\mathrm{d}P^T_{\theta_1}}\right)^{1/2}. \tag{3.4}$$

By Itô's formula, V_T can be represented as

$$V_T = 1 - \frac{1}{8}\int_0^T V_t\delta_t^2\,\mathrm{d}t + \int_0^T V_t\delta_t\,\mathrm{d}W_t. \tag{3.5}$$

The random process $\{V_t^2, \mathscr{F}_t, P_\theta^T,\ 0 \leqslant t \leqslant T\}$ is a martingale and from the $\mathscr{F}_t$-measurability of δ_t for each $t \in [0, T]$,

$$\begin{aligned} E^T_{\theta_1}\int_0^T V_t^2\delta_t^2\,\mathrm{d}t &= E^T_{\theta_1}\int_0^T E^T_{\theta_1}(V_T^2|\mathscr{F}_t)\delta_t^2\,\mathrm{d}t \\ &= E^T_{\theta_1}V_T^2\int_0^T \delta_t^2\,\mathrm{d}t \\ &= \int V_T^2\left(\int_0^T \delta_t^2\,\mathrm{d}t\right)\mathrm{d}P^T_{\theta_1} \end{aligned}$$

$$= \int \left(\int_0^T \delta_t^2 \, \mathrm{d}t \right) \mathrm{d}P_{\theta_2}^T$$

$$= E_{\theta_2}^T \left(\int_0^T \delta_t^2 \, \mathrm{d}t \right)$$

$$= E_{\theta_2}^T \int_0^T |f(\theta_2, X_t) - f(\theta_1, X_t)|^2 \, \mathrm{d}t \tag{3.6}$$

$$\leqslant E_{\theta_2}^T \int_0^T [J^2(X_t)] |\theta_2 - \theta_1|^2 \, \mathrm{d}t \quad \text{(by (A1))}$$

$$\leqslant (u_2 - u_1)^2 \frac{1}{T} \int_0^T E_{\theta_2}[J^2(\xi)] \, \mathrm{d}t$$

$$< C(u_2 - u_1)^2 < \infty. \tag{3.7}$$

Hence $E_{\theta_1}^T \int_0^T V_t \delta_t \, \mathrm{d}W_t = 0$. Therefore, using $|ab| \leqslant (a^2 + b^2)/2$, we obtain from (3.5)

$$E_{\theta_1}^T(V_T) = 1 - \frac{1}{8} \int_0^T E_{\theta_1}^T(\delta_t V_t . \delta_t) \, \mathrm{d}t$$

$$\geqslant 1 - \frac{1}{16} \int_0^T E_{\theta_1}^T \delta_t^2 \, \mathrm{d}t - \frac{1}{16} \int_0^T E_{\theta_1}^T V_t^2 \delta_t^2 \, \mathrm{d}t$$

$$= 1 - \frac{1}{16} E_{\theta_1}^T \int_0^T \delta_t^2 \, \mathrm{d}t - \frac{1}{16} E_{\theta_2}^T \int_0^T \delta_t^2 \, \mathrm{d}t \quad \text{(by (3.6))}. \tag{3.8}$$

Now

$$E_\theta^T [Z_T^{1/2}(u_1) Z_T^{1/2}(u_2)] = E_\theta^T \left\{ \left[\frac{\mathrm{d}P_{\theta + u_1 T^{-1/2}}^T}{\mathrm{d}P_\theta^T} \right]^{1/2} \left[\frac{\mathrm{d}P_{\theta + u_2 T^{-1/2}}^T}{\mathrm{d}P_\theta^T} \right]^{1/2} \right\}$$

$$= \int \left[\frac{\mathrm{d}P_{\theta_1}^T}{\mathrm{d}P_\theta^T} \right]^{1/2} \left[\frac{\mathrm{d}P_{\theta_2}^T}{\mathrm{d}P_\theta^T} \right]^{1/2} \mathrm{d}P_\theta^T$$

$$= \int \left[\frac{\mathrm{d}P_{\theta_2}^T}{\mathrm{d}P_{\theta_1}^T} \right]^{1/2} \mathrm{d}P_{\theta_1}^T = E_{\theta_1}^T(V_T). \tag{3.9}$$

Substituting (3.9) into (3.1) and using (3.8), we obtain

$$E_\theta^T [Z_T^{1/2}(u_1) - Z_T^{1/2}(u_2)]^2 \leqslant 2 - 2E_{\theta_1}(V_T)$$

$$\leqslant \frac{1}{8} E_{\theta_1}^T \int_0^T \delta_t^2 \, \mathrm{d}t + \frac{1}{8} E_{\theta_1}^T \int_0^T \delta_t^2 \, \mathrm{d}t$$

$$\leqslant C(u_2 - u_1)^2 \quad \text{(using arguments similar to (3.7))}.$$

This completes the proof of (1).

Let us now prove (2). By the Hölder inequality,

$$
\begin{aligned}
&E_\theta^T[Z_T^{1/2}(u)]\\
&\quad = E_\theta^T\left[\exp\left\{\frac{1}{2}\int_0^T g_t(u)\,\mathrm{d}W_t - \frac{1}{4}\int_0^T g_t^2(u)\,\mathrm{d}t\right\}\right]\\
&\quad = E_\theta^T\left[\exp\left\{\frac{1}{2}\int_0^T g_t(u)\,\mathrm{d}W_t - \frac{1}{6}\int_0^T (g_t(u))^2\,\mathrm{d}t\right\}\exp\left\{-\frac{1}{12}\int_0^T (g_t(u))^2\,\mathrm{d}t\right\}\right]\\
&\quad \leqslant \left\{E_\theta^T\left[\exp\left\{\frac{1}{2}\int_0^T g_t(u)\,\mathrm{d}W_t - \frac{1}{6}\int_0^T (g_t(u))^2\,\mathrm{d}t\right\}\right]^{4/3}\right\}^{3/4}\\
&\qquad \times\left\{E_\theta^T\left[\exp\left\{-\frac{1}{12}\int_0^T (g_t(u))^2\,\mathrm{d}t\right\}\right]^4\right\}^{1/4}\\
&\quad \leqslant \left[E_\theta^T\exp\left\{\frac{2}{3}\int_0^T g_t(u)\,\mathrm{d}W_t - \frac{2}{9}\int_0^T (g_t^2(u))\,\mathrm{d}t\right\}\right]^{3/4}\left[E_\theta^T\exp\left\{-\frac{1}{3}\int_0^T (g_t(u))^2\,\mathrm{d}t\right\}\right]^{1/4}. \qquad (3.10)
\end{aligned}
$$

Assumptions (A2) and (A3) yield

$$
\begin{aligned}
&E_\theta^T\exp\left\{-\frac{1}{3}\int_0^T (g_t(u))^2\,\mathrm{d}t\right\}\\
&\quad = E_\theta^T\exp\left\{-\frac{1}{3}\int_0^T [f(\theta+uT^{-1/2},X_t)-f(\theta,X_t)]^2\,\mathrm{d}t\right\}\\
&\quad = E_\theta^T\exp\left\{-\frac{u^2}{3T}\int_0^T (f_\theta^{(1)}(\overline{\theta},X_t))^2\,\mathrm{d}t\right\}\quad (\text{where } |\theta-\overline{\theta}|\leqslant uT^{-1/2})\\
&\quad \leqslant E_\theta^T\exp\left\{-\frac{u^2}{3T}\int_0^T \inf_{\phi\in\Theta}(f_\theta^{(1)}(\phi,X_t))^2\,\mathrm{d}t\right\}\\
&\quad \leqslant C\exp(-u^2C). \qquad (3.11)
\end{aligned}
$$

On the other hand, from Gikhman and Skorohod (1972)

$$
E_\theta^T\left[\exp\left\{\int_0^T \frac{2}{3}g_t(u)\,\mathrm{d}W_t - \frac{1}{2}\int_0^T\left(\frac{2}{3}g_t(u)\right)^2\mathrm{d}t\right\}\right]\leqslant 1. \qquad (3.12)
$$

Combining (3.10)–(3.12) completes the proof of (2). □

Proof of Theorem 3.1. Let $U_T=\{u:\theta+uT^{-1/2}\in\Theta\}$. Then

$$
\begin{aligned}
P_\theta^T\{\sqrt{T}|\theta_T-\theta|>\rho\} &= P_\theta^T\{|\theta_T-\theta|>\rho T^{-1/2}\}\\
&\leqslant P_\theta^T\left\{\sup_{|u|\geqslant\rho,\,u\in U_T} L_T(\theta+uT^{-1/2})\geqslant L_T(\theta)\right\}
\end{aligned}
$$

$$= P_\theta^T \left\{ \sup_{|u| \geqslant \rho} \frac{L_T(\theta + uT^{-1/2})}{L_T(\theta)} \geqslant 1 \right\}$$

$$= P_\theta^T \left\{ \sup_{|u| \geqslant \rho} Z_T(u) \geqslant 1 \right\}$$

$$\leqslant \sum_{r=0}^{\infty} P_\theta^T \left\{ \sup_{u \in \Gamma_r} Z_T(u) \geqslant 1 \right\}, \tag{3.13}$$

where $\Gamma_r = [\rho + r, \rho + r + 1]$.

Applying Lemma 3.2 with $\zeta(u) = Z_T^{1/2}(u)$, from Lemma 3.3 it follows that there exists a constant $B > 0$ such that

$$\sup_{\theta \in \Theta} E_\theta^T \left\{ \sup_{|u_1 - u_2| \leqslant h, \, |u_1|, \, |u_2| \leqslant l} |Z_T^{1/2}(u_1) - Z_T^{1/2}(u_2)| \right\} \leqslant B l^{1/2} h^{1/2} \log h^{-1}. \tag{3.14}$$

Divide Γ_r into subintervals of length atmost $h > 0$. The number n of subintervals is clearly less than or equal to $[1/h] + 1$. Let $\Gamma_r^{(j)}, 1 \leqslant j \leqslant n$ be the subintervals chosen. Choose $u_j \in \Gamma_r^{(j)}$. Then

$$P_\theta^T \left[\sup_{u \in \Gamma_r} Z_T(u) \geqslant 1 \right]$$

$$\leqslant \sum_{j=1}^{n} P_\theta^T \left[Z_T^{1/2}(u_j) \geqslant \frac{1}{2} \right] + P_\theta^T \left\{ \sup_{|u-v| \leqslant h, \, |u|, \, |v| \leqslant \rho + r + 1} |Z_T^{1/2}(u) - Z_T^{1/2}(v)| \geqslant \frac{1}{2} \right\}$$

$$\leqslant 2 \sum_{j=1}^{n} E_\theta^T [Z_T^{1/2}(u_j)] + 2B(\rho + r + 1)^{1/2} h^{1/2} \log(h^{-1}) \quad \text{(by Markov inequality and (3.14))}$$

$$\leqslant 2C \sum_{j=1}^{n} \exp(-Cu_j^2) + 2B(\rho + r + 1)^{1/2} h^{1/2} \log(h^{-1}) \quad \text{(by Lemma 3.3)}$$

$$\leqslant 2C \left(\left[\frac{1}{h} \right] + 1 \right) \exp\{-C(\rho + r)^2\} + 2B(\rho + r + 1)^{1/2} h^{1/2} \log(h^{-1}).$$

Let us now choose $h = \exp\{-C(\rho + r)^2/2\}$. Then

$$\sup_{\theta \in \Theta} P_\theta^T \left\{ \sup_{u > \rho} Z_T(u) \geqslant 1 \right\} \leqslant B \sum_{r=0}^{\infty} (\rho + r + 1)^{1/2} \exp\left\{ \frac{-C(\rho + r)^2}{4} \right\} \leqslant B \exp(-b\rho^2), \tag{3.15}$$

where B and b are positive generic constants independent of ρ and T.

Similarly, it can be shown that

$$\sup_{\theta \in \Theta} P_\theta^T \left[\sup_{u < -\rho} Z_T(u) \geqslant 1 \right] \leqslant B \exp(-b\rho^2). \tag{3.16}$$

Combining (3.15) and (3.16), we obtain

$$\sup_{\theta \in \Theta} P_\theta^T \left[\sup_{|u| > \rho} Z_T(u) \geqslant 1 \right] \leqslant B \exp(-b\rho^2). \tag{3.17}$$

The theorem follows from (3.14) and (3.17). □

4. Large deviation bounds for the Bayes estimators

Let $\wedge$ be a prior probability measure on $(\Theta, \mathscr{B})$ where $\mathscr{B}$ is the σ-algebra of Borel subsets of Θ. Suppose that $\wedge$ has a density $\lambda(\cdot)$ with respect to the Lebesgue measure on $\mathbb{R}$, which is continuous and positive on Θ and possesses in Θ a polynomial majorant.

Let $p(\theta\,|\,X_0^T)$ be the posterior density of θ given X_0^T. By Bayes theorem $p(\theta\,|\,X_0^T)$ is given by

$$p(\theta\,|\,X_0^T) = \frac{L_T(\theta)\lambda(\theta)}{\int_\Theta L_T(\theta)\lambda(\theta)\,\mathrm{d}\theta}.$$

Let $l(\cdot,\cdot) : \Theta \times \Theta \to \mathbb{R}$ be a loss function which satisfies the following properties:

1. $l(u,v) = l(u-v)$.
2. $l(u)$ is defined and nonnegative on $\mathbb{R}$, $l(0)=0$ and $l(u)$ is continuous at $u=0$ but is not identically equal to 0.
3. l is symmetric, i.e., $l(u) = l(-u)$.
4. $\{u\colon\, l(u) < c\}$ are convex sets and are bounded for all $c > 0$ sufficiently small.
5. There exists numbers $\gamma > 0, H_0 \geqslant 0$ s.t. for $H \geqslant H_0$

$$\sup\{l(u)\colon\, |u| \leqslant H^\gamma\} \leqslant \inf\{l(u)\colon\, |u| \geqslant H\}.$$

Clearly, all loss functions of the form $|u-v|^r$, $r > 0$ satisfy conditions 1–5. In particular, the quadratic loss function $|u-v|^2$ satisfies these conditions.

Then a Bayes estimator $\tilde{\theta}_T$ of θ with respect to the loss function $l(\theta,\theta')$ and prior density $\wedge(\theta)$ is one which minimizes the posterior risk and is given by

$$\tilde{\theta}_T := \operatorname*{argmin}_{u\in\Theta} \int_\Theta l(u,\theta)\, p(\theta|X_0^T)\,\mathrm{d}\theta. \tag{4.1}$$

In particular, for the quadratic loss function $l(u,v) = |u-v|^2$, the Bayes estimator $\tilde{\theta}_T$ becomes the posterior mean given by

$$\tilde{\theta}_T = \frac{\int_\Theta u\,p(u|X_0^T)\,\mathrm{d}u}{\int_\Theta p(v|X_0^T)\,\mathrm{d}v}.$$

We now state the large deviation inequality for the Bayes estimator $\tilde{\theta}_T$.

Theorem 4.1. *Suppose* (A1)–(A3) *and* 1–5 *hold. For* $\rho > 0$, *the Bayes estimator* $\tilde{\theta}_T$ *with respect to the prior* $\lambda(\cdot)$ *and a loss function* $l(\cdot,\cdot)$ *with* $l(u) = l(|u|)$ *satisfies*

$$\sup_{\theta\in\Theta} P_\theta^T\{\sqrt{T}|\tilde{\theta}_T - \theta| \geqslant \rho\} \leqslant B\exp(-b\rho^2)$$

for some positive constants B *and* b *independent of* ρ *and* T.

Corollary 4.1. *Under the conditions of Theorem* 4.1, *for arbitrary* $\varepsilon > 0$ *and all* $T > 0$, *we have*

$$\sup_{\theta\in\Theta} P_\theta^T\{|\tilde{\theta}_T - \theta| > \varepsilon\} \leqslant B\exp(-CT).$$

Proof of Theorem 4.1. To prove Theorem 4.1 we shall use Theorem 5.2 of Ibragimov and Has'minskii (1981, p. 45). In view of Lemma 3.3, conditions (1)–(3) of the said theorem are satisfied with $\alpha = 2$ and $g(u) = u^2$. Hence the result follows from Theorem 5.2 of Ibragimov and Has'minskii (1981, p. 45). □

As another application of Theorem 5.2 in Ibragimov and Has'minskii (1981, p. 45), we obtain the following result.

Theorem 4.2. *Under the assumptions* (A1)–(A3), *for any* N, *we have for the Bayes estimator* $\tilde{\theta}_T$ *w.r.t. the prior* $\lambda(\cdot)$ *and loss function* $l(\cdot,\cdot)$ *satisfying the conditions* 1–5,

$$\lim_{H\to\infty,\,T\to\infty} H^N \sup_{\theta\in\Theta} P_\theta^T\{\sqrt{T}|\tilde{\theta}_T-\theta|>H\}=0.$$

Acknowledgements

The author thanks his supervisor Prof. Arup Bose of Indian Statistical Institute for suggestions and helpful discussions. He would also like to thank the anonymous referee for careful reading of the paper and suggestions.

References

Bose, A., 1983a. Asymptotic theory of estimation in non-linear stochastic differential equations for the multiparameter case. Sankhyā Ser. A 45, 56–65.

Bose, A., 1983b. The Bernstein–von Mises theorem for a certain class of diffusion processes. Sankhyā Ser. A 45, 150–160.

Gikhman, I.I., Skorohod, A.V., 1972. Stochastic Differential Equations. Springer, Berlin.

Ibragimov, I.A., Has'minskii, R.Z., 1981. Statistical Estimation: Asymptotic Theory. Springer, Berlin.

Kallianpur, G., Selukar, R.S., 1993. Estimation of Hilbert space valued parameter by the method of sieves. In: Ghosh, J.K., Mitra, S.K., Parthasarathy, K.R., Prakasa Rao, B.L.S. (Eds.), Statistics and Probability, A R.R. Bahadur Festschrift. Wiley Eastern, New Delhi, pp. 325–347.

Kutoyants, Yu. A., 1977. Estimation of the trend parameter of a diffusion process in the smooth case. Theory Probab. Appl. 22, 399–406.

Kutoyants, Yu. A., 1984. Parameter Estimation for Stochastic Processes (Translated and edited by B.L.S. Prakasa Rao) Heldermann-Verlag, Berlin.

Lanksa, V., 1979. Minimum contrast estimation in diffusion processes. J. Appl. Probab. 16, 65–75.

Liptser, R.S., Shiryayev, A.N., 1977. Statistics of Random Processes I, General Theory. Springer, Berlin.

Prakasa Rao, B.L.S., 1982. Maximum probability estimation for diffusion processes. In: Kallianpur, G., Krishnaiah, P.R., Ghosh, J.K. (Eds.), Statistics and Probability, Essays in honour of C.R. Rao. North-Holland, Amsterdam, pp. 557–590.

Prakasa Rao, B.L.S., 1984. On the exponential rate of convergence of the least squares estimator in the nonlinear regression model with Gaussian errors. Statist. Probab. Lett. 2, 139–142.

Prakasa Rao, B.L.S., Rubin, H., 1981. Asymptotic theory of estimation in nonlinear stochastic differential equations. Sankhyā Ser. A 43, 170–189.

Schmetterer, L., 1974. Introduction to Mathematical Statistics, Springer, Berlin.