

THE BERNSTEIN-VON MISES THEOREM FOR STATIONARY PROCESSES

Kenichiro Tamaki*

This paper discusses the asymptotic properties of the posterior density under Whittle measure. The Bernstein-von Mises theorem is shown for short- and long-memory stationary processes. Applications to Bayesian inference for time series are provided.

Key words and phrases: Bernstein-von Mises theorem, short- and long-memory, stationary processes.

1. Introduction

In the literature of time series analysis since Whittle (1953), many authors (for example, Dunsmuir and Hannan (1976), Dunsmuir (1979), and Hosoya and Taniguchi (1982)) have considered an approach using Whittle's log-likelihood, which is an approximation of Gaussian log-likelihood of the data, and have developed the asymptotic properties of an estimator that maximizes Whittle's log-likelihood.

The Whittle likelihood is useful because it is easy to compute, and the use of the periodogram transforms dependent data into asymptotically independent data. Hence, there has been considerable interest in the further development of the theory in other directions. Monti (1997) applied the empirical likelihood approach to Whittle's likelihood for constructing confidence regions. Choudhuri *et al.* (2004) showed that the actual joint distribution of the periodograms, at certain frequencies for a Gaussian time series, is mutually contiguous with the corresponding Whittle measure. Contiguity plays vital roles in estimation and testing theory.

The Bernstein-von Mises theorem is one of the fundamental results in the asymptotic theory of Bayesian inference, and gives the convergence of the posterior density to normal. For Markov processes this result was obtained by Borwanker *et al.* (1971). Applications of this theorem lead to various results on the asymptotic behavior of Bayes estimates.

This paper discusses a Bayes approach to stationary time series. We give the asymptotic properties of the posterior density under Whittle measure. Then the Bernstein-von Mises theorems for short- and long-memory stationary processes are shown. In Section 2 we present our main results. These results enable us to elucidate the asymptotic behavior of Bayes estimates. Also some examples will be given. Proofs are relegated to Section 3.

Received May 14, 2007. Revised June 30, 2007. Accepted July 17, 2007.

*School of Political Science and Economics, Waseda University, 1-6-1, Nishiwaseda, Shinjuku-ku, Tokyo 169-8050, Japan.

2. Results

We consider a real-valued linear process $\{X(t)\}$ generated as

$$X(t) = \sum_{j=0}^{\infty} a_{\theta}(j)\varepsilon(t - j), \quad \theta \in \Theta,$$

where $\{\varepsilon(t)\}$ is a sequence of i.i.d. random variables satisfying $E[\varepsilon(t)] = 0$, $E[\varepsilon(t)^2] = \sigma^2$ and $E[\varepsilon(t)^8] < \infty$, and Θ is an open subset of a compact set $C \in \mathbf{R}$. Denote the spectral density of $\{X(t)\}$ by

$$f_{\theta}(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^{\infty} a_{\theta}(j)e^{ij\lambda} \right|^2.$$

Let

$$\begin{aligned} \lambda_j &= 2\pi j/(n + 1), & r_j &= \exp(i\lambda_j), \\ v_j &= n^{-1/2}(r_j, r_j^2, \dots, r_j^n)' & \text{for } j &= 1, \dots, n \end{aligned}$$

and let

$$c_j = (v_j + v_{n-j})/\sqrt{2}, \quad s_j = i(v_j - v_{n-j})/\sqrt{2} \quad \text{for } j = 1, \dots, n.$$

Define an $n \times n$ matrix

$$P_n = \begin{cases} (c_1, s_1, \dots, c_{n/2}, s_{n/2})', & \text{if } n \text{ is even,} \\ (c_1, s_1, \dots, c_{[n/2]}, s_{[n/2]}, 2^{-1/2}c_{[n/2]+1})', & \text{if } n \text{ is odd,} \end{cases}$$

where $[n]$ denotes the greatest integer less than or equal to n .

For a function $f > 0$ on $[-\pi, \pi]$ define an $n \times n$ diagonal matrix

$$D_n(f) = \begin{cases} 2\pi \text{diag}\{f(\lambda_1), f(\lambda_1), \dots, f(\lambda_{n/2}), f(\lambda_{n/2})\}, & \text{if } n \text{ is even,} \\ 2\pi \text{diag}\{f(\lambda_1), f(\lambda_1), \dots, f(\lambda_{[n/2]}), f(\lambda_{[n/2]}), f(\lambda_{[n/2]+1})\}, & \text{if } n \text{ is odd.} \end{cases}$$

For the stretch $X = (X(1), \dots, X(n))'$ define the Whittle measure $Q_{n,\theta}$ as the product measure of independent normals that gives rise to the Whittle likelihood. Then $Q_{n,\theta} = N\{0, D_n(f_{\theta})\}$ and the quasi (Gaussian) likelihood function based on $Z = P_n X$ under $Q_{n,\theta}$ is given by

$$(2.1) \quad L(\theta) = (2\pi)^{-n/2} \det\{D_n(f_{\theta})\}^{-1/2} \exp\{-(1/2)Z' D_n(f_{\theta})^{-1} Z\}$$

(see Section 4.5 of Brockwell and Davis (1991) and Choudhuri *et al.* (2004)).

We make the following assumption.

ASSUMPTION 1. The coefficients $a_{\theta}(j)$ satisfy

$$\sum_{j=0}^{\infty} |\partial^k a_{\theta}(j)| < \infty, \quad k = 0, 1, 2,$$

where $\partial^k = \partial^k / \partial \theta^k$.

ASSUMPTION 2.

- (i) $f_\theta > 0$ a.e. λ .
- (ii) If $f_{\theta_1} = f_{\theta_2}$, then $\theta_1 = \theta_2$ a.e. λ .
- (iii) There exists a positive constant d_0 such that

$$I(\theta) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left\{ \frac{\partial}{\partial \theta} f_\theta(\lambda) \right\}^2 f_\theta(\lambda)^{-2} d\lambda > d_0, \quad \text{for } \theta \in \Theta.$$

- (iv) Let θ_0 denote the true parameter and $K(t)$ be a nonnegative measurable function satisfying the following conditions: There exists $0 < \varepsilon < I(\theta_0)$ such that

$$\int_{-\infty}^{\infty} K(t) \exp[-\{I(\theta_0) - \varepsilon\}t^2/2] dt < \infty.$$

- (v) θ has the prior density function $\rho(\theta)$ which is continuous and positive in an open neighborhood of θ_0 .
- (vi) For every $h > 0$ and every $\delta > 0$

$$e^{-\delta n} \int_{|t|>h} K(n^{1/2}t) \rho(\hat{\theta} + t) dt \rightarrow 0 \quad \text{a.s. } n \rightarrow \infty,$$

where $\hat{\theta}$ is a maximum quasi-likelihood estimator which maximizes $L(\theta)$ in (2.1).

Assumption 1 implies that the spectral density $f_\theta(\lambda)$ of $\{X(t)\}$ is differentiable with respect to θ and satisfies

$$|\partial^k f_\theta(\lambda)| < \infty, \quad k = 0, 1, 2.$$

For the usual ARMA processes it can be shown that

$$|\partial^k a_\theta(j)| = O(|j|^k |r|^j), \quad k = 0, 1, 2,$$

for some $|r| < 1$. Hence we can see that Assumption 1 is satisfied by a wide class of time series models. The above ARMA process is referred to as a “short-memory process” because the autocovariance function decreases to zero geometrically.

Assumptions 1 and 2 (i) imply that $f_\theta(\lambda)^{-1}$ exists and has the Fourier series representation

$$(2.2) \quad f_\theta(\lambda)^{-1} = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \Delta_\theta(j) e^{ij\lambda}, \quad \sum_{j=-\infty}^{\infty} |\Delta(j)| < \infty.$$

We introduce D space of functions on $[-\pi, \pi]$ defined by

$$D = \left\{ g : g(\lambda) = \sum_{j=-\infty}^{\infty} b(j) \exp(-ij\lambda), b(j) = b(-j), \sum_{j=-\infty}^{\infty} |b(j)| < \infty \right\},$$

that is, D is the space of functions which have the Fourier series representations with absolutely summable Fourier coefficients. From (2.2) and $f_\theta(\lambda) = f_\theta(-\lambda)$, it is easily seen that $f_\theta(\lambda)^{-1} \in D$.

In what follows, we state the fundamental results on the periodogram.

LEMMA 1. Let $I_n(\lambda) = (2\pi n)^{-1} |\sum_{t=1}^n X(t) \exp(it\lambda)|^2$. For $\beta > 3/4$ and $g \in D$,

$$\left| n^{-\beta} \sum_{j=1}^n \{I_n(\lambda_j) - E_{\theta_0}[I_n(\lambda_j)]\} g(\lambda_j) \right| \rightarrow 0 \quad a.s.$$

Liggett (1971) shows that Lemma 1 holds for $\beta > 1/2$ in the Gaussian case because all of the moments exist. We define

$$\begin{aligned} l(\theta) &= \frac{1}{n} \log L(\theta) \\ &= -\log(2\pi) - \frac{1}{2n} \sum_{j=1}^n \log f_\theta(\lambda_j) - \frac{1}{2n} \sum_{j=1}^n f_\theta(\lambda_j)^{-1} I_n(\lambda_j). \end{aligned}$$

From Lemma 1 and $E_{\theta_0}[I_n(\lambda)] = f_{\theta_0}(\lambda) + o(1)$, we get

$$(2.3) \quad l(\theta) - l(\theta_0) \rightarrow \frac{1}{4\pi} \int_{-\pi}^{\pi} \left(1 - \frac{f_{\theta_0}}{f_\theta} + \log \frac{f_{\theta_0}}{f_\theta} \right) d\lambda < 0,$$

unless $f_\theta = f_{\theta_0}$, which implies $\hat{\theta} \rightarrow \theta_0$ a.s.

Now we discuss a Bayes approach to stationary time series. First we consider the posterior density of θ using the Whittle measure. The posterior density of θ given X is

$$f_n(\theta | X) = \frac{\exp\{nl(\theta)\} \rho(\theta)}{\int_{\Theta} \exp\{nl(\theta)\} \rho(\theta) d\theta}$$

Thus the posterior density of $t = n^{1/2}(\theta - \hat{\theta})$ is given by

$$f_n^*(t | X) = C_n^{-1} \nu_n(t) \rho(\hat{\theta} + n^{-1/2}t),$$

where

$$\nu_n(t) = \exp\{nl(\hat{\theta} + n^{-1/2}t) - nl(\hat{\theta})\}$$

and

$$C_n = \int_{-\infty}^{\infty} \nu_n(t) \rho(\hat{\theta} + n^{-1/2}t) dt.$$

In the following, we give the Bernstein-von Mises theorem for short-memory stationary processes.

THEOREM 1. Under Assumptions 1 and 2,

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} K(t) |f_n^*(t | X) - \phi(t; I(\theta_0)^{-1})| dt = 0 \quad a.s.,$$

where $\phi(t; V)$ is the normal density function with mean 0 and variance V .

This result enables us to elucidate the asymptotic behavior of Bayes estimators.

Example 1. We consider the Bayes estimator $\hat{\eta}$ which minimizes

$$B_n(\eta) = \int_{\Theta} l(\theta, \eta) f_n(\theta | X) d\theta,$$

where $l(\theta, \eta) = (\eta - \theta)^2$ is the loss function. Then $\hat{\eta}$ is given by

$$\hat{\eta} = \int_{\Theta} \theta f_n(\theta | X) d\theta.$$

Note that

$$\begin{aligned} \hat{\eta} &= \int_{-\infty}^{\infty} \left(\hat{\theta} + \frac{t}{\sqrt{n}} \right) f_n^*(t | X) dt \\ &= \hat{\theta} + \frac{1}{\sqrt{n}} \int_{-\infty}^{\infty} t f_n^*(t | X) dt. \end{aligned}$$

We obtain

$$\sqrt{n}(\hat{\eta} - \theta_0) = \sqrt{n}(\hat{\theta} - \theta_0) + \int_{-\infty}^{\infty} t f_n^*(t | X) dt.$$

From Theorem 1 we have

$$\int_{-\infty}^{\infty} t f_n^*(t | X) dt \rightarrow \int_{-\infty}^{\infty} t \phi(t; I(\theta_0)^{-1}) dt = 0, \quad \text{a.s.}$$

Hence it is seen that $\hat{\eta} \rightarrow \theta_0$ a.s. and $\sqrt{n}(\hat{\eta} - \theta_0) = \sqrt{n}(\hat{\theta} - \theta_0) + o_p(1)$, which means that the Bayes estimator $\hat{\eta}$ and the maximum quasi-likelihood estimator $\hat{\theta}$ have the same asymptotic distribution.

It is well known that the asymptotic distribution of $\sqrt{n}(\hat{\theta} - \theta_0)$ is normal with mean 0 and variance $V(\theta_0)$, where

$$\begin{aligned} V(\theta) &= I(\theta)^{-1} + V_0(\theta), \\ V_0(\theta) &= \frac{1}{8\pi} I(\theta)^{-2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \{ \partial^1 f_{\theta}(\lambda_1) \} \{ \partial^1 f_{\theta}(\lambda_2) \} f_{\theta}^{(4)}(-\lambda_1, \lambda_2, -\lambda_2) d\lambda_1 d\lambda_2 \end{aligned}$$

and $f_{\theta}^{(4)}(\lambda_1, \lambda_2, \lambda_3)$ is the forth-order cumulant spectral density of $\{X(t)\}$ (see Hosoya and Taniguchi (1982)). Hence if $V_0(\theta) = 0$, then the posterior density of t is asymptotically equal to the density function of the asymptotic distribution of $\sqrt{n}(\hat{\theta} - \theta_0)$.

Example 2. To compare the Bayes estimator $\hat{\eta}$ in Example 1 with the maximum quasi-likelihood estimator $\hat{\theta}$, we consider the following AR(1) model:

$$X(t) - \theta X(t - 1) = \varepsilon(t), \quad |\theta| < 1,$$

Table 1. The average values of $\hat{\theta}$ and $\hat{\eta}$.

θ_0	$\hat{\theta}$	$\hat{\eta}$	$\rho(\theta)$
-0.75	-0.728	-0.730	$[-0.99, -0.5]$
-0.25	-0.246	-0.248	$[-0.5, 0]$
0.25	0.237	0.242	$[0, 0.5]$
0.75	0.724	0.730	$[0.5, 0.99]$

Table 2. MSE of $\hat{\theta}$ and $\hat{\eta}$.

θ_0	$\hat{\theta}$	$\hat{\eta}$	$\rho(\theta)$
-0.75	0.00442	0.00375	$[-0.99, -0.5]$
-0.25	0.00967	0.00622	$[-0.5, 0]$
0.25	0.00962	0.00600	$[0, 0.5]$
0.75	0.00554	0.00441	$[0.5, 0.99]$

where $\varepsilon(t) \sim$ i.i.d. $t(10)$ and $t(p)$ is t -distribution with p degrees of freedom. Suppose that we have the prior information for θ , e.g., for the true value $\theta_0 = 0.25$, it is assumed that we know $\theta_0 \in [0, 0.5]$ in advance. Then let $\rho(\theta)$ be the density function of the uniform distribution on the interval $[0, 0.5]$. Similarly for the true value $\theta_0 = 0.75$, it is assumed that we know $\theta_0 \in [0.5, 0.99]$ in advance. The average values of $\hat{\theta}$ and $\hat{\eta}$ for $n = 100$, $\theta_0 = -0.75, -0.25, 0.25, 0.75$, and 1000 times simulations are given in Table 1, where the row of $\rho(\theta)$ expresses the interval on which $\rho(\theta)$ is the density function of the uniform distribution. Table 2 gives the mean square errors (MSE) of $\hat{\theta}$ and $\hat{\eta}$ for the same case as in Table 1. From Table 1, it is seen that the average value of $\hat{\eta}$ is closer to θ_0 than that of $\hat{\theta}$. Moreover Table 2 shows that the MSE of $\hat{\eta}$ is smaller than that of $\hat{\theta}$.

Recently much attention has been paid to “long-memory process” which appear in many fields (e.g., hydrology and economics). For these processes the autocovariance functions decrease to zero with order of power of lag. In what follows we consider a linear process with long-range dependence. First we impose the following assumption instead of Assumption 1.

ASSUMPTION 3.

- (i) For some $d = d(\theta)$ ($0 < d < 1/4$), the coefficients $a_\theta(j)$ satisfy

$$|\partial^k a_\theta(j)| = O\{|j|^{-1+d}(\log |j|)^k\}, \quad k = 0, 1, 2.$$

- (ii) Let $A_\theta(z) = \sum_{j=0}^{\infty} a_\theta(j)z^j$. Then $|A_\theta(z)| \neq 0$ for $|z| \leq 1$ and $A_\theta(z)$ can be expanded as follows:

$$A_\theta(z)^{-1} = 1 + \sum_{j=1}^{\infty} b_\theta(j)z^j,$$

where the coefficients $b_\theta(j)$ satisfy

$$|\partial^k b_\theta(j)| = O\{|j|^{-1-d}(\log |j|)^k\}, \quad k = 0, 1, 2.$$

Then similarly as in Lemma 1 we give

LEMMA 2. For $\beta > 3/4 + d$ and $g \in D$,

$$\left| n^{-\beta} \sum_{j=1}^n \{I_n(\lambda_j) - E_{\theta_0}[I_n(\lambda_j)]\} g(\lambda_j) \right| \rightarrow 0 \quad a.s.$$

Hence we have the Bernstein-von Mises theorem for stationary processes with long-range dependence.

THEOREM 2. Under Assumptions 2 and 3,

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} K(t) |f_n^*(t | X) - \phi(t; I_0)| dt = 0 \quad a.s.$$

3. Proofs

PROOF OF LEMMA 1. Setting

$$Z_n = n^{-\beta} \sum_{j=1}^n \{I_n(\lambda_j) - E_{\theta_0}[I_n(\lambda_j)]\} g(\lambda_j),$$

we obtain $E_{\theta_0}[Z_n] = 0$. Note that

$$\sum_{j=1}^n I_n(\lambda_j) g(\lambda_j) = \frac{1}{2\pi n} \sum_{j=1}^n \sum_{l=-(n-1)}^{n-1} \sum_{k=1+\underline{l}}^{n-\bar{l}} X_k X_{k+l} \sum_{m=-\infty}^{\infty} b(m) \exp\{-i(l+m)\lambda_j\},$$

where $\underline{l} = \max(0, -l)$ and $\bar{l} = \max(0, l)$ for $l \in \mathbf{Z}$. We get

$$\begin{aligned} \text{Var}_{\theta_0}[Z_n] &= \left(\frac{1}{2\pi n^{1+\beta}} \right)^2 \sum_{m_1, m_2 = -\infty}^{\infty} b(m_1) b(m_2) \\ &\times \sum_{l_1, l_2 = -(n-1)}^{n-1} \sum_{k_1 = 1+\underline{l_1}}^{n-\bar{l_1}} \sum_{k_2 = 1+\underline{l_2}}^{n-\bar{l_2}} \text{cum}[X_{k_1} X_{k_1+l_1}, X_{k_2} X_{k_2+l_2}] \\ &\times \sum_{j_1, j_2 = 1}^n \exp\{-i(l_1 + m_1)\lambda_{j_1}\} \exp\{-i(l_2 + m_2)\lambda_{j_2}\}. \end{aligned}$$

Since

$$\begin{aligned} \text{cum}_{\theta_0}[X_{k_1} X_{k_1+l_1}, X_{k_2} X_{k_2+l_2}] &= \text{cum}_{\theta_0}[X_{k_1}, X_{k_1+l_1}, X_{k_2}, X_{k_2+l_2}] \\ &\quad + \gamma_{\theta_0}(k_2 - k_1) \gamma_{\theta_0}(k_2 - k_1 + l_2 - l_1) \\ &\quad + \gamma_{\theta_0}(k_2 - k_1 + l_2) \gamma_{\theta_0}(k_2 - k_1 - l_1), \end{aligned}$$

where $\gamma_\theta(j) = E_\theta[X(t)X(t+j)]$, we have $\text{Var}_{\theta_0}[Z_n] = O(n^{1-2\beta})$. Similarly, $\text{cum}_{\theta_0}^{(4)}[Z_n] = O(n^{2-4\beta})$. Thus it is seen that $E_{\theta_0}[Z_n^4] = O(n^{2-4\beta})$. Since $2 - 4\beta < -1$, Lemma 1 follows from the Borel-Cantelli lemma. \square

The proof of Theorem 1 is based on the following three lemmas.

LEMMA 3.

(i) For every ε ($0 < \varepsilon < I(\theta_0)$) there exists a δ_0 and an integer N such that

$$\nu_n(t) \leq \exp \left[-\frac{1}{2} \{I(\theta_0) - \varepsilon\} t^2 \right],$$

for $|t| \leq \delta_0 n^{1/2}$ and $n \geq N$.

(ii) For every $\delta > 0$ there exists a positive ε and an integer N such that

$$\sup_{|t| > \delta n^{1/2}} \nu_n(t) \leq \exp \left(-\frac{1}{4} n \varepsilon \right)$$

for $n \geq N$.

(iii) For every fixed t

$$\lim_{n \rightarrow \infty} \nu_n(t) = \exp \left\{ -\frac{1}{2} I(\theta_0) t^2 \right\} \quad a.s.$$

PROOF OF LEMMA 3.

(i) Expanding $l(\hat{\theta} + n^{-1/2}t)$ in a Taylor series at $\theta = \hat{\theta}$, we obtain

$$(3.1) \quad \log \nu_n(t) = \sqrt{n} \partial^1 l(\hat{\theta}) t + \frac{1}{2} \partial^2 l(\theta^*) t^2,$$

where $|\theta^* - \hat{\theta}| \leq t n^{-1/2}$. The first order term on the right hand side of (3.1) equals zero. For the second order term we have

$$(3.2) \quad \frac{1}{2} \partial^2 l(\theta^*) t^2 = \frac{1}{2} \partial^2 l(\theta_0) t^2 + \frac{1}{2} \{ \partial^2 l(\theta^*) - \partial^2 l(\theta_0) \} t^2.$$

Note that

$$\begin{aligned} \partial^2 l(\theta) = -\frac{1}{2n} \sum_{j=1}^n \left[\frac{\{ \partial^2 f_\theta(\lambda_j) \} f_\theta(\lambda_j) - 2 \{ \partial^1 f_\theta(\lambda_j) \}^2}{f_\theta(\lambda_j)^3} \{ f_\theta(\lambda_j) - I_n(\lambda_j) \} \right. \\ \left. + \frac{\{ \partial^1 f_\theta(\lambda_j) \}^2}{f_\theta(\lambda_j)^2} \right]. \end{aligned}$$

From Lemma 1, the first order term on the right hand side of (3.2) converges a.s. to $-I(\theta_0) t^2 / 2$. Therefore it follows that for a positive ε , ($\varepsilon < I(\theta_0)$),

$$\frac{1}{2} \partial^2 l(\theta_0) t^2 < \frac{1}{2} \left\{ -I(\theta_0) + \frac{\varepsilon}{2} \right\} t^2$$

for $n \geq N_1$ (say). Now choose a positive δ such that $|\hat{\theta} - \theta_0| < \delta$ and $|\theta^* - \hat{\theta}| \leq tn^{-1/2} < \delta$ for $n \geq N_2$ (say). Hence if $n \geq N_2$, then $|\theta^* - \theta_0| < 2\delta$ and

$$\{\partial^2 l(\theta^*) - \partial^2 l(\theta_0)\} \leq \sup_{|\theta - \theta_0| < 2\delta} \{\partial^2 l(\theta) - \partial^2 l(\theta_0)\}.$$

From Lemma 1 there exists a $\delta_0 > 0$ such that

$$\sup_{|\theta - \theta_0| < 2\delta_0} \{\partial^2 l(\theta) - \partial^2 l(\theta_0)\} < \frac{\varepsilon}{2}.$$

Thus we get

$$\nu_n(t) \leq \exp \left[-\frac{1}{2} \{I(\theta_0) - \varepsilon\} t^2 \right].$$

(ii) We have

$$n^{-1} \log \nu_n(t) = l(\hat{\theta} + n^{-1/2}t) - l(\theta_0) + l(\theta_0) - l(\hat{\theta}).$$

If $|\hat{\theta} - \theta_0| < \delta/2$ for $n \geq N_3$, then $|tn^{-1/2}| > \delta$ implies that $|\hat{\theta} + tn^{-1/2} - \theta_0| > \delta/2$. Hence for $n \geq N_3$,

$$\begin{aligned} l(\hat{\theta} + n^{-1/2}t) - l(\theta_0) &\leq \sup_{|\theta - \theta_0| > \delta/2} \{l(\theta) - l(\theta_0)\} \\ &< 0 \quad \text{a.s.} \end{aligned}$$

Moreover, $l(\theta_0) - l(\hat{\theta})$ converges to zero a.s. Now choose $\varepsilon > 0$ such that

$$\sup_{|\theta - \theta_0| > \delta/2} \{l(\theta) - l(\theta_0)\} \leq -\varepsilon.$$

Thus, we obtain

$$\sup_{|t| > \delta n^{1/2}} \nu_n(t) \leq \exp \left(-\frac{1}{4} n\varepsilon \right).$$

(iii) For a fixed t and any $\varepsilon > 0$ choose an $\varepsilon_1 > 0$ such that $(t^2/2)\varepsilon_1 < \varepsilon$. From the proof of Lemma 3 (i), we have

$$\left| \log \nu_n(t) + \frac{1}{2} I(\theta_0) t^2 \right| < \frac{1}{2} t^2 \varepsilon_1 < \varepsilon$$

for $n \geq N_4$ (say), which implies the result. \square

LEMMA 4. *There exists a positive δ_0 such that*

$$\lim_{n \rightarrow \infty} \int_{|t| \leq \delta_0 n^{1/2}} K(t) |\nu_n(t) \rho(\hat{\theta} + n^{-1/2}t) - \rho(\theta_0) \exp\{-I(\theta_0)t^2/2\}| dt = 0 \quad \text{a.s.}$$

PROOF OF LEMMA 4.

$$(3.3) \quad \begin{aligned} & \int_{|t| \leq \delta_0 n^{1/2}} K(t) |\nu_n(t) \rho(\hat{\theta} + n^{-1/2}t) - \rho(\theta_0) \exp\{-I(\theta_0)t^2/2\}| dt \\ & \leq \int_{|t| \leq \delta_0 n^{1/2}} K(t) \rho(\theta_0) |\nu_n(t) - \exp\{-I(\theta_0)t^2/2\}| dt \\ & \quad + \int_{|t| \leq \delta_0 n^{1/2}} K(t) \nu_n(t) |\rho(\theta_0) - \rho(\hat{\theta} + n^{-1/2}t)| dt. \end{aligned}$$

Choose an $\varepsilon > 0$ such that $\int \exp[-\{I(\theta_0) - \varepsilon\}t^2/2] dt < \infty$. Then from Lemma 3 (i), there exist a $\delta_1 > \delta_0$ and an N such that

$$(3.4) \quad \nu_n(t) \leq \exp \left[-\frac{1}{2} \{I(\theta_0) - \varepsilon\} t^2 \right], \quad |t| \leq \delta_1 n^{1/2}, \quad n \geq N.$$

Hence, we have, by the dominated convergence theorem and Lemma 3 (iii),

$$(3.5) \quad \begin{aligned} & \int_{|t| \leq \delta_0 n^{1/2}} K(t) \rho(\theta_0) |\nu_n(t) - \exp\{-I(\theta_0)t^2/2\}| dt \\ & \rightarrow 0 \quad \text{as } n \rightarrow \infty, \quad \text{a.s.} \end{aligned}$$

For the second order term on the right hand side of (3.3), we obtain by (3.4)

$$\begin{aligned} & \int_{|t| \leq \delta_0 n^{1/2}} K(t) \nu_n(t) |\rho(\theta_0) - \rho(\hat{\theta} + n^{-1/2}t)| dt \\ & \leq \sup_{|\theta - \theta_0| \leq \delta_2} |\rho(\theta) - \rho(\theta_0)| \int_{|t| \leq \delta_0 n^{1/2}} K(t) \exp\{-I(\theta_0)t^2/2\} dt, \end{aligned}$$

where $\delta_2 > |\hat{\theta} - \theta_0| + \delta_0$. From $|\hat{\theta} - \theta_0| < \delta_1$ for $n \geq N_5$ (say), for a given δ , choose $\delta_0 < \delta_2 - \delta_1$ such that

$$(3.6) \quad \sup_{|\theta - \theta_0| \leq \delta_2} |\rho(\theta) - \rho(\theta_0)| \int_{|t| \leq \delta_0 n^{1/2}} K(t) \exp\{-I(\theta_0)t^2/2\} dt < \delta.$$

Combining (3.5) with (3.6), we have Lemma 4. \square

LEMMA 5. For every $\delta > 0$,

$$\lim_{n \rightarrow \infty} \int_{|t| > \delta n^{1/2}} K(t) |\nu_n(t) \rho(\hat{\theta} + n^{-1/2}t) - \rho(\theta_0) \exp\{-I(\theta_0)t^2/2\}| dt = 0 \quad \text{a.s.}$$

PROOF OF LEMMA 5. It is easily seen that

$$(3.7) \quad \begin{aligned} & \int_{|t| > \delta n^{1/2}} K(t) |\nu_n(t) \rho(\hat{\theta} + n^{-1/2}t) - \rho(\theta_0) \exp\{-I(\theta_0)t^2/2\}| dt \\ & \leq \int_{|t| > \delta n^{1/2}} K(t) \nu_n(t) \rho(\hat{\theta} + n^{-1/2}t) dt \\ & \quad + \int_{|t| > \delta n^{1/2}} K(t) \rho(\theta_0) \exp\{-I(\theta_0)t^2/2\} dt. \end{aligned}$$

For the first order term on the right hand side of (3.7), we obtain by Lemma 3 (ii),

$$\int_{|t| > \delta n^{1/2}} K(t) \nu_n(t) \rho(\hat{\theta} + n^{-1/2}t) dt \leq \exp\left(-\frac{1}{4}n\varepsilon\right) \int_{|t| > \delta n^{1/2}} K(t) \rho(\hat{\theta} + n^{-1/2}t) dt,$$

which, from Assumption 2 (vi), tends to zero a.s. Since

$$\rho(\theta_0) \int_{|t| > \delta n^{1/2}} K(t) \exp\{-I(\theta_0)t^2/2\} dt \rightarrow 0 \quad \text{a.s.},$$

we complete the proof. \square

PROOF OF THEOREM 1. From Lemmas 4 and 5, we obtain

$$(3.8) \quad \lim_{n \rightarrow \infty} \int K(t) |\nu_n(t) \rho(\hat{\theta} + n^{-1/2}t) - \rho(\theta_0) \exp\{-I(\theta_0)t^2/2\}| dt = 0 \quad \text{a.s.}$$

Putting $K(t) \equiv 1$, which satisfies the assumptions on the function K trivially, we get

$$(3.9) \quad C_n = \int_{-\infty}^{\infty} \nu_n(t) \rho(\hat{\theta} + n^{-1/2}t) dt \rightarrow \rho(\theta_0) \int_{-\infty}^{\infty} \exp\{-I(\theta_0)t^2/2\} dt \\ = \rho(\theta_0) (2\pi)^{1/2} I(\theta_0)^{-1/2}.$$

Note that

$$(3.10) \quad \int_{-\infty}^{\infty} K(t) |f_n^*(t | X) - \phi(t; I(\theta_0))| dt \\ \leq \int_{-\infty}^{\infty} K(t) |C_n^{-1} \nu_n(t) \rho(\hat{\theta} + n^{-1/2}t) \\ - C_n^{-1} \rho(\theta_0) \exp\{-I(\theta_0)t^2/2\}| dt \\ + \int_{-\infty}^{\infty} K(t) |\{C_n^{-1} \rho(\theta_0) - (2\pi)^{-1/2} I(\theta_0)^{1/2}\} \\ \times \exp\{-I(\theta_0)t^2/2\}| dt.$$

The first order term on the right hand side of (3.10) tends to zero from (3.8). The second order term tends to zero from (3.9). Hence we complete the proof. \square

PROOF OF LEMMA 2. A similar way to the proof of Lemma 1 yields the results. \square

PROOF OF THEOREM 2. First, we need to show that $\hat{\theta} \rightarrow \theta_0$ a.s. for long memory processes. To this purpose, we show that for $g \in D$,

$$(3.11) \quad \frac{1}{n} \sum_{j=1}^n E_{\theta_0} [I_n(\lambda_j)] g(\lambda_j) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f_{\theta_0}(\lambda) g(\lambda) d\lambda + o(1).$$

It is easily seen that

$$\begin{aligned} \frac{1}{n} \sum_{j=1}^n E_{\theta_0}[I_n(\lambda_j)]g(\lambda_j) &= \frac{1}{2\pi n} \sum_{m=-\infty}^{\infty} \sum_{l=-(n-1)}^{n-1} b(m) \left(1 - \frac{|l|}{n}\right) \gamma_{\theta_0}(l) \\ &\quad \times \sum_{j=1}^n \exp\{-i(m+l)\lambda_j\} \\ &= \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} \sum_{l=-(n-1)}^{n-1} b(m)\gamma_{\theta_0}(l) + o(1), \end{aligned}$$

where $m+l = k(n+1)$ ($k = 0, \pm 1, \pm 2, \dots$). Then we obtain

$$\begin{aligned} (3.12) \quad \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} \sum_{l=-(n-1)}^{n-1} b(m)\gamma_{\theta_0}(l) \\ &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \sum_{l=-(n-1)}^{n-1} b\{k(n+1) - l\}\gamma_{\theta_0}(l) \\ &= \frac{1}{2\pi} \sum_{k=-1}^1 \sum_{l=-(n-1)}^{n-1} b\{k(n+1) - l\}\gamma_{\theta_0}(l) + o(1) \end{aligned}$$

by $|k(n+1) - l| \geq n+3$ for $|k| \geq 2$. Note that for $k = \pm 1$

$$\begin{aligned} \left| \sum_{l=-(n-1)}^{n-1} b\{k(n+1) - l\}\gamma_{\theta_0}(l) \right| &\leq \sum_{|l| \leq [n/2]} |b\{k(n+1) - l\}\gamma_{\theta_0}(l)| \\ &\quad + \sum_{[n/2] < |l| \leq n-1} |b\{k(n+1) - l\}\gamma_{\theta_0}(l)| \\ &\rightarrow 0 \quad \text{as } n \rightarrow \infty. \end{aligned}$$

From (3.12) it follows that

$$\begin{aligned} \frac{1}{n} \sum_{j=1}^n E_{\theta_0}[I_n(\lambda_j)]g(\lambda_j) &= \frac{1}{2\pi} \sum_{l=-(n-1)}^{n-1} b(-l)\gamma_{\theta_0}(l) + o(1) \\ &= \frac{1}{2\pi} \sum_{l=-\infty}^{\infty} b(-l)\gamma_{\theta_0}(l) + o(1), \end{aligned}$$

which implies (3.11). From $f_{\theta}(\lambda)^{-1} \in D$, (2.3) and (3.11), we have $\hat{\theta} \rightarrow \theta_0$ a.s. Hence, Theorem 2 follows from Lemmas 2–5 in the same fashion as Theorem 1 follows from Lemmas 1 and 3–5. \square

Acknowledgements

The author would like to express his sincere thanks to Professor Masanobu Taniguchi for his encouragement and guidance. He also wishes to thank a referee for helpful comments.

REFERENCES

- Borwanker, J., Kallianpur, G. and Prakasa Rao, B. L. S. (1971). The Bernstein-von Mises theorem for Markov processes, *Ann. Math. Statist.*, **42**, 1241–1253.
- Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*, Springer-Verlag, New York.
- Choudhuri, N., Ghosal, S. and Roy, A. (2004). Contiguity of the Whittle measure for a Gaussian time series, *Biometrika*, **91**, 211–218.
- Dunsmuir, W. (1979). A central limit theorem for parameter estimation in stationary vector time series and its application to models for a signal observed with noise, *Ann. Statist.*, **7**, 490–506.
- Dunsmuir, W. and Hannan, E. J. (1976). Vector linear time series models, *Advances in Appl. Probability*, **8**, 339–364.
- Hosoya, Y. and Taniguchi, M. (1982). A central limit theorem for stationary processes and the parameter estimation of linear processes, *Ann. Statist.*, **10**, 132–153.
- Liggett, W. S., Jr. (1971). On the asymptotic optimality of spectral analysis for testing hypotheses about time series, *Ann. Math. Statist.*, **42**, 1348–1358.
- Monti, A. C. (1997). Empirical likelihood confidence regions in time series models, *Biometrika*, **84**, 395–405.
- Whittle, P. (1953). The analysis of multiple stationary time series, *J. Roy. Statist. Soc. B*, **15**, 125–139.