MAXIMUM LIKELIHOOD ESTIMATION FOR MARKOV PROCESSES

B. L. S. PRAKASA RAO

(Received March 20, 1970; revised Aug. 16, 1971)

1. Introduction and summary

In his monograph on Statistical inference for Markov processes, Billingsley [1] has proved the consistency and asymptotic normality of a maximum likelihood estimator (MLE) under some regularity conditions which involve the second and third derivatives of the transition densities. Prakasa Rao [8] and Daniels [3] considered the problem of maximum likelihood estimation under nonstandard conditions for independent and identically distributed random variables. Recently Huber [5] has given regularity conditions in the independent case, which do not involve the existence of second derivative of the likelihood function, under which MLE's are consistent and asymptotically normal. Another novel feature in Huber's result is that he does not assume that the underlying true distribution is a member of the assumed parametric family.

Our aim in this paper is to extend Huber's result to the case of Markov processes. We do assume that the true distribution is a member of the assumed parametric family. We shall prove that the MLE's are consistent and asymptotically normal under some regularity conditions which do not involve the existence of the second derivative of the transition density. In this connection, we would like to mention that Roussas [9], [10] also considered the problem of maximum likelihood estimation in the Markov case.

Section 2 contains some notations and assumptions. Sections 3 and 4 contain two different sets of conditions for strong consistency of estimators. Asymptotic normality of a MLE is proved in Section 5.

We chose to treat the case when the parameter is one-dimensional. Multi-dimensional case can be treated analogously as was done in Huber [5]. Proofs are similar to those given in Huber [5] but some modifications are required in view of the Markov dependence especially in Lemma 5.2.

2. Some notations and assumptions

Consider a measurable space (x, a) and for each $\theta \in H$, let P_θ be a probability measure on a. Assume that for every $\theta \in H$, $\{X_n, n \geqq 0\}$ is a Markov process taking values in the space (x, a, P_θ), with stationary transition measures $P_\theta(\xi, A) = P_\theta[X_{n+1} \in A \mid X_n = \xi]$. We assume that for each $\theta \in H$, $P_\theta(\xi, A)$ is a measurable function of ξ for fixed $A \in a$, and a probability measure on a for fixed ξ. Such a set of transition measures gives rise to a Markov process by Doob [4]. We shall suppose that there is a measure ν on a, not necessarily finite, with respect to which transition measures $P_\theta(\xi, \cdot)$ and the initial distribution $P_\theta(\cdot)$ have densities $f(\xi, \eta; \theta)$ and $f(\xi; \theta)$ respectively. Assume further that $f(\xi; \theta)$ is measurable in the pair (ξ, θ) and $f(\xi, \eta; \theta)$ is measurable in the triple (ξ, η, θ). If $(X_1, X_2, \cdots, X_{n+1})$ is an observation on the process, then it is easily seen that the log likelihood of $(X_1, X_2, \cdots, X_{n+1})$ is

$$\log f(X_1, \theta) + \sum_{k=1}^{n} \log f(X_k, X_{k+1}; \theta)$$

except possibly on a ν-null set. For reasons mentioned in Billingsley [1], it is convenient to assume the log-likelihood function to be

$$L_n(\theta) = \sum_{k=1}^{n} \log f(X_k, X_{k+1}; \theta) .$$

In all the later discussions, we assume that the following condition holds.

(2.1) For each $\theta \in H$, the stationary distribution exists and is unique and has the property that for each ξ in the state space, $P_\theta(\xi, \cdot)$ is absolutely continuous with respect to the stationary distribution.

Now we have the following theorem due to Billingsley [1].

THEOREM 2.1. *Under condition* (2.1), *if* $\varphi(x_1, x_2)$ *is* $a \times a$*-measurable and if* $E|\varphi(X_1, X_2)|$ *is finite when the initial distribution is stationary, then*

$$\lim_{n \to \infty} n^{-1} \sum_{k=1}^{n} \varphi(X_k, X_{k+1}) = E\varphi(X_1, X_2)$$

with probability one, no matter what the initial distribution is.

The second result which we will have occasion to use was proved independently by Billingsley [2] and Ibragimov [6]. This theorem is useful in proving the asymptotic normality.

THEOREM 2.2. *Let* $\{Z_n, n \geqq 1\}$ *be a strictly stationary ergodic process* *such that* $E[Z_1^2]$ *is finite and* $E[Z_n|Z_1,\cdots,Z_{n-1}]=0$ *with probability one for* $n>1$ *and* $E[Z_1]=0$. *Then the distribution of* $n^{-1/2}\sum_{k=1}^{n} Z_k$ *converges weakly, as* $n\to\infty$, *to the normal distribution with mean* 0 *and variance* $E[Z_1^2]$.

As a consequence of this theorem, we have the following lemma which will be used later.

LEMMA 2.3. *Under condition* (2.1), *for any* $a\times a$-*measurable function* $g(x_1, x_2)$ *with* $E[g(X_1, X_2)]^2$ *finite when* X_1 *has the stationary initial distribution, the distribution of*

$$n^{-1/2}\sum_{k=1}^{n} g(X_k, X_{k+1})$$

converges weakly, as $n\to\infty$, *to the normal distribution with mean* $E[g(X_1, X_2)]$ *and* $\mathrm{Var}\,[g(X_1, X_2)]$ *even if the distribution of* X_1 *is not the stationary one.*

The next lemma is very easy to prove and a proof of the same can be found in Loéve [7], p. 386.

LEMMA 2.4. *Let* $Z_1, Z_2, \cdots, Z_n$ *be* n *random variables such that* (i) $E[Z_1]=0$ (ii) $E[Z_i|Z_1,\cdots,Z_{i-1}]=0$ *with probability one,* $2\leqq i\leqq n$ *and* (iii) $E[Z_i^2]<\infty$ *for* $1\leqq i\leqq n$. *Then*

$$\mathrm{Var}\left[\sum_1^n Z_i\right]=\sum_1^n \mathrm{Var}\, Z_i\ .$$

As an immediate application of this lemma, we obtain the following result.

LEMMA 2.5. *For any* $a\times a$-*measurable function* $g(x_1, x_2)$ *with*

$$E[g(X_1, X_2)]^2$$

finite,

$$\mathrm{Var}\left[\sum_{k=1}^{n}(g(X_k, X_{k+1})-E\{g(X_k, X_{k+1})|X_k\})\right]$$
$$=\sum_1^n \mathrm{Var}\,[g(X_k, X_{k+1})-E\{g(X_k, X_{k+1})|X_k\}]\ .$$

PROOF. Let $Z_k=g(X_k, X_{k+1})-E\{g(X_k, X_{k+1})|X_k\}$. Clearly with probably one,

$$E[Z_k|Z_1,\cdots,Z_{k-1}]$$
$$=E[(g(X_k, X_{k+1})-E\{g(X_k, X_{k+1})|X_k\})|Z_1,\cdots,Z_{k-1}]$$
$$=E[(g(X_k, X_{k+1})-E\{g(X_k, X_{k+1})|X_k\})|X_1,\cdots,X_k]$$

$$=E[g(X_k, X_{k+1})|X_1,\cdots, X_k]-E[g(X_k, X_{k+1})|X_k]$$
$$=E[g(X_k, X_{k+1}|X_k]-E[g(X_k, X_{k+1})|X_k]=0$$

since the conditional distribution of X_{k+1} given $X_1,\cdots, X_k$ depends only on X_k. Since the transition density is stationary, it follows that for $1\leqq k\leqq n$, $\text{Var}[Z_k]=\text{Var}[Z_1]$ which is finite by the hypothesis since $E[g(X_1, X_2)]^2$ is finite. Hence we obtain Lemma 2.5 from Lemma 2.4.

3. Consistency: Case A

Let H be an open interval of the real line R^1. Suppose $\rho(x_1, x_2;\theta)$ is a function defined on $x\times x\times H$. Let $T_n=T_n(x_1,\cdots, x_n)$ be measurable with respect to the σ-field $a_n=a\times a\times\cdots\times a$ (n copies) such that

$$\text{(3.1)}\qquad n^{-1}\sum_{i=1}^{n}\rho(X_i, X_{i+1}; T_n)-\inf_{\theta}\left\{n^{-1}\sum_{i=1}^{n}\rho(X_i, X_{i+1};\theta)\right\}\to 0$$

almost surely. We shall now give sufficient conditions which ensure that T_n converges almost surely to some constant θ_0. Here and elsewhere in the paper the phrase "almost surely" refers to the underlying true probability measure.

ASSUMPTIONS A

(3.2) For each $\theta\in H$, $\rho(x_1, x_2;\theta)$ is $a\times a$-measurable and $\rho(x_1, x_i;\theta)$ is seperable with respect to closed intervals of H when considered as a process in θ.

(3.3) $\rho(x_1, x_2;\theta)$ is lower semi-continuous in θ for almost all pairs $(x_1, x_2)\in x\times x$.

(3.4) There exists a function $a(x_1, x_2)$ which is $a\times a$-measurable such that

$$E[\rho(X_1, X_2;\theta)-a(X_1, X_2)]^-<\infty \qquad \text{for all } \theta\in H$$

and

$$E[\rho(X_1, X_2;\theta)-a(X_1, X_2)]^+<\infty \qquad \text{for all } \theta\in H,$$

where expectations are taken with respect to the true underlying distribution. Let

$$\gamma(\theta)=E[\rho(X_1, X_2;\theta)-a(X_1, X_2)]\,.$$

(3.5) There exists $\theta_0\in H$ such that $\gamma(\theta)>\gamma(\theta_0)$ for all $\theta\neq\theta_0$.

(3.6) There exists a continuous function $b(\theta)$ such that

(i) $b(\theta)>0$ for $\theta\in H$,

(ii) $\inf_{\theta \in H} \{b^{-1}(\theta)[\rho(x_1, x_2; \theta) - a(x_1, x_2)]\} \geqq h(x_1, x_2)$ for some function $h(x_1, x_2)$ such that $E[|h(X_1, X_2)|] < \infty$,

(iii) $\liminf b(\theta) > \gamma(\theta_0)$ as $|\theta| \to \infty$, and

(iv) $E\left[\liminf_{|\theta| \to \infty} \{b^{-1}(\theta)(\rho(X_1, X_2; \theta) - a(X_1, X_2))\}\right] \geqq 1$.

THEOREM 3.1. *Under the assumptions* (3.2)-(3.6), *any sequence* T_n *satisfying* (3.1) *converges to a constant* θ_0 *almost surely.*

The proof of this theorem runs parallel to the proof given by Huber [5] except for the fact that Theorem 2.1 is used instead of the Strong law of large numbers at the appropriate step of the argument.

Remark. By taking $\rho(x_1, x_2; \theta) = -\log f(x_1, x_2; \theta)$, we obtain that a MLE is strongly consistent if it satisfies conditions (3.2)-(3.6) since a MLE $T_n = \hat{\theta}_n(X_1, \cdots, X_n)$ satisfies (3.1).

4. Consistency: Case *B*

As before H is an open interval of the real line R^1. Suppose $\Psi(x_1, x_2, \theta)$ is a real-valued function defined on $x \times x \times H$. Let $T_n = T_n(x_1, \cdots, x_n)$ be measurable with respect to the σ-field $a_n = a \times a \times \cdots \times a$ (n copies) such that

$$n^{-1} \sum_{i=1}^{n} \Psi(X_i, X_{i+1}; T_n) \to 0 \tag{4.1}$$

almost surely. We shall now give sufficient conditions which ensure that T_n converges almost surely to some constant θ_0.

ASSUMPTIONS *B*

(4.2) For each $\theta \in H$, $\Psi(x_1, x_2; \theta)$ is $a \times a$-measurable and $\Psi(x_1, x_2; \theta)$ is seperable with respect to closed intervals of H as a process in θ.

(4.3) $\Psi(x_1, x_2; \theta)$ is continuous in θ for almost all pairs $(x_1, x_2) \in x \times x$.

(4.4) $\lambda(\theta) = E[\Psi(X_1, X_2; \theta)]$ exists and is finite and has a unique zero at $\theta = \theta_0$, where E denotes the expectation with respect to the true underlying distribution.

(4.5) There exists a continuous function $b(\theta)$ such that

(i) $b(\theta) \geqq b_0 > 0$ for all $\theta \in H$,

(ii) $\sup_{\theta} \{b^{-1}(\theta) |\Psi(x_1, x_2; \theta)|\}$ is integrable,

(iii) $\liminf \{b^{-1}(\theta)|\lambda(\theta)|\} \geqq 1$ as $|\theta| \to \infty$, and

(iv) $E\left[\limsup_{|\theta|\to\infty} \{b^{-1}(\theta)|\Psi(X_1, X_2; \theta)-\lambda(\theta)|\}\right]<1$.

In the last assumption, separability of $\Psi(x_1, x_2; \theta)$ and the continuity of $b(\theta)$ with $b(\theta) \geqq b_0 > 0$ for all θ, imply the measurability of the functions involved in (ii) and (iii).

THEOREM 4.1. *Under the assumptions* (4.2)–(4.5), *any sequence* T_n *satisfying* (4.1) *will have the property that* T_n *converges to a constant* θ_0 *almost surely.*

The proof of this theorem is similar to the one given by Huber [5] and hence is omitted.

Remark. In view of Theorem 4.1, it follows that a MLE is consistent if assumptions (4.2)–(4.5) are satisfied by $\Psi(x_1, x_2; \theta) = \partial \log f(x_1, x_2; \theta)/\partial\theta$ since (4.1) is automatically satisfied in such a case.

5. Asymptotic normality

Assume that H is an open interval of the real line as was done earlier. Let $\Psi(x_1, x_2; \theta)$ be a real-valued function defined on $x \times x \times H$. Let $T_n = T_n(x_1, \cdots, x_n)$ be measurable with respect to the σ-field $a_n = a \times a \times \cdots \times a$ (n copies) such that

$$n^{-1/2} \sum_{i=1}^{n} \Psi(X_i, X_{i+1}; T_n) \to 0 \tag{5.1}$$

in probability. We shall now give conditions under which (5.1) implies that T_n is asymptotically normal. It is easy to see that any MLE $\hat{\theta}_n$ such that $\partial L_n(\theta)/\partial\theta = 0$ for $\theta = \hat{\theta}_n$ satisfies (5.1) by taking $\Psi(x_1, x_2; \theta) = \partial \log f(x_1, x_2; \theta)/\partial\theta$. Under the conditions given below, it follows that $\hat{\theta}_n$ is asymptotically normal if it is consistent estimator. We shall assume, through out this section, that T_n is a consistent estimator.

ASSUMPTIONS N

(5.2) For each fixed $\theta \in H$, $\Psi(x_1, x_2; \theta)$ is $a \times a$-measurable and $\Psi(x_1, x_2; \theta)$ is separable with respect to closed intervals of H. Let $\lambda(\theta) = E[\Psi(X_1, X_2; \theta)]$ and $g(x_1, x_2; \theta, d) = \sup \{|\Psi(x_1, x_2; \tau) - \Psi(x_1, x_2; \theta)| : |\tau - \theta| \leqq d\}$ where expectation is taken with respect to the true underlying distribution here and elsewhere in these assumptions.

(5.3) There exists $\theta_0 \in H$ such that $E[\Psi(X_1, X_2; \theta_0|X_1] = 0$ a.e. This implies in particular that $\lambda(\theta_0) = 0$.

(5.4) There exists positive constants a, b, c and d_0 such that

(i) $|\lambda(\theta)|\geqq a|\theta-\theta_0|$ for $|\theta-\theta_0|\leqq d_0$,

(ii) $\sup_{x_1 \in x}\int g(x_1, x_2; \theta, d)\mu(x_1, dx_2)\leqq bd$ for $|\theta-\theta_0|+d\leqq d_0$, $d\geqq 0$ where μ is the true transitional distribution function.

(iii) $\sup_{x_1 \in x}\int [g(x_1, x_2; \theta, d)]^2\mu(x_1, dx_2)\leqq cd$ for $|\theta-\theta_0|+d\leqq d_0$, $d\geqq 0$ where μ is as defined above.

(5.5) $E|\Psi(X_1, X_2; \theta_0)|^2$ is finite.

(5.6) T_n converges to θ_0 in probability.

THEOREM 5.1. *In addition to the above assumptions* (5.1)-(5.6), *suppose that* $\lambda(\theta)$ *has a nonzero derivative* Λ *at* θ_0. *Then* $n^{1/2}(T_n-\theta_0)$ *is asymptotically normal with mean zero and variance* $\Lambda^{-2}E[\Psi(X_1, X_2; \theta_0)]^2$.

Before we give a proof of this theorem, we shall state and prove some lemmas, which lead to its proof. Let

$$Z_n(\tau, \theta)=l_n(\tau)\left|\sum_1^n \{\Psi(X_i, X_{i+1}; \tau)-\Psi(X_i, X_{i+1}; \theta)-\lambda(\tau)+\lambda(\theta)\}\right|$$

where $l_n(\tau)=[n^{1/2}+n|\lambda(\tau)|]^{-1}$.

LEMMA 5.2. *Under the assumption* (5.2), (5.3) *and* (5.4)

$$\sup\{Z_n(\tau, \theta_0) : |\tau-\theta_0|\leqq d_0\}\to 0$$

in probability as n tends to infinity.

PROOF. Let us suppose without loss of generality that $\theta_0=0$ and $d_0=1$. Let $C_k=\{\theta : |\theta|\leqq(1-q)^k\}$, $1\leqq k\leqq k_0$ where k_0 and $q\leqq 1/2$ are to be chosen suitably later. Note that $C_0=\{\theta : |\theta|\leqq 1\}$ and $C_{k-1}-C_k=\{\theta : (1-q)^k<|\theta|<(1-q)^{k-1}\}$. $C_{k-1}-C_k$ consists of two subintervals such that the centre ξ of each of them has absolute value $2^{-1}(1-q)^{k-1}(2-q)$ and there are $N=2k_0$ such subintervals contained in $C_0-C_{k_0}$. Let them be $C_{(1)}, \cdots, C_{(N)}$. Let us observe that for any $\varepsilon>0$,

$$P\left\{\sup_{|\tau|\leqq 1} Z_n(\tau, 0)\geqq 2\varepsilon\right\}\leqq P\left\{\sup_{\tau \in C_{k_0}} Z_n(\tau, 0)\geqq 2\varepsilon\right\}+\sum_{j=1}^N P\left\{\sup_{\tau \in C_{(j)}} Z_n(\tau, 0)\geqq 2\varepsilon\right\} .$$

Choose γ in (1/2, 1) and let $k_0=k_0(n)$ be an integer depending on n such that

$$k_0(n)-1<\frac{\gamma \log n}{|\log(1-q)|}\leqq k_0(n) . \tag{5.7}$$

Since $N=2k_0$, it follows that $N=O(\log n)$. Let $\tau \in C_{(j)}$ and $C_{(j)}$ be a subinterval of $C_{k-1}-C_k$. Then by (5.4),

$$|\lambda(\tau)|\geqq a|\tau|\geqq a(1-q)^k$$

and hence

$$|\lambda(\tau)-\lambda(\xi)|=|E[\Psi(X_1, X_2; \tau)-\Psi(X_1, X_2; \xi)]| \tag{5.8}$$
$$\leqq E\left\{\sup_{|\tau-\xi|\leqq d}|\Psi(X_1, X_2; \tau)-\Psi(X_1, X_2; \xi)|\right\}$$
$$=E[g(X_1, X_2; \xi, d)]\leqq bd\leqq bq(1-q)^k$$

where $d=2^{-1}q(1-q)^{k-1}$ and $|\xi|=2^{-1}(1-q)^{k-1}(2-q)$. Clearly

$$Z_n(\tau, 0)\leqq Z_n(\tau, \xi)+l_n(\tau)\left|\sum_1^n\{\Psi(X_i, X_{i+1}; \xi)-\Psi(X_i, X_{i+1}; 0)-\lambda(\xi)\}\right| \tag{5.9}$$

and

$$Z_n(\tau, \xi)\leqq l_n(\tau)\left\{\sum_1^n[|\Psi(X_i, X_{i+1}; \tau)-\Psi(X_i, X_{i+1}; \xi)|+|\lambda(\tau)-\lambda(\xi)|]\right\} \tag{5.10}$$
$$\leqq l_n(\tau)\left\{\sum_1^n[g(X_i, X_{i+1}; \xi, d)+|\lambda(\tau)-\lambda(\xi)|]\right\}$$
$$\leqq l_n(\tau)\left\{\sum_1^n[g(X_i, X_{i+1}; \xi, d)+Eg(X_1, X_2; \xi, d)]\right\}.$$

Combining (5.9) and (5.10), we obtain that

$$\sup\{Z_n(\tau, 0): \tau\in C_{(j)}\}\leqq U_n+V_n \tag{5.11}$$

where

$$U_n=[na(1-q)^k]^{-1}\sum_1^n[g(X_i, X_{i+1}; \xi, d)+Eg(X_1, X_2; \xi, d)] \tag{5.12}$$

and

$$V_n=[na(1-q)^k]^{-1}\left|\sum_1^n[\Psi(X_i, X_{i+1}; \xi)-\Psi(X_i, X_{i+1}; 0)-\lambda(\xi)]\right|, \tag{5.13}$$

since $l_n(\tau)=(n^{1/2}+n|\lambda(\tau)|)^{-1}\leqq[na(1-q)^k]^{-1}$. Now for any $\varepsilon>0$,

$$P[U_n\geqq\varepsilon]=P\Big[\sum_1^n(g(X_i, X_{i+1}; \xi, d)-E\{g(X_i, X_{i+1}; \xi, d)|X_i\}) \tag{5.14}$$
$$\geqq\varepsilon na(1-q)^k-\sum_1^n E\{g(X_i, X_{i+1}; \xi, d|X_i\}$$
$$-nE\{g(X_1, X_2; \xi, d\}\Big].$$

But by (5.4)-(ii),

$$\varepsilon na(1-q)^k-\sum_1^n E\{g(X_i, X_{i+1}; \xi, d)|X_i\}-nE\{g(X_1, X_2; \xi, d)\} \tag{5.15}$$
$$\geqq\varepsilon na(1-q)^k-2bdn\geqq nbq(1-q)^k$$

provided $q\leqq(3b)^{-1}\varepsilon a$. Choose $q\leqq\min\{1/2, (3b)^{-1}\varepsilon a\}$ and let $m_n=nbq\cdot$

$(1-q)^k$. It now follows form (5.14) that

$$P[U_n \geqq \varepsilon] \leqq P\left[\sum_1^n \{g(X_i, X_{i+1}; \xi, d) - E[g(X_i, X_{i+1}; \xi, d) | X_i]\} \geqq m_n\right]$$
$$\leqq m_n^{-2} E\left[\sum_1^n \{g(X_i, X_{i+1}; \xi, d) - E[g(X_i, X_{i+1}; \xi, d) | X_i]\}\right]^2$$
$$= m_n^{-2} \sum_1^n \text{Var}\,\{g(X_i, X_{i+1}; \xi, d) - E[g(X_i, X_{i+1}; \xi, d) | X_i]\}$$

by Lemma 2.5, which in turn gives the inequality

$$\begin{aligned} P[U_n \geqq \varepsilon] &\leqq m_n^{-2} \sum_1^n E[g(X_i, X_{i+1}; \xi, d) - E\{g(X_i, X_{i+1}; \xi, d) | X_i\}]^2 \\ &\leqq 2m_n^{-2}\left\{\sum_1^n E[g(X_i, X_{i+1}; \xi, d)]^2 \right. \\ &\qquad \left. + \sum_1^n E[(E\{g(X_i, X_{i+1}; \xi, d) | X_i\})^2]\right\} \\ &\leqq 2m_n^{-2}[ncd + nb^2d^2] \\ &\leqq 2n(c+b^2)m_n^{-2}d \qquad \text{since } 0<d<1 \\ &= C_1[n(1-q)^{k-1}]^{-1} \end{aligned} \tag{5.16}$$

where $C_1 = (c+b^2)b^{-2}q^{-1}(1-q)^{-2}$. Let us now consider

$$P[V_n \geqq \varepsilon] = P\left[\left|\sum_1^n [\Psi(X_i, X_{i+1}; \xi) - \Psi(X_i, X_{i+1}; 0) - \lambda(\xi)]\right| \geqq \varepsilon na(1-q)^k\right]$$
$$\leqq P\left[\sum_1^n |\Psi(X_i, X_{i+1}; \xi) - \Psi(X_i, X_{i+1}; 0)| + n|\lambda(\xi)| \geqq \varepsilon na(1-q)^k\right]$$
$$\leqq P\left[\sum_1^n g(X_i, X_{i+1}; \xi, |\xi|) + nEg(X_1, X_2; \xi, |\xi|) \geqq \varepsilon na(1-q)^k\right]$$

which in turn can be shown to satisfy the following inequality by arguments similar to those given in (5.14), (5.15) and (5.16). Since $|\xi|<1$, it can be shown that

$$\begin{aligned} P[V_n \geqq \varepsilon] &\leqq 2n(c+b^2)m_n^{-2}|\xi| \\ &= C_2[n(1-q)^{k-1}]^{-1} \end{aligned} \tag{5.17}$$

where $C_2 = (c+b^2)(2-q)b^{-2}q^{-2}(1-q)^{-2}$. Combining (5.11), (5.16) and (5.17) we obtain that

$$P[\sup\{Z_n(\tau, 0) : \tau \in C_{(j)}\} \geqq 2\varepsilon] \leqq C^* n^{-1}(1-q)^{-k+1} < C^* n^{\gamma-1} \tag{5.18}$$

where

$$\begin{aligned} C^* &= C_1 + C_2 \\ &= (c+b^2)(1-q)^{-2}[b^{-2}q^{-1} + (2-q)b^{-2}q^{-2}] \\ &= (c+b^2)(1-q)^{-2}b^{-2}q^{-2}[q+2-q] \\ &= 2(c+b^2)(1-q)^{-2}b^{-2}q^{-2} \end{aligned} \tag{5.19}$$

which is independent of k and n. Clearly

$$\sup\{Z_n(\tau, 0) : \tau \in C_{k_0}\} \leqq n^{-1/2} \sum_1^n \{g(X_i, X_{i+1}; 0, d) + E[g(X_i, X_{i+1}; 0, d)]\}$$

and hence for $d=(1-q)^{k_0} \leqq n^{-\gamma}$, we have the following inequality.

$$\begin{aligned}(5.20)\qquad & P[\sup\{Z_n(\tau, 0) : \tau \in C_{k_0}\} \geqq 2\varepsilon] \\ & \leqq P\Big[\sum_1^n \{g(X_i, X_{i+1}; 0, d) - E(g(X_i, X_{i+1}; 0, d) \mid X_i)\} \\ & \qquad \geqq 2\varepsilon n^{1/2} - \sum_1^n E(g(X_i, X_{i+1}; 0, d \mid X_i) \\ & \qquad - \sum_1^n E(g(X_i, X_{i+1}; 0, d)\Big] \\ & \leqq P\Big[\sum_1^n g(X_i, X_{i+1}; 0, d) - E(g(X_i, X_{i+1}; 0, d) \mid X_i) \\ & \qquad \geqq 2\varepsilon n^{1/2} - 2bn^{1-\gamma}\Big] .\end{aligned}$$

The last inequality follows from the assumption (5.4). Since $1/2<\gamma<1$, it follows from (5.20), that for n sufficiently large

$$\begin{aligned} & P[\sup\{Z_n(\tau, 0) : \tau \in C_{k_0}\} \geqq 2\varepsilon] \\ & \leqq P\Big[\sum_1^n \{g(X_i, X_{i+1}; 0, d) - E(g(X_i, X_{i+1}; 0, d) \mid X_i)\} \geqq \varepsilon n^{1/2}\Big] \\ & \leqq n^{-1}\varepsilon^{-2} E\Big[\sum_1^n \{g(X_i, X_{i+1}; 0, d) - E(g(X_i, X_{i+1}; 0, d) \mid X_i)\}\Big]^2 \end{aligned}$$

which can be shown to give rise to the following inequality by Lemma 2.5 and by the arguments used in (5.16). It can be shown that

$$(5.20\text{a})\qquad P[\sup\{Z_n(\tau, 0) : \tau \in C_{k_0}\} \geqq 2\varepsilon] \leqq 2(c+b^2)\varepsilon^{-2} n^{-\gamma} .$$

Combining (5.6), (5.18) and (5.20a), we obtain that for n sufficiently large,

$$\begin{aligned}(5.21)\qquad & P[\sup\{Z_n(\tau, 0) : |\tau| \leqq 1\} \geqq 2\varepsilon] \\ & \leqq NC^* n^{\gamma-1} + 2\varepsilon^{-2} n^{-\gamma}(c+b^2) \\ & = 0(n^{-\gamma}) + 0(n^{\gamma-1} \log n)\end{aligned}$$

since $N=0$ $(\log n)$. This proves the result

$$\sup\{Z_n(\tau, 0) : |\tau| \leqq 1\} \to 0$$

in probability as n tends to infinity. This completes the proof of this lemma.

The next lemma is useful in proving our main result on asymptotic normality.

LEMMA 5.3. *Under the assumptions* (5.1) *and* (5.2)–(5.5), *for any* T_n *such that* $P(|T_n-\theta_0|\leqq d_0)\to 1$ *as* $n\to\infty$,

$$n^{-1/2}\sum_1^n \Psi(X_i, X_{i+1}; \theta_0)+n^{1/2}\lambda(T_n)\to 0 \tag{5.22}$$

in probability as n tends to infinity.

PROOF. We shall assume that $\theta_0=0$, $d_0=1$ as was done in the previous lemma. Let $W_n=(n^{1/2}+n|\lambda(T_n)|)^{-1}$. It is easy to see that

$$\begin{aligned}W_n&\left|\sum_1^n [\Psi(X_i, X_{i+1}; 0)+\lambda(T_n)]\right|\\ &\leqq W_n\left|\sum_1^n [\Psi(X_i, X_{i+1}; T_n)-\Psi(X_i, X_{i+1}; 0)-\lambda(T_n)]\right|\\ &\quad+n^{-1/2}\left|\sum_1^n \Psi(X_i, X_{i+1}; T_n)\right|\\ &\leqq \sup\{Z_n(\tau, 0); |\tau|\leqq 1\}+n^{-1/2}\left|\sum_1^n \Psi(X_i, X_{i+1}; T_n)\right|\end{aligned}$$

where the last inequality holds good with probability tending to one as n tends to infinity since $P(|T_n|\leqq 1)\to 1$ as $n\to\infty$. Now Lemma 5.2 together with assumption (5.1) imply that

$$W_n\left|\sum_1^n [\Psi(X_i, X_{i+1}; 0)+\lambda(T_n)]\right|\to 0 \tag{5.23}$$

in probability. Let $E[\Psi(X_1, X_2; 0)]^2=\sigma^2$ where expectation is taken with respect to the true underlying distribution. $E[\Psi(X_1, X_2; 0)]=0$ by (5.3). Hence by Lemma 2.3,

$$n^{-1/2}\sum_{i=1}^n \Psi(X_i, X_{i+1}; 0) \tag{5.24}$$

is asymptotically normal with mean 0 and variance σ^2. We shall now show that (5.23) and (5.24) together imply (5.22). Let $R_n=\sum_1^n \Psi(X_i, X_{i+1}; 0)$. We know that $n^{-1/2}R_n\xrightarrow{a}N(0, \sigma^2)$ and $W_n|R_n+n\lambda(T_n)|\xrightarrow{p}0$ as n tends to infinity. Choose any $\varepsilon>0$. We can find a $k=k(\varepsilon)$ such that for n sufficiently large,

$$P(|n^{-1/2}R_n|>k)<\frac{\varepsilon}{4}, \tag{5.25}$$

and

$$P(W_n|R_n+n\lambda(T_n)|>\varepsilon)<\frac{\varepsilon}{4}. \tag{5.26}$$

(5.25) and (5.26) imply that for n sufficiently large

$$P(|R_n|\leqq kn^{1/2} \text{ and } |R_n+n\lambda(T_n)|\leqq \varepsilon W_n^{-1})>1-\frac{\varepsilon}{2}\ .$$

It is easy to see that the inequalities $|R_n|\leqq kn^{1/2}$ and $|R_n+n\lambda(T_n)|\leqq \varepsilon W_n^{-1}$ imply that $n^{1/2}|\lambda(T_n)|\leqq(k+\varepsilon)(1-\varepsilon)^{-1}$. Hence for n sufficiently large,

$$P[n^{1/2}|\lambda(T_n)|\leqq(k+\varepsilon)(1-\varepsilon)^{-1}]>1-\frac{\varepsilon}{2}\ . \tag{5.27}$$

(5.26) and (5.27) together imply that for n sufficiently large

$$P(n^{-1/2}|R_n+n\lambda(T_n)|<\varepsilon+\varepsilon(k+\varepsilon)(1-\varepsilon)^{-1})>1-\frac{3\varepsilon}{4}$$

i.e.

$$P(n^{-1/2}|R_n+n\lambda(T_n)|<(k+1)\varepsilon(1-\varepsilon)^{-1})>1-\frac{3\varepsilon}{4}$$

where k depends on ε only. This proves that

$$n^{-1/2}\sum_1^n \Psi(X_i, X_{i+1}; 0)+n^{1/2}\lambda(T_n)\to 0$$

in probability which completes the proof of this lemma.

PROOF OF THEOREM 5.1. Since $T_n\to\theta_0$ in probability, $P(|T_n-\theta_0|\leqq d_0)\to 1$ as $n\to\infty$ and hence by Lemma 5.3, it follows that

$$n^{-1/2}\sum_1^n \Psi(X_i, X_{i+1}; \theta_0)+n^{1/2}\lambda(T_n)\to 0$$

in probability and by Lemma 2.3,

$$n^{-1/2}\sum_1^n \Psi(X_i, X_{i+1}; \theta_0)\xrightarrow{a} N(0, \sigma^2)$$

where $N(0, \sigma^2)$ denotes the normal distribution with mean 0 and variance $\sigma^2=E[\Psi(X_1, X_2, \theta_0)]^2$. Hence $n^{1/2}\lambda(T_n)$ is asymptotically normal with mean 0 and variance σ^2. But

$$\begin{aligned}\lambda(T_n)&=\lambda(T_n)-\lambda(\theta_0)\\&=\Lambda(T_n-\theta_0)+o_p(T_n-\theta_0)\ .\end{aligned}$$

Hence we obtain the result $n^{1/2}(T_n-\theta_0)$ is asymptotically normal with mean 0 and variance $\Lambda^{-2}\sigma^2$ since $T_n\to\theta_0$ in probability as $n\to\infty$.

INDIAN INSTITUTE OF TECHNOLOGY, KANPUR (INDIA)

REFERENCES

[1] Billingsley, P. (1961). *Statistical Inference for Markov Processes*, The University of Chicago Press.

[2] Billingsley, P. (1961). The Lindeberg-Levy theorem for martingales, *Proc. Amer. Math. Soc.*, **12**, 788-792.

[3] Daniels, H. (1961). The asymptotic efficiency of a maximum likelihood estimator, *Proc. Fourth Berkerley Symp. Math. Statist. Prob.*, **1**, 151-163.

[4] Doob, J. L. (1953). *Stochastic Processes*, Wiley, New York.

[5] Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions, *Proc. Fifth Berkeley Symp. Math. Statist. Prob.*, **1**, 221-233.

[6] Ibragimov, I. A. (1963). A central limit theorem for a class of dependent random variables, *Theory of Prob. and Appl.*, **8**, 83-89.

[7] Loeve, M. (1963). *Probability Theory*, Von Nostrand, Princeton.

[8] Prakasa Rao, B. L. S. (1968). Estimation of the location of the cusp of a continuous density, *Ann. Math. Statist.*, **39**, 76-87.

[9] Roussas, G. G. (1965). Extension to Markov Processes of a result by A. Wald about consistency of the maximum likelihood estimate, *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **4**, 69-73.

[10] Roussas, G. G. (1965). Asymptotic inference in Markov Processes, *Ann. Math. Statist.*, **36**, 978-992.