Ann. Inst. Statist. Math.
29 (1977), Part A, 203-219

ASYMPTOTIC PROPERTIES OF THE MAXIMUM PROBABILITY ESTIMATES IN MARKOV PROCESSES*

GEORGE G. ROUSSAS

(Received Apr. 15, 1975; revised Mar. 8, 1976)

1. Introduction

Let Θ be an open subset of the real line R and, for each $\theta \in \Theta$, consider the probability space (X, A, P_θ). Here (X, A) may be taken to be equal to $\prod_{j=0}^{\infty}(R_j, B_j)$, where $(R_j, B_j)=(R, B)$ denotes the Borel real line. The probability measure P_θ is the one induced on A by a probability measure $p_\theta(\cdot)$ defined on B and a transition probability measure $p(\cdot\,;\,\cdot)$ defined on $R\times B$. Then, for each $\theta \in \Theta$, the coordinate process $\{X_n\}$, $n\geqq 0$, n an integer, is a Markov process with initial measure $p_\theta(\cdot)$ and transition measure $p_\theta(\cdot\,;\,\cdot)$.

Let A_n be the σ-field induced by the first $n+1$ r.v.'s $X_0, X_1, \cdots, X_n$ of the process in question and let $P_{n,\theta}$ be the restriction of the probability measure P_θ to the σ-field A_n. It will be assumed below that, for any $n\geqq 0$ and any $\theta, \theta' \in \Theta$, the probability measures $P_{n,\theta}$ and $P_{n,\theta'}$, are mutually absolutely continuous. Then fix an arbitrary $\theta_0 \in \Theta$ and set

$$(1.1)\qquad \frac{dP_{0,\theta}}{dP_{0,\theta_0}}=q(X_0\,;\,\theta)\ , \qquad \frac{dP_{1,\theta}}{dP_{1,\theta_0}}=q(X_0, X_1\,;\,\theta)$$

for specified versions of the Radon-Nikodym derivatives involved. Also, set

$$(1.2)\qquad q(X_1\,|\,X_0\,;\,\theta)=\frac{q(X_0, X_1\,;\,\theta)}{q(X_0\,;\,\theta)}\ .$$

With the above notation the likelihood function, $L_n(\cdot\,;\,\theta)$, of the r.v.'s $X_0, X_1, \cdots, X_n$ is given by

$$L_n(Y_n\,;\,\theta)=q(X_0\,;\,\theta)\prod_{j=1}^{n} q(X_j\,|\,X_{j-1}\,;\,\theta)\ ,$$

where $Y_n=(X_0, X_1, \cdots, X_n)$. For each $\theta \in \Theta$, consider the interval $(\theta-$

* This research was supported by the National Science Foundation, Grant MPS 75-10373.

$1/\sqrt{n}, \theta+1/\sqrt{n})$, which will belong in Θ for sufficiently large n, and set

$$Z_n(Y_n;\theta)=\int_{\theta-1/\sqrt{n}}^{\theta+1/\sqrt{n}} L_n(Y_n;t)dl(t) . \tag{1.3}$$

The measure l is Lebesgue measure and assumptions to be made below will insure finiteness of the above integral. Then, any (measurable) function $\hat{d}_n=\hat{d}_n(Y_n)$ for which the integrated likelihood $Z_n(Y_n;\theta)$ is maximized (with respect to θ) is called a *maximum probability estimate* (MPE) of θ.

The concept of an MPE is due to Weiss and Wolfowitz. For more relevant information see [9] and the bibliography cited there.

In the present paper, it will be shown that, under rather weak regularity conditions, there exists at least one MPE, $\hat{d}_n$, for all sufficiently large n and with P_θ-probability $\rightarrow 1$, as $n\rightarrow\infty$. Any such estimates are $\sqrt{n}$-consistent, in the probability sense, and also asymptotically normal. Finally, they are asymptotically efficient in the sense of Weiss-Wolfowitz. By saying that an estimate, $\hat{d}_n$ is "$\sqrt{n}$-consistent" (in the probability sense) it is meant that, for each $\theta\in\Theta$, $\sqrt{n}(\hat{d}_n-\theta)$ is bounded in P_θ-probability. That is, for every $\varepsilon>0$, there is a positive number $M=M(\varepsilon)$ and a positive integer $N=N(\varepsilon)$ such that $P_\theta(|\sqrt{n}(\hat{d}_n-\theta)|<M)>1-\varepsilon$, $n\geqq N$.

The results just described have been established by this author for the i.i.d. case in [7] and [8]. Here they are being generalized to the Markov case.

In order to avoid unnecessary repetitions in the sequel, all limits will be taken as $n\rightarrow\infty$, unless otherwise explicitly stated.

2. Assumptions and notation

In this section, we introduce the necessary additional notation and also gather together the assumptions to be used in the various parts of the paper.

ASSUMPTIONS

(A1) The parameter set Θ is an open subset of the real line R.

For each $\theta\in\Theta$, let $\{X_n\}$, $n\geqq 0$, be a Markov process as described in Section 1. Then:

(A2) The process $\{X_n\}$, $n\geqq 0$, is strictly stationary and metrically transitive (ergodic). (See, for example, [1], pp. 94, 457).

For $n\geqq 0$ and $\theta\in\Theta$, let A_n and $P_{n,\theta}$ be as described in the Introduction. Also, let X and Y be r.v.'s such that X is distributed as X_0 and the 2-dimensional r.v. Y is distributed as the pair $Y_1=(X_0, X_1)$. Then:

(A3) The probability measures $\{P_{n,\theta};\ \theta\in\Theta\}$ are mutually absolutely continuous for all $n\geqq 0$.

Thus, the quantities

$$g(X;\theta)=\log q(X;\theta)\ ,\qquad g(Y;\theta)=\log q(X_1|X_0;\theta)$$

are well-defined with P_θ-probability one for all $\theta\in\Theta$, where $q(X;\theta)$ and $q(X_1|X_0;\theta)$ are given by (1.1) and (1.2), respectively.

(A4) (i) On a set of P_θ-probability one for every $\theta\in\Theta$, there exist the derivatives

$$\varphi(X;\theta)=\frac{\partial}{\partial\theta}g(X;\theta) \tag{2.1}$$

and

$$\Phi(Y;\theta)=\frac{\partial}{\partial\theta}g(Y;\theta)\ ,\qquad \Psi(Y;\theta)=\frac{\partial^2}{\partial\theta^2}g(Y;\theta) \tag{2.2}$$

for all $\theta\in\Theta$ and $\varphi(X;\theta)$ and $\Psi(Y;\theta)$ are continuous in $\theta\in\Theta$.
(ii) For every $\theta\in\Theta$,

$$\mathcal{E}_\theta[\Phi(Y;\theta)|X]=0\ ,\qquad \text{a.s. } [P_\theta]\ ,$$

and

$$\mathcal{E}_\theta[\Phi(Y;\theta)]^2=-\mathcal{E}_\theta\Psi(Y;\theta)=\sigma^2(\theta)>0\ .$$

(A5) For each $\theta\in\Theta$, there is an interval, $I(\theta)$, containing θ and contained in Θ, and a non-negative measurable function

$$H_\theta:\ (R^2, B^2)\to(R, B)$$

such that

$$|\Psi(Y;t)|\leqq H_\theta(Y)\ ,\qquad t\in I(\theta),\ \mathcal{E}_\theta H_\theta(Y)<\infty\ .$$

(A6) For each $\theta\in\Theta$ and any compact subset C of Θ, the functions

$$\inf\,[\Psi(Y;t)-\mathcal{E}_\theta\Psi(Y;t);\ t\in C]\ ,\qquad \sup\,[\Psi(Y;t)-\mathcal{E}_\theta\Psi(Y;t);\ t\in C]$$

are A_1-measurable.

For each $\theta\in\Theta$, set $Z_j=(X_{j-1}, X_j)$, $j=1, 2,\cdots, n$, and define $\Delta_n(\theta)$ by

$$\Delta_n(\theta)=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\Phi(Z_j;\theta)\ . \tag{2.3}$$

Also, for $h \in R$, let

$$\theta_n = \theta + \frac{h}{\sqrt{n}} . \tag{2.4}$$

Thus, $\theta_n \in \Theta$ for sufficiently large n. Finally, set

$$\Lambda(\theta, \theta_n) = \Lambda_n(\theta) = \log \frac{dP_{n,\theta_n}}{dP_{n,\theta}} . \tag{2.5}$$

Then:

(A7) For every $\theta \in \Theta$, the log-likelihood function $\Lambda_n(\theta)$ satisfies the following convergence property

$$\Lambda_n(\theta) - h\Delta_n(\theta) \to -\frac{1}{2} h^2\sigma^2(\theta) \qquad \text{in } P_\theta\text{-probability} . \tag{2.6}$$

Conditions on the probability density function involved which secure the validity of (2.6) may be found in [2], Theorem A.4. Also, a set of conditions of a somewhat different nature from the ones just referred to can be found in [5], p. 45.

It should be noted here that all results in this paper, except for those in Section 7, are derived under Assumptions (A1)–(A6). Thus, Assumption (A7) is used only in establishing the asymptotic efficiency of an MPE.

3. Some preliminary results

In this section, some auxiliary results are presented which play a key role in the sequel. To this end, recall that $Z_j = (X_{j-1}, X_j)$, $j = 1, 2, \cdots, n$, and that Z is a (2-dimensional) r.v. distributed as the pair (X_0, X_1). Next, for each $t \in \Theta$, let

$$W(Z_j; t) = \Psi(Z_j; t) - \mathcal{E}_\theta \Psi(Z_j; t) , \tag{3.1}$$

so that

$$\mathcal{E}_\theta W(Z_j; t) = 0 . \tag{3.2}$$

In this section, the parameter point θ is treated as if it were fixed. Also, let

$$W_n(t) = W_n(Y_n; t) = \frac{1}{n} \sum_{j=1}^{n} W(Z_j; t) . \tag{3.3}$$

Then, by the Ergodic theorem, it follows that, for every $t \in \Theta$,

$$W_n(t) \to 0 \qquad \text{a.s. } [P_\theta] .$$

According to the lemma below, the above convergence is uniform in t, provided t lies in a closed interval. Namely,

LEMMA 3.1. *Let Assumptions* (A1)–(A6) *hold and let* $W_n(t)$ *be defined by* (3.3). *Then, for each* $\theta \in \Theta$ *and any closed interval* $\bar{I}(\theta)$ *centered at* θ *and belonging in* Θ,

$$\sup[|W_n(t)|;\ t \in \bar{I}(\theta)] \to 0 \qquad a.s.\ [P_\theta].$$

PROOF. It is done as that of Lemma 3.1. in [7], where X and X_j there are replaced by X, distributed as X_0, and $Z_j = (X_{j-1}, X_j)$, respectively. Also, the Law of large numbers is replaced by the Ergodic theorem.

In the sequel, the following result, immediate from the previous lemma, will also be needed.

LEMMA 3.2. *Let Assumptions* (A1)–(A6) *hold and let* $\tau_n \in \Theta$ *with* $\tau_n \to 0$. *Then, for sufficiently large* n, *the r.v.'s* $W_n(\theta + \lambda\tau_n)$, $\lambda \in [0, 1]$, *given by* (3.3), *are well-defined and*

$$\int_0^1 |W_n(\theta + \lambda\tau_n)|\, dl(\lambda) \to 0 \qquad a.s.\ [P_\theta].$$

To this lemma, there is the following

COROLLARY 3.1. *Let Assumptions* (A1)–(A6) *hold and let* $\tau_n \in \Theta$ *with* $\tau_n \to 0$. *Then, for sufficiently large* n, *the r.v.'s* $\Psi(Z_j; \theta + \lambda\tau_n)$, $\lambda \in [0, 1]$, *given by* (2.2), *are well-defined and*

$$\int_0^1 \left[\frac{1}{n}\sum_{j=1}^n \Psi(Z_j;\ \theta + \lambda\tau_n)\right] dl(\lambda) \to -\sigma^2(\theta) \qquad a.s.\ [P_\theta].$$

PROOF. See Corollary 3.1 in [7].

4. Existence of roots which are consistent estimates

Before the results of this section are formulated and proved, it would be convenient to recall some of the notation employed so far. To this end, the likelihood function of the r.v.'s $X_0, X_1, \cdots, X_n$ is given by

$$L_n(\theta) = L_n(Y_n; \theta) = q_0(\theta) \prod_{j=1}^n q_j(\theta),$$

where

$$q_0(\theta) = q(X_0; \theta), \qquad q_j(\theta) = q(X_j | X_{j-1}; \theta), \quad j = 1, \cdots, n,$$

and $Y_n = (X_0, X_1, \cdots, X_n)$. The integrated likelihood function is given by

$$Z_n(\theta)=Z_n(Y_n;\theta)=\int_{\theta-1/\sqrt{n}}^{\theta+1/\sqrt{n}} L_n(t)dl(t)\ . \tag{4.1}$$

Now consider the equation

$$\frac{\partial}{\partial\theta}Z_n(\theta)=0\ . \tag{4.2}$$

In this section, it will be shown that there exists at least one root of (4.2) with P_θ-probability$\to 1$, which is a $\sqrt{n}$-consistent estimate of θ in the probability sense. (For the definition of the concept of the "$\sqrt{n}$-consistency" the reader is referred to the third paragraph from the end of Section 1.) More precisely, the following is true.

PROPOSITION 4.1. Let Assumptions (A1)–(A6) hold. Then, for all sufficiently large n and with P_θ-probability$\to 1$, there exists at least one root $\tilde{d}_n=\tilde{d}_n(Y_n)$ of equation (4.2) which is a $\sqrt{n}$-consistent estimate of θ in the probability sense.

The proof of this proposition is facilitated by two lemmas. Set

$$\Phi_j(\theta)=\Phi(Z_j;\theta)=\frac{\partial}{\partial\theta}g(Z_j;\theta)=\frac{\partial}{\partial\theta}\log q(X_j\,|\,X_{j-1};\theta) \tag{4.3}$$

and

$$\Psi_j(\theta)=\frac{\partial}{\partial\theta}\Phi_j(\theta)\ , \tag{4.4}$$

where $Z_j=(X_{j-1},X_j)$, $j=1,\cdots,n$. Then

LEMMA 4.1. *Suppose Assumptions* (A1)–(A6) *hold and let d be a parameter point close to θ (explicitly defined below). Then equation* (4.2) *is equivalent to the following relation*

$$4\Delta_n(\theta)+[\sqrt{n}(d-\theta)]^2(I_n^+-I_n^-)+2\sqrt{n}(d-\theta)(I_n^++I_n^-)+(I_n^+-I_n^-) \tag{4.5}$$
$$+2\left[g_0\left(d+\frac{1}{\sqrt{n}}\right)-g_0\left(d-\frac{1}{\sqrt{n}}\right)\right]=0\ ,$$

where

$$I_n^\pm=I_n\left[\theta+\left(d-\theta\pm\frac{1}{\sqrt{n}}\right)\right]=\int_0^1\left\{\frac{1}{n}\sum_{j=1}^n\Psi_j\left(\theta+\lambda\left(d-\theta\pm\frac{1}{\sqrt{n}}\right)\right)\right\}dl(\lambda) \tag{4.6}$$

and $\Delta_n(\theta)$ is given by (2.3).

PROOF. In all that follows, we shall always work on a set of P_θ-probability one for every $\theta\in\Theta$ without mentioning it explicitly. Differentiating in (4.1) and then taking logarithms, equation (4.2) becomes as follows

$$(4.7)\qquad g_0\left(\theta+\frac{1}{\sqrt{n}}\right)+\sum_{j=1}^{n} g_j\left(\theta+\frac{1}{\sqrt{n}}\right)=g_0\left(\theta-\frac{1}{\sqrt{n}}\right)+\sum_{j=1}^{n} g_j\left(\theta-\frac{1}{\sqrt{n}}\right),$$

where, it is recalled, that

$$(4.8)\qquad g_0(\theta)=\log q_0(\theta).$$

Upon replacing θ by d, where d will be given specific values in the next lemma, relation (4.7) becomes

$$(4.9)\qquad g_0\left(d+\frac{1}{\sqrt{n}}\right)+\sum_{j=1}^{n} g_j\left(d+\frac{1}{\sqrt{n}}\right)=g_0\left(d-\frac{1}{\sqrt{n}}\right)+\sum_{j=1}^{n} g_j\left(d-\frac{1}{\sqrt{n}}\right).$$

Now, for $j=1,\cdots,n$, the functions $g_j(d\pm 1/\sqrt{n})$ are expanded around θ, according to Taylor's formula and up to terms of second order, and give

$$(4.10)\qquad g_j\left(d\pm\frac{1}{\sqrt{n}}\right)=g_j(\theta)+\left(d-\theta\pm\frac{1}{\sqrt{n}}\right)\Phi_j(\theta)$$
$$+\frac{1}{2}\left(d-\theta\pm\frac{1}{\sqrt{n}}\right)^2\int_0^1 \Psi_j\left[\theta+\lambda\left(d-\theta\pm\frac{1}{\sqrt{n}}\right)\right]d\lambda.$$

In (4.10), sum over j, from 1 to n, and replace the expression $\sum_{j=1}^{n} g_j\left(d\pm\frac{1}{\sqrt{n}}\right)$ in (4.9). One then has, after the appropriate cancellations, and by also taking into consideration relation (4.6),

$$g_0\left(d+\frac{1}{\sqrt{n}}\right)+\Delta_n(\theta)+\frac{1}{2}[\sqrt{n}(d-\theta)]^2 I_n^+ +\frac{1}{2}I_n^+ +\sqrt{n}(d-\theta)I_n^+$$
$$=g_0\left(d-\frac{1}{\sqrt{n}}\right)-\Delta_n(\theta)+\frac{1}{2}[\sqrt{n}(d-\theta)]^2 I_n^- +\frac{1}{2}I_n^- -\sqrt{n}(d-\theta)I_n^-.$$

From this last expression, relation (4.5) follows in an obvious manner.

LEMMA 4.2. *Consider the left-hand side of relation* (4.5) *as a function of d and call it $\varphi_n(d)$ $(=\varphi_n(Y_n;d))$. Also, suppose that Assumptions* (A1)-(A6) *hold. Then, for every $\varepsilon>0$ and all sufficiently large $M>0$, there exists a set $A_n=A_n(\varepsilon, M)$ such that $P_\theta(A_n)>1-\varepsilon$ for $n\geqq N=N(\varepsilon, M)>0$ and on A_n,*

$$(4.11)\qquad \varphi_n\left(\theta+\frac{M}{\sqrt{n}}\right)<0,\qquad \varphi_n\left(\theta-\frac{M}{\sqrt{n}}\right)>0.$$

PROOF. Consider $\varphi_n(d)$ as defined above and, for some $M>0$, replace d by $\theta\pm M/\sqrt{n}$ successively. Then, relations (4.5) and (4.6) give

$$(4.12)\quad \varphi_n\left(\theta+\frac{M}{\sqrt{n}}\right)=4\Delta_n(\theta)+M^2\left[I_n\left(\theta+\frac{M+1}{\sqrt{n}}\right)-I_n\left(\theta+\frac{M-1}{\sqrt{n}}\right)\right]$$
$$+2M\left[I_n\left(\theta+\frac{M+1}{\sqrt{n}}\right)+I_n\left(\theta+\frac{M-1}{\sqrt{n}}\right)\right]$$
$$+\left[I_n\left(\theta+\frac{M+1}{\sqrt{n}}\right)-I_n\left(\theta+\frac{M-1}{\sqrt{n}}\right)\right]$$
$$+2\left[g_0\left(\theta+\frac{M+1}{\sqrt{n}}\right)-g_0\left(\theta+\frac{M-1}{\sqrt{n}}\right)\right]$$

and

$$(4.13)\quad \varphi_n\left(\theta-\frac{M}{\sqrt{n}}\right)=4\Delta_n(\theta)+M^2\left[I_n\left(\theta-\frac{M-1}{\sqrt{n}}\right)-I_n\left(\theta-\frac{M+1}{\sqrt{n}}\right)\right]$$
$$-2M\left[I_n\left(\theta-\frac{M-1}{\sqrt{n}}\right)+I_n\left(\theta-\frac{M+1}{\sqrt{n}}\right)\right]$$
$$+\left[I_n\left(\theta-\frac{M-1}{\sqrt{n}}\right)-I_n\left(\theta-\frac{M+1}{\sqrt{n}}\right)\right]$$
$$+2\left[g_0\left(\theta-\frac{M-1}{\sqrt{n}}\right)-g_0\left(\theta-\frac{M+1}{\sqrt{n}}\right)\right].$$

From the definition of the quantities $I_n^{\pm}$, by means of (4.6), and Corollary 3.1, it follows that all I_n's which appear on the right-hand side of relations (4.12) and (4.13) converge to $-\sigma^2(\theta)$ a.s. $[P_\theta]$. On the other hand,

$$(4.14)\qquad g_0\left(\theta\pm\frac{M\pm1}{\sqrt{n}}\right)-g_0\left(\theta\pm\frac{M\mp1}{\sqrt{n}}\right)\to 0$$

in P_θ-probability by Assumption (A4)-(i). Next, from Assumption (A4)-(ii),

$$(4.15)\qquad \mathcal{E}_\theta\Phi_1(\theta)=0\,,\quad \sigma_\theta^2[\Phi_1(\theta)]=\sigma^2(\theta)\qquad (0<\sigma(\theta)<\infty)\,.$$

Setting now C_n for the σ-field induced by the r.v.'s $\Phi_j(\theta)$, $j=1,\cdots,n$, it can be shown that the process $\left\{\sum_{j=1}^{n}\Phi_j(\theta)\right\}$, $n\geqq 1$, is a martingale with respect to the σ-fields $\{C_n\}$, $n\geqq 1$. (For details, see Lemma 1 in [4].) This result, along with (4.15), implies that the Central limit theorem for martingales (see, for example, Theorem 2.2A, p. 205, in [5]) applies and gives

$$(4.16)\qquad L[\Delta_n(\theta)\,|\,P_\theta]\Rightarrow N(0,\sigma^2(\theta))\,.$$

Now, combining the convergence to $-\sigma^2(\theta)$ in P_θ-probability of the I_n-quantities appearing on the right-hand side of (4.12) and (4.13), relations (4.14) and (4.16), as well as standard Slutsky type theorems (see, for example, Theorem 8, p. 152, in [6]), the r.v.'s $\varphi_n(\theta\pm M/\sqrt{n})$ converge as follows

$$L\left[\varphi_n\left(\theta+\frac{M}{\sqrt{n}}\right)\Big| P_\theta\right]\Rightarrow N(-4M\sigma^2(\theta),\, 16\sigma^2(\theta))$$

and

$$L\left[\varphi_n\left(\theta-\frac{M}{\sqrt{n}}\right)\Big| P_\theta\right]\Rightarrow N(4M\sigma^2(\theta),\, 16\sigma^2(\theta))\ .$$

Therefore, for $\varepsilon>0$ and all sufficiently large n,

$$P_\theta\left[\varphi_n\left(\theta+\frac{M}{\sqrt{n}}\right)<0\right]>1-\frac{\varepsilon}{2}$$

and

$$P_\theta\left[\varphi_n\left(\theta-\frac{M}{\sqrt{n}}\right)>0\right]>1-\frac{\varepsilon}{2}\ ,$$

provided, of course, M is sufficiently large. To be more precise, for every $\varepsilon>0$, there exists an $M=M(\varepsilon)>0$ sufficiently large, a set $A_n=A_n(\varepsilon, M)$ and $N=N(\varepsilon, M)>0$ such that for $n\geqq N$, $P_\theta(A_n)>1-\varepsilon$, and on A_n,

$$\varphi_n\left(\theta+\frac{M}{\sqrt{n}}\right)<0\ ,\qquad \varphi_n\left(\theta-\frac{M}{\sqrt{n}}\right)>0\ . \tag{4.17}$$

But this is relation (4.11). The proof of the lemma is completed.

We may now proceed with the proof of the proposition. Namely,

PROOF OF PROPOSITION 4.1. By Assumptions (A4) and (A5), the Dominated convergence theorem (see, for example, [3], pp. 125–126) applies in an obvious manner and provides continuity, with respect to d, of $I_n^\pm$'s as they are defined by relation (4.6). Also, by Assumption (A4)-(i), the r.v.'s $g_0(d\pm 1/\sqrt{n})$ are continuous in d. Then the r.v. $\varphi_n(d)$, defined by the left-hand side of relation (4.5), is continuous in d with P_θ-probability one for all $\theta\in\Theta$. This result, along with (4.17), implies that (for $n\geqq N$) there exists at least one quantity $\tilde{d}_n=\tilde{d}_n(Y_n)$ such that on A_n,

$$\tilde{d}_n\in\left[\theta-\frac{M}{\sqrt{n}},\, \theta+\frac{M}{\sqrt{n}}\right]\ ,\qquad \varphi_n(\tilde{d}_n)=0\ . \tag{4.18}$$

Relation (4.18), in conjunction with Lemma 4.1, implies that, for $n\geqq N$, there exists at least one root of (4.2) which is a $\sqrt{n}$-consistent estimate of θ in the probability sense. This completes the proof of the proposition.

5. Existence of maximum probability estimates

In this section, it is shown that there exists at least one root of the equation

$$\frac{\partial}{\partial\theta} Z_n(\theta)=0 \tag{5.1}$$

which is a $\sqrt{n}$-consistent, in the probability sense, MPE of θ. In other words, the following result is established.

THEOREM 5.1. *Let Assumptions* (A1)–(A6) *hold. Then, for all sufficiently large n and with P_θ-probability $\to 1$, equation* (5.1) *has at least one root $\hat{d}_n=\hat{d}_n(Y_n)$ which is a $\sqrt{n}$-consistent (in the probability sense) MPE of θ.*

PROOF. Let $\tilde{d}_n$ be as described in Proposition 4.1. Then, by Proposition 5.1 below, any such $\tilde{d}_n$ is also an MPE of θ. Then, employing the more conventional notation $\hat{d}_n$ for an estimate of θ, and utilizing the two properties just mentioned, the proof of the theorem is immediate.

Thus, it suffices to establish the following result, namely,

PROPOSITION 5.1. Let Assumptions (A1)–(A6) hold and let $\tilde{d}_n$ be an estimate of θ as described in Proposition 4.1. Then, for all sufficiently large n and with P_θ-probability $\to 1$,

$$\frac{\partial^2}{\partial\theta^2} Z_n(\theta)\,|_{\theta=\tilde{d}_n}<0 . \tag{5.2}$$

For the proof of this proposition, an auxiliary result is required. This will be formulated and proved below.

LEMMA 5.1. *Let Assumptions* (A1)–(A6) *hold and let $\tilde{d}_n$ be an estimate of θ as described in Proposition* 4.1. *Then, for all sufficiently large n and with P_θ-probability $\to 1$, the r.v.'s $W_n(\tilde{d}_n\pm\lambda/\sqrt{n})$, $\lambda\in[0,1]$, given by* (3.3), *are well-defined and*

$$\int_0^1 \left| W_n\!\left(\tilde{d}_n\pm\frac{\lambda}{\sqrt{n}}\right)\right| dl(\lambda)\to 0 \qquad \textit{in } P_\theta\textit{-probability}.$$

PROOF. Since Θ is open and $\tilde{d}_n\to\theta\in\Theta$ in P_θ-probability, it follows that, for every $\varepsilon>0$, there exists a set $A_n=A_n(\varepsilon)$ and $N=N(\varepsilon)>0$ such that, for all $n\geqq N$, $P_\theta(A_n)>1-\varepsilon$, and on A_n,

$$\left(\tilde{d}_n \pm \frac{\lambda}{\sqrt{n}}\right) \in \bar{I}(\theta), \qquad \lambda \in [0, 1].$$

At this point, recall that $\bar{I}(\theta)$ is the closed interval, centered at θ and belonging in Θ, employed in Lemma 3.1. Therefore, for n and A_n as above, the r.v.'s $W_n(\tilde{d}_n \pm \lambda/\sqrt{n})$, $\lambda \in [0, 1]$, are well-defined and

$$\begin{aligned} P_\theta&\left\{\left[\int_0^1 \left|W_n\left(\tilde{d}_n \pm \frac{\lambda}{\sqrt{n}}\right)\right| dl(\lambda) > \varepsilon\right] \cap A_n\right\} \\ &\leq P_\theta[\{\sup[|W_n(t)|;\ t \in \bar{I}(\theta)] > \varepsilon\} \cap A_n] \\ &\leq P_\theta\{\sup[|W_n(t)|;\ t \in \bar{I}(\theta)] > \varepsilon\} \end{aligned}$$

and this last expression$\to 0$ by Lemma 3.1.

The previous lemma has the following immediate corollary. This is used in the proof of Proposition 5.1.

COROLLARY 5.1. *Let Assumptions* (A1)–(A6) *hold and let* $\tilde{d}_n$ *be an estimate of* θ *as described in Proposition* 4.1. *Then, for all sufficiently large* n *and with* P_θ*-probability*$\to 1$, *the r.v.'s*

$$\frac{1}{n}\sum_{j=1}^{n} \Psi_j\left(\tilde{d}_n \pm \frac{\lambda}{\sqrt{n}}\right), \qquad \lambda \in [0, 1], \tag{5.3}$$

given by (2.2), *are well-defined and, with* P_θ*-probability*$\to 1$,

$$\int_0^1 \left[\frac{1}{n}\sum_{j=1}^{n} \Psi_j\left(\tilde{d}_n \pm \frac{\lambda}{\sqrt{n}}\right)\right] dl(\lambda) \to -\sigma^2(\theta).$$

PROOF. That the r.v.'s in (5.3) are well-defined follows as in the proof of the lemma in connection with the r.v.'s $W_n(\tilde{d}_n + \lambda/\sqrt{n})$, $\lambda \in [0, 1]$. Next, by relations (3.1), (3.3) and the lemma,

$$\int_0^1 \left[\frac{1}{n}\sum_{j=1}^{n} \Psi_j\left(\tilde{d}_n \pm \frac{\lambda}{\sqrt{n}}\right) - \mathcal{E}_\theta \Psi_1\left(\tilde{d}_n \pm \frac{\lambda}{\sqrt{n}}\right)\right] dl(\lambda) \to 0 \tag{5.4}$$

with P_θ-probability$\to 1$. Finally, it is seen, as in the proof of Corollary 3.1, that

$$\int_0^1 \left[\mathcal{E}_\theta \Psi_1\left(\tilde{d}_n \pm \frac{\lambda}{\sqrt{n}}\right)\right] dl(\lambda) \to -\sigma^2(\theta) \tag{5.5}$$

with P_θ-probability$\to 1$. Then relations (5.4) and (5.5) provide the desired result.

The proof of the proposition can now be presented.

PROOF OF PROPOSITION 5.1. On the basis of (4.1), it will have to be shown that, for sufficiently large n and P_θ-probability$\to 1$,

$$(5.6)\qquad \frac{\partial}{\partial\theta}L_n\left(\theta+\frac{1}{\sqrt{n}}\right)\bigg|_{\theta=\tilde{d}_n}<\frac{\partial}{\partial\theta}L_n\left(\theta-\frac{1}{\sqrt{n}}\right)\bigg|_{\theta=\tilde{d}_n}.$$

A differentiation (with respect to θ) of the likelihood function

$$L_n(\theta)=q_0(\theta)\prod_{j=1}^{n}q_j(\theta)$$

gives

$$\frac{\partial}{\partial\theta}L_n(\theta)=L_n(\theta)\left[\varphi(\theta)+\sum_{j=1}^{n}\Phi_j(\theta)\right],$$

where the $\Phi_j(\theta)$'s are given by (4.3) and $\varphi(\theta)=\varphi(X;\theta)$ is given by (2.1). Thus,

$$(5.7)\qquad \frac{\partial}{\partial\theta}L_n\left(\theta\pm\frac{1}{\sqrt{n}}\right)=L_n\left(\theta\pm\frac{1}{\sqrt{n}}\right)\left[\varphi\left(\theta\pm\frac{1}{\sqrt{n}}\right)+\sum_{j=1}^{n}\Phi_j\left(\theta\pm\frac{1}{\sqrt{n}}\right)\right].$$

Now, expanding $\Phi_j(\theta\pm1/\sqrt{n})$ around $\tilde{d}_n$ according to Taylor's formula and up to terms of first order, one has

$$\Phi_j\left(\theta\pm\frac{1}{\sqrt{n}}\right)=\Phi_j(\tilde{d}_n)+\left(\theta-\tilde{d}_n\pm\frac{1}{\sqrt{n}}\right)\int_0^1\Psi_j\left[\tilde{d}_n+\lambda\left(\theta-\tilde{d}_n\pm\frac{1}{\sqrt{n}}\right)\right]dl(\lambda).$$

Upon replacing θ by $\tilde{d}_n$ in this last expression and then summing over j, from 1 to n, the following relation results

$$(5.8)\qquad \sum_{j=1}^{n}\Phi_j\left(\tilde{d}_n\pm\frac{1}{\sqrt{n}}\right)=\sum_{j=1}^{n}\Phi_j(\tilde{d}_n)\pm\sqrt{n}\int_0^1\left[\frac{1}{n}\sum_{j=1}^{n}\Psi_j\left(\tilde{d}_n\pm\frac{\lambda}{\sqrt{n}}\right)\right]dl(\lambda).$$

On the basis of (5.7) (with θ replaced by $\tilde{d}_n$) and (5.8) and the fact that

$$L_n\left(\tilde{d}_n-\frac{1}{\sqrt{n}}\right)=L_n\left(\tilde{d}_n+\frac{1}{\sqrt{n}}\right),$$

as follows from the definition of $\tilde{d}_n$, inequality (5.6) becomes (for all sufficiently large n and with P_θ-probability$\to1$)

$$(5.9)\qquad \frac{1}{\sqrt{n}}\left[\varphi\left(\tilde{d}_n+\frac{1}{\sqrt{n}}\right)-\varphi\left(\tilde{d}_n-\frac{1}{\sqrt{n}}\right)\right]+\int_0^1\left[\frac{1}{n}\sum_{j=1}^{n}\Psi_j\left(\tilde{d}_n+\frac{\lambda}{\sqrt{n}}\right)\right]dl(\lambda)$$
$$<-\int_0^1\left[\frac{1}{n}\sum_{j=1}^{n}\Psi_j\left(\tilde{d}_n-\frac{\lambda}{\sqrt{n}}\right)\right]dl(\lambda).$$

But $\tilde{d}_n\to\theta$ in P_θ-probability. By this result, Assumption (A4)-(i) and Corollary 5.1, relation (5.9) gives then the desired result.

6. Asymptotic normality of maximum probability estimates

In this section, it is shown that any $\sqrt{n}$-consistent MPE of θ is also asymptotically normal. More precisely, the following result is true.

THEOREM 6.1. *Let Assumptions* (A1)–(A6) *hold and let the estimate* $\hat{d}_n$ *of* θ *be as in Theorem* 5.1. *Then,*

$$L[\sqrt{n}(\hat{d}_n-\theta)\,|\,P_\theta]\Rightarrow N(0,\sigma^{-2}(\theta))\,.$$

The following lemma will facilitate the proof of the theorem.

LEMMA 6.1. *Let Assumptions* (A1)–(A6) *hold and let the estimate* $\hat{d}_n$ *of* θ *be as in Theorem* 5.1. *Then, for all sufficiently large* n *and with* P_θ*-probability*$\rightarrow 1$, *the r.v.'s*

$$\frac{1}{n}\sum_{j=1}^{n}\Psi_j\left[\theta+\lambda\left(\hat{d}_n-\theta\pm\frac{1}{\sqrt{n}}\right)\right], \qquad \lambda\in[0,1]\,,$$

given by (2.2), *are well-defined and*

$$\hat{I}_n^{\pm}=\int_0^1\left\{\frac{1}{n}\sum_{j=1}^{n}\Psi_j\left[\theta+\lambda\left(\hat{d}_n-\theta\pm\frac{1}{\sqrt{n}}\right)\right]\right\}dl(\lambda)\rightarrow -\sigma^2(\theta) \tag{6.1}$$

in P_θ*-probability.*

PROOF. It is done as in Corollary 5.1.

We may now proceed with the proof of the theorem.

PROOF OF THEOREM 6.1. From the definition of $\hat{d}_n$, it follows that, for sufficiently large n and with P_θ-probability$\rightarrow 1$,

$$\frac{\partial}{\partial\theta}Z_n(\theta)\,|_{\theta=\hat{d}_n}=0$$

(see also (4.2)). By Lemma 4.1, however, this equation is equivalent to relation (4.5) with d replaced by $\hat{d}_n$. Thus, for sufficiently large n and P_θ-probability$\rightarrow 1$,

$$(\hat{I}_n^{+}+\hat{I}_n^{-})\sqrt{n}(\hat{d}_n-\theta)=-2\Delta_n(\theta)-\frac{1}{2}[\sqrt{n}(\hat{d}_n-\theta)]^2(\hat{I}_n^{+}-\hat{I}_n^{-})$$
$$-\frac{1}{2}(\hat{I}_n^{+}-\hat{I}_n^{-})-\left[g_0\left(\hat{d}_n+\frac{1}{n}\right)-g_0\left(\hat{d}_n-\frac{1}{n}\right)\right].$$

By (6.1), one may divide through in the above equation by the r.v. $\hat{I}_n^{+}+\hat{I}_n^{-}$. Thus, for sufficiently large n and P_θ-probability$\rightarrow 1$,

$$\sqrt{n}(\hat{d}_n-\theta)=-\frac{2}{\hat{I}_n^{+}+\hat{I}_n^{-}}\Delta_n(\theta)-\frac{\hat{I}_n^{+}-\hat{I}_n^{-}}{2(\hat{I}_n^{+}+\hat{I}_n^{-})}[\sqrt{n}(\hat{d}_n-\theta)]^2 \tag{6.2}$$
$$-\frac{\hat{I}_n^{+}-\hat{I}_n^{-}}{2(\hat{I}_n^{+}+\hat{I}_n^{-})}-\frac{1}{\hat{I}_n^{+}+\hat{I}_n^{-}}\left[g_0\left(\hat{d}_n+\frac{1}{n}\right)-g_0\left(\hat{d}_n-\frac{1}{n}\right)\right].$$

By (4.18), $\sqrt{n}(\hat{d}_n-\theta)$ is bounded in P_θ-probability. Then, by (6.1) and Assumption (A4)-(i), the second, third and fourth term on the right-hand side of (6.2) tend to zero in P_θ-probability. Finally, by (4.16) and (6.1) again, the desired result follows.

7. Asymptotic efficiency of maximum probability estimates

In this section, it is shown that, under certain regularity conditions, an MPE is asymptotically efficient in the sense of Weiss-Wolfowitz. For the precise formulation of this result, some additional notation is required. To this end, define the following sequences of parameter points, namely,

$$\{\theta_n\} \subset \Theta , \qquad \text{where } \theta_n=\theta+\frac{h}{\sqrt{n}},\ n\geqq 1,\ h \in R . \tag{7.1}$$

Now, let W_n be an MPE satisfying the following two conditions:

(C1) For any $h \in R$, let θ_n be given by (7.1). Then, there exists $\beta(\theta) >0$ such that,

$$\lim P_{\theta_n}[\sqrt{n}(W_n-\theta_n) \in (-1, 1)]=\beta(\theta) .$$

(C2) Let θ_n be as above and let ε, δ be arbitrary >0 numbers. Then, for sufficiently large $|h|$,

$$\liminf P_{\theta_n}[|\sqrt{n}(W_n-\theta_n)|<\delta|h|]\geqq 1-\varepsilon .$$

Next, let T_n be any (competing) estimate of $\theta \in \Theta$ such that

$$\lim \{P_{\theta_n}[\sqrt{n}(T_n-\theta_n) \in (-1, 1)]-P_\theta[\sqrt{n}(T_n-\theta) \in (-1, 1)]\} =0 \tag{7.2}$$

for any $h \in R$ and any θ_n as above. With the above notation, the asymptotic efficiency of the MPE W_n is defined as follows

DEFINITION 7.1. Let $\{W_n\}$ and $\{T_n\}$, $n\geqq 1$, be as above. Then $\{W_n\}$ is said to be asymptotically efficient (in the sense of Weiss-Wolfowitz), if

$$\limsup P_\theta[\sqrt{n}(T_n-\theta) \in (-1, 1)]\leqq \beta(\theta) . \tag{7.3}$$

The main result of this section is then

THEOREM 7.1. *Let Assumptions* (A1)–(A7) *hold and let the MPE* $\hat{d}_n$ *given by Theorem* 5.1. *Then the sequence* $\{\hat{d}_n\}$ *is asymptotically efficient in the sense of Definition* 7.1.

PROOF. It follows immediately by Theorem 3.1 in [9], provided the following proposition is established, namely, PROPOSITION 7.1. Under Assumptions (A1)-(A7), the MPE $\hat{d}_n$ satisfies conditions (C1) and (C2) in this section.

The proof of this proposition will be facilitated by the following lemma.

LEMMA 7.1. *Let Assumptions* (A1)-(A7) *hold and let* $\{\theta_n\}$ *be given by* (7.1) *for* $h \in R$. *Then*:
(i) *The sequences of probability measures* $\{P_\theta\}$ *and* $\{P_{\theta_n}\}$ *are contiguous, and*
(ii) *For any* $0 \neq h \in R$,

$$L[\Delta_n(\theta) \mid P_{\theta_n}] \Rightarrow N(h\sigma^2(\theta), \sigma^2(\theta)) .$$

PROOF. (i) For $\{\theta_n\}$ as above and $\Lambda_n(\theta) = \log(dP_{n,\theta_n}/dP_{n,\theta})$, recall that, by Assumption (A7),

$$\Lambda_n(\theta) - h\Delta_n(\theta) \to -\frac{1}{2} h^2\sigma^2(\theta) \quad \text{in } P_\theta\text{-probability} , \tag{7.4}$$

where, we recall, that

$$\Delta_n(\theta) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \Phi_j(\theta) .$$

By (4.16),

$$L[\Delta_n(\theta) \mid P_\theta] \Rightarrow N(0, \sigma^2(\theta)) ,$$

so that

$$L[h\Delta_n(\theta) \mid P_\theta] \Rightarrow N(0, h^2\sigma^2(\theta)) . \tag{7.5}$$

Relations (7.4) and (7.5) give

$$L[\Lambda_n(\theta) \mid P_\theta] \Rightarrow N\left(-\frac{1}{2} h^2\sigma^2(\theta), h^2\sigma^2(\theta)\right) = L , \tag{7.6}$$

and, as is easily seen by integration,

$$\int_R \exp \lambda dL(\lambda) = 1 . \tag{7.7}$$

Then relations (7.6) and (7.7) imply the contiguity of the sequences $\{P_{n,\theta}\}$ and $\{P_{n,\theta_n}\}$ by Theorem 6.1-(iv), p. 33, in [5].
(ii) The contiguity established above and Corollary 7.2, p. 35, in [5] imply that

$$\Lambda_n(\theta) - h\Delta_n(\theta) \to -\frac{1}{2} h^2\sigma^2(\theta) \quad \text{in } P_{\theta_n}\text{-probability}$$

and

$$L[\Lambda_n(\theta)\,|\,P_{\theta_n}] \Rightarrow N\left(\frac{1}{2}h^2\sigma^2(\theta),\ h^2\sigma^2(\theta)\right).$$

From these last two relationships the desired result follows in an obvious manner.

We may now proceed with the proof of the proposition, namely,

PROOF OF PROPOSITION 7.1. By the contiguity established in Lemma 7.1-(i) and relation (6.1), it follows that

$$\hat{I}_n^{\pm} \to -\sigma^2(\theta) \quad \text{in } P_{\theta_n}\text{-probability}. \tag{7.8}$$

Also, by said contiguity and Assumption (A4)-(i),

$$g_0\left(\hat{d}_n+\frac{1}{\sqrt{n}}\right)-g_0\left(\hat{d}_n-\frac{1}{\sqrt{n}}\right)\to 0 \quad \text{in } P_{\theta_n}\text{-probability}. \tag{7.9}$$

Next, the sequence $\{\sqrt{n}(\hat{d}_n-\theta)\}$ is bounded in P_θ-probability, as follows from (4.18). This fact, along with the above mentioned contiguity and Proposition 6.1, p. 31, in [5], implies that $\{\sqrt{n}(\hat{d}_n-\theta)\}$ is also bounded in P_θ-probability. Now, this result and relations (7.8), (7.9) and (6.2) imply, by means of Lemma 7.1-(ii), that

$$L[\sqrt{n}(\hat{d}_n-\theta)\,|\,P_{\theta_n}] \Rightarrow N(h,\sigma^{-2}(\theta)).$$

Therefore, letting Φ stand for the distribution function of the standard normal distribution,

$$\begin{aligned} &P_{\theta_n}[\sqrt{n}(\hat{d}_n-\theta_n)\in(-1,1)] \\ &\quad = P_{\theta_n}[-1<\sqrt{n}(\hat{d}_n-\theta)-h<1] \to 2\Phi[\sigma(\theta)]-1. \end{aligned}$$

It follows that (C1) holds true with $\beta(\theta)=2\Phi[\sigma(\theta)]-1$. For the proof of (C2), it is seen that

$$\begin{aligned} &P_{\theta_n}[|\sqrt{n}(\hat{d}_n-\theta_n)|<\delta|h|] \\ &\quad = P_{\theta_n}[-\delta|h|<\sqrt{n}(\hat{d}_n-\theta)-h<\delta|h|] \to 2\Phi[\delta|h|\sigma(\theta)]-1. \end{aligned}$$

Since $2\Phi[\delta|h|\sigma(\theta)]-1$ can be made $\geq 1-\varepsilon$, by choosing $|h|$ sufficiently large, Condition (C2) is also true and, in fact, with the lim inf being replaced by lim.

Acknowledgement

The author wishes to express his thanks to the referee for his thoughtful and constructive comments which helped to improve a previous version of this paper.

UNIVERSITY OF WISCONSIN AND UNIVERSITY OF PATRAS

References

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

[2] Hájek, J. (1972). Local asymptotic minimax and admissibility in estimation, *Proc. Sixth Berkeley Symp. Math. Statist. Prob.*, 1, 175-194, Univ. of California Press.

[3] Loève, M. (1963). *Probability Theory*, 3rd ed. Van Nostrand, Princeton.

[4] Roussas, G. G. (1968). Asymptotic normality of the maximum likelihood estimate in Markov processes, *Metrika*, **14**, 62-70.

[5] Roussas, G. G. (1972). *Contiguity of Probability Measures*, Cambridge University Press.

[6] Roussas, G. G. (1973). *A First Course in Mathematical Statistics*, Addison-Wesley, Reading, Massachussetts.

[7] Roussas, G. G. (1975). Asymptotic properties of maximum probability estimates in the i.i.d. case, *Proceedings of the Institute of Mathematical Statistics Summer Institute*, Indiana University, Bloomington, Indiana.

[8] Roussas, G. G. (1975). Asymptotic efficiency of the maximum probability estimate in the i.i.d. case, (Submitted for publication).

[9] Weiss, L. and Wolfowitz, J. (1975). *Asymptotic Methods in Statistics*, Springer-Verlag, New York, Heidelberg, Berlin.