ASYMPTOTIC NORMALITY OF THE MAXIMUM LIKELIHOOD ESTIMATE IN THE INDEPENDENT NOT IDENTICALLY DISTRIBUTED CASE

A. N. PHILIPPOU AND G. G. ROUSSAS*

(Received April 27, 1972; revised Oct. 29, 1973)

Abstract

In this paper, we assume the existence and consistency of the maximum likelihood estimate (MLE) in the independent not identically distributed (i.n.i.d.) case and we establish its asymptotic normality. The regularity conditions employed do not involve the third order derivatives of the underlying probability density functions (p.d.f.'s).

1. Introduction and summary

The asymptotic normality of the MLE in the independent identically distributed (i.i.d.) case is well established under a variety of conditions. For example, Cramér [3] assumes, among other things, the existence and boundedness of the third derivative of the underlying p.d.f., while Gurland [4] and Kulldorff [7] give regularity conditions, which involve the behavior of the first and second order derivatives only. The same problem has also been looked into by Huber [6] who established the asymptotic normality of the MLE under non-standard conditions. Recently non-i.i.d. cases have been considered by some authors. We mention Silvey [10] who treated processes with dependent random variables satisfying general regularity conditions; Billingsley [1] and Roussas [9] who dealt with Markov processes, which are stationary and ergodic; and, in particular, Bradley and Gart [2] who studied the i.n.i.d. case. The latter authors established asymptotic normality of the MLE under regularity conditions which include the existence and boundedness of the third derivatives of the underlying p.d.f.'s. Finally, in an interesting recent paper, Hoadley [5] established consistency and asymptotic normality of the MLE in the i.n.i.d. case. In this paper, we prove the asymptotic normality of the MLE in the i.n.i.d. case

* This research was supported by the National Science Foundation, Grant GR-20036, and the Office of Scientific Research and Development of the Greek Government.

under assumptions which involve the first and second order derivatives only. Our assumptions are related to those of Gurland [4], Hoadley [5] and Roussas [9].

In this note, by an MLE we mean a specified measurable function of $X_1, \cdots, X_n$ which is a root of the log-likelihood equation with P_θ-probability tending to 1, as $n \to \infty$, for all $\theta \in \Theta$, the underlying parameter space.

In Section 2, the assumptions to be used in this paper are spelled out, and the main result is stated in Section 3. Its proof is deferred to Section 5, after some auxiliary results have been established in Section 4. The paper is concluded with Section 6, where some simple examples are discussed for illustrative purposes.

In order to avoid unnecessary repetitions in the sequel, all limits are taken as $n \to \infty$ through positive integer values, unless otherwise explicitly specified.

2. Notation and assumptions

Let Θ be an open subset of R^k and for each $\theta \in \Theta$, let $p_{j,\theta}$, $j \geqq 1$ (integer) be probability measures on $(R_j, \mathscr{B}_j)$, where $(R_j, \mathscr{B}_j) = (R, \mathscr{B})$, the Borel real line. We assume that there is a σ-finite measure μ on $\mathscr{B}$ such that $p_{j,\theta} \ll \mu$, $\theta \in \Theta$, $j \geqq 1$, and we set $f_j(\cdot\,;\theta) = dp_{j,\theta}/d\mu$ for a specified version of the Radon-Nikodym derivative involved. Set $(\mathscr{X}, \mathscr{A}) = (R^\infty, \mathscr{B}^\infty)$ and let P_θ be the product measure of $p_{j,\theta}$, $j \geqq 1$, induced on $\mathscr{A}$. Then, if X_j, $j \geqq 1$, are the co-ordinate random variables (r.v.'s), it follows that, for each $\theta \in \Theta$, these r.v.'s are independent, and the p.d.f. of the jth r.v. is $f_j(\cdot\,;\theta)$.

The asymptotic normality of the MLE will be established under the following assumptions.

ASSUMPTIONS

(A1) Θ is an open subset of R^k.

(A2) For each $j \geqq 1$ and every $\theta \in \Theta$, the r.v. X_j has a p.d.f. $f_j(\cdot\,;\theta)$ with respect to a σ-finite measure μ on $\mathscr{B}$, and the set $\{x \in R;\ f_j(x;\theta) > 0\}$ is independent of θ.

(A3) Let $\hat{\theta}_n$ be the MLE of the parameter θ based on $X_1, \cdots, X_n$. Then $\hat{\theta}_n \xrightarrow{P_\theta} \theta$, for every $\theta \in \Theta$.

(A4) (i) For almost all $[\mu]$ $x_j \in R$ and for all $\theta \in \Theta$, $\phi_{jr}(\theta) = \partial/\partial\theta_r \cdot \log f_j(x_j;\theta)$ and $\psi_{jrs}(\theta) = -\partial^2/\partial\theta_r\partial\theta_s \log f_j(x_j;\theta)$ exist, $r, s = 1, \cdots, k$, $j \geqq 1$.

(ii) For almost all $[\mu]$ $x_j \in R$, every $\theta \in \Theta$ and each $r, s = 1, \cdots, k$,

$$\psi_{jrs}[\theta + \lambda(t-\theta)] \to \psi_{jrs}(\theta) \qquad \text{as } t \to \theta\,,$$

uniformly in $j \geqq 1$ and $\lambda \in [0, 1]$.

In some interesting cases, the uniformity assumption (A4)-(ii) is either not easy to verify or does not hold. For such cases, we purpose the following Lipschitz-type condition as a possible useful replacement.

(ii′) For every $\theta \in \Theta$, there exist a neighborhood of it N_θ in Θ and r.v.'s $Z_{jrs}(\theta)$ such that

$$|\psi_{jrs}(t)-\psi_{jrs}(\theta)| \leqq \|t-\theta\| Z_{jrs}(\theta) , \qquad t \in N_\theta$$

and

$$\left\{\frac{1}{n}\sum_{j=1}^{n} Z_{jrs}(\theta)\right\} \text{ is bounded in } P_\theta\text{-probability} ;$$

that is, for every $\varepsilon>0$, there exist $M_{rs}(\theta, \varepsilon)$ and $N_{rs}(\theta, \varepsilon)$, positive integers, such that

$$\mathrm{P}_\theta\left[\frac{1}{n}\sum_{j=1}^{n} Z_{jrs}(\theta) \geqq M_{rs}(\theta, \varepsilon)\right] < \varepsilon \qquad \text{for all } n \geqq N_{rs}(\theta, \varepsilon) ,$$

$r, s=1,\cdots, k$.

In the sequel, the notation $\phi_{jr}(\theta)$ and $\psi_{jrs}(\theta)$ will also be used for the r.v.'s resulting from $\phi_{jr}(\theta)$ and $\psi_{jrs}(\theta)$ when x_j is replaced by X_j.

Let $\phi_j(\theta)$ be the $k\times 1$ random vector and let $\psi_j(\theta)$ be the $k\times k$ random matrix given, respectively, by

$$\phi_j(\theta)=[\phi_{j1}(\theta),\cdots, \phi_{jk}(\theta)]' , \qquad j \geqq 1 \tag{2.1}$$

and

$$\psi_j(\theta)=\begin{bmatrix} \psi_{j11}(\theta) & \cdots & \psi_{j1k}(\theta) \\ \vdots & & \vdots \\ \psi_{jk1}(\theta) & \cdots & \psi_{jkk}(\theta) \end{bmatrix} , \qquad j \geqq 1 . \tag{2.2}$$

Then

(A5) $\mathcal{E}_\theta[\phi_j(\theta)]=0$ and $\mathcal{E}_\theta[\phi_j(\theta)\phi_j'(\theta)]=\mathcal{E}_\theta[\psi_j(\theta)]$ for each $j \geqq 1$ and all $\theta \in \Theta$.

For each $j \geqq 1$, $n \geqq 1$ and every $\theta \in \Theta$, define $\Gamma_j(\theta)$ and $\bar{\Gamma}_n(\theta)$ by

$$\Gamma_j(\theta)=\mathcal{E}_\theta[\phi_j(\theta)\phi_j'(\theta)] , \qquad \bar{\Gamma}_n(\theta)=\frac{1}{n}\sum_{j=1}^{n} \Gamma_j(\theta) . \tag{2.3}$$

Then

(A6) For every $\theta \in \Theta$, $\bar{\Gamma}_n(\theta) \to \bar{\Gamma}(\theta)$, some $\bar{\Gamma}(\theta)$ which is positive definite, and the convergence is convergence in any one of the usual norms for matrices.

(A7) For every $h \in R^k$ and every $\theta \in \Theta$, there exists $\delta\ (=\delta(h, \theta))>0$

such that

$$\frac{1}{n^{(2+\delta)/2}} \sum_{j=1}^{n} \mathcal{E}_\theta |h'\phi_j(\theta)|^{2+\delta} \to 0 .$$

(For example, this happens if $\mathcal{E}_\theta |h'\phi_j(\theta)|^3 \leqq M$ $(=M(h, \theta)) < \infty$.)

(A8) For every $\theta \in \Theta$, there exists δ $(=\delta(\theta))>0$ such that

$$\frac{1}{n^{1+\delta}} \sum_{j=1}^{n} \mathcal{E}_\theta |\psi_{jrs}(\theta) - \mathcal{E}_\theta[\psi_{jrs}(\theta)]|^{1+\delta} \to 0 , \qquad r, s=1, \cdots, k .$$

Under either one of the sets of Assumptions (A1)–(A4)-(i), (ii)–(A8) or (A1)–(A4)-(i), (ii′)–(A8), one may establish the asymptotic normality of the MLE.

3. Main result

The main result of this paper is the following.

THEOREM. *Let $\hat{\theta}_n$ denote the MLE of θ, based on X_j, $j=1, \cdots, n$ which are assumed to be i.n.i.d. r.v.'s with p.d.f.'s $f_j(x_j; \theta)$. Let θ_0 be the true (but unknown) value of the parameter θ and let $\bar{\Gamma}(\theta_0)$ be as in* (A6). *Then under either one set of Assumptions* (A1)–(A4)-(i), (ii)–(A8) *or* (A1)–(A4)-(i), (ii′)–(A8),

$$\mathcal{L}[\sqrt{n}(\hat{\theta}_n - \theta_0) | P_{\theta_0}] \Rightarrow N(0, \bar{\Gamma}^{-1}(\theta_0)) ;$$

(that is, as $n \to \infty$, the sequence of r.v.'s $\{\sqrt{n}(\hat{\theta}_n - \theta_0)\}$ converges in distribution, under P_{θ_0}, to an r.v. which is distributed as $N(0, \bar{\Gamma}^{-1}(\theta_0))$).

4. Some auxiliary results

The proof of the theorem will be given after some auxiliary results have been established.

PROPOSITION 4.1. Let $\hat{\theta}_n$ be the MLE of θ. Then under either one of the sets of Assumptions (A1)–(A4)-(i), (ii)–(A8), or (A1)–(A4)-(i), (ii′)–(A8), and for each $\theta \in \Theta$, and each $r, s=1, \cdots, k$, one has

$$\frac{1}{n} \sum_{j=1}^{n} \{\psi_{jrs}[\theta + \lambda(\hat{\theta}_n - \theta)] - \psi_{jrs}(\theta)\} \xrightarrow{P_\theta} 0 \qquad \text{uniformly in } \lambda \in [0, 1] ;$$

that is, for each $\theta \in \Theta$, each $r, s=1, \cdots, k$ and every $\varepsilon>0$, there exists $N(\theta, r, s; \varepsilon)$, to be denoted by $N(\varepsilon)$ since θ, r, s remain fixed, independent of $\lambda \in [0, 1]$, such that

$$\mathrm{P}_\theta \left(\left| \frac{1}{n} \sum_{j=1}^{n} \{\psi_{jrs}[\theta + \lambda(\hat{\theta}_n - \theta)] - \psi_{jrs}(\theta)\} \right| < \varepsilon \right) > 1 - \varepsilon , \qquad n \geqq N(\varepsilon) .$$

PROOF. Suppose first that (A1)-(A4)-(i), (ii)-(A8) hold. From (A4)-(ii), one has that for each $\theta \in \Theta$, each $r, s=1,\cdots,k$ and every $\varepsilon>0$, there exists $\delta(\theta, r, s; \varepsilon)>0$, to be denoted by $\delta(\varepsilon)$ since θ, r, s remain fixed, independent of $j\geqq 1$ and $\lambda \in [0, 1]$, such that

$$(4.1) \qquad |\psi_{jrs}[\theta+\lambda(t-\theta)]-\psi_{jrs}(\theta)|<\varepsilon\,, \qquad ||t-\theta||<\delta(\varepsilon)\,.$$

On the other hand, by (A3), $\hat{\theta}_n \xrightarrow{P_\theta} \theta$ so that for the above $\delta(\varepsilon)>0$ there exists $N_0(\delta(\varepsilon))=N(\varepsilon)$ such that

$$\mathrm{P}_\theta\,[||\hat{\theta}_n-\theta||<\delta(\varepsilon)]>1-\varepsilon\,, \qquad n\geqq N(\varepsilon)\,.$$

Thus, if $A_n(\theta;\varepsilon)$, to be denoted by $A_n(\varepsilon)$ since θ remains fixed, is the set for which $||\hat{\theta}_n-\theta||<\delta(\varepsilon)$, then

$$(4.2) \qquad \mathrm{P}_\theta\,[A_n(\varepsilon)]>1-\varepsilon\,, \qquad n\geqq N(\varepsilon)\,,$$

and on this set $A_n(\varepsilon)$,

$$(4.3) \qquad ||\hat{\theta}_n-\theta||<\delta(\varepsilon)\,, \qquad n\geqq N(\varepsilon)\,.$$

Relations (4.1)-(4.3) imply then that on the set $A_n(\varepsilon)$ with $\mathrm{P}_\theta\,[A_n(\varepsilon)]>1-\varepsilon$, provided $n\geqq N(\varepsilon)$, one has that, for each $\theta \in \Theta$ and each $r, s=1,\cdots,k$,

$$(4.4) \qquad |\psi_{jrs}[\theta+\lambda(\hat{\theta}_n-\theta)]-\psi_{jrs}(\theta)|<\varepsilon$$

simultaneously for all $j\geqq 1$ and all $\lambda \in [0, 1]$. From (4.4), it follows then that on the same set $A_n(\varepsilon)$ with $\mathrm{P}_\theta\,[A_n(\varepsilon)]>1-\varepsilon$, provided $n\geqq N(\varepsilon)$, one has that, for each $\theta \in \Theta$ and each $r, s=1,\cdots,k$,

$$(4.5) \qquad \left|\frac{1}{n}\sum_{j=1}^{n}\{\psi_{jrs}[\theta+\lambda(\hat{\theta}_n-\theta)]-\psi_{jrs}(\theta)\}\right|<\varepsilon$$

simultaneously for all $\lambda \in [0, 1]$.

Now, since the set of points of $\mathcal{X}$ for which (4.5) is valid, clearly, contains the set $A_n(\varepsilon)$, we have that, for every $\varepsilon>0$ there exists $N(\varepsilon)$ independent of $\lambda \in [0, 1]$ (but perhaps depending on θ, r, s which remain fixed throughout the proof) such that

$$\mathrm{P}_\theta\left(\left|\frac{1}{n}\sum_{j=1}^{n}\{\psi_{jrs}[\theta+\lambda(\hat{\theta}_n-\theta)]-\psi_{jrs}(\theta)\}\right|<\varepsilon\right)>1-\varepsilon\,, \qquad n\geqq N(\varepsilon)\,,$$

simultaneously for all $\lambda \in [0, 1]$.

This establishes then the desired result.

Next, suppose that (A1)-(A4)-(i), (ii′)-(A8) hold. Then, by (A3), $\hat{\theta}_n \in N_\theta$ on a set which depends only on n and θ and whose P_θ-probability tends to 1. Therefore on the above mentioned set, one has by means of (A4)-(ii′), (4.6) $\quad |\phi_{jrs}[\theta+\lambda(\hat{\theta}_n-\theta)]-\phi_{jrs}(\theta)|\leqq\|\lambda(\hat{\theta}_n-\theta)\|Z_{jrs}(\theta)\leqq\|\hat{\theta}_n-\theta\|Z_{jrs}(\theta)$

for each $\theta\in\Theta$, each $r, s=1,\cdots,k$ and simultaneously for all $j\geqq 1$ and all $\lambda\in[0,1]$.
From relation (4.6) and once again by means of (A3) and (A4)-(ii′), one has that for each $\theta\in\Theta$ and each $r, s=1,\cdots,k$,

$$\frac{1}{n}\sum_{j=1}^{n}\{\phi_{jrs}[\theta+\lambda(\hat{\theta}_n-\theta)]-\phi_{jrs}(\theta)\}\xrightarrow{P_\theta}0$$

uniformly in $\lambda\in[0,1]$.
This completes the proof of the proposition.

Let $\phi_j(\theta)$ be given by (2.2), and for each $\lambda\in[0,1]$ and t sufficiently close to θ, so that $[\theta+\lambda(t-\theta)]\in\Theta$, define $\bar{\phi}_n[\theta+\lambda(t-\theta)]$ by

$$\bar{\phi}_n[\theta+\lambda(t-\theta)]=\frac{1}{n}\sum_{j=1}^{n}\phi_j[\theta+\lambda(t-\theta)]\ . \tag{4.7}$$

Then we have the following result.

PROPOSITION 4.2. *Let $\bar{\phi}_n[\theta+\lambda(\hat{\theta}_n-\theta)]$ be given by (4.7), where $\hat{\theta}_n$ is the MLE of θ. Then under either one of the sets of Assumptions (A1)-(A4)-(i), (ii)-(A8) or (A1)-(A4)-(i), (ii′)-(A8), and for every $\theta\in\Theta$, one has*

$$\bar{\phi}_n[\theta+\lambda(\hat{\theta}_n-\theta)]\xrightarrow{P_\theta}\bar{\Gamma}(\theta)\qquad \text{uniformly in } \lambda\in[0,1]\ .$$

PROOF. Since $\bar{\Gamma}_n(\theta)\to\bar{\Gamma}(\theta)$, $\theta\in\Theta$, by (A6), it suffices to show that $\bar{\phi}_n[\theta+\lambda(\hat{\theta}_n-\theta)]-\bar{\Gamma}_n(\theta)\xrightarrow{P_\theta}0$, $\theta\in\Theta$, uniformly in $\lambda\in[0,1]$. By means of (4.7), (2.2) and (2.3), this last convergence is equivalent to the following convergence

$$\frac{1}{n}\sum_{j=1}^{n}\{\phi_{jrs}[\theta+\lambda(\hat{\theta}_n-\theta)]-\mathcal{E}_\theta\phi_{jrs}(\theta)\}\xrightarrow{P_\theta}0\ , \tag{4.8}$$

$$\theta\in\Theta,\ \text{uniformly in } \lambda\in[0,1],\ r, s=1,\cdots,k\ .$$

The left-hand side of the relation in (4.8) is written as follows

$$\begin{aligned}(4.9)\quad &\frac{1}{n}\sum_{j=1}^{n}\{\phi_{jrs}[\theta+\lambda(\hat{\theta}_n-\theta)]-\mathcal{E}_\theta\phi_{jrs}(\theta)\}\\ &=\frac{1}{n}\sum_{j=1}^{n}\{\phi_{jrs}[\theta+\lambda(\hat{\theta}_n-\theta)]-\phi_{jrs}(\theta)\}+\frac{1}{n}\sum_{j=1}^{n}[\phi_{jrs}(\theta)-\mathcal{E}_\theta\phi_{jrs}(\theta)]\ .\end{aligned}$$

But for every $\theta\in\Theta$ and $r, s=1,\cdots,k$,

$$\frac{1}{n}\sum_{j=1}^{n}\{\phi_{jrs}[\theta+\lambda(\hat{\theta}_n-\theta)]-\phi_{jrs}(\theta)\}\xrightarrow{P_\theta}0\qquad \text{uniformly in } \lambda\in[0,1]\ , \tag{4.10}$$

by Proposition 4.1, and

$$\frac{1}{n}\sum_{j=1}^{n}[\psi_{jrs}(\theta)-\mathcal{E}_{\theta}\psi_{jrs}(\theta)]\xrightarrow{P_\theta}0 \tag{4.11}$$

by the Weak law of large numbers (see, e.g., Loève [8], p. 275) which applies here on account of (A8). Relations (4.9)–(4.11) establish (4.8) and hence the proposition itself.

This section is closed with the following result.

PROPOSITION 4.3. Let $\tilde{\phi}_n(\theta)$ be given by

$$\tilde{\phi}_n(\theta)=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\phi_j(\theta)\ , \qquad \theta\in\Theta\ , \tag{4.12}$$

where $\phi_j(\theta)$ is given by (2.1). Then for every $\theta\in\Theta$,

$$\mathcal{L}[\tilde{\phi}_n(\theta)\,|\,P_\theta]\Rightarrow N(0,\bar{\Gamma}(\theta))\ .$$

PROOF. It suffices to show that for every $0\neq h\in R^k$,

$$\mathcal{L}[h'\tilde{\phi}_n(\theta)\,|\,P_\theta]\Rightarrow N(0,h'\bar{\Gamma}(\theta)h)\ ,\qquad \theta\in\Theta\ .$$

For each h as above, set $\phi^*_{nj}(\theta)=(1/\sqrt{n})h'\phi_j(\theta)$, $1\leqq j\leqq n$. Then $h'\tilde{\phi}_n(\theta)=\sum_{j=1}^{n}\phi^*_{nj}(\theta)$ by (4.12), and $\phi^*_{nj}(\theta)$, $1\leqq j\leqq n$ are independent r.v.'s with $\mathcal{E}_\theta\phi^*_{nj}(\theta)=0$, by (A5). Next, by means of (A6)

$$s_n^2=\sum_{j=1}^{n}\sigma_\theta^2[\phi^*_{nj}(\theta)]=h'\bar{\Gamma}_n(\theta)h\rightarrow h'\bar{\Gamma}(\theta)h>0\ ,$$

so that

$$\frac{1}{s_n^{2+\delta}}\sum_{j=1}^{n}\mathcal{E}_\theta|\phi^*_{nj}(\theta)|^{2+\delta}=\frac{1}{s_n^{2+\delta}}\,\frac{1}{n^{(2+\delta)/2}}\sum_{j=1}^{n}\mathcal{E}_\theta|h'\phi_j(\theta)|^{2+\delta}\rightarrow 0$$

on account of (A7). Therefore Liapounov's condition for the Central limit theorem to hold (see, e.g., Loève [8], p. 275) is satisfied and hence

$$\mathcal{L}[h'\tilde{\phi}_n(\theta)\,|\,P_\theta]\Rightarrow N(0,h'\bar{\Gamma}(\theta)h)\ ,$$

as was to be seen.

5. Proof of the main result

We are now ready to give the proof of the theorem.

PROOF OF THE THEOREM. We wish to show that

$$\mathcal{L}[\sqrt{n}(\hat{\theta}_n-\theta_0)\,|\,P_{\theta_0}]\Rightarrow N(0,\bar{\Gamma}^{-1}(\theta_0))\ . \tag{5.1}$$

Expanding $\phi_j(\theta)$ around θ_0 according to Taylor's formula, we obtain

$$\phi_j(\theta)=\phi_j(\theta_0)-\left\{\int_0^1 \dot{\phi}_j[\theta_0+\lambda(\theta-\theta_0)]d\lambda\right\}(\theta-\theta_0)\ . \tag{5.2}$$

In (5.2), we sum over j, $1\leqq j\leqq n$, divide through by $\sqrt{n}$ and replace θ by $\hat{\theta}_n$. Then by also utilizing (4.12) and (4.7), we get

$$\tilde{\phi}_n(\hat{\theta}_n)=\tilde{\phi}_n(\theta_0)-\left\{\int_0^1 \bar{\dot{\phi}}_n[\theta_0+\lambda(\hat{\theta}_n-\theta_0)]d\lambda\right\}\sqrt{n}(\hat{\theta}_n-\theta_0)\ .$$

But $\tilde{\phi}_n(\hat{\theta}_n)=0$ with P_{θ_0}-probability tending to 1. Thus with P_{θ_0}-probability tending to 1,

$$\left\{\int_0^1 \bar{\dot{\phi}}_n[\theta_0+\lambda(\hat{\theta}_n-\theta_0)]d\lambda\right\}\sqrt{n}(\hat{\theta}_n-\theta_0)=\tilde{\phi}_n(\theta_0)\ . \tag{5.3}$$

By Proposition 4.2, $\bar{\dot{\phi}}_n[\theta_0+\lambda(\hat{\theta}_n-\theta_0)]\xrightarrow{P_{\theta_0}}\bar{\Gamma}(\theta_0)$ uniformly in $\lambda\in[0, 1]$ and $\bar{\Gamma}(\theta_0)$ is positive definite, by (A6). Hence

$$\int_0^1 \bar{\dot{\phi}}_n[\theta_0+\lambda(\hat{\theta}_n-\theta_0)]d\lambda\xrightarrow{P_{\theta_0}}\bar{\Gamma}(\theta_0)\ . \tag{5.4}$$

From this result and (5.3), it follows that with P_{θ_0}-probability tending to 1, one has

$$\sqrt{n}(\hat{\theta}_n-\theta_0)=\left\{\int_0^1 \bar{\dot{\phi}}_n[\theta_0+\lambda(\hat{\theta}_n-\theta_0)]d\lambda\right\}^{-1}\tilde{\phi}_n(\theta_0)\ . \tag{5.5}$$

By taking now into consideration the fact that $\mathcal{L}[\tilde{\phi}_n(\theta_0)\,|\,P_{\theta_0}]\Rightarrow N(0, \bar{\Gamma}(\theta_0))$ (Proposition 4.3) and (5.4), relation (5.5) gives

$$\mathcal{L}[\sqrt{n}(\hat{\theta}_n-\theta_0)\,|\,P_{\theta_0}]\Rightarrow N(0, \bar{\Gamma}^{-1}(\theta_0))$$

by the standard Slutsky's theorems. The proof of the theorem is complete.

6. Some examples

For the sake of illustrating the validity of the assumptions made in this paper, we discuss the following simple examples.

Example 1. Let X_j, $j=1,\cdots,n$ be independent r.v.'s such that the r.v. X_j is distributed as $N(\lambda_j\theta, \sigma^2)$, where σ and $\lambda_j=\lambda_j(n)$, $1\leqq j\leqq n$ are assumed to be known. We also assume that the λ's satisfy the following conditions

$$\frac{1}{n}\sum_{j=1}^{n}\lambda_j^2\to\lambda>0 \qquad \text{and} \qquad \frac{1}{n^{3/2}}\sum_{j=1}^{n}|\lambda_j|^3\to 0\ . \tag{6.1}$$

(For example, these conditions are satisfied if $\lambda_j=j/n$.)

Here it is easily seen that $\hat{\theta}_n = \left(1 \Big/ \sum_{j=1}^{n} \lambda_j^2\right) \sum_{j=1}^{n} \lambda_j X_j$, and hence $\hat{\theta}_n \xrightarrow{P_\theta} \theta$, on account of (6.1). Next, $\phi_j(\theta) = (\lambda_j/\sigma^2)(X_j - \lambda_j\theta)$ and $\psi_j(\theta) = \lambda_j^2/\sigma^2$, so that both (A4)-(ii) and (A4)-(ii′) are, trivially, satisfied. Furthermore, $\mathcal{E}_\theta\phi_j(\theta) = 0$ and $\mathcal{E}_\theta\phi_j^2(\theta) = \lambda_j^2/\sigma^2 = \mathcal{E}_\theta\psi_j(\theta)$, so that (A5) is satisfied. Assumption (A6) is also satisfied on account of (6.1). Next, it is easily seen that

$$\mathcal{E}_\theta \left| \frac{\lambda_j}{\sigma^2}(X_j - \lambda_j\theta) \right|^3 \leqq \frac{\sqrt{3}}{\sigma^3} |\lambda_j|^3 ,$$

so that, by means of (6.1), (A7) holds true with $\delta = 1$. Finally, (A8) is, trivially, true, so that (A1)-(A8) are satisfied in the present example.

Example 2. Let X_j, $j = 1, \cdots, n$ be independent r.v.'s such that the r.v. X_j is distributed as $N(\mu_j, \theta)$, where μ_j, $1 \leqq j \leqq n$ are assumed to be known.

It is easily seen that $\hat{\theta}_n = 1/n \sum_{j=1}^{n} (X_j - \mu_j)^2$, so that $\hat{\theta}_n \xrightarrow{P_\theta} \theta$. Next, $\phi_j(\theta) = [(X_j - \mu_j)^2 - \theta]/2\theta^2$ and $\psi_j(\theta) = [2(X_j - \mu_j)^2 - \theta]/2\theta^3$. Thus for $\varepsilon > 0$ (such that $\theta - \varepsilon > 0$) and $t \in (\theta - \varepsilon, \theta + \varepsilon)$, we have

$$|\psi_j(\theta) - \psi_j(t)| \leqq |t - \theta| \left[\frac{2\theta + \varepsilon}{2\theta^2(\theta - \varepsilon)^2} + \frac{(\theta + \varepsilon)^2 + (\theta + \varepsilon)\theta + \theta^2}{(\theta - \varepsilon)^3\theta^3} (X_j - \mu_j)^2 \right]$$

and

$$\mathrm{P}_\theta \left[\frac{2\theta + \varepsilon}{2\theta^2(\theta - \varepsilon)^2} + \frac{(\theta + \varepsilon)^2 + (\theta + \varepsilon)\theta + \theta^2}{(\theta - \varepsilon)^3\theta^3} \frac{1}{n} \sum_{j=1}^{n} (X_j - \mu_j)^2 \geqq M \right] < \varepsilon$$

for sufficiently large M. Hence (A4)-(ii′) is satisfied. Furthermore, $\mathcal{E}_\theta\phi_j(\theta) = 0$ and $\mathcal{E}_\theta\phi_j^2(\theta) = 1/2\theta^2 = \mathcal{E}_\theta\psi_j(\theta)$, so that (A5) is true. Assumption (A6) is, trivially, true. Next, it is easily seen that

$$\mathcal{E}_\theta \left| \frac{(X_j - \mu_j)^2 - \theta}{2\theta^2} \right|^3 \leqq \frac{7}{2\theta^3}$$

which implies that (A7) is satisfied with $\delta = 1$. Finally, we have

$$\mathcal{E}_\theta \left[\frac{(X_j - \mu_j)^2 - \theta}{\theta^3} \right]^2 = \frac{2}{\theta^4}$$

which implies that (A8) also holds true with $\delta = 1$.

Example 3. Let X_j, $j = 1, \cdots, n$ be independent r.v.'s such that the r.v. X_j has the negative exponential distribution with parameter $\lambda_j\theta$, where $\lambda_j = \lambda_j(n) > 0$, $1 \leqq j \leqq n$ are assumed to be known.

It is easily seen that $\hat{\theta}_n = 1 \Big/ \left(1/n \cdot \sum_{j=1}^{n} \lambda_j X_j\right)$ and hence $\hat{\theta}_n \xrightarrow{P_\theta} \theta$. Next, $\phi_j(\theta) = 1/\theta - \lambda_j X_j$ and $\psi_j(\theta) = 1/\theta^2$. Hence $|\psi_j[\theta + \lambda(t - \theta)] - \psi_j(\theta)|$ is easily seen to be bounded above by $|\theta-t|((\theta+t)/\theta^4)$ for $t>\theta$ and by $|\theta-t|((3\theta-t)/\theta^2t^2)$ for $t<\theta$. Therefore (A4)-(ii) is satisfied. Also, for $\varepsilon>0$ (such that $\theta-\varepsilon>0$) and $t\in(\theta-\varepsilon, \theta+\varepsilon)$, $|\psi_j(\theta)-\psi_j(t)|\leqq|\theta-t|((2\theta+\varepsilon)/\theta^2(\theta-\varepsilon)^2)$, so that (A4)-(ii′) is satisfied as well. Furthermore, $\mathcal{E}_\theta\phi_j(\theta)=0$ and $\mathcal{E}_\theta\phi_j^2(\theta)=1/\theta^2=\mathcal{E}_\theta\psi_j(\theta)$, so that (A5) holds true. Assumption (A6) is, clearly, satisfied. Next, it is easily seen that $\mathcal{E}_\theta|1/\theta-\lambda_jX_j|^3\leqq16/\theta^3$, so that (A7) holds true for $\delta=1$. Finally, (A8) is, trivially, true, so that (A1)–(A8) are satisfied in this example.

Example 4. Let X_j, $j=1,\cdots,n$ be independent r.v.'s such that X_j is distributed as $P_0(\lambda_j\theta)$, where $0<\lambda_j=\lambda_j(n)$, $1\leqq j\leqq n$ are assumed to be known and to satisfy the following conditions

$$\text{(6.2)}\qquad \frac{1}{n}\sum_{j=1}^{n}\lambda_j\to\bar{\lambda}>0 \quad\text{and}\quad \frac{1}{n^{3/2}}\sum_{j=1}^{n}\lambda_j^m\to0 \quad\text{for } m=2,3\,.$$

(For example, these conditions are satisfied for $\lambda_j=\lambda_j(n)=j/n$.)

Here it is easily seen that $\hat{\theta}_n=\sum_{j=1}^{n}X_j\Big/\sum_{j=1}^{n}\lambda_j$, so that $\hat{\theta}_n\xrightarrow{P_\theta}\theta$, on account of (6.2). Next, $\phi_j(\theta)=(X_j-\lambda_j\theta)/\theta$ and $\psi_j(\theta)=X_j/\theta^2$. Thus for $\varepsilon>0$ (such that $\theta-\varepsilon>0$) and $t\in(\theta-\varepsilon, \theta+\varepsilon)$,

$$|\psi_j(\theta)-\psi_j(t)|\leqq|\theta-t|\frac{2\theta+\varepsilon}{\theta^2(\theta-\varepsilon)^2}X_j\ ,$$

and

$$\mathrm{P}_\theta\left[\frac{2\theta+\varepsilon}{\theta^2(\theta-\varepsilon)^2}\frac{1}{n}\sum_{j=1}^{n}X_j\geqq M\right]<\varepsilon$$

for sufficiently large M and sufficiently large n, by (6.2). Hence (A4)-(ii′) is satisfied. Furthermore, $\mathcal{E}_\theta\phi_j(\theta)=0$ and $\mathcal{E}_\theta\phi_j^2(\theta)=\lambda_j/\theta=\mathcal{E}_\theta\psi_j(\theta)$, so that (A5) is true. Assumption (A6) also holds on account of (6.2). Next, it is easily seen that

$$\mathcal{E}_\theta\left|\frac{X_j-\lambda_j\theta}{\theta}\right|^3\leqq\frac{1}{\theta^2}(8\lambda_j^3\theta^2+6\lambda_j^2\theta+\lambda_j)\ ,$$

so that, by means of (6.2), (A7) is satisfied with $\delta=1$.

Finally,

$$\mathcal{E}_\theta\left(\frac{X_j-\lambda_j\theta}{\theta^2}\right)^2=\frac{\lambda_j}{\theta^3}\ ,$$

so that, by means of (6.2), (A8) holds true with $\delta=1$.

We wish to thank the referee for a number of suggestions which helped improve a previous version of this paper.

UNIVERSITY OF TEXAS
UNIVERSITY OF WISCONSIN AND UNIVERSITY OF PATRAS

REFERENCES

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

[2] Bradley, R. A. and Gart, J. J. (1962). The asymptotic properties of ML estimators when sampling from associated populations, *Biometrika*, **49**, 205-214.

[3] Cramér, H. (1946). *Mathematical Methods of Statistics*, Princeton Univ. Press, Princeton.

[4] Gurland, J. (1954). On regularity conditions for maximum likelihood estimation, *Skand. Aktuartidskr.*, **37**, 71-76.

[5] Hoadley, B. (1971). Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case, *Ann. Math. Statist.*, **42**, 1977-1991.

[6] Huber, P. J. (1967). The behavior of maximum likelihood estimates under non-standard conditions, *Proc. Fifth Berkeley Symp. Math. Statist. and Prob.* I, 221-233, University of California Press, Berkeley.

[7] Kulldorff, G. (1957). On the conditions for consistency and asymptotic efficiency of maximum likelihood estimates, *Skand. Aktuartidskr.*, **40**, 129-144.

[8] Loève, M. (1963). *Probability Theory*, Van Nostrand, New Jersey.

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

[10] Silvey, S. D. (1961). A note on maximum likelihood in the case of dependent random variables, *J. Roy. Statist. Soc.* B**23**, 444-452.