YOKOHAMA MATHEMATICAL
JOURNAL VOL. 31, 1983

MAXIMUM PROBABILITY ESTIMATORS BASED ON ABSOLUTELY REGULAR PROCESSES

By

KEN-ICHI YOSHIHARA

(Received November 6, 1982)

I. Introduction.

Let $\{X_i\}$ be a sequence of random variables. For each positive integer n, let $f(x_1^n:\theta)$ be the density, with respect to a σ-finite measure ν^n, of $X_1^n=(X_1, \cdots, X_n)$ at the point x_1^n (of the appropriate space) when θ is the value of the (unknown to the statistician) parameter. The latter is a point of the known open set Θ which is a subset of the real line. An estimator (of θ) is a Borel measurable function of X_1^n with values in Θ.

For each n let $k(n)$ be a normalizing factor for the family $f(x_1^n:\cdot)$. Let R be a bounded, Borel measurable subsets of the real line. A maximum probability estimator (with respect to R) is one which maximizes, with respect to d,

$$\int f(x_1^n:\theta)d\theta ,$$

the integral being over the set $\left\{d-\dfrac{R}{k(n)}\right\}$. For simplicity we assume that there is a unique maximum.

Let $h>0$ be any number. We shall say that a sequence $\{\theta_n\}$ is in $H(h)$ (for a special point which below will always be θ_{00}), if $|k(n)(\theta_n-\theta_{00})|\leqq h$ for $n=1, 2, \cdots$. The symbols P_{θ}, and E_{θ}, are to denote, respectively, the probability (of an event) and expected value when θ' is the "true" value of the parameter.

The following theorem was proved by Weiss and Wolfowitz (1967).

Theorem. *Let M_n be a maximum probability estimator with respect to R such that:*

For any $h>0$ and any sequence $\{\theta_n\}$ in $H(h)$ we have

$$\lim_{n\to\infty} P_{\theta_n}(k(n)(M_n-\theta_n)\in R)=\rho(\theta_{00}),\ say. \tag{1.1}$$

Let ε and δ be arbitrary but positive. For h sufficiently large we have, for any sequence $\{\theta_n\}$ in $H(h)$

$$\liminf_{n\to\infty} P_{\theta_n}(|k(n)(M_n-\theta_n)|<\delta h)\geqq 1-\varepsilon \tag{1.2}$$

Let T_n be any (competing) estimators such that for any $h>0$ and any sequence $\{\theta_n\}$ in $H(h)$ we have

$$\lim |P_{\theta_n}(k(n)(T_n-\theta_n)\in R)-P_{\theta_{00}}(k(n)(T_n-\theta_{00})\in R)|=0 \tag{1.3}$$

Then

$$\limsup_{n\to\infty} P_{\theta_{00}}(k(n)(T_n-\theta_{00})\in R)\leqq\rho(\theta_{00})\,. \tag{1.4}$$

In a series of interesting papers of which [2] is the first, Ibragimov and Khas'minskii have developed a new approach to the asymptotic distribution theory of statistical estimators based on independent observations and in [4], Wolfowitz proved that if the conditions of Ibragimov and Khas'minskii [2] with two additional ones hold, then (1.1) and (1.2) are satisfied, and so the maxmum probability estimator M_n is asymptotically efficient in the sense of (1.4).

In this paper, we prove that under the conditions of Yoshihara [5] the analogous result to Wolfowitz's one in [4] holds for the maximum probability estimators constructed by a sequence $\{\xi_n\}$ of observations which satisfies some absolute regularity condition.

2. Regularity conditions.

Let Θ be an open set in the real line. Suppose that a family of probability spaces $\{(\Omega, \mathcal{F}, P_\theta), \theta\in\Theta\}$ is given. Let $\{\xi_n, -\infty<n<\infty\}$ be a strictly stationary sequence of random variables with values in the measurable space $(X, \mathcal{A})$. Let $\xi_m^n=(\xi_m, \cdots, \xi_n)$ denote a random element in the direct product $X^{n-m+1}=\times_{i=m}^{n} X_i(X_i=X,\ i=m, \cdots, n)$, let $\mathcal{F}_m^n$ denote the smallest σ-algebra generated by ξ_m^n, and let P_θ^n be the projection of the measure P_θ on $\mathcal{F}_1^n$.

We assume that the sequence $\{\xi_n\}$ satisfies an absolute regularity (a.r.) condition

$$\beta_n=\beta(n)=\sup_{\theta\in\Theta}\beta(n, \theta)\downarrow 0 \tag{2.1}$$

as $n\to\infty$ where for each $\theta\in\Theta$

$$\beta(n, \theta)=E_\theta(\sup_{A\in\mathcal{F}_n^\infty}|P_\theta(a\,|\,\mathcal{F}_{-\infty}^0)-P_\theta(A)|) \tag{2.2}$$

and $E_\theta(\cdot)$ denotes the expectation with respect to $P_\theta(\cdot)$.

Let $(X, \mathcal{A}, \nu)$ be a measurable space with σ-finite measure ν. Put $\mathcal{A}^n=\times_{i=1}^{n}\mathcal{A}_i(\mathcal{A}_i=\mathcal{A},\ i=1, \cdots, n)(n\geqq1)$. For each $\theta\in\Theta$ let P_θ and $P_\theta^n(n\geqq1)$ be the probability measures defined respectively by

$$P_\theta^n(A)=P_\theta(\xi_1^n\in A) \qquad (A\in\mathcal{A}^n)$$

and

$$P_\theta(A)=P_\theta(\xi_{-\infty}^{\infty}\in A) \qquad (A\in\mathcal{A}^\infty).$$

We assume that the measure P_θ^n is absolutely continuous with respect to the product measure $\nu^n=\nu\times\cdots\times\nu$ and defines the probability density

(2.3) $$f(x_1^n,\ \theta)=\frac{dP_\theta^n}{d\nu^n}(x_1^n)$$

which is $\mathcal{A}^n\times\mathcal{B}$-measurable. Here, $x_q^p\in X^{p-q+1}$ and $\mathcal{B}$ is the smallest σ-algebra of all Borel subsets of the real line. For $n\geqq p\geqq 1$ and $\theta\in\Theta$ let $f(x_p^n|x_1^{p-1},\ \theta)$ be the conditional probability density function or the probability density function, i. e.,

(2.4) $$P_\theta(\xi_1^{p-1}\in B,\ \xi_p^n\in A)$$
$$=\int_B f(x_1^{p-1},\ \theta)\nu^{p-1}(dx_1^{p-1})\int_A f(x_p^n|x_1^{p-1},\ \theta)\nu^{n-p+1}(dx_p^n)$$
$$(B\in\mathcal{A}^{p-1}\quad\text{and}\quad A\in\mathcal{A}^{n-p+1})$$

where $f(x_1^n|x_1^0,\ \theta)=f(x_1^n,\ \theta)$ denotes the probability density function.

We postulate the following regularity conditions which are those of Yoshihara [5].

Conditions of group. I. I_1. The parameter set Θ is an open interval (bounded or unbounded) of the real line.

I_2. For any n the functions $f(x_n|x_1^{n-1},\ \theta)$ are defined for all $x_1^n\in X^n$ and $\Theta\in\theta$ and are $\mathcal{A}^n\times\mathcal{B}$-measurable.

I_3. If $\theta\neq\theta'$, then $P_\theta\neq P_{\theta'}$. More precisely, if $\theta\neq\theta'$, then for all n and $p(n\geqq p)$ and for ν^{p-1}-almost all x_1^{p-1}

(2.5) $$\int_{X^{n-p+1}}|f(x_p^n|x_1^{p-1},\ \theta)-f(x_p^n|x_1^{p-1},\ \theta')|\nu^{n-p+1}(dx_p^n)>0$$

Whenever the integrations with respect to x_p^n are over all of X^{n-p+1} we shall agree to omit the region of integration and to write $\nu(dx_p^n)$ instead of $\nu^{n-p+1}(dx_p^n)$.

Conditions of group II. II_1. For any $n(n\geqq 1)$ and for any fixed x_1^n, the function $f(x_n|x_1^{n-1},\ \theta)$ is defined and continuously twice differentiable in the closure Θ^c of Θ.

We put

(2.6) $$f'(x_p^n|x_1^{p-1},\ \theta)=\frac{\partial}{\partial\theta}f(x_p^n|x_1^{p-1},\ \theta).$$

II_2. For all $n(\geqq 1)$ and for ν^{n-1}-almost all x_1^{n-1}

(2.7) $$\int f'(x_n|x_1^{n-1},\ \theta)\nu(dx_n)=\frac{\partial}{\partial\theta}\int f(x_n|x_1^{n-1},\ \theta)\nu(dx_n)=0$$

For any n and $k(n\geqq k)$, let

$$(2.8)\qquad U_{n,k}(\theta)=\begin{cases}\dfrac{f'(\xi_n|\xi_k^{n-1},\theta)}{f(\xi_n|\xi_k^{n-1},\theta)}, & \text{if } f(\xi_n|\xi_k^{n-1},\theta)>0\\ 0\qquad, & \text{otherwise}.\end{cases}$$

II$_3$. There exists a number $\delta(>0)$ such that

$$(2.9)\qquad \sup_{\theta\in\Theta}\sup_{n\geq 1} E_\theta|U_{n,1}(\theta)|^{4+\delta}<\infty$$

and

$$(2.10)\qquad \sup_{\theta\in\Theta}\sup_{n\geq 1} E_\theta\left|\frac{\partial U_{n,1}(\theta)}{\partial\theta}\right|^{2+\delta}<\infty.$$

For each $n(\geqq 1)$ and for each $\theta\in\Theta$, put

$$(2.11)\qquad I(\xi_n|\xi_1^{n-1},\theta)=E_\theta|U_{n,1}(\theta)|^2$$

and for each $\theta\in\Theta$

$$(2.12)\qquad I(\theta)=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}g_i(\theta)$$

where

$$(2.13)\qquad g_i(\theta)=I(\xi_i\|\xi_1^{i-1},\theta)+2\sum_{j=i+1}^{\infty}E_\theta\{U_{i,1}(\theta)U_{j,1}(\theta)\}\qquad (i=1,2,\cdots)$$

(The existence of limits in the right-hand side of (2.12) and (2.13), respectively, are verified in Propositions 3.2 and 3.3 (below)).

II$_4$. $I(\theta)$ is a positive and continuous function of $\theta\in\Theta^c$.

II$_5$. There exists a number $d(\geqq 0)$ such that

$$(2.14)\qquad \sup_{\theta\in\Theta}(I+|\theta|)^{-d}I(\theta)<\infty.$$

Conditions of group III. III$_1$. The sequence $\{\beta(n)\}$ defined by (2.1) satisfy the condition

$$(2.15)\qquad \beta(n)=O(e^{-\lambda_1 n})\qquad (n\to\infty) \text{ for some } \lambda_1>0.$$

III$_2$. For any $c>0$ and $k(\geqq 1)$, let

$$(2.16)\qquad \gamma(k,\theta,c)=\sup_{|\theta'-\theta|\leq c}\sup_{n\geq k}\{E_\theta|U_{n,1}(\theta')-U_{n,n-k}(\theta')|^4\}^{1/4}$$

There exists a positiwe number c_0 such that

$$(2.17)\qquad \gamma_k=\gamma(k)=\sup\gamma(k,\theta,c_0)=O(e^{-\lambda_2 n})\qquad (n\to\infty) \text{ for some } \lambda_2>0.$$

as $k\to\infty$.

Conditions of group IV. IV$_1$. There exists a positive number d_1 such that for all θ and θ' in Θ

$$\sup_{n\geq 1} \operatorname{ess\cdot sup}_{x_1^{n-1}} |\theta-\theta'|^{d_1} \int \sqrt{f(x_n | x_1^{n-1}, \theta) f(x_n | x_1^{n-1}, \theta')} \nu(dx_n) < \infty \tag{2.18}$$

IV$_2$. Let $\{a_m\}$ be any sequence with limit a, say. For any $h>0$ and any sequence $\{\theta_m\}$ in $H(h)$. we have

$$\sup_{n\geq 1} \Big| \int \sqrt{f(x_n | x_1^{n-1}, \theta_m) f(x_n x_1^{n-1}, \theta_m + a_m)} \nu(dx_n) - \int \sqrt{f(x_n | x_1^{n-1}, \theta_{00}) f(x_n | x_1^{n-1}, \theta_{00}+a)} \nu(dx_n) \Big| \longrightarrow 0 \tag{2.19}$$

Remark 2.1. If Conditions I-III are satisfied, then it is obvious that for any fixed p, q and $\theta \in \Theta$ $\{f'(\xi_{n+1}^{n+p} | \xi_{n-q}^{n}, \theta)/f(\xi_{n+1}^{n+p} | \xi_{n-q}^{n}, \theta)\}$, $n=0, \pm 1, \pm 2, \cdots$ is a strictly stationary sequence satisfying the a. r. condition with $\beta^*(n)=1$ if $n \leqq p+q$ and $=\beta(n-p-q)$ if $n>p+q$.

Remark 2.2. It is obvious that the condition $\gamma(n)=O(e^{-\lambda_2 n})$ is satisfied if $\{\xi_n\}$ possesses the r-th order Markov property, where $r \geqq 1$, i. e., for any $\theta \in \Theta$

$$P_\theta(\xi_{n+1} \in A | \xi_0, \cdots, \xi_n) = P_\theta(\xi_{n+1} \in A | \xi_{n-r+1}, \cdots, \xi_n) \qquad (n \geqq r-1).$$

In what follows, K(with or without subscript) will stand for a quantity not depending on the parameters occurring in the discussion and the same letter K will be used to denote different constants even within the same formula.

3. Examination of the proofs of [5].

In this section, we always suppose that Conditions I-IV hold.

For any $\{\theta_n\}$ in $H(h)$ we define

$$Y_n(\theta) = \log \frac{f(\xi_1^n, \theta_n + \theta n^{-1/2})}{f(\xi_1^n, \theta_n)} \tag{3.1}$$

For any positive integer n, let $p=p(n)=[n^\alpha]$ (α being a number such that $1/2 < \alpha < 1$), $q=q(n)=[n^{1/3}]$, $l=p+q$ and $k=[n/l]$, where as usual $[s]$ denotes the integer part of s. Further, let $\varepsilon=\varepsilon_n=bn^{-1/2}$ for some $b(>0)$. Let $\{\theta_n+\varepsilon_n\}$ be in $H(h)$. For any $\tau>0$, let

$$A_{n\tau}^* = \left\{ x_1^p : \left| \log \frac{f(x_1^p, \theta_n+\varepsilon_n)}{f(x_1^p, \theta_n)} \right| > \tau \right\} \tag{3.2}$$

(here we agree to set $0/0=1$). Further, let

$$a_m(s) = \frac{f'(x_1^m, s)}{\sqrt{f(x_1^m, s)}}. \tag{3.3}$$

The following lemma corresponds Conditions II$_5$ in Ibragimov and Khas'minskii

[2] or Lemma 4.2 in Yoshihara [5].

Lemma 3.1. *For any* $\tau>0$

$$\lim_{n\to\infty}\frac{1}{p\varepsilon}\int_{\theta_n}^{\theta_n+\varepsilon}\int_{A_{n\tau}^*}a_p^2(s)\nu(dx_1^p)ds=0 \tag{3.4}$$

Proof. By the proof of Proposition 3.3 in [5] and IV_2

$$I(\theta)=\lim_{n\to\infty}\frac{1}{n}E_\theta\left|\frac{f'(\xi_1^n,\theta)}{f(\xi_1^n,\theta)}\right|^2$$

holds uniformly in θ. So noting Condition IV_2 and applying the method of Lemma 4.2 in [5] we have (3.4), which is desired.

Lemma 3.2. *For any* $\tau>0$

$$\sup_{|\theta-\theta_n|<\varepsilon_n}\int_{A_{n\tau}^*}f(x_1^p,\theta)\nu(dx_1^p)=o(k^{-1})\qquad(n\to\infty)\cdot \tag{3.5}$$

The proof is obtained by the same method as that of Lemma 4.3 in [5] (cf. the proof of Lemma 2.1 in [2]).

Lemma 3.3. *Suppose Conditions* I-IV *hold. Then, as* $n\to\infty$

$$\int\sqrt{f(x_1^p,\theta_n+\varepsilon_n)f(x_1^p,\theta_n)}\nu(dx_1^p)=1-\frac{1}{8}I_{00}k^{-1}+o(k^{-1}) \tag{3.6}$$

and

$$\int\sqrt{f(x_1^p\,|\,x_{-q}^0,\theta_n+\varepsilon_n)f(x_1^p\,|\,x_{-q}^0,\theta_n)f(x_{-q}^0,\theta_n)}\nu(dx_{-q}^p)=1-\frac{1}{8}I_{00}k^{-1}+o(k^{-1}) \tag{3.7}$$

for any $\{\theta_n\}$ *in* $H(h)$.

Lemma 3.4. *For any positive* τ

$$\int_{\bar{A}_{n\tau}^*}\log\frac{f(x_1^p,\theta_n+\varepsilon_n)}{f(x_1^p,\theta_n)}f(x_1^p,\theta_n)\nu(dx_1^p)=-\frac{1}{2}k^{-1}I_{00}+o(k^{-1}) \tag{3.8}$$

and

$$\int_{\bar{A}_{n\tau}^*}\log^2\frac{f(x_1^p,\theta_n+\varepsilon_n)}{f(x_1^p,\theta_n)}f(x_1^p,\theta_n)\nu(dx_1^p)=k^{-1}I_{00}+o(k^{-1}) \tag{3.9}$$

as $n\to\infty$. *Here* $\bar{B}$ *denotes the complementary set of* B.

The proofs of Lemmas 3.3 and 3.4 are the same as those of Lemmas 4.6 and 4.7 and so are omitted.

Now, by Lemmas 3.2, 3.3 and the method used in the proof of Theorem 4.1 in [5], we have the following theorem.

Theorem 1. *Suppose Conditions* I-IV *hold. Then as* $n\to\infty$ *the finite dimensional distributions of the stochastic process* $Y_n(\theta)$ *converge to the finite-dimensional distributions of the process*

(3.10) $$Y(\cdot): Y(\theta)=\theta I_{00}\zeta-\frac{\theta^2}{2} I_{00}$$

where ζ is normally distributed with mean zero and variance one.

From Theorem 4.2 in [5] the following theorem easily follows.

Theorem 2. *If Conditions* I-IV *hold, then there exist two positive numbers b_0 and K_0 such that for any pair (θ_1, θ_2) $(\sqrt{n}|\theta_i-\theta_{00}|\leqq b_0,\ i=1, 2)$*

(3.11) $$E_{\theta_n}|Z_n^{1/2}(\theta_2)-Z_n^{1/2}(\theta_1)|\leqq K_0(\theta_2-\theta_1)^2 .$$

By Lemma 3.3 and the same method as the proof of Lemma 4.9 in [5], we have the following lemma.

Lemma 3.5. *For any K_1 and N there is a positive c_N such that in the region $|\theta|<K_1 n^{1/2}$*

(3.12) $$P_{\theta_n}\left(Z_n(\theta)>\frac{1}{|\theta|^N}\right)\leqq\frac{c_N}{|\theta|^N}$$

holds for all n sufficiently large.

By IV_2 and the proof of Lemma 4.10 we have the following lemma (cf. the proof of Lemma 2.7′ in [4]).

Lemma 3.6. *For any positive N, there are numbers n_0 and K_2 such that*

(3.13) $$P_{\theta_n}\left(Z_n(\theta)>\frac{1}{|\theta|^N}\right)\leqq\frac{c_N}{|\theta|^N}$$

for all $n>n_0$ and all $|\theta|>K_2 n^{1/2}$.

From Lemma 3.5 and 3.6 we have the following theorem.

Theorem 3. *Whatever be $N>0$, there exist an n_0 and a constant c_N depending only on N such that for $n>n_0$*

(3.14) $$P_{\theta_{00}}\left(\sup_{|\theta|>A} Z_n(\theta)>\frac{1}{A^N}\right)\leqq\frac{c_N}{A^N}$$

(3.15) $$P_{\theta_{00}}\left(\sup_{r\leq|\theta|\leq r+1} Z_n(\theta)\geqq\frac{1}{r^N}\right)\leqq\frac{c_N}{r^N} \qquad (r\geqq 1) .$$

Finally, let $C_0=C_0(-\infty, \infty)$ be the space of functions which are continuous on $(-\infty, \infty)$ and for which $\lim_{x\to\infty} f(x)=0$ endowed with the usual uniform metric, if $\Theta=(-\infty, \infty)$, put $Z_n(\theta)=Z_n(\theta)$ and if $\Theta=(a, b)\neq(-\infty, \infty)$, define a process $Z_n(\theta)$ as follows:

$$Z_n(\theta)=\begin{cases} Z_n(\theta) \text{ if } \sqrt{n}(a-\theta_{00})<\theta<n(b-\theta_{00}), \\ 0 \text{ if } \theta\leqq\sqrt{n}(a-\theta_{00})-1 \text{ or } \theta\geqq\sqrt{n}(b-\theta_{00})+1, \\ \text{linear and continuous in all other intervals.} \end{cases}$$

Then, by Theorems 1 and 2 we have the following theorem which corresponds to Theorem 4.4 in [5] (cf. Theorem 2.5 in [1]).

Theorem 4. *Suppose that Conditions* I-IV*hold. Then the distributions in* C_0 *generated by the process* Z_n *converge as* $n\to\infty$ *to the distribution generated by* Z. *In particular, if* h *is a continuous functional on* C_0, *thaen for all* x

$$\lim P_{\theta n}(h(Z_n)<x)=P_{\theta_{00}}(h(Z)<x). \tag{3.16}$$

4. Efficiency of estimators.

Let $R=(-r, r)$ and suppose Conditions I-IV hold. Exactly as in the argument of [4] we obtain from Theorem 3 that for any $h>0$ and any $\{\theta_n\}$ in $H(h)$

$$\lim_{n\to\infty} P_{\theta_n}(\sqrt{n}(M_n-\theta_n)<y)=\sqrt{\frac{I_{00}}{2\pi}}\int_{-\infty}^{y}\exp\left(-\frac{1}{2}I_{00}x^2\right)dx \tag{4.1}$$

for every y. So, (1.1) and (1.2) are easily obtained from (4.1). Hence, the maximum probability estimator M_n is asymptotically efficient in the sense of (1.4) for $R=(-r, r)$.

Now, let $\hat{\theta}_n$ be the maximum likelihood estimator. Then, by the same argument as that in Section 4 in [4] we can show that $\hat{\theta}_n$ is asymptotically efficient in the sense of (1.4) for $R=(-r, r)$.

References

[1] Dancunha-Castelle, D. (1978) : *Vittesse de convergence pour certains problêms statistiques*. Springer-Verlag, Berlin-Heiderberg-New York, Lecture Notes Mathematics. No. 678, 3-175.

[2] Ibragimov, I. A. and Khas'minskii, R. Z. (1975) : *Asymptotic behavior of statistical estimators in the smooth case I. Study of the likelihood ratio.* Theory Probab. Appl. 20 445-462.

[3] Weiss, L. and Wolfowitz, J. (1969) : *Maximum probability estimators with a general loss function.* Springer-Verlag, Berlin-Heidelberg New York, Lecture Notes in Mathematics, No. 89.

[4] Wolfowitz, J. (1975) : *Maximum probability estimators in the classical case and in the "almost smooth" case.* Theory Probab. Appl. 20 363-371.

[5] Yoshihara, K. (1982) : *Asymptotic behavior of statistical estimators constructed on absolutely regular sequences.* Yokohama Math. J. 30 121-139.

Yokohama National University
Department of Mathematics
Faculty of Engineering
156, Tokiwadai, Hodogaya,
Yokohama, Japan