CLASSICAL ASYMPTOTIC PROPERTIES OF A CERTAIN ESTIMATOR RELATED TO THE MAXIMUM LIKELIHOOD ESTIMATOR*

JAMES C. FU AND LEON JAY GLESER

(Received Oct. 14, 1972; revised April 11, 1973)

1. A general method for constructing point estimators

In a statistical point estimation problem, the goal is to use information obtained from a sample of observations to estimate an unknown parameter (or a function of an unknown parameter) of the probability distribution which governs the variability of sampling. Ideally, we would like to construct an estimator which has the property that with probability one a correct estimate is made of the true value θ of the unknown parameter. This goal is, of course, met only in trivial cases. More realistically, we hope to find an estimator having highest possible probability of being "close" to θ.

Suppose that we have independent and identically distributed (i.i.d.) random observations $X_1, X_2, \cdots, X_n$ having common distribution P_θ. Here we assume that each X_i is defined on a measure space $(\mathcal{X}, \beta)$, where $\mathcal{X}$ is any topological space and β is a sigma-field of measurable sets. We assume that P_θ is a member of a class $\{P_\theta,\ \theta \in \Theta\}$ indexed by a point θ in a subset Θ of k-dimensional Euclidean space. Further, we assume that the class $\{P_\theta,\ \theta \in \Theta\}$ is dominated by a σ-finite measure μ defined on β, so that each P_θ has a density (Radon-Nikodym derivative) $f(x|\theta)=dP_\theta/d\mu$ with respect to μ. The sample $(X_1, X_2, \cdots, X_n)$ is then defined on the Cartesian product space $(\mathcal{X}^{(n)}, \beta^{(n)})$ with respect to the product probability measure $P_\theta^{(n)}$ which has density $\prod_{i=1}^{n} f(x_i|\theta)$ with respect to the product measure $\mu^{(n)}$.

With this probability background in mind, we define an estimator T_n of θ based on $X_1, X_2, \cdots, X_n$ to be a measurable function mapping $\mathcal{X}^{(n)}$ into Θ. Since Θ is a subset of Euclidean k-dimensional space, we can measure distance by the usual Euclidean distance $|\theta-\theta'|$. One way of defining what we mean when we say that $T_n(x_1, x_2, \cdots, x_n)$ is "close"

* The work on this article was sponsored in part by the Air Force Office of Scientific Research, Air Force Systems Command, U.S.A.F. under AFOSR Contract F44620-70-C-0066. The United States Government is authorized to reproduce and distribute reprints for government purposes notwithstanding any copyright hereon.

to θ is to choose a small $\varepsilon>0$, and to say that $T_n(x_1, x_2, \cdots, x_n)$ is "close" to θ if $|T_n(x_1, x_2, \cdots, x_n)-\theta|<\varepsilon$. The probability $\alpha(T_n, \varepsilon, \theta)$ that T_n is "close" to θ is then

$$\alpha(T_n, \varepsilon, \theta)\equiv P_\theta\{|T_n(X_1, X_2, \cdots, X_n)-\theta|<\varepsilon\} . \tag{1.1}$$

Let τ_n be the class of all estimators $T_n: \mathscr{X}^{(n)}\to\Theta$. The estimator T_n^* has highest probability of being "close" to θ if

$$\alpha(T_n^*, \varepsilon, \theta)=\sup_{T_n\in\tau_n} \alpha(T_n, \varepsilon, \theta) . \tag{1.2}$$

We could define T_n^* to be a "best" estimator of θ if (1.2) holds for all $\theta\in\Theta$. Unfortunately, this definition of "best" (although clearly reasonable and meaningful) has the major disadvantage of almost never being satisfied in practical statistical problems. The estimator T_n^* satisfying (1.2) for a given ε and θ, may not satisfy (1.2) for other values of ε and θ; in other words, the optimal T_n^* under the criterion (1.2) depends on ε and θ.

To remove (or account for) the influence of θ in measuring goodness of estimators by means of the quantity $\alpha(T_n, \varepsilon, \theta)$ in (1.1), we could consider weighting different values of θ by means of any of a class of measures defined on the Borel subsets of Θ. Assume that we have such a positive measure G which has a Radon-Nikodym derivative $g(\theta)=dG/d\nu$ with respect to some standard σ-finite measure $\nu(\theta)$ (usually Lebesgue measure or counting measure). With respect to G, we might consider replacing the θ-specific measure (1.1) by the weighted probability

$$\alpha(T_n, \varepsilon, g)=\int_\Theta \alpha(T_n, \varepsilon, \theta)g(\theta)d\nu(\theta) . \tag{1.3}$$

We can then define a *maximum probability estimator with respect to* g *and* ε to be an estimator T_n^* which maximizes (1.3) over all $T_n\in\tau_n$. If

$$\int_\Theta g(\theta)d\nu(\theta)<\infty , \tag{1.4}$$

then a maximum probability estimator with respect to g and ε is a Bayes estimator with respect to $g(\theta)\Big/\int_\Theta g(\theta)d\nu(\theta)$ under the loss function

$$L_\varepsilon(\theta, a)=\begin{cases} 1 , & \text{if } |\theta-a|\geqq\varepsilon , \\ 0 , & \text{if } |\theta-a|<\varepsilon , \end{cases} \tag{1.5}$$

where $a\in\Theta$ is the action to estimate θ by a. When Θ is one-dimensional, such a Bayes estimator $T_n^*=T_n^*(X_1, X_2, \cdots, X_n)$ can be shown to satisfy

$$(1.6)\qquad \int_{T_n^*-\varepsilon}^{T_n^*+\varepsilon} \prod_{i=1}^{n} f(x_i|\theta)g(\theta)d\nu(\theta) = \sup_{\substack{T_n \in \tau_n \\ T_n = T_n(x_1, x_2, \cdots, x_n)}} \int_{T_n-\varepsilon}^{T_n+\varepsilon} \prod_{i=1}^{n} f(x_i|\theta)g(\theta)d\nu(\theta)$$

for all $(x_1, x_2, \cdots, x_n) \in \mathscr{X}^{(n)}$. Even if (1.4) does not hold, if

$$(1.7)\qquad \int_{\mathscr{X}^{(n)}} \int_{\Theta} L_\varepsilon(\theta, T_n(x_1, x_2, \cdots, x_n)) \cdot \prod_{i=1}^{n} f(x_i|\theta)g(\theta)d\nu(\theta)d\mu^{(n)}(x_1, x_2, \cdots, x_n)$$

is finite for some $T_n \in \tau_n$, and Θ is one-dimensional, then the maximum probability estimator T_n^* with respect to g and ε satisfies (1.6). In the special case where $G(\theta)$ and $\nu(\theta)$ are Lebesgue measure on one-dimensional Euclidean space (so that $g(\theta)\equiv 1$) and (1.7) is finite for some $T_n \in \tau_n$, we can see some similarity between the maximum probability estimator with respect to $g(\theta)\equiv 1$ and ε, the maximum likelihood estimator, and a somewhat different "maximum probability estimator" defined by Weiss and Wolfowitz [15].

The maximum probability estimator T_n^* with respect to a $g(\theta)$ and a given $\varepsilon>0$ can usually be shown to exist, and is an intuitively meaningful estimator for many statistical problems. In the case when $g(\theta)$ satisfies (1.4) and Θ is one-dimensional, T_n^* can be seen to be the midpoint of a modal interval of length 2ε for the posterior distribution

$$(1.8)\qquad h(\theta|x_1, x_2, \cdots, x_n) = \frac{g(\theta)\prod_{i=1}^{n} f(x_i|\theta)}{\int_{\Theta} g(\theta)\prod_{i=1}^{n} f(x_i|\theta)d\nu(\theta)} .$$

Although this property of the maximum probability estimator T_n^* with respect to g and ε is appealing (and of theoretical importance since it relates point estimation to Bayesian fixed-width confidence intervals), calculation of the estimator is not always a simple task and the estimator may not have a convenient closed mathematical form. Further, the estimator T_n^* depends upon the constant $\varepsilon>0$ (and upon the measurement of distance used). Since which value of $\varepsilon>0$ to use is not always clear in practical statistical problems, we would like to somehow avoid an estimation procedure in which an explicit choice of ε must be made. Presumably $\varepsilon>0$ would be a small constant in most cases, so it is reasonable to consider the limit of maximum probability estimators $T_n^*(\varepsilon)$ with respect to g and ε as ε tends to 0. If $T_n^*(\varepsilon)$ is unique for each $\varepsilon>0$ and if $T_n^*(\varepsilon)$ satisfies (1.6), then $T_n^*(\varepsilon)\to\theta_n^*$ as $\varepsilon\to 0$, where $\theta_n^*=\theta_n^*(x_1, x_2, \cdots, x_n)$ and

$$\prod_{i=1}^{n} f(x_i \mid \theta_n^*) g(\theta_n^*) = \sup_{\theta \in \Theta} \prod_{i=1}^{n} f(x_i \mid \theta) g(\theta) , \tag{1.9}$$

all $(x_1, x_2, \cdots, x_n) \in \mathcal{X}^{(n)}$. When $g(\theta)$ satisfies (1.4), $\theta_n^*(x_1, x_2, \cdots, x_n)$ is a mode of the posterior density $h(\theta \mid x_1, x_2, \cdots, x_n)$ defined in (1.8). We call $\theta_n^*(X_1, X_2, \cdots, X_n)$ satisfying (1.9) a *maximum probability estimator* with respect to $g(\theta)$. In many cases, θ_n^* is as easy or easier to compute than the maximum likelihood estimator $\hat{\theta}_n$. Of course, the maximum likelihood estimator is a maximum probability estimator with respect to $g(\theta) \equiv 1$, all θ.

It is important to note that the estimator θ_n^* is not the same as the maximum probability estimator of Weiss and Wolfowitz [15]. Their estimator is formed from $T_n^*(\varepsilon)$ by choosing a particular useful sequence of ε-values $\{\varepsilon_n\}$, $\varepsilon_n \to 0$, chosen to give certain asymptotic properties to the resulting sequence of estimators $T_n^*(\varepsilon_n)$. Our estimator θ_n^* is formed (in many cases) for *fixed n* by taking $\lim_{\varepsilon \to 0} T_n^*(\varepsilon)$, or can be defined directly from (1.9).

In this paper, we consider two classical asymptotic (as $n \to \infty$) properties of the maximum probability estimator θ_n^* with respect to g. In Section 2, we find regularity conditions under which θ_n^* exists and is measurable $(\mathcal{X}^{(n)}, \beta^{(n)})$. We also prove, under additional regularity conditions, that θ_n^* is a strongly consistent estimator of $\theta (\theta_n^* \xrightarrow{\text{a.s.}} \theta)$. In Section 3, we show (again under additional regularity conditions and under the assumption that Θ is a subset of the real line) that $\sqrt{n}(\theta_n^* - \theta) \xrightarrow{\mathcal{L}} N(0, 1/I(\theta))$, where $I(\theta)$ is Fisher's information, that $\sqrt{n}(\theta_n^* - \hat{\theta}_n) \xrightarrow{p} 0$, and that $n^{1/2-\rho}(\theta_n^* - \hat{\theta}_n) \xrightarrow{\text{a.s.}} 0$, all ρ, $0 < \rho \leq 1/2$. This shows that θ_n^* is a best asymptotic normal estimator in Fisher's classical sense. As a result of the asymptotic convergence and asymptotic efficiency properties of maximum probability estimators shown in Sections 2 and 3, and of the approximate Bayesian character of maximum probability estimators mentioned in the present section, we recommend that such estimators be considered and used in practical statistical problems. For large enough n, the choice of the weighting function $g(\theta)$ is irrelevant (provided $g(\theta)$ is positive everywhere and smooth enough—see Section 3). For small n, choice of $g(\theta)$ will influence the resulting estimates. In this paper, we do not intend to discuss choice of $g(\theta)$, since we feel that reasonable choices depend both upon the problem and upon prior judgements by the statistician.

2. Existence and consistency

In this section, we prove that the M.P.E. θ_n^* with respect to a prior density $g(\theta)$ exists and is a strongly consistent estimator of θ.

Our exposition of these results breaks naturally into two parts: (i) a proof of the existence and measurability of θ_n^* (Theorem 2.1), and (ii) a proof that $\{\theta_n^*\}$ is strongly consistent (Theorem 2.2).

We start by establishing sufficient conditions for the existence and measurability of θ_n^* for a given sample size $n \geqq 1$. We have assumed in Section 1 that we observe i.i.d. observations $X_1, X_2, \cdots, X_n$, and that each X_i takes values in the measure space $(\mathcal{X}, \beta)$ according to the probability measure P_θ, where θ is a point in a subset Θ of Euclidean k-dimensional space. Hence $(X_1, X_2, \cdots, X_n)$ takes values in the product measure space $(\mathcal{X}^{(n)}, \beta^{(n)})$ according to the product probability measure $P_\theta^{(n)}$; $\theta \in \Theta$. We write $\xi = (X_1, X_2, \cdots)$ and $s = (x_1, x_2, \cdots)$. The probability distribution of ξ in its sample space $(\mathcal{X}^{(\infty)}, \beta^{(\infty)})$ when θ obtains is denoted by $P_\theta^{(\infty)}$, but we shall usually abbreviate $P_\theta^{(\infty)}$ to P_θ.

Since Θ is a subset of Euclidean k-dimensional space, its closure $\bar{\Theta}$ can be represented as the union $\bigcup_{t=1}^{\infty} K_t$ of a countable number of compact subsets $\{K_t: t=1, 2, \cdots\}$, where for every $t=1, 2, \cdots$, $K_t \subseteq K_{t+1} \subseteq \bar{\Theta}$. Such a representation is by no means unique, but to prove that one such representation exists, note that the entire k-dimensional space $\mathcal{E}^{(k)}$ can be written as the union of the closed k-dimensional spheres $S_t = \{y: y \in \mathcal{E}^{(k)}, |y| \leqq t\}$, $t=1, 2, \cdots$. Let $K_t = \bar{\Theta} \cap S_t$. Then since S_t is compact and $\bar{\Theta}$ is closed, K_t is compact. Further, $\bar{\Theta} = \bigcup_{t=1}^{\infty} K_t$ and $K_t \subseteq K_{t+1}$.

In Section 1, it has been assumed that the collection $\{P_\theta, \theta \in \Theta\}$ of probability measures on $(\mathcal{X}, \beta)$ is dominated by the σ-finite measure μ, and that $dP_\theta/d\mu = f(x|\theta)$ is the Radon-Nikodym derivative of P_θ with respect to μ. Hence, the product measure $P_\theta^{(n)}$ has Radon-Nikodym derivative

$$L_n((x_1, x_2, \cdots, x_n)|\theta) = L_n(\boldsymbol{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta) \tag{2.1}$$

with respect to the product measure $\mu^{(n)}$ on $(\mathcal{X}^{(n)}, \beta^{(n)})$. We have also assumed that there is a non-negative σ-finite measure ν defined on Θ which has a Radon-Nikodym derivative $g(\theta) = d\nu/d\theta$ with respect to k-dimensional Lebesgue measure. Let

$$H_n(\boldsymbol{x}|\theta) = H_n((x_1, x_2, \cdots, x_n)|\theta) \equiv g(\theta) L_n(\boldsymbol{x}|\theta) . \tag{2.2}$$

We propose the following conditions:

CONDITION 2.1. For all $\boldsymbol{x} \in \mathcal{X}^{(n)}$ (except perhaps for a set of $\boldsymbol{x}$-values having $\mu^{(n)}$-measure equal to zero), $L_n(\boldsymbol{x}|\theta)$ is continuous in θ for all $\theta \in \Theta$, and can be extended to a function which is continuous over all $\theta \in \bar{\Theta}$. That is, for any sequence $\{\theta_j\} \subset \Theta$ such that $\lim_{j\to\infty} \theta_j = \theta_0$ exists,

$\lim_{j\to\infty} L_n(\boldsymbol{x}|\theta_j)$ exists and has a value dependent on the sequence $\{\theta_j\}$ only through its limit $\theta_0 \in \bar{\Theta}$. If $\theta_0 \in \Theta$, this limit is $L_n(\boldsymbol{x}|\theta_0)$. If $\theta_0 \in \bar{\Theta}\sim\Theta$, we denote the limit by $L_n(\boldsymbol{x}|\theta_0)$ for typographical convenience. Hence, $L_n(\boldsymbol{x}|\theta)$ in its extended definition is for all $\boldsymbol{x} \in \mathcal{X}^{(n)}$ now a continuous function of θ for all $\theta \in \bar{\Theta}$.

CONDITION 2.2. The prior density $g(\theta)$ is continuous in θ for all $\theta \in \Theta$ and can be extended to a continuous function of θ for all $\theta \in \bar{\Theta}$. Further, $g(\Theta) \equiv \sup_{\theta\in\Theta} g(\theta) < \infty$.

CONDITION 2.3. There exists an increasing sequence $\{K_t,\ t=1, 2, \cdots\}$ of compact subsets of $\bar{\Theta}$ such that $\bar{\Theta} = \bigcup_{t=1}^{\infty} K_t$ and $\lim_{t\to\infty} \sup_{\theta \in K_t^c \cap \bar{\Theta}} L_n(\boldsymbol{x}|\theta) = 0$ for all $\boldsymbol{x} \in \mathcal{X}^{(n)}$ (except perhaps for a set of $\boldsymbol{x}$-values having $\mu^{(n)}$-measure equal to zero).

CONDITION 2.4. For each $\theta \in \Theta$, $L_n(\boldsymbol{x}|\theta)$ is $\beta^{(n)}$-measurable (measurable with respect to $(\mathcal{X}^{(n)}, \beta^{(n)})$).*

Condition 2.4 implies that $L_n(\boldsymbol{x}|\theta)g(\theta) = H_n(\boldsymbol{x}|\theta)$ is $\beta^{(n)}$-measurable for each $\theta \in \Theta$. Since Θ is dense in $\bar{\Theta}$, it follows that for every $\theta \in \bar{\Theta}\sim\Theta$ there exists a sequence $\{\theta_j\} \subseteq \Theta$ such that $\lim_{j\to\infty} \theta_j = \theta$. However, Conditions 2.1 and 2.2 imply that $\lim_{j\to\infty} H_n(\boldsymbol{x}|\theta_j) = H_n(\boldsymbol{x}|\theta)$ for all $\boldsymbol{x} \in \mathcal{X}^{(n)}$, so that $H_n(\boldsymbol{x}|\theta)$ is the limit of $\beta^{(n)}$-measurable functions and therefore is $\beta^{(n)}$-measurable. Consequently, Conditions 2.1, 2.2 and 2.4 imply that $H_n(\boldsymbol{x}|\theta)$ is $\beta^{(n)}$-measurable for all $\theta \in \bar{\Theta}$. Conditions 2.1 and 2.2 also imply that for all $\boldsymbol{x} \in \mathcal{X}^{(n)}$, $H_n(\boldsymbol{x}|\theta)$ is continuous in θ for $\theta \in \bar{\Theta}$.

LEMMA 2.1. *Let S be a sample space with a sigma field Δ of subsets, and let Y be a compact subset of Euclidean k-dimensional space (a metric space under the usual Euclidean distance $|y-y'|$). Let Γ be the appropriate Borel sigma field of subsets of Y. Let $U(s, y)$ map $S \times Y$ into $[0, \infty)$. If for every $s \in S$, $U(s, \cdot)$ is a continuous function on Y, and if for every $y \in Y$, $U(\cdot, y)$ is a Δ-measurable function (measurable with respect to (S, Δ)), then*

(i) *$U(\cdot, \cdot)$ is a $\Delta \times \Gamma$-measurable function,*

(ii) *$\sup_{y\in Y} U(s, y)$ exists and is a Δ-measurable function,*

* We have called $L_n(\boldsymbol{x}|\theta)$ *the* Radon-Nikodym derivative of $P_\theta^{(n)}$ with respect to $\mu^{(n)}$. The Radon-Nikodym theorem asserts that there exists *a* $\beta^{(n)}$-measurable version of the Radon-Nikodym derivative, but this version of the Radon-Nikodym derivative need not satisfy Conditions 2.1 and 2.3. Condition 2.4 asserts that there is a version satisfying Conditions 2.1 and 2.3 which is also $\beta^{(n)}$-measurable for each $\theta \in \Theta$. (See also Wald [14], pp. 596–597).

(iii) *there exists a Δ-measurable function T: $S \to Y$ satisfying $U(s, T(s)) = \sup_{y \in Y} U(s, y)$ for all $s \in S$.*

PROOF. Since Y is a compact subset of k-dimensional space, Y is bounded. Thus, there exists a k-dimensional cube C containing Y. We may partition C into m^k subcubes, each subcube having volume m^{-k} times the volume of C. Let $C(p, m)$ be the pth such subcube formed in this fashion, and let $z(p, m)$ be any fixed point in $C(p, m) \cap Y$ (if $C(p, m) \cap Y = \phi$, this does not effect the proof), $p=1, 2, \cdots, m^k$; $m=1, 2, \cdots$. Define $U_m(s, y)$ to be equal to $U(s, z(p, m))$ whenever $y \in C(p, m)$, $p=1, 2, \cdots, m^k$; $m=1, 2, \cdots$. Then clearly $U_m(s, y)$ is $\Delta \times \Gamma$-measurable. Since $U(s, \cdot)$ is a continuous function on Y for every $s \in S$, it follows that $\lim_{m \to \infty} U_m(s, y) = U(s, y)$ for all $s \in S$ and for all $y \in Y$. Hence, $U(s, y)$ is the pointwise limit of $\Delta \times \Gamma$-measurable functions, and therefore $U(\cdot, \cdot)$ is $\Delta \times \Gamma$-measurable, proving (i).

Define t_m : $S \to [0, \infty)$ by

$$t_m(s) = \left[\int_Y (U(s, y))^m dy \right]^{1/m} \equiv || U(s, \cdot) ||_m ,$$

for $m=1, 2, \cdots$. Since Y is compact and $U(s, \cdot)$ is continuous on Y for all s, it follows that $|| U(s, \cdot) ||_m < \infty$ for all $s \in S$. Let

$$t(s) = \sup_{y \in Y} U(s, y) \equiv || U(s, \cdot) ||_\infty .$$

For each m, $t_m(s)$ is Δ-measurable (Fubini's Theorem), and since $\lim_{m \to \infty} || U(s, \cdot) ||_m = || U(s, \cdot) ||_\infty$ for all $s \in S$, it follows that $t(s)$ is a pointwise limit of Δ-measurable functions and consequently is Δ-measurable. This verifies (ii). The existence of T: $S \to Y$ which is Δ-measurable and satisfies $U(s, T(s)) = \sup_{y \in Y} U(s, y)$ for almost all $s \in S$ now follows as a direct consequence of Theorem 2 of Olech [13]. QED

LEMMA 2.2. *Let S be a sample space with a sigma field Δ of subsets, and let Y be a closed subset of Euclidean k-dimensional space with the appropriate Borel sigma field Γ of subsets. Let $U(s, y)$ map $S \times Y$ into $[0, \infty)$. If for every $s \in S$, $U(s, \cdot)$ is a continuous function on Y, and if for every $y \in Y$, $U(\cdot, y)$ is a Δ-measurable function, then for any σ-compact subset F of Y, the restriction of $U(s, y)$ to $S \times F$ is $\Delta \times \Gamma_F$-measurable (where Γ_F is the sub-Borel-field of Γ consisting of subsets of F which belong to Γ), and $U(s, F) \equiv \sup_{y \in F} U(s, y)$ is Δ-measurable. Further, if there exists an increasing sequence $\{K_t,\ t=1, 2, \cdots\}$ of compact subsets of Y such that $Y = \bigcup_{t=1}^{\infty} K_t$ and*

$$\lim_{t\to\infty} \sup_{y \in K_t^c \cap Y} U(s, y) = 0$$

for all $s \in S$, *then there exists a* Δ*-measurable function* $T: S \to Y$ *for which* $\sup_{y \in Y} U(s, y) = U(s, T(s))$, *all* $s \in S$.

PROOF. Since F is σ-compact, there exists a sequence $\{K_t: t=1, 2, \cdots\}$ of compact subsets of F (and hence of Y) such that $F = \bigcup_{t=1}^{\infty} K_t$. Let $U_t(s, y)$ be equal to $U(s, y)$ on $S \times K_t$ and be equal to 0 otherwise. Since $K_t \subseteq K_{t+1}$, all t, and $F = \bigcup_{t=1}^{\infty} K_t$, it follows that $U(s, y)$ restricted to $S \times F$ is equal to $\lim_{t\to\infty} U_t(s, y)$ for all $(s, y) \in S \times F$. Since $U(s, y)$ restricted to $S \times K_t$ is $\Delta \times \Gamma_{K_t}$-measurable for all $t=1, 2, \cdots$, by Lemma 2.1, it follows that $U_t(s, y)$ is $\Delta \times \Gamma_F$-measurable for each $t=1, 2, \cdots$. Hence, $U(s, y)$ is the limit of $\Delta \times \Gamma_F$-measurable functions, and therefore is $\Delta \times \Gamma_F$-measurable. Also, since by Lemma 2.1, $U(s, K_t) \equiv \sup_{y \in K_t} U(s, y)$ exists and is Δ-measurable for all $t=1, 2, \cdots$, and since $F = \bigcup_{t=1}^{\infty} K_t$, then $U(s, F) = \sup_{1 \leq t \leq \infty} U(s, K_t)$ is the supremum of a countable number of Δ-measurable functions, and thus is Δ-measurable.

For future reference, we remark that since Euclidean k-dimensional space $\mathcal{E}^{(k)}$ is itself σ-compact, every open set and every closed set in $\mathcal{E}^{(k)}$ is σ-compact. Hence Y is σ-compact, and $U(s, Y) = \sup_{y \in Y} U(s, y)$ exists and is Δ-measurable.

Now assume that $Y = \bigcup_{t=1}^{\infty} K_t$, where $\{K_t, t=1, 2, \cdots\}$ is an increasing sequence of compact subsets of Y and

$$\lim_{t\to\infty} \sup_{y \in K_t^c \cap Y} U(s, y) = 0\,, \qquad \text{all } s \in S\,. \tag{2.3}$$

By Lemma 2.1, for each $t=1, 2, \cdots$, $U(s, K_t) \equiv \sup_{y \in K_t} U(s, y)$ exists and is Δ-measurable. Also by Lemma 2.1, for each $t=1, 2, \cdots$, there exists a Δ-measurable function $T_t: S \to K_t$ satisfying

$$U(s, K_t) = U(s, T_t(s))\,, \qquad \text{all } s \in S\,. \tag{2.4}$$

Let $T_1^*(s) = T_1(s)$, all $s \in S$, and let

$$T_t^*(s) = \begin{cases} T_{t-1}(s)\,, & \text{if } U(s, Y) = U(s, K_t) = U(s, K_{t-1})\,, \\ T_t(s)\,, & \text{if } U(s, Y) = U(s, K_t) > U(s, K_{t-1})\,, \\ y_0\,, & \text{if } U(s, Y) > U(s, K_t)\,, \end{cases}$$

where y_0 is some fixed point in Y. Since $U(s, K_t)$ is non-decreasing in

t for each fixed $s \in S$, $T_t^*(s)$ is well-defined for all $s \in S$. By the Δ-measurability of $T_t(s)$ and $U(s, K_t)$, all $t=1, 2, \cdots$, by the Δ-measurability of $U(s, Y)$, and from the definition of $T_t^*(s)$, it follows that $T_t^*(s)$ is Δ-measurable for all $t=1, 2, \cdots$. Further, from the fact that $T_t(s) \in K_t \subseteq Y$, for all $s \in S$, all $t=1, 2, \cdots$, it follows that $T_t^*(s) \in Y$, for all $s \in S$, all $t=1, 2, \cdots$. Finally, for every $s \in S$, it follows from (2.3), and the fact that for all $s \in S$, $t=1, 2, \cdots$,

$$U(s, Y)=\max\{U(s, K_t), U(s, K_t^c \cap Y)\} ,$$

that for every $s \in S$, there exists integer $t_0=t_0(s) \geqq 1$ such that for all $t \geqq t_0$,

$$T_t^*(s)=T_{t_0}^*(s) . \tag{2.5}$$

Hence $\lim_{t\to\infty} T_t^*(s)$ exists in Y for each $s \in S$. Let $T(s)=\lim_{t\to\infty} T_t^*(s)$. Then $T: S \to Y$ is the pointwise limit of a sequence of Δ-measurable functions and hence is Δ-measurable. From (2.5) it follows that for each $s \in S$, there exists $t_0(s) \geqq 1$ such that for all $t \geqq t_0(s)$,

$$U(s, Y)=U(s, K_t)=U(s, K_{t_0(s)})=U(s, T_t^*(s)) .$$

Hence, since $U(s, y)$ is continuous in y for all $y \in Y$,

$$U(s, Y)=\lim_{t\to\infty} U(s, T_t^*(s))=U(s, T(s)) ,$$

for all $s \in S$. This completes the proof. QED

Let

$$H_n(\boldsymbol{x} \mid \bar{\Theta})=\sup_{\theta\in\bar{\Theta}} H_n(\boldsymbol{x} \mid \theta) , \qquad H_n(\boldsymbol{x} \mid \Theta)=\sup_{\theta\in\Theta} H_n(\boldsymbol{x} \mid \theta) . \tag{2.6}$$

By Lemma 2.2, $H_n(\boldsymbol{x} \mid \bar{\Theta})$ is $\beta^{(n)}$-measurable. Further, since $H_n(\boldsymbol{x} \mid \theta)$ is, for each $\boldsymbol{x} \in \mathcal{X}^{(n)}$, continuous in θ for all $\theta \in \bar{\Theta}$,

$$H_n(\boldsymbol{x} \mid \Theta)=H_n(\boldsymbol{x} \mid \bar{\Theta}) \qquad \text{all } \boldsymbol{x} \in \mathcal{X}^{(n)} ,$$

so that $H_n(\boldsymbol{x} \mid \Theta)$ is also $\beta^{(n)}$-measurable.

THEOREM 2.1. *If for a given integer $n \geqq 1$, Conditions 2.1 through 2.4 hold, there exists a $\beta^{(n)}$-measurable function $\theta_n^*(\boldsymbol{x}): \mathcal{X}^{(n)} \to \bar{\Theta}$ satisfying*

$$H_n(\boldsymbol{x} \mid \Theta)=H_n(\boldsymbol{x} \mid \bar{\Theta})=H_n(\boldsymbol{x} \mid \theta_n^*(\boldsymbol{x}))$$

for (almost) all $\boldsymbol{x} \in \mathcal{X}^{(n)}$. That is, for that n, the M.P.E. $\theta_n^*(\xi)$ exists and is $\beta^{(n)}$-measurable.

PROOF. Since by Conditions 2.1, 2.2 and 2.4, $H_n(\boldsymbol{x} \mid \theta)$ is $\beta^{(n)}$-measurable in $\boldsymbol{x}$ for each fixed θ, and continuous in θ for each fixed $\boldsymbol{x}\in\mathscr{X}^{(n)}$, and since by Condition 2.3, there exists an increasing sequence $\{K_t\colon t=1,2,\cdots\}$ of compact subsets of $\bar{\Theta}$ such that $\bar{\Theta}=\bigcup_{t=1}^{\infty}K_t$ and $\lim_{t\to\infty}\sup_{\theta\in K_t^c\cap\bar{\Theta}}H_n(\boldsymbol{x}|\theta)=0$ for all $\boldsymbol{x}\in\mathscr{X}^{(n)}$, the asserted result is a direct consequence of Lemma 2.2. QED

Before turning to a proof of strong consistency, we note that if for some $n_0\geqq 1$, Conditions 2.1, 2.3 and 2.4 are satisfied, then since

$$\prod_{i=1}^{n} f(x_i|\theta)=\prod_{i=1}^{n}\left(\prod_{j\neq i} f(x_j|\theta)\right)^{1/(n-1)},$$

it follows by induction on n that Conditions 2.1, 2.3 and 2.4 hold for all $n\geqq n_0$, and thus that $\theta_n^*(\boldsymbol{x})$ exists and is $\beta^{(n)}$-measurable for all $n\geqq n_0$.

For a proof of the strong consistency of θ_n^*, we add four additional conditions. In what follows, N equals a fixed integer n for which Conditions 2.1, 2.3 and 2.4 hold simultaneously.

CONDITION 2.5. For all $\theta\in\Theta$, $g(\theta)>0$.

CONDITION 2.6. If θ and θ', $\theta\neq\theta'$, are two distinct points, $\theta\in\Theta$, $\theta'\in\bar{\Theta}$, we have

$$\mu^{(N)}(\{\boldsymbol{x}\colon \boldsymbol{x}\in\mathscr{X}^{(N)},\ L_N(\boldsymbol{x}|\theta)\neq L_N(\boldsymbol{x}|\theta')\})>0 .$$

CONDITION 2.7. For every $\theta\in\bar{\Theta}\sim\Theta$,

$$\int_{\mathscr{X}^{(N)}} L_N(\boldsymbol{x}|\theta)d\mu(\boldsymbol{x})\leqq 1 .$$

CONDITION 2.8. For every $\theta\in\Theta$,

$$E_\theta \log\left[\frac{L_N((X_1,X_2,\cdots,X_N)|\Theta)}{L_N((X_1,X_2,\cdots,X_N)|\theta)}\right]<\infty .$$

It is important to note that we do not require that Conditions 2.6, 2.7 and 2.8 hold for the smallest $n=n_0$ such that Conditions 2.1, 2.3 and 2.4 hold simultaneously, but only that these conditions hold for some N greater than or equal to that smallest n.

For $n\geqq N$ and for fixed $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)\in\mathscr{X}^{(n)}$, define the sets

$$\text{(2.7)}\qquad \begin{aligned} \Lambda_n(\boldsymbol{x})&=\{\theta\colon \theta\in\Theta,\ H_n(\boldsymbol{x}|\theta)=H_n(\boldsymbol{x}|\Theta)\} ,\\ \Lambda_n^*(\boldsymbol{x})&=\{\theta\colon \theta\in\Theta,\ H_n(\boldsymbol{x}|\theta)\geqq 1/2\cdot H_n(\boldsymbol{x}|\Theta)\} .\end{aligned}$$

Note that for all $n\geqq N$, and all $\boldsymbol{x}\in\mathscr{X}^{(n)}$,

$$\text{(2.8)}\qquad \Lambda_n(\boldsymbol{x})\subset\Lambda_n^*(\boldsymbol{x}) .$$

Finally for any $\theta' \in \bar{\Theta}$, any $\varepsilon>0$, let $N(\theta', \varepsilon)$ denote the set

$$N(\theta', \varepsilon)=\{\theta'': \theta'' \in \bar{\Theta},\ |\theta'-\theta''|<\varepsilon\}\ . \tag{2.9}$$

LEMMA 2.3. *Assume that Conditions* 2.1 *through* 2.8 *hold. Then for any* $\varepsilon>0$ *and for each given* $\theta \in \Theta$,

$$P_\theta^{(\infty)}\{\Lambda_n^*(\boldsymbol{x})=\Lambda_n^*((X_1, X_2, \cdots, X_n))\subset N(\theta, \varepsilon) \text{ for all sufficiently large } n\}=1\ . \tag{2.10}$$

PROOF. Let $\theta \in \Theta$ be given. Let $\varepsilon>0$ be fixed. Let $\{K_t,\ t=1, 2, \cdots\}$ be the increasing sequence of compact sets guaranteed by Condition 2.3. There exists a large enough integer t_0 so that $N(\theta, \varepsilon)\subset K_{t_0}$. Since for any $t\geqq 1$, $K_t(\theta, \varepsilon)=K_t \sim N(\theta, \varepsilon)$ is the intersection of the compact set K_t and the closed set $N^c(\theta, \varepsilon)$, $K_t(\theta, \varepsilon)$ is compact.

Conditions 2.1 and 2.4 and Lemma 2.2 imply that $L_N(\boldsymbol{x}\,|\,N(\theta', \delta))=\sup_{\theta'' \in N(\theta', \delta)} L_N(\boldsymbol{x}\,|\,\theta'')$ is for any $\theta' \in \bar{\Theta}$, any $\delta>0$, a $\beta^{(N)}$-measurable function of $\boldsymbol{x}$. Further, Condition 2.1 also implies that

$$\lim_{\delta\to 0} \log\left[\frac{L_N(\boldsymbol{x}\,|\,N(\theta', \delta))}{L_N(\boldsymbol{x}\,|\,\theta)}\right]=\log\left[\frac{L_N(\boldsymbol{x}\,|\,\theta')}{L_N(\boldsymbol{x}\,|\,\theta)}\right], \tag{2.11}$$

for all $\boldsymbol{x} \in \mathcal{X}^{(N)}$. Since $L_N(\boldsymbol{x}\,|\,N(\theta', \delta))$ is for fixed θ', $\boldsymbol{x}$, an increasing function of δ always bounded above by $L_N(\boldsymbol{x}\,|\,\Theta)$, and since $\log[L_N(\boldsymbol{x}\,|\,\Theta)/L_N(\boldsymbol{x}\,|\,\theta)]$ is nonnegative for all $\boldsymbol{x} \in \mathcal{X}^{(N)}$, it follows from Condition 2.8, Equation (2.11), and the Lebesgue Monotone Convergence Theorem that

$$\lim_{\delta\to 0} E_\theta \log\left[\frac{L_N(\boldsymbol{X}\,|\,N(\theta', \delta))}{L_N(\boldsymbol{X}\,|\,\theta)}\right]=E_\theta \log\left[\frac{L_N(\boldsymbol{X}\,|\,\theta')}{L_N(\boldsymbol{X}\,|\,\theta)}\right]. \tag{2.12}$$

Since the logarithm function is strictly concave, it follows from Conditions 2.6 and 2.7, and from Jensen's inequality that $E_\theta \log[L_N(\boldsymbol{X}\,|\,\theta')/L_N(\boldsymbol{X}\,|\,\theta)]<0$ for all $\theta' \in K_t(\theta, \varepsilon)$, $t\geqq t_0$. Hence, given $\theta' \in K_t(\theta, \varepsilon)$ for fixed $t\geqq t_0$, there exists $\delta(\theta')$, $\eta(\theta')>0$ such that

$$E_\theta \log\left[\frac{L_N(\boldsymbol{X}\,|\,N(\theta', \delta(\theta')))}{L_N(\boldsymbol{X}\,|\,\theta)}\right]<-\eta(\theta')\ . \tag{2.13}$$

Let t be any integer $\geqq t_0$. Since $K_t(\theta, \varepsilon)$ is compact, there exist $\theta_1, \theta_2, \cdots, \theta_s \in K_t(\theta, \varepsilon)$ such that $K_t(\theta, \varepsilon)\subset \bigcup_{i=1}^{s} N(\theta_i, \delta(\theta_i))$. Hence since

$$L_N(\boldsymbol{x}\,|\,K_t(\theta, \varepsilon))=\sup_{\theta' \in K_t(\theta, \varepsilon)} L_N(\boldsymbol{x}\,|\,\theta')\leqq \max_{1\leqq i\leqq s} L_N(\boldsymbol{x}\,|\,N(\theta_i, \delta(\theta_i)))\ ,$$

it follows that

$$E_\theta \log\left[\frac{L_N(\boldsymbol{X}\,|\,K_t(\theta, \varepsilon))}{L_N(\boldsymbol{X}\,|\,\theta)}\right]\leqq \max_{1\leqq i\leqq s} -\eta(\theta_i)\equiv -\eta<0\ . \tag{2.14}$$

Consider $L_N(\boldsymbol{x}\,|\,K_t^c)=\sup_{\theta'\in K_t^c\cap\bar{\Theta}} L_N(\boldsymbol{x}\,|\,\theta')$. Since $K_t^c\cap\bar{\Theta}$ is σ-compact, it follows from Conditions 2.1 and 2.4 and from Lemma 2.2 that $L_N(\boldsymbol{x}\,|\,K_t^c)$ is $\beta^{(N)}$-measurable in $\boldsymbol{x}$. By Condition 2.3, $\lim_{t\to\infty} L_N(\boldsymbol{x}\,|\,K_t^c)=0$ for all $\boldsymbol{x}\in\mathcal{X}^{(N)}$. Further, for each $\boldsymbol{x}\in\mathcal{X}^{(N)}$, $L_N(\boldsymbol{x}\,|\,K_t^c)$ is decreasing in t and bounded above by $L_N(\boldsymbol{x}\,|\,\Theta)$. Hence, from the Lebesgue Monotone Convergence Theorem and Condition 2.8 it follows that $\lim_{t\to\infty} E_\theta \log\,[L_N(\boldsymbol{X}\,|\,K_t^c)/L_N(\boldsymbol{X}\,|\,\theta)]=-\infty$, and therefore that we can choose $t^*\geqq t_0$ so large that for the fixed $\eta>0$ defined in (2.14),

$$E_\theta \log\left[\frac{L_N(\boldsymbol{X}\,|\,K_{t^*}^c)}{L_N(\boldsymbol{X}\,|\,\theta)}\right]<-\eta\ . \tag{2.15}$$

From the preceeding results, we can conclude that

$$L_N(\boldsymbol{x}\,|\,N^c(\theta,\varepsilon))\equiv \sup_{\theta'\in N^c(\theta,\varepsilon)\cap\bar{\Theta}} L_N(\boldsymbol{x}\,|\,\theta')=\max\{L_N(\boldsymbol{x}\,|\,K_{t^*}^c),\ L_N(\boldsymbol{x}\,|\,K_{t^*}(\theta,\varepsilon))\}$$

is a $\beta^{(N)}$-measurable function of $\boldsymbol{x}$, and that

$$E_\theta \log\left[\frac{L_N(\boldsymbol{X}\,|\,N^c(\theta,\varepsilon))}{L_N(\boldsymbol{X}\,|\,\theta)}\right]<-\eta\ . \tag{2.16}$$

Let n be any integer $\geqq N$. We note that

$$\begin{aligned}\log H_n((x_1,x_2,\cdots,x_n)\,|\,\theta') \\ =\frac{n}{\binom{n}{N}}\sum \log L_N((x_{i_1},x_{i_2},\cdots,x_{i_N})\,|\,\theta')+\log g(\theta')\end{aligned} \tag{2.17}$$

where the summation is taken over the $\binom{n}{N}$ possible choices of N indices $i_1, i_2,\cdots, i_N$ from among the n indices $1,2,3,\cdots,n$. Thus by the superadditivity of the supremum function and the fact that the logarithm is a monotonic nondecreasing function, it follows that

$$\begin{aligned}&\frac{1}{n}\log\left[\frac{H_n((x_1,x_2,\cdots,x_n)\,|\,N^c(\theta,\varepsilon))}{H_n((x_1,x_2,\cdots,x_n)\,|\,\theta)}\right] \\ &\leqq \frac{1}{\binom{n}{N}}\sum\log\left[\frac{L_N(x_{i_1},x_{i_2},\cdots,x_{i_N}\,|\,N^c(\theta,\varepsilon))}{L_N((x_{i_1},x_{i_2},\cdots,x_{i_N})\,|\,\theta)}\right]+\frac{1}{n}\log\frac{g(\Theta)}{g(\theta)}\ .\end{aligned} \tag{2.18}$$

We recognize the first term on the right-hand side of inequality (2.18) as a U-statistic. Since the sequence of such statistics (indexed by n) is a reverse Martingale, it follows from the Martingale Convergence Theorem (see Berk [4]) that

$$(2.19)\quad \frac{1}{\binom{n}{N}} \sum \log\left[\frac{L_N((X_{i_1}, X_{i_2}, \cdots, X_{i_N}) \mid N^c(\theta, \varepsilon))}{L_N((X_{i_1}, X_{i_2}, \cdots, X_{i_N}) \mid \theta)}\right]$$

$$\xrightarrow{\text{a.s.}} E_\theta \log\left[\frac{L_N((X_1, X_2, \cdots, X_N) \mid N^c(\theta, \varepsilon))}{L_N((X_1, X_2, \cdots, X_N) \mid \theta)}\right] < -\eta < 0 .$$

Since by Conditions 2.2 and 2.5, $n^{-1}\log[g(\Theta)/g(\theta)] \to 0$ as $n \to \infty$, it follows that

$$(2.20)\quad P_\theta\left\{\frac{1}{n}\log\left[\frac{H_n((X_1, X_2, \cdots, X_n) \mid N^c(\theta, \varepsilon))}{H_n((X_1, X_2, \cdots, X_n) \mid \theta)}\right] < -\frac{\eta}{2} \text{ for all sufficiently large } n\right\} = 1 .$$

Hence, for all $\theta' \in N^c(\theta, \varepsilon)$

$$H_n(\xi \mid \theta') < e^{-n\eta/2} H_n(\xi \mid \theta) \leqq e^{-n\eta/2} H_n(\xi \mid \Theta) \leqq H_n(\xi \mid \Theta)/2$$

for all sufficiently large n with probability one under $P_\theta^{(\infty)}$, all $\theta \in \Theta$. This establishes (2.10). QED

THEOREM 2.2. *Under Conditions 2.1 through 2.8, there exists an integer $N \geqq 1$ such that $\theta_n^*(\xi)$ exists and is $\beta^{(n)}$-measurable for all $n \geqq N$ and such that*

$$\theta_n^*(\xi) \to \theta \quad \text{a.s. } P_\theta^{(\infty)} \qquad \text{as } n \to \infty$$

for all $\theta \in \Theta$.

PROOF. The first part of the assertion is a restatement of Theorem 2.1. Since θ_n^* exists and is measurable for $n \geqq N$, $\theta_n^*(s) \in \Lambda_n(s) \subset \Lambda_n^*(s)$, for all $s \in \mathcal{X}^{(\infty)}$, so that $\Lambda_n^*(s)$ is non-empty for $n \geqq N$. The asserted strong convergence of $\theta_n^*(\xi)$ now follows as a direct consequence of Lemma 2.3. QED

The regularity conditions which we have used in this section closely resemble Kiefer and Wolfowitz's [9] and Bahadur's [3] modifications of the regularity conditions originally adopted by Wald [14]. Of course, to cover the greater generality of our estimator $\theta_n^*(\xi)$, we have added Conditions 2.2 and 2.5, which are regularity conditions on the prior density $g(\theta)$. Since we nowhere used (or assumed) the requirement that $g(\theta)$ be a probability density on Θ, the density $g(\theta) \equiv 1$, all $\theta \in \Theta$, satisfies our conditions. Thus, our conditions cover the special case of the M.L.E. $\hat{\theta}_n$.

Our assumptions do differ in some respect from the regularity conditions of Kiefer and Wolfowitz [9]. The major differences between Conditions 2.1–2.8 and the conditions of Kiefer and Wolfowitz [9] are:

(a) Our Condition 2.1 requires continuity in θ of $L_N(\boldsymbol{x}|\theta)$. In contrast, Kiefer and Wolfowitz [9] require that $\sup_{\theta' \in N(\theta,\delta)} f(x|\theta') \rightarrow f(x|\theta)$ as $\delta \rightarrow 0$.

(b) A comment by Kiefer and Wolfowitz [9] that if their regularity conditions do not hold for $L_1(x|\theta)=f(x|\theta)$, everything still goes through if the regularity conditions hold for $L_N(\boldsymbol{x}|\theta)$, some $N \geqq 1$, has been directly incorporated into our regularity conditions (see also Berk [4]). This enables us, for example, to verify existence and strong consistency of the M.L.E. $\hat{\theta}_n$ and M.P.E. θ_n^* for the case of the normal distribution with unknown mean and variance.

(c) We have confined ourselves to parameters which are points in k-dimensional Euclidean space under the usual metric. Further, we have not tried to introduce special measures of distance or to transform our regularity conditions on the densities $f(x|\theta)$ to conditions on transformations of the density functions $f(x|\theta)$—devices mentioned by Kiefer and Wolfowitz [9], and used more explicitly by Huber [8].

The restriction in our Condition 2.1 pointed out in remark (a) above can probably be removed by appropriately modifying the Theorem 2 of Olech [13] used in proving Lemma 2.1. However, until we can remove this restriction, our conditions do not cover the case of estimation of the parameters for the uniform distribution and other similar distributions. The conditions of Kiefer and Wolfowitz, however, do cover these distributions. The restrictions noted in point (c) above may also limit the applicability of our results somewhat. These restrictions were needed to prove Lemma 2.1. On the other hand, as noted above, the incorporation of Kiefer and Wolfowitz's remark directly into our regularity conditions provides added flexibility and applicability to our results. Actually, Kiefer and Wolfowitz only mentioned looking at certain integrability results (Condition 2.8) for large enough n; we have extended their suggestion to continuity and measurability conditions. Our proof of strong consistency of the M.L.E. also differs somewhat from that indicated by Kiefer and Wolfowitz [9].

Perhaps the major novelty of our results is the rigorous proof (Theorem 2.1) we give for the existence and measurability of the M.P.E. θ_n^* for fixed sample size n. The existence of $\theta_n^*(s)$ for fixed $s=(x_1, x_2, \cdots)$ and large enough n has been argued by many authors (Bahadur [3], LeCam [10], etc.), but this does not prove that $\theta_n^*((x_1, x_2, \cdots, x_n))$ exists and is $\beta^{(n)}$-measurable for each n. The truth of this missing fact has been asserted by the authors mentioned above, but (probably due to considerations of space) they have omitted mentioning sufficient conditions or providing a proof for such a result. Without a proof of the existence and $\beta^{(n)}$-measurability of $\theta_n^*(s)$ for all large enough n, the usual proof of the consistency of such estimators (as given in our Lemma 2.3—although we have modified the usual proof slightly) is not completely rigorous (as Huber [8] remarks, it can probably be done by judicious use of inner and outer probabilities).

3. Asymptotic normality

In the previous section we established conditions sufficient to prove that the M.P.E. θ_n^* with respect to a prior density $g(\theta)$ is a strongly consistent estimator of θ. In the present section, we show that under certain additional regularity conditions when θ is a one-dimensional parameter, the M.P.E. is asymptotically normally distributed. We also derive a useful representation for the sequence $\{\theta_n^*\}$, and indicate a rate for the almost sure convergence of θ_n^* to θ.

Denote $\log f(x|\theta)$ by $l(x|\theta)$, and let

$$l^{(i)}(x|\theta')=\left(\frac{\partial}{\partial\theta}\right)^i l(x|\theta)|_{\theta=\theta'}$$

for $i=1, 2,\cdots$. Let

$$l_n^{(i)}(\xi|\theta)=\sum_{j=1}^n l^{(i)}(X_j|\theta) ,$$

where $\xi=(X_1, X_2,\cdots)$. In what follows we assume that the M.L.E. $\hat{\theta}_n(\xi)$ and the M.P.E. $\theta_n^*(\xi)$ with respect to a given prior density $g(\theta)$ exist for large enough n and are both strongly consistent estimators of θ. Further, we assume that for all $s=(x_1, x_2,\cdots)\in\mathcal{X}^{(\infty)}$, $\hat{\theta}_n(s)=\hat{\theta}_n((x_1, x_2,\cdots, x_n))$ is a solution of the equation

$$l_n^{(1)}(s|\theta)=0 , \tag{3.1}$$

and that $\theta_n^*(s)=\theta_n^*((x_1, x_2,\cdots, x_n))$ is a solution to the equation

$$l_n^{(1)}(s|\theta)+\frac{\partial}{\partial\theta}\log g(\theta)=0 , \tag{3.2}$$

for all large enough n. This will be the case if $l^{(1)}(x|\theta)$ exists and $\partial/\partial\theta \log g(\theta)$ exists for all $\theta\in\Theta$, all $x\in\mathcal{X}$, and if the maxima of $\prod_{i=1}^n f(x_i|\theta)$ and of $g(\theta)\prod_{i=1}^n f(x_i|\theta)$ exist within the interior of Θ for all (large enough) n. To obtain the main results of this section, we need the following added conditions:

CONDITION 3.1. For each $x\in\mathcal{X}$, $l^{(2)}(x|\theta)$ exists and is continuous in θ for all $\theta\in\Theta$.

CONDITION 3.2. For each $\theta \in \Theta$, we have $E_\theta l^{(1)}(X|\theta)=0$ and

$$0<E_\theta(l^{(1)}(X|\theta))^2=-E_\theta l^{(2)}(X|\theta)\equiv I(\theta)<\infty \ .$$

CONDITION 3.3. For each $\theta \in \Theta$, there exists a δ-neighborhood, say $N(\theta, \delta)=\{\theta': |\theta-\theta'|<\delta\}\subset\Theta$, of θ and a measurable function $M(x)$ on $(\mathcal{X}, \beta)$ such that

$$|l^{(2)}(x|\theta')-l^{(2)}(x|\theta)|<M(x)$$

for all $x \in \mathcal{X}$ and all $\theta' \in N(\theta, \delta)$, and such that $E_\theta M(X)<\infty$.

CONDITION 3.4. $\partial/\partial\theta \log g(\theta)$ exists for all $\theta \in \Theta$.

Conditions 3.1 through 3.3 have previously been adopted by Bahadur [2].

Before stating the main result of this section, we need to introduce some new notation. By the symbol $o_a(1/b_n)$ we represent any sequence of random variables $\{Z_n\}$ defined on $(\mathcal{X}^{(\infty)}, \beta^{(\infty)})$ which has the property that for the sequence $\{b_n\}$ of constants, $b_n\to\infty$ as $n\to\infty$, $b_nZ_n\to 0$ almost surely $(P_\theta^{(\infty)})$, $n\to\infty$. The symbol $o_a(1)$ represents any sequence of random variables $\{Z_n\}$ defined on $(\mathcal{X}^{(\infty)}, \beta^{(\infty)})$ which converges almost surely $(P_\theta^{(\infty)})$ to zero as $n\to\infty$. Addition, subtraction, multiplication, or division of the symbols $o_a(1/b_n)$ refer to the same operations performed on the corresponding sequences of random variables. These operations almost surely obey the calculus of $o(1/b_n)$. Also, in what follows, the symbols $o_p(1)$, $o_p(1/b_n)$, $O_p(1)$, $O_p(1/b_n)$ have their usual meanings (see Mann and Wald [12]).

THEOREM 3.1. *Under Conditions 3.1 through 3.4 and under the assumption that the M.L.E. $\hat{\theta}_n(\xi)$ and the M.P.E. $\theta_n^*(\xi)$ exist (for all sufficiently large n) and are strongly consistent estimators of θ satisfying Equation (3.1) and Equation (3.2) respectively, it follows that for each $\theta \in \Theta$,*

$$\hat{\theta}_n=\theta+\frac{1}{nI(\theta)}l_n^{(1)}(\xi|\theta)(1+o_a(1)) \ , \tag{3.3}$$

$$\theta_n^*=\theta+\frac{1}{nI(\theta)}\left[l_n^{(1)}(\xi|\theta)+\frac{\partial}{\partial\theta}\log g(\theta)|_{\theta=\theta_n^*}\right][1+o_a(1)] \ , \tag{3.4}$$

as $n\to\infty$.

PROOF. Since $\hat{\theta}_n(\xi)$ and $\theta_n^*(\xi)$ both are strongly consistent estimators of θ, we can assume that

$$\hat{\theta}_n(\xi)=\theta+k_n(\xi, \theta) \ , \qquad \theta_n^*(\xi)=\theta+h_n(\xi, \theta) \ , \tag{3.5}$$

where $k_n(\xi, \theta)=o_a(1)$, $h_n(\xi, \theta)=o_a(1)$ as $n\to\infty$. Since $\hat{\theta}_n(\xi)$ satisfies (3.1) and $\theta_n^*(\xi)$ satisfies (3.2), we have

$$l_n^{(1)}(\xi\,|\,\hat{\theta}_n(\xi))=0 \,, \tag{3.6}$$

and

$$l_n^{(1)}(\xi\,|\,\theta_n^*(\xi))+\frac{\partial}{\partial\theta}\log g(\theta)|_{\theta=\theta_n^*(\xi)}=0 \,, \tag{3.7}$$

for all large enough n. By Condition 3.1, we can expand $l^{(1)}(\xi\,|\,\hat{\theta}_n(\xi))$ and $l^{(1)}(\xi\,|\,\theta_n^*(\xi))$ each in a Taylor's expansion around the true θ, obtaining

$$0=l_n^{(1)}(\xi\,|\,\hat{\theta}(\xi))=l_n^{(1)}(\xi\,|\,\theta)+k_n(\xi,\theta)l_n^{(2)}(\xi\,|\,\eta_n) \,, \tag{3.8}$$

and

$$\begin{aligned}0&=l_n^{(1)}(\xi\,|\,\theta_n^*(\xi))+\frac{\partial}{\partial\theta}\log g(\theta)|_{\theta=\theta_n^*(\xi)}\\ &=l_n^{(1)}(\xi\,|\,\theta)+h_n(\xi,\theta)l_n^{(2)}(\xi\,|\,\alpha_n)+\frac{\partial}{\partial\theta}\log g(\theta)|_{\theta=\theta_n^*(\xi)} \,,\end{aligned} \tag{3.9}$$

where $\eta_n=\eta_n(\xi)$ lies between $\hat{\theta}_n(\xi)$ and θ, and where $\alpha_n=\alpha_n(\xi)$ lies between $\theta_n^*(\xi)$ and θ. Divide both sides of Equations (3.8) and (3.9) by $nI(\theta)$, where $I(\theta)>0$ is defined in Condition 3.2. We obtain

$$\frac{l_n^{(1)}(\xi\,|\,\theta)}{nI(\theta)}=(k_n(\xi,\theta))\left(-\frac{l_n^{(2)}(\xi\,|\,\eta_n)}{nI(\theta)}\right) \,, \tag{3.10}$$

and

$$\frac{1}{nI(\theta)}\left[l_n^{(1)}(\xi\,|\,\theta)+\frac{\partial}{\partial\theta}\log g(\theta)|_{\theta=\theta_n^*}\right]=(h_n(\xi,\theta))\left(-\frac{l_n^{(2)}(\xi\,|\,\alpha_n)}{nI(\theta)}\right) \,. \tag{3.11}$$

Since $\eta_n(\xi)$ is between $\hat{\theta}_n(\xi)$ and θ, and since $\alpha_n(\xi)$ is between $\theta_n^*(\xi)$ and θ, it follows from the fact that $\hat{\theta}_n(\xi)$ and $\theta_n^*(\xi)$ are strongly consistent estimators of θ that

$$\eta_n(\xi)-\theta=o_a(1) \,, \qquad \alpha_n(\xi)-\theta=o_a(1) \,.$$

Comparing (3.3) with (3.10), (3.4) with (3.11), and taking note of the definitions (3.5) of $k_n(\xi,\theta)$ and $h_n(\xi,\theta)$, we see that (3.3) and (3.4) are certainly verified if we can show that for any strongly consistent estimator $\{T_n(\xi)\}$ of θ, where $T_n(\xi)\equiv T_n(X_1,X_2,\cdots,X_n)$,

$$\frac{l_n^{(2)}(\xi\,|\,T_n)}{nI(\theta)}=-1+o_a(1) \,. \tag{3.12}$$

Since

$$\left|\frac{l_n^{(2)}(\xi|T_n)}{n}+I(\theta)\right| \leqq \left|\frac{l_n^{(2)}(\xi|\theta)}{n}+I(\theta)\right| + \left|\frac{l_n^{(2)}(\xi|T_n)-l_n^{(2)}(\xi|\theta)}{n}\right|$$

and since by Condition 3.2 and the strong law of large numbers

$$\left|\frac{l_n^{(2)}(\xi|\theta)}{n}+I(\theta)\right| = o_a(1) ,$$

to prove (3.12) we need only demonstrate that $n^{-1}|l_n^{(2)}(\xi|T_n)-l_n^{(2)}(\xi|\theta)|=o_a(1)$. But

$$n^{-1}|l_n^{(2)}(\xi|T_n)-l_n^{(2)}(\xi|\theta)| \leqq \frac{1}{n}\sum_{i=1}^{n}|l^{(2)}(X_i|T_n)-l^{(2)}(X_i|\theta)| , \tag{3.13}$$

so that if the right-hand side of (3.13) tends almost surely to 0 as $n\rightarrow\infty$, we are done.

Let

$$A(x, \delta)=\sup_{\theta'\in N(\theta,\delta)}\{|l^{(2)}(x|\theta')-l^{(2)}(x|\theta)|\}$$

and let

$$A(\delta)=E_\theta A(X, \delta) .$$

(Note. It can be shown using Condition 3.1 and Lemma 2.2 of Section 2 that $A(x, \delta)$ is measurable.)

It follows from Condition 3.1 that for every x, $A(x, \delta)\rightarrow 0$ as $\delta\rightarrow 0$. Hence, by Condition 3.3 and the Lebesgue Dominated Convergence Theorem, $A(\delta)\rightarrow 0$ as $\delta\rightarrow 0$. Given any $\varepsilon>0$, chose δ so small that $A(\delta)<\varepsilon$. Then since $P_\theta\{T_n \in N(\theta, \delta)$ for all large enough $n\}=1$, it follows that

$$P_\theta\left\{\frac{1}{n}\sum_{i=1}^{n}|l^{(2)}(X_i|T_n)-l^{(2)}(X_i|\theta)| \leqq \frac{1}{n}\sum_{i=1}^{n}A(X_i, \delta), \quad \text{for all large enough } n\right\}=1 .$$

Since by the strong law of large numbers,

$$\frac{1}{n}\sum_{i=1}^{n}A(X_i, \delta)\xrightarrow{\text{a.s.}}A(\delta)<\varepsilon , \qquad \text{as } n\rightarrow\infty ,$$

and since $\varepsilon>0$ can be chosen arbitrarily small, it follows that

$$\frac{1}{n}\sum_{i=1}^{n}|l^{(2)}(X_i|T_n)-l^{(2)}(X_i|\theta)|=o_a(1) .$$

This establishes (3.3) and (3.4). QED

COROLLARY 3.1. *Under the conditions of Theorem* 3.1, *and the additional condition that*

CONDITION 3.4′. Given $\theta \in \Theta$, there exists $\varepsilon>0$ such that

$$\sup_{\theta' \in N(\theta,\varepsilon)} \left| \frac{\partial}{\partial\theta} \log g(\theta)|_{\theta=\theta'} \right| < \infty ,$$

then the following relationships hold for any ρ, $0<\rho\leqq 1/2$:

$$n^{1/2-\rho}(\hat{\theta}_n-\theta)=o_a(1) , \qquad n^{1/2-\rho}(\theta_n^*-\theta)=o_a(1) , \tag{3.14}$$

$$n^{1/2-\rho}(\hat{\theta}_n-\theta_n^*)=o_a(1) . \tag{3.15}$$

Further,

$$n^{1/2}(\hat{\theta}_n-\theta_n^*)=o_p(1) , \tag{3.16}$$

and

$$n^{1/2}(\hat{\theta}_n-\theta)\xrightarrow{\mathcal{L}} N\left(0, \frac{1}{I(\theta)}\right) , \qquad n^{1/2}(\theta_n^*-\theta)\xrightarrow{\mathcal{L}} N\left(0, \frac{1}{I(\theta)}\right) . \tag{3.17}$$

It follows that under Condition 3.4′ *and the conditions of Theorem* 3.1, *the M.P.E.* θ_n^* *with respect to the prior density* $g(\theta)$ *is B.A.N.*

PROOF. By the law of the iterated logarithm for i.i.d. variables having zero mean and finite variance, and by Condition 3.2, it follows that for all ρ, $0<\rho\leqq 1/2$,

$$n^{1/2-\rho}\left(\frac{l_n^{(1)}(\xi\,|\,\theta)}{nI(\theta)}\right)=o_a(1) . \tag{3.18}$$

From Condition 3.4′ we know that there exists $\varepsilon>0$ for which

$$\sup_{\theta' \in N(\theta,\varepsilon)} \left| \frac{\partial}{\partial\theta} \log g(\theta)|_{\theta=\theta'} \right| < \infty .$$

Since $\theta_n^*-\theta=o_a(1)$, it follows that $P_\theta\{\theta_n^* \in N(\theta, \varepsilon)$ for all sufficiently large $n\}=1$ and thus that for all ρ, $0<\rho\leqq 1/2$,

$$\frac{n^{1/2-\rho}}{nI(\theta)} \frac{\partial}{\partial\theta} \log g(\theta)|_{\theta=\theta_n^*}=o_a(1) . \tag{3.19}$$

Assertion (3.14) now follows from (3.18) and (3.19). Assertion (3.15), of course, is an immediate consequence of (3.14).

Since almost sure convergence to zero implies convergence in probability to zero, and since by Condition 3.2 and the Central Limit Theorem

(3.20) $$\frac{\sqrt{n}}{nI(\theta)} l_n^{(1)}(\xi \mid \theta) \xrightarrow{\mathcal{L}} N\left(0, \frac{1}{I(\theta)}\right),$$

it follows from (3.3), (3.4), (3.19), and the calculus of $o_p(1)$ and $O_p(1)$, that (3.16) and (3.17) hold. This completes the proof. QED

Remark I. Results similar to Assertion (3.15) of Corollary 3.1 have been obtained by Bickel and Yahav [5]. These authors have shown that under certain regularity conditions, any Bayes estimator B_n under a convex loss function converges almost surely to the M.L.E. $\hat{\theta}_n$ at rate $1/\sqrt{n}$; i.e., that $\sqrt{n}(B_n - \hat{\theta}_n) = o_a(1)$. Their results have been generalized to the multiparameter case by Chao [6]. Although the results of Bickel and Yahav [5] and of Chao [6] are slightly stronger than our result (3.15), the regularity conditions assumed by these authors seem (as best we can compare them) to be somewhat more restrictive than ours. In any case, the results of these authors do not apply to our case since the M.P.E. θ_n^* is not necessarily a Bayes estimator for any convex loss function.

Remark II. The representation (3.3) for $\hat{\theta}_n$ has previously been used without proof by Weiss and Wolfowitz [15].

Remark III. The results in Theorem 3.1 and Corollary 3.1 can be rather straightforwardly generalized to the k-parameter case. Appropriate conditions for such an extension (except for conditions on the partial derivatives of $\log g(\theta)$) are given by Bahadur [2]. Essentially, what is needed is that the information matrix exists and is positive definite, that the second partials of $l(x \mid \theta)$ exist and are continuous in θ, and that we can apply the Lebesgue Dominated Convergence Theorem as in the proof of Theorem 3.1.

The results of this section show that under fairly weak regularity conditions on $g(\theta)$, the M.P.E. θ_n^* is asymptotically efficient in the classical Fisherian sense. In his Ph.D. dissertation (see Fu [7]), the first author of this paper has shown that θ_n^* has the maximum asymptotic probability defined by Bahadur [1], [3].

UNIVERSITY OF MANITOBA AND PURDUE UNIVERSITY

REFERENCES

[1] Bahadur, R. R. (1960). Asymptotic efficiency of tests and estimates, *Sankhyā*, **22**, 229-252.

[2] Bahadur, R. R. (1964). On Fisher's bound for asymptotic variances, *Ann. Math. Statist.*, **35**, 1545-1552.

[3] Bahadur, R. R. (1967). Rates of convergence of estimates and test statistics, *Ann. Math. Statist.*, **38**, 303-324.

[4] Berk, R. H. (1966). Limiting behaviour of posterior distributions when the model is incorrect, *Ann. Math. Statist.*, **37**, 51-58.

[5] Bickel, P. and Yahav, J. A. (1966). The asymptotic theory of Bayes' solutions, *Technical Report*, Department of Statistics, Stanford University.

[6] Chao, M. T. (1970). The asymptotic behavior of Bayes' estimators, *Ann. Math. Statist.*, **41**, 601-608.

[7] Fu, James C. (1973). On a theorem of Bahadur on the rate of convergence of point estimators, *Ann. Statist.*, **1**, 745-749.

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

[9] Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters, *Ann. Math. Statist.*, **27**, 887-906.

[10] LeCam, L. (1953). On some asymptotic properties of maximum likelihood and related Bayes estimates, *University of California Publications in Statistics.*

[11] LeCam, L. (1955). On the asymptotic theory of estimation and testing hypotheses, *Proc. Third Berkeley Symp. Math. Statist. Prob.*, **1**, 129-156.

[12] Mann, H. B. and Wald, A. (1943). On stochastic limit and order relationships, *Ann. Math. Statist.*, **3**, 217-226.

[13] Olech, C. (1965). A note concerning set-valued measurable functions, *Bulletin De L'academiè Polonaise Des Sciences Serie des Sciences Math., Astr. et Phys.*, No. 4, 317-322.

[14] Wald, A. (1949). Note on the consistency of maximum likelihood estimates, *Ann. Math. Statist.*, **20**, 595-601.

[15] Weiss, L. and Wolfowitz, J. (1966). Generalized maximum likelihood estimators, *Teoriya Vyeroyatnostey*, **11**, 68-98.

[16] Wolfowitz, J. (1953). The method of maximum likelihood and the Wald theory of decision functions, *Indag. Math.*, **15**, 114-119.

[17] Wolfowitz, J. (1965). Asymptotic efficiency of maximum likelihood estimator, *Theory Prob. Appl.*, **10**, 247-260.