ON THE STRONG CONSISTENCY OF APPROXIMATE MAXIMUM LIKELIHOOD ESTIMATORS

MICHAEL D. PERLMAN
UNIVERSITY OF MINNESOTA

1. Introduction and statement of problem

Wald's general conditions for strong consistency of Approximate Maximum Likelihood Estimators (AMLE) [11] have been extended by several authors, notably LeCam [9], Kiefer and Wolfowitz [8], Huber [7], Bahadur [1], and Crawford [4]. Except for mild identifiability and local regularity conditions these papers (except [9]) share two critical global assumptions, global in the sense that they concern the behavior of the Log Likelihood Ratio (LLR) over the entire parameter space Θ (which may be infinite dimensional). Crudely stated these are (a) there exists a "suitable compactification" $\bar{\Theta}$ of Θ (see [1], p. 320) to which the LLR may be extended in a continuous manner without altering the value of its supremum, and (b) the supremum of the LLR is integrable (dominance). Condition (b), however, is not satisfied in many common problems, especially multiparametric ones, where AMLE are known to be consistent. Kiefer and Wolfowitz and later Berk [3] suggested a method which seemed to overcome this difficulty in special cases, namely: consider the observations pairwise, or in groups of k. In the more general context of "maximum w" estimation described below, however, this method fails (see Example 2). Noticing this, Huber proposed that the LLR be divided by a function $b(\theta)$ such that this normalized LLR satisfies (a) and (b).

In Section 2 of this paper we show that under an extended global dominance assumption the method of Kiefer, Wolfowitz, and Berk is precisely the correct one. This idea is then extended to include Huber's modification. In Section 5 we show that (generalized versions of) LeCam's conditions are equivalent to those based on dominance.

The methods described above have several drawbacks, however: they require determination of a suitable group size k, normalizing functions $b(\theta)$, and a compactification $\bar{\Theta}$. As demonstrated by several examples below, these are not always naturally occurring quantities and may be difficult to determine. In Section 3 we introduce a new condition for strong consistency, based on a global uniformity assumption rather than dominance, which seems to present a more natural and straightforward method for determining strong consistency.

This research was supported by the National Science Foundation under Grant No. GP-9593.

It has the advantage that it is intrinsic, that is, it does not require searching for quantities k, $b(\theta)$, or $\bar{\Theta}$ which are not specified in the original problem.

Necessary and sufficient conditions for strong consistency are discussed in Section 4.

Throughout this paper we treat the problem of strong consistency of AMLE in the following generalized context.

Let $\mathscr{P}$ be a set of distinct probability distributions on a measurable space $(\mathscr{X}, \mathscr{A})$. Let $\theta = \theta(P)$ be a mapping of $\mathscr{P}$ onto a Hausdorff topological space Θ which satisfies the first axiom of countability. Let $X_1, X_2, \cdots$ be a sequence of independent, identically distributed (i.i.d.) random variables assuming values in $\mathscr{X}$, each distributed according to P_0, and let $\theta_0 = \theta(P_0)$. The symbol P_0 is also used to denote the product probability measure on the infinite product space of all sequences $(x_1, x_2, \cdots)$, and ${}_*P_0$ denotes the induced inner measure on this space. Let $w(x, \theta)$ be a real-valued function defined on $\mathscr{X} \times \Theta$ such that for each fixed θ, $w(\cdot, \theta)$ is measurable, and for $n = 1, 2, \cdots$, let

$$w_n(\theta) = w_n(x_1, \cdots, x_n, \theta) \equiv \frac{1}{n} \sum_{i=1}^{n} w(x_i, \theta). \tag{1.1}$$

(For any other function $y(x, \theta)$, $y_n(\theta)$ is defined in a similar manner.) In this paper we discuss the strong consistency of estimators which are based on maximizing $w_n(\theta)$.

Let $\mathscr{S}_1$ denote the class of all estimating sequences $\{T_n\} = \{T_n(x_1, \cdots, x_n)\}$ (T_n is a Θ-valued function and is *not necessarily measurable*) such that for all P_0 in $\mathscr{P}$,

$${}_*P_0[\sup_{\Theta} w_n(\theta) = w_n(T_n) \text{ a.a. } n] = 1 \tag{1.2}$$

(all suprema in this paper are taken with respect to θ over the indicated subset), where, if $\{A_n\}$ is any sequence of sets, $\{A_n \text{ a.a. } n\}$ is the set

$$\liminf_{n \to \infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k \geq n} A_k. \tag{1.3}$$

If $\{T_n\}$ is in $\mathscr{S}_1$, we call it a *Maximum w Estimator* (MWE). Since $\mathscr{S}_1$ may be empty (the supremum may not be attained), we shall mainly consider the larger class $\mathscr{S}_2$ consisting of all estimating sequences such that for all P_0 in $\mathscr{P}$

$${}_*P_0[H(\sup_{\Theta} w_n(\theta), w_n(T_n)) \to 0] = 1, \tag{1.4}$$

where

$$H(a, b) \equiv \begin{cases} a - b & \text{if } a < \infty \\ b^{-1} & \text{if } a = \infty \text{ and } b > 0 \\ 1 & \text{if } a = \infty \text{ and } b \leqq 0. \end{cases} \tag{1.5}$$

If $\{T_n\}$ is in $\mathscr{S}_2$, it is called an *Approximate Maximum w Estimator* (AMWE).

EXAMPLE 1. Suppose that $\theta(P)$ is one-to-one and that each P has a density $f(x, \theta)$ with respect to some measure μ. If $w(x, \theta) = \log f(x, \theta)$, then $\mathscr{S}_2$ contains all AMLE (in the sense of Wald [11], p. 600, Theorem 2).

EXAMPLE 2. Let $\mathscr{P}$ denote the set of all distributions on $(-\infty, \infty)$ which possess a unique population median. Let $\Theta = (-\infty, \infty)$ and $\theta(P) =$ median of P. If $w(x, \theta) = -|x - \theta|$, then $\mathscr{S}_1$ contains all sample medians (recall that the sample median may not be uniquely determined).

For each θ_0 in Θ let $\{V_r\} = \{V_r(\theta_0)\}$ be a decreasing sequence of neighborhoods of θ_0 which form a base for the neighborhood system at θ_0 (so $\cap V_r = \{\theta_0\}$), and let $\Omega_r = \Omega_r(\theta_0) = \Theta - V_r(\theta_0)$. Then an estimating sequence $\{T_n\}$ is *strongly consistent* if and only if for all P_0 in $\mathscr{P}$ and $r \geqq 1$,

$$_*P_0[T_n \text{ in } V_r(\theta_0) \text{ a.a. } n] = 1. \tag{1.6}$$

Note that if this is satisfied for one such sequence of neighborhoods $\{V_r\}$, it must be satisfied for any other such sequence.

A convenient starting point for this problem is the following obvious fact:

LEMMA 1.1. *A sufficient condition for the strong consistency of every* AMWE *is that*

$$_*P_0[\limsup_{n\to\infty} \sup_{\Omega_r} u_n(\theta; \theta_0) < 0] = 1 \tag{1.7}$$

for every P_0 in $\mathscr{P}$ and $r \geqq 1$, where

$$\begin{aligned} u(x, \theta) &\equiv u(x, \theta; \theta_0) \equiv w(x, \theta) - w(x, \theta_0), \\ u_n(\theta) &\equiv u_n(\theta; \theta_0) \equiv w_n(\theta) - w_n(\theta_0). \end{aligned} \tag{1.8}$$

(Under some additional assumptions (1.7) is also a necessary condition, see Section 4.)

The earlier papers [1] (p. 320), [4], [7] (p. 222), [8] (p. 890), [9] (pp. 302–304), and [11] all present conditions which imply Conditions 1 or 2 below, and therefore imply (1.7). (See discussion preceding Theorem 2.3 and see Section 5.) If Θ (hence Ω_r) is compact, Theorem 2.4 below is applicable. In this paper we are mainly concerned with the more interesting and difficult situation where Θ is not compact. In this case earlier papers (except LeCam [9]) assume that a "suitable compactification" of Θ exists (see [1], p. 320). Such a compactification is not always apparent if it exists; the one-point compactification is often unsuitable (see Example 4 and subsequent discussion). In this paper, as in LeCam ([9] pp. 302–304), we attempt to avoid the need to extend the parameter space. The lim sup in (1.7) is studied directly, first with no assumptions on Ω_r (Section 2) and then assuming Ω_r is σ-compact (Section 3).

2. Conditions for strong consistency of AMWE based on dominance and semidominance

Let Γ be a subset of Θ (so Γ is first countable and Hausdorff) and let $y(x, \theta)$ be a real-valued function, defined on $\mathscr{X} \times \Gamma$, which is measurable in x for each fixed θ. Let the sequence of $\mathscr{X}$-valued i.i.d. random variables $X_1, X_2, \cdots$

have probability distribution P (which need not be in $\mathscr{P}$). (Later we shall take $\Gamma = \Omega_r$, $y(x, \theta) = u(x, \theta; \theta_0)$, and $P = P_0$).

DEFINITION 1. *The function $y(x, \theta)$ is dominated (dominated by 0) on Γ with respect to P if there is a positive integer k and a real valued function $s(x_1, \cdots, x_k)$ on $\mathscr{X} \times \cdots \times \mathscr{X}$, measurable with respect to the product σ-field $\mathscr{A} \times \cdots \times \mathscr{A}$, such that*

(i) $\sup_\Gamma y_k(\theta) \leqq s(x_1, \cdots, x_k)$ *for all $x_1, \cdots, x_k$ in a set of probability one, and*
(ii) $Es(X_1, \cdots, X_k) < \infty$ (< 0).

(The subscript k will be used exclusively to refer to this definition.)

REMARK. Note that if $\sup_\Gamma y_k(\theta)$ is measurable, it can be used in place of $s(x_1, \cdots, x_k)$. In any case, note that s can be chosen to be a symmetric function of $x_1, \cdots, x_k$, for we may replace s by

$$s^*(x_1, \cdots, x_k) = \frac{1}{k!} \sum s(x_{i(1)}, \cdots, x_{i(k)}) \tag{2.1}$$

where the sum is taken over all permutations of $(1, \cdots, k)$. Also, s can be chosen such that $Es(X_1, \cdots, X_k) > -\infty$ for, if not, replace s by $\max(s, M)$ for any number M (or any M such that $E \max(s, M) < 0$).

DEFINITION 2. *$y(x, \theta)$ is semidominated (semidominated by 0) on Γ with respect to P if there exists a function $b(\theta)$ defined on Γ, $0 < b(\theta) < \infty$, such that $y(x, \theta)/b(\theta)$ is dominated (dominated by 0) on Γ with respect to P and* $\inf_\Gamma b(\theta) > 0$.

REMARK. Note that if $b_1(\theta)$, with $0 < b_1(\theta) < \infty$, is such that $y(x, \theta)/b_1(\theta)$ is dominated by 0 on Γ, it does not necessarily follow that $y(x, \theta)$ is semidominated by 0 on Γ, since replacing $b_1(\theta)$ by $b(\theta) = \max(b_1(\theta), a)$ with $a > 0$ will not necessarily preserve dominance by 0.

We first investigate some implications among Definitions 1, 2 and the condition (recall (1.7))

$$_*P[\limsup_{n \to \infty} \sup_\Gamma y_n(\theta) < 0] = 1. \tag{2.2}$$

THEOREM 2.1. *If y is dominated (or semidominated) by 0 on Γ then (2.2) holds.*

PROOF. We show that dominance by 0 implies (2.2) using an idea of Berk [3]. For any $n \geqq k$, let $\alpha = \{\alpha_1, \cdots, \alpha_k\}$ denote a selection of k indices from $\{1, 2, \cdots, n\}$. Then

$$y_n(\theta) = \binom{n}{k}^{-1} \sum_\alpha [k^{-1} \sum_{i \in \alpha} y(x_i, \theta)] \tag{2.3}$$

so that for all $x_1, \cdots, x_n$ in a set of probability one

$$\sup_\Gamma y_n(\theta) \leqq \binom{n}{k}^{-1} \sum_\alpha s(x_{\alpha_1}, \cdots, x_{\alpha_k}) \equiv S_{n,k}. \tag{2.4}$$

(We choose s to be a symmetric function of $x_1, \cdots, x_k$.) Berk ([3], pp. 55–56) shows that $\{S_{n,k}\}_{n=k}^\infty$ forms a reverse martingale sequence and $S_{n,k} \to Es < 0$ almost surely as $n \to \infty$, which implies (2.2).

Finally we show that semidominance by 0 implies (2.2). If y is semidominated by 0 on Γ, there is a function $b(\theta)$ such that, applying the above argument,

$$(2.5)\qquad {}_*P[\limsup_{n\to\infty}\sup_{\Gamma}\,(y_n(\theta)/b(\theta)) < 0] = 1.$$

Since $\inf_\Gamma b(\theta) > 0$, this implies (2.2). *Q.E.D.*

EXAMPLE 3. $X_1, X_2, \cdots$ are i.i.d. random variables, each with the normal distribution $N(-1, 1)$. Take $\Gamma = [1, \infty)$ and $y(x, \theta) = \theta x$, so $y_n(\theta) = \theta\bar{x}_n$. Since $\bar{X}_n \to -1$ a.s., $\sup_\Gamma y_n(\theta) \to -1$ a.s. so (2.2) is satisfied. However, for all n

$$(2.6)\qquad P[\sup_{\Gamma} y_n(\theta) = \infty] = P[\bar{X}_n > 0] > 0,$$

so y is not dominated on Γ. Choosing $b(\theta) = \theta$ we see that y is semidominated by 0 on Γ. Thus neither (2.2) nor semidominance by 0 necessarily implies dominance by 0.

A partial converse to Theorem 2.1 is presented in Theorem 2.2 (ii). Several preliminary results are needed.

LEMMA 2.1. *If X and Y are independent real valued random variables, then $E(X + Y)^+ < \infty \Rightarrow EX^+ < \infty$ and $EY^+ < \infty$.*

PROOF. Since $E(X + Y)^+ = E\{E[(X + Y)^+ \mid Y]\}$ it follows that $E(X + y)^+ < \infty$ for almost all y. But $X^+ \leqq (X + y)^+ + |y|$ so $EX^+ < \infty$, and similarly $EY^+ < \infty$. *Q.E.D.*

LEMMA 2.2. *If y is dominated or semidominated on Γ then for every θ' in Γ, $E[y_1(\theta')]^+ < \infty$. Thus for every θ' and n $Ey_n(\theta') = Ey_1(\theta')$ is well defined (possibly $= -\infty$) and $\sup_\Gamma Ey_1(\theta) < \infty$.*

PROOF. If y is dominated on Γ, Definition 1 implies that

$$(2.7)\qquad y_k(\theta') \leqq \sup_{\Gamma} y_k(\theta) \leqq s(x_1, \cdots, x_k)$$

so $E[y_k(\theta')]^+ \leqq Es^+ < \infty$. The result then follows from Lemma 2.1 (the semidominated case is treated similarly). *Q.E.D.*

LEMMA 2.3. *Suppose that for every θ' in Γ, $Ey_1(\theta')$ is well defined (possibly $\pm\infty$). Then*

$$(2.8)\qquad {}_*P[\sup_{\Gamma} Ey_1(\theta) \leqq \liminf_{n\to\infty}\sup_{\Gamma} y_n(\theta)] = 1$$

PROOF. For each θ' in Γ and all n, $y_n(\theta') \leqq \sup_\Gamma y_n(\theta)$. Letting $n \to \infty$ the result follows from the Strong Law of Large Numbers (SLLN).

REMARK. Lemma 2.3 implies that

$$(2.9)\qquad {}_*P[\sup_{\Gamma} Ey_1(\theta) \leqq \limsup_{n\to\infty}\sup_{\Gamma} y_n(\theta)] = 1,$$

and that

$$(2.10)\qquad {}_*P[\sup_{\Gamma} y_n(\theta) \to \sup_{\Gamma} Ey_1(\theta)] = 1$$

if and only if equality holds in (2.9), in fact, if and only if

$$(2.11)\qquad {}_*P[\sup_{\Gamma} Ey_1(\theta) \geqq \limsup_{n\to\infty}\sup_{\Gamma} y_n(\theta)] = 1.$$

LEMMA 2.4. (i) *If* $Ey_1(\theta') > -\infty$ *for some* θ' *in* Γ, *then* $E[\sup_\Gamma y_n(\theta)]^- < \infty$ *for all* n *such that* $\sup_\Gamma y_n(\theta)$ *is measurable, in which case* $E \sup_\Gamma y_n(\theta)$ *is well defined (possibly* $= +\infty$*).*

(ii) *If* y *is dominated on* Γ, *then* $E[\sup_\Gamma y_n(\theta)]^+ < \infty$ *for all* $n \geqq k$ *such that* $\sup_\Gamma y_n(\theta)$ *is measurable, in which case* $E \sup_\Gamma y_n(\theta)$ *is well defined (possibly* $= -\infty$*).*

PROOF. Part (i) is obvious; part (ii) follows from

$$\sup_\Gamma y_n(\theta) \leqq \binom{n}{k}^{-1} \sum_\alpha \sup_\Gamma \Big[k^{-1} \sum_{i\in\alpha} y(x_i, \theta)\Big] \equiv Y_{n,k}(\Gamma), \tag{2.12}$$

(see the proof of Theorem 2.1). *Q.E.D.*

The result need not hold if only semidominance is assumed.

LEMMA 2.5. *If* $Ey_1(\theta')$ *is well defined for all* θ' *in* Γ *and if* $\sup_\Gamma y_n(\theta)$ *is measurable and* $E \sup_\Gamma y_n(\theta)$ *is well defined for almost all* n *then*

$$\sup_\Gamma Ey_1(\theta) \leqq \downarrow \lim_{n\to\infty} E \sup_\Gamma y_n(\theta). \tag{2.13}$$

PROOF. From (2.12) with n, k replaced by $n + 1$, n,

$$\sup_\Gamma Ey_1(\theta) = \sup_\Gamma Ey_{n+1}(\theta) \leqq E \sup_\Gamma y_{n+1}(\theta) \leqq E \sup_\Gamma y_n(\theta), \tag{2.14}$$

which implies (2.13). *Q.E.D.*

Under the hypothesis of Lemma 2.5, (2.9) is valid with ${}_*P$ replaced by P, and should be compared with (2.13). The relationship between (2.9) and (2.13) is now clarified.

THEOREM 2.2. (i) *If* $\sup_\Gamma y_n(\theta)$ *is measurable and* $E \sup_\Gamma y_n(\theta)$ *well defined for almost all* n *then*

$$P[\limsup_{n\to\infty} \sup_\Gamma y_n(\theta) \leqq \downarrow \lim_{n\to\infty} E \sup_\Gamma y_n(\theta)] = 1. \tag{2.15}$$

(ii) *If* y *is dominated on* Γ *and* $\sup_\Gamma y_n(\theta)$ *is measurable for almost all* n *then*

$$P[\limsup_{n\to\infty} \sup_\Gamma y_n(\theta) = \downarrow \lim_{n\to\infty} E \sup_\Gamma y_n(\theta)] = 1. \tag{2.16}$$

Therefore under this measurability assumption, y *is dominated by* 0 *on* Γ *if and only if* (2.2) *is satisfied and* y *is dominated on* Γ.

PROOF. For part (i) assume that the right side of the inequality is $< \infty$ (in which case y is dominated on Γ; otherwise (2.15) is trivial). Referring to (2.12), for any q such that $E \sup_\Gamma y_q(\theta) < \infty$, $\{Y_{n,q}\}$ is a reverse martingale, $n = 1, 2, \cdots$ and

$$Y_{n,q} \to E \sup_\Gamma y_q(\theta) \text{ a.s.} \qquad \text{as } n \to \infty, \tag{2.17}$$

so $\limsup_{n\to\infty} \sup_\Gamma y_n(\theta) \leqq E \sup_\Gamma y_q(\theta)$ a.s. Letting $q \to \infty$ we obtain (2.15).

For part (ii) we use (2.12) and (2.17) to apply a well-known extension of the Fatou-Lebesgue theorem ([10], p. 162), obtaining

$$\downarrow \lim_{n\to\infty} E \sup_\Gamma y_n(\theta) \leqq E[\limsup_{n\to\infty} \sup_\Gamma y_n(\theta)]. \tag{2.18}$$

Combining this with (2.15) yields (2.16). *Q.E.D.*

REMARK. By the Hewitt-Savage zero-one law under the measurability assumption of Theorem 2.2 (ii), $\lim \sup_{n\to\infty} \sup_\Gamma y_n(\theta)$ is a constant a.s. (also lim inf), whether or not dominance holds.

Returning to the problem of strong consistency, we state the global

DOMINANCE AND MEASURABILITY ASSUMPTION $\mathscr{D}$. *For every P_0 in $\mathscr{P}$ and $r \geqq 1$, $u(x, \theta; \theta_0)$ is dominated on Ω_r and $\sup_{\Omega_r} u_n(\theta: \theta_0)$ is measurable for almost all n.*

Condition 1. For every P_0 and $r \geqq 1$, $u(x, \theta; \theta_0)$ is dominated by 0 on Ω_r. Then Theorem 2.2 (ii) implies: *if $\mathscr{D}$ holds then Condition* 1 *is necessary and sufficient for* (1.7), *and is thus sufficient for the strong consistency of all* AMWE.

The integer k needed to verify dominance of $u(x, \theta; \theta_0)$ may be $\geqq 2$, especially in multiparameter cases, as pointed out by Kiefer and Wolfowitz ([8], p. 904), Huber, and others. For example, in the context of Example 1, let $\theta = (\mu, \sigma)$ and consider AMLE of (μ, σ) in the location and scale family of densities $\sigma^{-1}f(\sigma^{-1}(x - \mu))$, f specified. If $f(0) > 0$ then $E \sup_{\Omega_r} u_1(\theta) = +\infty$ for r sufficiently large so if $u(x, \theta; \theta_0)$ is dominated, k must be $\geqq 2$. For example in the normal case, that is, $f(z) = (2\pi)^{-1/2} \exp\{-\frac{1}{2}z^2\}$, $E \sup_\Theta u_2(\theta) < \infty$.

Theorem 2.2 (ii) and Lemma 1.1 also imply that *a sufficient condition for the strong consistency of all* AMWE *is*

Condition 2. For every P_0 in $\mathscr{P}$ and $r \geqq 1$, $u(x, \theta; \theta_0)$ is semidominated by 0 on Ω_r.

The need for considering semidominance is illustrated by

EXAMPLE 2 (*continued*). Consistent estimation of the population median (see Huber [7], p. 223). Assuming without loss of generality that $\theta_0 = 0$, we have $u(x, \theta) = |x| - |x - \theta|$. With

$$\Omega_r = (-\infty, -r^{-1}] \cup [r^{-1}, \infty), \tag{2.19}$$

it can be shown that $u(x, \theta)$ is not dominated on Ω_r: for all r and n

$$\sup_{\Omega_r} u_n(\theta) \geqq h_r(x_1, \cdots, x_n) \equiv \begin{cases} n^{-1} \min |x_i| & \text{if } x_i \geqq r^{-1}, i = 1, \cdots, n, \\ -r^{-1} & \text{otherwise.} \end{cases} \tag{2.20}$$

Let $X_1, X_2, \cdots$ be i.i.d. real random variables, each distributed on $(-\infty, \infty)$ symmetrically about 0 and each $|X_i|$ having a cumulative distribution function (c.d.f.)

$$F(x) = \begin{cases} \dfrac{\log(1 + x)}{1 + \log(1 + x)} & \text{if } x \geqq 0, \\ 0 & \text{if } x < 0. \end{cases} \tag{2.21}$$

Since sgn X_i is independent of $|X_i|$,

$$\begin{aligned} E\{\min |X_i| \mid X_i \geqq r^{-1}, i = 1, \cdots, n\} &= E\{\min |X_i| \mid |X_i| \geqq r^{-1}, i = 1, \cdots, n\} \\ &= E\{\min |X_i| \mid \min |X_i| \geqq r^{-1}\} \\ &= +\infty, \end{aligned} \tag{2.22}$$

so $Eh_r(X_1, \cdots, X_n) = +\infty$. Thus $u(x, \theta)$ is not dominated on Ω_r. Setting $b(\theta) = |\theta|$ (or $|\theta - \theta_0|$ if $\theta_0 \neq 0$), however, Huber's arguments can be used to show that Condition 2 holds (see also our discussion preceding Theorem 2.3 below). (Huber takes $b(\theta) = |\theta| + 1$ but this is not necessary since $|\theta|$ is bounded away from 0 on Ω_r.) Thus all AMWE are strongly consistent even though dominance fails in this example.

It was stated above that the conditions of [1], [4], [8], [11] in fact imply Condition 1 and the conditions of Huber [7] imply Condition 2. These authors essentially assume that a compactification $\bar{\Theta}$ of Θ and an extension of $u(x, \theta)$ to $\bar{\Theta}$ exist where every $\theta' \neq \theta_0$ in $\bar{\Theta}$ possesses a sufficiently small neighborhood $V(\theta')$ such that $u(x, \theta; \theta_0)$ is dominated (or semidominated) by 0 on $V(\theta')$. Then using the compactness of $\bar{\Omega}_r$ and a by now well known argument (see Theorem 2.4 below) they deduce that (1.7) holds. The essential idea is, of course, to note that (1.7) holds with Ω_r replaced by $V(\theta')$, obtain a finite subcover $V(\theta_1), \cdots, V(\theta_h)$ of Ω_r, and conclude that (1.7) holds for Ω_r. Part (ii) of Theorem 2.3 isolates the key feature of the above method and shows that the conditions of the papers listed above do in fact imply Conditions 1 or 2.

THEOREM 2.3. (i) *If y is dominated on Γ and $\Gamma' \subset \Gamma$, then y is dominated on Γ'. This remains true if "dominated" is replaced by "semidominated," "dominated by 0," or "semidominated by 0."*

(ii) *If y is dominated on Γ_i, $i = 1, \cdots, h$, then y is dominated on $\cup \Gamma_i$. This remains true if "dominated" is replaced as above.*

PROOF. Part (i) is trivial. We prove part (ii) for $h = 2$ with "dominated" replaced by "dominated by 0," as this is the more difficult case. Using part (i) if necessary, we may assume that Γ_1 and Γ_2 are disjoint. Let $\Gamma = \Gamma_1 \cup \Gamma_2$ and let k_i and $s_i(x_1, \cdots, x_{k_i})$ be as in Definition 1, $i = 1, 2$. Let $m = k_1 k_2$,

$$S(i) = \frac{1}{k(i)} \sum_{j=0}^{k(i)-1} s_i(x_{jk_i+1}, \cdots, x_{(j+1)k_i}) \tag{2.23}$$

where $k(1) = k_2$, $k(2) = k_1$, and

$$y^*(\theta) = \begin{cases} S(1) & \text{if } \theta \in \Gamma_1 \\ S(2) & \text{if } \theta \in \Gamma_2. \end{cases} \tag{2.24}$$

Thus $y_m(\theta) \leqq y^*(\theta)$ and $y_{mn}(\theta) \leqq y_n^*(\theta)$. Since $ES(i) = Es_i < 0$,

$$\sup_{\Gamma} y_n^*(\theta) = \max\left(S_n(1), S_n(2)\right) \xrightarrow{\text{a.s.}} \max\left(Es_1, Es_2\right) < 0. \tag{2.25}$$

Applying Theorem 2.2 (ii) with y replaced by y^*, this implies that $E \sup_\Gamma y_p^*(\theta) < 0$ for some p. Setting $k = mp$ and $s(x_1, \cdots, x_k) = \sup_\Gamma y_p^*(\theta)$, we conclude that y is dominated by 0 on Γ. Note that we cannot set $k = m$ and $s = \sup_\Gamma y^*(\theta) = \max\left(S(1), S(2)\right)$, since $ES(i) < 0$ need not imply that $E \max\left(S(1), S(2)\right) < 0$. *Q.E.D.*

The preceding result is now used to show that, under local regularity assumptions only, (2.10) holds if Γ is compact. In Section 3 this is extended to σ-compact sets by imposing a uniformity assumption. We say $y(x, \theta)$ is *locally dominated* on Γ if every θ' in Γ possesses a neighborhood on which $y(x, \theta)$ is dominated. Recall that Γ is a first countable Hausdorff space.

THEOREM 2.4. *Let Γ be a compact subspace of Θ such that $y(x, \theta)$ is locally dominated on Γ. For each θ' in Γ suppose that $y(x, \cdot)$ is upper semicontinuous at θ' except for x in a P-null set possibly depending on θ' and that for some decreasing sequence $\{G_m\} = \{G_m(\theta')\}$ of subsets of Θ forming a base for the neighborhood system at θ', $\sup_{G_m} y_n(\theta)$ is measurable for almost all m and almost all n. Then* (2.11) *holds, implying* (2.10). *Also $Ey_1(\theta)$ is an upper semicontinuous function of θ.*

PROOF. Local dominance and compactness imply y is dominated on Γ by Theorem 2.3. Let k be an integer such that $E \sup_\Gamma y_k(\theta) < \infty$. Since $\bigcap G_m = \{\theta'\}$, upper semicontinuity of $y_k(\theta)$ implies $\sup_{G_m} y_k(\theta) \downarrow y_k(\theta')$ as $m \to \infty$ for almost all $x_1, \cdots, x_k$. Then by Lemmas 2.2 and 2.4 and the monotone convergence theorem, for each θ' in Γ,

$$E \sup_{G_m} y_k(\theta) \downarrow E y_1(\theta') \qquad \text{as } m \to \infty. \tag{2.26}$$

Thus, given $\delta > 0$, there is an integer $\mu = \mu(\theta')$ such that

$$E \sup_{G_\mu} y_k(\theta) \leqq \sup_\Gamma E y_1(\theta) + \delta. \tag{2.27}$$

There is a finite subset $\{\theta_1', \cdots, \theta_h'\}$ of Γ such that $F_1, \cdots, F_h$ covers Γ, where $F_i = G_{\mu(\theta_i')}$. Let

$$y^*(x, \theta) = y(x, \theta) - \sup_\Gamma E y_1(\theta) - 2\delta, \tag{2.28}$$

so y^* is dominated by 0 on each F_i. Then Theorem 2.3 implies that y^* is dominated by 0 on Γ. Applying Theorem 2.1 to y^* and then letting $\delta \to 0$ yields (2.11) and hence (2.10). Finally, since

$$E y_1(\theta') \leqq \sup_{G_m} E y_1(\theta) \leqq E \sup_{G_m} y_k(\theta), \tag{2.29}$$

(2.26) implies that $Ey_1(\theta')$ is upper semicontinuous at each θ'. *Q.E.D.*

The measurability assumption in Theorem 2.4 is satisfied, for example, if Γ is a separable space, each G_m is open, and $y(x, \cdot)$ is lower semicontinuous on Γ for almost all x. (For common spaces Γ, other criteria for measurability may be more useful, such as right continuity if θ is a real-valued parameter.) In particular, if Γ is separable and $\{y(x, \cdot)\}$ is an equicontinuous family of functions on Γ (except for x in a P-null set) then all the assumptions of Theorem 2.4 are satisfied.

We conclude this section with several remarks concerning the determination of a normalizing function $b(\theta)$ such that $y(x, \theta)$ is semidominated by 0. If $y(x, \theta)$ is not itself dominated by 0 on Γ no general conditions guaranteeing the existence of such a function $b(\theta)$ are known to the author and if such a function does exist no general formula for b(θ) is known. *Necessary* conditions for the existence of $b(\theta)$ are that $\sup_\Gamma Ey_1(\theta) < 0$ and $b(\theta) \leqq \delta|Ey_1(\theta)|$ for some $\delta > 0$ (by (2.2) and Lemma 2.3). This suggests choosing $b(\theta) = |Ey_1(\theta)|$ (as in Example 3) or more generally choosing $b(\theta)$ such that $b(\theta)/|Ey_1(\theta)|$ is bounded away from 0 and ∞ (as in Example 2, see the continuation of this example in Section 3). This "rule of thumb" seems to be satisfactory in most statistical applications but, at the level of generality of this paper, it is not universally valid. To see this we present an example where y is not dominated by 0 but is semidominated by 0, and where $b(\theta)$ cannot be $|Ey_1(\theta)|$. Let $\Gamma = [1, \infty)$ and let W be a random variable assuming the values 2 and -2 with probabilities 1/4 and 3/4 respectively. Let U be a stochastic process with parameter space Γ, U independent of W, such that for each θ, $U(\theta)$ is uniformly distributed on the interval $(-2\theta^2, -\theta)$ and $\{U(\theta): \theta \text{ in } \Gamma\}$ is a set of mutually independent random variables. Let $V(\theta) = \theta W$ and let $X = \{X(\theta)\} \equiv \{U(\theta) + V(\theta)\}$. Let $X_1, X_2, \cdots$ be a sequence of i.i.d. stochastic processes, each having the same distribution as X. Let $\mathscr{X}$ be the set of all real-valued functions on Γ, and for $(x, \theta) \in \mathscr{X} \times \Gamma$ set $y(x, \theta) = x(\theta)$. Then $P[\sup_\Gamma y_k(\theta) = \infty] > 0$ for every k, so y is not dominated. Setting $b(\theta) = \theta$ we have $y(X, \theta)/\theta = W + U(\theta)/\theta$ and $\sup_\Gamma [y_1(\theta)/\theta] \leqq W - 1$, so y is semidominated by 0. However, $2|Ey_1(\theta)| = 3\theta + \theta^2$ and for every k, we have $\sup_\Gamma[y_k(\theta)/(3\theta + \theta^2)] \geqq 0$ with probability 1.

This example shows that the choice of $b(\theta)$, if it exists, is very delicate: if $b(\theta)$ is too small, $y(x, \theta)/b(\theta)$ may not even be dominated, while if $b(\theta)$ is too large $y(x, \theta)/b(\theta)$ may be dominated but not dominated by 0. Also, there are situations where no such $b(\theta)$ exists: let $\Gamma = [1, \infty)$ and let X be a stochastic process with parameter space Γ such that $\{X(\theta): \theta \in \Gamma\}$ are mutually independent and each $X(\theta)$ is uniformly distributed on $(-1, -1/\theta)$. With $y(x, \theta)$ as defined in the preceding paragraph, $y(x, \theta)$ is dominated on Γ but not semidominated by 0 on Γ. Note that if $b(\theta) = 1/\theta$, $y(x, \theta)/b(\theta)$ is dominated by 0 but $\inf b(\theta) = 0$. If the example is changed slightly so that $X(\theta)$ is uniformly distributed on $(-1, 0)$ for each θ then there is no function $b(\theta) > 0$ such that $y(x, \theta)/b(\theta)$ is dominated by 0, even if we do not require that $\inf b(\theta) > 0$.

It should be clear by now that due to the attempt to achieve wide generality Conditions 1 and 2, although extending (perhaps "consolidating" is a better term) earlier conditions based on compactification of Θ, do not eliminate the usefulness of these conditions in actually verifying strong consistency. As seen by Example 2, direct determination of a suitable integer k to verify that $u(x, \theta; \theta_0)$ is dominated by 0 or semidominated by 0 on Ω_r may not be feasible. Also, no general method is known for determining suitable normalizing functions $b(\theta)$. Conditions 1 and 2 and earlier conditions do share a common drawback, however: these conditions all require global dominance or semidominance over the possibly noncompact set Ω_r which, it is felt, is not a natural restriction. We now amplify these remarks and introduce a new method based on a global uniformity assumption which requires neither dominance on Ω_r nor compactification.

3. A new condition for strong consistency of AMWE based on uniformity

Throughout this section and the next it is assumed that $E_0 u_1(\theta; \theta_0)$ is well defined (possibly infinite) for every θ in Θ and every P_0 in $\mathscr{P}$, where E_0 denotes expectation under P_0, and therefore that $Ey_1(\theta)$ is well defined for every θ in Γ. By Lemma 2.3 a *necessary condition for* (1.7), *therefore weaker than Conditions* 1 *or* 2, *is*

Condition 3. For every P_0 in $\mathscr{P}$ and $r \geqq 1$

$$(3.1) \qquad {}_*P_0[\sup_{\Omega_r} u_n(\theta; \theta_0) < \infty \text{ a.a. } n] = 1$$

and

$$(3.2) \qquad \sup_{\Omega_r} E_0 u_1(\theta; \theta_0) < 0.$$

(Notice that $\sup E_0 u_1$ is obviously easier to compute than $E_0 \sup u_n$.)

This condition has a natural interpretation. In Example 1, for instance, $u_1(\theta; \theta_0)$ is the log likelihood ratio and (3.2) is simply an identifiability condition stating that the topology of the parametrization is suited to the underlying probabilistic model. More precisely it states that if the density $f(x, \theta)$ converges to $f(x, \theta_0)$ in terms of Kullback-Leibler distance (information), so that the associated distributions converge, then θ must converge to θ_0. Clearly this is a minimal assumption which must be imposed.

An approach to the problem of strong consistency which seems natural, therefore, is to replace the assumption of dominance or semidominance of $u(x, \theta; \theta_0)$ on Ω_r by an assumption which implies that

$$(3.3) \qquad {}_*P_0[\limsup_{n \to \infty} \sup_{\Omega_r} u_n(\theta; \theta_0) = \sup_{\Omega_r} E_0 u_1(\theta; \theta_0)] = 1$$

for every P_0 and r. (The uniformity assumption is stated after Theorem 3.1.) That is, we are now concerned with verifying (2.10) rather than (2.16). Recall that Theorem 2.4 gave such conditions for a compact set Γ. A useful extension is based on the following result concerning equality of iterated limits, the proof of which is straightforward. When we say a limit exists we allow it to assume an infinite value.

LEMMA 3.1. *Let* $\{\beta(n, m)\}$ *be a double sequence of extended real numbers such that* $\beta(n, m)$ *is increasing in* m. *Suppose that the limit* $\beta(n, \infty) \equiv \uparrow \lim_{m \to \infty} \beta(n, m)$ *satisfies* $-\infty \leqq \beta(n, \infty) < \infty$ *for* a.a. n, *that the limit* $\beta(\infty, m) \equiv \lim_{n \to \infty} \beta(n, m)$ *exists for all* m, *and that* $-\infty \leqq \uparrow \lim_{m \to \infty} \beta(\infty, m) < \infty$. *Then*

$$(3.4) \qquad \lim_{m \to \infty} \beta(\infty, m) = \lim_{n \to \infty} \beta(n, \infty)$$

if and only if

$$(3.5) \qquad \beta(n, m) \to \beta(\infty, m)$$

uniformly in m,

and both (3.4) *and* (3.5) *are implied by*

$$\beta(n, m) \uparrow \beta(n, \infty) \tag{3.6}$$

uniformly in n (a.a. n).

If $-\infty < \lim_{m\to\infty} \beta(\infty, m) < \infty$ *then* (3.4), (3.5), *and* (3.6) *are mutually equivalent.*

THEOREM 3.1. *Suppose that* $\Gamma = \bigcup_{m=1}^{\infty} \Gamma_m$ *where* $\Gamma_m \subset \Gamma_{m+1}$ *for all* m. *Assume that*

$$-\infty \leqq \sup_{\Gamma} Ey_1(\theta) < \infty, \tag{3.7}$$

$${}_*P[\sup_{\Gamma} y_n(\theta) < \infty \text{ a.a. } n] = 1, \tag{3.8}$$

and

$${}_*P[\sup_{\Gamma_m} y_n(\theta) \to \sup_{\Gamma_m} Ey_1(\theta)] = 1 \qquad \textit{for all } m. \tag{3.9}$$

Then ${}_*P[\sup_\Gamma y_n(\theta) \to \sup_\Gamma Ey_1\theta] = 1$, *if and only if the convergence in* (3.9) *is uniform in* m, *that is,*

$${}_*P[\sup_{\Gamma_m} y_n(\theta) \to \sup_{\Gamma_m} Ey_1(\theta) \textit{ uniformly in } m] = 1, \tag{3.10}$$

and both (2.10) *and* (3.10) *are implied by*

$${}_*P[\sup_{\Gamma_m} y_n(\theta) \uparrow \sup_{\Gamma} y_n(\theta) \textit{ uniformly in } n \text{ (a.a. } n)] = 1. \tag{3.11}$$

If $-\infty < \sup_\Gamma Ey_1(\theta) < \infty$, *then* (2.10), (3.10), *and* (3.11) *are mutually equivalent.*

PROOF. Let $\beta(n, m) = \sup_{\Gamma_m} y_n(\theta)$ and apply Lemma 3.1. *Q.E.D.*

Using this theorem we can extend Theorem 2.4 to σ-compact sets.

THEOREM 3.2. *Let* Γ *be a* σ*-compact space, that is,* $\Gamma = \bigcup_{m=1}^{\infty} \Gamma_m$ *where each* Γ_m *is compact. Suppose that for each* θ' *in* Γ *the local dominance, upper semicontinuity, and measurability assumptions of Theorem* 2.4 *are satisfied, and that the conditions of* (3.7) *and* (3.8) *hold. Then all the conclusions of Theorem* 3.1 *hold, that is,* (3.11) $\Rightarrow$ (3.10) $\Leftrightarrow$ (2.10), *and these are mutually equivalent if* $\sup_\Gamma Ey_1(\theta)$ *is finite. Furthermore* $Ey_1(\theta)$ *is upper semicontinuous on* Γ.

PROOF. We can assume that $\Gamma_m \subset \Gamma_{m+1}$. Theorem 2.4 implies (3.9) since Γ_m is compact, so all hypotheses of Theorem 3.1 are satisfied. Also $Ey_1(\theta)$ is upper semicontinuous on each Γ_m, hence on Γ. *Q.E.D.*

REMARK. Suppose the hypotheses of Theorem 3.2 hold and that $\Gamma = \bigcup_{m=1}^{\infty} \Gamma'_m$ is another representation of Γ as an increasing union of compact sets, so that (3.9) holds for $\{\Gamma'_m\}$ as well as $\{\Gamma_m\}$. Since (2.10) does not depend on the decomposition of Γ, (3.10) holds for $\{\Gamma'_m\}$ if and only if it holds for $\{\Gamma_m\}$.

We can now state the global

UNIFORMITY ASSUMPTION $\mathscr{U}$. *For each* P_0 *in* $\mathscr{P}$ *and* $r \geqq 1$, $\Omega_r = \Omega_r(\theta_0)$ *can be expressed as* $\bigcup_{m=1}^{\infty} \Omega_{r,m}$ *where* $\Omega_{r,m} \subset \Omega_{r,m+1}$ *and*

$$(3.12) \qquad {}_*P_0[\sup_{\Omega_{r,m}} y_n(\theta) \to \sup_{\Omega_{r,m}} E_0 u_1(\theta; \theta_0) \text{ uniformly in } m] = 1.$$

Then Theorem 3.1 implies: *if* $\mathscr{U}$ *holds then Condition* 3 *is necessary and sufficient for* (1.7) *and is therefore sufficient for the strong consistency of all* AMWE.

In some cases it may be easier to verify a condition based on (3.11), rather than (3.12) which is based on (3.10). Notice also that the crucial aspect of (3.12) is the *uniformity* of the convergence: if $\Omega_{r,m}$ is compact the convergence itself is guaranteed by local regularity assumptions only, as in Theorem 3.2.

It is felt that this represents a more directly applicable approach to the problem of strong consistency than that contained in the statement following Condition 1. This is illustrated by Example 2 where it was shown earlier that dominance $\mathscr{D}$ fails but where we now show that $\mathscr{U}$ and Condition 3 hold.

EXAMPLE 2 (*continued*). Fix P_0 and r and assume $\theta_0 = 0$. With $\Gamma = \Omega_r$ as defined in Section 2 let

$$(3.13) \qquad \Gamma_m = \Omega_{r,m} = [-m, -r^{-1}] \cup [r^{-1}, m],$$

a compact set. Since $u_n(\theta) \leqq n^{-1} \Sigma_i |X_i|$, (3.1) is obviously true. Since $u(x, \theta)$ is continuous in θ (in fact equicontinuous) and $u_1(\theta) \leqq m$ on Γ_m, the hypotheses of Theorem 2.4 are satisfied so (3.8) holds. Thus to verify $\mathscr{U}$ we must only verify the uniformity of the convergence in (3.12).

Note that $u_n(\theta)$ is a unimodal function with mode at $X[(n+1)/2]$ (the (n + 1)/2th order statistic from a sample of size n) if n is odd, and mode "plateau" on the interval $(X[n/2], X[(n/2)+1])$ if n is even. Thus, if we choose δ such that

$$(3.14) \qquad P_0[-\delta \leqq X_1 \leqq \delta] > \tfrac{1}{2}$$

it follows that

$$(3.15) \qquad P_0[-\delta \leqq \text{mode (plateau) of } u_n(\theta) \leqq \delta \text{ a.a. } n] = 1,$$

which implies the uniformity in (3.11) or (3.12). It remains only to verify (3.2). If $\theta \geqq 0$ ($\theta \leqq 0$ is similar),

$$(3.16) \qquad \begin{aligned} E_0 u_1(\theta) &= 2\left\{\theta P_0[X_1 \geqq \theta] + \int_{0+}^{\theta-} x\, dP_0(x)\right\} - \theta \\ &\leqq \theta\{2P_0[X_1 > 0] - 1\} \\ &\leqq 0. \end{aligned}$$

The first inequality is strict if $P_0[0 < X_1 < \theta] > 0$ and the second is strict if $P_0[X_1 > 0] < 1/2$. Thus $P_0[X_1 > 0] < 1/2$ implies $\sup_{\Omega_r} E_0 u_1(\theta) < 0$. If $P_0[X_1 > 0] = 1/2$ then for all $\theta > 0$, $P_0[0 < X_1 < \theta] > 0$ since otherwise the population median would not be unique. This again implies $\sup_{\Omega_r} E_0 u_1(\theta) < 0$ since this expectation decreases monotonically in $|\theta|$, so (3.2) holds. Incidentally, $|Eu_1(\theta)| \leqq |\theta|$ since $|u_1(\theta)| \leqq |\theta|$, which shows that if $b(\theta) = |\theta|$, then $b(\theta)/|Eu_1(\theta)|$ is bounded away from 0 (and ∞, as seen in Section 2), thus verifying an earlier remark.

Of course we saw earlier that Condition 2, based on semidominance, and Huber's conditions, based on compactification, each apply in this example. The present method has the advantage, however, that we did not have to search for normalizing functions $b(\theta)$, integers k, or suitable compactifications, which seems to be a great advantage in general. Even when dominance and Condition 1 apply the present method retains these advantages, as illustrated by the next example.

EXAMPLE 4. In the context of Example 1, let $\Theta = [0, \infty)$, μ = Lebesgue measure, and

$$(3.17)\qquad f(x, \theta) = (2\pi)^{-1}\left[1 + \frac{a\theta}{1+\theta}\sin(x - \theta)\right] \qquad \text{if } 0 \leqq x \leqq 2\pi.$$

$f(x, \theta) = 0$ otherwise, where $0 < a < 1$ is a constant. We wish to show that all approximate maximum likelihood estimators are strongly consistent, although neither Wald's, Kiefer and Wolfowitz's nor Huber's conditions are satisfied here. Fix θ_0 and r, and let

$$(3.18)\qquad \Omega_r = [0, \theta_0 - r^{-1}] \cup [\theta_0 + r^{-1}, \infty).$$

Since $u(x, \theta) = \log f(x, \theta) - \log f(x, \theta_0) \leqq \log 2(1 + \theta_0)$, (3.1) is satisfied, and in fact, u is dominated on Ω_r. Setting $\Gamma = \Omega_r$ and $\Gamma_m = \Omega_{r,m} = \Gamma \cap [0, m]$, Γ_m is compact, u is dominated on Γ_m, and u is continuous so Theorem 2.4 implies (3.9). To verify (3.12) note that $\log t$ is uniformly continuous on $[1 - a, 1 + a]$, that for any x_i

$$(3.19)\qquad \left|\left[1 + \frac{a\theta}{1+\theta}\sin(x_i - \theta)\right] - [1 + a\sin(x_i - \theta)]\right| \leqq \frac{1}{1+\theta},$$

and that both terms in square brackets lie in $[1 - a, 1 + a]$. Therefore, given any $\delta > 0$ there exists $M > 0$ such that $\theta \geqq M$ implies

$$(3.20)\qquad \left|\frac{1}{n}\log\prod_{i=1}^{n}\left[1 + \frac{a\theta}{1+\theta}\sin(x_i - \theta)\right] - \frac{1}{n}\log\prod_{i=1}^{n}[1 + a\sin(x_i - \theta)]\right| \leqq \delta$$

independently of n and $x_1, x_2, \cdots$. Thus for θ sufficiently large, $u_n(\theta)$ can be approximated arbitrarily closely by a periodic function, uniformly in n and $x_1, x_2, \cdots$, which implies (3.12). Finally (3.2) follows from the information inequality and an easy limiting argument as $\theta \to \infty$, so Condition 3 is satisfied and all the AMLE are strongly consistent. Note that since dominance $\mathscr{D}$ holds this implies Condition 1 holds, but this would have been difficult to demonstrate directly.

In Example 4, conditions (a), (b), (c) of Bahadur ([1], p. 320) are satisfied, but this is not immediately evident. The difficulty arises when trying to find a suitable compactification, for the obvious one-point compactification $[0, \infty]$ is not adequate. We must adjoin to $\Theta = [0, \infty)$ an entire interval of length 2π, say, $I = [-2\pi, 0)$. Any θ in Θ can be uniquely represented as $\theta = 2\pi m + r$ where m is an integer and $-2\pi \leqq r < 0$. Then the topology in $\bar{\Theta} = \Theta \cup I$ must be defined so that if $\{\theta_n\} \subset \Theta$ and $\bar{\theta} \in I$, $\theta_n \to \bar{\theta}$ if and only if $m_n \to \infty$ and $r_n \to \bar{\theta}$.

However, it has been pointed out to the author by Professor Bahadur that if in this example (or more generally in Example 1) we redefine $\theta(P) \equiv P$ and $\Theta \equiv \mathscr{P}$ and consider the topology of weak convergence in $\mathscr{P}$, then there is a natural compactification of $\mathscr{P}$, namely the closure of $\mathscr{P}$ in the set of all measures on $(\mathscr{X}, \mathscr{A})$ with total mass $\leqq 1$. This is in fact the natural parameterization and topology to consider in Example 1 since we are interested primarily in estimating the underlying probability distribution. Using this parameterization Bahadur's conditions (a), (b), (c) are easily verified in Example 4.

A similar situation occurs in the context of Example 1, where $\theta = (\mu, \sigma)$ and $f(x, \theta)$ is the density of the normal distribution $N(\mu, \sigma^2)$ discussed earlier. Here it is again difficult to find a "suitable compactification" of the space Θ = the open half plane $\{\sigma > 0\}$. If, however, the natural parameterization $\theta(P) = P$ and the natural compactification described above are considered, Bahadur's conditions (a) and (c) are readily verified. Condition (b) fails, but as pointed out earlier it can be replaced by the assumption of dominance with $k = 2$.

4. Necessary and sufficient conditions for strong consistency of AMWE

We now add some mild local regularity assumptions to the underlying assumptions introduced in Section 1 and show that in this case our sufficient conditions become necessary as well. Recall that Θ is a first countable Hausdorff space.

LOCAL REGULARITY ASSUMPTIONS $\mathscr{L}$.

(a) Θ *is locally compact, so we can choose* $\{V_r(\theta_0)\}$ *to be a* compact *base for the neighborhood system at each* θ_0;

(b) *for every* P_0 *in* $\mathscr{P}$, $u(x, \theta; \theta_0)$ *is locally dominated on* Θ *with respect to* P_0;

(c) *for every* P_0 *in* $\mathscr{P}$ *and* θ' *in* Θ, $u(x, \theta; \theta_0)$ *is upper semicontinuous at* θ' *except for* x *in a* P_0*-null set possibly depending on* θ';

(d) *for each* θ' *in* Θ *there exists a decreasing sequence of subsets* $\{G_m\} = \{G_m(\theta')\}$ *forming a base for the neighborhood system at* θ' *such that* $\sup_{G_m} u_n(\theta: \theta_0)$ *is measurable for* a.a. m *and* a.a. n;

(e) *for every* P_0 *in* $\mathscr{P}$, $\sup_{V_r} E_0 u_1(\theta; \theta_0) = 0$ *for* a.a. r.

Assumptions (b), (c), and (d) already appeared in Theorem 2.4. Assumption (e) is a very weak local identifiability condition. Note that $\sup_{V_r} E_0 u_1 \geqq 0$ in any case since $u_1(\theta_0; \theta_0) \equiv 0$. In the case of maximum likelihood estimation (Example 1) (e) is always satisfied because of the information inequality.

THEOREM 4.1. *If assumptions* $\mathscr{L}$ *hold, then* (1.7) *is necessary as well as sufficient for the strong consistency of all* AMWE.

PROOF. From Theorem 2.4 it follows that if $\mathscr{L}$ holds then for every P_0 and r

$$ {}_*P_0[\sup_{V_r} u_n(\theta;\theta_0) \to 0] = 1 \tag{4.1} $$

so that $\sup_{V_r} w_n(\theta)$ is finite for almost all n with inner probability one and

$$ \begin{aligned} {}_*P_0\,[\limsup_{n\to\infty} \sup_{\Omega_r} u_n(\theta;\theta_0) < 0] &= {}_*P_0[\limsup_{n\to\infty} \{\sup_{\Omega_r} u_n(\theta;\theta_0) - \sup_{V_r} u_n(\theta;\theta_0)\} < 0] \\ &= {}_*P_0\,[\limsup_{n\to\infty} \{\sup_{\Omega_r} w_n(\theta) - \sup_{V_r} w_n(\theta)\} < 0]. \end{aligned} \tag{4.2} $$

Suppose that (1.7) is not satisfied. Then for some $\hat{P}_0$ and $\hat{r}$ the first term in (4.2), hence the last term, is < 1. Let $\hat{\theta}_0 = \theta(\hat{P}_0)$ and $\Omega_{\hat{r}} = \Omega_{\hat{r}}(\hat{\theta}_0)$. Now for every sample sequence $(x_1, x_2, \cdots)$ such that

$$ \lambda \equiv \limsup_{n\to\infty} \ \{\sup_{\Omega_{\hat{r}}} w_n(\theta) - \sup_{V_{\hat{r}}} w_n(\theta)\} \geqq 0 \tag{4.3} $$

there exists a sequence of integers $n_i \to \infty$ such that

$$ \sup_{\Omega_{\hat{r}}} w_{n_i}(\theta) - \sup_{V_r} w_{n_i}(\theta) \to \lambda \geqq 0 \qquad \text{as } i \to \infty \tag{4.4} $$

so there exists a sequence of points $\{\theta_{n_i}\} \subset \Omega_{\hat{r}}$ such that (see (1.4)) $H(\sup_\Theta w_{n_i}(\theta), w_{n_i}(\theta_{n_i})) \to 0$ as $i \to \infty$. Now choose an AMWE sequence $\{T_n(x_1, \cdots, x_n)\}$ in such a way that $T_{n_i}(x_1, \cdots, x_{n_i}) = \theta_{n_i}$ for all sequences $(x_1, x_2, \cdots)$ for which $\lambda \geqq 0$ (such a choice is always possible since we are not concerned with measurability of T_n). Then $\{T_n\}$ is not strongly consistent under $\hat{P}_0$ since

$$ {}_*\hat{P}_0[T_n \text{ in } V_{\hat{r}} \text{ a.a. } n] \leqq {}_*\hat{P}_0[\lambda < 0] < 1. \tag{4.5} $$

Q.E.D.

This result therefore implies: *if assumptions* $\mathscr{D}$ *and* $\mathscr{L}$ *hold then Condition* 1 *is necessary and sufficient for strong consistency of all* AMWE, *and if assumptions* $\mathscr{U}$ *and* $\mathscr{L}$ *hold then Condition* 3 *is necessary and sufficient for strong consistency of all* AMWE.

These statements can be strengthened if we add the natural global

IDENTIFIABILITY ASSUMPTION $\mathscr{I}$. *For every* P_0 *in* $\mathscr{P}$, $E_0 u_1(\theta;\theta_0) < 0$ *if* $\theta \neq \theta_0$.

Clearly this is weaker than (3.2) and is therefore a necessary condition for strong consistency of all AMWE. In the context of Example 1 it simply states that if $\theta \neq \theta_0$, then the two underlying distributions must be distinct. We now show that the necessary and sufficient condition (1.7) can be weakened to: for every P_0 there exists an $r_0 = r(P_0)$ such that

$$ {}_*P_0[\limsup_{n\to\infty} \sup_{\Omega_{r_0}} u_n(\theta;\theta_0) < 0] = 1. \tag{1.7$'$} $$

THEOREM 4.2. *If assumptions $\mathscr{L}$ are satisfied then* (1.7) *holds if and only if* (1.7′) *and $\mathscr{I}$ hold.*

PROOF. Assume $\mathscr{I}$ and (1.7′). For any P_0 and r choose an open neighborhood U of θ_0 such that $U \subseteqq V_r \cap V_{r_0}$ and let $W = V_{r_0} \sim U$. Then

$$\sup_{\Omega_r} u_n(\theta; \theta_0) \leqq \max \left[\sup_{\Omega_{r_0}} u_n(\theta; \theta_0), \sup_{W} u_n(\theta; \theta_0)\right]. \tag{4.6}$$

Now W is compact so by Theorem 2.4

$$_*P_0\left[\sup_W u_n(\theta; \theta_0) \to \sup_W E_0 u_1(\theta; \theta_0)\right] = 1. \tag{4.7}$$

However $E_0 u_1(\theta; \theta_0)$ is upper semicontinuous and thus achieves its maximum over W so $\sup_W E_0 u_1(\theta; \theta_0) < 0$ (by $\mathscr{I}$ which implies (1.4)). The converse is obvious. *Q.E.D.*

Therefore if assumptions $\mathscr{L}$ and $\mathscr{I}$ hold, the phrase "for every P_0 in $\mathscr{P}$ and every $r \geqq 1$" can be replaced by the phrase "for every P_0 in $\mathscr{P}$ there exists an $r_0 = r(P_0)$ such that" in assumptions $\mathscr{D}$, $\mathscr{U}$, Conditions 1, 2, 3, and (1.7), everywhere they appear in this paper.

Lastly, notice that if $\mathscr{L}$ holds and in addition Θ has a compact countable base for its topology (another global assumption) so Θ is second countable, then any open subset of Θ is σ-compact, in particular each Ω_r. Using the ideas of Theorem 3.2 this enables us to weaken slightly the uniformity assumption $\mathscr{U}$: the *convergence* in (3.12) is satisfied, so only the *uniformity* must be verified.

5. Relation to LeCam's condition for strong consistency

In Section 2 it was shown that the conditions of [1], [4], [7], [8], and [11], based on compactification of Θ, all imply either Condition 1 or 2. The conditions of LeCam ([9], pp. 302–304) are not based on compactification but rather on a form of dominance by Bochner-integrable random variables. We generalize LeCam's conditions slightly by introducing the following definitions. (The notation is that introduced in Section 2.)

DEFINITION 3. *Let $B(\Gamma)$ denote the Banach space of all bounded real-valued functions on Γ with the usual sup norm. We say $y(x, \theta)$ is Bochner-dominated on Γ with respect to P if there is a positive integer j and a function $v(\theta) \equiv v(x_1, \cdots, x_j, \theta)$ mapping $\mathscr{X} \times \cdots \times \mathscr{X}$ into $B(\Gamma)$ such that*

(i) *v is a strongly measurable mapping (with respect to the product σ-field $\mathscr{A} \times \cdots \times \mathscr{A}$),*

(ii) $\|v\| \equiv \sup_\Gamma |v(\theta)|$ *is integrable,*

(iii) *for all $(x_1, \cdots, x_j)$ in a set of probability one, $y_j(\theta) \leqq v(\theta)$ uniformly for θ in Γ.*

We say $y(x, \theta)$ is Bochner-dominated by 0 *on Γ if in addition*

(iv) $\sup_\Gamma Ev(\theta) < 0$.

Note that (i) and (ii) together are equivalent to Bochner-integrability of v. (For a definition of the terms used here see [6], also [5] and [9].) Also note that v

can be assumed to be a symmetric function of $x_1, \cdots, x_j$ (see the remark after Definition 1). Further, notice that (ii) implies $-\infty < \sup_\Gamma Ev(\theta)$.

DEFINITION 4. *$y(x, \theta)$ is semi-Bochner-dominated (semi-Bochner-dominated by 0) on Γ with respect to P if there exists a function $b(\theta)$ defined on Γ, $0 < b(\theta) < \infty$, such that $y(x, \theta)/b(\theta)$ is Bochner-dominated (Bochner-dominated by 0) on Γ and* $\inf_\Gamma b(\theta) > 0$.

LeCam's conditions for strong consistency of AMWE, in generalized form, are obtained from our Conditions 1 and 2 by replacing "dominated" by "Bochner-dominated" throughout. The following result shows that these conditions are in fact equivalent:

THEOREM 5.1. *$y(x, \theta)$ is Bochner-dominated on Γ if and only if it is dominated on Γ. Similarly, Bochner dominance by $0 \Leftrightarrow$ dominance by 0, semi-Bochner dominance $\Leftrightarrow$ semidominance, and semi-Bochner dominance by $0 \Leftrightarrow$ semidominance by 0.*

PROOF. We prove the second equivalence only: the other proofs are similar. If y is dominated by 0 on Γ, let $j = k$ and $v(x_1, \cdots, x_j, \theta) = s(x_1, \cdots, x_k)$. Clearly v is a strongly measurable mapping into $B(\Gamma)$. Since s may be chosen such that $-\infty < Es < 0$, (ii), (iii), and (iv) are satisfied, so y is Bochner-dominated by 0. Next suppose that y is Bochner-dominated by 0 on Γ. Letting

$$v_n(\theta) = \frac{1}{n} \sum_{i=0}^{n-1} v(x_{i,j+1}, \cdots, x_{(i+1),j}, \theta) \tag{5.1}$$

it follows from the Strong Law of Large Numbers (SLLN) for Bochner-integrable random variables taking values in a Banach space S that

$$P[\sup_\Gamma |v_n(\theta) - E_V(\theta)| \to 0] = 1. \tag{5.2}$$

(See Beck [2] or Hanš [5]). In stating the SLLN they assume that the Banach space S is separable. However, even if this is not the case—for example, if $S = B(\Gamma)$—strong measurability implies that v is almost separably-valued ([6], p. 72) so the range of v lies in a separable closed linear subspace of S.) Therefore

$$P[\sup_\Gamma v_n(\theta) \to \sup_\Gamma E_V(\theta) < 0] = 1. \tag{5.3}$$

(By Criterion 4 of Hanš [5], strong measurability implies that $v_n(\theta)$ is Borel measurable so $\sup_\Gamma v_n(\theta)$, being a continuous function of $v_n(\theta)$, is also Borel measurable, that is, a random variable.) However, we apply Theorem 2.2 (ii) with $y_n(\theta)$ replaced by $v_n(\theta)$ to see that there is an integer $m \geqq 1$ such that $E \sup_\Gamma v_m(\theta) < 0$. Thus Definition 1 is satisfied if we take $k = mj$ and $s(x_1, \cdots, x_k) = \sup_\Gamma v_m(\theta)$.

◇ ◇ ◇ ◇ ◇

I wish to thank Professors R. R. Bahadur, Robert Berk, and Lucien LeCam for helpful suggestions at various stages of this work.

REFERENCES

[1] R. BAHADUR, "Rates of convergence of estimates and test statistics," *Ann. Math. Statist.*, Vol. 38 (1967), pp. 303–324.

[2] A. BECK, "On the strong law of large numbers," *Ergodic Theory* (edited by F. B. Wright), New York, Academic Press, 1963, pp. 21–53.

[3] R. H. BERK, "Limiting behavior of posterior distributions when the model is incorrect," *Ann. Math. Statist.*, Vol. 37 (1966), pp. 51–58.

[4] G. B. CRAWFORD, "Consistency of non-parametric maximum likelihood estimates," 1958, unpublished.

[5] O. HANŠ, "Generalized random variables," *Conference on Information Theory, Statistical Decision Functions, Random Processes*, Prague, Publishing House of the Czechoslovak Academy of Science, 1956, pp. 61–104.

[6] E. HILLE and R. S. PHILLIPS, *Functional Analysis and Semi-Groups*, American Mathematical Society Colloquium Publication, Vol. 31, New York, American Mathematical Society, 1957 (revised ed.).

[7] P. J. HUBER, "The behavior of maximum likelihood estimates under nonstandard conditions," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1967, Vol. 1, pp. 221–233.

[8] J. KIEFER and J. WOLFOWITZ, "Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters," *Ann. Math. Statist.*, Vol. 27 (1956), pp. 884–906.

[9] L. LECAM, "On some asymptotic properties of maximum likelihood estimates and related Bayes estimates," *Univ. California Publ. Statistics*, Vol. 1 (1953), pp. 277–328.

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

[11] A. WALD, "Note on the consistency of the maximum likelihood estimate," *Ann. Math. Statist.*, Vol. 20 (1949), pp. 595–601.