The Annals of Statistics
1991, Vol. 19, No. 4, 2145–2159

ASYMPTOTIC SUPREMA OF AVERAGED RANDOM FUNCTIONS

BY ASAD ZAMAN

Columbia University

Suppose X_i are i.i.d. random variables taking values in $\mathscr{X}$, Θ is a parameter space and y: $\mathscr{X} \times \Theta \to \mathbf{R}$ is a map. Define the averages $S_n(y, \theta) = (1/n)\Sigma_{i=1}^{n} y(X_i, \theta)$ and the truncated expectations $T_m(y, \theta) = \mathbf{E}\max(y(X_1, \theta), -m)$. Under the hypothesis of global dominance [i.e., $\mathbf{E}\sup_{\Theta} y(X_1, \theta) < \infty$] and some regularity conditions, the main result of the paper characterizes the asymptotic suprema of S_n as follows. For any subset G of Θ, with probability 1,

$$\lim_{n\to\infty} \sup_{\theta \in G} S_n(y, \theta) = \lim_{m\to\infty} \sup_{\theta\in G} T_m(y, \theta).$$

This has immediate application to consistency of M-estimators. In particular, under global dominance, maxima of S_n must converge to the same limit as the maxima of $T_m(y, \theta)$ almost surely. We also obtain necessary and sufficient conditions for consistency resembling Huber's in the case of local dominance [i.e., each $\theta \in \Theta$ has a neighborhood $N(\theta)$ such that $\mathbf{E}\sup_{\psi \in N(\theta)} y(X_1, \psi) < \infty$]. In this case there must exist a function $b(\theta) \geq 1$ such that y/b is globally dominated and maxima of $T_m(y/b, \theta)$ converge.

1. Introduction. Let $(\Omega, \mathscr{B}(\Omega), \mathbf{P})$ be a probability space, $(\mathscr{X}, \mathscr{B}(\mathscr{X}))$ a measure space and X_i: $\Omega \to \mathscr{X}$ for $i = 1, 2, \dots$ an independent and identically distributed (i.i.d.) sequence of random variables. We will also use $\mathbf{P}$ to denote the common measure induced on $\mathscr{X}$ by the X_i; in particular, null subsets of $\mathscr{X}$ are with reference to this measure. Given a parameter set Θ and a function y: $\mathscr{X} \times \Theta \to \mathbf{R}$, define the averages $S_n(y, \theta) = (1/n)\Sigma_{i=1}^{n} y(X_i, \theta)$. Our object in this paper is to characterize the asymptotic supremum of S_n on subsets of Θ. These results are directly applicable to the problem of consistency of M-estimators. To avoid trivialities and degenerate cases, we will assume throughout, without explicit reference, that for each $\theta \in \Theta$, $\mathbf{E}y^+(X_1, \theta) < \infty$; here, $y^+ = \max(y, 0)$ as usual. Thus $\mathbf{E}y(X_1, \theta)$ is well defined, possibly $-\infty$. Furthermore, we will assume that for some $\theta^* \in \Theta$, $\mathbf{E}y(X_1, \theta^*) > -\infty$, so that $\sup_{\theta\in\Theta} \mathbf{E}y(X_1, \theta) > -\infty$.

We will say that y is globally dominated if $\mathbf{E}g(X_1) < \infty$ for some measurable function g such that $\sup_{\theta\in\Theta} y^+(X_1, \theta) < g(X_1)$. Perlman (1972) proved the following result about asymptotic suprema of S_n under the hypothesis of

Received October 1989; revised June 1990.
AMS 1980 *subject classifications*. Primary 62F12; secondary 60B99.
Key words and phrases. M-estimators, strong consistency, approximate discreteness, necessary and sufficient conditions for consistency, dominance, location-scale parameters, strong laws on function spaces.

global dominance:

$$(1)\qquad \sup_{\theta\in G}\mathbf{E}y(X_1,\theta) \le \limsup_{n\to\infty}\sup_{\theta\in G}S_n(y,\theta) \le \downarrow\lim_{n\to\infty}\mathbf{E}\sup_{\theta\in G}S_n(y,\theta).$$

He also showed that the second inequality is actually an equality under mild regularity conditions. Our main result provides a new characterization of the limiting value of the supremum of S_n over subsets of Θ. We show that for discrete random variables X_i with probability 1, for any $G \subseteq \Theta$,

$$\lim_{n\to\infty}\sup_{\theta\in G}S_n(y,\theta) = \lim_{m\to\infty}\sup_{\theta\in G}\mathbf{E}\max(y(X_1,\theta),-m).$$

This result also extends to a larger class of variables which can be suitably approximated by discrete ones. Under these hypotheses, it follows that the maxima of the functions $S_n(y,\theta)$ converge to the same limit as maxima of the truncated expectations $T_m(y,\theta) = \mathbf{E}\max(y(X_1,\theta),-m)$. Although our result requires stronger hypotheses than Perlman's, the quantity $T_m(y,\theta)$ is usually easier to compute than $\mathbf{E}\sup_{\theta\in G}S_n(y,\theta)$ and hence the corresponding conditions for consistency are easier to verify.

When global dominance fails, but local dominance holds, we show that for the class of variables which can be approximated by discrete ones, a condition similar to Huber's (1967) is actually necessary and sufficient. This condition involves finding a suitable function $b(\theta)$ such that y/b becomes globally dominated. Our results clarify both the choice of b and the range of applicability of this type of result.

2. Asymptotic suprema for the discrete case. In this section we will assume that $\mathscr{X} = \{1,2,3,\dots\}$, so that the X_i are integer valued. In this case, any function $y(X_i,\theta)$ can be written as

$$y(X_i,\theta) = \sum_{j=1}^{\infty} I\{X_i = j\}f_j(\theta).$$

Define $p_j = \mathbf{P}(X_1 = j)$ and $\hat{p}_{nj} = \hat{p}_{nj}(\omega) = (1/n)\sum_{i=1}^{n} I\{X_i = j\}$. We wish to explore asymptotic suprema of the sums $S_n(y,\theta)$. We will assume that global dominance holds so that

$$\mathbf{E}\sup_{\theta\in\Theta}y^+(X_1,\theta) = \sum_{j=1}^{\infty}p_j\sup_{\Theta}f_j^+ < \infty.$$

Our main result for the discrete case can now be stated as follows.

THEOREM 1. *Assume that $y(X_1,\theta)$ is globally dominated and the X_i are i.i.d. discrete variables. Then there exist a null subset N of Ω such that for all $\omega \notin N$ and for any subset G of Θ,*

$$\lim_{n\to\infty}\sup_{\theta\in G}\frac{1}{n}\sum_{i=1}^{n}y(X_i(\omega),\theta) = \lim_{m\to\infty}\sup_{\theta\in G}\sum_{j=1}^{\infty}p_j\max\big(f_j(\theta),-m\big).$$

PROOF. The condition of the following lemma is actually necessary and sufficient [see, e.g., Chatterji (1976)]. However, we will only need the implication stated below.

LEMMA 1. *If* $\sum_{j=1}^{\infty} p_j |M_j| < \infty$, *then for almost all* $\omega \in \Omega$,

$$\lim_{n\to\infty} \sum_{j=1}^{\infty} |\hat{p}_{nj}(\omega) - p_j|\,|M_j| = 0.$$

PROOF. In the separable Banach space l_1 of absolutely summable sequences, consider the random elements

$$\tilde{X}_i \equiv (M_1 I\{X_i = 1\}, M_2 I\{X_i = 2\}, \dots).$$

These are i.i.d. and satisfy

$$\mathbf{E}\|X_1\| = \sum_{j=1}^{\infty} p_j |M_j| < \infty.$$

It follows from Mourier's (1953) law of large numbers that

$$\left\| \frac{1}{n} \sum_{i=1}^{n} \tilde{X}_i - \mathbf{E}\tilde{X}_1 \right\|$$

converges to zero almost surely, where $\mathbf{E}\tilde{X}_1 = (p_1 M_1, p_2 M_2, \dots)$ is the Bochner integral of X_1. The lemma states this conclusion in a more explicit form.

Let $M_j = 1 + \sup_\Theta f_j^+$ and let $\Omega_1 \subset \Omega$ be the probability 1 subset of Ω for which the conclusion of the lemma holds. We will first show that for all $\omega \in \Omega_1$ and any subset G of Θ,

$$\limsup_{n\to\infty} \sup_{\theta\in G} S_n(y,\theta) \le \lim_{m\to\infty} \sup_{\theta\in G} T_m(y,\theta).$$

We have

$$\begin{aligned}\sup_{\theta\in G} \sum_{j=1}^{\infty} \hat{p}_{nj} f_j(\theta) &\le \sup_{\theta\in G} \sum_{j=1}^{\infty} \hat{p}_{nj} \max(f_j(\theta), -m)\\ &= \sup_{\theta\in G} \left\{ \sum_{j=1}^{\infty} (\hat{p}_{nj} - p_j)\max(f_j(\theta), -m) \right.\\ &\qquad \left. + \sum_{j=1}^{\infty} p_j \max(f_j(\theta), -m) \right\}\\ &\le \sum_{j=1}^{\infty} |\hat{p}_{nj} - p_j|(M_j + m) + \sup_{\theta\in G} T_m(y,\theta).\end{aligned}$$

Taking limits and noting that the sum converges to zero by the previous lemma yields the desired result. □

To complete the proof of our theorem, we must show that with probability 1, the reverse inequality also holds. The following lemma establishes this inequality without limitation to the discrete case.

LEMMA 2. *Suppose X_i are i.i.d. random variables taking values in sample space $\mathscr{X}$ and y: $\mathscr{X} \times \Theta \to \mathbf{R}_*$ is a measurable map for each fixed $\theta \in \Theta$. Suppose there exists measurable function $g(x)$ such that $\sup_{\theta \in \Theta} y^+(x,\theta) \le g(x)$ and $\mathbf{E}g(X) < \infty$. For any subset G of Θ with probability* 1,

$$\liminf_{n\to\infty} \sup_{\theta\in G} S_n(y,\theta) \ge \lim_{m\to\infty} \sup_{\theta\in G} \mathbf{E}\max(y(X_1,\theta),-m).$$

PROOF. Choose $\theta_m \in \Theta$ such that

$$\lim_{m\to\infty} \mathbf{E}\max(y(X_1,\theta_m),-m) = \lim_{m\to\infty} \sup_{\theta\in\Theta} \mathbf{E}\max(y(X_1,\theta),-m) \equiv \alpha > -\infty.$$

Define $y'_k(x) = \max(y(x,\theta_k),-k)$ and $y'(x) = \limsup_{k\to\infty} y'_k(x)$. Let $p_k = \mathbf{P}(y(X_1,\theta_k) < -k)$. Since $\alpha > -\infty$, we must have $\lim_{k\to\infty} p_k = 0$. For each integer n, we can choose integer $k(n)$ large enough so that $1-(1-p_{k(n)})^n < 2^{-n}$. This ensures that the probability of the event $y(X_j,\theta_{k(n)}) > -k(n)$ for all $j = 1,2,\dots,n$ is at least $1-2^{-n}$. By the Borel–Cantelli lemma, whenever $k'(n) \ge k(n)$, with probability 1, $S_n(y,\theta_{k'(n)}) = S_n(y')$ for almost all n:

$$\begin{aligned} S_n(y,\theta_{k'(n)}) &= \sum_{i=1}^{n} y(X_i,\theta_{k'(n)}) \\ &= \sum_{i=1}^{n} \max\bigl(y(X_i,\theta_{k'(n)}),-k'(n)\bigr) \equiv S_n(y'_{k'(n)}). \end{aligned}$$

Thus we have

$$\begin{aligned} \liminf_{n\to\infty} \sup_{\theta\in\Theta} S_n(y,\theta) &\ge \liminf_{n\to\infty} \limsup_{k\to\infty} S_n(y,\theta_k) \\ &\ge \liminf_{n\to\infty} S_n(y'_{k'(n)}) \\ &= \liminf_{n\to\infty} \limsup_{k\to\infty} S_n(y'_k) = \mathbf{E}y'. \end{aligned}$$

Applying Fatou's lemma, we conclude that

$$\mathbf{E}y' = \mathbf{E}\limsup_{k\to\infty} \max\bigl(y(X_1,\theta_k),-k\bigr) \ge \limsup_{k\to\infty} \mathbf{E}\max\bigl(y(X_1,\theta_k),-k\bigr) = \alpha.$$

This proves the lemma and our theorem. □

3. Approximable sets of functions. We plan to use discrete approximations to generalize our main Theorem 1 of the previous section. It is convenient to assume Θ is a separable metric space. Let $d = d(\theta,\theta')$ denote the distance function on Θ. For any $\theta \in \Theta$, let $S(\theta,r)$ be the sphere of radius r around θ: $S(\theta,r) = \{\theta' \in \Theta: d(\theta,\theta') < r\}$.

To permit approximation by discrete random variables, it is necessary to introduce an appropriate analog of separability in the function space

$F_B(\Theta, \mathbf{R}_*)$, the set of all functions from Θ to $\mathbf{R}_* = \mathbf{R} \cup \{-\infty\}$ which are bounded above. Let $\mathscr{B}_0$ be a countable basis for the topology of Θ. Our definition of approximability will be relative to this basis, which will remain fixed throughout this paper.

Given an $\varepsilon > 0$ and a subset $\mathscr{B}$ of $\mathscr{B}_0$ which forms a locally finite open cover of Θ, we will say that $f \in F(\Theta, \mathbf{R}_*)$ is a $(\mathscr{B}, \varepsilon)$ approximation to $g \in F(\Theta, \mathbf{R}_*)$ if for all $\theta \in \Theta$ and $B \in \mathscr{B}$ such that $\theta \in B$,

$$\max(g(\theta), -1/\varepsilon) \leq f(\theta), \tag{2}$$

$$\sup_{\theta' \in B} f(\theta') \leq \sup_{\theta' \in B} \max(g(\theta'), -1/\varepsilon) + \varepsilon. \tag{3}$$

Note that ε is used in two different ways in the above definition. It would be equivalent and clearer, though clumsier, to use two numbers: one for the truncation level and one for the closeness of the approximation.

A subset S of $F_B(\Theta, \mathbf{R}_*)$ will be called *approximable* if there exists a countable set $D = \{d_1, d_2, \dots\}$ of functions in $F_B(\Theta, \mathbf{R}_*)$ such that for every $s \in S$, $\varepsilon > 0$, and locally finite cover $\mathscr{B} \subset \mathscr{B}_0$, there exists $d \in D$ which is a $(\mathscr{B}, \varepsilon)$ approximation to s. It is useful to note that we can always choose the functions in D to be upper semicontinuous without loss of generality. We assume functions in D to be upper semicontinuous in the sequel.

It is obvious that separability in the uniform topology is a sufficient condition for approximability. To show that approximability is considerably weaker than separability in the uniform topology, we prove below that $F_B(\Theta, \mathbf{R}_*)$ is itself an approximable set when Θ is compact.

THEOREM 2. *If Θ is compact, then $F_B(\Theta, \mathbf{R}_*)$ is approximable.*

PROOF. It is easily checked that a locally finite cover $\mathscr{B} \subset \mathscr{B}_0$ of a compact set must actually be finite: $\mathscr{B} = \{B_1, \dots, B_m\}$. For each $\theta \in \Theta$, let $N(\theta)$ be a neighborhood such that for all j for which $\theta \in B_j$, we also have $N(\theta) \subset B_j$. Let $N_1, \dots, N_l$ be a finite subcover of the cover formed by the $N(\theta)$. Let $\psi_1, \dots, \psi_l$ be a partition of unity subordinate to the cover $N_1, \dots, N_l$. Partitions of unity exist because metric spaces are normal. We will show that every function $f \in F_B(\Theta, \mathbf{R}_*)$ can be $(\mathscr{B}, \varepsilon)$ approximated, by a continuous function. Define

$$c(\theta) = \sum_{j=1}^{l} \psi_j(\theta) \sup_{\theta' \in N_j} \max(f(\theta'), -1/\varepsilon).$$

Then it is easily verified that $\max(f(\theta), -1/\varepsilon) \leq c(\theta)$. Note that $c(\theta)$ is a weighted average of the quantities $\sup_{\theta' \in N_i} \max(f(\theta'), -1/\varepsilon)$, where $\theta \in N_i \subset B_i$. Since $\sup_{N_i} c = \sup_{B_i} \max(f, -1/\varepsilon)$, it follows that for any $\varepsilon > 0$,

$$\sup_{\theta \in B_i} \max(f(\theta), -1/\varepsilon) + \varepsilon \geq \sup_{\theta \in B_i} c(\theta).$$

Thus the continuous function c $(\mathscr{B}, \varepsilon)$ approximates f. The proof is completed using the fact that the continuous functions are separable in the uniform topology. □

4. Asymptotic suprema for the approximately discrete case. We now introduce the assumptions needed to approximate general functions $y(x, \theta)$ by discrete ones. We will need to assume that y is *locally upper semicontinuous*; that is, for all $\theta \in \Theta$, there exists a null set $N(\theta) \subset \mathscr{X}$ such that for all $x \notin N(\theta)$,

$$\lim_{r \to 0} \sup_{\psi \in S(\theta, r)} y(x, \psi) = y(x, \theta). \tag{4}$$

If $N(\theta)$ can be chosen independently of θ, then y is *globally upper semicontinuous*.

We will say that y is *approximately discrete* if there exists an approximable set $S \subset F(\Theta, \mathbf{R}_*)$ such that $y(x, \cdot) \in S$ for almost all $x \in \mathscr{X}$.

We will say that y is *supremely measurable* if there exists a countable basis $\mathscr{B}_0$ for the topology of Θ such that for each $B \in \mathscr{B}_0$, the function $\sup_{\theta \in B} y(X_1(\omega), \theta)$ (mapping Ω to $\mathbf{R}_*$) is measurable. This is the usual measurability assumption needed for results of this type; see, for example, Perlman (1972). We remark that when y is supremely measurable, we can set $g(X_1) = \sup_{\theta \in \Theta} y^+(X_1, \theta)$ in the definition of global dominance for y, since this must be measurable.

We now formulate the fundamental approximation result which permits the extension of our results for discrete variables to more general cases. Suppose $S \subset F(\Theta)$ is approximable and D is a countable set which can be used to approximate functions in S. Arrange the elements of D in a sequence so that $D = \{d_1, d_2, \dots\}$. Given $\varepsilon > 0$ and a locally finite cover $\mathscr{B} \subset \mathscr{B}_0$, let $J(s)$ be the first element of the sequence d_j which is a $(\mathscr{B}, \varepsilon)$ approximation to s. Fixing $x, y(x, \cdot)$ is an element of S and the notation $J(y(x, \cdot))$ represents a well defined element of D.

LEMMA 3 (Discrete approximation lemma). *Suppose y is supremely measurable and approximately discrete. The function $J(y(x, \theta))$ is measurable. More precisely, there exist measurable sets $E_j \subset \mathscr{X}$ forming a partition of $\mathscr{X}$ such that*

$$J(y(x, \theta)) = \sum_{j=1}^{\infty} I\{x \in E_j\} d_j(\theta).$$

PROOF. It suffices to show that the set E_1 of all x such that $J(y(x, \cdot)) = d_1$ is measurable. This is because the set E_j consists of intersections and unions of a finite number of sets of this type. To prove E_1 measurable, note that for any $B \in \mathscr{B}_0$, the set $F_B = \{x\colon \sup_{\theta \in B} \max(y(x, \theta), -1/\varepsilon) + \varepsilon \geq \sup_{\theta \in B} d_1(\theta)\}$ is measurable because y is supremely measurable. Then $J(y(x, \cdot))$ satisfies (3) for all $B \in \mathscr{B}$:

$$\sup_{\theta \in B} J(y(x,\theta)) \le \varepsilon + \sup_{\theta \in B} \max(y(x,\theta), -1/\varepsilon),$$

if and only if $x \in \bigcap_{B \in \mathscr{B}} F_B$. This set is measurable since $\mathscr{B}$ is at most countable, so the intersection is at most countable. To complete the proof, we must show that the set $G = \{x: \max(y(x,\cdot), -1/\varepsilon) \le d_1\}$ is also measurable. Define $G_B = \{x: \sup_{\theta \in B} y(x,\theta) \le \sup_{\theta \in B} d_1(\theta)\}$. We claim that $G = \bigcap_{B \in \mathscr{B}} G_B$. It is clear that if $x \in G$, then $x \in G_B$ for all $B \in \mathscr{B}_0$ so that $G \subset \bigcap_{B \in \mathscr{B}_0} G_B$. Now if $\max(y(x,\theta^*), -1/\varepsilon) > d_1(\theta^*)$ for some $\theta^* \in \Theta$, then, because d_1 is upper semicontinuous, the same inequality must hold for some neighborhood of θ^* and hence for some basis element $B \in \mathscr{B}_0$. Thus if $x \notin G$, then $x \notin \bigcap_{B \in \mathscr{B}} G_B$. This proves the lemma. □

Using the maps J to provide discrete approximations, we can generalize our results for the discrete case.

THEOREM 3. *Suppose $y(x,\theta)$ is approximately discrete, locally upper semicontinuous, supremely measurable and globally dominated. Then for almost all $\omega \in \Omega$,*

$$\lim_{n\to\infty} \sup_{\theta \in G} \frac{1}{n} \sum_{i=1}^{n} y(X_i(\omega),\theta) = \lim_{m\to\infty} \sup_{\theta \in G} \mathbf{E} \max(y(X_1,\theta), -m). \tag{5}$$

PROOF. Since the reverse inequality has already been established in Lemma 2, to prove the theorem it suffices to show that

$$\limsup_{n\to\infty} \sup_{\theta \in G} S_n(y,\theta) \le \lim_{m\to\infty} \sup_{\theta \in G} \mathbf{E}\max(y(X_1,\theta), -m).$$

For each $\theta \in \Theta$, find a neighborhood $N_k(\theta) \in \mathscr{B}_0$ of diameter less then $1/k$ such that

$$\mathbf{E} \sup_{\theta' \in N_k(\theta)} y(X_1,\theta') \le \mathbf{E} y(X_1,\theta) + 1/k.$$

Let $\mathscr{B}_k \subset \mathscr{B}_0$ be a locally finite refinement of the open cover of Θ consisting of the balls $N_k(\theta)$. Construct measurable maps J_k: $S \to D$ such that $J_k(s)$ is a $(\mathscr{B}_k, 1/k)$ approximation to s. Since $J_k(y)$ is globally dominated when y is, it follows from our results in the discrete case that

$$\begin{aligned}\lim_{n\to\infty} \sup_{\theta \in G} S_n(J_k(y)) &= \lim_{m\to\infty} \sup_{\theta \in G} \mathbf{E}\max(J_k(y(X_1,\theta)), -m)\\ &= \sup_{\theta \in G} \mathbf{E} J_k(y(X_1,\theta)).\end{aligned}$$

The second equality follows from that fact that $J_k(y) \ge -k$. Since $S_n(y) \le S_n(J_k(y))$ for all k, to prove (5) it suffices to show that $\lim_{k\to\infty} \sup_G \mathbf{E} J_k(y) \le \lim_{m\to\infty} \sup_G \mathbf{E}\max(y, -m)$. To prove this inequality, note that if $B(\theta) \in \mathscr{B}_k$

is such that $\theta \in B(\theta) \subset N_k(\theta)$, then we have

$$\begin{aligned}\sup_{\theta\in G} \mathbf{E} J_k(y(X_1,\theta)) &\le \sup_{\theta\in G} \mathbf{E} \sup_{\psi\in B(\theta)} J_k(y(X_1,\theta))\\ &\le \sup_{\theta\in G} \mathbf{E} \sup_{\psi\in B(\theta)} \max(y(X_1,\psi),-k) + \frac{1}{k}\\ &\le \sup_{\theta\in G} \mathbf{E} \sup_{\theta'\in N_k(\theta)} \max(y(X_1,\theta'),-k) + \frac{1}{k}\\ &\le \sup_{\theta\in G} \mathbf{E} \max(y(X_1,\theta),-k) + \frac{1}{k} + \frac{1}{k}.\end{aligned}$$

Taking limits as k increases to infinity yields the desired result. □

5. Consistency of *M*-estimators with global dominance. We first consider consistency with global dominance: $\mathbf{E}\sup_{\theta\in\Theta} y^+(X_1,\theta) < \infty$. The simplest case is when Θ is a compact parameter space. In this case our result is equivalent to the results of Wald (1949) and Bahadur (1967). The equivalence follows from the following lemma.

LEMMA 4. *Suppose Θ is compact and y is globally dominated and locally upper semicontinuous. Then y is approximately discrete and for all closed subsets G of Θ,*

$$\lim_{m\to\infty} \sup_{\theta\in G} \mathbf{E}\max(y(X_1,\theta),-m) = \sup_{\theta\in G} \mathbf{E} y(X_1,\theta).$$

PROOF. By Theorem 2, we know that y is approximately discrete. It is trivial that $\lim_m \sup_G T_m(y,\theta) \ge \sup_G \mathbf{E} y(X_1,\theta)$. To prove the lemma, we will prove the reverse inequality. Choose $\theta_m \in G$ such that $T_m(y,\theta_m) \ge \sup_G T_m(y,\theta) - 1/m$. Let θ^* be a cluster point of θ_m in G. Then local upper semicontinuity of y at θ^* ensures that $\mathbf{E} y(X_1,\theta^*) \ge \limsup_{m\to\infty} T_m(y,\theta_m)$ which implies the desired result.

The results of Wald (1949), Bahadur (1967) and Kiefer–Wolfowitz (1956) are all based on finding a suitable compactification of the original parameter space, and applying (an equivalent of) the result that if G is a closed set not including θ_0, the maximizing argument of $\mathbf{E} y(X_1,\theta)$, then

$$\lim_{n\to\infty} \sup_{\theta\in G} S_n(y,\theta) = \sup_{\theta\in G} \mathbf{E} y(X_1,\theta) < \mathbf{E} y(X_1,\theta_0).$$

None of these results can be used when the parameter space is not locally compact. In the locally compact case, it is unclear whether or not a suitable compactification always exists and how to find one even if it exists.

For the case where suitable compactifications of the parameter space cannot be found (in particular, when local compactness fails), the only available result is that of Perlman (1972). If

$$\limsup_{n \to \infty} \mathbf{E} \sup_{\theta \in E_k(\theta_0)} S_n(y, \theta) < \mathbf{E} y(X_1, \theta_0),$$

then by (1), consistency of maxima of S_n for θ_0 follows. Our results show that if y is approximately discrete (a condition not required by Perlman), then Perlman's condition for consistency can be replaced by the following:

$$\lim_{m \to \infty} \sup_{\theta \in E_k(\theta_0)} \mathbf{E} \max(y(X_1, \theta), -m) < \lim_{m \to \infty} \sup_{\theta \in \Theta} \mathbf{E} \max(y(X_1, \theta), -m).$$

The expressions used in our consistency result are typically easier to compute than the ones needed to apply Perlman's result. However, we do use the extra hypothesis of approximate discreteness for y. We present an example where approximate discreteness fails, to show that Perlman's result is valid in greater generality than ours.

Let B_{ij} for $i = 1, 2, \dots$ and $j = 1, 2, \dots$ be an array of i.i.d. Bernoulli random variables and define random variables X_i taking values in the space of sequences of reals by: $X_i = (B_{i1}, B_{i2}, \dots)$. Let $\Theta = \{1, 2, \dots\}$ and define $y(X_i, \theta) = B_{i\theta}$. It can be verified that $y(X_i, \theta)$ is not approximately discrete. It is easily checked that $\lim_{n \to \infty} \sup_{\theta \in \Theta} S_n(y, \theta) = 1$. It is also true that $\lim_{n \to \infty} \mathbf{E} \sup_{\theta \in \Theta} S_n(y, \theta) = 1$ verifying Perlman's result. However, $\sup_{\theta \in \Theta} \mathbf{E} \max(y(X_1, \theta) - m) = 1/2$ so that our result does not give the right value for the asymptotic supremum of S_n. □

6. Consistency with local dominance. We will say that y is *locally dominated* if every $\theta \in \Theta$ has a neighborhood $N(\theta)$ such that $\mathbf{E} \sup_{\psi \in N(\theta)} y(X_1, \psi) < \infty$. For the case that y is not globally dominated, Huber (1967) developed a new technique for proving consistency. His argument is based on an appropriate choice of a function $b(\theta)$ such that y/b is globally dominated. The scope of this condition and the appropriate choice for b remained unclear. We will now show that a modification of Huber's technique provides a necessary and sufficient condition for consistency whenever local dominance and approximate discreteness holds.

We will say that y is *normalized* if

$$\lim_{m \to \infty} \sup_{\theta \in \Theta} \mathbf{E} \max(y(X_1, \theta), -m) = 0.$$

One can always normalize a given y by adding or subtracting a constant. The following lemma shows that dividing by certain positive functions preserves normalization.

LEMMA 5. *If y is normalized and b: $\Theta \to [1, \infty)$ is an arbitrary function, then y/b is normalized.*

PROOF. If y is normalized, then given $\varepsilon > 0$ we can find m_0 such that for all $m > m_0$ and all $\theta \in \Theta$, $\mathbf{E} \max(y(X_1, \theta), -m) < \varepsilon$. Now note that

$\max(a, -m) = [a + m]^+ - m$. Thus we have for all θ,

$$\begin{aligned}\mathbf{E}\max(y/b, -m) &= (1/b)\mathbf{E}([y + bm]^+ - bm)\\ &\le \varepsilon/b \le \varepsilon.\end{aligned}$$

This shows that $\limsup_{m\to\infty} \sup_{\theta\in\Theta} \mathbf{E}\max(y/b, -m) \le 0$. For the reverse inequality, given $\varepsilon > 0$, choose θ_m such that $\mathbf{E}\max(y(X_1, \theta_m), -m) > -\varepsilon$. Let $b_m = b(\theta_m) \ge 1$. Then we have

$$\begin{aligned}\sup_{\theta\in\Theta} \mathbf{E}\max(y(X_1, \theta)/b, -m) &\ge \mathbf{E}(1/b_m)\max(y(X_1, \theta_m), -mb_m)\\ &\ge (1/b_m)\mathbf{E}\max(y(X_1, \theta_m), -m)\\ &\ge -\varepsilon/b_m \ge -\varepsilon.\end{aligned}$$

This completes the proof. □

A sufficient condition for consistency of M-estimators for some fixed $\theta_0 \in \Theta$ is the following.

THEOREM 4. *Suppose y is normalized and $b(\theta) \ge 1$ is a function satisfying the following conditions*:

(a) *y/b is globally dominated; that is, $\mathbf{E}\sup_{\theta\in\Theta} y^+(\theta)/b(\theta) < \infty$.*
(b) *y/b is locally upper semicontinuous and approximately discrete.*
(c) *For all closed subsets G of Θ which do not include $\theta_0 \in \Theta$,*

$$\lim_{m\to\infty} \sup_{\theta\in G} T_m(y/b, \theta) < 0.$$

Then all sequences of approximate maxima of $S_n(y, \theta)$ converge to θ_0.

REMARK. We define approximate maxima as in Perlman (1972): These are sequences, not necessarily measurable, of $\theta_n(\omega)$ satisfying

$$\lim_{n\to\infty} \sup_{\theta\in\Theta} S_n(y, \theta) - S_n(y, \theta_n) = 0 \quad \text{almost surely}.$$

PROOF. For any closed set G not including θ_0, since y/b is globally dominated, we must have

$$\lim_{n\to\infty} \sup_{\theta\in G} S_n(y/b, \theta) = \lim_{m\to\infty} \sup_{\theta\in G} T_m(y/b, \theta) < 0.$$

Since $b(\theta) \ge 1$, it follows that

$$\limsup_{n\to\infty} \sup_{\theta\in G} S_n(y, \theta) < 0. \tag{6}$$

On the other hand, we must have

$$\lim_{n\to\infty} \sup_{\theta\in\Theta} S_n(y/b, \theta) = \lim_{m\to\infty} \sup_{\theta\in\Theta} T_m(y/b, \theta) = 0.$$

This implies that

$$\liminf_{n\to\infty} \sup_{\theta\in\Theta} S_n(y,\theta) \ge 0. \tag{7}$$

Together (6) and (7) imply the consistency result of the theorem. □

The following converse shows that the conditions of Theorem 4 are necessary and sufficient.

THEOREM 5. *Suppose y is locally dominated, approximately discrete and all sequences of approximate maxima of $S_n(y,\theta)$ converge to θ_0. Let $y' = y \pm c$ be the normalization of y. We can find a function b such that* (a), (b) *and* (c) *of Theorem* 4 *hold* (*for* y'/b).

PROOF. Let $\{N_a\}$ for $a \in A$ an index set, be a locally finite open cover of Θ such that y is dominated on each of the sets N_a. Let ψ_a: $\Theta \to [0,1]$ be a partition of unity subordinate to the cover $\{N_a\}$. Let $b_a = \mathbf{E}\sup_{\theta\in N_a} y^+(X_1,\theta)$ and define $b(\theta) = 1 + \sum_{a\in A} b_a\psi_a(\theta)$. We claim that $\mathbf{E}\sup_\Theta y^+/b \le 1$. Denote by $\nu(dx)$ the common measure induced on the sample space $\mathscr{X}$ by the X_i. For any measurable subset B of $\mathscr{X}$, we must have for any index $a \in A$ such that $\theta \in N_a$,

$$\begin{aligned}\int_B y^+(x,\theta)\nu(dx) &\le \int_B \sup_{\theta'\in N_a} y^+(x,\theta')\nu(dx)\\ &\le \int_{\mathscr{X}} \sup_{\theta'\in N_a} y^+(x,\theta')\nu(dx) = \mathbf{E}\sup_{\theta'\in N_a} y^+(X_1,\theta') = b_a.\end{aligned}$$

It follows that for all measurable sets B,

$$\int_B y^+(x,\theta)\nu(dx) \le \sum_{a\in A} \psi_a(\theta) b_a < b(\theta).$$

This is because the weights are zero for those $a \in A$ for which $\theta \notin N_a$, so the sum is a weighted average of values b_a each of which is larger than the left-hand side. This implies that $\mathbf{E}\sup_{\theta\in\Theta} y^+(X_1,\theta)/b(\theta) \le 1$, as desired.

To verify (b), note that b is continuous by construction. Thus, y/b is locally upper semicontinuous. If S is an approximately discrete set of functions, then so is S/b. For if $s \le d$ and $b > 1$, then $s/b \le d/b$. Also if $\sup_B d \le \sup_B s + \varepsilon$, then $\sup_B d/b \le \sup_B s/b + \varepsilon/b \le \sup_B s/b + \varepsilon$. This proves that y/b is approximately discrete.

We now verify that condition (c) also holds for this choice of the function $b(\theta)$. Let G be a closed subset of Θ which does not include θ_0. Suppose, towards contradiction, that $\lim_{m\to\infty}\sup_{\theta\in G} T_m(y'/b,\theta) = 0$. Since y'/b is globally dominated, it follows that $\lim_{n\to\infty}\sup_{\theta\in G} S_n(y'/b,\theta) \ge 0$. Since y' and hence y'/b is normalized, it is also true that $\lim_{n\to\infty}\sup_{\theta\in\Theta} S_n(y/b,\theta) = 0$. From this we conclude that the supremum of $S_n(y,\theta)$ is asymptotically the same on the set G and on Θ. It follows that we can find approximate maxima of S_n in the set G asymptotically. Such sequences cannot converge to θ_0, contradicting the hypothesis of our theorem. □

7. Application to location-scale problems. We will demonstrate the power of our results by proving the consistency of maximum likelihood estimates for location-scale parameters. Let the parameter space be the half-plane $\Theta = \{(\theta_1, \theta_2): \theta_1 \in \mathbf{R},\ \theta_2 > 0\}$. Assume that X_i for $i = 1, 2, \dots$ are i.i.d. real valued random variables with common density $f^X(x)$ with respect to some measure $\nu(dx)$. We wish to explore whether maxima of the log-likelihood

$$l(\theta) = \sum_{i=1}^{n} \log\left[\frac{1}{\theta_2} f\left(\frac{X_i - \theta_1}{\theta_2}\right)\right],$$

converge to the true value $\theta_0 = (0, 1)$. After normalizing l by subtracting off its value at θ_0, we define $y(x, \theta)$ as

$$y(x, \theta) = \log\left[\frac{1}{\theta_2} f\left(\frac{X_i - \theta_1}{\theta_2}\right)\right] - \log f(x).$$

To state the result it is useful to define the set

$$S(\theta_1, \theta_2, k) = \{x \in \mathbf{R}: (1/\theta_2) f((x - \theta_1)/\theta_2) < k\}.$$

THEOREM 6. *Suppose X is a random variable with density $f(x)$ such that:*

(i) *$f(\cdot)$ is bounded and upper semicontinuous.*
(ii) $\mathbf{E} - \log f(X) < \infty$.
(iii) *For some $\alpha > 0$,*

$$\lim_{\theta_2 \to 0} \inf_{\theta_1 \in \mathbf{R}} \mathbf{P}(X \in S(\theta_1, \theta_2, \theta_2^{\alpha})) > 0.$$

Then the maximum likelihood estimates of θ in the location-scale family $1/\theta_2 f((x - \theta_1)/\theta_2)$ converge almost surely to the true values $(0, 1)$.

REMARK. The assumption (iii) is not transparent, but some version of it appears necessary for the result. Note that for fixed k, $S(\theta_1, \theta_2, k)$ expands to $\mathbf{R}$ as θ_2 goes to zero and hence $\mathbf{P}(X \in S)$ converges to 1. On the other hand, when k goes to zero, S shrinks to the null set. When k is set equal to θ^{α}, these two tendencies go in opposite directions. The smaller the α, the more slowly θ^{α} goes to zero, thus making (iii) more likely to hold. It is easily checked that (iii) does hold for all of the commonly used densities.

PROOF. It is easily verified that y is locally but not globally dominated. The problem arises for small values of θ_2. In fact for any $\varepsilon > 0$, y is globally dominated on the set $\Theta_{\varepsilon} = \{\theta \in \Theta: \theta_2 > \varepsilon\}$ and consistency is easily established by standard methods. Fix $\varepsilon > 0$ and define G_{ε} to be the complement of Θ_{ε}.

The hard part of proving consistency on Θ is to show that

$$\limsup_{n\to\infty} \sup_{\theta\in G_\varepsilon} S_n(y,\theta) < 0. \tag{8}$$

A natural candidate for the function $b(\theta)$ on the set G_ε is $b(\theta) = \max(-\log\theta_2, 1)$. With this b, it is easily checked that y/b is globally dominated:

$$\mathbf{E}\sup_{\theta\in\Theta} y^+(X_1,\theta)/b(\theta) < \infty.$$

Applying Theorem 4 to prove (8), it is enough to show

$$\lim_{m\to\infty} \sup_{\theta\in G_\varepsilon} \mathbf{E}\max\left(\frac{y(X_1,\theta)}{b(\theta)}, -m\right) < 0. \tag{9}$$ □

To apply our results, we must first verify that the set of all function $y(x,\theta)/b(\theta)$: $G_\varepsilon \to \mathbf{R}$ indexed by $x\in\mathbf{R}$ is approximable.

LEMMA 6. *The set S of functions of θ defined for each fixed $x\in\mathbf{R}$ by*

$$h_x(\theta) = \frac{y(x,\theta)}{b(\theta)} = 1 + \frac{\log f((x-\theta_1)/\theta_2)}{-\log\theta_2} - \frac{\log f(x)}{-\log\theta_2}$$

is approximable.

PROOF. It is enough to note that the functions $h_x(\theta)$ extend continuously to the one-point compactification of the set G_ε and apply Theorem 2 to conclude that the set of functions is separable.

Continuing our proof of the theorem, we now show that (9) holds for some $\varepsilon > 0$. It suffices to show that

$$\lim_{\theta_2\to 0} \sup_{\theta_1\in\mathbf{R}} \mathbf{E}\max(y/b, -m) < 0. \tag{10}$$

Note that since f is bounded and $\theta_2 \le \varepsilon$, if ε is small enough and m large enough, then

$$-\alpha - \frac{\log f(x)}{-\log\theta_2} > -m.$$

It follows that

$$\mathbf{E}\max(y/b, -m) \le \int_{\{x\in S_\alpha\}} \frac{\log(\theta_2^\alpha/f(x))}{-\log\theta_2} f(x)\,dx + \int_{\{x\notin S_\alpha\}} \frac{\log(M/f(x))}{-\log\theta_2} f(x)\,dx.$$

Under assumption (iii), the integral over $S(\theta_1,\theta_2,\theta_2^\alpha)$ approaches $-\alpha\mathbf{P}(X\in S_\alpha) < 0$. Assumption (ii) ensures that the integral over S_α^c converges to 0. This proves (10) and our theorem. □

Huber's (1967) original technique for proving consistency requires that for ε small enough,

$$\mathbf{E} \sup_{\theta \in G_\varepsilon} \frac{y(x,\theta)}{b(\theta)} < 0.$$

This is a crude sufficient condition which ensures (8). It fails to hold in the present case, so that Huber's result is inadequate for proving consistency. Perlman (1972), generalizing the work of Kiefer and Wolfowitz (1956) and others, showed that a necessary and sufficient condition for (8) is that there must exist an integer k such that

$$\mathbf{E} \sup_{\theta \in G_\varepsilon} \sum_{i=1}^{k} \frac{y(X_i,\theta)}{b(\theta)} < 0.$$

This condition was labelled semidominance by 0 by Perlman. For certain special densities, such as the normal, it is possible to verify this condition and hence prove consistency. In our case, with a general density f, it does not seem possible to verify, even though it must hold since it is necessary and sufficient. Because y is approximately discrete, our condition is also necessary and sufficient, and hence equivalent to Perlman's, but is easier to verify.

8. Concluding remarks. In the case that the function $y(X,\theta)$ can be discretely approximated, a modification of Huber's (1967) conditions provides necessary and sufficient conditions for consistency when local dominance holds. Our conditions are typically easier to check than Perlman's (1972); however, Perlman's results hold even when discrete approximation fails. Our conditions are also readily applicable to parameter spaces which are not locally compact. These results do not seem to be applicable to the case where local dominance fails. For some results in this case, see Wang (1985).

The main limitation on our result is of course the discrete approximation hypothesis. It is worth noting therefore that it is possible to weaken the requirements for approximation by discrete random elements. Given a $y(X_1,\theta)$ which is locally dominated, let $r_k(\theta)$ be a radius such that

$$\mathbf{E} \sup_{\{\theta' :\, d(\theta',\theta) < r_k(\theta)\}} y(X_1,\theta') < \mathbf{E} y(X_1,\theta) + 1/k.$$

The collection of spheres $S(\theta, r_k(\theta)) = \{\theta' \colon d(\theta',\theta) < r_k(\theta)\}$ forms a cover of Θ which is not locally finite. Let $\mathscr{B}_k$ be some locally finite refinement of this cover. In order to be able to prove Theorem 3, it is enough that there exist a countable set of functions D such that for almost all $x \in \mathscr{X}$ and all integers k, there exist $d \in D$ such that $\max(y(x,\theta), -k) \le d(\theta)$ and for all $B \in \mathscr{B}_k$,

$$\sup_{\theta \in B} d(\theta) \le 1/k + \sup_{\theta \in B} y(x,\theta).$$

Approximately discrete sets of functions as we have defined them originally must be approximable relative to all locally finite covers. Requiring approximation relative to a particular fixed sequence of locally finite covers only increase the approximable sets and enlarges the scope of applications of Theorem 3 and 4.

REFERENCES

BAHADUR, R. (1967). Rates of convergence and test statistics. *Ann. Math. Statist.* **38** 303–324.

CHATTERJI, S. D. (1976). Vector valued martingales and their applications. *Probability in Banach Spaces. Lecture Notes in Math.* **526** 33–51. Springer, New York.

HUBER, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **1** 221–233. Univ. California Press, Berkeley.

KIEFER, J. and WOLFOWITZ, J. (1956). Consistency of maximum likelihood estimator in presence of infinitely many nuisance parameters. *Ann. Math. Statist.* **27** 884–906.

LE CAM, L. (1953). On the asymptotic properties of maximum likelihood estimates and related Bayes estimates. *University of California Publications in Statist.* **1** 277–328.

MOURIER, E. (1953). Éléments aléatoires dans un espace de Banach. *Ann. Inst. H. Poincaré Sect. B* **13** 159–244.

PADGETT, W. L. and TAYLOR, R. L. (1972). *Laws of Large Numbers for Normed Linear Spaces and Certain Fréchet Spaces. Lecture Notes in Math.* **360**. Springer, New York.

PERLMAN, M. D. (1972). On the strong consistency of approximate maximum likelihood estimators. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **1** 263–281. Univ. California Press, Berkeley.

WALD, A. (1949). A note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.* **20** 595–601.

WANG, J.-L. (1985). Strong consistency of approximate maximum likelihood estimators with applications in nonparametrics. *Ann. Statist.* **13** 932–946.

DEPARTMENT OF ECONOMICS
INTERNATIONAL AFFAIRS BUILDING
COLUMBIA UNIVERSITY
NEW YORK, NEW YORK 10027