The Annals of Probability
1980, Vol. 8, No. 3, 630–635

MAXIMA OF PARTIAL SUMS AND A MONOTONE REGRESSION ESTIMATOR

BY R. T. SMYTHE

University of Oregon

Let $\{t_k\}$ be a sequence of points in d-dimensional Euclidean space. Let $\{X_k\}$ be a sequence of random variables with zero mean, i.i.d. or nearly so. If $\mathcal{Q}$ is a class of subsets of R^d, let

$$M_n(\omega) = \sup_{A \in \mathcal{Q}} \Sigma_{\{k<n\,:\,t_k \in A\}} X_k(\omega).$$

M_n is related to a commonly used estimator in monotone regression. Under various conditions on $\mathcal{Q}$ and the points $\{t_k\}$, we study the a.s. convergence to zero of M_n/n as $n \to \infty$.

0. Introduction. Let $\{t_k\}$ be a sequence of points (not necessarily distinct) in d-dimensional Euclidean space. Let $\{X_k\}$ be a sequence of random variables defined on a common probability space $(\Omega, \mathfrak{T}, P)$ and centered at their means; we think of X_k as being associated with the point t_k for $k = 1, 2, \cdots$. Let $F(y) = \sup_k P\{|X_k| \geqslant y\}$; for our purposes we will assume either that

(0.1) the $\{X_k\}$ are independent with mean zero, $F(y) \to 0$ as $y \to \infty$, and $\int_0^\infty y|dF(y)| < \infty$;

or that

(0.2) the $\{X_k\}$ form a stationary ergodic sequence with mean zero.

It is well known that the strong law of large numbers holds when either (0.1) or (0.2) is satisfied.

For $A \subset R^d$ let

(0.3) $$S_n(A, \omega) \equiv \Sigma_{\{k<n\,:\,t_k \in A\}} X_k(\omega) \qquad \text{where } \Sigma_\Phi = 0.$$

If $\mathcal{Q}$ is some collection of subsets of R^d we define, for $\omega \in \Omega$,

(0.4) $$M_n(\omega) = \sup_{A \in \mathcal{Q}} S_n(A, \omega).$$

The question we consider in this note is:

(0.5) Under what conditions on the class $\mathcal{Q}$ and the sequence $\{t_k\}$ does $M_n/n \to 0$ a.s.?

This question arises in proving the (strong) consistency of a commonly used estimator in monotonic regression problems. For the regression motivation the

Received April 27, 1978; revised February 1979.

AMS 1970 *subject classifications.* Primary 60F15; secondary 62G05.

Key words and phrases. Independent random variables, stationary ergodic sequences, maxima of partial sums, monotone regression, subadditive processes.

reader is referred to Hanson, Pledger and Wright (1973) or to Wright (1979). In the former paper the question (0.4) is resolved under condition (0.1) when $d = 1$ and $\mathcal{Q}$ is the class $\mathfrak{u}$ of *upper layers*, defined below (in the sequel the partial ordering $\leqslant$ on R^d is always taken to be the coordinatewise ordering):

DEFINITION 0.1. Let $U \subset R^d$. If whenever $s \in U$ and $s \leqslant t$ it follows that $t \in U$, U is called an *upper layer*. The complement L of an upper layer is called a *lower layer*; it clearly has the property that $s \in L$ and $s \geqslant t$ imply $t \in L$.

When $d = 1$ the upper layers are simply intervals half-infinite to the right. There are two natural generalizations of this class of sets to the case $d \geqslant 2$: the class $\mathfrak{R}$ of "half-infinite rectangles" $\{y : y \geqslant x\}$ for some $x \in R^d$, and the class $\mathfrak{u}$ of upper layers, which is considerably larger. For statistical purposes, it is again the class of upper layers which is used to define an estimator with optimal (least-squares) fit (Brunk, Ewing, and Utz (1957); see also [4], page 403).

For $d \geqslant 2$, a simple example of Wright (1979) shows that some restriction on the $\{t_k\}$ is essential when $\mathcal{Q} = \mathfrak{u}$, the upper layers, even when the $\{X_k\}$ are i.i.d.: Take $d = 2$ and let the points $\{t_k\}$ be distinct and lie on the line $y = -x$. Given any ω, there is an upper layer containing precisely the set of t_k, $k \leqslant n$, for which $X_k \geqslant 0$; hence $M_n(\omega) = \Sigma_{k=1}^n X_k^+(\omega)$ and $M_n/n \to E(X_1^+)$ a.s. by the strong law of large numbers. We show in Section 2 that if the class $\mathcal{Q}$ is taken to be $\mathfrak{R}$, and (0.1) holds, $M_n/n \to 0$ a.s. for *any* choice of points $\{t_k\}$. Although $\mathfrak{R}$ is not of statistical interest, this result is an exact analogue of that of Hanson, et al. for $d = 1$ (and gives a different approach to that case).

The case of perhaps the greatest interest, for any value of d, is that in which the points $\{t_k\}$ are a realization of an i.i.d. sequence $\{T_k\}$ of R^d-valued random vectors. In the case considered here, the $\{T_k\}$ are assumed to be defined on a probability space $(\tilde{\Omega}, \tilde{\mathfrak{F}}, \tilde{P})$ and $\{X_k\}$ and $\{T_k\}$ are taken to be independent sequences on the product space $\Omega \times \tilde{\Omega}$. Our problem is then to determine conditions under which (0.6) below holds:

(0.6) For a.e. $\tilde{\omega}$, $\sup_{A \in \mathcal{Q}} 1/n \Sigma_{i=1}^n 1_A(T_i) X_i(\omega) \to 0$ for a.e. ω.

Wright (1979) has shown that (0.6) holds under condition (0.1) for the class $\mathfrak{u}$, when the $\{T_k\}$ are i.i.d. with a distribution having no singular (with respect to Lebesgue measure on R^d) continuous part. In Section 1 we show that (0.6) holds under (0.1) or (0.2) for the class $\mathfrak{u}$ whenever the sequence $\{T_k\}$ is i.i.d., with a distribution whose continuous part does not charge the boundary of any upper layer; since all such boundaries are of Lebesgue measure zero (Brunk, et al. (1957)) this extends Wright's result. This condition appears more natural than Wright's and the proof (using a result of Steele (1978) on empirical discrepancies) is completely different.

In Section 3 we note briefly that when $(X_k\}$ satisfies (0.2) and the $\{T_k\}$ form an i.i.d. sequence, (0.6) may be cast as a problem of identifying the (constant) limit of a subadditive process.

1. Sufficient conditions for (0.6). We suppose now that the sequence $\{t_k\}$ is a realization of an i.i.d. sequence $\{T_k\}$ of random vectors in R^d. Denote by μ the probability measure of the T_k.

THEOREM 1.1. *Let* $\mathfrak{u}$ *be the class of upper layers, and let* $\{X_k\}$ *satisfy* (0.1) *or* (0.2). *If* $\{T_k\}$ *is an* i.i.d. *sequence independent of* $\{X_k\}$, *with the property that the continuous part of its measure* μ *does not charge the boundary of any upper layer, then* (0.6) *holds.*

PROOF. We note first that the X_k may be assumed to be uniformly bounded. For, given $N > 0$,

$$(1.1) \quad \frac{M_n}{n} \leqslant \sup_{A \in \mathfrak{u}} \frac{1}{n} \Sigma_{i=1}^n 1_A(T_i) 1_{\{|X_i| > N\}} X_i + \sup_{A \in \mathfrak{u}} \frac{1}{n} \Sigma_{i=1}^n 1_A(T_i) 1_{\{|X_i| < N\}} X_i.$$

The first term on the right-hand side of (1.1) is bounded above by $\frac{1}{n} \Sigma_{i=1}^n X_i^+ 1_{\{|X_i| > N\}}$ and it is easy to show that, given $\varepsilon > 0$,

$$(1.2) \quad \limsup_n \frac{1}{n} \Sigma_{i=1}^n X_i^+ 1_{\{|X_i| > N\}} < \varepsilon \text{ a.s.} \qquad \text{if } N > N_0.$$

Let $Y_i = X_i 1_{\{|X_i| \leqslant N\}}$; the second term on the right-hand side of (1.1) is bounded above by

$$(1.3) \quad \sup_{A \in \mathfrak{u}} \frac{1}{n} \Sigma_{i=1}^n 1_A(T_i)\{Y_i - E(Y_i)\} + \frac{1}{n} \Sigma_{i=1}^n E(Y_i),$$

and the second term in (1.3) can be made arbitrarily small (since the X_i have mean zero) by choosing N sufficiently large. So it will suffice to assume that for all k, $|X_k| \leqslant N$ a.s.

Next we observe that

$$(1.4) \quad \sup_{A \in \mathfrak{u}} \frac{1}{n} \Sigma_{i=1}^n 1_A(T_i) X_i \leqslant \sup_{A \in \mathfrak{u}} \frac{1}{n} \left| \Sigma_{i=1}^n (1_A(T_i) - \mu(A)) X_i \right| + \frac{1}{n} \left| \Sigma_{i=1}^n X_i \right|.$$

Under conditions (0.1) or (0.2), the second term on the right-hand side of (1.4) tends a.s. to zero as $n \to \infty$; the first term is (except for the presence of the X_i) the empirical discrepancy for the class $\mathfrak{u}$. If $\mathfrak{L}$ denotes the class of lower layers, it was shown by Blum (1955) that

$$(1.5) \quad \sup_{A \in \mathfrak{L}} \frac{1}{n} |\Sigma_{i=1}^n (1_A(T_i) - \mu(A)| \to 0 \quad \text{a.s.} \qquad \text{as } n \to \infty,$$

when μ is absolutely continuous with respect to Lebesgue measure and the $\{T_k\}$ are i.i.d.. Recently Steele (1978), drawing on a fundamental result of Vapnik and Chervonenkis (1971), showed that (1.5) holds for i.i.d. $\{T_k\}$, provided that the continuous part of μ does not charge the boundary of any lower layer ([8], Corollary 7.2); clearly the result then holds for upper layers as well.

Now define a sequence $\{\bar{X}_k\}$ of random variables such that $\bar{X}_k$ takes only a finite number $\{c_1, c_2, \cdots, c_m\}$ of values, and $|X_k(\omega) - \bar{X}_k(\omega)| < \varepsilon$ for all $\omega \in \Omega$ and all k (where $\varepsilon > 0$ is prescribed). Clearly $\{\bar{X}_k\}$ can be taken to satisfy (0.1) or (0.2).

Then

$$(1.6)\quad \sup_{A\in\mathfrak{u}}\frac{1}{n}|\Sigma_{i=1}^{n}(1_A(T_i)-\mu(A))X_i|$$
$$\leqslant \sup_{A\in\mathfrak{u}}\frac{1}{n}|\Sigma_{i=1}^{n}(1_A(T_i)-\mu(A))(X_i-\bar{X}_i)| + \sup_{A\in\mathfrak{u}}\frac{1}{n}|\Sigma_{i=1}^{n}(1_A(T_i)-\mu(A))\bar{X}_i|.$$

The first term on the right-hand side of (1.6) is less than 2ε. Consider the second term as a function of $\tilde{\omega}$, for ω fixed.

$$(1.7)\quad \sup_{A\in\mathfrak{u}}\frac{1}{n}|\Sigma_{i=1}^{n}(1_A(T_i)-\mu(A))\bar{X}_i(\omega)|$$
$$\leqslant \Sigma_{k=1}^{m}\frac{|c_k|}{n}\sup_{A\in\mathfrak{u}}|\Sigma_{\{i<n\,:\,\bar{X}_i(\omega)=c_k\}}(1_A(T_i)-\mu(A))|.$$

For each $k = 1, 2, \cdots, m$, $\{i : \bar{X}_i(\omega) = c_k\}$ is a (finite or infinite) subsequence of integers, independent of $\{T_k\}$. The result of Steele quoted above then shows that the right-hand side of (1.7) converges to 0 as $n \to \infty$, for almost all $\tilde{\omega} \in \tilde{\Omega}$; by (1.4) and (1.6), the proof is complete.

2. $\mathcal{Q} = \mathfrak{R}$ and $\{t_k\}$ arbitrary. We turn now to the case when $\mathcal{Q} = \mathfrak{R}$, the "half-infinite rectangles."

THEOREM 2.1. *Let $\mathcal{Q} = \mathfrak{R}$ and suppose that $\{X_k\}$ satisfies* (0.1). *Then for any sequence $\{t_k\}$ of points,*

$$\frac{M_n}{n} \to 0 \text{ a.s.} \qquad \textit{as } n \to \infty.$$

PROOF. We give the proof for the case $d = 2$ only; the extension to higher dimensions is straightforward. For each n, let

$$X_n'(\omega) = X_n(\omega) \qquad \text{if } |X_n(\omega)| \leqslant n$$
$$= 0 \qquad \text{if } |X_n(\omega)| > n.$$

Let X be a random variable defined on $(\Omega, \mathfrak{T}, P)$ with

$$(2.1)\qquad P\{|X| \geqslant n\} = F(n)$$

for every positive integer n, where $F(y)$ is defined just above (0.1). Then

$$(2.2)\qquad \Sigma_n P\{X_n \neq X_n'\} = \Sigma_n P\{|X_n| > n\} \leqslant \Sigma_n P\{|X| > n\} < \infty$$

by virtue of (0.1); so by Borel-Cantelli, $\{X_n\}$ and $\{X_n'\}$ are equivalent sequences, and it suffices to prove that

$$\frac{M_n'}{n} \to 0 \text{ a.s., where } M_n' = \sup_{A\in\mathfrak{R}}\Sigma_{\{k<n;\, t_k\in A\}}X_k'.$$

But clearly,

$$(2.3)\qquad \frac{M_n'}{n} \leqslant \sup_{A\in\mathfrak{R}}\frac{1}{n}\Sigma_{\{k<n\,:\,t_k\in A\}}\{X_k' - E(X_k')\} + \frac{1}{n}\Sigma_{k=1}^{n}|E(X_k')|,$$

and the second term on the right-hand side of (2.3) tends to zero as $n \to \infty$, since

$E(X_k') \to 0$ as $k \to \infty$. So it will be enough to prove the theorem in the case when $|X_n| \leqslant n$ a.s.

Fix positive integers N and N_0 with $N_0 > N$; there is a (finite) rectangular box B in the plane which contains the points $t_1, t_2, \cdots, t_{N_0}$ (not assumed distinct). At each of these points draw a horizontal and a vertical line through B so that the box B is partitioned into a rectangular grid.

Now consider the rectangular parallelopiped V with the box B as its base and height N_0. Move the point t_k (lying in B) vertically up so that its z-coordinate is k. Replicate the original grid on each plane $z = k$ and draw a line through each point t_k parallel to the z-axis; this partitions V into a cubic grid G. Place the random variable X_k at the new point t_k in the grid, and place random variables which are identically zero at every other point of G.

Consider the upper right corner of B as the origin and label each point of the grid with three integral coordinates (counting x from right to left and y from top to bottom of B). Given a point $\mathbf{k}$ in the grid, let $X_{\mathbf{k}}$ denote the (possibly zero) random variable at $\mathbf{k}$, and let

$$b_{\mathbf{k}} = \text{the } z\text{-coordinate of } \mathbf{k} \quad \text{if this coordinate exceeds } N$$
$$= N \qquad \text{otherwise.}$$

Then if $A \in \mathfrak{R}$ (in the original plane $z = 0$), to form $S_n(A)$ we sum over all $\mathbf{k}$ with z-coordinate not greater than n whose projections on B lie in A. Therefore

$$\text{(2.4)} \quad \max_{N \leqslant n \leqslant N_0} \frac{M_n}{n} = \max_{N \leqslant n \leqslant N_0} \sup_A \frac{S_n(A)}{n} \leqslant \max_{\mathbf{k} \in G,\, N_0 \geqslant k_3 \geqslant N} (b_{\mathbf{k}})^{-1} \Sigma_{\mathbf{j} \leqslant \mathbf{k}} X_{\mathbf{j}}$$
$$= \max_{\mathbf{k} \in G,\, N_0 \geqslant k_3 \geqslant N} S_{\mathbf{k}} / b_{\mathbf{k}}.$$

By the Hajek-Renyi inequality established in [7], for any $\lambda > 0$:

$$\text{(2.5)} \quad P\left\{\max_{N \leqslant n \leqslant N_0} \frac{M_n}{n} \geqslant \lambda\right\} \leqslant P\{\max_{N \leqslant k_3 \leqslant N_0} |S_{\mathbf{k}}| / b_{\mathbf{k}} \geqslant \lambda\}$$
$$\leqslant \frac{c}{\lambda^2}\left\{\frac{1}{N^2} \Sigma_{k=1}^N \sigma^2(X_k) + \Sigma_{k=N+1}^{N_0} \frac{\sigma^2(X_k)}{b_k^2}\right\},$$

where c is a constant depending only on the dimension d. But we have truncated the $\{X_k\}$ so that by standard arguments (cf. Chung (1974), page 126) the series $\Sigma_k \frac{\sigma^2(X_k)}{k^2}$ converges; by choosing N large enough we can therefore get

$$\text{(2.6)} \qquad P\left\{\sup_{n \geqslant N} \frac{M_n}{n} \geqslant \lambda\right\} < \varepsilon.$$

for any prescribed $\varepsilon > 0$, proving Theorem 2.1.

3. Subadditive processes. Assume now that $\{T_k\}$ is a stationary sequence in R^d, with measure μ; we assume also that $\{X_k\}$ is a stationary sequence with finite mean, independent of $\{T_k\}$.

Under these conditions it is not difficult to verify that, for any class $\mathcal{Q}$, the process

$$Y_{mn} \equiv \sup_{A \in \mathcal{Q}} \Sigma_{i=m+1}^{n} 1_A(T_i) X_i$$

is a *subadditive process* in the sense of Kingman (1968). By Kingman's ergodic theorem for subadditive processes,

$$\sup_{A \in \mathcal{Q}} \frac{1}{n} \Sigma_{i=1}^{n} 1_A(T_i) X_i \to C(\mathcal{Q}, \mu) \quad \text{a.s.} \qquad \text{as } n \to \infty;$$

under certain conditions (including, but not limited to, the case when both $\{X_k\}$ and $\{T_k\}$ are mixing, and $\mathcal{Q}$ and μ are arbitrary), the limit will be a constant. The problem then reduces to finding the value of $C(\mathcal{Q}, \mu)$.

When $\{X_k\}$ is ergodic and $\{T_k\}$ is i.i.d., we have seen that if $\mathcal{Q} = \mathfrak{u}$ and μ lives (and is continuous) on the line $y = x$, then $C(\mathcal{Q}, \mu) = E(X_1^+)$, and that if $\mathcal{Q} = \mathfrak{u}$ and the continuous part of μ does not charge the boundary of any upper layer, $C(\mathcal{Q}, \mu) = 0$. If $\mathcal{Q}$ is the class of convex sets and the continuous part of μ does not charge the boundary of any convex set, it again turns out that $C(\mathcal{Q}, \mu) = 0$; this follows as in Theorem 1.1 from a result of Ranga Rao (1962) for the empirical discrepancy. It would be of interest to know what values between 0 and $E(X_1^+)$ (if any) can be taken by $C(\mathcal{Q}, \mu)$ for other interesting choices of $\mathcal{Q}$ and μ.

The formulation above in terms of subadditive processes introduces an apparent simplification of the solution; for an ergodic subadditive process z_{mn}, convergence a.s. of z_{0n}/n to zero is equivalent to convergence to zero of $E(z_{0n}/n)$ (see [5]). The simplification is largely illusory, however, for the latter verification does not seem substantially simpler.

REFERENCES

[1] Blum, J. R. (1955). On convergence of empirical distribution functions. *Ann. Math. Statist.* **26** 527–529.

[2] Brunk, H. E., Ewing, G. M., and Utz, W. R. (1957). Some Helly theorems for monotone functions. *Proc. Amer. Math. Soc.* **7** 776–783.

[3] Chung, K. L. (1974). *A Course in Probability Theory* (Second Ed.). Academic Press, New York.

[4] Hanson, D. L., Pledger, Gordon, and Wright, F. T. (1973). On consistency in monotonic regression, *Ann. Statist.* **1** 401–421.

[5] Kingman, J. F. C. (1968). The ergodic theory of subadditive stochastic processes. *J. Roy. Statist. Soc.* B **30** 499–510.

[6] Ranga Rao, R. (1962). Relations between weak and uniform convergence of measures with applications. *Ann. Math. Statist.* **33** 659–680.

[7] Smythe, R. T. (1974). Sums of independent random variables on partially ordered sets. *Ann. Probability* **2** 906–917.

[8] Steele, J. Michael (1978). Empirical discrepancies and subadditive processes. *Ann. Probability* **6** 118–127.

[9] Vapnik, V. N., and Chervonenkis, A. Ya. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theor. Probability Appl.* **16** 131–138.

[10] Wright, F. T. (1979). A strong law for variables indexed by a partially ordered set with applications to isotone regression. *Ann. Probability* **7** 109–127.

Department of Mathematics
University of Oregon
Eugene, Oregon 97403