BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 21, Number 2, October 1989

AN ERGODIC THEOREM FOR CONSTRAINED SEQUENCES OF FUNCTIONS

JOHN C. KIEFFER

I. Introduction. For each integer $n = 1, 2, \ldots$, let S_n be a finite nonempty set, and let $\mathfrak{F}_n$ be a nonempty family of real-valued functions on S_n. This announcement is concerned with the asymptotic behavior of sequences of functions $\{f_n\}$ which are constrained by $\{\mathfrak{F}_n\}$ in the sense that $f_n \in \mathfrak{F}_n$ for every n. Specifically, given an S_n-valued random variable Y_n $(n \geq 1)$, an examination is made of the almost sure asymptotic behavior of sequences of random variables of the form $\{f_n(Y_n)/n\}$, where $\{f_n\}$ is constrained by $\{\mathfrak{F}_n\}$ is the above sense. (See Theorem 1.) Of particular interest for applications is the context in which for some finite set A, and some stationary sequence $X_1, X_2, \ldots$ of A-valued random variables, we have $S_n = A^n$ and $Y_n = (X_1, \ldots, X_n)$ for every n. In this context, our main result (Theorem 2) is an ergodic theorem which gives sufficient conditions on the constraining sequence $\{\mathfrak{F}_n\}$ so that $\{f_n(X_1, \ldots, X_n)/n\}$ will converge almost surely when $\{f_n\}$ is a certain sequence of functions constrained by $\{\mathfrak{F}_n\}$. The subadditive ergodic theorem **[4]** for stationary, ergodic processes with finite state space and the Shannon-McMillan-Breiman theorem **[2]** are special cases of our main result.

At this point, we mention examples from information theory and statistics illustrating the utility of results of the type just described.

EXAMPLE 1.1 (INFORMATION THEORY). Let S_n be a set of messages, each of which has length n. Let $\mathfrak{F}_n$ be the family of all functions $f\colon S_n \to \{1, 2, \ldots\}$ for which there is a uniquely decipherable code **[1]** which assigns to each message $m \in S_n$ a binary codeword of length $f(m)$. Let Y_n be a random message from S_n. One may want to select $f_n \in \mathfrak{F}_n$ so that, with probability one, the codeword length per message length $f_n(Y_n)/n$ does not exceed a certain bound in the limit as the message length $n \to \infty$.

EXAMPLE 1.2 (STATISTICS). Let S_n be the set of all sequences of length n that can be formed from a finite set A. Let $\Theta(\mu)$ be a real parameter of an unknown probability distribution μ on A. Let Y_n be a random sample of size n drawn according to the distribution μ. A family $\mathfrak{F}_n$ of functions on S_n is specified, consisting of the statistics that are to be allowed as possible parameter estimators. It is desired to select a statistic $f_n \in \mathfrak{F}_n$ $(n \geq 1)$ so that $f_n(Y_n) \to \Theta(\mu)$ with probability one as the sample size $n \to \infty$, no matter what may be the distribution of μ.

Received by the editors November 14, 1988 and, in revised form, May 25, 1989.

1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 28D99; Secondary 60G10, 94A15.

Author supported by NSF Grant NCR-8702176.

II. Log-convex families of functions. Let S be a finite nonempty set and $\mathfrak{F}$ a nonempty family of real-valued functions on S. We say that $\mathfrak{F}$ is *log-convex* if the family of functions $\{e^{-f}: f \in \mathfrak{F}\}$ is convex.

Constraining sequences $\{\mathfrak{F}_n\}$ considered later shall be required to consist of log-convex families. In what follows, we will see that the log-convex assumption on a family of functions allows us to select a certain function from the family in a natural way.

To see this, fix a log-convex family of functions $\mathfrak{F}$ on the finite set S, as well as a probability measure μ on S. Using the log-convex property of $\mathfrak{F}$, one can easily show there exists at most one function $f^*: S \to \mathbf{R}$ (in the sense of μ-almost everywhere equivalence) such that

$$\int f^* \, d\mu = \min\left\{\int f \, d\mu : f \in \mathfrak{F}\right\}. \tag{1}$$

We assume throughout that this function does exist. It does exist, for example, provided both of the following conditions are satisfied:

(i) There exists $C \in R$ such that $f(x) \geq C$, whenever $f \in \mathfrak{F}$, $x \in S$; and

(ii) If f is a real-valued function on S and there exists a sequence $\{f_n\}$ from $\mathfrak{F}$ such that

$$\mu\left\{s: \lim_{n\to\infty} f_n(s) = f(s)\right\} = 1, \quad \text{then } f \in \mathfrak{F}.$$

The following Lemma shows that f^* is an approximate lower bound for the functions in $\mathfrak{F}$, in a probabilistic sense.

LEMMA 1. *Let $\mathfrak{F}$ be a log-convex family of real-valued functions on the finite set S. Let μ be a probability measure on S. Let $f^* \in \mathfrak{F}$ be a function for which* (1) *holds. Then*

$$\mu\{s \in S: f(s) \geq f^*(s) - \varepsilon\} \geq 1 - e^{-\varepsilon}, \qquad f \in \mathfrak{F},\ \varepsilon > 0. \tag{2}$$

PROOF. Statement (2) is true if

$$\int e^{f^* - f} \, d\mu \leq 1. \tag{3}$$

To show this, we adapt a line of argument given in **[3]**. Let $\mathbf{R}^S$ denote the set of all real-valued functions on S (considered as a finite-dimensional Euclidean space) and let $\mathbf{R}^S_+$ denote the set of all $\alpha \in \mathbf{R}^S$ such that $\alpha > 0$ throughout S. Let $\Phi: \mathbf{R}^S_+ \to \mathbf{R}$ be the concave, differentiable function

$$\Phi(\alpha) = \sum_{s\in S} \mu(s) \log \alpha(s), \qquad \alpha \in \mathbf{R}^S_+.$$

Let $\beta: S \to \mathbf{R}$ be the gradient of Φ at e^{-f^*}, which is easily seen to satisfy

$$\beta(s) = \mu(s) e^{f^*(s)}, \qquad s \in S. \tag{4}$$

Let H be the hyperplane

$$H = \left\{\alpha \in \mathbf{R}^S: \sum_{s\in S} \alpha(s)\beta(s) = 1\right\}.$$

A moment's reflection will convince the reader that H must be the unique separating hyperplane for the convex subsets C_1, C_2 of $\mathbf{R}^S$ given by

$$C_1 = \{e^{-f} : f \in \mathfrak{F}\}$$
$$C_2 = \left\{\alpha \in \mathbf{R}_+^S : \Phi(\alpha) \geq -\int f^* \, d\mu\right\}.$$

It follows from this that

$$\sum_s \alpha(s)\beta(s) \leq 1, \qquad \alpha \in C_1. \tag{5}$$

Substituting into (5) the expression for β given in (4), we obtain (3).

Our first theorem is an easy corollary of Lemma 1.

THEOREM 1. *For each $n = 1, 2, \ldots$, let S_n be a finite nonempty set, let $\mathfrak{F}_n$ be a nonempty, log-convex family of real-valued functions on S_n, and let Y_n be an S_n-valued random variable. Suppose that for each n there exists $f_n^* \in \mathfrak{F}_n$ for which*

$$E[f_n^*(Y_n)] = \min\{E[f(Y_n)] : f \in \mathfrak{F}_n\}.$$

Then for any sequence of functions $\{f_n\}$ constrained by $\{\mathfrak{F}_n\}$

(a) $\underline{\lim}\, n^{-1} f_n(Y_n) \geq \underline{\lim}\, n^{-1} f_n^*(Y_n)$ *a.s.,*

and

(b) $\overline{\lim}\, n^{-1} f_n(Y_n) \geq \overline{\lim}\, n^{-1} f_n^*(Y_n)$ *a.s..*

III. Additive sequences of families of functions. Fix a finite set A throughout this section. For each $n = 1, 2, \ldots$, let $\mathfrak{F}_n$ be a family of functions from $A^n \to \mathbf{R}$. We say the sequence of families $\{\mathfrak{F}_n\}$ is *additive* if for each pair of integers $n, m \geq 1$ and each $f_n \in \mathfrak{F}_n$, $f_m \in \mathfrak{F}_m$, it is true that f_{n+m} is a member of $\mathfrak{F}_{n+m}$, where f_{n+m} is the function

$$f_{n+m}(x_1, \ldots, x_{n+m}) = f_n(x_1, \ldots, x_n) + f_m(x_{n+1}, \ldots, x_{n+m}), \qquad (x_1, \ldots, x_{n+m}) \in A^{n+m}.$$

We point out two examples of additive sequences $\{\mathfrak{F}_n\}$.

EXAMPLE 3.1. Let $\{g_n\}$ be a sequence of functions such that for $m, n \geq 1$, the functions $g_n : A^n \to \mathbf{R}$, $g_m : A^m \to \mathbf{R}$, and $g_{n+m} : A^{n+m} \to \mathbf{R}$ are related by

$$g_{n+m}(x_1, \ldots, x_{n+m}) \leq g_n(x_1, \ldots, x_n) + g_m(x_{n+1}, \ldots, x_{n+m}), \qquad (x_1, \ldots, x_{n+m}) \in A^{m+n}. \tag{6}$$

Let $\{\mathfrak{F}_n\}$ be the sequence of families in which, for each n, $\mathfrak{F}_n$ consists of all functions $f : A^n \to \mathbf{R}$ satisfying $f \geq g_n$. The sequence $\{\mathfrak{F}_n\}$ is additive.

EXAMPLE 3.2. *Let $\{\mathfrak{F}_n\}$ be the sequence of families in which, for each n, $\mathfrak{F}_n$ consists of all functions $f : A^n \to \mathbf{R}$ satisfying*

$$\sum_{x \in A^n} e^{-f(x)} \leq 1.$$

The sequence $\{\mathfrak{F}_n\}$ is additive.

We state now our main result.

THEOREM 2. *For each $n = 1, 2, \ldots$, let $\mathfrak{F}_n$ be a nonempty log-convex family of functions on A^n. Suppose the sequence $\{\mathfrak{F}_n\}$ is additive. Let $X_1, X_2, \ldots$ be a stationary, ergodic sequence of A-valued random variables. Assume that for each n, there exists $f_n^* \in \mathfrak{F}_n$ such that*

$$Ef_n^*(X_1, \ldots, X_n) = \min_{f \in \mathfrak{F}_n} Ef(X_1, \ldots, X_n).$$

Then, the sequence of random variables $\{f_n^*(X_1, \ldots, X_n)/n\}$ converges almost surely to the extended real number $M \in [-\infty, \infty)$ given by

$$M = \inf_{n \geq 1} n^{-1} \min_{f \in \mathfrak{F}_n} Ef(X_1, \ldots, X_n).$$

REMARKS. Before proceeding with the proof of Theorem 2, we point out two classic convergence theorems that follow from it, viz., the subadditive ergodic theorem **[4]** and Shannon-McMillan-Breiman theorem **[2]**. First, let $\{g_n\}$ be a sequence of functions satisfying (6). Then, applying Theorem 2 to the family $\{\mathfrak{F}_n\}$ of Example 3.1, we see that $\{g_n(X_1, \ldots, X_n)/n\}$ converges almost surely. This result is the *subadditive ergodic theorem* for stationary ergodic sequences $\{X_n\}$ with finite state space. Secondly, the *Shannon-McMillan-Breiman theorem* states that if H is the entropy rate **[1]** of the stationary, ergodic sequence $\{X_n\}$, then

$$\lim_{n \to \infty} \{-n^{-1} \log \Pr[X_1 = x_1, \ldots, X_n = x_n]\} = H$$

holds for almost every sequence $x_1, x_2, \ldots$ of observed values of $X_1, X_2, \ldots$; it can be obtained applying Theorem 2 to the sequence of families $\{\mathfrak{F}_n\}$ in Example 3.2.

LEMMA 2. *Let $\mathfrak{F}$ be a log-convex family of functions on the finite set S. Then, for any positive integer k, if $f_1, f_2, \ldots, f_k$ are functions from $\mathfrak{F}$, there exists a function $f \in \mathfrak{F}$ such that*

$$f \leq f_i + \log k, \qquad i = 1, \ldots, k.$$

PROOF OF LEMMA 2. Set $f = -\log[k^{-1} \sum_{i=1}^k e^{-f_i}]$.

PROOF OF THEOREM 2. Abbreviate $(X_1, \ldots, X_n)$ by $\mathbf{X}_n$, $n \geq 1$. The relation

$$\overline{\lim}\, n^{-1} f_n^*(\mathbf{X}_n) \leq M \quad \text{a.s.} \tag{7}$$

follows from Theorem 1(b) and the fact that $\{\mathfrak{F}_n\}$ is additive, using the pointwise ergodic theorem. Applying Theorem 1(a),

$$\begin{aligned} \underline{\lim}\, n^{-1} f_n^*(\mathbf{X}_n) &\leq \underline{\lim}[n^{-1} f_1^*(X_1) + n^{-1} f_{n-1}^*(X_2, \ldots, X_n)] \\ &= \underline{\lim}\, n^{-1} f_n^*(X_2, \ldots, X_n) \quad \text{a.s.}, \end{aligned}$$

from which it follows that $\underline{\lim}\, n^{-1} f_n^*(\mathbf{X}_n)$ is, with probability one, a shift-invariant function of $(X_1, X_2, \ldots)$. Since the sequence $\{X_n\}$ is ergodic, this means there is a constant $B \in [-\infty, M]$ such that

$$\underline{\lim}\, n^{-1} f_n^*(\mathbf{X}_n) = B \quad \text{a.s.}$$

In view of (7), the proof is complete once we show that $M \leq B$. Our demonstration that $M \leq B$ is an adaptation of a line of argument originated by Ornstein and Weiss **[5]** and modified by Shields **[6]**. Fix a real number $B' > B$, an $\varepsilon > 0$, and a positive integer N. For each $j \geq 1$, partition the block $\mathbf{X}_j$ into random sub-blocks $\{U_i\colon i = 1, \ldots, K_j\}$ ordered from left to right so that

(i) The length L_i of each U_i is either 1 or at least N.

(ii) If $L_i \geq N$, then $f^*_{L_i}(U_i) \leq L_i B'$.

(iii) Almost surely, $\{i\colon L_i = 1\}$ has no more than $j\varepsilon$ elements for sufficiently large j.

Using Lemma 2 and the fact that $\{\mathfrak{F}_n\}$ is additive, we may choose for each j a function $h_j \in \mathfrak{F}_j$ such that the sequence $\{h_j(\mathbf{X}_j)/j\colon j \geq 1\}$ is bounded above, and

$$h_j(\mathbf{X}_j) \leq \log[1 + C_j] + \sum_{i=1}^{K_j} f^*_{L_i}(U_i), \tag{8}$$

where C_j is the number of values of the random vector of random length $(L_1, L_2, \ldots, L_{K_j})$. Applying (i)–(iii) to (8), we see that with probability one,

$$h_j(\mathbf{X}_j) \leq \log[1 + C_j] + jB' + j\varepsilon + j\varepsilon(|B'| + \sup|f_1^*|),$$

for sufficiently large j. This yields

$$E\left[\overline{\lim_j}\, j^{-1} h_j(\mathbf{X}_j)\right] \leq \overline{\lim_j}\, j^{-1}\log[1 + C_j] + B' + \varepsilon + \varepsilon[|B'| + \sup|f_1^*|]. \tag{9}$$

Since $\{h_j(\mathbf{X}_j)/j\}$ is bounded above, Fatou's Lemma tells us that

$$E\left[\overline{\lim_j}\, j^{-1} h_j(\mathbf{X}_j)\right] \geq \overline{\lim_j}\, E[j^{-1} h_j(\mathbf{X}_j)] \geq M. \tag{10}$$

The integer C_j is no greater than $\sum_{1 \leq i \leq j/N} \binom{j}{i}^2$. Consequently, the right side of (9) is less than or equal to B in the limit as $N \to \infty$, $\varepsilon \to 0$, $B' \to B$. This observation, combined with inequality (10), yields the desired inequality $M \leq B$.

FINAL REMARKS. If $M > -\infty$, one can also obtain convergence of $\{f_n^*(X_1, \ldots, X_n)/n\}$ to M in L^1 mean. Also, Theorem 2 can be generalized to the case of a stationary random sequence $\{X_n\}$ which is not necessarily ergodic. (If $\{X_n\}$ is not ergodic, $\{f_n^*(X_1, \ldots, X_n)/n\}$ converges almost surely to a random variable whose expected value is M.) With some additional assumptions, one can also remove the requirement that the set A be finite. These improvements to Theorem 2 (and applications) shall appear elsewhere.

REFERENCES

1. R. Blahut, *Principles and practice of information theory*, Addison-Wesley, Reading, Mass., 1987.

2. L. Breiman, *The individual ergodic theorem of information theory*, Ann. Math. Statist. **28** (1957), 809–811.

3. I. Csiszár, J. Körner, L. Lovász, K. Marton, and G. Simonyi, *Entropy splitting for antiblocking pairs and perfect graphs* (submitted).

4. J. Kingman, *The ergodic theory of subadditive stochastic processes*, J. Roy. Statist. Soc. Ser. B **30** (1968), 499–510.

5. D. Ornstein and B. Weiss, *The Shannon-McMillan-Breiman Theorem for a class of amenable groups*, Israel J. Math. **44** (1983), 53–60.

6. P. Shields, *The ergodic and entropy theorems revisited*, IEEE Trans. Inform. Theory **IT-33** (1987), 263–266.

DEPARTMENT OF ELECTRICAL ENGINEERING, UNIVERSITY OF MINNESOTA, MINNEAPOLIS, MINNESOTA 55455