ISRAEL JOURNAL OF MATHEMATICS, Vol. 69, No. 2, 1990

A SHANNON–McMILLAN THEOREM FOR MOTLEY NAMES

BY
JIM PROPP†
Department of Mathematics, University of California, Berkeley, California, USA

ABSTRACT
The Shannon–McMillan theorem tells us the number of "typical" outcomes of an experiment in which one observes some discrete-valued function of an ergodic dynamical system for many time-steps; it also tells us that most such outcomes are roughly equally likely. Here it is shown that the same uniformity holds when several different measurements are made on the system over the course of time, provided that each sort of measurement is made many times in direct succession.

The Shannon–McMillan theorem ([S], [M]) gives a uniformity property for the masses of the components of the partition $\bigvee_{i=1}^{n} T^{-i}P$ as $n \to \infty$, where T is an ergodic measure-preserving transformation of a measure space $(X, \mu, \mathscr{A})$ and P is a finite partition of X. In particular, the theorem says that for all $\varepsilon > 0$, there exists N such that for all $n \geqq N$, all but ε in measure of the components of the partition $\bigvee_{i=1}^{n} T^{-i}P$ have μ-measure $e^{-n(h \pm \varepsilon)}$, where $h = h(T, P)$ is the entropy of the process T, P. (Here, as hereafter, $h \pm \varepsilon$ denotes some quantity between $h - \varepsilon$ and $h + \varepsilon$.)

Note that $(\bigvee_{i=1}^{n} T^{-i}P)(x)$ (the component of $\bigvee_{i=1}^{n} T^{-i}P$ containing the point x) determines a sequence of n "symbols" $k_1, k_2, \ldots, k_n$, with $T^i x$ belonging to the k_ith component of P for $1 \leqq i \leqq n$; this string is called the P-name of length n determined by the point x. We will freely pass back and forth between thinking of $P(x)$ as a "name" (a string of symbols) and an "atom" (a point in the finite probability space P, μ).

† Partially supported by an NSF Graduate Fellowship and an NSF Postdoctoral Fellowship during the writing of this article. Also supported as a Research Assistant under NSF Grant #DMS 84–03182.
Received May 29, 1989

In this article we prove a uniformity result for "motley" names, such as the names that correspond to components of the partition

$$P^* = \bigvee_{i=1}^{n_1} T^{-i}P \vee \bigvee_{i=n_1+1}^{n_2} T^{-i}Q,$$

with $1 \leqq n_1 \leqq n_2$. In general, we look at names consisting of a block of symbols from one alphabet, followed by a block of symbols from some other alphabet, and so on. We will show that all but a tiny fraction of the components of such a partition have the same mass, to within an exponentially small factor; for example, in the case of the partition P^* defined above, the theorem proved here implies that given any $\varepsilon > 0$, all but ε of the components of the partition P^* have mass $e^{-n_1 h(T,P)} e^{-(n_2-n_1)h(T,Q)}$ to within a factor of $e^{\pm n_2 \varepsilon}$ as long as n_2 is sufficiently large. Note that this is exactly the result we would get if we pretended the blocks were independent of one another and applied the standard Shannon–McMillan theorem to each block individually. What is more, our main theorem implies that this estimate holds for all the cylinder sets $P^*(x)$ as n_1 (and hence also P^*) varies, provided that a set of small measure (not depending on n_1) is disregarded. This strengthened result has been useful in showing that certain randomly constructed codes from one process to another are as close to being invertible as is consistent with entropy constraints; and such codes are in turn instrumental in the author's proof of the existence of unilateral (or "one-sided") factor maps from finite entropy independent processes onto mixing Markov chains of the same entropy [Pr]. This work on one-sided factor maps provides the motivation for proving all our results in a past-conditional way, whereby we estimate not just the absolute probabilities of long motley names but their conditonal probabilities given the infinite past of some measurement on the system.

Let $(X, T, \mathscr{A}, \mu)$ be an ergodic dynamical system with finite measurable partitions $P, Q, R, P^{(1)}, P^{(2)}, \ldots, P^{(r)}$. The process-entropy of P is equal to the common value of $H(P \mid P_{-\infty}^{-1})$ and $\lim_{n\to\infty}(1/n)H(P_1^n)$; we will write this quantity as $h(T, P)$, or sometimes just $h(P)$. For $x \in X$, we let $P(x)$ denote the component of P that contains x. We define $P^n = T^{-n}P$ ("the partition P shifted to time n"); it has the property that $P^n(x) = P(T^n x)$. We also introduce the useful abbreviation

$$P_a^b = \bigvee_{n=a}^{b} P^n$$

($=$ the trivial partition if $a > b$).

The partition R plays a special reserved role, serving as the "past" upon which events are conditioned; that is, we condition upon the σ-algebra $R^0_{-\infty}$. Unconditioned estimates can be gotten by taking R to be trivial. If $\mathscr{B}$ is $R^0_{-\infty}$ or any sub-σ-algebra of $R^0_{-\infty}$, we call its constituents *fibers*, and let $\mathscr{B}(x)$ denote the fiber containing the point x. Our first result concerns the conditional distribution of P_1^n on the fibers of $\mathscr{B}$.

THEOREM 1. *Let $\mathscr{B}$ be a sub-σ-algebra of $R^0_{-\infty}$. Fix $0 < \varepsilon < 1$. Then for all n sufficiently large, the function $\mu(P_1^n(x) \mid \mathscr{B}(x))$ has value $e^{-n(h(P) \pm \varepsilon)}$ on a set of measure $> 1 - \varepsilon$.*

PROOF. Put $h = h(P)$ for convenience. Take positive numbers $\delta_2 \ll \delta_1 \ll \delta \ll \varepsilon$; more specifically,

$$\delta < \frac{\varepsilon^2}{16}, \quad \delta_1 < \frac{\delta^2}{3}, \quad \delta_2 < \min\left(\delta_1, \frac{\delta_1}{2h}, \frac{\delta}{2(h+\delta)}\right).$$

By the Shannon–McMillan theorem, when n is large all but δ_2^2 in (unconditioned) measure of the components of P_1^n have measure $e^{-n(h \pm \delta_1/2)}$; call these the *typical* components of P_1^n, and let $\hat{P}$ be the partition whose components $\hat{P}_1, \hat{P}_2, \ldots$ are the typical components of P_1^n and whose special component $\hat{P}_0$ is the union of all the non-typical components of P_1^n. Note that P_1^n refines $\hat{P}$, so that $H(\hat{P}) \leqq H(P_1^n)$, which is less than $n(h + \delta_1)$ for large n. On the other hand, since all of the components of $\hat{P}$ (with the exception of a single component $\hat{P}_0$ of measure $< \delta_2^2 < \delta_1/2h$) have measure $e^{-n(h \pm \delta_1/2)}$, their joint contribution to $H(\hat{P})$ must be at least

$$(1 - \delta_1/2h)n(h - \delta_1/2) > n(h - \delta_1).$$

Hence $\hat{P}$ is a partition with $e^{n(h \pm \delta_1)}$ components that satisfies $H(\hat{P}) = n(h \pm \delta_1)$.

If n is large, then $(1/n)H((P \vee R)_1^n)$ is within δ_1 of $h(P \vee R)$, so that $H((P \vee R)_1^n)$ is within $n\delta_1$ of

$$nh(P \vee R) = H((P \vee R)_1^n \mid (P \vee R)^0_{-\infty}).$$

That is, the mutual information $I((P \vee R)_1^n; (P \vee R)^0_{-\infty})$ is less than $n\delta_1$ and *a fortiori* we have $I(\hat{P}; \mathscr{B}) < n\delta_1$. Thus

$$\begin{aligned} H(\hat{P} \mid \mathscr{B}) &= H(\hat{P}) - I(\hat{P}; \mathscr{B}) \\ &\geqq n(h - \delta_1) - n\delta_1 \\ &= n(h - 2\delta_1). \end{aligned}$$

That is, $H(\hat{P} \mid B)$ has average value at least $n(h - 2\delta_1)$ as B varies over $\mathscr{B}$. But on the other hand note that $\#(\hat{P}) < e^{n(h+\delta_1)}$, so that $H(\hat{P} \mid \mathscr{B})$ is less than $n(h+\delta_1)$ for all $B \in \mathscr{B}$. If we apply the usual "Fubini estimate" (and the condition $\delta_1 < \delta^2/3$), an easy calculation reveals that $H(\hat{P} \mid B)$ must exceed $n(h-\delta)$ on a set of B's of measure $> 1 - \delta$. Recall also that $\hat{P}_0$ has measure $< \delta_2^2$, so that $\hat{P}_0$ has conditional measure $< \delta_2$ on a set of B's of measure $> 1 - \delta_2$. Call a fiber B *good* if it satisfies both conditions, that is, if $H(\hat{P} \mid B) > n(h-\delta)$ and $\mu(\hat{P}_0 \mid B) < \delta_2$. The good fibers have total measure $> 1 - \delta - \delta_2 > 1 - \varepsilon/2$, and we hereafter restrict attention to them.

Let B be a good fiber of $\mathscr{B}$, and write $\mu_B(\cdot)$ instead of $\mu(\cdot \mid B)$. Taking stock of our progress thus far, we note that:

(a) $\#(\hat{P}) < e^{n(h+\delta_1)} < e^{n(h+\delta)}$;

(b) $H(\hat{P} \mid B) > n(h-\delta)$; and

(c) $\mu_B(\hat{P}_0) < \delta_2$.

We also note that no $\hat{P}$-measurable set of μ_B-measure c can contribute more than $cn(h+\delta)$ to $H(\hat{P} \mid B)$. To see why this continuity principle holds, note that the contribution made to $H(\hat{P} \mid B)$ by M atoms of total μ_B-measure c is as large as possible when the atoms all have mass c/M, yielding a contribution of

$$M \cdot \left[\frac{c}{M} \log \frac{M}{c}\right] = c \log M + c \log \frac{1}{c} .$$

Since $M \leqq \#(\hat{P}) < e^{n(h+\delta_1)}$ and $c \log(1/c) < 1$, the contribution is bounded above by $cn(h+\delta_1) + 1$, which for large n is less than $cn(h+\delta)$.

Let $\mathscr{C}$ be the collection of atoms $\hat{P}_i$ with $i \neq 0$ (that is, the collection of typical atoms of P_1^n), let

$$C_- = \bigcup \{A \in \mathscr{C} : \mu_B(A) < e^{-n(h+2\delta)}\},$$

$$C_0 = \bigcup \{A \in \mathscr{C} : e^{-n(h+2\delta)} \leqq \mu_B(A) \leqq e^{-n(h-\varepsilon)}\},$$

$$C_+ = \bigcup \{A \in \mathscr{C} : e^{-n(h-\varepsilon)} < \mu_B(A)\},$$

and let $\bar{C} = \hat{P}_0$, so that $\bar{C}$, C_-, C_0, C_+ are exclusive and exhaustive. Note that all the atoms A in C_0 satisfy $\mu_B(A) = e^{-n(h \pm \varepsilon)}$; we need to show that $\mu_B(C_0) > 1 - \varepsilon/2$. We already know that $\mu_B(\bar{C}) = \mu_B(\hat{P}_0) < \delta_2$. Also, since C_- contains at most $\#(\hat{P})$ atoms,

$$\begin{aligned}\mu_B(C_-) &\leqq \#(\hat{P})e^{-n(h+2\delta)} \\ &< e^{n(h+\delta)}e^{-n(h+2\delta)} \\ &= e^{-n\delta},\end{aligned}$$

which is less than δ_2 when n is large. It remains to bound $\mu_B(C_+)$, which we proceed to do in a roundabout way. Since $\mu_B(\bar{C})$, $m_B(C_-) < \delta_2$, our continuity principle implies that $\bar{C}$ and C_- each contribute at most $\delta_2 n(h+\delta)$ to $H(\hat{P} \mid B)$, yielding a joint contribution of at most $2\delta_2 n(h+\delta) < n\delta$. C_0 contributes at most

$$\mu_B(C_0)n(h+2\delta) \leqq (1-\mu_B(C_+))n(h+2\delta)$$

to $H(\hat{P} \mid B)$, and C_+ contributes at most $\mu_B(C_+)n(h-\varepsilon)$. Thus $H(\hat{P} \mid B)$ is at most

$$\begin{aligned} & n\delta + (1-\mu_B(C_+))n(h+2\delta) + \mu_B(C_+)n(h-\varepsilon) \\ &= n(\delta + (1-\mu_B(C_+))(h+2\delta) + \mu_B(C_+)(h-\varepsilon)) \\ &= n(\delta + (h+2\delta) - \mu_B(C_+)(h+2\delta) + \mu_B(C_+)(h-\varepsilon)) \\ &= n((h+3\delta) - \mu_B(C_+)(2\delta+\varepsilon)) \\ &\leqq n(h+3\delta-\mu_B(C_+)\varepsilon).\end{aligned}$$

But our atom B was chosen so that $H(\hat{P} \mid B) > n(h-\delta)$. Hence

$$h-\delta < h+3\delta-\mu_B(C_+)\varepsilon, \quad \mu_B(C_+)\varepsilon < 4\delta, \quad \text{and} \quad \mu_B(C_+) < \frac{4\delta}{\varepsilon} < \frac{\varepsilon}{4}.$$

It follows that

$$\begin{aligned}\mu_B(C_0) &= 1-\mu_B(\bar{C})-\mu_B(C_-)-\mu_B(C_+) \\ &> 1-\delta_2-\delta_2-\varepsilon/4 \\ &> 1-\varepsilon/2.\end{aligned}$$

Thus, we have shown that on a set of fibers B of total measure $> 1-\varepsilon/2$, the conditional measure of the set of points $x \in B$ for which $\mu_B(P_1^n(x)) = e^{-n(h\pm\varepsilon)}$ is $> 1-\varepsilon/2$. This implies the desired result. □

THEOREM 1. *Let $\mathscr{B}$ be a sub-σ-algebra of $R_{-\infty}^0$. Fix $0<\varepsilon<1$. Then there exists a positive constant C such that for all n, the function $\mu(P_1^n(x) \mid \mathscr{B}(x))$ has value $e^{-n(h(P)\pm\varepsilon)\pm C}$ on a set of measure $> 1-\varepsilon$.*

PROOF. Theorem 1 holds for all n exceeding some large integer N; if we choose C large enough, the new error-factor $e^{\pm C}$ will take care of the (finitely many) cases $n \leqq N$. □

We now derive consequences about motley names of total length n_2 consisting of a P-block of length n_1 and a Q-block of length $n_2 - n_1$. That is, we look at the components of the partition $P_1^{n_1} \vee Q_{n_1+1}^{n_2}$, which we denote $P_1^k \vee Q_{k+1}^n$ for convenience.

THEOREM 2. *Fix $0 < \varepsilon < 1$. Then for all n sufficiently large, and for all $0 \leqq k \leqq n$, there exists a set of fibers of $R_{-\infty}^0$ of total measure $> 1 - \varepsilon$, on each of which all but ε in conditional measure of the components of $P_1^k \vee Q_{k+1}^n$ have mass $e^{-n(\theta \pm \varepsilon)}$, with*

$$\theta = \frac{1}{n}(kh(P) + (n-k)h(Q))$$

(*note that θ depends on k*).

PROOF. Let $\delta = \varepsilon^2/2$, and write

$$f(x) = \mu(P_1^k(x) \mid R_{-\infty}^0(x)),$$

$$g(x) = \mu(Q_1^{n-k}(x) \mid (R_{-\infty}^{-k} \vee P_{-k+1}^0)(x)).$$

Applying Theorem 1′ to P with $\mathscr{B} = R_{-\infty}^0$, we find C such that $f(x)$ has value $e^{-k(h(P) \pm \delta) \pm C}$ on a set of measure $> 1 - \delta$. Also, applying Theorem 1′ to Q with $\mathscr{B} = R_{-\infty}^{-k} \vee P_{-k+1}^0 \subseteq (P \vee R)_{-\infty}^0$, we find C' such that $g(x)$ has value $e^{-(n-k)(h(Q) \pm \delta) \pm C'}$ on a set of measure $> 1 - \delta$. Thus the function

$$\mu((P_1^k \vee Q_{k+1}^n)(x) \mid R_{-\infty}^0(x)) = f(x)g(T^k x)$$

has value

$$e^{-k(h(P) \pm \delta) \pm C} e^{-(n-k)(h(Q) \pm \delta) \pm C'} = e^{-n\theta \pm (n\delta + C + C')}$$

on a set of measure $> 1 - 2\delta = 1 - \varepsilon^2$. The error exponent $n\delta + C + C'$ may be replaced by $2n\delta$ (which may in turn be replaced by $n\varepsilon$) provided $C + C' \leqq n\delta$; that is, provided $n \geqq (C + C')/\delta$. The fact that the set of "good" x's has measure $> 1 - \varepsilon^2$ implies that for a set of fibers $B \in R_{-\infty}^0$ of total measure $> 1 - \varepsilon$, the set of good points has conditional measure $> 1 - \varepsilon$ on B. □

The same argument more generally proves:

THEOREM 2′. *Consider partitions* $P^{(1)}, P^{(2)}, \ldots, P^{(r)}$, *with entropies* $h_j = h(T, P^{(j)})$ $(1 \leqq j \leqq r)$, *and fix* $0 < \varepsilon < 1$. *Then for all natural numbers* $k_1, k_2, \ldots, k_r$ *with sum n sufficiently large (where we put* $n_j = k_1 + \cdots + k_j$ *for convenience), there exists a set of fibers of* $R^0_{-\infty}$ *of total measure* $> 1 - \varepsilon$, *on each of which all but* ε *in conditional measure of the components of*

$$P^* = \bigvee_{j=1}^{r} \bigvee_{i=n_{j-1}+1}^{n_j} T^{-i}P^{(j)}$$

have mass $e^{-n(\theta \pm \varepsilon)}$, *with* $\theta = (1/n)\Sigma_{j=1}^{r} k_j h_j$.

Now we need to be able to "re-order the quantifiers" in Theorem 2, so that the set of exceptional x's does not depend on k. We cannot simply lump together the exceptional sets determined by different values of k; these sets are of size ε but there are n of them.

Here is the key idea of the proof, in the especially tractable case that P is a refinement of Q: Since the partitions $P_1^k \vee Q_{k+1}^n = P_1^k \vee Q_1^n$ $(k = 1, \ldots, n)$ form an increasingly refined nested family, we see that for each $x \in X$, the mass of $(P_1^k \vee Q_{k+1}^n)(x)$ is a decreasing function of k. Fix x, and suppose that some particular value k is "bad" with respect to x, so that for that k, $(P_1^k \vee Q_{k+1}^n)(x)$ is either too big or too small in mass by a factor of at least $e^{n\varepsilon}$. If the mass of $(P_1^k \vee Q_{k+1}^n)(x)$ is too small in comparison with its mandated value, then $k + 1, k + 2, \ldots$ are nearly as bad as k, since the mass of $(P_1^k \vee Q_{k+1}^n)(x)$ can only decrease as $k \to n$ while the mandated value of the mass decreases slowly (as a linear function of k); on the other hand, if the mass of $(P_1^k \vee Q_{k+1}^n)(x)$ is too big, then $k - 1, k - 2, \ldots$ are nearly as bad as k for the same sort of reason. In either case, bad behavior at one particular value of k implies almost-as-bad behavior for many neighboring values. If we "station" some special values of k along the way from 0 to n, then ensuring good behavior at the special values suffices to ensure almost-as-good behavior everywhere. The number of such equally-spaced stations required in order to guarantee good behavior everywhere is independent of n, so that we can lump together the exceptional sets that correspond to the special k and still keep control over the measure of the union of these sets.

The above idea works equally well when Q is a refinement of P. In the general case, when neither P nor Q is a refinement of the other, an extra trick is required: we need to put a "buffer zone" in between the two blocks, and appeal to Theorem 2′ with $r = 3$. The details appear below.

THEOREM 3. *Fix $0 < \varepsilon < 1$. Then for all n sufficiently large, there exists a set of fibers $B \in R^0_{-\infty}$ of total measure $> 1 - \varepsilon$, on each of which it is the case that for all but ε in conditional measure of the points $x \in B$,*

$$\mu_B((P_1^k \vee Q_{k+1}^n)(x)) = e^{-n(\theta_k \pm \varepsilon)}$$

for all $k < n$, where

$$\theta_k = \frac{1}{n}(kh(P) + (n-k)h(Q)).$$

PROOF. Let B be one of the good fibers from Theorem 2.† Let K be some positive integer exceeding $8h(P \vee Q)/\varepsilon$, and let $\delta = \varepsilon/2(K+1)$. For all n, let $L = L_n = [n/K]$, so that

$$L \leqq \frac{n}{K} < n\left(\frac{\varepsilon}{8h(P \vee Q)}\right)$$

and

$$L(2h(P \vee Q) - h(P) - h(Q)) \leqq L(2h(P \vee Q)) < n\varepsilon/4. \tag{1}$$

For each n, set

$$R^{(k)} = P_1^k \vee Q_{k+1}^n, \quad R_-^{(k)} = P_1^{k+L} \vee Q_{k-L+1}^n, \quad \text{and} \quad R_+^{(k)} = P_1^{k-L} \vee Q_{k+L+1}^n,$$

where L depends on n. For all x, we have

$$R_-^{(k)}(x) \subseteq R^{(k)}(x) \subseteq R_+^{(k)}(x)$$

(hence the $+$, $-$ notation); indeed, whenever $|k' - k| \leqq L$, we have

$$R_-^{(k)}(x) \subseteq R^{(k')}(x) \subseteq R_+^{(k)}(x). \tag{2}$$

Call $x \in B$ *good* at k if $R^{(k)}(x)$ has μ_B-mass $e^{-n(\theta_k \pm \varepsilon)}$, and *very good* at k if $R_-^{(k)}(x)$ and $R_+^{(k)}(x)$ both have μ_B-mass $e^{-n(\theta_k \pm \varepsilon/2)}$. We will find a sparse but evenly-spread subset S of $[0, n]$ such that nearly all points x in B are very good at all $k \in S$; we will then show that if x is very good at all $k \in S$, it is good at all $k \in [0, n]$.

Applying Theorem 2′ to the three-block name

$$R_-^{(k)}(x) = P_1^{k-L} \vee (P \vee Q)_{k-L+1}^{k+L} \vee Q_{k+L+1}^n$$

† *Added in proof.* This is not well-defined, since k has not been specified. It is necessary that the fiber B be simultaneously good with respect to $k = [in]$ for all $0 \leqq i \leqq K$; but since K is fixed, this causes no problems.

with n sufficiently large, we see that for each individual k, the negative of the logarithm of the μ_B-mass of $R_-^{(k)}(x)$ is, for all but δ in μ_B-measure of the points $x \in B$, equal to

$$\begin{aligned}(k-L)h(P) &+ 2Lh(P \vee Q) + (n-k-L)h(Q) \pm n\varepsilon/4 \\ &= kh(P) + (n-k)h(Q) + L(2h(P \vee Q) - h(P) - h(Q)) \pm n\varepsilon/4 \\ &= kh(P) + (n-k)h(Q) \pm n\varepsilon/2 \qquad \text{(by (1))} \\ &= n(\theta_k \pm \varepsilon/2);\end{aligned}$$

this is,

$$\mu_B(R_-^{(k)}(x)) = e^{-n(\theta_k \pm \varepsilon/2)}$$

for all but δ in μ_B-measure of the points x. Similarly, if n is large enough, it must be true that for each individual k,

$$\mu(R_+^{(k)}(x)) = e^{-n(\theta_k \pm \varepsilon/2)}$$

for all but δ of the points x (this time we use Theorem 2′ applied to a motley partition built from P, the trivial partition, and Q). Therefore, for each k, all but 2δ of the points of B are very good at k; and it follows that all but $(K+1)\cdot 2\delta = \varepsilon$ of the points $x \in B$ are very good at 0, $[n/K]$, $[2n/K]$, ..., and $[Kn/K]$ $(=n)$. (That is, we choose stations at values of k in the set $S = \{[in/K] : 0 \leqq i \leqq K\}$.)

It suffices now to show that if x is very good at $k = [in/K]$, it is good at every k' with

$$\left[\frac{(i-1)n}{K}\right] < k' < \left[\frac{(i+1)n}{K}\right].$$

Fix such a k', and note that since $|k - k'| \leqq n/K$,

$$\begin{aligned}|\theta_k - \theta_{k'}| &= |k - k'| \frac{|h(P) - h(Q)|}{n} \\ &\leqq \frac{n}{K}\frac{h(P \vee Q)}{n} \\ &= \frac{h(P \vee Q)}{K} \\ &< \frac{\varepsilon}{8}.\end{aligned}$$

Assuming x is very good at k, the inclusions (2) imply that

$$\begin{aligned}\mu_B(R^{(k')}(x)) &\leqq \mu_B(R_+^{(k)}(x)) \\ &\leqq e^{-n(\theta_k-\varepsilon/2)} \\ &\leqq e^{-n(\theta_{k'}-\varepsilon)}\end{aligned}$$

and

$$\begin{aligned}\mu_B(R^{(k')}(x)) &\geqq \mu_B(R_-^{(k)}(x)) \\ &\geqq e^{-n(\theta_k+\varepsilon/2)} \\ &\geqq e^{-n(\theta_{k'}+\varepsilon)},\end{aligned}$$

whence x is good at k'.

Thus, we have shown that for all $x \in B$ outside a set of μ_B-measure $<\varepsilon$, x is good at all k in $[0, n]$. Since this works on all good fibers B, and since the good fibers occupy all but ε of the measure of X, we are done. □

The same reasoning also proves

THEOREM 3′. *Consider $P^{(1)}, P^{(2)}, \ldots, P^{(r)}$, with entropies $h_1, h_2, \ldots, h_r$ as in Theorem 1′, and fix $0<\varepsilon<1$. Then for all n sufficiently large, there exists a set of fibers B of $R^0_{-\infty}$ of total measure $>1-\varepsilon$, on each of which it is the case that for all but ε in μ_B-measure of the points $x \in B$, it happens that for all natural numbers $k_1, k_2, \ldots, k_r$ with sum n, all of the sets $P^*(x)$ (defined as in Theorem 2′) have respective masses $e^{-n(\theta \pm \varepsilon)}$ with $\theta = (1/n)\Sigma_{j=1}^r k_j h_j$.*

ACKNOWLEDGMENT

I thank Jack Feldman for his advice and encouragement throughout the research that led to these results.

REFERENCES

[M] B. McMillan, *The basic theorems of information theory*, Ann. Math. Stat. **24** (1953), 196–219.

[Pr] J. Propp, *Coding from the Past*, Ph.D. dissertation, University of California at Berkeley, April 1987.

[S] C. Shannon, *A mathematical theory of communication*, Bell Syst. Tech. J. **27** (1948), 379–423, 623–656.