The Annals of Probability
1976, Vol. 4, No. 2, 335-337

ON THE MINIMUM NUMBER OF FIXED LENGTH SEQUENCES WITH FIXED TOTAL PROBABILITY

By John C. Kieffer

University of Missouri at Rolla

Let $X_1, X_2, \cdots$ be a stationary sequence of B-valued random variables, where B is a finite set. For each positive integer n, and number λ such that $0 < \lambda < 1$, let $N(n, \lambda)$ be the cardinality of the smallest set $E \subset B^n$ such that $P[(X_1, X_2, \cdots, X_n) \in E] > 1 - \lambda$. An example is given to show that $\lim_{n\to\infty} n^{-1} \log N(n, \lambda)$ may not exist for some λ, thereby settling in the negative a conjecture of Parthasarathy.

Let $B = \{0, 1\}$. Let Ω be the space of all sequences $(x_1, x_2, \cdots)$ from B. Let $X_1, X_2, \cdots$ be the coordinate mappings from Ω to B; that is $X_i(x_1, x_2, \cdots) = x_i$, $i = 1, 2, \cdots$. Let $\mathscr{F}$ be the smallest sigma-field of subsets of Ω with respect to which $X_1, X_2, \cdots$ are measurable. For each n, let $\mathscr{F}_n$ be the sub-sigmafield of $\mathscr{F}$ generated by $X_1, X_2, \cdots, X_n$. Let $T: \Omega \to \Omega$ be the measurable map which is the one-sided shift on Ω; that is, $T(x_1, x_2, \cdots) = (x_2, x_3, \cdots)$. Let $\mathscr{P}$ be the collection of all probability measures P on $\mathscr{F}$ which are stationary with respect to T and such that $P[(X_1, X_2, \cdots, X_n) = b] > 0$ for every block $b \in B^n$, $n = 1, 2, \cdots$.

If $P \in \mathscr{P}$, n is a positive integer, and $0 < \lambda < 1$, let $N(n, \lambda, P)$ be the minimum cardinality of those sets $E \subset B^n$ such that $P[(X_1, X_2, \cdots, X_n) \in E] > 1 - \lambda$. Parthasarathy [2] has shown that for each $P \in \mathscr{P}$, $\lim_{n\to\infty} n^{-1} \log N(n, \lambda, P)$ exists except for at most a countable number of λ, $0 < \lambda < 1$. It has been conjectured ([2], page 81) that if $P \in \mathscr{P}$, then $\lim_{n\to\infty} n^{-1} \log N(n, \lambda, P)$ exists for every λ, $0 < \lambda < 1$. It is the purpose of this paper to provide a counterexample to this conjecture. We construct after Lemma 3 a $P \in \mathscr{P}$ such that $\lim_{n\to\infty} n^{-1} \log N(n, \frac{1}{2}, P)$ does not exist.

If $P \in \mathscr{P}$, let $P(X_1, X_2, \cdots, X_n)$ be the random variable with domain Ω such that $P(X_1, X_2, \cdots, X_n)(\omega) = P[X_1 = X_1(\omega), X_2 = X_2(\omega), \cdots, X_n = X_n(\omega)]$, $\omega \in \Omega$. In [2] the following strong version of the Shannon–McMillan theorem is developed: There exists a T-invariant measurable function $h: \Omega \to [0, 1]$ such that $\lim_{n\to\infty} -n^{-1} \log P(X_1, X_2, \cdots, X_n) = h$ in $L^1(P)$ for every $P \in \mathscr{P}$, where the logarithm is to base 2. If $P \in \mathscr{P}$, let P^* be the Borel probability measure on $[0, 1]$ which is the distribution of h relative to P; that is, $P^*(E) = P[h \in E]$, E a Borel set in $[0, 1]$. The mapping $P \to P^*$ is linear on the convex set $\mathscr{P}$. If $0 \leqq p \leqq 1$, let $\delta(p)$ be the Borel probability measure on $[0, 1]$ with support $\{p\}$. If $P \in \mathscr{P}$, let $H(P) = \int h\, dP$, the entropy of P. If P is ergodic with respect to the shift T, then $P^* = \delta(H(P))$.

Received December 26, 1974.

AMS 1970 *subject classifications.* Primary 60B05, 28A65, 28A35; Secondary 94A15.

Key words and phrases. Stationary measures, Shannon–McMillan theorem, shift transformation on a product space.

LEMMA 1. *If $P \in \mathscr{P}$, and $0 < H \leqq H(P)$, and n is a positive integer, there exists $Q \in \mathscr{P}$ such that $P = Q$ over $\mathscr{F}_n$ and $Q^* = \delta(H)$.*

PROOF. A probability measure Q on $\mathscr{F}$, stationary and ergodic with respect to T, exists such that $P = Q$ over $\mathscr{F}_n$ and $H(Q) = H$ ([1], Theorem 4). An examination of the proof of Theorem 4 of [1] will show in addition that the Q constructed there is in $\mathscr{P}$. Since Q is ergodic, $Q^* = \delta(H(Q))$.

For the following lemma, see [2], Theorem 3.1.

LEMMA 2. *If $0 < \lambda < 1$ and $P \in \mathscr{P}$, then*

$$\liminf_{n\to\infty} n^{-1} \log N(n, \lambda, P) \geqq \sup\{\alpha : P^*[0, \alpha] < 1 - \lambda\},$$

and

$$\limsup_{n\to\infty} n^{-1} \log N(n, \lambda, P) \leqq \inf\{\alpha : P^*[0, \alpha] > 1 - \lambda\}.$$

LEMMA 3. *Let the numbers p_1, p_2, ε satisfy*

(3a) $\frac{1}{4} < p_1 < \frac{3}{8}$; $\frac{3}{4} < p_2 < 1$; $0 < \varepsilon < \frac{1}{6}$;
(3b) $\frac{1}{4} < (1 - 2\varepsilon)p_1 + \varepsilon < \frac{3}{8}$; $\frac{3}{4} < (1 - 6\varepsilon)p_2 + 3\varepsilon < 1$.

Let $P_1, P_2 \in \mathscr{P}$ satisfy $(P_i)^* = \delta(p_i)$, $i = 1, 2$. Let $P = (\frac{1}{2} + \varepsilon)P_1 + (\frac{1}{2} - \varepsilon)P_2$. Then for any positive integer n, there exist integers $n_2 > n_1 > n$, measures $P_1', P_2' \in \mathscr{P}$, and numbers p_1', p_2', ε' such that:

(3c) *p_1', p_2', ε' satisfy* (3a) *and* (3b) *with p_1, p_2, ε replaced by p_1', p_2', ε';*
(3d) $(P_i')^* = \delta(p_i')$, $i = 1, 2$;
(3e) *If $P' = (\frac{1}{2} + \varepsilon')P_1' + (\frac{1}{2} - \varepsilon')P_2'$, then $P = P'$ over $\mathscr{F}_n$, $n_1^{-1} \log N(n_1, \frac{1}{2}, P') < \frac{3}{8}$, $n_2^{-1} \log N(n_2, \frac{1}{2}, P') > \frac{2}{5}$.*

PROOF. By Lemma 2, since $P^* = (\frac{1}{2} + \varepsilon)\delta(p_1) + (\frac{1}{2} - \varepsilon)\delta(p_2)$ and $p_1 < \frac{3}{8}$, there exists $n_1 > n$ such that $n_1^{-1} \log N(n_1, \frac{1}{2}, P) < \frac{3}{8}$. Now $P = (\frac{1}{2} - \varepsilon)P_1 + 4\varepsilon(\frac{1}{2}P_1 + \frac{1}{2}P_2) + (\frac{1}{2} - 3\varepsilon)P_2$. Since $H(\frac{1}{2}P_1 + \frac{1}{2}P_2) = \frac{1}{2}p_1 + \frac{1}{2}p_2 > \frac{1}{2}$ by (3a), there exists by Lemma 1 a measure $P_3 \in \mathscr{P}$ such that $P_3 = \frac{1}{2}P_1 + \frac{1}{2}P_2$ over $\mathscr{F}_{n_1}$ and $(P_3)^* = \delta(\frac{1}{2})$. Let $P_4 = (\frac{1}{2} - \varepsilon)P_1 + 4\varepsilon P_3 + (\frac{1}{2} - 3\varepsilon)P_2$. Then $P_4 = P$ over $\mathscr{F}_{n_1}$. By Lemma 2, there exists $n_2 > n_1$ such that $n_2^{-1} \log N(n_2, \frac{1}{2}, P_4) > \frac{2}{5}$. Let $p_1' = (\frac{1}{2} - \varepsilon)(\frac{1}{2} + \varepsilon')^{-1}p_1 + (\varepsilon + \varepsilon')(\frac{1}{2} + \varepsilon')^{-1}\frac{1}{2}$ and $p_2' = (3\varepsilon - \varepsilon')(\frac{1}{2} - \varepsilon')^{-1}\frac{1}{2} + (\frac{1}{2} - 3\varepsilon)(\frac{1}{2} - \varepsilon')^{-1}p_2$, where the number ε' is chosen so that $0 < \varepsilon' < \min(\frac{1}{6}, 3\varepsilon)$, $\frac{1}{4} < p_1' < \frac{3}{8}$, $\frac{3}{4} < p_2' < 1$, $\frac{1}{4} < (1 - 2\varepsilon')p_1' + \varepsilon' < \frac{3}{8}$, $\frac{3}{4} < 3\varepsilon' + (\frac{1}{2} - 3\varepsilon')2p_2' < 1$. It is possible to choose such an ε' because of condition (b). Thus p_1', p_2', ε' satisfy (a) and (b) with p_1, p_2, ε replaced by p_1', p_2', ε'. Now $P_4 = (\frac{1}{2} + \varepsilon')Q_1 + (\frac{1}{2} - \varepsilon')Q_2$, where $Q_1 = (\frac{1}{2} - \varepsilon)(\frac{1}{2} + \varepsilon')^{-1}P_1 + (\varepsilon + \varepsilon')(\frac{1}{2} + \varepsilon')^{-1}P_3$ and $Q_2 = (3\varepsilon - \varepsilon')(\frac{1}{2} - \varepsilon')^{-1}P_3 + (\frac{1}{2} - 3\varepsilon)(\frac{1}{2} - \varepsilon')^{-1}P_2$. For $i = 1, 2$, $Q_i \in \mathscr{P}$ and $H(Q_i) = p_i' > 0$; thus by Lemma 1 there exist $P_1', P_2' \in \mathscr{P}$ such that $P_i' = Q_i$ over $\mathscr{F}_{n_2}$ and $(P_i')^* = \delta(p_i')$, $i = 1, 2$. Let $P' = (\frac{1}{2} + \varepsilon')P_1' + (\frac{1}{2} - \varepsilon')P_2'$. Then $P' = P_4$ over $\mathscr{F}_{n_2}$, so $N(n_2, \frac{1}{2}, P') = N(n_2, \frac{1}{2}, P_4)$. Also $P' = P_4 = P$ over $\mathscr{F}_{n_1}$ so $N(n_1, \frac{1}{2}, P') = N(n_1, \frac{1}{2}, P)$. Thus conditions (c)—(e) are satisfied.

THE COUNTEREXAMPLE. We can apply Lemma 3 to construct a sequence $\{P_i\}_1^\infty$ in $\mathscr{P}$ and a strictly increasing sequence $\{n_i\}_{i=1}^\infty$ of positive integers such that

$P_{i+1} = P_i$ over $\mathscr{F}_{n_{2i}}$, $(n_{2i-1})^{-1} \log N(n_{2i-1}, \frac{1}{2}, P_i) < \frac{3}{8}$, $(n_{2i})^{-1} \log N(n_{2i}, \frac{1}{2}, P_i) > \frac{2}{5}$, $i = 1, 2, \cdots$. It is easy to see, using the Kolmogorov extension theorem, that there exists a unique $P \in \mathscr{P}$ such that $P = P_i$ over $\mathscr{F}_{n_{2i}}$, $i = 1, 2, \cdots$. Thus for each i, $N(n_{2i-1}, \frac{1}{2}, P) = N(n_{2i-1}, \frac{1}{2}, P_i)$ and $N(n_{2i}, \frac{1}{2}, P) = N(n_{2i}, \frac{1}{2}, P_i)$. Consequently, $\lim_{n\to\infty} n^{-1} \log N(n, \frac{1}{2}, P)$ does not exist.

REFERENCES

[1] Kieffer, J. C. (1974). On the approximation of stationary measures by periodic and ergodic measures. *Ann. Probability* **2** 530–534.

[2] Parthasarathy, K. R. (1963). Effective entropy rate and transmission through channels with additive random noise. *Sankhyā Ser. A* **25** 75–84.

Department of Mathematics
326 Mathematics-Computer Science Building
University of Missouri
Rolla, Missouri 65401