AN EXPONENTIAL BOUND FOR FUNCTIONS OF A MARKOV CHAIN[1]

By Melvin Katz[2] Jr. and A. J. Thomasian

University of California, Berkeley

1. Introduction. An explicit and relatively simple exponential bound is obtained for $P\{|n^{-1}\sum_{k=1}^{n} f(X_k) - \mu| \geqq \epsilon \text{ for some } n \geqq m\}$, where $X_1, X_2, \cdots$ is a finite state ergodic Markov chain with arbitrary initial distribution, f is any real-valued function, and μ is the expected value of $f(X_1)$ computed under the unique initial stationary measure. Bounds for the one-sided inequalities are also given. Because the assumptions are weak and permit the transition matrix to contain zeroes, the result can be applied to multiple Markov chains (Doob [4], p. 185) and thus to sums of the form $S_n = \sum_{k=1}^{n} f(X_k, X_{k+1}, X_{k+2})$. The proof employs methods of recurrent event theory that have been used by Chung [2], and Doblin [3]. Asymptotic results for the one-sided inequalities have been obtained by Koopmans [7], under the restriction that the transition matrix contains no zeroes. Some possible applications for such bounds can be seen in Chernoff [1], Khinchin [6], and Koopmans [7].

2. Notation and summary. Let $P = (p_{ij})$ be an $r \times r$ stationary transition matrix with $r \geqq 2$ and, using the terminology in Doob [4], assume that there is only one ergodic class of states $E \subset R = \{1, 2, \cdots, r\}$. Let $T \subset R$ be the (possibly empty) class of transient states and let $p_1, \cdots, p_r$ be the stationary distribution for P. We denote the smallest positive element of P by p. Let X_1 $X_2, \cdots$ be the Markov chain determined by P and an arbitrary initial distribution for X_1; so that $X_n = j$ if the process is in state j at time n. Now let f be a real-valued function on R and let $S_n = \sum_{k=1}^{n} f(X_k)$, $\mu = \sum_{k=1}^{r} p_k f(k)$, and $M = \max_{k \epsilon R} f(k) - \min_{j \epsilon R} f(j)$.

The notation and assumptions which have been made will be used throughout the paper except for the countable state space example at the end. We can now state the following

THEOREM. *Let m be a positive integer and let $\epsilon > 0$. Then*

$$P\{|n^{-1}S_n - \mu| \geqq \epsilon \quad \text{for some} \quad n \geqq m\} \leqq 2Ae^{-B\epsilon^2 m}$$

$$P\{S_n \geqq n(\mu + \epsilon) \quad \text{for some} \quad n \geqq m\} \leqq Ae^{-B\epsilon^2 m}$$

$$P\{S_n \leqq n(\mu - \epsilon) \quad \text{for some} \quad n \geqq m\} \leqq Ae^{-B\epsilon^2 m}$$

Received October 8, 1959; revised January 6, 1960.

[1] This paper was prepared with the partial support of the Office of Naval Research (Nonr-222-53). This paper in whole or in part may be reproduced for any purpose of the United States Government.

[2] Now at the University of Chicago.

where

$$A = \frac{8r}{p^r} \frac{1}{1 - e^{-B\epsilon^2}}, \qquad B = \frac{p^{3r}}{2^8 M^2 r^2}.$$

3. Proof of the theorem. We first introduce some additional notation. For $i \varepsilon E$ let $V_0^i = 0$ and for $k = 1, 2, \cdots$ let V_k^i be the time to the kth occurrence of state i. Thus $V_k^i = l$ if $X_l = i$ and exactly k of $X_1, \cdots, X_l$ are equal to i. The V_k^i are defined with probability one. Now for $k = 1, 2, \cdots$ we let

$$v_k^i = V_k^i - V_{k-1}^i, \qquad u_k^i = \sum_{j=(V_{k-1}^i)+1}^{V_k^i} f(X_j), \qquad U_s^i = \sum_{k=1}^{s} u_k^i = \sum_{j=1}^{V_s^i} f(X_j).$$

It is well known that $v_1^i, v_2^i, \cdots$ are independent and $v_2^i, v_3^i, \cdots$ are identically distributed. Similarly, we have $u_1^i, u_2^i, \cdots$ are independent and $u_2^i, u_3^i, \cdots$ are identically distributed.

The basic idea behind the proof is to express the event $[S_n \geqq n(\mu + \epsilon)$ for some $n \geqq m]$ as a subset of a larger event and for the larger event employ the properties of the random variables U_s^i and V_s^i in order to obtain a bound on the probability of its occurrence. Once this bound is obtained the other two follow quite readily and the proof is complete. To obtain the initial result it is necessary to prove the following lemma. The bound and the method of proof of the lemma are similar to S. Bernstein's inequality, Uspensky [9].

LEMMA. *If* $i \varepsilon E$, $M \leqq 1/r$, $-4 \leqq \mu < 0$, $\delta \geqq 0$ *then*

$$P\{U_s^i \geqq -\mu\delta\} \leqq \frac{4}{p^r} e^{-\beta\mu^2(\delta+s-1)}$$

where $\beta = p^{3r}/2^6$.

PROOF. The state $i \varepsilon E$ will be arbitrary but fixed throughout the proof of the lemma so that we will not exhibit it in v_k^i, u_k^i and U_k^i. We take $t > 0$ and apply a known inequality (Loeve [8], p. 158, (1)) to obtain

$$P\{U_s \geqq -\mu\delta\} \leqq Ee^{t(\mu\delta + U_s)} = e^{t\mu\delta} Ee^{tu_1}(Ee^{tu_2})^{s-1}.$$

Now let $a = 1 - p^r$ so that $0 \leqq a < 1$ and for small enough $t > 0$ we have $ae^t < 1$. It is known (e.g., Feller [5], p. 378) that for $k = 0, 1, \cdots$ (defining $0^0 = 0$) $P\{v_1 > kr\} \leqq a^k$. Since the distribution of X_1 is arbitrary it is clear that the same inequality applies to v_2. We now bound Ee^{tu_1}

$$Ee^{tu_1} = \sum_{k=1}^{\infty} E(e^{tu_1} \mid v_1 = k)P(v_1 = k) \leqq \sum_{k=1}^{\infty} e^{(t/r)k} P(v_1 = k).$$

The inequality follows from the fact that $u_1 \leqq v_1 \max_{j\varepsilon R} f(j)$ and, since $\mu < 0$, $\min_{j\varepsilon R} f(j) < 0$. Thus $\max_{j\varepsilon R} f(j) < M \leqq 1/r$ and therefore $e^{tu_1} \leqq e^{(t/r)v_1}$ Further

$$\sum_{k=1}^{\infty} e^{(t/r)k} P(v_1 = k) = \sum_{k=1}^{r} e^{(t/r)k} P(v_1 = k) + \sum_{k=r+1}^{2r} e^{(t/r)k} P(v_1 = k) + \cdots$$

$$\leqq \sum_{k=1}^{\infty} e^{kt} a^{k-1} = \frac{e^t}{1 - ae^t}.$$

Therefore $Ee^{tu_1} \leqq e^t/(1 - ae^t)$. Taking a finite Taylor's expansion of Ee^{tu_2}, we get

$$Ee^{tu_2} = 1 + tEu_2 + \frac{t^2}{2} E(u_2)^2 e^{t'u_2} \qquad 0 < t' < t.$$

Using the same method as was used on Ee^{tu_1}, we obtain

$$\begin{aligned} E(u_2)^2 e^{t'u_2} &= \sum_{k=1}^{\infty} E[(u_2)^2 e^{t'u_2} \mid v_2 = k] P(v_2 = k) \\ &\leqq \sum_{k=1}^{\infty} \left(\frac{k}{r}\right)^2 e^{(t/r)k} P(v_2 = k) \\ &\leqq e^t \sum_{k=1}^{\infty} k^2 (e^t a)^{k-1} = e^t \frac{1 + ae^t}{(1 - ae^t)^3}. \end{aligned}$$

We now let $t = -2\mu\beta$ so that $t \leqq p^r/8 \leqq \frac{1}{8}$ and therefore $e^t \leqq 2$. Thus

$$e^t \leqq 1 + 2t \text{ and } ae^t = (1 - p^r)e^t \leqq 1 - \frac{3p^r}{4} + \frac{p^{2r}}{4} \leqq 1 - \frac{p^r}{2}.$$

Hence $ae^t < 1$ and $1 - ae^t \geqq p^r/2$. Using these facts we see that $Ee^{tu_1} \leqq 4/p^r$ and $E(u_2)^2 e^{t'u_2} \leqq 1/2\beta$. Now from Chung [2] we have $Eu_2 = \mu Ev_2 \leqq \mu < 0$ so that

$$Ee^{tu_2} \leqq e^{t\mu + (t^2/4\beta)} = e^{-\beta\mu^2}$$

and clearly

$$e^{t\mu\delta} = e^{-2\mu^2\beta\delta} \leqq e^{-\mu^2\beta\delta}.$$

So the proof of the lemma is complete.

We turn now to the proof of the theorem and note that $[S_n \geqq n(\mu + \epsilon)$ for some $n \geqq m]$ is contained in

$$\left\{\bigcup_{n \geqq m} [X_n \in T]\right\} \cup \left\{\bigcup_{i \in E} \bigcup_{n \geqq m} [S_n \geqq n(\mu + \epsilon), X_n = i]\right\}$$

so that

$$\begin{aligned} P\{S_n \geqq n(\mu + \epsilon) \quad \text{for some} \quad n \geqq m\} &\leqq P\{\bigcup_{n \geqq m} [X_n \in T]\} \\ &\quad + r \max_{i \in E} P\{\bigcup_{n \geqq m} [S_n \geqq n(\mu + \epsilon), X_n = i]\}. \end{aligned}$$

Now clearly

$$\bigcup_{n \geqq m} [S_n \geqq n(\mu + \epsilon), X_n = i] = \bigcup_{s \geqq 1} [U_s^i \geqq V_s^i(\mu + \epsilon), V_s^i \geqq m]$$

so that for $i \in E$

$$P\left\{\bigcup_{n\geqq m}[S_n \geqq n(\mu+\epsilon), X_n = i]\right\} \leqq \sum_{s=1}^{\infty} P\left\{U_s^i - V_s^i(\mu+\epsilon) \geqq 0, \frac{\epsilon}{2} V_s^i \geqq \frac{\epsilon}{2} m\right\}$$

$$\leqq \sum_{s=1}^{\infty} P\left\{U_s^i - V_s^i\left(\mu + \frac{\epsilon}{2}\right) \geqq \frac{\epsilon}{2} m\right\}$$

$$\leqq \sum_{s=1}^{\infty} P\left\{\frac{1}{Mr}\left[U_s^i - V_s^i\left(\mu + \frac{\epsilon}{2}\right)\right] \geqq \frac{\epsilon}{2Mr} m\right\}$$

$$\leqq \sum_{s=1}^{\infty} \frac{4}{p^r} e^{-\beta(\epsilon/2Mr)^2(m+s-1)} = \frac{1}{2r} A e^{-B\epsilon^2 m},$$

where for the last inequality we applied the lemma with $f(k)$ replaced by

$$g(k) = \frac{f(k) - (\mu + \epsilon/2)}{Mr}$$

and used

$$\max_{k\varepsilon R} g(k) - \min_{j\varepsilon R} g(j) = \frac{1}{r}, \qquad \sum_{k=1}^{r} g(k)p_k = -\frac{\epsilon}{2Mr} \geqq -\frac{1}{r},$$

where $\epsilon \leqq M$ is assumed, since the theorem is trivial for $\epsilon > M$. Since $(A/2) > 1$ we need only show that

$$P\left\{\bigcup_{n\geqq m}[X_n \varepsilon T]\right\} \leqq e^{-BM^2 m}$$

in order to obtain the bound for $P\{S_n \geqq n(\mu + \epsilon)$ for some $n \geqq m\}$.

Now we note that for $i \varepsilon E$

$$P\left\{\bigcup_{n\geqq m}[X_n \varepsilon T]\right\} \leqq P\{v_1^i > m\}$$

and, recalling that $P\{v_1^i > kr\} \leqq a^k = (1 - p^r)^k$, we obtain

$$P\left\{\bigcup_{n\geqq m}[X_n \varepsilon T]\right\} \leqq (1 - p^r)^{(m/r)-1}.$$

Now if $\epsilon \leqq M$ and $(m/r) \leqq 2^8$ we see that $Ae^{-B\epsilon^2 m} \geqq 16e^{-(m/2^8r^2)} \geqq 16e^{-1} > 1$, making the theorem trivial in this case. Thus we may assume that $(m/r) > 2^8$ so that $(m/r)(1 - 2^{-8}) \geqq 1$ and so

$$\frac{m}{r} - 1 \geqq \frac{m}{2^8 r} \geqq \frac{1}{p^r} BM^2 m.$$

Thus for $i \varepsilon E$

$$P\left\{\bigcup_{n\geqq m}[X_n \varepsilon T]\right\} \leqq (1 - p^r)^{(m/r)-1} \leqq e^{-p^r[(m/r)-1]} \leqq e^{-BM^2 m}$$

and we have the bound for $P\{S_n \geqq n(\mu + \epsilon)$ for some $n \geqq m\}$.

The other two inequalities of the theorem now follow immediately. We define the function $g(k) = -f(k)$ for $k = 1, \cdots, r$. Then $-\mu = \sum_{k=1}^{r} p_k g(k)$ and

$$P\left\{\sum_{j=1}^{n} f(X_j) \leqq n(\mu - \epsilon) \quad \text{for some} \quad n \geqq m\right\} =$$

$$P\left\{\sum_{j=1}^{n} g(X_j) \geqq n(-\mu + \epsilon) \quad \text{for some} \quad n \geqq m\right\} \leqq Ae^{-B\epsilon^2 m}.$$

For the remaining inequality we have that $P\{|n^{-1}S_n - \mu| \geqq \epsilon$ for some $n \geqq m\} \leqq P\{S_n \leqq n(\mu - \epsilon)$ for some $n \geqq m\} + P\{S_n \geqq n(\mu + \epsilon)$ for some $n \geqq m\} \leqq 2Ae^{-B\epsilon^2 m}$. This completes the proof of the theorem.

We close with an example of a Markov chain with a countable state space for which there exists no constants A and B such that $P\{S_n \geqq n(\mu + \epsilon)$ for some $n \geqq m\} \leqq Ae^{-B^2 m}$. Let the state space be $R = \{1, 2, \cdots\}$ and for $j \in R$ let $p_{1j} = c/j^3$ where $\sum_1^\infty j^{-3} = c^{-1}$. For $j > 1$ let $p_{j,j-1} = 1$ and $p_{jk} = 0$ for $k \neq j - 1$. Then this matrix $P = (p_{ij})$ admits a unique stationary distribution $p_1, p_2, \cdots$ (Feller [5]) with each $p_j > 0$, and we shall assume that it is our initial distribution. We define f on R by $f(1) = 0$ and $f(j) = 1$ for $j > 1$, so that $\mu = \sum_1^\infty f(j)p_j = 1 - p_1 < 1$. Just as before, we let $S_n = \sum_1^n f(X_j)$ and we shall show that if $\alpha = 1$ then lim sup $(1/n)$ log $P\{S_n \geqq n\alpha\} = 0$. We define a subsequence $\{m(k): k = 1, 2, \cdots\}$ of integers by

$$m(k) = 1 + 2 + \cdots + k = k(k + 1)/2.$$

For each k we define a sequence of $m(k)$ states by $w_{m(k)} = (1, 2, 1, 3, 2, 1, 4, 3, 2, 1, \cdots, k, k - 1, \cdots, 1)$. Clearly $(1/m(k))S_{m(k)}$ evaluated at $w_{m(k)}$ is equal to $(m(k) - k)/m(k)$ which converges to 1. Also $P(w_{m(k)}) = p_1 C^{k-1}(k!)^{-3}$ so that lim sup (n^{-1}) log $P(S_n \geqq n\alpha) \geqq$ lim $(1/m(k))$ log $P(w_{m(k)}) = 0$.

REFERENCES

[1] Herman Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 493–507.

[2] K. L. Chung, "Contributions to the theory of Markov chains. II," *Trans. Amer. Math. Soc.*, Vol. 76, No. 3 (1954), pp. 397–419.

[3] W. Doblin, "Sur deux problèmes de M. Kolmogoroff concernant les chaines denombrables," *Bull. Soc. Math. France*, Vol. 52 (1938), pp. 210–220.

[4] J. L. Doob, *Stochastic Processes*, John Wiley and Sons, New York, 1953.

[5] William Feller, *An Introduction to Probability Theory and Its Applications*, 2nd ed., John Wiley and Sons, New York, 1957.

[6] A. I. Khinchin, *Mathematical Foundations of Information Theory*, Dover Pub., New York, 1957.

[7] Lambert Herman Koopmans, "Asymptotic rate of discrimination for Markov processes," unpublished Ph.D. thesis, University of California, Berkeley, 1958.

[8] Michel Loeve, *Probability Theory*, D. Van Nostrand, New York, 1955.

[9] J. V. Uspensky, *Introduction to Mathematical Probability*, McGraw-Hill Book Co., New York, 1937.