

AN ARCSINE LAW FOR MARKOV CHAINS

DAVID A. FREEDMAN¹

1. **Introduction.** Suppose $\{x_n\}$ is a sequence of independent, identically distributed random variables with mean 0. Under certain mild assumptions [4], the arcsine law gives the limiting distribution of the relative frequency of positive partial sums; the probability that this relative frequency is $\leq x$ converges to $2\pi^{-1} \arcsin \sqrt{x}$, or more generally to an incomplete beta function.

This result may be extended to cover functionals of Markov chains with stationary transition probabilities. Precise conditions are given in Theorem 1 of §2 and Theorem 2 of §4.

The basic analytic technique, of partitioning the sample sequence according to the occurrence of a fixed state, is due to Doeblin [2]. This paper follows the notation and theory of [1]. A brief resume of the pertinent facts is given in §2; §3 gives proofs, and §4 generalizations.

2. **The arcsine law for Markov chains.** Let $\{x_n: n=0, 1, 2, \dots\}$ be a Markov chain with state space I . Let f be a real function on I . The process $y_n = f(x_n)$ is called functional on the chain $\{x_n\}$. Suppose $x_0 = i$. Let τ_ν be the time of the ν th i , so that $\tau_1 = 0$. Let $\rho_\nu = \tau_{\nu+1} - \tau_\nu$ for $\nu \geq 1$. Suppose the chain is positive, with stationary initial distribution $\{\pi_j: j \in I\}$; $m_i = E\rho_\nu = \pi_i^{-1}$, $0 < m_i < \infty$. The subscript ν will be used for random variables defined on $[\tau_\nu, \tau_{\nu+1})$, and for functions on the first ν elements of such sequences. Other types of sequence will be indexed by m, n, j . The letter i is reserved for states. By the strong Markov property, random variables defined on different $[\tau_\nu, \tau_{\nu+1})$ are independent. In particular, $\{\rho_\nu: \nu = 1, 2, \dots\}$ is a sequence of independent, identically distributed random variables, and any random variable with index ν is independent of $\{\rho_{\nu+1}, \rho_{\nu+2}, \dots\}$. Let

$$Y_\nu = \sum_{s=\tau_\nu}^{\tau_{\nu+1}-1} f(x_s); \quad \nu \geq 1.$$

Then if $\sum_{i \in I} \pi_i |f(i)| < \infty$, $M = \pi_i E(Y_\nu)$ is finite and independent of i . Let $\tilde{f} = f - M$, $Z_\nu = Y_\nu - M\rho_\nu$. The Z_ν are a sequence of independent, identically distributed random variables. Write $\hat{S}_\nu = Z_1 + \dots + Z_\nu$. Finally put

Received by the editors April 27, 1962 and, in revised form May 21, 1962.

¹ Prepared with the partial support of the National Science Foundation, Grant G-14648.

$$S_n = \sum_{j=0}^n f(x_j), \quad \bar{S}_n = \sum_{j=0}^n \bar{f}(x_j),$$

and let $r_j = 1$ when $\bar{S}_j > 0$, $= 0$ when $\bar{S}_j \leq 0$. Define

$$r(n) = n^{-1} \sum_{j=0}^{n-1} r_j.$$

Then

THEOREM 1. *Suppose $\{x_n\}$ is a Markov chain with stationary transition probabilities and state space I and f is a real function on I such that*

- (i) I is a positive recurrent class,
- (ii) $\sum_{i \in I} \pi_i |f(i)| < \infty$,
- (iii) $\lim_{n \rightarrow \infty} n^{-1} \sum_{j=1}^n P(\bar{S}_j > 0) = \alpha$.

Then

$$\lim_{n \rightarrow \infty} P(r(n) \leq x) = F_\alpha(x)$$

where

$$\left. \begin{aligned} F_\alpha(x) &= 0 & x < 0; & = 1, x \geq 0 \text{ for } \alpha = 0 \\ &= 0 & x < 0; & = 1, x \geq 1 \\ &= \pi^{-1} \sin \pi \alpha \int_0^x s^{\alpha-1} (1-s)^{-\alpha} ds, & 0 \leq x \leq 1 & \left. \right\} \text{ for } 0 < \alpha < 1 \\ &= 0, x < 1; & = 1, x \geq 1 \text{ for } \alpha = 1. \end{aligned} \right\}$$

3. Proofs. The proof of Theorem 1 proceeds via several lemmas.

LEMMA 1. *Let F be a distribution function not degenerate at 0; F^n its n th convolution with itself and $a < b$ real numbers. Then $F^n(b) - F^n(a) \rightarrow 0$.*

PROOF. Compute the integral characteristic function of F^n .

LEMMA 2. *Let $d_\nu = 1$ when $|\hat{S}_\nu| \leq k$, $= 0$ when $|\hat{S}_\nu| > k$. Define $d(n) = n^{-1} \sum_{\nu=1}^n d_\nu$. Then $d(n) \rightarrow P_0$.*

PROOF. $E d(n) = n^{-1} \sum_{\nu=1}^n P(|\hat{S}_\nu| \leq k)$.

Apply Lemma 1 and Cauchy's theorem.

LEMMA 3. *Let $\{\mathcal{F}_\nu: 1 \leq \nu < \infty\}$ be an increasing sequence of σ -fields such that $\{\rho_i, 1 \leq i \leq \nu-1\}$ is \mathcal{F}_ν -measurable and $\{\rho_\nu, \rho_{\nu+1}, \dots\}$ is independent of \mathcal{F}_ν . Let $\{U_\nu: 1 \leq \nu < \infty\}$ be a strictly increasing sequence of natural number-valued random variables, with $[U_\nu = k] \in \mathcal{F}_k$. Then $\hat{\rho}_\nu = \rho_{U_\nu}$ is a random variable, and $\{\hat{\rho}_\nu: \nu = 1, 2, \dots\}$ is distributed like $\{\rho_\nu: \nu = 1, 2, \dots\}$.*

PROOF. This is an easy generalization of Theorem 5.2 on p. 145 of [3].

LEMMA 4. Let $\{t_\nu: \nu=1, 2, \dots\}$ be a sequence of random variables taking the values 0 and 1, such that:

- (i) $\{t_1, \dots, t_\nu\}$ is $\mathfrak{F}_{\nu+1}$ -measurable;
- (ii) $\sum_{\nu=1}^n t_\nu \rightarrow \infty$ a.s.;
- (iii) $n^{-1} \sum_{\nu=1}^n t_\nu \rightarrow a.s.t.$

Then

$$n^{-1} \sum_{\nu=1}^n t_\nu \rho_{\nu+1} \xrightarrow[d]{a.s.} m_i.$$

PROOF. Let $T_n = \sum_{\nu=1}^n t_\nu$, and define $U_\nu = 1 + \text{smallest } n \text{ such that } T_n = \nu$. Then $\{\rho_\nu\}, \{U_\nu\}$ satisfy the conditions of Lemma 3, so that $\{\hat{\rho}_\nu\} = \{\rho_{U_\nu}\}$ behaves like $\{\rho_\nu\}$. Then

$$n^{-1} \sum_{\nu=1}^n t_\nu \rho_{\nu+1} = n^{-1} T_n T_n^{-1} \sum_{\nu=1}^{T_n} \hat{\rho}_\nu,$$

and by the strong law of large numbers,

$$T_n^{-1} \sum_{\nu=1}^{T_n} \hat{\rho}_\nu \xrightarrow[d]{a.s.} m_i.$$

This completes the argument.

The proof of the theorem is obtained by putting together these facts; the crucial one being Lemma 4.

To begin with, by the dominated convergence theorem, it suffices to consider only initial distributions concentrated at one point. Suppose then that $x_0 = i$. Let $n^* = [\pi_i n] + 1$. Define $t_\nu = 1$ or 0 according as $\hat{S}_\nu > 0$ or ≤ 0 ; $d_\nu = 1$ or 0 according as $|\hat{S}_\nu| \geq K$ or $< K$, where K will be chosen later; $g_\nu = 1$ or 0 according as

$$w_\nu = 1 + \sum_{s=\tau_\nu}^{\tau_{\nu+1}-1} |\bar{f}(x_s)| \leq K/2 \quad \text{or} \quad > K/2.$$

Fix $\epsilon, \delta > 0$, and consider the sets

$$\Lambda_0 = \{n^{-1} |(\tau_{n^*+1} - 1) - n| \leq \delta\},$$

$$\Lambda_1 = \left\{ n^{-1} \sum_{\nu=1}^{n^*-1} d_\nu \rho_{\nu+1} \leq \delta \right\},$$

$$\Lambda_2 = \left\{ n^{-1} \sum_{\nu=1}^{n^*} g_\nu \rho_\nu \geq 1 - \delta \right\},$$

$$\Lambda_3 = \{n^{-1} \rho_1 \leq \delta\}.$$

Now by choosing K sufficiently large, it will follow that there exist natural numbers n_j such that $n \geq n_j$ will imply $P(\Lambda_j) \geq 1 - \epsilon$, $j=0, 1, 2, 3$. Indeed, for $j=0$ observe that $n^{-1}\tau_{n^{*+1}} = (n^{*-1}\tau_{n^{*+1}})(n^{-1}n^*) = n^{*-1}(\rho_1 + \dots + \rho_{n^*})n^{-1}([\pi; n] + 1) \rightarrow m_i \pi_i = 1$ by the strong law of large numbers. When $j=1$ the estimate follows from Lemmas 2 and 3, with $\mathfrak{F}_\nu = \mathfrak{F}\{x_s, s \leq \tau_\nu\}$, in the sense of [1, p. 72]. For $j=2$, use the fact that $n^{-1} \approx \pi_i n^{*-1}$ while $\{g_\nu \rho_\nu\}$ is a sequence of independent, identically distributed L' random variables. Hence

$$\pi_i n^{*-1} \sum_{\nu=1}^{n^*} g_\nu \rho_\nu \xrightarrow{\text{a.s.}} \pi_i \int_{w_\nu \leq K/2} \rho_\nu dP.$$

For sufficiently large K , this is $\geq 1 - \delta/2$, and the result holds. The case $j = 3$ is clear. But $|\hat{S}_\nu| \geq K$ and $w_{\nu+1} \leq K/2$ implies $\{\bar{S}_j: \tau_{\nu+1} \leq j < \tau_{\nu+2}\}$ are all positive or negative with \hat{S}_ν . Hence on $\Lambda_0 \cap \Lambda_1 \cap \Lambda_2 \cap \Lambda_3$, which has probability $\geq 1 - 4\epsilon$, $r(n)$ differs by at most 4δ from $r^*(n^*) = \sum_{\nu=1}^{n^*-1} t_\nu \rho_{\nu+1} / \sum_{\nu=1}^{n^*} \rho_\nu$. Divide numerator and denominator by n^* and apply Lemma 3 to see that the limiting distribution of $r(n)$ is that of $n^{*-1} \sum_{\nu=1}^{n^*} t_\nu$.

Moreover $r^*(n^*)$ and $r(n)$ are both in $[0, 1]$; hence their expectations differ by at most $4\epsilon + 4\delta$. And for sufficiently large n , $(1 - \delta)m_i^{-1} \leq n^*(\sum_{\nu=1}^{n^*} \rho_\nu)^{-1} \leq (1 + \delta)m_i^{-1}$ except on a set of measure $\leq \epsilon$, so that

$$\begin{aligned} (1 - \delta)m_i^{-1} n^{*-1} \sum_{\nu=1}^{n^*-1} E(t_\nu \rho_{\nu+1}) - \epsilon &\leq E r^*(n^*) \\ &\leq (1 + \delta)m_i^{-1} n^{*-1} \sum_{\nu=1}^{n^*-1} E(t_\nu \rho_{\nu+1}) + \epsilon. \end{aligned}$$

But t_ν and $\rho_{\nu+1}$ are independent, so that $E(t_\nu \rho_{\nu+1}) = P(S_\nu > 0)m_i$. Thus

$$\begin{aligned} (1 - \delta)n^{*-1} \sum_{\nu=1}^{n^*-1} P(\hat{S}_\nu > 0) - 5\epsilon - 4\delta &\leq n^{-1} \sum_{j=1}^n P(\bar{S}_j > 0) \\ &\leq (1 + \delta)n^{*-1} \sum_{\nu=1}^{n^*-1} P(\hat{S}_\nu > 0) + 5\epsilon + 4\delta \end{aligned}$$

implying that

$$n^{*-1} \sum_{\nu=1}^{n^*} P(\hat{S}_\nu > 0) \quad \text{and} \quad n^{-1} \sum_{j=1}^n P(\bar{S}_j > 0)$$

have the same asymptotic behavior.

Applying Theorem 7.1 of [4], which is the best available arcsine law for independent random variables, completes the proof.

4. Generalizations. Clearly Lemma 2 holds for the partial sums of any sequence of independent, identically distributed random variables; Lemma 3 holds when $\{\rho_\nu: \nu = 1, 2, \dots\}$ is any sequence of independent, identically distributed random variables, as does Lemma 4—provided $E(\rho_\nu)$ is finite and nonzero. The lemma fails without this restriction. In Lemma 4, moreover, the $\{t_\nu\}$ may take any finite set of values, without a serious change in the proof. In addition, weighted averages of arcsine laws may arise as the limiting distribution, if there is more than one positive class. Finally, transient states are permitted if the process spends only a finite amount of time among them; M is then random, and computed for the positive class in which the process is absorbed.

To state these results precisely, suppose the process has positive recurrent classes $I_m, m = 1, 2, \dots$. Define $\{\pi_i^{(m)}: i \in I_m\}$ as the corresponding stationary initial distributions, M_m the corresponding M 's. Let $\pi_m = P(x_n \in I_m \text{ ultimately})$, $\bar{M} = M_m$ if $x_n \in I_m$ ultimately, $\bar{r}(n)$ be the relative frequency of the event $S_n - \bar{M}n > 0$ among the first n partial sums, and $P_i(\cdot) = P(\cdot | x_0 = i)$. Since the limiting behavior of $P(\bar{S}_n > 0)$ is independent of the initial distribution when there is only one positive class we have the following general result.

THEOREM 2. *Let $\{x_n\}$ be a Markov chain with stationary transition probabilities and state space I . Suppose*

- (i) $\{x_n\}$ has no null states,
 - (ii) $\{x_n\}$ is among its transient states only finitely often,
 - (iii) $\sum_{i \in I_m} \pi_i^{(m)} |f(i)| < \infty$ for all m ,
 - (iv) $\lim_{n \rightarrow \infty} n^{-1} \sum_{j=1}^n P_i(S_j - M_m j > 0) = \alpha_m$ for $i \in I_m$ exists for all m .
- Then $\{x_n\}$ obeys the weighted arcsine law

$$\lim_{n \rightarrow \infty} P(\bar{r}(n) \leq x) = \sum_m \pi_m F_{\alpha_m}(x).$$

These considerations generalize easily to the continuous time case. A much more interesting question is the behavior of a chain with one null recurrent class; for then the crucial Lemma 4 fails.

REFERENCES

1. Kai Lai Chung, *Markov chains with stationary transition probabilities*, Springer, Berlin, 1960.
2. W. Doeblin, *Sur deux problèmes de M. Kolmogoroff concernant les chaînes dénombrables*, Bull. Soc. Math. France **66** (1938), 210–220.
3. J. L. Doob, *Stochastic processes*, Wiley, New York, 1953.
4. F. Spitzer, *A combinatorial lemma and its applications to probability theory*, Trans. Amer. Math. Soc. **82** (1956), 323–339.

UNIVERSITY OF CALIFORNIA, BERKELEY