PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 96, Number 2, February 1986

A FINITELY ADDITIVE GENERALIZATION OF BIRKHOFF'S ERGODIC THEOREM[1]

S. RAMAKRISHNAN

ABSTRACT. A finitely additive generalization of Birkhoff's ergodic theorem is obtained which yields, in particular, strong laws of large numbers in the i.i.d. setting as well as for positive recurrent Markov chains.

Suppose X is a nonempty set and $\mathscr{A}$ a σ-field of subsets of X. Let T be a measurable transformation of X and let P be a finitely additive probability defined on $(X, \mathscr{A})$ which is invariant under T, i.e., $P(T^{-1}(A)) = P(A)$ for all $A \in \mathscr{A}$. Let $\{f_n\}$ and $\{g_n\}$ be sequences of real valued measurable functions defined on $(X, \mathscr{A})$.

DEFINITION. Say that $\{f_n\}$ is unlikely to be strongly dominated by $\{g_n\}$ if for every $\varepsilon > 0$, $\operatorname{Lim}_{N\to\infty} P\{x\colon\ g_n(x) \leqslant f_n(x) + \varepsilon \text{ for some } n \leqslant N\} = 1$. We shall denote this by $\{f_n\} \not\prec \{g_n\}$.

The following lemma is inspired by ideas in [**6**, **7**, **8** and **16**].

LEMMA 1. *Let f and g be measurable functions which are integrable with respect to P (meaning, functions for which the integral is well defined, not necessarily finite). If $\{(1/n)\Sigma_{i=0}^{n-1} f \circ T^i\} \not\prec \{(1/n)\Sigma_{i=0}^{n-1} g \circ T^i\}$, then $\int_A g\,dP \leqslant \int_A f\,dP$ for all $A \in \mathscr{I}$, where $\mathscr{I}$ is the σ-field of all invariant sets A in $\mathscr{A}$, i.e., sets A satisfying $A = T^{-1}(A)$.*

(See [**5** or **9**] for the definition of the finitely additive integral.)

PROOF. Fix $\varepsilon > 0$ and a positive integer M. By the hypothesis there exists a positive integer N such that

$$P\left\{x\colon \frac{1}{n}\sum_{i=0}^{n-1} g(T^i x) \leqslant \frac{1}{n}\sum_{i=0}^{n-1} f(T^i x) + \varepsilon \text{ for some } n \leqslant N\right\} \geqslant 1 - \frac{\varepsilon}{M}.$$

Call the set on the left side D. Define

$$\bar{g} = \min(g, M) \quad \text{and} \quad \bar{f} = \begin{Bmatrix} \max(f, -M) \text{ on } D \\ \max(f, M) \text{ on } D^c \end{Bmatrix}.$$

Clearly now, for *every* $x \in X$, $(1/n)\Sigma_{i=0}^{n-1} \bar{g}(T^i x) \leqslant (1/n)\Sigma_{i=0}^{n-1} \bar{f}(T^i x) + \varepsilon$ for some $n \leqslant N$. (For $x \in D^c$, n can be chosen to be 1.) For $x \in X$, let $n(x)$ denote the least such positive integer n. Define $n_1(x) = n(x)$ and $n_{k+1}(x) = n_k(x) + n(T^{n_k(x)}(x))$, $k \geqslant 1$ and $x \in X$. Choose a positive integer L such that $(N-1)M/L < \varepsilon$. For

Received by the editors July 2, 1984 and, in revised form, February 6, 1985.

1980 *Mathematics Subject Classification*. Primary 28D05, 60G05, 60F15.

Key words and phrases. Ergodic theorem, finitely additive probability, i.i.d. measure, Markov chain.

[1]Partially supported by NSF Grant No. MCS-8202069

0002-9939/86 $1.00 + $.25 per page

$x \in X$, let $k(x)$ be the largest positive integer such that $n_{k(x)}(x) \leqslant L$. By our choice, we have, for $1 \leqslant j \leqslant k(x)$,

$$\frac{1}{n_j(x) - n_{j-1}(x)} \sum_{i=n_{j-1}(x)}^{n_j(x)-1} \bar{g}(T^i x) \leqslant \frac{1}{n_j(x) - n_{j-1}(x)} \sum_{i=n_{j-1}(x)}^{n_j(x)-1} \bar{f}(T^i x) + \varepsilon,$$

where $n_0(x) = 0$. Also,

$$\sum_{i=n_{k(x)}(x)}^{L-1} \bar{g}(T^i x) \leqslant \sum_{i=n_{k(x)}(x)}^{L-1} \bar{f}(T^i x) + 2M\big(L - n_{k(x)}(x)\big),$$

since, by definitions, $\bar{g} \leqslant M$ and $\bar{f} \geqslant -M$. From all these inequalities, we get

$$\sum_{i=0}^{L-1} \bar{g}(T^i x) \leqslant \sum_{i=0}^{L-1} \bar{f}(T^i x) + L\varepsilon + 2M(N-1),$$

since, by definition of $n_{k(x)}(x)$, $L - n_{k(x)}(x)$ is at most $(N-1)$.

Therefore, by choice of L,

$$\frac{1}{L} \sum_{i=0}^{L-1} \bar{g}(T^i x) \leqslant \frac{1}{L} \sum_{i=0}^{L-1} \bar{f}(T^i x) + 3\varepsilon.$$

Now integrating both sides w.r.t. P and using the invariance of P under T, we get $\int_A \bar{g}\, dP \leqslant \int_A \bar{f}\, dP + 3\varepsilon$, for every $A \in \mathscr{I}$. Therefore, $\int_A \min(g, M)\, dP \leqslant \int_A \max(f, -M)\, dP + 2MP(D^c) + 3\varepsilon$ (using the fact that on D^c, $\bar{f} \leqslant \max(f, -M) + 2M$), and hence $\int_A \min(g, M)\, dP \leqslant \int_A \max(f, -M)\, dP + 5\varepsilon$. Since ε and M are arbitrary, the result follows by taking limits as $\varepsilon \to 0$ and $M \to \infty$ and using the definition of the finitely additive integral. □

DEFINITION. Say that a sequence $\{f_n\}$ of measurable functions is P-regular if, for every positive integer M, $\{f_n\} \not\prec \{\min(f^*, M)\}$ and $\{\max(f_*, -M)\} \not\prec \{f_n\}$ where $\{\min(f^*, M)\}$ and $\{\max(f_*, -M)\}$ are constant sequences with $f^* = \limsup_{n\to\infty} f_n$ and $f_* = \liminf_{n\to\infty} f_n$.

DEFINITION. Say that a nonnegative measurable function f is (P, T)-regular if the sequence $\{f_n\}$ defined by $f_n = (1/n)\sum_{i=0}^{n-1} f \circ T^i$ is P-regular.

Say that an integrable function f is (P, T)-regular if $f^+ = \max(f, 0)$ and $f^- = -\min(f, 0)$ are (P, T)-regular.

THEOREM. *Let T be a measurable transformation on the measurable space $(X, \mathscr{A})$ and P be a finitely additive probability on $(X, \mathscr{A})$ invariant under T.*

(A) *For every integrable (P, T)-regular function f on $(X, \mathscr{A})$, $\int_A f^*\, dP = \int_A f_*\, dP = \int_A f\, dP$ for every $A \in \mathscr{I}$, where*

$$f^* = \limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f \circ T^i, \qquad f_* = \liminf_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f \circ T^i$$

and $\mathscr{I}$ is the sub σ-field of T-invariant sets.

(B) *Suppose further that $P(\bigcup_{n=1}^{\infty} A_n) = 0$ for every sequence $\{A_n\}_{n\geqslant 1}$ of sets in $\mathscr{I}$ such that $P(A_n) = 0$ for all $n \geqslant 1$. Then $P\{f^* \neq f_*\} = 0$.*

PROOF. First, it is enough to prove Theorem A for nonnegative (P, T)-regular functions, because then

$$\int_A (f^+)^* \, dP = \int_A (f^+)_* \, dP = \int_A f^+ \, dP$$

and

$$\int_A (f^-)^* \, dP = \int_A (f^-)_* \, dP = \int_A f^- \, dP \quad \text{for all } A \in \mathscr{I},$$

where $f^+ = \max(f, 0)$ and $f^- = -\min(f, 0)$. This, together with the integrability of f, implies that either $(f^+)^*$ or $(f^-)^*$ is finite except on a P-null set. Hence, $[(f^+)^* - (f^-)_*]$ and $[(f^+)_* - (f^-)^*]$ are well defined except on a P-null set. We thus have, for $A \in \mathscr{I}$,

$$\begin{aligned}\int_A f^* \, dP &= \int_A (f^+ - f^-)^* \, dP \\ &\leqslant \int_A \left[(f^+)^* - (f^-)_*\right] dP \\ &= \int_A (f^+ - f^-) \, dP = \int_A \left[(f^+)_* - (f^-)^*\right] dP \\ &\leqslant \int_A (f^+ - f^-)_* \, dP = \int_A f_* \, dP.\end{aligned}$$

Since $f_* \leqslant f^*$, all inequalities above are actually equalities.

To prove Theorem A for a nonnegative (P, T)-regular function f, note that, by definitions and by the lemma, $\int_A \min(f^*, M) \, dP \leqslant \int_A f dP \leqslant \int_A f_* \, dP$ for every $A \in \mathscr{I}$ and hence $\int_A f^* \, dP \leqslant \int_A f dP \leqslant \int_A f_* \, dP$ for every $A \in \mathscr{I}$. Since $f_* \leqslant f^*$, all inequalities above are actually equalities and part (A) is proved. This implies that $P\{f^* - f_* > 1/n\} = 0$ for every n. Since all sets involved are in $\mathscr{I}$, part (B) follows by hypothesis. □

COROLLARY (BIRKHOFF'S ERGODIC THEOREM). *Let $(X, \mathscr{A}, P)$ be a countably additive probability space and T a measurable transformation. Suppose P is invariant under T. Then for every L^1 function f, $(1/n)\sum_{i=0}^{n-1} f \circ T^i$ converges almost surely to the conditional expectation of f given $\mathscr{I}$.*

PROOF. First, countable additivity of P implies that every integrable function f is (P, T)-regular. The proof is now immediate from the previous theorem. □

REMARK 1. In the general (finitely additive) case we cannot hope to improve the Theorem to include every L^1 function f as shown by the following example.

EXAMPLE 1. Let X be the set of positive integers, $\mathscr{A}$ be the power set of N, T: $n \to n + 1$, and P a finitely additive probability on X invariant under T. Such a P is known to exist (for example, is induced by a Banach limit). Let $\{a_n\}_{n \geqslant 1}$ be a sequence of zeros and ones such that $(1/n)\sum_{k=1}^{n} a_k$ does not converge.

Let $f(n) = a_n$, $n \geqslant 1$. Then $(1/n)\sum_{i=0}^{n-1} f(T^i x) = (1/n)\sum_{i=x}^{x+n-1} a_i$, which does not converge for any x.

REMARK 2. In the general (finitely additive) case, if $\int_A f^* \, dP = \int_A f_* \, dP$ for all $A \in \mathscr{I}$, it does not necessarily follow that $f^* = f_*$ almost surely although it is equivalent to $P\{f^* - f_* > \varepsilon\} = 0$ for every $\varepsilon > 0$.

Let $X = I^N$ where I is a nonempty set and N the set of positive integers. Let $\mathscr{A}$ be the σ-field generated by the open subsets of X when X is equipped with the product of discrete topologies. Given, for each $i \in I$, a finitely additive probability γ_i defined on all subsets of I and a finitely additive probability μ (also defined on all subsets of I), there exists a finitely additive probability P_μ defined on $\mathscr{A}$ (following Dubins and Savage [**3**], Dubins [**4**] and Purves and Sudderth [**10**]) which provides a reasonable framework to study a finitely additive Markov chain with stationary transition probabilities γ_i, $i \in I$, and initial distribution μ. (See [**12**, **13**, **14** and **15**].) We shall henceforth refer to P_μ defined above as the Markov measure on X with stationary transitions γ_i, $i \in I$. When $\gamma_i = \mu$ for all $i \in I$, we shall call P_μ the i.i.d. product measure with marginal μ. T henceforth will stand for the shift transformation on $X = I^N$.

LEMMA 2. *If P_μ is the i.i.d. product measure on $(X, \mathscr{A})$ with marginal μ, then every L^1 function f, defined on X, which depends only on the first coordinate, is (P_μ, T)-regular.*

PROOF. Let f_n be the function on X defined on X by $f_n(i_1, i_2, \ldots, i_n, \ldots) = f(i_n, i_{n+1}, \ldots)$ for all $n \geqslant 1$ and $(i_1, i_2, \ldots, i_n, \ldots)$ in X. Given $\varepsilon > 0$ and $\delta > 0$, we can use the standard technique of approximating by simple functions to obtain a sequence $\{g_n\}$ of finite-valued functions such that g_n depends only on the nth coordinate, and $P_\mu\{|f_n - g_n| < \varepsilon/2 \text{ for all } n\} \geqslant 1 - \delta/2$. This technique can and has been used in the finitely additive setting in [**1**, **10** and **14**] in case of independent product probability. Let $\mathscr{A}_n$ be the (finite) σ-field on I induced by g_n, $n \geqslant 1$. It is known by Theorem 2.1 in [**1**] that the i.i.d. measure P when restricted to $\prod_{n=1}^\infty \mathscr{A}_n$, the product σ-field, is countably additive.

Thus the sequence $\{g_n\}$ is (P_μ, T)-regular and by choice of our approximation it is easy to see that, for every positive integer M, $P\{(1/n)\sum_{k=1}^n f_k > \min(f^*, M) - \varepsilon$ for some $n \leqslant N\} \geqslant 1 - \delta$ for sufficiently large N, and $P\{(1/n)\sum_{k=1}^n f_k < \max(f_*, -M) + \varepsilon$ for some $n \leqslant N\} \geqslant 1 - \delta$ for sufficiently large N, where

$$f^* = \limsup_{n \to \infty} \frac{1}{n} \sum_{k=1}^n f_k \quad \text{and} \quad f_* = \liminf_{n \to \infty} \frac{1}{n} \sum_{k=1}^n f_k.$$

This proves that f is (P_μ, T)-regular.

The i.i.d. strong law of large numbers. (See [**2** and **9**].) Let $(X, \mathscr{A})$ be as defined above and P be the i.i.d. product measure on $(X, \mathscr{A})$ with marginal μ, where μ is a finitely additive probability defined on all subsets of I. Let f be an L^1-function on X depending only on the first coordinate. Then

$$P_\mu\left\{(i_1, i_2, \ldots, i_n, \ldots) \colon \frac{1}{n} \sum_{k=1}^n f(i_k, i_{k+1}, \ldots) \to \int f \, dP_\mu\right\} = 1.$$

PROOF. It follows from a result of Purves and Sudderth that P_μ is countably additive and 0-1 valued on the shift invariant σ-field in the i.i.d. case. (See Theorems 1 and 2 in §3 in [**11**] or see [**9**].) This together with Lemma 2 immediately gives the above Theorem as a corollary of Theorem B. $\square$

Suppose P_μ is the Markov measure on X with stationary transitions γ_i, $i \in I$, and initial distribution μ. If $\mu = \delta_i$, the point mass at i, we shall denote P_μ by P_i. For $j \in I$, let $t_{j,1}$ be the function on X defined by $t_{j,1}(x) = n$ if $i_n = j$ and $i_m \neq j$ for $1 \leqslant m < n$, $= \infty$ otherwise, where $x = (i_1, i_2, \ldots, i_n, \ldots)$. An element $i \in I$ is called positive recurrent if $m_{ii} = \int t_{i,1}\, dP_i < \infty$.

The set I is called a positive recurrent Markov chain under P_μ if (a) every $i \in I$ is positive recurrent, and (b) $P_i(t_{j,1} < \infty) > 0$ for all i, $j \in I$. For $j \in I$ and a function f on I, define

$$\mu_j(f) = \int \left[\sum_{k=1}^{\infty} f(i_k) 1_{\{t_{j,1} \geqslant k\}} \right] dP_j(x),$$

$x = (i_1, i_2, \ldots, i_k, \ldots)$, provided the integral exists. The integrand is the sum of f values up to the first occurrence of j. The function f on I can be used to define a function on X which depends only on the first coordinate. This function on X will also be denoted by f.

It has been proved in [**12**, §10, Theorem 1] that if I is a positive recurrent class under P_μ, there is a finitely additive probability λ, defined by $\lambda(E) = \mu_i(1_E)/m_{ii}$, $E \subseteq I$, $i \in I$, which is a canonical stationary initial distribution for the Markov chain (therefore, P_λ is invariant under the shift transformation T).

LEMMA 3. *Let I be a positive recurrent Markov chain. Let f be a function on I such that $\mu_i(|f|) < \infty$ for some $i \in I$. Then f, regarded as a function on X, is (P_λ, T)-regular, where λ is the canonical stationary initial distribution.*

PROOF. For $j \in I$, let $t_{n,j}$ (defined on X) be the time of nth occurrence of j, for $n \geqslant 1$. For the i in the statement of the lemma, let F be the set of all nonempty sequences of elements of I whose last coordinate is i and none of the other coordinates is i. An element of F will be called an i-block. Let G_i be the elements of X for which infinitely many coordinates are i. It is known (Theorem 9, §4 of [**12**]) that $P_j(G_i) = 1$ for all $j \in I$. Let $\{\beta_n\}$ be the sequence of functions defined on G_i into F by $\beta_1(x) = (w_1, \ldots, w_{t_{i,1}(x)})$ and $\beta_{n+1}(x) = (w_{t_{i,n}(x)+1}, \ldots, w_{t_{i,n+1}(x)})$, $n \geqslant 1$, where $x = (w_1, \ldots, w_n, \ldots)$.

The Blocks Theorem (Theorem 2, §5 of [**12**]) proves that, under the map Ψ: $x \to (\beta_1(x), \ldots, \beta_n(x), \ldots)$, the probability P_i is carried over to F^N as an i.i.d. product measure, i.e., $P_i \circ \Psi^{-1}$ is an i.i.d. measure. Let g be the function on F which is the sum of f values in an i-block. If g is regarded as a function on F^N depending only on the first coordinate, then the hypothesis $\mu_i(|f|) < \infty$ implies that g is an L^1-function on F^N. Let Λ be the block-length function defined on F. Once again the standard technique of approximating by finite valued functions together with Theorem 2.1 of [**1**] implies that the sequence $\{\sum_{k=1}^n g(\alpha_i)/\sum_{i=1}^n \Lambda(\alpha_i)\}$, $(\alpha_1, \ldots, \alpha_n, \ldots)$ in F^N, is $P_i \circ \Psi^{-1}$-regular. So by the Blocks Theorem the sequence $\{\sum_{k=1}^{t_{i,n}(x)} f(w_k)/t_{i,n}(x)\}$, $x = (w_1, \ldots, w_n, \ldots)$ in X, is (P_i, T)-regular. This suffices to prove the P_i-regularity of f (this can be seen, for example, by Lemmas 6 and 8 of [**14**]).

Let λ_n be defined on subsets of I by

$$\lambda_n(E) = \frac{1}{m_{ii}} \int 1_E(x_n) 1_{\{t_{i,1} \geqslant n\}}\, dP_i(x),$$

where $x = (x_1, \ldots, x_n, \ldots)$, $n \geqslant 1$.

Then $\lambda(E) = \sum_{n=1}^{\infty} \lambda_n(E)$. By the Strong Markov property (Theorem 4, §3 of [**12**]),

$$\begin{aligned} P_\lambda(t_{i,1} \geqslant K+1) &= \int_{\{j \neq i\}} P_j(t_{i,1} \geqslant K)\, d\lambda(j) \\ &\leqslant \int P_j(t_{i,1} \geqslant K)\, d\lambda(j) \\ &= \sum_{n=1}^{\infty} \int P_j(t_{i,1} \geqslant K)\, d\lambda_n(j) \quad \text{by Lemma 1, §8 of [12]} \\ &= \frac{1}{m_{ii}} \sum_{n=1}^{\infty} P_i(t_{i,1} \geqslant n+K) \quad \text{by the Strong Markov property.} \end{aligned}$$

Since the positive recurrence of i implies that $m_{ii} = \sum_{n=1}^{\infty} P_i(t_{i,1} \geqslant n)$ is finite, it follows from the above inequalities that $P_\lambda(t_{i,1} \geqslant K+1) \to 0$ as $K \to \infty$. Using the Strong Markov property again, and the (P_i, T)-regularity of f, the result follows.

Strong law of large numbers for a positive recurrent Markov chain. Let I be a positive recurrent Markov chain. Let f be a function on I such that $\mu_i(|f|) < \infty$ for some i. If λ is the canonical stationary initial distribution, then

$$P_\lambda\left\{ x = (w_1, \ldots, w_n, \ldots) \,\Big|\, \frac{1}{n} \sum_{k=1}^{n} f(w_k) \to \int f\, d\lambda \right\} = 1.$$

PROOF. First, by Theorem 1, §11 of [**12**], P_λ is 0-1 valued on the shift invariant σ-field $\mathscr{I}$. Also, by the Strong Markov property $P_\lambda(A) = P_\lambda(t_{i,1} < \infty) P_i(A) = P_i(A)$ for all $A \in \mathscr{I}$, the shift invariant σ-field. By the remark in the last paragraph of [**15**], P_i is countably additive on $\mathscr{I}$ and hence P_λ is countably additive on $\mathscr{I}$. Now an application of Lemma 3 and Theorem B completes the proof.

REFERENCES

1. R. Chen, *A finitely additive version of Kolmogorov's law of iterated logarithm*, Israel J. Math. **23** (1976), 209–220.

2. ______, *Some finitely additive versions of the strong law of large numbers*, Israel J. Math. **24** (1976), 244–259.

3. L. E. Dubins and L. J. Savage, *How to gamble if you must: Inequalities for stochastic processes*, McGraw-Hill, New York, 1965.

4. L. E. Dubins, *On Lebesgue-like extensions of finitely additive measures*, Ann. Probab. **2** (1974), 456–463.

5. N. Dunford and J. T. Schwartz, *Linear operators*, Part I, Interscience, New York, 1958.

6. T. Kamae, *A simple proof of the ergodic theorem using nonstandard analysis*, Israel J. Math. **42** (1982), 284–290.

7. Y. Katznelson and B. Weiss, *A simple proof of some ergodic theorems*, Israel J. Math. **42** (1982), 291–296.

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

9. R. A. Purves and W. D. Sudderth, *Some finitely additive probability*, Univ. of Minnesota School of Statistics Tech. Report No. 220, 1973.

10. ______, *Some finitely additive probability*, Ann. Probab. **4** (1976), 259–276.

11. ______, *Finitely additive zero-one laws*, Sankhya **45A** (1983), 32–37.

12. S. Ramakrishnan, *Finitely additive Markov chains*, Trans. Amer. Math. Soc. **265** (1981), 247–272.

13. ______, *Potential theory for finitely additive Markov chains*, Stochastic Process. Appl. **16** (1984), 287–303.

14. ______, *Central limits theorems in a finitely additive setting*, Illinois J. Math. **28** (1984), 139–161.

15. ______, *The tail σ-field of a finitely additive Markov chain starting from a recurrent state*, Proc. Amer. Math. Soc. **89** (1983), 493–497.

16. P. C. Shields, *A simple direct proof of Birkhoff's ergodic theorem*, 1982 (unpublished).

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MIAMI, P. O. BOX 249085, CORAL GABLES, FLORIDA 33124