ON THE APPLICATION OF THE INDIVIDUAL ERGODIC THEOREM TO DISCRETE STOCHASTIC PROCESSES

BY

H. D. BRUNK

1. **Introduction.** That Birkhoff's individual ergodic theorem provides a strong law of large numbers for stationary stochastic processes has been known since its applications to this end by Doob (cf. [3; 4]) and Hopf [9]. The conclusion of the ergodic theorem as it applies to a sequence of random variables, $\{x_n\}$, is from one point of view considerably stronger than that of the strong law of large numbers. The latter states (in part) that the sequence of arithmetic means $\{(1/k)\sum_{s=1}^{k} x_s\}$ converges with probability 1, while the individual ergodic theorem will be used in §§2 and 3 in giving necessary and sufficient conditions for the convergence with probability 1 of the sequence $\{(1/k)\sum_{s=0}^{k-1} z_s\}$ where $\{z_s\}$ is any of a wide class of sequences of random variables of the form $z_s=\phi(x_{1+s}, x_{2+s}, \cdots)$ and where $\phi(x_1, x_2, \cdots)$ is a measurable function on the infinite product space associated with the sequence $\{x_n\}$ (in particular, if $\phi(x_1, x_2, \cdots)\equiv x$, then $z_s=x_{s+1}$). Studies by Dunford and Miller [7] and by Dowker [6] involving averaged measures suggested the approach used.

2. **The individual ergodic theorem for totally finite measures.** Let Ω be a space of points ω. Let $\mathcal{J}$ be a σ-algebra of ω sets, that is, a Borel field containing Ω ($\mathcal{J}$ contains Ω, differences, countable unions, and countable intersections of sets of $\mathcal{J}$; [4, p. 599]). Let ν be a measure on $\mathcal{J}$ such that $\nu(\Omega)=1$. Let T be a point transformation from Ω into or onto Ω. For a set $F\in\mathcal{J}$, let $T^{-1}F$ denote the set of points ω for which $T\omega\in F$. Let T be measurable with respect to $\mathcal{J}$: for each $F\in\mathcal{J}$, we have $T^{-1}F\in\mathcal{J}$. Define the set function νT^{-s} for each nonnegative integer s so that $\nu T^{-s}(F)=\nu(T^{-s}F)$ for $F\in\mathcal{J}$. It is readily verified that νT^{-s} is a measure on F such that $\nu T^{-s}(\Omega)=1$. For each positive integer k, and for $a\in\mathcal{J}$, define

$$\mu_k(A) = (1/k)\sum_{s=0}^{k-1} \nu T^{-s}(A).$$

The set function μ_k is also a measure on $\mathcal{J}$ such that $\mu_k(\Omega)=1$.

Let $c_A(\omega)$ denote the characteristic function of the set A.

THEOREM 2.1. *A necessary and sufficient condition in order that*

$$\left\{(1/k)\sum_{s=0}^{k-1} c_A(T^s\omega)\right\} \qquad \textit{converge a.e. } (\nu) \tag{2.1}$$

Received by the editors March 29, 1954.

(*almost everywhere with respect to the measure* ν) *for each set* $A \in \mathcal{J}$ *is that*

$$\lim_{k\to\infty} \mu_k(A) \qquad \text{exists for each } A \in \mathcal{J}. \tag{2.2}$$

If (2.2) *holds then also*

$$\left\{(1/k) \sum_{s=0}^{k-1} f(T^s\omega)\right\} \qquad \text{converges a.e. } (\nu) \tag{2.3}$$

for each function f satisfying:

$$f \text{ is integrable } (\nu T^{-s}) \qquad (s = 0, 1, 2, \cdots), \tag{2.4}$$

and

$$\liminf_{k\to\infty} (1/k) \sum_{s=0}^{k-1} \int |f| \, d\nu T^{-s} < \infty. \tag{2.5}$$

This theorem is reminiscent of Dowker's result [6] to the effect that a necessary and sufficient condition that the sequence of averages $\{(1/k) \cdot \sum_{s=0}^{k-1} f(T^s\omega)\}$ converge a.e. for every integrable function is that the measure ν be potentially invariant. In that work, however, T denotes a ***biunique*** (1-1) transformation such that if $A \in \mathcal{J}$, then both $T^{-1}A \in \mathcal{J}$ and $TA \in \mathcal{J}$; these properties of T are essential to the argument that if $\lim_{k\to\infty} \mu_k(A) = 0$, then $\nu(A) = \nu(T^{-1}A) = \cdots = \nu T^{-s}(A) = \cdots = 0$. In Dunford's and Miller's paper [7], the transformation T^{-1} is required to preserve zero measure, a restriction which appears undesirable in making applications to stochastic processes.

Proof of sufficiency. Birkhoff's individual ergodic theorem as generalized by F. Riesz [14] states that if there is a σ-finite measure μ on $\mathcal{J}$ with respect to which T is measure preserving and f is integrable, then

$$\left\{\frac{1}{k} \sum_{s=0}^{k-1} f(T^s\omega)\right\}$$

converges a.e. (μ). Let $\mu(A) = \lim_{k\to\infty} \mu_k(A)$; this limit exists by (2.2). That μ is a measure on $\mathcal{J}$ follows from a known theorem to the effect that if a sequence of finite measures converges at each set of a Borel field, the limit function is a measure ([8, p. 170, Ex. 14]; [12, p. 107, Ex. 13]). One verifies immediately that $\mu(T^{-1}F) = \mu(F)$ for $F \in \mathcal{J}$. In order to complete the proof of (2.3) under hypotheses (2.4) and (2.5), it remains to show that f is integrable (μ) and that the set of points ω on which $\{(1/k) \sum_{s=0}^{k-1} f(T^s\omega)\}$ fails to converge has ν-measure 0 as well as μ-measure 0.

Since the domain of definition of each of the measures ν and μ is $\mathcal{J}$, and since, by hypothesis (2.4), f is an extended real-valued ν-measurable function, f is also μ-measurable. Suppose first that f is the characteristic function of a set $A \in \mathcal{J}$. Then

$$\int f d\mu = \mu(A) = \lim_{k\to\infty} \mu_k(A) = \lim_{k\to\infty} \int f d\mu_k.$$

It follows that for any simple function f (a linear combination of characteristic functions of sets in $\mathcal{J}$),

$$\int f d\mu = \lim_{k\to\infty} \int f d\mu_k$$

(indeed, this holds for f bounded, measurable; cf. [12, Exercises 23.d, 24.c]). Suppose now that f is an arbitrary function satisfying (2.4) and (2.5). The function $|f|$ is the limit of a nondecreasing sequence $\{g_i\}$ of non-negative simple functions, so that

$$\int |f| \, d\mu = \lim_{i\to\infty} \int g_i d\mu = \lim_{i\to\infty} \lim_{k\to\infty} \int g_i d\mu_k \leqq \liminf_{k\to\infty} \int |f| \, d\mu_k$$

$$= \liminf_{k\to\infty} (1/k) \sum_{s=0}^{k-1} \int |f| \, d\nu T^{-s} < \infty$$

(note that if f is bounded, $\lim_{k\to\infty} (1/k) \sum_{s=0}^{k-1} \int |f| \, d\nu T^{-s}$ necessarily exists and is finite). Thus a function satisfying hypotheses (2.4) and (2.5) is integrable (μ). It follows from the individual ergodic theorem as generalized by F. Riesz that the sequence $\{(1/k) \sum_{s=0}^{k-1} f(T^s\omega)\}$ converges a.e. (μ). It remains to show that the exceptional set has ν-measure 0 as well as μ-measure 0.

Let F denote the set of points ω at which $f(T^j\omega)$ is finite for all j $(j=0, 1, 2, \cdots)$:

$$F = \bigcap_{j=0}^{\infty} \{\omega: |f(T^j\omega)| < \infty\}.$$

Let E denote the set of points ω on which the sequence $\{(1/k) \sum_{s=0}^{k-1} f(T^s\omega)\}$ converges. Then

$$E = E \cap F = T^{-s}E \cap F.$$

For a set $A \in \mathcal{J}$, let $A' = \Omega - A$. We have $E' \cup F' = (T^{-s}E)' \cup F' = T^{-s}E' \cup F'$ $(s=0, 1, 2, \cdots)$ and, by hypothesis (2.4), $\nu(F')=0$. Therefore $\nu T^{-s}(E')$ $=\nu(T^{-s}E') = \nu(T^{-s}E' \cup F') = \nu(E' \cup F') = \nu(E')$ $(s=0, 1, 2, \cdots)$. It follows that $\mu_k(E') = \nu(E')$ for each k, and hence that $\mu(E') = \nu(E')$. But $\mu(E')=0$, so that also $\nu(E')=0$. This completes the proof of (2.3) under hypotheses (2.4) and (2.5). Since for each $A \in \mathcal{J}$, $c_A(\omega)$ is a function f satisfying (2.4) and (2.5), (2.1) follows immediately.

REMARK. It is clear from the proof that the sufficiency part of Theorem 2.1 is valid under the less restrictive hypotheses that ν is a σ-finite measure, that

$\nu T^{-1}(A) < \infty$ if and only if $\nu(A) < \infty$, and that $\{\mu_k(A)\}$ converges at each set A of finite ν-measure, provided the hypothesis is added that $\mu = \lim_{k\to\infty} \mu_k$ is a measure.

Proof of necessity. Let A be an arbitrary set of $\mathcal{J}$ and set $c_A^*(\omega) = \lim_{k\to\infty} (1/k) \sum_{s=0}^{k-1} c_A(T^s\omega)$. This function is bounded and measurable, hence integrable (ν). Define a set function

$$M(A) = \int c_A^*(\omega) d\nu.$$

By the bounded convergence theorem we have

$$\begin{aligned} M(A) &= \lim_{k\to\infty} (1/k) \sum_{s=0}^{k-1} \int c_A(T^s\omega) d\nu \\ &= \lim_{k\to\infty} (1/k) \sum_{s=0}^{k-1} \int c_A(\omega) d\nu T^{-s} \\ &= \lim_{k\to\infty} (1/k) \sum_{s=0}^{k-1} \nu T^{-s}(A). \end{aligned}$$

Thus (2.2) is satisfied. This completes the proof of Theorem 2.1.

Let $\mathcal{J}_0$ be a field containing Ω whose minimal Borel extension is $\mathcal{J}$.

THEOREM 2.2. *If $\{\mu_k(A)\}$ converges uniformly for $A \in \mathcal{J}_0$, then the limit function $\mu(A)$ is a meaure on $\mathcal{J}_0$, and $\{\mu_k(A)\}$ converges uniformly for $A \in \mathcal{J}$ to the unique extension of μ to $\mathcal{J}$ at A.*

It is easy to show by example, however, that the convergence of $\{\mu_k(A)\}$ for $A \in \mathcal{J}$ does not imply the uniformity of the convergence with respect to $A \in \mathcal{J}$ and hence, in view of Theorem 2.2, not even the uniformity with respect to $A \in \mathcal{J}_0$.

Proof. Let $\{A_i\}$ be a contracting sequence of sets in $\mathcal{J}_0$ whose limit is the void set. If $\epsilon > 0$, one can choose a positive integer k such that $|\mu(A) - \mu_k(A)| < \epsilon/2$ for $A \in \mathcal{J}_0$, in view of the uniformity of the convergence of the sequence $\{\mu_k(A)\}$. If then a positive integer i_0 is chosen so that $\mu_k(A_i) < \epsilon/2$ for $i > i_0$ (k fixed), we have $\mu(A_i) < \epsilon$ for $i > i_0$. Thus μ is continuous from above at the void set and is therefore a measure on $\mathcal{J}_0$. It has then a unique extension to $\mathcal{J}$, which we shall also denote by μ. Let $E \in \mathcal{J}$, and let ϵ be a positive number. Choose a positive integer K such that $|\mu_k(A) - \mu(A)| < \epsilon/3$ for $k \geqq K$, $A \in \mathcal{J}_0$. For *fixed* $k \geqq K$, choose a set $B = B(\epsilon, k) \in \mathcal{J}_0$ such that $|\mu(B) - \mu(E)| < \epsilon/3$ and $|\mu_k(B) - \mu_k(E)| < \epsilon/3$. Then $|\mu(E) - \mu_k(E)| \leqq |\mu(E) - \mu(B)| + |\mu(B) - \mu_k(B)| + |\mu_k(B) - \mu_k(E)| < \epsilon$.

The following simple example shows that convergence on $\mathcal{J}_0$ of the sequence $\{\mu_k\}$ of measures is not sufficient for (2.2). Let Ω be the real line, and $\mathcal{J}_0$ the field of finite unions of intervals $[a, b)$, (a, b extended real numbers,

$[a, b)$ being replaced by $(-\infty, b)$ when $a=-\infty$), so that $\mathcal{J}$ is the class of Borel sets. Let $\nu(A)=1$ if the point 1 belongs to A, otherwise $\nu(A)=0$ for $A\in\mathcal{J}$. Let the point transformation T take $1/n$ into $1/(n+1)$ for each positive integer n and be continuous (say, linear on $[1/(n+1), 1/n)$ $(n=1, 2, \cdots)$, linear on $[k, k+1)$, with $k+1\rightarrow k$ $(k=1, 2, \cdots)$, and the identity on $(-\infty, 0]$). The transformation T is measurable. If $[a, b)$ contains the origin, then for s sufficiently large, $T^{-s}[a, b)\supset[0, 1]$, so that $\mu([a, b))=\lim_{k\to\infty}(1/k)\cdot\sum_{s=0}^{k-1}\nu T^{-s}([a, b))=1$; otherwise $\mu([a, b))=0$. Hence $\mu(A)=\lim_{k\to\infty}\mu_k(A)$ exists for $A\in\mathcal{J}_0$ and is the measure assigning the entire mass 1 to the origin. Let B be the set of points $\{1/k\}$, $k=1, 4, 5, 6, 7, \cdots, 2^{2n}, 2^{2n}+1, \cdots, 2^{2n+1}-1, \cdots$ $(n=0, 1, 2, \cdots)$. Then $\nu T^{-s}(B)=0$ if $s=2^{2n+1}-1, 2^{2n+1}, \cdots, 2^{2n+2}-2$ $(n=0, 1, 2, \cdots)$, while $\nu T^{-s}(B)=1$ if $s=2^{2n}-1, 2^{2n}, \cdots, 2^{2n+1}-2$ $(n=0, 1, 2, \cdots)$. The mean value $\lim_{k\to\infty}(1/k)\sum_{s=0}^{k-1}\nu T^{-s}(B)$ does not exist, yet $B\in\mathcal{J}$. Thus (2.2) fails.

3. **Discrete stochastic processes.** Let $\mathbf{x}_1, \mathbf{x}_2, \cdots$, be a sequence of random variables, or more precisely, the representation of a discrete stochastic process based on the infinite-dimensional Cartesian space $\Omega=\mathsf{X}_{i=1}^{\infty}R_1$ (R_1 is the real line) of points ω: $(x_1, x_2, \cdots)$ [4, p. 12, ff.]. For a fixed positive integer n, let $\mathcal{C}_n$ denote the class of cylinder sets of the form $\{\omega: a_1<x_1\leqq b_1, a_2<x_2\leqq b_2, \cdots, a_n<x_n\leqq b_n\}$, for extended real numbers $a_1, a_2, \cdots, a_n, b_1, b_2, \cdots, b_n$ (the symbol $\leqq$ preceding b_i is to be replaced by $<$ if $b_i=\infty$). A set in $\mathcal{C}_n$ will be called a cylinder set based on an n-dimensional interval (all intervals referred to will be such half-open intervals). Let $\mathcal{J}_{0n}$ denote the field of finite unions of sets in $\mathcal{C}_n$ $(n=1, 2, \cdots)$; $\mathcal{J}_{0n}$ is an increasing sequence of fields. Let $\mathcal{J}_n$ denote the minimal Borel extension of $\mathcal{J}_{0n}$: the class of cylinder sets based on Borel sets in $\mathsf{X}_{i=1}^{n}R_1$. Let $\mathcal{J}_0=\bigcup_{n=1}^{\infty}\mathcal{J}_{0n}$, and let $\mathcal{J}$ denote the minimal Borel extension of $\mathcal{J}_0$.

Let T denote the shift transformation ([4, p. 455]; [9, p. 57]): if $\omega=(x_1, x_2, \cdots)$ then $T\omega=(x_2, x_3, \cdots)$. The transformation T is measurable with respect to $\mathcal{J}$. Let $T^{-s}\mathcal{J}_{0n}(T^{-s}\mathcal{J}_n, T^{-s}\mathcal{J}_0, T^{-s}\mathcal{J})$ denote the class of sets of the form $T^{-s}A$ for $A\in\mathcal{J}_{0n}(\mathcal{J}_n, \mathcal{J}_0, \mathcal{J})$ $(s=0, 1, 2, \cdots; n=1, 2, \cdots)$. We have $\mathcal{J}_{0n}\supset T^{-1}\mathcal{J}_{0,n-1}\supset T^{-2}\mathcal{J}_{0,n-2}\supset\cdots\supset T^{-(n-1)}\mathcal{J}_{01}$, $\mathcal{J}_n\supset T^{-1}\mathcal{J}_{n-1}\supset\cdots\supset T^{-(n-1)}\mathcal{J}_1$ $(n=1, 2, \cdots)$, $\mathcal{J}_0\supset T^{-1}\mathcal{J}_0\supset T^{-2}\mathcal{J}_0\supset\cdots$, and $\mathcal{J}\supset T^{-1}\mathcal{J}\supset\cdots$. Further, $T^{-s}\mathcal{J}$ is the minimal Borel extension of $T^{-s}\mathcal{J}_0$ $(s=0, 1, 2, \cdots)$.

For $A\in\mathcal{J}$, let $\nu(A)$ denote the probability that the combined random variable $(\mathbf{x}_1, \mathbf{x}_2, \cdots)$ will belong to A. The set function ν is a totally finite measure on $\mathcal{J}$, with $\nu(\Omega)=1$. As in §2, we shall denote by $\mu_k(A)$ the arithmetic mean $(1/k)\sum_{s=0}^{k-1}\nu T^{-s}(A)$, and by $\mu(A)$ the limit of the sequence $\{\mu_k(A)\}$ when this limit exists.

LEMMA 3.1. *If the sequence $\{\mu_k(A)\}$ converges for $A\in\mathcal{J}_0$ to a measure $\mu(A)$, then $\mu(T^{-1}F)=\mu(F)$ for $F\in\mathcal{J}$, where for $F\in\mathcal{J}$, $\mu(F)$ denotes the unique extension of μ to a measure on $\mathcal{J}$.*

Proof. We note first that under the hypothesis, $\mu(T^{-1}A)=\mu(A)$ for $A\in\mathcal{J}_0$. For $F\in\mathcal{J}$, we have $\mu(F)=\inf\sum\mu(A_i)$ for coverings of F by countable unions $\bigcup A_i$ of sets $A_i\in\mathcal{J}_0$, and since $T^{-1}F\in T^{-1}\mathcal{J}$, we have also $\mu(T^{-1}F)=\inf\sum\mu(B_i)$ for coverings of $T^{-1}F$ by countable unions $\bigcup B_i$ of sets $B_i\in T^{-1}\mathcal{J}_0$. Given a covering of F by a countable union $\bigcup A_i$ of sets of $\mathcal{J}_0$, the union $\bigcup T^{-1}A_i$ is a covering of $T^{-1}F$ by sets of $T^{-1}\mathcal{J}_0$, and further $\mu(A_i)=\mu(T^{-1}A_i)$ since $A_i\in\mathcal{J}_0$ $(i=1, 2, \cdots)$. Hence $\mu(T^{-1}F)\leqq\mu(F)$. On the other hand, given a covering of $T^{-1}F$ by a countable union $\bigcup B_i$ of sets of $T^{-1}\mathcal{J}_0$, we have $B_i=T^{-1}A_i$ for some set $A_i\in\mathcal{J}_0$ $(i=1, 2, \cdots)$ so that $\mu(B_i)=\mu(A_i)$ and $\bigcup A_i$ is a covering of F. Hence $\mu(F)\leqq\mu(T^{-1}F)$. The two inequalities complete the proof of Lemma 3.1.

Let $\mathcal{S}$ denote the class of sets each of which belongs to $T^{-s}\mathcal{J}$ for each s, $s=0, 1, 2, \cdots$. Let $\mathcal{S}^*$ denote the class of sets belonging to $\mathcal{J}$ which are invariant under T: $S^*\in\mathcal{S}^*$ if and only if $T^{-1}S^*=S^*$. It is clear that $\mathcal{S}^*\subset\mathcal{S}$.

If $f(\omega)$ is a ν-measurable function of ω, then the function f and the shift transformation T generate a stochastic process, $z_s=f(T^s\omega)$. In particular, the process $\{x_s\}$ itself is generated by the shift transformation and the function $f(\omega)=x_1$, where $\omega=(x_1, x_2, \cdots)$.

THEOREM 3.2. *A necessary and sufficient condition that*

$$\left\{(1/k)\sum_{s=0}^{k-1} z_s\right\} \tag{3.1}$$

converge with probability 1 *for every process* $z_s=f(T^s\omega)$ *generated by the shift transformation* T *and a function* $f(\omega)$ *such that*

$$\mathcal{E}(z_s) \text{ exists} \qquad (s = 0, 1, 2, \cdots) \tag{3.2}$$

and

$$\liminf_{k\to\infty}\ (1/k)\sum_{s=0}^{k-1} E(|z_s|) < \infty \tag{3.3}$$

is that both

(3.4) *for every positive integer* n, *the sequence* $\{\mu_k(A)\}$ *converges for* $A\in\mathcal{J}_{0n}$ *to a set function* $\mu(A)$ *which is a measure on* $\mathcal{J}_{0n}$, *and*

(3.5) $\nu<\mu$ *on* $\mathcal{S}^*$ (ν *is absolutely continuous with respect to* μ *as a measure on* $\mathcal{S}^*$). *In this case indeed*

(3.6) $\nu=\mu$ *on* $\mathcal{S}^*$.

Proof of necessity. Hypothesis (3.1) implies (2.1) of Theorem 2.1, on taking $f(\omega)=c_A(\omega)$. Hence (2.2) of Theorem 2.1 is satisfied, and we have $\lim_{k\to\infty}\mu_k(A)=\mu(A)$ for each $A\in\mathcal{J}$. It follows that if $S^*\in\mathcal{S}^*$ then $\nu(S^*)=\mu(S^*)$.

Proof of sufficiency. By hypothesis μ is a measure on $\mathcal{J}_{0n}$ for every positive integer n; it can therefore be extended uniquely to a measure on $\mathcal{J}$ ([11]; [4, pp. 609, 639]). Let now $c_F(\omega)$ be the characteristic function of a set $F \in \mathcal{J}$. It is then integrable with respect to μ. Further, by Lemma 3.1, the transformation T^{-1} preserves μ-measure (note that (3.4) implies $\mu(A) = \lim_k \mu_k(A)$ for $A \in \mathcal{J}_0$, since $\mathcal{J}_0 = \bigcup_{n=1}^{\infty} \mathcal{J}_{0n}$). By the individual ergodic theorem, it follows that the sequence $\{(1/k) \sum_{s=0}^{k-1} c_F(T^s\omega)\}$ converges a.e. (μ). Now the set E': $\{\omega: (1/k) \sum_{s=0}^{k-1} c_F(T^s\omega) \text{ diverges}\}$ belongs to $\mathcal{S}^*$, so that by (3.5), since $\mu(E')=0$, also $\nu(E')=0$, and (2.1) is satisfied. Conclusion (2.2) and (2.3) of Theorem 2.1 then follows, so that (3.1) and (3.6) hold.

REMARK 3.3. In view of Theorem 2.2, a sufficient condition in order that (3.4) may hold is:

(3.4′) *for every positive integer n, the sequence $\{\mu_k(A)\}$ converges uniformly for $A \in \mathcal{J}_{0n}$.*

For fixed integers $n \geqq 1$ and $s \geqq 0$, let $F_{ns}(y_1, y_2, \cdots, y_n)$ denote the joint distribution function of the random variables $\mathbf{x}_{s+1}, \mathbf{x}_{s+2}, \cdots, \mathbf{x}_{s+n}$:

$$\begin{aligned} F_{ns}(y_1, \cdots, y_n) &= \Pr\{\mathbf{x}_{s+1} \leqq y_1, \cdots, \mathbf{x}_{s+n} \leqq y_n\} \\ &= \nu(\{\omega: x_{s+1} \leqq y_1, \cdots, x_{s+n} \leqq y_n\}) \\ &= \nu T^{-s}(I_n) \end{aligned}$$

where I_n is the cylinder set based on the interval

$$\{(x_1, \cdots, x_n): x_1 \leqq y_1, \cdots, x_n \leqq y_n\}.$$

REMARK 3.4. In the statement of Theorem 3.2, condition (3.4) may be replaced by

(3.4″) *for each positive integer n and each set $(y_1, \cdots, y_n)$ of real numbers, the mean value $\overline{F}_n(y_1, \cdots, y_n) = \lim_{k\to\infty} (1/k) \sum_{s=0}^{k-1} F_{ns}(y_1, \cdots, y_n)$ exists and is right continuous in each of its arguments.*

Proof. The mean value with respect to s of the nth difference of F_{ns} over any interval then exists. If A is a set of $\mathcal{C}_n$, a cylinder set based on an interval, then the nth difference of F_{ns} over that interval is precisely $\nu T^{-s}(A)$. It follows that for each positive integer n and for each set $F \in \mathcal{J}_{0n}$, the mean value $\mu(F) = \lim_{k\to\infty} (1/k) \sum_{s=0}^{k-1} \nu T^{-s}(F)$ exists. Further, μ is a measure on $\mathcal{J}_{0n}$, by virtue of the right continuity of $\overline{F}_n$ in each of its arguments [12, pp. 120–124]. Hence condition (3.4″) implies (3.4). The converse is obvious on observing that $F_{ns}(y_1, \cdots, y_n) = \nu T^{-s}(A)$ and $\overline{F}_n(y_1, \cdots, y_n) = \mu(A)$, where A is the cylinder set $\{\omega: x_1 \leqq y_1, x_2 \leqq y_2, \cdots, x_n \leqq y_n\}$.

REMARK 3.5. Since each function $F_{ns}(y_1, \cdots, y_n)$ is right continuous in each of its arguments, a sufficient condition for the right continuity in each of its arguments of $\overline{F}_n(y_1, \cdots, y_n)$ is that

(3.4‴) *for fixed $y_1, y_2, \cdots, y_n$ and for each fixed j ($j = 1, 2, \cdots, n$), the convergence of the sequences $\{(1/k) \sum_{s=0}^{k-1} F_{ns}(y_1, \cdots, y_{j-1}, y, y_{j+1}, \cdots, y_n)\}$*

to $\overline{F}_n(y_1, \cdots, y_{j-1}, y, y_{j+1}, \cdots, y_n)$ *is uniform with respect to* y *for* $0<y-y_j$ *sufficiently small.*

REMARK 3.6. While Theorem 3.2 gives necessary and sufficient conditions for the validity of (3.1) under conditions (3.2) and (3.3), condition (3.5) can not readily be expressed in terms of the finite-dimensional joint distributions of the random variables $\mathbf{x}_1, \mathbf{x}_2, \cdots$. Nor does any of the subsequent remarks suggest how one may verify that (3.5) is satisfied from an examination of the finite-dimensional joint distributions alone, although they determine the probability measure ν on $\mathcal{J}$ completely. However, in view of Theorem 2.2, we have

(3.7) *a sufficient condition in order that* (3.4) *and* (3.5) *shall be satisfied is that* $\{\mu_k(A)\}$ *converge to* $\mu(A)$ *uniformly both with respect to* $A \in \mathcal{J}_{0n}$ *and with respect to* n. For then the convergence is uniform with respect to $A \in \mathcal{J}_0$ and hence with respect to $A \in \mathcal{J}$.

It is clear that if for each positive integer n the random variables $\mathbf{x}_{s+1}, \cdots, \mathbf{x}_{s+n}$ have a joint density function with respect to a probability measure on $R_n = \mathsf{X}_{i=1}^n R$ $(s=0, 1, 2, \cdots)$, then a sufficient condition for (3.4) and (3.5) may be formulated in terms of the uniformity with respect to n of the convergence of the arithmetic means (with respect to s) of these joint density functions.

Condition (3.5) of Theorem 3.2 requires that $\nu \prec \mu$ on $\mathcal{S}^*$, in which case, under the other conditions of the theorem, also $\nu=\mu$ on $\mathcal{S}^*$. A sufficient condition in order that (3.5) shall hold is clearly $\nu=\mu$ on $\mathcal{S}$. Dr. Jim Douglas, Jr., has kindly pointed out to the author that necessary and sufficient conditions in order that $\nu=\mu$ on $\mathcal{S}$ are simply expressed in terms of the quasi-metric introduced by Kakutani, with the aid of limit theorems of Doob [5, Theorems 1.2, 1.3] in a form given them by Andersen and Jessen ([1] and [2]). (The author is indebted to Professor J. L. Doob for calling his attention to these limit theorems.)

If m, n, r are measures defined on a Borel field $\mathcal{B}$, if $m \prec r$ (f) (m is absolutely continuous with respect to r with Radon-Nikodym derivative f) and $n \prec r(g)$, set [10]

$$\rho(m, n; \mathcal{B}) = \int (fg)^{1/2} dr.$$

($-\log \rho$ furnishes the quasi-metric suggested by Kakutani). We assume here the following known propertiesof ρ:

(3.8) ρ is independent of the choice of r subject to $m \prec r$, $n \prec r$; a possible choice is $r=(m+n)/2$;

(3.9) $0 \leqq \rho \leqq 1$;

(3.10) $\rho=1$ if and only if $m=n$;

(3.11) $\rho=0$ if and only if $m \perp n$ (m is orthogonal to n).

Set $p=(\mu+\nu)/2$ on $\mathcal{F}$, so that $\mu<p\ (f)$ and $\nu<p\ (g)$. Instead of introducing different symbols for the contractions of μ, ν, and p to various subfields, let us indicate the appropriate Borel field in parenthesis; thus $\mu<p\ (f;\mathcal{F})$, $\nu<p\ (g;\mathcal{F})$. Define the functions f_n, g_n, etc., by the following relations: $\mu<p\ (f_n;\mathcal{F}_n)$, $\nu<p\ (g_n;\mathcal{F}_n)$, $\mu<p\ (f_{kn};T^{-k}\mathcal{F}_n)$, $\nu<p\ (g_{kn};T^{-k}\mathcal{F}_n)$, $\mu<p\ (f^k;T^{-k}\mathcal{F})$, $\nu<p\ (g^k,\ T^{-k}\mathcal{F})$. For fixed k ($k=0, 1, 2, \cdots$) we have $T^{-k}\mathcal{F}_1\subset T^{-k}\mathcal{F}_2\subset\cdots$, and $T^{-k}\mathcal{F}$ is the smallest Borel field containing $\bigcup_n T^{-k}\mathcal{F}_n$. We then have $\rho(\mu,\nu;\mathcal{F}_n)=\int(f_n\,g_n)^{1/2}dp$, $\rho(\mu,\nu;T^{-k}\mathcal{F}_n)=\int(f_{kn}\,g_{kn})^{1/2}dp$, and $\rho(\mu,\nu;T^{-k}\mathcal{F})=\int(f^k g^k)^{1/2}dp$. From the first limit theorem of Andersen and Jessen ([1] or [2]) it follows that $\lim_{n\to\infty} f_n=f$ a.e. (p), $\lim_{n\to\infty} g_n=g$ a.e. (p), $\lim_{n\to\infty} f_{kn}=f^k$ a.e. (p), and $\lim_{n\to\infty} g_{kn}=g^k$ a.e. (p) $(k=0, 1, 2, \cdots)$. Further, since $p(A)=[\mu(A)+\nu(A)]/2=(1/2)\int_A(f+g)dp$, we have $f+g=2$ a.e. (p) and similarly $f_n+g_n=2$ a.e. (p), $f_{kn}+g_{kn}=2$ a.e. (p). Since each is non-negative, each is bounded by 2 a.e. (p), hence also $(f_n\,g_n)^{1/2}\leqq 2$ a.e. (p), etc. By Lebesgue's bounded convergence theorem, we then have $\lim_{n\to\infty}\rho(\mu,\nu;\mathcal{F}_n)=\rho(\mu,\nu;\mathcal{F})$, and $\lim_{n\to\infty}\rho(\mu,\nu;T^{-k}\mathcal{F}_n)=\rho(\mu,\nu;T^{-k}\mathcal{F})$ (cf. [13, Lemma 7.a]). Further, $\mathcal{F}\supset T^{-1}\mathcal{F}\supset T^{-2}\mathcal{F}\supset\cdots$ and $\mathcal{S}=\bigcap_k T^{-k}\mathcal{F}$. Using the second limit theorem of Andersen and Jessen ([1] or [2]) and an argument similar to that above we find that $\rho(\mu,\nu;\mathcal{S})=\lim_{k\to\infty}\rho(\mu,\nu;T^{-k}\mathcal{F})$. Thus $\rho(\mu,\nu;\mathcal{S})=\lim_{k\to\infty}\lim_{n\to\infty}\rho(\mu,\nu;T^{-k}\mathcal{F}_n)=\lim_{k\to\infty}\lim_{n\to\infty}\int(f_{kn}\,g_{kn})^{1/2}dp$. Thus we have the following theorem.

THEOREM 3.7. *A sufficient condition that $\mu=\nu$ on $\mathcal{S}^*$ is that*

$$\lim_{k\to\infty}\lim_{n\to\infty}\rho(\mu,\nu;T^{-k}\mathcal{F}_n) = 1,$$

that is, that $\lim_{k\to\infty}\lim_{n\to\infty}\int(f_{kn}\,g_{kn})^{1/2}dp=1$.

Indeed, this condition is necessary and sufficient in order that $\mu=\nu$ on $\mathcal{S}$, which contains $\mathcal{S}^*$.

We observe that the condition of Theorem 3.7 is a condition on the finite-dimensional distributions alone.

EXAMPLE 3.8. The roles played by the conditions set forth in the above remarks and Theorem 3.7 may be clarified by the following simple example. Let $\mathbf{x}_s$ denote the number of successes ($\mathbf{x}_s=0$ or 1) in the sth of a Poisson sequence of independent trials having probability p_s of success ($\mathbf{x}_s=1$) in the sth trial ($s=1, 2, \cdots$). The probability function ν is the direct product measure $\mathsf{X}_{s=1}^{\infty}\nu_s$, where ν_s assigns the measure p_s to a Borel set in R_1 (the real line) containing the point 1 but not the point 0, and the measure $q_s=1-p_s$ to a Borel set containing the point 0 but not the point 1. Suppose $\lim_{s\to\infty}p_s=0$. One knows from the Borel-Cantelli lemmas that $\mathbf{x}_s=0$ for s sufficiently large with probability 1 if $\prod q_s$ converges ($\sum p_s$ converges), while $\mathbf{x}_s=1$ for infinitely many s with probability 1 if $\prod q_s$ diverges ($\sum p_s$ diverges). However, it is instructive to examine in this context the conditions expressed in the above remarks and Theorem 3.7. The set functions $\mu_k(A)$ converge as $k\to\infty$ for $A\in\mathcal{J}_0$ to the measure μ which assigns the measure 1 to the set E_0 consisting of the single point $(0, 0, \cdots)$, so that (3.4) is satisfied. However, if $\prod q_s$ diverges then $\nu(E_0)=\nu T^{-1}(E_0)= \cdots =0$, so that $\lim_{k\to\infty}\mu_k(E_0)\neq\mu(E_0)$. On the other hand, if $\prod q_s$ converges, then the convergence of $\{\mu_k(A)\}$ to $\mu(A)$ is *uniform* both with respect to $A\in\mathcal{J}_{0n}$ and with respect to n, so that (3.7) is satisfied. We note also that μ is a direct product measure: $\mu=\mathrm{X}_{s=1}^{\infty}\mu_s$ where μ_s assigns the measure 1 to a set in R_1 containing the origin. It follows [10] that $\rho(\mu, \nu; T^{-k}\mathcal{J}_n)=\prod_{s=k+1}^{k+n}\rho(\mu_s, \nu_s; \mathcal{B})$ where $\mathcal{B}$ is the class of Borel sets in R_1. We have $\rho(\mu_s, \nu_s; \mathcal{B})=(q_s)^{1/2}$, and $\rho(\mu, \nu; \mathcal{S})=\lim_{k\to\infty}\lim_{n\to\infty}\rho(\mu, \nu; T^{-k}\mathcal{J}_n)$ which is 1 if $\prod q_s$ converges, 0 if $\prod q_s$ diverges.

References

1. E. S. Andersen and B. Jessen, *Some limit theorems on integrals in an abstract set*, Det. Kgl. Danske Videnskabernes Selskab, Matematisk-Fysiske Meddelelser vol. 22, no. 14 (1946) pp. 1–29.

2. ———, *Some limit theorems on set-functions*, Det. Kgl. Danske Videnskabernes Selskab, Matematisk-Fysiske Meddelelser vol. 25, no. 5 (1948) pp. 1–8.

3. J. L. Doob, *The law of large numbers for continuous stochastic processes*, Duke Math. J. vol. 6 (1940) pp. 290–306.

4. ———, *Stochastic processes*, New York, Wiley, 1953.

5. ———, *Regularity properties of certain families of chance variables*, Trans. Amer. Math. Soc. vol. 47 (1940) pp. 455–486.

6. Y. N. Dowker, *Invariant measure and the ergodic theorems*, Duke Math. J. vol. 14 (1947) pp. 1051–1061.

7. N. Dunford and D. S. Miller, *On the ergodic theorem*, Trans. Amer. Math. Soc. vol. 60 (1946) pp. 538–549.

8. P. R. Halmos, *Measure theory*, New York, Van Nostrand, 1950.

9. E. Hopf, *Ergodentheorie*, Ergebnisse der Mathematik, vol. 5, no. 2, 1937.

10. S. Kakutani, *On equivalence of infinite product measures*, Ann. of Math. vol. 49 (1948) pp. 214–224.

11. A. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Ergebnisse der Mathematik, vol. 2, no. 3, 1933.

12. M. E. Munroe, *Introduction to measure and integration*, Cambridge, Addison-Wesley, 1953.

13. P. E. Pfeiffer, *Equivalence of totally finite measures on infinite product spaces*, Ann. of Math. vol. 56 (1952) pp. 520–536.

14. F. Riesz, *Sur la théorie ergodique*, Comment Math. Helv. vol. 17 (1945) pp. 221–239.

University of Missouri,
Columbia, Mo.