

ISOMORPHISM AND APPROXIMATION OF GENERAL STATE MARKOV PROCESSES

BY
RICHARD ISAAC⁽¹⁾

1. **Summary.** Given below is a brief description of the main results of this paper:

§3. It is shown that the study of a Markov process $X_n(\Omega, \Sigma, \pi)$ where Ω is a general state space, Σ a separable σ -field of subsets of Ω , and π a σ -finite stationary measure for the process, can essentially be reduced to the study of a real-valued Markov process $Y_n(K, \Delta, \lambda)$ where K is a bounded interval or the entire real line, Δ the Borel sets of K , and λ Lebesgue measure, stationary for the process. To show this a notion of isomorphism of processes is introduced. The main tool is the geometric-isomorphism theorem of Halmos and von Neumann.

§4. Processes $X_n(R, \Delta, \lambda)$ are discussed where R is the real line, and Δ and λ are as defined in §3. The selection of R rather than a bounded interval K is made because we are particularly interested in infinite stationary measures and the infinite case introduces a few technical difficulties. Treatment of the finite case is practically identical to that of the infinite case except that it is simpler. The basic result is that the process $X_n(R, \Delta, \lambda)$ may be approximated in a certain way by processes (called k -processes) which are essentially Markov chains on a countable state space. The nature of the approximation is "weak" convergence of measures on function space. Such an approximation is useful because certain results difficult to prove directly for processes on a continuous state space may be easy to prove for Markov chains, and then carried over to the general process by using the k -processes. (Similar approaches have been used to approximate continuous time by discrete time processes; in this paper it is the state space rather than the time parameter which is discretized—time is discrete throughout.) Next, if the process is conservative and ergodic (that is, the shift T is; see [9] and Halmos' *Lectures on ergodic theory*) we derive a few conclusions about the nature of the k -processes.

§5. As an application of §4, we give a probabilistic proof of Birkhoff's ergodic theorem.

§6. Using the isomorphism of §3, we extend the results of the paper to processes $X_n(\Omega, \Sigma, \pi)$ for Σ separable.

2. Some notations and conventions.

Complement of E in the universe = $C(E)$.

Complement of E in F = $F - E$.

Closure of E in a given topology = $\text{Cl}(E)$.

Received by the editors November 15, 1967.

⁽¹⁾ Partially supported by NSF grant GP-3819.

If μ is a measure, x a point, $\langle a, b \rangle$ some interval of real numbers, we write $\mu(x)$ for $\mu(\{x\})$, $\mu\langle a, b \rangle$ for $\mu(\langle a, b \rangle)$. F is called μ -full, μ -null or μ -finite if $\mu(C(F))=0$, $\mu(F)=0$ or $\mu(F)<\infty$, respectively. $E_{\mu}f$ is the expectation of a function f with respect to the measure μ . μ/F or μ/Σ for a subset F or a sub- σ -field Σ is the restriction of μ to this set or this σ -field. $X_n(\Omega, \Sigma, \pi)$ refers to a Markov process with state space Ω , σ -field Σ of subsets of Ω , stationary measure π and stationary transition probabilities. All measures are understood to be σ -finite. If Σ_k is an increasing sequence of σ -fields, $\lim_k \Sigma_k$ is the smallest σ -field containing each Σ_k . \lim_k means $\lim_{k \rightarrow \infty}$. A set F is *stochastically closed* if $P(x, F)=1$ for every $x \in F$; F is a.e. stochastically closed if this relation holds only a.e. on F . For F a set, Σ a σ -field, $F \cap \Sigma$ is the restriction of Σ to subsets of F .

3. The isomorphism theorem. Topological terminology follows Kelley [14]. Σ is *separable* if it is countably generated. In [7] this word is used in a slightly weaker sense, the term *strict separability* employed where we write "separable." If Σ is separable, Σ is atomic, and the atoms can be written as countable intersections of generating sets [1]. A measure μ is continuous at x if $\mu(x)=0$; it is continuous if it is continuous at each $x \in \Omega$. In a metric space, the weight of a μ -atom must be concentrated at a point, so that if μ is continuous, μ is nonatomic [7, p. 334].

LEMMA 3.1. *Let Σ be a separable σ -field of subsets of a set Ω such that each singleton set is measurable. There is a topology \mathcal{T} on Ω such that*

- (i) Ω is a separable metric space under \mathcal{T} .
- (ii) the Borel σ -field Δ of \mathcal{T} is precisely Σ .

Proof. Let $\{E_i\}$ be a countable class generating Σ . Define \mathcal{T} on Ω by taking as subbase for \mathcal{T} the class $\mathcal{S}=\{E_i\} \cup \{C(E_i)\}$. Then \mathcal{S}^* , the smallest field of sets over \mathcal{S} , is a base for \mathcal{T} , is countable, and each set in \mathcal{S}^* is both open and closed. $\Delta \supseteq \Sigma$, for $\Delta \supseteq \{E_i\}$. Conversely, $\Sigma \supseteq \mathcal{S}$ and so $\Sigma \supseteq \Delta$. This proves (ii). Each point is expressible as an at most countable intersection of elements of \mathcal{S} , so that points are closed sets. An arbitrary neighborhood of a point contains a closed neighborhood of the point, namely a base element from \mathcal{S}^* , and so \mathcal{T} is regular. Urysohn's metrization theorem [14, p. 125] then proves Ω separable metric, yielding (i).

REMARK 1. In discussing Markov processes, one may automatically assume each singleton set measurable if Σ is separable. To see this, let $\{V_{\alpha}\}=\Omega^*$ be the class of atoms of Σ . A Markov process with Ω^* as state space may be defined in a natural way. Let Σ^* be the σ -field consisting of all subsets of Ω^* such that $E^* \in \Sigma^*$ and $E^* = \bigcup_{\alpha \in W} \{V_{\alpha}\}$ (W is some index set) if and only if $E \in \Sigma$ and $E = \bigcup_{\alpha \in W} V_{\alpha}$. Define $P^*(\{V_{\beta}\}, E^*) = P(x, E)$ where E and E^* are as related above and x is any point in the atom V_{β} . This yields a well-defined transition probability because $P(\cdot, E)$ is constant on atoms of Σ by measurability. The transition probability P^* induces a Markov process $\{X_n^*\}$ which is "probabilistically equivalent" to $\{X_n\}$ in an obvious interpretation of this phrase. Moreover, by construction all singleton sets of Ω^* are measurable.

In the following, when we write Tx and TE , it is always to be assumed that the statement makes sense; that is, the appropriate quantities are in the domain of T .

DEFINITION. Let $\{X_n\}$ and $\{X_n^*\}$ be Markov processes on (Ω, Σ, π) and $(\Omega^*, \Sigma^*, \pi^*)$ respectively. $X_n(\Omega, \Sigma, \pi)$ is said to be isomorphic to $X_n^*(\Omega^*, \Sigma^*, \pi^*)$ if there exists a one-one map T from almost all (π) of Ω on almost all (π^*) of Ω^* such that

- (i) $E \in \Sigma$ if and only if $TE \in \Sigma^*$ and then $\pi(E) = \pi^*(TE)$.
- (ii) If P^* is the transition probability function for $\{X_n^*\}$

$$P(x, E) = P^*(Tx, TE).$$

DEFINITION. A measure space (Ω, Σ, π) is *complete* if every subset of a set of measure zero is itself measurable. It is *properly separable* if there exists a separable σ -field $\Sigma_0 \subseteq \Sigma$ such that, to every $E \in \Sigma$ there corresponds an $F \in \Sigma_0$ with $E \subseteq F$ and $\pi(F - E) = 0$.

Let K be either a finite closed interval of real numbers, the entire real line or the empty set; let W be an at most countable index set, perhaps empty, $\{p_i\}$ an abstract set of points, and set $A = \bigcup_{i \in W} p_i$. Let \mathcal{M} be the σ -field consisting of the Borel subsets of K and all subsets of A .

THEOREM 3.1 (THE ISOMORPHISM THEOREM). *Let $X_n(\Omega, \Sigma, \pi)$ be a Markov process on separable Σ . Then $X_n(\Omega, \Sigma, \pi)$ is isomorphic to a process $Y_n(K \cup A, \mathcal{M}, \mu)$. If K is not void, $\mu|K$ is Lebesgue measure. If A is not void, $\mu(p_i) > 0$ for each $i \in W$. We always have $\pi(\Omega) = \mu(K) + \mu(A)$ so that if π is continuous, μ reduces to Lebesgue measure on a bounded interval or on the entire real line.*

Proof. Decompose Σ into an atomic part and a nonatomic part with respect to π [1]. Since the atomic part must be concentrated on points by Remark 1 and Lemma 3.1, there are at most a countable number of points $x_i \in \Omega$ with $\pi(x_i) > 0$. On the complement of this at most countable set, π is nonatomic (and may be identically zero). For notational convenience, suppose this π -nonatomic part is Ω itself and let $\pi(\Omega) = 1$. Let Σ' be the π -completion of Σ . Then (Ω, Σ', π) is properly separable with Σ as associated separable σ -field. Under the separable metric topology \mathcal{T} defined in Lemma 3.1, Ω may be embedded in its metric completion Ω^* . Extend π to a measure π^* on $\Sigma^* = \Sigma' \cup \mathcal{N}$, where \mathcal{N} is the class of all subsets of $\Omega^* - \Omega$, by putting $\pi^*(N) = 0$ for $N \subseteq \Omega^* - \Omega$. A Markov process $X_n^*(\Omega^*, \Sigma^*, \pi^*)$ may be defined by setting $P^*(x, E^*) = P(x, E^* \cap \Omega)$ for $x \in \Omega$, $E^* \in \Sigma^*$, and $P^*(x, \{x\}) = 1$ for $x \in \Omega^* - \Omega$. π^* is stationary for the process $\{X_n^*\}$, π^* is continuous, Σ^* is complete and properly separable, and Ω^* is topologically complete. Without loss of generality, then, it may be assumed at the outset that \mathcal{T} on Ω yields a complete separable metric space. Thus, Ω is absolutely closed [10, p. 142], hence absolutely Borel. It follows that if f is a one-one real-valued Σ -measurable function on Ω , $f(\Omega)$ is a Borel set in the reals [10, p. 301]. Now, Ω is a metric space under \mathcal{T} , Σ is its Borel field and π is continuous on Ω , so (Ω, Σ', π) is a *normal* measure space in the sense of Halmos and von Neumann [7, p. 336]. This is easily

checked by referring to their definitions. Let λ be Lebesgue measure. By the geometric isomorphism theorem [7, p. 339] there exists a one-one measure preserving map T from almost all $(\pi)\Omega$ on almost all $(\lambda)I$, the unit interval, such that $E \in \Sigma'$ if and only if TE is Lebesgue measurable. Moreover, a study of the proofs of Theorems 1 and 2 of [7] shows $E \in \Sigma$ if and only if TE is a Borel set of I . Let N be the π -null set on which T fails to be defined. From the stationarity of π , one easily proves the existence of a stochastically closed π -full set $\Lambda \subseteq C(N)$ such that T is defined everywhere on Λ . Let $\{x_i, i \in W\}$ be the set of π -atoms, put $A_1 = \bigcup_{i \in W} x_i$, and $\mathcal{M}_1 = \Sigma \cap (\Lambda \cup A_1)$. Let p_i be an abstract point for each $i \in W$, and set $Tx_i = p_i$. T is now defined on $\Lambda \cup A_1$; let us put

$$P^*(Tx, TE) = P(x, E); \quad x \in \Lambda \cup A_1, E \in \mathcal{M}_1.$$

P^* is clearly a transition probability defined on $T(\Lambda \cup A_1) \times \mathcal{M}$. To extend P^* to $I \cup A$, where $TA_1 = A$, put

$$P^*(y, \{y\}) = 1, \quad y \in I - T\Lambda.$$

Define μ on \mathcal{M} by putting $\mu(E) = \lambda(E)$ for $E \in \Delta$, the Borel field of I , $\mu(E) = \pi(T^{-1}E)$ for $E \subseteq A$. Then for $E \in \mathcal{M} \cap T(\Lambda \cup A_1)$, $\mu(E) = \pi(T^{-1}E)$ and

$$\begin{aligned} \int_{I \cup A} P^*(y, E)\mu(dy) &= \int_{T(\Lambda \cup A_1)} P^*(y, E \cap [T(\Lambda \cup A_1)])\mu(dy) \\ &= \int_{\Lambda \cup A_1} P(T^{-1}y, T^{-1}\{E \cap [T(\Lambda \cup A_1)]\})\pi(dT^{-1}y) \\ &= \pi(T^{-1}\{E \cap [T(\Lambda \cup A_1)]\}) \\ &= \mu(E \cap [T(\Lambda \cup A_1)]) = \mu(E), \end{aligned}$$

so that μ is stationary for the $\{Y_n\}$ process. This proves the theorem for $\pi(\Omega) = 1$, hence for any bounded interval. If $\pi(\Omega) = \infty$, using σ -finiteness and nonatomicity, partition Ω into sets $E_i, i = 0, \pm 1, \dots, \pi(E_i) = 1$, and map E_i into $[i, i + 1]$ by T_i according to the above procedure. This defines a map T on almost all Ω into almost all R , and the process is easily defined as before. This concludes the proof.

REMARK 2. For our purposes, the π -atomic part A is a nuisance and uninteresting. Therefore, we shall usually assume it is absent to simplify proofs. Trivial modifications of our arguments are necessary to adapt our demonstrations to processes with a nontrivial atomic part.

4. **k -process approximation.** Theorem 3.1 indicates the range of generality of stationary Markov processes on separable Σ : their study essentially reduces to the study of real-valued processes on the Borel field with Lebesgue measure as stationary measure. In this section we take processes $X_n(R, \Delta, \lambda)$ where R is the real line, Δ the Borel field, and λ Lebesgue measure, and prove that we can approximate the process in a certain way by simpler processes that look like Markov chains. The results of this section will be valid for the most general process

$Y_n(K \cup A, \mathcal{M}, \mu)$ described in Theorem 3.1; trivial modifications of proofs and statements are then necessary. By Remark 2, we are taking A to be empty and losing nothing thereby; therefore we have two distinct cases: K = a finite interval, say I , the unit interval, and $K = R$. This corresponds to μ finite in the first case, or μ infinite in the second. The treatment of the two cases in the study of k -process approximation is practically identical, except that the circumstance of an infinite measure presents a few technical nuisances in writing the proofs. Since this is so, we have chosen to work with R and let the reader simplify our arguments to get the finite case rather than working the other way around.

Let $X_n(R, \Delta, \lambda)$ be a Markov process as described above. Consider a sequence of interval partitions of R , each a refinement of the preceding. To fix ideas, let

$$\begin{aligned} \mathcal{G}_1 &= \{(a, a+1]; a = 0, \pm 1, \pm 2, \dots\}, \\ \mathcal{G}_2 &= \{(a, a+2^{-1}]; a = 0, \pm 1.2^{-1}, \pm 2.2^{-1}, \dots\}, \end{aligned}$$

and in general

$$\mathcal{G}_n = \{(a, a+2^{1-n}]; a = 0, \pm 1.2^{1-n}, \pm 2.2^{1-n}, \dots\}.$$

Let Δ_k be the σ -field generated by \mathcal{G}_k .

LEMMA 4.1. Δ_k is an increasing sequence of atomic σ -fields with \mathcal{G}_k the class of atoms of Δ_k and $\lim_k \Delta_k = \Delta$.

Proof. Only the last assertion needs proof. $\lim_k \Delta_k$ is, by definition, the smallest σ -field containing all semiclosed intervals with dyadic rational endpoints, that is with endpoints of the form $p2^{-q}$ for p and q integers. This set is dense in R , proving the lemma.

Let

$$(4.1) \quad P_k(\cdot, E) = E_\lambda\{P(\cdot, E) \mid \Delta_k\}$$

for each $E \in \Delta$. $P_k(\cdot, E)$ is defined a.e. (λ) on R by (4.1), is Δ_k -measurable, and satisfies

$$(4.2) \quad \int_A P_k(x, E) d\lambda_k = \int_A P(x, E) d\lambda$$

for all $A \in \Delta_k$, where λ_k is the restriction of λ to Δ_k . If A is an atom of Δ_k , i.e., if $A \in \mathcal{G}_k$, $P_k(x, E)$ reduces to a constant a.e. (λ) on A and (4.2) yields

$$(4.3) \quad P_k(x, E) = \left(\int_A P(z, E) d\lambda \right) (\lambda(A))^{-1}$$

for almost all $x \in A$. Now obtain a well-defined function $P_k(\cdot, \cdot)$ on $R \times \Delta$ by using (4.3) as definition for every $x \in A \in \mathcal{G}_k$, $E \in \Delta$.

LEMMA 4.2. For each $x \in R$, $P_k(x, \cdot)$ is a measure on Δ and $P_k(\cdot, \cdot)$ is a transition probability on $R \times \Delta_k$.

Proof. For each $x \in R$, x is an element of some unique atom A in \mathcal{G}_k . Defining $P_k(x, \cdot)$ by (4.3), it is clear that it is countably additive on Δ . For fixed $E \in \Delta$, $P_k(\cdot, E)$ is Δ_k -measurable and $P_k(x, R) = 1$ for all $x \in R$, so that $P_k(\cdot, \cdot)$ on $R \times \Delta_k$ is a transition function.

LEMMA 4.3. $\lim_k P_k(x, E) = P(x, E)$ a.e. (λ) for every fixed $E \in \Delta$; the exceptional null set may depend upon E .

Proof. (4.1) defines a martingale, so the conclusion is immediate by using Theorem 4.3 of [3, p. 331]. (Although λ on R is infinite, restrict attention to the trace on individual atoms of Δ_1 , getting a martingale on each.)

THEOREM 4.1. *There is a Markov process $\{X_n^k, n \geq 0\}$ defined on (R, Δ_k) with one-step transition probability function $P_k(x, E)$, $x \in R$, $E \in \Delta_k$. The process has the stationary measure $\lambda_k = \lambda/\Delta_k$.*

Proof. For $E \in \Delta_k$,

$$\int P_k(x, E) \lambda_k(dx) = \int P(x, E) \lambda(dx) = \lambda(E) = \lambda_k(E),$$

and so λ_k is stationary for the transition function $P_k(x, E)$ on $R \times \Delta_k$. Given this transition function and λ_k as initial measure (or any probability measure on Δ_k instead) a Markov process $\{X_n^k\}$ may be constructed in a standard way [see 3, p. 190 and 9].

REMARK 3. Let $\{V_n^k\}$ be random variables defined by $V_n^k = a$ if and only if $X_n^k \in (a - 2^{1-k}, a] = A$ for $A \in \mathcal{G}_k$. Let

$$P_k^*(a, b) = P_k(x, B) \quad \text{for } x \in A \in \mathcal{G}_k$$

and $(b - 2^{1-k}, b] = B \in \mathcal{G}_k$.

$\{V_n^k\}$ is a Markov chain for each fixed k with countable state space, one-step transition probability function P_k^* and stationary measure $\lambda_k^*(a) = \lambda_k(A) = 2^{1-k}$ where A is as defined above. The Markov chain $\{V_n^k\}$ and the process $X_n^k(R, \Delta_k, \lambda_k)$ "look exactly alike" where the point a in the state space of the chain corresponds to the atom A of the $\{X_n^k\}$ process. We call the processes $\{X_n^k\}$ " k -processes"; it is evident that Markov chain terminology ("irreducible," "persistent," etc., see [5]) can be employed for the k -processes with obvious meanings.

Let m be the measure on infinite-dimensional Euclidean space R'_∞ of points $\omega = (x_0, x_1, \dots)$ determined by λ and the transition probability $P(\cdot, \cdot)$ of the $\{X_n\}$ process. Thus, on cylinder sets of the form $\prod_{i=0}^n B_i \times \prod_{i=n+1}^\infty R^{(i)} = C$, $B_i \in \Delta$, $R^{(i)} = R$, we have [3, p. 190]

$$(4.4) \quad m(C) = \int_{B_0} \lambda(dz) \int_{B_1} P(z, dz_1) \cdots \int_{B_n} P(z_{n-1}, dz_n).$$

Since each $P_k(z, \cdot)$ is a measure on Δ , for each k there is a measure μ_k on R'_∞ such that on cylinders C of the above form one has

$$(4.5) \quad \mu_k(C) = \int_{B_0} \lambda_k(dz) \int_{B_1} P_k(z, dz_1) \cdots \int_{B_n} P_k(z_{n-1}, dz_n).$$

However, we shall be more interested in the measure m_k on R'_∞ determined by λ_k and $P_k(\cdot, \cdot)$ on $R \times \Delta_k$. This is the measure corresponding to the process $\{X_n^k\}$ and the related Markov chain $\{V_n^k\}$. m_k is induced on R'_∞ in the following way: Let C be a cylinder of the above type where $B_i = (-\infty, b_i]$, $0 \leq i \leq n$, then

$$(4.6) \quad m_k(C) = \int_{-\infty}^{c_0} \lambda_k(dz) \int_{-\infty}^{c_1} P_k(z, dz_1) \cdots \int_{-\infty}^{c_n} P_k(z_{n-1}, dz_n)$$

where $c_i \leq b_i$ and c_i is the unique right-hand endpoint of an interval in \mathcal{G}_k as close to b_i as possible. m_k is thus a measure induced by a discrete joint distribution related to the chain $\{V_n^k\}$. The measure m_k is identical to the measure on Borel sets induced by this chain. The integral in (4.6) is computed with respect to the measures $P_k(z, \cdot)$ as measures over Δ ; since the sets $(-\infty, c_i]$ are elements of Δ_k , the measures $P_k(z, \cdot)$ are restricted to Δ_k and then the integral computed as a simple sum.

Fix k throughout the following paragraph. Consider the n -dimensional distribution functions

$$(4.7) \quad F_k(x_1, x_2, \dots, x_n) = m_k(X_0^k \in (0, 1]; X_i^k \leq x_i, 1 \leq i \leq n).$$

Let R_∞ be infinite-dimensional Euclidean space of points $\omega = (x_1, x_2, \dots)$ and let P_k and P be the measures induced on R_∞ by F_k and F respectively, where F is defined by (4.7) by removing k everywhere. The measures P_k and P are introduced because they are finite; notice that if $B \in R_\infty$, then $(0, 1] \times B = B'$ is in R'_∞ , and

$$P_k(B) = m_k(B'); \quad P(B) = m(B').$$

By “ \Rightarrow ” we mean weak-* convergence of measures [4, p. 462; 15]; often called “weak convergence” by probabilists. Unfortunately this term has a different meaning in Banach space theory. Since both concepts play a role in this paper, we adhere to the terminology of functional analysis.

Our basic result is

THEOREM 4.2 (THE APPROXIMATION THEOREM). $P_k \Rightarrow P$.

Proof. $P_k(R_\infty) = P(R_\infty) = 1$, so to prove the theorem it suffices to show $F_k(x_1, x_2, \dots, x_n) \rightarrow F(x_1, x_2, \dots, x_n)$ on a dense set (x_1, x_2, \dots, x_n) in R_n , Euclidean n -space, for each $n = 1, 2, \dots$ [16]. Take the dense set D of dyadic rationals, i.e., endpoints of intervals in $\bigcup_k \mathcal{G}_k$. Fix n and x_1, x_2, \dots, x_n in D . For k sufficiently large, each $x_i, 1 \leq i \leq n$, will be endpoints of \mathcal{G}_k intervals. Assuming k is large enough to guarantee this, if $C = (0, 1] \times \prod_{i=1}^n (-\infty, x_i]$

$$(4.8) \quad \begin{aligned} F_k(x_1, x_2, \dots, x_n) &= m_k(C) = \mu_k(C) \\ &= \int_0^1 \lambda_k(dz) \int_{-\infty}^{x_1} P_k(z, dz_1) \cdots \int_{-\infty}^{x_n} P_k(z_{n-1}, dz_n) \end{aligned}$$

as can be seen by observing the equality of (4.5) and (4.6) because of the nature of x_1, x_2, \dots, x_n .

Define measures α_k^j and α^j on Δ for $1 \leq j \leq n+1$ by induction on j , satisfying

$$(4.9) \quad \alpha_k^j(R) \leq 1; \quad \alpha^j(R) \leq 1,$$

$$(4.10) \quad \alpha^j(E) \leq \lambda(E), \quad E \in \Delta,$$

$$(4.11) \quad \lim_k \alpha_k^j(E) = \alpha^j(E), \quad E \in \Delta.$$

Define α_k^1 and α^1 by

$$\alpha_k^1(E) = \int_0^1 \lambda_k(dz)P_k(z, E), \quad \alpha^1(E) = \int_0^1 \lambda(dz)P(z, E).$$

Since $(0, 1] \in \Delta_k$ for each k , $\alpha_k^1 \equiv \alpha^1$ by the definition of $P_k(\cdot, E)$ as a conditional expectation. Moreover $\alpha^1(R)=1$ and $\alpha^1(E) \leq \lambda(E)$ by stationarity. Thus (4.9), (4.10) and (4.11) are satisfied for α_k^1 and α^1 . Assume α_k^j and α^j have been defined. Set

$$\alpha_k^{j+1}(E) = \int_{-\infty}^{x_j} \alpha_k^j(dz)P_k(z, E) \quad \alpha^{j+1}(E) = \int_{-\infty}^{x_j} \alpha^j(dz)P(z, E).$$

These quantities are measures on Δ . (4.9) is satisfied, in fact, $\alpha_k^{j+1}(R) \leq \alpha_k^j(-\infty, x_j] \leq \alpha_k^j(R) \leq 1$, and similarly for $\alpha^{j+1}(R)$. (4.10) follows from

$$\alpha^{j+1}(E) = \int_{-\infty}^{x_j} \alpha^j(dz)P(z, E) \leq \int_{-\infty}^{x_j} \lambda(dz)P(z, E) \leq \lambda(E)$$

by the induction assumption. Now, (4.9) and (4.11) imply α_k^j converges to α^j weakly ([4, p. 308]; this is stronger than weak-* convergence). If f is any bounded measurable function on (Ω, Σ) ,

$$\int f d\mu = x^*(\mu)$$

defines a bounded linear functional x^* on the Banach space of bounded countably additive set functions on Σ (see [4]) and μ_k converging to μ weakly implies

$$\int f d\mu_k \rightarrow \int f d\mu.$$

Let l be fixed and $k > l$, then

$$(4.12) \quad \int_{-\infty}^{x_j} \alpha_k^j(dz)P_k(z, E) \leq \int_{-\infty}^{x_j} \alpha_k^j(dz) \sup_{k > l} P_k(z, E)$$

and

$$(4.13) \quad \int_{-\infty}^{x_j} \alpha_k^j(dz)P_k(z, E) \geq \int_{-\infty}^{x_j} \alpha_k^j(dz) \inf_{k > l} P_k(z, E).$$

The right-hand integrands in (4.12) and (4.13) are bounded, so the above remarks on weak convergence applied to (4.12) and (4.13) as $k \rightarrow \infty$ give

$$(4.14) \quad \limsup_k \int_{-\infty}^{x_j} \alpha_k^j(dz)P_k(z, E) \leq \int_{-\infty}^{x_j} \alpha^j(dz) \sup_{k>l} P'_k(z, E)$$

and

$$(4.15) \quad \liminf_k \int_{-\infty}^{x_j} \alpha_k^j(dz)P_k(z, E) \geq \int_{-\infty}^{x_j} \alpha^j(dz) \inf_{k>l} P_k(z, E).$$

Remembering that $\alpha^j(N) = 0$ whenever $\lambda(N) = 0$, Lemma 4.3 and letting $l \rightarrow \infty$ in (4.14) and (4.15), we conclude

$$\lim_{k \rightarrow \infty} \alpha_k^{j+1}(E) = \alpha^{j+1}(E),$$

which is (4.11), completing the construction by induction. In particular, if C is the set appearing in (4.8)

$$(4.16) \quad \begin{aligned} \lim_{k \rightarrow \infty} F_k(x_1, x_2, \dots, x_n) &= \lim_{k \rightarrow \infty} \mu_k(C) = \lim_{k \rightarrow \infty} \alpha_k^n(-\infty, x_n] \\ &= \alpha^n(-\infty, x_n] = m(C) = F(x_1, x_2, \dots, x_n). \end{aligned}$$

Since x_1, x_2, \dots, x_n are fixed but arbitrarily chosen elements of D , and n is arbitrary, (4.16) is valid on a dense set of R_n for each n , proving the theorem.

THEOREM 4.3. *Let $f(x_1, x_2, \dots)$ be a real-valued function jointly continuous in x_1, x_2, \dots . Set*

$$(4.17) \quad G_k(x) = P_k(f(x_1, x_2, \dots) < x)$$

and define $G(x)$ correspondingly by removing k in (4.17). Then $G_k \Rightarrow G$.

Proof. For the distribution functions G_k and G , the assertion of the theorem is equivalent to proving $G_k(x) \rightarrow G(x)$ on a dense set of R . By joint continuity of f , the set $[f(x_1, x_2, \dots) < x] = \mathcal{O}_x$ is open in R_∞ and $P(\text{Cl}(\mathcal{O}_x)) = P(\mathcal{O}_x)$ on the complement of a countable set of x . P is regular so by Alexandroff's theorem [4, p. 316] and Theorem 4.2, we have $P_k(\mathcal{O}_x) \rightarrow P(\mathcal{O}_x)$ on a dense set, proving the theorem.

COROLLARY 4.1. *If $f(x_1, x_2, \dots)$ can be made jointly continuous in x_1, x_2, \dots by altering its values on a set N^∞ with $\lim_k P_k(N^\infty) = P(N^\infty) = 0$, then Theorem 4.3 is still valid.*

Proof. Let $f^*(x_1, x_2, \dots)$ be a jointly continuous function agreeing with f on $C(N^\infty)$, let $[f^*(x_1, x_2, \dots) < x] = \mathcal{O}_x^*$, and \mathcal{O}_x be the corresponding set for f . $\mathcal{O}_x^* - \mathcal{O}_x \subseteq N^\infty$, and on a dense set of $x \in R$

$$\lim_k P_k(\mathcal{O}_x) = \lim_k [P_k(\mathcal{O}_x) + P_k(\mathcal{O}_x^* - \mathcal{O}_x)] = \lim_k P_k(\mathcal{O}_x^*) = P(\mathcal{O}_x^*) = P(\mathcal{O}_x),$$

concluding the proof.

COROLLARY 4.2. *If $f(x_1, x_2, \dots)$ can be made jointly continuous in x_1, x_2, \dots by altering its value on a P -null set of the form $N^\infty = (X_n \in N)$ for some $n, 1 \leq n < \infty$, then there is a k -process approximation such that Theorem 4.3 is still valid.*

Proof. Since $0 = P(N^\infty) = \int P^n(x, N)\lambda(dx)$ and λ is stationary, N is λ -null, and so has void interior on R . It is clear that by defining the original partition sets \mathcal{G}_k somewhat differently, the set of endpoints of \mathcal{G}_k sets can be made disjoint with N . It then follows for the related k -processes that $P_k(N^\infty) = 0$ for all k (remember that P_k is a discrete distribution jumping at the endpoints of \mathcal{G}_k sets). Finally, apply Corollary 4.1.

Introduce the two recurrence conditions (see, e.g., [11], [12], [13]):

CONDITION (B). $\lambda(E) > 0$ implies $P(X_n \in E \text{ infinitely often} \mid X_0 = x) = 1$ a.e. (λ) and

CONDITION (C). Same as Condition (B), except “a.e. (λ)” is replaced by “for every $x \in R$.”

It is not hard to see that Condition (B) is equivalent to requiring the Markov operator P to be conservative and ergodic [11]. Condition (C) is considerably more restrictive, and implies the indecomposability of the process in the sense of Doeblin [13].

THEOREM 4.4. *If Condition (B) holds for the $\{X_n\}$ process, each k -process is irreducible.*

Proof. Let $E \in \mathcal{G}_k$ and \mathcal{U}_1 be the class of atoms of \mathcal{G}_k such that $P_k(x, E) > 0$ for each $x \in U \in \mathcal{U}_1$. \mathcal{U}_1 is nonempty because $\lambda_k(E) > 0$ and λ_k is stationary for the k -process. By induction, define \mathcal{U}_j as the class of atoms in \mathcal{G}_k such that $P_k(x, U) > 0$ for each $x \in V \in \mathcal{U}_j$ for some $U \in \mathcal{U}_{j-1}$. Then $\bigcup_{n=1}^\infty \mathcal{U}_n = \mathcal{U}$ is the class of atoms of \mathcal{G}_k which can ever “reach” E . The complement of the union of elements in \mathcal{U} , say S , is stochastically closed for the k -process, so if S is nonempty, $P_k(x, S) = 1$ for all $x \in S$. Since S , if nonempty, has positive λ_k measure, we find $P(x, S) = 1$ a.e. (λ) on S , contradicting Condition (B).

If Condition (B) holds, it is reasonable to expect the k -processes to behave like persistent Markov chains. The precise statement is contained in

THEOREM 4.5. *If Condition (B) holds for the $\{X_n\}$ process, all but a finite number of the k -processes are persistent.*

Proof. Suppose that there is a subsequence k' such that the irreducible k' -processes are transient. Let $I[\cdot]$ be the indicator (characteristic) function of the event in brackets, let $E = (0, 1]$. The function

$$(4.18) \quad \limsup_n I[x_n \in E] = f(x_1, x_2, \dots)$$

is jointly continuous on R_∞ except on a subset of the set S of sequences $\{x_n\}$ containing a subsequence converging to 0 or 1 and no subsequence converging to

an interior point of E . If A is a bounded interval, $\sum_{n=1}^{\infty} P_k^n(x, A) < \infty$, and so transiency implies $\lim_n |x_n| = \infty$ a.e. (P_k) , so that $P_k(S) = 0$. Moreover, $P(S) = 0$ by Condition (B). On a dense set of x we have $P_k(f < x) \xrightarrow{k' \rightarrow \infty} P(f < x)$ by Corollary 4.1, where now f is (4.18). Since $P_k(f < x) =$ the distribution with unit jump at zero (since (4.18) = 0 a.e. for transient chains) and $P(f < x)$ has unit jump at $x = 1$ by Condition (B), the resulting contradiction proves that for $k > M$, say, (4.18) is equal to 1 a.e. (P_k) . Each k -process is irreducible so must then be persistent for $k > M$.

5. Probabilistic proof of the ergodic theorem. To fix ideas, let $X_n(\Omega, \Sigma, \pi)$ be a Markov process on separable Σ with stationary nonatomic measure π . Then, by Theorem 3.1, $X_n(\Omega, \Sigma, \pi)$ is isomorphic to a real process $Y_n(K, \Delta, \lambda)$ where $K = I = [0, 1]$ (for π a probability) or $K = R$ (for π infinite). Let $f(X_n)$ be a functional of the $\{X_n\}$ process, $E_\pi |f| < \infty$, and define $f^*(Y_n) = f(X_n)$ where T is the isomorphism mapping, that is f^* is an a.e. (λ) defined functional of the $\{Y_n\}$ process given by

$$f^*(y) = f(T^{-1}y),$$

for almost all $(\lambda) y \in R$. Clearly $E_\lambda |f^*| = E_\pi |f| < \infty$, and

$$(6.1) \quad \lim_n \frac{1}{n} \sum_{k=1}^n f(X_k) \quad \text{a.e. } (\pi) \text{ on } \Omega$$

if and only if

$$(6.2) \quad \lim_n \frac{1}{n} \sum_{k=1}^n f^*(Y_k) \quad \text{a.e. } (\lambda) \text{ on } R.$$

Now suppose $\{V_n\}$ is a *Markov chain*; without loss of generality $\{V_n\}$ may be assumed irreducible, and let f be a functional on $\{V_n\}$ with $E_\mu |f| < \infty$ where μ is a stationary measure for the chain. If $\{V_n\}$ is transient, irreducibility and stationarity imply $\{V_n\}$ has a countable state space and μ is infinite. Then $\{V_n\}$ eventually leaves the μ -finite set $[|f| \geq \varepsilon]$. Thus

$$(6.3) \quad \lim_n \frac{1}{n} \sum_{k=1}^n f(V_k) = 0 \quad \text{a.e.}$$

If $\{V_n\}$ is persistent, the left-hand side of (6.3) also exists a.e. This is an immediate consequence of results in Chung [17, part I, §§14 and 15] done completely by probability theory. The basic idea is to use the strong law of large numbers on the independent random variables obtained by stopping the chain every time it returns to a given point. We therefore see that the ergodic theorem holds for functionals of general Markov chains. Now let f be a functional on the real process $Y_n(K, \Delta, \lambda)$, where K is as defined in the first paragraph, and let $E_\lambda |f| < \infty$. Let $\{Y_n^k\}$ be a k -process approximation to $\{Y_n\}$; the functionals $f(Y_n^k)$ satisfy $E_{\lambda_k} |f| < \infty$. We have, by the result for Markov chains

$$(6.4) \quad \lim_n \sup \frac{1}{n} \sum_{j=1}^n f(Y_j^k) = s_1(Y_1^k, Y_2^k, \dots)$$

and

$$(6.5) \quad \liminf_n \frac{1}{n} \sum_{j=1}^n f(Y_j^k) = s_2(Y_1^k, Y_2^k, \dots)$$

are equal a.e. The functions $s_i(x_1, x_2, \dots)$, $i=1, 2$, are jointly continuous in their arguments, so by Theorem 4.3

$$(6.6) \quad P_k(s_i(x_1, x_2, \dots) < x) \rightarrow P(s_i(x_1, x_2, \dots) < x)$$

for $i=1, 2$. Since

$$(6.7) \quad P_k(s_1(x_1, x_2, \dots) = s_2(x_1, x_2, \dots)) = 1$$

for each k , (6.6) implies (6.7) for P_k replaced by P . This is the ergodic theorem for functionals of the process $\{Y_n\}$.

Let $\{Z_n, -\infty < n < \infty\}$ be a strictly stationary stochastic process with nonatomic stationary measure μ defined on bilateral coordinate space, and let $E_\mu|Z_1| < \infty$.

The following idea is related to one used by Harris [8]. Define

$$X_n = (Z_n, Z_{n-1}, Z_{n-2}, \dots).$$

We claim that we can make the random vectors $\{X_n\}$ into a strictly stationary Markov process with stationary transition probabilities and with unilateral infinite-dimensional coordinate space R_∞ as state space. First, due to σ -finiteness of μ , if E is Borel in R_∞

$$(6.8) \quad P(X_n \in E | X_j, j \leq t < n) = P[(Z_n, Z_{n-1}, \dots) \in E | (Z_j, Z_{j-1}, \dots), j \leq t < n]$$

can be defined a.e. on R_∞ as a probability function, i.e., bounded between 0 and 1. Then (6.8) is immediately observed to be the same as

$$P(X_n \in E | X_t)$$

proving the Markov property. The stationarity of the transition functions follows from (6.8) and the strict stationarity of the process $\{Z_n\}$. Finally, for $E \subseteq R_\infty$, set

$$\pi(E) = \mu[(Z_n, Z_{n-1}, \dots) \in E].$$

π is well defined by the strict stationarity of $\{Z_n\}$. We have

$$\begin{aligned} \int P(z, E)\pi(dz) &= \int P[(Z_1, Z_0, \dots) \in E | (Z_0, Z_{-1}, \dots)]\mu(dw) \\ &= \mu[(Z_1, Z_0, \dots) \in E] = \pi(E), \end{aligned}$$

thereby proving the stationarity of π for the $\{X_n\}$ process. Since the Borel field of R_∞ is separable, the isomorphism theorem can be applied to $\{X_n\}$. Define the functional

$$f(X_n) = Z_n.$$

$E_\mu|f| < \infty$, so that the functional f^* defined at the beginning of this section satisfies the ergodic theorem for functionals of real processes as we have seen, and by (6.1) and (6.2), the ergodic theorem holds for the functional f , i.e.,

$$\lim_n \frac{1}{n} \sum_{k=1}^n Z_k$$

converges a.e. This is the ergodic theorem for μ nonatomic; the general case requires modifications in line with our former comments.

6. Extensions. Our aim now is to extend our results from $X_n(R, \Delta, \lambda)$ to the general process on separable Σ by means of the isomorphism Theorem 3.1. This is an easy task. Since, by definition, the isomorphism is a transformation defined only almost everywhere, roughly speaking we may say that results not depending upon local behavior (that is, assertions not made about a specific starting point x , but rather about almost all starting points) may be carried over immediately to the general process, whereas facts about local behavior at points may require a further argument. For example, the results of §4 may be carried over immediately to the general process: since there is always a k -process approximation such that $P_k \Rightarrow P$ on $X_n(R, \Delta, \lambda)$, substituting $Y_n(\Omega, \Sigma, \pi)$ (where π is infinite) for $X_n(R, \Delta, \lambda)$ and making the appropriate replacements in §4 leaves all results intact, because loss of sets of measure zero makes no difference. The important observation is that a local result (depending upon a conditional probability at x) needs special consideration. That this should be so is natural since π -null sets are not necessarily $P(x, \cdot)$ -null for every x .

REFERENCES

1. D. Blackwell, *Idempotent Markov chains*, Ann. of Math. **43** (1942), 560–567.
2. R. V. Chacon, *Identification of the limit of operator averages*, J. Math. Mech. **11** (1962), 961–968.
3. J. L. Doob, *Stochastic processes*, Wiley, New York, 1953.
4. N. Dunford and J. T. Schwartz, *Linear operators*, Part I, Interscience, New York, 1958.
5. W. Feller, *An introduction to probability theory and its applications*, Vol. 1, 2nd ed., Wiley, New York, 1957.
6. P. R. Halmos, *Measure theory*, Van Nostrand, New York, 1950.
7. P. Halmos and J. von Neumann, *Operator methods in classical mechanics*. II, Ann. of Math. **43** (1942), 332–350.
8. T. E. Harris, *On chains of infinite order*, Pacific J. Math. **5** (1955), 707–724.
9. T. E. Harris and H. Robbins, *Ergodic theory of Markov chains admitting an infinite invariant measure*, Proc. Nat. Acad. Sci. U. S. A. **39** (1953), 860–864.
10. F. Hausdorff, *Set theory*, Chelsea, New York, 1957.
11. R. Isaac, *On regular functions for certain Markov processes*, Proc. Amer. Math. Soc. **17** (1966), 1308–1313.
12. T. E. Harris, *The existence of stationary measures for certain Markov processes*, Proc. Third Berkeley Sympos. Math. Stat. and Prob., 1954–1955, Vol. 2, Univ. of Calif. Press, Berkeley, 1956, pp. 113–124.

13. N. C. Jain, *Some limit theorems for a general Markov process*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **6** (1966), 206–223.
14. J. L. Kelley, *General topology*, Van Nostrand, New York, 1955.
15. J. Lamperti, *Probability*, Benjamin, New York, 1966.
16. Yu. V. Prokhorov, *Convergence of random processes and limit theorems in probability theory*, Theor. Probability Appl. **1** (1956), 157–214.
17. K. L. Chung, *Markov chains with stationary transition probabilities*, Springer-Verlag, Berlin, 1960.

GRADUATE CENTER, THE CITY UNIVERSITY OF NEW YORK,
NEW YORK, NEW YORK
HERBERT H. LEHMAN COLLEGE,
BRONX, NEW YORK