The Annals of Probability
1974, Vol. 2, No. 1, 51-75

NOTES ON CONSTRUCTIVE PROBABILITY THEORY[1]

BY Y. K. CHAN

University of Washington

This paper is part of the constructive program, initiated by E. Bishop, of systematic examination of classical mathematics for their computational content. From this constructive standpoint, basic concepts in probability theory are studied. Positive proofs are then given to some important theorems: Ionescu-Tulcea's theorem, a submartingale convergence theorem, and the construction of a Markov process from a strongly continuous semi-group of transition operators.

0. Introduction. While some mathematicians refuse to talk about meaning in mathematics and others regard it as a game of symbols, it is the constructivists' contention that the meaning of mathematical theorems is to be found in their numerical interpretation. E. Bishop [1] initiated a program of systematic examination of existing mathematical theorems and theories with this point of view, giving them (possibly more than one) numerical interpretations and finding out whether they are still valid under these interpretations. To illustrate, consider the theorem which asserts that every bounded non-decreasing sequence a_n of real numbers has a limit. Understood constructively, this theorem could mean that we are able to find, for every given $\varepsilon > 0$, an integer m such that $a_p - a_q < \varepsilon$ for all $p \geqq q \geqq m$. Interpreted this way, the theorem is not valid, as seen from the following counterexample in the style of Brouwer. Let $a_n = 0$ if for all nonnegative integers a, b, c, and d such that $d > 2$ and $a + b + c + d \leqq n$ we have $a^d + b^d \neq c^d$; let $a_n = 1$ otherwise. The sequence a_n is then non-decreasing and bounded by 1. Now let $\varepsilon = \frac{1}{2}$ and the reader is invited to try to find m. By no means is one interested in such pathological "fugitive" sequences. They serve however to convince us that it is hopeless to find a constructive proof for certain statements.

The question whether constructivitists prove new theorems is often asked. The answer is yes, because if a proof of a theorem is given by showing that the assertion can be deduced from certain axioms according to certain rules, and if a second proof is given which actually enables us to do certain computation, then we all agree that the second proof constitutes a new theorem. One can put it this way: a theorem = the assertion of the theorem + what is meant by the assertion. When a theorem is proved constructively, it has a new meaning.

Received June 28, 1972; revised June 25, 1973.

[1] Research supported in part by NSF Grant GP-29233.

AMS 1970 *subject classifications*. Primary 60J25, 60G45; Secondary 02E99.

Key words and phrases. Constructive analysis, construction of Markov processes, sample regularity of Markov processes with strongly continuous semi-groups, convergence of submartingales.

An exposition of the constructive view point and methods is in [1], where a major portion of analysis is constructivized.

The present part of the constructive program intends to (i) develop sufficient tools in probability theory and (ii) carry out some basic construction to prove some fundamental theorems in probability theory. The success in (i) should be measured by the success in (ii).

Section 1 is a summary of the measure theory developed by E. Bishop and H. Cheng [2]. One superficial difference of this theory from the standard ones is that a measure is defined via an integration. The reason for this difference is technical: there often are functions which should naturally be integrable, while no such thing can be said about subsets. A specific example: the space of bounded continuous functions on R, though too small for measure theoretic studies, is a natural start, while the collection of all intervals is not—an interval (a, b) is measurable with respect to the point mass δ_x concentrated at x only when we can prove $(x \leqq a)$ or $(a < x < b)$ or $(b \leqq x)$. In short, if we want to start with a family of subsets, the choice of that family depends on the measure or measures we have in mind. This is a nuisance. A more fundamental difference—and a natural one in view of our insistence on computational meaning—is that the Monotone Convergence Theorem as stated classically is not valid. Indeed, given a non-decreasing sequence of integrable functions X_n with bounded integrals $I(X_n)$, how can we say $\lim X_n$ is integrable if we can not even compute its integral (as would be possible classically by taking the limit of $I(X_n)$)? We need to assume that $\lim I(X_n)$ exists. In terms of events, this means that in order to generate an event by taking the union of a sequence of given ones A_n, we must first prove that $\lim P(A_1 \vee \cdots \vee A_n)$ exists. We hope to convince the reader that with such restriction, we still have enough events to study probability theory, and that the harder work in general is rewarded by clear, positive results.

In Section 2, we deviate from [2] and define measurable functions in a way more compatible with desirable properties of random variables. Section 3 deals with measurable spaces. As noted earlier, the measurable functions are too closely tied to the measures relative to which they are generated. So measurable spaces are defined in terms of measure spaces. Theorem 3.2 is suggested by J. Nuber. In Section 4, a transition function is defined. The main Theorem 4.2 will be used time and again in Section 5. We then prove Ionescu-Tulcea's theorem and use it to formulate the construction of a discrete parametered Markov process.

Section 5 contains the main result of this paper. We construct a time homogeneous Markov process with a compact metric space S as state space from a given strongly continuous semi-group of transition operators. Since the constructive counterpart of Doob's notion of separability and that of Kolmogorov's theorem [4] do not apply in the present situation where S has no linear structure, we rely heavily on a regularity of the sample functions in order to extend them from a dense subset of the parameter set $[0, \infty)$ to $[0, \infty)$. This regularity property corresponds classically to the lack of infinitely many oscillations. Again the computational content is emphasized (5.5). The assumption that S is compact is convenient rather than necessary. The results can be generalized to the case when S is locally compact.

Section 6 is independent of the rest of the paper. It extends a constructive martingale convergence theorem in [1] to submartingales.

1. Measure spaces. The present section is a summary of the constructive measure theory developed by Bishop and Cheng [2].

Throughout this paper, Ω will denote a basic set, and $F(\Omega)$ the family of functions defined partially on Ω. A function of (finitely or infinitely many) functions in $F(\Omega)$ is defined in the natural way. For instance, if X and Y are in $F(\Omega)$, then $X + Y$ is the function in $F(\Omega)$ whose domain is the (possibly empty) intersection of those of X and Y, and whose value on this domain is the sum of X and Y. A convention: we only write the sum of a series $\sum_1^\infty a_n$ of real numbers along with the implicit assumption that this series converges absolutely.

Suppose, now, that L is a linear subspace of $F(\Omega)$ which is closed relative to the operations of absolute value and truncation (i.e. if X belongs to L then so do $|X|$ and $X \wedge 1$). We may call functions in L primary integrable functions, to be assigned an integral in the following sense. A linear function I on L is called an *integration* (or a *measure*) if it satisfies the following continuity conditions.

(1.1) If $(X_n)_{n=0}^\infty$ is a sequence of nonnegative functions in L such that $\sum_{n=1}^\infty I(X_n) < I(X_0)$, then there exists an ω at which X_n is defined for every n, such that $\sum_{n=1}^\infty X_n(\omega) < X_0(\omega)$.

(1.2) For each X in L, $\lim_{n\to\infty} I(X \wedge n) = I(X)$ and $\lim_{n\to\infty} I(|X| \wedge n^{-1}) = 0$.

To rule out trivial cases, we also assume the existence of some $X \in L$ such that $I(X) > 0$. The triple (Ω, L, I) will be called a *measure space*. The family L of primary integrable functions is usually too small for measure theoretic studies. We can extend it by taking sums of series in L. To be precise, an *integrable function* X is an ordered pair $(X, (X_n)_{n=1}^\infty)$ where $X \in F(\Omega)$ and (X_n) is a sequence in L, such that $\sum I|X_n|$ converges and such that $X(\omega)$ is defined and equal to $\sum X_n(\omega)$ whenever the latter sum is defined—recall our convention for sums of series of real numbers. The *integral* of X is then defined by $I(X) = \sum I(X_n)$. When no confusion is likely, we write X for the integrable function. We say that the integrable function X is *represented by* the sequence (X_n). Note that a sequence (X_n) in L with $\sum I(X_n)$ defined (i.e. convergent absolutely) always represents an integrable function. The family of integrable functions will be denoted by $L(I)$. It turns out that $(\Omega, L(I), I)$ is again a measure space, and no further such extension is possible because $L(I)(I) = L(I)$.

As a consequence of (1.1), the domain of an integrable function must contain some point. Actually such a domain D is necessarily big in the sense that two integrable functions X and Y obeying $X \leqq Y$ on D must also satisfy $I(X) \leqq I(Y)$.

Furthermore, if $X \in F(\Omega)$ is equal to some integrable function on D, then X is also integrable. Hence, a subset of Ω is called a *full set* if it contains such a domain. Hereafter equality in $L(I)$ means equality on a full set; similarly for inequality. We see from the above remark that I is a function on $L(I)$ with this new equality relation. Note that the intersection of a sequence of full sets remains a full set.

A *complemented set* is a pair $A = (A^1, A^0)$ of subsets such that two points in different components of A are unequal. Conceptually we can identify a complemented set A with its *indicator*, the function χ_A which equals 1 on A^1 and 0 on A^0. Thus we could talk solely in terms of functions. For instance, the intersection (resp. union) of a sequence (A_n) of complemented sets is the complemented set $\bigwedge_n A_n$(resp. $\bigvee_n A_n$) whose indicator is $\bigwedge_n \chi_{A_n}$(resp. $\bigvee_n \chi_{A_n}$). Likewise $-A$ is the complemented set with indicator $1 - \chi_A$. If X is a member of $F(\Omega)$, we will write $\{\omega : X(\omega) \leqq a\}$, or simply $(X \leqq a)$ for the complemented set $(\{\omega : X(\omega) \leqq a\}, \{\omega : X(\omega) > a\})$. In general we will let the first component stand for a complemented set if it is clear what the second component is. If ω belongs to A^1, we write $\omega \in A$, and say ω belongs to A. Roughly speaking a complemented set is a subset of Ω with a built-in complement, so that the set operation of taking complements has a positive meaning.

Naturally, then, a complemented set A is called an *integrable set* if χ_A is integrable. In symbols $A \in L(I)$. The integral $I(A) = I(\chi_A)$ is then called the measure of A. Integrable sets abound. If X is an integrable function, then $(X \geqq a)$ and $(X > a)$ are integrable (and have the same measure) for all but countably many $a \in (0, \infty)$. From now on, we write $(X \geqq a)$ or $(X > a)$ only along with the implicit assumption that a has been so chosen as to make the set in question integrable. As expected, if A is integrable, then $A^1 \cup A^0$ is a full set. If moreover A has measure 0, then A^0 is a full set. The relation $A < B$ will stand for $\chi_A \leqq \chi_B$ for integrable sets A and B. If X is an integrable function and A an integrable set, then $X\chi_A$ is also integrable and we will write $I(X; A)$ or $I(A; X)$ for $I(X\chi_A)$.

The product of two measure spaces (Ω_i, L_i, I_i), $(i = 1, 2)$ is defined as usual. The primary integrable functions on $\Omega_1 \times \Omega_2$ are linear combinations of indicators of the form $X_1 \otimes X_2$ where X_ι is an integrable indicator in $L_\iota(I_i)$. Fubini's theorem is again valid.

In case the constant function 1 on a measure space (Ω, L, I) is integrable with integral 1, we call (Ω, L, I) a *probability space*. An important example of measure spaces is a nonnegative linear function on the space L of continuous functions with compact support in a locally compact metric space (Ω, d). If there is a bounded sequence X_n in L converging uniformly on compact sets to 1 such that $I(X_n) \to 1$, then (Ω, L, I) is a probability space.

2. Measurable functions. Let (Ω, L, I) be a measure space, A an integrable set, $X, X_n (n \geqq 1)$ elements of $F(\Omega)$. Then we way X_n *converges in measure on*

A to X if for all $\varepsilon > 0$ there exists a positive integer N such that for every $n \geqq N$ we have $|X_n - X|\chi_B < \varepsilon$ for some integrable set $B < A$ with $I(A - B) < \varepsilon$. (In symbols $X_n \rightarrow_m X$ on A). If B is independent of n, then we say X_n *converges almost everywhere on* A to X. (In symbols $X_n \rightarrow_{\text{a.e.}} X$ on A). Note that a.e. convergence is stronger than convergence on a full subset of A, the latter being a useless notion in the constructive theory. If $X_n \rightarrow_m X$ on A for every integrable set A we simply say X_n converges in measure to X. Similarly for a.e. convergence.

We now deviate from [2] and define a *measurable function* as a function X which is defined on a full set such that for every integrable set A there exists a sequence (X_n) of integrable functions converging to X in measure on A. Thus every integrable function is measurable. By a 3ε-argument it can be proved that limits in measure of measurable functions are also measurable.

PROPOSITION 2.1. *If $X_1, \cdots, X_k$ are measurable and f is a continuous function on R^k, then $f(X_1, \cdots, X_k)$ is measurable.*

PROOF FOR $k = 1$. Let X be a measurable function. Let A be any integrable set and let ε be a number in $(0, 1)$. First choose $Y \in L(I)$ such that $|Y - X|\chi_B < \varepsilon$ for some $B < A$ with $I(A - B) < \varepsilon$. Choose $a > 0$ so large that $I(|Y| \geqq a-1) < \varepsilon$. Write $C \equiv B \wedge (|Y| < a - 1)$. Clearly $C < A$, $I(A - C) < 2\varepsilon$ and $|X|\chi_C \leqq a$. Now subdivide $[-a - 1, a + 1]$ by numbers $-a - 1 = a_0 < a_1 < \cdots < a_n = a + 1$ such that $|f(r) - f(s)| < \varepsilon$ if $|r| \leqq a + 1$ and $|r - s| < \bigvee_1^n (a_i - a_{i-1})$. Next choose $Z \in L(I)$ such that $|X - Z|\chi_D < \bigvee_1^n (a_i - a_{i-1})$ for some integrable set D with $I(A - D) < \varepsilon$. Define

$$U \equiv \sum_1^n f(a_i)\chi_{(a_{i-1} \leq Z < a_i) \wedge C \wedge D} \in L(I) \, .$$

Then $|f(X) - U|\chi_{C \wedge D} < 2\varepsilon$ and $I(A - C \wedge D) < 3\varepsilon$. Since ε is arbitrarily small, $f(X)$ is measurable. □

As a corollary, sums, products, maxima, and minima of several measurable functions are measurable.

PROPOSITION 2.2. *If X is measurable, Y integrable, and $|X| \leqq Y$, then X is also integrable.*

PROOF. Let $a_n \downarrow 0$ be such that $A_n = (Y \geqq a_n)$ is integrable for every n. Then $\sum I(Y: A_{n+1} - A_n) < \infty$. Choose $X_n \in L(I)$ and $B_n < A_n$ such that

(a) $\sum I(A_n - B_n) < \infty$,
(b) $\sum I(Y; A_n - B_n) < \infty$,
(c) $|X - X_n|\chi_{B_n} < \varepsilon_n$ where $\Sigma \varepsilon_n I(A_n) < \infty$.

There is no loss of generality in assuming $|X_n| \leqq Y$. We will prove that $(X_1\chi_{B_1},$

$X_2\chi_{B_2} - X_1\chi_{B_1}, \cdots)$ represents an integrable function equal to X. Estimate

$$\begin{aligned} I|X_{n+1}\chi_{B_{n+1}} - X_n\chi_{B_n}| &\leqq I|X_{n+1} - X_n|\chi_{B_{n+1}\wedge B_n} + I|X_{n+1}|\chi_{B_{n+1}-B_n} + I|X_n|\chi_{B_n-B_{n+1}} \\ &\leqq (\varepsilon_{n+1} + \varepsilon_n)I(A_{n+1} \wedge A_n) + I(Y; A_{n+1} - B_n) + I(Y; A_{n+1} - B_{n+1}) \\ &\leqq \varepsilon_{n+1}I(A_{n+1}) + \varepsilon_n I(A_n) + I(Y; A_{n+1} - A_n) + I(Y; A_n - B_n) \\ &\qquad + I(Y; A_{n+1} - B_{n+1}) . \end{aligned}$$

Hence, in view of (a), (b), and (c), the series $\sum I|X_{n+1}\chi_{B_{n+1}} - X_n\chi_{B_n}|$ converges and so $(X_1\chi_{B_1}, X_2\chi_{B_2} - X_1\chi_{B_1}, \cdots)$ represents some integrable function Z. On a full set we have $|Z| = |\lim X_n\chi_{B_n}| \leqq Y$ because $|X_n| \leqq Y$. On the other hand, because of (a), the sequence $(\chi_{A_n} - \chi_{B_n})$ represents an integrable function also. In particular $\chi_{A_n} - \chi_{B_n} \to 0$ on a full set. We will now show that for every ω in the full set where $\chi_{A_n} - \chi_{B_n} \to 0$, $|Z| \leqq Y$, and $|X| \leqq Y$, we must have $X(\omega) = Z(\omega)$, whence follows the integrability of X. Suppose for such an ω that $X(\omega) \neq Z(\omega)$. Then there is $m \geqq 1$ such that $|X(\omega) - Z(\omega)| > 2a_m$. In particular $Y(\omega) \geqq a_m$, i.e. $\chi_{A_n}(\omega) = 1$ for all $n \geqq m$ and so $\chi_{B_n}(\omega) \to 1$. But $X_n\chi_{B_n}(\omega) \to Z(\omega)$ and $|X - X_n|\chi_{B_n}(\omega) \to 0$. Hence $X(\omega) = Z(\omega)$, a contradiction. □

As a consequence, $a \wedge (-a \vee X)\chi_A$ is integrable if A is an integrable set and a some positive real number, i.e. a measurable function X is also measurable in the sense of [2]. In particular, the Monotone Convergence Theorem and the Dominated Convergence Theorem proved in [2] remain valid. Since only the former is different from the classical version we state it here.

THEOREM 2.3 (Monotone Convergence). *Let (X_n) be a non-decreasing sequence of integrable functions. If $c = \lim I(X_n)$ exists, then X_n converges* a.e. *to some integrable function X with $I(X) = c$. Conversely, if X_n converges in measure to an integrable function X, then* $\lim I(X_n) = I(X)$.

A moment's reflection would make it apparent why the boundedness of the sequence $I(X_n)$ would not suffice, if one insists as we do on knowing how fast X_n converges to X (see the definition of a.e. convergence).

Corresponding to the concepts of convergence in measure and convergence almost everywhere are the respective notions of *Cauchyness*. The completeness of these two kinds of convergence is proved in [1] under a different set-up, but the reader will have no trouble adapting the proofs to the present situation. It should be remarked again that convergence on a full set (or even on the whole space Ω) does not imply convergence almost everywhere, the latter being a much more useful concept.

In case our measure space (Ω, L, I) turns out to be a probability space, i.e. if the constant function 1 is integrable with $I(1) = 1$, we call integrals *expectations*, measurable functions *random variables*, and will replace the word measure by the word probability whenever it appears. An integrable subset A of a probability space will be called an *event*. For a probability space we can prove the following criterion for measurability.

PROPOSITION 2.4. *A function X on a probability space (Ω, L, E) is a random variable* iff $(X \leqq a) \in L(E)$ *for all but countably many a and* $E(-a < X \leqq a) \uparrow 1$ *as* $a \uparrow \infty$.

3. Measurable spaces. A measurable space is a vehicle to carry various measures. An attempt to define measurable spaces, parallel to the classical development, would therefore be taking a set Ω and some family of functions on Ω which deserve to be called measurable functions. Later on, measures would be defined on this couple. However, there is in general not a natural family to play this role. For example, the family of continuous functions on the real line is too small for measure-theoretic studies, and to enlarge this family the only way is via convergence in measure—with respect to some measure. In short, the measurable functions are too closely related to the measure with respect to which they are generated to be studied alone. Accordingly, we make the following definition.

DEFINITION 3.1. A *measurable space* is a triple $(\Omega, L, \mathbf{E})$ where L is a family of functions on the set Ω, and $\mathbf{E}$ is a family of functions on L such that for every E in $\mathbf{E}$, the triple (Ω, L, E) is a probability measure space. For each E in $\mathbf{E}$ let $F(L, E)$ denote the family of measurable functions on (Ω, L, E). The intersection $\bigcap_{E \in \mathbf{E}} F(L, E)$ will be called the family of *measurable functions on the measurable space* $(\Omega, L, \mathbf{E})$, and will be denoted by $F(L, \mathbf{E})$.

A complemented set in Ω is said to be *measurable* if its indicator is a measurable function. A subset of Ω is called an $\mathbf{E}$-*full set* if it is an E-full set for every $E \in \mathbf{E}$. Two measurable functions are said to be equal if they coincide on some $(\mathbf{E})$-full set. We will let $L(\mathbf{E})$ denote the subset of $F(L, \mathbf{E})$ consisting of all functions which are integrable with respect to every E in $\mathbf{E}$. Clearly two members X and Y of $L(\mathbf{E})$ are equal iff $E|X - Y| = 0$ for all $E \in \mathbf{E}$.

We next introduce a substitute for Borel fields of subsets of Ω. A subfamily F' of $F(L, \mathbf{E})$ is called a *Borel family* if continuous functions of functions in F' remain in F', and if a function X belongs to F', provided that X is for every E in $\mathbf{E}$ the limit (in probability) of members in F'.

Let G be an arbitrary subfamily of $F(L, \mathbf{E})$. Let L' be the family of bounded continuous functions of (one or more) members of G. Then $(\Omega, L', \mathbf{E})$ is a measurable space. It can be shown that $F(L', \mathbf{E})$ is a Borel subfamily of $F(L, \mathbf{E})$ and that every Borel subfamily of $F(L, \mathbf{E})$ is generated in this way.

Let $X_1, \cdots, X_k$ be measurable functions on $(\Omega, L, \mathbf{E})$. They induce a measurable structure on (R^k, C_B) where C_B is the space of bounded continuous functions on R^k. To be precise, define $E' \equiv E_{X_1 \cdots, X_k}$ on C_B by

$$E'(f) \equiv E(f(X_1, \cdots, X_k)) \qquad \text{for} \quad f \in C_B, E \in \mathbf{E}.$$

Then $(R^k, C_B, \mathbf{E}')$, where $\mathbf{E}' \equiv \{E' : E \in \mathbf{E}\}$, is a measurable space. That E' is a probability measure follows from the observation that except on an arbitrarily-small-measured set, $X_1, \cdots, X_k$ are bounded (see 2.4) and so, as f converges boundedly to 1 uniformly on compact sets, $E'(f) \to 1$. The defining equation for E' can at once be generalized, by a continuity argument, to integrable functions. To be precise, if f is integrable on (R^k, C_B, E') then $f(X_1, \cdots, X_k)$ is integrable on (Ω, L, E) and $E'(f) = E(f(X_1, \cdots, X_k))$. A similar situation holds for measurable functions.

THEOREM 3.2. *For every measurable function f on the measurable space $(R^k, C_B, \mathbf{E})$, the function $f(X_1, \cdots, X_k)$ is measurable on $(\Omega, L, \mathbf{E})$.*

PROOF. For every positive integer n, the function $n \wedge (-n \vee f(X_1, \cdots, X_k))$ is equal to $n \wedge (-n \vee f)(X_1, \cdots, X_k)$ and so belongs to $L(E)$. Hence the set $(f(X_1, \cdots, X_k) \leqq a) \in L(E)$ for all but countably many a in R. Furthermore, as $a \to \infty$

$$E(-a < f(X_1, \cdots, X_k) \leqq a) = E'(-a < f \leqq a) \to 1 .$$

Hence 2.4 implies the measurability of $f(X_1, \cdots, X_k)$. □

This theorem will be used repeatedly in Section 5.

4. Transition functions. Consider a measurable space $(\Omega, L, \mathbf{E})$ and a member X of L. If $\mathbf{E}$ is countable, then it follows from the remark after Proposition 1.9 that for all but countably many a in R, the set $(X \geqq a)$ is integrable with respect to all E in $\mathbf{E}$. In general, if $\mathbf{E}$ is parametrized by a probability measure space, the situation will usually be such that for all but countably many a in R, the set $(X \geqq a)$ is integrable with respect to almost every E in $\mathbf{E}$. (This is the best that can be hoped for, as seen from the following example. Let L be the family of all continuous functions on R, and let $\mathbf{E}$ be the family of point masses δ_b with b ranging through R. Then the function $X(t) = t$ is a member of L. But given any a in R, the set $(X \geqq a)$ is integrable only with respect to those δ_b for which $a = b$ or $a \neq b$.) This will give us sufficiently many integrable sets for the construction and study of homogeneous Markov processes in a later section. The setting is, of course, furnished by transition functions.

DEFINITION 4.1. Let $(\Omega_i, L_i, \mathbf{E}_i)$ $(i = 1, 2)$ be measurable spaces. A linear mapping $E: L_2 \to L_1$ is said to be a *probability transition function* from $(\Omega_1, L_1, \mathbf{E}_1)$ to $(\Omega_2, L_2, \mathbf{E}_2)$ if the following two conditions are satisfied.

(i) The composite mapping $E_1 E$ is in $\mathbf{E}_2$ for every E_1 in $\mathbf{E}_1$.

(ii) For each ω in Ω_1, if we let L_2^ω denote the family of those functions X in L_2 for which $E^\omega(X) \equiv E(X)(\omega)$ is defined, then the triple $(\Omega_2, L_2^\omega, E^\omega)$ is a probability space.

Every ω in Ω_1 can be looked upon as a piece of information which leads us to select the probability E^ω on Ω_2. Classically no reference is made to $\mathbf{E}_1$ and $\mathbf{E}_2$; or equivalently, they are taken to be all possible probability measures. For us, however, the pair (Ω_i, L_i) is usually not sufficient to describe our measurable space.

We can extend E to a function on $F(\Omega_2)$ in the following way. For $X \in F(\Omega_2)$ let $E(X)$ be the function in $F(\Omega_1)$ whose domain consists of all $\omega \in \Omega_1$ such that $X \in L_2{}^{\omega}(E^{\omega})$, and which has value $E(X)(\omega) \equiv E^{\omega}(X)$ at such an ω.

THEOREM 4.2. *Suppose E is a transition function from $(\Omega_1, L_1, \mathbf{E}_1)$ to $(\Omega_2, L_2, \mathbf{E}_2)$. Then*

(a) *extended as above, E is a transition function from $(\Omega_1, L_1(\mathbf{E}_1), \mathbf{E}_1)$ to $(\Omega_2, L_2(\mathbf{E}_2), \mathbf{E}_2)$,*

(b) *for every measurable function X on $(\Omega_2, L_2, E_1 E)$ where E_1 is a given member of $\mathbf{E}_1$, for all but countably many $a \in R$, the sets $(X \geqq a)$ and $(X > a)$ are E^{ω}-integrable for all ω in some E_1-full set, and $E^{\omega}(X \geqq a) = E^{\omega}(X > a)$ for such ω's.*

PROOF. Take X in $L_2(\mathbf{E}_2)$. Let E_1 be an arbitrary member of $\mathbf{E}_1$. By Condition (i) of Definition 4.1 we have $E_1 E = E_2$ for some E_2 in $\mathbf{E}_2$. Therefore there exists some sequence (X_n) in L_2 such that $(X, (X_n))$ belongs to $L_2(E_2)$. In particular $\sum E_1 E|X_n| = \sum E_2|X_n| < \infty$. Hence the set of ω's for which $\sum E^{\omega}|X_n|$ is defined and converges constitutes an E_1-full set. By definition $E(X)$ equals $\sum E(X_n)$ on this full set. On the other hand we have $\sum E_1|E(X_n)| \ll \sum E_1 E|X_n| < \infty$ because $|E(X_n)| \leqq E|X_n|$ in view of condition (ii) of Definition 4.1. Consequently $\sum E(X_n)$ is E_1-integrable. Therefore $E(X)$ is also E_1-integrable. Since E_1 was arbitrary, we have $E(X) \in L_1(\mathbf{E})$. Moreover, in the above notation, $E_1 E(X) = \sum E_1 E(X_n) = \sum E_2(X_n) = E_2(X)$ and so $E_1 E$ and E_2 coincide on $L_2(\mathbf{E}_2)$. Finally $L_2(E_2)^{\omega} = L_2{}^{\omega}(E^{\omega})$ for every ω in Ω_1, as is clear from the definitions of $E(X)$ and $L_2(E_2)^{\omega}$. Hence E^{ω} is a probability measure on $(\Omega_2, L_2(E_2)^{\omega})$. Assertion (a) is thus proved. Assertion (b) is a consequence of (a). (See Section 1.) □

Two transition functions E and E' from $(\Omega_1, L_1, \mathbf{E}_1)$ to $(\Omega_2, L_2, \mathbf{E}_2)$ are said to be equal if they are equal as functions on L_2. It should be emphasized that the equality relation on $L_i (i = 1, 2)$ is the equality with respect to $\mathbf{E}_i$. By a continuity argument, we can prove that if E and E' are equal transition functions, so are their extensions.

If $\mathbf{E}_1'$ is a family of functions on L_1, we will let $\mathbf{E}_1'E$ denote the family $\{E_1'E : E_1' \in \mathbf{E}_1'\}$.

It can easily be verified that composites of transition functions are again transition functions.

In the rest of this section suppose for every nonnegative integer $n \geqq 0$ we are given a set Ω_n and a subset L_n of $F(\Omega_n)$. Let Ω denote the product $\prod_{n=0}^{\infty} \Omega_n$. We also regard L_n as a subset of $F(\Omega)$ by regarding a function X_n in L_n as a function on Ω whose domain consists of those $\omega \in \Omega$ for which ω_n belongs to the domain of X_n, and whose value at such an ω is $X_n(\omega_n)$. For arbitrary integers $p \leqq n$, define subsets of $F(\Omega)$:

$$M_{p,n} \equiv \{\textstyle\sum_{k=1}^{m} a_k X_p^k \cdots X_n^k : m \in N, a_k \in R, X_j^k \text{ some indicator in } L_j\}$$

$$M_n \equiv M_{0,n}$$

$$M_p' \equiv \textstyle\bigcup_{n \geqq p} M_{p,n}$$

$$M \equiv M_0' \equiv \textstyle\bigcup_{n \geqq 0} M_{0,n} = \bigcup_{n \geqq 0} M_n$$

Let $\Omega_p' \equiv \prod_{n \geqq p} \Omega_n$.

THEOREM 4.3. *Suppose for every n we are given*

(a) *a set* Ω_n, *a subset* L_n *of* $F(\Omega_n)$, *a family* $\mathbf{E}_n$ *of functions on* M_n *such that* $(\Omega, M_n, \mathbf{E}_n)$ *is a measurable space,*

(b) *a transition function* E_n *from* $(\Omega, M_{n-1}(\mathbf{E}_{n-1}), \mathbf{E}_{n-1})$ *to* $(\Omega, M_n(\mathbf{E}_n), \mathbf{E}_n)$ *such that*

$$E_n(XY) = XE_n(Y)$$

for all $X \in M_{n-1}$ *and* $Y \in M_n(\mathbf{E}_n)$.
Then the function E defined on M by

$$E(X) = E_1 \cdots E_n(X) \qquad \text{if} \quad X \in M_n$$

is a transition function from $(\Omega, M_0(\mathbf{E}_0), \mathbf{E}_0)$ *to* $(\Omega, M, \mathbf{E}_0 E)$.

For the proof of this theorem we need a lemma whose easy proof is omitted.

LEMMA. *Let X, Y, Z be nonnegative integrable functions on a probability space* (Ω, L, E). *Suppose* $X \leqq 1$, *and for some positive numbers* $\varepsilon_0 > \varepsilon_1$ *we have* $E(Y) < E(X) - 2\varepsilon_0$ *and* $E(Z) < \varepsilon_1^2$. *Then there exists an event A with positive measure such that* $Y(\omega) < X(\omega) - \varepsilon_0$ *and* $Z(\omega) < \varepsilon_1$ *for all* $\omega \in A$.

PROOF OF THEOREM 4.3. We may assume that the functions in L_n are bounded.

(i) To start, E is a well-defined function. For if $X \in M_{n-1} \cap M_n$, then

$$E_0 \cdots E_n(X) = E_0 \cdots E_{n-1}(X)$$

in view of condition (b).

(ii) Suppose $m < n$ and $X_j \in L_j (0 \leqq j \leqq n)$ are given. Then again by condition (b)

$$E_{m+1} \cdots E_n(X_0 \cdots X_n) = X_0 \cdots X_m E_{m+1} \cdots E_n(X_{m+1} \cdots X_n) .$$

Therefore, for α in an $\mathbf{E}_m$-full set, we have

$$E_\alpha E_{m+1} \cdots E_n(X_0 \cdots X_n) = E_\alpha E_{m+1} \cdots E_n(X_0(\alpha_0) \cdots X_m(\alpha_m) X_{m+1} \cdots X_n)$$

where E_α is the point mass concentrated at α. By linearity, for every $X \in M_n$, there exists an $\mathbf{E}_m$-full set of α's such that

$$E_\alpha E_{m+1} \cdots E_n(X) = E_\alpha E_{m+1} \cdots E_n(X \mid \alpha_0, \cdots, \alpha_m)$$

where $X \mid \alpha_0, \cdots, \alpha_m$ is the member of M_n defined by

$$(X \mid \alpha_0, \cdots, \alpha_m)(\omega) = X(\alpha_0, \cdots, \alpha_n, \omega_{m+1}, \omega_{m+2}, \cdots) .$$

(iii) We next prove that $(\Omega, M, \mathbf{E}_0 E)$ is indeed a measurable space, i.e. given $E_0 \in \mathbf{E}_0$, the function $E_0 E$ is a probability measure on (Ω, M). Only condition (1.1) in Section 1 needs be verified, the other conditions being trivial. Thus let $X_0, X_1, \cdots$ be a sequence of nonnegative functions in M such that $\sum_1^\infty E_0 E(X_j)$ converges and is less than $E_0 E(X_0) - 2\varepsilon_0$ for some $\varepsilon_0 > 0$. We may assume that $X_j \in M_{n(j)}$ with $n(j) < n(j+1)$ for $j \geqq 0$, and that X_0 and $3\varepsilon_0$ are bounded by 1. Choose a decreasing sequence ε_i of positive numbers such that $2 \sum_1^\infty \varepsilon_i < \varepsilon_0$. Then for some integer p large enough, we have

$$\sum_{1 \le j \le p} E_0 E_1 \cdots E_{n(j)}(X_j) < E_0 \cdots E_{n(0)}(X_0) - 2\varepsilon_0$$

and

$$\sum_{p<j} E_0 E_1 \cdots E_{n(j)}(X_j) < \varepsilon_1^2 .$$

Since $n(j) < n(j+1)$ for all j, the last two inequalities may be rewritten as

$$E_0 \cdots E_{n(p)}(\textstyle\sum_{1 \le j \le p} X_j) < E_0 \cdots E_{n(p)}(X_0) - 2\varepsilon_0$$

and

$$E_0 \cdots E_{n(p)}(\textstyle\sum_{p<j} E_{n(p)+1} \cdots E_{n(j)}(X_j)) < \varepsilon_1^2 .$$

The lemma, applied to the random variables X_0, $\sum_{1 \le j < p} X_j$, and $\sum_{p \le j} E_{n(p)+1} \cdots E_{n(j)}(X_j)$ on the probability space $(\Omega, M_{n(p)}, E_0 \cdots E_{n(p)})$, yields a positive measured subset A of $(\Omega, M_{n(p)}, E_0 \cdots E_{n(p)})$ such that for all $\alpha \in A$,

$$\sum_{1 \le j \le p} X_j(\alpha) < X_0(\alpha) - \varepsilon_0$$

and

$$\sum_{p<j} E_\alpha E_{n(p)+1} \cdots E_{n(j)}(X_j) < \varepsilon_1 .$$

Since $E_0 \cdots E_{n(p)}$ is a member of $\mathbf{E}_{n(p)}$, by (ii) we can choose $\alpha \in A$ such that

$$\sum_{1 \le j \le p} X_j(\alpha) < X_0(\alpha) - \varepsilon_0$$

and

$$\sum_{p<j} E_\alpha E_{n(p)+1} \cdots E_{n(j)}(X_j \mid \alpha_0', \cdots, \alpha_{n(p)}) < \varepsilon_1 = 3\varepsilon_1 - 2\varepsilon_1 .$$

The last displayed inequality has the same appearance as the first in this proof. Hence the argument above can be repeated to yield inductively integers $q < r < \cdots$ and elements $\beta, \gamma, \cdots$ of Ω such that

$$\sum_{p<j\le q} (X_j \mid \alpha_0, \cdots, \alpha_{n(p)})(\beta) < 3\varepsilon_1 - \varepsilon_1 = 2\varepsilon_1 ,$$
$$\sum_{q<j\le r} ((X_j \mid \alpha_0, \cdots, \alpha_{n(p)}) \mid \beta_0, \cdots, \beta_{n(q)})(\gamma) < 3\varepsilon_2 - \varepsilon_2 = 2\varepsilon_2 ,$$
$$\cdots .$$

Therefore, if we let $\omega = (\alpha_0, \cdots, \alpha_{n(p)}, \beta_{n(p)+1}, \cdots, \beta_{n(q)}, \gamma_{n(q)+1}, \cdots, \gamma_{n(r)}, \cdots)$, the above inequalities become

$$\sum_{1\le j\le p} X_j(\omega) < X_0(\omega) - \varepsilon_0 ,$$
$$\sum_{p<j\le q} X_j(\omega) < 2\varepsilon_1 ,$$
$$\sum_{q<j\le r} X_j(\omega) < 2\varepsilon_2 ,$$
$$\cdots$$

In short, we have constructed ω such that $\sum_{j=1}^{\infty} X_j(\omega)$ is well defined, convergent, and less than $X_0(\omega) - \varepsilon_0 + 2\sum_1^{\infty} \varepsilon_i < X_0(\omega)$. Condition (1.1) has thus been verified.

(iv) For arbitrary $\omega \in \Omega$, the argument in (iii) with E_0 replaced by E_ω shows that E_ω is a probability measure. Therefore condition 4.1 (ii) is satisfied by E. Condition 4.1 (i) being self evident, E is indeed a transition function. □

The above version of Ionescu–Tulcea's theorem is sufficient for the construction of a Markov process (discrete time) from its transition functions.

THEOREM 4.4. *Suppose for each n are given a measurable space* $(\Omega_n, L_n, \mathbf{I}_n)$ *and a transition function* D_n *from* $(\Omega_{n-1}, L_{n-1}(\mathbf{I}_{n-1}), \mathbf{I}_{n-1})$ *to* $(\Omega_n, L_n, \mathbf{I}_n)$ *such that* $\mathbf{I}_{n-1}D_n = \mathbf{I}_n$. *Then*

(i) *there exists a family* $\mathbf{E}$ *of probability integrations on* (Ω, M) *such that* $\mathbf{I}_n = \{E \mid L_n : E \in \mathbf{E}\}$,

(ii) *for every p there exists a transition function* E_p' *from* $(\Omega, L_p(\mathbf{E}), \mathbf{E})$ *to* $(\Omega, M_p'(\mathbf{E}), \mathbf{E})$ *such that*

$$E_p'(YZ) = YE_p'(Z)$$

for all bounded functions Y in $M_p(\mathbf{E})$ *and Z in* $M_p'(\mathbf{E})$.

The proof of this theorem is lengthy and omitted.

Let $(\Omega_i, L_i, \mathbf{E}_i)_{i=1,2}$ be measurable spaces, and let X be a function from Ω_1 to Ω_2. X is said to be a *measurable mapping*, or a random variable, on $(\Omega_1, L_1, \mathbf{E}_1)$ with values in $(\Omega_2, L_2, \mathbf{E}_2)$, if the conjugate map X^*, defined on L_2 by $X^*(f) = f \circ X$, constitutes a transition function from $(\Omega_1, L_1(\mathbf{E}_1), E_1)$ to $(\Omega_2, L_2, \mathbf{E}_2)$. The proof of the next proposition is routine and will not be given.

PROPOSITION 4.5. *If X is a measurable mapping from* $(\Omega_1, L_1, \mathbf{E}_1)$ *to* $(\Omega_2, L_2, \mathbf{E}_2)$, *then it is a measurable mapping from* $(\Omega_1, L_1(\mathbf{E}_1), \mathbf{E}_1)$ *to* $(\Omega_2, L_2(\mathbf{E}_2), \mathbf{E}_2)$.

In case $\mathbf{E}_2$ is the family of all probability measures on (Ω_2, L_2), we will simply say X is a random variable on $(\Omega_1, L_1, \mathbf{E}_1)$ with values in (Ω_2, L_2). Given a measurable mapping X as above, and a member E of E_1, the probability measure EX^* on (Ω_2, L_2) is called the *probability measure induced by X* relative to E, and is denoted by E_X.

5. Construction of a Markov process. Suppose we are given

(a) a compact metric space (S, d),

(b) a semi-group $(P_t)_{t\in[0,\infty)}$ of nonnegative linear operators on C, the space of continuous functions on (S, d) equipped with the supremum norm, such that $P_t(1) = 1$ and $P_0(f) = f$ for $f \in C$,

(c) a family $\mathbf{I}_0$ of probability measures on (S, C).

It is then well known classically that a time homogeneous Markov process can be constructed with parameter set $[0, \infty)$, state space S, transition functions P_t, and initial distributions $\mathbf{I}_0$. Indeed the construction is nothing more than an application of Kolmogorov's extension theorem. Now the constructive version of Kolmogorov's theorem [4] is proved only when S is some convex closed subset of the extended real line, and in any case some linear structure (in addition to the metric) on S is assumed. So it is not applicable here. Fortunately, the construction with a countable dense subset D of $[0, \infty)$ as parameter set is easy. It is also well known classically that under the condition

(d) for all $\varepsilon > 0$, there exists $\delta(\varepsilon) > 0$ such that $|P_t f(x)| < \varepsilon$ for all $t < \delta$, $x \in S$, and $f \in C$ with $||f|| \leqq 1$ and $f(y) = 0$ whenever $d(y, x) < \varepsilon$,

the process can be assumed to have sample functions which have no more than finitely many oscillations on every finite interval ([3], [6]), and so extendable from D to $[0, \infty)$; the process could then be extended correspondingly. The proof actually goes like this. Let $[\sigma, \sigma + \theta]$ be an interval where θ is some positive number depending only on the operation δ in (d). Let Δ be the complement of the set A of sample functions which have at most finitely many oscillations on $[\sigma, \sigma + \theta]$. It is proved that

$$\Delta < \bigwedge_{k=1}^{\infty} \bigvee_{n=1}^{\infty} \Delta_{k,n}$$

where $\Delta_{k,n}$ are measurable with

$$\Delta_{k,n} < \Delta_{k,n+1} \quad \text{and} \quad I(\Delta_{k,n}) \leqq 2^{-k}$$

for all k, n, and I corresponding to any initial distribution I_0. From this it is deduced that Δ is a null event. We maintain that the computational content of the above theorem is not as strong as should be. To be precise

(i) given an event B with $I(B) > 0$ it does not follow from the two formulas displayed above that a member of A can be constructed in B;

(ii) even if it could be proved that all the sample functions can be assumed to be members of A, it does not follow constructively from the lack of infinitely many oscillations that the left and right hand limits exist, and so the extension to $[0, \infty)$ is not valid.

The main task in our construction, technical questions aside, will therefore be proving a version of the above discussed regularity property of the sample functions as functions on D, with enough computational content so that the difficulties mentioned in (i) and (ii) do not arise.

Before proceeding we remark that condition (d) is equivalent to

(d') for all $f \in C$, we have $||P_t f - f|| \to 0$ as $t \to 0$.

Observe that, under condition (b), the condition (d') is hardly any restriction, because we have never seen a semi-group (P_t) where P_t is defined for *every*

$t \in [0, \infty)$ which is not strongly continuous in the sense of (d′). The same thing can therefore be said about (d).

Throughout this section assume that the objects in (a)–(d) are given. Let D denote the set of nonnegative dyadic numbers, with a fixed enumeration $(r_n)_{n=1}^{\infty}$. Let $\Omega = S^D$ be equipped with a (compact) product metric $\bar{d}(\omega, \omega') = \sum_{n=1}^{\infty} 2^{-n} d(\omega(r_n), \omega'(r_n))$. Let L denote the space of continuous functions on Ω, equipped with the supremum norm. Let L^*, C^* denote the families of all probability measures on L, C respectively. Define

$$X_t(\omega) = \omega(t) \qquad \text{for all} \quad t \in D, \omega \in \Omega\,,$$
$$L_s = \{g(X(t_1), \cdots, X(t_n)) : n \in N;\ g \ \text{ continuous on } \ S^n;\ t_1, \cdots, t_n \in D \cap [0, s]\}\,.$$

For convenience we sometimes write $X(t)$ for the function X_t and write $X(t, \omega)$ or $X(t)(\omega)$ for its value at ω. The easy construction of the Markov process on the parameter set D is carried out in the next theorem.

THEOREM 5.1. *There exists a transition function E from (S, C, C^*) to (Ω, L, L^*) such that for a given $I_0 \in C^*$ and $I \equiv I_0 E$ we have*

(i) *for every s in D, the probability measure $I_{X(s)}$ induced by $X(s)$ on (S, C) is the same as $I_0 P_s$,*

(ii) *for every $Z \in L(I_0 P_s E)$, if we let θ_s denote the shift map on Ω defined by $\theta_s(\omega)(t) = \omega(s + t)$ and let F_s denote $F(L_s, I)$, then $Z \circ \theta_s \in L(I)$ and*

$$I(Z \circ \theta_s \mid F_s) = E^{X(s)}(Z) \qquad \textit{on an I-full set.}$$

(Recall that the induced measure $I_{X(s)}$ is defined by $I_{X(s)}(f) \equiv I(f \circ X(s))$ for all $f \in C$. Recall also that by Theorem 4.2 the function $E(Z)$ belongs to $C(I_0 P_s)$. On the other hand, it follows from (i) that $X(s)$ is a measurable mapping from (Ω, L, I) to $(S, C, I_0 P_s)$. Therefore Proposition 4.5 applies and we see that the composite function $E^{X(s)}(Z)$ really belongs to $L(I)$. The intuitive meaning of assertion (ii) is as follows. $E^X(Z)$ represents the expectation of Z if our process starts at the initial position x. The Borel family F_s represents information about the process up to time s. So the displayed equality in (ii) means that the conditional expectation of a random variable $Z \circ \theta_s$ which depends only on the process after time s, if all information about the process up to time s is given, is as if the process had started at time s at the initial position of $X(s)$.)

PROOF. The linear closure L' of the subset of L consisting of functions of the form $f_1(X(t_1)) \cdots f_n(X(t_n))$ with $f_j \in C (j = 1, \cdots, n)$ is dense in L by Stone–Weierstrass' theorem. Define E on L' by

$$E(\textstyle\sum_i a_i f_1^i(X(t_1)) \cdots f_n^i(X(t_n))) = \sum_i a_i P_{t_1}(f_1^i P_{t_2 - t_1}(f_2^i \cdots P_{t_n - t_{n-1}}(f_n^i)))$$

for all $a_i \in R$, $f_j^i \in C$, and $t_1 < \cdots < t_n$ in D. To prove that E is well defined it suffices to show that if the sum in the left side of the above equality is the constant zero function, then so is the sum in the right hand side. But if the first sum is the constant function 0, then $\sum_i a_i f_1^i(x_1) \cdots f_n^i(x_n) = 0$ for all

$(x_1, \cdots, x_n) \in S^n$. In particular, when $x_1, \cdots, x_{n-1}$ are held fixed, this function as a function of x_n is the constant zero function. Consequently

$$\sum_i a_i f_1^i(x_1) \cdots f_{n-1}^i(x_{n-1}) P_{t_n - t_{n-1}}(f_n^i) = 0$$

for all $(x_1, \cdots, x_{n-1}) \in S^{n-1}$. Repeating this argument we see that the second sum in the defining formula for E is indeed the constant zero function if the first is. It is obvious that E is linear, nonnegative, and that $E(1) = 1$. Hence E can be extended by continuity to a similar function from L to C, which is *a fortiori* a transition function from (S, C, C^*) to (Ω, L, L^*). Let $I_0 \in C^*$ be given, and let $I = I_0 E$. Assertion (i) follows trivially from the defining formula of E. To prove (ii) take two arbitrary generating elements $Y' = f_1(X(t_1)) \cdots f_n(X(t_n))$ and $Z' = g_1(X(s_1)) \cdots g_k(X(s_k))$ with $t_1 \leqq \cdots \leqq t_n \leqq s$ for some $s \in D$ and $s_1 \leqq \cdots \leqq s_k$. Then

$$I(Y'(Z' \circ \theta_s)) = I(Y' E^{X(s)}(Z'))$$

as can be verified by a direct substitution. By linearity and continuity of E, the same equality therefore holds for all $Y' \in L_s$ and $Z' \in L$. The Markov property (ii) has thus been verified for all $Z' \in L$. In general take $(Z, (Z_n)) \in L(I_0 P_s E) = L(I_{X(s)})$. By Theorem 4.2 (or rather its proof) we have $(EZ, (EZ_n)) \in C(I_{X(s)})$; equivalently $(E^{X(s)}(Z), (E^{X(s)}(Z_n))) \in L(I)$. On the other hand $(Z \circ \theta_s, (Z_n \circ \theta_s)) \in L(I)$ because

$$\sum_n I|Z_n \circ \theta_s| = \sum_n I(|Z_n| \circ \theta_s) = \sum_n I E^{X(s)}|Z_n| = \sum_n I_0 P_s E|Z_n| < \infty$$

and the conditions (1.1) and (1.2) are fulfilled. Thus for $Y \in L_s$

$$I(Y(Z \circ \theta_s)) = \sum_n I(Y(Z_n \circ \theta_s)) = \sum_n I(Y E^{X(s)}(Z_n)) = I(Y E^{X(s)}(Z)) . \qquad \square$$

To extend the Markov process constructed on D to $[0, \infty)$ we need first to show that the sample functions $X(\cdot, \omega)$ are well behaved on D. To this end, the following scheme is followed. Fix $\varepsilon > 0$ and consider, heuristically, the first time T when $X(T)$ wanders out of the ε-neighborhood of $X(0)$. We note where $X(T)$ is, and observe the behavior after time T. The strong Markov property 5.2(b), if it were applicable to T, would say that this probabilistic behavior of X after time T would be as if the process had originally started at $X(T)$. So there is a first time T' after T when $X(T')$ wanders out of the ε-neighborhood of $X(T)$, and so on. The Markov property and the continuity condition (d) should then imply that, given any $s > 0$, for n large enough, the nth time $T^{(n)}$ should exceed s with a high probability. In particular, except a small subset of Ω, every sample function $X(\cdot, \omega)$ up to time s can be broken into n pieces each of diameter at most ε. This is the conclusion of Theorem 5.5.

There are two difficulties involved in the construction just described. First the function T might be constructively undefined, defined but not measurable, or measurable but with values not in D so that $X(T)$ does not make sense. Second, the strong Markov property might not a priori be applicable to T. To overcome these difficulties, we introduce a sequence of random times (T_k), where T_k is the first time the process X wanders out of the ε-neighborhood of $X(0)$ when X is sampled only on a finite subset D_k of D. On the one hand, since T_1 has only finitely many values the Markov property is easily applicable. On the other hand, it is hoped that as $D_k \uparrow D$ we have $T_k \downarrow T$ whence X differs on $D \cap [T_{k+1}, T_k]$ by at most some ε_k from $X(T_k)$, except perhaps on a subset of Ω with probability $< \varepsilon_k$. As a consequence X would differ on $D \cap (T, T_1]$ from $X(T_1)$ by at most $\sum \varepsilon_k$, except perhaps on a subset of Ω with probability at most $\sum \varepsilon_k$. This way we can sidestep the application of the Markov property to T by applying it to the T_k's, provided $\sum \varepsilon_k$ is very small. There is, however, no reason to expect $\sum \varepsilon_k$ to converge. It is even conceivable that the waiting time T_k for X to wander out of the fixed ε-neighborhood of $X(0)$, when X is sampled on D_k, to be much longer than T_{k+1}. For this reason we give up the fixed ε-neighborhood. Instead we let T_k be the first time in D_k when $X(T_k)$ is out of the $a(T_k)$-neighborhood of $X(0)$, where a is a function on D varying in $(\varepsilon, 2\varepsilon)$. How this will help is perhaps best explained by the proofs of Lemma 5.3 through Theorem 5.5, which make the above construction precise.

First we define a stopping time (with finitely many values) and state a lemma to the effect that the Markov property 5.1(ii) is applicable to such stopping time.

Let Ω, L, E, (X_s), (L_s), and (θ_s) be as constructed in the above theorem. Let $\mathbf{I}_0$ be a subset of C^* and let $\mathbf{I} = \mathbf{I}_0 E$. By Theorem 4.2 we can extend E to a transition function from $(S, C(\mathbf{I}_0), \mathbf{I}_0)$ to $(\Omega, L(\mathbf{I}), \mathbf{I})$. A measurable function T on $(\Omega, L(\mathbf{I}), \mathbf{I})$ with values in a finite subset D' of D will be called a *stopping time* for the Markov process $(\Omega, L, \mathbf{I}, E, (X_s), (\theta_s))$ relative to the initial distributions $\mathbf{I}_0$ if

$$(T = t) \in L_t(\mathbf{I}) \qquad \text{for every} \quad t \in D' .$$

For such a stopping time define

$$F(T, \mathbf{I}) \equiv \{X \in F(L, \mathbf{I}) : f(X)\chi_{(T=t)} \in F(L_t, \mathbf{I}) \text{ if } t \in D', f \in C\} ;$$
$$\theta_T(\omega) \equiv \theta_{T(\omega)}(\omega) ;$$
$$X_T(\omega) \equiv X_{T(\omega)}(\omega) .$$

Clearly $F(T, \mathbf{I})$ is a Borel subfamily of $F(L, \mathbf{I})$. Suppose T and T' are two stopping times as above. We will use without mention the easily proved fact that $T + T' \circ \theta_T$ is a stopping time relative to the same initial distributions $\mathbf{I}_0$.

LEMMA 5.2. *Let $I_0 \in C^*$ be given and let $I = I_0 E$. If T is a stopping time for the Markov process $(\Omega, L, \{I_0 P_u E\}_{u \in D}, E, (X_s), (\theta_s))$ and if $Z \in L(\{I_0 P_u E\}_{u \in D})$, then*

(a) $Z \circ \theta_T \in L(\{I_0 P_u E\}_{u \in D})$,
(b) $I(Z \circ \theta_T \mid F(T, I)) = E^{X(T)}(Z)$ *as measurable functions on* (Ω, L, I), *and*
(c) $E(Z \circ \theta_T) = E(E^{X(T)}(Z))$ *as measurable functions on* $(S, C, \{I_0 P_u\}_{u \in D})$.

PROOF. The proofs are straightforward. We give, for illustration, that of (a).

Since (P_s) is a semi-group, we have for every $u \in D$,

$$\{I_0 P_u P_s E\}_{s \in D} \subset \{I_0 P_s E\}_{s \in D} \quad \text{and so} \quad L(\{I_0 P_s E\}_{s \in D}) \subset L(\{I_0 P_u P_s E\}_{s \in D}) .$$

Hence $Z \in L(\{I_0 P_u P_s E\}_{s \in D})$. Therefore 5.1(ii) gives $Z \circ \theta_r \in L(I_0 P_u E)$ for all $r \in D$. Since u is arbitrary, we have $Z \circ \theta_r \in L(\{I_0 P_u E\}_{u \in D})$. Therefore (a) follows from

$$Z \circ \theta_T = \textstyle\sum_{r \in D'} Z \circ \theta_r \cdot \chi_{(T=r)} . \qquad \square$$

The next lemma is crucial to our construction. It also establishes the notations used in the theorem to follow.

LEMMA 5.3. *For arbitrary $s \in D$, the function $Y_s \equiv \sup \{d(X_0, X_t) : t \in D \cap [0, s]\}$ belongs to $L(C^*E)$.*

PROOF. Let $I_0 \in C^*$ and $I = I_0 E$. Conceivably Y_s could be undefined on a full set in (Ω, L, I). That this is not the case is part of the assertion. Let $0 \leqq a < b$ be arbitrary. Let a_k be a strictly increasing sequence in (a, b) converging to some a_∞ and such that $(d(X_0, X_t) > a_k)$ and $(d(X_0, X_t) > a_{k+1} - a_k)$ belong to $L_t(\{I_0 P_r E\}_{r \in D})$ for all $t \in D$, $k = 0, 1, 2 \cdots$ (this is possible by (2.4)). Construct

(i) a sequence $0 = p_0 < p_1 < \cdots < p_k < \cdots$ of strictly increasing integers such that $2^{-p_k} \cdot s < \delta((a_{k+1} - a_k)/2)$ for $k \geqq 1$,

(ii) $D_k \equiv \{j 2^{-p_k} \cdot s : 0 \leqq j \leqq 2^{p_k}\}$ for $k \geqq 0$, and $D_{-1} \equiv \phi$,

(iii) for each $t \in D \cap [0, s]$, the integer $j(t) = j$ such that $t \in D_j - D_{j-1}$,

(iv) in case $j(t) = k + 1$, the smallest member $r(t)$ of D_k such that $t < r(t)$,

(v) $A_k \equiv \bigwedge_{t \in D_k} (d(X_0, X_t) \leqq a_{j(t)})$,

(vi) $T_k \equiv \sum_{t \in D_k} t \chi_{(d(X_0, X_t) > a_{j(t)}); d(X_0, X_u) \leqq a_{j(u)} \text{ for all } u \in D_k \cap [0, t))} + s \chi_{A_k}$ (= first time in D_k when $d(X_0, X_t) > a_{j(t)}$).

Then T_k is a stopping time for $(\Omega, L, \{I_0 P_r E\}_{r \in D}, (X_r), (\theta_r))$. Also obvious is the relation: for all $k \geqq 0$, $A_k > A_{k+1}$ and

$$A_k - A_{k+1} < \textstyle\bigvee_{t \in D_{k+1}} (T_{k+1} = t; d(X_t, X_{r(t)}) > a_{k+1} - a_k) .$$

Hence

$$\begin{aligned} I(A_k - A_{k+1}) &\leqq \textstyle\sum_{t \in D_{k+1}} I(\chi_{(T_{k+1} = t)} E^{X(t)}(d(X_0, X_{r(t)-t}) > a_{k+1} - a_k)) \\ &\leqq \textstyle\sum_{t \in D_{k+1}} I(\chi_{(T_{k+1} = t)})(a_{k+1} - a_k)/2 \\ &= (a_{k+1} - a_k)/2 . \end{aligned}$$

The first inequality follows from the Markov property since $(T_{k+1} = t) \in F_t$; the second because $r(t) - t < 2^{-p_k} \cdot s < \delta((a_{k+1} - a_k)/2)$ for $t \in D_{k+1}$. Therefore the set

$$B \equiv \textstyle\bigwedge_{k=0}^{\infty} A_k = \bigwedge_{t \in D \cap [0,s]} (d(X_0, X_t) \leqq a_{j(t)})$$

is measurable in (Ω, L, I). It is obvious that for all $\omega \in B$, we have $d(X_0(\omega), X_t(\omega)) < b$ for all $t \in D \cap [0, s]$; and for all $\omega \in -B$, we have $d(X_0(\omega), X_t(\omega)) > a$ for some $t \in D \cap [0, s]$. Repeating the above construction we obtain, for every

n, measurable sets $\phi = B(0), B(1), \cdots, B(2^n)$ such that

$$\text{for all} \quad \omega \in B(k), \qquad d(X_0(\omega), X_t(\omega)) < k2^{-n}M \quad \text{for all} \quad t \in D \cap [0, s];$$
$$\text{for all} \quad \omega \in -B(k), \qquad d(X_0(\omega), X_t(\omega)) > (k-1)2^{-n}M \quad \text{for some} \quad t \in D \cap [0, s].$$

where M is the diameter of S. Define

$$Z_n = \sum_{k=1}^{2^n} K2^{-n}M\chi_{B(k)-B(k-1)}.$$

Then $Z_n - Z_{n+1} \leqq 2^{-n}M$ for all n, and so $Z \equiv \lim Z_n$ is in $L(I)$. It is easily verified that Z, whenever defined, is the supremum of $\{d(X_0, X_t) : t \in D \cap [0, s]\}$. In other words Y_s is equal to Z and so belongs to $L(I) = L(I_0 E)$. As $I_0 \in C^*$ is arbitrary, the lemma is proved. □

The next lemma says that with high probability, X will remain near $X(0)$ at least for a while. Perhaps one should bear in mind that the strong Markov property is not available; otherwise the proof would be much simpler.

LEMMA 5.4. *Suppose $\varepsilon > 0$, $s < \delta(\varepsilon)$, and $c > 2\varepsilon$. Then $E(Y_s > c) < 2\varepsilon$ on its domain. (Recall that x is in the domain of $E(Y_s > c)$ iff $(Y_s > c) \in L(E^x) = L(\delta_x E)$ where δ_x is the point mass concentrated at x.)*

PROOF. Let $I_0 = \delta_x$ and $I = I_0 E = E^x$. Let $a = \varepsilon$ and $b = 2\varepsilon$. Construct as in the proof of the previous lemma the objects (a_k), (p_k), (D_k), and (A_k). Then, as seen in that proof,

$$E^x(A_k - A_{k+1}) = I(A_k - A_{k+1}) < a_{k+1} - a_k.$$

If we let f be a continuous function on S^2 with $\chi_{(d>a_0)} \leqq f \leqq \chi_{(d>\varepsilon)}$, then

$$E^x(-A_0) = E^x(d(X_0, X_s) > a_0) \leqq E^x(f(X_0, X_s)) = P_s(f(x, \cdot))(x) \leqq \varepsilon.$$

The first equality is a consequence of the definition of E, while the last inequality is a direct consequence of condition (d). Combining,

$$\begin{aligned} E^x(-\textstyle\bigwedge_{k=0}^{\infty} A_k) &= E^x((-A_0) \vee (A_0 - A_1) \vee (A_1 - A_2) \vee \cdots) \\ &< \varepsilon + (a_1 - a_0) + (a_2 - a_1) + \cdots \\ &= \varepsilon + a_\infty - a_0 \\ &< 2\varepsilon. \end{aligned}$$

Since $(Y_s > c) < -\bigwedge_{k=0}^{\infty} A_k$ for all $c > 2\varepsilon$, the lemma is proved. □

The next theorem is the main result of this section.

THEOREM 5.5. *Let real numbers $\varepsilon \in (0, 1)$ and $s \in D$ be given. Then there exists an integer n which depends only on ε, s, and the operation δ in* (d) *such that for every $I_0 \in C^*$ there exist random variables*

$$0 = U_1 \leqq \cdots \leqq U_n$$

and a measurable set G in $(\Omega, L, I \equiv I_0 E)$ such that

$$I(G \vee (U_n \leqq s)) < 2\varepsilon,$$

and such that for all $\omega \in -G$ *we have*

$$d(X_t(\omega), X_r(\omega)) < 4\varepsilon \quad \textit{whenever} \quad t, r \in (U_i(\omega), U_{i+1}(\omega)) \quad \textit{for some } i\,.$$

(In particular, except for a set of measure no greater than 2ε, *the sample functions* $X(\cdot, \omega)$ *can be broken into n pieces on* $[0, s]$, *each of diameter no greater than* 4ε.*)*

PROOF. Let q be an integer greater than 2 and $2s \cdot \delta(\varepsilon/2)^{-1}(1-\varepsilon)^{-1}$. Let m be an integer so large that $(q-2)^m/(q-1)^m < \varepsilon$, and let $n = qm + 1$. We will prove that the integer n has the desired properties. Take any $I_0 \in C^*$ and let $I = I_0 E$. Let $a = \varepsilon$, $b = \varepsilon + \varepsilon/n$. As in the proof of Lemma 5.3 construct the objects (a_k), (p_k), (D_k), $j(t)$, (A_k) and (T_k). For $k \geqq 1$, define

$$Z_k \equiv \sup\{d(X_0, X_t) : t \in D \cap [0, 2^{-p_k} \cdot s]\}\,.$$

In view of Lemma 5.3, we can find a sequence (a_k') of numbers such that $a_k < a_k' < a_{k+1}$ and

$$(Z_k > a'_{k+1} - a_k) \in L(\{I_0 P_u E\}_{u \in D})$$

for every k. Define

$$B_k \equiv (Z_k \circ \theta_{T_{k+1}} > a'_{k+1} - a_k)\,.$$

Then $\chi_{B_k} = \chi_{(Z_k > a'_{k+1} - a_k)} \circ \theta_{T_{k+1}}$ and so, by Lemma 5.2,

$$B_k \in L(\{I_0 P_u E\}_{u \in D})$$

and

$$E(B_k) = E(E^{X(T_{k+1})}(Z_k > a'_{k+1} - a_k))\,.$$

Since $2^{-p_k} \cdot s < \delta((a_{k+1} - a_k)/2)$ and $a'_{k+1} - a_k > a_{k+1} - a_k$, Lemma 5.4 implies that the last displayed expression is always bounded by $a_{k+1} - a_k$. It follows that $B \equiv \bigvee_{k=1}^{\infty} B_k$ belongs to $L(\{I_0 P_u E\}_{u \in D})$ and

$$E(B) \leqq \sum_{k=1}^{\infty} (a_{k+1} - a_k) = a_\infty - a_1 < \varepsilon/n$$

on some full subset of $(S, C, \{I_0 P_u\}_{u \in D})$.

Likewise, since

$$T_{k+1} \leqq T_k \quad \text{in} \quad L(\{I_0 P_u E\}_{u \in D})$$

and

$$(T_k - T_{k+1} > 2^{-p_k} \cdot s) < B_k\,,$$

The function $T \equiv \lim_{k \to \infty} T_k$ belongs to $L(I_0 P_u E\}_{u \in D})$ also.

Now we start the process again at time T_1 and repeat the above construction n times. To be precise, define for $1 \leqq i \leqq n$

$$V_1 \equiv T_1\,; \qquad U_1 \equiv T\,; \qquad G_1 \equiv B\,;$$

$$V_i \equiv V_{i-1} + V_1 \circ \theta_{V_{i-1}}\,; \qquad U_i \equiv V_{i-1} + U_1 \circ \theta_{V_{i-1}}\,; \qquad G_i \equiv \theta^{-1}_{V_{i-1}}(G_1)\,.$$

$$G = G_1 \vee \cdots \vee G_n\,.$$

The set G will now be shown to be the exceptional set having the properties described in the conclusion of the theorem. To this end let ω be in $-G$ and in the common domain of the V_i and U_i's. Then

$$\omega' \equiv \theta_{V_{i-1}}(\omega) \in -G_1 = -B = -\bigvee_{k=1}^{\infty} B_k$$

and so

$$d(X(r), X(T_{k+1}))(\omega') < a'_{k+1} - a_k < a_{k+2} - a_k$$

for $k \geqq 1$ and $r \in [T_{k+1}(\omega'), T_{k+1}(\omega') + 2^{-p_k} \cdot s] \subset [T_{k+1}(\omega'), T_k(\omega')]$. Summing over k, we have for all $r \in (T(\omega'), T_1(\omega')]$,

$$d(X(r), X(T_1))(\omega') < 4(a_\infty - a_1) < 2\varepsilon .$$

Substituting $\theta_{V_{i-1}}(\omega)$ for ω' and simplifying,

$$d(X(t), X(V_i))(\omega) < 2\varepsilon \quad \text{for} \quad t \in (U_i(\omega), V_i(\omega)] .$$

On the other hand, if $t \in [V_i(\omega), U_{i+1}(\omega))$, we can write $t = r + V_i(\omega)$ where

$$r < U_{i+1}(\omega) - V_i(\omega) = T(\theta_{V_i}(\omega)) \leqq T_{j(r)}(\theta_{V_i}(\omega)) .$$

Therefore

$$d(X(t), X(V_i))(\omega) = d(X(r), X(0))(\theta_{V_i}(\omega)) \leqq a_{j(r)} < a_\infty < 2\varepsilon$$

because $T_{j(r)}$ is the first time when the process X, sampled on $D_{j(r)}$, wanders out of the $a_{j(r)}$-neighborhood of $X(0)$. The second conclusion of the theorem has been proved. It remains to see that the exceptional sets G and $(U_n \leqq s)$ have small measures as alleged. It is immediate that $I(G) < \varepsilon$ because for each i, by 5.2(c),

$$I(G_i) = I(\chi_B \circ \theta_{V_{i-1}}) = I(E^{X(V_{i-1})}(B)) < \varepsilon/n .$$

To show that $I(U_n \leqq s)$ is very small we first prove that $I(V_q \leqq s)$ is not too large and then that $I(U_n \leqq s) \leqq I(V_q \leqq s)^m$. Now for each i we have, for all $r \in D$ with $r > \delta(\varepsilon/2)$,

$$\begin{aligned} E(V_{i+1} - V_i) = E(T_1 \circ \theta_{V_i}) &= E(E^{X(V_i)}(T_1)) \geqq E(rE^{X(V_i)}(T_1 \geqq r)) \\ &\geqq E(rE^{X(V_i)}(Y_r \leqq a_1)) \geqq r(1 - \varepsilon) \end{aligned}$$

where $Y_r = \sup\{d(X_0, X_t) : 0 \leqq t \leqq r\}$ as defined in 5.4. Summing these inequalities over i, we see that $E(V_q) \geqq q\delta(\varepsilon/2)(1 - \varepsilon) > 2s$. Observing that $0 \leqq V_q \leqq qs$, we obtain by an elementary calculation

$$E(V_q \leqq s) \leqq (q - 2)/(q - 1) .$$

With $V_q, V_{2q}, \cdots, V_{(m-1)q}, V_{mq}$ abbreviated by $V, V', \cdots, V'', V'''$ respectively, an easy induction shows $V' - V = V \circ \theta_V, \cdots, V''' - V'' = V \circ \theta_{V''}$. Hence, using 5.2(b) repeatedly, we obtain

$$\begin{aligned} I(U_n \leqq s) &\leqq I(V_{n-1} \leqq s) = I(V''' \leqq s) \\ &\leqq I(V \leqq s;\ V' - V \leqq s;\ \cdots;\ V''' - V'' \leqq s) \\ &= I(V \leqq s;\ V' - V \leqq s;\ \cdots;\ V \circ \theta_{V''} \leqq s) \\ &= I(V \leqq s;\ V' - V \leqq s;\ \cdots;\ E^{X(V'')}(V \leqq s)) \\ &\leqq I(V \leqq s;\ V' - V \leqq s;\ \cdots) \cdot (q - 2)/(q - 1) \\ &\cdots \\ &\leqq (q - 2)^m/(q - 1)^m \\ &< \varepsilon . \end{aligned}$$

The proof is completed. □

The extension of the Markov process to the parameter set $[0, \infty)$ is now an easy corollary. Define X on $[0, \infty)$ by

$$X(t, \omega) = \lim_{r \in D, r \to t} X(r, \omega) ,$$

the domain of X composing of those (t, ω) for which the limit exists.

PROPOSITION 5.6. (a) *X is a measurable mapping from* $([0, \infty) \otimes \Omega, K, \mu \otimes C^*E)$ *to* (S, C). *Here K stands for the continuous functions on* $[0, \infty) \otimes \Omega$ *with compact supports, and μ for the Lebesgue measure on* $[0, \infty)$.

(b) *For every $t \in [0, \infty)$, the function $X_t \equiv X(t, \cdot)$ is a measurable mapping from* (Ω, L, C^*E) *to* (S, C). *In particular $X(t, \cdot)$ is defined on a full subset of* (Ω, L, C^*E)*—almost sure lack of discontinuity at t.*

PROOF. Let $s \in D$, $I_0 \in C^*$ be given. By the previous theorem, for each k there exists random variables $0 = U_0^k = V_0^k \leqq U_1^k \leqq V_1^k \leqq \cdots \leqq U_{n(k)}^k \leqq V_{n(k)}^k$ such that, except on a subset B_k of Ω with measure $I(B_k) = I_0E(B_k)$ at most 2^{-k}, we have $U_{n(k)}^k > s$ and the sample function $X(\cdot, \omega)$ differs on $(U_i^k, U_{i+1}^k)(\omega)$ from $X(V_i^k(\omega), \omega)$ by at most 2^{-k}. Define X^k on $[0, \infty) \otimes \Omega$ by

$$X^k(t, \omega) = \sum_{i=0}^{n(k)-1} \chi_{\{(t,\omega): U_i^k(\omega) < t < U_{i=1}^k(\omega)\}} X(V_i^k(\omega), \omega) .$$

Then X^k is clearly measurable. Moreover X^k is away from X^m by at most $2^{-k} + 2^{-m}$ on $[0, s] \times -\bigvee_{m \geqq k} B_m$ if $m \geqq k$. Since the last written set has measure at most $2^{-k}s$, the sequence X^k converges a.e. on $[0, s] \otimes \Omega$ relative to $\mu \otimes I$. The limit is readily verified to be X. In particular X is $\mu \otimes I$-measurable.

To prove (b) take $t \in [0, \infty)$ and $I \equiv I_0E$. Let $\varepsilon_k > 0$ be terms of a convergent series. Let (r_k) and (s_k) be sequences in D with $r_k \uparrow t$ and $s_k \downarrow t$, in such a way that the set

$$A_k = (\sup_{s \in D \cap [r_k, s_k]} d(X(r_k), X(s)) > \varepsilon_k)$$

has measure $I(A_k)$ less than ε_k. This is possible by 5.4. (Strictly speaking ε_k should be replaced by a slightly larger number so as to ensure the measurability of A_k.) Then $A \equiv \bigvee_{n=1}^{\infty} \bigwedge_{k=n}^{\infty} A_k$ is a null set. $X(t, \cdot)$ is clearly defined on $-A$. Moreover $d(X(t), X(r_k)) \leqq \varepsilon_k$ on $-A_k$ and so $X(r_k)$ converges a.e. (I) to X_t. In particular X_t is measurable as asserted. □

The next proposition merely extends 5.1 to cover the parameter set $[0, \infty)$. Its easy proof by continuity is omitted.

PROPOSITION 5.7. *Let the process X be defined as above. Again we write $X(t)$ for X_t when convenient.*

(i) Let $I_0 \in C^*$ and $I = I_0E$. Then the probability $I_{X(s)}$ induced by $X(s)$ on (S, C) relative to I is equal to I_0P_s.

(ii) For $s \geqq 0$ define the shift θ_s on Ω by $\theta_s(\omega) =$ the member of Ω whose value at $t \in D$ is given by $X(t + s, \omega)$. Then θ_s is a measurable mapping from (Ω, L, C^*E) to itself.

(iii) Let $s \geqq 0$, $I_0 \in C^*$, and $Z \in L(I_0P_sE)$ be given. Write I for I_0E and L_s

for the family

$$\{g(X(s_1), \cdots, X(s_n)) : n \in N;\ g \text{ continuous on } S^n;\ s_1 \leqq \cdots \leqq s_n \leqq s\}\,.$$

Then

$$I(Z \circ \theta_s \mid F(L_s, I)) = E^{X(s)}(Z) \text{ on an } I\text{-full set}\,. \qquad \square$$

6. A submartingale convergence theorem. In [1], Bishop proves that if certain "λ-norms" of a martingale converges, then the martingale converges a.e. The proof being constructive, bounds are given which also give numerical information concerning finite martingale sequences. This is in distinct contrast to the classical convergence theorem. The purpose of the present section is two-fold. First, we amend Bishop's definition for admissible functions for his "λ-norms", so that functions like $|x|^p (p > 1)$ are included. Second, Bishop's theorem is extended to submartingales.

A sequence of random variables (X_n) is called a *submartingale* if for every n

$$E(X_{n+1};\ A) \geqq E(X_n;\ A)$$

where A is any set measurable relative to $X_1, \cdots, X_n$ (in the terminology of Section 3, χ_A belongs to the Borel family generated by $\{X_1, \cdots, X_n\}$). If $\geqq$ is replaced by $=$ in the definition, (X_n) is called a martingale.

Let λ be a nonnegative even function with a continuous first derivative on R, and a continuous positive second derivative on $R - \{0\}$ and such that $\lambda(0) = 0$. (e.g., $|x|^p$ would be admissible if $p > 1$). For $r > 0$ define

$$\alpha(r) = \inf\{(r - x)^{-2} \textstyle\int_x^r \lambda'(t) - \lambda'(x)\, dt : |x| \leqq r\}\,,$$
$$\beta(r) = \inf\{(r - y)^{-2} \textstyle\int_y^r \lambda'(r) - \lambda'(t)\, dt : |y| \leqq r\}\,.$$

The infimums exist since the functions involved are continuous. A little calculation shows that these infimums are positive, and are continuous functions of $r \in (0, \infty)$. Moreover, if we define

$$\theta(r) = \alpha(r) \wedge \beta(r)\,,$$

then the function $\theta(|x| \vee |y|)(y - x)^2$ can be extended to a continuous function on R^2 which vanishes on the diagonal $x = y$. We have the following inequality.

LEMMA 6.1. *$\lambda(y) - \lambda(x) - \lambda'(x)(y - x) \geqq \theta(|x| \vee |y|)(y - x)^2$ for all $x, y \in R$.*

PROOF. The left side of the inequality can be written as $\int_x^y \lambda'(t) - \lambda'(x)\, dt$ which is equal to $\int_{-x}^{-y} \lambda'(t) - \lambda'(-x)\, dt$ since λ is an even function. In case $y > |x|$, the first integral is no less than $\alpha(y)(y - x)^2$ by the very definition of α. In case $-y > |x|$, the second integral is no less than $\alpha(-y)(y - x)^2$ for the same reason. Thus the left side is no less than

$$\alpha(|y|)(y - x)^2 = \alpha(|x| \vee |y|)(y - x)^2 \geqq \theta(|x| \vee |y|)(y - x)^2$$

when $|y| > |x|$. Similar reasoning will prove the inequality when $|y| < |x|$. Since both sides of the inequality are continuous in x and y, the lemma is proved. $\square$

From this inequality, Bishop [1] gives the following estimates for a martingale.

THEOREM 6.2. *Let K and ε be positive real numbers. Then there exists $\delta > 0$ such that for every finite martingale $X_0, \cdots, X_n$ with $E(\lambda(X_n)) \leqq K$ and $E(\lambda(X_n)) - E(\lambda(X_0)) < \delta$ we have*

(a) $E(\bigvee_{i=1}^n (|X_i - X_0| > \varepsilon)) < \varepsilon$.

(Strictly speaking, the first ε in the inequality should be replaced by a slightly larger number so as to ensure the measurability of the set involved.)

We now generalize this theorem to submartingales.

THEOREM 6.3. *Let ε be a positive real number and let α be an operation which assigns to every positive real number another positive real number. Then there exists $\delta' > 0$ such that for every finite submartingale $X_0, \cdots, X_n$ with*

(b) $E(\lambda(X_0); \lambda(X_0) > \alpha(\theta)) < \theta$ *for all* $\theta > 0$,

(c) $E(\lambda(X_n)) - E(\lambda(X_0)) < \delta'$,

(d) $E(X_n) - E(X_0) < \delta'$,

then $X_0, \cdots, X_n$ satisfy condition 6.2(a) *above.*

PROOF. Let ε, α be given and let $K \equiv \alpha(1) + 2$. Let δ be associated to K and $\varepsilon/3$ as in the previous theorem. Choose $r > 0$ so large that $\lambda(r) > \alpha(\delta/4)$. Then choose s so small that $|\lambda(x) - \lambda(y)| < \delta/4$ if $|x| \leqq r$ and $|x - y| < s$. Define $\delta' \equiv \delta'(\varepsilon) \equiv \varepsilon^2/9 \wedge s\delta/(4\lambda(r)) \wedge \delta/4 \wedge 1$. We will show that δ' has the desired properties.

So let $X_0, \cdots, X_n$ be a submartingale satisfying conditions (b) through (d). There is no loss of generality in assuming that $X_0, \cdots, X_n$ are such that for some countable subset C_n of the set $\{\text{rational numbers}\}^{n+1}$, we have

$$\textstyle\sum_{r \in C_n} E(X_0 = r_0, \cdots, X_n = r_n) = 1$$

and that all terms in the series are positive. Clearly, then, each subsequence $X_0, \cdots, X_k$ has the same property. In this case we can constructively define the *conditional expectations*

$$U_k \equiv E(X_n \mid X_0, \cdots, X_k) \equiv \textstyle\sum_{r \in C_k} \chi_{(X_0 = r_0, \cdots, X_k = r_k)} \dfrac{E(X_n; X_0 = r_0, \cdots, X_k = r_k)}{E(X_0 = r_0, \cdots, X_k = r_k)}\,.$$

Then it is well known that $(U_k)_{k=0}^n$ is a martingale, while $(V_k)_{k=0}^n$, defined by $V_k \equiv X_k - U_k$, is a non-positive submartingale. Also obvious is the relation $U_n = X_n$. Hence

$$0 \geqq E(V_0) = E(X_0) - E(U_0) = E(X_0) - E(U_n) = E(X_0) - E(X_n) > -\delta'\,.$$

Let ν be the first time k such that $V_k < -\varepsilon/3$. Then V_0, V_ν is again a submartingale and so

$$E\left(\bigvee_{i=0}^n \left(V_i < -\frac{\varepsilon}{3}\right)\right) = E\left(V_\nu < -\frac{\varepsilon}{3}\right) \leqq \frac{3}{\varepsilon} E(-V_\nu) \leqq \frac{3}{\varepsilon} E(-V_0)$$
$$< \frac{3\delta'}{\varepsilon} \leqq \frac{\varepsilon}{3}\,.$$

Next, observe that condition (b) with $\theta = 1$ implies $E(\lambda(X_0)) \leqq \alpha(1) + 1$ and so condition (c) implies $E(\lambda(X_n)) \leqq \alpha(1) + 1 + \delta' \leqq \alpha(1) + 2 = K$. Hence, for the martingale (U_k) we have

$$E(\lambda(U_k)) = E(\lambda(E(X_n \mid X_0, \cdots, X_k))) \leqq E(E(\lambda(X_n) \mid X_0, \cdots, X_k)) = E(\lambda(X_n)) \leqq K,$$

where the first inequality (Jensen) is a direct consequence of Lemma 6.1. At the same time

$$\begin{aligned} E(\lambda(U_n)) &= E(\lambda(X_n)) < E(\lambda(X_0)) + \delta' \leqq E(\lambda(X_0)) + \frac{\delta}{4} \\ &= E(\lambda(X_0); (|X_0| \leqq r) \wedge (|V_0| < s)) + E(\lambda(X_0); (|X_0| \leqq r) \wedge (|V_0| \geqq s)) \\ &\quad + E(\lambda(X_0); (|X_0| > r)) + \frac{\delta}{4} \\ &\leqq E\left(\lambda(U_0) + \frac{\delta}{4}\right) + \lambda(r)E(|V_0| \geqq s) + E(\lambda(X_0); (\lambda(X_0) \geqq \lambda(r))) + \frac{\delta}{4} \\ &\leqq E(\lambda(U_0)) + \frac{\delta}{4} + \lambda(r)s^{-1}E(-V_0) + \frac{\delta}{4} + \frac{\delta}{4} \\ &= E(\lambda(U_0)) + \delta . \end{aligned}$$

The hypothesis of the previous theorem is satisfied by K, $\varepsilon/3$, δ, and $U_0, \cdots, U_n$. Hence

$$E\left(\bigvee_{i=1}^{n}\left(|U_i - U_0| > \frac{\varepsilon}{3}\right)\right) < \frac{\varepsilon}{3} .$$

Combining this with

$$E\left(\bigvee_{i=0}^{n}\left(|V_i| > \frac{\varepsilon}{3}\right)\right) = E\left(\bigvee_{i=0}^{n}\left(V_i < -\frac{\varepsilon}{3}\right)\right) < \frac{\varepsilon}{3},$$

we obtain via the triangle inequality

$$E(\bigvee_{i=1}^{n} (|X_i - X_0| > \varepsilon)) < \frac{2\varepsilon}{3} < \varepsilon . \qquad \square$$

COROLLARY 6.4. *Suppose $(X_k)_{k=1}^{\infty}$ is a submartingale such that $(\lambda(X_k))_{k=1}^{\infty}$ is uniformly integrable (i.e. the random variables $\lambda(X_k)$ satisfy* (b) *for one and the same operation α) with $E(\lambda(X_k))$ and $E(X_k)$ convergent. Then X_k converges* a.e.

The corollary is proved by applying the previous theorem to successive blocks of the sequence $(\lambda(X_k))$. (See [1] page 225.)

We remark that in Theorem 6.3, the number δ' depends in an essential way on the operation α. Thus Theorem 6.2 is untrue for a submartingale even with the additional assumption that $E(X_n) - E(X_0) < \delta$. Suppose $\delta(K, \varepsilon)$ can be computed so that condition (a) is satisfied for every submartingale with $E(\lambda(X_k)) \leqq K$, $E(\lambda(X_n)) - E(\lambda(X_0)) < \delta$, and $E(X_n) - E(X_0) < \delta$. For illustration take $\lambda(x) = x^2$, $\varepsilon = \frac{1}{2}$, and $K = 1$. We may assume $\delta < \frac{1}{2}$. Let X_0 assume values 0 and $-\delta^{-1}$ with probabilities $1 - \delta^2$ and δ^2 respectively, and let X_1 be independent of X_0

and assume values -1 and 1 with probability $\frac{1}{2}$ for each. Then X_0, X_1 is a submartingale. Also $E(X_0^2) = \delta^{-2} \cdot \delta^2 = 1$, $E(X_1^2) = 1$, while $E(X_1) = 0$ and $E(X_0) = -\delta$. So (X_0, X_1) satisfies all the conditions laid down. However

$$E(|X_0 - X_1| > \tfrac{1}{2}) \geqq E(X_0 = 0) = 1 - \delta^2 > \tfrac{3}{4}.$$

By the same taken none of the three conditions in Corollary 6.4 can be dropped.

In case λ is convex monotone (but not necessarily smooth) an inequality with the same flavor as (a) has been derived for submartingales in [5].

Acknowledgment. The author wishes to thank Professor Errett Bishop for his valuable suggestions on the manuscript.

REFERENCES

[1] BISHOP, E. (1967). *Foundations of Constructive Analysis.* McGraw-Hill, New York.

[2] BISHOP, E. and CHENG, H. (1972). Constructive measure theory, *American Mathematical Society Memoirs* **116**.

[3] BLUMENTHAL, R. M. (1953). An extended Markov property. *American Mathematical Society Transactions* **85** 52–72.

[4] CHAN, Y. K. (1972). A constructive approach to the theory of stochastic processes. *American Mathematical Society Transactions* **165** 37–44.

[5] DINGES, H. (1971). Inequalities leading to a proof of the classical martingale convergence theorem. A report on a meeting at Oberwolfach, May 1970. Springer-Verlag, Berlin.

[6] KINNEY, J. R. (1953). Continuity properties of sample functions of Markov processes. *American Mathematical Society Transactions* **74** 280–302.

[7] MEYER, P.-A. (1966). *Probability and Potentials.* Blaisdell, Waltham.

[8] NEVEU, J. (1965). *Mathematical Foundations of the Calculus of Probability.* Holden-Day, San Francisco.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195