The contributions of the Russian school to the theory of rings are outstanding. Indeed the architects of much of the theory are the triumvirs Gelfand, Naimark, and Shilov. How fortunate for us that one of these three has put down in extenso his essence in writing. One can only purloin Dostoyevsky's famous last words about the original Three Brothers and exclaim in admiration: "Hurrah for Naimark."

EDGAR R. LORCH

Stationary processes and prediction theory. By H. Furstenberg. Annals of Mathematics Study No. 44. Princeton Univ. Press, Princeton, N. J., 1960. 283 pp. $5.00.

This work is an elaboration of the author's doctoral dissertation at Princeton. The limitations of the classical prediction theory of stochastic processes are first discussed. In the light of this discussion a new prediction theory for single time-sequences is formulated. The ideas uncovered in the course of this development are shown to have interesting ramifications outside prediction theory proper. In the author's opinion the discussion of these offshoots, for which prediction is more of an "excuse" than a "reason" (p. 7), constitutes the most important part of the book. In this review we shall touch upon the critique, the new theory as well as the offshoots, but greater emphasis will be placed on the second topic in relation to the third both from considerations of space and the reviewer's predilections. We shall conclude the review with some general remarks on the work.

The book abounds with strange terminology, which has to be understood to get any insight into it. It is also rather complex in structure. In this review we have thus been obliged to state definitions and to indulge in an abridged and sometimes over-simplified exposition of the author's theory. It is hoped that this exposition will serve as a guide to the prospective reader of the book.

I. LIMITATIONS OF CLASSICAL PREDICTION THEORY

It is well known that we are able to prognosticate the future in many realms in which strictly deterministic laws do not prevail. One scientific explanation of this ability is that such realms are governed by probabilistic laws in which the underlying probability measure is invariant under time-shifts. More precisely, underlying such a realm is a probability space $(\Omega, \mathcal{B}, P)$, and a P-measure-preserving transformation T on Ω onto Ω. We are interested in some $\mathcal{B}$-measurable function f on Ω or what amounts to the same thing, in a stationary stochastic process (S.P.) (f_n, $-\infty < n < \infty$), where $f_n(\omega) = f(T^n\omega)$.[1] As conceived classically, prediction consists in finding the conditional probability $P(\cdot \mid \mathcal{B}_0)$ or the conditional expectation $E(\cdot \mid \mathcal{B}_0)$ relative to the "past and present" Borel subalgebra $\mathcal{B}_0$ of $\mathcal{B}$, i.e. the algebra generated by the functions f_k, $k \leqq 0$. Specifically, the problem of prediction is to estimate $P(\cdot \mid \mathcal{B}_0)$ and $E(\cdot \mid \mathcal{B}_0)$ from past observations of a typical time-sequence ($f_n(\omega)$, $n \leqq 0$) of the S.P., $\omega \in \Omega$. Its solvability hinges on the ergodicity of T.[2]

In actual practice, however, we have often to predict the future values of a single time-sequence (x_n, $-\infty < n < \infty$); e.g. x_n may be the temperature at time n at a given place. In such cases we usually regard the x_n-sequence as being a typical time-sequence of an ergodic S.P. (f_n, $-\infty < n < \infty$), say $x_n = f_n(\omega_0)$, $\omega_0 \in \Omega$, and (naturally) take the predicted value of x_n to be $E(f_n \mid \mathcal{B}_0)(\omega_0)$. This procedure is questionable, however. For, since $E(f_n \mid \mathcal{B}_0)$ is an equivalence class of functions any two of which may differ on a (variable) set of zero P-measure, the symbol $E(f_n \mid \mathcal{B}_0)(\omega_0)$ has no clear-cut meaning. Except in special cases[3] it will be quite arbitrary to represent the equivalence class by one of its members. Thus in general there is no rational way of finding $E(f_n \mid \mathcal{B}_0)(\omega_0)$. In short, *it is impossible to derive directly from the prediction theory of stochastic processes a prediction theory for general individual time sequences.* The author makes a gallant attempt to develop the latter theory *ab initio* and *de novo.*

II. Prediction theory for a single time-sequence

A. **Regular sequence.** The author assumes that the time-sequence of interest has values in a compact metric space Λ. In many applications Λ will be a subset of the set $\mathcal{C}$ of complex numbers, but there are several cases in which $\Lambda \not\subseteq \mathcal{C}$, e.g. coin tossing. He denotes by Λ_∞ the set of bisequences $\xi = (\xi(n), -\infty < n < \infty)$ such that $\xi(n) \in \Lambda$. The set of one-sided sequences $\xi = (\xi(n), n \leqq 0)$ is denoted by Λ_∞^-. The spaces Λ_∞, Λ_∞^- are endowed with the weak product topology. Given $\xi \in \Lambda_\infty^-$, we wish to predict its future, i.e. for $\nu > 0$ to assign a function p_ν such that $p_\nu(\lambda)$ is the probability that $\xi(\nu) = \lambda$, $\lambda \in \Lambda$. This assignment has to be made on the basis of our knowledge of $\xi(n)$ for $n \leqq 0$. For such prediction to make sense the sequence ξ has, of course, to

[1] For simplicity of discussion, we are supposing that time is discrete. The author deals exclusively with this case.

[2] For details, see P. Masani and N. Wiener: *Non-linear prediction* in *Probability & Statistics*, The Harald Cramer volume, Almquist & Wiksell, Uppsala, 1959, pp. 190–212.

[3] E.g. when Ω is a topological space and a function in the family $E(f_n \mid \mathcal{B}_0)$ is continuous on Ω; or when Ω is a topological space for which Radon-Nikodym derivatives are definable as limits a.e. of quotients of measures of neighborhoods.

exhibit some statistical regularity and be free from erratic changes. The author effects the explication[4] of these vague requirements by means of the fundamental concepts of *stochastic sequence* and *regular sequence*. These concepts involve some preliminary notions:

1. DEFINITION. $\zeta \in \mathcal{C}_\infty$ *is called a* numerically-derived sequence *of* $\xi \in \Lambda_\infty$, *if*

$$(1) \qquad \zeta(n) = \psi\{\cdots, \xi(n-1), \xi(n)\}, \qquad -\infty < n < \infty,$$

where $\psi \in C(\Lambda_\infty)$, *i.e.* ψ *is a continuous complex-valued function on* Λ_∞. *This definition is also to hold when* $\mathcal{C}_\infty$, Λ_∞, "$-\infty < n < \infty$" *are replaced by* $\mathcal{C}_\infty^-$, Λ_∞^-, "$n \leqq 0$," *respectively.*

2. DEFINITION. (a) *The* upper (time-) average $E^-(\zeta)$ *of a sequence* $\zeta \in \mathcal{C}_\infty^-$ *is defined by* $\limsup_{n\to\infty} \{\sum_{n=0}^N \zeta(-n)/(N+1)\}$. *Similarly we define* $E_-(\zeta)$, *and in case* $E^-(\zeta) = E_-(\zeta)$, *the* (time-) average $E(\zeta)$.

(b) *The* upper density $D^-(S)$ *of a set* S *of integers is defined by* $E^-(\chi_S)$, *where* χ_S *is the restriction of the indicator-function of* S *to the set of integers* $n \leqq 0$.

The fundamental notions can now be defined:

3. DEFINITION.[4'] (a) $\xi \in \Lambda_\infty^-$ *is called* stochastic, *if every numerically derived sequence* ζ *of* ξ *has a time-average* $E(\zeta)$.

(b) $\xi \in \Lambda_\infty^-$ *is called* regular, *if* (i) ξ *is stochastic, and* (ii) *for all* $k \geqq 0$, *and all open* $\Delta_1, \cdots, \Delta_k \subseteq \Lambda$, *the set*

$$S = \{n: n \leqq 0,\ \xi(n) \in \Delta_1 \,\&\, \cdots \,\&\, \xi(n-k+1) \in \Delta_k\}$$

is either void or $D^-(S) > 0$.

As an example consider the case:

$$\xi(0) = \xi(-1) = 1, \qquad \xi(n) = (-1)^n, \qquad n \leqq -2.$$

It is easily seen that this ξ is stochastic but not regular. The irregularity stems from the change occurring in the definition of x_n when n crosses -2. It does not make sense to speak of "predicting the future" of such a sequence. On the other hand the regular sequences are amenable to such prediction as the author shows, cf. §§B–F below. There is a plentiful supply of regular sequences. For instance, every almost periodic sequence is regular (p. 41), and so are almost all time-sequences of one-sided stationary stochastic processes (5.3, p. 37).

B. **Prediction problem for a regular sequence.** A regular sequence

[4] Throughout this review the term "explication" will mean the transformation of an inexact pre-mathematical concept into a precise mathematical concept, cf. R. Carnap: *Logical foundations of probability*, Univ. of Chicago Press, Chicago, Ill., 1950, Ch. I.

[4'] Our definitions are variants of the author's (Def. 3.2, p. 24), convenient for purposes of this review.

$\xi \in \Lambda_\infty^-$ will have many quite arbitrary bisequential extensions $\bar{\xi} \in \Lambda_\infty$. For prediction purposes we must obviously confine attention to only those $\bar{\xi}$ which extend into the future the statistical features of ξ. The author explicates this vague requirement in terms of the following notion (p. 56):

4. DEFINITION. *$\bar{\xi} \in \Lambda_\infty$ is an L-extension*[5] *of $\xi \in \Lambda_\infty^-$, if $\bar{\xi}$ is an extension of ξ, and for all numerically derived sequences $\bar{\zeta}$ of $\bar{\xi}$,*

$$\sup_{-\infty < n < \infty} |\bar{\zeta}(n)| = \sup_{n \leqq 0} |\bar{\zeta}(n)|.$$

A heuristic or pre-mathematical version of the *prediction problem* can now be stated: *given a regular $\xi \in \Lambda_\infty^-$ to determine a probability measure μ on the space of all L-extensions $\bar{\xi}$ of ξ, so that $\mu(A)$ is the probability that the (actual) extension of ξ lies in A*. This problem is of course utterly trivial when ξ has only one L-extension $\bar{\xi}$, or as the author says when ξ is *deterministic* (p. 60). (An interesting class of regular, deterministic sequences are the $\xi \in C_\infty^-$ such that $\xi(n)$ is a uniform limit of $\sum_{\nu=1}^{k} c_\nu \exp\{2\pi i p_\nu(n)\}$, where p_ν is a polynomial with real coefficients (p. 83–). It follows that the almost periodic sequences are determinstic.)

In the author's theory the measure μ just referred to is to take the role played by the conditional probability $P(\cdot \mid \mathcal{B}_0)$ in the classical theory (cf. I above). Unfortunately μ cannot be obtained by a purely sequential analysis of the regular sequence ξ. To define it, the author has first to show that ξ is a "typical"[6] time-sequence of a one-sided, Λ-valued, stationary S.P. $X^-(\xi)$. Associated with $X^-(\xi)$ is a probability space $(\Omega^-(\xi), \mathcal{B}, P)$. The desired μ is constructed from this P by requiring that μ should behave like a conditional probability relative to the hypothesis that the event ξ has occurred (cf. p. 63). He is thus obliged to follow more or less the classical footsteps but at a more abstract level. The theory of Banach-algebras is the basic tool he uses to complete this analysis, which has many important features and so calls for a brief description (§§C–F below).

C. **The S.P. associated with a stochastic sequence.** With each stochastic $\xi \in \Lambda_\infty^-$ is associated the commutative linear algebra $A_0(\xi)$ of all numerically derived sequences ζ of ξ. Each such ζ has a pseudonorm

$$\|\zeta\|_\infty = \text{inf. } \{r : r \geqq 0 \,\&\, \bar{D}\{n : |\zeta(n)| > r\} = 0\}.$$

[5] The latter "L" is used because of a theorem of J. E. Littlewood (p. 59).

[6] "Typical" in the sense that the time-averages of all numerically derived sequences of ξ will be equal to the corresponding phase averages, cf. §D below.

This pseudo-norm becomes a norm on the quotient-algebra $A_1(\xi) = A_0(\xi)/N$, where N is the ideal of all ζ for which $\|\zeta\|_\infty = 0$, i.e. for which $E(|\zeta|) = 0$, E being the time-average, cf. Def. 3(a). The linear functional E on $A_0(\xi)$ therefore carries over to $A_1(\xi)$. Now the completion $A^-(\xi)$ of $A_1(\xi)$ is a commutative C^*-algebra with unit, to which the functional E can be extended by the Hahn-Banach Theorem. Moreover,

$$E(1) = 1, \; E(\zeta\bar{\zeta}) \geqq 0; \qquad E(\zeta\bar{\zeta}) = 0 \Rightarrow \zeta = 0.$$

In the author's terminology, $A^-(\xi)$ is an *E-algebra* (p. 12).

It follows that $A^-(\xi)$ is isomorphic to the algebra $C(\Omega^-(\xi))$ of all complex-valued, continuous functions on a compact Hausdorff space $\Omega^-(\xi)$. We may identify $\Omega^-(\xi)$ with the space of algebraic homomorphisms of $A^-(\xi)$ onto $\mathcal{C}$. By the Riesz-Markov Theorem, the functional E induces a probability measure P defined on the Borel sets of $\Omega^-(\xi)$ such that

$$(2) \qquad E(\zeta) = \int_{\Omega^-(\xi)} \zeta(\omega)P(d\omega), \qquad \zeta \in A^-(\xi) \simeq C(\Omega^-(\xi)).$$

Thus to each stochastic sequence $\xi \in \Lambda_\infty^-$ corresponds a probability space $(\Omega^-(\xi), \mathcal{B}, P)$, $\mathcal{B}$ being the family of Borel subsets of $\Omega^-(\xi)$.

Next, the author shows that to each stochastic $\xi \in \Lambda_\infty^-$ corresponds a one-sided, stationary, Λ-valued S.P. $X^-(\xi) = (x_k,\ k \leqq 0)$, $x_{n-1} = x_n T$, where the x_k are functions from $\Omega^-(\xi)$ to Λ, and T is a P-measure preserving transformation on $\Omega^-(\xi)$ into itself. The construction of the x_k (pp. 29, 30) is too technical to indicate here and is not entirely clear. As for T, we first define an operator T on Λ_∞^- by

$$(T\eta)(k) = \eta(k - 1), \qquad k \leqq 0, \qquad \eta \in \Lambda_\infty^-.$$

In an obvious way this induces a $\| \ \|_\infty$-preserving endomorphism on $A_0(\xi)$ and hence on its completion $A^-(\xi)$. Moreover T preserves E, i.e. $E(T\zeta) = E(\zeta)$, $\zeta \in A^-(\xi)$.[7] This T in turn induces a continuous mapping on $\Omega^-(\xi)$ to itself, which by (2) preserves P-measure.

This one-sided S.P. $X^-(\xi)$ is then shown to extend uniquely to a two-sided stationary S.P. $X(\xi) = (x_n', \ -\infty < n < \infty)$. Firstly, it is proved that the E-algebra $A^-(\xi) \simeq C(\Omega^-(\xi))$ can be embedded in a larger E-algebra $A(\xi)$, so that the endomorphism T extends to an automorphism T on $A(\xi)$, and for $\zeta' \in A(\xi)$, $E(T\zeta') = E(\zeta')$, this E

[7] The author speaks of an E-algebra A with an E-preserving endomorphism T as an *abstract process*. He calls A *two-sided* if T is an automorphism; otherwise *one-sided* (p. 15).

being an extension of the original E (Th. 2.1, p. 16).[8] Now $A(\xi)\simeq C(\Omega(\xi))$; hence from E, T we get as before a probability measure P over $\Omega(\xi)$ and a P-measure preserving homeomorphism T on $\Omega(\xi)$. Since $A^-(\xi)\subseteq A(\xi)$, every homomorphism ω' of $A(\xi)$ contracts to a homomorphism $\omega=\beta(\omega')$ of $A^-(\xi)$. We thus get a "*canonical mapping*" β on $\Omega(\xi)$ onto $\Omega^-(\xi)$ (p. 17). Now for $\omega'\in\Omega(\xi)$ let

$$x_n'(\omega') = x_n\{\beta(\omega')\},\ n \leqq 0; \qquad x_n'(\omega') = x_0(T^{-n}\omega'),\ n \geqq 0.$$

Then $X(\xi)=(x_n',\ -\infty<n<\infty)$ is the desired two-sided S.P. corresponding to ξ. Obviously $x_{n-1}'=x_n' T$, $-\infty<n<\infty$.

D. **A regular ξ as a typical time-sequence of $X^-(\xi)$.** Let $\xi\in\Lambda_\infty^-$ be stochastic, and $\zeta\in A_0(\xi)$. We recall, cf. (1) in §A, that to ζ corresponds a function $\psi\in C(\Lambda_\infty^-)$. On the other hand, on identifying $\zeta_1, \zeta_2\in A_0(\xi)$ for which $\zeta_1-\zeta_2\in N$, $A_0(\xi)$ becomes an everywhere dense subset of $A^-(\xi)\simeq C(\Omega^-(\xi))$. It is shown that these facts imply that $\Omega^-(\xi)$ *can be imbedded in* Λ_∞^-; in fact after making suitable identifications we can write $\Omega^-(\xi)\subseteq\Lambda_\infty^-$ (pp. 18, 19, 31). In case ξ is not only stochastic but *regular*, it is shown that $\xi\in\Omega^-(\xi)$ (Th. 5.1, pp. 33, 34). We can therefore write $\xi(n)=x_n(\xi)$, $n\leqq 0$, and regard ξ as a time-sequence of the associated one-sided S.P. $X^-(\xi)$. Moreover, ξ is a *typical time-sequence* of $X^-(\xi)$, or in the author's terminology, a *generic point* of $A^-(\xi)$, i.e.

$$\lim_{N\to\infty}\frac{1}{N+1}\sum_{n=0}^{N}\zeta(T^n\xi) = E(\zeta), \qquad \zeta\in A^-(\xi).$$

The notion of "generic point," defined more generally as follows, is quite important:

5. Definition (p. 38). *Let $A\simeq C(\Omega_A)$ be an E-algebra with an endomorphism T such that $E(Tf)=E(f)$, $f\in A$,*[9] *and let P be the probability measure on Ω_A induced by E, so that $E(f)=\int_{\Omega_A}f(\omega)P(d\omega)$. We call $\omega\in\Omega_A$ a generic point of A, if for all $f\in A$,*

$$E(f) = \lim_{N\to\infty}\frac{1}{N+1}\sum_{n=0}^{N}f(T^n\omega).$$

Roughly speaking this means that for any $f\in A$ the time-average of the sequence $(f(T^n\omega),\ n\geqq 0)$ equals the phase average of f.

The author proves (p. 57) that for a regular $\xi\in\Lambda_\infty^-$, $\tilde{\xi}$ is an L-extension of ξ, if and only if $\tilde{\xi}\in\Omega(\xi)$. It follows that *the set of all L-extensions of ξ is precisely the subset* $\beta^{-1}\{\xi\}$ of $\Omega(\xi)$, β being the canonical map-

[8] Actually the author shows that to every one-sided abstract process A corresponds a unique two-sided abstract process B. He calls B *the two-sided version of* A.

[9] I.e. A is an abstract process, cf. footnote 7.

ping from $\Omega(\xi)$ to $\Omega^-(\xi)$. Thus the measure μ referred to in the prediction problem in §B has to have the carrier $\beta^{-1}\{\xi\}$.

E. **Prediction measure and continuous predictability.** Our formulation of the Prediction Problem in §B for a regular sequence $\xi\in\Lambda_\infty^-$ is heuristic in that it merely indicates what the desired measure μ has to accomplish. In the light of the preceding investigation our requirements on μ can be expressed in a more mathematical way: μ has to behave like a conditional probability measure derived from the P-measure on $\Omega(\xi)$, relative to the hypothesis that the event ξ has occurred (p. 63). The author effects the explication of this still vague requirement in terms of the important notion of a *prediction measure*:

6. DEFINITION.[10] *Let* $A\simeq C(\Omega_A)$ *be an E-subalgebra*[11] *of the E-algebra* $B\simeq C(\Omega_B)$.

(a) *For* $g\in B$ *the* conditional expectation $E(g|A)$ *is the unique, bounded measurable function* ψ *on* Ω_A *such that* $E(fg)=E(f\psi)$, *for all* $f\in A$.[12]

(b) *A probability measure* μ *on* Ω_B *is called a* prediction measure *at* $\xi\in\Omega_A$, *if for all* $g\in B$,

$$E(g\mid A)\geqq 0 \textit{ in a nhbd. of } \xi \Rightarrow E_\mu(g)=\int_{\Omega_B} g(\omega)P(d\omega)\geqq 0.$$

The existence of such a prediction measure μ at each $\xi\in\Omega_A$, and the fact that the support of μ is contained in the set $\beta^{-1}\{\xi\}$, where β is the canonical mapping from Ω_B to Ω_A, are established (Cor., p. 68; 10.1, p. 64). The important question as to whether μ is unique is then considered:

7. DEFINITION (p. 64). (a) *Let* A, B *be as in Def.* 6. A *is* continuously predictable (c.p.) *to* B *at* $\xi\in\Omega_A$, *if there is just one prediction measure* μ *at* ξ.

(b) *A regular sequence* $\xi\in\Lambda_\infty^-$ *is c.p., if* $A^-(\xi)$ *is c.p. to* $A(\xi)$ *at* ξ.

The author shows (p. 69) that A is c.p. to B at ξ, if and only if for every $g\in B$, $E(g\mid A)$ is "continuous" at ξ in a certain sense which we shall not stop to explain (p. 65). This is the reason for the term "continuously predictable." Also, A is c.p. to B at *each* $\xi\in\Omega_A$, if and only if $E(g\mid A)\in A$ for all $g\in B$. In this case if μ_ξ is the unique prediction measure at ξ, then

$$E_{\mu_\xi}(g)=E(g\mid A)(\xi), \qquad \text{for almost all } \xi\in\Omega_A.$$

Several criteria are given for A to be c.p. to B at ξ. The most basic

[10] Once again our definition is an adaptation of the author's (pp. 63, 64), convenient for purposes of this review.

[11] I.e. the functional E for A is the restriction of the functional E for B.

[12] The existence and uniqueness of ψ are easy to establish (p. 13).

of these is stated in terms of the notion of a *predicting sequence at* ξ, i.e. a sequence $(f_n, n \geqq 1)$ such that each $f_n \in A$ and for each $g \in A$, $E(f_n g) \to g(\xi)$, as $n \to \infty$. A necessary and sufficient condition for A to be c.p. to B at ξ is that for every predicting sequence $(f_n, n \geqq 1)$ at ξ and every $g \in B$, $E(f_n g) \to F(g)$, where F is a linear functional on B, F turns out to be E_{μ_ξ} (Th. 10.1, Cor. 2, pp. 68, 69).

No nice direct criteria are given for a regular sequence $\xi \in \Lambda_\infty^-$ to be c.p. If ξ is deterministic, i.e. $\beta^{-1}\{\xi\}$ is a unit set, then obviously ξ is c.p. The only other c.p. sequences discussed by the author are typical time-sequences of standard stochastic processes. A regular sequence $\xi \in \Lambda_\infty^-$ is called a *random*, *Markoff* or *m-Markoff sequence*, according as the associated S.P. $X(\xi)$ is random,[13] Markoff or m-Markoff (p. 75). The author shows that a "random sequence" is c.p., and that so is a m-Markoff sequence with denumerable state space Λ. This is not the case when the state space is non-denumerable. Actually, the class of c.p. sequences turns out to be rather fragile. For instance, *a derived sequence of a c.p. sequence need not be c.p.* The author gives a simple but surprising example of such a sequence (§12, p. 78):

8. Example. Let ξ be the sequence of independent tosses of a biassed coin in which the experimenter records whether each toss gives the same or different outcome from the previous toss. Then ξ is not c.p., although it is derived from a "random" and therefore c.p. sequence η. More formally, letting $\Lambda = \{-1, 1\}$ we take $\eta \in \Lambda_\infty^-$ to be a regular sequence such that $X(\eta)$ is made up of independent random functions y_n, $-\infty < n < \infty$ for which $P(y_n = -1) = p$, $P(y_n = 1) = 1 - p$, where $0 < p < 1$, $p \neq 1/2$; we take $\xi(n) = \eta(n) \cdot \eta(n-1)$.

The author devotes a good deal of effort to finding conditions which will exclude the sort of pathology exemplified in Example 8. Ch. 4 is devoted to finding conditions under which a derived sequence of a Markoff sequence will be c.p. We do not have the space to report on this part of his work in any detail.

F. **Statistical predictability.** Consider now a regular sequence $\xi \in \Lambda_\infty^-$ which is not c.p., i.e. for which $A^-(\xi)$ is not c.p. to $A(\xi)$ at ξ. There is then more than one prediction measure μ at ξ. Are there non-c.p. sequences $\xi \in \Lambda_\infty^-$ for which one can single out one of these measures, say μ_0, as being more appropriate statistically than the others? If so, we could speak of μ_0 as being the "correct" prediction measure at ξ, and speak of the non-c.p. sequence ξ as being "statistically predictable" (s.p.).

[13] Unfortunately, the author uses the term "random process" to mean a process of *independent* random functions.

The author investigates this question in a slightly wider context. Let $A \simeq C(\Omega_B)$ be an E-subalgebra of the E-algebra $B \simeq C(\Omega_B)$, and let both have E-preserving endomorphisms denoted by T.[14] Let

$L(A)$ = *the E-algebra of all complex-valued bounded, measurable functions on* Ω_A,

$C = \{f: f = E(g|A),\ g \in B\}$,

$\tilde{A}$ = *the intersection of all E-algebras K such that*

$$A, C \subseteq K \subseteq L(A) \ \&\ T(K) \subseteq K.$$

The author shows that $\tilde{A}$ is an E-algebra with an E-preserving homomorphism T. He calls $\tilde{A}$ the *c.p. cover of A with respect to B* (p. 131).

Now let $\xi \in \Lambda_\infty^-$ be a regular but not c.p. sequence. Let $\tilde{A}^-(\xi)$ be the c.p. cover of $A^-(\xi)$ with respect to $A(\xi)$. As before (§C) we can imbed $\tilde{A}^-(\xi)$ in a (unique) algebra $\tilde{A}(\xi) \simeq C(\tilde{\Omega}(\xi))$ over which the extension of T is an E-preserving automorphism. This "two-sided" $\tilde{A}(\xi)$ is called *the c.p. cover of* ξ.

In general $\tilde{A}(\xi) \not\subseteq A(\xi)$; we define $(\tilde{A}(\xi), A(\xi))$ to be the smallest E-algebra containing both $\tilde{A}(\xi)$ and $A(\xi)$. The author shows that invariably $\tilde{A}(\xi)$ *is c.p. to* $(\tilde{A}(\xi), A(\xi))$ *at each* $\tilde{\xi} \in \tilde{\Omega}(\xi)$.[15] It follows that for each $\tilde{\xi} \in \tilde{\Omega}(\xi)$ there exists a unique prediction measure $\mu_{\tilde{\xi}}$ at $\tilde{\xi}$, this measure being over a subset of the space Ω such that $C(\Omega) \simeq (\tilde{A}(\xi), A(\xi))$. Since $A(\xi) \subseteq (\tilde{A}(\xi), A(\xi))$, this measure $\mu_{\tilde{\xi}}$ induces a measure on $\Omega(\xi)$ via the canonical mapping on Ω onto $\Omega(\xi)$. Thus *to each* $\tilde{\xi} \in \tilde{\Omega}(\xi)$ *corresponds a measure* $\mu_{\tilde{\xi}}$ *on* $\Omega(\xi)$.

Finally, since $A^-(\xi) \subseteq \tilde{A}(\xi)$, there is a canonical mapping β on $\tilde{\Omega}(\xi)$ onto $\Omega^-(\xi)$. Now suppose that the set $\beta^{-1}\{\xi\}$ has exactly one member $\tilde{\xi}$ which is a generic point for $\tilde{A}(\xi)$ (Def. 5 above). We can then associate with our $\xi \in \Lambda_\infty^-$ the measure $\mu_{\tilde{\xi}}$ on $\Omega(\xi)$. Such considerations lead the author to make the following definition (pp. 133–136, Defs. 22.1, 22.2):

9. Definition. *The regular sequence* $\xi \in \Lambda_\infty^-$ *is called* statistically predictable (s.p.),[16] *if* ξ *has exactly one extension* $\tilde{\xi} \in \tilde{\Omega}(\xi)$, *which is a generic point for the c.p. cover* $\tilde{A}(\xi) \simeq C(\tilde{\Omega}(\xi))$. *In this case the measure* $\mu_{\tilde{\xi}}$ *over* $\Omega(\xi)$ *is called* the determined prediction measure *at* ξ.

This concept of predictability is broader than that of continuous predictability as the following example shows.

10. Example. Let Λ, η, ξ be as in Example 8. We find that $\tilde{A}^-(\xi) = A^-(\eta)$ and so $\tilde{A}(\xi) = A(\eta)$. Thus $A(\eta)$ is the c.p. cover of the non-c.p. sequence ξ. Moreover, it can be shown that ξ has exactly one

[14] I.e. A, B are *abstract processes*, cf. footnote 7.

[15] Cf. proof of 21.1 on p. 131.

[16] Rather misleading terminology, since in the literature the term "statistical prediction" is used to refer to the classical concept.

extension to $\Omega(\eta)$ which is a generic point for the cover $A(\eta)$, viz. η itself. Thus ξ is s.p. and μ_η, confined to $\Omega(\xi)$, is the *determined* prediction measure at ξ.

A regular sequence $\xi \in \Lambda_\infty^-$ can fail to be s.p. either because it has no extension which is a generic point for the cover $\tilde{A}(\xi)$, or because it has more than one such extension. The latter possibility complicates the relationship between continuous and statistical predictability. Thus from the author's theorem (21.1, p. 131) that $\tilde{A}^-(\xi)$ is the minimal algebra A such that $A \subseteq L(\Omega^-(\xi))$ and A is c.p. to $(A, A(\xi))$, it follows that if ξ is c.p., we have $\tilde{A}^-(\xi) = A^-(\xi)$ and so $\tilde{A}(\xi) = A(\xi)$. But ξ may have more than one extension $\tilde{\xi}$ which is a generic point for the cover $A(\xi)$. Thus *a c.p. sequence need not be s.p.* Examples of this are provided by certain Markoff sequences with non-denumerable state space (p. 137).

The remainder of the author's investigation in prediction theory proper is devoted to showing that *if the sequence ξ is derived from a Markoff sequence η, and ξ, η are finite-valued, then ξ is s.p.* This is the last theorem in the book (p. 282). As such sequences are not particularly important for prediction, and the proof of the theorem is both long and difficult and involves ideas on which we shall comment in III, we shall here terminate our review of the author's prediction theory of time-sequences.

III. Ramifications

In II, although our interest was in predicting the future of a one-sided sequence, we followed the author in stating many definitions for general E-algebras endowed with E-preserving endomorphisms. This was done in accord with his view that *prediction from past to future should be treated as a special case of prediction from one E-algebra to another.* It turns out that important analytical and probabilistic problems fall within the framework of this wider outlook on prediction. An important instance is the study of sub-Markoff processes,[17] which occupies a good deal of the book (Chs. 3, 4, 8, 9). While such processes are uninteresting from the standpoint of pure prediction, their investigation fits well into the author's wider conception of the subject, and has in fact led him to the fruitful ideas of a *stochastic semi-group* (Ch. 5) and of an *inductive function* (Ch. 7). As regards the former concept we shall only say that to every stochastic semi-group S corresponds a stationary S.P. X, and vice versa, and that

[17] (y_n) is called a *sub-process* of the one-sided or two-sided process (x_n), if, for each admissible n, $y_n = \psi(\cdots, x_{n-1}, x_n)$ where ψ is a continuous function on Λ_∞^- to a compact metric space Λ' (p. 20).

when S consists of linear transformations the corresponding X is Markoff or sub-Markoff (Ch. 5). We now turn to the discussion of inductive functions.

With the aid of his theory of stochastic semi-groups the author proves (24.2, p. 149) that for a finitely-valued subprocess $X=(x_n, -\infty<n<\infty)$ of a m-Markoff process the c.p. cover $\tilde{X}$[18] is a composite (X, Z), i.e. the algebra $\tilde{X}$ is generated by functions of the form $\tilde{x}_n=(x_n, z_n)$, where $Z=(z_n, -\infty<n<\infty)$ is a process satisfying an equation of the form

$$z_n = \psi(x_n, z_{n-1}) \text{ a.e.} \tag{1}$$

This leads him to the study of such functional equations and the associated equation for sequences ξ, ζ:

$$\zeta(n) = \psi\{\xi(n), \zeta(n-1)\}. \tag{1'}$$

The author calls a Λ-valued process Z *an inductive function of* X, if X, Z are subprocesses of some process, Z satisfies equation (1) and ψ is continuous. Similarly he speaks of the Λ-valued sequence ζ being an *inductive function of* ξ when (1′) is satisfied (pp. 153, 154).

Let ψ and X be given and ξ be a generic sequence of X.[19] The author is interested in knowing if the inductive function ζ will be generic for the inductive function Z. He raises two questions:

Q_1: *Given a solution Z of* (1), *will it have a generic sequence ζ which satisfies* (1′)?

Q_2: *Will every solution ζ of* (1′) *be a generic sequence of a solution Z of* (1)?

These questions arise naturally within the wider context of prediction just referred to above.

The author first provides answers when ψ has simple forms (Ch. 7). For instance, for the special case

$$z_n = x_n + z_{n-1}, \qquad \zeta(n) = \xi(n) + \zeta(n-1)$$

he proves (p. 164) that Q_1 has an affirmative answer if X is topologically ergodic.[20] For the more difficult case in which Λ is a compact Abelian group G and

$$z_n = x_n \cdot z_{n-1}, \qquad \zeta(n) = \xi(n)\cdot\zeta(n-1),$$

[18] The c.p. cover $\tilde{X}$ of X is defined as follows. Let $A_{\bar{X}}$, A_X be the algebras generated by all functions $\psi(x_n)$ where $\psi\in C(\Lambda)$, and $n\leq 0$, $-\infty<n<\infty$, respectively. Then $\tilde{X}$ is the two-sided version (cf. footnote 8) of the c.p. cover $\tilde{A}_{\bar{X}}$ of $A_{\bar{X}}$ with respect to A_X (cf. Def. in §F).

[19] I.e. let ξ be a generic point of the algebra $A_X\simeq C(\Omega_X)$.

[20] I.e. if for all closed sets $\Delta\subseteq\Omega_X$ for which $T(\Delta)\subseteq\Lambda$, we have $P(\Delta)=0$ or 1.

he shows that if X is ergodic, "λ-ergodic" for every rational λ,[21] and "equidistributed in G,"[22] then (1) has a unique solution Z, and Q_2 has an affirmative answer (p. 166).

From the last result the author deduces Weyl's important theorem that if p is a polynomial with real coefficients at least one of which is irrational, then the points $\xi(n)=e^{2\pi i p(n)}$, $n \geqq 0$, are "equidistributed" on the circle $[|z|=1]$ (p. 171). In a similar vein he proves a general theorem (27.1, p. 175), from which emerges the theorem of Wiener and Wintner that if $f \in L_1(\Omega, B, P)$, T is ergodic and measure-preserving on Ω, then for almost all $\omega \in \Omega$, the limits

$$C_\lambda = \lim_{N\to\infty} \frac{1}{N+1} \sum_{n=0}^{N} f(T^n\omega) e^{-2\pi i n\lambda}$$

exist for *all* λ (p. 179).

The author is also able to answer the questions Q_1, Q_2 for a wide class of ψ by placing restrictions on X (Ch. 8). An important result of this sort is as follows (29.2, p. 202): if X is a finitely valued Markoff process and ψ is a "compact mapping,"[23] then Q_1 has an affirmative answer; as for Q_2, every solution ζ of (1′) is a generic sequence of some solution Z of (1). In the course of proving this, the writer establishes the interesting result (28.1, p. 183) that every ergodic, finitely-valued Markoff process is an inductive function of a "random process."[24] He also gets an interesting law of large numbers, viz. if $W=(w_k,\ k \geqq 1)$ is a finitely valued (non-stationary) Markoff process, then almost every time-sequence of W is a generating sequence (28.2, p. 187).

An immediate corollary of these results is the random ergodic theorem of von Neumann and Ulam, restricted to a finite number of measure-preserving transformations. The simplest version of this asserts that if T, T' are measure-preserving transformations on $(\Lambda, \mathcal{B}, P)$, $X=(x_n)_{n=0}^{\infty}$ is a "random process" with values T, T', and $f \in C(\Lambda)$, then for almost all time-sequences ξ of X, the sequence

$$(1) \qquad f(\lambda),\ f\{\xi(1)\lambda\},\ f\{\xi(2)\xi(1)\lambda\},\ \cdots,\ \lambda \in \Lambda$$

has an average. The proof rests on noting that X is a finitely valued Markoff process, and that therefore almost all time-sequences are

[21] I.e. $z \in L(\Omega_X)$ and $T(z)=e^{2\pi i\lambda}z$ implies $z=0$ (p. 21).

[22] I.e. for each $\sigma \in G$, the S.P.'s (x_n), (σx_n), $-\infty < n < \infty$ have the same joint distributions.

[23] I.e. for each π, the mapping $\psi(\pi, \cdot)$ on Λ to Λ is a contraction.

[24] Cf. footnote 13.

generic, and that what is involved in (1) is an inductive function ζ of ξ, viz.

$$\zeta(n) = \xi(n)\xi(n-1)\cdots\xi(1)\lambda = \xi(n)\cdot\zeta(n-1);$$

hence ζ is a generic and therefore stochastic sequence. This remarkable proof sheds new light on the theorem by showing that what makes it work is the inductive character of the underlying function.

IV. Concluding remarks

The limitations of the author's prediction theory for individual time-sequences should be evident. Finite-valued sub-Markoff sequences are of no particular interest for prediction, and so the author's final theorem (§F, above) is not very exciting. It is not yet clear if further work with the author's concept of "statistical prediction" will yield anything worthwhile in the actual analysis of the time-sequences which arise in science. As these limitations could not possibly have been foreseen, the author has of course rendered yeoman service to the subject by venturing on this new frontier.

The wider concept of prediction advocated by the author has already proved fruitful in revealing the essential core of certain deep theorems in analysis and probability. But it remains to be seen if significant tracts of mathematical territory can be illuminated by the development of his ideas and techniques. The reviewer would certainly hope that this may be possible. Recently it has become clear that many linear prediction problems are special (deficiency 1) cases of other quite important problems in functional analysis, and that linear prediction techniques extend to the latter. It would be very satisfying indeed, if the same situation were found to prevail at the non-linear level.

Reading the book is not an easy job. The subject is itself rather hard. But part of the difficulty stems from the peculiar organization of the book, which is very different from that of this review, for instance. Idiosyncrasies in its format also make for hard reading, especially the absence of an index and misuse of the decimal system of enumeration. (Theorem 10.2 occurs in §10.4!) But the work stands as a first-rate and highly original dissertation on a very difficult subject.

P. Masani

Strukturtheorie der Wahrscheinlichkeitsfelder und -räume. By D. A. Kappos. Ergebnisse der Mathematik, und Grenzgebiete, Heft 24. Springer, Berlin, 1960. 4+136 pp. DM 21.80.

This book is the first to be devoted to a systematic account of the applications of Boolean algebras to measure theory. The direct con-