Stochastic processes. By J. L. Doob. New York, Wiley, 1953. 8+654 pp. $10.00.

Chapter 1. Introduction and probability background (45 pages). The fundamental concepts of conditional probability and expectation are generalized over and somewhat different from those defined in Kolmogorov's *Grundbergriffe* (1933). Let y be a random variable whose expectation exists: let $\mathcal{J}$ be a Borel field of measurable ω sets and $\mathcal{J}'$ its "completion." The conditional expectation of y relative to $\mathcal{J}$, denoted by $E\{y|\mathcal{J}\}$, is any measurable $\mathcal{J}'$, integrable ω function which satisfies $\int_\Lambda E\{y|\mathcal{J}\}dP=\int_\Lambda ydP$ for every $\Lambda\in\mathcal{J}$. If $\mathcal{J}$ is the smallest Borel field $\mathcal{B}(x_t, t\in T)$ with respect to which the x_t's are measurable, the above definition specializes to the conditional expectation of y with respect to the x_t's. The conditional probability of a measurable set M, denoted by $P(M|\mathcal{J})$, is defined to be $E\{y|\mathcal{J}\}$ where $y(\omega)$ is the characteristic function of M. An important question arises: Is it possible to define an M, ω function $P(M, \omega)$ such that (i) for every ω, $P(\cdot, \omega)$ is a probability measure of M and for every M, $P(M, \cdot)$ is measurable $\mathcal{J}'$; (ii) for every M, $P(M, \cdot)=P\{M|\mathcal{J}\}$ with probability one. If such a $P(\cdot, \cdot)$ is defined for every $M \in\mathcal{B}(y_1, \cdots, y_n)$, it is called the conditional probability distribution of the y_j's relative to $\mathcal{J}$. A related question is as follows. Let Y denote a generic n-dimensional Borel set. Is it possible to define a Y, ω function $p(Y, \omega)$ such that (i) for every ω, $p(\cdot, \omega)$ is a probability measure of Y and for every Y, $p(Y, \cdot)$ is measurable $\mathcal{J}'$; (ii) for every Y, $p(Y, \omega)=P\{[y_1(\omega), \cdots, y_n(\omega)]\in Y\}$ with probability one. If such a $p(\cdot, \cdot)$ exists, it is called the conditional probability distribution of $y_1, \cdots, y_n$ in the wide sense relative to $\mathcal{J}$. Now the main result is: while a conditional probability distribution in the wide sense always exists, a conditional probability distribution may fail to exist. The point is that an ω set $M\in\mathcal{B}(y_1, \cdots, y_n)$ does not uniquely determine the Borel set Y and if $M=\{[y_1(\omega), \cdots, y_n(\omega)]\in Y\}$ is satisfied for $Y=Y_1$ and also for $Y=Y_2$ it does not necessarily follow that $p(Y_1, \omega)=p(Y_2, \omega)$ with probability one. A sufficient condition for this, and so for the existence of a $P(M, \omega)$, is e.g. that the range of $[y_1(\omega), \cdots, y_n(\omega)]$ be a Borel set; this condition is always satisfied if the y process is of "function space type" (see below). The lack of a conditional probability distribution however vitiates a couple of the author's more famous theorems (1938) (see the Bibliography for all references). A correct form of the extension theorem, due to Ionescu Tulcea, where the existence of conditional probability distributions is assumed, is given in the Appendix. Kolmogorov's version is not given explicitly but is implied by this theorem and the sufficient condition stated above. The lack of a conditional probability distribution also necessitates a detour in proving certain results involving conditional expectations. One way is to use "representation theory." This theory enables us, in the study of the family of (real) random variables x_t, $t \in T$, to replace the original basic space Ω by the space (of function space type) $\tilde{\Omega}$ of all (real) functions of $t \in T$. The point $\tilde{\omega}$ in $\tilde{\Omega}$ ranges over all functions defined on T, and $x_t(\omega)$ is mapped into $\tilde{x}_t(\tilde{\omega})$ which for $\tilde{\omega} = \xi(\cdot)$ has the value $\xi(t)$. Thus $\tilde{x}_t(\tilde{\omega})$ is a coordinate random variable in Ω, namely "the tth coordinate of the point ω." It is shown (in the Supplement) that such a measure-preserving transformation can be established, and we can therefore translate any problem involving ω random variables into the corresponding one involving $\tilde{\omega}$ random variables. In the simplest case this amounts to e.g. considering two random variables $x(\omega)$ and $y(\omega)$ as the two coordinate random variables (of the basic point $\tilde{\omega} = (x, y)$) in the Cartesian plane, thus substituting a 2-dimensional distribution function for an abstract probability measure. Such an approach was in fact commonly used before the advent of the basic space Ω, and has been tried again in a way by Cramér in his *Mathematical methods of statistics* (1945). Here it is treated as a device rather than the logical foundation.

Chapter 2. Definition of a stochastic process—Principal classes (56 pages). A stochastic process is defined as any family of random variables x_t, $t \in T$. Let $\mathcal{F}_T = \mathcal{B}(x_t, t \in T)$ and $\mathcal{F}'_T$ be its completion. If T is nondenumerable, many significant ω sets, e.g. $\sup_{t \in T} x_t(\omega) > \lambda$, in general do not belong to $\mathcal{F}'_T$. This circumstance was noted e.g. in Khintchine's *Asymptotische Gesetze* (1933) and called for a new foundation of the theory of stochastic processes beyond that given in Kolmogorov's Grundbegriffe. The author first gave such a theory in 1937. He took Ω to be the space of all functions of $t \in T$ and his method was to enlarge Ω by "adjoining" a set of outer measure one. This method is only briefly reviewed in the present book. The new method, first presented here, does not restrict Ω nor change $\mathcal{F}_T$; instead the x_t's are modified in such a way as to leave unaltered the probability of sets in $\mathcal{F}_T$ and yet to acquire the desirable property of "separability." Let $\mathcal{A}$ be a system of linear Borel sets. The stochastic process is said to be separable relative to $\mathcal{A}$ if there is a sequence t_j of parameter values and an ω set Λ of probability zero such that if $A \in \mathcal{A}$ and if I is any open interval, the ω sets $\{x_t \in A, t \in IT\}$ and $\{x_{t_j}(\omega) \in A, t_j \in IT\}$ differ by a subset of Λ. Two important cases are where $\mathcal{A}$ is the set of all closed sets and where $\mathcal{A}$ is the set of all closed intervals; in the latter case the stochastic process is simply said to be separable. Next, $\tilde{x}_t$ is called a standard modification of x_t if for each t, the ω set $\{\tilde{x}_t(\omega) \neq x_t(\omega)\}$ belongs to $\mathcal{F}'_T$ and has probability zero. Now the fundamental theorem (Theorem 2.4) is: Any stochastic process $\tilde{x}_t$ has a standard modification x_t which is separable relative to the set of closed sets. (The x_t's may take on the values $\pm\infty$.) This new approach deprives the theory of some of its erstwhile measure-theoretic halo. Thus, the proof of the above theorem depends on the clean-cut lemma: To each linear Borel set A there corresponds a finite or denumerable sequence t_j such that $P\{x_{t_n}(\omega)\in A,\ n\geqq 1;\ x_t(\omega)\notin A\}=0$ for each $t\in T$. Next the stochastic process x_t is called measurable if T is Lebesgue measurable, and $x_t(\omega)$ considered as a (t, ω) function is measurable in the product (t, ω) measure. A simple sufficient condition is given for the measurability of a separable stochastic process. For a separable measurable stochastic process the probabilities of significant ω sets such as the one mentioned above are defined and the integral $\int x_t(\omega)dt$ may be interpreted as ordinary Lebesgue integrals of the sample functions, etc. In essence separability reduces the considerations of sample functions as to their boundedness, continuity, etc. to those of their values on a denumerable set, and this will be repeatedly made use of throughout the book. The second half of Chapter 2 defines various stochastic processes to be discussed later together with their preliminary properties. Only one, the Gaussian process, is not discussed later for lack of further knowledge. It is used to define "strict sense" and "wide sense" concepts. If a stochastic process has a certain property P in terms of means and covariances and if the corresponding Gaussian process with the same means and covariances has the corresponding but stronger property P′, then P′ and P are the corresponding strict and wide sense properties.

Chapter 3. Processes with mutually independent random variables (46 pages). This chapter belongs to the classical theory of probability and treats such familiar topics as the 0–1 law, convergence of series, the law of large numbers, infinitely divisible distributions, and the central limit theorem. The material is presented frequently with a view to later applications. The criterion for the convergence of a series of independent random variables is given not only in the Khintchine-Kolmogorov form known as the three-series theorem but also in Lévy's form based on the idea of centering. In particular, convergence regardless of the order of terms is discussed, as will be needed later for processes with independent increments. Further criteria are given in terms of infinite products of characteristic functions, due to Wintner and others. The treatment of infinitely divisible distributions is direct and deals immediately with components which are infinitesimal but not necessarily equi-distributed. The central limit theorem due to Feller and Lévy is given in "finite terms," in the spirit of Lévy. The crisp treatment of the last two topics is made possible by a deft use of certain inequalities which are implicit in previous work but carefully sifted and dispatched in Chapter 2. The chapter ends with the author's youthful (1936) discovery to the effect that if a gambler chooses his play (in a stationary sequence of independent games) without clairvoyance, his chances are unaffected. The formal proofs of such "obvious" statements seem to be a source of the author's inspirations.

Chapter 4. Processes with mutually uncorrelated or orthogonal random variables (22 pages). There is a brief discussion of orthogonal series leading to Menshov's theorem. It is interesting to note that Menshov's condition for the strong law of large numbers of orthogonal random variables is within a factor of $\log^2 n$ of Kolmogorov's for that of independent random variables. However, it should be pointed out that Menshov's condition is not sufficient for the convergence of the corresponding infinite series, as Kolmogorov's is. This can be shown by an example based on Menshov's own counterexample.

Chapter 5. Markov processes—Discrete parameter (65 pages). The author defines a Markov chain to be a Markov process (discrete or continuous parameter) whose random variables assume values in a finite or denumerable set. Finite chains with stationary transition probabilities are first discussed. The ergodic properties are proved by studying the actual transitions, leading from various special cases to the general case. The choice of this method seems largely dictated by the generalization to the "general state space" which follows. This generalization is an exposition of Doeblin's Thesis (1937), fortified by measure theory, and strengthened and completed in important points. Here, as nowhere else in the book, the random variables assume values in an abstract space. (It is a tribute to the author's sense and conscience that he does not indulge in trivial generalizations.) Under an essential hypothesis (D), slightly generalized over Doeblin's, the decomposition of the space into a transient set and a finite number of ergodic sets each containing (possibly) cyclically moving subsets, and the convergence of the transition probabilities $p^{(n)}(\xi, E)$, uniformly in ξ and exponentially fast, are established. The main line of argument is an extension of the one used in the finite case, but it is considerably complicated by several measure-theoretic tricks and asides. While the tricks are Doeblin's, the asides, no less tricky to the average reader, are Doob's. Furthermore, the author adds a clarifying discussion of the hypothesis (D). The details of all this are awesome and smack of the *ad hoc*. It may be wondered if there is a better approach. Doeblin in his last great paper on Markov processes (1940; listed in the book but not used) indeed treated a more general case[1] where (D) is not postulated and used an apparently more powerful method. However, to present this paper in a shape comparable to what is given here will undoubtedly require even more space and labor, and further work will be necessary to adduce a special case like that under (D). (The reviewer has attempted to do this, but is still far short of the goal.) The condition (D) is even sufficient for a central limit theorem (within a noncyclic ergodic set) for $\sum_1^n f(x_m)$ where $\{x_n,\ n\geqq 1\}$ are the random variables of the Markov process, provided that $E[|f(x_1)|^{2+\delta}]<\infty$ for some positive δ (Doeblin postulated a bounded f.) The proof of this old-fashioned theorem, in which the author does all the antics of a computationist, is one of the longest (11 pages) in the book. It depends on the exponential speed of the convergence of $p^{(n)}$, and S. Bernstein's idea of grouping terms. It may be noted that the denumerable case is not treated separately but only by way of example under (D). As is well known there is a unified treatment of the finite and denumerable case, given by Kolmogorov in 1936 and now available in a somewhat different version in Feller's book. This treatment does not by itself yield the exponential speed needed here for the central limit theorem, but in the denumerable case, there is a much simpler method (see Doeblin, Bull. Soc. Math. France vol. 66 (1938) pp. 210–220).[2]

Chapter 6. Markov processes.—Continuous parameter (57 pages). Finite chains are treated first and a complete analysis of the behavior of the (stationary) transition probabilities $p_{ij}(t)$ at $t=0$ and $t=\infty$ is given. A description of the actual transitions follows, resulting in the theorem that the sample functions of a separable chain satisfying $\lim_{t\to 0} p_{ij}(t)=\delta_{ij}$ are almost all step functions. Next it is shown how to construct a chain with given $p'_{ii}\ (0+)$ and $p'_{ij}\ (0+)$. In this chapter we miss more acutely the absence of a detailed discussion of the denumerable case. The fact is that in the continuous parameter case this already offers a great challenge. The recent work of Lévy (1951) shows how much can be done, and how much remains to be done. A complete description of the sample functions becomes

[1] In a recent French book the authors stated to the effect that this extension presents no "major difficulties" and that it is only short of being "absolutely immediate." Anyone who reads the paper will see that these are gross understatements.

[2] Cf. a forthcoming paper by the reviewer, *Contributions to the theory of Markov chains*. II (to appear in Trans. Amer. Math. Soc.)

infinitely harder, but even a discussion of the simpler types along the lines of the author's 1942 and 1945 papers would have been valuable. In this connection it may be pointed out that a recent paper by Kolmogorov (Učenye Zapiski (Matem.), Moskov. Gov. Univ. (4) vol. 148, pp. 53–59) completes certain points of the behavior of $p_{ij}(t)$ at $t=0$ (cf. Ex. 1, pp. 265 and 271 of the book); also the behavior of $p_{ij}(t)$ at $t=\infty$ is established elegantly by Lévy (1951), both for the denumerable case. The generalization to a continuous state space (a linear Borel set) is accomplished by Doeblin's method (1939), similar to the one used in the finite case. The material in §3 is the author's version of the Ito process, and is of very recent date. It stems from the partial differential equations for the not necessarily stationary transition probabilities of Markov processes established by Kolmogorov in 1931 and studied by Feller (1936; 1945) and others. Ito was the first (1946; 1951) to show that these processes could be obtained constructively by solving the stochastic differential equation (*) $dx(t)=m[t, x(t)]dt+\sigma[t, x(t)]dy(t)$ where $y(t)$ is the Brownian motion process with variance parameter one. This equation is interpreted to mean $x(t)-x(a)=\int_a^t m[s, x(s)]ds+\int_a^t \sigma[s, x(s)]$ where the second integral (due to Ito) is defined in Chapter 9 of the book. It is shown that under reasonable assumptions on m and σ^2 a separable Markov process $x(t)$ exists whose sample functions are almost all continuous and which satisfies certain limit relations exhibiting m and σ^2 as the instantaneous mean and variance. Conversely, given m and σ, if $x(t)$ is a process whose sample functions are almost all continuous and which satisfies the stated limit relations, together with some auxiliary conditions, then $x(t)$ is a Markov process and there exists (or can be adjoined) a Brownian motion process $y(t)$ such that (*) is satisfied. This converse (Theorem 3.3) is new and made possible by the author's good work on martingales.

Chapter 8. Martingales (99 pages). This is perhaps the most original chapter of the book, much of the material being published here for the first time. Lévy in his 1937 book considered a class of dependent variables whose partial sums form a martingale, as a natural generalization of independent random variables in the sense that many classical limit theorems can be extended to them. (He discovered or anticipated a number of the results given here.) The name "martingale" was introduced by Ville (1939) who gave its present definition and obtained some preliminary results including the extension of Kolmogorov's inequality. It was, however, the author, a little later and under the unprepossessing name "processes with the property $\mathcal{E}$," who established martingale as a process *per se*, thus creating a new branch of stochastic processes much as Wiener did with the Brownian motion process and Lévy with the processes with independent increments. A martingale is a stochastic process $\{x_t, t\in T\}$ such that $E\{|x_t|\}<\infty$, $t\in T$, and $x_{t_n}=E\{x_{t_{n+1}}|x_{t_1}, \cdots, x_{t_n}\}$ with probability one whenever $t_1< \cdots <t_{n+1}$. If the $=$ sign is replaced by $\leqq$ then it is called a semi-martingale. The success of martingale is rooted in its interpretation as a fair game: if x_n is the gambler's fortune after the nth play, the above definition implies that his expected fortune is always equal to his present one. This gambling interpretation brings with it the notion of a gambling system, greatly exploited here, which is fundamental to the theory and its application. The intuitive background is obvious: if the game is fair, then it remains so if the gambler examines his fortune only at certain moments chosen without clairvoyance. Mathematically speaking, let $m_1, m_2, \cdots$ be a nondecreasing sequence of integer-valued random variables such that for every μ, the condition $m_j=\mu$ is a condition on the x_t with $t\leqq\mu$. Let $\check{x}_j=x_{m_j}$. Then the $\check{x}_j$ process is said to be obtained from the x_j process by "optional sampling." An important special case is that of "optional stopping": there is a random variable m such that $m_j=\text{Min}\ (m, j)$. It is intuitive that if x_j is a martingale (or semi-martingale), then so is $\check{x}_j$; it is, however, less intuitive that this is true only when certain conditions are imposed on x and/or m. Indeed were it not for the nuisance of such conditions a "winning system" would be possible and the author would have done better going to Las Vegas than writing this book. These conditions complicate the theorems which are often crammed (a style the author loves) but the reviewer will henceforth apply the excellent methods of optional sampling and skipping freely. The fundamental inequalities are the extension of Kolmogorov's mentioned above, and the following new one. Let x_j, $1\leqq j\leqq n$, be a semi-martingale and let $\beta(\omega)$ be the number of times the sample sequence $[x_1(\omega), \cdots, x_n(\omega)]$ passes from below r_1 to above r_2, then $E(\beta)\leqq(r_2-r_1)^{-1}(E\{|x_n|\}+|r_1|)$. This result, due to Doob and Snell, is here proved neatly by "optional skipping" (skipping some of the differences x_j-x_{j-1}). All convergence theorems for semi-martingales follow from this inequality. For a forward semi-martingale $\{x_n, n\geqq 1\}$ a condition is needed to ensure that $\lim_{n\to\infty}x_n=x_\infty$ exists (with probability one), and another to ensure that $\{x_n, 1\leqq n\leqq\infty\}$ is semi-martingale. For a backward semi-martingale $\{x_n, n\leqq -1\}$, $\lim_{n\to\infty} x_{-n}=x_{-\infty}$ exists always and $\{x_n, -\infty\leqq n\leqq -1\}$ is a semi-martingale. (The connection of these results with the related work of Andersen and Jessen is discussed in the Appendix.) Best illustrations of martingales are the conditional expectations

$E\{z|\mathcal{F}_n\}$ for a nondecreasing sequence of Borel fields $\mathcal{F}_n$; in particular $E\{z|y_1, \cdots, y_n\}$ and $E\{z|y_n, y_{n+1}, \cdots\}$, $n \geqq 1$. Various applications are given: the 0-1 law; the theorem that for a series of independent random variables convergence with probability one is equivalent to convergence in probability; generalization of some inequalities by Marcinkiewicz and Zygmund; theory of integration and differentiation; likelihood ratios; sequential analysis (Wald's equation). But the most delightful one is undoubtedly a new proof of the strong law of large numbers. If y_n', $n \geqq 1$, are independent random variables, and $y_n = y_1' + \cdots + y_n'$, this famous law follows from the fact that $E\{y_1'|y_n, y_{n+1}, \cdots\}$ is a (backward) martingale. The discussion of continuous parameter martingales is partly straightforward extensions of discrete parameter results and partly measure-theoretic complications. The main theorem (Theorem 11.5) states that almost all sample functions of a separable semi-martingale x_t, $t \in T$, are bounded in $[a, b]T$ if $a, b \in T$; that they have finite left-(right-) hand limits at every $t \in T$ which is a limit point of T from the left (right); and that their discontinuities are jumps except perhaps at the fixed points of discontinuity. This is proved by the two fundamental inequalities mentioned above and has several applications in the book. The discussion of optional sampling is peculiarly Doobian and is an example of painstaking modern rigor in probability theory. The chapter ends with applications to sample function continuity of Markov processes and processes with independent increments, and an important theorem (originated with Lévy) which states that if $\{x_t\}$ is a martingale whose sample functions are almost all continuous and such that $\{x_t^2 - t\}$ is also a martingale, then the process is a Brownian motion process. This theorem is used in Chapter 9 in connection with the Ito process.

Chapter 8. Processes with independent increments (34 pages). There is a brief discussion of Brownian motion process including the so-called "reflection principle," continuity of almost all sample functions and the expression of the variance as a stochastic limit (Lévy). That the author did not choose to enter into more details of this too, too popular process may be partly explained by the existence of a "profound study" in Lévy's 1948 book. Some readers will however miss the inclusion of material from the author's 1942 paper on the Ornstein-Uhlenbeck process rather than the desultory §3, obviously a concession to custom. There is a discussion of several useful formulations of the Poisson process and an application to the macroscopic equilibrium of molecular and stellar phenomena. (We hope physicists will appreciate the neat and yet rigorous derivation there!) The rest of the chapter is the author's version of what has been variously called "differential process," "additive process" and "integral with independent random elements," and is now renamed as given in the title. This theory, one of the crowning achievements in modern probability, is a natural generalization of the "addition of independent random variables" from the discrete to the continuous parameter case. It received its definitive form in the hands of Lévy (1934) and was further developed in his 1937 book. The re-discretization of this process, begun by Khintchine and carried on mainly by the Russian school, led to the ramified theory of infinitely divisible distributions. This latter theory, though of great importance in itself, must be regarded as a step toward retrenchment from the standpoint of stochastic processes. It is clear from Lévy's writing that he has always regarded the subject as one belonging to a (continuous-parameter) process and it was under this guidance that he was led by his extraordinary intuition to the discovery of all the main facts of the theory. That Khintchine and later authors chose the more formal analytical approach must be partly due to the fact that at the time the foundations of stochastic processes were hardly laid (cf. the last-mentioned dates with that of Khintchine's book cited in paragraph 2 above), and that mathematicians endowed with less intuition feared to tread the ground broken by Lévy. Thus we should indeed be grateful to the author for this account of Lévy's theory, taken strictly in the spirit of its creator, and embellished with the author's own reflections. Such is e.g. his construction of a process whose (almost all) sample functions have prescribed fixed points of discontinuity but are otherwise continuous. There is also a detailed discussion of the centering technique based on the results of Chapter 3. The final characterization of the sample functions of a separable centered process is obtained as a fast application of martingale theory, by noting that $\tilde{x}_t = e^{i\mu(x_t - x_a)}/E\left\{e^{i\mu(x_t-x_a)}\right\}$ is a martingale. Hence they have the same continuity properties as those of a martingale given in the preceding paragraph.

Chapter 9. Processes with orthogonal increments (27 pages). Integrals of the form $\int_A \Phi(t) d_t y(t)$, where $\{y(t),\ t \in T\}$ is a process with orthogonal increments and $A \in T$, are defined. It is followed by a discussion of processes which are Fourier transforms of each other, a subject which has been much blown up by "abstract" people. A more general integral of the form $\int_A \Phi(t,\ \omega) dy(t)$ is defined where the integrand now depends also on ω and $y(t)$ is a martingale, under appropriate conditions. These definitions specialize to Ito's if $y(t)$ is a Brownian motion process. The extension to a martingale $y(t)$ carries with it a closure property in the sense that the indefinite integral $x(t)=\int_a^t\Phi(s, \omega)dy(s)$ is then also a martingale and we can consider integrals of the form $\int\Phi(t, \omega)\, dx(t)$. Furthermore the $x(t)$ process has nice properties, in particular, its sample functions are almost all continuous if those of the $y(t)$ process are. If so and if roughly speaking the variance checks, then the $y(s)$ process in the integral representation of $x(t)$ may be taken to be the Brownian motion process. These results are used in the discussion of the Ito process in Chapter 6.

Chapter 10. Stationary processes—discrete parameter (55 pages); Chapter 11. Stationary processes—continuous parameter (53 pages). It is well known that the theory of strict sense (wide sense) stationary processes is equivalent to that of measure preserving (isometric, unitary) transformations. A detailed discussion is given to clarify the various notions of point, set, and random variable transformations in order to make this equivalence exact. G. D. Birkhoff's ergodic theorem thus becomes the strong law of large numbers for strictly stationary processes and is proved by F. Riesz's method. The usual corollaries are given. Turning to the wide sense stationary processes there is the inevitable identification of covariance function with positive definite function (Herglotz, Bochner, Khintchine) followed by specific examples of such processes to show the various possibilities. The spectral representation of the process (corresponding to the Hilbert space theory of von Neumann-Wintner and Stone) is given by Cramér's elegant method of setting up a distance preserving correspondence between the x_n's (the random variables of the process) and the numerical functions $e^{2\pi in\lambda}$. von Neumann's ergodic theorem, namely the law of large numbers in L^2, is now proved easily by this representation. A form of the strong law of large numbers (convergence with probability one), due to Loève and others, is proved under an order restriction. The author adds his own theorem on the convergence of $(n+1)^{-1}\sum_{j=0}^{n} x_{v+j}\bar{x}_j$ to the covarience $R(v)$, and related things. There are sections on "moving averages," "linear operations" (in the continuous parameter case they include differentiation and integration in L^2) and rational spectral densities in $e^{2\pi i\lambda}$. The continuous parameter case of these results is pretty much the same and both the reader and the reviewer, though not the author, may be spared the repetitious details. §11 in Chapter 11, on processes with stationary (wide sense) increments, is an exposition of Kolmogorov's theory on "curves in Hilbert space." The material in these two chapters, hallowed by tradition, somehow wears a worn look. The next generation of students of probability, who will read this probability version before its traditional counterpart, will probably find it more refreshing.

Chapter 12. Linear least squares prediction—stationary (wide sense) processes (39 pages). This chapter "is somewhat out of place in the book, since it discusses a rather specialized problem." The object of study is the wide sense conditional expectation $\hat{E}\{x_{n+v} | \cdots, x_{n-1}, x_n\}$, and is seen to reduce to linear approximations in L^2 with an arbitrary weighting function $F(\lambda)$, the spectral distribution function of the process. The corresponding analytic problem was discovered by Szegö in 1920; and the subject was treated by Wold and Kolmogorov before Wiener rediscovered and popularized it in this country. The present exposition claims to have made the material readily available to the American reader, in the usual language of probability.

The Supplement (24 pages) includes a "treatment of various aspects of measure theory with which the ordinary reader may not be familiar." The section headings are: fields of point sets; set functions; measure-preserving transformations. An Appendix (13 pages) collects references to the literature and historical remarks. A Bibliography (8 pages) collects the articles and books referred to in the Appendix. Relevance rather than excellence seems to be the criterion for inclusion there. There is a subject index at the end.

The following is a general appraisal of the book. The foundation of the theory of stochastic processes is largely the author's own contribution, some of which is published here for the first time. For the last 15 years the author has been almost alone in the treatment of stochastic processes from the purely probability or measure theory point of view, as distinct from the Laplace-Fourier-transform, or differential-integral-equation, or Hilbert-Banach-space standpoints. For him the study of a stochastic process is the study of its sample functions in their own right, not as the shadows behind analytical expressions nor in their weak-topological collectivizations. This book is an eloquent testimony to the success of this direct approach and will doubtless inspire and guide its further development.

Although much of the material on special processes is in the literature, the author's accounts of the results of Lévy, Doeblin, and others are far from being routine expositions. Indeed, the mathematical world is indebted to a man of the author's stature for rendering such unenviable services. He is apparently as willing to interpret the work of others as to give his own. This modesty is refreshing, although the reader may have been thereby deprived of some more thorough-going measure theory and several other interesting topics of the author's own which have been alluded to in the preceding paragraphs.

A few words must now be said of the style, inasmuch as the author enjoys, justly or not, a reputation on it. In judging the readability of this book, it must be borne in mind that it is frequently the substance rather than the form which makes for difficulty. It seems true that the author has been somewhat lax in allowing for human weaknesses on the part of the reader in his early writings, but in this book he seems to have made a sincere effort to be more accommodating. An example is his considerable pains in handling sets of measure zero, putting the horse before the cart in phrasing them and showing their dependence sometimes (cf. Theorem 3.1 of his 1937 paper). There is also a minimum amount of handwaving except that he occasionally refuses to be drawn out on measure theory. An instance where this seems unwarranted is as follows. On p. 51 a most trivial sort of example is mentioned but the reader is immediately waved off to the end of that section for further enlightenment. When he finally reaches p. 70 after a rugged 20 pages, or even if he turns at once to the promised place as this reviewer did, he finds only a number of statements of the "it-is-easy-to-see" type. Admittedly they are easy enough if the reader knows what's going on, but this is precisely what he (the average one!) does not at this stage of the game. E.g. it would take only half a page to substantiate the statement on p. 70 that "If X contains a second point, the x_t process is neither separable nor measurable," but this the author steadfastly refused to do. Indeed nowhere in the discussion did he ever indicate a proof (of the type attributed to Halmos) that a certain ω set may not be measurable. Such trivial omissions may cause some readers a disproportionate amount of labor (and complaint). However, let it be stated that this book should on the whole be quite accessible to a determined reader.

No serious error has been found by this reviewer, but there are some minor errors and a number of misprints. We mention only the following: p. 54, The last line is incorrect; part (ii) of Theorem 2.2 is proved only if "in probability" is substituted for "with probability one" in line 20 (noted by J. G. Wendel); p. 64, line 10: another term like $2\delta\int_0^{2b-a}|f(t)|\,dt$ is needed at the end; p. 285, line 2 and later reference to the formula: read h^2 for $h^{3/2}$; p. 357, the statement "If $t\in T_1(n)\cdots$" needs minor fixing. A longer list has been turned over to the author.

Finally, this is an honest mathematics book. It is not designed to sell stochastic processes cheap. It takes all sorts to make stochastic processes, but let Mr. Doob write only for the sake of mathematics. He has done it.

K. L. CHUNG