WHAT IS A MARTINGALE?

J. L. DOOB, University of Illinois

1. Introduction. Martingale theory illustrates the history of mathematical probability: the basic definitions are inspired by crude notions of gambling, but the theory has become a sophisticated tool of modern abstract mathematics, drawing from and contributing to other fields. Martingales have been studied systematically for about thirty years, and the newer probability texts usually devote some space to them, but the applications are so varied that there is no one place where a full account can be found. References [1] and [2] are the most complete sources.

The following account of martingale theory is designed to give a feeling for the subject with a minimum of technicality. The basic definitions are given at two levels, of which the first is more intuitive and elementary and suffices for some of the examples. The examples illustrate only the immediate applications requiring a minimum of background.

We recall that in probability theory one starts with a set called the **sample space,** that **events** are subsets of this space, and **random variables** are functions on this space. Suppose for simplicity that the sample space Ω has only countably many points $\omega_1, \omega_2, \cdots$ to which are assigned probabilities $p_1, p_2, \cdots$ respectively, with $p_j \geqq 0$ and $\sum_j p_j = 1$. If $x_1, \cdots, x_k$ are random variables, we write

$$\{x_1 = a_1, \cdots, x_k = a_k\} = \bigcap_m \{\omega_j : x_m(\omega_j) = a_m\}$$

for the set of points where the random variables have the indicated values. The **probability** $P\{A\}$ of the event A is defined as $\sum p_j$, where the sum is over those values of j with ω_j in A. If $P\{B\} > 0$, the **conditional probability** of the event A relative to B is defined by $P\{A \mid B\} = P\{A \cap B\}/P\{B\}$. If x is a random variable, its **expectation** is defined as

$$E\{x\} = \sum_j x(\omega_j) p_j, \tag{1.1}$$

(where it is supposed that the series converges absolutely) and the **conditional expectation** of x relative to B is defined correspondingly when $P\{B\} > 0$ as

$$E\{x \mid B\} = \sum_j{}' x(\omega_j) p_j / P\{B\}, \tag{1.2}$$

Professor Doob received his Harvard Ph.D. under J. L. Walsh, spent the next two years at Columbia on an N.R.C. Fellowship, a third year there on a Carnegie Corp. grant to Harold Hotelling, and began his long association with the University of Illinois in 1935. He is a member of the National Academy of Sciences, served as President of the American Mathematical Society in 1963–64, and served the Inst. of Math. Stat. as Vice-President in 1945 and as President in 1950. He is the author of *Stochastic Processes* (Wiley, 1953) and of numerous important research papers on probability theory and applications. *Editor.*

where the prime indicates that the sum is over the values of j with ω_j in B. If $P\{B\}=0$ the preceding conditional probability and expectation can be defined arbitrarily without affecting later work.

It has been found useful to make conditional expectations into functions, as follows. Let $\{B_n, n\geqq 1\}$ be a partition of Ω, that is, a countable class of disjoint sets with union Ω. This partition generates, and is in turn determined by, a σ-algebra $\mathfrak{F}$, namely the class of all unions of sets of the partition. If x is a random variable with an expectation, define $E\{x|\mathfrak{F}\}$, the **conditional expectation** of x relative to $\mathfrak{F}$, as the random variable with the constant value $E\{x|B_n\}$ on each set B_n. This definition is unambiguous except on the partition sets (if there are any) of probability 0. In particular if $y_1, \cdots, y_k$ are random variables, they induce the partition $\mathfrak{F}$ each of whose sets is determined by a condition of the form $\{y_1=a_1, \cdots, y_k=a_k\}$; in this case $E\{x|\mathfrak{F}\}$, also denoted by $E\{x|y_1, \cdots, y_k\}$, is the function with the value $E\{x|y_1=a_1, \cdots, y_k=a_k\}$ on the set $\{y_1=a_1, \cdots, y_k=a_k\}$.

Let $\mathfrak{F}_1\subset\mathfrak{F}_2\subset\cdots$ be a finite or infinite increasing sequence of σ-algebras (generated by partitions of the sample space as just described). The intuitive picture to keep in mind is that $\mathfrak{F}_n$ represents the class of all relevant past events up to and including time n. The monotoneity relation corresponds to the idea that the past to time $n+1$ includes more events than the past to time n. Let $x_1, x_2, \cdots$ be a sequence of random variables. We consider x_n as part of the relevant history to time n, interpreting this statement to mean that each event of the form $\{x_n=a\}$ is a set in the class $\mathfrak{F}_n$. The sequence of random variables is to be analyzed. In some applications $x_1, x_2, \cdots$ are specified and $\mathfrak{F}_n$ is the past as determined entirely by $x_1, \cdots, x_n$, that is $\mathfrak{F}_n$ is generated by the partition of Ω induced by $x_1, \cdots, x_n$. This choice of $\mathfrak{F}_n$ will be called **minimal** (relative to a specified sequence of random variables).

The sequence $\{x_n, n\geqq 1\}$ is called a **martingale** relative to $\{\mathfrak{F}_n, n\geqq 1\}$ if each x_n has an expectation, and if for $m<n$ the expected value of x_n given the past up to time m is x_m, that is

$$E\{x_n \mid \mathfrak{F}_m\} = x_m. \tag{1.3}$$

This is a relation between functions on the sample space and is to hold almost everywhere, that is everywhere except perhaps on a subset of the sample space of probability 0. If every p_j is strictly positive, the exceptional set is empty. If x_n is thought of as the fortune of a gambler at time n, the defining equality (1.3) corresponds to the idea that the game the gambler is playing is fair. If '$=$' in (1.3) is replaced by '$\geqq$' or '$\leqq$', the sequence of random variables is called a **submartingale** or **supermartingale** respectively, and the corresponding games are then respectively favorable or unfavorable to the gambler. Trivially (for specified σ-algebras) $\{x_n, n\geqq 1\}$ is a supermartingale if and only if $\{-x_n, n\geqq 1\}$ is a submartingale, and is a martingale if and only if it is both a supermartingale and a submartingale. The definitions imply that $E\{x_n\}$ increases with n in the submartingale case, decreases with n in the supermartingale case, and does not vary with n in the martingale case. (All monotoneity statements are to be interpreted in the wide sense.)

Equation (1.3) implies that

$$E\{x_n \mid x_1, \cdots, x_m\} = x_m, \tag{1.4}$$

equivalently that

$$E\{x_n \mid x_1 = a_1, \cdots, x_m = a_m\} = a_m, \tag{1.4'}$$

whenever the conditioning event has strictly positive probability, and in fact (1.4) is the same as (1.3) whenever every $\mathfrak{F}_k$ is minimal. In other words, a martingale relative to a given sequence of σ-algebras is also one relative to the minimal sequence of σ-algebras. A corresponding remark is valid for submartingales and supermartingales. If the sequence of σ-algebras is not mentioned, the minimal sequence is to be understood.

The definition of a martingale is applicable to complex-valued random variables, and we shall consider certain complex martingales below. Trivially, the real and imaginary parts of a complex martingale are real martingales.

2. Definitions in the general case. The definitions in Section 1 assumed countability of the sample space, a condition not satisfied for some of the applications to be described below. In this section definitions will be given in the general case, in non-probabilistic language to convince cynical readers that probability theory does not need an admixture of non-mathematical terms like *coin*, *event*, *gambler*, *urn*, $\cdots$, even though the ideas behind these terms have inspired much of the theory.

Let $\{\Omega, \mathfrak{F}, P\}$ be a measure space: Ω is a set, $\mathfrak{F}$ is a σ-algebra of subsets of Ω, and P is a measure defined on $\mathfrak{F}$. Assume further that $P\{\Omega\} = 1$. (Some non-probabilists accuse probabilists of seeking to mystify outsiders by disguising measurable functions and their integrals with the aliases *random variables* and *expectations*. Note however that probabilists were dealing with the integrals of functions on abstract sets before other analysts dreamed of measure theory. It is sardonic that, dually, some probabilists accuse others of obfuscating probability with measure theory.) Let $\mathfrak{F}_1 \subset \mathfrak{F}_2 \subset \cdots$ be an increasing sequence of σ-algebras of $\mathfrak{F}$ sets. Let $\{x_n, n \geqq 1\}$ be a sequence of complex functions on Ω satisfying the following conditions:

(a) x_n is measurable relative to $\mathfrak{F}_n$;
(b) x_n is integrable;
(c) If $m < n$ and if A is any set in $\mathfrak{F}_m$, then

$$\int_A x_n dP = \int_A x_m dP. \tag{2.1}$$

Then the sequence of random variables is said to be a **martingale** relative to $\{\mathfrak{F}_n, n \geqq 1\}$. If the random variables are real and if '=' is replaced in (2.1) by '$\leqq$'

or '$\geqq$', the sequence of functions is said to be a **supermartingale or submartingale,** respectively, relative to the sequence of σ-algebras. Conditional expectations relative to a σ-algebra (in the present general context) are defined in such a way that (2.1) and (1.3) (to hold P almost everywhere on Ω) are equivalent. Thus the present definitions include the earlier ones and the collateral remarks about martingale theory in Section 1 are valid in the general case also.

The definitions have been given for the parameter set $1, \cdots, k$ or $1, 2, \cdots$, ordered as usual, but they are obviously extendable to any simply-ordered parameter set.

3. Example: expectations knowing more and more. Let $\mathfrak{F}_1 \subset \mathfrak{F}_2 \subset \cdots$ be an increasing sequence (finite or infinite) of σ-algebras, as usual, and let x be a random variable with an expectation. Then if $x_n = E\{x \mid \mathfrak{F}_n\}$, the sequence $x_1, x_2, \cdots$ is a martingale relative to the given σ-algebra sequence. That is, successive conditional expectations of x, as we know more and more, yield a martingale. More generally, the parameter set $1, 2, \cdots$ can be replaced in this example by any simply ordered set. The essential condition is that $\mathfrak{F}_t$ increase with t. Every martingale whose parameter set has a last element is of this type, with x identified with the last element.

This example suggests the possibility of applications to statistics and information theory and induces such probabilistic extravagances as the statement that 'the game man plays with nature as he learns more and more is fair.' (Perhaps this statement indicates how realistically a martingale reproduces the idea of a fair game.)

4. Example: sums of independent random variables. Let $y_1, y_2, \cdots$ be independent random variables with expectations, and let $x_n = y_1 + \cdots + y_n$. Then it is intuitively obvious and easily proved that $x_1, x_2, \cdots$ is a martingale if $E\{y_j\} = 0$ for $j > 1$, a submartingale if $E\{y_j\} \geqq 0$ for $j > 1$, a supermartingale if $E\{y_j\} \leqq 0$ for $j > 1$.

5. Example: averages of independent random variables. In the preceding example suppose that $y_1, y_2, \cdots$ have a common distribution which has an expectation. Then a symmetry argument shows that

$$\cdots, \frac{x_3}{3}, \frac{x_2}{2}, x_1$$

is a martingale in the indicated order (left to right). This example suggests possible applications to the law of large numbers and therefore suggests ties between martingale theory and ergodic theory. In fact there are close relationships between the two theories, and sometimes it is said that one contains the other. Which is said to contain the other depends on the speaker. At any rate, it is true that in a reasonable sense there are only two qualitative convergence theorems in measure theory (aside from theorems of the form "convergence of type I implies convergence of type II"), the ergodic theorem and the martingale convergence theorem. The latter will be discussed below. Each is involved with finer and finer averaging.

6. Example: harmonic functions on a lattice. Let S be any subset of the set S' of points with integral coordinates in d-dimensional coordinate space, $d \geqq 1$. A point ξ of S will be called an **interior** point of S if S contains all $2d$ of the nearest neighbors in S' of ξ. Otherwise ξ will be called a **boundary** point of S. Unless $S=S'$ there will be boundary points. A function u on S will be called **harmonic** (**superharmonic**) if u at each interior point of S is equal (at least equal) to the average of u on its $2d$ nearest neighbors. For example, a linear function is harmonic, a concave function of a harmonic function is superharmonic. Define a **walk** on S, that is, a sequence $x_0, x_1, \cdots$ of random variables with values in S, as follows. Prescribe some initial point in S and set x_0 identically this point. If $x_0=a_0, \cdots, x_n=a_n$ and if a_n is an interior point of S, then x_{n+1} is to be (conditional probability) one of the $2d$ nearest neighbors of a_n, with (conditional) probability $1/(2d)$ for each one. If a_n is a boundary point of S, x_{n+1} is to be a_n. Thus the walk proceeds until the boundary is reached, if ever, and sticks at the first boundary point reached. It can be shown that such a walk exists. If u is harmonic (superharmonic) on S, the sequence of random variables $u(x_0)$, $u(x_1), \cdots$ is a martingale (supermartingale). If we are to consider the infinite sequence $x_0, x_1, \cdots$, the sample space must be uncountable.

7. Example: classical harmonic and analytic functions. A variation of the idea of Section 6 is the following. Let S be an open subset of d-dimensional coordinate space, $d \geqq 1$. A function u on S is said to be **harmonic** if u is continuous and if, whenever ξ is a point of S, and B is a ball with center ξ whose closure lies in S, the value of u at ξ is the average of its values on the boundary of B. For example linear functions are harmonic for all d, and are the only harmonic functions if $d=1$ and S is an interval. The harmonic functions are the infinitely differentiable functions whose Laplacians vanish. If $d=2$, the real part of an analytic function is harmonic. If ξ is in S and if S is the whole space, denote by $B(\xi)$ the boundary of the ball with center ξ and radius 1. If S is not the whole space, denote by $B(\xi)$ the boundary of the ball with center ξ and radius half the distance from ξ to the boundary of S. If $d>1$ and $A \subset B(\xi)$, let $p(\xi, A)$ be the $(d-1)$-dimensional "area" of A divided by that of $B(\xi)$. If $d=1$ let $p(\xi, A)$ be one-half the number of points in A. Now define a walk on S, that is, random variables $x_0, x_1, \cdots$ with values in S, as follows. Prescribe some initial point in S and set x_0 identically this point. If $x_0=a_0, \cdots, x_n=a_n$, then x_{n+1} is to be on $B(a_n)$, and in fact the conditional probability that x_{n+1} is in the subset A of $B(a_n)$ is to be $p(a_n, A)$. With this definition, if u is real and harmonic, or if $d=2$ and u is complex and analytic on S, the sequence of random variables $\{u(x_n), n \geqq 0\}$ is a martingale. If superharmonic and subharmonic functions are defined as usual in this context, there is a corresponding relation between superharmonic (subharmonic) functions and supermartingales (submartingales).

Sections 6 and 7 indicate connections between martingale theory and potential theory. In fact the probability theory of Markov processes and abstract potential theory are to a considerable extent different ways of looking at the same subject, and martingale theory is an essential tool of probabilistic potential theory.

8. Two basic principles. We shall need the concept of a **stopping time,** also called a **Markov time** and **optional time.** If $\mathfrak{F}_0 \subset \mathfrak{F}_1 \subset \cdots$ is an increasing sequence of σ-algebras and if ν is a random variable whose range is the set of positive integers together with $+\infty$, the random variable ν is called a **stopping time** (relative to the sequence of σ-algebras) if $\{\nu \leqq k\}$ is a set in $\mathfrak{F}_k$ for $k=0, 1, \cdots$, that is, in intuitive language, if the condition $\nu \leqq k$ is a condition involving only what has happened up to and including time k. For example, if $\{x_n, n \geqq 0\}$ is a martingale relative to the given sequence of σ-algebras and if $\nu(\omega)$ is the first integer j for which $x_j(\omega) > 0$, or $\nu(\omega) = \infty$ if there is no such integer, then ν is a stopping time.

Two basic principles are embodied in the following rough statements, which will be given exact versions and applied in various contexts. Let $x_0, x_1, \cdots$ be a supermartingale relative to $\mathfrak{F}_0, \mathfrak{F}_1, \cdots$.

P_1. *If $\nu_1 \leqq \nu_2 \leqq \cdots$ are finite stopping times which are not too large (with reference to $x_0, x_1, \cdots$), then the sequence $x_{\nu 1}, x_{\nu 2}, \cdots$ is a supermartingale; it is a martingale if $\{x_n, n \geqq 0\}$ is a martingale.*

P_2. *If* $\sup_n E\{|x_n|\}$ *is not large, then* $\sup_n |x_n|$ *is not large, and the sequences $\{x_0(\omega), x_1(\omega), \cdots\}$ of possible values of the supermartingale are not strongly oscillatory.*

The principle P_1 is suggested by the fact that a gambler playing an unfair (fair) game will still consider it unfair (fair) if he looks at his money not after every play, but only after plays number $\nu_1, \nu_2, \cdots$. The following version of P_1, in which all stopping times are finite, will be called the STOPPING TIME THEOREM below:

If $\{x_n, n \geqq 0\}$ is a supermartingale (martingale) and if each ν_j is bounded, then the sequence $\{x_{\nu_j}, j \geqq 1\}$ is also a supermartingale (martingale). With no restriction on the finite stopping times, the second sequence is a supermartingale if the first was, and if also $x_j \geqq 0$ for all j.

The principle P_2 is exemplified by the following theorem, which will be called the MARTINGALE CONVERGENCE THEOREM below:

If $\{x_n, n \geqq 1\}$ is a supermartingale with $\sup_n E\{|x_n|\} < \infty$, *then* $\lim_{n\to\infty} x_n$ *exists and is finite almost surely. If $\{\cdots, x_{-1}, x_0\}$ (ordered left to right) is a supermartingale, then (almost surely)* $\lim_{n\to-\infty} = x_{-\infty}$ *exists with* $-\infty < x_{-\infty} \leqq \infty$.

Here **almost surely** means everywhere on the sample space except possibly for a set of probability zero. (One of the most noticeable distinctions between probability and measure theory is that a probabilist frequently writes "almost surely" where a measure theorist writes "almost everywhere.")

The Martingale Convergence Theorem should be contrasted with the ERGODIC THEOREM, which we state in the following version:

Let $z_1, z_2, \cdots$ be random variables with expectations and having the property that for every $n \geqq 0$ the joint distribution of $z_j, \cdots, z_{n+j}$ does not depend on $j \geqq 1$. Then almost surely

$$\lim_{n\to\infty} (z_1 + \cdots + z_n)/n \tag{8.1}$$

exists and is finite.

Denote by z_n' the ratio in (8.1). Then z_n' is a weighted average of $z_1, z_2, \cdots$ and z_{n+1}' is a weighted average of $z_1', \cdots, z_n', z_{n+1}, \cdots$. Thus z_{n+1}' is a coarser average than z_n'. Now the defining equality of a martingale $\{x_n, n \geqq 1\}$ makes x_n a partial average of x_{n+1}. Thus if there is any relation between the Martingale Convergence Theorem and the Ergodic Theorem, one would conjecture that the Ergodic Theorem corresponds to the Martingale Convergence Theorem with decreasing index. The application in Section 11 verifies this conjecture in a special case.

9. Continuation of Section 3. In the example in Section 3, $E\{|x_n|\} \leqq E\{|x|\}$. Thus according to the Martingale Convergence Theorem, $\lim_{n\to\infty} E\{x|\mathfrak{F}_n\}$ exists almost surely. If $\mathfrak{F}_\infty$ is defined as the smallest σ-algebra containing every set of $\bigcup_n \mathfrak{F}_n$, then the limit can be identified as $E\{x|\mathfrak{F}_\infty\}$. We have obtained a kind of continuity theorem for conditional expectations. There is a corresponding theorem relative to a decreasing sequence of σ-algebras.

10. Continuation of Section 4. Suppose in Section 4 that $E\{y_j\} = 0$ for all j. Then the sequence $\{x_n, n \geqq 1\}$ is a martingale, and according to the Martingale Convergence Theorem, the series $\sum_j y_j$ converges almost surely if $\sup_n E\{|\sum_1^n y_j|\} < \infty$. Suppose for example that $E\{y_j^2\} < \infty$ for all j, so that the series $\sum_j y_j$ is a series of orthogonal random variables, and as such converges in the mean if and only if $\sum_j E\{y_j^2\} < \infty$, that is, if and only if $\sup_n E\{x_n^2\} < \infty$. But the finiteness of this supremum means that the hypothesis of the Martingale Convergence Theorem is satisfied, and we have proved that convergence in mean of a sum of independent random variables with zero expectations implies almost sure convergence.

To obtain a second application of martingale theory to sums of independent random variables, let $z_1, z_2, \cdots$ be mutually independent random variables with a common distribution having expectation α. If $x_0 = 0$ and $x_n = \sum_1^n (z_j - \alpha)$ for $n > 0$, the sequence $\{x_n, n \geqq 0\}$ is a sequence of sums of independent random variables with zero expectations and as such is a martingale. If ν is a finite stopping time (relative to the minimal σ-algebra sequence) with an expectation, and if $\nu_1=0$, $\nu_2=\nu$, it can be shown that a version of P_1 yields the fact that x_{ν_1}, x_{ν_2} is a martingale with two random variables, having common expectation 0. The fact that $E\{x_\nu\}=0$ means that

$$E\left\{\sum_1^\nu z_j\right\} = \alpha E\{\nu\}. \tag{10.1}$$

This equality is Wald's FUNDAMENTAL THEOREM OF SEQUENTIAL ANALYSIS, which has many applications in statistics.

11. Continuation of Section 5. An application of the Martingale Convergence Theorem to the martingale in Section 5 yields the almost sure existence of the limit

$$\lim_{n\to\infty} (y_1 + \cdots + y_n)/n. \tag{11.1}$$

The limit can be shown to be $E\{y_1\}$. This convergence result is known to probabilists as the STRONG LAW OF LARGE NUMBERS FOR INDEPENDENT RANDOM VARIABLES WITH A COMMON DISTRIBUTION, and the result can also be obtained as an application of the Ergodic Theorem. (See the discussion of the relation between the Ergodic Theorem and the Martingale Convergence Theorem in Section 8.)

12. Continuation of Section 6. We suppose that S is bounded and show first that almost every walk path from a point ξ of S reaches the boundary. There are elementary proofs of this fact and the following proof is given only to illustrate martingale theory. If u is defined and harmonic on S, then $\{u(x_n),\ n\geqq 0\}$ is a bounded martingale and is therefore almost surely convergent. In particular if u is a coordinate function, it is trivial that u on a walk sample sequence cannot be convergent unless u on the sequence is finally constant. Then almost every walk sample sequence must hit the boundary (where it sticks), as was to be proved.

If ξ is not a boundary point, there is an integer-valued random variable ν (the first hitting time of the boundary) such that x_ν is a boundary point, but x_j is not for $j<\nu$. The random variable ν is a stopping time relative to the sequence of minimal σ-algebras. If u is harmonic on S, if $\nu_1=0$ and $\nu_2=\nu$, the stopping time theorem (slightly extended) asserts that $u(\xi)$, $u(x_\nu)$ is a martingale with two random variables. But then

$$u(\xi) = E\{u(\xi)\} = E\{u(x_\nu)\}. \tag{12.1}$$

Denote by $\mu(\xi, \eta)$ the probability that $x_\nu=\eta$, that is, the probability that a walk starting at ξ hits the boundary at η. The distribution $\mu(\xi, \cdot)$ is called **harmonic measure** relative to ξ. In terms of harmonic measure, (12.1) becomes

$$u(\xi) = \sum u(\eta)\mu(\xi, \eta), \tag{12.2}$$

where the sum is over all boundary points η. Thus a harmonic function on S is determined by its values on the boundary, in fact, is the weighted average using harmonic measure of its values on the boundary. The rules for manipulating conditional expectations yield the fact that if u is an arbitrary function defined on the boundary of S, and if u is defined at interior points by (12.2), then u is harmonic on S. We have now shown the existence and uniqueness of a harmonic function with a specified boundary function. We omit the corresponding treatment of more general harmonicity defined using weighted averages, not necessarily at nearest neighbors. When $d=1$, the result obtained is particularly intuitive. In this case if S is the set of points $-a, \cdots, b$, where a and b are strictly positive integers, the walk is the random walk in which a step is either 1 or -1, with probability 1/2 each, independent of previous steps, until $-a$ or b is reached. In gambling language: a gambler is playing a fair game in which at each play he can win or lose a dollar with probability 1/2 each and the plays are independent; he starts with a dollars, his opponent with b dollars, and the game ends when he or his opponent has lost all his money; x_n is the gambler's total winnings (positive or negative) after the nth play. Since the play starts at time 0, we define $x_0=0$. If we take the harmonic function u to be the identity function, $u(\xi)=\xi$ on S, the sequence $\{x_n, n\geqq 0\}$ is seen to be a martingale, so (12.1) becomes

$$E\{u(x_\nu)\} = 0, \tag{12.1$'$}$$

that is, the expected final gain is 0 as it should be. Equation (12.1′) can also be obtained as a special case of (10.1). However obtained, this equation yields the standard result that the probability the gambler wins, that is, the probability that x_ν is b, is $a/(a+b)$.

13. Continuation of Section 7. If $d=2$, the properties of the random walk of Section 7 are intimately related to the properties of harmonic and analytic functions. We shall see this first in a simple application where S is the whole plane. It can be shown in this case that almost every sample walk starting from any point ξ is dense in the plane. This fact corresponds to Liouville's theorem that a bounded complex function which is analytic on the plane is a constant function, and we shall now prove Liouville's theorem probabilistically. Since the real and imaginary parts of an analytic function are harmonic, we shall obtain a stronger result if we prove that a function harmonic and bounded on the plane is constant. In fact we shall do even better and prove that a harmonic function on the plane which is bounded from above or below is a constant. By linearity we can suppose that the function u is positive, and we consider the martingale $\{u(x_n), n\geqq 0\}$. Trivially,

$$E\{u(x_0)\} = E\{u(x_n)\} = E\{\,|\,u(x_n)\,|\,\}.$$

According to the Martingale Convergence Theorem, this martingale is almost surely convergent. If we choose any sample walk which is dense and on which u has a limit, say c, the function u must be identically c by continuity, as was to be proved.

We proceed to the analog of the work in Section 12, assuming from now on that S is bounded and d arbitrary. The almost sure convergence of the bounded martingale $\{u(x_n),\ n\geqq 0\}$, when u is a coordinate function, implies that $\lim_{n\to\infty} x_n = x_\infty$ exists almost surely. This is impossible unless x_∞ has its values on the boundary of S. If ξ is the initial point of the walk and if A is a Borel boundary set, let $\mu(\xi, A)$ be the probability that x_∞ is in A. Then $u(\xi, \cdot)$ is a measure of boundary sets, *harmonic measure relative to* ξ. If u is a bounded harmonic function on S, or even if u is merely bounded above or below, the Martingale Convergence Theorem implies that $\lim_{n\to\infty} u(x_n)$ exists almost surely. That is, u has a limit at the boundary of S along almost every one of the sample walks. We shall continue this aspect of the discussion in Section 16. Suppose now that u is actually defined and continuous on the closure of S. Then trivially $\lim_{n\to\infty} u(x_n) = u(x_\infty)$ almost surely and it is straightforward to show that the ordered set of random variables

$$u(x_0),\ u(x_1),\ \cdots,\ u(x_\infty)$$

is a martingale. Since the expectations of the first and last random variables are equal,

$$u(\xi) = E\{u(x_\infty)\} = \int u(\eta)\mu(\xi, d\eta), \tag{13.1}$$

where the integration is over the boundary of S, that is, $u(\xi)$ is the average of its values on the boundary, weighted by harmonic measure. Conversely, if u is a function defined and continuous on the boundary of S, and if u is defined on S by (13.1), it can be shown, using the ideas in Section 16, that u is harmonic on S and has the assigned boundary function value as a limit at each boundary point near which the boundary is well-behaved (more precisely at each regular boundary point in the potential theoretic sense). Thus martingale theory solves the first boundary value problem for harmonic functions. This kind of analysis is applicable not only to classical harmonic functions, but also to the solutions of general elliptic and parabolic partial differential equations.

14. Example: application to derivation. The close relation between martingale theory and derivation theory is illustrated by the following example. Let Ω be the unit interval $[0, 1]$ and let the probability of a subset A be its Lebesgue measure, denoted by $|A|$. For each $n\geqq 1$, let $A_{n1}, \cdots, A_{nk_n}$ be a partition of $[0, 1]$ into disjoint intervals, and suppose that the $(n+1)$th partition is a refinement of the nth, that is, we suppose that each A_{nj} is a union of sets in the $(n+1)$th partition. If $\mathfrak{F}_n$ is the class of unions of sets in the nth partition, then $\mathfrak{F}_n \subset \mathfrak{F}_{n+1}$. If f is an integrable function on $[0, 1]$, define the random variable x_n by

$$x_n = \int_{A_{nj}} f(\omega)d\omega / |A_{nj}| \qquad \text{on } A_{nj},\quad j \geqq 1,$$

to get a martingale relative to $\{\mathfrak{F}_n,\ n\geqq 1\}$ for which

$$E\{|x_n|\} \leqq \int_0^1 |f(\omega)|\, d\omega. \tag{14.1}$$

If f is continuous and if $\lim_{n\to\infty}\max_j|A_{nj}| = 0$, it is trivial that $\lim_{n\to\infty}x_n(\omega) = f(\omega)$ for all ω. Without this restriction on f and on the partitions the Martingale Convergence Theorem implies that $\lim_{n\to\infty}x_n$ exists almost everywhere on $[0, 1]$. More generally, $[0, 1]$ can be replaced by any measure space $\{\Omega, \mathfrak{F}, P\}$ with $P\{\Omega\} = 1$. In this new context the partition is to be a partition of Ω into countably many disjoint measurable sets; the $(n+1)$th partition is to be a refinement of the nth; $\int_A f d\omega$ is replaced by $\mu(A)$, where μ is any finite signed measure on $\mathfrak{F}$; x_n is defined as $\mu(A_{nj})/P\{A_{nj}\}$ on A_{nj} if the denominator does not vanish, defined arbitrarily on A_{nj} if $P\{A_{nj}\} = 0$. With this definition the sequence of random variables is a martingale if

(a) $P\{A_{nj}\} > 0$ for all n, j,

or if

(b) $\mu(A_{nj}) = 0$, whenever $P\{A_{nj}\} = 0$.

Without either (a) or (b) the sequence of random variables is a supermartingale if

(c) $\mu \geqq 0$.

Under (a) or (b) or (c), $E\{|x_n|\}$ is at most the absolute variation of μ, and we conclude that the martingale or supermartingale converges almost surely. Since a finite signed measure is the difference between two finite measures, we have derived the fact that a finite signed measure has a derivative relative to any finite measure with respect to a net of partitions. (The extension to allow $P\{\Omega\}$ to be other than 1 is trivial.)

15. Example: functions on an infinite dimensional cube. The definition of a martingale implies that each martingale random variable is a partial averaging of the next one. The following example exhibits this fact very neatly. Let Ω be the coordinate space of two way infinite sequences $\cdots, \xi_{-1}, \xi_0, \xi_1, \cdots, 0 \leqq \xi_j \leqq 1$. Let P be the product measure on Ω for which each coordinate measure is Lebesgue measure, that is, P is infinite dimensional volume. Let f be a measurable integrable function on Ω. Then we define, in the obvious notation, the random variable x_n by

$$x_n = \int\int \cdots f(\cdots, \xi_0, \xi_1, \cdots) \prod_{n+1}^{\infty} d\xi_j. \tag{15.1}$$

Here we have integrated out $\xi_{n+1}, \xi_{n+2}, \cdots$. If $\mathfrak{F}_n$ is the smallest σ-algebra of subsets of Ω containing every set determined by a condition of the form $\xi_j < c$ for $j \leqq n$, it can be shown that $x_n = E\{f | \mathfrak{F}_n\}$ almost surely, as should be intuitively clear: we are calculating the expected value of f knowing all the coordinates up to and including the nth. Then (see Sections 3 and 9) the two way sequence $\{x_n, -\infty < n < \infty\}$ is a martingale which converges in both directions: $\lim_{n\to\infty}x_n = x_\infty$ and $\lim_{n\to-\infty}x_n = x_{-\infty}$ exist almost surely. These limits can be shown to be (almost surely) f and $E\{f\}$ respectively.

16. Continuous parameter case. Let $\{x_t, t\in I\}$ be a supermartingale, where I is an interval on the line. Then Principle P_2 is illustrated by the fact that each random variable $\omega \rightsquigarrow x_t(\omega)$ can be changed on an ω set of probability 0 in such a way that almost every sample function (that is for almost every ω the function $t\rightsquigarrow x_t(\omega)$) becomes continuous except for nonoscillatory discontinuities. Such a change does not affect the joint distributions of finite sets of the random variables or the fact that the family of random variables is a supermartingale. The theory of continuous parameter supermartingales is very rich and plays an essential role in probabilistic potential theory. We give only one application here, a continuation of Sections 7 and 13. Let $\{z_t, 0\leqq t<\infty\}$ be Brownian motion in d-dimensional space, with initial point ξ: z_0 is identically ξ; for each $t\geqq 0$ the random variable z_t has its values in d-space; every sample function $t\rightsquigarrow z_t(\omega)$ is continuous. We omit the exact specification of the distributions. It turns out that, in the notation of Section 7, if $T_0=0$, if T_1 is the first time t when z_t is a point of $B(\xi)$, if T_2 is the first time $t>T_1$ when z_t is a point of $B(z_{T_1})$ and so on, and if x_n is defined as z_{T_n}, then the sequence $\{x_n, n\geqq 0\}$ is precisely the walk discussed in Sections 7 and 13. Thus a walk sample sequence is a sequence of points on a Brownian path. If S is the whole space and if u is a real harmonic function (or, if $d=2$, a complex analytic function) on S and if $|u|$ is not too large (we omit the precise restriction), then $\{u(z_t), t\geqq 0\}$ is a martingale. Suppose from now on that S is bounded. (The exactly appropriate condition is more generally that the complement of S has zero capacity.) Let $T=\sup_n T_n$ be the first time t at which z_t is on the boundary of S. Then it can be shown that T is almost surely finite, and of course z_T can be identified with x_∞ (defined in Section 13). Then the distribution of z_T is harmonic measure relative to the initial point of the Brownian motion. A refinement of the analysis in Section 13 shows that if u is harmonic and positive on S, and if $u(z_t)$ is defined as 0 when $t>T$, the process $\{u(z_t), t\geqq 0\}$ is a supermartingale. (This result is valid even if u is only superharmonic and positive, excluding the parameter value 0 if $u(\xi)=\infty$, and the discussion here can be generalized correspondingly.) From this result it is then concluded, using the existence of left limits of supermartingale sample functions, that u has a limit at the boundary of S along almost every Brownian path: $\lim_{t\uparrow T} u(z_t)$ exists almost surely. The limit theorem we have obtained is a probabilistic generalization of the classical FATOU THEOREM that if S is a disk, then any positive harmonic function u on S has a limit along almost every radius: if the disk radius is a and if polar coordinates are used, $\lim_{r\to a} u(re^{i\theta})$ exists Lebesgue for almost every θ. (The result was extended later to balls of arbitrary dimensionality.) Note that Fatou's Theorem is much more special because in his theorem S is a ball. On the other hand Fatou's paths are undeniably more pleasant, or at least more traditional, and are individually identifiable. The probability theorem, however, when applied to a ball, can be used to deduce Fatou's theorem, and in fact when boundary limit theorems are extended to cover superharmonic functions and more general classes of functions it becomes clear that the probabilistic versions are intrinsic, not accidents of the geometry of the domains.

References

1. J. L. Doob, Stochastic Processes, Wiley, New York, 1953.
2. Paul A. Meyer, Probability and Potentials, Blaisdell, Waltham, Mass., 1966.

MATHEMATICAL FOUNDATIONS FOR MATHEMATICS

LEON HENKIN,[1] University of California, Berkeley

1. Introduction. Most mathematical papers deal with mathematics "in the small"—a few definitions, a few theorems, a few proofs. If the author has a modicum of boldness and compassion he may also include some account of the intuitive ideas from which these formal parts of his work were fashioned. This paper, however, will have a different character.

In wondering what subject to choose for this Chauvenet Symposium, I let my mind's eye wander over those areas of the foundations of mathematics in which I have worked or dabbled—completeness proofs, applications of logic to algebra, decision problems, infinitary logic, algebraic logic . . . somehow none of them seemed appropriate. I began to wonder why. Presently it seemed to me that the answer was bound up with what might be called the "sociological structure" of our contemporary American mathematical community.

[1] The point of view toward foundations developed here was first formulated by me in the IBM Lectures which I gave at Swarthmore in December, 1967. This viewpoint has been developed over an extended period, during which much of my work was supported by the National Science Foundation (most recently, Grant No. GP-6232X).

This paper constitutes the final paper prepared for the Chauvenet Symposium held at the U. S. Naval Academy in October 1969 (this Monthly, May 1970).

Professor Leon Henkin received his PhD at Princeton University in 1947 under the direction of Alonzo Church. His thesis included a proof of Gödel's completeness theorem for the predicate calculus which has since become the standard proof in almost every presentation of mathematical logic. In addition Professor Henkin developed the theory of cylindrification algebras which is an algebraic formulation of the theory of quantifiers. His principal work has been in the area of foundations and mathematical logic in which he has published many papers and is a recognized authority. He was awarded the Chauvenet Prize in 1964 for his paper "Are Logic and Mathematics Identical?" published in Science, 1962 (vol 138).

Professor Henkin was Fine Instructor and Jewett Fellow at Princeton from 1947 to 1949, having spent four previous years as a mathematician in industry. In addition he was a Fullbright Scholar in Amsterdam in 1954–55, a Visiting Professor at Dartmouth in 1960–61, and a Guggenheim Fellow and member of the Institute For Advanced Study in 1961–62, and a visiting Fellow at All Souls' College, Oxford in 1968–69. He taught at the University of Southern California and has been a member of the faculty of the University of California at Berkeley since 1953 where he has served as chairman twice. He has been Editor for the *Journal of Symbolic Logic* and served a three year term as President for the Association of Symbolic Logic. He has also been a member of the Council of the American Mathematical Society as well as active in CUPM. Besides his papers in foundations he is the author of "Retracing Elementary Mathematics," Macmillan, 1962, which indicates his keen interest in the teaching of mathematics. *J. C. Abbott.*