

## PROBABILITY IN FUNCTION SPACE

J. L. DOOB

**Introduction.** The mathematical theory of probability is now ordinarily formulated in terms of measure theory. The formalization of a succession of  $n$  trials is in terms of  $n$ -dimensional measure; that of a sequence of trials is in terms of (denumerably) infinite-dimensional measure; that of a continuous set of trials (a set dependent on a continuous time parameter say) is in terms of measure in function space. The latter case is the subject of this lecture.

There still is a residual tendency among mathematicians to deal with distribution functions rather than with the chance variables which have those distribution functions; in this way probability theorems are reduced to theorems on monotone functions, convolutions, and so on. This preference was once justified by a natural distrust of the mathematical foundations of probability theory, so that any discussion of chance variables seemed to be nonmathematical, physics, or at best mathematical physics rather than pure mathematics, and therefore not rigorous. By now this distrust of chance variables is no longer justified; in practice, although it has almost disappeared when finite or denumerably infinite sequences of chance variables are studied, its presence is still noticeable in studies of continuous families of chance variables. Thus there is now no hesitancy in evaluating the probability that a series whose terms depend on chance will converge, but there is still considerable hesitancy in discussing the probability that a function whose values depend on chance should have a limiting value at  $\infty$ , be continuous, or be Lebesgue integrable. It is the purpose of the present lecture to dispel some inhibitions by summarizing the basic knowledge on measure theory in function space, as applied to probability. In the course of doing this, some problems still to be solved will be discussed. Although the theory is mathematically rigorous, it cannot yet be considered fully satisfactory. Even mathematicians, although they can live on rigor alone, do not invariably enjoy it.

The general point of view adopted in analyzing a succession of trials, depending on the parameter  $t$  (usually representing time), is the following. A set  $T$  of values of  $t$  is preassigned, the times when events occur, or measurements are made. This set of times may be finite or infinite. At each such moment a number  $x(t)$  is obtained.

---

An address delivered at the invitation of the Committee to Select Hour Speakers for Annual and Summer Meetings before the Summer Meeting of the Society on August 22, 1946; received by the editors August 23, 1946.

This number depends on chance; the value obtained comes from a probability distribution whose distribution function

$$(0.1) \quad P\{x(t) < k\}$$

is preassigned.<sup>1</sup> The process is fully specified by the complex of all possible joint distributions: if  $t_1, \dots, t_n$  is any finite subset of  $T$ , and if  $-\infty \leq a_j < b_j \leq +\infty$ , the probability

$$(0.2) \quad P\{a_j < x(t_j) < b_j, j = 1, \dots, n\}$$

is preassigned. All probabilities are evaluated in terms of these elementary probabilities. (For alternative approaches see §4.)

This *nonmathematical set-up* is treated mathematically as follows. Let  $\Omega$  be the space of all functions  $x(t), t \in T, -\infty \leq x(t) \leq +\infty$ ;  $\Omega$  is the set of all possible successions of sample values, that is, the set of all possible sample functions. The set of functions of  $t$  defined by the inequalities

$$(0.3) \quad a_j < x(t_j) < b_j, \quad j = 1, \dots, n,$$

is given as measure (probability) the number assigned in (0.2). In other words, a set function is defined on certain sets in function space, and probability problems become measure problems. A chance variable, that is a function of sample values, becomes a function defined on function space. The change in language from probability to measure language and vice versa neither increases nor decreases the difficulty of problems, but on the one hand measure theory is the precise mathematical form of the probability calculus, and on the other hand the empirical background of probability theory provides one type of external criterion of the interest and importance of a given measure problem. In the following, a non-negative completely additive function of sets,  $P\{E\}$ , defined on a Borel field  $F$  of sets in a space  $\Omega$ , with  $P\{\Omega\}$  defined and  $P\{\Omega\} = 1$ , will be called a probability measure. It is usually convenient to complete the measure by defining  $P\{E\}$  on any set  $E_1$  which differs from a set  $E_0$  in  $F$  by at most a subset of a set of measure 0 as  $P\{E_1\} = P\{E_0\}$ . The enlarged field will be denoted by  $\bar{F}$ .

The set of functions in function space defined by the inequalities (0.3) will be called a neighborhood, as will the set of functions defined by (0.3) with equality allowed on the right whenever  $b_j = \infty$  and on the left whenever  $a_j = -\infty$ . Function space is topologized by the convention that each neighborhood is a topological neighborhood of any

<sup>1</sup> The notation  $P\{ \}$  will be used to denote the probability of the event, or measure of the point set, defined by the braces.

one of its points. With this definition,  $\Omega$  becomes a bicomact Hausdorff space.

Let  $F_0$  be the Borel field of sets in function space  $\Omega$  generated by the neighborhoods (that is,  $F_0$  is the smallest Borel field of  $\Omega$  sets which includes the neighborhoods) and let  $F_2$  be the Borel field generated by the open sets;  $F_2$  is the field of Borel sets. Then  $F_0 \subset F_2$  and if  $T$  is finite or denumerably infinite,  $F_0 = F_2$ . On the other hand if  $T$  is the entire  $t$  axis, if  $I$  is any open interval, and if  $k$  is any real number, the function space sets

$$(0.4) \quad \left\{ \text{L.U.B. } x(t) \leq k \right\}_{t \in I} = \prod_{t \in I} \{x(t) \leq k\},$$

$$\left\{ \text{G.L.B. } x(t) > k \right\}_{t \in I} = \sum_{t \in I} \{x(t) > k\}$$

are respectively closed and open sets in function space, which are in  $F_2$  but not in  $F_0$ .<sup>2</sup> The two sets are respectively the intersection and sum of nondenumerably many sets in  $F_0$ .

Now suppose that  $T$  is finite. Then  $\Omega$  is simply  $n$ -dimensional coordinate space, and a neighborhood is an  $n$ -dimensional open rectangle with sides parallel to the axes. It has long been known that in this case if the non-negative function of  $a_1, \dots, a_n, b_1, \dots, b_n$  defined in (0.2) obeys simple conditions of monotoneity, and so on, then this function, considered as a function of  $\Omega$  neighborhoods, can be extended to become a probability measure defined on  $F_0 = F_2$ , the field of all the Borel sets of  $n$ -space. In this case, then, probability problems become problems in measure theory in  $n$ -space. Next suppose that  $T$  is denumerably infinite, as required for the study of sequences of chance variables. In this case  $\Omega$  becomes denumerably infinite-dimensional coordinate space. Kolmogoroff proved [7, p. 27]<sup>3</sup> that if the non-negative set function (0.2) is defined for every finite  $t$  set  $(t_1, \dots, t_n)$ , and determines a probability measure in  $x(t_1), \dots, x(t_n)$   $n$ -dimensional coordinate space, and if the set functions so obtained are mutually compatible, that is, if whenever  $\nu$  pairs  $(a_j, b_j)$  are set equal to  $(-\infty, +\infty)$  in (0.2) the  $(n-\nu)$ -dimensional set function obtained in the other  $n-\nu$  variables is exactly the preassigned set function for those variables, then just as in the finite-dimensional case, the domain of definition of the set function  $P\{E\}$  in (0.2) can be extended to include all the sets of  $F_0 = F_2$  (and hence

<sup>2</sup> In fact it is clear that any set in  $F_0$  is obtained by imposing conditions on  $x(t)$  at finitely or denumerably many  $t$  values.

<sup>3</sup> Numbers in brackets refer to the bibliography at the end of the paper.

of  $\bar{F}_0 = \bar{F}_2$ ), in such a way that  $P\{E\}$  becomes a probability measure.

In the finite and denumerably infinite-dimensional cases, the measure so defined has proved adequate for all the needs of the calculus of probabilities. The central limit theorem, laws of large numbers, and so on, can always be expressed in terms of the properties of measurable sets and functions on the coordinate space  $\Omega$ . The situation changes, however, if the parameter  $t$  has a continuous range. In the following,  $T$  will be the full real line. The changes necessary if  $T$  is a line segment or half line will be obvious. The coordinate space  $\Omega$  then becomes the space of functions defined on the line. As before, (0.2) defines the measure of a certain  $\Omega$  set, that is, of the set of functions satisfying the stated inequalities, and Kolmogoroff's theorem, which is independent of dimensionality, states that this measure can be extended to become a probability measure defined on all the sets of the Borel field  $F_0$ . It turns out, however, that this definition of probability measure is not entirely suited to the needs of the calculus of probabilities. In fact it was noted above that each  $F_0$  set is defined by restrictions on  $x(t)$  for at most countably many  $t$  values. On the other hand, in many investigations probabilities like that in (0.4) are required; that is, probabilities of sets of functions defined by restrictions on  $x(t)$  at a continuum of  $t$  values. Such sets may not be in  $F_0$  or even in  $\bar{F}_0$ . For example the sets in (0.4) are never in  $F_0$  and in general are not in  $\bar{F}_0$ . More precisely, if  $k < \infty$ , the inner  $P_0$  measure of the first set in (0.4) is always 0 [3, p. 465]. Hence this set must have measure 0 if it is  $P_0$  measurable. This fact implies that if L.U.B.  $_{t \in I} x(t)$  is  $P_0$  measurable, it must be  $+\infty$  almost everywhere. Similarly if G.L.B.  $_{t \in I} x(t)$  is  $P_0$  measurable, it must be  $-\infty$  almost everywhere. These cases can actually occur. For example let  $F(x)$  be any distribution function with  $0 < F(x) < 1$  for all  $x$ . Suppose that for each value of  $t$  the chance variable  $x(t)$  has the distribution function  $F$ , and that for any  $n$  values  $t_1, \dots, t_n$  the chance variables  $x(t_1), \dots, x(t_n)$  are mutually independent. In other words, as  $t$  varies the family of chance variables  $x(t)$  consists of mutually independent chance variables with the common distribution function  $F$ . Then clearly if  $s_1, s_2, \dots$  is any denumerable  $t$  set,

$$(0.5) \quad \begin{aligned} P_0 \left\{ \text{L.U.B. } _j x(s_j) \leq k \right\} &= 0, \\ P_0 \left\{ \text{G.L.B. } _j x(s_j) > k \right\} &= 0 \end{aligned}$$

for  $k \neq \pm \infty$ . The sets in (0.4) are included in sets of  $P_0$  measure 0 (take

the sequence  $\{s_j\}$  in  $I$ ) and therefore have measure 0. Aside from such rather uninteresting cases, however, L.U.B. $_{t \in I} x(t)$  and G.L.B. $_{t \in I} x(t)$  are never  $P_0$  measurable.

The insufficiency of  $P_0$  measure is illustrated by the following example. Suppose that  $x(t)$  is intrinsically less than or equal to  $k$ . In this simple case, any satisfactory mathematical theory must contain the statement in some form or other that the sample functions  $x(t)$  are less than or equal to  $k$  with probability 1. This is trivial if  $T$  is denumerable,  $T = \{t_j\}$ , because then one only need conclude from the hypothesis  $P_0\{x(t_j) > k\} = 0$  that

$$(0.6) \quad P_0 \left\{ \text{L.U.B.}_j x(t_j) > k \right\} = P_0 \left\{ \sum_j \{x(t_j) > k\} \right\} \\ \leq \sum_j P_0 \{x(t_j) > k\} = 0.$$

This reasoning cannot be used if  $T$  is nondenumerable, and it has just been seen that the desired result is not even true in that case. The situation can be saved, however, by defining a probability measure in the space  $\Omega_1$  of functions less than or equal to  $k$  rather than in the space of all functions, and Kolmogoroff's extension theorem is applicable to  $\Omega_1$  as well as to  $\Omega$ . The restriction to  $\Omega_1$  is no sacrifice, because  $\Omega_1$  is the significant space in the problem, but in more complicated problems it may not be clear what the significant space is. That is, the word "significant" must be defined in this context, and defined in such a way that  $P_0$  measure induces a probability measure in any significant space. This will be done in the next section.

**1. Extension of Kolmogoroff measure.** The problem of the introduction is simply to extend  $P_0$  measure by assigning probabilities in a natural way to function space sets (0.4). The fact that these sets are Borel sets in function space suggests the possibility of extending  $P_0$  measure to  $F_2$ , the class of all Borel sets. This suggestion will be discussed below, in §2. If probabilities have been assigned to the sets (0.4), then if  $\{s_j\}$  is any denumerable point set in the open interval  $I$ ,

$$(1.1.1) \quad P \left\{ \text{L.U.B.}_{t \in I} x(t) \leq k \right\} \leq P \left\{ \text{L.U.B.}_j x(s_j) \leq k \right\}, \\ P \left\{ \text{G.L.B.}_{t \in I} x(t) > k \right\} \geq P \left\{ \text{G.L.B.}_j x(s_j) > k \right\}$$

for all  $k$ . It is also clear that as more points  $s_j$  are inserted on the right the probabilities decrease or remain the same, and increase or remain the same respectively, and that an extremal sequence  $\{s_j\}$  can be

found such that these probabilities are left unchanged by the insertion of more points. It is then natural to *define* the probabilities on the left as these extreme values. These definitions were in fact proposed by Khintchine [6, p. 69] and it appears reasonable to require that any extension of  $P_0$  measure for the purposes of probability analysis should agree with his definitions on the sets (0.4).

As noted in the introduction, in probability problems of the type discussed the investigator is commonly interested in probabilities defined in function space subsets of a given set, for example in subsets of the set of continuous functions, of the set of positive functions, of the set of finite-valued functions, and so on. It was therefore natural that the first systematic method of treating the problems described in the introduction was to define a probability measure in the sets of a suitable subspace  $\Omega_1$  of function space  $\Omega$ , with  $\Omega_1$  chosen so that sets of functions, such as the set of continuous functions, which are not  $P_0$  measurable become measurable relative to  $\Omega$ . This is done as follows.<sup>4</sup> Define a (Kolmogoroff) measure  $P_0\{E\}$  on the Borel field  $F_0$  as described above. The Kolmogoroff outer measure of any set is defined by

$$(1.1.2) \quad P_0^*\{A\} = \text{G.L.B.}_{A \subset E \in F_0} P_0\{E\} = \text{G.L.B.}_{A \subset G} P_0\{G\}$$

where  $G$  is a finite or denumerably infinite neighborhood sum. Let  $\Omega_1$  be any subset of  $\Omega$  with  $P_0^*\{\Omega_1\} = 1$ . Define  $\bar{F}_0/\Omega_1$  as the field of all subsets  $E_1$  of  $\Omega_1$  of the form

$$(1.1.3) \quad E_1 = E_0\Omega_1, \quad E_0 \in \bar{F}_0,$$

and define a set function  $P\{E_1/\Omega_1\}$  on  $\bar{F}_0/\Omega_1$  by

$$(1.1.4) \quad P\{E_1/\Omega_1\} = P_0\{E_0\}.$$

In order to show that this definition is unique, we must show that  $E_0$  is determined up to a set of  $P_0$  measure 0 by (1.1.3). In fact, if  $E_0\Omega_1 = G_0\Omega_1$  ( $E_0, G_0 \in \bar{F}_0$ ), it follows that  $E_0\Omega_1 = E_0G_0\Omega_1$ , so that  $(E_0G_0')\Omega_1 = 0$ .<sup>5</sup> Then  $P_0\{E_0G_0'\} = 0$ , since  $P_0^*\{\Omega_1\} = 1$ , and similarly  $P_0\{E_0'G_0\} = 0$ , as was to be shown. The set function is clearly a probability measure.

It is possible, using this relativization method, to drop the consideration of the total space  $\Omega$  in favor of a subspace  $\Omega_1$ , if such a transfer

<sup>4</sup> Cf. [2]. The notation used in the present paper differs from that in [2], and is more adapted to a general discussion and to the methodological simplification described below.

<sup>5</sup> In the following, the complement of a set  $G$  will be denoted by  $G'$ .

is convenient. The condition that  $\Omega_1$  have outer measure 1 is the condition that the subspace be large enough. This provides a definition of "significance" as desired in the introduction. All probabilities will look the same in the subspace as in the full space, in the sense that, for example, the probability number in (0.2) remains unchanged, but is reinterpreted as the probability that the stated inequalities are true for the functions in  $\Omega_1$ . The following theorem indicates one advantage that may be gained by the shift from  $\Omega$  to a properly chosen  $\Omega_1$ .

**THEOREM 1.2.**<sup>6</sup> *There is a subset  $\Omega_1$  of  $\Omega$ , with  $P_0^*\{\Omega_1\} = 1$ , such that in terms of the relativized measure  $P\{E/\Omega_1\}$ , L.U.B. $_{t \in I} x(t)$  and G.L.B. $_{t \in I} x(t)$  are measurable functions for every interval  $I$ , and in fact there is a denumerable sequence  $\{t_j\}$  such that for  $x(t) \in \Omega_1$  and  $I$  any open interval*

$$(1.2.1) \quad \begin{aligned} \text{L.U.B.}_{t \in I} x(t) &= \text{L.U.B.}_{t_j \in I} x(t_j), \\ \text{G.L.B.}_{t \in I} x(t) &= \text{G.L.B.}_{t_j \in I} x(t_j). \end{aligned}$$

In other words, relative to  $\Omega_1$  the sets of functions in (0.4) can be defined using only denumerably many  $t$  values, and are therefore  $F_0$  sets. The  $t$  set  $\{t_j\}$  will in general depend on the basic Kolmogoroff measure, but in the most important cases any denumerable dense set will serve. Note that the adjunction of more points to the sequence  $\{t_j\}$  cannot affect (1.2.1) and hence the distribution of the chance variables is in accordance with Khintchine's definition given above.

The proof of Theorem 1.2 will be sketched. If  $I$  is any open interval, there is a denumerable  $t$  set  $\{t_j(I)\}$  in  $I$  such that for any sequence  $\{s_j\}$  in  $I$ ,

$$(1.2.2) \quad \begin{aligned} \text{L.U.B.}_j x(s_j) &\leq \text{L.U.B.}_j x(t_j(I)), \\ \text{G.L.B.}_j x(s_j) &\geq \text{G.L.B.}_j x(t_j(I)), \end{aligned}$$

excluding possibly a set of functions  $x(t)$  in  $\Omega$  of  $P_0$  measure 0. (The extremal sequences can be taken as those giving extreme values to the integrals over  $\Omega$  of  $\arctan \text{L.U.B.}_j x(s_j)$  and  $\arctan \text{G.L.B.}_j x(s_j)$  for sequences  $\{s_j\}$  in  $I$ .) Let  $\{t_j\}$  be any sequence containing all the points  $t_j(I)$  for all open intervals  $I$  with rational end points. Then for each  $t$ , almost all functions in  $\Omega$  ( $P_0$  measure) satisfy

<sup>6</sup> [3, Theorem 2.4].

$$(1.2.3) \quad \lim_{\epsilon \downarrow 0} \text{G.L.B.}_{|t_j-t|<\epsilon} x(t_j) \leq x(t) \leq \lim_{\epsilon \downarrow 0} \text{L.U.B.}_{|t_j-t|<\epsilon} x(t_j).$$

Define  $\Omega_1$  as the set of all functions  $x(t)$  satisfying (1.2.3) simultaneously for all  $t$ . (The set  $\Omega_1$  may contain infinite-valued functions.) With this definition,  $\Omega_1$  satisfies (1.2.1). In order to prove that  $P_0^*\{\Omega_1\} = 1$  we show that if  $\Gamma$  is any finite or denumerably infinite sum of neighborhoods, and if  $\Omega_1 \subset \Gamma$ , then  $P_0\{\Gamma\} = 1$ . Let  $\{\tau_j\}$  be the  $t$  values involved in the definition of  $\Gamma$ , and let  $\Gamma_0$  be the  $\Omega$  set of functions  $x(t)$  satisfying (1.2.3) for  $t = \tau_1, \tau_2, \dots$ . Then

$$(1.2.4) \quad \Omega_1 \subset \Gamma_0 \subset \Gamma.$$

But since (1.2.3) is true for  $t = \tau_n$ , with  $P_0$  measure 1, for each  $n$ ,  $P_0\{\Gamma_0\} = 1$ . Hence  $P_0\{\Gamma\} = 1$ , as was to be proved.

Using this theorem, it is clear that many of the problems involving a continuum of conditions on the functions of  $\Omega$  are reducible to problems involving denumerably many conditions, relativizing the measure so that  $P_0\{E\}$  is replaced by  $P\{E/\Omega_1\}$ . For example the set  $C$  of continuous functions, which is not measurable relative to  $\Omega$ , becomes measurable relative to  $\Omega_1$ ; in fact the equality

$$(1.2.5) \quad \Omega_1 C = \Omega_1 \prod_{n=1}^{\infty} \sum_{\nu=1}^{\infty} \left\{ \text{L.U.B.}_{|t_j-t_k|<1/\nu, |t_j|<n} |x(t_j) - x(t_k)| < 1/n \right\},$$

where the set  $\{t_j\}$  is that described in Theorem 1.2, shows that the relative measure of  $C$  is the  $P_0$  measure of the multiplier of  $\Omega_1$  on the right. It is Theorem 1.2 or its equivalent that justifies the avoidance of measurability problems (beyond  $P_0$  measurability) by considering the functions in function space as if they were defined only on some appropriate everywhere dense denumerable set.

This method of attack by means of relative measure, although it has proved usable in practice, is rather inelegant. The following equivalent approach is more elegant. It is desired to have certain sets measurable which are not so in  $\bar{F}_0$ . This can be accomplished by extending the domain of measurability rather than by relativizing the measure. Let  $\Omega_1$  be any  $\Omega$  set of outer  $P_0$  measure 1, and let  $F_1$  be the Borel field generated by  $\Omega_1$  and the sets of  $F_0$ . If  $\Omega_1$  is a Borel  $\Omega$  set,  $F_0 \subset F_1 \subset F_2$ . For example, if  $\Omega_1$  is chosen as described in Theorem 1.2, it is a Borel set, the intersection of denumerably many Borel sets of the form

$$(1.3.1) \quad \text{L.U.B.}_{t \in I} x(t) = \text{L.U.B.}_{t_j \in I} x(t_j), \quad \text{G.L.B.}_{t \in I} x(t) = \text{G.L.B.}_{t_j \in I} x(t_j).$$

Any set in  $F_1$  can be written in the form

$$(1.3.2) \quad E_1 = E_0\Omega_1 + F_0\Omega_1', \quad E_0, F_0 \in F_0$$

and we define

$$(1.3.3) \quad P_1\{E_1\} = P_0\{E_0\}.$$

This definition of  $P_1\{E_1\}$  is unique if  $E_0$  is determined up to a set of  $P_0$  measure 0 by (1.3.1); that is, if the equation  $E_0\Omega_1 = G_0\Omega_1$ , when  $E_0, G_0 \in F_0$ , implies that  $E_0$  and  $G_0$  differ by at most a set of  $P_0$  measure 0. This has already been shown in the course of defining relativized measure. In particular, the equation  $\Omega_1 = \Omega\Omega_1$  shows that  $P_1\{\Omega_1\} = 1$ , and if  $E_0 \in F_0$  the equation  $E_0 = E_0\Omega_1 + E_0\Omega_1'$  shows that  $P_1\{E_0\} = P_0\{E_0\}$ . Thus (1.3.3) defines an extension of  $P_0\{E\}$ , extending the domain of definition to  $F_1$ . The extension to  $\bar{F}_1$  can be made in the usual way, or simply by allowing  $E_0$  and  $F_0$  be to in  $\bar{F}_0$  in (1.3.2). If  $\Omega_1$  is chosen so that (1.2.1) is satisfied (for a proper sequence  $\{t_j\}$ ), all functions in  $\Omega_1$ , and hence almost all ( $P_1$  measure) functions in function space, satisfy (1.2.1). The functions on the left in (1.2.1) are therefore  $P_1$  measurable functions since those on the right are even  $P_0$  measurable. In practice  $\Omega_1$  is always chosen to have this property. Applying  $P_1\{E\}$  to (1.3.2) we have  $P_1\{E_1\} = P_1\{E_0\}$ ; thus to any  $P_1$  measurable set  $E$  corresponds a  $P_0$  measurable set  $E_0$  differing from it by at most a set of  $P_1$  measure 0. This implies that if  $f_1$  is any  $P_1$  measurable function, there is a  $P_0$  measurable function  $f_0$  differing from  $f_1$  on at most a set of  $P_1$  measure 0.

**2. Discussion of solution.** The definition of a stochastic process (in the present case, with the parameter  $t$  running through the line or half line) can now be formulated. A stochastic process is the combination of function space  $\Omega$  with a probability measure defined on the sets of  $\bar{F}_1$ . The set  $\Omega_1$  adjoined to the field  $F_0$  to obtain  $F_1$  has outer  $P_0$  measure 1, and is chosen to satisfy whatever conditions compatible with this are dictated by the needs of the problem at hand.

This definition is somewhat unfortunate in that it makes the definition of probability depend essentially on the rather arbitrary choice of the set  $\Omega_1$ . For example suppose in (0.2) that the probability is zero unless the  $a_j$ 's are all negative and the  $b_j$ 's all positive, that is we suppose that for each  $t$ ,  $x(t) = 0$  with probability 1. It is easy to prove that  $\Omega_1$  can be taken as the single function  $x(t) \equiv 0$ ; this function, considered as a set in function space, has outer  $P_0$  measure 1. If  $\Omega_1$  is so chosen, then in terms of the final probabilities ( $P_1$  measure),  $x(t) \equiv 0$  with probability 1, and this is the desired result, since  $P_0$  measure has been defined as it would be for an experiment in which it is certain that the observed value at each moment of time is 0. On the other

hand, it is easy to show that  $\Omega_1'$  also has outer measure 1, and if this set instead of  $\Omega_1$  is adjoined to  $F_0$  to get  $F_1$ ,  $\Omega_1$  will have measure 0; the probability that  $x(t) \equiv 0$  is now zero! A theory which allows this choice is hardly satisfactory, unless the choice is shown to be inevitable, when judicious rationalization can be used to make the choice even desirable.

What has been done in the past is to choose  $\Omega_1$  to consist of the most regular functions possible, as the set of continuous functions, or the set continuous except for jumps, and so on. This has usually been accomplished by choosing  $\Omega_1$  to satisfy the conditions of Theorem 1.2. Suppose for example that  $P_0$  measure has the property that whenever  $s_n \rightarrow s$ ,  $x(s_n) \rightarrow x(s)$  for almost all functions of  $\Omega$ . This is true in a wide class of cases. Then it is easy to see that *any* everywhere dense denumerable set  $\{t_j\}$  will serve in Theorem 1.2. Suppose that for at least one such set  $\{t_j\}$  it is possible to prove that the  $P_0$  measurable  $\Omega$  set

$$(2.1.1) \quad \prod_{n=1}^{\infty} \sum_{\nu=1}^{\infty} \left\{ \begin{array}{l} \text{L.U.B.} \\ |t_j - t_k| < 1/\nu, |t_j| < n \end{array} \quad |x(t_j) - x(t_k)| < 1/n \right\}$$

has  $P_0$  measure 1, that is, that almost all functions  $x(t)$  in function space coincide on  $\{t_j\}$  with everywhere continuous functions. Then if  $\Omega_1$  is chosen as in Theorem 1.2, and if this sequence  $\{t_j\}$  is used, the bounds of each  $x(t)$  in  $\Omega$  in any open  $t$  interval become the bounds considering only the points of  $\{t_j\}$  in the interval, so that  $x(t)$  must be everywhere continuous. That is, in terms of the extended measure,  $P_1$  measure, almost all functions  $x(t)$  are everywhere continuous. Moreover let  $\{t'_j\}$  be *any* everywhere dense denumerable  $t$  set. Then, in terms of the  $P_1$  measure already introduced, almost all functions  $x(t)$  coincide on the set  $\{t'_j\}$  with everywhere continuous functions. Hence in terms of  $P_0$  measure, almost all functions coincide on the set  $\{t'_j\}$  with everywhere continuous functions. In fact, as is seen in (2.1.1), this coincidence is  $P_0$  measurable. Hence no matter what sequence  $\{t_j\}$  is used in Theorem 1.2 to define  $\Omega_1$ , the end result will be, in terms of relativized measure or  $P_1$  measure, that almost all functions  $x(t)$  will be everywhere continuous. In probability language this is stated in the form " $x(t)$  is everywhere continuous, with probability 1." Thus, at least if  $\Omega_1$  is chosen to satisfy Theorem 1.2,  $P_1$  measure has important properties independent of the set  $\Omega_1$  adjoined to  $F_0$  to get  $F_1$ . It is important to note, however, that the set  $\Lambda$  of nonmeasurable functions in function space always has outer  $P_0$  measure 1. Hence even in this example, if  $\Lambda$  is adjoined to  $F_0$  to obtain  $F_1$ , in terms of

this  $P_1$  measure almost *no* function  $x(t)$  is measurable, much less continuous.

The choice of  $\Omega_1$  to satisfy conditions dictated by common sense and the desire for simplicity has been satisfactory in practice, but this is not mathematically satisfactory, since mathematics and common sense are not in the same universe of discourse. Moreover more complicated examples than those given might make the common sense decision somewhat difficult.

Two ways out of the difficulty are conceivable. One possibility would be to accept the fact that any one of the possible  $P_1$  measures corresponding to a given  $P_0$  measure is as justified as any other, from a purely mathematical point of view, and that some new principle must be introduced if a choice is to be made. For example, from a physical point of view it may be considered desirable to accept only the  $P_1$  measures obtained using an adjoined set  $\Omega_1$  which satisfies Theorem 1.2. The mathematician might invoke the principle of least action to choose  $\Omega_1$  so that Theorem 1.2 is satisfied and so that  $(t, \omega)$  measurability, as discussed in §3, is obtained, if possible.

A second and more elegant way out would be to define a measure on a field of  $\Omega$  sets large enough to fulfill the needs of all problems. For example it would be satisfactory if a single  $\Omega_1$ , that is, a single field  $F_1$ , would do for all problems, and it might appear that a choice of  $\Omega_1$  as in Theorem 1.2 might well suffice. This is not true; no single denumerable set  $\{t_j\}$  is large enough, although any everywhere dense set suffices for most problems. Kakutani, in conversations with the speaker before the war, proposed extending  $P_0$  measure from  $F_0$  to  $F_2$ , the field of Borel sets of function space. When  $T$ , the range of  $t$ , is a finite or denumerable aggregate, the Borel  $\Omega$  sets have been seen above to be a sufficiently large class. The extension to  $F_2$  can be done as follows. Let  $G$  be any open  $\Omega$  set. Kakutani then defines  $P_2^*\{G\}$  by

$$(2.2.1) \quad P_2^*\{G\} = \text{L.U.B. } P_0\{E\}, \quad E \in F_0, \\ E \subset G$$

That is,  $P_2^*\{G\}$  is the inner  $P_0$  measure of  $G$ . If  $A$  is any  $\Omega$  set,  $P_2^*\{A\}$  is defined by

$$(2.2.2) \quad P_2^*\{A\} = \text{G.L.B. } P_2^*\{G\} \\ G \supset A$$

where  $G$  is open. It can be proved that  $P_2^*\{A\}$  has the usual properties of an outer measure. The (Carathéodory) measurable sets include all Borel sets and the measure,  $P_2$  measure, is an extension of  $P_0$  measure. It agrees with Khintchine's definition on sets of the type (0.4).

It would then appear that Kakutani's extension of Kolmogoroff measure has all the desired properties, but this is not quite true, or at least it has not been proved. To see what is lacking, the measurability of a process will be discussed.

**3. Measurable processes.** Let  $f(t, \omega)$  be a function of the two variables  $t, \omega$ , where  $t$  varies on a line or half line, and  $\omega$  varies in an abstract space. Then for each value of  $\omega$ ,  $f(t, \omega)$  is a function of  $t$ . Let  $\Omega_1$  be the aggregate of functions of  $t$  obtained in this way. If an  $\omega$  measure is defined, it induces a measure in function space, or more precisely, in subsets of  $\Omega_1$ . Assume that  $\omega$  measure is a probability measure, and that  $f(t, \omega)$  is measurable in  $\omega$  for each fixed value of  $t$ . Then if  $(t_1, \dots, t_n)$  is any finite  $t$  set, the  $\omega$  measure of the  $\omega$  set

$$(3.1.1) \quad \{a_j < f(t_j, \omega) < b_j, j = 1, \dots, n\}$$

can be identified with the probability, or measure, of the set of functions in function space defined by the inequalities (0.3). This approach was used to give rigorous meaning to measure in function space (or at least to measure in some of the subsets of a subset  $\Omega_1$  of function space) long before the direct attacks of the preceding section were developed. For purposes of the ergodic theorem, harmonic analysis, and similar work, where some sort of regularity in  $t$  was required, it was supposed, as a minimum condition, that  $f$  was measurable in the two variables  $t, \omega$ . That is, a measure in  $(t, \omega)$  space was defined in the usual multiplicative way as the direct product of Lebesgue  $t$  measure and the given  $\omega$  measure, and  $f$  was supposed  $(t, \omega)$  measurable. Note that the measure of a neighborhood in function space can be identified with the  $\omega$  measure of the set (3.1.1) so that this approach in terms of a function of  $t$  and an auxiliary probability parameter  $\omega$  actually determines a measure in function space  $\omega$ . More precisely,  $P_0$  probability measure is determined in the  $\Omega$  sets of the field  $F_0$  (Kolmogoroff's theorem). In terms of this  $P_0$  measure it is easily shown that  $\Omega_1$  has outer measure 1 [4, Theorem 1], so that a relativized measure can be defined on the sets of  $\bar{F}_0/\Omega_1$  as explained in §1. On the other hand, the given  $\omega$  measure determines a measure on  $\Omega_1$  sets by the map  $\omega \rightarrow x(t) = f(t, \omega)$ . By definition, these measures are identical on neighborhoods (relative to  $\Omega_1$ ) and therefore the relativized  $\Omega_1$  measure and mapped  $\Omega$  measures are identical on the sets of  $F_0/\Omega_1$ . It is frequently very convenient to discuss a stochastic process using an explicit formula  $x(t) = f(t, \omega)$  of this type for the sample functions. This has been frequently done in the past with  $f$  the sum of a series of functions of  $t$  with random coefficients. But it must be remembered that corresponding to the fact that a given  $P_0$  measure

in function space, the basic probability measure, may lead to many quite different extensions, so there may be many functions  $f(t, \omega)$  of two variables with quite different measure properties even though there are identical probability evaluations for inequalities of the type (3.1.1). In fact it will be seen below that each  $P_1$  measure leads to a function  $f(t, \omega)$ . Hence, corresponding to the different extensions of  $P_0$  measure in an example already discussed, two functions  $f_1(t, \omega)$ ,  $f_2(t, \omega)$  can be exhibited, with  $f_1(t, \omega) \equiv 0$  in  $t, \omega$ , and  $f_2(t, \omega) \not\equiv 0$  in  $t$  for any  $\omega$ , which agree as far as the probabilities of inequalities of the type (3.1.1) are concerned. Using  $f_1$  one would say that  $x(t) \equiv 0$  with probability 1; using  $f_2$  one would say that  $x(t) \equiv 0$  with probability 0; yet the basic probabilities are identical. One must watch out for such things in using these two variable representations of stochastic processes. The representations may contain implications not inherent in the basic probabilities.

Conversely any stochastic process can be put in the form of a two variable representation. In fact, denote the variable function  $x(t)$  in  $\Omega$  by  $\omega$ , and consider the product space of pairs  $(t, \omega)$ . Define the numerically valued function  $x(t, \omega)$  as the function of  $(t, \omega)$  which takes on the value  $x_0(t_0)$  when  $t = t_0$  and  $\omega = x_0(s)$ . Note that for each  $t$ ,  $x(t, \omega)$  is a measurable function of  $\omega$  ( $P_0, P_1$ , or  $P_2$  measurable, depending on which field of measurable sets is chosen); in fact  $x(t_0, \omega)$  has the distribution function  $P\{x(t_0) < k\}$ . The function  $x(t, \omega)$  thus has the role of the function  $f(t, \omega)$  above. Note that given a  $P_1$  measure, the function  $x(t, \omega)$  will have measure properties which may differ markedly from those due to another  $P_1$  measure extension of the same  $P_0$  measure. The consequences of this have already been commented on in the preceding paragraph.

The proper definition of measurability of a process is now clear; a process is measurable if  $x(t, \omega)$  is a measurable function of  $(t, \omega)$ , where  $(t, \omega)$  measure is defined as the direct product of Lebesgue measure in  $t$  and the given  $\omega$  measure. If a process is measurable, Fubini's theorem can be applied, to find that for almost all  $\omega$  (that is, with probability 1)  $x(t, \omega)$  is measurable in  $t$ ; in other words almost all functions in function space are measurable. The functions  $x(t)$ , that is  $x(t, \omega)$ , can then be integrated (in  $t$ ) if the function values are not too large. For example, if  $E\{|x(t)|\}$  is a bounded function of  $t$ , that is if the integral in  $\omega$  of  $|x(t, \omega)|$  is a bounded function of  $t$ , the integral  $\int_a^b x(t) dt = \int_a^b x(t, \omega) dt$  exists in the usual Lebesgue sense for almost all functions  $\omega$  in function space, and is itself a function of  $\omega$ ,  $P_0, P_1$ , or  $P_2$  measurable as the case may be, according to the  $\Omega$  measure used. The indirect definitions of this integral commonly found

<sup>7</sup> The expectation of a chance variable  $y$  will be denoted by  $E\{y\}$ .

in the literature, usually as some sort of averaged limit of Riemann sums, beg the question of integrability of the individual functions and complicate the mathematics.

Unfortunately three definitions of measurability of a stochastic process have been proposed in the preceding paragraph, according as the field of measurable  $\Omega$  sets is taken as  $\bar{F}_0$ ,  $\bar{F}_1$ , or  $\bar{F}_2$ . The first possibility is easily eliminated. In fact, no process is ever measurable if  $\bar{F}_0$  is used, for it was just noted that almost all functions  $x(t)$  are Lebesgue measurable if a process is measurable, whereas it is easily shown [2, p. 113] that the  $\Omega$  set of Lebesgue nonmeasurable functions has outer  $P_0$  measure 1. This is one more indication of the insufficiency of the field  $\bar{F}_0$ . Note that measurability using  $\bar{F}_1$  depends on the adjoined set  $\Omega_1$  in the sense that a process may be measurable or not depending on the choice of  $\Omega_1$ . For example if  $\Omega_1$  is the class of non-Lebesgue measurable functions, no process can be measurable using  $\Omega_1$ , according to the argument given above for  $\bar{F}_0$ . We shall accept the definition that a process is measurable if  $x(t, \omega)$  is  $(t, \omega)$  measurable if  $\bar{F}_1$  is used, for some choice of  $\Omega_1$ .<sup>8</sup>

It was shown in the beginning of this section how a function  $f(t, \omega)$  of  $t$  and an auxiliary probability parameter  $\omega$  can be used to determine a stochastic process. For a full parallelism it must be shown that whenever  $f(t, \omega)$  is  $(t, \omega)$  measurable, there is a corresponding measurable process, that is, the following theorem must be proved.

**THEOREM 3.1** [4, Theorem 6]. *Let  $f(t, \omega)$  be a function of  $(t, \omega)$ , where  $t$  varies on a line and  $\omega$  varies in any abstract space in which a probability measure is defined,  $m_\omega\{ \}$ . Suppose that  $f(t, \omega)$  is measurable in  $\omega$  for each fixed  $t$  and also measurable in  $(t, \omega)$ . Define a stochastic process  $P_0$  measure, a probability measure in function space, by setting*

$$(3.1.2) \quad \begin{aligned} P_0\{a_j < x(t_j) < b_j, j = 1, \dots, n\} \\ = m_\omega\{a_j < f(t_j, \omega) < b_j, j = 1, \dots, n\}. \end{aligned}$$

*Then this  $P_0$  measure can be extended to a  $P_1$  measure in terms of which the stochastic process is measurable.*

*Note 1.* This theorem is not true if the measurability of a stochastic process is defined in terms of  $P_0$  measure in function space; an extension of  $P_0$  measure is definitely needed, even in so regular an example as when  $f(t, \omega) \equiv 0$ , so that  $x(t) = 0$  for each  $t$ , with probability 1. On the other hand it is not known as yet (see the discussion below)

<sup>8</sup> This is equivalent to the definition in terms of relativized measure given in [2]. Necessary and sufficient conditions on the basic  $P_0$  measure that there be a  $P_1$  extension determining a measurable process have been given in [1] and [2].

whether the extension to  $P_2$  measure would always be sufficient.

*Note 2.* This theorem shows that no new types of measurable process would arise if extensions of  $P_1$  measure were used in the definition of measurability of a stochastic process. That is to say, if  $P_0$  measure is regular enough to make  $x(t, \omega)(t, \omega)$  measurable using any  $\Omega$  measure whatever which is an extension of the given  $P_0$  measure, then  $x(t, \omega)$  is  $P_1$  measurable also, using a properly chosen extended field  $\bar{F}_1$ . In fact if  $x(t, \omega)$  is  $(t, \omega)$  measurable using as  $\Omega$  measure some extension of the given  $P_0$  measure, then, using  $x(t, \omega)$  as the  $f(t, \omega)$  of the theorem, it follows that there is a  $P_1$  measure making  $x(t, \omega)(t, \omega)$  measurable.

The proof of this theorem, which will be omitted, consists in finding a set  $\Omega_1$  to adjoin to  $F_0$ , such that, in terms of  $P_1$  measure,  $x(t, \omega)$  is  $(t, \omega)$  measurable. This set  $\Omega_1$  can be chosen to satisfy the conditions of Theorem 1.2, so that measurability in  $(t, \omega)$  and these desirable conditions are always mutually compatible.

Unfortunately it has never been demonstrated that the  $\Omega_1$  described in the preceding paragraph can be chosen to be a Borel  $\Omega$  set. Thus the proof does not show that if the condition of the theorem is satisfied,  $x(t, \omega)$  will be  $(t, \omega)$  measurable when  $\Omega$  measure is  $P_2$  measure. In order that Kakutani's approach to these stochastic processes, using  $P_2$  measure, be fully satisfactory, it is necessary that whenever  $x(t, \omega)$  is  $(t, \omega)$  measurable using  $P_1$  measure, it should also be  $(t, \omega)$  measurable using  $P_2$  measure. It would be very desirable to have either a proof or a counterexample. At present, therefore, Kakutani's approach must be considered incomplete, although very elegant, and quite adequate in most problems. In [5] Kakutani exhibited a class of processes in which  $\bar{F}_0 = \bar{F}_2$ .

**4. Alternative approach.** The whole difficulty of defining a suitable probability measure in function space has been caused by the insistence on assigning measure (probability) to a neighborhood of functions, as in (0.2), first and going on from there. This automatically limits one to  $P_0$  measure and its extensions. An approach that would avoid all this would be to suppose right from the start that there is a probability measure defined on some field  $F$  of  $\Omega$  sets, and go on from there. The first and most obvious difficulty in this is that one must avoid being too general if theorems are to have any significance beyond their statement and proof. This means that  $F$  must be fairly strictly defined. No specific proposals in this direction have been made, unless Kakutani's proposal is considered in this light. In the second place  $P_0$  measure arises in a natural way because the many stochastic processes that are known and interesting are, at present,

all defined in terms of what happens at a finite number of  $t$  values. For example the processes with independent increments are simply described by the statement that if  $t_1 < \dots < t_n$ ,  $x(t_2) - x(t_1), \dots, x(t_n) - x(t_{n-1})$  are mutually independent. Moreover besides this simple description, it is possible to write down explicitly the most general distribution of the chance variables  $x(t_1), \dots, x(t_n)$  for any such process, and in this way describe the measure of the neighborhoods. Thus in this and other examples, up to the present at least, the approach through neighborhood measures to  $P_0$  measure and its extensions has been the natural one. It is conceivable, however, that new  $\Omega$  measures will be developed starting from characteristics of the curves as a whole, rather than their characteristics at finite sets of values of the argument.

#### BIBLIOGRAPHY

1. W. Ambrose, *On measurable stochastic processes*, Trans. Amer. Math. Soc. vol. 47 (1940) pp. 66–79.
2. J. L. Doob, *Stochastic processes depending on a continuous parameter*, Trans. Amer. Math. Soc. vol. 42 (1937) pp. 107–140.
3. ———, *Regularity properties of certain families of chance variables*, Trans. Amer. Math. Soc. vol. 47 (1940) pp. 455–486.
4. J. L. Doob and W. Ambrose, *On two formulations of the theory of stochastic processes depending upon a continuous parameter*, Ann. of Math. vol. 41 (1940) pp. 737–745.
5. Shizuo Kakutani, *Notes on infinite product measure spaces*, II, Proc. Imp. Acad. Tokyo vol. 19 (1943) pp. 184–188.
6. A. Khintchine, *Asymptotische Gesetze der Wahrscheinlichkeitsrechnung*, Ergebnisse der Mathematik, vol. 2, no. 4.
7. A. Kolmogoroff, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Ergebnisse der Mathematik, vol. 2, no. 3.

UNIVERSITY OF ILLINOIS