# COMPUTABLE PROBABILITY SPACES

BY

BAYARD RANKIN

*Case Institute of Technology, Cleveland, Ohio, U.S.A.*(1)

## 1. Introduction

One use of the axioms and definitions of this paper is found in the study of certain subclasses of random variables, for example those defined on a probability space ($R$, $\mathcal{B}$, $P$), $R$ real, that are finite and continuous almost everywhere as contrasted with finite and almost continuous (i.e. measurable). With this avenue open all expectations $EX$ of some important subclasses of random variables can be evaluated as asymptotic averages:

$$EX = \lim \frac{1}{N} \sum_{1}^{N} X(a_k) \tag{1}$$

on a single fixed sequence $S \equiv (a_1, a_2, a_3, \ldots)$, or obtained as simple extensions of such evaluations. In order that these subclasses can be studied without considering their behavior off the fixed sequence $S$—which behavior seems inconsequential from the probabilistic point of view—it is necessary to include them in a setting more general than measure-theoretic probability theory. This is the case because the class of functions for which (1) converges on a sequence $S$ may be more general than any class of random variables on a probability space of the form ($S$, $\mathcal{B}$, $P$) for which (1) is their mathematical expectation. The motivations for this approach are as follows.

Historically and intuitively the expectation $EX$ of a random variable $X$ is equated with an asymptotic average of the form (1), the variable $X$ being considered as a function on a sequence $S \equiv (a_1, a_2, a_3, \ldots)$ of elementary events or sample points. This equality is recovered in the strong law of large numbers and the ergodic theorem of modern measure-theoretic probability. In the modern version, of course, the points $a_k$ are obtained as images

(1) Acknowledgement: This paper has been made possible by a grant from the Case Research Fund and by the added support of the Case Computing Center.

of a point $a_1$ under a group of transformations on a probability space. The group must have special properties to insure that the equality holds and the initial point $a_1$ must be chosen outside a measurable set $M$ of probability measure zero, where $M$ in general depends on $X$.

The whole modern reformulation of the equality (1) is a natural one once a random variable is defined as a finite measurable function on a probability space. There are limitations, however. Primarily, it is intuitively unsatisfactory that the sequence $S$ for which (1) holds should be different for different random variables (a consequence of the fact that the exceptional set $M$ depends on $X$). One quite naturally asks, "Is there a probability space $(S, \mathfrak{B}, P)$ in which $S \equiv (a_1, a_2, a_3, \ldots)$ is a single abstract sequence and on which (1) holds for all random variables $X$?" The answer to the question is in the affirmative. At the same time any random variable on such a probability space must have at most a countable number of values (because $S$ is countable) and thus a distribution function which is at most a step function.

Another limitation is the absence of any computable structure in the space $S$ of elementary events, leaving only the expectation attainable by computable methods but not the sample values of the random variables. Again, one quite naturally asks, "Is there a probability space $(S, \mathfrak{B}, P)$ in which $S$ is composed of a subset of the real numbers computable in the Turing sense?" The answer is once again in the affirmative. At the same time, because there are at most countably many computable numbers, any random variable on such a probability space must again be elementary.

Thus, while the answers are not trivial—and, in fact, will require a detailed explanation—we are twice led to answers of limited generality after posing questions of basic interest. It is our present purpose to find answers of greater generality by asking the same two questions with more flexible phrasing. Inevitably this means formulating and accepting a flexible probability theory that augments the measure-theoretic approach.

## 2. The elementary case

Let $(S, \mathfrak{B}, P)$ be a probability space and let $I_A$ be an indicator function of the set $A \in \mathfrak{B}$ (i.e., a function defined on $S$ that takes the value 1 on $A$ and 0 elsewhere).

DEFINITION 1. *$(S, \mathfrak{B}, P)$ is called a sequence probability space if $S \equiv (a_1, a_2, a_3, \ldots)$ is an abstract sequence and*

$$PA = \lim \frac{1}{N} \sum_{1}^{N} I_A(a_k) \tag{2}$$

*for all $A \in \mathfrak{B}$.*

THEOREM 1. *There exists a sequence probability space* $(S, \mathcal{B}, P)$ *in which* $\mathcal{B}$ *is uncountably infinite.*

The proof of this theorem is accomplished by construction and is inspired, as is much of this paper, by H. Weyl's work on uniformly dense sequences [9]. A sequence $S \equiv (a_1, a_2, a_3, \ldots)$ is uniformly dense in the closed unit interval [0, 1] if for all intervals $i$ contained in [0, 1]

$$\mu(i) = \lim \frac{1}{N} \sum_1^N I_i(a_k) \tag{3}$$

$\mu$ being the Lebesque measure. Weyl has proved, for example, that the sequence $S \equiv (a_1, a_2, a_3, \ldots)$ defined by $a_k = \alpha k \pmod 1$, $\alpha$ irrational, has the property of being uniformly dense in [0, 1].

To construct a sequence probability space, let $S$ be a uniformly dense sequence in [0, 1]. Let $\Pi$ be a countable partition of [0, 1] composed of non-trivial intervals and let $\mathcal{B}$ be constructed by intersecting with $S$ the smallest $\sigma$-algebra, $\mathcal{A}$, of sets in [0, 1] containing $\Pi$: $\mathcal{B} = \mathcal{A} \cap S$. Every member $A'$ of $\mathcal{A}$ is at most a countable union of non-trivial, disjoint intervals and $\mathcal{A}$ is in one-to-one correspondence with $\mathcal{B}$. Because $A'$ is a countable union of intervals there are for $\varepsilon > 0$ two members $\bar{A}'$, $\underline{A}'$ of $\mathcal{A}$, each composed of finite disjoint unions of intervals such that

$$\text{i) } \underline{A}' \subset A' \subset \bar{A}'$$

$$\text{ii) } \mu\bar{A}' - \mu\underline{A}' < \varepsilon.$$

Thus,

$$\overline{\lim} \frac{1}{N} \sum_1^N I_{A'}(a_k) < \lim \frac{1}{N} \sum_1^N I_{\underline{A}'}(a_k) + \varepsilon = \mu\underline{A}' + \varepsilon$$

$$\underline{\lim} \frac{1}{N} \sum_1^N I_{A'}(a_k) > \lim \frac{1}{N} \sum_1^N I_{\bar{A}'}(a_k) - \varepsilon = \mu\bar{A}' - \varepsilon$$

and it follows that

$$\mu A' = \lim \frac{1}{N} \sum_1^N I_{A'}(a_k).$$

Because of the one-to-one correspondence between $\mathcal{A}$ and $\mathcal{B}$ we can define $PA = \mu A'$ for $A = A' \cap S$ with the consequence that for all $A \in \mathcal{B}$

$$PA = \lim \frac{1}{N} \sum_1^N I_A(a_k).$$

$\mathcal{B}$ is a $\sigma$-algebra of sets in $S$ and $P$ is a measure. $(S, \mathcal{B}, P)$ is then a sequence probability space. Q.E.D.

The property of one-to-one-ness between $\mathcal{A}$ and $\mathcal{B}$ played an essential role in the above proof and permitted the transfer of a measure $\mu$ on sets in [0, 1], to a measure $P$ on sets in $S$. If the $\sigma$-algebra $\mathcal{A}$ had been generated not from $\Pi$ but from the class of all intervals in [0, 1] the argument would have broken down. This is a point which will have bearing on the ways in which one can and cannot generalize the idea of a sequence probability space.

The theorem just proved tells in what sense a sequence probability space is non-trivial. The next theorem tells in what sense the concept is limited.

THEOREM 2. *Let $X$ be a random variable defined on the sequence probability space $(S, \mathcal{B}, P)$ and let $F_X(x)$ be its distribution function.*

*$F_X(x)$ is a step function.*

This theorem, which may be considered obvious, requires a little discussion. It is recalled that in the proof of the last theorem the $\sigma$-field, $\mathcal{B}$, in a sequence probability space could not be too rich, that is, could not include too many subsets of $S$. In particular, $\mathcal{B}$ could not include all points $a_k$ of an infinite sequence $S$ of distinct points, for in such a case $Pa_k = 0$ for each $k$, because of (2), and $PS = \sum Pa_k = 0$— a contradiction.

This proves that there is no random variable on a sequence probability space which has a different value at each point of $S$, unless $S$ has finitely many points.

The present theorem says more, namely, that each random variable on a sequence probability space is equivalent to one that induces a partition of $S$ into sets of positive measure, the elements of the partition being of the form $A = \{a_k : X = \text{const.}\}$. The proof, on the other hand, does not require the special form of measure given by (2), but is based simply on the fact that the range of a random variable on a sequence probability space is a countable set and no non-trivial continuous distribution function can have points of increase only on such a set. The proof follows.

*Proof.* Because $S$ is countable, $X$ can have at most a countable number of values. From the decomposition of distribution functions

$$F_X(x) = F_X^{(d)}(x) + F_X^{(s)}(x)$$

where $F_X^{(d)}(x)$ is a step function and $F_X^{(s)}(x)$ is a singular function which is continuous with points of increase belonging to a countable set. The measure $\mu^{(s)}$ on the real line that corresponds to $F_X^{(s)}(x)$ assigns the measure zero to every countable set andthus $F_X^{(s)}(x) \equiv 0$. Q.E.D.

The two theorems to follow display the special advantages of sequence probability spaces. For one of these advantages having to do with computability, it is convenient to define some notions.

THEOREM 3. *If $X$ is a bounded random variable defined on the sequence probability space $(S, \mathcal{B}, P)$, then*

$$EX = \lim \frac{1}{N} \sum_{1}^{N} X(a_k).$$

*Proof.* Let $\Pi_X$ be the partition of $S$ induced by $X$. That is $A \in \Pi_X$ if and only if for some real $r$,

$$A = \{a_k : X = r\}.$$

We may assume without loss of generality that $PA > 0$ for all $A$.

There exists for $\varepsilon > 0$ a set $A_\varepsilon$ composed of a finite union of elements of $\Pi_X$ such that $PA_\varepsilon > 1 - \varepsilon$. This can be demonstrated by letting $A_k$ be the union of all $A \in \Pi_X$ for which

$$2^{-(k-1)} \geqslant PA > 2^{-k}.$$

For each $k$, $A_k$ is at most a finite union sets in $\Pi_X$ and $\bigcup_{1}^{\infty} A_k \equiv S$. Because $P$ is competely additive, for some $M$ depending on $\varepsilon$, $P[\bigcup_{1}^{M} A_k] > 1 - \varepsilon$.

Now define for $\varepsilon > 0$

$$\overline{X}(a_k) = \begin{cases} X(a_k) & \text{for } a_k \in A_\varepsilon \\ B & \text{for } a_k \notin A_\varepsilon \end{cases}$$

$$\underline{X}(a_k) = \begin{cases} X(a_k) & \text{for } a_k \in A_\varepsilon \\ -B & \text{for } a_k \notin A_\varepsilon \end{cases}$$

where $B$ is any real number such that $|X| < B < \infty$. It follows that

i) $$\underline{X} \leqslant X \leqslant \overline{X}$$

ii) $$E\overline{X} - E\underline{X} \leqslant 2B\varepsilon$$

and since $\overline{X}$ and $\underline{X}$ have at most a finite number of values

iii) $$E\overline{X} = \lim \frac{1}{N} \sum_{1}^{N} \overline{X}(a_k)$$

iv) $$E\underline{X} = \lim \frac{1}{N} \sum_{1}^{N} \underline{X}(a_k).$$

Therefore,

$$\overline{\lim}\ \frac{1}{N}\sum_{1}^{N} X(a_k) \leqslant E\overline{X} \leqslant E\underline{X} + 2B\varepsilon \leqslant EX + 2B\varepsilon$$

$$\underline{\lim}\ \frac{1}{N}\sum_{1}^{N} X(a_k) \geqslant E\underline{X} \geqslant E\overline{X} - 2B\varepsilon \geqslant EX - 2B\varepsilon.$$

Q.E.D.

DEFINITION 2. *A sequence $S = (a_1, a_2, a_3, \ldots)$ is said to be computable, if it is real and there is a Turing machine*(1) *which, for any pair of positive integers $M$, $N$, will print out in order the first $M$ digits of all $a_k$, $k = 1, 2, \ldots, N$ in a finite number of steps.*

DEFINITION 3. *A sequence probability space $(S, \mathcal{B}, P)$ is called a computable probability space, if $S$ is computable.*

Because any Turing machine can be simulated by a modern all-purpose computer the notion of a computable probability space makes possible the investigation of whether the sample values of a sequence of statistically independent random variables can be computed. Such an investigation might yield results for actual statistical calculations, where at present random numbers generated by physical processes or pseudo-random numbers generated by computers must be used. Such an investigation might also yield results for the type of processes studied in statistical mechanics. More will be said about these implications later.

DEFINITION 4. *A random variable $X$ defined on the computable probability space $(S, \mathcal{B}, P)$ is called a computable random variable, if the sequence of sample values $X(a_1)$, $X(a_2)$, $X(a_3)$, ... is computable.*

DEFINITION 5. *A sequence of random variables $X_1, X_2, X_3, \ldots$ defined on the computable probability space $(S, \mathcal{B}, P)$ is called computable, if there is a Turing machine which, for any triplet of positive integers $M$, $N$, $Q$, will print out in some specified order the first $M$ digits of all $X_i(a_j)$, $j = 1, 2, 3, \ldots, N$, $i = 1, 2, 3, \ldots, Q$ in a finite number of steps.*

THEOREM 4. *There exists a computable probability space $(S, \mathcal{B}, P)$ in which $\mathcal{B}$ is uncountably infinite.*

The proof is the same as for the existence of a sequence probability space with the added remark that the sequence $a_k = \alpha k \pmod 1$ $k = 1, 2, 3, \ldots$, on which the space is

(1) For the definition of a Turing machine see [3, 8].

built, is computable if the irrational $\alpha$ is computable. Such an $\alpha$ might be chosen to be $\pi$, $e$ or $\sqrt{2}$.

THEOREM 5. *There exists for each positive integer $N$ a computable sequence of statistically independent random variables $X_1, X_2, \ldots X_N$ defined on a computable probability space $(S, \mathcal{B}, P)$.*

For a proof, it is sufficient to let $S \equiv (a_1, a_2, a_3, \ldots)$ be the uniformly dense sequence in the unit interval defined by $a_k = \pi k \pmod 1$. For given integer $N$, let $\mathcal{B}$ be the smallest $\sigma$-algebra containing all sets $A_{j,N}$ of the form

$$A_{j,N} = \left\{ a_k : \frac{j}{2^N} \leqslant a_k < \frac{j+1}{2^N} \right\}, j = 0, 1, \cdots, 2^N - 1$$

and let $P$ be defined by (2).

The $N$ identically distributed random variables

$$X_i = \sum_{j=0}^{2^{i-1}-1} I_{A_{2j,i}}, \; i = 1, 2, \cdots, N$$

are statistically independent and computable. The independence follows from direct computation and the computability follows from the fact that for each $k$ and each $i$ a finite number of digits of $a_k$ determines whether $X_i(a_k)$ has value 0 or 1. Q.E.D.

As will be seen later it is only for simplicity and not out of necessity that for the proof we have chosen to construct random variables which have only two values. In fact, the proof of the following theorem will be obvious from later results.

THEOREM 6. *For any distribution function $F_X(t)$ (of a random variable) with at most a finite number of points of increase, and for any positive integer $N$, there exist $N$ identically distributed, statistically independent, computable random variables $X_1, X_2, \ldots, X_N$ defined on a computable probability space $(S, \mathcal{B}, P)$ such that $F_{X_k}(t) = F_X(t)$, $k = 0, 1, \ldots, N$.*

As a special sequence that leads to simple computations the following sequence of rationals, $S^0 \equiv (a_1{}^0, a_2{}^0, a_3{}^0, \ldots)$ that is uniformly dense in the unit interval will be found extremely useful. $S^0$ will also be found useful in the proofs of some later theorems.

$$a_1{}^0 = 0$$

$$a^0_{2^N+i} = a^0_i + \frac{1}{2^{N+1}},$$

$$i = 1, 2, \cdots, 2^N, \; N = 0, 1, 2, \cdots$$

The first few terms of $S^0$ are:

$$0, \tfrac{1}{2}, \tfrac{1}{4}, \tfrac{3}{4}, \tfrac{1}{8}, \tfrac{5}{8}, \tfrac{3}{8}, \tfrac{7}{8}, \tfrac{1}{16}, \tfrac{9}{16}, \tfrac{5}{16}, \tfrac{13}{16}, \tfrac{3}{16}, \tfrac{11}{16}, \tfrac{7}{16}, \tfrac{15}{16}.$$

The proof that $S^0$ is uniformly dense in the unit interval is accomplished by showing first, that (3) holds for all diadic intervals, that is intervals whose end points are of the form $k/2^N$, where $k$ and $N$ are positive integers $0 \leqslant k \leqslant 2^N$. The second part of the proof uses the fact that all intervals in the unit interval can be approximated arbitrarily closely from above and below by diadic ones.

If in the proof of Theorem 5 we take as the computable sequence $S^0$ instead of $S \equiv (\pi (\text{mod } 1), 2\pi (\text{mod } 1), \ldots)$, the sample values of the statistically independent random variables defined there can immediately be displayed for any $N$:

*Table of ten sample values, $X_i(a^0{}_k)$, of five statistically independent and computable indicator random variables*

| $i$ \ $k$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |
| 2 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 |
| 3 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 |
| 4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| 5 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

The use of such sample values in monte Carlo calculations will be discussed later.

## 3. Probability functional spaces

It was chiefly to obtain theorems of the type 3 and 6 that computable probability spaces were introduced. Yet the random variables appearing in these theorems must be elementary as we learned from Theorem 2. Now because of a desire to reformulate Theorems 3 and 6 in a more general context, we define what is meant by "probability functional spaces". The specific use of probability functional spaces in extending the computability notions will be left for a later section.

Before closing this section we will show that the present ideas are completely consistent with measure-theoretic probability and, in fact, are more general by lacking only one postulate. The missing postulate is the familiar complete additivity, or continuity, postulate for an additive set function. One might look upon the definitions to follow as a system for defining a class of random variables without calling upon the complete additivity postulate, that is, without calling upon the whole mechanism of Lebesque theory. It is interesting to recall that the continuity axiom is the only axiom introduced by Kolmogorov in his basic paper of 1933 [4] of which he said, "It is almost impossible to elucidate its empirical meaning". It is because of its extreme importance in facilitating the mathematics that the axiom is kept throughout most of probability theory.

Let $S \equiv \{a\}$ be an abstract collection of elements $a$, $[0, 1]^N$ the closed $N$-dimensional unit cube, $\bar{R}^N$ the extended $N$-dimensional real space and $H^N$ a subset of $\bar{R}^N$. We write $\mathfrak{C}(H^N)$ for the class of all functions defined on $H^N$ to $[0, 1]$ and continuous on $H^N$. We will also write $\bar{R}$ for $\bar{R}^1$.

Let $f_1, f_2, \ldots, f_N$ be $N$ functions, each defined on $S$ to $\bar{R}$. If $H^N$ is the range of the vector function $(f_1, f_2, \ldots, f_N)$, a finite composition $c(f_1, f_2, \ldots, f_N)$ of the vector function $(f_1, f_2, \ldots, f_N)$ with the continuous function $c$ is defined for each $c \in \mathfrak{C}(G^N)$, $H^N \subset G^N$, by $c(f_1, f_2, \ldots, f_N)(a) = c(f_1(a), f_2(a), \ldots, f_N(a))$.

A class $\mathfrak{F}$ of functions, each defined on $S$ to $[0, 1]$ is said to *be closed under all finite compositions with continuous functions*, if $c(f_1, f_2, \ldots, f_N) \in \mathfrak{F}$ for all finite collections $f_1, f_2, \ldots, f_N$ chosen from $\mathfrak{F}$ and for all $c \in \mathfrak{C}([0, 1]^N)$.

Let us denote by "$\mathfrak{L}(\mathfrak{F})$" the smallest linear space containing a class $\mathfrak{F}$, of functions defined on $S$ to $[0, 1]$, that is, the space of complex valued functions with domain $S$ such that:

i) $f \in \mathfrak{L}(\mathfrak{F})$ if $f \in \mathfrak{F}$

ii) $k_1 g_1 + k_2 g_2 \in \mathfrak{L}(\mathfrak{F})$ if $g_1, g_2 \in \mathfrak{L}(\mathfrak{F})$ and $k_1, k_2$ are finite complex numbers

iii) $\mathfrak{L}(\mathfrak{F}) \subset \mathfrak{L}'(\mathfrak{F})$ if $\mathfrak{L}'(\mathfrak{F})$ is any other space satisfying i), ii).

DEFINITION 6. *Let $\mathfrak{F}$ be a class of functions defined on $S$ to $[0, 1]$ that contains $I_S$. A probability functional associated with the class $\mathfrak{F}$ is a real-valued functional $E$ defined on $\mathfrak{L}(\mathfrak{F})$ and satisfying the following three properties for all $f_1, f_2 \in \mathfrak{F}$ and all finite complex $k_1, k_2$:*

i) $0 \leqslant E f_1 \leqslant 1$

ii) $E I_S = 1$

iii) $E(k_1 f_1 + k_2 f_2) = k_1 E f_1 + k_2 E f_2$.

The following postulate will be essential for defining a probability functional space.

POSTULATE 1. *$\mathfrak{F}$ is a non-empty class of functions defined on $S$ to $[0, 1]$ and closed under all finite compositions with continuous functions.*

DEFINITION 7. *A probability functional space is the triplet $(S, \mathfrak{F}, E)$ in which $S$ is an abstract space, $\mathfrak{F}$ is a class of functions defined on $S$ to $[0, 1]$ that satisfies Postulate 1 and $E$ is a probability functional associated with $\mathfrak{F}$.*

DEFINITION 8. *A random variable defined on the probability functional space* $(S, \mathfrak{F}, E)$ *is a real, finite function* $X$ *whose domain is* $S$ *and for which*

$$\mathfrak{F}_X = \{c(X) : c \in \mathfrak{C}(\bar{R})\} \subset \mathfrak{F}.$$

Henceforth, we will always use the symbol "$\mathfrak{F}_{X_1, X_2, \ldots, X_N}$" for the class of compositions $\{c(X_1, X_2, \ldots, X_N) : c \in \mathfrak{C}(\bar{R}^N)\}$.

EXAMPLE 1. As an example of a probability functional space let $S$ be the unit interval $[0, 1]$, $\mathfrak{F}$ the class of continuous functions defined on and to $[0, 1]$ and $Ef$ the Riemann integral of $f \in \mathfrak{F}$. In this case the class of random variables $X$ defined on $(S, \mathfrak{F}, E)$ coincides with the class of real continuous functions defined on $[0, 1]$.

EXAMPLE 2. As another example let $S$ be the unit interval $[0, 1]$, $\mathfrak{F}$ the class of Riemann integrable functions defined on and to $[0, 1]$ and $Ef$ the Riemann integral of $f \in \mathfrak{F}$. In this case the class of random variables $X$ defined on $(S, \mathfrak{F}, E)$ coincides with the class of real finite functions defined on $S$ which are continuous almost everywhere.

We know from Lusin's Theorem that the class of real finite functions defined on $[0, 1]$ which are almost continuous coincide with a class of random variables (measurable functions) defined on a real probability space $([0, 1], \mathfrak{B}, P)$. From this fact it is clear how the random variables $X$ of this example compare with random variables defined on $([0, 1]$ $\mathfrak{B}, P)$.

EXAMPLE 3. As another example let $(S, \mathfrak{B}, P)$ be an abstract probability space, $\mathfrak{F}$ the class of measurable functions defined on $S$ to $[0, 1]$ and $Ef$ the integral or mathematical expectation of $f$ on the abstract space $(S, \mathfrak{B}, P)$. In this case the class of random variables defined on $(S, \mathfrak{F}, E)$ coincides with the class of random variables defined on $(S, \mathfrak{B}, P)$ and if $A \in \mathfrak{B}$, then $I_A \in \mathfrak{F}$ and $EI_A = PA$.

EXAMPLE 4. As in the proof of Theorem 1, let $S = (a_1, a_2, a_3, \ldots)$ be a uniformly dense sequence in the closed unit interval $[0, 1]$. Let $\mathfrak{F}$ be the class of Riemann integrable functions defined on and to $[0, 1]$ and restricted to $S$. For $f \in \mathfrak{F}$, let

$$Ef = \lim \frac{1}{N} \sum_1^N f(a_k).$$

A slight extension of the proof of Theorem 1 will show that $Ef$, as defined here, equals the Riemann integral of any Riemann integrable function on $[0, 1]$ whose restriction to $S$ is $f$. $E$ can immediately be extended to $\mathfrak{C}(\mathfrak{F})$ and $(S, \mathfrak{F}, E)$ is a probability functional space. The class of random variables defined on $(S, \mathfrak{F}, E)$ can be obtained by restricting to $S$ the class of real finite functions on $[0, 1]$ which are continuous almost everywhere.

An important thing displayed in the first two examples is that the class of random variables defined on a probability functional space may coincide with a subset of the class of random variables defined on a probability space. The third example shows that the concept of random variables on a probability functional space, though a generalization, is not inconsistent with the measure-theoretic concept. The fourth example suggests in what sense probability functional spaces permit a generalization of sequence and computable probability spaces. More will be said about this later.

The following theorem shows in what sense there is closure of the class of random variables defined on a probability functional space.

THEOREM 7. *Let $X_1, X_2, \ldots, X_N$ be random variables defined on the probability functional space $(S, \mathfrak{J}, E)$. The composition*

$$Y = c(X_1, X_2, \ldots, X_N)$$

*is a random variable if $c$ is a function defined on $\bar{R}^N$ to $\bar{R}^1$ for which*

i) *$c$ is finite on the finite part of $\bar{R}^N$*

ii) $\lim_{\bar{x} \to \bar{x}_0} c(\bar{x}) = c(\bar{x}_0)$ *for $\bar{x}, \bar{x}_0 \in \bar{R}^N$.*

*Proof.* $Y$ can be written $Y = \bar{c}(U_1, U_2, \ldots, U_N)$ where $U_k = c_k(X_k)$, $c_k$ is a continuous one-to-one function mapping $\bar{R}$ onto $[0, 1]$ and $\bar{c}$ is the function defined on $[0, 1]^N$ to $\bar{R}^N$ that is given by

$$\bar{c}(U_1, U_2, \cdots, U_N) = c[c_1^{-1}(U_1), c_2^{-1}(U_2), \cdots, c_N^{-1}(U_N)]$$

where $c_k^{-1}$ is the inverse of $c_k$. Thus $\bar{\bar{c}}(Y)$, $\bar{\bar{c}} \in C(\bar{R})$, is a finite composition of the vector function $(U_1, U_2, \ldots, U_N)$ with the continuous function $\bar{\bar{c}}(\bar{c})$ and belongs to $\mathfrak{J}$ because each $U_k$ belongs to $\mathfrak{J}$ and Postulate 1 is satisfied. Q.E.D.

Theorem 7 can be compared to the more general statement in measure-theoretic probability theory that Baire functions of random variables are random variables. The fact that limits of random variables as defined here may not be random variables prevents the more general statement. In fact, in the setting of Examples 1, 2 or 4 random variables can be displayed such that a composition with some function possessing a single discontinuity fails to produce a random variable.

THEOREM 8. *Let $X_1, X_2, \ldots, X_N$ and $Y = c(X_1, X_2, \ldots, X_N)$ be as in Theorem 7 with $c$ satisfying conditions* i) *and* ii).

i) $\mathfrak{J}_{X_1} \subset \mathfrak{J}_{X_1, X_2, \ldots, X_N}$

ii) $\mathfrak{J}_Y \subset \mathfrak{J}_{X_1, X_2, \ldots, X_N}$.

The Theorem is an immediate consequence of the definitions. It displays how the classes of functions $\mathfrak{J}_X$ behave relative to each other like the algebras of sets that are generated by random variables defined on a probability space.

We now wish to extend the functional $E$ to random variables $X$ defined on a probability functional space $(S, \mathfrak{J}, E)$ so that a mathematical expectation of random variables will be defined. To do this we consider two cases.

CASE 1: $X$ is non-negative. Let $X^{(n)} = \begin{cases} X & \text{for } X \leqslant n \\ n & \text{otherwise.} \end{cases}$

Each composition in the sequence $X^{(1)}, X^{(2)}, \ldots$ belongs to $\mathfrak{L}(\mathfrak{J}_X)$ and, thus, $EX^{(n)}$ exists and is finite for each integer $n$. $EX^{(1)}, EX^{(2)}, \ldots$ is a non-decreasing sequence of finite real numbers. We define the *mathematical expectation* of $X$ by

$$EX = \lim_n EX^{(n)}$$

CASE 2: $X$ is an arbitrary random variable defined on $(S, \mathfrak{J}, E)$.

In this case, $X$ can be written $X = X^+ - X^-$ where

$$X^+ = \begin{cases} X & \text{for } X \geqslant 0 \\ 0 & \text{otherwise} \end{cases}, \qquad X^- = \begin{cases} -X & \text{for } X < 0 \\ 0 & \text{otherwise} \end{cases}.$$

By Theorem 7, $X^+$ and $X^-$ are random variables, and they are non-negative. The mathematical expectations $EX^+$, $EX^-$ are defined by *Case 1*. If at least one of the numbers $EX^+$, $EX^-$ is finite, we define *the mathematical expectation* of $X$ by

$$EX = EX^+ - EX^-.$$

If $EX$ exists and is finite, $X$ is said to be *integrable*. Because of Theorem 7 and the above definitions all of the following expectations are defined in which $X$ and $Y$ are random variables, defined and integrable on a probability functional space:

$$E(X+Y),\ E|X|,\ E|X|^r,\ EX^r,\ r>0,\ E(XY).$$

Because $e^{i\theta X - \left|\frac{X}{r}\right|}$, $-\infty < \theta < \infty$, belongs to $\mathfrak{L}(\mathfrak{J}_X)$ for real $r$ and $e^{-\left|\frac{X}{r}\right|}$, $r > 0$, is monotonic in $r$, $\Phi_X(\theta) = \lim_{r\to\infty} Ee^{i\theta X - \left|\frac{X}{r}\right|}$, $-\infty < \theta < +\infty$, exists and is defined for all random variables $X$ defined on a probability functional space. The formal similarity between $\Phi_X(\theta)$ and characteristic functions of random variables on a probability space leads us to call this the *characteristic function* of $X$. The question of whether this function has the analytic properties of a characteristic function required in the measure theoretic case, that is, continuity and non-negative definiteness, remains to be answered by proof, however. We will answer this question later in special cases.

The similarity between the way the above extensions were made and the way the Lebesque integral or the Daniell integral is defined is obvious. Though $EX$ can be interpreted as an integral over a measure space for the proper choice of $(S, \mathfrak{J}, E)$, some choices of $(S, \mathfrak{J}, E)$ prevent this. Some exceptional cases in which $EX$ is not an integral over a measure space and cannot be extended to be an integral arise because no postulate corresponding to complete additivity of a set function (such as continuity of the linear functional $E$) has been assumed. It is precisely the exceptional case that is of greatest interest to this paper. It arises when, in generalizing computable probability spaces, $S$ is chosen to be a sequence.

Example 1. (continuation) In this case every random variable, $X$, is bounded and $EX$ is the Riemann integral of $X$. $\Phi_X$ is a characteristic function.

Example 2. (continuation) As stated earlier the class of random variables, $X$, in this example do not comprise the class of finite measurable functions on the real probability space $([0, 1], \mathfrak{B}, P)$ but are a subclass of them. On the other hand, $E$ extended to the random variables of this example is such that $EX$ is the Lebesque integral of $X$ over $([0, 1], \mathfrak{B}, P)$ and is thus a mathematical expectation in the measure-theoretic sense. $\Phi_X$ is a characteristic function.

Example 3. (continuation) In this case, where the class of random variables defined on $(S, \mathfrak{J}, E)$ coincides with the class of random variables defined on $(S, \mathfrak{B}, P)$, $EX$ coincides with the mathematical expectation or integral of $X$ over $(S, \mathfrak{B}, P)$, and $\Phi_X$ with the characteristic function.

Example 4. (continuation) Here $S$ is a sequence. Unlike Examples 1 and 2 the probability functional extended to $X$, $EX$, is not a Lebesque integral of a measurable function on a probability space. Take for example the sequence $X_1, X_2, \ldots$ where

$$X_n(a_k) = \begin{cases} 1 & k \leqslant n \\ 0 & \text{otherwise} \end{cases}$$

Each $X_n$ is a random variable on $(S, \mathfrak{J}, E)$ since it coincides with the restriction to $S$ of a Riemann integrable function whose domain is $[0, 1]$. $EX_n = 0$ for all finite $n$. Thus

$$\lim EX_n = 0.$$

However, at the same time, the sequence is monotone non-decreasing with $\lim X_n = I_S$. Thus

$$E \lim X_n = 1.$$

Monotone convergence does not hold and $E$ is not a Lebesque integral.

Many of the properties of expectation as defined in measure-theoretic probability theory are true of expectatation as defined in the present context. Some typical properties are stated in the next theorem.

THEOREM 9. *Let $X$ and $Y$ be random variables defined on the probability functional space $(S, \mathfrak{F}, E)$ and let $r$, $s$ be real finite numbers.*

i) *$X$ is integrable, if and only if $|X|$ is integrable*

ii) *If $X$ and $Y$ are integrable, $rX + sY$ is integrable and $E(rX + sY) = rEX + sEY$*

iii) *If $X \leqslant Y$, then $EX \leqslant EY$*

iv) *If $|X| \leqslant Y$ and $Y$ is integrable, then $X$ is integrable*

v) $|EX| \leqslant E|X|$

vi) *If $|X|^r$ is integrable, $|X|^s$ is integrable for $0 < s \leqslant r$*

vii) (*Hölder Inequality*) *If $1/r + 1/s = 1$, $r > 1$,*

$$E|XY| \leqslant E^{1/r}|X|^r E^{1/s}|Y|^s$$

vii) (*Minkowski Inequality*) *If $r \geqslant 1$,*

$$E^{1/r}|X+Y|^r \leqslant E^{1/r}|X|^r + E^{1/r}|Y|^r.$$

The proofs are similar to those of the corresponding theorems found in Loève's book [5] or Kolmogorov's monograph [4].

Because in the present context only the compositions $c(X)$ of random variables defined in Theorem 7 can be guaranteed to yield random variables, operations like $EI_{[X \leqslant t]}$ for real $t$ are undefined unless, as in the discrete case, $I_{[X \leqslant t]}$ happens to belong to $\mathfrak{F}$. It is also the case that the weak closure of the class of random variables on a probability functional space, as set forth in Theorem 7, restricts the type of limits that yield random variables.

A few of the most important ideas and theorems for probability functional spaces will now be developed.

DEFINITION 9. *Let $X$, $X_1$, $X_2$, ... be random variables defined on the probability functional space $(S, \mathfrak{F}, E)$. The sequence $X_1, X_2, \ldots$ converges in probability to $X$ and we write*

$$X_n \xrightarrow{P} X$$

*if and only if for arbitrary $a > 0$, and for any function $h \in C(\bar{R})$ such that $h(x) = 0$* for $|x| < a$

$$\lim_{n \to \infty} Eh(X_n - X) = 0.$$

It is easy to show that if "probability space" is substituted for "probability functional space" in this definition that the criterion is equivalent to convergence in probability of random variables defined on a probability space.

*Uniform convergence* and *convergence in the r-th mean* have the customary definitions and for these we write

$$X_n \xrightarrow{u} X$$

$$X_n \xrightarrow{r} X,$$

the latter being defined for random variables with finite $r$th absolute moments.

For sequences which are mutually convergent in any of the above senses we write

$$X_n - X_m \xrightarrow{P} 0$$

$$X_n - X_m \xrightarrow{u} 0$$

$$X_n - X_m \xrightarrow{r} 0$$

The basic inequalities of probability theory can be restated in the following form. (Compare Kolmogorov [4] or Loève [5].)

THEOREM 10. *Let $X$ be a random variable defined on the probability functional space $(S, \mathfrak{J}, E)$ and let $g$ be an even, non-decreasing, non-negative function satisfying* i) *and* ii) *of Theorem* 7, $N=1$. *Let $h \in \mathfrak{C}(\bar{R})$ be such that $h(X)=0$ if $|X|<a$, $a>0$.*

i) $$\frac{Eg(X)}{g(a)} \geqslant Eh(X)$$

ii) *If $g$ is bounded by $K$ and $1-g(a)/g(X) \leqslant h(X)$, then*

$$Eh(X) \geqslant \frac{Eg(X)-g(a)}{K}$$

iii) If $|X|$ is bounded by $L$ and $1-g(a)/g(X) \leqslant h(X)$, then

$$Eh(X) \geqslant \frac{Eg(X)-g(a)}{g(L)}.$$

*Proof*. Under the respective assumptions of i), ii) and iii) of the Theorem:

i) $Eg(X) \geqslant Eg(X)h(X) \geqslant g(a)Eh(X)$.

ii) $Eg(X) = Eg(X)h(X) + Eg(X)[1-h(X)] < KEh(X) + g(a)$.

iii) $Eg(X) = Eg(X)h(X) + Eg(X)[1-h(X)] \leqslant g(L)Eh(X) + g(a)$. Q.E.D.

THEOREM 11. *Let $X, X_1, X_2, \ldots$ be random variables defined on the probability functional space $(S, \mathfrak{J}, E)$.*

i) *If* $X_n \xrightarrow{r} X$ *then* $X_n \xrightarrow{P} X$

ii) *If the $X_n$ are uniformly bounded and if* $X_n \xrightarrow{P} X$ *then* $X_n \xrightarrow{r} X$

iii) *Let $g$ be an even bounded, non-negative function, monotonic increasing on $[0, \infty]$, satisfying* i), ii) *of Theorem* 7, $N=1$, *such that* $g(0)=0$.

$$X_n \xrightarrow{P} X \text{ if and only if } E\,g(X_N - X) \longrightarrow 0.$$

The proof follows directly from the previous theorem.

COROLLARY. *Let $X, X_1, X_2, \ldots$ be random variables on the probability functional space $(S, \mathfrak{J}, E)$.*

i) $$X_n \xrightarrow{P} X \text{ if and only if } E\frac{|X_n - X|}{1+|X - X_n|} \to 0$$

ii) $$X_m - X_n \xrightarrow{P} 0 \text{ if and only if } E\frac{|X_n - X_m|}{1+|X_m - X_n|} \to 0$$

iii) *If* $X_n \xrightarrow{1} X$ *or if* $X_1, X_2, \cdots$ *are uniformly bounded and* $X_n \xrightarrow{P} X$, *then*

$$E X_n \to E X$$

*and*

$$E|X_n| \to E|X|.$$

It is seen that if a sequence of $X_n$ converges in probability or in the $r$th mean to $X$ and also to $Y$ then

$$E\frac{|X - Y|}{1+|X - Y|} = 0.$$

Thus, as in measure-theoretic probability theory (see Loève [5]) where convergence is convergence of equivalence classes to equivalence classes of random variables, we defined equivalent random variables on a probability functional space in terms of the above metric.

DEFINITION 10. *Let $X$ and $Y$ be random variables defined on the probability functional space $(S, \mathfrak{J}, E)$. $X$ and $Y$ are called equivalent ($X \equiv Y$) if:*

$$E\frac{|X - Y|}{1+|X - Y|} = 0$$

Of course, this criterion implies that $X$ equals $Y$ almost everywhere if they happen also to be random variables defined on a probability space as in Example 3. The definition permits the equivalence classes of random variables on a probability functional space to be viewed as elements of a metric space in which distance is defined by

$$d(X, Y) = E\frac{|X - Y|}{1 + |X - Y|}.$$

The question of completeness of these metric spaces arises. Though in some important special cases the space is complete, one can discover counterexamples to completeness in the general case. Any sequence of random variables in Example 1 that converges everywhere to a discontinuous function converges mutually in the sense of the above metric but does not converge to a random variable defined on the space of Example 1.

THEOREM 12. *Let $X$ and $Y$ be random variables defined on the probability functional space $(S, \mathfrak{F}, E)$.*

i) $X \equiv Y$, *if and only if* $E|X - Y| = 0$

ii) $X \equiv Y$ *implies* $EX = EY$ *and* $E|X| = E|Y|$.

*Proof* of (i). Using the notation following Theorem 8,

$$\frac{E|X - Y|^{(n)}}{1+n} \leqslant E\frac{|X - Y|^{(n)}}{1 + |X - Y|^{(n)}} \leqslant E\frac{|X - Y|}{1 + |X - Y|} = 0$$

$$E|X - Y| = \lim E|X - Y|^{(n)} = 0$$

*Proof* of (ii).

$$|E|X| - E|Y|| \leqslant E||X| - |Y|| \leqslant E|X - Y|, \ |EX - EY| \leqslant E|X - Y|.$$ Q.E.D.

As will be seen in the following discussion, the concept of sets of probability measure zero, though definable in probability functional spaces, does not in general lead to a stronger form of convergence than convergence in probability or to a useful type of equivalence of random variables as happens in the special case of probability spaces.

DEFINITION 11. *Let $(S, \mathfrak{F}, E)$ be a probability functional space. A set $A \subset S$ is said to have probability functional measure zero, if*

i) $I_A \in \mathfrak{F}$

ii) $EI_A = 0$

*Convergence almost surely* for a sequence of random variables on a probability functional space can now be defined in the customary way, the only difference being that the exceptional set where the sequence may not converge must have the above definition. The notation for convergence to a random variable and mutual convergence in this sense is:

$$X_n \xrightarrow{a.s.} X$$

$$X_n - X_m \xrightarrow{a.s.} 0.$$

We will also say that a *relation* between random variables defined on a probability functional space *holds almost surely* (a.s.) if it holds except on a set of probability functional measure zero. These definitions coincide with the measure-theoretic definitions in the special case displayed in Example 3.

THEOREM 13. *Let $(S, \mathfrak{J}, E)$ be a probability functional space and let $A \subset S$ be a set of probability functional measure zero. For all random variables $X$, $Y$ defined on $(S, \mathfrak{J}, E)$:*

i) $E|X|I_A = 0$

ii) *If $X = Y$ a.s. then $X \equiv Y$.*

*Proof* of (i). For an arbitrary random variable $X$ on $(S, \mathfrak{J}, E)$, and $n > 1$,

$$E[|X|I_A]^{(n)} = nE\frac{[|X|I_A]^{(n)}}{n} = nE\frac{[|X|^{(n)}I_A]}{n} \leqslant nEI_A = 0$$

Thus,

$$E|X|I_A = \lim \mathrm{E}\,[|X|I_A]^{(n)} = 0.$$

*Proof* of (ii). If $X = Y$ a.s. then by (i):

$$E|X - Y| = E|X - Y|I_A = 0$$

and by Theorem 12, $X \equiv Y$. Q.E.D.

It is seen how equivalent random variables are obtained from $X$ by modifying $X$ on a set of measure zero. In measure-theoretic probability this method yields the entire class of variables equivalent to $X$. In the present context it is easy to show, however, that two random variables $X$ and $Y$ defined on a probability functional space may differ everywhere on $S$ and still be equivalent. Let $(S, \mathfrak{J}, E)$ be the space defined in Example 4, where for the sake of the present argument, $S$ is the special sequence defined in Section 2 and labeled "$S^0$". It is important that every point of $S^0$ is rational. Now take for $X$ the function identically zero on $S^0$ and for $X'$ the following:

$$X'(0) = 1$$
$$X'(a^0) = 1/n,\ a^0 \in S^0,\ a^0 \neq 0,$$

where $a^0 = m/n$, $n > 0$, and $m$ and $n$ are integers without any common divisor. $X'$ is the restriction to $S^0$ of a well known positive function on $[0, 1]$ which is continuous on all irrationals, discontinuous on all rationals, and whose Riemann integral over $[0, 1]$ vanishes. It is, thus, a random variable with $E|X'| = 0$ and is therefore equivalent to $X$.

In a similar example that will come later it will be seen that a sequence $X_1, X_2, \ldots$ of random variables can be defined on a probability functional space for which

$$X_n \xrightarrow{P} a > 0$$

and

$$X_n \xrightarrow{a.s.} 0$$

Therefore, convergence almost surely on a probability functional space does not imply convergence in probability.

EXAMPLE 1. (continued) In this example there are no sets of probability functional measure zero.

EXAMPLE 2. (continued) Here a set $A \subset S$ has probability functional measure zero if and only if it has Riemann content zero.

EXAMPLE 3. (continued) Here the concept of probability functional measure zero coincides with that of probability measure zero.

EXAMPLE 4. (continued) In this case a set $A$ chosen from the sequence $S \equiv (a_1\, a_2, \ldots)$ has probability functional measure zero if and only if it has density zero.

THEOREM 14. (Dominated Convergence) *If $Y, X, X_1, X_2, \ldots$ are random variables defined on the probability functional space $(S, \mathfrak{F}, E)$ and if*

$$|X_n| \leqslant Y \text{ a.s.}, \; |X| \leqslant Y \text{ a.s.}, \; E\,Y^r < \infty \text{ for some } r > 1,$$

*then $X_n \xrightarrow{P} X$ implies $E\,X_n \longrightarrow E\,X$ and $E|X_n| \longrightarrow E|X|$.*

*Proof.* We shall prove $X_n \xrightarrow{1} X$ and use the Corollary to Theorem 11. For arbitrary $a > 0$, let $h \in C(\bar{R})$ be such that

i) $h(X) = 0$ for $|X| < a$

ii) $1 - a/|X| \leqslant h(X)$.

Then,

$$E|X_n - X| = E|X_n - X|h(X_n - X) + E|X_n - X|[1 - h(X_n - X)]$$
$$\leqslant E^{1/r}|X_n - X|^r \cdot E^{1/s} h^s(X_n - X) + a$$

where $1/r + 1/s = 1$ and $r > 1$. We have used the Hölder inequality of Theorem 9. Because $|X_n - X|^r \leqslant (2\,Y)^r$ a.s., $E^{1/r}|X_n - X|^r$ is a finite number, say $K$. Consequently,

$$E|X_n - X| \leqslant K\,E^{1/s} h^s(X_n - X) + a.$$

Now $X_n \xrightarrow{P} X$ and $h^s \in C(\bar{R})$ with $h^s(X) = 0$ for $|X| < a$, and we have

$$E^{1/s} h^s(X_n - X) \to 0$$

a is arbitrary and the theorem is proved. Q.E.D.

DEFINITION 12. *Let $(S, \mathfrak{F}, E)$ be a probability functional space and $X_1, X_2, \ldots$ a sequence of random variables defined on it.*

i) *$X_1$ and $X_2$ are called statistically independent if $E f_1 f_2 = E f_1 E f_2$ for all $f_1 \in \mathfrak{F}_{X_1}$, $f_2 \in \mathfrak{F}_{X_2}$.*

ii) $X_1, X_2, \ldots$ *are called statistically independent if for any finite subcollection* $X_1', X_2', \ldots, X_N'$, $Ef_1 f_2 \ldots f_N = Ef_1 Ef_2 \ldots Ef_N$ *for all* $f_k \in \mathfrak{F}_{X_k'}$, $k = 1, 2, \ldots, N$.

LEMMA. *Let* $X$ *be a non-negative random variable defined on the probability functional space* $(S, \mathfrak{F}, E)$.

$$\lim_{\varepsilon \to 0} E\, X_{(\varepsilon)} = E\, X \text{ where } X_{(\varepsilon)} = \begin{cases} X & \textit{if } X \geqslant \varepsilon \\ \varepsilon & \textit{otherwise.} \end{cases}$$

*Proof.* $X_{(\varepsilon)} = X - X^{(\varepsilon)} + \varepsilon$
If $X$ is integrable, $0 \leqslant E\, X_{(\varepsilon)} - E\, X = \varepsilon - E\, X^{(\varepsilon)} \leqslant \varepsilon$.
If $X$ is not integrable, the Lemma is true by Theorem 9 (iii).

THEOREM 15. *Let* $X$ *and* $Y$ *be statistically independent random variables defined on the probability functional space* $(S, \mathfrak{F}, E)$,

i) $Eg_1 g_2 = Eg_1 Eg_2$ *for all* $g_1 \in \mathfrak{L}(\mathfrak{F}_X)$, $g_2 \in \mathfrak{L}(\mathfrak{F}_Y)$
ii) $c_1(X)$ *and* $c_2(Y)$ *are statistically independent for any functions* $c_1$ *and* $c_2$ *satisfying* i) *and* ii) *of Theorem* 7, *with* $N = 1$
iii) $\Phi_{X+Y}(\theta) = \Phi_X(\theta)\Phi_Y(\theta), -\infty < \theta < \infty$, *if* $XY \geqslant 0$
iv) *If* $X$ *and* $Y$ *are integrable,* $E\, X\, Y = E\, X\, E\, Y$
v) *If* $X$ *and* $Y$ *are integrable,* $\sigma^2_{X+Y} = \sigma_X{}^2 + \sigma_Y{}^2$ *where* $\sigma_X{}^2 = E[X - E\, X]^2$.

*Proof* of (iv).

*Case I.* $X \geqslant 0$, $Y \geqslant 0$. First,

$$X^{(\sqrt{n})}\, Y^{(\sqrt{n})} \leqslant [X\, Y]^{(n)}, n > 1.$$

Because $X^{(\sqrt{n})}$ and $Y^{(\sqrt{n})}$ belong to $\mathfrak{L}(\mathfrak{F}_X)$ and $\mathfrak{L}(\mathfrak{F}_Y)$, respectively, and because of (i):

$$E\, X^{(\sqrt{n})} E\, Y^{(\sqrt{n})} \leqslant E\,[X\, Y]^{(n)}$$

and

$$E\, X\, E\, Y \leqslant E\, X\, Y.$$

Next,

$$[X\, Y]^{(n\varepsilon)} \leqslant X^{(n)}_{(\varepsilon)}\, Y^{(n)}_{(\varepsilon)},\ n > \varepsilon > 0.$$

Again

$$X^{(n)}_{(\varepsilon)} \in \mathfrak{L}(\mathfrak{F}_X) \text{ and } Y^{(n)}_{(\varepsilon)} \in \mathfrak{L}(\mathfrak{F}_Y),$$

Therefore,

$$E\,[X\, Y]^{(n\varepsilon)} \leqslant E\, X^{(n)}_{(\varepsilon)}\, E\, Y^{(n)}_{(\varepsilon)}$$

and

$$E\, X\, Y \leqslant E\, X_{(\varepsilon)}\, E\, Y_{(\varepsilon)}.$$

From the Lemma, one obtains after letting $\varepsilon$ approach zero,

$$E\, X\, Y \leqslant E\, X\, E\, Y$$

and with the above reverse inequality the theorem follows.

*Case II.* $X$ and $Y$ are integrable.
From (ii) it is clear that the pairs $(X^+, Y^+)$, $(X^-, Y^-)$, $(X^+, Y^-)$, $(X^-, Y^+)$ $(|X|, |Y|)$ are statistically independent. It follows that $XY$ is integrable and we have:

$$\begin{aligned} EXY &= E[XY]^+ - E[XY]^- \\ &= E[X^+Y^+ + X^-Y^-] - E[X^-Y^+ + X^+Y^-] \\ &= (EX^+ - EX^-)(EY^+ - EY^-) \\ &= EXEY. \end{aligned}$$

Q.E.D.

In this paper the special case introduced in Example 4 is of particular interest and will provide the framework for the discussion remaining after the next section. There are a large number of definitions and theorems with which one could continue the discussion of probability functional spaces. As one might suspect, the majority of these are formulated by recasting the measure-theoretic ideas in terms of the above notions. Proofs must be carried out in the more general domain but often are suggested by the measure-theoretic ones.

## 4. A logical algebra of functions

It is immediately noticed that probability functional spaces, and random variables on them, have been introduced without any reference to logical operations on the class of functions $\mathfrak{F}$ analogous to the logical operations of complementation, union and intersection on a class of sets. This section introduces such operations.

NOTATION. The operations $f_1 + f_2, f_1 - f_2$ on real valued functions are the usual pointwise sum and difference. The relations $f_1 = f_2$, $f_1 \leqslant f_2$ etc. between functions defined on $S$ are understood to hold pointwise for all points in $S$. The functions min $[f_1, f_2]$, max $[f_1, f_2]$, where $f_1$, $f_2$ are functions defined to $[0, 1]$ are understood to be the compositions $c_1[f_1, f_2]$, $c_2$ $[f_1, f_2]$, where $c_1, c_2 \in \mathsf{C}([0, 1]^2)$ and $c_1[x, y] = \min [x, y]$, $c_2[x, y] = \max [x, y]$, $0 \leqslant x, y \leqslant 1$. "$\Leftrightarrow$" is a symbol for "if and only if". "$\Rightarrow$" is a symbol for "implies".

DEFINITION 13. *A logical algebra of functions* $(\mathfrak{F}, U, {}^c)$ *is a non-empty collection* $\mathfrak{F}$ *of functions defined on an abstract space* $S$ *to* $[0, 1]$, *a binary operation* $U$ *defined for each pair of functions in* $\mathfrak{F}$, *and a unary operation* ${}^c$ *defined for each element in* $\mathfrak{F}$, *for which the following eight postulates are satisfied for all* $f_1, f_2, f_3 \in \mathfrak{F}$. *In the postulates the definitions*

$$f_1 \cap f_2 = (f_1{}^c \cup f_2{}^c)^c$$
$$I_\phi = I_S{}^c$$

*are used.*

i) $f_1^c \in \mathfrak{F}$

ii) $f_1 \cup f_2 \in \mathfrak{F}$

iii) $f_1 \cup f_1^c = I_S$

iv) $f_1 \cap f_1^c = I_\phi$

v) $f_1 \cup f_2 = f_2 \cup f_1$

vi) $f_1 \cup (f_2 \cup f_3) = (f_1 \cup f_2) \cup f_3$

vii) $f_1 \cap I_\phi = I_\phi$

viii) $f_1 \cup f_2 = f_1 + f_2 \Leftrightarrow f_1 \cap f_2 = I_\phi$.

The operations $\cup$, $\cap$, $^c$ will be called union, intersection and complement, respectively. If the distributive postulates

ix) $f_1 \cup (f_2 \cap f_3) = (f_1 \cup f_2) \cap (f_1 \cup f_3)$

x) $f_1 \cap (f_2 \cup f_3) = (f_1 \cap f_2) \cup (f_1 \cap f_3)$

were added, $(\mathfrak{F}, \cup, ^c)$ would be Boolean, with or without postulate (viii). Examples of systems satisfying all postulates (i)–(x) can be displayed but do not include some important systems, as the following discussion will show. The formal similarity between postulate (viii) and the additivity property of additive set functions is worth noticing and, in fact, the postulate is brought in to make probability functionals behave formally on a logical algebra of functions like set functions on an algebra of sets.

A justification for the choice of the axioms will be seen in Theorem 19 and Theorem 21. These axioms lead to many of the properties of Boolean algebra, though the following theorems do not emphasis the Boolean characteristics.

DEFINITION 14. *Let $(\mathfrak{F}, \cup, ^c)$ be a logical algebra of functions. For all $f_1, f_2 \in \mathfrak{F}$*

$$f_1 \supset f_2 \Leftrightarrow f_1^c \cap f_2 = I_\phi$$

$$f_1 \subset f_2 \Leftrightarrow f_2 \supset f_1$$

THEOREM 16. *Let $(\mathfrak{F}, \cup, ^c)$ be a logical algebra of functions. For all $f_1, f_2 \in \mathfrak{F}$*

i) $f_1^c = I_S - f_1$

ii) $(f_1 + f_2)^c = f_1^c + f_2^c - I_S$ if $f_1 + f_2 \in \mathfrak{F}$

iii) $(f_1 - f_2)^c = f_1^c - f_2^c + I_S$ if $f_1 - f_2 \in \mathfrak{F}$

iv) $(\min [f_1, f_2])^c = \max [f_1^c, f_2^c]$ if $\min [f_1, f_2] \in \mathfrak{F}$

Theorem 16 permits us to state the following principle for a logical algebra of functions.

PRINCIPLE OF DUALITY. *Let $(\mathfrak{F}, \cup, ^c)$ be a logical algebra of functions. Any $=$, $\subset$ or $\supset$ relation that is universally true between elements of $\mathfrak{F}$ and that is formed with the use of the symbols*

$$\cup, \cap, {}^c, +, -, I_s, I_\phi, =, \supset, \subset, \text{min}, \text{max}$$

*becomes another true relation when these symbols are replaced, respectively, by*

$$\cap, \cup, {}^c, -I_S+, +I_S-, I_\phi, I_S, =, \subset, \supset, \text{max}, \text{min},$$

*complement and equality remaining unchanged.*

THEOREM 17. *Let* $(\mathfrak{F}, \cup, {}^c)$ *be a logical algebra of functions. For all* $f_1, f_2 \in \mathfrak{F}$

i) $f_1 \supset f_2 \Leftrightarrow f_1 \cap f_2^c = f_1 - f_2 \Leftrightarrow f_1 = f_2 \cup (f_2^c \cap f_1)$

ii) $f_1 \supset f_2 \Rightarrow f_1 \geqslant f_2$

iii) $f_1 \cup f_2 = f_1 + (f_1 \cup f_2) \cap f_1^c$

THEOREM 18. *Let* $(\mathfrak{F}, \cup, {}^c)$ *be a logical algebra of functions. For all* $f_1, f_2 \in \mathfrak{F}$

i) $f_1 \cup f_2 \leqslant \text{min}\,[f_1 + f_2, I_S]$

ii)(1) $f_1 \cup f_2 = f_1 + f_2 - f_1 \cap f_2$

The proof uses the statements of Theorem 17.

It should be remarked that Theorem 18 does not require $\text{min}\,[f_1 + f_2, I_S]$ or $f_1 + f_2$ to belong to $\mathfrak{F}$. Theorem 18, ii obviously implies

$$f = \tfrac{1}{2}(f \cup f + f \cap f)$$

for all $f \in \mathfrak{F}$ which is the replacement for idempotency in this algebra.

It will be of interest for us to investigate in the next theorem conditions under which equality holds in Theorem 18, i. Not only will such conditions provide us with an algorithm for computing $f_1 \cup f_2$ but they will also bring into the algebra other useful properties.

At this point it may be helpful to warn the reader of certain desirable properties familiar from Boolean algebra which are *not* in general true in a logical algebra of functions. These include among others the distribution, absorption and idempotency laws involving union and intersection.

THEOREM 19. *There is one and only one function $h$ on* $[0, 1]^2$ *to* $[0, 1]$ *such that for the binary operation* $f_1 \bar{\cup} f_2 = h(f_1, f_2)$ *and every class* $\mathfrak{F}$ *satisfying Postulate* 1 *(of Section* 2*) for some* $S$, $(\mathfrak{F}, \bar{\cup}, {}^c)$ *is a logical algebra of functions.*

*The unique binary operation may be written*

$$f_1 \bar{\cup} f_2 = \text{min}\,[f_1 + f_2, I_S].$$

*Proof.* For any class $\mathfrak{F}$ satisfying Postulate 1, $\text{min}\,[f_1 + f_2, I_S]$ will belong to $\mathfrak{F}$ when $f_1$ and $f_2$ belong, and it can easily be verified by direct computation that $(\mathfrak{F}, \bar{\cup}, {}^c)$ will be

(1) This relation has appeared in a number of different studies generally concerned with multivalued logies and valuated algebras.

a logical algebra of functions. It will suffice to prove uniqueness for any particular choice of the class $\mathfrak{F}$ satisfying Postulate 1.

We choose for the sake of the proof the class $\mathfrak{F}$ of all Riemann integrable functions defined on [0, 1] to [0, 1]. (See Example 2, Section 3.) For this class, consider a logical algebra of functions $(\mathfrak{F}, \cup, ^c)$ that is arbitrary except for the restriction $f_1 \cup f_2 = h(f_1, f_2)$ for all $f_1, f_2 \in \mathfrak{F}$ and some function $h$ on $[0, 1]^2$ to [0, 1]. Because of Postulate (ii) of Definition 13, $h(x, y)$ must be continuous when considered as a function on $[0, 1]^2$. We wish to show that $h(f_1, f_2)$ has the unique form $\min[f_1 + f_2, I_S]$ for all $f_1, f_2 \in \mathfrak{F}$. This will be proved if it can be shown that $h$ has the stated unique form for all constant functions in $\mathfrak{F}$.

Choose now any two constant functions $x, y$ in $\mathfrak{F}$ such that $x + y = I_S$. For those functions Postulate (iii) implies

$$x \cup y = x \cup x^c = I_S.$$

Because $y \cup I_\phi = y$ and $y \cup y^c = I_S$ and because $y \cup v$, as a function of constant arguments $v$, is continuous, there is for each constant function $z \geqslant y$, $z \in \mathfrak{F}$, some constant function $w \in \mathfrak{F}$ such that $z = y \cup w$. Therefore,

$$x \cup z = x \cup (y \cup w) = (x \cup y) \cup w = I_S.$$

We conclude that for the constant functions $x$ and $z$ in $\mathfrak{F}$

$$x \cup z = I_S \text{ provided } x + z \geqslant I_S.$$

From Postulate (viii), Theorem 16 and the definition of intersection we have for $x, z \in \mathfrak{F}$

$$x^c \cup z^c = x^c + z^c \Leftrightarrow x \cup z = I_S.$$

Adding the fact

$$x^c + z^c \leqslant I_S \Leftrightarrow x + z \geqslant I_S$$

we have for constants $x', z' \in \mathfrak{F}$

$$x' \cup z' = x' + z' \text{ provided } x' + z' \leqslant I_S.$$

Thus, if $f_1, f_2$ are constant functions in $\mathfrak{F}$

$$f_1 \cup f_2 = \min[f_1 + f_2, I_S]$$

and this implies the result. Q.E.D.

It can be seen that the inclusion relation for a logical algebra of function implies the $\geqslant$ relation and provides a partial ordering of $\mathfrak{F}$. In the special case where the union operation is chosen to be $\bar{\cup}$, the stronger results of the next theorem can be obtained.

THEOREM 20. *For a logical algebra of functions of the type* $(\mathfrak{F}, \bar{\cup}, ^c)$, *in which* $\mathfrak{F}$ *satisfies Postulate 1, we have for all* $f_1, f_2 \in \mathfrak{F}$

i) $f_1 \subset f_2 \Leftrightarrow f_1 \leqslant f_2$

ii) $(\mathcal{F}, \bar{\cup}, {}^c)$ *is a lattice under the partial ordering relation* $\subset$.

The following theorem displays in what sense a logical algebra of functions is consistent with a Boolean algebra of sets in the space $S$ on which the functions are defined.

THEOREM 21. *Let* $(\mathcal{F}, \cup, {}^c)$ *be a logical algebra of functions in which union has the special form* $\cup = \bar{\cup}$. *If* $\mathcal{D}$ *is the class of all indicator functions in* $\mathcal{F}$

i) $(\mathcal{D}, \cup, {}^c)$ *is a subalgebra*

ii) $(\mathcal{D}, \cup, {}^c)$ *is Boolean*

iii) *All functions in* $\mathcal{D}$ *are measurable with respect to an algebra* $(\mathcal{A}, \cup, {}^c)$ *of subsets of*

*$S$ and for all $A, B \in \mathcal{A}$:*

iv) $I_A, I_B \in \mathcal{D}$

v) $I_A \cup I_B = I_{A \cup B}$

vi) $I_A{}^c = I_{A^c}$

vii) $I_A \supset I_B \Leftrightarrow A \supset B$.

The proof of the theorem follows immediately from the postulates and the previous results. The same will be true of Theorem 22.

The last theorem of this section unites the properties of a logical algebra of functions with those of a probability functional. As mentioned earlier, the definition of probability functional spaces required no logical structure within the collection of functions $\mathcal{F}$. It is now seen that such a logical structure can always be assumed, however, and used if desired.

THEOREM 22. *Let* $(S, \mathcal{F}, E)$ *be a probability functional space and* $(F, \cup, {}^c)$ *a logical algebra of functions. For all* $f_1, f_2, \dots f_N \in \mathcal{F}$

i) $Ef^c = 1 - Ef$

ii) $f_1 \subset f_2 \Rightarrow Ef_1 \leqslant Ef_2$

iii) $E(f_1 \cup f_2) = Ef_1 + Ef_2 - E(f_1 \cap f_2)$

iv) $E(f_1 \cup f_2 \cup \dots \cup f_N) \leqslant Ef_1 + Ef_2 + \cdots + Ef_N$

v) $E(f_1 \cup f_2 \cup \dots \cup f_N) = Ef_1 + Ef_2 + \cdots + Ef_N$ *if all of the following are satisfied:*

$$f_1 \cap f_2 = I_\phi$$
$$(f_1 \cup f_2) \cap f_3 = I_\phi$$
$$\vdots$$
$$(f_1 \cup f_2 \cup \dots \cup f_{N-1}) \cap f_N = I_\phi$$

## 5. Computable probability functional spaces

It is of interest to investigate the generalized sequence probability spaces that result by letting $S$ and $E$, in a probability functional space $(S, \mathfrak{J}, E)$, be a sequence and asymptotic average, respectively. (Compare Definition 1, Section 2.) Such spaces might be called *sequence probability functional spaces*. Within this class of spaces there is the important case where $S$ is uniformly dense in [0, 1] and $\mathfrak{J}$ is the collection of Riemann integrable functions defined on and to [0, 1] and restricted to $S$. (See Example 4, Section 3.) *It is this case that we will be concerned with throughout the remainder of this paper.* We will designate a sequence probability functional space in this special case as *Riemann*. Furthermore, the sequence $S$ in [0, 1] may be computable and, if this is so, $(S, \mathfrak{J}, E)$ will be called *computable*. The adjective "*computable*" will also be applied to *random variables* and *sequences of random variables* on $(S, \mathfrak{J}, E)$, as in Section 2, with the understanding that their definitions are given by Definitions 4 and 5, modified to read 'computable probability functional space' in place of 'computable probability space'.

The special implications that a Riemann space $(S, \mathfrak{J}, E)$ has for number theory, Monte Carlo methods and statistical mechanics will be left for some papers to follow. We will now consider what further theorems are true in the Riemann case beside what has already been proved true in the general context of Section 3.

In Example 4, Section 3, it was seen that all random variables on a Riemann space are the restriction to $S$ of finite functions that are continuous almost everywhere. Moreover, this being the case, Theorem 7, Section 3, can be replaced by a stronger statement, namely, that any composition $c(X_1, X_2, \dots X_N)$ of such random variables, $X_1, X_2, \dots X_N$ with a function $c$ that is continuous on the finite part of $\bar{R}^N$ is again a random variable. The implications of this statement for characteristic functions of random variables on a Riemann space are given in the next theorem.

In the next theorem it is also seen that though random variables on a Riemann space are limited in the degree to which they can be discontinuous, their characteristic functions form as general a class as do those of random variables on any probability space.

THEOREM 23. i) *For any random variable $X$ defined on a Riemann space $(S, \mathfrak{J}, E)$,*

$$\Phi_X(\theta) = E\, e^{i\theta X}, \quad -\infty < \theta < \infty,$$

*exists and equals the characteristic function of some random variable on a probability space.*

ii) *Conversely, for any characteristic function, $\Phi(\theta)$, of a random variable on a probability space, there is a random variable $X$ defined on a Riemann space $(S, \mathfrak{J}, E)$ such that $\Phi_X(\theta) = \Phi(\theta)$.*

*Proof.* (i) follows from the fact that $X$ is the restriction to $S$ of a finite function that is continuous almost everywhere and thus $e^{i\theta X}$ is the restriction to $S$ of a Riemann integrable function.

To prove (ii), it is sufficient to show that for any characteristic function, $\Phi(\theta)$, of a random variable on a probability space, there is a random variable $X'$ on the probability space $([0, 1], \mathcal{B}, \mu)$ which is continuous almost everywhere and possesses $\Phi(\theta)$ as its characteristic function. The demonstration is immediate: Let $X'$ be (except for two points) an inverse of the distribution $F$ determined by $\Phi(\theta)$.

$$X'(\alpha) = x \Leftrightarrow F(x) < \alpha \leqslant F(x+0),\ \alpha \neq 0,\ 1$$
$$X'(\alpha) = 0,\ \alpha = 0,\ 1$$

The special evaluations at the points 0 and 1 simply complete the definition of unbounded functions, $X'$, otherwise undefined there. $X'$ restricted to $S$ is the desired random variable $X$. Q.E.D.

Though for each random variable $X$ on a Riemann space, $Ee^{i\theta X}$ exists and defines a unique characteristic function of a random variable on a probability space and, thus, a unique distribution function $F(x)$, we *cannot* in general conclude that $EI_{[X<x]} = F(x)$. Examples to the contrary have been given in Section 3. On the other hand, for any continuous function c,

$$Ec(X) = \int_{-\infty}^{\infty} c(x)\, dF(x),$$

where $F$ is determined by $Ee^{i\theta X}$. Thus, all theorems of the Helly-Bray type leading to convergence of moments and convergence of characteristic functions are meaningful and true in the context of Riemann spaces.

In the case of random variables $X$ on a Riemann space, we will speak of the distribution function $F$ determined by $Ee^{i\theta X}$ as the distribution function of $X$.

Before stating what theorems involving statistically independent random variables are true in the Riemann case, we find it necessary to prove an existence theorem. This theorem, that there exist an infinite number of statistically independent random variables would be trivially true in the measure-theoretic context. It has content here, because we define the random variables on a Riemann space and, thus, limit the degree of their discontinuity.

THEOREM 24. *For any sequence of distribution functions $F_k$, $k = 1, 2, 3, \ldots$, of random variables on a probability space, there is a sequence of statistically independent random variables $X_1, X_2, X_3, \ldots$ on a Riemann space such that $X_k$ has distribution function $F_k$.*

*Proof.* It is well known that a product measure $\mu_1 \times \mu_2 \times \cdots$ can be defined on the product Borel field $\mathcal{B}_1 \times \mathcal{B}_2 \times \cdots$ of sets in the infinite dimensional hypercube $[0, 1]^\infty$ to yield a measure space. For an arbitrary sequence of distribution functions $F_k$ and corresponding to each dimension, $k$, a random variable $Y_k$ can be defined on this space such that it is a function of the $k$th coordinate only $Y_k(x_1, x_2, \ldots, x_k, \ldots) = Z(x_k)$, it has at most a countable number of discontinuities when considered as a function of the $k$th coordinate, and it has distribution function $F_k$. The sequence $Y_1, Y_2, Y_3, \ldots$ is made up of statistically independent random variables, each with the prescribed distribution function. It will be sufficient for the proof to show that these random variables can be transferred to $([0, 1], \mathcal{B}, \mu)$ without destroying their independence or distributions and without introducing discontinuities outside of a set of measure zero.

We now consider the mapping $\tau$ of $[0, 1]^\infty$ onto $[0, 1]$ defined by expressing each coordinate $x_1, x_2, \ldots$ of a point $x$ in $[0, 1]^\infty$ in some binary form and letting $\tau(x)$ be that point in $[0, 1]$ whose binary representation is $\alpha_1\beta_1\alpha_2\gamma_1\beta_2\alpha_3 \ldots$, as obtained by summing diagonally over the digits of the array

$$
\begin{aligned}
x_1 &= \cdot\alpha_1\alpha_2\alpha_3 \ldots \\
x_2 &= \cdot\beta_1\beta_2\beta_3 \ldots \\
x_3 &= \cdot\gamma_1\gamma_2\gamma_3 \ldots
\end{aligned}
$$

This mapping has been studied by Wiener [10] and has been shown to be one-to-one up to a set of measure zero, measurable and measure preserving. The set of points in $[0, 1]^\infty$ with one or more rational coordinates of the form $k/2^n$ with $k$ and $n$ integers, and a like set of points in $[0, 1]$, must be excluded before $\tau$ is one-to-one.

Let us define $\tau^{-1}$ as the mapping from $[0, 1]$ to $[0, 1]^\infty$ that is the inverse of $\tau$ in the region of $[0, 1]$ where $\tau$ is one-to-one, that is determined on any point with a unique binary representation $\cdot\alpha_1\beta_1\alpha_2\gamma_1\beta_2\alpha_3 \ldots$ as that point in $[0, 1]^\infty$ whose coordinates have the binary representation given in the above array, and that is determined in like manner on points with non-unique binary representation by always selecting the representation that terminates in zeros.

We now define the following sequence of random variables on $([0, 1], \mathcal{B}, \mu)$.

$$
\begin{aligned}
X_1(\alpha) &= Y_1[\tau^{-1}(\alpha)] \\
X_2(\alpha) &= Y_2[\tau^{-1}(\alpha)] \\
X_3(\alpha) &= Y_3[\tau^{-1}(\alpha)]
\end{aligned}
$$

From the properties of $Y_k$ and $\tau$ it is clear that $X_1, X_2, X_3 \ldots$ is a sequence of statistically independent random variables and $X_k$ has distribution function $F_k$. The fact that each $X_k$ is continuous almost everywhere must now be demonstrated.

For fixed $k$, the set of points $\alpha$ in $[0, 1]$ for which $\tau^{-1}(\alpha)$ is a point of discontinuity o $Y_k$ constitutes a set of measure zero (being contained in the union of the set where $\tau$ fails to be one-to-one and the image under $\tau$ of the set where $Y_k$ is discontinuous). Let $\alpha_0$ be any irrational chosen outside this set of measure zero.

For fixed $k$, and for $\varepsilon > 0$, there exists an $n_0$ such that if the first $n_0$ binary digits in the $k$th coordinate of $\tau^{-1}(\alpha)$ coincide with those of $\tau^{-1}(\alpha_0)$ then

$$|Y_k(\tau^{-1}(\alpha) - Y_k(\tau^{-1}(\alpha_0)| < \varepsilon.$$

On the other hand, $\pi^{-1}$ is such that for any $n_0$ there exists a $\delta$ such that $|\alpha - \alpha_0| < \delta$ implies that the first $n_0$ binary digits in the $k$th coordinate of $\pi^{-1}(\alpha)$ and $\pi^{-1}(\alpha_0)$ coincide (This following from the irrationality of $\alpha_0$). Thus, $X$ is continuous at the point $\alpha_0$. Q.E.D.

The following theorem provides a link with measure-theoretic probability theory and helps us to construct true theorems for Riemann spaces from our knowledge of theorems in the measure-theoretic realm. In the statement of the theorem it is understood that independence, convergence, expectation and distribution have either the probability functional space or probability space definitions, the choice being consistent with the domain of the random variables involved.

THEOREM 25. *Let* $(S, \mathfrak{J}, E)$ *be a Riemann Space. Let* $Y, Y_1, Y_2, Y_3, \ldots$ *be random variables defined on the probability space* $([0, 1], \mathfrak{B}, P)$ – $P$ *being the Lebesque measure— which are continuous almost everywhere and whose restrictions to* $S$ *are* $X, X_1, X_2, X_3, \ldots$, *respectively. The following are true:*

i) $X_1, X_2, X_3, \ldots$ *are statistically independent if and only if* $Y_1, Y_2, Y_3, \ldots$ *are statistically independent.*

ii) $$X_n \xrightarrow{P} X \text{ if and only if } Y_n \xrightarrow{P} Y$$

iii) $$X_n \xrightarrow{r} X \text{ if and only if } Y_n \xrightarrow{r} Y$$

iv) $$EX = EY$$

v) *$X$ has distribution function $F$ if and only if $Y$ has distribution function $F$.*

*Proof* of (i). It is clear that the independence of $Y_1, Y_2, Y_3, \ldots$ implies the independence of $X_1, X_2, X_3, \ldots$ To prove the converse it is sufficient to show that if $Y_1', Y_2', \ldots Y_N'$ is any choice of random variables taken from $Y_1, Y_2, Y_3, \ldots$ and if $f_k \in \mathfrak{J}_{Y'_k}$, $k = 1, 2, \ldots N$, then

(A) $$Ef_1 f_2 \ldots f_N = Ef_1 Ef_2 \ldots Ef_N,$$

implies independence of $Y'_1, Y'_2, Y'_3, \ldots Y'_N$. This follows, however, by extending independence to the minimal $\sigma$-fields in $\mathcal{B}$ induced by $Y'_1, Y'_2, \ldots, Y'_N$ from the corresponding fields of sets in $\mathcal{B}$ whose characteristic functions, $I_k$, are of the form

$$I_k = \lim f_k^{(n)}, \ f_k^{(n)} \in \mathfrak{J}_{Y'_k}, \ f_k^{(n)} \leqslant f_k^{(n+1)}.$$

These latter fields are independent by virtue of $(A)$ and

$$\lim E f_1^{(n)} f_2^{(m)} \cdots f_N^{(g)} = E I_1 I_2 \cdots I_N$$

$$\lim E f_1^{(n)} E f_2^{(m)} \cdots E f_N^{(g)} = E I_1 \, E I_2 \cdots E I_N.$$

Q.E.D.

From the theorems of this section we can immediately conclude that all the theorems traditionally associated with the central limit problem are meaningful and true for sums $S_N = \sum_{k=1}^{N} X_{Nk}$ of independent random variables defined on a Riemann space. We refer to the various forms of the weak law of large numbers, the normal convergence criteria, the Poisson convergence, theorems relating to the infinitely decomposable laws and the central limit theorem, itself. See Loève [5, Chapter VI] for a discussion of the central limit problem in the measure theoretic setting.

In view of the fact that all of the theorems of this section remain meaningful and true when $S$ in the Riemann space $(S, \mathfrak{J}, E)$ and any random variable $X$ on the Riemann space are assumed computable, we can consider the results of this section to be generalizations of the results obtained in Section 2 for computable probability spaces.

## 6. Implications for number theory

It is recalled that in Section 2 we constructed, for arbitrary integer $N$, a finite collection of $N$ statistically independent random variables, the sample values of any one of which formed a computable sequence. These random variables were defined on a computable probability space. It is now clear from the generalizations developed since Section 2 that we can construct an infinite sequence of statistically independant random variables, with the property that each generates as its sample values a computable sequence. This can be done if we define the random variables on a computable Riemann space. In either the case of a finite collection or an infinite collection of random variables, the construction may be made consistent with sample values that are either 0 or 1 and, as a result, the computable sequences of sample values may be regarded as the binary representation of computable numbers in the unit interval.

The converse of this construction, namely, to select an infinite collection of computable numbers from the unit interval in such a way that the digits in their binary representation coincide with the sample values of an infinite collection of statistically independent random variables (on a computable Riemann space), is of particular interest in one case. This is the case where the numbers selected $a_1, a_2, a_3, \ldots$ are all related to one another through a translation of digits; specifically, $a_1 = \alpha 2^1 (\text{mod } 1)$, $a_2 = \alpha 2^2 (\text{mod } 1)$, $a_3 = \alpha 2^3 (\text{mod } 1)$, ..., for some computable $\alpha$. In this case the binary digits in the fractional part of $2\alpha$, namely, $\alpha_1, \alpha_2, \alpha_3, \ldots$ $(\cdot \alpha_1 \alpha_2 \alpha_3 \ldots = 2\alpha (\text{mod } 1))$ have many of the properties required of random numbers, for the relative frequency of occurrence among these digits of any specified sequence, $\alpha_1^0, \alpha_2^0, \ldots, \alpha_M^0$ of $M$ digits 0 or 1 always converges to $2^{-M}$:

$$\lim_{N\to\infty} \frac{1}{N} \sum_1^N I_k = 2^{-M}$$
$$I_k = \begin{cases} 1 & \text{if } (\alpha_k, \alpha_{k+1}, \cdots, \alpha_{k+M-1}) = (\alpha_1^0, \alpha_2^0, \cdots, \alpha_M^0) \\ 0 & \text{otherwise.} \end{cases} \tag{I}$$

The question of characterizing a computable number $\alpha$ whose binary digits have the property expressed by (I) is the basic one. In this section we will prove that almost all numbers $\alpha$ in the unit interval have the property (I) and that the binary digits of any such $\alpha$ provide all the sample values (via translation) of an infinite sequence of statistically independent random variables on a Riemann space. We will also formulate a necessary and sufficient condition that a given number $\alpha$ should satisfy (I).

It is proposed that if the necessary and sufficient condition can be verified for the fractional parts of $\pi$ and $e$, that the interesting statistical behavior in the digital structure of these numbers, as investigated on the high speed computing machines [1, 6, 7], will be explained in terms of statistically independent random variables on a computable Riemann space. Moreover, a positive result for $\pi$ or $e$ or any other computable number would show the way to a general analytical technique for computing, without recourse to construction methods, a table of digits $\alpha_1, \alpha_2, \alpha_3, \ldots$ possessing property (I).

To show that almost all numbers $\alpha$ in the unit interval have the property (I), it will be sufficient to show that almost all $\alpha$ in the unit interval generate a sequence

$$S_\alpha = (a_1 \equiv \alpha 2^1 (\text{mod } 1),\ a_2 = \alpha 2^2 (\text{mod } 1), \ldots)$$

that is uniformly dense there. Having shown this, property (I) will follow from the properties of the statistically independent random variables $X_1, X_2, X_3, \ldots$ defined on the Riemann space $(S_\alpha, \mathfrak{J}, E)$ according to

$$X_j(a_k) = \begin{cases} 1 & \text{if } \alpha_{j+k-1} = 0 \\ 0 & \text{otherwise} \end{cases}$$
$$j = 1, 2, 3, \cdots,\ k = 1, 2, 3, \cdots,\ a_k = \alpha_k \alpha_{k+1} \alpha_{k+2} \cdots.$$

(It should be noticed that a more elementary and direct proof would be possible if we did not wish to relate the sample values of statistically independent random variables on a Riemann space with the digits of numbers that possess property (I). It should also be noticed that the reasoning of probability functional spaces permits us to define an infinite number of statistically independent random variables all of whose sample values correspond to the digits of one number, whereas the reasoning of sequence probability spaces limits us to a finite number of random variables. Disregarding the structure in the domains of definition, however, the functions $X_1, X_2, X_3, \ldots$ given here are identical with those presented in the elementary case in the proof of Theorem 5.)

To show that for almost all $\alpha$ in the unit interval the above sequence $S_\alpha$ is uniformly dense, we appeal to previous results [2] which show that the kneading transformation $T$ of the unit square onto itself is metrically transitive. If $(\cdot\alpha_1\alpha_2\alpha_3\ldots, \cdot\beta_1\beta_2\beta_3\ldots)$ is the binary representation of a point $(a, b)$ in the unit square, the kneading transformation is defined by

$$T(\cdot\alpha_1\alpha_2\alpha_3\ldots, \cdot\beta_1\beta_2\beta_3\ldots) = (\cdot\alpha_2\alpha_3\alpha_4\ldots, \alpha_1\beta_1\beta_2\ldots)$$

This important result on metric transitivity implies that for any measurable set $A$ in the unit square

$$\lim \frac{1}{N}\sum_1^N I_A[T^k(a, b)] = \mu X \mu(A)$$

for almost all $(a, b)$ in the unit square, the exceptional set of measure zero for which equality does not hold possibly depending on $A$. In particular, we can let $A$ be any diadic interval $B$ in the unit square of the form:

$$\frac{j}{2^N} \leqslant a < \frac{j+1}{2^M},\ 0 \leqslant b \leqslant 1,\ \text{for some } j,\ 0 \leqslant j \leqslant 2^{M-1}$$

and obtain,

$$\lim \frac{1}{N}\sum_1^N I_B(a_k) = \mu(B),$$

with $a_k = \alpha 2^k \pmod 1$, for almost all $\alpha$ in the unit interval. Though for each diadic set $B$ there corresponds an exceptional set of points $\alpha$ of measure zero, there is for the totality of all such $B$ (the totality being countable) one exceptional set of measure zero for which equality might not hold. The last equality, therefore, holds for all diadic intervals, provided $\alpha$ is outside a set of measure zero, and, thus, can be shown to hold also for all intervals, provided $\alpha$ is outside the same set of measure zero. This proves that the sequence $S_\alpha$ is uniformly dense in the unit interval for almost all $\alpha$ and we have the following theorem.

THEOREM 26. *Almost all numbers $\alpha$ in the unit interval possess property* (I). *If $\alpha = \alpha_0 \alpha_1 \alpha_2 \ldots$ is the binary representation of a particular number possessing property* (I) *then for positive integer $k$ the digits $\alpha_k, \alpha_{k+1}, \alpha_{k+2}, \ldots$ correspond to the sample values of a random variable $X_k$ defined on the Riemann space $(S_\alpha, \mathfrak{J}, E)$, $S_\alpha = (\alpha 2^1 (\text{mod } 1), \alpha 2^2 (\text{mod } 1), \ldots)$ and $X_k$, $k = 1, 2, 3, \ldots$ are statistically independent.*

The method used in proving this theorem gives no indication of which specific numbers $\alpha$ have property (I), nor, in fact, whether any computable numbers $\alpha$ (of which there are at most countably many) possess the property. For this we must use the stronger results of H. Weyl [9].

First, let us notice that the number $\alpha$ possesses property (I) if and only if sequence $S_\alpha$ is uniformly dense in the unit interval. The sufficiency of the latter condition was proved in the last theorem and the necessity follows by similar reasoning. Thus any necessary and sufficient condition for a sequence $S_\alpha$ to be uniformly dense in the unit interval, such as Weyl gives [9], will yield a condition for $\alpha$ to possess property (I). In this way we obtain the following theorem.

THEOREM 27. *The number $\alpha$ possesses property* (I) *if and only if*

$$\sum_{k=1}^{N} \exp(2\pi i j \alpha 2^k) = o(N)$$

*for every positive integer $j$.* ($i = \sqrt{-1}$.)

As stated earlier, the interesting problem of characterizing the class of numbers $\alpha$ for which the necessary and sufficient condition of this theorem holds (and, in particular, of determining whether the class includes the numbers $\pi$ and $e$) is left as an unsolved problem in this paper.

We will not investigate in this paper the relationships between numbers $\alpha$ possessing property (I) and the Kollektiv of von Mises, the normal numbers of Borel, the admissible numbers of Copeland, the random sequences of Church, and other concepts occurring in frequency theories of probability.

## 7. Acknowledgements

We wish to acknowledge helpful discussions that were carried on during the preparation of this paper with Drs. Ernest B. Leach and Louis J. Green.

## 8. References

[1]. Geiringer, H., On the statistical investigation of transcendental numbers. *Studies in mathematics and mechanics presented to Richard von Mises*. Academic Press Inc. New York (1954), 310–322.

[2]. Hopf, E., *Ergodentheorie*. Erg. Math. 5, No. 2, Julius Springer, Berlin (1937). Also, Seidel, W., Note on a metrically transitive system. *Proc. Nat. Acad. U.S.A.* 19 (1933), 453.

[3]. Kleene, S. C., *Introduction to metamathematics*. D. van Nostrand, New York (1952).

[4]. Kolmogorov, A. N., *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Erg. Math. 2, No. 3, Julius Springer, Berlin (1933). Also trans. into English by Nathan Morrison, Chelsea Publishing Co., New York (1950).

[5]. Loève, M., *Probability theory. Foundations. Random Sequences*. D. van Nostrand Co., Inc., Toronto, New York, London (1955).

[6]. Reitwiesner, G. W., An Eniac determination of $\pi$ and $e$ to more than 2000 decimal places. *MTAC* 4 (1950), 11–15.

[7]. Stoneham, R. G., A study of 60,000 digits of $e$. *Notices, Amer. Math. Soc.* 5, No. 1 (1958), 64.

[8]. Turing, A. M., On computable numbers, with an application to the entscheidungsproblem. *Proc. London Math. Soc.* (2) 42 (1936), 230–265.

[9]. Weyl, H., Über die Gleichverteilung von Zahlen mod. eins. *Math. Ann.* 77 (1916), 313–352

[10]. Wiener, N. & Paley, R. E. A. C., *Fourier transforms in the complex domain*. Amer. Math. Soc. Colloq. Pub. 19, New York (1934).

*Received August 8, 1958*
*in revised form September 7, 1959.*