Теорія Ймовір. та Матем. Статист.
Вип. 71, 2004

Theor. Probability and Math. Statist.
No. 71, 2005, Pages 165–179
S 0094-9000(05)00656-3
Article electronically published on December 30, 2005

A STRUCTURAL APPROACH TO SOLVING THE 6TH HILBERT PROBLEM

UDC 519.21

YU. I. PETUNIN AND D. A. KLYUSHIN

Abstract. The paper deals with an approach to solving the 6th Hilbert problem based on interpreting the field of random events as a partially ordered set endowed with a natural order of random events obtained by formalization and modification of the frequency definition of probability. It is shown that the field of events forms an atomic generated, complete, and completely distributive Boolean algebra. The probability distribution of the field of events generated by random variables is studied. It is proved that the probability distribution generated by random variables is not a measure but only a finitely additive function of events in the case of continuous random variables (both rational- and real-valued).

Introduction

In 1900 in his lecture [1] David Hilbert formulated the problem of axiomatization of probability theory. Here is the corresponding quote from his lecture:

> "The investigations on the foundations of geometry suggest the problem: To treat in the same manner, by means of axioms, those physical sciences in which mathematics plays an important part; in the first rank are the theory of probabilities and mechanics.
>
> As to the axioms of the theory of probabilities, it seems to me desirable that their logical investigation should be accompanied by a rigorous and satisfactory development of the method of mean values in mathematical physics, and in particular in the kinetic theory of gases."

As is well known [2], there is no generally accepted solution of this problem so far. Therefore we will try to build a structural model of the empirical probability theory, that is, the model of the part of probability theory used in mathematical statistics as well as in related natural and social sciences.

First, let us consider an informal description of basic concepts of the empirical probability theory, with their properties and interplay. There exist four basic notions in modern probability theory: a random trial, a random experiment, a random event, and the probability of a random event.

We begin this description with the notion of a random trial whose position between other concepts is dominating: there was, there is, and there will be no probability theory without random trials, since one cannot rigorously define probability characteristics of random events and random variables, operations with them and any mathematical study of stochastic phenomena without having given a rigorous definition of random trials.

2000 *Mathematics Subject Classification.* Primary 60A05.

©2005 American Mathematical Society

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use

Moreover, the lack of a strict definition of a random event leads to confusion and to absurd statements.

We interpret as a trial T any action D carried out in a system S [3, 4] implying a certain response R. The trial T is usually identified with the action D. If the result of repeated accomplishing of the trial T is that, based on information on the system S, one can exactly predict the reaction R before carrying out (repeating) the trial T, the trial is called deterministic. Otherwise, if the reaction of the system cannot be determined in advance, before the trial is accomplished, we call it non-deterministic.

Let us give a mathematical definition of deterministic and non-deterministic trials. In the set of actions τ incoming to the system S, in the set σ of states of the system, and in the set $\mathfrak{R}$ of reactions of the system, introduce a distance between its elements (actions, states, and reactions). Therefore the sets τ, σ, and $\mathfrak{R}$ become metric spaces. A trial T is a function mapping each action and each state of the system to a reaction of the system $T\colon \tau \times \sigma \to \mathfrak{R}$. If this function is continuous, the trial is considered to be deterministic; if the function is discontinuous, the trial is said to be non-deterministic.

Non-deterministic trials can be divided into two classes: random and non-random. In order to distinguish between these two classes, we need some mathematical formalism.

Consider a trial T having two consequences: A and $\bar{A}$. Introduce the following quantity:

$$x_k = \begin{cases} 1 & \text{if the } k\text{th repetition of the trial } T \text{ leads to the consequence } A, \\ 0 & \text{otherwise.} \end{cases}$$

The numerical sequence $x_1, x_2, \dots$ of zeros and ones is called a Bernoulli sequence of order p, $0 \le p \le 1$, if

$$\lim_{n\to\infty} h_n(T, A) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} x_k = p.$$

It turns out that the knowledge of results of the one and only sequence

$$X = \{x_1, x_2, \dots\}$$

is not sufficient for a mathematical definition of a random trial. One needs to be aware of results of a series of trials $X_1, X_2, \dots$ conveniently represented by the following matrix:

$$\Theta(T) = \begin{pmatrix} x_{11} & \dots & x_{1n} & \dots \\ x_{21} & \dots & x_{2n} & \dots \\ \dots & \dots & \dots & \dots \\ x_{n1} & \dots & x_{nn} & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix}.$$

This matrix is called the characterization matrix of the trial T. Let

$$X_i = (x_{i1}, x_{i2}, \dots, x_{in}, \dots)$$

be rows and $X_j^* = (x_{1j}, x_{2j}, \dots, x_{nj}, \dots)$ be columns of the characterization matrix $\Theta(T)$.

It is easy to observe that each row X_n and each column X_n^* of the matrix $\Theta(T)$ generate real numbers α_n and α_n^* belonging to the interval $[0, 1]$. Indeed, put

$$\alpha_n = \overline{0.x_{n1}x_{n2}\dots x_{nn}\dots} \quad \text{and} \quad \alpha_n^* = \overline{0.x_{1n}x_{2n}\dots x_{nn}\dots}$$

and consider these expressions as binary fractions. Denote the sets of numbers α_n and α_n^* by M and M^*, respectively.

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use

Definition 1. A non-deterministic trial T is called random if the following conditions are satisfied:

1) all rows X_n and all columns X_n^* $(n = 1, 2, \dots)$ of the characterization matrix $\Theta(T)$ are Bernoulli sequences of the same order $p \in [0, 1]$;
2) the sets of numbers M and M^* generated by rows and columns of the characterization matrix $\Theta(T)$, respectively, are everywhere dense in the interval $[0, 1]$.

Definition 2. An infinite series of repetitions of a random trial T is called a random experiment E.

Definition 3. A result R of a random trial T generating a random experiment E is called the random event (E, R).

Remark 1. The result R can be a consequence either of A or of $\bar{A}$.

Definition 4. The order $p \in [0, 1]$ of a Bernoulli sequence formed by the results of a random trial T generating a random experiment E is called the probability $p(E, A)$ of the random event (E, A).

Remark 2. Let us emphasize that the concept of a random event and that of the probability of this event can only be defined given that a random experiment E is accomplished, not a random trial T, since we need an infinite series of repetitions of the trial T.

Remark 3. Generally speaking, the analysis of randomness is carried out in practice using finite matrices. The following criterion can be used for this purpose: a trial T is considered to be random if

1) all rows X_i and all columns X_i^* $(i = 1, 2, \dots, n)$ of the truncated characterization matrix $\Theta_n(T)$ are segments of Bernoulli sequences of the same order $p \in [0, 1]$;
2) for any $\varepsilon > 0$ there exists n such that the sets of numbers M_n and M_n^* generated by rows and columns of the truncated characterization matrix

$$\Theta_n(T) = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{pmatrix}$$

form an ε-net on the interval $[0, 1]$.

Definition 1 of a random trial T whose set of sample consequences is formed by two sets A and $\bar{A}$ can easily be extended to random trials having any set of sample consequences. Let A be a consequence of the trial T. Denote by T_A a non-deterministic trial having two sample consequences A and $\bar{A}$.

Definition 5. The trial T_A is called the restriction of a trial T to the events A and $\bar{A}$.

Definition 6. A non-deterministic trial T is called random if any restriction of this trial to arbitrary events A and $\bar{A}$ is a random trial.

Definition 7. An experiment is called pseudo-random if at least one of the sets M and M^* generated by rows and columns of the corresponding characterization matrix contains a finite number of different elements.

Theorem 1. *The probability that, in the Bernoulli scheme, the sets M and M^* formed by rows and columns of the characterization matrix $\Theta(T)$ are everywhere dense is equal to one.*

Proof. Let us prove the theorem for the set M.

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use

Assume that α is an arbitrary real number of the interval $[0,1]$ represented in the binary form $\alpha = \overline{0.i_1 i_2 \dots i_n \dots}$. Consider an arbitrarily small number $\varepsilon > 0$ and choose a positive integer n_0 to satisfy $1/2^{n_0} < \varepsilon/2$. Put

$$\tilde{\alpha} = \overline{0.0 \dots 0 i_{n_0+1} i_{n_0+2} \dots} \quad \text{and} \quad \hat{\alpha} = \alpha - \tilde{\alpha} = \overline{0.i_1 i_2 \dots i_{n_0} 0 \dots 0 \dots}.$$

Denote by A the random event consisting of the fact that the random trial T repeated n_0 times will produce the sequence $i_1, i_2, \dots, i_{n_0}$. Assume that the sequence $i_1, i_2, \dots, i_{n_0}$ contains k ones located in some fixed positions and $(n_0 - k)$ zeros. In this case, it can readily be seen that the probability of the event A is calculated as follows:

$$p(T_1, T_2, \dots, T_{n_0}, A) = p^k q^{n_0 - k} > 0,$$

where $q = 1 - p$. Then the probability that A does not occur is

$$\delta = p(T_1, T_2, \dots, T_{n_0}, \bar{A}) = 1 - p(T_1, T_2, \dots, T_{n_0}, A) < 1.$$

In the sequence of independent series of Bernoulli trials

$$\begin{gathered} i_1^{(1)}, i_2^{(1)}, \dots, i_{n_0}^{(1)}, \\ i_1^{(2)}, i_2^{(2)}, \dots, i_{n_0}^{(2)}, \\ \dots \\ i_1^{(n)}, i_2^{(n)}, \dots, i_{n_0}^{(n)}, \\ \dots \end{gathered}$$

the probability that A occurs at least once is equal to one. Indeed, the probability of the complement of A is at most δ^n for any positive integer n. Therefore this probability equals zero. Hence there exists with probability one a row of the characterization matrix $\Theta(T)$ whose first n_0 elements are $i_1, i_2, \dots, i_{n_0}$. This row defines the number $\bar{\alpha}$ belonging to M and satisfying

$$|\alpha - \bar{\alpha}| < \frac{1}{2^{n_0+1}} + \frac{1}{2^{n_0+2}} + \dots = \frac{1}{2^{n_0}} < \frac{\varepsilon}{2}. \tag{1}$$

Let $(\gamma, \eta) \subset [0,1]$ be an arbitrary interval. Put $\alpha = (\gamma + \eta)/2$ to obtain that there exists with probability one a number $\bar{\alpha} \in M$ satisfying (1); therefore $\bar{\alpha} \in [\gamma, \eta]$. Hence, in the Bernoulli scheme E, the set M of numbers generated by rows of the characterization matrix $\Theta(E)$ is everywhere dense in the interval $[0,1]$ with probability one.

The proof for the set M^* follows similar lines. □

Remark 4. By the strong law of large numbers (the Borel theorem), the probability that all rows X_n and all columns X_n^*, $n = 1, 2, \dots$, of the characterization matrix $\Theta(T)$ in the Bernoulli scheme are Bernoulli sequences of the same order $p \in [0,1]$ is equal to one [5, Appendix 1, p. 125].

According to our classification, the classical Bernoulli scheme is a random experiment.

A random experiment E should be distinguished from a pseudo-random experiment $E_{p.r.}$ for which the conditions E_A and E_B are not satisfied. Let us consider an example of a pseudo-random experiment. Assume that an experiment $E_{p.r.}$ consists in choosing positive integers $n \geq 2$ following their intrinsic order and determining whether the number is prime or not. Denote by A the event that the number is prime. If A occurs, we write one; otherwise we write zero. By infinite repeating of $E_{p.r.}$, the following sequence of zeros and ones is generated:

2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	...
1	1	0	1	0	1	0	0	0	1	0	1	0	0	0	1	0	...

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use

This is a Bernoulli sequence of order 1 (by the Legendre theorem, [6]); this sequence is irregular and non-deterministic. On the other hand, by explicitly writing the characterization matrix $\Theta(E_{p.r.})$, one can easily see that all rows of this matrix are identical and therefore the set M contains the one and only number.

1. The field of events

Assume that T is a random trial and $S(T)$ is the set of all random events that can occur if T is accomplished. Following the traditional definitions in classical probability theory, one can define addition and multiplication of events in the set $S(T)$ as well as complement of an event. Note that the first two operations are defined for arbitrary sets of events. Moreover, for any event A belonging to the set $S(T)$, the probability $p(A)$ is defined. Therefore $S(T)$ becomes a field called *the field of events* $S(E)$.

Let us emphasize that all these operations can be carried out for events belonging to *the same field of events* $S(E)$ only. It makes no sense to multiply or add events belonging to different fields of events similar to algebra where one cannot multiply elements belonging to different groups.

One can introduce a partial order in the field of events $S(E)$ generated by a random experiment E. We say that an event A implies an event B, and write

$$A \leq B,$$

if the event A occurring as the result of E necessarily implies that B should also occur. With this relation, the field of events $S\,(E)$ becomes a partially ordered set [7]. One can readily check that

$$A + B = \sup(A, B) = A \vee B \quad \text{and} \quad AB = \inf(A, B) = A \wedge B \tag{2}$$

for all events $A, B \in S(E)$. Therefore addition and multiplication can be expressed in terms of the partial order. These formulas can be extended to arbitrary sets of events:

$$\sum_{i \in J} A_i = \sup_{i \in J}\{A_i\} \quad \text{and} \quad \prod_{i \in J} A_i = \inf_{i \in J}\{A_i\}. \tag{3}$$

Since sums and products are well defined and belong to $S\,(E)$ by formula (3), the field of events $S(E)$ is a complete distributive lattice (structure) [7].

By the definition in [7], the complement of an element A belonging to a lattice containing zero (that is, $S(E)$ contains an element O satisfying $O \leq A$ for all $A \in S(E)$) is defined as an element $A' \in S$ such that $A \wedge A' = 0$ and $A \vee A' = I$. The lattice $S(E)$ is called a lattice containing complements if each element of the lattice has a complement. All these properties hold for the field of events, with the impossible event taken as O, the certain event taken as I and the negation of an event taken as the complement $\bar{A}$. Therefore $S(E)$ is a Boolean algebra containing complements.

A complete lattice $S(E)$ is called completely distributive [7] if it obeys two mutually dual distributive laws: for any non-empty family of index sets J_γ, $\gamma \in C$, we have

$$\begin{aligned} \bigwedge_{\gamma \in C}\Big[\bigvee_{\alpha \in J_\gamma} X_{\gamma,\alpha}\Big] &= \bigvee_{\varphi \in \Phi}\Big[\bigwedge_{\gamma \in C} X_{\gamma,\varphi(\gamma)}\Big], \\ \bigvee_{\gamma \in C}\Big[\bigwedge_{\alpha \in J_\gamma} X_{\gamma,\alpha}\Big] &= \bigwedge_{\varphi \in \Phi}\Big[\bigvee_{\gamma \in C} X_{\gamma,\varphi(\gamma)}\Big], \end{aligned} \tag{4}$$

where $X_{\gamma,\alpha}, X_{\gamma,\varphi(\gamma)} \in S(E)$, $\alpha \in J_\gamma$, and where Φ is the set of functions defined on C and satisfying $\varphi(\gamma) \in J_\gamma$.

Theorem 2. *The field of events $S(E)$ of an arbitrary random experiment E is a completely distributive complete Boolean algebra.*

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use

Proof. We proved already that $S(E)$ is a complete Boolean algebra. Let us prove that $S(E)$ is completely distributive. Indeed, put

$$X = \bigwedge_{\gamma\in C}\Big[\bigvee_{\alpha\in J_\gamma} X_{\gamma,\alpha}\Big] = \prod_{\gamma\in C}\Big[\sum_{\alpha\in J_\gamma} X_{\gamma,\varphi(\gamma)}\Big],$$

$$Y = \bigvee_{\varphi\in\Phi}\Big[\bigwedge_{\gamma\in C_\gamma} X_{\gamma,\varphi(\gamma)}\Big] = \sum_{\varphi\in\Phi}\Big[\prod_{\gamma\in C} X_{\gamma,\varphi(\gamma)}\Big],$$

$$X_{\gamma,\varphi(\gamma)}, X_{\gamma,\alpha} \in S(E).$$

The random events X and Y belong to the field of events $S(E)$ where

$$\alpha \in J_\gamma, \qquad \gamma \in C, \qquad \varphi\colon C \to J_\gamma, \qquad \varphi(\gamma) \in J_\gamma.$$

Assume that the random event X occurs in a random trial T whose repetitions form the random experiment E. Then for any $\gamma \in C$ there exists $\alpha_0^{(\gamma)} \in J_\gamma$ for which the random event $X_{\gamma,\alpha_0^{(\gamma)}}$ occurs. Define a function $\varphi\colon C \to J_\gamma$ as follows: $\varphi(\gamma) = \alpha_0^{(\gamma)} \in J_\gamma$ for all $\gamma \in C$. This implies that the event

$$\sum_{\varphi\in\Phi}\Big[\prod_{\gamma\in C} X_{\gamma,\varphi(\gamma)}\Big],$$

that is the event Y, occurs.

On the other hand, if the event Y occurs, then

$$\sum_{\varphi\in\Phi}\Big[\prod_{\gamma\in C} X_{\gamma,\varphi(\gamma)}\Big]$$

occurs, too. Thus there exists a function $\varphi_0 \in \Phi$ such that the event $\prod_{\gamma\in C} X_{\gamma,\varphi_0(\gamma)}$ occurs. Let $\varphi_0(\gamma) = \alpha_0^{(\gamma)} \in J_\gamma$. Then for all $\gamma \in C$ the event $X_{\gamma,\alpha_0^{(\gamma)}}$ occurs, and hence for all $\gamma \in C$ the event $\sum_{\alpha\in J_\gamma} X_{\gamma,\alpha}$ occurs. Therefore

$$\prod_{\gamma\in C}\Big[\sum_{\alpha\in J_\alpha} X_{\gamma,\alpha}\Big] = X$$

also occurs. This means that the event X occurs in the random trial T if and only if Y occurs, whence we derive that $X = Y$.

The second extended distributive law for the Boolean algebra $S(E)$ is proved along similar lines. This completes the proof of Theorem 2. □

Theorem 3 (Tarski)**.** *If a complete Boolean algebra S is completely distributive, then it is isomorphic to the algebra $2^{\aleph}$ of all subsets of a set M by structures of partially ordered spaces (or Boolean algebras)* [7].

By analyzing the proof of the Tarski theorem, one can see that the set M can be taken to be the collection of all atoms of the partially ordered space $S(E)$ [7].

Definition 8. A set of events $B = \{B_i\}_{i\in J}$ belonging to the field of events $S(E)$ is called basic if the following conditions are satisfied:

1) all sets B_i belonging to B are mutually disjoint: $B_iB_j = 0$ for $i \neq j$;
2) any event A in $S(E)$ can be represented as a sum of events B_i belonging to B:

$$A = \sum_{k\in K} B_{i_k}.$$

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use

Remark 5. The set P of all atoms of $S(E)$ is a basic set in $S(E)$. Therefore the algebra of events $S(E)$ is an atomically generated Boolean algebra. In classical probability theory, elements of the basic set $S(E)$ are called sample points.

2. Probability distributions in the field of events

A probability distribution $\mathsf{P}(E, A)$ in a field of events $S(E)$ is a function of two arguments: the random experiment E and the random event A. However, we will assume in the sequel that the random experiment E is fixed and therefore consider $\mathsf{P}(E, A)$ to depend on $A \in S(E)$ only.

By definition,

$$\mathsf{P}(A) = \lim_{n\to\infty} h_n(A), \tag{5}$$

where $h_n(A)$ is the frequency of the event A. The probability $\mathsf{P}(A)$ is a finitely additive function on $S(E)$. Therefore it is natural to ask whether or not the function $\mathsf{P}(A)$ is countably additive.

Assume that $A_1, A_2, \ldots, A_k, \ldots$ is a sequence of mutually disjoint events belonging to $S(E)$. Put $A = \sum_{k=1}^{\infty} A_k$. Since the function $h_n(A)$ is finitely additive, we have

$$p(A) = p\left(\sum_{k=1}^{\infty} A_k\right) = \lim_{n\to\infty} h_n\left(\sum_{k=1}^{\infty} A_k\right) = \lim_{n\to\infty}\lim_{k\to\infty}\left(\sum_{i=1}^{k} h_n(A_i)\right).$$

By changing the order of the limits, we would be able to prove that the probability $\mathsf{P}(A)$ is countably additive:

$$p(A) = \lim_{k\to\infty}\lim_{n\to\infty}\left(\sum_{i=1}^{k} h_n(A_i)\right) = \lim_{k\to\infty}\sum_{i=1}^{k}\lim_{n\to\infty} h_n(A_i) = \lim_{k\to\infty}\sum_{i=1}^{k} p(A_i) = \sum_{i=1}^{\infty} p(A_i).$$

However, there is no reason for this change of the order of the limits (see [8]).

Therefore it is impossible to assure that a probability distribution is countably additive. We will show in what follows that probability distributions, in general, are not measures.

3. Random variables and their probability characteristics

The notion of a random variable is one of the most important concepts in modern probability theory which, however, is not basic (for an axiomatic approach).

Indeed, a random variable x can be defined as a random experiment E having a numerical basic set $B(E) \subset \mathbf{R}^1$, or $B(E) \subset C$. In this case, particular values of the random variable x play the role of atoms of the partially ordered set $S(E)$.

In applications of probability theory, a random variable x is conveniently interpreted as a function defined on the basic set of the field of events $S(E)$, since any sample point $B_i \in B(E)$ is often mapped into the numerical quantity $x = x(B_i)$. It is easy to see that these two definitions are equivalent.

Let us switch to studying probability distributions on the field of events generated by values of a random variable. First, we consider random variables taking values in the set of rational numbers Q; this is the case coming from measurements in practical applications.

Definition 9. A random variable x is called rational-valued if it can take rational values only.

Denote by $B_E(x)$ the set of all possible values of a rational-valued random variable x whose values are observed in a random experiment E. We can assume that $B_E(x) = Q$ without loss of generality. Then $S(E)$ is formed by all possible subsets of the set Q.

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use

Definition 10. A rational-valued random variable x is called continuous if the distribution function $F_x(u)$ of this variable is continuous on $\mathbf{R}^1$. The corresponding probability distribution is called continuous.

Definition 11. A rational-valued random variable x is called singular if there exists a subset $\Psi = \{a_1, a_2, \dots, a_n, \dots\} \subset Q$ such that $p(E, \{a_n\}) = p_n > 0$ for all $n \in N$ and $\sum_{n=1}^{\infty} p_n = 1$. The corresponding probability distribution is called singular.

It is easy to observe that a continuous probability distribution $p(E, A)$ where $A \in S(E)$ corresponding to a rational-valued random variable in a fixed experiment E cannot be a countably additive set function, but only finitely additive. Indeed, we have in this case

$$p(E, (\alpha - \varepsilon, \alpha + \varepsilon)) = F(\alpha + \varepsilon) - F(\alpha - \varepsilon)$$

for an arbitrary rational number $\alpha \in B_E = Q$ and $\varepsilon > 0$.

Therefore

$$p(E, \{\alpha\}) \le p(E, (\alpha - \varepsilon, \alpha + \varepsilon)) = F(\alpha - \varepsilon, \alpha + \varepsilon) \to 0 \quad \text{as } \varepsilon \to 0.$$

Hence

$$p(E, \{\alpha\}) = 0.$$

If we assume that $p(E, A)$ is countably additive, then

$$p(E, Q) = \sum_{\alpha \in Q} p(E, \{\alpha\}) = 0,$$

since Q is a countable set.

On the other hand, $p(E, Q) = 1$. This contradiction shows that the assumption that $p(E, A)$ is countably additive is false. Therefore the probability distribution $p(E, A)$ on the field of events $S(E)$ is not a measure.

Theorem 4. *Assume that $F(u)$ is an arbitrary continuous distribution function on $\mathbf{R}^1$. Then there exists a random experiment E whose basic set (sample space) is $B_E = Q$ and whose probability distribution $p(E, A)$, $A \subset Q$, is such that*

$$p\left(E, Q_{(-\infty, u]}\right) = F(u), \tag{6}$$

for all $u \in \mathbf{R}^1$, where $Q_{(-\infty,u]} = Q \cap (-\infty, u]$.

Proof. The random experiment E mentioned in the theorem is obtained by repeating the random trial consisting in choosing at random an element from the set Q according to the probability distribution $p(E, A)$ satisfying condition (6). Therefore the main difficulty of the proof is to show that this probability distribution exists. The proof of this assertion is based on results in the theory of partially ordered vector spaces [9].

Recall that a vector space V is called a K-lineal if a class of positive elements $x > \theta$ is defined in V where θ denotes the zero element of the space V. We use this class to introduce an order relation $x > y \Leftrightarrow x - y > 0$ satisfying the following axioms:

1. If $x > \theta$, then $x \neq \theta$.
2. If $x > \theta$ and $y > \theta$, then $x + y > \theta$.
3. For two arbitrary elements $x, y \in V$ there exists $\sup\{x, y\} = x \vee y$.
4. If $x > \theta$ and $\lambda > 0$ is a number, then $\lambda x > \theta$.

Consider the vector space $V(Q)$ of all bounded functions defined on the set Q:

$$V(Q) = \{f(r), r \in Q, |f(r)| \le c\}.$$

It is easy to see that $V(Q)$ is a K-lineal if the set of all non-negative functions $f(r) \ge 0$ for all $r \in Q$, with some $r_0 \in Q$ such that $f(r_0) > 0$, is interpreted as the class of positive elements.

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use

A subset W of a K-lineal V is called majorizing with respect to V [9] if for all $f \in V$ there exists $f^* \in W$ satisfying $|f| \le f^*$. Choose a majorizing set W^* in V in the following way: consider the class K^* of all finite $(a, b]$ and infinite $(-\infty, b]$ half-closed intervals, and the infinite interval $(-\infty, \infty) = \mathbf{R}^1$. Let W^* be the set of all characteristic functions of intervals $(a, b] \in K^*$:

$$\chi_{(a,b]}(r) = \begin{cases} 1 & \text{if } r \in (a, b],\ r \in Q, \\ 0 & \text{if } r \notin (a, b],\ r \in Q, \end{cases}$$

and of their linear combinations. Then W^* is a linear subset of the K-lineal $V(Q)$; moreover, it is a majorizing set. Indeed, one can take as $f^*(r)$ the function

$$f^*(r) \equiv \sup_{r \in Q} |f(r)| \le c.$$

Denote by Θ the subset of W^* formed by the functions $\chi_{(-\infty,u]}(r)$ and $\chi_{(-\infty,\infty)}(r)$ where $r \in Q$, $u \in \mathbf{R}^1$. It is easy to see that the system of functions Θ is linearly independent and the vector subspace generated by this set coincides with the linear subset W^*, so that Θ is an algebraic basis in W^*. On each basic element $\chi_{(-\infty,u]}(r)$, we define a functional $\Phi(t)$ as follows:

$$\Phi\left(\chi_{(-\infty,u]}(r)\right) = F(u)$$

and extend this functional to the vector space W^* by linearity:

$$\Phi\left(\sum_{k=1}^{n} c_k \chi_{(-\infty,u_k]}\right) = \sum_{k=1}^{n} c_k F(u_k), \qquad u_k \ne u_i,\ i \ne k.$$

It is clear that

$$\Phi\left(\chi_{(a,b]}(r)\right) = \Phi\left(\chi_{(-\infty,b]}(r)\right) - \Phi\left(\chi_{(-\infty,a]}(r)\right) = F(b) - F(a).$$

By definition, the functional Φ is linear on the linear subset W^*. Let us show that this functional is also non-negative. Indeed, each function $f(r) \in W^*$ can be represented as follows:

$$f(r) = \sum_{k=1}^{n} c_k \chi_{(a_k,b_k]}(r),$$

where one of the numbers a_k can be equal to $-\infty$. Without loss of generality, we assume that $(a_i, b_i] \cap (a_k, b_k] = \varnothing$ for $i \ne k$. It is easy to see that $f(r) \ge 0$ if and only if $c_k \ge 0$. Then

$$\Phi(t) = \sum_{k=1}^{n} c_k \Phi\left(\chi_{(a_k,b_k]}(r)\right) = \sum_{k=1}^{n} c_k (F(b_k) - F(a_k)) \ge 0.$$

By the Kantorovich theorem [9], the non-negative additive functional $\Phi(f)$, $f \in W^*$, defined on the linear subset W^* majorizing $V(Q)$, can be extended to the whole of the K-lineal $V(Q)$; the extended functional will still be additive and non-negative. Denote this extension by $\tilde{\Phi}(f)$, $f \in V(Q)$. This functional defines a finitely additive function on the set of all subsets $A \subset Q$:

$$p(E, A) = \tilde{\Phi}(\chi_A(r)).$$

Indeed, assume that $A = \bigcup_{k=1}^{n} A_k$, $A_k \subset Q$, $A_k \cap A_i = \varnothing$, $i \ne k$. Then

$$\chi_A(r) = \chi_{\bigcup_{k=1}^{n} A_k}(r) = \sum_{k=1}^{n} \chi_{A_k}(r).$$

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use

Therefore

$$p(E,A)=\tilde{\Phi}\left(\chi_A(r)\right)=\tilde{\Phi}\Big(\sum_{k=1}^{n}\chi_{A_k}(r)\Big)=\sum_{k=1}^{n}\tilde{\Phi}\left(\chi_{A_k}(r)\right)=\sum_{k=1}^{n}p(E,A_k).$$

Furthermore, since the functional $\tilde{\Phi}(f)$ is non-negative, we have for all $A\subset Q$,

$$p(E,A)=\tilde{\Phi}\left(\chi_A(r)\right)\geq 0,$$

taking into account that $\chi_A(r)\geq 0$. Moreover, the function $p(E,A)$ is additive and therefore monotone. Hence for all $A\subset Q$,

$$0\leq p(E,A)\leq p(E,Q)=1$$

since $p(E,Q)=\Phi\left(\chi_{\mathbf{R}^1}(r)\right)=1$.

Thus we have extended the set function

$$p(E,(a,b])=\Phi\left(\chi_{(a,b]}(r)\right)=F(b)-F(a)$$

defined on all half-closed intervals $(a,b]$ in Q to the class of all subsets $A\subset Q$. The extended function takes values in the interval $[0,1]$ and is finitely additive. Therefore this function is a probability distribution on the field of events $S(E)$. Theorem 4 is proved. □

Remark 6. A rational-valued random variable satisfying the assumptions of Theorem 4 generates a probability distribution that is not countably additive and cannot be interpreted as a measure on a σ-ring of sets.

We emphasize that practical applications of probability theory are concerned with rational-valued random variables only, since an irrational number having an infinite number of decimals with an unknown regularity is an abstraction.

A continuous analog of Theorem 4 holds.

Theorem 5. *Assume that $F(u)$ is an arbitrary distribution function concentrated on the interval $[a,b]$:*

$$F(u)=\begin{cases}0 & \text{if } u\leq a,\\ 1 & \text{if } u\geq b.\end{cases}$$

Then there exists a random experiment E whose numerical basic set is $B_E=[a,b]$ and which generates the probability distribution $p(E,A)$ on all subsets $A\subset[a,b]$. Moreover

$$p(E,(a,u])=F(u) \tag{7}$$

for $u\in(a,b)$.

Proof. The random experiment E consists in choosing an element in $[a,b]$ at random according to the probability distribution $p(E,A)$ satisfying (7). The proof that this distribution exists is based on the following arguments.

Consider the vector field $V[a,b]$ of all bounded functions defined on the interval $[a,b]$:

$$V[a,b]=\{f(t)\colon t\in[a,b],|f(t)\leq c|\}.$$

It is easy to see that $V[a,b]$ is a K-lineal if the set of all non-negative functions $f(t)\geq 0$ for all $t\in[a,b]$ and such that there exists $t_0\in[a,b]$ satisfying $f(t_0)>0$ is taken as the set of positive elements. As the majorizing subset W^* of the K-lineal $V[a,b]$, we take the set of linear combinations of the indicators

$$\chi_{(\alpha,\beta]}(t)=\begin{cases}1 & \text{if } t\in(\alpha,\beta],\\ 0 & \text{if } t\notin(\alpha,\beta],\ (\alpha,\beta]\subset[a,b],\end{cases}$$

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use

of all half-closed intervals in $[a, b]$ and the indicator of the interval $[a, b]$. Then the subset $\Theta \subset W^*$ formed by the functions $\chi_{(a,u]}(t)$, $u \in (a, b]$, and $\chi_{[a,b]}(t)$ is linearly independent and the vector space generated by the set Θ coincides with the linear subset W^*. Therefore Θ forms an algebraic basis in W^*. Define a functional $\Phi(u)$ on each basic element $\chi_{(a,u]}(t)$ as follows:

$$\Phi(\chi_{(a,u]}(t)) = F(u)$$

and extend this functional to the whole of the vector space W^* by linearity. As was already done in the proof of Theorem 4, it is easy to check that the extended functional is also non-negative. Furthermore, the functional $\Phi(t)$, $t \in W^*$, is extended by the Kantorovich theorem (see [9]) to the whole of the K-lineal $V[a, b]$; the latter extension is still additive and non-negative. This extension $\tilde{\Phi}(t)$, $t \in V[a, b]$, of the functional $\Phi(t)$, $t \in W^*$, defines an additive non-negative function

$$p(E, A) = \tilde{\Phi}\left(\chi_A(t)\right)$$

on the set of all subsets $A \subset [a, b]$. Here, $\chi_A(t)$ is the characteristic function of the set A. The function $p(E, A)$ is the required probability distribution on the field of events $S(E)$. Theorem 5 is proved. □

Remark 7. Under the assumptions of Theorem 5, the probability distribution generated by a continuous function $F(u)$ is not a measure.

Indeed, assuming the contrary we obtain a measure defined on all subsets of the interval $[a, b]$, with the property that the measure of a one-point set equals zero. This contradicts the classical Ulam theorem [10] stating that, under the continuum hypothesis (or under the following weaker hypothesis: no cardinal number less than or equal to the continuum is weakly inaccessible) the above-mentioned measure is identically zero.

The fact that a probability distribution is always countably additive (as is assumed in the A. N. Kolmogorov model) has already been the matter of doubt of earlier researchers (de Finetti [11], Kac [12], and others), but no strict proof has been given. Theorems 4 and 5 imply that probability distributions generated by random variables (both rational- and real-valued) are not countably additive.

4. The notion of independence in the framework of structural approach

Two cases are of fundamental importance for studying the independence of random events and random experiments:

1) events A and B occur in the same random trial;
2) events A and B occur in two different random trials T_1 (where the event A occurs) and T_2 (where the event B occurs).

In case 2), the trials may coincide, that is, $T_1 = T_2 = T^*$. A trial is considered to be accomplished if an event being a result of this trial occurs.

We consider case 1) first. In this case, there is no importance to what event occurs first. Therefore we say that the event A is independent of the event B if the information on whether the event B occurs or not does not affect the probability of the event A. Otherwise, the event A is supposed to depend on B. The conditional probability of the event A in the experiment E generated by repeating the trial T is denoted by $p(E, A \mid B)$. Then the fact that the event A does not depend on B can be written as follows:

$$p(E, A \mid B) = p(E, A \mid \bar{B}).$$

Theorem 6. *Assume that E is a random experiment and let A and B be random events that can occur when E is accomplished. The random event A is independent of B if and only if*

$$p(E, A \mid B) = p(E, A).$$

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use

Corollary. *Events A and B are independent if and only if*

$$p(E \mid AB) = p(E \mid A)p(E \mid B). \tag{8}$$

Therefore formula (8) can be viewed as a definition of independent random events belonging to the same field of events.

Let us now switch to analyzing the notion of independence of two events belonging to two different fields of events $S(E_1)$ and $S(E_2)$ generated by random experiments E_1 and E_2, respectively. Assume that the random trial T_1, generating the random experiment E_1, is accomplished before the trial T_2 generating E_2 is. Then the random trials T_1 and T_2 can be combined to form a compound experiment $E = (E_1, E_2)$ if we agree to write first the trial T_1 that is accomplished first.

The order of appearance of the trials T_1 and T_2 is very important for the compound experiment since the fields of events $S(E)$ and $S\big(\widetilde{E}\big)$, where $E = (E_1, E_2)$ and $\widetilde{E} = (E_2, E_1)$, are Cartesian products of the fields $S(E_1)$ and $S(E_2)$ ($S(E_2)$ and $S(E_1)$, respectively). If we consider an isomorphism φ between these fields defined by the formula

$$\varphi(A, B) = (B, A),$$

where $A \in S(E_1)$ and $B \in S(E_2)$, the compound experiments E and $\widetilde{E}$ can generate probability distributions on the corresponding random fields $S(E)$ and $S\big(\widetilde{E}\big)$, which are not invariant with respect to this isomorphism, implying $p(E, (A, B)) \neq p\big(\widetilde{E}, (B, A)\big)$. These experiments are called *non-commutative*.

Definition 12. Let E_1 and E_2 be random experiments and let $E = (E_1, E_2)$ be the compound experiment. A random event $A_2 \in S(E_2)$ is called statistically independent of a random event $A_1 \in S(E_1)$ if

$$p(E, A) = p\big((E_1, E_2), (A_1, A_2)\big) = p(E_1, A_1)p(E_2, A_2). \tag{9}$$

Remark 8. If the compound experiment $\widetilde{E}$ consists in repeating the same random trial $\widetilde{T} = (T, T)$, that is $\widetilde{E} = (E, E)$, the definition of independence based on (8) is not the same as that based on (9).

Statistical independence of the event A_1 from the event A_2 in the random experiment $\widetilde{E} = (E_2, E_1)$ is defined in a similar manner:

$$p\big(\widetilde{E}, \widetilde{A}\big) = p\big((E_2, E_1), (A_2, A_1)\big) = p(E_2, A_2)p(E_1, A_1). \tag{10}$$

Observe that the concept of independence of two events A_1 and A_2 appearing in different random experiments is not reciprocal. The random event A_2 can depend on A_1 in the compound experiment $E = (E_1, E_2)$, while A_1 can be independent of A_2 in the experiment $\widetilde{E} = (E_2, E_1)$.

These definitions can easily be extended to an arbitrary finite set of random events $A_1 \in S(E_1), \ldots, A_k \in S(E_k)$.

Definition 13. If random events $A_1 \in S(E_1)$ and $A_2 \in S(E_2)$ are statistically independent of each other, that is formulas (9) and (10) hold, then we say that these events are statistically independent.

Definition 14. Two random experiments E_1 and E_2 are called statistically independent if arbitrary random events $A_1 \in S(E_1)$ and $A_2 \in S(E_2)$ are statistically independent.

Remark 9. The notion of independence of two random experiments E_1 and E_2 can be extended to an arbitrary finite set of experiments $E_1, \ldots, E_n$. If all experiments are independent and $E_i = E$, $i = 1, \ldots, n$, then these experiments form a Bernoulli scheme.

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use

5. Transformations of random variables

For random variables, one can introduce the operation of multiplying by a constant, as well as the standard arithmetic operations: addition, subtraction, multiplication, and division. In this case, a random variable is conveniently interpreted as an arbitrary function $x(B)$, $B \in B_T$, defined on the set of sample consequences of a random trial T. If random variables take their values as the result of accomplishing the same random trial T, there is no problem since all of them are functions defined on the same set of sample consequences of the random experiment T, that is on the set of atoms of the Boolean algebra $S(T)$. The situation becomes considerably different if random variables take values as the result of accomplishing different trials. In this case, random variables are defined on different sets of sample consequences, and the definition of arithmetic operations requires other tools, if ever possible.

Assume, for example, that the set of sample consequences of the random trial T_1 is the interval $[0,1]$ and that of the random trial T_2 is the interval $[2,3]$. Suppose that random variables x and y take values under the trials T_1 and T_2, respectively. Therefore the random variable x is interpreted as a function $x(B)$ defined on the interval $[0,1]$ and the random variable y is interpreted as a function $y(B)$ defined on the interval $[2,3]$ since $B_{E_1} = [0,1]$ and $B_{E_2} = [2,3]$. It is quite difficult to interpret the sum $x+y$ of these random variables.

There is no complete solution of this complicated problem so far. We can only describe a partial solution using the concept of compound random trials introduced in Section 4.

For simplicity, consider the solution of the problem in the case of two random variables $x(B)$, $B \in B_{T_1}$, and $y(B)$, $B \in B_{T_2}$, taking values as the result of two random trials T_1 and T_2. Let $T_c = \{T_1, T_2\}$ be the compound trial. Then $B_{T_c} = B_{T_1} \times B_{T_2}$ where $\times$ stands for the Cartesian product of the sets B_{T_1} and B_{T_2}. If the random trial T_c is accomplished, a random event $B_c = (B_1, B_2)$ occurs where $B_1 \in B_{E_1}$ and $B_2 \in B_{E_2}$, so that x takes the value $x(B_1)$ and y takes the value $y(B_2)$. Put

$$\begin{aligned}(x+y)(B_c) &= x(B_1) + y(B_2),\\ (x-y)(B_c) &= x(B_1) - y(B_2),\\ (xy)(B_c) &= x(B_1)y(B_2),\\ \left(\frac{x}{y}\right)(B_c) &= \frac{x(B_1)}{y(B_2)}.\end{aligned}$$

Unfortunately, these definitions are correct for arbitrary random variables $x(B_1)$ and $y(B_2)$ in the case of the so-called commutative random experiments E_1 and E_2 only.

Definition 15. Random experiments E_1 and E_2 are called commutative if

$$p\big((E_1,E_2),(A_1,A_2)\big) = p\big((E_2,E_1),(A_2,A_1)\big). \tag{11}$$

The fact that arithmetic operations with random variables are well defined actually means that addition and multiplication obey the commutative, associative, and distributive laws. However, a new problem emerges at this point: how to interpret the equality of two random variables, for example, $x+y$ and $y+x$ if E_1 and E_2 are different random experiments, since $x+y$ is a function defined on the set $B_{E_c} = B_{E_1} \times B_{E_2}$, while $y+x$ is defined on the set $B_{\widetilde{E}_c} = B_{E_2} \times B_{E_1}$, but $B_{E_c} \neq B_{\widetilde{E}_c}$? To answer this question, we must appeal to the first interpretation of random variables as random experiments having numerical basic sets. First we introduce the notion of isomorphic random experiments and isomorphic random events.

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use

Recall that a function $\Psi\colon P \to Q$ defined on an ordered set P and taking values in an ordered set Q is called an order-preserving, or isotonic, function if

$$x \le y \quad \text{implies that } \Psi(x) \le \Psi(y) \tag{12}$$

(see [7]). An isotonic function whose inverse is also isotonic is called an isomorphism. In other words, an isomorphism of two ordered sets is a one-to-one correspondence of these sets satisfying (12) and the condition

$$\Psi(x) \le \Psi(y) \quad \text{implies that } x \le y. \tag{13}$$

This property is called the inverse isotony of the mapping Ψ. The mapping Ψ itself is called a structural isomorphism of the ordered sets P and Q, since this mapping preserves orders in these sets.

Definition 16. Assume that E_1 and E_2 are two random experiments and let $S(E_1)$ and $S(E_2)$ be the fields of random events generated by the random experiments E_1 and E_2, respectively. The fields of events $S(E_1)$ and $S(E_2)$ are called isomorphic if there is a one-to-one correspondence Ψ between $S(E_1)$ and $S(E_2)$ such that Ψ is a structural isomorphism of the Boolean algebras $S(E_1)$ and $S(E_2)$ preserving probabilities of the corresponding events:

$$p(E_1, A) = p(E_2, \Psi(A)). \tag{14}$$

In this case, the random experiments E_1 and E_2 are also called isomorphic.

In other words, two fields of events $S(E_1)$ and $S(E_2)$ are called isomorphic if there exists a one-to-one correspondence

$$\Psi\colon S(E_1) \to S(E_2)$$

isotonic with respect to the ordering structures introduced in the Boolean algebras of corresponding events $S(E_1)$ and $S(E_2)$, and having the property of inverse isotony preserving probabilities of events, so that

$$p(E_1, A) = p(E_2, \Psi(A))$$

for all $A \in S(E_1)$. In this case, the mapping Ψ itself is called a probability isomorphism.

Theorem 7. *The fields of events $S(E_c)$ and $S(\widetilde{E}_c)$ generated by the combined compound experiments $E_c = \{E_1, E_2\}$ and $\widetilde{E}_c = \{E_2, E_1\}$ are isomorphic if the random experiments E_1 and E_2 are commutative.*

Now we come back to studying the commutative property of addition and multiplication of two random variables x and y generated by two different random experiments E_1 and E_2, respectively. As was already mentioned, $x+y$ and xy take values as the result of accomplishing the combined compound experiment $E_c = (E_1, E_2)$, while $y+x$ and yx take values as the result of accomplishing the combined compound experiment $\widetilde{E}_c = (E_2, E_1)$. We treat the random variable $x + y$ as a combined compound experiment E_c having the numerical basic set

$$B_{E_c} = B_{E_1} \oplus B_{E_2} = \{x + y\colon x \in B_{E_1}, y \in B_{E_2}\},$$

where B_{E_i}, $i = 1, 2$, is the numerical basic set of the random experiment E_i. In other words, B_{E_1} is the set of possible values of the random variable x and B_{E_2} is that of the random variable y. The random variable $y+x$ is identified with the random experiment $\widetilde{E}_c$ whose basic set is

$$B_{\widetilde{E}_c} = B_{E_2} \oplus B_{E_1} = B_{E_1} \oplus B_{E_2} = B_{E_c}.$$

As was already shown, the fields of events $S(E_c)$ and $S(\widetilde{E}_c)$ are isomorphic if the random experiments E_1 and E_2 commute, and therefore the random variables $x+y$ and $y+x$ are

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use

isomorphic. If we identify isomorphic objects, it is natural to accept that $x+y=y+x$ in this case; a similar identity holds for multiplication: $xy=yx$. One should note that the fields of events $S(E_c)$ and $S(\widetilde{E}_c)$ for the random variables $x+y$ and $y+x$ just coincide, since they are all possible subsets of the numerical sets $B_{E_c}=B_{\widetilde{E}_c}$ and since the random experiments E_c and $\widetilde{E}_c$ generate the same probability distribution on these fields in view of (14).

The other laws of arithmetical operations with random variables (associative, distributive, etc.) can be studied in a similar way.

Bibliography

1. D. Hilbert, *Lecture delivered before the International Congress of Mathematicians at Paris in 1900*, Bull. Amer. Math. Soc. **37** (2000), no. 4, 407–436. MR1779412
2. H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, N.J., 1946. MR0016588 (8:39f)
3. L. von Bertalanfy, *The theory of open systems in physics and biology*, Science **111** (1950), 23–29.
4. S. Beer, *Cybernetics and Management*, The English Universities Press, London, 1959.
5. V. I. Glivenko, *Probability Theory*, Gos. uchebno-pedagog. izd-vo, Moscow, 1937. (Russian)
6. K. Prachar, *Primzahlverteilung*, Springer-Verlag, Berlin–Göttingen–Heidelberg, 1957. MR0087685 (19:393b)
7. G. Birkhoff, *Lattice Theory*, 3rd. edn., American Mathematical Society, Providence, RI, 1967. MR0227053 (37:2638)
8. G. M. Fichtenholz, *A Course in Differential and Integral Calculus*, vol. 1, Gostekhizdat, Moscow–Leningrad, 1947; German transl., Differential- und Integralrechnung. I, VEB Deutscher Verlag der Wissenschaften, Berlin, 1986. MR0845555 (87j:00025)
9. B. Z. Vulikh, *Introduction to the Theory of Partially Ordered Spaces*, Figmatgiz, Moscow, 1961; English transl., Wolters-Noordhoff Scientific Publications, Ltd., Gröningen, 1967. MR0133668 (24 #A3494); MR0224522 (37:121)
10. John C. Oxtoby, *Measure and Category. A Survey of the Analogies between Topological and Measure Spaces*, 2nd. ed., Springer-Verlag, New York–Berlin, 1980. MR0584443 (81j:28003)
11. B. de Finetti, *Probability, Induction, and Statistics*, Wiley, New York, 1972. MR0440638 (55:13512)
12. M. Kac, *Probability and Related Topics in Physical Sciences*, Interscience Publishers, London–New York, 1959. MR0102849 (21:1635)

National Taras Shevchenko University, Faculty of Cybernetics, Volodymyrs'ka Street 64, Kyïv 01033, Ukraine

E-mail address: vm214@dcp.kiev.ua

National Taras Shevchenko University, Faculty of Cybernetics, Volodymyrs'ka Street 64, Kyïv 01033, Ukraine

Received 17/DEC/2001

Translated by V. ZAYATS

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use