ON A NEW AXIOMATIC THEORY OF PROBABILITY

By

ALFRÉD RÉNYI (Budapest), corresponding member of the Academy

Dedicated to Professors L. FEJÉR *and* F. RIESZ *on their* 75^{th} *birthday*

Introduction

The axiomatic foundation of probability theory given by A. N. KOLMOGOROV [1] in the year 1933 has been the starting point of a new and brilliant period in the development of probability theory. According to this theory to every situation (experiment, observation etc.) in which chance plays a role, there corresponds a *probability space* $[S, \mathcal{A}, \mathbf{P}]$, i. e. an abstract space S (the space of elementary events), a σ-algebra $\mathcal{A}$ (the set of events) of subsets of S, and a measure $\mathbf{P}(A)$ (the probability of the event A) defined for $A \in \mathcal{A}$ and satisfying $\mathbf{P}(S) = 1$. The theory of KOLMOGOROV furnished an appropriate and mathematically exact basis for the rapid development of probability theory, which took place in the last 30 years, as well as for its fruitfull application in a great number of branches of science, including other chapters of mathematics too. Nevertheless in the course of development there arose some problems concerning probability which can not be fitted into the frames of the theory of KOLMOGOROV.

The common feature of these problems is that in them *unbounded measures* occur, while in the theory of KOLMOGOROV probability is a bounded measure normed by the condition $\mathbf{P}(S) = 1$. Unbounded measures occur in statistical mechanics, quantum mechanics, in some problems of mathematical statistics (for example in connection with the application of the theorem of Bayes) as limiting distributions of Markov chains and Markov processes, in integral geometry, in connection with the applications of probability concepts in number theory etc.

In the theory of KOLMOGOROV it has, for instance, no sense to speak about a probability distribution which is uniform on the whole real axis or on the whole plane, further it has no sense to say that we choose an integer in such a way that all integers (or all non-negative integers) are equiprobable.

At the first glance it seems that unbounded measures can play no role in probability theory, because, in view of the connection between probability and relative frequency, probability clearly can not take any value greater than 1. But if we observe more attentively how unbounded measures are really used in all cases mentioned above, we see that unbounded measures are used only to calculate conditional probabilities as the quotient of the values of the unbounded measure of two sets (the first being contained in the second) and in this way reasonable values (not exceeding 1) are obtained. This is the reason why unbounded measures can be used with success in calculating (conditional) probabilities. But as the use of unbounded measures can not be justified in the theory of KOLMOGOROV, the necessity arises to generalize this theory. In the present paper such an attempt is made.

Clearly in a theory in which unbounded measures are allowed, *conditional probability must be taken as the fundamental concept.* This is natural also from an other point of view. In fact, the probability of an event depends essentially on the circumstances under which the event possibly occurs, and it is a commonplace to say that in reality every probability is conditional.

This has been realized by several authors; I mention here without aiming at completeness only H. JEFFREYS [2], H. REICHENBACH [3], J. KEYNES [4], R. KOOPMAN [5], A. COPELAND [6], G. A. BARNARD [7] and I. J. GOOD [8].

But none of the mentioned authors developed his theory on a measure-theoretic basic. The axiomatic theory developed in the present paper combines the measure-theoretic treatment of KOLMOGOROV with the idea proposed by the authors mentioned (and also by others) to consider conditional probability as the fundamental concept.

The novelty of the theory lies only in this combination. It follows from what has been said that in developing the theory proposed in this paper we follow the same way as KOLMOGOROV, only we try to go one step further. Thus, this new theory should be considered as a generalization of that of KOLMOGOROV. In fact, it contains the theory of KOLMOGOROV as a special case, but includes also cases which can not be fitted into the theory of KOLMOGOROV, namely cases in which conditional probabilities are calculated by means of unbounded measures.

Among the authors mentioned above only H. JEFFREYS uses explicitely unbounded probability distributions (especially random variables ξ for which $\log \xi$ is uniformly distributed on the whole real axis) but he does not give an exact mathematical meaning to such distributions, and restricts himself to the remark that it is a mere convention that to a certain event there corresponds the probability 1; he says that in some cases instead of 1 the value $+\infty$ can also be taken.

The attempt to extend the scope of the mathematical theory of probability with the aim to give a well founded basis for such calculating procedures which were successfully used in many fields of applications without being mathematically rigorously justified, has many analogues in the history of mathematics. I mention only, as one of the latest such successful attempts, that of S. L. SOBOLEV and L. SCHWARTZ, concerning the generalization of the notion of a function, to include, for instance, the function of DIRAC, together with its derivatives etc.[1]

The theory as presented in this paper is still far from being fully developed; we restrict ourselves to give here only the axioms and their immediate consequences (§ 1), further to discuss some examples (§ 2) and applications (§ 3), and finally to develop some "conditional" laws of large numbers (§ 4). As an application of the conditional strong law of large numbers, a generalization of BOREL's theorem on normal decimals for CANTOR's series is given.

I hope to return to the questions left unsolved in a forthcoming publication.

I had the occasion to lecture on the present theory in 1954 at several congresses: in Budapest [9] at the General Assembly of the Hungarian Academy in May 1954, at the Conference on probability and statistics in Prague in June 1954, at the International Congress in Amsterdam [10] in September 1954, at the Conference on probability and statistics in Berlin [11] in October 1954 and at the Conference on stochastic processes in Wrocław in December 1954. At these (and other) occasions many valuable remarks have been made. I mention only the following ones: B. V. GNEDENKO kindly informed me in Prague in June 1954 that in a lecture held some years ago in Moscow A. N. KOLMOGOROV himself has put forward the idea to develop his theory in such a manner that conditional probability should be taken as the fundamental concept, but he never published his ideas regarding this question. I was glad to learn from this information that my attempt, besides of being a continuation of the work of KOLMOGOROV, follows the lines which have been pointed out by himself. In § 3 an inequality of J. HÁJEK is used which he communicated to me in Prague in June 1954. A proof of this inequality is published in a joint paper of the author and J. HÁJEK [12] in this volume. Some interesting measure-theoretic problems, which arose in connection with the present paper, have been solved by Á. CSÁSZÁR: his results, which he

[1] The Dirac δ-"function" and the uniform distribution on the whole real axis (or the whole space etc.) arise in the same way in quantum mechanics and they are in a certain sense dual to each other. As a matter of fact, it is easy to verify the following fact connected with Heisenberg's uncertainty relation (for the sake of simplicity we restrict ourselves to the one dimensional case): if the wave function $\psi(x)$ of the position of a particle degenerates into a Dirac δ-"function", the wave function $\varphi(p)$ of the impulse of the particle degenerates into a function for which $|\varphi(p)|^2 \equiv \text{const.}$ is the "density" of a "uniform probability distribution in the whole phase space". This can be shown as follows: as it is known, denoting by $\varphi(p)$ the wave function of the position and by $\varphi(p)$ the wave function of the impulse of a particle, we have $\varphi(p) = \frac{1}{\sqrt{2\pi h}} \int_{-\infty}^{+\infty} \psi(x) e^{-\frac{ipx}{h}} dx$. This formula shows that if $\psi(x)$ degenerates into the Dirac function $\delta(x-x_0)$, we have $\varphi(p) = \frac{1}{\sqrt{2\pi h}} e^{-\frac{ipx_0}{h}}$ for which $|\varphi(p)|^2 = \text{const.}$

exposed at the first time as comments to my lecture [9], are published in his paper [13] in this volume and settle the question under what conditions can the conditional probability, introduced in the present paper as a set function of two set variables, be expressed in quotient form by means of (one or more) set functions of one variable. D. VAN DANTZIG called my attention to the work of KEYNES, COPELAND and KOOPMAN. JU. V. LINNIK called my attention to a paper of J. NEYMAN [14], where the foundations of the theory of probability are sketched in a form which has some points of contact with the point of view of the present paper. In Berlin A. N. KOLMOGOROV called my attention to those conditional probability spaces which I call "Cavalieri spaces" (§ 2). I hope to return to the thorough investigation of such spaces in a forthcoming paper. I owe to a discussion with E. MARCZEWSKI Theorem 14 of § 2. The result of 3. 5 has been suggested to the author by a remark of K. SARKADI. Á. CSÁSZÁR and J. CZIPSZER kindly read the manuscript of the present paper and made some valuable remarks which I have utilized in preparing the final form of this paper.

I am thankful to all those mentioned for their remarks and suggestions.

§ 1. The axioms and their immediate consequences

1. 1. Notations. In what follows if A and B are sets, we denote by $A+B$ the sum of the sets A and B (i. e. the set of those elements which belong at least to one of the sets A and B); AB denotes the product of the sets A and B (i. e. the set of those elements which belong to both of the sets A and B); to denote the sum resp. the product of a finite or infinite family of sets, we use also the notation Σ resp. Π. The empty set will be denoted by $\bigcirc$; $A \subseteq B$ expresses the fact that A is a subset of B; the subset of B consisting of those elements of B which do not belong to A will be denoted by $B-A$. If a is an element of the set A, this will be denoted by $a \in A$. If a does not belong to the set A, this will be denoted by $a \notin A$.

1. 2. Definitions and axioms. Let us be given an arbitrary set S; the elements of S which will be denoted by small letters $a, b, \ldots$ will be called elementary events. Let $\mathcal{A}$ denote a σ-algebra of subsets of S; the subsets of S which are elements of $\mathcal{A}$ will be denoted by capital letters $A, B, C, \ldots$ and called random events, or simply events. [The supposition that $\mathcal{A}$ is a σ-algebra means (see [15], p. 28) that 1. if $A_n \in \mathcal{A}$ $(n=1,2,\ldots)$, we have $\sum_{n=1}^{\infty} A_n \in \mathcal{A}$; 2. if $A \in \mathcal{A}$, we have $S-A \in \mathcal{A}$; 3. $\mathcal{A}$ is not empty. This implies that $\bigcirc \in \mathcal{A}$ and $S \in \mathcal{A}$.] Let us suppose further that a non empty subset $\mathcal{B}$

of $\mathcal{A}$ is given; we do not suppose any restrictions regarding the set $\mathcal{B}$.[2] We suppose finally that a set function $\mathbf{P}(A|B)$ of two set variables is defined for $A \in \mathcal{A}$ and $B \in \mathcal{B}$; $\mathbf{P}(A|B)$ will be called the conditional probability of the event A with respect to the event B. As the conditional probability of the event $A \in \mathcal{A}$ with respect to the event B is defined if and only if B belongs to $\mathcal{B}$, $\mathcal{B}$ may be called the set of possible conditions. We suppose that the set function $\mathbf{P}(A|B)$ satisfies the following axioms:

Axiom I. $\mathbf{P}(A|B) \geqq 0$, *if* $A \in \mathcal{A}$ *and* $B \in \mathcal{B}$; *further* $\mathbf{P}(B|B) = 1$, *if* $B \in \mathcal{B}$.

Axiom II. *For any fixed* $B \in \mathcal{B}$, $\mathbf{P}(A|B)$ *is a measure, i.e. a countably additive set function of* $A \in \mathcal{A}$, *i.e. if* $A_n \in \mathcal{A}$ $(n = 1, 2, \ldots)$ *and* $A_j A_k = \bigcirc$ *for* $j \neq k$ $(j, k = 1, 2, \ldots)$, *we have*

$$\mathbf{P}\left(\sum_{n=1}^{\infty} A_n \,\middle|\, B\right) = \sum_{n=1}^{\infty} \mathbf{P}(A_n|B).$$

Axiom III.[3] *If* $A \in \mathcal{A}$, $B \in \mathcal{A}$, $C \in \mathcal{B}$, *and* $BC \in \mathcal{B}$, *we have*

$$\mathbf{P}(A|BC) \cdot \mathbf{P}(B|C) = \mathbf{P}(AB|C).$$

If the Axioms I—III are satisfied, we shall call the set S, together with the σ-algebra $\mathcal{A}$ of subsets of S, the subset $\mathcal{B}$ of $\mathcal{A}$ and the set function $\mathbf{P}(A|B)$ defined for $A \in \mathcal{A}$, $B \in \mathcal{B}$, a *conditional probability space* and denote it for the sake of brevity by $[S, \mathcal{A}, \mathcal{B}, \mathbf{P}(A|B)]$.

1.3. Connection with Kolmogorov's theory. If $\mathbf{P}(A)$ is a measure (i. e. a countably additive and non-negative set function) defined on the σ-algebra $\mathcal{A}$ of subsets of the set S, if further $\mathbf{P}(S) = 1$, then the triple $[S, \mathcal{A}, \mathbf{P}(A)]$ is called a probability space in the sense of Kolmogorov, and if we define $\mathcal{A}^*$ as the set of those sets $B \in A$ for which $\mathbf{P}(B) > 0$ and put $\mathbf{P}(A|B) = \dfrac{\mathbf{P}(AB)}{\mathbf{P}(B)}$ for $A \in \mathcal{A}$, $B \in \mathcal{A}^*$, clearly $[S, \mathcal{A}, \mathcal{A}^*, \mathbf{P}(A|B)]$ is a conditional probability space which will be called the conditional probability space generated by the probability space $[S, \mathcal{A}, \mathbf{P}(A)]$.

Conversely, if $[S, \mathcal{A}, \mathcal{B}, \mathbf{P}(A|B)]$ is a conditional probability space and C is an arbitrary element of $\mathcal{B}$, putting $\mathrm{P}_C(A) = \mathbf{P}(A|C)$, $[S, A, \mathbf{P}_C(A)]$ will be a probability space in the sense of Kolmogorov. Thus a conditional probability space is nothing else than a set of ordinary probability spaces which

[2] It will be seen that our axioms imply that $\bigcirc \notin \mathcal{B}$, but it is possible that $\mathcal{B}$ contains all elements of $\mathcal{A}$ except $\bigcirc$; on the other hand, it is possible that $\mathcal{B}$ contains only one set. The theory is somewhat less general, but considerably simpler if it is supposed that $\mathcal{B}$ is an additive class of sets, i. e. if from $B_1 \in \mathcal{B}$ and $B_2 \in \mathcal{B}$ it follows $B_1 + B_2 \in \mathcal{B}$. Concerning the consequences of this supposition see [13].

[3] See 1.7 where an equivalent axiom (Axiom III') is discussed, and 1.8 where a stronger form of this axiom is mentioned.

are connected with each other by Axiom III. This connection is such that it is in conformity with the usual definition of conditional probability. Namely, if we put $\mathbf{P}_C(A) = \mathbf{P}(A|C)$ for $A \in \mathfrak{A}$ with $C \in \mathfrak{B}$ fixed, and define the conditional probability $\mathbf{P}^*(A|B)$ for a $B \in \mathfrak{B}$ for which $\mathbf{P}_C(B) > 0$, as usual in the theory of KOLMOGOROV, by $\mathbf{P}^*(A|B) = \frac{\mathbf{P}_C(AB)}{\mathbf{P}_C(B)}$, we have by Axiom III $\mathbf{P}^*(A|B) = \frac{\mathbf{P}(AB|C)}{\mathbf{P}(B|C)} = \mathbf{P}(A|BC)$. In case $S \in \mathfrak{B}$, clearly $[S, \mathfrak{A}, \mathbf{P}_S(A)]$ is a probability space in the sense of KOLMOGOROV. It must be mentioned that in this case $[S, \mathfrak{A}, \mathfrak{B}, \mathbf{P}(A|B)]$ may not be identical with the conditional probability space generated by $[S, \mathfrak{A}, \mathbf{P}_S(A)]$ because $\mathfrak{B}$ may contain sets B for which $\mathbf{P}(B|S) = 0$ and at the same time need not contain every set B for which $\mathbf{P}(B|S) > 0$, i. e. the system $\mathfrak{B}_S$ consisting of all sets $B \in \mathfrak{A}$ for which $\mathbf{P}(B|S) > 0$ need not be identical with $\mathfrak{B}$. However, if $\mathbf{P}^*(A|B)$ is defined by $\mathbf{P}^*(A|B) = \frac{\mathbf{P}(AB|S)}{\mathbf{P}(B|S)}$ for $B \in \mathfrak{B}_S$, we have $\mathbf{P}^*(A|B) = \mathbf{P}(A|B)$, provided that $B \in \mathfrak{B}$.

1. 4. Immediate consequences of the axioms. We shall prove some simple theorems which follow from our axioms. In what follows if $\mathbf{P}(A|B)$ occurs, it is always tacitly assumed that $A \in \mathfrak{A}$ and $B \in \mathfrak{B}$. We denote the set $S - A$ by $\bar{A}$.

THEOREM 1. $\mathbf{P}(A|B) = \mathbf{P}(AB|B)$.

PROOF. If in Axiom III $C = B$, we have $\mathbf{P}(A|B)\mathbf{P}(B|B) = \mathbf{P}(AB|B)$. Taking Axiom I into account Theorem 1 follows.

REMARK. It follows from Theorem 1 that $\mathbf{P}(S|B) = 1$; namely, by Theorem 1 $\mathbf{P}(S|B) = \mathbf{P}(SB|B) = \mathbf{P}(B|B)$ and thus, by Axiom I, $\mathbf{P}(S|B) = 1$.

THEOREM 2. $\mathbf{P}(A|B) \leqq 1$.

PROOF. According to Axiom II we have $\mathbf{P}(AB|B) + \mathbf{P}(\bar{A}B|B) = \mathbf{P}(B|B)$. As by Axiom I $\mathbf{P}(B|B) = 1$ and $\mathbf{P}(\bar{A}B|B) \geqq 0$, it follows $\mathbf{P}(AB|B) \leqq 1$ and thus by Theorem 1 we obtain $\mathbf{P}(A|B) \leqq 1$.

THEOREM 3. $\mathbf{P}(O|B) = 0$.

PROOF. According to Axiom II $\mathbf{P}(O|B) = \mathbf{P}(O + O|B) = 2\mathbf{P}(O|B)$ and thus $\mathbf{P}(O|B) = 0$.

REMARK 1. It follows from Theorem 3 that $O \notin \mathfrak{B}$, because if O would belong to $\mathfrak{B}$, we should have $\mathbf{P}(O|O) = 1$ by Axiom I and $\mathbf{P}(O|O) = 0$ by Theorem 3; thus the assumption $O \in \mathfrak{B}$ leads to a contradiction.

REMARK 2. It follows from Theorem 1 and 3 that if $AB = O$, then $\mathbf{P}(A|B) = 0$.

THEOREM 4. $\mathbf{P}(A|BC)\,\mathbf{P}(B|C)=\mathbf{P}(B|AC)\,\mathbf{P}(A|C)$.

PROOF. By Axiom III both expressions are equal to $\mathbf{P}(AB|C)$ and thus to each other.

THEOREM 5. *If* $A\subseteq A'\subseteq B\subseteq B'$, *we have*

$$\mathbf{P}(A|B')\leqq\mathbf{P}(A'|B).$$

PROOF. We have

$$\mathbf{P}(A|B')=\mathbf{P}(AA'B|B')=\mathbf{P}(AA'|BB')\,\mathbf{P}(B|B')\leqq$$
$$\leqq\mathbf{P}(AA'|B)=\mathbf{P}(A'|B)-\mathbf{P}(\bar{A}A'|B)\leqq\mathbf{P}(A'|B).$$

REMARK 1. If $A=A'$, we obtain the following special case of Theorem 5: if $A\subseteq B\subseteq B'$, we have $\mathbf{P}(A|B')\leqq\mathbf{P}(A|B)$.

REMARK 2. If $B=B'$, we obtain the following special case of Theorem 5: if $A\subseteq A'$, we have (without supposing that $A'\subseteq B$) $\mathbf{P}(A|B)\leqq\mathbf{P}(A'|B)$.[4] As a matter of fact, $\mathbf{P}(A|B)=\mathbf{P}(AB|B)$ and $\mathbf{P}(A'|B)=\mathbf{P}(A'B|B)$ by Theorem 1; if $A\subseteq A'$, we have $AB\subseteq A'B\subseteq B$, and Theorem 5 can be applied.

THEOREM 6. *If* $A_1+A_2\subseteq B_1B_2\in\mathfrak{B}$, *further* $\mathbf{P}(A_2|B_1)\,\mathbf{P}(A_2|B_2)>0$, *we have*

$$\frac{\mathbf{P}(A_1|B_1)}{\mathbf{P}(A_2|B_1)}=\frac{\mathbf{P}(A_1|B_2)}{\mathbf{P}(A_2|B_2)}.$$

PROOF. According to Axiom III

$$\mathbf{P}(A_1|B_1B_2)\,\mathbf{P}(B_1|B_2)=\mathbf{P}(A_1B_1|B_2)=\mathbf{P}(A_1|B_2) \tag{1}$$

and similarly

$$\mathbf{P}(A_2|B_1B_2)\,\mathbf{P}(B_1|B_2)=\mathbf{P}(A_2B_1|B_2)=\mathbf{P}(A_2|B_2). \tag{2}$$

Dividing (1) by (2) we obtain

$$\frac{\mathbf{P}(A_1|B_1B_2)}{\mathbf{P}(A_2|B_1B_2)}=\frac{\mathbf{P}(A_1|B_2)}{\mathbf{P}(A_2|B_2)}. \tag{3}$$

Interchanging B_1 and B_2 in (3) we obtain

$$\frac{\mathbf{P}(A_1|B_1B_2)}{\mathbf{P}(A_2|B_1B_2)}=\frac{\mathbf{P}(A_1|B_1)}{\mathbf{P}(A_2|B_1)}. \tag{4}$$

Comparing (3) with (4), Theorem 6 follows.

THEOREM 7. *If* $C\subseteq B=\sum_{k=1}^{\infty}B_k$ *and* $AB_jB_kC=O$ *for* $j\neq k$ $(j,k=1,2,\ldots)$, *we have*

$$\mathbf{P}(A|C)=\sum_{k=1}^{\infty}\mathbf{P}(A|B_kC)\,\mathbf{P}(B_k|C).$$

[4] It should be mentioned that this special case of Theorem 5 follows from Axioms I and II; Axiom III is not needed.

PROOF. By Axiom III $\mathbf{P}(A|B_kC)\mathbf{P}(B_k|C)=\mathbf{P}(AB_k|C)$ and thus by Axiom II and applying Theorem 1 twice, it follows

$$\sum_{k=1}^{\infty}\mathbf{P}(A|B_kC)\mathbf{P}(B_k|C)=\sum_{k=1}^{\infty}\mathbf{P}(AB_k|C)=\mathbf{P}(AB|C)=\mathbf{P}(ABC|C)=$$
$$=\mathbf{P}(AC|C)=\mathbf{P}(A|C).$$

REMARK. We mention the following consequences of Theorem 7: Let us suppose $B_k\in\mathfrak{B}$ and $B_jB_k=O$ for $j\neq k$, $B=\sum_{k=1}^{\infty}B_k$ and $B\in\mathfrak{B}$; if $\mathbf{P}(A|B_k)\leqq\lambda\mathbf{P}(A'|B_k)$ for $k=1,2,\ldots$ where $\lambda\geqq 0$, we have $\mathbf{P}(A|B)\leqq\leqq\lambda\mathbf{P}(A'|B)$.

Especially if $\mathbf{P}(A|B_k)=\lambda\mathbf{P}(A'|B_k)$ for $k=1,2,\ldots$, we have $\mathbf{P}(A|B)==\lambda\mathbf{P}(A'|B)$, further if $\mathbf{P}(A|B_k)=\lambda$ for $k=1,2,\ldots$, we have $\mathbf{P}(A|B)=\lambda$.

PROOF. Applying Theorem 7 twice with $C=B$ we obtain

$$\mathbf{P}(A|B)=\sum_{k=1}^{\infty}\mathbf{P}(A|B_k)\mathbf{P}(B_k|B)\leqq\lambda\sum_{k=1}^{\infty}\mathbf{P}(A'|B_k)\mathbf{P}(B_k|B)=\lambda\mathbf{P}(A'|B).$$

The two other assertions are evident consequences.

1.5. Representation of the conditional probability as a quotient. We shall give a sufficient condition under which the set function $\mathbf{P}(A|B)$ of two set variables can be represented in "quotient form", i. e. in the form $\mathbf{P}(A|B)=\dfrac{\mathbf{Q}(AB)}{\mathbf{Q}(B)}$ where the set function $\mathbf{Q}(A)$ is a measure on $\mathfrak{A}$ and satisfies $\mathbf{Q}(B)>0$ if $B\in\mathfrak{B}$.

THEOREM 8. *Let $[S,\mathfrak{A},\mathfrak{B},\mathbf{P}(A|B)]$ be a conditional probability space. Let us suppose that there exists a sequence of sets $B_n(n=0,1,\ldots)$ for which the following properties hold:*

a) $B_n\subseteq B_{n+1}$ $\quad(n=0,1,\ldots)$,

b) $\mathbf{P}(B_0|B_n)>0$ $\quad(n=1,2,\ldots)$,

c) *For any $B\in\mathfrak{B}$ there can be found a B_n for which $B\subseteq B_n$ and* $\mathbf{P}(B|B_n)>0$.

Then there exists a finite measure $\mathbf{Q}(C)$ defined for $C\in\mathfrak{A}^*$ where $\mathfrak{A}^*$ is the ring of those sets $C\in\mathfrak{A}$ for which there can be found a B_n with $C\subseteq B_n$, and this measure $\mathbf{Q}(C)$ has the following properties:

α) $\mathbf{Q}(B)>0$ *if* $B\in\mathfrak{B}$,

β) $\mathbf{P}(A|B)=\dfrac{\mathbf{Q}(AB)}{\mathbf{Q}(B)}$.

If the sequence B_n satisfies besides a), b), c) *also the following condition:*

d) $\lim_{n\to\infty}P(B_0|B_n)>0$,

then $\mathbf{Q}(C)$ *can be defined for all* $C \in \mathfrak{A}$ *and is a bounded measure on* $\mathfrak{A}$, *and thus, putting* $\mathbf{P}(C) = \frac{\mathbf{Q}(C)}{\mathbf{Q}(S)}$, *we have* $\mathbf{P}(S) = 1$. *Denoting by* $\mathfrak{B}^*$ *the set of those sets* $B \in \mathfrak{A}$ *for which* $\mathbf{P}(B) > 0$, *if* $\mathfrak{B}^*$ *is not identical with* $\mathfrak{B}$, *we may extend the definition* $\mathbf{P}(A|B)$ *to all* $B \in \mathfrak{B}^*$ *putting*

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(AB)}{\mathbf{P}(B)};$$

the conditional probability space $[S, \mathfrak{A}, \mathfrak{B}^*, \mathbf{P}(A|B)]$ *obtained in this way will be identical with the conditional probability space generated by the ordinary probability space* $[S, \mathfrak{A}, \mathbf{P}(A)]$.

PROOF. First we suppose only that the sequence B_n has the properties a), b) and c). Clearly $\mathfrak{A}^*$ is a ring and we have $\mathfrak{B} \subseteq \mathfrak{A}^* \subseteq \mathfrak{A}$. Let us consider a set $A \in \mathfrak{A}^*$, choose an index n for which $A \subseteq B_n$ and define $\mathbf{Q}(A)$ as follows:

$$\mathbf{Q}(A) = \frac{\mathbf{P}(A|B_n)}{\mathbf{P}(B_0|B_n)}. \tag{5}$$

It is clear that the value of $\mathbf{Q}(A)$ does not depend on the choice of n. As a matter of fact, if $A \subseteq B_n$ and $A \subseteq B_m$ where $n < m$, we have by Theorem 6

$$\frac{\mathbf{P}(A|B_n)}{\mathbf{P}(B_0|B_n)} = \frac{\mathbf{P}(A|B_m)}{\mathbf{P}(B_0|B_m)}.$$

Now, if $B \in \mathfrak{B}$, $\mathbf{Q}(B) > 0$, we have

$$\mathbf{P}(A|B) = \frac{\mathbf{Q}(AB)}{\mathbf{Q}(B)}. \tag{6}$$

This can be shown as follows: if $B \subseteq B_n$ and $\mathbf{P}(B|B_n) > 0$, we have

$$\frac{\mathbf{Q}(AB)}{\mathbf{Q}(B)} = \frac{\mathbf{P}(AB|B_n)}{\mathbf{P}(B_0|B_n)} \cdot \frac{\mathbf{P}(B_0|B_n)}{\mathbf{P}(B|B_n)} = \frac{\mathbf{P}(AB|B_n)}{\mathbf{P}(B|B_n)}.$$

Applying Axiom III we obtain

$$\frac{\mathbf{Q}(AB)}{\mathbf{Q}(B)} = \mathbf{P}(A|BB_n).$$

As $BB_n = B$, (6) is proved.

From (5) it can be seen that $\mathbf{Q}(A)$ is non-negative. To show that $\mathbf{Q}(A)$ is a measure, we have to prove that $\mathbf{Q}(A)$ is countably additive on $\mathfrak{A}^*$, i. e. if $A_k \in \mathfrak{A}^*$ $(k = 1, 2, \ldots)$ and $A_j A_k = \bigcirc$ for $j \neq k$ and $\sum_{k=1}^{\infty} A_k = A \in \mathfrak{A}^*$, then $\mathbf{Q}(A) = \sum_{k=1}^{k} \mathbf{Q}(A_k)$. This follows simply from the remark that if $A \subseteq B_n$, we have $A_k \subseteq B_n$ for $1, 2, \ldots$ and thus in the relations

$$\mathbf{Q}(A_k) = \frac{\mathbf{P}(A_k|B_n)}{\mathbf{P}(B_0|B_n)} \quad (k = 1, 2, \ldots); \qquad \mathbf{Q}(A) = \frac{\mathbf{P}(A|B_n)}{\mathbf{P}(B_0|B_n)}$$

the same B_n may be used, and therefore the countable additivity of $\mathbf{Q}(A)$ follows from that of $\mathbf{P}(A|B)$ for fixed $B \in \mathfrak{B}$ (Axiom II).

Thus the first part of Theorem 8 is proved. Now we suppose that the sequence B_n has besides a)—c) also the property d). By a well-known theorem (see p. e. [15], § 13, Theorem A) the definition of $\mathbf{Q}(A)$ can be extended to the smallest σ-ring $\mathfrak{A}^{**}$ containing $\mathfrak{A}^*$ in such a manner that $\mathbf{Q}(A)$ remains countably additive on $\mathfrak{A}^{**}$. Let us put $S^* = \sum_{n=0}^{\infty} B_n$; we shall show that $\mathfrak{A}^{**}$ is identical with the σ-algebra $\mathfrak{A}S^*$, i. e. with the set of all sets of the form AS^* where $A \in \mathfrak{A}$. As a matter of fact, $\mathfrak{A}^* \subseteq \mathfrak{A}S^*$ and $\mathfrak{A}S^*$ is a σ-ring, thus we have clearly $\mathfrak{A}^{**} \subseteq \mathfrak{A}S^*$. On the other hand, if $A \subseteq \mathfrak{A}$, we have

$$AS^* = \sum_{n=0}^{\infty} AB_n .$$

Now $AB_n \subseteq B_n$, and thus $AB_n \in \mathfrak{A}^*$, and therefore $AS^* \in \mathfrak{A}^{**}$; thus $\mathfrak{A}S^* \subseteq \mathfrak{A}^{**}$ which implies $\mathfrak{A}S^* = \mathfrak{A}^{**}$. Thus the definition of $\mathbf{Q}(A)$ can be extended to all $A \in \mathfrak{A}S^*$. We prove now that $\mathbf{Q}(A)$ is bounded on $\mathfrak{A}S^*$. To show this it is sufficient to prove that $\mathbf{Q}(S)$ is finite. But $S^* = \lim_{n\to\infty} B_n$ and $B_n \subseteq B_{n+1}$, and thus $\mathbf{Q}(S^*) = \lim_{n\to\infty} \mathbf{Q}(B_n)$ where $\mathbf{Q}(B_n)$ is non-decreasing. On the other hand,

$$\mathbf{Q}(B_n) = \frac{\mathbf{P}(B_n|B_n)}{\mathbf{P}(B_0|B_n)} = \frac{1}{\mathbf{P}(B_0|B_n)} .$$

Thus e) implies that $\mathbf{Q}(S^*) < +\infty$. Defining $\mathbf{Q}(A)$ by $\mathbf{Q}(A) = \mathbf{Q}(AS^*)$ for $A \in \mathfrak{A}$, $A \notin \mathfrak{A}S^*$, the definition of $\mathbf{Q}(A)$ is extended to the whole σ-algebra A. The final part of Theorem 8 concerning $\mathbf{P}(A)$ is obvious. Thus Theorem 8 is proved.

A necessary and sufficient condition for the existence of the quotient representation $\mathbf{P}(A|B) = \dfrac{\mathbf{Q}(AB)}{\mathbf{Q}(B)}$ is contained in the paper [13] of Á. CSÁSZÁR.

1.6. Random variables on a conditional probability space. Let $[S, \mathfrak{A}, \mathfrak{B}, \mathbf{P}(A|B)]$ be a conditional probability space. If $\xi = \xi(a)$ denotes a real-valued function defined for $a \in S$ which is measurable with respect to $\mathfrak{A}$, i. e. if A_x denotes the set of those $a \in S$ for which $\xi(a) < x$, we have $A_x \in \mathfrak{A}$ for all real x, we shall call ξ a random variable on $[S, \mathfrak{A}, \mathfrak{B}, \mathbf{P}(A|B)]$. Vector-valued random variables are defined similarly. The (ordinary) conditional probability distribution function of a random variable ξ with respect to an event $B \in \mathfrak{B}$ is defined by $F(x|B) = \mathbf{P}(A_x|B)$; if $F(x|B)$ is absolutely continuous, $F'(x|B) = f(x|B)$ is called the (ordinary) conditional probability density function of ξ with respect to B. The conditional mean value $\mathbf{M}(\xi|B)$ of ξ with respect to an event $B \in \mathfrak{B}$ is defined as the abstract Lebesgue integral

$$\mathbf{M}(\xi|B) = \int_S \xi(a)\, d\mathbf{P}(A|B)$$

of ξ with respect to the measure defined on S by $\mathbf{P}(A|B)$ with B fixed. Higher conditional moments, the conditional characteristic function etc. are defined similarly. The random variables ξ and η are called independent with respect to an event C, if denoting by A_x the set of those $a \in S$ for which $\xi(a) < x$ and by B_y the set of those $a \in S$ for which $\eta(a) < y$, we have $\mathbf{P}(A_x B_y|C) = \mathbf{P}(A_x|C)\mathbf{P}(B_y|C)$ for every real x and y.

As $[S, \mathfrak{A}, \mathbf{P}(A|B)]$ is for any fixed $B \in \mathfrak{B}$ a probability field in the sense of the theory of KOLMOGOROV, any theorem of ordinary probability theory remains valid when ordinary probabilities, distributions, mean values, independence etc. are replaced by conditional probabilities, conditional distributions, conditional mean values, conditional independence etc. with respect to the same $B \in \mathfrak{B}$. Problems which are peculiar to the theory presented in this paper are those in which conditional probabilities with respect to different conditions are figuring. We shall see in § 3 some examples of such problems.

Let us mention that if ξ is a random variable, and A_α^β denotes the set consisting of those elements $a \in S$ for which $\alpha \leq \xi(a) < \beta$, and if $A_\alpha^\beta \in \mathfrak{B}$ for a set X of intervals (α, β), then the conditional probabilities $\mathbf{P}(A_x^y|A_\alpha^\beta)$ can be considered for $(\alpha, \beta) \in X$ and thus ξ generates a conditional probability space on the real axis $\mathfrak{R}$, as the space of elementary events, the σ-algebra $\mathfrak{A}$ being the set of Borel subsets of $\mathfrak{R}$ and $\mathfrak{B}$ consisting of the intervals $(\alpha, \beta) \in X$.

This conditional probability space will be called the conditional probability distribution generated by ξ on the real axis.

Let $F(x)$ denote a non-decreasing function of x which is continuous to the left for $-\infty < x < +\infty$. If the set A_α^β belongs to $\mathfrak{B}$ whenever $F(\beta) - F(\alpha) > 0$, and we have for any subinterval (x, y) of such an interval (α, β)

$$\text{(7a)} \qquad \mathbf{P}(A_x^y|A_\alpha^\beta) = \frac{F(y) - F(x)}{F(\beta) - F(\alpha)},$$

we shall call $F(x)$ the generalized distribution function[5] of ξ; the function $F(x)$ is not uniquely determined, as together with $F(x)$ $G(x) = cF(x) + d$ where $c > 0$ is also a distribution function of ξ; but as $F(x)$ will be used only to calculate the conditional probabilities (7a), this will never lead to a misunderstanding. If the distribution function $F(x)$ of ξ is absolutely continuous, and $F'(x) = f(x)$, we shall call $f(x)$ the generalized density function of ξ; clearly $f(x)$ is determined only up to a positive constant factor. If $F(x) = x$ (i. e. $f(x) = 1$) for $-\infty < x < +\infty$, we shall say that the distribution of ξ is uniform in $(-\infty, +\infty)$.

[5] Conditions ensuring the existence of a generalized distribution function of a random variable ξ can be formulated by using the results of the paper [13].

If $f(x)$ is the generalized density function of ξ, we have

$$\mathbf{P}(A_x^y \mid A_\alpha^\beta) = \frac{\int_x^y f(u)\,du}{\int_\alpha^\beta f(u)\,du}. \tag{7b}$$

The generalized distribution function resp. density function of a random vector is defined similarly.

1. 7. An alternative form of Axiom III. We have already pointed out that our system of axioms can be characterised in the following manner: the set S, the σ-algebra $\mathcal{A}$ of subsets of S, the subset $\mathcal{B}$ of $\mathcal{A}$ and the set function of two set variables $\mathbf{P}(A|B)$ defined for $A \in \mathcal{A}$ and $B \in \mathcal{B}$ form a conditional probability space if $\mathcal{S}_B = [S, \mathcal{A}, \mathbf{P}(A|B)]$ is an ordinary probability space for every fixed $B \in \mathcal{B}$ and if the probability spaces $\mathcal{S}_C$ and $\mathcal{S}_{BC}$ are connected by Axiom III if $C \in \mathcal{B}$ and $BC \in \mathcal{B}$. Thus different probability fields can be combined to form a conditional probability field only if they are "compatible" in that sense that they satisfy Axiom III which can be considered as the condition of compatibility.

Now it is easy to prove

THEOREM 9. *Axiom III can be brought to the following equivalent form:*

AXIOM III′. *If $B \in \mathcal{B}$, $C \in \mathcal{B}$, $B \subseteq C$ and $\mathbf{P}(B|C) > 0$, we have for any $A \in \mathcal{A}$*

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(AB|C)}{\mathbf{P}(B|C)}.$$

PROOF. Clearly Axiom III′ is a special case of Axiom III. Conversely, if Axiom III′ is valid, Axiom III follows. This can be shown as follows: if $A \in \mathcal{A}$, $B \in \mathcal{A}$, $C \in \mathcal{B}$ and $BC \in \mathcal{B}$, two cases are possible: either $\mathbf{P}(B|C) = 0$ or $\mathbf{P}(B|C) > 0$. In the first case we have also[6] $\mathbf{P}(AB|C) = 0$ and thus $\mathbf{P}(A|BC)\mathbf{P}(B|C) = \mathbf{P}(AB|C)$ reduces to $0 = 0$. Now let us suppose $\mathbf{P}(B|C) > 0$. It is easy to see that Theorem 1 follows already from Axioms I—III′, and thus it can be applied in the present proof. But this means that $\mathbf{P}(BC|C) = \mathbf{P}(B|C) > 0$ and thus the conditions of Axiom III′ are satisfied with BC instead of B, and it follows from Axiom III′

$$\mathbf{P}(A|BC)\mathbf{P}(BC|C) = \mathbf{P}(ABC|C). \tag{8}$$

As $\mathbf{P}(BC|C) = \mathbf{P}(B|C)$ and $\mathbf{P}(ABC|C) = \mathbf{P}(AB|C)$, it follows from (8)

$$\mathbf{P}(A|BC)\mathbf{P}(B|C) = \mathbf{P}(AB|C).$$

Thus Axiom III follows from Axiom III′.

[6] See the footnote [4] to Remark 2 to Theorem 5.

REMARK. Theorem 9 means that Axiom III contains a compatibility condition for $\mathfrak{F}_B$ and $\mathfrak{F}_C$ where $B \subseteq C$ if and only if $\mathbf{P}(B|C) > 0$; if $\mathbf{P}(B|C) = 0$, $\mathfrak{F}_B$ and $\mathfrak{F}_C$ are compatible without any restriction. This fact is the basis of a general principle by use of which conditional probability spaces can be constructed. This principle is expressed in Theorem 14 of § 2.

1. 8. Extensions of a conditional probability space. If $[S, \mathfrak{A}, \mathfrak{B}, \mathbf{P}(A|B)]$ is a conditional probability space, it is natural to ask how could this space be extended, by including into $\mathfrak{B}$ sets $A \in \mathfrak{A}$ which are not contained in $\mathfrak{B}$. The most simple way is suggested by Axiom III, and is contained in the following

THEOREM 10. *Let B_1 denote a set for which $B_1 \in \mathfrak{A}$ and $B_1 \notin \mathfrak{B}$. If there exists at least one set B_2 with the following three properties: α) $B_2 \in \mathfrak{B}$, β) $B_1 \subseteq B_2$, γ) $\mathbf{P}(B_1|B_2) > 0$, further if for any other set B_3 which also has the properties α), β), γ), we have $B_2 B_3 \in \mathfrak{B}$, the definition of $\mathbf{P}(A|B)$ can be extended for $B = B_1$ by putting*

$$\mathbf{P}(A|B_1) = \frac{\mathbf{P}(AB_1|B_2)}{\mathbf{P}(B_1|B_2)}. \tag{9}$$

PROOF. If $\mathbf{P}(A|B_1)$ is defined by (9), Axioms I and II are clearly satisfied, therefore we have to verify only Axiom III. Three cases must be distinguished. a) If we put $B_1 = C'$ and B' is a set for which $B'C' \in \mathfrak{B}$, we must verify

$$\mathbf{P}(A|B'C')\mathbf{P}(B'|C') = \mathbf{P}(AB'|C').$$

But this means by (9)

$$\mathbf{P}(A|B_1B')\mathbf{P}(B_1B'|B_2) = \mathbf{P}(AB_1B'|B_2)$$

what is true by force of Axiom III, in view of $B_1B' = B'C' \in \mathfrak{B}$ and $B_2 \in \mathfrak{B}$. b) If $B_1 = B'C'$ where $C' \in \mathfrak{B}$, we must verify $\mathbf{P}(A|B_1)\mathbf{P}(B'|C') = \mathbf{P}(AB'|C')$. But this means by (9)

$$\mathbf{P}(B'|C')\mathbf{P}(AB_1|B_2) = \mathbf{P}(AB'|C')\mathbf{P}(B_1|B_2)$$

what reduces to $0 = 0$ if $\mathbf{P}(B'|C') = 0$ and to

$$\frac{\mathbf{P}(AB_1|B_2)}{\mathbf{P}(B_1|B_2)} = \frac{\mathbf{P}(AB'|C')}{\mathbf{P}(B'|C')} \tag{10}$$

if $\mathbf{P}(B'|C') > 0$. But as $B_1 = B'C'$, by applying Theorem 1 to both conditional probabilities on the right of (10) we find that (10) is equivalent to

$$\frac{\mathbf{P}(AB_1|B_2)}{\mathbf{P}(B_1|B_2)} = \frac{\mathbf{P}(AB_1|C')}{\mathbf{P}(B_1|C')}. \tag{11}$$

But (11) follows from Theorem 6, taking into account that

$$AB_1 + B_1 = B_1 \subseteq B_2 \quad \text{and} \quad AB_1 + B_1 = B_1 = B'C' \subseteq C',$$

and that clearly $\mathbf{P}(B_1|C')=\mathbf{P}(B'|C')>0$ and thus C' has the properties α), β) and γ) and therefore $B_2C'\in\mathfrak{B}$.

If $B'C'=B_1$ and $C'=B_1$ where $B'\in\mathfrak{A}$, we have to verify $\mathbf{P}(A|B'C')\mathbf{P}(B'|C')=\mathbf{P}(AB'|C')$. As $\mathbf{P}(AB'|C')=\mathbf{P}(AB'C'|C')$, this relation is trivially equivalent to $\mathbf{P}(A|B_1)=\mathbf{P}(AB_1|B_1)$, which is satisfied by force of (9).

It is easy to see that the definition of $\mathbf{P}(A|B_1)$ does not depend on the choice of B_2; as a matter of fact, if both B_2 and B_3 have properties α), β) and γ), it follows by Theorem 6 that

$$\frac{\mathbf{P}(AB_1|B_2)}{\mathbf{P}(B_1|B_2)}=\frac{\mathbf{P}(AB_1|B_3)}{\mathbf{P}(B_1|B_3)}.$$

It is also clear that $\mathbf{P}(AB_1|B_3)$ can not be defined otherwise as by (9) because if B_1 is included into $\mathfrak{B}$, (9) must hold by force of Axiom III.

Another possibility for including new sets into $\mathfrak{B}$ is yielded by passing to the limit; this procedure is described by the following

THEOREM 11. *Let us suppose that* $B_n\in\mathfrak{B}$, $B_n\subseteq B_{n+1}$, *further* $\mathbf{P}(B_n|B_{n+1})>0$ $(n=0,1,\ldots)$ *and that* $\prod_{n=0}^{\infty}\mathbf{P}(B_n|B_{n+1})$ *converges. If* $B_\infty=\sum_{n=0}^{\infty}B_n$ *does not belong to* $\mathfrak{B}$, *the definition of* $\mathbf{P}(A|B)$ *can be extended for* $B=B_\infty$, *by putting*

$$\mathbf{P}(A|B_\infty)=\lim_{n\to\infty}\mathbf{P}(A|B_n)\quad \textit{for any}\quad A\in\mathfrak{A}, \tag{12}$$

provided that the following condition is satisfied: if $B\in\mathfrak{B}$ *and* $\mathbf{P}(B|B_N)>0$ *for some* N, *we have* $BB_N\in\mathfrak{B}$.

PROOF. For an arbitrary $A\in\mathfrak{A}$ we put $A^{(0)}=AB_0$, $A^{(k)}=AB_k\bar{B}_{k-1}$ $(k=1,2,\ldots)$. As $A^{(k)}\subseteq B_n$ for $n\geqq k$, the sequence $\mathbf{P}(A^{(k)}|B_n)$ is by Theorem 5, Remark 1 monotonically decreasing for $n=k,k+1,\ldots$ and thus

$$\lim_{n\to\infty}\mathbf{P}(A^{(k)}|B_n)=\mathbf{P}(A^{(k)}|B_\infty)$$

exists.

Putting

$$\mathbf{P}^*(A|B_\infty)=\sum_{k=0}^{\infty}\mathbf{P}(A^{(k)}|B_\infty) \tag{13}$$

we have defined $\mathbf{P}^*(A|B_\infty)$ for every $A\in\mathfrak{A}$. To prove Theorem 11, we have to show that if $\mathbf{P}^*(A|B_\infty)$ is defined by (13), Axioms I—III remain valid, further that $\lim_{n\to\infty}\mathbf{P}(A|B_n)=\mathbf{P}(A|B_\infty)$ exists for all $A\in\mathfrak{A}$ and $\mathbf{P}(A|B_\infty)=\mathbf{P}^*(A|B_\infty)$, i. e. that (12) and (13) are equivalent. As regards Axiom I, it is clear that $\mathbf{P}^*(A|B_\infty)\geqq 0$; the validity of $\mathbf{P}^*(B_\infty|B_\infty)=1$ can be shown as follows: if $A = B_\infty$, $A^{(0)} = B_0$ and $A^{(k)} = B_k \bar{B}_{k-1}$ $(k = 1, 2, \dots)$ which implies

$$\mathbf{P}^*(B_\infty | B_\infty) = \mathbf{P}(B_0 | B_\infty) + \sum_{k=1}^{\infty} \mathbf{P}(B_k \bar{B}_{k-1} | B_\infty) = \lim_{N \to \infty} \mathbf{P}(B_N | B_\infty).$$

But

$$\mathbf{P}(B_N | B_\infty) = \lim_{n \to \infty} \mathbf{P}(B_N | B_n) = \lim_{n \to \infty} \prod_{k=N}^{n-1} \mathbf{P}(B_k | B_{k+1}) = \prod_{k=N}^{\infty} \mathbf{P}(B_k | B_{k+1})$$

and thus

$$\mathbf{P}^*(B_\infty | B_\infty) = \lim_{N \to \infty} \prod_{k=N}^{\infty} \mathbf{P}(B_k | B_{k+1}) = 1$$

because the remainder-products of a convergent infinite product are always tending to 1. Now let us verify Axiom II, i. e. that $\mathbf{P}^*(A | B_\infty)$ defined by (13) is countably additive. Let us suppose that $A_k \in \mathcal{A}$ and $A_j A_k = O$, if $j \neq k$ $(j, k = 1, 2, \dots)$, and $A_\infty = \sum_{n=1}^{\infty} A_n$. Let us put $A_n^{(0)} = A_n B_0$ and $A_n^{(k)} = A_n B_k \bar{B}_{k-1}$ for $k = 1, 2, \dots$, further $A_\infty^{(0)} = A_\infty B_0$ and $A_\infty^{(k)} = A_\infty B_k \bar{B}_{k-1}$ for $k = 1, 2, \dots$. As for $A \subseteq B_k$ we have

$$\mathbf{P}(A | B_\infty) = \mathbf{P}(A | B_k) \prod_{n=k}^{\infty} \mathbf{P}(B_n | B_{n+1}) \tag{14}$$

and $\mathbf{P}(A | B_k)$ is countably additive, further $A_\infty^{(k)} = \sum_{n=1}^{\infty} A_n^{(k)}$ and $A_n^{(k)} \cdot A_m^{(k)} = O$ for $n \neq m$, it follows

$$\mathbf{P}(A_\infty^{(k)} | B_\infty) = \sum_{n=1}^{\infty} \mathbf{P}(A_n^{(k)} | B_\infty). \tag{15}$$

But

$$\mathbf{P}^*(A_\infty | B_\infty) = \sum_{k=0}^{\infty} \mathbf{P}(A_\infty^{(k)} | B_\infty). \tag{16}$$

From (15) and (16) it follows

$$\mathbf{P}^*(A_\infty | B_\infty) = \sum_{k=0}^{\infty} \sum_{n=1}^{\infty} \mathbf{P}(A_n^{(k)} | B_\infty) = \sum_{n=1}^{\infty} \sum_{k=0}^{\infty} \mathbf{P}(A_n^{(k)} | B_\infty). \tag{17}$$

As we have from (13)

$$\sum_{k=0}^{\infty} \mathbf{P}(A_n^{(k)} | B_\infty) = \mathbf{P}^*(A_n | B_\infty), \tag{18}$$

it follows from (16) and (17) that

$$\mathbf{P}^*(A_\infty | B_\infty) = \sum_{n=1}^{\infty} \mathbf{P}^*(A_n | B_\infty), \tag{19}$$

i. e. that $\mathbf{P}^*(A | B_\infty)$ is countably additive.

Now we prove that

$$\mathbf{P}^*(A | B_\infty) = \lim_{N \to \infty} \mathbf{P}(A | B_N) \tag{20}$$

for any $A \in \mathcal{A}$. As a matter of fact, by (14)

$$\text{(21)} \qquad \mathbf{P}(A|B_N) = \sum_{k=0}^{\infty} \mathbf{P}(A^{(k)}|B_N) = \sum_{k=0}^{N} \mathbf{P}(A^{(k)}|B_N) = \frac{\sum_{k=0}^{N} \mathbf{P}(A^{(k)}|B_\infty)}{\prod_{n=N}^{\infty} \mathbf{P}(B_n|B_{n+1})}.$$

It follows from (21) that

$$\text{(22)} \qquad \lim_{N\to\infty} \mathbf{P}(A|B_N) = \sum_{k=0}^{\infty} \mathbf{P}(A^{(k)}|B_\infty)$$

and thus, by (13), (20) follows.

Now let us prove that Axiom III′ is valid. We distinguish three cases. The first case is when $C \in \mathcal{B}$, $B_\infty \subseteq C$ and $\mathbf{P}(B_\infty|C) > 0$; then we have $\mathbf{P}(B_N|C) > 0$ if N is sufficiently large and thus

$$\text{(23)} \qquad \mathbf{P}(A|B_N) = \frac{\mathbf{P}(AB_N|C)}{\mathbf{P}(B_N|C)}.$$

Passing to the limit $N \to \infty$ in (23), and using (20), it follows that

$$\text{(24)} \qquad \mathbf{P}(A|B_\infty) = \frac{\mathbf{P}(AB_\infty|C)}{\mathbf{P}(B_\infty|C)}.$$

The second case is when $B \in \mathcal{B}$, $B \subseteq B_\infty$ and $\mathbf{P}(B|B_\infty) > 0$; we have to prove that

$$\text{(25)} \qquad \mathbf{P}(A|B) = \frac{\mathbf{P}(AB|B_\infty)}{\mathbf{P}(B|B_\infty)}.$$

If there exists an index N for which $B \subseteq B_N$, we have for sufficiently large N

$$\frac{\mathbf{P}(AB|B_N)}{\mathbf{P}(B|B_N)} = \mathbf{P}(A|B)$$

and thus, passing to the limit $N \to \infty$, (25) follows.

In the general case, let us consider the two measures $\mu_1(A) = \mathbf{P}(A|B)$ and $\mu_2(A) = \mathbf{P}(A|B_\infty)$. Let $A \in \mathcal{A}$ denote an arbitrary set, for which $A \subseteq B_N B$ for some N. Then we have by Theorem 6, taking N sufficiently large, to ensure $\mathbf{P}(B|B_N) > 0$,

$$\frac{\mathbf{P}(A|B_N)}{\mathbf{P}(B|B_N)} = \frac{\mathbf{P}(A|B)}{\mathbf{P}(B|B)} \text{ and thus } \mu_2(A) = \mathbf{P}(B|B_\infty)\mu_1(A),$$

taking into account that $\lim_{N\to\infty} \mathbf{P}(B|B_N) = \mathbf{P}(B|B_\infty) = 1$ and thus $\mathbf{P}(B|B_N) > 0$ for sufficiently large N. Thus $\mu_2(A) = C\mu_1(A)$ where $C = \mathbf{P}(B|B_\infty)$ does not depend on A if A runs over all sets for which there exists an index N for which $A \subseteq B_N B$.

Clearly these sets A form a ring. As $\mu_2(A)$ and $C\mu_1(A)$ are finite measures, it follows that they coincide for the least σ-ring which contains all mentioned sets A. As $B_\infty = \sum_{N=0}^{\infty} B_N$, it follows that $\mu_2(A) = C\mu_1(A)$ if $A \subseteq B_\infty B$ and $A \in \mathfrak{A}$. As $B \subseteq B_\infty$, this means that $\mu_2(A) = C\mu_1(A)$ if $A \subseteq B$, i. e. we have proved that

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(A|B_\infty)}{\mathbf{P}(B|B_\infty)} \qquad \text{if } A \subseteq B.$$

As $AB \subseteq B$ for any $A \in \mathfrak{A}$, this means that for any $A \in \mathfrak{A}$ we have

$$\mathbf{P}(A|B) = \mathbf{P}(AB|B) = \frac{\mathbf{P}(AB|B_\infty)}{\mathbf{P}(B|B_\infty)}.$$

Thus (25) is proved. The third case is, when $B = B_\infty$ and $C = B_\infty$; in this case we have to prove that $\mathbf{P}(A|B_\infty) = \frac{\mathbf{P}(AB_\infty|B_\infty)}{\mathbf{P}(B_\infty|B_\infty)}$; but this is evident because we have shown that $\mathbf{P}(B_\infty|B_\infty) = 1$ and by (15) $\mathbf{P}(AB_\infty|B_\infty) = \mathbf{P}(A|B_\infty)$. Thereby the proof of Theorem 11 is accomplished.

REMARK. The assertion of Theorem 6 for the case $A_1 + A_2 \subseteq B_1 B_2$, but without the supposition $B_1 B_2 \in \mathfrak{B}$, can be considered as a stronger form of Axiom III, it shall be called Axiom III*.

AXIOM III*. *If* $A_1 \in \mathfrak{A}, A_2 \in \mathfrak{A}, B_1 \in \mathfrak{B}$ *and* $B_2 \in \mathfrak{B}$, *further* $A_1 + A_2 \subseteq B_1 B_2$ *and* $\mathbf{P}(A_2|B_1)\mathbf{P}(A_2|B_2) > 0$, *we have*

$$\frac{\mathbf{P}(A_1|B_1)}{\mathbf{P}(A_2|B_1)} = \frac{\mathbf{P}(A_1|B_2)}{\mathbf{P}(A_2|B_2)}.$$

As Theorem 1 is not a consequence of Axiom III*, in case we replace Axiom III by Axiom III*, we must suppose the validity of Theorem 1 as

AXIOM III**. $\mathbf{P}(A|B) = \mathbf{P}(AB|B)$ *for* $A \in \mathfrak{A}$ *and* $B \in \mathfrak{B}$.

Axiom III* and Axiom III** together imply Axiom III′ and thus Axiom III. As a matter of fact, choosing $A_1 = AB, A_2 = B, B_1 = B$ and $B_2 = C$ where $A \in \mathfrak{A}, B \in \mathfrak{B}, C \in \mathfrak{B}, B \subseteq C$ and $\mathbf{P}(B|C) > 0$, the conditions of Axiom III* are satisfied, and as by Axiom III** $\mathbf{P}(A|B) = \mathbf{P}(AB|B)$ it follows that

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(AB|B)}{\mathbf{P}(B|B)} = \frac{\mathbf{P}(AB|C)}{\mathbf{P}(B|C)},$$

i. e. Axiom III′ is really a consequence of Axiom III* and Axiom III**.

As Axiom III and Axiom III′ are equivalent, Axiom III is also a consequence of Axiom III*. Conversely, Axiom III* does not follow from Axiom III, only in the special case when $B_1 B_2 \in \mathfrak{B}$. If we would suppose Axioms III* and III** instead of Axiom III, the conditions of the extension theorems Theorem 10 and Theorem 11 could be reduced: in Theorem 10 the condition that for two sets B_2, B_3 with properties α), β) and γ), $B_2 B_3 \in \mathfrak{B}$ could be omitted, in Theorem 11 the condition that if $B \in \mathfrak{B}$ and $\mathbf{P}(B|B_N) > 0$, then

$BB_N \in \mathfrak{B}$ could be omitted. In what follows we do not suppose the validity of Axiom III*; it should be mentioned, however, that Axiom III* is contained as a special case in Axiom IV_2 introduced by Á. CSÁSZÁR [13].

If $\mathfrak{S} = [S, \mathfrak{A}, \mathfrak{B}, \mathbf{P}(A|B)]$ and $\mathfrak{S}' = [S', \mathfrak{A}', \mathfrak{B}', \mathbf{P}'(A|B)]$ are two such conditional probability spaces that $S \subseteq S'$, $\mathfrak{A} \subseteq \mathfrak{A}'$, $\mathfrak{B} \subseteq \mathfrak{B}'$ and $\mathbf{P}'(A|B) = \mathbf{P}(A|B)$ if $A \in \mathfrak{A}$ and $B \in \mathfrak{B}$, we shall say that the conditional probability space $\mathfrak{S}$ is imbedded into the conditional probability space $\mathfrak{S}'$. In the foregoing section we considered only such special imbeddings of a conditional probability space $\mathfrak{S}$ into a conditional probability space $\mathfrak{S}'$ for which $S' = S$, $\mathfrak{A}' = \mathfrak{A}$ and $\mathfrak{B} \subseteq \mathfrak{B}'$. These imbeddings of $\mathfrak{S}$ were obtained by extension of $\mathfrak{B}$ in several manners. If $\mathfrak{S}$ is imbedded into $\mathfrak{S}'$, we shall write $\mathfrak{S} \ll \mathfrak{S}'$.

1.9. Continuity properties of conditional probability. It follows from Axiom II that $\mathbf{P}(A|B)$ is for fixed $B \in \mathfrak{B}$ a bounded measure in A, and thus, of course, continuous in A, i. e. if $A_n \in A$ and $A_n \subseteq A_{n+1}$ (or $A_n \supseteq A_{n+1}$) for $n = 1, 2, \ldots$, we have for $B \in \mathfrak{B}$

$$\lim_{n\to\infty} \mathbf{P}(A_n|B) = \mathbf{P}(\lim_{n\to\infty} A_n|B).$$

As regards the continuity of $\mathbf{P}(A|B)$ as a function of B, we have

THEOREM 12. *If $B_n \in \mathfrak{B}$ and $B_n \subseteq B_{n+1}$ $(n = 1, 2, \ldots)$, further $\sum_{n=1}^{\infty} B_n = B \in \mathfrak{B}$, we have for $A \in \mathfrak{A}$*

$$\lim_{n\to\infty} \mathbf{P}(A|B_n) = \mathbf{P}(A|B).$$

PROOF. We have by Axiom III

$$\mathbf{P}(A|B_n) = \frac{\mathbf{P}(AB_n|B)}{\mathbf{P}(B_n|B)}$$

if $\mathbf{P}(B_n|B) > 0$. As $\lim_{n\to\infty} \mathbf{P}(B_n|B) = \mathbf{P}(B|B) = 1$, the last condition is satisfied for sufficiently large n, and thus

$$\lim_{n\to\infty} \mathbf{P}(A|B_n) = \frac{\lim_{n\to\infty} \mathbf{P}(AB_n|B)}{\lim_{n\to\infty} \mathbf{P}(B_n|B)} = \frac{\mathbf{P}(AB|B)}{\mathbf{P}(B|B)} = \mathbf{P}(A|B).$$

Thus Theorem 12 is proved.

The situation is somewhat more complicated if we consider a decreasing sequence of conditions. In this case we have

THEOREM 13. *If $B \in \mathfrak{B}$, $BC_n \in \mathfrak{B}$ and $C_n \supseteq C_{n+1}$ $(n = 1, 2, \ldots)$ further if putting $C = \prod_{n=1}^{\infty} C_n$ we have $BC \in \mathfrak{B}$ and $\mathbf{P}(C|B) > 0$, it follows that*

$$\lim_{n\to\infty} \mathbf{P}(A|BC_n) = \mathbf{P}(A|BC).$$

PROOF. We have by Axiom III

$$\mathbf{P}(A|BC_n) = \frac{\mathbf{P}(AC_n|B)}{\mathbf{P}(C_n|B)} \quad \text{if} \quad \mathbf{P}(C_n|B) > 0.$$

As $\mathbf{P}(C_n|B) \geqq \mathbf{P}(C|B) > 0$, the condition $\mathbf{P}(C_n|B) > 0$ is satisfied by every n, and thus it follows

$$\lim_{n\to\infty} \mathbf{P}(A|BC_n) = \frac{\lim\limits_{n\to\infty} \mathbf{P}(AC_n|B)}{\lim\limits_{n\to\infty} \mathbf{P}(C_n|B)} = \frac{\mathbf{P}(AC|B)}{\mathbf{P}(C|B)} = \mathbf{P}(A|BC)$$

what proves Theorem 13.

1. 10. Products of conditional probability spaces. The product of two conditional probability spaces $(S^{(k)}, \mathfrak{A}^{(k)}, \mathfrak{B}^{(k)}, \mathbf{P}^{(k)}) = \mathfrak{S}^{(k)}$ $(k=1, 2)$ is defined as follows: let $S = S^{(1)} * S^{(2)}$ denote the Cartesian product of the sets $S^{(k)}$ $(k=1, 2)$, i. e. let S denote the set of all ordered pairs $(a^{(1)}, a^{(2)})$ where $a^{(1)} \in S^{(1)}$ and $a^{(2)} \in S^{(2)}$. Let $\mathfrak{B}$ denote the set of all subsets $B_1 = B_1 * B_2$ of S where $B_1 \in \mathfrak{B}^{(1)}$ and $B_2 \in \mathfrak{B}^{(2)}$. Let $\mathfrak{A}$ denote the set of all subsets $A = A_1 * A_2$ where $A_1 \in \mathfrak{A}^{(1)}$ and $A_2 \in \mathfrak{A}^{(2)}$, and let $\overline{\mathfrak{A}}$ denote the least σ-algebra containing $\mathfrak{A}$. Let us define $\mathbf{P}(A|B)$ for $A = A_1 * A_2$ and $B = B_1 * B_2$, where $A_1 \in \mathfrak{A}^{(1)}$, $A_2 \in \mathfrak{A}^{(2)}$, $B_1 \in \mathfrak{B}^{(1)}$ and $B_2 \in \mathfrak{B}^{(2)}$ by

$$\mathbf{P}(A|B) = \mathbf{P}^{(1)}(A^{(1)}|B^{(1)})\,\mathbf{P}^{(2)}(A^{(2)}|B^{(2)}),$$

and let us extend the definition of $\mathbf{P}(A|B)$ for every fixed $B \in \mathfrak{B}$ to all $A \in \overline{\mathfrak{A}}$ in the usual way (see [15], § 13). Thus we obtain a conditional probability space $\mathfrak{S} = [S, \overline{\mathfrak{A}}, \mathfrak{B}, \mathbf{P}]$ which will be called the Cartesian product of the conditional probability spaces $\mathfrak{S}^{(1)}$ and $\mathfrak{S}^{(2)}$ and denoted by $\mathfrak{S} = \mathfrak{S}^{(1)} * \mathfrak{S}^{(2)}$.

The Cartesian product of a finite number of conditional probability spaces is defined similarly. The Cartesian product of a denumerable sequence of conditional probability spaces $P^{(k)} = [S^{(k)}, \mathfrak{A}^{(k)}, \mathfrak{B}^{(k)}, \mathbf{P}^{(k)}]$ $(k=1, 2, \ldots)$ is defined as follows: we denote by $S = S_1 * \cdots * S_n * \cdots$ the Cartesian product of the sets $S^{(k)}$ $(k=1, 2, \ldots)$ by $\mathfrak{B}$ the set of all sets $B = B_1 * \cdots * B_n * \cdots$, where $B_n \in \mathfrak{B}^{(n)}$ $(n=1, 2, \ldots)$ and by $\mathfrak{A}$ the set of all sets A of the form $A = A_1 * A_2 * \cdots A_n * S^{(n+1)} * \cdots$, i. e. $\mathfrak{A}$ is the set of all "cylinders" of S. We define $\mathbf{P}(A|B)$ for $A \in \mathfrak{A}$ and $B \in \mathfrak{B}$, where $A = A_1 \cdots A_n * S^{(n+1)} \cdots$ and $B = B_1 \cdots B_n$, by

$$\mathbf{P}(A|B) = \prod_{k=1}^{n} \mathbf{P}^{(k)}(A_k|B_k),$$

and extend the definition of $\mathbf{P}(A|B)$ in the usual way for any fixed $B \in \mathfrak{B}$ to all sets A belonging to the least σ-algebra $\overline{\mathfrak{A}}$ containing $\mathfrak{A}$. In this way we obtain a conditional probability space $\mathbf{P} = [S, \overline{\mathfrak{A}}, \mathfrak{B}, \mathbf{P}(A|B)]$ which will be called the Cartesian product of the conditional probability spaces $\mathfrak{S}^{(k)}$ and denoted by $\mathfrak{S} = \mathfrak{S}^{(1)} * \cdots * \mathfrak{S}^{(k)} * \cdots$. To prove that $\mathfrak{S}$ is really a conditional probability space, we have only to verify the validity of Axiom III, as Axioms I and II are clearly valid for $\mathfrak{S}$. As regards the validity of $\mathbf{P}(AB|B) = \mathbf{P}(A|B)$ (i. e. of Theorem 1) for $\mathfrak{S}$, it suffices to prove that

$$\mathbf{P}(A|BC)\,\mathbf{P}(BC|C) = \mathbf{P}(ABC|C) \tag{26}$$

for $A \in \mathfrak{A}$, $B \in \overline{\mathfrak{A}}$, $C \in \mathfrak{B}$ and $BC \in \mathfrak{B}$.

As both sides of (26) are measures as functions of A for fixed B and C, it suffices to verify (26) for the case when A is a cylinder, $A = A^{(1)} * \cdots * A^{(n)} * S^{(n+1)} \ldots$.

In this case

$$\mathbf{P}(A|BC) = \prod_{1}^{n} \mathbf{P}^{(k)}(A^{(k)}|(BC)^{(k)}).$$

Putting

$$P_N = \prod_{k=1}^{N} \mathbf{P}^{(k)}((BC)^{(k)}|C^{(k)}),$$

the sequence P_N is monotonically non-increasing, and thus $\lim_{N\to\infty} P_N = p$ exists. Two cases are to be distinguished. If $p = 0$, we have $\mathbf{P}(BC|C) = 0$ and thus $P(ABC|C) = 0$ and therefore (26) is satisfied. If $p > 0$, we have

$$\mathbf{P}(ABC|C) = \prod_{1}^{N} \mathbf{P}^{(k)}(A^{(k)}(BC)^{(k)}|C^{(k)}) \prod_{N+1}^{\infty} \mathbf{P}^{(k)}((BC)^{(k)}|C^{(k)})$$

and

$$\mathbf{P}(BC|C) = \prod_{k=1}^{\infty} \mathbf{P}^{(k)}((BC)^{(k)}|C^{(k)}),$$

and thus

$$\mathbf{P}(A|BC)\,\mathbf{P}(BC|C) = \prod_{k=1}^{N} \mathbf{P}^{(k)}(A^{(k)}|(BC)^{(k)})\,\mathbf{P}^{(k)}((BC)^{(k)}|C^{(k)}) \prod_{N+1}^{\infty} \mathbf{P}^{(k)}((BC)^{(k)}|C^{(k)})$$

and thus, as $\mathbf{P}^{(k)}$ satisfies Axiom III, we have

$$\mathbf{P}^{(k)}(A^{(k)}|(BC)^{(k)})\,\mathbf{P}^{(k)}((BC)^{(k)}|C^{(k)}) = \mathbf{P}^{(k)}(A^{(k)}(BC)^{(k)}|C^{(k)})$$

and thus (26) is valid.

§ 2. Examples of conditional probability spaces

2. 1. Conditional probability spaces of simple quotient type. An important class of conditional probability spaces is obtained as follows:

Let S denote a set, $\mathfrak{A}$ a σ-algebra of subsets of S, $\mu(A)$ a measure on $\mathfrak{A}$, $\mathfrak{B}$ the set of those sets $B \in \mathfrak{A}$ for which $\mu(B)$ is positive and finite and put $\mathbf{P}(A|B) = \dfrac{\mu(AB)}{\mu(B)}$ for $A \in \mathfrak{A}$, $B \in \mathfrak{B}$. Then $[S, \mathfrak{A}, \mathfrak{B}, \mathbf{P}(A|B)]$ is a conditional probability space which will be called a "conditional probability space of simple quotient-type". We mention two special cases.

a) Let us choose for S the n-dimensional Euclidean space E_n, for A the set of all measurable subsets of E_n, let $f(x_1, x_2, \ldots, x_n)$ be a non-negative measurable function on E_n. For $\mathcal{B}$ we choose the set of those subsets $B \in \mathcal{A}$ for which

$$\Phi(B) = \int_B f(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \ldots dx_n$$

is finite and positive and we suppose that $\mathcal{B}$ is not empty.

$$\mathbf{P}(A|B) = \frac{\Phi(AB)}{\Phi(B)}$$

for $A \in \mathcal{A}$ and $B \in \mathcal{B}$. Clearly $[S, \mathcal{A}, \mathcal{B}, \mathbf{P}(A|B)]$ is a conditional probability space of simple quotient-type, provided that B is not empty.

Especially, when $f(x_1, x_2, \ldots, x_n) \equiv 1$, we obtain a conditional probability space, for which $\mathbf{P}(A|B) = \dfrac{m(AB)}{m(B)}$, where $m(C)$ denotes the Lebesgue measure of the measurable set C; in this case $\mathcal{B}$ consists of all measurable sets which have a positive and finite Lebesgue measure. The conditional probability space thus obtained will be called the uniform conditional probability space in E_n.

b) Let S denote a finite or denumerable set, with elements $a_1, a_2, \ldots, a_k, \ldots$; let p_k $(k = 1, 2, \ldots)$ denote an arbitrary sequence of non-negative numbers. Let $\mathcal{A}$ denote the set of all subsets of S and $\mathcal{B}$ the set of those subsets B of S for which $\sum_{a_k \in B} p_k = \Pi(B)$ is positive and finite. Let us suppose that $\mathcal{B}$ is not empty and put

$$\mathbf{P}(A|B) = \frac{\Pi(AB)}{\Pi(B)} \quad \text{for} \quad A \in \mathcal{A} \quad \text{and} \quad B \in \mathcal{B}.$$

Clearly $[S, \mathcal{A}, \mathcal{B}, \mathbf{P}(A|B)]$ is a conditional probability space of simple quotient-type. Especially, if $p_k \equiv 1$ $(k = 1, 2, \ldots)$, we obtain a conditional probability space which will be called uniform on the denumerable set S.

2. 2. Conditional probability spaces of alternative quotient-type. A general type of conditional probability spaces is described by the following

THEOREM 14. *Let S denote a set, $\mathcal{A}_\gamma$ a σ-ring of subsets of S if $\gamma \in \Gamma$ where Γ is an arbitrary ordered set of "indices", and suppose that $\mathcal{A}_\beta \subseteq \mathcal{A}_\gamma$ if $\beta < \gamma$ according to the order relation of Γ. Let us put $\mathcal{A} = \sum_{\gamma \in \Gamma} \mathcal{A}_\gamma$. Let $\mathcal{B}_\gamma$ denote a subset of $\mathcal{A}_\gamma$ and put $\mathcal{B} = \sum_{\gamma \in \Gamma} \mathcal{B}_\gamma$; we suppose that $\mathcal{B}_\gamma$ is not empty for any $\gamma \in \Gamma$ and that $\mathcal{B}_\beta$ and $\mathcal{B}_\gamma$ are disjoint for $\beta \neq \gamma$. Let us suppose that for any $\gamma \in \Gamma$ a measure $\mu_\gamma(A)$ is given for $A \in \mathcal{A}$ (which can take also the value $+\infty$), and that if $B \in \mathcal{B}_\gamma$, we have $0 < \mu_\gamma(B) < +\infty$. Let us sup-*

pose that if $\beta \in \Gamma$, $\gamma \in \Gamma$, $\beta < \gamma$ (according to the order relation defined in Γ) and $\mu_\beta(A) < +\infty$, then $\mu_\gamma(A) = 0$.

Let $\mathbf{P}(A|B)$ be defined for $A \in \mathfrak{A}$ and $B \in \mathfrak{B}$ as follows:

$$\mathbf{P}(A|B) = \frac{\mu_\gamma(AB)}{\mu_\gamma(B)},$$

where γ is uniquely determined by the condition $B \in \mathfrak{B}_\gamma$. Then $[S, \mathfrak{A}, \mathfrak{B}, \mathbf{P}(A|B)]$ is a conditional probability space.

PROOF. Axioms I and II are clearly satisfied. Instead of Axiom III we verify the validity of Axiom III′ which has been shown to be equivalent to Axiom III. Clearly Axiom III′ is satisfied, because if $B \in \mathfrak{B}_\beta$ and $C \in \mathfrak{B}_\gamma$ and $B \subseteq C$, only three cases are possible: $\beta = \gamma$, $\beta < \gamma$ or $\beta > \gamma$. The case $\beta = \gamma$ is evident. If $\beta < \gamma$, $\mathbf{P}(B|C) = \frac{\mu_\gamma(BC)}{\mu_\gamma(C)} = \frac{\mu_\gamma(B)}{\mu_\gamma(C)}$ and as $\mu_\beta(B) < \infty$ and $\beta < \gamma$, we have by supposition $\mu_\gamma(B) = 0$ and therefore $\mathbf{P}(B|C) = 0$; and thus there is nothing to prove.

Finally $\gamma < \beta$ is impossible, because from $C \in \mathfrak{B}_\gamma$ it follows $\mu_\gamma(C) < +\infty$ and thus $\mu_\beta(C) = 0$, what implies $\mu_\beta(B) = 0$ because of $B \subseteq C$; but this contradicts $B \in \mathfrak{B}_\beta$. Thus Theorem 14 is proved.

The conditional probability spaces described by Theorem 14 will be called "conditional probability spaces of alternative quotient-type". If $B \in \mathfrak{B}_\gamma$, we may call γ the "dimension" of B, and say that the measures μ_γ are "dimensionally ordered".

We mention two special cases:

a) A special case of a conditional probability space of alternative quotient-type is obtained as follows: let $f(x_1, x_2, \ldots, x_n)$ denote a non-negative Borel measurable function in $S = E_n$; let $\mathfrak{A}$ denote the set of all measurable subsets of E_n, and $\mathfrak{B}$ the set of those measurable sets B, for which either $\Phi(B) = \int_B f(x_1, \ldots, x_n)\, dx_1\, dx_2 \cdots dx_n$ is finite and positive, or B is contained in a k-dimensional subspace $E_{i_1, i_2, \ldots, i_{n-k}}^{c_1, c_2, \ldots, c_{n-k}}$ of E_n $(k = 1, 2, \ldots, n-1)$ defined by $x_{i_1} = c_1, x_{i_2} = c_2, \ldots, x_{i_{n-k}} = c_{i_{n-k}}$, and the integral

$$\Phi_{i_1, i_2, \ldots, i_{n-k}}^{c_1, c_2, \ldots, c_{n-k}} = \int_B f(x_1, x_2, \ldots, x_n)\, dx_{j_1}\, dx_{j_2} \cdots dx_{j_k} \tag{27}$$

is finite and positive, where $j_1, j_2, \ldots, j_k$ denote those of the indices $1, 2, \ldots, n$ which are different from $i_1, i_2, \ldots, i_{n-k}$ and in (26) $x_{i_1} = c_1, x_{i_2} = c_2, \ldots, x_{i_{n-k}} = c_{n-k}$ are substituted; we suppose that $\mathfrak{B}$ is not empty; in the first case we put

$$\mathbf{P}(A|B) = \frac{\Phi(AB)}{\Phi(B)};$$

if $B \in \mathfrak{B}$ and B is contained in $E^{c_1, c_2, \ldots, c_{n-k}}_{i_1, i_2, \ldots, i_{n-k}}$, we put

$$\mathbf{P}(A|B) = \frac{\Phi^{c_1, \ldots, c_{n-k}}_{i_1, \ldots, i_{n-k}}(AB)}{\Phi^{c_1, \ldots, c_{n-k}}_{i_1, \ldots, i_{n-k}}(B)}$$

where $\Phi^{c_1, \ldots, c_{n-k}}_{i_1, \ldots, i_{n-k}}$ is defined by (27).

b) Another special case of a conditional probability space of alternative quotient-type is the following: let $S = E_n$ be the n-dimensional Euclidean space, $\mathfrak{A}$ the set of all Borel measurable subsets of E_n, $\mathfrak{B}_k$ the set of those Borel measurable subsets of E_n, which have finite positive k-dimensional measure[7] $(1 \leqq k \leqq n)$. Let us put $\mathfrak{B} = \sum_{k=1}^{n} \mathfrak{B}_k$ and $\mathbf{P}(A|B) = \frac{m_k(AB)}{m_k(B)}$, if $B \in \mathfrak{B}_k$, where $m_k(B)$ denotes the k-dimensional Lebesgue measure of the set B. Clearly this conditional probability space is an extension of the uniform conditional probability space mentioned in 2. 1, a).

2. 3. Cavalieri spaces. Now we introduce a general concept, that of a "Cavalieri space". Let $[S, \mathfrak{A}, \mathfrak{B}, \mathbf{P}(A|B)]$ be a conditional probability space and let us make the following assumptions:

C 1. To every real number t $(\alpha \leqq t < \beta)$ there corresponds a set $C_t \in \mathfrak{B}$.

C 2. For any numbers a and b for which $\alpha \leqq a < b \leqq \beta$

$$C_a^b = \sum_{a \leqq t < b} C_t \in \mathfrak{B}.$$

C 3. If $\alpha \leqq s < t < \beta$, we have $C_s \cdot C_t = O$.

C 4. If for $A \in \mathfrak{A}$, $A' \in \mathfrak{A}$, we have $\mathbf{P}(A|C_t) \leqq \lambda \mathbf{P}(A'|C_t)$, where $\lambda \geqq 0$, for every t lying in the interval $[a, b)$ $(\alpha \leqq a < b \leqq \beta)$, it follows $\mathbf{P}(A|C_a^b) \leqq \lambda \mathbf{P}(A'|C_a^b)$.

If C 1—C 4 hold, we shall say that $[S, \mathfrak{A}, \mathfrak{B}, \mathbf{P}(A|B)]$ is a Cavalieri space with respect to the family of sets $\{C_t; \alpha \leqq t < \beta\}$.

Let us mention that if instead of a family C_t of the power of the continuum we consider a denumerable set $\{C_n\}$ satisfying properties C 1, C 2, and C 3, then (as it has been mentioned in the remark to Theorem 7) C 4 is always satisfied. On the other hand, Theorem 14 shows that conditional probability spaces in general are not Cavalieri spaces with respect to a family $\{C_t\}$ of the power of the continuum, satisfying C 1, C 2 and C 3, because if $\mathbf{P}(C_t|C_a^b) = 0$ for $a \leqq t < b$, $\mathbf{P}(A|C_t)$ can be replaced by any other measure for every $t \in [a, b)$, by leaving $\mathbf{P}(A|C_a^b)$ unchanged.

The following simple consequences of the definition of a Cavalieri space may be mentioned: if $[S, \mathfrak{A}, \mathfrak{B}, \mathbf{P}(A|B)]$ is a Cavalieri space with respect to the family of sets $\{C_t; \alpha \leqq t < \beta\}$, then

[7] See [26], p. 108.

I.[8] If $\mathbf{P}(A|C_t)=\lambda\mathbf{P}(A'|C_t)$ for $a \leqq t < b$, we have $\mathbf{P}(A|C_a^b)=\lambda\mathbf{P}(A'|C_a^b)$.

II. If $\mathbf{P}(A|C_t)=\lambda$ for $a \leqq t < b$, we have $\mathbf{P}(A|C_a^b)=\lambda$.

As a matter of fact, if $\mathbf{P}(A|C_t)=\lambda\mathbf{P}(A'|C_t)$ for $a \leqq t < b$, we have $\mathbf{P}(A|C_t) \leqq \lambda\mathbf{P}(A'|C_t)$ and thus by C 4 $\mathbf{P}(A|C_a^b) \leqq \lambda\mathbf{P}(A'|C_a^b)$, further (if $\lambda > 0$) we have $\mathbf{P}(A'|C_t) \leqq \frac{1}{\lambda}\mathbf{P}(A|C_t)$ and thus $\mathbf{P}(A'|C_a^b) \leqq \frac{1}{\lambda}\mathbf{P}(A|C_a^b)$, i. e. $\lambda\mathbf{P}(A'|C_a^b) \leqq \mathbf{P}(A|C_a^b) \leqq \lambda\mathbf{P}(A'|C_a^b)$ (also for $\lambda=0$) and therefore $\mathbf{P}(A|C_a^b)= =\lambda\mathbf{P}(A'|C_a^b)$.

On the other hand, II follows from I by choosing $A'=C_a^b$. The notion of a Cavalieri space can be extended by replacing the one-parametric family $\{C_t\}$ by a k-parametric family $\{C_{t_1,t_2,\ldots,t_k};\ \alpha_i \leqq t_i < \beta_i;\ i=1,2,\ldots,k\}$.

2.4. Regular spaces. Let $\mathcal{S}=[S, \mathfrak{A}, \mathfrak{B}, \mathbf{P}(A|B)]$ denote a conditional probability space with the following properties:

R 1. To every real number t ($\alpha \leqq t \leqq \beta$) there corresponds a set $R_t \in \mathfrak{B}$.

R 2. For any pair of numbers a, b ($\alpha \leqq a < b \leqq \beta$) $R_a^b=\sum_{a \leqq t < b} R_t \in \mathfrak{B}$.

R 3. If $\alpha \leqq s < t < \beta$, we have $R_s \cdot R_t = \bigcirc$.

R 4. If $A \in \mathfrak{A}$, $\mathbf{P}(A|R_t)$ is a Borel measurable function of t for $\alpha \leqq t < \beta$.

R 5. There exists a strictly increasing and continuous function $F(t)$ in (α, β) with the property that if $\alpha \leqq a < b \leqq \beta$ we have for any $A \in \mathfrak{A}$

$$\mathbf{P}(A|R_a^b)=\frac{\int_a^b \mathbf{P}(A|R_t)\,dF(t)}{\int_a^b dF(t)}. \tag{28}$$

In this case we shall say that $\mathcal{S}$ is *regular* with respect to the family $\{R_t; \alpha \leqq t < \beta\}$. If $F(t)$ is absolutely continuous, we shall say that $\mathcal{S}$ is *absolutely regular*.

Clearly if $\mathcal{S}$ is regular with respect to the family $\{R_t\}$, it is also a Cavalieri space with respect to $\{R_t\}$.[9] As a matter of fact, C 1, C 2 and C 3 are identical with R 1, R 2 and R 3, and C 4 follows from R 4 and R 5.

It is easy to prove by using FUBINI's theorem, that if $0 < f(x_1, x_2, \ldots, x_n) < M$, the conditional probability space defined in 2. 2, a) is regular with respect to

[8] This consequence of the definition of a Cavalieri space (for $\lambda=1$) is analogous with the well-known "principle of Cavalieri" (according to which if the areas of the intersections of two solids with every horizontal plane are equal, the two solids have the same volume) introduced by BONAVENTURA CAVALIERI (1598—1647) in 1635. This justifies the name "Cavalieri space".

[9] J. CZIPSZER has shown that there exist Cavalieri spaces with respect to a family $\{R_t\}$ of sets which are not regular with respect to the same family $\{R_t\}$.

the family $\{R_t\}$ defined as follows: R_t is the set of points $(x_1, x_2, \ldots, x_n)$ for which $a_i \leqq x_i < b_i$ for $i \neq k$, $i = 1, 2, \ldots, n$, and $x_k = t$ $(\alpha \leqq t < \beta)$.

It is clear that if $\mathcal{S}$ is regular with respect to the family $\{R_t; \alpha \leqq t < \beta\}$, we have

$$\lim_{h>0,\, h\to 0} \mathbf{P}(A \mid R_t^{t+h}) = \mathbf{P}(A \mid R_t) \tag{29}$$

for almost every t in (α, β) with respect to the measure generated by $F(t)$ on (α, β) for any fixed A.

It is also evident that if $\mathcal{S}$ is regular with respect to the family $\{R_t; \alpha \leqq t < \beta\}$, we have $\mathbf{P}(R_t \mid R_a^b) = 0$ for $\alpha \leqq a \leqq t < b \leqq \beta$.

Let $\xi = \xi(a)$ $(a \in S)$ be a random variable on a conditional probability space $\mathcal{S} = [S, \mathcal{A}, \mathcal{B}, \mathbf{P}(A \mid B)]$, further let us denote by R_t the set of those $a \in S$ for which $\xi(a) = t$ $(-\infty \leqq t < +\infty)$. Let us suppose that $\mathcal{S}$ is regular with respect to the family $\{R_t\}$, i. e. we have for $A \in \mathcal{A}$

$$\mathbf{P}(A \mid R_a^b) = \frac{\int_a^b \mathbf{P}(A \mid R_t)\, dF(t)}{\int_a^b dF(t)} \tag{30}$$

for $-\infty < a < b < +\infty$. Evidently the conditional distribution generated by ξ on the real axis is the distribution

$$P(a \leqq \xi < b \mid c \leqq \xi < d) = \mathbf{P}(R_a^b \mid R_c^d) = \frac{F(b) - F(a)}{F(d) - F(c)} \quad \text{for} \quad c \leqq a < b \leqq d,$$

i. e. $F(x)$ is the (generalized) distribution function of ξ. If (30) holds, we shall say that ξ has a regular conditional distribution with the distribution function $F(x)$; if $F(x) = x$, we shall say that ξ has a regular uniform distribution.

The notion of a regular space can be extended by replacing the one-parametric family $\{R_t\}$ by a k-parametric family $\{R_{t_1, t_2, \ldots, t_k}\}$. We consider only the case when the following postulates are satisfied:

R 1. To every point $T = (t_1, t_2, \ldots, t_k)$ of a k-dimensional interval $I = (\alpha_i \leqq t_i < \beta_i;\ i = 1, 2, \ldots, k)$ there corresponds a set $R_t \in \mathcal{B}$.

R 2. For any subinterval $J = (a_i \leqq t_i < b_i;\quad i = 1, 2, \ldots, k)$ where $\alpha_i \leqq a_i < b_i \leqq \beta_i$ $(i = 1, 2, \ldots, k)$ we have $R_J = \sum_{T \in J} R_T \in \mathcal{B}$.

R 3. If $T_1 \neq T_2$, we have $R_{T_1} \cdot R_{T_2} = \bigcirc$.

R 4. If $A \in \mathcal{A}$, $\mathbf{P}(A \mid R_T)$ is a Borel-measurable function of the variables $t_1, t_2, \ldots, t_k$.

R 5. There exists a function $f(t_1, t_2, \ldots, t_k)$ which is positive and integrable on I, and if J is the interval $a_i \leqq t_i < b_i$ $(i = 1, 2, \ldots, k)$,

$(\alpha_i \leqq a_i < b_i \leqq \beta_i;\ i = 1, 2, \ldots, k)$, we have

$$\mathbf{P}(A \mid R_J) = \frac{\int\limits_J \mathbf{P}(A \mid R_T) f(t_1, t_2, \ldots, t_k)\, dt_1 \ldots dt_k}{\int\limits_J f(t_1, \ldots, t_k)\, dt_1 \ldots dt_k}.$$

If R 1—R 5 hold, and $f(t_1, \ldots, t_k) = \prod_{j=1}^{k} f_j(t_j)$, we say that *the space is absolutely regular with respect to the family* $\{R_T\}$ *with independent parameters.*

Especially if R_T is the set defined by $\xi_1 = t_1, \ldots, \xi_k = t_k$ where $\xi_1, \xi_2, \ldots, \xi_k$ are random variables defined on $[S, \mathcal{A}, \mathcal{B}, \mathbf{P}(A \mid B)]$ and $f(t_1, \ldots, t_n) = \prod_{j=1}^{k} f_j(t_j)$, the random variables $\xi_1, \ldots, \xi_k$ are independent with respect to any R_J, further ξ_j has a regular conditional distribution with the generalized density function $f_j(t)$ $(j = 1, 2, \ldots, k)$.

2. 5. Remark on Borel's paradox. The notions of Cavalieri spaces resp. regular spaces introduced in the preceding sections, are connected with the well-known paradox of BOREL. (Cf. [1], p. 44.) Let S denote the surface of the unit sphere $x^2 + y^2 + z^2 = 1$; let us introduce spherical coordinates φ and ϑ, defined by $x = \cos\varphi \sin\vartheta$, $y = \sin\varphi \sin\vartheta$, $z = \cos\vartheta$ $(0 \leqq \varphi < 2\pi;\ 0 \leqq \vartheta < \pi)$. Let us denote by $\mathcal{A}$ the set of all Borel subsets of S (i. e. sets A for which the set of all pairs of numbers (ϑ, φ) for which the corresponding point belongs to A, is a plane Borel set). Let us denote by $\mathcal{B}$ the set of all sets $B_{a,c}^{b,d}$ defined by the inequalities $a \leqq \varphi \leqq b$, $c \leqq \vartheta \leqq d$, where $0 \leqq a \leqq b \leqq 2\pi$ and $0 \leqq c < d \leqq \pi$. Let us put

$$\mathbf{P}(A \mid B_{a,c}^{b,d}) = \frac{\int\int\limits_{A B_{a,c}^{b,d}} \sin\vartheta\, d\vartheta\, d\varphi}{\int\limits_a^b \int\limits_c^d \sin\vartheta\, d\vartheta\, d\varphi} \quad \text{if} \quad a < b,$$

i. e. let $\mathbf{P}(A \mid B)$ for fixed B be proportional to the area of AB. Let us denote by B_t the set $\varphi = t$, $0 \leqq \vartheta \leqq \pi$ (i. e. put $B_t = B_{t,0}^{t,\pi}$); let us choose a bounded measure $\mu_t(B_t) = 1$ defined on the Borel subsets of B_t and normed by $\mu_t(B_t) = 1$, and put

$$\mathbf{P}(A \mid B_{t,c}^{t,d}) = \frac{\mu_t(A B_{t,c}^{t,d})}{\mu_t(B_{t,c}^{t,d})} \qquad (0 \leqq c < d \leqq \pi),$$

and let us suppose that $\mathbf{P}(B_{0,0}^{2\pi,x} \mid B_t) = F(x, t)$ is a continuous function of x and t, strictly increasing as a function of x. Clearly $[S, \mathcal{A}, \mathcal{B}, \mathbf{P}(A \mid B)] = \mathfrak{S}$ is a conditional probability space by any choice of the measures $\mu_t(A)$; we shall prove that it is a Cavalieri space with respect to the family $\{B_t; 0 \leqq t < 2\pi\}$ if and only if

$$\mu_t(A) = \frac{1}{2} \int_{AB_t} \sin \vartheta \, d\vartheta. \tag{31}$$

This can be shown as follows: let us suppose that $\mathcal{S}$ is a Cavalieri space with respect to $\{B_t; 0 \leqq t < 2\pi\}$; let us choose a number λ $(0 \leqq \lambda < 1)$ and define $x(t, \lambda)$ as the number for which, if $A_t(\lambda)$ denotes the interval $0 \leqq \vartheta \leqq$ $\leqq x(t, \lambda)$, we have $\mathbf{P}(A_t(\lambda) | B_t) = \lambda$, i. e. $F(x(t, \lambda), t) = \lambda$. Now it follows from our supposition, that $x(t, \lambda)$ is a continuous function of t and thus if $A(\lambda) = \sum_{0 \leqq t < 2\pi} A_t(\lambda)$, we have $A(\lambda) \in \mathcal{A}$, further $\mathbf{P}(A(\lambda) | B_t) = \lambda$ for $0 \leqq t < 2\pi$. It follows by II of 2. 3 that putting $B_{a, 0}^{b, \pi} = B_a^b$, we have $\mathbf{P}(A(\lambda) | B_a^b) = \lambda$ for $0 \leqq a < b < 2\pi$, i. e. $\frac{1}{2} \int_a^b \int_0^{x(t, \lambda)} \sin \vartheta \, d\vartheta \, dt = \lambda(b - a)$. By taking the derivative of both sides, with respect to b, by virtue of the continuity of $x(t, \lambda)$, we obtain that $\frac{1}{2} \int_0^{x(t, \lambda)} \sin \vartheta \, d\vartheta = \lambda$ for every t $(0 \leqq t < 2\pi)$. Thus, it follows easily that (31) holds.

Our result can be stated as follows: the uniform probability distribution on the surface of the sphere combined with arbitrary distributions on the meridians $\varphi = t$ gives always a conditional probability space, but this space is a Cavalieri space with respect to the set of meridians (under suitable continuity restrictions) if and only if the density of the distribution is on every meridian the same, and equal to $\frac{1}{2} \sin \vartheta$. In this case the conditional probability space is clearly also regular with respect to the set of meridians, as we have

$$\mathbf{P}(A | B_a^b) = \frac{1}{b - a} \int_a^b \left(\int_{AB_t} \frac{1}{2} \sin \vartheta \, d\vartheta \right) dt = \frac{\int_a^b \mathbf{P}(A | B_t) \, dt}{\int_a^b dt}.$$

2.6. Composition of conditional probability distributions. Let us suppose that $\xi_j = \xi_j(a)$ $(j = 1, 2, \ldots, k;\ a \in S)$ are positive random variables on the conditional probability space $\mathcal{S} = [S, \mathcal{A}, \mathcal{B}, \mathbf{P}(A | B)]$. Let us denote by $R_{t_1, t_2, \ldots, t_k}$ the set of those $a \in S$ for which $\xi_1(a) = t_1, \ldots, \xi_k(a) = t_k$ and suppose that the space $\mathcal{S}$ is absolutely regular with respect to the family $\{R_{t_1, \ldots, t_k};$ $0 \leqq t_i < +\infty;\ i = 1, 2, \ldots, k\}$, with independent parameters and thus the functions $F_j(T_j)$ figuring in R 5 are absolutely continuous and $F_j'(t) = f_j(t)$

is almost everywhere positive ($0 < t < +\infty$). Let us suppose further that if M is a Borel subset of the k-dimensional Euclidean space, $\sum_{(t_1,\ldots,t_k)\in M} R_{t_1,\ldots,t_k}$ belongs to $\mathfrak{B}$. In this case if I and $J \subseteq I$ are k-dimensional intervals,

$$I = (\alpha_i \leqq t_i < \beta_i;\; i = 1, 2, \ldots, k) \quad \text{and} \quad J = (a_i \leqq t_i < b_i;\; i = 1, 2, \ldots, k)$$

where $0 \leqq \alpha_i \leqq a_i < b_i \leqq \beta_i$ ($i = 1, 2, \ldots, k$), we have, putting $\vec{\xi} = (\xi_1, \ldots, \xi_k)$,

$$\mathbf{P}(\vec{\xi} \in J \mid \vec{\xi} \in I) = \prod_{j=1}^{k} \frac{\int_{a_j}^{b_j} f_j(t_j)\,dt}{\int_{\alpha_j}^{\beta_j} f_j(t)\,dt}. \tag{32}$$

Thus the random variables $\xi_1, \xi_2, \ldots, \xi_k$ are independent with respect to any R_I.

Let us calculate the conditional probability distribution of $\zeta_k = \xi_1 + \xi_2 + \cdots + \xi_k$. Let B_T be the interval $0 \leqq t_j < T$ ($j = 1, 2, \ldots, k$) and A_z the set defined by $\zeta_k < z$. Clearly, $A_z \in \mathfrak{B}$, as $A_z = \sum_{t_1+t_2+\cdots+t_k<z} B_{t_1,\ldots,t_k}$; we have by R 5

$$\mathbf{P}(A_z \mid B_T) = \frac{\int\limits_{\sum_{j=1}^{k} t_j < z,\; 0 \leqq t_j < T} \prod_{j=1}^{k} f_j(t_j)\,dt_j}{\prod_{j=1}^{k} \left(\int_0^T f_j(t)\,dt \right)}.$$

Now let us suppose that $0 \leqq z_1 < z_2 < T$; in this case $A_{z_1} \subseteq A_{z_2} \subseteq B_T$ and thus

$$\mathbf{P}(A_{z_i} \mid B_T) = \frac{\int\limits_{\sum_{j=1}^{k} t_j < z_i} \prod_{j=1}^{k} f_j(t_j)\,dt_j}{\prod_{j=1}^{k} \left(\int_0^T f_j(t)\,dt \right)}$$

and

$$\mathbf{P}(A_{z_1} \mid A_{z_2}) = \mathbf{P}(A_{z_1} \mid A_{z_2} B_T) = \frac{\mathbf{P}(A_{z_1} A_{z_2} \mid B_T)}{\mathbf{P}(A_{z_2} \mid B_T)} = \frac{\mathbf{P}(A_{z_1} \mid B_T)}{\mathbf{P}(A_{z_2} \mid B_T)}$$

and thus

$$\mathbf{P}(A_{z_1} \mid A_{z_2}) = \frac{\int\limits_{\sum_{j=1}^{k} t_j < z_1} \prod_{j=1}^{k} f_j(t_j)\,dt_j}{\int\limits_{\sum_{j=1}^{k} t_j < z_2} \prod_{j=1}^{k} f_j(t_j)\,dt_j}. \tag{33}$$

It follows that if we put

$$g_1(t) = f_1(t) \quad \text{and} \quad g_j(t) = \int_0^t g_{j-1}(t-u) f_j(u)\,du \quad \text{if} \quad t \geqq 0$$

and

$$g_j(t) = 0 \quad \text{if} \quad t < 0 \qquad (j = 2, 3, \ldots, k),$$

we have by (33)

$$\mathbf{P}(A_{z_1} \mid A_{z_2}) = \frac{\int_0^{z_1} g_k(t)\,dt}{\int_0^{z_2} g_k(t)\,dt}. \tag{34}$$

Thus $g_k(t)$ is the generalized density function of ζ_k.

If $\int_0^\infty f_j(t)\,dt < +\infty$ $(j = 1, 2, \ldots, k)$, we obtain as a special case the well-known law of composition of independent probability distributions, but clearly some or all of the integrals $\int_0^\infty f_j(t)\,dt$ may be divergent.

The method of Laplace transforms can be applied also in the case when $\int_0^\infty f_j(t)\,dt = +\infty$, but $\int_0^\infty e^{-st} f_j(t)\,dt = \varphi_j(s)$ exists for some $s > 0$ $(j = 1, 2, \ldots, k)$ and in this case we have, putting $\psi_j(s) = \int_0^\infty e^{-st} g_j(t)\,dt$, evidently $\psi_k(s) = \prod_{j=1}^{k} \varphi_j(s)$.

If e. g. ξ and η are independent, regularly distributed random variables which have the (generalized) density functions $x^{\alpha-1}$ resp. $x^{\beta-1}$ in $(0, +\infty)$, where $\alpha > 0$ and $\beta > 0$, the (generalized) density function of the random variable $\xi + \eta$ is $x^{\alpha+\beta-1}$. A striking property of the random variables with density functions $x^{\alpha-1}$ $(0 < x < +\infty)$ is the following: if ξ has the density function $x^{\alpha-1}$ in $(0, +\infty)$ and $C > 0$ is a constant, the random variable $C\xi$ has the same density function.

The problem of composition of distributions in the general case (i. e. if it is not supposed that the random variables considered are positive) is more intricate.[10] If ξ and η are independent random variables with regular joint distribution and density functions $f(x)$, $g(y)$, respectively, and the limit

$$\lim_{A \to \infty} \frac{\int_{-A}^{+A} f(x-y) g(y)\,dy}{\int_{-A}^{+A} f(x_0 - y) g(y)\,dy} = h(x)$$

[10] The theory of "distributions" of L. Schwartz has to be applied.

exists uniformly in x and is positive on a set of positive measure, then $h(x)$ is the density function of $\xi+\eta$. This implies for example that if the distribution of η is uniform on the real axis, the distribution of $\xi+\eta$ is also uniform, whatever the density function of ξ should be.

§ 3. Some applications of the notion of a conditional probability space

In this § we consider simple applications of the notions introduced in § 1 and § 2 to some problems of ordinary probability theory. The random variables considered in this § are thus random variables defined on an ordinary probability space $[S, \mathfrak{A}, \mathbf{P}]$, except when it is explicitly stated that they are defined on a conditional probability space.

3. 1. The limit of the Poisson distribution for $\lambda \to +\infty$. Let ξ_λ denote a random variable which is distributed according to Poisson's law with mean value $\lambda > 0$. The random variable ξ_λ generates the probability distribution on the set $\mathfrak{J}_+$ of non-negative integers

$$p_k(\lambda) = \mathbf{P}(\xi_\lambda = k) = \frac{\lambda^k e^{-\lambda}}{k!} \qquad (k = 0, 1, \ldots).$$

It follows easily from Stirling's formula that

$$\lim_{\lambda\to\infty} \frac{p_{[\lambda]+j}(\lambda)}{p_{[\lambda]+k}(\lambda)} = 1 \tag{35}$$

for $j, k = 0, \pm 1, \pm 2, \ldots$; thus if $\mathfrak{J}$ is the set of all integers, A and B are finite subsets of $\mathfrak{J}$ and B is not empty, then we have

$$\lim_{\lambda\to\infty} \mathbf{P}(\xi_\lambda - [\lambda] \in A \mid \xi_\lambda - [\lambda] \in B) = \frac{n(AB)}{n(B)},$$

where $n(B)$ denotes the number of elements of B. The result can be stated as follows: the conditional probability distribution generated by $\xi_\lambda - [\lambda]$ on $\mathfrak{J}$ tends to the uniform conditional probability distribution on $\mathfrak{J}$ if $\lambda \to +\infty$.

3. 2. The number of prime divisors. An interesting application of the result in 3. 1 to number theory is as follows: let $U(n)$ denote the number of different prime divisors of the integer n; let $\pi_k(N)$ denote the number of those natural numbers $n \leqq N$ for which $U(n) = k$. By a theorem of P. ERDŐS [16] we have

$$\frac{\pi_k(N)}{N} = \frac{(\log\log N)^k}{k!} e^{-\log\log N} \cdot (1 + o(1)) \tag{36}$$

for $|k - \log\log N| < c\sqrt{\log\log N}$, where $c > 0$ is constant, and $o(1)$ tends uniformly to 0 for $N \to \infty$, if c is fixed. It follows that if $D(1) = 0, D(2) = 1$ and $D(n) = U(n) - [\log\log n]$ for $n = 3, 4, \ldots$ and $P_k(N)$ denotes the number of positive integers $n \leqq N$ for which $D(n)=k$, we have by (35) and (36)

$$\lim_{N\to\infty} \frac{P_j(N)}{P_k(N)} = 1$$

for $j, k = 0, \pm 1, \pm 2, \ldots$; thus the conditional probability distribution of the number-theoretical function $D(n)$ $(1 \leqq n \leqq N)$ tends for $N \to \infty$ to the uniform conditional probability distribution on $\mathfrak{J}$.

3.3. The normal distribution for $\sigma \to +\infty$. Let ξ_σ be a random variable, normally distributed with mean value 0 and variance σ^2. Then we have for $c \leqq a < b \leqq d$

$$\mathbf{P}(a \leqq \xi_\sigma < b \mid c \leqq \xi_\sigma < d) = \frac{\int_{\frac{a}{\sigma}}^{\frac{b}{\sigma}} e^{-\frac{x^2}{2}} dx}{\int_{\frac{c}{\sigma}}^{\frac{d}{\sigma}} e^{-\frac{x^2}{2}} dx}$$

and thus

$$\lim_{\sigma\to+\infty} \mathbf{P}(a \leqq \xi_\sigma < b \mid c \leqq \xi_\sigma < d) = \frac{b-a}{d-c},$$

i. e. the conditional probability distribution, generated by ξ_σ on the real axis $\mathfrak{R}$ tends for $\sigma \to +\infty$ to the uniform conditional probability distribution on $\mathfrak{R}$.

3.4. Limiting distribution of the sum of independent random variables. Let $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ denote independent random variables; let us suppose that the variables ξ_n have the same distribution with the density function $f(x)$ for which $\int_{-\infty}^{+\infty} f^2(x)dx$ exists.[11] Let us suppose that $\mathbf{M}(\xi_n)=0$ and

$$\mathbf{M}(\xi_n^2) = \int_{-\infty}^{+\infty} x^2 f(x)dx = D^2 < +\infty.$$

Let us put $\zeta_n = \xi_1 + \xi_2 + \cdots + \xi_n$. Putting

$$\varphi(t) = \int_{-\infty}^{+\infty} e^{ixt} f(x)dx \qquad (-\infty < t < +\infty),$$

we have $\int_{-\infty}^{+\infty} |\varphi(t)|^n dt < +\infty$ for $n \geqq 2$ (see [17]), and thus denoting by $f_n(x)$ the density function of ζ_n, we have

$$f_n(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} (\varphi(t))^n e^{-itx} dt. \tag{37}$$

[11] These conditions can be replaced by weaker ones.

As by supposition $|\varphi(t)| \neq 1$ for $t \neq 0$ and $\lim_{t \to \pm\infty} |\varphi(t)| = 0$, it follows easily by the method of LAPLACE (see [17]) that

(38) $$f_n(x) = \frac{1}{D\sqrt{2\pi n}}(1 + o(1)),$$

where $o(1) \to 0$ if $n \to \infty$, uniformly for $|x| \leqq C$, where C does not depend on n; thus for $c \leqq a < b \leqq d$ we have

(39) $$\lim_{n \to \infty} \mathbf{P}(a \leqq \zeta_n < b \,|\, c \leqq \zeta_n < d) = \frac{b-a}{d-c},$$

i. e. the conditional probability distribution generated by ζ_n on $\mathfrak{R}$ tends to the uniform conditional probability distribution on $\mathfrak{R}$ for $n \to \infty$.

3. 5. A connection between the uniform conditional probability distribution on $\mathfrak{R}_+$ and on $\mathfrak{I}_+$. Let $[S, \mathfrak{A}, \mathfrak{B}, \mathbf{P}(A|B)]$ denote a conditional probability field, $\xi = \xi(a)$ $(a \in S)$ a random variable, which takes only non-negative integer values; let us denote by A_k the set of those $a \in S$ for which $\xi(a) = k$, put $B_n = \sum_{k=0}^{n} A_k$, and let us suppose that $A_k \in \mathfrak{B}$ and $B_n \in \mathfrak{B}$ $(k, n = 0, 1, 2, \ldots)$. Let us suppose that $\lambda = \lambda(a)$ is an other random variable which generates a regular uniform conditional probability distribution on the positive real axis $\mathfrak{R}_+$; let us denote by C_t the set of those $a \in S$ for which $\lambda(a) = t$ and put $D_x = \sum_{0 \leqq y < x} C_y$; let us suppose that $C_x \in \mathfrak{B}$, $D_x \in \mathfrak{B}$, further that $B_n D_x \in \mathfrak{B}$ $(0 < x < +\infty,\ n = 0, 1, 2, \ldots)$. We suppose

(40) $$\mathbf{P}(A_k | C_t) = \frac{t^k e^{-t}}{k!} \qquad (k = 0, 1, \ldots),$$

i. e. that ξ has a Poisson distribution with mean value t under the hypothesis that $\lambda = t$. We shall prove that in this case the conditional probability distribution generated by ξ on the set of non-negative integers $\mathfrak{I}_+$ is the uniform distribution, i. e.

$$\mathbf{P}(A_k | B_n) = \mathbf{P}(\xi = k \,|\, \xi \leqq n) = \frac{1}{n+1} \quad \text{for} \quad k = 0, 1, \ldots, n;\ n = 0, 1, \ldots.$$

This can be shown as follows: We have by Theorem 12

(41) $$\mathbf{P}(A_k | B_n) = \lim_{x \to \infty} \mathbf{P}(A_k | B_n D_x) = \lim_{x \to \infty} \frac{\mathbf{P}(A_k | D_x)}{\mathbf{P}(B_n | D_x)}.$$

As we have supposed that the distribution of λ is not only uniform on $\mathfrak{R}_+$ but also regular, we have

(42) $$\mathbf{P}(A | D_x) = \frac{1}{x} \int_0^x \mathbf{P}(A | C_t)\, dt$$

and thus by (41) and (42)

$$\mathbf{P}(A_k|B_n) = \frac{\int\limits_0^\infty \mathbf{P}(A_k|C_t)\,dt}{\sum\limits_{k=0}^{n}\int\limits_0^\infty \mathbf{P}(A_k|C_t)\,dt}. \tag{43}$$

As by (40)

$$\int\limits_0^\infty \mathbf{P}(A_k|C_t)\,dt = \int\limits_0^\infty \frac{t^k e^{-t}}{k!}\,dt = 1,$$

it follows from (43) that

$$\mathbf{P}(A_k|B_n) = \frac{1}{n+1}$$

which was to be proved.

A realisation of what has been said can be given as follows: let S denote the strip $0 \leqq x < +\infty$, $0 \leqq y < 1$ of the (x, y) plane, $\mathfrak{A}$ the set of all measurable subsets of S, $\mathfrak{B}$ the set of all Borel-measurable subsets of S with

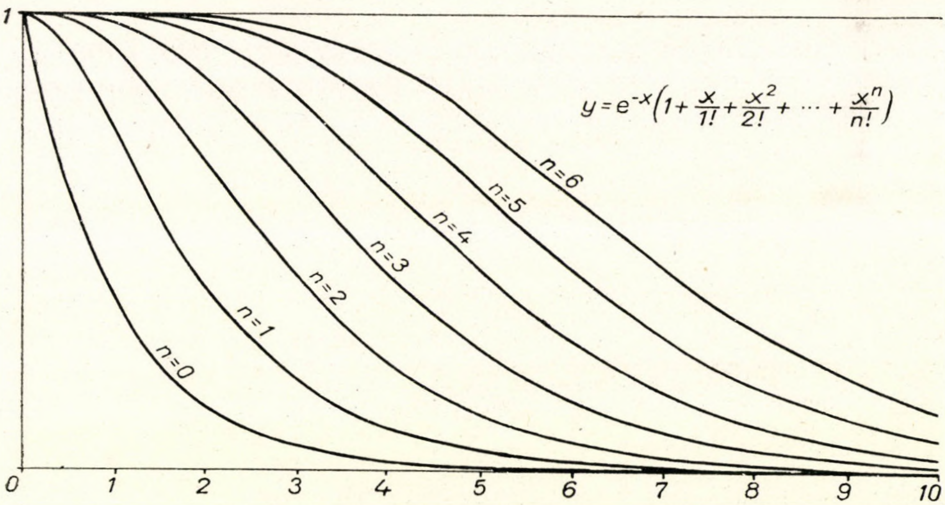

Fig. 1

positive and finite measure, and of all sets of positive linear measure lying on some line $x = t$ $(0 \leqq t < +\infty)$, and let us put $\mathbf{P}(A|B) = \dfrac{m_2(AB)}{m_2(B)}$, if $m_2(B) > 0$, where $m_2(B)$ and $m_2(AB)$ denote the plane Lebesgue measure of B and AB, respectively, and put $\mathbf{P}(A|B) = \dfrac{m_1(AB)}{m_1(B)}$, if B is lying on a line $x = t$, where

$m_1(B)$ denotes the linear Lebesgue measure of B. Define λ as follows: $\lambda = \lambda(x, y) = x$, and define $\xi = \xi(x, y)$ as follows: for a fixed value of x let us have $\xi = 0$ for $0 \leqq y < e^{-x}$, and $\xi = k$ for $\sum_{j=0}^{k-1} \frac{x^j e^{-x}}{j!} \leqq y < \sum_{j=0}^{k} \frac{x^j e^{-x}}{j!}$ $(k = 1, 2, \ldots)$.

The sets A_k are shown on Fig. 1.

Clearly each set A_k $(k = 0, 1, \ldots)$ has the area 1. Our result can be stated in the following suggestive form: if ξ has the Poisson distribution with mean value t under the hypothesis $\lambda = t$ and the conditional distribution of λ is regular and uniform on $\mathfrak{R}_+$, ξ is uniformly distributed on $\mathfrak{I}_+$. Clearly the result obtained is a consequence of the following remarkable property of the Poisson distribution: each member $\frac{\lambda^k e^{-\lambda}}{k!}$ of the Poisson distribution is as a function of λ for fixed k a probability density function in the interval $(0 \leqq \lambda < +\infty)$.

3.6. Remarks on the method of Bayes. Let $\mathfrak{S} = [S, \mathfrak{A}, \mathfrak{B}, \mathbf{P}(A \mid B)]$ denote a conditional probability space. Let $\xi = \xi(a)$ and $\eta = \eta(a)$ $(a \in S)$ be random variables on $\mathfrak{S}$ and suppose that $\mathfrak{S}$ is regular with respect to the family of sets R_{xy} defined by $\xi = x, \eta = y$ for $-\infty < x < +\infty$; $-\infty < y < +\infty$ and let us suppose that $R_{c,\gamma}^{d,\delta} = \sum_{c<x<d,\, \gamma \leqq y < \delta} R_{xy} \in \mathfrak{B}$ where $-\infty < c < d < +\infty$, $-\infty < \gamma < \delta < +\infty$, further if $-\infty \leqq c < d \leqq +\infty$ and $-\infty < \gamma < \delta < +\infty$, we have

$$\mathbf{P}(A \mid R_{c,\gamma}^{d,\delta}) = \frac{\int_c^d \int_\gamma^\delta \mathbf{P}(A \mid R_{xy}) h(x, y)\, dx\, dy}{\int_c^d \int_\gamma^\delta h(x, y)\, dx\, dy}, \tag{44}$$

where $h(x, y)$ is positive and integrable on any finite rectangle of the (x, y) plane.

Let us suppose further that $R_{c,y}^{d,y} \in \mathfrak{B}$ if $c < d$ and $R_{x,\gamma}^{x,\delta} \in \mathfrak{B}$ if $\gamma < \delta$, further that we have

$$\mathbf{P}(A \mid R_{c,y}^{d,y}) = \frac{\int_c^d \mathbf{P}(A \mid R_{xy}) h(x, y)\, dx}{\int_c^d h(x, y)\, dx} \qquad (c < d) \tag{45}$$

and

$$\mathbf{P}(A \mid R_{\gamma,x}^{\delta,x}) = \frac{\int_\gamma^\delta \mathbf{P}(A \mid R_{xy}) h(x, y)\, dy}{\int_\gamma^\delta h(x, y)\, dy} \qquad (\gamma < \delta). \tag{46}$$

In this case clearly for any fixed y, $h(x, y)$ is, as a function of x, the conditional density function of ξ under hypothesis $\eta=y$ and, for any fixed x, $h(x, y)$ is, as a function of y, the conditional density function of η under the hypothesis $\xi=x$. As a matter of fact, choosing $\gamma=\delta=y$ we have for $c \leqq a < b \leqq d$ from (45)

$$\mathbf{P}(a \leqq \xi < b \mid c \leqq \xi < d, \eta=y)=\frac{\int_a^b h(x, y)\,dx}{\int_c^d h(x, y)\,dx} \tag{47}$$

and choosing $\alpha=\beta=x$ we have for $\gamma \leqq \alpha < \beta \leqq \delta$ from (44)

$$\mathbf{P}(\alpha \leqq \eta < \beta \mid \gamma \leqq \eta < \delta, \xi=x)=\frac{\int_\alpha^\beta h(x, y)\,dy}{\int_\gamma^\delta h(x, y)\,dy}. \tag{48}$$

It should be mentioned that the set R_x defined by $\xi=x$ is not supposed to belong to $\mathfrak{B}$, and thus $h(x, y)$ is not an ordinary conditional density function, but only a generalized conditional density function, in the sense defined in 1.6. By other words, $h(x, y)$ is the generalized density function of the random variable η on the subspace $(R_x, \mathfrak{A}_x, \mathfrak{B}_x, \mathbf{P}_x)$ of $(S, A, B, \mathbf{P})$ where $\mathfrak{A}_x$ is the set of all sets of the form AB_x with $A \in \mathfrak{A}$, $\mathfrak{B}_x$ the set of all sets $B \in \mathfrak{B}$ which are subsets of R_x and $\mathbf{P}_x(A \mid B)$ is defined by

$$\mathbf{P}_x(A \mid B)=\mathbf{P}(A \mid B)$$

for $A \in \mathfrak{A}_x$ and $B \in \mathfrak{B}_x$. If $R_x \in \mathfrak{B}$, clearly $h(x, y)$ is up to a constant factor the ordinary conditional density function of η under the condition R_x and in this case $f(x)=\int h(x, y)\,dy$ is finite.

If $f(x)=\int_{-\infty}^{+\infty} h(x, y)\,dy$ is finite for every x, then $f(x)$ is the generalized density function of ξ. As a matter of fact, we have by Theorem 12

$$\mathbf{P}(a \leqq \xi < b \mid c \leqq \xi < d)=\lim_{\substack{\gamma \to -\infty \\ \delta \to +\infty}} \mathbf{P}(a \leqq \xi < b \mid c \leqq \xi < d, \gamma \leqq \eta < \delta)=\frac{\int_a^b f(x)\,dx}{\int_c^d f(x)\,dx}.$$

Putting

$$g(y \mid x)=\frac{h(x, y)}{f(x)} \tag{49}$$

besides $h(x, y)$, $g(y \mid x)$ is the normed (i. e. ordinary) conditional density function of η under condition $\xi = x$. Similarly, if we denote by $f(x|y)$ the generalized conditional density function of ξ under the condition $\eta = y$, we may put $f(x|y) = h(x, y)$ and thus we have

$$f(x|y) = g(y|x) f(x). \tag{50}$$

If $\int_{-\infty}^{+\infty} h(x, y)\, dx$ is also finite,

$$f^*(x|y) = \frac{g(y|x) f(x)}{\int_{-\infty}^{+\infty} g(y|x) f(x)\, dx} \tag{51}$$

is the normed (ordinary) conditional density function of ξ under the hypothesis $\eta = y$.

The result obtained in this § is a generalization of the well-known method of BAYES. The generalization consists in the possibility that $\int_{-\infty}^{+\infty} f(x)\, dx$ and $\int_{-\infty}^{+\infty} f(x|y)\, dx$ may be divergent.

Clearly if $h(x, y)$ vanishes in some points of the plane, all what we have said remains valid, only we may use as conditions only such sets $c \leqq \xi < d$, $\gamma \leqq \eta < \delta$ for which $\int_c^d \int_\gamma^\delta h(x, y)\, dx\, dy > 0$.

3. 7. Conditional ergodicity of Markov chains. Let us consider a temporally homogeneous Markov chain with a denumerable set of possible states; let us denote the states by the numbers $0, \pm 1, \pm 2, \ldots$ and put $\xi_n = k$ if the system is in state k at time n $(n = 0, 1, 2, \ldots)$. Let us denote by P_{jk} the probability of passing from state j to state k in one step, i. e. $P_{jk} = \mathbf{P}(\xi_{n+1} = k | \xi_n = j)$. We introduce the following notion: if there exists a sequence P_k of positive numbers, with the property that, putting $\mathbf{P}(\xi_n = k | \xi_0 = j) = P_{jk}^{(n)}$, we have

$$\lim_{n \to \infty} \frac{P_{hi}^{(n)}}{P_{jk}^{(n)}} = \frac{P_i}{P_k} \tag{52}$$

for all $h, i, j, k = 0, \pm 1, \pm 2, \ldots$, we shall say that the Markov chain is conditionally ergodic. If the chain is ergodic in the usual sense (see [18]), i. e. if

$$\lim_{n \to \infty} P_{ik}^{(n)} = P_k > 0 \qquad (i, k = 0, \pm 1, \pm 2, \ldots), \tag{53}$$

then (52) holds and we have $\sum_{k=-\infty}^{+\infty} P_k = 1$.

It is possible, however, that (52) holds without (53) taking place; especially, this is the case if (52) holds and $\sum_{k=-\infty}^{+\infty} P_k = +\infty$. If a Markov chain is conditionally ergodic, this implies that the conditional probability distribution generated by ξ_n on the set of integers converges to the conditional probability distribution defined by $\mathbf{P}(A|B)=\frac{\sum_{k\in AB} P_k}{\sum_{k\in B} P_k}$ where A and B are subsets of the set of integers and B is finite and not empty.

Let us suppose that the Markov chain considered is irreducible and all its states are recurrent null states (see [18]) in which case $\lim_{n\to\infty} P_{jk}^{(n)}=0$. If the numbers P_k in (52) satisfy the system of equations

$$P_k=\sum_{j=-\infty}^{\infty} P_j P_{jk} \qquad (k=0,\pm 1,\pm 2,\ldots), \tag{54}$$

we say that the conditional limiting distribution defined by the numbers P_k is a stationary distribution. If the conditional limiting distribution exists, it is always a stationary distribution, i. e. (52) implies (54). As a matter of fact, as we have supposed that all states are recurrent, the series $\sum_{n=1}^{\infty} P_{kk}^{(n)} (k=0,\pm 1,\ldots)$ diverges, and thus, it follows from (52) that

$$V_k=\lim_{n\to\infty}\frac{\sum_{r=1}^{n} P_{kk}^{(r)}}{\sum_{r=1}^{n} P_{00}^{(r)}}=\frac{P_k}{P_0}. \tag{55}$$

But it has been proved by C. DERMAN [19] that if all states of an irreducible chain are recurrent null states, the only solution $P_k=V_k$ of the system of equations (54) with $V_0=1$ is given by the limits in (55) which always exist.

The existence of the limits (52) has been proved by P. ERDŐS and K. L. CHUNG [20] in the special case of additive Markov chains. Let us consider an additive Markov chain, i. e. put $\xi_n=\xi_0+\delta_1+\delta_2+\cdots+\delta_n$ where the δ_k's are independent equidistributed random variables which take only integer values, further suppose that the following conditions are satisfied:

Let us put $\mathbf{P}(\delta_k=r)=W_r$ $(r=0,\pm 1,\pm 2,\ldots)$. We suppose that the greatest common divisor of the differences $r-s$ where r and s are such integers, for which $W_r>0$ and $W_s>0$, is equal to 1, further that $\sum_{r=-\infty}^{+\infty}|r|W_r<+\infty$ and $\sum_{r=-\infty}^{+\infty} rW_r=0$. It has been shown by K. L. CHUNG and P. ERDŐS [20] that in this case the chain is (in our terminology) conditionally ergodic, and the limiting conditional distribution is the uniform distribution over the set of all integers, i. e. $P_k\equiv 1$. If the variance $D^2=\sum_{r=-\infty}^{+\infty} r^2 W_r$ of the variables δ_k exists, this follows easily by LAPLACE'S method; as a matter of fact in this case, putting $\varphi(t)=\sum_{r=-\infty}^{+\infty} W_r e^{irt}$, we have

$$P_{jk}^{(n)}=\frac{1}{2\pi}\int_{-\pi}^{+\pi}(\varphi(t))^2 e^{it(j-k)}\,dt \tag{56}$$

and thus, by LAPLACE's method,

$$P_{ik}^{(n)}\sim\frac{1}{D\sqrt{2\pi n}} \tag{57}$$

what implies

$$\lim_{n\to\infty}\frac{P_{hj}^{(n)}}{P_{ik}^{(n)}}=1. \tag{58}$$

If $D=+\infty$, the proof is more intricate.

The existence of the limits (52) in the general case is still open, as has been pointed out by CHUNG [21].

3. 8. On a paradox of the renewal theory. Renewal theory (see [22]) deals with stochastic processes of the following type ("recurrent processes"): events occur at the moments $0<\tau_1<\tau_2<\cdots<\tau_n<\cdots$ where the differences $\tau_n-\tau_{n-1}=\xi_n$ are independent positive random variables with the same distribution function $F(x)=\mathbf{P}(\xi_n<x)$; for any $t>0$ let $\tau_{n(t)}$ denote the moment when the first event after time t occurs, i. e. we suppose $\tau_{n(t)-1}\leqq t<\tau_{n(t)}$ and put $\vartheta(t)=\tau_{n(t)}-t$; $\vartheta(t)$ is the waiting time at t, i. e. the time somebody arriving at time t has to wait until the next event occurs. Let us now suppose that somebody arrives at random in the time interval $(0, T)$ at time τ, where τ is uniformly distributed in $(0, T)$ and let us put $\delta(T)=\vartheta(\tau)$. It has been proved by W. FELLER [22] that if $m=\int_0^\infty x\,dF(x)$ (the mean value of the time interval between consecutive events) exists, and $F(x)$ is not a lattice distribution function, the distribution of $\vartheta(T)$ tends for $T\to\infty$ to the limiting distribution

$$H(x)=\lim_{T\to\infty}\mathbf{P}(\vartheta(T)<x)=\frac{1}{m}\int_0^x(1-F(u))\,du. \tag{59}$$

It has been shown by L. TAKÁCS [23] that if we consider $\delta(T)$ instead of $\vartheta(T)$, the limiting distribution exists for any $F(x)$ with finite mean value, and is the same as in the case of $\vartheta(T)$, i. e. we have

$$H(x)=\lim_{T\to\infty}\mathbf{P}(\delta(T)<x)=\frac{1}{m}\int_0^x(1-F(u))\,du. \tag{60}$$

It has been pointed out by L. TAKÁCS [23] that the fact that the mean value of the distribution function $H(x)$ is $\frac{m^2+\sigma^2}{2m}$ where σ^2 denotes the variance of ξ_n, i. e. $\sigma^2 = \int_0^\infty (x-m)^2 dF(x)$, implies that if $\sigma = +\infty$, the mean value of $\delta(T)$ tends to $+\infty$ in spite of the fact that $\delta(T)$ is the part of some time interval ξ_n having the finite mean value m.

The situation is still more paradox, if $m = \infty$; it can be shown (see [9]) that in this case

$$\lim_{T\to\infty} \mathbf{P}(\delta(T) < x) = 0 \tag{61}$$

for all finite x; nevertheless, if putting $G(x) = \int_0^x (1-F(t))\, dt$ the condition

$$\lim_{x\to\infty} \frac{x G'(x)}{G(x)} = 0 \tag{62}$$

is satisfied, there exists the limit of the conditional distribution of $\delta(T)$, and we have

$$\lim_{T\to\infty} \mathbf{P}(\delta(T) < x \mid \delta(T) < y) = \frac{G(x)}{G(y)} \quad \text{for} \quad 0 < x \leqq y < +\infty. \tag{63}$$

3.9. Deduction of Maxwell's law of velocity distribution. Let us consider an ideal gas consisting of N particles with equal masses m, let ξ_k, η_k, ζ_k denote the x, y, z-component of the velocity of the k-th particle; we suppose that ξ_k, η_k, ζ_k $(k = 1, 2, \ldots, N)$ are random variables defined on a conditional probability space $[S, \mathcal{A}, \mathcal{B}, \mathbf{P}(A|B)]$. Let us suppose that the conditional probability distribution generated by the variables ξ_k, η_k, ζ_k $(k = 1, 2, \ldots, N)$ in the $3N$-dimensional Euclidean space is uniform, further that the space is regular with respect to the sets defined by $\xi_k = x_k$, $\eta_k = y_k$, $\zeta_k = z_k$ $(-\infty < x_k < +\infty, -\infty < y_k < +\infty, -\infty < z_k < +\infty;\ k = 1, 2, \ldots, N)$, further that the sets $\sum_{(t_1, \ldots, t_{3N}) \in M} R_{t_1, \ldots, t_{3N}}$ belong to $\mathcal{B}$, where $R_{t_1, \ldots, t_{3N}}$ is the set defined by $\xi_1 = t_1, \eta_1 = t_2, \zeta_1 = t_3, \xi_2 = t_4, \ldots, \zeta_n = t_{3N}$ and M is a Borel subset of the $3N$-dimensional Euclidean space, further that the space is also regular with respect to the sets defined by $\xi_k^2 + \eta_k^2 + \zeta_k^2 = z_k$ $(k = 1, 2, \ldots, N;\ 0 < z_k < +\infty)$. All these conditions are satisfied, if S is the $3N$-dimensional Euclidean space, $\mathcal{A}$ the set of all Borel subsets of S, $\mathcal{B}_k$ the set of all Borel subsets of S which have positive and finite k-dimensional measure $(k = 1, 2, \ldots, 3N)$, $\mathcal{B}_0$ the set of all points of S and $\mathcal{B} = \sum_{k=0}^{3N} \mathcal{B}_k$ and we put $\mathbf{P}(A|B) = 1$ if $B \in \mathcal{B}_0$ and $B \subseteq A$, $\mathbf{P}(A|B) = 0$ if $B \in \mathcal{B}_0$ and $B \subseteq \bar{A}$, further $\mathbf{P}(A|B) = \frac{m_k(AB)}{m_k(B)}$ if $B \in \mathcal{B}_k$, where $m_k(C)$ is the k-dimensional measure of C $(k = 1, 2, \ldots, 3N)$, further if denoting by $P=(x_1, y_1, z_1, \ldots, x_N, y_N, z_N)$ a point of S, we put $\xi_1(P)=x_1$, $\eta_1(P)=y_1, \ldots, \zeta_N(P)=z_N$.

It follows that the kinetic energy $\varepsilon_k=\frac{1}{2}m(\xi_k^2+\eta_k^2+\zeta_k^2)$ of the k-th particle has the density function $\sqrt{x}$ for $0<x<+\infty$ and thus by 2.6 that the kinetic energy $\varepsilon=\sum_{k=1}^{N}\varepsilon_k$ of the whole gas has the density function $g_N(x)$ defined by the recursion

$$g_1(x)=\sqrt{x}, \quad g_k(x)=\int_0^x g_{k-1}(x-y)\sqrt{y}\,dy \quad \text{for} \quad x \geqq 0.$$

Thus

$$g_N(x)=\frac{\left(\frac{\sqrt{\pi}}{2}\right)^N}{\Gamma\left(\frac{3N}{2}\right)} x^{\frac{3N}{2}-1} \qquad (x \geqq 0),$$

and therefore, as any constant factor is irrelevant, we may take $x^{\frac{3N}{2}-1}$ $(x>0)$ for the density function of ε.

Similarly, the conditional density function of ε under condition $\varepsilon_k=t$ is clearly $(x-t)^{\frac{3N-5}{2}}$ for $x \geqq t$. It can be easily verified that the conditions for applying formula (48) are fulfilled, and thus we obtain that the conditional density function of ε_k with respect to the condition $\varepsilon=E$ is

$$f(x\,|\,E)=\frac{(E-x)^{\frac{3N-5}{2}}\sqrt{x}}{\int_0^E (E-x)^{\frac{3N-5}{2}}\sqrt{x}\,dx}=\frac{\left(1-\frac{x}{E}\right)^{\frac{3N-5}{2}}\sqrt{x}}{\int_0^E\left(1-\frac{x}{E}\right)^{\frac{3N-5}{2}}\sqrt{x}\,dx}.$$

If N is a large number, we have, putting $\beta=\frac{3N}{2E}$, approximately $\left(1-\frac{x}{E}\right)^{\frac{3N-5}{2}}\sim e^{-\beta x}$ and thus

$$f(x\,|\,E)\sim\frac{2\beta^{3/2}}{\sqrt{\pi}}\sqrt{x}\cdot e^{-\beta x}.$$

Thus the conditional density function $g(x\,|\,E)=mxf\left(\frac{m}{2}x^2\,|\,E\right)$ of the velocity $v_k=\sqrt{\xi_k^2+\eta_k^2+\zeta_k^2}$ of the k-th particle under condition $\varepsilon=E$ is approximately

$$g(x\,|\,E)\sim\sqrt{\frac{2}{\pi}}\frac{1}{\sigma^{3/2}}x^2 e^{-\frac{x^2}{2\sigma^2}}$$

where $\sigma=\frac{1}{\sqrt{m\beta}}$. As it is known that $\frac{E}{N}=\frac{3}{2}kT$ where T denotes the absolute temperature of the gas and k the constant of Boltzmann, we have $\beta=\frac{1}{kT}$ and $\sigma=\sqrt{\frac{kT}{m}}$. Thus finally we obtain

$$g(x|E)\sim\sqrt{\frac{2}{\pi}}\left(\frac{m}{kT}\right)^{\frac{3}{2}}x^2e^{-\frac{mx^2}{2kT}}\qquad(0<x<+\infty), \tag{64}$$

i. e. Maxwell's law of velocity distribution.

Summarising, we may state our result as follows: if the components of the velocity of each of N particles are independent of each other and of the components of the velocity of the other particles, further if all these random variables have regular uniform (a priori) distributions on the whole real axis, then the (a posteriori) distribution of the velocity of each particle under the condition that the kinetic energy of the whole system is given, is approximately, if N is large, the Maxwell distribution (64).

The proof given above shows clearly that the law of Maxwell is an approximation, which is valid only for a large number of particles. For a small number N we obtain the exact formula

$$g_N(x|E)=\sqrt{\frac{2}{\pi}}\left(\frac{m}{E}\right)^{3/2}\frac{\Gamma\left(\frac{3N}{2}\right)}{\Gamma\left(\frac{3N-3}{2}\right)}x^2\left(1-\frac{mx^2}{2E}\right)^{\frac{3N-5}{2}}. \tag{65}$$

It can also be shown by the same way that the conditional density function of each of the variables ξ_k, η_k, ζ_k under the condition $\varepsilon=E$ is exactly

$$h(x|E)=\frac{1}{\sqrt{2\pi}}\sqrt{\frac{m}{E}}\cdot\frac{\Gamma\left(\frac{3N}{2}\right)}{\Gamma\left(\frac{3N-1}{2}\right)}\left(1-\frac{mx^2}{2E}\right)^{\frac{3N-3}{2}}\sim\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{x^2}{2\sigma^2}} \tag{66}$$

where $\sigma=\sqrt{\frac{kT}{m}}$. Thus under the condition $\varepsilon=E$ all these variables are approximately normally distributed. This is the reason why the velocities have a Maxwell distribution, as the Maxwell distribution can be defined as the distribution of a variable $\sqrt{\xi^2+\eta^2+\zeta^2}$ where ξ, η, ζ are independent and normally distributed random variables with mean value 0 and variance σ^2.

§ 4. Conditional laws of large numbers

Conditional probability is in the same relation to conditional relative frequency as ordinary probability to ordinary relative frequency. This relation, which is well known as an empirical fact from everyday experience, is described mathematically by the laws of large numbers.

The laws of large numbers concerning the behaviour in the limit of the conditional relative frequency (and generalizations concerning conditional means of observations) shall be called "conditional laws of large numbers". These laws will be investigated in this §. Before going into details it should be emphasized that the conditional probability $\mathbf{P}(A|B)$ is considered in this paper as an objective characteristic of the random event A, under an objective condition B, and its value as that number in the near neighbourhood of which the corresponding conditional relative frequency $\frac{k_{AB}}{k_B}$ will be found in general, if a sufficiently great number n of observations is made, where k_B denotes the number of those under the n observations with respect to which the condition B has been realised, and k_{AB} the number of those observations which gave such results that, besides the condition B being realised, the event A occurred also, and it is supposed that $k_B > 0$.

4. 1. A strong law of large numbers. In this section we consider random variables defined on an ordinary probability space. We shall need the following

LEMMA 1. *Let $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ be mutually independent random variables with mean values $\mathbf{M}(\xi_n) = M_n \geqq 0$ and finite variances $D_n^2 = \mathbf{D}^2(\xi_n)$. Let us put $\zeta_n = \sum_{k=1}^{n} \xi_k$ and $A_n = \mathbf{M}(\zeta_n) = M_1 + M_2 + \cdots + M_n$, and suppose that the following conditions are fulfilled:*

a) $\lim_{n\to\infty} A_n = +\infty$,

b) $\sum_{n=1}^{\infty} \frac{D_n^2}{A_n^2} \leqq +\infty$.

It follows that

$$\mathbf{P}\left(\lim_{n\to\infty} \frac{\zeta_n}{A_n} = 1\right) = 1. \tag{67}$$

PROOF. Lemma 1 (see e. g. [27], p. 238) is a consequence of the well-known theorem of KOLMOGOROV, according to which if $\eta_1, \eta_2, \ldots, \eta_n, \ldots$ are mutually independent random variables and $\mathbf{M}(\eta_k) = 0$ $(k = 1, 2, \ldots)$, then $\sum_{k=1}^{\infty} \eta_k$ converges with probability 1, provided that the series $\sum \mathbf{D}^2(\eta_k)$ converges. Lemma 1 can be proved also directly by applying the following inequality which has been found recently by J. HÁJEK (see [12]):

LEMMA 2. *Let $\eta_1, \eta_2, \ldots, \eta_n, \ldots$ denote mutually independent random variables with mean values $\mathbf{M}(\eta_k)=0$ and variances $\mathbf{D}^2(\eta_k)=D_k^2$ $(k=1,2,\ldots)$. Let c_k denote a decreasing sequence of positive numbers, $c_k \leqq c_{k+1}$ $(k \geqq n)$; then we have for any $\varepsilon>0$*

$$\mathbf{P}(\max_{n\leqq k} c_k|\eta_1+\eta_2+\cdots+\eta_k| \geqq \varepsilon) \leqq \frac{1}{\varepsilon^2}\left(c_n^2\sum_{k=1}^{n} D_k^2 + \sum_{k=n+1}^{\infty} c_k^2 D_k^2\right). \tag{68}$$

Applying Lemma 2, Lemma 1 can be proved as follows: it follows from (68), applied for $\eta_k=\xi_k-M_k$, $c_k=\frac{1}{A_k}$ that

$$\mathbf{P}\left(\sup_{n\leqq k}\left|\frac{\zeta_k}{A_k}-1\right| \geqq \varepsilon\right) \leqq \frac{1}{\varepsilon^2}\left(\frac{\sum_{k=1}^{n} D_k^2}{A_n^2}+\sum_{k=n+1}^{\infty}\frac{D_k^2}{A_k^2}\right). \tag{69}$$

As a) and b) clearly imply that

$$\lim_{n\to\infty}\frac{\sum_{k=1}^{n} D_k^2}{A_n^2}=0, \tag{70}$$

it follows from (69) that

$$\lim_{n\to\infty}\mathbf{P}\left(\sup_{n\leqq k}\left|\frac{\zeta_k}{A_k}-1\right| \geqq \varepsilon\right)=0 \quad \text{for any} \quad \varepsilon>0. \tag{71}$$

and (71) is evidently equivalent to (67).

4.2. A conditional law of large numbers.

THEOREM 15. *Let $[S, \mathfrak{A}, \mathfrak{B}, \mathbf{P}(A|B)]$ denote a conditional probability space and $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ random variables on S which are mutually independent with respect to $C \in \mathfrak{B}$. Let $\mathfrak{J}$ denote the interval $a \leqq x < b$ $(a<b)$ of the real axis. Let B_n denote the set of those $a \in S$ for which $\xi_n(a) \in \mathfrak{J}$, and suppose that $B_n \subseteq C$ and $B_n \in \mathfrak{B}$; let us suppose that $\mathbf{M}(\xi_n|B_n)=M_n>0$ and $\mathbf{D}^2(\xi_n|B_n)=D_n^2$ exists $(n=1,2,\ldots)$. Let us put $\mathbf{P}(B_n|C)=p_n$ and suppose that the following conditions are satisfied:*

$$\sum_{n=1}^{\infty} p_n=+\infty \quad \text{and} \quad \sum_{n=1}^{\infty} p_n M_n=+\infty; \tag{72}$$

$$\lim_{n\to\infty}\frac{\sum_{k=1}^{n} p_k M_k}{\sum_{k=1}^{n} p_k}=M \tag{73}$$

exists;

(74) $$\sum_{k=1}^{\infty} \frac{p_k(D_k^2+(1-p_k)M_k^2)}{\left(\sum_{j=1}^{k} p_j M_j\right)^2} < +\infty.$$

Then we have

(75) $$\mathbf{P}\left(\lim_{n\to\infty} \frac{\sum\limits_{\substack{\xi_k \in \mathfrak{J} \\ 1\le k\le n}} \xi_k}{\sum\limits_{\substack{\xi_k \in \mathfrak{J} \\ 1\le k\le n}} 1} = M \,\middle|\, C \right) = 1.$$

PROOF. Let us define the random variable ε_k as follows: $\varepsilon_k = 1$ if $\xi_k \in \mathfrak{J}$ and $\varepsilon_k = 0$ if $\xi_k \notin \mathfrak{J}$; let us put $\xi_k^* = \xi_k \varepsilon_k$. Then we have

$$\mathbf{M}(\xi_k^* \mid C) = p_k M_k \quad \text{and} \quad \mathbf{M}(\xi_k^{*2} \mid C) = p_k(D_k^2 + M_k^2)$$

and thus

$$\mathbf{D}^2(\xi_k^* \mid C) = p_k(D_k^2 + (1-p_k)M_k^2).$$

Applying Lemma 1 to the sequence ξ_k^* of random variables on C, it follows that

(76) $$\mathbf{P}\left(\lim_{n\to\infty} \frac{\sum_{k=1}^{n} \xi_k^*}{\sum_{k=1}^{n} p_k M_k} = 1 \,\middle|\, C \right) = 1.$$

On the other hand, let us apply Lemma 1 to the sequence of random variables ε_k on C. As

$$\mathbf{M}(\varepsilon_k \mid C) = p_k \quad \text{and} \quad \mathbf{D}^2(\varepsilon_k \mid C) = p_k(1-p_k),$$

it follows that

(77) $$\mathbf{P}\left(\lim_{n\to\infty} \frac{\sum_{k=1}^{n} \varepsilon_k}{\sum_{k=1}^{n} p_k} = 1 \,\middle|\, C \right) = 1.$$

Combining (76), (77) and condition b) of the theorem, and taking into account that

$$\sum_{\substack{\xi_k \in \mathfrak{J} \\ 1\le k\le n}} \xi_k = \sum_{k=1}^{n} \xi_k^* \quad \text{and} \quad \sum_{\substack{\xi_k \in \mathfrak{J} \\ 1\le k\le n}} 1 = \sum_{k=1}^{n} \varepsilon_k,$$

(75) follows. Thus Theorem 15 is proved.

The statement of Theorem 15 can be expressed in words as follows: the conditional empirical mean value of those of the variables $\xi_1, \xi_2, \ldots, \xi_n$ which take on values lying in the interval $\mathfrak{J}$, converges with conditional probability 1 with respect to C to the limit M defined by (73). In the special case, when $M_n = M > 0$ and $D_n = D > 0$ do not depend on n, the conditions (72)—(74) reduce to the single condition that the series $\sum_{n=1}^{\infty} p_n$ diverges. Especially let us suppose further that $\mathfrak{J}$ is the closed interval [0, 1] and the only values in $\mathfrak{J}$ which the variables ξ_n can take on B_n are the values 0 and 1 (of course ξ_n can take also other values outside $\mathfrak{J}$). Suppose that A_n is the set on which $\xi_n = 1$, and $p = \mathbf{P}(A_n | B_n) > 0$ does not depend on n.

In this case the events A_n and B_n can be interpreted as the events consisting in the realisation of some events A and B, respectively, at the n-th experiment in a sequence of independent experiments, and we may put $p = \mathbf{P}(A|B)$; in this case

$$f_n(A|B) = \frac{\sum\limits_{\substack{\xi_k \in \mathfrak{J} \\ 1 \leq k \leq n}} \xi_k}{\sum\limits_{\substack{\xi_k \in \mathfrak{J} \\ 1 \leq k \leq n}} 1} = \frac{\sum\limits_{\xi_k = 1, 1 \leq k \leq n} 1}{\sum\limits_{\xi_k = 0 \text{ or } 1, 1 \leq k \leq n} 1}$$

is the conditional relative frequency of the event A with respect to the event B in course of the first n observations.[12] The statement of Theorem 15 gives for this special case

$$\mathbf{P}(\lim_{n \to \infty} f_n(A|B) = \mathbf{P}(A|B) | C) = 1, \tag{78}$$

if $\mathbf{P}(B|C) > 0$, i. e. *the conditional relative frequency of the event A with respect to the event B converges with the conditional probability* 1 *(with respect to C) to the conditional probability of A with respect to B.*

4. 3. Generalization of the theorem of Borel on normal decimals. The theorem of Borel [25] states that if $f_n(k; x)$ denotes the relative frequency of the number k between the first n digits of the decimal expansion of the real number x $(0 \leq x < 1)$, we have

$$\lim_{n \to \infty} f_n(k; x) = \frac{1}{10} \qquad (k = 0, 1, 2, \ldots, 9) \tag{79}$$

for almost every x; similarly all possible digits occur in the limit with equal frequency in the expansion, in the number system with any basis q of almost all real numbers.

We consider a more general representation of real numbers; namely expansion of real numbers into Cantor's series. Let q_n $(n = 1, 2, \ldots)$ denote an arbitrary sequence of integers, $q_n \geqq 2$. It is easy to prove (see [24]) that

[12] This interpretation suggests itself in the following special case: let us form the Cartesian product of a denumerable sequence of isomorphic conditional probability spaces $[S, \mathfrak{A}, \mathfrak{B}, \mathbf{P}]$ and denote by A_n resp. B_n the sets defined by $\overset{1}{\breve{S}} * \overset{2}{\breve{S}} * \cdots * \overset{n}{\breve{A}} * S * \cdots$ resp. $\overset{1}{\breve{S}} * \overset{2}{\breve{S}} * \cdots * \overset{n}{\breve{B}} * S * \cdots$ (see section 1.11).

any real number x $(0 \leqq x < 1)$ can be represented in the form

$$x = \sum_{n=1}^{\infty} \frac{\varepsilon_n(x)}{q_1 q_2 \cdots q_n},$$

where $\varepsilon_n(x)$ may take on the values $0, 1, \ldots, q_n - 1$; the digits $\varepsilon_n(x)$ are uniquely determined except for those rational numbers x which can be written in the form $x = \sum_{n=1}^{N} \frac{\varepsilon_n(x)}{q_1 q_2 \cdots q_n}$ $(\varepsilon_N(x) > 0)$, where besides this finite representation the alternative infinite representation

$$x = \sum_{n=1}^{N-1} \frac{\varepsilon_n(x)}{q_1 q_2 \ldots q_n} + \frac{\varepsilon_N(x) - 1}{q_1 q_2 \ldots q_N} + \sum_{n=N+1}^{\infty} \frac{q_n - 1}{q_1 q_2 \ldots q_n}$$

is also possible. We shall prove that in the case $q_n \to +\infty$, when all non-negative integers are possible "digits", all these digits occur in the limit with the same (conditional) relative frequency, provided that $\sum_{n=1}^{\infty} \frac{1}{q_n}$ diverges. This is contained in the following theorem which is a consequence of Theorem 15.

THEOREM 16. *Let the numbers q_n satisfy the following conditions:*

a) *q_n is an integer, $q_n \geqq 2$ for $n = 1, 2, \ldots$,*

b) *$q_n \geqq N$ for $n \geqq n_0$,*

c) $\sum_{n=1}^{\infty} \frac{1}{q_n} = +\infty$.

Let us consider the expansion of real numbers x $(0 \leqq x < 1)$ into Cantor's series with quotients q_n, i. e. the expansion

$$x = \sum_{n=1}^{\infty} \frac{\varepsilon_n(x)}{q_1 q_2 \ldots q_n} \tag{80}$$

where $\varepsilon_n(x)$ is one of the numbers $0, 1, \ldots, q_n - 1$.

Let $F_n(k; x)$ denote how many of the "digits" $\varepsilon_1(x), \varepsilon_2(x), \ldots, \varepsilon_n(x)$ are equal to k. Then for almost every x in $(0, 1)$ we have

$$\lim_{n \to \infty} \frac{F_n(j; x)}{F_n(k; x)} = 1 \quad \text{for} \quad 0 \leqq j \leqq k \leqq N-1. \tag{81}$$

REMARK 1. If $q_n \to \infty$, condition b) is fulfilled for every N, and thus (81) holds for any pair of non-negative integers j and k, i. e. *all digits $0, 1, 2, \ldots$ occur in the limit asymptotically with the same frequency, in the Cantor expansion of almost all real numbers.* This is the case, for instance, if $q_n = n + 1$, i. e. if we consider the expansion

$$x = \sum_{k=2}^{\infty} \frac{\varepsilon_k(x)}{k!} \tag{82}$$

where $\varepsilon_k(x)$ can have the values $0, 1, \ldots, k - 1$.

REMARK 2. It follows from (81) that for almost all x

$$\lim_{M\to\infty}\frac{F_M(k;x)}{\sum_{j=0}^{l}F_M(j;x)}=\frac{1}{l+1}\quad\text{for}\quad k=0,1,\ldots,l\quad\text{and}\quad l=1,2,\ldots,N-1. \tag{83}$$

This is an other way of expressing that the frequencies of all digits $0,1,\ldots,N-1$ are in the limit equal to another.

If $q_n\to\infty$, we have

$$\lim_{M\to\infty}\frac{F_M(k;x)}{M}=0\quad\text{for}\quad k=0,1,\ldots \tag{84}$$

for almost all x. This can be shown to be a consequence of Lemma 1 (see (86) below).

REMARK 3. Theorem 16 clearly reduces to the theorem of Borel if $q_n=10$ for $n=1,2,\ldots$.

PROOF. Theorem 16 is a special case of Theorem 15, but it is more simple to prove it directly by means of Lemma 1 of section 4. 1. Let us consider the ordinary probability space $[S,\mathcal{A},\mathbf{P}(A)]$ which we obtain if we choose for S the interval $(0,1)$, for $\mathcal{A}$ the set of all measurable subsets of S and put $P(A)=m(A)$ where $m(A)$ is the Lebesgue measure of the set A. Let us define the random variables $\xi_{nk}=\xi_{nk}(x)$ $(0\leqq x<1)$ for $0\leqq k\leqq n-1$ $(n=1,2,\ldots)$ as follows:

$$\xi_{nk}(x)=\begin{cases}1 & \text{if}\quad \varepsilon_n(x)=k,\\ 0 & \text{otherwise.}\end{cases} \tag{85}$$

The variables ξ_{nk} $(n=1,2,\ldots)$ are clearly mutually independent, further we have $M_{nk}=\mathbf{M}(\xi_{nk}(x))=\frac{1}{q_n}$ and $D^2_{nk}=\mathbf{D}^2(\xi_{nk}(x))=\frac{1}{q_n}\left(1-\frac{1}{q_n}\right)$ for $q_n\geqq k$, i. e. for $n\geqq n_0$ if $k<N$; the values of M_{nk} resp. D_{nk} for $n<n_0$ are irrelevant. Condition a) of Lemma 1 follows from supposition b) of Theorem 16, and condition b) of Lemma 1 follows also from supposition b) of Theorem 16. Thus we have

$$\mathbf{P}\left(\lim_{M\to\infty}\frac{\sum_{n=1}^{M}\xi_{nk}}{\sum_{n=n_0}^{M}\frac{1}{q_n}}=1\right)=1. \tag{86}$$

As (86) holds for all $k=0,1,\ldots,N-1$ and $\sum_{n=1}^{M}\xi_{nk}=F_M(k;x)$ the assertion of Theorem 16 follows.

The results of this section (as well as those of section 3. 2) show that there exist sequences $x_1,x_2,\ldots,x_n$, consisting of the numbers $0,1,2,\ldots$, in which the relative frequency $\frac{F_n(k)}{n}$ (where $F_n(k)$ denotes how many of the numbers $x_1, x_2, \ldots, x_n$ are equal to k) of every number $k=0, 1, \ldots$ tends to 0 for $n \to \infty$, but the conditional relative frequencies converge to definite limits, i. e.

$$\lim_{n\to\infty} \frac{F_n(h)}{\sum_{j=0}^{k} F_n(j)} = \frac{P_h}{\sum_{j=0}^{k} P_j} \qquad (h=0, 1, \ldots, k;\ k=1, 2, \ldots)$$

where $P_k > 0$ for $k=0, 1, \ldots$ and $\sum_{k=0}^{\infty} P_k = +\infty$. Such sequences may be considered as mathematical models of sequences of observations on a random variable ξ, defined on a conditional probability field, and having the conditional distribution

$$\mathbf{P}(\xi = h \mid 0 \leq \xi \leq k) = \frac{P_h}{\sum_{j=0}^{k} P_j} \qquad (h=0, 1, \ldots, k;\ k=1, 2, \ldots).$$

(Received 25 July 1955)

Bibliography

[1] A. N. Kolmogoroff, *Grundbegriffe der Wahrscheinlichkeitsrechnung* (Berlin, 1933).

[2] H. Jeffreys, *Theory of probability* (London, 1943).

[3] H. Reichenbach, *Wahrscheinlichkeitslehre* (Leiden, 1935).

[4] M. I. Keynes, *A treatise on probability theory.*

[5] B. O. Koopman, The axioms and algebra of intuitive probability, *Annals of Math.*, **41** (1940), pp. 269—292.

[6] A. H. Copeland, Postulates for the theory of probability, *Amer. Journal of Math.*, **63** (1941), pp. 741—762.

[7] G. A. Barnard, Statistical inference, *Journal of the Royal Statistical Soc.*, Ser. B, **11** (1949), pp. 115—139.

[8] I. J. Good, *Probability and the weighing of evidence* (London, 1950).

[9] A. Rényi, A valószínűségszámítás új axiomatikus felépítése, *MTA Mat. és Fiz. Oszt. Közl.*, **4** (1954), pp. 369—427.

[10] A. Rényi, On a new axiomatic foundation of the theory of probability. Under press in the Volume 1 of the *Proceedings of the International Mathematical Congress in Amsterdam, 1954.*

[11] A. Rényi, Über die axiomatische Begründung der Wahrscheinlichkeitsrechnung. Under press in the *Proceedings of the Conference on Probability Theory in Berlin, 1954.*

[12] J. Hájek and A. Rényi, Generalization of an inequality of Kolmogorov, *Acta Math. Acad. Sci. Hung.*, **6** (1955), pp. 281—283.

[13] Á. Császár, Sur la structure des espaces de probabilité conditionnelle, *Acta Math. Acad. Sci. Hung.*, **6** (1955), pp. 337—361.

[14] J. Neyman, L'estimation statistique traitée comme un problème classique de probabilité, *Act. Sci. et Ind.*, 739 (Paris, 1938), pp. 25—57.

[15] P. Halmos, *Measure theory* (New-York, 1951).

[16] P. Erdős, On the integers having exactly k prime factors, *Annals of Math.*, **49** (1948), pp. 53—66.

[17] B. V. Gnedenko, Über ein lokales Grenzwerttheorem für gleichverteilte unabhängige Summanden, *Wiss. Zeitschrift der Humboldt-Universität Berlin*, **3** (1954), pp. 287—293.

[18]. J. L. Doob, *Stochastic processes* (New-York, 1953).

[19] C. Derman, A solution to a set of fundamental equations in Markov chains, *Proc. Amer. Math. Soc.*, **5** (1954), pp. 332—334.

[20] Erdős P. and K. L. Chung, Probability limit theorems assuming only the first moment. I, *Memoirs of the Amer. Math. Soc.*, **6** (1952).

[21] K. L. Chung, Contributions to the theory of Markov chains, *Trans. Amer. Math. Soc.*, **76** (1954), pp. 397—419. See also K. L. Chung, An ergodic theorem for stationary Markov chains with a countable number of states, *Proc. Internat. Congress of Math. Cambridge (USA), 1950*, (Amer. Math. Soc., 1952), I, p. 568.

[22] W. Feller, On the integral equation of renewal theory, *Ann. Math. Stat.*, **12** (1941), pp. 243—267. See also S. Täcklind, Elementare Behandlung vom Erneuerungsproblem, *Skandinavisk Aktuarietidskrift*, (1944), pp. 1—15, and Fourier-analytische Behandlung vom Erneuerungsproblem, *Skandinavisk Aktuarietidskrift*, (1945), pp. 101—105; and J. L. Doob, Renewal theory from the point of view of the theory of probability, *Trans. Amer. Math. Soc.*, **63** (1948), pp. 422—438.

[23] L. Takács, Egy új módszer rekurrens sztochasztikus folyamatok tárgyalásánál, *MTA Alk. Mat. Int. Közleményei*, **2** (1953), pp. 135—163.

[24] O. Perron, *Irrationalzahlen* (Berlin, 1921), pp. 111—116.

[25] E. Borel, Les probabilités dénombrables et leurs applications arithmétiques, *Rend. Circ. Math. Palermo*, **27** (1909), pp. 247—271.

[26] H. Hahn and A. Rosenthal, Set functions (New Mexico, 1948).

[27] M. Loève, *Probability theory* (New-York, 1955).

НОВОЕ АКСИОМАТИЧЕСКОЕ ПОСТРОЕНИЕ ТЕОРИИ ВЕРОЯТНОСТЕЙ

А. Реньи (Будапешт)

(Резюме)

Работа содержит новое аксиоматическое построение теории вероятностей; основное понятие в этой новой теории, являющейся обобщением теории А. Н. Колмогорова, — понятие условной вероятности. Потребность в разработке новой теории была вызвана тем, что в теории А. Н. Колмогорова неограниченные меры исключены, так, например, не имеет смысла говорить о равномерном распределении, во всем n-мерном евклидовом пространстве, в то время, как приложения теории вероятности к физике, к интегральной геометрии, теории чисел и т. д. требуют рассмотрения таких не нормируемых распределений. Выбирая в качестве основного понятия теории вероятностей понятие условной вероятности, упомянутым не нормируемым распределениям можно дать точный математический смысл. Новая теория делает возможным обобщение целого ряда теорем теории вероятностей и новые приложения теории; работа занимается разработкой новой теории и знакомит с некоторыми ее типичными приложениями. Теория исходит из следующих предположений и аксиом:

Пусть S есть любое множество, которое мы будем называть пространством событий; пусть $\mathcal{A}$ есть σ-алгебра подмножеств пространства S, элементы $\mathcal{A}$ мы будем

называть событиями; пусть $\mathcal{B}$ есть некоторое не пустое подмножество от $\mathcal{A}$ ($\mathcal{B}$ есть множество всех допустимых условии), а $\mathbf{P}(A|B)$ функция множеств от двух переменных, определенная, если $A \in \mathcal{A}$ и $B \in \mathcal{B}$; число $\mathbf{P}(A|B)$ мы будем называть условной вероятностью события A относительно условия B. Предположим, что выполняются следующие 3 аксиомы:

Аксиома I. $\mathbf{P}(A|B) \geqq 0$, если $A \in \mathcal{A}$ и $B \in \mathcal{B}$, и $\mathbf{P}(B|B) = 1$, если $B \in \mathcal{B}$.

Аксиома II. Если $B \in \mathcal{B}$ фиксирован, $\mathbf{P}(A|B)$ есть мера на σ-алгебре $\mathcal{A}$.

Аксиома III. Если $A \in \mathcal{A}$, $B \in \mathcal{A}$, $C \in \mathcal{B}$ и $BC \in \mathcal{B}$, то

$$\mathbf{P}(A|BC)\,\mathbf{P}(B|C) = \mathbf{P}(AB|C).$$

Если аксиомы I—III выполненны, то совокупность множества S, σ-алгебры $\mathcal{A}$ системы множеств $\mathcal{B}$ и функции множеств $\mathbf{P}(A|B)$ назовем пространством условной вероятности и обозначим через $[S, \mathcal{A}, \mathcal{B}, \mathbf{P}]$.

Очевидно, что если B фиксированно, то $S, \mathcal{A}$ и $\mathbf{P}(A|B)$ образуют по теории А. Н. Колмогорова пространство вероятности. Поэтому пространство условной вероятности является ни чем иным, как совокупностью обыкновенных пространств вероятности, связанных аксиомой III.

В качестве непосредственного следствия аксиом получается равенство $\mathbf{P}(A|B) = \mathbf{P}(AB|B)$, а поэтому $\mathbf{P}(A|B) \leqq 1$. Раздел 1.4 занимается дальнейшими простыми следствиями аксиом. В разделе 1.5 исследуется при каких условиях условная вероятность $\mathbf{P}(A|B)$ может быть представлена в виде $\mathbf{P}(A|B) = \dfrac{\mathbf{Q}(AB)}{\mathbf{Q}(B)}$, где $\mathbf{Q}$ мера (не обязательно ограниченная) на σ-алгебре $\mathcal{A}$ и $\mathbf{Q}(B) > 0$ для $B \in \mathcal{B}$. Исчерпывающее исследование этого вопроса, а также следующего более общего вопроса: в каком случае условная вероятность $\mathbf{P}(A|B)$ может быть представлена в виде $\mathbf{P}(A|B) = \dfrac{\mathbf{Q}_\beta(AB)}{\mathbf{Q}_\beta(B)}$, где мера $\mathbf{Q}_\beta$ выбирается из некоторого множества мер в зависимости от B, можно найти в работе А. Часара [13]. В разделе 1.6 исследуются случайные величины, определенные в пространстве условной вероятности, и содержит определение их функции распределения и функции плотности.

Функцию $\xi = \xi(s)$ $(s \in S)$ мы называем случайной величиной, если она измерима относительно $\mathcal{A}$. Монотонно неубывающую и слева непрерывную функцию $F(x)$ мы называем (обобщенной) функцией распределения случайной величины ξ, если множество, где $a \leqq \xi(s) < b$, принадлежит $\mathcal{B}$, предполагая, что $F(b) - F(a) > 0$, и в этом случае

$$\mathbf{P}(c \leqq \xi < d \,|\, a \leqq \xi < b) = \frac{F(d) - F(c)}{F(b) - F(a)}, \text{ если } a \leqq c < d \leqq b.$$

Функция распределения $F(x)$ не определена однозначно, потому что вместе с $F(x)$ определению удовлетворяет и $\lambda F(x) + \mu$, где $\lambda > 0$ и μ вещественно; функция распределения $F(x)$ может принимать любые вещественные (значит и отрицательные) значения. Если $F(x)$ абсолютно непрерывна, функцию $f(x) = F'(x)$ назовем обобщенной функцией плотности случайной величины ξ; функция плотности определена лишь с точностью до положительного постоянного сомножителя и $\int_{-\infty}^{+\infty} f(x)\,dx$ не обязательно существует. Если, в частности, $f(x) \equiv 1$ $(-\infty < x < +\infty)$, то мы говорим, что случайная величина ξ равномерно распределена на всей вещественной оси. Раздел 1.7 занимается одной переформулировкой аксиомы III. В разделе 1.8 рассматривается расширение пространств условной вероятности, в 1.9 — непрерывность условной вероятности, в 1.10 — определение произведения пространств условной вероятности. Разделы 2.1 и

2. 2 содержат примеры построения пространств условной вероятности, в разделах 2. 3 и 2. 4 изучаются пространства условной вероятности, обладающие некоторыми дополнительными свойствами: так называемые пространства Кавальери и регулярные пространства. Раздел 2. 5 рассматривает парадокс Бореля, 2. 6 — распределение суммы независимых случайных величин и композицию введенных в разделе 1. 6 обобщенных функций распределений. В § 3 на нескольких примерах показывается, как новая теория приводит к открытию некоторых новых соотношений, которые в рамках обычной теории вероятности не могут быть сформулированны. Раздел 3. 6 содержит обобщение метода Бэйса. Раздел 3. 7 содержит определение условной эргодичности относительно цепей Маркова. В разделе 3. 8 рассматривается один парадокс теории возобновления. Раздел 3. 9 содержит простое и исходящее из более простых чем обычно предположений доказательство закона Максвелла о распределении скоростей, используя данное в разделе 3. 6 обобщение метода Бэйса. В § 4 речь идет об условных законах больших чисел, относящихся к сходимости с вероятностью 1 условного среднего наблюдений. В качестве приложения раздел 4. 3 содержит обобщение теоремы Бореля, относящейся к нормальным разложениям в десятичные дроби, на ряды Кантора.