# AXIOMS FOR CONDITIONAL PROBABILITY SPACES

by

Miklos Csorgo

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfilment of the requirements for the degree of Master of Arts.

Department of Mathematics,
McGill University,
Montreal.

April, 1961

I am greatly indebted to Professor W. A. D'N. Waugh for his help and criticism.

CONTENTS

# CHAPTER I

## INTRODUCTION

## FOUNDATIONS OF PROBABILITY

The calculus of probability is a branch of mathematics. Its foundations have so far not been fully investigated. There are, perhaps, many such branches of mathematics, but the calculus of probability is unique among them regarding the specific course of the development of its fundamental principles. This is bound up with what H. Steinhaus calls the "tavern" origin of probability. A theory of gambling games at first, it gradually extended its range of applicability, becoming finally a mathematical theory of great practical and theoretical importance.

It was at a very early stage of development of the calculus of probability that mathematicians felt the need of formulating its foundations more precisely. The first attempt in this direction was probably the definition of "classical probability" given by Laplace. However, it was the introduction of axiomatic methods, which made it possible to investigate the principles of probability along new lines.

The first axiomatic treatment of probability was given by Bohlmann [ 2 ] about the year 1904. This viewpoint developed in the twentieth century at the hands of such great probabilists as E. Borel, H. Steinhaus, P. Lévy and A. Kolmogorov. The first systematic

presentation of probability theory on an axiomatic basis was made in 1933 by A. Kolmogorov [13].

In principle it is the aim of every axiomatic theory of probability to answer the following two questions:

1) What are events, i.e. what are those objects supposed to be probable?

2) What kind of a function of events should probability be?

Experience shows that, in answering these two questions, certain parts of algebra (especially the theory of Boolean algebras) and certain parts of the theory of functions (measure theory) control the foundations of probability to such an extent that they almost absorb them. J. Łoś calls this (see [18]) a useful process of complete mathematization of the calculus of probability.

We proceed now to give a brief resumé of developments in the axiomatic foundations of probability theory from the logical point of view. All historical and philosophical problems connected with the subject are omitted.

1. What should an axiomatic foundation for the theory of probability be like? Suppose we are to choose a point at random from the interval $[0,1]$. What is the probability of this point belonging to a given set? Such a problem leads at once to the consideration of

Lebesgue measure in the interval $[0, 1]$ and to the consideration of a field of measurable sets. Let B be a subinterval of $[0, 1]$ It is natural to take the probability of any point belonging to this interval as the length of the interval. Let this length of B be denoted by P(B). If more generally the set B is a Borel subset of $[0, 1]$ the preceding choice leads us to take as the value of P(B) the Lebesgue measure of B. We are given then the Lebesgue measurable Borel sets of $[0, 1]$ to determine P(B) for any subinterval B of $[0, 1]$ Thus, in general, we have:

(a) a certain set S (e.g. interval $[0, 1]$ ),

(b) a certain algebra $\mathcal{A}$ of subsets of S (measurable sets)

(c) a certain measure P on $\mathcal{A}$ (e.g. Lebesgue measure).

The occurrence of the triplet $[S, \mathcal{A}, P]$ is observed in nearly all problems of probability theory and none of its elements is superfluous.

($a_o$) The set S, called the set of elementary events, is necessary to define random variables as real functions on S measurable with respect to the algebra $\mathcal{A}$.

($b_o$) The algebra $\mathcal{A}$, called the algebra (field) of events, is the set of those objects which are supposed to be probable. In this class of sets the set - theoretical operations correspond to the classical operations on events.

($C_o$) The measure P on $\mathcal{A}$ is the probability attributed to the events of $\mathcal{A}$.

There are certainly many such triplets $[\, S, \mathcal{A}, P \,]$ to which the probabilistic reasoning may refer. The aim of the axioms of probability is to select that class of them which is essential for probabilistic problems, i.e. triplets (which will be called probability spaces) with which the calculus of probability is concerned.

While selecting probability spaces from the triplets $[\, S, \mathcal{A}, P \,]$ we must proceed as follows:

(I.1) Every such triplet that appears in the problems of probability theory must be a probability space.

(I.2) The fundamental notions of probability theory should be definable for every probability space (e.g. notions like random variable, stochastic independence, mathematical expectation, etc.) and the fundamental theorems of probability theory, for instance the laws of large numbers, should be provable.

2. <u>First interpretation of Kolmogorov's Axiomatic Foundation of Probability Theory.</u> A.N. Kolmogorov published his "Grundbegriffe der Wahrscheinlichkeitsrechnung" in 1933, in which he gave not only an axiomatic foundation of probability theory but also showed how it satisfies the postulates (I.1) and (I.2).

There is no need to emphasize the decisive meaning of that work since, in order to avoid redundancy, we already adopted

Kolmogorov's standpoint in Chapter I above. When reading older textbooks and papers which deal with probability theory it becomes clear that Kolmogorov's work has indeed given mathematical foundations to this branch of knowledge. This has been achieved by an exact formulation of assumptions, a precise definition of notions and by establishment of the close connection of the calculus of probability with other mathematical theories, namely the theory of measure and integration which were already fully developed in those days by Lebesgue and Fréchet.

Kolmogorov's set of axioms demand that probability should be a normed measure (i.e. a non-negative and additive set function normed by the condition $P(S) = 1$) on an algebra $\mathcal{A}$ of subsets of S satisfying the axiom of continuity, i.e. for a decreasing sequence of events.

$$A_1 \supseteq A_2 \supseteq \ldots A_n \supseteq \ldots$$

of $\mathcal{A}$, for which $\bigcap_{n=1}^{\infty} A_n = O$, the following equation holds: $\lim P(A_n) = O$ as $n \longrightarrow \infty$. This is equivalent to the condition of denumerable additivity.

It is a consequence of the axiom of continuity that there exists a unique extension of the measure P to the denumerably additive measure on the smallest denumerably additive class of sets which contains the algebra of sets $\mathcal{A}$.

Therefore, we can always assume that $\mathcal{A}$ is already a denumerably additive class of sets ($\mathcal{C}$ -algebra) and that P is a denumerably additive measure ($\mathcal{C}$ -additive measure) on $\mathcal{A}$.

We are not going to discuss in detail how different probabilistic notions are defined on the basis of Kolmogorov's axioms. It is suffice to note that every real function X on S measurable with respect to the $\mathcal{C}$ -algebra $\mathcal{A}$ is called a random variable and the integral of X on S with respect to the measure P is the expected value of the random variable X.

From the intuitive point of view the essence of Kolmogorov's axiomatic theory is that only one kind of events is examined, namely those events which can be described as consisting of random points (elementary events) a which belong to a subset A of S ($a \in A \subseteq S$). It appears that such an interpretation is always possible.

3. Objections against Kolmogorov's axioms of probability Theory. Kolmogorov's axiomatic treatment of probability theory is accepted by the great majority of mathematicians of today. It has also been criticized and Kolmogorov has done this himself [14].

The first objection concerns the representation of every event in the form $a \in A$, which may be considered as an impoverishment of the formalism of probability theory, or at least of its intuitive side

and as a deviation from its tradition.

The second objection points out that his axioms do not admit the identification of almost identical events (i.e. events which are such that their symmetric difference is of measure O), or, which in fact is exactly the same, that in most of the cases considered, it does not permit the introduction of a strictly positive measure (i.e. a measure which is equal to zero only on the empty set).

The third objection points out that there are certain problems which give rise to probability spaces which cannot be normed; i.e. unbounded measures occur in them, while in the theory of Kolmogorov probability is a bounded measure normed by the condition $P(S) = 1$.

The first two objections are handled by introducing Boolean algebras in the axiomatic treatment of probability theory and by proving their isomorphism to an algebra of sets. This induced Kolmogorov [14] to suggest a somewhat different attitude towards the foundations of probability theory which is developed in paragraph 6 of this chapter.

The third objection was raised by A. Rényi in his paper: "On a New Axiomatic Theory of Probability" [22] where he presented a new set of axioms which use conditional probability as the fundamental concept.

Boolean algebras and their isomorphism to an algebra of sets, probability in Boolean algebras and Kolmogorov's suggestions in this regard will be touched only briefly here. In regard to A. Renyi's work, the whole thesis is devoted to presenting his theory in detail.

4. Boolean algebras and fields of sets. By a Boolean algebra we understand a class of objects furnished with operations governed by the same laws as the operations on sets. The notion of Boolean algebra is of essential importance for the foundations of probability. A set of events is a Boolean algebra, i.e. there is a correspondence between the operations of the Boolean algebra and operations among events. For the latter we refer to the first chapter of W. Feller's book. [ 5 ]. From the definition of Boolean algebras it follows that an algebra of sets is its particular case. M. H. Stone [26] has proved that the converse also holds and so we have: every Boolean algebra is isomorphic to an algebra of sets. Thus denoting by $\phi$ the impossible event, by C the certain event and any other events by the capital letters A, B, D, ..., we are given then a system of elements $\phi$ , A, B, D, ...., C and operations on them governed by the same laws as the operations on sets, $\phi$ playing the role of the empty set and C that of the entire space. Such a system is referred to as a Bodlean algebra and denoted by $\mathfrak{B}$.

Stone's construction of an algebra of sets $\mathcal{A}$

isomorphic to a given Boolean algebra $\mathfrak{B}$ consists of the subsets of the space S the elements (points) of which are the prime ideals of the Boolean algebra $\mathfrak{B}$.

Algebraically a prime ideal is defined as such a set of events $\mathfrak{J} \subset \mathfrak{B}$ that fulfills the following three conditions:

(4.1) $X \in \mathfrak{J}$, $Y \in \mathfrak{J}$ implies $X \cdot Y \in \mathfrak{J}$,

(4.2) $X \in \mathfrak{J}$, $Y \in \mathfrak{B}$ implies $X + Y \in \mathfrak{J}$

(4.3) from two complementary events $X$ and $\overline{X}$ one and only one belongs to $\mathfrak{J}$.

In the above conditions the event $X \cdot Y$ occurs if and only if both X and Y occur; the event X+Y occurs if and only if at least one of them occurs and finally the event $\overline{X}$ occurs if and only if the event X does not occur.

Suppose we carry out some trials on the occurrence of a certain physical phenomenon. Each trial in each case gives actually a set of events, $\mathfrak{J}$, which have occurred in the given case. It is never one event, because from the occurrence of an event X certainly follows the occurrence of the event X + Y (Y is an arbitrary event) generally different from X (e.g. rolling a die we get the number 1; it also means that the event "1 or 3" has also been realized) and so X + Y is also in $\mathfrak{J}$. The occurrence of the event X may also mean the occurrence of $X \cdot Y$. To continue with our previous example, when rolling a die the realization of the event "1" also means

the realization of the event "odd", i.e. $X \cdot Y$ has also occurred and so $X \cdot Y \in \mathfrak{J}$. If $X$ has occurred then $\overline{X}$ did not occur in a certain trial of an experiment. Therefore each trial gives rise to a set of events and this set of events is a prime ideal. This allows the prime ideals to be considered as elementary events.

The isomorphism, constructed by Stone, maps the Boolean algebra of events $\mathfrak{B}$ on an algebra of sets $\mathfrak{a}$ situated in the space S of all prime ideals of $\mathfrak{B}$. This mapping makes the event $X \in \mathfrak{B}$ correspond to the set $\varphi(X)$ of all those prime ideals $\mathfrak{J}$ to which X belongs. If we consider a prime ideal as a result of a trial, that is, in the role of an elementary event, then $\varphi(X)$ is the set of all those trials in which X occurs. The mapping thus defined proves to be an isomorphism; the algebra $\mathfrak{a}$ of all sets $\varphi(X)$ is isomorphic to the Boolean algebra $\mathfrak{B}$, i.e. $\mathfrak{a}$ consists of all sets $\varphi(X)$ and $\varphi(X)$ consists of all those prime ideals $\mathfrak{J}$ of $\mathfrak{B}$ to which X belongs. The points of the set $\varphi(X)$ are prime ideals of $\mathfrak{B}$ and, therefore, if we consider the space S of all the prime ideals of $\mathfrak{B}$ and accept the sets of $\mathfrak{a}$ as neighbourhoods in S, then S becomes a compact topological space and $\mathfrak{a}$ consists of all the subsets of S and is isomorphic to the Boolean algebra $\mathfrak{B}$.

By a Boolean $\sigma$-algebra we mean a Boolean algebra which, besides the operations discussed, is furnished with the operations of denumerable addition and multiplication. Similarly

to the finite operations, the infinite operations are also governed by the same laws as the infinite operations on sets. Two important things must be kept in mind:

1) Not every Boolean algebra is a Boolean $\sigma$-algebra; if we assume that B is one (i.e. if we assume countable additivity) then it is an essential restriction.

2) A Boolean $\sigma$-algebra need not be isomorphic to a $\sigma$-algebra of sets.

However, Loomis and Sikorski [17,23] have shown that each Boolean $\sigma$-algebra is isomorphic to a quotient $\sigma$-algebra of subsets of some space S, i.e. a $\sigma$-algebra of sets divided by a $\sigma$-ideal.

Here we shall briefly explain the operation of dividing an algebra by an ideal, in particular a $\sigma$-algebra by a $\sigma$-ideal.

A subset $J$ of a given Boolean algebra $\mathfrak{B}$ is called an ideal if it satisfies the conditions (4.1) and (4.2) (if it also satisfies the condition (4.3) it is called a prime ideal). For instance the set of events of probability one is an ideal. As we shall see, in this interpretation the construction of quotient algebra has a clear probabilistic meaning. In fact, this is the idea used by Kolmogorov in [14] to which we shall return later.

To continue the algebraic discussion of dividing

an algebra by an ideal, let $\mathfrak{J}$ be an ideal of the Boolean algebra of events $\mathfrak{B}$. In the quotient algebra $\mathfrak{B} \mid \mathfrak{J}$ two events $X, Y \in \mathfrak{B}$ whose simultaneous occurrence or non-occurrence is certain (i.e. the event $X \cdot Y$ belongs to $\mathfrak{J}$) are treated as identical.

An ideal $\mathfrak{J}$ is a $\sigma$-ideal (denumerably multiplicative ideal) if it satisfies in addition to (4.1) and (4.2) the following condition:

(4.4) if the events $X_1, X_2, X_3, \ldots$ belong to $\mathfrak{J}$ then the product $\prod_{n=1}^{\infty} X_n$ also belongs to $\mathfrak{J}$.

If $\mathfrak{J}$ is a $\sigma$-ideal of a Boolean $\sigma$-algebra $\mathfrak{B}$, then the quotient algebra $\mathfrak{B} \mid \mathfrak{J}$ is also a Boolean $\sigma$-algebra.

5. Probability in Boolean algebras. The first two objections against Kolmogorov's axiomatic treatment of probability, mentioned above in 3., can be handled by omitting the assumption that events supposed to be probable are sets and assuming only that they form a Boolean algebra. Such an attitude towards probability has been suggested by Glivenko [6] and Halmos [9]. In this sense the mathematical theory of probability consists of the study of Boolean $\sigma$-algebras and numerical probability is a measure function, that is a finite, non-negative, and countably additive function

P of elements in a Boolean $\sigma$ -algebra $\mathcal{B}$ , such that if the null and unit elements of $\mathcal{B}$ are $\phi$ and C respectively then P(A) = O is equivalent to A = $\phi$ and P(A) = 1 is equivalent to A = C. This is a return to the classical traditions, according to which events need not be sets and this also allows the introduction of a strictly positive measure; i.e. that P(A) = O if and only if A = $\phi$.

However, such an attitude deprives the probability fields of one element, namely of the space S. In a probability field we shall now have only $\mathcal{B}$ , the Boolean $\sigma$ -algebra and a measure P, in place of the triplet [ S, $\mathcal{a}$ , P] . This causes difficulties in defining many probabilistic notions and, in the first place, in defining random variables and their expected values.

Attempts have been made to eliminate this difficulty. They all reduce the notion of a random variable to the notion of a $\sigma$ -homomorphism of a field of Borel sets of the real axis into a Boolean $\sigma$ -algebra. A unification of these attempts has been developed by Sikorski [ 24 ] .

Let $\mathcal{B}$ be an algebra of Borel sets situated on the real axis. The mapping h of $\mathcal{B}$ into a Boolean algebra $\mathcal{B}$ is called a homomorphism if, for $A_1$ , $A_2 \in \mathcal{B}$

$$h(A_1 + A_2) = h(A_1)+h(A_2),$$

$$h(A_1 \cdot A_2) = h(A_1) \cdot h(A_2),$$

$$h(\bar{A}_1) = \overline{h(A_1)}$$

Moreover, if $\mathcal{B}$ is a Boolean $\sigma$-algebra and

$$h\left(\sum_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} h(A_i)$$

for any $A_1, A_2, \ldots \in \mathcal{B}$, then h is called a $\sigma$-homomorphism (a denumerably additive homomorphism).

In order to show how a homomorphism of an algebra of Borel sets into an algebra of events may, for probabilistic purposes, replace a random variable, we shall consider a real function f on the space S measurable with respect to a certain $\sigma$-algebra $\mathcal{A}$ of subsets of S. $\mathcal{A}$ is regarded as an algebra of events, f is regarded as a random variable; i.e. we are given a probability space $[\, S, \mathcal{A}, P \,]$, where P is a denumerably additive measure on $\mathcal{A}$ and a random variable f on S defined above.

Let A be an arbitrary Borel set on the real axis. Let h be a $\sigma$-homomorphism of the algebra $\mathcal{B}$ of Borel sets situated on the real axis into $\mathcal{A}$. Define h as

$$h(A) = \left\{ \text{all } x \in S \mid f(x) \in A \right\} = f^{-1}(A) \in \mathcal{A}$$

This homomorphism is closely connected with the distribution function of f. If A is the interval $[-\infty, a]$ and a $\sigma$ -additive measure P is given on $\mathfrak{A}$ , then putting $F_f(a) = P(h(A))$ we obtain the distribution function of f with respect to the measure P.

In order to define Lebesgue's integral of the function f with respect to the measure P, it is not necessary to know the function f itself, it suffices to know its distribution function or the homomorphism h. This allows us to replace the notion of a measurable function by the notion of homomorphism in the foundations of probability constructed on Boolean algebras provided the algebra in question is a $\sigma$ -algebra.

We thus obtain an equivalent to the usual description of random variables, in the form of these homomorphisms. If we have a homomorphism h of the algebra $\mathfrak{B}$ of Borel sets into a Boolean $\sigma$ -algebra $\mathfrak{B}$ furnished with a $\sigma$ -additive measure P, then taking the integral in the Lebesgue sense, the integral of the homorphism h can be defined, which acts as the expected value and has all the properties generally associated with the expected value.

6. Making use of the connections of Boolean algebras with algebras of sets. Kolmogorov's second interpretation.

We have seen above that the introduction of Boolean algebras into probability theory causes difficulties in defining the random variable and its expected value. The attitude described at the end of the former paragraph is possible, but the same results may be obtained by making use of the connections between Boolean algebras and algebras of sets. The treatment by means of these connections is based on the fact that each Boolean $\sigma$ -algebra is isomorphic to a quotient $\sigma$ -algebra of sets proved by Loomis and Sikorski.

Using this fact Halmos [ 9 ] proposed the following construction. If $\mathcal{B}$ is any Boolean $\sigma$ -algebra and P is a probability measure on $\mathcal{B}$ , then there exists a measure space S such that the system $\mathcal{B}$ is isomorphic to a $\sigma$ -algebra $\mathcal{A}$ of subsets of S reduced by identification according to sets of measure zero, and the value of P for any event A is identical with the value of the measure for the corresponding subset of S.

Reduction by identification according to sets of measure zero is meant as follows. P(A) = O should appear if and only if A = $\phi$ . It is proposed that we agree to consider as identical two sets of S whose symmetrical difference

has probability zero. Through this agreement we are committed in particular to identifying any set of probability zero with the empty set $\phi$, and it follows therefore that in the reduced $\sigma$-algebra of sets $\mathfrak{A}$ all the axioms of probability, which were introduced at the beginning of paragraph 5 on Boolean $\sigma$-algebras, are valid. We are also given back the space S of which we were deprived when it was supposed only that events form a Boolean $\sigma$-algebra.

The above construction culminates therefore in the same set of axioms as Kolmogorov's set of axioms, when starting with a $\sigma$-algebra of sets $\mathfrak{A}$ and with a $\sigma$-additive measure P on $\mathfrak{A}$, but starting with a Boolean $\sigma$-algebra isomorphic to a $\sigma$-algebra $\mathfrak{A}$ of subsets reduced by the suggested identification a strictly positive $\sigma$-additive measure is introduced on events which need not be looked upon as sets. It seems then that the first two objections, mentioned in paragraph 3., were taken care of.

But making use of the connections of Boolean algebras with algebras of sets more than that can be achieved. In the above discussions we always started with a Boolean $\sigma$-algebra furnished with a $\sigma$-additive measure when

making use of their isomorphism to algebras of sets to get back the space S, or, when the notion of $\sigma$- homomorphism was introduced to handle random variable problems, though it is remarked in paragraph 4. above that not every Boolean algebra is a Boolean $\sigma$-algebra; if we assume that $\mathcal{B}$ is one (i.e. if we assume countable additivity) then it is an essential restriction. In fact, we do not need a Boolean $\sigma$-algebra to start with.

We shall begin here with the description given by Kolmogorov [14]. He remarked, above all, that it was easier to apply probability on Boolean algebras as it allows us to assume that probability is a strictly positive measure.

He then remarked that in the case of Boolean algebras we did not have to assume the denumerable additivity of measure or the denumerable additivity of the algebra because for every Boolean algebra $\mathcal{B}_0$ with a strictly positive measure $P_0$ there exists a unique (with an exactitude to the isomorphism) $\sigma$-algebra $\mathcal{B}$ with strictly positive $\sigma$-additive measure P such that

$\mathfrak{B}$ is an extension of $\mathfrak{B}_0$,

P is an extension of $P_0$,

$\mathfrak{B}$ itself is the least $\sigma$-algebra of $\mathfrak{B}_0$ which contains $\mathfrak{B}_0$.

For a given algebra $\mathfrak{B}_0$ with a strictly positive measure $P_0$ the algebra $\mathfrak{B}$ and the measure P are constructed as follows:

As can be seen from Stone's construction, the algebra $B_0$ is isomorphic to an algebra of both closed and open sets $\mathcal{A}$ of a certain compact space S. By means of this isomorphism the measure $P_0$ can be transferred to $\mathcal{A}$. The measure $P_0$ in $\mathcal{A}$ satisfies the condition of continuity which follows from the compactness of the space S, and therefore it can be extended to a $\sigma$-additive measure on the least $\sigma$-algebra of sets $\mathcal{A}^*$ which includes $\mathcal{A}$. The measure in $\mathcal{A}^*$ need not be strictly positive, whereas dividing $\mathcal{A}^*$ by the ideal of sets whose symmetric difference is of measure zero (i.e. by the ideal of events whose simultaneous occurrence or non-occurrence is certain) we obtain a quotient $\sigma$-algebra of sets isomorphic to a Boolean $\sigma$-algebra $\mathfrak{B}$ and a strictly positive $\sigma$-additive measure P. This not only allows us to omit the condition of denumerable additivity but also gives a convenient foundation for

defining random variables as functions on the space S, the elements of which are the prime ideals in $\mathfrak{B}_0$ whose treatment as elementary events has been justified. Interpreted that way we can always start with the triplet [S, $\mathfrak{A}$ , P] assuming that $\mathfrak{A}$ is already a denumerably additive algebra and that P is a denumerably additive measure on $\mathfrak{A}$ . In what follows we shall always use these triplets in the sense of Kolmogorov's second interpretation.

7. The problem of unbounded measures. Conditional probabilities.

This thesis as a whole is devoted to presenting A. Rényi's axiomatic treatment of probability theory [21, 22] which uses conditional probability as the fundemental concept. A detailed discussion of his work follows in the next 2 chapters. Here, we shall only give a brief account of those ideas which can lead us to think about probability in his terms.

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 fruitful application in a great number of branches of science, including other parts of mathematics too. The second interpretation of his theory given above is free of the first two objections mentioned in paragraph 3 of this chapter. Nevertheless,

in the course of development there arose some problems concerning probability which cannot 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 $P(S) = 1$. Unbounded measures occur in statistical mechanics, in some problems of mathematical statistics, in connection with the applications of probability concepts in number theory etc. In the theory of Kolmogorov, for instance, 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 measurs can play no role in probability theory, because, in view of the connection between probability and relative frequencey, probability clearly cannot take on any value greater than 1. But if we observe how unbounded measures are 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 since

the use of unbounded measures cannot be justified in the theory of Kolmogorov, the necessity arises to generalize this theory. Such an attempt was made by Rényi in his paper [ 22 ].

In a theory of probability in which unbounded measures are to be allowed, which are used to calculate conditional probabilities anyway but their use cannot be justified in the theory of Kolmogorov, we should perhaps take conditional probability as the fundamental concept. Using conditional probability as the fundamental concept is also natural from another point of view, namely, that 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. H. Jeffreys [ 11 ], H. Reichenbach [ 19 ], J. Keynes [ 12 ], R. Koopman [ 15 ], A. Copeland [ 3 ], G.A. Barnard [ 1 ], and I J. Good [ 7 ], are mentioned by A. Rényi. None of these authors developed his theory on a measure theoretic basis.

The axiomatic theory developed by Rényi 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. 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 also includes cases which cannot be fitted into the theory of Kolmogorov, namely cases in which conditional probabilities are calculated by means of unbounded measures. According to information given by B.V. Gnedenko to A. Rényi in Prague in 1954, A.N. Kolmogorov himself has put forward the idea, in a lecture held some years ago in Moscow, to develop his theory in such a manner that conditional probability should be taken as the fundamental concept, but did not publish his ideas regarding this question. According to this information Rényi's attempt follows the lines which have been pointed out by Kolmogorov at that time. Some measure-theoretic problems, which arose in connection with Rényi's work, have been solved by Á. Császár [ 4 ]; his results settle the question under what conditions can the conditional probability, introduced by A. Rényi as a set function of two set variables, be expressed in quotient form by means of (one or more) set functions of one variable.

## CHAPTER II

## AXIOMS FOR CONDITIONAL PROBABILITY SPACES AND THEIR IMMEDIATE CONSEQUENCES

2.1. Notations. In what follows if A and B are sets, we denote by A + B the sum (union) of the sets A and B (i.e. the set of those elements which belong to at least one of the sets A and B; AB denotes the product (intersection) 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 and the product of a finite or infinite family of sets, we also use the notations $\sum$ and $\prod$ respectively. The empty set will be denoted by $\phi$ ; $A \subseteq B$ 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 $a \notin A$.

2.2. Definitions and axioms. Let there be given an arbitrary set S; the elements of S which will be denoted by small letters a, b, ... will be called elementary events. Let $\mathcal{A}$ denote a $\sigma$-algebra of subsets of S; the subsets of S which are elements of $\mathcal{A}$ will be denoted by capital letters A , B, C, ... and called random events, or simply events. ( The supposition that $\mathcal{A}$ is a $\sigma$-algebra of subsets of S means $\mathcal{A}$ is a non-empty class of sets closed under the formation of complements and countable unions;

i.e. 1. if $A_n \in \mathfrak{A}$ (n= 1,2, ...) we have $\sum_{n=1}^{\infty} A_n \in \mathfrak{A}$ ;

2. if $A \in \mathfrak{A}$, we have $S-A \in \mathfrak{A}$; 3. $\mathfrak{A}$ is not empty. This implies that $\phi \in \mathfrak{A}$ and so $S-\phi = S \in \mathfrak{A}$. Therefore a $\sigma$-algebra is a $\sigma$-ring containing S). Let us suppose further that a non empty subset $\mathfrak{B}$ of $\mathfrak{A}$ is given; we do not suppose any restrictions regarding the set $\mathfrak{B}$. (It will be seen that our axioms imply that $\phi \notin \mathfrak{B}$, but it is possible that $\mathfrak{B}$ contains all the elements of $\mathfrak{A}$ except $\phi$ ; it is also possible that $\mathfrak{B}$ contains only one set). We suppose finally that a set function $P(A \mid B)$ of two set variables is defined for $A \in \mathfrak{A}$ and $B \in \mathfrak{B}$ ; $P(A \mid B)$ will be called the conditional probability of the event A with respect to the event B. As the conditional probability of the event $A \in \mathfrak{A}$ with respect to the event B is defined if and only if $B \in \mathfrak{B}$, $\mathfrak{B}$ may be called the class of possible conditions. We suppose that the set function $P(A \mid B)$ satisfies the following axioms:

Axiom I. $P(A \mid B) \geq 0$ if $A \in \mathfrak{A}$ and $B \in \mathfrak{B}$;

further $P(B \mid B) = 1$, if $B \notin \mathfrak{B}$.

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

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

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

$$P(A \mid BC) \cdot P(B \mid C) = P(AB \mid C)$$

If the axioms I-III are satisfied, we shall call the set S, together with the $\sigma$-algebra $\mathcal{A}$ of subsets of S, the subset $\mathcal{B}$ of $\mathcal{A}$ and the set function $P(A \mid B)$ defined for $A \in \mathcal{A}$ , $B \in \mathcal{B}$ , a conditional probability space and denote it for the sake of brevity by

$[S, \mathcal{A}, \mathcal{B}, P(A \mid B)]$

2.3. Immediate consequences of the axioms. In what follows, if $P(A \mid B)$ occurs, it is always tacitly assumed that $A \in \mathcal{A}$ and $B \in \mathcal{B}$ . We denote the set S-A by $\overline{A}$.

Theorem 1. $P(A \mid B) = P(AB \mid B)$

Proof. Put $C = B$ in Axiom III. Then we have $P(A \mid B)\, P(B \mid B) = P(AB \mid B)$ and so $P(A \mid B) = P(AB \mid B)$ since by Axiom I $P(B \mid B) = 1$.

Remark 1. It follows from Theorem 1 that $P(S \mid B) = 1$; namely, by Theorem 1 $P(S \mid B) = P(SB \mid B) = P(B \mid B)$ and thus, by Axiom I, $P(S \mid B) = 1$.

Remark 2. If we have

$$\sum_{k=1}^{n} A_k = S, \text{ all } A_k \in \mathcal{A} \quad \text{and } A_j A_i = \phi \quad \text{if } i \neq j, \text{ i.e.}$$

if $A_1, A_2, \ldots, A_n$ are mutually exclusive and exhaustive events then

$$\sum_{k=1}^{n} P(A_k \mid B) = P\Big(\sum_{k=1}^{n} A_k \mid B\Big) = 1 \text{ for every fixed } B \in \mathcal{B}.$$

Remark 3. $P(A \mid B) = 1$ if $B \subseteq A$. By Theorem 1

$P(A \mid B) = P(AB \mid B) = P(B \mid B) = 1$. Note: From this remark the result of Remark I automatically follows.

Theorem 2. If $B \subseteq B'$, then $P(AB' \mid B) = P(A \mid B)$

Proof. Applying Theorem I twice we get

$P(AB' \mid B) = P(AB'B \mid B) = P(AB \mid B) = P(A \mid B)$.

Theorem 3. $P(A \mid B) \leq 1$.

Proof. According to Axiom II we have

$P(AB \mid B) + P(\bar{A}B \mid B) = P(AB + \bar{A}B \mid B) = P(B \mid B) = 1$. Also

$P(\bar{A}B \mid B) \geq 0$ by Axiom I. It follows then that $P(AB \mid B) = P(A \mid B) \leq 1$.

Remark. By Axiom I and by Theorem 3 we have:

$0 \leq P(A \mid B) \leq 1$.

Theorem 4. $P(\phi \mid B) = 0$

Proof. According to Axiom II $P(A \mid B) =$

$= P(\phi + A \mid B) = P(\phi \mid B) + P(A \mid B)$ and so $P(\phi \mid B) = 0$. Or,

$P(\phi \mid B) = P(\phi + \phi \mid B) = P(\phi \mid B) + P(\phi \mid B) = 2\, P(\phi \mid B)$ and thus again $P(\phi \mid B) = 0$.

Remark. It follows from Theorem 4 that $\phi \notin \mathcal{B}$. If $\phi$ belonged to $\mathcal{B}$, we should have $P(\phi \mid \phi) = 1$ by Axiom I and $P(\phi \mid \phi) = 0$ by Theorem 4; thus the assumption $\phi \in \mathcal{B}$ leads to a contradiction.

Theorem 5. If $AB = \phi$, then $P(A \mid B) = 0$.

Proof. $P(A \mid B) = P(AB \mid B) = P(\phi \mid B) = 0$ by Theorem 4.

Theorem 6. $P(A \mid BC)P(B \mid C) = P(B \mid AC)P(A \mid C)$, if C, BC, AC belong to $\mathcal{B}$.

Proof. Both expressions are equal to $P(AB \mid C)$ by Axiom III and thus to each other.

Theorem 7. If $A \subseteq A' \subseteq B \subseteq B'$, we have

$$P(A \mid B') \leq P(A' \mid B).$$

To prove this Theorem we need the following

Lemma. If $E \subseteq A$ then, for any fixed $B \in \mathcal{B}$, we have $P(A-E \mid B) = P(A \mid B) - P(E \mid B)$.

Proof. We have $A = E + (A-E)$ and E, (A-E) are disjoint. So $P(A \mid B) = P(E+(A-E) \mid B)$

$= P(E \mid B)+P(A-E \mid B)$ by Axiom II.

Now on both sides any $P(\cdot \mid \cdot)$ is such that $0 \leq P(\cdot \mid \cdot) \leq 1$ and therefore $P(A-E \mid B) = P(A \mid B) - P(E \mid B)$

Proof of Theorem 7. We have

$P(A \mid B') = P(AA'B \mid B)$, $AA'B = A$

$= P(AA' \mid BB') P(B \mid B')$, Axiom III

$\leq P(AA' \mid B)$, since $P(B \mid B') \leq 1$ and $BB' = B$

$= P(A' - \bar{A}A' \mid B)$ since $A' - \bar{A}A' = AA'$

$= P(A' \mid B) - P(\bar{A}A' \mid B)$, by $\bar{A}A' \subseteq A'$ and prev. Lemma.

$\leq P(A' \mid B)$, since $P(\bar{A}A' \mid B) \geq 0$ by Axiom I., i.e. we have $P(A \mid B') \leq P(A' \mid B)$

Remark 1. If $A = A'$, we obtain the following special

case of Theorem 7: If $A \subseteq B \subseteq B'$, we have

$$P(A \mid B') \leq P(A \mid B)$$

Remark 2. If $B = B'$, we obtain the following special case of Theorem 7.: If $A \subseteq A'$, we have, without supposing that $A' \subseteq B$, $P(A \mid B) \leq P(A' \mid B)$ and Axiom III is not needed in proof.

Proof of this remark. We have $P(A|B) = P(AA' \mid B), AA' = A$

$= P(A' - \bar{A}A' \mid B)$, $A' - \bar{A}A' = AA'$

$= P(A' \mid B) - P(\bar{A}A' \mid B)$, by $\bar{A}A' \subseteq A'$ and prev. Lemma.

$\leq P(A' \mid B)$, since $P(\bar{A}A' \mid B) \geq 0$ by Axiom I, and we did not use Axiom III.

Theorem 8. If $A_1 + A_2 \subseteq B_1 B_2 \in \mathcal{B}$, further $P(A_2 \mid B_1) P(A_2 \mid B_2) > 0$, we have

$$\frac{P(A_1 \mid B_1)}{P(A_2 \mid B_1)} = \frac{P(A_1 \mid B_2)}{P(A_2 \mid B_2)}$$

Proof. According to Axiom III

(1) $P(A_1 \mid B_1 B_2) P(B_1 \mid B_2) = P(A_1 B_1 \mid B_2) = P(A_1 \mid B_2)$

and similarly

(2) $P(A_2 \mid B_1 B_2) P(B_1 \mid B_2) = P(A_2 B_1 \mid B_2) = P(A_2 \mid B_2)$

since $A_1 B_1 = A_1$

and $A_2 B_1 = A_2$, for by hypothesis $A_1 + A_2 \subseteq B_1 B_2$ and this implies $A_1 \subseteq B_1 B_2$ and so $A_1 \subseteq B_1$

also $A_2 \subseteq B_1 B_2$ and so $A_2 \subseteq B_1$

Dividing (1) by (2) we obtain

$$(3) \qquad \frac{P(A_1 \mid B_1 B_2)}{P(A_2 \mid B_1 B_2)} = \frac{P(A_1 \mid B_2)}{P(A_2 \mid B_2)}$$

Interchanging $B_1$ and $B_2$ in (1) and (2) we obtain

$$(4) \qquad \frac{P(A_1 \mid B_1 B_2)}{P(A_2 \mid B_1 B_2)} = \frac{P(A_1 \mid B_1)}{P(A_2 \mid B_1)}$$

From (3) and (4) we have

$$\frac{P(A_1 \mid B_1)}{P(A_2 \mid B_1)} = \frac{P(A_1 \mid B_2)}{P(A_2 \mid B_2)}$$

Theorem 9. If $C \subseteq B = \sum_{k=1}^{\infty} B_k$ and $AB_jB_kC = \phi$ for $j \neq k$ $(j,k = 1, 2, \ldots)$ then,

$P(A \mid C) = \sum_{k=1}^{\infty} P(A \mid B_kC)P(B_k \mid C)$, where it is supposed that $C \in \mathcal{B}$ and $B_kC \in \mathcal{B}$ $(k = 1, 2, \ldots)$

This theorem corresponds to the total probability rule.

Proof. By Axiom III we have

$$P(A \mid B_kC)P(B_k \mid C) = P(AB_k \mid C).$$

Therefore

$$\begin{aligned}
\sum_{k=1}^{\infty} P(A \mid B_kC)P(B_k \mid C) &= \sum_{k=1}^{\infty} P(AB_k \mid C) \\
&= \sum_{k=1}^{\infty} P(AB_kC \mid C), \text{ by Theorem I} \\
&= P(A \sum_{k=1}^{\infty} B_kC \mid C), \text{ by Axiom II} \\
&= P(ABC \mid C), \text{ by hypothesis} \\
&= P(AC \mid C), \text{ by hypothesis } BC = C \\
&= P(A \mid C), \text{ by Theorem I}
\end{aligned}$$

Note. In proving Theorem 9 we do not suppose that $B \in \mathcal{B}$.

Remark. We mention the following consequences of Theorem 9. Let us suppose $B_k \in \mathcal{B}$ and $B_j B_k = \phi$ for $j \neq k$, $B = \sum_{k=1}^{\infty} B_k$ and $B \in \mathcal{B}$ ; if $P(A \mid B_k) \leq \lambda P(A' \mid B_k)$ for $k = 1, 2, \ldots$ where $\lambda \geq 0$ , we have $P(A \mid B) \leq \lambda P(A' \mid B)$

Proof. We have

$$P(A \mid B_k) \leq \lambda P(A' \mid B_k)$$

so $P(A \mid B_k)P(B_k \mid B) \leq \lambda P(A' \mid B_k)P(B_k \mid B)$

so $\sum_{k=1}^{\infty} P(A \mid B_k)P(B_k \mid B) \leq \lambda \sum_{k=1}^{\infty} P(A' \mid B_k)P(B_k \mid B)$

and applying Theorem 9 on both sides with $C = B$

$$P(A \mid B) = \sum_{k=1}^{\infty} P(A \mid B_k)P(B_k \mid B) \leq \lambda \sum_{k=1}^{\infty} P(A' \mid B_k)P(B_k \mid B) = \lambda P(A' \mid B)$$

Therefore $P(A \mid B) \leq \lambda P(A' \mid B)$.

Specifically a) if $P(A \mid B_k) = \lambda P(A' \mid B_k)$ for $k = 1, 2, \ldots$, we have $P(A \mid B) = \lambda P(A' \mid B)$

b) if $P(A \mid B_k) = \lambda$ for $k = 1, 2, \ldots$, we have

$$P(A \mid B) = \lambda$$

Proof of b) $P(A \mid B) = \sum_{k=1}^{\infty} P(A \mid B_k)P(B_k \mid B)$, by Theorem 9. with $C = B$

$$= \sum_{k=1}^{\infty} \lambda P(B_k \mid B), \text{ if } P(A \mid B_k) = \lambda$$

$$= \lambda \sum_{k=1}^{\infty} P(B_k \mid B)$$

$$= \lambda P\left(\sum_{k=1}^{\infty} B_k \mid B\right), \text{ by Axiom II}$$

$$= \lambda P(B \mid B)$$

$$= \lambda , \text{ by Axiom I .}$$

2.4. Connection with Kolmogorov's Theory. If P(A) is a measure (i.e. a countably additive and non-negative set function) defined on the $\sigma$-algebra $\mathcal{A}$ of subsets of the sets S, if further P(S) = 1, then the triple $[S, \mathcal{A}, P(A)]$ is called a probability space in the sense of Kolmogorov.

Theorem 10. Define $\mathcal{A}^*$ as the set of those sets $B \in \mathcal{A}$ for which $P(B) > 0$ and put $P(A \mid B) = \dfrac{P(AB)}{P(B)}$ for $A \in \mathcal{A}$, $B \in \mathcal{A}^*$. Then $[S, \mathcal{A}, \mathcal{A}^*, P(A \mid B)]$ is a conditional probability space which will be called the conditional probability space generated by the probability space $[S, \mathcal{A}, P(A)]$.

Proof. Axiom I is satisfied: $P(A \mid B) = \dfrac{P(AB)}{P(B)} \geq 0$ since $A \in \mathcal{A}$, $B \in \mathcal{A}^*$ and so $P(B) > 0$; further $P(B \mid B) = 1$ if $B \in \mathcal{A}^*$ since $P(B \mid B) = \dfrac{P(BB)}{P(B)} = \dfrac{P(B)}{P(B)} = 1, P(B) > 0$

Axiom II is satisfied: If $B \in \mathcal{A}^*$ and $A_j A_k = \phi$ for $j \neq k$ $(j, k = 1, 2, \ldots)$ we have $P\left( \sum_{n=1}^{\infty} A_n \,\middle|\, B \right) = \sum_{n=1}^{\infty} P(A_n B)$ for any fixed B.

Showing this:

$$P\left( \sum_{n=1}^{\infty} A_n \mid B \right) = \frac{P\left( \left( \sum_{n=1}^{\infty} A_n \right) B \right)}{P(B)} \quad \text{in } [S, \mathcal{A}, \mathcal{A}^*, P(A \mid B)$$

$$= \frac{P\left( \sum_{n=1}^{\infty} A_n B \right)}{P(B)}$$

$$= \frac{\sum_{n=1}^{\infty} P(A_n B)}{P(B)}, \quad P(A) \text{ being a } \sigma\text{-additive set function}$$

$$= \sum_{n=1}^{\infty} \frac{P(A_n B)}{P(B)}$$

$$= \sum_{n=1}^{\infty} P(A_n \mid B)$$

Axiom III is satisfied. If $A \in \mathcal{A}$, $B \in \mathcal{A}$, $C \in \mathcal{A}^*$, and $BC \in \mathcal{A}^*$, we have $P(A \mid BC)P(B \mid C) = P(AB \mid C)$.

For $P(A \mid BC) = \frac{P(ABC)}{P(BC)}$

$$P(B \mid C) = \frac{P(BC)}{P(C)}$$

and so $P(A \mid BC) \cdot P(B \mid C) = \frac{P(ABC)}{P(C)}$ ;

also $P(AB \mid C) = \frac{P(ABC)}{P(C)}$

and so $P(A \mid BC)\ P(B \mid C) = P(AB \mid C)$

Theorem II. If $[S, \mathcal{A}, \mathcal{B}, P(A \mid B)]$ is a conditional probability space and C is an arbitrary element of $\mathcal{B}$, putting $P_C(A) = P(A \mid C)$, $[S, \mathcal{A}, P_C(A)]$ will be a probability space in the sense of Kolmogorov for $C \in \mathcal{B}$ fixed.

Proof. Axioms I and II imply that $P_C(A)$ is a $\sigma$-additive non-negative set function for which $P_C(S) = 1$ by Remark 1 to Theorem 1.

Remark. $P_C(A) = P(A \mid C)$ for $A \in \mathcal{A}$ with $C \in \mathcal{B}$

fixed as before; i.e. we have $[S, \mathfrak{A}, P_C(A)]$ Define the conditional probability $P_C(A \mid B)$ for $B \in \mathfrak{B}$ for which $P_C(B) > 0$, as usual in the theory of Kolmogorov, by $P_C(A \mid B) = \frac{P_C(AB)}{P_C(B)}$.

If $BC \in \mathfrak{B}$, then we have by Axiom III.

$$P_C(A \mid B) = \frac{P_C(AB)}{P_C(B)} = \frac{P(AB \mid C)}{P(B \mid C)} = P(A \mid BC)$$

i.e. $P_C(A \mid B) = P(A \mid BC)$

Thus a conditional probability space is nothing else than a set of ordinary probability spaces (to each $C \in \mathfrak{B}$ there corresponds a probability space in the sense of Kolmogorov) which are connected with each other by Axiom III. This connection is such that it is in conformity with the usual definition of conditional probability demonstrated in this Remark.

Theorem 12. If $S \in \mathfrak{B}$, then $[S, \mathfrak{A}, P_S(A)]$ is a probability space in the sense of Kolmogorov on putting as before $P_S(A) = P(A \mid S)$ and $P_S(A \mid B) = \frac{P_S(AB)}{P_S(B)}$ if $P_S(B) > 0$, $B \in \mathfrak{B}$

Proof. Theorem 12 is a special case of Theorem 11.

Remark. In the sense of Theorem 10 $[S, \mathfrak{A}, P_S(A)]$ generates the conditional probability space $[S, \mathfrak{A}, \mathfrak{B}_S, P_S(A \mid B)]$, where $\mathfrak{B}_S$ is defined - in the sense of $\mathfrak{A}^*$ of Theorem 10 - as the

set of those sets $B \in \mathcal{A}$ for which $P_s(B) > 0$ , and

$P_s(A \mid B) = \dfrac{P_s(AB)}{P_s(B)}$ for $A \in \mathcal{A}$ , $B \in \mathcal{B}_s$.

It must be mentioned in this case that

$[S, \mathcal{A}, \mathcal{B}, P(A \mid B)]$ may not be identical with the $[S, \mathcal{A}, \mathcal{B}_s, P_s(A \mid B)]$ conditional probability space generated by $[S, \mathcal{A}, P_s(A)]$ because $\mathcal{B}$ may contain sets B for which $P_s(B) = 0$ and at the same time need not contain every set for which $P_s(B) > 0$ , i.e. the class $\mathcal{B}_s$ consisting of all sets $B \in \mathcal{A}$ for which $P_s(B) > 0$ need not be identical with $\mathcal{B}$ . However, if $P_s(A \mid B) = \dfrac{P_s(AB)}{P_s(B)}$ for $A \in \mathcal{A}$, $B \in \mathcal{B}_s$, then by Axiom III and by definition of $P_s(A)$

$$P_s(A \mid B) = \frac{P_s(AB)}{P_s(B)} = \frac{P(AB \mid S)}{P(B \mid S)} = P(A \mid BS) = P(A \mid B),$$

provided that $B \in \mathcal{B}$.

2.5. Conditional Independence of Events. The conditional probability of the event A with respect to the event B is given by $P(A \mid B) = \dfrac{P(AB)}{P(B)}$ for $B \in \mathcal{A}$ and $P(B) > 0$ in the Kolmogorov probability space $[S, \mathcal{A}, P(A)]$. This formula is often used in the form $P(AB) = P(A \mid B)P(B)$; this is the so called theorem on compound probabilities. If $P(A \mid B) = P(A)$ we say that A is stochastically independent or, simply, independent of B. The condition $P(A \mid B) = P(A)$ can be written in the form $P(AB) = P(A)P(B)$ and we use this as a

definition of independence of the events $A \nmid B$.

If we have the conditional probability space $[S, \mathcal{A}, \mathcal{B}, P(A \mid B)]$ then for $A \in \mathcal{A}$, $B \in \mathcal{A}$, $C \in \mathcal{B}$, and $BC \in \mathcal{B}$, we have $P(AB \mid C) = P(A \mid BC) \cdot P(B \mid C)$ (Axiom III). If we let $P_C(AB) = P(AB \mid C)$, $P_C(A \mid B) = P(A \mid BC)$ and $P_C(B) = P(B \mid C)$ then Axiom III reads $P_C(AB) = P_C(A \mid B) \cdot P_C(B)$. Writing Axiom III in this form it is indicated clearly that given the condition C Axiom III corresponds to $P(AB) = P(A \mid B) P(B)$ in $[S, \mathcal{A}, P(A)]$. It is just another indication of the fact that for any given $C \in \mathcal{B}$, putting $P_C(A) = P(A \mid C)$, $[S, \mathcal{A}, P_C(A)]$ will be a probability space in the sense of Kolmogorov. Therefore, for any given $C \in \mathcal{B}$ we can define independence of the events A and B with respect to the events C on the same way as we did it in $[S, \mathcal{A}, P(A)]$. We say that if $P(A \mid BC) = P(A \mid C)$ then A is stochastically independent or, simply, independent of B with respect to the event C. This condition can be written in the form $P(AB \mid C) = P(A \mid C) P(B \mid C)$ on using Axiom III, and we use this as a definition of independence of the events A and B with respect to the event C. This definition of conditional independence of events $A \nmid B$ readily extends to more than two events.

2.6. Representation of the conditional probability as a quotient. Sufficient conditions will be given here under which

the set function $P(A \mid B)$ of two set variables can be represented in "quotient form", i.e. in the form $P(A \mid B) = \frac{Q(AB)}{Q(B)}$ where the set function $Q(A)$ is a measure on $\mathfrak{A}$ and satisfies $Q(B) > 0$ if $B \in \mathfrak{B}$.

Theorem 13. Let $[S, \mathfrak{A}, \mathfrak{B}, P(A \mid B)]$ be a conditional probability space. Let us suppose that there exists a sequence of sets $B_n (n = 0, 1, \ldots)$ all in $\mathfrak{B}$ for which the following properties hold:

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

b) $P(B_0 \mid B_n) > 0$ $(n = 1, 2, \ldots)$,

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

Then there exists a finite measure $Q(C)$ 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 $Q(C)$ has the following properties:

$\alpha$) $Q(B) > 0$ if $B \in \mathfrak{B}$,

$\beta$) $P(A \mid B) = \frac{Q(AB)}{Q(B)}$

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

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

then $Q(C)$ can be defined for all $C \in \mathfrak{A}$ and is a bounded measure

on $\mathcal{A}$ , and thus, putting $P(C) = \frac{Q(C)}{Q(S)}$, we have $P(S) = 1$.

Denoting by $\mathcal{B}^*$ the set of those sets $B \in \mathcal{A}$ for which $P(B) > 0$, if $\mathcal{B}^*$ is not identical with $\mathcal{B}$ , we may extend the definition $P(A \mid B)$ to all $B \in \mathcal{B}^*$ putting

$$P(A \mid B) = \frac{P(AB)}{P(B)} \quad ;$$

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

Proof. First we suppose only that the sequence $B_n$ has the properties a), b) and c).

First we are going to prove that $\mathcal{A}^*$ is a ring and we have $\mathcal{B} \subseteq \mathcal{A}^* \subseteq \mathcal{A}$.

To show $\mathcal{A}^*$ is a ring, let $C_1 \in \mathcal{A}^*$ and $C_2 \in \mathcal{A}^*$. Suppose $C_1 \subseteq B_1$ and $C_2 \subseteq B_2$ (for any $C \in \mathcal{A}^*$ there can be found a $B_n$ such that $C \subseteq B_n$).

But $C_1 \subseteq B_1$ implies $C_1 - C_2 \subseteq B_1 \subseteq B_2$

i.e. $C_1 - C_2 \in \mathcal{A}^*$

Also $C_1 + C_2 \subseteq B_1 + B_2 = B_2$ since by a) $B_1 \subseteq B_2$

i.e. $C_1 + C_2 \in \mathcal{A}^*$;

i.e. $\mathcal{A}^*$ is a ring.

Now, for any $B \in \mathcal{B}$ we have

$B \subseteq B_n$ , by c)

therefore $B \in \mathcal{A}^*$ by definition of $\mathcal{A}^*$

i.e. $\mathcal{B} \subseteq \mathcal{A}^* \subseteq \mathcal{A}$ .

Now, we proceed to prove statements $\alpha$) and $\beta$). Let us consider a set $A \in \mathcal{A}^*$, choose an index n for which $A \subseteq B_n$ and define Q (A) as follows:

(5) $$Q(A) = \frac{P(A \mid B_n)}{P(B_o \mid B_n)} .$$

The value of Q(A) does not depend on the choice of n. For if $A \subseteq B_n$ and $A \subseteq B_m$ where $n < m$, we have by Theorem 8

$$\frac{P(A \mid B_n)}{P(B_o \mid B_n)} = \frac{P(A \mid B_m)}{P(B_o \mid B_m)} ;$$

using this definition (5) of Q(A) we show that if $B \in \mathcal{B}$ , $Q(B) > 0$ , we have

(6) $$P(A \mid B) = \frac{Q(AB)}{Q(B)}$$

This can be shown as follows: if $B \subseteq B_n$ and $P(B \mid B_n) > 0$ , we have by (5) and by Axiom III

$$\frac{Q(AB)}{Q(B)} = \frac{P(AB \mid B_n)}{P(B_o \mid B_n)} \cdot \frac{P(B_o \mid B_n)}{P(B \mid B_n)} = \frac{P(AB \mid B_n)}{P(B \mid B_n)} = P(A \mid BB_n)$$

As $BB_n = B$, by c)

$$\frac{Q(AB)}{Q(B)} = P(A \mid B) \text{ and (6) is proved,}$$

i.e. if $Q(A)$ is a measure for $A \in \mathfrak{A}^*$ then statements $\alpha)$ and $\beta)$ are proved. From (5) it can be seen that $Q(A)$ is non-negative since by b) $P(B_o \mid B_n) > 0$. Also $Q(B) > 0$ if $B \in \mathfrak{B}$ by $\mathfrak{B} \subseteq \mathfrak{A}^* \subseteq \mathfrak{A}$. Therefore statement $\alpha)$ is proved. Since $Q(A)$ is non-negative, to show that $Q(A)$ is a measure, we are left only to prove that $Q(A)$ is countably additive on $\mathfrak{A}^*$; i.e. if $A_k \in \mathfrak{A}^*$ $(k = 1, 2, \ldots)$ and $A_j A_k = \phi$ for $j \neq k$ and $\sum_{k=1}^{\infty} A_k = A \in \mathfrak{A}^*$, then $Q(A) = \sum_{k=1}^{\infty} Q(A_k)$.

This follows simply from the remark that if $A \subseteq B_n$, we have $A_k \subseteq B_n$ for $k = 1, 2, \ldots$ and thus in the relations

$$Q(A_k) = \frac{P(A_k \mid B_n)}{P(B_o \mid B_n)}, \quad (k = 1, 2, \ldots); \quad Q(A) = \frac{P(A \mid B_n)}{P(B_o \mid B_n)}$$

the same $B_n$ may be used, and therefore the countable additivity of $Q(A)$ follows from that of $P(A \mid B)$ for fixed $B \in \mathfrak{B}$ (Axiom II.). This proves the first part of Theorem 13.

Now suppose that the sequence $B_n$ has also the property d). We need here Theorem A. of section 13 of Halmos, Measure Theory, which goes as follows: If $\mu$ is a $\sigma$-finite measure on a ring $R$, then there is a unique measure $\bar{\mu}$ on the $\sigma$-ring $S(R)$ such that, for $E$ in $R$, $\bar{\mu}(E) = \mu(E)$; the measure $\bar{\mu}$ is $\sigma$-finite ($S(R)$ is the smallest $\sigma$-ring

containing R, generated by R).

By this Theorem the definition of Q(A) can be extended to the smallest $\sigma$-ring $\mathcal{A}^{**}$ containing $\mathcal{A}^{*}$, generated by $\mathcal{A}^{*}$ in such a manner that Q(A) remains countably additive on $\mathcal{A}^{**}$.

Put $S^{*} = \sum_{n=0}^{\infty} B_n$ .

We will show that $\mathcal{A}^{**} = \mathcal{A} S^{*}$ , i.e. $\mathcal{A}^{**}$ is identical with the set of all sets of the form $AS^{*}$ where $A \in \mathcal{A}$ .

First of all $\mathcal{A} S^{*}$ is a $\sigma$-ring since $\mathcal{A}$ is closed under complementation and on taking countable unions of the elements of $\mathcal{A}$ . As a matter of fact if $A_1 \in \mathcal{A}$ and $A_2 \in \mathcal{A}$ then

$A_1 S^{*} - A_2 S^{*} = S^{*}(A_1 - A_2) \in \mathcal{A} S^{*}$, since $A_1 - A_2 \in \mathcal{A}$ and $\sum_{i=1}^{\infty} A_i S^{*} = S^{*} \sum_{i=1}^{\infty} A_i \in \mathcal{A} S^{*}$, since $\sum_{i=1}^{\infty} A_i \in \mathcal{A}$ if $A_i \in \mathcal{A}$ , $i = 1, 2, \ldots$.

Therefore $\mathcal{A} S^{*}$ is a $\sigma$-ring.

To show $\mathcal{A}^{**} \subseteq \mathcal{A} S^{*}$ we have: if $A \in \mathcal{A}^{*}$ then $A \subseteq B_n$ and so $A \subseteq S^{*} = \sum_{n=0}^{\infty} B_n$. Also A is in $\mathcal{A}$ . Therefore $A \in \mathcal{A} S^{*}$ which implies

$\mathcal{A}^{*} \subseteq \mathcal{A} S^{*}$ and, as it was just shown, $\mathcal{A} S^{*}$ is a $\sigma$-ring. We have so far $\mathcal{A} S^{*}$ as a $\sigma$-ring containing

$\mathcal{A}^*$. But $\mathcal{A}^{**}$ being the smallest $\sigma$-ring containing $\mathcal{A}^*$, we have

(a) $$\mathcal{A}^{**} \subseteq \mathcal{A}S^*$$ .

On the other hand, if $A \in \mathcal{A}$, we have

$$AS^* = \sum_{n=0} AB_n \in \mathcal{A}S^*.$$

Now $AB_n \subseteq B_n$, and thus $AB_n \in \mathcal{A}^*$;

i.e. $AS^* \in \mathcal{A}^{**}$, $\mathcal{A}^{**}$ being the smallest $\sigma$-ring containing $\mathcal{A}^*$; and therefore we have

(b) $$\mathcal{A}S^* \subseteq \mathcal{A}^{**}$$

By (a) and (b) we have

$\mathcal{A}S^* = \mathcal{A}^{**}$; i.e. $\mathcal{A}^{**}$, the smallest $\sigma$-ring containing $\mathcal{A}^*$, is the set of all sets of the form $AS^*$ where $A \in \mathcal{A}$ . Thus the definition of $Q(A)$ can be extended to all $A \in \mathcal{A}S^*$ by the quoted Theorem A in Halmos, Measure Theory and because of the identity just proved.

We prove now that $Q(A)$ is bounded on $\mathcal{A}S^*$. To show this it is sufficient to prove that $Q(S^*)$ is finite since $S^* = \sum_{n=0}^{\infty} B_n$ is the set which take s part in the limiting process as $n \longrightarrow \infty$ . But $S^* = \lim_{n\to\infty} B_n$ since $B_n \subseteq B_{n+1}$, and thus $Q(S^*) = \lim_{n\to\infty} Q(B_n)$ where $Q(B_n)$ is non-decreasing. But $Q(B_n) = \dfrac{P(B_n \mid B_n)}{P(B_o \mid B_n)} = \dfrac{1}{P(B_o \mid B_n)}$

by (5) and by Axiom I, and $\lim_{n\to\infty} \mathbf{P}(B_o \mid B_n) > 0$ by property d). Therefore d) implies that $Q(S^*) < +\infty$. Defining $Q(A)$ by $Q(A) = Q(AS^*)$ for $A \in \mathfrak{A}$, $A \notin \mathfrak{A}S^*$, the definition of $Q(A)$ is extended to the whole $\sigma$-algebra $\mathfrak{A}$. The final part of Theorem I3 concerning $\mathbf{P}(A)$ is obvious. Thus Theorem 13 is proved.

2.7. Random variables on a conditional probability space. Let $[S, \mathfrak{A}, \mathfrak{B}, P(A \mid B)]$ 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 $\xi$ a random variable on $[S, \mathfrak{A}, \mathfrak{B}, \mathbf{P}(A \mid B)]$. Vector-valued random variables are defined similarly. The (ordinary) conditional probability distribution function of a random variable $\xi$ with respect to an event $B \in \mathfrak{B}$ is defined by $F(x \mid B) = P(A_x \mid B)$; if $F(x \mid B)$ is absolutely continuous, $F'(x \mid B) = f(x \mid B)$ is called the (ordinary) conditional probability density function of $\xi$ with respect to B. The conditional mean value $M(\xi \mid B)$ of $\xi$ with respect to an event $B \in \mathfrak{B}$ is defined as the abstract Lebesgue integral:

$$M(\xi \mid B) = \int_S \xi(a)\ d\,P(A_x \mid B)$$

of $\xi$ with respect to the measure defined on S by $P(A \mid B)$ with B fixed. Higher conditional moments, the conditional characteristic function etc. are defined similarly. The random variables $\xi$ and $\eta$ 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 $P(A_x B_y \mid C) = P(A_x \mid C) P(B_y \mid C)$ for every real $x$ and $y$.

As $[S, \mathfrak{A}, P(A \mid B)]$ is for any fixed $B \in \mathcal{B}$ 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 $B \in \mathcal{B}$.

Let us mention that if $\xi$ is a random variable, and $A_\alpha^\beta$ denotes the set consisting of those elements $a \in S$ for which $\alpha \leq \xi(a) \leq \beta$, and if $A_\alpha^\beta \in \mathcal{B}$ for a

set X of intervals $[\alpha, \beta)$, then the conditional probabilities $P(A_\chi^\gamma \mid A_\alpha^\beta)$ can be considered for $[\alpha, \beta) \in X$ and thus $\xi$ generates a conditional probability space on the real axis $\mathcal{R}$, as the space of elementary events, the $\sigma$-algebra $\mathfrak{A}$ being the set of Borel subsets of $\mathcal{R}$ and $\mathcal{B}$ consisting of the intervals $[\alpha, \beta) \in X$.

This conditional probability space will be called the conditional probability distribution generated by $\xi$ on the real axis.

Let $F(x)$ denote a non-decreasing function of $x$ which is continuous to the left for $-\infty < x < +\infty$ (i.e. $F(x)$ is defined here by $P(X < x)$ and not by $P(X \leq x)$).

If the set $A_\alpha^\beta$ belongs to $\mathcal{B}$ whenever $F(\beta) - F(\alpha) > 0$, and we have for any subinterval of such an interval $[\alpha, \beta)$ [i.e. $[x, y) \subseteq [\alpha, \beta)$]

$$(7(a)) \qquad P(A_\chi^\gamma \mid A_\alpha^\beta) = \frac{F(y) - F(x)}{F(\beta) - F(\alpha)},$$

we shall call $F(x)$ the generalized distribution function of $\xi$; 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 $\xi$; but as $F(x)$ will be used only to calculate the conditional probabilities 7(a), this will never lead to a misunderstanding. If the distribution function

$F(x)$ of $\xi$ is absolutely continuous, and $F'(x) = f(x)$, we shall call $f(x)$ the generalized density function of $\xi$; 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 $\xi$ is uniform in $(-\infty, +\infty)$.

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

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

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

2.8. An alternative form of Axiom III

Theorem 14. Axiom III can be stated in the following equivalent form:

Axiom III'. If $B \in \mathcal{B}$, $C \in \mathcal{B}$, $B \subseteq C$ and $P(B \mid C) > 0$, we have for any $A \in \mathcal{A}$

$$P(A \mid B) = \frac{P(AB \mid C)}{P(B \mid C)}$$

Proof. Clearly Axiom III' is a special case of Axiom III. Since all the requirements of Axiom III are fulfilled and since $B \subseteq C$ we have

$$P(A \mid B)\ P(B \mid C) = P(AB \mid C)$$

i.e. $$P(A \mid B) = \frac{P(AB \mid C)}{P(B \mid C)} \quad , P(B \mid C) > 0$$

Conversely, if Axiom III' is valid, Axiom III follows. This can be shown as follows: if

$A \in \mathfrak{A}$ , $B \in \mathfrak{A}$ $(B \in \mathfrak{B} \longrightarrow B \in \mathfrak{A})$, $C \in \mathfrak{B}$ and $BC \in \mathfrak{B}$

then two cases are possible: either $P(B \mid C) = 0$ or $P(B \mid C) > 0$. In the first case we have also $P(AB \mid C) = 0$. To show this we quote Remark 2. to Theorem 7 which says: if $A \subseteq A'$ , we have $P(A \mid B) \leq P(A' \mid B)$ and it is proved using Axioms I & II and Axiom III is not needed in proof. Therefore, we can apply that Remark 2. of Theorem 7. here when we wish to prove something in connection with Axiom III.

Now to show $P(AB \mid C) = 0$ when $P(B \mid C) = 0$ we note that $AB \subseteq B$. Therefore by that just quoted remark

$$P(AB \mid C) \leq P(B \mid C) = 0$$

so $$P(AB \mid C) = 0$$

and thus $P(A \mid BC)P(B \mid C) = P(AB \mid C)$ reduces to $0 = 0$.

Now let us suppose $P(B \mid C) > 0$ . It is easy to see that Theorem 1: $P(A \mid B) = P(AB \mid B)$ follows already from Axioms I - III' . As a matter of fact by Axiom III'

if $C = B$

$$P(A \mid B) = \frac{P(AB \mid B)}{P(B \mid B)}$$

$$= P(AB \mid B) \text{ by Axiom I and therefore}$$

Theorem I can be applied in the present proof where we suppose that Axiom III' is valid. But this means $P(BC \mid C) = P(B \mid C) > 0$ and thus the conditions of Axiom III' are satisfied with BC instead of B, and it follows from Axiom III' that

$$P(A \mid BC) = \frac{P(ABC \mid C)}{P(BC \mid C)}, \quad \text{i.e.}$$

$$(8) \quad P(A \mid BC)P(BC \mid C) = P(ABC \mid C).$$

As $P(BC \mid C) = P(B \mid C)$ and $P(ABC \mid C) = P(AB \mid C)$ by Theorem I, it follows from (8) that

$$P(A \mid BC)P(B \mid C) = P(AB \mid C).$$

Thus Axiom III follows from Axiom III'.

Remark. It was already pointed out that our system of axioms can be characterized in the following manner: the set S, the $\sigma$-algebra $\mathfrak{A}$ of subsets of S, the subset $\mathfrak{B}$ of $\mathfrak{A}$ and the set function of two set variables $P(A \mid B)$ defined for $A \in \mathfrak{A}$ and $B \in \mathfrak{B}$ form a conditional probability space if $\mathcal{P}_B = [\, S, \mathfrak{A}, P(A \mid B)\,]$ is an ordinary probability space for every fixed $B \in \mathfrak{B}$ and if the probability spaces $\mathcal{P}_C$ and $\mathcal{P}_{BC}$ are connected by Axiom III if $C \in \mathfrak{B}$ and $BC \in \mathfrak{B}$.

Thus different probability fields can be combined

to form a conditional probability field if they are "compatible" in the sense that they satisfy Axiom III which can be considered as the condition of compatability.

Theorem 14 means that Axiom III contains a compatability condition for $\mathcal{P}_B$ and $\mathcal{P}_C$ where $B \subseteq C$ if and only if $P(B \mid C) > 0$ ; if $P(B \mid C) = 0$, $\mathcal{P}_B$ and $\mathcal{P}_C$ are compatible without any restriction. This fact is the basis of a general principle by use of which conditional probability spaces can be constructed.

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

Theorem 15. Let $B_1$ denote a set for which $B_1 \in \mathfrak{A}$ and $B_1 \notin \mathcal{B}$. If there exists at least one set $B_2$ with the following three properties: $\alpha)$ $B_2 \in \mathcal{B}$, $\beta)$ $B_1 \subseteq B_2$, $\gamma)$ $P(B_1 \mid B_2) > 0$, further if for any other set $B_3$ which also has the properties $\alpha)$, $\beta)$, $\gamma)$, we have $B_2 B_3 \in \mathcal{B}$, the definition of $P(A \mid B)$ can be extended for $B = B_1$ by putting

$$P(A \mid B_1) = \frac{P(AB_1 \mid B_2)}{P(B_1 \mid B_2)} \tag{9}$$

Proof. To verify (9) we have to show that Axioms I, II and III are satisfied. Axiom I is clearly satisfied since by hypothesis $\gamma$) $P(B_1 \mid B_2) > 0$ and so $P(A \mid B_1) = \frac{P(AB_1 \mid B_2)}{P(B_1 \mid B_2)} \geq 0$. Also $P(B_1 \mid B_1) = \frac{P(B_1 \mid B_2)}{P(B_1 \mid B_2)} = 1$ by (9) and by hypothesis $\gamma$). Therefore, Axiom I is satisfied.

Axiom II is clearly satisfied for any $B_1 \in \mathcal{A}$ and $B_1 \in \mathcal{B}$ for which there exists at least one set $B_2$ with the properties mentioned, because for any such $B_2 \in \mathcal{B}$, fixed, $P(AB_1 \mid B_2)$, the numerator of (9), is a countably additive set function of $A \in \mathcal{A}$.

To verify Axiom III, three cases must be distinguished.

a) If we put $B_1 = C'$ and $B'$ is a set for which $B'C' \in \mathcal{B}$ we must verify

(10) $P(A \mid B'C')P(B' \mid C') = P(AB' \mid C')$ in order to verify Axiom III. Using $C' = B_1$, this can be written as

$P(A \mid B_1 B')P(B' \mid B_1) = P(AB' \mid B_1)$. By (9) we have

$$P(B' \mid B_1) = \frac{P(B_1 B' \mid B_2)}{P(B_1 \mid B_2)} \quad \& \quad P(AB' \mid B_1) = \frac{P(AB_1 B' \mid B_2)}{P(B_1 \mid B_2)}$$

using these expressions (10) can be written as

$$P(A \mid B_1 B')\, P(B_1 B' \mid B_2) = P(AB_1 B' \mid B_2)$$

or

$$P(A \mid B_1 B' B_2)P(B_1 B' \mid B_2) = P(AB_1 B' \mid B_2)$$

for $B_1 \subseteq B_2$ by hyp. and so $B_1 B_2 = B_1$. But $B_1 B' = B'C' \in \mathcal{B}$, $B_2 \in \mathcal{B}$, therefore the conditions of Axiom III are fulfilled and our last expression

is true by force of Axiom III., i.e. starting from (10) and using (9) we arrived to an expression which satisfies the conditions of Axiom III. Therefore (10) is true and Axiom III is satisfied in this case for expression (9).

b) If $B_1 = B' C'$ where $C' \in \mathcal{B}$, we must verify

(11) $$P(A \mid B_1)P(B' \mid C') = P(AB' \mid C')$$

Substituting $P(A \mid B_1) = \frac{P(AB_1 \mid B_2)}{P(B_1 \mid B_2)}$ from (9) we have

$$P(B' \mid C')P(AB_1 \mid B_2) = P(AB' \mid C')\, P(B_1 \mid B_2)$$

which reduces to $O = O$ if $P(B' \mid C') = O$ since, repeating the argument of Remark 2 to Theorem 7, $P(B' \mid C') = O$ implies $P(AB' \mid C') = O$. If $P(B' \mid C') > O$ we have

(12) $$\frac{P(AB_1 \mid B_2)}{P(B_1 \mid B_2)} = \frac{P(AB' \mid C')}{P(B' \mid C')}$$

But $B_1 = B' C'$ and applying Theorem I (again repeating here the argument that Theorem I follows from Axioms I-III', i.e. when $P(B' \mid C') > O$) on the right of (12) we have

$$\frac{P(AB' \mid C')}{P(B' \mid C')} = \frac{P(AB'C' \mid C')}{P(B'C' \mid C')} = \frac{P(AB_1 \mid C')}{P(B_1 \mid C')}\ ; \text{i.e.}$$

we find that (12) is equivalent to

(13) $$\frac{P(AB_1 \mid B_2)}{P(B_1 \mid B_2)} = \frac{P(AB_1 \mid C')}{P(B_1 \mid C')}$$

But (13) follows from Theorem 8 by taking into account that $AB_1 + B_1 = B_1 \subseteq B_2$ and $AB_1 + B_1 = B_1 = B'C' \subseteq C'$ which together imply that

$$AB_1 + B_1 \subseteq B_2 C'.$$

To have the conditions of Theorem 8 fulfilled we have to show $B_2 C' \in \mathcal{B}$. But we have here $C \in \mathcal{B}$ and we have just shown $B_1 = B'C' \subseteq C'$; also $P(B_1 \mid C') = P(B' C' \mid C') = P(B' \mid C') > 0$, i.e. $C'$ is such a set that it satisfies conditions $\alpha)$, $\beta)$, $\gamma)$ of the Theorem. Therefore, for any other set, say $B_2$, which also has the properties $\alpha)$, $\beta)$, $\gamma)$, we have $B_2 C' \in \mathcal{B}$. Summing up we have

$$AB_1 + B_1 \subseteq B_2 C' \in \mathcal{B} \quad \text{and further}$$

$P(B_1 \mid B_2) \cdot P(B_1 \mid C') > 0$; i.e. the conditions of Theorem 8 are fulfilled. Therefore by Theorem 8 (13) is true and that makes (11) verified; i.e. Axiom III is verified in this case too for expression (9)

c) If $B' C' = B_1$ and $C' = B_1$ where $B' \in \mathcal{A}$,

we have to verify

(14) $P(A \mid B'C')P(B' \mid C') = P(AB' \mid C')$ if we are to verify Axiom III for expression (9). This is equivalent to $P(A \mid B_1)P(B' \mid B_1) = P(AB' \mid B_1)$. Using expression (9) this can be written as

$$\frac{P(AB_1 \mid B_2)}{P(B_1 \mid B_2)} \cdot \frac{P(B'B_1 \mid B_2)}{P(B_1 \mid B_2)} = \frac{P(AB'B_1 \mid B_2)}{P(B_1 \mid B_2)}$$

But the conditions $B \subset -B$ and $C \subset B$ imply that $B \subset -B$

Therefore our last equality can be written as

$$\frac{P(AB_1 \mid B_2)}{P(B_1 \mid B_2)} \cdot \frac{P(B_1 \mid B_2)}{P(B_1 \mid B_2)} = \frac{P(AB_1 \mid B_2)}{P(B_1 \mid B_2)}$$

i.e. $$\frac{P(AB_1 \mid B_2)}{P(B_1 \mid B_2)} = \frac{P(AB_1 \mid B_2)}{P(B_1 \mid B_2)}$$ and hence

(14) is verified.

The cases a), b), c) discussed above exhaust the possible ways for Axiom III to have terms extendable in the sense of Theorem 15. Therefore our proof is complete.

It is easy to see that the definition of $P(A \mid B_1)$ does not depend on the choice of $B_2$; as a matter of fact, if both $B_2$ and $B_3$ have properties $\alpha), \beta)$ and $\gamma)$, it follows by Theorem 8 that

$$\frac{P(AB_1 \mid B_2)}{P(B_1 \mid B_2)} = \frac{P(AB_1 \mid B_3)}{P(B_1 \mid B_3)}$$

It is also clear that $P(AB_1 \mid B_3)$ cannot be defined otherwise as by (9) because if $B_1$ is included into $\mathcal{B}$ (9) must hold by force of Axiom III.

Remark. If $B_1 \subseteq A \in \mathcal{A}$ and $B_1 \notin \mathcal{B}$ and is such that it satisfies conditions of Theorem 15 then

$$P(A \mid B_1) = \frac{P(AB_1 \mid B_2)}{P(B_1 \mid B_2)} = \frac{P(B_1 \mid B_2)}{P(B_1 \mid B_2)} = 1 .$$

Another possibility for including new sets into $\mathcal{B}$

is yielded by passing to the limit; this procedure is described by the following

Theorem 16. Let us suppose that $B_n \in \mathcal{B}$, $B_n \subseteq B_{n+1}$, further $P(B_n | B_{n+1}) > 0 \quad (n = 0, 1, \ldots)$ and that $\prod_{n=0}^{\infty} P(B_n | B_{n+1})$ converges. If $B_\infty = \sum_{n=0}^{\infty} B_n$ does not belong to $\mathcal{B}$, the definition of $P(A \mid B)$ can be extended for $B = B_\infty$ by putting

(15) $P(A | B_\infty) = \lim_{n \to \infty} P(A | B_n)$ for any A provided that the following condition is satisfied:

if $B \in \mathcal{B}$, $B \subseteq B_\infty$ and $P(B | B_\infty) > 0$ then for some N we have $B \subseteq B_N$.

Proof. For an arbitrary $A \in \mathcal{A}$ we put $A^0 = AB_0$, $A^{(k)} = AB_k \bar{B}_{k-1}$. Then $A^{(k)} \subseteq B_n$ for $n \geq k$ and the sequence $\{P(A^{(k)} | B_n)\}$ is, by Theorem 7, Remark I, monotonic non-increasing for $n = k, k+1, \ldots$ and bounded below by O. Thus

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

exists.

Put

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

Doing this we have defined $P^*(A \mid B_\infty)$ for every $A \in \mathcal{A}$, for $A^{(k)}$ was defined for an arbitrary $A \in \mathcal{A}$. To prove

Theorem 16, we are going to show that if $P^*(A \mid B_\infty)$ is defined by (16), Axioms I-III remain valid and further that $\lim_{n\to\infty} P(A|B_n) = P(A|B_\infty)$ exists for all $A \in \mathfrak{A}$ and is equal to $P^*(A \mid B_\infty)$, i.e. that (15) and (16) are equivalent.

Regarding Axiom I, it is clear that $P^*(A|B_\infty) \geq 0$ since $P(A^{(k)} \mid B_\infty) = \lim_{n\to\infty} P(A^{(k)}|B_n)$ exists and $\geq 0$. Therefore $P^*(A \mid B_\infty) = \sum_{k=0}^{\infty} P(A^{(k)} \mid B_\infty) \geq 0$.

The validity of $P^*(B_\infty|B_\infty) = 1$ can be shown as follows: if $A = B_\infty = \sum_{n=0}^{\infty} B_n$ then $A = B_\infty \in \mathfrak{A}$, since $\mathfrak{A}$ is a $\sigma$-algebra, and $A^{(0)} = B_0 B_\infty = B_0$ and $A^{(k)} = B_\infty B_k \bar{B}_{k-1} = B_k \bar{B}_{k-1}$ $(k = 1, 2, \cdots)$ using our previous notation for $A^0$ and $A^{(k)}$ above. Accordingly, (16) can be written as follows:

$$P^*(B_\infty|B_\infty) = P(B_0|B_\infty) + \sum_{k=1}^{\infty} P(B_k \bar{B}_{k-1} \mid B_\infty)$$

$$= P(\underbrace{B_0 + B_1\bar{B}_0}_{B_1} + B_2\bar{B}_1 + \cdots + B_N\bar{B}_{N-1} + \cdots \mid B_\infty)$$

$B_N$ etc. since $B_N \subseteq B_{N+1}$

$$= \lim_{N\to\infty} P(B_N|B_\infty)$$

But

$$P(B_N|B_\infty) = \lim_{n\to\infty} P(B_N|B_n)$$

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

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

To show this last result consider , from the above expression, $P(B_N|B_n)$ and take $N<n$ . Then

$$P(B_N|B_n)=P(B_N B_{N+1}|B_n) \quad , \quad \text{since} \quad B_N \subseteq B_{N+1}$$

$$=P(B_N|B_{N+1}B_n)P(B_{N+1}|B_n) \quad , \quad \text{by Axiom III.}$$

$=P(B_N|B_{N+1})P(B_{N+1}|B_n)$ , since $N+1 \leq n$ and so $B_{N+1} \subseteq B_n$ . If $N+1<n$ then we take $P(B_{N+1}|B_n)$, the second factor of the LHS of the above expression; i.e.

$$P(B_{N+1}|B_n) = P(B_{N+1} B_{N+2}|B_n) \quad , \quad \text{since } B_N \subseteq B_{N+1}$$

$$= P(B_{N+1}|B_{N+2}B_n)P(B_{N+2}|B_n) \quad , \quad \text{by Axiom III.}$$

$$= P(B_{N+1}|B_{N+2})P(B_{N+2}|B_n) \quad , \quad \text{since } N+2 \leq n \text{ \& so } B_{N+2} \subseteq B_n$$

So far we have

$$P(B_N|B_n)= P(B_N|B_{N+1})P(B_{N+1}|B_{N+2})P(B_{N+2}|B_n)$$

Continuing this process up to $N=n-1$ we get

$$P(B_N|B_n) = P(B_N|B_{N+1})P(B_{N+1}|B_{N+2})P(B_{N+2}|B_{N+3})\cdots P(B_{n-2}|B_{n-1})P(B_{n-1}|B_n)$$

and here the process stops since if we take now $P(B_{n-1}|B_n)$ then

$$P(B_{n-1}|B_n)=P(B_{n-1}B_n|B_n) \quad , \quad \text{by } B_{n-1} \subseteq B_n$$

$$= P(B_{n-1}|B_nB_n)P(B_n|B_n) \quad , \text{ by Axiom III}$$

$$= P(B_{n-1}|B_n)$$

Therefore, we have

$$P(B_N|B_\infty) = \lim_{n\to\infty} P(B_N|B_n)$$

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

$$= \prod_{k=N}^{\infty} P(B_k|B_{k+1}) \quad \text{and thus}$$

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

because $\prod_{k=0}^{\infty} P(B_k \mid B_{k+1})$ is convergent by hypothesis and the remainder-product of a convergent infinite product does always tend to 1.

To verify Axiom II we have to verify that $P^*(A \mid B_\infty)$ defined by (16) is countably additive. Let us suppose that $A_k \in \mathfrak{A}$ and $A_j A_k = \phi$, if $j \neq k$ $(j,k = 1,2,\cdots)$ and $A_\infty = \sum_{n=1}^{\infty} A_n$ ($A_\infty \in \mathfrak{A}$ since $\mathfrak{A}$ is a $\sigma$-algebra). Let us put $A_n^{(0)} = A_n B_0$ and $A_n^{(k)} = A_n B_k \bar{B}_{k-1}$ for $k = 1,2,\cdots$, further $A_\infty^{(0)} = A_\infty B_0$ for $k = 1,2,\cdots$.

Reasoning as above, for $A \subseteq B_k$ (we can, without loss of generality, suppose that $A \subseteq B_k$ because by Remark 3 to Theorem I $P(A \mid B_k) = 1$ if $B_k \subseteq A$ and $P(A \mid B_k) = 0$ if $AB_k = \phi$ by Theorem 5) we have

(17) $$P(B \mid B_\infty) = P(A \mid B_k) \prod_{n=k}^{\infty} P(B_n \mid B_{n+1})$$

and $P(A \mid B_k)$ is countably additive. Further $A_\infty^{(k)} = \sum_{n=1}^{\infty} A_n^{(k)}$ and $A_n^{(k)} \cdot A_m^{(k)} = \phi$ for $n \neq m$ since

$$\begin{aligned} A_\infty^{(k)} &= A_\infty B_k \bar{B}_{k-1} \\ &= \left(\sum_{n=1}^{\infty} A_n\right) B_k \bar{B}_{k-1} \\ &= \sum_{n=1}^{\infty} A_n B_k \bar{B}_{k-1} \\ &= \sum_{n=1}^{\infty} A_n^{(k)} \end{aligned}$$

and $A_n^{(k)} A_m^{(k)} = \phi$ for $n \neq m$ since $A_n A_m = \phi$ for $n \neq m$.

This implies that

$$(18)\quad P\left(A_\infty^{(k)} \mid B_\infty\right) = P\left(\sum_{n=1}^{\infty} A_n^{(k)} \mid B_\infty\right) = \sum_{n=1}^{\infty} P\left(A_n^{(k)} \mid B_\infty\right).$$

But by (16) we have

$$(19)\quad P^*\left(A_\infty \mid B_\infty\right) = \sum_{k=0}^{\infty} P\left(A_\infty^{(k)} \mid B_\infty\right)$$

From (18) and (19) it follows that

$$(20)\quad P^*\left(A_\infty \mid B_\infty\right) = \sum_{k=0}^{\infty}\sum_{n=1}^{\infty} P\left(A_n^{(k)} \mid B_\infty\right) = \sum_{n=1}^{\infty}\sum_{k=0}^{\infty} P\left(A_n^{(k)} \mid B_\infty\right)$$

From (16) we have

$$\sum_{k=0}^{\infty} P\left(A^{(k)} \mid B_\infty\right) = P^*\left(A \mid B_\infty\right)$$

and putting that expression into (20) we have

$$(21)\quad P^*\left(A_\infty \mid B_\infty\right) = \sum_{n=1}^{\infty} P^*\left(A_n \mid B_\infty\right)$$

But $A_\infty$ was defined to be $\sum_{n=1}^{\infty} A_n$ and so (21) can be written as

$$P^*\left(\sum_{n=1}^{\infty} A_n \mid B_\infty\right) = \sum_{n=1}^{\infty} P^*\left(A_n \mid B_\infty\right)$$

which means that $P^*(A \mid B_\infty)$ is countably additive. So far we have Axioms I and II satisfied by (16) i.e. by $P^*(A \mid B_\infty) = \sum_{k=0}^{\infty} P\left(A^{(k)} \mid B_\infty\right)$.

Before we proceed to Axiom III we are going to show that

$\lim_{N\to\infty} P(A \mid B_N) = P(A \mid B_\infty)$ exists for all $A \in \mathcal{A}$

and that it is

$$(22)\quad P^*(A \mid B_\infty) = \lim_{N\to\infty} P(A \mid B_N)$$

i.e. $P^*(A \mid B_\infty) = P(A \mid B_\infty)$

for any $A \in \mathfrak{A}$ . If we can show that (22) is valid for any $A \in \mathfrak{A}$ then it also means that $\lim_{N\to\infty} P(A|B_N)$ exists for all $A \in \mathfrak{A}$ because $P^*(A|B_\infty)$ is defined by (16) for every $A \in \mathfrak{A}$ . To prove (22) we start with showing that

$$(24) \qquad P(A|B_N) = \sum_{k=0}^{\infty} P(A^{(k)}|B_N) = \sum_{k=0}^{N} P(A^{(k)}|B_N) .$$

To verify (24), consider

$$P\Big(\sum_{k=0}^{\infty} A^{(k)}|B_N\Big) = P(AB_0 + AB_1\bar{B}_0 + AB_2\bar{B}_1 + \cdots + AB_N\bar{B}_{N-1} + \cdots \,|\, B_N)$$

$$= P\Big(A\big(\underbrace{B_0 + B_1\bar{B}_0}_{B_1} + B_2\bar{B}_1 + \cdots + \underbrace{B_N\bar{B}_{N-1} + \cdots}_{B_N \cdots}\big)\,\Big|\, B_N\Big) \quad \text{since } B_N \subseteq B_{N+1}$$

$$= P(A(B_\infty)|B_N)$$

$$= P(A(B_N + B_\infty - B_N)|B_N)$$

$$= P(A(B_N + \bar{B}_N)|B_N)$$

$$= P(AB_N + A\bar{B}_N|B_N)$$

$$= P(AB_N|B_N) + P(A\bar{B}_N|B_N)$$

$$= P(A|B_N) + 0 \quad \text{for } (A\bar{B}_N|B_N) \text{ is an impossible}$$

event, i.e. we have

$$(25) \qquad P\Big(\sum_{k=0}^{\infty} A^{(k)}|B_N\Big) = P(A|B_N)$$

$$= P(AB_N|B_N)$$

$$= P\Big(\sum_{k=0}^{N} A^{(k)}|B_N\Big)$$

Therefore we have

(26) $$P(A|B_N) = P\left(\sum_{k=0}^{\infty} A^{(k)} | B_N\right), \quad \text{by (25)}$$

$$= \sum_{k=0}^{\infty} P\left(A^{(k)} | B_N\right), \quad \text{by Axiom II}$$

$$= \sum_{k=0}^{N} P\left(A^{(k)} | B_N\right), \quad \text{by (25), i.e. we}$$

justified (24) here. Now by (17)

$$P\left(A^{(k)} | B_\infty\right) = P\left(A^{(k)} | B_N\right) \prod_{n=N}^{\infty} P\left(B_n | B_{n+1}\right)$$

and using this result we get

(27) $$P\left(A^{(k)} | B_N\right) = \frac{P\left(A^{(k)} | B_\infty\right)}{\prod_{n=N}^{\infty} P\left(B_n | B_{n+1}\right)}$$ ;

combining (24) and (27) we get

(28) $$P(A|B_N) = \sum_{k=0}^{\infty} P\left(A^{(k)} | B_N\right) = \sum_{k=0}^{N} P\left(A^{(k)} | B_N\right) = \frac{\sum_{k=0}^{N} P\left(A^{(k)} | B_\infty\right)}{\prod_{n=N}^{\infty} P\left(B_n | B_{n+1}\right)}$$

It follows now immediately from (28) that

(29) $$\lim_{N \to \infty} P(A|B_N) = \sum_{k=0}^{\infty} P\left(A^{(k)} | B_\infty\right), \quad \text{for any } A \in \mathfrak{A}$$

$$= P^*\left(A | B_\infty\right), \quad \text{by (16) for every } A \in \mathfrak{A}.$$

Therefore (22) is proved.

To complete the proof of the theorem we still have to show that Axiom III is valid. We are going to show that Axiom III', the equivalent form of Axiom III, is valid. Axiom III' goes as follows: If $B \in \mathcal{B}$, $C \in \mathcal{B}$, $B \subseteq C$ and $P(B|C)>0$ we have for any $A \in \mathcal{A}$

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

We distinguish three cases here. The first case is when $C \in \mathcal{B}$, $B_\infty \subseteq C$ and $P(B_\infty | C) > 0$; then we have $P(B_N|C)>0$ if N is sufficiently large and thus

$$(30) \qquad P(A|B_N) = \frac{P(AB_N|C)}{P(B_N|C)} \quad , \text{ since } B_N \in \mathcal{B}$$

Passing to the limit $N \to \infty$ and using (22), it follows that

$$(31) \qquad P(A|B_\infty) = \frac{P(AB_\infty|C)}{P(B_\infty|C)}$$

which is Axiom III' for this case specified above.

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

$$(32) \qquad P(A|B) = \frac{P(AB|B_\infty)}{P(B|B_\infty)} \quad .$$

But when $B \in \mathcal{B}$, $B \subseteq B_\infty$ and $P(B|B_\infty)>0$ then there exists, by hypothesis of the theorem, an index N for which $B \subseteq B_N$. Then, by taking N sufficiently large,

$P(B|B_N) > 0$ and we have

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

since $B_N \in \mathcal{B}$ ; passing to the limit $N \to \infty$, (32) follows:

The third case is when $B = B_\infty$ and $C = B_\infty$ ; in this case we have to prove that

(33) $$P(A|B_\infty) = \frac{P(AB_\infty|B_\infty)}{P(B|B_\infty)}$$

But we have already shown that $P(B_\infty|B_\infty) = 1$ .

All we have to show now is that

$$P(A|B_\infty) = P(AB_\infty|B_\infty).$$

But by (17), for $A \subseteq B_k$ (we can suppose this without loss of generality) we have

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

Using this expression we have

(34) $$P(AB_\infty|B_\infty) = P(AB_\infty|B_k)\prod_{n=k}^{\infty} P(B_n B_\infty|B_{n+1})$$

But $A \subseteq B_k$ and so $A \subseteq B_\infty$ and therefore $AB_\infty = A$.

Also $B_n B_\infty = B_n$ and (34) can be written as

$$P(AB_\infty|B_\infty) = P(A|B_k)\prod_{n=k}^{\infty} P(B_n|B_{n+1})$$ which simply says that (34) and (17) are equivalent, i.e.

$$P(AB_\infty|B_\infty) = P(A|B_\infty)$$

and this concludes the proof of Theorem 16.

Remark. In his paper [22] A. Rényi states this theorem (Theorem 11 in [22]) as it is stated here but instead of the condition: "if $B \in \mathcal{B}$, $B \subseteq B_\infty$ and $P(B|B_\infty)>0$ for some N we have $B \subseteq B_N$" he puts down the condition which goes as follows: "if $B \in \mathcal{B}$ and $P(B|B_N) > 0$ for some N, we have $BB_N \in \mathcal{B}$". He uses this condition when he proves that Axiom III' is satisfied when $B \in \mathcal{B}$, $B \subseteq B_\infty$ and $P(B|B_\infty) > 0$ by applying Theorem 8. there, i.e. in the case when expression (32) of this paper is to be verified. A Rényi distinguishes two cases there. Case I is the case when there exists an index N for which $B \subseteq B_N$, and then we have

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

for N large enough and passing to the limit $N \to \infty$, (32) follows. In fact, this is the case that we have here as a condition to Theorem 16. The second case is the general case where he considers two measures $\mu_1(A) = P(A|B)$ and $\mu_2(A) = P(A|B_\infty)$ and lets $A \in \mathcal{A}$ denote an arbitrary set, for which $A \subseteq B_N B$ for some N. Then, by Theorem 8, taking N sufficiently large to ensure $P(B|B_N) > 0$,

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

and from here the same expression as of (32) is derived. But to use Theorem 8 here, the following conditions are to be satisfied: $A+B \subseteq B_N B$ and $P(B|B_N) \cdot P(B|B) > 0$.

But then $A \subseteq B_N B$ as it is supposed to be and $B \subseteq B_N B$ which is possible only if $B \subseteq B_N$ and with that we are back in Case 1 where it is supposed that there exists an index N for which $B \subseteq B_N$. Therefore it seems that using Theorem 8 and, because of this, supposing that we have $BB_N \in \mathcal{B}$ whenever $B \in \mathcal{B}$ and $P(B|B_N)>0$ for some N, we cannot achieve more than by supposing that if $B \in \mathcal{B}$, $B \subseteq B_\infty$ and $P(B|B_\infty)>0$ then for some N we have $B \subseteq B_N$. That is the reason we used this latter statement as a condition to Theorem 16 instead of the original condition used by A. Rényi in his paper [22].

Remark to Theorems 15 and 16. The assertion of Theorem 8 for the case $A_1 + A_2 \subseteq B_1 B_2$, but without the suppostion $B_1 B_2 \in \mathcal{B}$, can be considered as a stronger form of Axiom III. For the proof of Theorem 8 we needed the supposition $B_1 B_2 \in \mathcal{B}$ since it was used two times using Axiom III. Dropping the supposition $B_1 B_2 \in \mathcal{B}$ we now use the result, gained when it was a supposition, as an Axiom; it shall be called Axiom III*.

Axiom III*. If $A_1 \in \mathcal{A}$, $A_2 \in \mathcal{A}$, $B_1 \in \mathcal{B}$ and $B_2 \in \mathcal{B}$, further $A_1 + A_2 \subseteq B_1 B_2$ and $P(A_2|B_1)P(A_2|B_2)>0$,

we have

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

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

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

Then Axiom III* and Axiom III** together imply Axiom III$'$ and thus Axiom III since Axiom III$'$ and Axiom III are equivalent. To show this let $A_1 = AB$, $A_2 = B$, $B_1 = B$ and $B_2 = C$ where $A \in \mathfrak{A}$, $B \in \mathfrak{B}$, $C \in \mathfrak{B}$, $B \subseteq C$ and $P(B|C) > 0$. (These are the conditions of Axiom III$'$ but we cannot say right now that Axiom III$'$ holds. All we want to say is that if the above conditions are fulfilled then Axiom III$'$ is implied by Axiom III* and Axiom III** together). The conditions of Axiom III* are satisfied since we have

$$A_1 + A_2 = AB + B = B$$

$$B_1 B_2 = BC = B$$

and so $A_1 + A_2 \subseteq B_1 B_2$ since $B \subseteq B$.

Therefore we have $\frac{P(AB|B)}{P(B|B)} = \frac{P(AB|C)}{P(B|C)}$, by Axiom III*

Also $P(A|B) = P(AB|B)$, by Axiom III**

so $P(A|B) = \frac{P(AB|B)}{P(B|B)}$ , by Axiom I

$= \frac{P(AB|C)}{P(B|C)}$ , by above result;

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

and this is the assertion of Axiom III ;

i. e. Axiom III is a consequence of Axiom III* and Axiom III**. But Axiom III* follows from Axiom III only in the special case when $B_1 B_2 \in \mathcal{B}$ . In that case Axiom III* is Theorem 8 proved by using Axiom III.

If Axioms III* and III** were supposed instead of Axiom III then in Theorem 15 the condition that for two sets $B_2, B_3$ with properties $\alpha)$ , $\beta)$ and $\gamma)$ , $B_2 B_3 \in \mathcal{B}$ could be omitted. In the preceeding Remark we have already mentioned that A. Rényi uses the condition that if $B \in \mathcal{B}$ and $P(B|B_N) > 0$ for some N, we have $BB_N \in \mathcal{B}$ instead of the condition used here to Theorem 16 ( see Theorem 11 in [ ]) . Regarding that condition he notes that if Axioms III* and III** were supposed instead of Axiom III then the requirement $BB_N \in \mathcal{B}$ could be omitted. But omitting this requirement the application of Axiom III*, instead of the application of Theorem 8, would still require the condition $A+B \subseteq B_N B$ , discussed in the above Remark, which

leads us back to supposing that there exists an index N for which $B \subseteq B_N$ i.e. to the condition that if $B \in \mathcal{B}$, $B \subseteq B_\infty$ and $P(B|B_\infty) > 0$ for some N we have $B \subseteq B_N$, the condition used here to prove Theorem 16.

2.10. Continuity properties of conditional probability. For any fixed $B \in \mathcal{B}$, $P(A|B)$ is a countably additive set function of $A \in \mathcal{A}$ (Axiom II). By "A. Continuity Theorem for Additive Set Functions" of M. Loeve's [16], p. 84, $P(A|B)$ is continuous in A; i.e. if $A_n \in \mathcal{A}$ and $A_n \subseteq A_{n+1}$ or $A_n \supseteq A_{n+1}$ for $n = 1, 2, \cdots$, we have for $B \in \mathcal{B}$

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

Regarding the continuity of $P(A|B)$ as a function of B, we have

Theorem 17. If $B_n \in \mathcal{B}$ and $B_n \subseteq B_{n+1}$ $(n = 1, 2, \cdots)$ further $\sum_{n=1}^{\infty} B_n = B \in \mathcal{B}$, we have for $A \in \mathcal{A}$

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

Proof. $P(A|B_nB)P(B_n|B) = P(AB_n|B)$

But $B_nB = B_n$ and also $\lim_{n\to\infty} P(B_n|B) = P(B|B) = 1$, i.e. $P(B_n|B) > 0$ for sufficiently large n. Therefore, for n large enough, we have

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

and thus

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

which proves Theorem 17.

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

Theorem 18. 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 $P(C|B) > 0$, it follows that

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

Proof. We have by Axiom III

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

But $P(C_n|B) \geq P(C|B)$, by Remark 2 to Theorem 7 $(C \subseteq C_n)$ and $P(C|B) > 0$ by hypothesis, i.e. $P(C_n|B) > 0$ for every n. It follows that

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

which proves Theorem 18.

## 2.11. Products of conditional probability spaces.

Given a finite sequence of conditional probability spaces

$$\mathcal{P}^{(k)} = [S^{(k)}, \mathfrak{A}^{(k)}, \mathcal{B}^{(k)}, P^{(k)}] \quad (k = 1, 2, \cdots, N) .$$

The product of these conditional probability spaces is defined as follows: let $S = S^{(1)} \circ S^{(2)} \circ \cdots \circ S^{(N)}$ denote the Cartesian product of the sets $S^{(1)}, S^{(2)}, \cdots, S^{(N)}$, i.e., let S denote the set of all ordered N-tuples $(a^{(1)}, a^{(2)}, \cdots, a^{(N)})$ where $a^{(1)} \in S^{(1)}, a^{(2)} \in S^{(2)}, \cdots, a^{(N)} \in S^{(N)}$, i.e. where $a^{(k)} \in S^{(k)}$. Let $\mathcal{B}$ denote the set of all subsets $B = B^{(1)} \circ B^{(2)} \circ \cdots \circ B^{(N)}$, where $B^{(k)} \in \mathcal{B}^{(k)}$ $(k = 1, 2, \cdots, N)$. Accordingly, B consists of those N-tuples $(a^{(1)}, a^{(2)}, \cdots, a^{(N)})$ of S for which we have $a^{(k)} \in B^{(k)}$ $(k = 1, 2, \cdots, N)$. For brevity, let us denote $\mathcal{B}$ as follows: $\mathcal{B} = \mathcal{B}^{(1)} \circ \mathcal{B}^{(2)} \circ \cdots \circ \mathcal{B}^{(N)}$.

Let $\mathfrak{A}$ denote the set of all subsets $A = A^{(1)} \circ A^{(2)} \circ \cdots \circ A^{(N)}$, where $A^{(k)} \in \mathfrak{A}^{(k)}$. Accordingly, A consists of those N-tuples $(a^{(1)}, a^{(2)}, \cdots, a^{(N)})$ of S for which we have $a^{(k)} \in A^{(k)}$ $(k = 1, 2, \cdots, N)$. For brevity, let us denote $\mathfrak{A}$ as follows: $\mathfrak{A} = \mathfrak{A}^{(1)} \circ \mathfrak{A}^{(2)} \circ \cdots \circ \mathfrak{A}^{(N)}$ and let $\bar{\mathfrak{A}}$ denote the least $\sigma$-algebra containing $\mathfrak{A}$. Let us define $P(A|B)$ for $A = A^{(1)} \circ A^{(2)} \circ \cdots \circ A^{(N)}$ and $B = B^{(1)} \circ B^{(2)} \circ \cdots \circ B^{(N)}$ by

$$P(A|B) = P^{(1)}(A^{(1)}|B^{(1)}) P^{(2)}(A^{(2)}|B^{(2)}) \cdots P^{(N)}(A^{(N)}|B^{(N)})$$

$$= \prod_{k=1}^{N} P^{(k)}(A^{(k)}|B^{(k)})$$

if $A \in \mathfrak{A}$, $B \in \mathfrak{B}$, i.e. if $A^{(k)} \in \mathfrak{A}^{(k)}$, $B^{(k)} \in \mathfrak{B}^{(k)}$ $(k = 1, 2, \dots, N)$ and extend the definition of $P(A|B)$ for every fixed $B \in \mathfrak{B}$ to all $A \in \bar{\mathfrak{A}}$ in the usual way described by Theorem A p. 54 in [8], § 13. In other words, for any given $B \in \mathfrak{B}$ we obtain the product of the Kolmogorov probability spaces $[S^{(k)}, \mathfrak{A}^{(k)}, P^{(k)}(A^{(k)}|B^{(k)})]$ $(k = 1, 2, \dots, N)$ and perform this operation for every possible $B \in \mathfrak{B}$. Thus we obtain a conditional probability space $\mathcal{P} = [S, \bar{\mathfrak{A}}, \mathfrak{B}, P]$ which will be called the Cartesian product of the conditional probability spaces $\mathcal{P}^{(k)}$ $(k = 1, 2, \dots, N)$ and denoted by $\mathcal{P} = \prod_{k=1}^{N} \mathcal{P}^{(k)}$.

The Cartesian product of a denumerable sequence of conditional probability spaces $\mathcal{P}^{(k)} = [S^{(k)}, \mathfrak{A}^{(k)}, \mathfrak{B}^{(k)}, P^{(k)}]$ $(k = 1, 2, \dots)$ is defined as follows: we denote by $S = S^{(1)} \circ S^{(2)} \circ \dots \circ S^{(n)} \circ \dots$ the Cartesian product of the sets $S^{(k)}$ $(k = 1, 2, \dots)$, by $\mathfrak{B}$ the set of all sets $B = B^{(1)} \circ B^{(2)} \circ \dots \circ B^{(n)} \circ \dots$, where $B^{(k)} \in \mathfrak{B}^{(k)}$ $(k = 1, 2, \dots)$ and by $\mathfrak{A}$ the set of all sets A of the form $A = A^{(1)} \circ A^{(2)} \circ \dots \circ A^{(n)} \circ S^{(n+1)} \circ \dots$, where $A^{(k)} \in \mathfrak{A}^{(k)}$ $(k = 1, 2, \dots, n)$, i.e. $\mathfrak{A}$ is the set of all J-cylinders of S. We define $P(A|B)$ for

$A \in \mathfrak{A}$ and $B \in \mathfrak{B}$ by

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

and extend the definition of $P(A|B)$, in the usual way, for any fixed $B \in \mathfrak{B}$ to all sets A belonging to the least $\sigma$-algebra $\bar{\mathfrak{A}}$ containing $\mathfrak{A}$. In this way we obtain a conditional probability space $\mathcal{P} = [S, \bar{\mathfrak{A}}, \mathfrak{B}, P(A|B)]$ which will be called the Cartesian product of the conditional probability spaces $\mathcal{P}^{(k)}$ and denoted by $\mathcal{P} = \mathcal{P}^{(1)} \circ \ldots \circ \mathcal{P}^{(k)} \circ \ldots$ To prove that $\mathcal{P}$, defined that way, is a conditional probability space, we have only to verify the validity of Axiom III since, clearly, for $P(A|B)$ we have $P(A|B) \geq 0$ if $A \in \mathfrak{A}$ and $B \in \mathfrak{B}$; also $P(B|B) = 1$ i.e. Axiom I is satisfied. For any fixed $B$, $P(A|B)$ is a countably additive set function of $A \in \mathfrak{A}$, i.e. Axiom II is also satisfied. To show the validity of Axiom III for $\mathcal{P} = \prod_{k=1}^{\infty} \mathcal{P}^{(k)}$ we have to show $P(A|BC)\,P(B|C) = P(AB|C)$ for $A \in \bar{\mathfrak{A}}$, $B \in \bar{\mathfrak{A}}$, $C \in \mathfrak{B}$ and $BC \in \mathfrak{B}$. $\mathfrak{B}$ consists of all the sets $B = B^{(1)} \circ B^{(2)} \circ \ldots \circ B^{(k)} \circ \ldots$, where $B^{(k)} \in \mathfrak{B}$ $(k = 1, 2, \ldots)$. If $C \in \mathfrak{B}$ and $BC \in \mathfrak{B}$, then they are of the form

(35) $$C = C^{(1)} \circ C^{(2)} \circ \ldots \circ C^{(k)} \circ \ldots$$

and

(36) $$BC = (BC)^{(1)} \circ (BC)^{(2)} \circ \ldots \circ (BC)^{(k)} \circ \ldots$$

where $C^{(k)}$ and $(BC)^{(k)} \in \mathcal{B}^{(k)}$. Clearly, Axiom III is equivalent to

(37) $$P(A|BC)\,P(BC|C) = P(ABC|C)$$

for $A \in \bar{\mathcal{A}}$, $B \in \bar{\mathcal{A}}$, $C \in \mathcal{B}$ and $BC \in \mathcal{B}$ since the statement of Theorem I is valid in the product space $\mathcal{P}$ too. The validity of Theorem I follows immediately from the uniqueness of $P(A|B)$ if $A$ is a J-cylinder. For if A is a J-cylinder we also have $P(AB|B) = P(A|B)$ ; i.e. for any fixed $B \in \mathcal{B}$ of the form $B^{(1)} \circ B^{(2)} \circ \cdots \circ B^{(n)} \circ \cdots$ we have, for every $A \in \mathcal{A}$ $(A = A^{(1)} \circ A^{(2)} \circ \cdots \circ A^{(n)} \circ S^{(n+1)} \circ \cdots)$,

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

To state it explicitly, we have got two countably additive set functions $P(A|B)$ and $P(AB|B)$ completely equivalent on $\mathcal{A}$, the set of all sets of A of the form $A = A^{(1)} \circ A^{(2)} \circ \cdots \circ A^{(n)} \circ S^{(n+1)} \circ \cdots$ The definition of $P(A|B)$ was extended for any fixed $B \in \mathcal{B}$ to all sets A belonging to the least $\sigma$-algebra $\bar{\mathcal{A}}$ containing $\mathcal{A}$ in the usual sense; see for example Theorem A, p. 54, of [8]. If the set functions $P(A|B)$, $P(AB|B)$ were equivalent on $\mathcal{A}$ then they are also equivalent on the least $\sigma$-algebra $\bar{\mathcal{A}}$ containing $\mathcal{A}$. Therefore the relation of (37) is justified. To prove the validity of Axiom III it suffices, therefore, to prove (37). To do

this, let A be a J-cylinder, i.e. of the form

$$A = A^{(1)} \circ A^{(2)} \circ \dots \circ A^{(n)} \circ S^{(n+1)} \circ S^{(n+2)} \circ \dots .$$

Then by definition

$$P(A|BC) = \prod_{k=1}^{n} P^{(k)}\left(A^{(k)} \mid (BC)^{(k)}\right), \tag{38}$$

by (35) and (36)

$$P(BC|C) = \prod_{k=1}^{\infty} P^{(k)}\left((BC)^{(k)} \mid C^{(k)}\right), \tag{39}$$

by definition, (35) and (36)

$$P(ABC|C) = \prod_{k=1}^{n} P^{(k)}\left(A^{(k)}(BC)^{(k)} \mid C^{(k)}\right) \prod_{k=n+1}^{\infty} P^{(k)}\left((BC)^{(k)} \mid C^{(k)}\right) \tag{40}$$

Putting

$$P_n = \prod_{k=1}^{n} P^{(k)}\left((BC)^{(k)} \mid C^{(k)}\right),$$

the sequence $\{P_n\}$ is non-negative, monotonically non-increasing and thus $\lim_{n\to\infty} P_n = p$ exists.

Two cases, $p = 0$ and $p > 0$, are to be distinguished.

If $p = 0$ then $P(BC|C) = 0$ by (39) and

$P(ABC|C) = 0$ by (40), and so (37) is satisfied.

If $p > 0$, then by (38) and (39)

$$P(A|BC)P(BC|C) = \prod_{k=1}^{n} P^{(k)}\left(A^{(k)} \mid (BC)^{(k)}\right) \prod_{k=1}^{\infty} P^{(k)}\left((BC)^{(k)} \mid C^{(k)}\right) \tag{41}$$

$$= \prod_{k=1}^{n} P^{(k)}\left(A^{(k)} \mid (BC)^{(k)}\right) P^{(k)}\left((BC)^{(k)} \mid C^{(k)}\right) \prod_{k=n+1}^{\infty} P^{(k)}\left((BC)^{(k)} \mid C^{(k)}\right)$$

But $P^{(k)}\left(A^{(k)} \mid (BC)^{(k)}\right) \cdot P^{(k)}\left((BC)^{(k)} \mid C^{(k)}\right)$

$= P^{(k)}\left(A^{(k)} \mid (BC)^{(k)} C^{(k)}\right) \cdot P^{(k)}\left((BC)^{(k)} \mid C^{(k)}\right)$, $(BC)^{(k)} \subseteq C^{(k)}$

$= P^{(k)}\left(A^{(k)}(BC)^{(k)} \mid C^{(k)}\right)$,

by Axiom III which is satisfied by $P^{(k)}$. Using this result (41) can be written as

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

which is, by (40), equal to $P(ABC|C)$. This proves relation (37) and, by previous argument, Axiom III for every J-cylinder $A\in\mathfrak{A}$ and therefore, using the same argument as above for verifying $P(A|B)=P(AB|B)$ for every $A$ in $\bar{\mathfrak{A}}$, the least $\sigma$-algebra containing $\mathfrak{A}$.

We are going to examine now the special case when $S^{(k)}$ is the real line and $\mathfrak{A}^{(k)}$ is the class of all Borel sets of $S^{(k)}$ $(k=1,2,\cdots)$. Consider the space $S=S^{(1)}\circ S^{(2)}\circ\cdots$ consisting of points $X=(x_1,x_2,\cdots,x_k,\cdots)$, where $x_k\in S^{(k)}$ and define the random variables $\xi_k=\xi_k(X)$ as follows: $\xi_k(X)=x_k$ $(k=1,2,\cdots)$. If $A=A^{(1)}\circ A^{(2)}\circ\cdots\circ A^{(k)}\circ\cdots$ and $B=B^{(1)}\circ B^{(2)}\circ\cdots\circ B^{(k)}\circ\cdots$, where $A^{(k)}\in\mathfrak{A}^{(k)}$, $B^{(k)}\in\mathfrak{B}^{(k)}$, i.e. $A^{(k)}$ and $B^{(k)}$ are subsets of the real line $S^{(k)}$ which belong to $\mathfrak{A}^{(k)}$ and $\mathfrak{B}^{(k)}$ respectively. Then, clearly,

$$P(\xi_k\in A|B)=P^{(k)}(A^{(k)}|B^{(k)}),\quad k=1,2,\cdots$$

i.e. $$P(\xi_k\in A|B)=P^{(k)}(\xi_k\in A^{(k)}|B^{(k)})=P^{(k)}(A^{(k)}|B^{(k)})$$

If, therefore, $S^{(k)}$ is the real line then $P^{(k)}(A^{(k)}|B^{(k)})$ is the conditional distribution system on the real line of the random variables $\xi_k$, $k=1, 2, \cdots$, defined above. It is seen that

$$P(\xi_1 \in A^{(1)}, \xi_2 \in A^{(2)}, \cdots, \xi_n \in A^{(n)} | B) = \prod_{k=1}^{n} P^{(k)}(A^{(k)}|B^{(k)})$$

or, if we put $\xi = (\xi_1, \xi_2, \cdots, \xi_n)$, then

$$P(\xi \in A | B) = \prod_{k=1}^{n} P^{(k)}(A^{(k)}|B^{(k)});$$

i.e. the random variables $\xi_1, \xi_2, \cdots, \xi_n$ are conditionnally independent for every choice of $B \in \mathcal{B}$ with respect to B. We can say, therefore, that the random variables $\xi_k$, $k=1,2,3,\cdots,n$ are independent with respect to $\mathcal{B}$. We generalize and summarize our results in the following

Theorem 19. Let $\mathcal{P}^{(k)} = [S^{(k)}, \mathfrak{A}^{(k)}, \mathcal{B}^{(k)}, P^{(k)}(A^{(k)}|B^{(k)})]$, $k=1, 2, \cdots$, where $S^{(k)}$ is the real line, $\mathfrak{A}^{(k)}$ is the class of all Borel sets of $S^{(k)}$ and $\mathcal{B}^{(k)}$ is a subclass of this class of sets. Then, forming in the space $S = S^{(1)} \circ S^{(2)} \circ \cdots$ the least $\sigma$-algebra $\bar{\mathfrak{A}}$ of $\mathfrak{A} = \mathfrak{A}^{(1)} \circ \mathfrak{A}^{(2)} \circ \cdots$ and $\mathcal{B} = \mathcal{B}^{(1)} \circ \mathcal{B}^{(2)} \circ \cdots$, there can be given a set function $P(A|B)$, $A \in \bar{\mathfrak{A}}$, $B \in \mathcal{B}$ such that $[S, \bar{\mathfrak{A}}, \mathcal{B}, P(A|B)]$ is a conditional probability space; let $X = (x_1, x_2, \cdots)$ stand for any element of S and define the random variables $\xi_k = \xi_k(X)$ as follows: $\xi_k(X) = x_k$ $(k=1, 2, \cdots)$;

then the random variables $\xi_1, \xi_2, \ldots,$ are independent with respect to $\mathcal{B}$ and for any B of the form $B^{(1)}, B^{(2)}, \ldots$ in $\mathcal{B}$ the conditional probability distribution system of $\xi_K$ is given by $P^{(K)}(A^{(K)} \mid B^{(K)}), \quad K = 1, 2, \ldots$. In this case, therefore, we can speak of the conditional probability space $[S^{(K)}, \mathcal{A}^{(K)}, \mathcal{B}^{(K)}, P^{(K)}(A^{(K)} \mid B^{(K)})]$ as a conditional distribution system.

The case when the conditional probability system $P^{(K)} = (A^{(K)} \mid B^{(K)})$ of $\xi_K$, defined as above, does not depend on $K$, i.e. when the random variables $\xi_K$ have the same distribution, is of special interest. In this case we denote the conditional probability space $[S, \bar{\mathcal{A}}, \mathcal{B}, P(A \mid B)]$ constructed in the sense of Theorem 19, for the sake of brevity, by $[S', \mathcal{A}', \mathcal{B}', P'(A \mid B)]^{\infty}$. If the conditional distribution systems $[S^{(K)}, \mathcal{A}^{(K)}, \mathcal{B}^{(K)}, P^{(K)}(A^{(K)} \mid B^{(K)})]$ are not the same then we use the notation

$$[S, \bar{\mathcal{A}}, \mathcal{B}, P(A \mid B)] = \prod_{K=1}^{\infty} [S^{(K)}, \mathcal{A}^{(K)}, \mathcal{B}^{(K)}, P^{(K)}(A^{(K)} \mid B^{(K)})]$$

and say that the conditional probability space $[S, \bar{\mathcal{A}}, \mathcal{B}, P(A \mid B)]$ is the product of the conditional probability spaces $[S^{(K)}, \mathcal{A}^{(K)}, \mathcal{B}^{(K)}, P^{(K)}(A^{(K)} \mid B^{(K)})]$.

To conclude this chapter, the idea of <u>imbedding</u> conditional probability spaces will be introduced here.

If $\mathcal{P} = [S, \mathcal{A}, \mathcal{B}, P(A \mid B)]$ and $\mathcal{P}' = [S', \mathcal{A}', \mathcal{B}', P'(A \mid B)]$ are two conditional probability spaces such that $S \subseteq S', \mathcal{A} \subseteq \mathcal{A}',$

$\mathcal{B} \subseteq \mathcal{B}'$ and $P'(A \mid B) = P(A \mid B)$ if $A \in \mathcal{A}$, $B \in \mathcal{B}$, we shall say that the conditional probability space $\mathcal{P}$ is imbedded into the conditional probability space $\mathcal{P}'$. Theorems 15 and 16 are special imbeddings of a conditional $\mathcal{P}$ into a conditional probability space $\mathcal{P}'$ for which $S' = S$, $\mathcal{A}' = \mathcal{A}$ and $\mathcal{B} \subseteq \mathcal{B}'$. These imbeddings of $\mathcal{P}$ were obtained by extension of $\mathcal{B}$ in the manners discussed there. If $\mathcal{P}$ is imbedded into $\mathcal{P}'$, we shall write $\mathcal{P} \ll \mathcal{P}'$.

# CHAPTER III

## 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". It is emphasized that the conditional probability $P(A \mid B)$ is considered as an objective characteristic of the random event A, under an objective condition B, and its value is the number in the near neighbourhood of which the conditional relative frequency $\frac{k_{AB}}{k_B}$ will be found in general, if a sufficiently great number n of observations (experiments) is made, where $k_B$ denotes the number of those observations when condition B has been realised while making the n observations and $k_{AB}$ is the number of those observations where, besides the condition B being realized, the event A has also occured. It is supposed that $k_B > 0$.

3.1. Some laws of large numbers in ordinary probability space. In this section two theorems are presented, concerning random variables defined on an ordinary probability space, which will be needed later on when proving a conditional law of large numbers.

Theorem 1. Let $\xi_1, \xi_2, \ldots \xi_n, \ldots$ be mutually independent random variables with mean values $M(\xi_n) = M_n \geq 0$ and finite variances $D_n^2 = D^2(\xi_n)$. Put $\zeta_n = \sum_{K=1}^{n} \xi_K$ and $A_n = M(\zeta_n) = M_1 + M_2 + \ldots + M_n$ and suppose that the following conditions are fulfilled:

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

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

Then it follows that

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

Proof. Theorem 1 is a consequence of the Kolmogorov inequality according to which, if $\eta_1, \eta_2, \ldots \eta_K, \ldots$ are mutually independent random variables with mean values $M(\eta_K) = 0$ and finite variances $D^2(\eta_K) = M(\eta_K^2) = D_K^2$ $(K = 1, 2, \ldots)$ we have, for any $\epsilon > 0$,

(1) $$P\left(\max_{1 \leq K \leq m} |\eta_1 + \eta_2 + \ldots + \eta_K| \geq \epsilon\right) \leq \frac{1}{\epsilon^2} \sum_{K=1}^{m} D_K^2$$

Instead of using (1) we present here an inequality,

found by J. Hajek in 1953, which is a generalization of Kolmogorov's inequality and can be used to prove directly Theorem 1 (see [10]).

The following theorem will be proved:

Theorem 2. Let $\eta_1, \eta_2, \ldots \eta_k, \ldots$ denote mutually independent random variables with mean values $M(\eta_k) = 0$ and finite variances $M(\eta_k^2) = D_k^2$ and $C_k$ $(k = 1, 2, \ldots)$ is a non-increasing sequence of positive numbers ( $C_k \geq C_{k+1}$ ); then we have for any positive integers n and m $(n < m)$ and for any $\epsilon > 0$

$$(2)\quad P\left(\max_{n \leq k \leq m} C_k \left| \eta_1 + \eta_2 + \ldots + \eta_k \right| \geq \epsilon\right) \leq \frac{1}{\epsilon^2}\left(C_n^2 \sum_{k=1}^{n} D_k^2 + \sum_{k=n+1}^{m} C_k^2 D_k^2\right)$$

Proof. Put

$$(3)\quad \gamma = \sum_{k=n}^{m-1} (\eta_1 + \eta_2 + \ldots + \eta_k)^2 (C_k^2 - C_{k+1}^2) + C_m^2 (\eta_1 + \ldots \eta_m)^2$$

then

$$(4)\quad M(\gamma) = C_n^2 \sum_{k=1}^{n} D_k^2 + \sum_{k=n+1}^{m} C_k^2 D_k^2$$

To show (4) consider (2) and write it as follows:

$$\begin{aligned}
\gamma &= \sum_{k=n}^{m-1} (\eta_1 + \eta_2 + \ldots + \eta_k)^2 (C_k^2 - C_{k+1}^2) + C_m^2 (\eta_1 + \ldots + \eta_m)^2 \\
&= \sum_{k=n}^{m-1} (C_k^2 - C_{k+1}^2) \left(\sum_{i=1}^{k} \eta_i\right)^2 + C_m^2 \left(\sum_{i=1}^{m} \eta_i\right)^2 \\
&= \sum_{k=n}^{m-1} (C_k^2 - C_{k+1}^2) \left(\sum_{i=1}^{k} \eta_i^2 + \sum_{i \neq j}^{k} \eta_i \eta_j\right) + C_m^2 \left(\sum_{i=1}^{m} \eta_i^2 + \sum_{i \neq j}^{m} \eta_i \eta_j\right)
\end{aligned}$$

Taking the expectation of $\gamma$ the cross-product terms vanish since the random variables $\eta_1, \eta_2, \cdots, \eta_k, \cdots$ are mutually independent and $M(\eta_k) = 0$, $k = 1, 2, \cdots$.

Therefore

$$M(\gamma) = \sum_{k=n}^{m-1} (c_k^2 - c_{k+1}^2) \sum_{i=1}^{k} D_i^2 + c_m^2 \sum_{i=1}^{m} D_i^2$$

$$= \sum_{k=n}^{m-1} c_k^2 \sum_{i=1}^{k} D_i^2 - \sum_{k=n}^{m-1} c_{k+1}^2 \sum_{i=1}^{k} D_i^2 + c_m^2 \sum_{i=1}^{m} D_i^2$$

Writing out terms on R H S we get

$$M(\gamma) = \left( c_n^2 \sum_{i=1}^{n} D_i^2 + c_{n+1}^2 \sum_{i=1}^{n+1} D_i^2 + \cdots + c_{m-1}^2 \sum_{i=1}^{m-1} D_i^2 \right)$$

$$- \left( c_{n+1}^2 \sum_{i=1}^{n} D_i^2 + c_{n+2}^2 \sum_{i=1}^{n+1} D_i^2 + \cdots + c_{m-1}^2 \sum_{i=1}^{m-2} D_i^2 + c_m^2 \sum_{i=1}^{m-1} D_i^2 \right)$$

$$+ c_m^2 \sum_{i=1}^{m} D_i^2$$

Cancelling out terms we get

$$M(\gamma) = c_n^2 \sum_{i=1}^{n} D_i^2 + c_{n+1}^2 D_{n+1}^2 + c_{n+2}^2 D_{n+2}^2 + \cdots + c_{m-1}^2 D_{m-1}^2 + c_m^2 D_m^2$$

$$= c_n^2 \sum_{i=1}^{n} D_i^2 + \sum_{i=n+1}^{m} c_i^2 D_i^2$$

Using indices $k$ instead of $i$ we get (4)

$$M(\gamma) = c_n^2 \sum_{k=1}^{n} D_k^2 + \sum_{k=n+1}^{m} c_k^2 D_k^2$$

Denoting by $A_r$ $(r = n, n+1, \cdots, m)$ the event consisting in the simultaneous validity of the inequalities

$$c_s |\eta_1 + \cdots + \eta_s| < \epsilon \quad (n \le s < r) \quad \text{and} \quad c_r |\eta_1 + \cdots + \eta_r| \ge \epsilon$$

(if $r = n$ only the second inequality is supposed), inequality (2)

can be written as

$$(5)\quad \sum_{r=n}^{m} P(A_r) \le \frac{1}{\epsilon^2} M(\gamma)$$

To verify (5) we let $A = \sum_{r=n}^{m} A_r$ . Then A and $\bar{A}$ are mutually exclusive events. Therefore we have

$$M(\gamma) = \sum_{r=n}^{m} M(\gamma | A_r) P(A_r) + M(\gamma | \bar{A}) P(\bar{A})$$

where $M(\gamma) \ge 0$ and $M(\gamma | \bar{A}) \ge 0$

So we have

$$(6a)\quad M(\gamma) \ge \sum_{r=n}^{m} M(\gamma | A_r) P(A_r) .$$

By (3) we have

$$M(\gamma | A_r) = \sum_{k=n}^{m-1} M\left((\eta_1 + \cdots + \eta_k)^2 | A_r\right)(C_k^2 - C_{k+1}^2) + C_m^2 M\left((\eta_1 + \cdots + \eta_m)^2 | A_r\right)$$
$$\ge \sum_{k=r}^{m-1} M\left((\eta_1 + \cdots + \eta_k)^2 | A_r\right)(C_k^2 - C_{k+1}^2) + C_m^2 M\left((\eta_1 + \cdots + \eta_m)^2 | A_r\right)$$

where from definition of $A_r$ $n \le r$ ; i.e. we have

$$(6b)\quad M(\gamma | A_r) \ge \sum_{k=r}^{m-1} M\left((\eta_1 + \cdots + \eta_k)^2 | A_r\right)(C_k^2 - C_{k+1}^2) + C_m^2 M\left((\eta_1 + \cdots + \eta_m)^2 | A_r\right)$$

We also have

$$(6c)\quad M\left((\eta_1 + \cdots + \eta_k)^2 | A_r\right) \ge M\left((\eta_1 + \cdots + \eta_r)^2 | A_r\right) \ge \frac{\epsilon^2}{C_r^2} \quad (r \le k \le m)$$

To show this we note that according to the definition

of $A_r$ $\quad C_r|\eta_1+\cdots+\eta_r| \geq \epsilon$

so $\quad C_r^2(\eta_1+\cdots+\eta_r)^2 \geq \epsilon^2$

i.e. $\quad (\eta_1+\cdots+\eta_r)^2 \geq \frac{\epsilon^2}{C_r^2}$ ,

therefore $\quad M((\eta_1+\cdots+\eta_r)^2|A_r) \geq \frac{\epsilon^2}{C_r^2}$ , the second inequality of (6c).

The first inequality in (6c) comes from the fact that the random variables $\eta_1, \cdots, \eta_r, \cdots, \eta_k$ are mutually independent and remain so under the condition $A_r$ . Therefore $M((\eta_1+\cdots+\eta_r+\cdots+\eta_k)^2|A_r) = M(\eta_1^2|A_r)+\cdots+M(\eta_r^2|A_r)+\cdots+M(\eta_k^2|A_r)$ since, because of independence, the $\binom{k}{2}$ $M(\eta_i\eta_j|A_r)=0$ , $i\neq j$ . Therefore, being each $M(\eta_i^2|A_r) \geq 0$ and $r \leq k$ we have

$$M((\eta_1+\cdots+\eta_k)^2|A_r) \geq M((\eta_1+\eta_2+\cdots+\eta_r)^2|A_r) \geq \frac{\epsilon^2}{C_r^2} \qquad (r \leq k \leq m)$$

which is the required inequality of (6c).

Inequality (5) is the consequence of (6a), (6b) and (6c). To show this we write (6b) as follows

$$\begin{aligned} M(\tau|A_r) \geq\ & C_r^2 M((\eta_1+\cdots+\eta_r)^2|A_r)+C_{r+1}^2 M((\eta_1+\cdots+\eta_{r+1})^2|A_r)+\cdots+C_{m-1}^2 M((\eta_1+\cdots+\eta_{m-1})^2|A_r) \\ & -[C_{r+1}^2 M((\eta_1+\cdots+\eta_r)^2|A_r)+\cdots+C_{m-1}^2 M((\eta_1+\cdots+\eta_{m-2})^2|A_r)+C_m^2 M((\eta_1+\cdots+\eta_{m-1})^2|A_r)] \\ & + C_m^2 M((\eta_1+\cdots+\eta_m)^2|A_r) \\ \geq\ & C_r^2 M((\eta_1+\cdots+\eta_r)^2|A_r) , \end{aligned}$$

since leaving the first $+ve$ term of the R.H.S. of this inequality as it is and observing that each one of the remaining terms $(+ve \text{ or } -ve)$ is $\geq \epsilon^2$ by (6c) and by the hypothesis that $\{c_k\}$ is a non-increasing sequence of positive integers, we can take each one of them to be equal to $\epsilon^2$ without disturbing the validity of the inequality. In that case they cancel each other and we are left with $c_r^2 M(|\eta_1 + \ldots + \eta_r|^2 | A_r)$ on R.H.S. of this inequality which is again $\geq \epsilon^2$ by (6c).

I.e. (6a) by (6c) becomes

$$M(\gamma | A_r) \geq \epsilon^2$$

Using that result in (6a) we have

$$M(\gamma) \geq \sum_{r=n}^{m} M(\gamma | A_r)\, P(A_r)$$

$$\geq \epsilon^2 \sum_{r=n}^{m} P(A_r)$$

Therefore $\sum_{r=n}^{m} P(A_r) \leq \frac{1}{\epsilon^2} M(\gamma)$, which is (5).

Now $\sum_{r=n}^{m} P(A_r) = P\left(\sum_{r=n}^{m} A_r\right)$, where the union of the events $A_r$ means the realization of the event

$$\max_{n \leq k \leq m} |\eta_1 + \ldots + \eta_k| \geq \epsilon$$

Therefore $\sum_{r=n}^{m} P(A_r) = P\left(\sum_{r=n}^{m} A_r\right) = P\left(\max_{n \leq k \leq m} |\eta_1 + \eta_2 + \ldots + \eta_k| \geq \epsilon\right)$

and so we have (2)

$$P\left(\max_{n\le k\le m}|\eta_1+\eta_2+\cdots+\eta_k|\ge\epsilon\right)\le\frac{1}{\epsilon^2}M(\gamma)=\frac{1}{\epsilon^2}\left(C_n^2\sum_{k=1}^{n}D_k^2+\sum_{k=n+1}^{m}C_k^2D_k^2\right).$$

Remark 1. If we choose $n=1$ and $C_1=C_2=\cdots=C_m=1$, we obtain from (2) as a special case inequality (1). If we choose $C_k=\frac{1}{k}$ $(k=n,n+1,\cdots,m)$ we obtain the inequality

$$(7)\quad P\left(\max_{n\le k\le m}\frac{|\eta_1+\eta_2+\cdots+\eta_k|}{k}\ge\epsilon\right)\le\frac{1}{\epsilon^2}\left(\frac{\sum_{k=1}^{n}D_k^2}{n^2}+\sum_{k=n+1}^{m}\frac{D_k^2}{k^2}\right)$$

Remark 2. By means of passing to the limit $m\longrightarrow\infty$ it is easy to deduce from (2) the following inequality:

$$(8)\quad P\left(\sup_{n\le k}C_k|\eta_1+\cdots+\eta_k|\ge\epsilon\right)\le\frac{1}{\epsilon^2}\left(C_n^2\sum_{k=1}^{n}D_k^2+\sum_{k=n+1}^{\infty}C_kD_k^2\right),$$

since $\sup_{n\le k}C_k|\eta_1+\cdots+\eta_k|$, i.e. the least upper bound of the sequence $\{C_k|\eta_1+\eta_2+\cdots+\eta_k|\}$ exists $(k=n,n+1,\cdots)$. For $C_k|\eta_1+\cdots+\eta_k|\le C_k(|\eta_1|+\cdots+|\eta_k|)$ i.e. for any k, the sequence $\{C_k|\eta_1+\cdots+\eta_k|\}$ is bounded above, i.e. we have a non-empty subset of $R_e$ which is bounded above. Therefore, it has a least upper bound.

Now, if $C_k=\frac{1}{k}$, $(k=1,2,\cdots)$ we obtain the inequality

$$(9)\quad P\left(\sup_{n\le k}\frac{|\eta_1+\cdots+\eta_k|}{k}\ge\epsilon\right)\le\frac{1}{\epsilon^2}\left(\frac{\sum_{k=1}^{n}D_k^2}{n^2}+\sum_{k=n+1}^{\infty}\frac{D_k^2}{k^2}\right)$$

Remark 3. It follows from (9) immediately that the strong law of large numbers holds for the sequence of mutually independent random variables $\eta_1, \eta_2, \cdots, \eta_k, \cdots$ if the $\eta_k$'s have mean values 0, finite variances $D_k^2 = M(\eta_k^2)$ and

(10) $$\sum_{k=1}^{\infty} \frac{D_k^2}{k^2}$$

converges.

As a matter of fact, it follows from (9) and (10) that for any $\epsilon > 0$

(11) $$\lim_{n \to \infty} P\left(\sup_{n \leq k} \frac{|\eta_1 + \eta_2 + \cdots + \eta_k|}{k} \geq \epsilon\right) = 0$$

and therefore we have

(12) $$P\left(\lim_{n \to \infty} \frac{\eta_1 + \eta_2 + \cdots + \eta_n}{n} = 0\right) = 1$$

Applying Theorem 2, Theorem I can be proved as follows: consider the result of (8)

$$P\left(\sup_{n \leq k} C_k |\eta_1 + \cdots + \eta_k| \geq \epsilon\right) \leq \frac{1}{\epsilon^2}\left(C_n^2 \sum_{k=1}^{n} D_k^2 + \sum_{k=n+1}^{\infty} C_k D_k^2\right)$$

Put $\eta_k = \xi_k - M_k \longrightarrow M(\eta_k) = 0$ and $C_k = \frac{1}{A_k}$

Therefore, $\eta_1 + \eta_2 + \cdots + \eta_k = \xi_1 + \xi_2 + \cdots + \xi_k - (M_1 + M_2 + \cdots + M_k)$

$$= \zeta_k - A_k$$

and so $$C_k |\eta_1 + \eta_2 + \cdots + \eta_k| = \frac{1}{A_k} |\zeta_k - A_k|$$

$$= \left| \frac{\zeta_k}{A_k} - 1 \right|$$

Therefore we have

$$(13)\qquad P\left(\sup_{n\le K}\left|\frac{\zeta_K}{A_K}-1\right|\ge\epsilon\right)\le\frac{1}{\epsilon^2}\left(C_n^2\sum_{k=1}^{n}D_k^2+\sum_{K=n+1}^{\infty}C_K^2D_K^2\right)$$

As a) and b) of Theorem I imply that

$$(14)\qquad \lim_{n\to\infty}\frac{\sum_{k=1}^{n}D_k^2}{A_n^2}=0$$

it follows from (13) that

$$(15)\qquad \lim_{n\to\infty}P\left(\sup_{n\le K}\left|\frac{\zeta_K}{A_K}-1\right|\ge\epsilon\right)=0 \text{ for any } \epsilon>0 \text{ and so}$$

$$(16)\qquad P\left(\lim_{n\to\infty}\frac{\zeta_n}{A_n}=1\right)=1$$

which is the assertion of Theorem 1.

3.2. A conditional law of large numbers.

Theorem 3. Let $[S,\mathfrak{A},\mathfrak{B},P(A\mid B)]$ denote a conditional probability space and $\xi_1,\xi_2,\ldots,\xi_n,\ldots$ be random variables on S which are mutually independent with respect to $C\in\mathfrak{B}$. Let $\mathfrak{I}$ denote the interval $a\le x<b$ $(a<b)$ of the real axis. Let $B_n$ denote the set of those $a\in S$ for which $\xi_n(a)\in\mathfrak{I}$ and suppose that $B_n\subseteq C$ and $B_n\in\mathfrak{B}$; let us suppose that $M(\xi_n\mid B_n)=M_n>0$ and $D^2(\xi_n\mid B_n)=D_n^2$ exists $(n=1,2,\ldots)$. Let us put $P(B_n\mid C)=p_n$ and suppose that the following conditions are satisfied:

(17) $\sum_{n=1}^{\infty} p_n = +\infty \quad \text{and} \quad \sum_{n=1}^{\infty} p_n M_n = +\infty$

(18) $\lim_{n\to\infty} \frac{\sum_{k=1}^{n} p_k M_k}{\sum_{k=1}^{n} p_k} = M$

exists;

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

Then we have

(20) $P\left( \lim_{n\to\infty} \frac{\sum_{\substack{\xi_k \in \mathcal{Y} \\ 1 \le k \le n}} \xi_k}{\sum_{\substack{\xi_k \in \mathcal{Y} \\ 1 \le k \le n}} 1} = M \;\middle|\; C \right) = 1$

**Proof.** Let us define the random variable $\epsilon_k$ as follows: $\epsilon_k = 1$ if $\xi_k \in \mathcal{Y}$ and $\epsilon_k = 0$ if $\xi_k \notin \mathcal{Y}$; let us put $\xi_k^* = \xi_k \epsilon_k$. Then we have:

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

and thus

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

To show this we have:

$$M(\xi_k^* | C) = \int_S \xi_k^* \, d P(A_{\xi_k^*} | C)$$

$$= \int_S \int \xi_k \epsilon_k \, d P(A_{\xi_k} A_{\epsilon_k} | C)$$

Now, if $\epsilon_k = 1$ then $\xi_k \in \mathcal{Y}$ and so $a \in B_k$,

$\epsilon_k = 0$ otherwise,

$$\text{i.e. } M(\xi_k^* | C) = \int_{B_k} \int_S \xi_k \cdot 1 \, d P(A_{\xi_k} B_k | C)$$

By Axiom III $P(A_{\xi_k} | B_k C) P(B_k | C) = P(A_{\xi_k} B_k | C)$

therfore

$$M(\xi_k^* | C) = \int_{B_k} \int_S \xi_k \, d P(A_{\xi_k} | B_k C) \, d P(B_k | C)$$

$$= \int_{B_k} d P(B_k | C) \int_S \xi_k \, d P(A_{\xi_k} | B_k) \; ; \; B_k \subseteq C \text{ by hyp.}$$

$$= P(B_k | C) \int_S \xi_k \, d P(A_{\xi_k} | B_k)$$

$$= p_k \, M(\xi_k | B_k)$$

$$= p_k M_k .$$

Also

$$M(\xi_k^{*2} \mid C) = \int_S \xi_k^{*2} \, dP(A_{\xi_k^*} \mid C)$$

$$= \int_S \int_S \xi_k^2 \epsilon_k^2 \, dP(A_{\xi_k} A_{\epsilon_k} \mid C)$$

$$= \int_{B_k} \int_S \xi_k^2 \cdot 1 \, dP(A_{\xi_k} B_k \mid C)$$

$$= \int_{B_k} \int_S \xi_k^2 \, dP(A_{\xi_k} \mid B_k C) \, dP(B_k \mid C)$$

$$= \int_{B_k} dP(B_k \mid C) \int_S \xi_k^2 \, dP(A_{\xi_k} \mid B_k) \quad ; B_k \subseteq C$$

$$= P(B_k \mid C) \, M(\xi_k^2 \mid B_k)$$

$$= p_k \left( D^2(\xi_k \mid B_k) + M^2(\xi_k \mid B_k) \right)$$

$$= p_k \left( D_k^2 + M_k^2 \right)$$

and thus $$D^2(\xi_k^* | C) = M(\xi_k^{*2} | C) - M^2(\xi_k^* | C)$$

$$= p_k (D_k^2 + M_k^2) - p_k^2 M_k^2$$

$$= p_k (D_k^2 + (1-p_k) M_k^2)$$

I.e. The random variables $\xi_1^*, \xi_2^*, \ldots, \xi_k^*, \ldots$ on S are mutually independent with respect to $C \in \mathcal{B}$ by hypothesis of the theorem and have mean values and variances as above. We also have $\sum_{n=1}^{\infty} p_n M_n = +\infty$ and

$$\sum_{k=1}^{\infty} \frac{p_k (D_k^2 + (1-p_k) M_k^2)}{\left(\sum_{j=1}^{k} p_j M_j\right)^2} < +\infty$$ corresponding to conditions a) and b) of Theorem I. (All the conditions of Theorem I are satisfied).

Applying Theorem I to the sequence $\xi_k^*$ of random variables on S with respect to $C \in \mathcal{B}$ it follows that

$$(21) \qquad P\left( \lim_{n\to\infty} \frac{\sum_{k=1}^{n} \xi_k^*}{\sum_{k=1}^{n} p_k M_k} = 1 \,\middle|\, C \right) = 1$$

On the other hand, let us apply Theorem I to the sequence of random variables $\epsilon_k$ on S with respect to $C \in \mathcal{B}$. As.

$$M(\epsilon_k | C) = \int_S \epsilon_k \, d P(A_{\epsilon_k} | C)$$

Now, $\epsilon_k = 1$ if $\xi_k \in \mathcal{J}$, i.e. $\alpha \in B_k$ and $\epsilon_k = 0$ otherwise.

Therfore

$$M(\epsilon_k | C) = \int_{B_k} d P(B_k | C)$$

$$= P(B_k | C)$$

$$= p_k \quad ;$$

$$\text{also } M\left(\epsilon_k^2 | C\right) = \int_S \epsilon_k^2 \, d P(A_{\epsilon_k} | C)$$

$$= \int_{B_k} d P(B_k | C)$$

$$\text{i.e.} \quad M(\epsilon_k^2 | C) = P(B_k | C)$$

$$= p_k$$

Therefore

$$D^2(\epsilon_k | C) = M(\epsilon_k^2 | C) - M^2(\epsilon_k | C)$$

$$= p_k - p_k^2$$

$$= p_k(1 - p_k) ;$$

it follows that

$$(22) \quad P\left( \lim_{n \to \infty} \frac{\sum_{k=1}^{n} \epsilon_k}{\sum_{k=1}^{n} p_k} = 1 \;\middle|\; C \right) = 1$$

Combining (22) and (21) and also (18) we have

$$P\left( \lim_{n \to \infty} \frac{\sum_{k=1}^{n} \xi_k^*}{\sum_{k=1}^{n} \epsilon_k} = \lim_{n \to \infty} \frac{\sum_{k=1}^{n} p_k M_k}{\sum_{k=1}^{n} p_k} = M \;\middle|\; C \right) = 1$$

Taking into account that

$$\sum_{\substack{\xi_k \in \mathcal{Y} \\ 1 \le k \le n}} \xi_k = \sum_{k=1}^{n} \xi_k^* \qquad \text{and} \qquad \sum_{\substack{\xi_k \in \mathcal{Y} \\ 1 \le k \le n}} 1 = \sum_{k=1}^{n} \epsilon_k$$

we have

$$P\left(\lim_{n\to\infty} \frac{\sum_{\substack{\xi_k \in \mathcal{J} \\ 1 \le k \le n}} \xi_k}{\sum_{\substack{\xi_k \in \mathcal{J} \\ 1 \le k \le n}} 1} = M \;\middle|\; C\right) = 1$$ which is

the same as statement (20) which was to be proved.

The statement of this theorem 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 $\mathcal{J}$ converges with conditional probability I with respect to C to the limit M defined by (18).

In the special case, when $M_n = M > 0$ and $D_n = D > 0$ do not depend on n, the conditions (17), (18) and (19) reduce to the single condition that the series $\sum_{n=1}^{\infty} p_n$ diverges. Let us suppose further that $\mathcal{J}$ is the closed interval $[0,1]$ and the only values in $\mathcal{J}$ which the variables $\xi_n$ can take on $B_n$ are the values O and I (of course $\xi_n$ can take also other values outside $\mathcal{J}$ ).

Suppose that $A_n$ is the set on which $\xi_n = 1$ and $p = P(A_n|B_n) > 0$ does not depend on n.

In this case the events $A_n$ and $B_n$ can be in-

terpreted 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

$p = P(A|B)$ ; in this case

$$f_n(A|B) = \frac{\sum_{\substack{\xi_k \in \mathcal{Y} \\ 1 \le k \le n}} \xi_k}{\sum_{\substack{\xi_k \in \mathcal{Y} \\ 1 \le k \le n}} 1} = \frac{\sum_{\xi_k = 1,\ 1 \le k \le n} 1}{\sum_{\xi_k = 0 \text{ or } 1,\ 1 \le k \le n} 1}$$

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

$$(23) \quad P\left(\lim_{n \to \infty} f_n(A|B) = P(A|B) \,|\, C\right) = 1, \quad \text{if } P(B|C) > 0,$$

i.e. the conditional relative frequency of the event A with respect to the event B converges to the conditional prob. of A with respect to B with conditional probability I with respect to C. A special interpretation of the assertion (23) will be presented on product spaces in the next paragraph (see Theorem 5 there).

An important corollary of Theorem 3 of this chapter will be stated now:

Corollary: Let $[S, \mathfrak{A}, \mathfrak{B}, P(A|B)]$ be a conditional probability space, let the random variables $\xi_1, \xi_2, \ldots, \xi_n, \ldots$

be defined on it and let them be independent with respect to $\mathfrak{B}$; let each of these random variables have the same conditional distribution system $[S^{(\ )}, \mathfrak{A}^{(\ )}, \mathfrak{B}^{(\ )}, P^{(\ )}(A \mid B)]$, let $B \in \mathfrak{B}^{(\ )}$ and let $B_n \in \mathfrak{B}$ the set of those $a \in S$ for which $\xi_n(a) \in B$. Let $C \in \mathfrak{B}$ and suppose that $B_n \subseteq C \quad (n = 1, 2, \ldots)$; let $p_n = P(B_n \mid C)$ and suppose that $\sum_{n=1}^{\infty} p_n = +\infty$. Let us suppose that the conditional means $M(\xi_K \mid B_K) = M(\xi_K \mid \xi_K \in B) = M > 0$ and the conditional variances $D^2(\xi_K \mid B_K) = D^2(\xi_K \mid \xi_K \in B) = D^2$ exist and so they are identical for each $\xi_K$ because of the identical distribution of the random variables $\xi_K \quad (K = 1, 2, \ldots)$. Then

$$P\left( \lim_{n \to \infty} \frac{\sum\limits_{\substack{\xi_k \in B \\ 1 \le k \le n}} \xi_k}{\sum\limits_{\substack{\xi_k \in B \\ 1 \le k \le n}} 1} = M \;\middle|\; C \right) = 1$$

Proof: When $M_K = M > 0$ and $D_K = D > 0$ do not depend on $k$, the conditions (17), (18) and (19) of Theorem 3 reduce to the single condition that the series $\sum_{n=1}^{\infty} p_n$ diverges which is the condition of this corollary. All the conditions of Theorem 3. being satisfied, our result is verified.

Remark 1. The supposition $\sum_{n=1}^{\infty} p_n = \infty$ is a natural one since $\sum_{n=1}^{\infty} p_n$ means the conditional

expected value of the number of occurrence of the events $B_n$ with respect to C in the whole course of experiments. Therefore $\sum_{n=1}^{\infty} p_n < \infty$ would mean that the events $B_n$ on condition C occur on an average only finitely many times, and in this case it would be meaningless to speak about the limiting value of the relative frequency.

Remark 2. $[S, \mathcal{A}, \mathcal{B} P(A|B)]$, the conditional probability space of this corollary and of the theorem above is not specified at all, i.e. nothing is said about the construction of it. Theorems with specifically constructed conditional probability spaces will be discussed in the next chapter.

3.3. Particular cases of Theorem 3 of section 3.2.

To demonstrate the generality of Theorem 3 of section 3.2. and to show that such conditional probability spaces where the conditions of conditional laws of large numbers are satisfied do exist , we are going to present here particular cases of Theorem 3 of section 3.2 on probability spaces constructed in the sense of Theorem 19. of Chapter II. It is not necessary for Theorem 3 that the probability spaces that appear there have been constructed in this way; the reason we construct them in this way now is to verify the existence of conditional probability spaces in which the conditions of Theorem 3 of Section 3.2 are satisfied.

Let $E$ be an experiment which is performed and the possible results of it are observed $K$ times, $K = 1, 2, \ldots$. At the $K^{\underline{th}}$ experiment the possible results of $E$ are represented by certain subsets of $\mathfrak{A}^{(K)}$, the class of all Borel sets of the real line $S^{(K)}$. Let $\mathfrak{B}^{(K)}$ be a subclass of $\mathfrak{A}^{(K)}$ and let $[S^{(K)}, \mathfrak{A}^{(K)}, \mathfrak{B}^{(K)}, P^{(K)}(A^{(K)} | B^{(K)})]$ be a conditional probability space. Suppose that $[S^{(K)}, \mathfrak{A}^{(K)}, \mathfrak{B}^{(K)}, P^{(K)}(A^{(K)} | B^{(K)})]$ is independent of $K$, i.e. that the conditions of the experiment $E$ are kept fixed for each performance of $E$. To indicate independence of k we write $[S^{(\,)}, \mathfrak{A}^{(\,)}, \mathfrak{B}^{(\,)}, P^{(\,)}(A | B)]$ (In $P^{(\,)}(\cdot | \cdot)$ we should, perhaps, use $A^{(\,)}$ and $B^{(\,)}$ instead of $A \phi B$, but it will always be understood from the content whether we are talking about events $A \phi B$ belonging to $\mathfrak{A}^{(\,)}$ and $\mathfrak{B}^{(\,)}$ or to $\bar{\mathfrak{A}} \phi \mathfrak{B}$ of the product space, defined below, respectively). We can construct, in the sense of Theorem 19. of Chap. II, $[S, \bar{\mathfrak{A}}, \mathfrak{B}, P(A | B)] = [S^{(\,)}, \mathfrak{A}^{(\,)}, \mathfrak{B}^{(\,)}, P^{(\,)}(A|B)]^{\infty}$, the product of these identical conditional probability spaces and define in it the random variables $\xi_K$ such that they are independent with respect to $\mathfrak{B}$ and the conditional distribution system of $\xi_K$ is $[S^{(\,)}, \mathfrak{A}^{(\,)}, \mathfrak{B}^{(\,)}, P^{(\,)}(A|B)]$, $K = 1, 2, \ldots$. Any particular value of $\xi_K$ then indicates the result of the $K^{\underline{th}}$ performance of the experiment $E$. Let $\eta_N(A)$ denote the number of those random variables among $\xi_1, \xi_2, \ldots, \xi_N$ which belong to the set A,

where A is a subset belonging to $\mathfrak{A}^{(\ )}$. Then

$$f_N(A|B) = \frac{g_N(AB)}{g_N(B)} \quad , \quad (B \in \mathfrak{B}^{(\ )})$$

denotes the conditional relative frequency of A with respect to B in the course of the first N experiments. $g_N(A)$ is, therefore, a number expressing the frequency of the occurrence of the event which is represented by the set A. For the sake of brevity, the event which is represented by the set A can be identified with A and so we can say that $g_N(A)$ stands for the frequency of the occurrence of the event A in the course of the first N experiments. The theorems which are to be proved here are the corollaries of Theorem 3 of paragraph 3.2. but, because of their importance, will be presented as theorems. Continuing the numbering of the previous paragraph, we have

Theorem 4. Let the random variables $\xi_1, \xi_2, \ldots \xi_n, \ldots$ have identical distribution systems $[S^{(\ )}, \mathfrak{A}^{(\ )}, \mathfrak{B}^{(\ )}, P^{(\ )}(A|B)]$ and be defined on the conditional probability space $[S, \bar{\mathfrak{A}}, \mathfrak{B}, P(A|B)] = [S^{(\ )}, \mathfrak{A}^{(\ )}, \mathfrak{B}^{(\ )}, P^{(\ )}(A|B)]^{\infty}$ in the sense of Theorem 19 of Chap. II. and so they are independent with respect to $\mathfrak{B}$. Let $B \in \mathfrak{B}^{(\ )}$ and denote the identical conditional expectation value of the random variables $\xi_k$ $(k=1,2,\ldots)$ with respect to the condition B by $M(\xi|B)$; let $C = C^{(1)} \circ C^{(2)} \circ \ldots \in \mathfrak{B}$, where $C^{(k)} \in \mathfrak{B}^{(\ )}$ and, further, let $B \subseteq C^{(k)}$ $(k=1,2,\ldots)$. Let $p_N = P(B|C^{(N)})$

and suppose the condition $\sum_{N=1}^{\infty} p_N = +\infty$ is satisfied. Denote by $h_N(\xi|B)$ the conditional empirical mean value of the random variables $\xi_1, \xi_2, \ldots, \xi_N$ with respect to the condition B, i.e. let

$$h_N(\xi|B) = \frac{\xi_{i_1} + \xi_{i_2} + \ldots + \xi_{i_n}}{n}, \qquad n = q_N(B)$$

where $\xi_{i_1}, \xi_{i_2}, \ldots, \xi_{i_n}$ are those of the random variables $\xi_1, \xi_2, \ldots, \xi_N$ whose values fall in B, and $n = q_N(B)$ stands for the frequency of the occurrence of the event B in the course of the first N experiments $(B \in \mathcal{B}^{(\,)})$. Then, if $M(\xi|B) > 0$ and $D(\xi|B)$ exist,

$$P\left(\lim_{N\to\infty} h_N(\xi|B) = M(\xi|B)\,|\,C\right) = 1$$

Proof. According to the conditions of Corollary to Theorem 3 of Section 3.2. we have $B_N \subseteq C$ for $B_N \in \mathcal{B}$, $C \in \mathcal{B}$ where $\mathcal{B}$ is from $[S, \bar{\alpha}, \mathcal{B}, P(A|B)]$ and also $P(B_N|C) = p_N$, $\sum_{N=1}^{\infty} p_N = +\infty$. Since, in Theorem 4, we have an infinite product space of identical conditional probability distribution systems, any set in $\mathcal{B}$ of $[S, \bar{\alpha}, \mathcal{B}, P(A|B)]$ is of the form $B^{(1)} \circ B^{(2)} \circ \ldots$. If, in the Corollary mentioned, we had the conditional probability space constructed as it is in Theorem 4, then we would have for $B_N, C \in \mathcal{B}$ the forms $B^{(1)} \circ B^{(2)} \circ \ldots$, $C^{(1)} \circ C^{(2)} \circ \ldots$

respectively. Accordingly, by Theorem 19 of Chap. II. we have

$$P(\xi_N \in B_N | C) = P(B^{(N)} | C^N) \quad , \quad (B^{(N)}, C^{(N)} \in \mathcal{B}^{(\ )}).$$

According to the condition of Theorem 4, let $B^{(N)} \equiv B$

and then we have

$$P(\xi_N \in B_N | C) = P(B | C^{(N)}) = p_N \quad ,$$

by Corollary to Theorem 3 of Section 3.2. But this is the

$p_N$ of Theorem 4. We also have $\sum_{N=1}^{\infty} p_N = +\infty$

by the same Corollary, which is, again, also the condition of Theorem 4. To put it in short, if we have the construction of Theorem 4 for conditional probability spaces, then the conditions of Corollary to Theorem 3 of Section 3.2. become the conditions of Theorem 4. Therefore, the statement of this theorem follows from the Corollary mentioned.

Remark 1. If, in Theorem 4, we let $C^{(N)} = B$ for $N = 1, 2, \ldots,$ then we get the ordinary strong law of large numbers as a special case of Theorem 4.

Remark 2. If $\xi_N$ can take on only the values I and O and $\xi_n = \xi(a) = 1$ if $a \in A$ where $A \in \mathcal{A}^{(\ )}$ then Theorem 4 reduces to the following:

Theorem 5. Let $[S, \bar{\mathcal{A}}, \mathcal{B}, P(A|B)] = [S^{(\ )}, \mathcal{A}, \mathcal{B}^{(\ )}, P^{(\ )}(A|B)]^{\infty}$ be defined in the sense of

Theorem 19 of Chap. II. Let $A \in \mathfrak{A}^{(\ )}$, $B \in \mathfrak{B}^{(\ )}$ and, further, let $B \subseteq C^{(N)} \in \mathfrak{B}^{(\ )}$ $(N = 1, 2, \dots)$ and $C = C^{(1)} \circ C^{(2)} \circ \dots \in \mathfrak{B}$. Also let $P(B \mid C^{(N)}) = p_N$ $(N = 1, 2, \dots)$ and suppose $\sum_{N=1}^{\infty} p_N = +\infty$. Define the random variables $\xi_K$ on S as in Theorem 19 of Chap. II i.e. such that the conditional distribution system of $\xi_K$ is $[S^{(\ )}, \mathfrak{A}^{(\ )}, \mathfrak{B}^{(\ )}, P^{(\ )}(A \mid B)]$ and they are independent with respect to $\mathfrak{B}$. Denote by

$q_N(D)$ the number of those random variables from $\xi_1, \xi_2, \dots, \xi_N$ the values of which fall in $D \in \mathfrak{A}^{(\ )}$ and let

$$f_N(A \mid B) = \frac{q_N(A\,B)}{q_N(B)}$$

the conditional relative frequency of the event A with respect to B in the course of the first N experiments. Then

$$P\left(\lim_{N \to \infty} f_N(A \mid B) = P(A \mid B) \mid C\right) = 1.$$

Again, if B = C we have the ordinary strong law of large numbers.

## BIBLIOGRAPHY

[1] G.A. Barnard, Statistical Inference, Journal of the Royal Statistical Soc., Ser. B. 11 (1949), p. 115-139.

[2] G. Bohlmann, Lebensversicherungsmathematik, Enzyklopädie der Mathematischen Wissenschaften, Bd. I, Teil II, Leipzig 1900-1904, p. 857-917.

[3] A.H. Copeland, Postulates for the Theory of Probability, Amer. Journal of Math., 63 (1941), p. 741-762.

[4] Á. Császár, Sur la structure des espaces de probabilité conditionelle, Acta Math. Acad. Sci. Hung., 6 (1955) p. 337-361.

[5] W. Feller, An Introduction to Probability Theory and Its Applications, Vol. I, Second Ed., (New York, 1957).

[6] B. N. Гливенко, Курс теории вероятностей, Moskva - Ленинград, 1939.

[7] I.J. Good, Probability and the Weighing of Evidence, (London, 1950).

[8] P.R. Halmos, Measure Theory (New York, 1958).

[9] — " — , The Foundations of Probability, Amer. Math. Monthly 51 (1944), p. 497 - 510.

[10] J. Hajek and A. Rényi, Generalization of an Inequality of Kolmogorov, Acta Math. Acad. Sci. Hung., 6 (1955), p. 281-283.

Bibliography

[11] H. Jeffreys, Theory of Probability (London, 1943).

[12] J. M. Keynes, A Treatise on Probability Theory (London, 1948)

[13] A.N. Kolmogoroff, Grundbegriffe der Wahrscheinlichkeitsrechnung, Berlin 1933 (English ed. 1948).

[14] — " — , Algebres de Boole métriques completes, VI Zjard Matematyków Pol. 1948, Dodatek do Rocznika Pol. Tow. Mat. 20, p. 21-30.

[15] B.O. Koopman, The Axioms and Algebra of Intuitive Probability, Annals of Math., 41 (1940), p. 269-292.

[16] M. Loève, Probability Theory (New York, 1955).

[17] L.H. Loomis, On The Representation of $\sigma$-complete Boolean Algebras, Bull. of the Am. Math. Soc. (1947), p. 757-760.

[18] J. Łoś, On the Axiomatic Treatment of Probability, Colloquium Mathematicum III. 2 (1955), p. 125-137.

[19] H. Reichenbach, Wahrscheinlichkeitslehre (Leiden, 1935)

[20] A. Rényi, Valószínűségszámítás (Budapest, 1954).

[21] — " — , A valószínűségszámítás új axiomatikus felépítése, MTA Mat és Fiz. Oszt. Közl., IV, 3, p. 369-427.

Bibliography

[22] A. Rényi, , On a New Axiomatic Theory of Probability, Acta Math. Acad. Sci. Hungaricae 6 (1955), p. 285-334.

[23] R. Sikorski, On the Representations of Boolean Algebras as Fields of Sets, Fund. Math. 35 (1948), p. 247-258.

[24] — " — , On the Inducing of Homomorphisms by Mappings, Fund. Math. 36 (1949), p. 7-22.

[25] — " — , The Integral in a Boolean Algebra, Colloquium Mathematicum 2 (1949), p. 20-26.

[26] H.M. Stone, The Theory of Representations for Boolean Algebras, Trans. of the Am. Math. Soc. 40 (1936), p. 37-111.