ON CONDITIONAL PROBABILITY SPACES GENERATED BY A DIMENSIONALLY ORDERED SET OF MEASURES

By A. RÉNYI

Introduction

In his paper [1] the author has given a new axiomatic theory of probability, by introducing the notion of a *conditional probability space*. A conditional probability space \([S, \mathcal{A}, \mathcal{B}, P] = \mathfrak{P}\) is defined as follows: \(S\) is an arbitrary abstract space, \(\mathcal{A}\) is a \(\sigma\)-algebra of subsets of \(S\), \(\mathcal{B}\) a non-empty subset of \(\mathcal{A}\) and \(P = P(A|B)\) is a set-function of two set variables defined for all \(A \in \mathcal{A}\) and \(B \in \mathcal{B}\); the elements of \(\mathcal{A}\) are interpreted as events, and \(P(A|B)\) as the conditional probability of the event \(A\) with respect to the condition \(B\).

It is supposed that the following axioms are fulfilled:

I) \(P(A|B)\) is for any fixed \(B \in \mathcal{B}\) a measure (i.e. a non-negative and countably additive set function) on \(\mathcal{A}\),

II) \(P(B|B) = 1\) for any \(B \in \mathcal{B}\),

III) For any \(A \in \mathcal{A}\), \(B \in \mathcal{B}\) and \(C \in \mathcal{B}\), for which \(B \subset C\) and \(P(B|C) > 0\) we have * \(P(A|B) = P(AB|C)\). As regards the immediate consequences of the axioms we mention only that applying Axiom III with \(C = B\), and taking Axiom II into account, it follows that \(P(A|B) = P(AB|B)\), and thus, that \(P(A|B) < 1\) for any \(A \in \mathcal{A}\) and \(B \in \mathcal{B}\).

The theory of probability, based on the above axioms should be considered as a natural generalization of the theory of A. N. Kolmogorov given in [2]**. As a matter of fact, if the set \(\mathfrak{B}\) consists of the single element \(S\), putting \(P(A|S) = P_S(A)\) clearly \([S, \mathcal{A}, P_S(A)]\) is a probability space in the sense of [2]. More generally, if \(B \in \mathfrak{B}\) is fixed, and we put \(P_B(A) = P(A|B)\), then \(\mathfrak{B}_B = [S, \mathcal{A}, P_B(A)]\) is a probability space according to Kolmogorov. Thus a conditional probability space is nothing else than a set of ordinary probability spaces \(\mathfrak{B}_B = [S, \mathcal{A}, P_B(A)]\), some of which, namely \(\mathfrak{B}_B\) and \(\mathfrak{B}_C\) if \(P_C(B) > 0\) and \(B \subset C\), are connected by Axiom III, which should therefore be considered as a condition of compatibility. If \([S, \mathcal{A}, Q]\) is an ordinary probability space, i.e. \(Q = Q(A)\) a mea-

---

* We note by \(AB\) the intersection, and by \(A + B\) the union of the sets \(A\) and \(B\).

** After having worked out the theory as presented in [1], the author has been informed, that Kolmogorov himself has once, in a lecture mentioned the possibility of such a generalization of his theory, but he did not publish his ideas on this matter.
A. Rényi

sure defined on the \( \mathfrak{A} \) of subsets of \( S \) satisfying the condition \( Q(S) = 1 \) and we define \( \mathfrak{B} \) as the set of those \( B \in \mathfrak{A} \) for which \( Q(B) > 0 \) and put \( P(A | B) = \frac{Q(AB)}{Q(B)} \) for \( A \in \mathfrak{A} \) and \( B \in \mathfrak{B} \) we obtain a conditional probability space \( [S, \mathfrak{A}, \mathfrak{B}, P] \) which may be called the conditional probability space generated by the ordinary probability space \( [S, \mathfrak{A}, Q] \). However not all conditional probability spaces can be generated in this way by any ordinary probability space. As a matter of fact, if the conditional probability space \( [S, \mathfrak{A}, \mathfrak{B}, P] \) is generated by the ordinary probability space \( [S, \mathfrak{A}, Q] \) we have \( S \in \mathfrak{B} \), as \( Q(S) = 1 \). On the other hand there exist conditional probability spaces in which \( S \) does not belong to \( \mathfrak{B} \), and can in no way be adjoined to \( \mathfrak{B} \) by extending the conditional probability space. This can be seen as follows. From Axiom III it follows, that if \( A \in \mathfrak{A}, B \in \mathfrak{B}, C \in \mathfrak{B}, A \subset B \subset C \) and \( P(B | C) > 0 \), we have \( P(A | B) > P(A | C) \); further if \( P(B | C) = 0 \) then also \( P(A | C) = 0 \) and therefore \( P(A | B) > P(A | C) \) holds in this case also; thus it is clear that if \( S \in \mathfrak{B} \) we have \( P(A | S) \leq \inf_{A \subset B \subset \mathfrak{B}} P(A | B) \). Thus if for a denumerable sequence of disjoint sets for which \( \sum_{n=1}^{\infty} A_n = S \), we have \( \inf_{A_n \subset B \subset \mathfrak{B}} P(A_n | B) = 0 \), if \( S \) could be adjoined to \( \mathfrak{B} \) we would have \( P(A_n | S) = 0 \) for \( n = 1, 2, \ldots \) and thus \( P(S | S) = 0 \) which contradicts Axiom II. As an example let us consider the conditional probability space, which is obtained if we choose for \( S \) the \( n \)-dimensional Euclidean space \( E_n \), for \( \mathfrak{A} \) the set \( M \) of all measurable subsets of \( E_n \), for \( \mathfrak{B} \) the set \( M^* \) of all subsets of \( E_n \) having a finite and positive \( n \)-dimensional Lebesgue-measure, and put \( P(A | B) = \frac{\mu_n(AB)}{\mu_n(B)} \) for \( B \in M^* \) where \( \mu_n \) denotes the \( n \)-dimensional Lebesgue-measure. Thus we obtain a conditional probability space \( [E_n, M, M^*, P] \) which can not be extended in such a way that \( E_n \) should belong to \( M^* \). The situation is the same in every conditional probability space in which \( P(A | B) \) is defined by \( P(A | B) = \frac{\mu(AB)}{\mu(B)} \) for \( B \in \mathfrak{B} \), where \( \mathfrak{B} \) is defined as the class of those sets \( B \) for which \( \mu(B) > 0 \) where \( \mu(B) \) is a \( \sigma \)-finite measure on \( \mathfrak{A} \) which is not bounded. Thus the notion of a conditional probability space includes cases in which conditional probability is calculated by the formula \( P(A | B) = \frac{\mu(AB)}{\mu(B)} \) where \( \mu \) is an unbounded measure, while in the theory of Kolmogorov only bounded (normed) measures can play a similar role. The fact that in such conditional probability spaces \( S \) can not be adjoined to \( \mathfrak{B} \), i. e. \( P(A | S) \) can not be defined, can be interpreted by saying, that in such conditional probability spaces ordinary probability can not be defined, only conditional probabilities have a sense. As a matter of fact, one of the chief purposes of introducing the concept of a conditional probability space was to include such cases also, which occur frequently in the physical applications of probability theory, as well as in connection with the applications of probability theory to number theory, integral geometry, etc. Thus for instance in the generalized theory there has a sense to speak about a probability distribution which is uniform in the whole \( n \)-di-
On conditional probability spaces $E_n$. The conditional probability space $[E_n, M, M^+, P]$ described above corresponds exactly to this intuitive notion.

A general principle of constructing conditional probability spaces is the following. Let $S$ denote a set, $\mathfrak{A}$ a $\sigma$-algebra of subset of $S$ and $\mathfrak{B}$ a non-void subset of $\mathfrak{A}$. Let $\Gamma$ denote an arbitrary completely ordered set, and let us suppose that to every $\gamma \in \Gamma$ there corresponds a measure $\mu_\gamma$ defined on the $\sigma$-algebra $\mathfrak{A}$. Let us call the system of measures $\{\mu_\gamma; \gamma \in \Gamma\}$ dimensionally ordered if the following postulates are satisfied:

a) To every $B \in \mathfrak{B}$ there can be found a $\beta \in \Gamma$ for which $0 < \mu_\beta(B) < +\infty$.

b) If $\mu_\beta(B) < +\infty$ and $\beta < \gamma$ we have $\mu_\gamma(B) = 0$.

Clearly if the system $\{\mu_\gamma; \gamma \in \Gamma\}$ is dimensionally ordered, there exists to any $B \in \mathfrak{B}$ only one $\beta \in \Gamma$ for which $0 < \mu_\beta(B) < +\infty$ and if $\alpha < \beta$ we have $\mu_\alpha(B) = +\infty$. Let us define $P(A | B)$ for $A \in \mathfrak{A}$ and $B \in \mathfrak{B}$ as follows: $P(A | B) = \frac{\mu_\beta(AB)}{\mu_\beta(B)}$ where $\beta$ is the uniquely determined element of $\Gamma$ for which $0 < \mu_\beta(B) < +\infty$. Thus we obtain a conditional probability space $[S, \mathfrak{A}, \mathfrak{B}, P]$ which will be called the conditional probability space generated by the dimensionally ordered set of measures $\{\mu_\gamma; \gamma \in \Gamma\}$.

In § 1 we shall prove, that if $\mathfrak{B}$ is an additive class of sets, i.e. if from $B_1 \in \mathfrak{B}$ and $B_2 \in \mathfrak{B}$ it follows that $B_1 + B_2 \in \mathfrak{B}$, then there exists a dimensionally ordered set of measures, which generates the given conditional probability space. As $\mathfrak{B}$ is the system of all events which may be taken as conditions, the postulate that $\mathfrak{B}$ is an additive class is quite natural. Thus from the point of view of probability theory we can restrict ourselves to conditional probability spaces in which $\mathfrak{B}$ is an additive class, and therefore by Theorem 1 to conditional probability spaces which can be generated by a dimensionally ordered set of measures. The idea of generating a conditional probability space by a dimensionally ordered set of measures has been suggested to the author by E. Marczewski. If the conditional probability space $[S, \mathfrak{A}, \mathfrak{B}, P]$ is generated by an ordinary probability space $[\mathfrak{B}, \mathfrak{A}, Q]$ then $\mathfrak{B}$ is clearly an additive class; thus if the condition that $\mathfrak{B}$ should be an additive class is included into our system of axioms, the theory still remains a generalization of that of A. N. Kolmogorov. The condition that $\mathfrak{B}$ should be an additive class is not necessary for the existence of a dimensionally ordered set of measures which generates the given conditional probability space. If however such a system of measures exists, the conditional probability space can be extended by replacing $\mathfrak{B}$ by an additive class $\mathfrak{B}'$ containing $\mathfrak{B}$. As a matter of fact if the dimensionally ordered set $\{\mu_\gamma\}$ of measures generates the conditional probability space, and $\mathfrak{B}_\gamma$ is the set of all those sets $B$ for which $0 < \mu_\gamma(B) < +\infty$ then $\sum_{\gamma \in \Gamma} \mathfrak{B}_\gamma$ is clearly an additive class, further by putting

$$P(A | C) = \frac{\mu_\gamma(AC)}{\mu_\gamma(C)} \text{ for } C \in \mathfrak{B}_\gamma \quad (\gamma \in \Gamma)$$
the definition of $P(A \mid B)$ is thereby extended for $B \in \mathcal{B}^* = \sum_{\mathcal{B}_\gamma} \mathcal{B}_\gamma$. Thus the condition that it shall be possible to extend the conditional probability space in such a way that $\mathcal{B}$ shall become additive, is necessary and sufficient for the existence of a dimensionally ordered generating set of measures. A. Csaszar [3] has found other necessary and sufficient conditions for the existence of a dimensionally ordered set of measures which generates a given conditional probability space; especially he has generalized the theorem given in the present paper by proving that for the existence of a dimensionally ordered set of measures which generates the given conditional probability space it is sufficient, that $\mathcal{B}$ should be a quasi-additive class; the class $\mathcal{B}$ of sets is called quasi-additive by Csaszar, if to any $B_1 \in \mathcal{B}$ and $B_2 \in \mathcal{B}$ there exists a $B_3 \in \mathcal{B}$ such that $B_1 \subset B_2$, $B_2 \subset B_3$ and $P(B_1 \mid B_2) + P(B_2 \mid B_3) > 0$. An additive class is always quasi-additive, as $B_1 + B_2$ can be chosen for $B_3$. However the condition that $\mathcal{B}$ should be quasi-additive is also not necessary.

We shall show in § 1 that in case $\mathcal{B}$ is an additive class, any one of the measures $P_\gamma$ generating the given conditional probability space is uniquely determined, up to a positive constant factor, on a certain $\sigma$-ring $A_\gamma$. Without the supposition that $\mathcal{B}$ is an additive class this is not true in general. This is an other argument in favour of postulating the additivity of $\mathcal{B}$.

In § 2 some questions concerning the distribution or random variables defined on a conditional probability space are discussed.

The author is indebted to A. Csaszar and I. Czipszer for some valuable remarks.

§ 1. Conditional probability spaces in which the set of conditions is an additive class

In this § we shall make the following conventions concerning the symbol $+\infty$: $\frac{c}{0} = +\infty$, $\frac{c}{+\infty} = 0$, $c + \infty = +\infty$ if $c$ is a positive number, further $+\infty + +\infty = +\infty$.*

We shall prove the following

Theorem 1. If $\mathcal{B} = [\mathcal{S}, \mathcal{A}, \mathcal{B}, P]$ is a conditional probability space and $\mathcal{B}$ is an additive class, then $\mathcal{B}$ can be generated by a dimensionally ordered set of measures.

To prove Theorem 1 we first prove the following

Lemma 1. If $\mathcal{B}$ is an additive class and $B_i \in \mathcal{B}$ ($i = 1, 2, 3$), we have

$$P(B_1 \mid (B_1 + B_2) P(B_2 \mid B_2 + B_3) P(B_3 \mid B_3 + B_1) =$$

$$= P(B_2 \mid (B_1 + B_2) P(B_2 \mid B_2 + B_3) P(B_1 \mid B_3 + B_1).$$

(1)

Proof of Lemma 1. As it is supposed, that $\mathcal{B}$ is an additive class, putting $B = B_1 + B_2 + B_3$ we have $B \in \mathcal{B}$. Let $j$ denote the number of those of the numbers $P(B_i \mid B)$ ($i = 1, 2, 3$) which are positive. As $P(B_1 \mid B) + P(B_2 \mid B) + P(B_3 \mid B) \geq P(B \mid B) = 1$, we have $j \geq 1$. Thus

* Compare [4] p. 1. We follow [4] as regards the terminology of measure theory; thus especially we admit that a measure may take, on some sets, the value $+\infty$. 

\[\text{A. Rényi}\]
we have to distinguish three cases: 1) \( j = 3 \), 2) \( j = 2 \) and finally 3) \( j = 1 \). We shall prove that (1) holds in any of the three cases. In case 1) we have \( \mathbb{P}(B_h + B_1 | B) > 0 \) for \( i, h = 1, 2, 3 \) and thus by Axiom III

\[
\mathbb{P}(B_i | B_h + B_i) = \frac{\mathbb{P}(B_i | B)}{\mathbb{P}(B_h + B_i | B)} \quad \text{for} \quad i, h = 1, 2, 3;
\]

thus both sides of (1) are equal to

\[
\frac{\mathbb{P}(B_i | B) \mathbb{P}(B_1 | B) \mathbb{P}(B_3 | B)}{\mathbb{P}(B_i + B_1 | B) \mathbb{P}(B_i + B_3 | B) \mathbb{P}(B_i + B_1 + B_i | B)}.
\]

In case 2) we may suppose that \( \mathbb{P}(B_1 | B) = 0, \mathbb{P}(B_2 | B) > 0 \) and \( \mathbb{P}(B_3 | B) > 0 \). It follows, that \( \mathbb{P}(B_1 + B_2 | B) > 0 \) and \( \mathbb{P}(B_1 + B_3 | B) > 0 \) and thus by Axiom III that

\[
\mathbb{P}(B_1 | B) = \frac{\mathbb{P}(B_1 | B)}{\mathbb{P}(B_1 + B_2 | B)} = 0 \quad \text{and} \quad \mathbb{P}(B_1 | B) = \frac{\mathbb{P}(B_1 | B)}{\mathbb{P}(B_1 + B_3 | B)} = 0.
\]

and thus both sides of (1) are equal to 0. Thus (1) holds in all three cases.

Let us turn now to the proof of Theorem 1. Let us introduce the notation

\[
i(B_1, B_2) = \frac{\mathbb{P}(B_1 | B) \mathbb{P}(B_2 | B)}{\mathbb{P}(B_1 + B_2 | B)} \quad \text{(2)}
\]

(2) should be understood in the sense that if the denominator is equal to 0, we have \( i(B_1, B_2) = +\infty \); it should be noted that the nominator and the denominator of the fraction on the right-hand side of (2) can not both be equal to 0 as \( \mathbb{P}(B_1 | B) + \mathbb{P}(B_2 | B) + \mathbb{P}(B_1 + B_2 | B) > 1 \). We shall call \( i(B_1, B_2) \) the indicator of the two sets \( B_1 \) and \( B_2 \). If for two sets, \( B_1 \) and \( B_2 \) we have \( 0 < i(B_1, B_2) < +\infty \) we shall say that \( B_1 \) and \( B_2 \) are of the same dimension, and denote this by writing \( B_1 \sim B_2 \).

Now we prove the following

**Lemma 2.** The relation \( \sim \) has the following properties:

1) \( B_1 \sim B_1 \)

2) \( \text{If} \; B_1 \sim B_2 \; \text{then} \; B_2 \sim B_1 \)

3) \( \text{If} \; B_1 \sim B_2 \; \text{and} \; B_2 \sim B_3 \; \text{then} \; B_1 \sim B_3 \).

**Proof of Lemma 2.** 1) follows clearly from Axiom II. 2) is evident since if \( B_1 \sim B_2 \) we have

\[
i(B_2, B_1) = \frac{1}{i(B_1, B_2)} \quad \text{(3)}
\]

Let us remark that (3) remains valid also for any \( B_1 \in \mathfrak{B} \) and \( B_2 \in \mathfrak{B} \), as we agreed that \( \frac{c}{0} = +\infty \) and \( \frac{c}{+\infty} = 0 \) if \( 0 < c < +\infty \). 3) follows from the fact that if \( B_1 \sim B_2 \) and \( B_2 \sim B_3 \) we may write (1) in the form

\[
i(B_1, B_3) = i(B_1, B_2) i(B_2, B_3) \quad \text{(4)}
\]

Clearly (4) is valid always (i. e. not only if \( B_1 \sim B_2 \sim B_3 \)) except if one of the factors at the right of (4) is 0 and the other +\( \infty \) as we have agreed that \( +\infty = +\infty \) if \( 0 < c < +\infty \) and \( +\infty + +\infty = +\infty \). It follows from Lemma 2 that \( \sim \) is an equivalence-relation. Thus the set \( \mathfrak{B} \)

5 Теория вероятностей, № 1
is divided with respect to the relation $\sim$ in a system of mutually disjoint classes $\mathfrak{Y}$, every class consisting of sets which are mutually equivalent with each other. It is easy to show that a complete ordering of the classes $\mathfrak{Y}$ can be given as follows: let $\mathfrak{X}$ and $\mathfrak{Y}$ denote two disjoint classes. Let us choose an arbitrary element $B_1$ of $\mathfrak{X}$ and an arbitrary element $B_2$ of $\mathfrak{Y}$. Then only two cases are possible: either $i(B_1, B_2) = 0$ or $i(B_1, B_2) = +\infty$. As a matter of fact $0 < i(B_1, B_2) < +\infty$ is impossible, since this would imply that $B_1 \sim B_2$ and thus the classes $\mathfrak{X}$ and $\mathfrak{Y}$ would be identical, in contrary with our supposition that $\mathfrak{X}$ and $\mathfrak{Y}$ are different classes with respect to the relation $\sim$. If $i(B_1, B_2) = 0$ we shall put $\mathfrak{X} \subset \mathfrak{Y}$; if $i(B_1, B_2) = +\infty$ we shall put $\mathfrak{Y} \subset \mathfrak{X}$. To show that in this way a complete ordering of the classes $\mathfrak{Y}$ is obtained, it suffices to prove that the particular choice of the sets $B_1 \in \mathfrak{X}$ and $B_2 \in \mathfrak{Y}$ is irrelevant, further that the relation $\subset$ introduced between the classes $\mathfrak{X}$ is transitive, i.e. if $\mathfrak{X} \subset \mathfrak{Y}$ and $\mathfrak{Y} \subset \mathfrak{Z}$ we have $\mathfrak{X} \subset \mathfrak{Z}$. To prove the first assertion it suffices to show that if $B_1 \in \mathfrak{X}$, $B_2 \in \mathfrak{Y}$, and $B_3 \in \mathfrak{Z}$ further if $i(B_2, B_3) = 0$ we have also $i(B_1, B_3) = 0$. This however is a consequence of the relation (4) applied to the sets $B_1, B_2, B_3$, which is permissible as $B_1 \sim B_2$.

The transitivity of the relation follows also from (4). As a matter of fact if $\mathfrak{X} \subset \mathfrak{Y}$ and $\mathfrak{Y} \subset \mathfrak{Z}$, let us choose sets $B_1 \in \mathfrak{X}$, $B_2 \in \mathfrak{Y}$, $B_3 \in \mathfrak{Z}$. It follows that $i(B_1, B_2) = 0$ and $i(B_2, B_3) = 0$, and thus by (4) that $i(B_1, B_3) = 0$. It is clear that each $\mathfrak{Y}$ is an additive class, because if $B_1 \sim B_2$ we have

$$i(B_1 + B_2, B_1) = \frac{1}{P(B_1 | B_1 + B_2)}$$

and thus $0 < i(B_1 + B_2, B_1) < +\infty$, which means that $B_1 + B_2 \sim B_1$. If $B_1 \in \mathfrak{X}$ and $B_2 \in \mathfrak{Y}$ where $\mathfrak{X} \subset \mathfrak{Y}$ we have $B_1 + B_2 \in \mathfrak{Y}$; as a matter of fact if $P(B_1 | B_1 + B_2) = 0$ then $P(B_2 | B_1 + B_2) = 1$ and thus

$$i(B_1 + B_2, B_2) = 1.$$

Now let us introduce corresponding to every class $\mathfrak{Y}$ the set function $\mathfrak{Y}(B)(B \in \mathfrak{Y})$ by choosing an arbitrary element $B_0$ of $\mathfrak{Y}$ and putting

$$\mathfrak{Y}(B) = i(B, B_0) \text{ for } B \in \mathfrak{Y}. \quad (5)$$

Clearly we have $0 < \mathfrak{Y}(B) < +\infty$ for $B \in \mathfrak{Y}$. We define further $\mathfrak{Y}$ for any set which can be represented in the form $AB$ where $A \in \mathfrak{X}$ and $B \in \mathfrak{B}$, by putting

$$\mathfrak{Y}(AB) = P(A | B) \mathfrak{Y}(B). \quad (6)$$

It is easy to see that the value of $\mathfrak{Y}(C)$ does not depend on the particular representation of the set $C$ in the form $C = AB$, $(A \in \mathfrak{X}, B \in \mathfrak{Y})$ which has been chosen. As a matter of fact, if $C = AB = A'B'$ where $A \in \mathfrak{A}$, $B \in \mathfrak{B}$, and $A' \in \mathfrak{A}$, $B' \in \mathfrak{B}$, we have by Axiom III $P(C | B) = P(C | B' B') = P(C | B + B') = P(B | B + B')$, and thus, using the relation $P(A | B) = P(AB | B)$, we have

$$P(A | A) \mathfrak{Y}(B) = P(C | B) i(B, B_0) = P(C | B') i(B, B_0) = P(C | B') i(B', B_0) = P(A | B') \mathfrak{Y}(B').$$
Denoting by $\mathfrak{A}_Y$ the set of those sets $C \in \mathfrak{A}$ which can be represented in the form $C = AB$ with $A \in \mathfrak{A}$ and $B \in \mathfrak{A}_Y$, it is easy to see that $\mathfrak{A}_Y$ is a ring. Now we prove the following

**Lemma 3.** The set function $\nu_\tau(A)$ is a measure on the ring $\mathfrak{A}_Y$.

**Proof of Lemma 3.** First we prove that $\nu_\tau(A)$ is finitely additive on $\mathfrak{A}_Y$. Let us suppose that $C_1 = A_1 B_1$ and $C_2 = A_2 B_2$ where $A_i \in \mathfrak{A}$ and $B_1, B_2 \in \mathfrak{A}_Y$ (i = 1, 2) further that $C_1, C_2 = O$.\footnote{Here and in what follows $O$ denotes the empty set.}

As $A_i B_i = A_i (B_1 + B_2) = A_i B_1 + A_i B_2$ where $A_i \in \mathfrak{A}$ (i = 1, 2) and $B_1, B_2 \in \mathfrak{A}_Y$, we may suppose that $C_1 = A_1 B$ and $C_2 = A_2 B$ where $A_i \in \mathfrak{A}$ (i = 1, 2), $B \in \mathfrak{A}_Y$. We may suppose also that $A_i \subseteq B$ (i = 1, 2), as $A_i B = = (A_i B) B$ and $A_i B \subseteq B$. In this case $C_1 C_2 = O$ implies $A_1 A_2 = O$. Thus we have

$$\nu_\tau(C_1 + C_2) = \nu_\tau((A_1 + A_2) B) = \left( P(A_1 | B) + P(A_2 | B) \right) \nu_\tau(B) =$$

$$= \nu_\tau(C_1) + \nu_\tau(C_2).$$

Now let us prove that $\nu_\tau$ is also countably additive on $\mathfrak{A}_Y$. If $C_n \in \mathfrak{A}_Y$, $(n = 1, 2, \ldots)$ $C_k \cdot C_l = O$ if $k \neq l$ (k, l = 1, 2, \ldots) and $\sum_{n=1}^{\infty} C_n = C \in \mathfrak{A}_Y$ we may put $C_n = A_n B_n$ and $C = AB$ where $B_n \in \mathfrak{A}_Y$ and $A_n \in \mathfrak{A}$ (n = 1, 2, \ldots) further $B \in \mathfrak{A}_Y$ and $A \in \mathfrak{A}$. As $A_k B_n \subseteq AB$ we may put $C_n = A_n B_n = = A_n B_n AB = (A_n B_n AB) B = A_n B$ where $A_n \subseteq B$. It follows that $A_k \cdot A_l = O$ for $k \neq l$ and thus that

$$\nu_\tau(C) = \nu_\tau(AB) = P(A | B) \nu_\tau(B) = P(AB | B) \nu_\tau(B) =$$

$$= \sum_{n=1}^{\infty} P(A_n | B) \nu_\tau(B) = \sum_{n=1}^{\infty} \nu_\tau(A_n B) = \sum_{n=1}^{\infty} \nu_\tau(C_n),$$

which proves Lemma 3.

According to a well-known theorem (see [4], p. 54) the definition of $\nu_\tau$ can be extended to the least $\sigma$-ring $\mathfrak{A}_Y$ containing $\mathfrak{A}_r$. Thus $\nu_\tau$ is defined for every set $A \in \mathfrak{A}$ which can be covered by a countable sequence $\{B_k\}$ of sets $B_k \in \mathfrak{A}_Y$ (k = 1, 2, \ldots). As a matter of fact if $A \subseteq \bigcup_{n=1}^{\infty} B_k$

where $A \in \mathfrak{A}$ and $B_k \in \mathfrak{A}_Y$ (k = 1, 2, \ldots) we have $A = \sum_{k=1}^{\infty} AB_k$ and thus $A \in \mathfrak{A}_Y$. For those sets $A \in \mathfrak{A}$ for which such a covering does not exist, we put $\nu_\tau(A) = + \infty$. Thus $\nu_\tau$ is extended to the whole $\sigma$-algebra $\mathfrak{A}$. It is clear that if $B \in \mathfrak{A}_Y$ and $A \in \mathfrak{A}$ we have from (6)

$$\nu_\tau(AB) = P(A | B) \nu_\tau(B)$$

and thus

$$P(A | B) = \frac{\nu_\tau(AB)}{\nu_\tau(B)}. \quad (7)$$

To complete the proof of Theorem 1 we have only to prove that the system $\{\nu_\tau\}$ of measures satisfies all requirements of Theorem 1. Thus we have to prove that to any $B \in \mathfrak{B}$ there exists a unique class $\mathfrak{B}_B$ such that
A. Rényi

\[ \nu_\gamma(B) = \nu_\gamma(B(B + B_0)) = P(B | B + B_0) \nu_\gamma(B + B_0) = 0. \]

Thus Theorem 1 is proved.

It is easy to find a necessary and sufficient condition ensuring that the set \{\nu_\gamma\} should consist of a single measure \mu. As a matter of fact, it can be seen from the proof of Theorem 1, that this takes place if and only if any two elements of \mathfrak{B} are equivalent, i.e. if \( P(B_1 | B_1 + B_2) P(B_2 | B_1 + B_2) > 0 \) for any \( B_1, B_2 \in \mathfrak{B} \).

It can be easily shown that if \( \mathfrak{B} \) is an additive class, the measures \( \nu_\gamma \) generating the conditional probability space \( \mathfrak{A} \) are determined by \( \mathfrak{B} \) uniquely up to a constant factor. To show this we start from the fact, that the equivalence relation \( \sim \) and thus the classes \( \mathfrak{A} \) are uniquely determined by the conditional probability field, as well as by a set \{\nu_\gamma\} of measures generating the conditional probability field. As a matter of fact \( B_1 \sim B_2 \) if and only if \( P(B_1 | B_1 + B_2) P(B_2 | B_1 + B_2) > 0 \), further if and only if \( \nu_\gamma(B_1) \) and \( \nu_\gamma(B_2) \) are for the same \( \gamma \in \Gamma \) equal to 0, to some positive number, and to \( + \infty \). Thus in proving that the measures \( \nu_\gamma \) are uniquely determined up to a constant factor, we may restrict ourselves to a class \( \mathfrak{A} \).

If \( \mu \) and \( \nu \) are measures on \( \mathfrak{A} \), \( 0 < \mu(B) < + \infty \) and \( 0 < \nu(B) < + \infty \) for \( B \in \mathfrak{A} \) further \( P(A | B) = \frac{\nu(AB)}{\nu(B)} = \frac{\nu(AB)}{\nu(B)} \) for any \( B \in \mathfrak{A} \) and \( A \in \mathfrak{A} \), let us choose an arbitrary \( B_0 \in \mathfrak{A} \) and put \( \nu_\gamma(B) = \frac{\nu_\gamma(B)}{\nu_\gamma(B_0)} \) and \( \nu_\gamma(B) = \frac{\nu_\gamma(B)}{\nu_\gamma(B_0)} \). Thus we have \( P(A | B) = \frac{\nu_\gamma(AB)}{\nu_\gamma(B)} = \frac{\nu_\gamma(AB)}{\nu_\gamma(B)} \) and as \( \nu_\gamma(B_0) = \nu_\gamma(B_0) = 1 \) we have \( \nu_\gamma(C) = \nu_\gamma(C) \) for any \( C \in \mathfrak{A} \) for which \( C \subseteq B_0 \).

Now let us choose an arbitrary \( B \in \mathfrak{A} \); then

\[ P(B | B + B_0) = \frac{\nu_\gamma(B)}{\nu_\gamma(B + B_0)} = \frac{\nu_\gamma(B)}{\nu_\gamma(B + B_0)}. \]

But \( \nu_\gamma(B + B_0) = \nu_\gamma(B) + \nu_\gamma(B_0) - \nu_\gamma(BB_0) \) and \( \nu_\gamma(B + B_0) = \nu_\gamma(B) + \nu_\gamma(B_0) - \nu_\gamma(BB_0) \) and owing to \( BB_0 \subseteq B \) it follows that \( \nu_\gamma(BB_0) = \nu_\gamma(B) + \nu_\gamma(B_0) - \nu_\gamma(BB_0) \) which implies \( \nu_\gamma(B) = \nu_\gamma(B) \). Thus \( \nu_\gamma \) and \( \nu \) coincide for \( \mathfrak{A} \) and thus for \( \mathfrak{A} \) i.e. we have \( \mu_\gamma(A) = c \nu_\gamma(A) \) for \( A \in \mathfrak{A} \) where \( c = \frac{\nu_\gamma(B_0)}{\nu_\gamma(B_0)} > 0 \).

As regards sets not contained in \( \mathfrak{A} \), \( \mu \) and \( \nu \) may not coincide. This is shown by the following example: Let \( \mathcal{S} \) be the set of real numbers, \( \mathfrak{A} \) the set of all Borel-subsets of \( \mathcal{S} \) and \( \mathfrak{B} \) the set of all non-empty finite subsets of the set of integers. Let us define \( P(A | B) \) for \( B \in \mathfrak{B} \) and \( A \in \mathfrak{A} \) as the ratio of the number of elements of \( AB \) and the
number of elements of $B$ ($B$ and $AB$ both being finite subsets of the set of all integers). Now we have

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

where $\nu(A)$ is defined as follows: $\nu(A)$ is equal to the number of elements of the set $A$ if $A$ is a finite subset of the set of integers, and $\nu(A) = +\infty$ if $A$ is an infinite set of integers, or a set of real numbers which contains at least one not integer number, and $\mu(A)$ is defined for any set $A$ of real numbers as the number of integers contained in $A$. Clearly $\nu(A)$ and $\mu(A)$ coincide if $A$ is a subset of the set of integers, but are not identical, as for instance if $A$ is the set consisting of the two numbers 0 and $\frac{1}{2}$ we have $\nu(A) = +\infty$ and $\mu(A) = 1$. It should be mentioned, that if $\mathcal{B}$ is a conditional probability space which can be generated by a system $\{\nu_r\}$ of dimensionally ordered measures, but $\mathcal{B}$ is not an additive class, the measures $\nu_r$ are still less determined. For example let us consider the conditional probability space $\mathcal{B} = [S, \mathcal{A}, \mathcal{P}]$ where $S$ denotes the set of real numbers $\mathcal{A}$ the set of all Borel-subsets of $S$ and let us have $\mathcal{B} = \mathcal{B}_+ + \mathcal{B}_-$ where $\mathcal{B}_+$ is the set of all sets belonging to $\mathcal{A}$ which have a finite and positive Lebesgue-measure and are situated entirely in the interval $(0, +\infty)$, and $\mathcal{B}_-$ the set of all sets belonging to $\mathcal{A}$ which have a finite and positive Lebesgue-measure and are entirely situated in the interval $(-\infty, 0)$. Let us put further $P(A | B) = \frac{m(AB)}{m(B)}$, where $m$ denotes the Lebesgue-measure. The conditional probability space $\mathcal{B}$ can therefore be generated by a single measure. Now let $m_+(A)$ be equal to the Lebesgue-measure of that part of $A$ which is situated in the interval $(0, +\infty)$ and let $m_-(A)$ be equal to the Lebesgue-measure of the set $A$ if $m_+(A) = 0$, and $m_-(A) = +\infty$ if $m_+(A) > 0$. Then we have clearly $P(A | B) = \frac{m_+(AB)}{m_+(B)}$ for $B \in \mathcal{B}_+$ and $P(A | B) = \frac{m_-(AB)}{m_-(B)}$ if $B \in \mathcal{B}_-$, and thus our conditional probability space is also generated by the set $(m_-, m_+)$ of two dimensionally ordered measures; (that fact that $m_-$ and $m_+$ are dimensionally ordered, is evident as if $m_+(B) < +\infty$ we have $m_-(B) = 0$ by definition). Thus the space $\mathcal{B}$ can be generated either by the single measure $m$ or by the two measures $m_-$ and $m_+$ which are dimensionally ordered. This example shows that without supposing the additivity of $\mathcal{B}$ the number of elements of the set $\{\nu_r\}$ is also generally not determined.

§ 2. Random variables and their distributions

A random variable on a conditional probability space $\mathcal{B} = [S, \mathcal{A}, \mathcal{P}]$ is defined as a real-valued function $\xi = \xi(s)$ ($s \in S$) which is measurable with respect to $\mathcal{A}$. Clearly $\xi = \xi(s)$, when considered on the ordinary probability space $\mathcal{B}_B = [S, \mathcal{A}, \mathcal{P}_B(A)]$ where $B \in \mathcal{B}$ is fixed and $\mathcal{P}_B(A) = P(A | B)$, is an ordinary random variable. The probability distribution of $\xi$ with respect to the probability space $\mathcal{B}_B$ is called the conditional probability distribution of $\xi$ with respect to $B$. The probability distribu-
tion of \( \xi \) can however be investigated also from an other point of view. Let us denote by \( E \) the set of all real numbers, by \( A \) the set of all Borel-sets of \( E \) and if \( x \in A \), by \( \xi^{-1}(x) \) we denote the set consisting of those \( s \in S \) for which \( \xi(s) \in x \). Let \( B \) denote the set consisting of those \( \beta \in A \) for which \( \xi^{-1}(\beta) \in B \). If \( B \) is not empty, putting \( \pi(x|\beta) = P(\xi^{-1}(x)|\xi^{-1}(\beta)) \) if \( x \in A \) and \( \beta \in B \), we obtain a conditional probability space \([E, A, B, \pi]\).

If \( B \) is an additive class, then so is \( B \), and thus the conditional probability space \([E, A, B, \pi]\) can be generated by a dimensionally ordered set \( \{\nu_{\gamma}\} \) of measures. If the conditional probability space \([S, G, \mathfrak{B}, P]\) is generated by the set \( \{\nu_{\gamma} \in \Gamma\} \) of measures, let us denote by \( B_{\delta} \) the set of those \( \beta \in B \) for which \( \xi^{-1}(\beta) \in \mathfrak{B}_{\delta} \) and by \( \Delta \) the subset of \( \Gamma \) consisting of those \( \delta \in \Gamma \) for which \( B_{\delta} \) is not empty. In this case \( B_{\delta} \) is also an additive class, and we have \( \nu_{\delta}(\beta) = \pi(x|\beta)\nu_{\delta}(\xi^{-1}(\beta)) \), if \( x \in A \) and \( \beta \in B_{\delta} \), where \( \delta \in \Delta \). For an arbitrary \( \delta \in \Delta \) we may call the set function \( \nu_{\delta} \) the distribution of dimension \( \delta \) of the random variable \( \xi \). If \( \Delta \) consists of a single element, i.e. the space \([E, A, B, \pi]\) is generated by a single measure \( \nu \) and all sufficiently large intervals \((-\infty, a) \cup (a, +\infty) \) belong to \( B \) there exists a non-decreasing function \( F(x) \) defined for \(-\infty < x < +\infty \) such that if \( [a, b) \) denotes the interval \( a \leq x < b \), we have \( \nu([a, b)) = F(b) - F(a) \). In this case the function \( F(x) \) may be called the generalized distribution function of the random variable \( \xi \). If \( F(x) \) is absolutely continuous in any finite interval of the real axis, we may call its derivative \( f(x) = F'(x) \), the generalized density function of the random variable \( \xi \).

Every distribution \( \nu_{\delta} \) of \( \xi \) of given dimension \( \delta \) is determined only up to a positive constant factor; thus the generalized distribution function \( F(x) \) of \( \xi \) (if it exists) is determined only up to a linear transformation; moreover if \( F(x) \) is a distribution function of the random variable \( \xi \) then \( cF(x) + d \) where \( c > 0 \) and \( d \) is an arbitrary real number, is also a distribution function of \( \xi \). A generalized distribution function is a non-decreasing function, continuous to the right: \( F(x) \) may not be bounded. It is evident that if \( F(x) \) is a (generalized) distribution function of \( \xi \) and \([a, b) \) is an interval which belongs to \( B \) we have for \( a \leq a_1 \leq b_1 \leq b \)

\[
P(a_1 \leq \xi < b_1 | a \leq \xi \leq b) = \frac{F(b_1) - F(a_1)}{F(b) - F(a)}.
\]

Clearly if the generalized density function \( f(x) \) of \( \xi \) exists, it is determined only up to a positive constant factor.

A generalized density function \( f(x) \) is always a non-negative measurable function, which is integrable in any finite interval; of course

\[
\int_{-\infty}^{+\infty} f(x) \, dx \text{ may not exist. If } [a, b) \text{ belongs to } B \text{ we have for } a \leq a_1 \leq b_1 \leq b
\]

\[
P(a_1 \leq \xi < b_1 | a \leq \xi \leq b) = \frac{\int_{a_1}^{b_1} f(x) \, dx}{\int_{a}^{b} f(x) \, dx}.
\]
The generalized distribution function and density function of vector-valued random variables can be defined similarly.

REFERENCES