CHAPTER II

GENERAL LABEL SPACE

A. Distribution Function (Discrete Case). Measure-Theoretical Approach (Sections 1-3)

1. Introduction

In Chapter I we studied collectives with a discrete label space. To each point of a finite or countably infinite set S of labels a_i, a probability $p_i \geqslant 0$ was assigned (probability as explained in Sections 1–5) and such that $\Sigma_i p_i = 1$. We proved that for each (finite or countable) subset A of S, a probability $P(A)$ exists[1] and $P(A) = \Sigma_{a_i \in A} p_i$; in this way a completely additive set function P was defined on the σ-field of all subsets of S with $P(0) = 0$, $P(S) = 1$. For these collectives (whose consistency, in particular with respect to the randomness assumption, was the subject of Appendix One) four operations were defined which allow us to derive new collectives from given ones.

We shall also have to consider non-countable label spaces S. We do not attempt to assign probabilities to all possible subsets of S as was possible in the case of a discrete label space since this would clearly lead to difficulties. We also expect that the family of sets which have probabilities, in the sense of frequency limits, will not coincide with the family to be used in the measure-theoretical approach. We shall limit ourselves, in general, to the case of one dimension, a limitation which had no importance in discrete label spaces.

A convenient way of describing any probability distribution is by means of the *distribution function* (d.f.) or *cumulative distribution function* (c.d.f.). Hence, we shall begin this chapter by introducing (Section 2) the d.f. for a discrete label space. This does not present any difficulty.[2]

---

[1] We use now P rather than p as in Chapter I.

[2] This d.f. [defined by R. v. Mises, "Fundamentalsätze der Wahrscheinlichkeitsrechnung," *Math. Z.* 4 (1919), p. 20] could have been introduced in Chapter I. There, however, it would have been an essentially formal addition. A main value of the d.f. is to unify the discrete and the continuous distributions.

For a non-discrete label space, the most important particular case is the one in which a probability density exists (a concept quite parallel to mass density) and often this is taken as the starting point for the study of continuous label spaces. However, the way of introducing the density at a point x of an interval (a, b) as the limit of $P(A)/A$, as $A \to 0$ (where A is an interval always containing the point x) presupposes the definition of a sequence of set functions $P(A)$, each corresponding to a collective, which has the appropriate distribution and it does not seem easy to do this in a rigorous way. On the other hand, to introduce continuous distributions only as analytical approximations to discontinuous distributions does not lead to the desired understanding of the set function $P(A)$ [or the d.f. $F(x)$] in the continuous case.

For the sake of better understanding, we shall present briefly some of the main aspects and results belonging to the "axiomatic" or "measure-theoretical" approach[3] (Section 3). Some knowledge of this theory which is, in certain respects, mathematically simpler than ours and has been more fully investigated, will assist us and will also serve as a basis for comparison. *Our* approach is outlined in Sections 4–7.

Sections 3–7 are perhaps not as easy to read as most other parts of this book. A reader not interested in the foundations may choose to skip the material of Sections 3–7 and accept continuous distributions as convenient mathematical approximations to discontinuous distributions, defined by a continuous density $p(x)$ over an interval (a, b) (this density being analogous to a mass density) where then $\int_a^b p(x)\,dx = 1$ (in analogy to $\sum_i p_i = 1$ in Chapter I) and $\int_{-\infty}^{x} p(t)\,dt = F(x)$ is the cumulative distribution function, in analogy to Eq. (1) of this chapter. We hope, however, that most readers, and in particular future teachers, will not follow this course but will study Chapter II carefully; it contains an introduction to probability as a science and leads to a mathematical foundation which differs from the usual one.

2. *Cumulative Distribution Function for the Discrete Case*

Let $a_1, a_2, \ldots$ be the labels to which probabilities $p_1, p_2, \ldots$ belong. The *distribution function* (d.f.) or *cumulative distribution function* (c.d.f.), $F(x)$, is defined by

$$F(x) = \sum_{a_i \leqslant x} p_i . \tag{1}$$

[3] In Appendix Two we explained what we understand by frequency approach as opposed to the "measure-theoretical" or "abstract" approach.

This $F(x)$ which is continuous to the right: $F(x + 0) = F(x)$, has jumps of magnitude p_i at the points a_i, and $p_i = F(a_i) - F(a_i - 0)$. [The notation $f(x + 0)$ or $f(x - 0)$ means, as usual, with $h > 0$: $\lim_{h\to 0} f(x + h)$ and $\lim_{h\to 0} f(x - h)$].[1] In the interval between two jumps, $F(x)$ is constant. If there are only k values $a_1, a_2, \ldots, a_k$, $F(x)$ is a step line which is equal to zero for $x < a_1$, then equal to p_1 for $a_1 \leqslant x < a_2$, equal to $p_1 + p_2$ for $a_2 \leqslant x < a_3 \cdots$ and reaches the value one at $x = a_k$; thus, there are $k + 1$ intervals of constancy. If there are countably many a_i and if the a_i are, e.g., the positive integers, there are countably many intervals of constancy; if the a_i are everywhere dense in [0, 1], the only intervals of constancy are to the left of $x = 0$ and to the right of $x = 1$ (see Section 7.1). In general, however, if we speak of a discrete distribution we think of a "uniformly discrete" distribution for which the distance between any neighboring a_i remains above a given δ.

Another way of writing $F(x)$ is by means of the Heaviside *unit step function* $\epsilon(x)$ defined for all x by

$$\begin{aligned} \epsilon(x) &= 0 \qquad \text{for} \qquad x < 0 \\ &= 1 \qquad \text{for} \qquad x \geqslant 0. \end{aligned} \tag{2}$$

Then $\epsilon(x - a_i)$ has a jump of height one at $x = a_i$ and we obtain for $F(x)$

$$F(x) = \sum_i p_i \, \epsilon(x - a_i). \tag{3}$$

Here x appears in the usual way as the argument in ϵ.

According to the preceding definition, the d.f. $F(x)$ has the following meaning: *$F(x)$ is the probability of a result which is less than or equal to x.* We may also use the term *random variable* (r.v.) occasionally employed in Chapter I and denote by $F(x)$ the probability that the random variable ξ is less than or equal to x; this means the same as: "the result of the trial is less than or equal to x." *$F(x)$ may then be designated as the d.f. of the r.v. ξ.* One speaks of a "discrete r.v.," of a "normally distributed r.v.," etc., meaning that $F(x)$ is a step line or a normal d.f., etc. The terms *chance variable* or *stochastic variable* are synonymous with random variable[2] (Fig. 1).

---

[1] More explicitly, let $f(x)$ be defined on $a \leqslant x < b$. We write $f(x + 0) = A$ if there exists a number A such that for every $\epsilon > 0$ there is a $\delta > 0$ where $|f(t) - A| < \epsilon$ if $x < t < x + \delta$; similarly, for $f(x - 0)$.

[2] The term random variable was used first by F. P. Cantelli, "La tendeza ad un limite nel senso del calcolo della prohabilitá," *Rend. Circ. Mat. Palermo* **41** (1916), pp. 191–201. [In the same paper "convergence in probability" (see Chapter IV) is

In the discrete case ξ takes on the values $a_1, a_2, \ldots$ only and with probabilities $p_1, p_2, \ldots$, respectively. We denote by p_i the jumps or steps at a_i and *by* P (rather than by p, as in Chapter I), the set function which

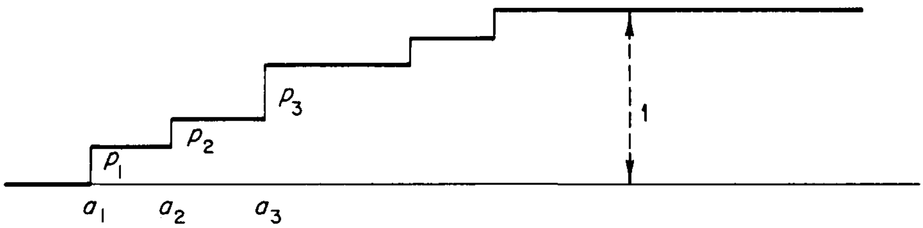

FIG. 1. Cumulative distribution function. Discrete case.

defines the probability distribution of the problem under consideration. Thus, if A_x denotes the semi-infinite interval $\xi \leqslant x$ open to the left, then

$$F(x) = P(A_x), \qquad p_i = F(a_i) - F(a_i - 0) \tag{4}$$

or

$$F(x) = \Pr\{\xi \leqslant x\}.^3 \tag{4'}$$

It follows that

$$F(-\infty) = 0, \qquad F(+\infty) = 1, \tag{5}$$

where

$$F(-\infty) = \lim_{x \to -\infty} F(x), \qquad F(+\infty) = \lim_{x \to +\infty} F(x);$$

in fact, $A_x \to 0$ as $x \to -\infty$, and A_x is the whole real line as $x \to +\infty$. Denoting by A the interval $x_1 < \xi \leqslant x_2$ we have

$$P(A) = F(x_2) - F(x_1). \tag{6}$$

Since $P(A) \geqslant 0$ it follows that $F(x_2) \geqslant F(x_1)$. Also, $F(x)$ is continuous to the right. In fact, within any interval of constancy, $F(x)$ is continuous, and at a_i

$$F(a_i) = F(a_i + 0). \tag{7}$$

We review our results in a form which (except for the final sentence) will remain correct in more general cases: *The d.f.* $F(x)$ *is a never*

introduced.] It is a matter of habit whether one prefers to describe certain properties as properties of the distribution $F(x)$ or as properties of the random variable ξ. We shall not necessarily always use Greek letters for random variables.

[3] By Pr{ } we mean the probability of the "event" in parentheses; we write, however, $P(A)$ or $p(A)$ rather than $\Pr\{A\}$ if the argument is A.

decreasing point-function, everywhere continuous to the right and tending toward zero as $x \to -\infty$, toward one as $x \to +\infty$. It has, at most, countably many points of discontinuity. It determines uniquely the discrete probability distribution, with probabilities p_i at points a_i, and vice versa.

3. Non-Countable Label Space. Measure-Theoretical Approach

3.1. *The first four axioms.* In the preceding section the label space S was denumerable. The family $\mathscr{S}$ of subsets of S to which probabilities were assigned consisted of *all* subsets of S. (see Chapter I and Appendix One). In the non-countable case, if S is, for example, a finite interval one cannot assign probabilities to all subsets of the interval since this would clearly lead to inconsistencies. In the axiomatic, measure-theoretical approach one postulates existence and essential properties of an additive set function $P(A)$ over an appropriate family T of subsets A of S. Subsets A are also called *events*. Our own approach and results will be better understood if we first explain some essentials of this approach, where the condition that $P(A)$ be a distribution over a collective does not appear.

Let S be a set of elements, T a set of subsets of S. Obviously one wants T to be a *field*. (1) T contains S as an element; (2) if A and B belong to T the same must be true for $A + B$, AB, and A' and B', where $A' = S - A$ denotes the complement of A.[1] As explained in Chapter I, $A + B$, also called *sum* or *union*, is the set which consists of all elements of S belonging either to A or to B or to both; AB, called *product* or *intersection*, is the set of all elements belonging to both A and B. A' contains all elements of S which do not belong to A. Since T contains S it must contain S', which is the *void* or *zero* set, occasionally denoted by v. From the above it follows that T contains the sum and the product of any finite number A_i of sets of T.

A set function P is defined[2] on a family T of subsets of S if to every set $A \in T$ a non-negative number $P(A)$ is assigned, the value of P at A. With respect to this P, which will be called the *probability distribution* on the field T, the following axioms are made:

(1) *To each A in T corresponds a non-negative (n.n.) number $P(A)$ called the probability of the event $x \in A$.*

(2) $P(S) = 1$.

---

[1] These requirements are not independent as seen by $(AB)' = A' + B'$. Also $(A + B)' = A'B'$.

[2] Kolmogorov [17], p. 2ff, p. 13ff.

(3) *If A and B are disjoint sets of T, then $P(A + B) = P(A) + P(B)$.*

(4) *$P(A)$ is a completely additive (or countably additive or σ-additive) set function; i.e., if a set A of T is the countable union of sets $A_i \in T$, such that $A = A_1 + A_2 + \cdots$, where $A_i A_j = 0$, $i \neq j$, $i, j = 1, 2, 3, \ldots$, then*

$$P(A) = P(A_1) + P(A_2) + \cdots .$$ [3]

The completely additive n.n. set function $P(A)$ on T forms the probability distribution on T.

Axiom (4) *of complete additivity* or *σ-additivity is equivalent to an axiom of continuity,* which may be regarded as a statement on the interchange of two passages to the limit:

Continuity axiom:

(4′) *Consider a decreasing sequence B_i of sets of T.*

$$B_1 \supset B_2 \supset B_3 \supset \cdots \supset B_n \supset \cdots ,$$

whose intersection approaches the zero set as $n \to \infty$; then

$$\lim_{n\to\infty} P(B_n) = 0.$$

It can easily be seen that (4′) as well as (4) follows from the other axioms in the case of T containing finitely many sets[4] $A_1, A_2, \ldots, A_n$ only. It can also be proved that (4) and (4′) are equivalent,[5] i.e., that a σ-additive n.n. set function has the above continuity property and that a n.n. finitely additive set function which has the continuity property is σ-additive.

3.2. *Extension*. The above axioms and assumptions do not imply that T is a *σ-field,* or *Borel field.* (See Chapter I, Section 7.3). A σ-field is defined by the property that all countable sums ΣA_i of *sets A_i of T likewise belong to T.*[6] If T (a system of subsets of S including S) is not a σ-field, then P is not necessarily defined on the union ΣA_i, although defined on each A_i. It seems desirable to Kolmogorov that the σ-additive set function P be defined on a σ-field as in the discrete case.

This is achieved by the use of the Banach-Caratheodory-Kolmogorov extension theorem which states that the distribution $P(A)$ on T where

[3] This equation may also be written: $P[\Sigma_n A_n] = \Sigma_n [P(A_n)]$.

[4] In this case we speak of finite additivity or just additivity.

[5] See, for example, Kolmogorov [17], p. 14ff [from (4′) to (4)], or de la Vallée Poussin [27], pp. 90–91, Gnedenko [10], p. 44.

[6] It follows that the intersection of countably many sets of T belongs to T.

T and P satisfy Kolmogorov's axioms can be extended to a distribution over a σ-field T_σ. A few words must suffice: if T is not a σ-field there exists a smallest σ-field T_σ containing T.[7] If the distribution $P(A)$, which we know to satisfy all axioms, is defined on T there exists a unique non-negative completely additive set function $P^*(A)$ on T_σ, which again satisfies the axioms such that for the sets A of T we have $P^*(A) = P(A)$.[8] In this way, a σ-additive set function is defined on the σ-field T_σ.[9]

Relating to this extension, we quote a significant comment of Kolmogorov (*loc. cit.*, p. 16):

"Even if the events[10] of T can be interpreted as actual and (perhaps only approximately) observable events, it does not follow from this that the events of the extended field admit of such an interpretation.[11] Thus, there is the possibility that while a probability distribution P on T may be regarded as the idealized image of actual random events the extension P^* on T_σ remains merely a mathematical construction ... . However, if a deduction which uses idealized events of T_σ leads to the determination of the probability P of an event of T, this determination will have a real meaning."[12]

A particularly important σ-field T_σ consists of all "Borel sets" on the real axis. The *Borel sets*,[13] defined by E. Borel, form a very important

---

[7] Hausdorff [13b], p. 79 ff. The set T_σ contains the countable unions $\sum A_i$ of sets A_i of T, and, in general, also other sets not contained in T.

[8] See Kolmogorov [17], pp. 15, 16, also Loève [20], p. 87 ff, Doob [6], p. 599.

[9] The reader should note that T_σ is not an arbitrary σ-field but a smallest σ-field over a field T. Not every σ-field F_σ contains a field F such that the smallest σ-field over F equals F_σ. Kolmogorov does not state (and it would not be true) that an additive set function can be constructed over an arbitrary σ-field. We could say equivalently: Kolmogorov's T_σ is a σ-field which admits a set function P satisfying his axioms. This restricts—adequately—the admitted σ-fields. In Kolmogorov's work it is, however, not always clear whether he uses his preferred term "Borel field" for any σ-field or only for those σ-fields for which a set function obeying his axioms exists. An example of a T_σ is the $\mathscr{S}_n$ of Chapter I, Section 7.3. An example of an F_σ which is not a T_σ is the field of all subsets of the closed unit interval: there exist non-measurable sets.

[10] "Event" is a subset A of S belonging to T.

[11] It is, however, left open by Kolmogorov what his T and P look like (see to this point also last subsection of Appendix Three) and why this probability field "may be regarded as the image of actual random events" in contrast to P^* on T_σ. There are σ-fields which admit a frequency interpretation and non-σ-fields which do not.

[12] We return to this last point in Section 6.4, pp. 91.

[13] The name has been proposed by Lebesgue [18], p. 117 ff. See also E. W. Hobson, *The Theory of Functions of a Real Variable*, Vol. 1, 3rd ed., p. 179ff, 1927. For the definition of Borel measurable sets see Cramér [4], p. 14, Lebesgue [18], p. 117, also de la Vallée Poussin [27], p. 33, etc. The term "Borel set" is often used in a more general sense than the one given here.

class of "Lebesgue measurable" or briefly "measurable" sets. They are constructed in the following fashion: Consider, first, an interval open or closed or half open; it is "measurable" in practically any sense; the sum of two intervals is also measurable, but it is not always an interval. Suppose we start with a group of finitely many or infinitely many intervals and apply to them a finite or an infinite number of times the operations of summation (union) and of multiplication (intersection). The resulting sets are interesting on account of their derivation from intervals; they have been called "Borel measurable" or simply *Borel sets*.[14] The Borel sets are Lebesgue measurable (but the converse is not true). It can be proven that *the infinite sum and the infinite product of a sequence of Borel sets is again a Borel set*, i.e., that the Borel sets form a σ-field (and actually the smallest σ-field over the intervals). If then, for example, S is a (finite or infinite) interval on the real number axis and T is the system of all Borel sets of S, and if $P(A)$ is a non-negative completely additive set function on T, then $P(A)$ is a σ-additive set function on a σ-field.[15]

3.3. *Converse question.* We shall now show, from another point of view, why the particular choice of T ($T =$ all Borel sets on the real axis) is mathematically advantageous.

Consider a c.d.f. $F(x)$, introduced as before [Eq. (4)] by

$$P(A_x) = F(x), \tag{4''}$$

where A_x is the interval $-\infty < \xi \leqslant x$, $F(x)$ is *non-decreasing, continuous to the right, and such that Eqs.* (5) *hold.* (We shall return to these properties in Section 7.) Now we want to discuss the important *converse question.* We ask *whether, given an arbitrary distribution function $F(x)$,* i.e., a function which possesses the just-named properties, *we can indicate a family of sets A such that $F(x)$ determines a completely additive set function $P(A)$ over the sets of this family and which for $A = A_x$ reduces to the $P(A_x)$ of* (4'').

[14] Denoting by subscripts σ and δ "sums" and "intersections" we have: the infinite intersection C_δ of closed intervals is closed, the infinite sum 0_σ of open intervals is open; the C_σ are in general not closed the 0_δ are in general not open. C_σ is, however, closed and 0_δ is open if the operations σ or δ apply to finitely many sets only. Next we can apply the operations δ and σ to the C_σ and 0_δ and obtain $C_{\sigma\delta}$ and $0_{\delta\sigma}$; all of these are Borel sets. The C_σ are complements of the 0_δ, the $C_{\sigma\delta}$ complements of the $0_{\delta\sigma}$.

[15] Kolmogorov's field T and therefore T_σ are left unspecified. Cramér decides to identify $T = T_\sigma$ with the σ-field of the Borel sets on the real axis. Many follow his lead.

If we know $F(x)$, we know $P(i)$ where i is an open, half open, or closed interval. Indeed, we have from our definition of $F(x)$

$$\Pr\{a \leqslant \xi \leqslant b\} = F(b) - F(a - 0), \qquad \Pr\{a < \xi \leqslant b\} = F(b) - F(a), \quad \text{etc.} \tag{4'''}$$

We thus know an interval function $P(i)$. Can we define a fairly general family T of sets A which contains the intervals, such that a completely additive set function is defined over T and reduces to the given interval function $P(i)$ if A is an interval? These problems have been studied by Lebesgue [18] and de la Vallée Poussin [27]. The latter calls such a family T of sets "normal" or, in a more suggestive way, "measurable with respect to $F(x)$." The result is as follows: given a distribution $F(x)$, we *can* construct, for a family of sets A, a completely additive set function $P(A)$ which reduces to the given interval function $P(i)$ if A is an interval, but this family depends on $F(x)$ (see [27], p. 108; [18], p. 168). *However, although it depends on* $F(x)$, *it contains always* (i.e., for any F) *all Borel sets*. Therefore, whatever F may be, $P(A)$ is uniquely determined on the Borel sets and is σ-additive over this σ-field. This fact recommends the Borel sets as a field of definition for the set function $P(A)$. (See footnote 15, p. 57.)

B. Non-Countable Label Space. Frequency Approach (Sections 4-7)

4. The Field of Definition of Probability in a Frequency Theory

4.1. *The problem of Chapter I. Projected extension.* We begin with some generalities. In Chapter I we considered a discrete sample space. This is the origin of probability theory and many very important problems are in a discrete sample space. But in other important and meaningful problems discrete distributions do not suffice.

The problems of Chapter I were concerned with a finite number n of trials and a finite or countable number of labels. We could ask, for example[1], for the probability of casting "6" at least once in three trials. The label space consists then of 6^3 points with coordinates x_1, x_2, x_3 where each $x_i = 1, 2, \ldots, 6$. To each of them there corresponds a probability $p(x_1, x_2, x_3)$. Each of the 216 possible results does or does not satisfy the conditions of the problem, which define a subset A_3 of the

[1] See many examples of problems of this type in Chapter IV.

label space. The result 1, 1, 6 belongs to A_3, the result 3, 1, 2 does not. The probability corresponding to A_3 is found by mixing. In the similar problem of finding the probability $P(A_n)$ of at least one 6 in n trials, the attribute space consists of 6^n points. If, finally, $p_1, p_2, \ldots$ with $\Sigma_i p_i = 1$ are probabilities for the countably many attributes 1, 2, ... and we again ask for the probability $P(C_n)$ of obtaining 6 at least once in n trials, the problem is still of the same type as the others and, if independence prevails, $P(C_n) = 1 - (1 - p_6)^n$. In general, i.e., without independence there is given an n-dimensional collective with probability distribution $p(x_1, x_2, \ldots, x_n)$, $x_i = 1, 2, 3, \ldots$ and we obtain $P(C_n)$ by the appropriate mixing (see Chapter I, Section 7.2.)

This label space S_n has been called discrete. *We have assigned probabilities to all subsets A of the discrete n-dimensional label space.* It has been proved that *these subsets form a field, even a σ-field,* which we denoted by $\mathscr{S}_n$. *In all problems heretofore considered, the decision whether a specific sequence of results does or does not satisfy the conditions of the problem is made after at most n trials.* Our probability theory would, however, be too narrow if we were bound to assign probabilities to such properties only when the above-named "decision" follows after a given number of trials. It is necessary to include problems for which n may increase without limit. This will lead us beyond the discrete label space. It is equally clear geometrically that we have to go beyond the discrete label space. Of course we shall be faced with the question to which non-countable label sets probabilities should be assigned, or, in other terms, with the question of what types of probabilistic problems may be considered.

Since we regard probability theory as a mathematical science (whose field of application in physics, biology, psychology, economics, etc., is increasing steadily) we postulate (see Appendix Two, p. 45) that, *corresponding to any probability statement, there exists, at least in principle, i.e., conceptually, an (approximate) verification by means of some frequency experiment.* This should be acceptable to the scientist and to the mathematician. If we say that the probability of casting "3" with a certain die is $\frac{1}{6}$, we anticipate that, if the die is tossed 1000 times or 10,000 times, we shall find, approximately, the corresponding proportion of three's. An analogous remark applies if we state that the death probability of a 40-year old American man has a certain value, and it applies likewise to any more general statement concerning insurance probabilities. Likewise, if we say that in Brownian motion the probability of finding at a certain time in a certain optically limited space a number x of particles equals $a^x e^{-a}/x!$ (with "a" an appropriate constant), such a statement is scientifically meaningful only inasmuch as it is in some way verifiable, *directly* *or indirectly by its consequences.* This "conceptual verifiability," which applied to all problems of Chapter I will be a guiding principle in our generalizations and will be reflected in our notion of a probability over a general label space.

Our program is extremely simple. Since we have to consider indefinitely increasing n, we must introduce a space B more general than $T = \Sigma \mathscr{S}_n$, the label space of the collective (Chapter I, Section 7.3). We have then to interpret the points of T and all the sets A consisting of such points in this space B. Such spaces, B, of countably many dimensions will be defined in the next subsection. Among the subsets of B are Borel sets, measurable (Lebesgue) sets which are not Borel sets, and non-measurable sets. From among all subsets of B we shall distinguish certain sets closely related to the sets of T, namely, those subsets M of B each of which lies between two sets of T whose probabilities differ arbitrarily little [see Eqs. (12) and (13), Section 5.2]. Such and only such sets M of B will be assigned probabilities; it will be seen that only those probabilities allow a frequency interpretation and are (conceptually) verifiable. We believe that this is as far as one should go, and no farther. *The sets M form a field F_1 which is the field for any frequency theory of probability.*[2] It is not a σ-field.

The reader may be interested in the relation between this proposed probability field F_1 and Kolmogorov's (Sections 3.1 and 3.2). Actually, his field T_σ remains unspecified except for the postulate that it should be a σ-field over which a set function obeying his axioms can be defined. Such σ-fields are, for example, our field $\mathscr{S}_n$ of the collective of Chapter I; or the σ-field of all Borel sets on the real line; the σ-field of all measurable sets, with distributions satisfying Kolmogorov's axioms. Kolmogorov repeatedly quotes as an example the field of Borel sets; Cramér, who adopts Kolmogorov's foundations decides, in [4] and [5], to choose as basic field $T = T_\sigma$, the field of the Borel sets, on the real axis, and others follow his lead. This is then a clear decision; but it is not Kolmogorov's.

Our field F_1 with its distribution P will be defined in Sect. 5. While the field T of the collective is denumerable, F_1 has a cardinality $f > c$ ($=$continuum). We can not go further, and, in particular, we shall not consider the smallest σ-field over F_1.

To see one aspect of the relation to Kolmogorov's program still more clearly, we add: (1) If Kolmogorov's field T does not contain all sets of F_1, then his extended field T_σ does not either. But T_σ will contain sets not contained in F_1, sets which admit no frequency interpretation. (2) Denote by F_2 the field of all (Lebesgue) measurable sets. If Kolmogorov's T is a proper part of F_2, then T_σ is also a proper part of F_2. If, in particular, T consists of all Borel measurable sets then $T = T_\sigma$ and it overlaps F_1. The cardinality of this T is that of the continuum.

---

[2] Our presentation will borrow decisive ideas from E. Tornier (see Appendix Three). We have also learned much from A. Wald (his theorems III and IV, pp. 46–47, of the paper quoted in Appendix One).

4.2. Sets of sequences. Basic sets. In the previously mentioned discrete space S_n of 6^n points these points have as coordinates all possible segments of length n which can be formed by means of the numbers 1, 2, ..., 6. For example, the set of points defined by the condition "the first three results are equal to 1, 2, and 3, respectively" consists of the 6^{n-3} points whose first three coordinates are equal to 1, 2, 3, in this order, while the remaining $n - 3$ coordinates are equal to any of the numbers 1, 2, ..., 6. The consideration of problems where n may increase without limit leads to the following generalization. We define a *set of sequences* or *space of sequences* (of the die), denoted by B or by B_6, as *the* non-countable *totality of all infinite sequences which can be formed by means of the numbers* 1, 2, ..., 6. This is an immediate generalization of the n-dimensional label space S_n. Each single infinite sequence forms a point in this space of sequences. We may consider B as a storehouse for the attribute spaces S_n belonging to any number n of trials. In the same way, we define the space of sequences $B = B_2$ of the "coin." It contains all infinite sequences which can be formed by means of the numbers 0 and 1 (hence all binary numbers). A very important B is the space $B = B^*$, whose points are all sequences formed by the positive integers.[3]

Compared with the definition of the label spaces (lattices) S_n, the definition of B is an abstraction, essentially of the same type as, for example, the definition of a continuous line in contrast to sets of discrete points. This comparison is actually more than an analogy, since we shall see that the inclusion of infinite n, as in the definition of B, is equivalent to the introduction of a continuous sample space. Many readers may also consider the latter to be an abstraction since, strictly speaking, all observations lead to discontinuous results; nevertheless, we cannot afford to forego the advantages of continuity.

A *basic set of order n* is the set of those and only those sequences (of the respective B) which have the same initial segment of length n.

A basic set of order three of B_6, for example, consists of the sequences

$$(461) = 4, 6, 1, x_4, x_5, \ldots, \qquad x_j = 1, 2, \ldots, 6, \qquad j = 4, 5, \ldots .$$

A basic set of order n in B is a point in S_n (the label space of Chapter I). Any basic set of order n can be expressed as the sum of basic sets of order $m > n$. If the order of a basic set is bounded, the basic set is called "of bounded order" or, briefly, "bounded." We see that a basic set is a neighborhood of each of its "points" in B. In agreement with usual

[3] Consider also in Appendix Three example (2) regarding the game with dice and tetrahedron (p. 99). There, a more specific B may be used to take care of the given dependence. See, however, Section 5.3, p. 76.

definitions, the sum of any number of neighborhoods (= basic sets) will be called open.[4]

There are countably many basic sets in any B. Each basic set contains non-countably many sequences = points. We consider B itself to be a basic set of order zero and we also include the empty set without assigning it an order.

The reader will easily prove that two basic sets of the same B can never overlap: *either they are disjoint or one is entirely contained in the other*. Hence, *the intersection of two basic sets of B is a basic set of B*. The analogous statement for the sum is, of course, not true.

We now repeat some facts, well known from Chapter I, in the new terminology. Consider the basic sets of order one. We denote by (i) the set of those and only those sequences of the appropriate B which begin with i. It is then clear that $B_6 = (1) + (2) + (3) + \cdots + (6)$, that $B_2 = (1) + (0)$, when the two attributes are 1, 0, and $B^* = (1) + (2) + \cdots$ when the attributes are the positive integers. To each basic set of order one we associate a non-negative number such that the sum of these numbers is one. We call these numbers the *contents* of the basic sets or the *probabilities*, denote them by P, and assign to B itself the number one, to the empty set v the number zero. With the notation $(1) = E^1$, $(2) = E^2$, ... we then have

$$1 = P(B) = P(E^1) + P(E^2) + \cdots, \tag{a}$$

which is, of course, the same as Eq. (3) of Chapter I. Any basic set of order one is the sum of basic sets of order two. For example, in the case of B_6, $(1) = (11) + (12) + (13) + \cdots + (16)$ and for B^*, $(1) = (11) + (12) + \cdots$. To each basic set E^{rs} of order two, the assigned probabilities must be such that if $E^r = \sum_s E^{rs}$, then

$$P(E^r) = \sum_s P(E^{rs}). \tag{a'}$$

(In the vocabulary of Chapter I this is the mixing rule.) In the same way we may decompose each E^{rs} into the sum of basic sets of order three, E^{rst}, and associate non-negative numbers with the E^{rst} such that

$$P(E^{rs}) = \sum_t P(E^{rst}), \tag{a''}$$

and so on. We obtain in this way the most varied probability distributions on B. Each such probability distribution defined by the compatible

[4] These explanations are given in order to relate our definition of "open" to the usual one. On p. 63 we give simplified definitions.

[i.e., satisfying (a), (a'), etc.] contents of the basic sets is called a *valuation* (see more on p. 83).

We repeat that the difference in approach as compared to Chapter I is that now we represent the result "1" in the tossing of a die by the basic set E^1 whose first result, "1," is fixed while all other infinitely many results could be any of the numbers 1, 2, 3, ..., 6. In Chapter I, the result "1" could, of course, also be considered as the totality of results 11, 12, 13, 14, 15, 16, or as the totality of results 111, 112, 113, ..., 166, and the corresponding relations for the probabilities are, in the notation of Chapter I: $p_1 = p_{12} + \cdots + p_{16} = p_{111} + \cdots + p_{166}$. In the present notation, the true die is given by $P(E^i) = \frac{1}{6}$, $i = 1, 2, ..., 6$, $P(E^{ik}) = \frac{1}{36}$, $i, k = 1, 2, ..., 6$, etc.

A set like A_3 in Section 4.1 now appears as the sum of $6^3 - 5^3 = 91$ basic sets, each of order three; the complement A_3' of A_3 is the sum of 5^3 basic sets of order three. To such a sum of basic sets we assign as content the sum of the contents of the basic sets, in agreement with Chapter I.

We defined before: any sum of basic sets (of neighborhoods) is called an *open set*. The complement of any open set is a *closed set*. The set B itself and the empty set are both open and closed. This is the same as in Euclidean space. *If O is any open set in B, there is one and only one system of greatest disjoint basic sets contained in B, such that O is their sum.* Indeed, two greatest basic sets contained in O are either identical or disjoint. The sum of the greatest basic sets contained in O is, therefore, O.

We call

$$O = E_1 + E_2 + \cdots, \tag{8}$$

where the E_i are the greatest disjoint basic sets[5] contained in O, *the normal form of O*. The normal form of the basic set of order n is its decomposition into basic sets of order $(n + 1)$ $(n = 0, 1, 2, ...)$.

While the decomposition (8) in a sum of greatest basic sets is unique it is, of course, possible to decompose an open set O in many ways into sums of basic sets. Regarding this, the following fundamental theorem holds,[6] valid under the above assumptions (a), (a'), ... of additivity:

If an open set O can be decomposed in two ways into a sum of disjoint basic sets

$$O = E_1 + E_2 + E_3 + \cdots = E_1' + E_2' + E_3' + \cdots,$$

[5] On p. 62 we denoted basic sets of order one, two, three, etc., by E^r, E^{rs}, E^{rst}, etc. In equations like (8) or in the relations of p. 63, etc., we shall denote a basic set by E_ν or E_i without implying anything about its order.

[6] See reference {1} at the end of Appendix Three.

then always

$$\sum_i P(E_i) = \sum_j P(E_j').$$

This is a deep lying theorem not proved here.

The following facts are well known.

(a) *The sum of any number of open sets is open* (from definition).

(b) *The intersection of finitely many open sets is open.* (Show this for two open sets.)

(c) *The intersection of infinitely many closed sets is closed* (or empty). Let $A_j = B - O_j$. The intersection of the A_j equals B minus the sum of the O_j.

(d) *The sum of finitely many closed sets is closed.* If $A_j = B - O_j$, $j = 1, 2$, are closed, then $A_1 + A_2 = B - O_1O_2$ is closed.

(e) *If $M \supset N$ and M is closed, N open, the difference is closed; if M is open, N closed, the difference is open.* In both cases, $M - N = M(B - N) = MN'$. In the first case this last product is closed; in the second case it is open.

So far, and in what follows in this chapter, the theory relates only to problems in the one-dimensional continuum, although no difficulties of principle are connected with the consideration of several dimensions. We would have to consider $B^{(m)}$ whose elements are ordered m-tuples of ordinary sequences. A basic set of order n of $B^{(m)}$ consists of those and only those elements of $B^{(m)}$ which have the same m beginnings of length n. To such basic sets probabilities are assigned, etc. The considerations of Section 4.4 generalize accordingly.

4.3. *The α-sets.* For any given B, we use the term *α-set* for *a sum A of basic sets whose complement $A' = B - A$ is also a sum of basic sets.*[7] We see immediately that all sets A considered in Chapter I (subsets of the field T) are α-sets, since they were sums of points in S_n or, in the present terminology, sums of *basic sets*, and the same was true of the complements A'. Hence all sets of $\mathscr{S}_n$ and of T are α-sets, but the converse does not always hold. It follows from our definitions that *an α-set is both open and closed.* (This property holds in a Euclidean space R_1 for the empty set and for R_1 only.) In fact, as a sum of open sets, A is open, and as a complement to a sum of open sets, it is closed. Conversely, we see that *any set A which is both open and closed is an α-set.* In fact, as A is open it is a sum of basic sets by (8), and since A is closed, A' is open and therefore a sum of basic sets by (8).

[7] The concept is due to Tornier.

We have seen that the basic sets do not form a field. However, *the α-sets form a field.* We call it F_0. In fact, let A_1, A_2 be two α-sets; then each is both open and closed, and hence, $A_1 + A_2$ is by (a) and (d) both open and closed. Let $A_1 \supset A_2$. If A_1 is considered open and A_2 closed, the difference is open; if A_1 is considered closed and A_2 open, the difference is closed by (e). Hence $A_1 - A_2$ is an α-set. To any α-set A we associate as content, or probability, the sum of the contents of the basic sets in the normal form of A. If the α-set is itself a basic set we recover the previous rules. If A is decomposed in several ways into a sum of basic sets and the content $P(A)$ of A is each time the sum of the contents of these basic sets which add up to A, the determination of $P(A)$ is unique (by the theorem quoted p. 64).

We introduce now *the field F of the α-sets of bounded order*, i.e., those α-sets whose normal form is of bounded order. We say that an α-set is of order n if n is the highest order among the basic sets of its normal form. If the α-set is of order n the same holds for its complement. The α-sets of order n form a field f_n which is obviously contained in f_{n+1} and, of course, in F_0; there is a sequence $f_1 \subset f_2 \subset f_3 \subset \cdots$, and we denote by F the system of all α-sets which belong to any f_ν. This F is a field and we have

$$f_1 \subset f_2 \subset f_3 \subset \cdots \subset F \subset F_0 .$$

All the f_n are σ-fields but $F = \Sigma f_n$ is not a σ-field. If A is an α-set of bounded order, the same holds for its complement A'.[8]

It can be proved that *when there are finitely many attributes in all trials, the fields F and F_0 coincide.* In other words: *in this case no unbounded α-sets exist.*

We indicate the proof by means of an example. Consider B_2 with labels 0 and 1 take A to be the infinite sum of basic sets

$$A = (1) + (001) + (00001) + (0000001) + \cdots$$

i.e., "1" stands in the $(2n + 1)$th place, $n = 0, 1, 2, \ldots$. Consider the sequence $P = 000000 \cdots$. It cannot belong to A: suppose it were contained in a basic set E of order n of the above expansion of A; this basic set would have to be $(00 \cdots 00)$, n times, and such a basic set of order n does not appear in A. Then P must belong to $A' = B - A$. Let the normal form of A' be $\Sigma E_\nu'$ and assume that P lies in E' of order n'. Then the initial segment of E' must consist of n' zeros. Any basic set of B_2 which begins with n' zeros must then lie in E'. But A contains basic sets with more than n' zeros. Hence A and A' could not be disjoint. This proves the impossibility of A being an α-set. The reader will have no difficulty in proving the theorem in the general case by constructing a sequence P which cannot lie in either A or A'.

[8] Note that "α-set of bounded order" is the same as "bounded sum of basic sets." But an arbitrary sum of basic sets is not always an α-set.

We saw at the beginning of this subsection that *the sets of* T, the field of the collective, *are* α*-sets*, and, of course, α*-sets of bounded order*, i.e., sets of F. *The converse also holds*: to a set A of F there corresponds a probability in the collective.

First, consider the case of a finite number of attributes. Then A, the α-set of bounded order, is always the sum of finitely many disjoint basic sets of a finite maximum order and the same holds for A'. The corresponding probability is that in an n-dimensional collective. We obtain $P(A)$ by adding those probabilities which correspond to the basic sets that add up to A. The reasoning is hardly modified if there are countably many attributes in at least one trial. The α-set A of bounded order is then the sum of countably many basic sets E_i of some maximum order n and we write A as a sum of basic sets, *all* of order n:

$$A = \sum E_j'. \tag{9}$$

Denote by K' the n-dimensional collective whose labels include the points E_j' (remember that a basic set of order n is equivalent to a point in n-space). The label set of K' consists of all possible n-tuples, and the sum of the corresponding probabilities is *one* since K' is a collective. The probability of A equals the (infinite) sum of the probabilities of the above E_j'. Indeed, each E_j' stands for a label point of K' and it follows from Eq. (9), Chapter I, that

$$P(A) = \sum_{j=1}^{\infty} P(E_j'), \tag{10}$$

where the E_j' are the same as in (9). *Hence, to the* α*-set* A *of* F *there corresponds a probability in a collective. For the* T *of* Chapter 1, Section 7.3 $T = F$.

If, however, the normal form[9] of a set A (α-set or not) contains basic sets of unbounded order (for this to happen it is necessary but not sufficient that the normal form be an *infinite* sum of greatest basic sets), no finite maximum order exists. The corresponding attribute space would have infinitely many dimensions and to sets in such spaces we have so far not assigned probabilities. Thus, *the collective of Chapter I assigns probabilities to the sets of the field F within B and only to them.* We have thus obtained an algebraic characterization of the field of the

[9] Note that an open set O which is not an α-set cannot appear as an attribute in a collective since the complement of O would not be a sum of basic sets, while, in a collective, together with the probability of a set A, the probability of its complement A' is always defined.

collective of Ch. I. From now on *we denote this field by F rather than by T.* By means of the extension mentioned at the end of Section 4.1 we shall reach F_1, the final field of our theory, directly from F.

We have introduced the α-sets since they form a natural generalization of the basic sets and of the sets considered in Chapter I. They have other interesting properties as we shall see later (p. 86). However, *for building up our theory we need only the field $T = F$ of Chapter I.* The α-sets of unbounded order, the sets of $F_0 - F$, may be obtained from F like all other sets of F_1.

4.4. *Relation between B and the linear continuum.* We wish to define a probability distribution over the real line. We use the distributions in the space B as a guide, since in B every step has a probabilistic meaning. We now want to establish a "mapping" of a space B onto the linear continuum, the unit interval [0, 1] which we denote by E, and take here as closed to the left and to the right. Consider first the simplest case: the space B_2 of the "coin." *We associate to any infinite sequence of B_2,* viz., $x_1, x_2, \dots$; $x_j = 0, 1$; $j = 1, 2, \dots$ *the binary number,* $\cdot x_1 x_2 x_3 \cdots$ between 0 and 1, *and the point of E defined by it.* We obtain in this way all points of E. (Clearly, if instead of the infinite sequences, which are the points in our various spaces of sequences, we were to consider points in n-dimensional space only as we did in Chapter I, then such finite segments $x_1, x_2, \dots, x_n$ would correspond to certain discrete points on E only (in our example, 2^n points in all). Note that to any infinite sequence of B_2 there corresponds one, and only one, point on E. The converse is, however, not true. The sequence 0, 1, 0, 1, 1, 1, 1, 1, ... and the sequence 0, 1, 1, 0, 0, 0, 0, ... both correspond to the point $\frac{3}{8}$ and the same holds for any sequence which from a certain place on contains zeros only, i.e., for the binary rational numbers.

The image of a basic set of B is a binary interval on E, closed at both ends. It contains the images of all "points" of the basic set. E. g., to the basic set (101) corresponds the closed interval between $0.101000 \cdots = \frac{5}{8}$ and $0.1011111 \cdots = \frac{6}{8}$. To a basic set of order n corresponds a *division interval of order n.* Two adjacent closed division intervals which have geometrically a point in common, are considered as images of two disjoint basic sets since the common point corresponds to two different sequences of B. This is in agreement with the theorem that two basic sets are disjoint unless one is entirely contained in the other. To each division interval (d.i.) we assign a probability equal to the probability of the respective basic set. (The disivion intervals on E are not "open as well as closed"; this property of a basic set is not conserved in the present mapping.)

The image of an α-set of B_2 is a finite sum of division intervals. The complement of such a sum of closed division intervals is a finite sum of division intervals which are, however, open, at least on one side. We add to each of them the missing endpoints which thus appear twice. For example, the complement of the sum of the two closed division intervals $[0, \frac{1}{2}]$ and $[\frac{5}{8}, \frac{6}{8}]$ are the intervals $(\frac{1}{2}, \frac{5}{8})$ and $(\frac{3}{4}, 1]$. We augment each by joining to it the missing points, assuming, for example, that the sequence 0, 1, 1, 1, ... corresponds to the $\frac{1}{2}$ in $[0, \frac{1}{2}]$ and the sequence 1, 0, 0, 0, 0, ... to the left endpoint of $(\frac{1}{2}, \frac{5}{8})$. Since an α-set as well as its complement consists of finitely many intervals, we are adding finitely many points only, and since a point has content = probability zero,[10] the addition of finitely many points does not influence any contents. We have so far obtained division intervals and sums of division intervals. This corresponds to the sets of F.

We proceed in an analogous way for B_6 or in general for any B with a finite number of labels.

A one-to-one mapping of a different type *may be defined for* B^*, whose attributes consist of all positive integers. Let us consider this case in more detail. Let the unit interval be closed to the left, open to the right, and call it E. We divide E into countably many parts which we denote by (1), (2), (3), ..., each closed to the left and open to the right, and with the point 1 as the only point of accumulation. We call these intervals *division intervals* of first order. For example:

$$[0, \tfrac{1}{2}), \quad [\tfrac{1}{2}, \tfrac{3}{4}), \quad [\tfrac{3}{4}, \tfrac{7}{8}), \ldots .$$

Each point of E then lies in exactly one such division interval of first order. To each of these intervals $(\nu) = \delta_\nu$ we assign a number (which we call its content or probability) $P(\delta_\nu)$ such that $\sum_{\nu=1}^{\infty} P(\delta_\nu) = 1$. We then divide each of these division intervals of first order, δ_ν or δ_{ν_1}, in a similar way into infinitely many division intervals of second order, δ_{ν_2}, each again closed to the left, open to the right; for example, the d.i. $[0, \frac{1}{2})$ may be divided into $[0, \frac{1}{4})$, $[\frac{1}{4}, \frac{3}{8})$, $[\frac{3}{8}, \frac{7}{16})$, ... and similarly for all others.[11] The corresponding basic sets are (11), (12), etc., and the contents of the division intervals of second order (which equal the contents of the respective basic sets) are such that $\sum_{\delta_{\nu_2} \in \delta_{\nu_1}} P(\delta_{\nu_2}) = P(\delta_{\nu_1})$ for a subdivision of δ_{ν_1} into intervals δ_{ν_2}. Then, in the same way, we subdivide the intervals of order two, then those of order three, etc., *ad infinitum*.[12] An interval

[10] This is really an assumption. If we were to consider "irregular valuations," see p. 83, where a point might have a positive content, the present considerations would not hold.

[11] It is, however, not necessary that the rule of subdivision into d.i.'s of second order be the same as that for d.i.'s of first order, etc.

[12] The construction may also be arranged in the following way: We take the content of the basic sets of first order as lengths of the intervals of the first subdivision. Since the sum of the contents of these basic sets is one, the unit interval will be exactly filled

of the *n*th subdivision corresponds to a basic set of order *n*. The construction is such that any sequence of intervals of orders 0, 1, 2, ..., for which $E \supset \delta_{\nu_1} \supset \delta_{\nu_2} \supset \cdots$ converges toward one and only one point of E; any point of E then lies in exactly one d.i. of order one, in one and only one of order two, etc. Hence, it corresponds to a sequence of integers $\nu_1, \nu_2, \nu_3, \ldots$. For example, the sequence 3, 7, 4, 1, ... means that the point lies in the third d.i. of order one; in this d.i. of order one, it lies in the seventh d.i. of order two, etc. Thus to any infinite sequence consisting of natural numbers there corresponds in a one-to-one way a point on E. *The linear continuum E corresponds one-to-one to the space B^* of sequences.* The division intervals form a particular countable set of intervals, by no means all rational intervals. For more details see Section 6.3.

We stress the geometrical arbitrariness of the chosen system of division intervals (arbitrary except for the additivity conditions explained above). (a) We may start with an arbitrary system on [0, 1) and assign to each interval its natural length as valuation (=probability). We have then to assign the same valuations to the corresponding basic sets in B^*. If we assign valuations not equal to the natural lengths of the chosen d.i.'s but given by a function $p(x) \geqslant 0$, $\int_0^1 p(x)\,dx = 1$, then we assign these valuations to the basic sets in B^*. (b) If, on the other hand, we consider as given the probabilities of the basic sets in B^*, *and* an arbitrary system of d.i.'s on [0, 1), then this defines a function $p(x) \geqslant 0$, etc., such that the value of the probability assigned to each basic set equals the integral of $p(x)$ over the corresponding d.i.

Let us return to a finite case, say B_6, and discuss some aspects of the "dual" mapping we have just described for B^*. Think of the interval E as divided first into six parts, then each of the six parts into six parts, etc. corresponding to the basic sets of B_6. Then any small division interval on E corresponds to a sequence of n numbers $\nu_1, \nu_2, \ldots, \nu_n$ since it lies in the ν_1-interval δ_{ν_1} of the first division, in the ν_2-interval δ_{ν_2} of the division of δ_{ν_1}, etc. Now, instead of E think of a roller[13] divided into six parts by six lines parallel to the axis of the roller, each part similarly divided into six parts, etc. Now roll the roller and consider as result the point on the roller where, say, a fixed pin touches its surface at the moment the roller stops. We obtain then at once, by one rolling, a result which corresponds to the repeated tossing of a die. The finer the

up by these d.i.'s. In the same way, in the subdivision of any interval of first order the lengths of the respective intervals of second order are chosen equal to the contents of the corresponding basic sets of second order. Then each d.i. has as content its natural length.

[13] An exactly analogous experiment relates to the roulette wheel divided in the same manner.

division on the roller, the larger the number n of trials represented by each "result," which reads: "pin in intervals δ_{ν_r}, $r = 1, 2, ..., n$." From this intuitive point of view, qualitative distinction between the "discrete" and the "continuous" disappears. This way of considering things corresponds well to our concept of the collective where a single trial really does not enter into any consideration. It is clear how the "roller," explained here for B_6, can be adapted to B^*.

We remember (see the beginning of Section 3) that our purpose is to assign probabilities to certain subsets of a label space in a manner consistent with verifiability. Since we can map the spaces B onto the linear continuum, we can now continue by studying either the spaces B or the linear continuum. We propose to do both, keeping their connection in mind.

4.5. Basic sets and cylinder sets. Place selection. As an application we give here the proof of a lemma used in Appendix One. We considered there the binary numbers in $[0, 1] = U$, or equivalently, the space of sequences B_2. A place selection, s, is given by a sequence of increasing natural numbers $\{\alpha_i\}$ such that a number x of U is transformed into x' by the rule $x_{\alpha_1} = x_1'$, $x_{\alpha_2} = x_2'$, ... It follows that a point x', i.e., a single infinite sequence, is mapped into countably many sequences x, a set of measure zero. We denote by A' a set formed of sequences x', by A the set of all sequences x corresponding to the $x' \in A'$ and prove *that* $|A| = |A'|$, *where A is a measurable together with A'.*

A basic set A' of order n is originated from a set A consisting of all numbers x which have the terms numbered $\alpha_1, \alpha_2, ..., \alpha_n$ in common; such a set is called a *cylinder set.* It is clearly a finite or countable sum of basic sets; for example, in B_6 we mean by the cylinder set $(a_2a_5a_6)$ the sum of all basic sets which have a_2 at the second place, a_5 at the fifth place, and a_6 at the sixth place; hence $(a_2a_5a_6) = \Sigma_{x_1}\Sigma_{x_3}\Sigma_{x_4}$ $(x_1a_2x_3x_4a_5a_6)$, hence the sum of eight basic sets of order six. Its probability is, in our simple case, equal to $8(\frac{1}{2})^6 = (\frac{1}{2})^3$. We may write $P[(a_2a_5a_6)]$ $= P[(a_2)]P[(a_5a_6)]$.

In our transformation any basic set of order n in the x', of probability $(\frac{1}{2})^n$, corresponds to a cylinder set which is the sum of $2^{\alpha_n - n}$ basic sets of order α_n each, hence with probability $2^{\alpha_n - n}(\frac{1}{2})^{\alpha_n} = (\frac{1}{2})^n$. Hence if A' is a basic set of order n, of probability $(\frac{1}{2})^n$, the same probability holds for A. On the unit interval, the set A' is a closed binary interval of length 2^{-n} and A is the sum of closed non-overlapping binary intervals with sum of their lengths equal to 2^{-n}. If, next, A' is a finite sum of basic sets of order n, A is a finite sum of basic sets of order α_n each. On U: If A' is a finite sum of open or closed binary intervals, A is a finite sum of binary intervals plus a set of measure zero, and again $|A| = |A'|$. If, finally, A' is measurable we cover the points of A' on U by non-overlapping binary intervals of a total length approximating $|A'|$ and $U - A'$ by similar intervals whose length approximates $1 - |A'|$. The images of the two systems of intervals will have the same total length as the respective originals. The first image covers A, the second $U - A$, and the respective lengths are approximately $|A'|$ and $1 - |A'|$. Hence A is measurable and $|A| = |A'|$.

5. Basic Extension

5.1. *Examples.* In Chapter I we restricted ourselves to a finite number n of trials. We shall now discuss a few simple problems where this restriction does not hold and which will guide us.

(a) What is the probability that in the continued throwing of a die the "six" will never appear?

(a') What is the probability that in the continued throwing of a die the "six" will appear at least 1000 times?

(b) In the example of the cylindrical roller, Chapter I (or in any problem whose labels are the positive integers) what is the probability of obtaining either 1 on the first trial or 2 on the first and second trials or 3 on the first and second and third trials, and so on?

In the first example, (a), with q the probability of "six" and p the probability of "non-six," and writing now Y_n (instead of A_n) for the set of sequences which contain the six at least once in n trials, and Z_n (instead of A_n') we have with $P_n = P(Y_n), Q_n = P(Z_n)$ and assuming independence

$$P_n = P(Y_n) = q + qp + qp^2 + qp^3 + \cdots + qp^{n-1} = q\frac{(1-p^n)}{(1-p)} = 1 - p^n,$$

$$Q_n = P(Z_n) = p^n.$$

Now let $n \to \infty$. Certainly the $\lim_{n\to\infty} P(Z_n)$ exists and equals zero; also Z_n is an α-set of bounded order. But what about the limit of the set Z_n, the set $Z_\infty = Z$? The set Z is contained in Z_n, and $Z_1 \supset Z_2 \supset Z_3 \supset \cdots \supset Z_n \supset \cdots$ and since the probability of Z_n becomes as small as we please as n increases, the probability of Z cannot be different from zero. Probabilities of sets like Z have *not been defined so far.*[1] If we

[1] The set $Y = B - Z$ is a sum of basic sets, but they are not of bounded order. Hence, neither $P(Y)$ nor $P(Z)$ is a probability in the collective. Since the number of labels is finite, the fields F and F_0 coincide; but since Y is not of bounded order it cannot be an α-set. (Also Z does not contain any basic set, since in each basic set there are sequences in which "six" appears.)

call P, Q the limits of P_n, Q_n, respectively, we have $P = 0, Q = 1$. The question is whether we have the right to call P and Q probabilities of Y and Z. The customary use of P and Q is absolutely correct as long as we consider P (or Q) merely as *analytic approximations* for large n to genuine probabilities.

Problem (a′) is similar to (a) with the role of Y and Z interchanged.

Occasionally it will be convenient to have *a notation for a certain field together with its probabilities*. We call it probability field. We denote by Φ the *probability field* which belongs to F. Then: neither P nor Q of problem (a) are probabilities in Φ.

In example (b) the space of sequences is B^*. We denote the subset of B^* whose "probability" concerns us by

$$\gamma = (1) + (22) + (333) + \cdots \tag{11}$$

where (1), (22), ... are basic sets. We shall show that $P(\gamma)$ is *not a probability in Φ*. We show first that γ is an α-set. Now γ itself is a sum of basic sets by Eq. (11). What about $\gamma' = B - \gamma$? Since $B^* = (1)+(2)+(3)+ \cdots$, we have

$$\gamma' = [(1) - (1)] + [(2) - (22)] + [(3) - (333)] + \cdots .$$

We consider (3), for example, as a sum of basic sets of order three; among them the set (333) appears, and hence (3) — (333) is a sum of basic sets. The same holds for every expression in brackets. Thus, γ' like γ, is a sum of (countably many) basic sets and γ is an α-set. It is, however, of unbounded order, and hence it does not belong to F but to $F_0 - F$.

In example (a) we had $F = F_0$ but Z was not an α-set. Here γ is an α-set but not in F. In both cases the set under consideration has not probability in Φ.

We have decided that our final criterion as to whether a probability can be meaningfully ascribed to a set is the question of "verifiability" of this number. Consider Y, the set of sequences with at least one "six." If we repeatedly make $n = 1000$ trials there will be at least one six in the overwhelming majority of these groups of 1000 trials each, and if $n = 10{,}000$, this will hold in an even more striking way. We cannot indicate an n large enough that *each* segment of length n may be specified as the initial segment of a sequence of Y or of Z. But for the majority of segments we can make this decision.

For a problem in Φ a precise *decision after a certain number n of trials* is possible *for each single sequence*; in other words, if $A \subset F$ it can always be decided after at most n trials whether a segment is the beginning of a sequence of A or not. For Y, this does not hold for each individual sequence. Nevertheless, the relative frequency of Y can be determined with ever increasing accuracy as n increases. We shall indeed show that for Y (and Z) a "probability" can be determined (and the method of determination will be given) although the "decision" after n trials is not possible for each individual sequence of B. In the case of the set γ in $F_0 - F$, there is a "decision" for every individual sequence after a *finite* number of trials, but we cannot indicate a number n of necessary terms as we could for sets of F. Further discussion of the examples is in Sections 5.2 and 6.2.

5.2. *Extension of the field of definition of the collective.* So far, the collective has assigned probabilities to the sets of F. We wish to assign probabilities to the sets of a more general field[2] F_1 which should contain sets like Y and γ of Section 5.1.

Given F together with its probabilities, hence Φ, we define:

A subset M of B which is not a set of F has probability if for any $\epsilon > 0$ there exist sets α_n and β_n of F (which, therefore, have probabilities) *such that*

$$\alpha_n \subset M \subset \beta_n \tag{12}$$

$$P(\beta_n) - P(\alpha_n) < \epsilon. \tag{13}$$

If this holds (for any arbitrarily small ϵ) we put

$$P(M) = \lim_{n\to\infty} P(\alpha_n) = \lim_{n\to\infty} P(\beta_n). \tag{14}$$

The probability of M so determined is unique and can be verified as accurately as we please by means of probabilities in Φ. In Section 6.1 we shall see that these new sets (which include the sets of F) form a field, which we shall call F_1. The new probability field F_1 *and* the probabilities corresponding to the sets of F_1 will be denoted by Φ_1.[3]

[2] This extension is completely different from the one discussed in Section 3 of this chapter (Kolmogorov-Banach extension theorem). The smallest σ-field over F or F_0 is without importance for us. (This smallest σ-field over F has the cardinality of the continuum, c, while the cardinality of the F_1 which we are going to define is $f > c$. If here and in the following cardinal numbers are occasionally mentioned this is meant as additional information for the reader familiar with these concepts. Such sentences may however be disregarded by other readers.)

[3] This extension is due to E. Tornier who has used the idea (which goes back to Jordan) since 1932 (see quotations at the end of Appendix Three) as an essential tool in his theory and suggested its application in the v. Mises theory. The extension (12)–(14), on the straight line, is by Jordan. The field F_1 appears in Tornier's book [25], in A. Wald's work (see his theorems III and IV in the paper quoted in Appendix One), in papers by de Finetti, and in a less outspoken way in papers by Copeland. For detailed quotations see Appendices One and Three.

Let us review our extremely simple procedure: *we start with the field T of the collective of Chapter I.* This field, in B, we call F remembering its characterization as the field of α-sets of bounded order and we write Φ for the corresponding probability field. *From Φ we derive—by means of* (12) *to* (14)*—the new probability field Φ_1 which will be the final probability field of our theory.*

It should be understood that an iteration of the operations (12)–(14) does not lead us beyond Φ_1. More explicitly: if in (12) to (14) we replace the α_n, β_n of F by sets γ_n, δ_n of $F_1 - F$ (i.e., sets which were obtained by (12) to (14) with the aid of some α_n, β_n of F), then the set N for which $\delta_n \supset N \supset \gamma_n$, $P(\delta_n - \gamma_n) < \epsilon$, again belongs to F_1. We can, in fact, replace γ_n and δ_n by sets $\bar{\gamma}_n$ and $\bar{\delta}_n$ of F which, in both size and content, differ from γ_n and δ_n by as little as we please; and we have then obtained N by means of sets of F without the detour via γ_n, δ_n.[4]

We shall see that the field formed by the new sets together with the previous sets of F is *the field of the sets of B which "have content in the sense of Peano and Jordan."*[5]

The term "content" is used here in the more general sense where we assign to an interval i an interval function $P(i)$ which is in general not equal to the length of the interval; to sets A we assign a set function $P(A)$ its p-content, which is not necessarily equal to the natural measure of this set. We shall, however, avoid as much as possible the heavy term p-content and speak of the "content" of a "Jordan set." Anticipating the most characteristic property of these Jordan sets, we state that for them and only for them is the measure of the boundary zero (see Section 6.1, p. 84).

As a first example of the application of (12) and (13), consider the set γ of Eq. (11) which we have seen to be a set of $F_0 - F$. Let $\gamma = E_1 + E_2 + \cdots$ where the E_j, $j = 1, 2, 3, \ldots$ are disjoint basic sets. Write, with γ' the complement of γ,

$$\gamma = \sum_{j=1}^{\infty} E_j , \qquad \gamma' = \sum_{i=1}^{\infty} \mathscr{E}_i . \tag{15}$$

[4] In particular, we obtain the same Φ_1, whether we start with sets of Φ or of Φ_0.

[5] C. Jordan, *J. Math.* (4), 8 (1892), pp. 76–79.

G. Peano, *Applicazione geometriche del calcolo infinitesimale*, pp. 154–156, 158. Turin, 1887.

H. Lebesgue [18].

For textbook presentations, see:

H. Kestelman, *Modern Theories of Integration.* Oxford, 1937.

F. Hausdorff [13a].

E. W. Hobson, *The Theory of Functions of a Real Variable*, Vol. 1, p. 179 ff. 1927.

W. Rudin, *Principles of Mathematical Analysis.* New York, 1953.

Both γ and γ' contain basic sets of unbounded order. We set now (with B for B^*)

$$\alpha_n = \sum_{\mu=1}^{n} E_\mu , \qquad \gamma_n' = \sum_{\nu=1}^{n} \mathscr{E}_\nu , \qquad \beta_n = B - \gamma_n' . \tag{16}$$

Here α_n and γ_n' are α-sets of F and β_n also belongs to F, since it is obtained by subtracting γ_n'—which is of bounded order—from B. Now, obviously,

$$\alpha_n \subset \gamma \subset \beta_n \tag{17}$$

as in (12) and α_n as well as β_n have probabilities in Φ. Also,

$$P(\beta_n) - P(\alpha_n) = P(\beta_n - \alpha_n) = P[B - (\gamma_n' + \alpha_n)] = 1 - P(\gamma_n' + \alpha_n). \tag{18}$$

On account of (16) and since $\gamma + \gamma' = B$, we have $P(\alpha_n + \gamma_n') \to 1$, as $n \to \infty$. Hence, the difference $P(\beta_n) - P(\alpha_n)$ can be made as small as we please by taking n sufficiently large. Thus, conditions (12) and (13) are satisfied. Hence, *in the extended domain of definition*, Φ_1, or briefly, *in* Φ_1, *a probability of* γ *is determined by means of probabilities in* Φ. We can write

$$\lim_{n\to\infty} P(\alpha_n) = P(\lim_{n\to\infty} \alpha_n) = P(\gamma). \tag{19}$$

Here γ is an α-set in $F_0 - F$ and $P(\gamma)$ exists in Φ_1.

Consider now example (a) where Y was not an α-set. Denote by Y_n the sum of the first n basic sets in the normal form of Y; this Y_n has probability in Φ, and $P(Y_n)$ approaches 1 as n increases. Therefore,

$$Y_n \subset Y \subset B \tag{20}$$

$$P(B - Y_n) < \epsilon, \tag{21}$$

for n sufficiently large. Hence, the probability of Y, and consequently of Z, exists in Φ_1 and we have

$$\lim_{n\to\infty} P(Y_n) = P(\lim_{n\to\infty} Y_n) = P(Y), \tag{22}$$

and similarly for Z. In Φ we could merely say that the set Z is contained in any Z_n of $F(=F_0)$. In Φ_1 a limit set Z exists and has a probability and this probability is determined by probabilities in Φ. We remember that for a set M of $F_1 - F$ there exist infinitely many individual sequences for which a decision as to whether such a sequence belongs to M or to M' is not possible. *However, the content, the probability of* M *is uniquely determined and with any desired accuracy, on account of* (13).

It should be well understood that we did not abandon our frequency definition of probability when we made the step from T, the field of the collective of Chapter I, to B, the space of sequences. In Chapter I we considered groups of n trials. Such a group of n trials can indeed be repeated N times and the frequency N_{A_n}/N of the property A_n in question tends toward the probability of A_n as N increases.

If A is a set of F_1, *an analogous interpretation of* $p(A)$ *as lim* N_A/N *holds mathematically*; but N and N_A relate then *to points of* B, the set of sequences, each point representing an *infinite* sequence of trials, and such a sequence of infinitely many trials cannot be considered "repeatable." By considering $p(A)$ as the $\lim_{N\to\infty} N_A/N$ (which is mathematically correct) we would really strain the idea of the collective beyond its clear meaning.

The situation remains the same if we interpret the points of B as points on the unit interval E (Section 4). As long as n is finite the label space consists of a fixed, finite or countable system of points on E, some of them belonging to A_n , and after at most n trials it can be decided whether a result belongs to A_n or not. But $n \to \infty$ corresponds to "picking an arbitrary point" of E and deciding whether it belongs to A or not. Some thinking shows that this "arbitrary picking and deciding" (both E and A being continua) is no longer clearly defined—apart from the difficulty of "approximating" the frequency of a continuum by N experiments.

We much prefer the clear-cut procedure given by us in this section where *the probability of a set of* $F_1 - F$ *has been defined with any desired accuracy*, not by the inconceivable repetition of infinite sequences, but *by means of probabilities of sets of* F [see Eqs. (12)–(14)]. The sets of F_1 are precisely those whose probabilities *can* be determined by means of the directly observable frequencies of sets of F.

5.3. *Further discussion of the extension.*

(a) *Arbitrary dependence.* We know that the n-dimensional collective in the discrete label space S_n comprises the most general dependence in n dimensions. A distribution assigns a probability to each single point of S_n and by choosing these appropriately we can realize any dependence. In the present case a valuation is defined by probabilities assigned to basic sets (each basic set comprising a continuum of sequences = points). Again any desired dependence can be taken care of.

To fix the ideas consider the following game: one throws a die (I) and a tetrahedron (II). The faces of II are denoted 1, 2, 3, 4. The result of each trial is the number appearing on the bottom face of either body. The first throw is with I. If the result of the nth throw is even the $(n+1)$th throw is with II, otherwise it is with I. In this example of dependence, a 5 or 6 can never follow an even number. Among the basic sets of order n of B_6, there are some which are thus "forbidden," namely, those whose defining sequence contains a forbidden succession. To such basic sets we assign probability zero, to the others we assign probabilities in accordance with the game. The sum of all basic sets which in this way obtain the value zero is an open set O of measure zero (there exist not more than countably many basic sets alltogether). Of course, a not-forbidden basic set of order n contains forbidden sequences. If the forbidden combination (a 5 or 6 after an even number) appears for the first time at the mth term, $m > n$, then this basic set of order m is contained in that of order n and to this smaller set is assigned probability zero.

Consider, for example the event A that 6 appears at most 15 times in the indefinitely continued game. A is a set of F_1. Denote by A_n the set which contains 6 at most 15 times in n games and by $P_n = P(A_n)$ the corresponding probability. Then, since $A_n \supset A_{n+1} \supset \cdots$ the intersection A of all the A_ν exists (cf. Section 5.4) and $P(A) = \lim_{\nu\to\infty} P(A_\nu)$ is defined.

We note that in the same way we can construct B_6, say, from any B_k, $k > 6$, and any B_k can be derived from B^*. Hence, *all we need* to obtain any variety of B's is B^*. This remark is, however, of a theoretical nature.

(b) *Content and measure.* We return to definitions (12) and (13) and give them another but equivalent form using now J instead of P for content. For an arbitrary set M of B the *outer content or exterior content* $\bar{J}(M)$ is defined as the greatest lower bound of contents $J(\beta)$ for all sets β in F which contain M. The *inner content* or *interior content* $\underline{J}(M)$ is defined as the least upper bound of the contents $J(\alpha)$ for all sets α of F contained in M.

$$\bar{J}(M) = \operatorname*{fin\,inf}_{\beta\supset M} J(\beta), \qquad \underline{J}(M) = \operatorname*{fin\,sup}_{\alpha\subset M} J(\alpha). \tag{23}$$

The set M *has content* if

$$\underline{J}(M) = \bar{J}(M). \tag{23'}$$

If (23') holds we can find to any ϵ two sets α and β in F such that $J(\beta) - J(\alpha) < \epsilon$, where $\beta \supset M \supset \alpha$; hence (12) and (13) hold. Assume on the other hand the validity of (12) and (13) for some M. Then, there are two sets, $\bar{\alpha}$, $\bar{\beta}$ in F such that $\bar{\beta} \supset M$, $\bar{\alpha} \subset M$, and $J(\bar{\beta}) - J(\bar{\alpha}) \leqslant \epsilon$. Then $J(\bar{\beta}) \leqslant J(\bar{\alpha}) + \epsilon$ and (23') follows since ϵ can be taken arbitrarily small. A set with content will be called a *J-set*. The J-sets form the field F_1.

Let us also consider the "measurable" sets in B which do not necessarily have content. [These sets are denoted by Cramer as "P-measurable" ([4], Chapter VI); also the term "weight" is used, and de la Vallée Poussin calls them "normal." "Normal" or "P-measurable" is the generalization of measurable (Lebesgue) and "P-measure" generalizes Lebesgue measure. The Lebesgue-Stieltjes integral takes the place of the Lebesgue integral. In analogy to "P-measurable," one may use *p-content* as defined in (23) and (23′), or in (12) and (13).]

Let us sketch the definition of P-measurable sets in B. We begin again by assigning non-negative numbers, $|E| \geqslant 0$ to the basic sets E of B and $|E|$ is briefly called the measure (rather than P-measure) of E. If $|E^r|$ is the measure of a basic set E^r of order one and $|E^{rs}|$ that of the basic sets E^{rs} of order two which add up to E^r, we have $|E^r| = \Sigma_{(s)} |E^{rs}|$, etc. Next, to any open set O of B, one assigns as measure $|O|$ the sum of the measures of the basic sets in the normal form of O. It can be shown that one obtains the same $|O|$ if O is decomposed in any other way into a sum of disjoint basic sets. Together with O, the complement $B - O$ is measurable. (The open sets played no role in our theory of content. There, the sets of F had the decisive importance.)

For any set L in B one then defines the *outer measure* $\bar{L}$ and the *inner* measure $\underline{L}$:

$$\bar{L} = \operatorname*{fin\,inf}_{O \supset L} |O| . \tag{a}$$

Thus: the *outer measure* of L is defined as the greatest lower bound of the measures $|O|$ of all open sets which contain L. Hence, any set L has an outer measure.

The inner measure L of $\underline{L}$ may be defined by

$$\underline{L} = |B| - \overline{B - L}$$

or, in analogy to (a),

$$\underline{L} = \operatorname*{fin\,sup}_{C \subset L} |C| . \tag{b}$$

The *inner measure* of any set L in B is the least upper bound of the measures of the closed sets C which are contained in L.

If for a set L the equality

$$\underline{L} = \bar{L} \tag{c}$$

holds, then L is called measurable ("P-measurable" if the distinction matters) and $\underline{L} = \bar{L} = |L|$ is called the measure ("P-measure") of L.

A second definition of P-measurability is analogous to our (12) and (13): A set is P-measurable (for a chosen valuation P, i.e., for chosen measures of the basic sets) if to any ϵ, no matter how small, one can find an open set $O_\epsilon \supset L$ and a closed set $C_\epsilon \subset L$ such that $|O_\epsilon - C_\epsilon| < \epsilon$. One can then prove that whatever valuation has been chosen, all Borel sets are P-measurable (see Section 3). We shall in general say "measurable" rather than "P-measurable" just as we say "content" rather than "p-content."

The measurable sets form a σ-field with we call F_2. It follows from the given definitions that F_1 is contained in F_2.[6]

[6] The field of Borel sets is contained in F_2 and overlaps F_1.

Let us now consider a set of $F_2 - F_1$, i.e., a measurable set without content. Such a set cannot be enclosed between sets of F or of F_0 (with "decision after a finite number of trials") or even of F_1; otherwise it would belong to F_1. It can merely be approximated "from above" by open sets O and "from below" by closed sets C, which are not both α-sets.[7] We see that *for sets of* $F_2 - F_1$ *conceptual verifiability is, in principle, impossible.* Examples follow in Section 5.5.

5.4. *Limits of sets and limits of measures of sets.* Before giving examples of sets of $F_2 - F_1$ we wish to discuss the following point of importance. Not only must we distinguish between (verifiable) content and (non-verifiable) measure. We must, first of all, distinguish between convergence of a sequence of sets and convergence of their respective contents or measures.

Consider a sequence of sets, $A_1, A_2, A_3, \ldots$. Denote by A_∞ *the set of those elements which belong to infinitely many* A_i and by B_∞ *the set of all elements which belong to almost all* A_i (i.e., to all except finitely many). If $A_\infty = B_\infty$ the sequence of sets is called *convergent*: $A_\infty = B_\infty = \lim_{n\to\infty} A_n$. It follows from our definitions that $A_\infty \supseteq B_\infty$ and that, A_∞, B_∞ remain unchanged if a finite number of sets A_i is added or subtracted or modified.

There are two cases in which the convergence is obvious. If the A_i form *a non-decreasing sequence*, $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$, the sequence[8] converges and $\lim_{n\to\infty} A_n = A_\infty = B_\infty = A_1 + A_2 + \cdots$. Likewise, if the A_i form a *non-increasing sequence*, $A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots$, then $\lim_{n\to\infty} A_n = A_\infty = B_\infty = A_1A_2A_3\cdots$, the intersection of the A_i.

In the general case, with

$$\sigma_n = A_n + A_{n+1} + A_{n+2} + \cdots, \qquad \delta_n = A_nA_{n+1}A_{n+2}\cdots,$$

we set

$$\overline{\lim_{n\to\infty}}\, A_n = (A_1 + A_2 + A_3 + \cdots)(A_2 + A_3 + \cdots)(A_3 + \cdots)\cdots = \sigma_1\sigma_2\sigma_3\cdots$$

$$\underline{\lim_{n\to\infty}}\, A_n = A_1A_2A_3\cdots + A_2A_3\cdots + A_3\cdots + \cdots = \delta_1 + \delta_2 + \delta_3 + \cdots.$$

This shows that $\overline{\lim}_{n\to\infty} A_n = A_\infty$ is the set of points which are contained in infinitely many A_n, while $\underline{\lim}_{n\to\infty} A_n = B_\infty$ is the set of points contained in all A_n from a certain n on; these definitions are the ones given at the start. (The reader may construct elementary examples of non-convergent sequences of sets.)

---

[7] It may however be approached from *one* side by α-sets.

[8] We write here $\subseteq$ and $\supseteq$ for extra clarity.

The relation between these concepts and the limits of the sequence of measures is contained in the following inequalities,[8] where $\underline{\lim}_{n\to\infty} \mid A_n \mid$ means the limit inferior of the $\mid A_n \mid$, etc.:

$$\mid \varliminf_{n\to\infty} A_n \mid \leqslant \varliminf_{n\to\infty} \mid A_n \mid \leqslant \varlimsup_{n\to\infty} \mid A_n \mid \leqslant \mid \varlimsup_{n\to\infty} A_n \mid . \tag{24}$$

If, in particular, the sequence $A_1, A_2, \ldots$ is convergent, then the following simple and important relation holds:

$$\lim_{n\to\infty} \mid A_n \mid = \mid \lim_{n\to\infty} A_n \mid . \tag{24'}$$

In the various limit procedures of probability calculus *it can happen that the sequence of measures converges without convergence of the corresponding sequence of sets*; if the sets converge towards a measurable limit set, we are assured by (24′) that the measures also converge, and these measures become contents if contents exist. But even if the A_n are sets of F_1 and converge toward a limit set, we cannot be sure that the limit set is in F_1. It could be in F_2 or have no measure. If it has measure and is in $F_2 - F_1$, the equality (24′) holds as a statement on measure only. We shall find this last situation in customary formulations of the strong law of large numbers (Chapter V, Section 4.4), in the so-called Borel-Cantelli theorems (Chapter V, Section 4.5), etc. The first-named possibility, where the measures converge but the sets do not converge, is typical of most of the so-called "limit theorems" of probability calculus (Chapter VI). We shall now analyze in this respect a very well-known limit theorem proved in Chapter V, Section 8.5.

Assume that a simple alternative, i.e., a trial with 0 and 1 as the only possibilities, like the tossing of a coin, is repeated n times. The probabilities for 0 and 1 are p and q, respectively. Denote by x the number of times 1 occurs in n trials. We ask for the probability of the inequality

$$nq - \lambda\sqrt{npq} \leqslant x \leqslant nq + \lambda\sqrt{npq} . \tag{25}$$

For each n this inequality holds on a set A_n of F and we set $P_n = P(A_n)$. What happens to A_n as $n \to \infty$? We are going to show that a limit does not exist. Let us first reason in the following intuitive way: The limit set of A_n, if it exists, cannot contain any basic set since each basic set also contains sequences which sometimes violate the inequalities (25).

[8] For details, the reader may wish to consult some book on measure theory; see also Loève [20].

Hence, the measure of this limit set would have to be zero and not equal to the Gaussian integral $\lim_{n\to\infty} P_n = (1/\sqrt{2\pi}) \int_{-\lambda}^{\lambda} e^{-z^2/2}\, dz$. Therefore, *this integral is certainly not the probability of the limit set* of the A_n.

We shall now show that a limit of the A_n *does not exist.* A limit of a sequence of sets exists only if each element which belongs to infinitely many of the A_n belongs to all of them, except perhaps finitely many, since we have defined convergence by $A_\infty = B_\infty$. If we can produce a single sequence which infinitely often remains within the bounds (25) and infinitely often violates (25) (i.e., which belongs to infinitely many A_n and is outside infinitely many others of them) the sequence $\{A_n\}$ is not convergent.

In the construction of such a sequence, we simplify our task in an unessential way by taking $p = q = \frac{1}{2}$, and $\lambda = 2$. Then (25) becomes: $n/2 - \sqrt{n} \leqslant x \leqslant n/2 + \sqrt{n}$. We make our construction by starting with a one and continuing with ones until the inequalities are violated for the first time; then we write zeros until we are again within the bounds of (25), etc. Thus we obtain (the stars denoting a segment where the inequality is violated)

1	$\frac{1}{2} - \sqrt{1} \leqslant 1 \leqslant \frac{1}{2} + \sqrt{1}$	
1, 1	$1 - \sqrt{2} \leqslant 2 \leqslant 1 + \sqrt{2}$	
1, 1, 1	$\frac{3}{2} - \sqrt{3} \leqslant 3 \leqslant \frac{3}{2} + \sqrt{3}$	
1, 1, 1, 1	$2 - \sqrt{4} \leqslant 4 \leqslant 2 + \sqrt{4}$	
1, 1, 1, 1, 1	$\frac{5}{2} - \sqrt{5} \leqslant 5 > \frac{5}{2} + \sqrt{5}$	*
1, 1, 1, 1, 1, 0	$3 - \sqrt{6} \leqslant 5 \leqslant 3 + \sqrt{6}$	
1, 1, 1, 1, 1, 0, 1	$\frac{7}{2} - \sqrt{7} \leqslant 6 \leqslant \frac{7}{2} + \sqrt{7}$	
1, 1, 1, 1, 1, 0, 1, 1	$4 - \sqrt{8} \leqslant 7 > 4 + \sqrt{8}$,	*

The next violation follows for $n = 11$, etc. By this procedure, we find a sequence which satisfies the inequalities (25) infinitely often and violates them infinitely often.

The $\lim_{n\to\infty} A_n$ does not exist. However the probabilities P_n converge toward a limit. Thus, the situation is as follows: *Purely analytical investigations* have shown that, under certain conditions, *the probability* $P_n = P(A_n)$ *of* (25) *is well approximated* by the Gaussian integral. The sets A_n are sets of F. The limit theorem gives an analytical approximation to the probability P_n, valid for large values of n. *But a limit set A, whose probability would be given by the Gaussian integral, does not exist.*

The same situation prevails for the Poisson formula (Chapter IV, Section 5): it is an analytic approximation to the Bernoulli formula as n increases and the event probability q decreases simultaneously. The result $a^x e^{-a}/x!$ is neither a probability nor a measure but an approximation for large n.

5.5. Examples of sets in $F_2 - F_1$. Earlier (in Section 5.1) we asked for the probability that in the continued tossing of a die, the 6 will appear at least 1000 times and we found that this probability exists in Φ_1 and is meaningful in a frequency theory: the set whose probability we want is in F_1 and its probability can be determined with any desired accuracy.

In contrast to such problems, consider the following ones:

(a) We ask for the probability $P(\omega)$ that in the continued tossing of a die, 6 appears a finite number of times only. Here, ω is the set of all those sequences which contain 6 only finitely often, a well-defined subset of B_6. However, for no finite segment of results, no matter how long, is it possible to decide whether it is the beginning of a sequence of ω or of $B - \omega$. Thus ω does not belong to F_1. No matter how large an n we choose and how many series of length n we observe, we cannot find an approximation to $P(\omega)$. The set ω is in $F_2 - F_1$.[9]

(b) As a second example, imagine a "die" with ten faces and a uniform distribution of the ten results 0, 1, 2, ..., 9. We ask for the probability that in the continued tossing of that "die," the resulting sequence represents a rational number. If we think of an irrational sequence as being an infinite non-periodic decimal fraction, the character of the problem becomes clear. No matter how many trials we make, we can never judge whether a segment of length n, consisting of the digits 0, 1, 2, ..., 9, is the beginning of a rational or of an irrational sequence, of a sequence belonging to R, the set of rationals, or to $B_{10} - R$. Both problems are such that "decisions" after a finite number of trials (and infinitely many trials we cannot make) are *in principle* impossible. Sets playing the role of the α_n and β_n in (12) and (13) do not exist. There seems to be no way even to formulate an analogous problem for finite n. The question which corresponds to our problem on the unit interval E is, of course, to ask for the "probability" of "hitting" a rational number on E. In no way can it ever be decided whether a number picked out is rational or not, since every point on E lies arbitrarily close to rational as well as irrational numbers. However, the lack of a frequency interpretation is more clearly evident in B_{10} than in E.

Both the set R of this problem and the set ω of the previous problem are not only measurable but even Borel sets. (In the last problem, the measure is zero. That of ω depends upon the convergence or divergence of the series $p_1 + p_2 + \cdots$, where p_n is the probability of 6 in the nth trial.) However, neither of these sets has content; neither admits a verifiable probability. We shall return to these examples in Section 6.2.

[9] The outer and the inner contents of ω are different. Should it be desired to assign a number to ω, any number between $\underline{J}(\omega)$ and $\bar{J}(\omega)$ can be chosen. But it is not a probability in the frequency sense.

6. The Field F_1

6.1. *Main properties of the sets of F_1.* The sets with content, the J-sets, have been defined by (12) and (13) and, equivalently, by (23) and (23′). These sets form a field, F_1.

We wish to stress that our definitions, (12) and (13), which depend explicitly on the contents of α_n and β_n, show that the field F_1 depends on the contents assigned to the original basic sets. (A similar remark applies to F_2. Nothing of this sort applies to F or F_0 or to the field of Borel sets.) The dependence is however very weak: consider any set B of sequences. Let us call a *valuation*, i.e., a choice of contents of the basic sets, *regular* if no single sequence (= point) in B has a positive content and no basic set has content zero. Then the statement holds: *All regular valuations of B lead to the same field F_1* (see proof at the end of this subsection).

The field F from which we started is denumerable (that means it contains denumerably many sets). *The field F_1 is not denumerable.*[1] As said before, F_1 is contained in the σ-field of all measurable sets, and it overlaps the σ-field of all Borel sets (which has cardinality c) as can be seen by examples.

We mention now a few further properties of the sets of F_1. (See proofs in the literature quoted on p. 74 or in Tornier's papers, quoted in Appendix Three). Some of the statements are almost obvious; some indications of their proofs will be given.

(a) *For any set M:* $\bar{J}(M) \geqslant \underline{J}(M)$

(b) *A basic set has content. An α-set has content.* The sets of F have content by definition; the contents of the α-sets of $F_0 - F$ are obtained from them by means of (12) and (13).

(c) *Let $M_1, M_2, \ldots, M_n$ be a finite number of disjoint J-sets. Then $\sum_{i=1}^{n} M_i$ is a J-set and*

$$J(M_1 + M_2 + \cdots + M_n) = J(M_1) + J(M_2) + \cdots + J(M_n).$$

(d) *The J-sets form a field* (this may also be concluded from theorem (h) of p. 84). The proof is omitted.

---

[1] F_1 has the cardinality $f > c$ (c = continuum). Consider the "Cantor set" A of B_6 which contains no "6" (see p. 71) in the independent continued throwing of a correct die. It has content (see pp. 75 and 86) it has the power of the continuum and measure zero (easily proved: see for example, Hausdorff [13a], pp. 245 and 416). Any subset of A has content. The set of these subsets—and therefore, *a fortiori* F_1—has cardinality $f > c$.

(e) *If a J-set is the countable union of disjoint J-sets M_i*, i.e., *if* $M = M_1 + M_2 + \cdots$, *then*

$$J(M) = J(M_1) + J(M_2) + \cdots .$$

This follows from the analogous theorem on measurable sets together with the fact that if a set has content it has measure and the two numbers coincide.

The following is a complement to (e), similar to the theorem of Chapter I, Eq. (9).

(e′) *Under the conditions of* (e) *any subsequence of the sum of the M_i has content.*

Denote now by M^* the greatest open set contained in a set M or, equivalently, the sum of all basic sets contained in M. It follows from (23) that $\underline{J}(M) = \underline{J}(M^*)$, and since an open set is measurable we have

$$\text{(f)} \qquad \underline{J}(M) = \underline{J}(M^*) = |\, M^* \,|.$$

The inner content of an open set equals its measure.

(g) Analogously: *The outer content of a closed set equals its measure.*

The most characteristic property of a J-set is that the measure of its boundary is zero. As before, denote by M^* the sum of all basic sets contained in M, that is the greatest open set in M, by $(B - M)^* = M'^*$, the greatest open set contained in M'. The *boundary* $\mathscr{B}(M)$ is then defined as

$$\mathscr{B}(M) = \mathscr{B}(M') = (M - M^*) + (M' - M'^*). \tag{26}$$

Our theorem reads:

(h) *A set M has content if and only if the measure of its boundary is zero.*

We state first that in this theorem the word "measure" may be replaced by "content." In fact, the boundary $\mathscr{B}$ is, by definition, closed. Then, by (g): $\bar{J}(\mathscr{B}) = |\, \mathscr{B} \,|$. Thus, from $|\, \mathscr{B} \,| = 0$, $\bar{J}(\mathscr{B}) = 0$ follows and hence by (a): $\underline{J}(\mathscr{B}) = 0$ and $J(\mathscr{B}) = 0$. We note that always, *a closed set of measure zero has content zero.*

A proof of theorem (h) runs as follows:

Let first $J(\mathscr{B}) = 0$, then by (26), $J(M - M^*) = 0$. Since the set of sequences B as well as the boundary $\mathscr{B}$ of M have content, $B - \mathscr{B}$ has content [by (d)]. Now

$$B - \mathscr{B} = B - M + M^* - B + M + (B - M)^* = M^* + (B - M)^*.$$

Here M^* and $(B - M)^*$ are disjoint sets, and since each is open, each is by (8) the sum of countably many basic sets. Since $M^* + (B - M)$ has content it follows from (e′), p. 84 that M^* has content and since $M - M^*$ has content the same holds for $(M - M^*) + M^* = M$.

Assume now that M has content. Then $B - M = M'$ has content and from (f):

$$J(M) = | M | = | M^* |, \qquad J(M') = | M' | = | M'^* |.$$

We show now that $| M - M^* | = 0$. In fact if $| M - M^* |$ were positive, we would have

$$| M | = | M^* + (M - M^*) | > | M^* |,$$

in contrast to the above equality. In the same way we see that $| M' - M'^* | = 0$. Consequently $| \mathscr{B} | = 0$. Thus this theorem is proved. A concise expression of our result is contained in the equation

$$\bar{J}(M) - \underline{J}(M) = | \mathscr{B}(M) |. \tag{27}$$

We now return to the previously mentioned statement (p. 83) regarding the dependence of the domain of F_1 upon the chosen contents of the basic sets. We have learned that a set has content if and only if its boundary has measure zero. Assume a regular valuation of all basic sets of B. Then a certain field F_1 is defined, the field of sets of B whose boundary has measure zero. Now consider the same B but a valuation where certain sequences (points) have a measure greater than zero. The boundary of any set M of B will be the same as under the first valuation. If a set M has for this first valuation $| \mathscr{B}(M) | = 0$, this property is now changed if and only if $\mathscr{B}(M)$ contains one of the points with measure greater than zero. If this is so, this set M no longer belongs to F_1 under the new irregular valuation. If, however, a new regular valuation different from the first one is adopted, the sets of F_1 will be exactly the same as for the first one since the boundaries remain the same and measure zero remains measure zero.

6.2. *Examples and further comments on F_1 and on $F_2 - F_1$.* Let us now apply the theorem on the zero measure of the boundary of a J-set to various examples. We consider:

(1) The set ω of sequences considered in Section 5.5 in which "6" appears a finite number of times only. ω contains no basic set since each basic set also contains sequences with infinitely many sixes. Hence ω^* is empty. Neither does $\omega' = B - \omega$ contain a basic set, since each

basic set also contains sequences with a finite number of sixes. Hence $(B - \omega)^*$ is empty. Thus $\mathscr{B}(\omega) = \omega + B - \omega = B$, and the content of B is one and not zero. Very similar reasoning applies to the example of the set R of rationals. Hence the Borel sets R as well as ω have no content and the frequency concept of probability does not apply to either. We have seen indeed that it is impossible to determine by any frequency experiment (directly or indirectly) an approximation to a "probability" of such a set.

(2) Recall the *Cantor set* Z of all sequences which contain no six (see Section 5.1). Again Z contains no basic set, hence Z^* is empty. $B - Z = Y$ is a sum of basic sets. Hence $Y^* = Y$ and $\mathscr{B}(Z) = \mathscr{B}(Y) = Z - Z^* + Y - Y^* = Z$. We have seen before that $| Z | = 0$ hence $| \mathscr{B} | = 0$. Both Y and Z have content.

Remark. Z has the cardinality of the continuum. Not a single point of Z admits finite verifiability; since Z^* is empty, Z consists of boundary points only. However, the content of Z can be verified indirectly, with any desired accuracy, by means of the verifiable content of Y. This behavior is unexceptional: the boundary of any J-set has content $(= 0)$ which can be verified indirectly.

For a basic set, or more generally *for an α-set, the boundary is empty.* All sequences (points) of an α-set admit finite verifiability. For a set of $F_1 - F_0$ all sequences with the exception of those forming a set of measure zero admit finite verifiability. It is instructive to show this independently of the theorem on the measure of the boundary. In fact, let M be a set of F_1. Then inequalities (12) and (13) hold. If M contains a sequence for which finite verification is impossible this sequence must lie in $M - \alpha_n$, since no sequence of α_n defies verification. Since $\beta_n - \alpha_n \supset M - \alpha_n$, this sequence must lie in $\beta_n - \alpha_n$. Also from (12), $B - \alpha_n \supset B - M \supset B - \beta_n$, where $B - \alpha_n$ and $B - \beta_n$ are α-sets. Hence any sequence of $M' = B - M$ for which finite verifiability fails must lie in $(B - M) - (B - \beta_n) = \beta_n - M$, which is $\subset \beta_n - \alpha_n$. Since from (13) $J(\beta_n - \alpha_n)$ is as small as we please we see that *all points of M and of M' for which finite verifiability fails form a set of measure zero.* This is the reason why the content of a set M of F_1 can be approximated with any desired accuracy. In contrast to this we saw that in our examples of sets of $F_2 - F_1$, not a single sequence admitted finite verification; the measure of the set of unverifiable sequences was one. (It is not essential that it equals one but it is always greater than zero.)

Each B has its specific F, F_0, and F_1. We have also seen that any B can be mapped onto the unit interval of the real axis, which we called E. What is the relation between the mappings of these various fields F_1?

6.3. *The Jordan sets on the real axis.* The introduction of Jordan sets is due to Jordan and to Peano. These sets might be defined as sets whose boundary is zero. A definition along the lines of (23) and (23′) is as follows. The *outer* (*inner*) *content* of a set M on (a, b) is the greatest lower bound (the least upper bound) of numbers u (of numbers v) equal to the content of any finite set of open intervals which contain M (which are contained in M). A set *has content* if its outer and inner contents are equal. We call such a set a Jordan set. We obtain the customary Jordan sets if in the above definitions the content of an interval is taken equal to its length.

We define J-sets in (any) B by means of (23) and (23′) and equivalently by (12) and (13). Our above definition of Jordan sets on the real line is equivalent (see p. 73) to a "squeezing-in" definition like (12) and (13) where for the Jordan sets finite sums of intervals play the role which the α-sets play for the J-sets.

We shall now show that *the images of the J-sets of any B are Jordan sets and vice versa.* We shall also answer the question posed at the end of Section 6.2 regarding the relation between mappings of different fields F_1.

That the image of a J-set of any B is a Jordan set and vice versa can be concluded from consideration of the boundaries, if we show that a set M of B whose boundary has measure zero is mapped onto a set $\bar{M}$ of E (the unit interval) whose boundary has measure zero, the characterization of J-sets and Jordan sets in terms of the boundary being necessary and sufficient. It is, however, more in line with our general approach to consider the relation between the definition, (12) and (13), of J-sets and the corresponding one for Jordan sets.

We begin with B^* which, as we know, maps one-to-one onto $E = [0, 1)$. For greater intuitivity we construct the division intervals (d.i.'s) on E in such a way that the length of each d.i. of order ν equals the content of the basic set of order ν whose image it is. Since the content of a basic set of order ν is equal to the sum of the contents of the basic sets of order $\nu + 1$ which add up to it, we see that any d.i. of order ν is covered completely by countably many d.i.'s of order $\nu + 1$ whose lengths add up exactly to the length of the d.i. of order ν. Each point of E lies exactly in one d.i. of order ν, $\nu = 1, 2, \ldots$ (see p. 69).

Let M be a J-set in B^*; then there exist α-sets α, β (we omit subscripts) such that $\alpha \subset M \subset \beta$, and $J(\beta) - J(\alpha) < \epsilon$. Denote by $\bar{\alpha}$, $\bar{\beta}$, $\bar{M}$ the respective images. We show that $\bar{\alpha}$, $\bar{\beta}$ can be replaced by sets α_1, β_1 each a sum of *finitely* many intervals (not necessarily d.i.'s) where $\alpha_1 \subset \bar{\alpha}$, $\beta_1 \supset \bar{\beta}$, and $J(\beta_1) - J(\alpha_1)$ as small as we please. Since $\alpha = E_1 + E_2 + \cdots$ (the E_i are disjoint basic sets) we can find an n such that the content of $E_1 + E_2 + \cdots + E_n$ is as close as we please to $J(\alpha)$. The projection α_1 of this finite sum is a sum of finitely many intervals on E, and, of course $\alpha_1 \subset \bar{\alpha}$. Since β is an α-set the same holds for its complement $\beta' = B^* - \beta = E_1' + E_2' + \cdots$. An m can be found such that the content of $B^* - E_1' - E_2' - \cdots - E_m'$ is arbitrarily close to $J(\beta)$. The projection β_1 of this difference is obtained by substracting from E the d.i.'s which are images of $E_1', E_2', \ldots, E_m'$. The remainder β_1 is a sum of not more than $m + 1$ intervals (not d.i.'s in general). And, of course $\beta_1 \supset \bar{\beta}$. Thus, we have shown that $\alpha_1 \subset \bar{M} \subset \beta_1$, where $J(\beta_1) - J(\alpha_1)$ is arbitrarily small and both α_1 and β_1 are *finite* sums of intervals. Hence the image of any J-set of B^* is a Jordan set of E.

To show the converse we assume for a moment that *any interval i of E is the image of a J-set* (this will be shown presently). It follows that any sum of finitely many disjoint intervals on E is the image of a finite sum of disjoint J-sets, hence of a J-set. Any Jordan set $\bar{M}$ on E can be "squeezed in" by such finite interval sums. Hence the corresponding M in B^* can be squeezed in by means of J-sets of B^*. Hence M is J-set.

It remains to show that any interval on E is the image of a J-set. Now, if ϵ is given an n can be chosen large enough that any d.i. of order n is less than ϵ in length.[2] Hence i can be approximated from within, as close as we please, by a sum δ of d.i.'s, where $\delta \subset i$, and the contents (lengths) of i and δ differ arbitrarily little. The (at most) two intervals i_1, i_2 to the left and right of i which are the complement of i in E can each again be approximated from within by sums of d.i.'s which we call δ_1, δ_2. Then $E - \delta_1 - \delta_2$ contains i and is a sum of d.i.'s (finitely many or infinitely many). Hence, the image I of i is squeezed in between α-sets of B^* and is therefore a J-set.[3]

Next, consider some B with finitely many attributes only. The mapping onto E of the basic sets of B leads to some countable set of division intervals (a different set for each B and different from the set of division intervals that correspond to B^*). It does not matter that the mapping of B onto E is not strictly one-to-one (see p. 67). In fact, let M be a J-set of B and $\alpha_n \subset M \subset \beta_n$, $J(\beta_n) - J(\alpha_n) < \epsilon_n$, in the usual way. The

[2] A sequence of d.i.'s each lying in one of next lower order is the image of a sequence of basic sets of this property (like: (i), (ij), (ijk), $(ijkl)$ $\cdots$) and we know that any such sequence of basic sets—of unbounded order—shrinks to exactly one "point" (= one single sequence) common to all these basic sets. (See p. 69.)

[3] What we have shown is that, *on E*, the customary definition of Jordan sets based on *finite* sums of arbitrary intervals is equivalent to one where we use "α-sets of d.i.'s"; by that we mean a sum of division intervals whose complement is again a sum of d.i.'s, i.e., the projection of our α-sets of B^*. Since, however, division intervals on E exist only in relation to B^* we preferred to give the whole proof in relation to B^*.

normal forms of α_n and β_n each contain a finite number of terms only. Their images on E consist of finitely many d.i.'s plus or minus, possibly, finitely many points. Since a point has content zero the contents of α_n and of the image $\bar{\alpha}_n$ are the same and similarly for β_n. Therefore the image $\bar{M}$ of M can be squeezed in by means of the $\bar{\alpha}_n$, $\bar{\beta}_n$ just as $\bar{M}$ was squeezed in in B, and here any $\bar{\alpha}_n$, $\bar{\beta}_n$ is a sum of finitely many intervals only.

We can now answer the question regarding the relation of the Jordan sets corresponding to different B's. The images of the basic sets of the various B's consist of different countable systems of division intervals. However, if for any of these systems we apply to its F the squeezing-in process, (12) and (13), we obtain one and the same system of sets: *the Jordan sets* (among them, all intervals of E). *They can be generated starting from very diverse countable sets of intervals.*[4]

6.4. Further remarks on Jordan sets. The theorems of pp. 83–85 are valid for the Jordan sets on E. Consider now the following question. We know that a finite sum of intervals has content. What about an infinite sum of intervals? To show that *such a sum may or may not have content* we consider an example, of interest in itself. Divide the closed interval $[0, 1] = E$ into three parts and remove part of the middle third, namely, a centered open interval of length $\frac{1}{3}\lambda$ $(0 \leqslant \lambda \leqslant 1)$. The remaining set consists of two disjoint closed intervals; from each of these remove a centered open interval of length $\frac{1}{9}\lambda$; the remaining set consists of four closed disjoint intervals, and from each of them we remove a centered open interval of length $\frac{1}{27}\lambda$, and so on, *ad inf.* Consider the open set $Y^{(\lambda)}$ consisting of the removed open intervals. We find

$$| Y^{(\lambda)} | = \lambda \left[\frac{1}{3} + \frac{2}{9} + \frac{4}{27} + \cdots\right] = \lambda.$$

The complement of $Y^{(\lambda)}$ is called $Z^{(\lambda)}$. $Z^{(\lambda)}$ is closed and does not contain any interval, like the Z considered on p. 71; therefore its interior content is zero while its exterior content equals $| Z^{(\lambda)} | = 1 - \lambda$. Hence $Z^{(\lambda)}$ has no content except for $\lambda = 1$, which gives our previous Z.[5] The

[4] We have seen that the Jordan sets on E are obtained as images of J-sets of the most varied B's. It should be understood that the same B but with different regular valuation leads also each time to the Jordan sets on E.

[5] The present $Z^{(1)}$ is not exactly the Z of p. 71. There we considered a die, i.e., 6 attributes rather than the present three. If for that die we mix the 6 labels to obtain three labels, each with probability $\frac{1}{3}$, then $Z = Z^{(1)}$. If we denote the three new labels by 0, 1, 2, then $Z^{(1)} = Z$ is the set of all sequences, in the ternary expansion of the numbers between 0 and 1, in which the 1 never appears.

interior content of $Y^{(\lambda)}$ is λ; its exterior content equals one. We thus see that *a countable sum of intervals may or may not have content.*

Closely connected with Jordan content is the concept of an *integrable group*, introduced by Lebesgue. It is defined as a set of points which can be enclosed in a finite number of intervals δ_i such that $\Sigma \mid \delta_i \mid$ is arbitrarily small. This is really nothing else than a Jordan set of exterior content zero, and therefore of content zero. A finite number of points forms an integrable group (if the content of each point is zero). The set $Z^{(1)} = Z$ just mentioned is an integrable group. We note for later use (p. 97) that "a necessary and sufficient condition for a bounded function $f(x)$ to be Riemann integrable is that, any ϵ being given, the points at which the oscillation[6] of $f(x)$ is $> \epsilon$ form an integrable group $G(\epsilon)$." Lebesgue [18, p. 42] has also proved that *the most general Jordan set consists of some infinite sum of intervals plus an integrable group.*[7]

6.5. *Probability distribution over the Jordan sets.*[8] We now return to our original problem—to define a system $\mathscr{S}$ of sets, which are subsets of a non-countable label space S such that the probability over $\mathscr{S}$ is an additive set function.

We take for S a bounded Jordan set (for example an interval) *on the real line. The family $\mathscr{S}$ of sets for which we define probabilities will be the Jordan sets which are subsets of S. Over $\mathscr{S}$ a non-negative content = probability P is defined for which* $P(S) = 1$. *By* (*e*), p. 84, *this distribution is completely additive*, i. e., if a Jordan set M is a sum of disjoint Jordan sets M_i

$$M = \sum_{i=1}^{\infty} M_i \tag{28}$$

then

$$P(M) = P\left(\sum_{i=1}^{\infty} M_i\right) = \sum_{i=1}^{\infty} P(M_i), \tag{28'}$$

where we have written P instead of J.

---

[6] Let x be a non-isolated point of the set E on which $f(x)$ is defined, let δ_1 be an interval that contains x; the points of E contained in δ_1 form a set e_1; on e_1 the function $f(x)$ admits an upper bound L_1 and a lower bound l_1, and an oscillation $L_1 - l_1 = \omega_1$. Let δ_2 be an interval contained in δ_1 and containing x; there are then defined L_2, l_2, ω_2 and $l_1 \leqslant l_2 \leqslant L_2 \leqslant L_1$, $\omega_2 \leqslant \omega_1$. We continue in the same fashion. As the lengths of the δ_i tend to zero, l_i, L_i tend toward limits l, L and $\omega = L - l$ is the *oscillation* of $f(x)$ at x.

[7] Hence a Jordan set of positive content must contain an interval. This corresponds somehow to the following fact: "Any Lebesgue measurable set is the sum of a Borel set and a set of measure zero." [See for example Cramér [4], p. 32.]

[8] See theorems III and IV in Wald's paper, quoted in Appendix One.

As a consequence of (24′) the following holds:

If for a sequence of sets $A_1, A_2, \ldots$ *of* $\mathscr{S}$ *the* $\lim_{n\to\infty} A_n = A$ *exists and belongs to* $\mathscr{S}$ *then* $\lim_{n\to\infty} P(A_n)$ *exists and*

$$P(A) = P(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} P(A_n). \tag{24''}$$

The last equality, with measures instead of contents, is (24′); if contents exist they are equal to the respective measures. [A particular case was Eq. (22).]

We have thus a probability over the field $\mathscr{S}$ consisting of the Jordan sets which are subsets of a Jordan set S. However, we cannot say that $\mathscr{S}$ is a σ-field since the sum of countably many Jordan sets need not be a Jordan set. As in Section 3, one could now apply the Banach extension theorem and extend P over the smallest σ-field that contains $\mathscr{S}$ and obtain in this way a σ-additive set function over a σ-field. On the Jordan sets of this σ-field, the probability P would be determined by relative frequencies. But for the sets which are not Jordan sets the assigned number P would merely be a formal extension of a probability without frequency meaning, a measure or a number assigned by some definition rather than a probability. (See several examples in Chapter V.) Of course, one may and should use measure theory as a computational help. But only inasmuch as our computations eventually lead back to statements regarding sets in F_1 are they verifiable statements in a frequency theory.[9] Actually, the aim of much of the measure-theoretical work on "probability" is not to arrive at statements regarding probabilities but rather to derive measure-theoretical theorems under the heading "probability."

Recalling Sections 5.2, 5.3, and 5.4 where we attempted to analyze the situation, we review: The contents of Jordan sets, sets of F_1, are called probabilities because they admit an approximate conceptual verification. To sets in $F_2 - F_1$ there correspond measures; these measures defy always, and in principle, any frequency interpretation. In particular, one should avoid saying that something "is expected with probability one," with the implication that it will happen almost always, if merely the respective measure equals one. Of course, *the important point is not so much the terminology as the distinction*; the latter is, however, helped by clear terminology. Finally, if a distribution is only an approximation to a sequence of probability distributions as a certain parameter tends toward a limit, then such distributions are neither probabilities nor measures if

[9] This is also the opinion of Kolmogorov (see the last lines of his statement quoted p 56). He does however not draw the consequences.

no limit set exists whose probability (or measure) the distribution would be (see Section 5.4). The distribution is then the limit of a sequence of probabilities (or of measures) and it serves as an analytic approximation.

The next step in the development of the theory will be the definition of the general distribution function or cumulative d.f., which we defined in Section 2 for discrete probabilities. We shall also have to consider the "converse question" discussed in Section 3.3 of this chapter.

7. Distribution Function. Riemann-Stieltjes Integral. Probability Density

7.1. Definition and properties of $F(x)$. Let S be an interval $(-X, +X)$.[1] Let $\mathscr{S}$ consist of all Jordan sets A which are subsets of S and let P be a probability defined over $\mathscr{S}$. If a set A is the interval $(a, b]$ we write $P(A) = P(a < \xi \leqslant b)$ and call $P(A)$ an *interval function,* which is a particular set function. The terms interval function or set function are used in contrast to ordinary *point functions* $P(x, y, ...)$ of one or more variables where the argument is the point with coordinates $x, y, ...$. If an interval function $P(A)$ and some constant k is given we define a point function $F(x; k)$ by putting

$$F(x; k) = \begin{cases} \Pr\{k < \xi \leqslant x\} & \text{for} \quad x > k \\ 0 & \text{for} \quad x = k \\ -\Pr\{x < \xi \leqslant k\} & \text{for} \quad x < k. \end{cases} \tag{29}$$

Whatever the value of k we find then for any finite interval $(a, b]$:

$$F(b;k) - F(a; k) = P(a < \xi \leqslant b) = P(A) \geqslant 0,$$

which shows that $F(x; k)$ *is a nondecreasing function* of x.

Two functions $F(x; k_1)$ and $F(x; k_2)$ differ only by a quantity independent of x; in fact, if $k_1 < k_2$ we obtain from (29)

$$F(x; k_1) - F(x; k_2) = P(k_1 < \xi \leqslant k_2).$$

We may thus choose a fixed but arbitrary value k_0 and denote: $F(x; k_0) = F(x)$. Any other $F(x; k)$ will then be of the form $F(x) +$ constant and we have

$$F(b) - F(a) = P(a < \xi \leqslant b) = \Pr\{a < \xi \leqslant b\}, \tag{30}$$

[1] An infinite interval is not Jordan measurable; however, for any chosen X the interval from $-X$ to $+X$ is Jordan measurable.

i.e., *the probability of a result between a and b*, and $F(x)$ is uniquely determined up to an additive constant. Take now x for b and y for a. We have from (30)

$$F(x) - F(y) = P(y < \xi \leqslant x) = P(A_{y,x}) = \Pr\{y < \xi \leqslant x\}, \tag{30'}$$

where $A_{y,x}$ is the interval $y < \xi \leqslant x$. As $x \to -\infty$, y, which is less than x, tends to the same limit. We set now

$$\lim_{x \to -\infty} F(x) = 0, \tag{31}$$

(which amounts to taking $k_0 = -\infty$). Next, let $y \to -\infty$, but x be arbitrary. From (30′) and (28) we obtain

$$\lim_{y \to -\infty} [F(x) - F(y)] = \lim_{y \to -\infty} P\{y < \xi \leqslant x\} = \Pr\{-\infty < \xi \leqslant x\},$$

and by (31)

$$F(x) = \Pr\{-\infty < \xi \leqslant x\} = \Pr\{\xi \leqslant x\}; \tag{32}$$

thus $F(x)$ *is the probability of obtaining a result which is less than or equal to* x.[2] We call $F(x)$ the *distribution function of the probability distribution* $P(A)$. If, with (31), we let $x \to +\infty$ then the interval $(-\infty < \xi \leqslant x)$ becomes the whole real line, which certainly contains the whole label space S, and therefore

$$\lim_{x \to +\infty} F(x) = 1. \tag{31'}$$

Since $F(x)$ is non-decreasing, the limits to the right and left,

$$F(x+0) = \lim_{\xi \to x+0} F(x), \qquad F(x-0) = \lim_{\xi \to x-0} F(x),$$

exist for all values of x and $F(x - 0) \leqslant F(x + 0)$ and from our definition (30′) we have

$$F(x+0) = F(x), \qquad F(x-0) \leqslant F(x). \tag{33}$$

Consider now a general *monotonically increasing* (or more precisely, "*never decreasing*") function $F(x)$. Such a function $F(x)$ which is everywhere continuous or continuous to the right, and for which (31) and (31′) hold, is called a *distribution function*, (d.f.).

The d.f. $F(x)$ has a *jump* or *saltus* at $x = x_0$ if

$$F(x_0) - F(x_0 - 0) = \gamma > 0.$$

[2] R. v. Mises ["Fundamentalsätze," *Math. Z.* 4 (1919), p. 20] introduced the concept of "Verteilung" with the above meaning.

Let us show that *a d.f. cannot have more than countably many jumps.* Indeed, it cannot have more than one jump of height greater than $\frac{1}{2}$, not more than three for which $\frac{1}{4} < \gamma \leqslant \frac{1}{2} \cdots$, not more than $2^n - 1$ such that $2^{-n} < \gamma \leqslant 2^{-n+1}$. Hence all jumps can be enumerated by arranging them in descending order of magnitude, and hence their total number is at most denumerable.

Thus, the discontinuity points, a_i, of $F(x)$ are finite or countable in number; in the last case they may be everywhere dense, occurring, for example at all rational points or at the points $1/2^k$, $2/2^k$, $3/2^k$, ..., $(2^k - 1)/2^k$, $k = 1, 2, \ldots$.[3] Of course the sum of all jumps must be $\leqslant 1$. Denote by $S(x) = \Sigma_{a_i \leqslant x}\, p(a_i)$, the *saltus function*[4] of $F(x)$; we see that it is a step function. With $S(+\infty) = \alpha$, $\alpha \leqslant 1$, $S(x)/\alpha = F_I(x)$ is a d.f. and $S(x) = \Sigma_{a_i \leqslant x}\, p(a_i) = \alpha F_I(x)$. Then $F(x) - S(x)$ is continuous and monotonically non-decreasing. *A d.f. $F(x)$ can be resolved into the sum of a step function and a continuous function*:

$$F(x) = S(x) + C(x). \tag{34}$$

Consider a simple example with two jumps only. Let $F(0) = 0, F(4) = 1$, $F(1 - 0) = 1/10, F(1) = 2/10, F(3 - 0) = 6/10, F(3) = 9/10$; between 0 and 1, between 1 and 3, and between 3 and 4 the function increases linearly. The corresponding step function $S(x)$ has two steps, one of height $1/10$ at $x = 1$, one of height $3/10$ at $x = 3$. The sum of the steps is $\alpha = \frac{2}{5} = S(+\infty)$. $F(x) - S(x)$ is a continuous polygonal line $C(x)$ with $C(4) = C(+\infty) = \frac{3}{5}$. From $S(x)$, $C(x)$ *the distributions* $F_I(x)$, $F_{II}(x)$ follow by partition; $S(x)/\alpha = F_I(x)$, $C(x)/(1 - \alpha) = F_{II}(x)$; they are conditional probabilities, and

$$F(x) = \alpha F_I(x) + (1 - \alpha) F_{II}(x). \tag{34'}$$

Since at its points of discontinuity, the step function $F_I(x)$ is continuous to the right, the same holds for $F(x)$.

If $F(x) = S(x) + C(x)$ has a countable number of discontinuity points, D, the step function $F_I(x) = S(x)/\alpha$ is the d.f. of a collective, and a corresponding set function may be defined over all subsets of D (Chapter I and Appendix One). If $F_{II}(x) = C(x)(1 - \alpha)$ is also the d.f. of a collective then the same holds for $F(x)$ in (34'). *Indeed, if $F_I(x)$, $F_{II}(x)$ are d.f.'s of two collectives K_1, K_2, then, with any α, $0 \leqslant \alpha \leqslant 1$, $F(x) = \alpha F_I(x) + (1 - \alpha) F_{II}(x)$ is the d.f. of a collective K.* To see this, consider a "coin" with probabilities α and $1 - \alpha$ for heads and tails.

[3] See, however, comment p. 52.

[4] See details in C Carathéodory, *Vorlesungen über reelle Funktionen*, p. 151. Leipzig, Berlin, 1918 (will be quoted as *Carathéodory*).

We start with an observation of K_1, say, note the result, then toss our coin; if heads appear we observe K_1 again, otherwise an observation of K_2 follows, and so on, *ad. inf.* In this way (by means of mixing and combining) a new collective K with the d.f. (34′) is obtained.[5]

We review:

A d. f. $F(x)$ is defined as a monotonically non-decreasing function, with $F(-\infty) = 0$, $F(+\infty) = 1$. It has, at most, denumerably many jumps p_i at points a_i, $i = 1, 2, 3, \ldots$; at a_i it is continuous to the right: $F(a_i + 0) = F(a_i)$, $F(a_i) - F(a_i - 0) = p_i$. It can be resolved into the sum (34′) *of two never decreasing functions, the saltus function $S(x) = \alpha F_I(x)$ and the continuous part $C(x) = (1 - \alpha)F_{II}(x)$, where $\alpha = \sum p_i$; $F_I(x)$ and $F_{II}(x)$ are distribution functions derived by partition from $S(x)$ and $C(x)$.*

7.2. *Converse question.* As in Section 3.3 we now consider the converse question. What kinds of d.f.'s $F(x)$ lead to a set function $P(A)$ over an adequate family $\mathcal{S}$ of sets A? We have mentioned (p. 57) that, whatever d.f. $F(x)$ we consider, a completely additive set function is defined over the Borel sets. This is mathematically a most satisfactory result. However (see Sections 4, 5, and 6, and in particular Section 5.5), the Borel sets are not an adequate field for a frequency theory.

Assume that $F(x)$ is continuous on $[a, b]$—the discontinuous part has been subtracted and we call the difference again $F(x)$—*and that everywhere in $[a, b]$ it has a continuous derivative $F'(x)$.* With a slight generalization of this case we assume that $F(x)$ is *absolutely continuous*; that means that $F(x)$ is an indefinite integral in the sense of Riemann. $F'(x)$ exists everywhere in $[a, b]$ except possibly on a set of measure zero. We then set $f(x) = F'(x)$ wherever $F'(x)$ exists and $f(x) = 0$ otherwise; this function $f(x)$ is integrable in Riemann's sense.

Let E be any set on an interval i which may be $[a, b]$: $E \in i$. The *characteristic* $\theta(x)$ of E, is a function equal to one on E, equal to zero on $i - E$, the complement of E. Now let E be a linear Jordan set on i and consider

$$\int_a^b \theta(x) f(x)\, dx.$$

The discontinuities of $\theta(x)$ are on the boundary of E and if E is a Jordan set, hence they form a set of measure zero; the discontinuities of $f(x)$ also form at most a set of measure zero; hence the Riemann integral exists and we obtain

$$\int_a^b \theta(x) f(x)\, dx = \int_{(E)} f(x)\, dx = P(E), \tag{35}$$

[5] Clearly this generalizes to $F(x) = \alpha_1 F_1(x) + \cdots + \alpha_k F_k(x)$, $0 \leqslant \alpha_i \leqslant 1$, $\sum_1^k \alpha_i = 1$; if the F_i are d.f.'s of collectives, then $F(x)$ is the d.f. of a collective.

where $P(E)$ is the Jordan content with respect to $p(x)$ of E.[6] We note *that* $P(E)$ *tends to zero as the (ordinary) content* (measure) *of* E *tends to zero* (see Lebesgue [18], p. 157). Indeed, the concept of absolute continuity has its origin in this property of the integral.

If, as in the present case, $F(x)$ is absolutely continuous, and in particular, if everywhere $F'(x) = f(x)$ exists and is continuous, the distribution $F(x)$ is briefly called *continuous* and sometimes for historical reasons, a "geometrical" distribution and $f(x)$ is called its *density*. From (31) and (31′),

$$\int_{-\infty}^{+\infty} f(x)\,dx = 1. \tag{31''}$$

The density is nowhere negative. It satisfies (31″) *and*

$$F(x) = \int_{-\infty}^{x} f(t)\,dt. \tag{36}$$

We review:

I. *If* $F(x) = C_a(x)$ *is absolutely continuous, then with the above explanation of* $f(x)$ *and (35), a completely additive set function* $P(E)$ *is defined on the linear Jordan sets. The probability distribution* $P(E)$ *corresponds to the generalized collective in* Φ_1.

If $F(x) = S(x) + C_a(x)$ *where* $S(x)$ *is the saltus function of* $F(x)$ *and* $C_a(x)$ *is absolutely continuous, then* $F(x)$ *is the d.f. of a collective in* Φ_1. *With* $S(\infty) = \alpha$, *the* $S(x)/\alpha = F_I(x)$, $C_a(x)/(1 - \alpha) = F_{II}(x)$, *are distribution functions.*

Hence, we have found so far that $F(x)$ may be the sum of a step function and of an absolutely continuous distribution.

The most general d.f. $F(x)$ is the sum of three non-decreasing functions

$$F(x) = S(x) + C_a(x) + C_s(x). \tag{37}$$

Here, $C_s(x)$ is the *singular component*[7] of the continuous $C(x) = C_a(x) + C_s(x)$. $C_s(x)$ is a continuous monotonically increasing function, which, without being constant, has a derivative equal to zero almost everywhere. An explicit example is given in Lebesgue [18], p. 56. In seeking a relation similar to (35), we are led to the investigation of the *Riemann-Stieltjes integral*.[8] We shall give here only the immediately needed definitions and results. (Others follow in Chapter III Section 2.

[6] Equation (35) will be compared with Eq. (39), p. 98.

[7] See *Carathéodory, loc. cit.*, p. 577, who has investigated $C_s(x)$, see p. 563 ff.

[8] STIELTJES, *Ann. Fac. Sci. Toulouse* (1), **8** (1894), pp. 1–122; see in particular pp. 68–75.

The reader to whom the Stieltjes integral is new may do well to read first this Section which is more elementary than the present considerations.)

Denote by $\phi(x)$ a bounded function in $(a, b]$ and by $F(x)$ a continuous d.f. Divide the interval $(a, b]$ into intervals δ_i by means of the points $a = x_0, x_1, x_2, \ldots, x_n, x_{n+1} = b$ and denote by L_i and l_i the upper and lower bounds of $\phi(x)$ in the interval δ_i: $x_i < x \leqslant x_i + 1, i = 0, 1, \ldots, n$. Put

$$\bar{S} = \sum_{i=0}^{n} L_i F\delta_i \qquad \text{and} \qquad \underline{S} = \sum_{i=0}^{n} l_i \, \delta F_i \, , \tag{38}$$

where $\delta F_i = F(x_{i+1}) - F(x_i)$. Under what conditions do the limits of $\bar{S}$ and $\underline{S}$ exist as $n \to \infty$ and the maximum length of any division interval tends to zero. These limits always exist (Lebesgue [18], p. 272) and are called the upper and lower Darboux-Stieltjes integrals, denoted by $\bar{D}$ and $\underline{D}$, respectively, and $\underline{D} \leqslant \bar{D}$. If $\underline{D} = \bar{D}$ this value is called the Riemann-Stieltjes integral of $\phi(x)$ with respect to $F(x)$. We shall see in Chapter III Section 2.2 that $\bar{D} = \underline{D}$ always if $\phi(x)$ is continuous; but here we need a more general result.

The condition for the equality of $\bar{D}$ and $\underline{D}$ has been investigated by Lebesgue [18], p. 276, and can be formulated in complete analogy to the condition for the existence of an ordinary Riemann integral (see p. 90): *A necessary and sufficient condition for the existence of the Riemann-Stieltjes integral* $\int_a^b \phi(x)\, dF(x)$ *is that the set of points, at which* $\phi(x)$ *has an oscillation* $> \epsilon$ *forms an integrable group with respect to* $F(x)$. An integrable group with respect to $F(x)$ is a set of points which can be enclosed in a finite number of intervals δ_i such that $\Sigma\, \delta F_i$ over these intervals is arbitrarily small. For $F(x) = x$ this reduces to the definition of an ordinary integrable group (p. 90).

We take now for $\phi(x)$ the characteristic $\theta(x)$ of a Jordan set E. The discontinuity points of $\theta(x)$ (the boundary points of E) form an ordinary integrable group, i.e., they can be enclosed in a finite number of intervals δ_i such that the sum of the lengths of these intervals, $\Sigma \mid \delta_i \mid$ is arbitrarily small. Lebesgue's condition, *which is necessary and sufficient* for the existence of $\int_a^b \theta(x)\, dF$, states that $\Sigma\, \delta F_i$ over those intervals should be arbitrarily small[9]; that means that to any given ϵ some γ can be found such that $\Sigma\, \delta F_i < \epsilon$ if $\Sigma \mid \delta_i \mid < \gamma$. This, however, is another form[10] of the definition of absolute continuity.[11]

[9] One can say: the condition is that an integrable group with respect to x should also be an integrable group with respect to $F(x)$ or: the set of discontinuity points of $\phi(x)$ has "weight" zero.

[10] It is a definition not using the notion of the indefinite integral.

[11] See *Carathéodory*, *loc. cit.*, pp. 510 and 513.

S. Saks, *Theory of the Integral*, pp. 30 and 96. Warshaw, New York, 1937.

Hence we see that with $F(x)$ continuous, $\theta(x)$ the characteristic of the set E, the Riemann-Stieltjes integral

$$\int_a^b \theta(x)\, dF(x) = \int_{(E)} dF(x) = P(E) \tag{39}$$

exists if and only if E is a Jordan set and $F(x)$ is absolutely continuous. In this case Eq. (39) reduces to Eq. (35).

If $F(x)$ contains an essentially singular part or, if $F(x) = C_s(x)$ the integral $\int_{(E)} dF(x)$ may exist as a Lebesgue-Stieltjes integral (see e.g. Cramér [4], Chapter 7), but this measure of E has no frequency-probability meaning. We thus see that, in the framework of our theory, *we cannot go beyond the result expressed in the theorem of p.* 96. We cannot affirm the existence of a (generalized) collective with respect to a distribution containing a non-vanishing singular part $C_s(x)$. But we have seen that a collective exists with respect to $F(x) = S(x) + C_a(x)$ where $S(x)$ is the saltus function (step function) of $F(x)$ and $C_a(x)$ is absolutely continuous.

APPENDIX THREE

TORNIER'S FREQUENCY THEORY

Probability theory as presented in this book is a mathematical theory and a science; the probabilities play the role of physical constants; to any probability statement an approximate verification should be at least conceivable (see also Appendix Two). The mathematical concepts and deductions of the theory have to provide appropriate formulations of these principles. Such a formulation, mathematically different from that of Mises, but, intrinsically, closely related to it has been given by Tornier.[1] Like v. Mises' theory, that of Tornier is a genuine frequency theory (see Appendix Two, p. 47). A more recent development—the introduction of operations (see {4}, pp. 28–61)—has brought it even closer to that of v. Mises. The role of sets with content as the labels of probabilities, which also appears in Wald's paper (see Appendix One)

[1] See bibliography at the end of this appendix. The papers will be quoted as {1}, ..., {5}. See also comments of v. Mises [22], pp. 101–103. Naturally, the presentation of an elaborate theory in a brief appendix remains sketchy. More space has been given to those parts of the theory which have no points of contact with Chapter II like subsections (c), (f), and (g) of the appendix.

and in papers by Copeland and de Finetti (see Appendix One for quotations; each of the authors has his specific approach), has been stressed by Tornier since 1932. Tornier's strictly measure theoretical formulations should make his valuable theory familiar to the mathematician and should thus contribute to a better understanding of frequency theories of probability.

(a) Experimental rule. Set of sequences. A main concept of the theory is that of the *experimental rule.* In an experimental rule one may distinguish two components: the *rule of procedure* and the *rule of notation* (or *of noting*), (called by Tornier "Versuchsvorschrift," "Ablaufregel," "Anschreibregel," respectively). We explain by means of some examples.

(1) A die, true or not, is cast repeatedly and each time the number appearing on the upper face is noted. Here, the rule of procedure tells us that a certain die is thrown in a certain way; the rule of notation states what is to be considered the result. Note that one can change the rule of notation without changing the rule of procedure and vice versa. The latter happens, for example, if we cast some other body with six faces. For the former case which plays a greater role we give example (1′):

(1′) The rule of procedure is the same as in (1). But now we note as result the sum of the numbers obtained in the first and second throw, in the second and third throw, ... and so on.

(2) Consider a die and a tetrahedron, whose faces are denoted by 1, 2, 3, 4. First, we throw the die. The $(n + 1)$th throw is with the tetrahedron, if in the nth throw the number on the lower face of either body was even ($n = 1, 2, \ldots$); otherwise the die is used. The result (rule of notation) is each time the number on the lower face (i.e., the face that touches the table) of either die or tetrahedron.

Again we can vary the rule of notation in various ways, for example as in (1′). The rule of procedure might be varied by prescribing that the $(n + 1)$th throw is with the tetrahedron if the result of the nth throw was 3 or 4 or 5 and with the die if it was 1 or 2 or 6.

(3) The number of labels (possible results) need not be finite. As an example remember the cylindrical roller, described Chapter I, p. 3.

(4) The result of each trial may consist of more than one number (as in the simultaneous throwing of several dice).

The rule of notation is idealized in the concept of the *set of sequences* ("Folgenmenge") also *space of sequences,* denoted by B. This *is the set of all infinite sequences which, logically, could be obtained by the given* *experimental rule*. (Tornier {2}). Each particular infinite sequence is considered a *point* in B. In example (1), $B = B_6$ is the set of all infinite sequences which can be formed with the numbers 1, 2, ..., 6. In (2), $B = B_6'$ is the set of all sequences which can be formed with the numbers 1, 2, ..., 6 with the restriction that a 5 or a 6 never follows an even number.[2] In (3), $B = B^*$ is the set of all sequences formed with the positive integers.

(b) The probability field. Preparatory concepts and theorems. The set of sequences B formalizes the rule of notation but it does not characterize the whole experimental rule. In (1), for example, the same B corresponds to a true die as well as to an arbitrarily distorted body with six faces. In order to imprint on B the rule of procedure we assign numbers, valuations or probabilities to certain subsets of B. In a way to be explained in subsection (f) these numbers will be defined as limits of relative frequencies and will in this way reflect the particular experimental rule (true die or biased die, etc). Right now, however, the construction is abstract; the chosen subsets and their valuations will only be bound to satisfy certain axioms which we shall give presently. The system of those subsets of B to which valuations are assigned, together with these valuations, is denoted as the *probability field* Φ. If we want to stress the physical meaning of the valuations we speak of the *probability field* Φ, *induced by the experimental rule* (see footnote 5, p. 101). The number (probability) associated to a set M of Φ is denoted by (Φ, M) or $J(M)$. We set, in particular, with v denoting the empty set

$$(\Phi, B) = 1, \qquad (\Phi, v) = 0. \tag{1}$$

A subset E of B which consists of all sequences of B and only of those which have the same initial segment of length n is called a *basic set of order n*. B itself is given the order zero. Obviously, *B contains countably many basic sets*. Another easily understood property is the following: *two basic sets are either mutually exclusive or one is entirely contained in the other*. To any basic set E_i we associate a content $J(E_i)$ in a way to be explained presently.

We consider E as the *neighborhood* of any sequence contained in E.[3] A finite or countable sum of basic sets is called an *open set*. The complement (with respect to B) of an open set is *closed*.

---

[2] In Section 5.3 of this Chapter we showed that the ordinary B_6 can also be used as representative of the experimental rule (2).

[3] It is easy to see (see Hausdorff [13a] for definitions) that B becomes in this way a topological space, a particularly simple one because any two neighborhoods are either disjoint or one is contained in the other.

If O is an open set in B there is one and only one system of greatest disjoint basic sets of B such that O is their sum. We call

$$O = E_1 + E_2 + \cdots, \tag{2}$$

the *normal form of O*.

For open sets and for closed sets certain well-known theorems quoted in Chapter II, Section 4.2 hold. We do not assign contents to all open (or closed) sets O.

The *α-sets of B* (Chapter II, Section 4.3) play a basic role in Tornier's theory. These sets, A, are both open and closed. An equivalent definition states that A is an α-set if both A and $A' = B - A$ are sums of basic sets; an α-set A is "of bounded order" if the basic sets which add up to A are all of bounded order. The field of the α-sets is denoted by F_0, and that of the α-sets of bounded order by F. The content (= valuation = probability) of an α-set A depends on the chosen contents of the basic sets which form A.

(c) *The axioms.* We state now in the form of axioms the rules to be satisfied by the probability field Φ, i.e., by the chosen subsets of B and their valuations.[4]

I. The sets of Φ form a field.[5]

II. All basic sets of B, including B and the empty set v, belong to Φ.

III. (a) To each set A of Φ we assign a number between 0 and 1 denoted by (Φ, A) or $J(A)$ and called its content. In particular, $J(v) = 0$, $J(B) = 1$.

(b) If $A = A_1 + A_2 + \cdots$, where the A_i are disjoint and A as well as the A_i belong to Φ, then

$$J(A) = \sum_n J(A_n).^{6} \tag{3}$$

IV. Let M be a set of B. Let A_n and B_n, $n = 1, 2, \ldots$ be sequences of sets of Φ such that for any n

$$B_n \supset M \supset A_n \tag{4}$$

[4] Our set of axioms is not a minimum set of requirements.

[5] The Φ used here is not identical with the Φ of Chapter II, but rather with the Φ_1 of Chapter II. There we denoted by Φ the probability field of the original collective and by Φ_1 that of the extended collective. The former plays no role in Tornier's theory.

[6] If for the normal form $E^{(n)} = \sum_i E_i^{(n+1)}$ of a basic set of order n ($n = 0, 1, 2, \ldots$) always $J(E^{(n)}) = \sum_i J(E_i^{(n+1)})$, then axiom III(b) holds as a theorem (see p. 63).

and such that for any ϵ, no matter how small, an n can be found for which

$$(\Phi, B_n) - (\Phi, A_n) < \epsilon. \tag{5}$$

Then M belongs to Φ and

$$(\Phi, M) = \lim_{n\to\infty} (\Phi, A_n) = \lim_{n\to\infty} (\Phi, B_n). \tag{6}$$

V. Φ is the smallest field which satifies the axioms I–IV.

These axioms differ essentially from those of Kolmogorov: Kolmogorov does not specify the "basic material" as Tornier does in axiom II; he has neither the "completeness axiom," IV, nor the "minimal axiom," V.[7] The extension applied by Kolmogorov (see Chapter II, Section 3.2) has no room in Tornier's frequency theory. (Further comments on Kolmogorov's program are at the end of Section 4.1 and end of this appendix.)

The *completeness axiom* IV provided the basic extension in Chapter II [Section 5.2, Eqs. (12)–(14)]. There, the $\{A_n\}$ and $\{B_n\}$ were specified as sets of F, the field of the collective of the discrete label space.

(*d*) *The sets of* Φ. Tornier has shown in {1} that Φ *contains the following types of sets and only these*: (1) *the* α*-sets*; (2) *the sets generated by means of* IV *if the* A_n *and* B_n *are* α*-sets*. The α-sets play a decisive role in Tornier's theory since their field F_0 *forms his basic material*. In v. Mises' theory, *the natural and meaningful basic material is the field of the sets of the discrete label space* (Chapter I) which we identified as α-sets of bounded order. The sets (1) and (2) form the field F_1, the field of sets with content.[8] The contents of all these sets depend on the chosen contents of the basic sets. If we reapply operations (4)-(6) with A_n and B_n replaced by sets of $F_1 - F_0$ we obtain no new sets. While there are countably many basic sets and countably many α-sets, the sets of F_1 have a cardinality greater than that of the continuum.

This theory of content in B is the exact analogue of the theory of the Jordan sets on the real axis, both with respect to the structure of the fields and with respect to the rules of computation for the contents.

We have defined the content of sets M by means of Eqs. (4) to (6), assuming that the sets A_n and B_n have been assigned content. An

[7] He has a "minimal axiom" with respect to his extension; his T_σ is the smallest σ-field containing T.

[8] One obtains exactly the same field F_1 whether the A_n, B_n in IV are sets of F_0 or of F (as in Chapter II), or of f, the smallest field over the basic sets.

equivalent definition is based on *exterior and interior contents of M* [Eqs. (23) and (23′), Chapter II). We have also seen that the sets M of Φ have the characteristic property that the *boundary $\mathscr{B}(M)$ has measure zero.*

For any α-set (set of F_0), $\mathscr{B}$ is empty; for a Jordan set of $F_1 - F_0$, $\mathscr{B}$ is not empty but $|\mathscr{B}| = 0$. All points of a set M of F_1 can be "verified" by a finite set of trials except the points of $\mathscr{B}$; but since the content of $\mathscr{B}$ is zero, this indeterminacy does not influence the content of M. We remember also the comments at the end of Section 6.1 regarding the dependence of the domain of F_1 upon the chosen content of the basic sets.[9]

The field F_1 can be extended in a certain way to F_2, the field of all measurable sets (see Tornier and Domizlaff {4}, p. 69 ff, and quotations given there).[10] In F_2 the simple rules of Lebesgue's theory of measure hold. They can be used to compute contents of sets of F_1 since for any set M that has content, its content is equal to its measure. However, the measure of a set of $F_2 - F_1$ *has no frequency meaning*, as we have seen in Chapter II and shall see right now from another particularly impressive aspect. We propose to consider and *denote as probabilities the verifiable contents of sets of F_1 only* and to use the term measure for those of $F_2 - F_1$.

(e) *Examples* have been given in Sections 5.1 and 5.5 of Chapter II. We pass now to a characteristic and basic piece of Tornier's theory.

(f) *Frequency models.* So far we have defined valuations = probabilities of the sets of Φ in a formal and abstract way, assuming arbitrary (additive) valuations for the basic sets and computing from them further valuations. We shall now assign "models" to such an abstract probability field Φ so that in the model the contents of the sets of the field appear as frequency limits.

The general procedure for determining approximately limits of relative frequencies for segments of length n (basic set of order n) is clear. A great number N of experiments, each consisting of n single trials, are performed. For each segment of length n a relative frequency within the N experiments is determined.[11] Tornier has proved: Suppose that for

[9] Many of the proofs and comments of Chapter II to which we here refer are actually from Tornier's work.

[10] The smallest σ-field over F_1 is not F_2 but contained in F_2.

[11] Here now appears a similarity to the collective, since for each segment of length n (point in n-dimensional label space) the limit of relative frequency is to exist. No "randomness" is explicitly postulated; it reappears however in Tornier's postulates for his "operations."

some B the contents of the basic sets of order n are given; one can then construct a sequence whose elements are segments of length n of B, such that for each segment a frequency limit in this sequence exists, and is equal to the given content of the corresponding basic set of order n. Therefore, one has the right to assume that, even for quite complicated experimental rules (reflected in Φ), contents of the basic sets can be assumed which can be considered as approximate frequencies. Tornier has however proved more.

Since basic sets can be of any order and since valuations are to be assigned to *all* sets of Φ he is led to the definition of more general "models" ({2}, §9; {3}, §5, §8).

If some B and a Φ in B are given, the following holds.

(1) In many ways, one can select countably many sequences of B and take them as columns of an infinite matrix $\bar{\Phi}$ so that in $\bar{\Phi}$ the frequency limits exist for all segments of order n (basic sets E) and are equal to the contents (probabilities) $J(E)$ in Φ.

Such a matrix is called a *frequency model* $\bar{\Phi}$ of Φ; the totality of all $\bar{\Phi}$ which belong to the same Φ is denoted by $\bar{\bar{\Phi}}$.

(2) If M is a set of Φ, i.e., a set with content, then its frequency limit exists in any $\bar{\Phi}$ (i.e., in any model which has the same contents of the basic sets as Φ) and it is equal to the content (= probability) $J(M)$ in Φ, which has been computed in Φ from the given contents of the basic sets.

It follows from (2) that in two different models $\bar{\Phi}$ and $\bar{\Phi}'$ of Φ the frequency limits for any set M of Φ are the same. We see that to any probability statement in Φ corresponds a frequency statement in $\bar{\Phi}$. Also to any computation and proof in Φ corresponds one in $\bar{\Phi}$.

The first n rows of any $\bar{\Phi}$ (for any n) form a kind of storehouse of all the possible results ("possible" means "in agreement with the axioms") which might be observed if one performed a great number of segments of length n.[12]

To obtain a better understanding of (2) and of a third statement that will follow, we remember (see p. 84) that *a set has content if and only if the measure of its boundary $\mathscr{B}$ is zero*. Any set M with content which is not an α-set contains *sequences which defy verification* (otherwise M would be an α-set); these are sequences (points) of $\mathscr{B}(M)$. For no point of $\mathscr{B}(M)$ can one decide after finitely many observations (terms)—and one cannot make infinitely many—whether it belongs to M or to M'. But all these points of $\mathscr{B}$ together have measure zero.

[12] Our $\bar{\Phi}$ has infinite columns, since we want to have n arbitrarily large. (We do not need infinite but only arbitrarily large n.)

Consider now a model $\bar{\Phi}$ and let M be a set of $F_1 - F_0$. One is free to choose in $\bar{\Phi}$ the sequences of $\mathscr{B}(M)$ as sequences of M or of M'. In $\bar{\Phi}$ they may be taken from M and in another model, $\bar{\Phi}'$, from M'. But since $|\mathscr{B}(M)| = 0$, that does not make any difference and we obtain the same frequency limits in $\bar{\Phi}$ and in $\bar{\Phi}'$.[13]

Let now L be a set of B which does not belong to F_1; the measure of $\mathscr{B}(L)$ is positive. If the sequences of $\mathscr{B}(L)$ are chosen from L in $\bar{\Phi}$, and from L' in a model $\bar{\Phi}'$, one can arrange it in such a way that frequency limits for L are different in $\bar{\Phi}$ and $\bar{\Phi}'$ or that a limit does not exist. This shows, together with (2), that *for the sets of F_1 and only for those, the numbers assigned in an abstract theory of content can be interpreted as frequencies*. Hence the statement:

(3) If L is a measurable set of B belonging to $F_2 - F_1$, then there exist among the models $\bar{\Phi}$, which belong to Φ, some for which the frequency limit of L does not exist and others for which it exists but is not equal to the measure of L and not equal to the limit in other models $\bar{\Phi}'$.

We see that a model $\bar{\Phi}$ is a mathematical image of the probability field Φ which corresponds to the experimental rule V. All probability statements which can be derived for this V are reproduced as frequency limits in $\bar{\Phi}$.

We also see that if, in any $\bar{\Phi}$, no frequency limit exists for a set L then no content can be assigned to L in the abstract theory. A measure of L may exist but it cannot be interpreted as a frequency in any $\bar{\Phi}$.

(g) *Continuity*. Tornier systematically reduces the study of continuity to that of B^* (see Chapter II, Section 4.4). To a basic set of order n corresponds a division interval d.i. of order n. We may assign to each d.i. its length as valuation. In the particular example of p. 68, the d.i. given by $(\nu_1\nu_2\ldots\nu_n)$ has the length $(\frac{1}{2})^{\nu_1+\cdots+\nu_n}$. Such a sequence of division intervals, each contained in one of lower order, contracts to one point x of $[0, 1) \equiv E$, as $n \to \infty$, and x lies in all these intervals. Conversely, the common point of a sequence of basic sets each contained in a basic set of next lower order defines a single sequence (point) of B^*. The linear continuum E is mapped one-to-one onto B^*.

We give a few details for the chosen example. For the points x_{ν_1} of the d.i. of first order ν_1 we have

$$\frac{1}{2} + \frac{1}{2^2} + \cdots + \frac{1}{2^{\nu_1-1}} \leqslant x_{\nu_1} < \frac{1}{2} + \frac{1}{2^2} + \cdots + \frac{1}{2^{\nu_1}}$$

[13] If M is in F_0, its boundary is empty and we obtain, of course, the same frequency limits in any model.

or

$$1 - \frac{1}{2^{\nu_1 - 1}} \leqslant x_{\nu_1} < 1 - \frac{1}{2^{\nu_1}} .$$

Likewise, for the points x_{ν_1,ν_2} of the d.i. of order two $(\nu_1\nu_2)$ we have

$$1 - \frac{1}{2^{\nu_1}} - \frac{1}{2^{\nu_1+\nu_2-1}} \leqslant x_{\nu_1,\nu_2} < 1 - \frac{1}{2^{\nu_1}} - \frac{1}{2^{\nu_1+\nu_2}} .$$

The ν_1, ν_2, ... are a sequence of B^*. If we put $\nu_1 + \nu_2 + \cdots + \nu_n = s_n$, we have for the points $x_{\nu_1,\nu_2,\ldots,\nu_n} = (x)$ of the d.i. $(\nu_1\nu_2\ldots\nu_n)$

$$1 - \sum_{i=1}^{n-1} \frac{1}{2^{s_i}} - \frac{1}{2^{s_n - 1}} \leqslant (x) < 1 - \sum_{i=1}^{n} \frac{1}{2^{s_i}} .$$

Hence to the infinite sequence ν_1, ν_2, ... of B^* corresponds the point

$$x = 1 - \sum_{i=1}^{\infty} \frac{1}{2^{s_i}} ;$$

to each point of B^* corresponds a point of [0, 1). Conversely, in order to find the sequence in B^* corresponding to a given x in [0, 1), we write down the binary expansion of $1 - x$:

$$1 - x = 2^{-s_1} + 2^{-s_2} + 2^{-s_3} + \cdots , \qquad s_\nu < s_{\nu+1} .$$

Then s_1, $s_2 - s_1$, $s_3 - s_2$, ... is the sequence of B^* which corresponds to x. It is seen that if two points of E are close to one other, the respective sequences of B^* coincide in many terms.

(*h*) *Operations.* Tornier's operations are considered by him as general analogs of v. Mises' operations (see Tornier and Domizlaff {4}).

The aim is to derive an experimental rule V′ with probability field Φ' from an experimental rule V with probability field Φ. We assume that the rule of procedure remains the same in V′ and that the rule of notation only is changed. The new rule of notation may be "weaker" for V′ than for V, or it may be "stronger." The first case contains Mises' selection and mixing, the second his partition. Consider the first case. The set of sequences B belonging to V is mapped onto the set B' of V′ by $B' = f(B)$, $B = g(B')$, where B' is uniquely determined by B, but in general not vice versa. Let V correspond to the customary casting of a die. Let a new rule of notation be that every second result only is noted, i.e., B' is obtained by deleting in every sequence of B the first, third, fifth, ... figure. This is as in v. Mises' selection. If the notation is changed so as to note only whether a result is odd or even, this amounts to mixing. These mappings, as well as all mappings of this first type, satisfy and have to satisfy the following postulates ({4}, pp. 28–54):

1. $B' = f(B)$ does not generate new probabilities. That means: if L' is a subset of B' belonging to Φ', then $g(L')$ belongs to Φ.

2. Likewise, if for L' in B', $g(L') \in \Phi$, then L' is in Φ'.

3. The mapping does not change the proportion of the probabilities: if L_1' and L_2' have probabilities $J(L_1')$, $J(L_2')$, then $J(L_1') : J(L_2') = J[g(L_1')] : J[g(L_2')]$.

These postulates correspond to our basic conception that probabilities are "material constants." We add the methodological postulate:

4. The mapping function f is independent of the values of the probabilities in Φ.

We obtain a frequency model of V′ by replacing any column of a frequency model of V by its image.

We wish now to characterize those mappings for which the above four postulates are satisfied. The answer is:

The four postulates are satisfied by those and only those mappings of B onto B' which map any basic set of B onto a sum of basic sets of B' and for which the original of any basic set of B' is a sum of basic sets in B.

Among the above characterized admissible mappings are mixing and selection and also transitions like that from (1) to (1′) (p. 99), etc.

We return once more to Tornier's selection (p. 106). For Tornier a selection is *always* an admissible operation in the sense of the above characterization (italics). His experimental rule V may be very general: the succession of trials need not be independent and the number of labels need not be the same in each trial. *If*, however, V consists of independent trials with the same number of labels for each trial, then, and even in a more general case,[14] any selection transforms V into a V′ with *identical valuations*. This corresponds, of course, to v. Mises' invariance under place selection. Thus Tornier's axioms for the operations may be regarded as the expression and generalization of the idea behind v. Mises' randomness concept: (1) probability is a physical constant; (2) the procedure of notation of the successive results cannot change the probabilities (= frequency limits).

In another type of operation which contains v. Mises' partition, the rule V′ is "stronger" than V. In this operation an α-set of B is mapped one-to-one onto the whole space B' ({4}, p. 40 ff). Again the validity of the four postulates is required. These rather meager indications will have to suffice.

---

[14] For every trial there are m possibilities; in a basic set E_n of order n, they may occur with frequencies $r_1, r_2, \ldots, r_m$, $r_1 + r_2 + \cdots + r_m = n$. Then all basic sets of order n with the same r_μ, $\mu = 1, \ldots, m$; $n = 1, 2, \ldots$ must have equal probabilities. These conditions are trivially satisfied in case of independent distributions, where the probability of an E_n is a product of n factors. More generally our condition postulates independence of the order of results: a generalization of independence.

Finally, there is a "combination" ({4}, p. 52 ff), where, in various ways, a new experimental rule V′ is derived from two rules V_1 and V_2 [an example was (2), p. 99]. In general, Φ is not determined by Φ_1 and Φ_2 except in the case of "independence."

(i) Rapid convergence. In Chapter II and in this appendix, we have talked much about verifiability. What exactly is its operational meaning?

Take a simple example. By tossing a coin we wish to determine the relative frequency of those sequences which contain at least two ones in the first three trials. The corresponding set A (an α-set) is the sum of four basic sets: $A = (111) + (110) + (101) + (011)$. In order to verify (or disprove) the assumption $p(A) = \frac{1}{2}$, we make N trials, each trial consisting of 3 throws and surmise that if N is reasonably large we shall find results belonging to A about $(N/2)$ times. In this sense we speak of verifiability no matter whether we wish to verify the simple probability $p = \frac{1}{6}$ for the "5" in the repeated tossing of a die, or $p(A) = \frac{1}{2}$ in the example given here, or the probability of any complicated set of F_1.

Most statisticians who have thought about such questions have probably realized that our belief (confirmed by experience), our hope to obtain a good approximation to a conjectured probability with a moderate N, is based on an assumption drawn from experience. *This assumption has nothing to do with the axioms of probability calculus* (of Tornier or of Kolmogorov or of v. Mises), and it is not explained by any results of theoretical statistics (see our Chapters VIII–XI) which actually rely on such an assumption. Tornier has examined this problem ({5}, §4).

Consider an infinite sequence s consisting of several results, among them the label "1", and assume that, with the usual notations, $\lim_{N\to\infty} N_1/N = p$. We form now a new sequence t identical with s, except that before the first term of s we write m labels among which the "1" appears as often as we please; clearly, the frequency limit of "1" exists again in t and equals p. The same holds if we omit in s the first m labels or if we add m new labels anywhere in s. Our silent assumption is that—in contrast to this example—*in certain known fields of application of probability theory* (games of chance, physics, biology, insurance, etc.) *the frequency limits are approached comparatively rapidly* (the rate of approach being different for different problems).

Tornier's models are very appropriate for precise considerations and conclusions of this type, since, in a model, all axioms of probability theory hold and consequently all statements which follow from the axioms hold. Consider a model $\bar{\Phi}$ of some probability field Φ. We form a new model $\bar{\Phi}'$ by writing m new columns which consist only of "ones" to the left of the first column of $\bar{\Phi}$. We see: the probability field Φ is not influenced by this change. One can add in a model, in an arbitrary way, any number m of columns—at the beginning or in any way dispersed—without changing a single probability in the model.

Let us stress our point also in terms of the collective, taking the simplest case of a collective K with two labels. We may modify K by interspersing randomly long groups of "ones" in such a way that in the resulting K' the frequency limit of "one" within interspersed groups is zero. K' is a collective just as K, with unchanged frequency limits and unviolated randomness.[15] But a statistical examination of such a K' —where rapid convergence does not hold—would lead to abstruse results.

Now anybody would say that even a single group consisting of 1000 ones in succession is "very improbable," its probability being $P = (\frac{1}{2})^{1000}$. But this improbability statement is again a limit statement and not a statement regarding a moderate number N of trials (each "trial" being a segment of length 1000). Without some assumption of "rapid convergence" a statement regarding the probability of such a group of trials has again no justification. On the other hand, the whole body of our experience in applications of probability theory seems to prove that rapid convergence indeed prevails (at least in some domains) as a physical fact, confirmed by an enormous number of observations.[16]

We have said before that the first n lines of a model (let $n = 1000$) contain all that could happen if a great many series of n trials each were performed. Now in the aforementioned modified model $\bar{\Phi}'$ the first m segments of length n consist each of n ones only. Any statistician would consider even *one* such group of 1000 ones highly significant against the assumption of a true coin or die, etc. However, without some assumption of rapid convergence, such a statement of significance has no theoretical justification, either in our theory or in any abstract theory. The advantage of the frequency theories is that they clearly show this state of affairs which remains hidden in the abstract theories.

Now, we all are convinced that a result like the above, showing a succession of many thousands of ones will not appear in the case of a true coin. It should, however, be understood that at the basis of this conviction there is no theorem of probability calculus but rather the above-explained silent assumption about games of chance, biological regularities, etc.

---

[15] It is not difficult to indicate an example of such a K'.

[16] If one would propose to avoid the limit concept and define probabilities as frequencies based on some (large but finite) N the *facts* which we are discussing would not be altered. If N is of the order 10^6, say, then 1000 ones at the beginning of a sequence would hardly change the quotient N_1/N and, for any statistical conclusion we would still have to rely on "rapid convergence."

Long experience in those domains of knowledge where probability theory is successfully used has taught us that probabilities *can* be approximated by means of comparatively short sequences of observations, and that, on the other hand, "significant" deviations from expected results (null hypotheses) *do* mean something, i.e., correspond to some property of the material under investigation. Hence we assume[17] that *in certain known fields of application the frequency limits are approached fairly rapidly.* We also assume that certain "privileged" sequences (to be expected by the law of large numbers, Chapter IV) appear right from the beginning and not only after millions of trials. Tornier {5} has proposed a formulation of those hypotheses to which we refer the reader.

Experience has taught us that in certain fields of research these hypotheses are justified,—we do not know the precise domain of their validity. In such domains, and only in them, statistics can be used as a tool of research. However, in "new" domains where we do not know whether "rapid convergence" prevails, "significant" results in the usual sense may not indicate any "reality." It seems, for example, a wrong approach to try to "prove" parapsychological phenomena by statistical methods.[18] Maybe, in such a domain the convergence is slow! Then situations like the ones discussed here (with the groups of 1000 ones) might lead us to assume a "reality" behind a seemingly significant result while all there is, is slow convergence.

(j) Relation of Tornier's theory to v. Mises' theory and to Kolmogorov's theory. Tornier emphasizes that his interest in probability theory was aroused by v. Mises' work, by Mises' conception of probability theory as a frequency theory and as a science with the probabilities playing the role of "physical constants." Being a pure mathematician, he expressed this scientific approach in an abstract language. He reached the full generality consistent with these ideas. His sample space is from the start the infinite space of sequences, B. His experimental rules comprise all kinds of stochastic dependence. His probabilities are defined in F_1, the field of sets with content. His models relate the abstract structure to frequency interpretations of formal relations. Randomness is implicit in Tornier's operations.

v. Mises' starting point is the collective in the n-dimensional discrete label space S_n, which he introduced into probability theory. The field of the collective, the field T of Chapter I, is the field F of α-sets of bounded order. It comprises dependence in n dimensions. The extension from F to F_1 (Chapter II) introduces then simultaneously the infinite

[17] This is, as always, the conceptual idealization of a physical fact.

[18] This has been said sometime ago by P. Bridgman.

label spaces B and the more general field of sets, namely the field F_1, which contains all kinds of dependence (see Section 5.3). The construction is very intuitive since everything relates eventually to trials in the simple discrete label space.

The probability field is thus the same in Tornier's and v. Mises' theories. Both are frequency theories. The mathematical construction while equally rigorous is quite different, the basic philosophy is the same.

Mises' as well as Tornier's purpose is to reproduce mathematically certain idealized experiences (Mises' collective; Tornier's experimental rule). On the other hand, Kolmogorov axiomatizes the mathematical methods of probability calculus. Therefore, in contrast to Tornier and to Mises, his basic field is of necessity indeterminate. Axioms I, III(a), III(b) of Tornier appear in both theories; but in axiom II, Tornier specifies his basic material while Kolmogorov only postulates a family of subsets T of S such that in T his axioms hold. He then extends his probability field from $P(A)$ on T to $P^*(A)$ on T_σ (smallest σ-field over T). He is free to choose S and T but, of course, it must then be proved that his axioms, in particular III(b), do actually hold for the chosen T.

Kolmogorov does not have axioms like Tornier's IV and V which assure verifiability; hence in general, his fields will contain sets with no conceivable relation to observation, sets which are logico-computational magnitudes.

Actually, there is hardly any intrinsic connection between Tornier and Kolmogorov. Tornier, following Mises, wants to lay the mathematical foundations of the science of probability while for Kolmogorov probability is a piece of mathematics whose analytical structure he explores. Both theories are rigourous mathematical constructions; neither is more general than the other: each has its own objective and it is that which determines its mathematics.

Bibliography

{1} W. Feller and E. Tornier, "Mass- und Inhaltstheorie des Baireschen Nullraums." *Math. Ann.* **107** (1932), pp. 166–187.

{2} E. Tornier, "Grundlagen der Wahrscheinlichkeitsrechnung," *Acta Math.* **60** (1933), pp. 239–380.

{3} E. Tornier, *Wahrscheinlichkeitsrechnung und allgemeine Integrationstheorie.* Leipzig, 1936; reprint, Ann Arbor, Michigan, 1944.

{4} E. Tornier and H. Domizlaff, *Theorie der Versuchsvorschriften der Wahrscheinlichkeitsrechnung.* Stuttgart, 1952.

{5} E. Tornier, "Die Arbeitshypothese 'Antizufallswahrscheinlichkeit,' ihr Ursprung und ihre Grenzen," *Z. Parapsycholog. Grenzgebiete* **3** (1960), pp. 90–119.