ALLE ZEIT WACH
1842

A. Kolmogoroff

Grundbegriffe der Wahrscheinlichkeits-rechnung

Reprint

Springer-Verlag
Berlin Heidelberg New York 1973

AMS Subject Classifications (1970): 60-02, 60-03, 60A05

ISBN 978-3-642-49596-0 ISBN 978-3-642-49888-6 (eBook)
DOI 10.1007/978-3-642-49888-6

Library of Congress Catalog Card Number 32-20335

ERGEBNISSE DER MATHEMATIK
UND IHRER GRENZGEBIETE
HERAUSGEGEBEN VON DER SCHRIFTLEITUNG
DES
„ZENTRALBLATT FÜR MATHEMATIK"
ZWEITER BAND
3

GRUNDBEGRIFFE DER WAHRSCHEINLICHKEITS-RECHNUNG

VON

A. KOLMOGOROFF

BERLIN
VERLAG VON JULIUS SPRINGER
1933

Vorwort.

Zweck des vorliegenden Heftes ist eine axiomatische Begründung der Wahrscheinlichkeitsrechnung. Der leitende Gedanke des Verfassers war dabei, die Grundbegriffe der Wahrscheinlichkeitsrechnung, welche noch unlängst für ganz eigenartig galten, natürlicherweise in die Reihe der allgemeinen Begriffsbildungen der modernen Mathematik einzuordnen. Vor der Entstehung der LEBESGUEschen Maß- und Integrationstheorie war diese Aufgabe ziemlich hoffnungslos. Nach den LEBESGUEschen Untersuchungen lag die Analogie zwischen dem Maße einer Menge und der Wahrscheinlichkeit eines Ereignisses sowie zwischen dem Integral einer Funktion und der mathematischen Erwartung einer zufälligen Größe auf der Hand. Diese Analogie ließ sich auch weiter fortführen: so sind z. B. mehrere Eigenschaften der unabhängigen zufälligen Größen den entsprechenden Eigenschaften der orthogonalen Funktionen völlig analog. Um, ausgehend von dieser Analogie, die Wahrscheinlichkeitsrechnung zu begründen, hätte man noch die Maß- und Integrationstheorie von den geometrischen Elementen, welche bei LEBESGUE noch hervortreten, zu befreien. Diese Befreiung wurde von FRÉCHET vollzogen.

Der diesen allgemeinen Gesichtspunkten entsprechende Aufbau der Wahrscheinlichkeitsrechnung war in den betreffenden mathematischen Kreisen seit einiger Zeit geläufig; es fehlte jedoch eine vollständige und von überflüssigen Komplikationen freie Darstellung des ganzen Systems (es befindet sich allerdings ein Buch von FRÉCHET (Literaturverzeichnis [2]) in Vorbereitung).

Ich möchte hier noch auf diejenigen Punkte der weiteren Darstellung hinweisen, welche außerhalb des erwähnten, den Kennern vertrauten Ideenkreises liegen. Diese Punkte sind die folgenden: Wahrscheinlichkeitsverteilungen in unendlich-dimensionalen Räumen (drittes Kap., § 4), Differentiation und Integration der mathematischen Erwartungen nach einem Parameter (viertes Kap., § 5), vor allem aber die Theorie der bedingten Wahrscheinlichkeiten und Erwartungen (fünftes Kap.). Es sei dabei hervorgehoben, daß diese neuen Fragestellungen notwendigerweise aus einigen ganz konkreten physikalischen Fragestellungen entstanden sind[1].

---

[1] Vgl. z. B. die in der Fußnote 1 auf der S. 41 zitierte Arbeit von Herrn LEONTOWITSCH und dem Verfasser sowie M. LEONTOWITSCH: Zur Statistik der kontinuierlichen Systeme und des zeitlichen Verlaufes der physikalischen Vorgänge. Phys. Z. Sowjetunion Bd. 3 (1933) S. 35–63.

Das sechste Kapitel enthält eine Übersicht (ohne Beweise) einiger Resultate von Herrn KHINTCHINE und dem Verfasser über die Anwendbarkeitsgrenzen des gewöhnlichen und des starken Gesetzes der großen Zahlen. In dem Literaturverzeichnis sind einige neuere Arbeiten angegeben, welche vom Standpunkte der Grundlagenfragen von Interesse sein dürften.

Herrn A. KHINTCHINE, der das ganze Manuskript sorgfältig durchgelesen und dabei mehrere Verbesserungen vorgeschlagen hat, danke ich an dieser Stelle herzlich.

Kljasma bei Moskau, Ostern 1933.

A. KOLMOGOROFF.

Inhaltsverzeichnis.

Erstes Kapitel.

Die elementare Wahrscheinlichkeitsrechnung.

Wir nennen elementare Wahrscheinlichkeitsrechnung denjenigen Teil der Wahrscheinlichkeitsrechnung, in welchem Wahrscheinlichkeiten von nur endlich vielen zufälligen Ereignissen vorkommen. Die Sätze, die hier gewonnen werden, werden natürlich angewandt auch auf Fragen, die mit unendlich vielen zufälligen Ereignissen verbunden sind, allerdings braucht man bei der Behandlung dieser letzteren Fragen auch wesentlich neue Prinzipien. Deshalb wird ein sich gerade auf den Fall unendlich vieler zufälliger Ereignisse beziehendes Axiom der mathematischen Wahrscheinlichkeitstheorie erst zu Beginn des zweiten Kapitels eingeführt (Axiom VI).

Die Wahrscheinlichkeitstheorie als mathematische Disziplin soll und kann genau in demselben Sinne axiomatisiert werden wie die Geometrie oder die Algebra. Das bedeutet, daß, nachdem die Namen der zu untersuchenden Gegenstände und ihrer Grundbeziehungen sowie die Axiome, denen diese Grundbeziehungen zu gehorchen haben, angegeben sind, die ganze weitere Darstellung sich ausschließlich auf diese Axiome gründen soll und keine Rücksicht auf die jeweilige konkrete Bedeutung dieser Gegenstände und Beziehungen nehmen darf.

Dementsprechend wird im § 1 der Begriff eines *Wahrscheinlichkeitsfeldes* als eines gewissen Bedingungen genügenden Mengensystems definiert. Was die Elemente dieser Mengen sind, ist dabei für die rein mathematische Entwicklung der Wahrscheinlichkeitsrechnung völlig gleichgültig (man vgl. die Einführung der geometrischen Grundbegriffe in HILBERTS „Grundlagen der Geometrie" oder die Definitionen von Gruppen, Ringen und Körpern in der abstrakten Algebra).

Jede axiomatische (abstrakte) Theorie läßt bekanntlich unbegrenzt viele konkrete Interpretationen zu. In dieser Weise hat auch die mathematische Wahrscheinlichkeitstheorie neben derjenigen ihrer Interpretationen, aus der sie aufgewachsen ist, auch zahlreiche andere. Wir kommen so zu Anwendungen der mathematischen Wahrscheinlichkeitstheorie auf Untersuchungsgebiete, die mit den Begriffen des Zufalls und der Wahrscheinlichkeit im konkreten Sinne dieser Begriffe nichts zu tun haben. Derartigen Anwendungen ist der Anhang gewidmet.

Die Axiomatisierung der Wahrscheinlichkeitsrechnung kann auf verschiedene Weisen geschehen, und zwar beziehen sich diese verschiedenen Möglichkeiten sowohl auf die Wahl der Axiome als auch auf die der Grundbegriffe und Grundrelationen. Wenn man allerdings das Ziel der möglichen Einfachheit des Axiomensystems und des weiteren Aufbaus der darauf folgenden Theorie im Auge hat, so scheint es am zweckmäßigsten, die Begriffe eines zufälligen Ereignisses und seiner Wahrscheinlichkeit zu axiomatisieren. Es gibt auch andere Begründungssysteme der Wahrscheinlichkeitsrechnung, nämlich solche, bei denen der Wahrscheinlichkeitsbegriff nicht zu den Grundbegriffen zählt, sondern durch andere Begriffe ausgedrückt wird[1]. Dabei wird jedoch ein anderes Ziel angestrebt, nämlich der größtmögliche Anschluß der mathematischen Theorie an die empirische Entstehung des Wahrscheinlichkeitsbegriffes.

§ 1. Axiome[2].

Es sei E eine Menge von Elementen $\xi, \eta, \zeta, \ldots$, welche man *elementare Ereignisse* nennt, und $\mathfrak{F}$ eine Menge von Teilmengen aus E; die Elemente der Menge $\mathfrak{F}$ werden weiter *zufällige Ereignisse* genannt.

I. *$\mathfrak{F}$ ist ein Mengenkörper*[3].

II. *$\mathfrak{F}$ enthält die Menge E.*

III. *Jeder Menge A aus $\mathfrak{F}$ ist eine nichtnegative reelle Zahl* $\mathsf{P}(A)$ *zugeordnet. Diese Zahl* $\mathsf{P}(A)$ *nennt man die Wahrscheinlichkeit des Ereignisses A.*

IV. $\mathsf{P}(E) = 1$.

V. *Wenn A und B disjunkt sind, so gilt*

$$\mathsf{P}(A + B) = \mathsf{P}(A) + \mathsf{P}(B).$$

Ein Mengensystem $\mathfrak{F}$ mit einer bestimmten Zuordnung der Zahlen $\mathsf{P}(A)$, welche den Axiomen I—V genügt, nennt man ein *Wahrscheinlichkeitsfeld.*

Unser Axiomensystem I—V ist *widerspruchsfrei.* Das zeigt folgendes Beispiel: E besteht aus einem einzigen Elemente ξ, $\mathfrak{F}$ aus E und der Nullmenge 0, wobei $\mathsf{P}(E) = 1$, $\mathsf{P}(0) = 0$ gesetzt ist.

[1] Vgl. z. B. R. von Mises [1] und [2] und S. Bernstein [1].

[2] Ein Leser, der den folgenden Axiomen sofort einen konkreten Sinn geben will, soll sogleich den § 2 lesen.

[3] Vgl. Hausdorff: Mengenlehre 1927 S. 78. Ein Mengensystem heißt ein Körper, wenn Summe Durchschnitt und Differenz von zwei Mengen des Systems wieder dem System angehören. Jeder nicht leere Mengenkörper enthält die Nullmenge 0. Wir bezeichnen mit Hausdorff den Durchschnitt von A und B mit AB, die Vereinigungsmenge von A und B im Falle $AB = 0$ mit $A + B$, allgemein aber mit $A \dot{+} B$, und die Differenz von A und B mit $A - B$. Das Komplement $E - A$ der Menge A wird durch A bezeichnet. Die elementaren Rechengesetze für Mengen und ihre Durchschnitte, Summen und Differenzen werden weiter als bekannt vorausgesetzt. Mengen aus $\mathfrak{F}$ werden weiter mit großen lateinischen Buchstaben bezeichnet.

Unser Axiomensystem ist aber *unvollständig:* in verschiedenen Problemen der Wahrscheinlichkeitsrechnung betrachtet man verschiedene Wahrscheinlichkeitsfelder.

Konstruktion von Wahrscheinlichkeitsfeldern. Einfachste Wahrscheinlichkeitsfelder konstruiert man folgendermaßen: Man nimmt eine beliebige endliche Menge $E = \{\xi_1, \xi_2, \ldots, \xi_k\}$ und eine beliebige Menge $\{p_1, p_2, \ldots, p_k\}$ von nichtnegativen Zahlen mit der Summe $p_1 + p_2 + \cdots + p_k = 1$, als $\mathfrak{F}$ nimmt man die Gesamtheit aller Untermengen von E und setzt $\mathsf{P}\{\xi_{i_1}, \xi_{i_2}, \ldots, \xi_{i_\lambda}\} = p_{i_1} + p_{i_2} + \cdots + p_{i_\lambda}$. Man sagt in diesem Falle, daß $p_1, p_2, \ldots, p_k$ die Wahrscheinlichkeiten der elementaren Fälle $\xi_1, \xi_2, \ldots, \xi_k$ oder die elementaren Wahrscheinlichkeiten sind. So erhält man alle möglichen *endlichen* Wahrscheinlichkeitsfelder mit der Gesamtheit aller Untermengen von E als $\mathfrak{F}$ (man nennt dabei ein Wahrscheinlichkeitsfeld endlich, wenn die Menge E endlich ist). Für weitere Beispiele vgl. zweites Kapitel, § 3.

§ 2. Das Verhältnis zur Erfahrungswelt [1].

Die Anwendung der Wahrscheinlichkeitsrechnung auf die reelle Erfahrungswelt geschieht nach dem folgenden Schema:

1. Es wird ein gewisser Komplex $\mathfrak{S}$ von Bedingungen vorausgesetzt, welcher unbeschränkter Wiederholung fähig ist.

2. Man untersucht einen bestimmten Kreis von Ereignissen, welche infolge der Realisation der Bedingungen $\mathfrak{S}$ entstehen können. In den einzelnen Fällen der Realisation der Bedingungen $\mathfrak{S}$ verlaufen die erwähnten Ereignisse im allgemeinen auf verschiedene Weisen. Es sei E die Menge der verschiedenen möglichen Varianten $\xi_1, \xi_2, \ldots$ des Verlaufes der besagten Ereignisse. Einige unter diesen Varianten brauchen dabei überhaupt nicht zur Realisation zu gelangen. Wir nehmen in die Menge E alle Varianten auf, die wir a priori für möglich erachten.

3. Wenn die nach der Realisation der Bedingungen $\mathfrak{S}$ praktisch aufgetretene Variante unserer Ereignisse zu der (durch irgendwelche Bedingungen definierten) Menge A gehört, so sagen wir, daß das Ereignis A stattgefunden hat.

[1] Ein Leser, der sich nur für die rein mathematische Entwicklung der Theorie interessiert, braucht diesen Paragraphen nicht zu lesen — die weitere Darstellung beruht auf den Axiomen des § 1 und benutzt nicht die Überlegungen des gegenwärtigen Paragraphen. In diesem wollen wir uns mit dem bloßen Hinweis auf die empirische Entstehung der Axiome der Wahrscheinlichkeitsrechnung begnügen und lassen deshalb eine eingehende philosophische Untersuchung des Wahrscheinlichkeitsbegriffes in der Erfahrungswelt bewußt beiseite. In der Darstellung der notwendigen Voraussetzungen für die Anwendbarkeit der Wahrscheinlichkeitsrechnung auf die Welt der reellen Geschehnisse folgt der Verfasser im hohem Maße den Ausführungen von Herrn von Mises (vgl. insbesondere [1] S. 21—27: „Das Verhältnis der Theorie zur Erfahrungswelt").

Beispiel. Der Komplex der Bedingungen $\mathfrak{S}$ besteht darin, daß man zweimal eine Münze wirft. Der Kreis von Ereignissen, von dem unter 2. die Rede war, besteht darin, daß bei jedem Wurf der Kopf bzw. der Adler zum Vorschein kommt. Daraus folgt, daß im ganzen vier verschiedene Varianten (*Elementarereignisse*) möglich sind, nämlich

Kopf—Kopf, Kopf—Adler, Adler—Kopf, Adler—Adler.

Als Ereignis A betrachte man eine Wiederholung. Dieses Ereignis besteht aus dem Inbegriff des ersten und vierten Elementarereignisses. Man kann also jedes Ereignis als Menge von Elementarereignissen betrachten.

4. Unter gewissen Bedingungen, auf die wir hier nicht näher eingehen wollen, kann man voraussetzen, daß einem Ereignis A, welches infolge der Bedingungen A auftritt oder nicht, eine gewisse reelle Zahl $\mathsf{P}(A)$ zugeordnet ist, welche folgende Eigenschaften besitzt:

A. Man kann praktisch sicher sein, daß, wenn man den Komplex der Bedingungen $\mathfrak{S}$ eine große Anzahl von n Malen wiederholt und dabei durch m die Anzahl der Fälle bezeichnet, bei denen das Ereignis A stattgefunden hat, das Verhältnis m/n sich von $\mathsf{P}(A)$ nur wenig unterscheidet,

B. Ist $\mathsf{P}(A)$ sehr klein, so kann man praktisch sicher sein, daß bei einer einmaligen Realisation der Bedingungen $\mathfrak{S}$ das Ereignis A nicht stattfindet.

Empirische Deduktion der Axiome. Gewöhnlich kann man voraussetzen, daß das System $\mathfrak{F}$ der in Betracht kommenden Ereignisse $A, B, C, \ldots$, denen gewisse Wahrscheinlichkeiten zugeschrieben sind, einen Mengenkörper bildet, welcher E als Element enthält (Axiome I und II sowie die erste Hälfte des Axiomes III — die Existenz der Wahrscheinlichkeiten). Es ist ferner evident, daß stets $0 \leqq \frac{m}{n} \leqq 1$ ist, so daß die zweite Hälfte des Axioms III durchaus als natürlich erscheint. Für das Ereignis E gilt immer $m = n$, weshalb man natürlicherweise $\mathsf{P}(E) = 1$ setzt (Axiom IV). Sind schließlich A und B miteinander unverträglich (d. h. sind die Mengen A und B disjunkt), so ist $m = m_1 + m_2$, wobei m, m_1, m_2 der Reihe nach die Anzahl der Versuche bezeichnet, in denen das Ereignis $A + B$ bzw. A bzw. B auftritt. Daraus folgt

$$\frac{m}{n} = \frac{m_1}{n} + \frac{m_2}{n}.$$

Es erscheint also als angebracht, $\mathsf{P}(A + B) = \mathsf{P}(A) + \mathsf{P}(B)$ zu setzen.

Bemerkung I. Aus der praktischen Sicherheit zweier Behauptungen folgt die praktische Sicherheit der Behauptung ihrer gleichzeitigen Richtigkeit, obwohl der Sicherheitsgrad sich dabei ein wenig erniedrigt. Ist jedoch die Anzahl der Behauptungen sehr groß, so lassen sich aus der praktischen Sicherheit jeder einzelnen dieser Behauptungen in bezug auf die Richtigkeit der simultanen Behauptung überhaupt keine Schlüsse ziehen. Deshalb folgt aus dem Prinzip A noch keineswegs, daß bei einer sehr großen Anzahl von Serien von Versuchen, von denen jede Serie aus n Versuchen besteht, in *jeder* Serie der Quotient m/n sich von $\mathsf{P}(A)$ wenig unterscheiden wird.

Bemerkung II. Dem unmöglichen Ereignis (der leeren Menge) entspricht kraft unserer Axiome die Wahrscheinlichkeit $\mathsf{P}(0) = 0$ *, während umgekehrt aus $\mathsf{P}(A) = 0$ die Unmöglichkeit des Ereignisses A durchaus nicht zu folgen braucht; nach dem Prinzip B folgt aus dem Nullwerden der Wahrscheinlichkeit nur, daß bei einer einmaligen Realisation der Bedingungen $\mathfrak{S}$ das Ereignis A praktisch unmöglich ist. Das bedeutet jedoch keineswegs, daß auch bei einer genügend langen Reihe von Versuchen das Ereignis A nicht auftreten wird. Andererseits kann man nach dem Prinzip A nur behaupten, daß bei $\mathsf{P}(A) = 0$ und sehr großem n der Quotient m/n sehr klein wird (er kann z. B. gleich $1/n$ sein).

§ 3. Terminologische Vorbemerkungen.

Wir haben die eigentlichen Objekte unserer weiteren Betrachtungen — die zufälligen Ereignisse — als Mengen definiert. Mehrere mengentheoretische Begriffe bezeichnet man aber in der Wahrscheinlichkeitsrechnung mit anderen Namen. Wir wollen hier ein kurzes Verzeichnis solcher Begriffe geben.

Mengentheoretisch.	*Im Falle der zufälligen Ereignisse.*
1. A und B sind disjunkt, d. h. $AB = 0$.	1. Die Ereignisse A und B sind unvereinbar.
2. $AB \ldots N = 0$.	2. Die Ereignisse $A, B, \ldots, N$ sind unvereinbar.
3. $AB \ldots N = X$.	3. Das Ereignis X besteht in der gleichzeitigen Realisation aller Ereignisse $A, B, \ldots, N$.
4. $A \dotplus B \dotplus \cdots \dotplus N = X$.	4. Das Ereignis X besteht in der Erscheinung mindestens eines unter den Ereignissen $A, B, \ldots, N$.
5. Die Komplementärmenge $\bar{A}$.	5. Das entgegengesetzte Ereignis $\bar{A}$ besteht in der Nichterscheinung des Ereignisses A.
6. $A = 0$.	6. A ist unmöglich.
7. $A = E$.	7. A muß notwendig vorkommen.

* Vgl. § 3, Formel (3).

8. Ein System $\mathfrak{A}$ der Mengen $A_1, A_2, \ldots, A_n$ bildet eine *Zerlegung* der Menge E, wenn $A_1 + A_2 + \cdots + A_n = E$ ist (das setzt bereits voraus, daß die Mengen A_i paarweise disjunkt sind).

8. Ein *Versuch* $\mathfrak{A}$ besteht darin, daß man feststellt, welches unter den Ereignissen $A_1, A_2, \ldots, A_n$ vorkommt. $A_1, A_2, \ldots, A_n$ sind die möglichen Ausgänge des Versuches $\mathfrak{A}$.

9. B ist eine Untermenge von A: $B \subset A$.

9. Aus der Realisation des Ereignisses B folgt notwendig dieselbe von A.

§ 4. Unmittelbare Folgerungen aus den Axiomen, bedingte Wahrscheinlichkeiten, der Satz von Bayes.

Aus $A + \bar{A} = E$ und den Axiomen IV und V folgt

$$\mathsf{P}(A) + \mathsf{P}(\bar{A}) = 1, \tag{1}$$

$$\mathsf{P}(\bar{A}) = 1 - \mathsf{P}(A). \tag{2}$$

Da $\bar{E} = 0$ ist, erhält man insbesondere

$$\mathsf{P}(0) = 0. \tag{3}$$

Wenn $A, B, \ldots, N$ unvereinbar sind, so folgt aus dem Axiom IV die Formel

$$\mathsf{P}(A + B + \cdots + N) = \mathsf{P}(A) + \mathsf{P}(B) + \cdots + \mathsf{P}(N) \tag{4}$$

(der Additionssatz).

Wenn $\mathsf{P}(A) > 0$ ist, so nennt man den Quotienten

$$\mathsf{P}_A(B) = \frac{\mathsf{P}(AB)}{\mathsf{P}(A)} \tag{5}$$

die *bedingte Wahrscheinlichkeit* des Ereignisses B unter der Bedingung A. Aus (5) folgt unmittelbar

$$\mathsf{P}(AB) = \mathsf{P}(A)\,\mathsf{P}_A(B). \tag{6}$$

Ein Induktionsschluß ergibt sodann die allgemeine Formel

$$\mathsf{P}(A_1 A_2 \ldots A_n) = \mathsf{P}(A_1)\,\mathsf{P}_{A_1}(A_2)\,\mathsf{P}_{A_1 A_2}(A_3) \ldots \mathsf{P}_{A_1 A_2 \ldots A_{n-1}}(A_n) \tag{7}$$

(der Multiplikationssatz).

Man beweist auch leicht folgende Formeln:

$$\mathsf{P}_A(B) \geqq 0, \tag{8}$$

$$\mathsf{P}_A(E) = 1, \tag{9}$$

$$\mathsf{P}_A(B + C) = \mathsf{P}_A(B) + \mathsf{P}_A(C). \tag{10}$$

Vergleicht man diese Formeln (8) bis (10) mit den Axiomen III bis V, so ergibt sich, daß *das Mengensystem* $\mathfrak{F}$ *mit der Mengenfunktion* $\mathsf{P}_A(B)$

(bei einer festen Menge A) *ein Wahrscheinlichkeitsfeld bildet.* Folglich *gelten alle für die Wahrscheinlichkeiten* $\mathsf{P}(B)$ *bewiesenen allgemeinen Sätze auch für die bedingten Wahrscheinlichkeiten* $\mathsf{P}_A(B)$ (mit einem festen Ereignisse A). Man bemerkt noch leicht, daß

$$\mathsf{P}_A(A) = 1 \tag{11}$$

ist.

Aus (6) und der symmetrischen Formel

$$\mathsf{P}(AB) = \mathsf{P}(B)\,\mathsf{P}_B(A)$$

ergibt sich die wichtige Formel

$$\mathsf{P}_B(A) = \frac{\mathsf{P}(A)\,\mathsf{P}_A(B)}{\mathsf{P}(B)}, \tag{12}$$

welche eigentlich den Satz von BAYES enthält.

Satz über die vollständige Wahrscheinlicheit. Es sei $A_1 + A_2 + \cdots + A_n = E$ (das setzt voraus, daß die Ereignisse A_1, $A_2, \ldots, A_n$ unvereinbar sind) und X beliebig. Dann ist

$$\mathsf{P}(X) = \mathsf{P}(A_1)\,\mathsf{P}_{A_1}(X) + \mathsf{P}(A_2)\,\mathsf{P}_{A_2}(X) + \cdots + \mathsf{P}(A_n)\,\mathsf{P}_{A_n}(X). \tag{13}$$

Beweis.

$$X = A_1X + A_2X + \cdots + A_nX,$$

nach (4) hat man somit

$$\mathsf{P}(X) = \mathsf{P}(A_1X) + \mathsf{P}(A_2X) + \cdots + \mathsf{P}(A_nX),$$

nach (6) gilt aber dabei

$$\mathsf{P}(A_iX) = \mathsf{P}(A_i)\,\mathsf{P}_{A_i}(X).$$

Satz von BAYES. Es sei $A_1 + A_2 + \cdots + A_n = E$ und X beliebig, dann gilt

$$\left\{\begin{aligned} &\mathsf{P}_X(A_i) = \frac{\mathsf{P}(A_i)\,\mathsf{P}_{A_i}(X)}{\mathsf{P}(A_1)\,\mathsf{P}_{A_1}(X) + \mathsf{P}(A_2)\,\mathsf{P}_{A_2}(X) + \cdots + \mathsf{P}(A_n)\,\mathsf{P}_{A_n}(X)},\\ &\qquad\qquad i = 1, 2, 3, \ldots, k. \end{aligned}\right. \tag{14}$$

Man nennt dabei öfters $A_1, A_2, \ldots, A_n$ „Hypothesen“ und sagt, daß die Formel (14) die Wahrscheinlichkeit $\mathsf{P}_X(A_i)$ der Hypothese A_i nach dem Auftreten des Ereignisses X angibt. [$\mathsf{P}(A_i)$ wird dabei als die Wahrscheinlichkeit a priori von A_i bezeichnet.]

Beweis. Nach der Formel (12) hat man

$$\mathsf{P}_X(A_i) = \frac{\mathsf{P}(A_i)\,\mathsf{P}_{A_i}(X)}{\mathsf{P}(X)}.$$

Um die Formel (14) zu gewinnen, bleibt nur die Wahrscheinlichkeit, $\mathsf{P}(X)$ durch ihren Ausdruck (13) nach dem Satze über die vollständige Wahrscheinlichkeit zu ersetzen.

§ 5. Unabhängigkeit.

Der Begriff der gegenseitigen *Unabhängigkeit* zweier oder mehrerer Versuche nimmt eine in gewissem Sinne zentrale Stellung in der Wahrscheinlichkeitsrechnung ein. In der Tat haben wir schon gesehen, daß die Wahrscheinlichkeitsrechnung vom mathematischen Standpunkte aus als eine spezielle Anwendung der allgemeinen Theorie der additiven Mengenfunktionen betrachtet werden kann. Man kann sich natürlich fragen, wie ist es dann möglich, daß die Wahrscheinlichkeitsrechnung sich in eine große, ihre eigenen Methoden besitzende selbständige Wissenschaft entwickelt hat?

Diese Frage zu beantworten, heißt die Spezialisierung anzugeben, die die allgemeinen Fragestellungen über additive Mengenfunktionen in der Wahrscheinlichkeitsrechnung erfahren.

Der Umstand, daß unsere additive Mengenfunktion $\mathsf{P}(A)$ nicht negativ ist und der Bedingung $\mathsf{P}(E) = 1$ genügt, bedingt natürlich noch keine tiefliegenden eigentümlichen Probleme. Zufällige Größen (vgl. drittes Kap.) sind vom mathematischen Standpunkte aus nichts anderes als in bezug auf $\mathsf{P}(A)$ meßbare Funktionen, während ihre mathematischen Erwartungen abstrakte LEBESGUEsche Integrale sind. (Diese Analogie wurde zum erstenmal in den Arbeiten von M. FRÉCHET völlig erklärt[1].) Die Einführung der genannten Begriffe könnte also auch noch keine Basis für die Entwicklung einer großen originellen Theorie liefern.

Geschichtlich ist die Unabhängigkeit von Versuchen und zufälligen Größen derjenige mathematische Begriff, welcher der Wahrscheinlichkeitsrechnung ihr eigenartiges Gepräge gibt. Die klassischen Arbeiten von LAPLACE, POISSON, TCHEBYCHEFF, MARKOFF, LIAPOUNOFF, v. MISES und BERNSTEIN sind in der Tat im wesentlichen der Untersuchung von Reihen unabhängiger Größen gewidmet. Wenn man in den neueren Untersuchungen (MARKOFF, BERNSTEIN usw.) öfters die Forderung der vollständigen Unabhängigkeit ablehnt, so sieht man sich immer gezwungen, um hinreichend inhaltreiche Resultate zu erhalten, abgeschwächte analoge Forderungen einzuführen. (Vgl. in diesem Kap. § 6 — MARKOFFsche Ketten.)

Man kommt also dazu, im Begriffe der Unabhängigkeit wenigstens den ersten Keim der eigenartigen Problematik der Wahrscheinlichkeitsrechnung zu erblicken — ein Umstand, welcher in diesem Buche nur wenig hervortreten wird, da wir hier hauptsächlich nur mit den logischen Vorbereitungen zu den eigentlichen wahrscheinlichkeitstheoretischen Untersuchungen zu tun haben werden.

Es ist dementsprechend eine der wichtigsten Aufgaben der Philosophie der Naturwissenschaften, nachdem sie die vielumstrittene Frage

---

[1] Vgl. [1] und [2].

über das Wesen des Wahrscheinlichkeitsbegriffes selbst erklärt hat, die Voraussetzungen zu präzisieren, bei denen man irgendwelche gegebene reelle Erscheinungen für gegenseitig unabhängig halten kann. Diese Frage fällt allerdings aus dem Rahmen unseres Buches.

Es seien n Versuche $\mathfrak{A}^{(1)}, \mathfrak{A}^{(2)}, \ldots, \mathfrak{A}^{(n)}$, d. h. n Zerlegungen

$$E = A_1^{(i)} + A_2^{(i)} + \cdots + A_{r_i}^{(i)} \qquad i = 1, 2, \ldots, n$$

der Grundmenge E gegeben. Im allgemeinen kann man $r = r_1 r_2 \ldots r_n$ Wahrscheinlichkeiten

$$p_{q_1 q_2 \ldots q_n} = \mathsf{P}\left(A_{q_1}^{(1)} A_{q_2}^{(2)} \ldots A_{q_n}^{(n)}\right) \geqq 0$$

beliebig unter der einzigen Bedingung

$$(1) \qquad \sum_{q_1, q_2, \ldots, q_n} p_{q_1 q_2 \ldots q_n} = 1$$

wählen[1].

Definition I. Man nennt n Versuche $\mathfrak{A}^{(1)}, \mathfrak{A}^{(2)}, \ldots, \mathfrak{A}^{(n)}$ gegenseitig unabhängig, wenn bei beliebigen $q_1, q_2, \ldots, q_n$ die Gleichung

$$(2) \qquad \mathsf{P}\left(A_{q_1}^{(1)} A_{q_2}^{(2)} \ldots A_{q_n}^{(n)}\right) = \mathsf{P}\left(A_{q_1}^{(1)}\right) \mathsf{P}\left(A_{q_2}^{(2)}\right) \ldots \mathsf{P}\left(A_{q_n}^{(n)}\right)$$

gilt.

Man hat unter den r Gleichungen (2) nur $r - r_1 - r_2 - \cdots - r_n + n - 1$ unabhängige[2].

Satz I. Wenn n Versuche $\mathfrak{A}^{(1)}, \mathfrak{A}^{(2)}, \ldots, \mathfrak{A}^{(n)}$ gegenseitig unabhängig sind, so sind auch irgendwelche m $(m < n)$ Versuche $\mathfrak{A}^{(i_1)}, \mathfrak{A}^{(i_2)}, \ldots, \mathfrak{A}^{(i_m)}$ unter ihnen unabhängig[3].

---

[1] Man konstruiert ein Wahrscheinlichkeitsfeld mit beliebigen, nur den erwähnten Bedingungen unterworfenen Wahrscheinlichkeiten $p_{q_1 q_2 \ldots q_n}$ folgendermaßen: die Menge E bestehe aus r Elementen $\xi_{q_1 q_2 \ldots q_n}$, die entsprechenden elementaren Wahrscheinlichkeiten seien $p_{q_1 q_2 \ldots q_n}$, endlich sei $A_q^{(i)}$ die Menge aller $\xi_{q_1 q_2 \ldots q_n}$ mit $q_i = q$.

[2] In der Tat kann man im Unabhängigkeitsfalle nur $r_1 + r_2 + \cdots + r_n$ Wahrscheinlichkeiten $p_q^{(i)} = \mathsf{P}(A_q^{(i)})$ beliebig wählen, und zwar unter der Geltung von n Bedingungen

$$\sum_q p_q^{(i)} = 1 .$$

Man hat also im allgemeinen Falle $r - 1$ Freiheitsgrade und im Falle der Unabhängigkeit nur $r_1 + r_2 + \cdots + r_n - n$.

[3] Zum Beweise genügt es, zu zeigen, daß aus der gegenseitigen Unabhängigkeit von n Zerlegungen die gegenseitige Unabhängigkeit von $n - 1$ ersten unter ihnen folgt. Es sei also angenommen, daß die Gleichungen (2) erfüllt sind. Dann ist

$$\mathsf{P}\left(A_{q_1}^{(1)} A_{q_2}^{(2)} \ldots A_{q_{n-1}}^{(n-1)}\right) = \sum_{q_n} \mathsf{P}\left(A_{q_1}^{(1)} A_{q_2}^{(1)} \ldots A_{q_n}^{(n)}\right)$$

$$= \mathsf{P}(A_{q_1}^{(1)}) \mathsf{P}(A_{q_2}^{(2)}) \ldots \mathsf{P}(A_{q_{n-1}}^{(n-1)}) \sum_{q_n} \mathsf{P}(A_{q_n}^{(n)}) = \mathsf{P}(A_{q_1}^{(1)}) \mathsf{P}(A_{q_2}^{(2)}) \ldots \mathsf{P}(A_{q_{n-1}}^{(n-1)}),$$

w. z. b. w.

Es gelten folglich im Falle der Unabhängigkeit die Gleichungen

$$\mathsf{P}\left(A_{q_1}^{(i_1)} A_{q_2}^{(i_2)} \ldots A_{q_m}^{(i_m)}\right) = \mathsf{P}\left(A_{q_1}^{(i_1)}\right) \mathsf{P}\left(A_{q_2}^{(i_2)}\right) \ldots \mathsf{P}\left(A_{q_m}^{(i_m)}\right) \tag{3}$$

(i_k sollen dabei alle verschieden sein).

Definition II. n Ereignisse $A_1, A_2, \ldots, A_n$ sind gegenseitig unabhängig, wenn die Zerlegungen (Versuche)

$$E = A_k + \bar{A}_k \qquad k = 1, 2, \ldots, n$$

unabhängig sind.

In diesem Falle ist $r_1 = r_2 = \cdots = r_n = 2, r = 2^n$; es gibt folglich unter den 2^n Gleichungen (2) nur $2^n - n - 1$ unabhängige. Die folgenden $2^n - n - 1$ Gleichungen sind auch für die Unabhängigkeit der Ereignisse $A_1, A_2, \ldots, A_n$ notwendig und hinreichend[1]:

$$\mathsf{P}(A_{i_1} A_{i_2} \ldots A_{i_m}) = \mathsf{P}(A_{i_1}) \mathsf{P}(A_{i_2}) \ldots \mathsf{P}(A_{i_m}), \tag{4}$$
$$m = 1, 2, \ldots, n,$$
$$1 \leqq i_1 < i_2 < \cdots < i_m \leqq n.$$

Diese Gleichungen sind alle gegenseitig unabhängig. Im Falle $n = 2$ erhält man nach (4) nur eine ($2^2 - 2 - 1 = 1$) Bedingung für die Unabhängigkeit von zwei Ereignissen A_1 und A_2:

$$\mathsf{P}(A_1 A_2) = \mathsf{P}(A_1) \mathsf{P}(A_2). \tag{5}$$

Das System der Gleichungen (2) besteht in diesem Falle außer (5) noch aus drei Gleichungen:

$$\mathsf{P}(A_1 \bar{A}_2) = \mathsf{P}(A_1) \mathsf{P}(\bar{A}_2),$$
$$\mathsf{P}(\bar{A}_1 A_2) = \mathsf{P}(\bar{A}_1) \mathsf{P}(A_2),$$
$$\mathsf{P}(\bar{A}_1 \bar{A}_2) = \mathsf{P}(\bar{A}_1) \mathsf{P}(\bar{A}_2),$$

welche in ersichtlicher Weise aus (5) folgen[2].

Man braucht dabei kaum zu bemerken, daß aus der *paarweisen* Unabhängigkeit der Ereignisse $A_1, A_2, \ldots, A_n$, d. h. aus der Relation

$$\mathsf{P}(A_i A_j) = \mathsf{P}(A_i) \mathsf{P}(A_j) \qquad i \neq j$$

im Falle $k > 2$ noch keineswegs die Unabhängigkeit dieser Ereignisse folgt[3] [dafür ist das Bestehen aller Gleichungen (4) notwendig].

[1] Vgl. S. Bernstein [1], S. 47—57. Der Leser kann das übrigens selbst mühelos nachprüfen (Induktionsschluß).

[2] $\mathsf{P}(A_1 \bar{A}_2) = \mathsf{P}(A_1) - \mathsf{P}(A_1 A_2) = \mathsf{P}(A_1) - \mathsf{P}(A_1) \mathsf{P}(A_2) = \mathsf{P}(A_1) \{1 - \mathsf{P}(A_2)\}$ $= \mathsf{P}(A_1) \mathsf{P}(\bar{A}_2)$ usw.

[3] Das zeigt auch das folgende einfache Beispiel (S. Bernstein): Die Menge E besteht aus vier Elementen $\xi_1, \xi_2, \xi_3, \xi_4$, die entsprechenden elementaren Wahrscheinlichkeiten p_1, p_2, p_3, p_4 werden alle gleich $\frac{1}{4}$ gesetzt und $A = \{\xi_1, \xi_2\}$, $B = \{\xi_1, \xi_3\}$, $C = \{\xi_1, \xi_4\}$ ist. Man berechnet leicht, daß dann $\mathsf{P}(A) = \mathsf{P}(B) = \mathsf{P}(C) = \frac{1}{2}$, $\mathsf{P}(AB) = \mathsf{P}(BC) = \mathsf{P}(A) = \frac{1}{4} = (\frac{1}{2})^2$, $\mathsf{P}(ABC) = \frac{1}{4} \neq (\frac{1}{2})^3$ ist.

Wir haben bei der Einführung des Unabhängigkeitsbegriffes die bedingten Wahrscheinlichkeiten nicht gebraucht. Unser Ziel war dabei, das Wesen dieses Begriffes rein mathematisch möglichst klar darzustellen. Seine Anwendungen beruhen aber meistens auf Eigenschaften gewisser bedingter Wahrscheinlichkeiten. Nehmen wir an, daß alle Wahrscheinlichkeiten $\mathsf{P}(A_q^{(i)})$ positiv sind, so folgt aus den Gleichungen (3), daß

$$\mathsf{P}_{A_{q_1}^{(i_1)} A_{q_2}^{(i_2)} \ldots A_{q_{m-1}}^{(i_{m-1})}}\left(A_{q_m}^{(i_m)}\right) = \mathsf{P}\left(A_{q_m}^{(i_m)}\right) \tag{6}$$

ist[1]. Aus dem Bestehen der Formeln (6) folgen umgekehrt nach dem Multiplikationssatz [Formel (7) aus § 4] die Formeln (2). Wir haben also den

Satz II. Im Falle positiver Wahrscheinlichkeiten $\mathsf{P}(A_q^{(i)})$ ist folgende Bedingung für die Unabhängigkeit der Versuche $\mathfrak{A}^{(1)}, \mathfrak{A}^{(2)}, \ldots, \mathfrak{A}^{(n)}$ notwendig und hinreichend: Die bedingte Wahrscheinlichkeit eines Ausganges $A_q^{(i)}$ des Versuches $\mathfrak{A}^{(i)}$ unter der Hypothese, daß einige andere Versuche $\mathfrak{A}^{(i_1)}, \mathfrak{A}^{(i_2)}, \ldots, \mathfrak{A}^{(i_k)}$ bestimmte Ausgänge $A_{q_1}^{(i_1)}, A_{q_2}^{(i_2)}, A_{q_3}^{(i_3)}, \ldots, A_{q_k}^{(i_k)}$ hatten, ist der absoluten Wahrscheinlichkeit $\mathsf{P}(A_q^{(i)})$ gleich.

Auf Grund der Formeln (4) beweist man ganz analog auch den folgenden Satz:

Satz III. Wenn alle Wahrscheinlichkeiten $\mathsf{P}(A_k)$ positiv sind, so ist das Bestehen der Gleichungen

$$\mathsf{P}_{A_{i_1} A_{i_2} \cdots A_{i_k}}(A_i) = \mathsf{P}(A_i) \tag{7}$$

bei beliebigen paarweise verschiedenen $i_1, i_2, \ldots, i_k, i$ für die gegenseitige Unabhängigkeit der Ereignisse $A_1, A_2, \ldots, A_n$ notwendig und hinreichend.

Im Falle $n = 2$ reduzieren sich die Bedingungen (7) auf zwei Gleichungen

$$\begin{cases} \mathsf{P}_{A_1}(A_2) = \mathsf{P}(A_2), \\ \mathsf{P}_{A_2}(A_1) = \mathsf{P}(A_1). \end{cases} \tag{8}$$

Man ersieht sogleich, daß schon die Gleichung (8) allein eine notwendige und hinreichende Bedingung für die Unabhängigkeit von A_1 und A_2 darstellt, wenn nur $\mathsf{P}(A_1) > 0$ ist.

§ 6. Bedingte Wahrscheinlichkeiten als zufällige Größen, Markoffsche Ketten.

Es sei $\mathfrak{A}$ eine Zerlegung der Grundmenge E:

$$E = A_1 + A_2 + \cdots + A_r$$

[1] Zum Beweise erinnere man sich an die Definition der bedingten Wahrscheinlichkeit [Formel (5) aus § 4] und ersetze die Wahrscheinlichkeiten der Durchschnitte durch Produkte der Wahrscheinlichkeiten nach der Formel (3).

und x eine reelle Funktion des Elementarereignisses ξ, welche auf jeder Menge A_q einer entsprechenden Konstante a_q gleich ist. Man sagt in diesem Falle, daß x eine *zufällige Größe* ist, und nennt die Summe

$$E(x) = \sum_q a_q \mathsf{P}(A_q)$$

die *mathematische Erwartung* der Größe x. Die Theorie der zufälligen Größen und ihrer Erwartungen wird in dem dritten und vierten Kapitel entwickelt, und zwar ohne Beschränkung auf zufällige Größen, welche nur endlich viele verschiedene Werte annehmen können.

Die zufällige Größe, welche auf jeder Menge A den Wert $\mathsf{P}_{A_{q_i}}(B)$ hat, nennen wir *die bedingte Wahrscheinlichkeit von B nach dem Versuch* $\mathfrak{A}$ und bezeichnen sie mit $\mathsf{P}_{\mathfrak{A}}(B)$. Zwei Versuche $\mathfrak{A}^{(1)}$ und $\mathfrak{A}^{(2)}$ sind also dann und nur dann unabhängig, wenn

$$\mathsf{P}_{A^{(1)}}(A_q^{(2)}) = \mathsf{P}(A_q^{(2)}) \qquad q = 1, 2, \ldots, r_2$$

ist.

Sind irgendwelche Zerlegungen (Versuche) $\mathfrak{A}^{(1)}, \mathfrak{A}^{(2)}, \ldots, \mathfrak{A}^{(n)}$ gegeben, so bezeichnen wir mit

$$\mathfrak{A}^{(1)}\mathfrak{A}^{(2)} \ldots \mathfrak{A}^{(n)}$$

die Zerlegung der Menge E in die Produkte $\mathfrak{A}_{q_1}^{(1)}\mathfrak{A}_{q_2}^{(2)} \ldots \mathfrak{A}_{q_n}^{(n)}$. Die Versuche $\mathfrak{A}^{(1)}, \mathfrak{A}^{(2)}, \ldots, \mathfrak{A}^{(n)}$ sind dann und nur dann gegenseitig unabhängig, wenn

$$\mathsf{P}_{\mathfrak{A}^{(1)}\mathfrak{A}^{(2)} \ldots \mathfrak{A}^{(k-1)}}(A_q^{(k)}) = \mathsf{P}(A_q^{(k)})$$

bei jeder Wahl von k und q ist[1].

Definition. Eine Folge

$$\mathfrak{A}^{(1)}, \mathfrak{A}^{(2)}, \ldots, \mathfrak{A}^{(n)}, \ldots$$

von Versuchen bildet eine MARKOFFsche Kette, wenn bei jedem n und q

$$\mathsf{P}_{\mathfrak{A}^{(1)}\mathfrak{A}^{(2)} \ldots \mathfrak{A}^{(n-1)}}(A_q^{(n)}) = \mathsf{P}_{\mathfrak{A}^{(n-1)}}(A_q^{(n)})$$

ist.

Die MARKOFFschen Ketten bilden also eine natürliche Verallgemeinerung der Folgen von gegenseitig unabhängigen Versuchen. Bezeichnet man

$$p_{q_m q_n}(m, n) = \mathsf{P}_{\mathfrak{A}_{q_m}^{(m)}}(A_{q_n}^{(n)}), \qquad m < n$$

so lautet die Grundformel der Theorie von MARKOFFschen Ketten

$$p_{q_k q_n}(k, n) = \sum_{q_m} p_{q_k q_m}(k, m)\, p_{q_m q_n}(m, n). \qquad k < m < n \tag{1}$$

---

[1] Die Notwendigkeit dieser Bedingungen folgt aus dem Satz II, § 5, daß sie auch hinreichend sind, schließt man unmittelbar aus dem Multiplikationssatz [Formel (7) des § 4].

Bezeichnet man die Matrix $\|p_{q_m q_n}(m, n)\|$ mit $p(m, n)$, so läßt sich (1) in der folgenden Form darstellen:

$$p(k, n) = p(k, m)\, p(m, n). \qquad k < m < n \text{ *} \tag{2}$$

Zweites Kapitel.

Unendliche Wahrscheinlichkeitsfelder.

§ 1. Das Stetigkeitsaxiom.

Wir bezeichnen wie üblich mit $\underset{m}{\mathfrak{D}} A_m$ den Durchschnitt der Mengen A_m (in einer endlichen oder unendlichen Anzahl) und mit $\underset{m}{\mathfrak{S}} A_m$ ihre Vereinigungsmenge. Nur im Falle disjunkter Mengen A_m schreibt man $\sum\limits_m A_m$ statt $\underset{m}{\mathfrak{S}} A_m$. Es ist also

$$\underset{m}{\mathfrak{S}} A_m = A_1 \dot{+} A_2 \dot{+} \cdots,$$

$$\sum_m A_m = A_1 + A_2 + \cdots,$$

$$\underset{m}{\mathfrak{D}} A_m = A_1 A_2 \cdots.$$

Bei allen weiteren Betrachtungen setzen wir voraus, daß außer I—V noch das folgende Axiom erfüllt ist:

VI. *Für eine abnehmende Folge*

$$A_1 \supset A_2 \supset \cdots \supset A_n \supset \cdots \tag{1}$$

von Ereignissen aus $\mathfrak{F}$ *mit*

$$\underset{n}{\mathfrak{D}} A_n = 0 \tag{2}$$

gilt die Gleichung

$$\lim \mathrm{P}(A_n) = 0. \qquad n \to \infty \tag{3}$$

In der ganzen weiteren Darstellung nennen wir *Wahrscheinlichkeitsfeld* nur ein Wahrscheinlichkeitsfeld im Sinne des ersten Kapitels, welches außerdem dem Axiom VI genügt. Wahrscheinlichkeitsfelder im Sinne des ersten Kapitels könnte man *Wahrscheinlichkeitsfelder im erweiterten Sinne* nennen.

Wenn das Mengensystem $\mathfrak{F}$ endlich ist, folgt VI aus den Axiomen I–V. In der Tat gibt es in diesem Falle nur endlich viele verschiedene Mengen in der Folge (1); es sei A_k die kleinste unter ihnen, dann fallen alle

* Über die weitere Entwicklung der Theorie der MARKOFFschen Ketten vgl. R. v. MISES [1] § 16 und B. HOSTINSKY: Méthodes générales du calcul des probabilités. Mém. Sci. math. Bd. 52. Paris 1931.

Mengen A_{k+p} mit A_k zusammen, und es gilt folglich

$$A_k = A_{k+p} = \mathfrak{D}_n A_n = 0\,,$$

$$\lim \mathsf{P}(A_n) = \mathsf{P}(0) = 0\,.$$

Alle Beispiele von *endlichen* Wahrscheinlichkeitsfeldern aus dem ersten Kapitel genügen folglich auch dem Axiom VI. Das Axiomensystem I—VI ist somit *widerspruchsfrei* und *unvollständig*.

Für die unendlichen Wahrscheinlichkeitsfelder ist dagegen das Stetigkeitsaxiom VI von den Axiomen I—V unabhängig. Da das neue Axiom nur für die unendlichen Wahrscheinlichkeitsfelder wesentlich ist, währe es kaum möglich, seine empirische Bedeutung zu erklären, etwa so, wie es für die Axiome I—V im § 2 des ersten Kapitels skizziert wurde. Bei einer Beschreibung irgendwelcher wirklich beobachtbarer zufälliger Prozesse kann man nur endliche Wahrscheinlichkeitsfelder erhalten. Unendliche Wahrscheinlichkeitsfelder erscheinen nur als idealisierte Schemata reeller zufälliger Prozesse. *Wir beschränken uns dabei willkürlich auf solche Schemata, welche dem Stetigkeitsaxiom VI genügen.* Diese Beschränkung erwies sich bis jetzt bei den verschiedensten Untersuchungen als zweckmäßig.

Der erweiterte Additionssatz. Wenn $A_1, A_2, \ldots, A_n, \ldots$ und A zu $\mathfrak{F}$ gehören, so folgt aus

$$A = \sum_n A_n \tag{4}$$

die Gleichung

$$\mathsf{P}(A) = \sum_n \mathsf{P}(A_n)\,. \tag{5}$$

Beweis. Es sei

$$R_n = \sum_{m>n} A_m$$

gesetzt. Dann gilt offenbar

$$\mathfrak{D}_n (R_n) = 0\,,$$

und folglich nach dem Axiom VI

$$\lim \mathsf{P}(R_n) = 0\,. \qquad n \to \infty \tag{6}$$

Nach dem Additionssatz ist andererseits

$$\mathsf{P}(A) = \mathsf{P}(A_1) + \mathsf{P}(A_2) + \cdots + \mathsf{P}(A_n) + \mathsf{P}(R_n)\,. \tag{7}$$

Aus (6) und (7) folgt unmittelbar (5).

Wir haben also bewiesen, daß *die Wahrscheinlichkeit $P(A)$ eine auf $\mathfrak{F}$ vollständig additive Mengenfunktion ist.* Umgekehrt gelten für jede auf einem Mengenkörper $\mathfrak{F}$ definierte vollständig additive Mengenfunktion $\mathsf{P}(A)$ die Axiome V und VI[1]. Man kann folglich den Begriff des Wahr-

[1] Vgl. z. B. O. Nikodym: Sur une généralisation des intégrales de M. J. Radon. Fundam. Math. Bd. 15 (1930) S. 136.

scheinlichkeitsfeldes folgendermaßen definieren: *Es seien E eine beliebige Menge, $\mathfrak{F}$ ein Körper der Untermengen von E, welcher E enthält, und $P(A)$ eine nichtnegative auf $\mathfrak{F}$ definierte vollständig additive Mengenfunktion; der Mengenkörper $\mathfrak{F}$ zusammen mit der Mengenfunktion $\mathsf{P}(A)$ bildet dann ein Wahrscheinlichkeitsfeld.*

Ein Überdeckungssatz. Wenn $A, A_1, A_2, \ldots, A_n, \ldots$ zu $\mathfrak{F}$ gehören und

(8) $$A \subset \underset{n}{\mathfrak{S}} A_n$$

ist, so gilt

(9) $$\mathsf{P}(A) \leqq \sum_n \mathsf{P}(A_n).$$

Beweis.

$$A = A\,\underset{n}{\mathfrak{S}}(A_n) = A A_1 + A(A_2 - A_2 A_1) + A(A_3 - A_3 A_2 - A_3 A_1) + \cdots,$$

$$\mathsf{P}(A) = \mathsf{P}(A A_1) + \mathsf{P}\{A(A_2 - A_2 A_1)\} + \cdots \leqq \mathsf{P}(A_1) + \mathsf{P}(A_2) + \cdots.$$

§ 2. Borelsche Wahrscheinlichkeitsfelder.

Ein Mengenkörper $\mathfrak{F}$ ist ein Borel*scher Körper,* wenn alle abzählbaren Summen $\sum_n A_n$ der Mengen A_n aus $\mathfrak{F}$ zu $\mathfrak{F}$ gehören. Man nennt Borelsche Körper auch vollständig additive Mengensysteme. Aus der Formel

(1) $$\underset{n}{\mathfrak{S}} A_n = A_1 + (A_2 - A_2 A_1) + (A_3 - A_3 A_2 - A_3 A_1) + \cdots$$

schließt man, daß ein Borelscher Körper auch alle Vereinigungsmengen $\underset{n}{\mathfrak{S}} A_n$ abzählbar vieler zu ihm gehörender Mengen A_n enthält. Aus der Formel

(2) $$\underset{n}{\mathfrak{D}} A_n = E - \underset{n}{\mathfrak{S}} \bar{A}_n$$

folgt dasselbe für Durchschnittsmengen.

Ein Wahrscheinlichkeitsfeld ist ein Borel*sches Wahrscheinlichkeitsfeld, wenn der entsprechende Mengenkörper $\mathfrak{F}$ ein* Borel*scher ist.* Nur in Borelschen Wahrscheinlichkeitsfeldern erhält die Wahrscheinlichkeitsrechnung eine vollständige Handlungsfreiheit, die mit keiner Gefahr verbunden ist, zu Ereignissen zu gelangen, welche keine Wahrscheinlichkeit besitzen.

Wir wollen jetzt zeigen, *daß man sich auf die Betrachtung der* Borel*schen Wahrscheinlichkeitsfelder beschränken kann.* Dies ergibt sich aus dem Erweiterungssatze, zu dem wir gleich übergehen.

Es sei ein Wahrscheinlichkeitsfeld $(\mathfrak{F}, \mathsf{P})$ gegeben. Bekanntlich[1] existiert ein kleinster Borelscher Körper $B\mathfrak{F}$ über $\mathfrak{F}$. Sodann gilt der

Erweiterungssatz. *Man kann immer die auf $\mathfrak{F}$ definierte nichtnegative, vollständig additive Mengenfunktion $\mathsf{P}(A)$ auf alle Mengen von*

[1] Vgl. Hausdorff: Mengenlehre 1927 S. 79.

$B\mathfrak{F}$ mit der Erhaltung der beiden Eigenschaften (nichtnegativ und vollständig additiv zu sein) *erweitern, und zwar auf eine einzige Weise.*

Der erweiterte Körper $B\mathfrak{F}$ bildet also mit der erweiterten Mengenfunktion $\mathsf{P}(A)$ auch ein Wahrscheinlichkeitsfeld $(B\mathfrak{F}, \mathsf{P})$; dieses Wahrscheinlichkeitsfeld $(B\mathfrak{F}, \mathsf{P})$ nennen wir *die* BOREL*sche Erweiterung des Feldes* $(\mathfrak{F}, \mathsf{P})$.

Der Beweis dieses Erweiterungssatzes, welcher der Theorie der additiven Mengenfunktionen angehört und in verschiedenen anderen Fassungen im wesentlichen bekannt sein dürfte, verläuft folgendermaßen:

Es sei A eine beliebige Untermenge von E, wir definieren $\mathsf{P}^*(A)$ als untere Grenze der Summen

$$\sum_n \mathsf{P}(A_n)$$

für alle Überdeckungen

$$A \subset \underset{n}{\mathfrak{S}} A_n$$

der Menge A mit endlich oder abzählbar vielen Mengen A_n aus $\mathfrak{F}$. Man beweist leicht, daß $\mathsf{P}^*(A)$ ein äußeres Maß im Sinne von CARATHÉODORY ist[1]. Für die Mengen A aus $\mathfrak{F}$ fällt $\mathsf{P}^*(A)$ nach dem Überdeckungssatz (§ 1) mit $\mathsf{P}(A)$ zusammen. Man beweist weiter, daß alle Mengen von $\mathfrak{F}$ im CARATHÉODORYschen Sinne meßbar sind. Da alle meßbaren Mengen einen BORELschen Körper bilden, so sind also alle Mengen von $B\mathfrak{F}$ meßbar. Die Mengenfunktion $\mathsf{P}^*(A)$ ist also auf $B\mathfrak{F}$ vollständig additiv, und wir können auf $B\mathfrak{F}$

$$\mathsf{P}(A) = \mathsf{P}^*(A)$$

setzen. Die Existenz der Erweiterung ist somit bewiesen. Die Eindeutigkeit der Erweiterung folgt unmittelbar aus der Minimaleigenschaft des Körpers $B\mathfrak{F}$.

Bemerkung. Wenn man die Mengen (Ereignisse) A aus $\mathfrak{F}$ als reelle und (vielleicht nur annäherungsweise) beobachtbare Ereignisse deuten kann, so folgt daraus natürlich noch nicht, daß die Mengen des erweiterten Körpers $B\mathfrak{F}$ eine solche Deutung als reelle beobachtbare Ereignisse vernünftigerweise gestatten. Es kann also vorkommen, daß das Wahrscheinlichkeitsfeld $(\mathfrak{F}, \mathsf{P})$ als ein (vielleicht idealisiertes) Bild reeller zufälliger Erscheinungen betrachtet werden kann, während das erweiterte Wahrscheinlichkeitsfeld $(B\mathfrak{F}, \mathsf{P})$ eine rein mathematische Konstruktion bleibt.

Mengen aus $B\mathfrak{F}$ betrachten wir also im allgemeinen nur als „ideelle Ereignisse", welchen in der Erfahrungswelt nichts entspricht. Wenn aber eine Deduktion, die die Wahrscheinlichkeiten solcher ideellen Ereignisse gebraucht, zur Bestimmung der Wahrscheinlichkeit eines reellen Ereignisses aus $\mathfrak{F}$ führt, wird diese Bestimmung offensichtlich automatisch auch vom empirischen Standpunkte aus einwandfrei sein.

[1] CARATHÉODORY: Vorlesungen über reelle Funktionen 1918 S. 237—258.

§ 3. Beispiele unendlicher Wahrscheinlichkeitsfelder.

I. Schon im ersten Kapitel, § 1 haben wir verschiedene endliche Wahrscheinlichkeitsfelder konstruiert. Es sei jetzt die Menge

$$E = \{\xi_1, \xi_2, \ldots, \xi_n, \ldots\}$$

abzählbar und $\mathfrak{F}$ mit der Gesamtheit der Untermengen der Menge E identisch.

Alle möglichen Wahrscheinlichkeitsfelder mit einer solchen Menge $\mathfrak{F}$ erhält man folgendermaßen: Man wählt eine Folge von nichtnegativen Zahlen p_n mit

$$p_1 + p_2 + \cdots + p_n + \cdots = 1$$

und setzt für jede Menge A

$$\mathsf{P}(A) = \sum_n{}' p_n,$$

wobei die Summation $\sum'$ sich auf alle Indizes n erstreckt, für welche ξ_n zu A gehört. Diese Wahrscheinlichkeitsfelder sind offensichtlich BORELsche.

II. Wir nehmen jetzt an, daß E die reelle Zahlengerade ist. Zuerst sei dabei $\mathfrak{F}$ aus allen endlichen Summen von halb offenen Intervallen $[a; b) = \{a \leqq \xi < b\}$ gebildet (dabei betrachten wir neben den eigentlichen Intervallen mit endlichen a und b auch die uneigentlichen $[-\infty; a)$, $[a; +\infty)$ und $[-\infty; +\infty)$). Man überzeugt sich leicht, daß $\mathfrak{F}$ ein Körper ist. Nach dem Erweiterungssatz kann man aber jedes Wahrscheinlichkeitsfeld auf $\mathfrak{F}$ in ein solches auf $B\mathfrak{F}$ erweitern. Das Mengensystem $B\mathfrak{F}$ ist in unserem Falle nichts anderes als das System aller BORELschen Punktmengen der Zahlengeraden E. Wir können also direkt zum folgenden Falle übergehen.

III. Es wird jetzt angenommen, daß E die reelle Zahlengerade ist und $\mathfrak{F}$ aus allen BORELschen Punktmengen dieser Gerade besteht. Um ein Wahrscheinlichkeitsfeld mit diesem Körper $\mathfrak{F}$ zu konstruieren, genügt es, eine beliebige nichtnegative vollständig additive Mengenfunktion $\mathsf{P}(A)$ auf $\mathfrak{F}$ zu definieren, welche die Bedingung $\mathsf{P}(E) = 1$ befriedigt. Eine solche Funktion ist bekanntlich[1] durch ihre Werte

$$\mathsf{P}[-\infty; x) = F(x) \tag{1}$$

für die speziellen Intervalle $[-\infty; x)$ eindeutig bestimmt. Die Funktion $F(x)$ nennt man *Verteilungsfunktion* von ξ. Es ist weiter bewiesen (drittes Kap., § 2), daß $F(x)$ nicht abnehmend und nach links stetig ist und die Limeswerte

$$\lim_{x \to -\infty} F(x) = F(-\infty) = 0, \qquad \lim_{x \to +\infty} F(x) = F(+\infty) = 1 \tag{2}$$

besitzt. Ist umgekehrt eine diesen Bedingungen genügende Funktion

[1] Vgl. z. B. LEBESGUE: Leçons sur l'intégration 1928 S. 152—156.

$F(x)$ gegeben, so bestimmt sie immer eine nichtnegative vollständig additive Mengenfunktion $\mathsf{P}(A)$ mit $\mathsf{P}(E) = 1$ *.

IV. Es sei jetzt der n-dimensionale euklidische Koordinatenraum R^n, d. h. die Menge aller geordneten n-Tupeln $\xi = \{x_1, x_2, \ldots, x_n\}$ reeller Zahlen, als die Grundmenge E gewählt. $\mathfrak{F}$ bestehe dabei aus allen Borelschen Punktmengen[1] von R^n. (Auf die Betrachtung verschiedener engerer Mengensysteme, z. B. des Systems aller n-dimensionalen Intervalle, können wir aus Gründen, die den in II angegebenen analog sind, verzichten.)

Als Wahrscheinlichkeitsfunktion $\mathsf{P}(A)$ kann man hier wie immer eine beliebige nichtnegative und vollständig additive Mengenfunktion wählen, welche auf $\mathfrak{F}$ definiert ist und der Bedingung $\mathsf{P}(E) = 1$ genügt. Eine solche Mengenfunktion wird eindeutig bestimmt, wenn man ihre Werte

$$\mathsf{P}(L_{a_1 a_2 \ldots a_n}) = F(a_1, a_2, \ldots, a_n) \tag{3}$$

für die speziellen Mengen $L_{a_1 a_2 \ldots a_n}$ angibt, wobei $L_{a_1 a_2 \ldots a_n}$ die Menge aller ξ mit $x_i < a_i$, $i = 1, 2, \ldots, n$, bezeichnet. Als Funktion $F(a_1, a_2, \ldots, a_n)$ kann man dabei eine beliebige in jeder Veränderlichen nicht abnehmende und nach links stetige Funktion wählen, welche den folgenden Bedingungen genügt:

$$\left\{\begin{aligned} &\lim_{a_i \to -\infty} F(a_1, a_2, \ldots, a_n) = F(a_1, \ldots, a_{i-1}, -\infty, a_{i+1}, \ldots, a_n) = 0, \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad i = 1, 2, \ldots, n \\ &\lim_{a_1 \to +\infty, a_2 \to +\infty, \ldots, a_n \to +\infty} F(a_1, a_2, \ldots, a_n) = F(+\infty, +\infty, \ldots, +\infty) = 1. \end{aligned}\right. \tag{4}$$

Man nennt $F(a_1, a_2, \ldots, a_n)$ die *Verteilungsfunktion* der Größen $x_1, x_2, \ldots, x_n$.

Die Betrachtung der Wahrscheinlichkeitsfelder des oben definierten Typus genügt vollständig für alle klassischen Probleme der Wahrscheinlichkeitsrechnung[2].

Insbesondere definiert man eine Wahrscheinlichkeitsfunktion in R^n, wenn man eine beliebige im R^n definierte nichtnegative Punktfunktion $f(x_1, x_2, \ldots, x_n)$ mit

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \ldots dx_n = 1$$

nimmt und

$$\mathsf{P}(A) = \iint \cdots \int_A f(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \ldots dx_n \tag{3}$$

* Vgl. die vorige Fußnote.

[1] Für die Definition der Borelschen Mengen des R^n vgl. Hausdorff: Mengenlehre 1927 S. 177—181.

[2] Vgl. z. B. R. v. Mises: [1], S. 13—19. Hier fordert man die Existenz der Wahrscheinlichkeiten $\mathsf{P}(A)$ für „alle praktisch vorkommenden" Mengen des n-dimensionalen Raumes.

setzt. Dann ist $f(x_1, x_2, \ldots, x_n)$ *die Wahrscheinlichkeitsdichte* im Punkte $(x_1, x_2, \ldots, x_n)$ (vgl. drittes Kap., § 2).

Ein anderer Typus der Wahrscheinlichkeitsfunktionen in R^n erhält man folgendermaßen: Es sei $\{\xi_i\}$ eine Folge von Punkten aus R^n und $\{p_i\}$ eine Folge von nichtnegativen reellen Zahlen mit $\sum p_i = 1$; dann setzt man wie im Beispiele I

$$\mathsf{P}(A) = \sum{}' p_i,$$

wobei die Summation $\sum{}'$ sich auf alle Indizes i erstreckt, für welche ξ zu A gehört. Zwei hier erwähnte Typen der Wahrscheinlichkeitsfunktionen in R^n erschöpfen nicht alle Möglichkeiten, obwohl man sich mit ihnen in den Anwendungen der Wahrscheinlichkeitsrechnung gewöhnlich begnügt.

Man kann sich jedoch außer diesem klassischen Gebiet auch für Anwendungen interessante Probleme vorstellen, in welchen Elementarereignisse mit Hilfe unendlich vieler Koordinaten definiert sind. Entsprechende Wahrscheinlichkeitsfelder wollen wir erst nach Einführung einiger hierzu notwendigen Hilfsbegriffe näher untersuchen (vgl. drittes Kap., § 3).

Drittes Kapitel.

Zufällige Größen.

§ 1. Wahrscheinlichkeitsfunktionen.

Es sei eine Abbildung der Menge E in eine Menge E' von irgendwelchen Elementen, d. h. eine auf E definierte eindeutige Funktion $u(\xi)$, deren Werte zur Menge E' gehören, gegeben. Jeder Untermenge A' von E' ordnen wir dann als ihr *Urbild* in E die Menge $u^{-1}(A')$ aller Elemente von E zu, welche auf ein Element von A' abgebildet sind. Es sei weiter $\mathfrak{F}^{(u)}$ das System aller Untermengen A' von E', deren Urbilder zum Mengenkörper $\mathfrak{F}$ gehören. $\mathfrak{F}^{(u)}$ ist dann auch ein Körper; ist dabei $\mathfrak{F}$ ein Borelscher Körper, so gilt dasselbe auch von $\mathfrak{F}^{(u)}$. Wir setzen jetzt

$$(1) \qquad \mathsf{P}^{(u)}(A') = \mathsf{P}\{u^{-1}(A')\}.$$

Diese auf $\mathfrak{F}^{(u)}$ definierte Mengenfunktion $\mathsf{P}^{(u)}$ genügt in bezug auf den Mengenkörper $\mathfrak{F}^{(u)}$ allen unseren Axiomen I—VI, ist folglich eine Wahrscheinlichkeitsfunktion auf $\mathfrak{F}^{(u)}$. Bevor wir zum Beweise aller soeben angegebenen Tatsachen übergehen, wollen wir schon jetzt die folgende Definition aussprechen:

Definition. Es sei eine eindeutige Funktion $u(\xi)$ des zufälligen Ereignisses ξ gegeben. Dann heißt die durch (1) definierte Mengenfunktion $\mathsf{P}^{(u)}(A')$ *die Wahrscheinlichkeitsfunktion* von u.

Bemerkung 1. Man nennt bei der Untersuchung des Wahrscheinlichkeitsfeldes $(\mathfrak{F}, \mathsf{P})$ die Funktion $\mathsf{P}(A)$ schlechthin die Wahrscheinlichkeitsfunktion und $\mathsf{P}^{(u)}(A')$ die Wahrscheinlichkeitsfunktion von u. Im Falle $u(\xi) = \xi$ fällt $\mathsf{P}^{(u)}(A')$ mit $\mathsf{P}(A)$ zusammen.

Bemerkung 2. Das Ereignis $u^{-1}(A')$ besteht darin, daß $u(\xi)$ zur Menge A' gehört. Folglich ist $\mathsf{P}^{(u)}(A')$ die Wahrscheinlichkeit dafür, daß $u(\xi) \subset A'$ gilt.

Es bleibt uns die obenerwähnten Eigenschaften von $\mathfrak{F}^{(u)}$ und $\mathsf{P}^{(u)}$ zu beweisen, Sie folgen aber aus einer einzigen fundamentalen Tatsache, und zwar aus der folgenden:

Hilfssatz. Vereinigungsmengen, Durchschnitte und Differenzen beliebiger Urbildmengen $u^{-1}(A')$ sind Urbilder der entsprechenden Vereinigungsmengen, Durchschnitte und Differenzen originaler Mengen A'.

Der Beweis dieses Hilfssatzes darf dem Leser überlassen werden.

Es seien jetzt A' und B' zwei Mengen aus $\mathfrak{F}^{(u)}$, ihre Urbilder A und B gehören dann zu $\mathfrak{F}$; da $\mathfrak{F}$ ein Körper ist, gehören auch die Mengen $AB, A + B$ und $A - B$ zu $\mathfrak{F}$, diese Mengen sind aber die Urbilder der Mengen $A'B', A' + B'$ und $A' - B'$, folglich gehören letztere Mengen zu $\mathfrak{F}^{(u)}$. Wir haben also bewiesen, daß $\mathfrak{F}^{(u)}$ ein Körper ist. Ebenso beweist man, daß, wenn F ein Borelscher Körper ist, dasselbe auch von $\mathfrak{F}^{(u)}$ gilt.

Es ist ferner klar, daß

$$\mathsf{P}^{(u)}(E') = \mathsf{P}\{u^{-1}(E')\} = \mathsf{P}(E) = 1$$

ist. Daß $\mathsf{P}^{(u)}$ immer nichtnegativ ist, ist selbstverständlich. Es bleibt folglich zu beweisen, daß $\mathsf{P}^{(u)}$ vollständig additiv ist (vgl. das Ende des § 1, zweites Kap.). Es seien also die Mengen A_n, folglich auch ihre Urbilder $u^{-1}(A'_n)$ disjunkt; daraus folgt aber, daß

$$\mathsf{P}^{(u)}(\sum_n A'_n) = \mathsf{P}\{u^{-1}(\sum_n A'_n)\} = \mathsf{P}\{\sum_n u^{-1}(A'_n)\}$$
$$= \sum_n \mathsf{P}\{u^{-1}(A'_n)\} = \sum_n \mathsf{P}^{(u)}(A'_n)$$

ist, womit die vollständige Additivität von P' bewiesen ist.

Zum Schluß wollen wir noch folgendes bemerken. Es sei $u_1(\xi)$ eine Funktion, welche E in E' abbildet, und $u_2(\xi')$ eine andere Funktion, welche E' in E'' abbildet. Dann bildet die zusammengesetzte Funktion $u_2 u_1(\xi)$ die Menge E in E'' ab. Wir betrachten jetzt die Wahrscheinlichkeitsfunktionen $P^{(u_1)}(A')$ und $P^{(u)}(A'')$ der Funktionen $u_1(\xi)$ und $u(\xi) = u_2 u_1(\xi)$. Diese zwei Wahrscheinlichkeitsfunktionen sind, wie man leicht berechnet, durch die folgende Relation verbunden:

$$\mathsf{P}^{(u)}(A'') = \mathsf{P}^{(u_1)}\{u_2^{-1}(A'')\}. \tag{2}$$

§ 2. Definition der zufälligen Größen, Verteilungsfunktionen.

Definition. Eine auf der Grundmenge E definierte eindeutige reelle Funktion $x(\xi)$ nennt man eine *zufällige Größe*, wenn bei jeder

Wahl einer reellen Zahl a die Menge $\{x < a\}$ aller ξ, für welche die Ungleichung $x < a$ gilt, zum Mengensystem $\mathfrak{F}$ gehört.

Die Funktion $x(\xi)$ bildet die Grundmenge E in die Menge R^1 aller reellen Zahlen ab. Diese Funktion bestimmt nach § 1 einen Körper $\mathfrak{F}^{(x)}$ von Untermengen der Menge R^1. Unsere Definition einer zufälligen Größe kann jetzt so formuliert werden: Eine reelle Funktion $x(\xi)$ ist dann und nur dann eine zufällige Größe, wenn $\mathfrak{F}^{(x)}$ jedes Intervall von der Form $(-\infty, a)$ enthält.

Da $\mathfrak{F}^{(x)}$ ein Körper ist, so enthält er mit den Intervallen $(-\infty; a)$ auch alle möglichen endlichen Summen von halboffenen Intervallen $[a; b)$. Wenn unser Wahrscheinlichkeitsfeld ein Borelsches ist, so sind $\mathfrak{F}$ und $\mathfrak{F}^{(x)}$ Borelsche Körper, *folglich enthält in diesem Falle $\mathfrak{F}^{(x)}$ alle* Borel*schen Mengen von R^1*.

Die Wahrscheinlichkeitsfunktion einer zufälligen Größe bezeichnen wir weiter mit $\mathsf{P}^{(x)}(A')$. Sie ist für alle Mengen des Körpers $F^{(x)}$ definiert. Im wichtigsten Falle des Borelschen Wahrscheinlichkeitsfeldes ist also $\mathsf{P}^{(x)}$ insbesondere für alle Borelschen Mengen des R^1 definiert.

Definition. Die Funktion

$$F^{(x)}(a) = \mathsf{P}^{(x)}(-\infty, a) = \mathsf{P}\{x < a\},$$

wobei $-\infty$ und $+\infty$ als Werte von a zugelassen sind, heißt *die Verteilungsfunktion der zufälligen Größe x.*

Es folgt unmittelbar aus unserer Definition, daß

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

ist. Die Wahrscheinlichkeit der Erfüllung beider Ungleichungen $a \leqq x < b$ ist offenbar durch die Formel

$$\mathsf{P}\{x \subset [a; b)\} = F^{(x)}(b) - F^{(x)}(a) \tag{2}$$

gegeben. Es folgt daraus, daß für $a < b$

$$F^{(x)}(a) \leqq F^{(x)}(b)$$

ist, wenn nur $a < b$ ist, d. h. daß $F^{(x)}(a)$ *eine nichtabnehmende Funktion ist.*

Es sei ferner $a_1 < a_2 < \cdots < a_n < \cdots < b$; dann ist

$$\mathfrak{D}_n\{x \subset [a_n; b)\} = 0,$$

folglich konvergiert wegen des Stetigkeitsaxioms

$$F^{(x)}(b) - F^{(x)}(a_n) = \mathsf{P}\{x \subset [a_n, b)\}$$

mit $n \to +\infty$ gegen Null. Man sieht also, daß $F^{(x)}(a)$ *linkseitig stetig ist.*

Es lassen sich analog die Formeln

$$\lim F^{(x)}(a) = F^{(x)}(-\infty) = 0, \qquad a \to -\infty, \tag{3}$$

$$\lim F^{(x)}(a) = F^{(x)}(+\infty) = 1, \qquad a \to +\infty, \tag{4}$$

beweisen.

Ist das Wahrscheinlichkeitsfeld $(\mathfrak{F}, \mathsf{P})$ ein BORELsches, so sind die Werte der Wahrscheinlichkeitsfunktion $\mathsf{P}^{(x)}(A)$ für alle BORELschen Mengen A des R^1 durch die Kenntnis der Verteilungsfunktion $F^{(x)}(a)$ eindeutig bestimmt (vgl. zweites Kap., § 3, III). Da man sich meistens nur für solche Werte von $\mathsf{P}^{(x)}(A)$ interessiert, spielen die Verteilungsfunktionen in der ganzen weiteren Darstellung eine wesentliche Rolle.

Wenn die Verteilungsfunktion $F^{(x)}(a)$ differentiierbar ist, so nennt man ihre Ableitung

$$f^{(x)}(a) = \frac{d}{da} F^{(x)}(a)$$

nach a *die Wahrscheinlichkeitsdichte* von x im Punkte a. Ist dabei für jedes a $F^{(x)}(a) = \int\limits_{-\infty}^{a} f^{(x)}(a)\, da$, so läßt sich die Wahrscheinlichkeitsfunktion $\mathsf{P}^{(x)}(A)$ für eine BORELsche Menge A folgendermaßen durch $f(a)$ ausdrücken:

$$\mathsf{P}^{(x)}(A) = \int\limits_{A} f^{(x)}(a)\, da\,. \tag{5}$$

In diesem Falle sagt man, daß *die Verteilung von x stetig ist.* Im allgemeinen Falle schreibt man analog

$$\mathsf{P}^{(x)}(A) = \int\limits_{A} dF^{(x)}(a)\,. \tag{6}$$

Alle angeführten Begriffsbildungen lassen sich auch auf den Fall der bedingten Wahrscheinlichkeiten verallgemeinern. Die Mengenfunktion

$$\mathsf{P}_B^{(x)}(A) = \mathsf{P}_B(x \subset A)$$

ist die bedingte Wahrscheinlichkeitsfunktion von x in der Hypothese B, die nichtabnehmende Funktion

$$F_B^{(x)}(a) = \mathsf{P}_B(x < a)$$

ist die entsprechende Verteilungsfunktion, und endlich [im Falle der Differentiierbarkeit von $F_B^{(x)}(a)$] ist

$$f_B^{(x)}(a) = \frac{d}{da} F_B^{(x)}(a)$$

die bedingte Wahrscheinlichkeitsdichte von x im Punkte a bei der Hypothese B.

§ 3. Mehrdimensionale Verteilungsfunktionen.

Es seien jetzt n zufällige Größen $x_1, x_2, \ldots, x_n$ gegeben. Der Punkt $x = (x_1, x_2, \ldots, x_n)$ des n-dimensionalen Raumes R ist eine Funktion des elementaren Ereignisses ξ. Man erhält folglich nach den allgemeinen Regeln des § 1 einen Mengenkörper $\mathfrak{F}^{(x_1, x_2, \ldots, x_n)}$ von Untermengen des

Raumes R^n und eine auf $\mathfrak{F}'$ definierte Wahrscheinlichkeitsfunktion $\mathsf{P}^{(x_1, x_2, \ldots, x_n)}(A')$. Diese Wahrscheinlichkeitsfunktion $\mathsf{P}^{(x_1, x_2, \ldots, x_n)}$ nennt man *die* n-dimensionale *Wahrscheinlichkeitsfunktion der zufälligen Größen* $x_1, x_2 \ldots, x_n$.

Der Körper $\mathfrak{F}'$ enthält, wie es direkt aus der Definition einer zufälligen Größe folgt, bei jeder Wahl von i und a_i $(i = 1, 2, \ldots, n)$ die Menge aller Punkte von R^n, für welche $x_i < a_i$ ist. Folglich enthält $\mathfrak{F}'$ auch den Durchschnitt der genannten Mengen, d. h. die Menge $L_{a_1 a_2 \ldots a_n}$ aller Punkte von R^n, für welche alle Ungleichungen $x_i < a_i$, $i = 1, 2, \ldots, n$, gelten[1]. Bezeichnet man als ein n-dimensionales halboffenes Intervall

$$[a_1, a_2, \ldots, a_n; \quad b_1, b_2, \ldots, b_n)$$

die Menge aller Punkte von R^n, für welche alle Ungleichungen $a_i \leqq x_i < b_i$ erfüllt sind, so sieht man sofort ein, daß jedes solches Intervall auch zum Körper $\mathfrak{F}'$ gehört, denn es ist

$$[a_1, a_2, \ldots, a_n; \quad b_1, b_2, \ldots, b_n)$$
$$= L_{b_1 b_2 \ldots b_n} - L_{a_1 b_2 \ldots b_n} - L_{b_1 a_2 b_3 \ldots b_n} - \cdots - L_{b_1 b_2 \ldots b_{n-1} a_n}.$$

Die BORELsche Erweiterung des Systems aller n-dimensionalen halboffenen Intervalle besteht aber aus allen BORELschen Mengen des R^n. Es folgt daraus, daß *der Körper $\mathfrak{F}'$ im Falle eines* BOREL*schen Wahrscheinlichkeitsfeldes alle* BOREL*schen Mengen des Raumes R^n enthält.*

Satz. *Im Falle eines* BOREL*schen Wahrscheinlichkeitsfeldes ist jede* BOREL*sche Funktion $x = f(x_1, x_2, \ldots, x_n)$ endlich vieler zufälliger Größen $x_1, x_2, \ldots, x_n$ wieder eine zufällige Größe.*

Zum Beweis genügt es, zu bemerken, daß die Menge aller Punkte $(x_1, x_2, \ldots, x_n)$ des R^n, für welche $x = f(x_1, x_2, \ldots, x_n) < a$ gilt, eine BORELsche ist. Insbesondere sind also endliche Summen und Produkte zufälliger Größen wieder zufällige Größen.

Definition. Die Funktion

$$F^{(x_1, x_2, \ldots, x_n)}(a_1, a_2, \ldots, a_n) = \mathsf{P}^{(x_1, x_2, \ldots, x_n)}(L_{a_1 a_2 \ldots a_n})$$

nennen wir *die* n-dimensionale *Verteilungsfunktion* der zufälligen Größen $x_1, x_2, \ldots, x_n$.

Man beweist wie im eindimensionalen Falle, daß $F^{(x_1, x_2, \ldots, x_n)}(a_1, a_2, \ldots, a_n)$ in jeder Variablen nicht abnehmend und nach links stetig ist. Analog zu den Gleichungen (3) bis (4) des § 2 hat man hier

$$\text{(7)} \quad \lim F(a_1, a_2, \ldots, a_n) = F(a_1, \ldots, a_{i-1}, -\infty, a_{i+1}, \ldots, a_n) = 0,$$

$$\text{(8)} \quad \lim_{a_1 \to +\infty,\, a_2 \to +\infty, \ldots,\, a_n \to +\infty} F(a_1, a_2, \ldots, a_n) = F(+\infty, +\infty, \ldots, +\infty) = 1.$$

Die Verteilungsfunktion $F^{(x_1 x_2 \ldots x_n)}$ gibt uns direkt nur die Werte von $\mathsf{P}^{(x_1, x_2, \ldots, x_n)}$ für die speziellen Mengen $L_{a_1 a_2 \ldots a_n}$. Wenn aber unser Wahrscheinlichkeitsfeld ein BORELsches ist, so ist $\mathsf{P}^{(x_1, x_2, \ldots, x_n)}$ *durch die*

[1] a_i können dabei auch unendliche Werte $\pm\infty$ annehmen.

Kenntnis der Verteilungsfunktion $F^{(x_1, x_2, \ldots, x_n)}$ *für alle* BORELschen *Mengen von* R^n *eindeutig bestimmt*[1].

Wenn die Ableitung

$$f(a_1, a_2, \ldots, a_n) = \frac{\partial^n}{\partial a_1 \partial a_2 \ldots \partial a_n} F^{(x_1, x_2, \ldots, x_n)}(a_1, a_2, \ldots, a_n)$$

existiert, so nennt man diese Ableitung f *die* n-dimensionale *Wahrscheinlichkeitsdichte* der zufälligen Größen $x_1, x_2, \ldots, x_n$ im Punkte $a_1, a_2, \ldots, a_n$. Ist dabei für jeden Punkt $(a_1, a_2, \ldots, a_n)$

$$F^{(x_1, x_2, \ldots, x_n)}(a_1 a_2 \ldots a_n) = \int_{-\infty}^{a_1} \int_{-\infty}^{a_2} \ldots \int_{-\infty}^{a_n} f(a_1, a_2, \ldots, a_n)\, d a_1\, d a_2 \ldots d a_n,$$

so nennt man die Verteilung von $(a_1, a_2, \ldots, a_n)$ *stetig*. Für jede BORELsche Menge $A \subset R^n$ gilt dann die Gleichung

$$\mathsf{P}^{(x_1, x_2, \ldots, x_n)}(A) = \iint \ldots \int_A f(a_1, a_2, \ldots, a_n)\, d a_1\, d a_2 \ldots d a_n. \tag{9}$$

Zum Schluß dieses Paragraphen machen wir noch eine Bemerkung über Relationen zwischen verschiedenen Wahrscheinlichkeits- und Verteilungsfunktionen. Es sei zuerst eine Permutation

$$S = \begin{pmatrix} 1, & 2, & \ldots, & n \\ i_1, & i_2, & \ldots, & i_n \end{pmatrix}$$

gegeben, und r_S bezeichne die Transformation

$$x'_k = x_{i_k} \qquad k = 1, 2, \ldots, n$$

des Raumes R^n in sich selbst. Dann gilt offenbar

$$\mathsf{P}^{(x_{i_1}, x_{i_2}, \ldots, x_{i_n})}(A) = \mathsf{P}^{(x_1, x_2, \ldots, x_n)}\{r_S^{-1}(A)\}. \tag{10}$$

Es sei jetzt $x' = p_k(x)$ „die Projektion" des Raumes R^n in den Raum $R^k (k < n)$, bei welcher der Punkt $(x_1, x_2, \ldots, x_n)$ in den Punkt $(x_1, x_2, \ldots, x_k)$ abgebildet wird. Dann gilt infolge der Formel (2) des § 1

$$\mathsf{P}^{(x_1, x_2, \ldots, x_k)}(A) = \mathsf{P}^{(x_1, x_2, \ldots, x_n)}\{p_k^{-1}(A)\}. \tag{11}$$

Für die entsprechenden Verteilungsfunktionen folgen aus (10) und (11) die Gleichungen

$$F^{(x_{i_1}, x_{i_2}, \ldots, x_{i_n})}(a_{i_1}, a_{i_2}, \ldots, a_{i_n}) = F^{(x_1, x_2, \ldots, x_n)}(a_1, a_2, \ldots, a_n), \tag{12}$$

$$F^{(x_1, x_2, \ldots, x_k)}(a_1, a_2, \ldots, a_k) = F^{(x_1, x_2, \ldots, x_n)}(a_1, \ldots, a_k, +\infty, \ldots, +\infty). \tag{13}$$

§ 4. Wahrscheinlichkeiten in unendlichdimensionalen Räumen.

Wir haben in § 3 des zweiten Kapitels gesehen, wie man verschiedene in der Wahrscheinlichkeitsrechnung übliche Wahrscheinlichkeitsfelder

[1] Vgl. zweites Kapitel, § 3, IV.

konstruiert. Man kann sich jedoch auch für Anwendungen interessante Probleme vorstellen, in welchen Elementarereignisse mit Hilfe unendlichvieler Koordinaten definiert sind. Es sei also eine Menge M — beliebiger Mächtigkeit $\mathfrak{m}$ — der Indizes μ gewählt. Die Gesamtheit aller Systeme

$$\xi = \{x_\mu\}$$

reeller Zahlen x_μ, wobei μ die ganze Menge M durchläuft, wird als der Raum R^M bezeichnet (um ein Element ξ des Raumes R^M zu bestimmen, soll man also jedem Elemente μ der Menge M eine bestimmte reelle Zahl x_μ zuordnen oder, was dasselbe ist, eine auf M definierte eindeutige reelle Funktion x_μ des Elementes μ angeben)[1].

Wenn die Menge M aus n ersten natürlichen Zahlen $1, 2, \ldots, n$ besteht, so ist R^M der gewöhnliche n-dimensionale Raum R^n. Wählt man als Menge M die Menge aller reellen Zahlen R^1, so besteht der entsprechende Raum $R^M = R^{R^1}$ aus allen reellen Funktionen

$$\xi(\mu) = x_\mu$$

des reellen Argumentes μ.

Die Menge R^M (mit einer beliebigen Menge M) nehmen wir jetzt zur Grundmenge E. Es sei $\xi = \{x_\mu\}$ ein Element von E; wir bezeichnen dann mit $p_{\mu_1 \mu_2 \ldots \mu_n}(\xi)$ den Punkt $(x_{\mu_1}, x_{\mu_2}, \ldots, x_{\mu_n})$ des n-dimensionalen Raumes R^n. Eine Untermenge A von E nennen wir *eine Zylindermenge,* wenn sie in der Form

$$A = \overset{-1}{p}_{\mu_1 \mu_2 \ldots \mu_n}(A')$$

darstellbar ist, wobei A' eine Untermenge von R^n ist. Die Klasse aller Zylindermengen fällt also mit der Klasse von allen solchen Mengen, welche durch eine Relation von der Form

$$f(x_{\mu_1}, x_{\mu_2}, \ldots, x_{\mu_n}) = 0 \tag{1}$$

definiert werden können, zusammen. Um eine beliebige Zylindermenge $p_{\mu_1 \mu_2 \ldots \mu_n}(A')$ durch eine solche Relation zu bestimmen, genügt es, als f eine Funktion zu nehmen, welche auf A' gleich Null und außerhalb A' gleich Eins ist.

Eine Zylindermenge ist *eine* BOREL*sche Zylindermenge,* wenn die entsprechende Menge A' eine BORELsche ist. *Alle* BOREL*schen Zylindermengen des Raumes R^M bilden einen Körper, welcher weiter mit $\mathfrak{F}^M$ bezeichnet wird*[2].

---

[1] Vgl. HAUSDORFF: Mengenlehre 1927 S. 23.

[2] Aus den oben angeführten Betrachtungen folgt, daß BORELsche Zylindermengen diejenigen Mengen sind, welche durch BORELsche Relationen (1) definiert werden können. Es seien jetzt A und B zwei BORELsche Zylindermengen, welche durch die Relationen

$$f(x_{\mu_1}, x_{\mu_2}, \ldots, x_{\mu_n}) = 0, \qquad g(x_{\lambda_1}, x_{\lambda_2}, \ldots, x_{\lambda_m}) = 0$$

definiert sind. Dann kann man die Mengen $A + B$, AB und $A - B$ bzw. durch

Die BORELsche Erweiterung des Mengenkörpers $\mathfrak{F}^M$ bezeichnen wir wie immer mit $B\mathfrak{F}^M$. Die Mengen des Mengensystems $B\mathfrak{F}^M$ nennen wir die BOREL*schen Mengen* des Raumes R^M.

Es wird weiter eine Methode angegeben, Wahrscheinlichkeitsfunktionen auf $\mathfrak{F}^M$ und folglich nach dem Erweiterungssatz auch auf $B\mathfrak{F}^M$ zu konstruieren und zu behandeln. Die so erhaltenen Wahrscheinlichkeitsfelder reichen im Falle einer abzählbaren Menge M für alle Zwecke hin. Man beherrscht also alle Fragen über das Verhalten einer abzählbaren Folge von zufälligen Größen. Wenn aber M unabzählbar ist, so bleiben manche einfache und interessante Untermengen von R^M noch außerhalb des Mengensystems $B\mathfrak{F}^M$. Z. B. die Menge aller Elemente ξ, für welche x_μ bei jeder Wahl des Indizes μ kleiner als eine feste Konstante bleibt, gehört im Falle einer unabzählbaren Menge M nicht zum Mengensystem $B\mathfrak{F}^M$.

Es ist folglich immer erstrebenswert, jedes Problem, wenn möglich, in einer solchen Form zu bringen, bei welcher der Raum aller elementaren Ereignisse ξ nur abzählbar viele Koordinaten hat.

Es sei eine Wahrscheinlichkeitsfunktion $\mathsf{P}(A)$ auf $\mathfrak{F}^M$ definiert. Jede Koordinate x_μ des elementaren Ereignisses ξ kann dann als eine zufällige Größe betrachtet werden. Folglich besitzt jede endliche Gruppe $(x_{\mu_1}, x_{\mu_2}, \ldots, x_{\mu_n})$ von diesen Koordinaten eine n-dimensionale Wahrscheinlichkeitsfunktion $\mathsf{P}_{\mu_1\mu_2\ldots\mu_n}(A)$ und eine entsprechende Verteilungsfunktion $F_{\mu_1\mu_2\ldots\mu_n}(a_1, a_2, \ldots, a_n)$. Man hat offensichtlich für jede BORELsche Zylindermenge

$$A = p^{-1}_{\mu_1\mu_2\ldots\mu_n}(A')$$

die Gleichung

$$\mathsf{P}(A) = \mathsf{P}_{\mu_1\mu_2\ldots\mu_n}(A'),$$

wobei A' eine BORELsche Menge von R^n ist. Die Wahrscheinlichkeitsfunktion P ist also auf dem Körper $\mathfrak{F}^M$ aller Zylindermengen durch die Kenntnis der Werte aller endlichdimensionalen Wahrscheinlichkeitsfunktionen $\mathsf{P}_{\mu_1\mu_2\ldots\mu_n}$ für alle BORELschen Mengen der entsprechenden Räume R^n eindeutig bestimmt. Für BORELsche Mengen sind aber die Werte der Wahrscheinlichkeitsfunktionen $\mathsf{P}_{\mu_1\mu_2\ldots\mu_n}$ eindeutig durch die entsprechenden Verteilungsfunktionen $F_{\mu_1\mu_2\ldots\mu_n}$ bestimmt. Wir haben also den folgenden Satz bewiesen:

---

die folgenden Relationen definieren:

$$f \cdot g = 0,$$
$$f^2 + g^2 = 0,$$
$$f^2 + \omega(g) = 0,$$

wobei $\omega(x) = 0$ für $x \neq 0$ und $\omega(0) = 1$ gesetzt ist. Wenn f und g BORELsche Funktionen sind, so gilt dasselbe für die Funktionen $f \cdot g$, $f^2 + g^2$ und $f^2 + \omega(g)$, folglich sind $A + B$, AB und $A - B$ BORELsche Zylindermengen. Somit ist die Körpereigenschaft des Mengensystems $\mathfrak{F}^M$ bewiesen.

Die Gesamtheit aller endlichdimensionalen Verteilungsfunktionen $F_{\mu_1\mu_2\ldots\mu_n}$ *definiert eindeutig die Wahrscheinlichkeitsfunktion* $\mathsf{P}(A)$ *für alle Mengen von* F^M. *Ist* $\mathsf{P}(A)$ *auf* $B\mathfrak{F}^M$ *definiert, so ist sie* (nach dem Erweiterungssatz) *auch auf* $B\mathfrak{F}^M$ *eindeutig durch die Kenntnisse der Verteilungsfunktionen* $F_{\mu_1\mu_2\ldots\mu_n}$ *bestimmt.*

Es liegt jetzt nahe, sich die Frage zu stellen, unter welchen Bedingungen ein a priori gegebenes System von Verteilungsfunktionen $F_{\mu_1\mu_2\ldots\mu_n}$ ein Wahrscheinlichkeitsfeld auf $\mathfrak{F}^M$ (und folglich auch auf $B\mathfrak{F}^M$) bestimmt.

Zuerst bemerken wir, daß jede Verteilungsfunktion $F_{\mu_1\mu_2\ldots\mu_n}$ den im zweiten Kapitel, § 3, III angegebenen Bedingungen genügen muß, was allerdings im Begriffe der Verteilungsfunktion enthalten ist. Außerdem gelten wegen den Formeln (13) und (14) des § 2 noch die folgenden Relationen:

(2) $$F_{\mu_{i_1}\mu_{i_2}\ldots\mu_{i_n}}(a_{i_1}, a_{i_2}, \ldots, a_{i_n}) = F_{\mu_1\mu_2\ldots\mu_n}(a_1, a_2, \ldots, a_n),$$

(3) $$F_{\mu_1\mu_2\ldots\mu_k}(a_1, a_2, \ldots, a_k) = F_{\mu_1\mu_2\ldots\mu_n}(a_1, a_2, \ldots, a_k, +\infty, \ldots, +\infty),$$

wobei $k < n$ und $\begin{pmatrix}1, & 2, & \ldots, & n\\ i_1, & i_2, & \ldots, & i_n\end{pmatrix}$ eine beliebige Substitution ist. Diese notwendigen Bedingungen sind aber auch hinreichend, wie es sich aus dem folgenden Satze ergibt:

Hauptsatz. *Jedes System von Verteilungsfunktionen* $F_{\mu_1\mu_2\ldots\mu_n}$, *welche den Bedingungen* (2) *und* (3) *genügen, bestimmt eine Wahrscheinlichkeitsfunktion* $\mathsf{P}(A)$ *auf* $\mathfrak{F}^M$, *welche allen Axiomen I—VI genügt. Diese Wahrscheinlichkeitsfunktion* $\mathsf{P}(A)$ *kann* (nach dem Erweiterungssatz) *auch auf* $B\mathfrak{F}^M$ *erweitert werden.*

Beweis des Hauptsatzes. Es seien also die Verteilungsfunktionen $F_{\mu_1\mu_2\ldots\mu_n}$ gegeben, welche den allgemeinen Bedingungen vom zweiten Kapitel, § 3, III und den Bedingungen (2) und (3) genügen. Jede Verteilungsfunktion $F_{\mu_1\mu_2\ldots\mu_n}$ definiert eindeutig die entsprechende Wahrscheinlichkeitsfunktion $\mathsf{P}_{\mu_1\mu_2\ldots\mu_n}$ für alle Borelschen Mengen des R^n (vgl. § 3). Weiter betrachten wir nur Borelsche Mengen des R^n und Borelsche Zylindermengen in E.

Für jede Zylindermenge

$$A = p^{-1}_{\mu_1\mu_2\ldots\mu_n}(A')$$

setzen wir

(4) $$\mathsf{P}(A) = \mathsf{P}_{\mu_1\mu_2\ldots\mu_n}(A').$$

Da dieselbe Zylindermenge A durch verschiedene Mengen A' definiert werden kann, muß man zuerst beweisen, daß die Formel (4) immer denselben Wert für $\mathsf{P}(A)$ ergibt.

Es sei $(x_{\mu_1}, x_{\mu_2}, \ldots, x_{\mu_n})$ ein endliches System der zufälligen Größen x_μ. Ausgehend von der Wahrscheinlichkeitsfunktion $\mathsf{P}_{\mu_1\mu_2\ldots\mu_n}$ dieser zufälligen Größen können wir nach den Regeln des § 2 die Wahrscheinlichkeitsfunktion $\mathsf{P}_{\mu_{i_1}\mu_{i_2}\ldots\mu_{i_k}}$ jedes Teilsystems $(x_{\mu_{i_1}}, x_{\mu_{i_2}}, \ldots, x_{\mu_{i_k}})$ bestimmen. Die Gleichungen (2) und (3) haben zur Folge, daß diese nach dem § 2 definierte Wahrscheinlichkeitsfunktion mit der a priori gegebenen Funktion $\mathsf{P}_{\mu_{i_1}\mu_{i_2}\ldots\mu_{i_k}}$ zusammenfällt. Wir setzen jetzt voraus, daß eine Zylindermenge A durch

$$A = p^{-1}_{\mu_{i_1}\mu_{i_2}\ldots\mu_{i_k}}(A')$$

und gleichzeitig durch

$$A = p^{-1}_{\mu_{j_1}\mu_{j_2}\ldots\mu_{j_m}}(A'')$$

definiert ist, wobei alle zufälligen Größen x_{μ_i} und x_{μ_j} zum System $(x_{\mu_1}, x_{\mu_2}, \ldots, x_{\mu_n})$ gehören, was offenbar keine wesentliche Einschränkung ist. Die Bedingungen

$$\left(x_{\mu_{i_1}}, x_{\mu_{i_2}}, \ldots, x_{\mu_{i_k}}\right) \subset A'$$

und

$$\left(x_{\mu_{j_1}}, x_{\mu_{j_2}}, \ldots, x_{\mu_{j_m}}\right) \subset A''$$

sind gleichbedeutend. Es gilt folglich

$$\mathsf{P}_{\mu_{i_1}\mu_{i_2}\ldots\mu_{i_k}}(A') = \mathsf{P}_{\mu_1\mu_2\ldots\mu_n}\left\{\left(x_{\mu_{i_1}}, x_{\mu_{i_2}}, \ldots, x_{\mu_{i_k}}\right) \subset A'\right\}$$
$$= \mathsf{P}_{\mu_1\mu_2\ldots\mu_n}\left\{\left(x_{\mu_{j_1}}, x_{\mu_{j_2}}, \ldots, x_{\mu_{j_m}}\right) \subset A''\right\} = \mathsf{P}_{\mu_{j_1}\mu_{j_2}\ldots\mu_{j_m}}(A''),$$

was unsere Behauptung bezüglich der Eindeutigkeit der Definition von $\mathsf{P}(A)$ beweist.

Wir wollen jetzt beweisen, daß das Wahrscheinlichkeitsfeld $(\mathfrak{F}^M, \mathsf{P})$ allen Axiomen I—VI genügt. Das Axiom I behauptet nur, daß $\mathfrak{F}^M$ ein Körper sein soll, diese Tatsache wurde schon oben bewiesen. Es gilt weiter bei einem beliebigen μ

$$E = p_\mu^{-1}(R^1),$$
$$\mathsf{P}(E) = \mathsf{P}_\mu(R^1) = 1,$$

was die Gültigkeit der Axiome II und IV beweist. Es folgt endlich unmittelbar aus der Definition von $\mathsf{P}(A)$, daß $\mathsf{P}(A)$ nicht negativ ist (Axiom III).

Nur wenig komplizierter ist der Beweis der Gültigkeit des Axioms V. Zu diesem Zwecke betrachten wir zwei Zylindermengen

$$A = p^{-1}_{\mu_{i_1}\mu_{i_2}\ldots\mu_{i_k}}(A')$$

und

$$B = p^{-1}_{\mu_{j_1}\mu_{j_2}\ldots\mu_{j_m}}(B').$$

Dabei setzen wir voraus, daß alle Größen x_{μ_i} und x_{μ_j} zu einem umfassenden endlichen System $(x_{\mu_1}, x_{\mu_2}, \ldots, x_{\mu_n})$ gehören. Wenn die Mengen A und B disjunkt sind, so sind die Relationen

$$\left(x_{\mu_{i_1}}, x_{\mu_{i_2}}, \ldots, x_{\mu_{i_k}}\right) \subset A'$$

und

$$\left(x_{\mu_{j_1}}, x_{\mu_{j_2}}, \ldots, x_{\mu_{j_k}}\right) \subset B'$$

unvereinbar. Es gilt folglich

$$\begin{aligned}\mathsf{P}(A+B) &= \mathsf{P}_{\mu_1\mu_2\ldots\mu_n}\left\{\left(x_{\mu_{i_1}}, x_{\mu_{i_2}}, \ldots, x_{\mu_{i_k}}\right) \subset A' \right. \\ &\qquad \left. \textit{oder} \quad \left(x_{\mu_{j_1}}, x_{\mu_{j_2}}, \ldots, x_{\mu_{j_m}}\right) \subset B'\right\} \\ &= \mathsf{P}_{\mu_1\mu_2\ldots\mu_n}\left\{\left(x_{\mu_{i_1}}, x_{\mu_{i_2}}, \ldots, x_{\mu_{i_k}}\right) \subset A'\right\} \\ &+ \mathsf{P}_{\mu_1\mu_2\ldots\mu_n}\left\{\left(x_{\mu_{j_1}}, x_{\mu_{j_2}}, \ldots, x_{\mu_{j_m}}\right) \subset B'\right\} = \mathsf{P}(A) + \mathsf{P}(B).\end{aligned}$$

Damit ist unser Beweis zu Ende geführt.

Es bleibt das Axiom VI. Es sei also

$$A_1 \supset A_2 \supset \cdots \supset A_n \supset \cdots$$

eine abnehmende Folge von Zylindermengen mit

$$\lim \mathsf{P}(A_n) = L > 0;$$

wir wollen beweisen, daß der Durchschnitt aller Mengen A_n nicht leer ist. Man kann, ohne die Fragestellung wesentlich einzuschränken, voraussetzen, daß in der Definition der ersten n Zylindermengen A_k nur die n ersten Koordinaten x_{μ_k} einer Folge

$$x_{\mu_1}, x_{\mu_2}, \ldots, x_{\mu_n}, \ldots$$

vorkommen, d. h. daß

$$A_n = p^{-1}_{\mu_1\mu_2\ldots\mu_n}(B_n)$$

ist. Wir setzen zur Abkürzung

$$\mathsf{P}_{\mu_1\mu_2\ldots\mu_n}(B) = \mathsf{P}_n(B);$$

dann ist offenbar

$$\mathsf{P}_n(B_n) = \mathsf{P}(A_n) \geqq L > 0.$$

Man kann in jeder Menge B_n eine abgeschlossene beschränkte Menge U_n finden mit

$$\mathsf{P}_n(B_n - U_n) \leqq \frac{\varepsilon}{2^n}.$$

Aus dieser Ungleichung folgt für die Menge

$$V_n = p^{-1}_{\mu_1\mu_2\ldots\mu_n}(U_n)$$

die Ungleichung

$$\mathsf{P}(A_n - V_n) \leqq \frac{\varepsilon}{2^n}. \tag{4}$$

Es sei weiter

$$W_n = V_1 V_2 \ldots V_n;$$

aus (4) folgt

$$\mathsf{P}(A_n - W_n) \leqq \varepsilon$$

und, da $W_n \subset V_n \subset A_n$ ist,

$$\mathsf{P}(W_n) \geqq \mathsf{P}(A_n) - \varepsilon \geqq L - \varepsilon.$$

Ist ε klein genug, so ist also $\mathsf{P}(W_n) > 0$ und W_n nicht leer. Wir wählen jetzt in jeder Menge W_n einen Punkt $\xi^{(n)}$ mit den Koordinaten $x^{(n)}_\mu$.

Jeder Punkt $\xi^{(n+p)}$, $p = 0, 1, 2, \ldots$, gehört zur Menge V_n, es gilt folglich

$$\left(x_{\mu_1}^{(n+p)}, x_{\mu_2}^{(n+p)}, \ldots, x_{\mu_n}^{(n+p)}\right) = p_{\mu_1 \mu_2 \ldots \mu_n}^{-1}(\xi^{(n+p)}) \subset U_n .$$

Da die Mengen U_n beschränkt sind, kann man (Diagonalverfahren!) aus der Folge $\{\xi^{(n)}\}$ eine Teilfolge.

$$\xi^{(n_1)}, \; \xi^{(n_2)}, \; \ldots, \; \xi^{(n_i)}, \; \ldots$$

wählen, für welche die entsprechenden Koordinaten $x_{\mu_k}^{(n_i)}$ bei jedem k gegen den bestimmten Limites x_k konvergieren. Es sei endlich ξ ein Punkt von E mit den Koordinaten

$$x_{\mu_k} = x_k ,$$
$$x_\mu = 0 , \qquad \mu \neq \mu_k . \qquad\qquad k = 1, 2, 3, \ldots$$

Der Punkt $(x_1, x_2, \ldots, x_k)$ gehört als Limes der Folge $(x_1^{(n_i)}, x_2^{(n_i)}, \ldots, x_k^{(n_i)})$, $i = 1, 2, 3, \ldots$, zur Menge U_k. Folglich gehört ξ zur

$$A_k \subset V_k = p_{\mu_1 \mu_2 \ldots \mu_k}^{-1}(U_k)$$

bei jedem k, also auch zur Durchschnittsmenge

$$A = \mathfrak{D}_k A_k .$$

§ 5. Äquivalente zufällige Größen, verschiedene Arten der Konvergenz.

Von diesem Paragraphen an betrachten wir ausschließlich BOREL*sche Wahrscheinlichkeitsfelder.* Dies bildet, wie schon im § 2, zweites Kapitel, erklärt wurde, keine wesentliche Einschränkung unserer Betrachtungen.

Zwei zufällige Größen x und y heißen *äquivalent*, wenn die Wahrscheinlichkeit der Relation $x \neq y$ gleich Null ist. Offenbar haben zwei äquivalente zufällige Größen dieselbe Wahrscheinlichkeitsfunktion:

$$\mathsf{P}^{(x)}(A) = \mathsf{P}^{(y)}(A) .$$

Die Verteilungsfunktionen $F^{(x)}$ und $F^{(y)}$ fallen folglich auch zusammen. In mehreren Fragen der Wahrscheinlichkeitsrechnung kann eine zufällige Größe durch jede äquivalente Größe ersetzt werden.

Es sei jetzt

(1) $$x_1, x_2, \ldots, x_n, \ldots$$

eine Folge von zufälligen Größen. Wir wollen die Menge A aller elementaren Ereignisse ξ betrachten, für welche die Folge (1) konvergiert. Wenn man mit $A_{np}^{(m)}$ die Menge aller ξ bezeichnet, in welchen alle Ungleichungen

$$|x_{n+k} - x_n| < \frac{1}{m} \qquad\qquad k = 1, 2, \ldots, p$$

gelten, so ergibt sich unmittelbar, daß

(2) $$A = \mathfrak{D}_m \mathfrak{S}_n \mathfrak{D}_p A_{np}^{(m)}$$

ist. Nach § 3 gehört die Menge $A_{np}^{(m)}$ immer zum Mengenkörper $\mathfrak{F}$. Die Relation (2) zeigt uns, daß auch die Konvergenzmenge A zu $\mathfrak{F}$ gehört. *Es hat also immer einen ganz bestimmten Sinn, über die Wahrscheinlichkeit der Konvergenz einer Folge zufälliger Größen zu sprechen.*

Es sei jetzt die Wahrscheinlichkeit $\mathsf{P}(A)$ der Konvergenzmenge A gleich *Eins*. Dann behaupten wir, daß die Folge (1) mit der Wahrscheinlichkeit Eins gegen eine zufällige Größe x konvergiert, dabei ist die zufällige Größe x bis auf Äquivalenz eindeutig bestimmt. Um eine solche zufällige Größe zu bestimmen, setzen wir auf A

$$x = \lim x_n \qquad n \to \infty$$

und $x = 0$ außerhalb A. Es bleibt uns zu beweisen, daß x eine zufällige Größe ist, d. h. daß die Menge $A(a)$ der Elemente ξ, für welche $x < a$ ist, zum Körper $\mathfrak{F}$ gehört. Es gilt aber

$$A(a) = A \mathop{\mathfrak{S}}_{n} \mathop{\mathfrak{D}}_{p} \{x_{n+p} < a\},$$

wenn $a \leqq 0$ ist, und

$$A(a) = A \mathop{\mathfrak{S}}_{n} \mathop{\mathfrak{D}}_{p} \{x_{n+p} < a\} + \bar{A}$$

im entgegengesetzten Falle, woraus unsere Behauptung unmittelbar folgt.

Wenn die Wahrscheinlichkeit der Konvergenz der Folge (1) gegen x gleich Eins ist, so sagen wir, daß (1) *fast sicher* gegen x konvergiert.

Doch ist für die Wahrscheinlichkeitsrechnung ein anderer Konvergenzbegriff vielleicht noch wichtiger.

Definition. Eine Folge $x_1, x_2, \ldots, x_n, \ldots$ von zufälligen Größen *konvergiert nach Wahrscheinlichkeit* (convèrge en probabilité) gegen eine zufällige Größe x, wenn bei jedem $\varepsilon > 0$ die Wahrscheinlichkeit

$$\mathsf{P}\{|x_n - x| > \varepsilon\}$$

mit $n \to \infty$ gegen Null konvergiert[1].

I. *Wenn die Folge* (1) *gegen x und gleichzeitig gegen x' konvergiert, so sind x und x' äquivalent.* Es gilt in der Tat

$$\mathsf{P}\left\{|x - x'| > \frac{1}{m}\right\} \leqq \mathsf{P}\left\{|x_n - x| > \frac{1}{2m}\right\} + \mathsf{P}\left\{|x_n - x'| > \frac{1}{2m}\right\};$$

da die letzten Wahrscheinlichkeiten bei einem genügend großen n beliebig klein sind, folgt es daraus, daß

$$\mathsf{P}\left\{|x - x'| > \frac{1}{m}\right\} = 0$$

ist; jetzt erhält man schon unmittelbar

$$\mathsf{P}\{x \neq x'\} \leqq \sum_m \mathsf{P}\left\{|x - x'| > \frac{1}{m}\right\} = 0.$$

[1] Dieser Begriff stammt im wesentlichen von BERNOULLI her, wurde jedoch in der vollen Allgemeinheit von SLUTSKY eingeführt (vgl. SLUTSKY [1]).

II. *Konvergiert die Folge* (1) *fast sicher gegen* x, *so konvergiert sie gegen* x *auch nach Wahrscheinlichkeit.* Es sei A die Konvergenzmenge der Folge (1); dann ist

$$1 = \mathsf{P}(A) \leqq \lim_{n\to\infty} \mathsf{P}\{|x_{n+p} - x| < \varepsilon, p = 0, 1, 2, \ldots\} \leqq \lim_{n\to\infty} \mathsf{P}\{|x_n - x| < \varepsilon\},$$

woraus die Konvergenz nach Wahrscheinlichkeit folgt.

III. *Für die Konvergenz nach Wahrscheinlichkeit der Folge* (1) *ist die folgende Bedingung notwendig und hinreichend: Bei jedem* $\varepsilon > 0$ *existiert ein solches* n, *daß für jedes* $p > 0$ *die Ungleichung*

$$\mathsf{P}\{|x_{n+p} - x_n| > \varepsilon\} < \varepsilon$$

gilt.

Es seien jetzt $F_1(a), F_2(a), \ldots, F_n(a), \ldots, F(a)$ die Verteilungsfunktionen der zufälligen Größen $x_1, x_2, \ldots, x_n, \ldots, x$. Wenn die Folge x_n nach Wahrscheinlichkeit gegen x konvergiert, so ist die Verteilungsfunktion $F(a)$ eindeutig durch die Kenntnis der Funktionen $F_n(a)$ bestimmt.

Es gilt in der Tat:

Satz. *Konvergiert die Folge* $x_1, x_2, \ldots, x_n, \ldots$ *nach Wahrscheinlichkeit gegen* x, *so konvergiert die Folge der entsprechenden Verteilungsfunktionen* $F_n(a)$ *in jedem Stetigkeitspunkte von* $F(a)$ *gegen die Verteilungsfunktion* $F(a)$ *von* x.

Daß $F(a)$ durch die $F_n(a)$ wirklich bestimmt ist, folgt daraus, daß $F(a)$ als eine nach links stetige monotone Funktion durch ihre Werte in den Stetigkeitspunkten eindeutig festgelegt ist[1]. Zum Beweise des Satzes setzen wir voraus, daß a ein Stetigkeitspunkt von F ist. Es sei $a' < a$, dann ist im Falle $x < a'$, $x_n \geqq a$ notwendig $|x_n - x| > a - a'$. Es gilt folglich

$$\lim \mathsf{P}(x < a', x_n \geqq a) = 0,$$

$$F(a') = \mathsf{P}(x < a') \leqq \mathsf{P}(x_n < a) + \mathsf{P}(x < a', x_n \geqq a) = F_n(a) + \mathsf{P}(x < a', x_n \geqq a),$$

$$F(a') \leqq \lim\inf F_n(a) + \lim \mathsf{P}(x < a', x_n \geqq a),$$

$$F(a') \leqq \lim\inf F_n(a). \tag{3}$$

Analog beweist man, daß aus $a'' > a$ die Relation

$$F(a'') \geqq \lim\sup F_n(a) \tag{4}$$

folgt. Da $F(a')$ und $F(a'')$ mit $a' \to a$ und $a'' \to a$ gegen $F(a)$ konvergieren, folgt aus (3) und (4), daß

$$\lim F_n(a) = F(a)$$

ist, womit unser Satz bewiesen ist.

[1] Sie hat in der Tat nur eine höchstens abzählbare Menge von Unstetigkeitspunkten (vgl. z. B. Lebesgue: Leçons sur l'intégration 1928 S. 50). Die Stetigkeitspunkte liegen infolgedessen überall dicht, und man bestimmt den Wert der Funktion $F(a)$ in einem Unstetigkeitspunkt als Limes ihrer Werte in den nach links liegenden Stetigkeitspunkten.

Viertes Kapitel.

Mathematische Erwartungen[1].

§ 1. Abstrakte LEBESGUEsche Integrale.

Es sei x eine zufällige Größe und A eine Menge aus $\mathfrak{F}$. Wir bilden dann für ein positives m die Summe

$$(1) \qquad S_\lambda = \sum_{k=-\infty}^{k=+\infty} k\lambda \, \mathsf{P}\{k\lambda \leqq x < (k+1)\lambda, \ \xi \subset A\}.$$

Wenn diese Reihe bei jedem λ absolut konvergiert, so strebt S_λ mit $\lambda \to 0$ gegen einen bestimmten Limes, welcher *nach Definition* das Integral

$$(2) \qquad \int_A x \, \mathsf{P}(dE)$$

ist. In dieser abstrakten Form wurde der Integralbegriff von M. FRÉCHET eingeführt[2], sie ist insbesondere für die Wahrscheinlichkeitsrechnung unentbehrlich (der Leser wird übrigens im folgenden Paragraphen sehen, daß die übliche Definition der bedingten mathematischen Erwartung der Größe x unter der Hypothese A bis auf einen konstanten Faktor mit der Definition des Integrals (2) zusammenfällt).

Wir geben hier ein kurzes Verzeichnis der wichtigsten Eigenschaften der Integrale von der Form (2). Der Leser findet ihre Beweise in jedem Lehrbuch über reelle Funktionen, obwohl sie meistens nur in dem Falle, daß $\mathsf{P}(A)$ das LEBESGUEsche Maß der Mengen im R^n ist, durchgeführt sind. Diese Beweise auf den allgemeinen Fall zu übertragen, ist eigentlich keine neue mathematische Aufgabe: sie bleiben meistens wörtlich dieselben.

I. Ist die zufällige Größe x auf A integrierbar, so ist sie auch auf jeder Teilmenge A' von A, welche zu $\mathfrak{F}$ gehört, integrierbar.

II. Ist x auf A integrierbar und A in höchstens abzählbar viele disjunkte Mengen A_n aus $\mathfrak{F}$ zerlegt, so gilt

$$\int_A x \, \mathsf{P}(dE) = \sum_n \int_{A_n} x \, \mathsf{P}(dE).$$

III. Mit x ist immer auch $|x|$ integrierbar, und es gilt dabei

$$\left| \int_A x \, \mathsf{P}(dE) \right| \leqq \int_A |x| \, \mathsf{P}(dE).$$

---

[1] Wie es im § 5, drittes Kapitel, erwähnt wurde, betrachten wir in diesem und in den folgenden Kapiteln nur BOREL*sche Wahrscheinlichkeitsfelder*.

[2] FRÉCHET: Sur l'intégrale d'une fonctionnelle étendue à un ensemble abstrait. Bull. Soc. Math. France Bd. 43 (1915) S. 248.

IV. Wenn in jedem Falle ξ die Ungleichungen $0 \leqq y \leqq x$ erfüllt sind, so ist mit x auch y integrierbar[1] und es gilt dabei

$$\int_A y \,\mathsf{P}(dE) \leqq \int_A x \,\mathsf{P}(dE)\,.$$

V. Wenn $m \leqq x \leqq M$ ist, wobei m und M zwei Konstanten sind, so ist

$$m\,\mathsf{P}(A) \leqq \int_A x \,\mathsf{P}(dE) \leqq M\,\mathsf{P}(A)\,.$$

VI. Sind x und y integrierbar und K und L zwei reelle Konstanten, so ist auch $Kx + Ly$ integrierbar, und es gilt dabei

$$\int_A (Kx + Ly)\,\mathsf{P}(dE) = K\int_A x\,\mathsf{P}(dE) + L\int_A y\,\mathsf{P}(dE)\,.$$

VII. Ist die Reihe

$$\sum_n \int_A |x_n|\,\mathsf{P}(dE)$$

konvergent, so konvergiert die Reihe

$$\sum_n x_n = x$$

in jedem Punkt der Menge A bis auf eine Menge B mit $\mathsf{P}(B) = 0$. Setzt man außerhalb $A - B$ überall $x = 0$, so ist

$$\int_A x\,\mathsf{P}(dE) = \sum_n \int_A x_n\,\mathsf{P}(dE)\,.$$

VIII. Sind x und y äquivalent $(\mathsf{P}\{x \neq y\} = 0)$, so ist für jede Menge A aus $\mathfrak{F}$

$$\int_A x\,\mathsf{P}(dE) = \int_A y\,\mathsf{P}(dE)\,. \tag{3}$$

IX. Gilt (3) für jede Menge A aus $\mathfrak{F}$, so sind x und y äquivalent.

Der obigen Definition des Integrals entnimmt man noch die folgende Eigenschaft, welche man in der gewöhnlichen LEBESGUEschen Theorie nicht hat:

X. Es seien $\mathsf{P}_1(A)$ und $\mathsf{P}_2(A)$ zwei Wahrscheinlichkeitsfunktionen, welche auf demselben $\mathfrak{F}$ definiert sind, $\mathsf{P}(A) = \mathsf{P}_1(A) + \mathsf{P}_2(A)$ und x in bezug auf $\mathsf{P}_1(A)$ und $\mathsf{P}_2(A)$ auf A integrierbar. Dann gilt

$$\int_A x\,\mathsf{P}(dE) = \int_A x\,\mathsf{P}_1(dE) + \int_A x\,\mathsf{P}_2(dE).$$

XI. Jede beschränkte zufällige Größe ist integrierbar.

---

[1] Man setzt dabei voraus, daß y eine zufällige Größe ist, d. h. in der Terminologie der allgemeinen Integrationstheorie in bezug auf $\mathfrak{F}$ meßbar ist.

§ 2. Absolute und bedingte mathematische Erwartungen.

Es sei x eine zufällige Größe. Das Integral

$$\mathsf{E}(x) = \int_E x\,\mathsf{P}(dE)$$

nennt man in der Wahrscheinlichkeitsrechnung *die mathematische Erwartung* der Größe x. Es folgt aus den Eigenschaften III, IV, V, VI, VII, VIII, XI, daß

I. $|\mathsf{E}(x)| \leqq \mathsf{E}(|x|)$;

II. $\mathsf{E}(y) \leqq \mathsf{E}(x)$, wenn immer $0 \leqq y \leqq x$ ist;

III. $\inf(x) \leqq \mathsf{E}(x) \leqq \sup(x)$;

IV. $\mathsf{E}(Kx + Ly) = K\mathsf{E}(x) + L\mathsf{E}(y)$;

V. $\mathsf{E}\left(\sum_n x_n\right) = \sum_n \mathsf{E}(x_n)$, wenn die Reihe $\sum_n \mathsf{E}(|x_n|)$ konvergiert.

VI. Sind x und y äquivalent, so ist

$$\mathsf{E}(x) = \mathsf{E}(y).$$

VII. Jede beschränkte zufällige Größe besitzt eine bestimmte mathematische Erwartung.

Man hat nach der Definition des Integrals

$$\begin{aligned}\mathsf{E}(x) &= \lim \sum_{k=-\infty}^{k=+\infty} km\,\mathsf{P}\{km \leqq x < (k+1)\,m\} \\ &= \lim \sum_{k=-\infty}^{k=+\infty} km\{F((k+1)\,m) - F(km)\}.\end{aligned}$$

Die zweite Zeile ist nichts anderes als die übliche Definition des Stieltjesschen Integrals

$$\int_{-\infty}^{+\infty} a\,dF^{(x)}(a) = \mathsf{E}(x). \tag{1}$$

Die Formel (1) könnte folglich auch als Definition der Erwartung $\mathsf{E}(x)$ dienen.

Es sei jetzt u eine Funktion des elementaren Ereignisses ξ und x eine zufällige Größe, welche als eine eindeutige Funktion $x = x(u)$ von u definiert ist. Dann ist

$$\mathsf{P}\{km \leqq x < (k+1)\,m\} = \mathsf{P}^{(u)}\{km \leqq x(u) < (k+1)\,m\},$$

wobei $\mathsf{P}^{(u)}(A)$ die Wahrscheinlichkeitsfunktion von u ist. Es folgt mithin aus der Definition des Integrals, daß

$$\int_E x\,\mathsf{P}(dE) = \int_{E^{(u)}} x\,\mathsf{P}^{(u)}(dE^{(u)})$$

und folglich

$$\mathsf{E}(x) = \int\limits_{E^{(u)}} x(u)\, \mathsf{P}^{(u)}(dE^{(u)}) \tag{2}$$

ist, wobei $E^{(u)}$ die Menge aller möglichen Werte von u bedeutet.

Wenn insbesondere u selbst eine zufällige Größe ist, so hat man

$$\mathsf{E}(x) = \int\limits_{E} x\, \mathsf{P}(dE) = \int\limits_{R^1} x(u)\, \mathsf{P}^{(u)}(dR^1) = \int\limits_{-\infty}^{+\infty} x(a)\, dF^{(u)}(a)\,. \tag{3}$$

Das letzte Integral in der Formel (2) ist im Falle einer stetigen Funktion $x(u)$ ein gewöhnliches STIELTJESsches Integral. Es sei aber dabei bemerkt, daß das Integral

$$\int\limits_{-\infty}^{+\infty} x(a)\, dF^{(u)}(a)$$

auch im Falle der Nichtexistenz der mathematischen Erwartung $\mathsf{E}(x)$ existieren kann: für die Existenz von $\mathsf{E}(x)$ ist die Endlichkeit des Integrals

$$\int\limits_{-\infty}^{+\infty} |x(a)|\, dF^{(u)}(a)$$

notwendig und hinreichend[1].

Ist u ein Punkt $(u_1, u_2, \ldots, u_n)$ des R^n, so ist wegen (2)

$$\mathsf{E}(x) = \iint\limits_{R^n} \cdots \int x(u_1, u_2, \ldots, u_n)\, \mathsf{P}^{(u_1, u_2, \ldots, u_n)}(dR^n)\,. \tag{4}$$

Wir haben schon gesehen, daß die bedingte Wahrscheinlichkeit $\mathsf{P}_B(A)$ alle Eigenschaften einer Wahrscheinlichkeitsfunktion besitzt. Das entsprechende Integral

$$\mathsf{E}_B(x) = \int\limits_{E} x\, \mathsf{P}_B(dE) \tag{5}$$

nennen wir *die bedingte mathematische Erwartung der zufälligen Größe x in bezug auf das Ereignis B*. Da

$$\mathsf{P}_B(\bar{B}) = 0\,, \qquad \int\limits_{\bar{B}} x\, \mathsf{P}_B(dE) = 0$$

ist, folgt aus (5) die Gleichung

$$\mathsf{E}_B(x) = \int\limits_{E} x\, \mathsf{P}_B(dE) = \int\limits_{B} x\, \mathsf{P}_B(dE) + \int\limits_{\bar{B}} x\, \mathsf{P}_B(dE) = \int\limits_{B} x\, \mathsf{P}_B(dE)\,.$$

Erinnert man sich, daß im Falle $A \subset B$

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

---

[1] Vgl. V. GLIVENKO: Sur les valeurs probables de fonctions. Rend. Accad. Lincei Bd. 8 (1928) S. 480—483.

ist, so erhält man

(6) $$\mathsf{E}_B(x) = \frac{1}{\mathsf{P}(B)}\int\limits_B x\,\mathsf{P}(dE),$$

(7) $$\int\limits_B x\,\mathsf{P}(dE) = \mathsf{P}(B)\,\mathsf{E}_B(x).$$

Aus (6) und der Gleichung

$$\int\limits_{A+B} x\,\mathsf{P}(dE) = \int\limits_A x\,\mathsf{P}(dE) + \int\limits_B x\,\mathsf{P}(dE)$$

folgt schließlich

(8) $$\mathsf{E}_{A+B}(x) = \frac{\mathsf{P}(A)\,\mathsf{E}_A(x) + \mathsf{P}(B)\,\mathsf{E}_B(x)}{\mathsf{P}(A+B)}.$$

Es gilt insbesondere die Formel

(9) $$\mathsf{E}(x) = \mathsf{P}(A)\,\mathsf{E}_A(x) + \mathsf{P}(\bar{A})\,\mathsf{E}_{\bar{A}}(x).$$

§ 3. Die TCHEBYCHEFFsche Ungleichung.

Es sei $f(x)$ eine nichtnegative Funktion eines reellen Argumentes x, welche bei $x \geqq a$ immer nicht kleiner als $b > 0$ bleibt. Dann ist für jede zufällige Größe x

(1) $$\mathsf{P}(x \geqq a) \leqq \frac{\mathsf{E}\{f(x)\}}{b},$$

wenn nur die mathematische Erwartung $\mathsf{E}\{f(x)\}$ existiert. Man hat in der Tat

$$\mathsf{E}\{f(x)\} = \int\limits_E f(x)\,\mathsf{P}(dE) \geqq \int\limits_{\{x \geqq a\}} f(x)\,\mathsf{P}(dE) \geqq b\,\mathsf{P}(x \geqq a),$$

woraus (1) unmittelbar folgt.

Es gilt z. B. für jedes positive e

(2) $$\mathsf{P}(x \geqq a) \leqq \frac{\mathsf{E}(e^{ex})}{e^{ea}}.$$

Es sei jetzt $f(x)$ nichtnegativ, gerade und bei positivem x nicht abnehmend. Dann gilt für jede zufällige Größe x und bei jeder Wahl der Konstante $a > 0$ die Ungleichung

(3) $$\mathsf{P}(|x| \geqq a) \leqq \frac{\mathsf{E}\{f(x)\}}{f(a)}.$$

Insbesondere gilt

(4) $$\mathsf{P}(|x - \mathsf{E}(x)| \geqq a) \leqq \frac{\mathsf{E}f\{x - \mathsf{E}(x)\}}{f(a)}.$$

Besonders wichtig ist der Fall $f(x) = x^2$. In diesem Fall erhält man aus (3) und (4)

(5) $$\mathsf{P}(|x| \geqq a) \leqq \frac{\mathsf{E}(x^2)}{a^2},$$

(6) $$\mathsf{P}(|x - \mathsf{E}(x)| \geqq a) \leqq \frac{\mathsf{E}\{x - \mathsf{E}(x)\}^2}{a^2} = \frac{\sigma^2(x)}{a^2}.$$

Man nennt dabei

$$\sigma^2(x) = \mathsf{E}\{x - \mathsf{E}(x)\}^2$$

die Streuung der Größe x. Es läßt sich leicht berechnen, daß

$$\sigma^2(x) = \mathsf{E}(x^2) - \{\mathsf{E}(x)\}^2$$

ist.

Ist $f(x)$ beschränkt:

$$|f(x)| \leqq K,$$

so kann man $\mathsf{P}(|x| \geqq a)$ auch nach unten abschätzen. Es gilt in der Tat

$$\mathsf{E}(f(x)) = \int_E f(x)\,\mathsf{P}(dE) = \int_{\{|x|<a\}} f(x)\,\mathsf{P}(d\mathsf{E}) + \int_{\{|x|\geqq a\}} f(x)\,\mathsf{P}(dE)$$

$$\leqq f(a)\,\mathsf{P}(|x| < a) + K\,\mathsf{P}(|x| \geqq a) \leqq f(a) + K\,\mathsf{P}(|x| \geqq a)$$

und folglich

$$\mathsf{P}(|x| \geqq a) \geqq \frac{\mathsf{E}\{f(x)\} - f(a)}{K}. \tag{7}$$

Ist statt $f(x)$ die zufällige Größe x selbst beschränkt:

$$|x| \leqq M,$$

so ist $f(x) \leqq f(M)$, und man erhält statt (7) die Formel

$$\mathsf{P}(|x| \geqq a) \geqq \frac{\mathsf{E}(f(x)) - f(a)}{f(M)}. \tag{8}$$

Im Fall $f(x) = x^2$ ergibt sich aus (8)

$$\mathsf{P}(|x| \geqq a) \geqq \frac{\mathsf{E}(x^2) - a^2}{M^2}. \tag{9}$$

§ 4. Einige Konvergenzkriterien.

Es sei

$$x_1, x_2, \ldots, x_n, \ldots \tag{1}$$

eine Folge von zufälligen Größen und $f(x)$ eine nichtnegative, gerade und bei positiven x monoton wachsende Funktion[1]. Dann gelten die folgenden Sätze:

I. Für die Konvergenz nach Wahrscheinlichkeit der Folge (1) ist die folgende Bedingung hinreichend: Zu jedem $\varepsilon > 0$ existiert ein solches n, daß für jedes $p > 0$ die Ungleichung

$$\mathsf{E}\{f(x_{n+p} - x_n)\} < \varepsilon \tag{2}$$

gilt.

II. Für die Konvergenz nach Wahrscheinlichkeit der Folge (1) gegen die zufällige Größe x ist die Bedingung

$$\lim_{n \to +\infty} \mathsf{E}\{f(x_n - x)\} = 0 \tag{3}$$

hinreichend.

[1] Es ist also $f(x) > 0$, falls $x \neq 0$ ist.

III. Ist $f(x)$ beschränkt und stetig und $f(0) = 0$, so sind die Bedingungen I und II auch notwendig.

IV. Ist $f(x)$ stetig, $f(0) = 0$ und alle $x_1, x_2, \ldots, x_n, \ldots, x$ in ihrer Gesamtheit beschränkt, so sind die Bedingungen I und II auch notwendig.

Aus II und IV folgt insbesondere

V. Für die Konvergenz nach Wahrscheinlichkeit der Folge (1) gegen x ist die Bedingung

$$\lim \mathsf{E}(x_n - x)^2 = 0 \tag{4}$$

hinreichend. Sind dabei $x_1, x_2, \ldots, x_n, \ldots, x$ in ihrer Gesamtheit beschränkt, so ist diese Bedingung auch notwendig.

Zum Beweis von I—IV vgl. SLUTSKY [1] und FRÉCHET [1]. Diese Sätze folgen übrigens fast unmittelbar aus den Formeln (3) und (8) des vorigen Paragraphen.

§ 5. Differentiation und Integration der mathematischen Erwartungen nach einem Parameter.

Ist jedem elementaren Ereignis ξ eine bestimmte reelle Funktion $x(t)$ eines reellen Argumentes t zugeordnet, so sagen wir, daß $x(t)$ eine *zufällige Funktion* ist, wenn bei jedem festen t die Größe $x(t)$ eine zufällige Größe ist. Es entsteht dann die Frage, unter welchen Bedingungen das Erwartungszeichen E mit den Integrations- und Differentiationszeichen vertauschbar ist. Zwei folgende Sätze, ohne die Frage zu erschöpfen, können in mehreren einfachen Fällen eine befriedigende Antwort auf diese Frage geben:

Satz I. Ist die mathematische Erwartung $\mathsf{E}[x(t)]$ für jedes t endlich, $x(t)$ immer für jedes t differentiierbar und die Ableitung $x'(t)$ von $x(t)$ nach t dem Betrage nach immer kleiner als eine feste Konstante M, so ist

$$\frac{d}{dt}\mathsf{E}(x(t)) = \mathsf{E}(x'(t)).$$

Satz II. Ist $x(t)$ dem Betrage nach immer kleiner als eine feste Konstante K und nach RIEMANN integrierbar, so ist

$$\int_a^b \mathsf{E}(x(t))\,dt = \mathsf{E}\left[\int_a^b x(t)\,dt\right],$$

wenn nur $\mathsf{E}(x(t))$ im RIEMANNschen Sinne integrierbar ist.

Beweis des Satzes I. Es ist zuerst zu bemerken, daß $x'(t)$ als Limes der zufälligen Größen

$$\frac{x(t+h) - x(t)}{h} \qquad h = 1, \frac{1}{2}, \ldots, \frac{1}{n}, \ldots$$

eine zufällige Größe ist. Da $x'(t)$ beschränkt ist, existiert die mathematische Erwartung $\mathsf{E}(x'(t))$ (Eigenschaft VII der mathematischen Erwartungen aus § 2). Wir wählen jetzt ein festes t und bezeichnen mit A das Ereignis

$$\left|\frac{x(t+h)-x(t)}{h} - x'(t)\right| > \varepsilon .$$

Die Wahrscheinlichkeit $\mathsf{P}(A)$ konvergiert mit $h \to 0$ gegen Null für jedes $\varepsilon > 0$. Da immer

$$\left|\frac{x(t+h)-x(t)}{h}\right| \leqq M, \qquad |x'(t)| \leqq M$$

ist und außerdem im Falle $\bar{A}$

$$\left|\frac{x(t+h)-x(t)}{h} - x'(t)\right| \leqq \varepsilon$$

gilt, so hat man

$$\left|\frac{\mathsf{E}x(t+h)-\mathsf{E}x(t)}{h} - \mathsf{E}x'(t)\right| \leqq \mathsf{E}\left|\frac{x(t+h)-x(t)}{h} - x'(t)\right|$$

$$= \mathsf{P}(A)\,\mathsf{E}_A\left|\frac{x(t+h)-x(t)}{h} - x'(t)\right| + \mathsf{P}(\bar{A})\,\mathsf{E}_{\bar{A}}\left|\frac{x(t+h)-x(t)}{h} - x'(t)\right|$$

$$\leqq 2M\,\mathsf{P}(A) + \varepsilon .$$

$\varepsilon > 0$ kann man dabei beliebig wählen, und $\mathsf{P}(A)$ ist bei jedem genügend kleinen h beliebig klein. Es gilt folglich

$$\frac{d}{dt}\mathsf{E}x(t) = \lim_{h\to 0}\frac{\mathsf{E}x(t+h)-\mathsf{E}x(t)}{h} = \mathsf{E}x'(t),$$

w. z. b. w.

Beweis des Satzes II. Es sei

$$S_n = \frac{1}{h}\sum_{k=1}^{k=h} x(t+kh), \qquad h = \frac{b-a}{n}.$$

Da S_n gegen $J = \int_a^b x(t)\,dt$ konvergiert, so kann man für jedes $\varepsilon > 0$ ein solches N wählen, daß aus $n \geqq N$ die Ungleichung

$$\mathsf{P}(A) = \mathsf{P}\{|S_k - J| > \varepsilon\} < \varepsilon$$

folgt. Setzt man

$$S_n^* = \frac{1}{h}\sum_{k=1}^{k=h} \mathsf{E}x(t+kh) = \mathsf{E}(S_n),$$

so wird

$$|S_n^* - \mathsf{E}(J)| = |\mathsf{E}(S_n - J)| \leqq \mathsf{E}|S_n - J|$$

$$= \mathsf{P}(A)\,\mathsf{E}_A|S_n - J| + \mathsf{P}(\bar{A})\,\mathsf{E}_{\bar{A}}|S_n - J| \leqq 2K\,\mathsf{P}(A) + \varepsilon \leqq (2K+1)\,\varepsilon .$$

Somit konvergiert S_n^* gegen $\mathsf{E}(J)$, woraus die Gleichung

$$\int_a^b \mathsf{E}x(t)\,dt = \lim S_n^* = \mathsf{E}(J)$$

folgt.

Der Satz II kann ohne jede neue Schwierigkeit für die zwei- und mehrfachen Integrale erweitert werden. Wir wollen eine Anwendung dieses Satzes auf ein Problem der geometrischen Wahrscheinlichkeiten geben. Es sei G ein quadrierbares Gebiet der Ebene, dessen Gestalt vom Zufall abhängt, d. h. es sei jedem Elementarereignis ξ eines Wahrscheinlichkeitsfeldes ein bestimmtes quadrierbares Gebiet G der Ebene zugeordnet. Wir bezeichnen mit J den Inhalt des Gebietes G und mit $\mathsf{P}(x, y)$ die Wahrscheinlichkeit, daß der Punkt (x, y) zum Gebiet G gehört. Dann ist

$$\mathsf{E}(J) = \iint \mathsf{P}(x, y)\, dx\, dy .$$

Um dies zu beweisen, genügt es, zu bemerken, daß

$$J = \iint f(x, y)\, dx\, dy ,$$
$$\mathsf{P}(x, y) = \mathsf{E} f(x, y)$$

ist, wobei $f(x, y)$ die charakteristische Funktion des Gebietes G ist ($f(x, y) = 1$ auf G und $f(x, y)$ außerhalb G)[1].

Fünftes Kapitel.

Bedingte Wahrscheinlichkeiten und Erwartungen.

§ 1. Bedingte Wahrscheinlichkeiten.

Wir haben im ersten Kapitel, § 6 die bedingte Wahrscheinlichkeit $\mathsf{P}_{\mathfrak{A}}(B)$ eines Ereignisses B nach einem Versuch $\mathfrak{A}$ definiert. Es wurde dabei im ersten Kapitel vorausgesetzt, daß $\mathfrak{A}$ nur endlichviele verschiedene mögliche Ausgänge hat. Man kann aber $\mathsf{P}_{\mathfrak{A}}(B)$ auch im Falle eines Versuches $\mathfrak{A}$ mit einer unendlichen Menge der möglichen Ausgänge, d. h. einer Zerlegung der Menge E in unendlichviele disjunkte Teilmengen, definieren. Eine solche Zerlegung erhält man insbesondere, wenn man eine beliebige Funktion u von ξ betrachtet und die Mengen $u = \text{const}$ als Elemente der Zerlegung $\mathfrak{A}_u$ definiert. Die bedingte Wahrscheinlichkeit $\mathsf{P}_{\mathfrak{A}_u}(B)$ wird auch mit $\mathsf{P}_u(B)$ bezeichnet. Eine beliebige Zerlegung A der Menge E erhält man als die Zerlegung $\mathfrak{A}_u$, welche durch eine Funktion u von ξ „induziert" ist, wenn man jedem ξ diejenige Menge der Zerlegung $\mathfrak{A}$ als $u(\xi)$ zuordnet, welche ξ enthält.

Zwei Funktionen u und u' von ξ bestimmen dann und nur dann dieselbe Zerlegung $\mathfrak{A}_u = \mathfrak{A}_{u'}$ der Menge E, wenn eine solche eineindeutige Zuordnung $u' = f(u)$ ihrer Wertebereiche $\mathfrak{F}^{(u)}$ und $\mathfrak{F}^{(u')}$ existiert, daß identisch $u'(\xi) = f u(\xi)$ ist. Der Leser kann leicht nachprüfen,

[1] Vgl. A. Kolmogoroff u. M. A. Leontowitsch: Physik. Z. Sowjetunion. Im Druck (1933).

daß die im folgenden definierten zufälligen Größen $\mathsf{P}_u(B)$ und $\mathsf{P}_{u'}(B)$ in diesem Falle zusammenfallen, sie sind also wirklich durch die Zerlegung $\mathfrak{A}_u = \mathfrak{A}_{u'}$ selbst bestimmt.

Man kann zur Definition von $\mathsf{P}_u(B)$ die folgende Gleichung brauchen:

$$\mathsf{P}_{\{u \subset A\}}(B) = \mathsf{E}_{\{u \subset A\}} \mathsf{P}_u(B)\,. \tag{1}$$

Man könnte leicht zeigen, daß im Falle einer endlichen Menge $\mathsf{E}^{(u)}$ der möglichen Werte von u die Gleichung (1) bei jeder Wahl der Menge A erfüllt ist (wobei $\mathsf{P}_u(B)$ nach § 6, erstes Kapitel definiert ist). Im allgemeinen Falle (in welchem $\mathsf{P}_u(B)$ noch nicht definiert ist) beweisen wir, daß immer eine und bis auf Äquivalenz nur eine zufällige Größe $\mathsf{P}_u(B)$ existiert, welche als Funktion von u sich bestimmen läßt und bei jeder Wahl von A mit $\mathsf{P}^{(u)}(A) > 0$ aus $\mathfrak{F}^{(u)}$ der Gleichung (1) genügt. *Die somit bis auf Äquivalenz bestimmte Funktion* $\mathsf{P}_u(B)$ *von* u *nennen wir die bedingte Wahrscheinlichkeit von* B *in bezug auf* u (oder bei bekanntem u). Den Wert von $\mathsf{P}_u(B)$ bei $u = a$ bezeichnen wir dabei mit $\mathsf{P}_u(a; B)$.

Beweis der Existenz und Eindeutigkeit von $\mathsf{P}_u(B)$. Multipliziert man (1) mit $\mathsf{P}\{u \subset A\} = \mathsf{P}^{(u)}(A)$, so erhält man links

$$\mathsf{P}\{u \subset A\}\,\mathsf{P}_{u \subset A}(B) = \mathsf{P}(B\{u \subset A\}) = \mathsf{P}(B\,u^{-1}(A))$$

und rechts

$$\mathsf{P}\{u \subset A\}\,\mathsf{E}_{\{u \subset A\}}\,\mathsf{P}_u(B) = \int\limits_{\{u \subset A\}} \mathsf{P}_u(B)\,\mathsf{P}(dE) = \int\limits_{A} \mathsf{P}_u(B)\,\mathsf{P}^{(u)}(dE^{(u)})\,.$$

Es ergibt sich also die Formel

$$\mathsf{P}(B\,u^{-1}(A)) = \int\limits_{A} \mathsf{P}_u(B)\,\mathsf{P}^{(u)}(dE^{(u)})\,. \tag{2}$$

Umgekehrt folgt aus (2) die Formel (1). Im Falle $\mathsf{P}^{(u)}(A) = 0$, in welchem (1) keinen Sinn hat, ist (2) in trivialer Weise richtig. Die Forderung (2) ist also mit (1) äquivalent.

Nach der Eigenschaft IX des Integrals (viertes Kap., § 1) ist eine zufällige Größe x durch die Werte des Integrals

$$\int\limits_{A} x\,\mathsf{P}(dE)$$

für alle Mengen aus $\mathfrak{F}$ bis auf Äquivalenz eindeutig bestimmt. Da $\mathsf{P}_u(B)$ eine auf dem Wahrscheinlichkeitsfelde $(\mathfrak{F}^{(u)}, \mathsf{P}^{(u)})$ bestimmte zufällige Größe ist, so folgt daraus, daß die Formel (2) diese Größe $\mathsf{P}_u(B)$ bis auf Äquivalenz eindeutig bestimmt.

Es bleibt uns die Existenz von $\mathsf{P}_u(B)$ zu beweisen. Zu diesem Zwecke brauchen wir den folgenden Satz von O. Nikodym[1]:

[1] Nikodym, O.: Fundam. Math. Bd. 15 (1930) S. 168 (Théorème III).

Es sei $\mathfrak{F}$ ein BORELscher Mengenkörper, $\mathsf{P}(A)$ eine auf $\mathfrak{F}$ definierte nichtnegative vollständig additive Mengenfunktion (in der wahrscheinlichkeitstheoretischen Terminologie — eine zufällige Größe auf $(\mathfrak{F}, \mathsf{P})$) und $Q(A)$ eine zweite auf $\mathfrak{F}$ definierte vollständig additive Mengenfunktion, wobei aus $Q(A) \neq 0$ die Ungleichung $\mathsf{P}(A) > 0$ folgt. Es existiert dann eine in bezug auf $\mathfrak{F}$ meßbare Funktion $f(\xi)$ (in der wahrscheinlichkeitstheoretischen Terminologie — eine zufällige Größe), welche für jede Menge A aus $\mathfrak{F}$ der Gleichung

$$Q(A) = \int_A f(\xi)\, \mathsf{P}(dE)$$

genügt.

Um diesen Satz in unserem Falle anwenden zu können, bleibt es zu zeigen: 1° daß

$$Q(A) = \mathsf{P}(B\, u^{-1}(A))$$

auf $\mathfrak{F}^{(u)}$ vollständig additiv ist, 2° daß aus $Q(A) \neq 0$ die Ungleichung $\mathsf{P}^{(u)}(A) > 0$ folgt.

Zunächst folgt 2° aus

$$0 \leqq \mathsf{P}(B\, u^{-1}(A)) \leqq \mathsf{P}(u^{-1}(A)) = \mathsf{P}^{(u)}(A).$$

Um 1° zu beweisen, setzen wir

$$A = \sum_n A_n.$$

Dann ist

$$u^{-1}(A) = \sum_n u^{-1}(A_n)$$

und

$$B\, u^{-1}(A) = \sum_n B\, u^{-1}(A_n).$$

Da P vollständig additiv ist, so folgt daraus, daß

$$\mathsf{P}(B\, u^{-1}(A_n)) = \sum_n \mathsf{P}(B\, u^{-1}(A_n))$$

ist, w. z. b. w.

Aus der Gleichung (1) folgt insbesondere (wenn man $A = E^{(u)}$ setzt) die wichtige Formel

$$\mathsf{P}(B) = \mathsf{E}(\mathsf{P}_u(B)). \tag{3}$$

Wir wollen jetzt zum Beweise der zwei folgenden fundamentalen Eigenschaften von bedingten Wahrscheinlichkeiten übergehen:

Satz I. Es ist fast sicher, daß

$$0 \leqq \mathsf{P}_u(B) \leqq 1 \tag{4}$$

ist.

Satz II. Ist B in die höchstens abzählbar vielen Summen den B_n zerlegt:

$$B = \sum_n B_n,$$

so gilt fast sicher die Gleichung

(5) $$\mathsf{P}_u(B) = \sum_n \mathsf{P}_u(B_n)\,.$$

Diese zwei Eigenschaften von $\mathsf{P}_u(B)$ entsprechen den zwei charakteristischen Eigenschaften der Wahrscheinlichkeitsfunktion $\mathsf{P}(B)$: es ist immer $0 \leqq \mathsf{P}(B) \leqq 1$, und $\mathsf{P}(B)$ ist vollständig additiv. Sie gestatten, mehrere weitere wesentliche Eigenschaften der absoluten Wahrscheinlichkeiten $\mathsf{P}(B)$ auf die bedingten Wahrscheinlichkeiten $\mathsf{P}_u(B)$ zu übertragen. Man soll dabei jedoch nicht vergessen, daß $\mathsf{P}_u(B)$ bei einer festen Menge B eine nur bis auf Äquivalenz bestimmte zufällige Größe ist.

Beweis des Satzes I. Setzen wir voraus — im Gegensatz zu der zu beweisenden Behauptung —, daß auf einer Menge $M \subset E^{(u)}$ mit $\mathsf{P}^{(u)}(M) > 0$ die Ungleichung $\mathsf{P}_u(B) \geqq 1 + \varepsilon$, $\varepsilon > 0$, erfüllt ist, so ist nach der Formel (1)

$$\mathsf{P}_{\{u \subset M\}}(B) = \mathsf{E}_{\{u \subset M\}}\,\mathsf{P}_u(B) \geqq 1 + \varepsilon\,,$$

was offenbar unmöglich ist. Ebenso beweist man, daß fast sicher $\mathsf{P}_u(B) \geqq 0$ ist.

Beweis des Satzes II. Aus der Konvergenz der Reihe

$$\sum_n \mathsf{E}\,|\mathsf{P}_u(B_n)| = \sum_n \mathsf{E}\,(\mathsf{P}_u(B_n)) = \sum_n \mathsf{P}(B_n) = \mathsf{P}(B)$$

folgt wegen der Eigenschaft V der mathematischen Erwartungen (viertes Kap., § 2), daß die Reihe

$$\sum_n \mathsf{P}_u(B_n)$$

fast sicher konvergiert. Da die Reihe

$$\sum_n \mathsf{E}_{\{u \subset A\}}|\mathsf{P}_u(B_n)| = \sum_n \mathsf{E}_{\{u \subset A\}}(\mathsf{P}_u(B_n)) = \sum_n \mathsf{P}_{\{u \subset A\}}(B_n) = \mathsf{P}_{\{u \subset A\}}(B)$$

bei jeder Wahl der Menge A mit $\mathsf{P}^{(u)}(A) > 0$ konvergiert, so folgt aus der erwähnten Eigenschaft V der mathematischen Erwartungen, daß für jede Menge A der erwähnten Art die Relation

$$\mathsf{E}_{\{u \subset A\}}\Big(\sum_n \mathsf{P}_u(B_n)\Big) = \sum_n \mathsf{E}_{\{u \subset A\}}(\mathsf{P}_u(B_n)) = \mathsf{P}_{\{u \subset A\}}(B) = \mathsf{E}_{\{u \subset A\}}(\mathsf{P}_u(B_n))$$

gilt; daraus folgt die Gleichung (5) unmittelbar.

Zum Schluß dieses Paragraphen wollen wir auf zwei Spezialfälle hinweisen. Ist erstens $u(\xi) = c$ eine Konstante, so gilt fast sicher $\mathsf{P}_c(A) = \mathsf{P}(A)$. Setzt man dagegen $u(\xi) = \xi$, so erhält man sogleich, daß $\mathsf{P}_\xi(A)$ auf A fast sicher gleich Eins und auf $\bar{A}$ fast sicher gleich Null ist. $\mathsf{P}_\xi(A)$ ist also *die charakteristische Funktion* der Menge A.

§ 2. Erklärung eines Borelschen Paradoxons.

Es sei die Menge aller Punkte einer Kugelfläche zur Grundmenge E gewählt. Als $\mathfrak{F}$ wählen wir die Gesamtheit aller Borelschen Mengen der Kugelflächen. Endlich sei $\mathsf{P}(A)$ dem Inhalt der Menge A proportional. Wir wählen jetzt zwei diametrale Punkte zu Polen, dann ist jeder Meridiankreis durch die entsprechende geographische Länge ψ, $0 \leqq \psi < \pi$, eindeutig bestimmt. Da ψ nur von 0 bis π läuft, wir also *volle* Meridiankreise (nicht Halbkreise) betrachten, muß auch die Breite Θ von $-\pi$ bis $+\pi$ $\left(\text{nicht von } -\frac{\pi}{2} \text{ bis } +\frac{\pi}{2}\right)$ laufen. Die von Borel gestellte Aufgabe ist: die „bedingte Wahrscheinlichkeitsverteilung" für die Breite Θ, $-\pi \leqq \Theta < +\pi$, bei bekannter Länge ψ zu bestimmen. Man berechnet leicht, daß

$$\mathsf{P}_\psi\{\Theta_1 \leqq \Theta < \Theta_2\} = \tfrac{1}{4}\int_{\Theta_1}^{\Theta_2} |\cos\Theta|\, d\Theta$$

ist. Die Wahrscheinlichkeitsverteilung für Θ bei bekanntem ψ ist also nicht homogen.

Setzt man voraus, daß die bedingte Wahrscheinlichkeitsverteilung für Θ „bei der Hypothese, daß ξ auf dem gegebenen Meridiankreis liegt", homogen sein soll, so erhält man einen Widerspruch.

Dieser Umstand zeigt, daß der Begriff der bedingten Wahrscheinlichkeit in bezug auf eine isoliert gegebene Hypothese, deren Wahrscheinlichkeit gleich Null ist, unzulässig ist: nur dann erhält man auf einem Meridiankreis eine Wahrscheinlichkeitsverteilung für Θ, wenn dieser Meridiankreis als Element der Zerlegung der ganzen Kugelfläche in Meridiankreise mit den gegebenen Polen betrachtet wird.

§ 3. Bedingte Wahrscheinlichkeiten in bezug auf eine zufällige Größe.

Ist x eine zufällige Größe und $\mathsf{P}_x(B)$ als Funktion von x im Borelschen Sinne meßbar, so kann man $\mathsf{P}_x(B)$ auch elementar definieren. Man könnte in der Tat der Formel (2), § 1 die folgende Gestalt geben:

$$\mathsf{P}(B)\,\mathsf{P}_B^{(x)}(A) = \int_A \mathsf{P}_x(B)\,\mathsf{P}^{(x)}(dE)\,. \tag{1}$$

In unserem Falle ergibt sich unmittelbar aus (1), daß

$$\mathsf{P}(B)\,F_B^{(x)}(a) = \int_{-\infty}^{a} \mathsf{P}_x(a;B)\,dF^{(x)}(a) \tag{2}$$

ist. Nach einem Satz von Lebesgue[1] folgt aus (2), daß

$$\mathsf{P}_x(a;B) = \mathsf{P}(B)\lim \frac{F_B^{(x)}(a+h) - F_B^{(x)}(a)}{F^{(x)}(a+h) - F^{(x)}(a)} \qquad h \to 0 \tag{3}$$

ist, bis auf eine Menge H der Punkte a mit $\mathsf{P}^{(x)}(H) = 0$.

[1] Lebesgue: Leçons sur l'intégration, S. 301—302. 1928.

$\mathsf{P}_x(a; B)$ wurde im § 1 nur bis auf eine Menge G mit $\mathsf{P}^{(x)}(G) = 0$ definiert. Betrachtet man jetzt die Formel (3) als Definition von $\mathsf{P}_x(a; B)$, wobei im Falle der Nichtexistenz des Limes auf der rechten Seite von (3) $\mathsf{P}_x(a; B) = 0$ gesetzt ist, so genügt diese neue Größe allen Forderungen des § 1.

Existieren noch die Wahrscheinlichkeitsdichten $f^{(x)}(a)$ und $f_B^{(x)}(a)$ und ist $f^{(x)}(a) > 0$, so verwandelt sich die Formel (3) in die folgende:

$$\mathsf{P}_x(a; B) = \mathsf{P}(B) \frac{f_B^{(x)}(a)}{f^{(x)}(a)}. \tag{4}$$

Es folgt noch aus der Formel (3), daß die Existenz des Limes (3) und der Wahrscheinlichkeitsdichte $f^{(x)}(a)$ die Existenz von $f_B^{(x)}(a)$ zur Folge hat. Es ist dabei

$$\mathsf{P}(B) f_B^{(x)}(a) \leqq f^{(x)}(a). \tag{5}$$

Ist $\mathsf{P}(B) > 0$, so folgt aus (4) die Gleichung

$$f_B^{(x)}(a) = \frac{\mathsf{P}_x(a; B) f^{(x)}(a)}{\mathsf{P}(B)}. \tag{6}$$

Im Falle $f^{(x)}(a) = 0$ ist nach (5) $f_B^x(a) = 0$, und folglich ist (6) auch richtig. Ist dabei die Verteilung von x stetig, so hat man

$$\mathsf{P}(B) = \mathsf{E}(\mathsf{P}_x(B)) = \int_{-\infty}^{+\infty} \mathsf{P}_x(a; B)\, dF^{(x)}(a) = \int_{-\infty}^{+\infty} \mathsf{P}_x(a; B) f^{(x)}(a)\, da. \tag{7}$$

Aus (6) und (7) ergibt sich

$$f_B^{(x)}(a) = \frac{\mathsf{P}_x(a; B) f^{(x)}(a)}{\int_{-\infty}^{+\infty} \mathsf{P}_x(a; B) f^{(x)}(a)\, da}. \tag{8}$$

In dieser Gleichung haben wir den sog. BAYES*schen Satz für die stetigen Verteilungen* erhalten. Die Voraussetzungen, unter welchen dieser Satz bewiesen ist, sind also die folgenden: $\mathsf{P}_x(B)$ ist im BORELschen Sinne meßbar und im Punkte a nach der Formel (3) bestimmt, die Verteilung von x ist stetig und die Wahrscheinlichkeitsdichte $f^{(x)}(a)$ im Punkte a existiert.

§ 4. Bedingte mathematische Erwartungen.

Es seien u eine beliebige Funktion von ξ und y eine zufällige Größe. Eine zufällige Größe $\mathsf{E}_u(y)$, welche als Funktion von u darstellbar ist und für jede Menge A aus $\mathfrak{F}^{(u)}$ mit $\mathsf{P}^{(u)}(A) > 0$ die Bedingung

$$\mathsf{E}_{\{u \subset A\}}(y) = \mathsf{E}_{\{u \subset A\}} \mathsf{E}_u(y) \tag{1}$$

erfüllt, heißt (falls sie existiert) *bedingte mathematische Erwartung der Größe y bei bekanntem Wert von u.*

Multipliziert man (1) mit $\mathsf{P}^{(u)}(A)$, so erhält man

$$\int\limits_{\{u \subset A\}} y \, \mathsf{P}(dE) = \int\limits_{A} \mathsf{E}_u(y) \, \mathsf{P}^{(u)}(dE^{(u)}) . \tag{2}$$

Umgekehrt folgt aus (2) die Formel (1). Im Falle $\mathsf{P}^{(u)}(A) = 0$, in welchem (1) keinen Sinn hat, ist (2) in trivialer Weise richtig. Ebenso wie im Falle der bedingten Wahrscheinlichkeit (vgl. § 1) beweist man, daß $\mathsf{E}_u(y)$ durch (2) bis auf Äquivalenz eindeutig bestimmt ist.

Der Wert von $\mathsf{E}_u(y)$ bei $u = a$ bezeichnen wir mit $\mathsf{E}_u(a; y)$. Es sei noch bemerkt, daß $\mathsf{E}_u(y)$ ebenso wie $\mathsf{P}_u(y)$ nur von der Zerlegung $\mathfrak{A}_u$ abhängig ist, und kann also als $\mathsf{E}_{\mathfrak{A}_u}(y)$ bezeichnet werden.

In der Definition von $\mathsf{E}_u(y)$ ist schon die Existenz von $\mathsf{E}(y)$ vorausgesetzt (man setze $A = \mathsf{E}^{(u)}$, dann ist $\mathsf{E}_{\{u \subset A\}}(y) = \mathsf{E}(y)$); wir wollen jetzt beweisen, daß *die Existenz von* $\mathsf{E}(y)$ *für die Existenz von* $\mathsf{E}_u(y)$ *auch hinreichend ist.* Dazu genügt es, nach dem Satz von O. Nikodym (vgl. § 1) zu beweisen, daß die Mengenfunktion

$$Q(A) = \int\limits_{\{u \subset A\}} y \, \mathsf{P}(dE)$$

auf $\mathfrak{F}^{(u)}$ vollständig additiv und in bezug auf $\mathsf{P}^{(u)}(A)$ total stetig ist. Die erste Tatsache beweist man wörtlich so wie im Falle der bedingten Wahrscheinlichkeiten (vgl. § 1). Die zweite Forderung der Totalstetigkeit besteht darin, daß aus $Q(A) \neq 0$ die Ungleichung $\mathsf{P}^{(u)}(A) > 0$ folgen soll. Setzt man voraus, daß $\mathsf{P}^{(u)}(A) = \mathsf{P}\{u \subset A\} = 0$ ist, so ist offenbar

$$Q(A) = \int\limits_{\{u \subset A\}} y \, \mathsf{P}(dE) = 0 .$$

Unsere zweite Forderung ist also auch erfüllt.

Setzt man in der Gleichung (1) $A = E^{(u)}$, so ergibt sich die Formel

$$\mathsf{E}(y) = \mathsf{E}\,\mathsf{E}_u(y) . \tag{3}$$

Man beweist weiter, daß fast sicher

$$\mathsf{E}_u(ay + bz) = a\,\mathsf{E}_u(y) + b\,\mathsf{E}_u(z) \tag{4}$$

ist, wobei a und b zwei beliebige Konstanten sind. (Die Durchführung des Beweises kann dem Leser überlassen bleiben.)

Sind u und v zwei Funktionen des Elementarereignisses ξ, so kann das Paar (u, v) auch als eine Funktion von ξ betrachtet werden. Es gilt dann die folgende wichtige Gleichung:

$$\mathsf{E}_u\,\mathsf{E}_{(u,v)}(y) \doteq \mathsf{E}_u(y) . \tag{5}$$

In der Tat ist $\mathsf{E}_u(y)$ durch die Relation

$$\mathsf{E}_{\{u \subset A\}}(y) = \mathsf{E}_{\{u \subset A\}}\,\mathsf{E}_u(y)$$

definiert. Man soll also beweisen, daß $\mathsf{E}_u \mathsf{E}_{(u,v)}(y)$ der Gleichung

$$\mathsf{E}_{\{u \subset A\}}(y) = \mathsf{E}_{\{u \subset A\}} \mathsf{E}_u \mathsf{E}_{(u,v)}(y) \tag{6}$$

genügt. Aus der Definition von $\mathsf{E}_{(u,v)}(y)$ folgt

$$\mathsf{E}_{\{u \subset A\}}(y) = \mathsf{E}_{\{u \subset A\}} \mathsf{E}_{(u,v)}(y). \tag{7}$$

Aus der Definition von $\mathsf{E}_u \mathsf{E}_{(u,v)}(y)$ folgt weiter, daß

$$\mathsf{E}_{\{u \subset A\}} \mathsf{E}_{(u,v)}(y) = \mathsf{E}_{\{u \subset A\}} \mathsf{E}_u \mathsf{E}_{(u,v)}(y) \tag{8}$$

ist. Die Gleichungen (7) und (8) haben die Gleichung (6) zur Folge, womit unsere Behauptung bewiesen ist.

Setzt man $y = \mathsf{P}_\xi(B)$ auf B gleich Eins und außerhalb B gleich Null, so ist

$$\mathsf{E}_u(y) = \mathsf{P}_u(B),$$

$$\mathsf{E}_{(u,v)}(y) = \mathsf{P}_{(u,v)}(B).$$

In diesem Falle ergibt sich aus (5) die Formel

$$\mathsf{E}_u \mathsf{P}_{(u,v)}(B) = \mathsf{P}_{(u)}(B). \tag{9}$$

Man kann die bedingte mathematische Erwartung $\mathsf{E}_u(y)$ auch direkt durch die entsprechenden bedingten Wahrscheinlichkeiten definieren. Man betrachtet zu diesem Zweck die folgenden Summen:

$$S_\lambda(u) = \sum_{k=-\infty}^{k=+\infty} k\lambda \mathsf{P}_u\{k\lambda \leqq y < (k+1)\lambda\} = \sum_k R_k. \tag{10}$$

Ist $\mathsf{E}(y)$ bestimmt, so konvergiert die Reihe (10) fast sicher. Es gilt in der Tat nach der Formel (3) des § 1

$$\mathsf{E}|R_k| = |k\lambda| \mathsf{P}\{k\lambda \leqq y < (k+1)\lambda\},$$

und die Konvergenz der Reihe

$$\sum_{k=-\infty}^{k=+\infty} |k\lambda| \mathsf{P}\{k\lambda \leqq y < (k+1)\lambda\} = \sum_k \mathsf{E}|R_k|$$

ist die notwendige Bedingung der Existenz von $\mathsf{E}(y)$ (vgl. viertes Kap., § 1). Aus dieser Konvergenz folgt, daß die Reihe (10) fast sicher konvergiert (vgl. viertes Kap., § 2, V). Man beweist weiter ebenso wie in der Theorie der LEBESGUEschen Integrale, daß aus der Konvergenz von (10) bei irgendeinem λ die Konvergenz bei jedem λ folgt und daß im Falle der Konvergenz der Reihe (10) $S_\lambda(u)$ mit $\lambda \to 0$ gegen einen bestimmten Limes konvergiert[1]. Man kann dann definitionsgemäß

$$\mathsf{E}_u(y) = \lim_{\lambda \to 0} S_\lambda(u) \tag{11}$$

setzen. Um zu beweisen, daß die durch (11) definierte bedingte Erwartung $\mathsf{E}_u(y)$ den früher aufgestellten Forderungen genügt, braucht man

[1] Dabei betrachten wir nur eine abzählbare Folge von λ-Werten: alle Wahrscheinlichkeiten $\mathsf{P}_u\{k\lambda \leqq y < (k+1)\lambda\}$ sind dann für alle diese Werte von λ fast sicher bestimmt.

nur die Geltung der Gleichung (1) für die nach (11) definierte Größe $\mathsf{E}_u(y)$ zu verifizieren. Dieser Beweis verläuft folgendermaßen:

$$\mathsf{E}_{\{u \subset A\}} \mathsf{E}_u(y) = \lim_{\lambda \to 0} \mathsf{E}_{\{u \subset A\}} S_\lambda(u)$$

$$= \lim_{\lambda \to 0} \sum_{k=-\infty}^{+\infty} k\lambda \, \mathsf{P}_{\{u \subset A\}}\{k\lambda \leqq y < (k+1)\lambda\} = \mathsf{E}_{\{u \subset A\}}(y).$$

Bei dieser Rechnung ist die Vertauschung des Erwartungs- und Limeszeichens berechtigt, da $S_\lambda(u)$ mit $\lambda \to 0$ gleichmäßig gegen $\mathsf{E}_u(y)$ konvergiert (eine einfache Folge der Eigenschaft V der mathematischen Erwartungen aus § 2). Die Vertauschung des Erwartungszeichens mit dem Summenzeichen ist ebenfalls berechtigt, da die Reihe

$$\sum_{k=-\infty}^{k=+\infty} \mathsf{E}_{\{u \subset A\}}\{|k\lambda| \, \mathsf{P}_u[k\lambda \leqq y < (k+1)\lambda]\}$$

$$= \sum_{k=-\infty}^{k=+\infty} |k\lambda| \, \mathsf{P}_{\{u \subset A\}}[k\lambda \leqq y < (k+1)\lambda]$$

konvergent ist (direkte Anwendung der Eigenschaft V der mathematischen Erwartungen!).

Man könnte statt (11)

$$\mathsf{E}_u(y) = \int_E y \, \mathsf{P}_u(dE) \tag{12}$$

schreiben, man darf aber dabei nicht vergessen, daß (12) kein Integral im Sinne des vierten Kapitels, § 1 ist, so daß (12) nur eine symbolische Bezeichnung ist.

Ist x eine zufällige Größe, so nennen wir die Funktion von x und a

$$F_x^{(y)}(a) = \mathsf{P}_x(y < a)$$

die bedingte Verteilungsfunktion von y bei bekanntem x. $F_x^{(y)}(a)$ ist bei jedem a fast sicher bestimmt. Ist $a < b$, so ist fast sicher

$$F_x^{(y)}(a) \leqq F_x^{(y)}(b).$$

Aus (11) und (10) folgt, daß fast sicher

$$\mathsf{E}_x(y) = \lim_{\lambda \to 0} \sum_{k=-\infty}^{k=+\infty} k\lambda [F_x^{(y)}((k+1)\lambda) - F_x^{(y)}(k\lambda)] \tag{13}$$

ist[1]. Diese Tatsache kann man symbolisch durch die Formel

$$\mathsf{E}_x(y) = \int_{-\infty}^{+\infty} a \, dF_x^{(y)}(a) \tag{14}$$

ausdrücken.

---

[1] Vgl. Fußnote 1 auf Seite 48.

Mit Hilfe der neuen Definition (10—11) der mathematischen Erwartung beweist man noch leicht, daß für eine reelle Funktion von u

$$\mathsf{E}_u[f(u)\, y] = f(u)\, \mathsf{E}_u(y) \tag{15}$$

ist.

Sechstes Kapitel.

Unabhängigkeit. Gesetz der großen Zahlen.

§ 1. Unabhängigkeit.

Definition 1. Zwei Funktionen u und v von ξ sind gegenseitig unabhängig, wenn für je zwei Mengen A aus $\mathfrak{F}^{(u)}$ und B aus $\mathfrak{F}^{(v)}$ die folgende Gleichung gilt:

$$\mathsf{P}(u \subset A,\, v \subset B) = \mathsf{P}(u \subset A)\, \mathsf{P}(v \subset B) = \mathsf{P}^{(u)}(A)\, \mathsf{P}^{(v)}(B)\,. \tag{1}$$

Bestehen die Mengen $E^{(u)}$ und $E^{(v)}$ nur aus endlichvielen Elementen

$$E^{(u)} = u_1 + u_2 + \cdots + u_n\,,$$
$$E^{(v)} = v_1 + v_2 + \cdots + v_m\,,$$

so fällt unsere Definition der Unabhängigkeit von u und v mit der Definition der Unabhängigkeit der Zerlegungen

$$E = \sum_k \{u = u_k\}\,,$$
$$E = \sum_k \{v = v_k\}$$

nach dem ersten Kapitel, § 5 zusammen.

Für die Unabhängigkeit von u und v ist die folgende Bedingung *notwendig und hinreichend:* Bei jeder Wahl der Menge A aus $\mathfrak{F}^{(u)}$ ist die Gleichung

$$\mathsf{P}_v(u \subset A) = \mathsf{P}(u \subset A) \tag{2}$$

fast sicher erfüllt.

Im Falle $\mathsf{P}^{(v)}(B) = 0$ sind in der Tat beide Gleichungen (1) und (2) erfüllt, es genügt folglich ihre Äquivalenz im Falle $\mathsf{P}^{(v)}(B) > 0$ zu beweisen. In diesem Falle ist (1) mit

$$\mathsf{P}_{\{v \subset B\}}(u \subset A) = \mathsf{P}(u \subset A) \tag{3}$$

und folglich mit

$$E_{\{v \subset B\}} \mathsf{P}_v(u \subset A) = \mathsf{P}(u \subset A) \tag{4}$$

äquivalent. Man sieht andererseits, daß aus (2) die Gleichung (4) folgt. Umgekehrt, da $\mathsf{P}_v(u \subset A)$ durch (4) bis auf die Wahrscheinlichkeit Null eindeutig bestimmt ist, folgt aus (4) die Gleichung (2) mit Fast-Sicherheit.

Definition 2. Es sei M eine Menge von Funktionen $u_\mu(\xi)$ von ξ, diese Funktionen heißen in ihrer Gesamtheit gegenseitig unabhängig, wenn die folgende Bedingung erfüllt ist: Es seien M' und M'' zwei disjunkte Untermengen von M, A' (bzw. A'') eine Menge aus $\mathfrak{F}$, welche durch eine Relation zwischen u_μ aus M' (bzw. M'') bestimmt ist; dann gilt die Gleichung
$$\mathsf{P}(A'A'') = \mathsf{P}(A')\,\mathsf{P}(A'').$$

Man könnte die Gesamtheit aller u_μ aus M' (bzw. aus M'') als Koordinaten einer Funktion u' (bzw. u'') betrachten. Die Definition 2 fordert nur die Unabhängigkeit von u' und u'' im Sinne der Definition 1 bei jeder Wahl teilfremder Mengen M' und M''.

Sind $u_1, u_2, \ldots, u_n$ gegenseitig unabhängig, so ist immer

$$(1)\qquad \left\{\begin{aligned} &\mathsf{P}\{u_1 \subset A_1, u_2 \subset A_2, \ldots, u_n \subset A_n\} \\ &= \mathsf{P}(u_1 \subset A_1)\,\mathsf{P}(u_2 \subset A_2)\ldots\mathsf{P}(u_n \subset A_n), \end{aligned}\right.$$

wenn nur die Mengen A_k zu den entsprechenden $\mathfrak{F}^{(u_k)}$ gehören (Beweis durch Induktion!). Diese Gleichung ist aber im allgemeinen keineswegs für die gegenseitige Unabhängigkeit von $u_1, u_2, \ldots, u_n$ hinreichend.

Man verallgemeinert (1) ohne Schwierigkeiten auf die *abzählbaren* Produkte.

Aus der gegenseitigen Unabhängigkeit von u_{μ_k} in jeder endlichen Gruppe $(u_{\mu_1}, u_{\mu_2}, \ldots, u_{\mu_k})$ folgt im allgemeinen noch nicht, daß alle u_μ gegenseitig unabhängig sind.

Man bemerkt endlich leicht, daß die gegenseitige Unabhängigkeit der Funktionen u_μ eigentlich eine Eigenschaft der entsprechenden Zerlegungen $\mathfrak{A}_{u_\mu}$ ist. Sind ferner die u'_μ eindeutige Funktionen der entsprechenden u_μ, so folgt aus der gegenseitigen Unabhängigkeit der u_μ diejenige von u'_μ.

§ 2. Unabhängige zufällige Größen.

Sind $x_1, x_2, \ldots, x_n$ gegenseitig unabhängige zufällige Größen, so folgt aus der Gleichung (2) des vorigen Paragraphen insbesondere die Formel

$$(1)\qquad F^{(x_1, x_2, \ldots, x_n)}(a_1, a_2, \ldots, a_n) = F^{(x_1)}(a_1)\,F^{(x_2)}(a_2)\ldots F^{(x_n)}(a_n).$$

Wenn dabei der Mengenkörper $\mathfrak{F}^{(x_1, x_2, \ldots, x_n)}$ nur aus Borel*schen Mengen des R^n besteht, so ist die Bedingung (1) für die gegenseitige Unabhängigkeit der Größen $x_1, x_2, \ldots, x_n$ auch hinreichend.*

Beweis. Es seien $x' = (x_{i_1}, x_{i_2}, \ldots, x_{i_k})$ und $x'' = (x_{j_1}, x_{j_2}, \ldots, x_{j_m})$ zwei disjunkte Teilsysteme der Größen $x_1, x_2, \ldots, x_n$. Man soll auf Grund der Formel (1) beweisen, daß für je zwei Borelsche Mengen A' und A'' des R^k (bzw. R^m) die Gleichung

$$(2)\qquad \mathsf{P}(x' \subset A', x'' \subset A'') = \mathsf{P}(x' \subset A')\,\mathsf{P}(x'' \subset A'')$$

erfüllt ist. Das folgt aus (1) unmittelbar für die Mengen von der Form

$$A' = \{x_{i_1} < a_1,\ x_{i_2} < a_2,\ \ldots,\ x_{i_k} < a_k\},$$
$$A'' = \{x_{j_1} < b_1,\ x_{j_2} < b_2,\ \ldots,\ x_{j_m} < b_m\}.$$

Man beweist weiter, daß diese Eigenschaft der Mengen A' und A'' bei der Differenz- und Summenbildung erhalten bleibt, woraus (2) für alle BORELschen Mengen folgt.

Es sei jetzt $x = \{x_\mu\}$ eine beliebige (im allgemeinen unendliche) Gesamtheit zufälliger Größen. *Ist der Mengenkörper* $\mathfrak{F}^{(x)}$ *mit dem Körper* $B\mathfrak{F}^M$ (M ist dabei die Menge aller μ) *identisch, so ist die Gesamtheit der Gleichungen*

$$F_{\mu_1 \mu_2 \ldots \mu_n}(a_1, a_2, \ldots, a_n) = F_{\mu_1}(a_1)\, F_{\mu_2}(a_2) \ldots F_{\mu_n}(a_n) \tag{3}$$

für die gegenseitige Unabhängigkeit der Größen x_μ *notwendig und hinreichend.*

Die Notwendigkeit dieser Bedingung folgt unmittelbar aus der Formel (1). Wir wollen beweisen, daß sie auch hinreichend ist. Es seien M' und M'' zwei disjunkte Untermengen der Menge M aller Indizes μ, A' (bzw. A'') eine Menge aus $B\mathfrak{F}^M$, welche durch eine Relation zwischen x_μ mit den Indizen μ aus M' (bzw. aus M'') bestimmt ist; man soll beweisen, daß dann die Gleichung

$$\mathsf{P}(A'A'') = \mathsf{P}(A')\,\mathsf{P}(A'') \tag{4}$$

erfüllt ist. Sind A' und A'' Zylindermengen, so handelt es sich in Wirklichkeit um Relationen zwischen endlichvielen Größen x_μ; die Gleichung (4) ist in diesem Falle eine einfache Folge der früheren Betrachtungen [Formel (2)]. Da aber die Relation (4) bei der Differenz- und Summenbildung der Mengen A' (bzw. A'') erhalten bleibt, so ist (4) auch für alle Mengen aus $B\mathfrak{F}^M$ bewiesen.

Es sei jetzt für jedes μ aus einer Menge M eine Verteilungsfunktion $F_\mu(a)$ a priori gegeben, *man kann dann ein solches Wahrscheinlichkeitsfeld konstruieren, daß in diesem Felde gewisse zufällige Größen* x_μ (*μ durchläuft die ganze Menge M*) *gegenseitig unabhängig sein werden, wobei* x_μ *die a priori gegebene Funktion* $F_\mu(a)$ *als Verteilungsfunktion haben wird.* Um das zu beweisen, genügt es, R^M als die Grundmenge E und $B\mathfrak{F}^M$ als den Körper $\mathfrak{F}$ zu wählen und die Verteilungsfunktionen $F_{\mu_1 \mu_2 \ldots \mu_n}$ (vgl. drittes Kap., § 4) durch die Gleichung (3) definieren.

Es sei noch bemerkt, daß, wie man oben gesehen hat, aus der gegenseitigen Unabhängigkeit jeder endlichen Gruppe von Größen x_μ [Gleichung (3)] die gegenseitige Unabhängigkeit aller x_μ auf $B\mathfrak{F}^M$ folgt. In umfassenderen Wahrscheinlichkeitsfeldern kann diese Eigenschaft verlorengehen.

Zum Schluß dieses Paragraphen geben wir noch einige Unabhängigkeitskriterien für zwei zufällige Größen:

Sind zwei zufällige Größen x und y gegenseitig unabhängig und ist $\mathsf{E}(x)$ ebenso wie $\mathsf{E}(y)$ endlich, so ist fast sicher

$$\left\{\begin{aligned} \mathsf{E}_x(y) &= \mathsf{E}(y), \\ \mathsf{E}_y(x) &= \mathsf{E}(x). \end{aligned}\right. \tag{5}$$

Diese Formeln bilden eine unmittelbare Folge der zweiten Definition der bedingten mathematischen Erwartung [Formeln (10) bis (11) des § 4, fünftes Kap.]. Folglich sind im Falle der Unabhängigkeit die beiden Größen

$$f^2 = \frac{\mathsf{E}[y - \mathsf{E}_x(y)]^2}{\sigma^2(y)}, \qquad g^2 = \frac{\mathsf{E}[x - \mathsf{E}_y(x)]^2}{\sigma^2(x)}$$

gleich *Null* [wenn nur $\sigma^2(x) > 0$ und $\sigma^2(y) > 0$ ist]. Die Zahl f^2 heißt das *Korrelationsverhältnis* von y zu x, g^2 dasjenige von x zu y (PEARSON).

Aus (5) folgt weiter, daß

$$\mathsf{E}(xy) = \mathsf{E}(x)\,\mathsf{E}(y) \tag{6}$$

ist. Zum Beweise wendet man die Formel (15) aus dem fünften Kapitel, § 4 an:

$$\mathsf{E}(xy) = \mathsf{E}\,\mathsf{E}_x(xy) = \mathsf{E}[x\,\mathsf{E}_x(y)] = \mathsf{E}[x\,\mathsf{E}(y)] = \mathsf{E}(y)\,\mathsf{E}(x).$$

Folglich ist im Unabhängigkeitsfalle

$$r = \frac{\mathsf{E}(x, y) - \mathsf{E}(x)\,\mathsf{E}(y)}{\sigma^2(x)\,\sigma^2(y)}$$

auch gleich *Null*; r ist bekanntlich der *Korrelationskoeffizient* von x und y.

Genügen zwei zufällige Größen x und y der Gleichung (6), so heißen sie *unkorreliert*. Für die Summe

$$S = x_1 + x_2 + \cdots + x_n$$

paarweise unkorrelierter Größen $x_1, x_2, \ldots, x_n$ berechnet man leicht, daß

$$\sigma^2(s) = \sigma^2(x_1) + \sigma^2(x_2) + \cdots + \sigma^2(x_n) \tag{7}$$

ist. Die Gleichung (7) gilt also insbesondere für die unabhängigen Größen x_k.

§ 3. Gesetz der großen Zahlen.

Zufällige Größen s einer Folge

$$s_1, s_2, \ldots, s_n, \ldots$$

heißen *stabil*, wenn es eine solche Zahlenfolge

$$d_1, d_2, \ldots, d_n, \ldots$$

gibt, daß für jedes positive ε

$$\mathsf{P}\{|s_n - d_n| \geqq \varepsilon\}$$

mit $n \to \infty$ gegen Null konvergiert. Wenn alle $E(s_n)$ bestimmt sind und man

$$d_n = \mathsf{E}(s_n)$$

wählen kann, so ist die Stabilität *normal*.

Sind alle s_n gleichmäßig beschränkt, so folgt aus

$$\mathsf{P}\{|s_n - d_n| \geqq \varepsilon\} \to 0 \qquad n \to +\infty \tag{1}$$

die Relation

$$|\mathsf{E}(s_n) - d_n| \to 0 \qquad n \to +\infty$$

und folglich

$$\mathsf{P}\{|s_n - \mathsf{E}(s_n)| \geqq \varepsilon\} \to 0. \qquad n \to +\infty \tag{2}$$

Die Stabilität einer beschränkten stabilen Folge ist also notwendig normal.

Es sei

$$\mathsf{E}(s_n - \mathsf{E}(s_n))^2 = \sigma^2(s_n) = \sigma_n^2.$$

Nach der TCHEBYCHEFFschen Ungleichung ist

$$\mathsf{P}\{|s_n - E(s_n)| \geqq \varepsilon\} \leqq \frac{\sigma_n^2}{\varepsilon^2}.$$

Folglich ist die MARKOFF*sche Bedingung*

$$\sigma_n^2 \to 0 \qquad n \to +\infty \tag{3}$$

für die normale Stabilität hinreichend.

Sind $s_n - E(s_n)$ gleichmäßig beschränkt:

$$|s_n - E(s_n)| \leqq M,$$

so ist nach der Ungleichung (9) aus § 3, viertes Kapitel

$$\mathsf{P}\{|s_n - E(s_n)| \geqq \varepsilon\} \geqq \frac{\sigma_n^2 - \varepsilon^2}{M^2}.$$

Folglich ist in diesem Falle die MARKOFFsche Bedingung (3) für die Stabilität von s_n auch notwendig.

Ist

$$s_n = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

und die Größen x_n paarweise unkorreliert, so hat man

$$\sigma_n^2 = \frac{1}{n^2}\{\sigma^2(x_1) + \sigma^2(x_2) + \cdots + \sigma^2(x_n)\}.$$

Folglich ist in diesem Falle für die normale Stabilität der Mittelwerte s_n die Bedingung

$$\sigma_n^2 = \sigma^2(x_1) + \sigma^2(x_2) + \cdots + \sigma^2(x_n) = o(n^2) \tag{4}$$

hinreichend (*Satz von* TCHEBYCHEFF). Insbesondere ist die Bedingung (4) erfüllt, wenn alle Größen x_n gleichmäßig beschränkt sind.

Man kann diesen Satz auf den Fall der schwach korrelierten Größen x_n verallgemeinern: Setzt man voraus, daß der Korrelationskoeffizient r_{mn}* von x_m und x_n der Ungleichung

$$r_{mn} \leqq c(|n-m|)$$

genügt und daß

$$C_n = \sum_{k=0}^{k=n-1} c(k)$$

ist, so ist für die normale Stabilität der Mittelwerte s die Bedingung

$$C_n \sigma_n^2 = o(n^2) \tag{5}$$

hinreichend[1].

Im Falle *unabhängiger* Summanden x_n kann man auch eine notwendige und hinreichende Bedingung für die Stabilität der Mittelwerte s_n geben. Für jedes x_n existiert eine Konstante m_n (Mediane von x_n), die den folgenden Ungleichungen genügt:

$$\mathsf{P}(x_n < m_n) \leqq \tfrac{1}{2},$$
$$\mathsf{P}(x_n > m_n) \leqq \tfrac{1}{2}.$$

Wir setzen nun

$$x_{nk} = x_k, \text{ wenn } |x_k - m_k| \leqq n \text{ ist,}$$
$$x_{nk} = 0 \text{ im entgegengesetzten Falle,}$$
$$s_n^* = \frac{x_{n1} + x_{n2} + \cdots + x_{nn}}{n}.$$

Dann sind die Relationen

$$\sum_{k=1}^{k=n} \mathsf{P}\{|x_n - m_n| > n\} = \sum_{k=1}^{k=n} \mathsf{P}(x_{nk} \neq x_k) \to 0, \qquad n \to +\infty \tag{6}$$

$$\sigma^2(s_n^*) = \sum_{k=1}^{k=n} \sigma^2(x_{nk}) = o(n^2) \tag{7}$$

für die Stabilität der Größen s_n *notwendig und hinreichend*[2].

Man kann dabei die Konstanten d_n gleich den Größen $E(s_n^*)$ nehmen, so daß im Falle

$$\mathsf{E}(s_n^*) - \mathsf{E}(s_n) \to 0 \qquad n \to +\infty$$

(und nur in diesem Falle) die Stabilität *normal* ist.

Man erhält eine weitere Verallgemeinerung des TCHEBYCHEFFschen Satzes, wenn man voraussetzt, daß s_n irgendwie von den Ausgängen irgendwelcher n Versuche

$$\mathfrak{A}_1, \mathfrak{A}_2, \ldots, \mathfrak{A}_n$$

* Es ist offenbar immer $r_{nn} = 1$.

[1] Vgl. A. KHINTCHINE: Sur la loi forte des grandes nombres, C. R. de l'acad. sci. Paris, Bd. 186 (1928) S. 285.

[2] Vgl. A. KOLMOGOROFF: Über die Summen durch den Zufall bestimmter unabhängiger Größen. Math. Ann. Bd. 99 (1928) S. 309—319 — Berichtigungen ebenda Bd. 102 (1929) S. 484—488, Satz VIII (und der entsprechende Zusatz S. 318).

abhängt, so daß nach jedem bestimmten Ausgang aller dieser n Versuche s_n einen bestimmten Wert erhält. Die allgemeine Idee aller Sätze, welche unter dem Namen *des Gesetzes der großen Zahlen* bekannt sind, besteht darin, daß, wenn die Abhängigkeit der Größe s_n bei einem großen n von jedem einzelnen Versuch $\mathfrak{A}_k$ $(k = 1, 2, \ldots, n)$ sehr klein ist, die Größen s_n stabil sind. Betrachtet man

$$\beta_{nk}^2 = \mathsf{E}[\mathsf{E}_{\mathfrak{A}_1 \mathfrak{A}_2 \ldots \mathfrak{A}_k}(s_n) - \mathsf{E}_{\mathfrak{A}_1 \mathfrak{A}_2 \ldots \mathfrak{A}_{k-1}}(s_n)]^2$$

als ein vernünftiges Maß der Abhängigkeit der Größe s_n von dem Versuch $\mathfrak{A}_k$, so läßt sich die obenerwähnte allgemeine Idee des Gesetzes der großen Zahlen durch die folgenden Betrachtungen konkretisieren[1]:

Es sei

$$z_{nk} \doteq \mathsf{E}_{\mathfrak{A}_1 \mathfrak{A}_2 \ldots \mathfrak{A}_k}(s_n) - \mathsf{E}_{\mathfrak{A}_1 \mathfrak{A}_2 \ldots \mathfrak{A}_{k-1}}(s_n) .$$

Dann ist

$$s_n - \mathsf{E}(s_n) = z_1 + z_2 + \cdots + z_n ,$$

$$\mathsf{E}(z_{nk}) = \mathsf{E}\,\mathsf{E}_{\mathfrak{A}_1 \mathfrak{A}_2 \ldots \mathfrak{A}_k}(s_n) - \mathsf{E}\,\mathsf{E}_{\mathfrak{A}_1 \mathfrak{A}_2 \ldots \mathfrak{A}_{k-1}}(s_n) = \mathsf{E}(s_n) - \mathsf{E}(s_n) = 0 ,$$

$$\sigma^2(z_{nk}) = \mathsf{E}(z_{nk}^2) = \beta_{nk}^2 .$$

Man berechnet weiter leicht, daß die zufälligen Größen z_{nk}, $k = 1, 2, \ldots, n$, unkorreliert sind. Es sei in der Tat $i < k$; dann ist[2]

$$\begin{aligned} \mathsf{E}_{\mathfrak{A}_1 \mathfrak{A}_2 \ldots \mathfrak{A}_{k-1}}(z_{ni} z_{nk}) &= z_{ni}\, \mathsf{E}_{\mathfrak{A}_1 \mathfrak{A}_2 \ldots \mathfrak{A}_{k-1}}(z_{nk}) \\ &= z_{ni}\, \mathsf{E}_{\mathfrak{A}_1 \mathfrak{A}_2 \ldots \mathfrak{A}_{k-1}}[\mathsf{E}_{\mathfrak{A}_1 \mathfrak{A}_2 \ldots \mathfrak{A}_k}(s_n) - \mathsf{E}_{\mathfrak{A}_1 \mathfrak{A}_2 \ldots \mathfrak{A}_{k-1}}(s_n)] \\ &= z_{ni}[\mathsf{E}_{\mathfrak{A}_1 \mathfrak{A}_2 \ldots \mathfrak{A}_{k-1}}(s_n) - \mathsf{E}_{\mathfrak{A}_1 \mathfrak{A}_2 \ldots \mathfrak{A}_{k-1}}(s_n)] = 0 \end{aligned}$$

und folglich

$$\mathsf{E}(z_{ni} z_{nk}) = 0 .$$

Wir haben also

$$\sigma^2(s_n) = \sigma^2(z_{n1}) + \sigma^2(z_{n2}) + \cdots + \sigma^2(z_{nn}) = \beta_{n1}^2 + \beta_{n2}^2 + \cdots + \beta_{nn}^2$$

Folglich ist die Bedingung

$$\beta_{n1}^2 + \beta_{n2}^2 + \cdots + \beta_{nn}^2 \to 0 \qquad n \to +\infty$$

für die normale Stabilität der Größen s_n hinreichend.

§ 4. Bemerkungen zum Begriff der mathematischen Erwartung.

Wir haben die mathematische Erwartung einer zufälligen Größe x als

$$\mathsf{E}(x) = \int_E x\, \mathsf{P}(dE) = \int_{-\infty}^{+\infty} a\, dF^{(x)}(a)$$

[1] Vgl. A. Kolmogoroff: Sur la loi des grands nombres. Rend. Accad. Lincei Bd. 9 (1929) S. 470—474.

[2] Anwendung der Formel (15), fünftes Kapitel, § 4.

definiert. Dabei versteht man rechts das Integral im Sinne

$$(1)\qquad \mathsf{E}(x) = \int_{-\infty}^{+\infty} a\, dF^{(x)}(a) = \lim \int_{b}^{c} a\, dF^{(x)}(a)\,. \qquad \begin{matrix} b \to -\infty \\ c \to +\infty \end{matrix}$$

Es liegt die Idee nahe, den Ausdruck

$$(2)\qquad \mathsf{E}^*(x) = \lim \int_{-b}^{+b} a\, dF^{(x)}(a) \qquad b \to +\infty$$

als die *verallgemeinerte* mathematische Erwartung zu betrachten. Man verliert dabei allerdings einige einfache Eigenschaften der mathematischen Erwartungen. Z. B. ist in diesem Falle die Formel

$$\mathsf{E}(x + y) = \mathsf{E}(x) + \mathsf{E}(y)$$

nicht mehr allgemeingültig. In dieser Form ist also die Verallgemeinerung kaum zulässig. Es ist aber bemerkenswert, daß unter einigen einschränkenden Nebenvoraussetzungen die Definition (2) als eine ganz natürliche und brauchbare erscheint.

Man kann nämlich die Frage folgendermaßen stellen: Es sei

$$x_1,\ x_2,\ \ldots,\ x_n,\ \ldots$$

eine Folge von gegenseitig unabhängigen Größen, welche dieselbe Verteilungsfunktion $F^{(x)}(a) = F^{(x_n)}(a)$, $n = 1, 2, \ldots$, haben wie x. Es sei weiter

$$s_n = \frac{x_1 + x_2 + \cdots + x_n}{n}\,.$$

Es wird gefragt, ob eine solche Konstante $E^*(x)$ existiert, daß bei jedem $\varepsilon > 0$

$$(3)\qquad \lim \mathsf{P}(|s_n - \mathsf{E}^*(x)|) > \varepsilon = 0 \qquad n \to +\infty$$

ist. Die Antwort lautet: *Existiert eine solche Konstante* $\mathsf{E}^*(x)$, *so ist sie durch die Formel* (2) *dargestellt.* Die notwendige und hinreichende Bedingung für die Gültigkeit der Formel (3) besteht dabei aus der Existenz des Limes (2) und der Relation

$$(4)\qquad \mathsf{P}(|x| > n) = o\left(\frac{1}{n}\right).$$

Zum Beweise braucht man den Satz, daß die Bedingung (4) für die Stabilität der Mittelwerte s_n notwendig und hinreichend ist, wobei im Falle der Stabilität

$$d_n = \int_{-n}^{+n} a\, dF^{(x)}(a)$$

gesetzt werden kann[1].

[1] Vgl. A. Kolmogoroff: Bemerkungen zu meiner Arbeit „Über die Summen zufälliger Größen". Math. Ann. Bd. 102 (1929) S. 484—488, Satz XII.

Existiert die mathematische Erwartung im früheren Sinne [Formel (1)], so ist die Bedingung (4) immer erfüllt[1]. Da in diesem Falle $\mathsf{E}(x) = \mathsf{E}^*(x)$ ist, so bestimmt die Forderung (3) wirklich eine Verallgemeinerung des Erwartungsbegriffes. Für die *verallgemeinerte mathematische Erwartung* bleiben die Eigenschaften I—VII (viertes Kap., § 2) erhalten, aus der Existenz von $\mathsf{E}^*(x)$ folgt aber die Existenz von $\mathsf{E}^*|x|$ im allgemeinen nicht.

Um zu beweisen, daß der neue Erwartungsbegriff wirklich allgemeiner ist als der frühere, genügt das folgende Beispiel: Man setzt die Wahrscheinlichkeitsdichte $f^{(x)}(a)$ gleich

$$f^{(x)}(a) = \frac{C}{(|a| + 2)^2 \ln(|a| + 2)},$$

wobei die Konstante C durch die Forderung

$$\int_{-\infty}^{+\infty} f^{(x)}(a)\, da = 1$$

bestimmt ist. Man berechnet leicht, daß in diesem Falle die Bedingung (4) erfüllt ist, die Formel (3) den Wert

$$\mathsf{E}(x) = 0$$

ergibt, das Integral

$$\int_{-\infty}^{+\infty} |a|\, dF^{(x)}(a) = \int_{-\infty}^{+\infty} |a|\, f^{(x)}(a)\, da$$

aber divergent ist.

§ 5. Starkes Gesetz der großen Zahlen, Konvergenz von Reihen.

Zufällige Größen s_n der Folge

$$s_1, s_2, \ldots, s_n, \ldots$$

sind *stark stabil*, wenn es eine solche Zahlenfolge

$$d_1, d_2, \ldots, d_n, \ldots$$

gibt, daß die zufälligen Größen

$$s_n - d_n$$

mit $n \to +\infty$ fast sicher gegen Null konvergieren. Aus der starken Stabilität folgt offenbar die gewöhnliche Stabilität. Wenn man

$$d_n = \mathsf{E}(s_n)$$

wählen kann, so ist die starke Stabilität *normal*.

[1] Vgl. A. Kolmogoroff: Bemerkungen zu meiner Arbeit „Über die Summen zufälliger Größen". Math. Ann. Bd. 102 (1929) S. 484—488, Satz XIII.

Im TCHEBYCHEFFschen Falle

$$s = \frac{x_1 + x_2 + \cdots + x_n}{n},$$

wobei die Größen x_n gegenseitig unabhängig sind, ist die Konvergenz der Reihe

$$\sum_{n=1}^{\infty} \frac{\sigma^2(x_n)}{n^2} \tag{1}$$

für die starke normale Stabilität der Mittelwerte s_n hinreichend[1]. Diese Bedingung ist in dem Sinne die beste, daß man für jede Reihe von Konstanten b_n mit

$$\sum_{n=1}^{\infty} \frac{b_n}{n^2} = +\infty$$

eine Reihe von gegenseitig unabhängigen zufälligen Größen x_n konstruieren kann, so daß

$$\sigma^2(x_n) = b_n$$

ist und die entsprechenden Mittelwerte s_n nicht stark stabil sind.

Haben alle x_n dieselbe Verteilungsfunktion $F^{(x)}(a)$, so ist die Existenz der mathematischen Erwartung

$$\mathsf{E}(x) = \int_{-\infty}^{+\infty} a\, dF^{(x)}(a)$$

für die starke Stabilität von s_n notwendig und hinreichend, die Stabilität ist in diesem Falle immer normal[2].

Es seien jetzt wieder

$$x_1, x_2, \ldots, x_n, \ldots$$

beliebige gegenseitig unabhängige zufällige Größen. Dann ist die Wahrscheinlichkeit der Konvergenz der Reihe

$$\sum_{n=1}^{\infty} x_n \tag{1}$$

entweder gleich *Eins* oder *Null*. Diese Wahrscheinlichkeit ist insbesondere gleich Eins, wenn die beiden Reihen

$$\sum_{n=1}^{\infty} \mathsf{E}(x_n) \quad \text{und} \quad \sum_{n=1}^{\infty} \sigma^2(x_n)$$

konvergent sind. Es sei weiter

$$y_n = x_n, \quad \text{im Falle} \quad |x_n| \leqq 1,$$
$$y_n = 0, \quad \text{im Falle} \quad |x_n| > 1$$

[1] Vgl. A. KOLMOGOROFF: Sur la loi forte des grandes nombres. C. R. Acad. Sci., Paris Bd. 191 (1930) S. 910—911.

[2] Der Beweis ist noch nicht publiziert.

gesetzt. Dann ist die gleichzeitige Konvergenz der Reihen

$$\sum_{n=1}^{\infty} \mathsf{P}\{|x_n| > 1\}, \quad \sum_{n=1}^{\infty} \mathsf{E}(y_n) \quad \text{und} \quad \sum_{n=1}^{\infty} \sigma^2(y_n)$$

für die Konvergenz der Reihe (1) mit der Wahrscheinlichkeit Eins notwendig und hinreichend[1].

Anhang.

Null- oder Eins-Gesetz in der Wahrscheinlichkeitsrechnung.

Man hat schon mehrere Fälle beobachtet, in welchen gewisse Grenzwahrscheinlichkeiten notwendigerweise gleich Null oder Eins sind. Zum Beispiel kann die Wahrscheinlichkeit der Konvergenz einer Reihe unabhängiger zufälliger Größen nur diese zwei Werte annehmen[2]. Wir wollen jetzt einen allgemeinen Satz beweisen, welcher mehrere solche Fälle umfaßt.

S a t z. Es seien $x_1, x_2, \ldots, x_n, \ldots$ irgendwelche zufällige Größen und $f(x_1, x_2, \ldots, x_n, \ldots)$ eine Baire sche Funktion[3] der Variablen $x_1, x_2, \ldots, x_n, \ldots$ derart, daß die bedingte Wahrscheinlichkeit

$$\mathsf{P}_{x_1, x_2, \ldots, x_n}\{f(x) = 0\}$$

der Relation

$$f(x_1, x_2, \ldots, x_n, \ldots) = 0$$

bei den bekannten n ersten Größen $x_1, x_2, \ldots, x_n$ der absoluten Wahrscheinlichkeit

$$\mathsf{P}\{f(x) = 0\} \tag{1}$$

für jedes n gleich bleibt. Unter diesen Bedingungen ist die Wahrscheinlichkeit (1) Null oder Eins.

Insbesondere sind die Voraussetzungen des Satzes erfüllt, wenn alle Größen x_n gegenseitig unabhängig sind und der Wert der Funktion $f(x)$ bei einer Änderung von endlich vielen Größen x_n unverändert bleibt.

---

[1] Vgl. A. Khintchine u. A. Kolmogoroff: Über Konvergenz von Reihen. Rec. math. Soc. math. Moscou Bd. 32 (1925) S. 668—677.

[2] Vgl. Kap. VI, § 5. Dasselbe gilt auch für die Wahrscheinlichkeit

$$\mathsf{P}\{s_n - d_n \to 0\}$$

im starken Gesetz der großen Zahlen, wenigstens in dem Falle, daß die Größen x_n gegenseitig unabhängig sind.

[3] Dabei versteht man unter einer Baireschen Funktion eine Funktion, welche ausgehend von Polynomen durch iterierte Grenzübergänge darstellbar ist.

Beweis des Satzes. Wir bezeichnen mit A das Ereignis

$$f(x) = 0.$$

Neben diesem Ereignisse betrachten wir den Körper $\mathfrak{K}$ aller Ereignisse, welche durch irgendwelche Relationen zwischen endlich vielen Größen x_n bestimmt werden können. Gehört ein Ereignis B zu $\mathfrak{K}$, so ist wegen der Bedingungen des Satzes

$$\mathsf{P}_B(A) = \mathsf{P}(A). \tag{2}$$

Im Falle $\mathsf{P}(A) = 0$ ist unser Satz schon bewiesen. Es sei also $\mathsf{P}(A) > 0$. Dann folgt aus (2) die Formel

$$\mathsf{P}_A(B) = \frac{\mathsf{P}_B(A)\,\mathsf{P}(B)}{\mathsf{P}(A)} = \mathsf{P}(B). \tag{3}$$

$\mathsf{P}(B)$ und $\mathsf{P}_A(B)$ sind also zwei vollständig additive Mengenfunktionen, welche auf $\mathfrak{K}$ zusammenfallen, sie müssen folglich auch für jede Menge der BORELschen Erweiterung $B\mathfrak{K}$ des Körpers $\mathfrak{K}$ einander gleich bleiben. Insbesondere ist deswegen

$$\mathsf{P}(A) = \mathsf{P}_A(A) = 1.$$

Womit unser Satz bewiesen ist.

Einige weitere Fälle, in welchen man über gewisse Wahrscheinlichkeiten behaupten kann, daß sie nur die Werte Null und Eins annehmen können, wurden von P. LÉVY entdeckt. Vgl. dazu P. LÉVY: Sur un théorème de M. KHINTCHINE, Bull. des Sc. Math. Bd. 55 (1931) S. 145—160, Théorème II.

Literaturverzeichnis.

[1]. BERNSTEIN, S.: Versuch einer axiomatischen Begründung der Wahrscheinlichkeitsrechnung (russisch). Mitt. math. Ges. Charkow 1917 S. 209—274.

[2]. — Wahrscheinlichkeitsrechnung (russisch). Moskau: Staatsverlag RSFSR. 1927.

[1]. BOREL, E.: Les probabilités dénombrables et leurs applications arithmétiques. Rend. Circ. mat. Palermo Bd. 27 (1909) S. 247—271.

[2]. — Principes et formules classiques, fasc. 1 du tome I du Traité des probabilités par E. BOREL et divers auteurs. Paris: Gauthier-Villars 1925.

[3]. — Applications à l'arithmetique et à la théorie des fonctions, fasc. 1 du tome II du Traité des probabilités par E. BOREL et divers auteurs. Paris: Gauthier-Villars 1926.

[1]. CANTELLI, F. P.: Una teoria astratta del Calcolo delle probabilità. Giorn. Ist. Ital. Attuari Bd. 3 (1932) S. 257—265.

[2]. — Sulla legge dei grandi numeri. Mem. Acad. Lincei Bd. 11 (1916).

[3]. — Sulla probabilità come limite della frequenza. Rend. Accad. Lincei Bd. 26 (1917) S. 39—45.

[1]. COPELAND, H.: The theory of probability from the point of view of admissible numbers. Ann. Math. Statist. Bd. 3 (1932) S. 143—156.

[1]. Dörge, K.: Zu der von R. von Mises gegebenen Begründung der Wahrscheinlichkeitsrechnung. Math. Z. Bd. 32 (1930) S. 232—258.

[1]. Fréchet, M.: Sur la convergence en probabilité. Metron Bd. 8 (1930) S. 1—48.

[2]. — Recherches théoriques modernes, fasc. 3 du tome I du Traité des probabilités par E. Borel et divers auteurs. Paris: Gauthier-Villars (in Vorbereitung).

[1]. Kolmogoroff, A.: Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. Math. Ann. Bd. 104 (1931) S. 415—458.

[2]. — Allgemeine Maßtheorie und Wahrscheinlichkeitsrechnung (russisch). Sbornik trudow sektii totschnych nauk K. A. Bd. 1 (1929) S. 8—21.

[1]. Lévy, P.: Calcul des probabilités. Paris: Gauthier-Villars.

[1]. Lomnicki, A.: Nouveaux fondements du calcul des probabilités. Fundam. Math. Bd. 4 (1923) S. 34—71.

[1]. Mises, R. v.: Wahrscheinlichkeitsrechnung. Leipzig u. Wien: Fr. Deuticke 1931.

[2]. — Grundlagen der Wahrscheinlichkeitsrechnung. Math. Z. Bd. 5 (1919) S. 52—99.

[3]. — Wahrscheinlichkeitsrechnung, Statistik und Wahrheit. Wien: Julius Springer 1928.

[1]. Reichenbach, H.: Axiomatik der Wahrscheinlichkeitsrechnung. Math. Z. Bd. 34 (1932) S. 568—619.

[1]. Slutsky, E.: Über stochastische Asymptoten und Grenzwerte. Metron Bd. 5 (1925) S. 3—89.

[2]. — Zur Frage von den logischen Grundlagen der Wahrscheinlichkeitsrechnung (russisch). Westnik Statistiki Bd. 12 (1922) S. 13—21.

[1]. Steinhaus, H.: Les probabilités dénombrables et leur rapport à la théorie de la mesure. Fundam. Math. Bd. 4 (1923) S. 286—310.

[1]. Tornier, E.: Wahrscheinlichkeitsrechnung und Zahlentheorie. J. reine angew. Math. Bd. 160 (1929) S. 177—198.

[2]. — Grundlagen der Wahrscheinlichkeitsrechnung. Acta math. Bd. 60 (1933) S. 239—380.