AXIOMATISCHER AUFBAU DER WAHRSCHEINLICHKEITSRECHNUNG

Von ALFRED RÉNYI, Budapest

Einführung

Die axiomatische Begründung der Wahrscheinlichkeitsrechnung, die von A. N. KOLMOGOROFF in seiner im Jahre 1933 erschienenen Arbeit[1] gegeben wurde, war die Grundlage der großartigen Entwicklung der Wahrscheinlichkeitsrechnung, die in den letzten zwei Jahrzehnten stattgefunden hat. Im Zuge der Entwicklung tauchten aber auch solche Probleme auf, die im Rahmen der KOLMOGOROFFschen Theorie nicht behandelt werden konnten. Wir denken in erster Linie daran, daß in der Physik (z.B. in der Quantenmechanik), ferner in der Theorie der MARKOFFschen Ketten und der stochastischen Prozesse, bei den Anwendungen der Wahrscheinlichkeitsrechnung in der Integralgeometrie, in der Zahlentheorie und auch auf anderen Gebieten oft solche Wahrscheinlichkeitsverteilungen (zum Teil als Grenzverteilungen) auftreten, die nicht „normiert"[2] werden können, d.h., es treten unbeschränkte Maße auf. Solche Verteilungen können in der KOLMOGOROFFschen Theorie nicht definiert werden; so hat es z.B. keinen Sinn, von einer im ganzen n-dimensionalen euklidischen Raum gleichmäßigen Verteilung zu sprechen; ebenso gibt es keine Wahrscheinlichkeitsverteilung, bei der abzählbar unendlich viele Zustände einer MARKOFFschen Kette gleich wahrscheinlich wären.

Da eine Wahrscheinlichkeit, die größer als 1 ist, prinzipiell sinnlos ist, scheint es auf den ersten Blick, als ob man an der Lage nichts ändern könnte. Dies ist zweifellos der Fall; hier handelt es sich aber darum, daß man mit Hilfe der in Rede stehenden nichtnormierbaren (also eigentlich sinnlosen) Verteilung für gewisse *bedingte* Wahrscheinlichkeiten durchaus sinnvolle, in den physikalischen Anwendungen mit der Erfahrung übereinstimmende Werte gewinnen kann[3]. Dies ist der Grund dafür, daß

[1] A. KOLMOGOROFF, Grundbegriffe der Wahrscheinlichkeitsrechnung. Springer, Berlin 1933.

[2] In der physikalischen Literatur hat sich der nicht sehr glückliche Gebrauch eingebürgert, daß die Dichtefunktionen der Wahrscheinlichkeitsverteilungen nicht normiert, also nur bis auf einen konstanten Faktor bestimmt werden. Ist das Integral der Dichtefunktion über den ganzen Raum endlich, so kann die nichtnormierte Funktion durch Division durch dieses Integral normiert werden; dadurch erhalten wir die wirkliche Dichtefunktion. Ist aber das oben erwähnte Integral der „Dichtefunktion" nicht endlich, so kann sie nicht normiert und daher nicht als gewöhnliche Dichtefunktion betrachtet werden. Diese Tatsache entgeht wegen des obigen Gebrauchs der Aufmerksamkeit mancher Autoren.

[3] Gemeint ist folgendes: Wenn $f(x)$ eine in jedem endlichen Intervall integrierbare und nichtnegative Funktion ist, für welche $\int_{-\infty}^{+\infty} f(x)\,dx = +\infty$ ist, so kann zwar $f(x)$ nicht die Dichtefunk-

solche „nicht existierenden“ Verteilungen gebraucht werden. Es entstand daher die Notwendigkeit, die Axiomatik der Wahrscheinlichkeitsrechnung derart weiterzuentwickeln, daß dadurch die Berechtigung der erwähnten Rechnungen — falls sie besteht — klargestellt wird.

In einer Arbeit[1], die ich im Sommer 1954 beendigt habe, habe ich einen Versuch in dieser Richtung veröffentlicht. Es handelt sich um eine neue axiomatische Begründung der Wahrscheinlichkeitsrechnung, in welcher der grundlegende Begriff gerade der Begriff der bedingten Wahrscheinlichkeit ist; absolute (unbedingte) Wahrscheinlichkeiten sind in dieser Theorie nicht notwendig definiert. Diese axiomatische Theorie ist eine logische Weiterentwicklung der Theorie von A. N. Kolmogoroff.

Wie Prof. B. W. Gnedenko mir freundlicherweise mitgeteilt hat[2], hat A. N. Kolmogoroff selbst die Idee einer solchen Weiterentwicklung seines Axiomensystems in einer Vorlesung erwähnt, aber über seine diesbezüglichen Gedanken nichts veröffentlicht. Es war mir eine große Freude zu erfahren, daß der Weg, den ich gewählt hatte, von Kolmogoroff selbst vorgeschlagen worden war. Während der Ausarbeitung der neuen Theorie (die noch weit davon entfernt ist, einigermaßen abgeschlossen zu sein) habe ich einige Teilergebnisse erhalten, die für sich — also abgesehen von der neuen Theorie — von Interesse sein dürften. Es handelt sich um eine Erweiterung des zentralen Grenzverteilungssatzes auf schwach abhängige Zufallsgrößen, ferner um eine Verallgemeinerung des Borelschen Satzes über normale Dezimalbrüche für Cantorsche Darstellungen von reellen Zahlen[3]. Durch die neue Theorie wird die Problematik der Grenzverteilungssätze der Wahrscheinlichkeitsrechnung wesentlich erweitert durch das Problem der Existenz von bedingten Grenzverteilungen. Wir sagen, daß

---

tion einer Wahrscheinlichkeitsverteilung sein, aber wenn $c \leqq a < b \leqq d$ ist, so kann

$$\frac{\int\limits_a^b f(x)\,dx}{\int\limits_c^d f(x)\,dx}$$

als die bedingte Wahrscheinlichkeit des Ereignisses angesehen werden, daß der Wert einer Zufallsgröße ξ in das Intervall (a, b) fällt, vorausgesetzt, daß dieser Wert zum Intervall (c, d) gehört. Es handelt sich darum, eine solche Berechnung von bedingten Wahrscheinlichkeiten mathematisch zu rechtfertigen. In der Geschichte der Mathematik ist es schon mehrmals vorgekommen, daß durch die Schaffung einer neuen mathematischen Theorie zunächst nicht exakte Verfahren, die sich in der Physik als wirksam erwiesen hatten, mathematisch gerechtfertigt wurden. Ein Beispiel ist die Diracsche δ-Funktion. Es scheint mir, daß die Sachlage in diesen beiden Fällen ziemlich ähnlich ist.

[1] A. Rényi, A valószinüségszámitás uj axiomatikus felépitése. Magyar Tudományos Akadémia III. o. Közleményei 4 (1954) 369–427. In englischer Sprache ist diese Arbeit in Acta Math. Acad. Sci. Hung. **VI** (1955), 285—335, erschienen. Ein kurzer Bericht wird in den Proceedings of the International Congress of Mathematicians, 1954, Bd. I erscheinen.

[2] Auf der Statistischen Konferenz im Juni 1954 in Prag, wo ich über dieses Thema einen Vortrag gehalten habe.

Es waren schon früher mehrere Axiomensysteme der Wahrscheinlichkeitsrechnung vorgeschlagen worden, z. B. von H. Reichenbach (Wahrscheinlichkeitslehre. Sijthoff, Leiden 1935), in welchem der Begriff der bedingten Wahrscheinlichkeit axiomatisiert ist, aber diese Theorien behandeln nicht solche Fälle, in denen gewöhnliche Wahrscheinlichkeiten überhaupt nicht, sondern nur bedingte Wahrscheinlichkeiten definiert sind.

[3] Nachdem man diese Sätze mit Hilfe der neuen Theorie gefunden hatte, konnte man sie nachträglich auch ohne diese Theorie beweisen.

eine Folge $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ von Zufallsveränderlichen die bedingte Grenzverteilungsfunktion $F(x)$ besitzt, wobei $F(x)$ eine monoton nichtabnehmende linksseitig stetige nichtnegative Funktion mit $\lim\limits_{x\to-\infty} F(x) = 0$ und $\lim\limits_{x\to+\infty} F(x) = +\infty$ ist, wenn

$$\lim_{n\to\infty} \frac{\mathsf{P}(\xi_n < x)}{\mathsf{P}(\xi_n < y)} = \frac{F(x)}{F(y)} \tag{1}$$

für alle $x < y$ gültig ist. Zum Beispiel sei ξ_1 gleichverteilt im Intervall $(0, n)$. Dann gilt (1) mit $F(x) = x$ für $x \geqq 0$.

Bedingte Grenzverteilungen spielen eine wichtige Rolle in der Theorie der Markoffschen Ketten mit abzählbar unendlich vielen Zuständen und bei gewissen stochastischen Prozessen.

Was die Beziehung der axiomatischen Theorie zur Wirklichkeit betrifft, möchte ich nur bemerken, daß der Zusammenhang zwischen *bedingter* relativer Häufigkeit und *bedingter* Wahrscheinlichkeit natürlich derselbe ist wie zwischen (nicht-bedingter) relativer Häufigkeit und absoluter Wahrscheinlichkeit.

Ich will nun etwas zu dem formalen mathematischen Gehalt der Theorie sagen. Die Theorie bezieht sich auf eine Mengenfunktion $\mathsf{P}(A \mid B)$ von zwei Veränderlichen, definiert für $A \in \mathsf{T}_1$ und $B \in \mathsf{T}_2$, wobei T_1 ein Borelscher Mengenkörper von Untermengen eines abstrakten Raumes H und T_2 eine *beliebige* nichtleere Untermenge von T_1 bedeutet; es wird vorausgesetzt, daß $\mathsf{P}(A \mid B)$ für jedes festgehaltene $B \in \mathsf{T}_2$ ein σ-additives nichtnegatives Maß auf T_1 ist, für welches $\mathsf{P}(B \mid B) = 1$ ist; so ist $\mathsf{P}(A \mid B)$ ein System von Maßfunktionen, die durch folgende Bedingung miteinander verknüpft sind:

$$\mathsf{P}(A \mid B\,C)\,\mathsf{P}(B \mid C) = \mathsf{P}(A\,B \mid C) \tag{2}$$

(hier bedeutet $A\,B$ bzw. $B\,C$ den Durchschnitt der Mengen A und B bzw. B und C).

Wir nennen H, T_1, T_2 und $\mathsf{P}(A \mid B)$ in ihrer Gesamtheit ein *bedingtes Wahrscheinlichkeitsfeld.*

Es erhebt sich die Frage, unter welchen Bedingungen die Mengenfunktion $\mathsf{P}(A \mid B)$ eine „Quotienten-Darstellung" $\mathsf{P}(A \mid B) = \frac{\mathsf{Q}(A\,B)}{\mathsf{Q}(B)}$ besitzt, wobei $\mathsf{Q}(C)$ ein nichtnegatives und σ-additives Maß auf T_1 ist, für welches $\mathsf{Q}(B) > 0$ für $B \in \mathsf{T}_2$ gilt. A. Császár hat die notwendigen und hinreichenden Bedingungen dafür gefunden. Es ist klar, daß wir ein Kolmogoroffsches Wahrscheinlichkeitsfeld als Spezialfall dann und nur dann erhalten, wenn $H \in \mathsf{T}_2$ und es eine solche Quotientendarstellung für $\mathsf{P}(A \mid B)$ gibt, daß die Maßfunktion $\mathsf{Q}(B)$ beschränkt ist, d.h., wenn $\mathsf{Q}(H) < +\infty$; in diesem Fall kann man freilich ohne Beschränkung der Allgemeinheit $\mathsf{Q}(H) = 1$ annehmen.

Im allgemeinen braucht aber $\mathsf{Q}(B)$ nicht beschränkt zu sein. Zum Beispiel sei $H = R_n$ der n-dimensionale euklidische Raum, T_1 die Menge aller meßbaren Mengen von R_n, $f(x)$ eine auf jeder beschränkten und meßbaren Menge $A \in \mathsf{T}_1$ integrierbare Funktion ($x = (x_1, x_2, \ldots, x_n)$ ist ein Punkt von R_n), ferner bedeute T_2 die Menge aller solchen meßbaren Teilmengen $B \in \mathsf{T}_1$, für welche $\int\limits_B f(x)\,d\,x$ positiv und endlich ist, und es sei

$$\mathsf{P}(A \mid B) = \frac{\int\limits_{AB} f(x)\,d\,x}{\int\limits_B f(x)\,d\,x}, \tag{3}$$

falls $A \in \mathsf{T}_1$ und $B \in \mathsf{T}_2$ ist.

Wir erhalten so ein bedingtes Wahrscheinlichkeitsfeld. Wenn zum Beispiel $f(x) \equiv 1$ ist, so erhalten wir ein bedingtes Wahrscheinlichkeitsfeld, welches dem intuitiven Begriff einer im ganzen Raume R_n gleichmäßigen Verteilung einen widerspruchslosen mathematischen Sinn entsprechen läßt.

Im allgemeinen gibt es keine Quotientendarstellung von $\mathsf{P}(A \mid B)$, nur eine Darstellung in der Form $\mathsf{P}(A \mid B) = \frac{\mathsf{Q}_k(A\,B)}{\mathsf{Q}_k(B)}$, wobei Q_k ein Element einer Menge von Maßfunktionen ist, dessen Wahl von B abhängt. Es scheint, daß durch solche bedingte Wahrscheinlichkeitsfelder manche Schwierigkeiten, die mit dem Begriff der bedingten Wahrscheinlichkeit zusammenhängen, beseitigt werden können. Doch muß vielleicht zu diesem Zwecke unser Axiomensystem durch weitere Axiome ergänzt werden.

§ 1. Das Axiomensystem und seine unmittelbaren Folgerungen

Es seien A und B zwei Mengen; dann bezeichnet $A + B$ die Vereinigung, $A\,B$ den Durchschnitt von A und B. Die leere Menge wird mit O bezeichnet.

Wir gehen von folgenden Definitionen und Axiomen aus:

Es sei eine Menge H gegeben, die wir „Ereignisraum" nennen; die Elemente der Menge H seien durch $a, b, \ldots$ bezeichnet und werden Elementarereignisse genannt. T_1 *sei ein* Borel*scher Mengenkörper von Teilmengen von H, dessen Elemente mit $A, B, C, \ldots$ bezeichnet und „Ereignisse" genannt werden. Wir nehmen an, daß $H \in \mathsf{T}_1$ gilt; es sei ferner ein nichtleeres Mengensystem T_2 gegeben, für welches $\mathsf{T}_2 \subseteq \mathsf{T}_1$ gilt. Es sei* $\mathsf{P}(A \mid B)$ *eine Mengenfunktion zweier Variablen, welche für $A \in \mathsf{T}_1$ und $B \in \mathsf{T}_2$ definiert ist.*

Die Zahl $\mathsf{P}(A \mid B)$ *heiße die bedingte Wahrscheinlichkeit des Ereignisses A in bezug auf das Ereignis B. Es sollen folgende Axiome gelten:*

I. *Für $A \in \mathsf{T}_1$, $B \in \mathsf{T}_2$ gilt*

$$\mathsf{P}(A \mid B) \geqq 0 \quad \textit{und} \quad \mathsf{P}(B \mid B) = 1 .$$

II. *Bei festem B ($B \in \mathsf{T}_2$) ist* $\mathsf{P}(A \mid B)$ *eine σ-additive Mengenfunktion von A, also für $A_k \in \mathsf{T}_1$ ($k = 1, 2, \ldots$) und $A_j\,A_k = O$ für $j \neq k$ gilt*

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

III. *Für $A \in \mathsf{T}_1$, $B \in \mathsf{T}_1$, $C \in \mathsf{T}_2$ und $B\,C \in \mathsf{T}_2$ gilt*

$$\mathsf{P}(A \mid B\,C)\,\mathsf{P}(B \mid C) = \mathsf{P}(A\,B \mid C) .$$

Sind die Axiome I, II, III erfüllt, so nennen wir den Ereignisraum H mit dem Borel*schen Mengenkörper* T_1, *dem Mengensystem* T_2 *und der Mengenfunktion* $\mathsf{P}(A \mid B)$ *ein bedingtes Wahrscheinlichkeitsfeld und bezeichnen es durch*

$$[H, \mathsf{T}_1, \mathsf{T}_2, \mathsf{P}(A \mid B)] .$$

Ist $\mathsf{P}^*(A)$ eine auf dem Borelschen Mengenkörper T_1^* definierte vollständig additive Mengenfunktion, für die $\mathsf{P}^*(H) = 1$ gilt (ist also $[H, \mathsf{T}_1^*, \mathsf{P}^*(A)]$ ein

KOLMOGOROFFsches Wahrscheinlichkeitsfeld), und gilt für $\mathsf{P}^*(B) > 0$ die Beziehung $\mathsf{P}^*(A \mid B) = \frac{\mathsf{P}^*(AB)}{\mathsf{P}^*(B)}$, so erhalten wir, wenn wir mit T_2^* diejenigen Elemente B von T_1 bezeichnen, für die $\mathsf{P}^*(B) > 0$ gilt, ein bedingtes Wahrscheinlichkeitsfeld $[H, \mathsf{T}_1^*, \mathsf{T}_2^*, \mathsf{P}^*(A \mid B)]$, welches wir das vom KOLMOGOROFFschen Wahrscheinlichkeitsfeld $[H, \mathsf{T}_1^*, \mathsf{P}^*(A)]$ erzeugte bedingte Wahrscheinlichkeitsfeld nennen.

Taucht im folgenden für Mengen A und B die Bezeichnung $\mathsf{P}(A \mid B)$ auf, so nehmen wir immer an, daß die Bedingungen $A \in \mathsf{T}_1$ und $B \in \mathsf{T}_2$ erfüllt sind. Es lassen sich einfach folgende fast triviale Folgerungen aus den Axiomen beweisen:

Satz 1. $\mathsf{P}(A \mid B) = \mathsf{P}(AB \mid B)$.

Satz 2. *Für* $B \subseteq C$ *gilt* $\mathsf{P}(AC \mid B) = \mathsf{P}(A \mid B)$.

Satz 3. $\mathsf{P}(A \mid B) \leqq 1$.

Satz 4. *Es gilt* $\mathsf{P}(O \mid B) = 0$.

Satz 5. *Für* $AB = O$ *gilt* $\mathsf{P}(A \mid B) = 0$.

Satz 6. *Ist* $A \subseteq C$, *so gilt* $\mathsf{P}(A \mid B) \leqq \mathsf{P}(C \mid B)$.

Satz 7. *Für* $A \subseteq B \subseteq C$ *gilt* $\mathsf{P}(A \mid B) \leqq \mathsf{P}(A \mid C)$.

Satz 8. *Es ist* $\mathsf{P}(A \mid BC)\, \mathsf{P}(B \mid C) = \mathsf{P}(B \mid AC)\, \mathsf{P}(A \mid C)$.

Satz 9. $\mathsf{P}(H \mid B) = 1$.

Satz 10. *Wenn* $C \subseteq \sum_{k=1}^{\infty} B_k = B$

und

$$ACB_j B_k = O \quad \textit{für} \quad j \neq k \quad (j, k = 1, 2, \ldots)$$

gilt, so ist

$$\mathsf{P}(A \mid C) = \sum_{k=1}^{\infty} \mathsf{P}(A \mid B_k C)\, \mathsf{P}(B_k \mid C).$$

§ 2. Das Gesetz der großen Zahlen

Es bezeichne a ein beliebiges Element von H, und es sei $\xi = \xi(a)$ eine auf H definierte Funktion, die in bezug auf T_1 meßbar ist, d. h., die Menge A_x der Elemente a, für die $\xi(a) < x$ (x beliebig reell), soll zu T_1 gehören. Wir nennen eine solche Funktion eine *Zufallsveränderliche*. Offenbar gehören T_1 auch die Mengen $A_{\mathfrak{B}}$ an, die aus den Elementen a mit $\xi(a) \in \mathfrak{B}$ bestehen, wobei $\mathfrak{B}$ eine beliebige BORELsche Menge auf der Zahlengeraden bedeutet. Die Funktion $\xi = \xi(a)$ erzeugt offenbar ein bedingtes Wahrscheinlichkeitsfeld auf der Zahlengeraden: Sind nämlich $\mathfrak{A}$ und $\mathfrak{B}$ zwei BORELsche Mengen auf der Zahlengeraden und ist A das Urbild von $\mathfrak{A}$ und B das Urbild von $\mathfrak{B}$ in bezug auf $\xi(a)$, ist also A bzw. B die Menge derjenigen $a \in H$, für die $\xi(a) \in \mathfrak{A}$ bzw. $\xi(a) \in \mathfrak{B}$, so genügt die durch

$$\mathfrak{P}(\mathfrak{A} \mid \mathfrak{B}) = \mathsf{P}(A \mid B)]$$

definierte Mengenfunktion (falls $B \in \mathsf{T}_2$) den Axiomen I—III.

Ist $\mathfrak{B}$ eine Menge, deren Urbild zu T_2 gehört und bedeutet $\mathfrak{A}_x$ das Intervall $(-\infty, x)$, so nennen wir die Funktion $F(x \mid \mathfrak{B}) = \mathfrak{P}(\mathfrak{A}_x \mid \mathfrak{B})$ die *bedingte Verteilungsfunktion der Zufallsvariablen* ξ; die Mengenfunktion $\mathfrak{P}(\mathfrak{A} \mid \mathfrak{B})$ nennen wir das *bedingte Verteilungssystem von* ξ. Im Raume R_n kann auf ähnliche Weise ein n-dimensionaler zufälliger Vektor definiert werden; wir nehmen an, daß er eine meßbare Abbildung von H auf den euklidischen Raum R_n vermittelt.

Der Begriff der bedingten Unabhängigkeit für zwei Zufallsvariablen kann folgendermaßen definiert werden: Wir sagen, daß $\xi = \xi(a)$ und $\eta = \eta(a)$, $a \in H$ unabhängig seien in bezug auf $C \in \mathsf{T}_2$, falls für alle reellen Werte von x und y

$$\mathsf{P}(A_y B_x \mid C) = \mathsf{P}(A_x \mid C)\,\mathsf{P}(B_y \mid C) \tag{4}$$

gilt, wobei A_x bzw. B_y die durch

$$\xi(a) < x,\ a \in H \quad \text{bzw.} \quad \eta(a) < y,\ a \in H$$

definierten Mengen bedeuten. Falls $\xi(a)$ und $\eta(a)$ unabhängig in bezug auf alle $B \in \mathsf{T}_2$ sind, so sagen wir, daß $\xi(a)$ und $\eta(a)$ in bezug auf T_2 unabhängig sind. Die Unabhängigkeit mehrerer Zufallsveränderlicher kann auf ähnliche Weise definiert werden.

Der Begriff des *bedingten Mittelwertes* soll folgendermaßen erklärt werden:

$$\mathsf{M}(\xi \mid B) = \int_H \xi(a)\, d\,\mathsf{P}(A \mid B), \tag{5}$$

wobei auf der rechten Seite ein bei festem B genommenes LEBESGUEsches Integral von $\xi(a)$ in bezug auf das Maß $\mathsf{P}(A \mid B)$ steht. Auf ähnliche Weise definiert man die *bedingte Streuung* von ξ in bezug auf B, die wir durch $\mathsf{D}(\xi \mid B)$ bezeichnen.

Es gilt folgende Verallgemeinerung des starken Gesetzes der großen Zahlen:

Satz 1. *Es sei $[H, \mathsf{T}_1, \mathsf{T}_2, \mathsf{P}(A \mid B)]$ ein bedingtes Wahrscheinlichkeitsfeld, $\xi_1, \xi_2, \ldots, \xi_n$ eine Folge von zufälligen Veränderlichen, die in bezug auf T_2 unabhängig sind. Es sei $\mathfrak{B}$ eine BORELsche Menge, und wir nehmen an, daß die Urbilder B_n bei der Abbildung ξ_n von $\mathfrak{B}$ alle zu T_2 gehören, $(n = 1, 2, \ldots)$. Es sei $\mathsf{M}(\xi_k \mid B_k) = M_k$, $\mathsf{D}(\xi_k \mid B_k) = D_k$ $(k = 1, 2, \ldots)$.*

Es sei $C \in \mathsf{T}_2$ und $B_n \subseteq C$ $(n = 1, 2, \ldots)$; ferner $\mathsf{P}(B_n \mid C) = p_n > 0$, $(n = 1, 2, \ldots)$. Es sei $\sum_{n=1}^{\infty} p_n = +\infty$. Es bedeute $\zeta_n(\mathfrak{B})$ den bedingten Mittelwert der Veränderlichen ξ_k $(k = 1, 2, \ldots)$ für die Werte, die zu $\mathfrak{B}$ gehören, d.h., es sei

$$\zeta_n(\mathfrak{B}) = \frac{\sum\limits_{\xi_k \in \mathfrak{B},\ 1 \leq k \leq n} \zeta_k}{\sum\limits_{\xi_k \in \mathfrak{B},\ 1 \leq k \leq n} 1} \qquad (n = 1, 2, \ldots).$$

Falls

$$\lim_{n \to \infty} \frac{\sum\limits_{k=1}^{n} M_k p_k}{\sum\limits_{k=1}^{n} p_k} = M$$

und

$$\sum_{k=1}^{\infty} \frac{p_k D_k^2}{\left(\sum\limits_{j=1}^{k} p_j\right)^2} < +\infty,$$

so ist

$$\mathsf{P}\left(\lim_{n\to\infty} \zeta_n(\mathfrak{B}) = M \mid C\right) = 1\,.$$

Falls das bedingte Wahrscheinlichkeitsfeld $[H, \mathsf{T}_1, \mathsf{T}_2, \mathsf{P}(A \mid B)]$ durch das KOLMOGOROFFsche Wahrscheinlichkeitsfeld $[H, \mathsf{T}_1, \mathsf{P}(A)]$ erzeugt ist und wir $C = H$ annehmen und für $\mathfrak{B}$ die ganze reelle Achse $(-\infty, +\infty)$ wählen, so ist $B_n = H$, $p_n = 1$, $(n = 1, 2, \ldots)$. Somit sind die Bedingungen $\sum_{n=1}^{\infty} p_n = +\infty$ und $B_n \subseteq C$ erfüllt, und es ist

$$\zeta_n = \zeta_n(\mathfrak{B}) = \frac{1}{n} \sum_{k=1}^{n} \xi_k\,.$$

So erhalten wir als Spezialfall den bekannten Satz von KOLMOGOROFF: Wenn die Bedingungen $\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} M_k = M$ und $\sum_{k=1}^{\infty} \frac{D_k^2}{k^2} < +\infty$ erfüllt sind, wobei $M_k = \mathsf{M}(\xi_k)$ und $D_k = \mathsf{D}(\xi_k)$ sind, so ist $\mathsf{P}\left(\lim_{n\to\infty} \zeta_n = M\right) = 1$.

Wir erwähnen einen anderen Spezialfall des Satzes 1. Es sei H das Intervall (0,1), T_1 die Menge aller meßbaren Untermengen von (0,1), T_2 die Menge aller Mengen von T_1, die ein positives LEBESGUEsches Maß besitzen, und es sei $\mathsf{P}(A \mid B) = \frac{m(AB)}{m(B)}$, wobei $m(C)$ das LEBESGUEsche Maß der Menge C bedeutet. Es sei $q_1, q_2, \ldots, q_n, \ldots$ eine beliebige Folge natürlicher Zahlen, die den Bedingungen $q_n \geqq 2$ genügen. Es sei

$$x = \sum_{n=1}^{\infty} \frac{\varepsilon_n(x)}{q_1 q_2 \cdots q_n} \quad (0 < x < 1)\,, \tag{6}$$

wobei $\varepsilon_n(x)$ einen der Werte $0, 1, 2, \ldots, q_n - 1$ annimmt, eine CANTORsche Reihe für x. Es bedeute $\mathfrak{B}$ die Menge der ganzen Zahlen $0, 1, 2, \ldots, s$; es sei $\xi_n(x) = 1$, falls $\varepsilon_n(x) = r$, und anderenfalls $\xi_n(x) = 0$.

Es sei $C = (0,1) = H$. Dann ist, falls $q_n \geqq s\,(n = 1, 2, \ldots)$ die Bedingung $p_n = \mathsf{P}(B_n \mid C) = \frac{s}{q_n} > 0$ erfüllt und falls die Reihe $\sum_{n=1}^{\infty} \frac{1}{q_n}$ divergiert, die Bedingung $\sum_{n=1}^{\infty} p_n = +\infty$ ebenfalls erfüllt. Es ist ferner

$$M_k = \mathsf{M}(\xi_k \mid B_k) = \frac{1}{s+1} \quad \text{und} \quad D_k^2 = \mathsf{D}^2(\xi_k \mid B_k) = \frac{1}{(s+1)^2}\,.$$

Also ist auch die Bedingung

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

erfüllt. So erhalten wir folgendes: Wenn

$$\zeta_n(x) = \frac{\sum\limits_{\substack{\varepsilon_k(x) = r \\ 1 \leqq k \leqq n}} 1}{\sum\limits_{\substack{0 \leqq \varepsilon_k(x) \leqq s \\ 1 \leqq k \leqq n}} 1}$$

gesetzt wird, so konvergiert $\zeta_n(x)$ für fast alle x gegen $\frac{1}{s+1}$.

Offenbar kann die Bedingung $q_n \geqq s$ abgeschwächt werden; es genügt, daß bis auf endlich viele Werte von n die Beziehung $q_n \geqq s$ gültig sein soll. Diese Bedingung ist offenbar für jeden Wert von s erfüllt, falls $q_n \to +\infty$. In diesem Falle kommen daher alle Ziffern asymptotisch gleich oft in der CANTORschen Entwicklung fast aller reellen Zahlen vor, in dem Sinne, der durch folgenden Satz ausgedrückt wird:

Satz 2. *Es sei $q_1, q_2, \ldots, q_n, \ldots$ eine beliebige Folge von natürlichen Zahlen, die den Bedingungen*

(A) $$q_n \geqq 2$$

(B) $$q_n \to +\infty$$

(C) $$\sum_{n=1}^{\infty} \frac{1}{q_n} \to +\infty$$

genügen. Es sei

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

die CANTORsche Reihe der reellen Zahl x $(0 < x < 1)$, wobei $\varepsilon_n(x)$ die Werte $0, 1, \ldots, q_n - 1$ annehmen kann.

Es bedeute $E_{nr}(x)$ die Anzahl, mit der die Ziffer r $(r = 0, 1, 2, \ldots)$ unter den ersten n Ziffern $\varepsilon_1(x), \ldots, \varepsilon_n(x)$ der CANTORschen Entwicklung von x vorkommt. Dann ist für alle Paare von Zahlen r und s $(r, s = 0, 1, 2, \ldots)$ und für fast alle x $(0 < x < 1)$

$$\lim_{n\to\infty} \frac{E_{nr}(x)}{E_{ns}(x)} = 1\,, \tag{7}$$

d.h., alle Ziffern $0, 1, 2, \ldots$ kommen in der CANTORschen Entwicklung von fast allen reellen Zahlen asymptotisch gleich oft vor.

§ 3. Der zentrale Grenzverteilungssatz

Der zentrale Grenzverteilungssatz der Wahrscheinlichkeitsrechnung kann auch für bedingte Wahrscheinlichkeitsfelder verallgemeinert werden. Eine Konsequenz dieser Verallgemeinerung kann auch im Rahmen der gewöhnlichen Theorie formuliert werden und ergibt eine Verallgemeinerung des gewöhnlichen Grenzverteilungssatzes für gewisse Folgen schwach abhängiger Größen. Wir beschränken uns auf die Formulierung dieses Satzes.

Zuerst geben wir folgende Definition: Eine Folge von zufälligen Veränderlichen $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ werde *quasi-unabhängig* genannt, falls

$$\sup_{(a,\, b,\, x_1,\, x_2,\, \ldots,\, x_n)} \frac{\mathsf{P}(a \leqq \xi_{n+1} < b \mid \xi_1 = x_1, \xi_2 = x_2, \ldots, \xi_n = x_n)}{\mathsf{P}(a \leqq \xi_{n+1} < b)} - 1 = \lambda_n \tag{8}$$

für alle $n = 1, 2, \ldots$ endlich ist und $\sum_{n=1}^{n} \lambda_n$ konvergiert[1]. Dann gilt folgender

Satz 3. *Es sei $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ eine quasi-unabhängige Folge von zufälligen Veränderlichen. Es bedeute $F_n(x)$ die Verteilungsfunktion von ξ_n, d. h., es sei $F_n(x) = \mathsf{P}(\xi_n < x)$.*

[1] Es ist leicht zu sehen, daß $\lambda_n \geqq 0$.

Nehmen wir an, daß $\mathsf{M}(\xi_n) = 0$ *ist* $(n = 1, 2, \ldots)$; *es sei ferner* $\mathsf{D}(\xi_k) = D_k$ *und* $B_n = \sqrt{\sum_1^n D_k^2}$. *Falls die* LINDEBERG*sche Bedingung für die Verteilungsfunktion* $F_n(x)$ *gültig ist, d.h., falls für jedes* $\varepsilon > 0$ *die Bedingung*

$$\lim_{n\to\infty} \frac{1}{B_n^2} \sum_{k=1}^{n} \int_{|x| > B_n \varepsilon} x^2 \, d F_k(x) = 0$$

erfüllt ist, so gilt für die Folge ξ_n *der zentrale Grenzverteilungssatz, d.h., wenn*

$$\zeta_n = \frac{\xi_1 + \xi_2 + \cdots + \xi_n}{B_n}$$

gesetzt wird, so ist

$$\lim_{n\to\infty} \mathsf{P}(\zeta_n < x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}} \, du .$$

Die LINDEBERGsche Bedingung ist hier für die Verteilungsfunktionen und nicht für die Veränderlichen ξ_n vorausgesetzt; das ist nicht dasselbe, weil nicht die Unabhängigkeit, sondern nur die Quasi-Unabhängigkeit der Veränderlichen ξ_n vorausgesetzt wurde, und so B_n nicht notwendig gleich der Streuung von $\xi_1 + \xi_2 + \cdots + \xi_n$ ist. Man kann aber leicht folgendes zeigen: Falls die Folge der Zufallsveränderlichen $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ quasi-unabhängig ist, ferner die Streuungen $D_k = \mathsf{D}(\xi_k)$ endlich sind und $\lim_{n\to\infty} \left(\sum_{k=1}^{n} D_k^2 \right) = +\infty$, so ist

$$\lim_{n\to\infty} \frac{\mathsf{D}^2(\xi_1 + \xi_2 + \cdots + \xi_n)}{\sum_{k=1}^{n} D_k^2} = 1 . \tag{9}$$

Das folgt daraus, daß aus Gleichung (8)

$$\mathsf{D}^2(\xi_1 + \xi_2 + \cdots + \xi_n) \leqq \sum_{k=1}^{n} D_k^2 \prod_{j=k}^{n} (1 + \lambda_j) \tag{10}$$

abgeleitet werden kann.