Fundamentalsätze der Wahrscheinlichkeitsrechnung.

Von

R. v. Mises in Frankfurt a. M.

Übersicht.

Wer die Entwicklung der Wahrscheinlichkeitsrechnung in den letzten Jahrzehnten verfolgt, wird sich dem Eindruck nicht verschließen, daß dieser Zweig der mathematischen Wissenschaft in zweierlei Hinsicht hinter allen anderen bedeutend zurückgeblieben ist. Es fehlt einmal — nur wenige Arbeiten russischer Mathematiker bilden da eine Ausnahme — den *analytischen Sätzen* der Wahrscheinlichkeitsrechnung jene *Präzision der Formulierung und Beweisführung*, die in anderen Teilen der Analysis längst zur Selbstverständlichkeit geworden ist. Und es besteht andererseits, trotz mancher wertvoller Ansätze, über die *Grundlagen der Wahrscheinlichkeitsrechnung als einer mathematischen Disziplin* heute noch so gut wie keine Klarheit: was um so erstaunlicher erscheint, als wir nicht nur in einer Zeit lebhaften Interesses für axiomatische Fragen innerhalb der Mathematik, sondern auch in einer Epoche stetig wachsender Ausbreitung der Wahrscheinlichkeitsrechnung auf den verschiedensten Anwendungsgebieten leben.

Die vorliegende Arbeit will nur dem erstangeführten Mangel Rechnung tragen, indem sie versucht, eine große Reihe von Einzelproblemen der Wahrscheinlichkeitsrechnung (das Bernoullische und Poissonsche Problem und seine Verallgemeinerung, das Gaußsche Fehlergesetz, das Gesetz der großen Zahl; dann das Bayessche und Laplace-Bienaymésche Problem und seine Verallgemeinerung, einen zweiten Satz der Fehlertheorie und ein zweites Gesetz der großen Zahl) von einem einheitlichen und allgemeinen analytischen Gesichtspunkt aus zu bewältigen. Dabei ließ es sich naturgemäß nicht vermeiden, daß die Darstellung, namentlich im zweiten Teil dieser Arbeit, auf einem bestimmten System von „Grundlagen" aufgebaut wurde, dessen zusammenhängende und erschöpfende Darstellung in diesem Rahmen unmöglich war. Nur das für die Formulierung der beiden „Fundamentalsätze" Notwendigste ist in den §§ 8 und 10 auseinandergesetzt. Wenn man mit Recht die an diesen Stellen gegebenen Andeutungen für unvollständig und mancher Aufklärung bedürftig erkennen wird, so habe ich doch andererseits alles getan, um die Ergebnisse auch im Anschluß an die heute üblichen Begriffsbildungen und Ausdrucksweisen der Wahrscheinlichkeitsrechnung verständlich zu machen. Eine zusammenfassende Darstellung der „Grundlagen" soll — wenn anders Zeit und Umstände es noch gestatten — bei späterer Gelegenheit vorgelegt werden.

Der *erste* Teil dieser Arbeit, enthaltend die „*analytischen Entwicklungen*", geht von einem einfachen, durchaus nicht tiefliegenden, aber wie es scheint, neuen und jedenfalls für die Wahrscheinlichkeitsrechnung sehr fruchtbaren Satz der Analysis aus, der im wesentlichen folgendes besagt[1]). Sei $f_1, f_2, f_3, \ldots$ eine unbeschränkte Folge von Funktionen einer reellen Variablen x, die an bestimmten, im Endlichen gelegenen Stellen $x = a_1, a_2, a_3, \ldots$ ein *reelles, endliches Maximum* und eine *reelle, nicht verschwindende zweite Ableitung* besitzen, so gilt, gleichmäßig in jedem endlichen Intervall von u:

$$\text{(a)} \qquad \lim_{n=\infty} \left[f_1\left(a_1 + \frac{u}{r_n}\right) \cdot f_2\left(a_2 + \frac{u}{r_n}\right) \cdots f_n\left(a_n + \frac{u}{r_n}\right) \right] = \text{konst.}\, e^{-u^2},$$

wo r_n^2 eine mit n wachsende positive Größe (die negative halbe Summe der genannten zweiten Ableitungen von $f_1 \ldots f_n$) bedeutet. Genügen überdies die f noch gewissen Bedingungen im Unendlichen, so gilt auch

---

[1]) Die *vollständigen* Voraussetzungen der Sätze usw. sind in dieser Einleitung durchwegs *nicht* angegeben.

$$\text{(b)}\quad \lim_{n=\infty} \int_a^b f_1\left(a_1 + \frac{u}{r_n}\right) \cdot f_2\left(a_2 + \frac{u}{r_n}\right) \ldots f_n\left(a_n + \frac{u}{r_n}\right) du = \text{konst.} \int_a^b e^{-u^2} du$$

für beliebige endliche oder unendliche a und b. Von dem speziellen Fall dieses Satzes, in dem alle f außerhalb des Bereiches $0 \leqq x \leqq 1$ verschwinden und innerhalb desselben durch $x^a(1-x)^{1-a}$, $(0 < a < 1)$ dargestellt werden, macht man in der heutigen Wahrscheinlichkeitsrechnung zum Beweis der Laplace-Bayesschen Formel Gebrauch. In den §§ 1 und 2 werden die beiden in (a) und (b) enthaltenen allgemeinen Behauptungen I und II unter dort näher angegebenen Voraussetzungen bewiesen und durch einige für die Anwendungen wichtige Korollare, namentlich die Ausdehnung auf mehrere unabhängige Variable, ergänzt.

Die Ergebnisse der beiden ersten Paragraphen dienen zunächst in § 3 dazu, um zwei im wesentlichen bekannte, aber niemals mit den richtigen Voraussetzungen formulierte, geschweige denn bewiesene Sätze IV und V unter Verwendung eines von Laplace herrührenden Gedankens („erzeugende Funktion") zu erledigen. Sei $v_1, v_2, v_3, \ldots$ eine unbeschränkte Folge reeller, nicht negativer Funktionen der reellen Variablen x, deren von $-\infty$ bis $+\infty$ erstrecktes Integral gleich 1 ist, und die im wesentlichen dadurch charakterisiert sind, daß *ihre Gesamtschwankungen eine endliche obere Schranke* besitzen. Dann gilt, gleichmäßig in jedem endlichen Bereich von u:

$$\text{(d)}\quad \lim_{n=\infty} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} v_1(x_1)\, v_2(x_2) \ldots v_{n-1}(x_{n-1})\, v_n(r_n u + b_n - x_1 - x_2 - \ldots - x_{n-1})\, dx_1 \ldots dx_{n-1} = \frac{1}{\sqrt{\pi}} e^{-u^2},$$

wo r_n und b_n gewisse, durch $v_1, v_2, \ldots, v_n$ bestimmte Größen (b_n die Summe der „Mittelwerte" $a_\varkappa = \int_{-\infty}^{\infty} x v_\varkappa(x)\, dx$, r_n^2 die doppelte Summe der „Streuungen" $\int_{-\infty}^{\infty} (x - a_\varkappa)^2 v_\varkappa(x)\, dx$, $\varkappa = 1, 2, 3, \ldots n$) bezeichnen. Hat man andererseits statt der Funktionen $v_1, v_2, v_3, \ldots$ einzelne Folgen nicht negativer Zahlen $\mathfrak{v}_1(x)$, $\mathfrak{v}_2(x)$, $\mathfrak{v}_3(x), \ldots$ mit $x = 0, 1, 2, \ldots$, wobei die über x erstreckte Summe jedesmal 1 beträgt und die $\mathfrak{v}_\varkappa(x)$ die wesentliche Eigenschaft aufweisen, daß *für jedes $\varkappa$ wenigstens ein Paar unmittelbar aufeinander folgender x-Werte durch nicht verschwindende $\mathfrak{v}$-Werte besetzt* ist, so gilt, ebenfalls gleichmäßig für endliche u:

$$\text{(e)}\quad \lim_{n=\infty} r_n \sum_{x_1} \sum_{x_2} \cdots \sum_{x_{n-1}} \mathfrak{v}_1(x_1) \cdot \mathfrak{v}_2(x_2) \ldots \mathfrak{v}_n(x_n) = \frac{1}{\sqrt{\pi}} e^{-u^2},$$

wenn bei der Summation über alle x-Werte $x_1 + x_2 + \ldots + x_n = r_n u_n + b_n$

gesetzt wird, mit $\lim u_n = u$ und bei analoger Bedeutung von b_n und r_n wie oben. Man erkennt, daß in (e) ein interessanter Satz über die Koeffizienten eines Polynoms, das als ein *Produkt sehr vieler Polynome mit positiven Koeffizienten* erscheint, enthalten ist (vgl. § 7, 4).

In beiden Fällen, sowohl in dem zu (d) führenden Fall „geometrischer“, wie in dem zu (e) führenden „arithmetischer“ Wahrscheinlichkeiten, kann man überdies eine analoge Gleichung für den *Grenzwert des unbestimmten Integrals* der linken Seite von (d) bzw. (e) angeben, wenn man statt der $v_\varkappa(x)$ bzw. $\mathfrak{v}_\varkappa(x)$ jene monotonen, für $x = -\infty$ verschwindenden Funktionen $V_\varkappa(x)$ einführt, für welche $V_\varkappa(b) - V_\varkappa(a)$ gleich dem von a bis b erstreckten Integral von $v_\varkappa(x)$, bzw. der von a bis b erstreckten Summe der $\mathfrak{v}_\varkappa(x)$ gleichkommt:

$$\text{(c)}\quad \lim_{n=\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} V_n(r_n u + b_n - x_1 - x_2 - \cdots - x_{n-1})\, dV_{n-1}(x_{n-1})\, dV_{n-2}(x_{n-2}) \cdots dV_1(x_1) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{u} e^{-x^2} dx,$$

wobei die Integrale links im Stieltjesschen Sinne zu bilden sind. Diese Gleichung (c) gilt gleichmäßig für *alle* Werte von u, ihre Geltung geht aber weit über das bisher Gesagte hinaus.

Definiert man nämlich $V_1(x)$, $V_2(x)$, $V_3(x)$, ... *ohne Rücksicht auf Eigenschaften ihrer Ableitungen*, als monotone, von null bis eins ansteigende Funktionen, die im wesentlichen nur der Bedingung genügen, daß die *Folge der „m-ten Momente“* $\int_{-\infty}^{\infty} x^m dV_\varkappa(x)$ *bei festem, aber beliebigem* m für $\varkappa = 1, 2, 3, \ldots$ *beschränkt* ist, so läßt sich (c) immer noch beweisen, während weder (d) noch (e) einen Sinn hat (Behauptung III). Der Beweis ist aber keineswegs mehr durch Zurückführung auf (a) und (b) zu erbringen, sondern erfordert ganz andere Hilfsmittel, mit denen wir ihn in den §§ 4 und 5 im Anschluß an einen von Tschebyschef und Markoff eingeschlagenen Weg[2]) erledigen. Durch die gemeinschaftlichen Bemühungen dieser beiden Mathematiker ist tatsächlich ein Beweis von (c), unter Heranziehung der Theorie der Kettenbruch-Entwicklungen, wenigstens für die Fälle, in denen die $V_\varkappa$ beschränkte Ableitungen haben oder durch einfache Treppenlinien dargestellt werden, zustande gekommen; unser allgemeiner Beweis arbeitet dagegen nur mit elementaren und anschaulichen Hilfsmitteln. Für die Sätze (d) und (e), die nicht etwa aus (c)

[2]) Vgl. A. A. Markoff, Wahrscheinlichkeitsrechnung, deutsch v. H. Liebmann, Leipzig 1912, S. 77ff., 259ff. P. L. Tschebyschef, Oeuvres, St. Pétersbourg 1899/1907, Bd. II, S. 481. Vgl. a. Fußnote 7.

abgeleitet werden können, scheinen Beweise bisher noch nicht versucht worden zu sein.

In § 6 werden die in den vorstehenden Gleichungen (c) bis (e) angedeuteten Behauptungen III bis V durch einige Korollare ergänzt, die namentlich die Ausdehnung auf mehrere unabhängige Veränderliche und verschiedene Varianten der Voraussetzungen zum Gegenstande haben.

Die im *zweiten* Teil der Arbeit enthaltenen „*Anwendungen in der Wahrscheinlichkeitsrechnung*" führen zu zwei allgemeinen, hier als „Fundamentalsätzen" bezeichneten Ergebnissen, von denen das erste sich auf die Behauptungen III bis V, das zweite unmittelbar auf die Behauptungen I und II des ersten Teiles stützt. Man kann die beiden großen Problemgruppen, von denen die Fundamentalsätze handeln, im Anschluß an die übliche Terminologie, als den Bernoullischen und den Bayesschen Ideenkreis charakterisieren.

Bernoullis Fragestellung, gehörig erweitert, aber immer unter Verwendung der heutigen Ausdrucksweise dargestellt, geht dahin (§ 8): Welches ist die Wahrscheinlichkeit eines Ergebnisses, das sich als die Summe sehr vieler, unabhängigen Zufallsgesetzen gehorchender, Einzelversuche darstellt? Der einfachste, von Bernoulli selbst behandelte Fall (§ 7, 2) ist bekanntlich der: Aus einer Urne, die lauter sonst gleiche, aber teils mit Null, teils mit Eins bezeichnete Kugeln im Mischungsverhältnis $(1-q):q$ enthält, wird in n-facher Wiederholung je eine Kugel gezogen; wie groß ist die Wahrscheinlichkeit $\mathfrak{w}_n(x)$, wenn n eine sehr große Zahl ist, gerade die Summe x (also x mit Eins bezeichnete Kugeln) zu ziehen? Laplace hat die Frage dahin beantwortet, daß $\mathfrak{w}_n(x)$ bei großem n asymptotisch durch das „Gaußsche Gesetz" e^{-u^2} mit $u=\left(\frac{x}{n}-q\right)\sqrt{\frac{n}{2q(1-q)}}$ dargestellt wird. Unsere Gleichung (e), die in die Aussage des ersten Fundamentalsatzes eingeht, besagt, daß dieses Gesetz viel allgemeiner gilt: nicht nur (Poissonscher Fall), wenn jeder der n Züge aus einer anderen Urne (mit anderem Mischungsverhältnis) erfolgt (§ 7, 3), sondern auch dann, wenn jede Urne in beliebiger Mischung mit beliebigen Zahlen 0, 1, 2, 3, ... bezeichnete Kugeln enthält (§ 7, 4), endlich auch analog, wenn jede einzelne Ziehung durch eine endliche Gruppe von Zahlen charakterisiert wird (z. B. Fünfnummernziehung beim Zahlenlotto, § 7, 5). Grundsätzlich das gleiche Problem liegt in der *Fehlertheorie* vor, wenn nach der Wahrscheinlichkeitsdichte $w_n(x)$ für das Auftreten eines Fehlers x gefragt wird, der als die Summe von n, unabhängigen Fehlergesetzen genügenden, Elementarfehlern aufgefaßt wird (§ 8, 5). Hier sagt unsere Gleichung (d), die ebenfalls einen Teil des ersten Fundamentalsatzes bildet, daß, wie immer die Gesetze der Elementarfehler beschaffen sein mögen, bei genügend großem n das Gaußsche Gesetz für den Gesamtfehler resultiert. Endlich kann man Gleichung (c) heranziehen, um eine, wenn auch eingeschränktere, Aussage für *beliebige* „Verteilungen“ $V_\varkappa(x)$ zu gewinnen. Wendet man aber das Augenmerk lediglich auf den Umstand, daß der Faktor von x^2 im Exponenten des Gaußschen Gesetzes mit $1:n$ unendlich klein wird, so erhält man (§ 8, 6) das sog. *Gesetz der großen Zahl*, das wir das „erste“ nennen, weil ihm in der Problemgruppe des zweiten Fundamentalsatzes ein analoges zur Seite tritt.

Überhaupt führt die Bayessche Fragestellung, deren Ausgestaltung im *zweiten* Fundamentalsatz (§ 10) ihren Ausdruck findet, zu zahlreichen Analogien mit den Folgerungen aus dem ersten. Das Ausgangsproblem ist folgendes (§ 9, 1): Man hat in n Zügen aus einer Urne mit schwarzen und weißen Kugeln an mal schwarz gezogen; gefragt ist nach der Wahrscheinlichkeitsdichte $w_n(x)$ dafür, daß die Wahrscheinlichkeit eines schwarzen Einzelergebnisses bei x liegt. Nach Laplace folgt $w_n(x)$, die sog. „a posteriori-Wahrscheinlichkeit“, für große n asymptotisch dem Gaußschen Gesetz e^{-u^2} mit $u = (x-a)\sqrt{\frac{n}{2a(1-a)}}$, sobald die „a priori-Wahrscheinlichkeit aller x-Werte *gleich* angenommen wird. Unser zweiter Fundamentalsatz, der sich auf die Gleichungen (a), (b) und die zugehörigen Korollare stützt, besagt zunächst, daß dieses Ergebnis *von dem Verlauf der a priori-Wahrscheinlichkeiten ganz unabhängig* ist (§ 9, 2), wodurch erst seine große praktische Bedeutung begründet wird; weiter, daß die analogen Aussagen möglich sind, wenn die Urne in beliebiger Mischung nicht nur zweierlei, sondern *beliebig vielerlei*, etwa mit 0, 1, 2, 3, ... bezeichnete, Kugeln enthält (§ 9, 3). Ein wichtiges Problem der *Fehlertheorie* fällt in diesen Ideenkreis (§ 10, 4): Man habe als Resultat von n Beobachtungen einer unbekannten Größe $a_1 n$ mal den Wert b_1, $a_2 n$ mal $b_2 \ldots a_k n$ mal b_k erhalten; wie groß ist die „Wahrscheinlichkeitsdichte a posteriori“ $w_n(x)$ dafür, daß der „wahre“ Wert der Beobachtung (d. i. derjenige, der sich als Durchschnitt bei einer ins Unendliche fortgesetzten Beobachtungsreihe ergeben müßte) bei x liegt? Der zweite Fundamentalsatz erteilt, in Erweiterung eines von Bienaymé bewiesenen Laplaceschen Satzes (§ 9, 4), die Antwort, daß — unabhängig von jeder Annahme über die a priori-Wahrscheinlichkeiten — $w_n(x)$ bei großem n proportional e^{-u^2} wird, mit $u = (x-A) : s\sqrt{2}$, wo A den Mittelwert und s^2 die Streuung der n Beobachtungs-Ergebnisse $\left(A = \sum\limits_m a_m b_m,\ s^2 = \sum\limits_m a_m (b_m - A)^2\right)$ bedeutet. Dieses Resultat, das von der Annahme des Gaußschen Fehlergesetzes ganz unabhängig ist, aus dieser Annahme aber auch keineswegs abgeleitet werden könnte, nennen wir den *zweiten Satz der Fehlertheorie* (§ 10, 4). Endlich ergibt sich (§ 10, 5) aus der Bemerkung, daß in der

Laplace-Bayesschen Formel der Faktor von x^2 im Exponenten mit n wächst, das schon erwähnte *zweite Gesetz der großen Zahl,* dessen einfache Formulierung man am Schlusse dieser Arbeit findet.

Die präzise Darstellung der Fundamentalsätze erforderte, wie oben bemerkt wurde, die Einführung einer ganzen Reihe neuer Begriffe und Bezeichnungen, die nach Ansicht des Verfassers bei einem logischen Aufbau der Wahrscheinlichkeitsrechnung *als eines Teiles der Mathematik* nicht entbehrt werden können. Da die darauf bezüglichen Erklärungen und Bemerkungen im Texte (§ 8, 1—3 und § 10, 1—2) ohnehin so kurz als möglich gehalten wurden, erübrigt sich eine auszugsweise Wiedergabe an dieser Stelle. Der wesentliche Inhalt der Sätze und ihrer Anwendungen bleibt natürlich von der Annahme oder Ablehnung des vorgebrachten Systems von „Grundlagen“ unabhängig.

I. Teil. Analytische Entwicklungen.

§ 1. Sätze über den Grenzwert eines Produktes von Funktionen.

1. Bezeichnungen. Es seien $f_1(x)$, $f_2(x)$, $f_3(x), \ldots$ reelle oder komplexe Funktionen der reellen Variablen x, eindeutig definiert für alle, endlichen oder unendlichen, reellen Werte von x. Es mögen ferner x_1, x_2, $x_3, \ldots$ reelle Zahlen bezeichnen, die in einer bestimmten, noch anzugebenden Weise *linear* von einer einzigen reellen Zahl u abhängen. Wir betrachten das Produkt

$$p_n(u) = f_1(x_1) \cdot f_2(x_2) \ldots f_n(x_n) \tag{1}$$

sowie sein bestimmtes Integral

$$P_n(a, b) = \int_a^b p_n(u)\, du \tag{2}$$

und untersuchen insbesondere die Grenzwerte

$$p(u) = \lim_{n=\infty} p_n(u) = f_1(x_1) \cdot f_2(x_2) \ldots \tag{1'}$$

und

$$P(a, b) = \lim_{n=\infty} P_n(a, b) = \lim_{n=\infty} \int_a^b p_n(u)\, du, \tag{2'}$$

wobei naturgemäß die $f_\varkappa$ gewissen Beschränkungen unterworfen werden. Diese Produktbildung genügt dem kommutativen Gesetz (Vertauschbarkeit der Faktoren), dagegen dem assoziativen (beliebige Zusammenfaßbarkeit der Faktoren) nur unter Einschränkungen für die Beziehung zwischen u und den $x_\varkappa$, die wir nicht als erfüllt annehmen.

2. Voraussetzungen. a) Von *jeder einzelnen* der Funktionen f_1, f_2, $f_3, \ldots$ setzen wir voraus:

1. Die Funktion $f_\varkappa(x)$ habe an einer im Endlichen gelegenen bestimmten Stelle $x = a_\varkappa$ den *reellen Wert* 1 und sei hier zweimal differenzierbar; der erste Differentialquotient sei *null*, der zweite *reell, endlich* und *nicht positiv*:

$$f_\varkappa(a_\varkappa) = 1, \quad f_\varkappa'(a_\varkappa) = 0, \quad f_\varkappa''(a_\varkappa) = -2s_\varkappa^2. \tag{a 1}$$

2. Der Betrag des *dritten Differenzenquotienten* von $f_\varkappa(x)$ und damit der des ersten von $f_\varkappa''(x)$ habe an der Stelle $x = a_\varkappa$ eine *endliche obere Schranke*. Danach gilt für die Funktion

$$\vartheta_\varkappa(y) = \frac{\log \operatorname{nat} f_\varkappa(a_\varkappa + y) + s_\varkappa^2 y^2}{y^2}, \tag{3}$$

die zufolge (a 1) an der Stelle $x = a_\varkappa$ den Grenzwert null annimmt, bei hinreichend kleinem $y_0 > 0$, für ein bestimmtes endliches $C_\varkappa'$:

$$|\vartheta_\varkappa(y)| < C_\varkappa' |y|, \quad \text{für} \quad |y| < y_0. \tag{a2}$$

3. An allen Stellen, die von $a_\varkappa$ verschieden sind, sei der Absolutwert von $f_\varkappa(x)$ *kleiner als* 1 und habe auch nicht den Grenzwert 1. Es existiert also zu jedem, noch so klein vorgegebenen Y_0 ein positives $\delta_\varkappa$, so daß

$$|f_\varkappa(a_\varkappa + y)| < 1 - \delta_\varkappa \quad \text{für} \quad |y| > Y_0. \tag{a3}$$

Es genügt übrigens, daß diese Ungleichung für alle y mit Ausnahme einer Menge vom Inhalt null besteht.

4. Für $x = +\infty$ und $x = -\infty$ verschwinde $|f_\varkappa(x)|$ mit einer positiven Potenz von $\frac{1}{x}$, d. h. es gibt ein $\alpha_\varkappa > 0$ derart, daß für genügend großes $X_\varkappa$:

$$|x^{\alpha_\varkappa} f_\varkappa(x)| < 1 \quad \text{für} \quad |x| > X_\varkappa. \tag{a4}$$

b) Von der *Gesamtheit* der Funktionen f_1, f_2, $f_3, \ldots$ setzen wir voraus:

1. Die Folgen der Zahlen $a_1, a_2, a_3, \ldots$ und $s_1^2, s_2^2, s_3^2, \ldots$ seien *beschränkt*, die Summe der letzteren wachse aber *stärker als mit der* $^2/_3$ *Potenz* von n ins Unendliche, d. h. es gebe zwei positive endliche Zahlen A, S^2 und eine für $n = \infty$ verschwindende Funktion $\varepsilon(n) > 0$, so daß für alle $\varkappa$

$$|a_\varkappa| < A, \quad |s_\varkappa^2| < S^2, \quad \frac{n^{\frac{2}{3}}}{\sum\limits_{\varkappa=1}^{n} s_\varkappa^2} < \varepsilon(n). \tag{b 1}$$

2. Es gebe *ein von $\varkappa$ unabhängiges C'*, so daß für alle $\varkappa$, bei hinreichend kleinem y_0:

$$|\vartheta_\varkappa(y)| < C' |y| \quad \text{für} \quad |y| < y_0. \tag{b 2}$$

Hieraus folgt u. a., daß das arithmetische Mittel $\frac{1}{n}\sum_{\varkappa=1}^{n}\vartheta_\varkappa(y)$ im Bereich $|y| < y_0$ beschränkt ist und stetig mit y gegen null geht.

3. Es gebe *ein von $\varkappa$ unabhängiges positives δ*, so daß für alle $\varkappa$, bei beliebig vorgegebenem Y_0:

$$(\text{b}\,3) \qquad |f_\varkappa(a_\varkappa + y)| < 1 - \delta \quad \text{für} \quad |y| > Y_0,$$

höchstens mit Ausnahme einer Menge von y-Werten vom Inhalt null.

4. Es gebe *ein von $\varkappa$ unabhängiges X* und ein ebensolches $\alpha > 0$, so daß für alle $\varkappa$

$$(\text{b}\,4) \qquad |x^\alpha f_\varkappa(x)| < 1 \quad \text{für} \quad |x| > X.$$

Man sieht ohne weiteres, daß, wenn alle $f_\varkappa$ untereinander gleich sind, die Voraussetzungen b) in den a) enthalten sind. Allgemein kann man sagen, daß durch die b) eine gewisse *Gleichmäßigkeit* in der Erfüllung der a) gefordert wird.

3. Behauptungen. I. *Genügen die $f_1, f_2, f_3, \ldots$ den eben angeführten Voraussetzungen* (a 1), (a2) *und* (b 1), (b2) *und setzt man*

$$(4) \qquad s_1^2 + s_2^2 + s_3^2 + \ldots + s_n^2 = r_n^2, \quad x_\varkappa = a_\varkappa + \frac{u}{r_n}, \qquad (\varkappa = 1, 2, \ldots, n)$$

so gilt für die mit (1) *und* (1′) *definierte Funktion* $p(u)$

$$(\text{I}) \qquad p(u) = e^{-u^2}$$

in jedem endlichen Bereich von u. Oder ausführlicher: *Es läßt sich stets ein n_0 finden, so daß*

$$(\text{I}') \qquad \left|f_1\left(a_1 + \frac{u}{r_n}\right)\cdot f_2\left(a_2 + \frac{u}{r_n}\right)\cdots f_n\left(a_n + \frac{u}{r_n}\right) - e^{-u^2}\right| < \varepsilon$$
$$\text{für} \quad |u| < u_0, \quad n > n_0$$

bei beliebig vorgegebenem ε und u_0.

II. *Genügen die Funktionen $f_1, f_2, f_3, \ldots$ sämtlichen Voraussetzungen* (a1) *bis* (a4) *und* (b1) *bis* (b4), *sind sie überdies integrabel und haben die $s_\varkappa^2$ eine von null verschiedene untere Schranke s^2, so gilt auch für die in* (2) *und* (2′) *definierte Funktion* $P(a, b)$

$$(\text{II}) \qquad P(a,b) = \int_a^b e^{-u^2}\,du$$

für beliebige endliche oder unendliche a, b. Oder ausführlicher: *Es gibt stets ein n_0, so daß*

$$(\text{II}') \quad \left|\int_a^b f_1\left(a_1 + \frac{u}{r_n}\right)\cdot f_2\left(a_2 + \frac{u}{r_n}\right)\cdots f_n\left(a_n + \frac{u}{r_n}\right)du - \int_a^b e^{-u^2}\,du\right| < \varepsilon \text{ für } n > n_0$$

bei beliebig vorgegebenem ε und endlichen oder unendlichen a, b. Zugleich gilt auch Gl. (I) *bzw.* (I′) *für die ganze u-Achse, höchstens mit Ausnahme einer Nullmenge von Punkten.*

Man sieht leicht ein, daß die Festsetzung (4) die Bedeutung hat, für jedes n

$$(4') \qquad p_n(0) = 1, \quad p_n'(0) = 0, \quad p_n''(0) = -2$$

zu bewirken. Denn es ist mit Rücksicht auf (a 1)

$$p_n(0) = f_1(a_1) \cdot f_2(a_2) \ldots f_n(a_n) = 1;$$

ferner

$$p_n'(0) = \frac{1}{r_n} \sum_{\varkappa=1}^{n} \frac{p_n(0)}{f_\varkappa(a_\varkappa)} f_\varkappa'(a_\varkappa) = 0;$$

endlich

$$p_n''(0) = \frac{1}{r_n^2} \sum_{\varkappa=1}^{n} \frac{p_n(0)}{f_\varkappa(a_\varkappa)} f_\varkappa''(a_\varkappa) = -\frac{2}{r_n^2} \sum_{\varkappa=1}^{n} s_\varkappa^2 = -2.$$

4. Beweis der ersten Behauptung. Aus der Definition von $\vartheta_\varkappa$ in Gleichung (3) folgt:

$$(5) \qquad \log\operatorname{nat} f_\varkappa(a_\varkappa + y) = -s_\varkappa^2 y^2 + y^2 \vartheta_\varkappa(y).$$

Setzt man hier $y = u : r_n$ und summiert die Gleichungen für $\varkappa = 1$ bis n, so erhält man mit Rücksicht auf (1) und (4):

$$(5') \qquad \log\operatorname{nat} p_n(u) = -u^2 + \frac{u^2}{r_n^2} \sum_{\varkappa=1}^{n} \vartheta_\varkappa\left(\frac{u}{r_n}\right).$$

Für den Absolutwert des zweiten Summanden rechts hat man nach Voraussetzung (b 1) und (b 2):

$$(5'') \qquad \left|\frac{u^2}{r_n^2} \sum \vartheta_\varkappa\left(\frac{u}{r_n}\right)\right| < C' n \left|\frac{u^3}{r_n^3}\right| < C' u_0^3 [\varepsilon(n)]^{\frac{3}{2}},$$

vorausgesetzt, daß $|u : r_n| < y_0$ bleibt. Dies letztere läßt sich für $|u| < u_0$, da r_n nach (b 1) mit n ins Unendliche wächst, durch Vergrößerung von n erreichen, womit also gezeigt ist, daß

$$(5''') \qquad |\log\operatorname{nat} p_n(u) + u^2| < \varepsilon', \quad \text{für} \quad |u| < u_0,$$

bei hinreichend großem n. Aus (5‴) folgt, daß $p_n e^{u^2}$ sich beliebig wenig von 1 unterscheidet, wenn nur n groß genug ist, und daraus die Behauptung I.

Man sieht, daß die Behauptung I lediglich auf dem Verhalten der Funktionen $f_\varkappa$ *in der Umgebung der Stellen* $a_\varkappa$ beruht.

5. Hilfssatz. Ehe wir zum Beweis der zweiten Behauptung übergehen, erledigen wir folgenden Hilfssatz. *Sei $\eta(u)$ eine, im allgemeinen komplexe, Funktion der reellen Variablen u, deren Absolutwert die Größe η_0 zur oberen Schranke hat. Dann gibt es zu jedem vorgegebenen ε ein $u_0 > 0$ und ein $\eta_0 < \frac{1}{2}$, so daß*

$$(6) \quad \left| \int_0^U e^{-u^2[1+\eta(u)]} du - \frac{\sqrt{\pi}}{2} \right| < \varepsilon \quad \text{für alle} \quad |\eta(u)| \leqq \eta_0 \quad \text{und} \quad U > u_0.$$

Beweis: Sei $\alpha(u)$ der reelle Bestandteil von $\eta(u)$, so daß auch $|\alpha(u)| \leqq \eta_0$, so ist

$$|e^{-u^2(1+\eta)}| = |e^{-u^2(1+\alpha)}| \leqq e^{-u^2(1-\eta_0)}.$$

Daraus folgt für beliebiges u_0 und $\eta_0 < 1$:

$$(6') \quad \left| \int_{u_0}^{\infty} e^{-u^2(1+\eta)} du \right| < \int_{u_0}^{\infty} e^{-u^2(1-\eta_0)} du = \frac{\sqrt{\pi}}{\sqrt{1-\eta_0}} [1 - \Phi(u_0\sqrt{1-\eta_0})],$$

wenn Φ das Gaußsche Integral

$$(7) \qquad \Phi(x) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{x} e^{-u^2} du$$

bezeichnet. Da bekanntlich $\Phi(x)$ mit $x = \infty$ monoton gegen 1 geht, kann man ein u_0 so wählen, daß, etwa für $\eta_0 = \frac{1}{2}$, und um so mehr für $\eta_0 < \frac{1}{2}$, der Ausdruck rechts in $(6')$ kleiner als $\frac{\varepsilon}{3}$ wird. Dann gilt für den Integranden in (6) bei $\eta_0 \leqq \frac{1}{2}$:

$$(6'') \qquad \left| \int_0^U - \int_0^{u_0} \right| < \frac{\varepsilon}{3} \quad \text{für} \quad U > u_0$$

und überdies, da auch $\eta_0 = 0$ gesetzt werden darf:

$$(6''') \qquad \left| \int_0^{u_0} e^{-u^2} du - \frac{\sqrt{\pi}}{2} \right| < \frac{\varepsilon}{3},$$

weil $\frac{\sqrt{\pi}}{2}$ der Wert des von null bis $+\infty$ erstreckten Integrals von e^{-u^2} ist.

Bezeichnet man mit $e_1(u)$ und $i\,e_2(u)$ den reellen bzw. komplexen Bestandteil von $e^{-u^2\eta(u)}$, so folgt aus dem ersten Mittelwertsatz der Integralrechnung:

$$\int_0^{u_0} e^{-u^2[1+\eta(u)]} du = [e_1(u_1) + i\,e_2(u_2)] \int_0^{u_0} e^{-u^2} du \qquad \begin{cases} 0 < u_1 < u_0 \\ 0 < u_2 < u_0 \end{cases}$$

Da man durch Herabdrücken von η_0 unter $\frac{1}{2}$ den Faktor vor dem Integralzeichen rechts beliebig nahe der reellen Einheit bringen kann, so ergibt sich weiter, daß durch entsprechende Wahl von η_0 — nach erledigter Verfügung über u_0 —

$$(6'''') \qquad \left|\int_0^{u_0} e^{-u^2[1+\eta(u)]}du - \int_0^{u_0} e^{-u^2}du\right| < \frac{\varepsilon}{3}$$

gemacht werden kann. Durch Addition von $(6'')$, $(6''')$ und $(6'''')$ folgt (6) und damit der Hilfssatz.

6. Beweis der zweiten Behauptung. Für endliche a, b läßt sich die Gleichung (II) aus der Behauptung I in einfachster Weise, etwa mit Hilfe des ersten Mittelwertsatzes der Integralrechnung, ableiten. Es genügt demnach, Gleichung (II) für den Fall $a = 0$, $b = \infty$ zu beweisen, da der Beweis für $a = -\infty$, $b = 0$ in ganz gleicher Weise zu führen ist. Wir haben also zu zeigen, daß

$$(8) \qquad \left|\int_0^{\infty} p_n(u)du - \frac{\sqrt{\pi}}{2}\right| < \varepsilon$$

wird, bei hinreichend großem n und bei der durch (1) und (4) festgelegten Bedeutung von $p_n(u)$. Zu diesem Zweck zerlegen wir das Intervall $0, \infty$ durch noch näher zu bestimmende Zwischenwerte U und U' in drei Teile und zeigen, daß bei entsprechender Wahl von U und U' für hinreichend große n:

$$(8') \qquad \left|\int_0^{U} p_n(u)\,du - \frac{\sqrt{\pi}}{2}\right| < \varepsilon',$$

$$(8'') \qquad \left|\int_U^{U'} p_n(u)du\right| < \varepsilon'',$$

$$(8''') \qquad \left|\int_{U'}^{\infty} p_n(u)du\right| < \varepsilon'''.$$

Um zunächst $(8')$ zu beweisen, beachten wir, daß nach $(5')$

$$(9) \qquad p_n(u) = e^{-u^2[1+\eta(u)]} \quad \text{mit} \quad \eta(u) = -\frac{1}{r_n^2}\sum_{\varkappa=1}^{n}\vartheta_\varkappa\left(\frac{u}{r_n}\right).$$

Nach dem oben dargelegten Hilfssatz bestimmen wir u_0 und η_0 zu dem in $(8')$ vorgeschriebenen ε'. Zufolge der Voraussetzung (b2) und wegen $s_\varkappa^2 > s^2$ gilt für das in (9) definierte $\eta(u)$

$$(9') \qquad |\eta(u)| < \left|C' n \frac{u}{r_n^3}\right| \leqq \frac{C'}{s^2}\frac{u_0}{r_n}, \quad \text{für} \quad \left|\frac{u_0}{r_n}\right| \leqq y_0.$$

Wir können also $|\eta(u)| < \eta_0$ erreichen, indem wir ein n_0 so groß wählen, daß $u_0 : r_{n_0}$ erstens kleiner wird als das nach (b2) bestimmte y_0 und zweitens kleiner als $\eta_0 s^2 : C'$. Damit ist $(8')$ zunächst für $n \geqq n_0$ und $U = u_0$ bewiesen; die Gleichung gilt aber auch zufolge des im vorausgehenden Abschnitt bewiesenen Hilfssatzes für größere U, wenn nur der Quotient

$U : r_n$ durch entsprechende Vergrößerungen von n unter $u_0 : r_{n_0}$ gehalten wird, weil dabei $|\eta|$ unter η_0 bleibt. Also ist allgemein:

$$(9'') \qquad \left| \int_0^{r_n y_0} p_n(u)\,du - \frac{\sqrt{\pi}}{2} \right| < \varepsilon' \quad \text{für} \quad n > n_0,$$

bei einem durch ε' und die Funktionen $f_1, f_2, f_3, \ldots$ bestimmten, endlichen y_0 (das nicht mit dem y_0 von (b 2) identisch sein muß).

In das Integral (8″) führen wir an Stelle von u die neue Integrations-Variable y durch $u = r_n y$ ein und erhalten so:

$$(10) \qquad r_n \int_{y_0}^{y_1} p_n(r_n y)\,dy = r_n \int_{y_0}^{y_1} f_1(a_1 + y) \cdot f_2(a_2 + y) \ldots f_n(a_n + y)\,dy,$$

wenn mit y_1 der zu U' gehörige Wert $U' : r_n$ bezeichnet wird. In (10) ist $|y|$ stets größer als y_0, es gibt also nach Voraussetzung (b 3) ein positives δ, so daß $|f_\varkappa| < 1 - \delta$. Daher ist nach dem Mittelwertsatz der Absolutwert von (10) kleiner als

$$(10') \qquad r_n (y_1 - y_0)(1 - \delta)^n < \sqrt{n} \cdot S \cdot (y_1 - y_0) \cdot (1 - \delta)^n,$$

wobei S die in (b 1) eingeführte obere Schranke der $|s_\varkappa|$. Die Nullmenge, in der die Ungleichung (b 3) für $f_\varkappa$ etwa nicht erfüllt ist, kommt bei der Integration nicht in Betracht. Für hinreichend große n wird (10′) beliebig klein, so daß für ein gewisses n'

$$(10'') \qquad \left| \int_{r_n y_0}^{r_n y_1} p_n(u)\,du \right| < \varepsilon'' \quad \text{für} \quad n > n'$$

gilt, bei beliebigen endlichen y_0, y_1.

Mit dem Integral (8‴) verfahren wir zunächst in gleicher Weise und schreiben dafür

$$(11) \quad r_n \int_{y_1}^{\infty} p_n(r_n y)\,dy = r_n \int_{y_1}^{\infty} f_1(a_1 + y) \cdot f_2(a_2 + y) \ldots f_n(a_n + y)\,dy.$$

Nun treffen wir die noch offene Wahl von y_1, indem wir $y_1 > A + X$ setzen, wobei A und X die in Voraussetzung (b 1) bzw. (b 4) eingeführten Größen bezeichnen. Es ist dann $|a_\varkappa + y| > X$ für alle Werte des Integrationsbereiches von (11), daher nach Voraussetzung (b 4) der Absolutwert von (11) kleiner als

$$(11') \quad r_n \int_{y_1}^{\infty} |(a_1 + y)(a_2 + y) \ldots (a_n + y)|^{-\alpha}\,dx < r_n \int_{y_1}^{\infty} (y - A)^{-n\alpha}\,dy$$
$$< \sqrt{n} \cdot S \cdot \frac{(y_1 - A)^{-n\alpha + 1}}{n\alpha - 1}.$$

Da der Ausdruck rechts mit wachsendem n unter jeden Betrag sinkt, ist damit bewiesen, daß für ein gewisses n''

$$(11'') \qquad \left|\int\limits_{r_n y_1}^{\infty} p_n(u)\,du\right| < \varepsilon''' \quad \text{für} \quad n > n''$$

gilt, bei der oben eingeführten Verfügung über y_1.

Die Gleichungen (9″), (10″) und (11″) sind gleichwertig mit (8′) bis (8‴) und diese geben zusammen (8), womit der Beweis für die Behauptung II über den Grenzwert eines Produktes von Funktionen erbracht ist.

§ 2. Korollare zu den Sätzen des § 1.

1. Schwächere und stärkere Divergenz von $\Sigma s_\varkappa^2$. In unserer Beweisführung für die erste Behauptung des § 1 war es wesentlich, daß die Summe $r_n^2 = s_1^2 + s_2^2 + \ldots s_n^2$ (d. i. bis auf einen Faktor die Summe der zweiten Ableitungen der $f_\varkappa$ an den ausgezeichneten Stellen $a_\varkappa$) stärker als mit der $\frac{2}{3}$ Potenz von n ins Unendliche geht (dritte der unter (b 1) zusammengefaßten Voraussetzungen). Man kann aber auch zulassen, daß r_n^2 *schwächer wächst*, etwa mit der Potenz n^μ, $\mu < \frac{2}{3}$, wenn dafür die unter (a 2) und (b 2) gemachten Voraussetzungen verschärft werden. Denn es kommt nur darauf an — siehe Gleichung (5″) des § 1 —, daß

$$(1) \qquad \frac{1}{r_n^2}\sum_{\varkappa=1}^{n} \vartheta_\varkappa\left(\frac{u}{r_n}\right)$$

für $|u| < u_0$ mit wachsendem n gegen null geht. Nimmt man an, daß für jede einzelne der Funktionen $\vartheta_\varkappa(y)$ mit $\nu_\varkappa > 0$

$$(\mathrm{a}2') \qquad |\vartheta_\varkappa(y)| \leqq C'_\varkappa\, |y|^{\nu_\varkappa} \quad \text{für} \quad |y| < y_\varkappa$$

und für die ganze Folge der $\vartheta_1, \vartheta_2, \vartheta_3, \ldots$ mit $\nu > 0$

$$(\mathrm{b}2') \qquad |\vartheta_\varkappa(y)| \leqq C'\, |y|^{\nu} \quad \text{für} \quad |y| < y_0$$

gilt, so sieht man, daß (1) mit wachsendem n verschwindet, sobald nur $\mu > \frac{2}{\nu+2}$. Andrerseits geht der Betrag von (1), da jedes $\vartheta_\varkappa(y)$ an der Stelle $y = 0$ stetig ist, auch dann mit wachsendem n gegen null, wenn r_n^2 mit der ersten Potenz von n wächst und die Stetigkeit der $\vartheta_\varkappa(y)$ als *gleichmäßig* vorausgesetzt wird. Wir haben so den Satz

I_1. *Man kann die dritte der Voraussetzungen* (b 1) *durch*

$$(2) \qquad \frac{n^{\frac{2}{\nu+2}}}{\sum\limits_{\varkappa=1}^{n} s_\varkappa^2} \leqq \varepsilon(n), \qquad \nu > 0,$$

ersetzen, wo $\varepsilon(n)$ *eine mit* $n=\infty$ *verschwindende Größe bezeichnet, wenn gleichzeitig die Voraussetzungen* (a2) *und* (b2) *durch die obenstehenden* (a2′) *bzw.* (b2′) *ersetzt werden; gilt aber statt* (2) *die Voraussetzung, daß* r_n^2 *mit* n *wächst (also z. B. eine von null verschiedene untere Schranke der* $s_\varkappa^2$ *existiert wie im Falle der Behauptung II), so kann man* (a2) *ganz fortlassen und an Stelle von* (b2) *fordern, daß zu einem beliebig vorgegebenen* $\Theta > 0$ *ein* y_0 *existiert, so daß für alle* $\varkappa$

$$(\mathrm{b}2^*) \qquad |\vartheta_\varkappa(y)| < \Theta \text{ bei } |y| \leqq y_0$$

gilt.

Von diesem Korollar wird man insbesondere dann Anwendung machen, wenn die $f_\varkappa$ an den ausgezeichneten Stellen „gleichmäßig" regulär sind, so daß die $\vartheta_\varkappa(y)$ durch Potenzreihen mit einer gemeinschaftlichen Majoranten darstellbar sind.

2. Veränderte Bedingungen für $f_\varkappa$ im Unendlichen. Die in (a4) und (b4) gestellte Forderung für das Verhalten der $f_\varkappa$ im Unendlichen kann, wie der Gang des Beweises von II zeigt, gemildert werden. Es genügt z. B., an Stelle von (b4) vorauszusetzen, daß $\sum_{\varkappa=1}^{n} \alpha_\varkappa$ etwa mit $\sqrt{n}$ ins Unendliche wächst, wenn $\alpha_\varkappa$ die mit (a4) eingeführten Exponenten bezeichnet — s. Gl. (11′) von § 1. Andrerseits kann man den Satz II erweitern, indem man das Produkt $p_n(u)$ statt über du, über $\psi(u)\,du$ integriert, wobei $\psi(u)$ von entsprechendem Verhalten im Unendlichen sein muß. Es ergibt sich so folgender Satz

II$_2$. *Bedeutet* $\psi(u)$ *eine für alle reellen Werte von* u *definierte, in jedem endlichen Intervall integrable Funktion und tritt an Stelle der Voraussetzungen* (a4) *und* (b4) *die Bedingung, daß*

$$(3) \qquad \lim_{n=\infty} \int_X^\infty |p_n(u)\,\psi(u)\,du| = 0$$

bei hinreichend großem X, *so gilt*

$$(\mathrm{II}_2) \qquad \lim_{n=\infty} \int_a^b p_n(u)\,\psi(u)\,du = \int_a^b e^{-u^2}\psi(u)\,du$$

für alle endlichen oder unendlichen a, b. *Ist* $\psi(u)$ *beschränkt, so gilt* (II$_2$) *unter den Bedingungen des* § 1. Von diesem Korollar machen wir in § 3 zum Beweis der Behauptungen IV und V Gebrauch. Setzt man in (3) $\psi=1$, so hat man hierin die allgemeinste Form der an Stelle von (a4) und (b4) zulässigen Bedingung für den Satz II.

Eine analoge Abänderung gestattet übrigens auch die Bedingung (a3) bzw. (b3), da es hier nur auf das Kleinwerden des *Produktes* und nicht auf das der einzelnen Faktoren ankommt.

3. Erweiterung durch Einführung einer Dichtefunktion $\psi(y)$. Eine für die Anwendung in der Wahrscheinlichkeitsrechnung sehr wesentliche Erweiterung erfahren die Behauptungen des § 1, wenn, ähnlich wie in dem eben behandelten Fall, ein Faktor zu $p_n(u)$ hinzutritt, der aber jetzt nicht als Funktion von u, sondern als Funktion $\psi(y)$ von $y = u : r_n$ gegeben sei. Da, wie wir gesehen haben, der *Wert* von $p(u)$ in jedem endlichen Bereich von u, und der von $P(a, b)$, wenn die entsprechenden Voraussetzungen erfüllt sind, in jedem Bereich der a, b überhaupt, nur von dem Verhalten der $f_\varkappa$ in der Umgebung von $y = 0$ abhängt, ist es klar, daß es jetzt auch im wesentlichen nur auf $\psi(0)$ bzw. das Verhalten von ψ für kleine y ankommen kann. Die Beweise des § 1 lassen sich fast unverändert für folgenden erweiterten Satz führen:

I_3, II_3. *Es bedeute $\psi(y)$ eine im ganzen Bereich der reellen Achse definierte, beschränkte Funktion von y, die für $y = 0$ stetig ist; dann gilt für die mit (1) und (4) in § 1 definierte Funktion $p_n(u)$:*

$$(\mathrm{I}_3) \qquad \lim_{n=\infty} \left[\psi\left(\frac{u}{r_n}\right) \cdot p_n(u)\right] = \psi(0)\, e^{-u^2}$$

für jeden endlichen Bereich $|u| < u_0$ unter den Voraussetzungen des Satzes I; *und*

$$(\mathrm{II}_3) \qquad \lim_{n=\infty} \int_a^b \psi\left(\frac{u}{r_n}\right) \cdot p_n(u)\, du = \psi(0) \int_a^b e^{-u^2} du$$

für alle endlichen oder unendlichen a, b unter den Voraussetzungen des Satzes II.

Man muß nur beachten, daß aus der Stetigkeit von ψ folgt, daß $|\psi(y) - \psi(0)|$ kleiner als jeder vorgegebene Betrag gemacht werden kann, wenn nur $|y| < y_0$ und y_0 hinreichend klein gewählt wird. Nun hindert aber nichts, die in § 1, 6 mit y_0 bezeichnete Größe beliebig klein zu wählen; durch Vergrößerung von n wird das zugehörige $u_0 = r_n y_0$ hinreichend vergrößert. Außerhalb des Bereiches $|u| < u_0$ — was nur für die Behauptung II in Frage kommt — wird aber $p_n(u)$ oder wenigstens das Integral über $p_n(u)$ nach § 1, 6 so klein, daß der Wert des Faktors $\psi(u : r_n)$ überhaupt gleichgültig ist, wofern er endlich bleibt.

4. Funktionen mehrerer Variablen. Die beiden Behauptungen von § 1 lassen sich auch auf den — für die Anwendung in der Wahrscheinlichkeitsrechnung sehr wichtigen — Fall ausdehnen, in dem die $f_\varkappa$ von *mehreren* reellen Veränderlichen $x^{(1)}, x^{(2)}, \ldots, x^{(k)}$ abhängen und dabei

ähnlichen Bedingungen unterworfen sind wie den in § 1 angeführten. Die Grenzübergänge sind genau dieselben wie im Falle einer Variablen und brauchen hier nicht wiederholt zu werden. Nur einige formale Rechnungen sind erforderlich. Zunächst folgende Bemerkung:

Wenn wir in die Aussage I an Stelle von u die Variable $y = u : r_n$ einführen, so nimmt sie die Form an:

$$\lim_{n=\infty} [f_1(a_1 + y) \cdot f_2(a_2 + y) \ldots f_n(a_n + y)] = e^{-r_n^2 y^2}.$$

Diese Gleichung gilt, wie wir gezeigt haben, wenn hinzugefügt wird, daß beim Grenzübergang nicht etwa y, sondern $r_n y = u$ konstant zu halten ist. (Für konstantes y ist die Aussage trivial, da beide Seiten gegen null gehen bzw. gleich 1 sind.) Unsere Ableitung wäre aber ganz ungeändert geblieben, wenn wir statt u eine andere Variable, die sich von u durch einen *endlich bleibenden Faktor* unterscheidet, gewählt hätten. Setzen wir z. B. in der Annahme, daß r_n^2 mindestens mit n wächst (wie es für II verlangt ist) also daß $r_n^2 : n > s^2$,

$$(4) \qquad \frac{r_n}{\sqrt{n}} = h_n, \qquad z = \frac{u}{h_n} = \sqrt{n}\, y,$$

so folgt aus unserer Annahme und der zweiten Ungleichung (b1):

$$(4') \qquad s^2 < h_n^2 < S^2.$$

Es gilt somit auch

$$(4'') \qquad \lim_{n=\infty} \left[f_1\left(a_1 + \frac{z}{\sqrt{n}}\right) \cdot f_2\left(a_2 + \frac{z}{\sqrt{n}}\right) \ldots f_n\left(a_n + \frac{z}{\sqrt{n}}\right) \right] = e^{-h_n^2 z^2}$$

bei festgehaltenem z, und analog die Behauptung II.

Dies vorausgeschickt, sei jetzt $f_\varkappa$ Funktion der k Variablen $x^{(\iota)}$, oder, wie wir kürzer schreiben wollen, des k-dimensionalen Vektors $\bar{x}$; sie habe an der Stelle $\bar{x} = \bar{a}_\varkappa$d en Wert 1 und sei hier zweimal differenzierbar; die ersten Differentialquotienten seien null, die zweiten durch die reelle Matrix $\bar{s}_\varkappa$ gegeben, nämlich:

$$(5) \qquad \frac{\partial^2 f_\varkappa}{\partial x^{(\iota)} \partial x^{(\lambda)}} = -2 s_\varkappa^{(\iota,\lambda)}, \qquad (\iota, \lambda = 1, 2, \ldots, k).$$

Der früheren Bedingung, daß $s_\varkappa^2$ nicht negativ sein durfte, entspricht jetzt die Forderung, daß die quadratische Form $\sum_{\iota,\lambda} s_\varkappa^{(\iota,\lambda)} y^{(\iota)} y^{(\lambda)}$ nicht negativ wird. Geometrisch heißt das, daß $\bar{x} = \bar{a}_\varkappa$ ein *elliptischer* oder mindestens *parabolischer* Punkt der Fläche $f_\varkappa$ ist, dessen hohle Seite nach unten gerichtet ist. Die Rolle der in Gl. (3) des § 1 definierten Funktion $\vartheta_\varkappa(y)$ übernimmt jetzt die Funktion

$$(5') \qquad \frac{\log \operatorname{nat} f_\varkappa(\bar{a}_\varkappa + \bar{y}) + \sum_{\iota,\lambda} s_\varkappa^{(\iota,\lambda)} y^{(\iota)} y^{(\lambda)}}{\sum_\iota y^{(\iota)^2}} = \vartheta_\varkappa(\bar{y}),$$

wo die Summationen von 1 bis k zu erstrecken sind. Zufolge (5) nähert sich $\vartheta_\varkappa$ mit $y^{(\iota)} = 0$ $(\iota = 1, 2, \ldots, k)$ dem Grenzwert null, und wir setzen wie in (b2*) voraus, daß alle $\vartheta_\varkappa$ sich *gleichmäßig* stetig verhalten. Bei Addition der für $\varkappa = 1, 2, \ldots, n$ angeschriebenen Gleichungen (5′) entstehen die Ausdrücke

$$(5'') \qquad r_n^{(\iota,\lambda)} = \sum_{\varkappa=1}^{n} s_\varkappa^{(\iota,\lambda)},$$

von denen wir im Sinne des ersten Korollars voraussetzen wollen, daß sie sich nur durch nicht verschwindende, für alle n endliche Faktoren $h_n^{(\iota,\lambda)}$ von n unterscheiden:

$$(5''') \qquad \frac{r_n^{(\iota,\lambda)}}{n} = h_n^{(\iota,\lambda)}$$

und daß *die mit ihnen gebildete quadratische Form positiv definit sei.* Hierauf führt dieselbe Überlegung wie die in § 1, 4 angestellte analog (4″) zu dem Satz:

$$(\mathrm{I}_4) \qquad \lim_{n=\infty} \left[f_1\left(\bar{a}_1 + \frac{\bar{z}}{\sqrt{n}}\right) \cdot f_2\left(\bar{a}_2 + \frac{\bar{z}}{\sqrt{n}}\right) \cdots f_n\left(\bar{a}_n + \frac{\bar{z}}{\sqrt{n}}\right) \right] = e^{-\sum_{\iota,\lambda} h_n^{(\iota,\lambda)} z^{(\iota)} z^{(\lambda)}}$$

Genügen die $f_\varkappa$ überdies den leicht auf den Fall mehrerer Variabler übertragbaren Bedingungen (a3), (a4) und (b3), (b4), so gilt auch das Analogon zu (II):

$$(\mathrm{II}_4) \qquad \lim_{n=\infty} \int_{(Z)} f_1\left(\bar{a}_1 + \frac{\bar{z}}{\sqrt{n}}\right) \cdot f_2\left(\bar{a}_2 + \frac{\bar{z}}{\sqrt{n}}\right) \cdots f_n\left(\bar{a}_n + \frac{\bar{z}}{\sqrt{n}}\right) dZ = \int_{(Z)} e^{-\sum_{\iota,\lambda} h_n^{(\iota,\lambda)} z^{(\iota)} z^{(\lambda)}} dZ,$$

wobei die Integrale über einen endlichen oder unendlichen Teil des Raumes mit dem Element $dZ = dz^{(1)} \cdot dz^{(2)} \ldots dz^{(n)}$ zu erstrecken sind. Wir kommen so zu dem Ergebnis:

I_4, II_4. *Wenn die Funktionen $f_\varkappa$ der k reellen Variablen $x^{(\iota)}$ $(\iota = 1, 2, \ldots, k)$ den hier angedeuteten, durch Übertragung von* (a1) *bis* (a4) *und* (b1) *bis* (b4) *aus* § 1 *hervorgehenden Forderungen genügen, gelten die obenstehenden, zu* (I′) *und* (II′) *analogen Gleichungen* (I_4) *und* (II_4).

Natürlich lassen auch die im vorliegenden Paragraph erörterten Abänderungen und Erweiterungen sinngemäß Anwendung auf den Fall mehrerer Variablen zu.

5. Verschwinden der zweiten Differentialquotienten der $f_\varkappa$. An den vorstehenden Sätzen über den Grenzwert des Produktes von Funktionen

wird es jedenfalls am auffälligsten erscheinen, daß man, ausgehend von fast ganz *willkürlichen*, nur unbedeutend eingeschränkten Funktionen $f_\varkappa$, zu einem so speziellen Resultat, wie es die Limes-Funktion e^{-u^2} darstellt, gelangt. Dies Ergebnis wird verständlicher, wenn wir uns überlegen, daß wir mit unseren Aussagen lediglich das Verhalten der $f_\varkappa$ in der unmittelbaren *Umgebung eines gewöhnlichen Scheitelpunktes* zur Darstellung gebracht haben. Denken wir uns die $f_\varkappa$ reell, was sie ja in der Nähe des ausgezeichneten Punktes sind, und der Einfachheit wegen — für die jetzige Überlegung — unbeschränkt differenzierbar, so ist das Kennzeichen eines Scheitelpunktes (d. i. eines Maximums) dies, daß mindestens der erste Differentialquotient verschwindet und der erste nicht verschwindende Differentialquotient von *gerader* Ordnung und *negativ* ist. Nehmen wir nun an, es sei

$$\frac{d^{2\nu} f_\varkappa}{dx^{2\nu}} = -(2\nu)!\, s_\varkappa^{2\nu} \tag{6}$$

die niedrigste, bei irgend einem $f_\varkappa$ (an der Stelle $x = a_\varkappa$) nicht verschwindende Ableitung und es wachse die Summe

$$r_n^{2\nu} = \sum_{\varkappa=1}^{n} s_\varkappa^{2\nu} \tag{6'}$$

mit n von gleicher Ordnung ins Unendliche. Dann kann man an die Stelle der in Gleichung (3), § 1 eingeführten Funktion $\vartheta_\varkappa(y)$ die folgende treten lassen:

$$\frac{\log\operatorname{nat} f_\varkappa + (s_\varkappa\, y)^{2\nu}}{y^{2\nu}} = \vartheta_\varkappa(y). \tag{6''}$$

Man sieht, daß jetzt in ähnlicher Weise wie früher und unter analogen Voraussetzungen geschlossen werden kann, daß

$$\lim_{n=\infty} \left[f_1\left(a_1 + \frac{u}{r_n}\right)\cdot f_2\left(a_2 + \frac{u}{r_n}\right)\cdots f_n\left(a_n + \frac{u}{r_n}\right)\right] = e^{-u^{2\nu}} \tag{I$_5$}$$

und ebenso hinsichtlich des Integrales über du. Es ist noch zu beachten, daß hier r_n mit der 2ν-ten Wurzel aus n wächst, so daß, wenn wir (I$_5$) auf die obenstehende Form (4'') bringen wollten, es heißen müßte:

$$f_1\left(a_1 + \frac{z}{\sqrt[2\nu]{n}}\right)\cdot f_2\left(a_2 + \frac{z}{\sqrt[2\nu]{n}}\right)\cdots f_n\left(a_n + \frac{z}{\sqrt[2\nu]{n}}\right) = e^{-(h_n z)^{2\nu}}. \tag{6'''}$$

Man sieht an (6''') deutlich, wie der Prozeß der *Abszissendehnung* für das Resultat wesentlich ist. Die weitere Verfolgung der hier angeregten allgemeineren Fragestellung können wir unterlassen, da sie für die Wahrscheinlichkeitsrechnung bisher keine Bedeutung erlangt hat. Es soll nur zur Verdeutlichung der eigentlichen Resultate festgestellt werden:

I_5, II_5. *Man gelangt zu dem allgemeineren Ausdruck* (I_5) *für den Grenzwert eines Produktes von Funktionen, in dem höhere, gerade Potenzen im Exponenten von e auftreten, wenn die ersten nicht verschwindenden Differentialquotienten in den Scheitelpunkten der* $f_\varkappa$ *von höherer als zweiter Ordnung sind.*

§ 3.

Sätze über den Grenzwert eines Integralproduktes von Funktionen.

1. Bezeichnungen und Definitionen. Es seien $V_1(x), V_2(x), V_3(x), \ldots$ reelle, *nirgends abnehmende*, im ganzen Bereich der reellen Variablen x definierte Funktionen, die für $x = -\infty$ den Wert *null*, für $x = \infty$ den Wert *eins* besitzen und an etwaigen Sprungstellen stets den Wert ihres *oberen* (rechten) Limes annehmen. Wir wollen derartige Funktionen, um nicht immer ihre Eigenschaften wiederholen zu müssen — im Hinblick auf ihre Bedeutung in der Wahrscheinlichkeitsrechnung — kurz als „*Verteilungen*" bezeichnen.

Da $V(x)$ monoton ist, existiert für jede beschränkte, stetige Funktion $\varphi(x)$ das „*Stieltjessche*" *Integral*

$$(*) \qquad \int_a^b \varphi(x)\, dV(x),$$

das in bekannter Weise als Grenzwert einer Summe über Ausdrücke von der Form $\varphi(\xi_\iota)[(V(x_{\iota+1}) - V(x_\iota)]$ definiert wird, wobei $a \leqq x_\iota < \xi_\iota < x_{\iota+1} \leqq b$, $\iota = 1, 2, 3 \ldots$ eine Intervall-Teilung des Integrationsbereiches a, b bedeutet, deren größte Intervall-Länge beim Grenzübergang gegen null geht. Wir wollen das Symbol (*) auch auf solche $\varphi(x)$ anwenden, die beschränkt, nicht notwendig stetig, aber jedenfalls *monoton* (oder wenigstens von beschränkter Schwankung) sind und hierzu festsetzen, daß unter dem Wert von (*) stets der *obere* Limes der Summe gemeint sei.

Die Zeichen $-\infty$ und $+\infty$ als untere bzw. obere Grenze eines bestimmten Integrales *lassen wir hinfort weg.* Es bedeutet also $\int$ schlechthin die Integration über die ganze reelle Achse, $\int^x$ die Integration von $-\infty$ bis x usw.[3]).

[3]) Die hier eingeführte Bezeichnungsweise $\int \varphi(x)\, dV(x)$ ist den in der Wahrscheinlichkeitsrechnung gebräuchlichen, wie $\mathfrak{D}[\varphi(x)]$ (Durchschnitt von φ, Bruns) oder $m \cdot H \cdot \varphi(x)$ (mathemat. Hoffnung von φ, Markoff) dadurch überlegen, daß die Verteilung, bezüglich deren der Durchschnitt gebildet wird, ausdrücklich angeführt erscheint. An Einfachheit dürfte sie auch nicht nachstehen und vor allem sich der sonstigen Schreibweise der Analysis besser anpassen.

Als *„Integralprodukt" der n Verteilungen* $V_1(x), V_2(x), \ldots, V_n(x)$ definieren wir das $(n-1)$fache, im Stieltjesschen Sinne und gemäß der eben gemachten Festsetzung zu bildende Integral

$$(1)\quad W_n(u) = \int\int\ldots\int V_n(x-x_1-x_2-\ldots-x_{n-1})\,dV_{n-1}(x_{n-1})\,dV_{n-2}(x_{n-2})\ldots dV_1(x_1),$$

wobei u in einer bestimmten, noch näher anzugebenden Weise *linear* mit *x wächst.* Die Existenz des Integrales im Hinblick auf die unendlichen Grenzen ist dadurch sichergestellt, daß $|V_n| \leqq 1$ und $\int dV_\varkappa = 1$ für jedes $\varkappa$.

Man sieht sofort, daß das *Integralprodukt wieder eine Verteilung* ist: Es verschwindet für $u = -\infty$, weil $V_n(-\infty) = 0$, es kann mit wachsendem u nicht abnehmen, da der einzige von u abhängige Faktor des Integranden, nämlich V_n, nicht abnimmt, während die anderen positiv sind; es geht für $u = \infty$ in das Produkt $\int dV_1(x_1) \cdot \int dV_2(x_2) \ldots \int dV_{n-1}(x_{n-1})$ über, dessen Faktoren sämtlich gleich 1 sind; und genügt schließlich an etwaigen Unstetigkeitsstellen der ausgesprochenen Bedingung, da das Integral als die obere Limes-Funktion definiert wurde.

Gegenstand unserer Untersuchung ist vor allem der *Grenzwert*

$$(1')\qquad W(u) = \lim_{n=\infty} W_n(u)$$

für eine unendliche Folge von Verteilungen, die gewissen Einschränkungen unterworfen werden wird.

Haben sämtliche $V_\varkappa(x)$ — was wir jedoch nicht allgemein voraussetzen wollen — eine *beschränkte Ableitung* $v_\varkappa(x)$, so hat auch $W_n(u)$ eine solche, und man erhält durch Differentiation von (1):

$$(2)\quad \begin{aligned} w_n(u) = \text{konst.} \int\int\ldots\int v_n(x-x_1-x_2-\ldots-x_{n-1})\,v_{n-1}(x_{n-1})\,v_{n-2}(x_{n-2})\ldots \\ v_1(x_1)\,dx_{n-1}\,dx_{n-2}\ldots dx_1. \end{aligned}$$

Die Konstante vor dem Integralzeichen ist der Koeffizient von u in dem linearen Ausdruck für x. Die nicht-negativen Funktionen $v_\varkappa(x)$ bzw. $w_n(u)$ bezeichnen wir als die zu $V_\varkappa$ bzw. W_n gehörigen *„Verteilungsdichten"*.

Haben die $V_\varkappa(x)$ — ein anderer Grenzfall — *nur für* $x = 0, 1, 2, \ldots$ *Wachstumstellen,* und zwar mit den Sprunggrößen $\mathfrak{v}_\varkappa(0), \mathfrak{v}_\varkappa(1), \mathfrak{v}_\varkappa(2), \ldots$, so haben auch die Integralprodukte W_n, zunächst als Funktion von x betrachtet, dieselbe Eigenschaft. Die Sprunggrößen $\mathfrak{w}_n(u_0), \mathfrak{w}_n(u_1), \mathfrak{w}_n(u_2), \ldots$, wobei die $u_0, u_1, u_2, \ldots$ vermöge der linearen Beziehung den $x = 0, 1, 2, \ldots$ entsprechen, sind gegeben durch $(n-1)$fache Summen:

$$(3)\qquad \mathfrak{w}_n(u) = \sum \mathfrak{v}_1(x_1) \cdot \mathfrak{v}_2(x_2) \ldots \mathfrak{v}_n(x_n),$$

zu erstrecken über alle ganzzahligen $x_1, x_2, \ldots, x_n$ mit der Summe x, unter Beachtung der Zuordnung zwischen x und u. Nur unwesentlich allgemeiner ist der später zu betrachtende Fall, in dem von den Wachstumstellen der $V_\varkappa$ vorausgesetzt wird, daß ihre *Abstände ganzzahlige Vielfache einer Einheit* $\mathfrak{e}$ sind.

In den beiden Sonderfällen (2) und (3) fragen wir nach den Grenzwerten

$$(2') \qquad w(u) = \lim_{n=\infty} w_n(u)$$

und

$$(3') \qquad \mathfrak{w}(u) = \lim_{n=\infty} [\text{konst. } \mathfrak{w}_n(u)],$$

wobei konst. wieder den in (2) auftretenden Wert hat. In (1′) und (2′) sind beim Grenzübergang die Werte von u festzuhalten; in (3′) wird der Grenzübergang so vollzogen, daß das bei jedem n nur diskreter Werte fähige u sich einer festen Grenze nähert. Daß dies möglich ist, wird sich später (§ 3, 3) zeigen.

Man überzeugt sich leicht, daß in allen Fällen die Faktoren des Integralproduktes *vertauschbar* sind. Dagegen genügt die Definition (1) dem assoziativen Gesetz nur bei besonderen Festsetzungen über die Beziehung zwischen x und u, die für uns nicht in Betracht kommen (vgl. § 1, 1).

Für $n = 1$ ergänzen wir die Definition durch die Festsetzung

$$(1'') \qquad W_1(u) = V_1(x),$$

aus der in den Fällen (2) und (3) notwendig folgt:

$$(2'') \qquad w_1(u) = \text{konst. } v_1(x),$$

$$(3'') \qquad \mathfrak{w}_1(u) = \mathfrak{v}_1(x).$$

2. Voraussetzungen. a) Von *jeder einzelnen* der Verteilungen $V_\varkappa(x)$ setzen wir voraus:

1. Es existiere eine positive Zahl $c_\varkappa^2$, für die das Integral

$$(\text{a}\,1) \qquad C = \int e^{c_\varkappa^2 x^2} d\,V_\varkappa(x)$$

einen endlichen Wert besitzt. Man erkennt leicht, daß es keine weitere Einschränkung bedeutet, wenn für C ein beliebiger Wert > 1 vorgeschrieben wird.

Aus (a 1) folgt, daß jedes Integral von der Form $\int \varphi(x)\,d\,V_\varkappa(x)$ existiert, wenn $\varphi(x)$ beschränkt ist oder im Unendlichen nicht stärker wächst, als etwa ein Polynom in x. Insbesondere erkennt man, daß jede (a 1) genügende Verteilung einen endlichen „*Mittelwert*“ $a_\varkappa$ und eine endliche „*Streuung*“ $s_\varkappa^2$ besitzt:

$$(4) \qquad a_\varkappa = \int x\,d\,V_\varkappa(x), \qquad s_\varkappa^2 = \int (x - a_\varkappa)^2\,d\,V_\varkappa(x).$$

2. Die *Streuung* $s_\varkappa^2$ der Verteilung $V_\varkappa(x)$ sei *von null verschieden:*

$$(a\,2)\qquad s_\varkappa^2 > 0.$$

Mit anderen Worten heißt das: Die Verteilung $V_\varkappa$ habe mehr als eine Wachstumsstelle.

b) Von der *Gesamtheit* der Verteilungen $V_\varkappa(x)$ setzen wir voraus:

1. Es existiere zu einem festen $C > 1$ eine positive Zahl c^2, so daß *unabhängig von* $\varkappa$:

$$(b\,1)\qquad \int e^{\frac{c^2(x-a_\varkappa)^2}{2 s_\varkappa^2}}\, d V_\varkappa(x) < C.$$

Dies erfordert nur (wenn die Folge der $a_\varkappa$ beschränkt ist), daß es für die in a) definierten Zahlen $c_\varkappa$ und $s_\varkappa$ ein von null verschiedenes kleinstes Produkt $c_\varkappa s_\varkappa$ gibt.

2. Die Reihe der in Gl. (4) definierten *Mittelwerte* $a_\varkappa$ und der *Streuungen* $s_\varkappa$ sei *beschränkt*, doch wachse die Summe der letzteren *stärker als mit der* $^2/_3$ *Potenz von* n ins Unendliche:

$$(b\,2)\qquad |a_\varkappa| < A, \quad s_\varkappa^2 < S^2, \quad \frac{n^{\frac{2}{3}}}{\sum\limits_{\varkappa=1}^{n} s_\varkappa^2} \leqq \varepsilon(n),$$

wo A und S beliebige endliche Beträge und $\varepsilon(n)$ eine mit $n = \infty$ verschwindende positive Funktion von n bedeutet.

Man erkennt wieder wie in § 1, daß die Voraussetzungen b) nur eine gewisse *Gleichmäßigkeit* in der Erfüllung der Voraussetzungen a) fordern. Sind alle $V_\varkappa$ untereinander gleich, so sind die b) in den a) enthalten.

Die Bedeutung von (b 1) liegt vornehmlich darin, daß dadurch die *Folge der mit dem Anfangspunkt* $a_\varkappa$ *und dem Maßstabfaktor* $s_\varkappa$ *gebildeten „m-ten Momente" der Verteilungen* $V_\varkappa$ *beschränkt* wird. Rechnet man nämlich das Maximum der Funktion $z^m e^{-c^2 z^2}$, so findet man leicht

$$(5)\qquad |z^m e^{-c^2 z^2}| \leqq \left(\frac{m}{2 c^2 e}\right)^{\frac{m}{2}}.$$

Daraus folgt, wenn

$$(5')\qquad z = \frac{x - a_\varkappa}{s_\varkappa \sqrt{2}}$$

gesetzt wird, mit (b 1) für den Betrag des auf z bezogenen m-ten Momentes von $V_\varkappa(x)$:

$$(6)\qquad \left|\int z^m\, d V_\varkappa(x)\right| < \int |z^m e^{-c^2 z^2}| \cdot e^{c^2 z^2}\, d V_\varkappa(x) < C \cdot \left(\frac{m}{2 c^2 e}\right)^{\frac{m}{2}},$$

unabhängig von $\varkappa$.

3. Behauptungen. III. *Genügen die Verteilungen* $V_1, V_2, V_3, \ldots$ *den eben angeführten Voraussetzungen* (a 1), (a 2) *und* (b 1), (b 2) *und setzt man*

$$(7) \qquad x = r_n u + b_n, \qquad u = \frac{x - b_n}{r_n}$$

mit

$$(7') \qquad b_n = a_1 + a_2 + \ldots + a_n, \qquad r_n^2 = 2(s_1^2 + s_2^2 + \ldots + s_n^2),$$

wo $a_\varkappa$ *und* $s_\varkappa^2$ *die in* (4) *definierten Mittelwerte und Streuungen der* $V_\varkappa$ *bezeichnen, so gilt für die mit* (1) *und* (1') *definierte Verteilung* $W(u)$:

$$(\text{III}) \qquad W(u) = \frac{1}{\sqrt{\pi}} \int^{u} e^{-x^2} dx \equiv \Phi(u)$$

im ganzen Bereich der u-Achse. Oder ausführlicher: *Es läßt sich stets ein* n_0 *so finden, daß für* $n > n_0$:

$$(\text{III}') \qquad \left| \int\int \ldots \int V_n(r_n u + b_n - x_1 - x_2 - \ldots - x_{n-1})\, dV_{n-1}(x_{n-1})\, dV_{n-2}(x_{n-2}) \ldots dV_1(x_1) - \Phi(u) \right| < \varepsilon$$

für jedes u, *bei beliebig vorgegebenem* ε.

IV. *Besitzen die Verteilungen* $V_1, V_2, V_3, \ldots$ *überdies*[4]) *durchweg beschränkte Ableitungen* $v_1, v_2, v_3, \ldots$ *von gleichmäßig beschränkter Variation* (also so, daß die Gesamtschwankungen $\sigma_1, \sigma_2, \sigma_3, \ldots$ von $v_1, v_2, v_3, \ldots$ eine obere Schranke σ haben), *so gilt außer Gleichung* (III) *auch für die mit* (2) *und* (2') *definierte Verteilungsdichte* $w(u)$ *in jedem endlichen Bereich von* u:

$$(\text{IV}) \qquad w(u) = \frac{1}{\sqrt{\pi}} e^{-u^2} \equiv \varphi(u)$$

an allen Stetigkeitsstellen von $w(u)$; *und an allen anderen Stellen:*

$$(\text{IV}') \qquad \tfrac{1}{2}[w(u+0) + w(u-0)] = \frac{1}{\sqrt{\pi}} e^{-u^2} \equiv \varphi(u).$$

Ausführlicher lautet (IV): *Es läßt sich zu jedem vorgegebenen* u_0 *und* ε *ein* n_0 *so finden, daß für* $n > n_0$:

$$(\text{IV}'') \qquad \left| r_n \int\int \ldots \int v_n(r_n u + b_n - x_1 - x_2 - \ldots - x_{n-1})\, v_{n-1}(x_{n-1})\, v_{n-2}(x_{n-2}) \ldots v_1(x_1)\, dx_{n-1}\, dx_{n-2} \ldots dx_1 - \varphi(u) \right| < \varepsilon$$

für jede Stelle $|u| \leqq u_0$, *an der das Integral stetig ist.*

V. *Besitzen die den Bedingungen* (a 1) (a 2) *und* (b 1) (b 2) *genügenden Verteilungen*[5]) $V_1, V_2, V_3, \ldots$ *nur diskret liegende Wachstumsstellen mit*

[4]) Durch die Korollare in § 6 werden die Voraussetzungen für die Fälle der Behauptungen IV und V wesentlich eingeschränkt werden.

[5]) Vgl. die vorangehende Fußnote.

den Sprunggrößen $\mathfrak{v}_1(x)$, $\mathfrak{v}_2(x)$, $\mathfrak{v}_3(x)$, ..., *deren Abstände (Differenzen der* x*) nur ganzzahlige Vielfache einer Einheit* $\mathfrak{e}$ *sind, wobei der Abstand* $\mathfrak{e}$ *selbst mindestens einmal bei jedem* $V_\varkappa$ *und zwar zwischen zwei endlichen, mit* $\varkappa = \infty$ *nicht gegen null gehenden Sprunggrößen vorkommt, so gilt außer Gleichung* (III) *auch für die mit* (3) *und* (3′) *definierte Zahlenfolge* $\mathfrak{w}(u)$:

$$\text{(V)} \qquad \mathfrak{w}(u) = \frac{1}{\sqrt{\pi}} e^{-u^2} \equiv \varphi(u),$$

oder ausführlicher: *Es läßt sich zu einem beliebig vorgegebenen* ε *ein* n_0 *finden, so daß für* $n > n_0$

$$\text{(V}'\text{)} \qquad \Big| r_n \sum_{(x_\varkappa)} \mathfrak{v}_1(x_1)\mathfrak{v}_2(x_2)\ldots\mathfrak{v}_n(x_n) - \varphi(u) \Big| < \varepsilon$$

$$\text{für } u = \frac{m_n \mathfrak{e} - b_n}{r_n}, \qquad (m_n = 0, \ \pm 1, \ \pm 2, \ldots),$$

wobei die Summation über alle Wertverbindungen $x_\varkappa = k_\varkappa \mathfrak{e}$ $(k_\varkappa = 0, \pm 1, \pm 2, \ldots)$ *zu erstrecken ist, für welche* $k_1 + k_2 + \ldots + k_n = m_n$*, und die* m_n *so zu wählen sind, daß* u *mit* $n = \infty$ *einer festen Grenze zustrebt.*

Da r_n^2 nach (7′) zufolge der dritten der Ungleichungen (b 1) mit n ins Unendliche wächst, kann man, wenn nur n genügend groß gewählt wird, immer ein ganzzahliges m_n so finden, daß der Unterschied zwischen einem beliebig vorgegebenen u und dem Quotienten $(m_n \mathfrak{e} - b_n) : r_n$ beliebig klein wird.

Die Bedeutung der Festsetzung (7) und (7′) ist analog (4′) in § 1 die, daß jetzt der *Mittelwert aller* $W_n(u)$ *null*, die *Streuung aller* $W_n(u)$ *gleich einhalb* wird. Denn führt man in (1) für x den Wert aus (7) ein und ersetzt bei Bildung von $\int u\, dW_n(u)$ die Integrations-Variable u durch die neue x_n:

$$(8) \quad x_n = r_n u + b_n - x_1 - x_2 - x_3 - \ldots - x_{n-1}, \quad u = \frac{1}{r_n} \sum_{\varkappa=1}^{n} (x_\varkappa - a_\varkappa),$$

so erhält man:

$$(9) \quad a_n' = \int u\, dW_n(u) = \frac{1}{r_n} \int\!\!\int \ldots \int \sum_{\varkappa=1}^{n} (x_\varkappa - a_\varkappa)\, dV_1(x_1) \cdot dV_2(x_2) \ldots dV_n(x_n)$$

$$= \frac{1}{r_n} \int (x_1 - a_1)\, dV_1(x_1) \int dV_2 \cdot \int dV_3 \ldots \int dV_n$$

$$+ \frac{1}{r_n} \int (x_2 - a_2)\, dV_2(x_2) \cdot \int dV_1 \cdot \int dV_3 \cdot \int dV_4 \ldots \int dV_n + \ldots = 0,$$

da in jedem der n Summanden rechts der erste Faktor vermöge (4) null ist. Analog behandelt man das n-fache Integral, das die Streuung von $W_n(u)$ liefert:

$$(10') \quad s_n'^2 = \int u^2 \, dW_n(u) = \frac{1}{r_n^2} \int\int \cdots$$
$$\int \Big[\sum_{\varkappa=1}^{n} (x_\varkappa - a_\varkappa)^2 + \sum_{\varkappa \neq \lambda = 1}^{n} (x_\varkappa - a_\varkappa)(x_\lambda - a_\lambda) \Big] dV_1(x_1)\, dV_2(x_2) \ldots dV_n(x_n).$$

Die von der Doppelsumme mit $\varkappa \neq \lambda$ herrührenden Glieder sind wieder null, die von der einfachen Summe stammenden einzeln gleich $s_\varkappa^2$ nach der zweiten Gleichung (4); demnach hat man mit Rücksicht auf (7'):

$$(10) \qquad s_n'^2 = \frac{1}{r_n^2} \sum_{\varkappa=1}^{n} s_\varkappa^2 = \tfrac{1}{2}.$$

Über das gegenseitige Verhältnis der Behauptungen III bis V ist zu sagen, daß zwar IV und V mehr aussagen als III (weil jedesmal auch Gl. (III) gilt), aber unter wesentlich einschränkenderen Voraussetzungen über die Verteilungen $V_\varkappa(x)$. Es erweist sich im folgenden als verhältnismäßig einfach, IV und V durch Zurückführung auf die früheren Sätze I und II zu beweisen, während der Beweis von Satz III umfassendere Vorbereitungen erfordert.

4. Zurückführung der Integralprodukte auf einfache Produkte. Die Bildung der hier als „*Integralprodukte*" bezeichneten Verteilungen läßt sich, wie schon Laplace im wesentlichen erkannt hat, auf die Bildung *einfacher Produkte* von Funktionen zurückführen. Man sieht zunächst in dem durch (3) gekennzeichneten Fall von Verteilungen mit nur ganzzahligen Wachstumstellen, daß $\mathfrak{w}_n(u)$ nichts anderes ist, als der Koeffizient von t^x in der Entwicklung des Produktes von Polynomen

$$\sum_z \mathfrak{v}_1(z) \cdot t^z \cdot \sum_z \mathfrak{v}_2(z) \cdot t^z \ldots \sum_z \mathfrak{v}_n(z) \cdot t^z,$$

wo die Summen über alle $z = 0, 1, 2, \ldots$ zu erstrecken sind und zwischen u und x die Beziehung (7) besteht. Im allgemeinen Falle beliebiger Verteilungen werden an Stelle dieser Summen Stieltjessche Integrale treten müssen, wobei noch eine gewisse Freiheit in der Verfügung über t und etwaige Faktoren gelegen ist.

Wir definieren zunächst (in § 4 wird eine andere „erzeugende Funktion" verwendet werden) als „*komplexe Adjunkte*" $f_\varkappa(x)$ *einer Verteilung* $V_\varkappa(x)$ *mit dem Mittelwert* $a_\varkappa$ die Funktion:

$$(11) \qquad f_\varkappa(x) = \int e^{(x-a_\varkappa)(z-a_\varkappa)i} \, dV_\varkappa(z).$$

Hier ist also, wenn von der Konstanten $a_\varkappa$ abgesehen wird, e^{xi} für t gesetzt. Das Integral existiert für jeden endlichen Wert von x und hat einen absoluten Betrag ≤ 1. Durch Differentiation findet man

$$(11') \qquad f_\varkappa^{(m)}(x) = i^m \int (z - a_\varkappa)^m \, e^{(x-a_\varkappa)(z-a_\varkappa)i} \, dV_\varkappa(z).$$

Daher ist für $x = a_\varkappa$, mit Rücksicht auf die Definition von $a_\varkappa$ und $s_\varkappa^2$ in Gl. (4):

$$(11'') \qquad f_\varkappa(a_\varkappa) = 1, \qquad f'_\varkappa(a_\varkappa) = 0, \qquad f''_\varkappa(a_\varkappa) = -s_\varkappa^2.$$

Die komplexe Adjunkte des n-ten Integralproduktes bezeichnen wir mit $p_n(x)$; sie ist nach (11) vermöge (9) und (10) definiert durch

$$(11\text{a}) \qquad p_n(x) = \int e^{xui}\, d\,W_n(u)$$

und hier gilt

$$(11''\text{a}) \qquad p_n(0) = 1, \qquad p'_n(0) = 0, \qquad p''_n(0) = -\tfrac{1}{2}.$$

Um das grundlegende *Multiplikationsgesetz* der komplexen Adjunkten zu erhalten, müssen wir in (11a) den Wert von $d\,W_n(u)$ aus der Definitionsgleichung (1) des Integralproduktes einsetzen. Verwendet man wieder die Substitution (8), so entsteht:

$$p_n(x) = \int\int \ldots \int e^{\frac{xi}{r_n} \sum\limits_{\varkappa=1}^{n} (x_\varkappa - a_\varkappa)} d\,V_1(x_1)\, d\,V_2(x_2) \ldots d\,V_n(x_n).$$

Hier lassen sich die Integrations-Variablen $x_1, x_2, \ldots, x_n$ trennen und man erhält für $p_n(x)$ ein Produkt von n Faktoren der Form

$$\int e^{(x_\varkappa - a_\varkappa)\frac{x}{r_n} i}\, d\,V_\varkappa(x_\varkappa) = f_\varkappa\left(a_\varkappa + \frac{x}{r_n}\right).$$

Demnach wird:

$$(12) \qquad p_n(u) = f_1\left(a_1 + \frac{u}{r_n}\right) \cdot f_2\left(a_2 + \frac{u}{r_n}\right) \ldots f_n\left(a_n + \frac{u}{r_n}\right);$$

in Worten:

Die komplexen Adjunkten der Verteilungen $V_\varkappa$ einerseits und die ihrer Integralprodukte W_n andererseits stehen in wesentlich derselben Beziehung zueinander wie die in § 1 betrachteten Funktionen $f_\varkappa$ und ihre Produkte p_n.

Nur in der Normierung liegt ein kleiner Unterschied: In § 1 bedeutete $-r_n^2$ die *halbe* Summe der zweiten Ableitungen der $f_\varkappa$ an den Stellen $x = a_\varkappa$, jetzt ist es nach (7′) und (11″) die *doppelte* Summe. Daraus folgt, daß der Limes von $p_n(u)$, falls wir die Erfüllung der Voraussetzungen des § 1 durch die komplexen Adjunkten der $V_\varkappa$ nachweisen können, den Wert

$$(13) \qquad p(u) = \lim_{n=\infty} p_n(u) = e^{-\frac{u^2}{4}}$$

annehmen muß.

Tatsächlich ist leicht zu sehen, daß unsere $f_\varkappa$ dem *ersten Teil* der Voraussetzungen des § 1, nämlich (a1) und (a2), sowie (b1) und (b2) genügen. Denn (a1) ist durch (11″) erledigt, während (b1) genau durch die Voraussetzung (b2) über die $V_\varkappa$ im vorliegenden Paragraph gedeckt wird. Es bleibt also nur noch (a2) und (b2) nachzuweisen, nämlich

$$(14) \qquad \vartheta_\varkappa(y)| = \left| \frac{\log \operatorname{nat} f_\varkappa(a_\varkappa + y) + \frac{1}{2} s_\varkappa^2 y^2}{y^2} \right| \leqq C' \, |y| \quad \text{für} \quad |y| < y_0.$$

Nun verhält sich $f_\varkappa$ und $\log \operatorname{nat} f_\varkappa$ an der Stelle $y = 0$ vollkommen regulär: der Zähler von $\vartheta_\varkappa(y)$ verschwindet samt seinen beiden ersten Ableitungen für $y = 0$, die dritte Ableitung ist vermöge (11′) und (11″):

$$\frac{d^3}{dy^3}\left[\log \operatorname{nat} f_\varkappa(a_\varkappa + y) + \tfrac{1}{2} s_\varkappa^2 y^2\right]_{y=0} = \frac{d^3 f_\varkappa(a_\varkappa)}{dy^3} = -i \int (x - a_\varkappa)^3 \, d V_\varkappa(x).$$

Dieser Ausdruck besitzt nach (6) und der zweiten Ungleichung (b2) die von $\varkappa$ unabhängige obere Grenze

$$C \cdot S^3 \left(\frac{3}{c^2 e}\right)^{\frac{3}{2}},$$

wo C und c^2 die mit (b1) eingeführten Konstanten bedeuten. Es ist somit für hinreichend kleine $|y|$:

$$(14') \qquad |\log \operatorname{nat} f_\varkappa(a_\varkappa + y) + \tfrac{1}{2} s_\varkappa^2 y^2| \leqq C' \, |y|^3 \quad \text{für} \quad C' > C S^3 \left(\frac{3}{c^2 e}\right)^{\frac{3}{2}},$$

womit (14) bewiesen ist. *Zufolge Satz* I *gilt sonach* (13) *in jedem endlichen Bereich von* u.

Setzt man in (11a) für $W_n(z)$ die Gaußsche Funktion $\Phi(z)$ ein, so ergibt sich die zugehörige komplexe Adjunkte

$$(15) \quad \int e^{xzi} d\Phi(z) = \frac{1}{\sqrt{\pi}} \int e^{-z^2 + xzi} \, dz = \frac{e^{-\frac{x^2}{4}}}{\sqrt{\pi}} \int e^{-\left(z - \frac{xi}{2}\right)^2} dz = e^{-\frac{x^2}{4}}.$$

Danach wird man wohl vermuten können, daß, wenn $p_n(u)$ nach (13) gegen $e^{-\frac{u^2}{4}}$ konvergiert, unter gewissen Umständen $W_n(z)$ gegen $\Phi(z)$ konvergiert. In den Spezialfällen der Sätze IV und V läßt sich dies, wie wir sehen werden, durch *nähere Betrachtung der* $f_\varkappa$ sowie durch Aufstellung einer *Umkehrformel*, die eine Verteilung aus ihrer komplexen Adjunkten zu finden gestattet, auch wirklich nachweisen.

5. **Beweis der Behauptung V.** Wir wollen annehmen, was offenbar keine wesentliche Einschränkung gegenüber den Voraussetzungen des Satzes V ist, daß Wachstumstellen der $V_\varkappa(x)$ nur bei $x = 0, \pm 1, \pm 2, \ldots$ liegen. Die komplexe Adjunkte ist sodann nach (11) durch

$$(16) \qquad f_\varkappa(a_\varkappa + y) = \sum_x \mathfrak{v}_\varkappa(x) \, e^{(x - a_\varkappa) y i}$$

definiert, wobei die Summation über alle ganzzahligen x zu erstrecken ist. Multipliziert man (16) beiderseits mit $e^{-(z - a_\varkappa) y i}$, wo z wieder eine ganze

Zahl, und integriert von $-\pi$ bis $+\pi$ nach y, so fallen rechts alle Summanden aus, für die $x \neq z$, und es bleibt nur für $x = z$ das Integral $2\pi\mathfrak{v}_\varkappa(z)$. Somit besteht die *Umkehrformel*:

$$(16') \qquad \mathfrak{v}_\varkappa(z) = \frac{1}{2\pi}\int\limits_{-\pi}^{\pi} f_\varkappa(a_\varkappa + y)\, e^{-(z-a_\varkappa)yi}\, dy, \qquad z = 0, \pm 1, \pm 2, \ldots.$$

Man kann dies noch etwas anders schreiben, indem man an Stelle von $f_\varkappa$ eine neue Funktion $f_\varkappa^*$ durch

$$(16'') \qquad \begin{aligned} f_\varkappa^*(a_\varkappa + y) &= f_\varkappa(a_\varkappa + y) \quad \text{für} \quad -\pi \leqq y \leqq \pi, \\ &= 0 \qquad\qquad\qquad\quad ,, \quad\; |y| > \pi \end{aligned}$$

einführt, womit (16′) lautet:

$$(16''') \qquad \mathfrak{v}_\varkappa(z) = \frac{1}{2\pi}\int f_\varkappa^*(a_\varkappa + y)\, e^{-(z-a_\varkappa)yi}\, dy.$$

Für die komplexe Adjunkte $p_n(y)$ des n-ten Integralproduktes hat man nach (11a):

$$(17) \qquad p_n(y) = \sum_u \mathfrak{w}_n(u)\, e^{uyi} = \sum_x \mathfrak{w}_n\left(\frac{x - b_n}{r_n}\right) e^{(x-b_n)\frac{y}{r_n}i},$$

wobei die Summation wieder über alle ganzzahligen x auszudehnen ist. Behandelt man (17) analog wie oben (16), multipliziert also beiderseits mit $e^{-(z-b_n)\frac{y}{r_n}i}$, so muß man von $-r_n\pi$ bis $+r_n\pi$ nach y integrieren, damit die Integrale rechts für $x \neq z$ verschwinden. Die Umkehrformel lautet mithin:

$$(17') \quad \mathfrak{w}_n(u) = \frac{1}{2 r_n \pi}\int\limits_{-r_n\pi}^{r_n\pi} p_n(y)\, e^{-uyi}\, dy, \quad u = -\frac{b_n}{r_n}, \; -\frac{b_n}{r_n} \pm \frac{1}{r_n}, \; -\frac{b_n}{r_n} \pm \frac{2}{r_n} \ldots,$$

sobald wieder u an Stelle von z substituiert wird. Setzt man hier für p_n den Wert aus (12) ein und berücksichtigt (16″), so erhält man:

$$(17'') \quad r_n \mathfrak{w}_n(u) = \frac{1}{2\pi}\int f_1^*\left(a_1 + \frac{y}{r_n}\right) \cdot f_2^*\left(a_2 + \frac{y}{r_n}\right) \cdots f_n^*\left(a_n + \frac{y}{r_n}\right) e^{-uyi}\, dy.$$

Die Funktionen $f_\varkappa^*$ erfüllen ebenso wie die $f_\varkappa$ (s. den vorangehenden Abschnitt) die Voraussetzungen (a 1), (a 2) und (b 1), (b 2) des § 1, die sich ja nur auf das Verhalten in der Umgebung von $y = 0$ erstrecken. Darüber hinaus genügen aber die $f_\varkappa^*$ auch den Bedingungen (a 4) und (b 4), da sie nach (16″) außerhalb eines endlichen Bereiches überhaupt verschwinden. Für das Verhalten gegenüber (a 3) und (b 3) sowie für die Existenz einer unteren Schranke der $s_\varkappa^2$ ist nun der in Satz V gemachte Vorbehalt entscheidend, wonach für jedes $\varkappa$ wenigstens ein Paar aufeinanderfolgender ganzer Zahlen $m_\varkappa$ und $m_\varkappa + 1$ existieren muß, für welche $\mathfrak{v}_\varkappa$ nicht verschwindet.

Da nämlich stets $\sum_x \mathfrak{v}_\varkappa(x) = 1$, kann $f_\varkappa$ nur dann den Betrag 1 erreichen, wenn alle Winkelargumente $(x - a_\varkappa)y$ in (16) bis auf ganze Vielfache von 2π einander gleich sind. Nun ist aber für zwei aufeinanderfolgende x der Winkelunterschied gleich y, und da (16) nur bis $|y| = \pi$ gilt, so sieht man, daß $|f_\varkappa^*|$ außerhalb der Stelle $y = 0$ niemals 1 werden kann, wie es (a3) verlangt. Ist $\mathfrak{v}$ der Betrag, der voraussetzungsgemäß von $\mathfrak{v}_\varkappa(m_\varkappa)$ und $\mathfrak{v}_\varkappa(m_\varkappa + 1)$ nicht unterschritten wird, so hat man

$$|f_\varkappa^*(a_\varkappa + y)| < 1 - 2\,\mathfrak{v}\,(1 - \cos y),$$

womit (a3) *und* (b3) *von* § 1 *Genüge getan ist.* Aus der Definition der Streuung folgt überdies:

$$s_\varkappa^2 \geqq \tfrac{1}{2}\mathfrak{v},$$

weil die beiden Stellen $x = m_\varkappa$ und $x = m_{\varkappa+1}$ zur Streuung mindestens den Beitrag $\mathfrak{v}[\xi^2 + (\xi - 1)^2]$ liefern, dessen Minimum für $\xi = \frac{1}{2}$ eben $\frac{1}{2}\mathfrak{v}$ ist.

Mithin erfüllen die $f_\varkappa^*$ *sämtliche Voraussetzungen des* § 1. Schreibt man (17″) in der Form

$$(17''')\quad \mathfrak{w}(u) = \lim_{n=\infty} r_n \mathfrak{w}_n(u) = \frac{1}{2\pi}\int p_n(y)\,\psi(y)\,dy \quad \text{mit} \quad \psi(y) = e^{-uyi},$$

wo $|\psi| = 1$, so folgt aus dem Korollar II_2 in § 2:

$$(18)\quad \mathfrak{w}(u) = \frac{1}{2\pi}\int e^{-\frac{y^2}{4} - uyi}\,dy = \frac{e^{-u^2}}{\pi}\int e^{-\left(\frac{y}{2}+ui\right)^2} d\left(\frac{y}{2}\right) = \frac{e^{-u^2}}{\sqrt{\pi}} = \varphi(u),$$

womit Gl. (V) *bewiesen ist.*

Aus (18) ergibt sich auch sofort, daß Gl. (III) für die jetzt betrachteten Verteilungen gilt. Denn zu einem beliebig vorgegebenen U und ε kann man nach (18) ein n_0 so bestimmen, daß für $n > n_0$

$$(18')\qquad \left|\mathfrak{w}_n(u) - \frac{\varphi(u)}{r_n}\right| < \frac{\varepsilon}{4 r_n U} \quad \text{für} \quad |u| < U.$$

Sind nun a, b zwei Punkte des Intervalls $|u| < U$, so besteht $W_n(b) - W_n(a)$ aus höchstens $r_n \cdot 2U$ Summanden $\mathfrak{w}_n(u_\iota)$. Demnach ist zufolge (18′):

$$(18'')\qquad \left|W_n(b) - W_n(a) - \sum_a^b \frac{\varphi(u_\iota)}{r_n}\right| < \frac{\varepsilon}{2}.$$

Die hier auftretende Summe geht aber im Sinne des Riemannschen Integralbegriffes gegen $\Phi(b) - \Phi(a)$.

Um einzusehen, daß die Ausdehnung dieser Überlegung auf den Fall $a = -\infty$ zulässig ist, braucht man nur zu beachten, daß sowohl $\Phi(-U)$ als $W_n(-U)$ kleiner als $1 : 2U^2$ sein muß, weil die Streuung beider Verteilungen $\frac{1}{2}$ ist. Wählt man also $U^2 > \frac{1}{\varepsilon}$, so gilt für $u < U$, und weiter mit

Rücksicht auf das Verhalten der Verteilungen im positiv Unendlichen, auch darüber hinaus:

$$(18a)\qquad |W_n(u) - \Phi(u)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Dies ist der Inhalt der Gl. (III).

6. Beweis der Behauptung IV. Wir schlagen einen analogen Weg ein, um IV zu beweisen, indem wir zuerst eine Umkehrformel für die Berechnung von $v_\varkappa$ bzw. w_n aus $f_\varkappa$ bzw. p_n aufstellen und sodann die besondere Natur der hier auftretenden $f_\varkappa$ untersuchen.

Die Umkehrung wird vermittelt durch das *Fouriersche Integraltheorem*, das für jede integrable, im Unendlichen verschwindende Funktion von beschränkter Schwankung angeschrieben werden kann. Es lautet für $v_\varkappa(x)$:

$$(19)\qquad \bar{v}_\varkappa(z) = \frac{1}{2\pi}\int du \int v_\varkappa(x)\cos u(z-x)\,dx$$

$$= \frac{1}{2\pi}\,\Re\int e^{-(z-a_\varkappa)ui}\,du \int v_\varkappa(x)\,e^{(x-a_\varkappa)ui}\,dx.$$

Durch den Querstrich über $v_\varkappa$ soll angedeutet werden, daß $\bar{v}_\varkappa$ nur an Stetigkeitstellen mit $v_\varkappa$ übereinstimmt, sonst aber das arithmetische Mittel zwischen dem rechten und linken Limes bezeichnet. Führt man in (19) den Wert von $f_\varkappa$ aus (11) ein, so gewinnt man die *Umkehrformel*:

$$(19')\qquad \bar{v}_\varkappa(z) = \frac{1}{2\pi}\,\Re\int f_\varkappa(a_\varkappa + u)\,e^{-(z-a_\varkappa)ui}\,du.$$

Denn für die komplexe Adjunkte $f_\varkappa$ einer Verteilung mit beschränkter Ableitung nimmt (11) die Form

$$(19'')\qquad f_\varkappa(a_\varkappa + y) = \int v_\varkappa(x)\,e^{(x-a_\varkappa)yi}\,dx$$

an. Analog hat man für die Ableitung des Integralproduktes:

$$(19''')\qquad \bar{w}_n(z) = \frac{1}{2\pi}\,\Re\int p_n(u)\,e^{-zui}\,du,$$

da auch w_n, das durch (2) aus den $v_\varkappa$ gewonnen wird, von beschränkter Schwankung sein muß.

Von den $f_\varkappa$ ist im Abschn. 4 allgemein gezeigt worden, daß sie dem *ersten Teil* der Voraussetzungen des § 1 genügen. Wir wollen jetzt beweisen, daß sie im vorliegenden Fall — in dem die Ableitungen $v_\varkappa$ von gleichmäßig beschränkter Schwankung sind — *auch die übrigen* Voraussetzungen der Sätze I und II erfüllen.

Bezeichnet $\sigma_\varkappa$ die Gesamtschwankung von $v_\varkappa(x)$, so läßt sich $v_\varkappa(x)$ als Differenz zweier monotoner Funktionen, die von null bis $\frac{1}{2}\sigma_\varkappa$ wachsen, darstellen. Daher ist nach dem zweiten Mittelwertsatz der Integralrechnung

$$\left|\int v_\varkappa(x)\cdot\cos y\,(x-a_\varkappa)\,dx\right| \leqq 2\cdot\frac{\sigma_\varkappa}{2}\left|\int_\alpha^\beta \cos y\,(x-a_\varkappa)\,dx\right| \leqq \frac{2\,\sigma_\varkappa}{|y|},$$

und da dieselbe Beziehung für den Sinus gilt und die $\sigma_\varkappa$ die obere Schranke σ haben:

$$(20) \qquad |f_\varkappa(a_\varkappa+y)| \leqq \frac{2\sqrt{2}\,\sigma_\varkappa}{|y|} \leqq \frac{2\sqrt{2}\,\sigma}{|y|};$$

womit den Bedingungen (a4) *und* (b4) *des* § 1 *Genüge getan ist.*

Die Bedingungen (a3) und (b3) verlangen, daß $|f_\varkappa|$ dauernd unterhalb 1 bleibt (abgesehen von der Stelle $y=0$), was für kleine y aus (20) nicht zu ersehen ist. Nun kann das Integral rechts in (19″) auf keinem Teilintervall α, β größeren Betrag haben als das zu demselben Intervall α, β gehörige Integral von $v_\varkappa$, das wir mit $V_\varkappa(\alpha,\beta)$ bezeichnen wollen. Es genügt daher, da das über die ganze Achse erstreckte Integral von $v_\varkappa(x)$ den Wert 1 hat, für *irgendein* $\alpha<\beta$ nachzuweisen, daß für $|y|>y_0$

$$(20') \qquad V_\varkappa(\alpha,\beta)-\left|\int_\alpha^\beta v_\varkappa(x)\,e^{(x-a_\varkappa)\,yi}\,dx\right|>\delta,$$

damit sicher auch $1-|f_\varkappa|>\delta$ wird. Wählen wir nun α, β so, daß $|y|(\beta-\alpha)=2\pi$, so bedeutet die linke Seite von (20′), durch $V_\varkappa(\alpha,\beta)$ dividiert, den Schwerpunktsabstand einer über den Umfang des Einheitskreises (mit der Dichte $|v_\varkappa:y|$) ausgebreiteten Masse $V_\varkappa(\alpha,\beta)$ von der Kreisperipherie. Dieser Abstand wird am kleinsten, wenn die Masse möglichst konzentriert wird: da aber $v_\varkappa$ nicht größer als $\frac{1}{2}\sigma_\varkappa$ werden kann, wird die stärkste Konzentration, also das Minimum von (20′) erreicht, durch gleichförmige Verteilung mit der Dichte $|\sigma_\varkappa:2y|$ über den Bogen $2\,y\,V_\varkappa(\alpha,\beta):\sigma_\varkappa$. Daraus ergibt sich

$$(20'') \qquad \left|\int_\alpha^\beta v_\varkappa(x)\,e^{(x-a_\varkappa)\,yi}\,dx\right| \leqq \left|\frac{\sigma_\varkappa}{y}\sin\frac{V_\varkappa(\alpha,\beta)\,y}{\sigma_\varkappa}\right| \leqq \frac{\sigma}{y_0}\sin\left|\frac{V_\varkappa(\alpha,\beta)\,y_0}{\sigma}\right|.$$

Der Wert von $V_\varkappa(\alpha,\beta)$ kann nicht etwa bei jeder Wahl von α mit wachsendem $\varkappa$ gegen null gehen, da sonst Streuung und Mittelwert, entgegen der Annahme (b2) dieses Paragraphen, nicht beschränkt sein könnten. Man weist leicht nach, daß auf das Intervall $\mp(A+2S)$ mehr als $\frac{3}{4}$ des $\int v\,dx$ entfallen muß, so daß hier die durchschnittliche Verteilungsdichte mindestens $3:8\,(A+2S)$ beträgt. Daher muß es innerhalb des Intervalls $\mp(A+2S)$ wenigstens ein Teilintervall von der Länge $\beta-\alpha=2\pi:|y|$ geben, für welches in (20″)

$$(20''') \qquad V_\varkappa(\alpha,\beta)>\frac{3\pi}{4\,|y|\,(A+2S)} \geqq \frac{3\pi}{16\,\sigma\,(A+2S)}$$

zu setzen ist, weil wir mit Rücksicht auf (20) nur $|y|$-Werte unterhalb 4σ zu betrachten brauchen. Endgültig ist daher für $y_0 < |y| \leqq 4\sigma$

$$(21) \qquad 1-|f_\varkappa| > \frac{3\pi}{16\sigma(A+2S)}\left[1-\frac{\sin\alpha}{\alpha}\right] \quad \text{mit} \quad \alpha = \frac{3\pi y_0}{16\sigma^2(A+2S)},$$

womit den Bedingungen (a3) *und* (b3) *des* § 1 *Genüge getan ist.*

Endlich haben die Streuungen $s_\varkappa^2$ auch eine durch σ bestimmte untere Grenze. Denn eine Verteilung, deren Dichte nirgends größer als $\frac{1}{2}\sigma_\varkappa$ sein kann, erreicht ihre kleinste Streuung bei stärkster Konzentration, also wenn die Dichte längs einer Strecke $2:\sigma_\varkappa$ den Maximalwert $\frac{1}{2}\sigma_\varkappa$ aufweist. Daher ist

$$s_\varkappa^2 \geqq \frac{1}{3\sigma_\varkappa^2} \geqq \frac{1}{3\sigma^2}.$$

Mithin sind alle Bedingungen des § 1 *erfüllt.*

Wenden wir auf (19‴) das Korollar II_2 des § 2 in derselben Weise, wie es oben bei (17‴) geschehen ist, an, so finden wir:

$$(22) \quad \overline{w}(x) = \lim_{n=\infty} w_n(x) = \frac{1}{2\pi}\Re\int e^{-\frac{u^2}{y}-xui}\,du = \frac{e^{-x^2}}{\sqrt{\pi}}\int e^{-\left(\frac{u}{2}+xi\right)^2} d\left(\frac{u}{2}\right) = \frac{e^{-x^2}}{\sqrt{\pi}}.$$

Dies ist der genaue Inhalt von Gl. (IV) *und* (IV′).

Aus (22) folgt auch leicht, daß im vorliegenden Fall überdies Gl. (III) gilt. Denn da die Streuung von W_n für jedes n gleich $\frac{1}{2}$ ist, hat man für $U>0$

$$(22') \qquad \int^{-U} w_n(u)\,du < \frac{1}{2U^2}, \qquad \int_U w_n(u)\,du < \frac{1}{2U^2}.$$

Man kann daher stets ein U so groß wählen, daß der Betrag rechts in (22′) kleiner als $\frac{\varepsilon}{2}$ wird; macht man dann n_0 so groß, daß für $n > n_0$

$$(22'') \qquad |\overline{w}_n(u) - \varphi(u)| < \frac{\varepsilon}{4U} \quad \text{für} \quad |u| \leqq U,$$

was nach (22) immer möglich ist, so ist im ganzen Bereich $-\infty, U$, und weiterhin, weil auch

$$(22''') \qquad \int_U \varphi(u)\,du < \frac{1}{2U^2}$$

gilt, auch im Bereich U, ∞

$$(23) \qquad \left|\int_a^b \overline{w}_n(u)\,du - \int_a^b \varphi(u)\,du\right| < \varepsilon,$$

was der Inhalt von Gl. (III) *ist.*

7. Ansatz zum Beweis der Behauptung III. Die Behauptung III besteht darin, daß Gl. (III) auch ohne eine der einschränkenden Voraussetzungen des Abschn. 5 oder 6 bewiesen werden soll. Bedenkt man, daß

jede Verteilung $V_\varkappa$ oder W_n eine monotone Funktion, daher die Differenz $W_n - \Phi$ von beschränkter Schwankung, überdies im Unendlichen null und mit Rücksicht auf die Voraussetzung (b1) auch absolut integrabel ist, so kann man für $W_n - \Phi$ das *Fouriersche Integraltheorem* ansetzen:

$$(24) \quad \overline{W}_n(z) - \Phi(z) = \frac{1}{2\pi}\int du \int [W_n(x) - \Phi(x)] \cos u(z-x)\, dx$$
$$= \frac{1}{2\pi}\Re \int e^{-zui} du \int [W_n(x) - \Phi(x)] e^{xui} dx.$$

Durch den Querstrich soll wieder angedeutet werden, daß $\overline{W}_n$ nur an Stetigkeitsstellen mit W_n übereinstimmt, sonst aber das arithmetische Mittel des rechten und linken Limes bezeichnet. Andrerseits findet man aus (11a), unter Beachtung von (15) und mit Rücksicht auf das starke Verschwinden von $W_n - \Phi$ für $\pm\infty$, durch partielle Integration:

$$(24') \quad p_n(u) - e^{-\frac{u^2}{4}} = \int e^{xui}[dW_n(x) - d\Phi(x)]$$
$$= -ui\int [W_n(x) - \Phi(x)] e^{xui} dx.$$

Setzt man dies in (24) ein, so erhält man die allgemeine *Umkehrformel:*

$$(25) \quad \overline{W}_n(z) - \Phi(z) = -\frac{1}{2\pi}\Re \int \frac{e^{-uzi}}{ui}\left[p_n(u) - e^{-\frac{u^2}{4}}\right] du.$$

In Abschn. 4 ist gezeigt worden, daß der Klammerausdruck rechts für wachsende n in jedem endlichen Bereich $|u| < U$ gleichmäßig gegen null geht. Um aus (25) schließen zu können, daß zugleich $\overline{W}_n$ gegen Φ konvergiert, müßte man noch sehen, daß das Integral rechts in (25) für den Integrationsbereich $|u| > U$ mit wachsendem n null wird. Dies gilt jedenfalls, wenn die $f_\varkappa$, deren Produkt p_n ist, den Bedingungen (a3) (a4) und (b3) (b4) des § 1 genügen, also z. B. im Falle des Satzes IV (Verteilungsdichte von beschränkter Schwankung).

Im Falle des Satzes V, und daher auch im allgemeinen Fall beliebiger Verteilungen, *bestehen jedoch die Voraussetzungen* (a3) (a4) *usw. nicht.* Nimmt man als Beispiel etwa die unter die Annahmen des Satzes V mit $\varepsilon = 2$ fallenden Verteilungen:

$$\begin{aligned} V_1(x) = V_2(x) = \ldots &= 0 \quad \text{für} \quad -1 > x, \\ &= \tfrac{1}{2} \quad \text{„} \quad +1 > x \geqq -1, \\ &= 1 \quad \text{„} \quad x \geqq 1, \end{aligned}$$

für welche $f_1(y) = f_2(y) = \ldots = \cos y$ wird, so sieht man, daß $f_\varkappa$ im allgemeinen *nicht einmal gegen null geht.* Unter den besonderen Voraussetzungen des Satzes V war es wohl möglich, die Funktionen $f_\varkappa$ durch die günstiger verlaufenden $f_\varkappa^*$ zu ersetzen und so auf dem Umweg über Gleichung (V) die Gleichung (III) zu beweisen. Für beliebige Verteilungen scheint ein derartiger Weg und damit eine direkte Ausnützung der Beziehung (25) und (13) *nicht möglich* zu sein. Um Gleichung (III) im vollen Umfang der Behauptung III zu beweisen, müssen wir weiter ausholen und im folgenden Paragraphen einige Hilfsmittel entwickeln, die in § 5 zum Beweise selbst führen werden.

§ 4.

Vorbereitungen zum Beweis der Behauptung III.

1. Reelle Adjunkte einer Verteilung, Es wird für das Folgende nützlich sein, an Stelle der in § 3, Gleichung (11) definierten „komplexen Adjunkten" eine *reelle* Funktion einzuführen, die sich hinsichtlich des Multiplikationsgesetzes ähnlich verhält, die allerdings nicht den Voraussetzungen des § 1 entspricht. *Wir definieren als reelle Adjunkte einer beliebigen Verteilung $V_\varkappa(x)$ mit dem Mittelwert $a_\varkappa$ und der Streuung $s_\varkappa^2$ die reelle Funktion.*

$$(1) \qquad g_\varkappa(u) = e^{-\frac{s_\varkappa^2 u^2}{2}} \int e^{-(x-a_\varkappa)u}\, dV_\varkappa(x).$$

Das Integral existiert für jeden endlichen Wert von u, falls die $V_\varkappa$ der Voraussetzung (a1) des § 3 genügen. Besitzt $V_\varkappa$ eine beschränkte Ableitung $v_\varkappa$, so ist $g_\varkappa$ im wesentlichen die sog. *Laplacesche Transformierte* von $v_\varkappa$. Für $u = 0$ hat $g_\varkappa(u)$ den Wert 1 und die erste Ableitung null; aber die zweite Ableitung ist, im Gegensatz zu Gleichung (11''), § 3 gleich $-s_\varkappa^2 + s_\varkappa^2 = 0$. Über das Verhalten für große u können wir vorerst nichts aussagen. Sicher ist nur, daß $g_\varkappa$ der Definition nach positiv ist, also keine Nullstellen im Endlichen besitzt.

Die reelle Adjunkte des n-ten Integralproduktes bezeichnen wir mit $q_n(u)$:

$$(1') \qquad q_n(u) = e^{-\frac{u^2}{4}} \int e^{-xu}\, dW_n(x);$$

sie besitzt natürlich alle eben erwähnten Eigenschaften der $g_\varkappa(u)$, sobald die Konvergenz des Integrales in (1') nachgewiesen ist. Wir beweisen diese, indem wir $q_n(u)$ aus den $g_1(u), g_2(u), \ldots, g_n(u)$ berechnen und damit zugleich die grundlegende Formel für die Anwendung der g und q ableiten.

Setzen wir in (1') so wie in § 3, 4 den Wert von $dW_n(x)$ aus der Definitionsgleichung (1) von § 3 ein und verwenden wieder an Stelle von x die neue Integrations-Variable x_n, die durch (8) in § 3 definiert ist, so entsteht durch Trennung der Variablen:

$$q_n(u) = e^{-\frac{u^2}{4}} \int\int \dots \int e^{-\frac{u}{r_n}\sum\limits_{\varkappa=1}^{n}(x_\varkappa - a_\varkappa)} dV_1(x_1) \cdot dV_2(x_2) \dots dV_n(x_n)$$

$$= e^{-\frac{u^2}{2r_n^2}\sum\limits_{\varkappa=1}^{n} s_\varkappa^2} \int e^{-(x_1-a_1)\frac{u}{r_n}} dV_1(x_1) \cdot \int e^{-(x_2-a_2)\frac{u}{r_n}} dV_2(x_2) \dots \int e^{-(x_n-a_n)\frac{u}{r_n}} dV_n(x_n).$$

Demnach mit Rücksicht auf (1):

$$q_n(u) = g_1\left(\frac{u}{r_n}\right) \cdot g_2\left(\frac{u}{r_n}\right) \dots g_n\left(\frac{u}{r_n}\right). \tag{2}$$

Es multiplizieren sich somit bei Bildung der Integralprodukte die reellen Adjunkten so wie die komplexen. Für den Grenzwert des q_n folgt daraus natürlich nichts, da die $g_\varkappa$ den wesentlichsten Bedingungen des § 1 nicht genügen. Der Weg, auf dem wir zu unserem Beweis gelangen, ist vielmehr der, daß wir eine *Potenzentwicklung* der $g_\varkappa$ aufsuchen, aus dieser eine Reihe für q_n gewinnen und schließlich *aus den Koeffizienten dieser Reihe* direkt auf W_n schließen.

2. Entwicklung der reellen Adjunkten. Wir setzen zunächst formal an:

$$g_\varkappa\left(\frac{u}{s_\varkappa\sqrt{2}}\right) = e^{-\frac{u^2}{4}} \int e^{-\frac{x-a_\varkappa}{s_\varkappa\sqrt{2}}u} dV_\varkappa(x) = c_\varkappa^{(0)} + c_\varkappa^{(1)} u + c_\varkappa^{(2)} u^2 + \dots, \tag{3}$$

worin nach dem oben Dargelegten stets

$$c_\varkappa^{(0)} = 1, \quad c_\varkappa^{(1)} = c_\varkappa^{(2)} = 0 \tag{3a}$$

sein muß, und wollen beweisen, daß die Reihe rechts für alle Werte von u konvergiert. Zu diesem Zwecke benutzen wir die bekannte Entwicklung:

$$e^{-\frac{u^2}{4} - uz} = \alpha_0(z) + \alpha_1(z)u + \alpha_2(z)u^2 + \dots, \tag{4}$$

in der die $\alpha_0, \alpha_1, \dots$ die *Hermiteschen Polynome:*

$$\left\{\begin{array}{ll} \alpha_0 = 1, \quad \alpha_1 = -z, \quad \alpha_2 = \frac{z^2}{2} - \frac{1}{4}, \quad \alpha_3 = -\frac{z^3}{6} + \frac{z}{4}, \dots & \\ \alpha_m = \sum \frac{(-1)^{m-k} z^{m-2k}}{4^k k! (m-2k)!} & \left(k = 0, 1, 2, \dots, \frac{m}{2} \text{ bzw. } \frac{m-1}{2}\right) \end{array}\right. \tag{4'}$$

bezeichnen. Die Reihe (4) konvergiert gleichmäßig in jedem endlichen Intervall von z und u, weil die Zahl der Summanden von α_m nur mit $\frac{m}{2}$ wächst, während der kleinste Nenner noch $\left(\frac{m}{2}\right)!$ enthält. Verwenden wir von nun an die beiden Variablen x und

$$z = \frac{x - a_\varkappa}{s_\varkappa\sqrt{2}}, \tag{5}$$

der kürzeren Schreibweise wegen, *nebeneinander*, so ergibt sich für die Koeffizienten von (3) der Ausdruck

$$(3') \qquad c_{\varkappa}^{(m)} = \int \alpha_m(z)\, d V_{\varkappa}(x)$$

und es ist nur zu zeigen, daß die mit diesen Koeffizienten gebildete Reihe (3) konvergiert und — mit Rücksicht auf den unendlichen Integrationsbereich — daß sie wirklich $g_{\varkappa}$ darstellt.

Das erstere folgt aus Gleichung (6) des § 3, nämlich:

$$(6) \qquad \int |z|^n\, d V_{\varkappa}(x) < C\left(\frac{n}{2c^2 e}\right)^{\frac{n}{2}}.$$

Denn nach (4′) ist bei Berücksichtigung von $m! > \left(\frac{m}{2}\right)!\left(\frac{m}{2}\right)^{\frac{m}{2}}$

$$(6') \qquad |\alpha_m| \leqq \sum_{k=0}^{\frac{m}{2}} \frac{|z|^{m-2k}}{4^k\, k!\left(\frac{m}{2}-k\right)!}\, \frac{1}{\left(\frac{m}{2}-k\right)^{\frac{m}{2}-k}},$$

daher nach (3′)

$$(3'') \qquad |c_{\varkappa}^{(m)}| < C \sum_{k=0}^{\frac{m}{2}} \frac{(c^2 e)^{-\frac{m}{2}+k} \cdot 4^{-k}}{k!\left(\frac{m}{2}-k\right)!} = \frac{C}{\left(\frac{m}{2}\right)!}\left(\frac{1}{c^2 e}+\frac{1}{4}\right)^{\frac{m}{2}} = c^{(m)},$$

womit die Konvergenz von (3) für alle u bewiesen ist. Man erkennt auch, daß die Folge der $c_{\varkappa}^{(m)}$ bei festem m beschränkt ist und $c^{(m)}$ zur oberen Schranke hat.

Um die Zulässigkeit der gliedweisen Integration als Übergang von (4) auf (3) zu beweisen, beachten wir, daß die mit den oberen Grenzen der Beträge der $c_{\varkappa}^{(m)} : C$ nach (3″) gebildete Reihe

$$(6'') \qquad \sum_{k=0}^{\infty} \frac{u^k}{\left(\frac{k}{2}\right)!}\left(\frac{1}{c^2 e}+\frac{1}{4}\right)^{\frac{k}{2}} = U(u)$$

konvergiert, und zwar gleichmäßig in jedem endlichen Bereich von u. Ist also ein ε und ein u_0 vorgeschrieben, so kann man dazu $U_0 = U(u_0)$ berechnen und auf Grund der Konvergenz des Integrals (b1) ein x_0 so bestimmen, daß für alle $|x| > x_0$

$$(6''') \qquad \int^{-x} e^{c^2 z^2}\, d V_{\varkappa}(x) + \int_{x} e^{c^2 z^2}\, d V_{\varkappa}(x) < \frac{\varepsilon}{3\, U_0}.$$

Ersetzt man hier $e^{c^2 z^2}$ durch $|z|^n$, so hat man rechts den Faktor von C aus (6) hinzuzufügen (vgl. § 3, Gl. (6)). Ersetzt man daher $e^{c^2 z^2}$ durch

$\alpha_m(z)$, so ist rechts der Faktor von C aus $(3'')$ beizufügen. Mithin wird mit Rücksicht auf $(6'')$ *bei beliebigem* m:

$$(7')\quad \left|\int_{-x}^{x}[\alpha_0(z)+\alpha_1(z)u+\ldots+\alpha_m(z)u^m]\,dV_\varkappa(x)-[c_\varkappa^{(0)}+c_\varkappa^{(1)}u+\ldots+c_\varkappa^{(m)}u^m]\right|<\frac{\varepsilon}{3},$$
$$\text{für } x>x_0,\quad |u|<u_0.$$

Weiter kann man, da das Integral in (3) existiert, ein $x_1>x_0$ finden, für welches bei $|u|<u_0$

$$(7'')\quad \left|\int_{-x_1}^{x_1}[\alpha_0(z)+\alpha_1(z)u+\alpha_2(z)u^2+\ldots]\,dV_\varkappa(x)-g_\varkappa\left(\frac{u}{s_\varkappa\sqrt{2}}\right)\right|<\frac{\varepsilon}{3}.$$

Endlich läßt sich wegen der gleichmäßigen Konvergenz von (4) zu diesem x_1 ein m_1 bestimmen, welches das Restglied der Reihe (4) für $|u|<u_0$ und $|x|\leqq x_1$ unter den Betrag $\varepsilon:6x_1$ herabdrückt, so daß für dieses m_1:

$$(7''')\quad \left|\int_{-x_1}^{x_1}[\alpha_0(z)+\ldots]\,dV_\varkappa(x)-\int_{-x_1}^{x_1}[\alpha_0(z)+\ldots+\alpha_{m_1}(z)u^{m_1}]\,dV_\varkappa(x)\right|<\frac{\varepsilon}{3}.$$

Wendet man $(7')$, das für jedes m und für jedes $x>x_0$ gilt, auf m_1 und x_1 an, so erhält man durch Zusammenhalten von $(7')$, $(7'')$ und $(7''')$ das verlangte Resultat:

$$(7)\quad \left|g_\varkappa\left(\frac{u}{s_\varkappa\sqrt{2}}\right)-[c_\varkappa^{(0)}+c_\varkappa^{(1)}u+\ldots+c_\varkappa^{(m_1)}u^{m_1}]\right|<\varepsilon\quad\text{für}\quad |u|<u_0.$$

Unsere Ergebnisse $(3''')$ und (7) lassen sich kurz so aussprechen:

Für Verteilungen, die unseren Voraussetzungen in § 3 genügen, ist die reelle Adjunkte eine ganze Funktion und die Folge der Koeffizienten $c_\varkappa^{(m)}$ *bei festem* m *beschränkt.*

3. Über die Nullstellen der Hermiteschen Polynome. Die wesentlichsten Eigenschaften der in $(4')$ angeführten Hermiteschen Polynome α_m — daß sie m reelle Zeichenwechsel besitzen usw. — setzen wir als bekannt voraus und wollen hier nur die Präzisierung eines (bekannten) Satzes über die *Verdichtung der Nullstellen*[6]) mit wachsendem m durchführen. Man beweist leicht, daß die Funktionen

$$(8)\qquad \psi_m(z)=e^{-\frac{z^2}{2}}\alpha_m(z),$$

die in den Nullstellen mit den α_m übereinstimmen, der Differentialgleichung

$$(8')\qquad \psi_m''=-(2m+1-z^2)\,\psi_m$$

genügen. Denn (8) gibt in $(8')$ eingesetzt die Differentialgleichung

[6]) A. A. Markoff, Bull. de l'Acad. St. Petersburg 5, série 9, 1898, p. 437; Wahrscheinlichkeitsrechnung, deutsch von H. Liebmann, Leipzig 1912, S. 258.

$\alpha_m'' - 2z\alpha_m' + 2m\alpha_m = 0$, und da man aus (4) (indem man beiderseits differenziert und dann links wieder die Reihe für den Exponentialausdruck setzt) leicht ableitet, daß $\alpha_m' = -\alpha_{m-1}$, so ist (8′) nur der etwas veränderte Ausdruck für die bekannte Rekursionsformel

$$(8'') \qquad \alpha_{m-1} + 2z\alpha_m + 2m\alpha_{m+1} = 0,$$

welche die Hermiteschen Polynome (4′) charakterisiert. Nun haben die Integrale von (8′) die Eigenschaft, daß sie in dem Bereiche, in dem $z^2 - (2m+1) > 0$ ist, ihre *konkave* Seite der Abszissenachse zukehren, also hier keine im Endlichen gelegene Nullstelle haben können, da sie sich sonst nicht asymptotisch der Abszissenachse nähern könnten. Andererseits verhalten sich die Integrale im Bereich $|z| < \sqrt{2m+1}$ wie Sinus-Linien, d. h. sie bilden Wellen wie die Integrale von $z'' = -c^2 z$ und diese Wellen sind im Bereich $|z| < z_0 < \sqrt{2m+1}$ mindestens so stark gekrümmt, wie die einer Sinus-Linie mit $c^2 = 2m + 1 - z_0^2$ und höchstens so stark, wie die einer Sinus-Linie mit $c^2 = 2m + 1$. Daraus folgt also der Satz:

Die m reellen Nullstellen der Hermiteschen Polynome $\alpha_m(z)$ liegen nich außerhalb des Gebietes $|z| \leqq \sqrt{2m+1}$ und verdichten sich bei wachsendem m derart, daß in einem Gebiet $|z| \leqq z_0 < \sqrt{2m+1}$ der Abstand d_m zweier aufeinander folgender nicht größer als $\frac{\pi}{\sqrt{2m+1-z_0^2}}$ ist. Andererseits kann der Abstand nicht kleiner als $\frac{\pi}{\sqrt{2m+1}}$ sein:

$$(8''') \qquad \frac{\pi}{\sqrt{2m+1}} < d_m < \frac{\pi}{\sqrt{2m+1-z_0^2}}.$$

4. Die einer Verteilung $V_\varkappa$ zugeordneten Treppen $S_\varkappa^{(m)}$. Zwischen den Momenten einer Verteilung und den Koeffizienten ihrer reellen Adjunkten besteht ein einfacher Zusammenhang. Da man die Gleichungen (4′), welche die Hermiteschen Polynome definieren, unschwer nach den z^m auflösen kann, sobald die α_m gegeben sind, nämlich:

$$(9) \qquad z^0 = \alpha_0, \quad z^1 = -\alpha_1, \quad z^2 = 2\alpha_2 + \tfrac{1}{2}, \quad z^3 = -6\alpha_3 - \tfrac{3}{2}\alpha_1 \ldots,$$

so sieht man, daß *das m-te Moment von $V_\varkappa$ sich durch $c_\varkappa^{(0)}, c_\varkappa^{(1)}, \ldots, c_\varkappa^{(m)}$ linear ausdrücken* läßt, z. B. für $m = 3$:

$$\int z^3 dV_\varkappa(x) = -6\int \alpha_3(z) dV_\varkappa(x) - \tfrac{3}{2}\int \alpha_1(z) dV_\varkappa(x) = -6c_\varkappa^{(3)} - \tfrac{3}{2}c_\varkappa^{(1)}.$$

Dies vorausgeschickt definieren wir als eine „*m-stufige Treppe*“ oder „Treppenlinie“ eine *Verteilung* (d. i. eine monoton von null bis 1 ansteigende Funktion), die *genau m im Endlichen gelegene Wachstumstellen* hat und sonst überall konstant verläuft. Das Bild einer solchen Funktion ist tatsächlich das des Querschnittes durch eine Treppe, die aus m Stufen besteht. Wir wollen den folgenden, für den Zusammenhang zwischen einer Verteilung und den Koeffizienten ihrer Adjunkten fundamentalen Satz beweisen:

Bedeutet m eine beliebige natürliche Zahl und $V_\varkappa$ eine unseren Voraussetzungen (a 1) *und* (a 2) *genügende Verteilung* (oder eine Verteilung, deren reelle Adjunkte eine ganze Funktion ist), *die selbst nicht weniger als m Wachstumstellen hat, so gibt es stets eine m-stufige Treppe $S_\varkappa^{(m)}$, die mit $V_\varkappa$ in den ersten $2m$ Momenten* (vom nullten bis zum $2m-1$-ten) *oder*, was dasselbe ist, *in den ersten $2m$ Koeffizienten der reellen Adjunkten, übereinstimmt; sie wird in jeder Stufe von $V_\varkappa$ doppelt, nämlich im horizontalen und vertikalen Teil der Stufe, geschnitten.*

Um dies zu beweisen, konstruieren wir zunächst die Treppenlinie, indem wir ihre Wachstumstellen (und zwar als Nullstellen eines Polynoms $\beta_\varkappa^{(m)}$) und dann die dazugehörigen Sprunggrößen $\mathfrak{s}_\varkappa^{(1)}, \mathfrak{s}_\varkappa^{(2)}, \ldots, \mathfrak{s}_\varkappa^{(m)}$ angeben. Die Wachstumstellen nennen wir auch einfach die *Stufen*, und die $\mathfrak{s}_\varkappa^{(1)}$ die *Stufenhöhen*.

Wir gehen von der quadratischen Form in den y_ι

$$(10) \qquad \int [\alpha_0 y + \alpha_1 y_1 + \cdots + \alpha_{m-1} y_{m-1}]^2 \, d V_\varkappa(x) = \sum_{\iota,\lambda} A_\varkappa^{(\iota,\lambda)} y_\iota y_\lambda$$

aus, die positiv definit sein muß, weil der Integrand nicht negativ ist und das Integral nur dann verschwinden könnte, wenn $V_\varkappa(x)$ gerade nur an den höchstens $m-1$ Nullstellen des Klammerausdruckes Wachstumstellen hätte, was der Voraussetzung zuwiderläuft. Setzen wir also nach (10)

$$(10') \qquad A_\varkappa^{(\iota,\lambda)} = \int \alpha_\iota(z)\, \alpha_\lambda(z)\, d V_\varkappa(x) \qquad (\iota, \lambda = 0, 1, 2, \ldots, m)$$

— immer mit dem in (5) definierten Zusammenhang zwischen z und x — so hat das Gleichungssystem

$$(10'') \qquad \sum_{\iota=0}^{m-1} A_\varkappa^{(\iota,\lambda)} \varrho_\varkappa^{(\iota)} = - A_\varkappa^{(m,\lambda)} \qquad (\lambda = 0, 1, 2, \ldots, m-1)$$

eine von null verschiedene Nenner-Determinante, also eine eindeutige Lösung $\varrho_\varkappa^{(0)}\ \varrho_\varkappa^{(1)} \ldots \varrho_\varkappa^{(m-1)}$. Bilden wir mit diesen $\varrho_\varkappa^{(\iota)}$ des Polynom m-ten Grades in z:

$$(11) \qquad \beta_\varkappa^{(m)} = \alpha_m + \varrho_\varkappa^{(m-1)} \alpha_{m-1} + \varrho_\varkappa^{(m-2)} \alpha_{m-2} + \cdots + \varrho_\varkappa^{(0)} \alpha_0,$$

so läßt sich vor allem zeigen, daß es *genau m reelle Nullstellen besitzt.*

Gleichung (10″) besagt nämlich mit Rücksicht auf (10′), daß zunächst jedes Integral der Form

$$(11') \qquad \int \alpha_\lambda \beta_\varkappa^{(m)} d V_\varkappa(x) = \sum_{\iota=0}^{m-1} \int \alpha_\lambda \alpha_\iota \varrho_\varkappa^{(\iota)} d V_\varkappa(x) + \int \alpha_\lambda \alpha_m d V_\varkappa(x)$$
$$= \sum_{\iota=0}^{m-1} \varrho_\varkappa^{(\iota)} A_\varkappa^{(\iota,\lambda)} + A_\varkappa^{(m,\lambda)} = 0 \qquad (\lambda = 0, 1, 2, \ldots, m-1)$$

verschwinden muß, weiterhin daher auch jedes Integral der Form

$$(11'') \qquad \int P_{m-1}(z) \cdot \beta_\varkappa^{(m)}(z) d V_\varkappa(x) = 0,$$

in dem $P_{m-1}(z)$ ein Polynom von höchstens $(m-1)$-ten Grade bezeichnet. Das letztere deshalb, weil nach dem, was oben gesagt wurde, jedes solche Polynom linear durch die ersten m Polynome α, also durch $\alpha_0, \alpha_1, \ldots, \alpha_{m-1}$, ausgedrückt werden kann. Aus dem Verschwinden von (11″) folgt nun, daß $\beta_\varkappa^{(m)}$ genau m reelle Zeichenwechsel besitzen muß, da man sonst für $P(z)$ das Differenzenprodukt $(z - z_1)(z - z_2) \ldots$, wo $z_1, z_2 \ldots$ die Nullstellen sind, setzen könnte, womit dann der Integrand konstantes Zeichen im ganzen Intervall erhielte.

Die m Nullstellen von $\beta_\varkappa^{(m)}$ sollen die Stufen der Treppe $S_\varkappa^{(m)}$ sein; wir bezeichnen sie mit $z_\varkappa^{(\iota)}$ bis $z_\varkappa^{(m)}$.

Als Stufenhöhe sei der Stelle $z_\varkappa^{(\iota)}$ zugeordnet:

$$(12) \qquad \mathfrak{s}_\varkappa^{(\iota)} = \int \frac{\beta_\varkappa^{(m)}(z) d V_\varkappa(x)}{(z - z_\varkappa^{(\iota)}) \beta_\varkappa'^{(m)}(z_\varkappa^{(\iota)})} \qquad (\iota = 1, 2, \ldots, m).$$

Dabei bedeutet β' die Ableitung von β nach z, und es bleibt jetzt zu beweisen: 1. daß die $\mathfrak{s}_\varkappa^{(\iota)}$ sämtlich positiv, 2. daß ihre Summe gleich 1, 3. daß die ersten $2m$ Koeffizienten der Adjunkten der durch sie bestimmten Verteilung mit den $c_\varkappa^{(0)}$ bis $c_\varkappa^{(2m-1)}$ übereinstimmen; 4. daß $V_\varkappa(x)$ und $S_\varkappa^{(m)}(x)$ genau $2m-1$ Schnittpunkte haben.

5. Nachweis der Eigenschaften von $S_\varkappa^{(m)}$. Die Formel (12) ist, wie bekannt, die Grundlage der Lagrangeschen Interpolationsformel, die zum Ausdruck bringt, daß für ein beliebiges Polynom P_{m-1} von höchstens $|m-1|$tem Grad:

$$(12') \qquad \sum_{\iota=1}^{m} P_{m-1}(z_\varkappa^{(\iota)}) \frac{\beta_\varkappa^{(m)}(z)}{(z - z_\varkappa^{(\iota)}) \cdot \beta_\varkappa'^{(m)}(z_\varkappa^{(\iota)})} \equiv P_{m-1}(z).$$

Dies gilt, gleichgültig wie die $z_\varkappa^{(\iota)}$ ausgewählt werden. Setzt man in (12′) für P_{m-1} der Reihe nach $\alpha_0, \alpha_1, \ldots, \alpha_{m-1}$ ein, so folgt sofort aus (12), daß für *diese* Indizes $k = 0$ bis $m-1$:

$$(13) \qquad \int \alpha_k(z) d S_\varkappa^{(m)}(x) = \sum_{\iota=1}^{m} \alpha_k(z_\varkappa^{(\iota)}) \mathfrak{s}_\varkappa^{(\iota)} = \int \alpha_k(z) d V_\varkappa(x) = c_\varkappa^{(k)}.$$

Damit ist ein Teil der Behauptung 3 und, wenn man $k=0$ in (13) einsetzt, die Behauptung 2 erwiesen.

Ist $k \geqq m$, höchstens gleich $2m-1$, so gilt für die oben mit (11) getroffene, spezielle Wahl von $\beta_\varkappa^{(m)}$ immer noch (13), wie folgende Überlegung zeigt. Man sucht zu jedem $\alpha_\varkappa$ für $m \leqq k \leqq 2m-1$ den Quotienten Q und Rest R bei Division durch $\beta_\varkappa^{(m)}$:

$$\alpha_k = Q\,\beta_\varkappa^{(m)} + R, \qquad (k = m, m+1 \ldots 2m-1)$$

wobei Q und R höchstens vom Grade $m-1$ sind und R mit α_k an allen Nullstellen von $\beta_\varkappa^{(m)}$ übereinstimmt. Wegen (11″) ist daher

$$\int \alpha_k(z)\,d\,V_\varkappa(x) = \int Q(z)\,\beta_\varkappa^{(m)}(z)\,d\,V_\varkappa(x) + \int R(z)\,d\,V_\varkappa(x) = \int R(z)\,d\,V_\varkappa(x),$$

d. h. das Integral über α_k ist auf ein solches über R, also über ein Polynom von höchstens $(m-1)$-ten Grade zurückgeführt. Auf R läßt sich wieder (12′) anwenden, und da $R(z_\varkappa^{(\iota)}) = \alpha_k(z_\varkappa^{(\iota)})$, so ist (13) auch für die weiteren Indizes $k = m$ bis $2m-1$ und damit die Behauptung 3 in vollem Umfang erwiesen.

Aus (13) folgt natürlich, mit Rücksicht auf den wiederholt erwähnten Zusammenhang zwischen den α_k und beliebigen Polynomen, daß der zweite Teil der Gleichung auch besteht, wenn an Stelle von α_k beiderseits des zweiten Gleichheitszeichens ein *beliebiges* Polynom von höchstens $(2m-1)$-tem Grad gesetzt wird. Wählen wir das Polynom $P(z) = [(z-z_\varkappa^{(1)})(z-z_\varkappa^{(2)}) \ldots (z-z_\varkappa^{(m-1)})]^2$, das vom $(2m-2)$-ten Grad und durchweg nicht negativ ist, so bleibt links in (13) lediglich $\mathfrak{z}_\varkappa^{(m)} \cdot P(z_\varkappa^{(m)})$ und rechts das Integral über das positive $P(z)$: Also muß $\mathfrak{z}_\varkappa^{(m)}$ positiv sein und in gleicher Weise jedes andere $\mathfrak{z}_\varkappa^{(\iota)}$. Dies ist der Beweis für die erste der oben angeführten Behauptungen.

Um endlich die letzte zu beweisen, leiten wir durch partielle Integration ab:

$$(14) \qquad 0 = \int \alpha_k(z)\left[d\,V_\varkappa(x) - d\,S_\varkappa^{(m)}(x)\right] = \int \alpha_k'(z)\left[S_\varkappa^{(m)}(x) - V_\varkappa^{(m)}(x)\right].$$

Auch diese Gleichung bleibt, da sie für $k=0$ bis $2m-1$ gilt, bestehen, wenn an Stelle von α_k' ein beliebiges Polynom von höchstens $(2m-2)$-tem Grad gesetzt wird. Hätte also $S_\varkappa^{(m)} - V_\varkappa$ nicht mehr als $2m-2$ Zeichenwechsel $z_1, z_2 \ldots$, so könnte man das Differenzenprodukt $(z-z_1)(z-z_2)\ldots$ in (14) an Stelle von α_k' einsetzen und bekäme einen Integranden von konstantem Vorzeichen, dessen Integral — im Gegensatz zu (14) — nicht verschwinden kann. Damit ist, über die Behauptung (4) hinaus, gezeigt, daß *zwei beliebige Verteilungen, die in den ersten $2m$ Momenten übereinstimmen, wenigstens $2m-1$ Schnittpunkte aufweisen müssen.*

Die besondere Bedeutung dieses Satzes gerade für die Treppen $S_\varkappa^{(m)}$ liegt darin, daß ja eine derartige Linie in jedem horizontalen und jedem vertikalen Stück nur *einmal* von einer andern, monoton ansteigenden Linie getroffen werden kann. Da nun eine m-fache Treppe nur $2m-1$ solche Stücke besitzt (die Teile der Horizontalen in der Höhe null und eins ungerechnet), so sieht man, daß sie *in jedem Stück* einmal geschnitten werden muß. Wir können diesem, für uns grundlegenden Ergebnis folgende Form geben:

Eine Verteilung $V_\varkappa(x)$, die selbst wenigstens m Wachstumstellen aufweist, wird durch Angabe ihrer ersten $2m$ Momente in ihrem ganzen Verlauf, an jeder Stelle x, auf ein bestimmtes beschränktes Gebiet eingeengt: Liegt der zu x nach (5) gehörige Wert von z zwischen $z_\varkappa^{(\iota)}$ und $z_\varkappa^{(\iota+1)}$, so liegt $V_\varkappa(x)$ zwischen $S_\varkappa^{(m)}(z_\varkappa^{(\iota)}-0)$ und $S_\varkappa^{(m)}(z_\varkappa^{(\iota+1)}+0)$; fällt z mit einem der $z_\varkappa^{(\iota)}$ zusammen, so liegt $V_\varkappa(x)$ zwischen $S_\varkappa^{(m)}(z_\varkappa^{(\iota)}+0)$ und $S_\varkappa^{(m)}(z_\varkappa^{(\iota)}-0)$. Der Ordinatenunterschied zwischen einer Verteilungslinie $V_\varkappa(x)$ und einer ihrer zugeordneten Treppen $S_\varkappa^{(m)}(x)$ kann nie größer sein als eine Stufenhöhe der letzteren.

6. Die einer Gaußschen Verteilung Φ zugeordneten Treppen $\Psi^{(m)}$. Die in § 3, Gl. (15) definierte Gaußsche Verteilung mit dem Mittelwert a und der Streuung s:

$$(15)\qquad V(x) = \Phi(z) = \Phi\left(\frac{x-a}{s\sqrt{2}}\right) = \frac{1}{s\sqrt{2\pi}}\int^{x} e^{-\frac{(x-a)^2}{2s^2}}\,dx$$

nimmt hinsichtlich der hier in Rede stehenden Probleme eine besondere Stellung ein. Zunächst reduziert sich ihre reelle Adjunkte $g(u)$ auf das *erste Glied*; denn nach (1) ist:

$$(15')\qquad g(u) = \frac{1}{s\sqrt{2\pi}}e^{-\frac{s^2u^2}{2}}\int e^{-(x-a)u-\frac{(x-a)^2}{2s^2}}\,dx = \frac{1}{\sqrt{\pi}}\int e^{-\left(\frac{su}{\sqrt{2}}+\frac{x-a}{s\sqrt{2}}\right)^2} d\left(\frac{x-a}{s\sqrt{2}}\right) = 1.$$

Es sind also alle zu Φ gehörigen Koeffizienten $c^{(1)} = c^{(2)} = c^{(3)} = \ldots = 0$. Hieraus ergeben sich naturgemäß einfache Beziehungen für die nach Abschnitt 4 der Verteilung $\Phi(z)$ zugeordneten Treppen $\Psi^{(m)}$.

Führen wir die Gaußsche Verteilungsdichte $\varphi^{(0)}$ und ihre Ableitungen $\varphi', \varphi'', \ldots, \varphi^{(m)}$ ein:

$$(16)\qquad \varphi^0(z) = \frac{1}{\sqrt{\pi}}e^{-z^2},\quad \varphi' = \frac{1}{\sqrt{\pi}}(-2z)e^{-z^2},\quad \varphi'' = \frac{1}{\sqrt{\pi}}(4z^2-2)e^{-z^2},\ldots;$$

so bestätigt man leicht, daß die in Klammer gesetzten Polynome sich nur durch konstante, d. i. von z unabhängige, Faktoren von den Hermiteschen unterscheiden:

$$(16')\qquad \varphi^{(m)} = 2^m\cdot m!\,\frac{1}{\sqrt{\pi}}e^{-z^2}\alpha_m(z).$$

Daraus folgt für die in (10′) zur Bestimmung der $\varrho^{(\iota)}$ eingeführten Koeffizienten $A^{(\iota,\lambda)}$ in unserm Fall $dV(x) = \varphi(z)dz$:

(16″) $$A^{(\iota,\lambda)} = \int \alpha_\iota \alpha_\lambda \varphi\, dz = \int \alpha_\iota \varphi^{(\lambda)} dz \cdot \frac{1}{2^\lambda \cdot \lambda!} = \int \alpha_\lambda \varphi^{(\iota)} \frac{1}{2^\iota\, \iota!}\, dz$$

und dieser Ausdruck ist, wie ι- bzw. λ-fache partielle Integration zeigt, null für $\iota \neq \lambda$. Demnach sind in den linearen Gleichungen (10″) sämtliche rechten Seiten null, daher auch alle ϱ und es bleibt $\beta^{(m)} = \alpha_m$, d. h. *die Stufen der Treppe $\Psi^{(m)}$ liegen an den Nullstellen des m-ten Hermiteschen Polynoms.*

Es kommt nun weiter gar nicht darauf an, auch die Stufenhöhen $\psi^{(\iota)}$ von $\Psi^{(m)}$ nach Gl. (12) zu bestimmen, wenn wir nur den Verlauf der $\Psi^{(m)}$ für große m verfolgen wollen. Wir erinnern uns an den in Abschnitt 2 dieses Paragraphen abgeleiteten Satz über die Verdichtung der Nullstellen der α_m. Da die Φ-Linie die Treppe $\Psi^{(m)}$ in jeder Stufe schneiden muß, kann eine Stufenhöhe nicht größer sein, als der Anstieg der Φ-Linie bis zur nächsten Stufe, also nicht größer als $\frac{1}{s\sqrt{2\pi}}$ mal dem in Abschn. 3 bestimmten größten Abstand zweier Nullstellen von α_m. Nach den Ergebnissen des vorangehenden Abschnittes kann sich aber die eine Verteilung darstellende Linie niemals mehr als um die Höhe einer Treppenstufe von der Treppe selbst unterscheiden. Wir haben also den Satz: *In dem Bereich $|z| \leqq z_0 < \sqrt{2m+1}$ gilt*

(17) $$|\Phi(z) - \Psi^{(m)}(x)| \leqq \psi^{(\iota)} < \frac{1}{2s}\sqrt{\frac{2\pi}{2m+1-z_0^2}}$$

für jedes m, mit der in (5) festgelegten Zuordnung zwischen z und x.

§ 5.

Beweis der Behauptung III.

1. Die Koeffizienten der Adjunkten von $W_n(x)$. Wenn wir die Gl. (2) und (3) des vorangehenden Paragraphen zusammenhalten, so gewinnen wir für die Adjunkte q_n des n-ten Integralproduktes W_n den Ausdruck:

(1) $$q_n(u) = \prod_{\varkappa=1}^{n}\left[c_\varkappa^{(0)} + c_\varkappa^{(1)}\left(\frac{s_\varkappa\sqrt{2}}{r_n}u\right) + c_\varkappa^{(2)}\left(\frac{s_\varkappa\sqrt{2}}{r_n}u\right)^2 + c_\varkappa^{(3)}\left(\frac{s_\varkappa\sqrt{2}}{r_n}u\right)^3 + \ldots\right].$$

Andererseits setzen wir $q_n(u)$ entsprechend (3) von § 4, da die Streuung von W_n gleich $^1/_2$ ist:

(1′) $$q_n(u) = k_n^{(0)} + k_n^{(1)} u + k_n^{(2)} u^2 + \ldots.$$

Den Koeffizienten $k_n^{(m)}$ von u^m in (1′) erhält man somit, indem man die Summe aller Produkte von der Form

(1″) $$c_1^{(\mu_1)} c_2^{(\mu_2)}, \ldots, c_n^{(\mu_n)}\, s_1^{\mu_1} s_2^{\mu_2}, \ldots, s_n^{\mu_n} \left(\frac{\sqrt{2}}{r_n}\right)^m$$

mit $\mu_1 + \mu_2 + \dots + \mu_n = m$ bildet. Eine obere Schranke für den Betrag von $k_n^{(m)}$ wird sich also dann ergeben, wenn man in dem Ausdruck rechts in (1) sowohl die $c_\varkappa^{(\mu)}$ als die $s_\varkappa$ durch obere Schranken für ihre absolute Beträge ersetzt. Mit Rücksicht auf die Bedingung (b 2) von § 3 und Gl. (3″) in § 4 ersetzen wir daher unter Beachtung von (3a) in § 4 das Produkt rechts in (1) durch

$$(2)\qquad \left[1 + c^{(3)}\left(\frac{S\sqrt{2}}{r_n}u\right)^3 + c^{(4)}\left(\frac{S\sqrt{2}}{r_n}u\right)^4 + \dots\right]^n = (1+D)^n$$
$$= 1 + \binom{n}{1}D + \binom{n}{2}D^2 + \dots + D^n.$$

Der Koeffizient von u^m in der Entwicklung des Ausdruckes (2) liefert die gesuchte obere Schranke für $|k_n^{(m)}|$, falls $m > 1$.

Da D mit einem Glied dritter Ordnung in u beginnt, braucht man, um den Koeffizient von u^m zu finden, in der Entwicklung der n-ten Potenz rechts in (2) nur bis zum Glied $D^{\frac{m}{3}}$ zu gehen. Wir nehmen nun an, es sei

$$(3')\qquad n > \frac{2m}{3},$$

dann wachsen die Binomialkoeffizienten in (2) bis zum letzten in Betracht kommenden und der Ausdruck (2) wird noch überschätzt, wenn wir den Koeffizienten des letzten beibehaltenen Gliedes vor alle setzen, also

$$(3'')\qquad \binom{n}{\frac{m}{3}}\left[1 + D + D^2 + \dots + D^{\frac{m}{3}}\right]$$

für (2) nehmen. Der Ausdruck in der eckigen Klammer in (3″) ist von n nur noch insofern abhängig, als r_n in D vorkommt. Fassen wir die Glieder, die u^m ergeben sollen, zusammen, so hebt sich r_n^m im Nenner heraus und wir erhalten:

$$(3)\qquad |k_n^{(m)}| < \binom{n}{\frac{m}{3}}\frac{R_m\sqrt{2}^m}{r_n^m} < \frac{R_m}{\left(\frac{m}{3}\right)!}\,\frac{n^{\frac{m}{3}}\sqrt{2}^m}{r_n^m} < \frac{R_m}{\left(\frac{m}{3}\right)!}\,[\varepsilon(n)]^{\frac{m}{2}}, \qquad (m > 1)$$

wo R_m eine von n unabhängige, $\varepsilon(n)$ eine nach Voraussetzung (b 2) mit wachsendem n gegen null gehende Größe bezeichnet. Man sieht also:

Der Koeffizient $k_n^{(m)}$ von u^m in der Entwicklung der reellen Adjunkten des n-ten Integralproduktes geht mit wachsendem n und konstantem $m > 1$ gegen null.

Es ist weiter nicht schwer, den Ausdruck R_m zu bestimmen, wenn man beachtet, daß jede Potenz D^k aus lauter Produkten der Form

$$(4')\qquad \left(\frac{S\sqrt{2}}{r_n}\right)^{\mu_1+\mu_2+\dots+\mu_k}\cdot c^{(\mu_1)}\,c^{(\mu_2)}\cdot\ldots\cdot c^{(\mu_k)}$$

besteht, wobei die μ die Werte 3, 4, 5, ... durchlaufen. Somit ist

$$(4'') \quad R_m = S^m \sum_{\varkappa=1}^{\frac{m}{3}} \sum_{\mu} c^{(\mu_1)} c^{(\mu_2)} \ldots c^{(\mu_k)} = S^m \left(\frac{1}{4} + \frac{1}{c^2 e}\right)^{\frac{m}{2}} \sum_{\varkappa=1}^{\frac{n}{3}} C^{\varkappa} \sum_{\mu} \frac{1}{\left(\frac{\mu_1}{2}\right)! \ldots \left(\frac{\mu_k}{2}\right)!}.$$

Da μ mindestens gleich 3, so ist jeder Summand der rechts zuletzt stehenden Summe sicher kleiner als 1! Die Zahl der Summanden ist gleich der Zahl der Kombinationen zur Summe $m - 2\varkappa$, gebildet aus $\varkappa$ Zahlen der Reihe 1, 2, 3, ..., d. i.

$$(4''') \qquad \binom{m - 2\varkappa - 1}{\varkappa - 1} < \binom{m}{\varkappa - 1} \qquad (m \geqq 3).$$

Setzt man diesen Wert für die Summe über die μ in (4″) ein und beachtet, daß $C^{\varkappa}$ höchstens gleich $C^{\frac{m}{3}}$ wird, so erhält man:

$$(4) \quad R_m < S^m \left(\frac{1}{4} + \frac{1}{c^2 e}\right)^{\frac{m}{2}} C^{\frac{m}{3}} \sum_{\varkappa=1}^{m+1} \binom{m}{\varkappa - 1} = S^m \left(1 + \frac{4}{c^2 e}\right)^{\frac{m}{2}} C^{\frac{m}{3}} = C'^m.$$

Dieser Ausdruck wächst im allgemeinen mit m, aber schwächer als $\left(\frac{m}{3}\right)!$, womit die aus Gl. (2) in § 4 folgende Konvergenz jeder einzelnen Entwicklung (1′) bestätigt wird. Man kann (3) und (4) zusammenfassen in

$$(5) \qquad |k_n^{(m)}| < \frac{C'^m}{\left(\frac{m}{3}\right)!} \varepsilon^{\frac{m}{2}}(n), \qquad \text{für} \quad m \geqq 3.$$

Für $m = 0, 1, 2$ gilt natürlich, unabhängig von n:

$$(5') \qquad k_n^{(0)} = 1, \qquad k_n^{(1)} = k_n^{(2)} = 0.$$

2. Die den W_n zugeordneten Treppen $T_n^{(m)}$. Nach § 4, 4 gehört zu jeder Verteilung, die mehr als m Wachstumstellen und eine ganze Funktion zur reellen Adjunkten besitzt, eine bestimmte m-stufige Treppe, die in den ersten $2m$ Koeffizienten der Adjunkten mit der gegebenen übereinstimmt. Von den Integralprodukten W_n ist nun leicht zu sehen, daß die Zahl ihrer Wachstumstellen mit n *ins Unendliche gehen* muß, wenn sie nicht überhaupt schon für jedes $n > n_0$ unendlich ist. Das letztere tritt ein, wenn V_{n_0} unendlich viel Wachstumstellen hat. Haben aber sämtliche $V_\varkappa$ nur endlich viele, und zwar $n_\varkappa$, so geht aus der Darstellung Gl. (3) in § 3 hervor, daß die Zahl der Wachstumstellen von W_n gleich der Summe $n_1 + n_2 + \cdots + n_n$ ist. Da nach Voraussetzung (a 2) jedes $V_\varkappa$ mindestens zwei Wachstumstellen haben muß, ist damit die Behauptung erwiesen. Es leuchtet auch ein, daß infolge des Wachsens von r_n mit $\sqrt{n}$ die Wachstumstellen von $W_n(x)$ sich im allgemeinen in jedem endlichen Bereich unbeschränkt verdichten.

Nehmen wir nun an, es sei ein festes m vorgeschrieben und zunächst $n > m$ gewählt, so daß es sicher eine zu W_n zugeordnete Treppe $T_n^{(m)}$ gibt: die Konvergenz der Entwicklung von q_n ist ja mit Gl. (5) (wie schon vorher durch die Darstellung (2) in § 4) nachgewiesen. Die m Stufen von $T_n^{(m)}$ sind die Nullstellen eines Polynoms $\gamma_n^{(m)}$ (vgl. § 4, 4):

$$(6) \qquad \gamma_n^{(m)} = \alpha_m + \sigma_n^{(m-1)} \alpha_{m-1} + \sigma_n^{(m-2)} \alpha_{m-2} + \dots + \sigma_n^{(0)} \alpha_0,$$

wobei die $\sigma_n^{(\iota)}$ die Lösungen der m linearen Gleichungen:

$$(6') \qquad \sum_{\iota=0}^{m-1} B_n^{(\iota,\lambda)} \sigma_m^{(\iota)} = -B_n^{(m,\lambda)} \qquad (\lambda = 0, 1, 2, \dots, m-1)$$

mit der Definition

$$(6'') \qquad B_n^{(\iota,\lambda)} = \int \alpha_\iota(z)\, \alpha_\lambda(z)\, d W_n(z) \qquad (\iota, \lambda = 0, 1, 2, \dots, m)$$

bedeuten. Die Größen z und x, die allgemein nach (5) in § 4 zusammenhängen, sind jetzt wegen $a' = 0$ und $s'^2 = \frac{1}{2}$ einander gleich.

Jedes der Produkte $\alpha_\iota \alpha_\lambda$ mit $\iota + \lambda \leqq 2m - 1$ kann man durch einen linearen Ausdruck in den $\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_{2m-1}$ darstellen, wobei aber α_0 nur in den Produkten mit $\iota = \lambda$ auftreten kann. Denn andernfalls könnten die in (16″), § 4 behandelten Ausdrücke von der Form $\int \alpha_\iota \alpha_\lambda \varphi\, dz$ unmöglich für $\iota \neq \lambda$ verschwinden, weil $\int \alpha_0 \varphi\, dz = 1$ und $\int \alpha_\iota \varphi\, dz = 0$ für $\iota \neq 0$ aus (16″), § 4 folgt. Da nun rechts in (6′) die oberen Indizes von B niemals gleich werden, kommen hier nur α_ι mit $\iota > 0$ und, nach Integration über $d W_n$, nur $k_n^{(\iota)}$ mit $\iota > 0$ in Betracht. Für diese Koeffizienten gilt aber Gl. (5) und (5′), d. h. man kann bei festem m durch Vergrößerung von n die rechte Seite der Gl. (6′) vermöge des Faktors $\varepsilon(n)$ beliebig klein machen.

In der Determinante des Gleichungssystems (6′) stehen nur in der Hauptdiagonale Elemente mit gleichen oberen Indizes. Diese liefern bei Entwicklung des Produktes $\alpha_\iota \cdot \alpha_\iota$, wie man sich, etwa durch Ausführung der in (16″), § 4 angedeuteten Integration überzeugen kann, das Glied $\alpha_0 : 2^\iota \cdot \iota!$. Alle anderen Elemente der Determinante führen wieder zu Größen, die den Faktor $\varepsilon(n)$ enthalten und sonst nur von m abhängig sind. Die Determinante unterscheidet sich also bei wachsendem n beliebig wenig von dem positiven Wert $2^{-(1+2+\dots+m-1)}[1 \cdot 1!\, 2! \dots (m-1)!]^{-1}$. Das Ergebnis dieser Überlegung ist, daß die $|\sigma_m^{(\iota)}|$ *bei festem m und wachsendem n unter jeden vorgegebenen Betrag* sinken.

Weitere Folge des Verhaltens der $\sigma_m^{(\iota)}$ ist, daß der Unterschied zwischen dem Polynom $\gamma_n^{(m)}$ und dem Hermiteschen Polynom α_m in der Weise mit wachsendem n sich der Null nähert, daß sämtliche Koeffizienten von $\gamma_n^{(m)}$ gegen die entsprechenden von α_m konvergieren. Dabei hat $\gamma_n^{(m)}$ nach dem in § 4, 4 gegebenen Beweis durchaus reelle Nullstellen. Da nun die Wurzeln einer algebraischen Gleichung *stetige* Funktionen der Gleichungskoeffizienten sind, so können wir schließen: Die m Wurzeln von $\gamma_n^{(m)}$, also die Stufen der Treppe $T_n^{(m)}$ nähern sich mit wachsendem n und festem m mehr und mehr den Nullstellen des Hermiteschen Polynoms α_m und erhalten den

Satz (1): *Man kann stets zu gegebenem m ein $n_1 > m$ so bestimmen, daß für $n > n_1$ der Abstand zwischen den Nullstellen von α_m und den Wachstumstellen von $T_n^{(m)}$ kleiner als ein beliebig vorgegebener Betrag wird.*

Die Stufen*höhen* der Treppe $T_n^{(m)}$ sind nach Gleichung (12) des § 4 durch

$$(6''') \qquad t_n^{(\iota)} = \int \frac{\gamma_n^{(m)}(z)\, dW_n(z)}{(z - z_n^{(\iota)})\, \gamma_n'^{(m)}(z)_n^{(\iota)}} \qquad (\iota = 1, 2, \ldots, m)$$

gegeben, worin $z_n^{(\iota)}$ die Nullstellen von $\gamma_n^{(m)}$ bezeichnen. Was hier als Integrand neben dem Differential dW_n erscheint, ist ein Polynom vom $(m-1)$-ten Grad, das sich wieder als Linearform in den $\alpha_0, \alpha_1, \ldots, \alpha_{m-1}$ ausdrücken läßt. Daher ist jedes $t_n^{(\iota)}$ ein linearer Ausdruck in den $k_n^{(0)}, k_n^{(1)}, \ldots, k_n^{(m-1)}$. Die Koeffizienten dieses Ausdruckes hängen lediglich von $\gamma_n^{(m)}$ ab und dieses ist nach (6) durch die $k_n^{(0)}, k_n^{(1)}, \ldots, k_n^{(2m-1)}$ vollständig bestimmt. Da beide Abhängigkeiten *stetig* sind, sind auch die $t_n^{(\iota)}$ *stetige Funktionen der Koeffizienten* $k_n^{(0)} \ldots k_n^{(2m)}$.

Oben ist gezeigt worden, daß jeder einzelne Koeffizient $k_n^{(\iota)}$ für $\iota > 1$ mit wachsendem n gegen null geht. Andrerseits sind nach § 4, 6 die Koeffizienten der Gaußschen Funktion Φ, bis auf den ersten, der immer konstant $= 1$ ist, gleich null. Wenn daher m eine feste Zahl ist, muß der Wert von $t_n^{(\iota)}$, der eine stetige Funktion der $k_n^{(\iota)}$ ist, mit wachsendem n stetig jenem Wert zustreben, den man erhält, wenn man $k_n^{(0)} = 1$ und $k_n^{(\iota)} = 0$ für $\iota > 1$ setzt, oder kurz dem Wert $\psi^{(\iota)}$, der sich für die Stufenhöhen der Treppe $\Psi^{(m)}$ einer Gaußschen Verteilung ergibt. Wir haben somit den

Satz (2): *Man kann stets zu einem gegebenen m ein $n_2 > m$ so bestimmen, daß für $n > n_2$ der Unterschied zwischen der ι-ten Stufenhöhe $t_n^{(\iota)}$ von $T_n^{(m)}$ und der analogen Stufenhöhe $\psi^{(\iota)}$ von $\Psi^{(m)}$, oder auch der zwischen der Treppenhöhe von $T_n^{(m)}$ in der ι-ten Stufe und der entsprechenden Treppenhöhe der Treppe $\Psi^{(m)}$ beliebig klein wird.*

3. Abschluß des Beweises. Um nunmehr zu zeigen, daß $W_n(x)$ mit wachsendem n gleichmäßig gegen $\Phi(x)$ konvergiert, wählen wir zunächst ein $x_0 > 0$ so groß, daß $\Phi(x)$ für $|x| > x_0$ sich um weniger als $\frac{\varepsilon}{2}$ von dem Grenzwert 0 bzw. 1 unterscheidet, also so daß

$$(7) \qquad |1 - \Phi(x_0)| < \frac{1}{4x_0^2} < \frac{\varepsilon}{2}, \qquad x_0 > \sqrt{\frac{1}{2\varepsilon}}.$$

Hierauf wählen wir ein m so, daß die in § 4, 6 ermittelte obere Grenze für die Stufenhöhe $\psi^{(\iota)}$ der Treppe $\Psi^{(m)}$ nicht größer als $\frac{\varepsilon}{8}$ wird; also, da jetzt $s = 1:\sqrt{2}$, nach (17) in § 4:

$$(7') \qquad \psi^{(\iota)} \leqq \sqrt{\frac{\pi}{2m+1-x_0^2}} < \frac{\varepsilon}{8}, \quad 2m > x_0^2 - 1 + \pi\left(\frac{8}{\varepsilon}\right)^2.$$

Liegt einmal m fest, so bestimmen wir nach dem ersten der im vorangehenden Abschnitt gewonnenen Sätze n_1 so, daß für $n > n_1$ der Abstand $d_n^{(\iota)}$ zwischen einer Nullstelle von α_m und der zugehörigen Stufe von $T_n^{(\iota)}$ kleiner ist als der kleinste Abstand zweier aufeinanderfolgender Nullstellen von α_m, also so, daß (vgl. § 4, 3):

$$(7'') \qquad d_n^{(\iota)} < \frac{\pi}{\sqrt{2m+1}} \quad \text{für} \quad n > n_1, \ \iota = 1, 2, \ldots, m.$$

Ferner sei nach dem zweiten Satz des vorangehenden Abschnittes n_2 derart gewählt, daß der Unterschied zwischen der ι-ten Stufenhöhe von $T_n^{(m)}$ und von $\Psi^{(m)}$ nicht größer als $\varepsilon : 8(m+1)$ ist, also

$$(7''') \qquad |\mathfrak{t}_n^{(\iota)} - \psi^{(\iota)}| < \frac{\varepsilon}{8(m+1)} \quad \text{für} \quad n > n_2, \ \iota = 1, 2, \ldots, m.$$

Bedeutet dann n_0 die größere der beiden Zahlen n_1 und n_2, so behaupten wir, daß

$$(8) \qquad |W_n(x) - \Phi(x)| < \varepsilon \quad \text{für} \quad n > n_0$$

gleichmäßig für alle Werte von x von $-\infty$ bis $+\infty$.

Denn zunächst kann sich innerhalb des Bereiches $|x| \leqq x_0$ die Höhe der Treppe $T_n^{(m)}$ von der Höhe der Treppe $\Psi^{(m)}$ an der gleichen Stelle, vermöge (7'') nicht um mehr unterscheiden, als um eine Stufenhöhe $\psi^{(\iota)}$, vermehrt um die Höhendifferenz der beiden Treppen in Punkten gleicher Stufenzahl. Es ist also nach (7') und (7''')

$$(8') \qquad |T_n^{(m)}(x) - \Psi^{(m)}(x)| < \frac{\varepsilon}{8} + \frac{m\varepsilon}{8(m+1)}.$$

Weiters kann nach dem Schlußsatz in § 4, 5 der Abstand zwischen W_n und $T_n^{(m)}$ nicht größer sein als das größte $\mathfrak{t}_n^{(\iota)}$ und der zwischen Φ und $\Psi^{(m)}$ nicht größer als das größte $\psi^{(\iota)}$, also nach (7') und (7'''):

$$(8'') \qquad |W_n(x) - T_n^{(m)}(x)| < \frac{\varepsilon}{8} + \frac{\varepsilon}{8(m+1)},$$

$$(8''') \qquad |\Phi(x) - \Psi^{(m)}(x)| < \frac{\varepsilon}{8}.$$

Durch Addition von (8'), (8'') und (8''') ergibt sich

$$(8\text{a}) \qquad |W_n(x) - \Phi(x)| < \frac{\varepsilon}{2} \quad \text{für} \quad |x| \leqq x_0,$$

womit ein Teil von (8) erledigt ist. Nun geht aber $W_n(x)$ ebenso wie $\Phi(x)$ außerhalb $|x| < x_0$ *monoton* gegen null bzw. gegen 1 und unterscheidet sich nach (7) und (8a) an der Stelle x_0 höchstens um ε von diesen Grenzwerten. Daher folgt (8) auch für alle Werte von x für die $|x| > x_0$.

Damit ist die Behauptung III *des* § 3 *in vollem Umfang erwiesen.*

4. Rückblick auf den Beweisansatz in § 3. In § 3 haben wir zum Beweise der Behauptung III die Gleichung (25) abgeleitet:

$$\overline{W}_n(x) - \Phi(x) = -\frac{1}{2\pi}\Re\int\frac{e^{-uxi}}{ui}\left[p_n(u) - e^{-\frac{u^2}{4}}\right]du, \tag{9}$$

worin $p_n(u)$ die komplexe Adjunkte von W_n bedeutete und nach (12) als Produkt der analog definierten komplexen Adjunkten $f_\varkappa$ der $V_\varkappa(x)$ erschien. Daß der Klammerausdruck rechts in (9) für jeden endlichen Bereich von u mit wachsendem n gleichmäßig gegen null geht, war eine Folgerung aus dem Satz I des § 1. Zum Nachweis der Behauptung III war es aber noch notwendig zu zeigen, daß das über den unendlichen Bereich von u erstreckte Integral (9) der Null zustrebt. Es fragt sich, in welchem Zusammenhang der von uns jetzt direkt geführte Beweis für III mit dieser Formulierung steht.

Zunächst ist zu sagen, daß die Übertragung der in § 4, 1 und 2 über die reellen Adjunkten angestellten Überlegungen sofort ergeben würde: Auch jedes $f_\varkappa$ und jedes p_n ist eine *ganze* Funktion der Variablen u. Die Koeffizienten von $f_\varkappa$ sind, bis auf Zahlenfaktoren, die die Konvergenz herbeiführen, unmittelbar die „Momente" der Verteilung $V_\varkappa(x)$. Weiter würde eine analoge Untersuchung wie die in § 5, 1 zeigen, daß die Koeffizienten der Entwicklung von p_n mit wachsendem n zwar nicht gegen null, aber *gegen die Koeffizienten der Entwicklung von* $e^{-\frac{u^2}{4}}$ konvergieren. Es liegt darin, wie man sieht, nur ein sehr spezieller Fall des Satzes I, nämlich der Fall, daß die dort behandelten $f_\varkappa$ in der Umgebung der ausgezeichneten Stellen $a_\varkappa$ Potenzentwicklungen zulassen (Konvergenz in der ganzen Ebene ist durchaus nicht erforderlich, vgl. § 6, 2), bei denen die Koeffizienten jeder Ordnung eine beschränkte Folge bilden. Unter dieser Voraussetzung kann man die Behauptung I geradezu dahin aussprechen, daß die Koeffizienten der Entwicklung des Produktes $p_n(u)$ gegen die der Entwicklung von $e^{-u^2:4}$ konvergieren. So ordnet sich der erste, im Abschnitt 1 des vorstehenden Paragraphen enthaltene Teil unseres Beweises den früheren Betrachtungen unter.

Der zweite Teil des Beweises, der sich auf die Beziehungen zwischen einer Verteilung $V_\varkappa(x)$ und den ihr zugeordneten Treppen $S_\varkappa^{(m)}(x)$, bzw.

zwischen $W_n(x)$ und $T_n^{(m)}(x)$ erstreckt, enthält nun gerade das, was zum Nachweis des Verschwindens von (9) noch ausstand. Denn im wesentlichen ist gezeigt worden: *Wenn die Momente m-ter Ordnung der Verteilungen $W_1(x), W_2(x), \ldots$ gegen das m-te Moment von $\Phi(x)$ konvergieren, so geht $W_n(x) - \Phi(x)$ gegen null*[7]). Im Anschluß an (9) ausgesprochen, lautet der Satz: Das Integral rechts in (9) hat den Grenzwert null, sobald die Potenzentwicklung des Klammerausdruckes gliedweise gegen null konvergiert. Der Wert des Klammerausdruckes selbst geht, wie wir wissen (Beispiel in § 3, 7), *nicht* im ganzen Integrationsbereich gegen null.

Man darf aber nicht außer acht lassen, daß der eben ausgesprochene Satz über das Integral (9) nur vermöge der besonderen Bedeutung des Klammerwertes für die gesuchte Funktion $W_n - \Phi$, nicht etwa für jedes Integral $\int e^{-xui} G(u)\,du$, wo G eine ganze Funktion ist, gilt. (Damit eine solche Funktion G die Entwicklung einer reellen Adjunkten sei, müssen die Koeffizienten so beschaffen sein, daß z. B. (10) in § 4 positiv definit wird.) Er besteht beispielsweise nicht ohne weiteres in dem Fall der Gl. (19‴) von § 3, die wir unter Verwendung von (15), § 3 in der Form schreiben:

$$(9''') \qquad w_n(x) - \frac{1}{\sqrt{\pi}} e^{-x^2} = \frac{1}{2\pi} \Re \int e^{-uxi} \left[p_n(u) - e^{-\frac{u^2}{4}} \right] du.$$

Denn wenn sämtliche $V_\varkappa(x)$ und daher auch die $W_n(x)$ Treppen von endlicher Stufenzahl sind, so hat die Ableitung w_n von W_n für jedes n an einer mit n wachsenden Anzahl von Stellen den Wert ∞. Andererseits haben wir für den besonderen Fall, in dem sämtliche $v_\varkappa(x)$ und daher auch die $w_n(x)$ Funktionen von beschränkter Schwankung sind, das Bestehen von (9‴) oben direkt nachgewiesen (§ 3, 6).

§ 6.

Korollare zu den Sätzen des § 3.

1. Schwächeres und stärkeres Anwachsen von $\Sigma s_\varkappa^2$. In der dritten der Voraussetzungen (b 2) in § 3 ist gefordert worden, daß die Summe der Streuungen stärker als mit der $^2/_3$ Potenz von n ins Unendliche wächst. Die Überlegungen in § 5, 1, insbesondere Gl. (3) zeigen,

---

[7]) Diesen Satz, der die wesentliche Schwierigkeit des Beweises von III in sich schließt, hat P. L. Tschebyschef 1873 ohne Ableitung angegeben, Journal de Liouville, sèr. II, t. 19. Einen Beweis auf Grund der Theorie der Kettenbruch-Entwicklung bestimmter Integrale gab zuerst Markoff 1884 (russisch), dann Tschebyschef selbst (Oeuvres, Bd. II, S. 421 und Bd. II, S. 443). Vgl. auch die Monographie von A. Wassilief und N. Delaunay, „P. L. Tschebyschef", Leipzig 1900, Kap. VI und VII.

daß für den Beweis der Behauptung III dieses Verhalten von r_n^2 nicht nur hinreichend, sondern auch notwendig ist. Dagegen wird diese Bedingung im Falle der Behauptungen IV und V gegenstandslos, da hier zufolge der besonderen Voraussetzungen (über das Vorhandensein eines Paares von endlichen Wachstumstellen bzw. über die beschränkte Gesamtschwankung) die $s_\varkappa^2$ eine von null verschiedene untere Schranke besitzen. Dies ist in den Abschn. 5 und 6 des § 3 ausdrücklich nachgewiesen worden. Wir können daher aussprechen:

IV$_1$, V$_1$. *Die Voraussetzung* (a 2) *und die letzte der drei Ungleichungen* (b 2) *kann im Falle der Behauptungen* IV *und* V (*auch für den Beweis der Gl.* (III)) *fortgelassen werden.*

2. Einschränkung der Bedingungen im Unendlichen. Mit den Voraussetzungen (a 1) und (b 1) in § 3 sind den Verteilungen $V_\varkappa$ bestimmte Einschränkungen für das Verhalten von $dV_\varkappa$ im Unendlichen auferlegt worden, deren Bedeutung hauptsächlich darin bestand, daß die Existenz der m-ten Momente jeder Verteilung und die Beschränktheit ihrer Folge bei festem m sichergestellt wurde. Nun ist bei dem Beweis der Behauptungen IV und V in § 3 überhaupt nur von den Momenten erster, zweiter und dritter Ordnung Gebrauch gemacht worden. Wir haben also zunächst das Korollar:

IV$_2$, V$_2$. *Die Behauptungen* IV *und* V *des* § 3 *gelten auch unter Weglassung der Voraussetzungen* (a 1) *und* (b 1), *also ohne jede weitere Einschränkung für* $dV_\varkappa$ *im Unendlichen, falls zu den* (*ersten beiden*) *Voraussetzungen* (b 2) *noch die hinzugefügt wird, daß die Folge der dritten Momente der* $V_\varkappa$ *beschränkt ist*:

$$\int (x-a_\varkappa)^3\, dV_\varkappa(x) < K\,. \tag{1}$$

Hierbei genügt es, die Momente, wie es hier angeschrieben ist, auf die ursprüngliche Variable x bzw. $x-a_\varkappa$ zu beziehen, statt auf $(x-a_\varkappa):s_\varkappa$, da ja die $s_\varkappa^2$ eine von null verschiedene untere Schranke haben.

Bei dem Beweis der Behauptung III in § 4 und 5 ist von den Voraussetzungen (a 1) und (b 1) vor allem zu dem Zwecke Gebrauch gemacht worden, um die Entwickelbarkeit der reellen Adjunkten $g_\varkappa$ einer Verteilung $V_\varkappa$ sowie die Beschränktheit der Koeffizientenfolge $c_\varkappa^{(m)}$ bei festem m nachzuweisen. Es ist klar, daß man in (a 1) und (b 1) noch etwas schwächer wachsende Funktionen statt $e^{c^2x^2}$ einsetzen könnte, ohne die Folgerungen zu beeinträchtigen. Darüber hinaus sieht man auch leicht, daß die Konvergenz der Entwicklung in der ganzen Ebene nicht wesentlich ist, sondern die in der Umgebung des Nullpunktes genügt. Aber man kann sogar durch eine kleine Abänderung des Beweisganges die Voraussetzung (a 1) darauf beschränken, daß die Momente beliebig hoher

Ordnung *endlich* sind. Denn dem Zusammenhang zwischen den Koeffizienten $c_\varkappa^{(m)}$ der $V_\varkappa$ und $k_n^{(m)}$ der Integralprodukte, wie er in § 5, 1 dargelegt wurde, liegt die folgende *Identität in den Hermiteschen Polynomen* zugrunde:

$$(2)\qquad \alpha_m\left(\sum_{\varkappa=1}^{n}\frac{x_\varkappa-a_\varkappa}{r_n}\right)=\left(\frac{\sqrt{2}}{r_n}\right)^m\sum_{\mu_1,\mu_2,\ldots,\mu_n} s_1^{\mu_1}\,\alpha_{\mu_1}(z_1)\cdot s_2^{\mu_2}\,\alpha_{\mu_2}(z_2)\ldots s_n^{\mu_n}\,\alpha_{\mu_n}(z_n),$$

wobei die Summe rechts so zu bilden ist, daß $\mu_1+\mu_2+\ldots+\mu_n=m$, und

$$(3)\qquad z_\varkappa=\frac{x_\varkappa-a_\varkappa}{s_\varkappa\sqrt{2}},\qquad r_n^2=2\,(s_1^2+s_2^2+\ldots+s_n^2)$$

gilt. Daraus folgt sofort für die Koeffizienten der Adjunkten von W_n:

$$(3')\ k_n^{(m)}=\int\alpha_m(u)\,dW_n(u)=\int\int\ldots\int\alpha_m\left(\sum_{\varkappa=1}^{n}\frac{x_\varkappa-a_\varkappa}{r_n}\right)dV_1(x_1)\ldots dV_n(x_n)$$

$$=\left(\frac{\sqrt{2}}{r_n}\right)^m\sum_{\mu} s_1^{\mu_1}s_2^{\mu_2}\ldots s_n^{\mu_n}\int\alpha_{\mu_1}(z_1)\,dV_1(x)\cdot\int\alpha_{\mu_2}(z_2)\,dV_2(x)\ldots\int\alpha_{\mu_n}(z_n)\,dV_n(z)$$

und daher, mit Rücksicht auf die Bedeutung von $c_\varkappa^{(m)}$, die Formel (1″) in § 5 für $k_n^{(m)}$. Man beweist (2) am einfachsten, indem man für Verteilungen, die den Voraussetzungen des § 3 genügen, so wie es oben geschehen ist, das Multiplikationsgesetz für die reellen Adjunkten ableitet und dann beiderseits die Koeffizienten gleicher Potenzen von u gleichsetzt. Hinterher kann man sich von den Voraussetzungen bezüglich der $V_\varkappa$ unabhängig machen, da (2) lediglich eine Eigenschaft der Hermiteschen Polynome darstellt.

Die Erweiterung des Geltungsbereiches unseres Satzes, zu der wir so gelangt sind, lautet:

III$_2$. *Die Behauptung* III *des* § 3 *gilt auch, wenn an Stelle der Voraussetzung* (a 1) *nur verlangt wird, daß für jede Verteilung* $V_\varkappa$ *die Momente* $\int z^m\,dV_\varkappa(x)$ *jeder Ordnung* m *existieren und an Stelle der Voraussetzung* (b 1), *daß die Folgen der Momente sämtlicher Verteilungen bei festem* m *beschränkt sind.*

Praktisch wird man von dieser Verallgemeinerung wohl selten Gebrauch machen, da die empirisch gegebenen Verteilungen im Endlichen begrenzt sind (und hierbei nicht mit dem Grenzwert null der Streuungen gerechnet werden muß), so daß die Voraussetzungen (a 1) und (b 1) von § 3 von vornherein erfüllt sind, während für „theoretische" Verteilungen in der Regel der Satz IV$_2$ oder V$_2$ eintreten wird.

3. Erweiterung durch Einführung linearer Verbindungen statt einfacher Summen. Die Definition des Integralproduktes in § 3, 1 kann in symmetrischer Form wie folgt geschrieben werden:

$$(4) \qquad dW_n(u) = \int\int \ldots \int dV_1(x_1)\, dV_2(x_2) \ldots dV_n(x_n),$$

wenn hinzugefügt wird, daß die Integration rechts *bei konstanter Summe* $x_1 + x_2 + \ldots + x_n = x$ durchzuführen ist und u eine bestimmte lineare Funktion von u darstellt. Wie wir gesehen haben, geht $W_n(u)$, wenn diese lineare Funktion die durch (7) und (7′) in § 3 gegebene Form besitzt, mit wachsendem n gegen $\Phi(u)$. Oft wird nun die Aufgabe in der Form gestellt, daß die Integration (4) nicht bei konstanter Summe, sondern unter Festhaltung einer gewissen *linearen Verbindung* der $x_\varkappa$ zu bilden ist:

$$(4') \qquad h_1^{(n)} x_1 + h_2^{(n)} x_2 + \ldots + h_n^{(n)} x_n = x, \qquad (n = 1, 2, \ldots).$$

Hierbei haben wir die n Koeffizienten auch noch von dem Index n des Integralproduktes abhängig angesetzt. Liegen sämtliche $h_\varkappa^{(n)}$ *in endlichen Grenzen*, so folgt durch Transformation der Variablen sofort, daß $\lim_{n=\infty} W_n(u) = \Phi(u)$, wenn jetzt

$$(4'') \qquad u = r_n x + b_n \quad \text{mit} \quad r_n^2 = 2\sum_{\varkappa=1}^{n} h_\varkappa^2 s_\varkappa^2, \quad b_n = \sum_{\varkappa=1}^{n} h_\varkappa a_\varkappa$$

gesetzt wird und $a_\varkappa$ bzw. $s_\varkappa^2$ Mittelwert und Streuung von $V_\varkappa$ bezeichnen. Natürlich kann man hier, wie auch in dem Fall, daß alle $h = 1$, auch eine beliebige *endliche* Transformation mit der Variablen u vornehmen.

Andererseits ist es klar, daß wenn die $h_\varkappa^{(n)}$ die genannte Bedingung nicht erfüllen, die Anwendung unserer Resultate allgemein überhaupt nicht möglich ist[8]). Nur der folgende Satz kann ausgesprochen werden:

III_3, IV_3, V_3: *Die Behauptungen* III, IV *und* V *des* § 3 *gelten auch für die in* (4) (4′) (4″) *gegebene erweiterte Definition des Integralproduktes, falls an Stelle der Voraussetzungen* (a 2) *und* (b 2) *folgende treten:*

$$(\mathrm{a}\,2''') \qquad h_\varkappa s_\varkappa \neq 0$$

$$(\mathrm{b}\,2''') \qquad |h_\varkappa a_\varkappa| < A, \qquad |h_\varkappa^2 s_\varkappa^2| < S^2, \qquad \frac{n^{\frac{2}{3}}}{\sum_{\varkappa=1}^{n} h_\varkappa^2 s_\varkappa^2} < \varepsilon(n).$$

4. Verteilungen in mehreren Dimensionen. Die Sätze des § 3 gestatten auch eine — für die Anwendungen sehr wichtige — Ausdehnung auf den Fall von Verteilungen mit *mehreren unabhängigen Variablen.* Es sei wieder wie in § 2, 4 das Zeichen $\bar{x}$ (spr. x-Vektor) ein Symbol für k Veränderliche $x^{(1)}, x^{(2)}, \ldots, x^{(k)}$. Unter einer *Verteilung* $V_\varkappa(\bar{x})$ *im*

---

[8]) Ein Beispiel für einen solchen Fall, in dem W_n nicht gegen Φ konvergiert, siehe bei F. Hausdorff, Leipz. Ber. 1901, S. 166.

k-dimensionalen Raum verstehen wir eine reelle, in bezug auf jede der k Variablen monotone, nicht abnehmende Funktion, die für $x^{(1)} = x^{(2)} = \ldots = x^{(k)} = -\infty$ den Wert null, für $x^{(1)} = x^{(2)} = \ldots = x^{(k)} = \infty$ den Wert 1 besitzt und an jeder Stelle ihren oberen Grenzwert annimmt. Durch zwei Vektoren $\bar{a}$, $\bar{b}$, die der Bedingung $a^{(\iota)} \leqq b^{(\iota)}$ $(\iota = 1, 2, \ldots, k)$ genügen, wird ein von Parallelen zu den Koordinatenebenen begrenzter Raumteil bestimmt, dessen Koordinaten $\xi^{(\iota)}$ die Ungleichungen $a^{(\iota)} < \xi^{(\iota)} \leqq b^{(\iota)}$ erfüllen. Zu jedem solchen Raumteil gehört ein gewisser durch $V_\varkappa(\bar{x})$ bestimmter Betrag, den wir mit $V_\varkappa(\bar{a}, \bar{b})$ bezeichnen wollen und der dadurch definiert ist, daß $V_\varkappa(\bar{x}) = V_\varkappa(-\infty, \bar{x})$ und daß für einen Raumteil $\bar{a}$, b, der in zwei oder mehrere andere Raumteile der betrachteten Art zerfällt, die Beträge sich addieren. Sind die Koordinaten $a^{(\iota)}$ und $b^{(\iota)}$ gegeben, so stellt sich $V(\bar{a}, \bar{b})$ als eine Summe von 2^k Funktionswerten $V(\bar{x})$ dar, wobei die Komponenten der $\bar{x}$ aus den sämtlichen Permutationen der $a^{(\iota)}$ und $b^{(\iota)}$ mit abwechselndem Vorzeichen bestehen, z. B. für $k = 3$:

$$\begin{aligned} V(\bar{a}, \bar{b}) = {} & V(b^{(1)}, b^{(2)}, b^{(3)}) - V(a^{(1)}, b^{(2)}, b^{(3)}) + V(a^{(1)}, a^{(2)}, b^{(3)}) - V(b^{(1)}, a^{(2)}, b^{(3)}) \\ & + V(a^{(1)}, b^{(2)}, a^{(3)}) - V(b^{(1)}, b^{(2)}, a^{(3)}) + V(b^{(1)}, a^{(2)}, a^{(3)}) - V(a^{(1)}, a^{(2)}, a^{(3)}). \end{aligned}$$

Das Stieltjessche Integral einer stetigen Funktion $\varphi(\bar{x})$ über den Raum $\bar{a}$, b, das wir durch $\int\limits_{(\bar{a}, \bar{b})} \varphi(\bar{x})\, d V_\varkappa(\bar{x})$ bezeichnen, ist dann definiert als der Grenzwert einer Summe, deren Glieder die Form $\varphi(\bar{\xi}_\iota) \cdot V_\varkappa(\bar{a}_\iota, \bar{b}_\iota)$ haben, wobei die Raumteile $\bar{a}_\iota$, $\bar{b}_\iota$ durch eine unbegrenzt enger werdende Intervallteilung des Bereiches $\bar{a}$, $\bar{b}$ entstehen. Hat $\varphi(\bar{x})$ Unstetigkeitsstellen erster Art, so soll unter dem Integral wieder der *obere* Grenzwert der Summe verstanden werden. Bei Integralen, die über den ganzen unendlichen Raum zu erstrecken sind, lassen wir die Bezeichnung der Grenzen wieder weg.

Mittelwert einer k-dimensionalen Verteilung ist der k-dimensionale Vektor $\bar{a}_\varkappa$:

$$(5) \qquad \bar{a}_\varkappa = \int \bar{x}\, d V_\varkappa(\bar{x}), \qquad a^{(\iota)} = \int x^{(\iota)}\, d V_\varkappa(\bar{x}) \qquad (\iota = 1, 2, \ldots, k);$$

Streuung die k-reihige symmetrische Matrix $\bar{\bar{s}}_\varkappa$ mit den Komponenten

$$(5') \qquad s_\varkappa^{(\iota, \lambda)} = \int (x^{(\iota)} - a^{(\iota)})(x^{(\lambda)} - a^{(\lambda)})\, d V_\varkappa(\bar{x}),$$

deren quadratische Form positiv definit ist, weil

$$(5'') \qquad \int \Big\{ \sum_\iota u^{(\iota)} (x^{(\iota)} - a_\varkappa^{(\iota)}) \Big\}^2 d V_\varkappa(\bar{x}) > 0,$$

wenn der Fall, daß $V_\varkappa(\bar{x})$ nur auf einer einzigen Ebene des k-dimensionalen Raumes Wachstumstellen hat, ausgeschlossen wird (Voraussetzung (a 2)).

Als *komplexe Adjunkte* definieren wir:

(6) $$f_n(\bar{a}_\varkappa + \bar{u}) = \int e^{\sum_\iota u^{(\iota)}\left(x^{(\iota)} - a_\varkappa^{(\iota)}\right) i} \, dV_\varkappa(\bar{x}).$$

Diese Funktionen entsprechen den in § 2, 4 gemachten Annahmen für $u = 0$ (abgesehen vom Faktor 2):

(6') $$f_\varkappa(0) = 1; \quad \frac{\partial f_\varkappa(0)}{\partial u^{(\iota)}} = 0; \quad \frac{\partial^2 f_\varkappa(0)}{\partial u^{(\iota)} \partial u^{(\lambda)}} = -s_\varkappa^{(\iota,\lambda)}; \quad \left| \frac{\partial^3 f_\varkappa(0)}{\partial u^{(\iota)} \partial u^{(\lambda)} \partial u^{(\mu)}} \right| < C'.$$

Den übrigen Voraussetzungen des § 1, nämlich $|f_\varkappa| < 1$ für $\bar{u} \neq 0$, und den Bedingungen hinsichtlich des Verhaltens im Unendlichen genügen die $f_\varkappa$ bei analogen Einschränkungen für die $V_\varkappa$, wie sie in § 3 angegeben wurden.

Unter dem n-ten *Integralprodukt* der $V_\varkappa(\bar{x})$ verstehen wir die k-dimensionale Verteilung

(7) $$W_n(\bar{z}) = \int\int \ldots \int V_n(\bar{x} - \bar{x}_1 - \bar{x}_2 - \ldots - \bar{x}_{n-1}) \, dV_{n-1}(\bar{x}_{n-1}) \, dV_{n-2}(\bar{x}_{n-2}) \ldots dV_1(\bar{x}_1),$$

wobei die Additionen nach den Regeln der Vektorrechnung, d. h. komponentenweise auszuführen sind. Den Zusammenhang von $\bar{z}$ und $\bar{x}$ wollen wir mit

(7') $$\bar{x} = \sqrt{n}\,\bar{z} + \bar{b}_n, \quad \bar{z} = \frac{x - \bar{b}_n}{\sqrt{n}}, \quad \bar{b}_n = \bar{a}_1 + \bar{a}_2 + \ldots + \bar{a}_n$$

ansetzen und dabei der Einfachheit halber annehmen, daß die Summe der $s_\varkappa^{(\iota,\lambda)}$ nicht nur, wie es der dritten Voraussetzung (b 2) in § 3 entsprechen würde, stärker als mit $n^{\frac{2}{3}}$, sondern genau mit n anwächst. Es seien also die Größen

(7'') $$h_n^{(\iota,\lambda)} = \frac{r_n^{(\iota,\lambda)}}{n} = \frac{1}{n} \sum_\varkappa s_\varkappa^{(\iota,\lambda)}$$

durchaus endlich (wie in § 2, 4). Aus (7') und (7'') folgt dann, daß der Mittelwert von W_n null, die Streuung gleich $\bar{\bar{h}}_n$, d. h. durch die Matrix der $h_n^{(\iota,\lambda)}$ gegeben ist, deren quadratische Form wieder positiv definit sei. Für die komplexe Adjunkte $p_n(\bar{u})$ des Integralproduktes von $W_n(\bar{x})$ ergibt sich danach:

(8) $$\begin{aligned} p_n(\bar{u}) &= \int e^{\sum_\iota u^{(\iota)} x^{(\iota)} i} \, dW_n(\bar{x}) \\ &= \int\int \ldots \int e^{\sum_{\varkappa,\iota} \left(x_\varkappa^{(\iota)} - a_\varkappa^{(\iota)}\right) \frac{u^{(\iota)}}{\sqrt{n}} i} \, dV_1(\bar{x}_1) \ldots dV_n(\bar{x}_n) \\ &= f_1\left(\bar{a}_1 + \frac{\bar{u}}{\sqrt{n}}\right) \cdot f_2\left(\bar{a}_2 + \frac{\bar{u}}{\sqrt{n}}\right) \ldots f_n\left(\bar{a}_n + \frac{\bar{u}}{\sqrt{n}}\right). \end{aligned}$$

Nach Satz I_4 von § 2 ist somit

$$(8') \qquad \lim_{n=\infty} p_n(\bar{u}) = e^{-\frac{1}{2}\sum_{\iota,\lambda} h_n^{(\iota,\lambda)} u^{(\iota)} u^{(\lambda)}}.$$

Wenn man die quadratische Form $\Sigma h_n^{(\iota,\lambda)} z^{(\iota)} z^{(\lambda)}$ durch Transformation der $z^{(\iota)}$ in $\zeta^{(\iota)}$ in eine Summe von Quadraten verwandelt:

$$(8'') \qquad \sum_{\iota,\lambda} h_n^{(\iota,\lambda)} z^{(\iota)} z^{(\lambda)} \equiv \sum_{\iota=1}^{k} H_n^{(\iota)} \zeta^{(\iota)^2},$$

so nimmt (8′) die Form an:

$$(8''') \qquad \lim_{n=\infty} p_n(\bar{z}) = e^{-\frac{1}{2}\sum_{\iota} H_n^{(\iota)} \zeta^{(\iota)^2}}$$

Andererseits findet man, daß für eine Gaußsche Verteilung in k Dimensionen:

$$(9) \quad \Phi\left(\frac{\xi_1}{\sqrt{2H_1}}\right)\Phi\left(\frac{\xi_2}{\sqrt{2H_2}}\right)\cdots\Phi\left(\frac{\xi_k}{\sqrt{2H_k}}\right) \equiv \sqrt{\frac{1}{(2\pi)^k H_1 H_2 \ldots H_k}} \int^{\bar{\xi}} e^{-\sum_{\iota}\frac{\xi_\iota^2}{2H_\iota}} d\xi_1 \ldots d\xi_k$$

die nach (6) gebildete komplexe Adjunkte den Wert

$$(9') \qquad e^{-\frac{1}{2}\sum_{\iota} H_\iota \xi_\iota^2}$$

besitzt. Demnach wird in Analogie zu den eindimensionalen Sätzen des § 3 zu schließen sein, daß — bei geeigneten Annahmen über das Verhalten der $V_\varkappa(\bar{x})$ — $W_n(\bar{x})$ gegen einen Ausdruck analog (9) konvergiert. Wir stellen daher den Satz hin:

III_4, IV_4, V_4. *Wenn man die Voraussetzungen* (a 1) (a 2) *und* (b 1) (b 2) *des* § 3 *auf* k*-dimensionale Verteilungen sinngemäß erweitert, so ergibt sich für das mit* (7) *definierte* n*-te Integralprodukt*:

$$(III_4) \qquad \lim_{n=\infty} W_n(\bar{x}) = \Phi\left(\frac{\xi^{(1)}}{\sqrt{2H_n^{(1)}}}\right)\Phi\left(\frac{\xi^{(2)}}{\sqrt{2H_n^{(2)}}}\right)\cdots\Phi\left(\frac{\xi^{(k)}}{\sqrt{2H_n^{(k)}}}\right);$$

im Falle die $V_\varkappa(\bar{x})$ *nach dem Volumelement* $dx^{(1)}\,dx^{(2)}\ldots dx^{(k)}$ *genommene Ableitungen* $v_\varkappa(\bar{x})$ *haben, die beschränkt und von gleichmäßig beschränkter Schwankung sind, auch für die analog gebildete Ableitung* $w_n(\bar{x})$ *von* $W_n(\bar{x})$:

$$(IV_4) \qquad \lim_{n=\infty} w_n(\bar{x}) = \frac{e^{-\sum_{\iota}\frac{\xi^{(\iota)^2}}{2H_n^{(\iota)}}}}{\sqrt{(2\pi)^k H_n^{(1)} H_n^{(2)} \ldots H_n^{(k)}}};$$

endlich, wenn die $V_n(\bar{x})$ *nur in den Punkten eines* k*-dimensionalen Gitters Wachstumstellen haben, wobei nach jeder Koordinatenrichtung* *wenigstens einmal zwei benachbarte Punkte endliche Sprunggrößen aufweisen, auch für die Sprunggrößen* $\mathfrak{w}_n(\bar{x})$ *von* $W_n(\bar{x})$:

$$(\mathrm{V}_4) \qquad \lim_{n=\infty} \sqrt{n}^k\, \mathfrak{w}_n(\bar{x}) = \frac{e^{-\sum\limits_{\iota} \frac{\xi^{(\iota)2}}{2 H_n^{(\iota)}}}}{\sqrt{(2\pi)^k H_n^{(1)} \ldots H_n^{(k)}}}.$$

Dabei gehen die Koordinatenrichtungen der $\xi^{(\iota)}$ *und die Werte* $H_n^{(\iota)}$ *durch Hauptachsen-Transformation aus der Matrix* $h_n^{(\iota,\lambda)}$ *hervor, deren Komponenten nach* (7″) *durch die Streuungen* $s_\varkappa^{(\iota,\lambda)}$ *der gegebenen Verteilungen bestimmt sind.*

Man kann die Ausdrücke rechts in (III_4), (IV_4) und (V_4) auch in den ursprünglichen Koordinaten $x^{(\iota)}$ schreiben, wenn man die Koeffizienten der Matrix bestimmt, deren Hauptachsenrichtungen mit denen der $h_n^{(\iota,\lambda)}$ übereinstimmen, während die Hauptachsenlängen die halben reziproken Werte haben. (Ein Beispiel hierzu s. § 7, 5.)

5. Geometrische Deutung der Sätze I bis IV. Denken wir uns in einem n-dimensionalen Koordinatensystem $x_1, x_2, \ldots, x_n$ eine skalare Funktion des Ortes p_n gegeben, die dadurch gekennzeichnet ist, daß sie sich als ein *Produkt von n Funktionen der einzelnen Koordinaten* $f_1(x_1) \cdot f_2(x_2) \ldots f_n(x_n)$ darstellt. Die $f_\varkappa(x_\varkappa)$ sollen den Voraussetzungen des § 1 genügen, also im wesentlichen die Eigenschaft besitzen, daß sie an einer regulären Stelle $x_\varkappa = a_\varkappa$ den Wert 1 und sonst überall kleinere Beträge annehmen. Verfolgt man dann *vom Punkte* $\bar{x} = \bar{a}$ *aus in der Richtung „unter* 45°" (nämlich in der durch $x_1 = x_2 = \ldots = x_n$ gegebenen Richtung) den Wert von p_n, so findet man nach Satz I, daß dieser Wert, als Funktion des in einem bestimmten Maßstab $u = \frac{r_n}{\sqrt{n}} z$ gemessenen Abstandes z von $\bar{a}$, sich um so mehr dem e^{-u^2} nähert, je größer die Zahl n ist — gleichgültig, wie die einzelnen Funktionen $f_\varkappa$ aussehen. Haben sämtliche $f_\varkappa$ an der Stelle $x_\varkappa$ die gleiche zweite Ableitung, so bleibt der Maßstabfaktor von n unabhängig. Er wird insbesondere gleich 1, wenn die $f_\varkappa$ so normiert werden, daß die zweiten Ableitungen gleich -2 sind. Satz II ergänzt die Aussage nur durch eine Hinzufügung über das Verhalten von $p_n(u)$ im Unendlichen.

Es ist klar, daß man statt der Richtung „unter 45°" *auch eine beliebige andere* wählen kann, die etwa durch $x_1 : x_2 : \ldots : x_n = h_1^{(n)} : h_2^{(n)} : \ldots : h_n^{(n)}$ gegeben wird, wobei sämtliche $h_\varkappa^{(n)}$ in endlichen Grenzen liegen. Dies bedeutet ja nur eine unwesentliche Transformation der unabhängig Veränderlichen der $f_\varkappa$ und hätte analog dem Korollar III_3 und IV_3 dieses Paragraphen als Korollar in § 2 formuliert werden können, wenn es sich nicht eben von selbst verstünde.

Nehmen wir nun an, es seien an Stelle der $f_\varkappa(x_\varkappa)$ Funktionen $v_\varkappa(x_\varkappa)$ der Koordinaten gegeben, die den Voraussetzungen des § 3 genügen, d. i. im wesentlichen: sie sind Ableitungen von „Verteilungen", die endliche Momente, wenigstens bis zur dritten Ordnung, besitzen, und von beschränkter Schwankung. Durch die Momente erster Ordnung, oder die Mittelwerte $a_\varkappa$, wird wieder ein Punkt $\bar{a}$ des n-dimensionalen Raumes bestimmt. Durch diesen Punkt, und dann in beliebigen Abständen z von diesem Punkt, legen wir *Ebenen „unter* 45°" (nämlich solche, die durch $x_1 + x_2 + \ldots + x_n = x$ = konst. bestimmt werden), bilden auf jeder Ebene das *Integral des Produktes der Funktionen* $v_\varkappa(x_\varkappa)$, nämlich die Größe

$$w_n = \int \ldots \int\int v_1(x_1)\, v_2(x_2) \ldots v_n(x_n)\, dx_1\, dx_2 \ldots dx_{n-1}$$

und betrachten dieses w_n als Funktion des Abstandes $z = \frac{x}{\sqrt{n}}$ vom Punkte $\bar{a}$. Unser Satz IV besagt, daß diese Funktion, bis auf einen veränderlichen Maßstabfaktor $r_n : \sqrt{n}$, sich den Werten von $\frac{1}{\sqrt{\pi}} e^{-u^2}$ $\left(\text{mit } u = \frac{\sqrt{n}}{r_n} z\right)$ nähert, sobald n hinreichend wächst. Der Maßstabfaktor bleibt wieder von n unabhängig, sobald alle Verteilungen auf $a_\varkappa$ bezogen gleich große Momente zweiter Ordnung (gleiche Streuungen) haben und wird insbesondere gleich 1, wenn alle Streuungen gleich $^1/_2$ sind. Satz III kann aufgefaßt werden als eine Ergänzung zu IV für den Fall, daß die $v_\varkappa(x)$ nicht beschränkt, aber integrabel sind und daß demgemäß nicht über die oben betrachtete Funktion, sondern nur über ihr unbestimmtes Integral etwas ausgesagt werden kann.

Mit dem Korollar III_3 und IV_3 dieses Paragraphen ist die Ausdehnung der Betrachtung auf *beliebig geneigte Ebenen*, an Stelle der „unter 45°" stehenden, gegeben. Zusammenfassend können wir sagen:

Schreitet man im n-dimensionalen Raum von dem (in beiden Fällen verschieden definierten) *ausgezeichneten Punkte $\bar{a}$ längs irgend einer Geraden G fort, so stellt sich einerseits das Produkt p_n der Funktionen $f_\varkappa(x_\varkappa)$,* nämlich $p_n(u) = f_1(x_1) \cdot f_2(x_2) \ldots f_n(x_n)$ *als Funktion des Abstandes von $\bar{a}$ betrachtet, andererseits das Integralprodukt w_n der Funktionen $v_\varkappa$, erstreckt über die zu G senkrechten Ebenen:*

$$w_n(u) = \int \ldots \int\int v_1(x_1)\, v_2(x_2) \ldots v_n(x_n)\, dx_1\, dx_2 \ldots dx_{n-1},$$

als Funktion des Abstandes der Ebenen von $\bar{a}$ betrachtet, im Limes für $n = \infty$ als e^{-u^2} bzw. $\frac{1}{\sqrt{\pi}} e^{-u^2}$ dar.

Die genauen Voraussetzungen, Formulierungen usw. dieser Sätze sind in den vorangehenden Ausführungen enthalten.

Eine einfache und lehrreiche Deutung gestatten auch die in den Korollaren § 2, 4 und § 6, 4 gegebenen Erweiterungen für die Fälle mit *mehreren unabhängig Veränderlichen.* Wir müssen jetzt einen Raum von kn Dimensionen betrachten, in dem die Komponenten des Vektors $\bar{x}$ durch $x_\varkappa^{(\iota)}$ $(\varkappa = 1, 2, \ldots, n;\ \iota = 1, 2, \ldots, k)$ dargestellt werden. Anstelle der Geraden „unter 45°" im Falle des *einfachen Produktes* tritt jetzt der durch $k(n-1)$ Gleichungen der Form $x_1^{(\iota)} = x_2^{(\iota)} = \ldots = x_n^{(\iota)}$ $(\iota = 1, 2, \ldots, k)$ herausgehobene k-fach ausgedehnte Unterraum. In diesem stellt sich das Produkt p_n nach geeigneter Regulierung der Maßstäbe und nach entsprechender Drehung des Koordinatenkreuzes (Transformation der quadratischen Form in § 2, 4 auf eine Summe von Quadraten) im Limes für $n = \infty$ als ein Produkt von k Funktionen $e^{-H^{(\iota)}\cdot u^{(\iota)2}}$ dar. Analog sind im Falle der *Integralprodukte* die Integrale der Produkte über $k(n-1)$ dimensionale Räume zu erstrecken, die durch k Gleichungen der Form $x_1^{(\iota)} + x_2^{(\iota)} + \ldots + x_n^{(\iota)} = x^{(\iota)} =$ konst. definiert werden und als Funktion der k Variablen $x^{(\iota)}$ zu betrachten. Hierauf ergeben sich die $w_n(\bar{u})$, wieder nach Regulierung des Maßstabes und Transformation auf Hauptachsen, im Limes als Produkte, wie es Gl. (IV_4) oben zeigt.

Man sieht, daß sich hier neue Fragen anknüpfen lassen, die bei den einfachen Produkten trivial sind, bei Integralprodukten aber zu neuen, schwierigen Problemen führen; nämlich die Fragen, was bei Hinzutreten *beliebiger weiterer linearer Beziehungen* zwischen den kn Variablen entsteht. So hat beispielsweise Markoff für $k = 2$ mit den neu hinzutretenden Gleichungen $x_\varkappa^{(2)} = x_{\varkappa+1}^{(1)}$ $(\varkappa = 1, 2, \ldots, n-1)$ das Integralprodukt „verketteter" Größen untersucht und unter gewissen Voraussetzungen wieder die Konvergenz gegen das Exponentialgesetz gefunden[9]). In der vorliegenden Arbeit bleiben diese Untersuchungen, die, allgemein gefaßt, auf höhere Produktbildungen führen, unberücksichtigt.

II. Teil. Anwendungen in der Wahrscheinlichkeitsrechnung.

§ 7.

Die Binomialzahlen, der Bernoullische und Poissonsche Fall und seine Verallgemeinerung.

1. Die Binomialzahlen. Der einfachste Fall einer Anwendung der Sätze über das Integralprodukt liegt vor bei Betrachtung der Binomialzahlen $\binom{n}{x}$ als Funktion von x, im Limes für $n = \infty$. In der Wahrscheinlich-

[9]) Vgl. A. A. Markoff, Wahrscheinlichkeits-Rechnung, Dtsch. v. H. Liebmann, Leipzig 1912, Anhang II und III.

keitsrechnung tritt diese Aufgabe z. B. beim „Kopf- und Adlerspiel" auf, wenn man voraussetzt, daß die Wahrscheinlichkeit, beim Werfen einer Münze die eine oder andere Seite zu treffen, gleich $\frac{1}{2}$ ist. Man fragt nach der Wahrscheinlichkeit $\mathfrak{w}_n(u)$, bei n Würfen eine bestimmte Anzahl x von Kopfwürfen zu erreichen, wo x und u in einer noch festzusetzenden Beziehung stehen.

Setzen wir

$$\tag{1} \begin{aligned} V_1(x) = V_2(x) = V_3(x) = \ldots &= 0 \text{ für } x < 0, \\ &= \tfrac{1}{2} \quad " \quad 0 \leqq x < 1, \\ &= 1 \quad " \quad 1 \leqq x, \end{aligned}$$

so entsprechen diese Verteilungen den allgemeinen Bedingungen des § 3 und den besonderen Annahmen der Behauptung V. Denken wir uns der Adlerseite den Wert null, der Kopfseite den Wert 1 zugeschrieben, so stellt $V_\varkappa(x)$ die Wahrscheinlichkeit dar, beim $\varkappa$-ten Versuch „höchstens x zu treffen". Mittelwert und Streuung ergeben sich nach (4) in § 3, da $V_\varkappa$ nur die zwei Wachstumstellen 0 und 1 mit den Sprüngen $\mathfrak{v}_\varkappa(0) = \mathfrak{v}_\varkappa(1) = \frac{1}{2}$ besitzt, zu

$$\tag{1'} \begin{aligned} a_1 = a_2 = a_3 = \ldots &= 0 \cdot \tfrac{1}{2} + 1 \cdot \tfrac{1}{2} = \tfrac{1}{2}, \\ s_1{}^2 = s_2{}^2 = s_3{}^2 = \ldots &= (-\tfrac{1}{2})^2 \cdot \tfrac{1}{2} + (\tfrac{1}{2})^2 \tfrac{1}{2} = \tfrac{1}{4}. \end{aligned}$$

Daraus folgt nach (7′) und (7) in § 3:

$$\tag{1''} b_n = \frac{n}{2}, \qquad r_n^2 = \frac{n}{2}; \quad x = \sqrt{\frac{n}{2}}\, u + \frac{n}{2}, \quad u = \frac{x - \frac{n}{2}}{\sqrt{\frac{n}{2}}}.$$

Das n-te Integralprodukt $W_n(u)$ der Verteilungen (1) liefert daher die Wahrscheinlichkeit, bei n Würfen „höchstens die Summe $x = \sqrt{\frac{n}{2}}\, u + \frac{n}{2}$ zu treffen". Man erkennt das unmittelbar an der Gl. (3) des § 3, die hier als Definition des Integralproduktes zu gelten hat. Die Größe $\mathfrak{w}_n(u)$ selbst bedeutet die Wahrscheinlichkeit, gerade die Summe x in n Würfen zu erreichen. Andererseits kann man diese Wahrscheinlichkeiten in bekannter Weise durch die Binomialzahlen n-ter Ordnung bzw. durch deren von null bis zu dem angegebenen Wert von x erstreckte Summe ausdrücken. Gl. (V) und (III) des § 3 liefern daher die folgenden beiden Sätze:

$$\tag{2a} \mathfrak{w}(u) = \lim_{n=\infty} \sqrt{\frac{n}{2}}\, \mathfrak{w}_n(u) = \lim_{n=\infty} \frac{\sqrt{n}}{2^{n+\frac{1}{2}}} \binom{n}{\sqrt{\frac{n}{2}}\, u + \frac{n}{2}} = \frac{1}{\sqrt{\pi}}\, e^{-u^2} \equiv \varphi(u);$$

$$\tag{2b} W(u) = \lim_{n=\infty} W_n(u) = \lim_{n=\infty} \frac{1}{2^n} \sum_{x=0}^{\sqrt{\frac{n}{2}} u + \frac{n}{2}} \binom{n}{x} = \frac{1}{\sqrt{\pi}} \int^{u} e^{-x^2}\, dx \equiv \Phi(u).$$

In Worten besagt (2a): *Die Binomialzahlen* $\binom{n}{x}$ *bei festen* n *als Funktion von* $x - \frac{n}{2}$, *in gehörigem, mit* n *veränderlichem, Abszissen- und Ordinatenmaßstab aufgetragen, ergeben in ihren Endpunkten im Limes für* $n = \infty$ *die Gaußsche Linie* $\frac{1}{\sqrt{\pi}} e^{-u^2}$, *wobei der Zusammenhang zwischen* u *und* $x - \frac{n}{2}$ *durch* $(1'')$ *gegeben ist.*

In der üblichen Gestalt von *Näherungsformeln für große* n lauten (2a) und (2b):

$$(2') \quad \binom{n}{x} \sim \frac{2^{n+\frac{1}{2}}}{\sqrt{n}} \varphi\left(\sqrt{\frac{2}{n}}\, x - \sqrt{\frac{n}{2}}\right); \quad \sum_{z=0}^{x} \binom{n}{z} \sim 2^n \Phi\left(\sqrt{\frac{2}{n}}\, x - \sqrt{\frac{n}{2}}\right).$$

Die genaue Bedeutung von $(2')$ kann natürlich nur aus (2a) und (2b) entnommen werden.

Es ist bekannt, wie man (2) und $(2')$ durch direkte Rechnung unter Heranziehung des *Stirling*schen asymptotischen Ausdruckes für die Fakultäten ableitet. Hierzu mag bemerkt werden, daß die *Stirlingsche Formel sich als ein sehr spezieller Fall unseres Satzes* II über den Grenzwert eines Produktes von Funktionen darstellt. Denn man hat mit der Substitution $z - n = nx$:

$$(3) \qquad n! = \int_0 z^n e^{-z} dz = n^{n+1} e^{-n} \int_{-1} (1+x)^n e^{-nx} dx.$$

Setzt man $f_1 = f_2 = \ldots f = (1+x) e^{-x}$ für $x \geqq -1$
$= 0$ „ $x < -1$,

so genügen diese $f_\varkappa$ sämtlichen Voraussetzungen des § 1 und insbesondere ist $f(0) = 1$, $f'(0) = 0$, $f''(0) = -1$, also $r_n^2 = \frac{n}{2}$. Demnach liefert (II) auf $a = -\infty$, $b = \infty$ angewendet:

$$(3') \quad \lim_{n=\infty} \int f^n\left(\sqrt{\frac{2}{n}}\, u\right) du = \lim_{n=\infty} \sqrt{\frac{n}{2}} \int f^n(x) dx = \lim_{n=\infty} \sqrt{\frac{n}{2}}\, n!\, n^{-(n+1)} e^n = \sqrt{\pi},$$

also die Stirlingsche Formel:

$$(3'') \qquad n! \sim n^n e^{-n} \sqrt{2\pi n}.$$

2. Der Bernoullische Fall. Aus einer Urne, die schwarze und weiße Kugeln (oder die Zahlen Null und Eins) enthält, wird mit der Wahrscheinlichkeit $p > 0$ eine schwarze, und mit der Wahrscheinlichkeit $q = 1 - p > 0$ eine weiße Kugel gezogen. Die Verteilungen

$$(4) \qquad \begin{aligned} V_1(x) = V_2(x) = V_3(x) = \ldots &= 0 \text{ für } x < 0, \\ &= p \;\text{ „ }\; 0 \leqq x < 1, \\ &= 1 \;\text{ „ }\; 1 \leqq x \end{aligned}$$

genügen den allgemeinen Voraussetzungen des § 3 und den besonderen der Behauptung V. Man findet:

$$(4') \qquad \begin{aligned} &a_1 = a_2 = a_3 = \ldots = 0 \cdot p + 1 \cdot q = q, \qquad b_n = n q, \\ &s_1{}^2 = s_2{}^2 = s_3{}^2 = \ldots = q^2 \cdot p + p^2 \cdot q = pq, \qquad r_n^2 = 2 n p q, \\ &x = \sqrt{2 n p q}\, u + n q, \qquad u = \frac{x - n q}{\sqrt{2 n p q}}. \end{aligned}$$

Das n-te Integralprodukt $W_n(u)$ der Verteilungen (4) gibt also die Wahrscheinlichkeit, in n Zügen „höchstens die Summe $x = \sqrt{2 n p q u + n q}$ zu treffen", $\mathfrak{w}_n(u)$ die Wahrscheinlichkeit, gerade die Summe x zu gewinnen. Mit Rücksicht auf die direkte Darstellbarkeit dieser Wahrscheinlichkeiten durch den binomischen Lehrsatz liefern unsere Gl. (III) und (V):

$$(5a) \qquad \mathfrak{w}(u) = \lim_{n=\infty} \sqrt{2 n p q}\, \mathfrak{w}_n(u)$$

$$= \lim_{n=\infty} \sqrt{2 n p q}\, p^{np - \sqrt{2npq}}\, q^{nq + \sqrt{2npq}} \binom{n}{\sqrt{2 n p q}\, u + n q} = \frac{1}{\sqrt{\pi}} e^{-u^2} \equiv \varphi(u);$$

$$(5b) \qquad W(u) = \lim_{n=\infty} W_n(u) = \lim_{n=\infty} \sum_{x=0}^{\sqrt{2npq}\,u + nq} p^{n-x} q^x \binom{n}{x} = \frac{1}{\sqrt{\pi}} \int^{u} e^{-x^2} dx \equiv \Phi(u).$$

Dies ist der exakte Ausdruck der von *Laplace* gefundenen Lösung des Bernoullischen Problems. Als *Näherungsformeln für große* n lauten (5a) und (5b):

$$(5a') \qquad \binom{n}{x} p^{n-x} q^x \sim \frac{1}{\sqrt{2 n p q}} \varphi\left(\frac{x - n q}{\sqrt{2 n p q}}\right),$$

$$(5b') \qquad \sum_{z=0}^{x} \binom{n}{z} p^{n-z} q^z \sim \Phi\left(\frac{x - n q}{\sqrt{2 n p q}}\right).$$

Auch diese Gleichungen lassen sich, wie bekannt, unschwer mit Hilfe der Stirlingschen Formel ableiten. Noch einfacher gewinnt man sie (und natürlich auch die spezielleren (2a) und (2b)) durch folgende Überlegung. Aus der Definition

$$(6) \qquad \mathfrak{w}_n(u) = \binom{n}{x} p^{n-x} q^x, \qquad (x = 0, 1, 2, \ldots),$$

in der u und x durch die Beziehung $x = r_n u + b_n$ verknüpft sind, leitet man durch Übergang von x auf $x + 1$ ab:

$$(6') \qquad r_n \frac{\mathfrak{w}_n\left(u + \frac{1}{r_n}\right) - \mathfrak{w}_n(u)}{\frac{1}{r_n}} = -2 u r_n \mathfrak{w}_n(u) \frac{1 + \frac{p}{r_n u}}{1 + \frac{2 p u}{r_n} + \frac{2 p}{r_n^2}}.$$

Der Bruch rechts geht mit wachsendem n gegen 1, der Bruch links nähert sich dem Differentialquotienten der Funktion $\mathfrak{w}(u)$. Man hat also im Limes für $n=\infty$: $\mathfrak{w}'(u) = -2u\,\mathfrak{w}(u)$, woraus sofort (5a) folgt. — Für uns sind (5a) und (5b) nur ganz spezielle Beispiele für die Behauptung V des § 3.

3. Der Poissonsche Fall. Poisson hat die Bernoullische Aufgabe dahin erweitert, daß die Wahrscheinlichkeiten p, q eines schwarzen bzw. weißen Zuges nicht konstant, sondern *von Zug zu Zug wechselnd* angenommen werden. Wir haben also die Verteilungen:

$$
(7) \qquad \begin{aligned} V_\varkappa &= 0 \text{ für } \quad x < 0, \\ &= p_\varkappa \;\;,, \;\; 0 \leqq x < 1, \\ &= 1 \;\;\; ,, \;\; 1 \leqq x; \end{aligned} \qquad (\varkappa = 1, 2, 3, \ldots).
$$

Hierbei wird angenommen, daß die $p_\varkappa$ und $q_\varkappa$ von null und eins verschieden sind, womit wieder allen Voraussetzungen des § 3 und auch den speziellen Annahmen des Satzes V Genüge getan ist. Man findet

$$
(7') \qquad \begin{aligned} a_\varkappa &= 0\cdot p_\varkappa + 1\cdot q_\varkappa = q_\varkappa, & b_n &= \sum_{\varkappa=1}^{n} q_\varkappa, \\ s_\varkappa^2 &= q_\varkappa^2\cdot p_\varkappa + p_\varkappa^2\cdot q_\varkappa = p_\varkappa q_\varkappa, & r_n{}^2 &= 2\sum_{\varkappa=1}^{n} p_\varkappa q_\varkappa, \\ x &= \sqrt{2\Sigma p_\varkappa q_\varkappa}\,u + \Sigma q_\varkappa, & u &= \frac{x-\Sigma q_\varkappa}{\sqrt{2\Sigma p_\varkappa q_\varkappa}}. \end{aligned}
$$

Erteilen wir wieder der schwarzen Kugel den Wert null, der weißen den Wert 1, so gibt $W_n(u)$, für die Verteilungen (7) berechnet, die Wahrscheinlichkeit, „höchstens die Summe $x = \sqrt{2\Sigma p_\varkappa q_\varkappa}\,u + \Sigma q_\varkappa$ zu treffen“, $\mathfrak{w}_n(u)$ die Wahrscheinlichkeit, gerade diese Summe x zu ziehen. Eine direkte Darstellung von $\mathfrak{w}_n(u)$ in geschlossenen Ausdrücken ist hier nicht mehr möglich. Die Poissonschen Resultate sind in den beiden folgenden, aus unseren Gl. (III) und (V) entspringenden Gleichungen präzisiert:

$$
(8a) \qquad \mathfrak{w}(u) = \lim_{u=\infty} \sqrt{2\Sigma p_\varkappa q_\varkappa}\,\mathfrak{w}_n(u) = \frac{1}{\sqrt{\pi}} e^{-u^2} - \varphi(u),
$$

$$
(8b) \qquad W(u) = \lim_{n=\infty} W_n(u) = \frac{1}{\sqrt{\pi}} \int^{u} e^{-x^2} dx = \Phi(u);
$$

die als *Näherungsformeln für große* n lauten:

$$
(8a') \qquad \mathfrak{w}_n(u) = \mathfrak{w}_n'(x) \sim \frac{1}{\sqrt{2\Sigma p_\varkappa q_\varkappa}}\,\varphi\left(\frac{x-\Sigma q_\varkappa}{\sqrt{2\Sigma p_\varkappa q_\varkappa}}\right),
$$

$$
(8b') \qquad W_n(u) = W_n'(x) \sim \Phi\left(\frac{x-\Sigma q_\varkappa}{\sqrt{2\Sigma p_\varkappa q_\varkappa}}\right),
$$

wenn $\mathfrak{w}_n'(x)$ bzw. $W_n'(x)$ die Wahrscheinlichkeiten bezeichnen, in n Zügen genau x bzw. höchstens x als Summe zu ziehen. Alle Σ sind von $\varkappa = 1$ bis n zu nehmen.

Die von Poisson herrührende Ableitung, die mit geringen Varianten auch in den heutigen Lehrbüchern der Wahrscheinlichkeitsrechnung wiedergegeben wird, benutzt die schon Laplace bekannte Zurückführung des Integralproduktes auf einfache Produkte einschließlich der Umkehrformel, im Sinne unserer allgemeinen Überlegungen in § 3, 5[10]). Die in unseren Behauptungen I und II enthaltenen allgemeinen Resultate werden dann durch umständliche Abschätzung des in Gl. (17′) des § 3 auftretenden Integrals, in seiner besondern, dem vorliegenden Fall entsprechenden Gestalt, ersetzt.

4. Allgemeiner eindimensionaler Fall. Eine weitere Verallgemeinerung des Bernoullischen und Poissonschen Problems entsteht, wenn die einzelnen Urnen *nicht nur schwarze und weiße*, oder mit Null und Eins bezeichnete, Kugeln, *sondern auch* 2, 3, ..., m-*wertige* enthalten, und nach der Wahrscheinlichkeit gefragt wird, in n Zügen die Summe x oder „höchstens die Summe x" zu ziehen. So wird z. B. beim gewöhnlichen Zahlenlotto aus einer Urne, die die Zahlen 1, 2, 3, ... bis 90 enthält, gezogen und man kann nach der Wahrscheinlichkeit fragen, in n Einzel-Ziehungen einen bestimmten Durchschnitt $x : n$ zu erhalten. Hier ist jede Verteilung durch m nicht-negative Zahlen gegeben, deren Summe $\leqq 1$, nämlich durch die „Einzelwahrscheinlichkeiten" $\mathfrak{v}_\varkappa(0), \mathfrak{v}_\varkappa(1), \ldots, \mathfrak{v}_\varkappa(m-1)$, zu denen $\mathfrak{v}_\varkappa(m) = 1 - \mathfrak{v}_\varkappa(0) - \mathfrak{v}_\varkappa(1) - \ldots - \mathfrak{v}_\varkappa(m-1)$ hinzutritt. Jedes $\mathfrak{v}_\varkappa(x)$ stellt die Wahrscheinlichkeit dar, beim $\varkappa$-ten Zug die Zahl x zu ziehen. Unsere $V_\varkappa(x)$ sind definiert durch:

$$
(9)\qquad \begin{aligned}
V_\varkappa(x) &= 0, && x < 0,\\
&= \mathfrak{v}_\varkappa(0), && 0 \leqq x < 1,\\
&= \mathfrak{v}_\varkappa(0) + \mathfrak{v}_\varkappa(1), && 1 \leqq x < 2,\\
&= \mathfrak{v}_\varkappa(0) + \mathfrak{v}_\varkappa(1) + \mathfrak{v}_\varkappa(2), && 2 \leqq x < 3\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&= 1, && m \leqq x.
\end{aligned}
$$

[10]) Eine Ausnahme bildet H. Bruns, Wahrscheinlichkeitsrechnung und Kollektivmaßlehre, Leipzig 1906, S. 191ff., der den Satz darauf stützt, daß in der Entwicklung von $W_n(x)$ nach den Ableitungen von $\Phi(x)$ die Koeffizienten der späteren Glieder mit wachsendem n gegenüber dem ersten zurücktreten. Diese Betrachtung ließe sich nicht ohne Schwierigkeit zu einem *Beweis* des Poissonschen Satzes ergänzen, indem man noch zeigt, daß auch die *Summe* der Glieder mit $n = \infty$ gegen null geht. Aber die Allgemeinheit der in den §§ 3–5 gegebenen Beweise läßt sich auf diesem Wege nicht erreichen, weil die Entwickelbarkeit willkürlicher Funktionen nach der Exponentialreihe an engere Voraussetzungen geknüpft ist (beschränkte Schwankung und Verschwinden im Unendlichen stärker als $x^{-3} e^{-\frac{x^2}{2}}$). Vgl. R. v. Mises, Jahresber. d. deutsch. Math.-Ver. 1912, S. 9.

Im Falle des Zahlenlottos ist nach der üblichen Annahme jedes $\mathfrak{v}_\varkappa(1) = \mathfrak{v}_\varkappa(2) = \ldots = \mathfrak{v}_\varkappa(90) = \frac{1}{90}$.

Man findet zu (9):

$$(9') \qquad \begin{aligned} a_\varkappa &= \sum_{x=0}^{m} x\,\mathfrak{v}_\varkappa(x); \quad s_\varkappa^2 = \sum_{x=0}^{m}(x-a_\varkappa)^2\,\mathfrak{v}_\varkappa(x) = \sum_{x=0}^{m} x^2\,\mathfrak{v}_\varkappa(x) - a_\varkappa^2; \\ b_n &= \sum_{\varkappa=1}^{n}\sum_{x=0}^{m} x\,\mathfrak{v}_\varkappa(x); \quad r_n^2 = 2\sum_{\varkappa=1}^{n}\left[\sum_{x=0}^{m} x^2\,\mathfrak{v}_\varkappa(x) - a_\varkappa^2\right], \end{aligned}$$

wodurch der Zusammenhang $x = r_n u + b_n$ zwischen dem Argument u der Gaußschen Funktion und der Summe x der Ziehungsergebnisse festgelegt ist.

In unserem Beispiel des Zahlenlottos ist $a_\varkappa = 45{,}5$, $s_\varkappa^2 = \frac{1}{90}[1^2 + 2^2 + \ldots + 90^2] - 45{,}5^2 = 674{,}9167$; daher $b_n = 45{,}5\,n$, $r_n^2 = 1349{,}83\,n$. Mithin ist die Wahrscheinlichkeit $\mathfrak{w}_n(u)$ in n Einzel-Ziehungen (vor denen jedesmal das zuletzt gezogene Los zurückgelegt wird) die Summe x oder den Durchschnitt $\frac{x}{n}$ zu ziehen, asymptotisch für große n:

$$\mathfrak{w}_n(u) = \mathfrak{w}_n'(x) \sim \frac{1}{\sqrt{\pi}}\,\frac{1}{36{,}74\,\sqrt{n}}\,e^{-\left(\frac{x-45{,}5\,n}{36{,}74\,\sqrt{n}}\right)^2}$$

und exakt, wenn α für den Zahlenwert $1349{,}83\ldots$ geschrieben wird:

$$(9'') \qquad \lim_{n=\infty} \sqrt{2\,n\,\alpha}\,\mathfrak{w}_n(u) = \frac{1}{\sqrt{\pi}}\,e^{-u^2}.$$

Die einzige Bedingung, der die $\mathfrak{v}_\varkappa(x)$ allgemein unterworfen werden müssen, damit Satz V gilt, ist die, daß für jedes $\varkappa$ mindestens ein Paar aufeinanderfolgender Koeffizienten von null verschieden sein muß. Denn die Beschränktheit der ersten drei Momente, die nach den Korollaren V_1 und V_2 außerdem noch allein zu fordern wäre, ist sichergestellt, wenn die Zahl m als endliche Größe unabhängig von $\varkappa$ gegeben ist. Da man nun die gesuchte Wahrscheinlichkeit $\mathfrak{w}_n(u)$ für das Treffen der Summe $x = r_n u + b_n$ in bekannter Weise durch Multiplikation der Polynome $\sum_x \mathfrak{v}_\varkappa(x)\,t^x$ $(\varkappa = 1, 2, 3, \ldots, n)$ findet, können wir folgenden allgemeinen Satz über die Koeffizienten eines Produktes von Polynomen aussprechen:

Besitzt jedes Polynom $P_\varkappa = \sum_x \mathfrak{v}_\varkappa(x)\,t^x$ der unendlichen Folge P_1, P_2, P_3, ... von Polynomen höchstens m-ter Ordnung lauter nicht-negative Koeffizienten, die Koeffizientensumme 1 und mindestens ein Paar unmittelbar aufeinanderfolgender (auch im Limes) nicht verschwindender Koeffizienten, so liefern die Koeffizienten $k_n(x)$ der Entwicklung des Produktes

$P_1 P_2 P_3 \ldots P_n = \sum\limits_x k_n(x) t^x$, *als Funktion des Index x, bei gehöriger (von n abhängiger) Wahl des Anfangspunktes, des Abszissen- und Ordinatenmaßstabes aufgetragen, im Limes für $n = \infty$ die Gaußsche Linie* $\frac{1}{\sqrt{\pi}} e^{-u^2}$; *es ist*

$$\lim_{n=\infty} r_n k_n(r_n u + b_n) = \frac{1}{\sqrt{\pi}} e^{-u^2} \tag{10}$$

mit den durch (9') *bestimmten Werten von r_n und b_n.* Der Grenzübergang ist so zu vollziehen, daß u konstant gehalten wird oder zumindest für $n = \infty$ einem festen Wert zustrebt.

Dieser, wie es scheint, neue Satz zeigt, daß die oben angeführte bekannte Eigenschaft der Binomialzahlen nicht etwa für diese kennzeichnend ist, sondern der Koeffizientenreihe eines *jeden Polynoms* zukommt, das im wesentlichen als *Produkt sehr vieler Polynome mit positiven Koeffizienten* entsteht.

Daß das Vorhandensein eines Paares aufeinanderfolgender Koeffizienten nicht nur hinreichend, sondern (wenigstens für „fast alle" Polynome) auch notwendig ist, sieht man leicht ein. Denn wenn z. B. die Koeffizienten aller ungeraden Indizes ganz ausfielen, könnten gar keine ungeraden Summen entstehen; wenn nun die ungeraden Indizes zwar vorkämen, aber nur in wenigen Faktoren, so müßten in der Entwicklung des Produktes die ungeraden Glieder verhältnismäßig viel kleinere Koeffizienten haben als die benachbarten geraden, was offenbar dem Resultat (10) widerspricht.

In unserem Satz V von § 3 ist nichts enthalten, was eine Beschränkung auf Polynome von endlicher Ordnung notwendig machen würde, wenn nur alle Voraussetzungen des § 3 erfüllt sind. Machen wir von den Korollaren V_1 und V_2 in § 6 Gebrauch, so können wir sagen:

Die in Gleichung (10) *zum Ausdruck kommende Eigenschaft der Koeffizienten besteht auch für das Produkt von Potenzreihen, die nach positiven und eventuell auch negativen Potenzen fortschreiten, sobald noch folgende Bedingungen erfüllt sind*: *es muß*

$$\left|\sum_x x \mathfrak{v}_\varkappa(x)\right| < A, \quad \left|\sum_x x^2 \mathfrak{v}_\varkappa(x)\right| < B, \quad \left|\sum_x x^3 \mathfrak{v}_\varkappa(x)\right| < C, \tag{10'}$$

d. h. die Folge der ersten, zweiten und dritten Momente der gegebenen Koeffizienten beschränkt sein.

Die beiden vorstehenden Sätze sind nur eine andere Form der Gleichung (V) unter Berücksichtigung der Korollare V_1 und V_2. Es erübrigt sich, Gleichung (III), die naturgemäß in analoger Weise als Satz über die Koeffizienten einer Potenzreihe gedeutet werden kann, noch besonders auszusprechen.

5. Mehrdimensionaler Fall. Die bisher abgeleiteten Sätze, namentlich die beiden letzten allgemeinen Behauptungen, lassen sich vermöge unseres Korollars V_4 in § 6 auch auf Fälle mehrdimensionaler Verteilungen ausdehnen. Um diese Anwendung zu erklären und zugleich den in § 6, 4 angedeuteten Rechnungsgang zu erläutern, wollen wir ein konkretes Beispiel aus dem Rahmen der klassischen Aufgaben der Wahrscheinlichkeitsrechnung voranstellen.

Beim „gewöhnlichen Zahlenlotto" werden aus einer Urne, die zu Anfang jede der Zahlen 1 bis 90 einmal enthält, hintereinander 5 Zahlen gezogen. Dabei wird angenommen, daß die Wahrscheinlichkeit eine bestimmte Zahl zu ziehen, für alle in der Urne enthaltenen Zahlen gleich groß ist. Die Wahrscheinlichkeit, bei einer Ziehung eine bestimmte Folge von 5 Zahlen $x^{(1)}, x^{(2)}, \ldots, x^{(5)}$ zu erhalten, wird durch eine 5-dimensionale Verteilung dargestellt, wobei

$$(11) \qquad \mathfrak{v}_\varkappa(x^{(1)}, x^{(2)}, \ldots, x^{(5)}) = \tfrac{1}{90}\cdot\tfrac{1}{89}\cdot\tfrac{1}{88}\cdot\tfrac{1}{87}\cdot\tfrac{1}{86} = p$$

$$\text{für } x^{(1)} \neq x^{(2)} \neq \ldots \neq x^{(5)}; \quad x^{(1)}, x^{(2)}, \ldots, x^{(5)} = 1, 2, 3, \ldots, 90,$$

während alle übrigen $\mathfrak{v}_\varkappa$ null sind. Diese, von $\varkappa$ unabhängige, Verteilung genügt allen Voraussetzungen der Behauptung V. Man erhält für den 5-dimensionalen Vektor des Mittelwertes $\bar{a}_\varkappa$ und die 5-dimensionale Matrix der Streuung $\overline{\overline{s}}_\varkappa$:

$$(11') \qquad \begin{aligned} a_\varkappa^{(\iota)} &= \tfrac{1}{90}[1+2+\ldots+90] = 45{,}5, & \iota &= 1, 2, \ldots, 5; \\ s_\varkappa^{(\iota,\iota)} &= \tfrac{1}{90}[1^2+2^2+\ldots+90^2] - 45{,}5^2 = \frac{89\cdot 91}{12} = \alpha, & \iota &= 1, 2, \ldots, 5; \\ s_\varkappa^{(\iota,\lambda)} &= \frac{1}{90\cdot 89}[1\cdot 2+1\cdot 3+\ldots+1\cdot 90+2\cdot 1+2\cdot 3+\ldots+2\cdot 90+\ldots] - 45{,}5^2 \\ &= -\tfrac{91}{12} = \beta & \iota, \lambda &= 1, 2, \ldots, 5, \quad \iota \neq \lambda. \end{aligned}$$

Da die $a_\varkappa^{(\iota)}$ und $s_\varkappa^{(\iota,\lambda)}$ für alle $\varkappa$ gleich sind, stimmen die letzteren mit den in (7″) des § 6 definierten $h_n^{(\iota,\lambda)}$ überein, während die ersteren $b_n^{(\iota)} = 45{,}5\,n$ ergeben. Die aus den $h_n^{(\iota,\lambda)}$ gebildete quadratische Form

$$(11'') \qquad \alpha(z_1^2 + z_2^2 + \ldots + z_5^2) + 2\beta(z_1 z_2 + z_1 z_3 + \ldots + z_4 z_5)$$

hat zur einen Hauptachse mit dem Wurzelwert $\lambda_1 = \alpha + 4\beta$ die Gerade $z_1 = z_2 = \ldots = z_5$; die anderen 4 Hauptachsen entsprechen dem Wurzelwert $\lambda_2 = \alpha - \beta$, können also in der Ebene $z_1 + z_2 + \ldots + z_5 = 0$ beliebig (orthogonal zueinander) angenommen werden. Die quadratische Form mit derselben Orientierung und reziproken Wurzelwerten ist

$$(11''') \qquad \begin{aligned} &\alpha'(z_1^2 + z_2^2 + \ldots + z_5^2) + 2\beta'(z_1 z_2 + z_1 z_3 + \ldots + z_4 z_5), \\ &\text{mit } \alpha' = \frac{\alpha + 3\beta}{(\alpha-\beta)(\alpha+4\beta)} = \frac{2\cdot 86}{15\cdot 85\cdot 91}, \quad \beta' = -\frac{\beta}{(\alpha-\beta)(\alpha+4\beta)} = \frac{2}{15\cdot 85\cdot 91}. \end{aligned}$$

Setzt man, wie in § 6, 4 $\bar{x} = \sqrt{n}\,\bar{z} + \bar{b}_n$, so ergibt die Anwendung von Satz V_4 für die Wahrscheinlichkeit $\mathfrak{w}_n(z^{(1)}, z^{(2)}, \ldots, z^{(5)})$, in n Ziehungen durch die an erster Stelle gezogenen Lose die Summe $x^{(1)} = \sqrt{n} z^{(1)} + 45{,}5\, n$ $\left(\text{bzw. den Durchschnitt } \frac{x^{(1)}}{n}\right)$, durch die an zweiter Stelle gezogenen die Summe $x^{(2)} = \sqrt{n}\, z^{(2)} + 45{,}5\, n$ usw. zu erhalten:

$$(12) \qquad \lim_{n=\infty} \sqrt{n}^5\, \mathfrak{w}_n(z^{(1)}, z^{(2)}, \ldots, z^{(5)})$$
$$= \frac{1}{\sqrt{(2\pi)^5 \lambda_1 \lambda_2^4}}\, e^{-\frac{\alpha'}{2}\left(z^{(1)^2} + z^{(2)^2} + \ldots + z^{(5)^2}\right) - \beta'\left(z^{(1)} z^{(2)} + \ldots + z^{(4)} z^{(5)}\right)}.$$

Es ist übersichtlicher, hier an Stelle von $\bar{z}$ als neue Variable $\bar{u} = \bar{z} : \sqrt{2\alpha}$ einzuführen, so daß dann $\bar{x} = \sqrt{2 n \alpha}\, \bar{u} + \bar{b}_n$ wird, weil man dadurch einen bequemen Vergleich mit dem oben § 7, 4 betrachteten eindimensionalen Fall gewinnt. Setzen wir teilweise gleich die Zahlenwerte ein, so wird:

$$(12') \qquad \lim_{n=\infty} \sqrt{2 n \alpha}^5\, \mathfrak{w}_n(u^{(1)}, u^{(2)}, \ldots, u^{(5)})$$
$$= \frac{1{,}0006}{\sqrt{\pi}^5}\, e^{-1{,}00052\left(u^{(1)^2} + \ldots + u^{(5)^2}\right) - 0{,}0232\left(u^{(1)} u^{(2)} + \ldots + u^{(4)} u^{(5)}\right)}.$$

Man sieht, daß der Koeffizient der Quadratsumme im Exponenten sich wenig von 1, der der übrigen Glieder wenig von null, der ganze Ausdruck nur unbedeutend von der fünften Potenz des oben gefundenen (9″) unterscheidet. Der Unterschied rührt ja nur von dem Nicht-Zurücklegen der Lose innerhalb einer Ziehung von 5 Nummern her.

Der allgemeine Satz, der hier dem im vorigen Abschnitt für Polynome einer Variablen angegebenen entspricht, lautet:

Besitzt jedes der Polynome in k Variablen

$$P_\varkappa = \sum \mathfrak{v}_\varkappa(x_1, x_2, \ldots, x_k)\, t_1^{x_1} t_2^{x_2} \ldots t_k^{x_k}$$

der unendlichen Folge von Polynomen $P_1, P_2, P_3, \ldots$ *von höchstens m-tem Grad in jeder Variablen lauter nicht-negative Koeffizienten, die Koeffizientensumme* 1 *und hinsichtlich jeder Variablen mindestens ein Paar unmittelbar aufeinanderfolgender* (*auch im Limes*) *nicht verschwindender Koeffizienten, so haben die Koeffizienten* $k_n(x_1, x_2, \ldots, x_k)$ *der Entwicklung des Produktes* $P_1 P_2 \ldots P_n = \sum k_n(x_1 x_2 \ldots x_k)\, t_1^{x_1} t_2^{x_2} \ldots t_k^{x_k}$ *als Funktionen der Indizes* $x_1 \ldots x_k$, *bei gehöriger, von n abhängiger Wahl des Koordinatensystems* (Anfangspunkt und Achsenrichtungen) *sowie der Maßstäbe für jede Achse und für die* k_n *selbst, die Form eines Produktes von k Gaußschen Funktionen* $\frac{1}{\sqrt{\pi}^k}\, e^{-u_1^2 - u_2^2 - \ldots - u_k^2}$. Die Aufsuchung der genannten Elemente, also der linearen Beziehungen zwischen den x_ι und u_ι, fällt, wie in § 6, 4 beschrieben, wesentlich mit der Aufgabe der *Reduktion einer Fläche zweiten Grades auf Mittelpunkt und Hauptachsen* zusammen.

Die Ausdehnung auf Potenzreihen, die Übertragung auf die Integrale W_n usw. bedarf keiner besonderen Erklärung. Der exakte Ausdruck für die Forderung hinsichtlich aufeinanderfolgender nicht verschwindender Koeffizienten lautet: Es gibt für jedes $\varkappa$ eine Gruppe von k Zahlen $m_\varkappa^{(\iota)}$ aus der Reihe $0, 1, 2, \ldots, m-1$ $(\iota = 1, 2, \ldots, k)$, so daß gleichzeitig

$$\mathfrak{v}_\varkappa(x_1', x_2', \ldots, x_{\iota-1}', m_\varkappa^{(\iota)}, x_{\iota+1}', \ldots, x_k') > \mathfrak{v},$$

$$\mathfrak{v}_\varkappa(x_1'', x_2'', \ldots, x_{\iota-1}'', m_\varkappa^{(\iota)}+1, x_{\iota+1}'', \ldots, x_k'') > \mathfrak{v}$$

für zwei Wertverbindungen $x_1' \ldots x_k'$ und $x_1'' \ldots x_k''$ bei einem von $\varkappa$ unabhängigen positiven Wert von $\mathfrak{v}$.

§ 8.

Der erste Fundamentalsatz.

1. Die Verteilungen. Den Gegenstand der Wahrscheinlichkeitsrechnung bildet die Untersuchung unendlicher Folgen von Elementen, die durch unterscheidende Merkmale gekennzeichnet sind (Würfe mit einer Münze, Ziehungen aus einer Urne usw.); als Merkmal betrachten wir immer einen dem Element zugeordneten Punkt im k-dimensionalen „Merkmalraum“ $(k \geqq 1)$, oder seine Koordinaten $x^{(1)}, x^{(2)}, \ldots, x^{(k)}$. Beispiele solcher unendlich gedachter Folgen sind die Ziehungen aus einer Urne, die etwa beim gewöhnlichen Lotto zu einem 5-dimensionalen Merkmalraum führen (Koordinaten sind die 5 Nummern einer Ziehung) oder die Moleküle eines Gases mit dem 3-dimensionalen Merkmalraum ihrer Geschwindigkeiten (Koordinaten sind die 3 Geschwindigkeitskomponenten). Irgendeinem Raumteil R des k-dimensionalen Merkmalraumes mögen die Merkmale von N_R unter den ersten N Elementen der Folge angehören; dann fordern wir die Existenz des Grenzwertes

$$\lim_{N=\infty} \frac{N_R}{N} = V_R, \tag{1}$$

der natürlich eine nicht-negative Zahl, höchstens gleich 1, sein muß. Wir nennen V_R die *Wahrscheinlichkeit für das Auftreten des „Merkmales R“ innerhalb der betrachteten Folge* oder kurz die Wahrscheinlichkeit von R. Eine Elementenfolge, für welche der Grenzwert (1) hinsichtlich eines jeden Raumteiles existiert (und gewisse weitere Axiome, von denen hier nicht die Rede sein soll, erfüllt sind) — nur solche Folgen werden in der Wahrscheinlichkeitsrechnung in Betracht gezogen — nennen wir im Anschluß an einen bekannten Ausdruck ein *Kollektiv.* Der Ausdruck „Wahrscheinlichkeit eines Merkmales R" hat demnach *nur innerhalb eines wohldefinierten Kollektivs* einen Sinn.

Der in der Regel an die Spitze aller Betrachtungen gestellte Alternativfall, ob ein „Ereignis eintrifft" oder „nicht eintrifft", ordnet sich unserer Auffassung in der Weise unter, daß hier ein 1-dimensionaler Merkmalraum vorliegt, in dem V_R null ist für jedes Gebiet R, das nicht einen von zwei festen Punkten (z. B. $x = 0$ und $x = 1$) enthält, und von null verschieden für jedes R, dem einer der beiden Punkte angehört. Ein einfaches Beispiel bildet der Ausgangspunkt des Bernoullischen Problems (§ 7, 2), wo man das Ziehen der weißen Kugel als „Eintreffen", das der schwarzen als „Nichteintreffen des Ereignisses" deuten kann, während wir es vorgezogen haben, dem einen Fall die Zahl null, dem anderen die Zahl eins zuzuordnen. Diese Bemerkung zeigt, daß auch von der „Wahrscheinlichkeit für das Eintreffen eines Ereignisses" nur dann gesprochen werden kann, wenn ein wohldefiniertes Kollektiv vorliegt.

Die Gesamtheit der V_R für alle Teile des Merkmalraumes bildet die *Verteilung* des Kollektivs. Bezeichnet R und R' zwei Raumteile ohne gemeinsame Punkte, so folgt aus der Definition (1) unmittelbar der Satz:

$$V_{R+R'} = V_R + V_{R'} \tag{2}$$

in Worten: Die Wahrscheinlichkeit für das Auftreten des Merkmales R *oder* R' innerhalb des betrachteten Kollektivs ist die Summe aus der Wahrscheinlichkeit für das Auftreten von R und jener für das Auftreten von R'. Man pflegt diesen Satz den von der *„Addition der Wahrscheinlichkeiten"* zu nennen und ihm einen zweiten über die Multiplikation zur Seite zu stellen. Bei der Multiplikation handelt es sich aber um etwas ganz anderes, nämlich um eine an zwei Kollektivs vorgenommene Operation durch die ein drittes Kollektiv gebildet wird. (Vgl. unter Abschn. 2.)

Die Beziehung (2) gestattet, die Verteilung eines beliebigen Kollektivs durch eine *einzige Funktion* von k Variablen $x^{(1)} \ldots x^{(k)}$ darzustellen, die hinsichtlich jeder Variablen monoton, nicht abnehmend ist, für $x^{(1)} = x^{(2)} = \ldots = x^{(k)} = -\infty$ den Wert null und für $x^{(1)} = x^{(2)} = \ldots = x^{(k)} = \infty$ den Wert eins annimmt. Eine solche Funktion ist nichts anderes als das, was wir allgemein in § 6, 4 und für den 1-dimensionalen Fall schon in § 3, 1 unter dem Namen „Verteilung" eingeführt haben. Der Zusammenhang zwischen der früheren Bezeichnung $V(\bar{x})$ und der jetzigen V_R wird dadurch hergestellt, daß $V(\bar{x}) = V_R$ für jenen Raumteil R, dessen Koordinaten $\xi^{(\iota)}$ den Bedingungen $\xi^{(\iota)} \leqq x^{(\iota)}$ (wo $x^{(\iota)}$ die Komponenten von $\bar{x}$) genügen. Durch die Aufnahme des Gleichheitszeichens in die Beziehung zwischen $\xi^{(\iota)}$ und $x^{(\iota)}$ wird festgelegt, daß die Funktion $V(\bar{x})$ mit ihrer *oberen* Limes-Funktion zusammenfallen muß. Es bedeutet daher $V(\bar{x})$ die

Wahrscheinlichkeit dafür, daß die erste Merkmalkomponente *höchstens gleich* $x^{(1)}$, die zweite *höchstens gleich* $x^{(2)}$ wird, usf. Die untere Limes-Funktion von $V(\bar{x})$ würde die Wahrscheinlichkeit dafür liefern, daß die erste Komponente *kleiner als* $x^{(1)}$ ist usw.

Unter allen möglichen Verteilungen heben wir zwei praktisch wichtigere Typen hervor, die wir im Anschluß an übliche Ausdrücke als „*arithmetische*" und „*geometrische*" Verteilungen bezeichnen wollen.

Eine Verteilung soll *arithmetisch* heißen, wenn V_R null ist für alle R, die nicht mindestens einen Gitterpunkt eines durch k Einheiten $\mathfrak{e}_1, \mathfrak{e}_2, \ldots, \mathfrak{e}_k$ bestimmten Gitters im Merkmalraum enthalten. Die Verteilung ist in diesem Fall vollständig bestimmt durch die Angabe der endlich oder abzählbar unendlich vielen von null verschiedenen Werte von V_R für alle solche R, die nur je einen Gitterpunkt umfassen. Wir nennen diese V_R die *Einzel- oder Punktwahrscheinlichkeiten* der betreffenden Gitterpunkte und schreiben dafür $\mathfrak{v}(\bar{x})$ oder $\mathfrak{v}(x^{(1)}, x^{(2)}, \ldots, x^{(k)})$, wo die $x^{(\iota)}$ die Komponenten von $\bar{x}$ und Koordinaten des Gitterpunktes bezeichnen. Die $\mathfrak{v}(\bar{x})$ sind nicht-negative Zahlen, deren Summe 1 ist. Der obenerwähnte Fall der Alternative, „ob ein Ereignis eintritt oder nicht", ist ein Beispiel für eine 1-dimensionale, arithmetische, zweiwertige Verteilung.

Eine Verteilung heiße *geometrisch*, wenn an jeder Stelle $\bar{x}$ des Merkmalraumes der Grenzwert

$$\lim_{|R|=0} \frac{V_R}{|R|} = v(\bar{x}) \tag{3}$$

existiert, endlich und beschränkt ist. Dabei bedeutet $|R|$ das Volumen des Raumteiles R und der Grenzübergang werde so vollzogen, daß unter Festhaltung von $\bar{x}$ die k Beträge $|x^{(\iota)} - \xi^{(\iota)}|$ gegen null gehen, wenn $\xi^{(\iota)}$ die Koordinaten eines Punktes von R sind. Die Verteilung ist in diesem Falle durch Angabe der Funktion $v(\bar{x})$ vollständig bestimmt, die wir als die *Wahrscheinlichkeitsdichte im Punkte* $\bar{x}$ bezeichnen. Die $v(\bar{x})$ sind nicht negativ und verschwinden derart im Unendlichen, daß das Raumintegral, erstreckt über den ganzen unendlichen Merkmalraum, gleich 1 wird.

Im allgemeinen Fall von Kollektivs mit beliebigen Verteilungen gibt es weder Einzelwahrscheinlichkeiten noch Wahrscheinlichkeitsdichte.

2. Die Grundoperationen. Die Aufgaben der Wahrscheinlichkeitsrechnung bestehen ausschließlich darin, aus den gegebenen Verteilungen gegebener Kollektivs die Verteilungen anderer Kollektivs zu berechnen, die aus den ersteren durch bestimmte Operationen abgeleitet werden. So sind z. B. beim Poissonschen Problem die gegebenen Kollektivs die Folgen der einfachen Züge aus je einer Urne mit den durch $p_\varkappa$, $q_\varkappa$ bestimmten, 1-dimensionalen, arithmetischen Verteilungen $V_\varkappa$ (§ 6, 3). Das abgeleitete

Kollektiv hat zum Element eine Ziehung aus sämtlichen n Urnen und als Merkmal die Summe der gezogenen weißen Kugeln; die gesuchte Verteilung W_n ist wieder 1-dimensional und arithmetisch.

Der systematische Aufbau der Wahrscheinlichkeitsrechnung erfordert das Aufsuchen aller jener Grundoperationen, auf die die Ableitung neuer Kollektivs aus gegebenen zurückgeführt werden kann. Ohne hier die Frage der Vollständigkeit zu berühren, betrachten wir die folgenden drei Grundoperationen, welche zur Formulierung der Fundamentalsätze notwendig sind: die *Mischung*, die *Teilung* und die *Verbindung*.

Unter *Mischung* verstehen wir die Veränderung eines Kollektivs, die dadurch bewirkt wird, daß jedem Punkt $\bar{x}$ des ursprünglichen Merkmalraumes ein Punkt $\bar{x}'$ eines neuen Merkmalraumes zugeordnet wird. Die Veränderung ist nur dann wesentlich, wenn die Zuordnung *nicht umkehrbar eindeutig* ist, z. B. wenn der neue Merkmalraum geringere Dimension hat und die Zuordnung etwa in der Streichung der überschüssigen Koordinaten besteht, oder wenn die ursprüngliche Verteilung geometrisch, die neue arithmetisch ist. Beispiele: Beim sog. „Nadelproblem" besteht das gegebene Kollektiv aus den Würfen einer Nadel auf ein von äquidistanten Parallelen gebildetes Strichgitter; Merkmal eines Elementes ist die durch drei Parameter bestimmte Lage der geworfenen Nadel (3-dimensionale, geometrische Verteilung). Ein neues Kollektiv entsteht, wenn wir die Elemente nur darnach unterscheiden, ob die Nadel einen Gitterstrich kreuzt oder nicht (1-dimensionale, arithmetische, zweiwertige Verteilung). Nur wenn die erste Verteilung gegeben ist, kann es Aufgabe der Wahrscheinlichkeitsrechnung sein, die zweite zu berechnen — womit sich das sog. Bertrandsche Paradoxon erledigt. Ein anderer Fall: Die Ziehungen des Zahlenlottos (vgl. § 7, 5) bilden die Elemente eines Kollektivs mit 5-dimensionaler, arithmetischer $90\cdot 89\cdot 88\cdot 87\cdot 86$ wertiger Verteilung. Sieht man aber als Merkmal einer Ziehung nur die 5 gezogenen Zahlen ohne Rücksicht auf ihre Reihenfolge oder gar nur ihre Summe bzw. (was auf dasselbe hinauskommt) das arithmetische Mittel an, so erhält man neue Kollektivs; das erste Mal ist die Verteilung wieder 5-dimensional, arithmetisch, aber nur $\frac{90\cdot 89\cdot 88\cdot 87\cdot 86}{5\cdot 4\cdot 3\cdot 2\cdot 1}$ wertig[11]), das zweite Mal wird sie 1-dimensional, arithmetisch und $(5\cdot 90-4) = 446$ wertig. In beiden Fällen ist also eine eindeutige Umkehrung der Zuordnung nicht möglich.

Wie im Falle einer Mischung die neue Verteilung V' gefunden wird,

---

[11]) Und zwar ist jedem Ziehungsergebnis $x_1, x_2, \ldots, x_5$ jener Punkt des Bereiches $\xi_1 > \xi_2 \ldots \xi_5 > 0$ im Merkmalraum zugeordnet, der entweder mit $\bar{x}$ identisch ist oder durch Spiegelungen an einer oder mehreren der Ebenen $\xi_1 = \xi_2, \ldots, \xi_4 = \xi_5$ in $\bar{x}$ übergeht.

ist aus der Definition der Mischung auf Grund der Axiome, die das Kollektiv im allgemeinen bestimmen, abzuleiten. Wir setzen, wie es in der Wahrscheinlichkeitsrechnung bisher immer üblich war, das Resultat einfach hin: Bezeichnet R' die Punktmenge, in die R durch die Transformation übergeht — und wir wollen, da dies für die vorliegenden Zwecke ausreicht, R' wieder als einen Raumteil ansehen[12]) — so ist

$$V'_{R'} = V_R = \int_{(R)} d\,V(\bar{\xi}). \tag{4}$$

Insbesondere erhält man $V'(\bar{x})$, indem man das Stieltjessche Integral von $d\,V$ über alle jene Raumteile erstreckt, deren Punkte nach der Transformation in Punkte $\xi^{(\iota)} \leqq x^{(\iota)}$ übergehen.

Die zweite Grundoperation, *Teilung* oder *Aussonderung*, besteht darin, daß der Merkmalraum eines Kollektivs in zwei Räume R_1 und R_2 zerlegt wird und dann alle Elemente, deren Merkmale R_2 angehören, ausgesondert werden. Es bleibt eine Folge von Elementen übrig, deren Merkmale alle in R_1 liegen und die neue Verteilung V' aufweisen, für die

$$V'_A = \frac{V_A}{V_{R_1}}, \qquad V'_B = 0, \tag{5}$$

wenn A ganz in R_1 enthalten ist und B keinen Punkt mit R_1 gemein hat. Daß (5) tatsächlich besteht, ist wieder aus den Axiomen usw. zu schließen. Beispiele: Beim obenerwähnten Nadelproblem kann man nach der Wahrscheinlichkeit dafür fragen, daß eine Nadel einen Gitterstrich senkrecht kreuzt; dies wäre nur eine andere als die oben angegebene Mischung der ursprünglichen Verteilung. Eine ganz andere Frage aber ist es, mit welcher Wahrscheinlichkeit eine Nadel, von der man schon weiß, daß sie senkrecht zum Gitterstrich fällt, das Gitter trifft. Hier liegt eine Teilung vor, indem alle jene Nadelwürfe aus der Betrachtung ausgeschlossen werden, bei denen die Nadel schief auffällt. — Ein anderer Fall: Man stelle sich eine unendliche Folge von Urnen mit schwarzen und weißen Kugeln vor, jede gekennzeichnet durch eine zwischen null und eins liegende Zahl $x^{(1)}$, gleich der Wahrscheinlichkeit eines weißen Zuges aus der betreffenden Urne. Das Element des Kollektivs sei nun ein dreimaliger Zug aus irgendeiner der Urnen; sein Merkmal zweidimensional, und zwar erstens die Zahl $x^{(1)}$, die der betreffenden Urne zukommt, zweitens die zwischen 0 und 3 inkl. liegende Anzahl $x^{(2)}$ der gezogenen weißen Kugeln. Als gegeben vorauszusetzen ist die Verteilung $V(x^{(1)}, x^{(2)})$; gefragt wird nach der Wahrscheinlichkeit $V'(x^{(1)})$ dafür, daß bei gegebenem $x^{(2)}$, also bei bekanntem Ziehungsergebnis, der Zug aus einer Urne mit irgendeinem

---

[12]) Bei exakter Durchführung der Grundlagen wird man statt von „Raumteilen" allgemein von *Punktmengen* sprechen.

$x^{(1)}$-Wert erfolgt. Auf diese Fragestellung, die zum zweiten Fundamentalsatz führt, kommen wir in den nächsten Paragraphen zurück.

Endlich erklären wir als die *Verbindung* zweier (unabhängiger) Kollektivs die folgende Art, aus zwei Kollektivs mit den Elementen e' bzw. e'' und den Verteilungen $V'(\bar{x}')$ und $V''(\bar{x}'')$ ein neues Kollektiv mit den Elementen e und der Verteilung $V(\bar{x})$ abzuleiten. Es sollen die Elemente $e_1, e_2, e_3, \ldots$ Kombinationen aus je einem Element e' und einem Element e'' sein, und zwar so, daß e_1 aus e_1' und e_1'', dann e_2 aus e_2' und e_2'' besteht usf. Das Merkmal eines solchen Elementes e sei die Zusammenfassung der Merkmale $\bar{x}'$ und $\bar{x}''$, d. h. wenn diese k' bzw. k''-dimensional sind, so sei $\bar{x}$ der $k = (k' + k'')$-dimensionale Vektor, dessen erste k' Komponenten mit denen von $\bar{x}'$, die restlichen mit denen von $\bar{x}''$ übereinstimmen. Ist ein Raumteil R' des ersten und ein Raumteil R'' des zweiten Merkmalraumes gegeben, so ist dadurch ein Raumteil R des neuen, k-dimensionalen Raumes bestimmt. Zwischen den drei Verteilungen V', V'' und V besteht dann die Beziehung:

$$V_R = V'_{R'}\, V''_{R''}, \tag{6}$$

deren Bestehen ebenfalls aus den Forderungen, denen ein Kollektiv genügen muß, zu begründen ist. Beispiele für diesen Fall sind leicht anzugeben: Besteht jedes der gegebenen Kollektivs aus den Würfen mit einem Würfel, deren Merkmal die Zahl der obenliegenden Augen ist, so hat das abgeleitete Kollektiv zum Element einen Wurf mit beiden Würfeln, dessen Merkmal von den beiden Augenzahlen gebildet wird. Man fragt nach der Wahrscheinlichkeit, mit dem ersten Würfel eine gewisse Augenzahl x', mit dem zweiten die Augenzahl x'' zu werfen. Alle drei Verteilungen sind arithmetisch, die beiden gegebenen 1-dimensional und 6-wertig, die der „Verbindung" 2-dimensional und 36-wertig.

Diese Andeutungen über die Grundoperationen müssen hier genügen. Auf die verschiedenen sich aufdrängenden Fragen, namentlich auf die der Berechtigung der Anwendung dieser Begriffskonstruktionen auf Vorgänge der Wirklichkeit, kann in diesem Zusammenhange nicht eingegangen werden.

3. Zusammengesetzte Operationen, Faltung. Aus der in mannigfacher Weise möglichen Zusammensetzung von Grundoperationen entstehen alle jene Ableitungen neuer Kollektivs aus gegebenen, deren Untersuchung den gesamten Inhalt der Wahrscheinlichkeitsrechnung bildet. So wie die rationelle Mechanik nicht etwa die Antwort auf die Frage gibt, welche Bewegung ein bestimmter Körper zu einer bestimmten Zeit besitzt, sondern auf die, wie man aus gegebener Anfangsgeschwindigkeit und gegebenem Kräfteverlauf die Bewegung zu irgendeinem Zeitpunkt berechnet, so liegt es auch hier: Aus den bekannten Verteilungen der gegebenen Kollektivs werden

die Verteilungen der abgeleiteten berechnet. Alle scheinbaren Paradoxien und Unbestimmtheiten verschwinden mit Annahme dieser Formulierung aus der Wahrscheinlichkeitsrechnung. Natürlich bleiben — ebenso wie in der Mechanik und allen anderen exakten Naturwissenschaften — gewisse Schwierigkeiten übrig, die in dem Übergang zwischen der Theorie und der Wirklichkeit liegen. Sie sind hier im wesentlichen doppelter Natur: Man bedarf einmal gewisser (experimenteller) Methoden, um die jedem Problem zugrundeliegenden „*Ausgangs-Wahrscheinlichkeiten*" zu ermitteln (in der Mechanik die Anfangsgeschwindigkeiten usw.), und man muß zweitens in jedem einzelnen Fall den „*Ansatz*" finden, d. h. das Problem auf die Grundoperationen zurückführen (in der Mechanik die kinematischen und dynamischen Eigenschaften des Systems feststellen).

Die beiden Fundamentalsätze, die wir in der vorliegenden Arbeit entwickeln, beziehen sich auf zwei ganz bestimmte zusammengesetzte Operationen, die wir als die *Faltung* und die *Kreuzung* von n Kollektivs bezeichnen. Die zweite der eben genannten Schwierigkeiten wird also dadurch ausgeschaltet, daß wir von vornherein *bestimmte Konstruktionen* betrachten, durch welche das neue Kollektiv aus den gegebenen hervorgeht. Die erste entfällt von selbst, weil es gerade den wesentlichsten Inhalt der Fundamentalsätze ausmacht, daß sie etwas über die Verteilung des abgeleiteten Kollektivs bei in weiten Grenzen *willkürlicher Wahl der ursprünglichen Verteilungen* aussagen.

Wir verstehen unter der *Faltung von n* (unabhängigen) *Kollektivs* mit k-dimensionalen Merkmalen die *Zusammensetzung aus einer n-fachen Verbindung* (durch die ein Kollektiv mit kn-dimensionalem Merkmal entsteht) und einer *darauffolgenden Mischung*, die das Merkmal wieder zu einem k-dimensionalen macht. Sei e_1 das Element des ersten gegebenen Kollektivs, $\bar{x}_1$ sein Merkmal, ebenso e_2, $\bar{x}_2$ Element und Merkmal des zweiten usf., so besteht ein Element E_n der n-ten Verbindung aus einer Kombination, für die wir symbolisch $e_1 e_2 e_3 \ldots e_n$ schreiben wollen. Das Merkmal des Elementes E_n setzt sich aus den kn Komponenten der Vektoren $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n$ zusammen. Die Mischung sei definiert durch eine k-wertige Gleichung:

$$\bar{x} = \bar{X}_n(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k), \tag{7}$$

wobei der k-dimensionale Vektor $\bar{x}$ das Merkmal des Elementes E_n der Faltung bezeichnet.

Sind $V_1(\bar{x}_1)$, $V_2(\bar{x}_2), \ldots$ die n Verteilungen der gegebenen Kollektivs, so erhält man die Verteilung $W_n(\bar{x})$ der n-ten Faltung in der Form:

$$d W_n(\bar{x}) = \int\int \ldots \int d V_1(\bar{x}_1)\, d V_2(\bar{x}_2) \ldots d V_n(\bar{x}_n), \tag{3}$$

wobei die $k(n-1)$-fache Integration rechts über alle Wertverbindungen der $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n$ zu erstrecken ist, welche bei festem $\bar{x}$ der Gleichung (7) genügen.

Hat die Gleichung (7) die Form einer *linearen* Beziehung

$$(7') \qquad \bar{x} = h_1(\bar{x}_1 - \bar{c}_1) + h_2(\bar{x}_2 - \bar{c}_2) + \ldots + h_n(\bar{x}_n - \bar{c}_n),$$

worin die $h_\varkappa$ und $\bar{c}_\varkappa$ gegebene Konstante sind, so sprechen wir von einer *linearen Faltung.* Sind alle $h_\varkappa = 1$, so nennen wir die Faltung eine *Summenbildung aus n Kollektivs*, sind alle $h_\varkappa = \frac{1}{n}$, so nennen wir sie *Durchschnittsbildung* [13]).

In jedem Fall der linearen Faltung kann man (8) formal nach $\bar{x}$ integrieren und erhält, wenn $h_n \neq 0$ vorausgesetzt wird:

$$(8') \quad W_n(\bar{x}) = \int\int\ldots\int V_n\left(\bar{c}_n + \frac{\bar{x} - h_{n-1}(\bar{x}_{n-1} - \bar{c}_{n-1}) - \ldots - h_1(\bar{x}_1 - \bar{c}_1)}{h_n}\right) dV_{n-1}(\bar{x}_{n-1}) \ldots dV_1(\bar{x}_1).$$

Man sieht hieraus, daß das in § 3, 1 eingeführte Integralprodukt von n Verteilungen, im 1-dimensionalen Fall und für $h_1 = h_2 = \ldots = h_n = \frac{1}{r_n}$, $c_1 + c_2 + \ldots + c_n = b_n$, gerade die Verteilung der n-ten linearen Faltung liefert. Die Ausdehnung auf andere $h_\varkappa$ und auf mehrere Dimensionen ist in § 6, 3 bzw. § 6, 4 erfolgt.

4. Der erste Fundamentalsatz. Der erste Fundamentalsatz der Wahrscheinlichkeitsrechnung enthält eine allgemeine Aussage über die Verteilung W_n, zu der im Limes für $n = \infty$ die fortgesetzte lineare Faltung von n gegebenen Kollektivs führt. Um den Satz in Kürze aussprechen zu können, führen wir den Begriff der „für die Faltung normalen Verteilungen" und einer „gleichmäßigen Folge" solcher Verteilungen ein.

Eine Verteilung soll als *für die Faltung normal* bezeichnet werden in folgenden Fällen:

a) Eine *arithmetische* Verteilung, wenn 1. nach jeder der k Koordinatenrichtungen wenigstens ein Paar benachbarter Gitterpunkte existiert, in denen die Einzelwahrscheinlichkeit nicht verschwindet; 2. das erste, zweite und dritte Moment der Verteilung (in jeder Komponente) endlich ist.

b) Eine *geometrische* Verteilung, wenn 1. die Wahrscheinlichkeitsdichte nach jeder Dimension von beschränkter Schwankung ist; 2. das erste, zweite und dritte Moment der Verteilung (in jeder Komponente) endlich ist.

c) Eine *allgemeine* Verteilung, wenn 1. ihre Streuung (die Determinante der Streuungsmatrix) nicht verschwindet; 2. jedes Moment m-ter Ordnung ($m = 1, 2, \ldots$) endlich ist.

---

[13]) An Stelle der skalaren Größen h könnten auch dyadische (Matrizen) treten, ohne daß sich im folgenden etwas wesentlich ändern würde.

Eine unendliche Folge derartiger Verteilungen soll *gleichmäßig* heißen, wenn folgende Bedingungen erfüllt sind: In den Fällen a), b) und c) sollen die unter 2. als endlich vorausgesetzten Momente beschränkte Folgen bilden; ferner soll im Falle

a) die Einzelwahrscheinlichkeit für je *ein* Paar benachbarter Gitterpunkte nach jeder Koordinatenrichtung eine von null verschiedene untere Schranke haben (das Gitter ist für alle Verteilungen als gleich vorausgesetzt);

b) die Gesamtschwankung der Wahrscheinlichkeitsdichte nach jeder Dimension eine obere Schranke besitzen;

c) die Summe der Streuungen (die Determinante der Summe der Streuungsmatrizes) stärker als mit der $^2/_3$ Potenz von n ins Unendliche wachsen.

Unter Verwendung dieser Bezeichnungen sprechen wir den *ersten Fundamentalsatz der Wahrscheinlichkeitsrechnung* wie folgt aus:

Die fortgesetzte lineare Faltung unendlich vieler k-dimensionaler Kollektivs, deren Verteilungen eine gleichmäßige Folge für die Faltung normaler Verteilungen bilden, bei konstanten oder wenigstens in endlichen Grenzen liegenden Koeffizienten $\bar{c}_\varkappa$ und Verhältnissen $h_1 : h_2 : h_3 \ldots$ in Gleichung (7′) *führt auf das Gaußsche Gesetz: Es läßt sich stets eine von n abhängige Koordinatentransformation im k-dimensionalen Raum von $\bar{x}$ auf $\bar{u}$* (Verschiebung des Anfangspunktes, Drehung der Achsen und Veränderung der Maßstäbe) *angeben, so daß*

a) *bei arithmetischen Verteilungen die Einzelwahrscheinlichkeit $\mathfrak{w}_n(\bar{x})$ der n-ten Faltung im Limes für $n = \infty$ proportional $e^{-u_1^2 - u_2^2 - \cdots - u_k^2}$ wird;*

b) *bei geometrischen Verteilungen die Wahrscheinlichkeitsdichte $w_n(\bar{x})$ der n-ten Faltung im Limes für $n = \infty$ proportional $e^{-u_1^2 - u_2^2 - \cdots - u_k^2}$;*

c) *im allgemeinen Falle, also auch für* a) *und* b), *die Verteilungsfunktion $W_n(\bar{x})$ im Limes für $n = \infty$ gleich dem Produkt $\Phi(u_1)$, $\Phi(u_2) \ldots \Phi(u_k)$, wo $\Phi(u)$ die Gaußsche Funktion $\frac{1}{\sqrt{\pi}} \int^u e^{-x^2} dx$ bezeichnet.*

Der Beweis dieses Satzes ist für den Fall $k = 1$ und $h_1 = h_2 = \ldots = h_n$ durch die Beweise der Behauptungen III bis V des § 3 gedeckt. Dabei ist hinsichtlich der in der Festlegung der „normalen" Verteilungen und ihrer „gleichmäßigen" Folgen eingeschlossenen Voraussetzungen von den Korollaren § 6, 1 und 2 Gebrauch gemacht. Die Ausdehnung auf k Dimensionen und von 1 verschiedene Verhältnisse $h_1 : h_2 : h_3 \ldots$ wird durch die Korollare § 6, 4 und § 6, 3 gerechtfertigt.

5. Summenbildung. Erster Satz der Fehlertheorie. Es bedeutet keine Einschränkung des Inhaltes des Fundamentalsatzes, wenn wir von vornherein die Koeffizienten $h_\varkappa$ der Gl. (7′) als untereinander gleich voraussetzen; denn in der Koordinatentransformation, die der Satz verlangt, ist eine Veränderung der Maßstäbe ausdrücklich eingeschlossen. Sprechen wir also von *Summenbildung* statt allgemein von „linearer Faltung“, so umfaßt, auf *arithmetische* Verteilungen angewendet, der erste Fundamentalsatz gerade die Aussagen des Bernoullischen und Poissonschen Gesetzes, einschließlich der weitgehenden Verallgemeinerungen, die im vorangehenden Paragraph behandelt wurden. In der üblichen Ausdrucksweise würde die allgemeine Aussage etwa lauten:

Die Wahrscheinlichkeit, in n unabhängigen Ziehungen eine Summe $\bar{x}$ *zu ziehen, wird für hinreichend große n proportional dem Ausdruck* $e^{-u_1^2-u_2^2-\cdots-u_k^2}$ (wo die u linear aus den Komponenten von $\bar{x}$ zu berechnen sind), *gleichgültig, welche Wahrscheinlichkeiten für die einzelnen Ziehungsergebnisse bestehen, wofern nur die sehr weiten, allgemeinen Voraussetzungen des Fundamentalsatzes* (der „normalen“ Verteilung und der „gleichmäßigen“ Folge) *erfüllt sind.*

In dieser Form und insbesondere auch dann, wenn man die einzelnen Verteilungen als „*geometrisch*“ voraussetzt, d. h. annimmt, daß überall eine endliche Wahrscheinlichkeitsdichte besteht, bietet der erste Fundamentalsatz die *allgemeine und präzise Formulierung des Gaußschen Fehlergesetzes im Sinne der Theorie der Elementarfehler.* Denn man kann unseren Satz b) wie folgt aussprechen:

Läßt sich ein Beobachtungsfehler als Summe von n unabhängigen Einzelfehlern auffassen, so wird die Wahrscheinlichkeitsdichte für eine gegebene Fehlergröße $\bar{x}$ *bei hinreichend großem n proportional dem Ausdruck* $e^{-u_1^2-u_2^2-\cdots-u_k^2}$ (wo die u linear aus den Komponenten von $\bar{x}$ zu berechnen sind), *gleichgültig, welchen Fehlergesetzen die einzelnen Elementarfehler genügen, wofern nur die sehr weiten, allgemeinen Bedingungen des Fundamentalsatzes* (Beschränktheit der Gesamtschwankung und der ersten drei Momente) *erfüllt sind.*

Die bekannten Ableitungen des Fehlergesetzes aus der Kombination von Elementarfehlern gehen von viel engeren Voraussetzungen aus. In der Regel wird (mit $k=1$) als Gesetz des Elementarfehlers angenommen, daß seine Wahrscheinlichkeitsdichte $v_\varkappa(x)$ für einen endlichen Teil der x-Achse gleich einer *konstanten*, von null verschiedenen Größe, im übrigen *null* sei[14]). In diesem speziellen Falle sind natürlich unsere Voraus-

[14]) Vgl. z. B. E. Czuber, Wahrscheinlichkeitsrechnung, Bd. I, 3. Aufl., Leipzig 1914, S. 173 ff.

setzungen über die $v_\varkappa(x)$ reichlich erfüllt. Die korrekte Ableitung des Ergebnisses wird aber kaum einfacher als die allgemeine in § 3, 6 gegebene Betrachtung.

Wir nennen den eben ausgesprochenen Satz, wonach der als Summe sehr vieler Elementarfehler erscheinende Fehler dem Gaußschen Gesetz genügt, den *ersten Satz der Fehlertheorie*, weil ihm ein zweiter ebenso wichtiger — als Folgerung aus dem zweiten Fundamentalsatz — zur Seite tritt (§ 10, 4).

6. Durchschnittsbildung. Erstes Gesetz der großen Zahl. Setzt man in (7′) alle Koeffizienten $h = \frac{1}{n}$ und die $\bar{c} = 0$, so bedeutet die durch (8′) definierte Größe $W_n(\bar{x})$ die Wahrscheinlichkeit dafür, daß das *arithmetische Mittel* oder der Durchschnitt der Merkmale von n unabhängigen Kollektivs „höchstens gleich $\bar{x}$“ wird (d. h. jede Komponente des Schwerpunktvektors der n Merkmalpunkte kleiner oder höchstens gleich der betreffenden Komponente von $\bar{x}$). Fassen wir, um uns konkreter ausdrücken zu können, den Fall einer eindimensionalen, arithmetischen Verteilung ins Auge, so haben wir in $\mathfrak{w}_n(x)$ die Wahrscheinlichkeit dafür, daß das arithmetische Mittel von n unabhängigen Ziehungsergebnissen gerade x beträgt. Nun ist für den allgemeinen eindimensionalen Fall, § 7, 4, der Zusammenhang zwischen dem Argument u der Gaußschen Funktion und der Summe x der Ziehungsergebnisse mit $x = r_n u + b_n$ im Anschluß an Gl. (9), § 7 festgelegt worden. Bezeichnen wir jetzt statt der Summe das arithmetische Mittel mit x, so müssen wir in die vorstehende Beziehung nx statt x einführen, erhalten also für große n:

$$(9) \qquad r_n \mathfrak{w}_n(x) \sim \frac{1}{\sqrt{\pi}} e^{-\left(\frac{nx-b_n}{r_n}\right)^2},$$

oder nach einfacher Umformung

$$(9') \qquad \mathfrak{w}_n\left(\frac{b_n}{n} + x\right) \sim \frac{1}{r_n\sqrt{\pi}} e^{-\frac{n^2}{r_n^2}x^2}.$$

Da r_n^2, die doppelte Streuungssumme der einzelnen Verteilungen, nicht stärker als mit n wachsen kann, weil die Streuungen in allen von uns betrachteten Fällen eine beschränkte Folge bilden, so erkennt man, daß der Faktor von x im Exponenten von (9′) mit n unbeschränkt wächst. D. h. *die Wahrscheinlichkeit dafür, daß das arithmetische Mittel der Ziehungsergebnisse um einen festen Betrag x von $\frac{b_n}{n}$ abweicht, geht mit wachsendem n gegen null.*

Man sieht leicht ein, daß dieser Satz für *jede* Durchschnittsbildung bei den von uns betrachteten Verteilungen gilt, aber auch, daß zu seiner

Ableitung keineswegs alle die Überlegungen notwendig sind, die zu dem Fundamentalsatz geführt haben. Es genügt, davon auszugehen, daß, wie in Gl. (10), § 3 festgestellt wurde, die Streuung des Integralproduktes konstant gleich $\frac{1}{2}$ ist. Denn die Bildung des Integralproduktes entspricht der Annahme $h_1 = h_2 = \ldots = \frac{1}{r_n}$, die Durchschnittsbildung der Annahme $h_1 = h_2 = \ldots = \frac{1}{n}$. Daher ist im letzteren Fall die Streuung $\frac{r_n^2}{n^2}$ mal so groß, also gleich

$$(9'') \qquad \frac{1}{2} \cdot \frac{r_n^2}{n^2} = \frac{\Sigma s_\varkappa^2}{n^2} \leqq \frac{S^2}{n},$$

wenn S^2 die obere Schranke der Streuungen bezeichnet. Man erhält so den allgemeinen Satz, der in der Regel als das „Gesetz der großen Zahl" bezeichnet wird und für uns als eine spezielle Folgerung aus dem ersten Fundamentalsatz *„erstes" Gesetz der großen Zahl* heißen soll:

Die fortgesetzte Durchschnittsbildung von Kollektivs mit Verteilungen, deren Streuungen eine beschränkte Folge bilden, führt zu Verteilungen mit unbeschränkt abnehmender Streuung.

Auf diesen einfachen Gedankengang haben erstmals Tschebyschef und Markoff die Ableitung des Gesetzes der großen Zahl zurückgeführt[15]). Geht die Streuung $s^2 = \int (x-a)^2 v(x) dx$ gegen null, so muß auch das Integral von $v(x)$ für jedes Intervall, das den Punkt $x = a$ nicht enthält, gegen null, für jedes andere gegen eins gehen. In engerem Anschluß an die übliche Ausdrucksweise der Wahrscheinlichkeitsrechnung wäre daher der Satz etwa so auszusprechen:

Die Wahrscheinlichkeit dafür, daß der Durchschnitt von n unabhängigen Ziehungsergebnissen höchstens um ε von seinem Mittelwert abweicht, geht bei festem, beliebig kleinem ε, aber unbeschränkt wachsendem n, gegen 1.

Mit fast unverändertem Wortlaut gilt der Satz auch für mehrdimensionale Verteilungen.

§ 9.

Das Bayessche Problem und seine Verallgemeinerung.

1. Bayessches Problem. In ähnlicher Beziehung wie die von Laplace gegebene Lösung des Bernoullischen Problems zum ersten Fundamentalsatz, steht die ebenfalls von Laplace angegebene Form des

[15]) Vgl. A. A. Markoff, Wahrscheinlichkeitsrechnung, dtsch. von H. Liebmann, Leipzig 1912, S. 58. — Es sei noch bemerkt, daß mitunter der ganze Inhalt des Laplaceschen Resultats, einschließlich der Konvergenz von $W_\varkappa$ gegen die Gaußsche Funktion als „Gesetz der großen Zahl" bezeichnet wird.

Bayesschen Theorems, die oft auch als eine „*Umkehrung des Bernoullischen Satzes*" bezeichnet wird, zum „*zweiten Fundamentalsatz*" der Wahrscheinlichkeitsrechnung. Wir geben zunächst eine Darstellung der Aufgabe und ihrer Lösung im Anschluß an die übliche Ausdrucksweise; die Einordnung in das durch „Grundoperationen" und ihre Zusammensetzung, nach § 8, festgelegte System der Wahrscheinlichkeitsrechnung soll im § 10 erfolgen.

Für eine Urne, die schwarze und weiße (oder mit null und eins bezeichnete) Kugeln enthält, betrage die Wahrscheinlichkeit eines weißen Zuges x $(0 < x < 1)$. In n Zügen (vor denen jedesmal die vorher gezogene Kugel zurückgelegt wurde), seien an mal $(0 < a < 1)$ weiße Kugeln gezogen worden. Die Wahrscheinlichkeit dieser Verbindung ist

$$(1) \qquad \binom{n}{a n} [x^a (1-x)^{1-a}]^n.$$

Die Fragestellung von Bayes geht nun dahin: Wenn a bekannt ist, welche Wahrscheinlichkeit besteht dafür, daß die n Züge aus einer Urne erfolgten, der gerade die Wahrscheinlichkeit x eines weißen Zuges entspricht? Laplace nimmt an, daß die „a priori-Wahrscheinlichkeit" eines jeden Wertes von x dieselbe ist, und setzt die gesuchte Wahrscheinlichkeitsdichte $w_n(x)$ an der Stelle x:

$$(2) \qquad w_n(x) = C_n [x^a (1-x)^{1-a}]^n, \qquad 0 \leqq x \leqq 1,$$

wo C_n so zu bestimmen ist, daß

$$(2') \qquad \int_0^1 w_n(x)\, dx = 1$$

wird.

Um den asymptotischen Wert von (2) für sehr große n zu erhalten, ziehen wir unsere Sätze I und II über den Grenzwert eines Produktes von Funktionen heran. Setzt man

$$(3) \qquad \begin{aligned} f_1(x) = f_2(x) = \ldots = f(x) &= \left(\frac{x}{a}\right)^a \left(\frac{1-x}{1-a}\right)^{1-a} && \text{für} \quad 0 < x < 1, \\ &= \quad 0 && \text{,,} \quad x \geqq 1, \quad x \leqq 0, \end{aligned}$$

so genügen diese $f(x)$ sämtlichen Voraussetzungen des § 1, sobald nur $0 < a < 1$. Insbesondere erreicht $f(x)$ an der Stelle $x = a$ sein Maximum in dem Wert 1 und man hat:

$$(3') \qquad f(a) = 1, \qquad f'(a) = 0, \qquad f''(a) = -\frac{1}{a(1-a)},$$

somit nach (a 1) und (4) in § 1:

$$(3'') \qquad s_1^2 = s_2^2 = \ldots = \frac{1}{2\,a(1-a)}, \qquad r_n^2 = \frac{n}{2\,a(1-a)}.$$

Da sich $w_n(x)$ nach (2) nur um einen konstanten Faktor von dem

Produkt $p_n = f_1 \cdot f_2 f_3 \ldots f_n$ unterscheidet und dieser Faktor gemäß (2′) zu wählen ist, so führt unser Satz I zu der Aussage:

$$(4\,\mathrm{a})\quad w(u) = \lim_{n=\infty} w_n\left(a + u\sqrt{\frac{2a(1-a)}{n}}\right) = \sqrt{\frac{n}{2\pi a(1-a)}}\, e^{-u^2} = \sqrt{\frac{n}{2a(1-a)}}\, \varphi(u),$$

und Satz II zu

$$(4\,\mathrm{b})\quad W(u) = \lim_{n=\infty} W_n\left(a + u\sqrt{\frac{2a(1-a)}{n}}\right) = \lim_{n=\infty} \int^{a+u\sqrt{\frac{2a(1-a)}{n}}} w_n(x)\,dx = \frac{1}{\sqrt{\pi}} \int^{u} e^{-x^2}\,dx = \Phi(u).$$

Als *Näherungsformeln für große* n geschrieben, lauten (4a) und (4b):

$$(4'\mathrm{a})\qquad w_n(x) \sim \sqrt{\frac{n}{2a(1-a)}}\, \varphi\left(\sqrt{\frac{n}{2a(1-a)}}\,(x-a)\right),$$

$$(4'\mathrm{b})\qquad W_n(x) \sim \Phi\left(\sqrt{\frac{n}{2a(1-a)}}\,(x-a)\right).$$

Dies ist die sog. Umkehrung des Bernoullischen Theorems, die in der üblichen Ausdrucksweise besagt:

Die aus einer sehr großen Anzahl von Ziehungen aus einer Urne mit zweierlei Kugeln erschlossene Wahrscheinlichkeit dafür, daß die Wahrscheinlichkeit eines Einzelzuges x beträgt, folgt dem Gaußschen Gesetz.

Die von Laplace herrührende und heute allgemein vorherrschende Ableitung dieses Satzes besteht darin, daß für die spezielle, in (3) gegebene Form der Funktionen $f_\varkappa$ die Überlegungen durchgeführt werden die uns in § 1 zu dem allgemeinen Satz I geführt haben[16]; die Geltung von (4 b) wird dann aus (4 a) in meist unzulässiger Weise erschlossen.

2. Einfluß ungleicher „a priori-Wahrscheinlichkeit". Bayes hat auch den Fall betrachtet, daß die Wahrscheinlichkeit der verschiedenen Werte von x von vornherein nicht dieselbe ist. Nehmen wir an, es sei eine beliebige *stetige* Funktion $\psi(x)$ gegeben, welche die „a priori"-Wahrscheinlichkeitsdichte an der Stelle x angibt, dann tritt an Stelle von (2):

$$(5)\qquad w_n(x) = C_n\, \psi(x)\left[x^a(1-x)^{1-a}\right]^n,$$

wo C_n wieder aus der Bedingung (2′) zu bestimmen ist. Nach dem in § 2, 2 abgeleiteten Korollar folgt aber, daß $\lim w_n$ sich nur um den konstanten Faktor $\psi(a)$ von dem früheren Wert unterscheiden kann. Da dieser Faktor, den wir von null verschieden voraussetzen, bei Bestimmung von C_n nach (2′) wieder herausfallen muß, so sieht man:

[16]) Für den Beweisansatz von H. Bruns, a. a. O. S. 232, gilt wieder das in Fußnote [10]) Gesagte.

Die Ergebnisse (4 a), (4 b) *und* (4′a), (4′b) *gelten unverändert, wenn eine beliebige stetige, an der Stelle $x = a$ nicht verschwindende a priori-Wahrscheinlichkeit für x vorausgesetzt wird.*

E. Czuber hat unter speziellen Annahmen für ψ erkannt, daß der Einfluß von $\psi(x)$ auf w_n bei großem n zurücktritt[17]). Der entscheidende, allgemeine Wert des Laplaceschen Resultates (4) liegt offenbar erst in seiner hier bewiesenen Unabhängigkeit von der Annahme über die „a priori-Wahrscheinlichkeit".

Wenn $\psi(x)$ an der Stelle $x = a$ eine Unstetigkeit erster Art hat, also einen rechten und einen linken Grenzwert $\psi(a + 0) = \psi_1$ und $\psi(a - 0) = \psi_2$, so ist

$$(5') \quad \lim_{n=\infty} w_n\left(a + u\sqrt{\frac{2a(1-a)}{n}}\right) = \frac{2\psi_1}{\psi_1+\psi_2}\sqrt{\frac{n}{2a(1-a)}}\,\frac{1}{\sqrt{\pi}}\,e^{-u^2} \quad \text{für } u > 0$$

$$= \frac{2\psi_2}{\psi_1+\psi_2}\sqrt{\frac{n}{2a(1-a)}}\,\frac{1}{\sqrt{\pi}}\,e^{-u^2} \quad \text{für } u < 0.$$

Hieraus folgt der Wert von W_n durch Integration; an der Stelle $x = a$ gibt es keine bestimmte Wahrscheinlichkeitsdichte $w(0)$.

Wird die a priori-Wahrscheinlichkeit durch eine arithmetische Verteilung dargestellt, so kann man wohl für endliches n

$$W_n(x) = C_n \sum \psi(x_\iota)\left[x_\iota^a (1 - x_\iota)^{1-a}\right]^n$$

ansetzen, wobei die Summe über alle $x_\iota \leqq x$ zu erstrecken und C_n wieder aus (2′) zu bestimmen ist. Aber man kann daraus nur schließen, daß die Einzelwahrscheinlichkeit jenes Wertes x_ι, der a zunächst liegt, mit zunehmendem n alle anderen überwiegt und unbeschränkt nahe an 1 heranrückt.

3. Mehrdimensionaler Fall. Wenn die Urnen, aus denen gezogen wird, nicht nur schwarze und weiße (oder mit Null und Eins bezeichnete) *sondern* 1, 2, ..., *k wertige Kugeln* oder Lose enthalten, so führt die Bayessche Fragestellung auf den durch die Korollare I_4 und II_4 in § 2 gedeckten Fall eines Produktes von Funktionen mehrerer Variabler. Das ist *nicht* analog den Verhältnissen bei der Bernoullischen Problemgruppe, wo das Vorhandensein mehrwertiger Lose noch nicht zum mehrdimensionalen Fall führte (§ 7, 4), sondern erst die Betrachtung von Ziehungsserien als Elementen (§ 7, 5).

Es seien $x_1, x_2, \ldots, x_k$ die Wahrscheinlichkeiten eines Einzelzuges, der „eins" bzw. „zwei" ... „k" ergibt, und es seien tatsächlich in n Zügen

[17]) E. Czuber, Wahrscheinlichkeitsrechnung, Bd. I, 3. Aufl., Leipzig 1914, S. 218.

$na_1, na_2, \ldots, na_k$ mit $1, 2, \ldots, k$ bezeichnete Kugeln gezogen worden, so daß

$$(6) \qquad x_1 + x_2 + \ldots + x_k = 1, \quad a_1 + a_2 + \ldots + a_k = 1.$$

Die Wahrscheinlichkeit dieses Resultates bei gegebenen $x_1, x_2, \ldots, x_k$ wäre proportional

$$(6') \qquad \left[x_1^{a_1} x_2^{a_2} \ldots x_{k-1}^{a_{k-1}} (1 - x_1 - x_2 - \ldots - x_{k-1})^{a_k}\right]^n.$$

Sind durch $\psi_1(x_1), \psi_2(x_2), \ldots, \psi_{k-1}(x_{k-1})$ die „a priori-Wahrscheinlichkeiten" der einzelnen x-Werte gegeben, so hat man als Beantwortung der Bayesschen Frage nach der Wahrscheinlichkeitsdichte $w_n(x_1, x_2, \ldots, x_{k-1})$ dafür, daß unter Annahme des durch $na_1, na_2, \ldots, na_k$ bestimmten Ziehungsresultates die Wahrscheinlichkeiten der einzelnen Kugelarten bei $x_1, x_2, \ldots, x_{k-1}, x_k$ liegen:

$$(6'') \qquad w_n(x_1, x_2, \ldots, x_{k-1}) = C_n \psi_1(x_1) \psi_2(x_2) \ldots \psi_{k-1}(x_{k-1}) \times \\ \left[x_1^{a_1} x_2^{a_2} \ldots x_{k-1}^{a_{k-1}} (1 - x_1 - x_2 - \ldots - x_{k-1})^{a_k}\right]^n,$$

wobei C_n durch die Bedingung

$$(6''') \qquad \int\int \ldots \int w_n(x_1 x_2 \ldots x_{k-1}) dx_1 dx_2 \ldots dx_{k-1} = 1$$

zu eliminieren ist.

Um den asymptotischen Ausdruck für $(6'')$ zu finden, setzen wir im Sinne von § 2, 4:

$$(7) \quad f_1 = f_2 = \ldots = f = \left(\frac{x_1}{a_1}\right)^{a_1} \left(\frac{x_2}{a_2}\right)^{a_2} \ldots \left(\frac{x_{k-1}}{a_{k-1}}\right)^{a_{k-1}} \left(\frac{1 - x_1 - x_2 - \ldots - x_{k-1}}{a_k}\right)^{a_k}, \\ \text{für } 0 < x_\iota < 1, \\ = 0, \quad \text{für} \quad x_\iota \leqq 0, \ x_\iota \geqq 1.$$

Die Funktion $f(x_1 \ldots x_{k-1})$ erreicht im Punkte $x_1 = a_1$, $x_2 = a_2 \ldots x_{k-1} = a_{k-1}$ ihr Maximum mit dem Wert 1. Man findet

$$(7') \qquad f(a_1, a_2, \ldots, a_{k-1}) = 1, \quad \frac{\partial f}{\partial x_\iota}(a_1 \ldots a_{k-1}) = 0,$$

$$\frac{\partial^2 f}{\partial x_\iota^2}(a_1 \ldots a_{k-1}) = -\left(\frac{1}{a_\iota} + \frac{1}{a_k}\right), \quad \frac{\partial^2 f}{\partial x_\iota \partial x_\varkappa}(a_1 \ldots a_{k-1}) = -\frac{1}{a_k} \qquad (\iota \neq \varkappa)$$

und daraus für die mit $(5''')$ in § 2 eingeführten Größen h:

$$(7'') \qquad h_{\iota,\varkappa} = \frac{1}{2a_k} \quad (\iota \neq \varkappa), \qquad h_{\iota,\iota} = \frac{1}{2}\left(\frac{1}{a_\iota} + \frac{1}{a_k}\right).$$

Setzt man

$$(7''') \qquad Q(z_1, z_2, \ldots, z_{k-1}) \equiv \sum_{\iota,\varkappa} h_{\iota,\varkappa} z_\iota z_\varkappa = \frac{1}{2}\left[\sum_{\iota=1}^{k-1} \frac{z_\iota^2}{a_\iota} + \frac{1}{a_k}\left(\sum_{\iota=1}^{k-1} z_\iota\right)^2\right],$$

so besagt unser Korollar I_4:

$$\text{(8 a)} \qquad \lim_{n=\infty} w_n\left(a_1 + \frac{z_1}{\sqrt{n}},\ a_2 + \frac{z_2}{\sqrt{n}}, \ldots, a_{k-1} + \frac{z_{k-1}}{\sqrt{n}}\right)$$

$$= \sqrt{\frac{n^{k-1}}{(2\pi)^{k-1}\, a_1 a_2 \ldots a_k}}\; e^{-Q(z_1, z_2, \ldots, z_{k-1})}$$

und analog II_4:

$$\text{(8 b)} \qquad \lim_{n=\infty} W_n\left(a_1 + \frac{z_1}{\sqrt{n}}, \ldots, a_{k-1} + \frac{z_{k-1}}{\sqrt{n}}\right)$$

$$= \int^{a_1 + \frac{z_1}{\sqrt{n}}} \cdots \int^{a_{k-1} + \frac{z_{k-1}}{\sqrt{n}}} w_n(x_1, x_2, \ldots, x_{k-1})\, dx_1\, dx_2 \ldots dx_{k-1}$$

$$= \sqrt{\frac{1}{2^{k-1}\, a_1, a_2, \ldots, a_k}} \int^{z_1}\int^{z_2} \cdots \int^{z_{k-1}} e^{-Q(x_1, x_2, \ldots, x_{k-1})}\, dx_1\, dx_2 \ldots dx_{k-1}.$$

Der Faktor vor dem Integralzeichen rechts in (8 b) ist die Quadratwurzel aus der Determinante der quadratischen Form Q. Wir haben somit als Verallgemeinerung des Laplace-Bayesschen Theorems in der üblichen Darstellungsform den Satz:

Die aus einer sehr großen Zahl von Ziehungen aus einer Urne mit k verschiedenen Kugelarten erschlossene Wahrscheinlichkeit dafür, daß die Wahrscheinlichkeit eines Einzelzuges jeder Kugelart bei x_1 bzw. $x_2, \ldots, x_k$ liegt, folgt einem $(k-1)$-dimensionalen Gaußschen Gesetz, gleichgültig, welche a priori-Wahrscheinlichkeiten für die $x_1, x_2, \ldots, x_k$ angenommen wurden.

Wir werden diesen Satz mit gewissen Verallgemeinerungen, im Rahmen der in § 8 angedeuteten rationellen Begriffsbildung der Wahrscheinlichkeitsrechnung, weiter unten als den „zweiten Fundamentalsatz der Wahrscheinlichkeitsrechnung" aussprechen. Hier soll noch eine Folgerung dargelegt werden, die einem von Laplace formulierten, von Bienaymé im wesentlichen bewiesenen, speziellen Resultat entspricht.

4. Laplace-Bienayméscher Satz. Laplace hat als allgemeine Umkehrung des Bernoullischen Problems folgende Frage behandelt. Aus einer Urne, die k verschiedene, etwa mit $1, 2, 3, \ldots, k$ bezeichnete Lose enthält, werden (wie oben) in n Zügen $a_1 n$ Lose der ersten, $a_2 n$ der zweiten, …, $a_k n$ der letzten Art gezogen; die unbekannte Wahrscheinlichkeit eines Zuges der ersten Art sei wieder x_1, die eines Zuges zweiter Art x_2, … usf. bis x_k, so daß wieder die beiden Gl. (6) gelten. Setzt man nun

$$(9) \qquad A = \sum_{m=1}^{k} m\, a_m, \qquad X = \sum_{m=1}^{k} m\, x_m,$$

wobei A der „scheinbare“, X der „wahre“ Mittelwert der in der Urne enthaltenen Werte heißt, so wird nach der Wahrscheinlichkeitsdichte $v_n(x)$ dafür gefragt, daß die Abweichung $X - A$ bei x liegt. Der „wahre“ Mittelwert X ist nichts anderes als die Größe, die nach der Definition des Wahrscheinlichkeitsbegriffes als arithmetisches Mittel einer ins Unendliche fortgesetzten Ziehungsserie sich ergeben müßte. Für große n folgt $v_n(x)$ nach Laplace dem Gaußschen Gesetz.

Wir erkennen unschwer in dieser Fragestellung eine spezielle, in dem allgemeinen mehrdimensionalen Bayesschen Problem enthaltene Frage, deren Beantwortung mit Hilfe von (8a) leicht möglich ist. Integriert man nämlich $w_n(x_1, x_2, \ldots, x_{k-1})$ über alle Wertverbindungen $x_1, x_2, \ldots, x_{k-1}$ bei konstantem X, so gewinnt man die Wahrscheinlichkeitsdichte für das Zutreffen des „wahren“ Mittelwertes X. Die Vertauschung der Reihenfolge des Grenzüberganges zu $n = \infty$ und der Integration ist zulässig, weil die f in Abschn. 3 auch alle Voraussetzungen des Satzes II, also auch des Korollars II_2 erfüllen. Schreiben wir

$$(9') \qquad x_m = a_m + \frac{z_m}{\sqrt{n}}, \qquad X = A + \frac{u}{\sqrt{n}}, \qquad \text{also} \quad u = \sum_{m=1}^{k} m\, z_m,$$

so erhalten wir unter Verwendung von (8a):

$$(9'') \quad \lim_{n=\infty} v_n\left(\frac{u}{\sqrt{n}}\right) = \sqrt{\frac{1}{(2\pi)^{k-1} a_1 a_2 \ldots a_k}} \int\int \ldots \int e^{-Q(z_1, z_2, \ldots, z_{k-1})} dz_1\, dz_2 \ldots dz_{k-2},$$

wobei in der durch $(7''')$ definierten quadratischen Form Q die Variable z_{k-1} durch die dritte Gleichung $(9')$ zu eliminieren ist, unter Beachtung von Gl. (6), die jetzt

$$(9''') \qquad \sum_{m=1}^{k} z_m = 0$$

ergibt.

Ersetzt man die Variablen $z_1 \ldots z_{k-1}$ durch (neue) $x_1 \ldots x_{k-1}$ mit Hilfe von

$$(10) \qquad (m - k)\, z_m = x_m,$$

so entsteht aus $(9''')$ und (6):

$$x_1 + x_2 + \ldots + x_{k-1} = u, \tag{10'}$$

und die quadratische Form $(7''')$ geht über in

$$Q = \sum_{\iota,\varkappa} h_{\iota,\varkappa} z_\iota z_\varkappa = \sum_{\iota,\varkappa} \frac{h_{\iota,\varkappa}}{(\iota-k)(\varkappa-k)} x_\iota x_\varkappa = \sum_{\iota,\varkappa} c_{\iota,\varkappa} x_\iota x_\varkappa, \tag{10''}$$

womit die neuen Koeffizienten $c_{\iota,\varkappa}$ zufolge $(7'')$ definiert sind durch:

$$c_{\iota,\varkappa} = \frac{1}{2a_k(\iota-k)(\varkappa-k)} \quad \text{für} \quad \iota \neq \varkappa; \qquad c_{\iota,\iota} = \frac{1}{2(\iota-k)^2}\left[\frac{1}{a_k} + \frac{1}{a_\iota}\right]. \tag{10'''}$$

Um das Integral $(9'')$ zu berechnen, muß man in den Ausdruck $(10'')$ von Q für x_{k-1} den Wert $u - x_1 - \ldots - x_{k-2}$ einführen. Dadurch entsteht eine Funktion $F(x_1, x_2, \ldots, x_{k-2})$ zweiten Grades in den $x_1 \ldots x_{k-2}$, die man durch „Transformation auf Mittelpunkt" zerlegen kann in eine quadratische Form Q' in den $k-2$ Variablen $x_\iota - \alpha_\iota$ (wo die α_ι bestimmte Konstante) und ein von den Variablen $x_1 \ldots x_{k-2}$ freies Glied Q_0 gleich $F(\alpha_1, \alpha_2, \ldots, \alpha_{k-2})$:

$$Q = F(x_1, x_2, \ldots, x_{k-2}) = Q'(x_1, x_2, \ldots, x_{k-2}) + Q_0. \tag{11}$$

Die Konstante Q_0 ist geometrisch dadurch definiert, daß die Fläche zweiten Grades $Q = Q_0$ von der Ebene $(10')$ berührt werden muß. Hieraus ergibt sich leicht für Q_0 der Wert

$$Q_0 = \frac{D}{D_1 + D_2 + \ldots + D_{k-1}} u^2, \tag{11'}$$

wo D die Determinante $|c_{\iota\varkappa}|$ der quadratischen Form Q, D_m diejenige Determinante bedeutet, die aus $|c_{\iota\varkappa}|$ hervorgeht, wenn darin die m-te Spalte durch lauter 1er ersetzt wird. Für die durch $(10''')$ gegebenen Werte von $c_{\iota\varkappa}$ findet man:

$$\frac{1}{D} = 2^{k-1}(k-1)!\,(k-1)!\, a_1 a_2 \ldots a_k, \quad \frac{D_m}{D} = 2a_m(k-m)(A-m) \tag{11''}$$

und daraus

$$u^2 = 2Q_0\left[\sum_{m=1}^{k} m^2 a_m - A^2\right] = 2Q_0 \sum_{m=1}^{k} (m-A)^2 a_m. \tag{11'''}$$

Da das Integral in $(9'')$ bis auf einen konstanten (von u unabhängigen) Faktor gleich e^{-Q_0} ist, hat man

$$\lim_{n=\infty} v_n\left(\frac{u}{r_n}\right) = \frac{r_n}{\sqrt{\pi}} e^{-u^2} \quad \text{mit} \quad r_n^2 = \frac{n}{2\sum_{m=1}^{k}(m-A)^2 a_m}. \tag{12}$$

Dies ist im wesentlichen das von Laplace angegebene Resultat, für das I. J. Bienaymé einen allgemeinen Beweisansatz gegeben hat[18]). Wir kommen auf die Bedeutung dieses Satzes und seiner Verallgemeinerung in § 10, 4 zurück. Ohne weiteres erkennt man, daß, wenn die k verschiedenen Kugelarten statt mit $1, 2, \ldots, k$, mit irgendwelchen Zahlen $b_1, b_2, \ldots, b_k$ bewertet würden, nur in der Definition von A, X und r_n^2, Gl. (9) und (12) b_m an Stelle von m zu setzen wäre.

§ 10.

Der zweite Fundamentalsatz.

1. Abhängige Kollektivs. Zwei Kollektivs können in verschiedener Weise so miteinander verknüpft sein, daß die Verteilung des einen eine Funktion auch des zweiten Merkmales ist. In diesem Falle sagen wir, die Kollektivs seien voneinander „*abhängig*“. Bei der Definition der „Verbindung“ in § 8, 2 wurde ausdrücklich bemerkt, daß es sich um unabhängige Kollektivs handelte; wir wollen jedoch die Definition jetzt auch auf abhängige ausdehnen. Ein einfaches *Beispiel* ist folgendes: In einer Urne mögen sich a weiße und b schwarze Kugeln befinden und es sei die übliche Annahme berechtigt, daß die Wahrscheinlichkeit eines weißen Zuges $a:(a+b)$ sei. Das erste Kollektiv, das wir hier betrachten, besteht aus den einzelnen Zügen aus der Urne, so daß seine Verteilung durch $\mathfrak{v}_1(0) = a:(a+b)$; $\mathfrak{v}_1(1) = b:(a+b)$ gegeben ist. Das zweite Kollektiv bestehe aus einer Reihe von Einzelzügen, die derart erfolgen, daß vor jedem Zug jedesmal eine Kugel, durch einen Zug als Element des ersten Kollektivs, aus der Urne entfernt wird. Gilt immer noch die Annahme, daß die Wahrscheinlichkeit der Anzahl der betreffenden Kugeln in der Urne proportional ist, so haben wir

$$
(1)\qquad \begin{aligned}
\mathfrak{v}_2(0) &= (a-1):(a+b-1) \quad \text{oder} \quad a:(a+b-1),\\
\mathfrak{v}_2(1) &= b:(a+b-1) \quad\quad\;\; \text{oder} \quad (b-1):(a+b-1).
\end{aligned}
$$

Die ersten Werte gelten, wenn beim ersten Zug eine weiße, die zweiten, wenn beim ersten Zug eine schwarze Kugel gezogen wurde. Nennen wir x_1 das Merkmal des ersten Kollektivs ($x_1 = 0$ weiß, $x_1 = 1$ schwarz), x_2 das des zweiten, so ist die Verteilung V_2 Funktion von x_1 und x_2:

$$
(1')\qquad \begin{gathered}
V_2 = 0 \qquad (x_2 < 0)\\
V_2 = 1 \qquad (1 \leqq x_2)\\
V_2 = \frac{a-1}{a+b-1} \text{ bei } x_1 = 0,\quad V_2 = \frac{a}{a+b-1} \text{ bei } x_1 = 1, \qquad (0 \leqq x_2 < 1).
\end{gathered}
$$

---

18) Mém. d. savants étrangers, Paris, t. V, 1838, S. 513 und t. XV, 1858, S. 615.

Allgemein betrachten wir jetzt Verteilungen, die *von einem k-dimensionalen Parameter* $\bar{y}$ *abhängen*, wobei der Querstrich wieder bedeuten soll, daß $\bar{y}$ für eine Reihe von k skalaren Variablen als Symbol gesetzt ist. Unter $V(\bar{x};\, \bar{y})$ wollen wir also eine Funktion verstehen, die für gewisse gegebene Werte (oder Wertsysteme) von $\bar{y}$, hinsichtlich $\bar{x}$ die Bedingungen des § 6, 4 erfüllt: nämlich in bezug auf jede Komponente von $\bar{x}$ monoton ansteigt, für $\bar{x} = -\infty$ den Wert null, für $\bar{x} = \infty$ den Wert 1 annimmt und überall gleich ihrem oberen Grenzwert ist. Ist der Parameter $\bar{y}$ einer Verteilung $V''(\bar{x}'';\, \bar{y})$ gegebene Funktion des Merkmales $\bar{x}'$ eines andern Kollektivs, so sagen wir das erstgenannte Kollektiv sei von dem zweiten abhängig. Andrerseits sprechen wir von einer *Schar* von Kollektivs, indem wir alle Kollektivs, deren Verteilungen $V''(\bar{x}'';\, \bar{y})$ durch die verschiedenen möglichen Werte von $\bar{y}$ festgelegt werden, ins Auge fassen.

Die *Verbindung eines Kollektivs* mit einem von ihm *abhängigen*, bzw. mit einer Schar von Kollektivs, erklären wir analog der Definition in § 7, 2 wie folgt. Es seien e_1', e_2', ... die Elemente, $\bar{x}'$ das Merkmal, $V'(\bar{x}')$ die Verteilung des ersten Kollektivs. Jedem Wert von $\bar{x}'$ ist eindeutig ein Wert des Parameters $\bar{y}$ zugeordnet, der in der Verteilung $V''(\bar{x}'';\, \bar{y})$ einer Schar von Kollektivs auftritt. Diese Schar habe die Elemente e_1'', e_2'', ..., deren Merkmale die Werte von $\bar{x}''$ sind. Jedes Element $e_\varkappa$ der Verbindung sei eine Kombination eines $e_\varkappa'$ mit einem $e_\varkappa''$, wobei das letztere aus jener Elementfolge (der Schar von Elementfolgen) entnommen ist, deren $\bar{y}$-Wert dem $\bar{x}'$-Wert des betreffenden e' entspricht; Merkmal von e sei die Zusammenfassung von $\bar{x}'$ und $\bar{x}''$. Die Verteilung $V(\bar{x})$, wo $\bar{x}$ die Zusammenfassung von $\bar{x}'$ und $\bar{x}''$ bezeichnet, ist dann

$$(2) \qquad V(\bar{x}) = \int^{\bar{x}'} V''(\bar{x}'';\, \bar{\eta}') \cdot d\,V'(\bar{\xi}'),$$

wenn $\bar{\eta}'$ der dem $\bar{\xi}'$ zugeordnete Wert von $\bar{y}$ ist. Aus (2) folgt im Falle der Unabhängigkeit, d. h. wenn V'' die zweite Variable nicht enthält, der speziellere Ansatz (6) von § 7. Sind die gegebenen Verteilungen „*geometrisch*" und die Wahrscheinlichkeitsdichten $v'(\bar{x}')$ und $v''(\bar{x};\, \bar{y})$, so folgt aus (2):

$$(2') \qquad v(\bar{x}) = v''(\bar{x}'';\, \bar{y}')\, v'(\bar{x}').$$

Sind die V' und V'' *arithmetisch*, so hat man analog:

$$(2'') \qquad \mathfrak{v}(\bar{x}) = \mathfrak{v}''(\bar{x}'';\, \bar{y}')\, \mathfrak{v}'(\bar{x}').$$

Beide Male bezeichnet $\bar{y}'$ den dem $\bar{x}'$ zugehörigen Wert von $\bar{y}$. In der üblichen Ausdrucksweise lautet (2''), wie bekannt, so: Die Wahrscheinlichkeit eines aus der Verbindung zweier abhängiger Ereignisse bestehenden Ereignisses ist das Produkt aus der Wahrscheinlichkeit des ersten mal der des zweiten, wobei diese unter der Annahme, daß das erste eingetroffen sei, berechnet werden muß[19]).

Auf die nähere Erörterung dieser Definitionen und Behauptungen, namentlich auf ihre Anwendbarkeit bei empirischen Beobachtungsreihen, gehen wir hier nicht ein.

2. Kreuzung von Kollektivs. Unter der *Kreuzung von n Kollektivs* mit $k_1, k_2, \ldots, k_n$-dimensionalem Merkmal, die alle von einem einzigen k-dimensionalen Kollektiv abhängen, *mit diesem* verstehen wir die *Zusammensetzung aus der Verbindung der $n+1$ Kollektivs* (durch welche ein $k+k_1+k_2+\ldots+k_n$ dimensionales entsteht) und einer *darauf folgenden Teilung*, die nur das k-dimensionale Merkmal übrig läßt. Nehmen wir, um uns einfacher ausdrücken zu können, zunächst an, sämtliche Verteilungen seien arithmetisch und durch die Einzelwahrscheinlichkeiten $\mathfrak{v}(\bar{x})$, $\mathfrak{v}_1(\bar{x}_1; \bar{y}_1)$, $\mathfrak{v}_2(\bar{x}_2; \bar{y}_2), \ldots, \mathfrak{v}_n(\bar{x}_n; \bar{y}_n)$ gegeben, so bedeutet die Einzelwahrscheinlichkeit $\mathfrak{w}_n(\bar{x})$ der Kreuzung, wo die $\bar{y}_\varkappa$ gegebene Funktionen von $\bar{x}$ sind:

$$\bar{y}_\varkappa = \bar{y}_\varkappa(\bar{x}) \tag{3}$$

folgendes: Die Wahrscheinlichkeiten $\mathfrak{v}_1, \mathfrak{v}_2, \ldots, \mathfrak{v}_n$ dafür, bei n verschiedenen Ziehungen die Ergebnisse $\bar{x}_1$, bzw. $\bar{x}_2 \ldots \bar{x}_k$ zu erhalten, hängen von einer einzigen Variablen $\bar{x}$, oder von einer „Ursache“ $\bar{x}$ ab, für welche eine „*a* priori-Wahrscheinlichkeit“ $\mathfrak{v}(\bar{x})$ gegeben ist. Man fragt nach der „*a posteriori-Wahrscheinlichkeit*“ $\mathfrak{w}_n(\bar{x})$ *der Ursache $\bar{x}$, gestützt auf bestimmte Ziehungsergebnisse*, die in einem gegebenen Wertsystem der $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n$ zum Ausdruck kommen. Es wird nach $(2'')$ mit Rücksicht auf (5) in § 8:

$$\mathfrak{w}_n(\bar{x}) = \text{konst.}\, \mathfrak{v}(\bar{x})\, \mathfrak{v}_1(\bar{x}_1; \bar{y}_1[\bar{x}])\, \mathfrak{v}_2(\bar{x}_2; \bar{y}_2[\bar{x}]) \ldots \mathfrak{v}_n(\bar{x}_n; \bar{y}_n[\bar{x}]), \tag{3'}$$

wobei man für $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n$ die gegebenen Werte einzusetzen und die konst. aus der Bedingung zu bestimmen hat, daß die über alle möglichen $\bar{x}$-Werte erstreckte Summe der $\mathfrak{w}_n(\bar{x})$ gleich 1 wird.

Im allgemeinen Fall *beliebiger Verteilungen* $V(\bar{x})$, $V_1(\bar{x}_1; \bar{y}_1)$, $V_2(\bar{x}_2; \bar{y}_2), \ldots, V_n(\bar{x}_n; \bar{y}_n)$ muß für jeden der n Merkmalräume der Verteilungen V_1 bis V_n ein Raumteil $R_1, R_2, \ldots, R_n$ vorgegeben sein, für den die „Teilung“ oder „Aussonderung“ zu bilden ist. Setzen wir das über $R_\varkappa$ erstreckte Integral von $dV_\varkappa$:

$$\int_{(R_\varkappa)} dV_\varkappa(\bar{x}_\varkappa; \bar{y}_\varkappa) = f_\varkappa(\bar{y}_\varkappa), \qquad (\varkappa = 1, 2, \ldots, n), \tag{3''}$$

---

[19]) Vgl. z. B. A. Markoff, Wahrscheinlichkeitsrechnung, Dtsch. v. H. Liebmann, Leipzig 1912, S. 12.

wo bei der Integration $\bar{y}_\varkappa$ festzuhalten ist, so wird die gesuchte Verteilung:

$$(4) \qquad W_n(\bar{x}) = \text{konst.} \int^{\bar{x}} f_1(\bar{\eta}_1) \cdot f_2(\bar{\eta}_2) \ldots f_n(\bar{\eta}_n)\, dV(\bar{\xi}),$$

die Integration erstreckt über alle $\bar{\xi}$, für welche $\xi^{(\iota)} \leqq x^{(\iota)}$. Dabei bedeutet $\bar{\eta}_\varkappa$ den nach (3) zu $\bar{x} = \bar{\xi}$ gehörige Wert von $\bar{y}_\varkappa$. In dem oben erwähnten Fall arithmetischer Verteilungen war angenommen, daß jedes $R_\varkappa$ nur *einen* Gitterpunkt des betreffenden Merkmalraumes einschließt.

In dieser Festlegung der „Kreuzung" als einer speziellen, zusammengesetzten Operation, durch die ein neues Kollektiv aus gegebenen abgeleitet wird, liegt ein wesentlicher Unterschied zwischen unserer Grundlegung der Wahrscheinlichkeitsrechnung und der sonst üblichen Auffassung, in der die Wahrscheinlichkeit „der Ursachen" od. dgl. ein ganz besonderes, auf *andere Erfahrungs- und Erkenntnisquellen* zurückgehendes Kapitel der Wahrscheinlichkeitsrechnung bildet. Doch soll hier auch darauf nicht weiter eingegangen werden.

Wenn die durch (3) dargestellten Abhängigkeiten durchaus *linear* sind:

$$(4') \qquad \bar{y}_\varkappa = \bar{c}_\varkappa + h_\varkappa \bar{x}$$

nennen wir die *Kreuzung* eine *„lineare"*. Auch hier könnte man an Stelle der skalaren Größe $h_\varkappa$ eine Matrix gesetzt denken.

3. Der zweite Fundamentalsatz. Der zweite Fundamentalsatz der Wahrscheinlichkeitsrechnung enthält eine allgemeine Aussage über die Verteilung $W_n(\bar{x})$, zu der im Limes für $n = \infty$ die fortgesetzte lineare Kreuzung von n Kollektivs mit einem, von dem sie alle abhängen, führt, vorausgesetzt, daß alle Abhängigkeiten an *derselben* Stelle ihre Extremwerte erreichen. Um den Satz kurz aussprechen zu können, führen wir den Begriff der „für die Kreuzung normalen Verteilungen" und einer „gleichmäßigen Folge" solcher Verteilungen ein.

Eine von einem Parameter $\bar{y}$ abhängige Verteilung $V_\varkappa(\bar{x}; \bar{y})$ soll als *für die Kreuzung normal* bezeichnet werden, wenn für einen Raumteil $R_\varkappa$ im $\bar{x}$-Raum, für den

$$(5) \qquad \int_{(R_\varkappa)} dV_\varkappa(\bar{x}; \bar{y}) = f_\varkappa(\bar{y})$$

nicht verschwindet, diese Funktion folgende Eigenschaften aufweist:

1. An einer im Endlichen gelegenen Stelle $\bar{y} = \bar{a}_\varkappa$ besitzt $f_\varkappa$ ein endliches Maximum vom Betrage $A_\varkappa$ und endliche Abteilungen der drei ersten Ordnungen (unter denen die der ersten Ordnung verschwinden, die der zweiten eine nicht positive quadratische Form mit den Koeffizienten $-2 s_\varkappa^{(\iota,\lambda)}$ bestimmen).

2. Außerhalb der Stelle $\bar{y} = \bar{a}_\varkappa$ bleibt $f_\varkappa$ unterhalb $A_\varkappa$ (d. h. für $|\bar{y} - \bar{a}_\varkappa| > \varepsilon$ ist $A_\varkappa - f_\varkappa > \delta_\varkappa$, wo $\delta_\varkappa > 0$ nur von ε abhängt) und im Unendlichen verschwindet $f_\varkappa$ mit einer positiven Potenz $\alpha_\varkappa$ von $|\bar{y}|$.

Eine unendliche Folge derartiger Verteilungen derselben Dimension soll *gleichmäßig* heißen, wenn folgendes zutrifft:

1. Die Folge der Komponenten von $\bar{a}_\varkappa$, der $A_\varkappa$ sowie aller Ableitungen der $f_\varkappa$ an der Stelle $\bar{y} = \bar{a}_\varkappa$ bis zur dritten Ordnung einschließlich, ist beschränkt; die Folge der aus den negativ genommenen zweiten Ableitungen gebildeten quadratischen Formen $\sum_{k=1}^{n} s_\varkappa^{(\iota,\lambda)} y^{(\iota)} y^{(\lambda)}$ hat eine von null verschiedene untere Schranke für jedes $\bar{y} \neq 0$.

2. Unter den positiven Zahlen $\delta_\varkappa$ und $\alpha_\varkappa$ gibt es je einen von null verschiedenen Kleinstwert δ bzw. α.

Unter Verwendung dieser Bezeichnungen läßt sich der *zweite Fundamentalsatz der Wahrscheinlichkeitsrechnung* wie folgt aussprechen:

> *Die fortgesetzte lineare Kreuzung unendlich vieler Kollektivs, deren Verteilungen eine gleichmäßige Folge für die Kreuzung normaler Verteilungen bilden, mit einem beliebigen k-dimensionalen Kollektiv von stetiger Verteilungsdichte, bei konstanten, oder wenigstens in endlichen Grenzen liegenden Verhältnissen $h_1 : h_2 : h_3 : \ldots$ in Gleichung (4′) und einem für alle $\varkappa$ gleichen Wert der Differenz $\bar{c}_\varkappa - \bar{a}_\varkappa$, führt auf das Gaußsche Gesetz: Es läßt sich stets eine von n abhängige Koordinatentransformation im k-dimensionalen Raum von $\bar{x}$ in $\bar{u}$* (Verschiebung des Anfangspunktes um $\bar{c}_\varkappa - \bar{a}_\varkappa$, Drehung der Achsen und Veränderung der Maßstäbe) *angeben, so daß die Wahrscheinlichkeitsdichte $w_n(\bar{x})$ der n-ten Kreuzung im Limes für $n = \infty$ proportional $e^{-u_1^2 - u_2^2 - \cdots - u_k^2}$ und die Verteilung selbst gleich dem Produkte $\Phi(u_1) \cdot \Phi(u_2) \ldots \Phi(u_k)$ wird, wo $\Phi(u)$ die Gaußsche Funktion $\frac{1}{\sqrt{\pi}} \int^{u} e^{-x^2} dx$.*

Der Beweis dieses Satzes für den Fall $k = 1$ und die Annahme $h_1 = h_2 = \ldots = h_n$ wird unmittelbar durch die in § 1 gegebenen Beweise für die Behauptungen I und II gedeckt. Die Voraussetzungen, die in der Festlegung der „für die Kreuzung normalen“ Verteilungen und ihrer „gleichmäßigen“ Folgen eingeschlossen sind, stimmen im wesentlichen mit den in § 1, 2 gegebenen überein: Daß das Maximum von $f_\varkappa$ nicht eins, sondern gleich $A_\varkappa$ angenommen wurde, ändert nichts, da bei der „Teilung“ der Kollektivs der Faktor herausfällt. Statt der Voraussetzung (a 2) bzw. (b 2) von § 1 haben wir jetzt der Einfachheit halber die Existenz bzw. Beschränktheit der dritten Ableitungen gefordert, was nur etwas strenger ist.

Die Ausdehnung auf mehrdimensionale Fälle ist in § 2, 4 behandelt, die naheliegende Erweiterung durch Einführung der untereinander verschiedenen Multiplikatoren $h_\varkappa$ in § 6, 5 erwähnt worden. Die $\bar{c}_\varkappa$ in Gleichung (4′) fallen aus der Rechnung völlig heraus, wenn man von vornherein statt $\bar{y}_\varkappa$ die Variable $\bar{y}_\varkappa + \bar{c}_\varkappa - \bar{a}_\varkappa$ einführt.

4. Anwendungen. Zweiter Satz der Fehlertheorie. Als Anwendungen des zweiten Fundamentalsatzes können vor allem die im vorangegangenen Paragraphen behandelten Probleme gelten. Der eindimensionale Bayessche Fall ist in der Weise aufzufassen, daß kleine Gruppen, etwa von n' Ziehungen, das Element eines Kollektivs bilden; das zugehörige Merkmal ist die Häufigkeit a_1 der gezogenen weißen Kugeln, Parameter die Wahrscheinlichkeit x eines weißen Einzelzuges. Die Funktionen $f_1, f_2, \ldots$ sind alle untereinander gleich und im Bereiche $0 < x < 1$ durch $[x^{a_1}(1-x)^{a_2}]^{n'}$ definiert. Die ganze weitere Rechnung zeigt nur eine unbedeutende Änderung gegen die in § 9, 1 durchgeführte. Für die Sache selbst ist es wesentlich, daß bei den aufeinanderfolgenden Gruppen von Ziehungen die relative Häufigkeit a_1 konstant bleibt. Dagegen ist der Fundamentalsatz ohne weiteres anwendbar und führt zu dem gleichen Ergebnis, wenn der Umfang der Gruppen *beliebig in endlichen Grenzen wechselt*, da man ja immer Funktionen f bekommt, die den Voraussetzungen genügen und an derselben Stelle $x = a_1$ ihr Maximum haben. Die Erweiterung bzw. Präzisierung des Bayes-Laplaceschen Theorems, zu der unsere Untersuchungen uns instand setzen, erstreckt sich somit nach folgenden drei Richtungen: Ausdehnung auf *mehrere Dimensionen* und damit Zurückführung des Laplace-Bienayméschen Satzes auf ein allgemeineres Theorem, Zulässigkeit *beliebiger a priori-Wahrscheinlichkeit*, endlich Festlegung des einzig wesentlichen Kriteriums, daß nämlich im Verlaufe der unbeschränkt fortgesetzten Ziehungen *immer wieder Gruppen mit der gegebenen oder angenommenen Häufigkeitsverteilung* $a_1, a_2, \ldots$ auftreten.

Wie der erste Fundamentalsatz zu dem für die *Gaußsche Fehlertheorie* grundlegenden Satz führte, daß die Wahrscheinlichkeit eines als Summe sehr vieler Elementarfehler sich bildenden Fehlers dem Exponentialgesetz genügt, so liefert der zweite ein Resultat, das für die Anwendung der Fehlertheorie unentbehrlich ist, und von dem man stets — allerdings meist unter Verzicht auf jede Beweisführung — Gebrauch macht. Die Frage liegt wie folgt. Man hat eine Gruppe von n' Beobachtungen für eine unbekannte Größe durchgeführt; $a_1 n'$ ergaben den Wert b_1, $a_2 n'$ den Wert $b_2 \ldots$, $a_k n'$ den Wert b_k; *gefragt wird nach der Wahrscheinlichkeit dafür, daß der „wahre“ Wert der Unbekannten bei x liegt.* Wie immer das Fehlergesetz angenommen wird, es gibt gewisse Wahrscheinlichkeiten $x_1, x_2, \ldots, x_k$ für das Auftreten der einzelnen Beobachtungsresultate

$b_1, b_2, \ldots, b_k$, und wenn wir den „wahren" Wert x mit dem Mittelwert $\sum_{m=1}^{k} x_m b_m$ der durch die x_m bestimmten (arithmetischen) Verteilung identifizieren, haben wir fast genau die in § 9, 4 behandelte Aufgabe. Setzen wir

$$A = \sum_{m=1}^{k} a_m b_m, \qquad s^2 = \sum_{m=1}^{k} a_m (b_m - A)^2, \tag{6}$$

so wird die Wahrscheinlichkeitsdichte dafür, daß der wahre Wert bei x liegt, falls die Gruppe von n' Versuchen hinreichend oft wiederholt wird, asymptotisch gleich

$$v_n(x) \sim \frac{1}{s\sqrt{2\pi}} e^{-\frac{(x-A)^2}{2s^2}}. \tag{6'}$$

Denn dies ist nichts anderes als Gleichung (12) von § 9, mit der Bezeichnung (6) und der am Schluß von § 9, 4 erwähnten Ergänzung, in Gestalt einer *Näherungsformel* geschrieben. Ebenso wie im einfachen Bayesschen Fall kann man von diesem Ergebnis immer dann Gebrauch machen, wenn eine große Zahl n von Einzelbeobachtungen vorliegt. Wie man sich diese in Gruppen von n' Beobachtungen geteilt denkt, ist gleichgültig, wenn man nur annimmt, daß in jeder Gruppe die gleichen Häufigkeiten $a_1, a_2, \ldots, a_k$ auftreten.

In der Regel sucht man in der Fehlertheorie das Resultat (6′) dadurch zu gewinnen, daß man schon von dem Fehlergesetz $e^{-h^2x^2}$ ausgeht. Auf diese Weise erhält man wohl durch eine ganz einfache Überlegung für v_n den Ausdruck

$$\frac{h}{\sqrt{\pi}} e^{-h^2(x-A)^2}, \tag{6''}$$

aber daß und wie hier das „Präzisionsmaß" h durch die aus den Beobachtungen abgeleitete Streuungsgröße s^2 zu bestimmen ist, bleibt eine ganz *willkürliche* Annahme. Erst die Anwendung des zweiten Fundamentalsatzes führt zu dem *zweiten*, entscheidenden *Satz der Fehlertheorie*:

Die aus einer großen Zahl von Beobachtungen zu folgernde Wahrscheinlichkeit des „wahren" Wertes x wird durch das Gaußsche Gesetz (6′) *dargestellt, wobei A den Mittelwert und s^2 die Streuung der Beobachtungsergebnisse bedeutet.*

5. Das zweite Gesetz der großen Zahl. Wie in § 8, 6 als ein Teilergebnis der Untersuchungen, die zum ersten Fundamentalsatz führten, das bekannte Gesetz der großen Zahl gewonnen wurde, so können wir jetzt — ebenfalls unter Beschränkung auf ganz einfache Überlegungen —

ein vollkommen *analoges Gesetz über die fortgesetzte lineare Kreuzung* von Kollektivs aussprechen.

Gehen wir etwa von der asymptotisch richtigen Näherungsformel (4′) in § 9 aus und führen wir dabei den Wert r_n aus (3″), § 9 teilweise wieder ein, so haben wir als Aussage des Laplace-Bayesschen Satzes:

$$(7) \qquad w_n(x) \sim \frac{r_n}{\sqrt{\pi}}\, e^{-r_n^2 (x-a)^2} = \frac{r_n}{\sqrt{\pi}}\, e^{-\frac{n(x-a)^2}{2a(1-a)}}.$$

Hier ist charakteristisch, daß der Faktor von $(x-a)^2$ im Exponenten mit n ins Unendliche wächst, ganz so wie in Gleichung (9) oder (9′) von § 8 bei der Durchschnittsbildung. Es folgt aus (7), daß das von $a-\varepsilon$ bis $a+\varepsilon$ erstreckte Integral von $w_n(x)$ bei beliebigem festem ε mit wachsendem *n gegen* 1 *gehen muß*, d. h. aber nichts anderes, als daß die Streuung der durch $w_n(x)$ dargestellten Verteilung unter jeden Betrag sinkt (§ 8, 6). Nun ist leicht zu sehen, daß dies ganz allgemein für jede Kreuzung von sehr vielen Verteilungen (mit gleicher Scheitelstelle) gelten muß wenn man wesentlich nur voraussetzt, daß die *Summe der zweiten Ableitungen* in den Scheitelpunkten mit n ins (negativ) *Unendliche* geht. Denn der zweite Fundamentalsatz beruht auf den im I. Teil dieser Arbeit bewiesenen Behauptungen I und II. Es ist aber in § 1, 6 ausdrücklich gezeigt, daß das Integral des Produktes $p_n(u)$ von n Funktionen für feste, endliche Grenzen von u gebildet, mit wachsendem n gegen $\sqrt{\pi}$ geht, während der Beitrag zum Integral, den die außerhalb liegenden u-Werte liefern, unter jeden Betrag sinkt. Da nun die Variable u sich wesentlich durch den Faktor r_n, gleich der Quadratwurzel aus der halben negativen Summe der zweiten Ableitungen, von $x-a$ unterscheidet, erkennt man, daß in allen Fällen, in denen die Behauptung II anwendbar ist, $w_n(x)$ die behauptete Eigenschaft besitzt. Wir haben so den folgenden, als *zweites Gesetz der großen Zahl* zu bezeichnenden Satz:

Die fortgesetzte lineare Kreuzung von n abhängigen Kollektivs mit gleicher Scheitelstelle, für welche die Summe der zweiten Ableitungen nach dem Parameter an der Scheitelstelle ins Unendliche geht, führt zu Verteilungen mit unbeschränkt abnehmender Streuung.

Wollte man sich völlig dem Gedankengang von § 8, 6 anschließen, so müßte man davon ausgehen, daß (§ 1, 3) die zweite Ableitung von $p_n(u)$ an der Scheitelstelle konstant -2 ist, daß daher die von $w_n(x)$, da sie sich um den Faktor r_n^2 unterscheidet, negativ unendlich wird. Dann kann man weiter schließen, daß auch die Werte von $w_n(x)$ in endlicher Umgebung des Scheitels — höchstens bis auf Punkte, die eine Nullmenge bilden — gegen null gehen usf. In der üblichen Ausdrucksweise der Wahrscheinlichkeitsrechnung wäre das *zweite Gesetz der großen Zahl* etwa so auszusprechen:

Die aus einer großen Zahl n von Beobachtungen erschlossene Wahrscheinlichkeit dafür, daß eine beobachtete Größe um höchstens ε von ihrem wahrscheinlichsten Wert abweicht, geht bei festem, beliebig kleinem ε, aber unbeschränkt wachsendem n, gegen 1.

Dieser Satz, wenn er auch nicht ausdrücklich formuliert, geschweige denn bewiesen wurde, ist seinem wesentlichen Inhalt nach der heutigen Wahrscheinlichkeitsrechnung doch nicht ganz fremd. Es gilt, wie man leicht sieht, mit fast unverändertem Ausdruck für den Fall mehrdimensionaler Verteilungen.

(Eingegangen am 31. August 1918.)