NL Hausdorff: Kapsel 21: Fasz. 64

# [Wahrscheinlichkeitsrechnung]

Vorlesung Univ. Bonn, SS 1923, SS 1931. Hs. Ms. – 227 Bll.

## § 1. Elementare Wahrscheinlichkeiten.

Die Wahrscheinlichkeitsrechnung betrachtet *ungewisse Aussagen* oder Urteile, d. h. solche, deren Richtigkeit oder Falschheit zur Zeit der Betrachtung nicht bekannt ist. Schon im praktischen Leben wird der Grad der Ungewissheit häufig geschätzt (eine Aussage als sehr wahrscheinlich, wahrscheinlicher als das Gegentheil u. dgl. taxirt); die Wahrscheinlichkeitsrechnung versucht diese Graduirung wissenschaftlich genau zu machen, indem sie unter geeigneten Umständen einer Aussage eine bestimmte Zahl, ihre mathematische Wahrscheinlichkeit, zuordnet.

Die ungewisse Aussage bezieht sich häufig auf die Zukunft: „*ein Ereignis $A$ wird eintreten*“. (Typus des Zufallsspiel: es wird mit einem Würfel 4 geworfen werden, es wird aus einer Urne eine weisse Kugel gezogen werden). Aber sie kann auch ein vergangenes Ereignis betreffen (es ist eine Kugel gezogen worden, sie ist aber noch verdeckt) oder einen zeitlosen Thatbestand, dessen Nachprüfung uns praktisch verwehrt ist (die 1000. Decimale von $e$ ist 7). Man spricht, mit Anlehnung an den ersten Aussagentypus, in der Regel von *ungewissen Ereignissen.* Die Wahrscheinlichkeit eines Ereignisses $A$ (oder einer Aussage $A$) werde $w(A)$ genannt.

Zur Einführung dienen uns *Axiome.* | Bl. 2

*Axiom ($\alpha$) Die Wahrscheinlichkeit eines gewissen Ereignisses ist 1.*

Dass man die Grenzfälle einer als richtig oder falsch bekannten Aussage eines gewissen oder unmöglichen Ereignisses nicht ausschliessen wird, ist ebenso Sache der Zweckmässigkeit, wie etwa, dass man eine Constante als Grenzfall einer Variablen ansieht. Die Wahrscheinlichkeit aller gewissen Ereignisse gleich zu setzen ist natürlich, wenn man die praktische Graduirung durch die mathematische Grössenbeziehung nachbilden will (aber man könnte sie natürlich = 100 setzen).

Aus zwei Ereignissen $A, B$ bilden wir zwei weitere, Summe und Produkt: (Aussagen- oder Ereigniscalcul) [1]

$$\begin{aligned} S &= A \cup B : \quad A \textit{ oder } B \text{ (oder beide) treten ein.} \\ P &= AB : \quad A \textit{ und } B \text{ treten ein.} \end{aligned}$$

Für die Summe schreiben wir $S = A + B$, wenn $A$ und $B$ einander ausschliessen ($AB$ unmöglich ist).

Entsprechend für drei oder mehr (endlich viele) Ereignisse. Für die Summe schreiben wir $S = A + B + C$, wenn die drei Ereignisse einander *paarweise* ausschliessen, also $AB, AC, BC$ unmöglich sind. Von Ereignissen $A_1, A_2, \ldots, A_m$, die einander *paarweise* ausschliessen, wird gesagt, dass sie eine *Disjunction* bilden, und zwar eine *vollständige Disjunction*, wenn eines von ihnen eintreten
Bl. 3 muss. Bezeichnet $G$ ein | gewisses Ereignis (von diesem Standpunkt aus sind alle gewissen Ereignisse als gleich zu betrachten), so ist

$$A_1 + A_2 + \cdots + A_m = G$$

die Formel für eine vollständige Disjunction, vorausgesetzt, dass $A_i A_k$ $(i \neq k)$ unmöglich ist. Ein unmögliches Ereignis kann passend mit 0 (Null) bezeichnet werden.

Ferner sei $\bar{A}$ die Negation von $A$, das Gegentheil oder Complement von $A$, das Nichteintreten von $A$ (die Aussage: $A$ ist falsch). $A, \bar{A}$ bilden eine vollständige Disjunction $A + \bar{A} = G$.

Bei zwei Ereignissen $A, B$ bilden $AB, A\bar{B}, \bar{A}B, \bar{A}\bar{B}$ eine vollständige Disjunction.

Wenn $S = A \cup B,\ P = AB,$ so ist $\bar{S} = \bar{A}\bar{B},\ \bar{P} = \bar{A} \cup \bar{B}$; das Complement der Summe ist das Produkt der Complemente, das Complement des Produktes die Summe der Complemente. Analog für drei oder mehr Ereignisse.

Wir fordern nun zweitens

*Axiom* $(\beta)$ (Additionspostulat). *Die Wahrscheinlichkeit, dass von zwei einander ausschliessenden Ereignissen eines eintrete, ist die Summe der Wahrscheinlichkeiten beider Ereignisse:*

$$w(A + B) = w(A) + w(B).$$

Bl. 4 Auch dies überträgt sich auf drei und mehr Ereignisse. | Für eine vollständige Disjunktion ist nach $(\alpha)(\beta)$ die Summe der Wahrscheinlichkeiten gleich 1; insbesondere ist

$$w(A) + w(\bar{A}) = 1, \qquad w(\bar{A}) = 1 - w(A).$$

Die Wahrscheinlichkeit des Gegentheils von $A$ ist 1 minus Wahrscheinlichkeit von $A$. *Die Wahrscheinlichkeit eines unmöglichen Ereignisses ist* 0.

Für zwei Ereignisse $A, B$ hat man

$$\begin{aligned} A &= AB + A\bar{B} \\ B &= AB + \bar{A}B \\ A \cup B &= AB + A\bar{B} + \bar{A}B \end{aligned}$$

und der Additionssatz liefert

$$w(A \cup B) + w(AB) = w(A) + w(B). \tag{1}$$

Bisher haben wir noch keine Wahrscheinlichkeiten ausser $0, 1$ kennen gelernt. Sei jetzt $A_1, A_2, \ldots, A_m$ eine vollständige Disjunktion, und alle ihre Glieder seien *gleichwahrscheinlich*; dann ergiebt sich $w(A_1) = \cdots = w(A_m) = \frac{1}{m}$. Ferner sei $A$ die Summe von $g$ Disjunctionsgliedern, etwa $A = A_1 + \cdots + A_g$, dann ist $w(A) = \frac{g}{m}$. Man sagt: $A_1, \ldots, A_m$ sind die *möglichen Fälle*, $A_1, \ldots, A_g$ die *für $A$ günstigen Fälle*, und erhält also: *Die Wahrscheinlichkeit eines Ereignisses ist das Verhältnis der Zahl der günstigen zur Zahl der möglichen Fälle, alle Fälle als gleichwahrscheinlich vorausgesetzt.* Bl. 5

Dies wird zuweilen als Definition an die Spitze gestellt. Die auf diese Weise erklärten Wahrscheinlichkeiten mögen *elementar* heissen; sie sind nothwendig rationale Zahlen des Intervalls $[0, 1]$. Später (§ 4) erweitern wir das Gebiet.

Beispiele (bei denen zugleich ersichtlich ist, welche Fälle als gleichwahrscheinlich angesehen werden).

Wahrscheinlichkeit, mit einer Münze Bild zu werfen. $w = \frac{1}{2}$.

Wahrscheinlichkeit, mit einem Würfel 4 zu werfen. $w = \frac{1}{6}$.

Wahrscheinlichkeit, mit einem Würfel ungerade zu werfen. $w = \frac{3}{6} = \frac{1}{2}$ .

Wahrscheinlichkeit, aus einem Piquetspiel (32 Karten) einen König zu ziehen. $w = \frac{4}{32} = \frac{1}{8}$.

Wahrscheinlichkeit, eine Figur (König, Dame, Bube) zu ziehen. $w = \frac{12}{32} = \frac{3}{8}$.

Wahrscheinlichkeit, aus einer Urne mit $m$ Kugeln, unter denen $g$ weisse, eine weisse Kugel zu ziehen. $w = \frac{g}{m}$ .

Wahrscheinlichkeit, mit zwei Münzen Bild, Schrift zu werfen. Achtet man auf das Ergebnis jeder einzelnen Münze, so sind die vier Fälle $BB$ (Bild-Bild), $BS$ (Bild-Schrift), $SB$, $SS$ möglich; $BS$ und $SB$günstig. $w = \frac{2}{4} = \frac{1}{2}$.

Wahrscheinlichkeit, mit zwei Würfeln 3, 4 oder 4, 3 zu werfen. $w = \frac{2}{36} = \frac{1}{18}$.

Wahrscheinlichkeit, mit zwei Würfeln 4, 4 zu werfen. w $= \frac{1}{36}$.

Wahrscheinlichkeit, mit zwei Würfeln einen Pasch zu werfen. $w = \frac{6}{36} = \frac{1}{6}$.

Wahrscheinlichkeit, mit zwei Würfeln die Augensumme $x$ zu werfen. $w_x$.

$$\begin{aligned} 2 &= 1+1,\ 3 = 1+2 = 2+1,\ 4 = 1+3 = 2+2 = 3+1, \ldots \\ 12 &= 6+6,\ 11 = 6+5 = 5+6,\ 10 = 6+4 = 5+5 = 4+6, \ldots \\ w_x &= \frac{6-|x-7|}{36} \text{ für } x = 2, 3, \ldots, 12. \end{aligned}$$

Über die gleichwahrscheinlichen Fälle ist viel diskutiert worden. Wenn man Bl. 6
die elementare Wahrscheinlichkeit = Anzahl der günstigen durch Anzahl der

möglichen Fälle *definirt*, falls diese Fälle gleichwahrscheinlich sind, so begeht man einen circulus vitiosus. Zu dessen Verschleierung spricht man von gleich möglichen, gleichberechtigten Fällen. Laplace: Les cas également possibles, c'est à dire tels que nous soyons également indécis sur leur existence. Diesem Prinzip des mangelnden Grundes stellen Andere (von Kries) das des zwingenden Grundes gegenüber und verlangen objektive Kriterien, z. B. die zentrische Lage des Würfelschwerpunkts. Die Wahrscheinlichkeit eines bestimmten Sterbetages wird man nicht $= \frac{1}{365}$ setzen, sondern klimatische Unterschiede zwischen Sommer und Winter machen. Der Mathematiker braucht alle diese empirischen Momente zunächst nicht zu berücksichtigen, sondern betrachtet eine Disjunktion gleichwahrscheinlicher Fälle als *gegeben* und entwickelt daraus die Konsequenzen.

*Relative Wahrscheinlichkeit.* Für $w(A) \neq 0$ nennen wir

$$w_A(B) = \frac{w(AB)}{w(A)} \tag{2}$$

die *Wahrscheinlichkeit von $B$ unter der Annahme, dass $A$ eintritt.*
Der Grund dieser Benennung ist dieser: gehören zu den Ereignissen

$$\begin{array}{cccc} AB & A\bar{B} & \bar{A}B & \bar{A}\bar{B} \\ \alpha & \beta & \gamma & \delta \end{array}$$

Bl. 7 gleichwahrscheinliche Fälle, im ganzen $m = \alpha + \beta + \gamma + \delta$, so ist | $w(A) = \frac{\alpha+\beta}{m}$, $w(AB) = \frac{\alpha}{m}$, also $w_A(B) = \frac{\alpha}{\alpha+\beta}$ und dies ergiebt sich auch so, indem man nur die $\alpha+\beta$ Fälle, wo $A$ eintritt, als mögliche berücksichtigt und hierunter die $\alpha$ Fälle, wo auch $B$ (d. h. $AB$) eintritt, als günstige zählt.

Per definitionem ist $w(AB) = w(A)w_A(B)$, d. h. die Wahrscheinlichkeit, dass $A$ und $B$ zugleich eintreten, ist die Wahrscheinlichkeit von $A$ mal der *unter der Voraussetzung $A$ gültigen* Wahrscheinlichkeit von $B$.

Z. B. sei beim Würfel $A$ das Werfen einer der Zahlen 1, 3, 5, $B$ das einer der Zahlen 4, 5, 6, $AB$ das von 5. $w(A) = \frac{1}{2}$, $w(AB) = \frac{1}{6}$, $w_A(B) = \frac{1}{3}$ (wenn $A$ eintritt, sind nur 3 Fälle 1, 3, 5, wovon einer für $B$ günstig ist). Zugleich ist $\bar{A}$ das Werfen von 2, 4, 6, $\bar{A}B$ das von 4, 6, $w_{\bar{A}}(B) = \frac{2}{3}$. $B$ hat also unter den Annahmen $A, \bar{A}$ zwei verschiedene relative Wahrscheinlichkeiten.

Die Ereignisse $A, B$ heissen *unabhängig*, wenn

$$w(AB) = w(A)\, w(B). \tag{3}$$

Für $w(A) \neq 0$ folgt also $w_A(B) = w(B)$, und da durch Subtraktion der Formel (3) von $w(B)$ | auch $w(\bar{A}B) = w(\bar{A})\, w(B)$ entsteht, so folgt für $w(\bar{A}) \neq 0$, $w_{\bar{A}}(B) = w(B)$. *D. h. wenn $A, B$ unabhängig sind, so ist die Wahrscheinlichkeit für $B$ unter beiden Annahmen $A, \bar{A}$ dieselbe* (falls $w(A)$ weder 0 noch 1 ist). Dasselbe gilt mit Vertauschung von $A, B$. Bl. 8

Wenn umgekehrt

$$\begin{aligned} w_A(B) &= w_{\bar{A}}(B) = w, \text{ so ist} \\ w(B) &= w(AB) + w(\bar{A}B) = w(A)w + w(\bar{A})w = w, \end{aligned}$$

also $w(AB) = w(A)\, w(B)$; $A$ und $B$ unabhängig.

Beispiele: $A$ Ziehen einer weissen Kugel aus einer Urne, $B$ das Werfen von Bild mit einer Münze. Hier ist es naturgemäss anzunehmen, dass $A, B$ unabhängig sind.

In einer Urne seien $s$ Kugeln, darunter $a$ weisse und $s-a$ schwarze; es werden zwei Kugeln gezogen, die erste nicht zurückgelegt; $A, B$ sei das Ziehen von Weiss beim 1. und 2. Zuge. Hier ist $w_A(B) = \frac{a-1}{s-1}$, $w_{\bar{A}}\ (B) = \frac{a}{s-1}$. $B$ ist von dem (früheren) Ereignis $A$ abhängig, aber auch $A$ von dem (späteren) $B$; es ist auch $w_B(A) = \frac{a-1}{s-1}$, $w_{\bar{B}}(A) = \frac{a}{s-1}$ (z. B. für $a = 9$, $s = 10$ : $w_{\bar{B}}(A) = \frac{9}{9} = 1$; wenn
beim 2. Zuge schwarz gezogen wird, so *muss* beim 1. | weiss gezogen worden Bl. 9
sein, da nur eine schwarze Kugel vorhanden ist.) Die Kugel des 2. Zuges muss für den 1. als nicht vorhanden betrachtet werden!

Ist $A$ die Aussage für einen Mann, $20-30$ Jahre alt zu sein (Statistik), $B$ die für eine Frau, $40-50$ Jahre alt zu sein, so wird für ein beliebig herausgegriffenes Paar $w(AB) = w(A)\, w(B)$ sein, nicht aber für ein Ehepaar.

Bei drei Ereignissen ist

$$w(ABC) = w(AB)\ w_{AB}(C) = w(A)\, w_A(B)\, w_{AB}(C).$$

Wenn $C$ von $AB$ und $B$ von $A$ unabhängig ist (aber auch noch in anderen Fällen), ist $w(ABC) = w(A)\, w(B)\, w(C)$.

Es kann $C$ von $A$ und von $B$ unabhängig und doch von $AB$ abhängig sein. Wenn z. B. nur die 4 gleichwahrscheinlichen Fälle

$$ABC, A\bar{B}\bar{C}, \bar{A}B\bar{C}, \bar{A}\bar{B}C$$

möglich sind (es tritt eines oder alle drei ein), so ist $w(A) = w(B) = w(C) = \frac{1}{2}$, $w(AB) = w(BC) = w(AC) = \frac{1}{4}$, $w(ABC) = \frac{1}{4}$; $C$ ist von $A$ und von $B$,
aber nicht von $AB$ unabhängig (sonst müsste $w(ABC) = \frac{1}{8}$ sein). | Bl. 11

Beim Roulette (vom 37. Feld abgesehen) haben wir die Disjunktionen [2]

| Pair | Impair | Manque | Passe | Rouge | Noir |
|---|---|---|---|---|---|
| $A$ | $\bar{A}$ | $B$ | $\bar{B}$ | $C$ | $\bar{C}$ |

(Manque = Nummern 1 – 18, Passe = 19 – 36); zu jeder Kategorie gehören 18 Felder und zu jedem Paar wie $AB, \bar{A}B, \ldots$ 9 Felder. Also $w(A) = \cdots = \frac{1}{2}$, $w(AB) = \cdots = \frac{1}{4}$; es ist also $C$ von $A$ und von $B$ unabhängig, kann aber nicht von $AB$ unabhängig sein, da sonst $w(ABC) = \frac{1}{8}$ sein müsste und doch auf die Klasse $ABC$ nicht $\frac{36}{8} = 4{,}5$ Felder entfallen können. Die Tripel $ABC$ usw. haben in Wahrheit teils die Wahrscheinlichkeit $\frac{4}{36}$, teils $\frac{5}{36}$.

Andererseits kann z. B. auch $C$ von $A$ und von $B$ abhängig, aber von $AB$ unabhängig sein. Die Glieder der vollständigen Disjunktionen mögen die Wahrscheinlichkeiten haben

$$\begin{array}{cccc|cccc} ABC & A\bar{B}C & \bar{A}BC & \bar{A}\bar{B}C & p_0 & p_1 & p_2 & p_3 \\ AB\bar{C} & A\bar{B}\bar{C} & \bar{A}B\bar{C} & \bar{A}\bar{B}\bar{C} & q_0 & q_1 & q_2 & q_3 \end{array}$$

so dass $p = p_0 + p_1 + p_2 + p_3 = w(C), \quad q = q_0 + q_1 + q_2 + q_3 = w(\bar{C})$. |

Man findet dann:

$$w_C(AB) = \tfrac{p_0}{p}, \qquad w_C(A) = \tfrac{p_0+p_1}{p}, \qquad w_C(B) = \tfrac{p_0+p_2}{p}$$

$$w_{\bar{C}}(AB) = \tfrac{q_0}{q}, \qquad w_{\bar{C}}(A) = \tfrac{q_0+q_1}{q}, \qquad w_{\bar{C}}(B) = \tfrac{q_0+q_2}{q}$$

und kann also leicht erreichen, dass $C$ von $AB$ (oder $AB$ von $C$) unabhängig wird, hingegen $A$ von $C$ und $B$ von $C$ abhängig; man nehme etwa $p = q = \frac{1}{2}$; $p_0 = q_0$, $p_1 \neq q_1$, $p_2 \neq q_2$.

*Reihe von Versuchen.* Wir machen $n$ Versuche; beim $k$. Versuch liege die vollständige Disjunktion $A_k + B_k = G_k$ ($G_k$ gewiss, $A_k$ der günstige Fall, $B_k$ der ungünstige) vor. Z. B. Werfen mit einem Würfel. $A_k$ sei jedesmal das Ereignis, dass 4 geworfen wird. (Man lässt dann manchmal den Index weg und sagt, dass es sich um Wiederholung eines und desselben Versuchs handelt). Zunächst habe der günstige Fall jedesmal dieselbe Wahrscheinlichkeit $p = w(A_k)$; $q = w(B_k) = 1 - p$ die des ungünstigen; und wir nehmen an, dass jedesmal die Wahrscheinlichkeit vom Ausfall der vorangehenden Versuche unabhängig ist. Die Wahrscheinlichkeit, dass in einer *bestimmten* Reihenfolge $\alpha$
Bl. 13 mal der günstige, | $\beta$ mal der ungünstige Fall eintrete, ($\alpha+\beta = n$), ist $p^\alpha q^\beta$, und die Wahrscheinlichkeit dass es in irgend einer, gleichviel welcher, Reihenfolge eintrete, ist

$$w_{\alpha\beta} = \binom{n}{\alpha} p^\alpha q^\beta = \frac{n!}{\alpha!\,\beta!}\, p^\alpha q^\beta$$

Es gilt in $x, y$ die Identität

$$(px + qy)^n = \sum w_{\alpha\beta}\, x^\alpha y^\beta \qquad (\textstyle\sum \text{ über } \alpha + \beta = n).$$

Z. B. Wahrscheinlichkeit, dass immer $A$ (der günstige Fall) eintrete: $p^n$; das wird (ausser für $p = 1$) mit wachsendem $n$ beliebig klein.

Wahrscheinlichkeit, dass wenigstens einmal $A$ eintrete: $1 - q^n$; das kommt (ausser für $q = 1$) der 1 beliebig nahe. Bei wieviel Versuchen ist diese Wahrscheinlichkeit $\geq \dfrac{1}{2}$? $\;q^n \leq \dfrac{1}{2}$, $\;\left(\dfrac{1}{q}\right)^n \geq 2$, $\;n \geq \dfrac{\log 2}{\log \frac{1}{q}}$. Z. B. für $p = \dfrac{1}{6}$ : $n \geq \dfrac{\log 2}{\log \frac{6}{5}} = 3,8$; also bei 4 oder mehr Versuchen.

Welches ist der *wahrscheinlichste* Wert von $\alpha$ oder der „Frequenz“ $\dfrac{\alpha}{n}$ (relative Häufigkeit) des günstigen Falles? Es ist

$$\frac{w_{\alpha\beta}}{w_{\alpha-1,\beta+1}} = \frac{n!}{\alpha!\,\beta!} p^\alpha q^\beta \; : \; \frac{n!}{(\alpha-1)!\,(\beta+1)!} p^{\alpha-1} q^{\beta+1} = \frac{\beta+1}{\alpha}\,\frac{p}{q} = \frac{n-\alpha+1}{\alpha}\,\frac{p}{q}$$

Bl. 14 | und dies ist $\begin{smallmatrix}\geq\\=\\<\end{smallmatrix}$ 1 für $(n-\alpha+1)p \begin{smallmatrix}\geq\\=\\<\end{smallmatrix} \alpha q$, d. h. $\alpha \begin{smallmatrix}\leq\\=\\>\end{smallmatrix} (n+1)p$.

Die $w_{\alpha\beta}$, in der Reihenfolge $\alpha = 0, 1, \ldots, n$ geschrieben, nehmen also erst zu, dann wieder ab. Ist $(n+1)p$ nicht ganz, so ist das $w_{\alpha\beta}$ mit $\alpha = [(n+1)p]$ das grösste; dann ist $\alpha < (n+1)p < \alpha+1, np-q < \alpha < np+p,\ p-\frac{q}{n} < \frac{\alpha}{n} < p+\frac{p}{n}$, so dass mit $n \to \infty$ die wahrscheinlichste Frequenz $\frac{\alpha}{n}$ nach $p$ strebt. Ist $(n+1)p$ ganz, so treten zwei benachbarte Maxima auf: für $\alpha = (n+1)p = np+p$ und $\alpha - 1 = np - q$; auch hier $\frac{\alpha}{n} \to p$.

**I**. Wichtiger ist folgendes: für beliebige $\varepsilon > 0$ konvergiert die Wahrscheinlichkeit $w_n(\varepsilon)$, dass $|\frac{\alpha}{n} - p| \geq \varepsilon$ sei, mit $n \to \infty$ nach 0, also die Wahrscheinlichkeit, dass $|\frac{\alpha}{n} - p| < \varepsilon$ sei, nach 1. Das ist die elementarste Form des „Gesetzes der grossen Zahlen“, wonach bei einer grossen Zahl von Versuchen die Frequenz annähernd gleich der Wahrscheinlichkeit zu erwarten ist. Es ist sogar $\sum_n \ w_n(\varepsilon)$ konvergent. | Bl. 15

Zum Beweise setzen wir in der Identität

$$\sum \ w_{\alpha\beta}\, x^\alpha y^\beta = (px+qy)^n$$

$$x = e^{qt}, \quad y = e^{-pt} \quad x^\alpha y^\beta = e^{[(1-p)\alpha - p\beta]t} = e^{(\alpha-np)t},$$

$$\sum \ w_{\alpha\beta}\, e^{(\alpha-np)t} = (pe^{qt} + qe^{-pt})^n = \left(1 + \frac{pq^2+qp^2}{2}t^2 + \cdots\right)^n$$

$$\text{oder } = \left(1 + \frac{pq}{2}t^2 + \cdots\right)^n = 1 + n\frac{pq}{2}t^2 + \cdots$$

also bei Entwicklung der linken Seite nach Potenzen von $t$ ausser
$\sum \ w_{\alpha\beta} = 1, \quad \sum \ w_{\alpha\beta}\,(\alpha - np) = 0:$

$$\begin{aligned} \sum \ w_{\alpha\beta}\,(\alpha-np)^2 &= npq \\ \sum \ w_{\alpha\beta}\,\left(\frac{\alpha}{n} - p\right)^2 &= \frac{pq}{n}. \end{aligned}$$

Die linke Seite ist, wenn wir $\overset{*}{\sum}$ nur über die $\alpha$ mit $|\frac{\alpha}{n} - p| \geq \varepsilon$ erstrecken, $\geq \overset{*}{\sum} \ w_{\alpha\beta}\ (\frac{\alpha}{n} - p)^2 \geq \varepsilon^2 \cdot \overset{*}{\sum} \ w_{\alpha\beta} = \varepsilon^2\, w(|\frac{\alpha}{n} - p| \geq \varepsilon)$. Also ist

$$w\left(\left|\frac{\alpha}{n} - p\right| \geq \varepsilon\right) \leq \frac{pq}{n\varepsilon^2}$$

und strebt (bei festem $\varepsilon$) mit $n \to \infty$ nach 0.

Nennen wir diese Wahrscheinlichkeit kurz $w_n(\varepsilon)$. Wegen einer späteren Anwendung wollen wir zeigen, dass sogar *die Reihe* $\sum_n \ w_n(\varepsilon)$ *konvergiert*, wozu die obige Abschätzung nicht ausreicht. Wir nehmen dann den Koeffizienten von $\frac{t^4}{4!}$ in der Identität, der links $\sum \ w_{\alpha\beta}(\alpha - np)^4$ ist; rechts erhält man | Bl. 16

$$\begin{aligned} pe^{qt} + qe^{-pt} &= 1 + \frac{pq}{2}\,t^2 + \frac{pq^3 - qp^3}{6}\,t^3 + \frac{pq^4 + qp^4}{24}\,t^4 + \cdots \\ (pe^{qt} + qe^{-pt})^n &= 1 + n\left[\frac{pq}{2}\,t^2 + \frac{pq^3 - qp^3}{6}\,t^3 + \frac{pq^4 + qp^4}{24}\,t^4 + \cdots\right] \\ &+ \frac{n(n-1)}{2}\ \left[\frac{pq}{2}\,t^2 + \cdots\right]^2 + \cdots, \end{aligned}$$

also als den fraglichen Koeffizienten

$$\begin{aligned}\sum w_{\alpha\beta}\,(\alpha - np)^4 &= n\cdot(pq^4+qp^4)+3n(n-1)p^2q^2\\ &= 3n^2p^2q^2+npq(q^3+p^3-3pq),\\ \sum w_{\alpha\beta}\left(\frac{\alpha}{n}-p\right)^4 &= 3\frac{p^2q^2}{n^2}+\frac{pq(1-6pq)}{n^3}\end{aligned}$$

und $w_n(\varepsilon)$ ergiebt sich in derselben Art wie oben $\leq$ diesem Ausdruck dividirt durch $\varepsilon^4$. Da $\sum \frac{1}{n^2}$, $\sum \frac{1}{n^3}$ konvergieren, ist die Behauptung bewiesen.

Wenn beim $k$. Versuch die Wahrscheinlichkeiten von $A_k$ und $B_k$ gleich $p_k, q_k$ sind $(p_k+q_k=1)$, immer aber unabhängig vom Ausfall der früheren Versuche, so ist die Wahrscheinlichkeit $w_{\alpha\beta}$, dass bei $n=\alpha+\beta$ Versuchen $\alpha$ mal der günstige Fall $A, \beta$ mal $B$ eintrete, der Koeffizient von $x^\alpha y^\beta$ in der Entwicklung des Produkts

$$(p_1x+q_1y)(p_2x+q_2y)\cdots(p_nx+q_ny)=\sum w_{\alpha\beta}\,x^\alpha y^\beta.$$

Bl. 17 | Hier nähert sich, wie wir später sehen werden, der Unterschied zwischen der relativen Häufigkeit $\dfrac{\alpha}{n}$ und dem arithmetischen Mittel $\dfrac{p_1+\cdots+p_n}{n}$ der Wahrscheinlichkeiten mit wachsendem $n$ der Null.

Wenn bei jedem Versuch eine dreifache Disjunktion $A_k+B_k+C_k=G_k$ mit den Wahrscheinlichkeiten $p, q, r$ betrachtet wird, so ist die Wahrscheinlichkeit, dass bei $n$ Versuchen $\alpha$ mal $A$, $\beta$ mal $B, \gamma$ mal $C$ $(\alpha+\beta+\gamma=n)$ in irgend einer Reihenfolge auftrete,

$$w_{\alpha\beta\gamma}=\frac{n!}{\alpha!\,\beta!\,\gamma!}\,p^\alpha q^\beta r^\gamma;$$

es ist

$$(px+qy+rz)^n=\sum w_{\alpha\beta\gamma}\,x^\alpha y^\beta z^\gamma.$$

Nehmen wir jetzt einen Fall, wo die Versuche von einander *abhängig* sind: in einer Urne sind $s=a+b$ Kugeln, $a$ weisse, $b$ schwarze; es werden $\sigma$ (an Stelle der vorigen $n$) Züge getan, ohne die gezogenen Kugeln zurückzulegen; $A_k, B_k$ das Ziehen von Weiss, Schwarz beim $k$. Zuge. Wir haben z. B. bei drei Zügen die Disjunktion

$$A_1A_2A_3,\ A_1A_2B_3,\ A_1B_2A_3,\ A_1B_2B_3,\ B_1A_2A_3,\ B_1A_2B_3,\ B_1B_2A_3,\ B_1B_2B_3;$$

Bl. 18 | die zugehörigen Wahrscheinlichkeiten haben immer den Nenner $s(s-1)(s-2)$ und die Zähler

$$\begin{gathered}a(a-1)(a-2),\ a(a-1)b,\ ab(a-1),\ ab(b-1),\\ ba(a-1),\ ba(b-1),\ b(b-1)a,\ b(b-1)(b-2);\end{gathered}$$

die Wahrscheinlichkeit für zwei mal Weiss, einmal Schwarz ist, in jeder der drei Reihenfolgen, $\dfrac{a(a-1)b}{s(s-1)(s-2)}$, bei Zusammenfassung $3\dfrac{a(a-1)b}{s(s-1)(s-2)}$.

Allgemein ist bei $\sigma = \alpha + \beta$ Zügen die Wahrscheinlichkeit für $\alpha$ weisse, $\beta$ schwarze Kugeln in bestimmter Reihenfolge immer

$$\frac{a(a-1)\cdots(a-\alpha+1)\,b(b-1)\cdots(b-\beta+1)}{s(s-1)\cdots(s-\sigma+1)} = \frac{a!\,b!}{(a-\alpha)!\,(b-\beta)!}\,\frac{(s-\sigma)!}{s!}$$

und in beliebiger Reihenfolge

$$w_{\alpha\beta} = \frac{\sigma!}{\alpha!\,\beta!}\,\frac{a!\,b!}{(a-\alpha)!\,(b-\beta)!}\,\frac{(s-\sigma)!}{s!} = \binom{a}{\alpha}\binom{b}{\beta} : \binom{s}{\sigma}$$

Die Relation $\sum w_{\alpha\beta} = 1$ giebt

$$\sum_{\alpha+\beta=\sigma} \binom{a}{\alpha}\binom{b}{\beta} = \binom{s}{\sigma}$$

oder

$$\sum_{\alpha=0}^{\sigma} \binom{a}{\alpha}\binom{s-a}{\sigma-\alpha} = \binom{s}{\sigma},$$

die man so verifizieren kann: wenn $\alpha, \beta$ unabhängig variieren, ist

$$\sum_{\alpha,\beta} \binom{a}{\alpha}\binom{b}{\beta} x^{\alpha+\beta} = (1+x)^{a+b} = \sum_{\sigma} \binom{s}{\sigma} x^{\sigma}.$$

Enthält die Urne $s = a + b + c$ Kugeln, $a$ weisse, $b$ schwarze, $c$ rote, so ist die Wahrscheinlichkeit, dass bei $\sigma = \alpha + \beta + \gamma$ Zügen ohne Zurücklegen der | Bl. 18v gezogenen Kugeln $\alpha$ weisse, $\beta$ schwarze, $\gamma$ rote gezogen werden:

$$w_{\alpha\beta\gamma} = \binom{a}{\alpha}\binom{b}{\beta}\binom{c}{\gamma} : \binom{s}{\sigma},$$

woraus man

$$\sum_{\alpha,\beta} \binom{a}{\alpha}\binom{b}{\beta}\binom{s-a-b}{\sigma-\alpha-\beta} = \binom{s}{\sigma}$$

erhält. | Bl. 19

*Die Bayessche Formel.* $C_1 + C_2 + \cdots + C_n$ sei eine vollständige Disjunktion; $C_i$ habe die Wahrscheinlichkeit $\gamma_i = w(C_i)$, und das Ereignis $A$ habe im Fall $C_i$ die Wahrscheinlichkeit $\alpha_i = w_{C_i}(A)$. Dann ist

$$w(AC_i) = \gamma_i \alpha_i \qquad (1)$$

$$w(A) = \sum_1^n \gamma_i \alpha_i \tag{2}$$

$$\text{also} \qquad w_A(C_i) = \gamma_i \alpha_i : \sum_1^n \gamma_i \alpha_i. \tag{3}$$

Man sagt, die $C_i$ seien *Ursachen*, die dem $A$ die Wahrscheinlichkeit $\alpha_i$ erteilen, oder *Hypothesen* (bei paarweise verschiedenen $\alpha_i$ kann man direkt sagen: $C_i$ ist die Hypothese, dass die Wahrscheinlichkeit von $A$ gleich $\alpha_i$ sei); die Hypothese $C_i$, die vor dem Eintritt von $A$ die Wahrscheinlichkeit $\gamma_i$ (Wahrscheinlichkeit a priori) hatte, hat dann nach Eintritt von $A$ die Wahrscheinlichkeit (3) (Wahrscheinlichkeit a posteriori).

Z. B. $A$ Ziehen eines Königs aus einem Kartenspiel; $C_1 + C_2$ sei die Disjunktion, dass das ein Piquet- oder Whistspiel sei; beide Annahmen gleich wahrscheinlich ($\gamma_1 = \gamma_2 = \frac{1}{2}$). $\alpha_1 = \frac{4}{32} = \frac{1}{8}$, $\alpha_2 = \frac{4}{52} = \frac{1}{13}$; die Wahrscheinlichkeit $w_A(C_1)$, dass es ein Piquetspiel sei, ist nach Ziehen eines Königs Bl. 20 $\frac{1}{8} : (\frac{1}{8} + \frac{1}{13}) = \frac{13}{21}$. – Ist $A$ das Ziehen einer | Drei (die es im Piquetspiel nicht giebt), so ist $\alpha_1 = 0$, $\alpha_2 = \frac{1}{13}$, $w_A(C_1) = 0$, $w_A(C_2) = 1$, es *muss* dann ein Whistspiel gewesen sein.

Ist $B$ noch ein Ereignis, das im Falle $AC_i$ die Wahrscheinlichkeit $\beta_i = w_{AC_i}(B)$ hat, so ist

$$w(AC_iB) = \gamma_i \alpha_i \beta_i \tag{4}$$

$$w(AB) = \sum_1^n \gamma_i \alpha_i \beta_i \tag{5}$$

$$w_A(B) = \sum_1^n \gamma_i \alpha_i \beta_i : \sum_1^n \gamma_i \alpha_i. \tag{6}$$

Wenn im Falle $C_i$ die Wahrscheinlichkeit von $B$ nicht vom Eintreten oder Ausbleiben von $A$ abhängt, ist $\beta_i = w_{C_i}(B)$ $(= w_{AC_i}(B) = w_{\bar{A}C_i}(B))$ zu setzen.

Sind a priori alle Hypothesen gleich wahrscheinlich ($\gamma_i = \frac{1}{n}$), so ist nach (3) $w_A(C_i) = \alpha_i : \sum\limits_1^n \alpha_i$ der Wahrscheinlichkeit $w_{C_i}(A) = \alpha_i$ proportional.

Beispiel. In einer Urne mit $s$ Kugeln befinden sich mit der Wahrscheinlichkeit $\gamma_a$ $a$ weisse und $b = s - a$ schwarze. ($C_a$ heisst: $a$ weisse Kugeln sind vorhanden. $a = 0, 1, \ldots . s$). $A$ sei das Ziehen einer weissen Kugel; also mit der Bl. 21 Wahrscheinlichkeit $w_{C_a}(A) = \alpha_a = \frac{a}{s}$. | Ferner sei $B$ das abermalige Ziehen von Weiss mit einem 2. Zug, nachdem beim 1. die gezogene Kugel nicht zurückgelegt wird. Also $\beta_a = w_{AC_a}(B) = \frac{a-1}{s-1}$. Wir haben also

$$\begin{aligned} w(A) &= \frac{1}{s}\sum_{0}^{s}\gamma_a a, \quad w_A(C_a) = \gamma_a a : \sum_{0}^{s}\gamma_a a \\ w(AB) &= \frac{1}{s(s-1)}\sum_{0}^{s}\gamma_a a(a-1), \quad w_A(B) = \frac{1}{s-1}\,\frac{\sum\limits_{0}^{s}\gamma_a a(a-1)}{\sum\limits_{0}^{s}\gamma_a a}. \end{aligned}$$

Machen wir verschiedene Annahmen:
(A) alle $\gamma_a = \frac{1}{s+1}$, jede Kugelverteilung a priori gleich wahrscheinlich.

$$\begin{aligned} w(A) &= \frac{1}{s(s+1)}\sum_{0}^{s} a = \frac{1}{2}, \quad w_A(C_a) = \frac{2a}{s(s+1)} \\ w(AB) &= \frac{1}{(s-1)s(s+1)}\sum_{0}^{s} a(a-1) = \frac{1}{3}, \quad w_A(B) = \frac{2}{3}. \end{aligned}$$

Nach Ziehen einer weissen Kugel ist die Wahrscheinlichkeit, dass $a$ Kugeln vorhanden waren, mit $a$ proportional; die wahrscheinlichste Annahme $a = s$ hat die Wahrscheinlichkeit $\frac{2}{s+1}$; die Wahrscheinlichkeit, dass mindestens soviele weisse wie schwarze Kugeln in der Urne waren, ist $> \frac{3}{4}$, nämlich

für $s = 2r : \frac{2}{s(s+1)} \sum\limits_{r}^{2r} a = \frac{1}{2r(2r+1)}\ (2r(2r+1) - r(r-1)) = \frac{3r+3}{4r+2} = \frac{3}{4} + \frac{3}{4(2r+1)}$,

für $s = 2r-1 : \frac{2}{s(s+1)} \sum\limits_{r}^{2r-1} a = \frac{1}{(2r-1)2r}\ [(2r-1)2r - (r-1)r] = \frac{3r-1}{4r-2} = \frac{3}{4} + \frac{1}{4(2r-1)}$.

(B) Man wisse, dass die Urne mindestens so viel weisse wie schwarze Kugeln enthält (die verbleibenden Fälle gleich wahrscheinlich). Dann ist für $s = 2r \quad \gamma_r = \gamma_{r+1} = \cdots = \gamma_{2r} = \frac{1}{r+1}$, die übrigen $\gamma_a = 0$. Bl. 22

$$w(A) = \frac{1}{(r+1)s}\cdot\sum_{r}^{2r} a = \frac{3r^2+3r}{4(r+1)r} = \frac{3}{4};$$

für $s = 2r-1 \quad \gamma_r = \cdots = \gamma_{2r-1} = \frac{1}{r}$, die übrigen $= 0$.

$$w(A) = \frac{1}{rs}\sum_{r}^{2r-1} a = \frac{3r^2-r}{2r(2r-1)} = \frac{3r-1}{4r-2} = \frac{3}{4} + \frac{1}{4s};$$

die Wahrscheinlichkeit, Weiss zu ziehen, $\geq \frac{3}{4}$.
(C) Die „kleine“ Urne sei aus einer „grossen“ Urne mit $S = A + B$ Kugeln, darunter $A$ weissen, durch Entnahme von $s$ Kugeln gefüllt worden.

$\gamma_a = \binom{A}{a}\binom{B}{b} : \binom{S}{s}$,

$$w(A) = \frac{1}{s} \sum_a \binom{A}{a}\binom{B}{b}\, a : \binom{S}{s} = \frac{A}{s} \sum \binom{A-1}{a-1}\binom{B}{b} : \binom{S-1}{s-1} = \frac{A}{S}$$

(d. h., wie vorauszusehen, = der Wahrscheinlichkeit, direkt aus der grossen Urne eine weisse Kugel zu ziehen).

Ferner die Wahrscheinlichkeit, dass $a$ weisse Kugeln in der kleinen Urne waren,

$$w_A(C_a) = \gamma_a \frac{a}{s} : \frac{A}{S} = \binom{A-1}{a-1}\binom{B}{b} : \binom{S-1}{s-1},$$

d. h. (nachdem eine weisse Kugel aus der grossen und kleinen Urne weggenommen ist) = der Wahrscheinlichkeit, dass aus einer grossen Urne mit $S-1 = (A-1)+B$ Kugeln ($A-1$ weisse) in die kleine mit $s-1=(a-1)+b$ Kugeln
Bl. 23 $a-1$ weisse gekommen sind. |

Ferner für das Ziehen von zwei weissen Kugeln

$$\begin{aligned} w(AB) &= \sum \frac{a(a-1)}{s(s-1)} \binom{A}{a}\binom{B}{b} : \binom{S}{s} \\ &= \sum \frac{A(A-1)}{S(S-1)} \binom{A-2}{a-2}\binom{B}{b} : \binom{S-2}{s-2} = \frac{A(A-1)}{S(S-1)} \end{aligned}$$

$w_A(B) = \frac{A-1}{S-1}$ wie beim direkten Ziehen aus der grossen Urne.

Wir machen unabhängige Versuche, wo bei jedem der günstige Fall eine und dieselbe Wahrscheinlichkeit $p$ hat, die gewisse Werte $p_i$ mit den Wahrscheinlichkeiten $\gamma_i$ annehmen kann ($C_i$ die Annahme, dass $p = p_i$). $A$ sei das Ereignis: bei $m$ Versuchen ist $l$ mal der günstige Fall eingetreten. Das hat bei der Wahrscheinlichkeit $p$ die Wahrscheinlichkeit

$$\alpha(p) = \binom{m}{l} p^l (1-p)^{m-l}$$

und dann ist $w(A) = \sum \gamma_i\, \alpha(p_i)$, $\quad w_A(C_i) = \gamma_i\, \alpha(p_i) : w(A)$. Bei gleichen $\gamma_i$ ist $w_A(C_i)$ mit $\alpha(p_i)$ oder $p_i^l(1-p_i)^{m-l}$ proportional. Die wahrscheinlichste Annahme ist $p = \frac{l}{m}$, da die Funktion $p^l(1-p)^{m-l}$ dort ein Maximum hat; ihre
Bl. 24 Ableitung ist $p^{l-1}(1-p)^{m-l-1}\,[l(1-p)-(m-l)p]$, $[\ \ ] = l - mp$. |

Machen wir für $p_i$ die $n$ Annahmen $p_i = \frac{i}{n}$ $(i = 1, 2, \ldots, n)$ mit gleicher Wahrscheinlichkeit, so konvergiert

$$w(A) = \frac{1}{n} \sum \alpha(p_i)$$

für $n \to \infty$ nach

$$w^*(A) = \int\limits_0^1 \alpha(p)\, dp = \binom{m}{l} \int\limits_0^1 p^l (1-p)^{m-l}\, dp.$$

Wir haben das Eulersche Integral

$$\int_0^1 p^l (1-p)^k \, dp = \frac{l!\,k!}{(l+k+1)!};$$

nämlich

$$\sum_{l=0}^{m} \binom{m}{l} p^l (1-p)^{m-l} x^l = (px+1-p)^m$$

nach $p$ integriert:

$$\begin{aligned}
\sum_{l=0}^{m} \binom{m}{l} \int_0^1 p^l (1-p)^{m-l} dp \cdot x^l &= \int_0^1 [p(x-1)+1]^m dp \\
&= \frac{1}{x-1}\,\frac{1}{m+1} \left[[p(x-1)+1]^{m+1}\right]_{p=0}^{p=1} \\
&= \frac{1}{x-1}\,\frac{1}{m+1}(x^{m+1}-1) = \frac{1}{m+1} \sum_{0}^{m} x^l \\
\int_0^1 p^l (1-p)^{m-l} dp &= 1:(m+1)\binom{m}{l} = \frac{l!\,(m-l)!}{(m+1)!}
\end{aligned}$$

Also ist $w^*(A) = \frac{1}{m+1}$ (bei sozusagen gänzlich unbestimmter Wahrscheinlichkeit $p$ ist jeder Versuchsausfall gleich wahrscheinlich). | Bl. 25

$B$ sei das Ereignis, dass bei weiteren $\mu$ Versuchen der günstige Fall $\lambda$ mal eintrete. Unter der Annahme $p$ ist

$$\beta(p) = \binom{\mu}{\lambda} p^\lambda (1-p)^{\mu-\lambda}$$

die Wahrscheinlichkeit dafür, $\alpha(p)\beta(p)$ die von $AB$, also

$$w(AB) = \sum \gamma_i \, \alpha(p_i)\beta(p_i);$$

das konvergiert wie oben nach

$$\begin{aligned}
w^*(AB) &= \int_0^1 \alpha(p)\beta(p)\, dp \\
&= \binom{m}{l}\binom{\mu}{\lambda} \frac{(l+\lambda)!\,(m+\mu-l-\lambda)!}{(m+\mu+1)!} \\
&= \binom{m}{l}\binom{\mu}{\lambda} : \left[(m+\mu+1)\binom{m+\mu}{l+\lambda}\right]
\end{aligned}$$

und $w_A(B)$ konvergiert nach

$$w_A^*(B) = \frac{m+1}{m+\mu+1} \cdot \frac{\binom{m}{l}\binom{\mu}{\lambda}}{\binom{m+\mu}{l+\lambda}}$$

Z. B. für $l = m, \lambda = \mu$ ist $\frac{m+1}{m+\mu+1}$ die (Grenz)-Wahrscheinlichkeit, dass ein bei $m$
Versuchen immer eingetretener Fall auch bei $\mu$ weiteren immer eintrete. Oder
Bl. 26 für $\mu = \lambda = 1$ ist | $\frac{l+1}{m+2}$ die Wahrscheinlichkeit, dass ein bei $m$ Versuchen $l$ mal
eingetretenes Ereignis beim nächsten Versuch wieder eintrete.

Bl. 27
## | § 2 Momente elementarer Vertheilungen.

$A_1, A_2, \ldots, A_m$ sei eine vollständige Disjunction, $w(A_i) = p_i$, $\sum p_i = 1$. Denken wir uns jedem Fall $A_i$ eine reelle Zahl $x_i$ zugeordnet (z. B. die Nummer $i$). Wenn die $p_i > 0$ und die $x_i$ paarweise verschieden sind, und $x$ eins von ihnen bedeutet, so nennen wir $x$ eine *Variable*, die die Werthe $x_1, \ldots, x_m$ annehmen kann; nur ist, gegenüber dem sonstigen Sprachgebrauch die Variable noch näher präcisirt, indem sie jeden Werth $x_i$ mit einer gewissen Wahrscheinlichkeit $p_i$ annimmt. Den Inbegriff der $p_i, x_i$ nennen wir eine Vertheilung der Variablen $x$.

Solche Vertheilungen spielen in den Anwendungen der Wahrscheinlichkeitsrechnung die Hauptrolle. Bei *Zufallsspielen* werden auf die verschiedenen möglichen Fälle $A_i$ Gewinne oder Preise $x_i$ gesetzt. Jede wissenschaftliche *Beobachtung* (Messung) ist wegen der damit verbundenen Fehlerquellen als ein Versuch anzusehen, bei dem nicht mit Gewissheit der wahre Werth der beobachteten Grösse, sondern mit irgend welchen Wahrscheinlichkeiten mehr oder minder abweichende Werthe $x_i$ gefunden werden.

Sei eine Vertheilung $(p_i, x_i)$ gegeben (wir können dabei die Voraussetzungen
Bl. 28 $p_i > 0$ und $x_i \neq x_k$ entbehren), so bilden wir für eine | Function $f(x)$ den
Ausdruck

$$Mf(x) = \sum_{1}^{m} p_i \, f(x_i) \tag{1}$$

Er heisst die *mathematische Hoffnung* oder *Erwartung* von $f(x)$, oder der *wahrscheinliche Werth* von $f(x)$, oder (Stieltjes) das *Moment* von $f(x)$; diesen Ausdruck verwenden wir in der Regel.

Ist $f$ der kleinste, $F$ der grösste unter den Werthen $f(x_i)$, so ist wegen $\sum p_i = 1$, $\quad f \leq Mf(x) \leq F$; $Mf(x)$ ist ein „Mittelwerth“ unter den von $f(x)$ angenommenen Werthen.

Die Momente der *Potenzen*

$$\mu_k = Mx^k = \sum_{i=1}^{m} p_i \, x_i^k \quad (k = 0, 1, \ldots) \tag{2}$$

werden kurz die *Momente* der Vertheilung schlechthin genannt, wobei $\mu_0 = \sum p_i = 1$. Das Moment eines Polynoms

$$\begin{aligned} f(x) &= \alpha_0 + \alpha_1 x + \cdots + \alpha_k x^k \\ \text{ist } \ Mf(x) &= \alpha_0\mu_0 + \alpha_1\mu_1 + \cdots + \alpha_k\mu_k. \end{aligned}$$

Der Ausdruck $M(x-\alpha)^2 = \mu_2 - 2\alpha\mu_1 + \alpha^2 = \mu_2 - \mu_1^2 + (\mu_1 - \alpha)^2$ giebt ein Mass
für die Abweichungen der $x_i$ vom Werthe $\alpha$; er ist dann 0 (und bei $p_i > 0$ nur
dann), wenn alle $x_i$ gleich $\alpha$ sind. Seinen kleinsten Werth hat er für $\alpha = \mu_1$: Bl. 29

$$M(x-\mu_1)^2 = \mu_2 - \mu_1^2; \tag{3}$$

man nennt $\sqrt{\mu_2 - \mu_1^2}$ die *Streuung* (Dispersion) der Variablen $x$, bei Beobachtungen auch den *mittleren Fehler*, bei Zufallsspielen das *mittlere Risico.* Ähnliche, für die Rechnung minder bequeme Abweichungsmasse sind

$$M(x-\mu_1)^4,\ M(x-\mu_1)^6, \ldots \quad M|x-\mu_1|,\ M|x-\mu_1|^3, \ldots,$$

während in $M(x-\mu_1) = 0,\ M(x-\mu_1)^3, \ldots$ auch die Vorzeichen der Abweichungen eine Rolle spielen.

Ist $t$ eine positive Zahl und wird die Summe $\overset{*}{\sum}$ nur über diejenigen $i$ erstreckt, wo $|x_i| \geq t$, so ist für $k = 2, 4, \ldots$

$$\mu_k = \sum p_i\, x_i^k \geq \overset{*}{\sum} p_i\, x_i^k \geq t^k \cdot \overset{*}{\sum} p_i,$$

$$\overset{*}{\sum} p_i \leq \frac{\mu_k}{t^k}$$

oder mit ersichtlicher Bezeichnung

$$w(|x| \geq t) \leq \frac{\mu_k}{t^k}, \quad w(|x| < t) \geq 1 - \frac{\mu_k}{t^k} \text{ (Tschebyscheff)} \tag{4}$$

Ersetzt man (den Fall, dass nur ein Werth mit Gewissheit angenommen wird, ausgeschlossen, also eine eigentliche Vertheilung mit $\mu_k > 0$ vorausgesetzt), $t$ durch $t\mu_k^{\frac{1}{k}}$, so ist

$$w(|x| \geq t\,\mu_k^{\frac{1}{k}}) \leq \frac{1}{t^k}, \qquad w(|x| < t\,\mu_k^{\frac{1}{k}}) \geq 1 - \frac{1}{t^k}. \tag{5}$$

Wird $\mu_1 = 0$ angenommen, so dass $\mu_2^{\frac{1}{2}}$ die Streuung ist, so ist z. B. die Wahr- Bl. 30
scheinlichkeit, dass die Variable (oder allgemein: die Abweichung von ihrem
wahrscheinlichen Werth) unter der 10-fachen Streuung bleibe, $\geq 1 - \frac{1}{100} = 0,99$.

Mit einer unbestimmten Zahl $u$ bilden wir

$$Me^{ux} \quad = \quad M(1 + xu + x^2\frac{u^2}{2!} + \cdots) = 1 + \mu_1 u + \mu_2 \frac{u^2}{2!} + \cdots$$

Für hinlänglich kleines $u$ können wir $\log(Me^{ux})$ in eine $\mathfrak{P}(u)$ entwickeln:

$$\log(Me^{ux}) = (\mu_1 u + \mu_2 \frac{u^2}{2} + \cdots) - \frac{1}{2}(\mu_1 u + \mu_2 \frac{u^2}{2} + \cdots)^2 + \frac{1}{3}(\quad)^3 - \cdots$$

$$= \lambda_1 u + \lambda_2 \frac{u^2}{2!} + \lambda_3 \frac{u^3}{3!} + \cdots \tag{6}$$

Diese Coefficienten $\lambda_k$ sollen die *logarithmischen Momente* heissen. Man findet:

$$\left.\begin{array}{lcl} \lambda_1 & = & \mu_1 \\ \lambda_2 & = & \mu_2 - \mu_1^2 \\ \lambda_3 & = & \mu_3 - 3\mu_1\mu_2 + 2\mu_1^3 \\ \lambda_4 & = & \mu_4 - 4\mu_1\mu_3 - 3\mu_2^2 + 12\mu_1^2\mu_2 - 6\mu_1^4 \\ & & \cdots \end{array}\right\} \tag{7}$$

und durch Auflösung oder aus $e^{\lambda_1 u + \lambda_2 \frac{u^2}{2} + \cdots} = 1 + \mu_1 u + \mu_2 \frac{u^2}{2} + \cdots$

$$\left.\begin{array}{lcl} \mu_1 & = & \lambda_1 \\ \mu_2 & = & \lambda_2 + \lambda_1^2 \\ \mu_3 & = & \lambda_3 + 3\lambda_1\lambda_2 + \lambda_1^3 \\ \mu_4 & = & \lambda_4 + 4\lambda_1\lambda_3 + 3\lambda_2^2 + 6\lambda_1^2\lambda_2 + \lambda_1^4 \\ & & \cdots \end{array}\right\} \tag{8}$$

Bl. 31 | Die $\mu_k, \lambda_k$ bestimmen einander also eindeutig. $\sqrt{\lambda_2}$ ist die Streuung.

Die $\mu_k$ und $\lambda_k$ sind *homogen*, d. h. für $x' = ax$ $(a \neq 0)$ ist $\mu'_k = a^k \mu_k$, $\lambda'_k = a^k \lambda_k$. Gegenüber einer Verschiebung des Nullpunktes, $x' = x - \alpha$, sind die $\lambda_k$ bis auf $\lambda_1$ invariant: $\lambda'_1 = \lambda_1 - \alpha$, $\lambda'_k = \lambda_k$ $(k = 2, 3, \ldots)$; denn es ist $Me^{ux'} = e^{-u\alpha} Me^{ux}$, $\log(Me^{ux'}) = -u\alpha + \log(Me^{ux})$.

*Beispiel.* Die logarithmischen Momente einer Vertheilung zu berechnen, wo $x$ mit den Wahrscheinlichkeiten $p$ und $1 - p$ die Werthe 1 und 0 annimmt. Das giebt

$$\sum_1^\infty \lambda_k \frac{u^k}{k!} = \log(pe^u + 1 - p); \tag{9}$$

die $\lambda_k$ sind Functionen von $p$. Durch Differentiation nach $p$ und $u$ folgt:

$$\begin{array}{rcl} \displaystyle\sum_1^\infty \frac{d\lambda_k}{dp} \frac{u^k}{k!} & = & \displaystyle\frac{e^u - 1}{pe^u + 1 - p}, \\ \displaystyle\sum_1^\infty \lambda_k \frac{u^{k-1}}{(k-1)!} & = & \displaystyle\frac{pe^u}{pe^u + 1 - p} \quad \text{oder} \\ \displaystyle\sum_0^\infty \lambda_{k+1} \frac{u^k}{k!} & = & \displaystyle p + p(1-p)\frac{e^u - 1}{pe^u + 1 - p} \\ & = & \displaystyle p + p(1-p) \sum_1^\infty \frac{d\lambda_k}{dp} \frac{u^k}{k!}, \end{array}$$

also $\lambda_1 = p$ und für $k \geq 1 \qquad \lambda_{k+1} = p(1-p)\frac{d\lambda_k}{dp}$, also

$$\left.\begin{array}{lcl} \lambda_2 & = & p(1-p) = p - p^2 \\ \lambda_3 & = & p(1-p)(1-2p) = p - 3p^2 + 2p^3, \\ \lambda_4 & = & p(1-p)(1-6p+6p^2) \\ & & \dots \end{array}\right\} \tag{10}$$

Das sind auch die logarithmischen Momente von $x - p$, bis auf das erste, das dann $= 0$ wird.

*Variablenpaare.* Man kann den Gliedern einer vollständigen Disjunction auch Variablenpaare oder Punkte der Ebene zuordnen; ein variabler Punkt durchläuft also eine endliche Punktmenge und nimmt darin jede Lage mit einer gewissen Wahrscheinlichkeit an. Sind $x_i$ die zugehörigen Abscissen, $y_k$ die zugehörigen Ordinaten, so ist es bequem, die Punktmenge aus *allen* Punkten $(x_i, y_k)$ bestehend anzunehmen; $w_{ik}$ sei die zugehörige Wahrscheinlichkeit (Für die Punkte, die in der ursprünglichen Menge nicht vorkamen, ist $w_{ik} = 0$ zu setzen). Die Ereignisse

$$A_i\ [x = x_i] \qquad B_k\ [y = y_k] \qquad A_iB_k\ [x = x_i, y = y_k]$$

haben die Wahrscheinlichkeiten $\quad p_i = \sum_k w_{ik}, \quad q_k = \sum_i w_{ik}, \quad w_{ik},$

wobei Bl. 33

$$\sum_i p_i = \sum_k q_k = \sum_{i,k} w_{ik} = 1.$$

Das Moment einer Function $f(x,y)$ ist durch

$$Mf(x,y) = \sum_{i,k} w_{ik}\ f(x_i, y_k)$$

zu definiren; die Momente der Potenzproducte

$$\mu_{m,n} = \sum_{i,k} w_{ik}\ x_i^m y_k^n$$

sind die Momente der Vertheilung. Insbesondere sind

$$\begin{array}{lcl} \mu_{m,0} & = & \sum\limits_{i,k} w_{ik}\ x_i^m = \sum\limits_i p_i\ x_i^m \\ \mu_{0,n} & = & \sum\limits_{i,k} w_{ik}\ y_k^n = \sum\limits_k q_k\ y_k^n \end{array}$$

die Momente der einzeln für sich betrachteten Variablen $x, y$.

Im Allgemeinen sind $A_i, B_k$ von einander abhängig. Ist jedes $A_i$ von jedem $B_k$ unabhängig,

$$w_{ik} = p_i\, q_k,$$

so nennen wir die *Variablen* $x, y$ (oder ihre Vertheilungen) *unabhängig*. Die Wahrscheinlichkeit $w_{A_i}(B_k)$, die im Allgemeinen $= \frac{w_{ik}}{p_i}$ und von $i$ abhängig ist, wird hier $= q_k$; die Wahrscheinlichkeit für $y = y_k$ hängt nicht von dem Werth $x_i$ ab, den $x$ annimmt (und ebenso die für $x = x_i$ nicht von $y_k$). Für Bl. 34 eine Function $f(x)g(y)$ wird dann das Moment |

$$\begin{aligned} Mf(x)g(y) &= \sum_{i,k} p_i q_k\, f(x_i) g(y_k) = \sum_i p_i\, f(x_i) \cdot \sum_k q_k\, g(y_k) \\ &= Mf(x) \cdot Mg(y) \end{aligned}$$

gleich dem Product der Momente der Einzelvertheilungen, insbesondere

$$\begin{aligned} \mu_{m,n} &= \mu_{m,0} \cdot \mu_{0,n} \\ \text{Ferner} \qquad Me^{u(x+y)} &= Me^{ux} \cdot Me^{uy} \\ \log Me^{u(x+y)} &= \log Me^{ux} + \log Me^{uy} \end{aligned}$$

und durch Entwicklung nach Potenzen von $u$ :

*Sind $x, y$ unabhängige Variable, so ist das $k^{te}$ logarithmische Moment von $x + y$ die Summe der logarithmischen Momente von $x$ und von $y$*:

$$\lambda_k(x + y) = \lambda_k(x) + \lambda_k(y) \tag{11}$$

Für $k = 1$ gilt dies auch bei abhängigen Variablen:

$$M(x + y) = \sum_{i,k} w_{ik}\,(x_i + y_k) = \sum_i p_i\, x_i + \sum_k q_k\, y_k = Mx + My.$$

Für $k = 2$ folgt: *sind $x, y$ unabhängig, so ist das Streuungsquadrat von $x + y$ gleich der Summe derer von $x$ und von $y$.*

Sind $x, y$ unabhängig und $a, b$ Constante, so ist

$$\lambda_k(ax + by) = a^k \lambda_k(x) + b^k \lambda_k(y).$$

Bl. 35 | Die Betrachtung dehnt sich auf drei und mehr Variable aus. $x$ möge die Werthe $x_i$, $y$ die Werte $y_k$, $z$ die Werthe $z_l$ durchlaufen; diese Ereignisse mögen $A_i, B_k, C_l$ heissen und $p_i, q_k, r_l$ ihre Wahrscheinlichkeiten. Der Punkt $(x, y, z)$ durchläuft die Punkte $(x_i, y_k, z_l)$ und dies Ereignis $A_i B_k C_l$ habe die Wahrscheinlichkeit $w_{ikl}$. Mit diesen $w_{ikl}$ sind auch

$$p_i = \sum_{k,l} w_{ikl}, \quad q_k = \sum_{i,l} w_{ikl}, \quad r_l = \sum_{i,k} w_{ikl}$$

sowie die Wahrscheinlichkeiten

$$w(B_k C_l) = \sum_i w_{ikl}, \quad w(A_i C_l) = \sum_k w_{ikl}, \quad w(A_i B_k) = \sum_l w_{ikl}$$

bekannt. Wenn stets

$$w_{ikl} = p_i\, q_k\, r_l,$$

also jedes $C_l$ von jedem $A_iB_k$ (und damit auch von jedem $A_i$ und jedem $B_k$) unabhängig ist, ebenso jedes $B_k$ von $A_iC_l$ und $A_i$ von $B_kC_l$, kurzum, wenn die Wahrscheinlichkeit für einen Werth einer Variablen nicht von den Werthen der beiden andern abhängt, so heissen $x, y, z$ unabhängig. In diesem Falle gilt

$$\lambda_k(x + y + z) = \lambda_k(x) + \lambda_k(y) + \lambda_k(z)$$

(für $k = 1$ auch ohne die Voraussetzung der Unabhängigkeit) und ebenso für beliebig viele unabhängige Variable. | Bl. 36

Dieses additive Verhalten der $\lambda_k$, insbesondere der Streuungsquadrate, ist von grosser Bedeutung: hierin liegt das Gesetz der grossen Zahlen, die Ausgleichung des Zufalls bei Massenerscheinungen. Werden $n$ gleichgute (derselben Vertheilung unterliegende) unabhängige Beobachtungen mit dem mittleren Fehler $\sqrt{\lambda_2}$ gemacht, so ist der mittlere Fehler für die Summe der beobachteten Variablen $\sqrt{n\lambda_2}$, für ihr arithmetisches Mittel $\sqrt{\frac{\lambda_2}{n}}$ ; dieses ist $\sqrt{n}$ mal genauer als die einzelne Beobachtung. Machen $n$ Spieler unabhängig von einander dasselbe Spiel mit dem Risico $\sqrt{\lambda_2}$, und treten sie dann zu einer Gesamtheit zusammen, die Gewinn und Verlust gleich vertheilt, so ist das Risico der Gesamtheit $\sqrt{n\lambda_2}$, das des einzelnen Spielers $\sqrt{\frac{\lambda_2}{n}}$.

Genauere Aussagen erhalten wir durch Verbindung mit der Tschebyscheffschen Abschätzung (4) oder (5). Nennen wir kurz für eine Variable $x$ die Abweichung $x - \mu_1$ des $x$ von seinem wahrscheinlichen Werth $\mu_1 = M\,x$ die *Abweichung* dieser Variablen.

Seien nun $\xi_1, \xi_2, \ldots$ von einander unabhängige Variable, $\alpha_n = M\xi_n$, $x_n = \xi_n - \alpha_n$ die Abweichung der Variablen $\xi_n$; ferner | Bl. 37

$$\left.\begin{array}{lcl} \eta_n & = & \xi_1 + \xi_2 + \cdots + \xi_n \\ \beta_n & = & \alpha_1 + \alpha_2 + \cdots + \alpha_n \\ y_n & = & \eta_n - \beta_n = x_1 + x_2 + \cdots + x_n \end{array}\right\} \tag{12}$$

also $\eta_n$ die Summe der ersten Variablen und $y_n$ ihre Abweichung. Nach (11) ist

$$\lambda_k(y_n) = \lambda_k(x_1) + \lambda_k(x_2) + \cdots + \lambda_k(x_n) = n\lambda_{k,n} \tag{13}$$

wo $\lambda_{k,n}$ also der Durchschnitt (das arithmetische Mittel) aus den $k^{\text{ten}}$ logarithmischen Momenten der ersten $n$ Variablen ist. (Für $k \geq 2$ ist $\lambda_k(x_n) = \lambda_k(\xi_n)$, $\lambda_k(y_n) = \lambda_k(\eta_n)$; dagegen $\lambda_1(x_n) = 0$, $\lambda_1(\xi_n) = \alpha_n$ ; $\lambda_1(y_n) = 0$, $\lambda_1(\eta_n) = \beta_n$). Haben z. B. alle Variablen dieselbe Vertheilung mit den logarithmischen Momenten $\lambda_k$, so ist $\lambda_{k,n} = \lambda_k$. Die Formeln (8) geben, da $\lambda_1(y_n) = 0$,

$$\left.\begin{array}{lcl} \mu_2(y_n) & = & \lambda_2(y_n) = n\lambda_{2,n} \\ \mu_4(y_n) & = & \lambda_4(y_n) + 3\lambda_2(y_n)^2 = n\lambda_{4,n} + 3n^2\lambda_{2,n}^2 \\ & & \cdots \end{array}\right\} \tag{14}$$

Die Formel (5) giebt, für $k = 2$ auf $y_n$ angewandt

$$w\left(|y_n| \geq t\sqrt{n\lambda_{2,n}}\right) \leq \frac{1}{t^2}$$

oder, wenn man $t = \sqrt{n}\,\varepsilon$ setzt ($\varepsilon > 0$)

$$w\left(\frac{|y_n|}{n} \geq \varepsilon\sqrt{\lambda_{2,n}}\right) \leq \frac{1}{n\varepsilon^2} \tag{15}$$

Bl. 37v | Ist $\sqrt{\lambda_{2,n}}$ beschränkt $\leq \mathcal{L}$, so ist

$$w\left(\frac{|y_n|}{n} \geq \varepsilon\mathcal{L}\right) \leq \frac{1}{n\varepsilon^2}$$

oder

$$w\left(\frac{|y_n|}{n} < \delta\right) \geq 1 - \frac{\mathcal{L}^2}{n\delta^2};$$

die Wahrscheinlichkeit, dass $\frac{y_n}{n}$ zwischen beliebig engen Grenzen $\pm\delta$ liege, con-
Bl. 38 vergirt für $n \to \infty$ nach 1. |
Also bei festem $\varepsilon$:

**I.** *Die Wahrscheinlichkeit, dass das Mittel*

$$\frac{y_n}{n} = \frac{1}{n}\left(\xi_1 - \alpha_1 + \cdots + \xi_n - \alpha_n\right)$$

*aus den Abweichungen der ersten* $n$ *Variablen absolut* $\geq \varepsilon\sqrt{\lambda_{2,n}}$ *sei, ist* $\leq \dfrac{1}{n\varepsilon^2}$ *und convergirt also mit* $n \to \infty$ *nach* 0.

Das ist die erste Form des *Gesetzes der grossen Zahlen.*

Wenn insbesondere alle Variablen dieselbe Streuung $\sqrt{\lambda_2}$ haben, so ist die Wahrscheinlichkeit, dass jenes Abweichungsmittel absolut $\geq \varepsilon\sqrt{\lambda_2}$ sei, $\leq \dfrac{1}{n\varepsilon^2}$ und convergirt mit $n \to \infty$ nach 0.

Beispiel. Beim $n^{\text{ten}}$ Versuch sei die vollständige Disjunction $A_n, B_n$ mit den Wahrscheinlichkeiten $p_n,\ q_n$ ($p_n + q_n = 1$) zu entscheiden; $A_n$ werde der *günstige* Fall des $n^{\text{ten}}$ Versuches genannt. Die Variable $\xi_n$ sei $= 1, 0$, jenachdem $A_n, B_n$ eintritt. Dann ist $\eta_n$ die Anzahl derjenigen unter den $n$ ersten Versuchen, bei denen $A_n$ eintritt, kurz die *Häufigkeit* der günstigen Fälle, $\dfrac{\eta_n}{n}$ die *relative Häufigkeit.* Ferner ist

$$\begin{aligned} M\xi_n = p_n \cdot 1 + q_n \cdot 0 &= p_n = M\xi_n^2, \\ \alpha_n = p_n, \quad \lambda_2(x_n) &= M\xi_n^2 - (M\xi_n)^2 = p_n(1 - p_n) = p_n q_n \end{aligned}$$

Bl. 39 (vgl. die erste Formel (10)). |

Also $\beta_n = p_1 + \cdots + p_n$ und $\lambda_{2,n} = \frac{p_1 q_1 + \cdots + p_n q_n}{n}$. So ergiebt sich:

**II.** *(Satz von Poisson). Es werde eine Folge unabhängiger Versuche gemacht; beim* $n^{ten}$ *Versuch sei* $p_n$ *die Wahrscheinlichkeit des günstigen Falls. Die Wahrscheinlichkeit, dass bei den ersten* $n$ *Versuchen die relative Häufigkeit* $\frac{\eta_n}{n}$ *des günstigen Falls vom arithmetischen Mittel* $\frac{\beta_n}{n} = \frac{p_1 + \cdots + p_n}{n}$ *der Wahrscheinlichkeiten um mindestens den Betrag* $\varepsilon\sqrt{\lambda_{2,n}}$ *abweiche* ($\lambda_{2,n} = \frac{p_1 q_1 + \cdots + p_n q_n}{n}$), *ist* $\leq \frac{1}{n\varepsilon^2}$ *und convergirt also für* $n \to \infty$ *nach* $0$.

Insbesondere, wenn alle $p_n = p$ :

**III.** *(Satz von Jakob Bernoulli). Es werde eine Folge unabhängiger Versuche gemacht; bei jedem Versuch sei* $p$ *die Wahrscheinlichkeit des günstigen Falls. Die Wahrscheinlichkeit, dass bei den ersten* $n$ *Versuchen die relative Häufigkeit des günstigen Falls von* $p$ *um mindestens den Betrag* $\varepsilon\sqrt{pq}$ *abweiche, ist* $\leq \frac{1}{n\varepsilon^2}$ *und convergirt also für* $n \to \infty$ *nach* $0$.

Hierbei ist $0 < p < 1$ vorausgesetzt (in (5) war ja $\mu_k > 0$ angenommen). Wir können mit $\delta = \varepsilon\sqrt{pq}$, indem wir das complementäre Ereigniss betrachten, auch sagen: | Bl. 40

**III*** *Die Wahrscheinlichkeit, dass bei den ersten* $n$ *Versuchen die relative Häufigkeit des günstigen Falles zwischen den (beliebig engen) Grenzen* $p \pm \delta$ *liege, ist* $\geq 1 - \frac{pq}{n\delta^2}$ *und convergirt also für* $n \to \infty$ *nach* $1$.

Diese Sätze sind wegen ihres elementaren Charakters werthvoll; allerdings lässt sich die Tschebyscheffsche Abschätzung verschärfen. Auf Grund von **III** kann man a priori angenommene Wahrscheinlichkeiten durch Versuchsreihen bestätigen (resp. aus Versuchsreihen Wahrscheinlichkeiten a posteriori bestimmen, Bayessche Regel). Z. B. fand Buffon unter 4040 Würfen mit einer Münze 2048 mal Wappen; Albert Fleck erzielte unter 12000 Würfen mit einem Würfel

| | | | | | | |
|---|---|---|---|---|---|---|
| | 2029 | 1988 | 1985 | 1964 | 2003 | 2031 |
| mal die Zahl | | | | | | |
| | Eins | Zwei | Drei | Vier | Fünf | Sechs |

Der Bernoullische Satz lässt sich dahin verschärfen:

**IV**. *Es werde eine Folge unabhängiger Versuche gemacht, bei jedem Versuch sei* $p$ *die Wahrscheinlichkeit des günstigen Falles. Die Wahrscheinlichkeit, dass die relative Häufigkeit des günstigen Falls bei den ersten* $n$ *Versuchen mit* $n \to \infty$ *nach* $p$ *convergirt, ist gleich* $1$. | Bl. 41

Aber dieser Satz ist von anderem Character als **III** und bezieht sich auf unendliche Ereignisfolgen (z. B. kommt die Wahrscheinlichkeit in Frage, dass bei *allen* Versuchen vom $n^{\text{ten}}$ ab die relative Häufigkeit zwischen $p \pm \delta$ liege), so dass wir ihn erst in § 4 beweisen können. Auch muss dazu (15) verschärft werden, so dass eine höhere Potenz von $n$ im Nenner steht. Wenden wir (5) für $k = 4$ auf $y_n$ an, wobei die zweite Formel (14) zu beachten ist, so kommt

$$w\left(|y_n| \geq t\sqrt[4]{3n^2\lambda_{2,n}^2 + n\lambda_{4,n}}\right) \leq \frac{1}{t^4}$$

oder wieder mit $t = \sqrt{n}\,\varepsilon$

$$w\left(\frac{|y_n|}{n} \geq \varepsilon\sqrt[4]{3\lambda_{2,n}^2 + \frac{1}{n}\lambda_{4,n}}\right) \leq \frac{1}{n^2\varepsilon^4} \tag{16}$$

Ist z. B. die $\sqrt[4]{\phantom{x}}$ für alle $n$ beschränkt,

$$3\lambda_{2,n}^2 + \frac{1}{n}\lambda_{4,n} \leq \mathcal{L}^4,$$

so ist erst recht

$$w\left(\frac{|y_n|}{n} \geq \varepsilon\mathcal{L}\right) \leq \frac{1}{n^2\varepsilon^4} \tag{17}$$

und dann wird sich in der That ergeben: die Wahrscheinlichkeit, dass $\frac{y_n}{n} \to 0$, ist gleich 1. Das gilt z. B., wenn alle Variablen dieselbe Vertheilung haben:

$$3\lambda_2^2 + \frac{1}{n}\lambda_4 \leq 3\lambda_2^2 + |\lambda_4|.$$

Bl. 42 | *Logarithmische Momente eines Variablenpaares.* Kehren wir zu einem Variablenpaar $(x, y)$ zurück, wo $x, y$ nicht mehr unabhängig zu sein brauchen. Durch die Potenzreihe ($u, v$ beliebige, hinlänglich kleine reelle Zahlen)

$$\log Me^{ux+vy} = \sum_{m,n=0}^{\infty} \frac{u^m v^n}{m!\,n!}\,\lambda_{m,n} \qquad (\lambda_{0,0} = 0)$$

definiren wir die logarithmischen Momente $\lambda_{m,n} = \lambda_{m,n}(x, y)$ des Variablenpaares. Ersetzt man $u, v$ durch $tu, tv$ und betrachtet $u, v$ als fest, so ist andererseits

$$\log Me^{t(ux+vy)} = \sum_{p=1}^{\infty} \frac{t^p}{p!}\,\lambda_p(ux + vy)$$

und das giebt:

$$\lambda_p(ux + vy) = \sum_{m=0}^{p} \binom{p}{m} u^m v^{p-m} \lambda_{m,p-m}(x, y).$$

Z. B. mit den früheren Bezeichnungen $\mu_{mn} = M(x^m y^n)$

$$\begin{aligned} \mu_1(ux+vy) &= \mu_{10}u + \mu_{01}v \\ \mu_2(ux+vy) &= \mu_{20}u^2 + 2\mu_{11}uv + \mu_{02}v^2 \\ \mu_3(ux+vy) &= \mu_{30}u^3 + 3\mu_{21}u^2v + 3\mu_{12}uv^2 + \mu_{03}v^3 \\ & \dots \end{aligned}$$

und nach (7)

$$\begin{aligned} \lambda_1(ux+vy) &= \mu_{10}u + \mu_{01}v \\ \lambda_2(ux+vy) &= (\mu_{20} - \mu_{10}^2)u^2 + 2(\mu_{11} - \mu_{10}\mu_{01})uv + (\mu_{02} - \mu_{01}^2)v^2 \\ & \dots \end{aligned}$$

also

$$\begin{aligned} \lambda_{10} &= \mu_{10}, \quad \lambda_{01} = \mu_{01}, \\ \lambda_{20} = \mu_{20} - \mu_{10}^2, \quad \lambda_{11} &= \mu_{11} - \mu_{10}\mu_{01}, \quad \lambda_{02} = \mu_{02} - \mu_{01}^2 \\ & \dots \end{aligned}$$

Allgemein ist $\lambda_{m,0} = \lambda_m(x)$, $\lambda_{0,n} = \lambda_n(y)$. Bei unabhängigen $x, y$ ist nach Bl. 43 einer früheren Formel (nach (11) ) $\lambda_n(ux+vy) = u^n\lambda_n(x) + v^n\lambda_n(y)$, also alle „gemischten“ logarithmischen Momente 0. Bei abhängigen ist im Allgemeinen schon

$$\lambda_{11} = \mu_{11} - \mu_{10}\mu_{01} = M(xy) - M(x)M(y) = \sum_{i,k} (w_{ik} - p_i\, q_k)\, x_i y_k$$

von 0 verschieden; man nennt $\dfrac{\lambda_{11}}{\sqrt{\lambda_{20}\lambda_{02}}}$ den *Correlationscoefficienten* von $x, y$, da er einen Massstab für die Abhängigkeit giebt (ohne dass etwa sein Verschwinden bereits die Unabhängigkeit garantirte). Er ist absolut $\leq 1$, da $\lambda_2(ux+vy) = \lambda_{20}u^2 + 2\lambda_{11}uv + \lambda_{02}v^2 \geq 0$, also $\lambda_{20}\lambda_{02} - \lambda_{11}^2 \geq 0$ ist. Er ist absolut $< 1$, da (uneigentliche Vertheilung mit nur einem Punkt ausgeschlossen) die quadratische Form $\lambda_2(ux+vy) = \lambda_{20}u^2 + 2\lambda_{11}uv + \lambda_{02}v^2$ definit positiv, also $\lambda_{20}\lambda_{02} - \lambda_{11}^2$ positiv ist.

Indem man von $x, y$ durch eine Coordinatentransformation (Bewegung) zu andern Coordinaten übergeht, kann man $\lambda_{10}, \lambda_{01}, \lambda_{11}$ zum Verschwinden bringen. Bl. 44

Für zwei von einander unabhängige Variablenpaare verhalten sich die logarithmischen Momente additiv:

$$\lambda_{m,n}(x+\xi, y+\eta) = \lambda_{m,n}(x,y) + \lambda_{m,n}(\xi,\eta).$$

## |§ 3. Zufallsspiele. Etwas über Versicherungsrechnung.

$A_1, A_2, \ldots$ mögen eine vollständige Disjunction bilden mit den Wahrscheinlichkeiten $p_1, p_2, \ldots$; wenn $A_i$ eintritt, empfängt der Spieler $X$ vom Spielgegner $Y$ den *Gewinn* $g_i$ und zahlt an ihn den *Einsatz* $e_i$ (der in der Regel $= e$, von $i$ unabhängig ist), hat also den *Reingewinn* $x_i = g_i - e_i$ (der $\gtreqless 0$ sein kann, für $x_i < 0$ also einen Reinverlust). Wir haben eine Vertheilung $(p_i, x_i)$ (die $x_i$ nicht nothwendig verschieden) mit den Momenten $Mf(x) = \sum p_i\, f(x_i)$, insbesondere $\mu_k = \sum\, p_i\, x_i^k$. $\mu_1$ ist der wahrscheinliche Reingewinn, $\mu_2 - \mu_1^2 = M(x-\mu_1)^2 = \lambda_2$ das Quadrat der Streuung oder des *Risicos*.

Das Spiel heisst *gerecht*, wenn $\mu_1 = 0$, für den Spieler $X$ günstig (ungünstig), wenn $\mu_1 > 0$ ($< 0$). Das hat folgenden Grund. Handelt es sich z. B. um eine Lotterie, wo $a = a_1 + a_2 + \cdots$ Loose mit den Gewinnen $g_1, g_2, \ldots$ gezogen werden, so ist für den Spieler eines Looses $p_i = \frac{a_i}{a}$, $x_i = g_i - e$, wenn $e$ den Preis eines Looses bedeutet. Werden sämtliche Loose verkauft, so hat der Lotterieunternehmer $ae$ erhalten und $\sum a_i g_i$ auszuzahlen; wenn er weder Nutzen noch Schaden haben will, ist $\sum a_i g_i = ae$, $\sum p_i g_i = e$, $\sum p_i (g_i - e) = \sum p_i x_i = 0$.

Ferner: wird das Spiel in unabhängigen Wiederholungen gespielt und ist $y_n$ Bl. 46 die Reingewinnsumme der ersten $n$ Spiele, $\frac{y_n}{n} - \mu_1 = \varepsilon_n$, | so besagt das Gesetz der grossen Zahlen (§ 2): $w(|\varepsilon_n| < \varepsilon) \to 1$ (oder nach der späteren Verschärfung: $w(\varepsilon_n \to 0) = 1$). Für $\mu_1 > 0$ macht also der Spieler voraussichtlich einen unbegrenzt wachsenden Reingewinn $y_n = n(\mu_1 + \varepsilon_n)$, für $\mu_1 < 0$ einen unbegrenzt wachsenden Verlust; für $\mu_1 = 0$ wird sich Gewinn und Verlust im Mittel ausgleichen, $\frac{y_n}{n}$ zwischen beliebig engen Grenzen liegen.

Ein ungerechtes Spiel kann man nach der Formel $Mx = \mu_1 + M(x - \mu_1)$ so auffassen: der Spieler macht ein gerechtes Spiel mit dem Reingewinn $x_i - \mu_1$ und hat ausserdem die Summe $\mu_1$ im Voraus zu empfangen (resp. zu zahlen, wenn $\mu_1 < 0$).

Macht ein Spieler zwei Spiele zugleich, gleichviel ob abhängig oder unabhängig, so ist $\mu_1(x+y) = \mu_1(x) + \mu_1(y)$. Bei unabhängigen Spielen addiren sich die Risicoquadrate: $\lambda_2(x+y) = \lambda_2(x) + \lambda_2(y)$.

Die Wahrscheinlichkeitsrechnung soll häufig Auskunft geben, ob ein Spiel rathsam ist. Sie kann nur feststellen, ob das Spiel günstig ist oder nicht ($\mu_1 \gtreqless 0$) und wie gross das Risico ist. Das Übrige ist Sache des Spielers. Ein Spiel mit hohen, selbst wenig wahrscheinlichen Verlusten ist auch bei geringem Gesamt-Risico unvernünftig. Es kommt auch der Zweck des Spieles (Versicherung!) in Bl. 47 Betracht, der selbst ungünstige Spiele rathsam machen kann. |

*Das Petersburger Paradoxon.* Eine Münze wird $n$ mal geworfen; zeigt sich Bild zuerst beim $i^{\text{ten}}$ Versuche, so erhält der Spieler den Preis $g_i = 2^i$, zeigt sie immer Schrift, so erhält er nichts. Bei gerechtem Spiel ist der Einsatz

$$e = \sum_1^n \frac{g_i}{2^i} + \frac{0}{2^n} = n.$$

Denkt man sich das Spiel unbegrenzt fortgesetzt (§ 4), so müsste $e$ unendlich sein. Dies Spiel ist natürlich trotz der Gerechtigkeit äusserst unvernünftig: hohe Verluste mit grosser, noch höhere Gewinne mit kleiner Wahrscheinlichkeit.

Beim *Roulettespiel* den Vortheil der Bank zu bestimmen. Von den 36 Nummern $1, \ldots, 36$ besetzt der Spieler $a (= 1, 2, 3, 4, 6, 9, 12, 18)$ mit dem Gesamteinsatz $e$ und erhält, wenn die Kugel in eine seiner Nummern fällt, den Gewinn $g = e \cdot \frac{36}{a}$. Die Gewinnwahrscheinlichkeit ist aber nur $p = \frac{a}{37}$, da noch eine Nummer 0 (Zéro) vorhanden ist. Also $\mu_1 = pg - e = e \cdot \frac{36}{37} - e = -\frac{e}{37}$, der Gewinn der Spielbank ist $\frac{1}{37}$ der Einsätze. Bisweilen wird das für den Spieler günstigere Verfahren geübt: wenn 0 fällt und beim nächsten Spiel eine der Nummern des Spielers kommt, erhält er den Einsatz zurück. Dann ist $\mu_1 = \frac{a}{37} g + \frac{1}{37} \cdot \frac{a}{37} e - e = -\frac{e}{37}(1 - \frac{a}{37})$, also z. B. für $a = 18$ nur etwa die Hälfte des vorigen Verlustes. Bl. 48

Wenn ein Spieler, immer auf 18 Nummern setzend, solange „doublirt" (d. h. $e, 2e, 4e, \ldots$ setzt), bis er einmal gewinnt, und wenn dies zuerst beim $n^{\text{ten}}$ Spiel der Fall ist, so hat er $2^n e$ gewonnen und $e(1 + 2 + \cdots + 2^{n-1}) = e(2^n - 1)$ verloren, also einen Reingewinn $= e$. Die Wahrscheinlichkeit, innerhalb der ersten $n$ Spiele wenigstens einmal zu gewinnen, ist $1 - (\frac{19}{37})^n$ und würde sich durch Vergrösserung von $n$ beliebig nahe an 1 bringen lassen, wenn $n$ nicht durch die Höhe seines Vermögens und durch den von der Bank festgesetzten Maximaleinsatz begrenzt wäre. Er spielt also ein unvernünftiges Spiel: ein hoher Verlust $e(2^n - 1)$ mit geringer Wahrscheinlichkeit, dem ein geringer Gewinn $e$ mit hoher Wahrscheinlichkeit gegenübersteht.

*Lotto* (genuesische Zahlenlotterie, im 17. Jahrhundert aus Wahlwetten entstanden). Von $s$ Nummern werden $\sigma$ gezogen. Die Wahrscheinlichkeit, dass eine im Voraus bestimmte Nummer gezogen wird, ist $p_1 = \sigma : s$. Der Spieler bestimmt im Voraus eine Nummer und setzt darauf $e$; bei gerechtem Spiel müsste er, wenn seine Nummer gezogen wird, den Gewinn $g = \frac{e}{p_1}$ erhalten. Besetzt er $l$ Nummern mit dem Gesamteinsatz $e$ (d. h. jede mit dem Einsatz $\frac{e}{l}$) und werden $\lambda$ seiner Nummern gezogen, so muss er für jede dieser den Gewinn $\frac{g}{l}$, also insgesamt den Gewinn $\frac{\lambda}{l} g$ erhalten. Bl. 49

Oder man spielt auf Nummernpaare (Amben). Es sind $\binom{s}{2}$ Amben vorhanden, $\binom{\sigma}{2}$ werden gezogen; die Wahrscheinlichkeit, dass eine im Voraus bestimmte Ambe gezogen wird, ist $p_2 = \binom{\sigma}{2} : \binom{s}{2}$. Ist der Einsatz für diese Ambe $e$, so muss der Gewinn $g = \frac{e}{p_2}$ sein. Besetzt der Spieler $l$ Nummern mit dem Gesamteinsatz $e$, so muss er den Gewinn $\frac{\binom{\lambda}{2}}{\binom{l}{2}} g$ erhalten, falls $\lambda$ seiner Nummern gezogen werden; denn er hat $\binom{l}{2}$ Amben, jede mit dem Einsatz $\frac{e}{\binom{l}{2}}$ besetzt, muss also für jede seiner gezogenen $\binom{\lambda}{2}$ Amben den Gewinn $\frac{g}{\binom{l}{2}}$ erhalten.

Ebenso kann man Terne u.s.w., allgemein für $m \leq \sigma$ Combinationen $C_m$ von $m$ Nummern spielen. $p_m = \binom{\sigma}{m} : \binom{s}{m}$, $g = \frac{e}{p_m}$, und wenn $\lambda$ von den $l$

Nummern des Spielers gezogen werden, muss er den Gewinn $\frac{\binom{\lambda}{m}}{\binom{l}{m}}\,g$ erhalten. Für $s = 90, \sigma = 5$ ist $p_1 = \frac{5}{90} = \frac{1}{18}$, $p_2 = \frac{5\cdot 4}{90\cdot 89} = \frac{1}{400{,}5}$ usw. Die Lottobanken geben natürlich nicht die Gewinne $18e$, $400,5 \cdot e$, usw., sondern viel kleinere.

Übrigens könnte diese Darstellung Zweifel lassen, ob z. B. $l$ Nummern soviel wie $\binom{l}{2}$ Amben bedeuten. Die Wahrscheinlichkeit, dass unter $l$ Nummern $\lambda$
Bl. 50 gezogen werden, ist $\binom{l}{\lambda}\binom{s-l}{\sigma-\lambda} : \binom{s}{\sigma}$, gleich der Wahrscheinlichkeit, aus einer |
Urne mit $s$ Kugeln, darunter $\sigma$ weissen, bei Ziehung von $l$ Kugeln $\lambda$ weisse zu finden. Hierauf soll nun der Gewinn $\frac{\binom{\lambda}{m}}{\binom{l}{m}}\,g$ gesetzt sein. Dann ist bei gerechtem Spiel der Einsatz

$$e = \frac{g}{\binom{l}{m}\binom{s}{\sigma}} \sum_{\lambda} \binom{l}{\lambda}\binom{s-l}{\sigma-\lambda}\binom{\lambda}{m} = \frac{g}{\binom{s}{\sigma}} \sum_{\lambda} \binom{l-m}{\lambda-m}\binom{s-l}{\sigma-\lambda}$$

$$= g \cdot \frac{\binom{s-m}{\sigma-m}}{\binom{s}{\sigma}} = g\frac{\binom{\sigma}{m}}{\binom{s}{m}} = gp_m \quad \text{wie oben.}$$

*Lotterie.* Unter $s = s_1 + s_2 + \cdots + s_n$ Loose werden $s_1$ Gewinne $g_1, \ldots,$ $s_n$ Gewinne $g_n$ verloost. Der Einsatz für ein Loos beträgt $g = \sum \frac{s_i}{s}\, g_i$, das Risico $r$ beim Spielen eines Looses ist gegeben durch

$$r^2 = \sum \frac{s_i}{s}(g_i - g)^2 = \frac{1}{s}\sum s_i g_i^2 - g^2.$$

Wie gross ist das Risico beim Spielen von $\sigma$ Loosen? Nicht $\sqrt{\sigma}\, r$, da ja die Loose nicht von einander unabhängig sind, sondern $\sqrt{\frac{\sigma(s-\sigma)}{s-1}}\, r$. Nämlich: Die Wahrscheinlichkeit, dass die $\sigma$ Loose des Spielers $\sigma_1$ mal den Gewinn $g_1, \ldots,$ $\sigma_n$ mal den Gewinn $g_n$ erzielen, ist $W = \binom{s_1}{\sigma_1} \cdots \binom{s_n}{\sigma_n} : \binom{s}{\sigma}$ $(\sigma_1 + \cdots + \sigma_n = \sigma)$. Sein Gewinn beträgt dann $G = \sigma_1 g_1 + \cdots + \sigma_n g_n$ (sein Reingewinn $G - \sigma g$), sein Risicoquadrat ist $R^2 = \sum W(G - \sigma g)^2 = \sum WG^2 - (\sum WG)^2$. Da nun
Bl. 51 (die Summe immer über die $\sigma_1, \ldots, \sigma_n$ mit $\sigma_1 + \cdots + \sigma_n = \sigma$ erstreckt) |

$$\sum \binom{s_1}{\sigma_1} \cdots \binom{s_n}{\sigma_n} = \binom{s}{\sigma},$$

$$\sum \binom{s_1}{\sigma_1} \cdots \binom{s_n}{\sigma_n} \frac{\sigma_1}{s_1} = \sum \binom{s_1 - 1}{\sigma_1 - 1}\binom{s_2}{\sigma_2} \cdots \binom{s_n}{\sigma_n} = \binom{s-1}{\sigma-1},$$

$$\sum \binom{s_1}{\sigma_1} \cdots \binom{s_n}{\sigma_n} \frac{\sigma_1(\sigma_1 - 1)}{s_1(s_1 - 1)} = \binom{s-2}{\sigma-2},$$

$$\sum \binom{s_1}{\sigma_1} \cdots \binom{s_n}{\sigma_n} \frac{\sigma_1}{s_1}\frac{\sigma_2}{s_2} = \binom{s-2}{\sigma-2},$$

so folgt

$$\sum W = 1, \qquad \sum W\sigma_1 = s_1 \frac{\binom{s-1}{\sigma-1}}{\binom{s}{\sigma}} = s_1 \frac{\sigma}{s}$$

$$\sum W\sigma_1(\sigma_1-1) = s_1(s_1-1)\frac{\sigma(\sigma-1}{s(s-1)},\ \sum W\sigma_1\sigma_2 = s_1 s_2\frac{\sigma(\sigma-1)}{s(s-1)}$$

$$\sum W\sigma_1^2 = s_1\frac{\sigma}{s} + s_1(s_1-1)\frac{\sigma(\sigma-1)}{s(s-1)} = s_1^2\frac{\sigma(\sigma-1)}{s(s-1)} + s_1\frac{\sigma(s-\sigma)}{s(s-1)}$$

und das Analoge, wenn man $\sigma_1$ durch $\sigma_i$ oder $\sigma_1\sigma_2$ durch $\sigma_i\sigma_k$ $(i \neq k)$ ersetzt. Demnach $\sum WG = \sum_i g_i \sum W\sigma_i = \sum g_i s_i \frac{\sigma}{s} = \sigma g$, wie schon bekannt.

$$\begin{aligned}\sum WG^2 &= \sum_i g_i^2 \sum W\sigma_i^2 + 2\sum_{i<k} g_i g_k \sum W\sigma_i\sigma_k \\ &= \sum_i g_i^2 \left[s_i^2\frac{\sigma(\sigma-1}{s(s-1)} + s_i\frac{\sigma(s-\sigma)}{s(s-1)}\right] + 2\sum_{i<k} g_i g_k\ s_i s_k\frac{\sigma(\sigma-1)}{s(s-1)} \\ &= \left(\sum s_i g_i\right)^2 \frac{\sigma(\sigma-1)}{s(s-1)} + \sum s_i g_i^2 \cdot \frac{\sigma(s-\sigma)}{s(s-1)}\end{aligned}$$

Mit $\sum s_i g_i = sg, \quad \sum s_i g_i^2 = s(g^2+r^2)$ wird also

$$\sum WG^2 = \frac{s}{s-1}\sigma(\sigma-1)g^2 + (g^2+r^2)\frac{\sigma(s-\sigma)}{s-1} = \sigma^2 g^2 + r^2\frac{\sigma(s-\sigma)}{s-1},$$

$R^2 = r^2 \cdot \frac{\sigma(s-\sigma)}{s-1}$, wie behauptet. Bl. 52

Wollen wir allgemeiner auf die Möglichkeit, halbe Loose und dgl. zu erwerben, Rücksicht nehmen, so behandeln wir die Aufgabe so (indem wir die Gruppen gleicher Gewinne in einzelne auflösen): unter die $s$ Loose werden die Gewinne $g_1, g_2, \ldots, g_s$ verloost; wir setzen wieder $sg = \sum g_i, \quad s(g^2+r^2) = \sum g_i^2$; so dass $g$ der Einsatz für ein Loos, $r$ das Risico für ein Loos ist. Der Spieler kauft von den Loosen $1, 2, \ldots, s$ die Bruchtheile $\vartheta_1, \vartheta_2, \ldots, \vartheta_s$ (d. h. er bezahlt für das $i^{\text{te}}$ Loos $\vartheta_i g$ und bekommt das $\vartheta_i$-fache des auf das Loos fallenden Gewinns ausbezahlt. Z. B. $\vartheta_i = \frac{1}{2}$ oder $0 \leq \vartheta_i \leq 1$, ev. auch $\vartheta_i$ beliebig). Setzen wir $s\vartheta = \sum \vartheta_i, \quad s(\vartheta^2+\rho^2) = \sum \vartheta_i^2$. Die Verloosung führt nun dazu, dass die Gewinne $g_1, \ldots, g_s$ auf die Loose $i_1, i_2, \ldots, i_s$ treffen, wo dies eine Permutation von $1, \ldots, s$ ist; die Wahrscheinlichkeit dafür ist $\frac{1}{s!}$. Der Spieler macht dabei den Gewinn $G = \sum_i g_k\vartheta_{i_k}$, den Reingewinn $G - s\,\vartheta\, g$, sein Risico $R$ ist gegeben durch

$$R = \frac{1}{s!}\sum \left(G - \frac{1}{s!}\sum G\right)^2 = \frac{1}{s!}\sum G^2 - \left(\frac{1}{s!}\sum G\right)^2,$$

wo die Summe $\sum$ über alle Permutationen zu erstrecken ist. Man hat danach

$$\begin{aligned}\sum \vartheta_{i_1} &= (\vartheta_1 + \cdots + \vartheta_s)(s-1)! = s!\,\vartheta, \quad \sum \vartheta_{i_1}^2 = s!(\vartheta^2+\rho^2) \\ \sum \vartheta_{i_1}\vartheta_{i_2} &= 2(\vartheta_1\vartheta_2 + \cdots + \vartheta_{n-1}\vartheta_n)(s-2)! = \\ &= \left[(\sum_i \vartheta_i)^2 - \sum_i \vartheta_i^2\right](s-2)! = [s^2\vartheta^2 - s(\vartheta^2+\rho^2)](s-2)! \\ &= s!\,\vartheta^2 - s!\,\frac{\rho^2}{s-1}\end{aligned}$$

Bl. 53 | Damit folgt

$$\begin{aligned}
\textstyle\sum G &= (g_1 + \cdots + g_s) s!\,\vartheta = s!\,s\,g\,\vartheta \\
\textstyle\sum G^2 &= (g_1^2 + \cdots + g_n^2)\; s!\,(\vartheta^2 + \rho^2) + 2[g_1 g_2 + \cdots + g_{n-1} g_n) \left[s!\,\vartheta^2 - s!\tfrac{\rho^2}{s-1}\right] \\
&= s \cdot s!\,(g^2 + r^2)(\vartheta^2 + \rho^2) + \left[s^2 g^2 - s(g^2 + r^2)\right] \left[s!\,\vartheta^2 - s!\tfrac{\rho^2}{s-1}\right] \\
\tfrac{1}{s!} \textstyle\sum G^2 &= s(g^2 + r^2)(\vartheta^2 + \rho^2) + \left[s(s-1)g^2 - sr^2\right] \left(\vartheta^2 - \tfrac{\rho^2}{s-1}\right) \\
&= s^2 g^2 \vartheta^2 + \tfrac{s^2}{s-1} r^2 \rho^2, \\
R^2 &= \tfrac{s^2}{s-1} r^2 \rho^2, \quad R = \tfrac{s}{\sqrt{s-1}} r\rho.
\end{aligned}$$

$\rho$ ist die zur Loosbesetzung $\vartheta_1, \ldots, \vartheta_n$ gehörige Streuung: $\rho^2 = \frac{1}{s} \sum (\vartheta_i - \vartheta)^2$. Der vorige Fall, Besetzung von $\sigma$ ganzen Loosen, ergiebt sich hier aus $\vartheta_1 = \cdots = \vartheta_\sigma = 1$, $\vartheta_{\sigma+1} = \cdots = \vartheta_s = 0 : s\vartheta = \sigma$, $s(\vartheta^2 + \rho^2) = \sigma$, $\rho^2 = \frac{\sigma(s-\sigma)}{s^2}$, $R^2 = \frac{\sigma(s-\sigma)}{s-1} r^2$ wie oben.
Folgerungen: das Risico ist dasselbe bei complementären Loosbesetzungen ($\vartheta_i$ und $1 - \vartheta_i$), z. B. bei Besetzung von $\sigma$ oder $s - \sigma$ ganzen Loosen. Bei Erwerbung von $2\sigma$ halben Loosen ist das Risico $\sqrt{\dfrac{\sigma(\frac{s}{2} - \sigma)}{s-1}} \cdot r$ kleiner als bei $\sigma$ ganzen, u.
Bl. 54 dgl. |

*Der Ruin der Spieler.* $A$ und $B$ machen eine aus einzelnen Spielen bestehende Partie; bei jedem Spiel gewinnt $A$ von $B$ den Betrag $\alpha$ mit Wahrscheinlichkeit $p$ oder verliert an ihn den Betrag $\beta$ mit Wahrscheinlichkeit $q = 1 - p$. Ein Spieler gilt als besiegt, wenn sein Vermögen $\leq 0$ geworden ist. Wahrscheinlichkeit, dass $A$ siege, wenn die Vermögen zu Beginn der Partie $a, b$ betragen?

Es ist sofort zu vermerken, dass nur drei Möglichkeiten vorliegen: $A$ siegt, $A$ wird besiegt, oder die Partie kommt bei *unbegrenzter* Fortsetzung zu keiner Entscheidung. (Wir überschreiten also damit schon den bisherigen Bereich der elementaren Wahrscheinlichkeiten, wo die Disjunction nur endlich viele Fälle umfasste). Ist $\varphi(x)$ die Wahrscheinlichkeit, dass $A$ siege, wenn sein Vermögen gerade $x$ beträgt, so ist mit Rücksicht auf die Möglichkeiten des nächsten Spieles

$$\varphi(x) = p\,\varphi(x + \alpha) + q\,\varphi(x - \beta)$$

und aus dieser Functionalgleichung ist $x$ für die Argumente $a + m\alpha - n\beta$ ($m, n$ ganze Zahlen $\geq 0$) des Intervalls $[0, a + b]$ zu ermitteln, während für $x \leq 0$, $\varphi(x) = 0$ und für $x \geq a + b$, $\varphi(x) = 1$ zu setzen ist. Die Behandlung dieser Gleichung ist im Allgemeinen ziemlich umständlich.

Bl. 55 Im Falle $\alpha = \beta = 1$, $a, b$ natürliche Zahlen erhält man |

$$\begin{aligned}
\varphi(x) &= p\varphi(x + 1) + q\varphi(x - 1) \\
p[\varphi(x + 1) - \varphi(x)] &= q[\varphi(x) - \varphi(x - 1)]
\end{aligned}$$

Das giebt wegen $\varphi(0) = 0$, wenn $\varphi(1)$ zunächst $= k$ gesetzt wird:

$$\varphi(2) - \varphi(1) = k\,\frac{q}{p}, \quad \varphi(3) - \varphi(2) = k \cdot \left(\frac{q}{p}\right)^2, \quad \ldots$$

$$\varphi(x) - \varphi(x-1) = k \cdot \left(\frac{q}{p}\right)^{x-1} \qquad (x \text{ ganze Zahl})$$

$$\varphi(x) = k\left[1 + \frac{q}{p} + \left(\frac{q}{p}\right)^2 + \cdots + \left(\frac{q}{p}\right)^{x-1}\right] = k \cdot \frac{1 - (\frac{q}{p})^x}{1 - \frac{q}{p}}$$

und wegen $\varphi(a+b) = 1$

$$\varphi(x) = \frac{1 - (\frac{q}{p})^x}{1 - (\frac{q}{p})^{a+b}}\ ,$$

zu Beginn der Partie also

$$\varphi(a) = \frac{1 - (\frac{q}{p})^a}{1 - (\frac{q}{p})^{a+b}} = P$$

(Hierbei wurde $p \neq q$ vorausgesetzt; für $p = q$ erhält man $P = \frac{a}{a+b}$).
Die Wahrscheinlichkeit $Q$, dass $B$ siege, ist (Vertauschung von $p, a$ mit $q, b$)

$$Q = \frac{1 - (\frac{p}{q})^b}{1 - (\frac{p}{q})^{a+b}}$$

(ev. Q $= \frac{b}{a+b}$). Es ist

$$P + Q = \frac{p^b(p^a - q^a)}{p^{a+b} - q^{a+b}} + \frac{q^a(q^b - p^b)}{q^{a+b} - p^{a+b}} = 1,$$

die Wahrscheinlichkeit $R$, dass die Partie unentschieden bleibt, also $= 0$, obwohl dieser Fall nicht unmöglich ist (§ 4).

Was geschieht, wenn $A$ gegen einen Spieler $B$ von ausserordentlich grossem Vermögen spielt? Das ist annäherungsweise der Fall des einzelnen Spielers gegenüber einer Spielbank, aber auch der Fall der Spielbank | gegenüber dem sich Bl. 56
beständig erneuernden Spielerpublicum. Man hat für $b \to \infty$

$$\begin{array}{llll} \text{falls} & p > q & : & P \to 1 - \left(\frac{q}{p}\right)^a, \quad Q \to \left(\frac{q}{p}\right)^a \\ \text{falls} & p \leq q & : & P \to 0, \qquad Q \to 1 \end{array}$$

$A$ hat also, wenn das Spiel für ihn ungünstig oder falls es gerecht ist, praktisch die Gewissheit ruinirt zu werden ($Q \to 1$); ist das Spiel für ihn günstig, so ist die Wahrscheinlichkeit seines Ruins nicht 0, sondern $(\frac{q}{p})^a$, also immerhin klein, wenn sein eigenes Vermögen gross ist. Der Spieler an einer Spielbank, wo er ja ungünstiges Spiel spielt, ruinirt sich also bei unbegrenzter Fortsetzung sicher; andererseits bedarf die Spielbank eines für sie günstigen Spieles, wenn sie nicht selbst mit Sicherheit „gesprengt“ werden will, und diese Möglichkeit besteht selbst dann noch.

Diese Ergebnisse widersprechen nicht dem Gesetz der grossen Zahlen. Die Summe $s_n$ der Reingewinne von $A$ bei den ersten Spielen konnten wir $=$ $n(p-q+\varepsilon_n)$ setzen, wo $w(|\varepsilon_n| < \varepsilon) \to 1$, oder sogar $w(\varepsilon_n \to 0) = 1$. Bei $p \geq q$ wird der durchschnittliche Reingewinn $\frac{s_n}{n}$ voraussichtlich 0 oder positiv sein; trotzdem kann die Wahrscheinlichkeit $Q$, dass *mindestens einmal* $s_n \leq -a$
Bl. 57 noch positiv oder sogar nahe an 1 sein. |

*Versicherungsrechnung.* Eine Andeutung über die Berechnung der Prämien bei der Lebensversicherung sei hier gegeben.

Für eine $x$-jährige Person sei $p_x$ die Wahrscheinlichkeit, nach 1 Jahr noch zu leben, $q_x = 1 - p_x$ die Wahrscheinlichkeit, im Laufe des nächsten Jahres zu sterben. Also $p_x p_{x+1}$ die Wahrscheinlichkeit, nach 2 Jahren noch zu leben, $p_x p_{x+1} \cdots p_{y-1}$ die Wahrscheinlichkeit, das Alter $y$ zu erreichen. Wir gehen von einem Anfangsalter $a$ aus und setzen, unter $l_a$ einen willkürlichen Proportionalitätsfactor verstanden, für $x > a$ ($x$ immer ganzzahlig angenommen) $l_x = l_a\, p_a p_{a+1} \cdots p_{x-1}$, so dass nun $\dfrac{l_y}{l_x}$ $(y \geq x)$ die Wahrscheinlichkeit für einen $x$-jährigen ist, das Alter $y$ zu erreichen. Die Zahlen

$$l_a, \quad l_{a+1}, \quad l_{a+2}, \quad \ldots$$

bilden die „Tafel der Lebenden“; man kann sich eine Urne mit $l_a$ Kugeln vorstellen, von denen im ersten Jahr $l_a - l_{a+1}$ herausgenommen wurden, also nach einem Jahr noch $l_{a+1}$ übrig sind, u.s.f. Diese Wahrscheinlichkeiten resp. die Zahlen $l_x$ sind durch statistische Erhebungen zu ermitteln.

Ferner ist hier, da es sich um Zahlungen zu verschiedenen Zeitpunkten handelt, die *Verzinsung* zu berücksichtigen. Eine heutige Zahlung 1 erlangt durch Verzinsung in einem Jahre den Werth $r$ (z. B. $r = 1,04$ bei 4 % jährlich), in 2 Jahren den Werth $r^2$, in $n$ Jahren den Werth $r^n$. (Es muss Zinseszins, die
Bl. 58 einzige mathematisch unanfechtbare Art, | gerechnet werden). Umgekehrt: eine in $n$ Jahren fällige Zahlung 1 hat heute nur den Werth $\frac{1}{r^n} = \rho^n$ ($\rho = \frac{1}{r}$, $r$ Aufzinsungs-, $\rho$ Abzinsungs- oder Discontirungsfactor). Wir betrachten nun einige einfache Versicherungen.

(A) Der $x$-jährige soll in $n$ Jahren, wenn er dann noch lebt, die Summe 1 erhalten.

Dieser „Gewinn“ hat heute den Werth $\rho^n$, die Wahrscheinlichkeit, ihn zu erlangen, ist $\dfrac{l_{x+n}}{l_x}$, also der dafür zu zahlende „Einsatz“ bei gerechtem Spiel $\dfrac{l_{x+n}}{l_x}\rho^n$. Man bringt das auf die Form $\dfrac{D_{x+n}}{D_x}$, indem man

$$D_x = l_x \rho^x$$

setzt (discontirte Zahlen der Lebenden).

(B) Der $x$-jährige soll (sofort beginnend) eine lebenslängliche Jahresrente 1 erhalten.

Der Einsatz dafür ist die Summe der vorigen für $n = 0, 1, 2, \dots$, d. h.

$$\begin{aligned} R_x &= \frac{1}{D_x}(D_x + D_{x+1} + D_{x+2} + \cdots) = \frac{E_x}{D_x}, \\ \text{wo} \quad E_x &= D_x + D_{x+1} + D_{x+2} + \cdots \end{aligned}$$

gesetzt ist (die Reihe bricht am Ende der Tafel ab).

(C) Der $x$-jährige (d. h. seine Erben) soll am Ende seines Sterbejahres die Summe 1 erhalten. Bl. 59

Der Tod kann im Laufe des $n^{\text{ten}}$ Jahres ($n = 1, 2, \dots$) eintreten; die Zahlung hat dann den heutigen Werth $\rho^n$, die Wahrscheinlichkeit, sie zu erlangen, ist $\frac{l_{x+n-1} - l_{x+n}}{l_x}$, also der heutige Gesamteinsatz

$$\begin{aligned} P_x &= \sum_1^\infty \frac{l_{x+n-1} - l_{x+n}}{l_x}\rho^n = \sum_1^\infty \frac{\rho D_{x+n-1} - D_{x+n}}{D_x} \\ &= \frac{\rho E_x - E_{x+1}}{D_x} = \frac{D_x - (1-\rho)E_x}{D_x} = 1 - (1-\rho)R_x. \end{aligned}$$

Soll diese Versicherung hingegen nicht durch einmalige Zahlung $P_x$, sondern durch eine (sofort beginnende), lebenslänglich zu zahlende Jahresprämie $p_x$ erworben werden,[1] wofür der heutige Einsatz $p_x R_x$ wäre, so ist

$$p_x = \frac{P_x}{R_x} = \frac{1}{R_x} - (1-\rho)$$

Entsprechend sind abgekürzte Versicherungen (die soeben besprochenen waren lebenslänglich), solche mit arithmetisch steigenden oder fallenden Prämien oder mit Rückgewähr bezahlter Jahresprämien (bei denen noch die Zahlen

$$F_x = D_x + 2D_{x+1} + 3D_{x+2} + \cdots = E_x + E_{x+1} + \cdots$$

einzuführen sind), Versicherungen auf verbundene Leben (Ehepaare) u.s.w. zu behandeln. Wir erwähnen nur noch einen Begriff, die *Prämienreserve* = Guthaben des Versicherten bei der Versicherungsgesellschaft. Hat ein $x$-jähriger Bl. 60
die Rentenversicherung ($B$) abgeschlossen und lebt er im Alter $y$ noch, so hat die Gesellschaft jetzt eine Verpflichtung gegen ihn, nämlich $R_y$. Denn soviel müsste ein neu Beitretender zahlen, während unser Versicherter durch seinen früheren Beitritt diesen Betrag guthat.

Ebenso, wenn er die Kapitalversicherung (C) gegen einmalige Zahlung $P_x$ abgeschlossen hat, hat er, wenn er im Alter $y$ noch lebt, dann das Guthaben $P_y$.

---

[1] nicht mit der anfangs erklärten Wahrscheinlichkeit $p_x$ zu verwechseln.

Wenn er jedoch gegen jährliche Prämie $p_x$ abgeschlossen hat, so hat auch er eine Verpflichtung, nämlich, als $y$-jähriger lebenslänglich die Prämie $p_x$ zu zahlen; diese Verpflichtung beläuft sich auf $p_x R_y$ und sein Guthaben ist $P_y - p_x R_y = (p_y - p_x) R_y$. Man kann auch sagen: er müsste, wenn er jetzt beiträte, die höhere Jahresprämie $p_y$ zahlen, hat also die Differenz, d. h. eine lebenslängliche Rente vom Betrage $p_y - p_x$ gut, deren jetziger Einsatz
$(p_y - p_x) R_y$ ist.

Bl. 61

## | § 4. Allgemeine Wahrscheinlichkeiten

Die elementaren Wahrscheinlichkeiten beruhten auf einer vollständigen Disjunction *endlich vieler*, gleichwahrscheinlicher Fälle; sie sind rationale Zahlen. Man kann aber auch ungewisse Aussagen oder Ereignisse in Betracht ziehen, bei denen die möglichen Fälle eine unendliche Menge $M$ bilden, die günstigen eine endliche oder unendliche Menge $G$, wo also eine Vergleichung beider Mengen durch Zählung ausgeschlossen ist.[2] Z. B. fragt man nach der Wahrscheinlichkeit, dass eine dem Intervall $[a, b]$ entnommene Zahl einem Theilintervall $[\alpha, \beta]$ angehöre, oder dass in einer Ziffernfolge $(x_1, x_2, \ldots) \quad (x_n = 0, \ldots, 9)$ nur endlich viele Nullen auftreten.

Die nächstliegende Art, den Wahrscheinlichkeitsbegriff zu erweitern, ist die Betrachtung abzählbarer *Ereignisfolgen* $A_1, A_2, \ldots$, wo also jeder natürlichen Zahl $n$ ein Ereignis $A_n$ zugeordnet ist. Z. B. werde eine entsprechende Folge von Würfelversuchen gemacht, $A_n$ sei das Ereignis, dass beim $n^{\text{ten}}$ Wurf, oder zuerst beim $n^{\text{ten}}$ Wurf, oder spätestens beim $n^{\text{ten}}$ Wurf eine Eins geworfen wird. Wir haben solche Folgen oder eigentlich nur ihre Abschnitte $A_1, \ldots, A_n$
Bl. 62 schon betrachtet (Gesetz der grossen Zahlen). |

Wir definiren auch hier Summe und Product:

$$S = A_1 \cup A_2 \cup \cdots \quad : \quad \text{wenigstens eins der } A_n \text{ tritt ein}$$
$$P = A_1 A_2 \cdots \quad : \quad \text{alle } A_n \text{ treten ein.}$$

Für die Gegentheile gilt wieder: $\bar{S} = \bar{A}_1 \bar{A}_2 \cdots , \quad \bar{P} = \bar{A}_1 \cup \bar{A}_2 \cup \cdots$.
Für Ereignisse, die einander paarweise ausschliessen ($A_n A_m$ unmöglich für $m \neq n$), schreiben wir auch $S = A_1 + A_2 + \cdots$ und sagen, das sie eine *Disjunction* bilden, eine *vollständige Disjunction*, wenn eins von ihnen eintreten muss. Wir erweitern sodann das Axion ($\beta$) zum

*Axiom* ($\gamma$). (Additionspostulat). *Die Wahrscheinlichkeit, dass von abzählbar vielen, einander paarweise ausschliessenden Ereignissen eins eintrete, ist die Summe der Wahrscheinlichkeiten der einzelnen Ereignissse:*

$$w(A_1 + A_2 + \cdots) \; = \; w(A_1) + w(A_2) + \cdots \tag{1}$$

---

[2] Die Cantorsche Mengenlehre bestimmt allerdings die Cardinalzahlen $m, g$ solcher Mengen, aber der Quotient $\frac{g}{m}$ hat keinen Sinn, ausser etwa in dem trivialen Falle $g < m$, wo man $\frac{g}{m} = 0$ setzen könnte. Oder für $\overline{G} = M - G : \quad \overline{g} < m, \quad \frac{g}{m} = 1$.

Darin ist das endliche Postulat ($\beta$) als spezieller Fall enthalten, wenn man nämlich $A_{n+1}, A_{n+2}, \dots$ als unmöglich annimmt.

Die Reihe rechts ist $\lim[w(A_1) + \cdots + w(A_n)] = \lim w(A_1 + \cdots + A_n)$, also gewiss convergent und ihre Summe im Intervall $[0, 1]$ gelegen, wenn bereits $0 \leq w(A_1 + \cdots + A_n) \leq 1$ vorausgesetzt ist. | Bl. 63

Beispiel: es wird eine unbegrenzte Folge von unabhängigen Versuchen gemacht, bei jedem ist $p > 0$ die Wahrscheinlichkeit des günstigen Falls (z. B. Würfe mit einer Münze, günstiger Fall = Werfen von Bild, $p = \frac{1}{2}$). Wie gross ist die Wahrscheinlichkeit, dass der günstige Fall mindestens einmal auftrete? $A_n$ sei das Ereignis, dass der günstige Fall beim $n^{\text{ten}}$ Versuche das erstemal eintrete, $A = A_1 + A_2 + \cdots$;

$$w(A_n) = q^{n-1}p \ (q = 1-p); \quad w(A) = \sum_{1}^{\infty} q^{n-1}p = p(1+q+q^2+\cdots) = \frac{p}{1-q} = 1$$

(weil $0 < q < 1$). Man sieht hier, dass ($\alpha$) nicht umkehrbar ist: *ein Ereignis mit der Wahrscheinlichkeit* 1 *braucht nicht gewiss, eins mit der Wahrscheinlichkeit* 0 *nicht unmöglich zu sein.*

Hier giebt es auch *irrationale* Werthe der Wahrscheinlichkeit. Die Wahrscheinlichkeit, dass der günstige Fall zuerst bei einer quadratischen Versuchsnummer eintrete, ist für $p = \frac{1}{2}$, $\frac{1}{2^1} + \frac{1}{2^4} + \frac{1}{2^9} + \cdots =$ irrational (nicht periodisch im dyadischen System). Man kann jede irrationale Zahl $\frac{1}{2^\alpha} + \frac{1}{2^\beta} + \cdots$ in $(0, 1)$ als $w(A_\alpha + A_\beta + \cdots)$ erhalten.

Wahrscheinlichkeit, dass der günstige Fall mindestens $k$ mal eintrete? ($k$ natürliche Zahl) $A_n$ sei das Ereignis, dass er das erste Mal nach $n$ Versuchen $k$ mal eingetreten sei, d. h. beim $n$. Versuche eintrete und vorher $k - 1$ mal.

$$w(A_n) = \binom{n-1}{k-1} p^k q^{n-k} \quad (n \geq k, \text{ sonst } A_n \text{ unmöglich}).$$

$A = A_k + A_{k+1} + \cdots$, | Bl. 63v

$$w(A) = \sum_{n \geq k} \binom{n-1}{k-1} p^k q^{n-k} = p^k[1 + \binom{k}{1} q + \binom{k+1}{2} q^2 + \cdots] = p^k(1-q)^{-k} = 1.$$

Wahrscheinlichkeit, dass der günstige Fall nur endlich oft eintrete? Da die Wahrscheinlichkeit, dass er mindestens $k$ mal eintrete, = 1 ist, ist die Wahrscheinlichkeit für das Ereignis $B_{k-1}$, dass er genau $k - 1$ mal eintrete, = 0. $B = B_0 + B_1 + \cdots =$ Ereignis, dass er nur endlich oft eintrete. $w(B) = w(B_0) + w(B_1) + \cdots = 0$. Die Wahrscheinlichkeit, dass der günstige Fall $\infty$ oft eintrete, ist = 1. | Bl. 64

Schreiben wir $A \subseteq B$ (oder $B \supseteq A$), wenn mit $A$ zugleich gewiss $B$ eintritt oder $B$ aus $A$ folgt, d. h. $A = AB$. Dann ist $B = AB + \bar{A}B = A + \bar{A}B$, oder $w(A) \leq w(B)$ (alle für $A$ günstigen Fälle sind auch für $B$ günstig, ausserdem aber können noch weitere Fälle für $B$ günstig sein). Mit $A \subseteq B$ ist $\bar{A} \supseteq \bar{B}$. Man kann, für $A \subseteq B$, statt $\bar{A}B$ auch $B - A$ schreiben: $B$ ohne $A$ tritt ein.

Sei $S_1 \subseteq S_2 \subseteq \cdots$ eine aufsteigende Ereignisfolge, d. h. jedes folgt aus dem vorhergehenden; $S = S_1 \cup S_2 \cup \cdots$ Dann ist

$$w(S) = \lim w(S_n), \tag{2}$$

wo rechts der Grenzwert einer monotonen (aufsteigenden) Folge steht. In der That setze man $A_1 = S_1$, für $n \geq 2 \quad A_n = \bar{S}_{n-1}S_n$, dann ist $S_n = S_{n-1}S_n + \bar{S}_{n-1}S_n = S_{n-1} + A_n$, $S_n = A_1 + A_2 + \cdots + A_n$ und $S = A_1 + A_2 + \cdots$, so dass (2) aus (1) folgt.

Ist $P_1 \supseteq P_2 \supseteq \cdots$ eine absteigende Ereignisfolge (jedes ist Consequenz des nachfolgenden), und $P = P_1P_2\cdots$, so ist

$$w(P) = \lim w(P_n). \tag{3}$$

Folgt aus (2) durch Complementbildung.

Für eine beliebige Ereignisfolge $A_1, A_2, \ldots$ sind

$$S_n = A_1 \cup \cdots \cup A_n, \qquad P_n = A_1 \cdots A_n$$

Bl. 65 solche auf- resp. absteigende Folgen, zugleich ist |

$$S = A_1 \cup A_2 \cup \cdots = S_1 \cup S_2 \cup \cdots, \quad P = A_1A_2\cdots = P_1P_2\cdots,$$

also

$$w(A_1 \cup A_2 \cup \cdots) = \lim w(A_1 \cup \cdots \cup A_n) \tag{4}$$

$$w(A_1A_2\cdots) = \lim w(A_1 \cdots A_n) \tag{5}$$

Mit einer Ereignisfolge $A_1, A_2, \ldots$ bilden wir noch folgende Ereignisse:

$A^\infty$ : unendlich viele $A_n$ treten ein.
$A_\infty$ : fast alle $A_n$ treten ein.

(fast alle, d. h. alle bis auf höchstens endlich viele Ausnahmen). Man nennt dies den *oberen und unteren Limes* der Ereignisfolge

$$A^\infty = \limsup A_n, \quad A_\infty = \liminf A_n.$$

Dabei ist $A_\infty \subseteq A^\infty$; wenn beide Ereignisse identisch sind, nennt man dies Ereigniss $A$ den *Limes* der Folge

$$A = \lim A_n$$

und bezeichnet die Folge als *convergent.*

Ist $B_n$ das Gegentheil von $A_n$, so ist $B^\infty = \limsup B_n$ das Gegentheil von $A_\infty$, $B_\infty = \lim \inf B_n$ das Gegentheil von $A^\infty$. Mit $A_n$ ist auch $B_n$ convergent

und $B = \lim B_n$ das Gegentheil von $A$.
Es gelten folgende Darstellungen:

$$\begin{aligned} A^\infty &= S_1 S_2 \cdots , \quad S_n = A_n \cup A_{n+1} \cup \cdots \\ A_\infty &= P_1 \cup P_2 \cup \cdots , \quad P_n = A_n A_{n+1} \cdots . \end{aligned}$$

Die zweite Formel ist evident (treten fast alle $A_n$ ein, so tritt ein $P_n$ ein, und umgekehrt), die erste folgt am einfachsten durch Complementbildung. Da hierbei die $S_n$ eine *absteigende*, die $P_n$ eine *aufsteigende* Folge bilden, ist

$$w(A^\infty) = \lim w(A_n \cup A_{n+1} \cup \cdots) \tag{6}$$

$$w(A_\infty) = \lim w(A_n A_{n+1} \cdots), \tag{7}$$

wo die rechts unter lim stehenden Wahrscheinlichkeiten nach (4)(5) zu ermitteln sind. Ferner ist $w(A_n \cup A_{n+1} \cup \cdots) \geq w(A_n) \geq w(A_n A_{n+1} \cdots)$ und das giebt

$$w(A^\infty) \geq \limsup w(A_n) \geq \liminf w(A_n) \geq w(A_\infty).$$

Ist die Mengenfolge convergent, $A = \lim A_n$, so folgt hieraus $w(A) = \lim w(A_n)$.

Eine aufsteigende (absteigende) Folge ist immer convergent und hat die Summe (Product) als Limes, so dass (2)(3) Sonderfälle hiervon sind.

Aus der Formel (§ 1) $\quad w(A \cup B) + w(AB) = w(A) + w(B)$ folgt $w(A \cup B) \leq w(A) + w(B)$, ebenso $w(A_1 \cup \cdots \cup A_n) \leq w(A_1) + \cdots + w(A_n)$ und nach (4)

$$w(A_1 \cup A_2 + \cup \cdots) \leq w(A_1) + w(A_2) + \cdots \tag{8}$$

(was trivial ist, wenn die Reihe rechts divergirt oder einen Werth $\geq 1$ hat). Bl. 67

*Wenn die Reihe* $\sum \; w(A_n)$ *convergirt, ist* $w(A^\infty) = 0$. Denn nach (8) convergirt $w(A_n \cup A_{n+1} \cup \cdots) \leq w(A_n) + w(A_{n+1}) + \cdots$ mit $n \to \infty$ nach 0.

Dies gestattet uns die am Schluss von § 2 angedeutete Verschärfung des *Gesetzes der grossen Zahlen* zu beweisen. Knüpfen wir an die damaligen Bezeichnungen (12) und an die Abschätzung (17) an: ist $A_n$ das Ereignis, dass $\dfrac{|y_n|}{n} \geq \varepsilon$, so ist $w(A_n) \leq \dfrac{\mathcal{L}^4}{n^2 \varepsilon^4}$ , also $\sum \; w(A_n)$ convergent. Danach ist $w(A^\infty) = 0$ und für das Gegentheil ($B_n$ bedeutet $\dfrac{|y_n|}{n} < \varepsilon$) $\quad w(B_\infty) = 1$, d. h. mit der Wahrscheinlichkeit 1 ist *schliesslich* $\dfrac{|y_n|}{n} < \varepsilon$, für irgend ein positives $\varepsilon$. Bringen wir nun noch die Abhängigkeit von $\varepsilon$ zum Ausdruck, indem wir für $A^\infty, B_\infty \quad A(\varepsilon),\ B(\varepsilon)$ schreiben; $A(\varepsilon)$ bedeutet: unendlich oft $\dfrac{|y_n|}{n} \geq \varepsilon$. $B(\varepsilon)$ bedeutet: schliesslich $\dfrac{|y_n|}{n} < \varepsilon$. Sei nun

$$A = A(1) \cup A\left(\frac{1}{2}\right) \cup \cdots , \quad B = B(1) B\left(\frac{1}{2}\right) \cdots .$$

$B$ bedeutet: für *jedes* $k = 1, 2, \ldots$ gilt schliesslich $\frac{|y_n|}{n} < \frac{1}{k}$, d. h. $\frac{y_n}{n}$ convergirt nach 0. Nun ist nach (8)

$$w(A) \leq w(A(1)) + w\left(A\left(\frac{1}{2}\right)\right) + \cdots = 0 + 0 + \cdots = 0,$$

Bl. 68 also $w(B) = 1$ : *mit der Wahrscheinlichkeit* 1 *convergirt* $\frac{y_n}{n}$ *nach* 0. | Dies gilt z. B., wenn alle Variablen $\xi_n$ die gleiche Vertheilung haben. Insbesondere lässt sich der Bernoullische Satz nun so formuliren (§ 2, IV):

Es werde eine Folge unabhängiger Versuche gemacht, bei jedem Versuch sei $p$ die Wahrscheinlichkeit des günstigen Falls. Die relative Häufigkeit des günstigen Falls bei den ersten $n$ Versuchen convergirt (für $n \to \infty$ ) mit der Wahrscheinlichkeit 1 nach $p$.

Z. B. wird unbegrenzt oft ein Würfel geworfen, so convergirt die relative Häufigkeit eines bestimmten Wurfes (z. B. 4) mit Wahrscheinlichkeit 1 nach $\frac{1}{6}$. Es besteht sogar die Wahrscheinlichkeit 1, dass dies für alle 6 Augenzahlen zugleich geschieht; denn das Product $B$ von endlich, sogar abzählbar vielen Ereignissen $B_i$, welche $w(B_i) = 1$ haben, hat auch noch die Wahrscheinlichkeit 1 (die Complemente $A_i$ haben $w(A_i) = 0$, also für $A = A_1 \cup A_2 \cup \cdots =$ Complement von $B$ : $w(A) \leq 0 + 0 + \cdots = 0$, $w(B) = 1$). Selbst wenn es sich also bei jedem Versuch um eine *unendliche* vollständige Disjunction $C_1, C_2, \ldots$ mit den Wahrscheinlichkeiten $p_1, p_2, \ldots$ handelt, so besteht die Wahrscheinlichkeit 1, dass *sämtliche* relativen Häufigkeiten von $C_i$ gleichzeitig nach $p_i$ convergiren.

Die Wahrscheinlichkeit 1 besteht, dass ein Decimalbruch *jede* Ziffer $0, 1, \ldots 9$

Bl. 69 mit einer relativen Häufigkeit $\to \frac{1}{10}$ enthalte (Borel). |

Bildet man aus den Ereignissen $A_n$ die Producte von je zweien und ordnet sie in einer Folge, so ist deren Summe

$$S^2 = A_1A_2 \cup A_1A_3 \cup A_2A_3 \cup \cdots$$

das Ereignis, dass mindestens zwei von den $A_n$ eintreten; analog ist $S^3, \ldots, S^k$ zu definiren ($S^1 = A_1 \cup A_2 \cup \cdots$ ; $S^0$ wäre = der Gewissheit $G$ zu setzen). Diese Summen bilden eine absteigende Folge $S^0 \supseteq S^1 \supseteq S^2 \supseteq \cdots \supseteq A^\infty$. Man kann auch das Ereignis $A^k$ betrachten, dass genau $k$ von den Ereignissen $A_n$ eintreten, wobei $S^k = A^k + S^{k+1}$ und

$$S^k = A^k + A^{k+1} + A^{k+2} + \cdots + A^\infty;$$

hierbei ist

$$A^0 = B_1B_2B_3 \cdots \quad (B_n \text{ Gegentheil von } A_n),$$

$$A^1 = (A_1B_2B_3 \cdots) + (B_1A_2B_3 \cdots) + (B_1B_2A_3 \cdots) + \cdots$$

analog $A^2, A^3, \ldots$

Nehmen wir an, *jedes $A_n$ sei von dem Resultat aller übrigen Versuche unabhängig* und habe die Wahrscheinlichkeit $p_n$, $B_n$ die Wahrscheinlichkeit $q_n =$

$1-p_n$; zur Vermeidung von Ausnahmefällen sei $0 < p_n < 1$. $A_n$ werde beim $n^{\text{ten}}$ Versuche der günstige Fall genannt, und wir sagen: es liege der *Convergenzfall* $C$ oder der *Divergenzfall* $D$ vor, jenachdem $\sum p_n$ convergirt oder divergirt. (Wenn z. B. alle $p_n = p$ sind, liegt der Divergenzfall vor). | Bl. 70

Dann ist $\quad w(A^0) = q_1 q_2 q_3 \cdots = (1-p_1)(1-p_2)(1-p_3)\cdots = w_0$
und dies unendliche Product ist

$$\begin{Bmatrix} \text{positiv} \\ 0 \end{Bmatrix} \text{ im Falle } \begin{Bmatrix} C \\ D \end{Bmatrix}.$$

Ferner hat $A_1 B_2 B_3 \cdots$ die Wahrscheinlichkeit $p_1 q_2 q_3 \cdots = w_0 \cdot \frac{p_1}{q_1}$ und also

$$w_1 = w(A^1) = w_0 \frac{p_1}{q_1} + w_0 \frac{p_2}{q_2} + \cdots .$$

Im Falle $C$ schreibe man

$$w_1 = w_0 \left( \frac{p_1}{q_1} + \frac{p_2}{q_2} + \cdots \right), \text{ da } \sum \frac{p_n}{q_n} = \sum \frac{p_n}{1-p_n}$$

wegen $p_n \to 0$ convergirt. Im Falle $D$ hat man $w_1 = 0 + 0 + \cdots = 0$. Ebenso ist $w_2 = w(A^2)$ im Falle (C)

$$w_2 = w_1 \left( \frac{p_1 p_2}{q_1 q_2} + \frac{p_1 p_3}{q_1 q_3} + \frac{p_2 p_3}{q_2 q_3} + \cdots \right),$$

im Falle $D \quad w_2 = 0$.
Kurzum: im Falle $\genfrac{}{}{0pt}{}{C}{D}$ sind alle $w_k \overset{>}{=} 0 \quad (k = 0, 1, 2, \ldots)$
Die Wahrscheinlichkeit, dass der günstige Fall nur endlich oft eintrete (aber der ungünstige fast immer), ist

$$w(B_\infty) = w_0 + w_1 + w_2 + \cdots$$

Im Falle ($C$) erhält man dafür 1, wie wir schon wissen; in der That ist

$$w_0 + w_1 + w_2 + \cdots = w_0 \left(1 + \frac{p_1}{q_1}\right)\left(1 + \frac{p_2}{q_2}\right) \cdots = w_0 : q_1 q_2 q_3 \cdots = 1.$$

| Im Falle $D$ ist $w(B_\infty) = 0$. Also: Bl. 71

Convergenzfall : $w(A^\infty) = 0, \quad w(B_\infty) = 1.$
Divergenzfall : $w(A^\infty) = 1, \quad w(B_\infty) = 0.$

Der günstige Fall tritt mit Wahrscheinlichkeit 1, im $\begin{Bmatrix} \text{Convergenzfall endlich} \\ \text{Divergenzfall unendlich} \end{Bmatrix}$ oft ein.

Weiteres hierüber: E Borel, sur les probabilités dénombrables et leurs applications arithmétiques, Rend. Palermo 27 (1909); *principiell ganz unklar.*

Die Behauptung für den Convergenzfall gilt immer, die für den Divergenzfall unter der Voraussetzung der *Unabhängigkeit.*

## § 5. Allgemeine Vertheilungen und ihre Momente. Additive Mengenfunktionen und Stieltjes-Integrale.

Die bisherigen Betrachtungen, insbesondere die von § 4, schweben insofern noch in der Luft, als sie sich auf *Axiome* stützen, denen wir übrigens noch das (bei elementaren Wahrscheinlichkeiten entbehrliche) Axiom

$$(\delta) \qquad w(A) \geq 0$$

anreihen müssen. Es ist noch zu zeigen, dass die Axiome erfüllbar sind, d. h. für einen gewissen Kreis von Ereignissen eine den Axiomen genügende Wahrscheinlichkeit $w(A)$ definirt werden kann. Z. B. kann ja ein Ereignis auf mehrere Weisen als Summe disjunkter Ereignisse darstellbar sein, $A_1 + A_2 + \cdots = B_1 + B_2 + \cdots$, und es bedarf des Nachweises, dass dann $w(A_1) + w(A_2) + \cdots = w(B_1) + w(B_2) + \cdots$ ist.

Wir machen jetzt den Übergang zur Mengenlehre, indem wir ein Ereignis als *Menge der ihm günstigen Fälle* auffassen; hierbei betrachten wir nur ein System von Ereignissen $A$, bei dem die *Menge $M$ der möglichen Fälle immer dieselbe* ist; $A$ ist Theilmenge von $M$. Der *Summe* und dem *Produkt* von Ereignissen entspricht *Summe* und *Durchschnitt* von Mengen; die bisherigen Bezeichnungen sind bereits die der Mengenlehre. Das Complement $\overline{A}$ von $A$ ist $M - A$; für $A \subseteq B$ ist $\overline{A}B = B - A$.

Bl. 72v Die Voraussetzung, dass alle betrachteten Ereignisse auf *dieselbe* Menge $M$ der möglichen Fälle zu beziehen sind, kann ev. durch Combination verwirklicht werden. Wenn man z. B. Würfe mit einem Würfel (mögliche Fälle $1, 2, \ldots, 6$) und mit einer Münze (mögliche Fälle $B$=Bild, $S$=Schrift) gleichzeitig betrachten will, so sind die Paare $(1, B) \cdots (6, B)(1, S) \cdots (6, S)$ als mögliche Fälle anzusehen. Würfelwurf 3 hat die günstigen Fälle $(3, B)(3, S)$; Münzenwurf Bild hat die günstigen Fälle $(1, B) \cdots (6, B)$; das Product beider Ereignisse hat als Menge der günstigen Fälle den Durchschnitt beider Mengen, d. h. den einen Fall $(3, B)$.

Allgemein: ist $M$ eine Menge möglicher Fälle $x$, $A$ eine Theilmenge von $M$, und $N$ eine Menge möglicher Fälle $y$, $B$ eine Theilmenge von $N$, so betrachte man die Menge $P$ der möglichen *Paare* $(x, y)$ (es braucht nicht jedes $x$ mit jedem $y$ combinirt ein mögliches Paar zu bilden) und nennen $A^*$ die Menge der Paare, für die $x$ zu $A$ gehört, $B^*$ die Menge der Paare, für die $y$ zu $B$ gehört (bei geometrischer Deutung: $A^*$ die grösste Theilmenge der ebenen Menge $P$, deren Projection auf die $x$-Achse $\subseteq A$ ist). Die Ereignisse $A, B$ werden dann durch die Theilmengen $A^*, B^*$ derselben Menge $P$ dargestellt. ($M^* = N^* = P$).

Bl. 73 Bei endlichen Mengen $M$, $A$ konnte $w(A) = \frac{a}{m}$ durch *Zählung* der (als gleichwahrscheinlich vorausgesetzten) Fälle erklärt werden. In § 4, theilweise schon früher (bei unbegrenzt fortgesetzten Spielpartien) trafen wir aber auf unendliche Mengen; z. B. bei unbegrenzt wiederholtem Werfen einer Münze mit den Seiten $0, 1$ sind alle aus Nullen und Einsen in irgendwelcher Anordnung bestehenden Ziffernfolgen $(x_1, x_2, \ldots)$ als mögliche Fälle anzusehen; sie bilden eine (unabzählbar) unendliche Menge. Hier wird $w(A)$ auf irgend einer anderen Vergleichung, Messung, Wägung der Mengen $A$, $M$ beruhen müssen. Jedenfalls ist $w(A)$ eine *Mengenfunction*, für alle oder gewisse Theilmengen von $M$ definirt, und eine der Bedingung $(\gamma)$

$$w(A_1 + A_2 + \cdots) = w(A_1) + w(A_2) + \cdots$$

genügende Mengenfunction nennt man *additiv* (auch absolut additiv); sie ist nichtnegativ, wegen $(\delta)$, und wegen $(\alpha)$ ist $w(M) = 1$. Diese letztere Bedingung ist für unseren jetzigen Standpunkt zunächst unwesentlich; ist $\mu$ eine positive Constante, so ist $\Phi(A) = \mu\, w(A)$ ebenfalls eine nichtnegative, additive Mengenfunction mit $\Phi(M) = \mu$. Umgekehrt, ist $\Phi(A)$ eine nichtnegative, additive Mengenfunction mit $\Phi(M) > 0$, so ist $w(A) = \dfrac{\Phi(A)}{\Phi(M)}$ eine unseren Postulaten | genügende Wahrscheinlichkeit, auch die relative Wahrscheinlichkeit Bl. 74
$w_A(B) = \dfrac{\Phi(AB)}{\Phi(A)}$ ($\Phi(A) > 0$ angenommen) ist in gleicher Weise erklärbar. *Die Wahrscheinlichkeit ist Quotient zweier Werthe einer nichtnegativen, additiven Mengenfunction.* Wir können $\Phi(A)$ als eine Art *Mass* (oder Gewicht) der Menge $A$ auffassen, und die Wahrscheinlichkeit beruht auf der Vergleichung dieser Masse. Nennen wir die Mengen $A$, deren $\Phi(A)$ definirt ist, *messbar*; das System der messbaren Mengen heisse $\mathfrak{M}$.

Über dies System $\mathfrak{M}$, das keineswegs alle Theilmengen von $M$ zu umfassen braucht, müssen wir auch noch eine Annahme machen. Wir haben doch, ohne es ausdrücklich zu erwähnen, vorausgesetzt, dass mit den Ereignissen $A_1, A_2, \ldots$ auch ihre Summe und ihr Produkt eine Wahrscheinlichkeit haben, ferner dass mit $A$ auch sein Gegentheil $\overline{A}$, mit $A$ und $B \supseteq A$ auch $\overline{A}B = B - A$ eine Wahrscheinlichkeit haben. Die entsprechende Eigenschaft von $\mathfrak{M}$ wollen wir kurz *Abgeschlossenheit* nennen: (Borelsches System)

*Ein Mengensystem heisst abgeschlossen, wenn die Differenz zweier Mengen des Systems, sowie die Summe und der Durchschnitt von endlich oder abzählbar vielen Mengen des Systems wieder dem System angehören.*

Wenn, wie hier, das System eine grösste Menge $M$ enthält, kann die Differenzforderung durch die specielle ersetzt werden, dass mit $A$ auch das Complement $M - A$ dem System angehören soll.

| Was wir brauchen, ist also: *ein abgeschlossenes Mengensystem* $\mathfrak{M}$*, bestehend aus allen oder gewissen Theilmengen $A$ von $M$ (wozu $M$ selbst gehört) und in ihm eine additive, nichtnegative Mengenfunction* $\Phi(A)$ (insbesondere $\Phi(M) > 0$). Bl. 75

Die Bestimmung $\Phi(A) \geq 0$ erwähnen wir künftig nicht besonders (die allgemeinen additiven Mengenfunctionen sind Differenzen von zwei nichtnegativen). *Beispiele*: Wenn es sich um Lage eines Punktes $x$ auf einer Geraden, also um lineare Mengen $A, M$ handelt, so giebt es eine additive Mengenfunction $m(A)$, die für ein Intervall die Intervalllänge giebt: das *Längenmass* von $A$ (im Sinne von H. Lebesgue; wir kommen darauf zurück. Die früheren Längenmassdefinitionen von C. Jordan und G. Peano sind nur beschränkt additiv; sie erfüllen nur ($\beta$), nicht ($\gamma$)). Es giebt aber unmessbare Mengen, für die nur ein äusseres und ein inneres Längenmass, beide verschieden, definirt sind; dann ist also $w(A)$ nicht definirt. Auch der Fall unendlichen Längenmasses ist auszuschliessen; z. B. ist die Wahrscheinlichkeit, dass eine reelle Zahl positiv sei ($M$ die Gerade, $A$ die Halbgerade $x > 0$) hiernach nicht definirt. Man muss ein messbares $M$ mit $0 < m(M) < \infty$, z. B. ein Intervall, und eine messbare Theilmenge $A$ von $M$ voraussetzen; dann ist $w(A) = \frac{m(A)}{m(M)}$ | eine zulässige, d. h. unseren Axiomen genügende Wahrscheinlichkeit. Ob sie die praktisch brauchbare ist, ist eine andere Frage. Bei einem Roulette (ohne Feldeintheilung) wird die Lage der Kugel in einem bestimmten Kreisintervall wohl die Wahrscheinlichkeit Intervalllänge: Kreisumfang haben. Ist aber $x$ etwa ein Beobachtungsfehler, so werden sehr grosse $x$ weniger wahrscheinlich sein als kleine, und die Wahrscheinlichkeit für Lage von $x$ in einem Intervall wird nicht einfach der Intervalllänge proportional sein. Die Wahl der geeigneten Mengenfunction $\Phi(A)$ ist, genau wir früher die Bewerthung der gleichwahrscheinlichen Fälle, von der Erfahrung vorzuschreiben oder an ihr zu prüfen. – Bl. 76

Dass nicht eine und dieselbe Function für alle Fälle zutreffen kann, ist schon durch die Möglichkeit der *Variablentransformation* klar; besteht zwischen der $x$- und der $y$-Geraden eine eineindeutige Transformation, so entsprechen den Mengen $A, M$ dort die Mengen $B, N$ hier, und das Ereigniss, dass $x$ zu $A$ gehört, ist identisch damit, dass $y$ zu $B$ gehört. Dies letztere hat also die Wahrscheinlichkeit $\frac{\Phi(A)}{\Phi(M)}$, was im Allgemeinen nicht gleich $\frac{\Phi(B)}{\Phi(N)}$ sein wird; man hat eben dann auf der $y$-Achse die Mengenfunction $\Psi(B) = \Phi(A)$ zu wählen.

Bl. 77 | Entsprechendes gilt für ebene Mengen, wo z. B. das Lebesguesche *Flächenmass* $m(A)$ eine additive Mengenfunction ist. Wenn auf eine Scheibe $M$ blindlings geschossen wird – jedoch so, dass sicher $M$ getroffen wird oder nur die $M$ treffenden Schüsse betrachtet werden –, so mag die Wahrscheinlichkeit, dass $A$ getroffen wird, gleich $\frac{m(A)}{m(M)}$ sein; wird aber auf einen bestimmten Punkt von $M$ gezielt, so wird eine andere Mengenfunction zu wählen sein.

Auf der Geraden $M$ seien abzählbar viele Punkte $x_1, x_2, \dots$ und zugeordnete positive Zahlen $p_1, p_2, \dots$ gegeben, die eine convergente Reihe $\sum\limits_1^\infty p_n$ bilden; es sei dann $\Phi(A) = \sum\limits_A p_n$, erstreckt über alle $n$, für die $x_n$ zu $A$ gehört. Das ist eine für alle linearen Mengen definirte additive Mengenfunction.

### *Vertheilung einer reellen Variablen.*

$M$ sei die ganze Gerade, $x$ ein Punkt auf ihr; es sei eine additive Mengenfunction $\Phi(A)$ $(\Phi(M) = \mu > 0)$ gegeben, definirt im abgeschlossenen Mengensystem $\mathfrak{M}$.

*Zu $\mathfrak{M}$ sollen insbesondere die Intervalle gehören* (mit und ohne Endpunkte; bekannte Bezeichnungen $[\xi,\eta], [\xi,\eta)$ usw.), folglich auch die *Halbgeraden* (mit und ohne Endpunkt). Zu der Mengenfunction $\Phi$ gehört eine *Punktfunction* (Vertheilungsfunction) | Bl. 78

$$\varphi(\xi) = \Phi(-\infty,\xi) \qquad (= \mu \cdot w(x < \xi))$$

die *monoton* ist, denn für $\xi < \eta$ ist nach $(\beta)$

$$\varphi(\eta) - \varphi(\xi) = \Phi[\xi,\eta) \geq 0.$$

Auf Grund der Additivität (vgl. § 4, (2)(3)) beweist man leicht:

$$\varphi(+\infty) = \lim_{\xi\to\infty} \varphi(\xi) = \mu\,, \quad \varphi(-\infty) = \lim_{\xi\to-\infty} \varphi(\xi) = 0\,,$$

$$\varphi(\xi-0) = \lim_{x\to\xi-0} \varphi(x) = \varphi(\xi)\ , \qquad \varphi(\xi) \textit{ nach links stetig},$$

während

$$\varphi(\xi+0) = \lim_{x\to\xi+0} \varphi(x) = \Phi(-\infty,\xi] = \varphi(\xi) + \Phi[\xi]$$

ist ($[\xi]$ die einpunktige Menge, die aus $\xi$ besteht). Die *Sprungstellen* von $\varphi$, wo $\varphi(\xi+0) - \varphi(\xi) = \Phi[\xi] > 0$, sind in höchstens abzählbarer Menge vorhanden und die Summe der Sprünge ist convergent $\leq \mu$.

Man hätte auch die Punktfunction

$$\psi(\xi) = \Phi(-\infty,\xi] \qquad (= \mu \cdot w(x \leq \xi))$$

betrachten können: $\psi(\xi) = \varphi(\xi+0) = \psi(\xi+0)$ , $\varphi(\xi) = \varphi(\xi-0) = \psi(\xi-0)$. $\psi$ ist nach rechts stetig; beide stimmen ausser an den Sprungstellen überein. Ist allgemein $\chi(\xi)$ eine monotone Function, die mit $\varphi(\xi)$ an überall dicht liegenden Stellen übereinstimmt und daher dieselben einseitigen Grenzwerthe hat, so ist $\varphi(\xi) = \chi(\xi-0)$ , $\psi(\xi) = \chi(\xi+0)$, $\varphi \leq \chi \leq \psi$. Für analytische Darstellungen ist $\chi = \frac{1}{2}(\varphi+\psi)$ wichtig. | Bl. 79

*Reine Sprungfunctionen.* $p(x)$ sei $\geq 0$ und an höchstens abzählbar vielen Stellen $> 0$, die Summe aller $p(x) > 0$ sei convergent $= \mu$ $(> 0)$. Dann definirt

$$\varphi(\xi) = \sum_{x<\xi} p(x)$$

eine links stetige Vertheilungsfunction ($\varphi(-\infty) = 0,\ \varphi(\infty) = \mu$) mit $\varphi(\xi+0) - \varphi(\xi) = p(\xi)$. Sie heisst eine reine Sprungfunction, da ihr Zuwachs gleich der Summe der Sprünge des betreffenden Intervalls ist:

$$\varphi(\eta) - \varphi(\xi) = \sum_{\xi \leq x < \eta} [\varphi(x+0) - \varphi(x)].$$

Hierzu gehören insbesondere die *Treppenfunctionen* (nur endlich viele $p(x) > 0$; elementare Vertheilungen). Aber die Sprungstellen können z. B. auch überall dicht liegen.

*Stetige Vertheilungsfunctionen* (ohne Sprünge). Die Ableitung $\varphi'(\xi)$, wo sie existirt, heisst die *Dichtigkeit* der Vertheilung (Wahrscheinlichkeitsdichte für $\mu = 1$). Unter ihnen sind besonders wichtig die *Integralfunctionen*

$$\varphi(\xi) = \int\limits_{-\infty}^{\xi} \vartheta(x)\,dx, \quad \vartheta(x) \geq 0;$$

etwa mit stetigem $\vartheta(x)$, das dann die Dichtigkeit ist. Z. B.

$$\varphi(\xi) = c \int\limits_{-\infty}^{\xi} \frac{dx}{1+x^2} = c\left(\frac{\pi}{2} + \operatorname{arctg}\xi\right);$$

$$\varphi(\xi) = c \int\limits_{-\infty}^{\xi} e^{-x^2} dx \quad \text{(Exponentialgesetz).}$$

Im Allgemeinen lässt sich $\varphi(\xi)$ als Summe einer Sprungfunction und einer stetigen Function (beide monoton) darstellen.

Bl. 80 | Sei nun umgekehrt eine links stetige monotone Function $\varphi(\xi)$ mit $\varphi(-\infty) = 0$, $\varphi(+\infty) = \mu$ gegeben: giebt es dann ein *abgeschlossenes*, die Intervalle enthaltendes Mengensystem und in ihm eine additive Mengenfunction $\Phi(A)$ derart, dass $\varphi(\xi) = \Phi(-\infty, \xi)$ die zugehörige Punktfunction ist? Wir zeigen, dass diese Frage zu bejahen ist. [3]

Betrachten wir zunächst die (links abgeschlossenen, rechts offenen) *Intervalle* $I = [\alpha, \beta)$ $(\alpha < \beta)$ und definiren für sie[3]

$$\Phi(I) = \varphi(\beta) - \varphi(\alpha) = \varphi_\alpha^\beta\,.$$

Die Nullmenge rechnen wir mit dazu (ebenso zu den $E, A$ im Folgenden) und setzen $\Phi(0) = 0$.

Bilden wir zunächst die Summen aus endlich vielen, paarweise fremden Intervallen

$$E = I_1 + I_2 + \cdots + I_n\,.$$

Da der Durchschnitt zweier $I$ ein $I$ ist (eventuell 0), so ist der Durchschnitt zweier $E$ ein $E$ (eventuell 0, so auch im Folgenden). Ferner ist $I - E$ offenbar ein $E$ und die Differenz $E - E_1$ zweier $E$ ein $E$, denn schliessen wir $E$ in ein $I$ ein, so ist $E - E_1 = E(I - E_1)$ Durchschnitt zweier $E$. Die Summe von zwei fremden $E$ ist ein $E$, aber auch die von zwei beliebigen: $E_1 \cup E_2 = E_1 + (E_2 - E_1 E_2)$.

---

[3] $\Phi$ Mass, $\overline{\Phi}$ äusseres Mass, $\underline{\Phi}$ inneres Mass bezüglich $\varphi$

Also: *Summe und Durchschnitt von zwei (oder endlich vielen) $E$ ist ein $E$, die Differenz zweier $E$ ist ein $E$.*

*Für $I = I_1 + I_2 + \cdots I_n$ ist* $\quad \Phi(I) = \Phi(I_1) + \Phi(I_2) + \cdots \Phi(I_n)$.

Denn nehmen wir alle Intervalle $\neq 0$ an und ordnen von links nach rechts, so ist $I_k = [\xi_{k-1}, \xi_k)$ $(k = 1, \ldots, n;\ \xi_0 < \xi_1 < \cdots < \xi_n)$ und $I = [\xi_0, \xi_n)$, $\varphi_{\xi_0}^{\xi_n} = \sum\limits_1^n \varphi_{\xi_{k-1}}^{\xi_k}$. Definiren wir nun für

$$E = I_1 + \cdots + I_n \ : \quad \Phi(E) = \Phi(I_1) + \cdots + \Phi(I_n),$$

*so ist dies von der Zerlegung unabhängig.* Denn ist gleichzeitig ($m, n$ mögen Gruppen verschiedener Indices durchlaufen) $E = \sum_m I_m = \sum_n I_n$, so sind auch die Durchschnitte $I_m I_n$ Intervalle, ferner

$$I_m = \sum_n I_m I_n, \quad I_n = \sum_m I_m I_n\ , \ \text{ also}$$

$$\Phi(I_m) = \sum_n \Phi(I_m I_n), \quad \Phi(I_n) = \sum_m \Phi(I_m I_n),$$
$$\sum_m \Phi(I_m) = \sum_{m,n} \Phi(I_m I_n) = \sum_n \Phi(I_n).$$

Damit ist $\Phi$ für die Mengen $E$ definirt. Für zwei fremde $E$ folgt

$$\Phi(E_1 + E_2) = \Phi(E_1) + \Phi(E_2)$$

unmittelbar; allgemein ist (Betrachtung der $E$-Mengen $E_1E_2$, $E_1 - E_1E_2$, $E_2 - E_1E_2$)

$$\Phi(E_1 \cup E_2) + \Phi(E_1E_2) = \Phi(E_1) + \Phi(E_2).$$

Also $\Phi(E_1 \cup E_2) \leq \Phi(E_1) + \Phi(E_2)$, was sich auf endlich viele Summanden überträgt.

Sodann betrachten wir die Summen

$$A = I_1 + I_2 + \cdots$$

aus endlich oder abzählbar vielen, paarweise fremden Intervallen. Bl. 82

*Der Durchschnitt zweier $A$ ist ein $A$*:

$$AA' = I_1I_1' + I_1I_2' + I_2I_1' + \cdots$$

Nun ist aber auch die Summe einer beliebigen Intervallfolge, $I_1 \cup I_2 \cup \cdots$, ein $A$; denn mit $E_n = I_1 \cup \cdots \cup I_n$ ist dies gleich

$$E_1 + (E_2 - E_1) + (E_3 - E_2) + \cdots\ ,$$

also Summe paarweise fremder $E$ und, wenn man diese in $I$ zerlegt, paarweise fremder $I$. *Die Summe endlich oder abzählbar vieler $A$ ist ein $A$*, denn für $A_1 = I_{11} \cup I_{12} \cup \cdots$ , $A_2 = I_{21} \cup I_{22} \cup \cdots$ , $\cdots$ ist

$$A_1 \cup A_2 \cup \cdots = I_{11} \cup I_{12} \cup I_{21} \cup \cdots$$

Die Differenz zweier $A$ und der Durchschnitt abzählbar vieler $A$ braucht kein $A$ zu sein.

*Für* $I = I_1 + I_2 + \cdots$ *ist* $\Phi(I) = \Phi(I_1) + \Phi(I_2) + \cdots$.
Dies bedarf eines genauen Beweises. Zunächst ist jedenfalls

$$\Phi(I) \geq \Phi(I_1) + \cdots + \Phi(I_n), \text{ also } \Phi(I) \geq \Phi(I_1) + \Phi(I_2) + \cdots .$$

Sei nun $I = [\alpha, \beta)$, $I_n = [\alpha_n, \beta_n)$. Wir verkleinern $I$ zu $I' = [\alpha, \beta - \delta)$, können dabei aber, *wegen* $\varphi(\beta - 0) = \varphi(\beta)$, $\delta$ so klein wählen, dass $\Phi(I') > \Phi(I) - \varepsilon$. Ebenso vergrössern wir $I_n$ zu ${I_n}' = [\alpha_n - \delta_n, \beta_n)$ und können $\delta_n$ so klein nehmen, Bl. 83 dass $\Phi(I'_n) < \Phi(I_n) + \frac{\varepsilon}{2^n}$, also |

$$\sum \Phi(I'_n) < \sum \Phi(I_n) + \varepsilon.$$

Endlich sei $K$ das *abgeschlossene* Intervall $[\alpha, \beta - \delta]$, $K_n$ das *offene* Intervall $(\alpha_n - \delta_n, \beta_n)$. $K$ ist in $I_1 + I_2 + \cdots$, also in $K_1 \cup K_2 \cup \cdots$ enthalten. Dann ist aber (Borelscher Satz) $K$ bereits in der Summe endlich vieler $K_n$, also für geeignetes $n$ in $K_1 \cup \cdots \cup K_n$ enthalten. Denn hätte für jedes $n$ $K$ einen Punkt $x_n$, der nicht zu $K_1 \cup \cdots \cup K_n$ gehört, so hätten diese einen Häufungspunkt $x$, der zu $K$ gehört, also einem $K_\nu$ angehört; in $K_\nu$ müssten dann unendlich viele $x_n$ liegen, während aber doch $x_\nu, x_{\nu+1}, \ldots$ gewiss nicht in $K_\nu$ liegen. Danach ist

$$I' \subseteq K \subseteq K_1 \cup \cdots \cup K_n \subseteq I'_1 \cup \cdots \cup I'_n,$$

$$\text{also} \qquad \Phi(I') \leq \Phi(I'_1 \cup \cdots \cup I'_n) \leq \Phi(I'_1) + \cdots + \Phi(I'_n)$$

und

$$\Phi(I) - \varepsilon < \Phi(I') \leq \sum_1^\infty \Phi(I'_n) < \sum_1^\infty \Phi(I_n) + \varepsilon\,,$$

$$\Phi(I) < \sum_1^\infty \Phi(I_n) + 2\varepsilon,$$

d. h. $\Phi(I) \leq \sum\limits_1^\infty \Phi(I_n)$ und damit $\Phi(I) = \sum\limits_1^\infty \Phi(I_n)$.

Definiren wir jetzt für

$$A = I_1 + I_2 + \cdots \quad : \quad \Phi(A) = \Phi(I_1) + \Phi(I_2) + \cdots$$

so ist dies von der Zerlegungsweise unabhängig. Denn ist (wie bei den $E$) gleich- Bl. 84 zeitig $A = \sum\limits_m I_m = \sum\limits_n I_n$, so erhalten wir |

$$\sum_m \Phi(I_m) = \sum_m \sum_n \Phi(I_m I_n) = \sum_n \sum_m \Phi(I_m I_n) = \sum_n \Phi(I_n),$$

da die (absolut) convergente Doppelreihe in beliebiger Anordnung summiert werden kann. Für paarweise fremde Mengen $A$ in endlicher oder abzählbarer Menge gilt offenbar

$$\Phi(A_1 + A_2 + \cdots) = \Phi(A_1) + \Phi(A_2) + \cdots \tag{1}$$

Ferner ist

$$\Phi(A_1 \cup A_2) + \Phi(A_1 A_2) = \Phi(A_1) + \Phi(A_2). \tag{2}$$

Dies ist so zu beweisen. Schreiben wir $A = E_1 \cup E_2 \cup \cdots$ mit aufsteigenden $E_1 \subseteq E_2 \subseteq \cdots$, so ist $A = E_1 + (E_2 - E_1) + \cdots$, also nach (1)

$$\Phi(A) = \Phi(E_1) + [\Phi(E_2) - \Phi(E_1)] + \cdots = \lim \Phi(E_n).$$

Sei ebenso $A' = E_1' \cup E_2' \cup \cdots$, so ist

$$A \cup A' = \bigcup (E_n \cup E_n'), \quad AA' = \bigcup E_n E_n'$$

(auch die 2. Formel leicht einzusehen), und da wir hier wieder aufsteigende $E$-Folgen haben, so folgt aus

$$\Phi(E_n \cup E_n') + \Phi(E_n E_n') = \Phi(E_n) + \Phi(E_n')$$

für $n \to \infty$

$$\Phi(A \cup A') + \Phi(AA') = \Phi(A) + \Phi(A').$$

Aus $\Phi(E_n \cup E_n') \geq \Phi(E_n)$ folgt ebenso $\Phi(A \cup A') \geq \Phi(A)$, d. h. für $A_1 \subseteq A_2$ ist $\Phi(A_1) \leq \Phi(A_2)$.

Endlich: ist $A = A_1 \cup A_2 \cup \cdots$ Summe *aufsteigender* $A_n$, so ist Bl. 85

$$\Phi(A) = \lim \Phi(A_n). \tag{3}$$

Denn zunächst ist $\Phi(A) \geq \Phi(A_n)$. Andererseits sei mit aufsteigenden Folgen

$$\begin{array}{rcl} A_1 & = & E_{11} \cup E_{12} \cup E_{13} \cup \cdots \\ A_2 & = & E_{21} \cup E_{22} \cup E_{23} \cup \cdots \\ A_3 & = & E_{31} \cup E_{32} \cup E_{33} \cup \cdots \\ & \cdots & \end{array}$$

Es bilden dann $E_1 = E_{11},\ E_2 = E_{12} \cup E_{22},\ E_3 = E_{13} \cup E_{23} \cup E_{33}, \cdots$ eine aufsteigende Folge mit $E_1 \cup E_2 \cup \cdots = A$, während zugleich $E_n \subseteq A_n$, also

$$\Phi(A) = \lim \Phi(E_n) \leq \lim \Phi(A_n) \leq \Phi(A),$$

womit (3) bewiesen ist. (Specialfall (1))

Aus (2) folgt $\Phi(A_1 \cup A_2) \leq \Phi(A_1) + \Phi(A_2)$, ebenso für endlich viele Summanden, aber auch für abzählbar viele, denn mit $A = A_1 \cup A_2 \cup \cdots$ ist

$$\begin{aligned} \Phi(A) &= \lim \Phi(A_1 \cup \cdots \cup A_n) \leq \lim [\Phi(A_1) + \cdots + \Phi(A_n)] \\ &= \Phi(A_1) + \Phi(A_2) + \cdots \end{aligned} \tag{4}$$

(wo die Reihe rechts allerdings divergiren kann und die Behauptung inhaltslos wird).

| Zu den Mengen $A$ zählt auch $M$, das als Summe aufsteigender Intervalle $[-n, n)$ dargestellt $\Phi(M) = \varphi(+\infty) - \varphi(-\infty) = \mu$ ergiebt.

Nun wird folgendermassen weitergegangen. Zu jeder Menge $X$ giebt es einschliessende $A$ (z. B. $M$); wir definiren dann (inf = Infimum = untere Grenze)

$$\overline{\Phi}(X) = \inf \Phi(A) \text{ für } A \supseteq X. \tag{5}$$

Für eine Menge $A$ selbst ist $\overline{\Phi}(A) = \Phi(A)$, insbesonders $\overline{\Phi}(M) = \mu$, $\overline{\Phi}(0) = 0$. Für $X \subseteq Y$ ist $\overline{\Phi}(X) \leq \overline{\Phi}(Y)$.

Seien $X_1, X_2$ zwei Mengen, $A_1 \supseteq X_1,\ A_2 \supseteq X_2$, also

$$A_1 \cup A_2 \supseteq X_1 \cup X_2, \quad A_1A_2 \supseteq X_1X_2.$$

Also auf Grund von (5) und (2)

$$\overline{\Phi}(X_1 \cup X_2) + \overline{\Phi}(X_1X_2) \leq \Phi(A_1 \cup A_2) + \Phi(A_1A_2) = \Phi(A_1) + \Phi(A_2).$$

Nimmt man die untere Grenze der rechten Seite, so folgt

$$\overline{\Phi}(X_1 \cup X_2) + \overline{\Phi}(X_1X_2) \leq \overline{\Phi}(X_1) + \overline{\Phi}(X_2). \tag{6}$$

Insbesondere für fremde Mengen

$$\overline{\Phi}(X_1 + X_2) \leq \overline{\Phi}(X_1) + \overline{\Phi}(X_2),$$

speciell für $X_1 + X_2 = M$

$$\mu = \overline{\Phi}(M) \leq \overline{\Phi}(X) + \overline{\Phi}(M - X).$$

Setzt man also

$$\underline{\Phi}(X) = \overline{\Phi}(M) - \overline{\Phi}(M - X) = \mu - \overline{\Phi}(M - X), \tag{7}$$

Bl. 87 | so ist $0 \leq \underline{\Phi}(X) \leq \overline{\Phi}(X)$. Insbesondere $\overline{\Phi}(0) = \underline{\Phi}(0) = 0$, $\overline{\Phi}(M) = \underline{\Phi}(M) = \mu$. Für $X \subseteq Y : \underline{\Phi}(X) \leq \underline{\Phi}(Y)$.

Aus (6) erhält man dann durch Übergang zu den Complementen

$$\underline{\Phi}(X_1 \cup X_2) + \underline{\Phi}(X_1X_2) \geq \underline{\Phi}(X_1) + \underline{\Phi}(X_2). \tag{8}$$

Ist $X = X_1 \cup X_2 \cup \cdots$ mit *aufsteigenden* $X_n$, so ist

$$\overline{\Phi}(X) = \lim \overline{\Phi}(X_n). \tag{9}$$

Denn zunächst $\overline{\Phi}(X) \geq \overline{\Phi}(X_n)$. Andererseits kann man $A_n \supseteq X_n$ mit $\Phi(A_n) < \overline{\Phi}(X_n) + \varepsilon$ bestimmen (Definition!) und zwar so, dass auch die $A_n$ eine aufsteigende Folge bilden. Man wähle zunächst $A_1 \supseteq X_1$ mit $\Phi(A_1) < \overline{\Phi}(X_1) + \varepsilon$, sodann $A \supseteq X_2$ mit

$$\Phi(A) < \overline{\Phi}(X_2) + [\overline{\Phi}(X_1) + \varepsilon - \Phi(A_1)],$$

dann ist wegen (2) und $AA_1 \supseteq X_1$

$$\begin{aligned}\Phi(A \cup A_1) + \Phi(AA_1) &= \Phi(A) + \Phi(A_1) < \overline{\Phi}(X_1) + \overline{\Phi}(X_2) + \varepsilon \\ &\leq \Phi(AA_1) + \overline{\Phi}(X_2) + \varepsilon, \\ \Phi(A \cup A_1) &< \overline{\Phi}(X_2) + \varepsilon\end{aligned}$$

und $A_2 = A \cup A_1 \supseteq A_1$ erfüllt unsere Forderungen. So kann man weitergehen. Mit $A = A_1 \cup A_2 \cup \cdots \supseteq X$ ist also

$$\overline{\Phi}(X) \leq \Phi(A) = \lim \Phi(A_n) \leq \lim \overline{\Phi}(X_n) + \varepsilon\,,$$

also $\overline{\Phi}(X) \leq \lim \overline{\Phi}(X_n) \leq \overline{\Phi}(X)$.

Durch Complementbildung folgt für den Durchschnitt $X = X_1X_2\ldots$ einer Bl. 88
*absteigenden* Folge

$$\underline{\Phi}(X) = \lim \underline{\Phi}(X_n)\,. \tag{10}$$

Für die Mengen $A$ war $\overline{\Phi}(A) = \Phi(A)$; es ist aber auch $\underline{\Phi}(A) = \Phi(A)$. Denn für eine Menge $E$ ist $M - E$ ein $A$, also

$$\underline{\Phi}(E) = \mu - \overline{\Phi}(M - E) = \Phi(M) - \Phi(M - E) = \Phi(E).$$

Und für $A = E_1 \cup E_2 \cup \cdots$ mit aufsteigenden $E_n$ ist $\underline{\Phi}(A) \geq \underline{\Phi}(E_n) = \Phi(E_n)$,

$$\underline{\Phi}(A) \geq \lim \Phi(E_n) = \Phi(A) = \overline{\Phi}(A) \geq \underline{\Phi}(A).$$

Demgemäss wollen wir jetzt *alle* Mengen $X$, für die $\overline{\Phi}(X) = \underline{\Phi}(X)$, messbar nennen und $\Phi(X) = \overline{\Phi}(X) = \underline{\Phi}(X)$ setzen. Aus (6) und (8) folgt sofort:
*Summe und Durchschnitt von zwei messbaren Mengen ist messbar* und zwar

$$\Phi(X_1 \cup X_2) + \Phi(X_1X_2) = \Phi(X_1) + \Phi(X_2)\,, \tag{11}$$

insbesondere für fremde Mengen $\quad \Phi(X_1 + X_2) = \Phi(X_1) + \Phi(X_2)$.
*Das Complement einer messbaren Menge ist messbar.* Folgt aus (7). Also auch die Differenz von zwei messbaren Mengen.

*Die Summe einer aufsteigenden Folge messbarer Mengen $X = X_1 \cup X_2 \cup \cdots$ ist messbar und zwar*

$$\Phi(X) = \lim \Phi(X_n). \tag{12}$$

Denn nach (9) $\quad \overline{\Phi}(X) = \lim \overline{\Phi}(X_n) = \lim \underline{\Phi}(X_n) \leq \underline{\Phi}(X) \leq \overline{\Phi}(X)$. Bl. 89

Ebenso ist der Durchschnitt $X = X_1X_2\ldots$ einer absteigenden Folge messbarer Mengen messbar und es gilt wieder (12).

*Summe und Durchschnitt einer Folge messbarer Mengen ist messbar.* Mit $X = X_1 \cup X_2 \cup \cdots$ ist $\Phi(X) = \lim \Phi(X_1 \cup X_2 \cup \cdots \cup X_n)$, mit $X = X_1X_2\ldots$ ist $\Phi(X) = \lim \Phi(X_1X_2\ldots X_n)$. Für paarweise fremde Mengen und ihre Summe $X = X_1 + X_2 + \cdots$ gilt

$$\begin{aligned}\Phi(X) = \lim \Phi(X_1 + \cdots + X_n) &= \lim\,[\Phi(X_1) + \cdots + \Phi(X_n)] \\ &= \Phi(X_1) + \Phi(X_2) + \cdots\end{aligned}$$

*Das System $\mathfrak{M}$ der messbaren Mengen ist also abgeschlossen und in ihm ist $\Phi(X)$ eine additive Mengenfunction*; sie stimmt für die Mengen $A$ mit $\Phi(A)$, insbesondere für die Halbgerade $(-\infty, \xi) =$ Summe der aufsteigenden Intervalle $[\xi - n, \xi)$ mit $\varphi(\xi) - \varphi(-\infty) = \varphi(\xi)$ überein.

Bemerken wir noch die Kette von Ungleichungen ($X_1, X_2$ fremd)

$$\begin{gathered}\underline{\Phi}(X_1) + \underline{\Phi}(X_2) \leq \underline{\Phi}(X_1 + X_2) \leq \left\{ \begin{matrix} \underline{\Phi}(X_1) + \overline{\Phi}(X_2) \\ \overline{\Phi}(X_1) + \underline{\Phi}(X_2) \end{matrix} \right\} \leq \\ \leq \overline{\Phi}(X_1 + X_2) \leq \overline{\Phi}(X_1) + \overline{\Phi}(X_2) \end{gathered} \tag{13}$$

Die äusseren sind Specialfälle von (6) (8); ferner für $M = X_1 + X_2 + X_3$ nach (6)

$$\overline{\Phi}(X_1 + X_2) + \overline{\Phi}(X_2 + X_3) \geq \overline{\Phi}(M) + \overline{\Phi}(X_2)$$

oder nach (7) $\overline{\Phi}(X_1 + X_2) \geq \underline{\Phi}(X_1) + \overline{\Phi}(X_2)$, ebenso aus (8) die andere innere Ungleichung. Ist z. B. $X_1 + X_2$ messbar, so ist $\underline{\Phi}(X_1) + \overline{\Phi}(X_2) = \Phi(X_1 + X_2)$,
Bl. 90 also etwa für ein Intervall $I = [a, b)$ und $X \subseteq I$ | $\underline{\Phi}(X) = \Phi(I) - \overline{\Phi}(I - X)$, so dass in (7) an Stelle von $M$ ein Intervall oder eine messbare Menge $\supseteq X$ stehen kann.

Für $X \subseteq I$ hängen $\overline{\Phi}(X)$ und $\underline{\Phi}(X)$ nur vom Verlauf der Function $\varphi(x)$ in $I$ ab. Denn bei der Berechnung von $\overline{\Phi}(X)$ nach (5) braucht man auch nur Mengen $A \subseteq I$ zu benutzen (indem man $A$ durch $A\,I$ ersetzt), und $\Phi(A)$ hängt nur von $\varphi(x|I)$ ab ($\varphi(b) = \varphi(b - 0)$ ist ja ebenfalls nur davon abhängig), also auch $\overline{\Phi}(X)$ und $\underline{\Phi}(X) = \Phi(I) - \overline{\Phi}(I - X)$. Man kann also bei Bestimmung von $\overline{\Phi}(X)$ und $\underline{\Phi}(X)$ für $X \subseteq I$ die Function $\varphi(x)$ etwa durch $\varphi^*(x) = \varphi(a)$, $\varphi(x)$, $\varphi(b)$ in $(-\infty, a)$, $[a, b)$, $[b, \infty)$ ersetzen. – Diese Modification gestattet auch den Fall zu behandeln, dass die Grenzwerthe $\varphi(-\infty)$, $\varphi(+\infty)$ nicht endlich sind: man bestimmt, um nur den Fall der Messbarkeit zu erwähnen, $\Phi(X\,I)$ mit Hülfe der Function $\varphi^*$ und setzt $\Phi(X) = \lim\limits_{\substack{b \to \infty \\ a \to -\infty}} \Phi(XI)$, falls dieser Limes existirt. So wird im Falle $\varphi = x$ das Lebesguesche Längenmass definirt. Wir nennen auch unser System $\mathfrak{M}$ das *Lebesguesche System* der (nach $\varphi$) messbaren Mengen.

Das *kleinste* abgeschlossene System $\mathfrak{M}_0$ (= dem Durchschnitt aller), das die Intervalle $I$ enthält, heisst das *Borelsche System*, seine Mengen die *Borelschen Mengen*. $\mathfrak{M}$ ist in der Regel umfassender als $\mathfrak{M}_0$. Um genauer den Bau des
Bl. 91 Borelschen Systems zu übersehen (das nicht von irgend einer | Function $\varphi(x)$ abhängt), bezeichnen wir, wenn $X$ eine Klasse von Mengen durchläuft, mit $X_\sigma = X_1 \cup X_2 \cup \cdots$ die Summe und mit $X_\delta = X_1 X_2 \ldots$ den Durchschnitt einer Folge von Mengen $X$. Zu den Borelschen Mengen gehören:

die Intervalle $I$
die Intervallsummen $A = I_\sigma$
die $A$-Durchschnitte $B = A_\delta = I_{\sigma\delta}$
die $B$-Summen $C = B_\sigma = A_{\delta\sigma} = I_{\sigma\delta\sigma}$
. . .

dann aber weiter, wenn etwa $P$ alle diese durch endliche Wiederholung des $\sigma$- und $\delta$-Processes entstehenden Mengen bedeutet,

die $P$-Summen $Q = P_\sigma$,
die $Q$-Durchschnitte $R = Q_\delta = P_{\sigma\delta}$ u.s.w.

(die präcise Beschreibung dieses unbegrenzt fortzusetzenden Verfahrens geschieht mit Hülfe der Cantorschen Ordnungszahlen). Obwohl wir hier nur $\sigma$- und $\delta$-Process angewandt haben, ist doch auch die Complementbildung schon berücksichtigt, denn $M - I$ ist ein $A$, $M - I_\sigma$ ein $A_\delta$, u. s. w. Zu den $A$ gehören die *offenen* Mengen $G$; denn die in $G$ enthaltenen *rationalen* Intervalle $[\alpha, \beta)$ (d. h. $\alpha, \beta$ rational) bilden eine Folge und $G$ ist ihre Summe. Die *abgeschlossenen* Mengen $F$, die Complemente der $G$, gehören zu den $A_\delta$.

| Zu jeder Menge $X$ giebt es Borelsche Mengen $B, C$ mit $B \supseteq X \supseteq C$, $\Phi(B) =$ Bl. 92
$\overline{\Phi}(X)$, $\Phi(C) = \underline{\Phi}(X)$. ($B$ eine massgleiche Hülle, $C$ ein massgleicher Kern von $X$). Denn wählt man nach (5) ein $A_n \supseteq X$ mit $\Phi(A_n) < \overline{\Phi}(X) + \frac{1}{n}$, so hat $B = A_1 A_2 \ldots$ (ein $A_\delta$) die geforderten Eigenschaften. $C$ kann als Complement der Hülle von $M - X$, also als $M - A_\delta$ oder $A_{\delta\sigma}$ gewählt werden. Man kann auch die $A_n$ (indem man ihre Intervalle nach links zu etwas grösseren *offenen* Intervallen erweitert) als offene Mengen $G$, $B$ als $G_\delta$, $C$ als $F_\sigma$ annehmen.

## *Messbare Functionen.*

Ist $y = f(x)$ eine für alle reellen $x$ definirte reelle eindeutige (nicht nothwendig umkehrbar eindeutige) Function, so entspricht jeder Menge $Y$ auf der $y$-Achse als ihr *Urbild* die Menge $X$ aller derjenigen $x$, für die $y = f(x)$ zu $Y$ gehört. Sind $X_1, X_2, \ldots$ die Urbilder von $Y_1, Y_2, \ldots$ (es kann sich um beliebig viele Mengen handeln, endlich oder abzählbar oder unabzählbar viele), so ist $X_1 \cup X_2 \cup \cdots$ das Urbild von $Y_1 \cup Y_2 \cup \cdots$, $X_1 X_2 \cdots$ das Urbild von $Y_1 Y_2 \cdots$; sind $Y_1, Y_2$ fremd, so auch $X_1, X_2$, und $X = X_1 + X_2$ ist das Urbild von $Y = Y_1 + Y_2$; ist $Y \supseteq Y_1$ und $X, X_1$ die Urbilder, so ist $X - X_1$ das Urbild von $Y - Y_1$. Wenn daher auf der $x$-Achse ein abgeschlossenes Mengensystem $\mathfrak{M}$ gegeben ist, so
bilden auf der $y$-Achse die Mengen $Y$, | deren Urbilder $X$ zu $\mathfrak{M}$ gehören, wieder Bl. 93
ein abgeschlossenes Mengensystem $\mathfrak{N}$. Nennen wir die Mengen von $\mathfrak{M}$ messbar und die *Function $f(x)$ messbar*, wenn zu $\mathfrak{N}$ die *Intervalle* gehören (inclusive Nullmenge und $y$-Gerade $N$; eo ipso gehören dann auch Nullmenge und $x$-Achse $M$ zu $\mathfrak{M}$.) Es genügt zu fordern, dass die offenen Halbgeraden $(-\infty, y)$ zu $\mathfrak{N}$ gehören, d. h. dass die Menge $[f < y]$ derjenigen $x$, für die $f(x) < y$, für jedes $y$ messbar sei. Dann gehören auch die abgeschlossenen Halbgeraden $(-\infty, y]$, die Intervalle mit und ohne Endpunkt, die Borelschen Mengen $Y$ zu $\mathfrak{N}$, ihre Urbilder sind messbar.

Mit $f(x)$ ist auch $kf(x)$, $|f(x)|$, $f(x)^2$ u. dgl. messbar, z. B. ist $[|f| < y] = [-y < f < y]$ für $y > 0$, sonst die Nullmenge.

*Summe und Differenz von zwei messbaren Functionen ist messbar.* Ist an einer Stelle $f_1(x) + f_2(x) < y$, so giebt es eine rationale Zahl $r$ mit $f_1(x) < r <$

$y - f_2(x)$ und umgekehrt. Hieraus folgt

$$[f_1 + f_2 < y] = \bigcup_r [f_1 < r][f_2 < y - r]$$

als messbare Menge. *Das Produkt von zwei messbaren Functionen ist messbar*:

$$f_1 f_2 = \left(\frac{f_1 + f_2}{2}\right)^2 - \left(\frac{f_1 - f_2}{2}\right)^2.$$

Bl. 94 | Die (als überall endlich vorausgesetzte) *untere Grenze* $f(x) = \inf f_n(x)$ einer Folge messbarer Functionen ist messbar, nämlich

$$[f < y] = \bigcup_n [f_n < y].$$

Ebenso die *obere Grenze* $F = \sup f_n = -\inf(-f_n)$, wo man erhält

$$[F \leq y] = \bigcap_n [f_n \leq y];$$

will man die Mengen $[F < y]$ durch die Mengen $[f_n < y]$ ausdrücken, so findet man zunächst

$$[F < y] \subseteq \bigcap_n [f_n < y] \subseteq [F \leq y]$$

und wenn man hierin $y$ durch $y - \frac{1}{m}$ $(m = 1, 2, \ldots)$ ersetzt und nach $m$ summirt

$$[F < y] = \bigcup_m \bigcap_n \left[f_n < y - \frac{1}{m}\right].$$

Die Mengen $[f < y]$ und $[F < y]$ gehen also aus den $[f_n < y]$ durch den $\sigma$- und $\delta$-Process hervor. Wiederholte Anwendung von inf und sup giebt $\underline{\lim}, \overline{\lim}$ und lim; ist z. B. $g = \underline{\lim}\, f_n$ (überall endlich), so bilden die Functionen $g_n = \inf[f_n, f_{n+1}, \ldots]$ eine aufsteigende Folge mit $g = \lim g_n = \sup g_n$. *Der Limes* $f = \lim f_n$ *einer convergenten Folge messbarer Functionen ist messbar*, und die Mengen $[f < y]$ gehen aus den $[f_n < y]$ durch wiederholten $\sigma$- und $\delta$-Process
hervor. – Für *stetiges* $f$ sind die Mengen $[f < y]$ offen; für die *Baireschen* |
Bl. 95 *Functionen* (d. h. die aus stetigen Functionen durch wiederholte Limesbildung
entstehen) sind es *Borelsche Mengen*. Stetige und Bairesche Functionen sind messbar, wenn auch $\mathfrak{M}$ die Intervalle enthält.

Eine Function $f$, die nur endlich oder abzählbar viele Werthe $y_1, y_2, \ldots$ annimmt, heisst eine *Skalenfunction*. Zu ihrer Messbarkeit ist nothwendig und *hinreichend*, dass alle Mengen $X_n = [f = y_n]$ messbar seien. Hinreichend, weil $[f < y] = \sum_{y_n < y} X_n$. Summe und Differenz von zwei messbaren Skalenfunctionen ist wieder eine.

*Jede messbare Function ist gleichmässiger Limes messbarer Skalenfunctionen.*

Man bilde eine Theilung der $y$-Achse: $\cdots < y_{-2} < y_{-1} < y_0 < y_1 < y_2 < \cdots$, für $n \to \pm\infty \quad y_n \to \pm\infty$, und setze $\bar{f} = y_n$ (oder $\underline{f} = y_{n-1}$) in $X_n = [y_{n-1} \leq f < y_n]$. Ist die Scala mit Intervallen $y_n - y_{n-1} \leq \varepsilon$ gewählt, so ist $0 < \bar{f} - f \leq \varepsilon$ oder $0 \leq f - \underline{f} < \varepsilon$.

*Momente oder Stieltjesintegrale.* $\mathfrak{M}$ sei nun wieder das Lebesguesche System, das zu $\varphi(x)$ gehört, mit der additiven Mengenfunction $\Phi(X)$. Wir werden dann naturgemäss das *Moment einer messbaren Skalenfunction*, die die Werthe $y_m$ in den Mengen $X_m$ annimmt, durch die Reihe

$$\int f \, d\varphi = \sum y_m \, \Phi(X_m) \tag{14}$$

definiren, die als convergent vorausgesetzt werden muss, aber alsbald als *absolut convergent* vorausgesetzt werden soll. Für eine Skalenfunction mit nur endlich vielen Werthen ist das ja das Moment einer elementaren Vertheilung. | Die integralmässige Bezeichnung $\int f \, d\varphi$ lässt zunächst die Abhängigkeit von $f$ und der zu Grunde liegenden Vertheilungsfunction $\varphi$ hervortreten. Integrale dieser Art (mit einem monotonen $\varphi(x)$ statt $x$) hat zuerst T. J. Stieltjes [4] eingeführt, allerdings mehr im Anschluss an den gewöhnlichen Riemannschen Integralbegriff, während diese hier dem Lebesgueschen Integralbegriff als Vorbild folgen (Stieltjes-Lebesguesche Integrale). Ausführlicher wäre $\int f \, d\varphi = \int f(x) \, d\varphi(x)$ zu schreiben. Eine Scalenfunction, für die das Integral existirt, d. h. die Reihe absolut convergirt, heisst *integrabel*. Z. B. ist jede beschränkte messbare Skalenfunction, $|f| \leq C$, integrabel und Bl. 96

$$\left| \int f \, d\varphi \right| \leq C \sum \Phi(X_m) = C \, \Phi(M) = C\mu.$$

Die Summe von zwei integrablen Skalenfunctionen ist wieder integrabel und zwar

$$\int (f+g) \, d\varphi = \int f \, d\varphi + \int g \, d\varphi \tag{15}$$

Denn nimmt $f, g, f+g$ die Werthe $y_m, y_n, y_p$ in den Mengen $X_m, X_n, X_p$ an, wo $X_p = \sum\limits_{y_m + y_n = y_p} X_m X_n$, so folgt aus der absoluten Convergenz (ohne die das nicht richtig wäre) wegen $X_m = X_m M = \sum\limits_n X_m X_n$

$$\int f \, d\varphi = \sum_m y_m \Phi(X_m) = \sum_m y_m \sum_n \Phi(X_m X_n) = \sum_{m,n} y_m \Phi(X_m X_n)$$

| mit beliebiger Summationsanordnung für die Doppelreihe, ebenso $\int g \, d\varphi$, also Bl. 97

---

[4] *Recherches sur les fractions continues*, Ann. Fac. Toulouse 8 (1894), S. 1-122, 9 (1895), S. 1-47.

$$\int f\,d\varphi + \int g\,d\varphi = \sum_{m,n} (y_m + y_n)\,\Phi(X_m X_n) = \sum_p y_p\,\Phi(X_p) = \int (f+g)\,d\varphi.$$

Zwei messbare Scalenfunctionen mit beschränkter Differenz sind gleichzeitig integrabel oder nicht.

Nennen wir übrigens zwei Functionen, deren Differenz beschränkt ist, *benachbart*. Jede beliebige messbare Function $f$ lässt sich nun als gleichmässiger Limes messbarer Scalenfunctionen $f_n$ darstellen: $|f - f_n| \leq \varepsilon_n$, $\varepsilon_n \to 0$. Alle $f_n$ sind benachbart, und ist eine unter ihnen integrabel, so sind es alle. In diesem Fall ist

$$\left|\int f_n\,d\varphi - \int f_m\,d\varphi\right| = \left|\int (f_n - f_m)\,d\varphi\right| \leq (\varepsilon_m + \varepsilon_n)\mu$$

und dies ist $< \varepsilon$, falls $m$ hinlänglich gross und $n > m$ ist. D. h. $\int f_n\,d\varphi$ convergirt nach einem Grenzwerth, und diesen *definiren* wir als

$$\int f\,d\varphi = \lim \int f_n\,d\varphi. \tag{16}$$

Er ist von der Wahl der approximirenden Folge $f_n$ unabhängig, wie unmittelbar einzusehen. Wir gelangen so zu der Definition:

*Die messbare Function $f(x)$ heisst integrabel, wenn es eine zu ihr benachbarte integrable Scalenfunction giebt; ist dann $f_n(x)$ eine gleichmässig nach $f(x)$ convergente Folge messbarer (also integrabler) Scalenfunctionen, so werde $\int f\,d\varphi = \lim \int f_n\,d\varphi$ definirt.*

Bl. 98 | Jede beschränkte messbare Function, $|f| \leq C$, ist integrabel und

$$\left|\int f\,d\varphi\right| \leq C\mu.$$

*Mit $f$ ist auch $|f|$ integrabel* und

$$\left|\int f\,d\varphi\right| \leq \int |f|\,d\varphi\,, \tag{17}$$

wie aus den approximirenden Scalenfunctionen folgt. Mit $f, g$ ist auch $f + g$ integrabel und es gilt wieder (15): Grenzübergang von den approximirenden Scalenfunctionen. Ebenso ist

$$\int (f-g)\,d\varphi = \int f\,d\varphi - \int g\,d\varphi\,, \quad \int cf\,d\varphi = c\int f\,d\varphi.$$

Benachbarte messbare Functionen sind gleichzeitig integrabel oder nicht. Wenn die integrablen Functionen $f_n$ gleichmässig nach $f$ convergiren, ist auch $f$ integrabel und es gilt (16).

Als Integrabilitätskriterium für $f$ (diese als messbar vorausgesetzt) kann die Existenz von $\int g\,d\varphi$ für irgend eine benachbarte Scalenfunction dienen. Nehmen wir z. B. eine *äquidistante* Scala $y_n = n\varepsilon$ ($\varepsilon > 0$, $n = 0, \pm 1, \ldots$); in $X_n =$

$[(n-1)\varepsilon < f \leq n\varepsilon]$ sei $g = n\varepsilon$, so ist $0 \leq g - f < \varepsilon$ und die absolute Convergenz von

$$\int g\,d\varphi = \varepsilon \sum_{-\infty}^{\infty} n\;\Phi(X_n)$$

ist zur Integrabilität von $f$ nothwendig und hinreichend; zugleich ist | dann Bl. 99

$$0 \leq \int g\,d\varphi - \int f\,d\varphi \leq \varepsilon\mu \tag{18}$$

(Als Kriterium genügt die Reihenconvergenz für ein festes $\varepsilon$, z. B. $\varepsilon = 1$). Ist $M(y) = [f > y]$ und zur Abkürzung $\mu(y) = \Phi(M(y))$, so ist

$$\int g\,d\varphi = \varepsilon \sum_{-\infty}^{\infty} n[\mu(\overline{n-1}\varepsilon) - \mu(n\varepsilon)]$$

Besonders einfach wird dies für $f \geq 0$; dann ist für $y < 0 \quad M(y) = M, \;\; \mu(y) = \mu$, so dass die Glieder für $n = 0, -1, -2, \ldots$ verschwinden und

$$\int g\,d\varphi = \varepsilon \sum_{1}^{\infty} n[\mu(\overline{n-1}\varepsilon) - \mu(n\varepsilon)]$$

wird, was sich zu

$$\int g\,d\varphi = \varepsilon \sum_{0}^{\infty} \mu(n\varepsilon) \tag{19}$$

umformen lässt in dem Sinne, dass die beiden Reihen gleichzeitig convergiren oder divergiren und im ersten Fall dieselbe Summe haben.
(Die Zahlen $u_n = \mu(n\varepsilon)$ convergiren abnehmend nach 0. Man hat nun

$$\begin{aligned} s_n &= u_0 + u_1 + \cdots + u_{n-1} \geq u_0 + \cdots + u_{n-1} - nu_n \\ &= (u_0 - u_1) + 2(u_1 - u_2) + \cdots + n(u_{n-1} - u_n) = \sigma_n, \end{aligned}$$

andererseits für $p > n$ | Bl. 100

$$\begin{aligned} \sigma_p &\geq (u_0 - u_1) + \cdots + n(u_{n-1} - u_n) + n(u_n - u_{n+1}) + \cdots + n(u_{p-1} - u_p) \\ &= u_0 + \cdots + u_{n-1} - nu_p = s_n - nu_p\,, \end{aligned}$$

$$\sigma_n \leq s_n \leq \sigma_p + nu_p.$$

Convergirt $s_n \to s$, so auch $\sigma_n \to \sigma \leq s$. Und convergirt $\sigma_p \to \sigma$, so ist auch (wegen $u_p \to 0$) $s_n \leq \sigma$, $s_n \to s \leq \sigma$. Mit der einen Reihe convergirt auch die andere und hat dieselbe Summe).

Die Convergenz von $\sum\limits_{0}^{\infty} \mu(n\varepsilon)$, z. B. $\sum\limits_{0}^{\infty} \mu(n)$, wo $\mu(n) = \Phi[f > n]$, ist also zur Integrabilität von $f \geq 0$ ($f$ messbar vorausgesetzt) nothwendig und hinreichend. Beiläufig ist $\int f\,d\varphi$ als Limes dieser Summe (19) für $\varepsilon \to 0$ durch das gewöhnliche Riemannsche Integral

$$\int f\,d\varphi = \int_0^\infty \mu(y)dy$$

darstellbar.

Aus diesem Kriterium folgt: ist $0 \leq g(x) \leq f(x)$, beide messbar, so ist mit $f$ auch $g$ integrabel. (Denn $[g > y]$ ist Theilmenge von $[f > y]$, die $f$-Reihe für die $g$-Reihe eine Majorante).

Ist $f$ messbar, so ist mit $|f|$ auch $f$ integrabel; denn die beiden Functionen $f_1 = \frac{1}{2}(|f| + f)$; $f_2 = \frac{1}{2}(|f| - f)$ sind $\geq 0$ und $\leq |f|$, also integrabel und ebenso $f = f_1 - f_2$. Umgekehrt war mit $f$ auch $|f|$ integrabel. Jede integrable Function ist Differenz von zwei nicht-negativen integrablen Functionen.

Bl. 101 | Ist $f$ integrabel, $g$ messbar und $|g| \leq |f|$, so ist auch $g$ integrabel. Folgender *Satz über monotone Folgen* ist wichtig:
I. *Es sei* $f_1 \leq f_2 \leq \ldots,\quad f_m \to f$, *die* $f_m$ *integrabel. Wenn* $\int f_m\,d\varphi$ *einen Grenzwerth hat, so ist* $f$ *integrabel und umgekehrt, und zwar ist dann* $\int f\,d\varphi = \lim \int f_m d\varphi$.

Zunächst ist $f$ messbar. Wir können $f_m \geq 0$ annehmen (indem wir $f_m - f_1$ betrachten). Sei wieder $M(y) = [f > y]$, $M_m(y) = [f_m > y]$ und $\mu(y)$, $\mu_m(y)$ die $\Phi$-Werthe dieser Mengen. Die $M_1(y), M_2(y), \ldots$ bilden eine aufsteigende Mengenfolge mit der Summe $M(y)$, also $\lim\limits_m \mu_m(y) = \mu(y)$. Auf Grund von (18)(19) haben wir für jedes $m$

$$\varepsilon \sum_0^\infty \mu_m(n\varepsilon) \leq \int f_m\,d\varphi + \varepsilon\mu.$$

Sei nun $\int f_m\,d\varphi \to \lambda$, so ist

$$\varepsilon \sum_0^\infty \mu_m(n\varepsilon) \leq \lambda + \varepsilon\mu,$$

$$\text{also erst recht} \qquad \varepsilon \sum_0^p \mu_m(n\varepsilon) \leq \lambda + \varepsilon\mu \ \text{ für jedes } p,$$

$$\text{daraus für } m \to \infty \qquad \varepsilon \sum_0^p \mu(n\varepsilon) \leq \lambda + \varepsilon\mu$$

$$\text{und wieder für } p \to \infty \qquad \varepsilon \sum_0^\infty \mu(n\varepsilon) \leq \lambda + \varepsilon\mu,$$

d. h. aber, $f$ ist integrabel und zugleich, wegen (18)(19):

$$\int f\,d\varphi \leq \lambda + \varepsilon\mu\,,$$

d. h. $\int f\,d\varphi \leq \lambda$.

| Ist umgekehrt $f$ integrabel, so ist $\int f_m\,d\varphi \leq \int f\,d\varphi$, der Grenzwerth $\lambda$ existirt und zwar $\lambda \leq \int f\,d\varphi$. I. überträgt sich sofort auf absteigende Folgen. Ferner folgt aus ihm unmittelbar:

II. Es sei $f_m \geq 0$ und integrabel, $f = f_1 + f_2 + \cdots$. Wenn die Reihe $\sum\limits_1^\infty \int f_m\,d\varphi$ convergirt, so ist $f$ integrabel und umgekehrt, und zwar ist $\int f\,d\varphi = \sum\limits_1^\infty \int f_m\,d\varphi$.

*Integrale über messbare Mengen.* Sei $f(x)$ in $M$ definirt, $X$ eine messbare Menge. Wir setzen dann

$$g = f \ \text{ in } X,\ g = 0 \ \text{ in } M - X$$

und erklären, falls $g$ integrabel ist,

$$\int_X f\,d\varphi = \int g\,d\varphi\ ; \tag{20}$$

$f$ heisst dann in $X$ integrabel.[5] (Die bisherigen Integrale waren also solche über die ganze Gerade $M$: $\int f\,d\varphi = \int_M f\,d\varphi$). Das ist gewiss der Fall, wenn $f$ schlechthin (in $M$) integrabel ist; denn | Bl. 103

$$[g < y] = [f < y]\,X \ \text{ für } y \leq 0$$

$$[g < y] = [f < y]\,X + (M - X) \ \text{ für } y > 0,$$

also ist $g$ messbar, überdies durchweg $|g| \leq |f|$. Ist $f$ in $X$ integrabel, $X_1$ messbare Theilmenge von $X$, so ist $f$ auch in $X_1$ integrabel, denn

$$\int_{X_1} f\,d\varphi = \int_{X_1} g\,d\varphi$$

und $g$ ist in $M$, also auch in $X_1$ integrabel. Ist ferner $X = X_1 + X_2$ (lauter messbare Mengen) und $f$ in $X_1, X_2$ integrabel, so ist es auch in $X$ integrabel und

$$\int_X f\,d\varphi = \int_{X_1} f\,d\varphi + \int_{X_2} f\,d\varphi,$$

denn für die bezüglichen Functionen $g$, $g_1$, $g_2$ ist $g = g_1 + g_2$. Dies lässt sich aber auf abzählbar viele Summanden ausdehnen, nämlich:

III. Sei $X = X_1 + X_2 + \cdots$ Summe paarweise fremder abzählbar vieler messbarer Mengen; $f \geq 0$ sei in jedem $X_m$ integrabel. Wenn die Summe $\sum \int\limits_{X_m} f\,d\varphi$ convergirt, so ist $f$ integrabel und umgekehrt, und zwar

$$\int_X f\,d\varphi = \int_{X_1} f\,d\varphi + \int_{X_2} f\,d\varphi + \cdots \tag{21}$$

---

[5] Speziell für $f = 1 : \int_X d\varphi = \Phi(X)$ ($g$ Skalenfunction).

Denn für die betreffenden $g$-Functionen gilt $g = g_1 + g_2 + \cdots$ mit $g_m \geq 0$, und die Behauptung folgt aus II.

Bl. 104 | Die Formel (21) gilt auch noch für ein $f$ beliebigen Vorzeichens, das in $X$ integrabel ist (Zerlegung $f = f_1 - f_2$ in zwei integrable Functionen $\geq 0$). Dass $f$ in $X$ integrabel ist, kann allerdings nicht aus der (selbst absoluten) Convergenz von $\sum \int\limits_{X_m} f\, d\varphi$, wohl aber aus der von $\sum \int\limits_{X_m} |f|\, d\varphi$ geschlossen werden.

Für ein in $M$ integrables $f$ definirt also

$$\Psi(X) = \int_X f\, d\varphi \tag{22}$$

selbst wieder eine *additive Mengenfunction* im System der messbaren Mengen ($\geq 0$ für $f \geq 0$, sonst Differenz nichtnegativer Mengenfunctionen). Es gilt also

$$\Psi(X) = \lim \Psi(X_n) \tag{23}$$

für die Summe $X = X_1 \cup X_2 \cup \cdots$ einer aufsteigenden und den Durchschnitt [4] $X = X_1 X_2 \cdots$ einer absteigenden Mengenfolge. $(*)$

Eine Menge $X$ mit $\Phi(X) = 0$ heisse eine *Nullmenge*, eine Function $f$, die bis auf eine Nullmenge verschwindet, eine *Nullfunction.* Eine Nullfunction ist stets integrabel und hat das Integral $f\, d\varphi = 0$. Umgekehrt, ist $f \geq 0$ und $\int f\, d\varphi = 0$, so ist $f$ eine Nullfunction. Denn ist $M(y) = [f > y]$ und $\mu(y) = \Phi(M(y))$, so ist für jedes $\varepsilon > 0$

$$0 = \int f\, d\varphi \geq \int_{M(\varepsilon)} f\, d\varphi \geq \varepsilon\, \mu(\varepsilon),$$

Bl. 105 | $\mu(\varepsilon) = 0$ und folglich $\mu(0) = \lim \mu(\frac{1}{n}) = 0$, da $M(\frac{1}{n})$ aufsteigend nach $M(0)$ convergirt; $[f > 0]$ ist eine Nullmenge.

Für jede beliebige Function ist $\int_X f\, d\varphi = 0$, wenn $X$ eine Nullmenge ist. Für eine absteigende Mengenfolge, deren Durchschnitt eine Nullmenge ist, ist also nach (23)

$$\int_{X_n} f\, d\varphi \to 0. \tag{24}$$

Das gilt z. B. für $X_n = [f > y_n]$ mit $y_1 < y_2 < \cdots \to \infty$, d. h. für jede integrable Function convergirt $\int\limits_{[f>y]} f\, d\varphi \to 0$ mit $y \to \infty$, ebenso $\int\limits_{[f<-y]} f\, d\varphi$.

Man kann zu jedem $\varepsilon > 0$ ein $\delta > 0$ bestimmen, so dass mit $\Phi(X) < \delta$ auch $\int_X f\, d\varphi < \varepsilon$ ($f \geq 0$ angenommen; im allgemeinen Fall betrachte man $|f|$). Denn setzt man $U = [f \leq y]$, $V = [f > y]$, so ist

$$\int_X f\, d\varphi = \int_{XU} f\, d\varphi + \int_{XV} f\, d\varphi \leq y\, \Phi(X) + \int_V f\, d\varphi$$

und hier kann man $y$ so gross wählen, dass $\int_V f\, d\varphi < \frac{\varepsilon}{2}$, sodann $\Phi(X) < \frac{\varepsilon}{2y}$. (24) gilt also stets für $\Phi(X_n) \to 0$.

| Sei $f \geq 0$ in $M$ integrabel; zu der nichtnegativen Mengenfunction (22) gehört wieder eine Punctfunction

$$\psi(x) = \Psi(-\infty, x) \tag{25}$$

die monoton, links stetig ist und Grenzwerthe $\psi(-\infty) = 0$, $\psi(\infty) = \int f\, d\varphi$ hat (alles wegen der Additivität). Man kann damit das Lebesguesche System der nach $\psi$ messbaren Mengen $X$ mit der additiven Mengenfunction $\Omega(X) = \int_X d\psi$ definiren; es ist leicht zu sehen, dass die nach $\varphi$ messbaren Mengen auch nach $\psi$ messbar sind und für sie $\Omega(X) = \Psi(X)$, d. h.

$$\int_X d\psi = \int_X f\, d\varphi \tag{26}$$

ist. Denn zunächst ist für Halbgeraden $(-\infty, x)$, demnach für die Intervalle und Borelschen Mengen die Gleichung $\Omega = \Psi$ erfüllt. Ist $X$ nach $\varphi$ messbar, so kann man Borelsche Mengen $B, C$ mit $B \supseteq X \supseteq C$, $\Phi(B) = \Phi(X) = \Phi(C)$ bilden, $B - C$ ist eine Nullmenge, also $\int_{B-C} f\, d\varphi = 0$, $\Psi(B) = \Psi(C)$ oder $\Omega(B) = \Omega(C)$, d. h. wegen $\Omega(B) \geq \bar{\Omega}(X) \geq \underline{\Omega}(X) \geq \Omega(C)$ | ist $X$ nach $\psi$ Bl. 107
messbar und $\Omega(X) = \Psi(B) = \Psi(C) = \Psi(X)$.

Allgemeiner gilt, wenn $g(x)$ nach $\varphi$ messbar ist:

$$\int g\, d\psi = \int fg\, d\varphi \tag{27}$$

in dem Sinne, dass mit dem einen Integral auch das andere existirt und ihm gleich ist. Wir können $g \geq 0$ annehmen (im allgemeinen Falle zerlege man $g = g_1 - g_2$, $g_1, g_2 = \frac{1}{2}|g| \pm \frac{1}{2}g$). Ist $g$ zunächst eine Scalenfunction, die die Werthe $y_n$ in den (nach $\varphi$) messbaren Mengen $X_n$ annimmt, so ist, wenn $g$ nach $\psi$ integrabel ist, nach (26) und III

$$\int g\, d\psi = \sum y_n \Psi(X_n) = \sum y_n \int_{X_n} f\, d\varphi = \sum \int_{X_n} gf\, d\varphi = \int gf\, d\varphi$$

und ebenso umgekehrt, wenn $fg$ nach $\varphi$ integrabel ist. – Ist sodann $g \geq 0$ nach $\varphi$ messbar, $h$ eine nach $\varphi$ messbare Scalenfunction $\geq 0$ mit $|g - h| \leq \varepsilon$, also $|gf - hf| \leq \varepsilon f$, so sind $g - h$ nach $\psi$, $gf - hf$ nach $\varphi$ integrabel, also $g$ und $h$ gleichzeitig nach $\psi$ integrabel oder nicht, $gf$ und $hf$ gleichzeitig nach $\varphi$ integrabel oder nicht. Aus

$$\int |g - h|\, d\psi \leq \varepsilon \Psi(M), \quad \int |gf - hf|\, d\varphi \leq \varepsilon \int f\, d\varphi = \varepsilon \Psi(M)$$

folgt dann im Integrabilitätsfalle, wo $\int h\, d\psi = \int hf\, d\varphi$,

$$\left| \int g\, d\psi - \int gf\, d\varphi \right| \leq 2\varepsilon \Psi(M), \ \text{ d. h. (27).}$$

*Oberes und unteres Integral.* $f$ lasse sich zwischen integrable Functionen $g, G$ einschliessen: $g \leq f \leq G$. Es ist dann stets $\int g\,d\varphi \leq \int G\,d\varphi$ und die beiden Zahlen

$$t = \sup \int g\,d\varphi, \quad T = \inf \int G\,d\varphi, \quad t \leq T$$

heissen *unteres und oberes* Integral: $t = \underline{\int} f\,d\varphi$, $T = \overline{\int} f\,d\varphi$. Ist $f$ selbst integrabel, also zugleich ein $g$ und $G$, so ist $t = T = \int f\,d\varphi$. Umgekehrt: *ist* $t = T$, *so ist* $f$ *integrabel.* Man bestimme nämlich eine Folge integrabler Functionen $g_n \leq f$ mit $\int g_n d\varphi \to t$; hierbei kann man $g_1 \leq g_2 \leq \cdots$ annehmen, indem man $g_n$ durch $\max[g_1, \ldots, g_n]$ ersetzt. ($\max[g_1, g_2] = \frac{1}{2}(g_1 + g_2) + \frac{1}{2}|g_1 - g_2|$ ist mit $g_1, g_2$ integrabel). Es ist dann $g_n \to g \leq f$ und $g$ nach I. integrabel mit $\int g\,d\varphi = t$; ebenso existirt eine integrable Function $G \geq f$ mit $\int G\,d\varphi = T$. Für $t = T$ ist $\int (G - g)\,d\varphi = 0$, $G - g$ eine Nullfunction; und da $G - g = (G - f) + (f - g)$, so sind auch $G - f$, $f - g$ Nullfunctionen, demnach integrabel mit Integral 0, also $f$ integrabel und $\int f\,d\varphi = \int G\,d\varphi = \int g\,d\varphi$.

Bl. 109 *Intervallintegrale.* Ist $f$ integrabel, so convergirt das Intervallintegral $\int_I f\,d\varphi$ $(I = [a, b))$ nach $\int f\,d\varphi$ für $b \to \infty$, $a \to -\infty$. Ist $f$ in jedem Intervall integrabel und $\int_I |f|\,d\varphi$ beschränkt, so ist $f$ über die ganze Gerade integrabel; man schliesst dies für $f \geq 0$ aus I., im Allgemeinen durch die Zerlegung $f = f_1 - f_2$.

Hier kann man wieder die Modification anbringen für den Fall, dass $\varphi(-\infty)$ oder $\varphi(+\infty)$ oder beide unendlich sind. $\int_I f\,d\varphi$ hängt nur vom Verlauf von $\varphi(x)$ in $I$ ab; man schliesst das aus der entsprechenden Thatsache über Mengen $\subseteq I$ durch Betrachtung von Skalenfunctionen. Demnach ist $\int_I f\,d\varphi = \int_I f\,d\varphi^*$ $(\varphi^*(x) = \varphi(a), \varphi(x), \varphi(b)$ in $(-\infty, a), [a, b), [b, \infty))$ und man definirt $\int f\,d\varphi = \lim \int_I f\,d\varphi$ für $b \to \infty, a \to -\infty$, vorausgesetzt, dass dieser Grenzwerth existirt – respective, um mit unseren bisherigen Betrachtungen in Einklang zu bleiben, dass $\int_I |f|\,d\varphi$ beschränkt ist. So ist für $\varphi = x$ das gewöhnliche Lebesguesche Integral $\int f\,dx$ zu erklären.

Die übrigen Intervall- und Halbgeradenintegrale erhält man durch Grenzwerthe, auf Grund der Additivität. Setzt man $F(x) = \int_{(-\infty, x)} f\,d\varphi$, so ist $(a < b)$

$$\int_{(a,b)} f\,d\varphi = F_a^b = F(b) - F(a), \quad \int_{[a,b]} f\,d\varphi = F_a^{b+0} = F(b+0) - F(a)$$

Bl. 110 und mit analogen Bezeichnungen

$$\int_{(a,b)} = F_{a+0}^b, \quad \int_{(a,b]} = F_{a+0}^{b+0}, \quad \int_{(-\infty,b)} = F_{-\infty}^b,$$

$$\int_{(-\infty,b]} = F_{-\infty}^{b+0}, \quad \int_{[a,\infty)} = F_a^\infty, \quad \int_{(a,\infty)} = F_{a+0}^\infty, \quad \int = F_{-\infty}^\infty$$

Die Grenzen $a - 0, b - 0$ sind durch $a, b$ zu ersetzen, weil $F(x)$ links stetig ist. Das Integral über einen Punkt ist $\int_{[a]} = F_a^{a+0} = f(a)\,\varphi_a^{a+0}$ (Skalenfunction!)

Wegen des Anschlusses an die ursprüngliche Stieltjessche Integraldefinition mag erwähnt werden, dass für eine monotone Function $\chi(x)$, die nicht links stetig zu sein braucht, die Integrale $\int f\,d\chi$ gleich den mit der entsprechenden linksstetigen Function $\varphi(x) = \chi(x-0)$ gebildeten $\int f\,d\varphi$ definirt werden; insbesondere gilt das auch für die zugehörige rechtsstetige Function $\psi(x) = \chi(x+0)$. Also

$$\int f\,d\chi = \int f\,d\varphi = \int f\,d\psi, \quad \int_X f\,d\chi = \int_X f\,d\varphi = \int_X f\,d\psi.$$

Z. B. ist $\int_{[a,b)} d\chi = \varphi_a^b = \chi_{a-0}^{b-0}$, das Integral über einen Punkt $\int_{[a]} f\,d\chi = f(a)\,\varphi_a^{a+0} = f(a)\,\chi_{a-0}^{a+0}$. Bei diesen Integralen spielen natürlich die Werthe $\chi(x)$ an den Sprungstellen keine Rolle, sondern immer nur die Grenzwerthe $\chi(x\pm 0)$. Dagegen definirt man jetzt auch, was | wir bisher nicht thaten, für $a < b$ [5] Bl. 112

$$\int_a^b f\,d\chi = f(a)\,\chi_a^{a+0} + \int_{(a,b)} f\,d\chi + f(b)\,\chi_{b-0}^b; \tag{28}$$

hierin treten auch $\chi(a), \chi(b)$ auf, so dass dieses $\int_a^b$ mit keinem der 4 Intervallintegrale übereinzustimmen braucht. (Z. B. $\int_a^b d\chi = \chi_a^b$, während $\chi_{a\pm 0}^{b\pm 0}$ die 4 Intervallintegrale liefert). Für $a < b < c$ ist

$$\int_a^b + \int_b^c = f(a)\,\chi_a^{a+0} + \int_{(a,b)} + \int_{[b]} + \int_{(b,c)} + f(c)\,\chi_{c-0}^c = \int_a^c .$$

Ist $\chi = \varphi$ links stetig, so ist $\int_a^b f\,d\varphi = \int_{[a,b)} f\,d\varphi$; ist $\chi = \psi$ rechts stetig, so $\int_a^b f\,d\psi = \int_{(a,b]} f\,d\psi$. Für $f \geq 0$ ist $\int_{(a,b)} \leq \int_a^b \leq \int_{[a,b]}$. Ist $f$ integrabel, so convergirt $(b \to \infty, a \to -\infty)$ $\int_a^b f\,d\chi$ nach $\int f\,d\chi$, und umgekehrt, wenn $\int_a^b f\,d\chi$ beschränkt ist, so ist $f$ integrabel. Für beliebige $f$ hat man $\int_a^b |f|\,d\chi$ beschränkt anzunehmen.

Für eine *stetige* (also messbare und in jedem Intervall beschränkte, also integrable) Function $f(x)$ kann man $\int_a^b f\,d\chi$ genau wie im gewöhnlichen Falle $\int_a^b f\,dx$ ermitteln (Riemannsches Integral). Es sei $a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$ eine Intervalltheilung, $m_i = \min f$ und $M_i = \max f$ in $[x_{i-1}, x_i]$, | $\omega_i = M_i - m_i$ die Schwankung. $\delta_i = \chi_{x_{i-1}}^{x_i}$ $(i = 1, \ldots, n)$. Es liegt dann Bl. 113 $\int_a^b f\,d\chi = \sum \int_{x_{i-1}}^{x_i} f\,d\chi$ zwischen der Untersumme $\sum m_i\delta_i$ und der Obersumme $\sum M_i\delta_i$, deren Unterschied $\sum \omega_i\delta_i$ durch Verkleinerung der Intervalle aber beliebig klein $(\omega_i \leq \varepsilon,\ \sum \omega_i\delta_i \leq \varepsilon\chi_a^b)$ gemacht werden kann; d. h. bei unbegrenzter Verkleinerung der Intervalle convergiren Untersumme wie Obersumme (wie auch die dazwischen liegenden $\sum f(\xi_i)\delta_i$, $\xi_i$ ein Werth aus $[x_{i-1}, x_i]$)nach $\int_a^b f\,d\chi$. | Ist $\chi(x) = \int_{-\infty}^x \vartheta(x)\,dx$ $\quad(\vartheta(x)$ stetig), so ist $\delta_i = \vartheta(\xi_i)(x_i - x_{i-1})$, Bl. 112v $\xi_i$ ein Wert in $[x_{i-1}, x_i]$, $\sum f(\xi_i)\delta_i = \sum f(\xi_i)\vartheta(\xi_i)(x_i - x_{i-1})$ konvergirt nach

$\int_a^b f(x)\vartheta(x)\,dx = \int_a^b f\,d\chi;$

$$\int f\,d\chi = \lim_{\substack{a\to-\infty\\ b\to\infty}} \int_a^b f(x)\vartheta(x)\,dx = \int_{-\infty}^{\infty} f(x)\vartheta(x)\,dx.$$

Bl. 113 | – Im Anschluss daran kann man für beschränkte Functionen auch unteres und oberes Riemannsches Integral definiren (Darbouxsche Integrale) und im Gleichheitsfalle das Stieltjes-Riemannsche Integral $\int_a^b f\,d\chi$, dessen Existenz die des Lebesgueschen (nicht umgekehrt) und die Gleichheit beider nach sich zieht; wir lassen dies beiseite, da dieser Integralbegriff respective der entsprechende Jordan-Peanosche Massbegriff für unseren Zweck nicht ausreicht; die messbaren Mengen bilden hier kein abgeschlossenes System. (Es gilt nur ($\beta$), nicht ($\gamma$). Die Grössen $\overline{\Phi}(X)$, $\underline{\Phi}(X)$ würden sich ergeben, wenn man $\overline{\Phi}(X) = \inf \Phi(E)$ für $E \supseteq X$ statt (5) definirt, $E$ endliche Intervallsumme).

Bl. 114 | Kehren wir zu links stetigen $\varphi(x)$ mit $\varphi(-\infty) = 0$, $\varphi(+\infty) = 1$ zurück. Sei $f(x)$ *stetig und beschränkt*, $|f| \leq M$, also $f(x+y)$ als Function von $x$ ebenfalls. Die Function

$$g(y) = \int_{-\infty}^{\infty} f(x+y)\,d\varphi(x) \tag{29}$$

(d. h. das Integral über die ganze Gerade erstreckt) ist wieder *beschränkt*, $|g| \leq M$; ausserdem ist sie *stetig*. Denn ($a < b$)

$$\begin{aligned} |g(y) - g(\eta)| &\leq \left[\int_{-\infty}^{a} + \int_a^b + \int_b^{\infty}\right] |f(x+y) - f(x+\eta)|\,d\varphi(x) \\ &\leq 2M(\varphi_{-\infty}^{a} + \varphi_b^{\infty}) + \int_a^b |f(x+y) - f(x+\eta)|\,d\varphi(x) \end{aligned}$$

und hier kann man erst $a, b$ so wählen, dass $\varphi_{-\infty}^{a} + \varphi_b^{\infty} < \varepsilon$, sodann wegen der gleichmässigen Stetigkeit im endlichen Intervall $y - \eta$ so klein, dass $|f(x+y) - f(x+\eta)| \leq \varepsilon$ für $a \leq x \leq b$, $\int_a^b |\ldots|\,d\varphi(x) \leq \varepsilon\,\varphi_a^b \leq \varepsilon$. Hat $f(x)$ ausserdem eine beschränkte stetige Ableitung $f'(x)$, so ist

$$g'(y) = \int_{-\infty}^{\infty} f'(x+y)\,d\varphi(x) \tag{30}$$

und also hat auch $g(y)$ eine beschränkte stetige Ableitung ($|g'| \leq M'$, wenn $|f'| \leq M'$). Denn es ist, wenn $g'(y)$ zunächst durch (30) und noch nicht als
Bl. 115 $\frac{dg(y)}{dy}$ erklärt wird, |

$$\begin{aligned} \frac{g(y) - g(\eta)}{y - \eta} - g'(y) &= \int \Big[\frac{f(x+y) - f(x+\eta)}{y - \eta} - f'(x+y)\Big]\,d\varphi(x) \\ &= \int [f'(x+\zeta) - f'(x+y)]\,d\varphi(x), \end{aligned}$$

wo $\zeta$ ein (auch von $x$ abhängiger) Werth zwischen $y$ und $\eta$ ist; dasselbe Raisonnement wie oben zeigt aber, dass für hinreichend kleines $y-\eta$ das Integral rechts beliebig klein wird, also $\lim_{\eta\to y}\frac{g(y)-g(\eta)}{y-\eta}=g'(y)$.

Hat $f(x)$ auch noch eine stetige und beschränkte zweite Ableitung $f''(x)$, so hat auch $g(y)$ eine und zwar

$$g''(y)=\int_{-\infty}^{\infty} f''(x+y)\,d\varphi(x)$$

($|g''|\leq M''$, wenn $|f''|\leq M''$) u.s.f.

## | § 6. Vertheilung eines Variablenpaares. Bl. 116

$z=(x,y)$ sei ein Variablenpaar oder ein Punkt in der Ebene $P$ mit den rechtwinkligen Coordinaten $x,y$. Unter den ebenen Punktmengen $Z$ sind die *Productmengen* $Z=(X,Y)$ bemerkenswerth; dies soll die Menge aller $z$ sein, wo $x$ die Menge $X$, $y$ die Menge $Y$ durchläuft (eine Art Productbildung, die auch mit $X\times Y$ bezeichnet wird; vom Durchschnitt zu unterscheiden.)
So ist $P=(M,N)$, wenn $M$ die $x$-Achse, $N$ die $y$-Achse ist. Ein Intervall $I=[\xi_0,\xi)$ der $x$-Achse mit einem Intervall $H=[\eta_0,\eta)$ der $y$-Achse liefert ein *ebenes Intervall*

$$K=(I,H)=(\xi_0\leq x<\xi,\ \ \eta_0\leq y<\eta),$$

d. h. ein Rechteck, zu dem von seiner Begrenzung nur die linke untere Ecke und die beiden darin zusammenstossenden Seiten, nicht die übrigen Ecken und Seiten zu rechnen sind. Hieraus entstehen Grenzfälle (mit dem obigen 16), wenn man für $I$ oder $H$ auch Halbgerade oder die ganze Gerade zulässt, d. h. $\xi_0$ oder $\eta_0$ durch $-\infty$, $\xi$ oder $\eta$ durch $+\infty$ ersetzt; wir führen nur die folgenden an: | Bl. 117

Halber Parallelstreifen: $(\xi_0\leq x<\xi,\ -\infty<y<\eta)$
Parallelstreifen: $(\xi_0\leq x<\xi,\ -\infty<y<\infty)$
Halbebene: $(-\infty<x<\xi,\ -\infty<y<\infty)$
Quadrant: $(-\infty<x<\xi,\ -\infty<y<\eta)$

Eine Zerlegung $I = I_1 + \cdots + I_m$, $H = H_1 + \cdots + H_n$ der linearen Intervalle liefert für $K = (I, H)$ eine *vollständige* Zerlegung

$$K = (I_1, H_1) + (I_1, H_2) + \cdots + (I_m, H_n).$$

Eine beliebige Zerlegung $K = K_1 + \cdots + K_p$ kann man durch Ausziehen aller Theilungslinien in eine vollständige Zerlegung von $K, K_1, \ldots K_p$ verwandeln.

Das kleinste abgeschlossene System $\mathfrak{P}_0$ ebener Mengen, das die $K$ enthält, heisst wieder das *Borelsche* System, seine Mengen die *ebenen Borelschen Mengen. Sind $A, B$ lineare Borelsche Mengen, so ist $(A, B)$ eine ebene Borelsche Menge.* Denn $\mathfrak{P}_0$ enthält ausser den $(I, H)$ die $(I_\sigma, H), (I_{\sigma\delta}, H), \ldots, (A, H)$, $(A, H_\sigma), (A, H_{\sigma\delta}), \ldots, (A, B)$. *Der Durchschnitt einer ebenen Borelschen Menge $C$ mit einer Geraden $D$ ist eine* | *lineare Borelsche Menge.* Denn $DK$ ist Bl. 119
[6] ein $I$ (d. h. ein lineares Intervall $[\alpha, \beta)$, eventuell 0), $DK_\sigma$ ein $I_\sigma$, $DK_{\sigma\delta}$ ein $I_{\sigma\delta}, \ldots, DC$ ein $A$.

Bedeutet $G$ eine (lineare oder ebene) offene, $F$ eine abgeschlossene Menge, $G_\delta$ den Durchschnitt abzählbar vieler $G$, $F_\sigma$ die Summe abzählbar vieler $F$, so gilt:

I. *$C$ sei ein ebenes $G_\delta$, $B$ ein lineares $F_\sigma$, $X$ beliebig. Ist $C \supseteq (X, B)$, so giebt es ein lineares $G_\delta$ $A \supseteq X$ derart, dass auch noch $C \supseteq (A, B)$.*

Bl. 120 | Beweis. ($\alpha$) $C$ sei *offen*, $B$ beschränkt und abgeschlossen. $C$ kann als Summe abzählbar vieler *offener* Rechtecke $k_n = (i_n, h_n)$ (wo $i_n, h_n$ offene Intervalle sind) dargestellt werden, z. B. der in $C$ enthaltenen mit rationalen Ecken. Jeder einzelne Verticalschnitt $(x, B)$ von $(X, B)$ ist nach dem Borelschen Satze in endlich vielen dieser $k_n$ enthalten; der Durchschnitt dieser $i_n$ sei $i_x$; $(i_x, B)$ ist $\subseteq C$. Die Summe aller $i_x$ liefert eine *offene* Menge $A \supseteq X$ mit $(A, B) \subseteq C$.

$(\beta)$ $C = C_1C_2\cdots$ sei ein $G_\delta$, die $C_n$ ebene offene Mengen; $B = B_1 \cup B_2 \cup \cdots$ ein $F_\sigma$, die $B_n$ lineare abgeschlossene Mengen. Man kann $C_1 \supseteq C_2 \supseteq \cdots$, $B_1 \subseteq B_2 \subseteq \cdots$ und die $B_n$ beschränkt annehmen; denn ist $N_n = [-n, n]$, so ist $N = N_1 \cup N_2 \cup \cdots$ und, wie leicht zu sehen, $B = BN = B_1N_1 \cup B_2N_2 \cup \cdots$. Da nun $C_n \supseteq (X, B_n)$, so giebt es eine offene Menge $A_n \supseteq X$ mit $C_n \supseteq (A_n, B_n)$. Sei $A = A_1A_2\cdots$ (ein lineares $G_\delta$); so ist $A \supseteq X$. Ferner $C_n \supseteq (A, B_n)$, für $m < n$ $\quad C_m \supseteq (A, B_n)$, also $C_m \supseteq (A, B)$, $\ C \supseteq (A, B)$.

Sei nun $\mathfrak{P}$ ein abgeschlossenes, die ebenen Intervalle (also auch ihre Grenzfälle) Bl. 121
enthaltendes System ebener Mengen $Z$, in dem eine additive Mengenfunction $\Omega(Z) \geq 0$ definirt ist. Hierzu gehört wieder eine Punktfunction

$$\omega(\xi, \eta) = \Omega(x < \xi, y < \eta),$$

das $w$-Mass für jene linken unteren Quadranten, mit folgenden Eigenschaften:
(A) *Sie ist nichtnegativ; nach* $\xi$ *einzeln, nach* $\eta$ *einzeln und nach* $\xi, \eta$ *zusammen monoton zunehmend.*
(B) *Sie ist nach links unten stetig.*
(C) *Sie ist beschränkt; für* $\xi \to -\infty$ *oder* $\eta \to -\infty$ *convergirt sie nach 0.*

(A) besagt, dass für $\xi_0 < \xi$, $\eta_0 < \eta$ die Differenzen

$$\omega(\xi, \eta) - \omega(\xi_0, \eta) \geq 0$$
$$\omega(\xi, \eta) - \omega(\xi, \eta_0) \geq 0$$
$$\omega(\xi, \eta) - \omega(\xi, \eta_0) - \omega(\xi_0, \eta) + \omega(\xi_0, \eta_0) \geq 0\,.$$

Die erste ist $\Omega(\xi_0 \leq x < \xi;\ y < \eta)$, das Mass für einen halben Parallelstreifen, ebenso die zweite; die dritte Doppeldifferenz, die wir mit

$$\omega_{\xi_0\eta_0}^{\xi\eta} = \omega(\xi, \eta) - \omega(\xi, \eta_0) - \omega(\xi_0, \eta) + \omega(\xi_0, \eta_0) \tag{1}$$

abkürzen, ist das Mass eines ebenen Intervalls $= \Omega(\xi_0 \leq x < \xi,\ \eta_0 \leq y < \eta)$.

(B) bedeutet, dass

$$\omega(\xi - 0, \eta - 0) = \lim_{h,k \to 0} \omega(\xi - h, \eta - k) = \omega(\xi, \eta). \tag{2}$$

In der That: ist $\xi_1 < \xi_2 < \cdots \to \xi$, $\quad \eta_1 < \eta_2 < \cdots \to \eta$, so bilden die Bl. 122
Quadranten $(x < \xi_n,\ y < \eta_n)$ eine aufsteigende Folge mit den Quadranten $(x < \xi,\ y < \eta)$ als Summe, also $\omega(\xi_n, \eta_n) \to \omega(\xi, \eta)$. Und ist $\xi_n \to \xi - 0$, $\eta_n \to \eta - 0$ eine beliebige Folge, so kann man aus ihr eine Theilfolge obiger Art herausgreifen; $\omega(\xi_n, \eta_n)$ enthält eine nach $\omega(\xi, \eta)$ convergente Theilfolge und muss, da für jede Theilfolge dasselbe gilt, selber nach $\omega(\xi, \eta)$ convergiren.

Beiläufig folgt aus (A) (B), dass alle vier *Quadrantengrenzwerthe* $\omega(\xi \pm 0,\ \eta \pm 0)$ und alle vier *Achsengrenzwerthe* $\omega(\xi \pm 0,\ \eta), \omega(\xi, \eta \pm 0)$ existiren und darin immer $\xi - 0, \eta - 0$ durch $\xi, \eta$ ersetzt werden können, so dass nur die 4 Werthe

$$\begin{array}{rcl} \omega(\xi,\eta) & = & \Omega(x < \xi,\ y < \eta) \\ \omega(\xi,\eta+0) & = & \Omega(x < \xi,\ y \leq \eta) \\ \omega(\xi+0,\eta) & = & \Omega(x \leq \xi,\ y < \eta) \\ \omega(\xi+0,\eta+0) & = & \Omega(x \leq \xi,\ y \leq \eta) \end{array}$$

übrig bleiben.

Zu (C): es ist $\omega(\xi,\eta) \leq \Omega(P) = \mu$, woraus man mittels aufsteigender Quadranten genau wie (2) die Formel

$$\omega(\infty,\infty) = \lim_{\substack{\xi\to\infty \\ \eta\to\infty}} \omega(\xi,\eta) = \mu \tag{3}$$

beweist. Mit $\omega(\xi,\eta)$ sind noch die beiden Functionen

$$\varphi(\xi) = \omega(\xi,\infty), \quad \psi(\eta) = \omega(\infty,\eta) \tag{4}$$

Bl. 123 | verknüpft, die Masse von Halbebenen: $\Omega(x < \xi)$, $\Omega(y < \eta)$; sie sind ihrerseits monoton und

$$\varphi(\infty) = \psi(\infty) = \mu, \quad \varphi(-\infty) = \psi(-\infty) = 0.$$

Endlich folgt aus $\omega(\xi,\eta) \leq \varphi(\xi)$, dass für $\xi \to -\infty \quad \omega(\xi,\eta) \to 0$, gleichviel was die Variable $\eta$ thut (d. h. $\omega(\xi_n,\eta_n) \to 0$ für $\xi_n \to -\infty$), ebenso für
$\eta \to -\infty \quad w(\xi,\eta) \to 0$.

Ist nun umgekehrt eine Function $\omega(\xi,\eta)$ mit diesen Eigenschaften gegeben, so existirt ein abgeschlossenes Mengensystem $\mathfrak{P}$ (das Lebesguesche), das die Intervalle $K$ enthält und in dem eine additive Mengenfunction $\Omega(Z)$ erklärt ist, die für Quadranten $(x < \xi,\ y < \eta)$ mit dem vorgegebenem $\omega(\xi,\eta)$ übereinstimmt. Der Beweis ist genau wie im linearen Falle, so dass es genügt, die Definitionen anzugeben:

$$\begin{array}{lcll} \text{für } K & = & (\xi_0 \leq x < \xi,\ \eta_0 \leq y < \eta) & \Omega(K) = \omega_{\xi_0\eta_0}^{\xi\eta} \\ \text{für } E & = & K_1 + \cdots + K_n & \Omega(E) = \Omega(K_1) + \cdots + \Omega(K_n) \\ \text{für } A & = & K_1 + K_2 + \cdots & \Omega(A) = \Omega(K_1) + \Omega(K_2) + \cdots \\ & & \overline{\Omega}(Z) = \inf \Omega(A) \text{ für } A \supseteq Z & \\ & & \underline{\Omega}(Z) = \mu - \overline{\Omega}(P - Z) & \end{array}$$

Dass für $K = K_1 + \cdots + K_n \quad \Omega(K) = \Omega(K_1) + \cdots + \Omega(K_n)$, ergiebt sich mit Hülfe vollständiger Zerlegungen; dass für $K = K_1 + K_2 + \cdots \quad \Omega(K) =$
Bl. 124 $\Omega(K_1) + \Omega(K_2) + \cdots$, | beruht wieder darauf, dass man vermöge (2) die Intervalle $K$ unter beliebig kleiner Änderung von $\Omega$ zu abgeschlossenen oder zu

offenen vergrössern kann. Die Momente oder *Stieltjes-Doppelintegrale* sind passend auch mit Doppeldifferential

$$\iint f\, d^2\omega = \iint f\, dd\omega$$

(ausführlicher $\iint f(x,y)\, d^2\omega(x,y)$) zu schreiben, da hier die Doppeldifferenz $\omega_{\xi_0\eta_0}^{\xi\eta}$ dieselbe Rolle spielt wie im linearen Fall die einfache $\varphi_{\xi_0}^{\xi}$; ihre Definition ist: für eine Skalenfunction $f$, die die endlich oder abzählbar vielen Werthe $u_m$ in den messbaren Mengen $Z_m$ annimmt,

$$\iint f\, d^2\omega = \sum u_m\, \Omega(Z_m), \tag{5}$$

die Reihe als absolut convergent vorausgesetzt, und für eine messbare Function $f = \lim f_n$ als gleichmässigen Limes integrabler Skalenfunctionen

$$\iint f\, d^2\omega = \lim \iint f_n\, d^2\omega \tag{6}$$

Auch $\iint_Z f\, d^2\omega$ wird wie damals erklärt, für ein in $P$ integrables $f$ definirt dies in $\mathfrak{P}$ eine additive Mengenfunction.

| *Reduction auf einfache Integrale.* Mit der monotonen Function $\varphi(\xi) = \omega(\xi, \infty)$ Bl. 125
bilden wir auf der $x$-Achse das Lebesguesche System $\mathfrak{M}$ der messbaren Mengen $X$, d. h. wir definiren $\overline{\Phi}(X), \underline{\Phi}(X)$ wie im vorigen Paragraphen und im Gleichheitsfalle $\Phi(X)$. Jeder Menge $X$ entspricht eine ebene Menge $Z = (X, N)$, gebildet von den durch alle Punkte von $x$ gehenden Verticalgeraden; nennen wir dies einen *Verticalstreifen*. Wir behaupten, dass

$$\overline{\Omega}(X,N) = \overline{\Phi}(X), \quad \underline{\Omega}(X,N) = \underline{\Phi}(X), \tag{7}$$

so dass also $(X, N)$ und $X$ gleichzeitig (nach $\omega$ resp. $\varphi$) messbar sind und in diesem Falle

$$\Omega(X,N) = \Phi(X) \tag{8}$$

besteht. Die messbaren $X$ und die messbaren Verticalstreifen $(X, N)$ bilden abgeschlossene Mengensysteme, zu denen die Intervalle $I$ und die Streifen $(I, H)$, daher die Borelschen Mengen $X$ und die entsprechenden Borelschen Streifen $(X, N)$ gehören; für diese also besteht die Gleichung (8), da sie per definitionem für die Halbgerade $(-\infty, \xi)$ und also für die $I, I_\sigma, \ldots$ besteht. Ist $X$ beliebig, $A \supseteq X$ ein $G_\delta$ (§ 5) mit $\Phi(A) = \overline{\Phi}(X)$, so ist

$$\overline{\Omega}(X,N) \leq \Omega(A,N) = \Phi(A) = \overline{\Phi}(X).$$

Andererseits sei $C$ ein ebenes $G_\delta \supseteq (X, N)$ mit $\Omega(C) = \overline{\Omega}(X, N)$; | nach I. Bl. 126
existirt, da $N$ ein $F_\sigma$ ist, ein lineares $G_\delta$ $A \supseteq X$ mit $C \supseteq (A, N)$, also

$$\overline{\Omega}(X,N) = \Omega(C) \geq \Omega(A,N) = \Phi(A) \geq \overline{\Phi}(X).$$

Hiermit ist die erste Gleichung (7) bewiesen; die andere folgt nach der Definition der unteren Masse unter Berücksichtigung von $\Omega(P) = \Phi(M) = \mu$.

Hieraus folgt weiter, wenn $f = f(x)$ Function von $x$ allein ist,

$$\iint f(x)\, d^2\omega(x,y) = \int f(x)\, d\varphi(x) \tag{9}$$

in dem Sinne, dass mit dem einen Integral auch das andere existirt und ihm gleich ist. Denn ist $f$ messbare Skalenfunction und nimmt die Werthe $u_m$ als Function von $x$ in den linearen Mengen $X_m$, als Function von $x, y$ aufgefasst in den zugehörigen Verticalstreifen $Z_m = (X_m, N)$ an, so ist $f$, wenn im einen, so auch im anderen Sinne messbar, ferner ist $\Omega(Z_m) = \Phi(X_m)$ und die das Integral definirende Reihe ist beidemal dieselbe. Jede messbare Function $f(x)$ kann durch messbare Scalenfunctionen $f_n(x)$ gleichmässig approximirt werden, woraus sich (9) allgemein ergiebt. Dasselbe Raisonnement gilt auch für Horizontal- und schräge Streifen; d. h. ist $s = \alpha x + \beta y$ ($\alpha, \beta$ nicht beide 0) und

$$\chi(\sigma) = \Omega(\alpha x + \beta y < \sigma) \tag{10}$$

Bl. 127 die Massfunction für diese, durch Parallelen $s = \sigma$ begrenzte offene Halbebenen, so ist

$$\iint f(\alpha x + \beta y)\, d^2\omega(x,y) = \int f(s)\, d\chi(s). \tag{11}$$

*Unabhängige Variable. Wiederholte Integrale.* Sei $A$ das Ereignis $(x < \xi)$, $B$ das Ereignis $(y < \eta)$, $AB$ das Ereignis $(x < \xi,\ y < \eta)$. Auf die Menge $P$ aller möglichen Fälle bezogen ist, wenn $X = (-\infty, \xi)$ und $Y = (-\infty, \eta)$ die entsprechenden linearen Halbgeraden bedeuten, $(X, N)$ die Menge der für $A$ günstigen Fälle, $(M, Y)$ die der für $B$ günstigen Fälle, ihr Durchschnitt $(X, Y)$ die Menge der für $AB$ günstigen Fälle. Die $w$-Masse dieser Mengen sind $\varphi(\xi), \psi(\eta), \omega(\xi, \eta)$, also wegen $\Omega(P) = \mu$

$$w(A) = \frac{\varphi(\xi)}{\mu}, \quad w(B) = \frac{\psi(\eta)}{\mu}, \quad w(AB) = \frac{\omega(\xi,\eta)}{\mu} \tag{12}$$

Wir nennen (§ 1) die beiden Variablen $x, y$ *unabhängig*, wenn stets

$$w(AB) = w(A)\, w(B),$$

also

$$\mu \cdot \omega(\xi, \eta) = \varphi(\xi)\ \psi(\eta),$$

d. h. $\omega(\xi, \eta)$ das Produkt zweier monotoner Functionen je einer Variablen ist. Zur Vereinfachung setzen wir $\mu = 1$

$$\omega(\xi, \eta) = \varphi(\xi)\ \psi(\eta) \tag{13}$$

so dass nunmehr

$$\omega_{\xi_0,\eta_0}^{\xi,\eta} = [\varphi(\xi) - \varphi(\xi_0)]\, [\psi(\eta) - \psi(\eta_0)] = \varphi_{\xi_0}^{\xi} \cdot \psi_{\eta_0}^{\eta}$$

die Doppeldifferenz das Produkt von zwei einfachen Differenzen wird. Entsprechend ersetzen wir das Doppeldifferential $d^2\omega(x,y)$ durch $d\varphi(x)\,d\psi(y)$, das Produkt zweier einfacher Differentiale:

$$\iint f\,d^2\omega = \iint f\,d\varphi d\psi$$

Hierher gehört der Fall $\omega(x,y) = xy$, wo $\Omega(K)$ das Flächenmass des Rechtecks $K$, $\Omega(Z)$ das *Lebesguesche Flächenmass* von $Z$, $\iint f\,dxdy$ das gewöhnliche Lebesguesche Doppelintegral wird. Die (wegen $\varphi(\pm\infty) = \pm\infty$ u.s.w.) nöthigen Modificationen sind wie im linearen Fall zu bewirken.

Auf der $x$-Achse $M$ definiren wir mit der Verteilungsfunction $\varphi(x)$ die äusseren und inneren Masse $\overline{\Phi}(X), \underline{\Phi}(X)$, wir erhalten insbesondere das Lebesguesche System $\mathfrak{M}$ der messbaren (nach $\varphi$) Mengen $X$ mit der additiven Mengenfunction $\Phi(X)$. Entsprechend auf der $Y$-Achse $N$: $\psi(y)$, $\overline{\Psi}(Y)$, $\underline{\Psi}(Y)$, $\Psi(Y)$, $\mathfrak{N}$. In der Ebene $P : \omega(x,y)$, $\overline{\Omega}(Z)$, $\underline{\Omega}(Z)$, $\Omega(Z)$, $\mathfrak{P}$. Sodann können wir mit $\varphi$ und $\psi$ wiederholte einfache Integrale bilden; ist z. B. $f(x,y)$ bei constantem $x$ nach $\psi$ integrabel, so ist

$$F = \int f\,d\psi, \quad F(x) = \int f(x,y)\,d\psi(y) \tag{14}$$

Bl. 129 Function von $x$ allein, mit der, falls sie nach $\varphi$ integrabel ist, das iterirte Integral

$$\int F\,d\varphi = \int d\varphi \int f\,d\psi \tag{15}$$

gebildet werden kann. Wir fragen, ob dies dem Doppelintegral gleich ist:

$$\iint f\,d\varphi d\psi = \int d\varphi \int f\,d\psi = \int F\,d\varphi, \tag{16}$$

und werden beweisen:

II. *Falls das Doppelintegral* $\iint f\,d\varphi d\psi$ *und für jedes* $x$ *das einfache Integral* $F = \int f\,d\psi$ *existirt (also* $f$ *nach* $\omega$ *und bei constantem* $x$ *nach* $\psi$ *integrabel ist), so ist* $F$ *nach* $\varphi$ *integrabel und es besteht* (16).

Ein entsprechender Satz gilt für Mengen: der Menge $Z$ ordnen wir die „charakteristische Function“ $f$ zu, die in $Z$ gleich 1, in $P - Z$ gleich 0 ist. $(x, Y)$ sei der Durchschnitt von $Z$ mit der Verticalen $(x, N)$, $Y = Y(x)$ seine Projection auf die $y$-Achse; der Kürze halber nennen wir $Y$ selbst (statt $(x, Y)$, was sprachlich

näher liegt), einen *Verticalschnitt* von $Z$. Dann lautet II: Bl. 130

III. *Falls $\Omega(Z)$ und für jeden Verticalschnitt $Y \quad \Psi(Y)$ existirt, (also $Z$ nach $\omega$ und jeder Verticalschnitt nach $\psi$ messbar ist), so ist $\Psi(Y)$ nach $\varphi$ integrabel und*

$$\Omega(Z) = \int \Psi(Y)\,d\varphi. \tag{17}$$

Nennen wir einstweilen Functionen und Mengen, für welche die Sätze gelten, (d. h. die sowohl die Voraussetzungen wie die Behauptungen erfüllen) *normal.* Summe und Differenz von zwei normalen Functionen ist normal, ferner: *Wenn die normalen Functionen $f_n$ aufsteigend nach $f$ convergiren und zugleich $I_n = \iint f_n\,d\varphi d\psi \to I, \quad F_n = \int f_n d\psi \to F$ (letzteres für jedes $x$), so ist auch $f$ normal.*

Denn nach § 5, I ist dann $\iint f\,d\varphi d\psi = I,\ \int f\,d\psi = F$; da ferner die nach $\varphi$ integrablen Functionen $F_n$ aufsteigend nach $F$ convergiren und $\int F_n\,d\varphi = I_n \to I$, so ist, wieder nach demselben Satz, auch $F$ nach $\varphi$ integrabel und $\int F\,d\varphi = I$.

Insbesondere ist die Differenz $Z_1 - Z_2$ normaler Mengen normal (charakteristische Function $f_1 - f_2$) sowie die Summe $Z_1 + Z_2$ von zwei fremden normalen Mengen (charakteristische Function $f_1 + f_2$. Dass allgemein $Z_1 \cup Z_2$ Bl. 131 normal ist, folgt hier | noch nicht; die charakteristische Function ist nicht $f_1 + f_2$, sondern $\max[f_1, f_2]$ oder $f_1 + f_2 - f_1 f_2$). Ferner ist die Summe $Z = Z_1 \cup Z_2 \cup \cdots$ *aufsteigender* normaler Mengen normal (charakteristische Function $f = \lim f_n$ mit aufsteigenden $f_n$, wobei $f_n, I_n, F_n \leq 1$ sind, also die Grenzwerthe $I, F$ sicher existiren); ebenso die Summe paarweise fremder normaler Mengen $Z = Z_1 + Z_2 + \cdots$. Da nun die Intervalle $K = (I, H)$ offenbar normal sind (der Verticalschnitt ist $H$ oder 0, wenn $x$ in $I$ oder $M - I$ liegt, $\Omega(K) = \int_I \Psi(H)\,d\varphi = \Phi(I)\Psi(H) = \varphi_{\xi_0}^{\xi}\psi_{\eta_0}^{\eta} = \omega_{\xi_0\eta_0}^{\xi\eta}$), so sind es auch die Mengen $A = K_1 + K_2 + \cdots$, ferner die $A_\delta$ (als Durchschnitte absteigender $A_n$; da das Complement jeder normalen Menge normal ist, ist mit der Summe aufsteigender auch der Durchschnitt absteigender normaler Mengen normal) und ihre Complemente.

Zu einer beliebigen Menge $Z$ bilden wir eine Borelsche Menge $C \supseteq Z$ mit $\Omega(C) = \overline{\Omega}(Z)$, wobei $C$ als $A_\delta$, also normal angenommen werden kann. Jeder Verticalschnitt $B$ von $C$ ist eine lineare Borelsche Menge und schliesst den entsprechenden Verticalschnitt $Y$ von $Z$ ein, also ist $\Psi(B) \geq \overline{\Psi}(Y)$ und nach Bl. 132 $\varphi$ integrabel, d. h. nach der Definition des oberen Integrals (§ 5) |

$$\overline{\Omega}(Z) = \Omega(C) = \int \Psi(B)\,d\varphi \geq \overline{\int} \overline{\Psi}(Y)\,d\varphi.$$

Ebenso können wir eine Menge $C \subseteq Z$ mit $\underline{\Omega}(Z) = \Omega(C)$ benutzen, die als Complement eines $A_\delta$ normal ist. So erhalten wir

$$\underline{\Omega}(Z) \leq \underline{\int} \underline{\Psi}(Y)\,d\varphi \leq \overline{\int} \overline{\Psi}(Y)\,d\varphi \leq \overline{\Omega}(Z). \tag{18}$$

Ist $\Omega(Z)$ vorhanden, so stimmen die inneren Integrale, also auch die zwischen ihnen liegenden $\underline{\int}\overline{\Psi}(Y)\,d\varphi$ und $\overline{\int}\underline{\Psi}(Y)\,d\varphi$, mit $\Omega(Z)$ überein, d. h. $\overline{\Psi}(Y)$ und $\underline{\Psi}(Y)$ sind nach $\varphi$ integrabel und

$$\Omega(Z) = \int \underline{\Psi}(Y)\,d\varphi = \int \overline{\Psi}(Y)\,d\varphi$$

Daraus folgt übrigens wegen $\overline{\Psi} \geq \underline{\Psi}$, dass bis auf eine Nullmenge $X$ (mit $\Phi(X) = 0$) $\overline{\Psi} = \underline{\Psi}$, die Verticalschnitte einer messbaren Menge „fast überall" nach $\psi$ messbar sind. (Dies giebt eine Verschärfung des Satzes III und eine analoge von II). Nehmen wir von vornherein die Existenz von $\Psi(Y)$ für jedes $x$ an, so folgt (17) und III ist bewiesen.

*Beispiele.* Sind $A, B$ lineare Borelsche Mengen, so ist $C = (A, B)$ eine ebene Borelsche Menge und (17) giebt

$$\Omega(A, B) = \Phi(A)\,\Psi(B).$$

Sind $X, Y$ beliebige lineare Mengen und $Z = (X, Y)$; so schliesse | man $X, Y$ in Bl. 133
Borelsche Mengen $A, B$ mit $\Phi(A) = \overline{\Phi}(X),\ \Psi(B) = \overline{\Psi}(Y)$ ein; da dann $Z$ in $C$ eingeschlossen ist, ist $\overline{\Omega}(Z) \leq \Omega(C) = \Phi(A)\,\Psi(B) = \overline{\Phi}(X)\,\overline{\Psi}(Y)$.
Ebenso ergiebt sich

$$\underline{\Phi}(X)\,\underline{\Psi}(Y) \leq \underline{\Omega}(Z) \leq \overline{\Omega}(Z) \leq \overline{\Phi}(X)\,\overline{\Phi}(Y).$$

Also: sind $X, Y$ nach $\varphi, \psi$ messbar, so ist $(X, Y)$ nach $\omega$ messbar und

$$\Omega(X, Y) = \Phi(X)\,\Psi(Y), \tag{19}$$

womit die Unabhängigkeitsbeziehung (13) verallgemeinert ist.

Ist $Z$ die Halbebene $(x + y < \sigma)$, ihr Mass $\Omega(Z) = \chi(\sigma)$ also die Vertheilungsfunction für die Variable $s = x + y$, so ist der Verticalschnitt $Y$ durch die Halbgerade $y < \sigma - x$ gegeben, also

$$\chi(\sigma) = \int \psi(\sigma - x)\,d\varphi(x) = \int \varphi(\sigma - y)\,d\psi(y). \tag{20}$$

Kommen wir nun zum Satz II; es genügt, ihn für $f \geq 0$ zu beweisen (im allgemeinen Fall ist die Zerlegung $f = f_1 - f_2$ in $f_1 = \frac{1}{2}|f| + \frac{1}{2}f,\ f_2 = \frac{1}{2}|f| - \frac{1}{2}f$ zu benutzen). Ist $f \geq 0$ zunächst eine Skalenfunction, die die Werthe $u_m$ in den Mengen $Z_m$, bei constantem $x$ in ihren Verticalschnitten $Y_m$ annimmt, so existirt nach Voraussetzung

$$\iint f\,d\varphi d\psi = \sum u_m\,\Omega(Z_m), \qquad F = \int f\,d\psi = \sum u_m\,\Psi(Y_m)$$

| und es ist $\Omega(Z_m) = \int \Psi(Y_m)\,d\varphi$. Nach § 5, II ist, wegen der Convergenz von Bl. 134
$\sum \int u_m \Psi(Y_m)\,d\varphi = \sum u_m \Omega(Z_m)$, auch $F$ nach $\varphi$ integrabel und $\int f\,d\varphi$ gleich

der Summe $= \iint f\, d\varphi d\psi$. – Eine beliebige Function $f \geq 0$, die nach $\omega$ und bei constantem $x$ nach $\psi$ integrabel ist, kann man als Limes aufsteigender Skalenfunctionen gleicher Eigenschaft darstellen (sei $g = m\varepsilon$ in

$$[m\varepsilon \leq f < (m+1)\varepsilon]$$

für $m = 0, 1, 2, \ldots$; setzt man $\varepsilon = 1, \frac{1}{2}, \frac{1}{4}, \ldots$, so convergirt $g$ aufsteigend nach $f$); diese sind bereits als normal erkannt, ihre Integrale nach $\omega$ und $\psi$ sind nach oben beschränkt ($\leq \iint f\, d\varphi d\psi$ resp. $\leq \int f\, d\psi$) und danach ist auch $f$ normal. Q.E.D

Ein zu II entsprechender Satz gilt bei Vertauschung von $x, y$, also

$$\iint f\, d\varphi d\psi = \int d\psi \int f\, d\varphi = \int G\, d\psi, \qquad G = \int f\, d\varphi. \tag{21}$$

Ist $f$ nach $\omega$, bei constantem $x$ nach $\psi$ und bei constantem $y$ nach $\varphi$ integrabel, so sind die beiden iterirten Integrale gleich, die Integration vertauschbar.

*Ist $f(x)$ nach $\varphi$ , $g(y)$ nach $\psi$ integrabel, so ist $f(x)g(y)$ nach $\omega$ integrabel.* Ist dies bewiesen, so folgt aus (16)

$$\iint fg\, d\varphi d\psi = \int f\, d\varphi \cdot \int g\, d\psi, \tag{22}$$

Bl. 135 | eine Verallgemeinerung von (19). Man kann sich wieder auf $f \geq 0$, $g \geq 0$ beschränken. $f$ ist nach $\varphi$, also nach $\omega$ messbar, vgl. (8), $g$ und $fg$ sind nach $\omega$ messbar. Sind $f (= u_m$ in $X_m)$ und $g$ $(= v_n$ in $Y_n)$ Skalenfunctionen, so folgt aus der Convergenz von $\sum u_m \Phi(X_m)$, $\sum v_n \Psi(Y_n)$ die der Doppelreihe

$$\sum u_m v_n \Phi(X_m)\Psi(Y_n) = \sum u_m v_n \Omega(X_m, Y_n) = \iint fg\, d\varphi d\psi.$$

Endlich: zu $f \geq 0$, $g \geq 0$ bilde man integrable Skalenfunctionen $\overline{f} \geq f$, $\overline{g} \geq g$; aus der Messbarkeit von $fg$ und der Integrabilität von $\overline{f}\overline{g}$ folgt wegen $0 \leq fg \leq \overline{f}\overline{g}$ die Integrabilität von $fg$.

*Momente und logarithmische Momente.* Für eine Variable $x$ mit der Vertheilungsfunction $\varphi(x)$ seien

$$\mu_k = \int x^k d\varphi \qquad (k = 0, 1, 2, \ldots)$$

die Momente ($\mu_0 = 1$); wir nehmen an, dass alle existiren. (Mit $\mu_k$ existirt auch $\nu_k = \int |x|^k d\varphi$, also $\nu_{k-1}$ und $\mu_{k-1}$. Im Allgemeinen werden $\mu_0, \ldots, \mu_{k-1}$ existiren, $\mu_k, \mu_{k+1}, \ldots$ nicht). Die logarithmischen Momente $\lambda_k$ definiren wir durch § 2, (7), um von der Voraussetzung

$$\int e^{xu} d\varphi = 1 + \mu_1 \frac{u}{1!} + \mu_2 \frac{u^2}{2!} + \cdots$$

unabhängig zu sein. (Es gilt aber: wenn für $|u| < \rho$ das Integral existirt, so convergirt auch die Reihe, und umgekehrt, und beide | sind gleich). Entsprechend sind für ein Variablenpaar $x, y$ die Momente

$$\mu_{k,l} = \iint x^k y^l d^2\omega \qquad (\mu_{0,0} = 1)$$

und die logarithmischen Momente (Ende von § 2) zu definiren. Sind $x, y$ *unabhängig* mit den Vertheilungsfunctionen $\varphi, \psi$ und den Momenten

$$\mu_{k,0} = \int x^k d\varphi, \qquad \mu_{0,l} = \int y^l d\psi;$$

so existiren nach (22) auch die Momente $\mu_{k,l} = \mu_{k,0} \cdot \mu_{0,l}$.

Für die Summe $s = x + y$ mit der Vertheilungsfunction $\chi = (20)$ ist nach (11)

$$\int f(s)\, d\chi(s) = \iint f(x+y)\, d\varphi(x)\, d\psi(y),$$

insbesondere sind ihre Momente

$$\mu_n = \int s^n d\chi = \iint (x+y)^n\, d\varphi d\psi = \sum_k \binom{n}{k} \mu_{k,0}\, \mu_{0,n-k}$$

und hieraus folgt algebraisch, dass die logarithmischen Momente sich additiv verhalten: $\lambda_k = \lambda_{k,0} + \lambda_{0,k}$ oder

$$\lambda_k(x+y) = \lambda_k(x) + \lambda_k(y).$$

| *Die Bayessche Regel.* $A$ sei das Ereignis $[x < \xi]$, $B_i$ das Ereignis $[y = y_i]$ Bl. 137
und $w_{B_i}(A) = \varphi(\xi,\ y_i)$, d. h. die Vertheilungsfunction $\varphi$ von $x$ hängt noch von einem Parameter $y$ ab, der vorläufig einzelne Werthe $y_i$ mit den Wahrscheinlichkeiten $w(B_i) = \beta_i$ annehmen kann. Ist $B = B_1 + \cdots + B_n$ (die $y_i$ paarweise verschieden), so ist

$$w(AB) = \sum w(AB_i) = \sum \beta_i\, \varphi(\xi, y_i).$$

Ist nun aber der Parameter $y$ selbst eine allgemeine Variable mit der Vertheilungsfunction $\psi(\eta) = w(y < \eta)$, so wird das Moment $\sum \beta_i\, \varphi(\xi,\ y_i)$ von $\varphi(\xi,\ y)$ nach $y$ nun durch ein Stieltjes-Integral zu ersetzen sein, z. B. für $B = (y < \eta)$

$$w(AB) = \int_{(-\infty,\eta)} \varphi(\xi,\ y)\, d\psi(y) = \int_{-\infty}^{\eta} \varphi(\xi, y)\, d\psi(y)$$

Dies ist das, was wir sonst $\omega(\xi, \eta)$ genannt haben; es ist also hier

$$w(AB) = \omega(\xi, \eta) = \int_{-\infty}^{\eta} \varphi(\xi,\ y)\, d\psi(y)$$

($\varphi(x,y)$ nach $x$ monoton und links stetig, $\varphi(-\infty,y) = 0$, $\varphi(\infty,y) = 1$), während

$$w(A) = \int_{-\infty}^{\infty} \varphi(\xi,\ y)\, d\psi(y) = \varphi(\xi), \qquad w(B) = \psi(\eta)$$

ist. Die Variablen $x, y$ sind hier nicht unabhängig. Ist $A$ eingetreten, so ist $w_A(B) = w(AB) : w(A) = \omega(\xi,\eta) : \varphi(\xi)$ die Wahrscheinlichkeit a posteriori von $B$.

Bl. 138

## § 7. Das Exponentialgesetz. Der Grenzwerthsatz von Liapunoff.

*Die Eulerschen Integrale* (vgl. Lehrbücher der Integralrechnung).

$$\Gamma(\alpha) = \int_0^{\infty} e^{-t} t^{\alpha-1}\, dt \qquad (\alpha > 0)$$

ist das Eulersche Integral 2. Gattung, die $\Gamma$-Function Legendres. Gauss setzt $\Pi(\alpha) = \Gamma(\alpha+1)$ für $\alpha > -1$. Aus

$$\frac{d}{dt}\left(e^{-t}t^{\alpha}\right) = \alpha e^{-t} t^{\alpha-1} - e^{-t} t^{\alpha} \qquad (\alpha > 0)$$

folgt durch Integration zwischen $0, \infty$

$$\Gamma(\alpha+1) = \alpha\,\Gamma(\alpha) \quad \text{oder} \quad \Pi(\alpha) = \alpha\,\Pi(\alpha-1) \qquad (\alpha > 0)$$

$$\text{und da} \quad \Gamma(1) = \Pi(0) = \int_0^{\infty} e^{-t}\, dt = 1\ : \qquad \Pi(n) = n! \quad \text{für } n = 1, 2, \dots$$

$$B(\alpha,\beta) = \int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\, du \qquad (\alpha, \beta > 0)$$

ist das Eulersche Integral 1. Gattung. Es drückt sich vermöge

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

durch $\Gamma$-Functionen aus. Kürzester Beweis, bei dem aber Theorie der Doppelintegrale vorausgesetzt wird: Bl. 139 Setzt man in

$$\Gamma(\alpha) = \int_0^{\infty} e^{-t} t^{\alpha-1}\, dt \qquad t = x^2,$$

so wird

$$\Gamma(\alpha) = 2\int_0^{\infty} e^{-x^2} x^{2\alpha-1}\, dx, \qquad \Gamma(\beta) = 2\int_0^{\infty} e^{-y^2} y^{2\beta-1}\, dy,$$

$$\Gamma(\alpha)\,\Gamma(\beta) = 4\int_0^{\infty}\int_0^{\infty} e^{-x^2-y^2} x^{2\alpha-1} y^{2\beta-1}\, dx dy.$$

In diesem Doppelintegral über den positiven Quadranten der $xy$-Ebene führe man Polarcoordinaten ein: $x = r\cos\varphi,\ y = r\sin\varphi$, wobei $dxdy$ durch $(x_r y_\varphi - y_r x_\varphi) dr d\varphi = r dr d\varphi$ zu ersetzen ist. Also

$$\Gamma(\alpha)\,\Gamma(\beta) = 4\int_0^\infty dr \int_0^{\frac{\pi}{2}} d\varphi \cdot e^{-r^2} r^{2\alpha+2\beta-1} (\cos\varphi)^{2\alpha-1} (\sin\varphi)^{2\beta-1}$$

$$= 2\,\Gamma(\alpha+\beta) \cdot \int_0^{\frac{\pi}{2}} (\cos\varphi)^{2\alpha-1} (\sin\varphi)^{2\beta-1}\, d\varphi$$

$$(\cos^2\varphi = u) \qquad = \Gamma(\alpha+\beta) \cdot \int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\, du. \qquad \text{q.e.d.}$$

Für ganze Zahlen $m, n \geq 0$ hatten wir die Formel

$$\int_0^1 u^m (1-u)^n\, du = \frac{\Gamma(m+1)\,\Gamma(n+1)}{\Gamma(m+n+2)} = \frac{m!\; n!}{(m+n+1)!}$$

bereits elementar abgeleitet (§ 1 am Ende).
Für $\alpha = \beta = \frac{1}{2}$ wird nach der Formel, in der noch $\varphi$ vorkommt

$$\Gamma\left(\frac{1}{2}\right)^2 = 2\int_0^{\frac{\pi}{2}} d\varphi = \pi, \qquad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

(pos. Wurzel). | Also Bl. 140

$$\sqrt{\pi} = 2\int_0^\infty e^{-x^2}\, dx = \int_{-\infty}^\infty e^{-x^2}\, dx.$$

Die monotone Function

$$\Phi(x) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^x e^{-x^2}\, dx \qquad (\Phi(-\infty) = 0,\ \Phi(+\infty) = 1) \tag{1}$$

oder allgemeiner

$$\Phi(h(\xi-\alpha)) = \frac{h}{\sqrt{\pi}} \int_{-\infty}^\xi e^{-h^2(\xi-\alpha)^2}\, d\xi \qquad (h > 0) \tag{2}$$

liefert die in der Wahrscheinlichkeitsrechnung wichtigste Vertheilungsfunction einer Variablen $x$ oder $\xi$, mit der Dichtigkeit

$$\frac{1}{\sqrt{\pi}} e^{-x^2} \quad \text{oder} \quad \frac{h}{\sqrt{\pi}} e^{-h^2(\xi-\alpha)^2};$$

*das Gausssche Exponentialgesetz.* Die Momente von (1) sind nach § 5 (27)

$$\mu_k = \frac{1}{\sqrt{\pi}} \int_{-\infty}^\infty e^{-x^2} x^k\, dx,$$

also $\mu_1 = \mu_3 = \cdots = 0$, während für gerade Indices $(x^2 = t)$

$$\mu_{2k} = \frac{2}{\sqrt{\pi}} \int_0^\infty e^{-x^2} x^{2k}\, dx = \frac{1}{\sqrt{\pi}} \int_0^\infty e^{-t}\, t^{k-\frac{1}{2}}\, dt$$

$$= \frac{\Gamma\left(k+\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)} = \frac{1}{2} \cdot \frac{3}{2} \cdot \,\cdots\, \cdot \frac{2k-1}{2} \qquad \text{ist;}$$

$$\mu_0 = 1, \quad \mu_2 = \frac{1}{2}, \quad \mu_4 = \frac{1 \cdot 3}{2 \cdot 2}, \quad \mu_6 = \frac{1}{2} \cdot \frac{3}{2} \cdot \frac{5}{2}, \; \cdots$$

Das Moment von $e^{xu}$ ($u$ eine Variable) ist

$$M\, e^{xu} = \frac{1}{\sqrt{\pi}} \int_{-\infty}^\infty e^{-x^2+xu}\, dx = \frac{1}{\sqrt{\pi}} \int_{-\infty}^\infty e^{-\left(x-\frac{u}{2}\right)^2+\frac{u^2}{4}}\, dx = e^{\frac{u^2}{4}},$$

$$\log M\, e^{xu} = \frac{u^2}{4},$$

Bl. 141 *alle logarithmischen Momente des Exponentialgesetzes* (1) *sind 0 bis auf* $\lambda_2 = \frac{1}{2}$.

Für $x = h(\xi - \alpha)$ ist, wie wir wissen, für $k \geq 2 \quad \lambda_k(x) = \lambda_k(h\xi) = h^k \lambda_k(\xi)$, während $\lambda_1(x) = h(\lambda_1(\xi) - \alpha)$ ist. Also: *alle logarithmischen Momente der durch die Vertheilung* (2) *bestimmten Variablen sind 0 bis auf* $\lambda_1 = \alpha$ *und* $\lambda_2 = \dfrac{1}{2h^2}$. $h$ das *Präcisionsmass* (je grösser $h$, desto kleiner die Streuung).

Zwei unabhängige Variable $\xi_1, \xi_2$ mögen die Vertheilungen $\Phi(h_1(\xi_1 - \alpha_1))$, $\Phi(h_2(\xi_2 - \alpha_2))$ haben. Wegen der Addition der logarithmischen Momente hat dann $\xi = \xi_1 + \xi_2$ die ersten beiden logarithmischen Momente $\alpha = \alpha_1 + \alpha_2$ und

$$\frac{1}{2h^2} = \frac{1}{2}\left(\frac{1}{h_1^2} + \frac{1}{h_2^2}\right),$$

die übrigen 0. Man ist hieraus versucht zu schliessen, dass $\xi$ die Vertheilung $\Phi(h(\xi - \alpha))$ hat (im Allg. ist jedoch eine Vertheilung durch ihre Momente oder logarithmischen Momente nicht bestimmt). In der That ist nach § 6, (20) die Vertheilungsfunction von $\xi$

$$\int_{-\infty}^\infty \Phi(h_2(\xi - \xi_1 - \alpha_2))\, d\Phi(h_1(\xi_1 - \alpha_1))$$

Bl. 141v oder mit $\xi - \alpha = x, \ \xi_1 - \alpha_1 = x_1$

$$= \int_{-\infty}^\infty \Phi(h_2(x - x_1))\, d\Phi(h_1 x_1) = \frac{h_1 h_2}{\pi} \int_{-\infty}^\infty e^{-h_1^2 x_1^2}\, dx_1 \int_{-\infty}^x e^{-h_2^2(x-x_1)^2}\, dx$$

$$= \frac{h_1 h_2}{\pi} \int_{-\infty}^x dx \int_{-\infty}^\infty e^{-(h_1^2+h_2^2)x_1^2 + 2h_2^2 x x_1 - h_2^2 x^2}\, dx_1$$

$$= \frac{h_1h_2}{\pi}\int_{-\infty}^{x} dx \int_{-\infty}^{\infty} e^{-\left(\frac{h_1h_2}{h}x_1 - \frac{h\,h_2}{h_1}x\right)^2 - h^2x^2}\,dx_1$$

$$= \frac{h}{\sqrt{\pi}}\int_{-\infty}^{x} e^{-h^2x^2}\,dx = \Phi(hx) = \Phi(h(\xi-\alpha)).$$

Also: *die Summe von zwei unabhängigen Variablen mit Exponentialgesetz folgt wieder dem Exponentialgesetz* (Erhaltung des Exponentialgesetzes). Dasselbe gilt natürlich für drei oder mehr Variablen, sowie für eine beliebige lineare Verbindung von ihnen mit constanten Coefficienten. Bl. 142

Auch die Absolutmomente der Vertheilung (1) sind von Interesse:

$$\nu_k = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} e^{-x^2}|x|^k\,dx = \frac{2}{\sqrt{\pi}}\int_0^{\infty} e^{-x^2}x^k\,dx = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)},$$

wobei $\nu_{2k} = \mu_{2k}$, hingegen Bl. 143

$$\nu_{2k+1} = \frac{\Gamma(k+1)}{\Gamma\left(\frac{1}{2}\right)} = \frac{k!}{\sqrt{\pi}}; \qquad \nu_1 = \frac{1}{\sqrt{\pi}}, \quad \nu_3 = \frac{1}{\sqrt{\pi}}, \quad \nu_5 = \frac{2}{\sqrt{\pi}}, \ \cdots$$

Neben $\Phi(x)$ führt man noch ein (da $\Phi(0) = \frac{1}{2}$)

$$\Theta(x) = 2\,\Phi(x) - 1 = 2\,[\Phi(x) - \Phi(0)] = \frac{2}{\sqrt{\pi}}\int_0^x e^{-x^2}\,dx \tag{3}$$

(Kramp'sche Function, Laplacesches Integral, Wahrscheinlichkeitsintegral); es ist auch $\Theta(x) = \Phi(x) - \Phi(-x)$ die Wahrscheinlichkeit, dass die Variable absolut $\leq x$ sei (für $x \geq 0$). $\Theta(x)$ ist eine ungerade Function, $\Theta(0) = 0$, $\Theta(\infty) = 1$. Für kleine $x$ ist die Potenzreihe

$$\Theta(x) = \frac{2}{\sqrt{\pi}}\left(x - \frac{x^3}{1!\,3} + \frac{x^5}{2!\,5} - \frac{x^7}{3!\,7} + \cdots\right)$$

zur Berechnung bequem.

Für grosse $x > 0$ schätzen wir das Restintegral

$$\int_x^{\infty} e^{-\xi^2}\,d\xi = \sqrt{\pi}(1-\Phi(x)) = \frac{\sqrt{\pi}}{2}(1-\Theta(x))$$

folgendermassen ab. Setzen wir

$$\Psi_n(x) = 2e^{x^2}\int_x^{\infty}\frac{e^{-\xi^2}}{\xi^{2n}}\,d\xi,$$

so dass

$$\Psi_0(x) = 2e^{x^2}\int_x^{\infty} e^{-\xi^2}\,d\xi = \sqrt{\pi}e^{x^2}[1-\Theta(x)]$$

ist. Aus

$$-\frac{d}{d\xi}\left(e^{-\xi^2}\xi^{-2n-1}\right) = 2e^{-\xi^2}\xi^{-2n} + (2n+1)e^{-\xi^2}\xi^{-2n-2}$$

| folgt durch Integration zwischen $x, \infty$ und Multiplikation mit $e^{x^2}$

$$\frac{1}{x^{2n+1}} = \Psi_n(x) + \frac{2n+1}{2}\Psi_{n+1}(x).$$

Also der Reihe nach

$$\begin{aligned} \Psi_0 &= \frac{1}{x} - \frac{1}{2}\Psi_1 \\ \Psi_1 &= \frac{1}{x^3} - \frac{3}{2}\Psi_2 \\ \Psi_2 &= \frac{1}{x^5} - \frac{5}{2}\Psi_3 \\ &\cdots \\ \Psi_n &= \frac{1}{x^{2n+1}} - \frac{2n+1}{2}\Psi_{n+1} \end{aligned}$$

$$\Psi_0(x) = \frac{1}{x} - \frac{1}{2}\frac{1}{x^3} + \frac{1}{2}\frac{3}{2}\frac{1}{x^5} - \cdots + (-1)^n\frac{1}{2}\frac{3}{2}\cdots\frac{2n-1}{2}\left[\frac{1}{x^{2n+1}} - \frac{2n+1}{2}\Psi_{n+1}\right]$$

Die Reihe

$$\frac{1}{x} - \frac{1}{2}\frac{1}{x^3} + \frac{1}{2}\frac{3}{2}\frac{1}{x^5} - \frac{1}{2}\frac{3}{2}\frac{5}{2}\frac{1}{x^7} + \cdots$$

ist nun zwar divergent, aber nennt man $\sigma_n(x)$ die Summe ihrer ersten $n$ Glieder, so ist (weil $\Psi_n > 0$)

$$0 < (-1)^n[\Psi_0(x) - \sigma_{n+1}(x)] = \frac{1}{2}\cdot\frac{3}{2}\cdots\frac{2n+1}{2}\Psi_{n+1} < \frac{1}{2}\frac{3}{2}\cdots\frac{2n+1}{2}\frac{1}{x^{2n+1}};$$

der Fehler, mit dem $\Psi_0(x)$ duch $\sigma_{n+1}(x)$ dargestellt wird, kleiner als das letzte Glied. Oder auch: $\Psi_0 - \sigma_n$ und $\Psi_0 - \sigma_{n+1}$ haben entgegengesetztes Zeichen, d. h. $\Psi_0$ liegt zwischen zwei aufeinanderfolgenden Partialsummen. Hierdurch lässt sich $\Psi_0$ zwischen zwei Grenzen bringen, deren Differenz bei festem $x$ zwar nicht unbegrenzt verkleinert werden kann, aber doch mit wachsendem $x$ beliebig klein wird, und

$$\Theta(x) = 1 - \frac{e^{-x^2}}{\sqrt{\pi}}\Psi_0(x)$$

Bl. 145 | lässt sich für einigermassen beträchtliches $x$ schon sehr genau abschätzen. Tafeln für $\Theta(x)$ finden sich in den meisten Lehrbüchern der Wahrscheinlichkeitsrechnung; das folgende Täfelchen giebt genügenden Aufschluss über die rasche Annäherung an 1.

| $x$ | $\Theta(x)$ | $x$ | $\Theta(x)$ | $x$ | $\Theta(x)$ |
|---|---|---|---|---|---|
| 0,0 | 0,000 | 1,0 | 0,843 | 2,0 | 0,995 |
| 0,1 | 112 | 1,1 | 880 | 2,1 | 997 |
| 0,2 | 223 | 1,2 | 910 | 2,2 | 998 |
| 0,3 | 329 | 1,3 | 934 | 2,3 | 999 |
| 0,4 | 428 | 1,4 | 952 | 2,4 | 999 |
| 0,5 | 520 | 1,5 | 966 | 2,5 | 1,000 |
| 0,6 | 604 | 1,6 | 976 | | |
| 0,7 | 678 | 1,7 | 984 | | |
| 0,8 | 742 | 1,8 | 989 | | |
| 0,9 | 797 | 1,9 | 993 | | |

Bl. 144v Wenn $x$ dem Exponentialgesetz $\Phi(hx)$ genügt, so ist die Streuung oder der *mittlere Fehler* $m = \frac{1}{\sqrt{2}\,h} = 0,7071 \cdot \frac{1}{h}$; $\Phi(hx) = \Phi(0,7071 \cdot \frac{x}{m})$. Die Wahrscheinlichkeit, dass $x$ absolut $\leq m,\ 2m,\ 3m,\ \ldots$ sei, ist $\Theta(0,7071) = 0,6821,\ \Theta(1,4142) = 0,9545,\ \Theta(2,1213) = 0,9973,\ \ldots$

*Wahrscheinlichen Fehler* $w$ nennt man bisweilen den, der von $|x|$ mit gleicher Wahrscheinlichkeit über- wie unterschritten werden kann; er bestimmt sich aus $\Theta(0,7071\,\frac{w}{m}) = \frac{1}{2},\ 0,7071\,\frac{w}{m} = 0,4769,\ w = 0,6745\,m$. *Durchschnittlicher Fehler* $d$ heisst das erste Absolutment

$$d = M\,|x| = \frac{1}{\sqrt{\pi}\,h} = \sqrt{\frac{2}{\pi}}\,m = 0,7979\,m. \quad w = 0,8453\,d.$$

Bl. 145 *Die Hermiteschen Polynome.* $\dfrac{d^n\,(e^{-x^2})}{d\,x^n}$ ist $= e^{-x^2}$ mal Polynom $n^{\text{ten}}$ Grades, z. B.

$$\frac{d\,(e^{-x^2})}{d\,x} = e^{-x^2}(-2x), \quad \frac{d^2\,(e^{-x^2})}{d\,x^2} = e^{-x^2}(4x^2-2),$$

$$\frac{d^3\,(e^{-x^2})}{d\,x^3} = e^{-x^2}(-8x^3+12x), \quad \ldots$$

setzen wir $(n = 0, 1, \ldots)$

$$\frac{d^n\,(e^{-x^2})}{d\,x^n} = (-1)^n\,2^n\,e^{-x^2}\,f_n(x), \tag{4}$$

so dass die Recursionsformel

$$f_{n+1} = xf_n - \frac{1}{2}f_n'$$

besteht. Diese $f_n$ (bisweilen noch mit andern Zahlfaktoren; hier sind diese so gewählt, dass $f_n = x^n + \cdots$ den höchsten Coefficienten 1 hat) heissen die Hermiteschen Polynome. Bl. 146

Nach dem Taylorschen Satz ist

$$e^{-(x-\frac{t}{2})^2} = \sum_0^\infty \frac{d^n\,(e^{-x^2})}{d\,x^n} \cdot \frac{(-1)^n\,t^n}{2^n\,n!} = e^{-x^2} \sum_0^\infty f_n(x) \frac{t^n}{n!},$$

also

$$e^{xt-\frac{t^2}{4}} = \sum_0^\infty f_n(x) \frac{t^n}{n!} \tag{5}$$

Durch partielle Integration erhält man für jedes Polynom $f(x)$ $(D = \frac{d}{d\,x})$

$$\int_{-\infty}^{\infty} D^n(e^{-x^2})\,f(x)\,dx = -\int_{-\infty}^{\infty} D^{n-1}(e^{-x^2})\,Df(x)\,dx$$

$$= \int_{-\infty}^{\infty} D^{n-2}(e^{-x^2})\,D^2f(x)\,dx = \cdots = (-1)^n \int_{-\infty}^{\infty} e^{-x^2}\,D^nf(x)\,dx,$$

also 0, wenn $f(x)$ vom Grade $< n$ ist, d. h.

$$\int_{-\infty}^{\infty} e^{-x^2}\,f_n(x)\,f(x)\,dx = 0,$$

$f_n(x)$ ist zu jedem Polynom niedern als $n^{\text{ten}}$ Grades „orthogonal“ (zwei Functionen heissen bezüglich der monotonen Function $\varphi$ orthogonal, wenn $M\,fg = \int f(x)g(x)\,d\varphi(x) = 0$); insbesondere ist für $n > m$

$$\int_{-\infty}^{\infty} e^{-x^2}\,f_n(x)f_m(x)\,dx = 0. \tag{6}$$

Bl. 147 | Für $f(x) = f_n(x)$ ergiebt sich

$$(-1)^n\,2^n \int_{-\infty}^{\infty} e^{-x^2} f_n(x)^2\,dx = (-1)^n\,n! \int_{-\infty}^{\infty} e^{-x^2}\,dx$$

$$\int_{-\infty}^{\infty} e^{-x^2} f_n(x)^2\,dx = \sqrt{\pi}\,\frac{n!}{2^n}, \tag{7}$$

also bei der Vertheilung (1) ist $M\,f_n(x)^2 = \dfrac{n!}{2^n}$.

Wenn eine Function $\vartheta(x)$ (z. B. eine Wahrscheinlichkeitsdichte) die Reihenentwicklung

$$\vartheta(x) = \frac{1}{\sqrt{\pi}} e^{-x^2} \sum_0^\infty c_n\,f_n(x)$$

nach Hermiteschen Polynomen (resp. nach $e^{-x^2}$ und seinen Ableitungen) in der Weise zulässt, dass die Reihe, auch nach Multiplication mit Potenzen von $x$, gliedweise von $-\infty$ bis $\infty$ integrirt werden kann, so ist

$$\int_{-\infty}^{\infty} \vartheta(x)\,f_n(x)\,dx = c_n\,\frac{n!}{2^n},$$

die Reihe lautet also

$$\begin{aligned}\vartheta(x) &= \frac{1}{\sqrt{\pi}} e^{-x^2} \sum_0^\infty \frac{2^n}{n!}\, f_n(x) \int_{-\infty}^{\infty} f_n(x)\, \vartheta(x)\, dx \\ &= \frac{1}{\sqrt{\pi}} \sum_0^\infty \frac{(-1)^n}{n!}\, D^n(e^{-x^2}) \int_{-\infty}^{\infty} f_n(x)\, \vartheta(x)\, dx, \qquad (8)\end{aligned}$$

$$\int_{-\infty}^{x} \vartheta(x)\, dx = \sum_0^\infty \frac{(-1)^n}{n!}\, D^n \Phi(x) \int_{-\infty}^{\infty} f_n(x)\, \vartheta(x)\, dx \qquad (9)$$

Entsprechend für eine monotone Function mit $\varphi(-\infty) = 0$ Bl. 148

$$\varphi(x) = \sum_0^\infty \frac{(-1)^n}{n!}\, D^n \Phi(x) \int_{-\infty}^{\infty} f_n(x)\, d\varphi(x) \qquad (10)$$

Die Gültigkeitsbedingungen für diese *Brunssche Reihe* sind noch nicht vollständig geklärt. Ihre Partialsummen, d. h. Ausdrücke der Form

$$\varphi_n(x) = \sum_0^n C_\nu\, D^\nu\, \Phi(x)$$

haben sich als Darstellungen von Vertheilungsfunctionen praktisch vielfach bewährt; ja bereits das einfache Exponentialgesetz $\Phi(x)$ (für eine Variable mit $\mu_1 = 0$ und $\mu_2 = \frac{1}{2}$) ist gewissermassen ein Normaltypus für die meisten Vertheilungen. Woher kommt das?

Gauss macht das Exponentialgesetz durch folgende Betrachtung plausibel, die er selbst allerdings später als „Metaphysik“ bezeichnet. $x$ sei etwa ein Beobachtungsfehler; d. h. bei einer Variablen $\xi$ (die durch Messungen bestimmt wird – Zufallsspiel) sei ein Werth $\xi_0$ als der „wahre Werth“ ausgezeichnet und $x = \xi - \xi_0$. $\vartheta(x)$ sei die Wahrscheinlichkeitsdichte; wir nehmen sie als positiv und mit stetiger Ableitung versehen an, ferner möge sie von $\xi_0$ unabhängig sein. Angenähert ist $\vartheta(x) \cdot \Delta$ die Wahrscheinlichkeit, dass der Fehler in einem kleinen die Stelle $x$ umgebenden Intervall von der Länge $\Delta$ liege. Werden $n$ unabhängige Messungen gemacht, so ist die Wahrscheinlichkeit

$$\prod_1^n \vartheta(x_i)\, \Delta_i,$$

dass die Fehler in kleinen Intervallen $\Delta_1, \ldots, \Delta_n$ um die Stellen $x_1, \ldots, x_n$ Bl. 149
liegen; hierbei ist $x_i = \xi_i - \xi_0$. Kennen wir nun $\xi_1, \ldots, \xi_n$ durch wirklich ausgeführte Messungen, nicht aber $\xi_0$, so wird nach der Bayesschen Regel die Wahrscheinlichkeit, dass $\xi_0$ der wahre Wert sei, mit

$$\prod_1^n \vartheta(x_i)$$

proportional, und der *wahrscheinlichste* Werth für $\xi_0$, für den dieses Produkt möglichst gross ist, muss der durch logarithmische Differentiation nach $\xi_0$ entstehenden Gleichung

$$\sum_1^n \psi(x_i) = 0 \quad (\psi(x) = \frac{\vartheta'(x)}{\vartheta(x)})$$

genügen. Nun machen wir die Annahme: der wahrscheinlichste Werth von $\xi_0$ sei das *arithmetische Mittel*

$$\frac{1}{n}(\xi_1 + \cdots + \xi_n).$$

Dann muss also $\sum_1^n \psi(x_i) = 0$ sein, sobald $\sum_1^n x_i = 0$ ist. Für $n = 2$ erhält man $\psi(x) + \psi(-x) = 0$, $\psi$ eine ungerade Function, für $n = 3$:

$$\psi(x) + \psi(y) = \psi(x + y),$$

und diese Functionalgleichung wird bei stetigem $\psi$ nur durch $\psi(x) = 2ax$ erfüllt ($a$ const.). Das giebt $\log \vartheta(x) = ax^2 + b$ oder $\vartheta(x) = c\, e^{ax^2}$. Damit $\int_{-\infty}^{\infty} \vartheta(x)\, dx$ existire und $= 1$ sei, muss $a = -h^2$ negativ und $c = \frac{h}{\sqrt{\pi}}$ sein[6],

$$\vartheta(x) = \frac{h}{\sqrt{\pi}}\, e^{-h^2 x^2}.$$

Aber weder dies noch die *Erhaltung* des Exponentialgesetzes würden seine Bedeutung erklären; diese liegt vielmehr in einem *Grenzwerthsatz*, der das früher elementar abgeleitete *Gesetz der grossen Zahlen* dahin verschärft, dass unter ähnlichen Voraussetzungen wie damals die *Summe unabhängiger Variabler* eine nach dem Exponentialgesetz convergirende Vertheilung hat. | Bl. 150

### *Der Grenzwerthsatz.*

Die allgemeinste Fassung, mit den geringsten Voraussetzungen, stammt von *A. Liapunoff* (C. R. 132 (1900), S. 126–127, S. 814–815; Bull. Ac. Petersb. (5) 13 (1900), S. 359–386); der einfachste Beweis dieses Liapunoffschen Satzes von *J. W. Lindeberg* (Math. Ztschr. 15 (1922), S. 211–225).

$x$ sei eine Variable mit der Vertheilungsfunction $\varphi(x)$ (links stetig, $\varphi(-\infty) = 0$, $\varphi(+\infty) = 1$) und

$$\int d\varphi = 1, \quad \int x\, d\varphi = 0, \quad \int x^2\, d\varphi = a^2, \quad \int |x|^3\, d\varphi = c^3, \qquad (11)$$

---

[6] Zum Vergleich: ist die Wahrscheinlichkeitsdichte $c\, e^{-h|x|}$, so ist der wahrscheinlichste Werth $\xi_0$ der, für den $|\xi_1 - \xi_0| + \cdots + |\xi_n - \xi_0|$ ein Minimum ist. Man sieht, wenn man die $\xi_1, \ldots, \xi_n$ paarweise verschieden annimmt und der Grösse nach ordnet: $\xi_0$ ist bei ungeradem $n$ der mittelste Werth, bei geradem irgend ein Werth im Intervall zwischen den beiden mittelsten.

höhere Momente brauchen nicht zu existiren ($a$ die Streuung). Ist $g(x)$ eine nebst ihren ersten drei Ableitungen stetige und beschränkte Function, kurz eine Function dritter Klasse, so ist nach § 5 (am Ende) auch

$$f(t) = \int g(t-x)\, d\varphi(x)$$

eine solche (und zwar $|f'''| \leq M$, wenn $|g'''| \leq M$). Wir haben nun

$$g(t-x) = g(t) - x\, g'(t) + \frac{x^2}{2}\, g''(t) - \frac{x^3}{6}\, g'''(t-\xi),$$

wo $\xi$ zwischen 0 und $x$ liegt (auch von $t$ abhängt) und also durch Integration

$$|f(t) - g(t) - \frac{a^2}{2}\, g''(t)| \leq \frac{1}{6}\, M\, c^3$$

$\Phi(x)$ sei eine weitere monotone Function derselben Art, *mit der gleichen Streuung* $a$, Bl. 151

$$\int d\Phi = 1, \quad \int x\, d\Phi = 0, \quad \int x^2\, d\Phi = a^2, \quad \int |x|^3\, d\Phi = C^3,$$

$$F(t) = \int g(t-x)\, d\Phi(x),$$

so ist auch

$$|F(t) - g(t) - \frac{a^2}{2}\, g''(t)| \leq \frac{1}{6}\, M\, C^3,$$

also

$$|f(t) - F(t)| \leq \frac{1}{6}\, M\, (c^3 + C^3).$$

Nehmen wir für $\Phi$ das Exponentialgesetz $\Phi(hx)$ (jetzt in der Bezeichnung (1)), dessen Streuung $= \dfrac{1}{2h^2} = a^2$, also $h = \dfrac{1}{\sqrt{2}\, a}$; nennen wir dies etwa das $\varphi$ osculirende Exponentialgesetz. Dann ist $C^3 = \dfrac{1}{\sqrt{\pi}\, h^3}$ (s. die obige Formel für $\nu_3$) $= \sqrt{\frac{2}{\pi}} \cdot 2a^3$. Überdies hat man für jedes $b > 0$

$$a^2 = \left( \int_{-\infty}^{-b} + \int_{-b}^{b} + \int_{b}^{\infty} \right) x^2\, d\varphi \leq \frac{1}{b} \left( \int_{-\infty}^{-b} + \int_{b}^{\infty} \right) |x^3|\, d\varphi + b^2 \leq \frac{c^3}{b} + b^2,$$

$$c^3 \geq b(a^2 - b^2)$$

z. B. für

$$b = \frac{a}{\sqrt{3}} : \quad c^3 \geq \frac{2}{\sqrt{27}}\, a^3, \quad C^3 \leq \sqrt{\frac{2}{\pi} \cdot 27}\, c^3 < 5c^3,$$

also

$$|f(t) - F(t)| \leq M\,c^3. \tag{12}$$

Hier ist also $\varphi$ eine monotone Function mit (11), $g(x)$ von dritter Klasse mit $|g'''| \leq M$, $\quad \Phi(hx)$ mit $h = \frac{1}{\sqrt{2}\,a}$ das $\varphi$ osculirende Exponentialgesetz,

$$f(t) = \int g(t-x)\,d\varphi(x), \qquad F(t) = \int g(t-x)\,d\Phi(hx). \tag{13}$$

Bl. 152 | Sodann seien $(n = 1, 2, \ldots)$ $\varphi_n$ lauter monotone Functionen dieser Art (11) mit

$$\int d\varphi_n = 1, \quad \int x\,d\varphi_n = 0, \quad \int x^2\,d\varphi_n = a_n^2, \quad \int |x^3|\,d\varphi_n = c_n^3; \tag{14}$$

$x_1, x_2, \ldots$ *unabhängige* Variable mit den Vertheilungsfunctionen $\varphi_n$; ferner $y_n = x_1 + x_2 + \cdots + x_n$ mit der Vertheilungsfunction $\psi_n$. Setzen wir

$$f_n(t) = \int g(t-x)\,d\psi_n(x) \tag{15}$$

so ist nach § 6, (23) ($y_n = y_{n-1} + x_n$ Summe von 2 unabhängigen Variablen) unter Verwandlung des Doppelintegrals in ein iterirtes, für $n > 1$

$$f_n(t) = \int g(t - y_n)\,d\psi_n(y_n) = \int\int g(t - y_{n-1} - x_n)\,d\psi_{n-1}(y_{n-1})\,d\varphi_n(x_n)$$

$$= \int f_{n-1}(t - x_n)\,d\varphi_n(x_n) = \int f_{n-1}(t-x)\,d\varphi_n(x),$$

welche Recursionsformel auch für $n = 1$ gilt, wenn man $f_0 = g$ setzt. Sind ferner $\Phi_n(x) = \Phi(h_n x)$ mit $h_n = 1 : \sqrt{2}\,a_n$ die $\varphi_n$ osculirenden Exponentialvertheilungen, $X_n$ unabhängige Variable mit diesen Vertheilungsfunctionen $\Phi_n$ und $Y_n = X_1 + \cdots + X_n$ mit der Vertheilungsfunction $\Psi_n$, so ist mit

$$F_n(t) = \int g(t-x)\,d\Psi_n(x) \tag{16}$$

ebenso

$$F_n(t) = \int F_{n-1}(t-x)\,d\Phi_n(x) \qquad\qquad (F_0 = g).$$

Bl. 153 | Alle Functionen $f_n$, $F_n$ sind von dritter Klasse und ihre dritten Ableitungen absolut $\leq M$. Wenn nun bereits die Abschätzung

$$|f_{n-1}(t) - F_{n-1}(t)| \leq M_{n-1}$$

bekannt ist, so erhält man unter Vermittlung der Functionen

$$g_n(t) = \int f_{n-1}(t-x)\,d\Phi_n(x)$$

$$|g_n(t) - F_n(t)| \leq M_{n-1},$$

nach (12) aber

$$|f_n(t) - g_n(t)| \leq M\, c_n^3,$$

$$|f_n(t) - F_n(t)| \leq M_{n-1} + M\, c_n^3$$

und also, da $M_0 = 0$

$$|f_n(t) - F_n(t)| \leq M \sum_1^n c_\nu^3 = M\, d_n^3 \tag{17}$$

Hierbei ist aber $\Psi_n(x) = \Phi(k_n x)$ selbst wieder das Exponentialgesetz mit dem Streuungsquadrat

$$b_n^2 = \sum_1^n a_\nu^2 \tag{18}$$

also $k_n = 1 : \sqrt{2}\, b_n$; es ist das $\psi_n$ osculirende Exponentialgesetz.

Die Abschätzung (17), wo $f_n$, $F_n$ durch (15) (16) erklärt sind, kann nun wegen der Willkürlichkeit von $g$ in eine für die Functionen $\psi_n, \Psi_n$ selbst verwandelt werden. Nehmen wir eine feste Function dritter Klasse $\gamma(x)$, die in $(-\infty, 0]$ gleich 0, in $[1, \infty)$ gleich 1 und in $(0, 1)$ zwischen 0 und 1 gelegen ist, z. B. | Bl. 154

$$\gamma(x) = \int_0^x x^3(1-x)^3\, dx : \int_0^1 x^3(1-x)^3\, dx \qquad \text{in } [0,1];$$

es ist dann $|\gamma'''(x)| \leq m$, wo $m$ eine numerische Constante ist. Für $l > 0$, $g(x) = \gamma(\frac{x}{l})$, welche Function 3. Klasse in $(-\infty, 0]$ gleich 0, in $[l, \infty)$ gleich 1 und in $(0, l)$ zwischen 0 und 1 gelegen ist, ist $g'''(x) = \dfrac{1}{l^3}\, \gamma'''(x)$, so dass die Constante $M$ in (17) durch $\dfrac{m}{l^3}$ ersetzt werden kann. Betrachtet man den Verlauf der Functionen $g(t-x)$, $g(t+l-x)$ und $g(t+l-x) - g(t-x)$ als Functionen von $x$, so erhält man für jede monotone Function $\psi(x)$ mit $\psi(-\infty) = 0$

$$\int g(t-x)\,d\psi(x) \leq \psi(t) \leq \int g(t-x+l)\,d\psi(x),$$

also für $\psi_n$ und $\Psi_n$

$$f_n(t) \leq \psi_n(t) \leq f_n(t+l) \qquad F_n(t) \leq \Psi_n(t) \leq F_n(t+l),$$

während zugleich

$$0 \leq F_n(t+l) - F_n(t) \leq \int_{t-l}^{t+l} d\Psi_n = \frac{k_n}{\sqrt{\pi}} \int_{t-l}^{t+l} e^{-k_n^2 x^2}\,dx < \frac{2}{\sqrt{\pi}}\,k_n\,l = \sqrt{\frac{2}{\pi}}\,\frac{l}{b_n}$$

Bl. 155 | ist. Diese Ungleichungen mit

$$|f_n(t) - F_n(t)| \leq \frac{m}{l^3}\,d_n^3, \qquad |f_n(t+l) - F_n(t+l)| \leq \frac{m}{l^3}\,d_n^3$$

verbunden geben

$$|\psi_n(t) - \Psi_n(t)| \leq m\left(\frac{d_n}{l}\right)^3 + \sqrt{\frac{2}{\pi}}\,\frac{l}{b_n}$$

Hier kann man noch über $l$ passend verfügen, etwa indem man das Minimum der rechten Seite aufsucht; man erhält

$$l = b_n^{\frac{1}{4}}\,d_n^{\frac{3}{4}}$$

mal einer numerischen Constante und schliesslich

$$|\psi_n(y) - \Phi(k_n y)| \leq \mu \cdot \left(\frac{d_n}{b_n}\right)^{\frac{3}{4}} \tag{19}$$

mit einer numerischen Constante $\mu$ oder mit $y = \dfrac{z}{k_n} = \sqrt{2}\,b_n\,z$, wobei

$$\psi_n(y) = \psi_n(\sqrt{2}\,b_n z) = \Phi_n(z) \tag{20}$$

die Wahrscheinlichkeit ist, dass $y_n < y$ oder dass die Variable

$$z_n = \frac{y_n}{\sqrt{2}\,b_n} < z \tag{21}$$

sei:

$$|\Phi_n(z) - \Phi(z)| \leq \mu \cdot \left(\frac{d_n}{b_n}\right)^{\frac{3}{4}}. \tag{22}$$

Es ist hier vielleicht folgende Benennung zweckmässig: eine Variable, deren erste Bl. 156 Momente $\mu_0 = 1,\ \mu_1 = 0,\ \mu_2 = \frac{1}{2}$ sind, heisse *normirt*; | aus jeder Variablen

$x$ kann man durch lineare Substitution eine und nur eine normirte Variable $\alpha x + \beta$ $(\alpha > 0)$ bilden, nämlich

$$\frac{x - \mu_1}{\sqrt{2(\mu_2 - \mu_1^2)}},$$

die zu $x$ gehörige normirte Variable.Wenn eine Variable einem Exponentialgesetz (2) folgt, so folgt die zugehörige normirte Variable dem normirten Exponentialgesetz

$$\Phi(x) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{x} e^{-x^2}\, dx.$$

Zu $y_n$ ist $z_n$ die zugehörige normirte Variable.

Wir können unsere Ergebnisse nun folgendermassen aussprechen:

I. (*Grenzwerthsatz von Liapunoff*) *Es seien* $x_1, x_2, \ldots$ *unabhängige Variable, deren Vertheilungsfunctionen* $\varphi_n$ (*d. h.* $\varphi_n(x) = w(x_n < x)$) *den Gleichungen* (14) *genügen,*

$$b_n^2 = a_1^2 + \cdots + a_n^2, \qquad d_n^3 = c_1^3 + \cdots + c_n^3,$$

*ferner* $y_n = x_1 + x_2 + \cdots + x_n$ *und* $z_n = \dfrac{y_n}{\sqrt{2}\, b_n}$ *die zugehörige normirte Variable,* $\Phi_n(z)$ *deren Vertheilungsfunction* (*d. h.* $= w(z_n < z)$). *Es gilt dann mit dem Exponentialgesetz*

$$\Phi(z) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{z} e^{-z^2}\, dz$$

*die Abschätzung*

$$|\Phi_n(z) - \Phi(z)| \leq \mu \cdot \left(\frac{d_n}{b_n}\right)^{\frac{3}{4}} \qquad (\mu \text{ eine num. Const.})$$

*und falls also* $\dfrac{d_n}{b_n} \to 0$, *so convergirt* $\Phi_n(z)$ *nach* $\Phi(z)$ (*gleichmässig für alle* $z$).

Aus der Gleichmässigkeit folgt, dass auch $\Phi_n(z + 0) = w(z_n \leq z)$ nach $\Phi(z + 0) = \Phi(z)$ convergirt. | Bl. 157

Die Voraussetzung $\dfrac{d_n}{b_n} \to 0$ trifft z. B. zu, wenn alle Variable dieselbe Vertheilung haben:

$$a_n = a, \quad c_n = c, \quad b_n = n^{\frac{1}{2}}\, a, \quad d_n = n^{\frac{1}{3}}\, c, \quad \frac{d_n}{b_n} = \frac{1}{n^{\frac{1}{6}}}\, \frac{c}{a}.$$

Wir haben in (14) die Voraussetzung gemacht, dass alle Variablen $x_n$ die ersten Momente $\int x\, d\varphi_n = 0$ haben. Handelt es sich um unabhängige Variable

$\xi_n$ mit den wahrscheinlichen Werthen $M(\xi_n) = \alpha_n$, so hat man natürlich wie in § 2, (12) zu setzen

$$\left.\begin{aligned} x_n = \xi_n - \alpha_n, \quad \eta_n = \xi_1 + \cdots + \xi_n, \\ \beta_n = \alpha_1 + \cdots + \alpha_n, \quad y_n = \eta_n - \beta_n = x_1 + \cdots + x_n. \end{aligned}\right\} \tag{23}$$

Über das Auftreten der dritten Absolutmomente $c_n^3 = \int |x|^3 \, d\varphi_n$ ist noch ein Wort zu sagen. Liapunoff hat bewiesen, dass man hierin den Exponenten 3 durch jeden (auch nicht ganzen) Exponenten $\alpha > 2$ ersetzen kann, d. h. in unserer Bedingung $\frac{d_n}{b_n} \to 0$

$$d_n^\alpha = c_1^\alpha + \cdots + c_n^\alpha, \qquad c_n^\alpha = \int |x|^\alpha \, d\varphi_n.$$

Diese Bedingung sagt um so weniger und der Liapunoffsche Satz ist also um Bl. 158 so inhaltreicher, je kleiner $\alpha$ genommen wird. Nämlich: | Sei $f(\alpha)$ in einem (offenen) Intervall zweimal differenzirbar und nach oben *concav*, d. h. $f''(\alpha) \geq 0$. Dann ist für $\alpha < \beta < \gamma$

$$\frac{f(\beta) - f(\alpha)}{\beta - \alpha} - \frac{f(\gamma) - f(\beta)}{\gamma - \beta} = f'(\xi) - f'(\eta) \leq 0$$

oder

$$f(\beta)(\gamma - \alpha) \leq f(\alpha)(\gamma - \beta) + f(\gamma)(\beta - \alpha).$$

Sind $p_1, \ldots, p_n, \; u_1, \ldots, u_n$ positiv,

$$g(\alpha) = p_1 \, u_1^\alpha + \cdots + p_n \, u_n^\alpha,$$

so ist

$$g'(\alpha) = \sum_{\nu=1}^{n} p_\nu \, u_\nu^\alpha \, \log u_\nu, \quad g''(\alpha) = \sum p_\nu \, u_\nu^\alpha \, (\log u_\nu)^2, \quad g g'' - g'^2 \geq 0$$

(weil $\sum x_\nu^2 \cdot \sum y_\nu^2 \geq (\sum x_\nu y_\nu)^2$; die quadratische Form $\sum (x_\nu \xi + y_\nu \eta)^2$ ist $\geq 0$), also $\log g(\alpha)$ concav, folglich

$$g(\beta)^{\gamma - \alpha} \leq g(\alpha)^{\gamma - \beta} \, g(\gamma)^{\beta - \alpha}$$

Durch Grenzübergang dehnt sich dies auch auf unendliche Summen

$$g(\alpha) = p_1 \, u_1^\alpha + p_2 \, u_2^\alpha + \cdots ,$$

ferner (Skalenfunctionen!) auf Stieltjesintegrale $g(\alpha) = \int u(x)^\alpha \, d\varphi(x)$ nach einer monotonen Function aus, z. B. für $\alpha > 0$ auf $g(\alpha) = \int |x|^\alpha \, d\varphi$, und ebenso auf Summen solcher $g_n(\alpha) = \int |x|^\alpha \, d\varphi_1 + \cdots + \int |x|^\alpha \, d\varphi_n$. Schreibt man dafür $g_n(\alpha)^{\frac{1}{\alpha}} = h_n(\alpha)$, so hat man also

$$h_n(\beta)^{\beta(\gamma-\alpha)} \leq h_n(\alpha)^{\alpha(\gamma-\beta)} \, h_n(\gamma)^{\gamma(\beta-\alpha)}$$

oder

$$\left[\frac{h_n(\beta)}{h_n(\alpha)}\right]^{\beta(\gamma-\alpha)} \leq \left[\frac{h_n(\gamma)}{h_n(\alpha)}\right]^{\gamma(\beta-\alpha)}$$

so dass ($n \to \infty$) mit $\dfrac{h_n(\gamma)}{h_n(\alpha)}$ auch $\dfrac{h_n(\beta)}{h_n(\alpha)}$ nach 0 convergirt. Unsere Voraus- Bl. 159
setzung war $\dfrac{h_n(3)}{h_n(2)} \to 0$, dann ist auch $\dfrac{h_n(\beta)}{h_n(2)} \to 0$ für $2 < \beta < 3$; aus dem Liapunoffschen Satz für $\beta$ folgt also der für den Exponenten 3. Umgekehrt folgt aus dem für 3 auch der für jeden Exponenten $> 3$; z. B. kann man also in I statt der dortigen $c_n$, $d_n$ setzen

$$c_n^4 = \int x^4 \, d\varphi_n, \qquad d_n^4 = c_1^4 + \cdots + c_n^4.$$

Der Grenzwerthsatz gilt gewiss, *wenn die Variablen $x_n$ gleichmässig beschränkt sind und die Quadrate ihrer Streuungen eine divergente Reihe bilden.* D. h. es soll eine feste Zahl $M$ geben derart, dass alle ausserhalb $[-M, M]$ gelegenen Mengen das Mass 0 haben, will sagen

$$\varphi_n(x) = 0 \text{ für } x \leq -M, \qquad \varphi_n(x) = 1 \text{ für } x \geq M,$$

und ausserdem $b_n \to \infty$ sein. Dann ist

$$c_n^3 = \int_{-M}^{M} |x|^3 \, d\varphi_n \leq M \int_{-M}^{M} x^2 \, d\varphi_n = M \, a_n^2, \;\; d_n^3 \leq M \, b_n^2 \text{ und } \frac{d_n^3}{b_n^3} \leq \frac{M}{b_n} \to 0$$

Sei z. B. (vgl. § 3, nach Satz I) $\xi_n$ eine Variable, die mit den Wahrscheinlich- Bl. 160
keiten $p_n$ und $q_n = 1 - p_n$ die Werthe 1 und 0 annimmt; $A_n = (\xi_n = 1)$ der „günstige Fall“ des $n^{\text{ten}}$ Versuches, $\eta_n$ die Häufigkeit des günstigen Falles bei den ersten $n$ Versuchen;

$$\alpha_n = p_n, \quad a_n^2 = p_n q_n, \quad x_n = \xi_n - p_n;$$

$$\beta_n = \sum_1^n p_\nu, \quad y_n = \eta_n - \beta_n, \quad b_n^2 = \sum_1^n p_\nu q_\nu.$$

Die Variablen $\xi_n$ oder $x_n$ sind gleichmässig beschränkt; $b_n \to \infty$ ist z. B. erfüllt, wenn stets $0 < \varepsilon \leq p_n \leq 1 - \varepsilon < 1$. Wir erhalten also folgende Verschärfung des *Poissonschen Satzes* § 3, II:

II. *Es werde eine Folge unabhängiger Versuche gemacht; beim* $n^{\text{ten}}$ *Versuch sei* $p_n$ *die Wahrscheinlichkeit des günstigen Falls und* $\sum_1^\infty p_n q_n$ *sei divergent. Ist* $\eta_n$ *die Häufigkeit des günstigen Falls bei den ersten* $n$ *Versuchen,*

$$\beta_n = \sum_1^n p_\nu, \quad b_n = \sqrt{\sum_1^n p_\nu q_\nu},$$

*so convergirt die Vertheilungsfunction* $\Phi_n(z)$ *der zu* $\eta_n$ *gehörigen normirten Variablen* $z_n = \dfrac{\eta_n - \beta_n}{\sqrt{2}\, b_n}$ *nach dem Exponentialgesetz* $\Phi(z)$.

Und für $p_n = p$ die Verschärfung des *Bernoullischen Satzes* § 3, III:

III. *Es werde eine Folge unabhängiger Versuche gemacht; bei jedem Versuch sei*
$p$ $(0 < p < 1)$ *die Wahrscheinlichkeit des günstigen Falls. Ist* $\eta_n$ *die Häufigkeit* |
Bl. 161 *des günstigen Falls bei den ersten* $n$ *Versuchen, so convergirt die Vertheilungs-*
*function* $\Phi_n(z)$ *der zu* $\eta_n$ *gehörigen normirten Variablen* $z_n = \dfrac{\eta_n - np}{\sqrt{2npq}}$ *nach*
Bl. 162 *dem Exponentialgesetz* $\Phi(z)$. |

*Beispiele* für approximative Gültigkeit des Exponentialgesetzes.
Vorbemerkung. Die wirkliche (theoretische oder beobachtete) Vertheilung einer Variablen $x$, die mit dem Exponentialgesetz $\Phi(x)$ verglichen werden soll, ist in der Regel *elementar* (§ 1), d. h. $x$ nimmt nur endlich viele Werthe $x_k$ mit rationalen Wahrscheinlichkeiten $p_k = \dfrac{s_k}{s}$ an, wo $s_k$ und $s = \sum s_k$ natürliche Zahlen sind; unter $s$ möglichen oder beobachteten Fällen wird $s_k$ mal der Werth $x_k$ angenommen. Ferner sind in der Regel die $x_k = k\,\delta$ $(k = 0, \pm 1, \ldots)$ ganze Vielfache einer Einheit $\delta$. Einen solchen Werth wird man zweckmässig als Repräsentanten des Intervalls $[(k - \frac{1}{2})\delta, (k + \frac{1}{2})\delta]$ ansehen, also $p_k$ mit

$$\Phi_{(k-\frac{1}{2})\delta}^{(k+\frac{1}{2})\delta} \quad \text{oder } s_k \text{ mit} \quad s \cdot \Phi_{(k-\frac{1}{2})\delta}^{(k+\frac{1}{2})\delta}$$

vergleichen.

Endlich wird man wegen der Symmetrie des Exponentialgesetzes (EG) die Werthe $\pm k$ zusammennehmen, wobei

$$\Phi_{(k-\frac{1}{2})\delta}^{(k+\frac{1}{2})\delta} + \Phi_{(-k-\frac{1}{2})\delta}^{(-k+\frac{1}{2})\delta} = \Theta_{(k-\frac{1}{2})\delta}^{(k+\frac{1}{2})\delta}$$

und für $k = 0$ $\quad \Phi_{-\frac{1}{2}\delta}^{\frac{1}{2}\delta} = \Theta(\frac{1}{2}\delta)$ ist; man wird also vergleichen:

| $\lvert k\rvert$ | 0 | 1 | 2 | $\cdots$ |
|---|---|---|---|---|
| wirkl. Vertheilung : | $s_0$ | $s_{-1} + s_1$ | $s_{-2} + s_2$ | $\cdots$ |
| Vertheilung nach d. EG : | $s\,\Theta(\frac{1}{2}\delta)$ | $s\,\Theta_{\frac{1}{2}\delta}^{\frac{3}{2}\delta}$ | $s\,\Theta_{\frac{3}{2}\delta}^{\frac{5}{2}\delta}$ | $\cdots$ |

*Beispiel zu III.* $\quad p = q = \frac{1}{2}$, $n = 10$. Die Wahrscheinlichkeit, dass der günstige Fall $m$ mal eintrete, ist $\binom{10}{m}\frac{1}{2^{10}}$. Die zu $m$ gehörige normirte Variable ist $z_{10} = \frac{m-5}{\sqrt{5}} = x$; sie nimmt wirklich nur die Werthe $\frac{k}{\sqrt{5}}$ $(k = m - 5,$ | Bl. 163
$k = 0, \pm 1, \ldots, \pm 5)$ und zwar unter $s = 1024 = 2^{10}$ Fällen $\binom{10}{k+5}$ mal an. Wir haben zu vergleichen:

| $\lvert k\rvert$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| wirkl. Vertheilung: | $\binom{10}{5}$ | $2\binom{10}{4}$ | $2\binom{10}{3}$ | $2\binom{10}{2}$ | $2\binom{10}{1}$ | $2\binom{10}{0}$ |
| Verth. nach d. EG: | $2^{10}\,\Theta\left(\frac{1}{2\sqrt{5}}\right)$ | $2^{10}\Theta_{\frac{1}{2\sqrt{5}}}^{\frac{3}{2\sqrt{5}}}$ |  | $\cdots$ |  |  |

Man erhält $\frac{1}{2\sqrt{5}} = 0,2236$

| $x$ | $0,2236$ | $0,6708$ | $1,118$ | $1,565$ | $2,013$ | $\infty$ |
|---|---|---|---|---|---|---|
| $\Theta(x)$ | $0,2482$ | $0,6572$ | $0,8861$ | $0,9731$ | $0,9956$ | $1$ |
| $2^{10}\,\Theta(x)$ | $254,2$ | $673,0$ | $907,4$ | $996,4$ | $1019,5$ | $1024$ |

|  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|
| Wirkl. Vertheilung : | 252 | 420 | 240 | 90 | 20 | 2 |
| EG : | $254,2$ | $418,8$ | $234,4$ | $89,0$ | $23,1$ | $4,5$ |

*Beispiel einer beobachteten Vertheilung.* Hagen zählte in einem Text die Häufigkeit des Buchstaben e. In $s = 110$ Zeilen fand sich e 886 mal, also durchschnittlich 8 mal (genau 8,055); versuchen wir, ob die Abweichung $k$ der Zeilenhäufigkeit von 8 das Gausssche Gesetz $\Phi(hk)$ annähernd befolgt oder ob angenähert

$$s\,\Phi_{(k-\frac{1}{2})h}^{(k+\frac{1}{2})h} = s_k.$$

Die beobachteten $\lvert k\rvert$ fanden sich so vertheilt

| $\lvert k\rvert$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Zeilenzahl | 26 | 43 | 24 | 11 | 4 | 0 | 2 |

Man hat hier erst $h$ (wie schon den Nullpunkt der Zählung) passend | zu be- Bl. 164
stimmen (worüber Näheres in § 9). Man wird zu diesem Zweck eine der Gleichungen $\Theta(\frac{1}{2}h) = \frac{26}{110}$, $\quad \Theta(\frac{3}{2}h) = \frac{26+43}{110}$, … benutzen und $h$ aus der Tafel der

$\Theta$-Function ermitteln; die verschiedenen Werthe stimmen natürlich nicht genau überein. Z. B.

$$\Theta(\tfrac{1}{2}h) = 0,2364, \quad \tfrac{1}{2}h = 0,2126, \quad h = 0,4252$$

$$\Theta(\tfrac{3}{2}h) = 0,6273, \quad \tfrac{3}{2}h = 0,6304, \quad h = 0,4203$$

$$\Theta(\tfrac{5}{2}h) = 0,8455, \quad \tfrac{5}{2}h = 1,0066, \quad h = 0,4026$$

$$\Theta(\tfrac{7}{2}h) = 0,9455, \quad \tfrac{7}{2}h = 1,3590, \quad h = 0,3884$$

Nehmen wir z. B. $h = 0,4$

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $x$ | $0,2$ | $0,6$ | $1,0$ | $1,4$ | $1,8$ | $2,2$ | $\infty$ |
| $\Theta(x)$ | $0,2227$ | $0,6039$ | $0,8427$ | $0,9523$ | $0,9891$ | $0,9981$ | $1$ |
| $110\,\Theta(x)$ | $24,5$ | $66,4$ | $93,7$ | $104,8$ | $108,8$ | $109,8$ | $110$ |

Also

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| wirkl. Vert. | 26 | 43 | 24 | 11 | 4 | 0 | 2 |
| EG | $24,5$ | $41,9$ | $27,3$ | $11,1$ | $4,0$ | $1,0$ | $0,2$ |

Bl. 165

## § 8. Das Momentproblem. Zweiter Grenzwerthsatz.

$\varphi(x)$ sei eine monotone Function, von der sämtliche Momente

$$\mu_k = M\,x^k = \int_{-\infty}^{\infty} x^k\,d\varphi \qquad (k = 0, 1, \ldots) \tag{1}$$

existiren (insbesondere $\mu_0 = \varphi(+\infty) - \varphi(-\infty)$ endlich). Die Bestimmung von $\varphi(x)$, wenn die Momente gegeben sind, heisst das Momentproblem. Dabei kann $\varphi$ um eine additive Constante geändert werden, ferner kommt es auf die Werthe an den Sprungstellen nicht an, sondern nur auf die Grenzwerthe $\varphi(x \pm 0)$ (wir hatten ja $\int f\,d\chi = \int f\,d\varphi$ definirt, wenn $\varphi(x) = \chi(x-0)$). Nehmen wir also etwa $\varphi(-\infty) = 0$ und $\varphi$ links stetig an; wenn es dann nur eine Lösung giebt, heisst das Momentproblem *bestimmt*, andernfalls unbestimmt. Von den beiden Fragen: Lösbarkeit und Bestimmtheit wollen wir nur die letztere und zwar nur durch Aufstellung *hinreichender* Bedingungen behandeln.

Eine Stelle $x$, die im Innern eines Intervalls liegt, wo $\varphi$ constant ist, heisst eine *Constanzstelle* ($\varphi(x-h) = \varphi(x+h)$ für ein $h > 0$); andernfalls heisst sie eine *Wachsthumsstelle* ($\varphi(x-h) < \varphi(x+h)$ für jedes $h > 0$). Eine Function $\varphi$ mit nur endlich vielen Wachsthumsstellen ist, wie man sofort sieht, zwischen zwei
Bl. 166 solchen constant, also eine *Treppenfunction*, die einer elementaren Vertheilung entspricht. Wir lassen dies beiseite und nehmen an, dass $\varphi$ unendlich viele Wachsthumsstellen hat.

Mit (1) ist das Moment $M\,F(x)$ jedes Polynoms gegeben; ist durchweg $F(x) \geq 0$, so ist $M\,F(x) \geq 0$, und zwar sogar $M\,F(x) > 0$, wenn $F(x)$ nicht identisch verschwindet. Denn ist $\xi$ eine von den (endlich vielen) reellen Nullstellen von $F$ verschiedene Wachsthumsstelle von $\varphi$, so ist in einer Umgebung

$[\xi - h, \xi + h] \quad F(x) \geq \varepsilon > 0,$

$$\int_{-\infty}^{\infty} F\,d\varphi \geq \int_{\xi-h}^{\xi+h} F\,d\varphi \geq \varepsilon\,[\varphi(\xi+h) - \varphi(\xi-h)] > 0.$$

Insbesondere sei $f(x) = u_0 + u_1\,x + \cdots + u_n\,x^n$ ein reelles Polynom (mit reellen Coefficienten $u_i$, die nicht alle 0), dann ist

$$M\,f(x)^2 = \sum_0^n \mu_{i+k}\,u_i u_k,$$

also positiv, d. h. eine positiv definite quadratische Form der reellen Variablen $u_0, \ldots, u_n$ und daher bekanntlich ihre Determinante

$$D_n = \begin{vmatrix} \mu_0 & \mu_1 & \cdots & \mu_n \\ \mu_1 & \mu_2 & \cdots & \mu_{n+1} \\ \cdots & \cdots & \cdots & \cdots \\ \mu_n & \mu_{n+1} & \cdots & \mu_{2n} \end{vmatrix} > 0 \qquad (n = 0, 1, \ldots). \tag{2}$$

Sie bleibt $(n \geq 1)$ auch noch definit positiv, wenn man $u_0 = 0$ setzt, also ist auch | Bl. 167

$$E_n = \begin{vmatrix} \mu_2 & \cdots & \mu_{n+1} \\ \cdots & \cdots & \cdots \\ \mu_{n+1} & \cdots & \mu_{2n} \end{vmatrix} > 0 \qquad (n = 1, 2, \ldots) \tag{3}$$

(Die Bedingungen (2) sind auch hinreichend zur Lösbarkeit des Momentproblems durch eine monotone Function $\varphi$ mit unendlich vielen Wachsthumsstellen.)

Wir betrachten jetzt alle Lösungen mit $\varphi(-\infty) = 0$, ohne sie jedoch links stetig vorauszusetzen. Wenn sie alle an einer Stelle $\xi$ denselben (nur von den $\mu_k$ und $\xi$ abhängigen) Werth haben, heisse $\xi$ eine *Bestimmtheitsstelle* des Momentproblems, andernfalls eine Unbestimmtheitsstelle. Z. B. ist jede Sprungstelle einer Lösung eine Unbestimmtheitsstelle, da man ja $\varphi(\xi)$ jeden Werth $\geq \varphi(\xi - 0)$ und $\leq \varphi(\xi + 0)$ beilegen kann. Wenn alle Stellen Bestimmtheitsstellen sind, ist das Momentproblem bestimmt (hinreichende, aber nicht nothwendige Bedingung). Untersuchen wir zuerst die Stelle $\xi = 0$.

Wir betrachten die Polynome $F(x)$, die *durchweg* $\geq 0$ und ausserdem respective für

$$(\beta) \quad x \leq 0 \qquad\qquad (\gamma) \quad x \geq 0 \qquad\qquad (\delta) \quad x = 0$$

auch noch $\geq 1$ sind. Die unteren Grenzen der (positiven) Momente dieser Polynome seien $\beta,\ \gamma,\ \delta\ (\geq 0)$. Ist z. B. $F$ ein $\beta$-Polynom, so ist für hinreichend

kleines $h > 0 \quad F \geq 1 - \varepsilon$ in $(-\infty, h]$ und daher für jede Lösung | Bl. 168

$$M\,F(x) = \int_{-\infty}^{\infty} F\,d\varphi \geq \int_{-\infty}^{h} (1-\varepsilon)\,d\varphi = (1-\varepsilon)\,\varphi(h) \geq (1-\varepsilon)\,\varphi(+0),$$

also $M\,F(x) \geq \varphi(+0)$, $\beta \geq \varphi(+0)$. Ebenso für die $\gamma$-Polynome

$$\gamma \geq \varphi(\infty) - \varphi(-0) = \mu_0 - \varphi(-0)$$

oder $\alpha \leq \varphi(-0)$, wenn $\alpha = \mu_0 - \gamma$ gesetzt wird (man könnte entsprechend $\alpha$-Polynome $= 1 - \gamma$-Polynome einführen; d. h. durchweg $F \leq 1$ und $F \leq 0$ für $x \geq 0$); für die $\delta$-Polynome $\delta \geq \varphi(+0) - \varphi(-0)$. Also

$$\alpha \leq \varphi(-0) \leq \varphi(+0) \leq \beta, \qquad \varphi(+0) - \varphi(-0) \leq \delta. \tag{4}$$

Wir werden zeigen, dass $\delta = \beta - \alpha$, so dass $\delta = 0$ eine *hinreichende* (übrigens auch nothwendige) Bedingung dafür wird, dass 0 Bestimmtheitsstelle ist.

Um die Zahlen $\alpha, \beta, \gamma, \delta$ zu bestimmen, betrachten wir $\beta$- $\gamma$- $\delta$-Polynome vom Grade $\leq 2n$ und nennen die unteren Grenzen ihrer Momente $\beta_n, \gamma_n, \delta_n$; bei Verwandlung von $n$ in $n+1$ erweitert sich der Kreis der zugelassenen Polynome, also sinken die unteren Grenzen ($\beta_n \geq \beta_{n+1}$ usw.) und man erhält

$$\beta = \lim \beta_n, \quad \gamma = \lim \gamma_n, \quad \delta = \lim \delta_n$$

als Grenzwerthe absteigender, $\alpha = \lim \alpha_n$ als Grenzwerth einer aufsteigenden Folge. Beginnen wir mit $\delta_n$. | Bl. 169

$F(x)$ sei ein $\delta$-Polynom, d. h. $\geq 0$ und $F(0) \geq 1$; zur Ermittlung der unteren Grenze können wir $F(0) = 1$ annehmen, da $\frac{F(x)}{F(0)}$ sonst kleineres Moment hätte als $F(x)$. Ist $F$ vom Grade $\leq 2n$, so ist

$$F(x) = f(x)^2 + g(x)^2$$

(weil alle reellen Linearfaktoren in gerader Vielfachheit auftreten), wo $f, g$ Polynome vom Grade $\leq n$ sind; überdies zeigt die Formel

$$F(x) = [g(0)f(x) - f(0)g(x)]^2 + [f(0)f(x) + g(0)g(x)]^2,$$

dass man $f(0) = 0$ und $g(0) = 1$ annehmen und wiederum $f(x)$ identisch 0 annehmen kann, da sonst $g(x)^2$ kleineres Moment hätte als $F(x)$. Also schliesslich: $\delta_n$ ist die untere Grenze von $M\,g(x)^2$ für die Polynome

$$g(x) = 1 + u_1\,x + \cdots + u_n\,x^n,$$

d. h. die untere Grenze der quadratischen Form

$$\sum_0^n \mu_{i+k}\,u_i u_k \qquad \text{für} \quad u_0 = 1.$$

Diese Form hat aber ein Minimum, das man durch Nullsetzen der Ableitungen nach $u_1, \ldots, u_n$ findet, d. h. aus den Gleichungen

$$\begin{array}{ccccccccc} \mu_1 & + & u_1\mu_2 & + & \cdots & + & u_n\mu_{n+1} & = & 0 \\ \mu_2 & + & u_1\mu_3 & + & \cdots & + & u_n\mu_{n+2} & = & 0 \\ & & \cdots & & & & \cdots & & \\ \mu_n & + & u_1\mu_{n+1} & + & \cdots & + & u_n\mu_{2n} & = & 0 \end{array}$$

mit der Determinante (3); das Polynom ist durch diese Bedingungen eindeutig bestimmt. Man kann sie auch

$$M\,(xg) = M\,(x^2g) = \cdots = M\,(x^ng) = 0$$

schreiben; $g(x)$ zu $x, \ldots, x^n$ *orthogonal* ($f$ zu $g$ orthogonal: $M\,(fg) = 0$); durch Bl. 170
sie und $g(0) = 1$ ist $g(x)$ eindeutig bestimmt. Der Ausdruck für dies $g$, das wir nunmehr $= g_n$ setzen wollen, ist

$$g_n(x) = \begin{vmatrix} 1 & x & \cdots & x^n \\ \mu_1 & \mu_2 & \cdots & \mu_{n+1} \\ & \cdots & & \\ \mu_n & \mu_{n+1} & \cdots & \mu_{2n} \end{vmatrix} : E_n \tag{5}$$

und

$$\delta_n = M\,g_n(x)^2 = M\,g_n(x) = D_n : E_n, \tag{6}$$

womit $\delta = \lim \delta_n$ bestimmt ist.

*$g_n(x)$ ist vom Grade $n$ oder $n-1$ und hat lauter reelle einfache Nullstellen.* Sei $g(x)$ das Produkt der verschiedenen reellen Linearfactoren $1 - \frac{x}{\xi}$, die in $g_n(x)$ in ungerader Vielfachheit auftreten, also $gg_n \geq 0$, $\quad M\,(x^2gg_n) > 0$. Daher kann $g$ nicht vom Grade $\leq n-2$ sein, da sonst $g_n$ zu $x^2g$ orthogonal wäre. $g_n$ hat also mindestens $n-1$ verschiedene reelle Nullstellen ungerader Vielfachheit, woraus die Behauptung unmittelbar folgt. Nennen wir den Index $n$ *regulär* oder *singulär*, jenachdem $g_n$ vom Grade $n$ oder $n-1$ ist. Im letzten Fall ist $g_n = g_{n-1}$, da ja $g_n$ die Forderungen erfüllt, durch die $g_{n-1}$ eindeutig bestimmt
war. Zwei benachbarte Indices $n$, $n+1$ können nicht zugleich singulär sein, Bl. 171
da $g_{n-1} = g_n = g_{n+1}$ für $g_{n+1}$ den zu niedrigen Grad $n-1$ ergeben würde; es giebt also *unendlich viele reguläre* Indices $n$. Da ferner für jedes Polynom

$$g = 1 + u_1\,x + \cdots + u_n\,x^n,$$

das $\neq g_n$ ist, $M\,g^2 > \delta_n$ ausfällt, so gilt in $\delta_{n-1} \geq \delta_n$ das Gleichheitszeichen nur, wenn $n$ singulär ist. In der Kette $\delta_1 \geq \delta_2 \geq \cdots$ kommt das Zeichen $>$ also unendlich oft vor und es ist $\delta_n > \delta$.

Beschränken wir uns auf *reguläre* $n$. Die $n$ Nullstellen von $g_n(x)$, zusammen mit $x = 0$ (welches keine Nullstelle ist), nennen wir $x_0 < x_1 < \cdots < x_n$, wobei etwa $x_i = 0$ (eigentlich wäre, wie auch im Folgenden, die Abhängigkeit von $n$ anzudeuten). Setzen wir

$$\left.\begin{array}{rcccl} f(x) & = & x g_n(x) & = & C(x-x_0)\cdots(x-x_n) \\ & & f_j(x) & = & \frac{f(x)}{(x-x_j)f'(x_j)} \qquad\qquad (j=0,1,\ldots,n) \\ & & p_j & = & M\, f_j(x) \end{array}\right\} \tag{7}$$

$f_j(x_k)$ verschwindet für $k \neq j$, während $f_j(x_j) = 1$. Speciell ist $f_i(x) = g_n(x), \quad p_i = \delta_n$.

Für jedes Polynom $F(x)$ vom Grade $\leq 2n$ ist

$$F(x) = f(x)\, q(x) + r(x)$$

Bl. 172 ($q$ Quotient, $r$ Rest), $q(x)$ vom Grade $\leq n-1$, $r(x)$ vom Grade $\leq n$; hierbei ist $M\,(fq) = M\,(g_n \cdot xq) = 0, \quad M\,F = M\,r$, ferner nach der Lagrangeschen Interpolationsformel

$$r(x) = \sum_0^n f_j(x)\, r(x_j)$$

$$M\,r = \sum_0^n p_j\, r(x_j) = \sum_0^n p_j\, F(x_j), \quad \text{also}$$

$$M\,F(x) = \sum_0^n p_j\, F(x_j) \tag{8}$$

für jedes Polynom vom Grade $\leq 2n$. Insbesondere

$$M\, f_j(x)^2 = p_j > 0 \tag{9}$$

(Wir haben hier eine elementare Vertheilung $(x_j, p_j)$, deren Momente $\mu_0, \ldots, \mu_{2n}$ nach (8) die *gegebenen* Werthe haben).

Ist nun $F(x)$ ein $\beta$-Polynom, so liefert (8) (9) $\quad M\,F \geq p_0 + \cdots + p_i$, also $\beta_n \geq p_0 + \cdots + p_i$. In Wahrheit (dies ist einer der wichtigsten Punkte dieser Tschebyscheffschen Theorie) ist $\beta_n = p_0 + \cdots + p_i$. Man kann nämlich ein Polynom $F(x)$ vom Grade $\leq 2n$ durch die $2n+1$ linearen Bedingungen

$$F(x_0) = \cdots = F(x_i) = 1, \quad F(x_{i+1}) = \cdots = F(x_n) = 0$$

$$F'(x_0) = \cdots = F'(x_{i-1}) = F'(x_{i+1}) = \cdots = F'(x_n) = 0$$

eindeutig bestimmen; denn schreibt man wieder $F = fq + r$, so ist

$$r(x_0) = \cdots = r(x_i) = 1, \quad r(x_{i+1}) = \cdots = r(x_n) = 0$$

und dadurch $r(x)$ bestimmt (Lagrange); sodann folgt aus

$$F' = fq' + f'q + r'$$

$$q(x_j) = -\frac{r'(x_j)}{f'(x_j)} \qquad (j \neq i),$$

wodurch auch $q(x)$ bestimmt ist. Dies Polynom ist nun ein $\beta$-Polynom, sein Moment $p_0 + \cdots + p_i$ (nach (8)), wodurch $p_0 + \cdots + p_i \geq \beta_n \geq p_0 + \cdots + p_i$, also Gleichheit bedingt ist. Denn $F'(x)$ hat ausser den $n$ vorgeschriebenen Nullstellen nach dem Rolleschen Satz noch $n-1$ weitere in den Intervallen

$$(x_0, x_1), \cdots (x_{i-1}, x_i), (x_{i+1}, x_{i+2}) \cdots (x_{n-1}, x_n);$$

dies sind also sämtliche Nullstellen und alle einfach, so dass $F'(x)$ dort jedesmal das Zeichen wechselt. Da $F(x_i) > F(x_{i+1})$, ist in $(x_i, x_{i+1})$ $F'(x) < 0$. Hierdurch ergiebt sich für $F'(x)$ das Vorzeichenschema

$F(x)$ nimmt in $(-\infty, x_0)$ und $(x_i, x_{i+1})$ ab, in $(x_n, \infty)$ zu; in den übrigen Intervallen nimmt es erst zu, erreicht ein Maximum und nimmt dann wieder ab.

Ebenso ergiebt sich $\gamma_n = p_i + \cdots + p_n$, wegen $\mu_0 = p_0 + \cdots + p_n$ also Bl. 174 $\alpha_n = p_0 + \cdots + p_{i-1}$. D. h.

$$\alpha_n = p_0 + \cdots + p_{i-1}, \quad \beta_n = p_0 + \cdots + p_i, \quad \delta_n = \beta_n - \alpha_n = p_i \tag{10}$$

und dadurch wirklich $\delta = \beta - \alpha$, und $\delta = 0$ ist, wie wir ankündigten, hinreichend für die Bestimmtheitsstelle 0. (Auch ist dann, nach (4), jede Lösung an der Stelle 0 stetig.)

Der Übergang von 0 auf eine Stelle $\xi$ erfolgt durch Coordinaten*verschiebung*: aus

$$\mu_k(\xi) = M\,(x - \xi)^k = \mu_k - \binom{k}{1}\mu_{k-1}\xi + \cdots + (-1)^k \mu_0 \xi^k$$

gewinnt man $\alpha(\xi), \beta(\xi), \delta(\xi)$ ebenso wie $\alpha, \beta, \delta$ aus $\mu_k$. Wir können auch sagen: es ist $\delta_n(\xi) = M\,g_n(x|\xi)$, wo $g_n(x|\xi)$ das zu $x - \xi, (x-\xi)^2, \ldots, (x-\xi)^n$ orthogonale Polynom vom Grade $\leq n$ ist, das an der Stelle $x = \xi$ den Werth $g_n(\xi|\xi) = 1$ annimmt. Oder: ist

$$h_n(x|\xi) = \frac{g_n(x|\xi)}{\delta_n(\xi)}$$

das zu $x-\xi, \cdots, (x-\xi)^n$ orthogonale Polynom vom Grade $\leq n$, während $M\,h_n = 1$, so ist $\delta_n(\xi) = 1 : h_n(\xi|\xi)$. $h_n$ ist | durch die Eigenschaft charakterisirt, dass für jedes Polynom vom Grade $\leq n$

$$f(x) = f(\xi) + (x-\xi)f'(\xi) + \cdots + (x-\xi)^n \frac{f^{(n)}(\xi)}{n!}$$

$$M\,f(x)h_n(x|\xi) = f(\xi). \tag{11}$$

Wir benutzen zur einfachsten Darstellung die *Orthogonalpolynome* der Momentfolge, d. h. diejenigen Polynome

$$f_n(x) = x^n + \cdots + u_0,$$

die zu $1, x, \ldots, x^{n-1}$ orthogonal sind ($f_0(x) = 1$); die betreffenden Gleichungen haben die Determinante $D_{n-1}$ und $f_n(x)$ ist also eindeutig bestimmt, nämlich

$$f_n(x) = \begin{vmatrix} \mu_0 & \mu_1 & \cdots & \mu_n \\ \mu_1 & \mu_2 & \cdots & \mu_{n+1} \\ & \cdots & & \cdots \\ \mu_{n-1} & \mu_n & \cdots & \mu_{2n-1} \\ 1 & x & \cdots & x^n \end{vmatrix} : D_{n-1} \qquad (f_0(x) = 1) \tag{12}$$

Hingegen wird

$$M\,f_n(x)^2 = M\,f_n(x)\,x^n = D_n : D_{n-1} = \frac{1}{b_n} \qquad \left(b_n = \frac{D_{n-1}}{D_n};\ b_0 = \frac{1}{\mu_0}\right). \tag{13}$$

Für die Momente des Exponentialgesetzes sind das gerade die Hermiteschen Polynome § 7; (4) und zwar ist dann $b_n = \dfrac{2^n}{n!}$. Setzt man dann

$$h_n(x|\xi) = \sum_0^n B_\nu f_\nu(x),$$

so ergiebt (11) (13)

$$M\,h_n(x|\xi)f_\nu(x) = f_\nu(\xi) = \frac{B_\nu}{b_\nu},$$

Bl. 176 | also

$$h_n(x|\xi) = \sum_0^n b_\nu f_\nu(x) f_\nu(\xi) \tag{14}$$

$$1 : \delta_n(\xi) = h_n(\xi|\xi) = \sum_0^n b_\nu f_\nu(\xi)^2, \tag{15}$$

so dass wir jetzt als *hinreichende Bedingung* für die *Bestimmtheitsstelle* $\xi$ die *Divergenz der Reihe*

$$\sum_0^\infty b_n \, f_n(\xi)^2$$

haben.

Hiernach ergiebt sich:

I. *$\varphi(x)$ sei eine monotone Function mit unendlich vielen Wachsthumsstellen, $\xi$ eine Bestimmtheitsstelle mit $\delta(\xi) = 0$, $\varphi_n(x)$ eine Folge monotoner Functionen, deren Momente*

$$M_n \, x^k = \int_{-\infty}^{\infty} x^k \, d\varphi_n$$

*nach den Momenten*

$$\mu_k = M \, x^k = \int_{-\infty}^{\infty} x^k \, d\varphi$$

*convergiren. Dann convergirt $\varphi_n(\xi)$ nach $\varphi(\xi)$.*

[Es ist $\varphi(-\infty) = \varphi_n(-\infty) = 0$ vorausgesetzt. Die Momente $M_n \, x^k$ brauchen bei festem $k$ nur für hinreichend grosse $n$ zu existiren. Für $\varphi_n(x)$ braucht $\xi$ keine Bestimmtheitsstelle zu sein und $\varphi_n(\xi)$ kann mit $\varphi_n(\xi \pm 0)$ oder jedem zwischenliegenden Werth identificirt werden.]

Der Beweis ist ganz einfach. Nehmen wir die Stelle $\xi = 0$. Für ein festes Polynom $F(x)$ ist $M_n \, F(x) \to MF(x)$. Sei $F(x)$ das $\beta$-Polynom vom Grade $\leq 2k$, für das $M \, F(x) = \beta_k$, so ist also *schliesslich* (für $n \geq n_0$) $M_n \, F(x) < \beta_k + \varepsilon$, also nach (4) $\varphi_n(0) < \beta_k + \varepsilon$, ebenso $\varphi_n(0) > \alpha_k - \varepsilon$, und | da zugleich Bl. 177
$\alpha_k \leq \varphi(0) \leq \beta_k$ ist:

$$|\varphi_n(0) - \varphi(0)| < \delta_k + \varepsilon.$$

Ist 0 so beschaffen, dass $\delta(0) = \delta = 0$, so kann man zuerst $\delta_k < \varepsilon$ machen. Also $\varphi_n(0) \to \varphi(0)$, und durch Translation für jede Stelle $\xi$.

Ist nun in I überall $\delta(x) = 0$, so convergirt überall $\varphi_n(x) \to \varphi(x)$. Das geschieht in diesem Falle, wo ja $\varphi(x)$ stetig ist, *gleichmässig* für alle $x$. Zunächst nämlich: wenn die in $[a, b]$ monotonen Functionen $\varphi_n(x)$ nach einer in $[a, b]$ *stetigen* (natürlich wieder monotonen) Function $\varphi(x)$ convergiren, so geschieht das gleichmässig. Denn man kann eine Intervalltheilung $a = x_0 < \cdots < x_n = b$ vornehmen derart, dass für $i = 1, \ldots, n$ $\varphi(x_i) - \varphi(x_{i-1}) < \varepsilon$, und sodann $n_0$ so gross wählen, dass für $i = 0, \ldots, n$ $|\varphi_n(x_i) - \varphi(x_i)| < \varepsilon$ für $n \geq n_0$. Man hat dann, wenn $x$ zu $[x_{i-1}, x_i]$ gehört

$$\varphi_n(x) - \varphi(x) \leq \varphi_n(x_i) - \varphi(x_{i-1}) < \varphi_n(x_i) - \varphi(x_i) + \varepsilon < 2\varepsilon,$$

$$\varphi_n(x) - \varphi(x) \geq \varphi_n(x_{i-1}) - \varphi(x_i) > \varphi_n(x_{i-1}) - \varphi(x_{i-1}) - \varepsilon > -2\varepsilon,$$

also im ganzen Intervall

$$|\varphi_n(x) - \varphi(x)| < 2\varepsilon \quad \text{für} \quad n \geq n_0.$$

Transformirt man $(a, b)$ in ein unendliches Intervall $(-\infty, \infty)$, so ergiebt sich: wenn die monotonen Functionen $\varphi_n(x)$ nach einer *stetigen* Function $\varphi(x)$ convergiren und zugleich

$$\varphi_n(-\infty) \to \varphi(-\infty), \quad \varphi_n(+\infty) \to \varphi(+\infty),$$

Bl. 178 so | ist die Convergenz gleichmässig. Das ist hier, wo $\varphi_n(-\infty) = \varphi(-\infty) = 0$ und $\varphi_n(+\infty) = M_n\,1 \to M\,1 = \varphi(+\infty)$, der Fall. Also:

II. *$\varphi(x)$ sei eine monotone Function, für die überall $\delta(x) = 0$, und $\varphi_n(x)$ eine Folge monotoner Functionen, deren Momente nach denen von $\varphi(x)$ convergiren. Dann convergirt $\varphi_n(x)$ nach $\varphi(x)$ und zwar gleichmässig für alle $x$.* $[\varphi_n(-\infty) = \varphi(-\infty) = 0$ vorausgesetzt$]$.

Wir wollen nun zeigen, dass beim Exponentialgesetz

$$\varphi(x) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{x} e^{-x^2}\,dx$$

die Voraussetzung $\delta(x) = 0$ erfüllt ist, d. h. die Reihe

$$\sum_0^\infty b_n\, f_n(x)^2 = \sum_0^\infty \frac{2^n}{n!}\, f_n(x)^2$$

überall divergirt, wo die $f_n(x)$ die Hermiteschen Polynome sind. Es ist

$$1 = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-(u-y)^2}\,du, \qquad e^{y^2} = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-u^2+2uy}\,du,$$

wir wollen als bekannt annehmen, dass dies auch für complexes $y$ gilt, z. B. wenn man $y$ mit $iy$ vertauscht

$$(\alpha) \qquad e^{-y^2} = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-u^2+2iuy}\,du$$

und $n$ mal differenzirt

$$(\beta) \qquad D^n(e^{-y^2}) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-u^2+2iuy}(2iu)^n\,du$$

Bl. 179 | Aus § 7, (5) erhält man, wenn man $t$ durch $2itu$ ersetzt

$$e^{2ixtu+t^2u^2} = \sum_0^\infty f_n(x)\,\frac{t^n}{n!}\,(2iu)^n$$

Dies multiplicire man mit $e^{-u^2+2iuy}\dfrac{du}{\sqrt{\pi}}$ und integrire nach $u$, so kommt nach $(\beta)$

$$\sum_0^\infty D^n(e^{-y^2}) f_n(x) \frac{t^n}{n!} = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-u^2(1-t^2)+2iu(xt+y)}\,du,$$

wobei $-1 < t < 1$ vorauszusetzen ist; das Integral rechts ist mit $u = \dfrac{v}{\sqrt{1-t^2}}$

$$= \frac{1}{\sqrt{\pi}\sqrt{1-t^2}} \int_{-\infty}^{\infty} e^{-v^2+2iv\frac{xt+y}{\sqrt{1-t^2}}}\, dv$$

oder nach $(\alpha)$

$$= e^{-\left(\frac{xt+y}{\sqrt{1-t^2}}\right)^2} : \sqrt{1-t^2}$$

Wir erhalten endlich, wenn wir

$$D^n(e^{-y^2}) = (-1)^n\, 2^n\, e^{-y^2} f_n(y)$$

setzen und $t$ mit $-t$ vertauschen

$$(\gamma) \qquad \frac{1}{\sqrt{1-t^2}} e^{-\frac{(y-xt)^2}{1-t^2}} = e^{-y^2} \sum_0^\infty \frac{2^n}{n!} f_n(x) f_n(y)\, t^n \qquad (|t| < 1)$$

insbesondere für $y = x$

$$(\delta) \qquad \frac{e^{\frac{2x^2 t}{1+t}}}{\sqrt{1-t^2}} = \sum_0^\infty \frac{2^n}{n!} f_n(x)^2\, t^n$$

Da für $t \to 1$ die linke Seite nach $+\infty$ divergirt, so muss auch

$$\sum_0^\infty \frac{2^n}{n!} f_n(x)^2$$

divergiren (Abelscher Satz), q.e.d. | Bl. 180

Hiernach können wir in II speciell das Exponentialgesetz

$$\varphi(x) = \Phi(x) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{x} e^{-x^2}\, dx$$

nehmen und, wenn wir statt der Momente die logarithmischen Momente nehmen:

III. (Grenzwerthsatz von Tschebyscheff). $\Phi_n(x)$ *sei eine Folge monotoner Functionen mit* $\Phi_n(-\infty) = 0, \quad \Phi_n(+\infty) \to 1$, *deren logarithmische Momente* $\lambda_k$ *sämtlich nach 0 convergiren bis auf* $\lambda_2 \to \frac{1}{2}$. *Dann convergirt* $\Phi_n(x)$ *gleichmässig nach*

$$\Phi(x) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{x} e^{-x^2}\, dx.$$

Dieser Satz, noch mehr II, ist in gewisser Hinsicht allgemeiner als der Liapunoffsche, der sich nur auf die Summe unabhängiger Variablen bezieht; für diesen Specialfall verlangt freilich der Tschebyscheffsche Satz erheblich mehr als der Liapunoffsche, der ja nicht einmal die Existenz höherer Momente (als des dritten) für die $x_n$ voraussetzt. | Bl. 181

## Das Exponentialgesetz

[7] Die Eulerschen Integrale (Integralrechnung).

Das Integral 1. Gattung

$$B(\alpha,\beta) = \int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\,du \quad (\alpha,\beta > 0).$$

Das Integral 2. Gattung

$$\Gamma(\alpha) = \int_0^\infty e^{-t}\,t^{\alpha-1}\,dt \quad (\alpha > 0);$$

die $\Gamma$-Funktion Legendres; Gauss setzt $\Pi(\alpha) = \Gamma(\alpha+1)$ für $\alpha > -1$. Es ist $\Gamma(1) = 1,\ \Gamma(n) = (n-1)!$ für $n = 2, 3, \ldots$ .

$$\Gamma(\alpha) = \alpha\,\Gamma(\alpha-1) \quad (\alpha > 1). \qquad \Gamma(\alpha+1) = (\alpha+1)\,\Gamma(\alpha) \quad (\alpha > 0).$$

Hier gilt

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

Insbesondere $B(\frac{1}{2},\frac{1}{2}) = \Gamma(\frac{1}{2})^2$, und da $(u = \sin^2\varphi)$

$$B\left(\frac{1}{2},\frac{1}{2}\right) = \int_0^1 \frac{du}{\sqrt{u(1-u)}} = \int_0^{\frac{\pi}{2}} 2\,d\varphi = \pi,$$

so ist $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, andererseits

$$\Gamma\left(\frac{1}{2}\right) = \int_0^\infty \frac{e^{-t}}{\sqrt{t}}\,dt = 2\int_0^\infty e^{-x^2}\,dx = \int_{-\infty}^\infty e^{-x^2}\,dx.$$

Bl. 182 | Die monotone Funktion

$$\Phi(x) = \frac{1}{\sqrt{\pi}}\int_{-\infty}^x e^{-x^2}\,dx \quad (\Phi(-\infty) = 0,\ \Phi(+\infty) = 1)$$

liefert die in der Wahrscheinlichkeitsrechnung wichtigste Verteilung einer Variablen $x$. Ihre Momente sind

$$\mu_k = \frac{1}{\sqrt{\pi}}\int_{-\infty}^\infty x^k\,e^{-x^2}\,dx;$$

also $\mu_1 = \mu_3 = \cdots = 0$, während für gerade Indizes $(x = \sqrt{t})$

$$\mu_{2k} = \frac{1}{\sqrt{\pi}}\int_{-\infty}^\infty e^{-x^2}\,x^{2k}\,dx = \frac{2}{\sqrt{\pi}}\int_0^\infty e^{-x^2}\,x^{2k}\,dx$$

$$= \frac{1}{\sqrt{\pi}} \int_0^\infty e^{-t}\, t^{k-\frac{1}{2}}\, dt = \frac{\Gamma\left(k+\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)} = \frac{1}{2} \cdot \frac{3}{2} \cdot \;\cdots\; \cdot \frac{2k-1}{2}$$

ist: $\mu_0 = 1$, $\mu_2 = \frac{1}{2}$, $\mu_4 = \dfrac{1 \cdot 3}{2 \cdot 2}, \ldots$

Das Moment von $e^{xu}$ ($u$ eine Variable) ist

$$M\, e^{xu} = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-x^2+xu}\, dx = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-(x-\frac{u}{2})^2+\frac{u^2}{4}}\, dx = e^{\frac{u^2}{4}},$$

$\log M\, e^{xu} = \dfrac{u^2}{4}$; *alle logarithmischen Momente*

$$(\log M\, e^{xu} = \lambda_1 u + \lambda_2 \frac{u^2}{2!} + \cdots)$$

*des Exponentialgesetzes verschwinden* bis auf $\lambda_2 = \frac{1}{2}$. | Mittels einer linearen Transformation $x = h(\xi - \alpha)$ ($h > 0,\ \alpha$ beliebig) erhält man die Verteilung Bl. 183

$$\Phi(h(\xi - \alpha)) = \frac{h}{\sqrt{\pi}} \int_{-\infty}^{\xi} e^{-h^2(\xi-\alpha)^2}\, d\xi$$

der Variablen $\xi$ mit den logarithmischen Momenten $\lambda_1 = \alpha$ und $\lambda_2 = \dfrac{1}{2h^2}$; die Streuung ist $\sqrt{\lambda_2} = \dfrac{1}{\sqrt{2}h}$, $h$ (Präzisionsmass) ist umso grösser, je kleiner die Streuung ist.

*Die Hermiteschen Polynome.* Die $n$. Ableitung $D^n\, e^{-x^2}$ ist $= e^{-x^2}$ mal Polynom $n$. Grades, z. B.

$$D\, e^{-x^2} = e^{-x^2}(-2x), \quad D^2\, e^{-x^2} = e^{-x^2}(4x^2 - 2),$$

$$D^3\, e^{-x^2} = e^{-x^2}(-8x^3 + 12x), \quad \ldots$$

Wir setzen

$$D^n\, e^{-x^2} = (-1)^n\, 2^n\, e^{-x^2} f_n(x),$$

wobei die Rekursionsformel

$$f_{n+1} = x f_n - \frac{1}{2} f_n'$$

besteht; $f_n(x) = x^n + \cdots$ . (Bisweilen werden andere Zahlenfaktoren gewählt). Diese $f_n(x)$ heissen die *Hermiteschen Polynome.* | Bl. 184

Nach Taylor ist

$$e^{-(x-\frac{t}{2})^2} = \sum_0^\infty D^n\, e^{-x^2} \cdot \frac{(-1)^n\, t^n}{2^n\, n!} = e^{-x^2} \sum_0^\infty f_n(x)\, \frac{t^n}{n!}$$

$$e^{xt-\frac{t^2}{4}} = \sum_0^\infty f_n(x)\,\frac{t^n}{n!}$$

Durch partielle Integration erhält man für ein Polynom $f(x)$

$$(*) \qquad \int_{-\infty}^{\infty} D^n(e^{-x^2})\,f(x)\,dx = -\int_{-\infty}^{\infty} D^{n-1}(e^{-x^2})\,Df(x)\,dx$$

$$= \int_{-\infty}^{\infty} D^{n-2}(e^{-x^2})\,D^2 f(x)\,dx = \cdots = (-1)^n \int_{-\infty}^{\infty} e^{-x^2}\,D^n f(x)\,dx,$$

also 0, wenn $f(x)$ vom Grade $< n$ ist

$$\int_{-\infty}^{\infty} e^{-x^2}\,f_n(x)\,f(x)\,dx = 0 \quad (f(x) \text{ vom Grade } < n)$$

$$\int_{-\infty}^{\infty} e^{-x^2}\,f_n(x)\,f_m(x)\,dx = 0 \qquad \text{für } m < n;$$

die Hermiteschen Polynome sind die Orthogonalpolynome unserer Verteilung $\Phi(x)$. Ist $f(x) = x^n + \cdots$, so ist nach $(*)$

$$(-1)^n\,2^n \int_{-\infty}^{\infty} e^{-x^2}\,f_n(x)\,f(x)\,dx = (-1)^n\,n! \int_{-\infty}^{\infty} e^{-x^2}\,dx = (-1)^n\,n!\,\sqrt{\pi},$$

Bl. 185 | also für $f(x) = f_n(x)$

$$M\,f_n(x)^2 = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-x^2}\,f_n(x)^2\,dx = \frac{n!}{2^n}$$

Das sind die vorhin mit $\dfrac{1}{b_n}$ bezeichneten Zahlen. Um nachzuweisen, dass für die Funktion $\Phi(x)$ jede Stelle $x$ Bestimmtheitsstelle mit $\delta(x) = 0$ ist, haben wir also die Divergenz der Reihe

$$\sum_0^\infty b_n\,f_n(x)^2 = \sum \frac{2^n}{n!}\,f_n(x)^2$$

zu zeigen.

Es ist

$$1 = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-(u-y)^2}\,du, \quad e^{y^2} = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-u^2+2uy}\,du.$$

Das gilt auch für komplexe $y$, also ($y$ durch $iy$ ersetzt)

$$(\alpha) \qquad e^{-y^2} = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-u^2+2iuy}\,du$$

Durch $n$-malige Differentiation

$$(\beta) \qquad D^n\, e^{-y^2} = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-u^2+2iuy} (2iu)^n\, du$$

Die vorhin angegebene Formel

$$e^{xt-\frac{t^2}{4}} = \sum f_n(x) \frac{t^n}{n!}$$

giebt, wenn man $t$ durch $2itu$ ersetzt,

$$e^{2ixtu+t^2u^2} = \sum f_n(x)\, \frac{t^n}{n!}\, (2iu)^n.$$

Wenn man mit Bl. 186

$$\frac{1}{\sqrt{\pi}}\, e^{-u^2+2iuy}$$

multipliziert, kommt

$$\frac{1}{\sqrt{\pi}} e^{-u^2(1-t^2)+2iu(xt+y)} = \sum f_n(x)\, \frac{t^n}{n!}\, \frac{1}{\sqrt{\pi}}\, e^{-u^2+2iuy} (2iu)^n.$$

Für $-1 < t < 1$ lässt sich die linke Seite nach $u$ über $-\infty$ bis $\infty$ integrieren, das giebt ($u = \dfrac{v}{\sqrt{1-t^2}}$)

$$\frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-u^2(1-t^2)+2iu(xt+y)}\, du = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-v^2+2iv\frac{xt+y}{\sqrt{1-t^2}}}\, \frac{dv}{\sqrt{1-t^2}}$$

$$\text{oder nach } (\alpha) \qquad = \frac{1}{\sqrt{1-t^2}}\, e^{-\frac{(xt+y)^2}{1-t^2}};$$

die rechte Seite gliedweise integriert giebt nach ($\beta$)

$$\sum f_n(x)\, \frac{t^n}{n!}\, D^n\, e^{-y^2} = e^{-y^2} \sum f_n(x) f_n(y)\, \frac{(-1)^n\, 2^n\, t^n}{n!}$$

Gleichsetzung beider Ausdrücke führt, wenn man noch $t$ mit $-t$ vertauscht, zu

$$\sum f_n(x) f_n(y)\, \frac{2^n}{n!}\, t^n = \frac{1}{\sqrt{1-t^2}}\, e^{y^2-\frac{(y-xt)^2}{1-t^2}} = \frac{1}{\sqrt{1-t^2}}\, e^{\frac{2xyt-(x^2+y^2)t^2}{1-t^2}}$$

Insbesondere Bl. 187

$$\sum \frac{2^n}{n!}\, f_n(x)^2\, t^n = \frac{1}{\sqrt{1-t^2}}\, e^{\frac{2x^2t}{1+t}} \quad (-1 < t < 1).$$

Da nun für $t \to 1$ die rechte Seite stets nach $+\infty$ strebt, kann (Abelscher Stetigkeitssatz) die linke Seite für $t = 1$ nicht konvergieren, also $\sum \dfrac{2^n}{n!}\, f_n(x)^2$ divergiert für jedes $x$.

Hiernach können wir in II speziell für $\varphi(x)$ die Funktion

$$\Phi(x) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{x} e^{-x^2}\, dx$$

setzen, und wenn wir statt der Momente die logarithmischen Momente nehmen:

III. $\Phi_n(x)$ sei eine Folge monotoner Funktionen mit $\Phi_n(-\infty) = 0,\ \Phi_n(+\infty) \to 1$, deren logarithmische Momente $\lambda_k$ alle nach 0 konvergieren bis auf $\lambda_2 \to \frac{1}{2}$. Dann konvergiert $\Phi_n(x)$ gleichmässig nach

$$\Phi(x) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{x} e^{-x^2}\, dx.$$

Insbesondere (resp. direkt aus $\delta(x) = 0$ zu schliessen): eine monotone Funktion, deren (logarithmische) Momente mit denen von $\Phi(x)$ oder $\Phi(h(\xi - \alpha))$
Bl. 188 übereinstimmen, ist mit $\Phi(x)$ oder $\Phi(h(\xi - \alpha))$ identisch. |

Man nennt zwei Variable $x, y$ mit den Verteilungsfunktionen

$$\varphi(x),\ \psi(y) \quad (W(x < \xi) = \varphi(\xi),\ W(y < \eta) = \psi(\eta))$$

*unabhängig*, wenn für das Variablenpaar die Wahrscheinlichkeit, dass gleichzeitig $x < \xi,\ y < \eta$ sei, gleich $\varphi(\xi)\, \psi(\eta)$ ist. Wir haben die auf zwei oder mehrere Variable bezügliche Mass- und Integraltheorie nicht ausführlich entwickeln können. Die logarithmischen Momente von $x + y$ sind die Summen der logarithmischen Momente von $x$ und $y$, falls $x, y$ unabhängig sind.

Wenn $\xi$ die Verteilungsfunktion $\Phi(h(\xi - \alpha))$, $\eta$ die Verteilungsfunktion $\Phi(k(\eta - \beta))$ hat, so ist $\lambda_1(\xi) = \alpha,\ \lambda_2(\xi) = \frac{1}{2h^2}$, die übrigen 0, $\lambda_1(\eta) = \beta,\ \lambda_2(\eta) = \frac{1}{2k^2}$, die übrigen 0, also bei unabhängigen $\xi, \eta$

$$\lambda_1(\xi + \eta) = \alpha + \beta, \quad \lambda_2(\xi + \eta) = \frac{1}{2}\left(\frac{1}{h^2} + \frac{1}{k^2}\right),$$

die übrigen 0; d. h. $\zeta = \xi + \eta$ hat wieder eine Gausssche Verteilung $\Phi(l(\zeta - \gamma))$. *Das Gausssche Gesetz erhält sich bei Addition unabhängiger Variablen.*

Eben auf demselben Grunde beruht es, dass das Exponentialgesetz unter gewissen Bedingungen bei Addition unabhängiger Variablen als *Grenzfunktion*
Bl. 189 auftritt. |

Nennen wir eine Variable $x$, deren erste Momente $\mu_0 = 1,\ \mu_1 = 0,\ \mu_2 = \frac{1}{2}$ sind, *normiert.* Aus einer Variablen $\xi$ mit den Momenten $\mu_0 = 1,\ \mu_1,\ \mu_2$ kann man durch lineare Transformation eine normierte Variable

$$x = \frac{\xi - \mu_1}{\sqrt{2(\mu_2 - \mu_1^2)}} \quad \text{ableiten:}$$

$$M\,x = 0,\ \ M\,x^2 = \frac{1}{2(\mu_2 - \mu_1^2)}\, M(\xi - \mu_1)^2 = \frac{1}{2(\mu_2 - \mu_1^2)}(\mu_2 - 2\mu_1^2 + \mu_1^2) = \frac{1}{2}.$$

Wenn $\xi$ einem Gaussschen Gesetz $\Phi(h(\xi-\alpha))$ folgt, so ist $x = h(\xi-\alpha)$ die normierte Variable und folgt dem Gaussgesetz $\Phi(x)$. Sind $\lambda_1, \lambda_2$ die ersten beiden logarithmischen Momente von $\xi$, so ist $x = \dfrac{\xi - \lambda_1}{\sqrt{2\lambda_2}}$ die zugehörige normierte Variable.

Seien nun wie in § 2 $\xi_1, \xi_2, \ldots$ unabhängige Variable, $M\,\xi_n = \alpha_n$ ihre ersten Momente, ferner

$$\eta_n = \xi_1 + \cdots + \xi_n, \quad M\,\eta_n = \beta_n = \alpha_1 + \cdots + \alpha_n;$$

setzen wir noch

$$x_n = \xi_n - \alpha_n, \quad y_n = \eta_n - \beta_n,$$

so dass $M\,x_n = M\,y_n = 0$. Ferner führen wir die logarithmischen Momente ein

$$\lambda_k(y_n) = \lambda_k(x_1) + \cdots + \lambda_k(x_n) = n\,\lambda_{kn} \quad (k \geq 2)$$

($\lambda_{kn}$ der Durchschnitt aus den logarithmischen Momenten der ersten $n$ Variablen, wobei für $k \geq 2$ ja $\lambda_k(x_n) = \lambda_k(\xi_n)$, $\lambda_k(y_n) = \lambda_k(\eta_n)$ | ist). Die zu $\eta_n$ Bl. 190
oder $y_n$ gehörige normierte Variable ist

$$z_n = \frac{y_n}{\sqrt{2\,\lambda_2(y_n)}} = \frac{y_n}{\sqrt{2n\,\lambda_{2n}}},$$

und für diese ist

$$\lambda_k(z_n) = \frac{1}{\left(\sqrt{2n\,\lambda_{2n}}\right)^k}\,\lambda_k(y_n) = \frac{1}{2^{\frac{k}{2}}\,n^{\frac{k-2}{2}}}\,\frac{\lambda_{kn}}{(\lambda_{2n})^{\frac{k}{2}}}$$

*Wenn* $\lambda_k(z_n)$ *für* $k \geq 3$ *mit* $n \to \infty$ *nach* 0 *konvergiert, so konvergiert die Verteilungsfunktion* $\Phi_n(z)$ *von* $z_n$ (*d. h. die Wahrscheinlichkeit, dass* $z_n < z$) *für* $n \to \infty$ *gleichmässig nach*

$$\Phi(z) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{z} e^{-x^2}\,dx.$$

Das trifft z. B. zu, falls alle Variablen dieselbe Verteilung haben, d. h. $\lambda_{kn} = \lambda_k$ von $n$ unabhängig ist; dann ist

$$\lambda_k(z_n) = \frac{1}{2^{\frac{k}{2}}\,n^{\frac{k-2}{2}}}\,\frac{\lambda_k}{\lambda_2^{\frac{k}{2}}} \to 0 \text{ für } k \geq 3.$$

Oder auch, wenn die Variablen $x_n$ gleichmässig beschränkt sind und $\sum \lambda_2(x_n)$ divergiert. Die erste Voraussetzung soll besagen, dass alle $x_n$ nur in einem festen Intervall $[-L, L]$ variieren, d. h. ihre Verteilungsfunktionen $\varphi_n(x)$ für $x \leq -L$ gleich 0, für $x \geq L$ gleich 1 sind. Unter dieser Annahme gilt | nämlich, wie wir Bl. 191
sogleich zeigen werden, eine Abschätzung der Form ($c_k$ nur von $k$ abhängig)

$$(*) \qquad |\lambda_k(x_n)| \leq c_k\,L^{k-2}\lambda_2(x_n) \quad (k > 2),$$

aus der folgt

$$|\lambda_k(y_n)| \leq c_k\, L^{k-2}\lambda_2(y_n),$$

demnach

$$\frac{|\lambda_k(y_n)|^{\frac{1}{k}}}{\lambda_2(y_n)^{\frac{1}{2}}} \leq \left(c_k L^{k-2}\right)^{\frac{1}{k}}\, \frac{1}{\lambda_2(y_n)^{\frac{1}{2}-\frac{1}{k}}} \to 0,$$

falls $\lambda_2(y_n) \to \infty$.

Die Abschätzung $(*)$ oder $|\lambda_k| \leq c_k L^{k-2}\lambda_2$ für eine Variable $x$, für die $\lambda_1 = 0$ und die nur in $[-L, L]$ variiert, kommt so zustande. Es ist für $k \geq 2$

$$\mu_k = \int_{-L}^{L} x^k\, d\varphi(x),$$

$$|\mu_k| \leq \int_{-L}^{L} |x|^k\, d\varphi(x) \leq L^{k-2} \int_{-L}^{L} x^2\, d\varphi(x) = L^{k-2}\mu_2 = L^{k-2}\lambda_2;$$

ebenso $\lambda_2 \leq L^2$. Nun sind die $\lambda$ für $\mu_1 = 0$ durch Formeln (§ 2) bestimmt wie

$$\lambda_3 = \mu_3, \quad \lambda_4 = \mu_4 - 3\mu_2^2,\ \ldots$$

Bl. 192 allgemein |

$$\lambda_k = \sum c\, \mu_{k_1}\mu_{k_2}\cdots\mu_{k_r}$$

$(k_1, \ldots, k_r \geq 2$, nicht notwendig verschieden, $k_1 + \cdots + k_r = k$, $c$ ganze Zahlen). Dann ist

$$|\mu_{k_1}\cdots\mu_{k_r}| \leq L^{k_1-2+\cdots+k_r-2}\,\lambda_2^r = L^{k-2r}\,\lambda_2^r$$

und für $r > 1$ (sonst haben wir bereits $L^{k-2}\lambda_2) \leq L^{k-2r}\,L^{2(r-1)}\lambda_2 = L^{k-2}\lambda_2$;

$$|\lambda_k| \leq L^{k-2}\,\lambda_2 \cdot \sum |c| = c_k\, L^{k-2}\,\lambda_2.$$

Beispiel. Die Variable $\xi_n$ nehme mit den Wahrscheinlichkeiten $p_n$ und $q_n = 1 - p_n$ die Werte 1 und 0 an; $\alpha_n = p_n$; $x_n = \xi_n - p_n$ nimmt die Werte $q_n$ und $-p_n$ mit den Wahrscheinlichkeiten $p_n$, $q_n$ an, ist gleichmässig beschränkt.

$$\lambda_2(x_n) = p_n \cdot q_n^2 + q_n(-p_n)^2 = p_n\, q_n.$$

Wenn also $\sum p_n q_n$ divergiert, was z. B. für $0 < \varepsilon \leq p_n \leq 1 - \varepsilon < 1$ der Fall ist, tritt das Exponentialgesetz als Grenzfall auf. Hier ist, wenn wir das Ereignis $A_n = [\xi_n = 1]$ als günstigen Fall bezeichnen, $\eta_n = \xi_1 + \cdots + \xi_n$ die Häufigkeit des günstigen Falls bei den ersten $n$ Versuchen;

$$\beta_n = p_1 + \cdots + p_n; \quad \lambda_2(y_n) = p_1 q_1 + \cdots + p_n q_n;$$

die zu $\eta_n$ gehörige normierte Variable ist

$$z_n = \frac{\eta_n - \beta_n}{\sqrt{2\lambda_2(y_n)}},$$

und die Wahrscheinlichkeit $\Phi_n(z)$, dass $z_n < z$, konvergiert gleichmässig nach

$$\Phi(z) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{z} e^{-z^2}\, dz.$$

Wenn insbesondere alle $p_n = p$ sind $(0 < p < 1)$, so ist Bl. 193

$$z_n = \frac{\eta_n - np}{\sqrt{2npq}}.$$

Numerische Beispiele für approximative Gültigkeit des Exponentialgesetzes.

Es sei $x$ eine Variable, die endlich viele Werte $k\delta$ ($\delta$ fest, $k = 0, \pm 1, \pm 2, \ldots$) mit Wahrscheinlichkeiten $p_k = \dfrac{s_k}{s}$ annimmt, wo $s_k$ und $s = \sum s_k$ natürliche Zahlen sind. Um diese Verteilung mit dem Exponentialgesetz zu vergleichen, wird man zweckmässig den Wert $k\delta$ als Repräsentanten des Intervalls

$$\left[\left(k - \frac{1}{2}\right)\delta, \left(k + \frac{1}{2}\right)\delta\right]$$

auffassen, also $p_k$ mit

$$\Phi\left(\left(k + \frac{1}{2}\right)\delta\right) - \Phi\left(\left(k - \frac{1}{2}\right)\delta\right)$$

vergleichen. Sodann wird man wegen der Symmetrie der Funktion $e^{-x^2}$ zum Nullpunkt die Werte $\pm k$ zusammennehmen. Man setzt, da $\Phi(0) = \frac{1}{2}$,

$$\Theta(x) = 2\,\Phi(x) - 1 = 2\,[\Phi(x) - \Phi(0)] = \frac{2}{\sqrt{\pi}} \int_0^x e^{-x^2}\, dx;$$

eine ungerade Funktion, die man nur für $x \geq 0$ zu betrachten braucht, und die von $\Theta(0) = 0$ nach $\Theta(+\infty) = 1$ wächst, wobei sie dem Wert 1 schon bald sehr nahe kommt:

| $x$ | $0$ | $0,5$ | $1$ | $1,5$ | $2$ | $2,5$ |
|---|---|---|---|---|---|---|
| $\Theta$ | $0,000$ | $0,520$ | $0,843$ | $0,966$ | $0,995$ | $1,000$ |

Ist $0 \leq \alpha < \beta$, so ist die Wahrscheinlichkeit, dass $|x|$ in $[\alpha, \beta]$ liege, Bl. 194

$$\Phi(\beta) - \Phi(\alpha) + \Phi(-\alpha) - \Phi(-\beta) = \Theta(\beta) - \Theta(\alpha).$$

Demnach wird man $p_k + p_{-k}$ $(k = 1, 2, \ldots)$ mit

$$\Theta\left(\left(k + \frac{1}{2}\right)\delta\right) - \Theta\left(\left(k - \frac{1}{2}\right)\delta\right)$$

und $p_0$ mit $\Theta(\frac{1}{2}\delta)$ vergleichen. Oder man wird vergleichen (Multipl. mit $s$)

| | | | | |
|---|---|---|---|---|
| wirkliche Verteilung | $s_0$ | $s_{-1} + s_1$ | $s_{-2} + s_2$ | $\cdots$ |
| Vert. nach d. Exp.ges. | $s\,\Theta(\frac{1}{2}\delta)$ | $s\,[\Theta(\frac{3}{2}\delta) - \Theta(\frac{1}{2}\delta)]$ | $s\,[\Theta(\frac{5}{2}\delta) - \Theta(\frac{3}{2}\delta)]$ | $\cdots$ |

Nehmen wir zum letzten Fall (Wiederholung eines Versuches, wo jedesmal der günstige Fall die Wahrscheinlichkeit $p$ hat) $p = q = \frac{1}{2}$, $n = 10$ an. Die Wahrscheinlichkeit, dass der günstige Fall $m$ mal eintrete, ist $\binom{10}{m}\frac{1}{2^{10}}$; dieses $m$ ist mit dem vorigen $\eta_n$, hier $\eta_{10} = \xi_1 + \cdots + \xi_{10}$ identisch. Die zu $m$ gehörige normierte Variable ist

$$z_{10} = \frac{m - 10 \cdot \frac{1}{2}}{\sqrt{2 \cdot 10 \cdot \frac{1}{2} \cdot \frac{1}{2}}} = \frac{m-5}{\sqrt{5}};$$

sie nimmt die Werte $\frac{k}{\sqrt{5}}$ $(k = 0, \pm 1, \ldots, \pm 5)$ und zwar den Wert $k$ mit der Wahrscheinlichkeit $p_k = \binom{10}{5+k} : 2^{10}$ an; setzen wir $s = 2^{10}$, $s_k = \binom{10}{5+k}$. Wir

Bl. 195 haben zu vergleichen:

| | | | | | | |
|---|---|---|---|---|---|---|
| wirkl. V.: | $\binom{10}{5}$ | $2\binom{10}{4}$ | $2\binom{10}{3}$ | $2\binom{10}{2}$ | $2\binom{10}{1}$ | $2\binom{10}{0}$ |
| EG: | $2^{10}\,\Theta(\frac{1}{2\sqrt{5}})$ | $2^{10}\,[\Theta(\frac{3}{2\sqrt{5}}) - \Theta(\frac{1}{2\sqrt{5}})]$ | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ |

Es ist $\frac{1}{2\sqrt{5}} = 0,2236$

| | | | | | | |
|---|---|---|---|---|---|---|
| $x$ | $0,2236$ | $0,6708$ | $1,118$ | $1,565$ | $2,013$ | $\infty$ |
| $\Theta(x)$ | $0,2482$ | $0,6572$ | $0,8861$ | $0,9731,$ | $0,9956$ | $1$ |
| $2^{10}\,\Theta(x)$ | $254,2$ | $673,0$ | $907,4$ | $996,4$ | $1019,5$ | $1024$ |

Also

| | | | | | | |
|---|---|---|---|---|---|---|
| wirkl. Vert. | 252 | 420 | 240 | 90 | 20 | 2 |
| Exp.gesetz | 254,2 | 418,8 | 234,4 | 89,0 | 23,1 | 4,5 |

Bl. 196

## § 9. Vergleich zwischen Theorie und Erfahrung. Methode der kleinsten Quadrate.

$\xi$ sei eine Variable mit der Vertheilungsfunction $\varphi(\xi)$ (= Wahrscheinlichkeit, dass die Variable $< \xi$ sei). Man macht $n$ unabhängige Versuche (Beobachtungen, Messungen); hierbei ergeben sich die Werthe $\xi_1, \xi_2, \ldots, \xi_n$. So entsteht, gegenüber der *theoretischen Vertheilung* nach $\varphi$, eine *empirische Vertheilung*, charakterisirt durch die Vertheilungsfunction $\psi(\xi)$ = Anzahl der $\xi_i < \xi$, dividirt durch $n$. Bei der Vergleichung beider kann man sowohl die Functionswerthe von $\varphi$, $\psi$ selbst confrontiren als auch die *theoretischen Momente*

$$M\,f(\xi) = \int f(\xi)\,d\varphi(\xi)$$

mit den *empirischen Momenten*

$$N\,f(\xi) = \int f(\xi)\,d\psi(\xi) = \frac{1}{n}\sum_{i=1}^{n} f(\xi_i)$$

vergleichen, wo $f(\xi)$ irgend eine integrable, etwa stetige Function ist.

Als „wahren Werth“ von $\xi$, wenn es sich um Messungen einer Grösse handelt, betrachten wir das Moment

$$\alpha = M\,\xi; \tag{1}$$

dies ist eine Definition analog der des „gerechten Spiels“, wo $\xi$ den | variablen Bl. 197
Gewinn, $\alpha$ den Einsatz bedeutet. Bei einer Abweichung würden wir die Beobachtungsmethode mit einem *systematischen* (nicht zufälligen) *Fehler* behaftet erklären, so wie wir beim Spiel von einem ungerechten Gewinn des Spielers oder Spielgegners sprechen. Die Differenz

$$x = \xi - \alpha \tag{2}$$

heisst der *wahre Fehler*; er ist eine Variable, deren wahrer Werth $M(\xi - \alpha) = 0$ ist. (Beim Spiel der Reingewinn).

Bisweilen ist die theoretische Vertheilung von vornherein bekannt, so in der Regel bei Zufallsspielen, wo sie sich aus dem vorausgesetzt exakten Bau der Spielapparate und aus den Spielbedingungen ergiebt. Der Vergleich mit der empirischen Vertheilung dient dann nur zur Controlle der letzteren und zur Bestätigung der theoretischen Vertheilung. – Das andere Extrem wäre, dass über $\varphi$ gar nichts bekannt ist, wo man dann ja einfach $\varphi = \psi$ setzen könnte, allerdings auf die Gefahr hin, dass eine neue Versuchsreihe dies wieder umstösst. – In der Regel ist es so: $\varphi$ ist theilweise bekannt, theilweise aus der Erfahrung zu ermitteln. Z. B. $\varphi(\xi) = \Phi(h(\xi - \alpha))$ wird als Exponentialgesetz angenommen, mit Parametern $\alpha$, $h$, die aus der empirischen Vertheilung so zu bestimmen sind, dass diese von der theoretischen möglichst wenig abweicht. Als Parameter,
| mag man nun diese oder jene Form von $\varphi(\xi)$ voraussetzen oder sich von Bl. 198
vornherein gar nicht binden wollen, empfehlen sich am meisten die *Momente* $M\,f(\xi)$ gewisser Functionen, so in erster Linie der wahre Werth $\alpha = M\,\xi$, dann die Momente

$$\mu_k = M\,(\xi - \alpha)^k = M\,x^k \tag{3}$$

($\mu_2$ Quadrat der Streuung) der Fehlerpotenzen, eventuell ihre Absolutmomente $\nu_k = M\,|x|^k$.

Das Princip dieser Bestimmung ist folgendes. Zunächst fassen wir bei den $n$ unabhängigen Beobachtungen die $\xi_1, \ldots, \xi_n$ als unabhängige Variable auf, jede derselben Vertheilungsfunction $\varphi(\xi)$ unterworfen. Ist nun

$$\eta = f(\xi_1, \xi_2, \ldots, \xi_n) \tag{4}$$

eine bekannte Function dieser Variablen, etwa ein *Polynom* mit numerischen Coefficienten, so ist ihr Moment

$$\beta = M\,\eta = \int \cdots \int f(\xi_1, \ldots, \xi_n)\, d\varphi(\xi_1) \cdots d\varphi(\xi_n) \tag{5}$$

ein (im Polynomfalle durch $\alpha$ und die $\mu_k$ ausdrückbarer) Parameter unserer Vertheilung. Denken wir uns in (4) aber die beobachteten $\xi_i$ eingesetzt, so wird

$\eta$ eine beobachtete Grösse. Sie wird natürlich durchaus nicht mit ihrem wahr-
Bl. 199 scheinlichen Werth $\beta = M\,\eta$ übereinzustimmen brauchen; | die Beobachtungen hätten ja auch andere Werthe $\xi_i$ liefern können. Wenn wir also $\beta = \eta$ setzen, so machen wir eine empirische Annahme, die wir, um sie von einer richtigen Gleichung zu unterscheiden, mit

$$\beta \sim \eta = f(\xi_1, \ldots, \xi_n) \tag{6}$$

bezeichnen wollen. Die Genauigkeit oder Zuverlässigkeit dieser Annahme bemisst Gauss nun nach der *Streuung* der Variablen $\eta$ oder ihres Fehlers $y = \eta - \beta$, (wobei man sich der Tschebyscheffschen Abschätzungen erinnern möge: je kleiner die Streuung, desto unwahrscheinlicher sind grosse Abweichungen einer Variablen von ihrem wahrscheinlichen Werth); das Quadrat dieser Streuung ist

$$\gamma = M\,(\eta - \beta)^2 = M\,\eta^2 - \beta^2 \tag{7}$$

Das ist nun zwar wieder ein neuer Parameter von derselben Art wie $\beta$, aber erstens kann man eventuell constatiren, dass $\gamma$ mit wachsender Beobachtungszahl $n$ klein wird, zweitens unter den verschiedenen Functionen $\eta$, die dasselbe $\beta$ als Moment haben, dasjenige suchen, wofür $\gamma$ ein Minimum wird, drittens auf $\gamma$ wieder dasselbe Verfahren anwenden, d. h. eine Function $\zeta = g(\xi_1, \ldots, \xi_n)$ wählen, deren Moment $M\,\zeta = \gamma$ wird, und approximativ $\gamma \sim \zeta$ setzen, die
Bl. 200 Streuung $M\,(\zeta - \gamma)^2 = M\,\zeta^2 - \gamma^2 = \delta$ bestimmen u.s.f. |

*Bestimmung von $\alpha = M\,\xi$.* Eine Function $\eta$, deren Moment gleich $\alpha$ ist, ist z. B. $\xi_i$ $(i = 1, \ldots, n)$; das Quadrat der Streuung oder des mittleren Fehlers der Annahme $\alpha \sim \xi_i$ ist

$$M\,(\xi_i - \alpha)^2 = M\,(\xi - \alpha)^2 = \mu_2.$$

Diese Annahme ist natürlich ganz schlecht; sie benutzt nur die $i^{\text{te}}$ Beobachtung, ignorirt die übrigen und ist mit dem vollen mittleren Fehler jeder einzelnen Beobachtung behaftet. Nehmen wir dagegen das arithmetische Mittel

$$\eta = \frac{1}{n}(\xi_1 + \xi_2 + \cdots + \xi_n), \tag{8}$$

so ist (Addition der logarithmischen Momente)

$$M\,(\eta - \alpha)^2 = \lambda_2(\eta) = \frac{1}{n^2} \sum_i \lambda_2(\xi_i) = \frac{1}{n}\,\mu_2$$

und die Bestimmung $\alpha \sim \eta$ ist nur noch mit dem mittleren Fehler $\sqrt{\dfrac{\mu_2}{n}}$ behaftet, der mit $n \to \infty$ nach 0 convergirt. Es ist dies auch die *günstigste* (plausibelste) Bestimmung in dem Sinne, dass ihr mittlerer Fehler ein *Minimum* wird; wenigstens wenn man sich auf Linearformen

$$\eta = \sum_i a_i\,\xi_i, \qquad \sum_i a_i = 1$$

beschränkt (die letzte Bedingung wegen $M\,\eta = \alpha$). Hier ist

$$M\,\eta^2 - \alpha^2 = \sum_i a_i^2 \cdot \mu_2$$

| und $\sum_i a_i^2$ hat unter der Bedingung $\sum_i a_i = 1$ sein Minimum für $a_i = \frac{1}{n}$. Bl. 201
(Bedeutet $d$ einen Multiplikator, so hat man die partiellen Ableitungen von $\frac{1}{2}\sum a_i^2 - d(\sum a_i - 1)$ gleich 0 zu setzen; das giebt $a_i = d$ also $= \frac{1}{n}$).

Während

$$x_i = \xi_i - \alpha \tag{9}$$

die *wahren* Fehler sind, nennt man

$$\varepsilon_i = \xi_i - \eta = x_i - \frac{1}{n}\sum_k x_k \tag{10}$$

die *scheinbaren* oder *angenommenen* Fehler (nämlich bei der Annahme $\alpha \sim \eta$).

Erinnern wir nochmals an die ursprüngliche Gausssche Ableitung des Exponentialgesetzes, die wir hier umkehren: wenn $\xi$ das Exponentialgesetz

$$\Phi(h(\xi - \alpha)) \quad \text{oder} \quad x = \xi - \alpha$$

das Exponentialgesetz $\Phi(hx)$ befolgt, so ist die Wahrscheinlichkeitsdichte an der Stelle $x_1, \ldots, x_n$ gleich

$$\left(\frac{h}{\sqrt{\pi}}\right)^n e^{-h^2 \sum x_i^2},$$

und diese wird bei gegebenen $\xi_i$, unbekanntem $\alpha$ ein Maximum, wenn $\sum x_i^2$ ein Minimum wird, d. h. (Differentiation nach $\alpha$) für

$$\sum x_i = 0, \quad \alpha = \frac{1}{n}\sum \xi_i = \eta.$$

Es lässt sich also die Bestimmung $\alpha \sim \eta$ so auffassen: man giebt dem $\alpha$ denjenigen Werth $\eta$, *für den die Summe $\sum x_i^2$ der Fehlerquadrate ein Minimum wird*; in diesem Sinne ist $\eta$ der *wahrscheinlichste* Werth von $\alpha$. (Daher der Name: Methode der kleinsten Quadrate). | Man muss dagegen mit Gauss selbst einwenden, Bl. 202
dass der wahrscheinlichste Werth an sich wenig Bedeutung hat (z. B. könnte eine Vertheilungsdichte $\vartheta(x)$ mehrere Maximalstellen haben) und dass diese Herleitung an die Gültigkeit des Exponentialgesetzes gebunden ist, während jene erste (Minimum der Streuung) über die Vertheilungsfunction nichts voraussetzt.

*Bestimmung von $\mu_2$.* Hier werden wir, wie bei $\alpha$ verfahrend, eine quadratische Form der $\xi_i$ suchen, deren Moment $= \mu_2$ ist; nehmen wir sie von vornherein (was sich nachträglich wieder aus der Forderung minimaler Streuung ergeben würde) in den $\xi_i$ symmetrisch an:

$$\zeta = a \sum_i \xi_i^2 + b \sum_{i \neq k} \xi_i \, \xi_k;$$

$$M\,\zeta = na\,(\alpha^2 + \mu_2) + n(n-1)b\,\alpha^2,$$

also

$$na = 1, \quad na + n(n-1)b = 0, \quad a = \frac{1}{n}, \quad b = -\frac{1}{n(n-1)};$$

$$\zeta = \frac{1}{n} \sum_i \xi_i^2 - \frac{1}{n(n-1)} \sum_{i \neq k} \xi_i \, \xi_k = \frac{1}{n} \sum \xi_i^2 - \frac{1}{n(n-1)} \left[ \left( \sum \xi_i \right)^2 - \sum \xi_i^2 \right]$$

$$= \frac{1}{n-1} \left( \sum \xi_i^2 - n\eta^2 \right) = \frac{1}{n-1} \sum (\xi_i - \eta)^2 = \frac{1}{n-1} \sum \varepsilon_i^2 \qquad (11)$$

Die Bestimmung $\mu_2 \sim \zeta$ hat das Quadrat des mittleren Fehlers

$$M\,(\zeta - \mu_2)^2 = M\,\zeta^2 - \mu_2^2,$$

Bl. 203 wofür sich durch eine Rechnung, auf die wir sofort zurückkommen, ergiebt

$$M\,(\zeta - \mu_2)^2 = \frac{1}{n}\,\lambda_4 + \frac{2}{n-1}\,\lambda_2^2 = \frac{1}{n}\,(\mu_4 - 3\mu_2^2) + \frac{2}{n-1}\,\mu_2^2 \qquad (12)$$

($\lambda_k$ logarithmische Momente von $\xi - \alpha$). Der mittlere Fehler von $\mu_2 \sim \zeta$ ist also abermals von der Ordnung $\frac{1}{\sqrt{n}}$ und convergirt mit $n \to \infty$ nach 0; um ihn genauer zu kennen, müsste man $\mu_4$ in gleicher Weise wie $\mu_2$ bestimmen. Befolgt $\xi$ das Exponentialgesetz, so ist $\lambda_4 = 0$ und der mittlere Fehler der Bestimmung $\mu_2 \sim \zeta$ ist $\sqrt{\frac{2}{n-1}}\,\mu_2$.

Die Berechnung von $M\,\zeta^2$ ist specieller Fall der Berechnung von Momenten beliebiger Formen in den $x_i$, oder $\varepsilon_i$. Am besten benutzt man dabei die logarithmischen Momente. Sind $\alpha_1, \ldots, \alpha_n$ beliebige Coefficienten, so bilden wir die Linearform

$$u = \sum_i \alpha_i \, \varepsilon_i = \sum_i a_i \, x_i, \qquad a_i = \alpha_i - \frac{1}{n} \sum_k \alpha_k$$

und führen die Potenzsummen

$$\sigma_k = \sum_i \alpha_i^k, \qquad s_k = \sum_i a_i^k$$

ein. Da die $x_i$ unabhängig sind (nicht die $\varepsilon_i$, z. B. ist $\sum \varepsilon_i = 0$), so ist

$$\lambda_k(u) = \sum_i a_i^k \, \lambda_k(x_i) = s_k \, \lambda_k.$$

Auf Grund der Übergangsformeln zwischen den Momenten und logarithmischen Momenten, die hier wegen $\lambda_1 = 0$ und $\lambda_1(u) = 0$ die einfachere Form

$$\mu_2 = \lambda_2, \qquad \mu_3 = \lambda_3, \qquad \mu_4 = \lambda_4 + 3\lambda_2^2, \quad \ldots$$

annehmen, hat man also

$$M\,u^2 = s_2\,\lambda_2, \qquad M\,u^3 = s_3\,\lambda_3, \qquad M\,u^4 = s_4\,\lambda_4 + 3s_2^2\,\lambda_2^2, \quad \ldots$$

Indem man hier beiderseitig die Coefficienten der Potenzprodukte der $\alpha_i$ vergleicht, erhält man die Momente der Potenzprodukte der $\varepsilon_i$ (die übrigens, weil alle $x_i$ gleiche Momente haben, bei Permutation ungeändert bleiben, z. B. $M\,\varepsilon_1^2 = M\,\varepsilon_2^2 = \cdots$, $\quad M\,\varepsilon_1\,\varepsilon_2 = M\,\varepsilon_1\,\varepsilon_3 = M\,\varepsilon_2\,\varepsilon_3 = \cdots$). Dabei ist wegen $a_i = \alpha_i - \frac{1}{n}\sigma_1$

$$s_2 = \sigma_2 - \frac{1}{n}\,\sigma_1^2, \qquad s_3 = \sigma_3 - \frac{3}{n}\,\sigma_2\sigma_1 + \frac{2}{n^2}\,\sigma_1^3,$$

$$s_4 = \sigma_4 - \frac{4}{n}\,\sigma_3\sigma_1 + \frac{6}{n^2}\,\sigma_2\sigma_1^2 - \frac{3}{n^3}\,\sigma_1^4, \quad \ldots$$

So ergiebt sich für die ersten Grade:

$$M\,\varepsilon_1 = 0, \qquad M\,\varepsilon_1^2 = \lambda_2\left(1 - \frac{1}{n}\right), \qquad M\,\varepsilon_1\varepsilon_2 = -\frac{1}{n}\,\lambda_2,$$

$$M\,\varepsilon_1^3 = \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\lambda_3, \quad M\,\varepsilon_1^2\varepsilon_2 = -\frac{1}{n}\left(1 - \frac{2}{n}\right)\lambda_3, \quad M\,\varepsilon_1\varepsilon_2\varepsilon_3 = \frac{2}{n^2}\lambda_3,$$

Bl. 205

$$\begin{aligned}
M\,\varepsilon_1^4 &= \left(1 - \frac{1}{n}\right)\left(1 - \frac{3}{n} + \frac{3}{n^2}\right)\lambda_4 + 3\left(1 - \frac{1}{n}\right)^2\lambda_2^2 \\
M\,\varepsilon_1^3\varepsilon_2 &= -\frac{1}{n}\left(1 - \frac{3}{n} + \frac{3}{n^2}\right)\lambda_4 - \frac{3}{n}\left(1 - \frac{1}{n}\right)\lambda_2^2 \\
M\,\varepsilon_1^2\varepsilon_2^2 &= \frac{1}{n^2}\left(2 - \frac{3}{n}\right)\lambda_4 + \left(1 - \frac{2}{n} + \frac{3}{n^2}\right)\lambda_2^2 \\
M\,\varepsilon_1^2\varepsilon_2\varepsilon_3 &= \frac{1}{n^2}\left(1 - \frac{3}{n}\right)\lambda_4 - \frac{1}{n}\left(1 - \frac{3}{n}\right)\lambda_2^2 \\
M\,\varepsilon_1\varepsilon_2\varepsilon_3\varepsilon_4 &= -\frac{3}{n^3}\lambda_4 + \frac{3}{n^2}\lambda_2^2
\end{aligned}$$

(Controllen: aus $M\,\varepsilon_1^{k-1}(\varepsilon_1 + \cdots + \varepsilon_n) = 0:\quad M\,\varepsilon_1^k + (n-1)M\,\varepsilon_1^{k-1}\varepsilon_2 = 0$; $M\,\varepsilon_1^{k-2}\varepsilon_2(\varepsilon_1 + \cdots + \varepsilon_n) = 0:\quad M\,\varepsilon_1^{k-1}\varepsilon_2 + M\,\varepsilon_1^{k-2}\varepsilon_2^2 + (n-2)M\,\varepsilon_1^{k-2}\varepsilon_2\varepsilon_3 = 0$ u. dgl.)

Für

$$\zeta = \frac{1}{n-1}\sum_i \varepsilon_i^2 \quad \text{ist} \quad M\,\zeta = \frac{n}{n-1}M\,\varepsilon_1^2 = \lambda_2,$$

wie wir schon wissen;

$$M\,\zeta^2 = \frac{1}{(n-1)^2}\,[n\,M\,\varepsilon_1^4 + n(n-1)\,M\,\varepsilon_1^2\varepsilon_2^2]$$

$$= \frac{1}{n-1}\left(1 - \frac{3}{n} + \frac{3}{n^2}\right)\lambda_4 + \frac{3}{n}\,\lambda_2^2 + \frac{1}{n(n-1)}\left(2 - \frac{3}{n}\right)\lambda_4 +$$

$$+\frac{n}{n-1}\left(1 - \frac{2}{n} + \frac{3}{n^2}\right)\lambda_2^2 = \frac{1}{n}\,\lambda_4 + \frac{n+1}{n-1}\,\lambda_2^2,$$

$$M\,\zeta^2 - \lambda_2^2 = \frac{1}{n}\,\lambda_4 + \frac{2}{n-1}\,\lambda_2^2\,, \quad \text{vgl. (12).}$$

Bl. 206 Bisher nahmen wir alle Beobachtungen als derselben Vertheilung unterliegend an. Bisweilen ist aber eine Beobachtung unter günstigeren Umständen gemacht als eine andere; man kann das z. B. so zum Ausdruck bringen, dass man sich die $i^{\text{te}}$ Beobachtung $p_i$ mal gemacht oder gleichwerthig mit $p_i$ einfachen Beobachtungen denkt, ihnen also Vielfachheiten oder *Gewichte* $p_i$ zugeschrieben denkt. Dann würde man statt des einfachen arithmetischen Mittels das unter Berücksichtigung der Gewichte gebildete

$$\eta = \sum p_i\,\xi_i : \sum p_i \tag{13}$$

setzen, oder statt $\sum x_i^2$ die Summe $\sum p_i\,x_i^2$ zum Minimum machen (das für $\alpha = \eta$ eintritt). Dasselbe ergiebt sich auch so: wir denken uns die $i^{\text{te}}$ Variable $\xi_i$ einer Vertheilungsfunction $\varphi_i(\xi)$ unterworfen, für die zwar auch noch $M\,\xi_i = \alpha$, aber

$$\lambda_2(\xi_i) = M\,x_i^2 = M\,(\xi_i - \alpha)^2 = m_i^2$$

nicht mehr für alle $i$ dasselbe ($= \mu_2$) zu sein braucht. Die Linearform

$$\eta = \sum a_i\,\xi_i, \qquad \sum a_i = 1,$$

für die

$$\lambda_2(\eta) = M\,(\eta - \alpha)^2 = \sum a_i^2\,m_i^2$$

(unter der Bedingung $\sum a_i = 1$) ein Minimum wird, wird jetzt durch

$$a_i\,m_i^2 = d, \qquad d = 1 : \sum \frac{1}{m_i^2},$$

also

$$\eta = \sum \frac{\xi_i}{m_i^2} : \sum \frac{1}{m_i^2}$$

Bl. 207 geliefert. Definirt man daher das Gewicht umgekehrt proportional mit $\dfrac{1}{m_i^2}$, also

$$p_1\,m_1^2 = p_2\,m_2^2 = \cdots = p_n\,m_n^2 = m^2,$$

so dass $m$ der mittlere Fehler einer einfachen Beobachtung (vom Gewicht 1) ist, so liefert (13) die kleinste Streuung, nämlich

$$M\,(\eta-\alpha)^2 = \frac{\sum p_i^2\, m_i^2}{\left(\sum p_i\right)^2} = \frac{m^2}{\sum p_i}.$$

Um die Bestimmung $\alpha \sim \eta$ anzuwenden, muss man die Gewichte als bekannt ansehen (Schätzung nach den Umständen der Beobachtung); es bleibt dann noch $m^2$, das frühere $\mu_2$, zu bestimmen. Eine Function, deren Moment $m^2$ ist, ist

$$\zeta = \frac{1}{n-1} \sum p_i\, \varepsilon_i^2,$$

wo $\varepsilon_i = \xi_i - \eta$ jetzt mit dem Werth (13) zu bilden ist; die Berechnung von $M\,(\zeta - m^2)^2$ wollen wir unterlassen, resp. auf den allgemeinen Fall der Ermittlung eines Systems von Unbekannten verschieben. | Bl. 208

*Bestimmung eines Systems von Variablen.* Denken wir uns mehrere voneinander unabhängige Variable $\eta_1, \eta_2, \dots, \eta_r$. Jede Variable $\eta_\lambda$ (die Indices $\lambda, \mu, \nu$ sollen von 1 bis $r$ gehen, die Indices $i, k, h, j$ von 1 bis $n$) hat eine Vertheilungsfunction $\psi_\lambda(\eta_\lambda)$. Es werden $n$ unabhängige Beobachtungen gemacht und bei der $i^{\text{ten}}$ Beobachtung eine lineare homogene Function mit gegebenen Coefficienten

$$\sum_\lambda b_{i\lambda}\, \eta_\lambda$$

der $\eta_\lambda$ gemessen; für sie ergebe sich der Werth $\xi_i$. Zunächst haben wir die Variable $\eta_\lambda$ des $i^{\text{ten}}$ Versuchs (so wie früher das $\xi$ beim $i^{\text{ten}}$ Versuch) besonders zu bezeichnen; nennen wir es $\eta_{\lambda i}$, wo also $\eta_{\lambda 1}, \dots, \eta_{\lambda n}$ dieselbe Vertheilungsfunction $\psi_\lambda(\eta_\lambda)$ haben. Beim $i^{\text{ten}}$ Versuch ist dann die Variable

$$\xi_i = \sum_\lambda b_{i\lambda}\, \eta_{\lambda i} \tag{14}$$

beobachtet worden; ihre Vertheilungsfunction sei $\varphi_i(\xi_i)$. Die wahren Werthe der $\eta_\lambda$ seien $\beta_\lambda = \int \eta_\lambda\, d\psi_\lambda$, die der $\xi_i$ $\quad \alpha_i = \int \xi_i\, d\varphi_i$, so dass

$$\alpha_i = \sum_\lambda b_{i\lambda}\, \beta_\lambda \tag{15}$$

| sein muss. Wir nehmen $n > r$ an; für $n \leq r$ würden die Gleichungen (15) im Allgemeinen zur Bestimmung der $\beta_\lambda$ aus den $\alpha_i$ nur eben hinreichen (oder nicht einmal das) und für die $\alpha_i$ müsste man die beobachteten $\xi_i$ setzen ($\alpha_i \sim \xi_i$), ohne über die Vertheilungen etwas ermitteln zu können. Bl. 209

Wir werden jetzt (analog dem früheren $\eta = \sum a_i \xi_i$, $\sum a_i = 1$) solche Functionen der $\xi_i$ zu suchen haben, welche die gesuchten $\beta_\lambda$ zu Momenten haben; nehmen wir sie wieder als Linearformen an

$$\eta_\lambda = \sum_i a_{\lambda i}\, \xi_i \tag{16}$$

(mit neuer Bedeutung der $\eta_\lambda$) mit noch zu bestimmenden Coefficienten $a_{\lambda i}$. Es soll also

$$\beta_\lambda = \sum_i a_{\lambda i}\, \alpha_i \tag{17}$$

sein; daraus folgt wegen (15)

$$\beta_\lambda = \sum_{i,\mu} a_{\lambda i}\, b_{i\mu}\, \beta_\mu$$

und da die $\beta_\lambda$ durch keine lineare Relation verknüpft zu denken sind,

$$\sum_i a_{\lambda i}\, b_{i\mu} = \delta_{\lambda\mu} = \frac{1}{0} \quad \text{für} \quad \lambda \underset{\neq}{\overset{=}{}} \mu. \tag{18}$$

(Dies entspricht dem früheren $\sum_i a_i = 1$).

Für die wahren Fehler $y_\lambda = \eta_\lambda - \beta_\lambda, \quad x_i = \xi_i - \alpha_i$ folgt aus (16) (17)

$$y_\lambda = \sum_i a_{\lambda i}\, x_i \tag{19}$$

Bl. 210 und danach für die Streuungsquadrate

$$M\, y_\lambda^2 = \sum_i a_{\lambda i}^2\, M\, x_i^2$$

oder, wenn wir wieder Gewichte einführen

$$q_\lambda \cdot M\, y_\lambda^2 = p_i \cdot M\, x_i^2 = m^2 \tag{20}$$

($m$ mittlerer Fehler für die Gewichtseinheit)

$$\frac{1}{q_\lambda} = \sum_i \frac{a_{\lambda i}^2}{p_i}. \tag{21}$$

Wir wählen die $a_{\lambda i}$ wieder so, dass diese sämtlichen Ausdrücke Minima werden, wobei die Gleichungen (18) zu berücksichtigen sind; man hat also die Gleichungen (18) mit einem Lagrangeschen Multiplicator $d_{\lambda\mu}$ zu versehen und die Ableitungen von

$$\frac{1}{2} \sum_i \frac{a_{\lambda i}^2}{p_i} - \sum_\mu d_{\lambda\mu} \cdot \left( \sum_i a_{\lambda i}\, b_{i\mu} - \delta_{\lambda\mu} \right)$$

nach den $a_{\lambda i}$ gleich 0 zu setzen (und zwar für jedes $\lambda$); das giebt

$$a_{\lambda i} = p_i \sum_\mu d_{\lambda\mu}\, b_{i\mu}; \tag{22}$$

wenn man mit $b_{i\nu}$ multiplicirt und nach $i$ summirt, wegen (18)

$$\delta_{\lambda\nu} = \sum_{i,\mu} d_{\lambda\mu}\, p_i\, b_{i\mu}\, b_{i\nu} = \sum_{\mu} d_{\lambda\mu}\, c_{\mu\nu}, \tag{23}$$

indem man

$$c_{\mu\nu} = \sum_{i} p_i\, b_{i\mu}\, b_{i\nu} = c_{\nu\mu} \tag{24}$$

| setzt. Diese $c_{\mu\nu}$ sind bekannt (mit den $b_{i\lambda}$ und den Gewichten $p_i$) und, falls ihre Determinante $c \neq 0$[7], nach (23) auch die $d_{\lambda\mu}$; die Matrix der $d_{\lambda\mu}$ ist zur Matrix der $c_{\lambda\mu}$ reciprok. Daher ist auch $d_{\mu\lambda} = d_{\lambda\mu}$. Danach sind auch die $a_{\lambda i}$ nach (22) bekannt. Die gesuchten Minima der Ausdrücke (21) werden dann nach (22) (18) Bl. 211

$$\sum_{i} \frac{a_{\lambda i}^2}{p_i} = \sum_{i,\mu} d_{\lambda\mu}\, a_{\lambda i}\, b_{i\mu} = \sum_{\mu} d_{\lambda\mu}\, \delta_{\lambda\mu} = d_{\lambda\lambda}$$

$$\frac{1}{q_\lambda} = d_{\lambda\lambda}, \tag{25}$$

die Gewichte der $\eta_\lambda$ gleich den Diagonalelementen der Matrix $d$.

Diese $\eta_\lambda$ sind also die Linearformen der $\xi_i$, die mit $M\,\eta_\lambda = \beta_\lambda$ die kleinsten Streuungen aufweisen, und indem man nun statt der $\xi_i$ ihre beobachteten Werthe setzt, erhält man eine angenäherte Bestimmung

$$\beta_\lambda \sim \eta_\lambda$$

für die Unbekannten. Die mit diesen $\eta_\lambda$ gebildeten Zahlen

$$\varepsilon_i = \xi_i - \sum_{\lambda} b_{i\lambda}\, \eta_\lambda \tag{26}$$

heissen die *scheinbaren* Beobachtungsfehler. Man findet nun

$$\sum_{i} p_i\, \varepsilon_i\, b_{i\mu} = \sum_{i} p_i\, \xi_i\, b_{i\mu} - \sum_{\lambda} c_{\lambda\mu}\, \eta_\lambda = 0, \tag{27}$$

weil

$$\sum_{\lambda} c_{\lambda\mu}\, \eta_\lambda = \sum_{\lambda,i} c_{\lambda\mu}\, a_{\lambda i}\, \xi_i = \sum_{i,\lambda,\nu} c_{\lambda\mu}\, p_i \xi_i\, d_{\lambda\mu}\, b_{i\nu} = \sum_{i,\nu} \delta_{\mu\nu}\, p_i \xi_i\, b_{i\nu} = \sum_{i} p_i \xi_i\, b_{i\mu}$$

| und kann auch sagen: Die Linearformen $\eta_\lambda$ sind aus den Gleichungen (27) Bl. 212

---

[7] Da $c_{\mu\nu}$ das Product der $\mu^{\text{ten}}$ und $\nu^{\text{ten}}$ Spalte der Matrix $(\sqrt{p_i}\, b_{i\mu})$ ist, so ist nach einem bekannten Satz $c$ gleich der Quadratsumme aller $r$-reihigen Determinanten dieser Matrix, also

$$c = \sum p_{i_1} p_{i_2} \cdots p_{i_r} \begin{vmatrix} b_{i_1 1} & b_{i_1 2} & \cdots & b_{i_1 r} \\ b_{i_2 1} & b_{i_2 2} & \cdots & b_{i_2 r} \\ & \cdots & & \cdots \\ b_{i_r 1} & b_{i_r 2} & \cdots & b_{i_r r} \end{vmatrix}^2$$

$(i_1, \ldots, i_r$ durchläuft alle Combinationen von $r$ Ziffern aus der Reihe $1, 2, \ldots, n)$. $c$ ist also dann und nur dann $\neq 0$, wenn die Matrix $(b_{i\mu})$ vom Range $r$ ist, wenn also die Linearformen $\sum_\lambda b_{i\lambda}\, v_\lambda$ sich nicht durch weniger als $r$ Variable ausdrücken lassen.

mit der Determinante $c$ zu bestimmen. Diese Gleichungen ergeben sich durch Nullsetzen der Ableitungen von

$$\sum_i p_i\, \varepsilon_i^2 \tag{28}$$

nach den $\eta_\mu$; d. h. die $\eta_\mu$ machen den Ausdruck (28) zum Minimum.

Man kann diese Forderung wieder, wie im Fall einer Variablen, aus der Annahme des Exponentialgesetzes herleiten. Wenn die wahren Beobachtungsfehler

$$x_i = \xi_i - \alpha_i = \xi_i - \sum_\lambda b_{i\lambda}\, \beta_\lambda \tag{29}$$

sämtlich das Exponentialgesetz $\Phi(h_i x_i)$ befolgen, wobei

$$M\, x_i^2 = \frac{1}{2h_i^2} = \frac{m^2}{p_i},$$

$h_i^2$ mit $p_i$ proportional, so wird die Wahrscheinlichkeitsdichte für das Fehlersystem $x_i$

$$\frac{h_1 \cdots h_n}{\pi^{\frac{n}{2}}}\, e^{-\sum_i h_i^2 x_i^2}$$

und dasjenige System der $\beta_\lambda$ (bei gegebenen $\xi_i$) das Wahrscheinlichste, für das $\sum h_i^2\, x_i^2$ oder $\sum p_i\, x_i^2$ ein Minimum wird, d. h. eben das System $\beta_\lambda = \eta_\lambda$.

Zur Bestimmung sämtlicher mittlerer Fehler fehlt nach (20) nur noch die von $m$; wir haben eine quadratische Form $\zeta$ in den $\xi_i$ zu suchen, deren Moment $m^2$ ist, und dann eventuell wieder $m^2 \sim \zeta$ zu setzen mit Berechnung der Streuung $M\,(\zeta - m^2)^2$. Rationell, aber mit grösserer Rechnung verknüpft ist die Aufsuchung desjenigen $\zeta$, wofür Bl. 213 | die Streuung ein Minimum ist (H. Bruns, Über die Ableitung des mittleren Fehlers, Leipziger Decanatsschrift 1892/93). Gauss giebt ein $\zeta$, das zwar im Allgemeinen nicht das günstigste ist, aber doch genügt, da seine Streuung mit $n \to \infty$ nach 0 convergirt.

Aus (26) erhält man mit (15) und (19)

$$\varepsilon_i = x_i - \sum_\lambda b_{i\lambda}\, y_\lambda = x_i - \sum_{\lambda,k} b_{i\lambda}\, a_{\lambda k}\, x_k \tag{30}$$

und wenn wir

$$g_{ik} = p_i \left(\delta_{ik} - \sum_\lambda b_{i\lambda}\, a_{\lambda k}\right) \qquad \left(\delta_{ik} = \begin{matrix}1\\0\end{matrix}\ \text{ für }\ i \mathrel{\substack{=\\ \neq}} k\right) \tag{31}$$

setzen

$$p_i \varepsilon_i = \sum_k g_{ik}\, x_k. \tag{32}$$

Die Matrix der $g_{ik}$ ist symmetrisch, da

$$p_i \sum_\lambda b_{i\lambda}\, a_{\lambda k} = p_i p_k \sum_{\lambda,\mu} b_{i\lambda}\, b_{k\mu}\, d_{\lambda\mu}$$

bei Vertauschung von $i, k$ ungeändert bleibt. Aus (30) (27) folgt

$$\sum_i p_i\, \varepsilon_i\, (\varepsilon_i - x_i) = 0,$$

also

$$\sum_i p_i\, \varepsilon_i^2 = \sum_i p_i\, \varepsilon_i\, x_i = \sum_{i,k} g_{ik}\, x_i x_k \tag{33}$$

und hieraus ($M\, x_i = 0, \quad M\, x_i^2 = \frac{m^2}{p_i}$)

$$M\,(\sum p_i\, \varepsilon_i^2) = \sum_i g_{ii}\, M\, x_i^2 = m^2 \cdot \sum_i \frac{g_{ii}}{p_i} = m^2(n-r),$$

da

$$\sum_i \frac{g_{ii}}{p_i} = \sum_i \left(1 - \sum_\lambda a_{\lambda i}\, b_{i\lambda}\right) = n - \sum_\lambda \delta_{\lambda\lambda} = n - r.$$

Also hat die Function Bl. 214

$$\zeta = \frac{1}{n-r} \sum_i p_i\, \varepsilon_i^2 \tag{34}$$

das Moment $m^2$ und kann zur Bestimmung von $m^2 \sim \zeta$ dienen. Um die Streuung zu berechnen, bilden wir

$$\left(\sum_i p_i\, \varepsilon_i^2\right)^2 = \sum_{i,k,h,j} g_{ik}\, g_{hj}\, x_i x_k x_h x_j$$

Bei der Momentbildung liefern nur folgende Glieder Beiträge:

$$\begin{array}{lcl} i = k = h = j & : & M\, x_i^4 = \mu_4(x_i) \\ i = k \neq h = j & : & M\, x_i^2 x_h^2 = \mu_2(x_i)\, \mu_2(x_h) \\ i = h \neq k = j & : & M\, x_i^2 x_k^2 = \mu_2(x_i)\, \mu_2(x_k) \\ i = j \neq k = h & : & M\, x_i^2 x_k^2 = \mu_2(x_i)\, \mu_2(x_k) \end{array}$$

Also (wegen der Symmetrie der $g_{ik}$)

$$M \left(\sum_i p_i\, \varepsilon_i^2\right)^2 = \sum_i g_{ii}^2\, \mu_4(x_i) + \sum_{i \neq h} g_{ii} g_{hh}\, \mu_2(x_i) \mu_2(x_h) + 2 \sum_{i \neq k} g_{ik}^2\, \mu_2(x_i) \mu_2(x_k),$$

und wenn man die Beschränkung der Doppelsummen wieder aufhebt

$$M \left(\sum_i p_i\, \varepsilon_i^2\right)^2 = \sum_i g_{ii}^2\, [\mu_4(x_i) - 3\mu_2(x_i)^2] + \sum_{i,k} (g_{ii}\, g_{kk} + 2 g_{ik}^2)\, \mu_2(x_i) \mu_2(x_k)$$

|

$$= \sum_i g_{ii}^2 \, \lambda_4(x_i) + m^4 \sum_{i,k} \frac{1}{p_i p_k} \, (g_{ii} g_{kk} + 2 g_{ik}^2)$$

Hier ist, wie schon gefunden war,

$$\sum_i \frac{g_{ii}}{p_i} = n - r,$$

ferner nach (32)

$$M \, p_i \, \varepsilon_i^2 = m^2 \sum_k \frac{g_{ik}^2}{p_i p_k},$$

$$M \sum_i p_i \, \varepsilon_i^2 = m^2 \sum_{i,k} \frac{g_{ik}^2}{p_i p_k} = m^2 (n-r), \qquad \sum_{i,k} \frac{g_{ik}^2}{p_i p_k} = n - r$$

Also

$$M \left( \sum_i p_i \, \varepsilon_i^2 \right)^2 = \sum_i g_{ii}^2 \, \lambda_4(x_i) + m^4 \left[ (n-r)^2 + 2(n-r) \right]$$

$$M \, \zeta^2 = \frac{1}{(n-r)^2} \sum_i g_{ii}^2 \, \lambda_4(x_i) \ + m^4 \ + 2 \frac{m^4}{n-r}$$

$$M \, (\zeta - m^2)^2 = \frac{1}{(n-r)^2} \sum_i g_{ii}^2 \, \lambda_4(x_i) + \frac{2}{n-r} \, m^4 \tag{35}$$

Dies ist also das Streuungsquadrat von $\zeta$; um es zu kennen, braucht man die $\lambda_4(x_i)$. Nimmt man etwa an (was bei Beobachtungen der Fall sein wird), dass die auf gleiche Streuung reduzirten $x_i$, d. h. die Variablen $\dfrac{x_i}{\sqrt{\mu_2(x_i)}}$ gleichmässig beschränkt sind, so wird

$$\frac{\mu_4(x_i)}{\mu_2(x_i)^2} \quad \text{oder} \quad \frac{\mu_4(x_i)}{\mu_2(x_i)^2} - 3 = \frac{\lambda_4(x_i)}{\mu_2(x_i)^2}$$

beschränkt,

$$|\lambda_4(x_i)| \leq C \, (M \, x_i^2)^2 = C \, \frac{m^4}{p_i^2},$$

und das erste Glied in (35) rechts absolut

$$\leq \frac{C m^4}{(n-r)^2} \cdot \sum_i \frac{g_{ii}^2}{p_i^2} \leq \frac{C m^4}{(n-r)^2} \sum_{i,k} \frac{g_{ik}^2}{p_i p_k} = \frac{C}{n-r} \, m^4,$$

Bl. 216 also wie das 2. von der Ordnung $\frac{1}{n}$. | Gilt für alle $x_i$ ein Exponentialgesetz, so ist $\lambda_4(x_i) = 0$ und die Streuung von $\zeta$ oder der mittlere Fehler der Annahme $m^2 \sim \zeta$ wird $\sqrt{\frac{2}{n-r}} \, m^2$.

Handelt es sich um die Bestimmung einer einzigen Unbekannten ($r = 1$), so tritt an Stelle der $r$ Linearformen die einzige

$$\eta = \frac{1}{p} \sum p_i \, \xi_i \qquad (p = \sum p_i),$$

es wird

$$\varepsilon_i = \xi_i - \eta = x_i - \frac{1}{p} \sum_k p_k \, x_k,$$

also

$$g_{ik} = p_i \left[ \delta_{ik} - \frac{p_k}{p} \right],$$

die Function

$$\zeta = \frac{1}{n-1} \sum p_i \, \varepsilon_i^2$$

hat das Moment $m^2$ (der mittlere Fehler der Annahme $\alpha \sim \eta$ war $\frac{m}{\sqrt{p}}$) und das Streuungsquadrat von $\zeta$ wird

$$\frac{1}{(n-1)^2} \sum_i g_{ii}^2 \, \lambda_4(x_i) + \frac{2}{n-1} \, m^4.$$

Nehmen wir alle Gewichte $p_i = 1$ und alle $\lambda_4(x_i) = \lambda_4$, so ist $g_{ii} = 1 - \frac{1}{n}$ und wir erhalten als Streuungsquadrat von $\zeta$ wie früher

$$\frac{1}{n} \lambda_4 + \frac{2}{n-1} \, m^4 \, .$$

## Übungsbeispiele zur Wahrscheinlichkeitsrechnung

Bl. 218

(1) (Zu § 1) Unter $n$ Urnen wird eine (die $i^{\text{te}}$ mit der Wahrscheinlichkeit $\gamma_i$) [8] gewählt; wenn die $i^{\text{te}}$ gewählt wird, ist $\alpha_i$ die Wahrscheinlichkeit, eine weisse Kugel zu ziehen. Welches ist die Wahrscheinlichkeit, eine weisse Kugel zu ziehen? $\sum_{i=1}^{n} \gamma_i \alpha_i$; wenn alle Urnen gleichwahrscheinlich sind, $\frac{1}{n} \sum_1^n \alpha_i$.

(2) (§ 1) In einer Urne befinden sich $s$ Kugeln, die Wahrscheinlichkeit, dass darunter $a$ weisse sind ($b = s - a$ schwarze), sei $\gamma_a$. Wahrscheinlichkeit, eine weisse Kugel zu ziehen? $w = \sum_{a=0}^{s} \gamma_a \cdot \frac{a}{s}$. Ist jede Vertheilung gleichwahrscheinlich, so $w = \frac{1}{s+1} \sum_0^s \frac{a}{s} = \frac{1}{2}$. Ist die Urne aus einer grösseren Urne mit $A$ weissen, $B = S - A$ schwarzen Kugeln durch Ziehung gefüllt worden, so ist $\gamma_a = \binom{A}{a} \binom{B}{b} : \binom{S}{s}$. Es ist vorauszusehen, dass $w$ dann = der Wahrscheinlichkeit sein muss, direkt aus der grossen Urne eine weisse Kugel zu ziehen, also $w = \frac{A}{S}$. Wirklich ist ja

$$\sum_0^s \binom{A}{a} \binom{S-A}{s-a} = \binom{S}{s};$$

wenn man $a$, $s$, $A$, $S$ um 1 vermindert

$$\sum_1^s \binom{A-1}{a-1}\binom{S-A}{s-a} = \binom{S-1}{s-1} \quad \text{oder} \quad \sum_0^s \binom{A}{a}\frac{a}{A}\binom{S-A}{s-a} = \binom{S}{s}\frac{s}{S},$$

also thatsächlich $\sum_0^s \gamma_a \frac{a}{A} = \frac{s}{S}$ oder $w = \frac{A}{S}$.

(3) Beim Beispiel (2) sei bekannt, dass die Urne mindestens soviel weisse
wie schwarze Kugeln enthält; die noch verbleibenden Fälle seien gleichwahr-
Bl. 219 scheinlich. Wahrscheinlichkeit, eine weisse Kugel zu ziehen?

$$\underline{s=2r} \quad w = \sum_r^{2r} \gamma_a \frac{a}{s} = \frac{1}{r+1}\cdot\frac{1}{2r}\cdot\frac{2r(2r+1)-(r-1)r}{2} = \frac{3r^2+3r}{4r(r+1)} = \frac{3}{4}$$

$$\underline{s=2r-1} \quad w = \sum_r^{2r-1} \gamma_a \frac{a}{s} = \frac{1}{r}\cdot\frac{1}{2r-1}\cdot\frac{(2r-1)2r-(r-1)r}{2}$$

$$= \frac{3r^2-r}{2r(2r-1)} = \frac{3r-1}{2(2r-1)} > \frac{3}{4}$$

(4) (§ 1) Aus $n$ Urnen, deren $i^{\text{te}}$ die Wahrscheinlichkeit $p_i$ für eine weisse Kugel liefert, werden $k$ willkürlich gewählt und aus jeder eine Kugel gezogen. Wahrscheinlichkeit, dass sämtliche Kugeln weiss sind?

Die Wahrscheinlichkeit, dass $k$ bestimmte Urnen, z. B. $1, 2, \ldots, k$, in bestimmter Reihenfolge ausgewählt werden, ist $\frac{1}{n}\cdot\frac{1}{n-1}\cdot\cdots\cdot\frac{1}{n-k+1} = \frac{(n-k)!}{n!}$, dass sie überhaupt ausgewählt werden, $\frac{(n-k)!}{n!}k!$ $(= 1 : \binom{n}{k}$, $\binom{n}{k}$ gleichwahrscheinliche Combinationen von $k$ Urnen). In diesem Falle ist $p_1 \cdots p_k$ die Wahrscheinlichkeit für Weiss. Also

$$w = \sum p_1 p_2 \cdots p_k : \binom{n}{k} = c_k : \binom{n}{k},$$

$c_k$ die Summe über alle Combinationen von $k$ Grössen $p_1 \cdots p_k = k^{\text{te}}$ elementarsymmetrische Funktion der $p_1, \ldots, p_n$, d. h. Coefficient des Polynoms

$$(x+p_1)(x+p_2)\cdots(x+p_n) = x^n + c_1\,x^{n-1} + c_2\,x^{n-2} + \cdots + c_n.$$

(5) (§ 1) Problème des parties (Pascal).

$A$ und $B$ spielen eine aus einzelnen Spielen bestehende Partie; bei jedem Spiel
gewinnt $A$ oder $B$ mit den Wahrscheinlichkeiten $p, q$ $(p+q=1)$; wenn $A$ $\alpha$
Spiele gewinnt, gewinnt er die Partie; wenn $B$ $\beta$ Spiele gewinnt, gewinnt er die
Bl. 220 Partie. Welche Wahrscheinlichkeit haben $A$ und $B$, die Partie zu gewinnen?
($w(A)$, $w(B)$.)

Nach $\alpha+\beta-1=\gamma$ Spielen ist die Partie spätestens entschieden (und soviel Spiele können auch noch, selbst nach inzwischen erfolgter Entscheidung der

Partie, gespielt werden, ohne dass über Gewinn oder Verlust der Partie ein Zweifel entsteht). $A$ gewinnt die Partie, wenn er von diesen $\gamma$ Spielen mindestens $\alpha$ gewinnt; verliert, wenn er weniger als $\alpha$ (also $B$ mindestens $\beta$) gewinnt. Demnach

$$w(A) = \sum_{\alpha}^{\gamma} \binom{\gamma}{m} p^m q^{\gamma-m}, \quad w(B) = \sum_{\beta}^{\gamma} \binom{\gamma}{n} q^n p^{\gamma-n} = \sum_{0}^{\alpha-1} \binom{\gamma}{m} p^m q^{\gamma-m},$$

$w(A) + w(B) = 1$.

Die Betrachtung gilt auch noch, wenn eine Anzahl von Spielen bereits erfolgt ist und dem $A$ noch $\alpha$ Spiele zum Gewinn fehlen, dem $B$ noch $\beta$ Spiele. Z. B. ursprünglich war $p = q = \frac{1}{2}$, und für jeden Spieler sollten 5 gewonnene Spiele zum Gewinn der Partie erforderlich sein (dann ist natürlich $w(A) = w(B) = \frac{1}{2}$). Nachdem aber $A$ die ersten zwei Spiele gewonnen hat, fehlen ihm noch $\alpha = 3$, dem $B$ noch $\beta = 5$ Spielgewinne; $\gamma = 7$,

$$w(A) = \frac{1}{2^7} \sum_{3}^{7} \binom{7}{m} = \frac{1}{128}(35 + 35 + 21 + 7 + 1) = \frac{99}{128},$$

$$w(B) = \frac{1}{2^7} \sum_{0}^{2} \binom{7}{m} = \frac{1}{128}(1 + 7 + 21) = \frac{29}{128}.$$

(6) (§ 1) Wahrscheinlichkeit, dass eine Permutation von $1, 2, \ldots, n$ an keiner Stelle mit $1, 2, \ldots, n$ übereinstimmt.

Ist $\pi_{k,n}$ die Anzahl der Permutationen, deren erste $k$ Ziffern mit $1, 2, \ldots, k$ nicht übereinstimmen, so ist

$$\pi_{k,n} = \pi_{k,n-1} + \pi_{k+1,n},$$

denn | diese Permutationen $x_1, \ldots, x_n$ ($x_1 \neq 1, \ldots, x_k \neq k$) zerfallen in die, wo $x_{k+1} = k+1$ (das sind also, von der $k+1^{\text{ten}}$ Stelle abgesehen, die Permutationen $(x_1, \ldots, x_k, x_{k+2}, \ldots, x_n)$ mit $x_1 \neq 1, \ldots, x_k \neq k$, ihre Anzahl $\pi_{k,n-1}$) und diejenigen, wo $x_{k+1} \neq k + 1$ (Anzahl $\pi_{k+1,n}$). Aus Bl. 221

$$\pi_{k+1,n} = \pi_{k,n} - \pi_{k,n-1}$$

erhält man mit $\pi_{0,n} = n!$

$$\begin{aligned} \pi_{1,n} &= n! - (n-1)! \\ \pi_{2,n} &= n! - 2(n-1)! + (n-2)! \\ \pi_{k,n} &= n! - \tbinom{k}{1}(n-1)! + \tbinom{k}{2}(n-2)! - \cdots \end{aligned}$$

Die gesuchte Wahrscheinlichkeit ist

$$w = \frac{\pi_{n,n}}{n!} = \frac{1}{n!}\left[n! - \binom{n}{1}(n-1)! + \binom{n}{2}(n-2)! - \cdots + (-1)^n\right]$$

$$= 1 - \frac{1}{1!} + \frac{1}{2!} - \cdots + (-1)^n \frac{1}{n!};$$

sie convergirt für $n \to \infty$ nach $e^{-1} = 0,3678794\cdots$

(7) (§ 1). Zwei Zeugen stimmen in einer Aussage $B$ überein; die Wahrscheinlichkeit, dass sie die Wahrheit sagen (nicht lügen oder sich nicht irren), seien $p_1$, $p_2$ ($q_1 = 1 - p_1$, $q_2 = 1 - p_2$). Wie gross ist die Wahrscheinlichkeit, dass die Aussage wahr ist?

Beobachtet ist das Ereignis $A$: beide Zeugen sagen dasselbe. Fraglich ist $B$. Gesucht $w_A(B)$.

$$w(AB) = p_1 p_2, \quad w(A\overline{B}) = q_1 q_2, \quad w(A) = p_1 p_2 + q_1 q_2.$$

Also

$$w_A(B) = \frac{p_1 p_2}{p_1 p_2 + q_1 q_2}$$

Bl. 222 (Das ist natürlich Spielerei und ohne praktischen Werth.) |

Von $m + n = l$ Zeugen, mit den Glaubwürdigkeiten $p_1, \ldots, p_l$, haben sich die $m$ ersten für, die $n$ übrigen gegen einen Thatbestand $B$ ausgesprochen. Wahrscheinlichkeit, dass $B$ wahr sei? $(q_i = 1 - p_i)$

$A$ ist wieder das beobachtete Ereignis der Zeugenaussagen.

$$w(AB) = p_1 \cdots p_m \, q_{m+1} \cdots q_l, \quad w(A\overline{B}) = q_1 \cdots q_m \, p_{m+1} \cdots p_l$$

$$w_A(B) = \frac{p_1 \cdots p_m \, q_{m+1} \cdots q_l}{p_1 \cdots p_m \, q_{m+1} \cdots q_l + q_1 \cdots q_m \, p_{m+1} \cdots p_l} \quad .$$

Bei gleicher Glaubwürdigkeit

$$w = \frac{p^m q^n}{p^m q^n + q^m p^n} = \frac{p^{m-n}}{p^{m-n} + q^{m-n}}$$

Ist $m - n$ (die Majorität) $> 0$, so ist dennoch $w < \frac{1}{2}$, wenn $p < \frac{1}{2}$, bei wenig glaubwürdigen Zeugen das Urtheil der Majorität weniger wahrscheinlich als das Gegentheil. Ist $p = \frac{2}{3}$, so ist

$$w = \frac{2^{m-n}}{2^{m-n} + 1};$$

die Wahrscheinlichkeit eines mit 4 Stimmen Majorität ergangenen Urtheils $= \frac{16}{17}$.

(8) (§ 1) Auf wieviele Arten kann man mit $n$ Würfeln die Augensumme $s$ erzielen?

Diese Anzahl $s_n$ ist der Coefficient von $t^s$ in der Entwicklung von

$$(t^1 + t^2 + \cdots + t^6)^n = \left(t \cdot \frac{1 - t^6}{1 - t}\right)^n = t^n \, (1 - t)^{-n} \, (1 - t^6)^n$$

$$= t^n + \binom{n}{1} t^{n+1} + \binom{n+1}{2} t^{n+2} + \cdots + \binom{n+4}{5} t^{n+5} + \left[\binom{n+5}{6} - \binom{n}{1}\right] t^{n+6} + \cdots$$

Die Wahrscheinlichkeit, die Augensumme $s$ mit $n$ Würfeln zu erzielen, ist $\dfrac{s_n}{6^n}$. Bl. 223 Wird zunächst mit einem Würfel die Zahl $n$ der Würfel willkürlich ermittelt, so ist die Wahrscheinlichkeit, die Summe $s$ zu erzielen,

$$= w_s = \frac{1}{6}\left[\frac{s_1}{6} + \frac{s_2}{6^2} + \cdots + \frac{s_6}{6^6}\right].$$

Bernoulli (Ars conjectandi) berechnet

$$w(s < 12) = \frac{7530}{15552}, \quad w_{12} = \frac{749}{15552}, \quad w(s > 12) = \frac{7273}{15552}$$

Obwohl $w(s > 12) < \frac{1}{2}$ ist, ist die wahrscheinliche Augensumme (§ 2), das Moment $Ms = \sum w_s \cdot s, \quad > 12$. Diese ergibt sich für 1 Würfel $= \frac{1+2+\cdots+6}{6} = \frac{7}{2}$, für $n$ Würfel $\frac{7}{2}n$, also für unser Spiel

$$= \frac{7}{2} \cdot \frac{1+2+\cdots+6}{6} = \left(\frac{7}{2}\right)^2 = \frac{49}{4}$$

(9) (§ 2). $x, y$ seien Variable mit abhängigen Vertheilungen. Wie drückt sich $\lambda_2(x+y)$ (Quadrat der Streuung) aus?

$w_{ik}$ Wahrscheinlichkeit des Grössenpaares $(x_i, y_k)$,

$$p_i = \sum_k w_{ik}, \qquad q_k = \sum_i w_{ik}$$

und nicht durchweg $w_{ik} = p_i\, q_k$. Man hat

$$\mu_2(x+y) = \sum_{i,k} w_{ik}(x_i^2 + 2x_iy_k + y_k^2) = \mu_2(x) + \mu_2(y) + 2\sum_{i,k} w_{ik}\, x_iy_k$$

$$\mu_1(x+y) = \sum w_{ik}(x_i + y_k) = \mu_1(x) + \mu_1(y)$$

$$\lambda_2(x+y) = \mu_2(x+y) - [\mu_1(x+y)]^2 = \lambda_2(x) + \lambda_2(y) + 2\sum_{i,k}(w_{ik} - p_i\, q_k)\, x_iy_k.$$

Wir haben allgemeiner ($u, v$ zwei Parameter) Bl. 224

$$M(ux+vy)^2 = u^2\, Mx^2 + 2uv\, Mxy + v^2\, My^2$$

$$[M(ux+vy)]^2 = u^2(Mx)^2 + 2uv\, MxMy + v^2(My)^2$$

$$\lambda_2(ux+vy) = a_{11}\, u^2 + 2a_{12}\, uv + a_{21}\, v^2$$

$$a_{11} = Mx^2 - (Mx)^2 = \lambda_2(x), \quad a_{12} = M(xy) - Mx\, My,$$

$$a_{22} = My^2 - (My)^2 = \lambda_2(y)\,, \qquad a_{11}a_{22} - a_{12}^2 \geq 0\,,$$

da die Form $\geq 0$ ist.

$$\frac{a_{12}}{\sqrt{a_{11}a_{22}}}$$

heisst der *Correlationscoefficient* (er ist absolut $\leq 1$); wenn $x, y$ unabhängige Vertheilungen haben, verschwindet er. (Natürlich nicht umgekehrt, da $a_{12} = 0$ nur eine von den Bedingungen $M(x^m y^n) = Mx^m \cdot My^n$ ist, oder da

$$a_{12} = \sum_{i,k} (w_{ik} - p_i\, q_k)\, x_i y_k$$

verschwinden kann, ohne dass einzeln alle $w_{ik} - p_i\, q_k$ verschwinden).

(10) Ein Schachspieler 1 habe die Wahrscheinlichkeiten $p_1, q_1, r_1$ zu gewinnen, zu verlieren, Remis zu machen; ein Spieler 2 entsprechend $p_2, q_2, r_2$. (Diese Zahlen $p, q, r$ mit $p + q + r = 1$ charakterisiren die „Spielstärke“). Wahrscheinlichkeit, dass im Spiele der beiden gegeneinander 1 gewinnt, verliert, Remis macht?

An sich, z. B. im Spiel mit einem Dritten, können alle Möglichkeiten eintreten: 1 und 2 gewinnen (Wahrscheinlichkeit $p_1 p_2$), 1 gewinnt, 2 verliert ($p_1 q_2$), 1 gewinnt, 2 macht remis ($p_1 r_2$). Als bekanntes Ereignis $A$ hat man anzusehen, dass nur die drei Fälle eintreten können: 1 gewinnt, 2 verliert; 1 verliert, 2 gewinnt; 1 und 2 machen remis. $w(A) = p_1 q_2 + q_1 p_2 + r_1 r_2$. $B$ ist eins dieser 3 Ereignisse, $w_A(B) = \frac{w(B)}{w(A)}$. | Also sind die drei Wahrscheinlichkeiten Bl. 225

$$\frac{p_1 q_2}{s}, \quad \frac{q_1 p_2}{s}, \quad \frac{r_1 r_2}{s} \qquad (s = p_1 q_2 + q_1 p_2 + r_1 r_2).$$

Zur Illustration diene das Beispiel zweier Urnen, die weisse, schwarze, rothe Kugeln mit den Wahrscheinlichkeiten $p_1, q_1, r_1$ und $p_2, q_2, r_2$ enthalten, und wo nach den Wahrscheinlichkeiten für Weiss Schwarz, Schwarz Weiss, Roth Roth gefragt wird *unter Voraussetzung A*, dass nur einer dieser 3 Fälle eintrete.

Die betr. Wahrscheinlichkeit kann sinnlos ($\frac{0}{0}$) werden, z. B. wenn $p_1 = p_2 = 1$, beide Spieler sicher gewinnen.

Bei einem Spielturnier, wo von $n$ Personen jede mit jeder andern spielt, ist die Wahrscheinlichkeit eines Gesamtergebnisses $W = w_{12} w_{13} w_{23} \cdots w_{n-1,n}$, wo für $w_{ik}$ $(i < k)$ die Wahrscheinlichkeit des betr. Ergebnisses zu setzen ist. Man könnte die Aufgabe stellen (Bayessche Regel), auf Grund der *thatsächlichen* Ergebnisse eines Spielturniers die Zahlen $p_i, q_i, r_i$ so zu bestimmen, dass $W$ möglichst gross wird.

(11) Wahrscheinlichkeit, dass ein (dekadischer) Logarithmus an der $k^{\text{ten}}$ Decimalstelle ($k = 1, 2, \ldots$) die Ziffer $i$ ($= 0, \ldots, 9$) habe. Die Wahrscheinlichkeit soll durch die Summe der betr. Intervalllängen des Numerus gemessen werden.

Bedeutet

$$\xi = x_0 + \frac{x_1}{10} + \cdots + \frac{x_{k-1}}{10^{k-1}}$$

einen $(k-1)$ stelligen Decimalbruch, so wird das Intervall

$$\xi \leq \log x < \xi + \frac{1}{10^{k-1}}$$

durch die $k^{\text{te}}$ Ziffer in die Intervalle

$$\xi + \frac{i}{10^k} \leq \log x < \xi + \frac{i+1}{10^k}$$

getheilt; das Theilintervall

$$\left(10^{\xi+\frac{i}{10^k}} \leq x < 10^{\xi+\frac{i+1}{10^k}}\right) \text{ hat die Länge } 10^{\xi+\frac{i}{10^k}} \cdot \left(10^{\frac{1}{10^k}} - 1\right),$$

das Gesamtintervall die Länge

$$10^{\xi} \cdot \left(10^{\frac{1}{10^{k-1}}} - 1\right);$$

der Quotient ist Bl. 226

$$10^{\frac{i}{10^k}} \cdot \frac{10^{\frac{1}{10^k}} - 1}{10^{\frac{1}{10^{k-1}}} - 1} = \delta^i \frac{\delta - 1}{\delta^{10} - 1} = \frac{\delta^i}{1 + \delta + \cdots + \delta^9} \qquad (\delta = 10^{\frac{1}{10^k}}),$$

und dies ist die gesuchte Wahrscheinlichkeit $w_{k,i}$ (werden z. B. die Zahlen $1 \leq x < 10$ betrachtet, so zerfällt dies in Intervalle, den verschiedenen $\xi$ entsprechend, und die gesuchte Wahrscheinlichkeit ist

$$\sum_{\xi} 10^{\xi} \cdot \delta^i(\delta - 1) : \sum_{\xi} 10^{\xi}(\delta^{10} - 1) = w_{k,i}).$$

Bei festem $k$ wachsen die $w_{k,i}$ in geometrischer Reihe $(w_{k,0} : w_{k,1} : \cdots : w_{k,9} = 1 : \delta : \cdots : \delta^9)$; z. B. (vgl. Urban, S. 83)

| | $k = 1$ | $k = 2$ | $k = 3$ |
|---|---|---|---|
| $i = 0$ | 0,02877 | 0,08996 | 0,09897 |
| 1 | 03622 | 09206 | 09920 |
| 2 | 04560 | 09420 | 09942 |
| 3 | 05740 | 09639 | 09965 |
| 4 | 07227 | 09864 | 09988 |
| 5 | 0,09097 | 0,10094 | 0,10011 |
| 6 | 11453 | 10329 | 10034 |
| 7 | 14419 | 10569 | 10058 |
| 8 | 18152 | 10816 | 10081 |
| 9 | 22852 | 11068 | 10104 |

für $k \to \infty$ convergirt jedes $w_{k,i}$ nach $\frac{1}{10}$. $(\delta \to 1)$

Verallgemeinerung: $y = f(x)$ sei für $a \leq x \leq b$ stetig und habe eine stetige positive Ableitung; es gilt das Gleiche für die inverse Funktion $x = \varphi(y)$ im Intervall $A \leq y \leq B$ $(A = f(a),\ B = f(b))$. Es sei $0 \leq \alpha < \beta \leq 1$ und die Wahrscheinlichkeit $w_n$ gesucht, dass $\alpha \leq ny - [ny] < \beta$, wo $n$ eine natürliche Zahl, $[z]$ die grösste ganze Zahl $\leq z$ bedeutet; $w_n$ soll erklärt sein als Summe der Intervalllängen derjenigen $x$, wo die fragliche Ungleichung erfüllt ist, dividirt durch $b - a$. Setzt man $[ny] = m$, so lautet die fragliche Ungleichung

$$\frac{m+\alpha}{n} \leq y < \frac{m+\beta}{n}$$

(Theilintervall $I_m$ von $K_m: \ \frac{m}{n} \leq y < \frac{m+1}{n}$) und die zugehörige Intervallsumme von $x$ wird

$$\sum_m \int_{\frac{m+\alpha}{n}}^{\frac{m+\beta}{n}} \varphi'(y)\, dy = \frac{\beta - \alpha}{n} \sum_m \varphi'(y_m),$$

$y_m$ ein Werth in $I_m$. Die Summe $\frac{1}{n} \sum_m \varphi'(y_m)$ ist ein Näherungswerth von

$$\sum_m \int_{\frac{m}{n}}^{\frac{m+1}{n}} \varphi'(y)\, dy = \int_A^B \varphi'(y)\, dy = \int_a^b dx = b - a$$

und convergirt nach ihr für $n \to \infty$. Also convergirt $w_n$ nach $\beta - \alpha$.

Nimmt man speciell $n = 10^{k-1},\ \alpha = \frac{i}{10},\ \beta = \frac{i+1}{10}\ \ (i = 0, \ldots, 9)$, so findet man: die Wahrscheinlichkeit, dass $y$ an der $k^{\text{ten}}$ Decimalstelle die Ziffer $i$ habe, convergirt für $k \to \infty$ nach $\frac{1}{10}$.

## Anmerkungen

[1] Hausdorff schreibt $\dot{+}$ für $\cup$, $\underset{n}{\mathfrak{S}}$ für $\bigcup\limits_n$ und $\underset{n}{\mathfrak{D}}$ für $\bigcap\limits_n$. Hausdorffs Bezeichnungen sind hier und im Folgenden der besseren Lesbarkeit halber durch die heute üblichen ersetzt worden.

[2] Blatt 10 ist komplett getilgt.

[3] Hier hat Hausdorff eingefügt: J. Radon, *Theorie und Anwendungen der absolut additiven Mengenfunctionen*, Wiener Ak. Ber. 122 (1913), S. 1295–1438.

[4] An der Stelle $(*)$ steht „s. umst.“ und auf Bl. 104v heißt es: „Abkürzung im SS 1931: Intervallintegrale; $\int f\, d\chi = \int_{-\infty}^{\infty} f'(x)\vartheta(x)\, dx$, falls $\chi(x) = \int_{-\infty}^{x} \vartheta(x)\, dx$, $f$ und $\vartheta$ stetig. Dann § 8: Momentproblem u. (zweiter) Grenzwertsatz. Exponentialgesetz.“

[5] Bl. 111 ist im Original komplett getilgt.

[6] Bl. 118 ist komplett getilgt.

[7] Die Bll. 181–195 enthalten eine etwas abgekürzte Version von „Das Exponentialgesetz“ (§ 7). Die modernere Orthographie (Funktion, Wert, Verteilung etc.) deutet darauf hin, daß diese Version später verfaßt wurde und beim Vortrag im Sommersemester 1931 Verwendung fand.

[8] Bl. 217 ist komplett getilgt.