

Zeitschrift: Mitteilungen / Vereinigung Schweizerischer Versicherungsmathematiker
= Bulletin / Association des Actuaires Suisses = Bulletin / Association of
Swiss Actuaries

Herausgeber: Vereinigung Schweizerischer Versicherungsmathematiker

Band: 19 (1924)

Artikel: Die Wahrscheinlichkeitsansteckung : ein Beitrag zur theoretischen
Statistik

Autor: Eggenberger, Florian

DOI: <https://doi.org/10.5169/seals-550865>

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. [Mehr erfahren](#)

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. [En savoir plus](#)

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. [Find out more](#)

Download PDF: 22.07.2025

ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>

Die Wahrscheinlichkeitsansteckung.

Ein Beitrag zur theoretischen Statistik.

Von Dr. Florian Eggenberger.

Einleitung.

§ 1. Problemstellung.

Eine ganz genaue Übereinstimmung der effektiven mit der erwartungsmässigen Bernoullischen Dispersion ist eine Seltenheit in der Statistik. Sind diese Resultate der Statistik ein negativer Beweis der Anwendbarkeit der Wahrscheinlichkeitsrechnung in der Statistik? Wir müssen uns bei der Beantwortung dieser Frage klar vor Augen halten, dass man zur theoretischen Ermittlung der Dispersion in ausgedehntem Masse von dem Bernoullischen Urnenschema Gebrauch macht. Daraus folgt, dass die gefundenen Relationen unmittelbar nur bei solchen Aufgaben Anwendung finden können, die sich auf das genannte Schema reduzieren lassen. Der Anwendungsbereich beschränkt sich somit auf Reihen, die analog aufgebaut sind wie solche, die vom Bernoullischen Schema herrühren. Auf die oben aufgeworfene Frage werden wir also erklären:

Die beobachtete Stabilität stimmt deswegen bei weitem nicht mit der erwartungsmässigen überein, weil sich die Vorgänge im statistischen Material nicht durch das einfachste Bernoullische Urnenschema charakterisieren lassen. Bevor wir bei einem bestimmten statistischen Material die Anwendbarkeit der Wahrscheinlichkeitsrechnung verneinen, müssen wir vor allem die Struktur von diesem Material studieren und können dann höchstens erklären: Mit den vorliegenden Mitteln der Wahrscheinlichkeitsrechnung kann das betreffende

statistische Material nicht auf den methodologisch wichtigen Kern untersucht, und dieser von der Schlacke des Zufälligen gereinigt werden.

Betrachten wir etwa als Beobachtungsmaterial die Bevölkerung eines Schweizerkantons, als Ereignis das Eintreffen eines Unfalles, der bei einem Individuum mindestens einen Tag Arbeitsunfähigkeit hervorruft. Den Umfang einer Beobachtungsserie begrenzen wir mit einem Kalenderjahr. Wir betrachten 40 Serien (1880—1919). Wird die effektive mit der erwartungsmässigen Dispersion übereinstimmen? Die Antwort wird verneinend ausfallen. Beim wiederholten Ziehen aus einer Urne wird sich die Einzelwahrscheinlichkeit nicht ändern; bei unserm Beispiel ändert aber die Unfallwahrscheinlichkeit von Serie zu Serie (Entwicklung der Industrie). Wir müssen ein Urnenschema konstruieren, das unserem Material besser entspricht, d. h. wir müssen die Einzelwahrscheinlichkeiten für die Urne von Serie zu Serie ändern.

Es erscheint uns sofort begreiflich, dass das gewöhnliche Urnenschema etwa nicht auf die Statistik der Todesfälle infolge Dampfkesselexplosionen angewendet werden kann. Die Züge aus der gewöhnlichen Urne sind unabhängig von einander, was bei den oben erwähnten Todesfällen nicht behauptet werden kann, da bei einer Dampfkesselexplosion gewöhnlich mehrere Personen ums Leben kommen. Die Leben der einzelnen Personen sind miteinander solidarisch. Ein Todesfall infolge Dampfkesselexplosion vergrössert für die übrigen Beteiligten die Wahrscheinlichkeit, bei einer solchen Explosion zu sterben. Das ideale Urnenschema für dieses Material wäre nun ein solches, bei welchem die einzelnen „Züge“ in derselben Weise von einander abhängig sind wie die eben erwähnten Todesfälle. Die ganze Sache

würde aber so kompliziert, dass damit für die Statistik nichts gewonnen wäre. Als wichtige Bedingung müssen wir noch stellen: Das Schema muss einfach sein.

Das Hauptziel der vorliegenden Arbeit ist, ein einfaches wahrscheinlichkeitstheoretisches Schema, das die Verkettung der Einzelfälle bei Explosionen, Epidemien usw. in annehmbarer Annäherung beschreibt (Definition in § 3), mathematisch und statistisch zu untersuchen.

§ 2. Statistisches Material.

Haben wir nun ein bestimmtes Schema, das einfach ist und von dem wir vermuten, dass es die Vorgänge der Wirklichkeit in der Hauptsache gut wiedergibt, so ist es dann einzig und allein Sache der Erfahrung, und damit der Statistik, den Beweis zu erbringen, dass jene Sätze, die wir aus unserem Schema herleiten, auch wirklich brauchbar sind. Trifft das zu, dann haben wir mehr gewonnen, als wenn wir für bestimmte, beobachtete Ereignisse eine möglichst einfache Funktion gefunden haben, die uns dieselben mit der gewünschten Genauigkeit wiedergibt. In der Tat entspricht das Urnenschema der Struktur der Ereignisse; von der Funktion können wir aber nichts weiter aussagen, als dass sie innerhalb der in Betracht kommenden Grenzen die Erfahrungsergebnisse wiedergibt. Natürlich ist den Sätzen, soweit sie sich in einem gegebenen Falle als brauchbar erweisen, immer nur die Bedeutung einer Annäherung zuzusprechen.

Um das Urnenschema von Hrn. Prof. Pólya, auf das ich im nächsten Paragraphen zurückkommen werde, auf seine Anwendbarkeit zu prüfen, habe ich alle Sätze, die daraus abgeleitet wurden, an verschiedenem statistischem Material, der Bevölkerungsstatistik entnommen, geprüft. (Betreffend Quellen vgl. Zahlentafel 1 bis 11.)

In den ersten Beispielen befasse ich mich mit der Statistik der Todesfälle bei Explosionen. Bei allen Beispielen wurde zuerst untersucht, ob das einfache Urnenschema von Bernoulli, d. h. in unserem Falle, wegen der Seltenheit der Ereignisse, das Poissonsche Gesetz Anwendung finde. Erst wenn wir festgestellt haben, dass die Reihen der beobachteten Ereignisse nicht analog aufgebaut sind wie solche, die vom gewöhnlichen Schema herrühren, schreiten wir zu weiteren Untersuchungen. Wie ich schon im vorangehenden Paragraphen bemerkt habe, müssen wir annehmen, dass die Todesfälle bei Explosionen (nicht aber die Explosionen) von einander abhängig sind. Diese Annahme wurde durch genauere Untersuchungen bestätigt.

Im weitern habe ich die Statistik der Todesfälle an einzelnen Infektionskrankheiten (Pocken, Scharlach) in der Schweiz untersucht. Das Resultat der Untersuchungen ergab die Nichtanwendbarkeit des gewöhnlichen Urnenschemas auf diese Infektionskrankheiten. Die Sätze hingegen, die wir aus dem Urnenschema von Pólya herleiten, finden sich in sehr befriedigender Weise bestätigt, und dadurch ist die Brauchbarkeit dieses Schemas bewiesen. Eine Untersuchung der Statistik der *Erkrankungen* an Pocken im Kanton Zürich (15. Febr. 1919 bis 22. September 1923) hat ergeben, dass die Reihen der Erkrankungen auch diesem Schema gemäss aufgebaut sind.

Ferner wurden dann noch an zwei Beispielen, der weiblichen Selbstmorde in 8 deutschen Staaten und der tödlichen Unfälle bei 11 Berufsgenossenschaften, einige Sätze über Inhomogenität geprüft.

Kapitel I.

**Theorie der Wahrscheinlichkeits-
ansteckung.**

§ 3. Die Struktur der Wahrscheinlichkeitsansteckung.

Die Konstruktion eines Urnenschemas, das einfach ist und in der Statistik auf den ersten Blick unerwartet gute Resultate liefert, sobald es sich um „ansteckende Wahrscheinlichkeiten“ handelt, verdanken wir Herrn Prof. Dr. Pólya in Zürich. Die Resultate der Untersuchungen über die Anwendbarkeit des Urnenschemas von Pólya auf epidemische Krankheiten, bei denen die Todesfälle in gewisser Hinsicht solidarisch sind, wurden von mir bereits veröffentlicht (vgl. Pólya 2).

In den ersten zwei Kapiteln der vorliegenden Arbeit soll die Theorie der Ansteckung zunächst weiter ausgebaut werden, und dann sollen vor allem die Konsequenzen dieser Theorie für die Statistik diskutiert werden. Der Vollständigkeit wegen finden sich in diesem, wie in dem nächstfolgenden Kapitel zum Teil Resultate, die bereits a. a. O. veröffentlicht wurden (vgl. Pólya 2).

In einer Urne befinden sich zu Beginn des Spieles R rote und S schwarze, insgesamt $R + S = N$ Kugeln. Wir ziehen nun eine Kugel aus der Urne und legen hierauf $1 + \Delta$ Kugeln von der Farbe der Gezogenen in die Urne zurück. Aus der Urne, die nun $N + \Delta$ Kugeln enthält, machen wir einen zweiten Zug und legen wiederum $1 + \Delta$ Kugeln von der Farbe der Gezogenen in die Urne zurück. Diese Operation wiederholen wir n -mal. Nach dem n -ten Zug befinden sich in der Urne $N + n\Delta$ Kugeln. Haben wir in diesen n Zügen r rote und s schwarze Kugeln gezogen,

$r + s = n$, so setzt sich der Inhalt der Urne aus $R + r\Delta$ roten und $S + s\Delta$ schwarzen Kugeln zusammen ¹⁾. Die Wahrscheinlichkeit,

im 1. Zug rot bzw. schwarz zu ziehen ist $\frac{R}{N}$ bzw. $\frac{S}{N}$,

im 2. Zug rot bzw. schwarz zu ziehen ist $\frac{R + \Delta}{N + \Delta}$
bzw. $\frac{S}{N + \Delta}$, wenn im 1. Zug rot gezogen wurde,

im 2. Zug rot bzw. schwarz zu ziehen ist $\frac{R}{N + \Delta}$
bzw. $\frac{S + \Delta}{N + \Delta}$, wenn im 1. Zug schwarz gezogen wurde,

im $(n + 1)$. Zug rot bzw. schwarz zu ziehen ist $\frac{R + r\Delta}{N + n\Delta}$
bzw. $\frac{S + s\Delta}{N + n\Delta}$, wenn in den ersten n Zügen r rote und s schwarze Kugeln herausgekommen sind.

Falls $\Delta > 0$ ist, vergrößert der Zug einer roten Kugel die Chancen für rot, der Zug einer schwarzen Kugel diejenigen für schwarz. Man kann daher dieses Schema auch *Urnschema der Chancenvermehrung* nennen.

Im Falle $\Delta \geq 0$ lässt sich die Ziehungsreihe ins Unendliche fortsetzen. Ist hingegen $\Delta < 0$, so verschlechtern sich die Chancen mit dem Erfolg. Die Ziehungsreihe ist nicht mehr unbeschränkt fortsetzbar. Soll die Ziehung im Falle $\Delta < 0$ bis zum n -ten Zug fortgesetzt werden können, so müssen auch bei den extremsten Eventualitäten nach dem $(n - 1)$ -ten Zug noch rote und schwarze Kugeln in der Urne vorhanden sein, d. h. es müssen folgende Ungleichungen gelten

¹⁾ Vom Zufall abhängige Größen sind durch Fettdruck hervorgehoben.

$$R + (n-1) \Delta > 0, \quad S + (n-1) \Delta > 0$$

und somit

$$\Delta > -\frac{R}{n-1}, \quad \Delta > -\frac{S}{n-1}.$$

§ 4. Die Wahrscheinlichkeiten.

Wir fragen nun nach der Wahrscheinlichkeit $p_{r,s}$ in n Zügen aus der im vorhergehenden Paragraphen definierten Urne der Chancenvermehrung r rote und s schwarze Kugeln zu erhalten, wobei $s = n - r$ ist, und für r die Werte $0, 1, 2, \dots, n$ in Betracht kommen. Nach dem Multiplikationssatz der Wahrscheinlichkeiten ist die Wahrscheinlichkeit in den ersten r Zügen nur rot, in den folgenden $s = n - r$ Zügen nur schwarz zu ziehen, gleich

$$\frac{R(R+\Delta)(R+2\Delta)\dots(R+[r-1]\Delta)S(S+\Delta)\dots(S+[s-1]\Delta)}{N(N+\Delta)(N+2\Delta)\dots(N+[n-1]\Delta)} \quad (1).$$

Dieser Quotient stellt aber zugleich auch die Wahrscheinlichkeit dafür dar, dass r bestimmte unter den n Zügen rot, die übrigen schwarz ergeben. Denn wenn man den entsprechenden Produktausdruck der Wahrscheinlichkeiten der einzelnen Proben bildet, so erhält man wieder n Brüche, deren Nenner ebenso lauten wie die Faktoren des Nenners von (1), die Zähler hingegen sind nur in der Reihenfolge vertauscht. Die Berechnung von $p_{r,s}$ kommt somit nur auf eine Permutation der Faktoren im Zähler unseres Quotienten (1) hinaus. Da man aber aus n Elementen r auf $\binom{n}{r}$ Arten herausheben kann, ist die gesuchte Wahrscheinlichkeit

$$p_{r,s} = \binom{n}{r} \frac{R(R+\Delta)(R+2\Delta)\dots(R+[r-1]\Delta) S(S+\Delta)\dots(S+[s-1]\Delta)}{N(N+\Delta)(N+2\Delta)\dots(N+[n-1]\Delta)} \quad (2).$$

Setzen wir zur Abkürzung

$$\frac{R}{N} = \varrho, \quad \frac{S}{N} = \sigma, \quad \frac{\Delta}{N} = \delta \quad (3),$$

so erhalten wir

$$p_{r,s} = \binom{n}{r} \frac{\varrho(\varrho+\delta)(\varrho+2\delta)\dots(\varrho+[r-1]\delta) \sigma(\sigma+\delta)\dots(\sigma+[s-1]\delta)}{1(1+\delta)(1+2\delta)\dots(1+[n-1]\delta)} \quad (4)$$

oder auch

$$p_{r,s} = \frac{\binom{-\frac{\varrho}{\delta}}{r} \binom{-\frac{\sigma}{\delta}}{s}}{\binom{-\frac{1}{\delta}}{n}}, \quad s = n - r \quad (5).$$

Wir werden nun häufig von folgenden bekannten Identitäten zwischen den Binominalkoeffizienten Gebrauch machen:

$$\sum_{r=0}^n \binom{a}{r} \binom{b}{n-r} = \binom{a+b}{n}$$

$$\sum_{r=0}^n \binom{-a}{r} \binom{-b}{n-r} = \binom{-(a+b)}{n} \quad (6)$$

$$\binom{-a}{k} = \binom{a+k-1}{k} (-1)^k \quad (7)$$

$$\binom{a}{r} \binom{r}{b} = \binom{a}{r-b} \binom{a+b-r}{b} \quad (8).$$

Nun bilden wir zur Kontrolle die Summe der Wahrscheinlichkeiten, die bekanntlich 1 sein muss. Nach (5) ist

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

$$\sum_{r=0}^n p_{r, n-r} = \frac{\sum_{r=0}^n \binom{-\frac{\sigma}{\delta}}{r} \binom{-\frac{\sigma}{\delta}}{n-r}}{\binom{-\frac{1}{\delta}}{n}}$$

also nach (6)

$$\sum_{r=0}^n p_{r, n-r} = \frac{\binom{-\frac{1}{\delta}}{n}}{\binom{-\frac{1}{\delta}}{n}} = 1.$$

Mit Hilfe von (7) geht (5) über in

$$p_{r, s} = \frac{\binom{\frac{\sigma}{\delta} + r - 1}{r} \binom{\frac{\sigma}{\delta} + s - 1}{s}}{\binom{\frac{1}{\delta} + n - 1}{n}}, \quad r + s = n. \quad (9)$$

§ 5. Verlauf der Wahrscheinlichkeiten.

In diesem Paragraphen sei n als fest gedacht, zur Abkürzung

$$p_{r, n-r} = p_r$$

gesetzt und $\sigma < \sigma$ angenommen (was keine wesentliche Einschränkung bedeutet). Aus § 3 folgt ohne weiteres, dass $\delta > -\frac{\sigma}{n-1}$ ist. Wir wollen nun den Verlauf der Reihe der Wahrscheinlichkeiten

$$p_0, p_1, p_2, \dots, p_n \quad (10)$$

untersuchen.

Aus der Annahme $\varrho < \sigma$ folgt unmittelbar

$$\begin{aligned} & \sigma(\sigma + \delta)(\sigma + 2\delta) \dots (\sigma + [n-1]\delta) \\ & > \varrho(\varrho + \delta)(\varrho + 2\delta) \dots (\varrho + [n-1]\delta) \end{aligned}$$

also gemäss (4)

$$p_0 > p_n.$$

Um weiteres zu erfahren, bilden wir nun den Quotienten

$$\begin{aligned} \frac{p_r}{p_{r-1}} &= \frac{n-r+1}{r} \cdot \frac{\varrho + (r-1)\delta}{\sigma + (n-r)\delta} \\ &= \frac{r-1 + \frac{\varrho}{\delta}}{r} \cdot \frac{n+1-r}{n + \frac{\sigma}{\delta} - r} = f(r) \end{aligned}$$

$$r = 1, 2 \dots n.$$

Es findet im Übergang von p_{r-1} zu p_r Abnahme oder Zunahme statt, je nachdem $f(r) < 1$ oder $f(r) > 1$ ist.

Man beachte nun zweierlei: Erstens ist die Funktion $f(r)$ eine stetige Funktion von r , wenn r als eine im Intervalle $1 \leq r \leq n$ stetige Veränderliche aufgefasst wird. Zweitens hat die Gleichung

$$f(r) = 1$$

nur eine Wurzel

$$r_0 = \frac{(n+1)(\varrho - \delta)}{\varrho - \delta + \sigma - \delta} = \frac{(n+1)(\varrho - \delta)}{1 - 2\delta}.$$

Es folgt hieraus: die Funktion $f(r)$ kann entweder keinmal oder einmal durch den Wert 1 hindurchgehen. Wenn $f(1) = 1$ und $f(n) = 1$ von gleichem Vorzeichen

sind, so liegt r_0 ausserhalb des Intervalles $(1, n)$, $f(r) - 1$ bewahrt daselbst ein festes Vorzeichen, und die Reihe der Wahrscheinlichkeiten (10) ist monoton. Falls $f(1) - 1$ und $f(n) - 1$ von entgegengesetztem Vorzeichen sind, liegt r_0 innerhalb des Intervalles $(1, n)$, $f(r) - 1$ ändert das Vorzeichen im Punkte $r = r_0$. Die Reihe der Wahrscheinlichkeiten (10) ist sowohl links von r_0 als auch rechts von r_0 monoton, aber links in anderem Sinne als rechts: ist $f(1) > 1 > f(n)$, so nimmt die Reihe (10) anfänglich zu, später ab (glockenförmiger Verlauf), ist dagegen $f(1) < 1 < f(n)$, so nimmt die Reihe anfänglich ab, später zu (U-förmiger Verlauf). Es ist

$$f(1) - 1 = \frac{n\varrho - \sigma - \delta(n-1)}{n\delta - \delta + \sigma},$$

$$f(n) - 1 = \frac{\varrho + (n-1)\delta - n\sigma}{n\sigma},$$

und die möglichen Fälle sind in folgender Tafel zusammengefasst:

δ zwischen	$f(1) - 1$	$f(n) - 1$	Verlauf
$-\frac{\varrho}{n-1}, \varrho - \frac{\sigma - \varrho}{n-1}$	+	-	glockenförmig
$\varrho - \frac{\sigma - \varrho}{n-1}, \sigma + \frac{\sigma - \varrho}{n-1}$	-	-	monoton fallend
$\sigma + \frac{\sigma - \varrho}{n-1}, \infty$	-	+	U-förmig

Es ist bezeichnend, dass die U-förmige Kurve durch grosse Chancenvermehrung hervorgerufen wird. Die U-förmige Kurve, mit zugehörigen statistischen Beispielen, findet sich in den Lehrbüchern der Statistik

verschiedentlich erwähnt (vgl. Yule, S. 103). Pearson hat auch eine Funktion ermittelt, die den Verlauf von solchen statistischen Reihen genügend gut wiedergibt. Ein Schema von Wahrscheinlichkeiten, die eine U-förmige Kurve erzeugen, ist meines Wissens bis anhin nicht erwähnt worden. Wir haben hier also eine mögliche Erklärung für den U-förmigen Verlauf der Reihe der Wahrscheinlichkeiten gefunden.

Es sei beim glockenförmigen Verlauf p_μ das Maximum der Reihe der Wahrscheinlichkeiten (10). Dann ist

$$f(u) = \frac{p_\mu}{p_{\mu-1}} > 1, \quad f(u+1) = \frac{p_{\mu+1}}{p_\mu} < 1.$$

Für r_0 erhalten wir somit die Ungleichung

$$\mu < r_0 < \mu + 1.$$

Bezeichnen wir wie üblich die zu a links benachbarte ganze Zahl mit $[a]$, dann ist

$$\mu = [r_0] = \left[\frac{(n+1)(\varrho - \delta)}{1 - 2\delta} \right] \quad (11).$$

Bei den Untersuchungen dieses Paragraphen wurde die Diskussion des Gleichheitszeichens in den Ungleichungen weggelassen. Der Fall des Gleichheitszeichens ist in der Tat praktisch belanglos und sofort durch Kontinuität erhältlich.

Wie wir aus obiger Tabelle sehen, kommen wir mit wachsendem δ von der Glockenform zur U-Form, indem wir das Gebiet des monotonen Fallens durchschreiten. Dieses Gebiet wird kleiner, je näher ϱ an σ kommt und fällt für $\varrho = \sigma$ weg. In diesem Fall haben wir glockenförmigen Verlauf für $\delta < \frac{1}{2}$, U-förmigen Verlauf für $\delta > \frac{1}{2}$. Dem Übergang von der Glocken-

form zur U-Form entspricht das Passieren des Wertes δ durch $\frac{1}{2}$. In diesem bemerkenswerten Spezialfall $\varrho = \sigma = \delta = \frac{1}{2}$ sind alle Wahrscheinlichkeiten einander gleich, nämlich $= \frac{1}{h+1}$

§ 6. Berechnung der Erwartungen.

Bezeichnen wir die mathematische Erwartung von \mathbf{a} mit $\{\mathbf{a}\}$ (der Gebrauch der spitzen Klammern $\{\}$ wird für diesen Zweck reserviert), so wird

$$\left\{ \binom{r}{k} \right\} = \sum_{r=0}^n \binom{r}{k} p_{r, n-r},$$

also nach Formel (9)

$$\left\{ \binom{r}{k} \right\} = \sum_{r+s=n}^{0 \dots n} \binom{r}{k} \frac{\binom{\varrho+r-1}{r} \binom{\sigma+s-1}{s}}{\binom{\frac{1}{\delta}+n-1}{n}}.$$

Nun ist aber nach (8)

$$\binom{r}{k} \binom{\varrho+r-1}{r} = \binom{\varrho+r-1}{r-k} \binom{\varrho+k-1}{k}$$

und daraus folgt

$$\begin{aligned} \left\{ \binom{r}{k} \right\} &= \frac{\sum_{r+s=n}^{0 \dots n} \binom{\varrho+r-1}{r-k} \binom{\varrho+k-1}{k} \binom{\sigma+s-1}{s}}{\binom{\frac{1}{\delta}+n-1}{n}} \\ &= \frac{\binom{\varrho+k-1}{k} \sum_{r+s=n}^{0 \dots n} \binom{\sigma+s-1}{s} \binom{\varrho+r-1}{r-k}}{\binom{\frac{1}{\delta}+n-1}{n}}; \end{aligned}$$

nach (6) und (7) erhalten wir somit

$$\begin{aligned}
 \left\{ \binom{r}{k} \right\} &= \frac{\binom{\frac{\varrho}{\delta} + k - 1}{k}}{\binom{\frac{1}{\delta} + n - 1}{n}} \binom{\frac{1}{\delta} + n - 1}{n - k} \\
 &= \frac{\left(\frac{\varrho}{\delta} + k - 1\right)! \left(\frac{1}{\delta} + n - 1\right)! n! \left(\frac{1}{\delta} - 1\right)!}{\left(\frac{1}{\delta} + n - 1\right)! k! \left(\frac{\varrho}{\delta} - 1\right)! (n - k)! \left(\frac{1}{\delta} + k - 1\right)!} \\
 &= \frac{\frac{\varrho}{\delta} \left(\frac{\varrho}{\delta} + 1\right) \left(\frac{\varrho}{\delta} + 2\right) \dots \left(\frac{\varrho}{\delta} + k - 1\right)}{\frac{1}{\delta} \left(\frac{1}{\delta} + 1\right) \dots \left(\frac{1}{\delta} + k - 1\right)} \cdot \frac{n!}{k! (n - k)!} \\
 &= \binom{n}{k} \frac{\varrho (\varrho + \delta) (\varrho + 2\delta) \dots (\varrho + [k - 1]\delta)}{1 (1 + \delta) (1 + 2\delta) \dots (1 + [k - 1]\delta)} \\
 &\quad \{ r (r - 1) (r - 2) \dots (r - k + 1) \} \\
 &= \frac{\varrho (\varrho + \delta) (\varrho + 2\delta) \dots (\varrho + [k - 1]\delta)}{1 (1 + \delta) (1 + 2\delta) \dots (1 + [k - 1]\delta)} n (n - 1) \dots (n - k + 1). \quad (12)
 \end{aligned}$$

Für $k = 1$ bzw. 2 erhalten wir die Formeln für den Mittelwert und das Quadrat des Schwankungsmasses

$$\{ r \} = n \varrho \quad (13)$$

$$\{ r (r - 1) \} = n (n - 1) \frac{\varrho (\varrho + \delta)}{1 + \delta}$$

$$\begin{aligned}
 \xi^2 &= \{ (r - \{r\})^2 \} = \{ r^2 \} - \{ r \}^2 \\
 &= n \varrho (1 - \varrho) \frac{1 + n \delta}{1 + \delta} \quad (14).
 \end{aligned}$$

§ 7. Diskussion des Schwankungsmasses.

Wenn man bei einer statistischen Untersuchung Serien von verschiedener Länge betrachtet, indem man Woche, Monat oder Jahr als Beobachtungsreihe wählt, variiert man eigentlich die Grösse, die wir in der Wahrscheinlichkeitstheoretischen Betrachtung mit n bezeichnen und Versuchszahl nennen. Welchen Einfluss hat es auf das Schwankungsmass, wenn wir die Serienlänge um ein k -faches vergrössern, d. h. wenn wir zum Beispiel statt Monate Jahre als Serieneinheiten nehmen? Um dies genauer studieren zu können, betrachten wir eine Urne der Chancenvermehrung, wie wir sie im § 3 kennen gelernt haben. (Zu Beginn seien R rote und S schwarze Kugeln in der Urne, $R + S = N$.) Wir machen nach den Vorschriften des § 3 n_1 Züge aus der Urne. Nach diesen n_1 Proben führen wir n_2 solche aus nach derselben Methode usw. Im ganzen machen wir $n_1 + n_2 + \dots + n_k = k \cdot n$ Züge. Bei diesem System der aufeinanderfolgenden Serien unterscheidet sich zwei Fälle:

1. *Nicht zusammenhängende Serien*: Wenn wir nach den ersten n_1 Zügen die Urne wieder in ihren Ausgangszustand versetzen (Inhalt R rote und S schwarze Kugeln), so ist das Resultat der n_2 Proben der zweiten Serie unabhängig vom Ergebnis der ersten Serie. Die einzelnen Serien sind unabhängig voneinander und hängen in keiner Weise zusammen. Es sei die Wahrscheinlichkeit, in der γ -ten Serie r „rote“ Züge zu machen $p_{r, n_\gamma - r}^{(\gamma)}$. Wir erhalten somit

$$\begin{aligned} \{r\} &= \sum_{\gamma=0}^k \sum_{r=0}^{n_\gamma} r p_{r, n_\gamma - r}^{(\gamma)} = n_1 q + n_2 q + \dots \\ &\dots + n_k q = k \cdot n q \end{aligned}$$

$$\begin{aligned}
 s^2 &= \sum_{\gamma=0}^k \sum_{r=0}^{n_\gamma} (r - \{r\})^2 p_{r, n_\gamma - r}^{(\gamma)} \\
 &= \frac{\varrho(1-\varrho)}{1+\delta} [n_1(1+n_1\delta) + \dots + n_k(1+n_k\delta)] \\
 s^2 &= kn\varrho(1-\varrho) \frac{1+n\delta}{1+\delta} \text{ für } n_1 = n_2 = \dots = n_k = n.
 \end{aligned}
 \tag{15}.$$

Der Mittelwert der Ereigniszahl und das Quadrat des Schwankungsmasses sind *proportional* der Anzahl der vereinigten Serien, sobald wir gleich lange Serien vereinigen ($n_1 = n_2 = \dots = n_k$).

2. *Zusammenhängende Serien*: Lassen wir nach den ersten n_1 Proben die Urne unverändert, so befinden sich zu Beginn der zweiten Serie $R + r_1\Delta$ rote und $S + s_1\Delta$ schwarze Kugeln in der Urne. Das Ergebnis der zweiten Serie ist abhängig vom Inhalt der Urne zu Beginn dieser Serie. Die Serien hängen zusammen. Das Ergebnis der letzten Serie ist von den Resultaten aller vorangehenden abhängig. Der Effekt der Proben ist derselbe, wie wenn wir es nur mit einer Serie zu tun haben und aus der Anfangsurne (R rote und S schwarze Kugeln) $kn = n_1 + n_2 + \dots + n_k$ Züge machen. Aus dieser Überlegung erhalten wir nach (13) und (14), indem wir n durch $n \cdot k$ ersetzen

$$\begin{aligned}
 \{r\} &= k \cdot n\varrho \\
 s^2 &= kn\varrho(1-\varrho) \frac{1+kn\delta}{1+\delta}.
 \end{aligned}
 \tag{16}.$$

Bei zusammenhängenden gleichartigen Serien ist das Quadrat des Schwankungsmasses eine *quadratische* Funktion der Anzahl der vereinigten Serien, nicht bloss eine lineare wie im vorangehenden Fall.

Der Vergleich der effektiven mit den theoretischen (Bernoullischen) Schwankungen führte auf die Berechnung des Lexisschen Dispersionskoeffizienten. Im Fall der homogenen Chancenvermehrung wird dieser Koeffizient nach (14) gleich

$$\frac{\sqrt{nq(1-q)} \frac{1+n\delta}{1+\delta}}{\sqrt{nq(1-q)}} = \sqrt{\frac{1+n\delta}{1+\delta}}.$$

Im gewöhnlichen Fall der unabhängigen Ereignisse ($\delta = 0$) ist also der Dispersionskoeffizient gleich 1 (normale Dispersion). Wir machen nun die Annahme, dass sich die Streuung nur unwesentlich von derjenigen der unabhängigen Wahrscheinlichkeiten unterscheidet, d. h. dass der *Dispersionskoeffizient* bei wachsendem n endlich bleibt.

Im allgemeinen hat man es bei allen statistischen Untersuchungen mit einer sehr grossen Versuchszahl zu tun. Dementsprechend lässt man, in sachgemässer mathematischer Abstraktion, n gegen ∞ streben. Zwei Grenzfälle haben sich hierbei als wichtig erwiesen:

I. *Der Laplacesche Grenzfall*, den man mit Herrn Prof. Pólya auch als „*Grenzfall der gewöhnlichen Ereignisse*“ bezeichnen könnte. In diesem Fall ist q fest, n strebt gegen ∞ .

II. *Der Poissonsche Grenzfall*, den man nach Herrn v. Mises treffend als den „*Grenzfall der seltenen Ereignisse*“ bezeichnet. In diesem Fall ist $nq = h$ gesetzt, h fest, n strebt gegen ∞ , $q = \frac{h}{n}$ gegen 0.

Analoge Grenzfälle können auch dann betrachtet werden, wenn die *Wahrscheinlichkeiten nicht unabhängig* sind. Wir wollen dieselben in den nächsten

Paragraphen bei Annahme der Chancenvermehrung untersuchen. Wir können nach den Ausführungen am Anfang dieses Paragraphen nur dann ein vernünftiges Resultat erwarten, wenn wir für beide Grenzfälle die Bedingung hinzufügen, dass $n\delta = d$ gesetzt, d fest bleibt, während $\delta = \frac{d}{n}$ gegen 0 strebt („schwache Chancenvermehrung“). Aus § 3 erhalten wir für die Fortsetzbarkeit der Ziehungsreihen die Bedingung

$$\begin{aligned} \delta &> -\frac{\varrho}{n-1}, & \delta &> -\frac{\sigma}{n-1}, \\ d &> -\varrho \frac{n}{n-1}, & d &> -\sigma \frac{n}{n-1}. \end{aligned}$$

In dem zum Laplaceschen analogen Grenzfall sind die Reihen sowohl für Chancenvermehrung wie für Chancenverminderung (negatives d) fortsetzbar, sofern

$$d > -\varrho, \quad d > -\sigma$$

ist. d muss also immer grösser als $-\frac{1}{2}$ sein. Das Analogon zum Grenzfall der seltenen Ereignisse beschränkt sich auf schwache Chancenvermehrung, da $d \geq 0$ sein muss.

§ 8. Der Grenzfall gewöhnlicher Ereignisse.

Hilfssatz: Sind $a_1, a_2, \dots, a_k, b_1, b_2, \dots, b_l$ positive und $\alpha_1, \alpha_2, \dots, \alpha_k, \beta_1, \dots, \beta_l$ beliebige konstante Grössen, so gilt für ins Unendliche strebendes n , sofern $a_1 + a_2 + \dots + a_k = b_1 + b_2 + \dots + b_l$ und $\alpha_1 + \alpha_2 + \dots + \alpha_k = \beta_1 + \beta_2 + \dots + \beta_l$ ist

Bel

$$\frac{\Gamma(a_1 n + \alpha_1 \sqrt{n}) \Gamma(a_2 n + \alpha_2 \sqrt{n}) \dots \Gamma(a_k n + \alpha_k \sqrt{n})}{\Gamma(b_1 n + \beta_1 \sqrt{n}) \Gamma(b_2 n + \beta_2 \sqrt{n}) \dots \Gamma(b_l n + \beta_l \sqrt{n})}$$

$$\sim \left(\frac{a_1^{a_1} \dots a_k^{a_k}}{b_1^{b_1} \dots b_l^{b_l}} \right)^n \left(\frac{a_1^{a_1} \dots a_k^{a_k}}{b_1^{\beta_1} \dots b_l^{\beta_l}} \right)^{\sqrt{n}} \left(\frac{2\pi}{n} \right)^{\frac{k-l}{2}} \sqrt{\frac{b_1 \dots b_l}{a_1 \dots a_k}} e^{\sum_{\gamma=0}^k \frac{\alpha_\gamma^2}{2a_\gamma} - \sum_{\gamma=0}^l \frac{\beta_\gamma^2}{2b_\gamma}}$$

In der Tat ist $\Gamma(s) \sim \sqrt{2\pi} s^{s-\frac{1}{2}} e^{-s}$, wenn s gegen $+\infty$ strebt.

$$\Gamma(an + a\sqrt{n}) \sim \sqrt{\frac{2\pi}{an}} \left(1 + \frac{a}{a\sqrt{n}}\right)^{an + a\sqrt{n} - \frac{1}{2}} \left(\frac{a}{e}\right)^{an + a\sqrt{n}} n^{an + a\sqrt{n}}$$

$$\sim n^{an + a\sqrt{n}} a^{an + a\sqrt{n}} e^{-an} \sqrt{\frac{2\pi}{an}} e^{\frac{a^2}{2a}},$$

weil

$$\log \left[\left(1 + \frac{a}{a\sqrt{n}}\right)^{an + a\sqrt{n} - \frac{1}{2}} e^{-a\sqrt{n}} \right] \sim \left(an + a\sqrt{n} - \frac{1}{2}\right) \left(\frac{a}{a\sqrt{n}} - \frac{a^2}{2a^2 n} + \dots\right) - a\sqrt{n}$$

$$\sim a\sqrt{n} + \frac{a^2}{a} - \frac{a^2}{2a} - a\sqrt{n} \sim \frac{a^2}{2a},$$

somit ist unser Hilfssatz bewiesen.

Zum Grenzfall gewöhnlicher Ereignisse mit Chancenvermehrung, der demjenigen von Laplace entspricht, gelangen wir, wie schon im § 7 erwähnt wurde, indem wir ϱ und $n\delta = d$ konstant annehmen, d. h. ϱ und d sind fest, während n gegen ∞ konvergiert. Nach Formel (9) ist

$$p_{r,s} = \frac{\binom{\frac{\varrho}{\delta} + r - 1}{r} \binom{\frac{\varrho}{\delta} + s - 1}{s}}{\binom{\frac{1}{\delta} + n - 1}{n}}$$

Ferner lässt sich $\binom{a}{b}$ wie folgt durch die Gammafunktion ausdrücken:

$$\binom{a}{b} = \frac{\Gamma(a+1)}{\Gamma(b+1)\Gamma(a-b+1)} = \frac{a\Gamma(a)}{(a-b)b\Gamma(b)\Gamma(a-b)},$$

$$p_{r,s} = \frac{n \cdot \Gamma(n)}{r \cdot s \cdot \Gamma(r)\Gamma(s)} \cdot \frac{\Gamma\left(\frac{\varrho}{\delta} + r\right) \Gamma\left(\frac{\sigma}{\delta} + s\right) \Gamma\left(\frac{1}{\delta}\right)}{\Gamma\left(\frac{\varrho}{\delta}\right) \Gamma\left(\frac{\sigma}{\delta}\right) \Gamma\left(\frac{1}{\delta} + n\right)}.$$

Wir betrachten nun das Verhalten der Wahrscheinlichkeiten in der Umgebung des Mittelwertes und setzen $r = n\varrho + x\sqrt{n}$ und $s = n\sigma - x\sqrt{n}$, dabei halten wir ϱ , d und x fest und lassen n variieren und ins Unendliche wachsen. Wir erhalten

$$p_{r,s} = \frac{n \Gamma(n) \Gamma\left(n\varrho \frac{1+d}{d} + x\sqrt{n}\right) \Gamma\left(n\sigma \frac{1+d}{d} - x\sqrt{n}\right) \Gamma\left(\frac{n}{d}\right)}{(n\varrho + x\sqrt{n})(n\sigma - x\sqrt{n}) \Gamma(n\varrho + x\sqrt{n}) \Gamma(n\sigma - x\sqrt{n}) \Gamma\left(\frac{n\varrho}{d}\right) \Gamma\left(\frac{n\sigma}{d}\right) \Gamma\left(n \frac{1+d}{d}\right)}$$

Wir betrachten nun die in der folgenden Tabelle aufgeführten Werte der Grössen a_γ , α_γ , b_γ , β_γ .

γ	1	2	3	4	5
a_γ	1	$\frac{\varrho}{d} + \varrho$	$\frac{\sigma}{d} + \sigma$	$\frac{1}{d}$	
α_γ	0	x	$-x$	0	
b_γ	ϱ	σ	$\frac{\varrho}{d}$	$\frac{\sigma}{d}$	$\frac{1}{d} + 1$
β_γ	x	$-x$	0	0	0

Es ist

$$\sum_{\gamma=1}^4 a_{\gamma} = \sum_{\gamma=1}^5 b_{\gamma} = 2 \frac{1+d}{d}, \quad \sum_{\gamma=0}^4 a_{\gamma} = \sum_{\gamma=0}^5 b_{\gamma} = 0$$

$$\frac{a_1 \dots a_4}{b_1^{\sigma_1} \dots b_5^{\sigma_5}} = 1, \quad \frac{a_1^{\alpha_1} \dots a_4^{\alpha_4}}{b_1^{\beta_1} \dots b_5^{\beta_5}} = 1;$$

nach unserm Hilfssatz erhalten wir somit

$$p_{r,s} \sim \frac{1}{n \varrho \sigma} \left(\frac{2\pi}{n} \right)^{-\frac{1}{2}} \sqrt{\frac{\varrho \sigma}{1+d}} e^{\frac{x^2}{2} \left[\frac{d}{\varrho(1+d)} + \frac{d}{\sigma(1+d)} - \frac{1}{\varrho} - \frac{1}{\sigma} \right]}$$

$$p_{r,s} \sim \frac{1}{\sqrt{2\pi n \varrho \sigma (1+d)}} e^{-\frac{x^2}{2\varrho\sigma(1+d)}}$$

Setzen wir nun noch

$$r = n\varrho + l = n\varrho + x\sqrt{n}, \quad x = \frac{l}{\sqrt{n}},$$

so erhalten wir

$$p_{r,s} \sim \frac{1}{\sqrt{2\pi n \varrho \sigma (1+d)}} e^{-\frac{l^2}{2n\varrho\sigma(1+d)}} \quad (17),$$

für $d = 0$ ergibt sich die bekannte Formel

$$q_{r,s} \sim \frac{1}{\sqrt{2\pi n \varrho \sigma}} e^{-\frac{l^2}{2n\varrho\sigma}},$$

was zur Kontrolle dienen mag.

Als Anwendung von (17) betrachten wir das „Schema der nicht zurückgelegten Kugeln“. Aus einer Urne, in der sich $R + S = N$ Kugeln befinden, machen wir n Züge, ohne jeweilen die gezogene

Kugel wieder zurückzulegen. Wie gross ist, unter der Voraussetzung, dass N und n grosse Zahlen sind, die Wahrscheinlichkeit, dass $n\rho + l$ rote und $n\sigma - l$ schwarze Kugeln gezogen werden, wenn

$$\rho = \frac{R}{N}, \quad \sigma = \frac{S}{N}$$

ist? In dem vorliegenden Fall ist

$$A = -1, \quad \delta = -\frac{1}{N}, \quad d = -\frac{n}{N};$$

die gesuchte Wahrscheinlichkeit wird daher nach (17) gleich

$$p_{r,s} = \frac{1}{\sqrt{2\pi n\rho\sigma \frac{N-n}{N}}} e^{-\frac{l^2}{2n\rho\sigma \frac{N-n}{N}}}.$$

Als Bedingung der Fortsetzbarkeit der Versuchsreihe erhalten wir nach dem vorangehenden Paragraphen

$$d = -\frac{n}{N} > -\rho, \quad -\frac{n}{N} > -\sigma.$$

Damit die Reihe fortsetzbar sei, ist auf alle Fälle erforderlich, dass

$$d > -\frac{1}{2}, \quad \frac{n}{N} < \frac{1}{2}, \quad \text{d. h. } N-n > n \text{ ist}^1).$$

¹⁾ Diese Bemerkung scheint Herrn E. Czuber entgangen zu sein (vgl. Bd. 1, S. 183), da er nur fordert $N-n$ muss gross sein. Er hat damit stillschweigend eine Abänderung der Spielregel angenommen. Diese Abänderung ist aber für das Endergebnis unwesentlich, was auch aus der Übereinstimmung der Formeln für die gesuchte Wahrscheinlichkeit hervorgeht.

§ 9. Der Grenzfall seltener Ereignisse.

Das Analogon zum Grenzfall seltener Ereignisse, d. h. zum Poissonschen Grenzfall, erhalten wir bei Annahme der Chancenvermehrung, wie wir im § 7 gesehen haben, indem wir $n\varrho = h$ und $n\delta = d$ setzen, h und d fest halten und n gegen ∞ konvergieren lassen. Wir erhielten im § 8 für die Einzelwahrscheinlichkeit

$$p_{r,s} = \frac{n \cdot \Gamma(n) \Gamma\left(\frac{\varrho}{\delta} + r\right) \Gamma\left(\frac{\sigma}{\delta} + s\right) \Gamma\left(\frac{1}{\delta}\right)}{r \cdot s \cdot \Gamma(r) \Gamma(s) \Gamma\left(\frac{\varrho}{\delta}\right) \Gamma\left(\frac{\sigma}{\delta}\right) \Gamma\left(\frac{1}{\delta} + n\right)};$$

da aber

$$n\varrho = h, \quad n\delta = d, \quad \varrho = \frac{h}{n}, \quad \sigma = 1 - \frac{h}{n}, \quad \delta = \frac{d}{n}$$

ist, wird

$$p_{r,s} = \frac{1}{r} \frac{\Gamma\left(\frac{h}{d} + r\right)}{\Gamma\left(\frac{h}{d}\right) \Gamma(r)} \cdot \frac{n}{n-r} \cdot \frac{\Gamma(n)}{\Gamma(n-r)} \frac{\Gamma\left(\frac{n}{d}\right)}{\Gamma\left(\frac{n-h}{d}\right)} \frac{\Gamma\left(\frac{1+d}{d} n - \frac{h}{d} - r\right)}{\Gamma\left(\frac{1+d}{d} n\right)}.$$

Im Falle gewöhnlicher Ereignisse (vgl. § 8) waren bei ins Unendliche strebendem n alle Argumente der Gammafunktionen, die in $p_{r,s}$ auftraten, bei der gemachten Annahme über r , grosse Zahlen, die auf die Form $an + a\sqrt{n}$ gebracht werden konnten. Dadurch fand der Hilfssatz Anwendung.

Im Falle der seltenen Ereignisse sind r und $\frac{h}{d}$ feste Zahlen (wir betrachten nur Werte von r in der Umgebung des Mittelwertes). $\Gamma\left(\frac{h}{d} + r\right)$, $\Gamma\left(\frac{h}{d}\right)$ und $\Gamma(r)$ sind auch fest. Die übrigen Gammafunktionen hängen

von n in der Art $\Gamma(an + \beta)$ ab, wobei a, β fest sind. Dass $\beta \neq 0$ ist, macht den eigentlichen Unterschied zu dem in § 8 betrachteten Fall aus. Es genügt jetzt die folgende einfache Formel:

$$\frac{\Gamma(an + a)}{\Gamma(an + \beta)} \sim \frac{(an + a)^{an+a} e^{-a}}{(an + \beta)^{an+\beta} e^{-\beta}} \sim (an)^{a-\beta}.$$

In Anwendung hiervon erhält man

$$p_{r, n-r} \sim \frac{1}{r} \frac{\Gamma\left(\frac{h}{d} + r\right)}{\Gamma\left(\frac{h}{d}\right) \Gamma(r)} \cdot n^r \left(\frac{1}{d}n\right)^{\frac{h}{d}} \left(\frac{1+d}{d}n\right)^{-\frac{h}{d}-r}$$

$$= \frac{\frac{h}{d} \cdot \left(\frac{h}{d} + 1\right) \dots \left(\frac{h}{d} + r - 1\right) (1+d)^{-\frac{h}{d}-r} d^r}{1 \cdot 2 \cdot 3 \dots r}$$

$$\lim_{n \rightarrow \infty} p_{r, n-r} = \frac{h(h+d)(h+2d) \dots (h+[r-1]d)}{r! (1+d)^{\frac{h}{d}+r}} \quad (18).$$

Kapitel II.

Seltene Ereignisse.

§ 10. Die Wahrscheinlichkeiten.

In diesem Kapitel befasse ich mich nur mit dem Grenzfall seltener Ereignisse mit „schwacher“ Chancenvermehrung, auf den ich im § 9 hingewiesen habe. Wir setzen wie dort

$$n\varrho = h \qquad n\delta = d \qquad (19)$$

und halten h und d fest, während wir n gegen ∞ konvergieren lassen.

Wir bezeichnen ferner mit P_r den Grenzwert $\lim_{n \rightarrow \infty} p_{r, n-r}$ und mit Q_r den Grenzwert $\lim_{n \rightarrow \infty} q_{r, n-r}$, wobei bei $p_{r, n-r}$ dieselbe Bedeutung hat wie im vorhergehenden Kapitel, $q_{r, n-r}$ aber die Wahrscheinlichkeit im Falle unabhängiger Wahrscheinlichkeiten mit endlicher Versuchszahl bezeichnet.

Es ist somit

$$q_{r, n-r} = \lim_{d \rightarrow 0} p_{r, n-r}.$$

Nach den Entwicklungen des § 9 erhalten wir aus (18)

$$P_r = \frac{h(h+d)(h+2d) \dots (h+[r-1]d)}{r!(1+d)^{\frac{h}{d}+r}}; \quad (20)$$

bekanntlich ist

$$Q_r = \frac{e^{-h} h^r}{r!}, \quad (21)$$

und somit

$$Q_r = \lim_{d \rightarrow 0} P_r;$$

für $r = 0$ erhalten wir

$$P_0 = (1+d)^{-\frac{h}{d}}.$$

(Vgl. Formel (14) „Über die Statistik verketteter Vorgänge“, Pólya 2, wo diese direkt hergeleitet wurde, mit Hilfe einer andern Methode). Mit Hilfe der Gammafunktionen geht (20) über in

$$\begin{aligned}
 P_r &= (1+d)^{-\frac{h}{d}} \left(\frac{d}{1+d}\right)^r \frac{\left(\frac{h}{d}+r-1\right)!}{r! \left(\frac{h}{d}-1\right)!} \\
 &= (1+d)^{-\frac{h}{d}} \left(\frac{d}{1+d}\right)^r \cdot \frac{\Gamma\left(\frac{h}{d}+r\right)}{\Gamma(r+1) \Gamma\left(\frac{h}{d}\right)}.
 \end{aligned}$$

Nun ist aber bei grossem n , angenähert schon für $n > 10$, genügend genau

$$n \cdot \Gamma(n) = n! = n^{n+\frac{1}{2}} e^{-n} \left(\sqrt{2\pi} + \frac{1}{12n} + \dots \right)$$

$$\begin{aligned}
 P_r &= \frac{(1+d)^{-\frac{h}{d}} \cdot \left(\frac{d}{1+d}\right)^r \cdot \left(\frac{h}{d}+r\right)^{\frac{h}{d}+r+\frac{1}{2}}}{\Gamma\left(\frac{h}{d}\right) \cdot \left(\frac{h}{d}+r\right)} \\
 &\quad \cdot \frac{e^{-\frac{h}{d}-r} \cdot \left(\sqrt{2\pi} + \frac{1}{12\left(\frac{h}{d}+r\right)} + \dots \right)}{r^{r+\frac{1}{2}} e^{-r} \cdot \left(\sqrt{2\pi} + \frac{1}{12r} + \dots \right)} \\
 P_r &\sim \frac{(1+d)^{-\frac{h}{d}} \left(\frac{d}{1+d}\right)^r \left(\frac{h}{d}+r\right)^{\frac{h}{d}+r-\frac{1}{2}}}{\Gamma\left(\frac{h}{d}\right) e^{\frac{h}{d}} r^{r+\frac{1}{2}}} \quad (22).
 \end{aligned}$$

Diese Formel habe ich bei grösserem statistischem Material stets als Kontrollformel benutzt. P_r kann aber auch durch Rekursion aus P_{r-1} erhalten werden. Es ist

$$P_r = \frac{h+(r-1)d}{r(1+d)} P_{r-1}. \quad (23)$$

Es sei W_r die Wahrscheinlichkeit, dass ein Ereignis genau r -mal eintritt. Unter der *erzeugenden Funktion* der Wahrscheinlichkeitsreihe

$$W_0, W_1, W_2 \dots$$

versteht man die Potenzreihe

$$W_0 + W_1 z + W_2 z^2 + \dots = G(z).$$

Insbesondere erhalten wir für die erzeugenden Funktionen $E(z)$ und $F(z)$ der Wahrscheinlichkeitsreihen

$$P_0, P_1, P_2 \dots, \text{ und } Q_0, Q_1, Q_2 \dots$$

$$\begin{aligned} E(z) &= \sum_{r=0}^{\infty} z^r P_r = \sum_{r=0}^{\infty} z^r (1+d)^{-\frac{h}{d}} \frac{h(h+d) \dots (h+[r-1]d)}{(1+d)^r r!} \\ &= (1+d)^{-\frac{h}{d}} \sum_{r=0}^{\infty} \left(-\frac{zd}{1+d} \right)^r \binom{-\frac{h}{d}}{r} = (1+d)^{-\frac{h}{d}} \left(1 - \frac{zd}{1+d} \right)^{-\frac{h}{d}} \end{aligned}$$

$$E(z) = (1+d[1-z])^{-\frac{h}{d}} \quad (24).$$

$$F(z) = \sum_{r=0}^{\infty} z^r Q_r = \sum_{r=0}^{\infty} \frac{z^r e^{-h} h^r}{r!} = e^{-h} \sum_{r=0}^{\infty} \frac{(zh)^r}{r!} = e^{-h(1-z)} \quad (25).$$

Dieselbe Formel ergibt sich auch als Grenzfall für $\lim d = 0$ aus (24).

Auf eine sehr elegante Art erhalten wir aus den erzeugenden Funktionen die mathematischen Erwartungen von $r(r-1)(r-2) \dots (r-k+1)$. (In der Folge werden dieselben häufig als Momente bezeichnet.) Es ist nämlich für Chancenvermehrung

$$\begin{aligned} \{r(r-1)(r-2)\dots(r-k+1)\} &= \sum_{r=0}^{\infty} r(r-1)(r-2)\dots(r-k+1) P_r \\ &= \left[\sum_{r=0}^{\infty} \frac{d^k}{dz^k} z^r P_r \right]_{z=1} = \left[\frac{d^k}{dz^k} \sum_{r=0}^{\infty} z^r P_r \right]_{z=1} = \left[\frac{d^k}{dz^k} (1+d[1-z])^{-\frac{h}{d}} \right]_{z=1} \\ &= \left[h(h+d)(h+2d)\dots(h+[k-1]d) (1+d[1-z])^{-\frac{h}{d}-k} \right]_{z=1} \end{aligned}$$

$$\{r(r-1)(r-2)\dots(r-k+1)\} = h(h+d)\dots(h+[k-1]d) \quad (26).$$

Durch Grenzübergang hätten wir dies direkt aus (12) erhalten. Speziell ergibt sich der Mittelwert $\{r\}$ und das Quadrat des Schwankungsmasses s^2

$$\{r\} = h, \quad s^2 = \{(r-h)^2\} = \{r^2\} - \{r\}^2 = h(1+d) \quad (27).$$

Das erste Studium, das ich an meinem *statistischen Material* über Chancenvermehrung unternahm, bezog sich auf die Einzelwahrscheinlichkeiten. Die Frage ist folgende: Stimmen die theoretischen erwartungsmässigen Anzahlen innerhalb der erwartungsmässigen Grenzen mit den beobachteten überein?

In Zahlentafel 1 betrachten wir als erstes Beispiel die Todesfälle bei *Dampffässerexplosionen* in Preussen in den Jahren 1890—1909. Den Hinweis auf diesen Fall verdanke ich Herrn Prof. v. Bortkiewicz (durch die freundliche Vermittlung von Herrn Prof. Pólya). Die 20 Beobachtungsjahre wurden zunächst in Kategorien mit gleicher Anzahl Todesfälle pro Jahr eingeteilt und in den Spalten I und II registriert.

Insgesamt wurden 101 Todesfälle bei Dampffässerexplosionen verzeichnet (Summe der Produkte der Stellen der Spalte I mit den entsprechenden Stellen der Spalte II), also im Durchschnitt pro Jahr $5.05 = h$.

Zuerst wurde nun geprüft, ob die Theorie der Unabhängigkeit (Poissonsches Schema) befriedigende Resultate liefere. Zu diesem Zweck habe ich die der Spalte II entsprechenden Werte in Spalte III nach (21) $20Q_r = 20 \frac{e^{-h} h^r}{r!}$ berechnet.

Aus der systematischen Abweichung der Spalte III von der Spalte II (die berechneten Werte sind für kleines und grosses r zu klein, für mittleres r zu gross) überzeugen wir uns von der Nichtübereinstimmung der Theorie der unabhängigen Ereignisse mit der Wirklichkeit. In den Summenzahlen der Spalten II und III, die wir in V und VI finden, kommt dieser Gang der Abweichung noch mehr zum Ausdruck, indem die Abweichung, die in VIII verzeichnet ist, zuerst negativ, dann positiv ist. Noch deutlicher springt die Nichtübereinstimmung der Theorie, die zu den Wahrscheinlichkeiten Q führt, in die Augen, wenn man bedenkt, dass das Quadrat der mittleren Abweichung nach (27) (für $d = 0$) $h = 5.05$ betragen sollte. Wir multiplizieren nun $(r-5.05)^2$ mit der Anzahl derjenigen Jahre, in denen genau r Todesfälle aufgetreten sind (Spalte II), summieren diese Produkte und dividieren die erhaltene Summe mit der um 1 verminderten Anzahl der Beobachtungsjahre, d. h. mit 19 (vgl. Formel 36), und erhalten 24.576, also bedeutend mehr als 5.05.

Diese Abweichung liegt, wie man sich leicht überzeugt, ausserhalb des erwartungsmässigen Schwankungsbereiches. Nun wollen wir zusehen, ob die Theorie der Chancenvermehrung bessere Resultate liefere. Nach (27) ist $s^2 = h(1+d)$, also $= 24.576$, und daraus ist $d = 3.867$; nun berechnen wir die Grössen $20P_r$ nach (20), die in Spalte IV aufgezeichnet sind. Die Summation der Spalte IV wurde in Spalte VII durch-

geführt, während in der Spalte IX die Abweichung von V und VII, d. h. die Abweichung der Theorie der Chancenvermehrung von der Beobachtung registriert wurde. Die relativ gute Übereinstimmung der Spalten II und IV bzw. V und VII, die in dem günstigen Verlauf der Spalte IX (absolut klein, bald positiv, bald negativ) ihren Ausdruck findet, bekräftigt uns in der Annahme, dass die Todesfälle bei Dampffässerexplosionen in ähnlichem Zusammenhang stehen wie die Züge aus der Urne der Chancenvermehrung.

Die praktische Berechnung der Spalten III und IV wurde durchweg mit der Rekursionsformel durchgeführt:

$$P_r = \frac{h + (r-1)d}{r(1+d)} P_{r-1}, \quad Q_r = \frac{h}{r} Q_{r-1}.$$

P_0 und Q_0 wurden bei diesem, wie auch bei allen ähnlichen Beispielen, direkt bestimmt (auf 4 Dezimalstellen) und daraus die Spalten III und IV mit der Rechenmaschine (Rekursion) konstruiert. Ausserdem wurden bei den Tabellen 1, 3, 4, 5 und 6 zur Kontrolle jeweils in Spalte IV einige Zwischenwerte mit Logarithmentafel und einer Tafel der Gammafunktion direkt bestimmt.

Wie schon Herr Prof. v. *Bortkiewicz* (durch die freundliche Vermittlung von Herrn Prof. *Pólya*) bemerkt hat, liefert uns die Annahme der Unabhängigkeit bei den Dampffässerexplosionen (die *Explosionen* als Ereignisse registriert) befriedigende Resultate. Die Statistik dieser Explosionen finden wir in Zahlentafel 2, die ganz entsprechend Zahlentafel 1 angelegt ist, nur wurden die Kolonnen betreffend Chancenvermehrung weggelassen. Es ist

$$h = 3.3 \pm 0.41 \quad \text{und} \quad s^2 = 3 \pm 1.12$$

(Über die Berechnung der erwarteten Fehler s. L. v. Bortkiewicz 1) Wie zu erwarten ist, sind die *Todesfälle* bei Dampfässerexplosionen in gewisser Hinsicht solidarisch, die *Explosionen* selbst hingegen voneinander unabhängig.

Ein weiteres Beispiel der Explosionstodesfälle finden wir in Zahlentafel 3, die gleich angelegt ist wie Tafel 1. Diese Zahlentafel 3 der Todesfälle bei den *Dampfkesselexplosionen in Preussen 1883—1907* gibt uns zusammen mit dem ersten Beispiel die Überzeugung, dass die bessere Übereinstimmung der Theorie der Chancenvermehrung mit der Wirklichkeit nicht bloss davon herrührt, dass in P_r ein Parameter mehr vorhanden ist als in Q_r , sondern dass das zugrunde gelegte Urnenschema die Explosionstodesfälle besser charakterisiert.

Ein Beispiel ganz anderer Art bildet das in Zahlentafel 4 aufgeführte und von mir bereits veröffentlichte *Pockenbeispiel* (vgl. F. Eggenberger, S. 5 und ff.). Die Berechnungen bieten gegenüber Zahlentafel 1 nichts Neues. Obschon wir hier ausser der Ansteckung einen anderen störenden Faktor (die Inhomogenität) haben, gibt uns das Schema der Chancenvermehrung ein befriedigendes Resultat.

In Zahlentafel 5, die analog aufgebaut ist wie Zahlentafel 1, finden wir die *Todesfälle an Scharlach in der Schweiz in den Jahren 1877—1900* (monatsweise Beobachtungen). Im ersten Moment wird man, beeinflusst durch die günstigen Resultate der Pockenuntersuchungen, vermuten, dass die theoretischen Wahrscheinlichkeiten P_r in befriedigender Weise mit den Beobachtungen übereinstimmen werden. Der systematische Verlauf von Spalte IX, die die Abweichung der Theorie der Chancenvermehrung von der Wirklichkeit angibt, lehrt uns aber das Gegenteil. Wohl stimmt die

Theorie der Chancenvermehrung besser mit der Wirklichkeit überein als die Theorie der unabhängigen Ereignisse. Das unbefriedigende Resultat rührt, wie wir sofort sehen werden, von der Inhomogenität des statistischen Materials her. Wir vergleichen nun die Anzahl der Todesfälle pro Monat im gleichen Jahr miteinander und legen in der folgenden Tabelle fest, wie oft auf einen bestimmten Monat das jeweilige Jahresmaximum und Jahresminimum der Jahre 1877—1900 entfallen ist.

Monat	Anzahl der Jahresmaxima in den Jahren 1877—1900	Anzahl der Jahresminima in den Jahren 1877—1900
Januar .	3	1
Februar .	3	—
März . .	5	—
April . .	9	—
Mai . . .	2	1
Juni . . .	—	1
Juli . . .	1	—
August .	—	7
September	—	2
Oktober .	—	5
November	1	6
Dezember	—	1

Wir stellen somit fest: die Scharlachtodesfälle in der ersten Jahreshälfte (Januar bis Juni) sind bedeutend häufiger als diejenigen der zweiten Jahreshälfte. Die starke Inhomogenität beherrscht das Material.

In Zahlentafel 6 sind die Resultate der Statistik der *Todesfälle an Scharlach in der Schweiz in den Monaten des zweiten Halbjahres (Juli bis Dezember) der Jahre 1879—1900* verwertet worden. Diese Tabelle

wurde analog berechnet wie Zahlentafel 1. Die vollständig befriedigende Übereinstimmung der Theorie der Chancenvermehrung mit der Beobachtung, die in Spalte IX zum Ausdruck kommt, zeigt uns, dass, sobald wir nur Monate der zweiten Jahreshälfte betrachten, die Inhomogenität kaum mehr in Betracht kommt gegenüber der Chancenvermehrung und bestärkt uns in der Annahme, dass die Todesfälle, verursacht durch die betrachteten Infektionskrankheiten, in analoger Weise solidarisch sind wie die aus einer Urne der Chancenvermehrung gezogenen roten Kugeln.

§ 11. Verlauf der Wahrscheinlichkeiten.

Die Reihe der Wahrscheinlichkeiten

$$P_0, P_1, P_2, P_3, \dots \quad (28)$$

ist entweder monoton abnehmend (für $d \geq h$) oder hat glockenförmigen Verlauf (für $d < h$). Dieses Resultat erhalten wir ohne weiteres aus § 5, da (28) ein Spezialfall von (10) ist, können aber auch direkt aus (23) dazu kommen.

Ist P_μ das Maximum der Reihe der Wahrscheinlichkeiten (28), dann bestimmt sich μ aus der Gleichung

$$\mu = [h - d] \quad (29).$$

Diese Relation erhalten wir durch Grenzübergang aus (11). Ist $h - d > 0$ eine ganze Zahl, so ist $P_{\mu-1} = P_\mu$.

Bezeichnen wir den Rest der Reihe (28) mit

$$R_i = P_{i+1} + P_{i+2} + \dots,$$

so gilt für diesen Rest die Ungleichung

$$\left. \begin{aligned} P_i \frac{h+id}{i+1+d-h} \leq R_i \leq P_i \frac{h+id}{i+1} \text{ für } h \leq d \\ P_i \frac{h+id}{i+1} \leq R_i \leq P_i \frac{h+id}{i+1+d-h} \text{ für } h \geq d \end{aligned} \right\} (30).$$

Nach der Rekursionsformel (23) können wir nämlich das Restglied darstellen durch

$$\begin{aligned} R_i &= P_{i+1} + P_{i+2} + P_{i+3} + \dots \\ &= P_{i+1} \left(1 + \frac{1}{i+2} \frac{h+(i+1)d}{1+d} + \frac{1}{i+2} \frac{h+(i+1)d}{1+d} \cdot \frac{1}{i+3} \right. \\ &\quad \left. \cdot \frac{h+(i+2)d}{1+d} + \dots \right). \end{aligned}$$

Nun vergleichen wir den Rest R_i mit der Summe der geometrischen Reihen A_i und B_i , die beide P_{i+1} zum Anfangsglied haben und deren eine als Quotienten $\frac{d}{1+d}$, die andere $\frac{h+id}{(i+1)(1+d)}$ aufweist.

Es ist

$$A_i = P_{i+1} \left[1 + \frac{d}{1+d} + \left(\frac{d}{1+d} \right)^2 + \dots \right] = P_{i+1} (1+d) = P_i \frac{h+id}{1+i}$$

$$\begin{aligned} B_i &= P_{i+1} \left[1 + \frac{h+id}{(i+1)(1+d)} + \left(\frac{h+id}{(i+1)(1+d)} \right)^2 + \dots \right] = P_{i+1} \frac{(i+1)(1+d)}{i+1-h+d} \\ &= P_i \frac{h+id}{i+1-h+d} \end{aligned}$$

Es gilt die Ungleichung:

$$\frac{h+(i+l)d}{(1+d)(l+i+1)} \begin{matrix} \leq \\ > \end{matrix} \frac{d+(i+l)d}{(1+d)(l+i+1)} = \frac{d}{1+d} \text{ für } h \begin{matrix} \leq \\ > \end{matrix} d.$$

Ferner nimmt der Quotient

$$\frac{h + (i+l)d}{(1+d)(l+i+1)} = \frac{1}{1+d} \left[d + \frac{h-d}{1+i+l} \right]$$

mit wachsendem l zu oder ab oder bleibt konstant, je nachdem $h < d$, $h > d$ oder $h = d$ ist. Also gilt die Ungleichung

$$\frac{h + id}{(i+1)(1+d)} \begin{matrix} \leq \\ \geq \end{matrix} \frac{h + (i+l)d}{(1+d)(l+i+1)} \begin{matrix} \leq \\ \geq \end{matrix} \frac{d}{1+d} \quad \text{für } h \begin{matrix} \leq \\ \geq \end{matrix} d.$$

Somit folgt ohne weiteres

$$B_i \begin{matrix} \leq \\ \geq \end{matrix} R_i \begin{matrix} \leq \\ \geq \end{matrix} A_i \quad \text{für } h \begin{matrix} \leq \\ \geq \end{matrix} d,$$

und damit ist die Ungleichung (30) bewiesen.

Nun zu unseren Anwendungen. In Spalte IV der Tabellen 1, 3, 4, 5, 6 finden wir die Folgen der Wahrscheinlichkeiten P_0, P_1, \dots (mit einem konstanten Faktor multipliziert). Bei den in Zahlentafeln 1, 3, 6 betrachteten Beispielen ist $h > d$, es tritt in Spalte IV ein Maximum auf (Fettdruck). Im Pocken- und Scharlachbeispiel der Zahlentafeln 4 und 5 ist $h < d$; Spalte IV ist monoton abnehmend. Nun wollen wir beim *Pockenbeispiel* (Zahlentafel 4) die Resultate dieses Paragraphen prüfen. Ohne zuerst ausführliche Berechnungen zu machen, können wir aus den Beobachtungen (Spalte II) schliessen, dass die Theorie der Chancenvermehrung die Beobachtungen besser wiedergibt als die Theorie, die die Unabhängigkeit der Einzelereignisse voraussetzt. In der Tat liegt es nahe, aus dem Verlauf der Spalte der Beobachtungen (II) anzunehmen, dass die Theorie (Chancenvermehrung), die ein monoton abnehmen der Reihe P_0, P_1, \dots ergibt, der Wirklichkeit näher kommt als diejenige (Unabhängigkeit), die dem Index des Maximalgliedes den Wert h gibt.

Die Restformeln (30) liefern uns bei der numerischen Auswertung des statistischen Materials eine gute Kontrolle für die Berechnungen. In Zahlentafel 4 des Pockenbeispiels wurde mit dem 62. Glied abgebrochen. Aus Spalte VII ersehen wir, dass $288 (P_0 + P_1 + \dots + P_{62}) = 287.4$ ist. Haben wir Spalte IV richtig berechnet, so ergibt sich für $288 R_{62} = 0.6$, nach unsern Ungleichungen (30) ist aber

$$0.6 \leq 288 R_{62} \leq 0.7.$$

Nach Spalte VII ist ferner $288 R_{30} = 8.3$, nach (30) ist

$$7.7 \leq 288 R_{30} \leq 9.8.$$

Auf diese Weise können wir die Berechnung von (IV) durch Stichproben kontrollieren.

§ 12. Berechnung von $\{|r-h|\}$. Die erwartungsmässigen Fehler.

Es sei $r-h$ der Fehler des Einzelergebnisses oder, kürzer, der Fehler von r . Die mathematische Erwartung dieses Fehlers ist gleich 0.

$$\{r-h\} = \{r\} - h = h - h = 0.$$

Wir fragen nun nach der mathematischen Erwartung des absoluten Betrages dieses Fehlers. Es sei

$$[h] = \gamma,$$

dann wird

$$\sum_{r=0}^{\gamma} (h-r) P_r = \sum_{r=\gamma+1}^{\infty} (r-h) P_r \quad \text{weil} \quad \sum_{r=0}^{\infty} (r-h) P_r = 0 \quad \text{ist.}$$

$$E = \{|\mathbf{r}-h|\} = \sum_{r=0}^{\gamma} (h-r) P_r + \sum_{r=\gamma+1}^{\infty} (r-h) P_r,$$

$$E = 2 \sum_{r=0}^{\gamma} (h-r) P_r,$$

$$\frac{d}{1+d} E = \frac{2d}{1+d} \sum_{r=0}^{\gamma} (h-r) P_r.$$

Durch Subtraktion folgt

$$\begin{aligned} E \frac{1}{2(1+d)} &= \sum_{r=0}^{\gamma} (h-r) P_r - \frac{d}{1+d} \sum_{r=0}^{\gamma} (h-r) P_r \\ &= \sum_{r=0}^{\gamma} (h-r) P_r - \frac{d}{1+d} \sum_{r=1}^{\gamma+1} (h-[r-1]) P_{r-1} \\ &= \sum_{r=1}^{\gamma+1} \left[(h-r) P_r + \frac{(r-1)d}{1+d} P_{r-1} - \frac{hd}{1+d} P_{r-1} \right] \\ &\quad + h P_0 - (h-[\gamma+1]) P_{\gamma+1}. \end{aligned}$$

Nun ist aber

$$P_r = \frac{1}{r} \frac{h+(r-1)d}{1+d} P_{r-1},$$

$$r P_r - \frac{(r-1)d}{1+d} P_{r-1} = \frac{h}{1+d} P_{r-1}$$

und somit

$$\begin{aligned} E \frac{1}{2(1+d)} &= \sum_{r=1}^{\gamma+1} \left[h P_r - \frac{h}{1+d} P_{r-1} - \frac{hd}{1+d} P_{r-1} \right] \\ &\quad + h P_0 - h P_{\gamma+1} + (\gamma+1) P_{\gamma+1} \\ &= (\gamma+1) P_{\gamma+1} + h(P_0 - P_{\gamma+1}) + \sum_{r=1}^{\gamma+1} h(P_r - P_{r-1}), \\ E = \{|\mathbf{r}-h|\} &= 2(1+d)(\gamma+1) P_{\gamma+1}, \end{aligned}$$

und weil $\gamma = [h]$ ist,

$$\{\mathcal{P} - h\} = 2(h + [h]d)P_{[h]}. \quad (31)$$

Wie beim gewöhnlichen Urnenschema von Bernoulli die effektiven Resultate nicht genau mit den erwartungsmässigen übereinstimmen, werden wir auch in der Theorie der Chancenvermehrung gewisse Abweichungen haben. Die Resultate, die wir bei irgendeiner auf Grund der Theorie der Chancenvermehrung gemachten Untersuchung erhalten, können wir nur richtig beurteilen, wenn wir die Schwankungen kennen, die die Theorie zulässt. In der folgenden Tabelle finden wir eine Zusammenstellung der gebräuchlichsten der Formeln für die mathematische Erwartung, das Quadrat des Schwankungsmasses und den mittleren Fehler. m bedeutet dabei die Anzahl der betrachteten Serien.

	Mathe- matische Erwartung	Quadrat des Schwankungsmasses	Mittlerer Fehler
Mittelwert	h (32)	$\frac{h(1+d)}{m}$ (33)	$\sqrt{\frac{h(1+d)}{m}}$ (33)
Quadrat des Schwankungs- masses	$h(1+d)$ (34)	$\frac{h(1+d)}{m} [(2h+6d)(1+d)+1]$ (35)	$\sqrt{\frac{h(1+d)}{m} [(2h+6d)(1+d)+1]}$ (35)

Die Ermittlung der in obiger Tabelle aufgeführten Werte erfolgt nach bekannten Methoden (man vergleiche etwa Czuber **1** und v. Bortkiewicz **1**). Wir wollen hier beispielsweise Formel (35) herleiten. Es ist

$$\begin{aligned}
 F^2 &= \left\{ \left(\frac{(\mathbf{r}_1 - h)^2 + (\mathbf{r}_2 - h)^2 + \dots + (\mathbf{r}_m - h)^2}{m} - h(1+d) \right)^2 \right\} \\
 &= \left\{ \left(\frac{[(\mathbf{r}_1 - h)^2 - h(1+d)] + [(\mathbf{r}_2 - h)^2 - h(1+d)] + \dots + [(\mathbf{r}_m - h)^2 - h(1+d)]}{m} \right)^2 \right\} \\
 &= \frac{1}{m^2} \sum_{i=1}^m \{([\mathbf{r}_i - h]^2 - h[1+d])^2\} \\
 &\quad + \frac{2}{m^2} \sum_{i=1}^m \sum_{\substack{k=1 \\ k \neq i}}^m \{(\mathbf{r}_i - h)^2 - h(1+d)\} \{(\mathbf{r}_k - h)^2 - h(1+d)\} \\
 &= \frac{1}{m^2} \sum_{i=1}^m \{(\mathbf{r}_i - h)^4\} + \frac{h^2}{m} (1+d)^2 - \frac{2h}{m^2} (1+d) \sum_{i=1}^m \{(\mathbf{r}_i - h)^2\} \\
 &= \frac{1}{m} \{(\mathbf{r} - h)^4\} - \frac{h^2}{m} (1+d)^2.
 \end{aligned}$$

Nun ermitteln wir die mathematische Erwartung von $(\mathbf{r} - h)^4$. Das Wesentliche der Ermittlung der mathematischen Erwartung ähnlicher Ausdrücke besteht in folgendem: Man ordnet die rationale ganze Funktion, deren mathematische Erwartung auszurechnen ist, nicht wie üblich nach Potenzen von r , d. h. nach $1, r, r^2, r^3, \dots$, sondern nach den Funktionen

$$1, r, r(r-1), r(r-1)(r-2), \dots,$$

wobei $r = 0, 1, 2, 3 \dots$ ist.

$$(\mathbf{r} - h)^4 = r(r-1)(r-2)(r-3) + ar(r-1)(r-2) + br(r-1) + cr + d$$

Die Koeffizienten a, b, c, d bestimmen wir am besten durch Einsetzen der Werte $r = 0, 1, 2, 3$ in obige Gleichung und erhalten

$$a = \frac{(2-h)^4 - 3(2-h)^4 + 3(1-h)^4 - h^4}{6} = 6 - 4h$$

$$b = \frac{(3-h)^4 - 2(1-h)^4 + h^4}{2} = 7 - 12h + 6h^2$$

$$c = (1-h)^4 - h^4 = 1 - 4h + 6h^2 - 4h^3$$

$$d = h^4.$$

Wir erhalten somit nach (26)

$$F^2 = \frac{1}{m} \left[h(h+d)(h+2d)(h+3d) + (6-4h)h(h+d)(h+2d) \right. \\ \left. + (7-12h+6h^2)h(h+d) + (1-4h+6h^2-4h^3)h+h^4 - h^2(1+d)^2 \right]$$

$$F = \sqrt{\frac{h(1+d)}{m} [(2h+6d)(1+d) + 1]}. \quad (35)$$

In den Formeln (32) bis (35) wird man überall an Stelle von h und d aus dem statistischen Material berechnete, d. h. angenäherte Werte h' und d' einsetzen müssen, sobald es sich um numerische Kontrollen handelt, da man weder d noch h absolut genau kennt. Bei der Herleitung dieser Formeln setzt man dennoch h als bekannt voraus. Mit der gleichen Methode erhält man die entsprechenden Ausdrücke, unter der Voraussetzung, dass man h nicht genau kennt, sondern nur den angenäherten Mittelwert h' berechnet hat. Wie man sich leicht überzeugt, sind (bei einigermaßen grosser Serienzahl) die Werte, die man unter Voraussetzung eines bekannten h ermittelt, nahezu gleich denjenigen, die auf h' aufgebaut sind.

In der Praxis wird man, da es sich hier vor allem um Abschätzungen handelt, voraussetzen, h sei bekannt.

Setzt man z. B. bei der Ermittlung von (34) voraus, dass h unbekannt sei, so erhält man durch einfache Rechnungen (vgl. etwa bei v. Bortkiewicz **1** den Fall ohne Chancenvermehrung) für die mathematische Erwartung des Quadrates des Schwankungsmasses

$$\frac{m-1}{m} h(1+d). \quad (36)$$

Von dieser Formel, aus der wir in den praktischen Beispielen d bestimmen, werden wir häufig Gebrauch machen. Zur Bestimmung des Quadrates des Schwankungsmasses ist auf die bekannte Art die Summe der quadratischen Abweichungen durch die um 1 verminderte Serienzahle zu dividieren.

Im Spezialfall $d = 0$ geht (35) über in

$$F = \sqrt{\frac{h}{m} (2h + 1)}.$$

Dies gibt uns das von L. v. Bortkiewicz für die seltenen unabhängigen Ereignisse ermittelte Fehlermass (vgl. **1** § 4, Formel (5)), während (33) für $d = 0$ in Formel (4) übergeht.

Für d , im folgenden oft *Ansteckung* genannt, erhalten wir

$$d = \frac{(\text{Schwankungsmass})^2}{h} - 1 = \frac{s^2}{h} - 1.$$

Der Fehler von d wird gleich

$$\begin{aligned} & \frac{h \cdot (\text{Fehler von } s^2) + h(1+d) (\text{Fehler von } h)}{h^2} \\ &= \sqrt{\frac{(1+d)}{h \cdot m} [(2h+6d)(1+d)+1]} + \sqrt{\frac{(1+d)^3}{h \cdot m}} \quad (37). \end{aligned}$$

Für sämtliche Beispiele habe ich die sogenannte durchschnittliche Abweichung $\{|r-h|\}$ aus der Beobachtung direkt berechnet. Es ist, wie zu Beginn dieses Paragraphen gezeigt wurde,

$$\{|r-h|\} = 2 \sum_{r=0}^{\gamma} (h-r) P_r$$

und sofern X_r die Anzahl der beobachteten Serien mit genau r seltenen Ereignissen bedeutet

$$\{|r-h|\} = \frac{2}{m} \left[h \sum_{r=0}^{\gamma} X_r - \sum_{r=0}^{\gamma} r X_r \right], \quad \gamma = [h].$$

Für das Beispiel der Todesfälle durch Dampfkessel-
explosionen erhalten wir auf diese Art (vgl. Zahlen-
tafel 1) 3.665. Berechnen wir hingegen für dieses Beispiel
 $\{|r-h|\}$ aus (31), so erhalten wir 3.658. Wir finden
eine sehr gute Übereinstimmung der Theorie mit der
Beobachtung; setzen wir aber die Theorie von Poisson
voraus (Unabhängigkeit), so wäre für $\{|r-h|\}$ nach
(31) ($d=0$) 1.77 zu erwarten, also ein Betrag, der
nur ein Bruchteil der Beobachtung ist.

Von Formel (36) haben wir bei allen Beispielen
Gebrauch gemacht.

Bekanntlich sind bei allen Anwendungen die Para-
meter h und d aus dem ersten und zweiten Moment
bestimmt worden. Beim gewöhnlichen Fall der unab-
hängigen Ereignisse wird jeweilen untersucht, ob das
theoretische Quadrat des Schwankungsmasses, berechnet
aus dem durch das erste Moment bestimmten einzigen
Parameter, mit dem direkt berechneten übereinstimmt.
Mit andern Worten, man untersucht, ob der Lexische

Dispersionskoeffizient gleich 1 sei. Wir wollen nun analog vorgehen und mit Hilfe der durch das erste und zweite Moment bestimmten zwei Parameter h und d das dritte Moment direkt berechnen. Nach (26) ist

$$\{r(r-1)(r-2)\} = h(h+d)(h+2d).$$

Für unser Pockenbeispiel ist

$$\{r(r-1)(r-2)\} = h(h+d)(h+2d) = 3673.06 \text{ (theoretisch)}$$

$$\{r(r-1)(r-2)\} = \{r^3\} - 3\{r^2\} + 2\{r\} = 3892.63 \text{ (beobachtet).}$$

Die Übereinstimmung der Theorie mit der Beobachtung scheint uns befriedigend zu sein.

Dieselbe Untersuchung machen wir nun noch an einem andern statistischen Material.

In den wöchentlichen „Bulletins des Schweizerischen Gesundheitsamtes“ werden jeweilen die neu zur Anzeige gelangten Erkrankungen an ansteckenden Krankheiten veröffentlicht, nach Kantonen und Krankheiten getrennt.

Aus diesen Angaben habe ich eine *Statistik der Erkrankungen an Pocken im Kanton Zürich* vom 15. Febr. 1919 bis 22. September 1923 verarbeitet. Die Resultate wurden ganz gleich wie in Spalte II der Zahlentafeln 1, 2, 3 etc. zusammengestellt. In den beobachteten 240 Wochen wurden im Kanton Zürich 981 Pockenerkrankungen amtlich angezeigt, im Durchschnitt pro Woche 4.09. (Die Serienlänge setzen wir hier entsprechend den Bulletins auf eine Woche fest.)

$$\{r\} = 4.09 = h, \quad \{r^2\} = 73.17, \quad \{r^3\} = 2044.70$$

$$d = \frac{\{r^2\} - h^2}{h} - 1 = 12.80, \quad s^2 = 56.44.$$

Es ist

$$\{r(r-1)(r-2)\} = h(h+d)(h+2d) = 2050.99 \text{ (theoretisch)}$$

$$\{r(r-1)(r-2)\} = \{r^3\} - 3\{r^2\} + 2\{r\} = 1833.37 \text{ (beobachtet).}$$

Auch hier finden wir die Differenz zwischen Theorie und Beobachtung unwesentlich.

§ 13. Diskussion des Schwankungsmasses bei Serienverbindung.

Das Quadrat des Schwankungsmasses ist bei Chancenvermehrung grösser als der Mittelwert, somit grösser als bei der Voraussetzung der Unabhängigkeit der Ereignisse. Wir fragen nun nach der Abhängigkeit des Schwankungsmasses von der Serienlänge.

Aus den Formeln (15) und (16) folgt ohne weiteres, dass die *mathematische Erwartung proportional zur Serienlänge* ist (zur Anzahl der Versuche). Dies ist auch selbstverständlich (vgl. § 7).

Ebenso schliessen wir, dass das Quadrat des Schwankungsmasses eine *lineare* bzw. *quadratische* Funktion der Serienlänge ist, je nachdem es sich um nicht zusammenhängende oder zusammenhängende Serien handelt. Wird eine Serie um den k -fachen Betrag verlängert (Vereinigung von k aufeinanderfolgenden Serien) und bezeichnen wir die mathematische Erwartung und das Quadrat des Schwankungsmasses mit h und s^2 bzw. h_k und s_k^2 , je nachdem wir die ursprüngliche oder die verlängerte Serie betrachten, so erhalten wir nach (15) und (16)

$$h_k = kh \tag{38}$$

für nicht zusammenhängende und zusammenhängende Serien, ferner für *nicht zusammenhängende Serien*

$$\left. \begin{aligned} \varsigma_k^2 &= h_k(1+d_k) = kh(1+d) = k\varsigma^2 \\ h_k &= kh, d_k = d \end{aligned} \right\} (39),$$

für *zusammenhängende* Serien hingegen

$$\left. \begin{aligned} \varsigma_k^2 &= h_k(1+d_k) = kh(1+kd) = k\varsigma^2 \frac{1+kd}{1+d} \\ h_k &= kh, d_k = k \cdot d \end{aligned} \right\} (40).$$

In Zahlentafel 3 haben wir die Todesfälle bei Dampfkesselexplosionen in 25 Jahren näher betrachtet. Als Serie wählten wir das Jahr (ich konnte keine kürzere Serie wählen, weil ich keine Statistik dafür fand). Nun verlängern wir die Serie auf das 5-fache. Je 5 aufeinanderfolgende Jahre vereinigen wir zu einer Serie, und zwar derart, dass alle 25 Jahre vertreten sind, aber kein Jahr in zwei Serien auftritt. Wir erhalten 5 Serien zu 5 Jahre. Wie wir früher h , ς^2 und daraus d berechnet haben, bestimmen wir nun h_5 , ς_5^2 und d_5 . Für h_5 erhalten wir $68.6 = 5 \cdot 13.72 = 5h$, was mit (38) übereinstimmt.

Ferner wird

$$\varsigma_5^2 = 432.75 \pm 303.94 \text{ (beobachtet).}$$

Der Fehler (± 218.69) von ς_5^2 wurde nach (35) berechnet. Aus Zahlentafel 3 erhalten wir $\varsigma^2 = 62.48 \pm 24.05$. Nach unserer Formel (39) für nicht zusammenhängende Serien ist somit

$$\varsigma_5^2 = 312.40 \pm 120.25$$

(theoretisch berechnet für nicht zusammenhängende Serien).

Nach Formel (40) hingegen

$$\varsigma_5^2 = 1284 \pm 927$$

(theoretisch berechnet für zusammenhängende Serien).

Daraus schliessen wir: Die Anzahl der Todesfälle bei Dampfkesselexplosionen in einem Monat ist unabhängig von derjenigen des vorangehenden Monats. Dies entspricht auch der unbefangenen Vermutung, da ja die einzelnen Dampfkesselexplosionen, die voneinander unabhängig sind (vgl. Zahlentafel 2), sich nicht über längere Zeitperioden erstrecken, und vor allem wird eine Explosion keine Todesfälle in zwei verschiedenen Monaten zur Folge haben.

Ganz anders steht es mit den in Zahlentafel 4, 5 und 6 betrachteten Infektionskrankheiten. Beim Pockenbeispiel der Zahlentafel 4 war die Serienlänge gleich ein Monat. Es werden nun nacheinander Serien zu 2, 3, 4, 6, 12 und 24 Monaten nach demselben Prinzip wie oben (bei den Dampfkesselexplosionstodesfällen) gebildet. Die Resultate dieser Untersuchung sind aus Zahlentafel 7 ersichtlich.

Die aufeinanderfolgenden Monate sind ganz ähnlich miteinander verwachsen wie die Serien bei der Theorie der „zusammenhängenden Serien“. In der Tat wird sich eine Pockenepidemie nicht auf einen Monat beschränken und mit dem letzten Tag des Monats erlöschen. Wir sehen dies auch wieder bei den Pockenepidemien der neuesten Zeit. In den Kriegsjahren waren in der Schweiz keine Pockenfälle aufgetreten, seit 1919 aber vergeht keine Woche, in der nicht ein Pockenfall zur Anzeige gelangt. Die Anzahl der Pockentodesfälle in einem Monat ist in hohem Masse vom Verlauf der Pocken im Vormonat abhängig.

Analoge Berechnungen habe ich für die Todesfälle an Scharlach in der Schweiz in den Jahren 1877 bis 1900 vorgenommen (vgl. Zahlentafel 5) und die Resultate in Zahlentafel 8 registriert. Wir erhalten folgendes Resultat: Die Annahme der nicht zusammenhän-

genden Serien müssen wir auch hier fallen lassen, da das Schwankungsmass verhältnismässig stärker zunimmt als die Serienlänge (vgl. Spalte IV der Zahlentafel 8). Aufeinanderfolgende Serien stecken einander an.

Die Ansteckung erfolgt ganz analog, wie sie die Theorie der zusammenhängenden Serien vorschreibt, solange wir nicht mehr als 6 Monate zu einer Serie vereinigen. Sobald wir aber mehr als 6 Monate vereinigen, so tritt eine Unstimmigkeit ein, die allerdings noch im Fehlerbereich liegt. Der Schritt von 12 auf 24 Monate ist wieder normal. Wir schliessen daraus (obschon die Serienzahl klein ist), dass die Beeinflussung der Winter- und Sommermonate nicht dieselbe ist wie bei gleichartigen Monaten. Wir haben 2 Serientypen vor uns, Monate des ersten und des zweiten Halbjahres (vgl. § 10).

Verlängern wir hingegen die Serien bei den Todesfällen an Scharlach im zweiten Halbjahr (vgl. Zahlentafel 6), so liefern die Berechnungen für den Verlauf des Schwankungsmasses die gewünschten Resultate, wovon man sich in Zahlentafel 9 überzeugen kann.

Wie steht es nun, wenn wir bei unserem Pockenbeispiel nicht zwei aufeinanderfolgende Monate, sondern zwei zeitlich weit auseinanderliegende zu einer Serie vereinigen? Wir erwarten, dass sich hier die am Anfang des Paragraphen entwickelte Theorie der Vereinigung unabhängiger Serien bewähre. Um dies zu prüfen, vereinigen wir das Resultat des

Januar 1877	mit dem Resultat des	Januar 1889
Februar 1877	„ „ „ „	Februar 1889
März 1877	„ „ „ „	März 1889
⋮		⋮
Januar 1888	„ „ „ „	Januar 1900
Dezember 1888	„ „ „ „	Dezember 1900.

Wir erhalten auf diese Art 144 Serien zu 2 Monaten (jeder Monat tritt nur einmal auf).

Es wird

$$\varsigma_2^2 = 146.85,$$

$$d_2 = 12.35 \sim d = 14.20 \text{ (Grössenordnung),}$$

während wir oben für $d_2 = 31.44$ erhalten hatten (vgl. Zahlentafel 7). Unsere Erwartungen haben sich somit bestätigt.

Die Resultate dieses Paragraphen wollen wir nun noch an unserem Beispiel der Erkrankungen an Pocken in Zürich untersuchen (vgl. § 12). Wir vereinigen 2 aufeinanderfolgende Wochen zu einer Serie und erhalten 120 neue 2-Wochenserien. Wir erhalten

$$h_2 = 8.18, \quad d_2 = 18.77 > d(12.80), \quad \frac{d_2}{2} = 9.39 \sim d$$

$$\varsigma_2^2 = 161.65 > 2\varsigma^2, \quad (\varsigma^2 = 56.44).$$

Die Vereinigung von je 3 aufeinanderfolgenden Wochen zu 80 3-Wochenserien ergibt

$$h_3 = 12.26, \quad \varsigma_3^2 = 348.12 > 3\varsigma^2$$

$$d_3 = 27.39 > d, \quad \frac{d_3}{3} = 9.13 \sim d.$$

Das System der Serienvereinigung lässt uns hier an dem Verlauf von d_k und damit des Schwankungsmasses auf *Chancenvermehrung* durch Erfolg mit Ansteckung der aufeinanderfolgenden Serien schliessen.

Die Theorie der Chancenvermehrung findet somit auch in der Statistik der *Erkrankungen* an Infektionskrankheiten Anwendung.

Wir sehen also, dass das Urnenschema der Chancenvermehrung in befriedigender Weise bei zwei charakteristischen statistischen Typen (Unfallstatistik und Epidemiostatistik, Todesfälle und Erkrankungen) Anwendung finden kann. Die Resultate dieses sowie der vorangehenden Paragraphen berechtigen uns zur Annahme, dass wir mit Hilfe des Urnenschemas der Chancenvermehrung die Statistik der Todesfälle bei Pocken, Explosionen, Eisenbahnunglücken etc. auf den allein wichtigen Kern untersuchen können, und dass der Schritt von der Wirklichkeit zum Pólyaschen Schema nichts anderes bedeutet als das Entfernen der Schlacke vom Kern.

Kapitel III.

Inhomogenität.

§ 14. Die Wahrscheinlichkeiten und ihre erzeugenden Funktionen.

Aus den Entwicklungen des vorangehenden Kapitels sehen wir, dass die übernormale Dispersion von der Abhängigkeit der Einzelfälle, die zu einer statistischen Masse zusammengezogen worden sind, herrühren kann.

Der Fall der übernormalen Dispersion lässt aber auch eine Entstehungsweise zu, welche keine Verkettung der Einzelfälle voraussetzt. Er kann auch davon herrühren, dass den einzelnen statistischen Massen

nicht dieselben Ursachen zugrunde liegen. In diesem Fall entspricht jedem Einzelglied der statistischen Reihe eine Spezialwahrscheinlichkeit, zum Unterschied zur Durchschnittswahrscheinlichkeit, die den Mittelwert der Spezialwahrscheinlichkeiten darstellt. Es ist einleuchtend, dass dieses modifizierte Schema grössere Schwankungen ergeben muss als das ursprüngliche, weil die zufälligen Abweichungen der Einzelglieder der Reihe von ihrem Mittelwert durch die Ungleichheit der Spezialwahrscheinlichkeiten vergrössert werden. Lexis zerlegt die Schwankung in eine erste Komponente, die „Normal zufällige Schwankungskomponente“, die aus der Wirkung der zufälligen Ursachen entsteht, und in eine zweite Komponente, die „Physische Schwankungskomponente“, die ihren Ursprung in der Verschiedenheit der einzelnen Spezialwahrscheinlichkeiten hat. In der Bevölkerungsstatistik hat man es fast immer mit Material, das sich über grosse Zeiträume erstreckt, und daher mit zeitlich sich ändernden Wahrscheinlichkeiten zu tun. Der Fall mit übernormaler Dispersion, herührend von der Verschiedenheit der Serienwahrscheinlichkeiten, wird daher nicht selten sein.

Selbstverständlich liegt es uns ferne zu behaupten, ein bestimmtes Material entspreche genau einer bestimmten Annahme der Inhomogenität. Es handelt sich hier auch wieder darum, den *Hauptcharakter* der Ereignisse zu erfahren. Betrachten wir etwa die im vorhergehenden Kapitel genau untersuchte Pockenstatistik. Wir konstatierten, dass das gewöhnliche Urnenschema ein sehr schlechtes Bild der Wirklichkeit gibt. Aus der Statistik entnehmen wir, dass in den fünfjährigen Perioden 1881—1885, 1886—1890, 1891—1895, 1896—1900: 703 bzw. 248, 128, 44 Todesfälle an Pocken verzeichnet wurden.

Die Pockentodesfälle nehmen mit der Zeit stark ab. Durch die Jahre des 20. Jahrhunderts (die nicht mehr zu unserer in Zahlentafel 4 aufgeführten Statistik gehören) wird diese Behauptung auch bestätigt. Wir konstatieren somit eine Inhomogenität von Serie zu Serie. Rührt die Unstimmigkeit des gewöhnlichen Urnenschemas zum grössten Teil oder ganz von der Inhomogenität her? Diese Frage muss, gestützt auf die Resultate der Zahlentafeln 4 und 7, verneint werden (vgl. § 18). Wir haben es mit Chancenvermehrung und Inhomogenität zu tun. Der Chancenvermehrungscharakter scheint zu prävalieren, und darum gibt vermutlich die Annahme der reinen Chancenvermehrung ein mit der Wirklichkeit so gut übereinstimmendes Resultat. Wenn wir aber Zahlentafel 4 genauer betrachten, so sehen wir, dass in der Kolonne IX, der Abweichung der Theorie der Chancenvermehrung von der Beobachtung, ein systematischer Gang vorhanden ist (zuerst negativ, dann positiv). Dieser Gang könnte seinen Ursprung in der Inhomogenität der Serien haben. Dass die Inhomogenität tatsächlich einen solchen Gang hervorruft, soll weiter unten gezeigt werden.

Beim Studium der Inhomogenität sind zwei Hauptklassen zu unterscheiden:

A. Inhomogenität innerhalb der Serie:

Die Wahrscheinlichkeiten ändern innerhalb einer Serie nach einem bestimmten Gesetz, das unabhängig ist von den Resultaten und für alle Serien dasselbe ist. So werden z. B. die Selbstmorde innerhalb eines Landes ein statistisches Material liefern, das dieser Kategorie angehört, weil die Selbstmordwahrscheinlichkeiten bei beiden Geschlechtern recht verschieden sind. Wir

werden aber später sehen, dass sich dieses Material trotzdem durch das einfache Urnschema beschreiben lässt.

B. Inhomogenität von Serie zu Serie:

Die Proben innerhalb ein und derselben Serie sind den gleichen Bedingungen unterworfen; diese Bedingungen ändern aber von Serie zu Serie, z. B. Sterblichkeit für eine bestimmte Altersgruppe während einer langen Zeitperiode.

In Wirklichkeit werden die Klassen A und B nicht immer getrennt auftreten, sondern es wird meistens sowohl Inhomogenität innerhalb der Serien wie auch von Serie zu Serie vorhanden sein. Unsere erste Aufgabe ist dann die, das Unwesentliche vom Wesentlichen zu trennen. Wir werden im folgenden die Klassen A und B vom Gesichtspunkt der Theorien der seltenen Ereignisse und der Chancenvermehrung näher betrachten.

Bei Inhomogenität innerhalb der Serien (Klasse A) bedeuten $Q', Q'', \dots Q^{(s)}$ bzw. $P', P'', \dots P^{(s)}$ die Wahrscheinlichkeiten innerhalb der homogenen Teilserien, während Q^* und P^* die Wahrscheinlichkeiten in den vollen inhomogenen Serien darstellen. (Analog wie im vorhergehenden Kapitel ist wieder $\lim_{\delta=0} P^* = Q^*$.)

Bei Inhomogenität von Serie zu Serie bezeichnen wir die Wahrscheinlichkeiten in den einzelnen Serien mit $Q', Q'', \dots Q^{(s)}$ bzw. $P', P'', \dots P^{(s)}$, die Wahrscheinlichkeit für das ganze Material wird mit \bar{Q} bzw. \bar{P} dargestellt.

Wahrscheinlichkeiten, die wir durch Charliersche Reihen darstellen (vgl. § 16), sind mit W bzw. w (wenn es sich nicht um den Grenzfall handelt) bezeichnet.

Die folgende Tabelle stellt uns die Bezeichnung der verschiedenen Wahrscheinlichkeiten dar, wie ich sie in dieser Arbeit gebrauche:

		Ohne Grenzfall	Seltene Ereignisse	Inhomogenität: Teilerien, einzelne Serien	Inhomogenität innerhalb der Serie	Inhomogenität von Serie zu Serie
Chancenvermehrung	$\delta > 0$	$p_{r,s}$	P_r	$P', P'', \dots P^{(s)}$	P^*	\bar{P}
Unabhängigkeit	$\delta = 0$	$q_{r,s}$	Q_r	$Q', Q'' \dots Q^{(s)}$	Q^*	\bar{Q}
Allgemein, Charlier		$w_{r,s}$	W_r			

Wie bei den meisten theoretischen Untersuchungen der Wahrscheinlichkeitsrechnung erweist sich auch hier die Anwendung der erzeugenden Funktionen als vorteilhaft. Mit diesen charakteristischen Funktionen erhalten wir verschiedene Beziehungen viel leichter als durch die gewöhnlichen Berechnungen und dringen so tiefer in den Stoff ein.

Im besonderen wird mit Hilfe der charakteristischen Funktionen eine Brücke von der Theorie der Chancenvermehrung zur Theorie der Variation der Spezialwahrscheinlichkeiten geschlagen. Auf diesen Zusammenhang, auf den ich weiter unten zurückkommen werde, wurde ich von Herrn Prof. Dr. *Pólya* aufmerksam gemacht.

In den Entwicklungen der nächsten Paragraphen werde ich für die erzeugenden Funktionen der verschiedenen Wahrscheinlichkeiten folgende Bezeichnungen einführen:

Wahrscheinlichkeiten	Erzeugende Funktionen	Analytischer Ausdruck der erzeugenden Funktion
Q	$F(z)$	$e^{-h(1-z)}$ vgl. (25)
P	$E(z)$	$(1+d(1-z))^{\frac{-h}{d}}$ vgl. (24)
Q^*	$F^*(z)$	$e^{-\sum_{\gamma=1}^s h_{\gamma}(1-z)}$ (41)
P^*	$E^*(z)$	$\prod_{\gamma=1}^s (1+d_{\gamma}(1-z))^{\frac{-h_{\gamma}}{d_{\gamma}}}$ (42)
\bar{Q}	$\bar{F}(z)$	$\frac{1}{s} \sum_{\gamma=1}^s e^{-h_{\gamma}(1-z)}$ (43)
\bar{P}	$\bar{E}(z)$	$\frac{1}{s} \sum_{\gamma=1}^s (1+d_{\gamma}(1-z))^{\frac{-h_{\gamma}}{d_{\gamma}}}$ (44)

In dieser Tabelle bedeutet etwa Formel (44):

$$\begin{aligned} \bar{E}(z) &= \bar{P}_0 + \bar{P}_1 z + \bar{P}_2 z^2 + \dots \\ &= \frac{1}{s} \left\{ (1+d_1(1-z))^{\frac{-h_1}{d_1}} + (1+d_2(1-z))^{\frac{-h_2}{d_2}} + \dots + (1+d_s(1-z))^{\frac{-h_s}{d_s}} \right\}. \end{aligned}$$

Die Formeln (41) und (42) ergeben sich ohne weiteres aus der Bedeutung der erzeugenden Funktionen (vgl. § 10) und aus dem Multiplikationssatz der Wahrscheinlichkeitsrechnung, da ja die Wahrscheinlichkeit,

dass in n Versuchen r seltene Ereignisse auftreten, gleich ist der Wahrscheinlichkeit, dass in sämtlichen Teilserien zusammen r seltene auftreten. Da \bar{P} bzw. \bar{Q} nichts anderes als Mittelwerte der Serienwahrscheinlichkeiten sind, so lassen sich auch die erzeugenden Funktionen als Mittelwerte darstellen, und wir erhalten (43) und (44). (Vgl. übrigens Czuber 1, Seite 120, und Charlier 2, Seite 2.)

Bei der Bildung der Formeln (42) und (44) wurde vorausgesetzt, dass sich die Chancenvermehrung nur über die Teilserien bzw. über die einzelnen homogenen Serien erstreckt, mit andern Worten, wir schliessen die Fälle aus, bei denen die Teilserien bzw. die Serien einander anstecken. Dieser Fall der *totalen Ansteckung* ist weit schwieriger zu bewältigen und benötigt umfangreiche Berechnungen. Trotzdem er nicht in den Rahmen dieser Arbeit gehört, wollte ich ihn hier erwähnen, da er, wie wir gesehen haben (vgl. Kapitel II), bei epidemischen Krankheiten eine grosse Rolle spielt.

§ 15. Verlauf der Wahrscheinlichkeiten.

Sofern es keine zwingenden Gründe verbieten, wird der Statistiker sein Material zuerst immer auf den einfachsten Fall von Bernoulli untersuchen. Liefert ihm die Theorie mit der Beobachtung keine genügende Übereinstimmung, so wird er nach den Ursachen der Abweichung forschen. Wir wollen nun sehen, welche Unstimmigkeiten die Inhomogenität gegenüber dem gewöhnlichen Fall der unabhängigen seltenen Ereignisse hervorruft.

Die Inhomogenität innerhalb der Serien macht sich beim gewöhnlichen Fall der seltenen Ereignisse nicht bemerkbar. Wir haben es also hier mit s homo-

genen, unter sich verschiedenen Teilserien zu tun, die zusammen eine Serie bilden. Da die mathematische Hoffnung für die ganze Serie gleich ist der Summe der mathematischen Hoffnungen der Teilserien, erhalten wir

$$h = \sum_{\gamma=1}^s h_{\gamma}.$$

Nach (41) und (25) ist

$$F^{**}(z) = e^{-\sum_{\gamma=1}^s h_{\gamma}(1-z)} = e^{-(1-z)\sum h_{\gamma}} = e^{-h(1-z)} = F(z),$$

womit unsere Behauptung bewiesen ist, da die erzeugende Funktion die Wahrscheinlichkeitsverhältnisse eindeutig bestimmt.

Wird die tatsächlich vorhandene *Inhomogenität von Serie zu Serie* nicht berücksichtigt, so sind die theoretisch berechneten Wahrscheinlichkeiten $Q_0, Q_1, Q_2, \dots, Q_r, \dots$ für kleines und grosses r zu klein und für mittleres r (in der Umgebung des Mittelwertes) zu gross. Kürzer ausgedrückt: Bei Inhomogenität von Serie zu Serie verläuft die Verteilungskurve *flacher* als bei voller Homogenität.

Zum Beweise des vorstehenden Satzes betrachten wir die Funktion

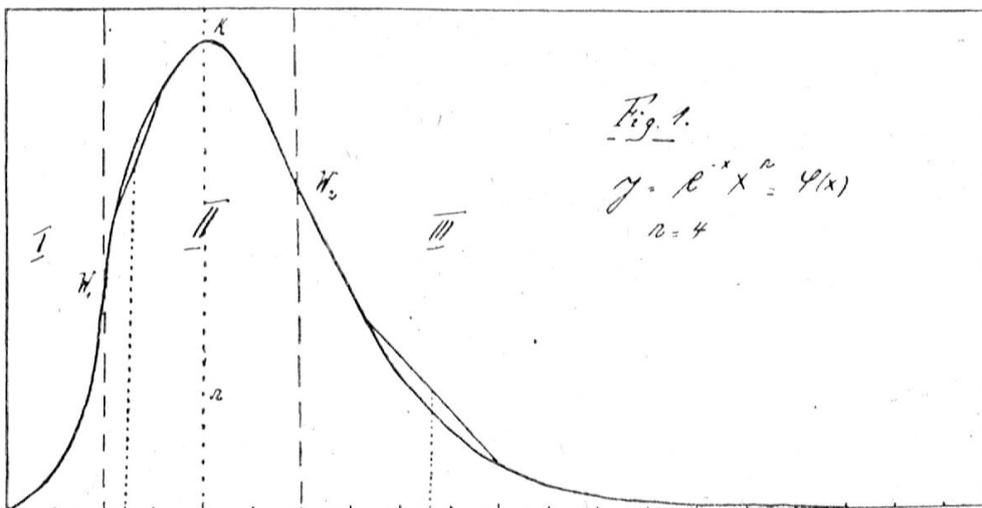
$$\varphi(x) = e^{-x} x^r, \quad \varphi'(x) = e^{-x} x^{r-1}(r-x).$$

$\varphi'(x)$ ist > 0 , solange $x < r$, wird negativ für $x > r$. Der Punkt K mit den Koordinaten $r, e^{-r} r^r$ ist Kulminationspunkt (vgl. Figur 1).

$$\varphi''(x) = e^{-x} x^{r-2}(x^2 - 2rx + r[r-1]).$$

Die Wendepunkte W_1 und W_2 haben die Abszissen $r - \sqrt{r}$ und $r + \sqrt{r}$. Durch diese beiden Punkte W_1 und W_2 wird die Kurve in 3 Teile zerlegt:

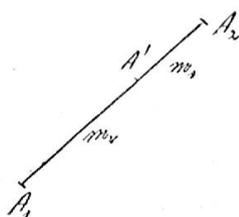
- I: $0 \leq x < r - \sqrt{r}$ in diesem Gebiet ist die Kurve konvex nach unten.
 II: $r - \sqrt{r} < x < r + \sqrt{r}$ in diesem Gebiet ist die Kurve konkav nach unten.
 III: $r + \sqrt{r} < x$ in diesem Gebiet ist die Kurve konvex nach unten.



Es gilt somit (vgl. Figur 1):

Im Gebiet I und III ist die Ordinate eines Sehnenpunktes grösser als die Ordinate des Kurvenpunktes mit gleicher Abszisse. Im Gebiet II ist die Ordinate eines Sehnenpunktes kleiner als die Ordinate des entsprechenden Kurvenpunktes.

Es teile der Sehnenpunkt A' die durch die Kurvenpunkte A_1 und A_2 begrenzte Sehne $A_1 A_2$ im Verhältnis von $m_2 : m_1$, dann ist, die entsprechenden Koordinaten mit (x', y') , (x_1, y_1) , (x_2, y_2) bezeichnet,



$$\frac{A_1 A'}{A' A_2} = \frac{m_2}{m_1} = \frac{x' - x_1}{x_2 - x'}$$

$$x' = \frac{m_1 x_1 + m_2 x_2}{m_1 + m_2} \text{ und analog}$$

$$y' = \frac{m_1 y_1 + m_2 y_2}{m_1 + m_2}.$$

Es sei A_3 der Kurvenpunkt mit den Koordinaten x_3, y_3 , wobei

$$x_3 = x' = \frac{m_1 x_1 + m_2 x_2}{m_1 + m_2}$$

ist, dann wird

$y_3 < y'$ für die Gebiete I und III

$y_3 > y'$ für das Gebiet II (vgl. Figur 1).

A_1 und A_2 müssen in demselben Gebiet liegen.

Sofern $x_1, x_2 < r - \sqrt{r}$ oder $x_1, x_2 > r + \sqrt{r}$ ist, erhalten wir

$$\frac{m_1 e^{-x_1} x_1^r + m_2 e^{-x_2} x_2^r}{m_1 + m_2} > e^{-\frac{m_1 x_1 + m_2 x_2}{m_1 + m_2}} \left(\frac{m_1 x_1 + m_2 x_2}{m_1 + m_2} \right)^r \quad (45),$$

für $r - \sqrt{r} < x_1, x_2 < r + \sqrt{r}$ wird

$$\frac{m_1 e^{-x_1} x_1^r + m_2 e^{-x_2} x_2^r}{m_1 + m_2} < e^{-\frac{m_1 x_1 + m_2 x_2}{m_1 + m_2}} \left(\frac{m_1 x_1 + m_2 x_2}{m_1 + m_2} \right)^r \quad (45').$$

Wir betrachten ein statistisches Material, von dem wir $m_1 + m_2$ Serien zur Verfügung haben. In m_1 Serien

soll die Spezialwahrscheinlichkeit durchwegs q_1 , in den übrigen m_2 Serien q_2 betragen.

$$\lim_{n \rightarrow \infty} n q_1 = h_1, \quad \lim_{n \rightarrow \infty} n q_2 = h_2.$$

Für die Wahrscheinlichkeit \bar{Q}_r erhalten wir

$$\bar{Q}_r = \frac{m_1 e^{-h_1} h_1^r + m_2 e^{-h_2} h_2^r}{r! (m_1 + m_2)}$$

und für die Wahrscheinlichkeit, die innerhalb der Totalserie Homogenität voraussetzt und somit die vorhandene Inhomogenität vernachlässigt,

$$Q_r = \frac{e^{-\frac{h_1 m_1 + h_2 m_2}{m_1 + m_2}} \left(\frac{h_1 m_1 + h_2 m_2}{m_1 + m_2} \right)^r}{r!}.$$

Wir erhalten somit nach obigen Formeln (45), (45')

$$\bar{Q}_r > Q_r \text{ für } h_1, h_2 < r - \sqrt{r} \text{ oder } h_1, h_2 > r + \sqrt{r} \quad (46)$$

$$\bar{Q}_r < Q_r \text{ für } r - \sqrt{r} < h_1, h_2 < r + \sqrt{r}. \quad (46')$$

Es sei $h_1 \leq h_2$, dann genügt es, die 4 Fälle: h_1 oder h_2 gleich $r - \sqrt{r}$ oder $r + \sqrt{r}$ zu diskutieren, die auf die quadratische Gleichung führen

$$(\pm \sqrt{r})^2 = (h - r)^2,$$

deren beide Wurzeln $h + \frac{1}{2} \pm \sqrt{h + \frac{1}{4}}$ sind.

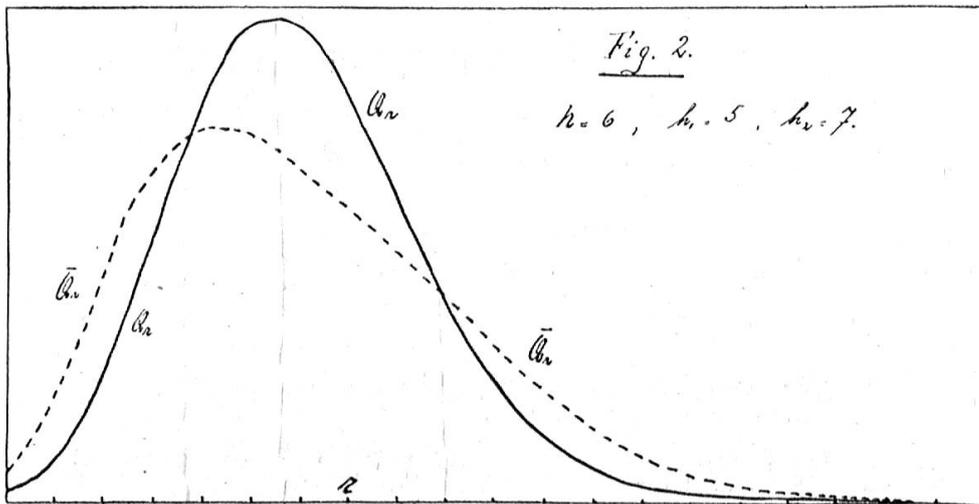
Für $r < h_1 + \frac{1}{2} - \sqrt{h_1 + \frac{1}{4}}$ ist $h_1, h_2 > r \pm \sqrt{r}$,

für $r > h_2 + \frac{1}{2} + \sqrt{h_2 + \frac{1}{4}}$ ist $h_1, h_2 < r + \sqrt{r}$.

Unter Berücksichtigung von dieser Beziehung geht (46) über in

$$\bar{Q}_r > Q_r \text{ für } r < h_1 + \frac{1}{2} - \sqrt{h_1 + \frac{1}{4}} \text{ oder } r > h_2 + \frac{1}{2} + \sqrt{h_2 + \frac{1}{4}}$$

$$\bar{Q}_r < Q_r \text{ für } h_2 + \frac{1}{2} - \sqrt{h_2 + \frac{1}{4}} < r < h_1 + \frac{1}{2} + \sqrt{h_1 + \frac{1}{4}} \quad (47).$$



Im vorstehenden haben wir die zwei Gebiete

$$h_1 + \frac{1}{2} - \sqrt{h_1 + \frac{1}{4}} < r < h_2 + \frac{1}{2} - \sqrt{h_2 + \frac{1}{4}}$$

und
$$h_1 + \frac{1}{2} + \sqrt{h_1 + \frac{1}{4}} < r < h_2 + \frac{1}{2} + \sqrt{h_2 + \frac{1}{4}}$$

nicht berücksichtigt. In diesen Zwischengebieten können wir über das Vorzeichen von $\bar{Q}_r - Q_r$ nichts aussagen. Es ist dies auch erklärlich, da wir über das Verhältnis der Ordinaten eines Schnenpunktes und des Kurvenpunktes mit gleicher Abszisse auch nichts Bestimmtes wissen, solange die Sehne in zwei Gebieten

der Figur 1 liegt. Ist $h_1 \sim h_2 \sim h$, d. h. sind die Differenzen $h_1 - h$, $h_2 - h$ klein, so geht (47) über in

$$\left. \begin{array}{l} \bar{Q}_r > Q_r \\ \text{für } r < h + \frac{1}{2} - \sqrt{h + \frac{1}{4}} \\ \text{oder } r > h + \frac{1}{2} + \sqrt{h + \frac{1}{4}}; \\ \bar{Q}_r < Q_r \\ \text{für } h + \frac{1}{2} - \sqrt{h + \frac{1}{4}} < r < h + \frac{1}{2} + \sqrt{h + \frac{1}{4}} \end{array} \right\} (48).$$

Bisher haben wir nur den Fall mit zwei Serientypen betrachtet. Der allgemeinste Fall ergibt sich aus diesem ohne weiteres durch Induktionsschluss. Wir erhalten

$$\left. \begin{array}{l} \bar{Q}_r > Q_r \quad \text{für } r < r_1 \quad \text{oder } r > r_4 \\ \bar{Q}_r < Q_r \quad \text{für } r_2 < r < r_3 \end{array} \right\} (49)$$

wobei

r_1 das Minimum der Grössen $h_1 + \frac{1}{2} - \sqrt{h_1 + \frac{1}{4}}$,

$\dots h_s + \frac{1}{2} - \sqrt{h_s + \frac{1}{4}}$,

r_2 das Maximum von diesen Grössen darstellt.

r_3 ist das Minimum der Grössen $h_1 + \frac{1}{2} + \sqrt{h_1 + \frac{1}{4}}$,

$h_2 + \frac{1}{2} + \sqrt{h_2 + \frac{1}{4}}, \dots h_s + \frac{1}{2} + \sqrt{h_s + \frac{1}{4}}$,

r_4 ist das Maximum derselben.

Die summatorische Reihe. Ein besseres Urteil als die Einzelwahrscheinlichkeiten bieten uns, besonders bei kleiner Serienzahl, die summatorischen Reihen (vgl. Zahlentafel 1—6). Es ist

$$S_r = \sum_{v=0}^r Q_v = e^{-h} \left(1 + h + \frac{h^2}{2!} + \dots + \frac{h^r}{r!} \right).$$

Nun fragen wir nach dem Verhalten der summatorischen Reihen, wenn wir die Inhomogenität von Serie zu Serie vernachlässigen. Wir vergleichen, analog wie oben, \bar{S}_r mit S_r . Es ist

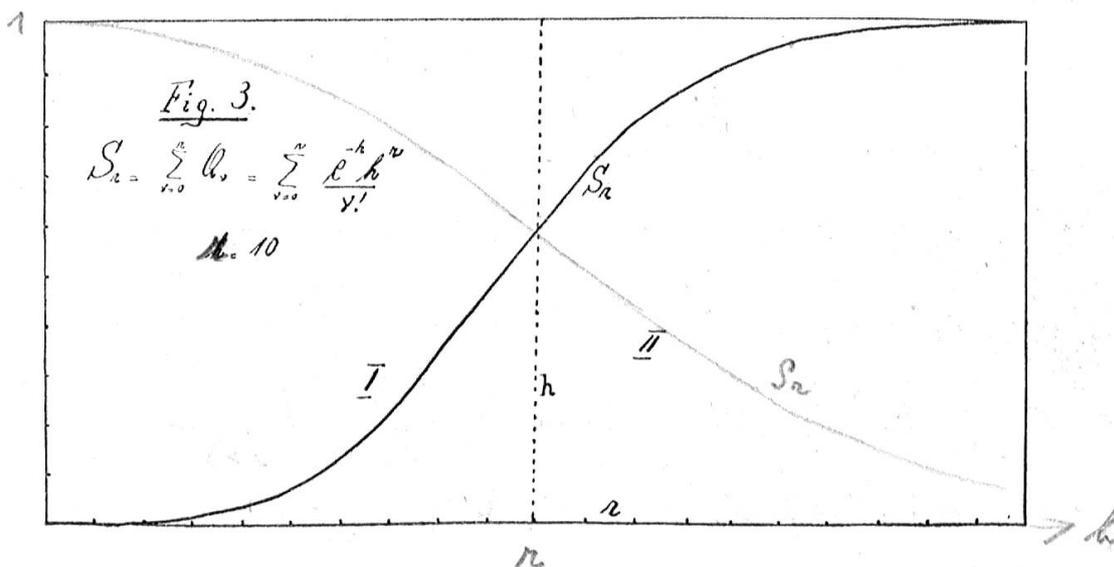
$$S_r = e^{-h} \left(1 + h + \frac{h^2}{2!} + \dots + \frac{h^r}{r!} \right) = \Phi(h),$$

$$\Phi'(h) = -\frac{e^{-h} h^r}{r!} < 0, \quad \Phi''(h) = \frac{e^{-h} h^{r-1}}{r!} (h-r),$$

$\Phi''(h) > 0$, für $h > r$, die Kurve $y = \Phi(h)$ ist ^{vex} konkav nach unten,

$\Phi''(h) < 0$, für $h < r$, die Kurve $y = \Phi(h)$ ist ^{conv} konvex nach unten,

für $h = r$ hat die Kurve $y = \Phi(h)$ einen Wendepunkt.



Liegt eine Sehne ganz im Gebiet $h < r$, so ist die Ordinate des Kurvenpunktes grösser als die Ordinate des Sehnenpunktes mit gleicher Abszisse. Liegt die Sehne aber im Gebiet $h > r$, so ist die Ordinate des Kurvenpunktes kleiner als die Ordinate des Sehnenpunktes mit gleicher Abszisse. Es sei nun

$$\bar{S}_r = \frac{1}{2} \left[e^{-h_1} \left(1 + h_1 + \frac{h_1^2}{2!} + \dots + \frac{h_1^r}{r!} \right) + e^{-h_2} \left(1 + h_2 + \frac{h_2^2}{2!} + \dots + \frac{h_2^r}{r!} \right) \right] = \frac{\Phi(h_1) + \Phi(h_2)}{2}$$

dann gilt für $h = \frac{h_1 + h_2}{2}$, aus analogen Überlegungen, wie beim Vergleich von Q_r mit \bar{Q}_r ,

$$\begin{aligned} \bar{S}_r &> S_r && \text{für } h_1, h_2 > r \\ \bar{S}_r &< S_r && \text{für } h_1, h_2 < r \text{ (vgl. Fig. 3).} \end{aligned}$$

Ist nun S_r die summatorische Reihe bei Inhomogenität von Serie zu Serie, sind h_1, h_2, \dots, h_s die mathematischen Erwartungen der homogenen einzelnen Serien und ist $h_1 \leq h_2 \leq \dots \leq h_s$, dann erhalten wir

$$\left. \begin{aligned} \bar{S}_r &> S_r && \text{für } r < h_1 \\ \bar{S}_r &< S_r && \text{für } r > h_s \end{aligned} \right\} (50).$$

Betrachten wir nun noch den Fall der schwachen Inhomogenität, in welchem $h_1 \sim h_2 \sim h_3 \dots \sim h_s \sim h$ ist. Wir erhalten

$$\left. \begin{aligned} \bar{S}_r &> S_r && \text{für } r < h \\ \bar{S}_r &< S_r && \text{für } r > h \\ \bar{S}_r &= S_r && \text{für } r = h \end{aligned} \right\} (51).$$

Nun zur statistischen Ausnutzung der vorstehenden Resultate. Wir betrachten nach dem Vorgang von L. v. Bortkiewicz in Zahlentafel 10 die *Statistik der weiblichen Selbstmorde* in 8 deutschen Staaten [Bortkiewicz 1, § 10]. Vergleichen wir die durchschnittlichen jährlichen Selbstmorde in den verschiedenen Staaten, die zwischen 1.4 und 5.6 liegen, so springt uns die Ungleichheit sofort in die Augen. Wir haben es mit Inhomogenität von Serie zu Serie zu tun. Zum Vergleich der Wahrscheinlichkeiten, die die Inhomogenität vernachlässigen, mit denjenigen, die ihr Rechnung tragen, habe ich Zahlentafel 10 ganz analog wie Zahlentafel 1, 3 etc. angelegt. An Stelle der Spalten der Chancenvermehrung (IV, VII und IX) treten hier die Spalten der Inhomogenität. Vergleichen wir nun die Spalte III der Wahrscheinlichkeiten mit den Beobachtungen, Spalte II, so finden wir unsere Theorie bestätigt. $112 Q_r$ ist für kleines und grosses r zu klein, für mittleres zu gross. Der Zeichenumschlag hat nach der Theorie in der Umgebung von $r = 3.47$ (genau zwischen $h_1 = 1.4$ und $h_s = 5.6$) zu erfolgen. In Wirklichkeit ist Spalte VIII für $r = 3$ negativ, für $r = 4$ positiv, also genau wie es zu erwarten ist. Ferner sind die nach (49) zu bestimmenden Grössen $r_1 = 0.6$, $r_2 = 3.2$, $r_3 = 3.7$, $r_4 = 8.5$, somit ist zu erwarten, dass

$$\overline{Q_r} > Q_r \quad \text{für } r < 0.6$$

$$\overline{Q_r} > Q_r \quad \text{für } r > 8.5$$

$$\overline{Q_r} < Q_r \quad \text{für } 3.2 < r < 3.7.$$

Auch diese Vermutung finden wir, wenn wir die Kolonnen III und II miteinander vergleichen, gut bestätigt.

In derselben Arbeit von L. v. Bortkiewicz ist in § 11 eine *Statistik der tödlichen Unfälle* bei 11 Berufsgenossenschaften aufgeführt. In Zahlentafel 11, analog angelegt wie Zahlentafel 10, wird dieses inhomogene Material verarbeitet. Auch hier finden wir unsere Theorie bestätigt. Der Zeichenwechsel in Spalte VIII erfolgt bei $r = 4$, nach der Theorie bei $r \sim 4.4$. Da die Spalte III der Zahlentafeln 10 und 11 ihr Maximum bei h haben, ist nach der Theorie zu erwarten, dass der Zeichenumschlag in VIII ungefähr dort erfolgt, wo III das Maximum erreicht hat.

§ 16. Fortsetzung. Die Charlierschen Reihen.

Eine andere Ableitung der im vorangehenden Paragraphen gewonnenen Sätze geht aus der Umformung der erzeugenden Funktionen hervor. Man hat nach (43)

$$\bar{F}(z) = \frac{1}{s} \sum_{\gamma=1}^s e^{-h_{\gamma}(1-z)} = \frac{1}{s} e^{-h(1-z)} \sum_{\gamma=1}^s e^{\sum_{\gamma=1}^s (h-h_{\gamma})(1-z)} \quad \sum h_{\gamma} = h \cdot s$$

$$\bar{F}(z) = \frac{1}{s} e^{-h(1-z)} \left[s - \sum_{\gamma=1}^s (h_{\gamma} - h)(1-z) + \frac{\sum_{\gamma=1}^s (h-h_{\gamma})^2 (1-z)^2}{2!} \right. \\ \left. + \frac{\sum_{\gamma=1}^s (h-h_{\gamma})^3 (1-z)^3}{3!} + \dots \right]$$

$$\bar{F}(z) = e^{-h(1-z)} \left[1 + \frac{\sum_{\gamma=1}^s (h-h_{\gamma})^2 (1-z)^2}{2! s} + \frac{\sum_{\gamma=1}^s (h-h_{\gamma})^3 (1-z)^3}{3! s} + \dots \right] \quad (52)^1).$$

¹⁾ Diese Reihe konvergiert für alle Werte von z .

Ferner ist nach (24)

$$\begin{aligned}
 E(z) &= (1+d(1-z))^{-\frac{h}{d}} = e^{-\frac{h}{d} \log(1+d(1-z))} \\
 &= e^{-h(1-z) + \frac{hd(1-z)^2}{2} - \frac{hd^2(1-z)^3}{3} + \dots} \\
 &= e^{-h(1-z)} \left[1 + \frac{hd(1-z)^2}{2} - \frac{hd^2(1-z)^3}{3} + \dots \right] \quad (53)^1.
 \end{aligned}$$

Aus (44) erhalten wir

$$\begin{aligned}
 \bar{E}(z) &= \frac{1}{s} \sum_{\gamma=1}^s (1+d_{\gamma}(1-z))^{-\frac{h_{\gamma}}{d_{\gamma}}} = \frac{1}{s} \sum_{\gamma=1}^s e^{-h_{\gamma}(1-z)} \left(1 + \frac{h_{\gamma} d_{\gamma} (1-z)^2}{2} \right. \\
 &\quad \left. - \frac{h_{\gamma} d_{\gamma}^2}{3} (1-z)^3 + \dots \right) \\
 &= e^{-h(1-z)} \frac{1}{s} \sum_{\gamma=1}^s e^{-(h_{\gamma}-h)(1-z)} \left(1 + \frac{h_{\gamma} d_{\gamma} (1-z)^2}{2} - \frac{h_{\gamma} d_{\gamma}^2}{3} (1-z)^3 + \dots \right) \\
 &= e^{-h(1-z)} \left[1 + \frac{\sum_{\gamma=1}^s (h-h_{\gamma})^2 (1-z)^2}{2! s} + \frac{\sum_{\gamma=1}^s h_{\gamma} d_{\gamma}}{2 s} (1-z)^2 \right. \\
 &\quad \left. + \frac{\sum_{\gamma=1}^s (h-h_{\gamma})^3 (1-z)^3}{3! s} - \frac{\sum_{\gamma=1}^s h_{\gamma} d_{\gamma}^2}{3 s} (1-z)^3 + \dots \right] \quad (54)^1.
 \end{aligned}$$

¹⁾ Konvergenz, solange $|(z-1)| < \frac{1}{d}$.

Bei Inhomogenität innerhalb einer Serie mit gleichzeitiger Chancenvermehrung erhalten wir aus (42)

$$E^*(z) = \prod_{\gamma=1}^s (1 + d_\gamma(1-z))^{-\frac{h_\gamma}{d_\gamma}} = e^{-\sum_{\gamma=1}^s \frac{h_\gamma}{d_\gamma} \log(1 + d_\gamma(1-z))}, \quad \sum_{\gamma=1}^s h_\gamma = h$$

$$E^*(z) = e^{-\sum_{\gamma=1}^s h_\gamma(1-z) + \sum_{\gamma=1}^s \frac{h_\gamma d_\gamma}{2}(1-z)^2 - \frac{1}{3} \sum_{\gamma=1}^s h_\gamma d_\gamma^2 (1-z)^3 + \dots}$$

$$E^*(z) = e^{-h(1-z)} \left[1 + \frac{1}{2} \sum_{\gamma=1}^s h_\gamma d_\gamma (1-z)^2 - \frac{1}{3} \sum_{\gamma=1}^s h_\gamma d_\gamma^2 (1-z)^3 + \dots \right] \quad (55).$$

Der *allgemeine Typus der Reihen*, wie er in (52) bis (55) auftritt, ist

$$G(z) = W_0 + W_1 z + W_2 z^2 + \dots = e^{-h(1-z)} [1 + \gamma_2(1-z)^2 + \gamma_3(1-z)^3 + \dots] \quad (56).$$

Die Wahrscheinlichkeiten W_0, W_1, W_2, \dots , die durch eine solche Reihe erzeugt werden, hängen ausser von h von den Konstanten $\gamma_2, \gamma_3, \gamma_4$ ab. Sie werden mit den Wahrscheinlichkeiten $Q_0, Q_1, \dots, Q_n, \dots$ durch folgende wichtige Formel verbunden:

$$W_r = \sum_{\nu=0}^{\infty} \gamma_\nu \Delta^\nu Q_\nu \quad (57),$$

wobei das Zeichen Δ^ν die in der Differenzenrechnung übliche Bedeutung hat und $\gamma_0 = 1, \gamma_1 = 0$ zu setzen ist. Wir erhalten nämlich aus dem Ansatz (56)

$$G(z) = F(z) + \gamma_2 (F(z) - 2zF(z) + z^2 F(z)) + \gamma_3 (F(z) - 3zF(z) + 3z^2 F(z) - z^3 F(z)) + \dots$$

und daraus ist

$$\begin{aligned} W_r &= Q_r + \gamma_2(Q_r - 2Q_{r-1} + Q_{r-2}) \\ &\quad + \gamma_3(Q_r - 3Q_{r-1} + 3Q_{r-2} - Q_{r-3}) + \dots \\ &= Q_r + \gamma_2[(Q_r - Q_{r-1}) - (Q_{r-1} - Q_{r-2})] \\ &\quad + \gamma_3[((Q_r - Q_{r-1}) - (Q_{r-1} - Q_{r-2})) \\ &\quad - ((Q_{r-1} - Q_{r-2}) - (Q_{r-2} - Q_{r-3}))] + \dots \end{aligned}$$

$$W_r = Q_r + \gamma_2 \Delta^2 Q_r + \gamma_3 \Delta^3 Q_r + \dots$$

$$Q_r - Q_{r-1} = \Delta Q_r, \quad \Delta Q_r - \Delta Q_{r-1} = \Delta^2 Q_r$$

$$W_r = \sum_{\nu=0}^{\infty} \gamma_{\nu} \Delta^{\nu} Q_r$$

womit Gleichung (57) bewiesen ist.

Diese *Zerlegung der Wahrscheinlichkeiten*, wie sie in (57) angegeben ist, rührt von Charlier her (vgl. Charlier **1**, **2**, **3** und **4**). Sie ergibt, wie wir sehen werden, eine interessante Analyse der Natur der Kurve.

1. *Fall: γ_2 überwiegt.* Wir nehmen $\gamma_3 = \gamma_4 = \dots = 0$ an. Ist die *zweite Charakteristik*, d. h. γ_2 , die wir mit Charlier als *Exzentrizität* bezeichnen wollen, ungleich null und verschwinden alle höheren Charakteristiken, so wird

$$\begin{aligned} W_r &= (1 + \gamma_2)Q_r - 2\gamma_2 Q_{r-1} + \gamma_2 Q_{r-2} \\ &= Q_r \left[1 + \gamma_2 - 2\gamma_2 \frac{r}{h} + \gamma_2 \frac{r(r-1)}{h^2} \right], \end{aligned}$$

$$W_r - Q_r = \gamma_2 \left[1 - \frac{2r}{h} + \frac{r(r-1)}{h^2} \right].$$

Ist

$$1 - \frac{2r}{h} + \frac{r(r-1)}{h^2} = 0, \text{ so wird } r = h + \frac{1}{2} \pm \sqrt{h + \frac{1}{4}},$$

somit

$$\left. \begin{aligned} W_r &> Q_r \\ &\text{für } r < h + \frac{1}{2} - \sqrt{h + \frac{1}{4}} \\ &\text{oder } r > h + \frac{1}{2} + \sqrt{h + \frac{1}{4}}; \\ W_r &< Q_r \\ &\text{für } h + \frac{1}{2} - \sqrt{h + \frac{1}{4}} < r < h + \frac{1}{2} + \sqrt{h + \frac{1}{4}} \end{aligned} \right\} (58).$$

Positive Exzentrizität bewirkt eine Vergrößerung der Wahrscheinlichkeiten bei grossem und kleinem r , bei mittlerem r (in der Nähe der mathematischen Erwartung) hat sie den umgekehrten Einfluss. *Positive Exzentrizität flacht die Wahrscheinlichkeitskurve ab.*

Nun betrachten wir die *Summenfunktion*: Es ist im allgemeinen, sofern wir die erzeugende Funktion der Wahrscheinlichkeiten (57) mit $G(z)$ bezeichnen,

$$\left. \begin{aligned} \frac{G(z)}{1-z} &= \sum_{r=0}^{\infty} z^r \sum_{v=0}^r W_v = W_0 + z(W_0 + W_1) \\ &\quad + z^2(W_0 + W_1 + W_2) + \dots \\ &= \frac{1}{1-z} \sum_{r=0}^{\infty} z^r W_r = \frac{e^{-h(1-z)}}{1-z} [1 + \gamma_2(1-z)^2 + \dots] \end{aligned} \right\} (59).$$

Der Koeffizient von z^r ist nichts anderes als die Summe der Wahrscheinlichkeiten W_0 bis W_r .

Für den besonderen Fall $\gamma_2 \neq 0, \gamma_3 = \gamma_4 = \dots = 0$ wird

$$\frac{G(z)}{1-z} - \frac{F(z)}{1-z} = \gamma_2(1-z) e^{-h(1-z)}$$

$$\sum_{v=0}^r W_v - \sum_{v=0}^r Q_v = \gamma_2(Q_r - Q_{r-1}) = \gamma_2 \frac{Q_r}{h} (h-r)$$

$$\left. \begin{aligned} \sum_{v=0}^r W_v &\geq \sum_{v=0}^r Q_v \quad \text{für } \gamma_2 > 0 \quad \text{und } r \leq h \\ \sum_{v=0}^r W_v &\leq \sum_{v=0}^r Q_v \quad \text{für } \gamma_2 < 0 \quad \text{und } r \geq h \\ \sum_{v=0}^r W_v &= \sum_{v=0}^r Q_v \quad \text{für } r = h \end{aligned} \right\} (60).$$

Positive Exzentrizität bewirkt Vergrößerung der summatorischen Reihen, solange $r < h$ ist; für $r > h$ wird diese Reihe durch positive Exzentrizität verkleinert. *Es findet keine Verschiebung über den Mittelwert statt*, d. h. die Anzahl der Elemente links vom Mittelwert wird durch die Exzentrizität nicht geändert.

Der Quotient $\frac{W_h}{W_{h-1}}$ ist gleich

$$\frac{h^2 + h^2\gamma_2 - 2h^2\gamma_2 + h^2\gamma_2 - h\gamma_2}{h^2 + h^2\gamma_2 - 2h^2\gamma_2 + 2h\gamma_2 + h^2\gamma_2 - 3h\gamma_2 + 2\gamma_2}$$

$$\frac{W_h}{W_{h-1}} = \frac{h(h-\gamma_2)}{h(h-\gamma_2) + 2\gamma_2}$$

Wir haben hier der Einfachheit halber angenommen, h sei eine ganze Zahl.

1. $\gamma_2 < 0$, $\frac{W_h}{W_{h-1}} > 1$, rechts von $r = h - 1$ gibt es noch Werte, die grösser als W_{h-1} sind.
2. $\gamma_2 = 0$, $\frac{W_h}{W_{h-1}} = 1$, $W_h = W_{h-1} = \text{Maximum}$ (Poisson).
3. $\gamma_2 > 0$, $\frac{W_h}{W_{h-1}} < 1$, Maximum ist nach links verschoben.

Positive Exzentrizität verschiebt den wahrscheinlichsten Wert nach links.

2. Fall: γ_2, γ_3 überwiegen. Wir setzen: $\gamma_4 = \gamma_5 = \dots = 0$.

Die dritte Charakteristik, die ich *Verschiebung* nennen will, bewirkt eine Deplacierung der Elemente über den Mittelwert hinaus. Die Anzahl der Elemente links vom Mittelwert wird durch die Verschiebung geändert.

In der Tat erhalten wir aus (59)

$$\frac{G(z)}{1-z} - \frac{F(z)}{1-z} = e^{-h(1-z)} (\gamma_2(1-z) + \gamma_3(1-z)^2)$$

$$\sum_{v=0}^r W_v - \sum_{v=0}^r Q_v = Q_r \left[\gamma_2 - \gamma_2 \frac{r}{h} + \gamma_3 - 2\gamma_3 \frac{r}{h} + \frac{\gamma_3 r(r-1)}{h^2} \right]$$

$$\sum_{v=0}^h W_v - \sum_{v=0}^h Q_v = -\gamma_3 \frac{Q_h}{h}$$

Positive Verschiebung verkleinert die Anzahl der Elemente links vom Mittelwert.

Nun kehren wir zurück zu unsern erzeugenden Funktionen (52) bis (55). In der folgenden Tabelle finden wir die Werte der Exzentrizität und Verschiebung für die entsprechenden Wahrscheinlichkeitsreihen zusammengestellt.

Statistisches Material	$\gamma_2 =$ Exzentrizität	$\gamma_3 =$ Verschiebung	
Unabhängige Ereignisse	homogen	—	
	inhomogen innerhalb Serie	—	
	inhomogen von Serie zu Serie	$\frac{1}{2s} \sum_{\gamma=1}^s (h-h_\gamma)^2$	$\frac{1}{3!s} \sum_{\gamma=1}^s (h-h_\gamma)^3$
Chancenvermehrung	homogen	$\frac{h \cdot d}{2}$	$-\frac{h \cdot d^2}{3}$
	inhomogen innerhalb Serie	$\frac{1}{2} \sum_{\gamma=1}^s h_\gamma d_\gamma$	$-\frac{1}{3} \sum_{\gamma=1}^s h_\gamma d_\gamma^2$
	inhomogen von Serie zu Serie	$\frac{1}{2s} \left[\sum_{\gamma=1}^s (h-h_\gamma)^2 + \sum_{\gamma=1}^s h_\gamma d_\gamma \right]$	$\frac{1}{3!s} \sum_{\gamma=1}^s (h-h_\gamma)^3 - \frac{1}{3s} \sum_{\gamma=1}^s h_\gamma d_\gamma^2$

Bemerkenswert ist an obiger Tabelle, dass bei gleichem Mittelwert und gleicher Exzentrizität die Verschiebung bei homogener Chancenvermehrung kleiner ist als bei innerhalb der Serien inhomogener Chancenvermehrung. Es ist nämlich nach der bekannten Ungleichung

$$\left(\sum_{\gamma=1}^s a_{\gamma} b_{\gamma}\right)^2 < \left(\sum_{\gamma=1}^s a_{\gamma}^2\right) \left(\sum_{\gamma=1}^s b_{\gamma}^2\right)$$

$$\text{für } a_{\gamma} = \sqrt{h_{\gamma}}, \quad b_{\gamma} = \sqrt{h_{\gamma}} d_{\gamma}$$

$$\left(\sum_{\gamma=1}^s h_{\gamma} d_{\gamma}\right)^2 < \sum_{\gamma=1}^s h_{\gamma} \sum_{\gamma=1}^s h_{\gamma} d_{\gamma}^2.$$

Nun ist aber bei gleichem Mittelwert und bei gleicher Exzentrizität

$$h = \sum_{\gamma=1}^s h_{\gamma}, \quad hd = \sum_{\gamma=1}^s h_{\gamma} d_{\gamma},$$

somit

$$hd^2 = \frac{\left(\sum_{\gamma=1}^s h_{\gamma} d_{\gamma}\right)^2}{\sum_{\gamma=1}^s h_{\gamma}} < \sum_{\gamma=1}^s h_{\gamma} d_{\gamma}^2.$$

Unsere Formeln (58) und (60) stimmen mit (48) und (51) genau überein. Bei diesen hatten wir vorausgesetzt, dass $h_{\gamma} \sim h$ sein soll, bei (58) und (60) hingegen ist $\gamma_3 = \gamma_4 = \dots = 0$ gesetzt worden, was dasselbe bedeutet ($[h - h_{\gamma}]^3 \sim 0$).

Die im vorangehenden Paragraphen abgeleiteten *Sätze für Inhomogenität* von Serie zu Serie gelten auch für *kleine Chancenvermehrung* $\left(\gamma_3 = -\frac{hd^2}{3} \sim 0\right)$.

Bei Chancenvermehrung macht sich die Inhomogenität innerhalb der Serien nicht bemerkbar, solange sich diese nur über die mathematischen Erwartungen erstreckt.

Aus der Tabelle sehen wir, dass sich in diesem Fall weder die Exzentrizität noch die Verschiebung ändern, damit ist unser Satz für kleine Chancenvermehrung bewiesen. Die allgemeine Gültigkeit folgt aber aus

$$\begin{aligned} E^*(z) &= \prod_{\gamma=1}^s (1+d_{\gamma}(1-z))^{-\frac{h_{\gamma}}{d_{\gamma}}} = \prod_{\gamma=1}^s (1+d(1-z))^{-\frac{h_{\gamma}}{d}} \\ &= (1+d(1-z))^{-\frac{\sum_{\gamma=1}^s h_{\gamma}}{d}} = (1+d(1-z))^{-\frac{h}{d}} = E(z). \end{aligned}$$

Inhomogenität und Chancenvermehrung üben, wenn sie in *kleinem Masse* vorhanden sind, auf den statistischen, wahrnehmbaren Verlauf der Häufigkeiten dieselbe Wirkung aus.

Sind in der Tat d^2 und $(h-h_{\gamma})^3$ verschwindend kleine Grössen, so haben wir sowohl bei Inhomogenität wie auch bei Chancenvermehrung eine verschwindende Verschiebung; die höheren Charakteristiken sind gleich Null, während die Exzentrizität

$$\left(\frac{1}{2s} \sum_{\gamma=1}^s (h-h_{\gamma})^2 \text{ bzw. } \frac{hd}{2} \right) \text{ positiv ist.}$$

Insbesondere erhalten wir aus (60) für kleine Chancenvermehrung

$$\sum_{\nu=0}^r P_{\nu} \cong \sum_{\nu=0}^r Q_{\nu}, \quad \text{für } r \cong h.$$

Dass der Zeichenumschlag in Spalte VIII der Zahlentafeln 1, 3, 4, 5 und 6 immer unmittelbar nach dem Maximum der Spalte III eintritt, wird hierdurch in Analogie zu (51) bis zu einem gewissen Grade erklärt.

Die theoretisch berechnete Wahrscheinlichkeit des Nichteintreffens des seltenen Ereignisses ist zu klein, wenn vorhandene positive Chancenvermehrung vernachlässigt wird. Es ist in der Tat

$$1 + d < 1 + d + \frac{d^2}{2!} + \frac{d^3}{3!} + \dots = e^d$$

$$(1+d)^{\frac{h}{d}} < e^h$$

$$(1+d)^{-\frac{h}{d}} > e^{-h},$$

somit $P_0 > Q_0$ für $d > 0$.

Diese Beziehung finden wir bei allen unseren Beispielen beim Vergleich des Anfangs der Spalten II und III bestätigt.

§ 17. Berechnung der Erwartungen.

In diesem Paragraphen wollen wir die verschiedenen Formeln der mathematischen Erwartung des Produktes $\{r(r-1)(r-2)\dots(r-k+1)\}$ für die Wahrscheinlichkeiten Q, Q^*, \bar{Q}, P, P^* und \bar{P} ermitteln. Wir benützen dazu die erzeugenden Funktionen dieser Wahrscheinlichkeiten, ähnlich wie in § 12, unter Anwendung der Gleichung

$$\{r(r-1)(r-2)\dots(r-k+1)\} = \left[\frac{d^k}{dz^k} G(z) \right]_{\text{für } z=1}$$

1. *Im Fall der unabhängigen seltenen Ereignisse* ist

$$\begin{aligned} \{r(r-1)(r-2)\dots(r-k+1)\} &= \left[\frac{d^k}{dz^k} F(z) \right]_{\text{für } z=1} \\ &= h^k \text{ [vgl. Formel (26)]} \end{aligned} \quad (61).$$

2. Da sich die erzeugende Funktion der Wahrscheinlichkeiten für *seltene Ereignisse mit Inhomogenität innerhalb der Serien* von denjenigen für gewöhnliche seltene Ereignisse nicht unterscheidet, gilt Formel (61) auch für diesen Fall.
3. *Bei Inhomogenität von Serie zu Serie ohne Chancenvermehrung* ist

$$\bar{F}(z) = \frac{1}{s} e^{-\sum_{\gamma=1}^s h_{\gamma}(1-z)}$$

$$\{r(r-1)(r-2)\dots(r-k+1)\} = \frac{1}{s} \sum_{\gamma=1}^s h_{\gamma}^k \quad (62).$$

4. Nach Formel (26) erhalten wir für *homogene Chancenvermehrung*

$$\{r(r-1)(r-2)\dots(r-k+1)\} = h(h+d)(h+2d)\dots(h+[k-1]d)$$

5. Nun betrachten wir den *Fall der Inhomogenität innerhalb der Serien bei Chancenvermehrung* und setzen voraus, dass die Ansteckung nicht über die Teilserien hinausgeht, mit andern Worten, dass die Teilserien voneinander unabhängig sind.

Nach (42) ist

$$E^*(z) = \prod_{\gamma=1}^s (1+d_{\gamma}(1-z))^{-\frac{h_{\gamma}}{d_{\gamma}}}.$$

Setzen wir, soweit keine Verwechslung möglich, $E^*(z) = E^*$, dann ist $E^*(1) = 1$

$$\{r\} = \left[\frac{d E^*}{d z} \right]_{z=1} = \sum_{\gamma=1}^s h_{\gamma} = h \quad (63)$$

$$\frac{E^{*''}}{E^*} - \left(\frac{E^{*'}}{E^*}\right)^2 = \sum_{\gamma=1}^s \frac{h_\gamma d_\gamma}{(1+d_\gamma(1-z))^2}$$

$$\{r(r-1)\} = \sum_{\gamma=1}^s h_\gamma d_\gamma + h^2 \quad (63')$$

$$\frac{E^{*'''}}{E^*} - 3 \frac{E^{*''} E^{*'}}{E^{*2}} + 2 \frac{(E^{*'})^3}{E^{*3}} = \sum_{\gamma=1}^s \frac{2 h_\gamma d_\gamma^2}{(1+d_\gamma(1-z))^3}$$

$$\{r(r-1)(r-2)\} - 3 \{r(r-1)\} \{r\} + 2 \{r\}^3$$

$$= 2 \sum_{\gamma=1}^s h_\gamma d_\gamma^2$$

$$\{r(r-1)(r-2)\} = 3h \sum_{\gamma=1}^s h_\gamma d_\gamma + h^3 + 2 \sum_{\gamma=1}^s h_\gamma d_\gamma^2 \quad (63'')$$

6. Bei der Inhomogenität von Serie zu Serie mit Chancenvermehrung setzen wir voraus, dass die einzelnen Serien nicht zusammenhängen. Es ist nach (44)

$$\bar{E}(z) = \frac{1}{s} \sum_{\gamma=1}^s (1+d_\gamma(1-z))^{-\frac{h_\gamma}{d_\gamma}}$$

$$\bar{E}^{(k)}(z) = \frac{1}{s} \sum_{\gamma=1}^s h_\gamma (h_\gamma + d_\gamma) \dots$$

$$\dots (h_\gamma + (k-1)d_\gamma) (1+d_\gamma(1-z))^{-\frac{h_\gamma}{d_\gamma} - k}$$

und daraus ist

$$\{r(r-1)(r-2) \dots (r-k+1)\}$$

$$= \frac{1}{s} \sum_{\gamma=1}^s h_\gamma (h_\gamma + d_\gamma) (h_\gamma + 2d_\gamma) \dots (h_\gamma + (k-1)d_\gamma) \quad (64).$$

7. Bei Wahrscheinlichkeiten, die sich nach Charlier'schen Reihen entwickeln lassen, ist nach (56), unter Berücksichtigung, dass $\gamma_0 = 1$, $\gamma_1 = 0$ ist,

$$G(z) = e^{-h(1-z)} [1 + \gamma_2(1-z)^2 + \gamma_3(1-z)^3 + \dots]$$

$$= e^{-h(1-z)} \sum_{\nu=0}^{\infty} \gamma_{\nu} (1-z)^{\nu}$$

$$G'(z) = h e^{-h(1-z)} \sum_{\nu=0}^{\infty} \gamma_{\nu} (1-z)^{\nu}$$

$$- e^{-h(1-z)} \sum_{\nu=1}^{\infty} \nu \gamma_{\nu} (1-z)^{\nu-1}$$

$$G''(z) = h^2 e^{-h(1-z)} \sum_{\nu=0}^{\infty} \gamma_{\nu} (1-z)^{\nu}$$

$$- 2h e^{-h(1-z)} \sum_{\nu=1}^{\infty} \nu \gamma_{\nu} (1-z)^{\nu-1}$$

$$+ e^{-h(1-z)} \sum_{\nu=2}^{\infty} \nu(\nu-1) \gamma_{\nu} (1-z)^{\nu-2}$$

$$G'''(z) = h^3 e^{-h(1-z)} \sum_{\nu=0}^{\infty} \gamma_{\nu} (1-z)^{\nu}$$

$$- 3h^2 e^{-h(1-z)} \sum_{\nu=1}^{\infty} \nu \gamma_{\nu} (1-z)^{\nu-1}$$

$$+ 3h e^{-h(1-z)} \sum_{\nu=2}^{\infty} \nu(\nu-1) \gamma_{\nu} (1-z)^{\nu-2}$$

$$- e^{-h(1-z)} \sum_{\nu=3}^{\infty} \nu(\nu-1)(\nu-2) \gamma_{\nu} (1-z)^{\nu-3},$$

daraus ist

$$\begin{aligned} \{r\} &= h, \quad \{r(r-1)\} = h^2 + 2\gamma_2, \quad \{r(r-1)(r-2)\} \\ &= h^3 + 6h\gamma_2 - 6\gamma_3 \end{aligned} \quad (65)$$

$$s^2 = h + 2\gamma_2$$

$$\{(r-h)^3\} = h + 6\gamma_2 - 6\gamma_3.$$

Unter Benützung der im vorhergehenden Paragraphen aufgestellten Tabelle für die Exzentrizität und Verschiebung könnten wir die Werte der ersten drei Momente in (61), (62), (26), (63) und (64) aus (65) direkt erhalten.

§ 18. Statistische Unterscheidung zwischen Wahrscheinlichkeitsansteckung und Inhomogenität. Die Serienverbindung.

Im Paragraphen 16 haben wir gesehen, dass Chancenvermehrung und Inhomogenität, sofern beide in kleinem Masse auftreten, an dem Verlauf der Häufigkeiten nicht unterschieden werden können. Wir wissen ferner, dass sowohl bei Chancenvermehrung, wie auch bei Inhomogenität übernormale Dispersion auftritt. Wir konstatieren also, dass Inhomogenität und Chancenvermehrung in gewisser Hinsicht statistisch gleich wahrnehmbar sind.

Es ist daher berechtigt zu fragen, ob Inhomogenität und Chancenvermehrung überhaupt denselben Effekt haben. Mit andern Worten: Sind zwei statistische Materialien, die dieselbe Erwartung und dasselbe Schwankungsmass aufweisen, beobachtbar voneinander verschieden, wenn wir von dem einen Material Inhomogenität von Serie zu Serie voraussetzen, beim andern aber Chancenvermehrung?

Zur Veranschaulichung der Verhältnisse habe ich Zahlentafel 12 berechnet. Spalte I gibt die Anzahl der seltenen unter $n(n=\infty)$ Ereignissen an. In Spalte II, III und IV wurden die Wahrscheinlichkeiten Q , Q^* und \bar{Q} für seltene Ereignisse ohne Chancenvermehrung berechnet, während wir in Spalte V, VI und VII die Wahrscheinlichkeiten P , P^* und \bar{P} mit Chancenvermehrung finden. Die Parameter der Wahrscheinlichkeiten wurden durchweg so gewählt, dass normale Dispersion vorhanden ist, oder, sofern dies das Wahrscheinlichkeitsschema nicht zulässt, wurde der Lexische Quotient auf $\sqrt{3.5}$ festgesetzt. Mit andern Worten: Für alle in Zahlentafel 12 berechneten Wahrscheinlichkeiten beträgt die mathematische Erwartung 10, das Quadrat des Schwankungsmasses 10 (für die Wahrscheinlichkeiten der Spalten II und III) bzw. 35 (für die Wahrscheinlichkeiten der Spalten IV bis VII). Der Einfachheit halber habe ich nur zwei inhomogene Serien bzw. Teilsereien vorausgesetzt.

Vergleichen wir nun Spalte IV (Inhomogenität von Serie zu Serie ohne Chancenvermehrung) mit Spalte V (homogene Chancenvermehrung), so sehen wir, dass diese trotz der Übereinstimmung sowohl in der mathematischen Erwartung wie auch im Quadrat des Schwankungsmasses bei weitem nicht identisch sind. Noch klarer tritt dies beim Studium der Figur 4 in die Augen, in welcher Zahlentafel 12 graphisch dargestellt ist. Mit dieser Figur ist der Beweis erbracht, dass die Wahrscheinlichkeitskurven der Inhomogenität und Chancenvermehrung wohl in ähnlicher Weise vom homogenen unabhängigen Fall abweichen, unter sich aber verschieden sind.

Um ein statistisch wahrnehmbares Unterscheidungskriterium zwischen Inhomogenität und Chancen-

vermehrung zu erhalten, studieren wir das *Verhalten des Schwankungsmasses bei Serienverlängerung*. Wie wir im § 13 gesehen haben (vgl. auch § 7), wird eine Serienverlängerung statistisch durch die Verbindung von Serien erreicht. Bei dieser Serienverbindung kann man unterscheiden zwischen *Vereinigung von berührenden* Serien (kontinuierliche Verlängerung) und *nichtberührenden* Serien (unterbrochene Verlängerung), je nachdem zeitlich aufeinanderfolgende oder auseinanderliegende Serien verbunden werden.

Nach (62) erhalten wir für Inhomogenität von Serie zu Serie

$$\begin{aligned} \varsigma^2 &= \frac{1}{s} \sum_{\gamma=1}^s h_{\gamma}^2 + \frac{1}{s} \sum_{\gamma=1}^s h_{\gamma} - \left(\frac{1}{s} \sum_{\gamma=1}^s h_{\gamma} \right)^2 \\ &= \frac{1}{s} \sum_{\gamma=1}^s h_{\gamma}^2 + h - h^2 = \sum_{\gamma=1}^s \frac{(h - h_{\gamma})^2}{s} + h \\ \varsigma^2 &= h \left(1 + \frac{\sum_{\gamma=1}^s (h - h_{\gamma})^2}{hs} \right) \end{aligned} \quad (66)$$

$$\varsigma^2 = h(1 + \lambda^2) \quad (67).$$

λ nennen wir mit L. v. Bortkiewicz den relativen Fehlerexzedenten. Aus 66 ist ersichtlich (vgl. Bortkiewicz 1, Seite 33), dass das Quadrat des relativen Fehlerexzedenten direkt proportional zur Serienlänge ist. Vermehrt man die Versuche in sämtlichen Serien um das k -fache, ohne dass sich dabei die Grundwahrscheinlichkeiten in den einzelnen Serien ändern, so

vergrössern sich alle h_γ -Werte um das k -fache, und somit wird auch h k -mal grösser. Wir erhalten daher aus (66) für das Quadrat des Schwankungsmasses bei dieser Verlängerung der Serien um das k -fache

$$s_k^2 = kh(1 + k\lambda^2) = kh + k^2h\lambda^2 \quad (68).$$

Das Quadrat des Schwankungsmasses ist also eine quadratische Funktion der Serienlänge. Bei Inhomogenität hat die Serienverlängerung statistisch so zu erfolgen, dass man nur gleichartige Serien vereinigt. (68) ist sowohl für die *Vereinigung von berührenden* wie *nichtberührenden* Serien gültig.

Für das Verhalten bei Serienverlängerung gilt daher nach (39), (40) und (68) :

Das Quadrat des Schwankungsmasses ist eine lineare Funktion der Serienlänge, sofern es sich um Chancenvermehrung und nichtzusammenhängende Serien oder um Chancenvermehrung und zusammenhängende Serien mit unterbrochener Verlängerung handelt.

Das Quadrat des Schwankungsmasses ist eine quadratische Funktion der Serienlänge bei Chancenvermehrung mit zusammenhängenden Serien, sofern kontinuierliche Verlängerung stattfindet, sowie bei Inhomogenität von Serie zu Serie, gleichgültig, ob man kontinuierliche oder unterbrochene Verlängerung vornimmt.

Konstatiert man also bei übernormaler Dispersion, dass das Quadrat des Schwankungsmasses eine lineare Funktion der Serienlänge ist, so kann nicht reine Inhomogenität, wohl aber Chancenvermehrung ohne Serienzusammenhang vorliegen. Ist hingegen das Quadrat des Schwankungsmasses eine quadratische Funktion

der Serienlänge, so ist weder die Inhomogenität noch die Chancenvermehrung mit Serienzusammenhang ausgeschlossen; je nachdem bei unterbrochener Verlängerung dieser quadratische Charakter bleibt oder durch den linearen ersetzt wird, wird die Chancenvermehrung oder die Inhomogenität unmöglich.

Nach dem vorliegenden ist es also ausgeschlossen, nach den Untersuchungen des § 13 bei unserem Pockenbeispiel (vgl. Zahlentafel 4) bloss Inhomogenität oder Chancenvermehrung ohne Serienzusammenhang anzunehmen.

Zum Schluss wollen wir vorliegende Erörterungen noch an unserem Beispiel der weiblichen Selbstmorde in 8 deutschen Staaten diskutieren. Es ist

$$\{r\} = h = 3.47$$

$$s^2 = 6.43 = h(1 + 0.85) = h(1 + a).$$

Nun vereinigen wir je zwei aufeinanderfolgende Serien, mit andern Worten: statt der Jahresergebnisse studieren wir das Resultat in zwei aufeinanderfolgenden Jahren, wir erhalten 56 Serien und berechnen

$$h_2 = 6.95,$$

$$\begin{aligned} s_2^2 &= 18.24 = h_2(1 + 1.63) \sim 2h + (2)^2 ha \\ &= \text{Quadratische Funktion von } 2. \end{aligned}$$

Reine Chancenvermehrung ohne Serienzusammenhang ist somit ausgeschlossen. Nun vereinigen wir das Beobachtungsergebnis eines bestimmten Jahres in einem Staate mit den um 7 Jahre entfernten desselben Staates

und erhalten so wieder 56 Doppeljahreserien. Wir notieren z. B. für das Doppeljahr 1881/1888 in Lübeck 5 Selbstmorde. Wir erhalten

$$h_2 = 6.95,$$

$$\begin{aligned} \varsigma_2^2 &= 17.91 = h_2 (1 + 1.58) \sim 2h + (2)^2 ha \\ &= \text{Quadratische Funktion von } 2. \end{aligned}$$

Das Resultat der Berechnungen macht also bei diesem Beispiel die einfache Chancenvermehrung mit und ohne zusammenhängende Serien unmöglich, während kein zwingender Grund vorliegt, anzunehmen, es handle sich hier nicht um reine Inhomogenität.

Zahlentafel 1.

Die Todesfälle bei Dampffässerexplosionen in Preussen in den Jahren 1890 bis 1909.

Vgl. Seite 58.

I	II	III	IV	V	VI	VII	VIII	IX
Zahl der Todesfälle	Zahl der Jahre			Summenzahlen			Abweichung	
	tatsächlich	nach Gleichung (21)	nach Gleichung (20)	zu Spalte II	zu Spalte III	zu Spalte IV	zwischen Spalte V und Spalte VI	Spalte V und Spalte VII
0	2	0.1	2.5	2	0.1	2.5	— 1.9	+ 0.5
1	4	0.6	<u>2.6</u>	6	0.7	5.1	— 5.3	— 0.9
2	1	1.6	<u>2.4</u>	7	2.3	7.5	— 4.7	+ 0.5
3	2	2.8	2.1	9	5.1	9.6	— 3.9	+ 0.6
4	3	3.5	1.8	12	8.6	11.4	— 3.4	— 0.6
5	1	3.5	1.5	13	12.1	12.9	— 0.9	— 0.1
6	2	3.0	1.3	15	15.1	14.2	+ 0.1	— 0.8
7	1	2.1	1.0	16	17.2	15.2	+ 1.2	— 0.8
8	1	1.3	0.9	17	18.5	16.1	+ 1.5	— 0.9
9	—	0.8	0.7	17	19.3	16.8	+ 2.3	— 0.2
10	—	0.4	0.6	17	19.7	17.4	+ 2.7	+ 0.4
11	—	0.2	0.5	17	19.9	17.9	+ 2.9	+ 0.9
12	1	0.1	0.4	18	20.0	18.3	+ 2.0	+ 0.3
13	—	—	0.3	18	20.0	18.6	+ 2.0	+ 0.6
14	—	—	0.3	18	20.0	18.9	+ 2.0	+ 0.9
15	1	—	0.2	19	20.0	19.1	+ 1.0	+ 0.1
16	—	—	0.2	19	20.0	19.3	+ 1.0	+ 0.3
17	—	—	0.1	19	20.0	19.4	+ 1.0	+ 0.4
18	1	—	0.1	20	20.0	19.5	—	— 0.5
19	—	—	0.1	20	20.0	19.6	—	— 0.4
*20	—	—	0.1	20	20.0	19.7	—	— 0.3
21	—	—	0.1	20	20.0	19.8	—	— 0.2

- I: Anzahl der Todesfälle in einem Jahr = r .
- II: Anzahl der Jahre, in denen effektiv r Todesfälle aufgetreten = M_r .
- III: Anzahl der Jahre mit r Todesfällen, theoretisch, bei Annahme der Unabhängigkeit = $20 Q_r$.
- IV: Anzahl der Jahre mit r Todesfällen, theoretisch, bei Annahme der Chancenvermehrung durch Erfolg = $20 P_r$.
- V: $\sum_{r=0}^r M_r$.
- VI: $20 \sum_{r=0}^r Q_r$.
- VII: $20 \sum_{r=0}^r P_r$.
- VIII: $\sum_{r=0}^r (20 Q_r - M_r)$; Abweichung von Spalten VI und V.
- IX: $\sum_{r=0}^r (20 P_r - M_r)$; Abweichung von Spalten VII und V.

Das Maximalglied ist in Spalten III und IV durch Fettdruck hervorgehoben.

Quelle: Stat. Jahrbuch für den Preussischen Staat 1910. VIII. Jahrgang.

Durchschnittliche Anzahl der Todesfälle pro Jahr = $h = 5.05$.

Quadrat des Schwankungsmasses $s^2 = 24.576$, $d = 3.867$.

Durchschnittliche Abweichung: $\{ |r-h| \}$ beobachtet: 3.67, berechnet: 3.66 [nach Formel (31)].

Zahlentafel 2.

Die Dampffässerexplosionen in Preussen in den Jahren 1890 bis 1909.

Vgl. Seite 60.

I Zahl der Explosionen	II		III		VI Abweichung zwischen Spalte IV und V
	Zahl der Jahre		Summenzahlen		
	tatsächlich	nach Gleichung (21)	zu Spalte II	zu Spalte III	
0	1	0.7	1	0.7	— 0.3
1	2	2.4	3	3.1	+ 0.1
2	3	4.0	6	7.1	+ 1.1
3	3	4.4	9	11.5	+ 2.5
4	7	3.6	16	15.1	— 0.9
5	3	2.4	19	17.5	— 1.5
6	1	1.3	20	18.8	— 1.2
7	—	0.6	20	19.4	— 0.6
8	—	0.3	20	19.7	— 0.3
9	—	0.1	20	19.8	— 0.2

I: Anzahl der Explosionen in einem Jahr = r .

II: Anzahl der Jahre, in denen effektiv r Explosionen aufgetreten = M_r .

III: Anzahl der Jahre mit r Explosionen theoretisch, bei Annahme der Unabhängigkeit = $20 Q_r$.

$$\text{IV: } \sum_{r=0}^r M_r.$$

$$\text{V: } 20 \sum_{r=0}^r Q_r.$$

$$\text{VI: } \sum_{r=0}^r (20 Q_r - M_r); \text{ Abweichung von Spalten IV und V.}$$

Das Maximalglied ist in Spalte III durch Fettdruck hervorgehoben.

Quelle: Vgl. Zahlentafel 1.

Durchschnittliche Anzahl der Explosionen pro Jahr = $h = 3.3$.

Quadrat des Schwankungsmasses = 3.

Zahlentafel 3.

Die Todesfälle bei Dampfkesselexplosionen in Preussen in den Jahren 1883 bis 1907.

Vgl. Seite 61.

I	II	III	IV	V	VI	VII	VIII	IX
Zahl der Todesfälle	Zahl der Jahre			Summenzahlen			Abweichung	
	tatsächlich	nach Gleichung (21)	nach Gleichung (20)	zu Spalte II	zu Spalte III	zu Spalte IV	zwischen Spalte V und Spalte VI	zwischen Spalte V und Spalte VII
0	—	—	0.1	—	—	0.1	—	+ 0.1
1	—	—	0.2	—	—	0.3	—	+ 0.3
2	1	—	0.4	1	—	0.7	— 1.0	— 0.3
3	1	—	0.6	2	—	1.3	— 2.0	— 0.7
4	—	—	0.8	2	—	2.1	— 2.0	+ 0.1
5	1	0.1	1.0	3	0.1	3.1	— 2.9	+ 0.1
6	1	0.3	1.2	4	0.4	4.3	— 3.6	+ 0.3
7	1	0.5	1.3	5	0.9	5.6	— 4.1	+ 0.6
8	3	0.8	1.4	8	1.7	7.0	— 6.3	— 1.0
9	1	1.3	1.4	9	3.0	8.4	— 6.0	— 0.6
10	1	1.8	1.4	10	4.8	9.8	— 5.2	— 0.2
11	1	2.2	1.4	11	7.0	11.2	— 4.0	+ 0.2
12	—	2.5	1.4	11	9.5	12.6	— 1.5	+ 1.6
13	3	2.7	1.3	14	12.2	13.9	— 1.8	— 0.1
14	1	2.7	1.2	15	14.9	15.1	— 0.1	+ 0.1
15	1	2.4	1.1	16	17.3	16.2	+ 1.3	+ 0.2
16	1	2.1	1.0	17	19.4	17.2	+ 2.4	+ 0.2
17	1	1.7	0.9	18	21.1	18.1	+ 3.1	+ 0.1
18	1	1.3	0.9	19	22.4	19.0	+ 3.4	—
19	1	1.0	0.8	20	23.4	19.8	+ 3.4	— 0.2
20	2	0.6	0.7	22	24.0	20.5	+ 2.0	— 1.5
21	—	0.4	0.6	22	24.4	21.1	+ 2.4	— 0.9
22	—	0.3	0.5	22	24.7	21.6	+ 2.7	— 0.4
23	—	0.2	0.5	22	24.9	22.1	+ 2.9	+ 0.1
24	—	0.1	0.4	22	25.0	22.5	+ 3.0	+ 0.5

I	II	III	IV	V	VI	VII	VIII	IX
Zahl der Todesfälle	Zahl der Jahre			Summenzahlen			Abweichung	
	tatsächlich	nach Gleichung (21)	nach Gleichung (20)	zu Spalte II	zu Spalte III	zu Spalte IV	zwischen Spalte VI und Spalte VII	Spalte V und Spalte VII
25	—	—	0.4	22	25.0	22.9	+ 3.0	+ 0.9
26	1	—	0.3	23	25.0	23.2	+ 2.0	+ 0.2
27	1	—	0.3	24	25.0	23.5	+ 1.0	— 0.5
28	—	—	0.2	24	25.0	23.7	+ 1.0	— 0.3
29	—	—	0.2	24	25.0	23.9	+ 1.0	— 0.1
30	—	—	0.2	24	25.0	24.1	+ 1.0	+ 0.1
31	—	—	0.1	24	25.0	24.2	+ 1.0	+ 0.2
32	—	—	0.1	24	25.0	24.3	+ 1.0	+ 0.3
33	—	—	0.1	24	25.0	24.4	+ 1.0	+ 0.4
34	—	—	0.1	24	25.0	24.5	+ 1.0	+ 0.5
35	1	—	0.1	25	25.0	24.6	—	— 0.4
36	—	—	0.1	25	25.0	24.7	—	— 0.3
37	—	—	0.1	25	25.0	24.8	—	— 0.2

I: Anzahl der Todesfälle in einem Jahr = r .

II: Anzahl der Jahre, in denen effektiv r Todesfälle aufgetreten = M_r .

III: Anzahl der Jahre mit r Todesfällen, theoretisch, bei Annahme der Unabhängigkeit = $25 Q_r$.

IV: Anzahl der Jahre mit r Todesfällen, theoretisch, bei Annahme der Chancenvermehrung durch Erfolg = $25 P_r$.

V: $\sum_{v=0}^r M_v$.

VIII: $\sum_{v=0}^r (25 Q_v - M_v)$; Abweichung von Spalten VI und V.

VI: $25 \sum_{v=0}^r Q_v$.

IX: $\sum_{v=0}^r (25 P_v - M_v)$; Abweichung von Spalten VII und V.

VII: $25 \sum_{v=0}^r P_v$.

Das Maximalglied ist in Spalten III und IV durch Fettdruck hervorgehoben.

Quelle: Stat. Jahrbuch für das Deutsche Reich 38 ff.

Durchschnittliche Anzahl der Todesfälle pro Jahr = $h = 13.72$.

Quadrat des Schwankungsmasses = 62.48, $d = 3.55$.

Durchschnittliche Abweichung: $\{ |r-h| \}$ beobachtet: 6.086, berechnet (theoretisch) 6.226 [nach Formel (31)].

Zahlentafel 4.

Die Todesfälle an Pocken in der Schweiz in den Jahren 1877 bis 1900.

Vgl. Seite 61.

I	II	III	IV	V	VI	VII	VIII	IX
Zahl der Todesfälle	Zahl der Monate			Summenzahlen			Abweichung	
	tatsächlich	nach Gleichung (21)	nach Gleichung (20)	zu Spalte II	zu Spalte III	zu Spalte IV	zwischen Spalte V und Spalte VI	Spalte V und Spalte VII
0	100	1.2	100.4	100	1.2	100.4	— 98.8	+ 0.4
1	39	6.5	36.3	139	7.7	136.7	— 131.3	— 2.3
2	28	17.8	23.5	167	25.5	160.2	— 141.5	— 6.8
3	26	32.6	17.5	193	58.1	177.7	— 134.9	— 15.3
4	13	44.9	13.8	206	103.0	191.5	— 103.0	— 14.5
5	6	49.4	11.3	212	152.4	202.8	— 59.6	— 9.2
6	11	45.2	9.5	223	197.6	212.3	— 25.4	— 10.7
7	5	35.5	8.1	228	233.1	220.4	+ 5.1	— 7.6
8	5	24.5	7.0	233	257.6	227.4	+ 24.6	— 5.6
9	6	15.0	6.1	239	272.6	233.5	+ 33.6	— 5.5
10	1	8.2	5.3	240	280.8	238.8	+ 40.8	— 1.2
11	6	4.1	4.7	246	284.9	243.5	+ 38.9	— 2.5
12	2	1.9	4.2	248	286.8	247.7	+ 38.8	— 0.3
13	2	0.8	3.7	250	287.6	251.4	+ 37.6	+ 1.4
14	3	0.3	3.3	253	287.9	254.7	+ 34.9	+ 1.7
15	3	0.1	3.0	256	288.0	257.7	+ 32.0	+ 1.7
16	—	—	2.7	256	288.0	260.4	+ 32.0	+ 4.4
17	—	—	2.4	256	288.0	262.8	+ 32.0	+ 6.8
18	4	—	2.2	260	288.0	265.0	+ 28.0	+ 5.0
19	1	—	2.0	261	288.0	267.0	+ 27.0	+ 6.0
20	2	—	1.8	263	288.0	268.8	+ 25.0	+ 5.8
21	4	—	1.6	267	288.0	270.4	+ 21.0	+ 3.4
22	1	—	1.5	268	288.0	271.9	+ 20.0	+ 3.9
23	3	—	1.3	271	288.0	273.2	+ 17.0	+ 2.2
24	—	—	1.2	271	288.0	274.4	+ 17.0	+ 3.4

I	II	III	IV	V	VI	VII	VIII	IX
Zahl der Todesfälle	Zahl der Monate			Summenzahlen			Abweichung	
	tatsächlich	nach Gleichung (21)	nach Gleichung (20)	zu Spalte II	zu Spalte III	zu Spalte IV	zwischen Spalte VI und Spalte VII	Spalte V und Spalte VII
25	2	—	1.1	273	288.0	275.5	+ 15.0	+ 2.5
26	1	—	1.0	274	288.0	276.5	+ 14.0	+ 2.5
27	1	—	0.9	275	288.0	277.4	+ 13.0	+ 2.4
28	1	—	0.8	276	288.0	278.2	+ 12.0	+ 2.2
29	—	—	0.8	276	288.0	279.0	+ 12.0	+ 3.0
30	3	—	0.7	279	288.0	279.7	+ 9.0	+ 0.7
31	—	—	0.6	279	288.0	280.3	+ 9.0	+ 1.3
32	1	—	0.6	280	288.0	280.9	+ 8.0	+ 0.9
33	—	—	0.5	280	288.0	281.4	+ 8.0	+ 1.4
34	1	—	0.5	281	288.0	281.9	+ 7.0	+ 0.9
35	1	—	0.5	282	288.0	282.4	+ 6.0	+ 0.4
36	—	—	0.4	282	288.0	282.8	+ 6.0	+ 0.8
37	—	—	0.4	282	288.0	283.2	+ 6.0	+ 1.2
38	1	—	0.4	283	288.0	283.6	+ 5.0	+ 0.6
39	—	—	0.3	283	288.0	283.9	+ 5.0	+ 0.9
40	—	—	0.3	283	288.0	284.2	+ 5.0	+ 1.2
41	—	—	0.3	283	288.0	284.5	+ 5.0	+ 1.5
42	—	—	0.3	283	288.0	284.8	+ 5.0	+ 1.8
43	1	—	0.2	284	288.0	285.0	+ 4.0	+ 1.0
44	1	—	0.2	285	288.0	285.2	+ 3.0	+ 0.2
45	—	—	0.2	285	288.0	285.4	+ 3.0	+ 0.4
46	—	—	0.2	285	288.0	285.6	+ 3.0	+ 0.6
47	—	—	0.2	285	288.0	285.8	+ 3.0	+ 0.8
48	—	—	0.2	285	288.0	286.0	+ 3.0	+ 1.0
49	—	—	0.1	285	288.0	286.1	+ 3.0	+ 1.1
50	—	—	0.1	285	288.0	286.2	+ 3.0	+ 1.2
51	—	—	0.1	285	288.0	286.3	+ 3.0	+ 1.3
52	—	—	0.1	285	288.0	286.4	+ 3.0	+ 1.4
53	—	—	0.1	285	288.0	286.5	+ 3.0	+ 1.5
54	2	—	0.1	287	288.0	286.6	+ 1.0	— 0.4

I	II	III	IV	V	VI	VII	VIII	IX
Zahl der Todesfälle	Zahl der Monate			Summenzahlen			Abweichung	
	tatsächlich	nach Gleichung (21)	nach Gleichung (20)	zu Spalte II	zu Spalte III	zu Spalte IV	zwischen Spalte V und Spalte VI	zwischen Spalte V und Spalte VII
55	—	—	0.1	287	288.0	286.7	+ 1.0	— 0.3
56	—	—	0.1	287	288.0	286.8	+ 1.0	— 0.2
57	—	—	0.1	287	288.0	286.9	+ 1.0	— 0.1
58	—	—	0.1	287	288.0	287.0	+ 1.0	—
59	—	—	0.1	287	288.0	287.1	+ 1.0	+ 0.1
60	1	—	0.1	288	288.0	287.2	—	— 0.8
61	—	—	0.1	288	288.0	287.3	—	— 0.7
62	—	—	0.1	288	288.0	287.4	—	— 0.6

I: Anzahl der Todesfälle in einem Monat = r .

II: Anzahl der Monate, in denen effektiv r Todesfälle aufgetreten = M_r .

III: Anzahl der Monate mit r Todesfällen, theoretisch, bei Annahme der Unabhängigkeit = $288 Q_r$.

IV: Anzahl der Monate mit r Todesfällen, theoretisch, bei Annahme der Chancenvermehrung durch Erfolg = $288 P_r$.

$$V: \sum_{r=0}^r M_r$$

$$VI: 288 \sum_{r=0}^r Q_r$$

$$VII: 288 \sum_{r=0}^r P_r$$

$$VIII: \sum_{r=0}^r (288 Q_r - M_r); \text{ Abweichung von Spalten VI und V.}$$

$$IX: \sum_{r=0}^r (288 P_r - M_r); \text{ Abweichung von Spalten VII und V.}$$

Das Maximalglied ist in Spalten III und IV durch Fettdruck hervorgehoben.

Quelle: Schweizerische Statistik, Ehe, Geburt und Tod in der schweizerischen Bevölkerung. Fünfter Teil. Seite 131.

Durchschnittliche Anzahl der Todesfälle pro Monat = $h = 5.5$.

Quadrat des Schwankungsmasses = 83.589, $d = 14.20$.

Durchschnittliche Abweichung: $\{ |r-h| \}$ beobachtet: 6.326, berechnet: 6.003.

Zahlentafel 5.

Die Todesfälle an Scharlach in der Schweiz in den Jahren 1877 bis 1900.

Vgl. Seite 61.

I	II	III	IV	V	VI	VII	VIII	IX
Zahl der Todes- fälle	Zahl der Monate			Summenzahlen			Abweichung	
	tatsäch- lich	nach Gleichung (21)	nach Gleichung (20)	zu Spalte II	zu Spalte III	zu Spalte IV	zwischen Spalte V und Spalte VI	Spalte V und Spalte VII
0	5	—	36.5	5	—	36.5	— 5.0	+ 31.5
1	7	—	18.1	12	—	54.6	— 12.0	+ 42.6
2	13	—	13.4	25	—	68.0	— 25.0	+ 43.0
3	15	—	11.0	40	—	79.0	— 40.0	+ 39.0
4	18	—	9.5	58	—	88.5	— 58.0	+ 30.5
5	19	—	8.4	77	—	96.9	— 77.0	+ 19.9
6	8	—	7.6	85	—	104.5	— 85.0	+ 19.5
7	10	—	6.9	95	—	111.4	— 95.0	+ 16.4
8	7	—	6.4	102	—	117.8	— 102.0	+ 15.8
9	8	—	5.9	110	—	123.7	— 110.0	+ 13.7
10	2	—	5.5	112	—	129.2	— 112.0	+ 17.2
11	6	—	5.2	118	—	134.4	— 118.0	+ 16.4
12	11	—	4.9	129	—	139.3	— 129.0	+ 10.3
13	6	0.1	4.6	135	0.1	143.9	— 134.9	+ 8.9
14	7	0.2	4.4	142	0.3	148.3	— 141.7	+ 6.3
15	8	0.3	4.2	150	0.6	152.5	— 149.4	+ 2.5
16	9	0.6	4.0	159	1.2	156.5	— 157.8	— 2.5
17	7	1.1	3.8	166	2.3	160.3	— 163.7	— 5.7
18	2	1.8	3.6	168	4.1	163.9	— 163.9	— 4.1
19	5	2.8	3.5	173	6.9	167.4	— 166.1	— 5.6
20	1	4.2	3.3	174	11.1	170.7	— 162.9	— 3.3
21	4	5.9	3.2	178	17.0	173.9	— 161.0	— 4.1
22	8	8.0	3.1	186	25.0	177.0	— 161.0	— 9.0
23	4	10.3	3.0	190	35.3	180.0	— 154.7	— 10.0
24	4	12.8	2.9	194	48.1	182.9	— 145.9	— 11.1

I	II	III	IV	V	VI	VII	VIII	IX
Zahl der Todesfälle	Zahl der Monate			Summenzahlen			Abweichung	
	tatsächlich	nach Gleichung (21)	nach Gleichung (20)	zu Spalte II	zu Spalte III	zu Spalte IV	zwischen Spalte VI und Spalte VII	Spalte V und Spalte VII
25	5	15.3	2.8	199	63.4	185.7	— 135.6	— 13.3
26	6	17.5	2.7	205	80.9	188.4	— 124.1	— 16.4
27	3	19.3	2.6	208	100.2	191.0	— 107.8	— 17.0
28	2	20.5	2.5	210	120.7	193.5	— 89.3	— 16.5
29	3	21.1	2.4	213	141.8	195.9	— 71.2	— 17.1
30	2	20.9	2.3	215	162.7	198.2	— 52.3	— 16.8
31	2	20.1	2.2	217	182.8	200.4	— 34.2	— 16.6
32	1	18.7	2.2	218	201.5	202.6	— 16.5	— 15.4
33	3	16.9	2.1	221	218.4	204.7	— 2.6	— 16.3
34	—	14.8	2.0	221	233.2	206.7	+ 11.2	— 14.3
35	1	12.6	2.0	222	245.8	208.7	+ 23.8	— 13.3
36	3	10.4	1.9	225	256.2	210.6	+ 31.2	— 14.4
37	1	8.4	1.9	226	264.6	212.5	+ 38.6	— 13.5
38	1	6.6	1.8	227	271.2	214.3	+ 44.2	— 12.7
39	2	5.0	1.8	229	276.2	216.1	+ 47.2	— 12.9
40	3	3.7	1.7	232	279.9	217.8	+ 47.9	— 14.2
41	—	2.7	1.7	232	282.6	219.5	+ 50.6	— 12.5
42	—	1.9	1.6	232	284.5	221.1	+ 52.5	— 10.9
43	1	1.3	1.6	233	285.8	222.7	+ 52.8	— 10.3
44	1	0.9	1.5	234	286.7	224.2	+ 52.7	— 9.8
45	—	0.6	1.5	234	287.3	225.7	+ 53.3	— 8.3
46	—	0.4	1.4	234	287.7	227.1	+ 53.7	— 6.9
47	—	0.2	1.4	234	287.9	228.5	+ 53.9	— 5.5
48	—	0.1	1.4	234	288.0	229.9	+ 54.0	— 4.1
49	2	—	1.3	236	288.0	231.2	+ 52.0	— 4.8
50	2	—	1.3	238	288.0	232.5	+ 50.0	— 5.5
51	1	—	1.3	239	288.0	233.8	+ 49.0	— 5.2
52	1	—	1.2	240	288.0	235.0	+ 48.0	— 5.0
53	—	—	1.2	240	288.0	236.2	+ 48.0	— 3.8
54	1	—	1.2	241	288.0	237.4	+ 47.0	— 3.6

I	II	III	IV	V	VI	VII	VIII	IX
Zahl der Todesfälle	Zahl der Monate			Summenzahlen			Abweichung	
	tatsächlich	nach Gleichung (21)	nach Gleichung (20)	zu Spalte II	zu Spalte III	zu Spalte IV	zwischen Spalte V und Spalte VI	zwischen Spalte V und Spalte VII
55	2	—	1.1	243	288.0	238.5	+ 45.0	— 4.5
56	1	—	1.1	244	288.0	239.6	+ 44.0	— 4.4
57	1	—	1.1	245	288.0	240.7	+ 43.0	— 4.3
58	2	—	1.0	247	288.0	241.7	+ 41.0	— 5.3
59	—	—	1.0	247	288.0	242.7	+ 41.0	— 4.3
60	1	—	1.0	248	288.0	243.7	+ 40.0	— 4.3
61	1	—	1.0	249	288.0	244.7	+ 39.0	— 4.3
62	—	—	0.9	249	288.0	245.6	+ 39.0	— 3.4
63	1	—	0.9	250	288.0	246.5	+ 38.0	— 3.5
64	—	—	0.9	250	288.0	247.4	+ 38.0	— 2.6
65	1	—	0.9	251	288.0	248.3	+ 37.0	— 2.7
66	—	—	0.9	251	288.0	249.2	+ 37.0	— 1.8
67	2	—	0.8	253	288.0	250.0	+ 35.0	— 3.0
68	—	—	0.8	253	288.0	250.8	+ 35.0	— 2.2
69	—	—	0.8	253	288.0	251.6	+ 35.0	— 1.4
70	—	—	0.8	253	288.0	252.4	+ 35.0	— 0.6
71	1	—	0.8	254	288.0	253.2	+ 34.0	— 0.8
72	—	—	0.7	254	288.0	253.9	+ 34.0	— 0.1
73	—	—	0.7	254	288.0	254.6	+ 34.0	+ 0.6
74	3	—	0.7	257	288.0	255.3	+ 31.0	— 1.7
75	1	—	0.7	258	288.0	256.0	+ 30.0	— 2.0
76	1	—	0.7	259	288.0	256.7	+ 29.0	— 2.3
77	—	—	0.7	259	288.0	257.4	+ 29.0	— 1.6
78	2	—	0.6	261	288.0	258.0	+ 27.0	— 3.0
79	—	—	0.6	261	288.0	258.6	+ 27.0	— 2.4
80	1	—	0.6	262	288.0	259.2	+ 26.0	— 2.8
81	—	—	0.6	262	288.0	259.8	+ 26.0	— 2.2
82	1	—	0.6	263	288.0	260.4	+ 25.0	— 2.6
83	—	—	0.6	263	288.0	261.0	+ 25.0	— 2.0
84	2	—	0.6	265	288.0	261.6	+ 23.0	— 3.4

I	II	III	IV	V	VI	VII	VIII	IX
Zahl der Todesfälle	Zahl der Monate			Summenzahlen			Abweichung	
	tatsächlich	nach Gleichung (21)	nach Gleichung (20)	zu Spalte II	zu Spalte III	zu Spalte IV	zwischen Spalte VI und Spalte VII	Spalte V und Spalte VII
85	—	—	0.6	265	288.0	262.2	+ 23.0	— 2.8
86	—	—	0.5	265	288.0	262.7	+ 23.0	— 2.3
87	2	—	0.5	267	288.0	263.2	+ 21.0	— 3.8
88	—	—	0.5	267	288.0	263.7	+ 21.0	— 3.3
89	—	—	0.5	267	288.0	264.2	+ 21.0	— 2.8
90	—	—	0.5	267	288.0	264.7	+ 21.0	— 2.3
91	—	—	0.5	267	288.0	265.2	+ 21.0	— 1.8
92	—	—	0.5	267	288.0	265.7	+ 21.0	— 1.3
93	—	—	0.5	267	288.0	266.2	+ 21.0	— 0.8
94	—	—	0.5	267	288.0	266.7	+ 21.0	— 0.3
95	1	—	0.4	268	288.0	267.1	+ 20.0	— 0.9
96	—	—	0.4	268	288.0	267.5	+ 20.0	— 0.5
97	—	—	0.4	268	288.0	267.9	+ 20.0	— 0.1
98	—	—	0.4	268	288.0	268.3	+ 20.0	+ 0.3
99	—	—	0.4	268	288.0	268.7	+ 20.0	+ 0.7
100	—	—	0.4	268	288.0	269.1	+ 20.0	+ 1.1
101	—	—	0.4	268	288.0	269.5	+ 20.0	+ 1.5
102	—	—	0.4	268	288.0	269.9	+ 20.0	+ 1.9
103	—	—	0.4	268	288.0	270.3	+ 20.0	+ 2.3
104	1	—	0.4	269	288.0	270.7	+ 19.0	+ 1.7
105	1	—	0.4	270	288.0	271.1	+ 18.0	+ 1.1
106	1	—	0.3	271	288.0	271.4	+ 17.0	+ 0.4
107	—	—	0.3	271	288.0	271.7	+ 17.0	+ 0.7
108	—	—	0.3	271	288.0	272.0	+ 17.0	+ 1.0
109	—	—	0.3	271	288.0	272.3	+ 17.0	+ 1.3
110	—	—	0.3	271	288.0	272.6	+ 17.0	+ 1.6
111	3	—	0.3	274	288.0	272.9	+ 14.0	— 1.1
112	—	—	0.3	274	288.0	273.2	+ 14.0	— 0.8
113	—	—	0.3	274	288.0	273.5	+ 14.0	— 0.5
114	1	—	0.3	275	288.0	273.8	+ 13.0	— 1.2

I	II	III	IV	V	VI	VII	VIII	IX
Zahl der Todesfälle	Zahl der Monate			Summenzahlen			Abweichung	
	tatsächlich	nach Gleichung (21)	nach Gleichung (20)	zu Spalte II	zu Spalte III	zu Spalte IV	zwischen Spalte V und Spalte VI	Spalte V und Spalte VII
115	—	—	0.3	275	288.0	274.1	+ 13.0	— 0.9
116	—	—	0.3	275	288.0	274.4	+ 13.0	— 0.6
117	—	—	0.3	275	288.0	274.7	+ 13.0	— 0.3
118	—	—	0.3	275	288.0	275.0	+ 13.0	—
119	—	—	0.3	275	288.0	275.3	+ 13.0	+ 0.3
120	—	—	0.3	275	288.0	275.6	+ 13.0	+ 0.6
121	—	—	0.3	275	288.0	275.9	+ 13.0	+ 0.9
122	—	—	0.2	275	288.0	276.1	+ 13.0	+ 1.1
123	—	—	0.2	275	288.0	276.3	+ 13.0	+ 1.3
124	—	—	0.2	275	288.0	276.5	+ 13.0	+ 1.5
125	—	—	0.2	275	288.0	276.7	+ 13.0	+ 1.7
126	—	—	0.2	275	288.0	276.9	+ 13.0	+ 1.9
127	—	—	0.2	275	288.0	277.1	+ 13.0	+ 2.1
128	—	—	0.2	275	288.0	277.3	+ 13.0	+ 2.3
129	—	—	0.2	275	288.0	277.5	+ 13.0	+ 2.5
130	—	—	0.2	275	288.0	277.7	+ 13.0	+ 2.7
131	—	—	0.2	275	288.0	277.9	+ 13.0	+ 2.9
132	—	—	0.2	275	288.0	278.1	+ 13.0	+ 3.1
133	—	—	0.2	275	288.0	278.3	+ 13.0	+ 3.3
134	—	—	0.2	275	288.0	278.5	+ 13.0	+ 3.5
135	—	—	0.2	275	288.0	278.7	+ 13.0	+ 3.7
136	—	—	0.2	275	288.0	278.9	+ 13.0	+ 3.9
137	—	—	0.2	275	288.0	279.1	+ 13.0	+ 4.1
138	1	—	0.2	276	288.0	279.3	+ 12.0	+ 3.3
139	—	—	0.2	276	288.0	279.5	+ 12.0	+ 3.5
140	—	—	0.2	276	288.0	279.7	+ 12.0	+ 3.7
141	1	—	0.2	277	288.0	279.9	+ 11.0	+ 2.9

- I: Anzahl der Todesfälle in einem Monat = r .
- II: Anzahl der Monate, in denen effektiv r Todesfälle aufgetreten = M_r .
- III: Anzahl der Monate mit r Todesfällen, theoretisch, bei Annahme der Unabhängigkeit = $288 Q_r$.
- IV: Anzahl der Monate mit r Todesfällen, theoretisch, bei Annahme der Chancenvermehrung durch Erfolg = $288 P_r$.
- V: $\sum_{r=0}^r M_r$.
- VI: $288 \sum_{r=0}^r Q_r$.
- VII: $288 \sum_{r=0}^r P_r$.
- VIII: $\sum_{r=0}^r (288 Q_r - M_r)$; Abweichung von Spalten VI und V.
- IX: $\sum_{r=0}^r (288 P_r - M_r)$; Abweichung von Spalten VII und V.

Das Maximalglied ist in Spalten III und IV durch Fettdruck hervorgehoben.

Quelle: Schweizerische Statistik. Ehe, Geburt und Tod in der schweizerischen Bevölkerung. Fünfter Teil, S. 137.

Durchschnittliche Anzahl der Todesfälle pro Monat = $h = 29.778$.

Quadrat des Schwankungsmasses = 1788.125 , $d = 59.049$.

Durchschnittliche Abweichung: $\{ |r-h| \}$ beobachtet: 27.574 , berechnet: 29.037
[nach Formel (31)].

Zahlentafel 6.

Die Todesfälle an Scharlach in der Schweiz in den Monaten des 2. Halbjahres
(Juli bis Dezember) der Jahre 1879 bis 1900.

Vgl. Seite 62.

I	II	III	IV	V	VI	VII	VIII	IX
Zahl der Todes- fälle	Zahl der Monate			Summenzahlen			Abweichung	
	tatsäch- lich	nach Gleichung (21)	nach Gleichung (20)	zu Spalte II	zu Spalte III	zu Spalte IV	zwischen Spalte V und Spalte VI	Spalte V und Spalte VII
0	4	—	4.8	4	—	4.8	— 4	+ 0.8
1	4	—	6.4	8	—	11.2	— 8	— 3.2
2	9	0.1	7.2	17	0.1	18.4	— 16.9	+ 1.4
3	10	0.3	7.5	27	0.4	25.9	— 26.6	— 1.1
4	8	0.7	7.6	35	1.1	33.5	— 33.9	— 1.5
5	10	1.8	7.4	45	2.9	40.9	— 42.1	— 4.1
6	6	3.5	7.1	51	6.4	48.0	— 44.6	— 3.0
7	5	6.0	6.8	56	12.4	54.8	— 43.6	— 1.2
8	6	8.9	6.4	62	21.3	61.2	— 40.7	— 0.8
9	5	11.8	6.0	67	33.1	67.2	— 33.9	+ 0.2
10	2	14.1	5.6	69	47.2	72.8	— 21.8	+ 3.8
11	5	15.2	5.2	74	62.4	78.0	— 11.6	+ 4.0
12	8	15.1	4.8	82	77.5	82.8	— 4.5	+ 0.8
13	5	13.8	4.5	87	91.3	87.3	+ 4.3	+ 0.3
14	5	11.7	4.1	92	103.0	91.4	+ 11.0	— 0.6
15	4	9.3	3.8	96	112.3	95.2	+ 16.3	— 0.8
16	4	6.9	3.4	100	119.2	98.6	+ 19.2	— 1.4
17	4	4.8	3.2	104	124.0	101.8	+ 20.0	— 2.2
18	1	3.2	2.9	105	127.2	104.7	+ 22.2	— 0.3
19	3	2.0	2.6	108	129.2	107.3	+ 21.2	— 0.7
20	—	1.2	2.4	108	130.4	109.7	+ 22.4	+ 1.7
21	1	0.7	2.2	109	131.1	111.9	+ 22.1	+ 2.9
22	4	0.4	2.0	113	131.5	113.9	+ 18.5	+ 0.9
23	1	0.2	1.8	114	131.7	115.7	+ 17.7	+ 1.7
24	4	0.2	1.6	118	131.9	117.3	+ 13.9	— 0.7

I	II	III	IV	V	VI	VII	VIII	IX
Zahl der Todesfälle	Zahl der Monate			Summenzahlen			Abweichung	
	tatsächlich	nach Gleichung (21)	nach Gleichung (20)	zu Spalte II	zu Spalte III	zu Spalte IV	zwischen Spalte V und Spalte VI	und Spalte VII
25	2	0.1	1.5	120	132	118.8	+ 12	— 1.2
26	2	—	1.3	122	132	120.1	+ 10	— 1.9
27	1	—	1.2	123	132	121.3	+ 9	— 1.7
28	1	—	1.1	124	132	122.4	+ 8	— 1.6
29	1	—	1.0	125	132	123.4	+ 7	— 1.6
30	1	—	0.9	126	132	124.3	+ 6	— 1.7
31	—	—	0.8	126	132	125.1	+ 6	— 0.9
32	—	—	0.7	126	132	125.8	+ 6	— 0.2
33	—	—	0.6	126	132	126.4	+ 6	+ 0.4
34	—	—	0.6	126	132	127.0	+ 6	+ 1.0
35	—	—	0.5	126	132	127.5	+ 6	+ 1.5
36	2	—	0.5	128	132	128.0	+ 4	—
37	—	—	0.4	128	132	128.4	+ 4	+ 0.4
38	1	—	0.4	129	132	128.8	+ 3	— 0.2
39	—	—	0.3	129	132	129.1	+ 3	+ 0.1
40	1	—	0.3	130	132	129.4	+ 2	— 0.6
41	—	—	0.3	130	132	129.7	+ 2	— 0.3
42	—	—	0.2	130	132	129.9	+ 2	— 0.1
43	—	—	0.2	130	132	130.1	+ 2	+ 0.1
44	—	—	0.2	130	132	130.3	+ 2	+ 0.3
45	—	—	0.2	130	132	130.5	+ 2	+ 0.5
46	—	—	0.2	130	132	130.7	+ 2	+ 0.7
47	—	—	0.1	130	132	130.8	+ 2	+ 0.8
48	—	—	0.1	130	132	130.9	+ 2	+ 0.9
49	1	—	0.1	131	132	131.0	+ 1	—
50	—	—	0.1	131	132	131.1	+ 1	+ 0.1
51	—	—	0.1	131	132	131.2	+ 1	+ 0.2
52	—	—	0.1	131	132	131.3	+ 1	+ 0.3
53	—	—	0.1	131	132	131.4	+ 1	+ 0.4
54	—	—	0.1	131	132	131.5	+ 1	+ 0.5

I	II	III	IV	V	VI	VII	VIII	IX
Zahl der Todesfälle	Zahl der Monate			Summenzahlen			Abweichung	
	tatsächlich	nach Gleichung (21)	nach Gleichung (20)	zu Spalte II	zu Spalte III	zu Spalte IV	zwischen Spalte VI und Spalte VII	Spalte V und Spalte VII
55	—	—	0.1	131	132	131.6	+ 1	+ 0.6
56	—	—	0.1	131	132	131.7	+ 1	+ 0.7
57	1	—	0.1	132	132	131.8	—	— 0.2

I: Anzahl der Todesfälle in einem Monat = r .

II: Anzahl der Monate, in denen effektiv r Todesfälle aufgetreten = M_r .

III: Anzahl der Monate mit r Todesfällen, theoretisch, bei Annahme der Unabhängigkeit = $132 Q_r$.

IV: Anzahl der Monate mit r Todesfällen, theoretisch, bei Annahme der Chancenvermehrung durch Erfolg = $132 P_r$.

$$V: \sum_{r=0}^r M_r.$$

$$VI: 132 \sum_{r=0}^r Q_r.$$

$$VII: 132 \sum_{r=0}^r P_r.$$

$$VIII: \sum_{r=0}^r (132 Q_r - M_r); \text{ Abweichung von Spalten VI und V.}$$

$$IX: \sum_{r=0}^r (132 P_r - M_r); \text{ Abweichung von Spalten VII und V.}$$

Das Maximalglied ist in Spalten III und IV durch Fettdruck hervorgehoben.

Quelle: Vgl. Zahlentafel 5.

Durchschnittliche Anzahl der Todesfälle pro Monat = $h = 11.90$.

Quadrat des Schwankungsmasses = 104.48 , $d = 7.78$.

Durchschnittliche Abweichung: $\{ |r-h| \}$ beobachtet: 7.691 , berechnet: 7.680

[nach Formel (31)].

Zahlentafel 7.

Das Quadrat des Schwankungsmasses bei Serienvereinigung in der Statistik der Pockentodesfälle in den Jahren 1877 bis 1900. (Vgl. Zahlentafel 4.)

Vgl. Seite 76.

I	II	III	IV	V	VI	VII	VIII
Zahl der Serien	Monate in einer Serie	Mittlere Anzahl der Todesfälle pro Serie	Quadrat des Schwankungsmasses			Ansteckung	
			beobachtet	berechnet, wenn Serienzusammenhang		ermittelt, wenn Serienzusammenhang	
				nicht angenommen	angenommen	nicht angenommen	angenommen
288	1	5.5	83.59 ± 20.61	83.59	83.59	14.20	14.20
144	2	11.0	356.84 ± 112.12	167.18	323.40	31.44	15.72
96	3	16.5	770.63 ± 307.11	250.76	719.40	45.71	15.23
72	4	22.0	1287.94 ± 541.13	334.35	1271.60	57.54	14.38
48	6	33.0	3007.40	501.53	2844.60	90.13	15.02
24	12	66.0	9209.04	1003.07	11253.00	138.53	11.54
12	24	132.0	30341.45	2006.13	34080.00	228.86	9.54

I: Anzahl der beobachteten Serien = $\frac{288}{k}$.

II: Länge der einzelnen Serien in Monaten = k .

III: Durchschnittliche Anzahl der Todesfälle pro Serie = h_k .

IV: Quadrat des Schwankungsmasses für Serien mit k Monaten = s_k^2 .

V: Quadrat des Schwankungsmasses, theoretisch, bei Annahme der nicht zusammenhängenden Serien = $h_k(1+d_k) = kh(1+d)$.

VI: Quadrat des Schwankungsmasses, theoretisch, bei Annahme der zusammenhängenden Serien = $h_k(1+d_k) = kh(1+kd)$.

VII: d ermittelt aus der Beobachtung (IV) bei Annahme der nicht zusammenhängenden Serien, $d = \frac{s_k^2}{kh} - 1$, nach Formel (39).

VIII: d ermittelt aus der Beobachtung (IV) bei Annahme der zusammenhängenden Serien, $d = \left(\frac{s_k^2}{kh} - 1\right) \frac{1}{k}$, nach Formel (40).

Bemerkung: Je nachdem die Annahme der nicht zusammenhängenden oder zusammenhängenden Serien richtig ist, muss Spalte VII oder VIII konstant sein (theoretisch).

Quelle: Vgl. Zahlentafel 4.

Zahlentafel 8.

Das Quadrat des Schwankungsmasses bei Serienvereinigung in der Statistik der Scharlachtodesfälle in den Jahren 1877 bis 1900. (Vgl. Zahlentafel 5.)

Vgl. Seite 76.

I	II	III	IV	V	VI	VII	VIII
Zahl der Serien	Monate in einer Serie	Mittlere Anzahl der Todesfälle pro Serie	Quadrat des Schwankungsmasses			Ansteckung	
			beobachtet	berechnet, wenn Serienzusammenhang		ermittelt, wenn Serienzusammenhang	
				nicht angenommen	angenommen	nicht angenommen	angenommen
288	1	29.778	1788.13	1788.13	1788.13	59.05	59.05
144	2	59.556	6740.00	3576.26	7093.00	112.17	56.09
96	3	89.333	15360.22	5364.39	15914.41	170.94	56.98
72	4	119.111	26733.43	7152.52	28252.65	223.44	55.86
48	6	178.667	61503.47	10728.78	63479.31	343.24	57.21
24	12	357.333	203487.26	21457.56	253559.21	568.46	47.37
12	24	714.667	831814.81	42915.12	1013523.59	1162.92	48.46

I: Anzahl der beobachteten Serien = $\frac{288}{k}$.

II: Länge der einzelnen Serien in Monaten = k .

III: Durchschnittliche Anzahl der Todesfälle pro Serie = h_k .

IV: Quadrat des Schwankungsmasses für Serien mit k Monaten = s_k^2 .

V: Quadrat des Schwankungsmasses, theoretisch, bei Annahme der nicht zusammenhängenden Serien = $h_k(1+d_k) = kh(1+d)$.

VI: Quadrat des Schwankungsmasses, theoretisch, bei Annahme der zusammenhängenden Serien = $h_k(1+d_k) = kh(1+kd)$.

VII: d ermittelt aus der Beobachtung (IV) bei Annahme der nicht zusammenhängenden

• Serien, $d = \frac{s_k^2}{kh} - 1$, nach Formel (39).

VIII: d ermittelt aus der Beobachtung (IV) bei Annahme der zusammenhängenden

Serien, $d = \left(\frac{s_k^2}{kh} - 1 \right) \frac{1}{k}$, nach Formel (40).

Bemerkung: Je nachdem die Annahme der nicht zusammenhängenden oder zusammenhängenden Serien richtig ist, muss Spalte VII oder VIII konstant sein (theoretisch).

Quelle: Vgl. Zahlentafel 5.

Zahlentafel 9.

Das Quadrat des Schwankungsmasses bei Serienvereinigung in der Statistik der Scharlach Todesfälle in den Monaten des 2. Halbjahres (Juli bis Dezember) der Jahre 1879 bis 1900. (Vgl. Zahlentafel 6.)

Vgl. Seite 77.

I	II	III	IV	V	VI	VII	VIII
Zahl der Serien	Monate in einer Serie	Mittlere Anzahl der Todesfälle pro Serie	Quadrat des Schwankungsmasses			Ansteckung	
			beobachtet	berechnet, wenn Serienzusammenhang		ermittelt, wenn Serienzusammenhang	
				nicht angenommen	angenommen	nicht angenommen	angenommen
132	1	11.90	104.48	104.48	104.48	7.78	7.78
66	2	23.80	344.37	208.96	394.13	13.47	6.73
44	3	35.70	702.96	313.44	868.94	18.69	6.23
33	4	47.60	1173.40	417.92	1528.91	23.65	5.91
22	6	71.40	2667.80	626.88	3404.35	36.36	6.06
11	12	142.80	10665.28	1253.76	13474.61	73.69	6.14

I: Anzahl der beobachteten Serien = $\frac{132}{k}$.

II: Länge der einzelnen Serien in Monaten = k .

III: Durchschnittliche Anzahl der Todesfälle pro Serie = h_k .

VI: Quadrat des Schwankungsmasses für Serien mit k Monaten = s_k^2 .

V: Quadrat des Schwankungsmasses, theoretisch, bei Annahme der nicht zusammenhängenden Serien = $h_k(1+d_k) = kh(1+d)$.

VI: Quadrat des Schwankungsmasses, theoretisch, bei Annahme der zusammenhängenden Serien = $h_k(1+d_k) = kh(1+kd)$.

VII: d ermittelt aus der Beobachtung (IV) bei Annahme der nicht zusammenhängenden Serien, $d = \frac{s_k^2}{kh} - 1$, nach Formel (39).

VIII: d ermittelt aus der Beobachtung (IV) bei Annahme der zusammenhängenden Serien, $d = \left(\frac{s_k^2}{kh} - 1\right) \frac{1}{k}$, nach Formel (40).

Bemerkung: Je nachdem die Annahme der nicht zusammenhängenden oder zusammenhängenden Serien richtig ist, muss Spalte VII oder VIII konstant sein (theoretisch).

Quelle: Vgl. Zahlentafel 4.

Zahlentafel 10.

Statistik der weiblichen Selbstmorde in 8 deutschen Staaten
in den Jahren 1881 bis 1894.

Vgl. Seite 94.

I	II	III	IV	V	VI	VII	VIII	IX
Zahl der Todes- fälle	Zahl der Jahre			Summenzahlen			Abweichung	
	tatsäch- lich	nach Gleichung (21)	nach Gleichung (43)	zu Spalte II	zu Spalte III	zu Spalte IV	zwischen Spalte V und Spalte VI	und Spalte VII
0	9	3.5	8.0	9	3.5	8.0	— 5.5	— 1.0
1	19	12.1	16.9	28	15.6	24.9	— 12.4	— 3.1
2	17	20.9	20.3	45	36.5	45.2	— 8.5	+ 0.2
3	20	24.2	18.7	65	60.7	63.9	— 4.3	— 1.1
4	15	21.0	15.1	80	81.7	79.0	+ 1.7	— 1.0
5	11	14.6	11.5	91	96.3	90.5	+ 5.3	— 0.5
6	8	8.4	8.3	99	104.7	98.8	+ 5.7	— 0.2
7	2	4.2	5.6	101	108.9	104.4	+ 7.9	+ 3.4
8	3	1.8	3.6	104	110.7	108.0	+ 6.7	+ 4.0
9	5	0.7	2.1	109	111.4	110.1	+ 2.4	+ 1.1
10	3	0.5	1.1	112	111.9	111.2	— 0.1	— 0.8
11	—	0.1	0.8	112	112.0	112.0	—	—

I: Anzahl der Todesfälle in einem Jahr ¹⁾ = r .

II: Anzahl der Jahre, in denen effektiv r Todesfälle aufgetreten = M_r .

III: Anzahl der Jahre mit r Todesfällen, theoretisch, bei Annahme der Homogenität und Unabhängigkeit = $112 Q_r$.

IV: Anzahl der Jahre mit r Todesfällen, theoretisch, bei Annahme der Inhomogenität von Serie zu Serie und der Unabhängigkeit = $112 \bar{Q}_r$.

$$V: \sum_{r=0}^{\infty} M_r$$

$$VI: 112 \sum_{r=0}^{\infty} \bar{Q}_r$$

¹⁾ Jahr = Beobachtungsjahr, auf ein Kalenderjahr entfallen entsprechend den 8 Staaten 8 Beobachtungsjahre.

$$\text{VII: } 112 \sum_{\nu=0}^r \bar{Q}_{\nu}$$

$$\text{VIII: } \sum_{\nu=0}^r (112 Q_{\nu} - M_{\nu}); \text{ Abweichung von Spalten VI und V.}$$

$$\text{IX: } \sum_{\nu=0}^r (112 \bar{Q}_{\nu} - M_{\nu}); \text{ Abweichung von Spalten VII und V.}$$

Das Maximalglied ist in Spalten III und IV durch Fettdruck hervorgehoben.

Quelle: Allgemeines Statistisches Archiv. 4. Jahrgang, II. Hbd. 1896. Art. „Der Selbstmord“, von G. v. Mayr, Seite 718. (Vgl. L. v. Bortkiewicz 1: Das Gesetz der kleinen Zahlen, Seite 20.)

Durchschnittliche Anzahl der Todesfälle in allen 8 Staaten zusammen 27.786, pro Staat 3.473 = h .

In Schaumburg-Lippe	1.429
„ Waldeck	2.214
„ Lübeck	2.571
„ Reuss ä. L.	2.643
„ Lippe	2.857
„ Schwarzburg-Rudolstadt	5.143
„ Mecklenburg-Strelitz	5.286
„ Schwarzburg-Sondershausen	5.643

Zahlentafel 11.

Statistik der tödlichen Unfälle bei 11 Berufsgenossenschaften
in den Jahren 1886 bis 1894.

Vgl. Seite 95.

I	II	III	IV	V	VI	VII	VIII	IX
Zahl der Todes- fälle	Zahl der Jahre			Summenzahlen			Abweichung	
	tatsäch- lich	nach Gleichung (21)	nach Gleichung (43)	zu Spalte II	zu Spalte III	zu Spalte IV	zwischen Spalte V und Spalte VI	und Spalte VII
0	5	1.3	3.7	5	1.3	3.7	— 3.7	— 1.3
1	9	5.5	9.6	14	6.8	13.3	— 7.2	— 0.7
2	14	12.0	13.9	28	18.8	27.2	— 9.2	— 0.8
3	13	17.5	15.2	41	36.3	42.4	— 4.7	+ 1.4
4	14	19.0	14.3	55	55.3	56.7	+ 0.3	+ 1.7
5	16	16.6	12.3	71	71.9	69.0	+ 0.9	— 2.0
6	7	12.1	9.8	78	84.0	78.8	+ 6.0	+ 0.8
7	7	7.5	7.3	85	91.5	86.1	+ 6.5	+ 1.1
8	8	4.4	5.1	93	95.9	91.2	+ 2.9	— 1.8
9	2	2.1	3.3	95	98.0	94.5	+ 3.0	— 0.5
10	1	0.9	2.0	96	98.9	96.5	+ 2.9	+ 0.5
11	1	0.1	1.2	97	99.0	97.7	+ 2.0	+ 0.7
12	1	—	0.7	98	99.0	98.4	+ 1.0	+ 0.4
13	—	—	0.3	98	99.0	98.7	+ 1.0	+ 0.7
14	1	—	0.2	99	99.0	98.9	—	— 0.1
15	—	—	0.1	99	99.0	99.0	—	—

I: Anzahl der Todesfälle in einem Jahr ¹⁾ = r .

II: *Anzahl der Jahre, in denen effektiv r Todesfälle aufgetreten = M_r .

III: Anzahl der Jahre mit r Todesfällen, theoretisch, bei Annahme der Homogenität und Unabhängigkeit = $99 Q_r$.

IV: Anzahl der Jahre mit r Todesfällen, theoretisch, bei Annahme der Inhomogenität von Serie zu Serie und der Unabhängigkeit = $99 \bar{Q}_r$.

¹⁾ Jahr = Beobachtungsjahr, entsprechend den 11 Genossenschaften entfallen auf ein Kalenderjahr 11 Beobachtungsjahre.

$$\text{V: } \sum_{\nu=0}^r M_{\nu}.$$

$$\text{VI: } 99 \sum_{\nu=0}^r Q_{\nu}.$$

$$\text{VII: } 99 \sum_{\nu=0}^r \bar{Q}_{\nu}.$$

$$\text{VIII: } \sum_{\nu=0}^r (99 Q_{\nu} - M_{\nu}); \text{ Abweichung von Spalten VI und V.}$$

$$\text{IX: } \sum_{\nu=0}^r (99 \bar{Q}_{\nu} - M_{\nu}); \text{ Abweichung von Spalten VII und V.}$$

Das Maximalglied ist in Spalten III und IV durch Fettdruck hervorgehoben.

Quelle: Statistisches Jahrbuch für das Deutsche Reich. (Vgl. L. v. Bortkiewicz 1: Das Gesetz der kleinen Zahlen, Seite 22.)

Durchschnittliche Anzahl der Todesfälle pro Jahr in allen Berufsgenossenschaften zusammen 48, somit pro Genossenschaft = 4.36 = h .

Die Anzahl der Todesfälle pro Jahr für die Genossenschaften

Nr. 27	ist gleich	1.89	=	h_1
" 12	" "	2.56	=	h_2
" 14	" "	2.56	=	h_3
" 40	" "	2.89	=	h_4
" 20	" "	4.33	=	h_5
" 55	" "	4.33	=	h_6
" 42	" "	4.56	=	h_7
" 41	" "	5.11	=	h_8
" 29	" "	5.89	=	h_9
" 23	" "	6.22	=	h_{10}
" 13	" "	7.67	=	h_{11} .

Zahlentafel 12.

Verlauf der Wahrscheinlichkeiten bei Inhomogenität und Chancenvermehrung.

Vgl. Seite 110, Tabelle Seite 83.

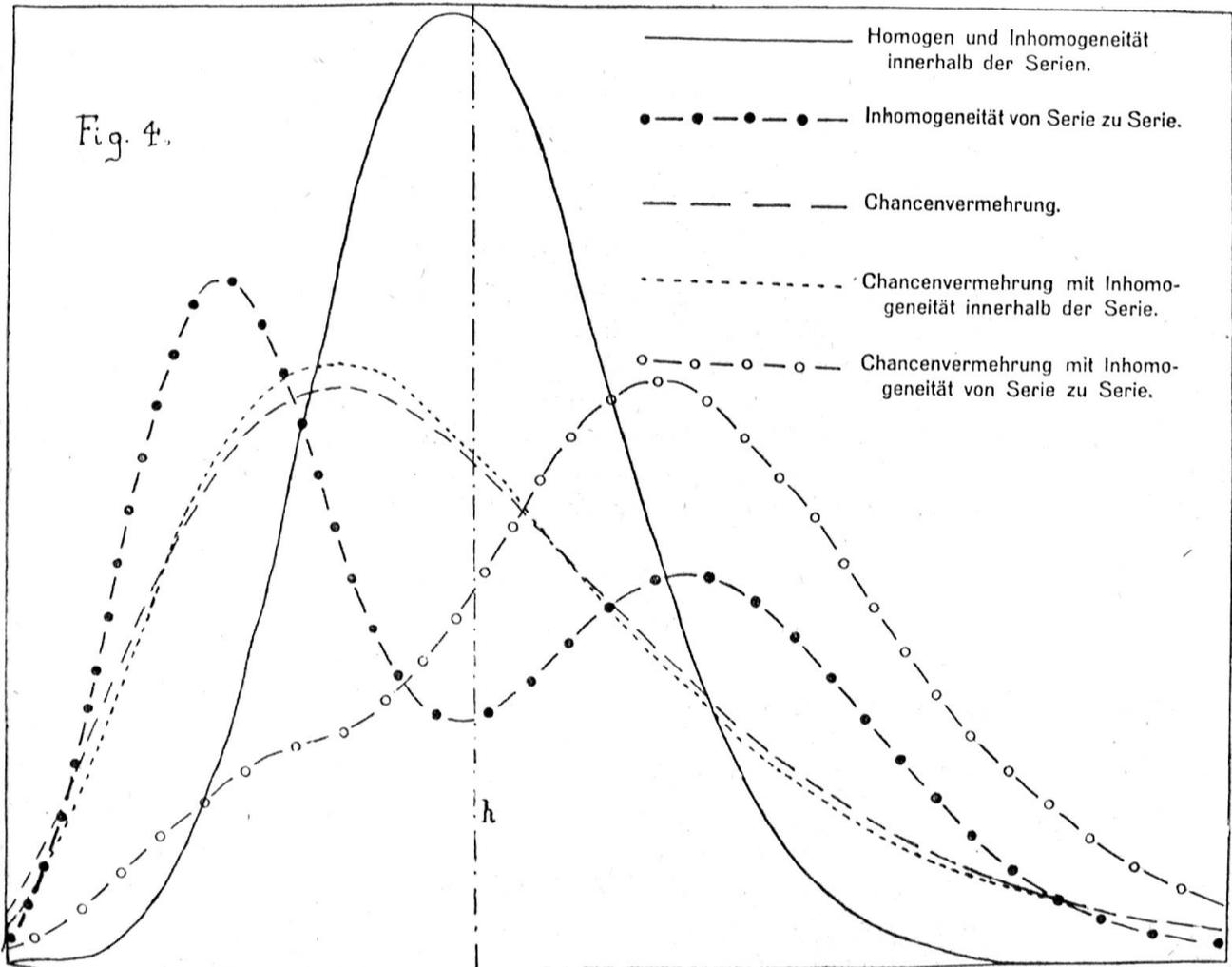
I	II	III	IV	V	VI	VII
Anzahl der seltenen Ereignisse	Mathematische Erwartung = $h = 10$					
	Quadrat des Schwankungs- masses = 10		Quadrat des Schwankungsmasses = 35			
	Q_r	Q_r^*	\bar{Q}_r	P_r	P_r^*	\bar{P}_r
0	—	—	0.003	0.007	0.005	0.002
1	0.001	0.001	0.017	0.019	0.016	0.005
2	0.002	0.002	0.042	0.034	0.031	0.010
3	0.008	0.008	0.070	0.049	0.047	0.016
4	0.019	0.019	0.088	0.061	0.062	0.021
5	0.038	0.038	0.089	0.069	0.072	0.026
6	0.063	0.063	0.076	0.074	0.078	0.029
7	0.090	0.090	0.057	0.076	0.079	0.031
8	0.113	0.113	0.042	0.075	0.078	0.035
9	0.125	0.125	0.034	0.071	0.073	0.041
10	0.125	0.125	0.033	0.066	0.067	0.050
11	0.114	0.114	0.037	0.060	0.060	0.060
12	0.095	0.095	0.043	0.053	0.053	0.069
13	0.073	0.073	0.048	0.047	0.046	0.075
14	0.052	0.052	0.051	0.041	0.040	0.077
15	0.035	0.035	0.051	0.035	0.034	0.074
16	0.022	0.022	0.048	0.030	0.028	0.068
17	0.013	0.013	0.042	0.025	0.024	0.061
18	0.007	0.007	0.035	0.021	0.020	0.052
19	0.004	0.004	0.028	0.017	0.016	0.043
20	0.002	0.002	0.021	0.014	0.013	0.035
21	0.001	0.001	0.015	0.012	0.011	0.028
22	—	—	0.010	0.009	0.009	0.022
23	—	—	0.007	0.007	0.007	0.017
24	—	—	0.004	0.006	0.006	0.013

I	II	III	IV	V	VI	VII
Anzahl der seltenen Ereignisse	Mathematische Erwartung = $h = 10$					
	Quadrat des Schwankungs- masses = 10		Quadrat des Schwankungsmasses = 35			
	Q_r	Q_r^*	\bar{Q}_r	P_r	P_r^*	\bar{P}_r
25	—	—	0.003	0.005	0.005	0.010
26	—	—	0.002	0.004	0.004	0.007
27	—	—	0.001	0.003	0.003	0.006
28	—	—	0.001	0.002	0.002	0.004
29	—	—	—	0.002	0.002	0.003
30	—	—	—	0.002	0.002	0.003
31	—	—	—	0.001	0.001	0.002
32	—	—	—	0.001	0.001	0.001
33	—	—	—	0.001	0.001	0.001
34	—	—	—	0.001	—	0.001
35	—	—	—	—	—	0.001
36	—	—	—	—	—	0.001

- I: Anzahl der seltenen Ereignisse r .
- II: Wahrscheinlichkeiten Q_r nach Poisson $h = 10$, $\sigma^2 = 10$.
- III: Wahrscheinlichkeiten Q_r^* bei Inhomogenität innerhalb der Serien ohne Chancenvermehrung $h = 10$, $\sigma^2 = 10$.
- IV: Wahrscheinlichkeiten \bar{Q}_r bei Inhomogenität von Serie zu Serie ohne Chancenvermehrung. Annahme: 2 Serien mit $h_1 = 5$, $h_2 = 15$, $h = 10$, $\sigma^2 = 35$.
- V: Wahrscheinlichkeiten P_r bei homogener Chancenvermehrung $h = 10$, $d = 2.5$, $\sigma^2 = 35$.
- VI: Wahrscheinlichkeiten P_r^* bei Chancenvermehrung und Inhomogenität innerhalb der Serien. Annahme: 2 homogene Teilsereien $h_1 = 4$, $d_1 = 1$, $h_2 = 6$, $d_2 = 3.5$. Daraus ist $h = 10$, $\sigma^2 = 35$.
- VII: Wahrscheinlichkeiten \bar{P}_r bei Chancenvermehrung und Inhomogenität von Serie zu Serie. Annahme: 2 Serientypen mit $h_1 = 8$, $d_1 = 1$, $h_2 = 12$, $d_2 = 2.833$, daraus ist $h = 10$, $\sigma^2 = 35$.

Das Maximalglied ist in den Spalten II bis VII durch Fettdruck hervorgehoben.

Vgl. Fig. 4.



Inhaltsverzeichnis.

Einleitung.		Seite
§ 1.	Problemstellung	31
§ 2.	Statistisches Material	33
Kapitel I: Theorie der Wahrscheinlichkeitsansteckung.		
§ 3.	Die Struktur der Wahrscheinlichkeitsansteckung	35
§ 4.	Die Wahrscheinlichkeiten	37
§ 5.	Verlauf der Wahrscheinlichkeiten	39
§ 6.	Berechnung der Erwartungen	43
§ 7.	Diskussion des Schwankungsmasses	45
§ 8.	Der Grenzfall gewöhnlicher Ereignisse	48
§ 9.	Der Grenzfall seltener Ereignisse	53
Kapitel II: Seltene Ereignisse.		
§ 10.	Die Wahrscheinlichkeiten	54
§ 11.	Verlauf der Wahrscheinlichkeiten	63
§ 12.	Berechnung von $\{ \nu - h \}$. Die erwartungsmässigen Fehler	66
§ 13.	Diskussion des Schwankungsmasses bei Serienverbindung	74
Kapitel III: Inhomogenität.		
§ 14.	Die Wahrscheinlichkeiten und ihre erzeugenden Funktionen	79
§ 15.	Verlauf der Wahrscheinlichkeiten	85
§ 16.	Fortsetzung. Die Charlierschen Reihen	95
§ 17.	Berechnung der Erwartungen	105
§ 18.	Statistische Unterscheidung zwischen Wahrscheinlichkeitsansteckung und Inhomogenität. Die Serienverbindung	109

Zahlentafeln 1--12.

Literaturverzeichnis.

E. Blaschke: Vorlesungen über mathematische Statistik.

L. v. Bortkiewicz:

1. Das Gesetz der kleinen Zahlen. Leipzig 1898.

2. Homogenität und Stabilität in der Statistik. Upsala 1918.

H. Bruns: Wahrscheinlichkeitsrechnung und Kollektivmasslehre.
Leipzig 1906.

Bulletin des Schweizerischen Gesundheitsamtes 1919/22.

Guido Castelnuovo: Calcolo delle Probabilità. Milano 1919.

C. V. L. Charlier:

1. Vorlesungen über die Grundzüge der mathematischen Statistik.
Hamburg — Lund 1920.

2. Über das Fehlergesetz. Arkiv for Mathematik, Astronomi och
Fysik. II/8.

3. Die zweite Form des Fehlergesetzes. Arkiv for Mathematik,
Astronomi och Fysik. II/15.

4. Researches into the theory of probability. Lund Observatorium
Meddelanden, Serie II, 1906.

Czuber:

1. Wahrscheinlichkeitsrechnung. 3. Auflage, Leipzig 1914.

2. Die statistischen Forschungsmethoden. Wien 1921.

F. Eggenberger: Über die Statistik verketteter Vorgänge (F. Eggenberger
und G. Pólya), Zeitschrift für angewandte Mathematik
und Mechanik, 1923.

A. Markoff: Wahrscheinlichkeitsrechnung. Übersetzt von Liebmann.
Leipzig 1912.

Pearson: Tables for statisticians and biometricians. Cambridge 1914.

G. Pólya:

1. Anschauliche und elementare Darstellung der Lexisschen
Dispersionstheorie. Zeitschrift für schweizerische Statistik
und Volkswirtschaft, 55. Jahrgang, 1919.

2. Über die Statistik verketteter Vorgänge (F. Eggenberger und
G. Pólya).

Udny Yule: An Introduction to the Theory of Statistics. London 1922.

