COMPOSITIO MATHEMATICA

HILDA GEIRINGER

Methoden der theoretischen Statistik

Compositio Mathematica, tome 2 (1935), p. 276-320

<http://www.numdam.org/item?id=CM_1935__2__276_0>

Methoden der theoretischen Statistik

von

Hilda Geiringer

Istanbul

INHALT.

Allen Methoden der theoretischen Statistik ist es gemeinsam, daß sie eine Beziehung herstellen zwischen einer konkreten statistischen Aufnahme und theoretischen Wahrscheinlichkeitskollektivs. Hat die Beobachtung etwa m Zahlen $X_1, X_2, \ldots, X_m$ ergeben (z.B. die Anzahlen der Todesfälle in Holland in $m = 20$ Jahren), so sucht man eine geeignete Annahme zu machen, wonach irgendwelche m Verteilungen vorhanden sind $v_1(x)$, $v_2(x)$, $\ldots$, $v_m(x)$, derart, daß das i-te Ergebnis der i-ten Verteilung entspricht. Die Wahrscheinlichkeit des gleichzeitigen Auftretens von $x_1, x_2, \ldots, x_m$ ist dann $v_1(x_1) \cdot v_2(x_2) \cdot \ldots \cdot v_m(x_m) = w(x_1, x_2, \ldots, x_m)$.

In der Annahme solcher Verteilungen besteht die *Hypothese*, die zwecks Erklärung des empirischen Ergebnisses gemacht wird. Die theoretische Statistik hat die Aufgabe, Kriterien zur Prüfung solcher Hypothesen aufzustellen.

Eine recht allgemeine in dieser Richtung gehende Anweisung formuliert R. v. Mises[1]) etwa so: *Aus den Beobachtungsgrößen* $X_1, X_2, \ldots, X_m$ *bildet man eine geeignete Funktion* $F(X_1, X_2, \ldots, X_m)$ *und vergleicht diese mit ihrem „Erwartungswert"* $\mathfrak{E}(F)$ *innerhalb des m-dimensionalen Kollektivs* $w(x_1, x_2, \ldots, x_m)$ *unter Berücksichtigung der „Streuung"* $\mathfrak{Str}(F)$ *dieser Funktion.* Liegt dann $F(X_1, X_2, \ldots, X_m)$ innerhalb der Grenzen

$$(a) \qquad \mathfrak{E}(F) - \sqrt{\mathfrak{Str}(F)} \text{ und } \mathfrak{E}(F) + \sqrt{\mathfrak{Str}(F)},$$

so spricht das für die gemachte Hypothese hinsichtlich der $v_i(x)$, (wobei natürlich das Zutreffen dieser Tatsache nur „im Durchschnitt" zu erwarten ist).

Eine unmittelbare Illustration zu dieser allgemeinen Anweisung kann man in der sogenannten Pearsonschen χ^2-Methode sehen. Hier wird eine bestimmte quadratische Funktion der Beobachtungen und der angenommenen Verteilungen $v_i(x)$ als Prüffunktion gewählt und diese unter Berücksichtigung der Streuung mit ihrem Erwartungswert verglichen. — Weniger evident ist es, daß auch die sog. Lexis'sche Theorie sich im gleichen Sinne auffassen läßt. Bei dieser besteht der charakteristische Unterschied z.B. gegen die χ^2-Methode darin, daß die hypothetischen Verteilungen $v_i(x)$ *nicht als vollständig bekannt* angesehen werden, wobei es gelingt, ein Kriterium so anzugeben, daß der Vergleich trotz der nur unvollständig bekannten Verteilungen $v_i(x)$ ermöglicht wird.

An diesen eigentlichen Kern der Lexis'schen Theorie, der bei manchen allzu rohen Darstellungen allerdings nicht zu Tage tritt, knüpfen die folgenden Überlegungen an. Es kann nämlich sein, daß eine vernünftige Hypothese zwar hinsichtlich *der Form*

---

[1]) R. v. Mises, Vorlesungen über angewandte Mathematik. Band I: Wahrscheinlichkeitsrechnung. Leipzig-Wien 1932. Neben dieser ist noch eine gewissermaßen duale Auffassung berechtigt, welche sich zu der obengenannten verhält wie der Gedankenkreis des sogenannten „Zweiten Fundamentalsatzes" zu dem des „Ersten Fundamentalsatzes".

der vermuteten Verteilungen $v_i(x)$ gemacht wird, — indem man diese z.B. als Bernoullische Verteilungen vermutet oder sie als Gaußsche Verteilungen voraussetzt — daß aber die Parameter dieser angenommenen Verteilungen zunächst unbekannt sind. Vielleicht sollen diese erst, wenn die Hypothese sich prinzipiell bestätigt hat, hinterher irgendwie ermittelt und geprüft werden! Als ein erstes Beispiel für die strenge Behandlung eines solchen schwierigeren Falles läßt sich, wie gesagt, die richtig gedeutete Lexis'sche Theorie auffassen. Dies — mit einigen Erweiterungen — zeigt § 1. — Es läßt sich aber darüber hinaus in dieser Richtung zu sehr allgemeinen Resultaten kommen, die zu einer systematischen *Methode der theoretischen Statistik* ausgebaut werden, wie dies in § 2 dargestellt und an einfachen Anwendungen erörtert wird. — Da die Berechnung von Erwartungswerten und Streuungen nicht ganz einfacher Funktionen oft mühselig und manchmal unmöglich ist, werden in § 3, ausgehend von dem naheliegenden Gedanken der Entwicklung der in Rede stehenden Funktion in eine Taylorreihe, für diese Erwartungswerte und Streuungen *Näherungsformeln für große m angegeben*, die für besonders allgemeine Fälle „gebrauchsfertige" Formeln bereitstellen. Insbesondere kommt man auf diesem Wege zur näherungsweisen Lösung einer Art von Umkehraufgabe, die im wesentlichen darin besteht, eine Funktion zu ermitteln, die einen bestimmten, formelmäßig gegebenen Erwartungswert besitzt. Die direkte Behandlung dieser Aufgabe, für einen praktisch wichtigen Spezialfall (§ 2, 6), dürfte an sich von Interesse sein. — Mit Hilfe der Formeln des § 3 werden schließlich in § 4 nach der in § 2 entwickelten Methode einige schwierigere Aufgaben behandelt.

§ 1. Zur Systematik der theoretischen Statistik.

1. *Bezeichnungen und Definitionen.*

Bei einer statistischen Aufnahme, der eine „empirische Verteilung" entspricht, können die Angaben in folgender Weise vorliegen.

1.) m sei die Anzahl der wirklich gemachten Beobachtungen

$$X_1, X_2, \ldots, X_m, \tag{1}$$

z.B. bezeichne X_m die Anzahl von Knabengeburten auf je 100 Geburten in $m = 20$ Gruppen, etwa

$$X_1 = 45, \ X_2 = 52, \ \ldots, \ X_{20} = 51,$$

2.) es bestehen k wirklich voneinander verschiedene Möglichkeiten

(2) $$y_1, y_2, \ldots, y_k;$$

in dem Beispiel (1) wäre $k = 101$; $y_1 = 0$, $y_2 = 1, \ldots, y_k = 101$; wieder sind m Beobachtungen gemacht worden, und man gibt die relativen Häufigkeiten an

(2') $$\frac{m_1}{m}, \frac{m_2}{m}, \ldots, \frac{m_k}{m},$$

mit denen die einzelnen Möglichkeiten wirklich eingetreten sind, z.B.

$$\frac{m_1}{m} = 0, \frac{m_2}{m} = 0, \ldots, \frac{m_{46}}{m} = \frac{1}{20}, \frac{m_{52}}{m} = \frac{1}{20}, \frac{m_{53}}{m} = \frac{1}{20}, \ldots, \frac{m_{100}}{m} = 0.$$

In diesem Beispiel ist $k > m$, natürlich kann k auch $\leqq m$ sein.[2])

Die zwecks „Erklärung" einer solchen Aufnahme zugrundegelegte Hypothese besteht nun darin, daß es m theoretische Verteilungen

(3) $$v_1(x), v_2(x), \ldots, v_m(x)$$

gibt, derart daß das Ergebnis x_μ der μ-ten Verteilung zuzuschreiben ist. In unserm Beispiel kann man etwa die Hypothese machen, daß sich die Anzahl der Knabengeburten auf je 100 so verhalten wie die Anzahl der Kopf-Würfe bei 100-maligen „Kopf-Adler"-Spiel mit einer Münze, deren Kopfwahrscheinlichkeit q_μ (nahezu $= \frac{1}{2}$) ist. Dann wäre

$$v_\mu(x) = \binom{100}{x} q_\mu^x (1 - q_\mu)^{100-x} \qquad (\mu = 1, \ldots, 20).$$

Die zu den Argumenten (0, 1, ..., 100, allgemein $y_1, y_2, \ldots, y_k$) gehörigen $v_\mu(y_\varkappa)$ nennen wir $q_\varkappa^{(\mu)}$ und ihr arithmetisches Mittel $q_\varkappa$ also

(4) $$v_\mu(y_\varkappa) = q_\varkappa^{(\mu)}, \quad \frac{1}{m}\sum_{\mu=1}^{m} q_\varkappa^{(\mu)} = q_\varkappa \qquad (\varkappa = 1, \ldots, k;\ \mu = 1, \ldots, m).$$

Es entsprechen also bei der einen Auffassung den empirischen $X_1, \ldots, X_m$ die theoretischen $v_1(x), \ldots, v_m(x)$, bei der anderen den empirischen $\frac{m_1}{m}, \ldots, \frac{m_k}{m}$ die theoretischen $q_1, \ldots, q_k$.

In derartiger Weise werden eine empirische und eine theoretische Verteilung einander zugeordnet. Zu jeder dieser beiden

---

[2]) Man spielt $m = 1000$-mal mit einem Würfel und beobachtet die jeweils auftretenden Zahlen 1, 2, 3, 4, 5, 6. Dann ist $k = 6$, $m = 1000$.

gehören *Momente* beliebigen Grades. Wir bezeichnen die Momente der theoretischen Verteilung mit lateinischen, die der empirischen mit griechischen Buchstaben. Ist z.B. die Verteilung $v(x)$ arithmetisch, so ist zunächst

$$(5)\qquad \Sigma v(x)x^\nu = M_\nu^0, \text{ mit } M_1^0 = a;\ \Sigma(x-a)^\nu v(x) = M_\nu \qquad (\nu = 1, 2, \ldots);$$

und für die empirische Verteilung

$$(6)\qquad \Sigma \frac{m_\varkappa}{m} y_\varkappa^\nu = \mathbf{M}_\nu^0 \text{ mit } \mathbf{M}_1^0 = \alpha; \quad \Sigma(y_\varkappa - \alpha)^\nu \frac{m_\varkappa}{m} = \mathbf{M}_\nu,$$

bzw.

$$(7)\qquad \frac{1}{m}\sum_{\mu=1}^{m} x_\mu^\nu = \mathbf{M}_\nu^0 \text{ mit } \mathbf{M}_1^0 = \alpha; \quad \frac{1}{m}\Sigma\,(x_\mu - \alpha)^\nu = \mathbf{M}_\nu.$$

Für die Momente ersten und zweiten Grades sind besondere Namen üblich, α heißt *Durchschnitt*, $\mathbf{M}_2 = \sigma^2$ *mittleres Abweichungsquadrat, a Mittelwert* (oder *Erwartungswert* von x hinsichtlich der Verteilung $v(x)$), M_2 oder s^2 *Streuung* von x (Dispersion oder Erwartungswert von $(x-a)^2$ hinsichtlich der Verteilung $v(x)$). Diese Definitionen, insbesondere die von Erwartungswert und Streuung, erweitern sich, wenn an Stelle von x eine beliebige Funktion $f(x)$ tritt. Es ist

$$(8)\qquad \begin{aligned} &\mathfrak{E}(f) = \Sigma f(x)v(x),\\ &\mathfrak{Str}(f) = \Sigma\,[f(x) - \mathfrak{E}(f)]^2 v(x) = \\ &\qquad\qquad \mathfrak{E}\{[f - \mathfrak{E}(f)]^2\} = \mathfrak{E}(f^2) - [\mathfrak{E}(f)]^2 \end{aligned}$$

der Erwartungswert bzw. die Streuung von $f(x)$ hinsichtlich der Verteilung $v(x)$. Ferner ist für eine Funktion mehrerer Variablen $F(x_1, x_2, \ldots, x_m)$ und eine Verteilung $w(x_1, \ldots, x_m)$

$$(9)\qquad \begin{aligned} &\mathfrak{E}(f) = \Sigma \ldots \Sigma\, F(x_1, \ldots, x_m)\, w(x_1, \ldots, x_m),\\ &\mathfrak{Str}(f) = \Sigma \ldots \Sigma\,[F - \mathfrak{E}(F)]^2\, w(x_1, \ldots, x_m). \end{aligned}$$

2. *Die Pearsonsche χ^2-Methode und die Lexis'sche Theorie.*

a) die χ^2-Methode; ein empirisches Material sei gegeben durch die zu den irgendmöglichen Werten $y_1, y_2, \ldots, y_k$ gehörigen relativen Häufigkeiten $\frac{m_1}{m}, \frac{m_2}{m}, \ldots, \frac{m_k}{m}$. Man macht die Hypothese, daß die „verursachenden" theoretischen Verteilungen $v_\mu(x)$ bekannt seien und wie oben (4) $v_\mu(y_\varkappa) = q_\varkappa^{(\mu)}$. Eine besondere hier übliche Vereinfachung ist es,

$$(10)\qquad v_1(x) = v_2(x) = \ldots = v_m(x) = v(x)$$

zu setzen, dann ist natürlich $q_\varkappa^{(\mu)} = q_\varkappa$. Als Prüffunktion wählt man nach Pearson eine aus den Beobachtungen und den gegebenen $q_\varkappa$ zusammengesetzte; nämlich

$$\chi^2 = \sum_{\varkappa=1}^{k} \frac{m}{q_\varkappa}\left(\frac{m_\varkappa}{m} - q_\varkappa\right)^2, \tag{11}$$

deren Erwartungswert $\mathfrak{E}(\chi^2)$ besonders einfach wird. Auch die Streuung läßt sich berechnen. Man hat

$$\mathfrak{E}(\chi^2) = k - 1; \tag{12}$$

$$\mathfrak{Str}(\chi^2) = 2(k-1) + \frac{1}{m}\left[\sum_{\varkappa=1}^{k} \frac{1}{q_\varkappa} - k^2 - 2k + 2\right].$$

Man untersucht, wie eingangs erläutert, ob das berechnete empirische χ^2 innerhalb des Intervalls

$$\mathfrak{E}(\chi^2) \pm \sqrt{\mathfrak{Str}(\chi^2)} \tag{12'}$$

zu liegen kommt.

b) *die Lexis'sche Dispersionstheorie*; diese wird oft unpräzise dahin erklärt, daß man das empirische mittlere Abweichungsquadrat σ^2 mit „seinem theoretischen Wert" s^2 vergleiche. Das hieße etwa, man nähme als Prüffunktion die aus den empirischen Größen gebildete Größe σ^2 und vergleiche diese mit ihrem Erwartungswert. Das geht aber nicht ohneweiteres, denn es ist für die Lexis'sche Theorie charakteristisch, daß man die theoretische Verteilung und daher auch s^2 nicht (vollständig) kennt.

Bei strenger Darstellung hat die Lexis'sche Theorie allerdings einen etwas spezielleren Charakter. Wesentlich ist, daß man trotz unvollständiger Kenntnis der theoretischen Verteilungen $v_\mu(x)$ zu Kriterien kommt. Die Theorie macht eine ganz bestimmte spezielle Annahme über die $v_\mu(x)$ als *Bernoullische Verteilungen*

$$v_\mu(x) = \binom{n_\mu}{x} q_\mu^x \left(1 - q_\mu\right)^{n_\mu - x}. \tag{13}$$

Ausgehend von den durch die Praxis gebotenen Problemen wird nur die *Form* (13) als gegeben betrachtet und eventuell auch der Gruppenumfang n_μ, *nicht aber die* q_μ. Aus diesem $v_\mu(x)$ bildet man wie immer hier:

$$w(x_1, x_2, \ldots, x_m) = v_1(x_1) v_2(x_2) \ldots v_m(x_m). \tag{14}$$

Wenn man dann für σ^2 den $\mathfrak{E}(\sigma^2)$ und $\mathfrak{Str}(\sigma^2)$ rechnet, so sieht man, daß diese Größen die unbekannten Parameter q_μ enthalten und daher das oben genannte Kriterium nicht anwendbar ist. Man hat aus dieser Schwierigkeit herausgefunden, indem man spezielle und ad hoc konstruierte Vorstellungen heranzog; man kann nämlich bei dem besonderen Charakter der Aufgabe das empirische Ergebnis (1) in zweierlei Weisen auffassen. Einmal *als m Beobachtungsergebnisse, jedes gewonnen durch eine Beobachtung an einer aus n Individuen bestehenden Gruppe*, ein andermal als *die Ergebnisse von $n \cdot m$ Einzelbeobachtungen.* Indem man beide Male das mittlere Abweichungsquadrat bildet und jeweils den Erwartungswert desselben hinsichtlich der entsprechenden Verteilung, gelangt man zu Größen, die von den unbekannten Parametern q_μ frei sind. Diese Betrachtungsweise erscheint aber als „Kunstgriff", denn von vornherein war es garnicht sicher, daß man durch diese zweifache Auffassung zu vergleichbaren Größen gelangt, abgesehen von der nicht verallgemeinerungsfähigen, auf das spezielle Problem zugeschnittenen Art dieses Gedankens, dessen korrekte Durchführung übrigens langwierig ist. Das Endresultat besteht darin, die Lexis'schen Größen

$$\frac{m}{m-1}\,\sigma^2 \text{ und } \frac{nm}{mn-1}\,\alpha\left(1-\frac{\alpha}{n}\right) \tag{15}$$

miteinander zu vergleichen. Diese besitzen im Fall (10) gleiche Erwartungswerte. Andernfalls bestehen gewisse Ungleichungen zwischen ihren Erwartungswerten. Man berechnet nun noch die Streuung beider Größen und beurteilt danach die erlaubte Abweichung von der Gleichheit.

Wir wollen nun zeigen, 1.) daß man die Lexis'sche Theorie dem allgemeinen Gedanken von S. [2] 277 unterordnen kann; 2.) daß ein Gedankengang ganz frei von Kunstgriffen und außerordentlich verallgemeinerungsfähig uns zum Ziel führen wird und zwar auf einem besonders einfachen Wege.

3. *Die „Eliminationsmethode" erläutert an der Lexis'schen Theorie.*

Was soll bei der Lexis'schen Fragestellung eigentlich „geprüft" werden? Offenbar nichts anders als, ob die Hypothese zulässig ist, daß die Beobachtungsergebnisse $X_1, X_2, \ldots, X_m$ gerade durch Verteilungen Bernoullischer Form (13) interpretiert werden können. Die Hypothese besteht nur in der Annahme dieser *Form,* nicht in Voraussetzungen über die Parameter q_μ. Ob nun die Annahme einer bestimmten Verteilungsform, z.B. einer Bernoullischen, zulässig ist, das läßt sich an *einem einzigen skalaren Kriterium* gewiß nicht restlos prüfen, man kann aber z.B. irgendeine bei einer Bernoullischen Verteilung bestehende Beziehung herausgreifen und sehen, ob sie erfüllt ist. Genauer gesagt kann man etwa folgendermaßen vorgehen.

Da in *alle* Momente M_ν der Bernoullischen Verteilung mit den Parametern n und q doch nur diese beiden Größen eingehen können, so muß es möglich sein, den einen oder anderen oder auch beide Parameter zwischen zwei oder drei Momenten zu *eliminieren* und so eine von dem unerwünschten Parameter freie reine Momentengleichung zu erhalten. Natürlich gibt es nicht *eine* solche Gleichung, sondern unendlich viele, man kann aber zunächst — wir werden noch andere Gesichtspunkte erörtern — darauf ausgehen, die *einfachste* derartige Beziehung bzw. die zwischen den Momenten niedrigster Ordnung bestehende als Kriterium zu nehmen. *In dem Aufsuchen einer solchen Beziehung besteht der erste Schritt.*

Auf die Bernoullische Verteilung und auf die Elimination von q angewendet, ist er sogleich erledigt. Mit den Bezeichnungen (5) gilt, wenn wir noch s^2 für M_2 schreiben, bekanntlich:

$$(16) \qquad a = nq, \qquad s^2 = nq(1-q).$$

Elimination von q gibt

$$(17) \qquad s^2 = a - \frac{a^2}{n} \quad \text{oder} \quad (17') \quad s^2 - a + \frac{a^2}{n} = 0.$$

Das ist die einfachste der genannten Momentbeziehungen.

Die Verbindung von (17) mit den Beobachtungen wird nun hergestellt, indem man *eine solche Funktion* $F(x_1, x_2, \ldots, x_m)$ *der Beobachtungen zu gewinnen sucht, daß* (17′) *ihr Erwartungswert hinsichtlich der Verteilung* (14) *ist* (meist mit der Nebenbedingung (10)); oder, wie wir gelegentlich kurz sagen wollen, indem man (17) in eine Beziehung zwischen Erwartungswerten empirischer Größen „übersetzt". Die so gewonnene Funktion der $x_1, x_2, \ldots, x_m$ ist dann eine geeignete Prüffunktion F. Denn ihr Erwartungswert ist gegeben durch die linke Seite von (17′), diese ist aber so gewählt, daß sie im Fall der Bernoullischen Verteilung gleich Null ist. Wir haben auf diese Art dann eine Funktion F der Beobachtungen, deren Erwartungswert von den unerwünschten Parametern frei ist, wenn die Hypothese hinsichtlich der $v_\mu(x)$ zu Recht besteht, er ist bei unserer Wahl insbesondere gleich Null.

Die Prüfung besteht dann schließlich darin, in die $F(x_1, x_2, \ldots, x_m)$ die wirklichen Beobachtungen $X_1, X_2, \ldots, X_m$ einzuführen und *zu sehen, ob* $F(X_1, X_2, \ldots X_m)$ *von* 0 *nur innerhalb der Grenzen*

$$\pm \sqrt{\mathfrak{S}\mathrm{tr}(F)}$$

abweicht.

Wir kehren zu (17) zurück. Das „Übersetzen" ist hier ungemein einfach, wenn man nur die Erwartungswerte (und Streuungen vgl. (9)) der einfachsten Funktionen der $x_1, x_2, \ldots, x_m$ berechnen kann, nämlich des Durchschnittes α, seines Quadrates α^2 und des mittleren Abweichungsquadrates $\sigma^2 = \mathbf{M}_2$. Wir beschränken uns zunächst auf (10). Die Verteilung $w(x_1, x_2, \ldots, x_m)$ ist durch (14) gegeben. Die benötigten Formeln lauten

$$(18) \qquad \mathfrak{E}(\alpha) = a, \quad \mathfrak{E}(\sigma^2) = \frac{m-1}{m} s^2, \quad \mathfrak{E}(\alpha^2) = a^2 + \frac{s^2}{m}.$$

Es ist also, wenn man benutzt, daß der Erwartungswert einer Summe gleich der Summe der Erwartungswerte ist,

$$(19) \qquad a = \mathfrak{E}(\alpha), \quad s^2 = \frac{m}{m-1}\mathfrak{E}(\sigma^2), \quad a^2 = \mathfrak{E}\left(\alpha^2 - \frac{\sigma^2}{m-1}\right).$$

Wir haben somit die in (17) auftretenden Größen a, s^2, a^2 als Erwartungswerte empirischer Größen ausgedrückt. Setzt man aus (19) in (17) ein, so erhält man

$$\frac{m}{m-1}\mathfrak{E}(\sigma^2) = \mathfrak{E}(\alpha) - \frac{1}{n}\mathfrak{E}(\alpha^2) + \frac{1}{n(m-1)}\mathfrak{E}(\sigma^2)$$

oder

$$(20) \qquad \mathfrak{E}\left[\frac{m}{m-1}\sigma^2\right] = \mathfrak{E}\left[\frac{nm}{nm-1}\left(\alpha - \frac{\alpha^2}{n}\right)\right].$$

In der Gleichheit dieser Erwartungswerte besteht aber genau das bekannte Lexis'sche Kriterium (15), das wir hier durch die in (16)—(20) enthaltenen wenigen Zeilen begründet haben.

Die Prüffunktion F, deren Erwartungswert bei der gemachten Hypothese gleich Null ist, lautet also hier

$$(21) \qquad \frac{m}{m-1}\sigma^2 - \frac{nm}{nm-1}\left(\alpha - \frac{\alpha^2}{n}\right).$$

Immer ist der Erwartungswert von (21) durch die linke Seite von (17′) gegeben. Im Fall der Bernoullischen Verteilung aber ist diese und somit der fragliche Erwartungswert gleich Null. Die Lexis'sche Theorie besteht im Falle (10), wie man sieht, nur darin, das für eine Bernoullische Verteilung notwendige Bestehen der Gleichung (17) zu prüfen. — Durch unsere Auffassung ist die Lexis'sche Theorie, soweit sie auf der Annahme (10) beruht, der allgemeinen „Methode der Prüffunktion" untergeordnet.

4. *Erweiterungen und Zusätze.*

1.) Wenn die Funktion (21) erst einmal gefunden ist, ist es eine zwangsläufige Ausrechnung, ihre *Streuung*, immer hinsichtlich (14), *zu rechnen.* Dies ist aber oft mühevoll, wenn man es genau macht. Man findet hier mit $s^2 = M_2$

$$
(22)\quad
\begin{aligned}
\mathfrak{Str}\left(\frac{m}{m-1}\,\sigma^2\right) &= \frac{s^2}{m(m-1)}\left[2ms^2 + (m-1)(1-6q(1-q))\right] \\
&= \frac{1}{m(m-1)}\left[(m-1)\,M_4 + (3-m)\,M_2\right], \\
\mathfrak{Str}\left(\frac{nm}{nm-1}\left(\alpha - \frac{\alpha^2}{n}\right)\right) & \\
&= \frac{s^2}{m(nm-1)}\left[(nm-1) - 2q(1-q)(2nm-3)\right].
\end{aligned}
$$

Das ist schon hinlänglich unübersichtlich. Noch komplizierter wird es, wenn man nicht die Streuung der rechten und der linken Seite von (15) sondern die Streuung von (21) sucht! Denn die Streuung ist nicht *additiv*: Die Streuung einer Summe unterscheidet sich von der Summe der Streuungen noch um einen Zusatzbestandteil. Hingegen vereinfacht sich alles, wenn man — ausgehend von dem Gedanken, daß für praktisch verwendbare Fälle m ja doch eine verhältnismäßig große Zahl sein muß — Glieder von der Ordnung $\frac{1}{m^\nu}$ $(\nu > 1)$ vernachlässigt. (Die nähere Begründung dieser Überlegungen und der Formel (23) folgt in § 2.) Dann findet man für die gesuchte Streuung von (21)

$$
(23)\quad
\begin{aligned}
\mathfrak{Str}\left[\frac{m}{m-1}\,\sigma^2 - \frac{nm}{nm-1}\,\alpha + \frac{m}{nm-1}\,\alpha^2\right] & \\
= \frac{1}{m}\left[M_2\left(\frac{2a}{n}-1\right)^2 + 2M_3\left(\frac{2a}{n}-1\right) + \left(M_4 - M_2^2\right)\right]&.
\end{aligned}
$$

Die rechts in (23) auftretenden theoretischen Momente sind aber alle unbekannt: Es wird sich nun in § 2 zeigen, daß der Fehler, den man macht, wenn man diese unbekannten M_ν durch die bekannten entsprechenden empirischen $\mathbf{M}_\nu$ ersetzt, von der Ordnung $\frac{1}{m^2}$ ist, und wir dies daher — bei Vernachlässigung von Gliedern dieser Ordnung — tun dürfen! Andererseits zeigt dies, daß die *genaue* Berechnung der $\mathfrak{Str}(F)$ von diesem Standpunkt aus nicht viel Sinn hat, da wir bei der tatsächlichen Abschätzung (wenn wir weiteren Komplikationen ausweichen wollen) ja schließlich doch die M_ν durch die $\mathbf{M}_\nu$ ersetzen müssen.

Noch eine Bemerkung läßt sich an (22) oder (23) anknüpfen: Die drei Summanden in (23) sind in n von verschiedener Größenordnung. Der letzte, der von $\mathfrak{Str}(\sigma^2)$ herrührt, ist von der Ordnung n^2, der erste, herrührend von $\mathfrak{Str}\left(\alpha - \frac{\alpha^2}{n}\right)$ ist von der Ordnung n, der mittlere Zusatzbestandteil, der von der *Summierung* von σ^2 und $\left(\alpha - \frac{\alpha^2}{n}\right)$ kommt, ist auch von der Ordnung n. Es ist also der Erwartungswert von σ^2 mit einer größeren Streuung behaftet als der von $\left(\alpha - \frac{\alpha^2}{n}\right)$. Es hat also zwar wenig Sinn, σ^2, wie das gelegentlich geschieht, als „empirische" Streuung in Gegensatz zu einer „theoretischen" Streuung $\left(\alpha - \frac{\alpha^2}{n}\right)$ zu stellen; doch ist im erörterten Sinn der eine empirische Wert allerdings „*genauer*" als der andere, da er mit geringerer Streuung behaftet ist und die Wiederholungszahl n in den Anwendungen meist eine recht große Zahl, zum Beispiel die Bevölkerungszahl eines Landes ist.

2.) Wir haben hier (10) angenommen, und daraus die *Gleichung* zwischen den Erwartungswerten (20) abgeleitet. Auf den allgemeinen Fall (13) voneinander verschiedener $v_\mu(x)$, wobei also in (14) statt 1 bzw. 2 jetzt m bzw. $2m$ unbekannte Parameter eingehen, ist die S. [8] 283 dargelegte Methode nicht unmittelbar anwendbar, insbesondere was das „Übersetzen" der theoretischen Beziehung in eine zwischen Erwartungswerten empirischer Größen betrifft. Man kann aber stets, nach Erledigung des Hauptfalles (10), sehen, was aus dem Erwartungswert der so aufgefundenen Prüffunktion — hier (21) — bei Annahme des allgemeinen Falles (14) wird. *Diese* Überlegung führt dann zu weiteren Ergebnissen der Lexis'schen Theorie: Wenn, wie in (13), voneinander verschiedene $v_\mu(x)$ vorliegen, so gilt für *jede* (17), d.h. wenn alle $n_\mu = n$ und nur die q_μ voneinander verschieden:

$$(24) \qquad s_\mu^2 - a_\mu + \frac{a_\mu^2}{n} = 0 \quad (\mu = 1, 2, \ldots, m)$$

$$\text{oder } \sum_1^m s_\mu^2 - \sum_1^m a_\mu + \frac{1}{n} \sum_1^m a_\mu^2 = 0.$$

An Stelle von (18) ist dann, wenn $a = \frac{1}{m}(a_1 + a_2 + \ldots + a_m)$:

$$(25)\quad \mathfrak{E}(\alpha) = \frac{1}{m}\sum_1^m a_\mu;\quad \mathfrak{E}(\sigma^2) = \frac{m-1}{m^2}\sum_1^m s_\mu^2 + \frac{1}{m}\sum_1^m (a_\mu - a)^2,$$

$$\mathfrak{E}(\alpha^2) = \frac{1}{m^2}\sum_1^m (s_\mu^2 + a_\mu^2) + \frac{2}{m^2}\sum_{(\mu\neq\nu)}^{1\ldots m} a_\mu a_\nu = \frac{1}{m^2}\left[\sum_1^m s_\mu^2 + \left(\sum_1^m a_\mu\right)^2\right].$$

Bildet man dann den Erwartungswert von (21), so kommt:

$$\frac{1}{m}\sum s_\mu^2 + \frac{1}{m-1}\sum (a_\mu - a)^2 - \frac{n}{nm-1}\sum a_\mu + \frac{1}{m(nm-1)}\left[\sum s_\mu^2 + (\sum a_\mu)^2\right].$$

Unter Benutzung von (24) erhält man dafür

$$(26)\quad \mathfrak{E}\left[\frac{m}{m-1}\sigma^2 - \frac{nm}{mn-1}\alpha + \frac{m}{nm-1}\alpha^2\right] = \frac{m(n-1)}{(m-1)(nm-1)}\left[\sum a_\mu^2 - m a^2\right] > 0\,.$$

Damit ist der sogenannte *Lexis'sche Fall* erledigt, der besagt, daß für verschiedene q_μ in (13) der Erwartungswert der Prüffunktion (21) positiv ist, was man als Fall „übernormaler Dispersion" zu bezeichnen pflegt, da die linke Seite von (20) in diesem Fall einen *größeren* Erwartungswert hat als die rechte.

Hingegen gehört der sog. *Poissonsche Fall* von unserem systematischen Standpunkt aus an eine andere Stelle. Es ist der Fall, daß die einzelnen nach der Hypothese anzunehmenden Verteilungen garnicht Bernoullische sind; sondern sie sind zwar alle *untereinander gleich*, aber jede ist eine sogenannte *Poissonsche Verteilung*, die (zum Unterschied von der Bernoullischen) durch einfache Summenbildung oder Faltung aus *von einander verschiedenen* einfachen Alternativversuchen mit den Grundwahrscheinlichkeiten $q^{(\nu)}$ $(\nu = 1, 2, \ldots, n)$ entstehen.

Die in der Lexis'schen Theorie bekannte Ungleichung ergibt sich hier noch einfacher als im Lexis'schen Fall: Der Erwartungswert von (21) ist ja *immer* durch die linke Seite von (17'), d.h. durch $s^2 - a + \frac{a^2}{n}$ gegeben. Nun ist für die Poissonsche Verteilung anstelle von (16)

$$(27)\quad a = \sum_1^n q^{(\nu)};\quad s^2 = \sum_1^n q^{(\nu)}(1 - q^{(\nu)}) = \sum_1^n q^{(\nu)} - \sum_1^n (q^{(\nu)})^2.$$

Somit

$$(28)\qquad \mathfrak{E}\left[\frac{nm-1}{n(m-1)}\sigma^2-\alpha+\frac{\alpha^2}{n}\right]$$
$$=s^2-a+\frac{a^2}{n}=\sum_1^n q^{(\nu)}-\sum_1^n (q^{(\nu)})^2-\sum_1^n q^{(\nu)}+\frac{1}{n}(\sum_1^n q^{(\nu)})^2$$
$$=\frac{1}{n}(\sum_1^n q^{(\nu)})^2-\sum_1^n (q^{(\nu)})^2<0.$$

Es ist also *hier* der Erwartungswert von (21) *negativ*, was man als Fall „unternormaler Dispersion" zu bezeichnen pflegt. Auch hier liegt keine Anwendung unserer eigentlichen Methode vor, obgleich diese hier prinzipiell möglich wäre, da (10) gilt und nur eben n unbekannte Parameter $q^{(\nu)}$ auftreten. Sondern es wird lediglich — zwecks Gewinnung der bekannten Resultate — der Erwartungswert der einmal gewonnenen Prüffunktion (21) unter der Voraussetzung betrachtet, daß die untereinander *gleichen* $v_\mu(x)=v(x)$ statt Bernoullischer Verteilungen Poissonsche sind. [3])

§ 2. Die Eliminationsmethode als allgemeines Verfahren der Statistik.

5. *Darstellung der Methode.*

Die im vorhergehenden an dem Spezialfall der Lexis'sche Theorie erläuterte „Eliminationsmethode" ist offenbar in keiner Weise auf den speziellen Problemkreis dieser Methode beschränkt. Wenn es als Grundaufgabe der theoretischen Statistik betrachtet wird, Kriterien aufzustellen zur Prüfung der Frage, ob ein gewisses empirisches Material $X_1, \ldots, X_m$ durch gewisse Verteilungen $v_\mu(x)$ „erklärt" werden kann, so kann man die folgende Unterscheidung treffen. Entweder die $v_\mu(x)$ sind vollständig gegeben, wie z.B. im Fall der χ^2-Methode, oder sie sind zwar der Form nach vermutet, aber gewisse unbekannte Parameter sind noch offengelassen, wie im Fall der Lexis'schen Theorie. Im ersten Fall wird man mit Prüffunktionen von der Art des Pearsonschen χ^2, des v. Mises'schen ω^2 oder dergl. arbeiten, die aus den Beobachtungen und aus den ja vollständig bekannten hypothetischen (theoretischen) Verteilungen in geeigneter Weise zusammengesetzt sind. Den zweiten, schwierigeren Fall, als dessen klassi-

[3]) Zwecks Elimination der unbekannten Parameter $q^{(\nu)}$, zwecks Erzielung einer *Gleichung*, beachte man, daß der in (28) gerechnete Erwartungswert von m *unabhängig* ist. Die Differenz zweier solcher Erwartungswerte für gleiches n und verschiedenes m muß daher verschwinden.

sches Beispiel die Lexis'sche Theorie erscheint, kann man mit der „Eliminationsmethode" behandeln. Zunächst sei (10) erfüllt, was wir als „Hauptfall" bezeichnen wollen.

Man wähle also irgendeine geeignete Folge von Größen, die aus den theoretischen Verteilungen abgeleitet werden. Hier wollen wir — als geeignet und einfach — die Momente wählen. Dabei bleibt es dahin gestellt, welche Art von Momenten genommen wird, Momente auf den Nullpunkt bezogen, auf den Schwerpunkt bezogen, faktorielle Momente, Semiinvarianten od. dgl. Denken wir z.B. an die Nullmomente. Die Parameter der Verteilung gehen im allgemeinen in sämtliche Momente ein. Aus der unbegrenzten Folge der Gleichungen, welche die unendlich vielen Momente als Funktionen der Parameter ausdrücken, greift man dann geeignete und soviele als notwendig heraus und eliminiert zwischen diesen die unbekannten Parameter. Sagen wir, es seien $\varkappa$ unbekannte Parameter vorhanden, so erhält man im allgemeinen eine Beziehung, in die Momente *mindestens* bis zur $(\varkappa+1)$-ten Ordnung eingehen. Wir wählen unter den vielen existierenden Beziehungen dieser Art eine *zwischen den Momenten niedrigster Ordnung*, sofern nicht irgendwelche Gesichtspunkte dagegen sprechen.

Es besteht hier offenbar eine gewisse Freiheit, die auch nicht von vornherein eingeschränkt werden soll. Z.B. könnte im Gegensatz zu der eben gegebenen, im allgemeinen zu beachtenden Vorschrift, zwischen gewissen höheren Momenten eine viel einfachere Beziehung bestehen als zwischen solchen niedrigerer Ordnung, so daß trotz der höheren Ordnung die einfachere Beziehung empfehlenswerter schiene. Auch wird man es vermeiden, eine solche Beziehung als „charakteristische Gleichung" auszuwählen, die auch für andere einfache bekannte Verteilungen in gleicher Weise gilt, und man würde in diesem Fall unter Heranziehung höherer Momente eine andere Beziehung aufsuchen, die von diesem Mangel frei ist. — Endlich könnte es geschehen, daß die aus der Beziehung niedrigster Ordnung abgeleitete Prüffunktion eine verhältnismässig große Streuung aufweist. Dann wäre sie zu unpräzise für eine Prüffunktion, und man würde ev. eine Beziehung zwischen Momenten höherer Ordnung heranziehen [4]).

Jedenfalls aber kommt man schließlich dazu, eine Wahl zu treffen, d.h. eine Gleichung herauszugreifen, die für Verteilungen der betrachteten Art gewiß erfüllt ist. (Das Umgekehrte gilt natürlich nicht: Zu jeder einzelnen solchen Gleichung

[4]) Vgl. dazu auch den Begriff des sogen. „Gütemaßes", der in verschiedener Weise erklärt werden kann, z.B. l. c. [1]). S. 280.

kann man ganz andere Verteilungen als die betrachtete angeben, zwischen deren Momenten dieselbe Beziehung besteht.) So eine Gleichung für die Bernouillische Verteilung war die Beziehung (17).

Es folgt nun der zweite Teil der Aufgabe, die „*Übersetzung*" *dieser theoretischen Gleichung in eine Beziehung zwischen den Erwartungswerten der empirischen Größen.* Gegeben ist also eine Beziehung zwischen den theoretischen, hinsichtlich einer gegebenen Verteilung $v(x)$ gebildeten Momenten, etwa

$$G(a, M_2^0, M_3^0, M_4^0, \ldots) = 0. \tag{29}$$

Es handelt sich darum, eine *Funktion der empirischen Größen* $F(\alpha, \mathbf{M}_2^0, \mathbf{M}_3^0, \mathbf{M}_4^0, \ldots)$ *anzugeben, deren Erwartungswert* (immer hinsichtlich einer gegebenen Verteilung (14), (10)) *gerade* (29) *ist.* Eine allgemeine und einfache Näherungslösung dieser Aufgabe wird bei Annahme großer m in § 3 gegeben. Die Aufgabe läßt sich aber (wie in Abschn. 6 gezeigt werden wird) ohne Vernachlässigung behandeln, falls (29) ein *Polynom* in den Momenten ist, womit man ja sehr oft auskommen wird.

Ist sodann die gesuchte Funktion der empirischen Größen $F(\alpha, \mathbf{M}_2^0, \ldots)$ mit dem Erwartungswert (29) gefunden, so rechnet man schließlich ihre Streuung $\mathfrak{Str}(F)$. Auch diese kann man immer genau rechnen, wenn F ein Polynom in den Momenten ist; doch wird diese Rechnung — wie wir schon in 4 gesehen haben — selbst in verhältnismäßig einfachen Fällen kompliziert und unübersichtlich. Daher treten besonders hierbei dann die Näherungsformeln von § 3 in Kraft.

Hat man dies erledigt, so folgt der letzte Schritt; man rechnet aus den beobachteten Größen $X_1, X_2, \ldots, X_m$ die beobachteten Momente $\alpha_{\text{beob.}}, M^0_{2\text{beob.}}, \ldots$ etc. und setzt dies in F und in $\mathfrak{Str}(F)$ ein. Man sieht sodann nach, ob dieses so erhaltene $F(\alpha_{\text{beob.}}, M^0_{2\text{beob.}}, \ldots) = F_{\text{beob.}}$ innerhalb der Grenzen $0 \pm (\sqrt{\mathfrak{Str}(F)})_{\text{beob.}}$, also im Intervall

$$-\left(\sqrt{(\mathfrak{Str}(F)}\right)_{\text{beob.}} \text{ und } +\left(\sqrt{\mathfrak{Str}\,(F)}\right)_{\text{beob.}}$$

gelegen ist.

In diesen drei Schritten

1.) Elimination der unbekannten Parameter zwecks Gewinnung einer von diesen freien Beziehung zwischen den Momenten,

2.) „Übersetzung" dieser Beziehung in eine zwischen den Erwartungswerten und Berechnung der Streuung der so gewonnenen Funktion, 3.) Berechnung von $F_{\text{beob.}}$ und $(\sqrt{\mathfrak{Str}(F)})_{\text{beob.}}$ und Prüfung, ob $F_{\text{beob.}}$ von 0 nur um $\pm\ (\sqrt{\mathfrak{Str}(F)})_{\text{beob.}}$ abweicht, besteht die hier entwickelte Methode. Sie dient — um es noch einmal zusammenzufassen — dazu, auch den Fall, daß die hypothetisch angenommene Verteilung $v(x)$ unbekannte Parameter enthält, einer korrekten und systematischen Behandlung einzuordnen. Dabei sind der zweite und der dritte Schritt zwangsläufig. Im ersten Schritt aber besteht eine gewisse Willkür für die Wahl der in Rede stehenden Beziehung.

Jedenfalls wird durch das ganze Verfahren niemals mehr geprüft als das Bestehen einer einzigen skalaren Gleichung, der die hypothetische Verteilung $v(x)$ genügt. Aber andererseits ist es ja stets der Wunsch, sich an *einem zusammenfassenden Kriterium* zu orientieren. Ist die Bedingung (10) nicht erfüllt, so ist die hier dargestellte Methode in dieser Weise nicht anwendbar, da sich die unter 2.) erwähnte „Übersetzung" in dieser Weise nicht leisten läßt.

Man kann aber stets das folgende tun: Man betrachtet die im Hauptfall (10) gefundene Prüffunktion $F(\alpha, \mathbf{M}_2^0, \ldots) = 0$ und untersucht *ihren* Erwartungswert unter der allgemeinen Annahme *verschiedener* $v_i(x)$. Dieser Erwartungswert wird im Allgemeinen von Null verschieden sein, und man wird evtl. zu *Ungleichungen* gelangen, die den Fall verschiedener $v_i(x)$ charakterisieren. Diese Vorschrift ist dem Verfahren im sogen. Lexis'schen Fall nachgebildet. Die Berechnung von Erwartungswerten im Falle verschiedener $v_i(x)$, also für die Verteilung (14), bereitet keine Schwierigkeit prinzipieller Art. Die Näherungsmethoden von § 3 umfassen auch diesen Fall.

6. *Die Umkehraufgabe für Polynome in den Momenten.*

Wir wollen nun die S. [15] ... angedeutete Umkehraufgabe genauer untersuchen. Wir ändern die Bezeichnung hierzu vorübergehend, indem wir die Nullmomente statt mit M^0 mit M bezeichnen (bezw. $\mathbf{M}$ statt $\mathbf{M}^0$) und die Schwerpunktmomente zum Unterschied mit $\overline{M}$ bezw. $\overline{\mathbf{M}}$. Es gelten also wenn $v(x)$ irgendeine — etwa arithmetische — Verteilung ist, die Bezeichnungen

$$(5') \qquad M_i = \Sigma\, x^i v(x) \quad \text{und} \quad (7') \quad \mathbf{M}_i = \frac{1}{m} \Sigma\, x_\mu^i\,,$$

wenn die $x_1, x_2, \ldots, x_m$ gewisse m Beobachtungergebnisse darstellen. Wir bilden ferner

$$(14') \qquad w(x_1, x_2, \ldots x_m) = v(x_1) v(x_2) \ldots v(x_m).$$

Bezüglich der Verteilung (14′) rechnen wir nach (9) die Erwartungswerte beliebiger Funktionen $F(x_1, x_2, \ldots, x_m)$, also

$$(9) \qquad \mathfrak{E}(F) = \Sigma \ldots \Sigma F(x_1, \ldots, x_m) w(x_1, \ldots, x_m).$$

Aus den M bzw. den $\mathbf{M}$ bilden wir nun Ausdrücke:

$$(30) \qquad M_i^{\nu_i} M_j^{\nu_j} \ldots M_\varkappa^{\nu_\varkappa} \text{ bzw. } (30') \quad \mathbf{M}_i^{\nu_i} \mathbf{M}_j^{\nu_j} \ldots \mathbf{M}_\varkappa^{\nu_\varkappa}.$$

Wir nennen

$$(31) \qquad \nu = i\nu_i + j\nu_j + \ldots + \varkappa\nu_\varkappa$$

das *Gewicht* von (30) bzw. (30′) und

$$(31') \qquad \nu' = \nu_i + \nu_j + \ldots + \nu_\varkappa \text{ mit } \nu' \leqq \nu$$

die *Faktorenzahl* von (30) bzw. (30′).

Versieht man Ausdrücke wie (30) (bzw. (30′)) mit konstanten Koeffizienten und addiert sie, so erhält man ein *Polynom in den* (theoretischen, bzw. den empirischen) *Momenten.* Haben alle Glieder eines solchen das gleiche Gewicht ν, so sprechen wir von einem *homogenen* Polynom vom Gewicht ν in den Momenten. (Z.B. ist das ν-te Schwerpunktmoment $\overline{M}_\nu$ ein homogenes Polynom vom Gewicht ν in den Nullmomenten.)

Dann gilt mit den Bezeichnungen (5′), (7′), (14′), (9) der folgende Satz:

Ein Ausdruck vom Gewicht ν der Form (30) *in den theoretischen Momenten, für den die Faktorenzahl $\nu' \leqq m$ ist, ist der Erwartungswert eines homogenen Polynoms vom Gewicht ν in den empirischen Nullmomenten.*

Selbstverständlich gilt dann auch das gleiche für ein aus Ausdrücken der Form (30) zusammengesetztes *Polynom* und auch, falls die gegebenen theoretischen Momente Schwerpunktmomente $\overline{M}$ sind, da ein Ausdruck $\overline{M}_i^{\nu_i} \overline{M}_j^{\nu_j} \ldots \overline{M}_\varkappa^{\nu_\varkappa}$ in diesen gleich einem homogenen Polynom vom Gewicht ν in den Nullmomenten ist. (Aber es wird nicht behauptet, daß ein aus den theoretischen Schwerpunktmomenten gebildeter Ausdruck sich als Erwartungswert eines Polynoms in den empirischen Schwerpunktmomenten ausdrückt.)

Wir werden obigen Satz beweisen, indem wir zur Auffindung des gesuchten Polynoms in den empirischen Momenten eine Rechenvorschrift geben. Zunächst rechnen wir den Erwartungswert eines Produktes

$$(32) \qquad x_1^i x_2^k \ldots x_m^r,$$

wo die $i, k, \ldots, r$ nicht negative ganze Zahlen sind:

$$(33)\qquad \mathfrak{E}(x_1^i x_2^k \ldots x_m^r) = \sum x_1^i x_2^k \ldots x_m^r w(x_1, x_2, \ldots, x_m) = \\ = \sum x_1^i v(x_1) \sum x_2^k v(x_2) \ldots \sum x_m^r v(x_m) = M_i M_k \ldots M_r.$$

Ein Ausdruck (30′) mit den gemäß (7′) erklärten $\mathbf{M}_i$ ergibt aber ausmultipliziert eine Summe von endlich vielen Ausdrücken der Form (32), und somit ist zunächst gesehen, daß *der Erwartungswert von* (30′) *sicher ein Polynom in den theoretischen Momenten ist*; und zwar ein *homogenes Polynom* vom Gewicht ν.

Wir müssen diese einfache Tatsache nun genauer verfolgen. Eine Zahl ν sei gegeben. Dann kann man verschiedene Ausdrücke in den $\mathbf{M}$ alle vom Gewicht ν bilden, nämlich

$$(34')\qquad \mathbf{M}_\nu;\ \mathbf{M}_{\nu-1}\alpha,\ \mathbf{M}_{\nu-2}\mathbf{M}_2, \ldots;\ \mathbf{M}_{\nu-2}\alpha^2,\ \mathbf{M}_{\nu-3}\mathbf{M}_2\alpha, \ldots; \ldots;\ \alpha^\nu.$$

$\mathbf{M}_\nu$ hat die Faktorenzahl 1, die nächstfolgende Gruppe zwischen den Strichpunkten die Faktorenzahl 2, dann folgt eine Gruppe mit der Faktorenzahl 3, schließlich ist die Faktorenzahl für α^ν gleich ν. Innerhalb jeder Gruppe mit bestimmter Faktorenzahl ist die Reihenfolge gleichgültig.

Nun betrachten wir die Erwartungswerte der Ausdrücke (34′). Irgendwo in der Reihe (34′) kommt der allgemeine Ausdruck (30′) vor; er lautet abgesehen von dem Faktor $\dfrac{1}{m^{\nu'}}$

$$(30'')\qquad (x_1^i + x_2^i + \ldots + x_m^i)^{\nu_i} \cdot (x_1^j + x_2^j + \ldots + x_m^j)^{\nu_j} \cdot \ldots \cdot \\ \cdot (x_1^\varkappa + x_2^\varkappa + \ldots + x_m^\varkappa)^{\nu_\varkappa}.$$

Beim Ausmultiplizieren von (30″) treten zunächst Ausdrücke auf mit *lauter gleichen* x_μ, z.B.

$$(35')\qquad x_1^{i\nu_i} x_1^{j\nu_j} \ldots \cdot x_1^{\varkappa\nu_\varkappa} = x_1^\nu \quad \text{oder etwa } x_2^\nu \text{ usf.},$$

ferner solche mit *zwei* verschiedenen x_μ, wie etwa

$$(35'')\qquad x_1^i x_2^{i(\nu_i-1)} x_2^{j\nu_j} \ldots x_2^{\varkappa\nu_\varkappa} = x_1^i x_2^{\nu-i} \text{ oder etwa } x_2^{i+3j} x_3^{\nu-i-3j} \text{ usf.},$$

schließlich solche mit $\nu' = \nu_i + \nu_j + \ldots \nu_\varkappa$ verschiedenen x_μ (noch *mehr* von einander verschiedene x_μ sind offenbar unmöglich), also etwa:

$$(35)\qquad x_1^i x_2^i \ldots x_{\nu_i}^i x_{\nu_i+1}^j x_{\nu_i+2}^j \ldots x_{\nu_i+\nu_j}^j \ldots x_{\nu'}^\varkappa \quad \text{oder etwa} \\ x_1^j x_2^i \ldots x_{\nu_i}^i x_{\nu_i+1}^i x_{\nu_i+2}^j x_{\nu_i+\nu_j}^j \ldots x_{\nu'}^\varkappa \text{ usf.}$$

Alle Ausdrücke der Form (35′) haben den Erwartungswert M_ν, alle der Form (35) den Erwartungswert (30), die von (35″) haben Erwartungswerte wie $M_i M_{\nu-i}$ usf. (Die Koeffizienten all der verschiedenen Ausdrücke (35), (35′) usf. in der Entwicklung von (30″) interessieren uns hier nicht, sie lassen sich mit Hilfe des polynomischen Satzes angeben.) Der Erwartungswert des empirischen Ausdrucks (30′) mit der Ordnung ν und der Faktorenzahl ν' ist also ein Polynom ν-ter Ordnung in den M, wobei *eines* der Glieder dieses Polynoms die Faktorenzahl ν' aufweist und aus (30′) *durch Ersetzen der* $\mathbf{M}$ *durch die entsprechenden* M entsteht; alle anderen Glieder haben eine Faktorenzahl kleiner ν'.

Alle auftretenden Glieder entstehen gemäß (30″) in folgender Weise: In dem gegebenen Ausdruck (30′) seien verschiedene Indices $i, j, \ldots k$ vorhanden, der erste ν_i-mal, der zweite ν_j-mal, der letzte $\nu_\varkappa$-mal. Aus diesen ν' Zahlen wählt man irgend eine Teilgruppe und ersetzt sie durch ihre Summe, z.B. wählt man dreimal den Index i und einmal den Index j und schreibt dafür den Index $(3i+j)$, ferner zweimal den Index i und schreibt dafür den Index $2i$, u.s.f. So entsteht etwa mit den Indices

$$(36) \qquad (3i+j),\ 2i,\ 4\varkappa,\ \underbrace{i,\ i,\ \ldots,\ i}_{(\nu_i-5)\,\mathrm{mal}},\ \underbrace{j,\ldots,j}_{(\nu_j-1)\,\mathrm{mal}},\ldots,\ \underbrace{\varkappa,\ldots,\varkappa}_{(\nu_\varkappa-4)\,\mathrm{mal}}$$

ein Glied: $M_{3i+j}\, M_{2i} M_{4\varkappa}\, M_i^{\nu_i-5}\, M_j^{\nu_j-1}, \ldots, M_\varkappa^{\nu_\varkappa-4}$.

(Z.B. hat der Erwartungswert von $\mathbf{M}_2^2\alpha^2$ mit $\nu = 6$, $\nu' = 4$ entsprechend den Zusammenfassungen von 2, 2, 1, 1,

$$2, 2, 1, 1 \mid 4, 1, 1 \mid 3, 2, 1 \mid 2, 2, 2 \mid 4, 2 \mid 5, 1 \mid 3, 3 \mid 6 \mid,$$

die Gestalt

$$\mathfrak{E}(\mathbf{M}_2^2\, \alpha^2) = c_1 M_6 + c_2 M_3^2 + c_3 M_5 a + c_4 M_4 M_2 + \\ + c_5 M_2^3 + c_6 M_3 M_2 a + c_7 M_4 a^2 + c_8 M_2^2 a^2.$$

Nur der letzte Summand, der dem $\mathbf{M}_2^2\alpha^2$ auf der linken Seite entspricht, hat $\nu' = 4$ Faktoren, alle anderen weniger.)

Nun denken wir uns schließlich für *alle* Ausdrücke (34′) vom Gewicht ν nacheinander die Erwartungswerte angeschrieben und fassen in der Folge der so erhaltenen Gleichungen die dabei auftretenden Ausdrücke in den M, also

$$(34) \qquad M_\nu;\ M_{\nu-1}a,\ M_{\nu-2}M_2, \ldots;\ M_{\nu-2}a^2,\ M_{\nu-3}M_2 a, \ldots; \ldots a^\nu$$

als die „Unbekannten" auf. Dann haben wir ein System linearer Gleichungen mit ebenso vielen Unbekannten, und wenn dessen Determinante nicht verschwindet, so können wir es auflösen. Das ist aber nun sofort zu erledigen. Nennen wir kurz *die* Gleichung, die den Erwartungswert eines Ausdrucks (30′) angibt, die der Unbekannten (30) entsprechende Gleichung. Um dann alle Unbekannten (34′) zu bestimmen, hat man eine *Gleichungskette* zur Verfügung, indem *jede Unbekannte direkt aus der ihr entsprechenden Gleichung gerechnet wird.* In der etwa der Unbekannten (30) entsprechenden Gleichung sind nämlich alle anderen Unbekannten von niedrigerer Faktorenzahl als (30); und die einer solchen Unbekannten mit kleinerer Faktorenzahl entsprechende Gleichung befindet sich bereits unter den vorher angeschriebenen Gleichungen, da ja für die Gleichungen die nach wachsender Faktorenzahl fortschreitende Reihenfolge (34′) gilt.

Um eine bestimmte Unbekannte, etwa (30) zu rechnen, brauchen wir auch nicht *alle* vorhergehenden Gleichungen heranzuziehen; sondern zunächst einmal nur die den Unbekannten mit kleinerer (also nicht gleicher) Faktorenzahl entsprechenden Gleichungen; und unter diesen wieder nur die gemäß den möglichen Zusammenfassungen der in (30) auftretenden Indices. Daraus ergibt sich die folgende Anweisung:

Soll ein Ausdruck (30) *als Erwartungswert eines Polynoms vom Gewicht ν in den* $\mathbf{M}$ *dargestellt werden, so bestimme man zunächst den Erwartungswert von* (30′), *d.h. man schreibe die* (30) *entsprechende Gleichung an. Dann schreibe man zu allen weiteren in dieser Gleichung auftretenden M-Ausdrücken die jeweils diesen Ausdrücken entsprechenden Gleichungen an, alles in der Reihenfolge steigender Faktorenzahl. Dann erhält man soviel Gleichungen als Unbekannte und kann nacheinander jede Unbekannte für sich an der ihr entsprechenden Gleichung rechnen.*

Soll z.B. der Ausdruck $M_1 M_2 M_3$ (wo $\nu = 6$, $\nu' = 3$) als Erwartungswert dargestellt werden, so sieht man, daß gemäß den Zusammenfassungen $1, 2, 3\,|\,3, 3\,|\,4, 1\,|\,1, 5\,|\,6\,|$ die folgenden Gleichungen bestehen: (die Reihenfolge der 2., 3., 4. Gleichung ist gleichgültig) (die $c_{\iota\varkappa}$ ergeben sich mittels des polynomischen Satzes)

$$\begin{aligned}
&\mathfrak{E}(\mathbf{M}_6) &&= c_{11} M_6 \\
&\mathfrak{E}(\mathbf{M}_1 \mathbf{M}_5) &&= c_{21} M_6 + c_{22} M_1 M_5 \\
&\mathfrak{E}(\mathbf{M}_4 \mathbf{M}_2) &&= c_{31} M_6 \qquad\qquad + c_{33} M_4 M_2 \\
&\mathfrak{E}(\mathbf{M}_3^2) &&= c_{41} M_6 \qquad\qquad\qquad\qquad + c_{44} M_3^2 \\
&\mathfrak{E}(\mathbf{M}_1 \mathbf{M}_2 \mathbf{M}_3) &&= c_{51} M_6 + c_{52} M_1 M_5 + c_{53} M_4 M_2 + c_{54} M_3^2 + c_{55} M_1 M_2 M_3.
\end{aligned}$$

Daraus werden nacheinander M_6; $M_1 M_5$, $M_2 M_4$, M_3^2; $M_1 M_2 M_3$ gefunden. — Insbesondere kann man für jedes Gewicht ν die entsprechenden Formelsysteme *ein für allemal* bereitstellen. Z.B. hat man für $\nu = 1$; $\nu = 2$; $\nu = 3$:

$$(37) \qquad \mathfrak{E}(\alpha) = a \qquad\qquad \text{daraus } a = \mathfrak{E}(\alpha)$$

$$(\overline{37'}) \quad \begin{aligned} &\mathfrak{E}(\mathbf{M}_2) = M_2 \\ &\mathfrak{E}(\alpha^2) = \frac{M_2}{m} + \frac{m-1}{m}\, a^2 \end{aligned} \qquad \text{daraus } (37') \quad \begin{aligned} &M_2 = \mathfrak{E}(\mathbf{M}_2) \\ &a^2 = \mathfrak{E}\left(\frac{m}{m-1}\alpha^2 - \frac{1}{m-1}\mathbf{M}_2\right) \end{aligned}$$

$$(\overline{37''})\qquad \begin{aligned}&\mathfrak{E}(\mathbf{M}_3) = M_3\\ &\mathfrak{E}(\mathbf{M}_2\,\alpha) = \frac{M_3}{m} + \frac{m-1}{m}\,a M_2\\ &\mathfrak{E}(\alpha^3) = \frac{M_3}{m^2} + 3\,\frac{m-1}{m}\,aM_2 + \frac{(m-1)(m-2)}{m^2}\,a^3\end{aligned}$$

$$\text{daraus }(37'')\qquad \begin{aligned}&M_3 = \mathfrak{E}(\mathbf{M}_3)\\ &M_2 a = \mathfrak{E}\left(\frac{m}{m-1}\,\alpha\,\mathbf{M}_2 - \frac{1}{m-1}\,\mathbf{M}_3\right)\\ &a^3 = \mathfrak{E}\left[\frac{1}{(m-1)(m-2)}\,(m^2\alpha^3 - 3m\alpha\mathbf{M}_2 + 2\mathbf{M}_3)\right].\end{aligned}$$

7. *Anwendungen.*

1) *Die beiden „Funktionen seltener Ereignisse."*

So nennt man die Poissonsche ψ-Funktion

$$v(x) \equiv \psi(x) = \frac{c^x}{x!}\,e^{-c}$$

und die Rückschlußwahrscheinlichkeit:

$$\chi(x) = \frac{x^\nu}{\nu!}\,e^{-\nu}.$$

Es bestehe die Hypothese, daß in (10)

$$v_1(x) = v_2(x) = \ldots = v_m(x) = \psi(x).$$

Für ψ gilt bekanntlich als einfachste Beziehung zwischen den Momenten $a = s^2$ oder $M_1^0 = M_2$, ehe wir das „übersetzen". Betrachten wir auch die χ-Funktion; für diese ist

$$a = \nu + 1,\ s^2 = M_2^0 - a^2 = (\nu+1)(\nu+2) - (\nu+1)^2 = a,$$

also *auch*

$$s^2 = a \text{ oder } M_1^0 = M_2.$$

Da dies die gleiche Bedingung für $\psi(x)$ und für $\chi(x)$ ist, ziehen wir noch die dritten Momente heran. Wir finden (unter Benutzung des dritten faktoriellen Momentes für $\psi(x)$) für das auf den Schwerpunt bezogene M_3

$$\text{für } \psi(x)\text{:}\quad M_3 = c \text{ und für } \chi(x)\text{:}\quad M_3 = 2(\nu+1)$$

(letzteres daraus, daß $M_3^0 = (\nu+1)(\nu+2)(\nu+3)$); also gelten die Gleichungen

$$(38)\qquad \text{für } \psi(x)\text{:}\quad a = M_3 \text{ und für } \chi(x)\text{:}\quad 2a = M_3.$$

Unter Vorwegnahme von (43′) folgt daraus die „Übersetzung"

$$\text{(39)} \qquad \text{für } \psi(x)\colon \quad \mathfrak{E}(\alpha) = \frac{m^2}{(m-1)(m-2)}\,\mathfrak{E}(\mathbf{M}_3)$$

$$\text{und für } \chi(x)\colon \quad 2\,\mathfrak{E}(\alpha) = \frac{m^2}{(m-1)(m-2)}\,\mathfrak{E}(\mathbf{M}_3).$$

2) *Summenbildung aus dreiwertiger Grundverteilung.*

Wir betrachten nun eine weniger einfache Aufgabe. Es seien p, q, r $(p+q+r=1)$ die Wahrscheinlichkeiten aus einer „Urne" 0, bzw. 1, bzw. 2 zu ziehen. Aus dieser Grundverteilung bilden wir durch n-fache Summenbildung eine der Bernoullischen analoge Verteilung, die die *Wahrscheinlichkeit dafür gibt, in n Zügen aus einer Urne mit einer durch p, q, r gegebenen Grundwahrscheinlichkeit die Summe x zu ziehen.* Rechnet man für die einfache Grundverteilung die Schwerpunktmomente, die wir vorübergehend mit M' bezeichnen, so erhält man

$$a' = q + 2r\,; \quad M_2' = (q+4r) - (q+2r)^2 = a' - a'^2 + 2r,$$

$$\text{(40)} \qquad M_3' = a' + 6r - 3a'M_2' - a'^3.$$

Nun ist es ein bekannter Satz, daß *Mittelwert und Streuung sich bei Summenbildung addieren.* Man zeigt aber in ganz analoger Weise die weniger bekannte Tatsache, daß dasselbe auch für das auf den Schwerpunkt bezogene dritte Moment gilt. *Auch die dritten Schwerpunktmomente addieren sich bei Summenbildung.* Es gilt also für die betrachtete Summenverteilungen mit den Momenten a, M_2, M_3

$$a = n(q+2r)\,; \quad M_2 = a - \frac{a^2}{n} + 2rn,$$

$$\text{(41)} \qquad M_3 = a + 6rn - \frac{3aM_2}{n} - \frac{a^3}{n^2}\,;$$

durch Elimination von r aus den zwei letzten Gleichungen (41) erhält man:

$$\text{(42)} \qquad 3M_2 - M_3 - \frac{3aM_2}{n} - 2a + \frac{3a^2}{n} - \frac{a^3}{n^2} = 0.$$

Das ist also eine Gleichung, die für diese Summenbildung aus dreiwertiger Grundverteilung charakteristisch ist, und die wir im Sinne unseres Verfahrens als das zu „übersetzende" Kriterium annehmen. Um dies durchzuführen, benutzen wir (37″) (wobei auf die vorübergehend geänderte Bezeichnung in Abschn.

6 zu achten ist) und erhalten unter Anwendung des „Verschiebungssatzes"

$$
\begin{aligned}
\mathfrak{E}(\mathbf{M}_3) &= \frac{(m-1)(m-2)}{m^2} M_3 \\
\mathfrak{E}(\alpha \mathbf{M}_2) &= \frac{m-1}{m^2} M_3 + \frac{m-1}{m} a M_2 \qquad (43) \\
\mathfrak{E}(\alpha^3) &= \frac{1}{m^2} M_3 + \frac{3}{m} a M_2 + a^3
\end{aligned}
$$

und daraus

$$
\begin{aligned}
M_3 &= \frac{m^2}{(m-1)(m-2)} \mathfrak{E}(\mathbf{M}_3) \\
a M_2 &= \frac{m}{m-1} \mathfrak{E}\left(\alpha \mathbf{M}_2 - \frac{\mathbf{M}_3}{m-2}\right) \qquad (43') \\
a^3 &= \mathfrak{E}\left(\alpha^3 - \frac{3\alpha \mathbf{M}_2}{m-1} + \frac{2\mathbf{M}_3}{(m-1)(m-2)}\right).
\end{aligned}
$$

Außerdem brauchen wir die schon bei der Lexis'schen Theorie benutzten Gleichungen (19). Setzt man aus (19) und (43′) in (40) ein, so ergibt sich, daß

$$
(44) \quad \mathfrak{E}\left[\frac{nm-1}{n(m-1)}\left(3\mathbf{M}_2 - \frac{3\alpha\mathbf{M}_2}{n} - \frac{nm-2}{n(m-2)}\mathbf{M}_3\right) - 2\alpha + \frac{3\alpha^2}{n} - \frac{\alpha^3}{n^2}\right] = 0.
$$

Denn es ist, wie man mittels (18) und (43) verifizieren kann, der Erwartungswert der in (44) gegebenen Prüffunktion gleich der linken Seite von (42), also wenn die Hypothese zutrifft, gleich Null.

Das ist eine erste *inhaltliche Erweitering der Lexis'schen Theorie.* Eine Anwendung von (44) kann man sich etwa so denken: In m Städten (Ländern) werden Gruppen von je n Familien betrachtet, die höchstens 2 Kinder haben, und in jeder Gruppe die Anzahl X_μ aller Kinder gezählt; aus diesen $X_1, X_2, \ldots, X_m$ bildet man α_{beob}, $\mathbf{M}_{2\text{beob}}$, $\mathbf{M}_{3\text{beob}}$, rechnet daraus den Klammerausdruck in (44) und sieht nach, ob dies entsprechend nahe an Null zu liegen kommt. Wenn ja, so spricht das für die Hypothese, daß es für alle m Städte, unabhängig davon, welche man ins Auge faßt, Grundwahrscheinlichkeiten für eine Familie gibt, 0, 1, 2 Kinder zu haben. (Die Einschränkung auf Familien mit höchstens 2 Kindern stört nicht, denn wenn es Wahrscheinlichkeiten $p', q', r', s', t', \ldots$ gibt, 0, 1, 2, 3, 4, ... Kinder zu haben, so gibt es auch Wahrscheinlichkeiten $p = \dfrac{p'}{p'+q'+r'}$, $q = \dfrac{q'}{p'+q'+r'}$,

$r = \frac{r'}{p'+q'+r'}$ dafür, 0, 1, 2 Kinder zu haben, wenn nur Familien mit höchstens zwei Kindern in Betracht gezogen werden, und umgekehrt.)

Zur genauen Beurteilung fehlt noch die Streuung von (44). Diese wäre nach den elementaren Methoden schon schwer zu rechnen. Nach den Näherungsmethoden von § 3 (64) ergibt sich, wenn man zur Abkürzung für den in der eckigen Klammer in (44) stehenden Ausdruck $F(\alpha, \mathbf{M}_2, \mathbf{M}_3)$ und

$$(45) \qquad 2 + \frac{3M_2}{n} - \frac{6a}{n} + \frac{3a^2}{n^2} = f_1; \quad \frac{3a}{n} - 3 = f_2 \text{ setzt:}$$

$$(46) \qquad \begin{aligned} m\mathfrak{S}\mathrm{tr}[F(\alpha, \mathbf{M}_2, \mathbf{M}_3)] &= M_2 f_1^2 + 2f_1(f_2 M_3 + M_4 - 3M_2^2) \\ &+ f_2^2(M_4 - M_2^2) + 2f_2(M_5 - 4M_2 M_3) + \\ &+ (M_6 - M_3^2 - 6M_4 M_2 + 9M_2^3). \end{aligned}$$

Dies ist kompliziert, und die Summanden sind in Bezug auf n von verschiedener Größenordnung. Infolgedessen kann es wünschenswert erscheinen, anzugeben, wie sich der in (46) gegebene Wert der Streuung abschätzen läßt. Teilt man (44) in zwei Teile, so daß auf der einen Seite nur $\mathbf{M}_3$ steht, so hat das Kriterium (44) die Form

$$(44') \qquad \frac{(nm-1)(nm-2)}{n^2(m-1)(m-2)} \mathfrak{E}(\mathbf{M}_3) = \mathfrak{E}\left[2\alpha - \frac{3\alpha^2}{n} + \frac{\alpha^3}{n^2} - \frac{3(nm-1)}{n(m-1)} \mathbf{M}_2 \left(1 - \frac{3\alpha}{n}\right)\right].$$

Es ist dann, immer als Näherungsformel bei Vernachlässigung von Gliedern mit $\frac{1}{m^2}$,

$$(46') \qquad m\mathfrak{S}\mathrm{tr}\,(\mathbf{M}_3) = M_6 - M_3^2 - 6M_4 M_2 + 9M_2^3,$$

dieser Ausdruck ist von der Ordnung n^3. Hingegen ist die Streuung der rechten Seite von (44′) von der Ordnung n^2, und man könnte hier also wieder ähnlich wie bei der Lexis'schen Theorie in (44′) die linke Seite, als „ungenauer", der als konstant angesehenen andern Seite gegenüberstellen, da für einigermaßen großes n die Streuung von $\mathbf{M}_3$ groß gegen die des links in (44′) stehenden Klammerausdrucks ist. Für unbegrenzt wachsendes n konvergiert die in (46) gegebene Gesamtstreuung gegen die von $\mathbf{M}_3$, und man kann daher näherungsweise (46′) an Stelle von (46) verwenden.

Wir wollen schließlich noch überlegen, was sich aussagen läßt, wenn man bei der in Rede stehenden Verteilung nur Beziehungen zwischen den Momenten bis zur zweiten Ordnung benutzt. Bilden wir die analogen Ausdrücke wie zu (17), so erhält man aus den ersten zwei Formeln (41) $s^2 - a + \frac{a^2}{n} = 2rn$, das ist für $r > 0$ stets positiv. Außerdem gilt $s^2 - 2a + \frac{a^2}{n} = -nq < 0$, wenn $q > 0$. Ist also r und q positiv, d.h. ist die Grundverteilung wirklich dreiwertig, so gilt für sie

$$(47)\qquad 2a - \frac{a^2}{n} > s^2 > a - \frac{a^2}{n}$$

oder nach „Übersetzung"

$$(48)\qquad \frac{nm}{nm-1}\,\mathfrak{E}\left(\alpha - \frac{\alpha^2}{n}\right) < \frac{m}{m-1}\,\mathfrak{E}(\sigma^2) < \frac{nm}{nm-1}\,\mathfrak{E}\left(2\alpha - \frac{\alpha^2}{n}\right).$$

Benutzt man also bloß die Momente erster und zweiter Ordnung, so kann man immer noch das Bestehen der Ungleichung (48) kontrollieren und darin ein Kriterium für die hinsichtlich der $v(x)$ gemachten Hypothese sehen.

§ 3. Näherungsweise Berechnung von Erwartungswerten. Umkehraufgabe.

8. *Stellung der Aufgabe.*

Wir haben Erwartungswerte — Streuungen sind ja auch Erwartungswerte — bisher nur für Polynome in den x_μ berechnet, nämlich für Ausdrücke (32) und daraus dann für Polynome in den Momenten und dergl. Abgesehen von dieser Beschränkung ist *auch in diesen Fällen* die Ausrechnung insbesondere der Streuungen, wenn man über das Einfachste hinausgeht, kompliziert und unübersichtlich. Es ist nun der Gedanke naheliegend, für die Funktion, deren Erwartungswert gesucht ist, eine Taylor-Entwicklung anzusetzen, die man nach einigen Gliedern abbricht. Dies Verfahren findet sich wiederholt in der Literatur, z.B. für die Gewinnung von Formeln für die Erwartungswerte von Potenzen einer Funktion (z.B. Wurzel) von Produkten, Quotienten zweier Funktionen usw. Es werden dabei aber auch gelegentlich falsche Formeln angegeben, bei denen die vernachlässigten Glieder keineswegs das implicite vorausgesetzte Verhalten zeigen.

Die hier folgenden, prinzipiell einfachen Überlegungen werden im Anschluß an die von R. v. Mises herrührenden „Ergänzungssätze zu den Gesetzen der großen Zahl" [5]) durchgeführt, zu deren erstem sie ein Korollar darstellen. Tätsächlich steht die näherungsweise Berechnung von Erwartungswerten in einer engen Beziehung zu dem formalen Kernpunkt des ersten Gesetzes der großen Zahlen.

Betrachtet man nämlich den gewogenen Durchschnitt aus m Beobachtungen

$$\alpha = \frac{c_1 x_1 + c_2 x_2 + \ldots + c_m x_m}{m},$$

wobei die Hypothese besteht, dass die μ-te Beobachtung x_μ der Verteilung $v_\mu(x)$ entspricht, so gilt, wenn, wie in (5), a_μ der Erwartungswert, s_μ^2 die Streuung von $v_\mu(x)$ ist und der Erwartungswert wie immer durch (9) und (14) erklärt ist:

$$\mathfrak{E}(\alpha) = \frac{c_1 a_1 + c_2 a_2 + \ldots + c_m a_m}{m},$$

$$\mathfrak{Str}(\alpha) = \mathfrak{E}(\alpha) - [\mathfrak{E}(\alpha)]^2 = \frac{c_1 s_1^2 + c_2 s_2^2 + \ldots + c_m s_m^2}{m^2}.$$

Es entsteht also $\mathfrak{E}(\alpha)$ aus α, indem man die x_μ durch ihre Erwartungswerte a_μ ersetzt. Die $\mathfrak{Str}(\alpha)$ ist klein von der Ordnung $\frac{1}{m}$, wenn die s_μ^2 beschränkt sind, derart daß ihre gewogene Summe mit m wächst. Fügt man zu diesen Tatsachen noch die sogenannte „Streuungsungleichung", so erhält man das „Erste Gesetz der großen Zahlen" in sehr allgemeiner Form.

Daß $\mathfrak{E}(\alpha)$ aus α entsteht, indem man nur die x_μ durch ihre Erwartungswerte ersetzt, überträgt sich nicht in dieser Einfachheit auf allgemeinere Funktionen der x_μ. Mann kann im allgemeinen nur zeigen, daß dies der Fall ist bis auf einen Fehler, der mit $\frac{1}{m}$ unendlich klein wird. Die Streuung aber zeigt in sehr allgemeinen Fällen ganz dasselbe Verhalten wie $\mathfrak{Str}(\alpha)$. Beispiele für beides haben die vorhergegangenen Ausführungen geboten. So war der Erwartungswert von (21) durch die linke Seite von (17') gegeben. Vernachlässigt man in (21) Glieder der Ordnung $\frac{1}{m}$, so erhält man genau das, was in (17') bei Ersatz von a und s^2 durch α und σ^2 resultiert. Analog stand es mit (42) und (44). Die Streuungen (22) und (46) gingen mit $\frac{1}{m}$ gegen Null usf.

---

[5]) l. c. [1]), S. 192 ff.

Um den Tatbestand allgemein zu untersuchen, benutzt man eine Taylorsche Entwicklung, aber nicht für *beliebige* Funktionen der Beobachtungen — dann entstehen unter Umständen ganz falsche Formeln. — Man kann dies Verfahren aber immer anwenden, wenn es sich, wie z.B. in den Ergänzungssätzen, um *Funktionen der relativen Häufigkeiten der Beobachtungen handelt.*

Wir greifen für die Angabe einer statistischen Beobachtung auf die Vorstellungen und Bezeichnungen von § 1 zurück. Es bestehen k voneinander verschiedene Möglichkeiten (2), auf die sich die m Beobachtungen mit den relativen Häufigkeiten (2′) verteilen. Es gibt m theoretische Verteilungen (3). Es ist wie dort durch (4) $q_\varkappa^{(\mu)}$ und $q_\varkappa$ erklärt. Die Erwartungswerte werden hinsichtlich der aus den $v_\mu(x)$ nach (14) zu bildenden, m-dimensionalen Verteilung $w(x_1, x_2, \ldots, x_m)$ gerechnet.

Man untersucht nun Erwartungswerte und Streuungen von, nur geringen Einschränkungen unterliegenden, k-dimensionalen Funktionen der relativen Häufigkeiten

$$f\left(\frac{m_1}{m}, \frac{m_2}{m}, \ldots, \frac{m_k}{m}\right). \tag{49}$$

Dann geht nach den erwähnten Ergänzungssätzen mit $m \to \infty$ *der Erwartungswert einer solchen Funktion gegen* $f(q_1, q_2, \ldots q_m)$, *also gegen die Funktion der Erwartungswerte und die Streuung von* (49) *gegen Null.*

Wir wollen hier aber, da wir gerade das brauchen, über dies Resultat der Ergänzungssätze *hinaus,* die Abweichung zwischen (49) und $f(q_1, q_2, \ldots q_k)$ *wirklich angeben* und analog die mit $\frac{1}{m}$ verschwindende *Streuung von* (49), wie dies für Spezialfälle schon oft geschehen ist[6]). Es sei noch erwähnt, daß in den Ergänzungssätzen die $v_\mu(x)$ einander gleich genommen werden. Man kann aber ohne weiteres die allgemeineren *Annahmen verschiedener* $v_\mu(x)$ *treffen.*

9. *Erwartungswerte und Streuungen von Funktionen der relativen Häufigkeiten.*

---

[6]) Vgl. z.B. die bekannten Formeln für Erwartungswert und Streuung von Produkten, ferner R. v. Mises: Über einige Abschätzungen von Erwartungswerten [Journal für reine u. angew. Math. **165** (1931), § 5 ff]; F. Bernstein: Über den mittleren Fehler der Potenzmomente [Blätter für Versicherungsmathematik **30** (1930), 365—77]; ferner: A. A. Tschuprow, Grundbegriffe und Grundprobleme der Korrelationstheorie [Leipzig 1925].

Es sei eine Funktion (49) gegeben, die so beschaffen ist, daß sie sich im Punkte $(q_1, q_2, \ldots, q_k)$ in eine Taylorsche Reihe entwickeln läßt. Wir bezeichnen

$$(50) \quad \frac{m_\varkappa}{m} = z_\varkappa; \quad \frac{\partial f}{\partial z_j}(z_1 = q_1, z_2 = q_2, \ldots, z_k = q_k) = f_j; \text{ usf.}$$

und setzen

$$(51) \quad f(z_1, \ldots, z_k) = f(q_1, \ldots, q_k) + \sum_\varkappa f_\varkappa(z_\varkappa - q_\varkappa)$$
$$+\tfrac{1}{2}\sum_{j,\varkappa} f_{j\varkappa}(z_j - q_j)(z_\varkappa - q_\varkappa) + \tfrac{1}{6}\sum_{j,\varkappa,\lambda} f_{j\varkappa\lambda}(z_j - q_j)(z_\varkappa - q_\varkappa)(z_\lambda - q_\lambda) + \ldots$$

Wir rechnen links und rechts in (51) den Erwartungswert. Dazu benötigen wir den Erwartungswert von $z_\varkappa$; von $(z_\varkappa - q_\varkappa)^2$, $(z_\varkappa - q_\varkappa)(z_j - q_j)$; von $(z_\varkappa - q_\varkappa)^3$, usf. Nach bekannten Formeln, die für die Summenbildung aus m verschiedenen Alternativen gelten (Poissonscher Fall vgl. S. [12] 287) ist:

$$(52) \quad \mathfrak{E}\left(\frac{m_\varkappa}{m}\right) = \frac{1}{m}\sum_1^m q_\varkappa^{(\mu)} = q_\varkappa; \quad \mathfrak{E}\left(\frac{m_\varkappa}{m} - q_\varkappa\right)^2 = \frac{q_\varkappa}{m} - \frac{\sum\limits_1^m \left(q_\varkappa^{(\mu)}\right)^2}{m^2};$$

$$\mathfrak{E}\left[\left(\frac{m_\varkappa}{m} - q_\varkappa\right)\left(\frac{m_j}{m} - q_j\right)\right] = -\frac{1}{m^2}\sum_1^m q_\varkappa^{(\mu)} q_j^{(\mu)};$$

$$\mathfrak{E}[(z_\varkappa - q_\varkappa)]^3 = \frac{1}{m^3}\sum_1^m q_\varkappa^{(\mu)}\left(1 - q_\varkappa^{(\mu)}\right)\left(1 - 2q_\varkappa^{(\mu)}\right) \quad \text{usf.}$$

Gemäß dem Bau der Momente einer Bernoullischen (bzw. Poissonschen) Verteilung sind die Erwartungswerte der Glieder ν-ter Ordnung, wie etwa von $\left(\frac{m_k}{m} - q_k\right)^\nu$ etc. von der Größenordnung $m^{-\lambda}$ wobei je nach dem Wert von ν $\lambda = \frac{\nu}{2}$ bzw. $\lambda = \frac{\nu + 1}{2}$.

Wenn wir daher bei Gliedern der Ordnung $\frac{1}{m}$ abbrechen wollen, so können wir die von den kubischen Gliedern herrührenden Ausdrücke bereits vernachlässigen. Und wir erhalten

$$\mathfrak{E}[f(z_1, \ldots, z_k)] = f(q_1, \ldots, q_k) + 0 + \frac{1}{2m}\sum_1^k f_{\varkappa\varkappa}\sum_1^m \frac{q_\varkappa^{(\mu)}\left(1 - q_\varkappa^{(\mu)}\right)}{m}$$
$$- \frac{1}{2m}\sum_{j \neq \varkappa}^{1 \ldots k} f_{j\varkappa}\sum_1^m \frac{q_j^{(\mu)} q_\varkappa^{(\mu)}}{m}$$

oder

$$(53') \quad \mathfrak{E}\left[f\left(\frac{m_1}{m},\ldots,\frac{m_k}{m}\right)\right] = f(q_1,\ldots q_k) + \frac{1}{2m}\left[\sum_1^k f_{\varkappa\varkappa} q_\varkappa - \sum_1^k f_{j\varkappa} \sum_1^m \frac{q_j^{(\mu)} q_\varkappa^{(\mu)}}{m}\right].$$

Dies ist eine sehr verwendbare *Näherungsformel.* Falls alle $q_\varkappa^{(\mu)}$ $(\mu = 1,\ldots,m)$ einander gleich sind, ergibt sich:

$$(53) \quad \mathfrak{E}\left[f\left(\frac{m_1}{m},\ldots,\frac{m_k}{m}\right)\right] = f(q_1,\ldots,q_k) + \frac{1}{2m}\left[\sum_1^k f_{\varkappa\varkappa} q_\varkappa - \sum_1^k f_{j\varkappa} q_j q_\varkappa\right],$$

eine Formel, van der wir mehrfach Gebrauch machen werden.

Um den jeweils beim Abbrechen entstehenden Fehler abzuschätzen, braucht man die Taylorsche Formel. Setzt man z.B. die Taylorsche *Formel* zweiter Ordnung an

$$(51') \quad f(z_1,\ldots,z_k) = f(q_1,\ldots,q_k) + \sum_1^k f_\varkappa (z_\varkappa - q_\varkappa) + \tfrac{1}{2}\sum_1^k f_{j\varkappa}(z_j - q_j)(z_\varkappa - q_\varkappa),$$

wobei jetzt $f_{j\varkappa}$ einen Zwischenwert bedeutet, und die $\frac{\partial^2 f}{\partial z_j \partial z_\varkappa}$ im ganzen k-dimensionalen Intervall (0, 1) beschränkt sind,

$$(55') \quad \left|\frac{\partial^2 f}{d z_j\, \partial z_\varkappa}\right| < F_{j\varkappa},$$

so ist:

$$(55'') \quad 2\left|(z_j - q_j)(z_\varkappa - q_\varkappa)\right| \leqq (z_j - q_j)^2 + (z_\varkappa - q_\varkappa)^2.$$

Daraus folgt

$$(54) \quad \left|\mathfrak{E}(f) - f(q_1,\ldots,q_k)\right| < \frac{1}{2m}\sum_{\varkappa=1}^k q_\varkappa(1 - q_\varkappa)\sum_{j=1}^k F_{j\varkappa}$$

oder im allgemeinen Fall verschiedener $q_\varkappa^{(\mu)}$

$$(54') \quad \left|\mathfrak{E}(f) - f(q_1,\ldots,q_k)\right| < \frac{1}{2m}\sum_{\varkappa=1}^k \sum_{\mu=1}^m \frac{q_\varkappa^{(\mu)}(1 - q_\varkappa^{(\mu)})}{m} \sum_{j=1}^k F_{j\varkappa}$$

Die rechte Seite geht mit wachsendem m gegen Null. Ebenso zeigt man mittels der Taylorschen Formel 3. Ordnung, wenn man voraussetzt, daß die Ableitungen 3. Ordnung im k-dimensionalen Einheitswürfel beschränkt sind, das Resultat, daß die Differenz der linken und rechten Seite in (53) mit $\frac{1}{m^2}$ gegen Null geht.

Wir wollen nun eine zu (53) analoge Näherungsformel für die *Streuung* von f angeben, wobei wir wieder Glieder von höherer als erster Ordnung in $\frac{1}{m}$ vernachlässigen. Es ist

$$\mathfrak{Str}[f(z_1, \ldots, z_k)] = \mathfrak{E}\{[f(z_1, \ldots, z_k) - \mathfrak{E}(f)]^2\}.$$

Setzt man dafür angenähert

$$\mathfrak{E}\{[f(z_1, \ldots, z_k) - f(q_1, \ldots, q_k)]^2\},$$

so hat man nach (53′) in der eckigen Klammer das Glied

$$\frac{1}{2m}\left[\Sigma f_{\varkappa\varkappa} q_\varkappa - \Sigma f_{j\varkappa} \Sigma \frac{1}{m} q_j^{(\mu)} q_\varkappa^{(\mu)}\right]$$

weggelassen. Dieser Fehler führt aber bei der Berechnung der Streuung nur zu einem Unterschied der Ordnung $\frac{1}{m^2}$, denn beim Quadrieren diese Zusatzgliedes tritt eo ipso $\frac{1}{m^2}$ auf, und beim Erwartungswert des doppelten Produkts multipliziert sich dieses Zusatzglied der Ordnung $\frac{1}{m}$ mit $\mathfrak{E}(f) - f(q_1, \ldots, q_k)$ und diese Differenz ist wieder von der Ordnung $\frac{1}{m}$, also auch das doppelte Produkt von der Ordnung $\frac{1}{m^2}$.

Schreiben wir nun nach (51)

$$(56) \qquad [f(z_1, \ldots, z_k) - f(q_1, \ldots, q_k)]^2 \sim [\sum_\varkappa f_\varkappa (z_\varkappa - q_\varkappa)]^2,$$

so erhält man daraus nach (52)

$$\mathfrak{E}\{[f - f(q_1, \ldots, q_k)]^2\} = \frac{1}{m}(\Sigma q_\varkappa f_\varkappa^2 - \Sigma f_j f_\varkappa q_j q_\varkappa) = \frac{1}{m}[\Sigma q_\varkappa f_\varkappa^2 - (\Sigma f_\varkappa q_\varkappa)^2].$$

Somit erhalten wir die *Näherungsformeln für die Streuung*:

$$(57) \quad \mathfrak{Str}(f) = \frac{1}{m}\left(\sum_1^k q_\varkappa f_\varkappa^2 - \sum_1^k f_j f_\varkappa q_j q_\varkappa\right) = \frac{1}{m}\left[\sum_1^k q_\varkappa f_\varkappa^2 - \left(\sum_1^k q_\varkappa f_\varkappa\right)^2\right]$$

und allgemeiner

$$(57') \quad \mathfrak{Str}(f) = \frac{1}{m}\left[\sum_1^k f_\varkappa^2 \sum_1^m \frac{1}{m} q_\varkappa^{(\mu)} - \sum_1^k f_j f_\varkappa \frac{\sum_1^m q_j^{(\mu)} q_\varkappa^{(\mu)}}{m}\right] = \frac{1}{m}\left[\sum_1^k f_\varkappa^2 \frac{\sum_1^m q_\varkappa^{(\mu)}}{m} - \frac{1}{m}\sum_1^m\left(\sum_1^k f_\varkappa q_\varkappa^{(\mu)}\right)^2\right].$$

Man sieht, daß es keinen Sinn hat, bei Berechnung der Streuung von f auf der rechten Seite in (51) Glieder von höherer Ordnung in $\frac{1}{m}$ zu berücksichtigen, da man ja doch auf der linken Seite in (57) bereits die analoge Vernachlässigung gemacht hat.

Praktisch verwendet wird vor allem (57) und (53) werden. Man sieht aus (57), daß es auch keinen rechten Zweck hätte (53) noch mit Gliedern höherer Ordnung in $\frac{1}{m}$ zu rechnen, da die *zufällige oder statistische* Abweichung, die durch $\sqrt{\mathfrak{S}\mathfrak{tr}(f)}$ gegeben ist, ja sogar von der Ordnung $\frac{1}{\sqrt{m}}$ ist, also sogar „größer'' als das Zusatzglied der Ordnung $\frac{1}{m}$ in (53). Diese Tatsache wird natürlich auch nicht geändert, wenn man (57) noch bis zu Gliedern nächster Ordnung verfolgen würde, da die Entwicklung von $\sqrt{\mathfrak{S}\mathfrak{tr}(f)}$ ja doch mit $\frac{1}{\sqrt{m}}$ anfängt.

Durch das Vorhergehende ist u. a. auch gezeigt, daß die v. Misesschen *Ergänzungsgesetze zu den Gesetzen der großen Zahlen richtig bleiben, wenn die theoretische Grundverteilung in den m Beobachtungen von Beobachtung zu Beobachtung eine andere ist,* also wenn $q_\varkappa^{(\mu)}$ $(\mu=1,\ldots,m;\ \varkappa=1,\ldots,k)$ statt der $q_\varkappa$ auftreten.

10. *Formeln für Funktionen der Nullmomente.*

Diese folgen natürlich aus den in 9 abgeleiteten Formeln, da die Momente spezielle Funktionen der relativen Häufigkeiten sind. Es ist jedoch bequem, diese Formeln hier direkt herzuleiten.

Es sei f eine Funktion der auf den Nullpunkt bezogenen empirischen Momente bis zur r-ten Ordnung also $f(\mathbf{M}_1^0, \mathbf{M}_2^0, \ldots, \mathbf{M}_r^0)$ mit den Bezeichnungen 1. Sind dann die gegebenen theoretischen Verteilungen voneinander verschieden, so gilt

$$(58') \qquad \mathfrak{E}\left(\frac{x_1^r + x_2^r + \ldots + x_m^r}{m}\right) = \frac{1}{m}\left(M_r^{0(1)} + M_r^{0(2)} + \ldots + M_r^{0(m)}\right),$$

wenn $M_\varrho^{0(\mu)} = \sum\limits_x x^\varrho v_\mu(x)$ das ϱ-te Moment der μ-ten Verteilung $v_\mu(x)$ bezeichnet. Insbesondere gilt, wenn alle $v_\mu(x)$ gleich sind:

$$(58) \qquad \mathfrak{E}(\mathbf{M}_\varrho^0) = M_\varrho^0.$$

Setzen wir

$$(59) \qquad \frac{1}{m}\left(M_\varrho^{0(1)} + M_\varrho^{0(2)} + \ldots + M_\varrho^{0(m)}\right) = M_\varrho^0,$$

dann gilt (58) *allgemein,* wie aus (58′) folgt. Wir setzen an

$$f(\mathbf{M}_1^0, \ldots, \mathbf{M}_r^0) = f(M_1^0, \ldots, M_r^0) + \sum_{\varrho=1}^{r} f_\varrho(\mathbf{M}_\varrho^0 - M_\varrho^0) + \\ + \tfrac{1}{2} \sum_{\varrho,\sigma}^{1..r} f_{\varrho\sigma}(\mathbf{M}_\varrho^0 - M_\varrho^0)(\mathbf{M}_\sigma^0 - M_\sigma^0),$$

wenn

$$(59') \qquad f_\varrho = \frac{\partial f}{\partial \mathbf{M}_\varrho} (\mathbf{M}_1^0 = M_1^0, \ldots, \mathbf{M}_r^0 = M_r^0), \text{ usf.}$$

Da die Nullmomente nichts anderes als lineare Kombinationen der relativen Häufigkeiten sind, so bedeutet das Abbrechen nach dem zweiten Glied hier dasselbe wie in **9** und auch bez. der Restabschätzung gilt das gleiche wie dort.

Wir brauchen nun auch hier die Erwartungswerte der Ausdrücke zweiter Ordnung $\mathbf{M}_\varrho^0 \mathbf{M}_\sigma^0$ und $(\mathbf{M}_\varrho^0)^2$. Ist zunächst (10) erfüllt, so gilt, wenn man (33) in Abschn. 6 benutzt

$$(60') \quad \mathfrak{E}(\mathbf{M}_\varkappa^0 \mathbf{M}_j^0) = \frac{1}{m^2}(mM_{\varkappa+j}^0 + m(m-1)M_\varkappa^0 M_j^0) = \\ M_\varkappa^0 M_j^0 + \frac{1}{m}(M_{\varkappa+j}^0 - M_\varkappa^0 M_j^0)$$

$$\mathfrak{E}\{(\mathbf{M}_\varkappa^0)^2\} = (M_\varkappa^0)^2 + \frac{1}{2m}(M_{2\varkappa}^0 - (M_\varkappa^0)^2)$$

und daraus

$$(60) \qquad \mathfrak{E}[(\mathbf{M}_\varkappa^0 - M_\varkappa^0)(\mathbf{M}_j^0 - M_j^0)] = \frac{1}{m}(M_{\varkappa+j}^0 - M_\varkappa^0 M_j^0),$$

somit die Näherungs-Formel, die zusammen mit (62) neben (53) und (57) tritt

$$(61) \quad \mathfrak{E}[f(\mathbf{M}_1^0, \ldots, \mathbf{M}_r^0)] = f(M_1^0, \ldots, M_r^0) + \\ + \frac{1}{2m} \sum_{\varrho,\sigma}^{1\ldots r} f_{\varrho\sigma}(M_{\varrho+\sigma}^0 - M_\varrho^0 M_\sigma^0)$$

und ganz analog wie früher für die Streuung

$$(62) \quad m\mathfrak{Str}(f) = \sum_{\varrho,\sigma}^{1..r} f_\varrho f_\sigma (M_{\varrho+\sigma}^0 - M_\varrho^0 M_\sigma^0) = \\ \sum_{\varrho,\sigma}^{1..r} f_\varrho f_\sigma M_{\varrho+\sigma}^0 - \left(\sum_{\varrho=1}^{r} f_\varrho M_\varrho^0\right)^2.$$

Als Beispiel rechnen wir

$$\mathfrak{E}(\alpha^n) = a^n + \frac{n(n-1)}{2m} a^{n-2}(M_2^0 - a^2) = a^n + \frac{n(n-1)}{2m} a^{n-2} M_2$$

in Übereinstimmung mit der 3. Formel (42).

Sind die $v_\mu(x)$ untereinander verschieden, so ist

$$m\mathfrak{E}(\mathbf{M}_i^0\mathbf{M}_j^0)=\sum_\mu M_{i+j}^{0(\mu)}+\sum_{\mu\neq\nu}M_i^{0(\mu)}M_j^{0(\nu)}$$

und daraus, wenn man für M_i^0, M_j^0 den Wert (59) setzt,

$$(60')\quad \mathfrak{E}[(\mathbf{M}_i^0-M_i^0)(\mathbf{M}_j^0-M_j^0)]=\frac{1}{m^2}\Big\{\sum_\mu M_{i+j}^{0(\mu)}-\sum_\mu M_i^{0(\mu)}M_j^{0(\mu)}\Big\}$$

und damit

$$(61')\quad \mathfrak{E}[f(\mathbf{M}_1^0,\ldots,\mathbf{M}_r^0)]=f(M_1^0,\ldots,M_r^0)+$$
$$+\frac{1}{2m^2}\sum_{\varrho,\sigma}^{1..r}f_{\varrho\sigma}\Big(\sum_\mu M_{\varrho+\sigma}^{0(\mu)}-\sum_\mu M_\varrho^{0(\mu)}M_\sigma^{0(\mu)}\Big)$$

und für die Streuung:

$$(62')\qquad m^2\,\mathfrak{S}\mathrm{tr}(f)=\sum_{\varrho,\sigma}^{1..r}f_\varrho f_\sigma\sum_\mu\big(M_{\varrho+\sigma}^{0(\mu)}-M_\varrho^{0(\mu)}M_\sigma^{0(\mu)}\big).$$

11. *Formeln für Funktionen der Schwerpunktmomente.*

Gegeben sei nun eine Funktion f der Schwerpunktmomente $\mathbf{M}_\varrho$ $(\varrho=2,\ldots,r)$ und des ersten Nullpunktmomentes $\alpha=\mathbf{M}_1^0$, das wir der Bequemlichkeit der Anwendung halber dazu nehmen. Wenn man dann die Erwartungswerte und Streuungen nur bis auf Glieder in $\frac{1}{m^2}$ haben will, so kann man, statt von den Formeln (53), (57) auszugehen, wie folgt vorgehen. Wir setzen die ersten Glieder der Taylorschen Entwicklung von f an und werden rechts für die Erwartungswerte die Formeln von Abschn. 2 verwenden. Setzt man analog wie früher

$$\frac{\partial f}{\partial\alpha}(\alpha=a,\ \mathbf{M}_2=M_2,\ldots\ \mathbf{M}_r=M_r)=f_1,$$

$$\frac{\partial f}{\partial\mathbf{M}_\varrho}(\alpha=a,\ \mathbf{M}_2=M_2,\ldots\ \mathbf{M}_r=M_r)=f_\varrho\qquad(\varrho=2,\ldots,r)\ \text{usf.},$$

so gilt, wenn man nach Gliedern zweiter Ordnung abbricht,

$$(63)\ f(\alpha,\mathbf{M}_2,\ldots,\mathbf{M}_r)=f(a,M_2,\ldots,M_r)+(\alpha-a)f_1+\sum_{\varrho=2}^{r}f_\varrho(\mathbf{M}_\varrho-M_\varrho)$$
$$+\tfrac{1}{2}f_{11}(\alpha-a)^2+(\alpha-a)\sum_{\varrho=2}^{r}f_{1\varrho}(\mathbf{M}_\varrho-M_\varrho)$$
$$+\tfrac{1}{2}\sum_{\varrho,\sigma}^{2\ldots r}f_{\varrho\sigma}(\mathbf{M}_\varrho-M_\varrho)(\mathbf{M}_\sigma-M_\sigma).$$

Beginnen wir mit der Streuung: Es ist

$$[f(\alpha, \mathbf{M}_2, \ldots, \mathbf{M}_r) - f(a, M_2, \ldots, M_r)]^2$$
$$= [(\alpha - a) f_1 + \sum_{\varrho=r}^{r} f_\varrho (\mathbf{M}_\varrho - M_\varrho)]^2.$$

Wir suchen den Erwartungswert der beiden Seiten dieser Gleichung.

Es ist nun

$$(64') \qquad \mathfrak{E}(\alpha - a)^2 = \mathfrak{Str}(\alpha) = \frac{M_2}{m}, \text{ das gilt genau.}$$

Nun brauchen wir

$$\mathfrak{E}[(\alpha - a)(\mathbf{M}_\varrho - M_\varrho)] \text{ und } \mathfrak{E}[(\mathbf{M}_\varrho - M_\varrho)(\mathbf{M}_\sigma - M_\sigma)].$$

Wir setzen wie früher, um (53) anzuwenden, $\frac{m_k}{m} = z_k$ und die Merkmalwerte gleich y_k, dann ist

$$\mathbf{M}_\varrho = z_1(y_1 - \alpha)^\varrho + \ldots + z_k(y_k - \alpha)^\varrho \text{ mit } \alpha = y_1 z_1 + \ldots + y_k z_k.$$

Es ist dann, wenn man in (53) mit $f = (\alpha - a) \cdot (\mathbf{M}_\varrho - M_\varrho)$ (als Funktion der z_k betrachtet) eingeht,

$$\frac{\partial f}{\partial z_j} = y_j (\mathbf{M}_\varrho - M_\varrho) + (\alpha - a) \frac{\partial \mathbf{M}_\varrho}{\partial z_j},$$
$$\frac{\partial^2 f}{\partial z_j \partial z_\varkappa} = y_j \frac{\partial \mathbf{M}_\varrho}{\partial z_\varkappa} + y_\varkappa \frac{\partial \mathbf{M}_\varrho}{\partial z_j} + (\alpha - a) \frac{\partial^2 \mathbf{M}_\varrho}{\partial z_j d z_\varkappa}.$$

Weiters:

$$\frac{\partial \mathbf{M}_\varrho}{\partial z_j} = (y_j - \alpha)^\varrho - \varrho y_j \sum_{\varkappa=1}^{k} z_\varkappa (y_\varkappa - \alpha)^{\varrho-1} = (y_j - \alpha)^\varrho - \varrho y_j \mathbf{M}_{\varrho-1},$$

somit:

$$f_{j\varkappa} = y_j (y_\varkappa - a)^\varrho + y_\varkappa (y_j - a)^\varrho - 2\varrho y_j y_\varkappa M_{\varrho-1},$$
$$f_{jj} = 2 y_j \cdot \{(y_j - a)^\varrho - y_j \varrho M_{\varrho-1}\}.$$

Das Zusatzglied rechts in (53) wird also:

$$\frac{1}{2m} \{ \sum_j 2 q_j y_j [(y_j - a)^\varrho - \varrho y_j M_{\varrho-1}]$$
$$- \sum_{j\varkappa} q_j q_\varkappa [y_j (y_\varkappa - a)^\varrho + y_\varkappa (y_j - a)^\varrho - 2\varrho y_j y_\varkappa M_{\varrho-1}] \}.$$

Nun ist, wenn man in diesem Ausdruck die Faktoren von $M_{\varrho-1}$ beachtet,

$$\sum_j y_j^2 q_j - \sum_{j,\varkappa} y_j y_\varkappa q_j q_\varkappa = M_2,$$

und für die Summe der anderen Glieder kommt $2M_{\varrho+1}$ (unter

Beachtung von $a = \sum_{\varkappa} y_\varkappa q_\varkappa$). Es ist also näherungsweise:

$$(64'') \qquad \mathfrak{E}[(\alpha - a)(\mathbf{M}_\varrho - M_\varrho)] = \frac{1}{m}(M_{\varrho+1} - \varrho M_{\varrho-1} M_2).$$

Nun setzen wir weiter in (53) $f = (\mathbf{M}_\varrho - M_\varrho)(\mathbf{M}_\sigma - M_\sigma)$, dann ist:

$$\frac{\partial f}{\partial z_j} = \frac{\partial \mathbf{M}_\varrho}{\partial z_j}(\mathbf{M}_\sigma - M_\sigma) + \frac{\partial \mathbf{M}_\sigma}{\partial z_j}(\mathbf{M}_\varrho - M_\varrho),$$

$$\frac{\partial^2 f}{\partial z_j \partial z_\varkappa} = \frac{\partial^2 \mathbf{M}_\varrho}{\partial z_j \partial z_\varkappa}(\mathbf{M}_\sigma - M_\sigma)$$
$$+ \frac{\partial \mathbf{M}_\varrho}{\partial z_j}\frac{\partial \mathbf{M}_\sigma}{d z_\varkappa} + \frac{\partial \mathbf{M}_\sigma}{d z_j}\frac{\partial \mathbf{M}_\varrho}{\partial z_\varkappa} + \frac{\partial^2 \mathbf{M}_\sigma}{\partial z_j \partial z_\varkappa}(\mathbf{M}_\varrho - M_\varrho),$$

$$f_{j\varkappa} = \{(y_j - a)^\varrho - \varrho y_j M_{\varrho-1}\} \cdot \{(y_\varkappa - a)^\sigma - \varrho y_\varkappa M_{\sigma-1}\} + \{\quad\} \cdot \{\quad\}.$$

Setzt man das in das Zusatzglied in (53) ein und vereinfacht genau analog wie in dem eben durchgerechneten Fall, so ergibt sich

$$(64''') \quad \mathfrak{E}[(\mathbf{M}_\varrho - M_\varrho)(\mathbf{M}_\sigma - M_\sigma)] =$$
$$\frac{1}{m}(M_{\varrho+\sigma} - M_\varrho M_\sigma - \varrho M_{\varrho-1} M_{\sigma+1} - \sigma M_{\sigma-1} M_{\varrho+1} + \varrho\sigma M_{\varrho-1} M_{\sigma-1} M_2).$$

Damit erhalten wir mit einem Fehler der Ordnung $\frac{1}{m^2}$ die wichtige Formel:

$$(64) \quad \mathfrak{Str}[f(\alpha, \mathbf{M}_2, \ldots, \mathbf{M}_r)]$$
$$= \frac{1}{m}\Big\{M_2 f_1^2 + 2 f_1 \sum_{\varrho=2}^{r} f_\varrho (M_{\varrho+1} - \varrho M_2 M_{\varrho-1})$$
$$+ \sum_{\varrho,\sigma}^{2..r} f_\varrho f_\sigma (M_{\varrho+\sigma} - M_\varrho M_\sigma - \varrho M_{\varrho-1} M_{\sigma+1}$$
$$- \sigma M_{\sigma-1} M_{\varrho+1} + \varrho\sigma M_{\varrho-1} M_{\sigma-1} M_2)\Big\}.$$

Nun wollen wir mit der gleichen Annäherung den *Erwartungswert* von $f(\alpha, \mathbf{M}_2, \ldots, \mathbf{M}_r)$ rechnen; aus (63) erhalten wir:

$$(63') \quad \mathfrak{E}(f) = f(a, M_2, \ldots, M_r) + f_1 \mathfrak{E}(\alpha - a)$$
$$+ \sum_\varrho f_\varrho \mathfrak{E}(\mathbf{M}_\varrho - M_\varrho) + \tfrac{1}{2} f_{11} \mathfrak{E}(\alpha - a)^2$$
$$+ \sum_\varrho f_{1\varrho} \mathfrak{E}[(\alpha - a)(\mathbf{M}_\varrho - M_\varrho)] + \tfrac{1}{2}\sum_{\varrho,\sigma} \mathfrak{E}[\mathbf{M}_\varrho - M_\varrho)(\mathbf{M}_\sigma - M_\sigma)].$$

Hier ist nur $\mathfrak{E}(\mathbf{M}_\varrho - M_\varrho)$ neu zu rechnen. Da brauchen wir neben $\frac{\partial \mathbf{M}_\varrho}{\partial z_j}$ auch die *zweiten* Ableitungen

$$\frac{\partial^2 \mathbf{M}_\varrho}{\partial z_j \partial z_\varkappa} = \varrho(\varrho-1) y_j y_\varkappa \sum_\varkappa (y_\varkappa - \alpha)^{\varrho-2} z_\varkappa - \varrho(y_j-\alpha)^{\varrho-1} y_\varkappa - \varrho(y_\varkappa-\alpha)^{\varrho-1} y_j$$

und somit, mit $f = \mathbf{M}_\varrho - M_\varrho$

$$f_{j\varkappa} = \varrho(\varrho-1) y_j y_\varkappa M_{\varrho-2} - \varrho\{y_\varkappa(y_j-a)^{\varrho-1} + y_j(y_\varkappa-a)^{\varrho-1}\},$$
$$f_{jj} = \varrho(\varrho-1) y_j^2 M_{\varrho-2} - 2\varrho y_j (y_j-a)^{\varrho-1}.$$

Setzt man in (53) ein und benutzt analoge Vereinfachungen wie früher, so folgt:

$$(64^{\mathrm{IV}}) \qquad \mathfrak{E}(\mathbf{M}_\varrho - M_\varrho) = \frac{1}{m}\left[\binom{\varrho}{2} M_{\varrho-2} M_2 - \varrho M_\varrho\right].$$

Alle anderen in (63′) auftretenden Erwartungswerte sind bereits gerechnet. Man erhält im ganzen die neben (64) tretende gebrauchsfertige Formel:

$$(65) \qquad \mathfrak{E}(f) = f(a, M_2, \ldots, M_r) + \frac{1}{2m} f_{11} M_2$$
$$+ \frac{1}{m} \cdot \left\{ \sum_{\varrho=2}^{r} f_\varrho \left[\binom{\varrho}{2} M_{\varrho-2} M_2 - \varrho M_\varrho\right] + f_{1\varrho}[M_{\varrho+1} - \varrho M_2 M_{\varrho-1}] \right\}$$
$$+ \frac{1}{2m} \sum_{\varrho,\sigma}^{2..r} f_{\varrho\sigma} \big(M_{\varrho+\sigma} - M_\varrho M_\sigma - \varrho M_{\varrho-1} M_{\sigma+1} - \sigma M_{\sigma-1} M_{\varrho+1} + \varrho\sigma M_{\varrho-1} M_{\sigma-1} M_2\big).$$

Zur teilweisen Probe wählen wir $f = \mathbf{M}_\varrho \mathbf{M}_\sigma$, dann ist $f_\sigma = M_\varrho$, $f_\varrho = M_\sigma$, $f_{\sigma\varrho} = 1$, und man erhält aus (65)

$$(66) \qquad \mathfrak{E}(\mathbf{M}_\varrho \mathbf{M}_\sigma)$$
$$= M_\varrho M_\sigma + \frac{1}{m}\{[M_{\varrho+\sigma} - \sigma M_{\sigma-1} M_{\varrho+1} - \varrho M_{\varrho-1} M_{\sigma+1} - (\varrho+\sigma+1) M_\varrho M_\sigma]\}$$
$$+ M_2\left[\varrho\sigma M_{\varrho-1} M_{\sigma-1} + \binom{\varrho}{2} M_{\varrho-2} M_\sigma + \binom{\sigma}{2} M_{\sigma-2} M_\varrho\right].$$

Andererseits ist:

$$\mathfrak{E}[(\mathbf{M}_\varrho - M_\varrho)(\mathbf{M}_\sigma - M_\sigma)] = \mathfrak{E}(\mathbf{M}_\varrho \mathbf{M}_\sigma) - M_\varrho \mathfrak{E}(\mathbf{M}_\sigma) - M_\sigma \mathfrak{E}(\mathbf{M}_\varrho) + M_\varrho M_\sigma,$$

und bei Verwendung von (66) und (64^{IV}) stimmen die Ergebnisse (66) und (64‴) überein.

Wir haben hier in (64) und (65) außerordentlich allgemeine Formeln, die in § 4 zur Anwendung kommen werden. Auf die Verallgemeinerungen für den Fall verschiedener $v_\mu(x)$ gehen wir hier nicht ein.

12. *Näherungslösung der Umkehraufgabe.*

In Abschn. 2, 3, 4 haben wir die näherungsweise Berechnung von Erwartungswerten durchgeführt. Die für unsere Eliminationsmethode aber eigentlich charakteristische Aufgabe ist die umgekehrte, der *Bestimmung der Gestalt einer Funktion, die hinsichtlich einer gegebenen Verteilung einen gegebenen Erwartungswert hat.*

Vorgelegt sei eine eindimensionale Verteilung $v(x)$, aus der wir die m-dimensionale Verteilung (14′) (vgl. § 2, 6) ableiten, ferner beliebige nach (5′) bezüglich $v(x)$ gebildete theoretische Momente (Nullmomente oder auch Schwerpunktmomente) a, M_2^0, $M_3^0, \ldots$, M_2, $M_3, \ldots$ und empirische Momente α, $\mathbf{M}_2^0$, $\mathbf{M}_3^0, \ldots$, $\mathbf{M}_2$, $\mathbf{M}_3, \ldots$, die aus m Beobachtungen gemäß (7′) usf. gebildet sind. Gegeben sei nun eine Funktion der theoretischen Momente (Nullmomente oder Schwerpunktmomente)

$$f(a, M_2, M_3, \ldots). \tag{67}$$

Wir *suchen eine Funktion der empirischen Momente*

$$F(\alpha, \mathbf{M}_2, \mathbf{M}_3, \ldots), \tag{67'}$$

deren Erwartungswert bezüglich (14′) *sich von* (67) *nur um Größen höherer als erster Ordnung in* $\frac{1}{m}$ *unterscheidet.* Die *genaue* Lösung der Umkehr-Aufgabe für einen wichtigen Spezialfall wurde in § 2, 6 behandelt.

Es gilt nach (61) oder (65) mit einem Fehler der Ordnung $\frac{1}{m^2}$:

$$\mathfrak{E}[f(\alpha, \mathbf{M}_2, \mathbf{M}_3, \ldots)] = f(a, M_2, M_3, \ldots) + \frac{1}{m} G(a, M_2, \ldots), \tag{68}$$

daraus näherungsweise

$$f(a, M_2, M_3, \ldots) = \mathfrak{E}[f(\alpha, \mathbf{M}_2, \mathbf{M}_3, \ldots)] - \frac{1}{m} G(a, M_2, \ldots). \tag{68'}$$

Ersetzt man rechts den Ausdruck $G(a, M_2, \ldots)$ durch $\mathfrak{E}[G(\alpha, \mathbf{M}_2, \ldots)]$ so begeht man einen Fehler der Ordnung $\frac{1}{m}$, und da in (68′) $G(a, M_2, \ldots)$ noch mit $\frac{1}{m}$ multipliziert ist, ist der Fehler im Ganzen nur von der Ordnung $\frac{1}{m^2}$. Wir erhalten also die gesuchte Näherungsfunktion F in der Form:

$$F(\alpha, \mathbf{M}_2, \ldots) = f(\alpha, \mathbf{M}_2, \ldots) - \frac{1}{m} G(\alpha, \mathbf{M}_2, \ldots), \tag{69}$$

wobei G durch (68) bestimmt ist. Es ist dann näherungsweise

$$\mathfrak{E}[F(\alpha, \mathbf{M}_2, \ldots)] = f(a, M_2, \ldots).$$

Damit sind die Hilfsmittel zur Anwendung der Eliminationsmethode auf schwierige Fälle bereitgestellt.

§ 4. Anwendungen der Eliminationsmethode.

13. *Die Gaußsche Verteilung.*

Nach K. Pearson ist die notwendige und hinreichende Bedingung dafür, daß eine Pearsonsche Kurve insbesondere eine Gaußsche ist, gegeben durch

$$M_3 = 0, \quad M_4 = 3M_2^2. \tag{70}$$

Außerdem hat Steffensen die Frage untersucht, wann eine nicht notwendig Pearsonsche Kurve eine Gaußsche ist, und erhält die Bedingungen:

$$M_{2n+1} = 0, \quad M_{2n} = (2n-1)\, M_2 M_{2n-2}.$$

Wir wollen ein Kriterium dafür aufstellen, daß die Resultate von m Beobachtungen $X_1, X_2, \ldots, X_m$ als Ergebnisse der Ziehung aus m Urnen mit untereinander gleichen *Gaußschen Verteilungen* aufgefaßt werden können. $M_3 = 0$ gilt für *jede* symmetrische Verteilung. Es kommt daher vor allem auf die zweite Gleichung (70^{II}) an. Wir wollen erst (70) „übersetzen".

Man weiß, daß

$$\mathfrak{E}(\mathbf{M}_3) = \frac{(m-1)(m-2)}{m^2} M_3, \quad \text{also } M_3 = \frac{m^2}{(m-1)(m-2)}\, \mathfrak{E}(\mathbf{M}_3).$$

Es muß also zunächst sein:

$$\mathfrak{E}(\mathbf{M}_3) = 0. \tag{71}$$

Die zweite Gleichung behandeln wir nach der Näherungsmethode des Abschn. 12. Es ist nach (64^{IV}) bzw. (66)

$$\begin{aligned} \mathfrak{E}(\mathbf{M}_4) &= M_4 + \frac{1}{m}(6M_2^2 - 4M_4), \\ \mathfrak{E}(\mathbf{M}_2^2) &= M_2^2 + \frac{1}{m}(M_4 - 3M_2^2), \\ \mathfrak{E}(\mathbf{M}_4 - 3\mathbf{M}_2^2) &= (M_4 - 3M_2^2) - \frac{1}{m}(7M_4 - 15M_2^2). \end{aligned} \tag{72}$$

Diese Beziehungen gelten allgemein für *jede* Verteilung. Im Fall des Bestehens von (70) aber vereinfacht sich (72^{III}) zu:

$$\mathfrak{E}(\mathbf{M}_4-3\mathbf{M}_2^2) = -\frac{2}{m} M_4.$$

Die Prüffunktion F des vorigen Abschn. lautet daher

$$F = (\mathbf{M}_4-3\mathbf{M}_2^2) + \frac{2}{m}\mathbf{M}_4.$$

Ihr Erwartungswert muß verschwinden, wenn die Hypothese einer Gaußschen Verteilung stimmt. Das gesuchte Kriterium, das neben (71) tritt, ist also:

$$\mathfrak{E}\left(\frac{m+2}{m}\mathbf{M}_4 - 3\mathbf{M}_2^2\right) = 0. \tag{73}$$

Wir wollen nun die Streuungen rechnen. Nach (64) gilt:

$$\mathfrak{S}\mathrm{tr}(\mathbf{M}_3) = \frac{1}{m}(M_6 - M_3^2 - 6M_2M_4 + 9M_2^3). \tag{71'}$$

Bei der Berechnung der Streuung von (73) kommt es bei der hier vorliegenden Näherungsbehandlung auf das Zusatzglied $\frac{2}{m}\mathbf{M}_4$ nicht an. Es ist nur die Streuung von ($\mathbf{M}_4-3\mathbf{M}_2^2$) zu rechnen: Diese ist nach (64)

$$\begin{aligned} m\mathfrak{S}\mathrm{tr}(\mathbf{M}_4-3\mathbf{M}_2^2) = M_3 &- 6M_2M_6 \\ &- 8M_3M_5 - M_4^2 + 42M_4M_2^2 + 40M_2M_3^2 - 36M_2^4. \end{aligned} \tag{73'}$$

Da man die theoretischen Momente nicht kennt, weil ja die zwei Parameter der Gaußschen Verteilung unbekannt sind, so wird man, wie wiederholt ausgeführt, bei der Abschätzung von (71′) und (73′) die M durch die beobachteten $\mathbf{M}$ ersetzt denken, dies beinhaltet einen Fehler der Ordnung $\frac{1}{m}$, und da die Streuung ohnehin wie $\frac{1}{m}$ klein ist, ist der Fehler von der Ordnung $\frac{1}{m^2}$, also wieder von der erlaubten Größenordnung.

14. *Die Pearsonschen Kurven.*

Die sogenannten Pearsonschen Kurven enthalten 4 Konstante; in der Bezeichnung von W. P. Elderton [6a]) gilt, falls $a=0$, zwischen den Momenten derselben die Beziehung:

$$M_n[1+(n+1)c_2] + M_{n-1}(nc_1-c) + M_{n-2}(n-1)c_0 = 0.$$

[6a]) W. P. Elderton, Frequency curves and correlation [London 1927]. Vgl. dazu auch v. Mises [l. c. [1])], 271, 3. Formel von oben.

Daraus ergeben sich, wenn man nacheinander $n = 1, 2, 3, 4$ setzt und $M_{-1} = 0$ nimmt, die c_i zu

$$c_0 = -\frac{M_2(4M_2M_4 - 3M_3^2)}{10M_2M_4 - 18M_2^3 - 12M_3^2} = -\frac{M_2(4M_2M_4 - 3M_3^2)}{N},$$

$$c = c_1 = -\frac{1}{N}M_3(M_4 + 3M_2^2),$$

$$c_2 = -\frac{1}{N}(2M_2M_4 - 6M_2^3 - 3M_3^2).$$

Geht man mit diesen Werten c_i in die obige Rekursionsformel ein, so findet man eine *Beziehung, der die Momente einer beliebigen Pearsonschen Kurve genügen müssen.* Diese Gleichung findet sich bei Herrn A. Guldberg. [7])

$$\begin{aligned}(74)\quad & M_n[(8-2n)M_2M_4 - (12-6n)M_2^3 - (9-3n)M_3^2],\\ & - M_{n-1}(n-1)[M_3M_4 + 3M_3M_2^2]\\ & \qquad - M_{n-2}(n-1)(4M_2^2M_4 - 3M_2M_3^2) = 0.\end{aligned}$$

Wir betrachten nun den *Spezialfall,* daß die *sogenannte „Schiefe"* $\beta = \frac{M_3^2}{M_2^3}$ *verschwindet.* Die niedrigste, nicht identische aus (74) resultierende Beziehung lautet:

$$\begin{aligned}(74')\quad 0 = {} & M_5(-2M_2M_4 + 18M_2^3 + 6M_3^2)\\ & - 4M_4(M_3M_4 + 3M_3M_2^2) - 4M_3(4M_4M_2^2 - 3M_2M_3^2).\end{aligned}$$

Wenn nun die Schiefe verschwindet, wird aus (74'):

$$(74'')\qquad M_2M_5(9M_2^2 - M_4) = 0.$$

Man kann annehmen $M_2 \neq 0$, denn die Streuung verschwindet nur bei Verteilungen mit *einer* Sprungstelle. Es ist also entweder

$$(74''')\qquad M_5 = 0 \text{ oder } 9M_2^2 - M_4 = 0.$$

M_5 ist mit der hier üblichen Näherung der Erwartungswert von $\left(\mathbf{M}_5\frac{m+5}{m}\right)$, denn es ist

$$\mathfrak{E}\left(\mathbf{M}_5\frac{m+5}{m}\right) = \frac{m+5}{m}\left[M_5 + \frac{1}{m}(10M_3M_2 - 5M_5)\right].$$

Da $M_3 = 0$, steht tatsächlich rechts

$$M_5\left(1 + \frac{5}{m} - \frac{5}{m} - \frac{25}{m^2}\right).$$

[7]) A. Guldberg [Annales de l'Institut H. Poincaré 3 (1933), 229—278].

Wenn der Erwartungswert von $\mathbf{M}_5 \frac{m+5}{m}$ verschwindet, so auch der von $\mathbf{M}_5$ selbst. Die zu $\mathbf{M}_5$ gehörige Streuung ist

$$(75') \qquad \frac{1}{m}\left(M_{10} - M_5^2 - 10 M_4 M_6 + 25 M_2 M_4^2\right).$$

Die Prüffunktion für die zweite der Gleichungen (74''') ist:

$$(75) \qquad \left(\mathbf{M}_4 - 9\mathbf{M}_2^2\right) + \frac{2}{m}\left(5\mathbf{M}_4 - 3\mathbf{M}_2^2\right).$$

Denn der Erwartungswert dieses Ausdrucks ist bis auf Größen höherer Ordnung in $\frac{1}{m}$:

$$\left(M_4 - 9 M_2^2\right) + \frac{1}{m}\left(6 M_2^2 - 4 M_4 - 9 M_4 + 27 M_2^2\right) + \frac{1}{m}\left(10 M_4 - 6 M_2^2\right).$$

Dies aber ist, wenn $M_4 - 9 M_2^2 = 0$, gleich

$$\frac{1}{m}\left(-3 M_4 + 27 M_2^2\right) = 0.$$

Die zugehörige Streuung von $\left(\mathbf{M}_4 - 9\mathbf{M}_2^2\right)$ lautet:

$$(75'') \qquad \frac{1}{m}\Big\{\left(M_8 - M_4^2\right) + 8\left(20 M_3^2 M_2 - M_3 M_5\right) \\ + 36\left(10 M_2^2 M_4 - 9 M_2^4 - M_2 M_6\right)\Big\}.$$

Es muß also entweder $\mathfrak{E}(\mathbf{M}_5) = 0$ sein mit der Streuung (75') oder der Erwartungswert von (75) verschwinden mit der Streuung (75''), wenn das beachtete Material einer Pearsonschen Kurve mit verschwindender Schiefe angehört. Ist die Kurve als schlechthin symmetrisch vorauszusetzen, so verschwindet M_5 sicher, und es kommt nun auf (75), (75'') als Kriterien an.

15. *Die hypergeometrische Verteilung.*

Als letztes Beispiel betrachte ich die sogenannte hypergeometrische Verteilung. Diese gibt die Wahrscheinlichkeit $w(x)$ dafür an, aus einer Urne, die $(h+k)$ Kugeln und zwar k weiße und h schwarze enthält, in n Zügen, bei denen die jeweils gezogene Kugel *nicht* zurückgelegt wird, x weiße Kugeln zu erhalten. Es gilt:

$$(76) \qquad w(x) = \frac{\binom{k}{x}\binom{h}{n-x}}{\binom{h+k}{n}}.$$

Diese klassische Verteilung wurde in neuerer Zeit von A. Guldberg näher studiert, der insbesondere ihre Momente angibt und eine

Differenzengleichung mitteilt, die von den Konstanten der Verteilung frei ist, und der $w(x)$ genügt. Diese Differenzengleichung wird von Guldberg als statistisches Kriterium herangezogen, und dies bildet einen gewissen Berührungspunkt mit der hier dargelegten Methode [8]).

Es gelten bei dieser Verteilung besonders einfache Ausdrücke für die *faktoriellen* Momente. Diese faktoriellen Momente $M^{(\nu)}$ sind bekanntlich definiert durch

$$M^{(\nu)} = \sum_x x(x-1)\dots(x-\nu+1)v(x) = \sum_x x^{(\nu)}v(x).$$

Mit dieser Bezeichnung gilt

$$M^{(\nu)} = \frac{n^{(\nu)}k^{(\nu)}}{(h+k)^{(\nu)}}. \tag{77}$$

Aus den $M^{(\nu)}$ folgen natürlich die Nullmomente M_ν^0 (die übrigens — ebenso wie auch bei der Bernoullischen Verteilung — keine übersichtliche Gestalt haben). Einfache Ausdrücke resultieren erst wieder für die *Schwerpunktmomente*. Nehmen wir $M_1^0 = a$ dazu, so haben wir das System

$$\begin{aligned} a &= n\frac{k}{h+k}, \\ M_2 &= n\frac{h+k}{(h+k)^2}\cdot\frac{h+k-n}{h+k-1}, \\ M_3 &= n\frac{hk(h-k)}{(h+k)^3}\cdot\frac{h+k-n}{h+k-1}\cdot\frac{h+k-2n}{h+k-2}. \end{aligned} \tag{76}$$

Man sieht, daß für $(h+k)\to\infty$, n fest, diese Formeln in die bekannten analogen der Bernoullischen Verteilung übergehen. Hier sind drei unbekannte Parameter.

Wir wollen zunächst n und $(h+k)$, also die Anzahl der Züge und die Anzahl der Kugeln als gegeben voraussetzen und nur das Mischungsverhältnis als unbekannt. Dann folgt aus der ersten und zweiten der Gleichungen (76):

---

[8]) So eine Differenzengleichung stellt natürlich nicht *ein* zusammenfassendes, skalares Kriterium dar, was man — meines Erachtens — doch eigentlich wünschen muß. In dem Gedanken, ein von den Parametern der Verteilung freies Kriterium zu gewinnen, liegt ein Berührungspunkt zwischen dem Guldbergschen und dem in dieser Untersuchung dargelegten Verfahren. Es unterscheidet sich unser Verfahren von dem Guldbergschen durch die bei uns wesentliche systematische „*Übersetzung*" in Gleichungen zwischen Erwartungswerten unter *Beachtung der Streuungen.*

$$M_2 = a\left(1 - \frac{a}{n}\right)\frac{h+k-n}{h+k-1}, \tag{77}$$

rechts ist $a - \frac{a^2}{n}$ mit einem echten Bruch multipliziert. Es herrscht also bei der hypergeometrischen Verteilung *unternormale Dispersion* im Sinne der Lexis'schen Theorie, d.h.

$$\mathfrak{E}\left(\mathbf{M}_2 \frac{m}{m-1}\right) < \frac{mn}{mn-1}\mathfrak{E}\left(\alpha - \frac{\alpha^2}{n}\right). \tag{77'}$$

Will man also ohne Kenntnis der Parameter zwischen einer Bernoullischen und hypergeometrischen Verteilung entscheiden, so kann (20) bzw. (77′) dazu dienen.

Wir wollen nun auch noch $(h+k)$ eliminieren. Es folgt aus der 2. und 3. Gleichung (76)

$$(h+k)\left(M_3 - M_2 + \frac{2a}{n}M_2\right) = 2M_3 - 2nM_2 + 4M_2. \tag{78'}$$

Andererseits folgt $(h+k)$ aus (77) gemäß:

$$(h+k)\left(M_2 - a + \frac{a^2}{n}\right) = M_2 - na + a^2. \tag{78''}$$

Durch Vergleich von (78′) und (78″) folgt schließlich die „charakteristische Gleichung"

$$\frac{M_2 - na + a^2}{nM_2 - na + a^2} = 2\,\frac{M_3 - nM_2 + 2aM_2}{nM_3 - nM_2 + 2aM_2}. \tag{78}$$

Das ist eine Beziehung, die nur noch den Parameter n enthält; n auch zu eliminieren, wäre nicht sachgemäß.

Da die Ausdrücke für die faktoriellen Momente so einfach waren, könnte man daran denken, da vielleicht eine noch bequemere Beziehung zu erhalten. Wir schreiben für $M^{(\nu)}$ der Kürze halber m_ν. Dann gilt

$$a = m_1 = n\frac{k}{h+k},$$

$$m_2 = \frac{n(n-1)k\cdot(k-1)}{(h+k)(h+k-1)},$$

$$m_3 = \frac{n(n-1)(n-2)k(k-1)(k-2)}{(h+k)(h+k-1)(h+k-2)};$$

daraus folgt, mit $h + k = x$,

$$m_3 = m_2(n-2)\frac{ax-2n}{nx-2n},$$

$$m_2 = a\,(n-1)\frac{ax-n}{nx-n}$$

und schließlich

$$(79)\qquad \frac{m_2 - a(n-1)}{nm_2 - a^2(n-1)} = 2\,\frac{m_3 - m_2(n-2)}{nm_3 - am_2(n-2)}.$$

(79) ist wohl etwas einfacher als (78). Wir könnten (79) mittels der Formeln von § 3, 10, für die Nullmomente bearbeiten, da $m_2 = M_2^0 - a$, $m_3 = M_3^0 - 3M_2^0 + 2a$. Wir wollen aber hier lieber die weitere Durchführung an (78) anknüpfen. Die *beiden* Nenner in (78) verschwinden, im Fall, daß die Verteilung eine Bernoullische ist. Für die hypergeometrische sind sie von Null verschieden. Durch Ausmultiplizieren folgt:

$$(78')\qquad nM_2M_3 + (2-n)(a^2M_3 - naM_3) + (2n-1)(2aM_2^2 - nM_2^2) + (an^2M_2 - 3a^2nM_2 + 2a^3M_2) = 0.$$

Berechnung der erforderlichen Erwartungswerte nach (65) ergibt:

$$\mathfrak{E}(\mathbf{M}_2\mathbf{M}_3) = M_2M_3 + \frac{1}{m}(M_5 - 8M_2M_3),$$

$$\mathfrak{E}(\alpha\mathbf{M}_3) = aM_3 + \frac{1}{m}(M_4 - 3aM_3 - 3M_2^2),$$

$$\mathfrak{E}(\alpha^2\mathbf{M}_3) = a^2M_3 + \frac{1}{m}(M_2M_3 - 3a^2M_3 + 2a(M_4 - 3M_2^2)),$$

$$\mathfrak{E}(\mathbf{M}_2^2) = M_2^2 + \frac{1}{m}(M_4 - 5M_2^2),$$

$$\mathfrak{E}(\alpha\mathbf{M}_2^2) = aM_2^2 + \frac{1}{m}(2M_2M_3 + aM_4 - 3aM_2^2),$$

$$\mathfrak{E}(\alpha\mathbf{M}_2) = aM_2 + \frac{1}{m}(M_3 - aM_2),$$

$$\mathfrak{E}(\alpha^2\mathbf{M}_2) = a^2M_2 + \frac{1}{m}(M_2^2 - a^2M_2 + 2aM_3),$$

$$\mathfrak{E}(\alpha^3\mathbf{M}_2) = a^3M_2 + \frac{1}{m}(3aM_2^2 - a^3M_2 + 3a^2M_3).$$

Man erhält somit, wenn man die linke Seite von (78′) kurz mit

$f(a, M_2, M_3)$ bezeichnet und zur Vereinfachung (78′) addiert:

$$m\,\mathfrak{E}\,[f(\alpha, \mathbf{M}_2, \mathbf{M}_3)] = (-2M_2M_3 + nM_5 + n^2M_3) \\ + (n+1)(2a^2M_3 - 2naM_3 - 2aM_2^2 + nM_2^2 + 2aM_4 - nM_4).$$

Es muß also nach unserer Theorie *der Erwartungswert verschwinden von*:

$$(79)\quad n\mathbf{M}_2\,\mathbf{M}_3 + \alpha\mathbf{M}_3\,(2-n)\,(\alpha-n) \\ + \mathbf{M}_2^2\,(1-2n)\,(n-2\alpha) + \alpha\mathbf{M}_2\,(2\alpha^2 - 3n\alpha + n^2) \\ - \frac{1}{m}\left[2\alpha\mathbf{M}_3\,(n+1)\,(\alpha-n) \right. \\ \left. + (\mathbf{M}_2^2 - \mathbf{M}_4)(n+1)(n-2\alpha) - 2\mathbf{M}_2\mathbf{M}_3 + n\mathbf{M}_5 + n^2\mathbf{M}_3\right].$$

Auch die Streuung können wir nach unsern Formeln angeben. Es ist mit den Bezeichnungen von § 3, 11

$$(80')\quad \begin{aligned} f_1 &= (2-n)(2a-n)M_3 + (4n-2)M_2^2 + n^2M_2 + 6aM_2(a-n) \\ f_2 &= nM_3 + (2n-1)(4aM_2 - 2nM_2) + (an^2 - 3a^2n + 2a^3) \\ f_3 &= nM_2 + (2-n)(a^2 - an), \end{aligned}$$

und es folgt aus (64) *für die Streuung von* (79), die wir $\mathfrak{Str}(F)$ nennen:

$$(80)\quad m\,\mathfrak{Str}(F) = f_1^2M_2 + 2f_1[f_2M_3 + f_3(M_4 - 3M_2^2)] \\ + f_2^2(M_4 - M_2^2) + 2f_2f_3(M_5 - 4M_2M_3) \\ + f_3^2(M_6 - M_3^2 - 6M_2M_4 + 9M_2^3).$$

In (79) und (80) liegt die Übertragung der einfachen Lexis'schen Kriterien (20), (23) auf die hypergeometrische Verteilung. In einer folgenden Arbeit [9]) soll eine Übertragung der Lexis'schen Theorie auf mehrdimensionale Probleme — mit der bloßen Analogie zu den eindimensionalen Resultaten kommt man dabei nicht aus — unter Benutzung der hier dargelegten allgemeinen Betrachtungsweise durchgeführt werden.

(Eingegangen den 24. Februar 1934.)

---

[9]) Zusatz während der Korrektur:
Die Arbeit ist indessen erschienen: Une nouvelle methode de statistique théorique. (Problémes á deux dimensions.) [Bulletin de l'académie royale de Belgique **21** (1935).]