

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

Herausgeber: Vereinigung Schweizerischer Versicherungsmathematiker

Band: 34 (1937)

Artikel: Analytische Ausgleichung durch Polynome mit einer Anwendung auf die
schweizerischen Volkssterbetafeln

Autor: Ruchti, Werner

DOI: <https://doi.org/10.5169/seals-555051>

Nutzungsbedingungen

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

Conditions d'utilisation

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

Terms of use

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

Download PDF: 16.01.2026

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

Analytische Ausgleichung durch Polynome

mit einer Anwendung auf die schweizerischen Volkssterbetafeln

Von Dr. Werner Ruchtli

I.

Theoretischer Teil.

1. Einleitung.

Es ist die ursprünglichste Aufgabe der Statistik, einfache, messbare Vorgänge durch eine Reihe von Beobachtungen in ihrem Verlaufe festzuhalten. Denken wir uns eine Anzahl solcher Beobachtungspunkte gegeben. Um einen ersten Überblick über die Verteilung zu erhalten, können wir die aufeinanderfolgenden Punkte durch gerade Strecken verbinden, wodurch wir einen gebrochenen Linienzug herstellen. Mehr als zu einer ersten, primitiven Veranschaulichung kann uns aber dieser nicht dienen, denn trotz des anscheinenden Zusammenhangs bleibt es noch bei der diskontinuierlichen Punktverteilung. Die Aufgabe des mathematischen Statistikers ist es, diesen Polygonzug durch eine stetige Funktion zu ersetzen, so dass der Funktionswert nicht nur an den beobachteten Stellen angegeben werden kann, sondern an jeder beliebigen Stelle zwischen den äussersten Grenzen. Diese Aufgabe heisst Interpolation.

In einer solchen beobachteten Verteilung sind aber neben den wesensmässigen, grossen Schwankungen auch immer kleinere, zufällige enthalten, und zwar um so mehr, je kleiner die Gesamtzahl der Beobachtungen ist. Denken wir uns dazu noch eine grosse Anzahl von Beobachtungspunkten, so führt das zu einer sehr komplizierten Polygonlinie. Wenn der Mathematiker auch hier noch diese diskontinuierliche Verteilung durch eine stetige ersetzen will, so kann es sich nicht mehr darum handeln, eine Funktion zu finden, die für alle die gegebenen Punkte erfüllt ist. Es handelt sich vielmehr darum, eine Kurve zu finden, die den Gesamtverlauf möglichst naturgemäss wiedergibt, ohne den zufallsbedingten Schwankungen zweiter und dritter Ordnung zu folgen.

Gegeben sei eine Reihe von n Beobachtungswerten. Diese n Punkte tragen wir möglichst genau auf Millimeterpapier ab und versuchen, aus dem Verlauf den Grad für die Ausgleichungsparabel zu ermitteln. Es sei der m . Grad. Folglich stellen wir das Polynom m . Grades mit den $m + 1$ unbekannten Koeffizienten auf. Es wäre nun möglich, n Gleichungen zur Bestimmung dieser $m + 1$ Unbekannten aufzustellen. Statt dass wir aber $m + 1$ beliebige von den n auswählen, zeichnen wir die gewünschte Kurve möglichst genau durch die Beobachtungspunkte ein und greifen dann $m + 1$ für die betreffende Kurve typische Punkte zur Bestimmung der Koeffizienten heraus.

Für komplizierte Verteilungen ist natürlich diese empirische Methode zu ungenau. Aber wir finden durch sie den Anschluss an eine erste Verbesserung. Sicher möchten wir die provisorische Kurve so durch die n Punkte einzeichnen, dass die Summe der absoluten oder quadratischen Fehler minimal wird. Eine Methode, bei der die Koeffizienten nach diesem Gesichtspunkt bestimmt werden, existiert. Sie heisst die «Methode der kleinsten Quadrate».

2. Die Methode der kleinsten Quadrate.

1. Ausgehend von den n beobachteten Werten y_1, y_2, \dots, y_n an den Stellen x_1, x_2, \dots, x_n stellen wir folgendes Polynom auf:

$$f(x) = a_0 P_0(x) + a_1 P_1(x) + a_2 P_2(x) + \dots + a_m P_m(x)$$

Dabei sind die $P_i(x)$ bekannte, passend gewählte Funktionen von x . Z. B. kann man wählen $P_i(x) = x^i$. Die a_i sind wieder zu bestimmende Parameter. Durch Einsetzen der n Punkte werden wir auf folgendes System geführt:

$$\begin{aligned} y_1 &= a_0 P_0(x_1) + a_1 P_1(x_1) + a_2 P_2(x_1) + \dots + a_m P_m(x_1) \\ &\dots\dots\dots \\ y_n &= a_0 P_0(x_n) + a_1 P_1(x_n) + a_2 P_2(x_n) + \dots + a_m P_m(x_n) \end{aligned}$$

Es muss natürlich gelten $n \geq m + 1$. Interessant ist nur der Fall, wo $n > m + 1$, sonst hat man einfach n inhomogene Gleichungen zur Bestimmung von n Unbekannten vor sich.

2. Wir begnügen uns nun aber nicht damit, diese $m + 1$ Unbekannten aus irgendwelchen $m + 1$ von den n Gleichungen zu bestimmen, sondern wir wollen dazu alle n Gleichungen verwenden.

Zur Berechnung der a_i stellen wir folgende Forderung:

$$\sum_1^n (a_0 P_0(x_k) + a_1 P_1(x_k) + \dots + a_m P_m(x_k) - y_k)^2 \equiv \sum_1^n (f(x_k) - y_k)^2$$

soll ein Minimum sein. Da diese Summe als Funktion der Parameter a_i aufzufassen ist, müssen die partiellen Ableitungen nach den a_i verschwinden.

$$\sum_1^n P_0(x_k) (f(x_k) - y_k) = 0$$

$$\sum_1^n P_1(x_k) (f(x_k) - y_k) = 0$$

.....

$$\sum_1^n P_m(x_k) (f(x_k) - y_k) = 0$$

Setzt man nun $\sum_1^n P_i(x_k) \cdot P_j(x_k) = \sigma_{ij} = \sigma_{ji}$

und $\sum_1^n P_i(x_k) \cdot y_k = \varrho_i$

so erhält man zur Bestimmung der a_i folgendes System

$$\sigma_{00} a_0 + \sigma_{01} a_1 + \dots + \sigma_{0m} a_m = \varrho_0$$

$$\sigma_{10} a_0 + \sigma_{11} a_1 + \dots + \sigma_{1m} a_m = \varrho_1$$

.....

$$\sigma_{m0} a_0 + \sigma_{m1} a_1 + \dots + \sigma_{mm} a_m = \varrho_m$$

Das sind $m + 1$ inhomogene Gleichungen mit $m + 1$ Unbekannten. Die einfache Matrix ist symmetrisch. In diesen $m + 1$ Gleichungen sind aber nun alle n Beobachtungen verwertet. Durch Auflösen des Systems findet man diejenigen Koeffizienten a , die vereint mit den einmal gewählten Funktionen $P_i(x)$ bei der Ausgleichung die kleinsten quadratischen Fehler ergeben.

3. Das sind nun nicht die kleinsten Fehler, die überhaupt erreicht werden können, denn die Güte des Resultats ist neben den a_i auch abhängig von den $P_i(x)$. Diese haben wir jedoch nur durch Probieren ermitteln können. Es wäre ein Zufall, wenn man dabei gerade auf die allerbesten gestossen wäre. So ist die Genauigkeit der Ausgleichung schon beschränkt, ehe man nur zur Bestimmung der Koeffizienten nach der Methode der kleinsten Quadrate schreitet. Wir brauchen deshalb noch ein Kriterium für diese Funktionen $P_i(x)$. Diese Frage findet durch die Theorie der orthogonalen Polynome, die in einem der folgenden Kapitel berührt werden soll, eine praktische Lösung.

Zuerst wollen wir uns noch einer andern, mit der der kleinsten Quadrate verwandten Methode zuwenden, die man dem Mathematiker *Cauchy* zu verdanken hat.

3. Die Methode von Cauchy ¹⁾.

Zu den Abszissenstellen x_1, x_2, \dots, x_n seien die Beobachtungswerte y_1, y_2, \dots, y_n gegeben. Wir setzen wieder:

$$f(x) = a_1 P_1(x) + a_2 P_2(x) + a_3 P_3(x) + \dots \text{ ad. lib.}$$

Die Methode von Cauchy geht etappenweise vor: Die Resultate der n . Näherung dienen zur Berechnung der Konstanten der Funktion für die $(n+1)$. Näherung.

Erste Näherung.

Man nimmt vorläufig nur $f(x) = a_1 P_1(x)$ oder kurz $f(x) = a \cdot P(x) \dots (1)$.

Damit hat man zur Berechnung der einen Unbekannten a folgendes System von n Gleichungen:

$$\left. \begin{array}{l} f(x_1) = a \cdot P(x_1) \\ f(x_2) = a \cdot P(x_2) \\ \dots\dots\dots \\ f(x_n) = a \cdot P(x_n) \end{array} \right\} \quad \begin{array}{l} y_1 \\ y_2 \\ \dots\dots\dots \\ y_n \end{array}$$

Nun trägt man in einem Koordinatensystem $\{P(x), 0, f(x)\}$ folgende n Punkte ab:

$$M_i (P(x_i), f(x_i)), \quad i = 1, 2, \dots, n$$

y_i

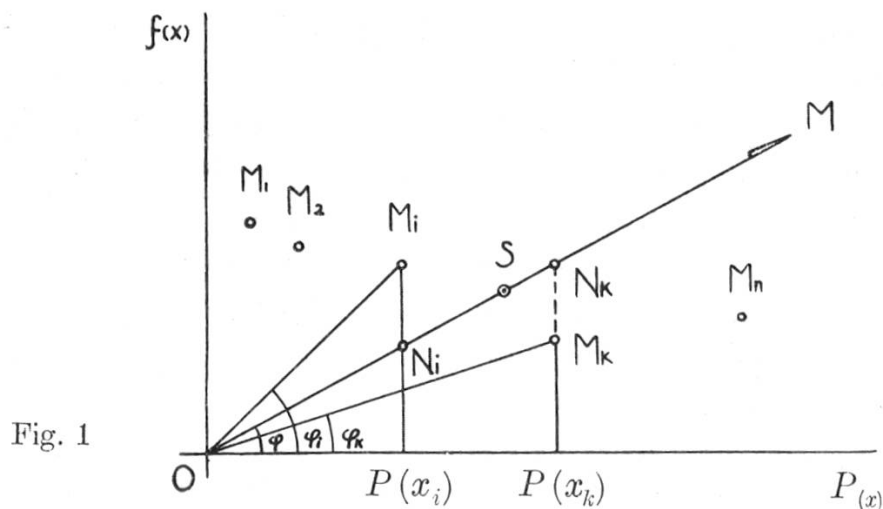


Fig. 1

$$f(x_i) = P(x_i) \cdot \operatorname{tg} \varphi_i$$

$$\operatorname{tg} \varphi_i = a_{(i)} \text{ gesetzt.}$$

Wir haben damit n Gerade durch den Nullpunkt mit den Richtungskoeffizienten $a_{(1)}, a_{(2)}, \dots, a_{(n)}$ festgelegt.

Die Methode von Cauchy besteht darin, die Resultierende der n Vektoren $\overrightarrow{OM_i}$ zu berechnen und den Richtungskoeffizienten des resultierenden Vektors, dessen Länge für uns keine Rolle spielt, als das gesuchte a anzunehmen. Das bedeutet, dass man den Schwerpunkt S der n Punkte M_i bestimmt und auf der Geraden OS den Vektor \overrightarrow{OS} n mal abträgt, was zu dem Punkt M führt.

Es sind also in gewissem Sinne die M_i die beobachteten und die N_i , die auf \overline{OM} liegen, die berechneten Werte. Die Abweichungen sind von der Form

$$\overline{N_i M_i} = f(x_i) - P(x_i) \cdot \operatorname{tg} \varphi$$

Infolge der charakteristischen Schwerpunktseigenschaft ist aber die Summe dieser Fehler, wenn man das Vorzeichen berücksichtigt, gleich null.

$$\sum_{i=1}^n \{f(x_i) - P(x_i) \cdot \operatorname{tg} \varphi\} = 0$$

Folglich ergibt sich für die Resultierende:

$$(2) \quad \operatorname{tg} \varphi = a = \frac{\sum_{i=1}^n i f(x_i)}{\sum_{i=1}^n i P(x_i)}$$

Bemerkung: Dieses Vorgehen verlangt eine Ergänzung. Wären nämlich die Punkte M_i gerade so beidseitig des Ursprungs verteilt, dass S und somit auch M die Abszisse 0 hätte, dann wäre:

$$a = \operatorname{tg} (90^\circ) = \infty.$$

Somit würde jeder Fehler ∞ . Je mehr sich φ einem rechten Winkel nähert, desto schlechter wird diese Methode.

Um zu verhüten, dass in (2) der Nenner null wird, addiert man nicht die n Gleichheiten $f(x_i) = a \cdot P(x_i)$ direkt, um dann nach a aufzulösen, sondern man multipliziert zuerst jede mit ± 1 , je nachdem der entsprechende Wert von $P(x_i)$ positiv oder negativ ist. Die so erhaltenen Summen wollen wir mit Σ' bezeichnen. Diese Abänderung wirkt sich also nur dann aus, wenn der Punkt M_i eine negative Abszisse hat.

Begründung dieser neueingeführten Summen Σ' :

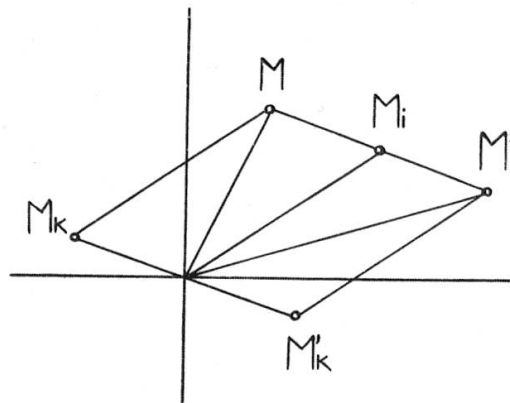


Fig. 2.

Nach der Figur ist $a_{(i)} > 0$, während $a_{(k)} < 0$, wo $a_{(j)} = \operatorname{tg} \varphi_j$. Also muss das resultierende a kleiner sein als $a_{(i)}$. Statt M_k hat man also M'_k zu berücksichtigen, für das $a_k = \operatorname{tg} \varphi_k$ gleich gross ist wie für M_k . Damit gelangt man schliesslich nach M' statt nach M .

Diese Überlegung kann man leicht auf den dritten Quadranten ausdehnen und sieht, dass auch da eine Rückführung, und zwar in den ersten Quadranten, nötig ist.

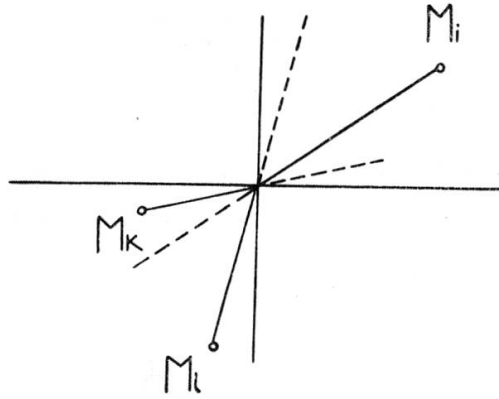


Fig. 3.

$a_{(k)}$ verlangt eine Verkleinerung des $a_{(i)}$, $a_{(l)}$ dagegen eine Vergrößerung.

Für den Parameter a erhält man das endgültige Ergebnis:

$$(2') \quad a = \frac{\sum' f(x_i)}{\sum' P(x_i)}$$

Damit nehmen die Fehler zwischen Beobachtung und Berechnung die Form an:

$$\Delta f(x_i) = f(x_i) - \frac{\sum' f(x_i)}{\sum' P(x_i)} \cdot P(x_i)$$

wobei
$$\sum \Delta f(x_i) = \sum f(x_i) - \frac{\sum' f(x_i)}{\sum' P(x_i)} \cdot \sum P(x_i) = 0$$

Haben diese Abweichungen $\Delta f(x_i)$ nicht den Charakter von Zufallsfehlern, so nimmt man eine

Zweite Näherung.

$$(3) \quad f(x) = a_1 P_1(x) + a_2 P_2(x)$$

Man geht wieder gleich vor. Je nachdem $P_1(x)$ positiv oder negativ ist, multipliziert man mit $+1$ oder -1 und addiert. So findet man:

$$\begin{aligned} \Sigma' f(x_i) &= a_1 \Sigma' P_1(x_i) + a_2 \Sigma' P_2(x_i) \\ (4) \quad a_1 &= \frac{\Sigma' f(x_i)}{\Sigma' P_1(x_i)} - a_2 \frac{\Sigma' P_2(x_i)}{\Sigma' P_1(x_i)} \quad \text{in (3) eingesetzt:} \end{aligned}$$

$$\begin{aligned} f(x_i) &= \left(\frac{\Sigma' f(x_i)}{\Sigma' P_1(x_i)} - a_2 \frac{\Sigma' P_2(x_i)}{\Sigma' P_1(x_i)} \right) \cdot P_1(x_i) + a_2 P_2(x_i) \\ f(x_i) - \frac{\Sigma' f(x_i)}{\Sigma' P_1(x_i)} \cdot P_1(x_i) &= a_2 \left(P_2(x_i) - \frac{\Sigma' P_2(x_i)}{\Sigma' P_1(x_i)} \cdot P_1(x_i) \right) \end{aligned}$$

oder $\Delta f(x_i) = a_2 \cdot \Delta P_2(x_i)$

Darauf üben wir wieder jene Manipulation mit ± 1 aus, je nachdem $\Delta P_2(x_i)$ positiv oder negativ ist. So wie wir die Summen, die vom Vorzeichen von $P_1(x_i)$ abhängig sind, mit Σ' bezeichnet haben, so wollen wir jetzt die Summen, die dem Vorzeichen von $\Delta P_2(x_i)$ Rechnung tragen, Σ'' nennen.

$$\Sigma'' \Delta f(x_i) = a_2 \Sigma'' \Delta P_2(x_i)$$

woraus

$$(4') \quad a_2 = \frac{\Sigma'' \Delta f(x_i)}{\Sigma'' \Delta P_2(x_i)}$$

und von früher

$$(4) \quad a_1 = \frac{\Sigma' f(x_i)}{\Sigma' P_1(x_i)} - a_2 \frac{\Sigma' P_2(x_i)}{\Sigma' P_1(x_i)}$$

Die Fehler ergeben sich nun in der Form:

$$\Delta^2 f(x_i) = \Delta f(x_i) - \frac{\Sigma'' \Delta f(x_i)}{\Sigma'' \Delta P_2(x_i)} \cdot \Delta P_2(x_i)$$

Wegen der Schwerpunktseigenschaft gilt auch hier:

$$\Sigma \Delta^2 f(x_i) = 0$$

Zeigen die Fehler $\Delta^2 f(x_i)$ noch immer systematischen Charakter, so versucht man eine

Dritte Näherung.

$$(5) \quad f(x) = a_1 P_1(x) + a_2 P_2(x) + a_3 P_3(x)$$

Je nachdem $P_1(x)$ positiv oder negativ ist, multipliziert man mit ± 1 und addiert.

$$\begin{aligned} \Sigma' f(x_i) &= a_1 \Sigma' P_1(x_i) + a_2 \Sigma' P_2(x_i) + a_3 \Sigma' P_3(x_i) \\ (6) \quad a_1 &= \frac{\Sigma' f(x_i)}{\Sigma' P_1(x_i)} - a_2 \frac{\Sigma' P_2(x_i)}{\Sigma' P_1(x_i)} - a_3 \frac{\Sigma' P_3(x_i)}{\Sigma' P_1(x_i)} \\ f(x_i) &= \left(\frac{\Sigma' f(x_i)}{\Sigma' P_1(x_i)} - a_2 \frac{\Sigma' P_2(x_i)}{\Sigma' P_1(x_i)} - a_3 \frac{\Sigma' P_3(x_i)}{\Sigma' P_1(x_i)} \right) P_1(x_i) + a_2 P_2(x_i) + \\ &\quad + a_3 P_3(x_i) \end{aligned}$$

$$\begin{aligned} f(x_i) - \frac{\Sigma' f(x_i)}{\Sigma' P_1(x_i)} \cdot P_1(x_i) &= a_2 \left(P_2(x_i) - \frac{\Sigma' P_2(x_i)}{\Sigma' P_1(x_i)} \cdot P_1(x_i) \right) + \\ &\quad + a_3 \left(P_3(x_i) - \frac{\Sigma' P_3(x_i)}{\Sigma' P_1(x_i)} \cdot P_1(x_i) \right) \\ \Delta f(x_i) &= a_2 \Delta P_2(x_i) + a_3 \Delta P_3(x_i) \end{aligned}$$

Man macht wieder die bekannte Multiplikation, jetzt in bezug auf $\Delta P_2(x_i)$:

$$\begin{aligned} \Sigma'' \Delta f(x_i) &= a_2 \Sigma'' \Delta P_2(x_i) + a_3 \Sigma'' \Delta P_3(x_i) \\ (6') \quad a_2 &= \frac{\Sigma'' \Delta f(x_i)}{\Sigma'' \Delta P_2(x_i)} - a_3 \frac{\Sigma'' \Delta P_3(x_i)}{\Sigma'' \Delta P_2(x_i)} \end{aligned}$$

in $\Delta f(x_i)$ eingesetzt:

$$\begin{aligned} \Delta f(x_i) &= \left(\frac{\Sigma'' \Delta f(x_i)}{\Sigma'' \Delta P_2(x_i)} - a_3 \frac{\Sigma'' \Delta P_3(x_i)}{\Sigma'' \Delta P_2(x_i)} \right) \Delta P_2(x_i) + a_3 \Delta P_3(x_i) \\ \Delta f(x_i) - \frac{\Sigma'' \Delta f(x_i)}{\Sigma'' \Delta P_2(x_i)} \Delta P_2(x_i) &= a_3 \left(\Delta P_3(x_i) - \frac{\Sigma'' \Delta P_3(x_i)}{\Sigma'' \Delta P_2(x_i)} \cdot \Delta P_2(x_i) \right) \\ \Delta^2 f(x_i) &= a_3 \Delta^2 P_3(x_i) \end{aligned}$$

Dieselbe Erweiterung mit ± 1 führt man aus für $\Delta^2 P_3(x_i)$ und bezeichnet die entsprechenden Summen mit Σ'''

$$\Sigma''' \Delta^2 f(x_i) = a_3 \Sigma''' \Delta^2 P_3(x_i)$$

$$(6'') \quad a_3 = \frac{\Sigma''' \Delta^2 f(x_i)}{\Sigma''' \Delta^2 P_3(x_i)}$$

Die Gleichungen (6''), (6') und (6) ergeben sukzessive die gesuchten Parameter.

Abweichungen:

$$\Delta^3 f(x_i) = \Delta^2 f(x_i) - \frac{\Sigma''' \Delta^2 f(x_i)}{\Sigma''' \Delta^2 P_3(x_i)} \Delta^2 P_3(x_i)$$

wobei: $\Sigma \Delta^3 f(x_i) = 0.$

Würde eine dritte Näherung noch nicht genügen, so könnte man analog weiterfahren.

Im Falle von n beobachteten Werten und m Gliedern oder Koeffizienten kann der zu $f(x)$ gehörende mittlere quadratische Fehler ε wie folgt definiert werden:

$$\varepsilon = \pm \sqrt{\frac{\Sigma (\Delta^m f(x))^2}{n - m}} \quad 1)$$

4. Rückführung der Methode von Cauchy auf die Methode der kleinsten Quadrate nach Carvallo.

Die von Cauchy verwendeten Gewichte ± 1 können durch andere ersetzt werden. Multipliziert man z. B. jede der n Gleichungen

$$f(x_i) = a_1 P_1(x_i) + a_2 P_2(x_i) + a_3 P_3(x_i)$$

mit dem zugehörigen $P_1(x_i)$, so bedeutet das hinsichtlich der Vorzeichen dasselbe wie die Erweiterung mit ± 1 . Die neuen Summen, die bei der Addition entstehen, wollen wir mit $'\Sigma$ bezeichnen.

Das Analoge kann durchgeführt werden für $P_2(x_i)$ und $P_3(x_i)$. Es seien $''\Sigma$ und $'''\Sigma$ die entsprechenden Summen.

$$\Sigma f(x_i) P_1(x_i) = a_1 \Sigma P_1^2(x_i) + a_2 \Sigma P_2(x_i) \cdot P_1(x_i) + a_3 \Sigma P_3(x_i) \cdot P_1(x_i):$$

$$(1) \quad '\Sigma f(x_i) = a_1 '\Sigma P_1(x_i) + a_2 '\Sigma P_2(x_i) + a_3 '\Sigma P_3(x_i)$$

und analog

$$(2) \quad ''\Sigma f(x_i) = a_1 ''\Sigma P_1(x_i) + a_2 ''\Sigma P_2(x_i) + a_3 ''\Sigma P_3(x_i)$$

$$(3) \quad '''\Sigma f(x_i) = a_1 '''\Sigma P_1(x_i) + a_2 '''\Sigma P_2(x_i) + a_3 '''\Sigma P_3(x_i)$$

Das sind drei Gleichungen für die drei Unbekannten a . Auf dasselbe System stösst man aber auch, wenn man die partielle Ableitungen von

$$\Sigma \{f(x_i) - a_1 P_1(x_i) - a_2 P_2(x_i) - a_3 P_3(x_i)\}^2$$

nach den a gleich null setzt.

Aus (1) folgt

$$a_1 = \frac{'\Sigma f(x_i)}{'\Sigma P_1(x_i)} - a_2 \frac{'\Sigma P_2(x_i)}{'\Sigma P_1(x_i)} - a_3 \frac{'\Sigma P_3(x_i)}{'\Sigma P_1(x_i)} \dots (\alpha)$$

Eingesetzt in (2) und (3)

$$(2') \quad ''\Sigma f(x_i) = \left(\frac{'\Sigma f(x_i)}{'\Sigma P_1(x_i)} - a_2 \frac{'\Sigma P_2(x_i)}{'\Sigma P_1(x_i)} - a_3 \frac{'\Sigma P_3(x_i)}{'\Sigma P_1(x_i)} \right) \cdot ''\Sigma P_1(x_i) + \\ + a_2 ''\Sigma P_2(x_i) + a_3 ''\Sigma P_3(x_i)$$

$$(3') \quad '''\Sigma f(x_i) = (\dots) '''\Sigma P_1(x_i) + a_2 '''\Sigma P_2(x_i) + a_3 '''\Sigma P_3(x_i)$$

Aus (2') folgt sodann:

$$(2'') \quad ''\Sigma f(x_i) - \frac{'\Sigma f(x_i)}{'\Sigma P_1(x_i)} ''\Sigma P_1(x_i) = \\ = a_2 \left(''\Sigma P_2(x_i) - \frac{'\Sigma P_2(x_i)}{'\Sigma P_1(x_i)} ''\Sigma P_1(x_i) \right) + \\ + a_3 \left(''\Sigma P_3(x_i) - \frac{'\Sigma P_3(x_i)}{'\Sigma P_1(x_i)} ''\Sigma P_1(x_i) \right)$$

Nun ist:

$$''\Sigma f(x_i) - \frac{'\Sigma f(x_i)}{'\Sigma P_1(x_i)} ''\Sigma P_1(x_i) = \Sigma f(x_i) \cdot P_2(x_i) - \\ - \frac{'\Sigma f(x_i)}{'\Sigma P_1(x_i)} \Sigma P_1(x_i) \cdot P_2(x_i) \\ = \Sigma P_2(x_i) \left\{ f(x_i) - \frac{'\Sigma f(x_i)}{'\Sigma P_1(x_i)} \cdot P_1(x_i) \right\} \\ = \Sigma P_2(x_i) \cdot \Delta f(x_i) \\ = ''\Sigma \Delta f(x_i)$$

Analoge Bezeichnungen führen wir auch für die Ausdrücke auf der rechten Seite von (2'') ein.

Dann geht (2'') über in

$$\text{I.} \quad {}''\Sigma \Delta f(x_i) = a_2 {}''\Sigma \Delta P_2(x_i) + a_3 {}''\Sigma \Delta P_3(x_i)$$

Analog verfährt man mit (3') und findet

$$\text{II.} \quad {}'''\Sigma \Delta f(x_i) = a_2 {}'''\Sigma \Delta P_2(x_i) + a_3 {}'''\Sigma \Delta P_3(x_i)$$

Aus I. folgt

$$a_2 = \frac{{}''\Sigma \Delta f(x_i)}{{}''\Sigma \Delta P_2(x_i)} - a_3 \frac{{}''\Sigma \Delta P_3(x_i)}{{}''\Sigma \Delta P_2(x_i)} \dots \beta$$

In II. eingesetzt

$$\begin{aligned} & {}'''\Sigma \Delta f(x_i) = \\ & = \left(\frac{{}''\Sigma \Delta f(x_i)}{{}''\Sigma \Delta P_2(x_i)} - a_3 \frac{{}''\Sigma \Delta P_3(x_i)}{{}''\Sigma \Delta P_2(x_i)} \right) {}'''\Sigma \Delta P_2(x_i) + a_3 {}'''\Sigma \Delta P_3(x_i) \end{aligned}$$

$${}'''\Sigma \Delta f(x_i) - \frac{{}''\Sigma \Delta f(x_i)}{{}''\Sigma \Delta P_2(x_i)} {}'''\Sigma \Delta P_2(x_i) =$$

$$= a_3 \left({}'''\Sigma \Delta P_3(x_i) - \frac{{}''\Sigma \Delta P_3(x_i)}{{}''\Sigma \Delta P_2(x_i)} {}'''\Sigma \Delta P_2(x_i) \right)$$

$${}'''\Sigma \Delta^2 f(x_i) = a_3 {}'''\Sigma \Delta^2 P_3(x_i)$$

$$a_3 = \frac{{}'''\Sigma \Delta^2 f(x_i)}{{}'''\Sigma \Delta^2 P_3(x_i)} \dots (\gamma)$$

(α), (β) und (γ) zeigen, dass wir tatsächlich das gleiche Lösungssystem erhalten wie bei der Multiplikation mit ± 1 ; nur sind darin die alten Summen durch die neuen ersetzt.

Damit sind wir wieder bei der Methode der kleinsten Quadrate angelangt. Ihr Vorteil liegt in der Bestimmung der Koeffizienten, während ein Kriterium für die Funktionen $P(x)$ noch fehlt. Man muss sie einfach ausprobieren. Ändert man aber nachträglich eine davon ab oder fügt eine neue bei, so verlangt das jedesmal wieder eine vollständig neue Berechnung der Koeffizienten.

Hier setzt nun die Arbeit von *Tchebychef* ein. Er ersetzte diese willkürlichen Funktionen $P(x)$ durch ganz bestimmte, die selbst wieder Polynome sind. Sie gehören zur Klasse der orthogonalen Funktionen und werden als «Orthogonale Polynome» bezeichnet. In der Literatur werden sie auch etwa «Tchebychefsche Polynome» genannt. Das mag daher stammen, dass sie Tchebychef zur Ausgleichung heranzog. Die eigentlichen Tchebychefschen Polynome, die eine Kernfunktion enthalten und auch orthogonal sind, werden dabei nicht verwendet.

Die orthogonalen Polynome, die, wie wir sehen werden, ausgezeichnete Eigenschaften besitzen, werden nun, vereint mit der Methode der kleinsten Quadrate, die erwähnten Mängel beheben.

5. Orthogonale Polynome.

1. Bei der Methode der kleinsten Quadrate stiessen wir auf Produkte von der Form:

$$\sum_{x_0}^{x_{n-1}} P_i(x) \cdot P_k(x)$$

Die Funktionen $P_i(x)$, über die nichts vorausgesetzt war, wollen wir ersetzen durch Polynome $T_i(x)$, wo der Index den Grad angeben soll. Diese Polynome $T_i(x)$ wollen wir so definieren, dass wir folgende fundamentale Eigenschaft von ihnen verlangen:

$$(1) \quad \sum_{x_0}^{x_{n-1}} T_l(x) \cdot T_m(x) = 0 \quad \text{wenn } l \neq m$$

Polynome, die diese Eigenschaft besitzen, heissen *orthogonal* für das angegebene Summationsintervall.

2. Wir wollen versuchen, solche orthogonale Polynome zu finden. Dazu gibt es mehrere Wege, wie ja auch die Kategorie dieser Funktionen recht vielgestaltig ist. Sehr systematisch ist der folgende Weg, der sich auf die Differenzenrechnung stützt und eine Formel für

$$\Sigma (u_x \cdot v_x) \text{ gibt } ^2).$$

$E_1 \equiv 1 + \Delta_1$; $\Sigma_1 \equiv \Delta_1^{-1}$ sollen sich nur auf u_x beziehen

$E_2 \equiv 1 + \Delta_2$; $\Sigma_2 \equiv \Delta_2^{-1}$ sollen sich nur auf v_x beziehen

$E \equiv 1 + \Delta$; $\Sigma \equiv \Delta^{-1}$ sollen sich auf die Produktfunktion $(u_x v_x)$ beziehen.

$$\begin{aligned} \Delta (u_x v_x) &= (E_1 E_2 - 1) u_x v_x, & E_2 &= 1 + \Delta_2 \text{ eingesetzt:} \\ &= (E_1 + E_1 \Delta_2 - 1) u_x v_x \\ &= (E_1 \Delta_2 + \Delta_1) u_x v_x \\ &= E_1 \Delta_2 (1 + E_1^{-1} \Delta_1 \Delta_2^{-1}) u_x v_x \end{aligned}$$

Durch Inversion folgt daraus:

$$\begin{aligned} \Sigma (u_x v_x) &= \Delta^{-1} u_x v_x \\ &= E_1^{-1} \Delta_2^{-1} (1 + E_1^{-1} \Delta_1 \Delta_2^{-1})^{-1} u_x v_x \\ &= E_1^{-1} \Delta_2^{-1} (1 - E_1^{-1} \Delta_1 \Delta_2^{-1} + E_1^{-2} \Delta_1^2 \Delta_2^{-2} - + \dots) u_x v_x \\ &= (E_1^{-1} \Delta_2^{-1} - E_1^{-2} \Delta_1 \Delta_2^{-2} + E_1^{-3} \Delta_1^2 \Delta_2^{-3} - + \dots) u_x v_x \\ &= u_{x-1} \Sigma v_x - \Delta u_{x-2} \Sigma^2 v_x + \Delta^2 u_{x-3} \Sigma^3 v_x - + \dots \end{aligned}$$

Nun setzen wir $\overset{\Sigma_1}{u_x} \equiv T_l(x)$ und $\overset{\Sigma_2 + \Sigma_2 \Delta_1^{-1}}{v_x} \equiv T_m(x)$, wobei wir annehmen $m \geq l$

$$\begin{aligned} \Sigma T_m(x) \cdot T_l(x) &= T_l(x-1) \Sigma T_m(x) - \Delta T_l(x-2) \Sigma^2 T_m(x) + \dots + \\ &\quad + (-1)^l \underbrace{\Delta^l T_l(x-l-1) \Sigma^{l+1} T_m(x)}_{\text{konstant}} \end{aligned}$$

Diese unbestimmte Summe nennen wir $S(x)$. Die Summe zwischen den Grenzen $x_0 \rightarrow x_{n-1}$ ist dann

$$\sum_{x_0}^{x_{n-1}} T_m(x) T_l(x) = S(x_n) - S(x_0)$$

Verschwenden nun, wenn $m > l$, die Summen

$$(2) \quad \Sigma T_m(x), \Sigma^2 T_m(x), \dots \Sigma^m T_m(x)$$

alle für die Grenzen $x = x_0$ und $x = x_n$, dann wird auch $S(x_n) - S(x_0)$ verschwinden. Ist aber $m = l$, so wird das letzte Glied nicht null. Das ist gerade, was wir wünschten.

3. Wir haben somit die ursprüngliche Bedingung (1) in (2) übergeführt. Es wird also unsere Aufgabe sein, solche $T_m(x)$ zu finden, die der Bedingung (2) genügen.

Wir setzen:

$$\begin{aligned}\Sigma^m T_m(x) &= G(x) \\ \Sigma^{m-1} T_m(x) &= \Delta G(x) \\ &\dots\dots\dots \\ T_m(x) &= \Delta^m G(x)\end{aligned}$$

Es wird nun verlangt, dass die Funktion $G(x)$ und ihre m ersten Differenzen verschwinden für $x = x_0$ und $x = x_n$. Das bedeutet, dass $G(x)$ selbst verschwinden muss für

$$x = x_0, x = x_0 + 1, x = x_0 + 2, \dots x = x_0 + m - 1$$

$$x = x_n, x = x_n + 1, x = x_n + 2, \dots x = x_n + m - 1$$

Das ist der Fall, wenn $G(x)$ folgende Faktoren enthält:

$$(x - x_0), (x - x_0 - 1), \dots (x - x_0 - m + 1)$$

und $(x - x_n), (x - x_n - 1), \dots (x - x_n - m + 1)$

Da $G(x) = \Sigma^m T_m(x)$ vom Grade $2m$ in x ist, so kann es neben diesen Faktoren höchstens noch eine Konstante als weiteren Faktor enthalten. Die Wahl dieser Konstanten ist beliebig. Wir wählen dafür $\frac{1}{(m!)^2}$, damit die absteigenden Fakultätenfunktionen in Binomialkoeffizienten übergehen.

$$\begin{aligned}G(x) &= \frac{(x - x_0)(x - x_0 - 1) \dots (x - x_0 - m + 1)}{m!} \\ &\quad \cdot \frac{(x - x_n)(x - x_n - 1) \dots (x - x_n - m + 1)}{m!} \\ &= \binom{x - x_0}{m} \binom{x - x_n}{m}\end{aligned}$$

$$(3) \quad T_m(x) = \Delta^m \binom{x-x_0}{m} \binom{x-x_n}{m}$$

Da es nur eine einfache Verschiebung des Intervalls bedeutet, die man jederzeit rückgängig machen kann, dürfen wir ohne weiteres setzen $x_i = i$. Das gefundene orthogonale Polynom lautet dann einfacher

$$(4) \quad T_m(x) = \Delta^m \binom{x}{m} \binom{x-n}{m}$$

4. Umformung von (4).

Es sei $\binom{x-m}{m} = f(x)$ und folglich $\binom{x-n}{m} = f(x+m-n)$

Wir wollen die Differenzenformel von Newton

$$f(x+n) = f(x) + \binom{n}{1} \Delta f(x) + \binom{n}{2} \Delta^2 f(x) + \dots + \Delta^n f(x)$$

anwenden und gerade berücksichtigen, dass

$$\Delta^q \binom{x-m}{m} = \binom{x-m}{m-q}$$

und

$$\binom{m-n}{q} = (-1)^q \binom{n-m+q-1}{q}$$

$$f[x+(m-n)] = f(x) + (m-n) \Delta f(x) + \dots + \binom{m-n}{m} \Delta^m f(x)$$

$$\begin{aligned} \binom{x-n}{m} &= \binom{x-m}{m} + (m-n) \binom{x-m}{m-1} + \binom{m-n}{2} \binom{x-m}{m-2} + \\ &\quad + \dots + \binom{m-n}{m} \end{aligned}$$

$$\begin{aligned} &= \binom{x-m}{m} - (n-m) \binom{x-m}{m-1} + \\ &\quad + \binom{n-m+1}{2} \binom{x-m}{m-2} - \dots + (-1)^m \binom{n-1}{m} \end{aligned}$$

$$\begin{aligned} \binom{x}{m} \binom{x-n}{m} &= \binom{x}{m} \binom{x-m}{m} - (n-m) \binom{x}{m} \binom{x-m}{m-1} + \dots + \\ &\quad + (-1)^m \binom{x}{m} \binom{n-1}{m} \end{aligned}$$

$$\begin{aligned} \binom{x}{m} \binom{x-m}{m} &= \frac{x(x-1) \dots (x-m+1)}{m!} \cdot \\ &\quad \cdot \frac{(x-m)(x-m-1) \dots (x-2m+1)}{m!} \end{aligned}$$

$$= \binom{x}{2m} \frac{(2m)!}{(m!)^2} = \binom{x}{2m} \binom{2m}{m}$$

$$\begin{aligned} \binom{x}{m} \binom{x-m}{m-1} &= \frac{x(x-1) \dots (x-m+1)}{m!} \cdot \\ &\quad \cdot \frac{(x-m)(x-m-1) \dots (x-2m+2)}{(m-1)!} \end{aligned}$$

$$= \binom{x}{2m-1} \frac{(2m-1)!}{m!(m-1)!} =$$

$$= \binom{x}{2m-1} \frac{(2m-1)(2m-2) \dots m}{m!} = \binom{x}{2m-1} \binom{2m-1}{m}$$

$$\begin{aligned} \binom{x}{m} \binom{x-m}{m-2} &= \frac{x(x-1) \dots (x-m+1)}{m!} \cdot \\ &\quad \cdot \frac{(x-m)(x-m-1) \dots (x-2m+3)}{(m-2)!} \end{aligned}$$

$$= \binom{x}{2m-2} \frac{(2m-2)!}{m!(m-2)!} =$$

$$= \binom{x}{2m-2} \frac{(2m-2)(2m-3) \dots (m-1)}{m!} = \binom{x}{2m-2} \binom{2m-2}{m}$$

.....

Eingesetzt:

$$\binom{x}{m} \binom{x-n}{m} = \binom{2m}{m} \binom{x}{2m} - \binom{2m-1}{m} (n-m) \binom{x}{2m-1} + \\ + \binom{2m-2}{m} \binom{n-m+1}{2} \binom{x}{2m-2} - + \dots + (-1)^m \binom{n-1}{m} \binom{x}{m}$$

In (4) eingesetzt:

$$(5) \quad T_m(x) = \binom{2m}{m} \binom{x}{m} - \binom{2m-1}{m} (n-m) \binom{x}{m-1} + \\ + \binom{2m-2}{m} \binom{n-m+1}{2} \binom{x}{m-2} - + \dots + (-1)^m \binom{n-1}{m}$$

*W = Grad
n = Anzahl der
Pol. v. 5*

(5) ist die endgültige Form des gesuchten orthogonalen Polynoms

$$T_0 = 1$$

$$T_1(x) = 2x - (n-1)$$

$$T_2(x) = 6 \binom{x}{2} - 3 \cdot (n-2) x + \binom{n-1}{2}$$

.....

$$\text{Aus } \sum_0^{n-1} x T_m(x) T_l(x) = 0 \quad \text{folgt } \sum_0^{n-1} x T_m(x) \underbrace{T_0(x)}_1 = \sum_0^{n-1} x T_m(x) = 0$$

Das heisst, die orthogonalen Polynome liegen so in ihrem Intervall, dass sich die positiven und die negativen Ordinaten bei der Summation gegenseitig aufheben. Sie liegen entweder punktsymmetrisch oder axialsymmetrisch zur Mitte des Intervalls, je nachdem sie von ungerader oder gerader Ordnung sind.

5. Aus der gefundenen Entwicklung für $T_m(x)$ lässt sich sofort eine solche für $\Delta^k T_m(x)$ angeben; diese Reihe bricht mit dem Glied, das $\binom{x}{0}$ als Faktor enthält, ab.

$$(6) \quad \Delta^k T_m(x) = \binom{2m}{m} \binom{x}{m-k} - \binom{2m-1}{m} (n-m) \binom{x}{m-k-1} + \\ + \dots + (-1)^{m-k} \binom{m+k}{k} \binom{n-k-1}{m-k}$$

Setzen wir in (5) und (6) $x = 0$, so erhalten wir zwei Hilfsformeln, die wir bei der Anwendung gebrauchen werden

$$(5') \quad T_m(0) = (-1)^m \binom{n-1}{m}$$

$$(6') \quad \Delta^k T_m(0) = (-1)^{m-k} \binom{m+k}{k} \binom{n-k-1}{m-k}$$

6. Bemerkenswert ist die Verwandtschaft dieser orthogonalen Polynome mit den *Legendreschen* Polynomen

$$P_m(x) = C \frac{d^m}{dx^m} \{x^m (x-n)^m\}$$

so dass
$$\int_0^n P_m(x) \cdot P_l(x) dx = 0 \quad \text{wenn } m \neq l$$

Die Summation wird also hier kontinuierlich durchgeführt.

7. Wir können aber auch von den früher besprochenen, willkürlichen Funktionen $P(x)$, bei denen der Index nicht den Grad anzugeben braucht, zu orthogonalen Polynomen gelangen.

Es war
$$f(x) = a_0 P_0(x) + a_1 P_1(x) + \dots + a_m P_m(x)$$

Die Ausrechnung führte dann auf die Produktsummen

$$\sum_{x_0}^{x_{n-1}} P_i(x) \cdot P_k(x) = \sigma_{ik} = \sigma_{ki}$$

Nun substituieren wir für die $P(x)$ neue Funktionen $\Phi(x)$ ¹⁾ mit der Erläuterung

$$\Phi_0(x) = \sigma_{00} \quad \Phi_1(x) = \begin{vmatrix} \sigma_{00} & P_0 \\ \sigma_{10} & P_1 \end{vmatrix} \quad \Phi_2(x) = \begin{vmatrix} \sigma_{00} & \sigma_{01} & P_0 \\ \sigma_{10} & \sigma_{11} & P_1 \\ \sigma_{20} & \sigma_{21} & P_2 \end{vmatrix}$$

$$\Phi_m(x) = \begin{vmatrix} \sigma_{00} & \sigma_{01} & \dots & \sigma_{0m-1} & P_0 \\ \sigma_{10} & \sigma_{11} & \dots & \sigma_{1m-1} & P_1 \\ \sigma_{20} & \sigma_{21} & \dots & \sigma_{2m-1} & P_2 \\ \dots & \dots & \dots & \dots & \dots \\ \sigma_{m0} & \sigma_{m1} & \dots & \sigma_{mm-1} & P_m \end{vmatrix}$$

Damit erhalten wir ein neues Polynom:

$$(7) \quad f(x) = A_0 \Phi_0(x) + A_1 \Phi_1(x) + \dots + A_m \Phi_m(x)$$

Diese Funktionen $\Phi(x)$ sind selbst Polynome, und wir werden jetzt beweisen, dass sie orthogonal sind, dass sie also folgender Bedingung genügen:

$$\Sigma \Phi_i(x) \Phi_k(x) = 0 \quad \text{wenn } i \neq k$$

Es sei

$$D_k = \begin{vmatrix} \sigma_{00} & \dots & \sigma_{0k} \\ \dots & \dots & \dots \\ \sigma_{i0} & \dots & \sigma_{ik} \\ \dots & \dots & \dots \\ \sigma_{k0} & \dots & \sigma_{kk} \end{vmatrix} \quad i < k$$

und D_k^{ik} die zu σ_{ik} gehörende Unterdeterminante in D_k . Dann ist:

$$\Phi_i(x) = \begin{vmatrix} \sigma_{00} & \dots & \sigma_{0i-1} & P_0 \\ \dots & \dots & \dots & \dots \\ \sigma_{i0} & \dots & \sigma_{ii-1} & P_i \end{vmatrix} = D_i^{0i} \cdot P_0 + D_i^{1i} \cdot P_1 + \dots + D_i^{ii} \cdot P_i$$

$$\Sigma \Phi_i(x) \Phi_k(x) = D_i^{0i} \sum x P_0 \Phi_k + D_i^{1i} \sum x P_1 \Phi_k + \dots + D_i^{ii} \sum x P_i \Phi_k$$

Darin lautet der Koeffizient von D_i^{pi} ($p \leq i < k$)

$$\begin{aligned} \sum x P_p \Phi_k &= D_k^{0k} \sum x P_0 P_p + D_k^{1k} \sum x P_1 P_p + \dots + D_k^{kk} \sum x P_k P_p \\ &= D_k^{0k} \sigma_{0p} + D_k^{1k} \sigma_{1p} + \dots + D_k^{kk} \sigma_{kp} \end{aligned}$$

Dieser Ausdruck ist nichts anderes als die nach der letzten Spalte entwickelte Determinante D_k . Jedoch steht darin an Stelle der k . Spalte noch einmal die p . Also ist:

$$\sum x P_p \Phi_k = \begin{vmatrix} \sigma_{00} & \dots & \sigma_{0p} & \dots & \sigma_{0k-1} & \sigma_{0k} \\ \sigma_{10} & \dots & \sigma_{1p} & \dots & \sigma_{1k-1} & \sigma_{1k} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \sigma_{k0} & \dots & \sigma_{kp} & \dots & \sigma_{kk-1} & \sigma_{kk} \end{vmatrix}$$

Da diese Determinante zweimal dieselbe Spalte enthält, verschwindet sie. Damit ist wirklich

$$\sum \Phi_i(x) \Phi_k(x) = 0$$

da darin alle Glieder null werden.

Ist aber $i = k$, so wird p , das die Werte $0, 1, 2, \dots i$ annimmt, für $p = i = k$ automatisch der höchste Spaltenindex, so dass in dieser Determinante nicht zwei gleiche Spalten vorkommen:

$$\sum \Phi_i^2(x) \neq 0$$

Die $\Phi(x)$ sind tatsächlich orthogonale Polynome, wenn der Grad mindestens 1 ist.

Beispiel ¹⁾:

$$x_i = 0, 1, 2, 3, 4.$$

Wir nehmen ein Polynom 3. Grades und wählen $P_i(x) = x^i$. Dann ist:

$$\sigma_{00} = \sum_0^4 P_0^2 = \sum_0^4 1 = 5$$

$$\sigma_{01} = \sum_0^4 P_0 P_1 = \sum_0^4 1 \cdot x = 0 + 1 + 2 + 3 + 4 = 10$$

$$\sigma_{02} = \sum_0^4 P_0 P_2 = \sum_0^4 1 \cdot x^2 = 0 + 1 + 4 + 9 + 16 = 30$$

.....

Man findet so das System.

$\sigma_{00} = 5$	$\sigma_{01} = 10$	$\sigma_{02} = 30$	$\sigma_{03} = 100$
$\sigma_{10} = 10$	$\sigma_{11} = 30$	$\sigma_{12} = 100$	$\sigma_{13} = 354$
$\sigma_{20} = 30$	$\sigma_{21} = 100$	$\sigma_{22} = 354$	$\sigma_{23} = 1300$
$\sigma_{30} = 100$	$\sigma_{31} = 354$	$\sigma_{32} = 1300$	$\sigma_{33} = 4890$

Beachtenswert ist die vollständige Symmetrie in bezug auf die Hauptdiagonale.

$$\Phi_0 = \sigma_{00} = 5$$

$$\Phi_1 = \begin{vmatrix} \sigma_{00} & P_0 \\ \sigma_{10} & P_1 \end{vmatrix} = \begin{vmatrix} 5 & 1 \\ 10 & x \end{vmatrix} = 5(x-2)$$

$$\Phi_2 = \begin{vmatrix} \sigma_{00} & \sigma_{01} & P_0 \\ \sigma_{10} & \sigma_{11} & P_1 \\ \sigma_{20} & \sigma_{21} & P_2 \end{vmatrix} = \begin{vmatrix} 5 & 10 & 1 \\ 10 & 30 & x \\ 30 & 100 & x^2 \end{vmatrix} = 50(x^2 - 4x + 2)$$

$$\Phi_3 = \begin{vmatrix} \sigma_{00} & \sigma_{01} & \sigma_{02} & P_0 \\ \sigma_{10} & \sigma_{11} & \sigma_{12} & P_1 \\ \sigma_{20} & \sigma_{21} & \sigma_{22} & P_2 \\ \sigma_{30} & \sigma_{31} & \sigma_{32} & P_3 \end{vmatrix} = \begin{vmatrix} 5 & 10 & 30 & 1 \\ 10 & 30 & 100 & x \\ 30 & 100 & 354 & x^2 \\ 100 & 354 & 1300 & x^3 \end{vmatrix} = 140(5x^3 - 30x^2 + 43x - 6)$$

8. Schliesslich sei noch auf einen dritten Weg hingewiesen, der zu orthogonalen Polynomen führt. Es ist der Weg, den Tchebychef hauptsächlich gegangen ist. Man benützt dazu die Kettenbruchrechnung.

Es seien wieder x_0, x_1, \dots, x_{n-1} die n Beobachtungsstellen. Dann stellen wir folgende provisorische Funktion auf:

$$F(x) = (x - x_0)(x - x_1) \dots (x - x_{n-1})$$

Diese Funktion differenzieren wir und entwickeln $\frac{F'(x)}{F(x)}$ in einen Kettenbruch.

$$\begin{aligned} F &= F' \cdot Q_0 + R_0 \\ F' &= R_0 \cdot Q_1 + R_1 \\ R_0 &= R_1 \cdot Q_2 + R_2 \\ &\dots \dots \dots \\ R_{n-3} &= R_{n-2} \cdot Q_{n-1} \end{aligned}$$

$$\frac{F}{F'} = Q_0 + \frac{R_0}{F'} = Q_0 + \frac{1}{\frac{F'}{R_0}}$$

Wir setzen $\frac{F'}{R_0} = Q_1 + \frac{R_1}{R_0}$ ein:

$$\frac{F}{F'} = Q_0 + \frac{1}{Q_1 + \frac{1}{\frac{R_0}{R_1}}}$$

schliesslich erhält man:

$$\frac{F}{F'} = Q_0 + \frac{1}{Q_1 + \frac{1}{Q_2 + \frac{1}{Q_3 + \frac{1}{\ddots + \frac{1}{Q_{n-2} + \frac{1}{Q_{n-1}}}}}}$$

$$Q_0 = \frac{Q_0}{1} \quad \text{Nenner } N_0 = 1$$

$$Q_0 + \frac{1}{Q_1} = \frac{Q_0 Q_1 + 1}{Q_1} \quad \text{Nenner } N_1 = Q_1$$

$$Q_0 + \frac{1}{Q_1 + \frac{1}{Q_2}} = \frac{Q_0 Q_1 Q_2 + Q_0 + Q_2}{Q_1 Q_2 + 1} \quad \text{Nenner } N_2 = Q_1 Q_2 + 1 \quad \text{etc.}$$

Diese aufeinanderfolgenden Nenner N sind orthogonal, und es gilt:

$$\sum_{x_0}^{x_{n-1}} x N_i N_k = 0 \quad \text{für } i \neq k$$

und

$$\sum_{x_0}^{x_{n-1}} x N_i = 0$$

Damit stellen wir endgültig das Polynom auf

$$(8) \quad f(x) = \alpha_0 N_0 + \alpha_1 N_1(x) + \dots + \alpha_m N_m(x)$$

Beispiel ¹⁾:

Wir nehmen dasselbe Intervall $0 \leq x \leq 4$ wie im vordern Beispiel.

Hilfsfunktion $F(x) = x(x-1)(x-2)(x-3)(x-4)$

Da die Abszissen symmetrisch um die Mitte ($x-2$) liegen, ist es vorteilhaft zu substituieren $(x-2) = z$

$$\begin{aligned} F(z) &= (z+2)(z+1)z(z-1)(z-2) \\ &= z(z^2-1)(z^2-4) \\ &= z^5 - 5z^3 + 4z \end{aligned}$$

$$F'(z) = 5z^4 - 15z^2 + 4$$

Nun entwickeln wir $\frac{F(z)}{F'(z)}$ in einen Kettenbruch.

$$z^5 - 5z^3 + 4z : 5z^4 - 15z^2 + 4 = \frac{z}{5} = Q_0$$

$$\begin{array}{r} z^5 - 5z^3 + \frac{4}{5}z \\ \hline -2z^3 + \frac{16}{5}z = R_0 \end{array}$$

$$5z^4 - 15z^2 + 4 : -2z^3 + \frac{16}{5}z = -\frac{5}{2}z = Q_1$$

$$\begin{array}{r} 5z^4 - 8z^2 \\ \hline -7z^2 + 4 = R_1 \end{array}$$

$$-2z^3 + \frac{16}{5}z : -7z^2 + 4 = \frac{2}{7}z = Q_2$$

$$\begin{array}{r} -2z^3 + \frac{8}{7}z \\ \hline \frac{72}{35}z = R_2 \end{array}$$

$$-7z^2 + 4 : \frac{72}{35}z = -\frac{245}{72}z = Q_3$$

$$\begin{array}{r} -7z^2 \\ \hline 4 = R_3 \end{array}$$

$$\frac{72}{35}z : 4 = \frac{18}{35}z = Q_4$$

Also:

$$\frac{F}{F'} = \frac{z}{5} + \frac{1}{-\frac{5}{2}z + 1} - \frac{\frac{2}{7}z + 1}{-\frac{245}{72}z + 1} - \frac{\frac{18}{35}z}{\frac{18}{35}z}$$

Die aufeinanderfolgenden Nenner lauten:

$$N_0 = 1$$

denn $\frac{z}{5} = Q_0$ wird als Ganzes aufgefasst.

$$N_1 = -\frac{5}{2}z$$

$$N_2 = -\frac{5}{7}z^2 + 1$$

$$N_3 = \frac{175z^3 - 425z}{72}$$

$$N_4 = \frac{35z^4 - 105z^2 + 28}{28}$$

$z = x - 2$ rückwärts ersetzt:

$$N_0 = 1$$

$$N_1 = -\frac{5}{2}(x-2)$$

$$N_2 = \frac{-5x^2 + 20x - 13}{7}$$

$$N_3 = \frac{175x^3 - 1050x^2 + 1675x - 550}{72}$$

$$N_4 = \frac{35x^4 - 280x^3 + 735x^2 - 700x + 168}{28}$$

Diese Kettenbruchentwicklung braucht man aber nicht jedesmal auszuführen, denn es gibt für die Nenner eine Rekursionsformel, die man leicht herleiten kann.

Führen wir für den Kettenbruch mit den Quotienten Q_0, Q_1, \dots, Q_{n-1} die abkürzende Bezeichnung $|Q_0, Q_1, \dots, Q_{n-1}|$ ein, so können wir schreiben, wenn wir diesen Kettenbruch in einen einfachen Bruch zurückführen:

$$|Q_0, Q_1, \dots, Q_{n-1}| = \frac{Z_{n-1}}{N_{n-1}}$$

$$\frac{Z_0}{N_0} = Q_0 \quad \frac{Z_1}{N_1} = \frac{Q_0 Q_1 + 1}{Q_1} \quad \frac{Z_2}{N_2} = \frac{Q_0 Q_1 Q_2 + Q_0 + Q_2}{Q_1 Q_2 + 1}$$

In diesen Brüchen besteht folgendes Rekursionsgesetz:

$$Z_p = Q_p Z_{p-1} + Z_{p-2}$$

$$N_p = Q_p N_{p-1} + N_{p-2}$$

Die Allgemeingültigkeit kann durch vollständige Induktion bewiesen werden. Da das Gesetz für die niedrigsten Indizes offensichtlich gilt, zeigt man, dass aus der Gültigkeit für $p - 1$ sofort auch die für p folgt. Wir setzen also voraus, es gelte:

$$\frac{Z_{p-1}}{N_{p-1}} = \frac{Q_{p-1} Z_{p-2} + Z_{p-3}}{Q_{p-1} N_{p-2} + N_{p-3}}$$

Um daraus $\frac{Z_p}{N_p}$ zu erhalten, muss man für Q_{p-1} setzen $Q_{p-1} + \frac{1}{Q_p}$

$$\begin{aligned} \frac{Z_p}{N_p} &= \frac{\left(Q_{p-1} + \frac{1}{Q_p}\right) Z_{p-2} + Z_{p-3}}{\left(Q_{p-1} + \frac{1}{Q_p}\right) N_{p-2} + N_{p-3}} \\ &= \frac{Q_p (Q_{p-1} Z_{p-2} + Z_{p-3}) + Z_{p-2}}{Q_p (Q_{p-1} N_{p-2} + N_{p-3}) + N_{p-2}} \\ &= \frac{Q_p \cdot Z_{p-1} + Z_{p-2}}{Q_p \cdot N_{p-1} + N_{p-2}} \end{aligned}$$

Damit ist die Allgemeingültigkeit dieser Rekurrenz bewiesen, so dass man die aufeinanderfolgenden Nenner nach der Formel

$$(9) \quad N_i = Q_i N_{i-1} + N_{i-2}$$

berechnen kann.

10. Die Polynome N sind aber nicht gerade einfach, wie wir gesehen haben. Tchebychef fand eine andere Rekursionsformel, die zugleich eine Verallgemeinerung darstellt. Bis jetzt haben wir für die Äquidistanz der Beobachtungspunkte, das heisst für $x_i - x_{i-1}$, immer den Betrag 1 vorausgesetzt. Hat man bei einer Serie von Beobachtungen die Äquidistanz d

$$x_0, x_1 = x_0 + d, x_2 = x_0 + 2d, x_3 = x_0 + 3d, \dots$$

so kann man setzen

$$x_i - x_0 = d \xi_i$$

Damit ist auf der ξ -Achse die Äquidistanz 1 wieder hergestellt und der Anfangspunkt in den Ursprung verschoben.

x_0	$x_1 = x_0 + d$	x_2	x_3
$\xi_0 = 0$	$\xi_1 = 1$	$\xi_2 = 2$	$\xi_3 = 3$

Am Schluss hat man diese Verzerrung wieder rückgängig zu machen.

Diese Äquidistanz d hat Tchebychef in seiner Rekursionsformel berücksichtigt. Bezeichnet man mit $\psi_p(x)$ das orthogonale Polynom p . Grades, so lautet diese Formel:

$$(10) \quad \psi_{p+1}(x) = x \cdot \psi_p(x) - \frac{d^2 \cdot p^2 (n^2 - p^2)}{4 (4p^2 - 1)} \cdot \psi_{p-1}(x)$$

Dabei ist n die Anzahl der Beobachtungen x_0, x_1, \dots, x_{n-1} und d also die Äquidistanz $x_i - x_{i-1}$.

Voraussetzung für Formel (10) ist, dass man den Ursprung in die Mitte des Intervalls $x_0 \rightarrow x_{n-1}$ verschoben hat. Ist das nicht der Fall, so kann man die Verschiebung auch in der Formel nachholen.

Für die Abszissen 1, 2, \dots n z. B. lautet sie:

$$(10') \quad \psi_{p+1}(x) = \left(x - \frac{n+1}{2}\right) \psi_p(x) - \frac{d^2 \cdot p^2 (n^2 - p^2)}{4(4p^2 - 1)} \cdot \psi_{p-1}(x)$$

und für 0, 1, 2, ... $n-1$:

$$(10'') \quad \psi_{p+1}(x) = \left(x - \frac{n-1}{2}\right) \psi_p(x) - \frac{d^2 \cdot p^2 (n^2 - p^2)}{4(4p^2 - 1)} \cdot \psi_{p-1}(x)$$

Beispiel: $x = 0, 1, 2, 3, 4$ $d = 1$ $n = 5$

$$\psi_0 = 1$$

$$\psi_1 = x - 2$$

$$\psi_2 = x^2 - 4x + 2$$

$$\psi_3 = \frac{5x^3 - 30x^2 + 43x - 6}{5}$$

.....

Das Glied vom höchsten Grade hat immer den Koeffizienten 1.

Diese Rekursionsformeln werden zur Berechnung der orthogonalen Polynome häufig benutzt. Wir werden es aber bei unseren Anwendungen vorziehen, die früher gefundene Formel (5) zu verwenden, da für diese die Rechnung in ein sehr bequemes Schema gebracht werden kann.

Nachdem wir gesehen haben, wie man die orthogonalen Polynome finden kann, wollen wir sehen, welche Vorteile sie uns bei der analytischen Ausgleichung bieten.

6. Anwendung der orthogonalen Polynome bei der analytischen Ausgleichung.

Es sei irgendeine Verteilung $(x_0, y_0), (x_1, y_1) \dots (x_{n-1}, y_{n-1})$ beobachtet.

Scheint es uns zweckmässig, zur Ausgleichung dieser Beobachtungsreihe ein Polynom m . Grades zu wählen, so schreiben wir

$$f(x) = a_0 T_0 + a_1 T_1(x) + \dots + a_m T_m(x)$$

Dabei sind also die $T(x)$ orthogonale Polynome, gleichgültig, nach welcher der angegebenen Methoden sie berechnet wurden.

Wir setzen

$$\begin{aligned} R_{(m)}^2 &= \sum_{x_0}^{x_{n-1}} \{a_0 T_0 + a_1 T_1(x) + \dots + a_m T_m(x) - y(x)\}^2 \\ &= \Sigma (f(x) - y(x))^2 \end{aligned}$$

Zur Bestimmung der a verlangen wir, dass $R_{(m)}^2$ minimal werde. Die partiellen Ableitungen nach den a müssen verschwinden. Für die Ableitung nach a_i erhält man:

$$\sum_x T_i(x) (f(x) - y(x)) = 0$$

Nun macht sich die orthogonale Eigenschaft geltend, und es bleibt einzig

$$a_i \sum_x T_i^2(x) - \sum_x \{T_i(x) \cdot y(x)\} = 0$$

woraus

$$(11) \quad a_i = \frac{\sum_x T_i(x) \cdot y(x)}{\sum_x T_i^2(x)}$$

Man sieht, dass a_i nur abhängig ist von $T_i(x)$, an das es gebunden ist, während es unberührt bleibt von allen andern a und T . Das bedeutet, dass man im Polynom $f(x)$ eine Abänderung vornehmen kann, z. B. das Anbringen eines weiteren Gliedes, ohne dadurch, wie es bei den willkürlichen $P(x)$ der Fall war, die bereits gemachte Rechnung zu annullieren.

Man kann das auch direkt zeigen:

$$\begin{aligned} R_{(m)}^2 &= \Sigma \{a_0 + a_1 T_1(x) + \dots + a_m T_m(x) - y(x)\}^2 \\ &= \Sigma a_0^2 + \Sigma a_1^2 T_1^2(x) + \dots + \Sigma a_m^2 T_m^2(x) \\ &\quad - 2 \{a_0 \Sigma y(x) + a_1 \Sigma T_1(x) y(x) + \dots + a_m \Sigma T_m(x) y(x) \\ &\quad + \Sigma y^2(x)\} \end{aligned}$$

Es ist aber

$$a_0 = \frac{\sum y(x)}{n}$$

$$a_0^2 = a_0 \frac{\sum y(x)}{n}$$

$$n \cdot a_0^2 = \sum a_0^2 = a_0 \sum y(x)$$

und allgemein

$$\sum a_i^2 T_i^2(x) = a_i \sum T_i(x) y(x)$$

Eingesetzt in $R_{(m)}^2$

$$R_{(m)}^2 = \sum y^2(x) - a_0 \sum y(x) - a_1 \sum T_1(x) y(x) - \dots - a_m \sum T_m(x) y(x)$$

$${}^m/m+1 \left[R_{(m+1)}^2 = R_m^2 - a_{m+1} \sum T_{m+1}(x) y(x) \right]$$

So kann R^2 Schritt für Schritt berechnet werden, und jede Vergrößerung des Grades vermindert den Wert von R^2 , bis es für ein Polynom vom $(n-1)$. Grade null wird, da in diesem Falle gleichviel Unbekannte a vorhanden sind wie Gleichungen, so dass die Methode der kleinsten Quadrate nicht mehr nötig ist.

Die Güte einer Ausgleichung wird allerdings nicht durch den blossen Wert von R^2 charakterisiert, sondern es kommt auch darauf an, mit welchem Grad man diese Fehler erzielt hat. Lidstone gibt folgendes Kriterium an, das den Grad m mitberücksichtigt:

$$\frac{R^2(m)}{n-m-1} \quad ^2)$$

Schreiten wir z. B. von einem Polynom 3. Grades zu einem solchen 4. Grades, so muss danach, wenn das Resultat mit dem 4. Grad besser sein soll, gelten:

$$\frac{R_4^2}{n-5} < \frac{R_3^2}{n-4}$$

Bemerkung.

Wir haben immer stillschweigend vorausgesetzt, allen Beobachtungen komme dasselbe Gewicht zu. Ist das nicht der Fall, be-

zeichnet z. B. θ_x^2 das Gewicht an der Stelle x , so muss man die Koeffizienten aus der Formel

$$(12) \quad a_i = \frac{\sum T_i(x) \cdot y(x) \cdot \theta_x^2}{\sum T_i^2(x) \cdot \theta_x^2} \quad 1)$$

berechnen. Wir werden unsere Anwendungen jedoch auf den Fall beschränken, wo allen Beobachtungen dasselbe Gewicht anhaftet.

Gliederung der Rechnung.

Trotz der Eleganz der Formel (11) erfordert die Berechnung der a eine beträchtliche Arbeit. Es ist jedoch eine systematische Gliederung möglich, die die Ausrechnungen in ein einfaches Schema bringt und dadurch eine gute Kontrolle schafft.

Dazu definieren wir folgende Summen ²⁾

x	$y(x)$	S^1	S^2	...
0	y_0	$S_0^1 = y_0 + y_1 + y_2 + \dots + y_{n-1}$	$S_0^2 = S_0^1 + S_1^1 + \dots + S_{n-1}^1$...
1	y_1	$S_1^1 = y_1 + y_2 + \dots + y_{n-1}$	$S_1^2 = S_1^1 + \dots + S_{n-1}^1$...
2	y_2	$S_2^1 = y_2 + \dots + y_{n-1}$	$S_2^2 = S_2^1 + \dots + S_{n-1}^1$...
.
.
.
$n-1$	y_{n-1}	$S_{n-1}^1 = y_{n-1}$	$S_{n-1}^2 = S_{n-1}^1 = y_{n-1}$	

Aus (11) folgt:

$$\begin{aligned}
 a_i \sum T_i^2(x) &= \sum T_i(x) y(x) \\
 &= \sum \left\{ y(x) \left(T_i(0) + \binom{x}{1} \Delta T_i(0) + \binom{x}{2} \Delta^2 T_i(0) + \dots + \binom{x}{i} \Delta^i T_i(0) \right) \right\} \\
 &= T_i(0) \sum y(x) + \Delta T_i(0) \sum \binom{x}{1} y(x) + \dots + \Delta^i T_i(0) \sum \binom{x}{i} y(x) \\
 &= T_i(0) S_0^1 + \Delta T_i(0) S_1^1 + \Delta^2 T_i(0) S_2^1 + \dots + \Delta^i T_i(0) S_i^{i+1}
 \end{aligned}$$

$$(13) \quad a_i = \frac{T_i(0) S_0^1 + \Delta T_i(0) S_1^2 + \Delta^2 T_i(0) S_2^3 + \dots + \Delta^i T_i(0) S_i^{i+1}}{\Sigma T_i^2(x)}$$

Nach dieser Formel werden wir bei den Anwendungen die Konstanten berechnen. Dabei brauchen wir also nur die Summen von der Form S_k^{k+1} .

Auch $f(x)$ stellen wir indirekt dar:

$$(14) \quad \begin{cases} f(0) = a_0 + a_1 T_1(0) + a_2 T_2(0) + \dots + a_m T_m(0) \\ \Delta f(0) = a_1 \Delta T_1(0) + a_2 \Delta T_2(0) + \dots + a_m \Delta T_m(0) \\ \Delta^2 f(0) = a_2 \Delta^2 T_2(0) + \dots + a_m \Delta^2 T_m(0) \\ \dots \dots \dots \end{cases}$$

$$(15) \quad f(x) = f(0) + \binom{x}{1} \Delta f(0) + \binom{x}{2} \Delta^2 f(0) + \dots + \binom{x}{m} \Delta^m f(0)$$

Newton

Für die praktische Durchführung ist diese Darstellung die zweckmässigste.

Zusammenstellung der nötigen Formeln:

$$\left\{ \begin{array}{l} (5) \quad T_i(x) = \binom{2i}{i} \binom{x}{i} - \binom{2i-1}{i} (n-i) \binom{x}{i-1} + \\ \quad + \binom{2i-2}{i} \binom{n-i+1}{2} \binom{x}{i-2} + \dots + (-1)^i \binom{n-1}{i} \\ (5') \quad T_i(0) = (-1)^i \binom{n-1}{i} \\ (6') \quad \Delta^k T_i(0) = (-1)^{i-k} \binom{i+k}{k} \binom{n-k-1}{i-k} \\ (13) \quad a_i = \frac{T_i(0) S_0^1 + \Delta T_i(0) S_1^2 + \dots + \Delta^i T_i(0) S_i^{i+1}}{\Sigma T_i^2(x)} = \frac{Z_i}{\Sigma T_i^2(x)} \\ (14) \quad \Delta^i f(0) = a_i \Delta^i T_i(0) + a_{i+1} \Delta^i T_{i+1}(0) + \dots + a_m \Delta^i T_m(0) \\ (15) \quad f(x) = f(0) + \binom{x}{1} \Delta f(0) + \binom{x}{2} \Delta^2 f(0) + \dots + \binom{x}{m} \Delta^m f(0) \end{array} \right.$$

Schema für die Tabellierung von $\Delta^k T_i(0)$ und $\sum_0^{n-1} T_i^2(x)$

$k \backslash i$	0	1	2	3	...
0	1	*	*	*	...
1	0	2 ⁽²⁾	*	*	...
2	0	0	6 ⁽⁴⁾	*	...
3	0	0	0	20	...
.	
.	
$\sum_0^{n-1} T_i^2(x)$	n	*	*	*	

In der Tabelle sind die Diagonalglieder 1, 2, 6, 20, ... für jedes n dieselben. Einzig die mit * bezeichneten Werte sind zu berechnen.

So ist die ganze Arbeit systematisch gruppiert.

Bemerkung.

Bevor wir an die eigenen Versuche herantreten, sei erwähnt, welche Anwendung die Methode kürzlich in Frankreich gefunden hat ⁸⁾.

Es handelt sich um eine Ausgleichung der einjährigen Sterbenswahrscheinlichkeit der französischen Bevölkerung in der Periode von 0 bis 25 Jahren.

Bevor diese Aufgabe nach der Methode von Tchebychef gelöst wurde, war ein besonderer Ausschuss damit beauftragt worden, empirisch ein Polynom aufzustellen und auf dieses die Methode der kleinsten Quadrate anzuwenden. Dieser Ausschuss stellte ein Polynom 6. Grades zusammen, das aber nicht bessere Resultate ergab als das von den angeführten Autoren später hergeleitete Polynom 3. Grades mit Orthogonalfunktionen, ganz abgesehen von der grösseren Anzahl an Wendepunkten.

II. TEIL.

Anwendung auf die schweizerischen Volkssterbetafeln 1921—1930 und 1929—1932.

Bei der Ausgleichung von Sterbetafeln stellen sich oft grosse Schwierigkeiten ein, namentlich, soweit das erste Lebensdrittel in Betracht fällt. In den ersten 2—3 Lebensjahren zeigt die Sterblichkeitskurve einen überaus steilen Abfall, erreicht ungefähr bei $x = 12$ ein Minimum und nimmt später bei $x = 20$ noch einmal vorübergehend ab. Dann steigt q_x monoton an, so dass die Beobachtungsreihe von da an mit Vorliebe durch das Gesetz von Makeham ersetzt wird.

Es soll unsere Aufgabe sein, den Verlauf von q_x in diesem ersten, etwas komplizierten Intervall nach der behandelten Methode von Tchebychef auszugleichen. Dabei werden wir uns der Orthogonalfunktionen von der Form $T(x)$ bedienen, da mit ihnen die besten Erfahrungen gemacht wurden.

Bei den ersten Versuchen stellte es sich heraus, dass ein Polynom 3. Grades zur Darstellung der Sterblichkeit im ersten Lebensdrittel für die schweizerischen Verhältnisse vollständig ungenügend ist. Erstens ist der Abfall am Anfang überaus steil und zweitens hat er sich ungefähr bei $x = 3$ oder 4 schon fast ganz ausgewirkt, so dass eine starke Verflachung folgt. Um diesem Umstand gerecht zu werden, bedürfte man eines so hohen Kurvengrades, dass dadurch der praktische Wert in Frage gestellt würde.

Es bestände die Möglichkeit, mit der Ausgleichung erst beim dritten oder vierten Lebensjahre anzufangen und für die früheren einfach die unausgeglichenen Werte zu nehmen, wie es in der Schweiz bei der Anwendung der Kingschen Methode gemacht wurde ⁹⁾. Jedoch wird ein Polynom dritten Grades auch dem weiteren Verlaufe nicht gerecht.

Nach diesen Überlegungen schien es vorteilhaft, eine Parabel zweiten Grades zu verwenden und diese am Anfang wie auch am Ende durch eine möglichst einfache Korrektur zu ergänzen. Die Korrektur am Anfang muss den steilen Abfall bewirken, und die am Ende muss den Anstieg bei $x = 20$ herum unterbrechen und für später hinausschieben.

Damit diese Korrekturen nur da wirken, wo sie erwünscht sind, und den übrigen Verlauf nicht stören, wurden sie so berechnet, dass sie an gewünschter Stelle aus der reellen Zahlenachse verschwinden.

Auf diese Art und Weise konnten besonders bei den Beobachtungsergebnissen der Periode 1921—1930 recht befriedigende Resultate erzielt werden. Die neuesten Messungen 1929—1932 weisen nach dem Alter 20 jenes erwähnte Abnehmen der Sterblichkeit auf, das durch die neue hygienische und sportlerische Lebensweise hervorgerufen zu sein scheint und jedenfalls bei der Ausgleichung nicht schlechthin übergangen werden darf. Deshalb war es hier nicht möglich, mit einer einzigen Kurve die Sterblichkeit ebensovieler Altersjahre wiederzugeben wie bei der Beobachtungsperiode 1921—1930, wenigstens, wenn die Korrektur nicht zu kompliziert werden sollte.

1. Tafel für die Männer 1921—1930.

Die unausgeglichenen Werte der q_x findet man in der Schlusstabelle neben den ausgeglichenen. Bequemlichkeitshalber führen wir die Rechnung nicht für q_x selbst durch, sondern für $10^3 \cdot q_x$.

Im Intervall 3—19 wollen wir die besprochene Parabel 2. Grades berechnen: $n = 17$

$$S_0^1 = 37,49$$

$$S_1^2 = 303,22$$

$$S_2^3 = 1643,79$$

Tabelle für $\Delta^k T_i(0)$

$k \backslash i$	0	1	2
0	1	— 16	120
1	0	2	— 45
2	0	0	6
$\Sigma T_i^2(x)$	17	1632	69768

$$Z_0 = 37,49$$

$$a_0 = 2,2053$$

$$f(0) = 3,37306$$

$$Z_1 = 6,60$$

$$a_1 = 0,00404$$

$$\Delta f(0) = - 0,45407$$

$$Z_2 = 716,64$$

$$a_2 = 0,01027$$

$$\Delta^2 f(0) = 0,06162$$

Die Parabelgleichung lautet also gemäss (15)

$$f(x) = 3,37306 - 0,45407 x + 0,06162 \binom{x}{2}$$

Der Ursprung liegt aber bei $x = 3$, und um diese Verschiebung rückgängig zu machen, substituieren wir $x/x-3$:

$$(1) \quad f^*(x) = 5,105 - 0,66974 x + 0,03081 x^2$$

Nun tabelliert man die Werte von $f^*(x)$ (siehe Schlusstabelle). Die obere Grenze ist vorläufig noch nicht festgelegt. Für die ersten Stellen werden oft die unausgeglichene Werte gesetzt. Wir wollen davon absehen und für die Alter $x = 0, 1, 2$ eine Korrektur $\varphi(x)$ berechnen, die aber mehr nur den Zweck haben soll, die Kurve zu vervollständigen. Sie haben die Form:

$$(2) \quad \varphi(x) = \frac{35,53 \sqrt{3-x}}{1 + 7,88 x^2}$$

Die zweite Korrektur $\psi(x)$ ist schwieriger zu berechnen. Erstens ist es sehr wichtig zu erkennen, von wo an sie wirken soll. Es schien hier am besten, wenn $\psi(x)$ etwa von $x = 18,8$ ab zu bremsen beginnt. Als Zähler von $\psi(x)$ kann man somit schreiben

$$\sqrt{(x-18,8)^3}$$

Um den Nenner zu bestimmen, nimmt man eine Anzahl Punkte heraus, durch die die ausgeglichene Kurve gehen sollte. Diese können nach einer beliebigen Methode ermittelt werden. An jeder dieser Stellen berechnet man sodann, wie gross der Nenner sein müsste, um eine genaue Übereinstimmung zu liefern. So ergibt sich die Abhängigkeit des Nenners von x , und man kann dafür eine Funktion aufstellen, die, wenn nötig, nach der Methode der kleinsten Quadrate berechnet werden kann.

Als Nenner ergab sich so folgende lineare Funktion:

$$(3) \quad \psi(x) = - \frac{3,6 + 0,025x}{\sqrt{(x-18,8)^3}}$$

$$(4) \quad F(x) = 5,105 - 0,66974x + 0,03081x^2 + \frac{35,53 \sqrt{3-x}}{1 + 7,88 x^2} - \frac{\sqrt{(x-18,8)^3}}{3,6 + 0,025x}$$

Bemerkung: Die Werte der Parabel $f^*(x)$ berechnet man am schnellsten, indem man die zwei ersten Differenzen berechnet. Die zweite ist konstant, und die erste verändert sich stets um diesen konstanten Betrag.

Tabelle Nr. 1.

x	Beobachtete Werte $1000 \cdot q_x$	$f^*(x)$	$\varphi(x)$	$-\psi(x)$	Ausgegliche ne Werte $F(x)$	Abweichungen $F(x) - 1000 \cdot q_x$
0	66,65	5,11	61,54		66,65	0
1	10,13	4,47	5,66		10,13	0
2	4,96	3,89	1,09		4,98	+ 0,02
3	3,42	3,37			3,37	— 0,05
4	2,88	2,92			2,92	+ 0,04
5	2,35	2,52			2,52	+ 0,17
6	2,22	2,20			2,20	— 0,02
7	1,95	1,93			1,93	— 0,02
8	1,87	1,72			1,72	— 0,15
9	1,72	1,57			1,57	— 0,15
10	1,55	1,49			1,49	— 0,06
11	1,41	1,46			1,46	+ 0,05
12	1,38	1,50			1,50	+ 0,12
13	1,51	1,60			1,60	+ 0,09
14	1,76	1,76			1,76	0
15	1,90	1,99			1,99	+ 0,09
16	2,27	2,27			2,27	0
17	2,81	2,62			2,62	— 0,19
18	2,99	3,03			3,03	+ 0,04
19	3,50	3,50		0,02	3,48	— 0,02
20	3,70	4,03		0,32	3,71	+ 0,01
21	3,79	4,62		0,79	3,83	+ 0,04
22	3,86	5,28		1,38	3,90	+ 0,04
23	3,91	6,00		2,06	3,94	+ 0,03
24	3,94	6,78		2,82	3,96	+ 0,02
25	3,99	7,62		3,65	3,97	— 0,02
26	3,85	8,52		4,55	3,97	+ 0,12
27	4,05	9,48		5,49	3,99	— 0,06
28	4,01	10,51		6,49	4,02	+ 0,01
29	3,84	11,59		7,53	4,06	+ 0,22
30	4,29	12,74		8,61	4,13	— 0,16
31	4,36	13,95		9,73	4,22	— 0,14
32	4,41	15,22		10,89	4,33	— 0,08
33	4,35	16,55		12,09	4,46	+ 0,11
34	4,44	17,95		13,32	4,63	+ 0,19
35	4,82	19,41		14,57	4,84	+ 0,02
36	5,08	20,95		15,85	5,10	+ 0,02
37	5,57	22,50		17,15	5,35	— 0,22
38	5,73	24,14		18,49	5,65	— 0,08
39	5,85	25,84		19,85	5,99	+ 0,14
40	6,37	27,61		21,22	6,39	+ 0,02
41	6,83	29,44		22,61	6,83	0
42	7,65	31,33		24,03	7,30	— 0,35
43	7,88	33,27		25,46	7,81	— 0,07
44	8,41	35,29		26,91	8,38	— 0,03
45	9,08	37,36		28,38	8,98	— 0,10

S.M. 1921-30.

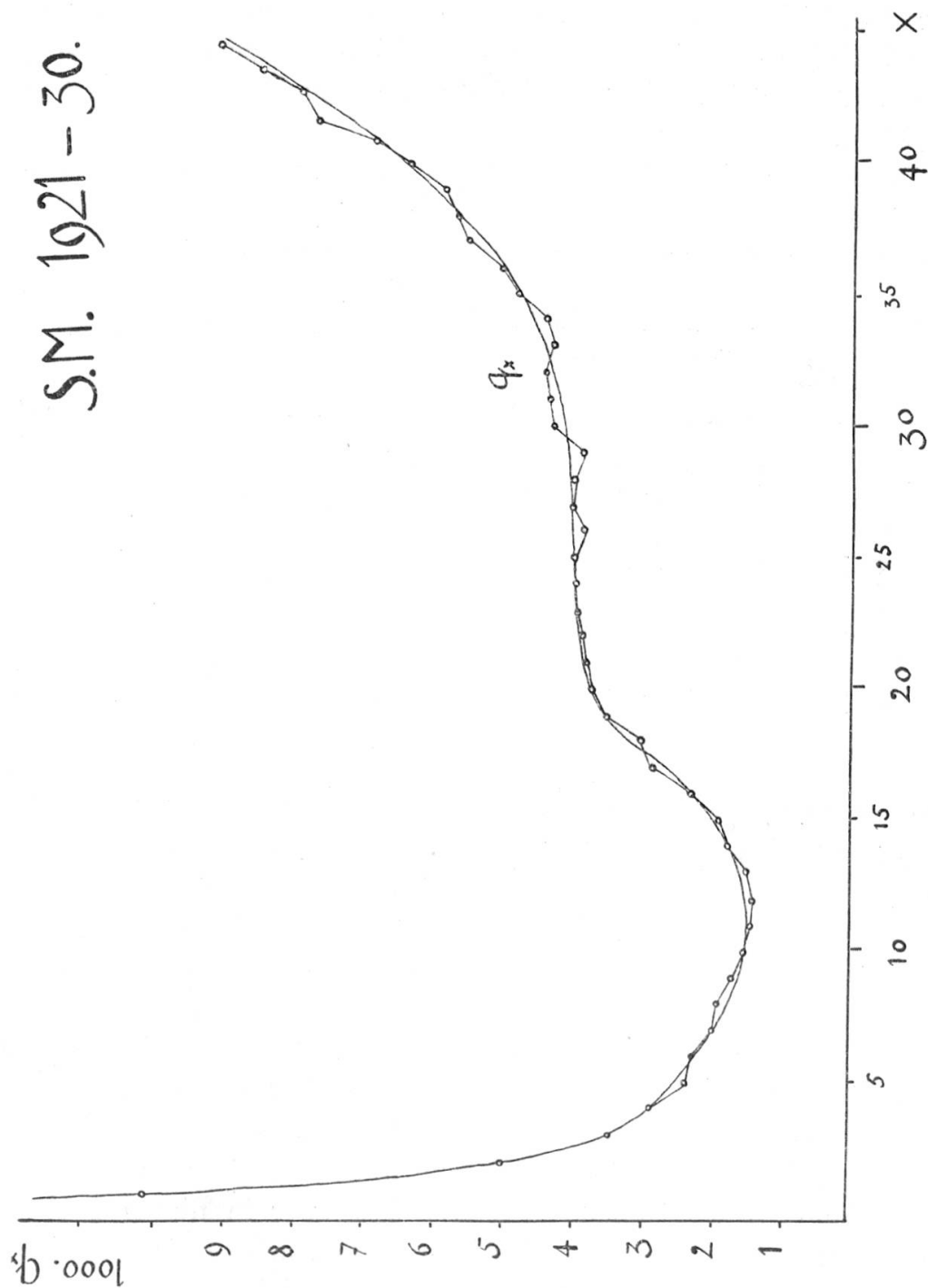


Fig. 4.

Bewertung:

Bei den Versicherungsgesellschaften ist folgende Würdigung einer Ausgleichung üblich: Man wendet die ausgeglichenen Wahrscheinlichkeiten auf die Zahl der unter Risiko gestandenen Personen L_x an und vergleicht die Anzahl der erwarteten Todesfälle mit den wirklichen.

Dieser Vergleich wird hier gebracht für Gruppen von 5 zu 5 Altersjahren.

Altersgruppen	(1) Wirkliche Zahl von Todesfällen	(2) Nach der Ausgleichung des Verfassers erwartete Zahl von Todesfällen	Differenz (2)—(1)
0— 4	29252	29256	+ 4
5— 9	3121	3070	— 51
10—14	2466	2539	+ 73
15—19	4574	4537	— 37
20—24	6159	6183	+ 24
25—29	5628	5686	+ 58
30—34	5469	5443	— 27
35—39	6267	6233	— 34
40—44	8303	8203	— 100
45	1998	1973	— 25

Die Differenzen überschreiten also nur selten 1% der wirklichen Quote. Ausgenommen am Schluss, wo die ausgeglichene Kurve infolge der Korrektur die Tendenz annimmt, zu tief zu verlaufen, sind keine systematischen Fehler vorhanden. Man ersieht auch aus der Tabelle Nr. 1, dass das Vorzeichen der Abweichungen stark wechselt, was naturgemäss ein gutes Zeichen für eine Ausgleichung ist. Je häufiger der Zeichenwechsel, desto besser die Bewertung.

Aus diesen Erwägungen und aus der Betrachtung der graphischen Darstellung ist man geneigt, die durchgeführte Ausgleichung als sehr befriedigend zu bezeichnen. Man vermisst aber in diesen Proben etwas Handfestes, Sicheres, an Stelle des Gefühlsmäßigen. In zweifelhaften Fällen ist es auf diese Weise nicht möglich, einen Entscheid zu fällen.

Dieser Frage betreffend die Güte einer Ausgleichung wird der letzte Teil unserer Arbeit gewidmet werden.

2. Tafel für die Frauen 1921—1930.

Die unausgeglichene Werte findet man wieder in der Schlusstabelle.

Parabel 2. Grades im Intervall 3—16; $n = 14$

$$S_0^1 = 26,50 \quad S_1^2 = 159,69 \quad S_2^3 = 662,41$$

Tabelle für $\Delta^k T_i(0)$

$k \backslash i$	0	1	2
0	1	— 13	78
1	0	2	— 36
2	0	0	6
$\Sigma T_i^2(x)$	14	910	26208

$$Z_0 = 26,50 \quad a_0 = 1,89286 \quad f(0) = 3,12292$$

$$Z_1 = -25,12 \quad a_1 = -0,02760 \quad \Delta f(0) = -0,45732$$

$$Z_2 = 292,62 \quad a_2 = 0,01117 \quad \Delta^2 f(0) = 0,06702$$

$$f(x) = 3,12292 - 0,45732x + 0,06702 \binom{x}{2}$$

$$x/x-3 \quad (1) \quad f^*(x) = 4,897 - 0,69189x + 0,03351x^2$$

$$(2) \quad \varphi(x) = \frac{26,58 \sqrt{3,2-x}}{1 + 7,06x^2}$$

$$(3) \quad \psi(x) = -\frac{\sqrt{(x-17)^3}}{6,7 - 0,07x}$$

$$(4) \quad F(x) = 4,897 - 0,69189x + 0,03351x^2 + \frac{26,58 \sqrt{3,2-x}}{1 + 7,06x^2} - \frac{\sqrt{(x-17)^3}}{6,7 - 0,07x}$$

Tabelle Nr. 2.

x	Beobachtete Werte $1000 \cdot q_x$	$f^*(x)$	$\varphi(x)$	$-\psi(x)$	Ausgegliche- ne Werte $F(x)$	Abweichungen $F(x) - 1000 \cdot q_x$
0	52,45	4,90	47,55		52,45	0
1	9,13	4,24	4,89		9,13	0
2	4,59	3,65	0,99		4,64	+ 0,05
3	3,28	3,12	0,18		3,30	+ 0,02
4	2,44	2,66			2,66	+ 0,22
5	2,28	2,28			2,28	0
6	1,95	1,95			1,95	0
7	1,62	1,69			1,69	+ 0,07
8	1,64	1,51			1,51	— 0,13
9	1,45	1,39			1,39	— 0,06
10	1,42	1,33			1,33	— 0,09
11	1,31	1,34			1,34	+ 0,03
12	1,26	1,42			1,42	+ 0,16
13	1,56	1,56			1,56	0
14	1,77	1,78			1,78	+ 0,01
15	2,13	2,06			2,06	— 0,07
16	2,39	2,40			2,40	+ 0,01
17	2,88	2,82			2,82	— 0,06
18	3,02	3,30		0,18	3,12	+ 0,10
19	3,44	3,85		0,53	3,32	— 0,12
20	3,41	4,46		0,98	3,48	+ 0,07
21	3,59	5,14		1,53	3,61	+ 0,02
22	3,79	5,89		2,16	3,73	— 0,06
23	3,83	6,71		2,88	3,83	0
24	3,93	7,59		3,68	3,91	— 0,02
25	4,10	8,54		4,57	3,97	— 0,13
26	4,02	9,56		5,53	4,03	+ 0,01
27	3,89	10,64		6,57	4,07	+ 0,18
28	4,05	11,80		7,70	4,10	+ 0,05
29	4,11	13,01		8,90	4,11	0
30	4,05	14,30		10,19	4,11	+ 0,06
31	4,07	15,65		11,56	4,09	+ 0,02

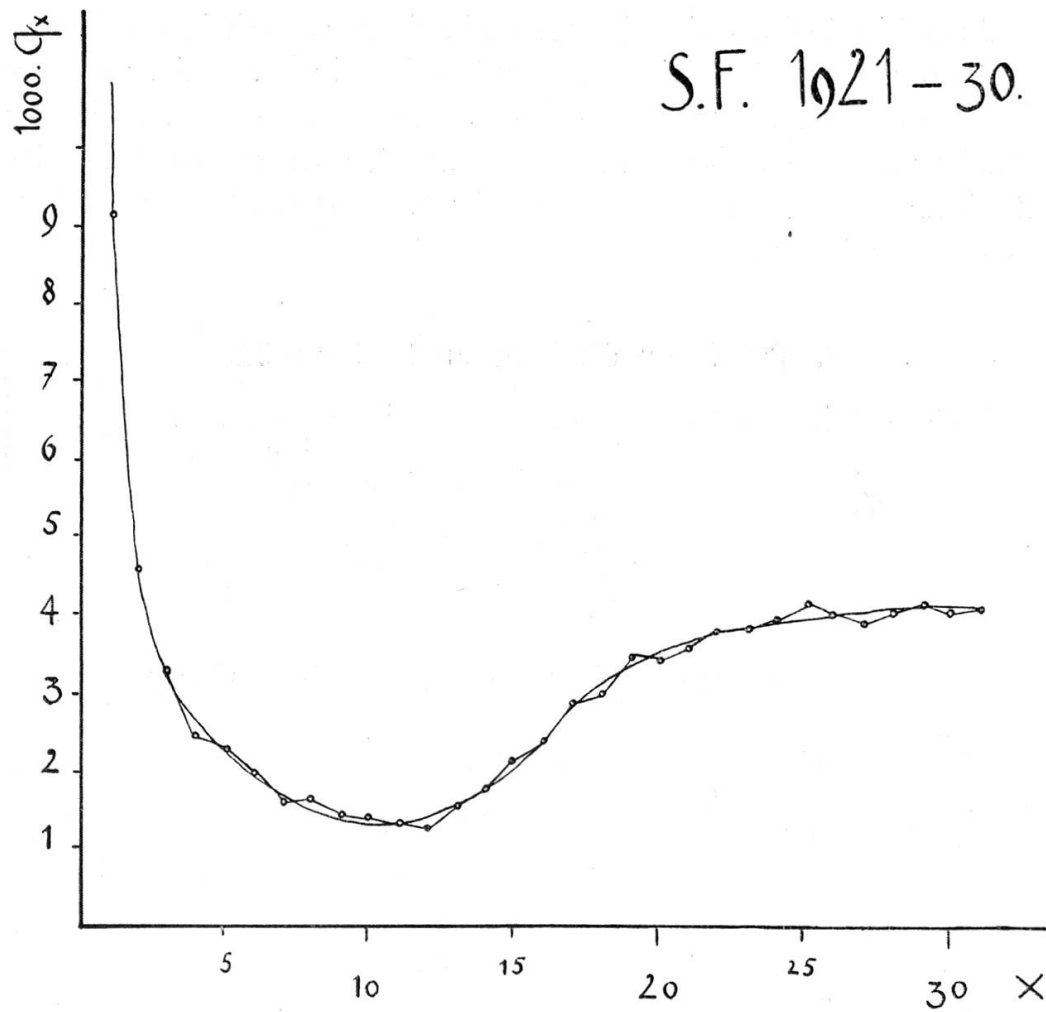


Fig. 5.

Altersgruppen	(1) Wirkliche Zahl von Todesfällen	(2) Nach der Ausgleichung des Verfassers erwartete Zahl von Todesfällen	Differenz (2)—(1)
0— 4	22880	22969	+ 89
5— 9	2708	2674	— 34
10—14	2349	2407	+ 58
15—19	4831	4784	— 47
20—24	6356	6355	— 1
25—29	6396	6416	+ 20
30—31	2371	2390	+ 19

Auch hier bleiben die Differenzen in einem erträglichen Rahmen. Der Erfolg dieser Ausgleichung liegt darin, eine komplizierte Verteilung in einem ziemlich grossen Intervall ohne nennenswerte Fehler bewältigt zu haben. Legt man Wert darauf, dem ganzen Abschnitt mit einer einzigen Funktion zu genügen, so darf man mit diesen Resultaten zufrieden sein.

3. Tafel für die Männer 1929—1932.

Man geht immer nach demselben Arbeitsschema vor:

Quadratische Parabel von $x = 3$ bis $x = 19$:

$$f(x) = 3,12275 - 0,42810 x + 0,06036 \binom{x}{2}$$

$$x/x-3 \quad (1) \quad f^*(x) = 4,769 - 0,63936 x + 0,03018 x^2$$

$$(2) \quad \varphi(x) = \frac{28,3 \sqrt{3,3-x}}{1 + 12,52 \sqrt{x^3}}$$

$$(3) \quad \psi(x) = - \frac{\sqrt{(x-19)^3}}{4,05}$$

$$(4) \quad F(x) = 4,769 - 0,63936 x + 0,03018 x^2 + \frac{28,3 \sqrt{3,3-x}}{1 + 12,52 \sqrt{x^3}} - \frac{\sqrt{(x-19)^3}}{4,05}$$

Da die Beobachtungsdauer dreimal kürzer ist als bei den Messungen 1921—1930, so sind die Beobachtungswerte der q_x ungleichmässiger verteilt, so dass die Abweichungen im Durchschnitt etwas grösser werden.

Dazu kommt noch das besprochene Zurückgehen der Sterblichkeit zwischen 24 und 30, das nicht zufälliger Natur zu sein scheint. Dieses Stück wird von der Funktion $F(x)$ nicht gerade gut ausgeglichen, und es wäre empfehlenswert, ungefähr ab 23 mit einer neuen Ausgleichung anzufangen.

Tabelle Nr. 3.

x	Beobachtete Werte $1000 \cdot q_x$	$f^*(x)$	$\varphi(x)$	$-\psi(x)$	Ausgegliche ne Werte $F(x)$	Abweichungen $F(x) - 1000 \cdot q_x$
0	56,19	4,77	51,42		56,19	0
1	7,33	4,16	3,17		7,33	0
2	4,57	3,61	0,91		4,52	— 0,05
3	3,32	3,12	0,23		3,35	+ 0,03
4	2,40	2,69			2,69	+ 0,29
5	2,24	2,32			2,32	+ 0,08
6	1,99	2,01			2,01	+ 0,02
7	1,94	1,77			1,77	— 0,17
8	1,55	1,58			1,58	+ 0,03
9	1,64	1,46			1,46	— 0,18
10	1,46	1,39			1,39	— 0,07
11	1,41	1,39			1,39	— 0,02
12	1,35	1,44			1,44	+ 0,09
13	1,67	1,56			1,56	— 0,11
14	1,51	1,73			1,73	+ 0,22
15	1,83	1,97			1,97	+ 0,14
16	2,44	2,27			2,27	— 0,17
17	2,64	2,62			2,62	— 0,02
18	2,80	3,04			3,04	+ 0,24
19	3,72	3,52			3,52	— 0,20
20	3,60	4,05		0,25	3,80	+ 0,20
21	4,02	4,65		0,70	3,95	— 0,07
22	3,98	5,31		1,28	4,03	+ 0,05
23	3,94	6,03		1,98	4,05	+ 0,11
24	4,28	6,81		2,76	4,05	— 0,23
25	4,08	7,65		3,63	4,02	— 0,06
26	3,76	8,54		4,57	3,97	+ 0,21
27	3,66	9,50		5,59	3,91	+ 0,25
28	3,93	10,53		6,66	3,87	— 0,06
29	3,80	11,61		7,81	3,80	0

Altersgruppen	(1) Wirkliche Zahl von Todesfällen	(2) Nach der Ausgleichung des Verfassers erwartete Zahl von Todesfällen	Differenz (2)—(1)
0— 4	7827	7853	+ 26
5— 9	984	962	— 22
10—14	729	737	+ 8
15—19	1469	1468	— 1
20—24	2163	2169	+ 6
25—29	1961	1995	+ 34

4. Tafel für die Frauen 1929—1932.

Quadratische Parabel von $x = 3$ bis $x = 19$:

$$f(x) = 2,7713 - 0,37243 x + 0,05238 \binom{x}{2}$$

$$x/x-3 \quad (1) \quad f^*(x) = 4,203 - 0,55576 x + 0,02619 x^2$$

$$(2) \quad \varphi(x) = \frac{26,33 \sqrt{2,3-x}}{1 + 7,16 x}$$

$$(3) \quad \psi(x) = - \frac{\sqrt{(x-18)^3}}{6,25}$$

$$(4) \quad F(x) = 4,203 - 0,55576 x + 0,02619 x^2 + \frac{26,33 \sqrt{2,3-x}}{1 + 7,16 x} - \frac{\sqrt{(x-18)^3}}{6,25}$$

Altersgruppen	(1) Wirkliche Zahl von Todesfällen	(2) Nach der Ausgleichung des Verfassers erwartete Zahl von Todesfällen	Differenz (2)—(1)
0— 4	6077	6086	+ 9
5— 9	878	848	— 30
10—14	622	656	+ 34
15—19	1331	1297	— 34
20—24	1942	1951	+ 9
25—26	850	848	— 2

Tabelle Nr. 4.

x	Beobachtete Werte $1000 \cdot q_x$	$f^*(x)$	$\varphi(x)$	$-\psi(x)$	Ausgegliche- ne Werte $F(x)$	Abweichungen $F(x) - 1000 \cdot q_x$
0	44,12	4,20	39,92		44,12	0
1	7,35	3,67	3,68		7,35	0
2	4,10	3,19	0,94		4,13	+ 0,03
3	2,81	2,77			2,77	— 0,04
4	2,30	2,40			2,40	+ 0,10
5	2,08	2,08			2,08	0
6	1,93	1,81			1,81	— 0,12
7	1,49	1,59			1,59	+ 0,10
8	1,56	1,43			1,43	— 0,13
9	1,46	1,32			1,32	— 0,14
10	1,16	1,26			1,26	+ 0,10
11	1,09	1,25			1,25	+ 0,16
12	1,25	1,30			1,30	+ 0,05
13	1,31	1,40			1,40	+ 0,09
14	1,61	1,56			1,56	— 0,05
15	1,78	1,76			1,76	— 0,02
16	2,15	2,02			2,02	— 0,13
17	2,55	2,32			2,32	— 0,23
18	2,51	2,68			2,68	+ 0,17
19	3,04	3,10		0,16	2,94	— 0,10
20	2,92	3,56		0,45	3,11	+ 0,19
21	3,09	4,08		0,83	3,25	+ 0,16
22	3,57	4,65		1,28	3,37	— 0,20
23	3,49	5,28		1,79	3,49	0
24	3,67	5,95		2,35	3,60	— 0,07
25	3,72	6,68		2,96	3,72	0
26	3,86	7,46		3,62	3,84	— 0,02

Die Genauigkeit der erzielten Resultate kann nach diesen einfachen Proben als sehr befriedigend bezeichnet werden, wofür wir im letzten Teil noch eine genauere Bestätigung finden werden. Grosse Fehler von systematischem Charakter sind nicht vorhanden.

Mit Rücksicht auf das von Lidstone gegebene Kriterium ist es durch eine Erhöhung des Grades der Parabel nicht möglich, eine Verbesserung zu erzielen, denn die Abweichungen würden nicht stark genug verkleinert, dass der Ausdruck

$$\frac{R^2(m)}{n - m - 1}$$

vermindert würde. Schliesslich ist es ja auch gar nicht der Sinn einer Ausgleichung, die Ersatzfunktion in allen Einzelheiten den beobachteten Werten anzupassen; sie soll vielmehr unter Wahrung der charakteristischen Züge die vereinzelt Punkte in einen glatten Verlauf überführen.

* *

Unter Umständen kann jedoch noch auf einem andern Wege eine Verbesserung der Resultate versucht werden.

Erscheint die Reihe der Beobachtungspunkte, die man ausgleichen will, als eine Zusammensetzung von einzelnen, in sich selbst typischen Teilstücken, so kann es sich eventuell lohnen, an Stelle des Gesamtintervalles jedes einzelne dieser Teilstücke für sich mit einem Polynom auszugleichen ³⁾.

Diesem Versuch ist der folgende Abschnitt gewidmet. Es schien mir nämlich, dass der Verlauf um das Minimum einen spezifischen Charakter hat, dem man durch diese Aufteilung noch besser gerecht werden kann.

5. Partielle Tafel für die Männer 1929—1932.

Intervall 0 → 4.

Quadratische Parabel durch $x = 2, 3, 4$:

$$f(x) = 4,57 - 1,25x + 0,33 \binom{x}{2}$$

$$x/x-2 \quad (1) \quad f^*(x) = 8,06 - 2,075x + 0,165x^2$$

$$(2) \quad \varphi(x) = \frac{34,04 \sqrt{2-x}}{1 + 27,85x}$$

$$(3) \quad F(x) = 8,06 - 2,075x + 0,165x^2 + \frac{34,04 \sqrt{2-x}}{1 + 27,85x}$$

x	Beobachtete Werte $1000 \cdot q_x$	$f^*(x)$	$\varphi(x)$	Ausgegliche ne Werte $F(x)$	Abweichungen $F(x) - 1000 \cdot q_x$
0	56,19	8,06	48,13	56,19	0
1	7,33	6,15	1,18	7,33	0
2	4,57	4,57		4,57	0
3	3,32	3,32		3,32	0
4	2,40	2,40		2,40	0

Intervall 4 \rightarrow 12.

Quadratische Parabel:

$$f(x) = 2,4242 - 0,21816x + 0,024 \binom{x}{2}$$

$$x/x-4 \quad f^*(x) = 3,536 - 0,326x + 0,012x^2$$

x	Beobachtete Werte $1000 \cdot q_x$	Ausgegliche ne Werte $f^*(x)$	Abweichungen $f^*(x) - 1000 \cdot q_x$
(4)	2,40	2,42	
5	2,24	2,21	— 0,03
6	1,99	1,99	0
7	1,94	1,84	— 0,10
8	1,55	1,69	+ 0,14
9	1,64	1,57	— 0,07
10	1,46	1,47	+ 0,01
11	1,41	1,40	— 0,01
12	1,35	1,35	0

Intervall 12 \rightarrow 23.

Parabel 3. Grades:

$$f(x) = 1,4318 + 0,0316x + 0,1412 \binom{x}{2} - 0,034 \binom{x}{3}$$

$$x/x-12 \quad f^*(x) = 24,448 - 4,6021x + 0,2917x^2 - 0,00567x^3$$

x	Beobachtete Werte $1000 \cdot q_x$	Aus- gegliche ne Werte $f^*(x)$	Abweichungen $f^*(x) - 1000 \cdot q_x$
(12)	1,35	1,43	
13	1,67	1,46	— 0,21
14	1,51	1,63	+ 0,12
15	1,83	1,91	+ 0,08
16	2,44	2,27	— 0,17
17	2,64	2,66	+ 0,02
18	2,80	3,06	+ 0,26
19	3,72	3,43	— 0,29
20	3,60	3,73	+ 0,13
21	4,02	3,94	— 0,08
22	3,98	4,02	+ 0,04
23	3,94	3,93	— 0,01

Intervall	(1) Wirkliche Zahl von Todesfällen	(2) Nach der Ausgleichung des Verfassers erwartete Zahl von Todesfällen	Differenz (2)—(1)
0— 4	7827	7827	0
5—12	1407	1402	— 5
13—23	3482	3470	— 12

Es ist noch interessant, einen Vergleich mit den offiziellen ausgeglichenen Zahlen des Eidgenössischen Statistischen Amtes zu machen.

Dazu muss voraus bemerkt werden, wie diese Zahlen hergeleitet wurden. Von 0 bis 3 wurden die unausgeglichenen Werte übernommen. Der schwierige Teil von 4 bis 20 wurde stückweise graphisch ausgeglichen, wobei man auf eine möglichst genaue Anpassung an die rohen Werte achtete. Deshalb sollen diese offiziellen Zahlen mit unserer partiellen Ausgleichung verglichen werden. Für Alter über 20 wandte das statistische Amt die Methode von King an.

x	$1000 \cdot q_x$			Abweichungen	
	(1) Rohe Werte	Ausgeglichene Werte		(2) — (1)	(3) — (1)
		(2) Eidg. Stat. Amt	(3) Verfasser		
0	56,19	56,19	56,19	0,00	0,00
1	7,33	7,33	7,33	0	0
2	4,57	4,57	4,57	0	0
3	3,32	3,32	3,32	0	0
4	2,40	2,54	2,40	+ 14	0
5	2,24	2,20	2,21	— 4	— 3
6	1,99	1,99	1,99	0	0
7	1,94	1,82	1,84	— 12	— 10
8	1,55	1,68	1,69	+ 13	+ 14
9	1,64	1,57	1,57	— 7	— 7
10	1,46	1,47	1,47	+ 1	+ 1
11	1,41	1,40	1,40	— 1	— 1
12	1,35	1,38	1,35	+ 3	0
13	1,67	1,48	1,46	— 19	— 21
14	1,51	1,68	1,63	+ 17	+ 12
15	1,83	1,94	1,91	+ 11	+ 8
16	2,44	2,25	2,27	— 19	— 17
17	2,64	2,61	2,66	— 3	+ 2
18	2,80	2,99	3,06	+ 19	+ 26
19	3,72	3,37	3,43	— 35	— 29
20	3,60	3,65	3,73	+ 5	+ 13
21	4,02	3,86	3,94	— 16	— 8
22	3,98	4,02	4,02	+ 4	+ 4
23	3,94	4,06	3,93	+ 12	— 1
				$\Sigma \text{ abs.} = 2,15$	$\Sigma \text{ abs.} = 1,77$
				$\Sigma = - 0,17$	$\Sigma = - 0,17$

Die erzielte Genauigkeit ist also befriedigend.

6. Partielle Tafel für die Frauen 1929—1932.

Intervall 0 → 10.

Parabel von 3 → 10:

$$f(x) = 2,7353 - 93373 x + 0,042 \binom{x}{2}$$

$$x/x-3 \quad (1) \quad f^*(x) = 4 - 0,4843 x + 0,021 x^2$$

$$(2) \quad \varphi(x) = \frac{23,05 \sqrt{3,03 - x}}{1 + 4x + 3,63x^2}$$

$$(3) \quad F(x) = 4 - 0,4843 x + 0,021 x^2 + \frac{23,05 \sqrt{3,03 - x}}{1 + 4x + 3,63x^2}$$

x	Beobachtete Werte $1000 \cdot q_x$	$f^*(x)$	$\varphi(x)$	Aus- gegliche Werte $F(x)$	Ab- weichungen $F(x) - 1000 \cdot q_x$
0	44,12	4	40,12	44,12	0
1	7,35	3,54	3,81	7,35	0
2	4,10	3,11	0,99	4,10	0
3	2,81	2,74	0,08	2,82	+ 0,01
4	2,30	2,40		2,40	+ 0,10
5	2,08	2,10		2,10	+ 0,02
6	1,93	1,85		1,85	— 0,08
7	1,49	1,64		1,64	+ 0,15
8	1,56	1,47		1,47	— 0,09
9	1,46	1,34		1,34	— 0,12
(10)	1,16	1,26		1,26	

Intervall 10 → 17.

Parabel:

$$f(x) = 1,1376 + 0,00432 x + 0,0657 \binom{x}{2}$$

$$x/x-10 \quad f^*(x) = 4,7079 - 0,68553 x + 0,03285 x^2$$

x	Beobachtete Werte $1000 \cdot q_x$	Aus- gegliche Werte $f^*(x)$	Abweichungen $f^*(x) - 1000 \cdot q_x$
10	1,16	1,14	— 0,02
11	1,09	1,14	+ 0,05
12	1,25	1,21	— 0,04
13	1,31	1,35	+ 0,04
14	1,61	1,55	— 0,06
15	1,78	1,81	+ 0,03
16	2,15	2,15	0
17	2,55	2,55	0

Intervall 17 → 23.

Gerade: $f(x) = 2,4897 + 0,1782x$

$$x/x-17 \quad f^*(x) = - 0,5397 + 0,1782x$$

x	Beobachtete Werte $1000 \cdot q_x$	Aus- gegliche Werte $f^*(x)$	Abweichungen $f^*(x) - 1000 \cdot q_x$
(17)	2,55	2,49	
18	2,51	2,67	+ 0,16
19	3,04	2,85	— 0,19
20	2,92	3,02	+ 0,10
21	3,09	3,20	+ 0,11
22	3,57	3,38	— 0,19
23	3,49	3,56	+ 0,07

Intervall	(1) Wirkliche Zahl von Todesfällen	(2) Nach der Ausgleichung des Verfassers erwartete Zahl von Todesfällen	Differenz (2)—(1)
0— 9	6955	6954	— 1
10—17	1318	1318	0
18—23	2158	2164	+ 6

x	$1000 \cdot q_x$			Abweichungen	
	(1) Rohe Werte	Ausgeglichene Werte		(2) — (1)	(3) — (1)
		(2) Eidg. Stat. Amt	(3) Verfasser		
0	44,12	44,12	44,12	0,00	0,00
1	7,35	7,35	7,35	0	0
2	4,10	4,10	4,10	0	0
3	2,81	2,81	2,82	0	+ 1
4	2,30	2,36	2,40	+ 6	+ 10
5	2,08	2,08	2,10	0	+ 2
6	1,93	1,84	1,85	— 9	— 8
7	1,49	1,65	1,64	+ 16	+ 15
8	1,56	1,50	1,47	— 6	— 9
9	1,46	1,36	1,34	— 10	— 12
10	1,16	1,20	1,14	+ 4	— 2
11	1,09	1,11	1,14	+ 2	+ 5
12	1,25	1,19	1,21	— 6	— 4
13	1,31	1,35	1,35	+ 4	+ 4
14	1,61	1,57	1,55	— 4	— 6
15	1,78	1,82	1,81	+ 4	+ 3
16	2,15	2,11	2,15	— 4	0
17	2,55	2,39	2,55	— 16	0
18	2,51	2,64	2,67	+ 13	+ 16
19	3,04	2,86	2,85	— 18	— 19
20	2,92	3,05	3,02	+ 13	+ 10
21	3,09	3,23	3,20	+ 14	+ 11
22	3,57	3,40	3,38	— 17	— 19
23	3,49	3,54	3,56	+ 5	+ 7
				$\Sigma \text{ abs.} = 1,71$	$\Sigma \text{ abs.} = 1,63$
				$\Sigma = - 0,09$	$\Sigma = - 0,05$

Es ist deutlich ersichtlich, dass man durch diese partielle Ausgleichung dem wirklichen Verlaufe besser gerecht wird als mit einer einzigen Funktion. Dafür bedeutet diese Methode gewissermassen eine Vergewaltigung der Verhältnisse bei den Zusammensetzstellen,

indem dort die q_x -Funktion sowie die Intensität unstetig werden. Dadurch wird die Brauchbarkeit bei diesen Stellen fragwürdig.

Es hängt vom vorliegenden Beispiel ab, zu entscheiden, ob sich diese Aufteilung lohnt. Bemerkenswert ist, dass es Methoden gibt, die es erlauben, bei den Ansatzstellen die Stetigkeit wieder herzustellen.

Über solche Methoden hat J. P. Gram Veröffentlichungen gemacht, und es soll hier eine davon zitiert werden ³⁾:

x sei ein Punkt, wo zwei Intervalle zusammenstossen. Dabei wird man natürlich einen solchen Punkt nicht da wählen, wo die Verteilung grosse Schwankungen macht. Die Umgebung von x , z. B. das Stück $x - h$ bis $x + h$ sei zweimal ausgeglichen worden durch $f_1(x)$ und $f_2(x)$.

Lineare Kombination: $F(x) = \lambda(x) f_1(x) + (1 - \lambda(x)) f_2(x)$

Für $\lambda(x)$ wählt man passende echte Brüche. Diese werden so bestimmt, dass sie nur von h abhängig sind, für $h = 0 \frac{1}{2}$ werden und für positive h gegen 1, für negative gegen 0 streben.

Solche Brüche findet man z. B., wenn man in $(1 + 1)^m$ die Glieder allmählich addiert und durch 2^m dividiert.

$$\begin{array}{l} \text{Z. B. ergibt } m = 3 \text{ die Reihe} \quad 8\lambda = 0, 1, 4, 7, 8 \\ h = 2 \quad \quad \quad 8(1 - \lambda) = 8, 7, 4, 1, 0 \\ \\ f_1(x) \dots f_1(x-2), f_1(x-1), f_1(x), f_1(x+1) \\ f_2(x) \dots \quad \quad f_2(x-1), f_2(x), f_2(x+1), f_2(x+2) \\ \\ F(x) \dots f_1(x-2), \frac{7f_1(x-1) + f_2(x-1)}{8}, \frac{4f_1(x) + 4f_2(x)}{8}, \\ \quad \quad \quad \frac{f_1(x+1) + 7f_2(x+1)}{8}, f_2(x+2) \end{array}$$

Je mehr überschobene Werte vorliegen, desto glatter wird der Übergang. Vollständige Kontinuität kann aber nie erreicht werden.

Wegen der Methode der kleinsten Quadrate war

$$\sum T_i(x) \cdot f(x) = \sum T_i(x) \cdot y(x)$$

Momentensumme der ausgeglichenen Werte = Momentensumme der unausgeglichenen Werte.

Diese Eigenschaft, die für $f_1(x)$ und $f_2(x)$ nach Herleitung erfüllt ist, geht bei der Ausglättung verloren, so dass um die Stelle x herum unter Umständen eine Verschlechterung eintritt.

In derselben Arbeit hat Gram noch andere Methoden angegeben. Oft wird es aber schon genügen, wenn man bei den Grenzstellen die beiden Kurven genügend übereinandergreifen lässt.

Bei den vorausgegangenen Anwendungen wurden diese Betrachtungen nicht verwirklicht. Es wurde hauptsächlich danach getrachtet, die einzelnen typischen Stücke gut herauszupräparieren.

III. TEIL.

Die Güte einer Ausgleichung.

1. Kriterium für den Grad der Ausgleichsfunktion.

Es ist schon ein Kriterium erwähnt worden, das über die Güte der Ausgleichung entscheidet, wenn man hinsichtlich des zu wählenden Parabelgrades im Zweifel ist. In diesem Falle entscheidet also nicht die Summe der quadratischen Fehler allein, sondern man hat sie zu dividieren durch $(n-m-1)$, wo n die Anzahl der beobachteten Merkmale und m der Grad der Parabel ist.

Es sei als Beispiel folgende Verteilung gegeben:

x	0	1	2	3	4	5	6	7	8	9	10	11
y	135	167	151	183	244	264	280	372	360	402	398	394

Diese Zahlen stammen aus $10^5 \cdot q_x$ für Männer 1929—1932, vom 12. bis 23. Altersjahr. Wir wollen sie mit folgenden vier Polynomen ausgleichen:

$$f_0 = a_0$$

$$f_1(x) = a_0 + a_1 T_1(x)$$

$$f_2(x) = a_0 + a_1 T_1(x) + a_2 T_2(x)$$

$$f_3(x) = a_0 + a_1 T_1(x) + a_2 T_2(x) + a_3 T_3(x)$$

Durch den üblichen Prozess findet man:

$$f_0 = 279,17$$

$$f_1(x) = 125,84 + 27,86 x$$

$$f_2(x) = 115,01 + 33,77x - 1,18 \binom{x}{2}$$

$$f_3(x) = 143,18 + 3,16x + 14,12 \binom{x}{2} - 3,4 \binom{x}{3}$$

x	y	f_0	$f_1(x)$	$f_2(x)$	$f_3(x)$
0	135	279	126	115	143
1	167	279	154	149	146
2	151	279	181	181	163
3	183	279	209	213	191
4	244	279	237	243	227
5	264	279	265	272	266
6	280	279	293	300	306
7	372	279	321	327	343
8	360	279	349	352	373
9	402	279	377	377	394
10	398	279	404	399	402
11	394	279	432	421	393

$$R_0^2 = 118056 \quad R_1^2 = 6872 \quad R_2^2 = 6433 \quad R_3^2 = 2773$$

$$\frac{R_0^2}{11} = 10732 \quad \frac{R_1^2}{10} = 687 \quad \frac{R_2^2}{9} = 715 \quad \frac{R_3^2}{8} = 347$$

Man erkennt, dass z. B. die Gerade besser bewertet wird als die quadratische Parabel, da ihre Fehler nur unbedeutend grösser sind. Wie man sieht, ist der 3. Grad hier am Platze. Eine weitere Graderhöhung würde nach Lidstone wieder zu einer Verschlechterung führen.

Dieses Kriterium ist jedoch in seiner Anwendungsmöglichkeit beschränkt, indem es nur gerade für ganze rationale Funktionen gilt. Ferner bedingt es, dass das gegebene Material durch mindestens zwei Polynome verschiedenen Grades ausgeglichen worden ist. In diesem Falle gibt es nur eine relative Beziehung zwischen der Güte dieser Ausgleichungen, sagt aber nicht aus, welche davon in ein erlaubtes Mass fallen und welche es überschreiten. Es fehlt also die absolute Beurteilung.

Wir wollen nun ein anderes Kriterium heranziehen, das diese Mängel nicht mehr aufweist.

2. Die χ^2 -Methode.

Mit der Entwicklung dieser Theorie sind die Namen Pearson ¹⁰⁾, Elderton ⁵⁾, R. A. Fisher ⁴⁾, Cramer ⁶⁾, Anderson ⁷⁾, Helmert ⁷⁾ und Bortkiewicz ⁷⁾ verbunden.

Gegeben sei irgendeine Verteilung, gleichgültig, an welchem Material sie beobachtet worden sei. Sie werde dann ersetzt durch eine gesetzmässige, theoretische, sei es durch analytische Ausgleichung oder auf einem andern Weg. Dann drängt sich sofort die Frage auf, ob diese theoretische Verteilung auch ruhig an Stelle der beobachteten verwendet werden dürfe, das heisst, ob sie das Gesetz, das die unausgeglichenen Zahlen schon durchblicken lassen, gut wiedergebe oder ob sie systematische Abweichungen davon zeige. Schliesslich ist auch von Bedeutung zu wissen, ob die beobachtete Verteilung überhaupt homogen genug ist, um durch eine zuverlässige theoretische ersetzt werden zu können. Die Grösse der einzelnen Fehler oder die Summe ihrer Absolutbeträge gibt uns nur in den extremsten Fällen zuverlässigen Aufschluss.

Ein interessantes Kriterium zur Beurteilung der Güte der Anpassung ist die χ^2 -Methode. Wie wir sehen werden, macht sie fundamentale Voraussetzungen über die Struktur des Materials, die nicht in jedem praktischen Falle erfüllt sind, so dass ihre Anwendbarkeit dadurch beschränkt wird.

Das Wesentliche der χ^2 -Methode besteht darin, dass man nicht die beobachtete Verteilung, sondern die theoretische als die primäre auffasst. Sie stellt die zu erwartende Verteilung dar, das heisst, das Gesetz des Vorganges. Von ihr ausgehend, berechnet man sodann die Wahrscheinlichkeit für das Eintreffen einer gleich guten oder schlechteren Verteilung als die beobachtete. Die Wahrscheinlichkeit für eine genau gleich gute wäre praktisch verschwindend klein, weshalb man die Wahrscheinlichkeit für eine höchstens gleich gute sucht.

An jeder Stelle wird also der ausgeglichene Wert als der zu erwartende aufgefasst. Nun setzt die χ^2 -Methode voraus, dass sich bei oftmaliger Wiederholung derselben Messungen die Häufigkeiten in jeder einzelnen Merkmalklasse nach der Gauss'schen Art, das heisst in Glockenform, um den betreffenden theoretischen Mittelwert anordnen.

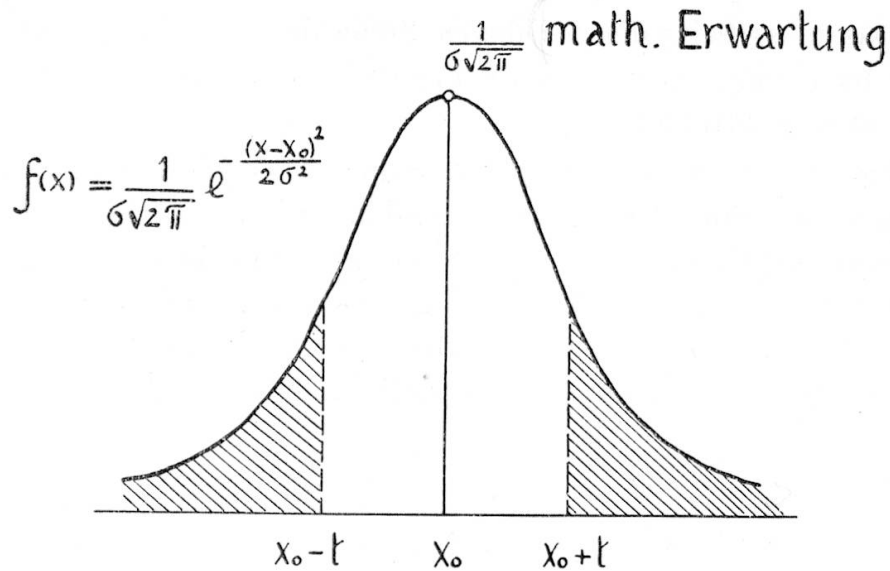


Fig. 6.

Normalverteilung wenn $\sigma = 1$.

Die Wahrscheinlichkeit für eine Abweichung $\geq t$ vom Mittelwert x_0 wird gegeben durch die beiden äusseren Flächenstücke, dividiert durch die Gesamtfläche unter der Kurve:

$$P_1 = \frac{\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x_0-t} e^{-\frac{(x-x_0)^2}{2\sigma^2}} dx + \frac{1}{\sigma\sqrt{2\pi}} \int_{x_0+t}^{\infty} e^{-\frac{(x-x_0)^2}{2\sigma^2}} dx}{\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(x-x_0)^2}{2\sigma^2}} dx} =$$

$$= \frac{2 \int_{x_0+t}^{\infty} e^{-\frac{(x-x_0)^2}{2\sigma^2}} dx}{2 \int_{x_0}^{\infty} e^{-\frac{(x-x_0)^2}{2\sigma^2}} dx}$$

$$P_1 = \frac{\int_{x_0+t}^{\infty} e^{-\frac{(x-x_0)^2}{2\sigma^2}} dx}{\int_{x_0}^{\infty} e^{-\frac{(x-x_0)^2}{2\sigma^2}} dx}$$

Das muss bei jeder beliebigen Stelle der Verteilung gelten, bei jeder nach einer andern unabhängigen Veränderlichen x und mit einer andern Streuung.

Fassen wir nun zwei Stellen auf einmal ins Auge und betrachten die Resultierende der beiden Einzelabweichungen als totale Abweichung, so erhalten wir eine zusammengesetzte Gauss'sche Funktion $z = f(x, y)$, die sich also nach zwei unabhängigen Variablenrichtungen verändern kann. Wir erhalten als Bild nicht mehr eine Kurve, sondern eine Fläche, eine sogenannte Korrelationsfläche.

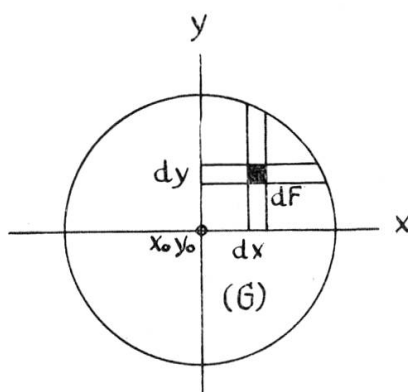


Fig. 7.

Die Mittelwerte der beiden Verteilungen legen wir in den Punkt x_0, y_0 und wählen diesen als Basispunkt der Fläche $f(x, y)$.

Die Masse der Verteilung, die über dem Gebiet (G) liegt, wird gegeben durch das Doppelintegral

$$\iint_{(G)} z dF = \iint_{(G)} f(x, y) dx dy \dots \text{Volumen.}$$

Erklärt man $f(x, y)$ als Wahrscheinlichkeit ^{sdichte} und integriert in den beiden Richtungen von $-\infty$ bis $+\infty$, so wird dieses Doppelintegral analog wie das einfache, bei einer einzigen Variablen, $= 1$.

Jetzt handelt es sich darum, die Funktion $z = f(x, y)$ zu finden. Die unabhängigen Funktionen lauten:

$$f(x) = \frac{1}{\sigma_1 \sqrt{2\pi}} e^{-\frac{(x-x_0)^2}{2\sigma_1^2}} \quad f(y) = \frac{1}{\sigma_2 \sqrt{2\pi}} e^{-\frac{(y-y_0)^2}{2\sigma_2^2}}$$

Die zusammengesetzte Gauss'sche Funktion ist das Produkt der unabhängigen. Sie lautet also:

$$Z = f(x, y) = \frac{1}{2\pi \sigma_1 \sigma_2} e^{-\frac{1}{2} \left\{ \frac{(x-x_0)^2}{\sigma_1^2} + \frac{(y-y_0)^2}{\sigma_2^2} \right\}}$$

$$\frac{(x-x_0)^2}{\sigma_1^2} + \frac{(y-y_0)^2}{\sigma_2^2} = h^2$$

sind konzentrische Ellipsen um (x_0, y_0) . Wenn $\sigma_1 = \sigma_2$ ist, gehen sie in Kreise über.

Die Wahrscheinlichkeit P_2 , für das Eintreffen gewisser Mindestabweichungen t_1 und t_2 vom Mittel $f(x_0, y_0)$ erscheint nun als Quotient zweier Doppelintegrale.

Haben wir nicht nur eine zweifach zusammengesetzte Funktion, sondern eine solche mit n unabhängigen Variablen, so werden wir folglich für P_n n -fache Integrale zu lösen haben.

Pearson gelang es, diese n -fachen Integrale auf einfache zurückzuführen ¹⁰⁾, und er erhielt für die Wahrscheinlichkeit P_n den Ausdruck:

$$P_n = \frac{\int_0^\chi e^{-\frac{1}{2}x^2} \cdot x^{n-1} dx}{\int_0^\infty e^{-\frac{1}{2}x^2} \cdot x^{n-1} dx}$$

Die untere Grenze χ ist eine Funktion der einzelnen Abweichungen in den n verschiedenen Richtungen. Sie drückt also noch die Abhängigkeit von den n Variablen aus. Bezeichnen wir mit t_i die ins Auge gefasste Abweichung vom Mittelwert T_i in der i . Variablenrichtung, so ist:

$$\chi^2 = \sum_{i=1}^n \frac{t_i^2}{T_i}$$

Sehen wir nun zu, welche Form χ^2 bei unserem Ausgleichungsproblem annimmt! An den n Stellen x_0, x_1, \dots, x_{n-1} seien die Werte y_0, y_1, \dots, y_{n-1} beobachtet worden und ausgeglichen durch eine Funktion $f(x)$.

Nun nimmt man an, $f(x)$ stelle bei jedem betrachteten Punkte den Mittelwert dar, und y_i z. B. sei, wenn man viele Messungen machte, normal um das Mittel $f(x_i)$ verteilt. Die Gesamtheit der n Beobachtungsstellen zusammen liefert die Problemstellung, die von Pearson behandelt wurde.

Damit ist:

$$\chi^2 = \sum_{x_0}^{x_{n-1}} \frac{(f(x) - y(x))^2}{f(x)}$$

$$= \sum \frac{(\text{ausgeglicherer, d. h.: erwarteter Wert} - \text{beobachteter Wert})^2}{\text{ausgeglicherer Wert}}$$

Wächst χ von 0 bis ∞ , so nimmt P ab von 1 bis 0. Es ist also diejenige Ausgleichung die beste, die das kleinste χ beziehungsweise χ^2 liefert; denn für diesen Wert wird P_n am grössten. P bedeutet aber die Wahrscheinlichkeit für gleich grosse oder, was wichtig ist, für noch grössere Fehler.

Elderton ¹²⁾ und Fisher ⁴⁾ haben die Grössen n , χ^2 , P_n tabuliert, so dass man die entsprechenden Werte nur abzulesen braucht.

Bei der praktischen Anwendung ist es schwer, einen Wert für P anzugeben, der die Grenze der Brauchbarkeit einer Ausgleichung darstellt. Fisher und Anderson nehmen dafür ungefähr den Wert $P = 0,05$ an; das heisst, nur in einem Fall von 20 sind gleich grosse oder noch grössere Fehler als die vorliegenden zu erwarten. Es kommt übrigens gar nicht auf eine genaue Kenntnis von P an. Beträgt dieses etwa 0,05 oder noch weniger, so bedeutet das, dass Grund vorhanden ist, dem aufgestellten Ausgleichungsgesetz zu misstrauen. Voraussetzung ist natürlich immer, dass genügend Beobachtungen gemacht wurden, so dass die gegebenen Punkte das Verteilungsgesetz durchblicken lassen.

Der Willkürlichkeitsgrad einer Verteilung.

Man muss bei der Anwendung der χ^2 -Methode noch einen weiteren Faktor berücksichtigen, den sogenannten Willkürlichkeitsgrad oder Freiheitsgrad der Verteilung:

Es seien N Beobachtungen auf n Merkmalgruppen verteilt, so, dass den Punkten x_0, x_1, \dots, x_{n-1} die Mengen y_0, y_1, \dots, y_{n-1} ent-

sprechen. Diese gegebene Verteilung werde nun ersetzt durch ein Gesetz $f(x)$. Es gibt Fälle, bei denen gelten muss:

$$\sum_{x_0}^{x_{n-1}} f(x) = \sum_{x_0}^{x_{n-1}} y(x) = N$$

das heisst, das Gesetz $f(x)$ muss sich an die vorgegebene Elementenzahl N halten. Es hat also die Aufgabe, genau die N Elemente, nicht mehr und nicht weniger, in die n Gruppen zu verteilen. Man könnte also unter Umständen die Mengen in $n-1$ Gruppen willkürlich wählen, natürlich nicht $> N$; dann ist aber die Menge in der n . Gruppe festgelegt. Man kann nicht mehr willkürlich über sie verfügen, denn sie beträgt genau N , vermindert um die Zahl der schon verfügbaren Elemente. Man sagt, der Willkürlichkeitsgrad sei in diesem Falle gleich $n-1$.

Bedingt man etwa aus, dass nicht nur die Gesamtzahl N in beiden Verteilungen übereinstimmen muss, sondern auch der häufigste Wert, so schränkt das die Verfügungsfreiheit weiter ein, und der Willkürlichkeitsgrad sinkt auf $n-2$.

Es ist nun klar, dass, je mehr solcher Bedingungen man voraussetzt, die die beobachtete und die berechnete Verteilung aneinanderknüpfen, desto kleiner von selbst die Abweichungen werden. Dieser Umstand muss berücksichtigt werden, wenn man die Güte einer Ausgleichung beurteilen will.

Es kann gerade an dieser Stelle bemerkt werden, dass bei Sterbetafeln der Willkürlichkeitsgrad identisch ist mit n , denn es existieren keine Postulate, die die Freiheit der Ausgleichung einschränken. Die Summe der ausgeglichenen q_x -, l_x - oder d_x -Werte braucht nicht mit der der unausgeglichenen übereinzustimmen.

Elderton ⁵⁾ wies darauf hin, dass man diesen Faktor beobachten müsse, um das berechnete P richtig einzuschätzen. Er meint aber, es sei ohne Verwirrung nicht möglich, diese Grösse systematisch in die Rechnung einzubeziehen. In seiner Tabelle bedeutet n die Anzahl der Merkmalklassen. Fisher ⁴⁾ dagegen misst dieser Tatsache grössere Bedeutung zu. In seiner Tabelle bedeutet n nicht mehr die Anzahl der Merkmale, sondern die Anzahl derjenigen von ihnen, die man beliebig besetzen kann, also den Willkürlichkeitsgrad. Da es aber nicht auf den genauen, sondern nur auf den angenäherten Wert von

P ankommt, so hat dieser Unterschied in den meisten Fällen keine allzugrosse Bedeutung.

Es folgt ein Ausschnitt aus der Tabelle von Fisher ⁴⁾ (Seiten 96/97):

$n \backslash P_n$	0,99	0,98	...	0,05	...	0,01
1	0,000157	0,000628		3,841		6,635
2	0201	0404		5,991		9,210
3	115	185		7,815		11,341
4	297	429		9,488		13,277
5	554	752		11,070		15,086
6	872	1,134		12,592		16,812
7	1,239	564		14,067		18,475
.
.
.
.
.
30	14,953	16,306		43,773		50,892

Übersteigt n den Wert 30, so kann man die Kolonnen ein Stück weit extrapolieren und erkennt dann leicht, in welche das berechnete χ^2 fällt. Ferner gibt Fisher ⁴⁾ eine Ersatzmethode an, die um so genauer wird, je mehr n wächst: Für eine grosse Anzahl von Messungen an demselben Material nimmt nämlich χ angenähert die Form einer Normalverteilung an, und zwar um so genauer, je grösser n ist. Eine gute Näherung ist die Annahme, dass $\sqrt{2}\chi^2$ normal um das Mittel $\sqrt{2n-1}$ verteilt ist. $\sqrt{2n-1}$ ist also der Erwartungswert für $\sqrt{2}\chi^2$. $\sqrt{2}\chi^2 - \sqrt{2n-1}$ ist normal um 0 verteilt. Je grösser n ist, desto besser ist die Annäherung an die Normalverteilung.

Als Übergangsstelle, die ungefähr dem Wert $P_n = 0,05$ entspricht, gibt Fisher folgenden Ausdruck an:

$$\sqrt{2}\chi^2 - \sqrt{2n-1} = + 2 \text{ ⁴⁾ }$$

Man kann sich leicht selbst davon überzeugen, dass diese Gleichung, für irgendwelche konkreten n , χ^2 -Werte liefert, denen angenähert $P = 0,05$ entspricht. Wird die linke Seite grösser als 2, so bedeutet das, dass wir der gemachten Ausgleichung misstrauen müssen. Ein gutes Zeichen ist es, wenn sie negativ wird.

Aus der Definition von χ^2 erkennt man, dass es im wesentlichen nichts anderes ist als die Bedingung der kleinsten Quadrate. Neu ist einzig, dass die Fehler nicht absolut gewertet werden, sondern im Verhältnis zur zugehörigen Ordinate. Eine Abweichung von 4 an einer Ordinate von 50 fällt z. B. weniger ins Gewicht als an einer Ordinate von 10. Das ist ohne weiteres vernünftig. Im ersten Falle ist eine Abweichung ≥ 4 wahrscheinlicher als im zweiten.

Die χ^2 -Methode kann deshalb neben ihrer Verwendung als nachträgliche Probe auch benutzt werden, wie die Methode der kleinsten Quadrate, zur Bestimmung des Polynoms. Dadurch wird die Summe der quadratischen Fehler grösser, jedoch sind dafür die Abweichungen proportional zu den Ordinaten verteilt.

Es sei q_x eine durch Beobachtung erhaltene Serie von Sterbenswahrscheinlichkeiten. Die Ausgleichung soll gemacht werden mit dem Polynom

$$a_0 + a_1 P_1(x) + a_2 P_2(x) + \dots + a_m P_m(x)$$

Dann stellt man zur Bestimmung der Koeffizienten die Forderung auf:

$$\chi^2 = \sum_x \frac{\{a_0 + a_1 P_1(x) + \dots + a_m P_m(x) - q_x\}^2}{a_0 + a_1 P_1(x) + \dots + a_m P_m(x)}$$

soll ein Minimum sein. Eine bedeutende Komplikation stellt nun der Nenner dar. H. Cramer ⁶⁾ hat aber gezeigt, dass die Fehler ganz unbedeutend sind, wenn man im Nenner die unausgeglichenen Werte setzt.

Damit lautet die Bestimmungsgleichung:

$$\bar{\chi}^2 = \sum \frac{\{a_0 + a_1 P_1(x) + \dots + a_m P_m(x) - q_x\}^2}{q_x}$$

Die Auflösung ist nun ganz analog wie bei der blossen Methode der kleinsten Quadrate.

Derselbe Autor weist darauf hin, dass es beim Gesetz von Makeham bequemer ist, mit der Intensität μ_x statt mit der Wahrscheinlichkeit q_x zu rechnen. Sind l_x die Personen unter Risiko und d_x die beobachteten Todesfälle, so ergibt sich:

$$\begin{aligned}\chi^2 &= \sum_x \frac{(d_x - l_x \mu_x)^2}{l_x \mu_x} \\ &= \sum_x \frac{\{d_x - l_x (A + Bc^x)\}^2}{l_x (A + Bc^x)} \\ \bar{\chi}^2 &= \sum_x \frac{\{d_x - Al_x - Bc^x l_x\}^2}{d_x}\end{aligned}$$

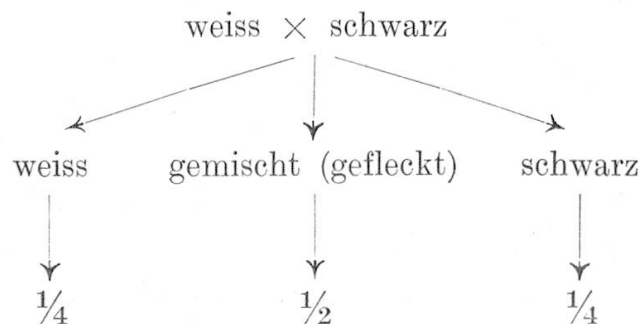
Nun setzt man einige Werte für den Parameter c ein und berechnet A und B . Dann stellt man die resultierenden χ^2 -Werte zu den Abszissen $\log. c$ graphisch dar. Daraus erkennt man, dass χ^2 als Funktion von $\log. c$ ziemlich genau als Parabel 2. oder 3. Grades erscheint. Man bestimmt das Minimum dieser Parabel und liest seine Abszisse ab. Damit ist c bestimmt. Mit diesem Wert von c berechnet man sodann endgültig die Parameter A und B .

Diese stark vereinfachte Methode von H. Cramer gab befriedigende Resultate für χ^2 .

3. Anwendung der χ^2 -Methode.

Zuerst sei ein ganz einfaches Beispiel vorausgeschickt, das auch den Sinn des Willkürlichkeitsgrades zeigt ¹³⁾:

Das Vererbungsgesetz von Mendel lehrt, dass die Hälfte der Nachkommen von zwei Kaninchen z. B., von denen das eine weiss, das andere schwarz ist, von gemischter Farbe sind. Nur ein Viertel ist rein weiss und ein Viertel rein schwarz. Dies sind die Erwartungswerte. Sie spielen die Rolle wie bei unserem Problem die ausgeglichenen Werte.



Bei einer Anzahl solcher Paare sind 158 Junge beobachtet worden. $N = 158$, $n = 3$.

Die Verteilung ergab sich folgendermassen:

	schwarz	gefleckt	weiss
erwartet	39,5	79	39,5
beobachtet . . .	39	78	41
$\chi^2 = \frac{(0,5)^2}{39,5} + \frac{1^2}{79} + \frac{(1,5)^2}{39,5} = 0,0759$			

Die Anzahl der Klassen ist also $n = 3$. Die theoretische Verteilung ist aber nicht frei. Die Summe der 3 Posten muss 158 ergeben. Also könnte man nur zwei von den drei Klassen willkürlich ausfüllen. Die Anzahl in der dritten wäre dadurch festgelegt. Also ist der Willkürlichkeitsgrad nur 2. In der Tabelle von Fisher entspricht den Werten $n = 2$, $\chi^2 = 0,0759$ ungefähr $P = 0,96$. Das heisst, nur in vier Fällen von hundert ist ein so gutes oder noch besseres Resultat zu erwarten. Dieser Versuch spricht also stark für das von Mendel aufgestellte Gesetz.

Nun kehren wir zurück zu unserer Ausgleichungsaufgabe. Bevor wir die χ^2 -Methode anwenden können, müssen wir noch eine Bemerkung machen.

Multipliziert man die Zahl N der Beobachtungen mit irgendeiner Zahl Q , so wird auch χ^2 mit Q multipliziert, wenn man auch die Abweichungen mit demselben Faktor vergrössert. Das ist einleuchtend; denn wenn man die Anzahl N der beobachteten Elemente vervielfacht, so sollte die resultierende Verteilung gleichmässiger ausfallen, da die Zufälligkeiten nach und nach der Gesetzmässigkeit zustreben. Ist das nicht der Fall, so ist es ein Zeichen dafür, dass das vorliegende Material heterogener ist und dass von der theoretischen Verteilung aus die Aussicht für eine gleich oder weniger wahrscheinliche Verteilung wie die beobachtete sinkt.

Wollten wir nun direkt die gegebene Serie der q_x prüfen, so würde uns der zugrunde liegende Massstab fehlen. Es kommt sicher auf dasselbe heraus, ob wir q_x berücksichtigen und dabei gewisse Abweichungen erhalten oder etwa $1000 \cdot q_x$ und dementsprechend auch 1000mal grössere Abweichungen. Beide Ausgleichungen sind genau gleich gut. Trotzdem würde χ^2 bei der zweiten 1000mal grösser. Die χ^2 -Probe ist empfindlich gegenüber einem Massstab, von dem die Wahrscheinlichkeiten q_x selbst gelöst sind. Es ist von grosser Bedeutung, aus wievielen Beobachtungen diese q_x berechnet worden sind.

Also müssen wir, um eine Anwendung machen zu können, zu der Reihe der l_x oder d_x zurückkehren.

Männer 1929—1932.

x	d_x beobachtet	d_x berechnet	x	d_x beobachtet	d_x berechnet
0	5923	5923	12	128	137
1	871	871	13	160	149
2	454	449	14	146	166
3	335	337	15	186	201
4	244	273	16	260	241
5	230	239	17	295	293
6	207	209	18	312	338
7	203	186	19	416	395
8	166	169	20	397	419
9	178	159	21	446	438
10	153	146	22	436	441
11	142	139	23	428	439

Der Willkürlichkeitsgrad ist hier identisch mit der Anzahl Klassen: $n = 24$.

Aus diesen Werten ergibt sich $\chi^2 = 19,3$.

Und nach der Tabelle von Fisher $P_{24} \sim 0,73$.

Das Resultat kann als befriedigend bezeichnet werden, besonders, wenn man noch einen wichtigen Umstand berücksichtigt, auf den wir noch zu sprechen kommen werden.

Frauen 1929—1932.

x	d_x beobachtet	d_x berechnet	x	d_x beobachtet	d_x berechnet
0	4458	4458	13	123	132
1	718	718	14	153	148
2	397	400	15	179	177
3	276	272	16	229	215
4	228	238	17	288	262
5	210	210	18	285	304
6	197	185	19	350	339
7	153	163	20	337	359
8	163	150	21	364	382
9	155	140	22	417	394
10	120	130	23	405	405
11	108	123	24	419	411
12	118	123	25	421	421
			26	429	427

Woraus $\chi^2 = 16,8$
und $P_{27} \sim 0,93$.

Geht man von der berechneten Verteilung aus, so wäre danach in 93 von 100 Fällen mit schlechteren Werten zu rechnen als die, welche tatsächlich beobachtet wurden. Umgekehrt bedeutet es, dass die gemachte Ausgleichung das Verteilungsgesetz der rohen Zahlen sehr gut getroffen hat.

Die Übereinstimmung ist also noch grösser als bei den q_x der Männer.

Nun wäre es auch interessant, ein Kriterium dafür zu besitzen, wann sich die Aufteilung in Intervalle, wie wir es für die Messungen von 1929—1932 versucht haben, lohnt und wann nicht. Es ist allerdings sehr schwer, ein solches, absolut gültiges Gesetz aufzustellen, und man wird die Hilfe der rein anschaulichen Überlegung nie ausschalten können. Jedoch ist es naheliegend, dass man versuchsweise die Wahrscheinlichkeiten P für die einzelnen Teilintervalle berechnet und daraus über das ganze Intervall das gewogene arithmetische Mittel bildet.

Für die q_x bei den Männern gibt das:

$$P_{24}^* \sim 0,73 \quad \text{also gleich wie } P_{24}.$$

Bei den Frauen dagegen

$$P_{24}^* \sim 0,85 \quad \text{gegenüber } P_{27} \sim 0,93.$$

Stellt man auf diese Probe ab, so hätte also die Aufteilung in keinem der beiden Fälle eine lohnende Verbesserung gebracht. Dies kommt daher, dass in einzelnen Intervallen die Summe der absoluten Fehler kaum verkleinert werden konnte. In diesen wird dann P sofort bedeutend kleiner, und ein solches Intervall vermag das ganze Resultat bedeutend zu beeinträchtigen. In andern, besonders typischen Teilstücken dagegen ist der Vorteil unzweifelhaft.

* * *

Bei den Messungen von 1921—1930 erstreckte sich die Beobachtungsdauer über 6 Jahre mehr als bei denen von 1929—1932. Will man hier die χ^2 -Methode anwenden, so offenbart sich eine äusserst wichtige Tatsache. Man erhält nämlich für die Alter von 0 bis 45 Jahren einen Wert von χ^2 bei 80 herum. Extrapoliert man die

Kolonnen der Tabelle, so sieht man, dass das entsprechende P kleiner ist als 0,05, wonach die Ausgleichung vollständig ungenügend wäre. Auch mit der von Fisher angegebenen Ersatzmethode kommt man zum selben Schluss, denn $\sqrt{2\chi^2} - \sqrt{2n-1}$ nimmt einen Wert um $+3$ herum an.

Betrachtet man aber die graphische Darstellung, so darf man ruhig behaupten, dass die Ausgleichung auf keinen Fall so schlecht ist, sondern im Gegenteil recht befriedigend.

Der Fehler liegt wirklich nicht in der Ausgleichung, sondern in der Annahme, dass man die χ^2 -Methode anwenden könne.

Die χ^2 -Methode geht von der fundamentalen Voraussetzung aus, dass in jeder Gruppe die Häufigkeit bei oftmaliger Wiederholung der Messungen nach dem Gauss'schen Gesetz, das heisst, in Glockenform, um einen wahrscheinlichsten Mittelwert herum verteilt sei. Diese Voraussetzung trifft nun aber für die q_x nicht zu. Es ist eine nachgewiesene Tatsache, dass in letzter Zeit die Sterblichkeit für jedes einzelne Alter im Laufe der Kalenderjahre abnimmt. So stellt das q_x für jedes bestimmte Alter eine mit der Zeit irgendwie abnehmende Funktion dar.

Einen wahrscheinlichsten Mittelwert im Sinne der Zufallsverteilung gibt es da nicht. Diskontinuierlich aufgefasst hat eigentlich jedes Kalenderjahr seinen eigenen Mittelwert, so dass bei einer langen Beobachtungsdauer an Stelle eines einzigen wahrscheinlichsten Wertes eine ganze Anzahl auftritt. Statt des Verlaufes (1) nimmt also die Kurve schematisch den Verlauf (2) an.

Die Stärke der Abflachung ist proportional zu der Länge der Beobachtungszeitspanne. Das Fehlerintegral ist somit hier nicht brauchbar und ebensowenig die χ^2 -Methode. Die Fläche unter der Kurve ausserhalb einer gewissen Grenze $\pm t$ ist grösser als bei der reinen Zufallsverteilung. Damit ist auch das Verhältnis gebildet aus der Fläche ausserhalb dieser Grenzen, dividiert durch die Gesamtfläche grösser als normal. Wendet man nun doch die χ^2 -Methode an, so benutzt man die den Tatsachen nicht entsprechende Verteilung (1). Damit wird ein q_x , das vielleicht noch maximale Wahrscheinlichkeit besitzt, als zu weit vom Mittel entfernt und daher als unwahrscheinlich taxiert.

Verlängert man eine Beobachtungsdauer von u auf $k \cdot u$ Jahre, so kann daher χ^2 niemals k mal kleiner werden. Die Tendenz, durch

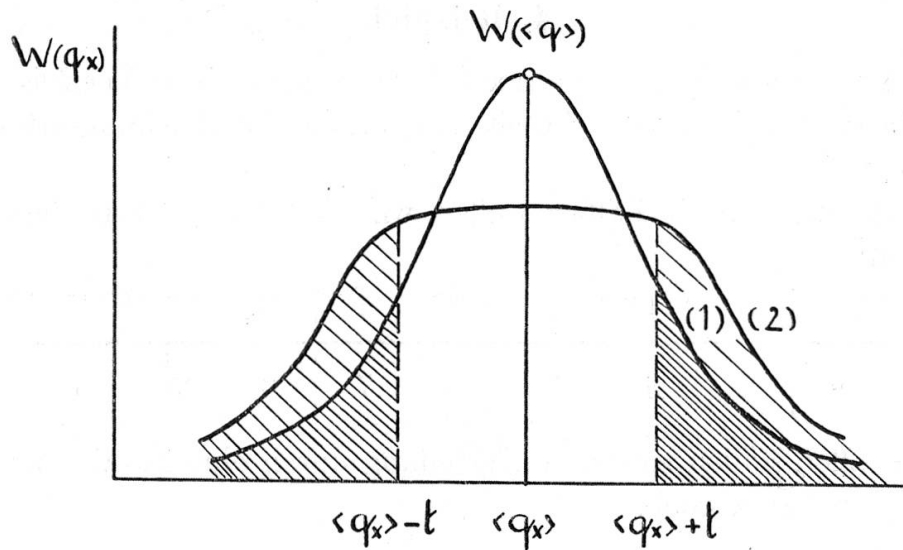


Fig. 8.

die Vergrößerung der Anzahl der beobachteten Elemente einer gesetzmässigeren Verteilung zuzustreben, ist vorhanden, jedoch wirkt ihr eine zweite Kraft entgegen, eben dieses Abnahmegesetz der q_x , die das Material wieder heterogener macht.

Wie man aus einem Vergleich der beiden Messungen sieht, hat bei diesem Entgegenwirken der zwei genannten Kräfte doch die ausgleichende die Oberhand behalten, denn das Material von 1921—1930 ist tatsächlich etwas homogener als dasjenige von 1929—1932, jedoch nicht in dem von der χ^2 -Methode geforderten Masse.

Die Voraussetzung der χ^2 -Methode ist ein rein abstraktes Postulat und scheint auch in vielen andern Fällen nicht restlos erfüllt zu sein, wenn auch nicht in dem Masse wie bei den q_x .

Die erwähnte Verfälschung der Tatsachen wirkte natürlich auch schon bei der kürzeren Beobachtungsperiode 1929—1932. Jedoch ist dort die Verflachung des Maximums noch nicht so bedeutend, so dass die Gauss'sche Verteilung noch als Näherung benutzt werden kann. Jedenfalls wäre aber auch hier, wenn es auf den genauen Wert von P ankäme, das Ergebnis ein wenig besser zu bewerten, als das berechnete P angab.

Am besten wäre die χ^2 -Methode anwendbar, wenn sich die Beobachtungen auf ein einziges Jahr beschränkten, was aber eine grössere Bevölkerung voraussetzt, als dies in der Schweiz der Fall ist.

4. Beispiel.

Die χ^2 -Methode soll uns noch dazu dienen, einen Vergleich der verschiedenen, in dieser Arbeit erwähnten Ausgleichungsarten zu machen.

Der Einfachheit halber wollen wir eine fiktive Verteilung annehmen:

x	0	1	2	3	4	5	6	7
y	2	4	5	7	7	6	5	1

In allen Fällen soll diese Verteilung durch eine Parabel zweiten Grades ersetzt werden.

a) Empirische Methode.

Wir zeichnen die Parabel möglichst gut von Hand ein und lesen ab, dass sie durch die Punkte (0,1), (3,7) und (7,1) gehen soll. Durch diese drei Punkte ist folgende Parabel bestimmt:

$$f_1(x) = 1 + 3,5x - 0,5x^2$$

x	0	1	2	3	4	5	6	7
$f_1(x)$	1	4	6	7	7	6	4	1

$$\chi_1^2 = 1,4167$$

$$\Sigma (f_1(x) - y)^2 = R_1^2 = 3$$

b) Methode der kleinsten Quadrate.

Es sei

$$P_i(x) = x^i$$

$$f_2(x) = a_0 + a_1 x + a_2 x^2$$

$$\sigma_{00} = 8$$

$$\sigma_{01} = 28$$

$$\sigma_{02} = 140$$

$$\varrho_0 = 37$$

$$\sigma_{10} = 28$$

$$\sigma_{11} = 140$$

$$\sigma_{12} = 784$$

$$\varrho_1 = 130$$

$$\sigma_{20} = 140$$

$$\sigma_{21} = 784$$

$$\sigma_{22} = 4676$$

$$\varrho_2 = 578$$

Man hat also das System zu lösen:

$$8a_0 + 28a_1 + 140a_2 = 37$$

$$28a_0 + 140a_1 + 784a_2 = 130$$

$$140a_0 + 784a_1 + 4676a_2 = 578$$

Daraus erhält man: $a_0 = 1,5418(5)$

$$a_1 = 3,0534$$

$$a_2 = -0,4345$$

$$f_2(x) = 1,5418 + 3,0534x - 0,4345x^2$$

x	0	1	2	3	4	5	6	7
$f_2(x)$	1,542	4,161	5,911	6,792	6,803	5,946	4,220	1,625

$$\chi_2^2 = 0,6799$$

$$R_2^2 = 2,1496$$

c) Methode von Cauchy.

$$f_3(x) = a_1 x + a_2 x^2$$

Um nicht noch eine Konstante a_0 berechnen zu müssen, wollen wir das Intervall 0—7 verschieben nach 1—8, was wir nachträglich durch die Substitution $x/x+1$ wieder rückgängig machen können. Nun setzen wir die 8 Punkte ein:

$$\begin{array}{rcl}
 x = 1 & . & 2 = a_1 + a_2 \\
 x = 2 & . & 4 = 2a_1 + 4a_2 \\
 x = 3 & . & 5 = 3a_1 + 9a_2 \\
 x = 4 & . & 7 = 4a_1 + 16a_2 \\
 x = 5 & . & 7 = 5a_1 + 25a_2 \\
 x = 6 & . & 6 = 6a_1 + 36a_2 \\
 x = 7 & . & 5 = 7a_1 + 49a_2 \\
 x = 8 & . & 1 = 8a_1 + 64a_2 \\
 \hline
 & & 37 = 36a_1 + 204a_2
 \end{array}$$

Die Koeffizienten von a_1 sind alle positiv. Also konnten wir direkt addieren.

$$a_1 = \frac{37}{36} - \frac{204}{36} a_2$$

$$f_3(x) = \left(\frac{37}{36} - \frac{204}{36} a_2 \right) \cdot x + a_2 x^2$$

$$\underbrace{f_3(x) - \frac{37}{36}x}_{\Delta f_3(x)} = a_2 \underbrace{\left(x^2 - \frac{204}{36}x\right)}_{\Delta P_2(x)} \quad \text{wo } P_2 \equiv x^2$$

Das gibt wieder 8 Gleichungen:

$x = 1 \dots 2 - \frac{37}{36} = a_2 \left(1 - \frac{204}{36}\right)$	$0,9722 = -4,6667 a_2$
$x = 2 \dots 4 - \frac{37}{18} = a_2 \left(4 - \frac{204}{18}\right)$	$1,9444 = -7,3333 a_2$
$x = 3 \dots 5 - \frac{37}{12} = a_2 \left(9 - \frac{204}{12}\right)$	$1,9167 = -8,0000 a_2$
$x = 4 \dots 7 - \frac{37}{9} = a_2 \left(16 - \frac{204}{9}\right)$	$2,8889 = -6,6667 a_2$
$x = 5 \dots 7 - \frac{185}{36} = a_2 \left(25 - \frac{1020}{36}\right)$	$1,8611 = -3,3333 a_2$
$x = 6 \dots 6 - \frac{37}{6} = a_2 \left(36 - \frac{204}{6}\right)$	$-0,1667 = +2,0000 a_2$
$x = 7 \dots 5 - \frac{259}{36} = a_2 \left(49 - \frac{1428}{36}\right)$	$-2,1944 = +9,3333 a_2$
$x = 8 \dots 1 - \frac{296}{36} = a_2 \left(64 - \frac{1632}{36}\right)$	$-7,2222 = +18,6667 a_2$

Vor der Addition müssen wir hier diejenigen Gleichungen, die einen negativen Koeffizienten von a_2 haben, mit -1 erweitern. Die Summe Σ'' ergibt dann:

$$-19,1666 = 60,0000 a_2$$

$$a_2 = -0,3194$$

$$a_1 = 2,8377$$

$$f_3(x) = 2,8377x - 0,3194x^2 \quad x/x_{+1}$$

$$f_3^*(x) = 2,5183 + 2,1989x - 0,3194x^2$$

x	0	1	2	3	4	5	6	7
$f_3^*(x)$	2,518	4,398	5,638	6,240	6,204	5,528	4,213	2,260

$$\chi_3^2 = 1,2991$$

$$R_3^2 = 4,4748$$

Es wäre mindestens eine dritte oder vierte Näherung nötig, um ein brauchbares Resultat zu erhalten. Dafür hat die Methode von Cauchy den Vorteil, dass sie nicht auf die Äquidistanz der Beobachtungspunkte angewiesen ist, denn sie stellt ja nur auf den Schwerpunkt der Verteilung ab.

d) Methode der kleinsten Quadrate mit Orthogonalpolynomen.

$$n = 8$$

$$S_0^1 = 37$$

$$S_1^2 = 130$$

$$S_2^3 = 224$$

$$\Delta^k T_i(0):$$

$k \backslash i$	0	1	2
0	1	— 7	21
1	0	2	— 18
2	0	0	6
$\Sigma T_i^2(x)$	8	168	1512

$$Z_0 = 37$$

$$a_0 = 4,625$$

$$f_4(0) = 1,5422$$

$$Z_1 = 1$$

$$a_1 = 0,006$$

$$\Delta f_4(0) = 2,6184$$

$$Z_2 = -219$$

$$a_2 = -0,1448$$

$$\Delta^2 f_4(0) = -0,8688$$

$$f_4(x) = 1,5422 + 2,6184 x - 0,8688 \binom{x}{2}$$

x	0	1	2	3	4	5	6	7
$f_4(x)$	1,542	4,160	5,910	6,791	6,803	5,946	4,221	1,626

$$\chi_4^2 = 0,6793$$

$$R_4^2 = 2,1473$$

e) χ^2 -Minimum-Methode.

Es kommt natürlich nur die von H. Cramer angegebene Vereinfachung in Betracht, wonach im Nenner an Stelle der theoretischen Werte die beobachteten stehen. Damit lässt sich aber die Anwendung der orthogonalen Polynome nicht mehr ohne weiteres verbinden, denn man würde bei der Auflösung für den Koeffizienten a_i die Gleichung erhalten:

$$a_i \sum_x \frac{T_i^2(x)}{y(x)} = \sum_x \frac{T_i(x) \cdot y(x)}{y(x)} = \sum T_i(x) = 0$$

Somit würden alle Koeffizienten ausser a_0 verschwinden. Deshalb setzen wir:

$$f_5(x) = a_0 + a_1 x + a_2 x^2$$

Dann soll

$$\bar{\chi}^2 = \frac{(a_0 + a_1 x + a_2 x^2 - y(x))^2}{y(x)}$$

ein Minimum sein.

Analog wie bei der Methode der kleinsten Fehlerquadrate setzen wir bei der Auflösung

$$\sigma'_{ik} = \sum_0^7 \frac{x^i x^k}{y} \quad \text{und} \quad \varrho'_i = \sum_0^7 x^i \frac{y}{y} = \sum x^i$$

So erhalten wir das System:

$\sigma'_{00} = 2,6024$	$\sigma'_{01} = 10,6833$	$\sigma'_{02} = 64,9880$	$\varrho'_0 = 8$
$\sigma'_{10} = 10,6833$	$\sigma'_{11} = 64,9880$	$\sigma'_{12} = 421,8833$	$\varrho'_1 = 28$
$\sigma'_{20} = 64,9880$	$\sigma'_{21} = 421,8833$	$\sigma'_{22} = 2815,9595$	$\varrho'_2 = 140$

Auflösen sind also folgende drei Gleichungen:

$$\begin{aligned} 2,6024 a_0 + 10,6833 a_1 + 64,9880 a_2 &= 8 \\ 10,6833 a_0 + 64,9880 a_1 + 421,8833 a_2 &= 28 \\ 64,9880 a_0 + 421,8833 a_1 + 2815,9595 a_2 &= 140 \end{aligned}$$

Daraus ergibt sich:

$$\begin{aligned} a_0 &= 1,7285 \\ a_1 &= 3,0238 \\ a_2 &= -0,4432 \end{aligned}$$

$$f_5(x) = 1,7285 + 3,0238x - 0,4432x^2$$

x	0	1	2	3	4	5	6	7
$f_5(x)$	1,729	4,309	6,003	6,811	6,733	5,768	3,916	1,178

$$\chi_5^2 = 0,5841$$

$$R_5^2 = 2,5425$$

* *

Anhand dieser fünf Resultate kann man nun leicht die verschiedenen Methoden, wenigstens oberflächlich, beurteilen. In erster Linie sieht man, dass die χ^2 - und die R^2 -Proben wie erwartet ganz verschiedene Resultate geben. Ferner ist an diesem einfachen Beispiel der Vorteil der orthogonalen Polynome nicht deutlich genug zum Ausdruck gekommen, abgesehen von der grossen Vereinfachung der Rechnung.

Die χ^2 -Methode geht dem Problem tiefer auf den Grund als sämtliche andern, indem sie ganz bestimmte Voraussetzungen über die Struktur der Verteilung macht. Jedoch sind diese Voraussetzungen nicht in allen Fällen erfüllt, so dass die Anwendbarkeit der χ^2 -Methode beschränkt ist. Wir haben schon den Fall angetroffen, wo das χ^2 -Gesetz als Probe für die Güte einer Ausgleichung nicht zulässig ist, und nun sei auch noch darauf hingewiesen, dass es sich ebenfalls nicht immer zur Ausgleichung als χ^2 -Minimum-Methode eignet:

Bei vielen Häufigkeitsverteilungen sind die äussersten Werte, da sie nur selten auftreten, ziemlich regellos verteilt, während die häufigeren, um den Erwartungswert herum, das Gesetz schon deutlich zeigen. Nun verlangt aber die χ^2 -Methode eine maximale Anschmiegung an die kleinen Werte auf Kosten der grossen. Auf diese Weise würde man wegen den äussersten, unzuverlässigen Mengen systematische Fehler bei den regelmässigeren in der Nähe des Erwartungswertes heraufbeschwören. Eine Anwendung ist nur dann am Platze, wenn die Zahl N der Beobachtungen so gross ist, dass auch die seltensten Werte von zufälligen Schwankungen so weit als möglich befreit sind.

Die χ^2 -Probe kann auch über die Homogenität oder Heterogenität eines Materials entscheiden. Wenn man trotz bestmöglicher Ausgleichung unbefriedigende Werte für die Wahrscheinlichkeit P

erhält, so darf man schliessen, dass das gegebene Material zu heterogen sei und dass das Verteilungsgesetz zu wenig durchblicke.

Bei der Beurteilung eines P -Wertes muss man die Homogenität des beobachteten Materials mitberücksichtigen. Z. B. kann $P = 0,5$ bei einer sehr regelmässigen Verteilung ein schlechtes Zeichen für eine Ausgleichung sein, während es diese bei einem heterogenen Material gutheissen kann.

IV. TEIL.

Zusammenfassung der Ergebnisse.

1. Das Prinzip der *Methode von Cauchy* ist sehr elegant, jedoch ist die rechnerische Durchführung etwas umständlich. Auch die Genauigkeit des Resultates wird von einzelnen anderen Methoden überboten.

2. Die *Methode der kleinsten Quadrate*, die von einem ganz anderen Gesichtspunkt ausgeht, ist der Grundstein für die analytische Ausgleichung. Sie stellt nicht ein abgeschlossenes Ganzes dar, sondern vielmehr einen Ansatzpunkt, der auf anderem Wege noch vervollständigt werden kann.

3. Ein solcher Weg der Vervollständigung ist die Einführung der *orthogonalen Polynome*. Sie schalten die Zufälligkeit, die in der Aufstellung der Ausgleichungsfunktion besteht, aus und bringen eine wesentliche Vereinfachung der Rechnung mit sich, so dass der Arbeitsumfang nicht grösser wird als bei den häufig verwendeten elementaren Methoden. Bei der Anwendung auf die schweizerischen Sterbetafeln hat diese Methode in der schwer auszugleichenden Periode von 0 bis 25 Jahren sehr befriedigende Resultate gezeigt.

4. Die wichtige Frage, wie die Güte einer Ausgleichung zu beurteilen sei, erlaubt schwerlich eine allgemeingültige Antwort. Die χ^2 -Methode ist etwas vom Interessantesten, was in dieser Hinsicht geschaffen wurde, doch ist sie ein hochempfindliches Kriterium. Die Anwendung auf die schweizerischen Volkssterbetafeln hat gezeigt, dass in jedem Fall zuerst die Voraussetzungen geprüft werden müssen, da man sonst leicht zu irreführenden Folgerungen gelangen kann.

Literatur.

- 1) *R. Risser*: Applications de la Statistique à la Démographie et à la Biologie, Paris, 1932.
 - 2) *G. J. Lidstone*: Notes on Orthogonal Polynomials and their Application to the Least-Square Method. J. I. A. Vol. LXIV, N° 309. London, 1933.
 - 3) *J. P. Gram*: Über partielle Ausgleichung mittelst Orthogonalfunktionen. M. V. M. Heft 10, 1915.
 - 4) *R. A. Fisher*: Statistical Methods for Research Workers. London, 1932.
 - 5) *W. P. Elderton*: Frequency Curves and Correlation. London, 1927.
 - 6) *H. Cramer* and *H. Wold*: Mortality Variations in Sweden. A Study in Graduation and Forecasting. Uppsala, 1935.
 - 7) *O. N. Anderson*: Einführung in die mathematische Statistik. Wien, 1935.
 - 8) *Frechet, Perrenoud et Mahrer*: Sur l'ajustement des Tables de Mortalité par la Méthode de Tchebychef. Congrès des Sociétés savantes; France, 1924.
 - 9) Schweizerische Volkssterbetafeln 1876—1932. Eidgenössisches Statistisches Amt, Beiträge zur schweizerischen Statistik, Heft 4. Bern, 1935.
 - 10) *K. Pearson*: On the criterion, that a given system of deviations from the probable in the case of a correlated system of variables is such, that it can be reasonably supposed to have arisen from random sampling. Phil. Mag. July, 1900.
 - 11) *H. E. Soper*: Frequency Arrays. Cambridge Univ. Press.
 - 12) *W. P. Elderton*: Tables for Testing the Goodness of Fit of Theory to Observation. Biometrika I, S. 155—163. 1901—1902.
 - 13) *G. Darmoy*: Vorlesung über mathematische Statistik an der Universität Paris. S. S. 1935. Nach Vorlesungsnotizen des Verfassers.
 - 14) *W. Schoeb*: Ein Beitrag zur Theorie der Ausgleichsrechnung. Diss. Bern, 1935.
 - 15) Das nötige statistische Material wurde in verdankenswerter Weise vom Eidgenössischen Statistischen Amt zur Verfügung gestellt.
-

