Untersuchungen über den Wahrscheinlichkeitscharakter der Sterblichkeit

Autor(en): **Niedermann, Hans**

Objekttyp: **Article**

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

Band (Jahr): **46 (1946)**

PDF erstellt am: **28.05.2024**

Persistenter Link: https://doi.org/10.5169/seals-966877

Nutzungsbedingungen

Die ETH-Bibliothek ist Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Inhalten der Zeitschriften. Die Rechte liegen in der Regel bei den Herausgebern.
Die auf der Plattform e-periodica veröffentlichten Dokumente stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Einzelne Dateien oder Ausdrucke aus diesem Angebot können zusammen mit diesen Nutzungsbedingungen und den korrekten Herkunftsbezeichnungen weitergegeben werden.
Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die systematische Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber.

Haftungsausschluss

Alle Angaben erfolgen ohne Gewähr für Vollständigkeit oder Richtigkeit. Es wird keine Haftung übernommen für Schäden durch die Verwendung von Informationen aus diesem Online-Angebot oder durch das Fehlen von Informationen. Dies gilt auch für Inhalte Dritter, die über dieses Angebot zugänglich sind.

Ein Dienst der *ETH-Bibliothek*
ETH Zürich, Rämistrasse 101, 8092 Zürich, Schweiz, www.library.ethz.ch

http://www.e-periodica.ch

Untersuchungen über den Wahrscheinlichkeitscharakter der Sterblichkeit

Von *Hans Niedermann*, Genf

I. Wahrscheinlichkeitstheorie und Versicherungsmathematik

Es ist eine alte Streitfrage, ob die einjährigen Sterbenswahrscheinlichkeiten, wie sie in den Sterbetafeln zusammengestellt werden, als Wahrscheinlichkeiten im Sinne der Wahrscheinlichkeitstheorie betrachtet und behandelt werden dürfen, eine Streitfrage, welche gerade in den letzten Jahren im Rahmen der Diskussion um die Grundlagen der Versicherungsmathematik wieder in den Vordergrund getreten ist. Es ist ein Spezialfall jenes allgemeinern Problems, welches nach den Voraussetzungen fragt, unter denen statistische Häufigkeitszahlen als Wahrscheinlichkeiten aufgefasst werden können.

Um den Ausgangspunkt der nachstehenden Überlegungen festzulegen und ihren Sinn klarzustellen, ist es bei der immer noch herrschenden Vielfalt der Meinungen nicht unnötig, auf die Frage der Anwendbarkeit der Wahrscheinlichkeitsrechnung im Versicherungswesen im grundsätzlichen Sinne etwas einzugehen.

1. Die Stellung der Wahrscheinlichkeitstheorie im allgemeinen

Die Wahrscheinlichkeitstheorie muss, wenn sie überhaupt als mathematische Theorie gelten will, die strengen Anforderungen erfüllen, welche die Mathematik stellt. Sie muss deshalb axiomatisch begründet werden. Ihre Begriffe dürfen nicht mit der Unschärfe der Wörter des täglichen Sprachgebrauchs behaftet sein, sondern ihre Eigenschaften müssen sich in strenger und eindeutiger Weise aus den Axiomen ergeben. Es werden also vollkommen abstrakte Begriffe sein, ebenso wie die Begriffe Punkt, Gerade, Ebene in der Geometrie oder Massenpunkt, Wellenpaket, Energie in der theoretischen Physik.

Diese Begriffe sind reine Produkte unseres Geistes. Es wird in der Aussenwelt, also der Welt unserer sinnlichen Erfahrung, nichts geben, das mit ihnen vollkommen isomorph wäre, d. h. dieselben Eigenschaften und Beziehungen besässe wie sie selbst. Aber wir dürfen hoffen, dass es in der Natur Dinge gibt, welche diese Eigenschaften und Beziehungen wenigstens mit einer gewissen Annäherung aufweisen; denn unser Geist hat die Begriffe nicht aus sich allein heraus geschaffen, sondern er ist dazu durch Eindrücke, die er aus der Aussenwelt aufgenommen hat, angeregt worden.

«Unser Verstand hat die Fähigkeit, Symbole zu schaffen, ...aber er macht von dieser Fähigkeit nur Gebrauch, wenn die Erfahrung ihm dazu Veranlassung gibt» (H. Poincaré in «Wissenschaft und Hypothese»).

Wir versuchen in unseren Gedankenkonstruktionen, das nachzubilden, was wir aufgenommen haben. Unser Ziel ist es, diese Nachbildung so genau zu machen als nur möglich, so genau, dass wir mit ihrer Hilfe Voraussagen über unsere zukünftigen Erfahrungen machen können. Aber niemals wird sie vollkommen sein. Es ist deshalb wesentlich, dass man sich über den Grad der Approximation immer Rechenschaft ablegt, sonst werden Fehlurteile und Fehlschlüsse unvermeidlich sein. Dass aber in solchen Fällen nicht die Theorie falsch sein kann — vorausgesetzt, dass sie logisch einwandfrei abgeleitet wurde —, sondern der Fehler immer nur in der unvollkommenen Übereinstimmung zwischen Theorie und Aussenwelt zu suchen ist, betont auch de Finetti in seinem Kongressbeitrag «Il calcolo delle probabilità nel dominio dell' assicurazione»:

«Secondo la concezione più ampia e più chiara della teoria delle probabilità, le relazioni fra le probabilità di diversi eventi, ossia i principii del calcolo di probabilità, hanno un carattere puramente logico, e quindi una validità assolutamente generale. Non si può discutere pertanto la legittimità delle loro applicazioni a un qualunque problema, nè chiedere a risultati sperimentali una conferma o smentita della teoria; soltanto le particolari valutazioni delle probabilità di determinati eventi (e quindi delle proprietà che ne discendono, come ad es. quella dell' indipendenza, dell' equiprobabilità, ecc. di diversi eventi) sono soggette all' opinione e alla discussione» [1]).

[1]) Berichte des XII. Internationalen Kongresses der Versicherungsmathematiker, Band I (Luzern 1940). Im folgenden zitiert als «Kongressband I».

Welches nun die grundlegenden Axiome der Wahrscheinlichkeitstheorie sein sollen, darüber ist man sich noch nicht einig geworden. Im grossen ganzen stehen sich zwei Auffassungen gegenüber: die klassische und die Misessche. Ich brauche mich auf deren Erörterung hier nicht einzulassen. Wichtig für uns ist nur, dass beide im wesentlichen zu denselben Resultaten führen. So konnte Anderson in der Einleitung zu seinem Buch «Einführung in die mathematische Statistik» feststellen: «Welche Definition des Wahrscheinlichkeitsbegriffes man auch wählt, der bei weitem grössere Teil des rein mathematischen Inhaltes der Wahrscheinlichkeitsrechnung — von einigen ziemlich unwesentlichen Ausnahmen abgesehen — bleibt hiervon beinahe ganz unberührt.»

Auch eine spätere Verbesserung der axiomatischen Grundlage wird diese Hauptresultate nicht tangieren; denn es wird sich ja nicht darum handeln, die Wahrscheinlichkeitstheorie neu zu schaffen, sondern die bereits vorhandene Theorie streng zu begründen. An diese allgemein anerkannten Sätze werden wir uns also zu halten haben, wenn wir Theorie und Erfahrung vergleichen wollen. Damit werden direkte Verifikationen problematischer Voraussetzungen umgangen, wie der Gleichmöglichkeit in der klassischen oder der Existenz von Grenzwerten in unendlichen, regellosen Reihen (die dann in praxi doch immer nur endlich sind) in der Misesschen Definition.

2. Die Wahrscheinlichkeitstheorie in der Versicherungsmathematik

«Eine Anhäufung von Tatsachen ist so wenig eine Wissenschaft, als ein Steinhaufen ein Haus ist.»
(H. Poincaré: «Wissenschaft und Hypothese».)

Über dieses Problem ist in den letzten Jahren ein ziemlich heftiger Streit entbrannt, wenn es auch früher schon gelegentlich zur Diskussion gestellt wurde. Um eine Abklärung herbeizuführen, setzte das Organisationskomitee des XII. Internationalen Kongresses der Versicherungsmathematiker diese Frage als Verhandlungsgegenstand I auf die Traktandenliste. Es wurden denn auch eine Reihe von Abhandlungen eingereicht, die auf eine angeregte Diskussion hoffen liessen. Leider konnte dann die abklärende Aussprache nicht stattfinden, weil der Kongress infolge des inzwischen ausgebrochenen zweiten Weltkrieges abgesagt werden musste. Vielleicht hätten sich

bei der persönlichen Kontaktnahme die verschiedenen extremen Standpunkte, wie sie in den Kongreßschriften zur Darstellung kommen, etwas genähert. Gewissermassen als Ersatz für die ausgefallene Diskussion ist dann von verschiedenen Autoren versucht worden, aus den Kongressabhandlungen die Quintessenz zu ziehen und jenen Standpunkt herauszuschälen, der für den unvoreingenommenen Beurteiler nach sorgfältiger Prüfung aller Argumente und Einwände als der richtige erscheint [1]). Dies ist auch der Zweck der nachfolgenden kurzen Darstellung.

Ich möchte bei den Gegnern der Anwendung der Wahrscheinlichkeitsrechnung im Versicherungswesen zwei Gruppen unterscheiden: die absoluten Empiristen und die bedingten Empiristen. Wir wollen uns mit ihnen kurz auseinandersetzen.

a) Der absolute Empirismus

Hieher gehören diejenigen, welche die Wahrscheinlichkeitsrechnung in der Versicherungsmathematik vollständig ablehnen. Sie verweisen meist darauf, dass die Formeln für Prämien und Deckungskapitalien ohne Verwendung des Wahrscheinlichkeitsbegriffes einfach unter Zugrundlegung einer fiktiven Gesamtheit, welche einer gegebenen Absterbeordnung gehorcht, berechnet werden können. Das ist formal betrachtet wohl richtig; aber es wird dabei übersehen, dass die Berechtigung dieses Verfahrens noch nicht sichergestellt ist. Der Hinweis auf die Erfahrung, durch welche die Methode gerechtfertigt werde, stellt nicht nur keine ausreichende Begründung dar, sondern macht die Unzulänglichkeit der Grundlage erst recht deutlich, kann doch nicht bestritten werden, dass die wirklichen Versichertenbestände niemals genau die fiktive Absterbeordnung befolgen; vielmehr treten kleinere oder grössere Abweichungen auf. Diese Abweichungen werden von der Methode der fiktiven Gesamtheiten vollständig ignoriert. Sie setzt einfach stillschweigend voraus, dass sich die Schwankungen nach oben und unten ausgleichen. Aber damit ist das Problem nicht gelöst, sondern nur beiseite geschoben. Der Wissenschafter, aber auch der verantwortungsbewusste Praktiker, will doch wissen, worauf diese Abweichungen beruhen, er möchte ihre mutmassliche Grösse abschätzen können, und er will darüber Aufschluss haben, unter welchen Voraus-

[1]) H. Jecklin (7); H. Wyss: MVSV Bd. 41, S. 212; H. Wiesler (9).

setzungen er erwarten kann, dass sich die positiven und negativen Abweichungen ausgleichen werden. Mit andern Worten: er braucht eine Theorie dieser Abweichungen. Ich betone hier ausdrücklich das Wort «Theorie»; denn mit einer blossen Beschreibung der bisherigen Erfahrungen ist es nicht getan. Die Wissenschaft will doch Voraussagen machen. Die nackten Erfahrungstatsachen lassen aber für sich allein betrachtet niemals Schlüsse auf zukünftige Ereignisse zu; dazu bedarf es einer Theorie, welche sich auf diese Erfahrungen stützt und sie interpretiert. Max Planck sagt mit Recht von den reinen Empiristen:

«Ein Forscher, der sich bei seinen Arbeiten nicht von irgendeiner Hypothese leiten lässt, sei sie auch so vorsichtig und provisorisch als nur möglich, verzichtet von vornherein auf ein tieferes Verständnis seiner eigenen Resultate.» (Max Planck: «Wege zur physikalischen Erkenntnis.»)

Und auch Ivo Lah, der in seiner Kongressabhandlung «Wahrscheinlichkeitsrechnung und Versicherungswesen» eine empiristische Auffassung vertritt (wenn er auch eher der zweiten Gruppe der bedingten Empiristen zuzuordnen ist), muss zugeben, dass diese Methode an einem Mangel an wissenschaftlicher Grundlage krankt [1]).

b) Der bedingte Empirismus

Neben den absoluten Empiristen, welche die Notwendigkeit einer Theorie überhaupt bestreiten, ist noch eine zweite Gruppe von Bedeutung, die ich «bedingte Empiristen» nennen möchte. Sie geben zwar die Wünschbarkeit einer theoretischen Interpretation der Erfahrungstatsachen zu, bestreiten aber, dass die Wahrscheinlichkeitsrechnung (oder eine andere Theorie) dazu in der Lage sei. Insbesondere verneinen sie die Möglichkeit von irgendwelchen ernsthaften Prognosen. «Und», so fügen sie hinzu, «lieber keine Theorie als eine schlechte!» Dass ein Praktiker, durch böse Erfahrungen enttäuscht, zu dieser Einstellung gelangt, kann man verstehen; aber dass nun auch Wissenschafter sich diesen unbefriedigenden, pessimistischen Standpunkt zu eigen machen und das Kind mit dem Bade ausschütten, d. h. auf jede Theorie verzichten wollen, weil ein erster oder auch zweiter und dritter Versuch misslang, ist nicht verzeihlich. Sie sollten im Gegenteil um so ent-

---

[1]) Kongressband I, S. 421.

schlossener daran gehen, die Theorie so auszubauen, dass sie besser wird als «keine Theorie». Nur dann ist ein Fortschritt möglich. So sagt auch P. Nolfi [1]):

«Es wäre verfehlt, die Risikotheorie mit ihren wertvollen Errungenschaften deswegen zu verwerfen, weil sie nicht in allen Teilen zu befriedigen vermag. Es ist Aufgabe der Wissenschaft, zu zeigen, in welcher Weise eine Theorie, die den wirklichen Verhältnissen noch nicht genügend nahe kommt, zweckmässig ausgebaut bzw. umgebaut werden kann.»

Die Vorwürfe richten sich natürlich vor allem gegen die Wahrscheinlichkeitstheorie, über der erbarmungslos der Stab gebrochen wird:

«Damit sind wir zu einer Anschauung gelangt, die in der Statistik schon allgemein bekannt ist, nämlich, dass die Wahrscheinlichkeitsrechnung für statistische Forschungsverfahren unbrauchbar sei» [2]).

«Die Wahrscheinlichkeitstheorie kann uns niemals einen Rechenprozess liefern, mittels dessen aus den Eigenschaften der vergangenen Erscheinungen auf solche der zukünftigen Erscheinungen geschlossen werden darf, da man die zukünftigen Begleitumstände nicht kennen kann» [3]).

Simon Shannon [4]) hält eine Risikotheorie in der Lebensversicherung nicht für genügend begründet, weil die zufälligen Sterblichkeitsschwankungen neben andern, stärker störenden Einflüssen, welche theoretisch nicht erfasst werden können (wie Zinsfuss, Storno, Unkosten), nur eine geringe Rolle spielten. Einen ausgezeichneten Ersatz für sie böten die eingehenden Statistiken über die Gewinn- und Verlustquellen (Gain and Loss Exhibits), wie sie die konzessionierten Versicherungsgesellschaften der Vereinigten Staaten pflegen.

Trotz diesen vernichtenden Urteilen kann nicht bezweifelt werden, dass die Wahrscheinlichkeitstheorie besser ist als ihr Ruf. Dafür zeugen schon die vielen ausgezeichneten Resultate, die man mit ihrer Hilfe

---

[1]) P. Nolfi: «Die jährlichen Sterblichkeitsschwankungen und ihre wahrscheinlichkeitstheoretische Erfassung», Kongressband I, S. 405.

[2]) Ivo Lah: «Wahrscheinlichkeitsrechnung und Versicherungswesen», Kongressband I, S. 419.

[3]) ten Pas: «Wahrscheinlichkeitsrechnung und Statistik in der Versicherungsmathematik», Kongressband I, S. 264.

[4]) Simon Shannon: «Some Assumptions and Hypotheses underlying Actuarial Calculations», Kongressband I, S. 137.

in der Physik, in der Biologie und auch im Versicherungswesen erzielt hat. Wenn sie auch noch nicht allen Fragestellungen gewachsen ist, so bietet sie uns eben doch das einzige theoretische Hilfsmittel dar, mit dem Gesetzmässigkeiten in Massenerscheinungen erfasst werden können, wenn Aussagen über den Einzelfall aus prinzipiellen oder praktischen Gründen unmöglich sind. Einen ähnlichen Gedanken spricht auch Jean Baptist in der Zusammenfassung zu seinem Kongressbeitrag aus:

«Die Wahrscheinlichkeitstheorie scheint jedenfalls das einzige Mittel zu sein, das uns in der praktischen Forschung nach dem wahren Sachverhalt zur Verfügung steht. ... Wie immer man diese Theorie auch benennen mag, sie bildet das Kettenglied, das die rein deduktive Wissenschaft mit der experimentellen verbindet [1]).»

Die Gefahr eines einseitigen *Rationalismus* war in der Versicherungsmathematik nie besonders gross. Die Abhängigkeit jeder versicherungsmathematischen Theorie von der Erfahrung ist zu augenfällig. Es erübrigt sich deshalb, näher auf diesen andern extremen Standpunkt einzugehen.

Zusammenfassend können wir also sagen: Wenn die Versicherungsmathematik eine Wissenschaft und nicht nur ein blosses Handwerk (wenn auch ein mit geistigem Werkzeug betriebenes) sein soll, dann müssen ihre Methoden theoretisch begründet werden. Die Mittel zu dieser Begründung kann nur die Wahrscheinlichkeitstheorie liefern. Die praktischen Erfahrungen sind ständig mit den Resultaten der Theorie zu vergleichen, und wenn keine befriedigende Übereinstimmung besteht, ist den Ursachen der Diskrepanz nachzugehen. Dies wird dazu führen, dass einerseits die Theorie immer mehr verfeinert und anderseits Mängel in den statistischen Erhebungs- und Verarbeitungsmethoden aufgedeckt werden. H. Wiesler bezeichnet diese Verbindung der Standpunkte des Empirismus und des Rationalismus treffend als Komplementaritätsprinzip, in Analogie zum bekannten Komplementaritätsprinzip der theoretischen Physik [2]).

Im Sinne dieser Ausführungen wollen wir uns jetzt einer wahrscheinlichkeitstheoretischen Untersuchung der relativen Sterbehäufigkeiten zuwenden.

---

[1]) Jean Baptist: «Le calcul des probabilités dans le domaine de l'assurance», Kongressband I, S. 133.

[2]) H. Wiesler: «Über die Grundlagen der Lebensversicherungsmathematik», MVSV Bd. 44, 1944, S. 193.

II. Die Streuung der relativen Sterbehäufigkeiten

A. Allgemeines

1. Die Sterbefläche

Für die einjährigen relativen Sterbehäufigkeiten, wie sie uns die Statistik liefert, ist zuallererst ihre Abhängigkeit vom Alter der beobachteten Personen charakteristisch. Zwar spielen auch noch andere Umstände eine Rolle, wie Beruf, soziale Stellung, abgelaufene Versicherungsdauer etc.; aber sie sind lange nicht so wichtig und sollen uns hier auch nicht weiter beschäftigen. Aber einen zweiten variablen Parameter müssen wir doch noch einführen: die Zeit. Die Erfahrungen der letzten Jahrzehnte lehrten uns ja, dass die Sterblichkeit nicht konstant bleibt, sondern in allen Altersklassen mehr oder weniger stark abnimmt. Wir setzen also

$$q = q(x, t).$$

Diese Funktion, welche uns den Bruchteil der x-jährigen angibt, die im Jahre t gestorben sind, ist nur für ganzzahlige Werte von x und t erklärt, also für die Gitterpunkte der x, t-Ebene. Um aber ein anschaulicheres Bild zu gewinnen, mag man sie sich für beliebige reelle Werte der Intervalle

$$0 \leqq x \leqq \omega$$

$$0 \leqq t \leqq T$$

interpoliert denken.

Die so entstehende Fläche über der x, t-Ebene wird aber dem Bild, das wir uns in Gedanken von ihr gemacht haben, nur wenig ähnlich sehen. Sie wird in gänzlich unübersichtlicher Weise von Einbuchtungen und Gipfeln, Gräben und Gräten überstreut und durchzogen sein, während uns doch eine möglichst glatt verlaufende Fläche vorschwebt; denn wir suchen ja hinter jedem Naturgeschehen einfache Gesetzmässigkeiten. Dort, wo sie nicht offen zutage treten, versuchen wir, sie durch Befreiung der Beobachtungsdaten von zufälligen störenden Einflüssen herauszuschälen.

Wir machen deshalb die Hypothese, dass unsere beobachteten Sterbehäufigkeiten sich aus zwei Komponenten zusammensetzen: der «wahren» Sterblichkeit und einer Schwankungskomponente. Dabei soll dem Wörtchen «wahr» keinerlei mystisch-philosophische Bedeutung zugeschrieben werden. Wir definieren die wahren Sterblichkeiten rein mathematisch durch die beiden Forderungen, dass sie 1. durch eine glatte Fläche sollen verbunden werden können und 2. möglichst nahe bei den beobachteten Werten liegen sollen. Dass dies zu ihrer Festlegung genügt, werden wir im III. Kapitel sehen. Vorher wollen wir die zweite, die Schwankungskomponente untersuchen.

2. Das Urnenschema

Entsprechend unserer Auffassung vom Naturgeschehen möchten wir auch die Schwankungskomponente nicht für gänzlich unberechenbar halten, sondern auch sie durch Gesetzmässigkeiten zu beschreiben und zu erfassen suchen. Es ist naheliegend, die beobachteten Schwankungen mit einer Art Standard- oder Normalschwankungen zu vergleichen. Diese Rolle spielen die Schwankungen, die bei Ziehungen aus einer Urne mit Kugeln verschiedener Farbe auftreten und die zur Binomialverteilung (und daraus durch Grenzübergang abgeleitet zur Gauss-Verteilung und Poisson-Verteilung) führen. Man nennt Schwankungen dieser Art Wahrscheinlichkeitsschwankungen und schreibt den zugrunde liegenden relativen Häufigkeiten den Charakter von echten Wahrscheinlichkeiten zu.

Wir wollen unsere Betrachtungen auf folgendes Urnenschema stützen: Jedem Gitterpunkt (x, t) sei eine Urne $U_{x,t}$ zugeordnet, welche schwarze und weisse Kugeln im Verhältnis

$$Q(x, t) : 1 - Q(x, t) \qquad [0 < Q(x, t) < 1]$$

enthält. Das sind im ganzen $(\omega + 1)(T + 1)$ Urnen, die wir in einem rechteckigen Schema mit $\omega + 1$ Zeilen und $T + 1$ Kolonnen anordnen. Als Versuchsvorschrift setzen wir fest: Aus jeder Urne wird je eine Stichprobe entnommen, und zwar aus der Urne $U_{x,t}$ $\lambda(x, t)$ Kugeln, worunter sich $\delta(x, t)$ schwarze befinden mögen. Wir wollen annehmen, dass in jeder Urne so viele Kugeln seien, dass man auch ohne Zurücklegen der Kugeln mit der Bernoullischen Verteilung rechnen darf.

Die Streuung der empirischen Sterbehäufigkeiten $q(x, t)$ vergleichen wir nun mit der theoretischen Streuung der Quotienten

$$\varkappa(x, t) = \frac{\delta(x, t)}{\lambda(x, t)}.$$

Dieser Vergleich kann in verschiedenen Richtungen erfolgen. So können wir bei festem x die Streuung der Reihe $q(x, 0), q(x, 1), \ldots, q(x, T)$ in Beziehung setzen mit der entsprechenden theoretischen Streuung der $\varkappa(x, 0), \varkappa(x, 1), \ldots, \varkappa(x, T)$. Ich nenne das die Streuung in der t-Richtung. Dabei können wir uns insbesondere noch fragen, ob sich bei nicht zu grossem T das Mischungsverhältnis dieser Urnen derselben Zeile als konstant annehmen lässt, mit andern Worten, ob die säkulare Änderung der Sterblichkeit während kürzerer Zeitperioden vernachlässigt werden kann.

Anderseits können wir t festhalten und die Streuung der Grössen $q(0, t), q(1, t), \ldots, q(\omega, t)$ vergleichen mit der Streuung der $\varkappa(0, t)$, $\varkappa(1, t), \ldots, \varkappa(\omega, t)$. Dies ist die Streuung in der x-Richtung.

Die bisherigen Untersuchungen beschäftigen sich sozusagen ausschliesslich mit der Streuung in der t-Richtung. H. Wiesler hat sie in seiner Arbeit «Über die Grundlagen der Lebensversicherungsmathematik» (9) zusammengestellt und eigene Rechnungen angeschlossen. Ich brauche deshalb nicht mehr im Detail darauf einzugehen, sondern will nur kurz an die Hauptergebnisse erinnern.

3. Die Streuung in der t-Richtung

Für Zeitperioden bis ca. 20—25 Jahre erweist sich die Streuung in allen Altersklassen von ca. 10 Jahren an aufwärts ziemlich gut normal, kriegs- und epidemienfreie Jahre vorausgesetzt, und zwar darf hier $Q(x, t)$ bei festem x für alle t konstant angenommen werden. Bei der Kindersterblichkeit ist das Bild nicht einheitlich. Die ältern Untersuchungen von Lexis und Peek zeigen deutlich übernormale Dispersion, ebenso auch die Berechnung von Wiesler für die männliche Gesamtbevölkerung der Schweiz in der 22jährigen Beobachtungsperiode von 1889 bis 1910. Bei den Volksversicherten der Basler Lebensversicherungsgesellschaft dagegen fand er für die Nulljährigen eine Dispersion von 1,41 (Beobachtungsdauer 14 Jahre) und für die

1—9jährigen eine solche von 1,12 (Beobachtungsdauer 25 Jahre). Hier ist also ein abschliessendes Urteil noch nicht möglich.

Übersteigt die Beobachtungsdauer 25 Jahre, so wird die Dispersion in allen Altersklassen ausgesprochen übernormal. Dann lässt sich also die wesentliche Schwankungskomponente nicht mehr vernachlässigen. So bezog C. Lange in seiner Arbeit «Untersuchungen über die jährlichen Schwankungen der Schadensquotienten in der Lebensversicherung und in der Feuerversicherung» (10), bei der er sich auf die Statistik des Staates Preussen der Jahre 1867—1929, also eine 63jährige Beobachtungsperiode, stützte, seine Rechnungen auf linear abfallende Trendlinien der Sterblichkeit. Er untersuchte dann nicht nur die Streuung, sondern die ganzen Verteilungen um diese Trendlinien und fand in allen Altersklassen mit sehr guter Annäherung Gauss-Verteilungen. Die Dispersion allerdings war überall deutlich übernormal. Dies mag zum Teil darauf zurückzuführen sein, dass der Trend nicht genau eine Gerade zu sein braucht, was eine willkürliche Annahme darstellt, wenn sie auch in diesem Fall plausibel scheint. Ferner hat Lange für die theoretische Streuung einen Ausdruck benutzt, der keineswegs gleich dem Erwartungswert der empirischen Streuung bei Normalverteilung ist, so dass sich auch hieraus möglicherweise gewisse Korrekturen ergeben könnten.

In der neuesten Arbeit, die zu diesem Gegenstand erschienen ist, gibt der Verfasser, H. Ammeter (11), ebenfalls der Vermutung Ausdruck, dass die Dispersion in allen Altersklassen übernormal ist, wenn auch nur schwach. (Eine Ausnahme bilden nur die Altersklassen über 65, welche in dieser Untersuchung mit stark unternormaler Dispersion merkwürdig aus dem Rahmen der übrigen Resultate fallen.) H. Ammeter stützt sich dabei auf seine Rechnungen, die er anhand der Sterblichkeitserfahrungen der Schweizerischen Lebensversicherungs- und Rentenanstalt bei den Einzelkapitalversicherungen des Hauptzweiges in den Jahren 1869—1938 und der Volksversicherungen in den Jahren 1908—1938 anstellte. Die Dispersion, die er dabei fand, ist in allen Fällen (mit der erwähnten Ausnahme) so wenig von 1 verschieden, dass die Abweichung ohne weiteres als zufällig gelten könnte. Dem steht aber entgegen, dass sämtliche Abweichungen positiv sind, während sie doch bei Zufallsschwankungen ungefähr gleich oft positiv wie negativ auftreten sollten. Mir scheint aber die übernormale Dispersion, wie sie H. Ammeter gefunden hat, in der besondern Art seiner

Berechnung eine natürliche Erklärung zu finden. Die Sterblichkeit wird nicht für jedes Altersjahr gesondert betrachtet, sondern es werden Gruppen zu je fünf Jahresklassen gebildet; auch werden beide Geschlechter zusammengefasst. Dadurch können zusätzliche Schwankungen entstehen, als Folge der wechselnden Zusammensetzung der Fünfergruppen nach Alter und Geschlecht. Ein zweiter Grund besteht wie bei Lange in der willkürlichen Annahme des Trends. Ammeter geht hier zwar weiter als Lange, indem er nicht nur eine einzige Gerade, sondern deren drei verwendet, um die säkulare Sterblichkeitsabnahme zu beschreiben. Trotzdem kann nicht mit Sicherheit behauptet werden, dass damit die wesentliche Schwankungskomponente vollständig eliminiert sei.

Auf Grund dieser Untersuchungen scheint es somit berechtigt, die relativen Sterbehäufigkeiten als Wahrscheinlichkeiten zu betrachten, mit dem Vorbehalt allerdings, dass das Vorliegen von normaler Dispersion an und für sich noch keine streng hinreichende Begründung für eine Bernoulli- oder Gauss-Verteilung darstellt. Tatsächlich begnügt man sich aber damit, weil die höhern Momente der Verteilung mit einer zu grossen Unsicherheit behaftet sind und deshalb keine zuverlässigen Schlüsse zulassen.

Unter der Annahme einer Gauss-Verteilung

$$w(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

beträgt nämlich der Erwartungswert für das $k$te Moment

$$E(x^k) = \int_{-\infty}^{\infty} x^k w(x)\, dx = \begin{cases} 0 & \text{für ungerade } k \\ 1 \cdot 3 \cdot 5 \cdot \ldots \cdot (k-1) & \text{für gerade } k \end{cases}$$

Die Streuung des $k$ten Momentes um seinen Erwartungswert wird:

$$\text{Str. } (x^k) = E(x^{2k}) - [E(x^k)]^2 = \begin{cases} 1 \cdot 3 \cdot 5 \ldots (2k-1) & \text{für ungerade } k \\ 1 \cdot 3 \cdot 5 \ldots (2k-1) - [1 \cdot 3 \cdot 5 \ldots (k-1)]^2 & \text{für gerade } k \end{cases}$$

Also für $k = 3 : E(x^3) = 0$ Streuung $(x^3) = 15$

$k = 4 : E(x^4) = 3$ Streuung $(x^4) = 96$

usf.

Diese Zahlen beziehen sich auf die einzelnen Beobachtungen. Hat man T Beobachtungen mit gleicher Wahrscheinlichkeitsverteilung, so wird die Streuung des empirischen $k$ten Moments:

$$\text{Streuung}\left(\frac{1}{T}\sum_{i=0}^{T} x_i^k\right) = \frac{1}{T}\ \text{Streuung}\ (x^k).$$

Die mittleren Fehler werden also um den Faktor $\frac{1}{\sqrt{T}}$ verkleinert. Da aber bei uns $T \lessapprox 25$, bleiben sie immer noch beträchtlich, nämlich ca. $^4/_5$ für das dritte und ca. 2 für das vierte Moment.

Die Berücksichtigung der höhern Momente wäre also unter diesen Umständen ziemlich problematisch. Aber auch ein direkter Vergleich der beiden Verteilungen hat seine Schwierigkeiten, weil sich über eine Häufigkeitsverteilung mit nur 25 Elementen noch nicht viel Bestimmtes aussagen lässt. Wählt man aber, um mehr Elemente zu haben, $T > 25$, wie es C. Lange getan hat, so darf man den Trend nicht mehr vernachlässigen. Dann wird aber die Verteilung wesentlich von der gewählten Trendlinie abhängig und wie diese in weitem Masse willkürlich.

4. Die Streuung in der x-Richtung

Mit der Streuung in der x-Richtung hat man sich bis jetzt nur wenig beschäftigt [1]). Dies mochte zum Teil davon herrühren, dass man eine Methode vermisste, welche erlaubte, die Streuung unabhängig von der Ausgleichung der Sterbehäufigkeiten zu berechnen, was unumgänglich ist, wenn jede Willkür ausgeschaltet sein soll. Zum andern Teil aber war der Grund vielleicht auch der, dass man annahm, durch die normale Dispersion in der t-Richtung wäre der Wahrscheinlichkeitscharakter der Sterblichkeit genügend belegt, so dass auf die Berechnung der Streuung in der x-Richtung verzichtet werden könne.

Die normale Streuung in der t-Richtung zeigt ja tatsächlich (mit dem erwähnten Vorbehalt), dass jedes einzelne q_x als echte Wahrscheinlichkeit aufgefasst werden kann. Damit ist aber unser rechteckiges

---

[1]) Z. B. Ronald L. Michaelson und Frank M. Redington in ihrer Kongressarbeit «An Aspect of the „a priori" Probability Theory of Mortality», Kongressband I, wo für eine Selektionstafel die Dispersion 1,09 und für die zugehörige Schlusstafel eine solche von 1,5 berechnet wird.

Urnenschema noch nicht genügend begründet. Denn nach unserer Versuchsvorschrift sind die Ziehungen in den verschiedenen Zeilen voneinander unabhängig. Ob das aber auch für die Sterbehäufigkeiten zutrifft, das können wir aus den Dispersionsberechnungen in der t-Richtung allein noch nicht schliessen; denn diese betrachten ja immer nur eine Zeile für sich allein, unabhängig davon, was in den übrigen Zeilen geschieht. Eine allfällige Abhängigkeit der verschiedenen Zeilen voneinander würde sich also in den Divergenzquotienten in keiner Weise äussern.

Um zu zeigen, dass dies nicht nur eine theoretische Spitzfindigkeit, sondern eine Frage von praktischem Interesse ist, wollen wir den Ursachen, welche die Sterblichkeitsschwankungen bewirken, etwas nachgehen und dabei einen Vergleich anstellen zwischen der Streuung in der x-Richtung und jener in der t-Richtung.

5. Die Ursachen der Sterblichkeitsschwankungen

Es sind zwei Ursachen, welche die Sterblichkeitsschwankungen bewirken. Die erste besteht in der stets wechselnden zufälligen Zusammensetzung des Beobachtungsmaterials, welches das eine Mal vielleicht besonders viele kränkliche, anfällige Personen, ein anderes Mal wieder mehr gesunde, widerstandsfähige Naturen aufweist. Diese Schwankungsursache wirkt sich in der x-Richtung wie in der t-Richtung in gleicher Weise aus, und sie ist von der Art, wie sie die Fehlertheorie für das Zustandekommen von «zufälligen» Fehlern fordert: Schwankungen nach oben und unten sind gleich wahrscheinlich, und kleine Schwankungen sind wahrscheinlicher als grosse [1]). In der t-Richtung, also bei festem x, sind diese Schwankungen auch sicher unabhängig voneinander; denn ein Beobachtungsbestand von x-jährigen im Jahr t kann weder in einem vorangehenden noch in einem spätern Jahr ein zweites Mal mit dem Alter x unter Risiko stehen. Die einzelnen Beobachtungsgesamtheiten sind deshalb alle fremd zueinander und deshalb auch unabhängig, wenigstens soweit diese erste, rein stochastische Schwankungsursache in Frage steht.

Auch in der x-Richtung, bei festem t, ist aus dem gleichen Grund diese Unabhängigkeit gegeben, nicht aber in der Richtung $t-x=$ const., also bei den sogenannten Generationssterbetafeln. Hier ist es ja, wenn

[1]) Vgl. Czuber (1), Bd. I, S. 289 f.

wir von Wanderungen absehen, immer wieder derselbe Bestand, nämlich eine Jahresklasse, der beobachtet wird, und es ist deshalb zwischen den Schwankungen der Sterblichkeit in den verschiedenen Altern, vor allem der benachbarten, eine gewisse Korrelation zu erwarten. Dabei zeigt sich der merkwürdige Umstand, dass sowohl Gründe für positive als für negative Korrelation sprechen. Nehmen wir einmal an, dass die x-jährigen im Jahre t eine auffallend geringe Sterblichkeit hatten. Man kann dann annehmen, dass dieser Bestand aus besonders robusten Leuten zusammengesetzt sei, und deshalb auch für die $x+1$jährigen im Jahre $t+1$ eine geringe Sterblichkeit voraussagen. Anderseits aber könnte man auch vermuten, dass, nachdem aus unserm Bestand im Jahre t nur wenige Personen weggestorben sind, im nächsten Jahre eine Tendenz zur Nachholung der ausgebliebenen Todesfälle besteht und deshalb die Sterblichkeit unseres Bestandes im Jahre $t+1$ eher überdurchschnittlich ausfallen werde. Diese beiden Tendenzen zu positiver und negativer Korrelation können sich gegenseitig aufheben, so dass eine scheinbar korrelationslose Reihe entsteht. Die eine kann aber auch die andere so stark übertreffen, dass sie nachweisbar wird. So stellten E. M. Elderton und K. Pearson in zwei Arbeiten (14 und 15) unter Verwendung von zwei verschiedenen Methoden zwischen $q(x, t)$ und $q(x+1, t+1)$ bei Säuglingen und Kleinkindern eine deutliche negative Korrelation fest: in der ersten Arbeit für $x = 0, 1, 2, 3$, je ca. $-0{,}7$ und in der zweiten Arbeit für $x = 0$ ca. $-0{,}5$. Die Beobachtungen erstreckten sich über die Periode von 1850 bis 1912 bzw. 1850 bis 1919 am Material der Registrar-General's Returns for Births in England and Wales. Leider fehlen für die übrigen Alter entsprechende Untersuchungen, vor allem wohl, weil genügend umfangreiches, zuverlässiges statistisches Material nur schwer aufzutreiben wäre. Es ist immerhin zu vermuten, dass bei den Erwachsenen die Antikorrelation weniger ausgeprägt sein dürfte, wenn sie überhaupt nachweisbar ist [1]).

Diese Frage der Korrelation bei Generationssterbetafeln ist deshalb von Interesse, weil bei der Berechnung der üblichen Sterbetafeln,

[1]) Die Korrelationskoeffizienten, die H. Ammeter in seiner Arbeit (11) berechnet, beziehen sich nicht auf die Korrelation zwischen $q(x, t)$ und $q(x+1, t+1)$, sondern auf jene zwischen $q(x, t)$ und $q(x+1, t)$, wo x und $x+1$ zwei benachbarte Gruppen von je fünf Altersjahren bedeuten. Sie erweisen sich in keinem Fall als wesentlich von null verschieden.

um möglichst grosse Gesamtheiten zu erhalten, ja oft die Beobachtungen mehrerer Jahre zusammengefasst werden. So entsteht eine Kombination einer Tafel gleichzeitig Lebender mit Abschnitten von Generationstafeln, und es erhebt sich die Frage, ob nicht die allfällige Korrelation der Generationstafeln auf die kombinierte Tafel abfärben könnte.

Die zweite Schwankungsursache der Sterblichkeit ergibt sich aus dem An- und Abschwellen der verschiedenen Krankheitshäufigkeiten. Wie aus den Beobachtungen hervorgeht, zeigen diese Schwankungen in der t-Richtung das mehr oder weniger ausgeprägte Bild einer Wellenbewegung [1]). Aus physiologisch plausiblen Gründen sind diese Wellen besonders deutlich bei den epidemischen Krankheiten, wie Typhus [2]), Scharlach und Diphtherie [3]), Poliomyelitis [4]), Meningitis [5]), für welche spezielle Untersuchungen vorliegen. Bei gewissen andern Krankheiten dagegen, z. B. beim Krebs, ist eine Wellenbewegung kaum wahrzunehmen.

Überall dort, wo die Schwankungen durch solche Wellenbewegungen bestimmt oder wenigstens mitbestimmt werden, darf man nicht mehr mit vollem Recht von Zufallsschwankungen sprechen; denn sie unterliegen ja einer, wenn vielleicht auch schwachen Gesetzmässigkeit, und zwar einer Gesetzmässigkeit, welche sich auf die Einzelbeobachtung bezieht und nicht nur auf die Mittelwerte, wie es bei den sogenannten Zufallsgesetzen der Fall ist.

Aus zwei Gründen wirken aber diese Wellenbewegungen auf die Gesamtsterblichkeit ähnlich wie zufällige Fehlerquellen: erstens variieren Wellenlänge und Amplitude der Wellen ziemlich stark und unberechenbar, und zweitens ergibt sich ja die Gesamtsterblichkeit durch Superposition aller dieser Wellen verschiedener Wellenlänge und Amplitude. Daraus resultiert ein Gesamtbild, das sich von einer Zufallsschwankungen unterworfenen Reihe nicht mehr wesentlich unterscheidet.

---

[1]) Braun (12). Ferner: Braun: «Der Rhythmus der Lebensvorgänge und seine Verursachung», Blätter für Versicherungsmathematik, Bd. 5, Juni 1941; Burckhardt: «Neuere versicherungswissenschaftliche Probleme und ihre Bedeutung für die Versicherungspraxis», Neumanns Zeitschrift für Versicherungswesen, Nr. 27, 1937.

[2]) Ziezold: «Weitere Ergebnisse der Analyse von Seuchenkurven», Zeitschrift für Hygiene, Bd. 124, 1942.

[3]) Kupfermann: Diss. Berlin 1942.

[4]) Fischer: Diss. Berlin 1941.

[5]) Mrugowsky: Zeitschrift für Hygiene, Bd. 123, 1941.

(Alle zitiert nach Braun.)

In der eben geschilderten Weise wirkt diese zweite Schwankungsursache auf die Streuung in der t-Richtung. In der x-Richtung dagegen liegt der Fall anders. Hier ist es die ungleich starke Anfälligkeit der einzelnen Altersklassen für die verschiedenen Krankheiten, welche die Schwankungen verursacht. Diese Anfälligkeit variiert aber von Altersjahr zu Altersjahr nichts weniger als in zufälliger Weise; vielmehr bietet sie in der graphischen Darstellung das Bild einer ziemlich regelmässig verlaufenden Kurve. Benachbarte Jahrgänge werden im gleichen Sinne beeinflusst, so dass zwischen den Schwankungen aufeinander folgender Reihenglieder starke positive Korrelation besteht. Ist z. B. die Sterblichkeit an Diphtherie in einem bestimmten Jahr besonders gross, so wird dies für die Sterbeziffern aller Jahrgänge bis ca. 15 einen Ausschlag nach oben zur Folge haben, wenn wir von den übrigen Einflüssen absehen. Schwankungen dieser Art werden bei der Ausgleichung nicht als solche erkannt, sondern geben dazu Anlass, dass die ausgeglichene Kurve in der betreffenden Altersgruppe durchgehend zu hoch (bzw. zu tief) angesetzt wird. Ob die Gesamtstreuung wesentlich beeinflusst wird, hängt vom Stärkeverhältnis der beiden Schwankungsursachen ab, über das auf direktem Wege nur schwer Aufschlüsse zu gewinnen sein werden. Dagegen wird ein numerischer Vergleich der Streuung in der x-Richtung mit derjenigen in der t-Richtung Rückschlüsse auf dieses Verhältnis zulassen.

Damit ist wohl genügend dargetan, dass die Streuung in der x-Richtung eine eigene Untersuchung erfordert. Mit ihr wollen wir uns jetzt beschäftigen.

B. Die Berechnung der Streuung in der x-Richtung

1. Das Problem

Bei der Berechnung der Streuung in der x-Richtung zeigt sich eine Schwierigkeit, welche in der t-Richtung vernachlässigt werden konnte, wenigstens bei kürzeren Perioden. Dort durften wir den Erwartungswert der Sterblichkeit eines bestimmten Alters während Zeitperioden bis zu 25 Jahren konstant annehmen, weil sich die säkulare Änderung nur langsam bemerkbar macht. Dieser Erwartungswert lässt sich durch das arithmetische Mittel der Beobachtungswerte sehr gut approximieren, von dem dann die Abweichungen berechnet werden.

Zur Berechnung der Streuung in der x-Richtung aber lässt sich dieses einfache Verfahren nicht mehr verwenden, weil sich der Erwartungswert der Sterblichkeit von Altersjahr zu Altersjahr so bedeutend ändert, dass von Konstanz nicht entfernt die Rede sein kann. Die wesentliche Schwankungskomponente darf also hier nicht vernachlässigt werden, vielmehr müssen wir sie auf irgendeine Weise zu eliminieren suchen.

Der nächstliegende Gedanke wäre, die Tafel zuerst auszugleichen und dann die Abweichungen von den ausgeglichenen Werten zu berechnen. Die letzteren würden somit als Erwartungswerte für die Sterblichkeit des betreffenden Alters betrachtet. Dann stellt sich aber sofort die Frage: Wie soll ausgeglichen werden? Nach welcher der vielen Methoden? Die ausgeglichene Kurve ist ja gar nicht eindeutig gegeben, sondern mit gewissen Einschränkungen frei variierbar. Wir können durch eine geeignet gewählte Ausgleichung ohne weiteres erreichen, dass die Abweichungen um die ausgeglichene Kurve eine beliebig vorgegebene Verteilung, z. B. eine Normalverteilung, bilden. Dieser Weg ist also nicht gangbar, weil er zu ganz willkürlichen und deshalb wertlosen Resultaten führt.

Es handelt sich also darum, die unwesentliche Schwankungskomponente unabhängig von der Ausgleichung zu berechnen. Dies gelingt uns mit Hilfe der Differenzenmethode (Variate Difference Method). Da die Bücher und Zeitschriften, in denen diese Methode ausführlich behandelt ist, hierzulande nur wenig verbreitet sind, ist es wohl nicht unangebracht, wenn ihre Grundideen an dieser Stelle kurz dargelegt werden, wenigstens soweit wir sie nachher brauchen.

2. Die Differenzenmethode (Variate Difference Method)

Die ersten Grundlagen dieser Methode wurden in Anknüpfung an Gedanken von Cave und Hooker beinahe gleichzeitig und unabhängig voneinander durch «Student» (Pseudonym von W. S. Gosset) und O. Anderson geschaffen. Ihre Arbeiten sind im gleichen Band der Biometrika erschienen:

«Student»: «The Elimination of Spurious Correlation due to Position in Time or Space» (Biometrika, Bd. X, S. 179); O. Anderson:

Nochmals über «The Elimination of Spurious Correlation due to Position in Time or Space» (Biometrika, Bd. X, S. 269).

Die Methode wurde von der Pearsonschen Schule sofort aufgenommen und mit Erfolg zu statistischen Untersuchungen auf den verschiedensten Gebieten verwendet. Eine zusammenfassende Darstellung gab Anderson in seinem kleinen Buch «Die Korrelationsrechnung in der Konjunkturforschung», erschienen 1929 als Heft 4 der Veröffentlichungen der Frankfurter Gesellschaft für Konjunkturforschung.

Ausgangspunkt der Methode ist die Tatsache, dass alle ganzen rationalen Funktionen beim «endlichen Differenzieren» (d. h. bei der Bildung der Differenzen 1., 2., 3., ff. Ordnung bei äquidistanten Argumentenwerten) nach endlich vielen Schritten verschwinden (nämlich in der $n+1$. Differenzenreihe, wenn der Grad der Funktion n beträgt) und dass bei vielen andern geläufigen Funktionen (z. B. der Exponentialfunktion, dem Logarithmus, den Potenzfunktionen mit negativen und gebrochenen Exponenten u. a.) die Differenzen mit steigender Ordnung immer kleiner werden, d. h. es gelten die Ungleichungen:

$$\left.\begin{array}{l} |\Delta^k x_i| > |\Delta^{k+1} x_i| \\ \text{oder} \quad |\Delta^k x_i| > |\Delta^{k+1} x_{i-1}| \end{array}\right\} \text{ für alle } i \text{ und } k.$$

Bei einer Funktion dagegen, deren Funktionswerte eine zufällige Reihe bilden (man denke, um ein Beispiel zu haben, etwa an eine Funktion, deren Werte der Reihe nach durch Würfeln oder durch Ziehungen aus einer Urne bestimmt werden), nehmen die Differenzen steigender Ordnung dem absoluten Betrage nach immer mehr zu. Setzen wir nämlich

$$\Delta x_i = x_{i+1} - x_i$$

$$\Delta^2 x_i = \Delta x_{i+1} - \Delta x_i$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\Delta^k x_i = \Delta^{k-1} x_{i+1} - \Delta^{k-1} x_i$$

so erhalten wir durch sukzessives Einsetzen

$$\Delta^k x_i = \sum_{j=0}^{k} (-1)^j \binom{k}{j} x_{i+k-j}.$$

Der Einfachheit halber wollen wir annehmen, dass der Erwartungswert für alle x_i gleich Null ist, dass alle x_i dasselbe Verteilungsgesetz besitzen und voneinander unabhängig sind. Es sei also:

$$\begin{aligned} E x_i &= 0 \text{ für alle } i \\ E x_1^2 &= E x_2^2 = \ldots = E x_N^2 = E x^2 \\ E(x_i x_j) &= 0 \text{ für } i \neq j. \end{aligned}$$

Dann wird der Erwartungswert für die Quadrate der $k$ten Differenzen (statt der absoluten Beträge betrachten wir die Quadrate, was einfacher ist):

$$E(\Delta^k x_i)^2 = E\left\{\sum_{j=0}^{k} (-1)^j \cdot \binom{k}{j} \cdot x_{i+k-j}\right\}^2 = \sum_{j=0}^{k} \binom{k}{j}^2 \cdot E x^2.$$

Nach einer aus der Kombinatorik bekannten Beziehung ist aber:

$$\sum_{j=0}^{k} \binom{k}{j}^2 = \binom{2k}{k}.$$

Also:

$$E(\Delta^k x_i)^2 = \binom{2k}{k} \cdot E x^2. \tag{I}$$

Die Quadrate der Differenzen einer zufälligen Reihe nehmen also mit steigender Ordnung k zu wie $\binom{2k}{k}$.

Dank diesem unterschiedlichen Verhalten der Differenzenreihen von glatten und von zufälligen Funktionen ist es möglich, bei Funktionen, die sich aus einer glatten und einer zufälligen Komponente zusammensetzen, die erstere zu eliminieren und damit die letztere einer besonderen Betrachtung zugänglich zu machen, ohne dass über die Form der glatten Komponente irgendwelche speziellen Voraussetzungen gemacht werden müssten.

Es sei nun X_i eine zufällige Variable, die noch von einem Parameter i abhängen soll, der die Werte $1, 2, \ldots, N$ annehmen kann. (Unter einer zufälligen Variabeln versteht man eine Grösse, welche verschiedene Werte mit gewissen Wahrscheinlichkeiten annehmen kann.) Zu jedem Werte i dieses Parameters gehört also ein Verteilungsgesetz $w_i(X_i)$, welches die Wahrscheinlichkeit angibt, dass X_i zwischen X_i und $X_i + dX_i$ liegt. Diese Wahrscheinlichkeitsverteilungen sollen gegenseitig unabhängig, sonst aber ganz beliebig sein. (Der Wert von X_i soll also nicht von den Werten der übrigen X_j abhangen.)

Die X_i spalten wir nun in zwei Komponenten auf: ihren Erwartungswert m_i und eine Schwankungskomponente x_i:

$$X_i = m_i + x_i$$

$$\text{wo } m_i = EX_i = \int_{-\infty}^{\infty} X_i w_i(X_i)\, dX_i$$

$$\text{und deshalb } Ex_i = 0.$$

Ferner wird

$$E(x_i x_j) = 0, \text{ wenn } i \neq j$$

wegen der Unabhängigkeit der Verteilungen.

Da bei den m_i das Zufallselement nun ausgeschaltet ist, dürfen wir annehmen, dass die Abhängigkeit von m_i vom Parameter i durch eine glatte Funktion dargestellt werde. (Unter einer glatten Funktion wollen wir jetzt immer eine Funktion verstehen, deren Differenzen zunehmender Ordnung dem absoluten Betrage nach nicht grösser werden.) Diese Annahme entspricht unserer Auffassung, dass dem Naturgeschehen einfache Gesetze zugrunde liegen, und «einfach» heisst hier eben, dass die Werte m_i nicht zufällig verteilt sein sollen, sondern durch eine einfache, d. h. glatte Kurve verbunden werden können.

Diese Annahme ist auf jeden Fall viel allgemeiner als jene, die man gewöhnlich macht, nämlich dass die Werte m_i auf einer Kurve mit vorgegebener analytischer Gleichung liegen (z. B. auf einer Parabel $n$ter Ordnung oder einer Exponentialkurve oder einer Makehamkurve), von der dann nur noch wenige Parameter frei verfügbar sind, welche nach der Methode der kleinsten Quadrate oder einer ähnlichen Methode bestimmt werden.

Wir bilden nun bei unserer Reihe $X_1, X_2, \ldots, X_N$ die erste bis $k$te Differenzenreihe:

$$\Delta X_i = X_{i+1} - X_i = (m_{i+1} - m_i) + (x_{i+1} - x_i) = \Delta m_i + \Delta x_i$$

$$\Delta^2 X_i = \Delta X_{i+1} - \Delta X_i = (\Delta m_{i+1} - \Delta m_i) + (\Delta x_{i+1} - \Delta x_i) = \Delta^2 m_i + \Delta^2 x_i$$

. . . . . .

$$\Delta^k X_i = \Delta^{k-1} X_{i+1} - \Delta^{k-1} X_i = (\Delta^{k-1} m_{i+1} - \Delta^{k-1} m_i) + (\Delta^{k-1} x_{i+1} - \Delta^{k-1} x_i) = \Delta^k m_i + \Delta^k x_i$$

Diese Differenzen quadrieren wir und berechnen in jeder Differenzenreihe den Mittelwert der Quadrate, also die Grösse

$$\frac{1}{N-k} \sum_{i=1}^{N-k} (\Delta^k X_i)^2 .$$

Diesen Ausdruck wollen wir mit seinem Erwartungswert vergleichen:

$$E\left\{\frac{1}{N-k} \sum_{i=1}^{N-k} (\Delta^k X_i)^2\right\} = E\left\{\frac{1}{N-k} \sum_{i=1}^{N-k} (\Delta^k m_i + \Delta^k x_i)^2\right\} =$$

$$= \frac{1}{N-k} \sum_{i=1}^{N-k} (\Delta^k m_i)^2 + \frac{2}{N-k} \sum_{i=1}^{N-k} \Delta^k m_i \, E(\Delta^k x_i) + \frac{1}{N-k} \sum_{i=1}^{N-k} E(\Delta^k x_i)^2 .$$

Nun wollen wir annehmen, k sei genügend gross, so dass die glatte Komponente soweit zurückgedrängt ist, dass wir die $\Delta^k m_i$ gegenüber den $\Delta^k x_i$ vernachlässigen können. Das gilt dann um so mehr für deren Quadrate, und wir können die erste der drei obigen Summen weglassen. Die mittlere von ihnen verschwindet ohnehin; denn es ist ja

$$E(\Delta^k x_i) = E \sum_{j=0}^{k} (-1)^j \binom{k}{j} x_{i+k-j} = \sum_{j=0}^{k} (-1)^j \binom{k}{j} E x_{i+k-j} = 0$$

weil alle $E x_i$ gleich Null sind.

Damit wird:

$$E\left\{\frac{1}{N-k}\sum_{i=1}^{N-k}(\Delta^k X_i)^2\right\} = \frac{1}{N-k}\sum_{i=1}^{N-k}E(\Delta^k x_i)^2$$

$$= \frac{1}{N-k}\sum_{i=1}^{N-k}E\left[\sum_{j=0}^{k}(-1)^j\binom{k}{j}x_{i+k-j}\right]^2$$

$$= \frac{1}{N-k}\sum_{i=1}^{N-k}\sum_{j=0}^{k}\binom{k}{j}^2 E x_{i+k-j}^2 .$$

Nun müssen wir zwei Fälle unterscheiden. Im ersten wollen wir annehmen, dass $E x_i^2$ für alle i denselben Wert haben soll (was z. B. richtig ist, wenn alle x_i dasselbe Verteilungsgesetz besitzen). Dann liegt genau der Fall vor, den wir schon zu Beginn unserer Ausführungen über die Differenzenmethode betrachtet haben (S. 150), und wir können die dort abgeleitete Formel (I) übernehmen. Es wird also:

$$E\left\{\frac{1}{N-k}\sum_{i=1}^{N-k}(\Delta^k X_i)^2\right\} = \binom{2k}{k} E x^2 .$$

Bezeichnen wir noch:

$$\frac{1}{\binom{2k}{k}\cdot(N-k)}\sum_{i=1}^{N-k}(\Delta^k X_i)^2 = \sigma_k^2 = \text{empirische Streuung der } k\text{ten Differenzenreihe}$$

und $\qquad E x^2 = \mu^2 =$ a priori-Streuung der X_i um ihre Erwartungswerte,

dann wird: $\qquad \underline{E\sigma_k^2 = \mu^2 \quad \text{für } k \geqq k_0 .}$

Dabei soll die $k_0$te Differenzenreihe die erste sein, bei der wir die $(\Delta^k m_i)^2$ gegenüber den $(\Delta^k x_i)^2$ vernachlässigen dürfen. k_0 lässt sich wie folgt ungefähr bestimmen:

Bilden wir der Reihe nach $\sigma_1^2, \sigma_2^2, \sigma_3^2, \ldots$ so wird zunächst

$$\sigma_1^2 > \sigma_2^2 > \sigma_3^2 > \ldots$$

weil die wesentliche Schwankungskomponente schrittweise eliminiert wird. Sobald diese genügend zurückgedrängt ist, werden eine Reihe von ungefähr gleich grossen σ_k^2 auftreten:

$$\sigma_k^2 \approx \sigma_{k+1}^2 \approx \sigma_{k+2}^2 \approx \ldots .$$

Der Index k des ersten von ihnen wird der wahrscheinlichste Wert von k_0 sein.

Man darf nun aber nicht etwa annehmen, dass die σ_k^2 mit zunehmendem k gegen μ^2 konvergieren würden, so dass man mit steigender Differenzenordnung immer bessere Näherungswerte für μ^2 bekäme. Das ist leider nicht der Fall. Anderson hat die mittleren Fehler der Grössen σ_k^2 gegenüber ihrem Erwartungswert μ^2 bestimmt und gefunden, dass sie mit wachsendem k immer grösser werden. Für den Fall, dass die x_i die Normalverteilung mit der Streuung μ^2 befolgen, lauten diese Formeln für die ersten vier Differenzenreihen:

$$\text{mittlerer Fehler von } (\sigma_1^2) = \frac{\sqrt{3{,}000\,N - 4{,}000}}{N-1}\,\mu^2$$

$$\text{mittlerer Fehler von } (\sigma_2^2) = \frac{\sqrt{3{,}889\,N - 9{,}778}}{N-2}\,\mu^2$$

$$\text{mittlerer Fehler von } (\sigma_3^2) = \frac{\sqrt{4{,}620\,N - 16{,}860}}{N-3}\,\mu^2$$

$$\text{mittlerer Fehler von } (\sigma_4^2) = \frac{\sqrt{5{,}253\,N - 25{,}012}}{N-4}\,\mu^2 .$$

Aus diesem Grunde wird $\sigma_{k_0}^2$, also das erste der ungefähr gleich grossen σ_k^2, der wahrscheinlich beste Näherungswert für μ^2 sein, und zwar ist dieser Näherungswert um so zuverlässiger, je länger die Folge der ungefähr konstanten σ_k^2 ist. Es hat demnach keinen Sinn, die Bildung der Differenzenreihen allzu weit zu treiben. Normalerweise wird man mit $k = 4$ bis 6 auskommen und nur in seltenen Fällen vielleicht bis $k = 10$ gehen müssen.

Mit Hilfe der Grössen σ_k^2 lassen sich jetzt auch *Divergenzquotienten* bilden, indem wir sie durch die Bernoullische Streuung μ_B^2 dividieren:

$$\underline{Q_k^2 = \frac{\sigma_k^2}{\mu_B^2}}$$

Es ist also:

$$EQ_k^2 = \frac{\mu^2}{\mu_B^2}$$

$$= 1\,, \text{ wenn } \mu^2 = \mu_B^2\,.$$

Es sei noch einmal ausdrücklich daran erinnert, dass in diese Divergenzquotienten nur die unwesentliche Schwankungskomponente eingeht und dass diese vollständig unabhängig von der Ausgleichung berechnet werden konnte.

Jetzt wollen wir zum zweiten Fall übergehen, bei dem die Ex_i^2 nicht als gleich gross vorausgesetzt werden. Wir setzen dann

$$Ex_i^2 = \mu_i^2 = \text{a priori-Streuung von } X_i\,,$$

und es wird:

$$E\left\{\frac{1}{N-k}\sum_{i=1}^{N-k}(\Delta^k X_i)^2\right\} = \frac{1}{N-k}\sum_{i=1}^{N-k}\sum_{j=0}^{k}\binom{k}{j}^2\mu_{i+k-j}^2$$

$$= \frac{1}{N-k}\left\{\left|\binom{k}{0}^2\right|\mu_1^2 + \left|\binom{k}{0}^2 + \binom{k}{1}^2\right|\mu_2^2 + \ldots +\right.$$

$$+ \left|\binom{k}{0}^2 + \binom{k}{1}^2 + \ldots + \binom{k}{k-1}^2\right|\mu_k^2 +$$

$$+ \left|\binom{2k}{k}\right|(\mu_{k+1}^2 + \ldots + \mu_{N-k}^2) +$$

$$+ \left|\binom{k}{0}^2 + \binom{k}{1}^2 + \ldots + \binom{k}{k-1}^2\right|\mu_{N-k+1}^2 + \ldots +$$

$$\left. + \left|\binom{k}{0}^2 + \binom{k}{1}^2\right|\mu_{N-1}^2 + \left|\binom{k}{0}^2\right|\mu_N^2\right\}.$$

Dividieren wir noch beide Seiten durch die Summe der in den eckigen Klammern stehenden Faktoren: $(N-k)\binom{2k}{k}$, so steht rechts ein gewogenes Mittel der μ_i^2, das wir mit $\overline{\mu_{(k)}^2}$ bezeichnen. Darin erhalten fast alle μ_i^2 dasselbe Gewicht $\binom{2k}{k}$, nur bei den ersten und letzten k von ihnen ist es kleiner. In der nachstehenden Tabelle sind diese Gewichte für die ersten 10 Differenzenreihen zusammengestellt:

$k =$	1	2	3	4	5	6	7	8	9	10
$\binom{k}{0}^2 =$	1	1	1	1	1	1	1	1	1	1
$\binom{k}{0}^2+\binom{k}{1}^2 =$	2	5	10	17	26	37	50	65	82	101
$\binom{k}{0}^2+\binom{k}{1}^2+\binom{k}{2}^2 =$		6	19	53	126	262	491	849	1378	2126
$\binom{k}{0}^2+\binom{k}{1}^2+\binom{k}{2}^2+\binom{k}{3}^2 =$			20	69	226	662	1716	3985	8434	16526
$\binom{k}{0}^2+\binom{k}{1}^2+\binom{k}{2}^2+\binom{k}{3}^2+\binom{k}{4}^2 =$				70	251	887	2941	8885	24310	60626
$\binom{k}{0}^2+\binom{k}{1}^2+\binom{k}{2}^2+\binom{k}{3}^2+\binom{k}{4}^2+\binom{k}{5}^2 =$					252	923	3382	12021	40186	124130
$\binom{k}{0}^2+\binom{k}{1}^2+\binom{k}{2}^2+\binom{k}{3}^2+\binom{k}{4}^2+\binom{k}{5}^2+\binom{k}{6}^2 =$						924	3431	12805	47242	168230
$\binom{k}{0}^2+\binom{k}{1}^2+\binom{k}{2}^2+\binom{k}{3}^2+\binom{k}{4}^2+\binom{k}{5}^2+\binom{k}{6}^2+\binom{k}{7}^2 =$							3432	12869	48538	182630
$\binom{k}{0}^2+\binom{k}{1}^2+\binom{k}{2}^2+\binom{k}{3}^2+\binom{k}{4}^2+\binom{k}{5}^2+\binom{k}{6}^2+\binom{k}{7}^2+\binom{k}{8}^2 =$								12870	48619	184655
$\binom{k}{0}^2+\binom{k}{1}^2+\binom{k}{2}^2+\binom{k}{3}^2+\binom{k}{4}^2+\binom{k}{5}^2+\binom{k}{6}^2+\binom{k}{7}^2+\binom{k}{8}^2+\binom{k}{9}^2 =$									48620	184755
$\binom{k}{0}^2+\binom{k}{1}^2+\binom{k}{2}^2+\binom{k}{3}^2+\binom{k}{4}^2+\binom{k}{5}^2+\binom{k}{6}^2+\binom{k}{7}^2+\binom{k}{8}^2+\binom{k}{9}^2+\binom{k}{10}^2 =$										184756

Jeweils die unterste Zahl jeder Kolonne ist $\binom{2k}{k}$, also dasjenige Gewicht, welches fast allen μ_i^2 zukommt.

Setzen wir wiederum:

$$\sigma_k^2 = \frac{1}{\binom{2k}{k}(N-k)} \sum_{i=1}^{N-k} (\Delta^k X_i)^2$$

so wird jetzt:

$$\underline{E\sigma_k^2 = \overline{\mu_{(k)}^2} \quad \text{für} \quad k \geq k_0 .}$$

Auch in diesem Fall können wir zu jeder Differenzenreihe einen Divergenzquotienten bestimmen, indem wir durch den in gleicher Weise gebildeten gewogenen Mittelwert $\overline{\mu_{B,(k)}^2}$ der Bernoullischen Streuungen $\overline{\mu_{i,B}^2}$ dividieren:

$$\underline{Q_k^2 = \frac{\sigma_k^2}{\overline{\mu_{B,(k)}^2}} .}$$

Dabei ist also

$$\overline{\mu_{B,(k)}^2} = \sum_{i=1}^{N} g_{ik} \mu_{i,B}^2$$

wo

$$g_{ik} = \begin{cases} \sum_{j=0}^{i-1} \binom{k}{j}^2 & \text{wenn} \quad 1 \leq i \leq k \\ \sum_{j=0}^{N-i} \binom{k}{j}^2 & \text{wenn} \quad N-k+1 < i \leq N \\ \binom{2k}{k} & \text{sonst} . \end{cases}$$

Die Zahlen g_{ik} lassen sich aus obiger Tabelle entnehmen.

Infolge der wechselnden Gewichte ändern sich die Erwartungswerte von σ_k^2 und Q_k^2 von Differenzenreihe zu Differenzenreihe etwas. Nur wenn alle μ_i^2 mit den Bernoullischen Streuungen $\mu_{i,B}^2$ identisch sind, wird derjenige von Q_k^2 von k unabhängig, und zwar gleich 1.

3. Anwendung auf die Berechnung der Streuung in der x-Richtung

Diese Methode wollen wir nun verwenden, um die Dispersion der empirischen Sterbehäufigkeiten in der x-Richtung zu berechnen. Wir ersetzen also die X_i durch die empirischen Sterbehäufigkeiten q_x. Der variable Parameter ist hier das Alter x; die m_i entsprechen den wahren Sterblichkeiten.

Wenn für das Alter x bei L_x unter Risiko stehenden Personen T_x Todesfälle beobachtet wurden, dann beträgt nach der Bayesschen Formel der Erwartungswert für die zugrunde liegende Wahrscheinlichkeit:

$$q_x = \frac{T_x + 1}{L_x + 2} \text{ [1]}.$$

Die a priori-Streuung um diesen Erwartungswert ist unter den gleichen Annahmen

$$\mu^2_{x,B} = \frac{1}{L_x + 3} \cdot q_x \cdot (1 - q_x).$$

Es liegt also der zweite Fall vor, wo die Streuung für jedes Reihenglied verschieden ist.

Als *Beobachtungsmaterial* diente:

1. «Die Erfahrungen der Schweizerischen Lebensversicherungs- und Rentenanstalt im Hauptgeschäft, Alter 3—71», entnommen der Jubiläumsschrift «Fünfundsiebzig Jahre Schweizerische Lebensversicherungs- und Rentenanstalt» (23), also die Zahlen, die der Tafel RAH $\frac{1894/1930}{1921/1931}$ zugrunde liegen.

[1]) Es ist zwar allgemein üblich, hier nicht mit dem Erwartungswert, sondern mit dem wahrscheinlichsten Wert $\frac{T_x}{L_x}$ zu rechnen, obwohl meines Erachtens für die Verwendung des erstern die stärkern Gründe sprechen. Vor allem hat der Erwartungswert den Vorzug, dass die Streuung um ihn zu einem Minimum wird und zudem durch eine sehr einfache Formel dargestellt werden kann (s. Text). Der Erwartungswert liefert ausserdem in unserm Fall gerade dann sicher die bessern Zahlen, wenn der Unterschied ins Gewicht fällt, nämlich wenn T_x sehr klein oder gar null ist. Vgl. Czuber-Burckhardt (2), S. 309, oder Gebelein (5).

2. «Die Erfahrungen der Rentenanstalt in der Volksversicherung, Alter 3—71», ebenfalls entnommen der Jubiläumsschrift, also das Urmaterial der Tafel RAV $\frac{1894/1930}{1921/1931}$.

3. «Die Erfahrungen der Basler Lebens-Versicherungs-Gesellschaft in der Volksversicherung, Alter 3—69», entnommen der Jubiläumsschrift «Fünfundsiebzig Jahre Basler Lebens-Versicherungs-Gesellschaft», also die Zahlen, aus denen die Tafel BaV $\frac{1900—1936}{1927—1937}$ berechnet wurde.

Bei den Tafeln der Rentenanstalt wurden die ersten und letzten paar Jahrgänge weggelassen, weil dort die Streuung infolge der kleinen Besetzungszahlen so gross ist, dass die Zahlen der übrigen Jahrgänge bei der Mittelbildung vollständig überdeckt würden. Bei der Tafel der Basler wurden dann aus Gründen der Vergleichbarkeit die ersten drei Jahrgänge ebenfalls gestrichen.

In der folgenden Tabelle sind die Werte für $Q_k = \sqrt{\frac{\sigma_k^2}{\mu_{B,(k)}^2}}$ für die ersten vier Differenzenreihen zusammengestellt. Es war nicht nötig, weiterzugehen, weil schon hier die Werte von Q_k genügend konstant sind.

	Q_1	Q_2	Q_3	Q_4
RAH. .	1,181	1,099	1,115	1,115
RAV. .	1,151	1,158	1,397	1,285
BaV . .	1,296	1,133	1,198	1,185

Diese Zahlen zeigen mit sehr guter Annäherung normale Dispersion. Überraschenderweise sind sie schon für $k = 1$ nicht viel grösser als 1 und nehmen auch nachher nicht mehr wesentlich ab oder sogar leicht zu. Dies zeigt, dass die Komponente m_i schon in der ersten Differenzenreihe zu einem guten Teil eliminiert ist. Das bedeutet nicht, dass die m_i auf einer Geraden zu liegen brauchen, sondern nur, dass sie sich von Schritt zu Schritt so wenig ändern, dass die Δm_i im Vergleich zur Streuung der x_i sehr klein sind.

Es wäre nun möglich, dass die normale Dispersion so zustande kommt, dass die Streuung in einem Teil der Sterbetafel grösser als 1,

in einem andern aber kleiner als 1 ist. Um darüber Klarheit zu erhalten, wollen wir die Reihe der q_x unterteilen und die Dispersion für jeden der Teile gesondert berechnen. Allzu weit dürfen wir diese Unterteilung aber nicht treiben, weil sonst die einzelnen Teilreihen so wenig Reihenglieder erhalten, dass man keine zuverlässigen Resultate mehr erwarten kann. Es wurden deshalb drei Gruppen gebildet, so dass jede von ihnen 23 bzw. 21 Glieder aufweist: eine erste vom Alter 3 bis 25, die zweite vom Alter 26 bis 48 und die dritte vom Alter 49 bis 71 bzw. 69. Die Ergebnisse finden sich in der nachstehenden Tabelle:

	Q_1	Q_2	Q_3	Q_4
RAH 3—25 . .	0,922	0,813	0,645	0,599
26—48 . .	1,304	1,039	0,922	0,853
49—71 . .	1,184	1,102	1,118	1,119
RAV 3—25 . .	0,865	0,659	0,667	0,688
26—48 . .	0,873	0,741	0,716	0,706
49—71 . .	1,154	1,162	1,240	1,291
BaV 3—25 . .	1,557	1,153	1,084	1,057
26—48 . .	1,407	1,331	1,322	1,314
49—69 . .	1,281	1,116	1,178	1,158

Auch hier ist die Streuung annähernd normal, obwohl wegen der viel kleinern Zahl der Reihenglieder von vornherein weniger genaue Ergebnisse erwartet werden konnten. Einzig die unternormale Dispersion in der 1. Altersgruppe der beiden Tafeln der Rentenanstalt fällt etwas auf, vor allem weil unternormale Dispersion in der Statistik sonst nur selten beobachtet wird. Aber die Abweichung gegenüber der normalen Dispersion ist immer noch kleiner als der doppelte mittlere Fehler [1]), so dass es sich auch um eine zufällige Abweichung handeln kann.

[1]) Dies zeigen die Andersonschen Formeln für die mittleren Fehler von σ_k^2 (S. 154), die sich zwar nur auf den Fall beziehen, wo alle Reihenglieder dieselbe a priori-Streuung besitzen, die aber zu einer groben Abschätzung auch hier verwendet werden mögen.

Wir können somit im ganzen feststellen, dass unser Urnenschema auch in der x-Richtung durch unsere Rechnungen in sehr befriedigender Weise gerechtfertigt wurde.

Auf Grund dieser Resultate scheint es also, dass die für die Risikotheorie wichtige Frage, ob ein Ausgleich der Sterblichkeitsschwankungen nicht nur im Verlaufe der Zeit, sondern auch innerhalb eines Versicherungsjahres durch die verschiedenen Altersklassen hindurch zu erwarten ist, in positivem Sinne beantwortet werden darf.

III. Ein allgemeines Ausgleichungsprinzip

Entsprechend unserer Zielsetzung, den Wahrscheinlichkeitscharakter der Sterblichkeit zu untersuchen, haben wir uns im II. Kapitel eingehend mit der Schwankungskomponente der relativen Sterbehäufigkeiten beschäftigt. Von der wahren Komponente m_i interessierte uns dabei nur die Frage, wie sie zu eliminieren sei. Das ist uns mit Hilfe der Differenzenmethode gelungen, ohne dass wir sie in irgendeiner speziellen Form, z. B. des Makehamschen Gesetzes, anzunehmen brauchten, ja ohne dass wir sie überhaupt berechnen mussten. In diesem Schlusskapitel wollen wir nun noch kurz untersuchen, was für Folgerungen sich aus unseren Überlegungen für die Berechnung der m_i ergeben.

Die Berechnung der wahren Komponente ist Aufgabe der Ausgleichungsrechnung, welche zu diesem Zwecke in ihrer Rüstkammer eine unübersehbare Zahl verschiedener Methoden bereithält. Allen wohnt ein ziemlich grosses Mass von Willkür inne, teils weil zu ihrer Begründung Voraussetzungen oder Annahmen gemacht werden, die durchaus nicht zwingend, ja oft nicht einmal plausibel scheinen, teils weil sie überhaupt auf jede Begründung verzichten und nichts als rein handwerkliche Regeln darstellen. Damit sollen diese Methoden keineswegs verurteilt werden; sie können alle zu sehr guten Resultaten führen. Die Schwierigkeit liegt gerade darin, zu urteilen, ob eine vorgelegte Ausgleichung gut ist und wie gut sie ist, d. h. ein Mass für ihre Güte zu finden.

Man hat schon oft behauptet, eine solche Beurteilung sei deshalb nicht in eindeutiger Weise möglich, weil die Aufgabe, eine gegebene empirische Reihe möglichst gut auszugleichen, keine bestimmte Lösung

besitze. Im folgenden soll gezeigt werden, dass es im Gegensatz zu dieser Behauptung eine beste Ausgleichung einer gegebenen Reihe wirklich gibt. Es wird sich dabei allerdings zeigen, dass die numerische Berechnung dieser Ausgleichung äusserst langwierig wäre, so dass sie vielleicht praktisch nicht in Frage kommt, wenigstens nicht in dieser Form. Es ist deshalb auch nicht zu befürchten, dass all die vielen bisher entwickelten Ausgleichungsmethoden nun zum alten Eisen geworfen werden müssten. Trotzdem stellt dieses allgemeine Ausgleichungsprinzip mehr als ein blosses Existenztheorem dar, das nur akademisches Interesse bietet; denn es zeigt uns, worauf wir bei einer Ausgleichung zu achten haben, und gestattet ein Urteil über die Güte einer vorgelegten Ausgleichung.

Bevor wir das Prinzip aufstellen können, durch das die gesuchte beste Ausgleichung definiert wird, müssen wir uns darüber klar sein, was wir von einer guten Ausgleichung verlangen.

An die Spitze stellen müssen wir die Tatsache, dass bei allen praktisch in Betracht fallenden Verteilungsgesetzen die empirisch gefundenen unausgeglichenen Werte die besten sind, wenn wir sie einzig und allein vom Standpunkt ihrer Wahrscheinlichkeit aus betrachten. Bei der Ausgleichung werden sie durch schlechtere, d. h. weniger wahrscheinliche ersetzt, und zwar sind sie um so schlechter, je weiter entfernt sie vom beobachteten Wert sind. Von unserer besten Ausgleichung müssen wir deshalb in erster Linie verlangen:

1. Die ausgeglichenen Werte sollen möglichst nahe bei den unausgeglichenen liegen.

Um für die Abweichung der ausgeglichenen Grössen von den unausgeglichenen ein Mass zu bekommen, denken wir uns eine Wahrscheinlichkeitsdichte, welche zu jedem möglichen Wertesystem $(m_1, m_2, \ldots, m_N)$ dessen Wahrscheinlichkeit angibt:

$W(m_1, \ldots, m_N)\, dm_1 \ldots dm_N =$ Wahrscheinlichkeit dafür, dass gleichzeitig der richtige Wert von X_1 zwischen m_1 und $m_1 + dm_1$, von X_2 zwischen m_2 und $m_2 + dm_2$,

. . . . . .

und von X_N zwischen m_N und $m_N + dm_N$ liegt.

Falls die einzelnen Reihenglieder voneinander unabhängig sind, hat diese Funktion die Form

$$W(m_i, \ldots, m_N) = \prod_{i=1}^{N} w_i(m_i)$$

Wenn die x_i ausserdem noch den Charakter von Wahrscheinlichkeitsschwankungen haben, wie wir es für die Sterblichkeitsschwankungen festgestellt haben, können wir für die w_i Gauss-Verteilungen einsetzen.

Wesentlich ist nun, welche Wertesysteme $(m_1, \ldots, m_N)$ wir zur Konkurrenz zulassen. Denn selbstverständlich dürfen wir nicht beliebige Wertesysteme erlauben, sonst wäre ja jenes der unausgeglichenen Reihe das beste. Anderseits wollen wir auch nicht zu streng sein und etwa nur Systeme gestatten, deren Werte auf einer Geraden oder auf einer Parabel bestimmter Ordnung oder sonst auf irgendeiner Kurve mit vorgegebener analytischer Gleichung

$$m_i = f(i)$$

liegen, wie das bei der Methode der kleinsten Quadrate gemacht wird. (Es sei denn, man wüsste aus theoretischen Gründen zum vornherein, dass die Werte auf einer solchen Kurve liegen müssen, welcher Fall ausserhalb der Physik und Astronomie nur selten vorkommt.)

Um zur richtigen einschränkenden Bedingung zu kommen, brauchen wir uns nur zu überlegen, weshalb wir überhaupt ausgleichen. Die Ausgleichung bezweckt doch, die Ausschaltung der zufälligen Schwankungen der beobachteten Werte, die ein so unregelmässiges Bild ergeben, das unserer Erwartung von einem glatten Verlauf der Kurve nicht entspricht. Wir fordern deshalb:

2. Die ausgeglichene Reihe soll keine Zufallskomponente mehr enthalten.

Woran sollen wir aber erkennen, ob bei einem Wertesystem die Zufallskomponente ausgeschaltet ist oder nicht? Hier hilft uns wieder die Differenzenmethode. Wie wir bemerkt haben, zeichnen sich die glatten Funktionen dadurch aus, dass sie beim endlichen Differenzieren langsam verschwinden, während die Differenzen zufälliger Reihen mit steigender Ordnung wie $\binom{2k}{k}$ unbegrenzt wachsen. Diese Eigenschaft

können wir verwenden, um Wertesysteme, die unserer Auffassung von einem glatten Verlauf der Kurve widersprechen, auszuschliessen, indem wir nur diejenigen zulassen, deren Differenzen steigender Ordnung absolut nicht zunehmen.

Unsere gesuchte beste Ausgleichung ergibt sich somit als Lösung der folgenden Extremalaufgabe:

Die Funktion $W(m_1, m_2, \ldots, m_N)$ ist durch geeignete Wahl der Werte $m_1, m_2, \ldots, m_N$ zu einem Maximum zu machen, unter Berücksichtigung der folgenden Nebenbedingungen:

$$\left.\begin{array}{c} (\Delta^{k+1} m_i)^2 \leqq (\Delta^k m_i)^2 \\ \text{oder} \\ (\Delta^{k+1} m_i)^2 \leqq (\Delta^k m_{i+1})^2 \end{array}\right\} \text{ für alle } i \text{ und } k.$$

Deuten wir die Variabeln $m_1, m_2, \ldots, m_N$ als Koordinaten in einem N-dimensionalen Raum, dann wird durch die Nebenbedingungen ein abgeschlossener, aber nicht beschränkter Teilbereich definiert, den wir mit Ω bezeichnen. Wir haben also das Maximum der Ortsfunktion W in Ω zu bestimmen.

Um die Existenz eines solchen Maximums beweisen zu können, müssen wir die Funktion $W(m_1, \ldots, m_N)$ als stetig in allen Variabeln voraussetzen. Wegen ihrer Eigenschaft als Wahrscheinlichkeitsdichte gilt ferner

$$\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} W(m_1, \ldots, m_N)\, dm_1 \ldots dm_N = 1.$$

Das ist nur möglich, wenn W in allen Richtungen gegen Null strebt, und zwar ist diese Konvergenz, wie man leicht einsieht, gleichmässig. Zu jedem $\varepsilon > 0$ gibt es also ein R, so dass die Funktion W ausserhalb der N-dimensionalen Kugel vom Radius R überall kleiner ist als ε.

Es sei nun $P^0 = (m_1^0, m_2^0, \ldots, m_N^0)$ irgendein Punkt aus Ω und W^0 der Wert von W in diesem Punkt. Dann können wir, wie wir eben bemerkt haben, um P^0 eine Kugel konstruieren, in deren Äussern W überall kleiner ist als W^0. Bilden wir den Durchschnitt D dieser Kugel mit dem Bereich Ω, dann muss also das Maximum der Funktion W in Ω sicher in diesem Durchschnitt liegen, vorausgesetzt, dass es

existiert. D ist aber ein abgeschlossener und beschränkter Bereich, in dem nach dem Satz von Bolzano-Weierstrass jede stetige Funktion ihren Maximalwert wirklich annimmt. (Unter gewissen einschränkenden Voraussetzungen über die Funktion W lässt sich sogar behaupten, dass dieses Maximum auf dem Rande von Ω angenommen wird, was uns aber hier nicht interessiert.)

Damit ist die Existenz einer besten Ausgleichung bewiesen. Sie ist allerdings noch nicht ganz eindeutig bestimmt; denn W kann ja seinen Maximalwert an mehreren Stellen von Ω annehmen. Um auch diese Mehrdeutigkeit noch auszuschalten, kann man unter allen Wertesystemen, welche zum gleichen maximalen W gehören, dasjenige auswählen, dessen Differenzen am raschesten abnehmen.

Der tatsächlichen Berechnung dieser besten Ausgleichung stellen sich allerdings infolge der grossen Zahl von Variabeln und Nebenbedingungen fast unüberwindliche praktische Schwierigkeiten entgegen. Trotzdem sind unsere Überlegungen nicht nur eine müssige Spielerei; denn abgesehen von ihrem theoretischen Interesse zeigen sie uns, worauf wir bei der Beurteilung einer Ausgleichung zu achten haben:

1. ist zu untersuchen, ob die Zufallskomponente eliminiert ist. Dazu müssen wir die Differenzenreihen bilden und verifizieren, ob die Differenzen mit zunehmender Ordnung nicht grösser werden. Man beachte dabei, dass es keinen Sinn hat, allzu viele Differenzenreihen zu berechnen, weil sich von einer gewissen Ordnung an die Auf- und Abrundungsfehler bemerkbar machen werden; denn auch sie wirken natürlich wie zufällige Fehler;

2. haben wir $W^* = W(m_1^*, \ldots, m_N^*)$ zu berechnen, wo $m_1^*, \ldots, m_N^*$ die ausgeglichenen Werte sind. Der absolute Wert von W^* sagt aber noch nicht viel. Wir müssen ihn mit dem W der unausgeglichenen Reihe vergleichen. Ferner können wir die Wahrscheinlichkeit dafür berechnen, dass W einen noch kleineren Wert als den gefundenen annimmt. Falls W die Form

$$W(m_1, \ldots, m_N) = \prod_{i=1}^{N} w_i(m_i)$$

hat und die $w_i(m_i)$ Gauss-Verteilungen sind, ist dies das bekannte χ^2-Kriterium. Korrekterweise müsste man auch hier sich auf den Bereich Ω beschränken, d. h. man hätte zu bilden

$$\frac{\int\limits_{\Omega^*} \dots \int W(m_1, \dots, m_N) \cdot dm_1 \dots dm_N}{\int\limits_{\Omega} \dots \int W(m_1, \dots, m_N) \cdot dm_1 \dots dm_N}.$$

Dabei ist Ω^* der Durchschnitt von Ω mit jenem Teil des Raumes, in dem W kleiner ist als der beobachtete Wert W^*. Tatsächlich scheitert man auch hier an der praktischen Undurchführbarkeit und ist deshalb gezwungen, das Integral im Nenner über den ganzen Raum und jenes im Zähler über den ganzen Bereich $W < W^*$ zu ersetzen. (Falls man, wie das sehr oft auch gemacht wird, die ausgeglichene und nicht die beobachtete Reihe als die primäre ansieht, ist das sogar korrekt.)

Die Literatur kennt noch andere Kriterien für die Güte der Ausgleichung, wie das Helmertsche Kriterium für die Anzahl der Zeichenwechsel und Zeichenfolgen [1]) u. ä. Diese hangen aber nicht so sehr von der Ausgleichung als von der Art der Schwankungen der empirischen Werte ab und lassen deshalb weniger ein Urteil über die Güte der Ausgleichung als über den Charakter der Schwankungskomponente, also der x_i, zu.

---

[1]) F. R. Helmert: Über die Genauigkeit der Kriterien des Zufalls bei Beobachtungsreihen (Sitzungsberichte der königlichen preussischen Akademie der Wissenschaften, 1905).

Literaturverzeichnis

(MVSV = Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker.)

1. E. Czuber: Wahrscheinlichkeitsrechnung und ihre Anwendung auf Fehlerausgleichung, Statistik und Lebensversicherung (Leipzig und Berlin 1928).
2. Czuber-Burkhardt: Die statistischen Forschungsmethoden (Wien 1938).
3. E. Blaschke: Vorlesungen über mathematische Statistik (Leipzig und Berlin 1906).
4. O. Anderson: Einführung in die mathematische Statistik (Wien 1935).
5. H. Gebelein: Zahl und Wirklichkeit (Leipzig 1943).
6. Berichte des XII. Internationalen Kongresses der Versicherungsmathematiker, Band I (Luzern 1940).
7. H. Jecklin: Die Wahrscheinlichkeitstheorie im Versicherungswesen (MVSV, Bd. 41, 1941).
8. P. Nolfi: Versicherungsmathematik und Wirklichkeit (MVSV, Bd. 37, 1939).
9. H. Wiesler: Über die Grundlagen der Lebensversicherungsmathematik (MVSV, Bd. 44, 1944).
10. Curt Lange: Untersuchungen über die jährlichen Schwankungen der Schadensquotienten in der Lebensversicherung und in der Feuerversicherung (Wirtschaft und Recht der Versicherung, Nr. 2, 1932).
11. H. Ammeter: Untersuchungen über die jährlichen Sterblichkeitsschwankungen in einem Versicherungsbestand (MVSV, Bd. 45, 1945).
12. H. Braun: Der relative Wert der Vergleichung der wirklichen Sterblichkeit oder Invalidität mit der wahrscheinlichen (Blätter für Versicherungsmathematik, 6. Bd., 1. Heft, 1944).
13. «Student»: The Elimination of Spurious Correlation due to Position in Time or Space (Biometrika, Bd. X).
14. O. Anderson: Nochmals über «The Elimination of Spurious Correlation due to Position in Time or Space» (Biometrika, Bd. X).
15. E. M. Elderton und K. Pearson: Further Evidence of Natural Selection in Man (Biometrika, Bd. X).
16. E. M. Elderton und K. Pearson: On the Variate Difference Method (Biometrika, Bd. XIV).
17. O. Anderson: Über ein neues Verfahren bei Anwendung der «Variate Difference Method» (Biometrika, Bd. XV).

18. O. Anderson: Über die Anwendung der Differenzenmethode («Variate Difference Method») bei Reihenausgleichungen, Stabilitätsuntersuchungen und Korrelationsmessungen (Biometrika, Bd. XVIII und XIX).

19. O. Anderson: Die Korrelationsrechnung in der Konjunkturforschung (Heft 4 der Veröffentlichungen der Frankfurter Gesellschaft für Konjunkturforschung, Bonn 1929).

20. J. Simonett: Beiträge zur Ausgleichung von Massenerscheinungen nach der Methode von King (MVSV, Bd. 28, 1933).

21. W. Schöb: Ein Beitrag zur Theorie der Ausgleichungsrechnung (MVSV, Bd. 30, 1935).

22. W. Ruchti: Analytische Ausgleichung durch Polynome (MVSV, Bd. 34, 1937).

23. Fünfundsiebzig Jahre Schweizerische Lebens-Versicherungs- und Rentenanstalt, 1857—1932.

24. Fünfundsiebzig Jahre Basler Lebens-Versicherungs-Gesellschaft Basel, 1864 bis 1939.

Inhaltsverzeichnis

Seite

I. Wahrscheinlichkeitstheorie und Versicherungsmathematik 131
1. Die Stellung der Wahrscheinlichkeitstheorie im allgemeinen . . . 131
2. Die Wahrscheinlichkeitstheorie in der Versicherungsmathematik . . 133
a) Der absolute Empirismus 134
b) Der bedingte Empirismus 135

II. Die Streuung der relativen Sterbehäufigkeiten 138
A. Allgemeines 138
1. Die Sterbefläche 138
2. Das Urnenschema 139
3. Die Streuung in der t-Richtung 140
4. Die Streuung in der x-Richtung 143
5. Die Ursachen der Sterblichkeitsschwankungen 144
B. Die Berechnung der Streuung in der x-Richtung 147
1. Das Problem 147
2. Die Differenzenmethode (Variate Difference Method) 148
3. Anwendung auf die Berechnung der Streuung in der x-Richtung 158

III. Ein allgemeines Ausgleichungsprinzip 161