Stochastik

Barth, Friedrich

München, [20]03

17. 1. Zur Geschichte und Aufgabe der Statistik

17. Das Testen von Hypothesen

17.1. Zur Geschichte und Aufgabe der Statistik

Στοχαστικὴ τέχνη, *Stochastik*, ins Lateinische übersetzt *ars conjectandi* (der Titel von *Jakob Bernoulli*s Buch über unseren Gegenstand) ist die Kunst, im Falle von Ungewißheit auf geschickte Weise Vermutungen anzustellen. Ursprünglich entwickelte sich die Stochastik aus dem Bedürfnis, die Gewinnchancen bei Glücksspielen in den Griff zu bekommen (Seite 71ff.). Wenn dieser Gesichtspunkt auch heute noch interessant ist, so würde er allein es doch kaum rechtfertigen, daß Stochastik in der Schule gelehrt wird! Die wichtigste Anwendung findet die »Kunst des Vermutens« heute als *mathematische Statistik* in allen Zweigen der Wirtschaft, der Technik, der Politik und der Wissenschaften.

Verstand man Statistik schon immer in diesem Sinn? Nein; denn die mathematische Statistik entstand erst in diesem Jahrhundert und speist sich aus mehreren geschichtlichen Quellen.
Am Anfang steht die *Amtliche Statistik*, die bevölkerungsstatistischen Erhebungen. Überliefert sind uns Volkszählungen aus dem Alten Reich der Ägypter (um 2600 v. Chr.) und aus China. Der 6. König Roms, *Servius Tullius* (Regierungszeit 577–534), bestimmte in seiner Verfassung, alle 5 Jahre den census durchzuführen, eine Volkszählung, verbunden mit einer Erhebung über die Vermögensverhältnisse der Bürger und einer Einteilung für den Waffendienst. Eine solche Einteilung in Zensusklassen war auch in Griechenland üblich. Unter dem im Jahre 27 v. Chr. von *Augustus* (63 v. Chr. bis 14 n. Chr.) eingerichteten Prinzipat fanden die ersten Volkszählungen in den Provinzen des Römischen Reichs statt, so 27 v. Chr. in Gallien und 14 n. Chr. in Germanien. Die berühmteste Volkszählung ist wohl jener Provinzialcensus, der in Judäa im Jahre 6 n. Chr. durchgeführt wurde, als es römische Provinz wurde, und den *Lukas* in seinem Evangelium (2,1) irrtümlich für eine Reichszählung hält. Der 70. und letzte census fand 73 n. Chr. statt. Aber schon aus dem Alten Testament sind Volkszählungen bekannt. So künden das 2. Buch Mose (30,11) und das 4. Buch Mose (1), das auf lateinisch bezeichnenderweise *numeri* heißt, von einer von Gott angeordneten Volkszählung (um 1200 v. Chr.). König *David* (1004–965) hingegen verführte der Satan zu einer Volkszählung, wie im 2. Buch Samuel (24,2) und im 1. Buch der Chronik (21,2) berichtet wird. Für diesen Fürwitz wurde das Volk Israel mit der Pest bestraft. *Helmut Swoboda* meint:

> »Diese biblische Warnung bestimmte bis in die Neuzeit das Verhältnis zur statistischen Erhebung: Es war zweifellos sträfliche Neugier oder vorwitzige Vermessenheit, durch Volkszählungen oder gar durch systematische Beobachtungen von Geburten, Krankheiten und Todesfällen in die unerforschlichen Absichten Gottes Einsicht nehmen zu wollen.«

Die mittelalterlichen Erhebungen wie das karolingische *Capitulare de villis* und das *Domesday Book* (1086) *Wilhelms des Eroberers* sind daher fast ausschließlich Vermögens- oder Ständeerhebungen, so auch 1250 in Asti und 1288 in Mailand. Alle Seelen zählte erstmals Venedig 1422, und 1444 Straßburg, als eine Belagerung drohte. 1449 tat es Nürnberg.
Die Erweiterung des geographischen Horizonts, die wachsende Verflechtung der Staaten untereinander und die Ausweitung der Wirtschaftsbeziehungen zu Beginn der Neuzeit ließen eine weitere Quelle der heutigen Statistik entstehen, die *Staatskunde* als *Lehre von den Staatsmerkwürdigkeiten*, auch *Universitätsstatistik* genannt. So ist *Francesco Sansovinos* (1521–1586) Werk *Del governo et amministratione di diversi regni, et republiche*, [...] (1562) eine Sammlung von Staatsbeschreibungen. *Hermann Conring* (1606–1681) führte diese beschreibende Staatswissen-

schaft als Lehrfach an der Universität Helmstedt ein. *Gottfried Achenwall* (1719–1772) führte *Conring*s Arbeiten in Göttingen weiter. In seiner *Staatsverfassung** definierte er 1748 das Wort »Statistik« im Sinne von Staatskunde, wohl durch Rückgriff auf den lateinischen Begriff des *status rei publicae*, des Zustands des Staates. Er schreibt:

> »Der Inbegriff der wirklichen Staatsmerkwürdigkeiten eines Reichs, oder einer Republik, macht ihre Staatsverfassung im weitern Verstande aus: und die Lehre von der Staatsverfassung eines oder mehrerer einzelner Staaten, ist die Statistik [Staatskunde], oder Staatsbeschreibung.«

Er grenzt Statistik gegen die philosophische Staatslehre und gegen das Staatsrecht ab. Sein Schüler *August Ludwig von Schlözer* (1735–1809) in Göttingen und *Anton Friedrich Büsching* (1724–1793) in Berlin waren bedeutende Vertreter dieser Statistik.

Die dritte Quelle der modernen Statistik, die *Bevölkerungsstatistik* oder *Politische Arithmetik*, entsprang in England. Der Tuchhändler *John Graunt* (1620–1674) legte die Sterbelisten der Stadt London, beginnend mit dem Jahre 1603, seiner 1662 erschienenen Studie *Natural and Political Observations, mentioned in a Following Index, and Made upon the Bills of Mortality* zugrunde, dem ersten Werk über Bevölkerungsstatistik. Er wurde zum Begründer der Biometrie und der Bevölkerungsstatistik, die der Nationalökonom Sir *William Petty* (1623–1687) *Politische Arithmetik* nannte. Man sammelte bevölkerungsstatistische Massentatsachen und fragte nach ihren Ursachen und Regelmäßigkeiten. Eine erste Anwendung fanden solche Untersuchungen in der Ermittlung der Prämien für Leibrenten mittels einer Statistischen Mortalitätstheorie durch *Edmond Halley*** (1656–1742) in *An Estimate of the Degrees of Mortality of Mankind, drawn from curious Tables of the Births and Funerals at the City of Breslaw; with an Attempt to ascertain the Price of Annuities upon Lives* (1693). Sein Freund *Abraham de Moivre* (1667–1754) führte diese Untersuchungen weiter in seinen *Annuities on lives* (1725, 1743, 1750, 1752). *John Arbuthnot* (1667–1735) versuchte 1710 einen mathematischen Gottesbeweis auf statistischer Grundlage, ausgehend von der Tatsache der zahlenmäßigen Gleichheit der Geschlechter, obwohl in den letzten 82 Jahren in London fast konstant 18 Knabengeburten auf 17 Mädchengeburten kamen. Der bekannteste Vertreter der Politischen Arithmetik ist *Thomas Robert Malthus* (1766–1834). In Deutschland findet die Politische Arithmetik durch die Leistungen des Feldpredigers und späteren Oberkonsistorialrats *Johann Peter Süßmilch* (1707–1767) Anerkennung. Bevölkerungsstatistik dient auch bei ihm dem Nachweis, daß Gott die Welt weise eingerichtet hat, wie der Titel seines Werks zeigt: *Die göttliche Ordnung in den Veränderungen des menschlichen Geschlechts aus der Geburt, dem Tode und der Fortpflanzung desselben erwiesen* (1741).

Aber schon 1666 wurde die alte biblische Warnung in den Wind geschlagen; in La Nouvelle France (Quebec) fand die erste Volkszählung eines ganzen Landes in der Neuzeit statt. Deutsche Staaten begannen ab 1720 mit Volkszählungen. Schweden ordnete als erstes Land der Neuzeit 1749 regelmäßige Volkszählungen an; 1756 schuf es als erstes Land ein Statistisches Zentralamt, das sich mit der fortlaufenden Analyse der Bevölkerungszahlen beschäftigen sollte. 1790 begannen die USA mit regelmäßigen Volkszählungen, wie sie die Unionsverfassung als Grundlage für Wahlen verlangte. 1800 entstand in Paris das Bureau de Statistique, 1801 fanden erste Volkszählungen in Frankreich und Großbritannien (beschlossen bereits 1753) statt. Frankreich verwendete dabei Methoden, die *Laplace* vorgeschlagen hatte.

* Die 1. Auflage von 1748 trug den Titel *Vorbereitung zur Staatswissenschaft der Europäischen Reiche*. 1749 hieß sie dann *Abriß der Staatswissenschaft der Europäischen Reiche* und schließlich 1752 *Staatsverfassung der heutigen vornehmsten Europäischen Reiche und Völker im Grundrisse*.

** Gesprochen hæli. – In der zitierten Arbeit findet man neben analytischen Beweisen wahrscheinlichkeitstheoretischer Formeln zum ersten Mal auch geometrische Beweisverfahren, wie sie 1733 *Buffon* (1707–1788) verwendete (siehe dazu Anhang I). Als erster benützte *Newton* (1643–1727) geometrische Wahrscheinlichkeiten in einem Manuskript, geschrieben zwischen 1664 und 1666.

Die neue Amtliche Statistik, die Universitätsstatistik und die Politische Arithmetik verschmolzen im 19. Jahrhundert zur *Deskriptiven Statistik*. Diese untersucht eine Gesamtheit nach bestimmten, ihr wesenseigenen Merkmalen. Statistik in diesem Sinne ist also eine Kunst des geschickten Zählens und der Handhabung von Zählergebnissen. Von Vermutungen oder vom Zufall ist dabei nicht die Rede. Man rechnet im Gegenteil damit, daß durch das Erheben einer sehr großen Anzahl von Daten sich die Besonderheiten des Einzelfalls »herausmitteln« und dafür die allgemeinen Gesetzmäßigkeiten, der »Trend«, zutage treten.
Das Eindringen erster Vorstellungen aus der Wahrscheinlichkeitstheorie führte bei *Adolphe Quetelet* (1796–1874) zur Schaffung des statistischen Idealtyps, des *homme moyen*.* Sir *Francis Galton* (1822–1911) verfeinerte u.a. diese Begriffsbildung und begründete zusammen mit *Karl Pearson* (1857–1936) und Sir *Ronald Aylmer Fisher* (1890–1962) die biometrische Schule der Statistik.
Zu Beginn dieses Jahrhunderts zeichnete sich jedoch eine große Wende in der Statistik ab, die in den 30er Jahren zur Geburt der modernen Statistik, der *Mathematischen Statistik* oder auch der *Analytischen Statistik*, führte. Man erkannte, daß es vielfach unmöglich war, eine Gesamtheit durch eine Vollerhebung zu erfassen. Denken wir nur an die Qualitätskontrolle in der Industrie. Es wäre finanziell nicht tragbar und auch technisch oft unmöglich, *alle* Produkte einer Serienfertigung peinlich genau zu prüfen. Statt dessen schlug in den zwanziger Jahren *W.A. Shewhart* von den Bell Telephone Laboratories vor, eine *Zufallsstichprobe* von verhältnismäßig wenigen Stücken aus der laufenden Produktion zu entnehmen und diese um so sorgfältiger zu prüfen. Vom Prüfergebnis schließt man dann auf den Zustand der gesamten Ware und entscheidet, ob die Produktion weiterlaufen darf oder gestoppt werden muß. Dabei können natürlich Irrtümer vorkommen. Mit Hilfe der Mathematik ist es aber möglich, das Risiko des Irrtums zu kalkulieren und von vorneherein in gewünschten Grenzen zu halten**.
Das Ziel der Mathematischen Statistik ist also nicht mehr die *Vollerhebung*. Statt ihrer sollen *Zufallsstichproben* Aufschluß geben über die Eigenschaften der Gesamtheit; Vermutungen, sog. *statistische Hypothesen*, sollen durch Stichproben entschieden werden. Die darauf basierenden Folgerungen heißen *statistische Schlüsse*, die natürlich im Sinne der klassischen Logik nie zwingend sein können. Unter Verwendung von Methoden der Höheren Mathematik entstand eine Vielfalt von Testverfahren zur Entscheidung von Hypothesen. Die von *R.A. Fisher* und anderen begründeten Verfahren wurden von *Egon Sharpe Pearson* (1895–1980) und *Jerzy Neyman* (1894–1981) zu einer Theorie der Stichproben ausgebaut. Während des 2. Weltkriegs entwarf *Abraham Wald* (1902–1950) die *Sequentialanalyse*, die als Kriegsgeheimnis galt und erst 1947 veröffentlicht werden konnte. Nach dem Kriege entwickelte er die *statistische Entscheidungstheorie*, die es erlaubt, auch in Situationen großer Ungewißheit noch vernünftig begründbare Entscheidungen zu fällen. Und so wird Statistik heute aufgefaßt, wenngleich die Amtliche Statistik immer noch das Material für viele Entscheidungen liefern muß.

Worin unterscheidet sich nun die Mathematische Statistik von der gewöhnlichen Wahrscheinlichkeitsrechnung, die wir bisher ausgiebig betrieben haben? Wir erläutern dies am wohlvertrauten Urnenbeispiel. Die Urne enthalte schwarze und andersfarbige Kugeln. In der Wahrscheinlichkeitsrechnung gehen wir davon aus, daß der Anteil p der schwarzen Kugeln *bekannt* ist. Man betrachtet ein Zufallsexperiment und *berechnet* die Wahrscheinlichkeiten dabei auftretender Ereignisse.
Anders ist jedoch die Ausgangslage in der Statistik. Nun ist der Anteil p der schwarzen Kugeln in der Urne *unbekannt*. Man führt ein Zufallsexperiment aus –

* Den Begriff des *homme moyen* prägte *Buffon* (1707–1788) in seinem *Essai d'arithmétique morale:* »[...] *l'homme moyen*, c'est-à-dire les hommes en général, bien portans ou malades, sains ou infirmes, vigoureux ou foibles.«
** Auf die große Bedeutung der Teilerhebung wies 1895 als erster der Norweger *Anders Nicolai Kiaer* (1838–1919) hin.

es handelt sich um das Ziehen einer Stichprobe – und *schließt* auf Grund des eingetretenen Ereignisses zurück auf den Anteil p der schwarzen Kugeln. Dabei unterscheidet man zwei Situationen.

1. Das Schätzproblem. Man hat keinerlei Vermutung über den Anteil p der schwarzen Kugeln in der Urne. In diesem Fall *schätzt* man den Anteil p auf Grund des eingetretenen Ereignisses (**Hochrechnung**). Man gibt als Schätzergebnis entweder einen einzigen Wert für p an (**Punktschätzung**) oder ein ganzes Intervall, in dem p liegen soll (**Intervallschätzung**). Das so abgegebene Urteil über den Anteil p ist mit einer gewissen Unsicherheit behaftet. Die Berechnung dieses Unsicherheitsgrades ist eine der wesentlichen Aufgaben der *Beurteilenden Statistik*.

2. Das Testproblem. Man hat von vornherein gewisse Vermutungen, Hypothesen genannt, über den Anteil p der schwarzen Kugeln in der Urne. Auf Grund des eingetretenen Ereignisses wird nun *entschieden*, welche dieser Hypothesen man beibehält oder verwirft. Auch hier ist es wesentlich, sich darüber klarzuwerden, mit welcher Sicherheit ein solches Urteil ausgesprochen werden kann.

Bevor wir uns diesen beiden Problemen zuwenden, wollen wir erst den Begriff der Stichprobe klären.

17.2. Stichproben

Die Grundlage aller Anwendungen der Stochastik ist die Möglichkeit, einen Versuch unter gleichen Bedingungen mehrmals zu wiederholen. Wollen wir z. B. über die Einkommensverteilung in einer Bevölkerung Ω etwas erfahren, so nützt es so gut wie nichts, wenn wir nur von einem zufällig ausgewählten Bürger ω das Einkommen wissen. Wir müssen eine Stichprobe von mehreren Personen ziehen. Dabei muß jede Person die gleiche Chance haben, in die Stichprobe aufgenommen zu werden. Man spricht dann von einer **Zufallsstichprobe***.
Nun sei X die Zufallsgröße »Einkommen der ausgewählten Person in DM«. Sie habe die Wertemenge $\mathfrak{S} := \{x_1, \ldots, x_s\}$ und die Wahrscheinlichkeitsverteilung W mit

$$W(x_j) = \frac{\text{Zahl der Personen mit } x_j \text{ DM Einkommen}}{\text{Zahl aller Personen}} \qquad \text{für } j = 1, \ldots, s.$$

Um über W etwas zu erfahren, wählen wir n-mal eine Person aus der Gesamtbevölkerung aus. Wir erhalten als Ergebnis ein n-Tupel von Zahlen, die sämtlich der Wertemenge $\mathfrak{S}$ von X angehören. Diese n Zahlen hängen vom Zufall ab. Wir haben es also mit n verschiedenen Zufallsgrößen $X_1, \ldots, X_n$ zu tun:

$X_i :=$ »Einkommen der i-ten ausgewählten Person in DM«

mit $i = 1, \ldots, n$. Die Wertemengen aller X_i stimmen mit der von X überein; die

* Das Wort *Stichprobe* entstammt der Bergmannssprache. Die alten Schmelzöfen wurden angestochen, um die Schmelze auf ihren Zustand zu prüfen.