

Gigerenzer, Gerd

Die Evolution des statistischen Denkens

Unterrichtswissenschaft 32 (2004) 1, S. 4-22



Quellenangabe/ Reference:

Gigerenzer, Gerd: Die Evolution des statistischen Denkens - In: Unterrichtswissenschaft 32 (2004) 1, S. 4-22 - URN: urn:nbn:de:0111-opus-58059 - DOI: 10.25656/01:5805

<https://nbn-resolving.org/urn:nbn:de:0111-opus-58059>

<https://doi.org/10.25656/01:5805>

in Kooperation mit / in cooperation with:

BELTZ JUVENTA

<http://www.juventa.de>

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@djpf.de
Internet: www.pedocs.de

Digitalisiert

Mitglied der


Leibniz-Gemeinschaft

Unterrichtswissenschaft

Zeitschrift für Lernforschung
32. Jahrgang / 2004 / Heft 1

8. April 2004

Thema

Stochastisches Denken

Verantwortliche Herausgeber

Jürgen Baumert, Gerd Gigerenzer, Laura Martignon

Editorial	2
Einleitung	3
<i>Gerd Gigerenzer</i>	
Die Evolution des statistischen Denkens	4
<i>Klaus Fiedler, Henning Plessner</i>	
Die Stichprobenfalle: Lässt sich eine Sensibilität für metakognitive Probleme beim stochastischen Denken vermitteln?	23
<i>Stefan Krauss, Silke Atmaca</i>	
Wie man Schülern Einsicht in schwierige stochastische Probleme vermitteln kann. Eine Fallstudie über das „Drei-Türen-Problem“	38
<i>Christoph Wassner, Laura Martignon, Rolf Biehler</i>	
Bayesianisches Denken in der Schule	58

Die Evolution des statistischen Denkens

The Evolution of Statistic Thinking

Lernen mit Unsicherheit zu leben - statistisches Denken - ist der wichtigste Teil der Mathematik im wirklichen Leben. Denken ist das Hinterfragen von Gewissheiten, und man lernt es anhand von guten Beispielen. Zu den besten gehören jene Probleme, welche die Entwicklung des statistischen Denkens tatsächlich geprägt haben. Genau dies ist das Programm meines Artikels.

Learning to live with uncertainty - statistical thinking - is the most important aspect of mathematics for everyday life. To think means to call certainties into question; it can be learned through good examples, the best of which are problems that have actually influenced the historical development of statistical thinking. These are the focus of my article.

Der Beginn der mathematischen Theorie der Wahrscheinlichkeit wird auf 1654 datiert. Anders als die meisten großen Ideen, die bereits in der griechischen Antike entwickelt worden sind, ist das Konzept der mathematischen Wahrscheinlichkeit eine ungewöhnlich späte Entdeckung. Der Philosoph Ian Hacking hat dies als den „Skandal der Philosophie“ bezeichnet. Die Geschichte der Wahrscheinlichkeit ist also relativ kurz, und sie ist bestens dokumentiert (z.B. Daston, 1988; Gigerenzer et al., 1999; Hacking, 1975, 1990). Ich werde diese Entwicklung hier nicht nacherzählen, sondern einen anderen Weg gehen: eine kurze Geschichte in Form klassischer Denkprobleme und der Bedeutung des statistischen Denkens als dem Einmaleins des skeptischen Denkens, damals und heute. Ich beginne mit einem fanatischen Spieler und zwei großen Mathematikern.

1 Die Wette des Chevalier

Der Chevalier de Méré war ein leidenschaftlicher Spieler und lebte im Frankreich des 17. Jahrhunderts. Eines der Spiele, mit denen er seine Mitspieler verführte, war das folgende: „Wir werfen einen Würfel viermal. Wenn dabei eine oder mehrere Sechsen sind, gewinne ich. Wenn keine Sechsen dabei ist, gewinnen Sie.“ Soweit wir wissen, waren seine Würfel fair; dennoch gewann der Chevalier mit diesem Spiel regelmäßig Geld. Schließlich fand er keine Opfer mehr, oder das Spiel wurde auf die Dauer eintönig - was immer der Grund war, er dachte sich eine Variante aus, die ebenso lukrativ sein sollte. Hier ist das neue Spiel, das der Chevalier seinen Mitspielern anbot:

Doppel-Sechs: Wir werfen ein Paar von Würfeln 24 Mal. Wenn dabei eine Doppel-Sechs oder mehrere sind, gewinne ich. Wenn keine Doppel-Sechs dabei ist, gewinnen Sie.

Würden Sie das Angebot annehmen? De Méré's Intuition ist durchsichtig. Er wusste aus Erfahrung, dass es von Vorteil ist, darauf zu wetten, dass mindestens eine Sechs in einer Serie von 4 Würfeln auftritt. Eine Doppel-Sechs ist aber 6-mal so selten wie eine einfache Sechs. Daraus schloss er, dass es von Vorteil ist, darauf zu wetten, dass er mindestens eine Doppel-Sechs in 24 (also 6 mal 4) Würfeln erhält. Fortuna jedoch enttäuschte den Chevalier; er begann zu verlieren. War er glücklos, obgleich er richtig dachte, oder war er glücklos, weil er falsch dachte? Der Chevalier konnte diese Frage nicht entscheiden, seine Intuition sprach für Ersteres, seine Erfahrung für Letzteres.

De Méré wandte sich an die berühmten Mathematiker Blaise Pascal und Pierre Fermat, die im Jahre 1654 eine Reihe von Briefen über dieses und ähnliche Probleme austauschten und einen allgemeinen Lösungsweg entwickelten. Deshalb wird 1654 als das Geburtsjahr der mathematischen Theorie der Wahrscheinlichkeit angenommen. Die Enttäuschung des Chevalier de Méré war der Anlass für eine der größten intellektuellen Revolutionen. Hier ist die Analyse von Pascal und Fermat, in moderner Terminologie.

Beginnen wir mit dem ersten Spiel. Wie hoch ist die Wahrscheinlichkeit von mindestens einer Sechs in einer Serie von vier Würfeln? Die Wahrscheinlichkeit $p(\text{Sechs})$ von einer Sechs in einem Wurf eines fairen Würfels ist $1/6$. Daher ist die Wahrscheinlichkeit von „keine Sechs“

$$p(\text{keine Sechs}) = 5/6.$$

Die Wahrscheinlichkeit von „keine Sechs“ in einer Serie von 4 Würfeln ist daher:

$$p(\text{keine Sechs in 4 Würfeln}) = (5/6)^4 = (5/6)(5/6)(5/6)(5/6) = .482.$$

Also ist die Wahrscheinlichkeit $p(\text{mindestens eine Sechs in 4 Würfeln}) = .518$. Wir verstehen nun, warum de Méré mit dem ersten Spiel Geld verdient. Seine Chance zu gewinnen, war etwas höher als 50%. Die gleiche Logik lässt sich auf das Doppel-Sechs Spiel anwenden. Wenn Sie die Antwort noch nicht sehen, geben Sie nicht auf. Wir lösen jetzt ein Problem, das vor 1654 noch niemand gelöst hat. Nochmals, die Frage ist: Wie hoch ist die Wahrscheinlichkeit, mindestens eine Doppel-Sechs in 24 Würfeln zu erhalten? Die Wahrscheinlichkeit $p(\text{Doppel-Sechs})$ in einem Wurf mit einem Paar von Würfeln ist $1/36$. Daher ist die Wahrscheinlichkeit von „Keiner Doppel-Sechs“

$$p(\text{keine Doppel-Sechs}) = 35/36.$$

Die Wahrscheinlichkeit von „keine Doppel-Sechs“ in einer Serie von 24 Würfeln ist daher:

$p(\text{keine Doppel-Sechs in 24 W\u00fcrfen}) = (35/36)^{24} = .509$.

Also ist die Wahrscheinlichkeit daf\u00fcr, mindestens eine Doppel-Sechs in 24 W\u00fcrfen zu erhalten gleich .491. Jetzt sehen wir, dass die Chance, das Doppel-Sechs-Spiel zu gewinnen, tats\u00e4chlich leicht unter 50% liegt. Der Grund warum de M\u00e9r\u00e9 verlor, war also nicht ein Mangel an Gl\u00fcck, sondern eine falsche Intuition. Doch die Genauigkeit seiner Erfahrung am Spieltisch ist faszinierend. Er muss reichlich Mitspieler gefunden und lange Zeit mit diesem Spiel verbracht haben, um den kleinen Unterschied zu 50% bemerken zu k\u00f6nnen. Dieser Widerspruch zwischen genauer Erfahrung und falscher Intuition inspirierte Pascal und Fermat, die Gesetze der Wahrscheinlichkeit zu suchen und zu finden. Hier sind sie, in moderner Terminologie:

1. Die Wahrscheinlichkeit eines unm\u00f6glichen Ereignisses ist 0 und jene eines sicheren Ereignisses ist 1.
2. Die Summe der Wahrscheinlichkeiten aller m\u00f6glichen Ereignisse ist 1.
3. Wenn A und B unabh\u00e4ngige Ereignisse sind, dann ist die Wahrscheinlichkeit $p(A\&B)$ daf\u00fcr, dass A und B eintreten, gleich dem Produkt der individuellen Wahrscheinlichkeiten: $p(A\&B) = p(A)p(B)$.

Zum Beispiel: Die Wahrscheinlichkeit, eine „7“ mit einem regul\u00e4ren W\u00fcrfel zu erhalten ist 0 und jene, eine Zahl zwischen 1 und 6 zu erhalten, ist 1. Die Summe aller Wahrscheinlichkeiten f\u00fcr die Ergebnisse 1 bis 6 betr\u00e4gt 1, und die Wahrscheinlichkeit, eine „6“ im ersten Wurf und eine „1“ im zweiten zu erhalten, also von zwei unabh\u00e4ngigen Ereignissen, betr\u00e4gt $1/6$ mal $1/6$, das ergibt $1/36$.

2 Pascals Wette

Die Gesetze der Wahrscheinlichkeit waren eine Antwort auf Erfahrungen mit Gl\u00fccksspielen, aber dies war nur eine von mehreren Wurzeln. Die Entwicklung des Denkens in Wahrscheinlichkeiten war vielmehr Teil einer gro\u00dfen intellektuellen Revolution: die Aufgabe des Ideals des sicheren Wissens und die Entwicklung von Formen des rationalen Umgangs mit einer unsicheren Welt.

Aristoteles teilte unsere Welt einst in zwei Reiche auf: in die himmlische Welt der unver\u00e4nderlichen Ordnungen und des gesicherten Wissens und die ungeordnete Welt voller Ver\u00e4nderungen und Ungewissheiten. Jahrhundertlang glaubten Mathematiker wie auch Theologen und ihre gl\u00e4ubigen Anh\u00e4nger, sie lebten in einer Welt absoluter Gewissheit. Doch die Reformation und die Gegenreformation unterh\u00f6hlten weitgehend das Reich der Gewissheit. Allm\u00e4hlich setzte sich ein bescheideneres Ideal durch. Man fand sich damit ab, dass vollst\u00e4ndige Gewissheit des Wissens unerreichbar ist, hielt aber trotzdem daran fest, dass das verf\u00fcgbare Ma\u00df an Wissen ausreicht, um vern\u00fcnftige Menschen in Theorie und Praxis zu lenken.

Religiöse Überzeugungen waren und sind noch heute von emotionaler Gewissheit gefärbt, und das gilt für Gläubige wie auch für Atheisten. Man weiß mit absoluter Sicherheit, dass Gott existiert. Oder es erscheint unbezweifelbar, dass er nicht existiert. In seinen *Pensées* aber stellt Blaise Pascal (1669, Bd. 2, S. 141-55) die religiöse Frage in einem völlig anderen Licht. Es geht nicht mehr um die Wahrheit, sondern um die Erwartung. Eine Erwartung ist nicht sicher, sondern wie eine Wette. Pascals Wette kann man so zusammenfassen:

Pascals Wette: Ich weiß nicht, ob Gott existiert. Aber ich weiß, wenn ich an ihn glaube und er nicht existiert, dann werde ich einige Momente weltlicher Lust und Laster versäumen. Wenn ich aber nicht an ihn glaube und er dennoch existiert, dann werde ich mit ewiger Verdammung und ewigem Elend dafür bezahlen. Worauf soll ich wetten?

Für Pascal ist die Antwort klar: Auch wenn man die Chance, dass Gott existiert, für beliebig gering hält, werden doch, falls er existiert, die Folgen unabsehbar hoch sein: unendlich die Seligkeit der Erlösten, aber auch unendlich das Elend der Verdammten. Unter diesen Umständen, so Pascals Argument, verlangt rationales Eigeninteresse, dass wir unsere sicheren, aber nur endlichen weltlichen Freuden dem ungewissen, aber unendlichen Gewinn der Erlösung opfern.

Pascals Wette illustriert eine radikal neue Denkweise, die mit der Entwicklung der Wahrscheinlichkeitstheorie einhergeht. Religiosität ist eine Frage der Erwartung, nicht des unbedingten Glaubens, und diese Erwartung ist unsicher. Es ist wohl kein Zufall, dass zur selben Zeit, als das neue Denken in Wahrscheinlichkeiten und Erwartungen Fuß fasste, der Gebrauch der Folter in Europa zurückging. In der Inquisition war die Folter das Werkzeug, um die eindeutige Wahrheit herauszufinden - ein Zweck, der die Mittel heiligte.

Pascals Grundbegriff war nicht die Wahrscheinlichkeit, sondern die Erwartung, die später als Produkt aus der Wahrscheinlichkeit p_i eines Ereignisses i und seinem Wert x_i bestimmt wurde:

$$E = \sum p_i x_i.$$

Beispielsweise beträgt am Rouletttisch die Wahrscheinlichkeit von „Rot“ und „Schwarz“ jeweils $18/37$ und jene von „Grün“ (null) beträgt $1/37$. Wenn man 100 Euro auf „Rot“ setzt, beträgt also die Erwartung $18/37 \times 200 \text{ Euro} + 18/37 \times 0 \text{ Euro} + 1/37 \times 0 \text{ Euro} = 97,79 \text{ Euro}$. Die Definition von rationalem Verhalten durch die Erwartung wurde zum Grundstein des neuen Verständnisses dafür, mit Unsicherheiten umzugehen, statt sie zu verleugnen, und mit falschen Sicherheiten zu leben. Aber die Definition von rationalem Handeln als Maximierung der Erwartung war noch nicht das Ende der Geschichte. Die mathematische Erwartung geriet bald in unerwartete Schwierigkeiten.

3 Das St.-Petersburg-Paradox

Das St.-Petersburg-Paradox brachte den ersten großen Konflikt zwischen dem Konzept der rationalen Erwartung und dem gesunden Menschenverstand (Jorland, 1987). Nicholas Bernoulli wies als Erster in einem Brief an Pierre de Montmort auf das Problem hin; dieser veröffentlichte es in der zweiten Auflage seines *Essai d'analyse sur les jeux de hasard* (1713). Daniel Bernoulli, ein Vetter von Nicholas, veröffentlichte im Jahre 1738 eine mögliche Lösung in den Jahrbüchern der Petersburger Akademie, daher kommt der Name des Problems.

St.-Petersburg-Spiel: Pierre und Paul spielen ein Glücksspiel mit einer fairen Münze. Wenn das Ergebnis des ersten Wurfs „Ziffer“ ist, muss Pierre an Paul einen Euro zahlen und das Spiel ist beendet. Wenn „Ziffer“ erst beim zweiten Wurf kommt, erhält Paul zwei Euro; wenn dies erst beim dritten Wurf geschieht, gewinnt er vier Euro und so weiter. Wie hoch ist der faire Preis, den Paul zahlen sollte, um das Spiel zu spielen?

Wie viel würden Sie bieten? Der faire Preis ist jene Summe, bei der eine Person unentschieden ist, die Rolle von Pierre oder die von Paul zu spielen. (Wenn ein Kind ein Stück Kuchen in zwei Teile teilt, und das andere Kind die Wahl hat, handelt es sich um dasselbe Prinzip von Fairness.) Nach der klassischen Theorie der Rationalität ist der faire Preis durch die mathematische Erwartung definiert:

$$E = (1/2 \times 1 \text{ Euro}) + (1/4 \times 2 \text{ Euro}) + (1/8 \times 4 \text{ Euro}) + \dots + [(1/2)^n \times (2^{n-1} \text{ Euro})] + \dots = \infty$$

In Worten, mit der Wahrscheinlichkeit $p = 1/2$ gewinnt Paul 1 Euro, mit $p = 1/4$ gewinnt er 2 Euro, mit $p = 1/8$ gewinnt er 4 Euro und so weiter. Man kann sehen, dass jedes der Glieder auf der rechten Seite der Gleichung einer Erwartung von $1/2$ Euro entspricht, und da deren Anzahl unendlich ist, ist der Erwartungswert ebenfalls unendlich groß. Gemäß der Theorie, dass die Erwartung der faire Preis ist, sollte jeder von uns all sein Vermögen einsetzen, um dieses Spiel zu spielen - und dies würde sogar noch von Vorteil sein, da das Vermögen ja nur endlich groß ist. Kein vernünftiger Mensch ist jedoch bereit, mehr als eine kleine Summe, vielleicht 5 bis 10 Euro, für dieses Spiel zu bieten.

Die Mathematiker nannten diesen Widerspruch zwischen Theorie und gesundem Menschenverstand das „St.-Petersburg-Paradox“. Nach unserem heutigen Verständnis liegt aber kein Paradox vor: Es gibt keinen Widerspruch zwischen Resultaten, die sich aus gleich validen Annahmen ableiten lassen. Nach dem klassischen Verständnis war die Theorie der Wahrscheinlichkeit jedoch keine reine, inhaltsfreie Theorie, sondern untrennbar von ihrem Gegenstand. Und dieser Gegenstand war die menschliche Vernunft. Der Widerspruch zwischen Theorie und Vernunft wurde daher als ein Paradox interpretiert.

Daniel Bernoulli versuchte diesen Widerspruch zu lösen. Er argumentierte, dass es beim St.-Petersburg-Problem nicht alleine um Fairness gehe und man anstelle der mathematischen Erwartung die „moralische“ Erwartung des umsichtig abwägenden Kaufmanns einführen sollte. Diese definierte er als das Produkt aus der Wahrscheinlichkeit des Ergebnisses und dem, was später sein Nutzen genannt wurde. Bernoulli argumentierte, dass ein Gewinn von 200 Euro (in moderner Währung) nicht notwendigerweise doppelt soviel wert ist wie ein Gewinn von 100 Euro, und dass ein Spieler umso mehr Geld gewinnen muss, um glücklich zu werden, je reicher er schon ist. Nehmen wir an, die Beziehung zwischen Euro (x) und Nutzen N ist logarithmisch, $N(x) = \ln(x)$, und ihr derzeitiges Vermögen V beträgt 50.000 Euro. Dann berechnet sich der sichere Gewinn G , der denselben Nutzen hat wie die Teilnahme am St.-Petersburg-Spiel, wie folgt:

$$N(V+G) = 1/2N(V+1) + 1/4N(V+2) + 1/8N(V+4) + \dots$$

Die Berechnung ergibt einen Wert von etwa 9 Euro. In Worten ausgedrückt, für jemanden, der ein Vermögen von 50.000 Euro besitzt, beträgt der erwartete Nutzen dieses Spiels nur 9 Euro. Dieser Wert liegt im Bereich dessen, was jemand mit einem gesunden Menschenverstand bereit ist, für das Spiel zu zahlen.

Mit dieser Lösung des St.-Petersburg-Paradox transformierte Daniel Bernoulli das Konzept des erwarteten Werts von Pascal-Fermat in das Konzept des erwarteten Nutzens, das auch heute noch die Wirtschaftswissenschaften dominiert. Nicholas Bernoulli, der Professor für Römisches und Kanonisches Recht an der Universität Basel war, hielt dagegen weiterhin an Fairness als Modell der menschlichen Vernunft fest, da er dieses als die Grundlage für rechtliche Verträge ansah. Daniel Bernoulli nahm dagegen seine Inspiration aus der Welt von Handel und Gewerbe, nicht aus dem Recht. Für ihn war Vernünftigkeit kluger Geschäftssinn, und es war sicherlich nicht wirtschaftlich klug, eine hohe Summe in das St.-Petersburg-Spiel zu investieren. Für sein neues Verständnis von rationalem Verhalten war der Prototyp eines vernünftigen Menschen nicht mehr der unparteiische Richter, sondern der umsichtig abwägende Kaufmann. Die mathematische Theorie der menschlichen Vernunft hatte eine Wende vollzogen.

4 Pro Monogamie: Der erste Nullhypothesentest

Die mathematische Wahrscheinlichkeit hat drei Quellen: Glücksspiel, Gerichtshof und statistische Tabellen. Die drei Hauptinterpretationen des Konzepts der Wahrscheinlichkeit entstammen diesen drei Anwendungen: Propensität, subjektive Wahrscheinlichkeit und Häufigkeit. Mit Propensität (Englisch: propensity) ist das Design eines Würfels oder eines Roulettrads gemeint, das die Wahrscheinlichkeit bestimmt. Das Konzept der subjektiven Wahrscheinlichkeit kommt aus rechtlichen Fragen, etwa wie hoch ein Richter die Glaubwürdigkeit von Zeugen einschätzen soll, zum Beispiel

wenn diese mit dem Angeklagten verwandt oder nicht verwandt sind. Die Interpretation von Wahrscheinlichkeit als relative Häufigkeit auf lange Sicht basiert auf statistischen Informationen wie Geburten- und Sterblichkeitstabellen. Diese Tabellen waren eine der ersten systematischen Datenbanken, die in der westlichen Welt angelegt wurden, und zugleich die Basis für den ersten Nullhypothesentest.

John Arbuthnot (1710) stellte eine alte Frage: Gibt es einen aktiven Gott? Seine Methode, eine Antwort zu finden, aber war neu und revolutionär. Er suchte nach Fakten statt Rhetorik und nach einem empirischen Test. Arbuthnot beobachtete, dass Männer gefährlicher leben als Frauen und dass mehr junge Männer als Frauen durch Unfälle sterben. Wenn es einen aktiven Gott gibt und dieser Monogamie vorsieht, so argumentierte er, wird er mehr Jungen als Mädchen erzeugen, um diesen Verlust zu kompensieren. Arbuthnot testete diese Hypothese göttlicher Vorsehung gegen die Nullhypothese von blindem Zufall. Dazu untersuchte er die Geburtenstatistiken, die damals seit 82 Jahren in London geführt wurden. Er stellte fest, dass in jedem der 82 Jahre mehr Jungen als Mädchen geboren wurden, und berechnete die „Erwartung“ dieses Ergebnisses (D) unter der Nullhypothese (H_0):

$$p(D|H_0) = (1/2)^{82}.$$

Weil diese Wahrscheinlichkeit so außerordentlich klein war, schloss er blinden Zufall aus und sah das Ergebnis als Beweis für die göttliche Vorsehung. Hier ist seine Folgerung in seinen eigenen Worten:

„*Scholium.* From hence it follows, that Polygamy is contrary to the Law of Nature and Justice, and to the Propagation of the human Race; for where Males and Females are in equal number, if one Man takes Twenty Wives, Nineteen Men must live in Celibacy, which is repugnant to the Design of Nature; nor is it probable that Twenty Women will be so well impregnated by one Man as by Twenty.“

Arbuthnots Idee, eine Behauptung gegen statistische Daten zu testen, war revolutionär und seiner Zeit weit voraus. Nullhypothesentests kamen erst im späten 19. Jahrhundert und frühen 20. Jahrhundert in Gebrauch und wurden durch die Arbeiten des Statistikers und Genetikers Sir Ronald Fisher popularisiert. Arbuthnots Test macht zugleich die Möglichkeiten und Grenzen eines Nullhypothesentests (Arbuthnot verwandte diesen Begriff noch nicht) so klar wie kaum ein späteres Beispiel. Dieser Test liefert die Wahrscheinlichkeit einer Serie von Beobachtungen (in jedem von 82 Jahren mehr Jungen als Mädchen), falls die Nullhypothese gilt. Die erste Beschränkung liegt darin, dass in dieser Methode per Definition die Forschungshypothese (hier: göttliche Vorsehung) selbst nicht in statistischer Form formuliert wird. Die göttliche Vorsehung gewinnt immer, falls die Nullhypothese verliert, die einzige Bedingung hier ist, dass mehr Jungen geboren werden. Welches Verhältnis aber die Forschungshypothese vorher-

sagt, wird nicht spezifiziert. Die zweite Beschränkung liegt darin, dass keine Alternativhypothese in statistischer Form getestet wird. Beispielsweise könnte eine Alternativhypothese besagen, dass 3% aller weiblichen Neugeborenen unmittelbar nach der Geburt illegal ausgesetzt oder getötet werden und deshalb nicht in der Statistik auftauchen. Selbst wenn das der Fall ist, würde Arbuthnots Test dennoch diesen Effekt der göttlichen Vorsehung zuschreiben. Der Mangel an präzisen Forschungs- und Alternativhypothesen ist das größte Problem in Arbuthnots Test, und das gilt ebenfalls für die heutigen Anwendungen von Nullhypotesentests in den Sozialwissenschaften (Gigerenzer, 1993).

Die Bedeutung von Arbuthnots Vorgehen liegt jedoch nicht in der besonderen Struktur dieses Nullhypotesentests, sondern in der bahnbrechenden Idee, Fragen durch Heranziehen empirischer Daten zu entscheiden. Dies war damals ein revolutionäres Vorgehen, doch die Revolution selbst musste noch fast 2 Jahrhunderte warten. Arbuthnots Test fand keine Beachtung. Wenn wir uns darüber wundern, sollten wir nicht vergessen, dass es auch heute keine Ausnahme ist, dass Ideologien statt empirischer Evidenz über religiöse, pädagogische und politische Fragen entscheiden.

5 Wer entdeckte die Regel von Bayes?

Das Testen einer Nullhypothese blieb nicht die einzige Methode, eine Hypothese zu testen. Die Regel von Bayes ist eine der bekanntesten Alternativen. Thomas Bayes (1702-1761) war ein Reverend der „Nonconformist Church“. Er hat seine berühmte Abhandlung über das Problem der „inversen Wahrscheinlichkeit“ - die Wahrscheinlichkeit einer Hypothese gegeben Evidenz - nie selbst veröffentlicht. Dafür hat ihm R. A. Fisher (1935) später gratuliert, denn Fisher war der Meinung, dass die Regel von Bayes für das Testen wissenschaftlicher Hypothesen nutzlos sei. In Fishers Augen hatte Bayes dies erkannt, doch seine Bewunderer nicht. Die Regel von Bayes wurde nach seinem Tod von Richard Price im Jahre 1763 publiziert.

Für den einfachsten Fall mit binären Hypothesen, H_1 und H_2 , und einem Datum (Evidenz) D ergibt sich die Regel von Bayes (in moderner Terminologie) als:

$$p(H_1 | D) = \frac{p(H_1)p(D | H_1)}{p(H_1)p(D | H_1) + p(H_2)p(D | H_2)}$$

In Worten ausgedrückt, die Aposteriori-Wahrscheinlichkeit $p(H_1|D)$ ergibt sich aus der Apriori-Wahrscheinlichkeit $p(H_1)$ und den Wahrscheinlichkeiten $p(D|H_1)$ und $p(D|H_2)$. Betrachten wir eine moderne Anwendung, HIV-Screening für Personen ohne Risikoverhalten, das heißt, HIV-Tests für Personen, die z.B. nicht intravenös Drogen spritzen (Gigerenzer, 2002). Mit einem positiven Testergebnis (D) ist hier gemeint, dass eine Blutprobe so-

wohl in dem Suchtest (Elisa) als auch in dem Bestätigungstest (Western-Blot-Test) positiv getestet.

In Deutschland ist etwa einer von je 10.000 Männern ohne Risikoverhalten HIV-infiziert. Die Wahrscheinlichkeit $p(D|H_1)$, dass der Test positiv (D) wird, wenn der Mann infiziert ist, beträgt .999. Die Wahrscheinlichkeit $p(D|H_2)$, dass der Test positiv wird, wenn der Mann nicht infiziert ist, beträgt .0001. Ein Mann getestet positiv. Wie hoch ist die Wahrscheinlichkeit $p(H_1|D)$, dass er wirklich mit dem Virus infiziert ist?

Wenn man diese Werte in die Regel von Bayes einsetzt, dann erhält man einen Wert von $p(H_1|D) = .5$. Nur jeder zweite, der positiv getestet, hat demnach tatsächlich den Virus. Die meisten professionellen AIDS-Berater an deutschen Gesundheitsämtern teilen Klienten dagegen fälschlicherweise mit, dass es absolut sicher oder zumindest zu 99,9% sicher sei, dass man infiziert sei (Gigerenzer, Hoffrage & Ebert, 1998). Die Ausbildung dieser Berater im statistischen Denken lässt sehr zu wünschen übrig. Die Folgen dieser Fehlinformation können vom Verlust des Arbeitsplatzes bis zum Selbstmord reichen.

Die Regel von Bayes ist vom Testen von Nullhypothesen zu unterscheiden. Betrachten wir das HIV-Screening einmal aus der Perspektive des Nullhypothesentestens. Die H_0 postuliert hier, dass eine Person nicht durch HIV infiziert ist. Man erhält dennoch ein positives Ergebnis (D) und stellt die Frage, ob dieses Ergebnis signifikant ist. Die Wahrscheinlichkeit $p(D|H_0)$ des positiven Testergebnisses gegeben, dass die Nullhypothese wahr ist, beträgt .0001. Dies ist ein signifikantes Ergebnis. Daher wird die Nullhypothese zurückgewiesen und zwar mit einem hohen Grad von „Vertrauen“. Wie die Regel von Bayes jedoch zeigt, ist die Wahrscheinlichkeit einer HIV-Infektion gegeben ein positives Testergebnis nur .5. Warum kommen beide Methoden zu verschiedenen Aussagen? Die Regel von Bayes berücksichtigt die drei Wahrscheinlichkeiten $p(H_1)$, $p(D|H_1)$ und $p(D|H_2)$, während der Nullhypothesentest nur die letztere Wahrscheinlichkeit berücksichtigt.

Beim HIV-Screening oder anderen Reihenuntersuchungen kennt man die Grundrate einigermaßen genau und kann daher diese als Apriori-Wahrscheinlichkeit einsetzen. Wenn es sich jedoch um eine wissenschaftliche Hypothese handelt, hat man in der Regel keine statistische Information, welche dieser Grundrate entspricht, und die Apriori-Wahrscheinlichkeit wird notwendig subjektiven Charakter haben. Die mögliche Beliebigkeit subjektiver Wahrscheinlichkeiten war das Ziel von R. A. Fishers Attacke gegen die Verwendung der Regel von Bayes für die Bestimmung der Wahrscheinlichkeit von Hypothesen.

Zurück zu Thomas Bayes. Der Historiker der Statistik, Stephen M. Stigler (2001), hat einmal ein Gesetz über den Ursprung von Erkenntnissen aufge-

stellt, das er *Law of Eponymy* nannte. Dieses Gesetz besagt, dass keine wissenschaftliche Entdeckung nach ihrem ursprünglichen Entdecker benannt ist. Das Theorem von Pythagoras wurde nicht von Pythagoras entdeckt, Pascals Dreieck stammt nicht von Pascal, und die Gaußsche Verteilung wurde auch nicht von Gauß entdeckt. Die Regel von Bayes scheint keine Ausnahme zu sein. In einer spannenden Detektiv-Geschichte berechnete Stigler eine Wahrscheinlichkeit von 3 zu 1, dass tatsächlich Nicholas Saunderson und nicht Thomas Bayes die Regel entdeckt hat. Saunderson, der seit seinem ersten Lebensjahr vollständig blind war und Optik lehrte, hatte den begehrten „Lucasian Chair of Mathematics“ in Cambridge inne, den vor ihm Newton hatte. Er starb im Jahre 1739. Man kann Bayes jedoch nicht die weniger höfliche (und falsche) Interpretation von Stiglers Gesetz vorwerfen, die besagt, dass jede wissenschaftliche Entdeckung nach der letzten Person benannt sei, die ihre Vorgänger nicht zitiert hat. Wie bereits erwähnt hat Bayes seine Abhandlung ja nie veröffentlicht. Bleibt noch die Frage: Wer entdeckte Stiglers Gesetz?

6 Die erste Nacht im Paradies

Der erste Tag im Paradies geht zu Ende. Adam und Eva legen sich zur Ruhe. Sie hatten am Tag die Sonne aufgehen sehen und sie bewundert, wie sie am Himmel ihre Bahn zog und all die herrlichen Bäume, Blumen und Vögel beschien. Irgendwann wurde es aber kühler, während die Sonne unter dem Horizont verschwand. Würde es jetzt auf ewig dunkel bleiben? Adam und Eva fragen sich sorgenvoll, welche Chance sie wohl haben, dass die Sonne wieder aufgeht?

Im Nachhinein könnten wir vermuten, dass Adam und Eva sicher waren, dass die Sonne wieder aufgeht. Aber sie hatten die Sonne ja erst einmal am Firmament emporsteigen sehen. Was also konnten sie erwarten? Die klassische Antwort auf dieses Problem wurde im Jahre 1812 von dem französischen Mathematiker Pierre Simon de Laplace gegeben.

Wenn Adam und Eva die Sonne niemals hätten aufgehen sehen, würden sie für beide möglichen Ereignisse (das Wieder-Aufgehen und das Dunkel-Bleiben) gleiche Wahrscheinlichkeiten ansetzen. Daher würden sie - für das Wieder-Aufgehen der Sonne - einen weißen Stein in einen Beutel stecken, außerdem einen schwarzen Stein für das Dunkel-Bleiben. Aber sie hatten ja schon einmal gesehen, wie die Sonne aufging, und legten deshalb einen weiteren weißen Stein in den Beutel. Dort lagen jetzt also zwei weiße Steine und ein schwarzer. Das bedeutet, ihr Überzeugungsgrad, dass die Sonne wieder aufgehen wird, war von $1/2$ auf $2/3$ gestiegen. Nach dem folgenden Tag - also nach dem zweiten Sonnenaufgang, den sie erlebten - gaben sie einen dritten weißen Stein hinzu; nun war für sie die Wahrscheinlichkeit für einen Sonnenaufgang von $2/3$ auf $3/4$ angewachsen. Laplaces *Sukzessionsregel* (Regel der Folge) gibt allgemein die Wahrscheinlichkeit $p(E|n)$ an, dass ein Ereignis E wieder auftritt, nachdem es n mal eingetreten ist:

$$p(E|n) = (n+1)/(n+2).$$

Ein 27-Jähriger hat in seinem Leben ungefähr 10.000 Sonnenaufgänge erlebt. Daher beträgt für ihn der Überzeugungsgrad, dass die Sonne auch am nächsten Tag wieder aufgeht, $10.001/10.002$. Diese Anwendung von Laplaces Regel ist auf Kritik gestoßen. Da die Regel von Laplace aus der Regel von Bayes hergeleitet ist, zeigt diese Kritik zugleich Probleme mit der Anwendung der Regel von Bayes auf. Anders als bei dem HIV-Screening, wo der Grundanteil der Infektion in der betreffenden Population bekannt ist, konnten Adam und Eva anfangs keinen Grundanteil der Sonnenaufgänge kennen. Sie konnten also nicht wissen, wie viele weiße oder schwarze Steine sie am ersten Abend in den Beutel stecken mussten. Als Pessimisten hätten sie vielleicht einen weißen und zehn schwarze Steine genommen, als Optimisten dagegen zehn weiße und nur einen schwarzen. Wenn man keine Informationen zum Abschätzen der Wahrscheinlichkeiten hat, kann man den möglichen Ereignissen oder Ergebnissen gleich hohe Wahrscheinlichkeiten zuschreiben. Diese Faustregel nennt man *Indifferenzprinzip*. Seine Befürworter verteidigen es damit, dass die anfängliche Annahme gleicher Wahrscheinlichkeiten umso geringere Auswirkungen hat, je mehr Beobachtungen man einfließen lässt. Beispielsweise erhält man für die Wahrscheinlichkeit, dass die Sonne morgen aufgeht, nach zehn Jahren, also nach über 3.650 Sonnenaufgängen, praktisch denselben Wert, unabhängig davon, ob man als Pessimist oder Optimist begann.

Die problematische Annahme der Indifferenz hat Laplace tatsächlich an zwei Stellen gemacht, und man kann das sehen, wenn man die Regel von Bayes verwendet. H_1 und H_2 stehen für die Hypothesen, dass die Sonne jeden Morgen aufgeht bzw. nicht jeden Morgen aufgeht, und D dafür, dass Adam und Eva einen Sonnenaufgang beobachtet haben. Dann ergibt sich die gesuchte Wahrscheinlichkeit $p(H_1|D)$ wie folgt:

$$p(H_1 | D) = \frac{p(H_1)p(D | H_1)}{p(H_1)p(D | H_1) + p(H_2)p(D | H_2)}$$

Nun kennen aber Adam und Eva die Grundrate nicht. Die Faustregel, das *Indifferenzprinzip*, nimmt an: $p(H_1) = p(H_2) = 1/2$. Damit vereinfacht sich die Regel von Bayes zu:

$$p(H_1 | D) = \frac{p(D | H_1)}{p(D | H_1) + p(D | H_2)}$$

Die Wahrscheinlichkeit $p(D|H_1)$ ist per Definition 1, aber die Wahrscheinlichkeit $p(D|H_2)$ ist nicht bekannt. Hier wird das *Indifferenzprinzip* nochmals angewandt: $p(D|H_2) = 1/2$. Daraus erhält man schließlich:

$$p(H_1|D) = 2/3.$$

Die erste Nacht im Paradies illustriert die Problematik, die Regel von Bayes in Situationen anzuwenden, in denen keine oder unzureichende empirische Daten vorliegen. Das Indifferenzprinzip, so umstritten es ist, spielt auch heute eine Rolle, beispielsweise in Vaterschaftsprozessen.

Um die Wahrscheinlichkeit zu bestimmen, dass ein Mann wirklich der Vater eines Kindes ist, braucht man, wie die Regel von Bayes zeigt, eine Apriori-Wahrscheinlichkeit oder Grundrate. Aber was könnte diese Apriori-Wahrscheinlichkeit sein? Viele Laboratorien verwenden hier das Indifferenzprinzip und nehmen eine Apriori-Wahrscheinlichkeit von 50% dafür an, dass der Angeklagte tatsächlich der Vater ist (Gigerenzer, 2002). Diese Praxis ist umstritten, da sie voraussetzt, dass der Angeklagte genauso wahrscheinlich der Vater ist wie alle anderen Männer zusammen. Die Frage ist, wie man die beiden Fehler gewichten soll, die ein Richter machen kann: Einen Schuldigen freizusprechen oder einen Unschuldigen verurteilen. Die französischen Mathematiker Dennis Poisson und Pierre Laplace vertraten eine konservative Linie - der Schutz der Gesellschaft vor Kriminellen ist wichtiger als der Schutz des Individuums vor falscher Verurteilung - gegen die früheren liberalen Reformen des Philosophen und Politikers Condorcet. Die Antwort auf diese Frage trennt Liberale von Konservativen, heute wie in den vergangenen Jahrhunderten.

7 Die Illusion von Gewissheit

Die klassische Theorie der Wahrscheinlichkeit überkam das Streben nach absoluter Gewissheit mit einer epistemischen Interpretation von Wahrscheinlichkeit. Das bedeutet, dass die Ursache der Ungewissheit in der Unkenntnis des Menschen, nicht aber in der Natur selbst gesehen wurde. Die Vertreter der klassischen Theorie, von Pascal bis Laplace, waren Deterministen: Sie hielten die Welt selbst für vollständig determiniert. Gott oder seine säkulare Version, Laplaces Superintelligenz, braucht keine Statistik, nur wir Menschen können ohne diese nicht auskommen. Das war auch noch Albert Einsteins Sicht: Gott würfelt nicht. Eine ontische Interpretation der Wahrscheinlichkeit musste bis zur zweiten Hälfte des 19. Jahrhunderts warten, als der Begründer der Psychophysik, Gustav Theodor Fechner, und der Philosoph Charles Sanders Peirce den Zufall als eine Eigenschaft der Natur postulierten.

Die Evolution des statistischen Denkens, die ich an sechs klassischen Problemen illustriert habe, ist vor allem eine Entwicklung des skeptischen Denkens. Dieses Einmaleins der Skepsis hat mehrere Seiten. Wie Pascals Wette veranschaulicht, ist die Basis für skeptisches Denken der Übergang vom Streben nach Gewissheit zum vernünftigen Umgang mit einer unsicheren Welt. Benjamin Franklin sagte einmal, „nichts ist sicher in dieser Welt außer der Tod und die Steuern“. Diese scherzhaft formulierte, aber tiefe Ein-

sicht ist auch heute noch für viele Menschen zu schmerzhaft, um sie als Lebensgefühl zu akzeptieren.

Das Streben nach trügerischer Gewissheit ist ein Teil unseres emotionalen und kulturellen Erbes. Es erfüllt die Sehnsucht nach Sicherheit und Autorität. Die Esoterik-Abteilungen der heutigen Buchhandlungen zeugen davon, dass viele Menschen sich nach schnellem Glauben sehnen. Zu allen Zeiten erdachte man Glaubenssysteme, wie Religion, Astrologie und Weissagung, die Sicherheit und Gewissheit versprechen und in denen die Menschen Trost finden können - vor allem jene, die Schweres erleiden müssen. Gewissheit ist inzwischen zu einer Ware geworden. Sie wird weltweit vermarktet: durch Versicherungsgesellschaften, Anlageberater und Wahlkämpfer, aber auch in Medizin und Pharmazie.

Im 17. Jahrhundert bestand in Europa der Erwerb einer Lebensversicherung darin, eine Wette auf die Lebensdauer eines prominenten Bürgers abzuschließen. Zum Beispiel ging es darum, ob der Bürgermeister von Paris innerhalb einer bestimmten Zeitspanne sterben würde; wenn ja, dann konnte derjenige, der darauf gesetzt hatte, womöglich ein kleines Vermögen gewinnen (Daston, 1987, 1988). Das war ein Spiel wie dasjenige des Chevalier de Méré, nicht aber eine sittliche Verpflichtung. Heutzutage reden uns die Versicherungsvertreter ein, dass es bei einer Lebensversicherung um eine Absicherung gehe und es moralisch geboten sei, sozusagen Geld auf unser eigenes Leben zu setzen, damit die Hinterbliebenen im Ernstfall versorgt seien. Auch politische Parteien schüren den Drang nach Sicherheit. Vor den Bundestagswahlen 1998 warb die CDU mit dem Slogan „Sicherheit statt Risiko“. Solche Versprechungen waren nicht nur vom damaligen Kanzler Helmut Kohl und seinen Parteifreunden zu hören - auch andere Parteien behaupteten im Wahlkampf, Sicherheit zu bieten.

Die Illusion der Gewissheit kann erzeugt werden, um politische oder wirtschaftliche Ziele zu erreichen. Nehmen wir als Beispiel die Rinderkrankheit BSE. BSE griff im Jahr 2000 in Großbritannien, Irland, Portugal, Frankreich und der Schweiz um sich, während die deutsche Regierung ihr Land für BSE-frei erklärte. „Deutsches Rindfleisch ist sicher“ - diese Phrase wiederholten der Präsident des Bauernverbandes, der Landwirtschaftsminister und eine ganze Beamtenschar unentwegt. Die Deutschen hörten das nur zu gerne. Der Import von englischem Rindfleisch wurde verboten, und man empfahl den Verbrauchern, beim Metzger nur Fleisch von Rindern zu verlangen, die in Deutschland gezüchtet worden waren. In anderen Ländern, so wurde verbreitet, seien mangelnde Sorgfalt und Kontrolle an der Tagesordnung. Als man schließlich doch zahlreiche BSE-Tests an deutschen Rinderherden vornahm, zeigte sich die Erkrankung auch hier. Die Öffentlichkeit war völlig überrascht, Minister mussten zurücktreten, die Preise für Rindfleisch fielen drastisch, und andere Länder verboten nun ihrerseits den Import von deutschem Rindfleisch. Die Regierung gestand schließlich ein,

sich zu lange an die Illusion geklammert zu haben, deutsches Vieh sei von dieser Krankheit überhaupt nicht betroffen. Das Spiel mit dem Versprechen von Sicherheit ging indes weiter, nur waren die Akteure jetzt andere. Supermärkte und Metzger hängten Plakate auf und verteilten Broschüren, in denen sie ihren Kunden versicherten: „Unser Rindfleisch ist garantiert BSE-frei.“ Einige begründeten diese Aussage damit, dass ihre „glücklichen Kühe“ auf ökologischen Wiesen weiden konnten, und andere behaupteten, ihre Rinder seien sämtlich überprüft worden - kaum jemand erwähnte, dass bei diesen Tests zahlreiche Fehler auftreten. Als die Medien schließlich von einer Kuh berichteten, die negativ getestet worden war und dennoch BSE hatte, war die Öffentlichkeit erneut schockiert. Wieder war eine Illusion der Gewissheit dahin. Regierung und Supermärkte hatten vor allem die Beruhigung der Verbraucher im Sinn und weniger die Information über BSE.

Die Illusion von Gewissheit ist nicht immer für alle bestimmt; zuweilen wird sie nur für ein ausgewähltes Publikum heraufbeschworen. So schilderte Jay Katz, Juraprofessor an der Yale University, einmal eine Diskussion mit einem befreundeten Chirurgen. Das Gespräch drehte sich um die Ungewissheiten bei der Behandlung von Brustkrebs. Beide waren sich darin einig, dass niemand weiß, wie die optimale Therapie aussieht. Katz fragte seinen Freund, wie er seine Patientinnen berät. Der Chirurg antwortete, er habe erst kürzlich einer Patientin mit Brustkrebs dringend eine Radikaloperation als beste Therapie empfohlen. Katz hielt seinem Freund vor, sich widersprüchlich zu verhalten: Wie könne er plötzlich so sicher sein, was die optimale Therapieform sei? Der Chirurg gab zu, die Patientin kaum zu kennen, beharrte aber darauf, dass seine Patientinnen die Ungewissheit über die beste Therapie weder verstehen noch hinnehmen würden, wenn sie davon wüssten. Wenn man dieser Ansicht folgt, dann wünschen die Patientinnen die Illusion der Gewissheit, und diese Patientin bekam sie auch.

Können moderne Technologien den Rest an Unsicherheit nicht bald beseitigen und Sicherheit endlich herstellen? Auch dies ist eine verbreitete Illusion. Moderne HIV-Tests zählen beispielsweise zu den besten medizinischen Tests. Wie wir gesehen haben, ist dennoch - wegen der kleinen Grundrate von HIV-Infizierten ohne Risikoverhalten - nur etwa eine von zwei Personen, die positiv testen, tatsächlich infiziert. Die Falsch-Negativ-Rate liegt bei HIV-Tests bei nur etwa 0,1%; dennoch ist in der Literatur der Fall eines amerikanischen Bauarbeiters bekannt, der 35mal negativ testete, obgleich er mit dem Virus infiziert war. Medizinische Tests sind nicht absolut sicher, und das gleiche gilt für forensische Evidenz wie Fingerabdrücke, DNS-Profile und andere genetische Tests (Gigerenzer, 2002).

8 Empirisches Denken als Lebensgefühl

Das Grundmotiv des skeptischen Denkens ist die Abnabelung vom Ideal sicheren Wissens. Die zweite Motivation ist eine intellektuelle Neugierde,

die nicht bereit ist, Überzeugungen einfach beizubehalten oder abzulehnen, sondern diese aufgrund empirischer Evidenz bewerten möchte. Dies erfordert den Übergang von einer politisch-emotionalen Lebenshaltung, in der Meinungen durch die soziale Gruppe bestimmt sind, zu einem statistischen Lebensgefühl. Dieses ist relativ neu, da für viele Bereiche des menschlichen Lebens empirische Daten früher kaum vorhanden waren oder auch nicht gesucht wurden. John Arbuthnots statistischer Test war, trotz seiner offensichtlichen Mängel, ein früher Schritt in die Richtung, Überzeugungen durch Evidenz zu testen.

Im 18. und 19. Jahrhundert waren statistische Informationen meist Staatsgeheimnisse, die nur einer Elite bekannt waren und der Öffentlichkeit vorenthalten wurden. Die Bedeutung statistischer Informationen, etwa von Bevölkerungszahlen, wurde aber von den politisch Verantwortlichen erkannt. Napoleons Gier nach Daten aus seinem *bureau de statistique* war legendär (Bourget, 1987). Und er wollte die Zahlen immer sofort haben. In seiner Umgebung hieß es: Wenn du etwas von Napoleon willst, gib ihm Statistiken. Die Bereitschaft, wirtschaftliche und demographische Daten der Öffentlichkeit zugänglich zu machen, ist dagegen jüngerer Datums. Erst ab etwa 1830 wurden Statistiken veröffentlicht, zumindest einige. Seitdem hat eine „Lawine gedruckter Zahlen“, wie sich der Philosoph Ian Hacking ausdrückte, die heutige Welt in einen gewaltigen Ozean von Informationen verwandelt, der von Medien wie Fernsehen und Zeitschriften sowie vom Internet gespeist wird. Die zunehmende Verbreitung statistischer Informationen im 19. und 20. Jahrhundert korrelierte mit dem Aufkommen der Demokratien in der westlichen Welt.

9 Statistisches Denken statt statistischer Rituale

Statistisches Denken ist nicht nur skeptisch gegenüber der Illusion von Gewissheit, sondern auch gegenüber dem Umgang mit Statistik selbst. Es ist nützlich, zwei Umgangsweisen zu unterscheiden: statistisches Denken und statistische Rituale. Statistisches Denken ist selbstreflektiv; es beinhaltet die Abwägung, welche Methode oder welches Modell für eine Situation die beste ist und unter welchen Annahmen das gilt. Beispielsweise macht Laplaces Geschichte von der ersten Nacht im Paradies deutlich, dass die Anwendung der Regel von Bayes besser zu rechtfertigen ist, wenn man empirische Informationen über die Grundraten und die Wahrscheinlichkeiten hat, als wenn das nicht der Fall ist. Richter lassen heute die Regel von Bayes in Strafprozessen nur zu, wenn empirische Informationen vorhanden sind, und dann auch nicht immer. John Arbuthnots Gottesbeweis illustriert dagegen die Probleme und Grenzen des Nullhypothesentestens.

Statistische Rituale sind heute in den Sozialwissenschaften weit verbreitet – anders als in der molekularen Biologie, der Kernphysik oder den anderen naturwissenschaftlichen Disziplinen, wo diese nie Fuß fassen konnten.

Wenn auf jedes Problem ein- und dieselbe Methode unreflektiert angewendet wird, dann haben wir ein statistisches Ritual vor uns. Beispielsweise berechnet man in manchen Bereichen der pädagogischen Psychologie mechanisch Pfadanalysen, in Bereichen der Persönlichkeitspsychologie ebenso mechanisch Faktorenanalysen und in der experimentellen Sozialpsychologie erinnert das ständige Nullhypothesentesten an zwanghaftes Händewaschen. Jede Disziplin hat ihr eigenes statistisches „Überich“, das bei ihren Mitgliedern Verhaltensweisen erzeugt, die an eine Neurose erinnern. Man fühlt den Druck, die Methode anwenden zu müssen, denn ohne sie fühlt man sich nackt. Nur ein geringer Prozentsatz der akademischen Psychologen versteht überhaupt, was ein signifikantes Ergebnis bedeutet oder was man daraus schließen kann. Das gilt selbst für Dozenten, die Statistik für Psychologen lehren (Oakes, 1986; Haller & Krauß, 2002). Viele glauben irrtümlicherweise, ein signifikantes Ergebnis - wie bei Arbutnot - würde die Wahrscheinlichkeit angeben, dass die Nullhypothese richtig sei oder dass die Alternativhypothese falsch sei. Anders als die Regel von Bayes kann jedoch ein Nullhypothesentest keine Wahrscheinlichkeit für Hypothesen erbringen, lediglich eine Wahrscheinlichkeit für die Daten unter der Annahme, dass die Nullhypothese wahr ist. Alternative statistische Methoden, wie Neyman-Pearson Hypothesentests, Walds Sequentielle Tests, Tukeys „exploratory data analysis“ oder Bayes' Statistik, sind kaum bekannt, und es besteht auch geringes Interesse, diese kennen zu lernen. Statistiker wie R. A. Fisher und Jerzy Neyman haben die gedankenlose Anwendung ein- und derselben statistischen Methode immer wieder kritisiert, aber die Betroffenen scheinen dies nicht bemerkt oder schlicht verdrängt zu haben (Gigerenzer, 1993; Gigerenzer et al., 1999).

Die Evolution des statistischen Denkens hat nicht nur Probleme gelöst, sie hat auch neue Probleme erzeugt. Mit der Entwicklung der Theorie der Wahrscheinlichkeit wurde die Vielfalt von Konzepten, die im begrifflichen Umfeld von Zufall und Erwartung stand, auf einige ganz wenige eingengt. Der Begriff „probabilitas“ bedeutete ursprünglich eine durch Autorität gesicherte Meinung. Dieses Konzept wurde nicht zum Gegenstand der Theorie, genauso wenig wie die nahe liegenden Kandidaten Glück und Schicksal. Die drei Interpretationen von Wahrscheinlichkeit, mit der die Theorie begann, sind noch heute oft miteinander in Konflikt. Ist Wahrscheinlichkeit eine relative Häufigkeit in einer Referenzklasse von Ereignissen wie in statistischen Tabellen? Oder ist sie der Grad der subjektiven Überzeugung, die eine vernünftige Person hat? Oder ist sie durch das Design bestimmt wie die Konstruktion eines Würfels? Diese Frage trennt die Frequentisten, wie Richard von Mises und Jerzy Neyman, von den subjektiven Bayesianern, wie De Finetti und Leonard Savage, und von den Vertretern von Propensitäten wie Karl Popper. Die Antwort auf diese Frage bestimmt den Gegenstandsbereich der Theorie. Für einen Subjektivisten ist dieser unbegrenzt; alles in der Welt, wozu Menschen Überzeugungen haben, die den Gesetzen der

Theorie folgen, ist möglicher Gegenstand. Dies schließt Wahrscheinlichkeiten für Einzelfälle mit ein, selbst für Ereignisse, die noch nie beobachtet wurden, wie die erstmalige Anhebung der Lebenserwartung auf mehr als 100 Jahre. Für einen Frequentisten bezieht sich die Theorie nur auf Aussagen über Elemente einer Referenzklasse, die als untereinander gleich angesehen werden können und für die hinreichend viel statistische Information vorliegt. Aus dieser Sicht ist die Theorie auf Situationen wie dem HIV-Screening anwendbar, wo genügend Daten vorliegen, nicht aber auf die erste Nacht im Paradies. Aus der Sicht der Wahrscheinlichkeit als Propensität (Design) ist der Anwendungsbereich der Theorie noch kleiner: Sie betrifft nur Gegenstände, deren Bauplan oder kausale Struktur wir kennen.

Diese verschiedenen Auslegungen des Begriffs Wahrscheinlichkeit können unterschiedliche Abschätzungen des jeweiligen Risikos hervorrufen. Vor einigen Jahren nahm ich an einer Führung durch ein Werk der DASA (Daimler Benz Aerospace) teil, in dem die Ariane-Rakete hergestellt wird, die Satelliten in ihre Umlaufbahn befördert. Ich stand mit dem Führer vor einem großen Plakat, auf dem alle bis dahin abgeschossenen 94 Raketen (Ariane, Modelle 4 und 5) aufgeführt waren, und fragte ihn, wie hoch das Risiko eines missglückten Starts sei. Er erwiderte, der Sicherheitsfaktor betrage etwa 99,6%. Das erschien mir überraschend hoch, denn auf dem Plakat sah ich acht Sterne, die für acht Unfälle standen. Ich fragte also, wie acht Unfälle von 94 Starts einem Sicherheitsfaktor von 99,6% entsprechen könnten. Daraufhin erklärte er, die DASA zähle nicht die Fehlstarts, sondern berechne den Sicherheitsfaktor aus der Konstruktion der einzelnen Teile der Rakete. Die Fehlstarts zu zählen würde menschliches Versagen einbeziehen. Er fügte hinzu, dass beispielsweise einer dieser Sterne letztlich auf ein Missverständnis zurückgehe, nämlich zwischen einem Arbeiter, der eine Schraube weggelassen hatte, und seinem Kollegen von der nächsten Schicht, der annahm, sein Vorgänger habe die Schraube eingesetzt. Somit beruhte das genannte Risiko von Fehlstarts der Ariane-Raketen auf einer Design-Interpretation und nicht auf den tatsächlichen Häufigkeiten.

10 Vom Mut, sich seines eigenen Verstandes zu bedienen

Statistisches Denken ist ein Produkt der Zeit der Aufklärung. Der Philosoph Immanuel Kant begann seinen im Jahre 1784 verfassten Aufsatz „Zur Beantwortung der Frage: Was ist Aufklärung?“ folgendermaßen:

Aufklärung ist der Ausgang des Menschen aus seiner selbst verschuldeten Unmündigkeit. Unmündigkeit ist das Unvermögen, sich seines Verstandes ohne Leitung eines anderen zu bedienen. Selbstverschuldet ist diese Unmündigkeit, wenn die Ursache derselben nicht am Mangel des Verstandes, sondern der Entschließung und des Mutes liegt, sich seiner ohne Leitung eines anderen zu bedienen. Sapere aude! Habe Mut, dich deines eigenen Verstandes zu bedienen! ist also der Wahlspruch der Aufklärung.

Das sind klare und tiefe Gedanken. Der Schlüsselbegriff ist „Mut“. Dieser ist notwendig, weil man sich mit Hilfe des eigenen Verstandes nicht nur die Gefühle der Befreiung und der Selbstständigkeit verschaffen kann, sondern weil auch Strafe und Schmerz die Folge sein können. Kant selbst musste das erfahren. Einige Jahre nachdem er diese Sätze niedergeschrieben hatte, verbot ihm die Obrigkeit - aus Furcht, sein rationales Denken könne die Sicherheit der christlichen Lehre untergraben - weiterhin über religiöse Themen zu schreiben oder zu lehren. Ganz allgemein kann das Überwinden der Unmündigkeit bedeuten, dass man Lücken oder Widersprüche in Berichten, Tatsachen und Wertvorstellungen findet, an die man immer geglaubt hatte. Das Hinterfragen von Gewissheiten bedeutet oft das Hinterfragen von gesellschaftlicher Autorität.

Mit Ungewissheiten leben zu lernen, stellt für Einzelne wie auch für Gesellschaften eine große Herausforderung dar. Ein großer Teil unserer Geschichte wurde von Menschen geprägt, die sich völlig sicher waren, dass ihre Sippe, Rasse oder Religion die von Gott oder vom Schicksal erwählte war - und die für sich daraus das Recht ableiteten, abweichende Ideen zu bekämpfen wie auch die Menschen, die davon „befallen“ waren. Es war ein langer Weg zu den heutigen Gesellschaftsformen mit größerer Toleranz gegenüber Ungewissheit und Vielfalt. Trotzdem sind wir noch weit davon entfernt, die mutigen und informierten Menschen zu sein, die Kant vor Augen hatte - ein Ziel, das sich in zwei schlichten lateinischen Wörtern ausdrücken lässt: *Sapere aude*. Habe den Mut, selbst zu denken.

Literatur

- Arbuthnot, J. (1710). An argument for divine providence, taken from the constant regularity observ'd in the birth of both sexes. *Philosophical Transactions of the Royal Society*, 27, 186-190.
- Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society*, 53, 370-418. Übersetzung von: H. E. Timmerding (1908). Versuch zur Lösung eines Problems der Wahrscheinlichkeitsrechnung. Leipzig: Wilhelm Engelmann (Oswalds Klassiker der exakten Wissenschaften, 169).
- Bernoulli, D. (1738). Specimen theoriae novae de mensura sortis. *Commentarii academiae scientiarum imperialis Petropolitanae*, 5, 175-192. Englische Übersetzung von: L. Sommer (1954). Exposition of a new theory on the measurement of risk. *Econometrica*, 22, 23-36.
- Bourguet, M.-N. (1987). Décrire, compter, calculer: The debate over statistics during the Napoleonic period. In L. Krüger, L. Daston & M. Heidelberger (Eds.), *The probabilistic revolution: Vol I. Ideas in history* (pp. 305-316). Cambridge, MA: MIT Press.
- Daston, L. (1987). The domestication of risk: Mathematical probability and insurance 1650-1830. In L. Krüger, L. Daston & M. Heidelberger (Eds.), *The probabilistic revolution: Vol. I. Ideas in history* (pp. 237-260). Cambridge, MA: MIT Press.

- Daston, L. (1988). *Classical probability in the enlightenment*. Princeton, NJ: Princeton University Press.
- Fisher, R. A. (1935). *The design of experiments*. Edingborgh: Oliver and Boyd.
- Gigerenzer, G. (1993). Über den mechanischen Umgang mit statistischen Methoden. In E. Roth (Hg.), *Sozialwissenschaftliche Methoden* (3. Aufl., S. 607-618). München: Oldenbourg.
- Gigerenzer, G. (2002). *Das Einmaleins der Skepsis*. Berlin: Berlin Verlag.
- Gigerenzer, G., Hoffrage, U., & Ebert, A. (1998). AIDS counseling for low-risk clients. *AIDS CARE*, 10, 197-211.
- Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., & Krüger, L. (1999). *Das Reich des Zufalls*. Heidelberg: Spektrum.
- Hacking, I. (1975). *The emergence of probability*. Cambridge, MA: Cambridge University Press.
- Hacking, I. (1990). *The taming of change*. Cambridge, MA: Cambridge University Press.
- Haller, H., & Krauss, S. (2002). Misinterpretations of significance: A problem students share with their teachers? *Methods of Psychological Research Online*, 7, 1-20.
- Jorland, G. (1987). The St.-Petersburg-Paradox, 1713-1937. In L. Krüger, L. Daston & M. Heidelberger (Eds.), *The probabilistic revolution: Vol. I. Ideas in history* (pp. 157-190). Cambridge, MA: MIT Press.
- Krüger, L., Gigerenzer, G., & Morgan, M. (Eds.). (1987). *The probabilistic revolution: Vol. II. Ideas in the sciences*. Cambridge, MA: MIT Press.
- Laplace, P. S. (1812). *Théorie analytique des probabilités*. Paris: Courcier.
- Montmort, P. R. de. (1713). *Essai d'analyse sur les jeux de hasard* (2. Aufl.). Paris.
- Pascal, B. (1654/1970). Briefwechsel Pascal-Fermat. In B. Pascal, *Œuvres complètes* (S. 1136-1158), hg. von Jean Mesnard (*Œuvres diverses*, Bd. 1). Paris: Bibliothèque Européenne - Desclès de Brouwer.
- Pascal, B. (1669/1904). *Pensées*. Hg. von Léon Brunschwig (3 Bde.). Paris: Librairie Hachette. Deutsche Übersetzung von: W. Rüttenauer (1937). *Gedanken*. Leipzig: Dieterich'sche Verlagsbuchhandlung.
- Oakes, M. (1986). *Statistical inference: A commentary for the social and behavioral sciences*. New York: Wiley.
- Stigler, S. M. (1983). Who discovered Bayes theorem? *American Statistician*, 37, 290-196.

Anschrift des Autors

Prof. Dr. Gerd Gigerenzer, Max-Planck-Institut für Bildungsforschung,
Lentzeallee 94, 14195 Berlin
Email: sekgigerenzer@mpib-berlin.mpg.de