**Zeitschrift:** Bulletin de la Société Fribourgeoise des Sciences Naturelles = Bulletin der Naturforschenden Gesellschaft Freiburg

**Herausgeber:** Société Fribourgeoise des Sciences Naturelles

**Band:** 79 (1990)

**Heft:** 1-2

**Artikel:** Zufällige oder nicht zufällige Abweichungen? : Beispiele aus der Vorgeschichte der mathematischen Statistik

**Autor:** Ineichen, Robert

**DOI:** https://doi.org/10.5169/seals-308696

**Nutzungsbedingungen**

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

**Conditions d'utilisation**

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

**Terms of use**

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

**Download PDF:** 15.07.2025

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

# Zufällige oder nicht zufällige Abweichungen? Beispiele aus der Vorgeschichte der mathematischen Statistik

von ROBERT INEICHEN,
Mathematisches Institut der Universität Freiburg,
CH-1700 Freiburg

Die moderne mathematische Statistik stellt dem experimentell arbeitenden Forscher – sei er nun Naturwissenschafter, Mediziner, Pädagoge, Psychologe oder Betriebswirtschafter usw. – heutzutage ein umfangreiches Instrumentarium für die *Planung* und die *Auswertung* von Versuchen zur Verfügung. Oft spielt dabei die Durchführung eines statistischen Tests eine wichtige Rolle. Die allgemeine Theorie der statistischen Testverfahren, die heute zur Verfügung stehen, ist im wesentlichen in diesem Jahrhundert geschaffen worden, Beispiele von mehr oder weniger korrekt durchgeführten statistischen Tests findet man jedoch bereits in Publikationen des frühen 18. Jahrhunderts. Es handelt sich dabei in der Regel um Beispiele für sogenannte *Signifikanztests.* Solche Signifikanztests gehören wohl auch heute noch zu den wichtigsten statistischen Tests. Es sind dies Prüfverfahren, Tests, bei denen wir uns für die Ablehnung (Verwerfung) oder die (vorläufige) Beibehaltung einer *einzigen* Hypothese zu entscheiden haben. In unseren Ausführungen soll jenen frühen Beispielen und damit der *Vorgeschichte der mathematischen Statistik*[1] etwas nachgegangen werden. Mehrere in den

[1] Die *mathematische Statistik* baut auf der Wahrscheinlichkeitsrechnung auf. Sie umfaßt vor allem die Lehre von den Testverfahren und die Lehre von den Vertrauensintervallen (Konfidenzintervallen). In beiden Fällen werden *statistische Schlüsse,* Schlüsse aufgrund von Stichproben, gezogen, und es ist dabei möglich, das Risiko eines Irrtums zu berechnen und dieses Risiko in gewissen Grenzen zu halten. Es gibt andere Gebiete der Statistik, die nicht oder viel weniger auf mathematischen Methoden aufbauen, so etwa gewisse Teile der *amtlichen* Statistik, die oft nicht auf Stichproben, sondern auf Vollerhebungen beruhen. Im 18. Jahrhundert hatte Statistik auch die Bedeutung von «Staatsbeschreibung» oder «Staatskunde»; sie war Lehre von den «Staatsmerkwürdigkeiten», vom Zustand des Staates (status rei publicae).

Bull. Soc. Frib. Sc. Nat. *79 (1/2),* 96–127 (1990)

letzten Jahren erschienene Arbeiten, die einzelne derartige Beispiele beschreiben, ermöglichen es, heute einen zusammenhängenden Rückblick zu geben.

## 1. Die heutigen Überlegungen bei der Durchführung eines Signifikanztests

Stellen wir zunächst an Hand eines sehr einfachen Beispiels die Überlegungen zusammen, die heute gemacht werden – oder gemacht werden sollten –, wenn ein Signifikanztest durchgeführt werden muß. Gehen wir dazu von der folgenden Situation aus: Für die Messung eines gewissen Effektes sind zwei sehr verschiedene Meßmethoden entwickelt worden. Es sei möglich, diesen Effekt zu erzeugen und dann gleichzeitig seine Intensität mit beiden Methoden zu messen.

*1. Schritt:* Weil unsere beiden Meßmethoden sehr verschieden sind und weil bereits gewisse Vorerfahrungen zur Verfügung stehen, vermuten wir, die beiden Methoden würden auch zu wesentlich verschiedenen Resultaten führen, zu Resultaten, die *signifikant* (bedeutsam) voneinander abweichen. Können wir diese *Vermutung bestätigen?* Wir haben also eine *Vermutung* (wesentliche Unterschiede in den Meßresultaten) und eine *plausible Erklärung* dafür (nämlich die Verschiedenheit der Meßmethoden).

*2. Schritt:* Nehmen wir an, die beiden Methoden würden zu Resultaten führen, die nicht wesentlich voneinander abweichen. Dies müßte sich doch in einer genügend umfangreichen Serie von Messungen, in einer Stichprobe also, darin zeigen, daß die Resultate der beiden jeweils durchgeführten Messungen eng beieinander liegen. Prüfen wir, *testen* wir, ob dies wirklich der Fall ist! Mit dieser *Annahme* «Keine wesentlichen Abweichungen» haben wir nun eigentlich eine «Gegenthese» zu unserer ursprünglichen Vermutung formuliert; man nennt sie in der Fachsprache bekanntlich *Nullhypothese. Diese* Nullhypothese *testen* wir. Sind dann die Resultate so, daß sie die Nullhypothese widerlegen, daß wir die *Nullhypothese verwerfen* müssen, dann ist dadurch unsere *Vermutung indirekt bestätigt.*

*3. Schritt:* Wir planen ein entsprechendes *Zufallsexperiment.* Wir beschließen z. B., den zu messenden Effekt zehnmal zu erzeugen und ihn jedesmal nach beiden Methoden zu messen. Wir werden dann zwei Reihen von Meßresultaten erhalten, zum Beispiel:

| | $x_i$ (Resultate mit der ersten Methode) | $y_i$ (Resultate mit der zweiten Methode) | $d_i = x_i - y_i$ |
|---|---|---|---|
| | 558 | 560 | – 2 |
| | 559 | 558 | 1 |
| usw. für i = 1 bis 10 | ... | ... | ... |
| | 562 | 562 | 0 |
| | 564 | 561 | 3 |

Wenn unsere Nullhypothese zutrifft, so müßten die Differenzen $d_i$ im Idealfall den Mittelwert 0 ergeben. Natürlich müssen wir aber damit rechnen, daß der berechnete Mittelwert $\bar{d}$ etwas von Null abweicht. Nur wenn er *zu stark* von Null abweicht, werden wir die Nullhypothese als widerlegt betrachten, werden wir sie verwerfen. Dieses «zu stark» müssen wir jetzt präzisieren: Die «zu starken» Abweichungen dürften natürlich nur mit einer sehr kleinen Wahrscheinlichkeit $\alpha$ auftreten, wenn die Nullhypothese doch zutreffen würde. Je nachdem, was auf dem Spiel steht, wählt man für $\alpha$ den Wert 5% oder 1% oder gar nur 1‰. $\alpha$ heißt *Signifikanzniveau,* manchmal auch Irrtumswahrscheinlichkeit. Die Gesamtheit der Resultate, die zum Verwerfen der Nullhypothese – und damit zur Bestätigung der Vermutung – führen könnten, bilden den *Verwerfungsbereich.* Er läßt sich leicht angeben, wenn eine entsprechende Formelsammlung mit den zugehörigen Tafeln zur Verfügung steht und verursacht – dank der Vorarbeit der Mathematiker – keine große Rechenarbeit [2].

*4. Schritt:* Wir führen das Zufallsexperiment durch (oder ziehen die Resultate eines bereits früher durchgeführten Zufallsexperimentes heran [3]). Nun gibt es zwei Möglichkeiten:
*Erste Möglichkeit:* Das Ergebnis (hier also $\bar{d}$) fällt *in den Verwerfungsbereich.* Jetzt verwerfen wir die Nullhypothese und betrachten die *ursprüngliche Vermutung,* die wir im ersten Schritt genannt haben, als *bestätigt.* – Wir verwerfen dabei die Nullhypothese aus zwei Gründen:

[2] Sammlungen von entsprechenden Tafeln findet man zum Beispiel in
– DMK/DPK: Formeln und Tafeln, Unterrichtswerk der deutschschweiz. Mathematikkommission (DMK) und der deutschschweiz. Physikkommission (DPK). Zürich 1988; oder in
– SACHS, LOTHAR: Statistische Methoden – Ein Soforthelfer. Berlin 1970.

[3] Das Material eines «bereits früher durchgeführten Zufallsexperimentes» sollte indessen nicht jenes sein, das uns auf die Vermutung (auf die Arbeitshypothese) geführt hat. Der Test sollte eine *echte Frage* an die Natur oder an die technische oder soziale Umwelt sein.

Erstens, weil die *Wahrscheinlichkeit* für unser Ergebnis nur *sehr klein* wäre, wenn die Nullhypothese zutreffen würde, *und*
zweitens, weil wir eine *andere, eine plausible Erklärung* für unser Ergebnis haben.
Es ist dabei sehr wichtig, die *beiden* Gründe für die Verwerfung zu beachten. Hätten wir nur den ersten dieser Gründe, so müßten wir zum Beispiel die Lotteriegesellschaft sofort einklagen, wenn uns morgen in einer Lotterie das große Los zufallen würde: Die Nullhypothese «faire Ziehung» wäre zu verwerfen! Auf diesen Gesichtspunkt wird übrigens selbst in der Fachliteratur oft viel zu wenig hingewiesen. Besonders deutlich wird etwa von W. STEGMÜLLER (1973) die Wichtigkeit auch dieses zweiten Grundes betont:
«Verwirf eine Hypothese h nicht bereits dann, wenn das, was sich tatsächlich ereignet hat, sich unter der Annahme der Richtigkeit von h sehr selten ereignet: *verwirf* h *nur dann, wenn du eine bessere Hypothese hast.*» Es scheint, daß P.S. de Laplace 1814 als erster explizit auf diesen Imperativ hingewiesen hat [4].
Eine kleine Irrtumswahrscheinlichkeit $\alpha$ nehmen wir bei diesem Verwerfen der Nullhypothese in Kauf. $\alpha$ stellt die Wahrscheinlichkeit dar, daß das Resultat in den Verwerfungsbereich fällt, obwohl die Nullhypothese zutrifft. Wir würden in diesem Falle also die Nullhypothese verwerfen, sollten sie aber eigentlich beibehalten.

[4] So schreibt Laplace im *Essai philosophique des probabilités,* den er 1814 der zweiten Auflage seiner grundlegenden *Théorie analytique des probabilités* als Einleitung voranstellt: «Wenn wir dort, wo wir eine Symmetrie bemerken, nach einer Ursache forschen, geschieht dies nicht deshalb, weil wir ein symmetrisches Ereignis für weniger möglich als die anderen halten; aber vor die Wahl gestellt, ob wir dieses Ereignis als Wirkung einer regelmäßigen Ursache oder des Zufalls ansehen sollen, erscheint uns die erste dieser Ansichten wahrscheinlicher als die zweite. Wir erblicken auf einem Tisch Lettern in dieser Ordnung zusammengestellt: KONSTANTINOPEL, und wir betrachten diese Anordnung nicht als Wirkung des Zufalls, nicht, weil sie weniger möglich ist als die anderen, da wir diesem Wort, falls es in keiner Sprache verwendet würde, keine besondere Ursache zumuten würden; da aber dieses Wort bei uns im Gebrauch ist, so ist es unvergleichlich wahrscheinlicher, daß irgendeine Person jene Lettern so angeordnet hat, als daß diese Anordnung nur dem Zufall zu verdanken ist» (LAPLACE, 1932).

Deutlich weist z. B. NOETHER (1971) in dieselbe Richtung: «Let us return briefly to the example where we observe 10 heads in 10 tosses of a coin. While it is true that such an occurence has only a very small probability when tossing a fair coin, *this by itself is not sufficient reason to reject the null hypothesis. Equally important is the fact that a much more plausible explanation exists,* namely cheating.»

*Zweite Möglichkeit:* Das Ergebnis (also $\bar{d}$) fällt *nicht* in den Verwerfungsbereich. Damit ist nun allerdings die Nullhypothese nicht etwa bewiesen. Wir haben einfach keinen Grund, keine Handhabe, sie zu verwerfen; wir behalten sie deshalb – bis auf weiteres – bei. Wir sind in der Lage eines Richters, der einen Angeklagten mangels Beweisen freisprechen muß. Es kann sein, daß sich später zeigt, daß der Angeklagte doch schuldig ist. Weitere, vielleicht umfangreichere Versuche zeigen uns eventuell nachher, daß die Nullhypothese doch nicht zutrifft. Auch dieser Sachverhalt wird oft zu wenig genau erfaßt. Dabei wird er schon im grundlegenden Buch von Ronald Aylmer Fisher (1890–1962), einem der erfolgreichsten Begründer der modernen Testtheorie und ihrer Anwendungen, in seinem *Design of Experiments,* sehr deutlich ausgesprochen (FISHER, 1935):
«... it should be noted that the null hypothesis is never proved, but is possibly disproved in the course of experimentation. Every experiment may be said to exist only to give the facts a chance of disproving the null hypothesis ...».

Man ist also bei einem Signifikanztest eigentlich nur dann auf der guten Seite, nur dann in einer starken Position, wenn man die Nullhypothese verwerfen kann. Dies ist übrigens in den Naturwissenschaften meistens der Fall, wenn man eine Hypothese durch ein Experiment prüft, auch dann, wenn man die Hilfsmittel der mathematischen Statistik nicht ausdrücklich benützt: Man ist sehr oft eher dann auf der sicheren Seite, wenn das Experiment eine Hypothese widerlegt, als wenn es uns veranlaßt, sie – vorläufig! – beizubehalten. Dazu zwei Zitate:

– Der französische Mathematiker und Wissenschaftsphilosoph Henri Louis Poincaré (1853–1912) sagt (nach JAKOB, 1948): «Die Erfahrungen durch das Experiment dienen dazu, Hypothesen zu widerlegen, nicht etwa, sie zu beweisen.»

– Und Albert Einstein schreibt (nach PITSCHMANN, 1980): «Wir stellen viele Fragen an die Natur; die Natur sagt meistens ‹nein›, manchmal ‹vielleicht›, niemals ‹ja›.» Hier liegt auch ein Grund für das typische Vorgehen bei der Durchführung eines Signifikanztests: Wir möchten gerne unsere Vermutung bestätigen (die vielleicht unsere Arbeitshypothese ist) und können dies oft nicht direkt tun. Deshalb stellen wir eine «Gegenthese», die Nullhypothese, auf und versuchen, sie zu *widerlegen,* um so in die eben erwähnte «starke Position» zu gelangen.

An der Schaffung der notwendigen mathematischen Grundlagen der Mathematischen Statistik, vor allem auch der Theorie der statistischen Tests, sind außer R. A. Fisher noch zahlreiche andere Mathematiker beteiligt. Zu nennen wären wohl in erster Linie der Engländer Egon Sharpe Pearson (1895–1980), weiter der in Polen, dann in England und in den USA wirkende Jerzy Neyman (1894–1981) und schließlich Abraham Wald (1902–1950), der in den USA während des Zweiten Weltkrieges entscheidende

Fortschritte in der mathematischen Statistik erzielte, die erst 1947 publiziert werden konnten. Auf unserem Kontinent haben sich die Methoden der modernen mathematischen Statistik relativ spät, auf alle Fälle erst nach dem Zweiten Weltkrieg, langsam durchzusetzen begonnen: das erste Lehrbuch der modernen mathematischen Statistik in deutscher Sprache publizierte ARTHUR LINDER, damals Professor für angewandte mathematische Statistik an der Universität Genf und Dozent an der ETH in Zürich, im Jahre 1945 unter dem Titel «Statistische Methoden für Naturwissenschafter, Mediziner und Ingenieure». – Und noch 1960 gab Walter Saxer, damals Professor an der ETH in Zürich, an der Tagung der Schweizerischen Naturforschenden Gesellschaft zu bedenken: «Das immer stärkere Überhandnehmen der mathematischen Statistik auf den Gebieten der Physik, der Technik, der Biologie und der Medizin erfordert von den Schweizer Hochschulen vermehrtes Eingehen auf die Rechenverfahren der mathematischen Statistik ... Die heutige Entwicklung auf dem Gebiet der mathematischen Statistik und speziell deren Übergreifen auf Medizin und Biologie erfordert aber dringend, daß an sämtlichen Schweizer Gymnasien Grundlagen der Wahrscheinlichkeitsrechnung vermittelt werden» (W. NIEDERBERGER, 1961). Das Postulat Saxer dürfte heute einigermaßen erfüllt sein. [5]

Gehen wir nun also der Frage nach, ob und wie die Überlegungen, die wir bei unseren Signifikanztests machen, auch schon in früheren Jahrhunderten ausgeführt worden sind.

[5] Zum Beispiel sind an der Mathematisch-Naturwissenschaftlichen Fakultät der Universität Freiburg (Schweiz), gegründet 1896, zwischen 1909 und 1961 nur sporadisch Vorlesungen zur Wahrscheinlichkeitsrechnung, zur Fehlerrechnung und zur mathematischen Statistik angeboten worden (zunächst durch M.F. Daniels, dann durch P. Lambossy). Erst seit 1964 sind regelmäßig Einführungen in die Wahrscheinlichkeitsrechnung und Statistik angeboten worden (R. Ineichen), die vor allem auch für Naturwissenschafter bestimmt waren. 1972 wurde erstmals ein Ordinarius für mathematische Statistik an die Fakultät berufen: P. Thullen; sein Nachfolger ist A. Antille (seit 1976).

Eine der ersten deutschsprachigen Darstellungen dieser Gebiete für den Gebrauch an höheren Schulen dürfte die Schrift *Einführung in die elementare Wahrscheinlichkeitsrechnung und Statistik* von R. INEICHEN (Luzern 1962) sein (7., vollständig überarbeitete Auflage unter dem Titel *Stochastik,* Luzern 1989).

## 2. Erste Spuren

### *2.1. Analogien und Parallelen*

Die folgende Tatsache darf nicht außer acht gelassen werden: Wenn wir einen statistischen Schluß in Form eines Signifikanztestes *mit* Verwerfen der Nullhypothese durchführen, so arbeiten und argumentieren wir über weite Strecken so, wie wir im Alltag – unbelastet von moderner statistischer Methodik – auch verfahren würden. Wir geben im Alltag bloß kein Signifikanzniveau an und präzisieren den Verwerfungsbereich nicht; das wäre dann eine Angelegenheit der modernen mathematischen Statistik.

So versuchen wir doch oft, eine Aussage zu widerlegen, indem wir ein *Gegenbeispiel* vorlegen. Sobald es präsentiert wird, wird die Aussage als falsch erkannt. Jenes Stichprobenresultat eines Testes, das nur mit geringer Wahrscheinlichkeit eintritt, wenn die Nullhypothese zutrifft, spielt doch – *wenn* es tatsächlich eintrifft – die Rolle des Gegenbeispiels. Doch tritt an Stelle des «nein» ein «fast unmöglich». Und sobald ein solches Stichprobenresultat vorliegt, zweifeln wir so stark an unserer Nullhypothese, daß wir sie verwerfen, verwerfen allerdings *auch deshalb,* weil meistens eine andere, plausible Erklärung zur Verfügung steht.

Oder wir neigen im Alltag schließlich dazu, *Erklärungen* (Begründungen, eben Nullhypothesen!) für ein Geschehnis zu verwerfen, wenn dieses Geschehnis aufgrund dieser Erklärung äußerst unwahrscheinlich scheint. Dazu findet man schon bei M. T. Cicero (106–43 v. Chr.) Gedanken ausgesprochen, die durchaus an entsprechende Formulierungen in Lehrbüchern der modernen mathematischen Statistik erinnern könnten. So etwa in *De divinatione,* 1.23: «Vier geworfene Astragale können durch Zufall einen Venuswurf bewirken. Glaubst du etwa, daß auch hundert Venuswürfe durch Zufall entstehen werden, wenn du vierhundert Astragale wirfst?» [6]

Schließlich müssen wir auch die Analogien beachten, die zwischen dem *indirekten Beweis,* den wir ja in der Mathematik immer wieder verwenden, und unseren statistischen Schlüssen in Form eines *Signifikanztests mit Verwerfen der Nullhypothese* bestehen: In beiden Fällen liegt zunächst eine *Vermutung* vor. Man versucht, die Vermutung *indirekt zu beweisen,* bzw. beim Signifikanztest *indirekt zu bestätigen.* Beim indirekten Beweis nimmt man nun das (kontradiktorische) *Gegenteil* der zu beweisenden Vermutung

[6] Cicero, *De divinatione* 1.23: Quicquam potest casu esse factum quod omnes habet in se numero veritatis? Quattuor tali iacti casu venerium efficiunt; num etiam centum venerios, si quadringentos talos ieceris, casu futuros putas?
Venuswurf: Jeder der vier Astragale zeigt eine andere Seite.

an und zeigt, daß diese Annahme auf einen Widerspruch führt. Somit kann unsere Annahme nicht richtig sein. Die Vermutung muß also zutreffen; sie ist damit indirekt bewiesen. Und beim Signifikanztest formuliert man – ganz analog – eine *«Gegenhypothese»* zur ursprünglichen Vermutung, eben die *Nullhypothese.* Wenn die Nullhypothese zutreffen würde, so würde das Resultat unseres Zufallsexperimentes *fast sicher nicht* in den Verwerfungsbereich fallen. Nun fällt es aber *doch* in den Verwerfungsbereich. Also ist es «praktisch unmöglich», daß die Nullhypothese zutrifft. Da wir zudem eine plausible Erklärung für das eingetretene Ereignis zur Verfügung haben, wird die Nullhypothese verworfen und die Vermutung wird als indirekt bestätigt angesehen. Indirekte Beweise wurden bereits in der *Antike* sehr oft verwendet; auch in den Elementen Euklids (ca. 300 v. Chr.) findet man häufig indirekte Beweise; «... diese Schlußweise galt zu Platons Zeit ... als die mathematische Beweisform schlechthin» (SZABO, 1969 [7]).

Betrachten wir alle diese Analogien und Parallelen, die sich noch vermehren ließen, so dürfen wir uns eigentlich nicht allzu sehr wundern, auch lange vor dem Aufkommen der *modernen* mathematischen Statistik mit ihren Schätz- und Testverfahren bereits frühe Spuren dieser Methodik zu finden.

### *2.2. Verwerfen von Hypothesen*

Man findet aber noch deutlichere Spuren, Überlegungen, in denen ausdrücklich eine Hypothese verworfen wird, weil ein Ereignis eingetreten ist, das bei Gültigkeit dieser Hypothese nur eine sehr kleine Wahrscheinlichkeit gehabt hätte und – wiederum! – weil eine andere Hypothese plausibel erscheint.

Was zunächst die wiederholte Durchführung einer Messung anbelangt, so weisen INVREA (1936) und unabhängig von ihm BOLZAN (1972) auf ein sehr frühes Beispiel einer *mehrfachen Messung* bzw. einer mehrfachen Schätzung hin, um einen möglichst genauen Wert zu erhalten. Thukidides erzählt nämlich in seiner *Geschichte des Peloponnesischen Krieges,* wie die eingeschlossenen und belagerten Platäer Leitern «von der Höhe der feindlichen Mauer» benötigten, um einen Ausbruchsversuch zu unternehmen (428 v. Chr.). Sie nahmen dazu «das Maß nach den Lagen der Backsteine an solchen Stellen, wo dieselben ... nicht verputzt waren ... *mehrere mußten zählen,* damit, wenn der eine oder andere unrichtig zählte, doch die meisten die Rechnung richtig treffen möchten. Dieses konnte um so viel eher geschehen, da sie solche zu *wiederholten Malen* nachzählten, auch die Entfernung nicht eben groß war ...» (THUKIDIDES, o.Jg.). Sicher eine für den mathema-

[7] Vergleiche dazu auch Platons Dialog Theätet (162e, ff.).

tischen Statistiker bemerkenswerte Stelle im Werk von Thukidides! INVREA (1936) sieht darin geradezu *«un applicazione avveduta della legge dei grandi numeri»!*

In der *Historia Francorum* von Gregor von Tours aus dem 6. Jahrhundert (HØYRUP, 1983) wird erzählt wie zwanzig Krieger, «die Gott nicht fürchteten», über einen Fluß setzten, um ein dem Hl. Martin geweihtes Kloster zu plündern und die Mönche zu töten. Sie luden die Beute in ihr Boot und wollten ans andere Ufer zurückfahren. Doch das Boot wurde ständig hin und her gewirbelt. Sie verloren ihre Ruder und versuchten schließlich, das Boot mit ihren Speerschäften ans Ufer zu manövrieren. Aber das Boot zersplitterte, und sie wurden von ihren eigenen Lanzen durchbohrt und getötet.» Und Gregor fährt fort: «Unus tantum ex ipsis, qui eos increpabat ne ista committerent, remansit inlaesus. Quod si hoc quis *fortuitu* evenisse iudicat, *cernat, unum insontem plurimis evasisse de noxiis.»* Dies besagt etwa: «Nur einer von ihnen blieb unverletzt, jener, der die andern wegen ihres Vorhabens getadelt hatte. Wenn jemand denkt, dies sei durch Zufall geschehen, so möge er bedenken, daß *ein unschuldiger* Mann gerettet wurde *inmitten so vieler Verbrecher.»* Natürlich wird auch hier noch nicht gerechnet; es wird *kein* Signifikanzniveau angegeben. Aber man könnte doch in moderner Sprechweise sagen: Es wird ausdrücklich eine Nullhypothese (fortuitu: durch Zufall) formuliert, und für Gregor scheint es offensichtlich, daß sie verworfen werden muß.

Ganz ähnliche Überlegungen findet man bei Girolamo Cardano (1501–1576) in seiner Autobiographie *Hieronymi Cardani Mediolanensis De propria vita liber* im 37. Kapitel (CARDANO, 1969): «Wenn ich in irgendeiner großen Sache und wichtigen Angelegenheit würfle, und die guten und ehrlichen drei Würfel werfen nur drei Punkte, so ist dies etwas ganz Natürliches und darf auch für etwas Natürliches gehalten werden, auch dann noch, wenn sich das nämliche beim zweiten Wurf wiederholt. Wenn aber derselbe Wurf ein drittes und ein viertes Mal nacheinander fällt, so darf dies auch der vernünftigste Mensch für verdächtig halten. Und ebenso ist es in den Fällen, wo schon alle Hoffnung aufgegeben war und dann so plötzlich eine Rettung aufleuchtete, daß man ganz deutlich sehen mußte, daß es durchaus Gottes sichtbarer Wille war.» An sich hätte Cardano die hier in Frage kommenden Wahrscheinlichkeiten berechnen können: in seinem *Liber de ludo aleae* (ca. 1564) berechnet er völlig richtig die Anzahl der Variationen beim Spiel mit drei Würfeln (216); er bestimmt öfters Wahrscheinlichkeiten, indem er die Anzahl der günstigen Fälle durche die gesamte Zahl der gleichmöglichen Fälle dividiert (allerdings ohne die heutige Terminologie zu verwenden), und er kennt die Formel für die Wahrscheinlichkeit, in drei unabhängigen Wiederholungen drei Treffer zu erhalten (CARDANO, 1663; ORE, 1953).

SCHNEIDER (1976) weist darauf hin, «daß die später bei Arbuthnot und de Moivre erkennbare Idee des Tests einer statistischen Hypothese mindestens schon durch Kepler vorbereitet ist.» Er bezieht sich dabei auf den *«gründlichen Bericht über den neuen Stern»*, den Johannes Kepler (1571–1630) im Anschluß an die Beobachtung einer Nova anläßlich der Konjunktion von Saturn, Jupiter und Mars (Oktober 1604) verfaßte. Um die *Zufälligkeit* des Zusammentreffens von großer Konjunktion und Nova zu *widerlegen,* vergleicht Kepler in diesem Bericht das zeitliche und örtliche Zusammentreffen mit dem Auftreten eines Vierer- und Fünferpaschs beim Werfen gewöhnlicher Spielwürfel: «... dan ob wol war/ das (zum exempel) ein jeder gerader wolgemachter würffel sechs felder hat/ und eins so wol fallen khan/ als das andere/ jedoch wan ein anzahl spieler jeder mit vier oder fünff Würffeln nur einen einigen Wurff thuen sollten/ und einem under jnen füele das Sechsen auff allen würffeln/ so würde man ein sollichen nit unbbilich wegen einer verborgenen kunst verdacht haben/ und es schwärlich dem glück zuschreiben: angesehen/ das wol hundert tausendt würffe geschehen möchten/ ehe wieder einer auff diese weise geriehte.»

Vier Sechser in einem Wurf mit vier Würfeln ist in der Tat ein seltenes Ereignis; die Wahrscheinlichkeit dafür ist $1/1296 \approx 1‰$. Um zum Beispiel mit einer Wahrscheinlichkeit von 99% beim Werfen von jeweils vier Würfeln mindestens einmal ein Viererpasch zu erzeugen, müßte man mindestens etwa 6000 Würfe mit jeweils vier Würfeln ausführen. (Für ein Fünferpasch wären entsprechend rund 35 800 Würfe mit fünf Würfeln notwendig.) Kepler überschätzt also die notwendige Anzahl der Würfe sehr stark. Seine Argumentation hat trotzdem einiges für sich: «Derowegen wie gesagt/ ich diese wunderbarliche eintreffung der zeit und ort/ nit gern dem blinden glück zuschreiben wolte: zumahl weil die erscheinung selbsten eines neuen Sternens für sich allein (auch ohne betrachtung der zeit und ort) nit ein gemein ding ist/ wie ein spielwurff/ sondern ein grosses wunder/ desgleichen vor unsern zeiten nie erhört oder gelesen worden.»

Aus heutiger Sicht der Dinge könnte man sagen: Kepler entscheidet sich *gegen die Nullhypothese der Zufälligkeit* und möchte das seltene und seltsame Ereignis der doppelten Konjunktion verbunden mit dem Auftreten einer Nova «nit gern dem blinden glück zuschreiben»; er neigt zur Alternative: «ein grosses wunder». Er versucht zudem durch den Vergleich mit dem Vierer- oder Fünferpasch das Risiko abzuschätzen, das darin bestehen würde, die Nullhypothese zu verwerfen, obwohl sie richtig wäre: Er gibt einen *geschätzten Wert für das Signifikanzniveau,* allerdings einen viel zu kleinen.

STIGLER (1977) berichtet über eine andere, weit zurückführende und sehr interessante Spur in England, die ebenfalls auf eine Art Signifikanztest hinweist: Seit rund 800 Jahren wurde in England «The Trial of the Pyx» zur

Überprüfung der Qualität der Gold- und Silbermünzen durchgeführt, die in der Königlichen Münzstätte hergestellt wurden. Aus 15 Pfund Gold- und 60 Pfund Silbermünzen der Produktion wurde jeweils je eine Münze herausgegriffen und in die «Pyx» gelegt (gr.πυξίς, Dose). In unregelmäßigen Zeitabständen wurde die Pyx durch die Aufsichtskommission geöffnet, die Münzen wurden gezählt, gewogen und auf ihren Gehalt geprüft. Die Resultate verglich man mit den Standards, die in der Westminster Abbey wohl verwahrt wurden. Eine gewisse Toleranz wurde jeweils zugestanden, zu große Abweichungen führten zu entsprechenden Maßnahmen. Leider scheinen die nähern Umstände dieser Kontrolle nicht bekannt zu sein: Wie wurde die Stichprobe erhoben, wie groß waren die zulässigen Toleranzen usw.?

N.L. Rabinovitch ist in mehreren Arbeiten unter anderem jenen Spuren stochastischen Denkens nachgegangen, die sich im *Talmud und seinen Kommentaren* zeigen; er hat solche zahlreich gefunden. In unserem Zusammenhang interessieren Beispiele, in welchen von zwei Hypothesen eine abgelehnt wird. Abgelehnt, weil ein Ereignis eingetroffen ist, für das nur eine kleine Wahrscheinlichkeit bestehen würde, wenn die Hypothese richtig wäre *und* weil eben eine andere Hypothese zur Verfügung steht, die plausibler, einleuchtender erscheint. So etwa nach RABINOVITCH (1977): Der Jerusalemer Talmud erzählt, wie Moses einen Rat der Ältesten von siebzig Mitgliedern zu bestimmen hatte. Aus jedem der zwölf Stämme wurden zunächst sechs Kandidaten ausgewählt; aus der Schar dieser 72 Kandidaten waren nun zwei zu eliminieren. Dazu wurden 72 Kugeln vorbereitet; 70 davon wurden mit «Ältester» markiert, zwei blieben unmarkiert. Die Kugeln wurden in eine Urne gelegt und gemischt, und jeder Kandidat hatte eine Kugel zu ziehen; jene beiden, die die unmarkierte erwischten, wurden eliminiert. Um 300 n. Chr. kommentierte Rabbi Abbahu: Wenn die nicht markierten Kugeln in gleichen Intervallen erscheinen – und zwar in einem einzelnen Wahlakt –, so ist dies sehr bemerkenswert und kann kaum durch Zufall erklärt werden: man hat dann in diesem Ereignis das Eingreifen Gottes zu sehen. (Weitere Beispiele auch bei RABINOVITCH, 1973.)

Selbstverständlich ist weder im Talmud selbst noch im Kommentar die entsprechende Rechnung durchgeführt. Doch kann man wohl – mit N.L. Rabinovitch – den Kommentator nur begreifen, wenn man ihm stochastisches Denken zubilligt. Und in der Tat: Die 72 Kugeln, wovon 70 unter sich gleiche und zwei unter sich gleiche, lassen sich auf 2556 Arten anordnen. Gleiche Intervalle ergeben sich nur, wenn die unmarkierten Kugeln an der ersten und der 37. Stelle oder wenn sie an der 36. und der 72. Stelle stehen. Die Wahrscheinlichkeit, daß bei einem einzelnen Wahlakt gleiche Intervalle entstehen, ist somit $2/2556 = 1/1278 \approx 1‰$!

Es sind uns keine Beispiele bekannt, daß auch die schlechten Würfel von den guten schon im Mittelalter durch ein Verfahren getrennt worden sind, das wenigstens der Idee nach an einen modernen Signifikanztest erinnern würde. Die einfachsten kombinatorischen Grundlagen dazu wären in den ersten Jahrzehnten des 14. Jahrhunderts auf alle Fälle vorhanden gewesen: Man findet sie zum Teil bei den ersten Dante-Kommentatoren oder im sogenannten Pseudoovidius *De Vetula* aus der Mitte des 13. Jahrhunderts (INEICHEN, 1988, 1990). Fest steht aber, daß man den guten Würfel schon seit dem 13. Jahrhundert in verschiedenen Darstellungen ausdrücklich gefordert (INEICHEN, 1991) und daß man Spieler mit falschen Würfeln hart bestraft hat. So verordnet zum Beispiel das *Zürcher Stadtrecht* bereits um 1300 (TAUBER, 1987): «Wer dem andern mit valschen Würfeln sin guot angewinnet und das erzüget wird/.../ der sol es wider geben, das er im angewinnet und darzu also vil geben und der stat V pfunt. Were es aber ein buobe, den sol man swemmen und sol V jar von der stat sin.» (Buobe: ein notorischer Spieler, swemmen: in einem Korb unter Wasser tauchen.) Das Brünner Stadtrecht aus dem 14. Jh. verlangt viel drakonischer: «... dem schol man den daumen abslahen ...».

Zum Schluß wollen wir noch auf eine eigenartige wahrscheinlichkeitstheoretische Betrachtung in den *Pensées* von Blaise Pascal (1623–1662) hinweisen. Auch hier wird eine Hypothese verworfen; es werden dabei aber nicht jene Überlegungen gemacht, die für einen Signifikanztest charakteristisch sind: Man entscheidet sich für jene von zwei Hypothesen, deren Annahme die *Gewinnerwartung maximalisiert.* Pascals Überlegungen finden sich in einem Fragment seiner *Pensées,* überschrieben mit *Infini-rien: Le Pari* (Nr. 451, PASCAL, 1954). Pascals Überlegungen sind kompliziert und nicht leicht zu überblicken. Wir wollen nur *einen* Aspekt herausgreifen, eine ausführliche Diskussion der ganzen Argumentation gibt HACKING (1975). Pascal will beweisen, daß man sich für den Gott der Christen zu entscheiden hat mit all' seinen Konsequenzen, die dieser Entscheid impliziert: «Dieu est ou il n'est pas». Wenn der Gott der Christen wirklich existiert und du also die Hypothese «Dieu est» annimmst, so hast du «eine Unendlichkeit unendlich glücklichen Lebens zu gewinnen; eine Möglichkeit (unendlichen) Gewinnes gegen eine (nur) endliche Zahl von Verlustmöglichkeiten, und das, was du einsetzest ist (selbst) endlich. Das hebt alle (wirkliche) Gegenüberstellung (wägbaren Gewinnes und Verlustes) auf. Überall, wo das Unendliche (Gegenstand möglichen Gewinnes) ist und wo keine Unendlichkeit von Verlustmöglichkeiten gegen die Möglichkeit des Gewinnes steht, gibt es kein Schwanken, man muß alles einsetzen» (Übersetzung von GUARDINI, 1962).

Schematisch und mit moderner Terminologie können wir diesen Aspekt von Pascals Argumentation so darstellen:

- Hypothese $H_1$: «Dieu est»; die Wahrscheinlichkeit dafür sei p, wobei $0 < p << 1$.
- Hypothese $H_2$: «Dieu n'est pas»; die Wahrscheinlicheit dafür ist dann (1-p).

Nun *müssen* wir auf $H_1$ oder $H_2$ setzen: «... il faut parier. Cela n'est pas volontaire, vous êtes embarqué» schreibt Pascal. Setzen wir auf $H_1$, so impliziert dies nach Pascal auch, daß wir so fromm leben, daß wir zum Glauben kommen.
Tragen wir nun Gewinn und allfällige Verluste (Höllenstrafen?) in eine Matrix ein, so erhalten wir

| | $H_1$ wahr (mit Wahrsch. p) | $H_2$ wahr (mit Wahrsch. 1–p) |
|---|---|---|
| ich setze auf $H_1$ | $\infty$ | a |
| ich setze auf $H_2$ | b | c |

Das Symbol $\infty$ steht dabei für den unendlichen Gewinn; a, b, c stehen für allfällige endliche Gewinne oder Verluste. Wie klein nun auch p ist, die Gewinnerwartung wird maximalisiert, wenn wir auf $H_1$ setzen: In mathematisch sehr fragwürdiger Form könnte man diese unendlich große Gewinnerwartung durch $p \cdot \infty + (1-p) \cdot a$ darstellen – *wenn* man mit dem Symbol $\infty$ einfach wie mit einer Zahl rechnen dürfte!

GUARDINI (1962) gibt eine ausführliche Darstellung von Pascals Überlegungen und ordnet sie auch in den geschichtlichen Zusammenhang ein. Pascals Gedanken in *Infini-rien* stellen jedoch nicht den Unterbau von Pascals eigenem Glauben dar; Pascal will damit die *Skeptiker* überzeugen. Ob ihm dies gelungen ist? HACKING (1975) meint: «These... arguments are all valid. None are convincing. All rely on dubious premises». Und doch sind sie bemerkenswert: Sie enthalten wesentliche Grundgedanken der modernen Entscheidungstheorie, jener Theorie also, die es heute in vielen Fällen ermöglicht, auch bei großer Ungewißheit noch rational begründbare Entscheidungen zu fällen.

## 3. Die ersten Signifikanztests im frühen 18. Jahrhundert

### *3.1. Zufall oder göttliche Vorsehung – Der Test von John Arbuthnot*

In den *Philosophical Transactions of the Royal Society* des Jahrgangs 1710 stellte John Arbuthnot[8] ein Argument für die göttliche Vorsehung dar, gegründet auf das während vieler Jahre festgestellte konstante Verhältnis zwischen männlichen und weiblichen Geburten (präziser: zwischen männlichen und weiblichen Getauften) in London (ARBUTHNOT, 1710; KENDALL, 1977). Bezeichnen wir Jahre mit mehr männlichen Geburten kurz als «männliche Jahre» und solche mit mehr weiblichen Geburten als «weibliche Jahre», so können wir in *moderner Sprechweise* sagen, John Arbuthnot testet die Nullhypothese: «Die Wahrscheinlichkeitsverteilung für männliche Jahre und für weibliche Jahre ist eine Binomialverteilung mit dem Parameter p, wobei

p: = P (männliches Jahr) = P (weibliches Jahr) = 1/2 ist.»

Arbuthnot geht dabei von der folgenden Überlegung aus: Wenn die Wahrscheinlichkeit, daß ein Kind, das geboren wird, von männlichem Geschlecht ist, gleich der Wahrscheinlichkeit ist, daß es von weiblichem Geschlecht ist – also 1/2 –, so müßte die Verteilung der männlichen und der weiblichen Jahre über die insgesamt 82 aufeinanderfolgenden Jahre, aus denen ihm Beobachtungen zur Verfügung stehen, der Wahrscheinlichkeitsverteilung von «Kopf» und «Zahl» in 82 Würfen mit einer symmetrischen Münze entsprechen. (Die Richtigkeit dieser Entsprechung begründet er übrigens nicht.)

Nun zeigen Beobachtungen – es handelt sich dabei wie gesagt nicht um das Register der Geburten, sondern um die Zahlen der Taufen, die an Knaben und Mädchen von 1629 bis 1710 vorgenommen worden sind – für jedes der 82 Jahre mehr männliche als weibliche Geburten. Wäre nun die oben formulierte Nullhypothese richtig, so würde die Wahrscheinlichkeit für 82 aufeinanderfolgende männliche Jahre

$(1/2)^{82}$, «which will be found easily by the table of logarithms to be 1: 4836 000 000 000 000 000 000 000».

Arbuthnot sagt dann weiter, daß wir nun noch beachten können, daß der Überschuß an männlichen Geburten (bzw. Getauften) jeweils in einem nahezu konstanten Verhältnis zur Gesamtzahl steht, daß die Differenzen in fixierten Grenzen liegen, und daß wir dasselbe Ereignis wohl auch für andere Zeiten und für andere Gegenden vorfinden würden. So wird für ihn dann die

[8] Arbuthnot hat seine Überlegungen 1710 vorgetragen. Der zugehörige Band der Philosoph. Transactions von 1710 dürfte etwa 1712 erschienen sein.

obige Wahrscheinlichkeit zu einer «infinitely small quantity, at least less than any assignable fraction»; Arbuthnot verwirft deshalb die Nullhypothese.

Dazu nun die folgenden Bemerkungen:
(1) Mit BIRNBAUM (1967), HACKING (1965, 1975), HALD (1989), SHOESMITH (1987) und andern Autoren können wir feststellen, daß hier ein Signifikanztest durchgeführt wird, und zwar in Form eines *Vorzeichentests:* Arbuthnot sieht primär einfach den Überschuß an männlichen Geburten; hätte er z. B. jedes männliche Jahr mit einem Pluszeichen versehen, jedes weibliche mit einem Minuszeichen, so hätte er eben 82 mal ein Pluszeichen erhalten.

(2) Dieser als Vorzeichentest durchgeführte Signifikanztest von 1710 ist nicht der erste: Arbuthnot selbst hat ähnliche Überlegungen schon sechzehn Jahre früher angestellt. BELLHOUSE (1989) hat dies in einem Manuskript von Arbuthnot, das höchstwahrscheinlich aus dem Jahr 1694 stammt, festgestellt. Dieses Manuskript enthält übrigens noch – rudimentär – einen Signifikanztest, der auf eine andere Art, also nicht als Vorzeichentest, durchgeführt wird.

(3) Selbstverständlich hat Arbuthnot noch nicht die heute gebräuchliche Terminologie verwendet. Vor allem formuliert er auch nichts, was als Verwerfungsbereich interpretiert werden könnte. Dies ist besonders gefährlich: Er hätte ja auch für *jede* andere *bestimmte* Anordnung für Plus- und Minuszeichen – zum Beispiel 41mal plus und 41mal minus in einer bestimmten Anordnung – dieselbe Wahrscheinlichkeit erhalten und die Nullhypothese verwerfen können.

(4) Arbuthnot hat sich nicht mit der Verwerfung der Nullhypothese begnügt. Er folgert aus der oben erwähnten kleinen Wahrscheinlichkeit, daß hier nicht der Zufall regiert, «that is Art, not Chance, that governs». Und in dieser Festellung ist wohl mitenthalten, daß an Stelle des bloßen Zufalls eben die göttliche Vorsehung die Verteilung der Geschlechter regelt. Dies kommt ja schon im Titel seiner Arbeit sehr deutlich zum Ausdruck: *An argument for Divine providence, taken from the constant regularity observ'd in the birth of both sexes.* Durch diese weise Planung – so überlegt Arbuthnot – wird der Verlust an Männern ausgeglichen, der dadurch entsteht, daß die Männer ein viel gefährlicheres Leben führen müssen als die Frauen ... Und in einem Scholium fügt er bei, daß daraus auch folge, daß Polygamie gegen das Naturgesetz und gegen die Gerechtigkeit sei: «Wenn schließlich Männer und Frauen in derselben Zahl vorhanden sind und ein Mann dann zwanzig Frauen hätte, so müßten neunzehn Männer ledig bleiben.» Für Arbuthnot ist

also die Annahme, daß hier die göttliche Vorsehung wirke, die plausible, die bessere Hypothese, wie sie in Stegmüllers «Imperativ», den wir ja oben erwähnt haben, verlangt wird; eine Ansicht, die übrigens ganz dem Zeitgeist im Umfeld von Arbuthnot entsprochen hat.

(5) Gegen Ende des 17. Jahrhunderts und weit ins 18. Jahrhundert hinein spielte der Schluß von der Erkenntnis einer weisen Planung des irdischen Geschehens auf die Existenz eines höchsten Wesens – die Engländer sprachen vom «argument from design» – eine sehr wichtige Rolle, und dies nicht etwa nur im Kreise der Royal Society, in der Arbuthnot seinerzeit seine Argumentation vorgetragen hatte: «Wie nicht anders zu erwarten, wurde Arbuthnot's Argument begierig aufgegriffen und durch mehrere Jahrzehnte von den Kanzeln Oxfords bis herab zu denen Münchens verkündet» (STEGMÜLLER, 1973). Einen Kulminationspunkt dieser Gedankengänge stellt wohl die Publikation des Werkes von I. P. SÜSSMILCH (1741) dar mit dem Titel *Die göttliche Ordnung in den Veränderungen des menschlichen Geschlechts aus der Geburt, dem Tode und der Fortpflanzung desselben erwiesen.* (Siehe dazu z. B. SHOESMITH, 1987.) K. Pearson schreibt in diesem Zusammehang, daß die «idea of an omnipresent activating deity, who maintains mean statistical values, formed the foundation of statistical development through Derham, Süssmilch, Niewentyt, Price to Quetelet and Florence Nightingale» und fährt fort mit: «The causes which led de Moivre to his ‹Approximatio› or Bayes to his theorem were more theological and sociological than purely mathematical, and until one recognizes that post-Newtonian English mathematicians were more influenced by Newton's theology than by his mathematics, the history of science in the eighteenth century – in particular that of the scientists who were members of the Royal Society – must remain obscure» (PEARSON, 1926; vgl. dazu BIRNBAUM, 1967, und PEARSON, 1978, p. 279 ff.).

### *3.2. S'Gravesande verschärft die Argumentation von Arbuthnot*

Der holländische Mathematiker W. J. s'Gravesande (1688–1742), der später als Professor an der Universität Leiden wirkte, verfaßte bereits 1712 eine Verbesserung der Argumentation von John Arbuthnot: *Démonstration mathématique du soin que Dieu prend de diriger ce qui se passe dans le monde, tirée du nombre des Garçons et des Filles qui naissent journellement.* Die wichtigsten Resultate seiner Überlegungen wurden durch den Mathematiker und Philosophen B. Nieuwentyt um 1715 in seinem Buch *Het regt*

*gebruik der Wereldbeschouwingen*[9] publiziert; die ganze Arbeit erschien dann in den *Œuvres* von S'GRAVESANDE (1774).

S'Gravesande stellt zunächst fest, daß die relative Häufigkeit der Knabengeburten schwankt, zwischen 7765/15 448, also 50,27%, im Jahre 1703 und 4748/8855, also 53,62%, im Jahre 1661. Nun sind aber die effektiven Zahlen der Knaben- und Mädchengeburten bei Arbuthnot natürlich von Jahr zu Jahr verschieden. S'Gravesande arbeitet deshalb mit einer fiktiven Anzahl von jährlich 11 429 Geburten, was nahezu der mittleren Anzahl von Geburten pro Jahr in der von Arbuthnot publizierten Liste entspricht (der genaue Mittelwert wäre 11 442). Damit erhält s'Gravesande das folgende Modell für seine weiteren Überlegungen:

| | |
|---|---|
| Fiktive Anzahl Geburten pro Jahr | = 11 429, |
| fiktive minimale Anzahl von Knabengeburten pro Jahr = (7765/15 448) · 11 429 | = 5 745, |
| fiktive maximale Anzahl von Knabengeburten pro Jahr = (4748/8855) · 11 429 | = 6 128. |

Dann berechnet er in mühseliger Arbeit mit Hilfe der binomischen Verteilung die Wahrscheinlichkeit, daß die Zahl X der Knabengeburten zwischen den eben erwähnten Grenzen liegt. Einige Einzelheiten dieser Berechnung findet man bei SHOESMITH (1987). Er arbeitet dabei wie Arbuthnot mit dem Parameter $p = 1/2$ als Wahrscheinlichkeit für eine Knabengeburt und findet

$$P(5745 \leq X \leq 6128) = 0{,}292.$$

(Eine Berechnung dieser Wahrscheinlichkeit mit Hilfe der Normalverteilung, die jedoch s'Gravesande noch nicht zur Verfügung stand, würde 0,284 ergeben.)

Nun ist die Wahrscheinlichkeit, daß X in 82 aufeinanderfolgenden Jahren stets zwischen diesen Grenzen 5745 und 6128 liegt,

$$(0{,}292)^{82} \approx 1{,}45 \cdot 10^{-44},$$

also mehr als $10^{19}$ mal kleiner als jene von Arbuthnot. Die Überlegungen von s'Gravesande führen so zu einem Resultat, das eine sehr viel signifikantere Abweichung von der Nullhypothese aufweist, als das Ergebnis von Arbuthnot sie zeigt. Deshalb schließt s'Gravesande seine Überlegungen mit den Worten: «Que conclure de tout ce qu'on vient de dire? C'est que celui qui a créé les Cieux & la Terre, dirige ce qui s'y passe non seulement par les loix

[9] Der Herausgeber der Werke von s'Gravesande 1774, J.N.S. Allamand, schreibt: S'Gravesande «en donna cette démonstration, qu'il se contenta de communiquer à ses amis sans la faire imprimer. Mr. B. Nieuwentyt fut un de ceux à qui il en fit part, et l'on trouve le résultat dans l'ouvrage de cet auteur...».

générales qu'il a établies dès le commencement de leur existence, mais encore par des loix particulières dont l'effet se fait sentir journellement. Il n'y a qu'un Etre intelligent qui puisse faire naître des garçons & des filles précisément autant qu'il en faut des uns & des autres, pour que tout reste dans l'ordre, malgré la prodigieuse probabilité qui s'y oppose, si on ne fait attention qu'à ce qui peut découler des loix générales & physiques. Si l'on examinoit de la même manière tout ce qui arrive sur notre globe, on se convaincroit que non seulement il n'y a rien qui échappe à la connoissance de Dieu; mais qu'encore des choses, qui semblent dépendre de ces loix générales, sont dirigées tous les jours d'une façon particulière par cet Etre suprême, à la conservation & au bonheur de ses créatures...» (S'GRAVESANDE, 1774).

### *3.3. Nikolaus Bernoulli entwickelt ein besseres Modell*

Arbuthnot und s'Gravesande wollten also mit ihren Folgerungen aus den ihnen zur Verfügung stehenden Daten die Annahme widerlegen, daß in diesen Daten einfach das Wirken des bloßen Zufalls zum Ausdruck komme: «Art, not chance» will Arbuthnot durch das Verwerfen der Nullhypothese bestätigen, und s'Gravesande möchte eine «Démonstration que Dieu prend soin...» geben.

Nikolaus Bernoulli (1687–1759) scheint sich weniger für die theologische Seite der ganzen Angelegenheit interessiert zu haben. Seine Überlegungen kann man den Briefwechseln mit s'Gravesande und mit Montmort entnehmen (S'GRAVESANDE, 1774; MONTMORT, 1713). Aus dem letzten Brief an s'Gravesande (vom 30.12.1712) kann man zunächst klar entnehmen, daß es ihm nicht etwa darum geht, gegen die Vorsehung zu kämpfen, schreibt er doch: «Il est vrai que si la naissance d'un garçon et d'une fille est également possible, il est fort peu probable que le nombre des garçons surpasse celui des filles plusieurs fois de suite: or comme cela est arrivé pourtant à Londres, je conclus et tout le monde conclura aussi, qu'à Londres la naissance d'un garçon est plus facile que celle d'une fille; et que cela soit un effet de la Providence de Dieu, c'est ce que je n'ai jamais nié.» In Bezug auf die Folgerungen von Arbuthnot ist der Briefwechsel von Nikolaus Bernoulli mit s'Gravesande nicht ganz frei von Mißverständnissen; klar sind jedoch die mathematischen Darlegungen von Nikolaus Bernoulli, die darin zum Ausdruck kommen. Diese mathematischen Aspekte werden vor allem in der Beilage dargestellt, die Bernoulli dem zweiten seiner Briefe (vom 9.11.1712) beifügt. Es handelt sich dabei um die Kopie eines Briefes an den englischen Mathematiker Burnet; mit ihm und andern Mitgliedern der Royal Society hat Bernoulli während seines Aufenthaltes in London (1712) die Arbeit von Arbuthnot diskutiert.

Diese Briefkopie bringt zunächst den ausführlichen Beweis eines Satzes, den man gelegentlich als *Theorem von Nikolaus Bernoulli* bezeichnet. Es handelt sich um eine Abschätzung der Endglieder bei der Berechnung von $(p + q)^n$ für großes n und $p + q = 1$, die bei der Arbeit mit der binomischen Verteilung gute Dienste leistet (HALD, 1984, 1989; SHOESMITH, 1987). Dieses Theorem von Nikolaus Bernoulli stellt eine Verschärfung der Approximation dar, die sein Onkel Jacob Bernoulli im vierten Teil der Ars conjectandi gegeben hatte und die meistens als Satz von Bernoulli[10] bezeichnet wird (JAKOB BERNOULLI, 1899).

Weiter bestimmt Nikolaus Bernoulli aus den Daten von Arbuthnot für jedes Jahr die relativen Häufigkeiten einer Knabengeburt und erarbeitet das folgende Modell:

Fiktive Anzahl der Geburten pro Jahr = 14 000,
fiktive minimale Anzahl von Knabengeburten
pro Jahr = (7765/15 448) · 14 000 = 7 037,
fiktive maximale Anzahl von Knabengeburten
pro Jahr = (4748/8855) · 14 000 = 7 507.

Als Mittelwert erhält er in seinem Modell 7237; daraus berechnet er den Wert der Wahrscheinlichkeit p einer Knabengeburt, also den Parameter p seiner binomischen Verteilung, zu $p = 7237 : 14\,000 = 0{,}5169$.
Mit Hilfe des oben erwähnten Theorems findet er dann die Wahrscheinlichkeit, daß die Zahl X der Knabengeburten zwischen $7037 = 7237 - 200$ und $7437 = 7237 + 200$ liegt; es ist

$$P(7037 \leq X \leq 7437) > 303/304 = 0{,}9967.$$

(Auf die vier Werte, die außerhalb dieses Intervalls liegen, geht er nicht ein.) Weiter berechnet er nun $(303/304)^{100} = 0{,}7193$. Diese Wahrscheinlichkeit ist für Nikolaus Bernoulli genügend groß um festzustellen, daß es eigentlich kein Wunder sei, daß die Zahl der Knabengeburten während 82 aufeinanderfolgenden Jahren in so engen Grenzen liege: «Ergo non est miraculum numerum masculorum limitem hunc 82 annis non excessisse...».

[10] Der *Satz von Bernoulli,* genauer *Satz von Jakob Bernoulli,* lautet in der ursprünglichen Formulierung der *Ars conjectandi* von 1713 (J. BERNOULLI, 1899): «Es möge sich die Zahl der günstigen Fälle zu der Zahl der ungünstigen Fälle genau oder näherungsweise wie r/s, also zu der Zahl aller Fälle wie $r/(r + s) = r/t$ – wenn $r + s = t$ gesetzt wird – verhalten, welches letzteres Verhältnis zwischen $(r + 1)/t$ und $(r - 1)/t$ enthalten ist. Nun können, wie zu beweisen ist, so viele Beobachtungen gemacht werden, daß es beliebig oft (z. B. c mal) wahrscheinlicher wird, daß das Verhältnis der günstigen zu allen angestellten Beobachtungen innerhalb dieser Grenzen liegt als außerhalb, also weder größer als $(r + 1)/t$, noch kleiner als $(r - 1)/t$ ist.»

Ähnlich äußert er sich auch in zwei Briefen an Montmort in dieser Angelegenheit. Die beiden Briefe vom 11.10.1712 und vom 23.1.1713 sind in Montmorts «Essai d'Analyse sur les jeux de hasard» abgedruckt (MONTMORT, 1713). Der zweite dieser Briefe enthält auch wieder den Beweis des bereits erwähnten «Theorems von Nikolaus Bernoulli». Nikolaus Bernoulli erinnert dabei auch, daß sein Onkel Jakob Bernoulli «une semblable chose dans son Traité *De Arte conjectandi,* qui s'imprime à présent à Bâle», bewiesen hat. Die *Ars conjectandi* von Jakob Bernoulli erschien postum in Basel 1713 (BERNOULLI J., 1899, 1975).

In diesen Briefen an Montmort arbeitet Bernoulli mit einem leicht veränderten Modell: «J'ai trouvé en examinant le catalogue des enfants nés à Londres depuis 1629 jusqu'à 1710 inclusivement, qu'il y a plus de mâles que de femelles; et qu'en prenant un milieu, la raison des mâles aux femelles est fort près de la raison de 18 à 17, un peu plus grande; d'où j'ai conclus que la probabilité, pour qu'il naisse un garçon est à la probabilité que naisse une fille environ comme 18 à 17» (Brief vom 11.10.1712).

Somit rechnet Bernoulli also in diesen beiden Briefen mit $p = 18/35 = 0{,}5143$. Wiederum legt er seinen Beobachtungen eine fiktive Gesamtzahl von 14 000 Geburten zugrunde. Der Mittelwert für die Anzahl der Knabengeburten ist dann $14\,000 \cdot 18/35$, also 7200. Die fiktive minimale Anzahl der Knabengeburten ist wie oben 7037, die maximale ist wieder 7507. Damit erhält er im Brief vom 23.1.1713 für die Wahrscheinlichkeit, daß die Anzahl der Knabengeburten zwischen $7037 = 7200 - 163$ und $7363 = 7200 + 163$ liegt, mit Hilfe seines bereits erwähnten Theorems

$$P(7037 \leq X \leq 7363) > 43{,}58/44{,}58 = 0{,}9776.$$

Er schließt daraus: «Donc on peut parier avec avantage qu'en 82 fois le nombre des mâles ne tombera pas trois fois hors de ces limites.» Und weiter stellt er fest, daß in 11 Beobachtungsjahren die fiktiven Anzahlen der Knabengeburten außerhalb des Intervalls [7037; 7363] liegen und sagt: «... vous trouverez aussi aisément qu'on peut parier plus que 226 contre 1 que le nombre des mâles ne tombera pas en 82 ans onze fois hors de ces limites.» Bernoulli schreibt nicht, *wie* er die Wahrscheinlichkeiten berechnet hat, die diese Feststellungen rechtfertigen. Wir würden die erste dieser beiden Wahrscheinlichkeiten so berechnen: Es sei E das Ereignis «Höchstens zwei Anzahlen fallen nicht ins Intervall [7037; 7363]». Dann gilt mit $w = 0{,}9776$

$$P(E) > w^{82} + \binom{82}{1} w^{81}(1-w) + \binom{82}{2} w^{80}(1-w)^2 = 0{,}72.$$

Diese Wahrscheinlichkeit ist somit größer als 1/2, «donc on peut parier avec avantage»! In derselben Art kann man die zweite Aussage verifizieren: Die

Nachrechnung liefert 0,9999; Bernoulli gibt 226/227 = 0,9956 für die Wahrscheinlichkeit des Ereignisses. «Höchstens zehn Anzahlen fallen nicht ins Intervall [7037; 7363].»

Zu diesen Überlegungen von Nikolaus Bernoulli nun die folgenden Bemerkungen:

(1) Wenn man aus den von Arbuthnot verwendeten Originalzahlen den Quotienten aus der gesamten Zahl der Knabengeburten und der gesamten Zahl der Geburten beider Geschlechter berechnet, so erhält man 0,5163. Es wäre also naheliegend gewesen, mit p = 0,5163 als Schätzwert für den Parameter einer Binominalverteilung zu arbeiten. Wie wir gesehen haben, verwendet Bernoulli im Brief an W. Burnet p = 0,5169 und in den Briefen an Montmort 18/35 = 0,5143. Den ersten Wert hat er aus seinem Modell mit den fiktiven Anzahlen berechnet; beim zweiten Wert könnte der Wunsch mitgespielt haben, im Nenner von p einen Teiler von 14 000 zu haben. Weiter wählt Bernoulli in seinen Modellen eine fiktive Gesamtzahl von jährlich 14 000 Geburten. Den Originaldaten von Arbuthnot kann man entnehmen, daß in den 82 Beobachtungsjahren 56 mal eine kleinere – zum Teil eine wesentlich kleinere – gesamte Zahl von Geburten vorliegt; die Grenze von 14 000 wird erst gegen Ende der Tabelle in der Regel überschritten.

(2) Sowohl die Wahl des Wertes von p = 18/35 als auch die Wahl von n = 14 000 für die fiktive Gesamtzahl aller Geburten tragen dazu bei, daß das Modell die Beobachtungen nicht ganz befriedigend wiedergibt. «... il y a une grande probabilité que les nombres des mâles et de femelles tombent entre des limites encore *plus petites que celles qu'on a observées*», schreibt Bernoulli im Brief vom 23.1.1712. Dies ist tatsächlich der Fall: Er stellt selbst fest, daß in 11 Beobachtungsjahren die fiktiven Anzahlen der Knabengeburten außerhalb des Intervalls [7037; 7363] liegen, und dies trotz der sehr kleinen Wahrscheinlichkeit, die sich aus seinem Modell für dieses Ereignis ergeben würde. Nach unseren heutigen Maßstäben würden wir sein Modell nicht mehr akzeptieren; wir würden die entsprechende Nullhypothese verwerfen, denn die beobachtete Schwankung der Daten ist signifikant größer als sie nach dem Modell sein sollte. Weitere Einzelheiten dazu findet man bei SHOESMITH (1985) und HALD (1989). ANSCOMBE (1981) hat übrigens gezeigt, daß die Untersuchung der entsprechenden 2 mal 82-Felder-Tafel mit dem Chi-Quadrat-Test zeigt, daß die Binomialverteilung kein geeignetes Modell für die Beschreibung der Daten von Arbuthnot liefert; er fand $\chi^2 = 170$ bei 81 Freiheitsgraden, was die Feststellung einer höchst signifikanten Abweichung zwischen den Beobachtungsdaten und einem auf der Binomialverteilung

beruhenden Modell erlaubt. Anscombe kommentiert: «In fact his table[11] shows more variability, not less, than the Bernoulli-trials hypothesis would imply: $\chi^2 = 170$ with 81 degrees of freedom[12]. A possible explanation is counting or copying errors in the data.» Die Binomialverteilung liefert also kein geeignetes Modell für die Beschreibung der Daten von Arbuthnot.

(3) Trotzdem sind die Beiträge von Nikolaus Bernoulli zu unserer ganzen Frage vom Standpunkt der mathematischen Statistik aus als *sehr hoch* einzuschätzen. HALD (1989) charakterisiert Bernoullis Verdienste mit den Worten: «... his attitude is pragmatic, like that of a modern statistician. The main points of his letters are (1) to estimate the probability of male birth from the observation; (2) to compare the distribution of the observations with the binomial distribution to determine whether the observed variation can be explained by this model; and (3) to provide the mathematical tool of this comparison by finding an approximation to the binomial for large n». Dazu muß in diesem Zusammenhang noch betont werden, daß Bernoulli auch klar erkannt hat, daß *nicht* bloß dann «Zufall» herrscht, wenn $p = 1/2$ gesetzt wird, wenn also die Wahrscheinlichkeiten für Knaben- und Mädchengeburten *gleich* wären, sondern auch dann, wenn *andere* geeignete Werte von p das Geschehen beschreiben. Wie bereits oben bemerkt, wollte er damit nicht das Wirken der «Providence de Dieu» ausschalten. Es ist ihm viel eher darum gegangen zu zeigen, daß man durchaus erwarten darf, daß die jährliche Anzahl der Knabengeburten innerhalb relativ enger Grenzen liegt, wenn man ein geeignetes Modell zur Beschreibung dieses Phänomens heranzieht. Es scheint aber, daß man seine Kritik an Arbuthnot eher so aufgefaßt hat, als ob er sich gegen das Wirken einer Vorsehung aussprechen möchte. Die bereits genannten Mißverständnisse im Briefwechsel mit s'Gravesande weisen in diese Richtung, ebenso eine Bemerkung, mit der Abraham de Moivre (1667–1754) in seiner Doctrine of Chance seine Hinweise auf die Ausführungen von Nikolaus Bernoulli zur Argumentation von Arbuthnot schließt (DE MOIVRE, 1756): «As if we were shewn a number of Dice, each with 18 white and 17 black faces, which is Mr. Bernoulli's supposition, we should not doubt but that those Dice had been made by some Artist, and that their form was not owing to *Chance,* but was adapted to the particular purpose he had in View.» Auf das, was Nikolaus Bernoullis Überlegungen ganz wesentlich von jenen Arbuthnots unterscheidet, geht de Moivre allerdings nicht ein; Niko-

[11] Mit «his table» ist die Tafel in der Publikation von Arbuthnot gemeint, die die Zahl der männlichen und weiblichen Geburten (bzw. Getauften) in London in den 82 aufeinanderfolgenden Jahren 1629–1710 gibt.

[12] Es handelt sich um eine 2-mal-82-Felder-Tafel. Die Anzahl der Freiheitsgrade ist dann $(2-1)\cdot(82-1) = 81$.

laus Bernoullis Ausführungen zum Problem von Arbuthnot scheinen auch sonst keine besonders große Wertschätzung von Seiten seiner Zeitgenossen erfahren zu haben.

### *3.4. Daniel Bernoulli testet die Lage der Planetenbahnen*

Zunächst ist festzuhalten, daß sich auch Daniel Bernoulli (1700–1782) mit den Schwankungen des Geschlechtsverhältnisses von Neugeborenen beschäftigt hat. Und auch bei ihm findet man wieder Überlegungen, die implizit jenen entsprechen, die man bei der Durchführung eines Signifikanztestes zu machen hat. So berechnet er zum Beispiel aus vorliegenden Daten die Wahrscheinlichkeit einer Knabengeburt, berechnet damit für eine Anzahl Jahre die Erwartungswerte und vergleicht diese Zahlen mit den wirklich vorliegenden Daten. Die Abweichungen vergleicht er mit dem wahrscheinlichen Fehler. Dieselbe Untersuchung führt er nachher mit einem hypothetischen Wert der Wahrscheinlichkeit einer Knabengeburt durch. Es soll hier nicht weiter auf diese und analoge Untersuchungen eingegangen werden; Einzelheiten dazu findet man bei HALD (1989). Wir wollen unsere Zusammenstellung der ersten Untersuchungen im frühen 18. Jahrhundert, die wenigstens implizit als Durchführung eines Signifikanztests gewertet werden können, mit Überlegungen schließen, die Daniel Bernoulli bezüglich der gegenseitigen Lage der Planetenbahnen gemacht hat.

Für das Jahr 1734 stellte die Académie Royale des Sciences de Paris zum zweiten Mal eine Preisaufgabe, die sie erstmals bereits 1732 gestellt hatte: «Quelle est la cause physique de l'inclinaison des Plans des Orbites des Planètes par rapport au plan de l'Equateur de la révolution du Soleil autour de son axe; et d'où vient que les inclinaisons de ces Orbites sont différentes entre elles?»

In den (verdoppelten) Preis konnten sich Daniel Bernoulli und sein Vater Johann I Bernoulli (1667–1748) teilen. Die Arbeit von Daniel Bernoulli ist im *Recueil des pièces qui ont remporté le prix de l'Académie Royale des Sciences* erschienen (D. BERNOULLI, 1734). Sie ist zunächst lateinisch abgefaßt und nachträglich ins Französische übersetzt worden, «traduite en français par son autheur». Daniel Bernoulli stellt darin fest, daß die Bahnen von Merkur und Erde den größten Winkel miteinander bilden, nämlich einen Winkel von der Größe 6°54', während die Bahnen von Saturn und Merkur einen Winkel von 6°24' und jene von Jupiter und Merkur einen Winkel von 6°8' einschließen. Bevor nun Daniel Bernoulli die physikalische Seite des Problems anpackt, stellt er sich die Frage: «... pourrait-on, pour ne pas alléguer d'autres raisons, attribuer à un pur hazard le peu d'inclinaison de toutes ces orbites au plan de l'Equateur solaire? Ou si cela pouvoit paroître encore douteux [vû le peu de précision et de certitude dans la position de

l'Equateur solaire] du moins ne pourra-t-on pas disconvenir que les Orbites planétaires ne tendent vers un plan commun, puisque sans cela il auroit été moralement impossible, que les Orbites fussent renfermées dans des limites aussi serrées qu'il le sont.»[13] Dieser Frage will er auf die folgende Art nachgehen: «Je chercherai de toutes les Orbites planétaires, les deux qui se coupent sous le plus grand angle, après quoi *je calculerai quelle probabilité il y a, que toutes les autres Orbites soient renfermées par hazard dans les limites de ces deux orbites.* On verra par-là que *cette probabilité est si petite* qu'elle doit passer pour une *impossibilité morale.*»

Es ist offensichtlich, daß in Daniel Bernoullis Überlegungen wieder die Idee eines *Signifikanztests* steckt. Er vermutet – entsprechend der Aufgabenstellung – das Vorliegen physikalischer Ursachen und stellt als Gegenhypothese eine Nullhypothese auf, die er zunächst testen will. Diese Nullhypothese lautet: Es ist Zufall, daß die genannten Planetenbahnen so nahe beieinander liegen. Nun zieht er die bereits bekannten Beobachtungen heran und versucht, die Wahrscheinlichkeit zu berechnen, daß die Planetenbahnen – bei Gültigkeit der Nullhypothese! – so nahe beieinander liegen. Sie ist nach seinen Berechnungen sehr klein. Deshalb *und* weil er eine plausiblere Erklärung zur Verfügung hat (jene von den physikalischen Ursachen) verwirft er die Nullhypothese. Anschließend geht er ausführlich auf die physikalischen Ursachen ein (DANIEL BERNOULLI, 1734, 1982).

Es ist eigentlich nicht schwierig, die gesuchte Wahrscheinlichkeit zu berechnen. B.L. van der Waerden formuliert in seinem Kommentar (D. BERNOULLI, 1982) die Fragestellung und die Lösungsidee folgenderma-

[13] Moralement impossible: Jakob Bernoulli definiert in seiner *Ars conjectandi* von 1713 die Begriffe *moraliter certum* und *moraliter impossibile: «Moralisch gewiß* ist etwas, dessen Wahrscheinlichkeit nahezu der vollen Gewißheit gleichkommt, so daß ein Unterschied nicht wahrgenommen werden kann. *Moralisch unmöglich* dagegen ist das, was nur soviel Wahrscheinlichkeit besitzt, als dem moralischen Gewissen an der vollen Gewißheit mangelt. Wenn man also das, was 999/1000 der Gewißheit für sich hat, als moralisch gewiß betrachtet, so ist das, was nur 1/1000 der Gewißheit für sich hat, moralisch unmöglich.»

Und etwas später: «Es würde also nützlich sein, wenn auf Veranlassung der Obrigkeit bestimmte Grenzen für die moralische Gewißheit festgesetzt würden, wenn z. B. entschieden würde, ob zur Erzielung dieser 99/100 oder 999/1000 der Gewißheit verlangt werden müssen ...» (J. BERNOULLI, 1899).

Die mathematische Statistik konnte sich erst richtig entfalten, nachdem man eingesehen hatte, daß man sich mit einem vernünftigen Grad von Zuverlässigkeit begnügen muß. So wählt man heute bei der Angabe von Vertrauensintervallen (Konfidenzintervallen) meistens eine Konfidenzzahl von 95% oder 99% und nur selten 99,9%; beim Testen von Hypothesen wählt man sehr häufig 5% oder 1% als Signifikanzniveau und wohl nur in seltenen Fällen 0,1%.

ßen: «Was ist die Wahrscheinlichkeit, daß die Bahnen der fünf übrigen Planeten Winkel von höchsten 6°54' mit der Erdbahn bilden? – Wenn diese Bahnebenen rein zufällig verteilt wären, so wären auch ihre Pole über die ganze Kugel gleich verteilt. Die Nordpole wären dann gleich verteilt auf einer Halbkugel. Was ist nun die Wahrscheinlichkeit, daß sie alle in einem Kreis vom Radius 6°54' um den Nordpol der Ekliptik liegen?»

Nun ist der Flächeninhalt der entsprechenden Kugelkalotte (Radius der Kugel r, Höhe der Kalotte h) gegeben durch $A_1 = 2\pi\, hr$, wobei offensichtlich $h = r(1\text{-}\cos 6°54')$ ist; der Flächeninhalt der Halbkugel ist $A_2 = 2\pi\, r^2$. Die Wahrscheinlichkeit, daß ein bestimmter dieser Nordpole in jener Kalotte liegt, ist also $A_1/A_2 = 0{,}007243$; die gesuchte Wahrscheinlichkeit, daß alle fünf Planetenbahnen ihren Nordpol in jener Kalotte haben, ist somit $(A_1/A_2)^5 \approx 2 \cdot 10^{-11}$.

Daniel Bernoulli hat diese Wahrscheinlichkeit falsch berechnet: Er schneidet die Planetenbahnen mit einer Sphäre. Seine erste Berechnung beruht auf der Annahme, alle Planetenbahnen würden in einer Kugelzone von der Breite 6°54' liegen. Er stellt fest, daß der Flächeninhalt dieser Kugelzone ungefähr 1/17 der ganzen Kugelfläche ausmacht. (In der Tat wäre der Flächeninhalt dieser Kugelzone gegeben durch $A_1' = 2\pi r^2 \sin 6°54'$; $A_1'$ dividiert durch den Flächeninhalt der ganzen Kugeloberfläche ergibt ungefähr $0{,}06 \approx 1/17$.) Weiter berechnet er «par les règles ordinaires, le nombre des cas, qui fassent tomber les 5 Orbites dans ladite zone, au nombre des cas contraires comme 1 à $17^5$-1; c'est-à-dire 1 à 1 419 856». Er stellt selber fest: «Je ne donne pas à cette méthode toute la précision géométrique, ..., parce qu'il ne s'agit ici que d'avoir quelque idée générale de la chose.» Und er nennt auch eine der hauptsächlichsten Schwierigkeiten: Die Bewegung der Knoten der Planetenbahnen. Er gibt anschließend noch zwei weitere Berechnungsmöglichkeiten dieser Wahrscheinlichkeit; sie führen beide nicht zum richtigen Resultat, das wir oben angegeben haben (D. BERNOULLI, 1734, 1982; TODHUNTER, 1865).

Zusammenfassend kann man sagen:

(1) Daniel Bernoullis Behauptung, daß eine rein zufällige Verteilung der Planetenbahnen «moralisch unmöglich» sei, ist gerechtfertigt; die zugehörige Wahrscheinlichkeit hat er allerdings nicht richtig berechnet.

(2) Trotzdem berechtigt uns seine Schlußweise festzustellen, daß er ähnlich wie Arbuthnot und s'Gravesande eine Untersuchung durchgeführt hat, die man heute als Signifikanztest bezeichnen würde. Man kann ihn indessen wohl nicht als «Begründer der modernen statistischen Prüfverfahren»

bezeichnen, wie dies noch 1961 am Biometrischen Kolloquium der Gruppe Schweiz der Biometric Society in Bern gesagt worden ist (vgl. den Bericht NIEDERBERGER, 1961).

(3) Weiter ist anzumerken, daß Daniel Bernoulli hier mit *geometrischen* Wahrscheinlichkeiten gearbeitet hat. In späteren Werken über Wahrscheinlichkeitsrechnung sind solchen Problemen oft eigene Kapitel gewidmet worden, so z. B. noch von LOUIS BACHELIER (1912), der geometrische Wahrscheinlichkeiten so umschreibt: «On dit qu'un problème est relatif aux probabilités géométriques lorsqu'il consiste à déterminer la probabilité pour qu'un ensemble de points, de lignes ou de surfaces dépendant d'une certaine façon du hasard possède une propriété géométrique donnée.» Heute findet man in der Regel nur noch in elementaren Darstellungen Probleme, die ausdrücklich mit «geometrische Wahrscheinlichkeiten» überschrieben sind; in höheren, maßtheoretischen Darstellungen werden solche Probleme von einem etwas andern Gesichtspunkte aus betrachtet. Zur Zeit von Daniel Bernoulli indessen ist mit diesen geometrischen Problemen Neuland erschlossen worden: Das wohl älteste Beispiel findet sich in einem Manuskript von Newton aus dem Jahre 1665 (NEWTON, 1976). SHEYNIN (1976) stellt fest, daß E. Halley als erster in einem 1693 publizierten Werk geometrische Wahrscheinlichkeiten verwendet hat. Vier Jahrzehnte später hat G. L. L. Buffon vor der Pariser Académie des Sciences über ein solches Problem vorgetragen; die oben erwähnte Arbeit von Daniel Bernoulli stammt aus dem Jahre 1734[14].

(4) D'Alembert (1717–1783) hat übrigens die Argumentation von Daniel Bernoulli als wertlos betrachtet (d'ALEMBERT, 1821; DASTON, 1979). Er ist zwar ebenfalls der Meinung, die geringen Neigungen der Planetenbahnen gegeneinander würden auf eine Ursache hinweisen; indessen hätten *andere* mögliche Anordnungen auch nur eine sehr kleine Wahrscheinlichkeit für sich: «... mathématiquement parlant, elle soit aussi possible que toute autre combinaison où l'on ne verrait aucun ordre ni aucune singularité....» Dies trifft natürlich zu. Aber d'Alembert scheint eben nicht zu beachten, daß Daniel Bernoulli eine *andere, plausible* Erklärung für diese bestimmte Anordnung der Planetenbahnen hat, nämlich das Vorliegen physikalischer Ursachen. In heutiger Sprechweise könnten wir sagen: Er verwirft die Nullhypothese «Zufall», weil ein Ereignis eingetreten ist, dem nur eine kleine

[14] Ähnliche Überlegungen wie sie Daniel Bernoulli für die Planetenbahnen gemacht hat, sind später von John Michell gemacht worden betreffend die Existenz von Doppelsternen: Man findet viel häufiger Sterne nahe beieinander – eben Doppelsterne – als dies bei Gleichverteilung wahrscheinlich wäre (MICHELL, 1767).

Wahrscheinlichkeit zukommen würde, wenn die Nullhypothese gelten würde, *und* weil er eine andere, plausible Erklärung zur Verfügung hat. Man vergleiche dazu den oben, im ersten Abschnitt, zitierten «Imperativ» von Stegmüller.

## 4. Abschließende Bemerkungen

Die Entscheidungen, die von Arbuthnot, s'Gravesande, Nikolaus Bernoulli und Daniel Bernoulli in den oben dargelegten Beispielen getroffen worden sind, beruhen offensichtlich auf Überlegungen, die der Idee nach unserem heutigen Vorgehen bei der Durchführung eines Signifikanztests entsprechen. Irgendeine Theorie solcher Tests und die zugehörige Terminologie lagen damit natürlich noch nicht vor. Die Theorie des Testens wurde – wie wir eingangs unserer Ausführungen festgestellt haben – in der ersten Hälfte des 20. Jahrhunderts geschaffen (vgl. z. B. BIEHLER, 1985). Die modernen Testtheoretiker haben dabei kaum direkt an die Überlegungen jener Mathematiker angeknüpft, die als erste solche Prüfverfahren angewendet haben. Von großem Einfluß dürfte viel eher die Fehlerrechnung, allgemein die Theorie der Beobachtungsfehler, gewesen sein, die z. B. in der Astronomie des 19. Jahrhunderts eine sehr große Rolle gespielt hat. Daneben gibt es natürlich weitere Brücken und Wege, die von jenen ersten tastenden Versuchen im frühen 18. Jahrhundert, die ja beinahe als Kuriosa anmuten, ins 19. und dann ins 20. Jahrhundert führen. Sie können hier nicht dargestellt werden. Drei Namen sollen immerhin noch erwähnt werden: P. S. de Laplace (1749–1827), S. D. Poisson (1781–1840) und der französische Arzt Louis-Dominique-Jules Gavarret (1809–1890).

P. S. de Laplace bringt in seinem *Essai philosophique des probabilités* von 1814, vor allem aber in seiner umfangreichen *Théorie analytique des probabilités,* außerordentlich viele Beispiele, die eigentlich die Durchführung eines Signifikanztests zeigen, natürlich immer noch ohne die heutige Terminologie: Er nimmt die oben erwähnten Probleme früherer Autoren wieder auf, ergänzt sie und erweitert Problemstellung und Lösung. So findet man darunter auch wieder Untersuchungen über das Geschlechtsverhältnis der Neugeborenen; aber nun werden die Verhältnisse auch für *verschiedene* Orte untersucht, und es werden *signifikante Unterschiede,* so etwa zwischen London und Paris, festgestellt und nachträglich auch begründet. Ähnlich ist es mit dem Problem der Planetenbahnen: Laplace zieht nicht nur die Bahnneigungen in Betracht, sondern auch die Umlaufsrichtungen aller damals bekannten Planeten und Monde und die Exzentrizitäten; schließlich zieht er

auch die Kometen heran und stellt signifikante Unterschiede zwischen den Verhältnissen bei den Planeten und ihren Monden einerseits und den Kometen andererseits fest (LAPLACE, 1886).

S.D. Poisson bringt in seinem Buch *Recherches sur la probabilité des jugements en matière criminelle et en matière civile* jene Formeln, die die Grundlagen geben können, um Signifikanzbetrachtungen durchzuführen oder um Konfidenzintervalle zu berechnen (POISSON, 1837). Auf ihnen baut der Arzt und spätere Professor für Physique médicale an der Medizinischen Fakultät von Paris L.D.J. Gavarret in seinem Buch *Principes généraux de statistique médicale* auf (GAVARRET, 1840). Geburtenziffern, Geschlechtsverhältnis der Neugeborenen, Sterbeziffern, Mortalität einer Krankheit bei verschiedenen Behandlungsmethoden und ähnliche Fragen sind seine Themen. Für die Berechnung von Konfidenzintervallen und für seine Signifikanzbetrachtungen verwendet er die von S.D. Poisson im oben genannten Buch angegebenen Formeln. S.D. Poisson war der Lehrer von Gavarret: Gavarret besuchte vor der Aufnahme des Medizinstudiums die Ecole Polytechnique von Paris. Gavarret arbeitet stets mit einer großen Zahl von Beobachtungen; er kann deshalb Formeln benützen, die aufgrund der Normalverteilung hergeleitet worden sind. Er verwendet natürlich noch nicht die heutige übliche Terminologie: So umschreibt er zum Beispiel «Konfidenzintervall» mit dem Begriff des *erreur possible.* Sein Konfidenzniveau (wie wir heute sagen würden) ist stets «P = 0.9953, à parier 212 contre 1». Einen großen Raum nehmen in seinem Buch jene Darlegungen ein, in denen er den Vorurteilen der Ärzteschaft gegen die Verwendung probabilistischer Methoden in der medizinischen Statistik – vor allem bei der Beurteilung von Therapie-Erfolgen – entgegentritt. Seine Ausführungen sind von musterhafter Klarheit, und die Behandlung der vielen Beispiele zeugt von Fachkenntnis und großem didaktischen Geschick. Nach der Lektüre seines Buches kann man dem Urteil von FREUDENTHAL und STEINER (1966) sehr wohl zustimmen, wenn sie schreiben: *«Mit der gemeinverständlichen Darstellung dieser Mathematik hat Gavarret eine neue Epoche in der mathematischen Statistik eingeleitet.»*[15]

[15] Zur Illustration nur einige Sätze aus dem Buch von Gavarret: «Lors donc qu'un savant entreprend une série d'observations relatives à un phénomène déjà soumis à l'expérience, ..., il doit s'attendre à rencontrer une différence entre les résultats actuels et les résultats déjà obtenus... Le premier travail d'un observateur qui constate une différence dans deux longues série d'observations consiste donc à chercher *si l'anomalie n'est qu'apparante ou si elle est réelle* et accuse l'intervention d'une cause perturbatrice...» (p. 193/194).

**Zusammenfassung**

Die allgemeine Theorie der statistischen Testverfahren ist im wesentlichen in der ersten Hälfte dieses Jahrhunderts geschaffen worden. Beispiele von mehr oder weniger korrekt durchgeführten Tests findet man jedoch bereits zu Beginn des 18. Jahrhunderts, und erste Spuren solcher Überlegungen gibt es bereits viel früher.

**Résumé**

La théorie générale des tests statistiques est essentiellement une contribution de la première moitié de notre siècle. On trouve cependant des exemples plus ou moins corrects de telles procédures au début du 18e siècle. Il faut remarquer que les premières traces de cette problématique sont encore bien antérieures.

**Summary**

The general theory of statistical tests is basically a contribution of the first part of our century. However, one can find examples of more or less correct test procedures at the beginning of the 18th century and traces of this problematic much earlier.

**Literatur**

D'ALEMBERT, J.: Œuvres complètes *I.* Bélin: Paris 1821.

ANSCOMBE, F.J.: Computing in statistical science through APL. Springer: Berlin 1981.

ARBUTHNOT, J.: An argument for Divine Providence, taken from the constant regularity observ'd in the births of both sexes. Phil. Trans. *27,* 186–190 (1710).

BACHELIER, L.: Calcul des probabilités *1.* Gauthier-Villars: Paris 1912.

BELLHOUSE, D.R.: A manuscript on chance written by John Arbuthnot. Internat. Stat. Rev. *57,* 249–259 (1989).

BERNOULLI, D.: Recherches physiques et astronomiques sur le problème posé pour la seconde fois par l'Académie Royale des Sciences de Paris. Recueil des pièces qui ont remporté le prix de l'Académie Royale des Sciences *3,* 93–122, 123–144 (1734).

– – : Die Werke von Daniel Bernoulli *2,* bearb. u. kommentiert von L.P. BOUKKAERT und B.L. VAN DER WAERDEN, unter Benützung von Vorarbeiten von H. STRAUB. Birkhäuser: Basel 1982.

BERNOULLI, J.: Wahrscheinlichkeitsrechnung (Ars conjectandi). Ostwalds Klassiker der exakten Wissenschaften *107* und *108.* Akad. Verlagsges.: Leipzig 1899.

– – : Die Werke von Jakob Bernoulli *3,* bearb. von B.L. VAN DER WAERDEN. Birkhäuser: Basel 1975.

BIEHLER, R.: Zur historischen Entwicklung der Theorien statistischen Testens – Mathematische Forschungsprogramme und ihr sich wandelnder Anwendungsbe-

zug. In: STEINER, H. G., und WINTER, H. (Hrsg.): Mathematikdidaktik – Bildungsgeschichte – Wissenschaftsgeschichte. Aulis: Köln 1985.

BIRNBAUM, A., und EISENHART, C.: Anniversaires in 1966–67 of interest to statistician. The Americ. Statistician *21,* 22–29 (1967).

BOLZAN, J. E.: An ancient anticipation of the probability calculus. Scientia *107,* 876–878 (1972).

CARDANO, G.: Liber de ludo aleae. In: Cardani opera omnia, Hrsg. C. SPON, *1.* Huguetan et Ravaud: Lyon 1663 (Reprint: Frommann: Stuttgart-Bad Canstatt 1966).

– – : Des Girolamo Cardano von Mailand eigene Lebensbeschreibung, übers. von H. HEFELE. Kösel: München 1969.

DASTON, L. J.: D'Alembert's critique of probability theory. Hist. Math. *6,* 259–279 (1979).

FISHER, R. A.: The design of experiments. Oliver and Boyd: London 1935.

FREUDENTHAL, H., und STEINER, H. G.: Aus der Geschichte der Wahrscheinlichkeitstheorie und Statistik. In: BEHNKE, H., BERTRAM, G., und SAUER, R.: Grundzüge der Mathematik *IV.* Vandenhoeck & Ruprecht: Göttingen 1966.

GAVARRET, L. D. J.: Principes généraux de statistique médicale. Bechet et Labé: Paris 1840.

S'GRAVESANDE, G. J.: Œuvres philosophiques et mathématiques. Rey: Amsterdam 1774.

GUARDINI, R.: Christliches Bewußtsein – Versuche über Pascal. dtv: München 1962.

HACKING, I.: Logic of statistical inference. Univ. Press: Cambridge 1965.

– – : The emergence of probability. Univ. Press: Cambridge 1975.

HALD, A.: Nicholas Bernoulli's Theorem. Internat. Stat. Rev. *52,* 93–99 (1984).

– – : A history of probability and statistics and their applications before 1750. Wiley: New York 1989.

HØYRUP, J.: Sixth-century intuitive probability: The statistical significance of a miracle. Hist. Math. *10,* 80–84 (1983).

INEICHEN, R.: Dante-Kommentare und die Vorgeschichte der Statistik. Hist. Math. *15,* 264–269 (1988).

– – : «Wenn auseinandergehn die Würfelspieler ...». Praxis der Math. *32,* 63–66 (1990).

– – : Der schlechte Würfel – ein selten behandeltes Problem in der Geschichte der Stochastik. Erscheint in Hist. Math. *18* (1991).

INVREA, R.: La legge dei grandi numeri era nota a Tucidide. Giornale dell'Ist. Ital. Attuari *7,* 229–230 (1936).

JAKOB, J.: Die Grundlagen unserer naturwissenschaftlichen Erkenntnis. Büchergilde: Zürich 1948.

KENDALL, M., and PLACKET, R. L.: Studies in the history of statistics and probability *II.* Griffin: London 1977.

LAPLACE, P. S.: Œuvres complètes de Laplace, publiés sous les auspices de l'Académie des Sciences. Gauthier-Villars: Paris 1886.

– – : Philosoph. Versuch über die Wahrscheinlichkeit. Ostwalds Klassiker der exakten Wissenschaften *233.* Akad. Verlagsges.: Leipzig 1932. (Reprint. Akad. Ver-

lagsges.: Leipzig 1986).
LINDER, A.: Statistische Methoden für Naturwissenschafter, Mediziner und Ingenieure. Birkhäuser: Basel 1945.
MICHELL, J.: An inquiry into the probable parallax, and magnitude of the fixed stars, from the quantity of light which they afford us, and the particular circumstances of their situations. Phil. Trans. *52* (1767).
MOIVRE, A. DE: The doctrine of chances. Millar: London 1756.
MONTMORT, P.R. DE: Essai d'analyse sur les jeux de hasard. Quillau: Paris 1713.
NIEDERBERGER, W.: Daniel Bernoulli als Begründer der modernen statistischen Prüfverfahren. Basler Nachrichten *512* (1961).
NEWTON, I.: Mathematical papers, Hrsg. WHITESIDE, D.T. Univ. Press: Cambridge 1967.
NOETHER, G.: Introduction to statistics – A fresh approach. Houghton & Mifflin: Boston 1971.
ORE, O.: Cardano, the gambling scholar, with a translation from the latin of Cardano's book on games of chance by S.H. GOULD. Univ. Press: Princeton 1953.
PASCAL, B.: Œuvres complètes, Hrsg. CHEVALIER, J. Gallimard: Paris 1954.
PEARSON, K.: A Letter to the Editor. Nature *117,* 551–552 (1926).
– – : The history of statistics in the 17th and 18th centuries against the changing background of intellectual, scientific and religious throught, Hrsg. E.S. PEARSON. Griffin: London 1978.
PITSCHMANN, H.: Das Ende des naturwissenschaftlichen Zeitalters. Wien 1980.
POISSON, S.D.: Recherches sur la probabilité des jugements en matière criminelle et en matière civile. Bachelier: Paris 1837.
RABINOVITCH, N.L.: Probability and statistical inference in ancient and medieval Jewish literature. Univ. of Toronto Press: Toronto 1973.
– – : The One and the Many: Early stochastic reasoning in philosophy. Ann. Sci. *34,* 331–344 (1977).
SCHNEIDER, I.: Wahrscheinlichkeit und Zufall bei Kepler. Philosoph. naturalis *16,* 40–63 (1976).
SHEYNIN, O.B.: Newton and the classical theory of probability. Arch. Hist. Exact Sci. *7,* 217–243 (1971).
SHOESMITH, E.: Nicholas Bernoulli and the argument of Divine Providence. Internat. Stat. Rev. *53,* 255–259 (1985).
– – : The continental controversy over Arbuthnot's argument for Divine Providence. Internat. Stat. Rev. *14,* 133–146 (1987).
STEGMÜLLER, W.: Statistisches Schließen. In: Probleme der Wissenschaftsgeschichte und Analytischen Philosophie *IV.* Springer: Berlin 1973.
STIGLER, S.M.: Eight centuries of sampling inspection: The trial of the Pyx. Americ. Statist. Assoc. *72,* 493–500 (1977).
SÜSSMILCH, J.P.: Die Göttliche Ordnung in den Veränderungen des menschlichen Geschlechts, aus der Geburt, Tod und Fortpflanzung desselben erwiesen. Spener: Berlin 1741.
SZABO, A.: Anfänge der griechischen Mathematik. Oldenbourg: Wien 1969.
TAUBER, W.: Das Würfelspiel im Mittelalter und in der frühen Neuzeit. Lang: Bern 1987.

THUKIDIDES: Geschichte des Peloponnesischen Krieges. Deutsch nach J. D. HEIMANN. Neubearbeitung. Schneider: Berlin o. Jg. (? 1940).

TODHUNTER, I.: A history of the mathematical theory of probability. Chelsea Publ.: Cambridge 1865. (Reprint: Chelsea Publ.: New York 1969.)