Hans-Joachim GIRLICH, Leipzig

John Arbuthnott und das statistische Schließen

Mathematikgeschichte kann für die Ausbildung von Mathematikern und Gymnasiallehrern hilfreich sein, wenn sie durch Beispiele und einfache Prinzipien zur Motivation und zum Verständnis scheinbar komplizierter Sachverhalte beiträgt, deren Kern durch überzogene Formalisierung verdeckt wird. Als ein Exempel betrachten wir „den Vorzeichentest, der schon zu Beginn des 18. Jahrhunderts zur Untersuchung der Häufigkeit von Knaben- und Mädchengeburten angewendet worden sein soll.“ wie in [1], Seite 159 behauptet wird.

1. Ein neuartiger Gottesbeweis

Gottfried Wilhelm Leibniz (1646-1716) eröffnete seine richtungsweisende Leipziger „Dissertatio de Arte combinatoria“ von 1666 mit einem strengen Beweis der Existenz Gottes als eines Wesens mit unendlicher Machtfülle, der auf 3 Definitionen, einem Postulat und 4 Axiomen über die Bewegung beruht. John Arbuthnott (1667-1735) erweiterte in [2] Leibniz' logische Schlussweise, indem er numerische Beobachtungsdaten in den Kalkül einbrachte, die eine statistische Schlussweise bedingten. Er führte die quantitative Ausgeglichenheit der Geschlechter auf die göttliche Vorsehung zurück. Sein „Beweis“ basiert auf der Ablehnung von deren Alternative, eines Zufallsmodells, auf Grund von dazu im Widerspruch stehenden Londoner Geburtenzahlen. Dieses statistische Schließen fand erst in der Testtheorie im 20. Jahrhundert mit vergleichbaren Hypothesen eine befriedigende Anwendung. Das Beleuchten der Ursprünge und der damals zur Verfügung stehenden Hilfsmittel vor einer Darstellung aus heutiger Sicht erscheint im Sinne der Einleitung gerechtfertigt. Auf die Interpretation im Sinne des Zeitgeistes, inkarniert durch den Präsidenten der Royal Society, der den Leibarzt der englischen Königin Anna wohl dazu animiert hat, werden wir nicht zurückkommen.

2. Das Binomialmodell

Christiaan Huygens (1629-1695) fasste 1657 seine Pariser Untersuchungen zur Wahrscheinlichkeitsrechnung in der Arbeit „Van Rekeningh in Spelen van Geluck“ zusammen, deren lateinische Fassung von Arbuthnott ins Englische übersetzt und mit einem Vorwort, Anmerkungen und eigenen Lösungen zu einigen von Huygens nur gestellten Aufgaben versehen wurde (vgl.[3]). Aus dieser Zeit als Privatlehrer für Mathematik stammt auch ein Manuskript, das Arbuthnott 1694 David Gregory (1659-1708) ausgehändigt hatte, das bereits Wesentliches von [2] vorwegnimmt. Als ein Modell des Zufalls hat er in [2] den Münzwurf mit einer idealen Münze gewählt, die bei jedem Wurf mit gleicher Wahrscheinlichkeit von 0,5 auf eine der beiden Seiten fällt. Unter Verwendung des binomischen Satzes berechnete er die Wahrscheinlichkeit für das gleich häufige Fallen beider Seiten bei einer geraden Anzahl von Würfen (explizit bis n=10) und stellte fest, dass diese Wahrscheinlichkeit mit wachsendem n gegen 0 geht. Dieses Ergebnis folgt mittels der Stirlingschen Formel von 1730 sofort. Weiterhin bestimmt er die Wahrscheinlichkeit, dass n-mal hintereinander dieselbe Münzseite fällt, zu $(0.5)^n$, speziell für n = 82 zu 1/ 4836 0000 0000 0000 0000.

3. Die Geburtenstatistik

John Graunt (1620-1674) hatte in seinem berühmten Buch „Natural and Political Observations...“, das 1676 bereits in 5. Auflage erschienen war, eine Liste mit den Anzahlen der in London in den Jahren 1629 bis 1664 getauften Knaben und Mädchen zusammengestellt und daraus die Stabilität der geburtlichen Sexualproportion von 14 Knaben auf 13 Mädchen erkannt.

Arbuthnott veröffentlichte als Anhang seiner Arbeit [2] eine analoge Liste, die sich auf die 82 Jahre 1629 bis 1710 erstreckt. Ihr entnahm er nur den Sachverhalt, den wir für später als das Ereignis A_{82} bezeichnen wollen, dass in allen diesen Jahren mehr Knaben als Mädchen geboren worden waren („and that in almost a constant proportion“). Er bediente sich nicht nur der Daten von Graunt, sondern übernahm auch dessen Argumentation, insbesonder gegen die Polygamie, ohne ihn selbst zu erwähnen (vgl.[4]).

4. Das Prinzip von Arbuthnott

Wird A_{82} als ein Ereignis im Binomialmodell aufgefaßt, so besitzt es nach Abschnitt 2 die Wahrscheinlichkeit $P(A_{82}) = (0.5)^{82} \approx 2 \cdot 10^{-25}$, die so verschwindend klein ist, dass das Ereignis A_{82} unmöglich eintreten kann. Hieraus schließt Arbuthnott, dass das beobachtete Geburtengeschehen nicht vom Zufall (-smodell) beherrscht wird, da es sonst ein praktisch unmögliches Ereignis hervorgebracht hätte. Diese statistische Schlussweise läßt sich nun einfach als Grundidee eines statistischen Hypothesentests formulieren, wobei eine Hypothese als eine Verteilungsannahme eines Zufallsmodells zu verstehen ist.

Prinzip von Arbuthnott:

Eine Hypothese wird verworfen, wenn unter dieser Verteilungsannahme ein praktisch unmögliches Ereignis eingetreten ist.

Arbuthnott hat damit berechtigt die Annahme abgelehnt, dass die Wahrscheinlichkeit p für eine Knabengeburt gleich 0,5 ist. Allerdings war damit bereits Graunt vertraut. Niklaus Bernoulli (1687-1759) bemühte sich um die Anpassung der Beobachtungswerte an das Binomialmodell mit dem geschätzten Wert p = 0,5163. Der Weg zum Signifikanztest mit einem kritischen Bereich, der an die Stelle von praktisch unmöglichen Ereignissen tritt, wurde erst durch Ronald A. Fisher (1890-1962), Jerzy Neyman (1894-1981) und Egon S. Pearson (1895-1980) konsequent beschritten.

5. Der Vorzeichentest

Bezeichnen x_k und y_k die Anzahl der Knaben- bzw. Mädchengeburten im k-ten Jahr, so interessiert bei Arbuthnott insbesondere der beobachtete Knabenüberschuß $d_k = x_k - y_k > 0$ für $k = 1,...,82$. Der erste Schritt zu einem statistischen Test erfordert die Interpretation der Beobachtungswerte als Realisierungen von Zufallsgrößen und damit die Einführung gewisser Verteilungsannahmen.

Basisannahme der klassischen Statistik (BS):

$(X_1,Y_1),...,(X_n,Y_n)$ sind unabhängige identisch verteilte Zufallsvektoren.

Wir untersuchen mittels $D_k = X_k - Y_k$, $k = 1,...,n$ zufällige Unterschiede. Die Annahme, es gäbe verteilungsmäßig keinen Unterschied, bezeichnen wir als Nullhypothese H_0 :

$$P(D_k \leq t) = P(-D_k \leq t), \text{ für alle } t \,\varepsilon\, \mathbf{R} ,$$

also $\quad F(t) = 1 - F(-t)$ für eine stetige Verteilungsfunktion F. (1)

Die Annahme eines stochastisch positiven Unterschieds bezeichnen wir als Gegenhypothese H :

$$P(D_k > t) \geq P (-D_k > t), \text{ für alle } t \,\varepsilon\, \mathbf{R} ,$$

wobei für gewisse t die strenge Ungleichung gilt.

Im zweiten Schritt ist mittels einer Testgröße ein kritischer Bereich zu konstruieren. Beim Vorzeichentest wird der kritische Bereich **K** über die Testgröße S bestimmt: $S = \| \{k: D_k > 0\} \|$ - Anzahl der positiven Differenzen.

Nach dem Prinzip von Arbuthnott sollte $P_{Ho}(S\varepsilon\mathbf{K})$ sehr klein sein, wenn wir die Hypothese H_0 zu Gunsten der Alternative H verwerfen. Wählen wir $\mathbf{K} = (z,\infty)$, so wird H_0 abgelehnt, wenn S die kritische Zahl z übertrifft. Den Term $P_{Ho}(S\varepsilon\mathbf{K})$ können wir auch als Irrtumswahrscheinlichkeit dafür interpretieren, die Hypothese H_0 zu verwerfen, obwohl sie wahr ist (Fehler 1.Art). Wird ein kleines, dem Problem angepasstes Signifikanzniveau a vorgegeben, so wählen wir dieses als obere Schranke für den Fehler 1.Art und erhalten damit eine Vorschrift, die bisher noch offene kritische Zahl z zu ermitteln: $\quad P_{Ho} (S > z) \leq a$. (2)

Unter (BS) bilden die Differenzen $D_1,...,D_n$ ein Bernoulli-Schema mit der Erfolgswahrscheinlichkeit $p = P(D_1 > 0) = 1 - F(0)$. Wenn H_0 wahr ist, so folgt aus (1) gerade $F(0) = 0{,}5$. Damit ist S unter H_0 binomialverteilt

Bi(n ; 0,5) mit den Parametern n und p = 0,5 und wir können (2) auswerten, indem wir das kleinste natürliche $z = z_{1-a}$ bestimmen gemäß

$$P_{Ho}(S > z) = (0,5)^n \sum_{j>z}^{n} \binom{n}{j} \leq a \quad ,$$

dabei ist z_{1-a} das Quantil von Bi(n ; 0,5) der Ordnung 1-a, das zum Beispiel in [5] tabelliert vorliegt.

Auf den Fehler 2.Art, bei wahrer Gegenhypothese die Nullhypothese anzunehmen, kann hier nicht mehr eingegangen werden.

Neben dem gerade konstruierten einseitigen Test von { p = 0,5 } mit der Alternative {p > 0,5} lässt sich analog ein zweiseitiger Test mit der Alternative { p ≠ 0,5 } angeben (vgl. [6]).

Unsere Herleitung des allgemeinen Vorzeichentests kann nicht unmittelbar auf den von Arbuthnott untersuchten Fall angewendet werden, da die Verteilungsfunktion der Geburtenzahlen nicht stetig ist. Eine naheliegende Modifizierung, auch für diskrete Verteilungen, ist in [7] zu finden.

Literatur

[1] Jürgen Lehn, Helmut Wegmann: Einführung in die Statistik. B.G.Teubner, Stuttgart 1992

[2] John Arbuthnott: An Argument for Divine Providence, taken from the constant Regularity observ'd in the Birth of both Sexes. Philosophical Transactions 27 (1712), 186-190

[3] J. A. (anonym): Of the Laws of Chance. B.Motte, London 1692

[4] Anders Hald: A History of Probability and Statistics and Their Applications before 1750. John Wiley, New York 1990

[5] Paul H. Müller, Peter Neumann, Regina Storm: Tafeln der mathematischen Statistik. Fachbuchverlag, Leipzig 1973

[6] Peter J. Bickel, Kjell Doksum: Mathematical Statistics. Holden-Day, Inc., San Francisco 1977

[7] J. Hemelrijk: A theorem on the sign test when ties are present. Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen, Ser.A, Mathematical sciences Vol. 55, No. 3,p. 322-326, 1952