



STATISTICIANS of the CENTURIES



C.C. HEYDE • E. SENETA • *Editors*

STATISTICIANS
OF THE
CENTURIES



Springer Science+Business Media, LLC

STATISTICIANS OF THE CENTURIES



C.C. HEYDE • E. SENETA
Editors

P. CRÉPEL • S.E. FIENBERG • J. GANI
Associate Editors



Springer

C. C. HEYDE
School of Mathematical Sciences
Australian National University
Canberra, ACT 0200
Australia
and
Department of Statistics
Columbia University
New York, NY 10027
USA
chris@maths.anu.edu.au

E. SENETA
School of Mathematics and Statistics
University of Sydney
Sydney, NSW 2006
Australia
eseneta@maths.usyd.edu.au



LIBRARY OF CONGRESS CATALOGING-IN-PUBLICATION DATA

Statisticians of the centuries / C.C. Heyde, E. Seneta, editors.

p. cm.

Includes bibliographical references and index.

ISBN 978-0-387-95283-3 ISBN 978-1-4613-0179-0 (eBook)

DOI 10.1007/978-1-4613-0179-0

1. Statisticians—Biography. I. Heyde, C.C. (Christopher Charles), 1939–

II. Seneta, E. (Eugene), 1941–

QA276.156 .S73 2001

519.5'092'2—dc21

[B]

2001031422

Printed on acid-free paper.

© 2001 Springer Science+Business Media New York
Originally published by Springer-Verlag New York, Inc. in 2001
Softcover reprint of the hardcover 1st edition 2001

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use of general descriptive names, trade names, trademarks, etc., in this publication, even if the former are not especially identified, is not to be taken as a sign that such names, as understood by the Trade Marks and Merchandise Marks Act, may accordingly be used freely by anyone.

Manufacturing supervised by Jerome Basma.
Book design by Steven Pisano.
Typeset by MacroTeX Book Publishing Services, Savoy, IL.

9 8 7 6 5 4 3 2 1

ISBN 978-0-387-95283-3

PROBABILITY PRIOR TO PASCAL



Summary. Early calculations in probability prior to the Pascal-Fermat correspondence of 1654 were related either to combinatorial calculations for simple games of chance or to attempts at a solution to the problem of points.

The generally accepted birth date for the probability calculus is 1654, the year of the Pascal-Fermat correspondence. Prior to that time some probability calculations were made and published. Elementary concepts of the probability calculus were extant since at least the Late Middle Ages. For the most part early probability calculations were motivated by games of chance. One of the great conundrums in the history of probability is that games of chance have provided simple models under which many probability calculations have been made, but the gambling that is associated with games of chance has had little impact on the development of the probability calculus. Until it was demonstrated in the mid-eighteenth century that probability could be used to improve the strategy of play, gambling manuals focused mainly on how to combat cheating and on the rules of play.

There are two separate strands to the history of probability prior to the mid-seventeenth century. Both strands are connected through games of chance. One strand follows elementary combinatorial calculations to evaluate events in a finite sample space, almost invariably connected with cards or dice. The second strand follows a long standing problem among mathematicians. This is the problem of points. Simply put, this problem addresses the question of how the stakes should be divided when a game of chance is interrupted in the middle. Along this strand, until Pascal (q.v.), Fermat (q.v.) and Huygens (q.v.) entered the scene, very little in terms of probabilistic arguments was used to attempt a solution to this problem.

Following the first strand, the earliest known probability calculations appear in a poem entitled *De Vetula*. The poem was purported to have been written by the Roman poet Ovid (43 BC–17 AD), but modern scholars place it in thirteenth century France, Paris in particular, and some scholars attribute the poem to Richard de Fournival. The poem was written in three books, the first of which is devoted to love affairs and amusements in youth. Among these amusements are chess and dicing. The second book recounts the ups and downs of a love affair and the woman in the affair becomes the *vetula* (old woman) of the poem's title. In the third book there is an account of philosophical studies in later life and in the course of this the writer becomes convinced of the truths of Christian faith. It is in the first book that the probability calculations appear. Here, calculations are given for the probability of each kind of throw of three dice and for the sum of the faces which show on three dice.

De Vetula was a medieval bestseller. It was widely read and several medieval writers quoted from it, among them Roger Bacon, the English philosopher and scientist, and Thomas Bradwardine, the English theologian, mathematician and philosopher who became Archbishop of Canterbury in the last year of his life. Another one of those who had read *De Vetula* was Nicole Oresme. Some of his work contains probability calculations and probabilistic arguments related to the question of the commensurability or incommensurability of celestial motions.

Over the next three or four centuries several others occupied themselves at one time or another with the throw of three dice. A 1477 commentary on Dante's *Purgatorio* written by Benvenuto da Imola contains a discussion about the sum of the faces which show on three dice. It is noted that 3, 4, 17 and 18 are each less likely than each of the sums in between, but no calculations are given and it is unclear whether the writer understood that there is a difference in probabilities between the sums 3 and 4. In the 1520s the physician and mathematician Girolamo Cardano wrote a gambling manual *Liber de Ludo Aleae*. In manuscript form for over one hundred years, it was not published until 1663 when Cardano's collected works were published. Among many other things Cardano calculates probabilities for the sums of the faces which show on two or three dice and applies his results to some dicing games. Finally, sometime between 1613 and 1623 Galileo made very detailed calculations of the probability of the sums of the faces which show on three dice.

Cardano's manuscript also contains probability calculations related to the card game *primero*, the forerunner of modern poker, and to the throw of *astragali* or knucklebones. *Liber de Ludo Aleae* is much more than an book on probability. Cardano was an inveterate gambler and at times supported himself through gambling. The book is actually a gambling manual, the most thor-

ough one produced up to its time. This includes the printed version of 1663, several decades thereafter. In addition to the probability calculations there are informative and detailed discussions of methods of cheating at cards and dice so that the player might beware.



GIROLAMO CARDANO

b. 24 September 1501 • d. 21 September 1576

There are other hints in several places that a rudimentary probability calculus was well-known prior to 1654. For example, the Pardoner in Chaucer's *Canterbury Tales* states two events which are the sums of throws of two dice. The events are assigned as the chances between the Pardoner himself and another person. Both events, throwing a seven and throwing a three or five, are equiprobable.

The problem that led to the development of the probability calculus by Pascal and Fermat, followed three years later by Huygens's *De Ratiociniis in Ludo Aleae*, was the problem of points. This is the second strand to the history of the development of the probability calculus. The problem of points, or the division of stakes in a game which is terminated early, was a long-standing mathematical problem rather than a practical gambling problem. Most serious gamblers would continue until the game was finished or would cast lots for the stake. Discussions of the problem of points have all so far been found in mathematical works while the strictly combinatorial probabilities associated with dicing and cards have been found in other works. This assumes that Cardano's *Liber de Ludo Aleae* is treated as a gambling manual rather than a work in mathematics.

The earliest known result for the problem of points is from a manuscript dated to about 1400. It is part of a collection of arithmetic and algebraic works.

The author tackles the problem for the situation in which the first player to win three games, among two playing the game, wins the stake. Using algebraic, rather than combinatorial or probabilistic arguments, the unknown author correctly solves the problem when the game is stopped after one player is leading the other by a score of two to zero. However, when the score is three to zero and the object is to win four games, the author has trouble with the solution and gives up without completing it. The first printed reference to the problem of points is in a 1494 work by Luca Pacioli, a Franciscan friar and itinerant mathematics teacher, who eventually became a professor of mathematics at the University of Milan. His solution to the problem of points, which is completely fallacious, also involves no probability arguments or combinatorics. Throughout the sixteenth century several other Italian mathematicians, including Cardano, Tartaglia, Peverone and Forestani tried their hands at this problem and published what little success they had with the problem in their mathematical works. Within this group the problem of points appears to be a mathematical curiosity of intellectual rather than practical interest.

The few probability calculations that have been made prior to 1654 have been reviewed in Coumet (1965), David (1955), David (1962), Hald (1990), Kendall (1956), Schneider (1988), and Todhunter (1865). Some other discussions about early probabilistic concepts and possible probability calculations are found in Bellhouse (1993), Bellhouse and Franklin (1997), Hacking (1975), and Rabinovitch (1973).

— DAVID BELLHOUSE

Bibliography

- BELLHOUSE, D.R. (1993). The role of roguery in the history of probability. *Statistical Science*, **8**, 410–420.
- BELLHOUSE, D.R. AND J. FRANKLIN (1997). The language of chance. *International Statistical Review*, **65**, 73–85.
- COUMET, E. (1965). Le problème des partis avant Pascal. *Archives internationales d'histoire des sciences*, **18**, 245–272.
- DAVID, F.N. (1955). Dicing and gaming (a note on the history of probability). *Biometrika*, **42**, 1–15.
- DAVID, F.N. (1962). *Gods, Games and Gambling*. Griffin, London.
- HACKING, I. (1975). *The Emergence of Probability*. Cambridge University Press, Cambridge.
- HALD, A. (1990). *A History of Probability and Statistics and Their Applications before 1750*. Wiley, New York.



THOMAS BAYES



b. c. 1701

d. 7 April 1761

Summary. The problem of passing from a population to the properties of a sample was one of the first studied in probability. Thomas Bayes, a nonconformist minister, was the first to solve the inverse problem of passage from sample to population, using ideas that are widely used today.

Thomas Bayes, born in London, the son of a nonconformist minister, spent most of his adult life in a similar position in Tunbridge Wells, England. He was educated at Edinburgh University and was elected a fellow of the Royal Society in 1742. During his lifetime he published a few mathematical papers, of which the best-known is a 1736 defence of Newton's ideas against an attack by Bishop Berkeley. He is today remembered for a paper that his friend Richard Price claimed to have found amongst his possessions after death. It appeared in the Society's Transactions in 1763 and has often been republished. Apart from these bare facts, surprisingly little is known of Bayes' life.

By the middle of the 18th century it was well-understood that if, to use modern terminology, in each of n independent trials, the chance of success had the same value, θ say; then the probability of exactly r successes was given by the binomial distribution

$$P(r|\theta, n) = \binom{n}{r} \theta^r (1 - \theta)^{n-r}.$$

Jakob Bernoulli (q.v.) had established the weak law of large numbers and de Moivre (q.v.) had found the normal approximation to the binomial. The passage from a known value of θ to the empirical observation of r was therefore

extensively appreciated. Bayes studied the inverse problem; what did the data (r, n) say about the chance θ ? Partial answers already existed. For example, Arbuthnot (q.v.) had observed r male, and $n - r$ female, births with r considerably in excess of $1/2n$. He argued that, on the basis of the binomial with $\theta = 1/2$, a value of r as high as this was so improbable that θ could not be $1/2$. That idea has been much extended into the modern form of a significance test and its associated P -value or significance level.

Bayes proceeded differently, using the theorem that nowadays always bears his name, though it does not appear explicitly in the 1763 paper,

$$P(A|B) = P(B|A)P(A)/P(B)$$

for events A, B with $P(B) \neq 0$. The theorem permits the inversion of the events in $P(B|A)$ into $P(A|B)$. Applied when A refers to θ and B to the empirical r , we have

$$P(\theta|r, n) \propto P(r|\theta, n)P(\theta|n).$$

(The missing constant of proportionality does not depend on θ . It is $P(r|n)^{-1}$ but is most easily found by making the product integrate to one by multiplying by the constant). The result effects the passage from the binomial, on the right, to a probability statement about the change, on the left. It therefore becomes possible to pass from the data to a statement about what are probable, and what are improbable, values of the chance.

This elegantly and simply solves the problem, except for one difficulty. It requires a value for $P(\theta|n)$, a probability distribution for the chance before the result of the trials has been observed. It is usual to describe this as the prior distribution (prior, that is, to r) and the final result as the posterior distribution. Thus the theorem describes how your views of θ change, from prior to posterior, as a result of data r . Bayes discussed the choice of prior but his approach is ambiguous. He is usually supposed to have taken $P(\theta|n)$ uniform in $(0,1)$ —the so-called Bayes's postulate—but an alternative reading suggests he took $P(r|n)$ to be uniform. Mathematically these lead to the same result.

Little notice was taken of the 1763 paper at the time. It was first appreciated by Laplace (q.v.) in the early years of the next century, who used the ideas in his eclectic approach to probability. The theorem is of basic importance because it provides a solution to the general problem of inference or induction. Let H be a universal hypothesis and E empirical evidence bearing on H . A simple example might be H , all swans are white, and E the observation of the colour of a swan. A more sophisticated one would have H as Newton's laws and E observation of the motions of the planets. In either case, $P(E|H)$ can

be calculated. Bayes's theorem says

$$P(H|E) \propto P(E|H)P(H),$$

expressing a view about the hypothesis, given the evidence, in terms of the known probability of the evidence, given the hypothesis, and the prior view about H . As more evidence supporting H accrues, having large probability on H , so even the sceptic, with low $P(H)$, will become convinced, $P(H|E)$ will approach one and the hypothesis will be accepted. Many people, following Jeffreys (q.v.), who extensively developed these ideas into a practicable scientific tool, hold that this provides a description of the scientific method. This view differs from that of Popper, who only admits refutation of a hypothesis and whose attitude to probability is regarded as unsound by supporters of Jeffrey's ideas.

Recently Bayes's theorem has been used as a means of processing evidence in a court of law. Let G be the hypothesis that the defendant is truly guilty of the offence with which he or she has been charged and E a new piece of evidence. Then applying the theorem both to G and to $\bar{G}$, denoting innocence,

$$\frac{P(G|E)}{P(\bar{G}|E)} = \frac{P(E|G)}{P(E|\bar{G})} \frac{P(G)}{P(\bar{G})}.$$

The expression on the left is the odds on guilt, given E ; that on the right is the same odds without E . The remaining term is the likelihood ratio, being a comparison of the probability of the evidence, supposing guilt, to the same probability supposing innocence. Forensic scientists present evidence to the court in the form of a likelihood ratio. The court can then multiply their former (prior) odds by the likelihood ratio to obtain new (posterior) odds as a result of hearing the evidence.

Bayes's result has attained increased importance as a result of work by Ramsey, de Finetti and Savage between 1925 and 1955, which demonstrated that our knowledge had to be based on probability and that our beliefs must obey the rules of the probability calculus, of which Bayes's is essentially the multiplication rule. In this view, the significance levels of the classical school are unsound because they do not express opinions about hypotheses like H , or parameters like θ , in terms of direct probabilities of H or θ . The resulting methodology is called Bayesian statistics and has rather different procedures and results from those of the classical school. Bayesians regard probability as a measure of a person's belief, whereas the classical school only admits probability as a frequency concept.

Bayes's result is also central to modern ideas on decision-making under uncertainty. Suppose there is a choice to be made amongst a set d of decisions

in the presence of uncertainty about a parameter θ . The work of Ramsey and others leads to the introduction of a utility function $u(d, \theta)$, describing the worth of decision d when the parameter has the value θ , and the choice of that d which maximizes the expected utility $\sigma_{\theta}u(d, \theta)P(\theta)$. Additional evidence E updates $P(\theta)$ to $P(\theta|E)$, by the theorem, and improves the decision-making. All this is a long way from Bayes's original problem and its resolution. He would doubtless be astonished were he to realize how his wonderful idea has been extended and his name used.

— D.V. LINDLEY

Bibliography

The original paper appeared in *The Philosophical Transactions of the Royal Society of London* (1763). **53**, 370–418. There is a reprint in *Biometrika* (1958). **45**, 296–315. An illuminating commentary on it is provided by S.M. STIGLER (1982). Thomas Bayes's Bayesian inference. *Journal of the Royal Statistical Society, Series A*, **145**, 250–258. The most complete biography is provided by A.W.F. EDWARDS in the latest edition of *The Dictionary of National Biography*. Two recent books on modern Bayesian methods are A. O'HAGAN (1994). *Bayesian Inference*. Vol. 2B of Kendall's *Advanced Theory of Statistics*. Edward Arnold, London; John Wiley, New York. J.M. BERNARDO AND A.F.M. SMITH (1994). *Bayesian Theory*. John Wiley, Chichester. The latter is part of a forthcoming 3-volume work and has an extensive bibliography. The modern "classic" is B. DE FINETTI (1974/5). *Theory of Probability*. John Wiley, London, in 2 volumes, translated from the Italian. C.G.G. AITKEN (1995). *Statistics and the Evaluation of Evidence for Forensic Scientists*. John Wiley, Chichester, deals with legal applications. D.V. LINDLEY (1985). *Making Decisions*. John Wiley, London, extends Bayesian ideas to decision-making.