CHANCE AS A TOOL

Jan Coufal – Petra Dotlačilová – Jiří Tobíšek

Abstract

In the 19th century, something significant began to happen in statistics, among other sciences. It no longer limited itself to merely registering data on the state of the treasury and the state in general, but began to deal with what we now call statistical inference. Statistics from applications in economics also made its way into the natural sciences. Among the first steps was the development of the χ^2 test in 1900. It is a statistical test for verifying the significance of the difference between two sets of random data, proposed by K, Pearson. Soon after him, in 1908, W. S. Gossett proposed his tests of probable error in estimating the mean. A significant contribution is played by the work of R. A. Fisher, who has moved the industry forward significantly. The true boom of statistical methods in biology occurred in the middle of the twentieth century with the increasingly easy availability of powerful computers. C. R. Rao was one of the founders of modern statistics. He became famous for his anthropometric research, solving meteorological problems, statistical data for the development of Indian agriculture and also for the economic concept of India's development.

Key words: K. Pearson, statistical inference, W. S. Gossett, R. A. Fisher, C. R. Rao

JEL Code: B16, C10

Introduction

The history of mathematical statistics (as well as other scientific disciplines) usually proceeds relatively smoothly, in small steps, which gradually add new knowledge and context. Sometimes there is a flash of lightning somewhere, something unexpected happens, suddenly we know how to put an egg on the tip, or that there are infinitely many different infinities. It is about such moments when there was a flash, and about people who had ideas that no one else enlightened, that is this text. A lot of time has passed between the moment when our ancestors first noticed that there is more of something and less of something, and today, with the ubiquitous stochastic models, – exact thinking and knowledge have evolved, sometimes smoothly, sometimes with surprising turns. Many people think that statistics started with counting, but statistics appeared even earlier. It was then when people felt the need to talk and think about the world around them in contradictory terms such as small–large, much–little, near–far and so on. Only later came the terms more–less, followed by how many–so much. Practical counting, which left the first traces of statistics in archaeological finds, came later. Statistics did not start at the beginning and still do not start with numbers. She has been much more concerned with quantitative and qualitative relationships. People didn't start inventing statistical models for fun. They needed them to live, and they needed them so much that they devoted time and intellectual capacity to them, which they could seemingly use more effectively to ensure sustenance and warmth in the cave. It should be noted that it is not only statisticians who need to be familiar with statistical models. They make our world what it is for us, and what we can sometimes not get lost in, and humanity has not invented a simulator to understand such models that would be roughly comparable to statistics. After this excursion into prehistory and the beginnings of history, because the beginnings of statistics were (at least as far as time is concerned) very long – the first material monuments that can have a statistical interpretation can be dated to the period between 30,000 and 35,000 years ago. The article will go to the time when medieval science, which needs arithmetic, geometry and logic in an exact form, gradually began to become a modern science, which was about half a millennium ago. The world began to change. The Renaissance was reminiscent of antiquity, trade and production developed, and space opened up for natural sciences. The development of discoveries from seafaring and the exploration of the inland of unknown continents. All this forced people to look for answers to questions – what does something look like and how does it behave that way, then why does it look like that and why does it behave that way. The first two questions were mostly satisfied with observation or hunting expeditions, the other ones necessarily needed some generalizing and usually at least a little quantitative model of objects and events that were the focus of interest. The Renaissance and the modern age brought efforts to liberate from the dogmas of the Church to explain many phenomena.

1 Probability

The origin of something like the calculus of probability is usually placed in the period of the reign of Louis XIV in France. By Mošna (2022) it is usually associated with the names of important scholars Blaise Pascal (1623–1662) and Pierre de Fermat (1601–1665). However, marking the beginning of a field and linking it to a certain person and country is always very problematic and misleading, and evidence of older occurrence or mentions often appears later. The beginning of probability calculation in the modern period is generally considered to be the correspondence between Pascal and de Fermat in the summer and autumn of 1654. Another mathematician dealing with chance, Christiaan Huygens (1629–1695), was a lawyer and also a mathematician, physicist and astronomer. During his stay in Paris, he became acquainted with the correspondence of de Fermat and Pascal. He developed and summarized their ideas in *De ratiociniis in ludo aleae* (Reflections on the Dice Game, 1657). If we talk about the introduction of probability, it is necessary to mention another of the great mathematicians dealing with this discipline, the Frenchman Abraham de Moivre (1667–1754). As a Protestant, he was imprisoned after the revocation of the Edict of Nantes and forced to leave France for Great Britain three years later. Despite his undeniable genius, as a foreigner he was forced to make a living as an advisor to gamblers. His reflections on gambling and his own experience are contained in his gradually expanded textbook on probability theory, *The doctrine of chances: a method of calculating the probabilities of events in play* (1718, 1738, 1756). Among the first to bring mathematical precision to relatively intuitive statistics was T. Bayes and P.–S. Laplace. The proverbial red thread running through the history of probability and statistics is the concept of probability by Thomas Bayes (1701–1761), which is published in his work on probability An essay towards solving a problem in the doctrine of chances, which was published by R. Price in 1763, i.e. two years after Bayes' death. This process was completed and summarized in the work of Pierre-Simon Laplace (1749–1827), who laid the foundations of classical probability in his *Théorie analytique des probabilités* (Analytical Theory of Probability, 1812, 1820). It summarizes the knowledge of this field and thus completes the efforts of many mathematicians. By Coufal at all (2019) the Prague theologian, philosopher and mathematician Bernard Bolzano (1781–1848) dealt with probability in his works *Lehrbuch der Religionswissenschaft* (Textbook of Religion, 1834) and *Wissenschaftslehre* (Science, 1837). His conception of probability is close to the logical theories that were later brought by philosophers in later times, from Wittgenstein to Carnap. Probability gradually began to assert itself in fields other than mathematics and physics. It turned out that statistically collected data from different fields have properties known from the calculus of probability. Jacob Bernoulli came up with the so-called Golden Theorem, or the law of large numbers, which represented an important combination of statistics and probability. The relative frequencies were close to the average. Furthermore, central limit theorems and normal distributions were an important discovery for statistics and were also used in other fields. In addition to measuring errors in astronomy, the normal distribution is also reflected in statistical surveys in the social sciences. By Mošna (2022) the importance of this division in biometrics was strongly realized by the Belgian astronomer and versatile scientist Lambert Adolphe Quételet (1796–1874) and the inventor of the calculating machine Charles Babbage (1791–1871). The idea of a normal division began to be applied more and more often in science. The idea that some kind of strict law is applied in confusion and chaos has become generally accepted in the world of science. The idea that some kind of strict law is applied in confusion and chaos has become generally accepted in the world of science. The great spirit of Victorian England, statistician, biologist, anthropologist, psychologist, geographer and founder of eugenics, Francis Galton (1822–1911), made significant strides in the formulation and application of statistical methods. In statistics, in addition to the role of the mean value, he also recognized the importance of variance and the median value, i.e. the median (for the first time in 1874), and he also introduced percentiles and interquartile deviation (which is half of the interquartile range). In the following sections, the text is devoted to four men (K. Pearson, R. A. Fisher, W. S. Gossett and C. R. Rao), whose work had a major impact on the development of mathematical statistics, as they significantly developed methods of statistical induction.

2 Karl Pearson

By Pearson (1938, 1968, 1970), Norton (1978) and Porter (2004) Karl Pearson (27 March 1857, London – 27 April 1936, Coldharbour) was an English mathematician and philosopher. In 1911, he founded the first ever department of statistics at University College London. In his fundamental philosophical work The Grammar of Science (1892) he deals with questions of the methodology of science. According to him, the role of science is based on the classification and description of facts. Material things represent groups of sense perceptions and laws of nature. Space and time are creations of human reason. He studied at University College London, then entered the Royal University of Cambridge to study mathematics. Between 1879 and 1880 he studied medieval literature and 16th-century German literature at the universities of Berlin and Heidelberg. He graduated from the University of Cambridge in 1879. He then travelled to Germany to study physics at the University of Heidelberg, where he attended lectures by the famous physiologist Emile du Bois-Reymond on Darwinism. He also studied Roman law, medieval and 16th-century literature as well as socialism in Berlin. He then returned to London to study law, as his father had done, so that he could work at the bar. His next career was to join one of the important associations for lawyers, where he lectured law until 1881 (although he never practiced law). He subsequently returned to mathematics when he became Professor of Mathematics at King's College in 1881 and Professor at University College in 1883. Pearson's work was applied in the development of mathematical statistics for the fields of biology, epidemiology, anthropometry, medicine and social history. In 1901, with Weldon and Galton, he founded the journal Biometrika, the subject of which was the development of statistical theories, which he himself edited until his death. He also founded the journal Annals of Eugenics (now Annals of Human Genetics). Pearson created many of the classical statistical methods that are commonly used today. His contributions include:

The moment method consists in comparing the first general moments with the values of their sample counterparts; the parametric correlation coefficient is a statistical test (assuming a normal distribution) to determine how close the relationship between variables is; χ^2 distribution is derived from the sum of independent random variables with a normalized normal distribution; χ^2 test (or Pearson's χ^2 goodness-of-fit test) is any test of a statistical hypothesis whose test criterion has a χ^2 distribution, assuming the null hypothesis is valid; χ distance is based on the correlation between variables that can be identified and analysed using patterns; The *p*-value is a numerical value used in statistical hypothesis testing. All this is related to the basics of statistical hypothesis testing theory and statistical decision theory, where Pearson proposed testing the validity of the assumed values of the evaluation χ of the distance between the hypothesis and the empirically determined value using the *p*-value that was proposed in the same paper.

2.1 Pearson's More Important Publications

Pearson, K (1896). Mathematical Contributions to the Theory of Evolution. III. Regression, Heredity and Panmixia. *Philosophical Transactions of the Royal Society of London* 187. pp. 253–318.

Pearson, Karl, & Whiteley, M.A. (1899). Data for the Problem of Evolution in Man, I: A First Study of the Variability and Correlation of the Hand. *Proceedings of the Royal Society of London*, Vol. LXV, pp. 126–151.

Pearson, Karl; Beeton, M., & Yule, G.U. (1900). On the Correlation Between Duration of Life and the Number of Offspring, *Proceedings of the Royal Society of London*, Vol. LXVII, pp. 159–179.

Pearson, Karl (1904). *On the Theory of Contingency and its Relation to Association and Normal Correlation*. London: Dulau & Co.

Pearson, Karl (1906). *A Mathematical Theory of Random Migration*. London: Dulau & Co.

Pearson, Karl (1913). *On the Correlation of Fertility with Social Value*: *A Cooperative Study*. London: Dulau & Co.

3 Ronald Aylmer Fisher

By Box (1978), Bennett (1990) and Pearson (1968, 1970) Ronald Aylmer Fisher (17 February 1890, London – 29 July 1962, Adelaide) was an English statistician, evolutionary biologist and geneticist.For his work in statistics, he has been described as "a genius who almost single-handedly created the foundations for modern science" and "the single most important figure in 20th century statistics". Fisher was born in London in 1890. In 1912 he obtained a B.A. degree in mathematics at the University of Cambridge. On Fisher's initiative, Cambridge University Eugenics Society was founded in May 1911, whose chairman and spokesman Fisher became. He advocated eugenics and represented a position which is sometimes called positive eugenics, after which upper classes should receive incentives for a higher number of children; At the same time, members of other classes should not be held away. His work on the errors in astronomical calculations along with his interest in genetics led to his work in statistics. From 1919 he worked at the Rothamsted Experimental Station. In 1933 he became professor of Eugenik at University College London, and from there he moved to the Balfour Chair of Genetics in Cambridge in 1943. Fisher was elected to the American Academy of Arts and Sciences in 1934, to the American Philosophical Society in 1941 and to the National Academy of Sciences in 1948. He received numerous awards for his works, such as the Royal Medal of the Royal Society in 1938 and the Weldon Memorial Prize in 1930. In 1952 he was defeated by Queen Elisabeth II. In 1959 he received the Darwin plaque. In 1960 he was elected a member of the Leopoldina.Already in retirement, he spent some time in Adelaide, Australia, where he died in 1962. Fisher introduced the maximum-Likelihood principle and the Fisher information and also the statistical method of variance analysis goes back essentially to it. He gave significant contributions to statistical experimental planning and postulated the esteemed theoretical concepts of sufficiency and freedom of distribution (English ancillary statistic). This made him one of the most important statisticians of the 20. Century. His article On a distribution yielding the error functions of several well known statistics presents Karl Pearsons' χ^2 distribution and the Student's t-distribution in the same probabilitiestheoretical framework as the normal distribution and the F-distribution (or Fisher's t-distribution) named after him. Fisher's book Statistical methods for research workers describes how these distributions can be used. With the use of multiple measurements in taxonomic problems (1936), he introduced the Fisherian discriminance function, which is the basis for the development of discriminal analysis. He also intervened in biology when he used mathematics to develop Mendelian genetics and natural selection theory. This contributed to the revival of Darwinism and the revision of the evolution theory from the early 20th century known as modern synthesis. He's the founder of biostatistics and population genetics. He studied sexual selection (his basic concepts are Fisher's principle, Runaway selection and the hypothesis of "sexy sons"). He did his research for the sake of eugenics. Fisher also conducted experimental agricultural research at the Rothamsted Research Station, which saved millions of people from hunger through analysis of variance.

3.1 Fisher's More Important Publications

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population, *Biometrika* **10.** pp. 507–521

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society A* **222.** pp. 309–368,

Fisher, R. A. (1922a). The goodness of fit of regression formulae, and the distribution of regression coefficients. *Journal of the Royal Statistical Society*. Band 85, Nr. 4. pp. 597–612

Fisher, R. A. (1925). *Statistical methods for research workers*. Edinburgh : Oliver & Boyd,
Fisher, R. A. (1925a). Applications of Student's distribution. *Metron* **5**. pp. 90–104.
Fisher, R. A. (1928). The general sampling distribution of the multiple correlation coefficient. *Proceedings of Royal Society A*. Band 121. pp. 654–673. doi:10.1098/rspa.1928.0224
Fisher, R. A. (1934). Two new properties of mathematical likelihood. *Proceedings of Royal Society A* 144. pp. 285–307. doi:10.1098/rspa.1934.0050.
Fisher, R. A. (1935). *The design of experiments.* Edinburgh : Oliver & Boyd,
Fisher, R. A. (1949). *The theory of inbreeding*. Edinburgh : Oliver & Boyd,
Fisher, R. A. (1950). *Contributions to mathematical statistics*. New York : Wiley,
Fisher, R. A. (1956). *Statistical methods and statistical inference*. Edinburgh : Oliver & Boyd,

4 William Sealy Gossett

By Pearson (1968, 1970, 1990), Ziljak (2011, 2019) and Zabel (2008) William Sealy Gosset (13 June 1876. Canterbury – 16 October 1937, Beaconsfield) was an English statistician. He published under the pseudonym Student. His main work resulted in the student's *t*-distribution and the *t*-test, which allows the statistician to check whether the two averages of two samples differ significantly or not. The term studentisation also goes back to him: Studentisation (after the pseudonym Student of Gosset) is understood in mathematical statistics as a transformation of the realizations of a random variable, so that the resulting values possess the arithmetic mean zero and the empirical variance one.

4.1 Gosset's (Student's) biography

He studied chemistry and mathematics at New College in Oxford. After completing his studies in 1899 he began working in the Dublin brewery Arthur Guinness & Son. Guinness was an advanced agro-chemical plant, and Gosset applied its statistical knowledge both in brewery and in agriculture to achieve the best barley quality for beer production. William Gosset developed these skills in 1906/1907 in studies and experiments in the biometric laboratory of Pearson. Gosset and Pearson had a good relationship with each other, and Pearson helped with the mathematical work in Gosset's writings. Pearson also helped his studies in 1908, but had little attention to their importance: This work dealt with small sample sizes – a typical problem of a brewery, while a biometric is usually able to draw hundreds of samples and therefore requires no special methods for small sample sizes. Another scientist at the Guinness Brewery had published a work which – for the damage of the brewery – contained business secrets. In order to avoid further betrayal of confidential information, the brewery banned its employees from publishing any work. This meant for Gosset that he had to publish under a pseudonym; he chose Student. Its greatest achievement, the *t*-distribution, is therefore known as Student's *t*-distribution". Almost all his publications published Gosset under his pseudonym, also The probable error of a mean, which appeared in Pearsons Journal Biometrika. However, it was not

Pearson, but the statistician Fisher, who recognized the importance of Gosset's work on small sample sizes. In 1935 Gosset left Dublin to take over a scientific leadership position in the new Guinness brewery in London. Gosset was a friend of both Pearson and Fisher, a noteworthy, achievement for each had a massive ego and a loathing for the other.

4.2 Gosset's (Student's) More Important Publications

COSSETT, W. S. (1907). On the error of counting with hæmacytometer. *Biometrika*. 5 (3-February). Pp. 351–360. doi:10.1093/biomet/5.3.351.

COSSETT, W. S. (1908). The probable error of a mean. *Biometrika*. 6 (1-March). pp. 1–25. doi:10.1093/biomet/6.1.1.

COSSETT, W. S. (1908a). Probable error of a correlation coefficient. *Biometrika.* 6 (2/3-September): 302–310. September 1908. doi:10.1093/biomet/6.2-3.302.

COSSETT, W. S. (1909). The distribution of the means of samples which are not drawn at random. *Biometrika.* 7 (1/2- July–October): 210–214. 1909. doi:10.1093/biomet/7.1-2.210.

COSSETT, W. S. (1921). An experimental determination of the probable error of Dr Spearman's correlation coefficients. *Biometrika*. 13 (2/3-Juli). pp. 263–282. doi:10.1093/biomet/13.2-3.263.

5 Calyampudi Radhakroshnan Rao

By Banks et all (2023), Putanen et all (1996) Calyampudi Radhakrishna Rao (10 September 1920, Hadagali – 22 August 2023, Buffalo) was an Indian American mathematician and statistician. He was born as an eighth child of a Telugu family, which is why he received the first name Radhakrishna. He visited Mrs. A. V. N. College in Visakhapatnam before he studied mathematics in Andhra (M.A. 1940). He then studied one year at the Indian Statistical Institute (ISI) in Kolkata, where he published his first research work in 1941 (with K. Raghavan Nair) and acquired his Master in Statistics at the University of Calcutta (Final 1943). After that he was a research assistant at ISI and also lecturer at the University of Calcutta. In 1946 he moved to Cambridge, where he worked on statistical problems during the day at the Anthropological Museum and studied the chromosomes of mice in the evening at Ronald Fisher's Genetic Laboratory. He received his doctorate in 1948 with a dissertation on statistical problems of biological classification at the University of Cambridge. In the same year he returned to ISI where he became professor in 1949 and remained until his retirement in 1979. Srinivasa S. R. Varadhan, K. R. Parthasarathy and Varadarajan were his students there. In 1979 he went to the United States, where he was university teacher until 1987 at the University of Pittsburgh and then at Pennsylvania State University, at the latter from 1988 to 2001, and held the Eberly Chair in Statistics since 2001. From 2010 he was also a research professor at the Department of Biostatistics of the University of Buffalo. His perhaps most important work was published in 1945. It contained three basic results, which are regarded as groundbreaking for the research field of statistics and led to statistical tools which are now used in many scientific disciplines. The first is now known as Cramér-Rao-Unification and allows to determine when an estimation method cannot be further improved. The second result is referred to as the set of Rao-Blackwell and allows an estimation to be transformed into an optimal estimate. The third result showed a connection between information theory and geometry and led to the interdisciplinary research area of information geometry. Rao was a member of several national academies of sciences, including India and Italy, in the USA Fellow of the National Academy of Sciences and a member of the American Academy of Arts and Sciences and in the UK Fellow of the Royal Society. He was President of the International Statistical Institute, the International Biometric Society and the Institute of Mathematical Statistics.

5.1 Rao's Publications

By Putanen et all (1996) C. R. Rao is the author of 477 research papers published in prestigious journals and 15 books, one of which, Statistics and Truth were translated into French, German, Japanese, Main Land, Taiwan Chinese, Turkish, and Korean, and another book Linear Statistical Inference, used as a textbook, has been in the market for over 50 years and is available in Russian, German, Czech, Polish, Chinese, and Japanese languages. He has edited 39 volumes of Handbook of Statistics dealing with the latest methodologies in Statistics and practical applications.

Conclusion

They are suitable representatives of the period when mathematics and mathematical statistics have changed into a consistently exact science in the last two centuries. The progress of statistics (very briefly summarized in this article) was certainly due to many people, but four of them, who were remarkable not only scientifically, but also for various reasons humanly, is the content of this contribution. Perhaps there are appropriate representatives of what has changed in the last two centuries of statistics into a rigorously exact science, which, precisely due to its exactify, can afford to intervene in yet recently un-mattized fields and which is cultivated by an international community of experts, for whom borders and continents are of secondary importance, if any.

Acknowledgment

This contribution is a follow-up to the project of University of Economics and Management Prague.

References

BANKS, D., & CLARKE. J. L. (2023). C. R. Rao (1920–2023). Pioneering statistician and father of information geometry. *Science*. Band 382, Nr. 6672. p. 771, doi:10.1126/science.adl1762.

BENNETT, J. H. (ed.) (1990). *Statistical Inference and Analysis: Selected Correspondence of R. A. Fisher*. Oxford : Clarendon Press.

BOX, J. F. (1978). *R. A, Fisher, The life of a scientist.* New York : Wiley.

COUFAL, J., & BREZINA, P. (2019). Bernard Bolzano and Theory of Probability. *Ekonomické listy.* Vol. 10. pp. 27–46. ISSN 1804-4166. URL: https://www.vsem.cz/data/data/ces-soubory/ekonomicke_listy/ekonomicke_listy_01_2019.pdf.

MOŠNA, F. (2022). *Pojetí pravděpodobnosti a statistiky ve výuce,* Praha : Vydavatelství UK PedF. ISBN 9788076033399.

NORTON, B. J (1978). Karl Pearson and Statistics: The Social Origins of Scientific Innovation. *Social Studies of Science*. **8** (1). pp. 3–34. doi:10.1177/030631277800800101

PEARSON, E. S. (1938). *Karl Pearson*: *An Appreciation of Some Aspects of his Life and Work*. Cambridge : Cambridge University Press.

PEARSON, E. S. (1968). Some early correspondence between W S Gosset, R A Fisher and Karl Pearson, *Biometrika* 55. pp. 445-457.

PEARSON, E. S. (1970). Some early correspondence between W S Gosset, R A Fisher and Karl Pearson, *Studies in the History of Statistics and Probability. p*p. 405-418.

PEARSON, E. S. (1990). *Student – A Statistical Biography of William Sealy Gosset.* Oxford : Oxford University Press

PORTER, T. M. (2004). *Karl Pearson*: *The Scientific Life in a Statistical Age*. Princeton : Princeton University Press. ISBN 978-0-691-12635-7.

PUNTANEN, S., & STYAN, G. P. H. (1996) A Brief Biography and Appreciation of Calyampudi Radhakrishna Rao, with a Bibliography of His Books and Papers. *Linear Algebra and its Applications*. Band 237–238, April. pp. 1–40, doi:10.1016/0024-3795(95)00027-5

ZILIAK, S. T. (2011). W.S. Gosset and Some Neglected Concepts in Experimental Statistics: Guinnessometrics II. *Journal of Wine Economics*. 6 (2). pp. 252–277. doi:10.1017/S1931436100001632

ZILIAK, S. T. (2019). How Large Are Your G-Values? Try Gosset's Guinnessometrics when a Little "p" is Not Enough. *The American Statistician*. 73. pp. 281–290. doi:10.1080/00031305.2018.1514325

ZABELL, S. L (2008). On Student's 1908 Article The Probable Error of a Mean. *Journal of the American Statistical Association.* 103 (481). pp. 1–7. doi:10.1198/016214508000000030

Contact

Jan Coufal & Jiří Tobíšek

University of Economics and Management, Nárožní 2600/9a, 158 00 Prague 5, Czechia

jan.n.coufal@seznam.cz & jiri.tobisek@vsem.cz

Petra Dotlačilová

Prague University of Economics and Business, Faculty of Informatics and Statistics,

W. Churchill Sq. 4, 130 67 Prague 3, Czechia

petra.dotlacilova@vse.cz