History of statistics

The history of statistics starts with something called "Political Arithmetic" in the 17th century, when the course was set which led to our modern science and technology.

Political arithmetic

Political Arithmetic was launched in the second half of the 17th century, at about the same time as the theory of game of chance. The expression *Political Arithmetic* was introduced by William Petty (1623-1687), who was a physician, a musician, a physicist, a mathematician, a solicitor, and one of the first econometricians. His intellectual legacy is his masterpiece "Political Arithmetic" which appeared in 1690, three years after he had passed away. It starts with the following words:

> *Political Arithmetic or A Discourse concerning the extent and value of lands, people, buildings: husbandry, manufacture, commerce, fishery, artizans, seamen, soldiers; publick revenues, interest, taxes, superlucration, registries, banks, valuation of men, increasing of seamen, of militia's, harbours, situation, shipping, power at sea, etc. As the same relates to every country in general, but more particularly to the territories of His Majesty of Great Britain, and his neighbours of Holland, Zealand, and France.*

Petty's message is to document all natural, social and political components of a state by numbers in order to have a sound basis for decisions and, thus, avoid political controversies. It is rather likely that Petty was much influenced by his friend John Graunt, who had published in 1662 his famous book *Natural and Political Observations on the London Bills of Mortality*, which earned him the reputation of being the founder of demography.

Graunt calculates relative frequencies, i.e. averages, and detects certain stabilities. He notes (see Hauser):

> *We shall observe, that among the several Casualties some bear a constant proportion unto the whole number of* Burials*; such are* Chronicle *Diseases, and the Diseases whereunto the City is most subject; as for Example,* Consumptions, Dropsies *(...): nay, some Accidents, as* Grief, Drowning *(...), whereas* Epidemical *and* Malignant *Diseases, as the* Plaque, Purples *(...) do not keep that equality: so as in some Year, or Months, there died ten times as many as in others.*

Graunt's main idea was averaging and this remained the main idea of statistics until our days. In fact, this idea is not only central for statistics, but also for physics and even for all quantified branches of science. But, as Pearson (see Pearson) notes, Graunt is neither aware of the relationship between 'sample size' and 'stability' nor does he have the concept of 'probability' at his disposal. Nevertheless, Graunt determined the general approach to be taken when dealing with randomly generated data. The approach was based on 'averages' aiming at detecting 'stable laws', also frequently called 'statistical laws'.

Statistics

The ideas of Graunt and Petty were taken up by the mathematicians Edmund Halley and Abraham de Moivre in England and (see Pearson) spread to continental Europe. Poisson and Laplace combined them with the fundamental ideas of the calculus of probability developing the foundation of a mathematical science. However, Graunt's main idea to search for 'stable laws' by means of averaging was maintained except that the methods used became much more mathematically sophisticated. Besides simple arithmetic means, limit theorems and laws of large numbers were used for detecting 'stable laws' and removing variability.

This development from a discipline investigating real problems to a discipline solving mathematical problems did not at all pass unchallenged. John Venn judges the development in his seminal treatise "The Logic of Chance" as follows:

> *Probability has been very much abandoned to mathematicians, who as mathematicians have generally been unwilling to treat it thoroughly. They have worked out its results, it is true, with wonderful acuteness, and the greatest ingenuity has been shown in solving various problems that arose, and deducing subordinate rules. And this was all that they could in fairness be expected to do. Any subject which has been discussed by such men as Laplace and Poisson, and on which they have exhausted all their power of analysis, could not fail to be profoundly treated, so far as it fell within their province. But from this province the real principles of science have generally been excluded, or so meagrely discussed that they had better have been omitted altogether.*

During the 18th century, empirical Political Arithmetic turns to a highly mathematical calculus based on the concept of probability. However, just as Venn described it, the real-world meaning of 'probability', which constitutes not a mathematical problem, remained obscure and became ambiguous. Consequently, the interpretation of obtained results became difficult and applicability doubtful. Anyhow, Political Arithmetic was somehow clouded by the calculus of probability and something had to be done to regain the decreased impetus.

In this critical situation an extra-ordinary 'marketing' event happened, which is described (see Pearson) as follows:

> *A Scotsman steals the words 'Statistics' and 'Statistik' and applies them to the data and methods of 'Political Arithmetic'. It was certainly a bold, bare-faced act of robbery which Sir John Sinclair committed in 1798. It would be exactly paralleled if somebody stole our word biometry and applied it in a totally different sense to that of its creators. Well, we have to bless Sinclair, ...*

Sinclair himself describes his 'bold deed' in the following way (see Pearson):

> *Many people were at first surprised at my using the new word, 'Statistics' and 'Statistical', as it was supposed that some term in our own language, might have expressed the same meaning. But in the course of a very extensive tour through the northern parts of Europe, which I happened to take in 1786, I found that in Germany they were engaged in a species of political inquiry to which they had given the name of 'Statistics', and though I apply a different meaning to that word, for by 'Statistical' is meant in Germany, an inquiry for the purpose of ascertaining the political strength of a country or questions respecting matters of state; whereas, the idea, I annex to them is an inquiry into the states of a country, for the purpose of ascertaining the quantum of happiness enjoyed by its inhabitants, and the means of its future improvement, but as I thought that a new word, might attract more public attention, I resolved on adopting it, and I hope that it is now completely naturalised and incorporated with our language.*

The change of name at the end of the 18th century restored the old enthusiasm and started an amazing development during the 19th century.

Homme Moyen

The tremendous success of physics during the 17th and 18th century had raised the belief that the secrets of nature could completely be disclosed. Everything seemed to be reducible to 'laws of nature' at least for the inanimate nature. The observed stability of averages of many demographic phenomena seemed to indicate that there were also 'laws of nature' ruling the animate nature. Consequently, they were not any more ascribed to divine providence.

The most prominent representative of the new discipline statistics in the 19th century was the mathematician, physicist and astronomer Adolphe Quetelet (1796-1874) with his masterpiece 'Sur l'homme et le développement de ses facultés ou Essai de physique sociale'. Quetelet applies the new science to social phenomena, defines the 'average man' and thus inaugurates sociology, which he called 'social physics'.

Graunt's idea of averaging reached a new climax in Quetelet's 'social physics' centering around the 'average man'. Any deviation from the average human was defined as random error or perturbation and explained by means of the 'error law', i.e. the normal probability measure. However, Quetelet used the normal probability law not only for describing the distribution of 'errors', but in a more general context and, thereby, paved the way for the dominating role of the normal distribution in statistics.

Quetelet has undoubtedly contributed a great deal to the success of statistics. In 1829, his collection of empirical social data led to the first national census in Belgium and Holland. As president of the Belgian Central Statistical Commission, Quetelet did much to inspire the creation of statistical bureaus all over Europe and worked hard to promote internationally uniform methods and terminology in data collection and presentation. Under his leadership, the first of a long series of International Statistical Congresses was held in Brussels in 1853.

With Quetelet and his 'social physics', statistics approached in its methodology natural sciences and, thus, achieved a better reputation. Stigler notes:

> *Quetelet made two important advances toward the statistical analysis of social data: the first of these was formulating the concept of the average man, the second the fitting of distributions.*

Contemporaries considered Quetelet as the ultimate authority of statistics, but Menges remarks:

> *Certainly his approach is modern; Quetelet had left Political Arithmetic behind and followed completely the intellectual trends of modern statistics. But his 'modernity' is excessive, is exaggerated to absurdism, sometimes grotesqueness. Quetelet thought to be the Newton of sociology, he believed to have found the formula to calculate society.*

Another interesting view on the general statistical approach is due to Francis Galton, who notes under the heading *The Charms of Statistics*:

> *It is difficult to understand why statisticians commonly limit their inquiries to averages, and do not revel in more comprehensive views. Their souls seem as dull to the charm of variety as that of the native of one of our flat English counties, whose retrospect of Switzerland was that, if its mountains could be thrown into its lakes, two nuisances would be got rid of at once.*

Thus again, the success story seemed to approach gradually its end. However, during the second half of the 19th century mathematics advanced and at the end of the century the problem of measuring sets came to the fore.

Mathematical statistics

At the turn to the 20th century, a mathematical result was obtained, which should become decisive for probability calculus and for statistics. In France, Henri Lebesgue derived the 'theory of measure', based on results obtained by Emile Borel, and measure theory made the development of modern probability theory as a genuine branch of mathematics possible.

In 1933, the Russian mathematician Andrey Kolmogorov (see Kolmogoroff) published a seminal work on probability theory about which Vitanyi notes:

> *Much important work on probability theory had already been done without benefit of foundations, but this little book "Foundations of the Calculus of Probabilities", published in German in 1933, immediately became the definitive formulation of the subject.*

and

> *According to Gnedenko: In the history of probability theory it is difficult to find other works that changed the established points of view and basic trends in research work in such a decisive way. In fact, this work could be considered as the beginning of a new stage in the development of the whole theory.*

Kolmogorov himself states:

> *The theory of probability, as a mathematical discipline, can and should be developed from axioms in exactly the same way as geometry and Algebra. This means that after we have defined the elements to be studied and their basic relations, and have stated*

> *the axioms by which these relations are to be governed, all further exposition must be based exclusively on these axioms, independent of the usual concrete meaning of these elements and their relations.*

With Kolmogorov's axiomatisation of probability theory, it had become an genuine branch of mathematics and, thus, did not deal with real phenomena anymore and not with randomness in particular.

It did not take long for a discipline "Mathematical Statistics" to emerge. Jerzy Neyman, one of the founders of mathematical statistics, said:

> *The origins of probability theory and of mathematical statistics are also empirical. Now, both have reached the stage of maturity and live their own lives. However, because of certain circumstances recounted below, we witness a seemingly endless sequence of feedbacks to empirical sciences and, simultaneously, a comparable sequence of stimuli from empirical sciences to mathematical theories of both probability and statistics.*

The interchange between mathematics and science was widely performed on the basis of data. Science provided data and statisticians took on the analysis by means of methods derived from mathematical statistics. Among these statisticians the catchword 'Let data speak!' became popular. Since the 1930s, this type of dialogue between the two branches of mathematics, 'probability theory' and 'mathematical statistics', and all branches of science has expanded to an unimagined extent. Nowadays, many branches of science, like psychology or sociology, but also biology and medicine and others, are completely penetrated by statistics and even resemble more branches of statistics than independent sciences.

Methods based on results obtained in mathematical statistics are used wherever decisions have to be made or situations involving uncertainty have to be analysed. Decisions based on statistics accompany human life continuously from the beginning to the end. A majority of theses in many branches of science deals with statistics or the application of statistics.

This tremendous success is surprising because there is no satisfactory real-world interpretation of the axiomatic elements mathematical probability theory or mathematical statistics are based on. The axiomatic elements have either opposing interpretations, which are often used simultaneously, or they have no counterpart in real world at all.

The question arises how statistics could acquire its present ubiquitous position. The answer is almost obvious. Any research in science has to cope with uncertainty caused by ignorance and randomness and aims at reducing ignorance and controlling randomness. Therefore, the demand for formal rules how to handle uncertainty and solve the related problems is immense.

Since the overwhelming success of classical astronomy and physics, mathematics has coined science. Mathematics, however, is logic applied to a set of axioms and logic has also played a major role in deriving and verifying the 'laws of nature'. Thus, logic was assumed to be not only the doctrine providing the rules for developing abstract mathematics, but also the rules for developing science.

Consequently, the concepts and methods derived in probability theory and mathematical statistics according to the rule of logic were uncritically accepted by scientists of all branches, not noticing that "the real principles of science have generally been excluded, or so meagrely discussed that they had better have been omitted altogether", as John Venn put it.

Probability theory and mathematical statistics, as special branches of mathematics, necessarily follow the rules of logic, and there is no need at all to extend the rules from the viewpoint of mathematics. When some more discerned statisticians felt the discrepancies between reality and statistical theories, statistics split into several statistical denominations, each of which believes in different principles and interpretations.

Summary

Statistics began in the second half of the 17th century as 'Political Arithmetic' aiming at collecting data in order to find stable laws as a rational foundation of making decisions. Soon afterwards, this kind of primitive data analysis was merged with probability, and arithmetic means were 'justified' by limit theorems, and sophisticated mathematics were applied for data analysis. During the 18th century, almost any mathematician of reputation was involved in the further development of mathematical probability theory, and political arithmetic became overshadowed.

At the end of the 18th century political arithmetic was revived by changing its name to 'statistics'. Based on the 'concept of error' and the 'normal law', statistics reached a first climax in the 19th century by the definition of the 'average man' and its subsequent application in the newly established social sciences.

New advancement in mathematics and the establishment of measure theory by Borel and Lebesgue at the turn to the 20th century enabled the axiomatisation of probability theory as a genuine branch of mathematics by Kolmogorov. Since then, probability theory and its application 'mathematical statistics' got rid of all so far existing relationships to reality and developed according to the intrinsic laws of logic to a remarkable mathematical theory.

Johann Adam Schall von Bell Bibliography:

• Galton, F. (1889): *Natural Inheritance.* MacMillan, London.

• Hauser, W. (1997). *Die Wurzeln der Wahrscheinlichkeitsrechnung.* Franz Steiner, Stuttgart.

• Kolmogoroff, A. (1933): *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Springer Berlin.

• Menges, G. (1982): *Die Statistik. Zwölf Stationen des statistischen Arbeitens.* Gabler, Wiesbaden.

• Neyman, J. (1976): The Emergence of Mathematical Statistics. A Historical Sketch with Particular Reference to the United States. In: *On the History of Statistics and Probability.* Ed. D.B. Owen, Marcel Dekker, New York.

• Pearson, K. (1978): *The History of Statistics in the 17th and 18th Centuries against the changing background of intellectual, scientific and religious thought.* Lectures by Karl Pearson given at the University College London during the academic sessions 1921-1933. Edited by E.S. Pearson. Charles Griffin, London.

• Stigler, S.M. (1986). *The History of Statistics. The Measurement of Uncertainty before 1900.* Havard University Press, Cambridge.

• Vitanyi, M.B. (1988): Andrei Nikolaevich Kolmogorov. *CWI Quarterly* 1, 3-18.

• http://www.swlearning.com/quant/kohler/stat/biographical_sketches/bio.2.html