GOMER PRESS, LLANDYSUL



# THE DEVELOPMENT OF STATISTICAL IDEAS and their APPLICATIONS

*Inaugural Lecture of the Professor of Statistics, delivered at the University College of Swansea on 28 January, 1975*

*by*

PROFESSOR ALAN G. HAWKES
B.SC., PH.D.

GWEDDW CREFFT HEB EI DAWN

UNIVERSITY COLLEGE OF SWANSEA

UNIVERSITY COLLEGE OF SWANSEA

# THE DEVELOPMENT OF STATISTICAL IDEAS and their APPLICATIONS

*Inaugural Lecture of the Professor of Statistics, delivered at the University College of Swansea on 28 January,* 1975

*by*

PROFESSOR ALAN G. HAWKES
B.SC., PH.D.

*Inaugural Lecture of the Professor of Statistics delivered at the University College of Swansea on 28 January, 1975*

*by*

PROFESSOR ALAN G. HAWKES
B.SC., PH.D.



# THE DEVELOPMENT OF STATISTICAL IDEAS AND THEIR APPLICATIONS

Statistics has a very long history, and yet it can fairly be said that the subject as we know it today only began to be systematically developed towards the end of the nineteenth century. There had been isolated examples of "modern" statistical ideas before then, but it had not begun to be thought of as a unified theory (or set of theories) which could be applied to an enormous variety of problems. During this century the theory and application of statistics has developed rapidly. Tonight I want to say something of the origins and historical development of the subject, the forces which have moulded it, and to touch upon some of the ideas which give the subject its essential flavour. I can, of course, give only the briefest sketch of so complex a subject, and many important aspects will be left untouched. I aim, therefore, not at a definitive and comprehensive study, but a biased sample of material which I hope will inform and entertain you.

It is only in recent years that any extensive attempt to study the history of statistics has taken place, indeed the first reasonably comprehensive book on the subject was published as recently as 1970, edited by Egon Pearson and Maurice Kendall.[1] I commend this book to any serious student of statistics for the picture it gives of great men stumbling and groping in the twilight to formulate ideas which we now take for granted. They may find it encouraging to see how often these pioneers made the kind of blunders we find inexcusable if committed by our undergraduates. For it is one of the delights and difficulties of statistics that, in addition to a wealth of precise mathematical axioms, theorems and lemmas, there are a great many ill-formed, half-understood and illogical principles over which there has been (and I think always will be) a great deal of heated debate. The non-specialist statistician would derive some entertainment also from the extensive details of some of the very colourful characters who have a prominent role in the subject, in contrast to the stereotype statistician who is popularly thought of as a rather dull fellow who simply collects large quantities of facts.

It is true that collection of data is an important part of statistics, but far more important are *ideas* of statistical theory. How can we build statistical models which *explain* why the data are what they are ? How can we use the data most efficiently to make scientific inferences or to make decisions ? Perhaps one of the main points I would like to make tonight is best illustrated in a passage from a little known work by K. A. C. Manderville called "The undoing of Lamia Gurdleneck".

*"You haven't told me yet", said Lady Nuttal, "what your fiancé does for a living".*

*"He's a statistician", replied Lamia, with an annoying sense of being on the defensive.*

*Lady Nuttal was obviously taken aback. It had not occurred to her that statisticians entered into normal social relationships. The species, she would have surmised, was perpetuated in some collateral manner, like mules.*

*"But Aunt Sara, it's a very interesting profession", said Lamia warmly.*

*"I don't doubt it", said her aunt, who obviously doubted it very much. "To express anything important in mere figures is so plainly impossible that there must be endless scope for well-paid advice on how to do it. But don't you think that life with a statistician would be rather, shall we say, humdrum" ?*

*Lamia was silent. She felt reluctant to discuss the surprising depth of emotional possibility which she had discovered below Edward's numerical veneer.*

*"It's not the figures themselves", she said finally, "it's what you do with them that matters".*

What then is statistics ? I think it is useful to begin with a decomposition into four broad components.

(i) The collection, tabulation and graphical presentation of data and calculations based on them in such a way as to facilitate the understanding of the objects under study.



(ii) The mathematical theory of probability.
(iii) The philosophy of scientific inference.
(iv) Decision making.

Each of the components has its own history, in some cases a very long one, but it was only when these components became fused together that the modern theory of statistics began to evolve. Today a compact definition of statistics might read

*That part of scientific method which is concerned with drawing inferences or making decisions in situations involving variability or uncertainty.*

A hundred years ago that definition would not have been understood.

I will begin with a brief discussion of the history of the separate components, and follow with some comments on the development of the modern theory.

### *Data collection presentation and political arithmetic*

The collection of data about states, or state-istics, has been practised for a long time, usually to enable governments to tax their subjects more effectively. One may recall famous examples such as the numbering of the people of Israel, Augustus's balance sheets of the Roman Empire or the Doomsday book. The first recorded use of the word statistics as far as I know was by an Italian historian, Girolamo Chilini who refers in 1589 to "*statistica*". Indeed, the so-called "political arithmetic" necessitated by the importance of trade was largely begun in fourteenth and fifteenth century Italy, but then stagnated until about the middle of the seventeenth century.

Progress was then rapid. For the first time one can discern a modern approach. Men began to reason about their data, to seek explanations and make predictions rather than using data merely as a description or record of the existing state of affairs. The subject matter was largely that of political economy, vital statistics, demography and actuarial science. People like John Graunt, William Petty, Ludwig Huyghens and Edmund Halley (the comet man) were concerned to calculate life-tables and fertility rates. These were for purely medical reasons, or to calculate annuities or to make economic judgements. For example, they were concerned to discover if one place was more healthy to live in than another, or to estimate how much it would be worth spending on sanitation in order to combat the plague. Much of the material they used were the bills of mortality regularly compiled in the various parishes of London and elsewhere.

This kind of study of vital statistics continued throughout the eighteenth century. Mathematicians such as Daniel Bernoulli and D'Alembert formulated the numerical calculations into mathematical formulae involving the calculus, although that did not really help much. In 1801 the first official population census, which was to take place every ten years, was made. With much detailed and accurate data available, the methods of vital statistics became more refined, notably in the hands of William Farr, but still largely non-probabilistic.

At about this time graphical methods of presenting statistics were introduced, the most important originator being a colourful character called William Playfair, who was born near Dundee in 1759. He was involved in a number of dubious enterprises, including helping to capture the Bastille, but he managed to write some quite respectable books on economic matters which included a considerable amount and variety of graphical devices. He drew histograms, bar-charts, pie charts and plotted time series in a beautifully executed manner which presumably owed much to his early training as an engineering draughtsman. Figure 1 shows an example of a combination of a graph (for weekly wages) and a bar-chart for the price of wheat.

Collecting and tabulating large amounts of data can be a tedious business. Even today, some people still have the idea that this is all statistics consists of (and that statisticians are very dull people). Playfair, writing in 1801, gives some idea of what it was like.[2]

. . . *for no study is less alluring or more dry and tedious than*



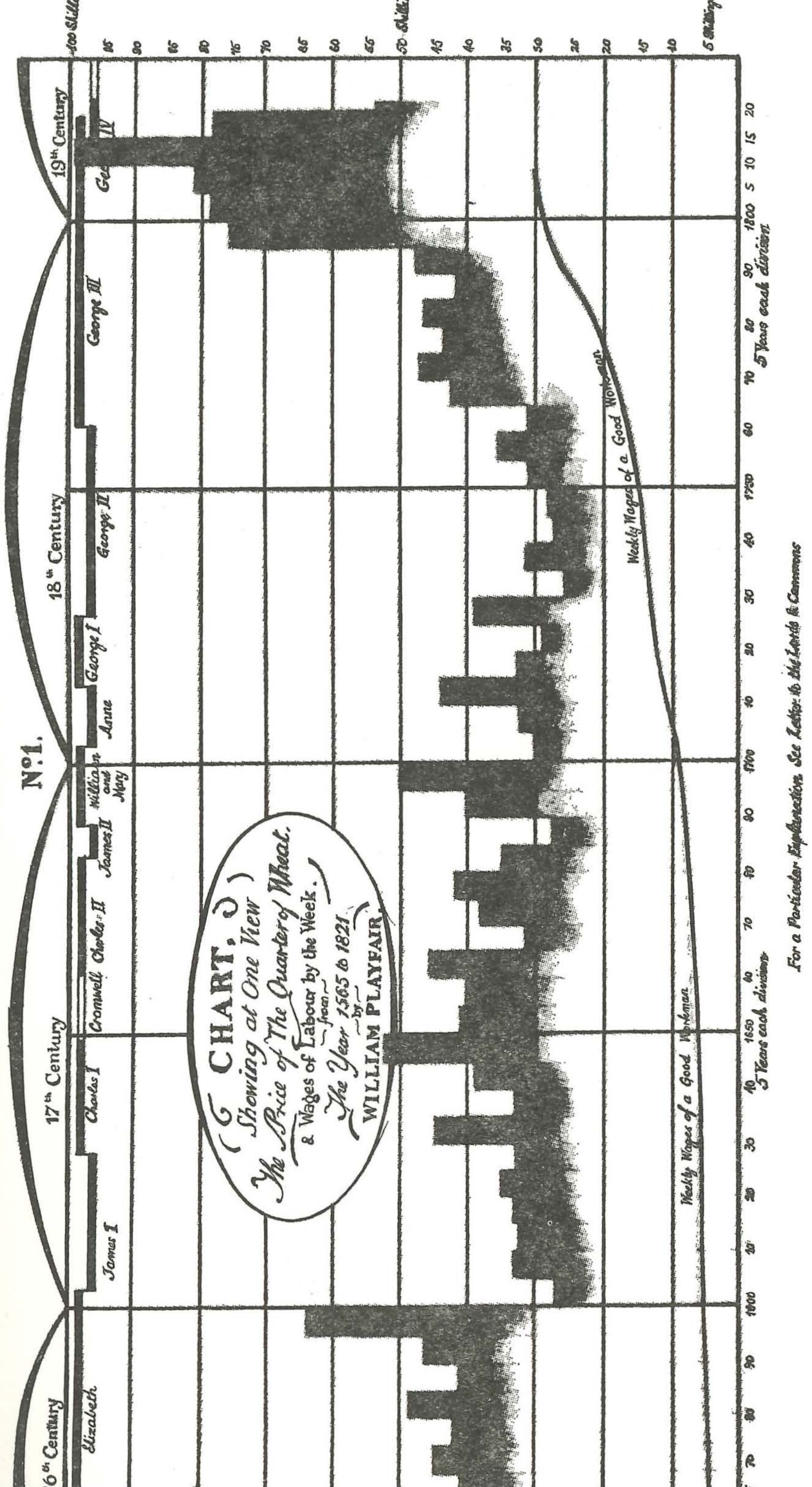



Fig. 1 Extract from Playfair's *A Letter on our Agricultural Distress*, 1821.

*statistics, unless the mind and imagination are set to work or that the person studying is particularly interested in the subject ; which is seldom the case with young men in any rank in life.*

In the same volume he gives us some idea of the growth in interest in statistics, from the point of view of data collection.

*Statistical knowledge, though in some degree searched after in the most early ages of the world, has not till within these last* 50 *years become a regular object of study.*

The industrial revolution brought a greater need for facts. In 1832 the Board of Trade set up a Statistical Office. In 1834 the Statistical Society of London (later to become the Royal Statistical Society) was founded. in 1838 the Society first published its Journal ; it is interesting to note that in the introduction to the first volume it says

*The science of statistics seeks only to collect, arrange and compare facts . . . It does not discuss causes, nor reason upon probable effect .*

Half a century later the position had started to change rapidly. People were asking quite sophisticated questions which they hoped a study of social statistics might answer. Florence Nightingale was a very strong advocate of the setting up of a Professorship of Statictics at Oxford in the cause of trying to answer some of the many questions she felt needed to be answered.

It took about seventy years for a Professor of Statistics to be appointed at Oxford (he is not called that), but she would no doubt have been glad to know that in 1963 one of her great-nieces, Florence Nightingale David, became the first woman Professor of Statistics in this country. To take but one example of the problems that interested her from a letter to Francis Galton.[3]

*What effect has education on crime ? Some people answer unhesitatingly—As education increases crime decreases. Others as unhesitatingly—Education only teaches to escape conviction, or to steal better when released. Others again—Education has nothing to do with it either way . . .*

Towards the end of the nineteenth century the biologists



also took part in the data explosion, and this was to prove critical in the history of statistics. Gregor Mendel had already diligently grown, classified and counted thousands of peas. Now two men in particular, W. F. R. Weldon (appointed to the Chair of Zoology at University College London in 1891) and Francis Galton, who was already at University College, developed a prodigious capacity for counting and measuring everything biological they could lay their hands on. They did much more than just collect data, we shall hear more of that later on. For the moment one little anecdote will suffice to show how eager they were to measure everything, and what great originality they had.

Galton when in Africa was particularly struck by substantial buttocks of the Hottentot women. On seeing one Venus among Hottentots he was exceedingly anxious to obtain measurements of her shape. He did not know one word of Hottentot and did not feel himself able to[3]

*explain to the lady what the object of my footrule could be ; and I dared not ask my worthy missionary host to interpret for me. The object of my admiration stood under a tree, and was turning herself about to all points of the compass, as ladies who wish to be admired usually do. Of a sudden my eye fell upon my sextant ; the bright thought struck me, and took a series of observations upon her figure in every direction, up and down, crossways, diagonally, and so forth, and I registered them carefully upon an outline drawing for fear of any mistake, this being done, I boldly pulled out my measuring tape, and measured the distance from where I was to the place where she stood, and having thus obtained both base and angles, I worked out the result by trigonometry and logarithms.*

However, to return to social and economic statistics, the number of facts which governments and industry appear to want to know has continued to grow rapidly in this century, most especially since 1945. In 1972 the President of the Royal Statistical Society, J. Harold Wilson, in his presidential address[4] described the information on which government was based in the early post war years as like

*looking up train times in last year's Bradshaw,*
a quotation from Harold Macmillan. He went on to describe the growth of the Government Statistical Service into the efficient organisation it now is. We now have a vast array of official statistics, indices of costs, wages, production, social trends, housing surveys etc. etc. Much of the data and the calculations and judgements made on them are in the tradition of the old political arithmetic, involving no probability statements. Statistical controversy about origin, construction and interpretation of such statistics is now commonplace in politics ; Dennis Healey's inflation rate statements and Sir Keith Joseph's demographic calculations spring to mind as notable examples in the past year.

All this shows a certain amount of distrust of statistics. I think this is a good thing if it means that people are going to look carefully and critically at the sources of statistics and the arguments based on them, but a bad thing if it means a rejection of statistical arguments altogether. I would like to end this section with a quotation which I found in the Times Higher Education Supplement towards the end of 1973, just before I was interviewed for the Chair of Statistics in this College. It was a report of the annual address to the Court of Governors by the Principal, who was at that time of course Professor Llewellyn-Jones. He was reported as complaining about the University Grants Committee's method of assessing university needs using the cost per full-time-equivalent student as

*the results of esoteric calculations emerging from the appropriate Ministry, no doubt based on statistics.*

*Probability*

The history of the mathematical theory of probability is quite well documented. The classic book on the subject was Isaac Todhunter's[5] scholarly, though rather dull, account published in 1865. A more entertaining account by Florence David appeared in 1962.[6]



Probability theory can be considered as a purely abstract mathematical exercise of proving some very difficult and powerful theorems, starting from some relatively simple axioms, and with very little reference to problems of the real world. Although some very eminent mathematicians appear to think in this way, nevertheless it is quite clear that the overwhelming stimulus to the development of the theory has always been the need to solve practical problems. It is also true, that often we later find practical applications of the abstract theory which the pure mathematicians have produced.

It is well known that the main stimulus to the beginning of a theory of probability was gambling, and the desire to calculate probabilities of complicated events. For a very long time people had an intuitive notion that in simple symmetrical situations, such as drawing lots or throwing a die, the outcomes had the same chance of happening (if the system were fair). The first known record of a more complicated calculation was in the sixteenth century when the Italian Gerolamo Cardano, the illegitimate son of a geometer, calculated certain probabilities concerned with throwing two or three dice. He did a similar calculation with astragali, but got the wrong answer because the basic events are not equally probable. The astragalus is a small bone in the ankle. Hooved animals, such as a sheep, provide the best examples ; the bones have four faces on which they can land with probabilities approximately 1/10, 1/10, 4/10 and 4/10. They have been used for gambling in ancient Egypt, Greece and Rome, and probably long before that.

Later that century Galileo answers the same question for a gambler who had experimentally observed that in throwing three dice the total ten occurs more frequently than the total nine. It says a lot about the degree to which these gamblers played when you calculate the difference of these two probabilities to be 1/108. In the seventeenth century mathematicians, such as Fermat and Pascal, with growing confidence calculated probabilities in much more complicated gambling situations. One of these was the celebrated problem of the Chevalier de Méré who gambled so often that he was able to discover experimentally that if he bet on getting at least one six in four throws of a die he won more often than he lost, whereas if he gambled on getting at least one double six in twenty-four throws of two dice then he lost more often than he won. The mathematicians were able to show that these two events had probabilities 0.518 and 0.491 respectively. Out of these solutions to individual problems there began to emerge a general theory of probability, which was first systematized and published in 1657 by Christian Huygens.[7]

This theory was not difficult, although some of the particular problems were, and it is intriguing to ask why it has not been developed before. Astragali and dice are known to have been used for games of chance in classical times ; the earliest known die of pottery comes from Iraq and is dated at the beginning of the third millenium B.C. The Greeks, with their mastery of solid geometry, made several polyhedral dice, in addition to the usual cube.

The Romans were passionate gamblers. Suetonius[8] tells us that Augustus and Claudius were addicts, the latter to such an extent that he used to play while driving, throwing on to a board fitted into his carriage. I don't recommend anyone to follow his example. The situation got so bad that at one time the Romans forbade gaming at certain times of the year. They were not above cheating either ; there exist examples of deliberately loaded Roman dice.

There is abundant evidence that games of chance continued to abound throughout the period up to the time of Cardano and that Church and State were often concerned about the idleness, crime, drinking, swearing and generally bad behaviour which often accompanied them.

Participants in the third crusade (A.D. 1190) had careful instructions of the extent to which they might gamble ; no person below the rank of knight was allowed



to play for money ; knights and the clergy were not allowed to lose more than twenty shillings in twenty-four hours. Louis IX banned dicing, or even the manufacture of dice, in 1255. Henry VIII also prohibited dice and cards, although he set a very bad example himself.

Why then, with all this gambling activity, did we have to wait so long for the appearance of an elementary theory of probability ? The Greeks, for example were certainly good enough at mathematics one would have thought. One major factor may be that astragali, dice and other random devices also had important religious uses. The Greeks and Romans seemed to have some idea of chance, but also believed that the gods had some influence over affairs. It was common when contemplating some major enterprise, such as getting married or starting a war, to go to the temple and cast dice or astragali. The gods, by exercising influence over the fall of the dice, would indicate whether the omens were good or not ; the interpretation being made by the priest by reference to the oracle tablets. The Jews also used the drawing of lots to decide some matters, and they regarded the outcome as an expression of God's will, especially if the drawing was commanded by God himself, as happened in the division of the land of Israel. (Numbers 26, 55-56).

*Notwithstanding the land shall be divided by lot : According to the names of the tribes of their fathers they shall inherit.*

Another example occurs in the book of Joshua (Chapter 7) in which someone had been looting from the Babylonians. Under God's instruction Joshua went to find out who the thief was

*So Joshua rose up early in the morning, and brought Israel by their tribes ; and the tribe of Judah was taken : And he brought the family of Judah ; and he took the family of Zarhites man by man ; and Zabdi was taken : And he brought his household man by man ; and Achan, the son of Carmi, the son of Zabdi, the son of Zerah, of the tribe of Judah, was taken.*

And so it was discovered that Achan was the thief. It is not clear from this what is meant by "taken", but from the Talmud we find that the Lord said to Joshua

"Go and cast lots". Thereupon he went and cast lots, and the lot fell to Achan. Putting these two together it looks as if what happened was a four-stage random sample. Since I doubt whether they were clever enough to sample their units at each stage with "probabilities proportional to size" as we would do today, this may not have been a completely fair way to do it (people in small families within small tribes having an unfairly high chance of selection).

Divination is still practised today in parts of the world. The Tibetan Buddhists, for example, are said to use a number of random devices. One such obtained by L. A. Waddell in 1893 consisted of a sacred board with squares and a die which is thrown onto it, the result of the throw being used to predict a future state of reincarnation He remarks[9]

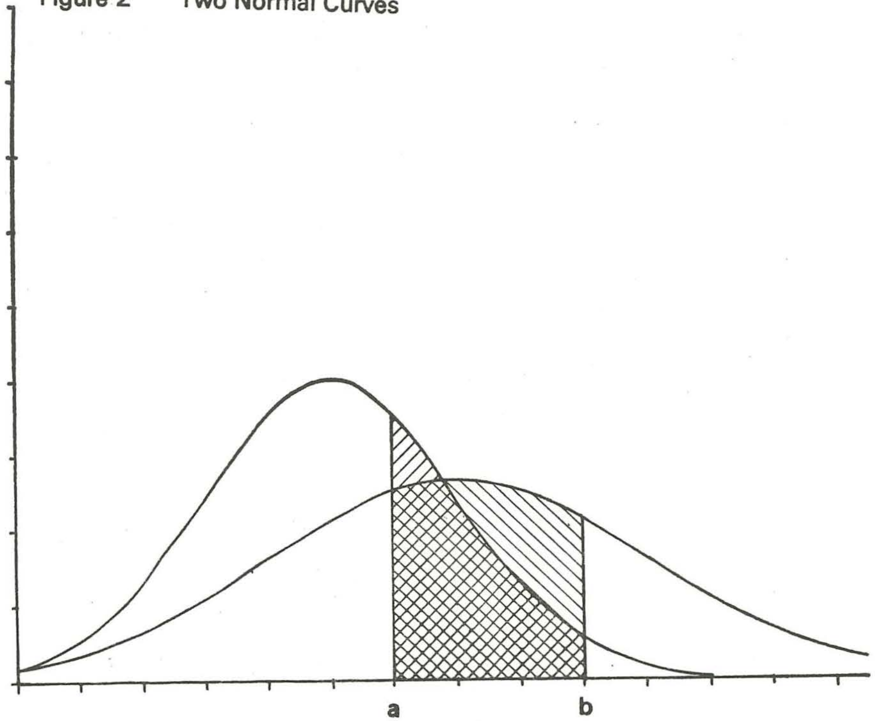
*The dice accompanying my board seems to have been loaded so as to show up the letter Y, which gives a ghostly existence, and thus necessitates the performance of many expensive rites to counteract so undesirable a fate.*

With such uses of random devices it is perhaps not surprising that people did not try too hard to understand or explain such phenomena in a mathematical way. The situation was no more favourable in early Christian times when every outcome, no matter how small, tended to be thought of as a pre-ordained expression of God's will. If anything appeared to be due to chance that was an expression of man's ignorance, not the nature of things.

Well now, let us return to our chronological narrative. Probability continued to develop throughout the eighteenth century, with most of the eminent mathematicians of the day making some contribution, notably James Bernoulli (and most of the other members of the Bernoulli family), de Moivre, Laplace and Gauss. During this time they moved from the classical probability situation of a finite number of equally likely outcomes, as in games of chance, to more flexible systems. One of the main motivations for this was in accounting for the fact that repeated observations of astronomical quantities were not



Figure 2 Two Normal Curves




all exactly the same, and for this reason they developed continuous probability distributions of errors ; the prime example being the famous Normal curve which was discovered several times throughout this period. Of course there is in fact a whole family of Normal curves, of which we see two examples in Figure 2. The interpretation of such a curve is that if we take a large number of observations then the proportion of the observations falling in the interval (a, b) is given by the area under the curve between the values a and b or, for the more theoretically minded, the probability of an observation falling in (a, b) is given by this area. The two curves shown in the figure differ in their average values and in that in the first case the observations cluster more closely about their average than they do in the second case. In fact a Normal curve is completely specified by just two parameters, the mean (or average) and the standard deviation which measures this clustering effect.

To the astronomers these random errors were just a nuisance to be removed in trying to estimate some physical quantity. In the first half of the nineteenth century the Belgian Quetelet realised that Normal curves gave a very good representation of the veriability of observations of all kinds of biological and social measurements, and in this context it is the variability itself which is of interest ; it is not simply a nuisance anymore. Weldon, Galton and Pearson at the end of the century were likewise fascinated by this ; they collected masses of biological data and found, time and again, that the Normal curve fitted remarkably well. We can get some idea of Galton's enthusiasm from two extracts from his great work "Natural Inheritance" published in 1889, at the age of 67.

*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 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. An Average is but a solitary fact, whereas if a single other fact be added to it, an entire Normal Scheme, which nearly corresponds to the observed one, starts potentially into existence.*

*Some people hate the very name of statistics, but I find them full of beauty and interest. Whenever they are not brutalized, but delicately handled by lhe higher methods, and are warily interpreted, their power of dealing with complicated phenomena is extraordinary. They are the only tools by which an opening can be cut through the formidable thicket of difficulties that bars the path of those who pursue the Science of man.*

A further extract

*I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the "Law of Frequency of Error". The law would have been personified by the Greeks and deified, if they had known of it. It reigns with*



*serenity and in complete self-effacement amidst the wildest confusion. The huger the mob, and the greater the apparent anarchy the more perfect is its sway. It is the supreme law of Unreason. Whenever a large sample of chaotic elements are taken in hand and marshalled in the order of their magnitude, an unsuspected and most beautiful form of regularity proves to have been latent all along.*

Soon, however, they began to find data which seemed to have different shaped distributions. Weldon measured the relative frontal breadth of Naples crabs and seemed to get a distribution with two humps instead of one. He wrote to Karl Pearson, who was Professor of Applied Mathematics at University College London, and this problem led to Pearson's first statistical paper which he presented to the Royal Society in 1893.[10] At about the same time Edgeworth, the economist, showed Pearson some skew price curves and Pearson set about constructing whole new families of probability distributions including many of those most commonly used in statistics today.

And so, in biology and social science, probability models began to become a basic part of the scientist's attitude to natural phenomena, it was no longer just an error of observation. At about this time the same became true in physics, for example in statistical mechanics, as physicists began to consider microscopic problems involving very large aggregates of particles.

I will not attempt to discuss the enormous progress in probability theory in this century, David Williams has already discussed that in his inaugural lecture. As a statistician my interest is more in application than the mathematical theory itself. What I would like to do is to say something of the progress, equally enormous, in the extent to which probability has entered into the models which scientists make of the world about them.

To quote the physicist Schrödinger, writing in nature (153) in 1944

*In the course of the last sixty or eighty years, statistical methods and the calculus of probability have entered one branch of science after another. Independently, to all appearance, they* *acquired more or less rapidly a central position in biology, physics, chemistry, meteorology, astronomy, let alone such political sciences as national economy, etc. At first, that may have seemed incidental : a new theoretical device had become available, and was used whenever it could be helpful, just as the microscope, the electric current, X-rays or integral equations. But in the case of statistics, it was more than this kind of coincidence.*

*On its first appearance the new weapon was mostly accompanied by an excuse : it was only to remedy our shortcomings, our ignorance of details or our inability to cope with vast observational material. In the study of heredity we might prefer to be able to record the individual processes of meiosis, and thus to know how the hereditary treasure of a particular individual is composed from those of its grandparents. In textbooks on gas-theory it has become a stock phrase, that statistical methods are imposed on us by our ignorance of the initial co-ordinates and velocities of the single atom—and by the unsurmountable intricacy of integrating* $10^{23}$ *simultaneous differential equations, even if we knew the initial values.*

*But inadvertently, as it were, the attitude changes. It dawns upon us that the individual case is entirely devoid of interest, whether detailed information about it is obtainable or not, whether the mathematical problem it sets can be coped with or not. We realise that even if it could be done we should have to follow up thousands of individual cases and could eventually make no better use of them than compound them into one statistical enunciation. The working of the statistical mechanism itself is what we are really interested in.*

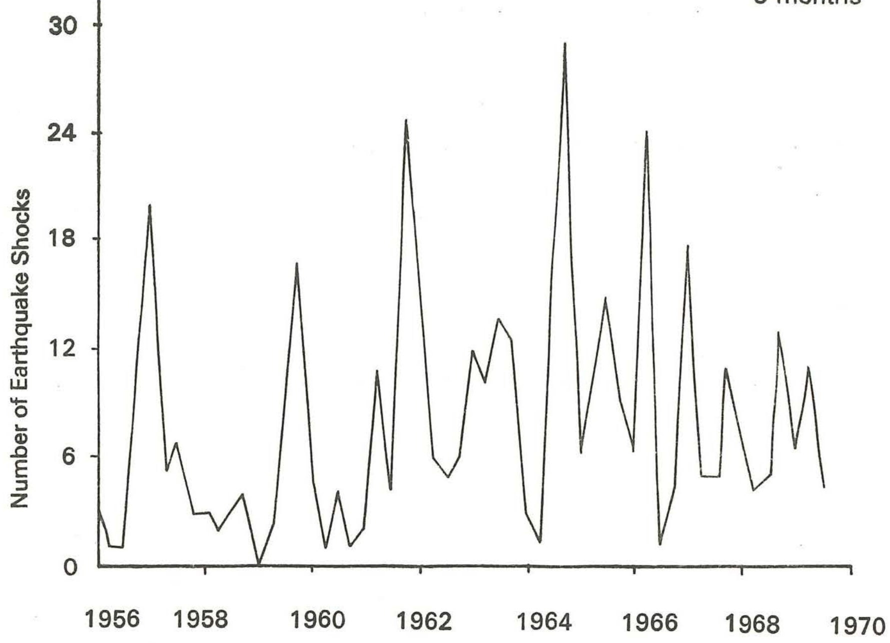
A much shorter quotation from Jacob Marschak[11]

*Briefly, empirical social science consists of statements about probability distributions.*

One of the features of this growth in the application of probability models, particularly since 1940, has been the study of dynamic rather than static situations, the so-called stochastic processes. The scope for the application of stochastic process models is immense and so, as a biased sample of material I shall mention briefly three examples on which I have done some research myself. My doctoral



**Figure 3** North Atlantic : Number of Earthquake Shocks in Periods of 3 months



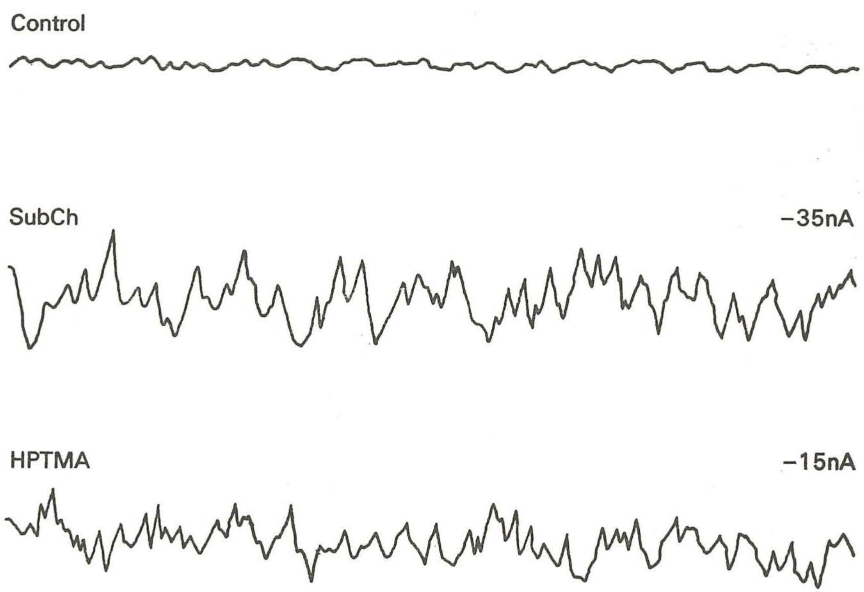

thesis was concerned with queueing models, most of which were applied to the behaviour of vehicles at various kinds of traffic intersection.

More recently I was involved in an attempt to fit probability models to describe the fluctuating intensity with which earthquakes occur. A typical example is given in Figure 3 which shows the rate of occurrence of earthquakes plotted over a period of fifteen years in the North Atlantic region.

Finally, one of my current interests is in collaboration with David Colquhoun, a pharmacologist. We are trying to understand how nerve stimuli are passed from nerve endings into a muscle. This is essentially a molecular process involving acetylcholine in the body, but we can study the effect of various drugs as well. You can see that this is a very random process from Figure 4, which

Figure 4 Samples of current through frog end-plate clamped at −73mV

Scale 1nA




shows the fluctuation with time of the potential difference between the inside and outside of a nerve membrane.

*The Philosophy of Scientific Inference*

I shall make this section brief, because I know very little about it. However, it can be no accident that Karl Pearson, was a considerable scholar in that area. In 1891, before his contributions to statistics began, he published "The Grammar of Science". There was a lucid exposition of scientific philosophy of the Machian school and Lenin, who strongly disagreed with much of what Pearson wrote, nevertheless had considerable respect for him and clearly considered him superior to Mach himself.

The components of general scientific methodology may be classified as



(i) formulation of an hypothesis, or model

(ii) experimentation and collection of relevant data

(iii) interpretation of the data so collected as evidence for or against the hypothesis

(iv) estimation of parameters in the model.

The peculiar nature of statistical inference is that it deals explicitly with models which include a probabilistic component. It follows that when we come to interpret our data the conclusions which we draw from them must themselves be, in some sense, random. It is the question of how to describe this randomness of our conclusions that lies at the heart of the modern theories of statistical inference.

Before leaving the subject of experimentation we might note that Todhunter, whose history of probability we mentioned earlier, was said to be opposed to the establishment of the first laboratories at Cambridge. He thought it unnecessary for students to see experiments performed, since the results could be vouched for by their teachers, all of them men of the highest character, and many of them clergymen of the Church of England.

*Decision making*

Men have made decisions ever since there have been men to make them. Most decisions of consequence involve uncertainty and statisticians are interested in situations where this uncertainty can be specified in terms of probabilities, or where a random mechanism may be part of the decision making process itself.

I have already mentioned the random devices by which the oracle would be used to give advice. The Jews also used random devices for settling disputes between conflicting parties. We read in Proverbs 18. 18

*The lot causeth contentions to cease, and parteth between the mighty.*

As an example of this I have already mentioned the division of Israel. In the Talmud we find that this was done by drawing out the names of the twelve tribes from one urn, and the descriptions of pieces of land from another urn.

In 1737 John Wesley drew lots to decide whether to marry or not.[12]

The object of the early probability calculations for games of chance was obviously intended to help make betting decisions, and moreover there were certain losses or gains involved. The same elements of costs, probability and decision exist in statistical quality control which began to be used in the 1920's, notably by Shewhart in America. Since 1940 a rigorous theory of decision making has emerged which involves

(i) a set of possible "states of nature"

(ii) assessment of the decision maker's "degree of belief" about which of these states is actually true, as measured by a "subjective probability"

(iii) a set of possible decisions, which will normally include making some experiment to help the decision maker find out more about the true state of nature

(iv) a measure of the benefits (or costs) which would arise from any particular combination of experiment decision and state of nature.

To illustrate these notions consider a simplified model for deciding whether to drill for oil in a particular section of the Celtic Sea. There are two possible "states of nature", oil is present or it is not. The prospecting company must evaluate its beliefs about the relative chances of these two states by assigning prior probabilities. The actions to be taken could be to drill or not to drill and before deciding they may take some kind of seismic survey. This data is used to modify the prior beliefs into posterior probabilities (by Bayes' Theorem) which are then used, in conjunction with an assessment of the costs and rewards, to make the decision about drilling.

This theory incorporates two old ideas. First the measurement of prior beliefs about nature by probabilities and the manner in which they may be modified by evidence supplied from experimental data. This was



first put forward in a celebrated paper by Thomas Bayes which was published in 1763, two years after his death.[13] The second idea is that of utility which had been current among economists for some time.

It is no accident that these ideas enter into the theory. It was shown by Frank Ramsey[14] in 1926 (and later by Savage)[15] that if you accept some very simple and natural axioms about the way in which one should choose among a set of consequences in a consistent or coherent way, then it follows that you must act as if you had a utility function and a prior probability distribution. There is no other way in which you can act coherently.

This is a very important theoretical point but, sad to say, there are great practical difficulties in applying the theory. How do you calculate your utility function and measure your prior beliefs by a probability distribution ? Just you try, and see how difficult it is ! Of course we all would like to think that we were behaving coherently, so that the theory is very attractive.

A closely related, but different, theory involving decision making under uncertainty arose over a similar time period. The theory of games, in which two or more players are competing to maximise their own reward which is determined by the strategies of each player, has grown rapidly out of the classic book by von Neumann and Morgenstern published in 1947.[16]

*Modern Statistics, theory and applications*

Now let us put these four components together.

About 1890 there was a great leap forward in which new ideas abounded, old ones were rediscovered, and it was generally recognised that they were applicable to a very wide range of problems indeed. All this largely came about through the close collaboration of the biologists Weldon and Galton, the economist Edgeworth and the mathematician Karl Pearson.

This illustrates a point which I hope has already been apparent throughout my lecture. Mathematicians, and statisticians in particular, thrive on problems brought to them by other research workers. This is true not only because we like solving problems, but because practical problems are a great driving force for the development of new mathematical theory. In return we hope we can sometimes help to solve those immediate problems and, perhaps more importantly, in the process we may impart some ideas to the applied researcher which will influence his future approach to research.

Weldon and Galton were interested in heredity, and problems in genetics have been important ever since in the development of statistical theory. Galton was responsible for introducing the ideas of correlation and regression. The theory developed rapidly in the hands of Pearson, while the applications mushroomed in biology, psychology and many other fields. By 1906 it must have been very widespread because we find G. B. Shaw writing in the introduction to "The Doctor's Dilemma",

*It is easy to prove that the wearing of tall hats and the carrying of umbrellas enlarges the chest, prolongs life, and confers comparative immunity from disease . . . A university degree, a daily bath, the owning of thirty pairs of trousers, a knowledge of Wagner's music, a pew in church, anything, in short, that implies more means and better nurture . . . can be statistically palmed off as a magic-spell conferring all sorts of privileges . . . The mathematician whose correlations would fill a Newton with admiration, may, in collecting and accepting data and drawing conclusions from them, fall into quite crude errors by just such popular oversights as I have been describing.*

I said ealier that in statistical inference our conclusions will always be in some sense random, and we must find ways of describing this randomness. Pearson *et al* recognised this and coped with it in a number of ways. Up to this time it had been the custom to compare the distribution of data with a theoretical probability distribution simply by seeing if they looked similar. Pearson recognised that some formal criterion for testing the goodness of fit was needed and that this criterion would



itself have a probability distribution, the famous chi-squared distribution. This has received tremendous use ever since, unfortunately often erroneously.

In trying to estimate a theoretical quantity from data, it was recognised that this estimator was itself a random variable and that if possible its probability distribution should be calculated.

Simpson had already realised that in 1755.

In the early years of this century W. S. Gossett (who wrote under the pseudonym of Student) came into contact with Pearson. Working as he did for a commercial enterprise, Guiness of Dublin, he did not have a large enough research budget to obtain the large amounts of data which were customarily obtained at the time. Consequently he was forced to make his inferences from small samples and also to pay great attention to careful design of his experiments, thereby causing something of a revolution in statistical thinking. Subsequently agricultural experimentation was to provide an important motivation for the development of new statistical theory both for Gossett and Ronald Fisher, who was appointed to Rothamstead agricultural research station in 1919. The analysis of variance technique, so common today, was invented and the theory of design of experiments became a complex and sophisticated tool. One of the fundamental ideas of valid experimentation, that of randomization, was expounded by Fisher. For example if you wish to compare the effect of using fertilizer on the yield of some crop, and having obtained a number of plots on which to grow the crop, then the choice of which plots whould receive the fertilizer and which should not must be made by a suitable random device.

Starting in about 1915 Fisher also made great contributions to the theory of statistical inference with his recognition that the *likelihood function* was of fundamental significance. This is so important a concept that a simple example is called for. Suppose there is some event A which has probability p and I perform n independent and identical experiments out of which the event A occurs on r occasions, then the probability of obtaining this particular experimental result is given by

$p^r(1 - p)^{n-r}$

Considering this as a function of p, you have the likelihood function for this experiment. To be more explicit suppose that in a random sample of 20 people from Swansea you found 3 Welsh speakers, then you have the likelihood $p^3 (1 - p)^{17}$, where p stands for the proportion of Welsh speakers in Swansea. This function is plotted in Figure 5. Now looking at this it seems obvious that p is likely to be approximately 3/20, (or 5%) the so-called *maximum likelihood estimate*. Also it is very implausible that p should be less than about 0.01 or greater than about 0.5

Figure 5 Likelihood Function for Proportion of Welsh Speakers in Swansea

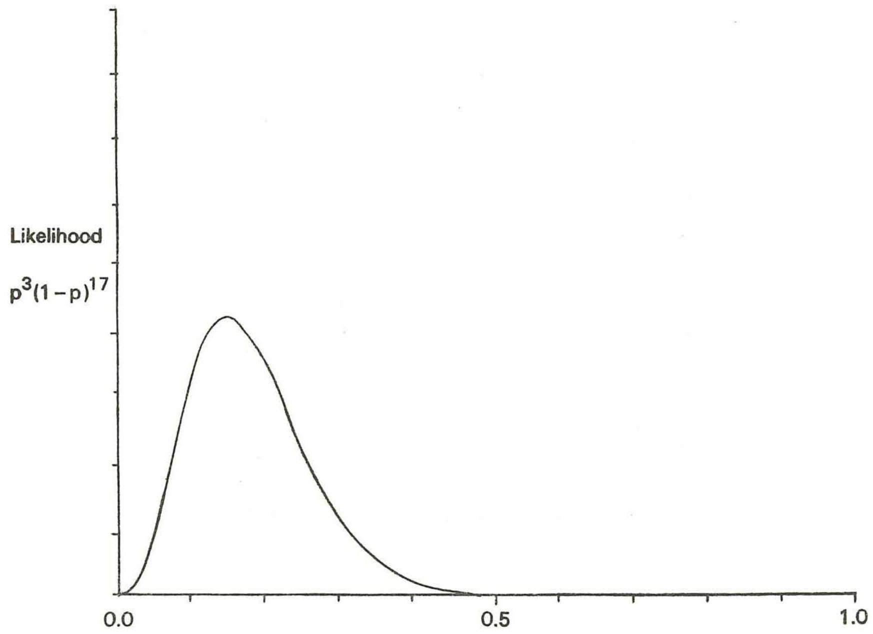
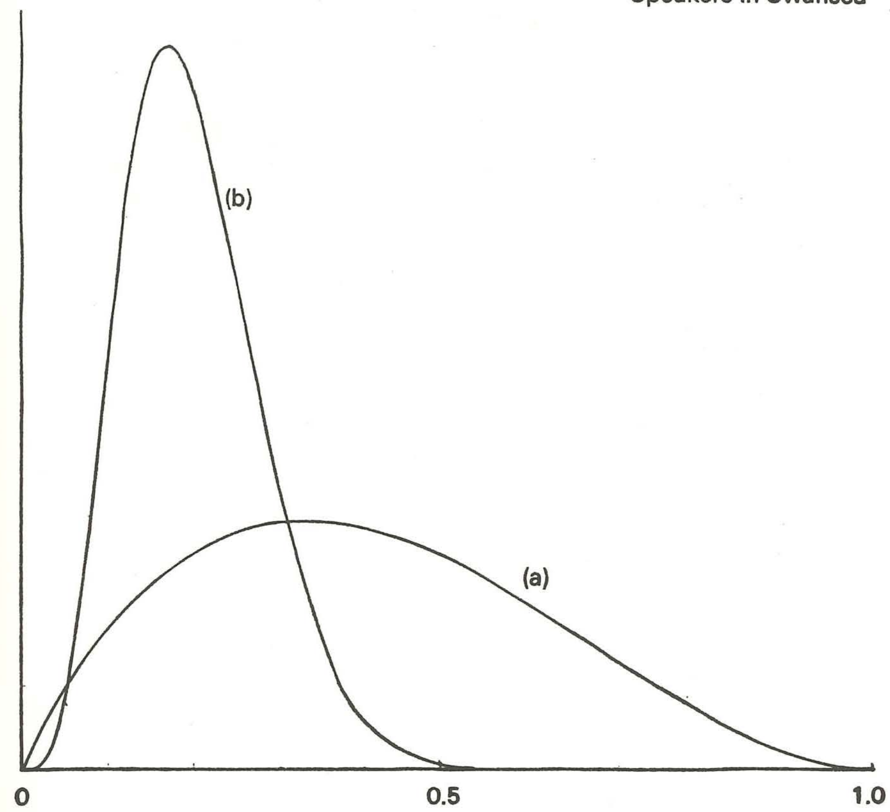





Figure 6 Prior (a) and Posterior (b) Distributions for Proportion of Welsh Speakers in Swansea




Fisher studied the properties of maximum likelihood estimators. He found that very often they are the best estimators (though not always) and that, even in situations where they are not best in small samples, they usually become fully efficient as the sample size increases. The idea of a maximum likelihood estimator was not new, Lambert had used it in 1760, but its importance had not been fully realised before.

Our example also shows that it is a good idea to estimate a parameter, not by a single value like 3/20, but by an interval. Fisher introduced the notion of a *fiducial interval* whose logic was obscure and caused bitter controversy for years. It is now almost dead, but refuses to lie down completely. He also contributed to the logic of hypothesis testing.

This problem was to be clarified considerably in a series of papers published jointly by Jerzy Neyman and Egon Pearson (son of Karl) starting in 1928. The problem of accepting or rejecting hypotheses was formally very similar to that of accepting or rejecting batches of items in quality control, and in this they were influenced to some extent by the work of Shewhart. Once again they found the likelihood function played an important role in constructing the most powerful tests. They constructed *confidence intervals*, instead of fiducial intervals, in which one can calculate the probability that the interval contains the true value of the parameter being estimated. This should be interpreted in the sense that if you always use a probability of 0.95 say, then 95% of the intervals you construct will contain the true parameter value and 5% of them won't.

In our example on Welsh speakers in Swansea we obtained a 95% confidence interval ranging from 3Æ to 36% of Welsh speakers. This may not seem very accurate, but it is not at all bad when you consider it is based on a sample of only 20 people. With a larger sample you would get more accurate estimation, e.g. if you sampled 5,000 out of the total population of about 170,000 your estimate would be correct to within plus or minus 1%. I am sure that would be sufficiently accurate for all practical purposes, and obtained at a fraction of the cost of a complete enumeration.

Hypothesis tests are certainly illogical ; estimation is a much sounder and more important activity ; after all almost all hypotheses are false, the question is not are they true or false, but how false are they ? To illustrate the illogicality suppose you have the following data from 100 two-child families.



| | 1st Child: Boy | 1st Child: Girl | |
|---|---|---|---|
| 2nd Child: Boy | 30 | 20 | 50 |
| 2nd Child: Girl | 20 | 30 | 50 |
| | 50 | 50 | 100 |

Consider hypotheses

A — The probability of a boy is 1/2 and the sexes of the children are independent

B — The sexes of the two children are independent

Using a 5% significance level we find that B is rejected ($x_1^2$) while A is not rejected ($x_3^2$), even though A logically implies B.

Despite such behaviour the hypothesis test, used with care, is a useful tool in the sense that it can help to prevent people making extravagant claims. This is not without value. For example in 1966 Schor and Karten found that in 72% of a sample of 149 articles from highly regarded medical journals, the conclusions drawn were not justified by the results presented, as judged by the usual statistical norms.[17]

Since 1930 the theory and practice of sample surveys has grown. There is a new enthusiasm for non-parametric methods (in which we make as few assumptions as possible). Since 1940 there have been big advances in multivariate analysis, time series and econometrics.

All of these ideas can be classified as *frequentist or sampling theory*. Other ideas involve the use of *subjective probability* or *Bayesian* ideas, in which the utility function component of decision theory is not used, but each individual must supply his own prior beliefs in the form of a probability distribution on the parameters, or the probability of an hypothesis being true, and on obtaining the data use Bayes' Theorem to modify those beliefs (likelihood has a role in this too).

For example, in our problem of estimating Welsh speakers, suppose our prior beliefs about p are represented by the prior probability density $12\, p(1 - p)^2$, shown in Figure (6a). Multiplying by the likelihood, and putting in a suitable constant, we find the posterior probability density of p to be $212520\, p^4\, (1 - p)^{19}$ as hown in Figure (6b). Thus we have a direct measure of the probability that an hypothesis is true, but that probability is unique to the person who calculated the prior probabilities. This of course is controversial. Some would argue it is unscientific to allow personal prejudice to enter in this way, but if you allow that possibility then this is a coherent way to do it. Logic aside, however, I find it difficult in a practical situation to give any such precise quantification of my prior beliefs.

Bayes and Gauss used those methods. Karl Pearson did too at first but later rejected them. These pioneers used, without justification, uniform priors but the neo-Bayesians are more self-confident ; everyone is entitled to his own individual beliefs, although Jeffreys tried to derive rational principles for choosing impersonal priors. Prominent among the subjectivists have been Ramsey, J. M. Keynes, and more recently Savage and Dennis Lindley (formerly Professor of Statistics at Aberystwyth).[18]

In George Eliot's "Middlemarch" we find this approach described as follows :

. . . *before anything was known of Lydgate's skill, the judgements on it had naturally been divided, depending on a sense of likelihood, situated perhaps in the pit of the stomach, or in the pineal gland, and differing in its verdict, but not less valuable as a guide in the total deficit of evidence.*

*Decision theory* itself may be applied directly to statistical inference if costs are allocated to incorrect decisions about the truth of hypotheses or to the distance of an estimator from the true value of the quantity being estimated. There are some interesting consequences of this approach but on the whole my view is that in scientific inference we are merely accumulating and evaluating evidence, not making decisions. More recently there has arisen a movement called *empirical-Bayes* and another, the *likelihood school*, and also methods of *structural inference.*



In conclusion, the central problems of theoretical statistical inference lie in producing a general method of comparing a probability model with a set of data. There are a number of schools of thought, none of which is without logical or practical difficulties, and the merits of which will continue to be debated with much enthusiasm. I do not think that a single unified theory will ever be achieved ; different problems are best treated by different theories. Thus, while my main allegiance is to the Frequentist Faith, I am content to use other approaches for particular problems. While the Philosophical debate continues the practising statistician must continue to do commonsense things with his data, even if he is a little bit illogical. After all, when Wellington constructed a telephone out of two cans and a piece of string he modestly said

*"I'm sorry but I don't know exactly how this invention works, you know".*

*"Don't let it worry you", said Great Uncle Bulgaria. "As long as it does work, that's the important part".*[19,20]

## REFERENCES

1. E. S. Pearson and M. G. Kendall (Ed.) (1970). *Studies in the History of Statistics and Probability.* Griffin.
2. W. Playfair (1801). *Statistical Breviary.* London.
3. K. Pearson (1924). *The Life, Letters and Labours of Francis Galton.* C.U.P.
4. J. Harold Wilson (1973). *Statistics and Decision Making in Government—Bradshaw Revisited : The Address of the President.* Journal of the Royal Statistical Society, B, 136, 1-20.

5. I. Todhunter (1865). *A History of the Mathematical Theory of Probability from the time of Pascal to that of Laplace*. Macmillan.
6. F. N. David (1962). *Gods, Games and Gambling*. Griffin.
7. C. Huygens (1657). *Tractatusde ratiociniis in aleae ludo.*
8. Suetonius. *Life of Augustus* and *Life of Claudius.*
9. L. A. Waddell (1934). *The Buddhism of Tibet* (2nd *Edition*). Heffer.
10. K. Pearson (1894). *Contributions to the Mathematical Theory of Evolution.* Phil. Trans. A185, 71-110.
11. J. Marschak from P. Lazarsfeld (Ed.) (1954). *Mathematical Thinking in the Social Sciences.* Free Press of Glencoe, New York.
12. J. Wesley (1737). *Journal—Vol.* 1, *Friday* 4 *March.*
13. T. Bayes (1763). *An Essay Towards Solving A Problem in the Doctrine of Chances.* Phil. Trans. Royal Society.
14. F. Ramsey (1931). *The Foundations of Mathematics and Other Essays.* Kegan, Paul, Trench, Trubner & Co.
15. L. J. Savage (1954). *The Foundations of Statistics.* Wiley.
16. J. Von Neumann and O. Morgenstern (1947). *Theory of Games and Economic Behaviour.* Princeton University Press.
17. S. Schor and I. Karten (1966). *Statistical Evaluation of Medical Journal Manuscripts.* J. Am. Med. Ass. 195, 1123-8.
18. D. V. Lindley (1971). *Bayesian Statistics, a Review.* SIAM, Philadelphia.
19. E. Beresford (1973). "Peep-Peep-Peep" in *The Invisible Womble and Other Stories.* Puffin.
20. D. J. Bartholemew (1974). *Social Probability.* Inaugural lecture at the London School of Economics.