the problem rather than of statistics, but even if we view it as such, the consequent difficulty in empirical model building and model selection is very much part of statistics and a statistician's life. This discussion also contains a warning that statisticians cannot work in a vacuum, since most of the relevant factors that create inductive uncertainties in a problem are subject matter specific.

1.4   The emergence of statistics

It is impossible to calculate accurately events which are determined by chance. – Thucydides (c. 400BC)

There were two strands in the emergence of statistics. One was the development of the theory of probability, which had its original motivation in the calculation of expectation or uncertainties in gambling problems by Pascal (1623–1662) and Fermat (1601–1665). The theory was later developed on the mathematical side by Huygens (1629–1695), the Bernoulli brothers, in particular James Bernoulli (1654–1705), de Moivre (1667–1754) and Laplace (1749–1827), and on the logical side by Bayes (1701–1761), Boole (1815–1864) and Venn (1834–1923).

The growth of probability theory was an important milestone in the history of science. Fisher liked to comment that it was unknown to the Greek and the Islamic mathematicians (Thucydides was a historian); Persi Diaconis once declared that our brain is not wired to solve probability problems. With probability theory, for the first time since the birth of mathematics, we can make rigorous statements about uncertain events. The theory, however, is mostly deductive, which makes it a true branch of mathematics. Probability statements are evaluated as consequences of axioms or assumptions rather than specific observations. Statistics as the child of probability theory was born with the paper of Bayes in 1763 and was brought to maturity by Laplace.

The second strand in the emergence of statistics was an almost parallel development in the theory of errors. The main emphasis was not on the calculation of probabilities or uncertainties, but on summarizing observational data from astronomy or surveying. Gauss (1777–1855) was the main contributor in this area, notably with the principle of *least squares* as a general method of estimation. The important ingredient of this second line of development was the data-rich environment. In this connection Fisher noted the special role of Galton (1822-1911) in the birth of modern statistics towards the end of the 19th century. A compulsive data gatherer, Galton had a passionate conviction in the power of quantitative and statistical methods to deal with 'variable phenomena'.

Further progress in statistics continues to depend on data-rich environments. This was first supplied by experiments in agriculture and biometry, where Fisher was very much involved. Later applications include: industrial quality control, the military, engineering, psychology, business, medicine and health sciences. Other influences are found in data gathering and analysis for public or economic policies.

Bayesians and frequentists

The Bayesian and frequentist schools of statistics grew in response to problems of uncertainty, in particular to the way probability was viewed. The early writers in the 18th and 19th centuries considered it both a (subjective) degree of belief and (objective) long-run frequency. The 20th century brought a strong dichotomy. The frequentists limit probability to mean only a long-run frequency, while for the Bayesians it can carry the subjective notion of uncertainty.

This Bayesian–frequentist divide represents the fundamental tension between the need to say something relevant on a specific instance/dataset and the sense of objectivity in long-run frequencies. If we toss a coin, we have a sense of uncertainty about its outcome: we say the probability of heads is 0.5. Now, think about the *specific* next toss: can we say that our sense of uncertainty is 0.5, or is the number 0.5 meaningful only as a long-term average? Bayesians would accept both interpretations as being equally valid, but a true frequentist allows only the latter.

Since the two schools of thought generate different practical methodologies, the distinction is real and important. These disagreements do not hinder statistical applications, but they do indicate that the foundation of statistics is not settled. This tension also provides statistics with a fruitful dialectical process, at times injecting passion and emotion into a potentially dry subject. (Statisticians are probably unique among scientists with constant ponderings of the foundation of their subject; physicists are not expected to do that, though Einstein did argue with the quantum physicists about the role of quantum mechanics as the foundation of physics.)

Inverse probability: the Bayesians

The first modern method to assimilate observed data for quantitative inductive reasoning was published (posthumously) in 1763 by Bayes with his *Essay towards Solving a Problem in the Doctrine of Chances*. He used an inverse probability, via the now-standard Bayes theorem, to estimate a binomial probability. The simplest form of the Bayes theorem for two events A and B is

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\bar{A})P(\bar{A})}. \tag{1.1}$$

Suppose the unknown binomial probability is θ and the observed number of successes in n independent trials is x. Then, in modern notation, Bayes's solution is

$$f(\theta|x) = \frac{f(x,\theta)}{f(x)} = \frac{f(x|\theta)f(\theta)}{\int f(x|\theta)f(\theta)d\theta}, \tag{1.2}$$

where $f(\theta|x)$ is the conditional density of θ given x, $f(\theta)$ is the so-called prior density of θ and $f(x)$ is the marginal probability of x. (Note that we have used the symbol $f(\cdot)$ as a generic function, much like the way we use $P(\cdot)$ for probability. The named argument(s) of the function determines

what the function is. Thus, $f(\theta, x)$ is the joint density of θ and x, $f(x|\theta)$ is the conditional density of x given θ, etc.)

Leaving aside the problem of specifying $f(\theta)$, Bayes had accomplished a giant step: he had put the problem of inductive inference (i.e. learning from data x) within the clean deductive steps of mathematics. Alas, 'the problem of specifying $f(\theta)$' a priori is an equally giant point of controversy up to the present day.

There is nothing controversial about the Bayes theorem (1.1), but (1.2) is a different matter. Both A and B in (1.1) are random events, while in the Bayesian use of (1.2) only x needs to be a random outcome; in a typical binomial experiment θ is an unknown fixed parameter. Bayes was well aware of this problem, which he overcame by considering that θ was generated in an *auxiliary physical experiment* – throwing a ball on a level square table – such that θ is expected to be uniform in the interval $(0, 1)$. Specifically, in this case we have $f(\theta) = 1$ and

$$f(\theta|x) = \frac{\theta^x (1-\theta)^{n-x}}{\int_0^1 u^x (1-u)^{n-x} du}. \tag{1.3}$$

Fisher was very respectful of Bayes's seeming apprehension about using an axiomatic prior; in fact, he used Bayes's auxiliary experiment to indicate that Bayes was not a Bayesian in the modern sense. If θ is a random variable then there is nothing 'Bayesian' in the use of the Bayes theorem. Frequentists do use Bayes theorem in applications that call for it.

Bayes did, however, write a *Scholium* (literally, a 'dissertation'; see Stigler 1982) immediately after his proposition:

> ... the same rule [i.e. formula (1.3) above] is a proper one to be used in the case of an event concerning the probability of which we absolutely know nothing antecedently to any trial made concerning it.

In effect, he accepted the irresistible temptation to say that if we know nothing about θ then it is equally probable to be between zero and one. More significantly, he accepted that the uniform prior density, which now can be purely axiomatic, can be processed with the objective binomial probability to produce a posterior probability. So, after all, Bayes was a Bayesian, albeit a reluctant one. (In hindsight, probability was then the only available concept of uncertainty, so Bayes did not have any choice.)

Bayes's paper went largely unnoticed until Pearson (1920). It was Laplace, who, after independently discovering Bayes theorem, developed Bayesian statistics as we understand it today. Boole's works on the probability theory (e.g. *Laws of Thought*, published in 1854), which discussed Bayes theorem in the 'problem of causes', clearly mentioned Laplace as the main reference. Laplace's *Théorie Analytique des Probabilités* was first published in 1812 and became the standard reference for the rest of the century. Laplace used the flat or uniform prior for all estimation problems, presented or justified as a reasonable expression of ignorance. The principle of inverse probability, hence Bayesian statistics, was an integral part of

the teaching of probability until the end of the 19th century. Fisher (1936) commented that that was how he learned inverse probability in school and 'for some years saw no reason to question its validity'.

Statistical works by Gauss and others in the 19th and early 20th centuries were largely Bayesian with the use of inverse probability arguments. Even Fisher, who later became one of the strongest critics of axiomatic Bayesianism, in his 1912 paper 'On an absolute criterion for fitting frequency curves', erroneously called his maximum likelihood the 'most probable set of values', suggesting inverse probability rather than likelihood, although it was already clear he had distinguished these two concepts.

Repeated sampling principle: the frequentists

A dominant section of statistics today views probability formally as a long-run frequency based on repeated experiments. This is the basis of the frequentist ideas and methods, where the truth of a mathematical model must be validated through an objective measure based on externally observable quantities. This feels natural, but as Shafer (1990) identified, 'the rise of frequentism' in probability came only in the mid-19th century from the writings of empiricist philosophers such as John Stuart Mill. Population counting and classification was also a factor in the empirical meaning of probability when it was used for modelling.

The *repeated sampling principle* specifies that procedures should be evaluated on the basis of repeat experimentation under the same conditions. The sampling distribution theory, which expresses the possible outcomes from the repeated experiments, is central to the frequentist methodology. Many concepts in use today, such as bias, variability and standard error of a statistic, P-value, type I error probability and power of a test, or confidence level, are based on the repeated sampling principle. The dominance of these concepts in applied statistics today proves the practical power of frequentist methods. Neyman (1894–1981) and Wald (1902–1950) were the most influential exponents of the frequentist philosophy. Fisher contributed enormously to the frequentist methodology, but did not subscribe fully to the philosophy.

True frequentism states that measures of uncertainties are to be interpreted only in a repeated sampling sense. In areas of statistical application, such as medical laboratory science or industrial quality control, where procedures are naturally repeated many times, the frequentist measures are very relevant.

The problem arises as the requirement of repeat experimentation is allowed to be hypothetical. There are many areas of science where experiments are unlikely to be repeated, for example in archaeology, economics, geology, astronomy, medicine, etc. A reliance on repeated sampling ideas can lead to logical paradoxes that appear in common rather than esoteric procedures.

Extreme frequentism among practical statisticians is probably quite rare. An extremist will insist that an observed 95% confidence interval,

say $1.3 < \theta < 7.1$, either covers the parameter or it does not, we do not know which, and there is no way to express the uncertainty; the 95% applies only to the procedure, not to the particular interval. That is in fact the orthodox interpretation of the confidence interval. It neglects the evidence contained in a particular interval/dataset, because measures of uncertainty are only interpreted in hypothetical repetitions.

Most scientists would probably interpret the confidence interval intuitively in a subjective/Bayesian way: there is a 95% probability the interval contains the true parameter, i.e. the value 95% has some evidential attachment to the observed interval.

Bayesians versus frequentists

A great truth is a truth whose opposite is also a great truth. – Thomas Mann (1875–1955)

In Bayesian computations one starts by explicitly postulating that a parameter θ has a distribution with prior density $f(\theta)$; for example, in a problem to estimate a probability θ, one might assume it is uniformly distributed on (0,1). The distinguishing attitude here is that, since θ does not have to be a random outcome of an experiment, this prior can be specified axiomatically, based on thinking alone. This is the methodological starting point that separates the Bayesians from the frequentists, as the latter cannot accept that a parameter can have a distribution, since such a distribution does not have an external reality. Bayesians would say there is an uncertainty about θ and insist any uncertainty be expressed probabilistically. The distribution of θ is interpreted in a subjective way as a degree of belief.

Once one accepts the prior $f(\theta)$ for θ and agrees it can be treated as a regular density, the way to proceed is purely deductive and (internally) consistent. Assuming that, given θ, our data x follows a statistical model $p_\theta(x) = f(x|\theta)$, then the information about θ contained in the data is given by the *posterior* density, using the Bayes theorem as in (1.2),

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{f(x)}.$$

In Bayesian thinking there is no operational difference between a prior density $f(\theta)$, which measures belief, and $f(x|\theta)$, which measures an observable quantity. These two things are conceptually equal as measures of uncertainty, and they can be mixed using the Bayes theorem.

The posterior density $f(\theta|x)$, in principle, captures from the data all the information that is relevant for θ. Hence, it is an update of the prior $f(\theta)$. In a sequence of experiments it is clear that the current posterior can function as a future prior, so the Bayesian method has a natural way of accumulating information.

When forced, most frequentists would probably admit that a degree of belief does exist subjectively. The disagreement is not that a parameter

can assume a density, since frequentists could also think of $f(\theta)$ as a prior likelihood (the likelihood of the parameter before we have any data). Two genuine concerns exist:

(i) the practical problem of choosing an appropriate prior. Leaving aside the problem of subjective interpretation, there is an ongoing controversy on how we should pick $f(\theta)$. Several early writers such as Boole (1854, pages 384, 392) and Venn (1876) had criticized the arbitrariness in the axiomatic choice of $f(\theta)$; Fisher was also explicit in his rejection of any axiomatic prior, although *he did not rule out* that some applications, such as genetics, may have physically meaningful $f(\theta)$. Modern Bayesians seem to converge toward the so-called 'objective priors' (e.g. Gatsonis *et al.* 1997), but there are many shades of Bayesianism (Berger 2000).

(ii) the 'rules of engagement' regarding a subjective degree of belief. There is nothing really debatable about how one feels, and there is nothing wrong in thinking of probability in a subjective way. However, one's formal action based on such feeling is open to genuine disagreement. Treating a subjective probability density like a regular density function means, for example, that it can be integrated out, and it needs a Jacobian term when transformed to a different scale. The latter creates a lack of invariance in the choice of prior: seeming ignorance in one scale becomes information in another scale (see Section 2.8). □

Efron (1998) compares the psychological differences between the two schools of thought. A comparative study highlights the strengths and weaknesses of each approach. The strength of the Bayesian school is its unified approach to all problems of uncertainty. Such unity provides clarity, especially in complex problems, though it does not mean Bayesian solutions are practical. In fact, until recently Bayesians could not solve complex problems because of computational difficulties (Efron 1986a). While, bound by fewer rules, the strength of a frequentist solution is usually its practicality.

Example 1.2: A new eye drug was tested against an old one on 10 subjects. The two drugs were randomly assigned to both eyes of each person. In all cases the new drug performed better than the old drug. The P-value from the observed data is $2^{-10} = 0.001$, showing that what we observe is not likely due to chance alone, or that it is very likely the new drug is better than the old one. □

Such simplicity is difficult to beat. Given that a physical randomization was actually used, very little extra assumption is needed to produce a valid conclusion. And the final conclusion, that the new drug is better than the old one, might be all we need to know from the experiment. The achieved simplicity is a reward of focus: we are only interested in knowing if chance alone could have produced the observed data. In real studies, of course, we might want to know more about the biological mechanism or possible side effects, which might involve more complicated measurements.

The advent of cheap computer power and Monte Carlo techniques (e.g. Gilks *et al.* 1995) have largely dismantled the Bayesian computational wall. Complex problems are now routinely solved using the Bayesian methodology. In fact, being pragmatic, one can separate the Bayesian numerical methods from the underlying philosophy, and use them as a means of obtaining likelihood functions. This is a recent trend, for example, in molecular genetics. In Section 10.6 we will see that the Bayesian and likelihood computations have close numerical connections.

Luckily, in large-sample problems, frequentist and Bayesian computations tend to produce similar numerical results, since in this case the data dominate the prior density and the level of uncertainty is small. In small-to medium-sized samples, the two approaches may not coincide, though in real data analysis the difference is usually of smaller order of magnitude than the inductive uncertainty in the data and in the model selection.

The following 'exchange paradox', discussed in detail by Christensen and Utts (1992), illustrates how our handling of uncertainty affects our logical thinking. To grasp the story quickly, or to entertain others with it, replace x by 100.

Example 1.3: A swami puts an unknown amount of money in one envelope and twice that amount in another. He asks you to pick one envelope at random, open it and then decide if you would exchange it with the other envelope. You pick one (randomly), open it and see the outcome $X = x$ dollars. You reason that, suppose Y is the content of the other envelope, then Y is either $x/2$ or $2x$ with probability 0.5; if you exchange it you are going to get $(x/2 + 2x)/2 = 5x/4$, which is bigger than your current x. 'With a gleam in your eye', you would exchange the envelope, wouldn't you?

The reasoning holds for any value of x, which means that you actually *do not need to open the envelope* in the first place, and you would still want to exchange it! Furthermore, when you get the second envelope, the same reasoning applies again, so you should exchange it back. A discussion of the Bayesian and frequentist aspects of this paradox is left as an exercise. □

1.5   Fisher and the third way

The likelihood approach offers a distinct 'third way', a Bayesian-frequentist compromise. We might call it Fisherian as it owes most of its conceptual development to Fisher (1890–1962). Fisher was clearly against the use of the axiomatic prior probability fundamental to the Bayesians, but he was equally emphatic in his rejection of long-run frequency as the only way to interpret probability. Fisher was a frequentist in his insistence that statistical inference should be objectively verifiable; however, his advocacy of likelihood inference in cases where probability-based inference is not available puts him closer to the Bayesian school.

In a stimulating paper on Fisher's legacies, Efron (1998) created a statistical triangle with Fisherian, Bayesian and frequentist nodes. He then placed various statistical techniques within the triangle to indicate their flavour.

Fisher's effort for an objective inference without any use of prior probability led him to the idea of *fiducial probability* (Fisher 1930, 1934). This concept prompted the confidence interval procedure (Neyman 1935). It appears that Fisher never managed to convince others what fiducial probability was, despite his insistence that, conceptually, it is 'entirely identical with the classical probability of the early writers' (Fisher 1973, page 54). In some models the fiducial probability coincides with the usual frequentist/long-run-frequency probability. The problems occur in more complex models where exact probability statements are not possible.

From his last book *Statistical Methods and Scientific Inference* (1973, in particular Chapter III) it is clear that Fisher settled with the idea that

- whenever possible to get exact results we should base inference on probability statements, otherwise it should be based on the likelihood;

- the likelihood can be interpreted subjectively as a rational degree of belief, but it is weaker than probability, since it does not allow an external verification, and

- in large samples there is a strengthening of likelihood statements where it becomes possible to attach some probabilistic properties ('asymptotic approach to a higher status' – Fisher 1973, page 78).

These seem to summarize the Fisherian view. (While Fisher's probability was fiducial probability, let us take him at his own words that it is 'entirely identical with the classical probability'.) About 40 years elapsed between the explicit definition of the likelihood for the purpose of estimation and Fisher's final judgement about likelihood inference. The distinguishing view is that *inference is possible directly from the likelihood function*; this is neither Bayesian nor frequentist, and in fact both schools would reject such a view as they allow only probability-based inference.

These Fisherian views also differ from the so-called 'pure likelihood view' that considers the likelihood as the sole carrier of uncertainty in statistical inference (e.g. Royall 1997, although he would call it 'evidence' rather than 'uncertainty'). Fisher recognized two 'well-defined levels of logical status' for uncertainty about parameters, one supplied by probability and the other by likelihood. A likelihood-based inference is used to 'analyze, summarize and communicate statistical evidence of types too weak to supply true probability statements' (Fisher 1973, page 75). Furthermore, when available, a probability statement must allow for an external verification (a verification by observable quantities), so it is clear that frequentist consideration is also an important aspect of the Fisherian view.

Fisher's requirement for an exact probability inference is more stringent than the so-called 'exact inference' in statistics today (Fisher 1973, pages 69–70). His prototype of an exact probability-based inference is the confidence interval for the normal mean (even though the term 'confidence interval' is Neyman's). The statement

$$P(\overline{x} - 1.96\sigma/\sqrt{n} < \mu < \overline{x} + 1.96\sigma/\sqrt{n}) = 0.95$$

is unambiguous and exactly/objectively verifiable; it is an ideal form of inference. However, the so-called 'exact 95% confidence interval' for the binomial proportion (see Section 5.8) in fact does not have exactly 95% coverage probability, so logically it is of lower status than the exact interval for the normal model. It is for this situation the likelihood is indicated.

For Fisher, both likelihood and probability are measures of uncertainty, but they are on a different footing. This is a non-Bayesian view, since for Bayesians all uncertainty is measured with probability. The subjective element in the interpretation of likelihood, however, is akin to a Bayesian/non-frequentist attitude. It is worth noting that, when backed up with large-sample theory to supply probability statements, the mechanics and numerical results of likelihood inference are generally acceptable to frequentist statisticians. So, in their psychology, Fisherians are braver than the frequentists in saying that inference is possible from the likelihood function alone, but not as brave as the Bayesians to admit an axiomatic prior into the argument.

Legacies

By 1920 the field of statistics must have been a confusing place. Yates (1990) wrote that it was the age of correlation and coefficients of all kinds. To assess association in 2×2 tables there were the coefficient of association, coefficient of mean square contingency, coefficient of tetrachoric correlation, equiprobable tetrachoric correlation, and coefficient of colligation, but the idea of estimating the association and its test of significance were mixed up. There were many techniques available, such as the least squares principle, the method of moments, the inverse probability method, the χ^2 test, the normal distribution, Pearson's system of curves, the central limit theorem, etc., but there was no firm logical foundation.

The level of confusion is typified by the title of Edgeworth's paper in 1908 and Pearson's editorial in *Biometrika* in 1913: '*On the probable errors of frequency constants*', which in modern terminology would be 'the standard error of fixed parameters'. There was simply no logical distinction or available terms for a parameter and its estimate. On the mathematical side, the χ^2 test of association for the 2×2 tables had 3 degrees of freedom!

A more serious source of theoretical confusion seems to be the implicit use of inverse probability arguments in many early statistical works, no doubt the influence of Laplace. The role of the prior distribution in inverse probability arguments was never seriously questioned until early 20th century. When explicitly stated, the arbitrariness of the prior specification was probably a stumbling block to a proper appreciation of statistical questions as objective questions. Boole (1854) wrote in the *Laws of Thoughts* (Chapter XX, page 384) that such arbitrariness

> seems to imply, that definite solution is impossible, and to mark the point where inquiry ought to stop.

Boole discussed the inverse probability method at length and identified its

weakness, but did not see any alternative; he considered the question of inductive inference as

> second to none other in the Theory of Probabilities in importance, [I hope it] will receive the careful attention which it deserves.

In his works on the theory of errors, Gauss was also aware of the problem, but he got around it by justifying his method of estimation in terms of the least-squares principle; this principle is still central in most standard introductions to regression models, which is unfortunate, since (i) in itself it is devoid of inferential content and (ii) it is not natural for general probability models, so it creates an unnecessary conceptual gap with the far richer class of generalized linear models.

Fisher answered Boole's challenge by clearly identifying the likelihood as the key inferential quantity that is free of subjective prior probabilities. He stressed that if, prior to the data, we know absolutely nothing about a parameter (recall Bayes's *Scholium* in Section 1.4) then all of the information from the data is in the likelihood. In the same subjective way the Bayesians interpret probability, the likelihood provides a 'rational degree of belief' or an 'order of preferences' on possible parameter values; the fundamental difference is that *the likelihood does not obey probability laws*. So probability and likelihood are different concepts available to deal with different levels of uncertainty.

There were earlier writers, such as Daniel Bernoulli or Venn, who had used or mentioned the idea of maximum likelihood in rudimentary forms (see Edwards 1992, Appendix 2). It usually appeared under the name of 'most probable value', indicating the influence of inverse probability argument. Even Fisher in 1912 used that name, even though it was clear from the discussion he had likelihood in mind. The confusion was only cleared in 1921 when Fisher invented the term 'likelihood'.

In a series of the most influential papers in statistics Fisher (in particular in 1922 and 1925) introduced order into the chaos by identifying and naming the fundamental concepts such as 'parameter', 'statistic', 'variance', 'sufficiency', 'consistency', 'information', and 'estimation','maximum likelihood estimate', 'efficiency' and 'optimality'. He was the first to use Greek letters for unknown parameters and Latin letters for the estimates. He set up the agenda for statistical research by identifying and formulating the important questions.

He 'fixed' the degree of freedom of the χ^2 test for the 2×2 tables in 1922. He recognized the paper by 'Student' in 1908 on the t-test, which was ignored by the large-sample-based statistical world at the time, as a milestone in the history of statistics: it was the first exact test. He emphasized the importance of inference based on exact distribution and identified 'the problem of distribution' as a respectable branch of theoretical statistics. Fisher was unsurpassed in this area, being the first to derive the exact distribution of the t and F statistics, as well as that of the sample correlation and multiple correlation coefficient.

Fisher's influence went beyond the foundation of statistics and the likelihood methods. His *Statistical Methods for Research Workers*, first published in 1925, brought the new ideas to generations of practical research workers. Fisher practically invented the field of experimental design, introducing the fundamental ideas of randomization, replication, blocking, factorial experiments, etc., and its analysis of variance. His *Design of Experiments*, first published in 1935, emphasized the importance of carefully collected data to simplify subsequent analysis and to arrive at unambiguous conclusions. He contributed significantly to areas of sampling distribution theory, regression analysis, extreme value theory, nonparametric and multivariate analysis. In a careful study of Fisher's legacy, Savage (1976) commented that it would be a lot faster to list areas in statistics where Fisher did *not* contribute fundamentally, for example sequential analysis and time series modelling.

Outside statistics, many geneticists consider Fisher as the most important evolutionary biologist after Darwin. In 1930 Fisher was the first to provide a key synthesis of Mendelian genetics and Darwin's theory of evolution, thus giving a quantitative basis for the latter. Fisher was never a professor of statistics: he was Galton Professor of Eugenics at University College London, then Balfour Professor of Genetics at Cambridge University.

For a statistician, his writings can be inspirational as they are full of conviction on the fundamental role and contributions of statistical methods in science and in 'refinement of human reasoning'. Fisher (1952) believed that

> Statistical Science was the peculiar aspect of human progress which gave to the twentieth century its special character. ... it is to the statistician that the present age turns for what is most essential in all its more important activities.

The 'important activities' include the experimental programmes, the observational surveys, the quality control engineering, etc. He identified the crucial contribution of statistical ideas to the fundamental scientific advances of the 19th century such as in Lyell's *Principles of Geology* and Darwin's theory of evolution.

It is an unfortunate turn of history that Fisher's articles and books are no longer standard reading in the study of statistics. Fisher was often criticized for being obscure or hard to read. Savage (1976), however, reported that his statistical mentors, which included Milton Friedman and W. Allen Wallis, gave the advice: 'To become a statistician, practice statistics and mull Fisher over with patience, respect and scepticism'. Savage closed his 1970 Fisher Memorial Lecture with 'I do hope that you won't let a week go by without reading a little bit of Fisher'.

Fisher's publications were collected in the five-volume *Collected Papers of R.A. Fisher*, edited by Bennett and Cornish (1974). His biography, entitled *R.A. Fisher, The Life of a Scientist*, was published by his daughter Joan Fisher Box in 1978. Other notable biographies, memoirs or reviews

of his works were written by Barnard (1963), Bartlett (1965), Yates and Mather (1963), Kendall (1963), Neyman (1961, 1967), Pearson (1974) and Savage (1976). Recent articles include Aldrich (1997), Efron (1998) and Hald (1999). Edwards's (1992) book on likelihood was largely influenced by Fisher and the Appendices contain useful accounts of the history of likelihood and Fisher's key contributions. Fienberg and Hinkley (1980) contains a wide-ranging discussion of Fisher's papers and his impact on statistics.

1.6 Exercises

Exercise 1.1: Discuss the stochastic and inductive uncertainty in the following statements:

(a) A study shows that children of mothers who smoke have lower IQs than those of non-smoking mothers.

(b) A report by Interpol in 1994 shows a rate of (about) 55 crimes per 1000 people in the USA, compared to 100 in the UK and 125 in Sweden. ('Small' note: the newspaper that published the report later published a letter by an official from the local Swedish Embassy saying that, in Sweden, if a swindler defrauds 1000 people the case would be recorded as 1000 crimes.)

(c) Life expectancy in Indonesia is currently 64 years for women and 60 years for men. (To which generation do these numbers apply?)

(d) The current unemployment rate in Ireland is 4.7%. (What does 'unemployed' mean?)

(e) The total fertility rate for women in Kenya is 4.1 babies.

(f) The population of Cairo is around 16 million people. (Varies by a few million between night and day.)

(g) The national clinical trial of aspirin, conducted on about 22,000 healthy male physicians, established the benefit of taking aspirin. (To what population does the result apply?)

Exercise 1.2: What is wrong with the reasoning in the exchange paradox in Example 1.3? Discuss the Bayesian and frequentist aspects of the paradox, first assuming the 'game' is only played once, then assuming it is played repeatedly.

In All Likelihood

Statistical Modelling and Inference Using Likelihood

Yudi Pawitan

Department of Medical Epidemiology and Biostatistics
Karolinska Institutet
Stockholm, Sweden
`yudi.pawitan@mep.ki.se`