Statistics notes

Bayesians and frequentists

J Martin Bland, Douglas G Altman,

There are two competing philosophies of statistical analysis: the Bayesian and the frequentist. The frequentists are much the larger group, and almost all the statistical analyses which appear in the *BMJ* are frequentist. The Bayesians are much fewer and until recently could only snipe at the frequentists from the high ground of university departments of mathematical statistics. Now the increasing power of computers is bringing Bayesian methods to the fore.

Bayesian methods are based on the idea that unknown quantities, such as population means and proportions, have probability distributions. The probability distribution for a population proportion expresses our prior knowledge or belief about it, before we add the knowledge which comes from our data. For example, suppose we want to estimate the prevalence of diabetes in a health district. We could use the knowledge that the percentage of diabetics in the United Kingdom as a whole is about 2%, so we expect the prevalence in our health district to be fairly similar. It is unlikely to be 10%, for example. We might have information based on other datasets that such rates vary between 1% and 3%, or we might guess that the prevalence is somewhere between these values. We can construct a prior distribution which summarises our beliefs about the prevalence in the absence of specific data. We can do this with a distribution having mean 2 and standard deviation 0.5, so that two standard deviations on either side of the mean are 1% and 3%. (The precise mathematical form of the prior distribution depends on the particular problem.)

Suppose we now collect some data by a sample survey of the district population. We can use the data to modify the prior probability distribution to tell us what we now think the distribution of the population percentage is; this is the posterior distribution. For example, if we did a survey of 1000 subjects and found 15 (1.5%) to be diabetic, the posterior distribution would have mean 1.7% and standard deviation 0.3%. We can calculate a set of values, a 95% credible interval (1.2% to 2.4% for the example), such that there is a probability of 0.95 that the percentage of diabetics is within this set. The frequentist analysis, which ignores the prior information, would give an estimate 1.5% with standard error 0.4% and 95% confidence interval 0.8% to 2.5%. This is similar to the results of the Bayesian method, as is usually the case, but the Bayesian method gives an estimate nearer the prior mean and a narrower interval.

Frequentist methods regard the population value as a fixed, unvarying (but unknown) quantity, without a probability distribution. Frequentists then calculate confidence intervals for this quantity, or significance tests of hypotheses concerning it. Bayesians reasonably object that this does not allow us to use our wider knowledge of the problem. Also, it does not provide what researchers seem to want, which is to be able to say that there is a probability of 95% that the population value lies within the 95% confidence interval, or that the probability that the null hypothesis is true is less than 5%. It is argued that researchers want this, which is why they persistently misinterpret confidence intervals and significance tests in this way.

A major difficulty, of course, is deciding on the prior distribution. This is going to influence the conclusions of the study, yet it may be a subjective synthesis of the available information, so the same data analysed by different investigators could lead to different conclusions. Another difficulty is that Bayesian methods may lead to intractable computational problems. (All widely available statistical packages use frequentist methods.)

Most statisticians have become Bayesians or frequentists as a result of their choice of university. They did not know that Bayesians and frequentists existed until it was too late and the choice had been made. There have been subsequent conversions. Some who were taught the Bayesian way discovered that when they had huge quantities of medical data to analyse the frequentist approach was much quicker and more practical, although they may remain Bayesian at heart. Some frequentists have had Damascus road conversions to the Bayesian view. Many practising statisticians, however, are fairly ignorant of the methods used by the rival camp and too busy to have time to find out.

The advent of very powerful computers has given a new impetus to the Bayesians. Computer intensive methods of analysis are being developed, which allow new approaches to very difficult statistical problems, such as the location of geographical clusters of cases of a disease. This new practicability of the Bayesian approach is leading to a change in the statistical paradigm—and a rapprochement between Bayesians and frequentists.[1 2] Frequentists are becoming curious about the Bayesian approach and more willing to use Bayesian methods when they provide solutions to difficult problems. In the future we expect to see more Bayesian analyses reported in the *BMJ*. When this happens we may try to use Statistics notes to explain them, though we may have to recruit a Bayesian to do it.

We thank David Spiegelhalter for comments on a draft.

1 Breslow N. Biostatistics and Bayes (with discussion). *Statist Sci* 1990;5: 269-98.
2 Spiegelhalter DJ, Freedman LS, Parmar MKB. Bayesian approaches to randomized trials (with discussion). *J R Statist Soc A* 1994;157:357-416.

Department of Public Health Sciences, St George's Hospital Medical School, London SW17 0RE
J Martin Bland, *professor of medical statistics*
ICRF Medical Statistics Group, Centre for Statistics in Medicine, Institute of Health Sciences, Oxford OX3 7LF
Douglas G Altman, *head*
Correspondence to: Professor Bland

BMJ 1998;317:1151

Correction

North of England evidence based guidelines development project: guideline for the primary care management of dementia

An editorial error occurred in this article by Martin Eccles and colleagues (19 September, pp 802-8). In the list of authors the name of Moira Livingston [not Livingstone] was misspelt.