

## Bayesian statistics in action

Analytical Methods Committee, AMCTB No 52

Received 23rd May 2012

DOI: 10.1039/c2ay90023h

While data evaluation methods based on significance tests are familiar to analytical scientists, it is not so widely appreciated that alternative Bayesian methods are now also commonly used in a range of fields. This Technical Brief outlines some examples of the importance of Bayesian methods in a variety of areas of applied analytical science.

The limitations of conventional significance testing are illustrated by a simple example. Suppose we test a sample of tonic water to ensure that it contains no more than the maximum permitted level of quinine, taken to be 80 mg L<sup>-1</sup>. After making some fluorimetric measurements of the quinine level and determining their mean and variance we apply a *t*-test (a one-tailed one, as we are only concerned with quinine levels >80 mg L<sup>-1</sup>) using the null hypothesis  $H_0: \mu = 80$  mg L<sup>-1</sup>. The outcome of this test provides us with the probability of getting the experimental (or more extreme) results if the null hypothesis is true. We reject  $H_0$  if this probability is too low. But what we really need is the converse of this: we want the probability that the quinine level exceeds 80 ppm, given the experimental results. This difficulty was tackled by the Presbyterian minister Thomas Bayes (1702–1761), who introduced the idea of attaching probabilities to hypotheses. The Bayesian method involves making an initial assumption on the nature of the data distribution: this is called the *prior distribution*. When this is multiplied by the probability of the experimental data the result is proportional to a new distribution – the *posterior distribution*.

Posterior distribution  $\propto$  probability of observed values  $\times$  prior distribution

We might assume that the *prior distribution* of the quinine level  $\mu$  is (say) a normal one with a given mean and variance, or that it lies between 60 and 100 ppm with all values in that range being equally likely. The *probability of the observed values* is provided by a normal distribution with its mean and variance found from the experimental measurements. We can thus find the *posterior distribution* (a numerical example is given in *AMC Technical Briefs*

no. 14). A clear benefit of the Bayesian method is that it can be used iteratively to yield probabilities for any population parameter: additional measurements of the quinine level in the tonic water, perhaps using another method (*e.g.*, absorption spectrometry), would allow us to calculate a further posterior distribution. Equally clearly, the main problem is that the choice of prior distribution will affect the calculated posterior one. This can be a big problem if the observed values are poor in quality or few in number, but if the probability of the observed values is very well defined, the selection of the prior is less critical. Some examples of practical analytical applications will show the value of the Bayesian approach.<sup>1–4</sup>

## Radiocarbon dating in archaeology

This famous technique for dating organic artefacts originated in 1950 (radiocarbon years are still reported as *before present* [BP], present being defined as 1950!). It is based on the fact that the fixation of carbon dioxide by plants in photosynthesis incorporates the atmospheric level of <sup>14</sup>C at the time. This isotope, a  $\beta$ -emitter with a half-life of  $\sim 5730$  years, which is formed by cosmic ray neutrons interacting with <sup>14</sup>N, has an abundance of roughly one part per trillion relative to the total carbon level. Once a plant dies, or is consumed by humans or other animals, the decay of the <sup>14</sup>C can be used to estimate the age of a sample. (Levels of <sup>14</sup>C are now normally estimated using accelerator mass spectrometry, allowing milligram samples to be studied). But the age estimation calculation is complicated by the changes in atmospheric <sup>14</sup>C levels that have occurred over time, due to factors such as climate variation, volcanic eruptions, and fluctuations in the Earth's magnetosphere that influence cosmic ray activity. Such factors are accounted for by a *calibration* process, using agreed standards that allow the dating of artefacts up to about 50 000 years BP. In a recent example some cattle bones were dated initially to 4650 BP (*i.e.*, 2700 BC) with a 95% confidence interval of  $\pm 31$  years, but with the acknowledged possibility of a large bias. After calibration to correct the bias (which was not allowed for in the initial confidence interval) this was amended to  $3490 \pm 65$  BC. The calibration process thus shifted the



**amc technical briefs**  
www.rsc.org/amc

AMC Technical Briefs are produced by the Analytical Methods Committee, the technical committee of the Analytical Division of the Royal Society of Chemistry.

**About the AMC Statistical Subcommittee.**

Statistics is the framework supporting all aspects of data quality in analytical science. The aim of the Statistical Method Subcommittee is to optimise the use of statistical methods by analytical scientists, by:

- providing information about good basic statistical practice;
- investigating the benefits and limitations of both traditional and more modern statistical methods in analytical science;
- facilitating the use of newer statistical techniques by providing the necessary information in a readily usable form;
- applying statistical principles to undecided problems relating to quality in analytical data.

estimated age of the bones by some 800 years, reducing the bias considerably at the expense of poorer precision. Recently, however, Bayesian analysis has been used to reverse the loss of precision by combining the  $^{14}\text{C}$  result with other information: the prior beliefs are obtained from other evidence, such as the stratigraphy of an excavation site, the presence of coins, pottery or tools of known types datable by other methods, and so on. Combining this evidence with the  $^{14}\text{C}$  dating results gives a posterior estimate superior to the latter alone. This application area has been elegantly summarised in an illustrated lecture by T.S. Dye.<sup>1</sup>

**Bayesian methods in forensic science**

The value of the Bayesian approach in forensic science has often been expounded but its application has proved controversial in practice. This may be due to the difficulty of explaining the methodology to lay jurors and judges, and to understandable concerns that the liberty of any individual might depend solely on a probabilistic argument. (There is also concern that juries might give undue weight to evidence with a mathematical basis compared with other evidence of a more qualitative nature, though some research suggests that if anything the opposite is true). In one important case, a rape victim failed to identify her assailant at an identity parade, and agreed that the suspect did not fit her description. However he was convicted solely on the basis of DNA evidence, the likelihood ratio for a match between his DNA and the crime scene evidence being stated to be 200 million to 1. The defence challenged this figure, and an expert witness explained to the jurors how to use Bayes' Theorem to combine the DNA evidence with the probabilities of the other evidence, which was clearly more in the defendant's favour. After his conviction a re-trial was ordered but he was again convicted, although the Appeal judges were critical of the use of Bayes' Theorem in court. Nonetheless the application of Bayesian methods in forensic work is widespread. Judicial systems are clearly willing in principle to accept evidence based on probabilities, especially in the way that DNA profiling evidence is provided. The Bayesian method provides a flexible and important route to

combining and evaluating such probabilities, and is discussed in detail in two important texts.<sup>2,3</sup>

**Bayesian methods in astronomy**

Modern astronomical research is largely based on spectroscopic studies across the entire range of the electromagnetic spectrum. Many of the measurements face major problems of sensitivity, background signals *etc.* However there is a long history of the application of statistical methods to astronomy (*e.g.* Halley is regarded as a founder of actuarial science and Legendre applied least squares regression to compute comet orbits) and in recent years Bayesian methods have been applied with enthusiasm to many astronomical problems. A major factor has been the accumulation of large databases derived from telescopic survey projects: about a billion astronomical objects have been catalogued. These databases facilitate the formulation of realistic prior distributions to combine with observational data in the Bayesian manner to give better information on the object studied. Bayesian analysis has been applied to the possible detection of gravitational waves, the estimation of galaxy red shifts using photometric methods, astronomical imaging processing, planetary geology studies, and the hunt for extra-solar planets.

**References**

- 1 www.tsdy.com/research/ua/ua-bayesian-lecture.pdf.
- 2 C. G. G. Aitken, *Statistics and the Evaluation of Evidence for Forensic Scientists*, John Wiley and Sons, 1995.
- 3 D. Lucy, *Introduction to Statistics for Forensic Scientists*, John Wiley and sons, 2005.
- 4 C. E. Buck, W. G. Kavanagh, and C. Litton, *Bayesian Approach to Interpreting Archaeological Data*, Wiley-Blackwell, 1996.

*This Technical Brief has been compiled by Professor James N Miller on behalf of the Statistical Sub-Committee of the Analytical Methods Committee.*

You can help the AMC in the production of **AMC Technical Briefs** by posting comments, queries, and ideas for new topics on "MyRSC".

You can find this via [www.rsc.org/myrsc](http://www.rsc.org/myrsc). After you have logged in, go to the "Group" called "Analytical Methods Committee—Announcements and Discussions".

**CPD Certification** I certify that I have studied this document as a contribution to Continuing Professional Development.

Name.....  
Signature.....Date.....

Name of supervisor.....  
Signature.....Date.....