Supported by sanofi-aventis

What is Bayesian statistics?

John W Stevens BSc
Director, Centre for Bayesian Statistics in Health Economics, University of Sheffield

- Statistical inference concerns unknown **parameters** that describe certain population characteristics such as the true mean efficacy of a particular treatment. **Inferences** are made using data and a statistical model that links the data to the parameters.

- In frequentist statistics, parameters are fixed quantities, whereas in Bayesian statistics the true value of a parameter can be thought of as being a random variable to which we assign a probability distribution, known specifically as **prior information**.

- A Bayesian analysis synthesises both sample data, expressed as the **likelihood function**, and the prior distribution, which represents additional information that is available.

- The **posterior distribution** expresses what is known about a set of parameters based on both the sample data and prior information.

- In frequentist statistics, it is often necessary to rely on large-sample approximations by assuming **asymptomatic normality**. In contrast, Bayesian inferences can be computed exactly, even in highly complex situations.

For further titles in the series, visit:
www.whatisseries.co.uk

What is Bayesian statistics?

Statistical inference

Statistics involves the collection, analysis and interpretation of data for the purpose of making statements or inferences about one or more physical processes that give rise to the data. Statistical inference concerns unknown **parameters** that describe certain population characteristics such as the true mean efficacy of a treatment for cancer or the probability of experiencing an adverse event. Inferences are made using data and a **statistical model** that links the data to the parameters. The statistical model might be very simple such that, for example, the data are normally distributed with some unknown but true population mean, μ say, and known population variance, σ^2 say, so that our objective is to make inferences about μ through a sample of data. In practice, statistical models are much more complex than this. There are two main and distinct approaches to inference, namely frequentist and Bayesian statistics, although most people, when they first learn about statistics, usually begin with the frequentist approach (also known as the classical approach).

The nature of probability

The fundamental difference between the Bayesian and frequentist approaches to statistical inference is characterised in the way they interpret probability, represent the unknown parameters, acknowledge the use of prior information and make the final inferences.

The frequentist approach to statistics considers probability as a limiting long-run frequency. For example, the toss of a fair die an infinite number of times would result in the numbers one to six arising with equal frequency, and the probability of any particular event – 1/6 – is the long-run frequency relative to the number of tosses of the die. It is clear then that in frequentist statistics probability applies only to events that are (at least in principle) repeatable. In contrast, the Bayesian approach regards probability as a measure of the degree of personal belief about the value of an unknown parameter. Therefore, it is possible to ascribe probability to any event or proposition about which we are uncertain, including those that are not repeatable, such as the probability that the Bank of England Monetary Policy Committee will reduce interest rates at their next meeting.

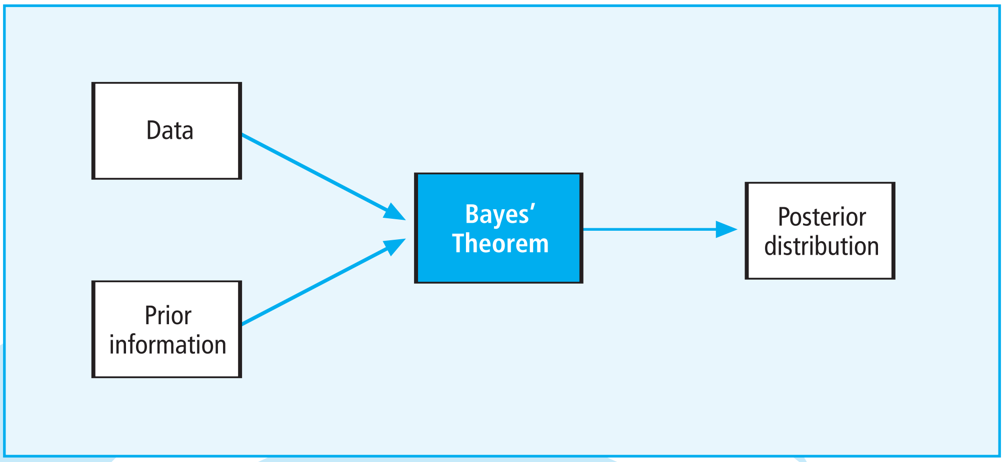
Prior information

A random variable can be thought of as a variable that takes on a set of values with specified probability. In frequentist statistics, parameters are not repeatable random things but are fixed (albeit unknown) quantities, which means that they cannot be considered as random variables. In contrast, in Bayesian statistics anything about which we are uncertain, including the true value of a parameter, can be thought of as being a random variable to which we can assign a probability distribution, known specifically as **prior information**.

A fundamental feature of the Bayesian approach to statistics is the use of prior information in addition to the (sample) data. A proper Bayesian analysis will always incorporate genuine prior information, which will help to strengthen inferences about the true value of the parameter and ensure that any relevant information about it is not wasted.

In general, the argument against the use of prior information is that it is intrinsically subjective and therefore has no place in science. Of particular concern is the fact that an unscrupulous analyst can concoct any desired result by the creative specification of prior distributions for the parameters in the model. However, the potential for manipulation is not unique to Bayesian statistics. The scientific community[1] and regulatory agencies[2] have developed sophisticated safeguards and guidance to avoid conscious or unconscious biases. An example is the use of double-blind,

Figure 1. The Bayesian method



randomised, controlled trials for the rigorous comparison of interventions. This, and similar requirements for the (statistical analysis) protocol to be established before a trial begins, are necessary to obviate the potential for manipulation or bias that already exists in the use of frequentist statistics. A serious Bayesian statistician will spend reasonable, and sometimes substantial, effort to develop probability distributions that genuinely represent prior beliefs. The process should be formal and transparent so that the basis for the resulting probability distribution is understandable and justifiable.

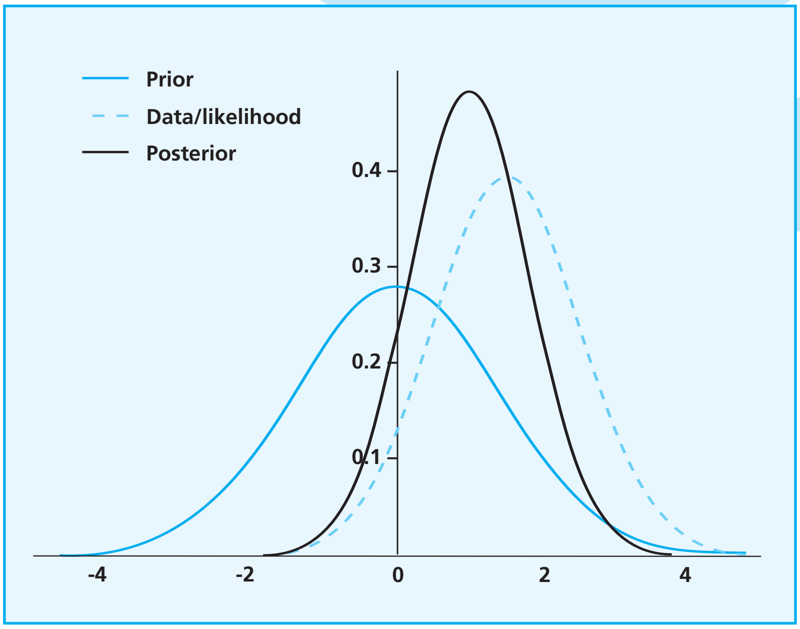
The Bayesian method

A Bayesian analysis synthesises two sources of information about the unknown parameters of interest. The first of these is the sample data, expressed formally by the **likelihood function**. The second is the prior distribution, which represents additional (external) information that is available to the investigator (Figure 1). Whereas the likelihood function is also fundamental to frequentist inference, the prior distribution is used only in the Bayesian approach. If we represent the data by the symbol D and denote the set of unknown parameters by θ, then the likelihood function is $f(D|\theta)$; the probability of observing the data D being conditional on the values of the parameter θ. If we further represent the prior distribution for θ as $\pi(\theta)$, giving the probability that θ takes any particular value based on whatever additional information might be available to the investigator, then, with the application of Bayes's theorem,[3] an elementary result about conditional probability named after the Reverend Thomas Bayes, we synthesise these two sources of information through the equation:

Equation 1. $p(\theta|D) \propto f(D|\theta)\pi(\theta)$

The proportionality symbol $\propto$ expresses the fact that the product of the likelihood function and the prior distribution on the right hand side of Equation 1 must be scaled to integrate to one over the range of plausible θ values for it to be a proper probability distribution. The scaled product, $p(\theta|D)$, is then called the **posterior** distribution for θ (given the data), and expresses what is now known about θ based on both the sample data and prior information (Figure 2).

Figure 2. Example of a triplot



The posterior distribution for θ is a weighted compromise between the prior information and the sample data. In particular, if for some value of θ the likelihood in the right-hand side of Equation 1 is small, so that the data suggests that this value of θ is implausible, then the posterior distribution will also give small probability to this θ value. Similarly, if for some value of θ the prior distribution in the right-hand side of Equation 1 is small, so that the prior information suggests that this value of θ is implausible, then, again, the posterior distribution will also give small probability to this θ value. In general, the posterior probability will be high for some θ only when *both* information sources support that value. The simple and intuitive nature of Bayes' theorem as a mechanism for synthesising information and updating personal beliefs about unknown parameters is an attractive feature of the Bayesian method.

The nature of inference

Classical inference is usually based on unbiased estimators defined to have expected value equal to the parameter being estimated, a significance test of some null hypothesis and confidence intervals. As with frequentist probability, such inferences are justified through long-run repetition of the data. However, they do not – although it may appear that they do – make direct statements about parameters. For example, consider the statement, 'we reject the null hypothesis at the 5% level of significance'. This means that if we were to repeat the experiment a large number of times then in 5% of occasions *when the null hypothesis is true* we will reject it. However, nothing is stated about this particular occasion. In contrast, the Bayesian approach allows direct probability statements to be made about the truth of the null hypothesis on the basis of *this* sample of data (as a personal degree of belief). In fact, many other probabilistic statements about parameters can be made from the posterior distribution and it is of particular value to simply plot the (posterior) distribution of the parameter of interest given the data (Figure 2).

Advantages of the Bayesian approach

The arguments that are made in favour of the Bayesian approach are that it offers more intuitive and meaningful inferences, that it gives the ability to tackle more complex problems and that it allows the incorporation of prior information in addition to the data.

The Bayesian approach enables direct probability statements to be made about parameters of interest, whereas frequentist methods make indirect inferences by considering the data and more extreme but unobserved situations conditional on the null hypothesis being true; that is, p-values. Also, classical 100(1–α)% confidence intervals do not – although they appear to – provide an interval estimate containing the true value of the parameter with probability 100(1–α)%. To understand this it is important to recognise that once a confidence interval has been generated it either does or does not contain the true value of the parameter, in which case we say that the confidence interval has 100(1–α)% **coefficient**. In contrast, the Bayesian approach allows the construction of interval estimates, known as credible intervals, which do have a probabilistic interpretation.

Statistical modelling can often generate quite complex problems and these can quickly become difficult to deal with or to construct exact test statistics from using a frequentist approach. Often it is necessary to rely on large-sample approximations by assuming **asymptotic normality**. In contrast, Bayesian inferences can be computed exactly, even in highly complex situations. A simple example is the estimation of the probability of survival beyond one year for patients given a new treatment for cancer. Suppose that with standard treatment 40% of patients survive beyond one year. Prior information suggests that the new treatment improves survival. An expert investigator gives a prior estimate of 45% and expresses her uncertainty as a standard deviation of 7%. We define the prior distribution to be Beta(22.28, 27.23) to give the required mean and standard deviation. After treating 70 patients, 34 (48.6%) survive. The posterior distribution is Beta(56.28, 63.23), which has a mean of

47.1%, which is a compromise between the prior estimate and the sample estimate. Further examples of varying complexity can be found on The BUGS Project website.[4]

The use of prior information in addition to sample data is fundamental to the Bayesian approach. Prior information of some degree almost always exists and can make important contributions to strengthen inferences about unknown parameters and/or to reduce sample sizes.

The main 'controversies'

To avoid having to use informative prior distributions while still being able to use Bayesian tools, some authors have suggested using non-informative prior distributions to represent a state of prior ignorance.[5] In so doing the analyst obtains some of the benefits of a Bayesian approach, particularly that the results are presented in the intuitive way in which one would like to make inferences. Another way to represent prior information is to specify sceptical prior distributions for the parameters in the model. In this context, the prior distribution is specified in such a way that it automatically favours the standard treatment. Such a proposal is tempting, particularly in a regulatory framework where rigorous standards and safeguards are demanded. However, both ideas suffer from serious objections, and both fail to exploit the full potential of the Bayesian approach. In addition, there is no unique way to implement either idea, and hence subjectivity is not removed. For all but the simplest of situations, there is no general agreement regarding non-informative prior distributions, and the posterior answers depend on the (subjective) way in which the model is specified. There can be even less agreement over what constitutes a 'sceptical' prior distribution.

Of course, the prior distribution in a Bayesian analysis is not the only place where subjective judgements are in danger of entering the analysis. Any statistical model, whether formulated for a frequentist or Bayesian analysis, is a matter of subjective judgement, and it is commonplace that different statisticians make different choices. Furthermore, the choice of which estimator, significance test or confidence interval to employ is a subjective matter in frequentist statistics. Such subjectivity is reduced but not removed by the requirement that the basic form of the analysis should be prespecified. There is no such choice to make in Bayesian statistics, as once the posterior distribution has been obtained there is a unique (objective) answer to any properly specified question about the parameters.

Application of the Bayesian approach

In principle, the posterior distribution is always available, although in realistically complex problems it cannot be represented analytically. This presented a barrier to the implementation of the Bayesian approach until the development of numerical methods and powerful computers during the late 20th century. Now, posterior distributions can be constructed for highly complex problems using Markov chain Monte Carlo (MCMC) simulation. MCMC involves simulating a sample from the (joint) posterior distribution of the unknown parameters using one of three main algorithms: the Metropolis–Hastings algorithm, Gibbs sampling and slice sampling. With a sufficiently large sample, we are able to numerically generate the whole distribution from which we can make any inferences of interest. (Suggested further reading on MCMC can be found at the end of this document.) WinBUGS is a software package that implements MCMC algorithms without the analyst having to write their own sampling algorithms and is able to analyse highly complex problems.[6]

References

1. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001; **357:** 1191–1194.
2. European Medicines Agency. *ICH Topic E 9: Statistical Principles for Clinical Trials (CPMP/ICH/363/96)*. www.emea.europa.eu/pdfs/human/ich/036396en.pdf (last accessed 22 January 2009)
3. Senn S. *Dicing with Death: Chance, Risk and Health*. Cambridge: Cambridge University Press, 2003.
4. *The BUGS Project*. www.mrc-bsu.cam.ac.uk/bugs/ (last accessed 22 January 2009)
5. Briggs AH. A Bayesian approach to stochastic cost-effectiveness analysis. *Health Econ* 1999; **8:** 257–261.
6. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing* 2000; **10:** 325–337.

What is Bayesian statistics?

Further reading

1. Bolstad WM. *Introduction to Bayesian Statistics*. Chichester: Wiley, 2004.
2. Gilks WR, Richardson S, Spiegelhalter DJ. *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall, 1996.
3. Lee PM. *Bayesian Statistics: An Introduction*, 3rd edn. London: Arnold, 2004.
4. O'Hagan A, Forster J. *Kendall's Advanced Theory of Statistics: Volume 2B: Bayesian Inference*, 2nd edn. Oxford: Oxford University Press, 2004.
5. Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Chichester: Wiley, 2004.

Published by Hayward Medical Communications, a division of Hayward Group Ltd.

Supported by *sanofi-aventis*