amc technical brief

Analytical Methods Committee No.14. Oct 2003

A glimpse into Bayesian statistics

Many analytical chemists find the logic of hypothesis tests and confidence intervals hard to follow. What looks like a probability statement about a true concentration is in fact an assertion about random intervals, involving data we did not observe but might have. There is another way. Bayesian statistics allows, indeed insists on, probabilities for hypotheses.

An example

Consider the simple example of analysing a material to test it against a specification. Suppose there is an upper limit $c_L = 10$ units for an acceptable concentration of an impurity in the material, and by analysis we obtain a single measurement $c_m = 10.7$ of the concentration in this particular sample. The analytical method is unbiased and has known precision (standard deviation) 0.4 units. Thus the variance of the measurement is $v_m = 0.4^2 = 0.16$. What is the strength of the evidence that the true concentration θ in this sample exceeds the allowable limit?

A standard statistical treatment

This argues as follows. If the true value θ = 10, then the measurement is drawn from a normal distribution with mean 10 and standard deviation 0.4. The probability that such a measurement is 10.7 or greater is the same as the probability that an observation from the standard normal distribution exceeds (10.7 – 10)/0.4 = 1.75, which is 0.04 from tables. If θ < 10, this probability will be even smaller. The small probability for the observed (or more extreme) data under the hypothesis θ ≤ 10 is taken as evidence against the hypothesis. Either we quote 0.04 as a p-value measuring the strength of this evidence or, noting that 0.04 is less than the magic 0.05, announce that the hypothesis has been rejected at the 5% level. All this should seem fairly familiar. What may also be familiar is the common practice of interpreting the p-value as though it is the probability that the hypothesis is true. It is not. It is the probability of observing particular data given that the hypothesis is true. If we want to attach probabilities to hypotheses then we have to work in a Bayesian framework.

A Bayesian analysis

The Bayesian approach requires us to quantify our beliefs about the true value θ in the form of a probability distribution. These beliefs will change when we see the result of the measurement, and the main tool in Bayesian statistics is the recipe

posterior ∝ likelihood × prior

for updating beliefs in the light of new evidence. The workings of this formula are most easily followed in the case when θ may only take one of a finite set of values, $\theta_1, \theta_2, \ldots, \theta_k$, and 'prior' attaches a probability to each θ_i. This prior distribution expresses our beliefs about θ before observing the data. The likelihood, which also has a value for each θ_i, is the probability of observing the data given $\theta = \theta_i$. Multiplying the prior probability and the likelihood for each θ_i and then scaling so that the resulting numbers add to one over the k values of θ gives us a new set of probabilities, the posterior distribution, which expresses our updated beliefs about θ. When, as in our example, it is more natural to think of θ as continuous rather than discrete, prior and posterior beliefs are represented by probability density functions (pdfs), and the likelihood becomes a continuous function of θ, but the idea is essentially the same.

Sometimes the updating has to be done numerically, just as described above, possibly after discretising a continuous distribution. Sometimes, if the prior distribution and likelihood have compatible mathematical forms, it can be done algebraically.

Suppose that in our example our prior beliefs about θ may be described by a normal distribution with mean m_p and variance v_p. This combines with the normal likelihood to give a normal posterior distribution with mean

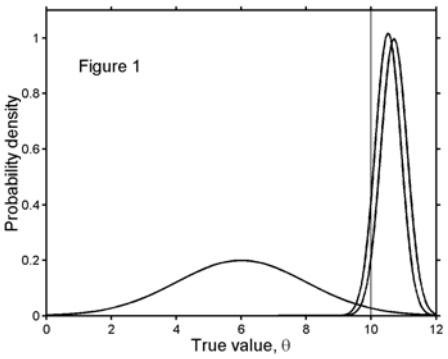
$$m = (v_m^{-1} + v_p^{-1})^{-1}(v_m^{-1}c_m + v_p^{-1}m_p) ,$$

a linear combination of the prior mean m_p and the measurement c_m with weights inversely proportional to the respective variances, and variance

$$v = (v_m^{-1} + v_p^{-1})^{-1} .$$

An 'informative prior' distribution

To get any further we need to specify the values of m_p and v_p, the prior mean and variance. If the sample of material under test comes from a manufacturing process that we have experience of, we may be able to use this experience to specify, for example, a prior mean of $m_p = 6$ and variance of $v_p = 4$. The corresponding distribution is shown in Figure 1, where it is the one cantered on 6 and spreading across the whole range.



What we are saying here is that before taking account of the measurement we are prepared to regard the material under test as a randomly chosen sample from a process that produces material with an average impurity concentration of 6 units and a spread such that about 2.5% of the material will exceed the allowable limit of 10 units.

Plugging these numbers and the values $c_m = 10.7$, $v_m = 0.16$ into the formulae above gives us a mean of

$m = (0.16^{-1} + 4^{-1})^{-1}(0.16^{-1}\times 10.7 + 4^{-1}\times 6)$
$= (6.25 + 0.25)^{-1}(6.25\times 10.7 + 0.25\times 6)$
$= 10.52$

and a variance of

$v = (0.16^{-1} + 4^{-1})^{-1} = 0.154$,

which corresponds to a standard deviation of 0.392. Both the likelihood and the posterior distribution are shown in Figure 1.

They are the two sharp peaks on the right, with the likelihood the furthest right, and the posterior shifted slightly to the left, because of the effect of the prior.

The posterior distribution for θ, normal with mean 10.52 and standard deviation 0.392, can be used to make probability statements about θ. For example, the probability that $\theta < 10$ is equal to the probability that a standard normal is less than $(10 - 10.52)/0.392 = -1.33$, which is 0.09 from tables. This is a little greater than the p-value of 0.04, because we have taken into account the extra evidence that material failing the specification is relatively unusual, but more importantly it is qualitatively different. It is a probability statement about θ, not a probability statement about the data asking to be mistaken for one about θ.

A 'reference prior' distribution

In some, though not all, situations it is possible to reproduce the same numerical results as in the standard analysis, whilst still keeping the Bayesian interpretation. If in the example we let the prior variance v_p become very large, the posterior mean will approach c_m and the posterior variance will approach v_m. In Figure 1, the pdf of the prior distribution becomes more and more spread, until it is effectively flat, and the likelihood and posterior distributions coincide. Then the posterior distribution of θ is normal with mean c_m and variance 0.16, and for $c_m = 10.7$ the probability that $\theta < 10$ is 0.04, the same as the standard p-value. To get this posterior distribution in a formal way we may take the prior distribution as uniform over all values of θ. It is hard to imagine any real case in which we believe that *all* values of θ are equally likely, and a uniform distribution over an infinite range cannot be normalised to integrate to 1, so this is a mathematical convenience rather than a proper expression of belief. However the resulting posterior distribution is commonly taken as the appropriate one for 'vague prior beliefs'.

Opportunities

In this example and in many others, the Bayesian approach has two advantages. It provides a formal and coherent way to take account of the other information that will often be available, and it provides probabilities for hypotheses in a straightforward way. The second aspect is particularly important if the purpose of the analysis is to inform a decision, e.g. to accept or reject a batch of material. The Bayesian approach extends naturally to a theory for optimal decision making[1] in which the probability distributions for uncertain quantities are combined with costs for outcomes.

Problems

The role of the prior is open to criticism. Two analysts with different prior distributions may reach different conclusions from the same data. This is logical: they are basing their inferences on different information. However it can give rise to argument in legal or regulatory contexts. Whose prior distribution is appropriate, that of the analyst or the regulator? One can, of course, use both.

Further reading

There is at present no book on Bayesian Statistics for the Analytical Chemist and there are relatively few analytical applications in the literature. There is a book on archaeological applications that has a little chemistry in it,[2] and at least one book on forensic applications.[3] There are general introductory texts, of which the book by Berry[4] is one of the most accessible and that by Lee[5] a little more mathematical. The more adventurous might like to try O'Hagan's First Bayes software.[6]

References

1. Lindley, D.V. *Making Decisions*. Wiley: London (1985).
2. Buck, C.E., Cavanagh, W.G and Litton, C.D. *Bayesian Approach to Interpreting Archaeological Data.* Wiley: Chichester (1996).
3. Aitken, C.G.G. *Statistics and the Evaluation of Evidence for Forensic Scientists.* Wiley: Chichester (1995).
4. Berry, D.A. *Statistics, A Bayesian Perspective.* Duxbury: London (1996).
5. Lee, P.M. *Bayesian Statistics: an Introduction.* Arnold: London (1997).
6. www://sunsite.univie.ac.at/statlib/DOS/general/ first-bayes/

About the AMC: The Analytical Methods Committee handles matters that are of technical importance to the Analytical Division of the RSC and the analytical community in general. The aim of the AMC is "to participate in national and international efforts to establish a comprehensive framework for appropriate quality in chemical measurement". It achieves this aim through the activities of its expert subcommittees, which handle:

- the development, revision and promulgation of validated, standardised and official methods of analysis;
- the development and establishment of suitable performance criteria for methods and instruments;
- the use and development of appropriate statistical methods;
- the identification and promulgation of best analytical practice, including aspects relating to sampling, equipment, instrumentation and materials;
- the generation of validated compositional data of natural products for interpretative purposes.

About the AMC Statistical Subcommittee. The aim of the Statistical Methods Sub-committee is to optimise the usage 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.

Other AMC Technical Briefs can be found on: www.rsc.org/lap/rsccom/amc/amc_index.htm