What is the use of evidence?

Brian Newman

A little bit of Bayes and a little bit of mathematics, but why in Radical Statistics? Because I find the outcome of my analysis politically unacceptable and hope the reader may be able to find a flaw in my argument. The conclusion appears to be that there is no point in testing a new drug for safety, and that either a prisoner is always innocent, by presumption, or that the policeofficer's hunch is as good as a detailed examination of the evidence. I emphasise that I and other authors quoted are only considering the statistical model, and I am not commenting on any particular case or drug.

The Safety of a Contraceptive and the Guilty Party

This analysis originated while assessing the safety of an oral contraceptive, using a sample of approximately 100 women (1). A very large number of different effects, were considered, e.g. change in blood pressures, serum lipids, metals, clotting factors etc., For the contraceptive to be 'safe' all these effects have to be within 'acceptable' limits. I do not have to be precise about what is acceptable at this stage, the important feature is that each and every effect has to satisfy some criteria. Tests are applied to the sample data to provide evidence of acceptability. The conclusion must be based on these tests and any relevant prior information about the contraceptive. A similar problem occurs in assessing certain types of evidence in law suits. In a civil suit the plaintiffs contention may depend on two or more component issues both of which must be established for the case to succeed, for example the plaintiff may need to establish that they were injured by the defendant and that the defendant behaved in a negligent manner. In a murder case it may be necessary to establish the weapon used and the prisoners access to both victim and weapon. Throughout this discussion we will only consider independent issues or features.

The issues and an analysis of the probabilities in the civil legal case has been examined by Dawid(2) in response to Cohen(3), however Dawid and Cohen appear to have limited their analysis to a specific set of evidence, which is most likely to occur if the plaintiff has chosen to present only positive evidence and suppressed negative evidence.

All evidence has some degree of unreliability. Cohen (3) gave the example of two independent witnesses, each 70% reliable, who gave evidence supporting the case. The probability that both were correct, he said, was 0.7 x 0.7 = 49%, which is just below the 50% level of proof required to establish a civil liability in English law. Cohen dismisses the use of probability in assessing legal evidence, on the grounds that, both witnesses would have made the case individually, yet their evidence was not enough to make the case together. Dawid examines Cohen's example with a Bayesian approach. He considers n issues all of which must support the plaintiff for the case for the plaintiff to be made. Each

issue has a witness testifying in support of the plaintiff. Dawid defines more precisely the meaning of 70% reliable, then he shows that each new witness giving evidence increases the probability of the plaintiff's case being true. He shows that the probability of the case being true as *n* tends to infinity equals the prior probability of the case being true before the evidence, raised to the power 3/7.

The Whole Truth

With an infinite number of unreliable witnesses, some of them must appear to support the defence. They cannot all give evidence in the same direction unless the evidence of some witnesses has been suppressed. There was a tradition that each side in a legal argument would only present favourable evidence. In drug safety studies, it is unusual to deliberately suppress evidence, unless you regard correction for multiple siginificance tests as a form of supression. The mathematics below is Dawid's analysis with an extension to some witnesses giving evidence for the plaintiff and some for the defendant. It is this extension which leads to the strange conclusion mentioned above.

The analysis is presented in the form of a drug safety study, the feature SAFE, corresponds to GUILTY or the case for the PLAINTIFF in the legal analogy.

Assume, initially, that the safety of a drug depends on one variable only, e.g. change in cholesterol. Let A represent the situation that any changes in cholesterol caused by the drug is within acceptable limits and A the situation that it is not within acceptable limits. A or A represents the true, but unknown situation. The data in the clinical trial are examined, and declared to support A or to support A. This evidence is denoted by a or a, respectively. For example, A could represent the hypothesis that the mean change in cholesterol induced by the drug was zero, the data are examined using a significance test and a represents a significant difference from zero mean.

Bayes Theorem gives:

$$\frac{Pr(A \mid a)}{Pr(\overline{A} \mid a)} = \frac{Pr(A)}{Pr(\overline{A})} \times \frac{Pr(a \mid A)}{Pr(a \mid \overline{A})}$$

Substitution of λ for the last ratio in this equation and p for the prior probability that the effect is acceptable gives:

$$\mathbf{Pr(A \mid a) = \lambda p/(1 - p + \lambda p)}$$

λ represents the reliability of the test procedure.
Example: If I had examined the data with a significance test which had type I error of α and type II error of β, then λ = (1 - α)/ β: a typical value of λ used in small scale clinical research might be 0.95/0.1 = 9.5. If ethical considerations required that I was 99% certain that the drug was safe before starting a human study, then after getting a non-significant

result I would be 99.8938% certain that the drug is safe, (I need this accuracy to demonstrate the point using realistic values of p and λ).

I now consider another independent variable e.g. change in vitamin B, with similar notation to cholesterol substituting B for A. I also assume vitamin B has the same values of λ and p as cholesterol. For the drug to be safe I require that both variables are acceptable i.e A&B. If the evidence supports acceptance for both variables i.e. a & b then:

$$Pr(\text{Drug is safe} \mid \text{evidence}) = Pr(A \,\&\, B \mid a \,\&\, b) = \{ \lambda p/(1 - p + \lambda p) \}^2$$

The prior probability of cholesterol being acceptable i.e. p, and the prior probability of vitamin B being acceptable i.e. p, must be multiplied together to give the prior probability of the drug being safe.
Example: λ = 9.5, and if the prior probability of the drug being safe is 99%, then p^2 must be 0.99. If both a and b yield non-significant results then the probability that the drug is safe after the evidence is 99.8940%. In general this will be greater than the computations for one variable.

If the prior probability that the drug is safe is π and we have an infinite number of independent variables all of which give non-significant results then Dawid has shows that in the limit :

$Pr(\text{Drug is safe} \mid \text{evidence}) = \pi^{1/\lambda}$ In the example this is 99.8943%.

However, in an infinite number of independent but unreliable assessments then, if the drug is safe, I can expect, on average, α of them to give significant results, each of which appears to indicate that the change produced by the drug is unacceptable and the drug should be classed as unsafe.

In the case of a drug which depends on a single test A and the evidence, a, points to an unacceptable change, then, using Bayes theorem:

$$Pr(A \mid) = \mu p/(1 - p + \mu p) \text{ where } \mu = \alpha/(1 - \beta).$$

If the true situation is that the drug is safe, we examine a large number *n* of different effects and, as expected, α*n* of them appear unacceptable then in the limit:

$$Pr(\text{Drug is safe} \mid \text{evidence with } \alpha \text{ unacceptable}) = \pi^{(1-\alpha)/\lambda}\, \pi^{\alpha/\mu} = \pi$$

i.e the probability returns to the prior belief !

If the drug has an unacceptable effect on a fixed number of features, then as we increase the number of features examined , any unacceptable results will be lost in the α*n* produced by the unreliability of the testing procedure and the probability will still tend to π.

Thus on average we can expect that a detailed examination of all the component effects of a

drug will leave us back with our starting probability that it was safe, so why bother ?

Before the Evidence

Dawid's argument depends on a fixed prior probability that the case as a whole should support the plaintiff, while the prior probability for a specific issue supporting the plaintiff depends on the number of issues. T.Fearn(4) has shown that assuming Dawid's model for the prior distribution, then the prior probability distribution of the number of issues supporting the defence is Poisson with mean $\ln(1/\pi)$ for a large number of issues, i.e. is independant of the number of issues. In the drug safety example with a prior of 99% that the drug is safe then this implies a prior belief that not more than one feature will prove unsafe. In a civil lawl case with a prior of 50%, to represent even handedness between the parties, then division into equal prior probabilities for each issue, implies a prior belief that less than 4 or 5 of the issues support the defence.

If the prior for each issue is regarded as constant, the case can be made virtually impossible to prove, by deliberately subdividing each issue into a large number of sub-issues. Alternatively issues could be subdivided because the tests of sub-issues are more reliable than the test of the main issue. It seems likely that there will be a critical increase in reliability at which it will pay to subdivide the issues, but this has not yet been computed.

References

1. M.Nichols et al. (1993), Effect of four combined oral contraceptives on blood pressure in the pill-free interval. Contraception 47:367-376. Other factors may be reported probably with J.Guillebaud in the list of authors. 2. A.P.Dawid. (1987), The difficulty about conjunction. The Statistician 36:91-97 .
3. L.J.Cohen. (1977) The provable and the probable. Oxford, Clarendon Press.
4. T.Fearn (1993) personal communication.

Interested in writing book reviews?

The following books are available for review:

Analyzing Qualitative Data Edited by Alan Bryman & Robert G Burgess, 232pp
Health and Development Edited by David R Phillips & Yola Verhasselt, 331pp
Population Matters - The Local Dimension Edited by Tony Champion, 189pp
Spatial Analysis & GIS Edited by Stewart Fotheringham & Peter Rogerson, 281pp
Modelling Biological Populations in Space & Time Author Eric Renshaw, 401pp
Error Propagation in Quantitative Spatial Modelling - Applications in Geographical Information Systems Author Gerard B M Heuvelink, 151pp

Books allocated on a first come first served basis. Reviews should be 500-700 words, unpaid, but you get to keep the book! Please contact the Book Review Editor Nicki Dennis, Tel: 071 873 6360 - full address on the inside front cover.