even when a small proportion of contamination by the wider component occurs. The contaminated normal is used to exemplify a 'good' distribution affected by outliers rather than as a 'typical' distribution with long tails.

References

DAVID, H. A. (1968). Gini's mean difference rediscovered. *Biometrika* **55**, 573–5.
DOWNTON, F. (1966). Linear estimates with polynomial coefficients. *Biometrika* **53**, 129–41.
HARTER, H. L. (1961). Expected values of normal order statistics. *Biometrika* **48**, 151–65.
HARTER, H. L. & MOORE, A. H. (1966). Iterative maximum-likelihood estimation of the parameters of normal populations from singly and doubly censored samples. *Biometrika* **53**, 205–13.
TEICHROEW, D. (1956). Tables of expected values of order statistics and products of order statistics from samples of size 20 and less from the normal distribution. *Ann. Math. Statist.* **27**, 410–26.

[*Received January* 1978. *Revised May* 1978]

Biometrika (1978), **65**, 3, *pp.* 646–8

On a problem in forensic science

BY ALLAN SEHEULT
Department of Mathematics, University of Durham

Summary

A Neyman–Pearson test of identification in forensic science is shown to reflect all the properties of a Bayes factor approach presented by D. V. Lindley in a previous volume of this journal.

Some key words: Bayes factor; Forensic science; Hypothesis testing; Identification.

Recently Lindley (1977) discussed a problem in forensic science of deciding whether two sets of fragments have come from a common source. A typical situation is where measurements are made of the refractive indices of pieces of window glass at the scene of a crime and of fragments of glass found on a suspect's clothing. It is argued that evidence of identity should depend not only on the measurements but also on the distribution of refractive indices of window glass, an additional objective source of information. Although Lindley considers the more realistic case of a nonnormal distribution for the refractive indices, he shows that the essential features of his solution are embodied in the following simplified situation. A measurement x, with normal error having known standard deviation σ, is made on the unknown refractive index θ_1 of the glass at the scene of the crime. Another measurement y, made on the glass found on the suspect, is also assumed to be normal but with mean θ_2 and the same standard deviation as x. The refractive indices θ are assumed to be normally distributed with known mean μ and known standard deviation τ. If I is the event that the two pieces of glass come from the same source ($\theta_1 = \theta_2$) and $\bar{I}$ the contrary event, Lindley suggests that the odds on identity should be multiplied by the factor

$$p(x, y \mid I)/p(x, y \mid \bar{I}). \tag{1}$$

In this special case, it follows from Lindley's equation (6) that the factor is

$$\frac{1+\lambda^2}{\lambda(2+\lambda^2)^{\frac{1}{2}}} \exp\left\{-\frac{1}{2(1+\lambda^2)}(u^2-v^2)\right\}, \tag{2}$$

where

$$\lambda = \frac{\sigma}{\tau}, \quad u = \frac{x-y}{\sigma\sqrt{2}}, \quad v = \frac{z-\mu}{\tau(1+\frac{1}{2}\lambda^2)^{\frac{1}{2}}}, \quad z = \tfrac{1}{2}(x+y).$$

Note that in the derivation of (2) it is not necessary to assume that λ^2 is negligible, although this is typically the case. The result is exact and remains so for arbitrary but equal numbers of measurements for both sets of fragments, when x and y are interpreted as the two sample means, σ^2 is replaced by σ^2/n, and n is the common sample size. If λ^2 is assumed to be negligible, then (2) simplifies to

$$\frac{1}{\lambda\sqrt{2}} e^{-st}, \tag{3}$$

where $s = (u+v)/\sqrt{2}$ and $t = (u-v)/\sqrt{2}$.

Lindley then compares the factor in (3) with the naive significance test based only on u, that ignores the distribution of refractive indices θ, and also with a solution suggested by Evett (1977) and Parker (1966) which, at least, implicitly acknowledges the distribution of θ. He argues convincingly that the Bayes factor approach is superior to both significance tests in that it successfully balances the various features of the situation.

The main purpose of this note is to point out that, regardless of any philosophical convictions, one can always use the objective information contained in the distribution of refractive indices. In the case of the significance tester, or hypothesis tester, I and $\bar{I}$ each become simple hypotheses when the joint distribution of x and y is compounded with that of θ, and the Neyman–Pearson lemma is applicable. It is shown below that the most powerful test also successfully balances the various features of the problem and does not conflict with the inferences using the Bayes factor. The likelihood ratio is precisely the factor in (1), so that if λ^2 is negligible and (3) is appropriate large values of st lead us to reject I in favour of $\bar{I}$. But under the identity hypothesis I, s and t are independent $N(0, 1)$ variates and thus Table 3 in Lindley's paper gives the required significance levels. Consider, for example, the case suggested by Evett and discussed by Lindley in the third paragraph of § 2. Here, $\lambda = 0{\cdot}01$, $v = 0$ and $u = 2$, so that $st = 2$ and, from Table 3, the observed significance level is found to be about 3%. Evett's test concurs with this result, rejecting the identity hypothesis at the 5% level, whereas the Bayes factor is about 10. In the case of Evett's solution Lindley argues:

What the test fails to take into account is the extraordinary coincidence of X and Y being so close together were the pieces of glass truly different. If $\sigma = 1$ and $\tau = 100$ the difference is only 2·83 in a distribution of $X-Y$ of variance $2(\sigma^2+\tau^2)$, or standard deviation about 145. Within 0·020 of the mean of a $N(0, 1)$ variable the probability is only about 0·008, much smaller than the 0·05 of the significance level. The argument we have given balances the various features of a situation, some pulling some way, some another, combining them all into a single factor. This the significance level, with its concentration on the null hypothesis, here that of identity, does not do.

To see that the Neyman–Pearson test is not in conflict with the Bayes factor it is instructive to consider in more detail the joint distribution of s and t under both hypotheses. Under I, s and t are independent $N(0, 1)$ variates, whereas under $\bar{I}$ the distribution is markedly different, and this is best appreciated when λ^2 and higher powers are ignored. To this order of approximation, s and t are identically distributed normal variates with zero means but having very large variance, $\frac{1}{2}\lambda^2$. Moreover, the correlation coefficient between s and t is unity! Thus, to all intents and purposes, the joint distribution under $\bar{I}$ is a singular uniform distribution concentrated on the line $s = t$. It follows that the test statistic st is approximated by s^2 which is such that $2\lambda^2 s^2$ has a chi-squared distribution on one degree of freedom. Consider again the Neyman–Pearson test applied to Evett's example, $st = 2$ and $\lambda = 0{\cdot}01$. At the observed significance level of 0·031 the power is one minus the chance that a standard normal variate is between plus and minus 0·02, a probability of 0·984. This result indicates that not only is this test most powerful but very powerful. Suppose, for example, that it is required that the two types of errors have equal probabilities. The critical level corresponding to this requirement is about 1·8%, so that in the above example $st = 2$ is not such clear evidence against identity. Of course, within the context of criminal law an error of the second kind, convicting an innocent person, is considered more serious than the error of setting a criminal free. Lehmann (1958) has given a rule of thumb for choosing critical levels in such cases. For example, if, somewhat conveniently, it is required that $\alpha = 14\beta$, where α and β are the errors of the first and second kind respectively, then $\alpha = 0{\cdot}14$ and $\beta = 0{\cdot}01$ when $\lambda = 0{\cdot}01$. Both values are unacceptable in law but are, once again, in agreement with the analysis in § 3 of Lindley's paper.

Finally, consider what happens when λ is halved to 0·005 and $st = 2$ as before. The Bayes factor is doubled to 20 and the critical level to achieve equal error probabilities is reduced to about 0·9%. The power of the test is thus increased, which is intuitively reasonable since the distributions under the two hypotheses become even more widely separated, the standard deviation of the distribution under $\bar{I}$ being doubled.

The author would like to thank Professor Lindley for his helpful and thought-provoking correspondence on several aspects of this note.

References

Evett, I. W. (1977). The interpretation of refractive index measurements. *Forensic Sci.* **9**, 209–17.
Lehmann, E. L. (1958). Significance level and power. *Ann. Math. Statist.* **29**, 1167–76.
Lindley, D. V. (1977). A problem in forensic science. *Biometrika* **64**, 207–13.
Parker, J. (1966). A statistical treatment of identification problems. *J. For. Sci. Soc.* **6**, 33–9.

[*Received January* 1978. *Revised May* 1978]

Biometrika (1978), **65**, 3, *pp*. 648–9

A remark on likelihood, Bayes and affinity for a preconceived value

By J. B. COPAS

Department of Mathematics, University of Salford

and J. M. DICKEY

Department of Statistics, University College of Wales, Aberystwyth

Summary

As judged by large sample efficiency, one can make use of a preconceived value in a considerably less modest way than Bayes's theorem permits, provided one respects the likelihood function.

Some key words: Bayesian theory; Likelihood; Point estimate; Prior knowledge.

Let $\bar{X}$ be the average of a random sample of size n from a normal distribution with unknown mean μ and known variance σ^2. We consider large n.