Consider how we might change p_i and p_j in order to increase the value of this expression, however, keeping $p_i + p_j$ constant so that $\Sigma p_i = 1$. If the expression in square brackets is positive, the maximum occurs when $p_i = p_j$. If it is negative, then the maximum occurs when one of p_i and p_j is 0. Thus the expression is maximised when all the p_i are equal and $p_i = 1/n$.

Substituting these values into equation (1) gives

$$P^* = \frac{1}{n} \cdot \frac{n-1}{n} \cdots \frac{n-m}{n}. \tag{4}$$

It is simple to check that

$$-\log\left(1-\frac{1}{n}\right) - \cdots - \log\left(1-\frac{m}{n}\right) \geq \frac{m(m+1)}{2n}. \tag{5}$$

Thus,

$$\max P^* \leq \frac{e^{-m(m+1)/2n}}{n}. \tag{6}$$

This is maximised by $n = m(m+1)/2$ giving

$$\max P^* \leq \frac{2e^{-1}}{m(m+1)}. \tag{7}$$

This is an appropriate p-value for the test of H: the accused is innocent. Even with a sample of size 100, the p-value is less than 1 in 10,000, and this seems sufficient for forensic purposes. It should be noted that under the alternative hypothesis H_1: the accused is guilty, max $P^* = 1$.

The analysis is not so simple when the number of matches in the sample is not 0, but the work is in progress. As is clear, the argument in no way depends on the categories being defined by DNA typing, but applies to any method of classification.

It is obviously of importance to choose a suitable criterion for a match. Usually these criteria have been based on some number of standard deviations of the error, without any stronger argument than that this should give a small probability of a mismatch. However, the most obvious course is to derive the criterion directly from a database. If this contains duplicate profiles, then it should be possible to devise a criterion which allows a very small percentage of false matches and a very high probability that two profiles from the same person will be declared a match. This has been shown to be possible by Herrin (1993) and Sudbury, Marinopoulos and Gunn (1993). A blanket criterion of allowing a 2 or 5% error for each band independently, neither takes into account band-shift nor the number of loci that have been successfully probed.

I enjoyed reading Kathryn Roeder's review of the DNA fingerprinting controversy and found it a fair-minded and comprehensive survey of the area. But has all this work really been necessary? Could we not have saved the courts a lot of trouble by keeping things simple?

Comment

William C. Thompson

To determine the value of forensic DNA evidence for proving two samples have a common source, one must take into account three sources of uncertainty. First, there is uncertainty about the interpretation of laboratory results. Were the bands in the DNA prints scored correctly? Has the analyst adequately accounted for any discrepancies between the "matching" prints? How likely are such discrepancies if the samples have a common source? Second, there is uncertainty about laboratory error. Could an error, such as inadvertant switching, mixing or cross-contamination of samples, have accounted for the incriminating results? How common are such errors? Third, there is uncertainty about the probability of a coincidental match. How rare are the matching genotypes?

Kathryn Roeder's review of the controversy over DNA fingerprinting focuses primarily on estimation of the frequency of matching genotypes. Her discussion of this intricate issue is helpful, although she might be faulted for failing to cite and discuss the arguments and data presented by other scholars who take a different point of view (e.g., Slimowitz and Cohen, 1993; Krane et al., 1992; Mueller, 1993; Geisser and Johnson, 1993). A more important complaint is that Roeder fails to take adequate account

William C. Thompson is Associate Professor of Criminology, Law and Society at University of California, Irvine, California 92717.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *Statistical Science*. www.jstor.org