quence among the "repeats" in the VNTR loci, so that each person can be recognized by a unique signature. If our interest is, indeed, to correctly identify the perpetrators of violent crimes, then it is unclear why we continue to argue about probability calculations and statistical artefacts in place of carrying out the necessary research to create a real "DNA fingerprinting."

Comment

Aidan Sudbury

Imagine a scientific world in which there is no theory of population genetics, but in which a clever technique has been devised which associates with each individual a set of six characteristics. Let us give this technique a name, say, "DNA fingerprinting." It is claimed that these fingerprints identify someone with a high degree of reliability. To test this viewpoint, databases are assembled and it is found that matches are indeed very rare—in fact, that a match between different individuals is found on average 1 in 10,000 times. Compared to other evidence accepted by the courts, such as identity parades, alibis, motives for the crime, this is considered very reliable and has become accepted.

Now, some years later, the theory of population genetics evolves, and a new method of determining match probabilities is based on this theory. Two things may happen. First, the calculations suggest match probabilities of the order of 1 in 10,000. In which case we may say "How interesting! But I don't think we want to burden the courts with the considerable complexities involved with these calculations. We're quite happy with the way we're doing things." Second, the calculations may suggest probabilities of the order of 1 in 100,000, in which case we shall just assume they are wrong.

To return to the real world: it seems that the undoubted charms of population genetics, with its Hardy–Weinberg and linkage equilibrium, have led us into confusing the primary with the secondary evidence. If the observed match probabilities in databases were not small, no amount of testing of databases for independence, or discussion as to just how different allele frequencies are in different races could persuade us that the theory we were using was correct.

As far as I know, all investigations of databases (see, e.g., Risch and Devlin, 1992a, b; Herrin, 1993; Sudbury, Marinopoulos and Gunn, 1993) have shown that matches between unrelated individuals are extremely unlikely. Among related individuals, only the immediate family (brother, sister, father, mother) are sufficiently close to give a probability of a match that is not forensically significant. What perhaps remains to be shown is that matches within small communities are still rare even though there has been a degree of inbreeding in the past. Nichols and Balding (1991) have treated this problem theoretically, but some data covering these situations would be welcome.

Now, let us see how knowledge about the number of matches in a database may be used to make statements about the probability of guilt. Suppose the population can be classified into an unknown number of categories $C_1, \ldots, C_n$ and that these have unknown frequencies $p_1, \ldots, p_n, \sum p_i = 1$. Further, a sample of size m has been taken and none have been found to be from the same category (there have been no matches). Now, a sample taken from the accused has been found to be in the same category as a crime sample, but both are different from any in the original sample. The aim is to use this data to test the hypothesis H: the accused is innocent.

The probability that the crime samples should match, but no others, under H is

$$P^* = \sum_{i=1}^{n} p_i^2 \sum_{j_l \neq j_k \neq i} p_{j_1} \cdots p_{j_m}. \tag{1}$$

An appropriate p-value of the test is the maximum of this expression over all sets $\{p_i\}$. Consider the terms involving p_i and p_j. They are of the form

$$A(p_i^2 p_j + p_j^2 p_i) + B(p_i^2 + p_j^2) + C p_i p_j + D(p_i + p_j), \tag{2}$$

where A, B, C and D are functions of the other p_l. This expression can be written

$$\left[A(p_i + p_j) + C - 2B\right] p_i p_j + B(p_i + p_j)^2 + D(p_i + p_j). \tag{3}$$

Aidan Sudbury received a Ph.D. in Astrophysics from Monash University, where he is now a Senior Lecturer. His address is: Department of Mathematics, Monash University, Clayton 3168, Australia.

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 $\sum p_i = 1$. 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.