PROBLEMS ASSOCIATED WITH THE DETERMINATION OF BAND MATCH PROBABILITIES

P Gill, D J Werrett and I W Evett

Central Research and Support Establishment, Home Office Forensic Science Service, Aldermaston, Reading, Berkshire, RG7 4PN

INTRODUCTION

Recently Lander (1989), and Evett et al (1989a) have addressed the problem of determining whether two bands match each other in suspect and crime samples. Whereas the statistical treatments discussed by Gill et al (1989) in general provide a conservative answer to the problem of allele frequency estimation, the approach does not give information regarding the probability that 2 bands match, taking into account errors in measuring band position. Gjertson et al (1988) suggested a Bayesian model which addressed the problem of band match determination, although this model did not take account of the correlation of shift in band position.

Gjertson et al (1988) introduced a Bayesian analysis which incorporated measurement error associated with fragment size determination. Evett et al (1989b) introduced a model which took account of the correlation of shift in band position.

A Bayesian Approach

Any attempt to define a clear cut off point in deciding whether two bands match each other is problematical for the following reasons: first, criteria used to decide a cut off point are arbitrary. Second, the probability that the samples are from one individual has occurred must decrease the further away the bands are from each other. Intuitively, the pattern in Figure 1, profile b is a 'better' match than that in Figure 1, profile c yet regardless of this, probabilities are usually assigned to be the same, i.e. 1. If an arbitrary cut off point of 3 SDs is given, this could mean that if bands are greater than 3 SDs apart, the probability that samples have come from one individual suddenly plummets to 0. Clearly the declaration of a match or an exclusion cannot be a binary event. A moderator needs to be introduced into the analysis to take account of bands which are not recorded at precisely the same positions. One approach would be to define a match as occurring when bands are <3 SDs apart. If reported as inconclusive. Alternatively, cut-off points can be avoided by introducing a Bayesian analysis, where

Advances in Forensic Haemogenetics 3
Edited by H. F. Polesky and W. R. Mayr

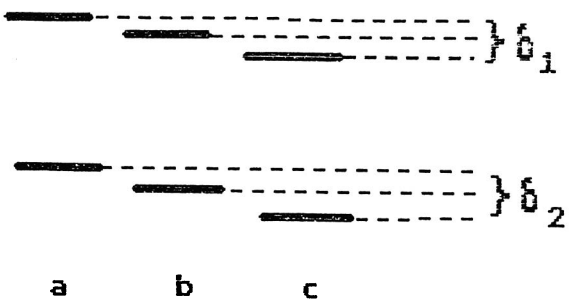

Fig. 1. Unless the correlation coefficient is taken account of, the probability of profiles b and c matching control a are the same.

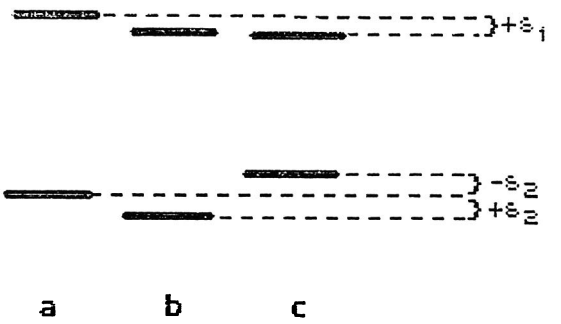

Fig. 2. If profile a is the control, then intuitively, b is a better match than c. This is allowed for in Bayesian statistics but using 'binning' or 'sliding window' routines Gill et al (1989) gives equal probabilities to the results.

the likelihood ratio assesses the relative probabilities that two profiles have originated either from the same or from different individuals:-

$$\frac{\text{The probability of the evidence given that two profiles have the same origin}}{\text{The probability of the evidence given that two profiles have different origins}}$$

The numerator is the probability of obtaining the two profiles given that they have originated from the same source. The demoninator measures the probability that the stain has come from someone else. If the likelihood ratio is greater than 1 it will tend to support the hypothesis that the control and the crime stain have the same origin, whereas a value of less than 1 will support the hypothesis that the stain has come from someone else.

The denominator assesses the probability of the evidence if the profile has come drom a randomly chosen individual, and this can be calculated from another probability density function derived from the distribution of bands in a given population. The density function in Figure 3 has been produced using Kernel density

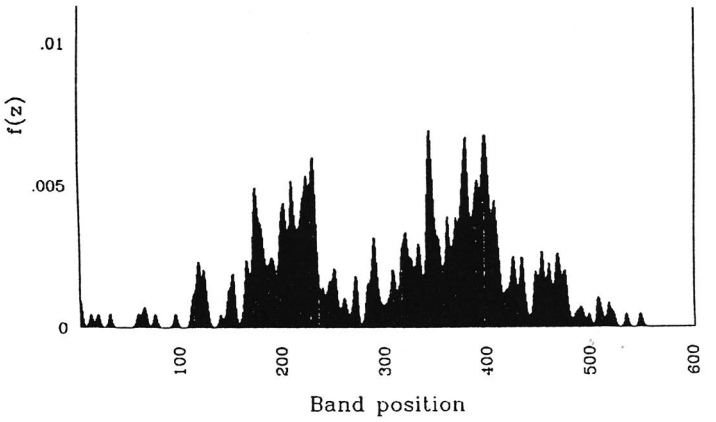

Fig. 3. Frequency histogram of D2S44 produced by Kernel density smoothing.

smoothing. This routine takes raw data and simulates an expected distribution if an infinitely large population had been sampled. The area under the curve is 1. If the recorded band position falls close to a sharp peak then a conservative value could be taken (i.e. the value of the peak itself).

If the profile contains 2 bands at positions y1 and y2 and the probability of obtaining a profile with 2 bands is p(2) then the probability of finding two bands at those positions can be given:-

$$p(2)p(y1,y2 \mid 2)$$

If the profile has only one band at position y1 and the probability of obtaining a profile with 1 band is p(1) then the denominator is:-

$$p(1)p(y1 \mid 1)$$

The advantage of using these expressions rather than the equivalent genetic expressions $2p(y1)p(y2)$ for heterozygotes and $p(y1)^2$ for homozygotes is that the former do not require unambiguous identification of homozygotes and heterozygotes. Theoretically, two bands may appear so close together, that they cannot be distinguished and appear as a homozygote. Alternatively, bands may be scored as homozygotes if a low molecular weight band has run off the end of a gel or it too faint to be visualised. Evett et al (1989) has introduced a correction term to allow for the fact that if a

band appears at position y1, then the second at y2 must be greater than the resolution interval of the electrophoretic system. In Figure 4 a plot of $\delta_1 = -\delta_2$ is compared with $\delta_1 = \delta_2$. As the

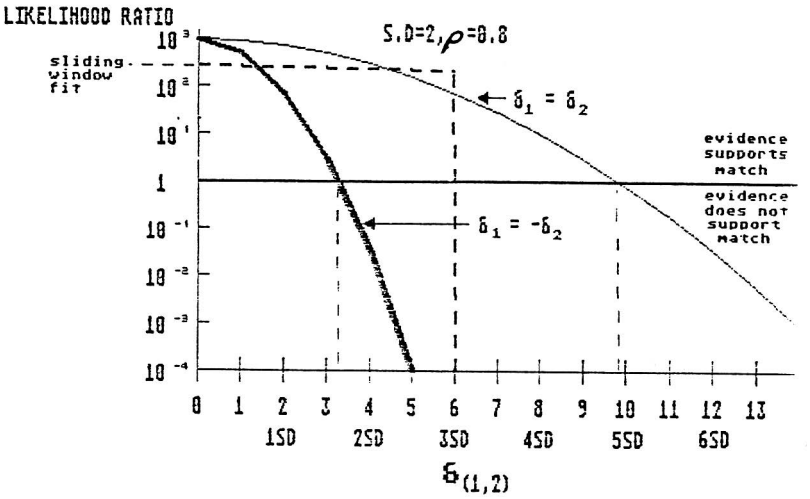

Fig. 4. Plots of likelihood ratio versus delta (1,2) when SD = 2, ρ = 0.8; frequency of 2 bands = 0.005 each. Likelihood ratio calculated using a sliding window Gill et al (1989) is superimposed.

distance between bands in the control and suspect profiles increases, the likelihood ratio falls. The likelihood ratio = 1 when $\delta_1 = \delta_2 = 10$ units (5 SD's separation), contrasting with the alternative plot of $\delta_1 = -\delta_2 = 3$ units (<2 SD's). The equivalent likelihood ratio obtained from a sliding window fit is superimposed, emphasising the problem of using an arbitrary cut-off point where the value drops from 167 to 0 at 3 SD's separation distance. The lower standard deviation, the greater the potential maximum likelihood ratio, demonstrating the need to optimise electrophoretic systems, perhaps by introducing internal markers, for example.

The importance of the correlation coefficient is illustrated in Figure 5. When $\delta_1 = \delta_2$, the likelihood ratio steadily increases with the correlation coefficient for any given value.

CONCLUSION

To conclude, some of the problems of band matching probabilities have been discussed. Lander (1989) has already pointed out the problems of using an arbitrary cut-off point in deciding whether

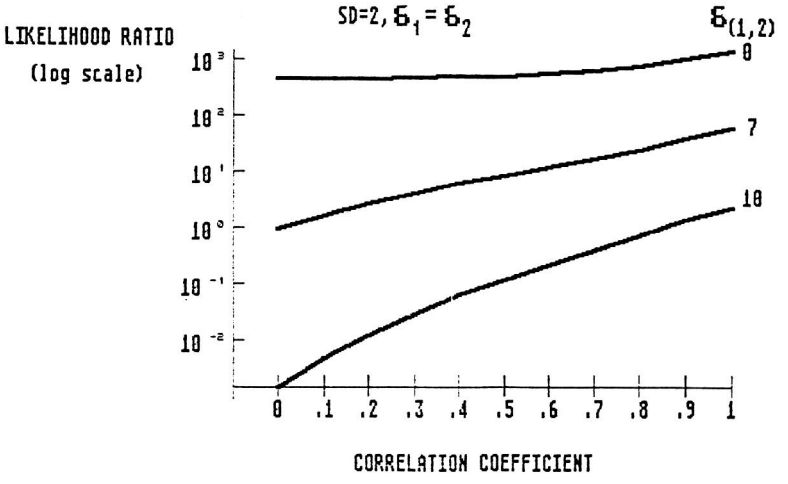

Fig. 5. Likelihood ratio versus correlation coefficient of delta 1 = delta 2 (0, 7, 10 units separation). Band frequencies = 0.005, SD = 2.

two bands match. Using a Bayesian approach the decision of whether 2 profiles could have had the same origin is no longer binary and does not depend upon arbitrary decisions regarding cut-off point. The likelihood ratio is dependent upon band separation between control and suspect profiles. The strength of the analysis relies upon a detailed analysis of characteristics and errors in an electrophoretic system.

REFERENCES

Evett IW, Werrett DJ, Gill P and Buckleton JS (1989a) DNA fingerprinting on trial. Nature, 340: 435.

Evett IW, Werrett DJ, Pinchin R, Gill P and Buckleton JS (1989b) Bayesian analysis of single locus DNA profiles (in preparation).

Gill P, Jeffreys AJ and Werrett DJ (1985) Forensic application of DNA 'fingerprints'. Nature 318: 577-579.

Gill P, Sullivan KS, Werrett DJ (1989) Problems associated with the objective determination of band matching frequencies in hypervariable DNA. Hum Genet (in press).

Gjertson DW, Mickey R, Hopfield J, Toshinao T, Terasaki PI (1988) Calculation of probability of paternity using DNA sequences. Am J Hum Genet 43: 860-869.

Lander ES (1989) DNA fingerprinting on trial. Nature 339: 501-559.