DNASYS: A User-Friendly Computer Program for Evaluating Single Locus Probe Data in Forensic Casework

I.W. Evett, R. Pinchin and J. Scranage

The Forensic Science Service, Central Research and Support Establishment,
Aldermaston, Reading, Berkshire, RG7 4PN, UK

INTRODUCTION

Match/binning is not an efficient method for evaluating single locus probe (SLP) data and there are several reasons why a Bayesian approach is preferable. See, for example, Evett et al. (1990), Berry (1991), Berry et al. (1991) and Evett et al. (1991).

We have been using Bayesian analysis, mainly as a research tool, in the Forensic Science Service (FSS) for over two years, but recently we have been able to develop a computer program known as DNASYS for use by forensic caseworkers. This will be made available worldwide in early 1992 through new marketing initiatives of the FSS.

In this note we briefly describe the method and give the results of just two of the many experiments we have carried out to test its robustness. We give an overview of DNASYS and of some of the support services which we will be able to provide for those who are interested in implementing the program for their own use.

METHOD

Whereas match/binning is a two stage process, it is much more sensible to evaluate evidence through a function which takes account of all of the relevant information in one step. This is the likelihood ratio (LR), which takes account of how close the measurements on the crime sample are to those on the suspect's sample and also of how rare or common those measurements are in the population of interest. Instead of a match function which switches from "match" to "non-match" at some arbitrary threshold it is a continuous function which is large when measurements are close, becoming progressively smaller as they become further apart. A critical feature of the rate at which the ratio falls off is an allowance for the correlation in band weight measurement errors - band shift. This increases discrimination and hence efficiency.

Values of the LR less than one support the hypothesis that crime and suspect samples come from different people. Values greater than one support the hypothesis that they came from the same person: the larger the ratio, the greater the support.

DATA

The following experiments were carried out on data for the probes MS1 and MS31 collected at the Metropolitan Police Forensic Science Laboratory (MPFSL) as

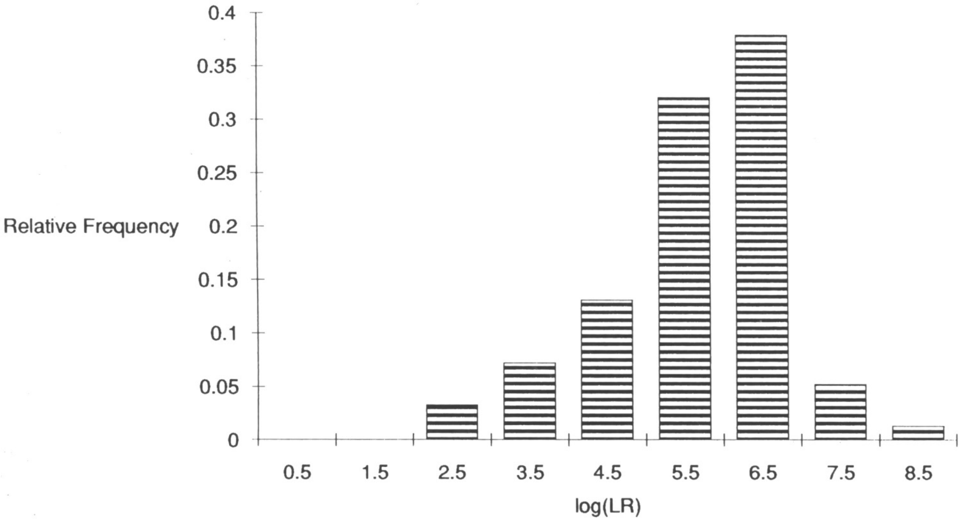
Advances in Forensic Haemogenetics 4
Edited by Ch. Rittner and P. M. Schneider

described by Buffery et al. (1991). The reproducibility parameters (standard deviations and correlation coefficients) were estimated from duplicate profilings carried out by the MPFSL.

WITHIN-PERSON COMPARISONS

A total of 153 Caucasians had been profiled twice and we calculated the LR using the data from the two probes for each of the 153 pairs. Figure 1 is a histogram of the logarithm of the LR and the histogram intervals are labelled by their mid points; so 6.5, for example, corresponds to LR's from one million to ten million. Note that approximately 97% of the comparisons gave LR's in excess of 1000.

Fig.1. Within person comparisons



Of course, it is not sufficient to know that the method gives such high LR's when two samples from the same person are compared. We must also establish what can happen when samples from two different people are compared with each other.

BETWEEN-PERSON COMPARISONS

For this experiment we took a file of data on 1056 Caucasians and carried out all of the 557,040 between-person comparisons. The vast majority of these resulted in infinitesimal likelihood ratios but we are only concerned with those occasions when the LR exceeded one. These are summarised in Figure 2, again a histogram of the log of the LR.

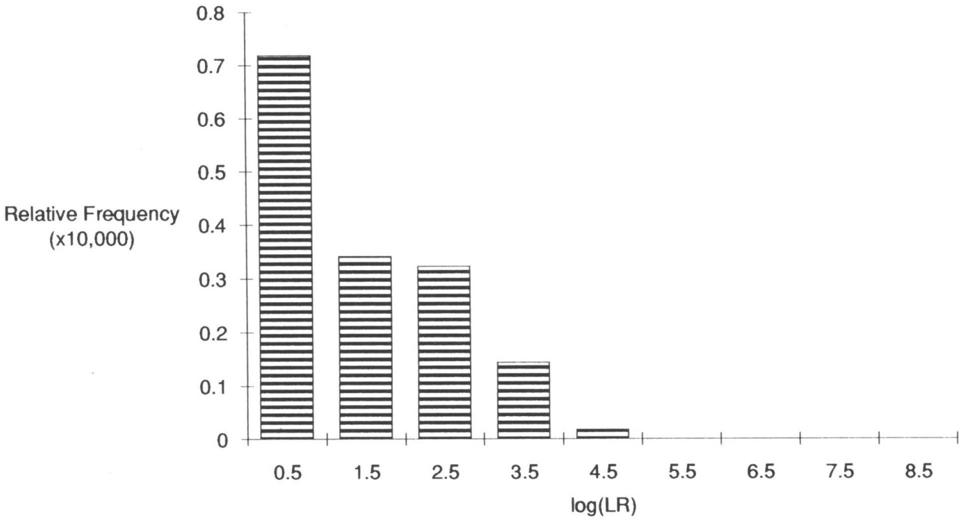


Fig.2. Between person comparisons

It is important to note that the vertical scale has been expanded by a factor 10,000 and to remember that this is the extreme right hand tail of the distribution. The probability of incorrectly giving associative evidence is very small. Note, in particular, that the chance of a LR in excess of 1000 is less than 1 in 56,000. This is very conservative.

DNASYS

The program has been written in PASCAL to run on IBM PC's and clones. We have placed particular emphasis on the user interface so the program is easy to use. Data from up to four probes can be accommodated: we are aware that many laboratories use more but, for crime work at least, there is little need to do so given the statistical power of the method.

INTELLIGENCE SEARCHING

Several organisations have indicated an intention to maintain DNA databases on known criminal offenders to assist in the solution of serious crimes. Databases of unsolved sex crimes have considerable value because linking crimes provides valuable information to investigators.

Retrieval from such databases by simple window searching is a clumsy and highly inefficient procedure - particularly when the database is large. To ensure successful retrieval of the correct entry the search window must be wide and this means a substantial number of false hits - all of which need to be followed up manually, costing time and money.

DNASYS has a search module which is much more efficient than window searching, mainly because it takes account of the fact that measurement errors in the band weights of one profile are highly correlated. Retrieval lists are much shorter than from window searching and a further great advantage is that any hits are ranked by the magnitude of the LR. This can save hours of manual checking.

FUTURE PLANS

Work is in hand to create a module for paternity analysis. Here, too, it is our aim to reduce the number of probes necessary to achieve good evidential results and also to create a good user interface. The effort that we devote to development of this module will be determined by the amount of interest that practitioners demonstrate. This conference is an excellent opportunity for us to gauge the level of that interest.

There is also scope for more advanced modules for dealing with mixed crime cases and for cases where relatives of the suspect may be suggested as hypothetical alternatives. Here again, the manpower which we devote to such developments will be determined by the level of interest.

The program clearly cannot run without databases and we are able to provide a service for converting users' databases to the correct format. We can also carry out analyses for determining standard deviations and correlation coefficients assuming that the users can provide suitable data.

We also have a suite of programs which we have used to test the robustness of our own procedures. We have studied, in particular, the robustness of the various independence assumptions which are much debated in the literature. We have also investigated the consequences of using non-representative databases and find that they have been exaggerated. In 1992 we will be making such analysis services available to other organisations on a consultancy basis.

CONCLUSION

The progress of DNA profiling has been hindered by poor quality statistical reasoning. Efficient statistical methods such as those embodied in DNASYS enable the real power of single locus profiling to be achieved. Good statistics mean that more information can safely be gleaned through the use of fewer probes: this saves time and it saves money.

REFERENCES

Berry DA (1991) Inferences using DNA profiling in forensic identification and paternity cases. Statistical Science. In press

Berry DA, Evett IW, Pinchin R (1991) Statistical inference in crime investigations using DNA profiling. Applied Statistics. In press

Buffery C, Burridge F, Greenhalgh M, Jones S, Willott G. Allele frequency distributions of four VNTR loci in the London area. Forensic Science International. In press

Evett IW, Werrett DJ, Pinchin R, Gill P (1990) Bayesian analysis of single locus DNA profiles. The International Symposium on Human Identificfation 1989. Promega, Madison, WI, USA

Evett IW, Scranage JK, Pinchin R (1991) An efficient statistical procedure for interpreting DNA single locus profiling data in crime cases. Journal of the Forensic Science Society. Submitted