MIXTURE TRACES PROBABILISTIC ANALYSIS IN THE COMPARISON OF SEVERAL HYPOTHESES

Marina Andrade[1], Manuel Alberto M. Ferreira[2]

*[1]Instituto Universitário de Lisboa (ISCTE-IUL), Information Sciences, Technologies and Architecture Research Center (ISTAR-IUL) and Business Research Unit (BRU-IUL), **PORTUGAL***
*[2]Instituto Universitário de Lisboa (ISCTE-IUL), Business Research Unit (BRU-IUL) and Information Sciences, Technologies and Architecture Research Center (ISTAR-IUL), **PORTUGAL***

E-mails: marina.andrade@iscte.pt, manuel.ferreira@iscte.pt

ABSTRACT

The objective of this work is to discuss various approaches in the analysis of DNA mixture profiles, starting with a mixture trace with only two contributors and extending the analysis to a larger number of contributors. Both the algebraic treatment and the use of OOBN are considered concluding with the inevitability of using the latter in more complex situations.

***Key words:** Mixture traces, Likelihood ratio, Bayesian networks, OOBN*

INTRODUCTION

Nowadays the use of DNA profiles in forensic identification problems is a very common procedure, in many and different situations. In this work it is intended to discuss various approaches in the analysis of DNA mixture profiles, starting with a mixture trace with only two contributors and extending the analysis to a larger number of contributors. In the next section is presented the set of different hypotheses to test according to the possible mixture traces pointed out. Beyond the two hypotheses emergent in court, for each case, others may also be of interest. Therefore it is needed a form to evaluate those ones in an efficient way. Thus, in the beginning, an algebraic approach is considered and then the use of Bayesian networks, particularly important in complex cases of mixture profiles. In complex cases this tool allows to compute easily the likelihood ratio for the set of all hypotheses suggested. In last section the discussion comprises possible analysis of the cases and also the potential use of Bayesian networks in this context. In a real case the number of hypotheses to test may become a hard and a long work to execute algebraically.

This subject has already been considered in Andrade and Ferreira (2007, 2007a) and Andrade, Ferreira, and Filipe, J. A. (2009). Now in the return to the ideas exposed in those works it is sought to update and improve the approach then followed.

48 | *Vol.4. No.1 (2018)*   www.actaint.com

HYPOTHESES AND DATA

The observation of mixture traces generally occurs in criminal cases. When a DNA mixture profile is mentioned it means that more than two bands were observed for one or more loci in the set of known genetic markers used to analyze the trace found. Here is discussed for illustration the set of hypotheses of a mixture with two contributors. After this are presented the data of a more complex mixture. For the purpose intended here will only be presented data for two markers.

ALGEBRAIC HANDLING

Before proceeding to a more complex case, is briefly discussed the hypotheses to test in a case to which a mixture trace was found and connected with a certain crime. In this is admitted that there were two donors, a victim *v* and a suspect *s*. In such a case the competing hypotheses are:

$$i)\ s\ \&\ v$$
$$ii)\ s\ \&\ u$$
$$iii)\ v\ \&\ u$$
$$iv)\quad 2u$$

with *u* an unknown individual in the population. Where *i)* means that the mixture is composed with DNA of the victim and the suspect; *ii)* the mixture composition is formed with DNA of the suspect and an unknown individual; *iii)* the mixture is composed with DNA of the victim and an unknown individual; and *iv)* the mixture is formed with DNA of two unknown individuals.

Assuming Hardy-Weinberg equilibrium, independence between markers, and given the allele frequencies it is possible to determine algebraically the values of the likelihood ratio for each hypothesis, Weir *et al.* (1997). That can be more or less heavy depending on the mixture observed and the known individual genotypes involved. In court the hypotheses in dispute are the prosecution hypothesis: H_P, stating that the mixture is composed with genetic material from the victim and the suspect $\{v, s\}$ *versus* the defence hypothesis: H_D arguing that the mixture results of biological material from the victim and an unknown person $\{v, U\}$. With the likelihood ratio for each of the four hypotheses one may wish to compare

$$s\ \&\ v\ vs\ v\ \&\ U, \text{or}\ s\ \&\ v\ vs\ 2U, \text{or even}\ s\ \&\ U\ vs\ 2U.$$

Now admit the following set of data:

Table 1. *Mixture trace data*

Marker	$G_{v_1}(f)$	$G_{v_2}(m)$	G_{s_2}	E_c[i] (mixture)
FES	A, C	C, C	B, B	A, B, C
FGA	B, E	B, C	A, C	A, B, C, E

An excerpt of a criminal case with two victims (v_1, v_2) and a suspect (s_2). As any criminal case the court has to answer the question $P(Guilty|E)$? Therefore it seems natural just want to compare that probability with the following $P(Not\ guilty|E)$. To compare the hypotheses may be done through the ratio of these two hypotheses determination as follows:

$$\frac{P(Guilty|E)}{P(Not\ guilty|E)} = \frac{P\left(E_c|G_{v_1}, G_{v_2}, G_{s_2}, H_P\right)}{P\left(E_c|G_{v_1}, G_{v_2}, G_{s_2}, H_D\right)} \times \frac{P(H_P)}{P(H_D)}$$

Supported data one can determine the likelihood ratio (the first factor in the right side of the equation above). Thus, the probability of the evidence given the prosecution hypothesis is one. The probability of the evidence conditional to data and the defence hypothesis can be obtained as the product of the last column of Table 2 below:

Table 2. *Probability of the evidence given the defence hypothesis*

Marker	$P\left(E_c|G_{v_1}, G_{v_2}, G_{s_2}, H_D\right)$
FES	$p_B^2 + 2p_A p_B + 2p_B p_C$
FGA	$p_A^2 + 2p_A p_B + 2p_B p_C + 2p_A p_E$

With these values one can compare the hypotheses. But, in such a case it is reasonable to be interested in a comparison of a larger set of hypotheses viewing the possible origin of the mixture - a source level proposition according to Cook *et al.* (1998). One of the complexities in the interpretation and evaluation of a mixture trace is to assign the number of total contributors. The different number of alleles present in the mixture suggest a minimum for that number but says nothing about a maximum to consider. For this Lauritzen and Mortera (2002) gave a useful low upper bound to the number of contributors to consider, and it can give some clues.

In a case like the mentioned, in which it was thought that three persons were involved, the set of hypotheses to test may admit up to three unknown individuals to consider. Therefore considering up to six contributors in the mixture. Consequently, the set of hypotheses to test will have a total of 32 states. A mixture

---

[i] *E* means evidence

with three contributors has eight hypotheses to test. In this case must be considered those eight states for the known individuals plus those eight states combined with one, two and three unknowns (see Table 3). Naturally to determine the expression and the value for each hypothesis becomes a difficult and slow task, making it difficult to proceed with the analysis. In court or during the investigation process of a real case it is extremely important to define, in reasonable time, the weight of each hypothesis in evaluation.

Table 3. *Set of 32 hypotheses to test*

0	$\emptyset$	8	u	16	$2u$	24	$3u$
1	v_1	9	v_1,u	17	$v_1,2u$	25	$v_1,3u$
2	v_2	10	v_2,u	18	$v_2,2u$	26	$v_2,3u$
3	v_1,v_2	11	v_1,v_2,u	19	$v_1,v_2,2u$	27	$v_1,v_2,3u$
4	s_2	12	s_2,u	20	$s_2,2u$	28	$s_2,3u$
5	s_2,v_1	13	s_2,v_1,u	21	$s_2,v_1,2u$	29	$s_2,v_1,3u$
6	s_2,v_2	14	s_2,v_2,u	22	$s_2,v_2,2u$	30	$s_2,v_2,3u$
7	s_2,v_1,v_2	15	s_2,v_1,v_2,u	23	$s_2,v_1,v_2,2u$	31	$s_2,v_1,v_2,2u$

BAYESIAN NETWORKS TOOL

The complex identification problems raised in the forensic area encourage the interest for the development of inference mechanisms that allow the search and attainment of answers for this problems.

The use of Bayesian networks to analyse DNA mixture profiles in criminal cases had its beginning with the works of Mortera (2003) and Mortera *et al*. (2003). Since then a more common discussion and the computational improvements achieved within the software grant a good support to the authorities, whether they are the courts or the polices.

In the previous subsection the 32 hypothesis advanced intend to cover different assumptions, starting with the involvement of the three nominated individuals till the involvement of three unknown persons. As it was highlighted the problems appear when it is needed to determine the expression and the correspondent value of each hypothesis. For the case in discussion Andrade and Ferreira (2007) have performed the analysis with **o**bject-**o**riented **B**ayesian **n**etworks (OOBN), there considering a total of five markers. After having the networks built, supported with a software program, and the insertion of the evidence (data) the results were obtained in a simple and quick way. An example of the results is given in Table 4, referring the values of each hypothesis for marker FGA. As it can be observed some values are null. This happens when the hypothesis is not consistent with the minimum number of individuals necessary to generate the mixture inserted. Obviously some hypotheses present a larger value for the likelihood ratio than others. That is one of the discussions in the next section.

Table 4. *Results for marker FGA and the given data*

State	FGA	State	FGA	State	FGA	State	FGA
0	0.0000	**8**	0.0000	**16**	0.0006	**24**	0.0004
1	0.0000	**9**	0.0083	**17**	0.0052	**25**	0.0015
2	0.0000	**10**	0.0017	**18**	0.0015	**26**	0.0006
3	0.0000	**11**	0.0207	**19**	0.0067	**27**	0.0016
4	0.0000	**12**	0.0042	**20**	0.0029	**28**	0.0009
5	0.3768	**13**	0.0714	**21**	0.0135	**29**	0.0026
6	0.0000	**14**	0.0101	**22**	0.0036	**30**	0.0009
7	0.3768	**15**	0.0714	**23**	0.0135	**31**	0.0026

DISCUSSION

The analysis of mixture traces and the evaluation of important hypotheses connected to criminal context present obvious troubles. The algebraic treatment becomes more complex as it is admitted one more person in a mixture. Wanting to question a bigger number of contributors the complexities increase largely. An important resource to deal with these problems is the one mentioned.

In a criminal case of forensic identification, before its evaluation in court, usually it is necessary to test and compare a certain number of hypotheses connected with the inherent conjectures. At an earlier period the policies during the investigation processes have to define the reasonable scenarios and to determine the important ones that will be evaluated by the courts. And even in court beyond the main comparison some others may be important. Whatever the circumstances are, to perform those comparisons as quick and efficiently as possible is an exigency of all the parts involved in the judicial area. The results of Table 4 give an illustration of what can be tested and that some conjectures should not be left to appreciate. For example, some of the hypotheses in which it is considered the presence of an unknown person are not to be depreciated.

Also worth to be mentioned is the modularity and flexibility of OOBN, which allow its possible use in cases with similar details and the extension to more complex cases. The different modules or instances can be reused to analyze different problems. In the new problem one can define the necessary new objects and combine them with the already defined, and deal with the singularity of each case.

REFERENCES

1) Andrade, M. and Ferreira, M. A. M. (2007), Analysis of a DNA mixture sample using object-oriented Bayesian networks. *Proceedings of the 6th International Conference APLIMAT 2007*, Feb 6-9, Bratislava, Slovak Republic.
2) Andrade, M. and Ferreira, M. A. M. (2007a), Mixture traces: comparison of several hypotheses*. Bulletin of the International Statistical Institute LXII*, 3097-3100.
3) Andrade, M., Ferreira, M. A. M. and Filipe, J. A. (2009), Evidence evaluation in DNA mixture traces. *Journal of Mathematics Statistics and Allied Fields*, 2(2).

4) Cook, R., Evett, I. W., Jackson, G., Jones, P. J. and Lambert, J. A. (1998), A hierarchy of propositions: Deciding which level to address in casework. *Science and Justice*, 38, 151-156.

5) Lauritzen, S.L. and Mortera, J. (2002), Bounding the number of contributors to mixed stains. *Forensic Sci. Internat.* 130, 125-126.

6) Mortera, J. (2003), Analysis of DNA mixtures using probabilistic expert systems. In: Green, P.J., Hjort, N.L., Richardson, S. (Eds.), Highly Structured Stochastic Systems. Oxford University Press.

7) Mortera, J., Dawid, A. P., and Lauritzen and S. L. (2003), Probabilistic expert systems for DNA mixture profiling. *Theoretical Population Biology*, 63, 191-205.

8) Weir, B. S., Triggs, C. M., Starling, L., Stowell, L. I., Walsh, K. A. J., and Buckleton, J. S. (1997), Interpreting DNA mixtures. *Journal of Forensic Sciences*, 42, 213-22.