EasyChair Preprint

№ 7897

Investigation of a Crime Scene with Two Victims and a Perpetrator Through DNA Traces

Marina Andrade and Manuel Alberto M. Ferreira

May 3, 2022

Investigation of a Crime Scene with Two Victims and a Perpetrator through DNA Traces

Marina Andrade

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

Manuel Alberto M. Ferreira

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

To deal with crime identification problems, that are examples of situations in which forensic approach the DNA profiles is frequent, it is needed an introduction to present and explain the various concepts involved. So the use of object-oriented Bayesian networks (OOBN), examples of probabilistic expert systems (PES), is shown and exemplified.

Key words *Probabilistic expert systems, Bayesian networks, DNA profiles, Identification problems*

INTRODUCTION

The use of networks transporting probabilities began with Sewall Wright in the beginning of the 20th century (1921). Then they assumed different forms in several areas in which the models are, in general, linear. It is the case of the so named Path Diagrams or Structural Equations Models (SEM). But in artificial intelligence generally non-linear models, named Bayesian networks, are used, also called Probabilistic Expert Systems (PES).

Bayesian networks are graphical structures for representing the probabilistic relationships among a large number of variables and for doing probabilistic inference with those variables, (Neapolitan, 2004).

To approach the use of Bayesian networks to the problems of interest, some aspects of PES with uncertainty problems must be studied, see for instance (Cowell *et al*., 1999).

Here the use of Bayesian networks is illustrated with an example, in the field of forensic identification, of an investigation of a crime scene with two victims and a perpetrator where DNA profiles are considered.

This material was already presented at 4th IISMES-International Institute of Statistics and Management Engineering Symposium, Dalian, China, July 24-29 2011.The present

work is a version corrected, revised and updated of the one published in the 4th IISMES Proceedings, see (Andrade and Ferreira, 2011a).

THE USE OF BAYESIAN NETWORKS

In (Dawid *et al*., 2002) it is described a new approach to the problems mentioned above. The building and use of Bayesian networks to analyse complex problems of forensic identification inference was initially done there, followed by (Evett *et al*., 2002), (Dawid *et al*., 2002), (Mortera, 2003) and (Mortera *et al*., 2003) among others.

Mixtures

The achieved advances in the forensic biology have certainly encouraged the interest in problems of forensic identification, allowing also a much more rigorous treatment of the problems in analysis. That is the case of DNA mixtures problems – (Mortera, 2003) and (Mortera *et al*., 2003).

One of the complexities in the interpretation of the mixture traces is assigning the number of the contributors to the mixture. In general, the trace suggests a lower bound for the total number of contributors but not an upper bound. In (Lauritzen and Mortera, 2002) it is given an useful low upper bound for the number of contributors worth considering.

In what follows it is described a complex mixture case and presented the data important in the analysis. After the hypotheses formulation, the analysis is performed for one marker considering the information from one trace. Then the two traces are considered and finally the analysis is generalized considering two mixture traces and the three markers.

The case under appreciation

A crime has been committed, and two persons were murdered, V_1 and V_2. At the crime scene two different mixture traces were found: T_1 in the toilet and T_2 in the victims' car. A potential suspect is identified, S_2. And his/her DNA profile was measured and found to be compatible with the mixture traces.

Assuming that a fight occurred during the assault producing some material, it is obvious that the individual who perpetrated the crime could left some of his/her material in some but not in the whole traces. The non-DNA evidence indicates that two persons could be involved in the crime.

Excerpt of data

To summarize the evidence the DNA profiles of the victims' and the suspect, S_2 are presented in Table 1. In Table 2 the profiling results for the mixtures traces (T_1 and T_2), for the STR markers studied, respectively, and the allele frequencies for each marker are presented.

The traces contain biological material that must belong to some person other than the two victims. The allele frequencies used in this work are the Portuguese population frequencies collected in the worldwide database “The Distribution of Human DNA-PCR Polymorphisms”, since the case mentioned took place in Portugal.

Table 1. Two victims and suspect DNA profiles

Crime scene interveners / Marker	V_1 (f)	V_2 (m)	S_2
TH01	*D;E*	*D;E*	*B;C*
FES	*A;C*	*C;C*	*B;B*
FGA	*B;E*	*B;C*	*A;C*

It must be considered that the crime traces can contain DNA from up to three unknown contributors, in addition to the victims and/or the suspect.

If the DNA from S_2 is present in at least one of the traces, this will place him/her at the crime scene and consequently as one of the possible perpetrators. Consideration of whether or not the suspect was a contributor to any of the mixture traces will give a measure of the evidence strength.

Table 2. DNA mixture traces and allele frequencies[1]

	TH01	FES	FGA
*T*1	*B; C; D; E*	*A; B; C*	*A; B; C; E*
*T*2	*B; C; D; E*	*B; C*	*A; B; C*
p_A	*	0.0129	0.0684
p_B	0.1696	0.3287	0.1740
p_C	0.1386	0.3664	0.1606
p_D	0.1984	*	*
p_E	0.2748	*	0.0321

Hypotheses

The court has to determine if the suspect is or is not guilty. These are described as the level III, or offence, propositions. However the forensic scientist does not typically address such propositions. In this case it appears more appropriate to address source level propositions, as follows:

[1] The use of * refers values that are of no concern in the analysis.

H_1: S_2 is one of the contributors to T_1 but not T_2.
H_2: S_2 is one of the contributors to T_2 but not T_1.
H_3: S_2 is one of the contributors to both T_1 and T_2.
H_4: S_2 did not contribute to trace T_1 or T_2.

What interests to measure is

$$P(S_2 \text{ contr. to at least one of the traces} \mid \xi)$$

where ξ is the vector comprising the profiles observed of the traces found at the crime scene, the victims' and the suspect profiles. This is equivalent to

$$P(H_1 \cap H_2 \cap H_3 \mid \xi) = 1 - P(H_4 \mid \xi).$$

One mixture trace and a single marker

The network for one trace and a single marker follows (Mortera *et al*., 2003), an OOBN version considering up to three unknown contributors: **marker** network, Figure 1. Here it is presented the network for the marker, FES[2].

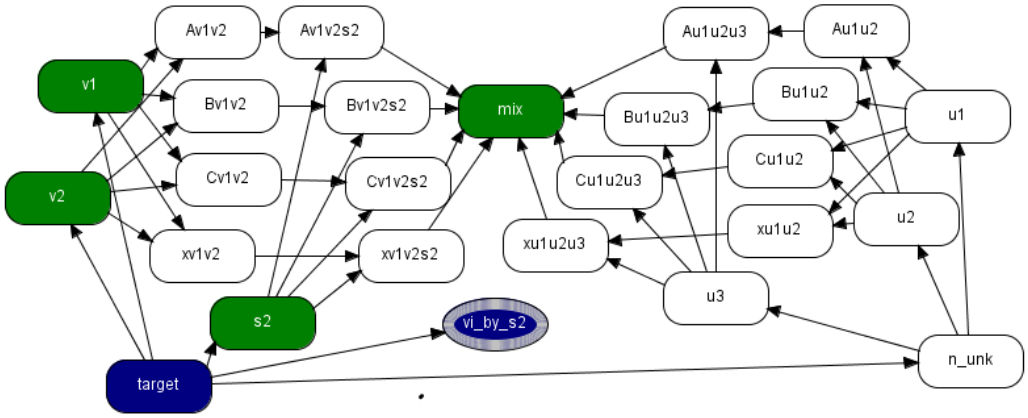

Figure 1. Marker network

Two mixture traces and a single marker

As described above, there were two different traces at the crime scene. So it is necessary to combine the information from both traces. To do so define an instance **combine**, Figure 2. This instance has as parents the output nodes *vi_by_s*2 of the instance **marker** for trace T_1 and trace T_2. The node $T_1\_T_2$ combines the results obtained in the parent instances for node *vi_by_s*2, expressing the result values of the one-to-one correspondence with the eight joint configurations of its parents nodes for the considered marker.

[2]The marker networks differ only in the number of alleles to consider, whether it is the space of states of the nodes referring the alleles or in the presence of one more allele to consider in the network. Since Hugin software does not allow modification of the state of a node in order to reuse a network, for markers TH01 and FGA a codification in the space of states of the node *gene* was performed and put it in accordance with the alleles of each marker under consideration so that it could used the same network.

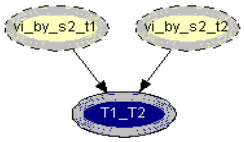

Figure 2. Combine network

Therefore, the node $T_1\_T_2$ assumes values 0, 1, 2, 3 corresponding to the hypothesis H_4, H_1, H_2 and H_3, respectively. $T_1\_T_2$ is 0 if *vi_by_s*2 is less than 4 in T_1 and T_2; assumes value 1 if *vi_by_s*2 is equal to 4 or more in T_1 and less than 4 in T_2; takes value 2 if *vi_by_s*2 is less than 4 in T_1 and equal to 4 or more in T_2; and is 3 if *vi_by_s*2 is equal to 4 or more in both T_1 and T_2. In the start an uniform prior distribution for node $T_1\_T_2$ is assumed.

Now it is possible to put the networks for each trace together and compute the interest information, Figure 3. The instances **FES trace_t1** and **FES trace_t2** are of class **marker** in which all the individuals in any of the networks have the same structure (**individual**). His/her differentiation is made when the evidence is inserted.

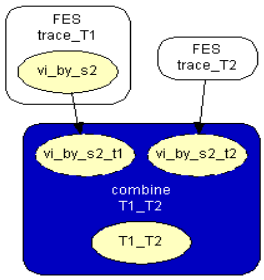

Figure 3. Combine_T1_T2 network

When combining the two traces, in order to obtain a measure of the evidential weight associated to the possible presence of genetic material from the suspect in the traces found at the crime scene, the results listed in the Tables below are obtained. For marker FES with different mixture traces:

Table 3. Results of the node *vi_by_s2*

S_2, V_2, V_1	trace T_1	trace T_2
0 (FFF)	0.0048	0.1470
1 (FFT)	0.1334	0.0000
2 (FTF)	0.0068	0.1791
3 (FTT)	0.1334	0.0000
4 (TFF)	0.0072	0.1881
5 (TFT)	0.3526	0.0000
6 (TTF)	0.0092	0.4857
7 (TTT)	0.3526	0.0000

where the state 0 corresponds to *s*2*_in_mix*? = *False*, *v*2*_in_mix*? =*False* and *v*1*_in_mix*? = *False* (FFF), and for simplicity the state 0 is read as S_2; V_2; V_1 = *FFF*.

In Table 4 it is shown the combined information of the two traces for marker FES.

Table 4. Results for the node *T1_T2*

H_1	0.2353
H_2	0.1876
H_3	0.4862
H_4	0.0908

Thus,

$$P(S_2 \ contr.to\ at\ least\ \text{one}\ of\ the\ traces \mid \xi) \cong 0.91 \cdot$$

Generalization for two mixture traces and three markers

Given the results obtained for one marker it is necessary to extend the reasoning in order to consider the information for the three markers, FES, TH01 and FGA.

The instances **combine_T1_T2** express the results for each marker accounting the information for the two traces. The node $T_1\_T_2$ in each of these instances computes the results for each marker. The respective tables, similar to Table 4, can be extracted for the other two markers.

The instance **accumulate** having as inputs the output nodes of the instances combine $T_1\_T_2$, with the results of each marker, incorporates the information for the two traces obtained separately, Figure 4. The node *multi_markers* combines the information from the different instances **combine_T1_T2**, i.e., *multi_markers* gives the results synthesizing the results of $T_1\_T_2$ for the three markers. The node *multi_markers* with states 0, 1, 2 and 3 assumes the state 0 if all the input nodes are 0. Takes value 1 if all the input nodes are 1 or at least one of the input nodes has state 1 and the others have the state 0[3]. The node *multi markers* is 2 if all the input nodes have state 2 or this state 2 is combined after the states 0 and 2 of the input nodes. The node assumes state 3 if all the input nodes have state 3 or if the inputs are combining state 0, state 1 and state 2.

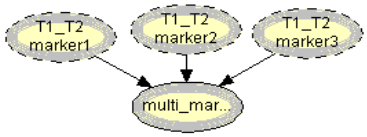

Figure 4. Accumulate network

Joining the networks for the three markers, each of which accounts for the two traces, it is obtained the **accumulate_three_markers** network, Figure 5.

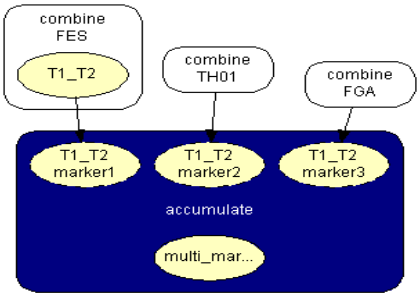

Figure 5. Accumulate three markers network

[3] e.g., *multi markers*=1 if

$T_1\_T_2$ =1 for *marker*1, *marker*2 and *marker*3; or $T_1\_T_2$ =1 for *marker*1 and *marker*2 and $T_1\_T_2$ =0 for *marker*3; or $T_1\_T_2$ =1 for *marker*1 and *marker*3 and $T_1\_T_2$ =0 for *marker*2; or $T_1\_T_2$ =1 for *marker*2 and *marker*3 and $T_1\_T_2$ =0 for *marker*1; or $T_1\_T_2$ =1 for *marker*1 and $T_1\_T_2$ =0 for *marker*2 and *marker*3; or $T_1\_T_2$ =1 for *marker*2 and $T_1\_T_2$ =0 for *marker*1 and *marker*3; or $T_1\_T_2$ =1 for *marker*3 and $T_1\_T_2$ =0 for *marker*1 and *marker*2.

Tables 5 and 6 display the results for the marker FGA and TH01 and the cumulative result for all three markers, rescaled to sum up to 1. This aims at the question of interest.

Table 5. Results for the eight configurations for markers FGA and TH01

S_2, V_2, V_1	trace T_1	trace T_2	trace T_1	trace T_2
0 (FFF)	0.0010	0.0084	0.0134	0.0134
1 (FFT)	0.0150	0.0000	0.0342	0.0342
2 (FTF)	0.0037	0.0476	0.0342	0.0342
3 (FTT)	0.0290	0.0000	0.0342	0.0342
4 (TFF)	0.0079	0.0977	0.0599	0.0599
5 (TFT)	0.4644	0.0000	0.2748	0.2748
6 (TTF)	0.0146	0.8463	0.2748	0.2748
7 (TTT)	0.4644	0.0000	0.2748	0.2748

Table 6. Results for the node *T1_T2* for markers FGA and TH01

H_1	0.002114
H_2	0.001568
H_3	0.996313
H_4	0.000003

Therefore,

$$P(S_2 \text{contr. to at least one of the traces} \mid \xi) = 1 - 0.000003 \cong 0.999997$$

when the whole information for the two traces on the three markers is taken into account a very significant value for the interest quantity is obtained.

CONCLUDING REMARKS

The use of DNA evidence analysis is commonly accepted nowadays in the whole courts. However, the presentation, interpretation and evaluation of this type of evidence sometimes raise some problems. And it is far the day when a total incorporation of this kind of evidence is achieved, although in some cases it has been decisive for the conviction or absolution of the individuals. This is already a good support for Justice.

The statistical treatment of criminal evidence has raised new challenges to those who have to decide, in the basis of the presented results. Independently of the methodology used, the great difficulty inhabits in the interpretation of the evidence, which is summarized in a number – what does that value means?

In the most complex problems, as the mentioned ones, the use of Bayesian networks for the analysis and interpretation of the evidence can be of great help. In a Bayesian network the complex inter-relations between the variables are transformed into modular units.

This technology – which use is everyday more and more common in different areas – supplies, as a support to the decision, a number. It does not give the decision; it is a decision support instrument. Consequently it is important that the legal system knows how to evaluate and interpret correctly the information contained in it. However, there is still much to do.

REFERENCES

1) R. G. Cowell, A. P. Dawid, S. L. Lauritzen, D. J. Spiegelhalter (1999). Probabilistic Expert Systems. Springer New Yorl.

2) A. P. Dawid, J. Mortera, V. L. Pascali, D. W. van Boxe (2002). Probabilistic expert systems for forensic inference from genetic markers . Scandinavian Journal of Statistics, 29, p577-595.

3) I. W. Evett, P. D. Gill, G. Jackson, J. Whitaker, C. Champod (2002). Interpreting small quantities of DNA: the hierarchy of propositions and the use of Bayesian networks. Journal of Forensic Sciences, 47, p520-530.

4) S. L. Lauritzen, J. Mortera (2002). Bounding the number of contributors to mixed DNA stains. Forensic Science International, 130, p125-126.

5) J. Mortera (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.

6) J. Mortera, A. P. Dawid, S. L. Lauritzen (2003). Probabilistic expert systems for DNA mixture profiling. Theoretical Population Biology, 63, p191-205.

7) R. E. Neapolitan (2004). Learning Bayesian Networks. Pearson Prentice Hall.

8) M. Andrade, M. A. M. Ferreira, J. A. Filipe (2008). Evidence evaluation in DNA mixture traces. Journal of Mathematics, Statistics and Allied Fields, 2 (2).

9) M. Andrade, M. A. M. Ferreira (2009). Bayesian networks in forensic identification problems. APLIMAT - Journal of Applied Mathematics, 2 (3), p13-30.

10) M. Andrade, M. A. M. Ferreira (2009). Criminal and Civil Identification with DNA Databases Using Bayesian Networks. International Journal of Security-CSC Journals, 3 (4), p65-74.

11) M. Andrade, M. A. M. Ferreira (2010). Solving civil identification cases with DNA profiles databases using Bayesian networks. Journal of Mathematics and Technology, 1(2), p37-40.

12) M. Andrade, M. A. M. Ferreira (2011). Some considerations about forensic DNA evidences. International Journal of Academic Research, 3 (1, Part I), p7-10.

13) M. Andrade, M. A. M. Ferreira (2011a). Crime Scene Investigation with Two Victims and a Perpetrator. In : K. Zhu, H. Zhang (Eds.), RECENT ADVANCE IN STATISTICS APPLICATIONS AND RELATED AREAS, PTS 1 AND 2 (Proceedings of 4th IISMES-International Institute of Statistics and Management Engineering Symposium, Dalian, China, July 24-29 2011). Aussino Academic Publishing House, Australia, p763-767.

14) M. Andrade, M. A. M. Ferreira (2014). Bayesian networks use in simple maternity problems. Applied Mathematical Sciences, 8(137-140), p6963-6967. **DOI**:10.12988/ams.2014.48617