

CRIME SCENE INVESTIGATION THROUGH DNA TRACES USING BAYESIAN NETWORKS

ANDRADE Marina, (PT), FERREIRA Manuel Alberto M., (PT)

Abstract. The use of biological information in crime scene identification problems is becoming more and more common. In this work, a crime scene is presented and used to exemplify the use of Bayesian networks, to analyse the information contained in a mixture DNA trace, referring to a crime in which there are two victims' and two suspect's involved. It is also made some discussion about the hypotheses to be considered in order to make them compatible with the possible information extracted from the mixture trace, that constitutes the evidence.

Key words: Mixture traces, forensic identification, Bayesian networks, DNA profiling.

Mathematics Subject Classification: 62C10.

1 Introduction

A crime has been committed and two persons, V_1 and V_2 , were murdered. One mixture trace was found and S_1 and S_2 are potential suspects. Their DNA profiles were measured and considered to be compatible with the mixture trace.

Possibly a fight occurred during the assault and some material was produced. It is acceptable that the individuals who perpetrated the crime could have left some of their genetic material in the trace.

The crime scene is analysed in section 2. It will be presented the evidence, E , and explained the hypotheses to be considered.

In section 3 the Bayesian network – example of a probability expert system - built expressly to perform the necessary calculations is shown¹. Then are discussed the number, quite large, and the significance of the results that is possible to obtain.

¹ To perform the calculations it is mandatory to apply repeatedly the Bayes' Law. This leads to very complicated computations impossible to perform algebraically. So the use of a probabilistic expert system is needed.

In section 4 the numerical results will be seen. In section 5 a brief discussion is outlined.

With the geneticist Sewall Wright, in the beginning of the 20th century, began the use of networks transporting probabilities. In (Dawid *et al.*, 2002) it is described this new approach to problems of the kind of the one described above. The conception and use of Bayesian networks to analyse problems in forensic identification inference, was initially done there, followed by (Evetts *et al.*, 2002), (Mortera, 2003) and (Mortera *et al.*, 2003). The analysis of a crime scene analogous to the considered in this work, but with two victims’ and one perpetrator and two mixture traces was presented in (Andrade and Ferreira, 2009). About this subject see (Andrade and Ferreira, 2011a).

2 Evidence and Hypotheses

To summarize the evidence it is presented in Table 1 the DNA profiles of the victims’ and the suspect’s, V_1 , V_2 , S_1 , S_2 , and the trace found at the crime scene, E .

	V_1	V_2	S_1	S_2	E
VWA	14,15	16,17	14,14	14,17	14,15,16,17
D21S11	29,31.2	28,28	30,31.2	28,32.2	28,29,30,31.2,32.2
SE33	19,30.2	17,30.2	17,18	16,19	16,17,18,19,30.2

Table 1: **Two Victims’ and Two Suspect’s DNA Profiles and Evidence.**

In Table 2 the allele frequencies, for each marker found in the trace, are presented.

	p_{14}	p_{15}	p_{16}	p_{17}	
VWA	0.1101	0.1197	0.1827	0.2753	
	p_{28}	p_{29}	p_{30}	$p_{31.2}$	$p_{32.2}$
D21S11	0.1674	0.2136	0.2437	0.1138	0.0894
	p_{16}	p_{17}	p_{18}	p_{19}	$p_{30.2}$
SE33	0.0590	0.0660	0.0833	0.0868	0.0140

Table 2: **Allele frequencies.**

The allele frequencies in Table 2 were collected in the database “The Distribution of Human DNA-PCR Polymorphisms”.

The crime trace can contain DNA from up to four unknown contributors, in addition to the victims and/or the suspects.

If the DNA of S_i with $i = 1, 2$ is presented in the trace this will place him/her at the crime scene and consequently as one of the possible perpetrators.

The court has to determine if each suspect is or is not guilty. The hypotheses to be evaluated are:

H_1 : S_1 is a contributor to the trace but S_2 is not, given the evidence.

H_2 : S_2 is a contributor to the trace but S_1 is not, given the evidence.

H_3 : S_1 and S_2 are both contributors to the trace, given the evidence.

H_4 : Neither S_1 nor S_2 are contributors to the trace, given the evidence.

The respective events probabilities are called $p_{10}, p_{02}, p_{12}, p_{00}$, where 0 mentions the absence of the respective, in order, individual DNA in the trace. So:

If $p_{00} > p_{10} + p_{02} + p_{12}$ the two suspects are acquitted. If not it must be seen if $p_{12} > p_{10} + p_{02}$ case at which the two suspects are both placed at the crime scene. If not p_{10} must be compared with p_{02} . If $p_{10} > p_{02}$ the evidence favours the presence of S_1 at the crime scene and the acquaintance of S_2 . The contrary happens when $p_{02} > p_{10}$.

3 Bayesian Network for One Marker

The probabilities referred above will be computed using the Bayesian network of Figure 1.

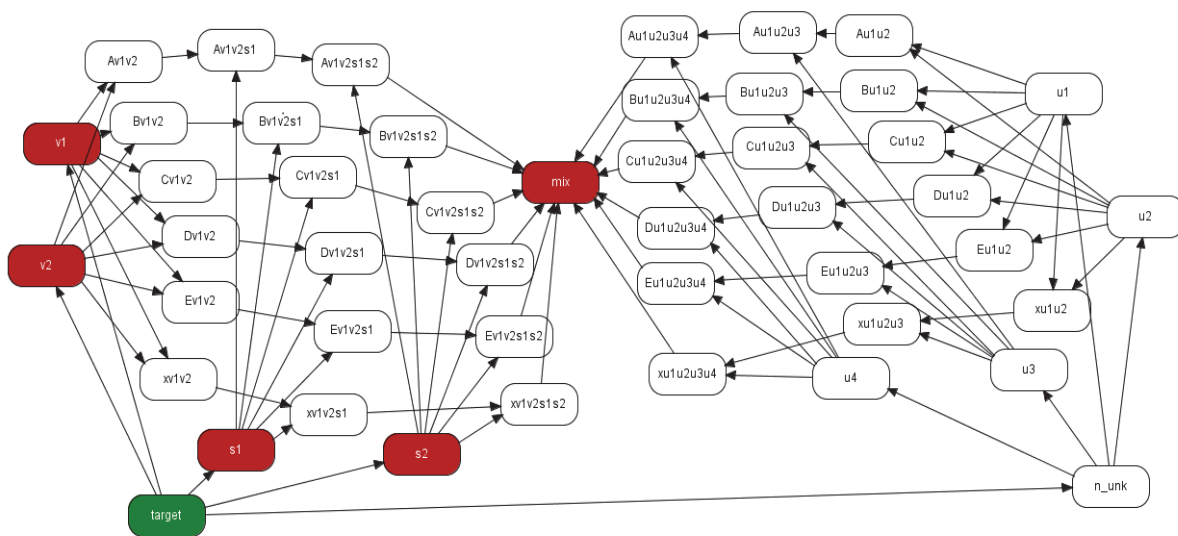


Figure 1: Marker network.

Nodes $v_i, i = 1, 2, s_j, j = 1, 2$ and $u_k, k = 1, 2, 3, 4$, in Figure 1 are themselves Bayesian networks that represent the genetic structure and inheritance of each individual - the victims, the suspects and the unknowns, respectively - and have all the same structure. The $v_i, i = 1, 2$ and $s_j, j = 1, 2$ constitute data of the problem. The nodes in white, at the left of the node mix, that represents the mixture and it is also comprised by known data (E), represent the relations in which the nodes $v_i, i = 1, 2$ or $s_j, j = 1, 2$ may contribute to the mixture. The nodes in white, at the right of the node mix, except the $u_k, k = 1, 2, 3, 4$ and n_unk - that is a counter for the number of unknowns in the mixture - represent the relations in which the $u_k, k = 1, 2, 3, 4$ may contribute to the mixture. Node target collects the states and the respective probabilities.

As it is mandatory to consider the possible contribution of till four unknown individuals to the mixture, the number of admissible states jumps to 80, numbered from 0 - no one in the mixture - to 79 - the two victims, the two suspects and the four unknowns are all in the mixture. Of course these two states are unrealistic and there are other ones also unrealistic because are incompatible with the minimum number of contributors to the mixture, according to the evidence inserted. These unrealistic states are discarded by the network but have to be considered conceptually in its building.

Among the realistic states only a few ones are interesting to the problem: the corresponding to the hypotheses events defined above. The network was implemented using Hugin² software³.

4 Results

For marker VWA, alleles 14, 15, 16, 17 are considered, Table 1. And so they are represented in the Figure 1 Bayesian network by A, B, C, D, respectively. E is considered with 0 frequency. When considering marker D21S11, the alleles are 28, 29, 30, 31.2 and 32.2 corresponding to A, B, C, D, E. In marker SE33 the alleles are 16, 17, 18, 19, 30.2 corresponding to A, B, C, D, E. In the whole cases x accumulates the remaining frequencies of the non considered alleles for each marker.

The results obtained are presented in Table 3. The values in line rescale are constituted by the ratios of the products of the values in the respective column⁴ by the total sum of the four products. The values in this line are the used ones in the tests described in section 2.

	p_{00}	p_{12}	p_{10}	p_{02}
VWA	0.2134	0.2699	0.2473	0.2699
D21S11	0.0752	0.5354	0.1335	0.2554
SE33	0.0091	0.8868	0.0454	0.0585
Rescale	0.0011	0.9576	0.0112	0.0301

Table 3: **Results.**

Following the procedure recommended in section 2 the conclusion is that both suspects are placed at the crime scene – note the great value of $p_{12} = 0.9576$. For marker VWA, alone, S_2 is placed at the crime scene but S_1 is not. But note that the probability is not very convincing: p_{02} is only slightly greater than p_{10} . For markers D21S11 and SE33, each of them alone, both suspects are placed together at the crime scene, in a much more convincing way for SE33.

5 Discussion

The problems and the difficulties posed in the interpretation and evaluation of DNA evidence are very well outlined, for instance, in (Andrade and Ferreira, 2011) and (Lauritzen, 2003). As in general they are stated in probabilistic terms leads to some confusion to the judges, when they have to issue a decision, because of its difficulty in interpreting the meaning of the measure of

² www.hugin.com

³ To compute the interesting probabilities there must be considered the following states probabilities:

- p_{00} : 1, 2, 3, 16, 17, 18, 19, 32, 33, 34, 35, 48, 49, 50, 51, 64, 65, 66 and 67,
- p_{12} : 12, 13, 14, 15, 28, 29, 30, 31, 44, 45, 46, 47, 60, 61, 62, 63, 76, 77, 78 and 79,
- p_{10} : 8, 9, 10, 11, 24, 25, 26, 27, 40, 41, 42, 43, 56, 57, 58, 59, 72, 73, 74 and 75,
- p_{02} : 4, 5, 6, 7, 20, 21, 22, 23, 36, 37, 38, 39, 52, 53, 54, 55, 68, 69, 70 and 71

from the output given by Hugin after the inserted evidence.

⁴ It is possible to multiply the respective probabilities, for each marker, because it is assumed independence between and across marker, i.e., linkage and Hardy-Weinberg Equilibrium (Andrade, 2007).

probability⁵. In this situation the Bayesian approach is the most clear to explain the significance of the evidence through the comparison of the hypotheses likelihood ratios see (Ferreira and Andrade, 2009). So the use of Bayesian networks to compute the interesting probabilities is a natural option, due to the intractable algebraic manipulation, when attempting to use successively the Bayes' Law in very complicated situations.

When the inference tool is the hypotheses tests, in these problems, they must be defined for each type of problem. This was exemplified with the four hypotheses suggested for this crime scene investigation. Only after this definition results clear which probabilities must be computed among a huge of possible ones.

Note finally that this methodology allows to conclude for the absolution of a suspect but not for the conviction. Eventually it only places the suspect in the crime scene. In this case further police work must be made.

References

- [1] ANDRADE, M.: *A Estatística Bayesiana na Identificação Forense: Análise e avaliação de vestígios de DNA com redes Bayesianas*. Phd Thesis, ISCTE, 2007.
- [2] ANDRADE, M.: *A Note on Foundations of Probability*. Journal of Mathematics and Technology, vol. 1 (1), pp 96-98, 2010.
- [3] ANDRADE, M. and FERREIRA, M. A. M.: *Bayesian networks in forensic identification problems*. Aplimat - Journal of Applied Mathematics, vol. 2 (3), pp. 13-30, 2009.
- [4] ANDRADE, M. and FERREIRA, M. A. M.: *Some considerations about forensic DNA evidences*. International Journal of Academic Research, vol. 3, (1, Part I), pp 7-10, 2011.
- [5] ANDRADE, M. and FERREIRA, M. A. M.: *Evidence evaluation in DNA mixture traces possibly resulting from two victims and two suspects*. Portuguese Journal of Quantitative Methods, vol 2 (1), pp 99-103, 2011a
- [6] ANDRADE, M. and FERREIRA, M. A. M.: *Crime scene investigation with probabilistic expert systems*. International Journal of Academic Research, vol. 3 (3, Part I), pp 7-15, 2011b.
- [7] ANDRADE, M. and FERREIRA, M. A. M.: *Crime scene investigation with two victims and a perpetrator*. 4 th IISMES Conference, Dalian, P. R. China. 2011c. Forthcoming.
- [8] DAWID, A. P., MORTERA, J., PASCALI, V. L. and BOXEL, D. W.: Probabilistic expert systems for forensic inference from genetic markers. Scandinavian Journal of Statistics vol. 29, pp 577-595, 2002.
- [9] EVETT, I. W., GILL, P. D., JACKSON, G., WHITAKER, J. and CHAMPOD, C.: *Interpreting small quantities of DNA: the hierarchy of propositions and the use of Bayesian networks*. Journal of Forensic Science, vol. 47, pp 520-539, 2002.
- [10] FERREIRA, M. A. M. and ANDRADE, M.: *A note on Dawnie Wolfe Steadman, Bradley J. Adams, and Lyle W. Konigsberg, Statistical Basis for Positive Identification in Forensic Anthropology. American Journal of Physical Anthropology 131: 15-26 (2006)*. International Journal of Academic Research, vol. 1 (2), pp 23-26, 2009.
- [11] LAURITZEN, S. L.: *Bayesian networks for forensic identification Problems*. Tutorial 19th Conference on Uncertainty in Artificial Intelligence, Mexico, 2003.
- [12] MORTERA, J.: *Analysis of DNA mixtures using probabilistic expert systems*. In: P. J. Green, N. L. Hjort and S. Richardson (Eds.), Highly Structured Stochastic Systems. Oxford University Press, 2003.

⁵ This confusion is originated by the conflict: probability frequentist interpretation vs subjectivist probability interpretation (Andrade and Ferreira, 2009) and (Andrade, 2010).

- [13] MORTERA, J., DAWID, A. P. and LAURITZEN, S. L.: *Probabilistic expert systems for DNA mixture profiling*. Theoretical Population Biology, vol. 63, pp 191-205, 2003.

Current address

Marina Andrade, Professor Auxiliar

ISCTE – Lisbon University Institute

UNIDE - IUL

Av. Das forças armadas

1649-026 Lisboa

Telefone: + 351 21 790 34 05

Fax: + 351 21 790 39 41

e-mail: marina.andrade@iscte.pt

Manuel Alberto M. Ferreira, Professor Catedrático

ISCTE – Lisbon University Institute

UNIDE - IUL

Av. Das forças armadas

1649-026 Lisboa

TELEFONE: + 351 21 790 37 03

FAX: + 351 21 790 39 41

e-mail: manuel.ferreira@iscte.pt