Aplimat

Journal of Applied Mathematics and Engineering
volume 6 (2014)
ISSN: 1337-6365

A BAYESIAN METHODOLOGY TO STUDY THE SIMPLE MATERNITY SEARCH PROBLEM

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

Abstract. The use of DNA profiles in forensic identification problems are common procedure nowadays. It was in England, in 1995, where the first DNA profiles database was built. It gave rise to new challenges concerning the forensic identification. Here it is intended to exemplify how to use the analysis of DNA profiles to solve the problem of simple maternity search. For this aim it is necessary to make use of a probabilistic expert system (PES), in this case an object-oriented Bayesian network (OOBN).

Key words. *DNA profiles, Bayesian networks, maternity search.*
Mathematics Subject Classification: MSC2010 62F15

1 Introduction

The use of networks that "carry" probabilities began with the geneticist Sewall Wright in 1921. Their used was widespread in several forms in various areas as social sciences and economics. The models used in these sciences are, in general, linear. Examples are the Path Diagrams or Structural Equation Models (SEM). Non-linear models, called Bayesian networks or Probabilistic Expert Systems (PES), are usually used in artificial intelligence.

This work target is the simple maternity problem approach: to recognize if a woman is the mother of a child – in general a dead child. To solve the real problems data analysis is performed through object-oriented Bayesian networks, which are a PES example, using Hugin[1] software.

[1] www.hugin.com – OOBN a resource available in the Hugin 6.4 software.

2 Simple Maternity Search

From Jornal PÚBLICO 19.01.2012 (translated):

> According to Francisco Corte-Real, INML[2], last year were carried out 5709 kinship biological research exams, concerning 1217 judicial processes. In the previous year (2010) had been carried out 5595 exams in the context of 1379 examinations processes.
> The specialist stated that through these tests, paternities and maternities are looked for, being confirmed that the first are the most frequent. In the case of maternity examinations, these are requested in cases like crimes of infanticide when a newborn is found dead and it is necessary to identify the mother, but, according to Francisco Corte-Real, are much less frequent.
>
> According to this charge, approximately 90% of examinations are referring to doubts raised by parents about the paternity of the children and the remaining 10% are made in their private capacity.
> In the INML took place, only last year, a total of 4405 forensic examinations for biological processes involving criminal cases as 1165 rapes and murders. In 2010, had been carried out 4795 concerning 1105 examinations processes.
> Under 60 processes were conducted, for individual examinations of genetic identification, 203 (142 in 2010, to 88 processes), usually requested by prosecutors and for identification of corpses or parts.

End of citation[3].

The news above show how often, and in what context, the analysis of DNA profiles is used for forensic proposes in Portugal. In this paper it will be analysed the situation of maternity examinations cases, much less usual than the paternity ones, particularly the requested in cases like crimes of infanticide, when a new-born is found dead and it is imperative to identify the mother. It is supposed that are available the genotypes from the putative mother (*pmgt*) and from the child (*chgt*). This is generally called the simple maternity search.

So the hypotheses in dispute are:

H_P: The putative mother is the true mother of the child

vs.

H_D: The true mother of the child is another individual chosen randomly from the population, not related with the putative mother.

The evidence is $E = (chgt, pmgt)$ – the child genotype and the putative mother genotype. The *posterior odds* is, as in (Ferreira and Andrade, 2009),

$$\frac{P(H_P \mid E)}{P(H_D \mid E)} = \frac{P(E \mid H_P)}{P(E \mid H_D)} * \frac{P(H_P)}{P(H_D)} \tag{1}$$

and assuming $P(H_P) = P(H_D)$ as usual,

[2] INML – Instituto Nacional de Medicina Legal (National Legal Medicine Institute)

[3] http://www.publico.pt/Sociedade/mais-de-5700-testes-de-paternidade-realizados-em-2011-1529758

$$\frac{P(H_P \mid E)}{P(H_D \mid E)} = \frac{P(E \mid H_P)}{P(E \mid H_D)} \tag{2}$$

As only a single likelihood ratio is being considered it may be computed using a Bayesian network, Fig. 1, where the where the nodes **pg** and **mg** are of class *founder* (single node network which states are the observed alleles with the observed population frequencies). Nodes **pmgt** and **chgt** are of class *genotype* (representing the individuals genotypes). Nodes **tpg** and **tmg** specify if the corresponding allele belongs or not to the putative mother. If **ch_match_pm?** is true the child allele is identical to the one of the putative mother, if not it is chosen randomly in the population. Node **chmg** defines the Mendel inheritance being the allele of the individual chose randomly after the ancestral alleles. Node **chpg** is the other element of the child genotype pair chosen randomly from the population.

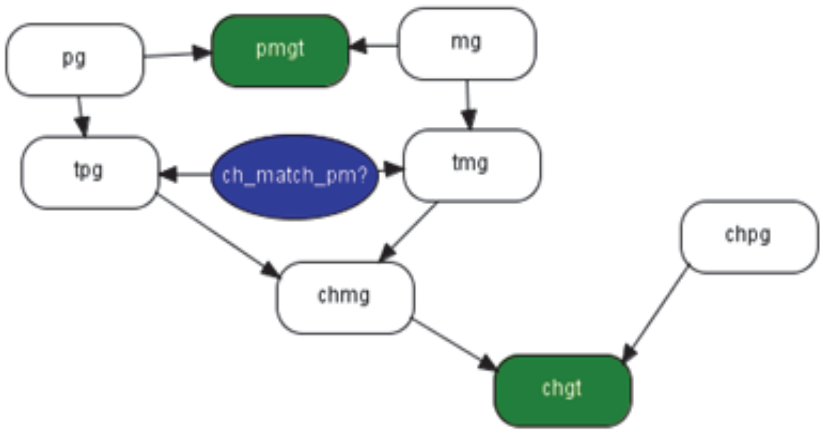

Figure 1: Simple maternity network.

3 Data and Results

To exemplify the application of the described tools, three markers (FGA, D21S11 and PENTA D) were chosen. In Table 1 are presented the genotypes and the allelic frequencies in the population.

Marker	**Allele Frequencies**				***chgt, pmgt***
FGA	p_{20}	p_{21}	p_{24}	p_{25}	(20, 24); (24,25)
	0.1421	0.1768	0.1325	0.0718	
D21S11	p_{27}	p_{29}	p_{30}	$p_{31.2}$	(27,31.2); (29,31.2)
	0.0246	0.2136	0.2437	0.1138	
PENTA D	p_{9}	p_{11}	p_{13}	p_{15}	(11, 13); (9,11)
	0.1984	0.1777	0.2066	0.0250	

Table 1: Genetic profiles and population frequencies for the chosen markers

The obtained results are presented in Table 2. Among many possible conclusions it must be highlighted that when there is a share of an allele with low frequencies in the population the probability of the yes hypothesis is very high, as it is expected.

Marker		**FGA**	**D21S11**	**PENTA D**	**Rescaled**
ch_match_pm	yes	0.7848	0.6872	0.5845	0.9185
	no	0.2152	0.3128	0.4155	0.0815

Table 2: Analysis results

References

[1.] ABRANTES, D., PONTES, M. L., PINHEIRO, M. F., ANDRADE, M. AND FERREIRA, M. A. M.. "Towards a systematic probabilistic evaluation of parentage casework in forensic genetics: A modest attempt to define a general standardized approach to simple and complex cases", Forensic Science International: Genetics Supplement Series 1, pp. 635-637, 2008.

[2.] ANDRADE, M.. "A Estatística Bayesiana na Identificação Forense – análise e avaliação de vestígios de DNA com redes Bayesianas", PhD Thesis, ISCTE, Lisboa, 2007.

[3.] ANDRADE, M.."A Note on Foundations of Probability", Journal of Mathematics and Technology, vol.. 1 (1), pp 96-98, 2010.

[4.] 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.

[5.] ANDRADE, M. and FERREIRA, M. A. M.. "Civil Identification Problems with Bayesian Networks Using Official DNA Databases", Aplimat-Journal of Applied Mathematics, vol. 3 (3), pp. 155-162, 2010.

[6.] ANDRADE, M. and FERREIRA, M. A. M.. "Evaluation of Paternities with less usual Data using Bayesian Networks", BMEI 2010 – 2010 3rd International Conference on Biomedical Engineering and Informatics. Proceedings, vol.6, pp. 2475-2477, 2010.

[7.] ANDRADE, M. and FERREIRA, M. A. M.. "Crime Scene Investigation with Two Victims and a Perpetrator", Recent Advance in Statistics Application and Related Areas – Conference Proceedings of the 4th International Institute of Statistics & Management Engineering Symposium, Dalian, China, 24-29 July, PART I, pp. 762-767, 2011.

[8.] ANDRADE, M., FERREIRA, M. A. M. and FILIPE, J. A.. "Evidence evaluation in DNA mixture traces", Journal of Mathematics and Allied Fields (Scientific Journals International-Published online), vol. 2 (2), 2008.

[9.] ANDRADE, M., FERREIRA, M. A. M., FILIPE, J. A. and COELHO, M.. "Paternity dispute: is it important to be conservative?", Aplimat – Journal of Applied Mathematics, vol. 1 (2), 2008.

[10.] DAWID, A. P., MORTERA, J., PASCALI, V. L. and van BOXEL, D. W.. "Probabilistic expert systems for forensic inference from genetic markers". Scandinavian Journal of Statistics vol. 29, pp. 577-595, 2002.

[11.] 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.

[12.] LAURITZEN, S. L.. "Bayesian networks for forensic identification Problems", Tutorial 19th Conference on Uncertainty in Artificial Intelligence, Mexico, 2003.

Current address

ANDRADE, Marina
Department of Quantitative Methods
Instituto Universitário de Lisboa (ISCTE-IUL), BRU-IUL
Lisboa, Portugal
e-mail: marina.andrade@iscte.pt

FERREIRA, Manuel Alberto M.
Department of Quantitative Methods
Instituto Universitário de Lisboa (ISCTE-IUL), BRU-IUL
Lisboa, Portugal
e-mail: manuel.ferreira@iscte.pt