Dipartimento di Statistica
"Giuseppe Parenti"

WORKING PAPER 2003/09

Bayesian Networks for Forensic Identification Via a Data Base Search of DNA Profiles

David Cavallini, Fabio Corradi

Università degli Studi
di Firenze

Statistics

Dipartimento di Statistica "G. Parenti" – Viale Morgagni 59 – 50134 Firenze - www.ds.unifi.it

Bayesian Networks for Forensic Identification Via a Data Base Search of DNA Profiles

David Cavallini[*]

Department of Statistics
University of Florence
50134, Italy

and

Fabio Corradi[†]

Department of Statistics
University of Florence
50134, Italy

November, 2003

Keywords:
Bayesian Network, Forensic Identification, database search

SUMMARY. In this paper we evaluate the evidence for pairs of competitive and exhaustive hypotheses obtained considering a characteristic observed on a crime sample and on individuals contained in a database. The subject considered here takes into account a debate which recently appeared in the literature concerning the appropriateness of different sets of hypotheses. First we demonstrate the problem via a computational efficient Bayesian Network (BN) obtained transforming some recognized conditional specific independencies into conditional independencies. Moreover in the proposed BN the sets of hypotheses proposed in the literature are included in the BN so that their role is better understood. Our BN is first proposed for a generic dichotomous characteristic but we are particularly interested in considering inheritable DNA traits. In this respect we show how to use the BN to evaluate the hypotheses that some individuals, who genetically related to the members of the database, are the donors of the crime sample.

1. Introduction

The forensic identification problem considered in this paper arises when a crime sample has been found but there is no clue about its origin. One possible step is to move on consists in a search a database (DB) of identified people, in the hope of finding suggestions about the origin of the trace.

[*] *email:* cavallin@ds.unifi.it
[†] *email:* corradi@ds.unifi.it

The topic has already been examined in the literature achieving surprisingly different conclusions. Two approaches seem to appear leading to a debate about the choice of the mutually exclusive hypotheses of interest.

To clarify the positions, consider the simple but relevant case occurring when only one of the elements in the DB matches the crime sample.

The National Research Council I Report (NRCI, 1992), followed by a revised version (NRCII, 1996), extended the standard solution to identification provided when there is a suspect to a DB search. Their proposal is based on the probability of a chance finding of an individual identical to the crime sample in a population of suspects (a so-called "match"). To cope with a DB search, NRCI and NRCII started from the true assertion that the probability of finding a match by chance is more probable in a DB than in a single drawn from the population; so, the probability that one person in the DB matches the crime sample by chance is considered as the probability of the data conditional to the defendant hypothesis. Conversely, the probability that one person in the DB matches because he/she is the origin of the trace, is regarded as the probability of the data conditional to the prosecutor hypothesis. From these premises the weight of evidence (WE), i.e., the ratio between the latter and the former of probabilities, is found to be inversely proportional to the DB size (NRCII, pag. 161). This is acceptable if the hypotheses under which the WE is evaluated are:

- H_p : "One of the people in the DB matches since he/she is the origin of the trace, i.e., the origin of the trace is in the DB";
- H_d : "One of the people in the DB matches by chance, i.e., the origin of the trace is not in the DB".

As a corollary, if a match is found and the DB contains almost the entire population of suspects, $P(H_p) \approx 1$ and $P(H_p \mid Evidence)$ cannot be very different. This means that nevertheless we are almost sure to have found the origin of the trace, a single match has a very low impact on $P(H_p)$, so the WE is small.

The second stream of contributions originated from the general approach to forensic identification provided by Balding and Donnelly (1995) and Dawid and Mortera (1996). In their papers, for each generic observed individual, say i, a pair of hypotheses are considered:

- $H^{'}_{i,p}$: "The origin of the trace is i, a well identified person";
- $H^{'}_{i,d}$: "Someone else, with respect to i, is the origin of the trace".

Starting from a prior for the hypotheses for all the members of the DB a posterior odds and a WE are derived.

This very detailed solution was apparently lost in the Balding and Donnelly (1996) contribution which was expressly devoted to the evaluation of a WE in a DB search. Focusing again on the single match case, they considered those individuals who were not compatible excluded from the search. This led to a pair of alternative hypotheses:

- $H_p^{'}$: "The origin of the trace is the matching person".
- $H_d^{'}$: "Someone else, out of the DB, is the origin of the trace".

Their conclusions, with explicit focus on the influence of the DB size on the WE, were exactly opposite to those reached by the NRCI and NRCII Reports. Balding and Donnelly (1996) argued that, the greater the DB size compared to the suspect population, the greater the number of the excluded individuals and the greater the WE provided by only one match.

Stockmarr (1999), strongly opposed to $H_p^{'}$ (and $H_d^{'}$) since he considers it impossible to set these hypotheses in advance with respect to the data. Thus, he considers the hypotheses H_p and H_d proposed by the NRCI and NRCII Reports to be valid. Moreover, he extended the analysis to the case involving more than one match, also distinguishing between finite and infinite populations. Obviously his results are on the track of those reached by NRCI and NRCII.

This further contribution to the debate has received many reactions. Among others, Dawid (2001) considered H_p too generic to bring before a judge, since it refers to the possibility that the DB contains the origin of the trace and the DB is not obviously on trial; in other words H_p does not address the problem in a useful way. Still he proposed the detailed set of hypotheses $H_{i,p}^{'}$ for all the observed individuals in the DB and made a very convincing logical distinction between the prosecutor hypotheses H_p and $H_p^{'}$. He noted that a difference between them appears only *before* the evidence is available: at that stage H_p considers the possibility that the origin of the trace may be in the DB, while $H_p^{'}$ considers the possibility that the matching person is the origin of the trace. After the evidence has been collected, H_p collapses into $H_p^{'}$ since, having excluded from suspicion all the people in the DB but one, this latter corresponds to the individual considered by $H_p^{'}$. Hypotheses having this behavior are called conditionally equivalent. Obviously, also that $H_i^{'}, p$ corresponding to the matching individual in the DB is conditionally equivalent to H_p. Moreover, if the prior on H_p, $H_p^{'}$ and $H_{i,p}^{'}$ are coherently specified by a unique set of assumptions, the same posterior odds are obtained.

Judging by this result, one might wrongly assume that the controversy is solved. The problem is that the evaluation of evidence is based on the WE which is usually considered an "objective measure" and is preferred to the posterior odds which require the elicitation of the priors. This delicate task is reserved for the jury. In this case, however, the point is different and concerns the choice from among the possible relevant hypotheses. Since, by definition, there is no clue about any of the persons in the DB, a no-informative prior is acceptable, however the prosecutor hypotheses have different priors as they refer to different sets of individuals: just one person for $H_p^{'}$ (and each $H_{i,p}^{'}$) and the size of the DB for H_p.

Since a WE is derived via the relation $WE = O(H|E)^{-1}O(H)$, obviously different WEs derive if we consider H_p instead of $H_p^{'}$ since, for these hypotheses, their posterior odds coincide but their corresponding priors do not.

Starting from an intuitive BN representation of the DB search problem for a binary characteristic, we provide a computationally efficient network obtained transforming some recognized conditional specific independencies (Geiger and Heckerman, 1996) into conditional independencies (Section (3)). Then, we extend the model to more complex genetic traits (Section (4)) and we exploit the inheritance between individuals of the same lineage to extend the search to some relatives of the individuals in the DB (Section (4)).

Finally, we propose a simulation study using a real DB (Section (5)) and some conclusions are drawn (Section (6)).

2. Background and definitions

A BN, $\mathfrak{B}_{\mathbf{U}}(\mathcal{D},\mathbb{P})$ or more succintly $\mathfrak{B}_{\mathbf{U}}$, is defined as a pair of objects: a Directed Acyclic Graph (DAG), $\mathcal{D}$, whose nodes, $\mathbf{U}$, represent discrete random variables, and a set $\mathbb{P}$ of Conditional Probability Tables (CPT) which define the conditional distributions of each vertex given the parents.

Every node is independent of its non-descendants conditional to the parents, so the joint distribution of $\mathbf{U}$ can be factorized as a product of CPTs (Pearl, 1988). Many other conditional independence assertions can be read from the network using the *d-separation* criterion (Pearl, 1988).

One of the main advantages of codifying a probabilistic model through a BN is the reduction of the computational efforts for calculating the conditional probability of the interesting unobserved nodes (*query* variables) given the observed ones (*evidence*). This task can be achieved using some different *Propagation Algorithms* such as the Junction Tree (Jensen, 2001) and the Bucket Elimination Variable (Dechter, 1999).

In regard to notation, upper-case letters denote random variables and corresponding lower-case letters are used to indicate a specified event or state. The vectors of random variables are denoted with bold upper-case letters and a particular realization or configuration is indicated with bold lower-case letters. Finally, lower-case Greek letters represent parameters.

In order to proceed in our formal discussion the following definitions are needed.

DEFINITION 2.1. *Let* $\mathbf{X}$ *and* $\mathbf{Z}$ *be two disjoined sets of random variables. The BN* $\mathfrak{B}_{\mathbf{X}\cup\mathbf{Z}}(\mathcal{D}^\star,\mathbb{P}^\star)$ *is* Probabilistic Equivalent*, (PE), to* $\mathfrak{B}_{\mathbf{X}}(\mathcal{D},\mathbb{P})$ *with respect to* $\mathbf{X}$*, if and only if*

$$P(\mathbf{X}) = \sum_{\mathbf{Z}} P^\star(\mathbf{X},\mathbf{Z}). \qquad (1)$$

DEFINITION 2.2. *Let* $\mathbf{X}$ *and* $\mathbf{Z}$ *be two disjoined sets of random variables. The BN* $\mathfrak{B}_{\mathbf{X}\cup\mathbf{Z}}(\mathcal{D}^\star,\mathbb{P}^\star)$ *is* Specific Probabilistic Equivalent*, (SPE), to* $\mathfrak{B}_{\mathbf{X}}(\mathcal{D},\mathbb{P})$

with respect to $\mathbf{X}$ *and a configuration* $\mathbf{e}$ *of* $\mathbf{E} \subset \mathbf{X}$*, if and only if*[1]

$$P(\mathbf{X} \backslash \mathbf{E}, \mathbf{e}) \propto \sum_{\mathbf{Z}} P^{\star}(\mathbf{X} \backslash \mathbf{E}, \mathbf{e}, \mathbf{Z}). \tag{2}$$

DEFINITION 2.3. *Let* $\mathbf{X}$*,* $\mathbf{T}$*,* $\mathbf{Y}$ *and* $\mathbf{Z}$ *be four disjoined sets of random variables. The BN* $\mathfrak{B}_{\mathbf{U}^{\star}}(\mathcal{D}^{\star}, \mathbb{P}^{\star})$*, where* $\mathbf{U}^{\star} = \mathbf{X} \cup \mathbf{T} \cup \mathbf{Z}$*, is* Artificial Probabilistic Equivalent*, (APE), to* $\mathfrak{B}_{\mathbf{X} \cup \mathbf{Y}}(\mathcal{D}, \mathbb{P})$ *with respect to* $\mathbf{X}$ *and a configuration* $\mathbf{y}$*, if and only if there exists a realization* $\mathbf{z}$ *such that*

$$P(\mathbf{X}, \mathbf{y}) \propto \sum_{\mathbf{T}} P^{\star}(\mathbf{X}, \mathbf{T}, \mathbf{z}). \tag{3}$$

The above definitions establish some probabilistic relations among the BNs defined on different domains. The relevance of these concepts concerns the probability of updating of a set of shared query variables when a set of nodes, that is fixed in advance, receives evidence. In fact, if (1) holds, then for any evidence on a subset of $\mathbf{X}$, the posterior probability of the unobserved nodes of $\mathbf{X}$ can be obtained indifferently using the BNs $\mathfrak{B}_{\mathbf{X}}$ or $\mathfrak{B}_{\mathbf{X} \cup \mathbf{Z}}$. Instead, condition (2) provides a milder relation; the result of the propagation, with respect to the set $\mathbf{X} \backslash \mathbf{E}$, on the two BNs $\mathfrak{B}_{\mathbf{X}}$ and $\mathfrak{B}_{\mathbf{X} \cup \mathbf{Z}}$ is the same if and only if the evidence on $\mathbf{E}$ is $\mathbf{e}$. Obviously, (1) implies (2) but the opposite is not true. Finally, a more sophisticated scenario is proposed in **DEFINITION (2.3)**. There, the set $\mathbf{X}$ is the intersection between the domains on which the two BNs, $\mathfrak{B}_{\mathbf{X} \cup \mathbf{Y}}$ and $\mathfrak{B}_{\mathbf{U}^{\star}}$, are built. Given a piece of evidence $\mathbf{y}$, the propagation on $\mathfrak{B}_{\mathbf{X} \cup \mathbf{Y}}$ and on $\mathfrak{B}_{\mathbf{U}^{\star}}$ provides the same posterior distribution on $\mathbf{X}$ if a particular configuration $\mathbf{z}$ exists such that (3) holds.

3. BN and forensic identification in a simplified setting

3.1 *The Island problem*

The evaluation of generic evidence in the forensic setting has received considerable attention in the last ten years. A very careful analysis is provided by Dawid (1994) and Dawid and Mortera (1996) with respect to the so-called *Island problem* originally proposed by Egglestone (1983): there, a certain binary characteristic or attribute, X, is observed on the crime scene (X_c) and the population of the possible donors is restricted to the $N + 1$ persons, the Island inhabitants.

For the j-th individual of the Island population a random variable X_j, with $j \in \mathbb{I} = \{1, 2, \ldots, N + 1\}$, is defined and the main focus of the analysis is to evaluate the probability that each inhabitant is the origin of the trace, given the evidence on $\mathbf{X} = \{X_j : j \in \mathbb{I}\}$ or a subset of it.

The most natural way to proceed is to define a discrete random variable H with $N + 1$ states, representing the originator status of each single inhabitant and to calculate its posterior distribution.

[1] The symbol \ denotes the topological subtraction operator among the sets.

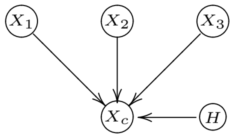


Figure 1. A DAG for the Island problem with only three inhabitants.

Moreover, the following assumptions are taken:

i. among the individuals, the characteristics distributions are independent given $\theta = P(X = 1)$, i.e., $\forall j \neq t$, $X_j \perp\!\!\!\perp X_t \mid \theta$;

ii. the characteristic is *pure*, i.e. is stochastically independent of the originator status of each inhabitant, $\forall j$, $X_j \perp\!\!\!\perp H \mid \theta$;

iii. for $H = j$ the characteristics of the rest of the inhabitants, $\mathbf{X}_{-j}$, are independent of the attribute observed on the crime scene, $\forall j$, $X_c \perp\!\!\!\perp \mathbf{X}_{-j} \mid H = j$ where $\mathbf{X}_{-j} = \{X_i : i \in \mathbb{I} \backslash \{j\}\}$;

iv. the observations are *symmetric error free*, $\forall j$, $\mathbf{Pr}(X_c = 1 \mid X_j = 1, H = j) = \mathbf{Pr}(X_c = 0 \mid X_j = 0, H = j) = 1$;

v. no other clue is available in advance, so the prior probability on H is no informative, $\forall j$, $\mathbf{Pr}(H = j) = 1/(N + 1)$;

Note that (*iii*) is a whole set of $N + 1$ independence statements: for each value of H a different assertion of independence holds. This form of independence is known as *Conditional Specific Independence* (CSI) (Geiger and Heckerman, 1996), which differs from the usual definition of conditional independence (Dawid, 1979), since, in the latter, the independence assertions between variables do not vary according to the values of the conditioning sets.

3.2 *A BN for the Island problem*

A BN for the Island problem can be built in a simple way. The domain $\mathbf{U}$ is comprised of $\mathbf{X}$, X_c and H. The graphical structure, which derives from the assumptions (*i*) and (*ii*), is summarized in the following four statements and by way of example in Figure (1).

A. $\forall j \neq t, X_j \nsim X_t$. B. $\forall j, X_j \nsim H$.
C. $\forall j, X_j \rightarrow X_c$. D. $H \rightarrow X_c$.

A Bernoulli distribution with parameter θ determines the marginal probability table of each node X_j and the distribution of H is defined with respect to the assumption (*v*). Furthermore, the CPT for X_c has a repetitive structure according to the CSI assumptions (*iii*) and the symmetric error free hypothesis (*iv*). The proposed naive network for the Island problem does not feature any conditional independence, so, for some evidence, the probability updating does

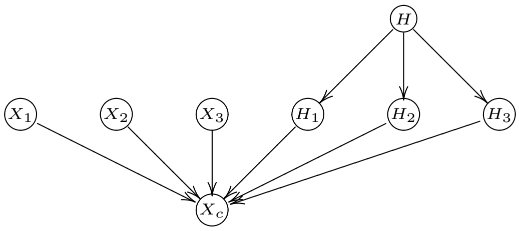


Figure 2. The augmented DAG obtained from Figure (1).

not take advantage of the graphical representation. Moreover, the size of the CPT of X_c increases exponentially with respect to the number of inhabitants so that the propagation becomes rapidly unfeasible. To improve the efficiency of the algorithm, some modifications are required. The task is achieved by introducing a set of instrumental nodes in such way that the new structure, built on the augmented domain, is APE to $\mathfrak{B}_{\mathbf{U}}$.

The result is attained in three steps.

Step 1. First, a set of binary random variables $\mathbf{H} = \{H_j : j \in \mathbb{I}\}$ is added and a new network, $\mathfrak{B}^\star_{\mathbf{U}^\star}$, is defined on the augmented domain $\mathbf{U}^\star = \mathbf{U} \cup \mathbf{H}$.

The new DAG, $\mathfrak{D}^\star$, is built considering the graph-theoretical statements (A)-(C) and on the following:

E. $H \not\rightarrow X_c$.
F. $\forall j, H \rightarrow H_j$.
G. $\forall j \neq i, H_j \not\rightarrow H_i$.
H. $\forall j$ and $i, H_j \not\rightarrow X_i$.
I. $\forall j, H_j \rightarrow X_c$.

The augmented network derived from Figure (1) is illustrated in Figure (2).

The marginal distribution of the variables X_j and H are the same as in the original network and the remaining CPTs are defined as follows:

a. $\forall j$, $P^\star(H_j = 1 \mid H = i)$ is equal to 1 when $i = j$, otherwise is 0;

b. $\forall j$, $P^\star(X_c \mid \mathbf{X}, \mathbf{H} = \mathbf{0}_j) = P(X_c \mid \mathbf{X}, H = j)$ and $P^\star(X_c \mid \mathbf{X}, \mathbf{H} \neq \mathbf{0}_j) \neq 0$ where $\mathbf{0}_j$ is a vector with dimension $N + 1$, whose i-th element is 0 for $i \neq j$ and 1 for $i = j$.

Finally, the following proposition (Appendix A) holds:

PROPOSITION 3.1. *The BN $\mathfrak{B}^\star_{\mathbf{U}^\star}$ is PE to $\mathfrak{B}_{\mathbf{U}}$ with respect to* $\mathbf{U}$.

The CPTs attached to each node H_j, specified as in (a), is the probabilistic translation of the deterministic logical `if-then` relation, i.e., $\forall j$ `if` $H = j$ `then` $H_j = 1$ and $\forall i \neq j$, $H_i = 0$. Thus, each variable H_j represents the originator status for the j-th inhabitant and the deterministic relation is a consequence of the assumption that the characteristic observed on the crime scene was left by only one individual.

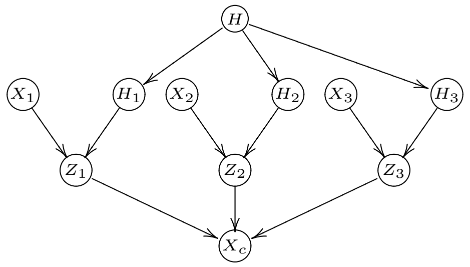


Figure 3. The augmented DAG of Figure (2) after the divorce.

Remark 1. This step has to be interpreted as preparatory because there are no computational motivations to use $\mathfrak{B}^\star_{\mathbf{U}^\star}$. The dimension of CPT attached to X_c, $2^{(3+2N)}$, is larger than the original one, $2^{(N+2)} \cdot (N+1)$, so that the calculation of the posterior probability of H given the evidence becomes more demanding.

Step 2. Here a *divorcing* technique (Jensen, 2001) is applied. The idea is to introduce a set of mediating variables between the parents and their child of a large converging connection. The role of the mediating variables is to lead some parents to divorce. The main advantage of this method is the reduction of the computational efforts because the original clique, $\{\mathbf{X}, X_c, \mathbf{H}\}$, is broken into a tree of smaller cliques.

In the Island problem, a reasonable way to divorce the parents of node X_c in the network $\mathfrak{B}^\star_{\mathbf{U}^\star}$, is to add $N+1$ mediating variables $\mathbf{Z} = \{Z_j : j \in \mathbb{I}\}$, so that each pair of variables X_j and H_j are married.

More formally, the new DAG, $\mathfrak{D}^+$, is built on the set of nodes $\mathbf{U}^+ = \mathbf{U}^\star \cup \mathbf{Z}$ and is formulated according to the statements (A)-(B), (E)-(H) and to the following ones:

L. $\forall j, Z_j \not\to H$. M. $\forall j \neq i, Z_j \not\to Z_i$.
N. $\forall j, H_j \to Z_j$ and $X_j \to Z_j$. O. $\forall j \neq i$ and $\forall j \neq t, Z_j \not\to X_i$ and $Z_j \not\to H_t$.
P. $\forall j, Z_j \to X_c$.

In Figure (3), the network of Figure (2) is divorced.

Concerning the CPTs, the marginal distributions of the variables X_j and H are unchanged with respect to the previous BN, the probabilistic assertion (a) still holds as well as the following:

c. $\forall j$, Z_j is binary,

d. $\forall j$, $P^+(Z_j = 1 \mid X_j = 1, H_j = 1) = P^+(Z_j = 0 \mid X_j = 0, H_j = 1) = 1$,

e. $\forall j$, $Z_j \perp\!\!\!\perp X_j \mid H_j = 0$ and $P^+(Z_j = \{0,1\} \mid H_j = 0) \neq 0$.

Finally, the CPT related to X_c has two different specifications:

f1. $P^a(X_c = 1 \mid \mathbf{Z} = \mathbf{1}) = 1$ and $P^a(X_c = 1 \mid \mathbf{Z} \neq \mathbf{1}) = 0$ where $\mathbf{1}$ is the unitary vector with dimension $N+1$;

f2. $P^o(X_c = 0 \mid \mathbf{Z} = \mathbf{0}) = 1$ and $P^o(X_c = 0 \mid \mathbf{Z} \neq \mathbf{0}) = 0$ where $\mathbf{0}$ is a vector with dimension $N+1$ whose elements are zero.

The symbols $\mathfrak{B}^a_{\mathbf{U}+}$ and $\mathfrak{B}^o_{\mathbf{U}+}$ are used for denoting respectively the BN built in accordance with the constraints (f1) and (f2). Note that (f1) is the probabilistic translation of the deterministic `and` relation and (f2) is a probabilistic representation of the logical `or` relation.

The following proposition (Appendix B) establishes the probabilistic relation between the pair of BNs $(\mathfrak{B}^a_{\mathbf{U}^+}, \mathfrak{B}^o_{\mathbf{U}^+})$ and the network $\mathfrak{B}^\star_{\mathbf{U}^\star}$.

PROPOSITION 3.2. *The BN $\mathfrak{B}^a_{\mathbf{U}+}$ ($\mathfrak{B}^o_{\mathbf{U}+}$) is SPE to $\mathfrak{B}^\star_{\mathbf{U}^\star}$ with respect to $\mathbf{U}^\star$ and the evidence $X_c = 1$ ($X_c = 0$).*

At this point the task of building a network able to perform local computations is achieved but the result is unwieldy since the BN to be used depends on the evidence on X_c. This drawback is overcome in the next step.

Step 3. In both networks, $\mathfrak{B}^a_{\mathbf{U}+}$ and $\mathfrak{B}^o_{\mathbf{U}+}$, when an evidence on X_c is entered, the relative reduced CPT can be written as a product of $N+1$ potentials, that is,

$$P^a(X_c = 1 \mid Z) = \prod_{j=1}^{N+1} \phi^a(Z_j) \tag{4}$$

$$P^o(X_c = 0 \mid Z) = \prod_{j=1}^{N+1} \phi^o(Z_j) \tag{5}$$

where, $\forall j$,

$$\phi^a(Z_j) = \begin{cases} 1 & \text{if } Z_j = 1 \\ 0 & \text{if } Z_j = 0 \end{cases} \tag{6}$$

$$\phi^o(Z_j) = \begin{cases} 0 & \text{if } Z_j = 1 \\ 1 & \text{if } Z_j = 0. \end{cases} \tag{7}$$

The potentials (6) and (7) can be interpreted as *findings*, i.e., tables whose elements are zeros or ones. Relations (6) and (7) establish that each variable Z_j takes values one or zero with probability 1. In other words, during the propagation, the evidence on X_c is transferred to each mediating variable Z_j.

The new DAG, $\mathfrak{D}^-$, is defined from $\mathfrak{D}^+$ simply by dropping the variable X_c and its incidental arcs. More formally, $\mathfrak{D}^-$ is built on the domain $\mathbf{U}^- = \mathbf{U}^\star \cup \mathbf{Z}$ considering the graph-theoretical statements (A)-(B), (E)-(H) and (L)-(O). Finally, the CPTs of the related network, $\mathfrak{B}^-_{\mathbf{U}^-}$, are specified with respect to the assumptions (a) and (c)-(e). As usual, the marginal distributions of the variables X_j and H remain unchanged.

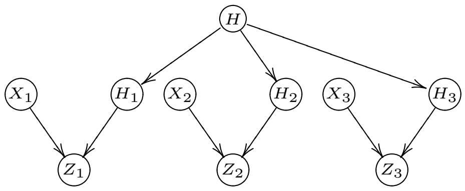


Figure 4. The network obtained after dropping the X_c node and the related incidental arcs from the DAG in Figure (3).

Figure (4) depicts the network of Figure (3) after the node X_c and the related incidental arcs have been dropped.

Finally, the next proposition (Appendix C) holds:

PROPOSITION 3.3. *For any evidence on* $X_c = \{0, 1\}$ *there exists only one correspondent realization of* $\mathbf{Z} = \{\mathbf{0}, \mathbf{1}\}$ *such that*

$$P(\mathbf{X}, H, X_c = \{0, 1\}) \propto \sum_{\mathbf{Y}} P^{-}(\mathbf{X}, H, \mathbf{Y}, \mathbf{Z} = \{\mathbf{0}, \mathbf{1}\}). \tag{8}$$

In other words, for every evidence on X_c*,* $\mathfrak{B}^{-}_{\mathbf{U}^{-}}$ *is APE to* $\mathfrak{B}_{\mathbf{U}}$ *with respect to* $\{\mathbf{X}, H\}$.

The **PROPOSITION 3.3** assures that, for any possible evidence on $\mathbf{X}$ and an observation on X_c (0 or 1) only one configuration of $\mathbf{Z}$ ($\mathbf{0}$ or $\mathbf{1}$) exists such that the posterior distribution of the variable H obtained using the networks $\mathfrak{B}_{\mathbf{U}}$ and $\mathfrak{B}^{-}_{\mathbf{U}^{-}}$ is the same.

The graph $\mathfrak{D}^{-}$ is featured by a repetitive structure with respect to the inhabitants of the Island. For each of them the same BN is built and all the networks are mixed by the hypothesis variable H which is the only parent of every H_j. Therefore, in the resulting network $\mathfrak{B}^{-}_{\mathbf{U}^{-}}$, a set of conditional independence assertions appears, i.e., given H, each triple (Z_j, H_j, X_j) is independent of the rest of the variables and so, for calculating the posterior distributions of H, local computations are allowed.

Finally, in the next proposition (Appendix D) the marginal distributions of the mediating variables are found.

PROPOSITION 3.4. *The marginal distribution of each variable* Z_j *is*

$$\begin{aligned} P(Z_j \mid \theta) = \frac{1}{N+1} \cdot [N \cdot P(Z_j \mid H_j = 0) + \theta \cdot P(Z_j \mid X_j = 1, H_j = 1) \\ +(1-\theta) \cdot P(Z_j \mid X_j = 0, H_j = 0)] \,. \end{aligned} \tag{9}$$

Table 1

Posterior probabilities for the identification hypothesis and implied WEs for some piece of evidence.

Evidence	**Posterior**	**WE**
$X_i = 1$	$(1 + N \cdot \theta)^{-1}$	θ^{-1}
$X_i = X_j = 1$	$(2 + (N-1) \cdot \theta)^{-1}$	$N/(1 + (N-1) \cdot \theta)$
$X_i = 1, X_{j \neq i \in \mathcal{I}} = 0$	$(1 + \theta \cdot (N-k+1))^{-1}$	$N/(N-k+1) \cdot \theta$

From (9) it is clear that the marginal distribution of each variable Z_j depends on $P(Z_j \mid H_j = 0)$ which, according to the condition (e), needs only to be greater than 0. Under the hypothesis of symmetric error free, if $P(Z_j = 1 \mid H_j = 0) = \theta$ then $P(Z_j = 1 \mid \theta) = \theta$, that is, the marginal distribution of each mediating variable is a Bernoulli with parameter θ. So, in this context, considering that the evidence on the vertex Z_j is set equal to the characteristic observed on the crime scene, the mediating nodes can be interpreted as a replication of X_c, but this interpretation is valid only if the symmetric error free assumption holds.

3.3 *First applications of the proposed BN to the Island problem*

In the Island problem, X_c is always observed but several different scenarios can arise with regards to the availability of the observed individuals forming a subset $\mathcal{I} \subseteq \mathbb{I}$, so that $\{X_c, X_{\mathcal{I}}\}$ is the evidence.

To illustrate some preliminary uses of the BN derived in Section (3.2), we consider three possibilities.

1. Only one individual is observed, $\mathcal{I} = \{i\}$ and he/she is found to have the characteristic under consideration. This possibility arises if the DB has only one element or a clue suggests that individual i be investigated, or i is drawn randomly from $\mathbb{I}$.
2. Two individuals are observed, $\mathcal{I} = \{i, j\}$. Both of them are found to have the characteristic under consideration and the observation mechanism follows one of the schemes depicted above.
3. Among k observed individuals only one of them matches the characteristic under consideration.

The posteriors for the identification hypothesis concerning individual i are provided in Dawid and Mortera (1996) and summarized in Table (1).

The same results can be obtained using the proposed network $\mathfrak{B}^{-}_{\mathbf{U}^{-}}$ which can be used to process every other possible piece of evidence.

Since sometimes the crime sample is observed with error, the assumption (iv) could not be valid, so more generally, instead of assumption (iv), we consider

$$iv^{\star}.\ \forall j,\ \mathbf{Pr}(X_c = 1 \mid X_j = 1, H = j) = \mathbf{Pr}(X_c = 0 \mid X_j = 0, H = j) = \beta < 1,$$

where symmetry is maintained. This scenario can be easily accommodated in the proposed BN, modifying (d) as follows:

d*. $\forall j$, $P^+(Z_j = 1 \mid X_j = 1, H_j = 1) = P^+(Z_j = 1 \mid X_j = 1, H_j = 1) = \beta$.

Remark 2. Dawid and Mortera (1996)) address the Island problem allowing for error-prone observations via $\mathbf{Pr}(X_j = 1 \mid X_c = 1, H_j = 1) = P^\star < 1$. It is not hard to establish a probabilistic connection between the two approaches, in fact, with respect to the BN proposed in this paper, the results obtained by Dawid and Mortera (1996) can be derived easily posing $P^\star = \beta{\cdot}\theta/(\beta{\cdot}\theta+(1-\theta){\cdot}(1-\beta))$.

3.4 *From the Island problem to a generic reference population*

The representation of all the inhabitants in the BN has illustrative purposes but it is not compulsory when the number of individuals for which the evidence is available is $k < N+1$. Without loss of generality, we assume to have observations for the first k individuals. Let $\mathcal{I} = \{1, 2, \ldots, k\}$.

Since $\forall i \neq j$, with $i, j \in \mathbb{I}\backslash\mathcal{I}$

$$P(H = i \mid X_c, X_{\mathcal{I}}) = P(H = j \mid X_c, X_{\mathcal{I}}), \tag{10}$$

a more parsimonious representation can be obtained by dropping the subnetworks related to the unobserved individuals and collapsing their correspondent H states in a residual one, which represents the hypothesis that the originator of the trace is in $\mathbb{I}\backslash\mathcal{I}$. In keeping with the assumption (v) the prior for the new state is equal to $1 - k/(N + 1)$.

Remark 3. The BN just obtained is able to embed the two different set of hypotheses (H_p,H_d) and ($H^{'}_{i,p}$, $H^{'}_{i,d}$). Also, looking at the matching person ($H^{'}_p$,$H^{'}_d$) can be monitored. In fact the first k states of H contains the whole set of hypotheses suggested by Dawid (2001), moreover the $k + 1$-th residual state is exactly the Stockmarr H_d hypothesis, so that, $P(H_p) = 1 - P(H_d) = 1 - P(H = k + 1)$.

A reference population is always difficult to define since the number of potential donors varies from case to case. When we are dealing with a DB search, usually, the extension of the reference population is not well-defined. Nevertheless, the introduction of a new node N and a directed arc $N \to H$, permits this kind of uncertainty to be taken into account. The CPT of H still follows condition (v) while the states of N and their prior probabilities depend on the specific case.

With regard to the uncertainty of the population characteristic, Corradi et al. (2003) show that if the sample from which the inference about θ is derived is large (> 300) then the effect on the hypothesis posteriors and the related WE is negligible, but care must be taken with regard to the possible heterogeneity of the examined characteristic. Note that the sample, drawn from the reference population, must not be confused with the DB, since the former has to be collected (approximately) at random and, typically, the identity of the contributors is not recorded.

4. BN and forensic identification through a search of DNA profiles in a DB

In Section (1) we introduced the DB search theme by means of the one-match-case. Actually, this outcome is especially valuable when the characteristic of interest is rare and the DB size is some order of magnitude smaller than the number of individuals in the population, since it is most usual to find no matches and there is only a very small probability of finding more than one match. This comment is still true when, instead of considering a generic dichotomous characteristic, the crime sample is a DNA profile. It follows that, depending on the set of assumed hypotheses, the proposals of Stockmarr (1999) and Balding and Donnelly (1996) solve the problem in the DNA setting as well. The possibility to recover these results as shown in section (3) does not seem essential, even if some assumptions adopted in those solutions could be easily relaxed making use of the modular BN structure.

To become effective, the detailed representation of all the individuals in the DB and the related hypotheses $H^{'}_{i,p}$ need to be considered jointly with the transmission of genetic information among relatives in a pedigree. This inheritance allows us to consider, as the possible donors of the crime sample, also individuals never typed but genetically related to the members of the DB. In this way, the most common but unfortunately also the last useful outcome of the DB search (the no-match case) could produce WEs different from zero for some *compatible* individuals. Compatible individuals are defined as those having a positive probability for the characteristic observed on the crime sample, conditional to all the available evidence. For instance, a member of a DB not matching the crime sample has a *compatible* child if he/she has an allele in common with the crime sample at each locus. Following this track, an augmented DB is defined and explored with respect to the set of hypotheses that one of its members is the origin of the crime sample.

4.1 *Background and notation for DNA evidence*

A DNA profile concerns measurements on several well specified locations of the DNA, called *loci*. At each locus we observe two alleles, one inherited from the father and the other from the mother, even if their origin is not recoverable. In this paper we assume independence of the alleles within each locus and between the loci, i.e., we assume Hardy Weinberg and linkage equilibrium. These assumptions are a simplification of reality but are also commonly assumed as an acceptable approximation.

For each individual $i \in \mathcal{I}$, we consider a pedigree constituted by the parents (i_0 and i_1), a child (i_c), a partner (i_p) and a brother (i_b). Here, the labels 0 and 1 are referred to a generic parent and not specifically to the mother or father because the information concerning inheritance is not available. Since this pedigree is built around i, we call it a *one-generation-around* pedigree and will be denoted with $i^{\star}$. A traditional pedigree representation of the outlined family is in Figure (5)

For a generic locus and for each $j \in i^{\star}$ we define two random variables A^0_j and A^1_j whose states, $a_1, a_2, \ldots, a_m$, are the inheritable alleles. In addition,

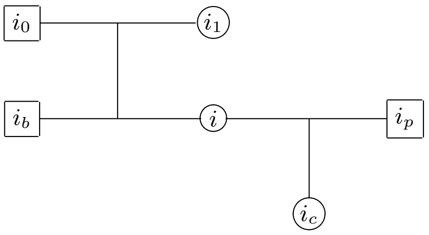


Figure 5. The one-generation-around pedigree.

$\forall j \in i^\star$ we consider a further random variable X_j whose states represent the genotypes, i.e., order pairs of alleles (a_t, a_u) with $t \leq u$.

Then, the one-generation-around DAG, pictured in Figure (6), is built explicitly considering the alleles transmitted from parents to children.

In order to obtain a full specification of the BN we need to specify three different kinds of CPTs.

1. The marginal distribution of the root nodes A_j^0 and A_j^1 with $j \in i^\star$. The probability of a generic allele a_t is estimated by means of the relative frequency of alleles observed in a sample taken from a specified reference population. The sample is often not random, nevertheless, the observations are considered to be exchangeable if weak information on the genetic structure of the population does not allow further structuring of observations. Anyway, care is typically taken to avoid the selection of strictly related individuals in the sample, such as siblings who are excluded from the sample in order to avoid the over-representation of highly correlated genotypes (Evett and Weir, 1998).

2. The conditional distribution of each no root node, A_j^0 and A_j^1, with $j \in i^\star$, given its parents in the graph, i.e., a pair of variables (A_v^0, A_v^1) with $v \in i^\star$ and $v \neq j$. By the first Mendelian, the related CPT is defined as follows

$$\mathbf{Pr}(A_j^s = a_r \mid A_v^0 = a_t, A_v^1 = a_u) = \begin{cases} 1 & \text{if } r = t = u \\ 0.5 & \text{if } r = t \text{ and } r \neq u \\ 0.5 & \text{if } r = u \text{ and } r \neq t \\ 0 & \text{otherwise.} \end{cases} \quad (11)$$

3. The conditional probability distribution of each variable X_j given A_j^0 and A_j^1 with $j \in i^\star$. This CPT is specified considering the deterministic relation between genotype and alleles, such that

$$\mathbf{Pr}(X_j = (a_r, a_u) \mid A_j^0 = a_h, A_j^1 = a_t) = \begin{cases} 1 & \text{if } h = r \text{ and } t = u \\ 1 & \text{if } h = u \text{ and } t = r \\ 0 & \text{otherwise.} \end{cases} \quad (12)$$

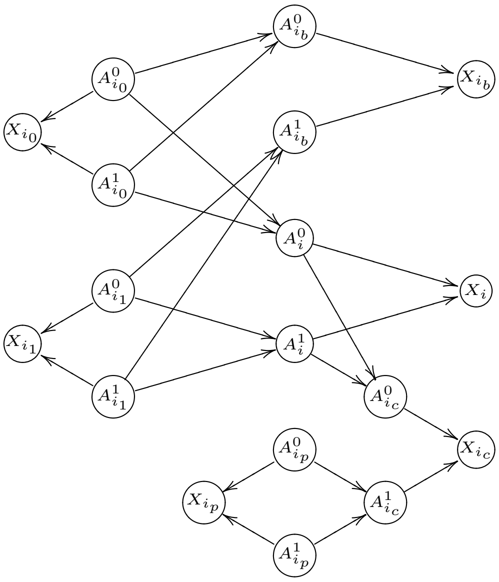


Figure 6. The one-generation-around network.

The one-generation-around pedigree can be represented as BN in yet two other ways. The first one, called Segregation Network, provides a more detailed representation of the genetic mechanism since a segregation indicator is introduced as the parent of each genotype node. The other possibility, Genotype Network, considers only the genotypes (X_j); the topology of the graph is very similar to the parents-child biological relationships depicted in Figure (5). Like Dawid et al. (2002), our choice to consider the BN of Figure (6) is motivated by computational considerations and by the fact that the loci used for identification purposes have codominant alleles. For more details and further comments about possible alternatives see Jensen (1997) and Lauritzen and Sheehan (2002).

4.2 *A search on the augmented DB*

To provide a one-generation-around search an augmented network, as shown in Figure (6), is built for each $i \in \mathcal{I}$ not recently related with other individuals in the DB. Then, in accord with the theory developed in Section (3) the genotype node of each individual in $\mathcal{I}^* = \{i^\star : i \in \mathcal{I}\}$ is linked to a mediating node and a dichotomous hypothesis variable is added.

In Figure (7) we represent the network, providing details only for the generic i-th family. Obviously, every node H_j with $j \in i^\star$ must be connected to the general hypothesis variable H (not represented in the figure) that also includes the possibility that the origin of the trace is outside $\mathcal{I}^*$.

For each individual in $\mathcal{I}^*$, the result of the search is the computation of the WE supporting the hypothesis that he/she is the origin of the trace. Under

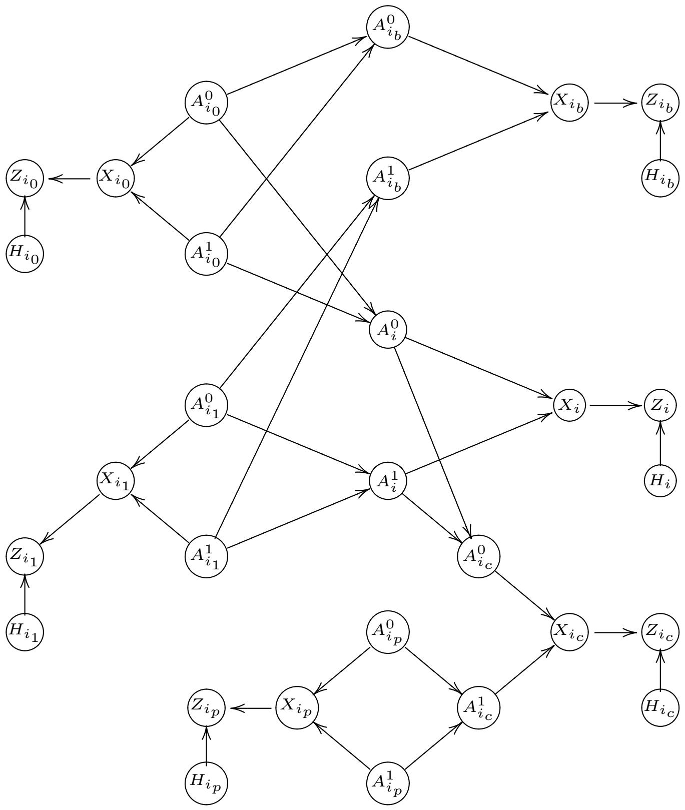


Figure 7. The extended network of Figure (6) with a related hypothesis system.

the H-W assumption, the product of the single WEs evaluated for each locus constitutes an overall measure of the genetic evidence. A WE cannot be read directly by the net, but it is a simple matter to derive it from the prior and posterior probabilities of the dichotomous hypotheses variables.

Some comments are in order. First, the adoption of the one-generation-around pedigree for each of the $\mathcal{I}$ members is a compromise between the inclusion of only direct descendants or ancestors (i.e., relatives for which an exclusion is possible) and to choose a more extensive search. A further concern arises with regard to apparent rigidity of the one-generation-around pedigree choice. This decision could not absorb some pieces of information available for some of the individuals in $\mathcal{I}^*$. This is a general problem concerning the relation between the elements in the DB and the population of possible donors of the crime sample. In fact, the DB is acquired independently of each single investigative case and, in the comparison with a crime sample originating from a specific case, we must ask if there are some pieces of information acts to discard some of the members of the DB as members of the donor population. This consideration held, either we consider the set $\mathcal{I}$ or its augmented version $\mathcal{I}^*$. In fact, considering $\mathcal{I}$, some of its members could remain excluded from the donor population; this exclusion could be reversible (e.g., they were in jail when the crime was perpetrated) or permanent (e.g. they are dead). The advantage to consider $\mathcal{I}^*$ is that we are in the position to discard i because he is out of suspect but not his relatives if they are considered belonging to the donors population. This refined state of information can be easily incorporated in the proposed BN simply working on the H node's CPT. If we want to exclude someone from the search we set to zero his corresponding prior probability. In this case, independently of the evidence, his posterior is also zero and the WE takes an undetermined form.

There are several ways to refine the analysis making use of the detailed and modular BN structure. One is to replace the uniform prior probabilities for each of the states of H linked to the $\mathcal{I}^*$ with some specific probabilities concerning the existence of each not observed individual, (e.g. the probability that a man has a child). Another possible refinement should be the introduction of more specific allele probabilities in a family for which we have information about the ethnicity of some of its members. Finally, we have assumed all the members of the DB to be not recently related: obviously if we know that some of them are in a well specified relative relation, instead of considering two or more different networks as specified in Figure (7), we instantiate the only one identified family network with all the available evidence.

5. An application using a real DB

Now we give account of some results obtained simulating a search on a real DB. The DB contains 100 observations on 13 loci. The members of the DB are assumed to be unrelated and we also assume that all the one-generation-around individuals belong to the donor population.

The size of the donor population was set to one million and the prior on H is assumed to be uniform.

For each observed individual, we generated two crime samples obtained respectively from the posterior marginal distribution of the child's and the sibling's genotypes. We call them the child-crime-samples and the sibling-crime-samples.

For every child-crime-sample, we evaluate two hypotheses: one strictly related to the identification of the child for each member of the DB, the other related to the possibility that the crime sample comes from a generic member of each one-generation-around family. Similar computations are provided if the sibling-crime-samples are used.

Results are stored in 100×100 matrices, where the columns label the crime samples and the rows the hypotheses. The matrices are shown as images (Figures 8(a)-8(b) and 9(a)-9(b)) on the gray scale of 65,535 levels. Darker gray levels correspond to higher WE values.

Another way to summarize the results is to provide for the main diagonal elements of the matrices, the ranks computed for each column (Table 2). The higher the rank, the higher the relative position of the *correct* identification hypothesis.

Concerning the identification of a child 98 out of 100 of the WEs supporting the *correct* identification hypothesis have the highest values; the remaining 2 have the second highest values. Identification of the family was a slightly less successful which was also the case for the identification of a sibling and the related family. In Table (2) we report the distribution of the ranks for the diagonal elements of each matrix. The first two columns concern the child-crime-sample with respect to the child hypothesis and to the more generic familiar hypothesis; the third and fourth columns concern the sibling-crime-sample and the appropriate hypotheses.

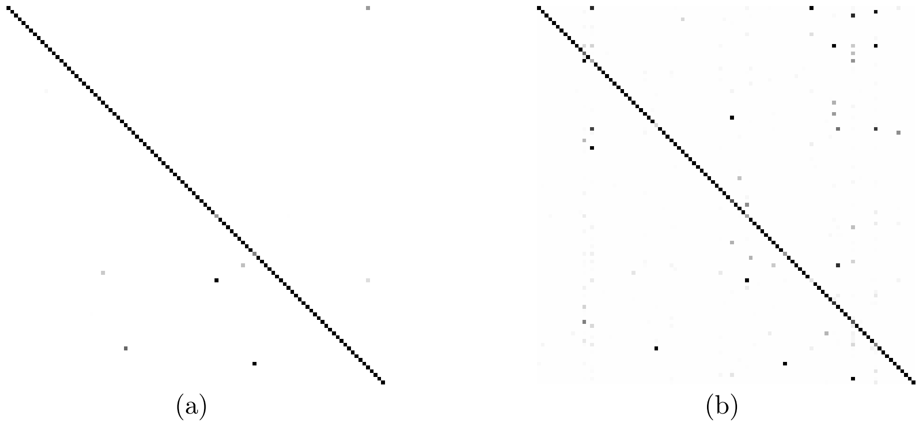
Whatever the evaluation of these results, it must be noted that our simulation is conservative in nature since, for instance, in sampling sibling-crime-sample we do not know the relatives' genotypes but we sample from their posterior distribution conditional to the genotype of just one of their children. In real cases where the "nature" knows the relatives' genotypes, brothers' genotypes are often very similar: for each locus, it suffices that only one of the parents is homozygote that the probability they share one allele is 1 and the probability they are identical is 0.5.

6. Conclusions

The use of BN to provide an evaluation of the weight of evidence for forensic identification purposes is a new but already well established approach. The consideration of several mutation models in paternity identification (Dawid and Pueschel, 1999), the possibility of considering mixtures of traces despite uncertainty as to the number of contributors, (Mortera et al., 2003), and the possibility to retain uncertainty on population parameters (Corradi et al., 2003) are just a few important examples of the potential of the techniques based on graphical models for solving the forensic identification issue.

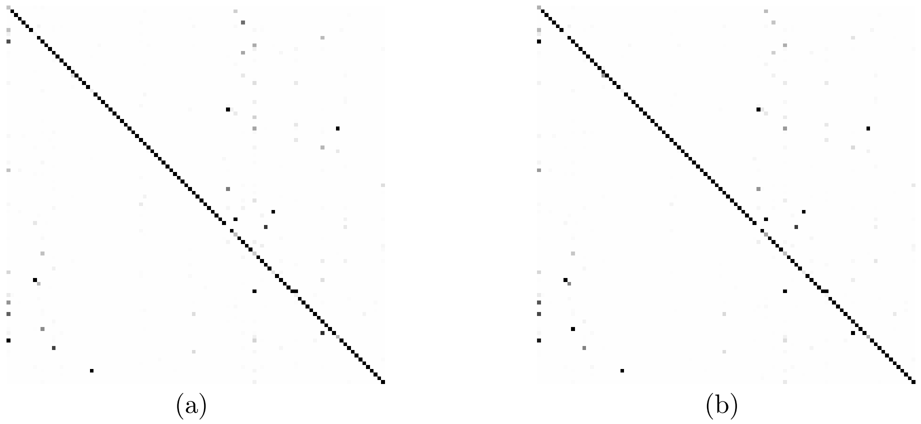
Here, the BN technology is invoked when there is no clue about the origin of the trace so that a suspect, or a group of suspects, is not available. When this occurs two approaches can be followed.

One, not considered here, consists in constructing a classification of the ref-



(a) (b)

Figure 8. WEs supporting the identification of a child: on the main diagonal the child-crime-sample is compared with the correct identification hypothesis, darker levels of gray correspond to higher WE values.



(a) (b)

Figure 9. WEs supporting the identification of a sibling: on the main diagonal the child crime sample is compared with the correct identification hypothesis, darker levels of gray correspond to higher WE values.

Table 2

The rank distributions of the WE supporting the correct identification hypothesis.

	Child-Crime-Sample		Brother-Crime-Sample	
Rank	Child	Family	Brother	Family
100	0.98	0.9	0.91	0.9
99	0.02	0.05	0.04	0.05
98	0	0.03	0.01	0.01
97	0	0.02	0.01	0.01
96	0	0	0.01	0.01
95	0	0	0	0.01
94	0	0	0.01	0
93	0	0	0.01	0.01

erence population in sub-groups and providing the probability that the crime sample belongs to each of them and can be considered the last resort of forensic identification, since only a geographical/ethnical response is expected. Experience in deconvolving a population making use of genetic data is provided, with different aims, by (Roeder et al., 1998, Dawid and Pueschel, 1999 and Pritchard et al., 2000).

A more favorable situation arises when there is the availability of a list of well identified individuals, the DB, not apparently related to the crime. This is the issue we have considered here and among the different positions held in the literature and summarized in the introduction. The first result we achieved was to embed the different hypotheses characterizing each of the approaches in only one model.

This result becomes effective when an augmented DB is introduced, having assumed that all its members belong to the population of possible donors of the crime sample, even if some of them are not observed. In this new perspective, and even if the one-match-case holds, H_p and $H_p^{'}$ are no longer conditionally equivalent since, in addition to the matching person, other unobserved individuals could have the same characteristic expressed by the crime sample; in other words, the posterior probability of finding the donor of the trace in the (augmented) DB is not, *a posteriori*, concentrated on only one person. This is *a fortiori* true if no match is found. Here, *a posteriori*, H_p concerns the possibility that, in the augmented database, one of the individuals is the origin of the trace but this (now perhaps interesting) probability does not coincide with any of the $H_{i,p}^{'}$.

Meester and Sjerps (2003) recently made a distinction between the evaluation of the case, measured by the posterior odds, and the pure evaluation of the evidence, provided by the WE. Since, for different sets of conditionally equivalent hypotheses, the evaluation of the case is not sensitive to a choice between members of that class but the WE is, they prefer the former result to the latter. These conclusions are formulated keeping in mind the one-match-case and a no-augmented DB. In that setting there is a clear indication towards a well identified person and the hypothesis of legal interest concerns this individual.

In the zero-match-case we do not expect to find strong evidence against a single member of the augmented DB but rather many different people having a positive WE. If the population size is much greater than the size of the DB, this produce very small priors for each of the considered individuals that finally leads to non-relevant posteriors. But nevertheless, if few WEs are considerably greater than one, this stimulates the acquisition of new evidence in a well-defined direction.

The extension of the DB search to inherited traits on an augmented DB was motivated by a real case study provided by the Raggruppamento Carabinieri Investigazioni Scientifiche (Italy). There, no match was found in the available DB but a striking similarity between the crime sample and one of the elements in the DB motivated the proposed extension. Incidentally, the results in terms of the obtained weights of evidence indicating the right track to carry on the investigative work.

A final remark concerns the fact that all the results are obtained in closed form, a compulsory requirement when the number of elements in the DB becomes in the order of thousands or more.

Appendix A
PROPOSITION 3.1

Proof. Showing that the following equation

$$\sum_{\mathbf{H}} P^\star(X_c \mid \mathbf{X}, \mathbf{H}) \cdot \prod_{j=1}^{N+1} P^\star(H_j \mid H = i) = P(X_c \mid \mathbf{X}, H = i) \tag{A.1}$$

holds $\forall i$, is equivalent to prove **PROPOSITION 3.1** because the marginal distributions of the variables X_j and H are the same in the two BNs, $\mathfrak{B}_{\mathbf{U}}$ and $\mathfrak{B}^\star_{\mathbf{U}^\star}$.

Furthermore, for the condition (a) the following result is straightforward

$$\sum_{H_t} P^\star(X_c \mid \mathbf{X}, \mathbf{H}) \cdot \prod_{j=1}^{N+1} P^\star(H_j \mid H = i) = \\ \prod_{j \neq t} P^\star(H_j \mid H = i) \cdot \begin{cases} P^\star(X_c \mid \mathbf{X}, \mathbf{H}_{-t}, Y_t = 1) & \text{for } t = i \\ P^\star(X_c \mid \mathbf{X}, \mathbf{H}_{-t}, Y_t = 0) & \text{for } t \neq i \end{cases} \tag{A.2}$$

where $\mathbf{H}_{-t} = \{H_j : j \in \mathbb{I} \text{ and } j \neq t\}$. So performing the marginalization of equation (A.1) in a recursive way with respect to each H_t we obtain that

$$P^\star(X_c \mid \mathbf{X}, \mathbf{H} = \mathbf{0}_i) = P(X_c \mid \mathbf{X}, H = i). \tag{A.3}$$

The proof is complete because equation (A.3) is true for condition (b).

Appendix B
PROPOSITION 3.2

Proof. Since the joint marginal distribution of the variables X_j, H_j and H is unchanged with respect to $\mathfrak{B}^\star_{\mathbf{U}^\star}$ then, considering the evidence $X_c = 1$, **PROPOSI-**

TION 3.2 is proved if and only if the following proportional equation holds:

$$P^{\star}(X_c = 1 \mid \mathbf{X}, \mathbf{H}) \propto \sum_{\mathbf{Z}} P^{+}(X_c = 1 \mid \mathbf{Z}) \cdot \prod_{j=1}^{N+1} P^{+}(Z_j \mid X_j, H_j) \tag{B.4}$$

Since the table $P^{+}(X_c = 1 \mid \mathbf{Z})$, for condition (f1), can be expressed as the product of the following $N+1$ potentials

$$\phi(Z_j) = \begin{cases} 1 & \text{if } Z_j = 1 \\ 0 & \text{if } Z_j = 0, \end{cases} \tag{B.5}$$

equation (B.4) can also be written in this way

$$\begin{aligned} P^{\star}(X_c = 1 \mid \mathbf{X}, \mathbf{H}) &\propto \prod_{j=1}^{N+1} \sum_{Z_j} P^{+}(Z_j \mid X_j, H_j) \cdot \phi(Z_j) \\ &\propto \prod_{j=1}^{N+1} P^{+}(Z_j = 1 \mid X_j, H_j). \end{aligned} \tag{B.6}$$

Given a valid configuration $(\mathbf{0}_i)$ for the $\mathbf{H}$ vector, for assumptions (b) and (*iii*), the left side of (B.6) is equal to $P(X_c \mid X_i, H = i)$, so, considering hypothesis (e), (B.6) becomes

$$P(X_c \mid X_i, H = i) \propto \prod_{j \neq i} P^{+}(Z_j = 1 \mid H_j = 0) \cdot P^{+}(Z_j = 1 \mid X_i, H_i = 1). \tag{B.7}$$

Finally, since the first term of the right side of (B.7) is a constant, for condition (e), comparing hypotheses (d) and (*iv*) the proof is complete. With similar arguments **PROPOSITION 3.2** can be proved if the network $\mathfrak{B}^{o}_{\mathbf{U}^+}$ and evidence $X_c = 0$ are considered.

Appendix C
PROPOSITION 3.3

Proof. From the comparison between the networks $\mathfrak{B}^{a}_{\mathbf{U}^+}$ and $\mathfrak{B}^{-}_{\mathbf{U}^-}$ the derivation of the next equation is straightforward,

$$P^{a}(\mathbf{X}, H, \mathbf{H}, \mathbf{Z}, X_c) = P^{a}(X_c \mid \mathbf{Z}) \cdot P^{-}(\mathbf{X}, H, \mathbf{H}, \mathbf{Z}). \tag{C.8}$$

Posing $X_c = 1$, from (4) and (6), (C.8) can be written in this way

$$\begin{aligned} P^{a}(\mathbf{X}, H, \mathbf{H}, \mathbf{Z}, X_c = 1) &= \prod_{j=1}^{N+1} \phi^{a}(Z_j) \cdot P^{-}(\mathbf{X}, H, \mathbf{H}, \mathbf{Z}) \\ &= P^{-}(\mathbf{X}, H, \mathbf{H}, \mathbf{Z} = \mathbf{1}). \end{aligned} \tag{C.9}$$

Note that, in the second equation of (C.9) all variables Z_j are set to 1 as each potential $\phi^{a}(Z_j)$ is a finding representing the evidence $Z_j = 1$.

Marginalizing both terms of (C.9) with respect to $\mathbf{Z}$, from **PROPOSITION 3.2**, the following proportional equation holds,

$$P^{\star}(\mathbf{X}, H, \mathbf{H}, X_c = 1) \propto P^{-}(\mathbf{X}, H, \mathbf{H}, \mathbf{Z} = \mathbf{1}). \tag{C.10}$$

So, considering **PROPOSITION 3.1**, the marginalization of (C.10) with respect to $\mathbf{H}$ completes the proof, that is, (8) is obtained. With similar arguments **PROPOSITION 3.3** can be proved when $X_c = 0$, simply using the network $\mathfrak{B}^o_{\mathbf{U}+}$ instead of $\mathfrak{B}^a_{\mathbf{U}+}$.

Appendix D
PROPOSITION 3.4

Proof. In order to obtain the marginal distribution of Z_j $P^-(\mathbf{X}, \mathbf{Z}, \mathbf{H}, H \mid \theta)$ must be marginalized with respect to $\mathbf{Z}_{-j} = \{Z_i : j \in \mathbb{I} \text{ and } i \neq j\}$, $\mathbf{X}$, $\mathbf{H}$ and H.

As each variable in $\mathbf{Z}_{-j}$ is a sink of $\mathfrak{D}^-$, the marginalization with respect to those variables is equal to one. For similar arguments every node X_i and H_i with $i \neq j$ disappears, so

$$P^-(Z_j \mid \theta) = \sum_H \sum_{X_j} \sum_{H_j} P^-(Z_j \mid X_j, H_j) \cdot P^+(H_j \mid H) \cdot P(X_j \mid \theta) \cdot P(H) \qquad \text{(D.11)}$$

Marginalizing (D.11) with respect to H and Y_j, for hypothesis (v), the following equation holds

$$\begin{aligned} P^-(Z_j \mid \theta) = \frac{1}{N+1} \cdot \sum_{X_j} P(X_j \mid \theta) \cdot & \\ \Bigg[P^-(Z_j \mid X_j, H_j = 1) \cdot \sum_{i=1}^{N+1} P^+(H_j = 1 \mid H = i) & \\ + P^-(Z_j \mid X_j, H_j = 0) \cdot \sum_{i=1}^{N+1} P^+(H_j = 0 \mid H = i) \Bigg] . & \qquad \text{(D.12)} \end{aligned}$$

Considering conditions (a) and (e), (D.11) becomes

$$\begin{aligned} P^-(Z_j \mid \theta) = \frac{1}{N+1} \cdot \sum_{X_j} P(X_j \mid \theta) \cdot \big[P^-(Z_j \mid X_j, H_j = 1) & \\ + P^-(Z_j \mid H_j = 0) \cdot N \big] . & \qquad \text{(D.13)} \end{aligned}$$

Finally, (9) is obtained marginalizing (D.13) with respect to variable X_j.

References

Balding, D. J. and Donnelly, P. (1995). Inference in Forensic Identification. *Journal of the Royal Statistical Society* **A158**, 21–53.

Balding, D. J. and Donnelly, P. (1996). Evaluating DNA Profile Evidence When the Suspect Is Identified Through a Database Search. *Journal of Forensic Science* **41**, 603–607.

Corradi, F., Lago, G. and Stefanini, F. M. (2003). The Evaluation of DNA Evidence in Pedigrees Requiring Population Inference. *Journal of the Royal Statistical Society* **A166**, 425–440.

Dawid, A. P. (1979). Conditional Independence in Statistical Theory. *Journal of the Royal Statistical Society* **B41**, 1–31.

Dawid, A. P. (1994). The Island Problem: Coherent Use of Identification Evidence. In Freeman, P. R. and Smith, A. F. M., editors, *Aspects of Uncertainty: A Tribute to D. V. Lindley*, pages 159–170. J. Wiley.

Dawid, A. P. (2001). Comment on Stockmarr's Likelihood Ratios for Evaluating DNA Evidence When the Suspect is Found Through a Database Search'. *Biometrics* **57**, 976–980.

Dawid, A. P. and Mortera, J. (1996). Coherent Analysis of Forensic Identification Evidence. *Journal of the Royal Statistical Society* **B58**, 425–443.

Dawid, A. P., Mortera, J., Pascali, V. L. and Boxel, D. V. (2002). Probabilistic Expert Systems for Forensic Inference from Genetic Markers. *Scandinavian Journal of Statistics* **29**, 577–595.

Dawid, A. P. and Pueschel, J. (1999). Hierarchical Models for DNA Profiling Using Heterogenous Databases (with Discussion). In *Bayesian Statistics 6: Proceedings of the Sixth Valencia International Meeting*, pages 187–212.

Dechter, R. (1999). Bucket Elimination: A Unifying Framework for Probabilistic Inference. In Jordan, M. I., editor, *Learning in Graphical Model*, pages 75–104. Kluwer Academic Publishers.

Egglestone, R. (1983). *Evidence Proof and Probability.* Weidenfeld and Nicholson.

Evett, I. W. and Weir, B. S. (1998). *Interpreting DNA Evidence.* Sinauer Associates.

Geiger, D. and Heckerman, D. (1996). Knowledge Representation and Inference in Similitary Networks and Bayesian Multinets. *Artificial Intelligence* **82**, 45–74.

Jensen, C. S. (1997). *Blocking Gibbs Sampling for Inference in Large and Complex Bayesian Networks with Applications in Genetics.* PhD thesis, University of Aalborg, Denmark.

Jensen, F. V. (2001). *Bayesian Network and Decision Graphs.* Springer-Verlag.

Lauritzen, S. L. and Sheehan, N. A. (2002). Graphical Model for Genetic Analyses. Technical Report R-02-2020, Department of Mathematical Sciences, University of Aalborg.

Meester, R. and Sjerps, M. (2003). The Evidential Value in the DNA Database Search Controversy and the Two Stain Problem. *Biometrics* **59**, 727–732.

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

NRCI (1992). National Research Council Committee on DNA Forensic Science. DNA Technology in Forensic Science. National Academy Press.

NRCII (1996). National Research Council Committee on DNA Forensic Science. An Update: The Evaluation of Forensic DNA Evidence. National Academy Press.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann Publishers.

Pritchard, J. K., Stephens, M. and Donnelly, P. (2000). Inference in Population

Structures Using Multilocus Genotype Data. *Genetics* **155**, 945–959.

Roeder, K. M., Escobar, M. and Kanade, J. (1998). Measuring Heterogeneity in Forensic Databases Using Hierarchical Bayes Models. *Biometrika* **85**, 268–287.

Stockmarr, A. (1999). Likelihood Ratios for Evaluating DNA Evidence When the Suspect is Found Through a Database Search. *Biometrics* **55**, 671–677.