Bayesian Networks for Forensic Analysis and The Law

Haya Toumy

March-26-2015

Contents

List of Figures

1 Abstract and Motivation

In most cases in forensics, scientists are faced with a large quantity of information coming from various observations, data, or more generally, findings related to the case under investigation. In many occasions, knowing some of these observations will change our belief about others. The scientists task is to help express a probabilistic conclusion on the joint value of such a quantity items of information or help the court of justice in expressing a belief on a judicial question of interest, typically expressed in terms of a proposition compared to a particular alternative.
The use of Bayesian networks is starting to be used to help lawyers understand the strength of a case when there are multiple dependent and conflicting pieces of evidence. The crucial graphical feature of a BN is that it tells us which variables are *not* linked, and hence, it captures our assumptions about which pairs of variables are not directly dependent.

2 Mathematical Framework

The statement of changing beliefs about some observations upon knowledge of others, are translated into conditional probabilities. However, setting up all the aspects of a case having large amounts of information we know about it along with the need of associating a reasoning process, can be a complicated task; especially under uncertainty. To approach this delicate inference problems in a coherent way, we need a probabilistic framework enjoying clarity of formulation and a thorough computational architecture. Since the picture is worth a thousand words, a graphical representation will be a concise description of a challenging practical problem; all that is the Bayesian networks.

A simple example: [1].

Assume we have a coin that can be either fair X_1, or having two heads X_2 or two tails X_3. Where X_i are random variables; and H is the hypothesis of having a head upon flipping the coin. We observe two things, the outcome of each toss in the sequence depends upon which hypothesis is true, and it is relevant to the degree of belief in the outcome of the next toss. These dependence relationships can be represented by means of the upper directed graph,

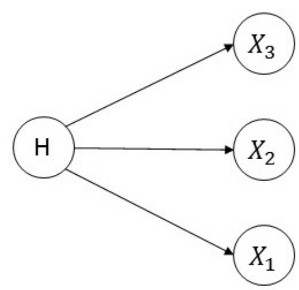


Figure 1: Simple example of a Bayesian network

3 Glossary

Each node in the graph is associated with one random variable in the Bayesian network. Nodes like H in the figure above, are called a *parent node* for X_1 and X_2 and X_3, and so these latter are called *child nodes* for H. A node can have no or many parents and/or many or no children. A farther parent in the graph is called *ancestor*; likewise, a farther child in the network is called a *descendant*. A node that has no parents is called **root** node. Undirected arcs (arrows) are called *edges*.
Root nodes are given marginal distributions. Other nodes have distributions conditional only on the

values of the respective parents. [3]
Cycles: A path forms a cycle if it starts and ends on the same node.
DAG: Directed Acyclic Graph, describes a graphical model where all nodes (random variables) are connected with directed arcs (arrows), and has no cycles.
The DAGs expresses only the conditional independence structure of a BN; and it is very useful to discuss the construction or the interpretation of a BN. Where it is meant by the BN structure: the set of its conditional independencies [3]
Leaf node is a node in a DAG without any child.
A **path**, is a set of successive links (arcs). It is usually described by an ordered sequence of linked nodes. All the arcs that comprises the path must follow the same direction.
Hard evidence is the information that fixes the state of a variable (node), otherwise it is called *'soft evidence'*.

4 Logical Relations

There are three possible connections by which information can travel through a variable in a directed graph: diverging, serial, converging. [1]

4.1 Diverging Connection

Also called: **Common Cause**.

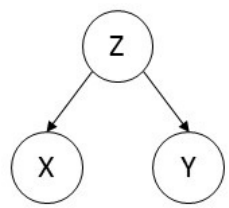


Figure 2: A common cause connection

This type is appropriate graphical model whenever it's believed that: 1) Z is relevant for both X and Y, and 2) X and Y are conditionally independent given Z (that is, when Z is known and fixed). This means that if the state of Z in known, then knowing state of X does not affect our belief about possible states of Y, and vice versa; i.e. when Z is fixed, no flow of information can be transmitted between X and Y. 3) Diverging connection is also appropriate when is believed that if the state of Z is unknown, then the knowledge of X provides information about the possible states of Y, and vice versa; i.e. information can be transmitted through the diverging connection. It is then said that X and Y are *d-separated* (directionally separated) given Z.

4.2 Serial Connection

Also called: **Causal and Evidential Trail**.

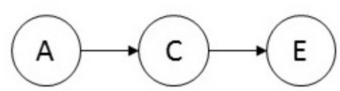


Figure 3: A causal and evidential trail

This one typically applies to 'chain' reasoning; e.g. a Markov chain. A serial connection is an appropriate graphical model whenever it's believed that 1) A is relevant for C, and 2) C is relevant for E, and 3) A and E are conditionally independent given C. This means that, as above, if the state of C is known, then the knowledge of state of A does not change the belief about the possible states of E, and vice versa; and it also means that if the state C is unknown, then information can flow between A and E, i.e. knowledge of A provides information about the possible states of E, and vice versa. In other words, fixation (knowledge) of C blocks the information transmission between A and E, i.e. A and E will be directionally separated. Therefore, also in this type of connections, it is said that A and E are *d-separated* given C.

Example:
Let A be the proposition *The suspect is the offender.* C be the proposition *The blood stain found on the crime scene comes from the suspect.* And E be the proposition *The suspect's blood sample and the blood stain from the crime scene share the same DNA profile.* Then A is relevant for C; and C is relevant for E; but there might be an explanation of the occurrence of the event described by C different from the occurrence of the event described by A. And given that C occurred, then the observed correspondence in DNA profiles does not depend on which explanation of C is true.

4.3 Converging Connection

Also called: **Common Effect**.

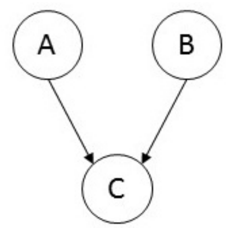


Figure 4: A common effect connection

The best way to characterize this type of connections in one's mind is to remember that it is 'to the contrary' of the diverging connection, in the context of the middle state being known and its effect on the flow of information through the graph. A converging connection is an appropriate graphical model whenever it is believed that 1) A and B are both relevant for C, and 2) A in not relevant for B; but it does become relevant if the state of C is known. In other words, when it is believed that A and B are *un*conditionally independent, but are conditionally dependent given C (when C is known). As before, this means that if we know C, then knowledge of the state of A does provide information about the possible states of B, and vice versa. But it also means that if we do not know C, then knowing A does *not* provide information about possible states of B, and vice versa. i.e. the channel (the flow of information) is blocked if the state of the middle variable is unknown (to the contrary to the diverging connection).

Example:
Let, A be the proposition that *The suspect is the offender*, C be the proposition that *The blood stain found on the crime scene comes from the suspect*, and B *The blood stain found on the crime scene comes from the offender*. Then knowledge of either A or B occurred does not provide information about the occurrence of the other. The fact that the source of the blood stain is the offender has no bearing at all, if taken alone, on the guilt of the suspect. But if C is true, then A and B become obviously related.

Another peculiarity of the converging connections is that, to open the channel (allowing information transmission), it is not necessary for the state in the middle to be known for certain, but it suffices that there is some information, even if not such as to achieve certainty.
Truth, i.e. information that fixes the state of a variable, is called 'hard evidence', otherwise it is

called 'soft evidence'.
In other words, hard evidence bearing on a given variable allows its true state to be known, whereas soft evidence allows only the assignment of probability values less than one to the states of the variable.
Thus, for converging connections, it can be said that information may only be transmitted if there is some evidence (hard or soft) bearing on either the middle variable, or on one of its descendants.

This last condition can easily be understood by an example; putting together events A,B,C and E as described above, we will arrive to the graph to the right,

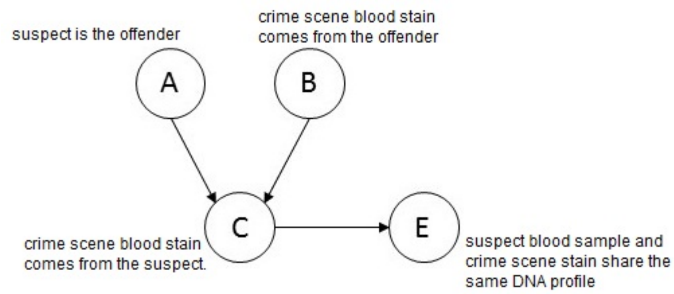


Figure 5: Blood stain example

The formal development of the likelihood ratio associated with this Bayesian network is presented later.
Note that, knowing the state of E (which is the descendant of C), opens the channel between A and B because this information, in turn, changes the degree of belief in C, even though this degree is still lower than certainty, (allowing for the possibility of a correspondence by chance).

5 Building the Network

The construction of a BN involves three major steps [2]. First, we must decide on the set of the relevant variables and their possible values. Next, we must build the network structure by connecting the variables into a DAG. Finally, we must define the CPT (Conditional Probability Table) for each network variable (node); which are called NPT (Node Probability Tables in [4]). The last step is the quantitative part of this process and can be the most involved in certain situations.

To build up a Bayesian network step by step given a set of variables, we use the above rules in the following way: If X and Y are conditionally independent give Z, then X and Y are *d*-separated given Z in the network. Conversely, it is also possible to decide for any pair of variables X and Y in a given Bayesian network whether they are independent given another variable Z. Thus, a decision rule for *d*-separation can be formulated in general for subsets of variables in a a directed acyclic graph, DAG.
Let **X**, **Y** and **Z** be disjoint subsets of variables in a DAG; then **X** and **Y** are d-separated given **Z** *if and only if* every path between a variable in **X** and a variable in **Y** contains either 1) a serial connection $\rightarrow \mathbf{Z} \rightarrow$, or a diverging connection $\rightarrow \mathbf{Z} \leftarrow$, such that the middle node belongs to **Z**. Or, 2) a converging connection $\rightarrow W \leftarrow$ such that the middle node does not belong to **Z** and *no* descendant of W belongs to **Z**.

Examples:
In figure 5,

nodes A and B are not d-separated given $\mathbf{Z}=\{E\}$, because E is a descendant of C.
In the figure 6, nodes B and D are d-separated given $\mathbf{Z}=\{A,E\}$ because there are two paths going between B and D, namely, $B \leftarrow A \rightarrow E \rightarrow D$ and $B \rightarrow C \leftarrow D$. Where in the first path, there is a diverging connection at A (and a serial connection at E). In the second path, there is a converging connection at C and neither is it the case that C belongs to the set $\{A,E\}$, nor any descendant of C belongs to $\{A,E\}$ (in fact C has no descendants). On the other hand, B and D are not d-separated given $\mathbf{Z}=\{A,C\}$ because in this case, the middle node C of the converging connection belongs to the set $\mathbf{Z}=\{A,C\}$.

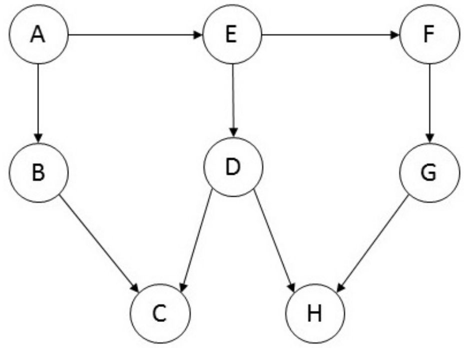


Figure 6: A directed acyclic graph example

It can be proved that for Bayesian networks that if sets of nodes **X** and **Y** are d-separated given **Z**, then they are conditionally independent given **Z**. Therefore, it is not necessarily the case that the graphical structure represents all the independence relationships that hold among the variables, but it ensures that, at least, no false relationship of independence can be read from the it, and that all the dependencies are represented.

6 Bayesian Networks and The Law

The hypothesis H that the defendant is or in not guilty is often referred to by lawyers as **the ultimate hypothesis**. In general, a legal case may consist of many additional hypotheses. [4]
Probabilistic reasoning of legal evidence often boils down to the simple BN model shown in figure 7 below. We start with an ultimate hypothesis, and observe some evidence E (such as an expert witness testimony that the defendant's blood does or does not match the blood stain found at the crime scene).

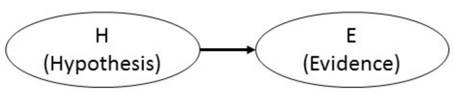


Figure 7: Causal view of the evidence

The direction of the causal structure makes sense here because the defendant's guilt increases the probability of finding incriminating evidence. Conversely, such evidence cannot 'cause' guilt. And vice versa for innocence.
Lawyers and jurors do not formally use Bayes' theorem, but they would normally use the following Bayesian inference about the evidence:

- Start with some (unconditional) prior assumption about guilt. For example, the assumption that the defendant is innocent until proven guilty, which means the defendant is no more likely to be guilty than any other member of the population on importance in the studied case.
- Update this prior belief about H once you observe evidence E. This updating takes account of the likelihood of the evidence, which is the chance of seeing the evidence E if H is true.

The first item in the formal Bayes context is the prior $p(H)$, which we assume known. Also we assume we know the conditional probability of E given H, $P(E|H)$. Then the second item is the posterior that we want to know, $P(H|E)$, which reads as, the probability of observing this evidence knowing that the defendant is guilty.
Now consider the example of a blood trace found at the scene of a crime that must have come from the murderer. The blood is tested against the DNA of the defendant and and result is presented (whether true or false). This is certainly an important piece of evidence that will adjust our prior belief about the the defendant's guilt.
Using the approach described in this section and figure 7, we could model this case using BN with the prior beliefs (i.e. marginals, that is, before the evidence is known). We assume that the false DNA match probability is *one in a million*, and the following overly **simplifying assumptions**: i) We will definitely establish a match if the defendant is guilty, i.e. the DNA testing is perfect, in particular there is no possibility of wrongly finding a match; note that this is very different from the probability of random match. ii) The blood tested really that found at the scene of the crime. iii) The blood did not become contaminated at any time, and iv) The person presenting the DNA evidence in the court does so in a completely truthful and accurate way. Finally, we assume that the prior probability of guilt is 0.01 (which would be the case if the defendant is one of 100 people that were at the crime scene around the time of the murder, in accordance with the 'innocent till proven guilty' code)
Then if we discovered a match, the posterior will be, using Bayes' theorem and these prior beliefs:
$P(\text{guilt}) = P(H) = 0.01$,
$P(\text{match given guilt}) = P(E|H) = 1$, and
$P(\text{match given innocence}) = P(E|\text{not } H) = \frac{1}{10^6}$
We will have the wanted posterior, P(guilt given match):

$$\begin{aligned} P(H|E) &= \frac{P(E|H).P(H)}{P(E|H).P(H) + P(E|\text{not } H).P(\text{not } H)} \\ &= \frac{(1)(0.01)}{(1)(0.01) + (\frac{1}{10^6})(0.99)} \\ &= 0.9999 \end{aligned}$$

Which means in this case, that if we discovered a match in the DNA between the crime scene and the defendant's, our belief of his guilt should become 0.9999 (i.e. 99.99% sure of his guilt). Therefore, P(innocence given match)= $P(\text{not } H|E)$ is *one in a 10'000*. However, note that this is still significantly greater than P(match given innocence)= $P(E|\text{not } H) = \frac{1}{10^6}$. This is to stress the importance to differentiate between $P(E|\text{not } H)$ and $P(\text{not } H|E)$ which is a common error among prosecutors, and is called *the prosecutor's fallacy* when they claim something like: 'The chances that an innocent person has the matching DNA is one in a million, therefore the chances that the defendant is innocent is one in a million'.
There is reluctance in the court of law to use subjective priors to the ultimate hypothesis as the above because it is a delicate and controversial issue. It is possible to avoid this matter if we instead focused on the probabilistic 'value' of the evidence; the impact of any single evidence E on the hypothesis H can be determined by considering only the *likelihood ratio* of E, which is the odds of $P(E|H)$ (probability of seeing this E if H is true), that is $\frac{P(E|H)}{P(E|\text{not } H)}$. However, we still need to

use Bayes' theorem in its odds form to draw definitive conclusions, which is:
Posterior odds = likelihood ration $\times$ prior odds
From this formula, it is apparent that we must report the prior probabilities as well, as they weigh the likelihood. However, we could state something like: "Whatever you believed before about the possible guilt of the defendant, the evidence is one million times more likely if the defendant is guilty than if he is innocent. So, if you believed at the beginning that there was 50:50 chance that the defendant was innocent, then it is only rational for you to conclude with the evidence that there is only one to a million chance the defendant is really innocent. On this basis you should return a 'guilty' verdict. But if you believed at the beginning that there are a million other people in the area who are just as likely to be guilty of this crime, then it is only rational for your to conclude with the evidence that there is a 50:50 chance the defendant really is innocent. On that basis you should return a 'not guilty' verdict".
Note that such an approach does not attempt to force particular prior probabilities on the jury, the judiciary would always reject such an attempt. But it simply ensures that the correct conclusions are drawn from what may be very different subjective priors.

7 Multiple Pieces of Evidence

7.1 Adams' Case

In the case of R.V. Adams [4], convicted of rape, a DNA match was the only prosecution evidence against the defendant, but it was also the only evidence that was presented in probabilistic terms in the original trial, even though there was actually great uncertainty about the its value (the match probability was disputed, but it was accepted to be between 1 in 2 million and 1 in 200 million). This has a powerful impact on the jury. The other two pieces of evidence favored the defendant: failure of the victim to identify Adams and an unchallenged alibi. In the Appeal, the defense argued that it was wrong to consider the impact of the DNA probabilistic evidence alone without combining it with the other evidence. Figure 8 below shows a BN that incorporates both the prosecution evidence and the defense evidence, with the priors and the node probability tables (NPTs) shows based on discussion during the trial. Using the assumption of a 1 in 2 million random match probability for the DNA, the impact of the DNA match evidence alone is shown if the figure 8 below.

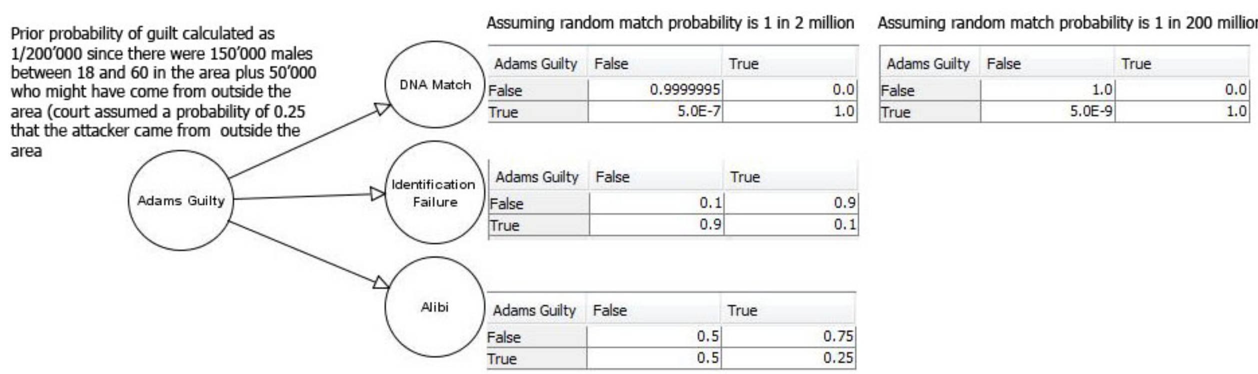


Figure 8: Hypothesis and evidence in the Adams' case

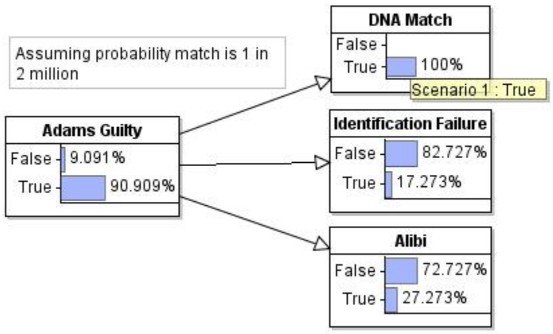
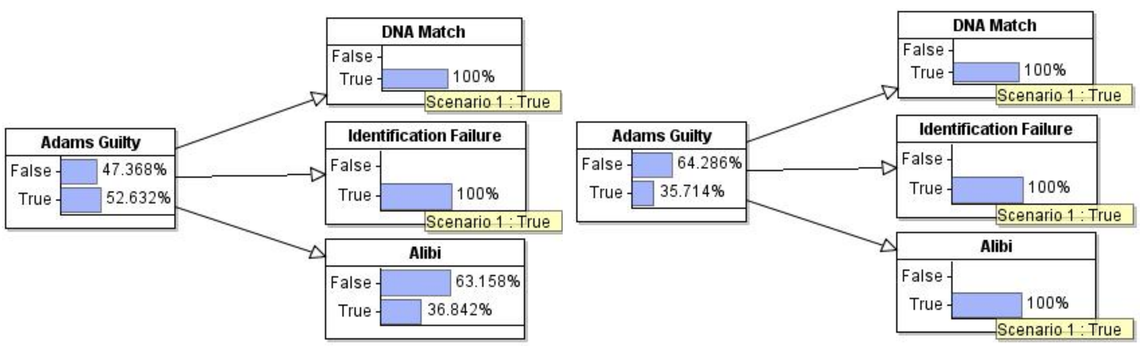


Figure 9: Result of entering DNA match evidence

In figure 9, note that because of the (1/200'000) prior probability of guilt, the DNA evidence is nothing like as conclusive as was implied by the prosecutor. The probability of innocence is still over 9%. When we enter the identification failure evidence, figure 10a, the probability of innocence goes up to 47% and it goes above 60% when we further add the alibi evidence, figure 10b.



(a) Identification Failure evidence added

(b) Alibi evidence added

Figure 10: Defense evidence added

Finally, we can return the model with the other extreme assumption about the match probability of 1 in 200 million. Figure 11 shows the result when all the evidence is entered. In this case, the probability of guilt is much higher (98%). Of course, it would be up to the jury to decide, not just if the assumption in the model are reasonable, but also whether the resulting probability of guilt leaves room for doubt. What the jury should certainly NOT have to do is understand the complex Bayesian calculations from first principles (and expect the jury, with the aid of calculators, to be able to arrive at the correct results); the judge ruled this approach overly complex.

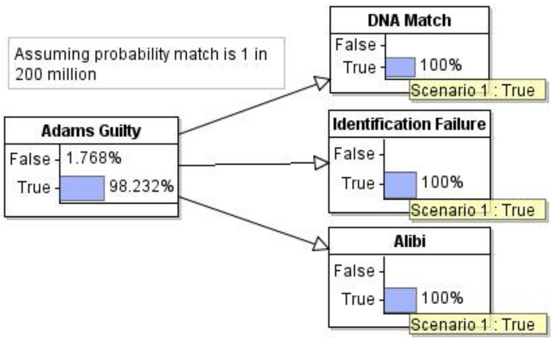


Figure 11: Effect of all evidence in case of 1 in 200 million match probability

7.2 Adding The Accuracy/Measure Factor

Back to the simplifying assumptions we made in the blood stain example in section 6 earlier,

if any of them is uncertain, then the presentation of evidence of blood match DNA being true or false cannot be simply accepted unconditionally as before. It must be conditioned on the overall *accuracy/reliability* of the evidence. In general, the validity of any piece of evidence has uncertainty associated with it, just as there is uncertainty associated with the ultimate hypothesis. Therefore, the simple model in figure 7 should now be as shown in figure 12 here,

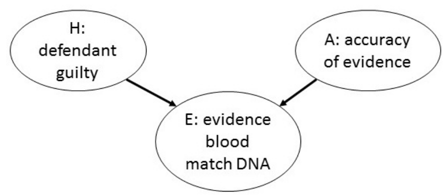


Figure 12: One evidence conditioned on its accuracy

We have gathered all possible sources of inaccuracy in one node, but more complete solutions will be presented later. Also, now the node E is more complicated because it is conditional on two parents being true or false, rather than one parent. We can have the earlier simple model of figure 7 by fixing the node A in figure 12 to *one* for being *true*.
If we assigned P(evidence accurate)=0.9, so P(evidence inaccurate)=0.1=10%, and P(guilt)=0.01=1%, so P(innocence)=0.99; and the following NPT (node probability table) for E, which now conditional on two parents:

H: Defendant guilty	False		True	
A: Accuracy of evidence	False	True	False	True
False	0.5	0.999999	0.5	0.0
True	0.5	1.0E-6	0.5	1.0

Figure 13: NPT of the evidence E

Then computations will show that $P(\text{guilt}|\text{match}) \approx 16\%$, still low (see figure 14); but that makes perfect sense because the model looks for the *most likely* explanation of the blood match evidence; since the prior of guile was 1 in 100, which is low compared to the prior of inaccuracy 10 in 100, so when the blood match evidence is presented, the model points to inaccurate evidence as being the more likely explanation of the result; indeed $P(\text{inaccurate}|\text{match})$ is $\approx 85\%$.

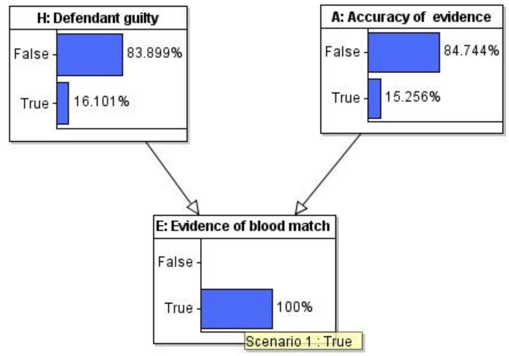


Figure 14: Adding accuracy example

However, if the evidence is accurate, and we got a match, then $P(\text{guilt}|\text{accurate,match}) = 99.99\%$. But if we determined that the evidence is inaccurate, then $P(\text{guilt}|\text{inaccurate,match}) = 1\%$ which is the same as the prior of guilt, and that means that the evidence is worthless if it is inaccurate.
We should mention that there are several ways to express evidence accuracy rather than the Boolean false or true; for instance, we could refine it in a scale of 5 increments from 'completely inaccurate' to 'completely accurate'; or even a continuous scale. We can also decompose the accuracy into three components: *competence, objectivity* and *veracity*. [4]
This example stresses the fact that both priors of guilt and accuracy are important to computing the probability of guilt given the evidence (DNA matching). Moreover, this example clarifies what inferences should be drawn from a positive testing result.

7.3 Adding Motive and Opportunity

So far we have considered the ultimate hypothesis (guilt) to be a root node (has no parents), but there are two very common types of evidence which support hypotheses that are *causes* rather than *consequences* of guilt. Namely, the "motive" and "opportunity" which change the fundamental structure of the studied causal model.

Opportunity for lawyers mean a necessary requirement for the defendant's guilt. A most common example of it is 'being present at the crime scene (at the time of the crime)'. A figure describing the general Bayesian network now is figure 15 here,

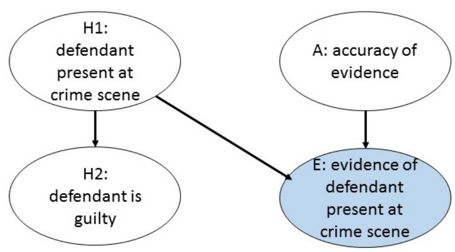


Figure 15: Incorporating opportunity example and accuracy

Note that the ultimate hypothesis: defendant is guilty, is not known to the jury, neither is the opportunity hypothesis. Just like any other hypothesis in a trial, its truth-value must be determined on the basis of evidence. Also, there might be multiple types of evidence for the opportunity hypothesis, each of a different level of accuracy. e.g. eye witness, surveillance camera, payment receipt, physical traces, etc.
Figure 16 below shows an example of multiple opportunity hypothesis evidence with their individual accuracy.
Also, assigning priors to the opportunity hypothesis cancels the need of the tricky and controversial problem of assigning priors to the guilt hypothesis itself.

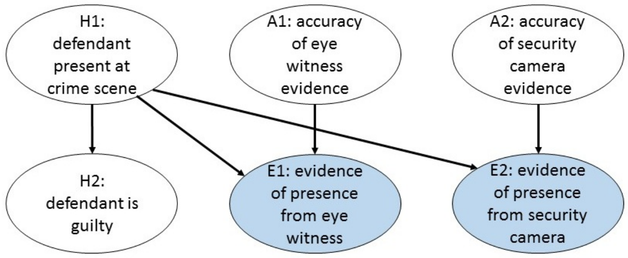


Figure 16: Multiple types of evidence for opportunity hypothesis

Motive is a widely accepted notion by police and legal community as a (normal) requirement for crime. The motive covers the notions of "intention" and "premeditation". Although, unlike opportunity, a motive is not necessary to commit a crime, the existence of it increases the chances of the crime happening. Since opportunity and motive are both evidence, then it is logical to add the accuracy of them to the model. Figure 17 below shows an example of multiple opportunity hypothesis evidence with their individual accuracy.

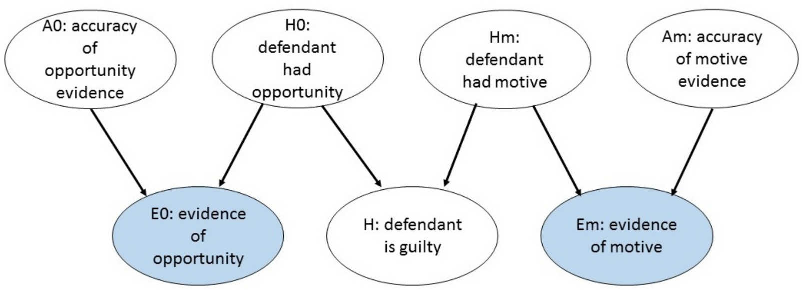


Figure 17: Incorporating opportunity and motive, and their accuracy

Note: Now the probability of guilt is conditioned on both the opportunity and the motive. What we must avoid though, is conditioning H directly on *multiple* motives, i.e. having multiple motive parents of H. Instead, if there are multiple motives, we consider the accuracy of each motive separately, and jointly as the overall strength of the motive. The apporopirate model for that is shown below:

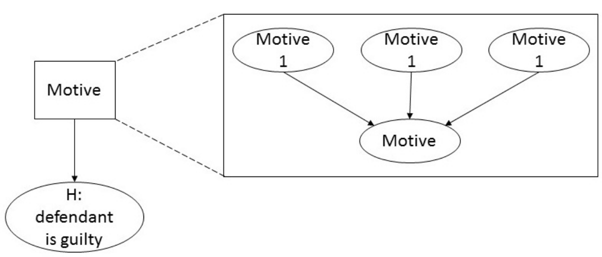


Figure 18: Appropriate model for multiple motives

7.4 Multiple Dependent Evidences

We have assumed so far that the multiple pieces of evidence are independent (like the Adams' case above).

But this is not true always, because in case we had several surveillance cameras for instance, then there is clear dependence between them, as if the first camera spotted a person that looked like the defendant, then there is good possibility that the second camera did too, and so on. Figure 19 shows an example for dependent evidences,

Note: There are other types of dependencies between pieces of evidence that include: i) Dependent evidence through confirmation bias. e.g. two experts determining whether there is forensic match (the type of forensics could be different, such as fingerprints and DNA). It has been shown that the second expert's conclusion will be biased if he/she knows the conclusion of the first expert. ii) Dependent evidence through common biases, assumptions, and sources of inaccuracies. [4]

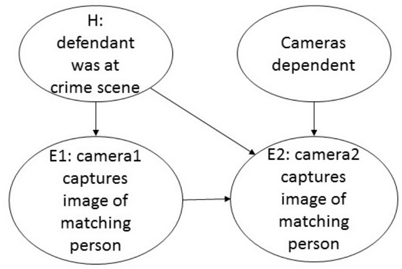


Figure 19: Dependency between multiple pieces of evidence example

The Alibi Evidence is a special case of additional direct dependency within the model, as it, generally, directly contradicts the prosecutor's hypothesis of guilt. However, not all alibi types of evidence are accurate alike, e.g. an eye witness who is known to the defendant; in this case, the hypothesis of guilt itself may have influence on the accuracy of the alibi. Also, because we tend to believe that people lie, especially, if they are in support or an opponent to the defendant.

Constraint Nodes are added as the only solution to problems when: i) There are two or more mutually exclusive causes of an event E, each with separate causal pathways. ii) The mutually exclusivity acts as a constraint on the state-space of the model and can be modeled as a constraint. Then, when running the model, soft evidence must be entered for the constraint node to ensure that impossible states cannot be realized in the model.
An example for that would be: Suppose we have the evidence E that blood found on the defendant's shirt matches the victim's blood. There is a small chance, say 1% that the defendant's blood is the same type as the victim's. Assume there are only two possible causes for the blood stain on defendant's shirt, *cause 1*: This stain comes from the victim. *cause 2*: This stain comes from the defendant himself.

Note

It is possible that the stain comes from the defendant because he either got cut either in a recent accident or in a fight (with the victim while committing the crime). Therefore, the usual approach of BNs will not work in this case, and so we add a constraint node conditioned on both causes and impossibility, while the causes have soft evidence. Then the general sketch in this case will be the figure 20.

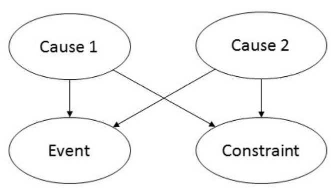


Figure 20: Using constraint node example

8 Hypothetical Illustrative Example

To put all the above together, we present the following example, L. Vole is charged with murdering his new friend, a rich elderly lady, Miss French by a strike to the back of her head. Motives included Vole being poor, and that French has changed her will to leave him all her money. Opportunity: Him visiting her regularly including the night of her death (backed by eye witness, French's maid). Incriminating evidence: Vole inquiring about luxurious cruises soon after French has changed her will; there was no force entry into the house; Vole had blood stains on his cuffs that matched French's blood type. Exonerating evidence: The maid admitted she did not like Vole; the maid was previously the sole benefactor in French's will; Vole's blood is of the same type as French's; Vole admitted he cut his wrist in the kitchen, and he had the scare to assert his claim. More exonerating pieces of evidence: Vole's wife, Romaine was to testify that her husband came back home before the time of crime, then in the trial she changed her story and testified that her husband came home late (which will make time for him to commit the crime), also she added that he told her that he have killed French; However, an unknown woman provided the defense lawyer with a bundle of letters allegedly written by Romaine to her overseas lover, in one letter of them Romaine planned on fabricating her testimony to incriminate Vole. Accordingly, Vole was acquitted. However, Romaine later revealed that she forged the letters herself, and that there was no lover overseas, and that she did this to discredit herself as she was the prosecutor's key witness.

We now build the model from scratch for this case:
Step 1: Identify the prosecution hypothesis, including opportunity and motive,

- The ultimate hypothesis: H0: Vole is guilty
- Opportunity: Vole present at crime scene at the time of the crime.
- Motive: Vole was in French's will and Vole is poor

Step 2: What evidence is available for each of the above, and what is the accuracy of each evidence:
Evidence for H0: There is no direct evidence for H0 because no record or witness to the crime happening. But we have evidence for two hypotheses that depend on H0:
H1: Vole admits guilt to his wife Romaine
H2: Blood on Vole's cuffs is from French. (Neither of these hypotheses is certainly true if H0 is true)
The prosecution evidence to support H1 is Romaine testimony in the trial. Note also that Romaine's statement puts Vole in crime scene at crime time (opportunity)
The issue of evidence accuracy is especially important for Romaine's evidence, therefore H1 influences her accuracy, because she is a witness that is related to the defendant.
Another evidence to support the opportunity is the maid's statement.
Step 3: Considering defense evidence:
The evidence to challenge H1 is the presentation of the love letters, and the introduction of a new hypothesis, H5: Romaine has lover.
To challenge the opportunity: i) Debate the accuracy of the maid's statement, and ii) Vole's own alibi evidence.
To negate H2: i) Vole and French have the same blood type, and ii) providing an additional hypothesis H3 that depends on H2, that is H3: Blood on Vole's cuffs is from a recent accident.

We add some simplifying assumptions, that the blood match evidence is perfectly accurate, and that the motives as stated are accepted without evidence.
Hence, we arrive to the following BN after these analysis,

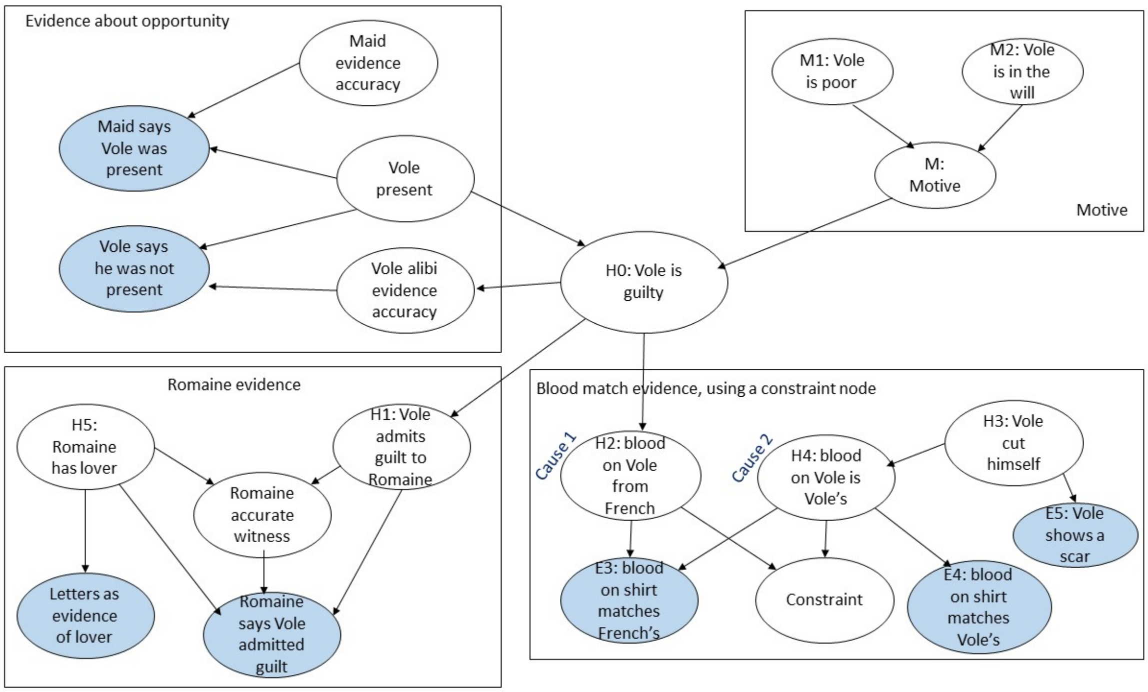


Figure 21: Full Vole example model

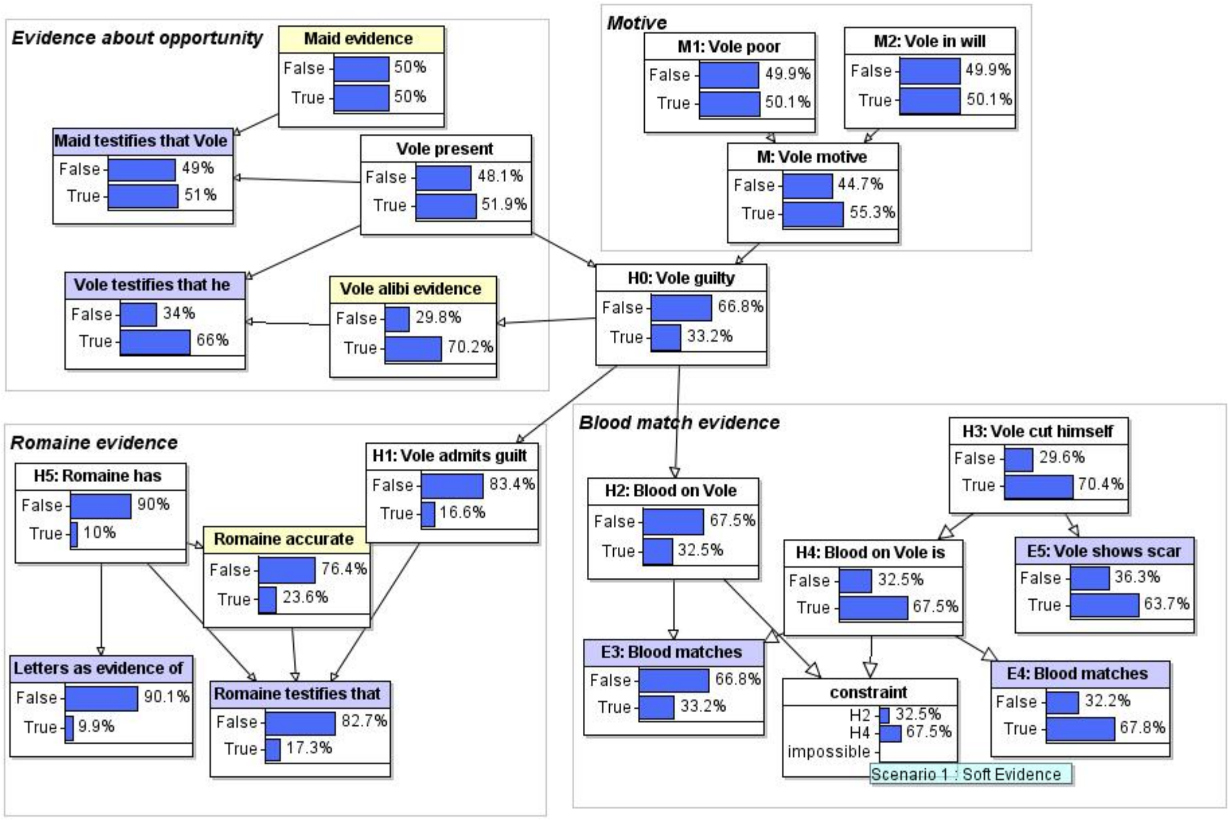


Figure 22: Full Vole example model after computations

It is worth mentioning that performing a sensitivity analysis shows that the immediate impact of different pieces of evidence *individually* is very much determined by the order in which the evidence is presented.

9 How To Assign Priors

We mentioned earlier, that it is very important to report to the judiciary, the priors we have used, along with the results we found about the posterior. Also, we must report not only the probability of the ultimate hypothesis, but the alternative as well. [4]
In all examples above, the reader might got the gist, but not a clear statement regarding how to assign the priors.

The Bayesian expert, is *not* to assign priors by him/herself [5]. We must rely on a scientist or an expert prior (opinion), or otherwise provided by the prosecutor and the defense attorneys. However, although that is not entirely avoidable, we expect to be given priors by the judiciary that are as non-subjective as possible, which falls into one or more of the following categories:

Scientific facts such as the probability of having two matching DNA profiles in a given population (random match). Or to express statements like *"innocent till proven guilty"* by assuming the defendant is no more guilty than any other individual of the considered population of people might have committed the crime, e.g. adult males who live in 100 miles of the crime scene, provided that there is no doubt about an adult male being the criminal.

Empirical or frequentest beliefs that may or may not rely on scientific facts, such as accuracy of an evidence. e.g. accuracy of a surveillance camera, or a blood testing machine. Note that accuracy of an eye witness depends on further evidence like the eye witness relationship to the defendant and his/her own interest of exonerating or incriminating the defendant.

Premises-like logical beliefs and/or intuitive ones that makes sense upon contemplating and agreed upon by both parties and/or the jury.

10 Conclusion

We have seen that Bayes theorem is the natural way to reason about legal evidences, and it also provides a shield against many common fallacies [4]. And that the likelihood ratio is a powerful method for evaluating the (probative) value of an evidence.
Also that the Bayesian networks are the ideal tool as a graphical representation of everything we know and need to consider about a given case for a decision that is as close to the truth as possible.

11 Appendix

11.1 Computations Involved in a Bayesian Network

To clarify the underlying computations for each node, consider the following toy example of an opportunity evidence with its accuracy, we will think of it not as part of a bigger network, but as a very simple complete BN example merely for sak

Assume we have the following NPTs to start with, (aligned with their positions in figure 23),

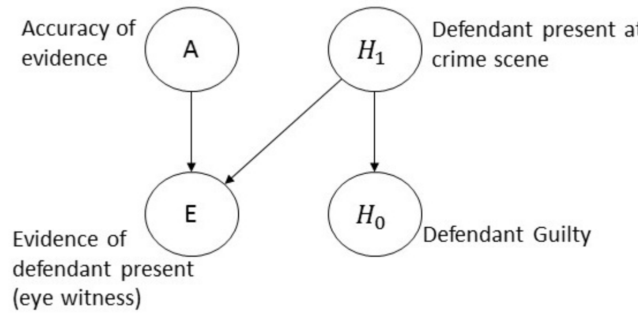


Figure 23: Simplistic BN

False	0.6
True	0.4

(a) Accuracy

False	0.9
True	0.1

(b) H1, hypothesis of presence

Accurate	False		True	
defendant present	False	True	False	True
False	0.7	0.4	0.4	0.2
True	0.3	0.6	0.6	0.8

(c) Evidence of Opportunity

defendant present	False	True
False	0.9	0.2
True	0.1	0.8

(d) H0, hypothesis of guilt

Figure 24: NPTs for Figure 23

First: Suppose now we want to confirm there is an evidence for opportunity, say an eye witness, this means, we want to find the *marginal* probability that E is true, then using *the law of total probability*, and noting that A and H_1 are independent when E is unknown, we get:

$$
\begin{aligned}
P(E=\text{true}) = \sum_{A,H_1} & P(E=\text{true}|A,H_1).P(A).P(H_1) \quad (1)\\
&= P(E=\text{true}|A=\text{true},H_1=\text{true}).P(A=\text{true}).P(H_1=\text{true})\\
&+ P(E=\text{true}|A=\text{true},H_1=\text{false}).P(A=\text{true}).P(H_1=\text{false})\\
&+ P(E=\text{true}|A=\text{false},H_1=\text{true}).P(A=\text{false}).P(H_1=\text{true})\\
&+ P(E=\text{true}|A=\text{false},H_1=\text{false}).P(A=\text{false}).P(H_1=\text{false})\\
&= (0.8)\times(0.4)\times(0.1)\\
&+ (0.6)\times(0.4)\times(0.9)\\
&+ (0.6)\times(0.6)\times(0.1)\\
&+ (0.3)\times(0.6)\times(0.9)\\
&= 0.446
\end{aligned}
$$

Second: Suppose we are interested in $P(E = \text{true}|H0 = \text{true})$ which would help us to get an idea about the honesty of the eye witness. But note that we have a common parent for the two nodes H_0 and E, which means if we *know* the state of H_0 being true, then we must *update* our parent node probability H_1 accordingly; *before* we compute our probability of interest; and we do that using Bayes' theorem. Thus, this is a two-stage computation process, first:

$$
\begin{aligned}
P(H_1=\text{true}|H_0=\text{true}) &= \frac{P(H_1=\text{true}).P(H_0=\text{true}|H_1=\text{true})}{\sum_{H_1} P(H_1).P(H_0=\text{true}|H_1)}\\
&= \frac{P(H_1=\text{true}).P(H_0=\text{true}|H_1=\text{true})}{P(H_1=\text{true}).P(H_0=\text{true}|H_1=\text{true})+P(H_1=\text{f}).P(H_0=\text{true}|H_1=\text{f})}\\
&= \frac{(0.1)\times(0.8)}{(0.1)\times(0.8)+(0.9)\times(0.1)}\\
&= 0.47059
\end{aligned}
$$

And so,

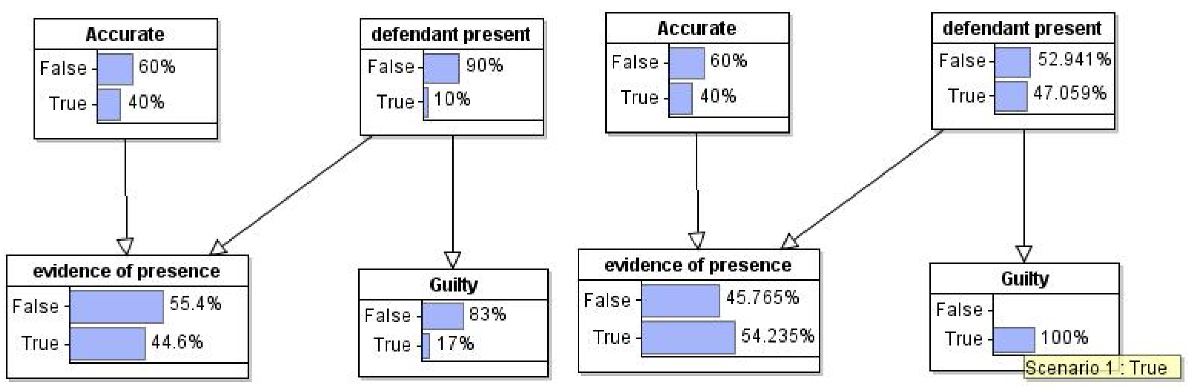
$$
\begin{aligned}
P(H_1=\text{false}|H_0=\text{true}) &= 1-0.47059\\
&= 0.52941
\end{aligned}
$$

Now when we marginalize the joint distribution to get $P(E|H_0 = \text{t})$ we use this $P(H_1|H_0)$ we just found now that we have *fixed* the state of H_0:

$$
\begin{aligned}
P(E=\mathrm{t}|H_0=\mathrm{t}) &= \sum_{A,H_1} P(E,A,H_1,H_0) \\
&= \sum_{A,H_1} P(E|A,H_1).P(A).\underbrace{P(H_1).P(H_0|H_1)} \\
&= \sum_{A,H_1} P(E|A,H_1).P(A).\overbrace{P(H_1|H_0).P(H_0)}^{\text{cond. prob. def}} \\
&= \sum_{A,H_1} P(E|A,H_1).P(A).P(H_1|H_0=\mathrm{t}).1, \quad \{P(H_0)=1\} \text{ (because fixed)} \\
&= P(E=\mathrm{t}|A=\mathrm{t},H_1=\mathrm{t}).P(A=\mathrm{t}).P(H_1=\mathrm{t}|H_0=\mathrm{t}) \\
&+ P(E=\mathrm{t}|A=\mathrm{t},H_1=\mathrm{f}).P(A=\mathrm{t}).P(H_1=\mathrm{f}|H_0=\mathrm{t}) \\
&+ P(E=\mathrm{t}|A=\mathrm{f},H_1=\mathrm{t}).P(A=\mathrm{f}).P(H_1=\mathrm{t}|H_0=\mathrm{t}) \\
&+ P(E=\mathrm{t}|A=\mathrm{f},H_1=\mathrm{f}).P(A=\mathrm{f}).P(H_1=\mathrm{f}|H_0=\mathrm{t}) \\
&= (0.8)\times(0.4)\times(0.47059) \\
&+ (0.6)\times(0.4)\times(0.52941) \\
&+ (0.6)\times(0.6)\times(0.47059) \\
&+ (0.3)\times(0.6)\times(0.52941) \\
&= 0.5423534
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow P(E=\mathrm{f}|H_0=\mathrm{t}) &= 1-0.5423534 \\
&= 0.4576466 \approx 0.45765
\end{aligned}
$$

If we multiplied these probabilities by 100, we get the percentage expression shown in the BN NPTs in Figures 25,



(a) Initial NPTs, when all true states are unknown (b) Assuming P(Guilt)=1

Figure 25: Change in NPTs when we know the true state of a node

Note how the joint probability is hugely reduced when we incorporate the BN structure and the properties of each connection type.
Initially we assume we do not know the true state of any node, and we broke down the joint accordingly. Now when we want to find the probability of a particular node, we assume its parents are given and so we reduce the joint further, recall equation (1). It is true the full joint is $P(E, A, H_1, H_0) = P(E|A, H_1).P(H_0|H_1).P(A, H_1) = P(E|A, H_1).P(H_0|H_1).P(A).P(H_1)$ because initially, E is unknown, then A and H_1 are independent due to the *converging* connection between them. Hence we could write $P(E|A, H_1)$ as $P(A).P(H_1)$.
Likewise, H_1 is unknown, then H_0 and E are dependent due to the *diverging* connection between them.
But note that when we stated $P(E|H_1, A)$, we did not include dependency on H_0 because we are saying in this probability that H_1 is *given* which makes E and H_0 independent, hence $P(E|A, H_1, H_0) = P(E|A, H_1)$ by a property of independence between random variables.

References

[1] Franco Taroni, Alex Biedermann, Silvia Bozza, Paolo Garbolino, Colin Aitken. *Bayesian Networks for Probabilistic Inference and Decision Analysis in Forensic Science, 2nd edition.* Wiley, 2014.

[2] James Allen. Using entropy for evaluating and comparing probability. http://www.cs.rochester.edu/u/james/CSC248/Lec6.pdf, 2003.

[3] G. Casella , R. L. Berger. *Statistical Inference, 2nd edition.* Duxbury , Thomson Learning, 2002.

[4] Niko Brummer. Application-independent evaluation of speaker detection. *In: Proceedings of Odyssey-04: The ISCA Speaker and Language Recognition Workshop , Toldeo*, 2004.

[5] Adnan Darwiche. *Modeling and Reasoning with Bayesian Networks.* Cambridge University Press, 2009.

[6] Marco Scutari, Jean-Baptiste Denis. *Bayesian Networks: With Examples in R.* CRC Press, 2015.

[7] Niko Brummer , Johan du Preez. Application-independent evaluation of speaker detection. *Computer Speech and Language*, 20(2-3):230275, 2006.

[8] Christian Muller (Ed.). *Speaker Classification I: Fundamentals, Features, and Methods.* Springer, 2007.

[9] Amy Neustein , Hemant A. Patil (Editors). *Forensic Speaker Recognition. Law Enforcement and Counter-Terrorism.* Springer, 2012.

[10] Kasper Daniel Hansen, Jeff Gentry, Li Long, Robert Gentleman, Seth Falcon, Florian Hahne, and Deepayan Sarkar. *Rgraphviz: Provides plotting capabilities for R graph objects.* R package version 2.10.0.

[11] Criminal Justice Information, Services Division, and Operational Technology Division. Investigatory Voice Biometrics Committee Report Investigatory Voice Biometrics Committee Report Development of a Draft Type-11 Voice Record. pages 1–61, 2012. http://www.fbi.gov/about-us/cjis/fingerprints$_biometrics/biometric - center - of - excellence/files/type - 11_ivbcreport_v2 - 1b - 09262012.pdf$.

[12] Masud Mansuripur. *Introduction to Information Theory.* Prentice-Hall, Inc. Upper Saddle River, NJ, USA, 1987.

[13] Oliver Pourret, Patrick Naim, Bruce Marcot. *Bayesian Networks, A Practical Guide to Applications.* John Wiley and Sons, Ltd, 2008.

[14] Vasileios Maroulas. Computational bayesian statistics. *Lecture Notes for Math 525-526*, 2014.

[15] Virginia Maureen C. Bottrell , Quantico. Forensic glass comparison: Background information used in data interpretation. *Forensic science communications*, 11(2), 2009. http://www.fbi.gov/about-us/lab/forensic-science-communications/fsc/april2009/review/2009$_04_review01.htm$.

[16] Geoffrey Stewart Morrison. Measuring the validity and reliability of forensic likelihood-ratio systems. *Science Justice*, 51(3):9198, 2011.

[17] Norman Fenton, Martin Neil. *Risk Assessment and Decision Analysis with Bayesian Networks.* Cambridge University Press, 2013.

[18] Finn V. Jensen, Thomas D. Nielsen. *Bayesian Networks and Decision Graphs.* Springer, 2007.

[19] J. K. Ghosh , M. Delampady , T. Samanta. *An Introduction to Bayesian Analysis Theory and Methods.* Springer, 2006.

[20] Rudolph Verderber , Kathleen Verderber , Deanna Sellnow. *The Challenge of Effective Speaking.* Thomson Wadsworth, 2008.

[21] Jim Q. Smith. *Bayesian Decision Analysis: Principles and Practice.* Cambridge University Press, 2010.

[22] William C. Thompson, Joelle Vuille, Alex Biedermann, Franco Taroni. The role of prior probability in forensic assessments. *Frontiers in Genetics*, 4(220):175–191, 2013.

[23] Thomas M. Cover , Joy A. Thomas. *Elements of Information Theory, 2nd edition.* Wiley-Interscience, 2006.