Bayesian Inference for Probabilistic Knowledge Models

Ramón Béjar

Departament d'Informàtica
Universitat de Lleida

In a No Perfect World, No Perfect Information

So far, we have been thinking about intelligent systems in a perfect world, where the information needed by the system to reason about and make decisions:

- Is complete: We have all the needed knowledge
- Is exact: for any assertions/properties such that we get information about them, we know their exact truth value

But in many real world applications, not all the information we will need to consider to run our intelligent systems will satisfy these conditions

Can we still work with incomplete and inexact information ?

Inference with Random Variables

We will consider that the variables that our system wants to discover their value are random variables

We may not be able to directly observe the actual value for some of these variables, but we will use probabilistic knowledge models that will allow us to infer the most probable values for our target random variables, given some partial knowledge we have about the world: our evidence

The Family Out, Dog Out Problem

Consider a house with a garden, where the family has a dog
Suppose we want to understand, and predict, the state of the two following binary variables:

- FO: The family living in the house is out of the house (1) or inside (0)
- DO: The dog is outside of the house (1) or inside (0)

Is there any relation between the state of both variables, such that having knowledge about one of them will help us to get knowledge about the other ?

The Family Out, Dog Out Problem

When we do not know the state of both variables, what can we discover about them ?
If we have the joint probability distribution for

$$P(FO, DO)$$

we can perform probabilistic inference for the value of any of the variables, given some partial knowledge:

1. No knowledge at all: There is maximal uncertainty, but the joint probability distribution can provide probabilities for each possible situation or for each individual variable
2. We know the knowledge of one of the variables: We can infer the conditional probability for the value of the other variable

The Family Out, Dog Out Problem

Suppose we have the joint probability distribution for both variables

v1	v2	P(FO = v1, DO = v2)
0	0	0.25
0	1	0.25
1	0	0.125
1	1	0.375

We will discuss later how we are supposed to learn such a probability distribution ...

The Family Out, Dog Out Problem

What can we infer when we there is complete uncertainty about the value of both variables ?

Probability of different events:

1. For a particular combination of variables with values (given by one particular entry of the table)
2. For an individual variable=value (marginal probability)
3. For an individual variable=value assuming certain hypotheses are true, using conditional probabilities. For example:
   - What is the probability of having the dog out, supposing the family is out ?

   $$\mathrm{P}(\mathrm{DO} = 1 | \mathrm{FO} = 1)$$

   - What is the probability of having the family out, supposing the dog is out ?

   $$\mathrm{P}(\mathrm{FO} = 1 | \mathrm{DO} = 1)$$

The Family Out, Dog Out Problem

We compute marginal (unconditional) probabilities for any desired variable with the sum rule:

For $FO = 1$:

$P(FO=1) =$
$\sum_{DO=v_i} P(FO=1, DO=v_i) =$
$0.375 + 0.125 = 0.5$

For $FO = 0$:

$P(FO=0) =$
$\sum_{DO=v_i} P(FO=0, DO=v_i) =$
$0.25 + 0.25 = 0.5$

And $P(FO=1)+P(FO=0) = 1.0$

For $DO = 1$:

$P(DO=1) =$
$\sum_{FO=v_i} P(FO=v_i, DO=1) =$
$0.25 + 0.375 = 0.625$

For $DO = 0$:

$P(DO=0) =$
$\sum_{FO=v_i} P(FO=v_i, DO=0) =$
$0.25 + 0.125 = 0.375$

And $P(DO=1)+P(DO=0) = 1.0$

The Family Out, Dog Out Problem

Conditional probabilities can be computed from joint probabilities using factorizations of the joint distribution:

Joint Probability Factorization for events A and B

$$\mathrm{P}(\mathrm{A},\mathrm{B}) = \mathrm{P}(\mathrm{A}|\mathrm{B})\mathrm{P}(\mathrm{B}) = \mathrm{P}(\mathrm{B}|\mathrm{A})\mathrm{P}(\mathrm{A})$$

So:

$$\mathrm{P}(\mathrm{A}|\mathrm{B}) = \frac{\mathrm{P}(\mathrm{A},\mathrm{B})}{\mathrm{P}(\mathrm{B})} \text{ and } \mathrm{P}(\mathrm{B}|\mathrm{A}) = \frac{\mathrm{P}(\mathrm{A},\mathrm{B})}{\mathrm{P}(\mathrm{A})}$$

So, from $\mathrm{P}(\mathrm{A},\mathrm{B})$ we can compute any desired conditional probability, because $\mathrm{P}(\mathrm{A})$ or $\mathrm{P}(\mathrm{B})$ can also be computed from the joint probability distribution $\mathrm{P}(\mathrm{A},\mathrm{B})$

The Family Out, Dog Out Problem

Computing $P(DO=1|FO=1)$:

$$P(DO=1|FO=1) = \frac{P(DO=1, FO=1)}{P(FO=1)} = \frac{0.375}{0.375 + 0.125} = 0.75$$

v1	v2	$P(FO = v1, DO = v2)$
0	0	0.25
0	1	0.25
1	0	0.125
1	1	0.375

Target cases: rows with $FO = 1$

The Family Out, Dog Out Problem

Computing $P(FO=1|DO=1)$:

$$P(FO=1|DO=1) = \frac{P(FO=1, DO=1)}{P(DO=1)} = \frac{0.375}{0.375 + 0.25} = 0.6$$

v1	v2	$P(FO=v1, DO=v2)$
0	0	0.25
0	1	0.25
1	0	0.125
1	1	0.375

Target cases: rows with DO $= 1$

The Family Out, Dog Out Problem

Exercise:
Compute $P(DO=1|FO=0)$ and $P(FO=1|DO=0)$

Observe that from the law of total probability:

$$P(DO=0|FO=1) = 1 - P(DO=1|FO=1)$$

$$P(FO=0|DO=1) = 1 - P(FO=1|DO=1)$$

$$P(DO=0|FO=0) = 1 - P(DO=1|FO=0)$$

$$P(FO=0|DO=0) = 1 - P(FO=1|DO=0)$$

That is, for a binary random variable it is enough to compute the probability for one of its values, the other is obtained from this law

The Family Out, Dog Out Problem

Do we really need the complete table of the joint distribution to perform probabilistic inference ?

Suppose we have instead the following information:

- $\mathrm{P}(\mathrm{DO}|\mathrm{FO})$: that is, we know the values of $\mathrm{P}(\mathrm{DO}|\mathrm{FO}=1)$ and $\mathrm{P}(\mathrm{DO}|\mathrm{FO}=0)$
- $\mathrm{P}(\mathrm{FO})$: that is, we know the value of $\mathrm{P}(\mathrm{FO}=1)$

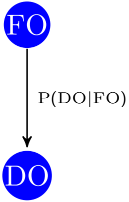
Then, we have the following factorization:

$$\mathrm{P}(\mathrm{FO},\mathrm{DO}) = \mathrm{P}(\mathrm{DO}|\mathrm{FO})\mathrm{P}(\mathrm{FO})$$

But we do not have the table with the complete joint probability distribution

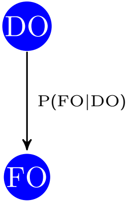
The Family Out, Dog Out Problem

However, once we have one factorization for the joint probability distribution, we can use it to compute the probability for all the possible outcomes of any conditional event P(A|B). We are going to represent Factorizations with Bayesian Networks



Factorization:

$$P(FO, DO) = P(DO|FO)P(FO)$$



Factorization:

$$P(FO, DO) = P(FO|DO)P(DO)$$

Both factorizations are equally good, in terms of the information we can infer using them !

The Family Out, Dog Out Problem

What do we need for computing any value $P(FO = v_1, DO = v_2)$ from the factorization $P(FO, DO) = P(DO|FO)P(FO)$?

We need the probability tables for $P(DO|FO)$ and $P(FO)$:

v1	v2	P(DO = v1|FO = v2)
0	0	0.5
1	0	0.5
0	1	0.25
1	1	0.75

v1	P(FO = v1)
0	0.5
1	0.5

Basic Bayes Theorem

If we only have available one of the two factorizations, for example: P(B|A)P(A), can we perform conditional probability inference for the inverse event P(A|B) ?

Bayes theorem

From $\mathrm{P(A, B) = P(A|B)P(B) = P(B|A)P(A)}$ we have that:

$$\mathrm{P(A|B)} = \frac{\mathrm{P(B|A)P(A)}}{\mathrm{P(B)}}$$

But it seems that to compute $\mathrm{P(A|B)}$ from $\mathrm{P(B|A)P(A)}$ we need $\mathrm{P(B)}$

Bayesian Inference - Two Variables

Once we have one complete factorization for the joint probability distribution, we can perform inference for any particular event, conditioned on any given evidence

Probabilities for event and evidence:

- Probability for target event $P(A|E)$: what we want to compute, conditioned to the case of being E true
- Probability for evidence $P(E)$: needed to compute $P(A|E)$ from a factorization of the joint probability distribution

But do we really need the value of $P(E)$?

Bayesian Inference - Two Variables

Given factorization $\mathrm{P(FO, DO) = P(DO|FO)P(FO)}$, compute $\mathrm{P(FO\!=\!1|DO\!=\!v_i)}$ and $\mathrm{P(FO\!=\!0|DO\!=\!v_i)}$, for a particular evidence $\mathrm{DO\!=\!v_i}$

For example, for $\mathrm{DO\!=\!1}$:

$$\mathrm{P(FO\!=\!1|DO\!=\!1)} = \frac{\mathrm{P(DO\!=\!1|FO\!=\!1)P(FO\!=\!1)}}{\mathrm{P(DO\!=\!1)}} = \frac{0.75 \cdot 0.5}{\mathrm{P(DO\!=\!1)}}$$

$$\mathrm{P(FO\!=\!0|DO\!=\!1)} = \frac{\mathrm{P(DO\!=\!1|FO\!=\!0)P(FO\!=\!0)}}{\mathrm{P(DO\!=\!1)}} = \frac{0.5 \cdot 0.5}{\mathrm{P(DO\!=\!1)}}$$

Bayesian Inference - Two Variables

But we do not know $P(DO=1)$. However, we know that:

$$P(FO=1|DO=1) + P(FO=0|DO=1) = 1$$

So, it turns out that:

$$\frac{0.75 \cdot 0.5}{P(DO=1)} + \frac{0.5 \cdot 0.5}{P(DO=1)} = \frac{0.75 \cdot 0.5 + 0.5 \cdot 0.5}{P(DO=1)} = 1$$

So: $(0.75 \cdot 0.5) + (0.5 \cdot 0.5) = P(DO=1)$

Then, we can infer what will be the most probable value for FO, given the known value for DO:

$$P(FO=1|DO=1) = \frac{0.375}{0.375+0.25} = 0.6$$

$$P(FO=0|DO=1) = \frac{0.25}{0.375+0.25} = 0.4$$

What is the Right Model ?

Consider a pair of twin brothers, Alice and Bob, such that they can be in two basic states: good mood (1) and bad mood (0)

Suppose we learn the following joint probability distribution for both random variables:

v1	v2	P(A=v1, B=v2)
0	0	0.5
0	1	0
1	0	0
1	1	0.5

It seems obvious that there is some correlation between the state of both brothers

What possible factorization should we prefer ?

What is the Right Model ?

In the causal interpretation of Bayesian Networks, the direction of links indicate a cause-effect relationship

$$A \xrightarrow{P(B|A)} B$$

Factorization (the mood of Bob is affected by the mood of Alice):

$$P(A, B) = P(B|A)P(A)$$

$$B \xrightarrow{P(A|B)} A$$

Factorization (the mood of Alice is affected by the mood of Bob):

$$P(A, B) = P(A|B)P(B)$$

But, is in this example a clear cause-effect relationship between the state of our twin brothers ?

Model : $P(A, B) = P(A|B)P(B)$

We need the probability tables for $P(A|B)$ and $P(B)$:

v1	v2	$P(A=v1 \mid B=v2)$
0	0	1
1	0	0
0	1	0
1	1	1

v1	$P(B=v1)$
0	0.5
1	0.5

Model : $P(A, B) = P(B|A)P(A)$

This model is analogous to the previous one (probability tables have the same values)

But even if the conditional probability tables were different, if they give a correct factorization for $P(A, B)$, then we can use that model too for answering the same questions like with the previous model !

So, both models are equally good for answering inference questions, because they are exact models (they give a perfect factorization for the joint distribution)

What is the Right Model ?

Who is the cause and who is the effect ?

So, even if we need to pick up some ordering for the dependence link between Alice and Bob, in order to provide a factorization for being able to perform inference, whether there is a real cause-effect relationship it does not mind too much for our goal of performing inference for events !

Suppose next that investigating further the reasons behind our entangled twins, we discover that it seems to be a relationship with the state of the moon (M). We consider two states for the moon: full moon (1) or not full moon (0).

What is the Right Model ?

So, after observing for many days the state of our twins and the state of the moon, we obtain the following joint probability distribution for $P(A, B, M)$:

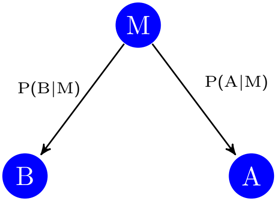
v1	v2	v3	$P(A = v1, B = v2, M = v3)$
0	0	0	0.5
0	1	0	0
1	0	0	0
1	1	0	0
0	0	1	0
0	1	1	0
1	0	1	0
1	1	1	0.5

What is the Right Model ?

Looking at that table, we suspect that a good model for explaining what happens with the mood of our twins is given by this factorization for $\mathrm{P(A, B, M)}$:

$$\mathrm{P(A, B, M) = P(A|M)P(B|M)P(M)}$$

And this would be his corresponding Bayesian Network:



What is the Right Model ?

And with the following conditional probability tables:

Table for Alice:

v1	v2	P(A=v1|M=v2)
0	0	1
1	0	0
0	1	0
1	1	1

The same table for Bob:

v1	v2	P(B=v1|M=v2)
0	0	1
1	0	0
0	1	0
1	1	1

v1	P(M=v1)
0	0.5
1	0.5

Can you check if this factorization is totally correct ?

Conditional Independence Relations

Suppose our factorization $P(A, B, M) = P(A|M)P(B|M)P(M)$ is correct

Then, taking into account that this alternative factorization:

$$P(A, B, M) = P(A|B, M)P(B, M) = P(A|B, M)P(B|M)P(M)$$

is also valid, we have the following interesting conditional independence result:

$$P(A|B, M) = P(A|M)$$

This last equation tells us that our original factorization captures the fact that given M, the probability of A is independent of the probability of B

Conditional Independence Relations

Anologously, taking into account that this other factorization:

$$P(A, B, M) = P(B|A, M)P(A, M) = P(B|A, M)P(A|M)P(M)$$

is also valid, we also have this conditional independence result:

$$P(B|A, M) = P(B|M)$$

This last equation tells us that our original factorization captures the fact that given M, the probability of B is independent of the probability of A

Bayesian Inference - Three Variables

If we now want to use our factorization $P(A, B, M) = P(A|M)P(B|M)P(M)$ to compute $P(M=1|A=1)$, we should compute:

$$P(M=1|A=1) = \frac{P(A=1, B, M=1)}{P(A=1)}$$

where $P(A=1, B, M=1)$ is the probability of having $A=1, M=1$ with any value for B. This can be obtained from our factorization in this way:

$$\begin{aligned} P(A=1, B, M=1) &= P(A=1, B=1, M=1) + P(A=1, B=0, M=1) \\ &= P(A=1|M=1)P(B=1|M=1)P(M=1)+ \\ &\quad\; P(A=1|M=1)P(B=0|M=1)P(M=1) \\ &= (1 \cdot 1 \cdot 0.5) + (1 \cdot 0 \cdot 0.5) = 0.5 \end{aligned}$$

Bayesian Inference - Three Variables

We also have to compute:

$$P(M=0|A=1) = \frac{P(A=1, B, M=0)}{P(A=1)}$$

That is:

$$\begin{aligned} P(A=1, B, M=0) &= P(A=1, B=1, M=0) + P(A=1, B=0, M=0) \\ &= P(A=1|M=0)P(B=1|M=0)P(M=0) + \\ &\quad P(A=1|M=0)P(B=0|M=0)P(M=0) \\ &= (0 \cdot 0 \cdot 0.5) + (0 \cdot 1 \cdot 0.5) = 0 \end{aligned}$$

So:

$$P(M=1|A=1) = \frac{P(A=1, B, M=1)}{P(A=1, B, M=1) + P(A=1, B, M=0)} = \frac{0.5}{0.5+0} = 1$$

Bayesian Inference - Three Variables

The more information we have in our evidence, the less computations we will have to do to perform inference
For example, for computing:

$$P(M=1|A=1,B=1) = \frac{P(A=1,B=1,M=1)}{P(A=1,B=1,M=1)+P(A=1,B=1,M=0)}$$

we perform the following computations:

$$\begin{aligned} P(A=1,B=1,M=1) &= P(A=1|M=1)P(B=1|M=1)P(M=1) \\ &= (1 \cdot 1 \cdot 0.5) = 0.5 \\ P(A=1,B=1,M=0) &= P(A=1|M=0)P(B=1|M=0)P(M=0) \\ &= (0 \cdot 0 \cdot 0.5) = 0 \end{aligned}$$

That is, as all the variables are instantiated with particular values, in this case each one of the probabilities depends only on one of the possible cases of the distribution

A Cancer Diagnosis Example

Consider the problem of diagnosing whether a patient suffers a particular cancer C from the results of a series of three clinical tests: T1, T2 and T3

We believe that the right factorization is:

$$P(C, T_1, T_2, T_3) = P(T_1|C)P(T_2|C)P(T_3|C)P(C)$$

With the following conditional probability tables:

Table for any T_i:

v1	v2	$P(T_i = v1 \| C = v2)$
0	0	0.8
1	0	0.2
0	1	0.1
1	1	0.9

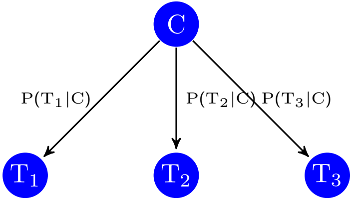
Marginal Probability for C:

v1	$P(C = v1)$
0	0.99
1	0.01

Exercise: What are the conditional independence relations that we are assuming with our factorization being right ?

A Cancer Diagnosis Example

So, this would be its corresponding Bayesian Network:



How does the probability of having cancer change as more tests give a positive result ?

A Cancer Diagnosis Example

First case: Three tests positive. Then, to compute:

$$P(C=1|T_1=1,T_2=1,T_3=1) = \frac{P(C=1,T_1=1,T_2=1,T_3=1)}{P(T_1=1,T_2=1,T_3=1)}$$

we perform the following computations:

$$\begin{aligned} P(C=1,T_1=1,T_2=1,T_3=1) &= \\ P(T_1=1|C=1)P(T_2=1|C=1)P(T_3=1|C=1)P(C=1) &= \\ (0.9 \cdot 0.9 \cdot 0.9 \cdot 0.01) &= 0.00729 \end{aligned}$$

We also need to compute the complementary event:

$$\begin{aligned} P(C=0,T_1=1,T_2=1,T_3=1) &= \\ P(T_1=1|C=0)P(T_2=1|C=0)P(T_3=1|C=0)P(C=0) &= \\ (0.2 \cdot 0.2 \cdot 0.2 \cdot 0.99) &= 0.00792 \end{aligned}$$

A Cancer Diagnosis Example

So, with the evidence of three positive tests, we have the following probability of having cancer:

$$P(C=1|T_1=1, T_2=1, T_3=1) = \frac{0.00729}{0.00729 + 0.00792} = 0.4792$$

Are things less clear when we only have an evidence with two positive tests ?

A Cancer Diagnosis Example

Exercise: Perform all the needed computations to check that for two positive tests the answer is:

$$P(C=1|T_1=1, T_2=1) = \frac{0.0081}{0.0081 + 0.0396} = 0.1698$$

This is considerably lower than when the evidence has three positive tests !

So, more evidence can change dramatically the answers for our queries !
Exercise: Perform all the needed computations to check that for two positive tests and one negative test the answer is:

$$P(C=1|T_1=1, T_2=1, T_3=0) = 0.024$$

General Bayesian Inference

Given a factorization for a model $P(V_1, \ldots, V_n)$ and:

- The query variable Q for which we want to measure the probability for each of its values
- An evidence E, given by a subset of variables with particular values

We can generalize the inference method we have seen in the previous examples to compute $P(Q|E)$
The next algorithm gives a general decomposition method for computing $P(Q, E)$ from our factorization of $P(V_1, \ldots, V_n)$

General Bayesian Inference

The following function recursively computes P(E), calling it with the Bayesian Network associated with our factorization

Function ExpandProb(BNet ,Vars, E)

if Vars=[] then return 1.0 $Y := \text{extractFirst}(\text{Vars}, \text{BNet})$;

if $Y = y_i \in E$ then
  return $P(Y = y_i|\text{Parents(Y,BNet,E)}) \times$ ExpandProb(BNet,Vars,E)

else
  // Y is free: sum up over all its possible values y_j
  return
  $\sum_{y_j} P(Y{=}y_j|\text{Parents(Y,BNet,E)}) \times \text{ExpandProb}(\text{BNet,Vars,E} \cup \{Y = y_j\})$
  ;

The variable ordering must satisfy that the parents of any node Y are always processed before Y, so when Parents(Y,BNet,E) is evaluated, all the parents of Y have a value assigned in E

General Bayesian Inference

Then, we compute $P(Q = v_j | E)$ computing $P(E \cup \{Q = v_i\})$ with the previous function for all the values v_i of Q, and then normalizing $P(E \cup \{Q = v_j\})$ by the sum of all these values

---

Function QueryVar(BNet ,Vars, E, $Q = v_j$)

$total_Q := 0$;

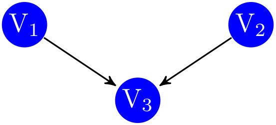
for each v_i of Q **do**

  $P_{Q=v_i} :=$ ExpandProb(BNet , Vars, $E \cup \{Q = v_i\}$) ;

  $total_Q := total_Q + P_{Q=v_i}$;

return $\frac{P_{Q=v_j}}{total_Q}$;

General Bayesian Inference - Example 1



We label the variables following the order in which they will be processed by the Inference algorithm. All the variables are binary, domain: $\{0, 1\}$. Query:

$$P(V_2 = 1 | E = \{V_1 = 1\})$$

General Bayesian Inference - Example 1

So, we need to compute $P(E' = \{V_1 = 1\} \cup \{V_2 = 1\})$:

$$
\begin{aligned}
&= \mathrm{ExProb}(\mathrm{BNet}, [V_1, V_2, V_3], E') \\
&= P(V_1{=}1)\mathrm{ExProb}(\mathrm{BNet}, [V_2, V_3], E') \\
&= P(V_1{=}1)(\ P(V_2{=}1)\mathrm{ExProb}(\mathrm{BNet}, [V_3], E') \\
&= P(V_1{=}1)(P(V_2{=}1)(P(V_3{=}0|V_1{=}1, V_2{=}1)\mathrm{ExProb}(\mathrm{BNet}, [], E' \cup \{V_3 = 0\}) + \\
&\qquad P(V_3{=}1|V_1{=}1, V_2{=}1)\mathrm{ExProb}(\mathrm{BNet}, [], E' \cup \{V_3 = 1\})) \\
&= P(V_1{=}1)(P(V_2{=}1)(P(V_3{=}0|V_1{=}1, V_2{=}1)1 + \\
&\qquad P(V_3{=}1|V_1{=}1, V_2{=}1)1) \\
&= P(V_1{=}1)(P(V_2{=}1)(P(V_3{=}0|V_1{=}1, V_2{=}1) + P(V_3{=}1|V_1{=}1, V_2{=}1)))
\end{aligned}
$$

General Bayesian Inference - Example 1

And we also need to compute $P(E' = \{V_1 = 1\} \cup \{V_2 = 0\})$:

$$
\begin{aligned}
&= \mathrm{ExProb}(\mathrm{BNet}, [V_1, V_2, V_3], E') \\
&= P(V_1 = 1)\mathrm{ExProb}(\mathrm{BNet}, [V_2, V_3], E') \\
&= P(V_1 = 1)(\ P(V_2 = 0)\mathrm{ExProb}(\mathrm{BNet}, [V_3], E') \\
&= P(V_1 = 1)(P(V_2 = 0)(P(V_3 = 0 | V_1 = 1, V_2 = 0)\mathrm{ExProb}(\mathrm{BNet}, [], E' \cup \{V_3 = 0\}) + \\
&\qquad P(V_3 = 1 | V_1 = 1, V_2 = 0)\mathrm{ExProb}(\mathrm{BNet}, [], E' \cup \{V_3 = 1\})) \\
&= P(V_1 = 1)(P(V_2 = 0)(P(V_3 = 0 | V_1 = 1, V_2 = 0)1 + \\
&\qquad P(V_3 = 1 | V_1 = 1, V_2 = 0)1) \\
&= P(V_1 = 1)(P(V_2 = 0)(P(V_3 = 0 | V_1 = 1, V_2 = 0) + P(V_3 = 1 | V_1 = 1, V_2 = 0)))
\end{aligned}
$$

So, the final conditional probability is:

$$P(V_2 = 1 | V_1 = 1) = \frac{P(V_1 = 1, V_2 = 1)}{P(V_1 = 1, V_2 = 1) + P(V_1 = 1, V_2 = 0)}$$

General Bayesian Inference - Example 1

We can understand how the recursive computation proceeds by noticing that it can be associated with the depth-first traversal of a computation tree associated with the recursive function call

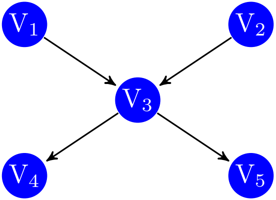
Computation tree for $P(V_2 = 1, V_1 = 1)$:

$$\begin{tikzcd}[column sep=small] & P(V_1 = 1) \arrow[d, "\times"'] & \\ & P(V_2 = 1) \arrow[d, "\times"'] & \\ & + \arrow[dl] \arrow[dr] & \\ P(V_3 = 0 | V_1 = 1, V_2 = 1) & & P(V_3 = 1 | V_1 = 1, V_2 = 1) \end{tikzcd}$$

Computation tree for $P(V_2 = 0, V_1 = 1)$:

$$\begin{tikzcd}[column sep=small] & P(V_1 = 1) \arrow[d, "\times"'] & \\ & P(V_2 = 0) \arrow[d, "\times"'] & \\ & + \arrow[dl] \arrow[dr] & \\ P(V_3 = 0 | V_1 = 1, V_2 = 0) & & P(V_3 = 1 | V_1 = 1, V_2 = 0) \end{tikzcd}$$

General Bayesian Inference - Example 2



We label the variables following the order in which they will be processed by the Inference algorithm. All the variables are binary, domain: $\{0, 1\}$. Query:

$$P(V_2 = 1 | V_1 = 1, V_4 = 1)$$

General Bayesian Inference - Example 2

So, we need to compute: $P(E' = \{V_2 = 1\} \cup \{V_1 = 1, V_4 = 1\})$:

$$= \mathrm{ExProb}(\mathrm{BNet}, [V_1, V_2, V_3, V_4, V_5], E')$$

$$= P(V_1 = 1)\mathrm{ExProb}(\mathrm{BNet}, [V_2, V_3, V_4, V_5], E')$$

$$= P(V_1 = 1)P(V_2 = 1)\mathrm{ExProb}(\mathrm{BNet}, [V_3, V_4, V_5], E'))$$

$$= P(V_1 = 1)P(V_2 = 1)(P(V_3 = 0|V_1 = 1, V_2 = 1)\mathrm{ExProb}(\mathrm{BNet}, [V_4, V_5], E' \cup \{V_3 = 0\}) +$$
$$P(V_3 = 1|V_1 = 1, V_2 = 1)\mathrm{ExProb}(\mathrm{BNet}, [V_4, V_5], E' \cup \{V_3 = 1\}))$$

$$= P(V_1 = 1)\ P(V_2 = 1)(P(V_3 = 0|V_1 = 1, V_2 = 1)P(V_4 = 1|V_3 = 0)\mathrm{ExProb}(\mathrm{BNet}, [V_5], E' \cup \{V_3 = 0\}) +$$
$$P(V_3 = 1|V_1 = 1, V_2 = 1)P(V_4 = 1|V_3 = 1)\mathrm{ExProb}(\mathrm{BNet}, [V_5], E' \cup \{V_3 = 1\}))$$

$$= P(V_1 = 1)\ P(V_2 = 1)(P(V_3 = 0|V_1 = 1, V_2 = 1)P(V_4 = 1|V_3 = 0)(P(V_5 = 0|V_3 = 0) + P(V_5 = 1|V_3 = 0)) +$$
$$P(V_3 = 1|V_1 = 1, V_2 = 1)P(V_4 = 1|V_3 = 1)(P(V_5 = 0|V_3 = 1) + P(V_5 = 1|V_3 = 1)))$$

And also $P(E' = \{V_2 = 0\} \cup \{V_1 = 1, V_4 = 1\})$:

$$= P(V_1 = 1)\ P(V_2 = 0)(P(V_3 = 0|V_1 = 1, V_2 = 0)P(V_4 = 1|V_3 = 0)(P(V_5 = 0|V_3 = 0) + P(V_5 = 1|V_3 = 0)) +$$
$$P(V_3 = 1|V_1 = 1, V_2 = 0)P(V_4 = 1|V_3 = 1)(P(V_5 = 0|V_3 = 1) + P(V_5 = 1|V_3 = 1)))$$

General Bayesian Inference - Example 2

Computation tree for $P(V_2=0, V_1=1, V_4=1)$:

- $P(V_1=1)$
  - $\downarrow \times$ $P(V_2=0)$
    - $\downarrow \times$ $+$
      - $P(V_3=0 \mid V_1=1, V_2=0)$
        - $\downarrow \times$ $P(V_4=1 \mid V_3=0)$
          - $\downarrow \times$ $+$
            - $P(V_5=0 \mid V_3=0)$
            - $P(V_5=1 \mid V_3=0)$
      - $P(V_3=1 \mid V_1=1, V_2=0)$
        - $\downarrow \times$ $P(V_4=1 \mid V_3=1)$
          - $\downarrow \times$ $+$
            - $P(V_5=0 \mid V_3=1)$
            - $P(V_5=1 \mid V_3=1)$

Computation tree for $P(V_2=1, V_1=1, V_4=1)$:

- $P(V_1=1)$
  - $\downarrow \times$ $P(V_2=1)$
    - $\downarrow \times$ $+$
      - $P(V_3=0 \mid V_1=1, V_2=1)$
        - $\downarrow \times$ $P(V_4=1 \mid V_3=0)$
          - $\downarrow \times$ $+$
            - $P(V_5=0 \mid V_3=0)$
            - $P(V_5=1 \mid V_3=0)$
      - $P(V_3=1 \mid V_1=1, V_2=1)$
        - $\downarrow \times$ $P(V_4=1 \mid V_3=1)$
          - $\downarrow \times$ $+$
            - $P(V_5=0 \mid V_3=1)$
            - $P(V_5=1 \mid V_3=1)$

General Bayesian Inference - Complexity

Worst-case time complexity

In the worst case, when the number of variables in the evidence is bounded by a constant, the total time to traverse the recursive computation tree will be exponential in the number of variables

Worst-case space complexity

The space needed is the one needed to store the final computed value plus the space needed for the current branch on the recursive computation tree.
So, it is lineal in the size of the factorization

And if the number of parents per variable is bounded by a constant, the total space needed will be polynomial in the number of variables

General Bayesian Inference - Complexity

A better algorithm for exact inference ?

But is this only a deficiency of the particular algorithm we have analyzed ?

The answer seems to be : No !!!

Given that the problem of exact inference with bayesian networks is NP−hard, and assuming that NP $\neq$ P, there is no polynomial time algorithm for solving this problem

However, there some algorithms for approximate inference (computing probability values that approximate the real value) that work on polynomial time. For example: sampling algorithms like Gibbs sampling and Particle Filters