Chapter 3
Probabilistic Reasoning

Paragraph 1
Bayesian Inference

Example

- You got a baby, wrapped it up nicely, and went for a walk in the fresh winter air.
- After walking for an hour, it suddenly occurs to you that the diaper may need to get changed so that the baby's skin does not get sore.
- Given the cold air, you would rather not have to check directly if the diaper needs changing.
- You observe: The baby cries, but you do not smell anything.

Limits of Determinism

- One could of course try and set up a deterministic knowledge base KB for the problem of inferring whether the diapers are full or not.
- Consider the fact "baby cries." Of course, a full diaper could be the reason, but there are many other reasons as well: the baby may be cold, the baby might be bored, the baby might be ill, the baby may be hungry…
- Consider the fact "no smell." A full diaper may cause a stench, but there might be occasions when it is not recognizable: maybe the baby is just wet (no poo), or the clothing may be too thick, or you might have a beginning cold…

Limits of Determinism

- First of all, we do not want to have to enumerate all potential causes for the baby's crying or for the fact that you do not smell anything.
- Secondly, in the knowledge base we could only ever infer that the fact "diapers are full" holds when this is 100% guaranteed to be the case. In most realistic scenarios, there exist both the option that the diapers are full and that they are not.

Probabilistic Inference

- What we want is an inference mechanism that allows us to judge how likely it is that the diapers are full.
- As an experienced parent, you estimate that there is a 60% chance that the baby has wet itself, and a 25% chance that it did even more.
- In the latter case, the diaper will need to be changed in any case. If the baby is just wet, with 20% chance the diaper is able to manage and needs no changing. Even if the baby stayed clean, there is a 5% chance that the diaper needs changing anyway (e.g. because of sweating).

Conditional Probability

- Given just the prior probability of pee and poo and their effects on the diaper, what is the conditional probability that the diaper needs changing?

Conditional Probability

	Poo (0.25)		No Poo (0.75)		
	Pee (0.6)	no Pee (0.4)	Pee (0.6)	no Pee (0.4)	Sum
New Diaper	0.15	0.1	0.36 = 0.8x0.45	0.015 = 0.05x0.3	0.625
No New Diaper	0	0	0.09 = 0.2x0.45	0.285 = 0.95x0.3	0.375
Sum	0.15 = 0.25x0.6	0.1 = 0.25x0.4	0.45 = 0.75x0.6	0.3 = 0.75x0.4	1

Full Joint Probability Distribution

- P(diaper) = 0.625
- P(diaper | pee)
  = (0.15+0.36)/0.6
  = 0.85
- P(¬diaper | pee)
  = 0.09/0.6 = 0.15
- P(diaper | ¬pee)
  = 0.115/0.4 = 0.2875
- P(poo | diaper,pee)
  = 0.15/(0.15+0.36)
  = 0.294

	Poo (0.25)		No Poo (0.75)		
	Pee (0.6)	no Pee (0.4)	Pee (0.6)	no Pee (0.4)	Sum
New Diaper	0.15	0.1	0.36 = 0.8x0.45	0.015 = 0.05x0.3	0.625
No New Diaper	0	0	0.09 = 0.2x0.45	0.285 = 0.95x0.3	0.375
Sum	0.15 = 0.25x0.6	0.1 = 0.25x0.4	0.45 = 0.75x0.6	0.3 = 0.75x0.4	1

X,Y independent iff $P(X|Y) = P(X)$

$P(Y) = \Sigma_z P(Y,z) = \Sigma_z P(Y|z) P(z)$

$P(X|Y) = P(X \cap Y) / P(Y)$

Bayes' Rule: $P(Y|X) = P(X|Y)P(Y)/P(X)$

Bayesian Inference

- We can infer all (conditional) probabilities from the full joint probability distribution.
- However, the corresponding table has a lot of entries. Can you tell how many?
- For n Boolean variables, in the worst case the full table has $\theta(2^n)$ entries!
- How long does it take to answer a query from the table?
- For n Boolean variables, we can answer queries in $O(n2^n)$ steps.

An Efficient Representation

- The first thing that we need is an efficient compact representation of a probabilistic problem.
- We can achieve a much smaller representation of the full joint probability distribution if we factor out independent random variables (such as poo and pee in our baby example).
- Note that sparsely connected independent variables frequently occur in real-world scenarios since components of real systems are usually only locally influenced by the few immediate neighbors. We speak of locally structured and sparse systems in this case.

An Efficient Representation of the complete Diaper Problem

Pee

P(pee)	P(¬pee)
0.6	0.4

Poo

P(poo)	P(¬poo)
0.25	0.75

Query Variable

Full Diaper

pee	poo	P(d)	P(¬d)
true	true	1	0
true	false	0.8	0.2
false	true	1	0
false	false	0.95	0.05

Crying

d	P(c)	P(¬c)
true	0.9	0.1
false	0.6	0.4

Stench

d	P(s)	P(¬s)
true	0.9	0.1
false	0.6	0.4

Hidden Variables

Observations

An Efficient Representation of the complete Diaper Problem

Pee

P(pee)	P(¬pee)
0.6	0.4

Poo

P(poo)	P(¬poo)
0.25	0.75

Query Variable

Full Diaper

pee	poo	P(d)	P(¬d)
true	true	1	0
true	false	0.8	0.2
false	true	1	0
false	false	0.95	0.05

Crying

d	P(c)	P(¬c)
true	0.9	0.1
false	0.6	0.4

Stench

d	P(s)	P(¬s)
true	0.9	0.1
false	0.6	0.4

$P(d|pee,c) = ?$

An Efficient Representation of the complete Diaper Problem

Pee

P(pee)	P(¬pee)
0.6	0.4

Poo

P(poo)	P(¬poo)
0.25	0.75

Query Variable

Full Diaper

pee	poo	P(d)	P(¬d)
true	true	1	0
true	false	0.8	0.2
false	true	1	0
false	false	0.95	0.05

Crying

d	P(c)	P(¬c)
true	0.9	0.1
false	0.6	0.4

Stench

d	P(s)	P(¬s)
true	0.9	0.1
false	0.6	0.4

$$P(d|pee,c) = \alpha \Sigma_{p\in Poo}\Sigma_{s\in S} P(pee)P(p)P(d|pee,p)P(c|d)P(s|d)$$

An Efficient Representation of the complete Diaper Problem

Pee

P(pee)	P(¬pee)
0.6	0.4

Poo

P(poo)	P(¬poo)
0.25	0.75

Query Variable

Full Diaper

pee	poo	P(d)	P(¬d)
true	true	1	0
true	false	0.8	0.2
false	true	1	0
false	false	0.95	0.05

Crying

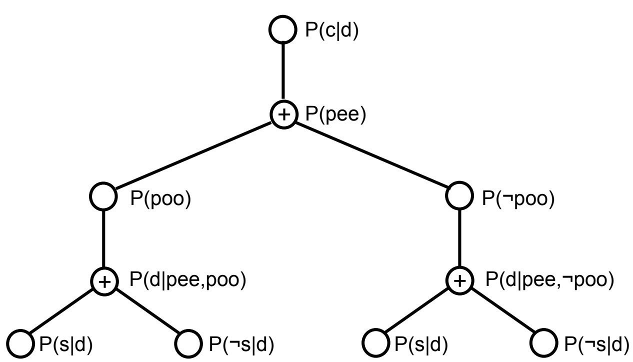
d	P(c)	P(¬c)
true	0.9	0.1
false	0.6	0.4

Stench

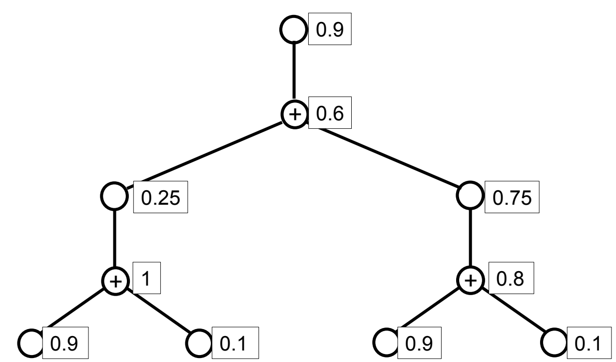
d	P(s)	P(¬s)
true	0.9	0.1
false	0.6	0.4

$$P(d|pee,c) = \alpha P(c|d) P(pee) \Sigma_{p\in Poo}P(p) P(d|pee,p) \Sigma_{s\in S} P(s|d)$$

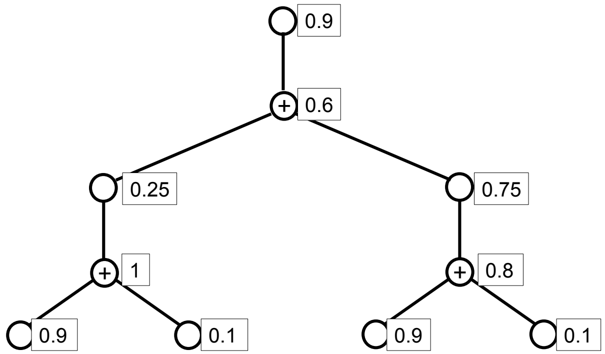
Answering Queries



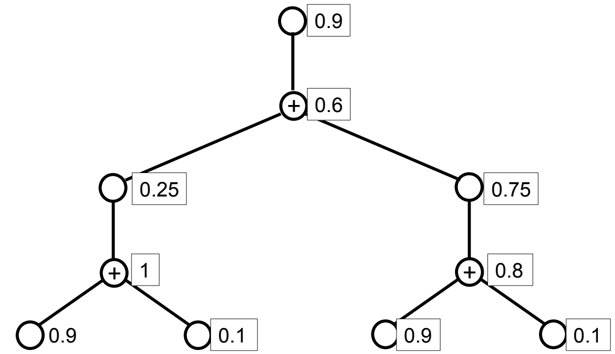
Answering Queries



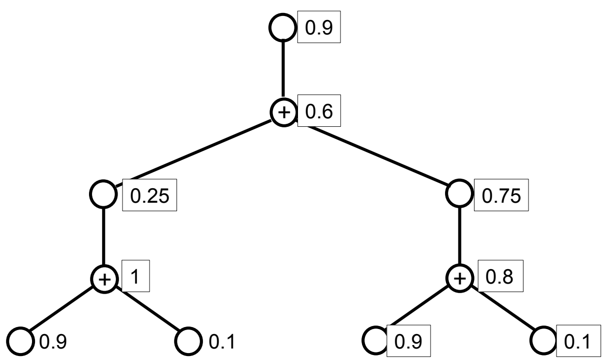
Answering Queries



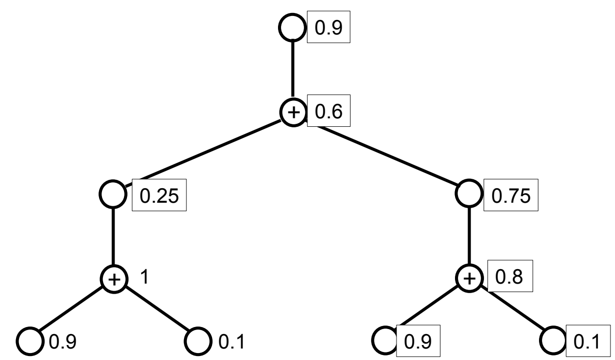
Answering Queries



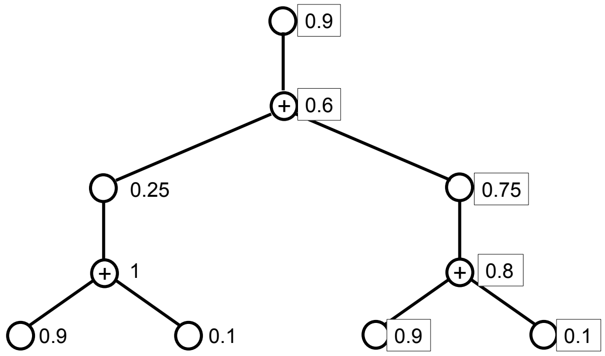
Answering Queries



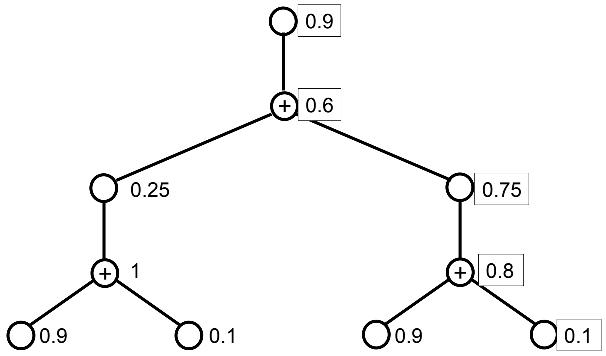
Answering Queries



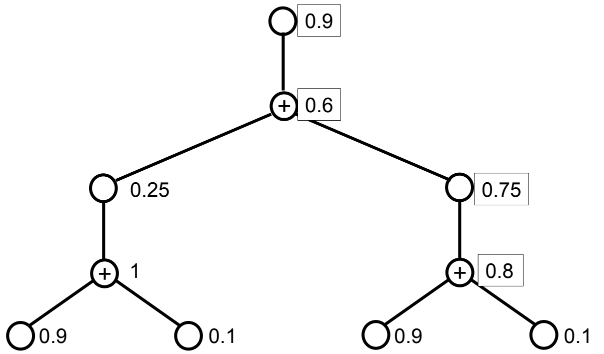
Answering Queries



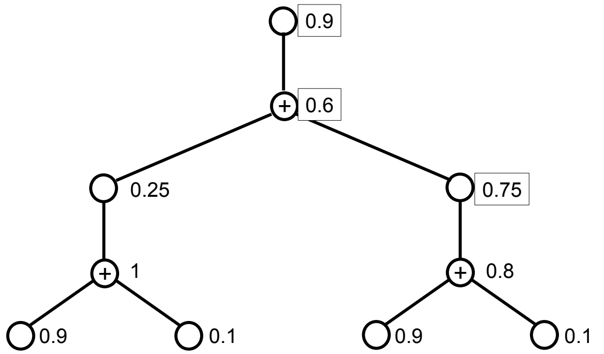
Answering Queries



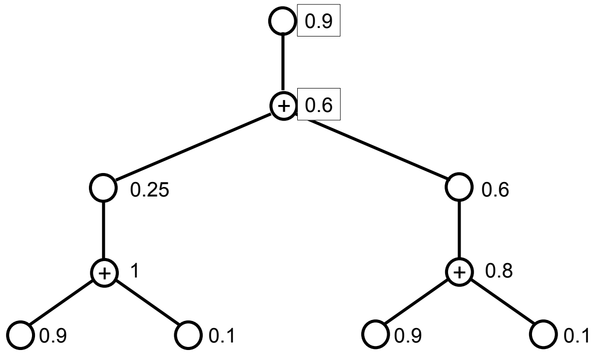
Answering Queries



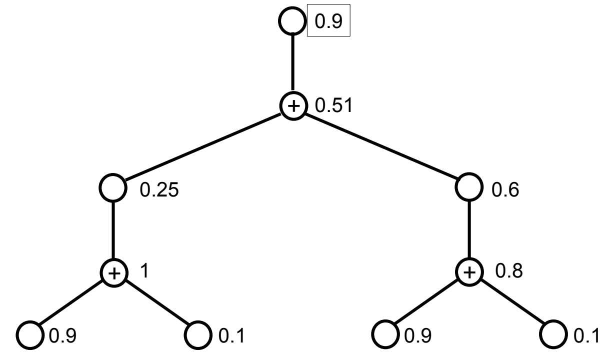
Answering Queries



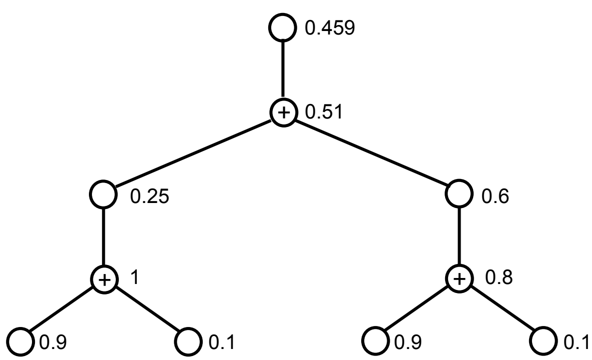
Answering Queries



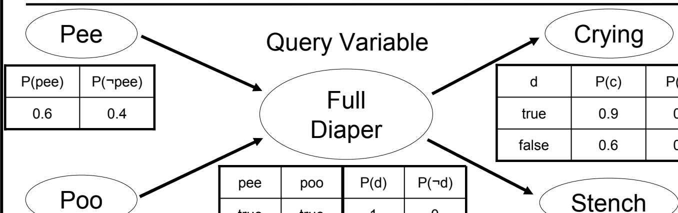
Answering Queries



Answering Queries



An Efficient Representation of the complete Diaper Problem



P(pee)	P(¬pee)
0.6	0.4

P(poo)	P(¬poo)
0.25	0.75

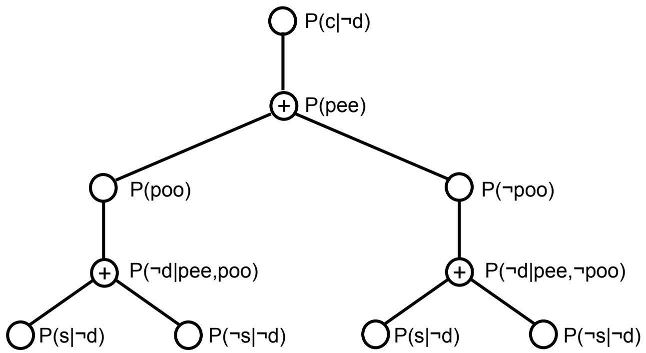
pee	poo	P(d)	P(¬d)
true	true	1	0
true	false	0.8	0.2
false	true	1	0
false	false	0.95	0.05

d	P(c)	P(¬c)
true	0.9	0.1
false	0.6	0.4

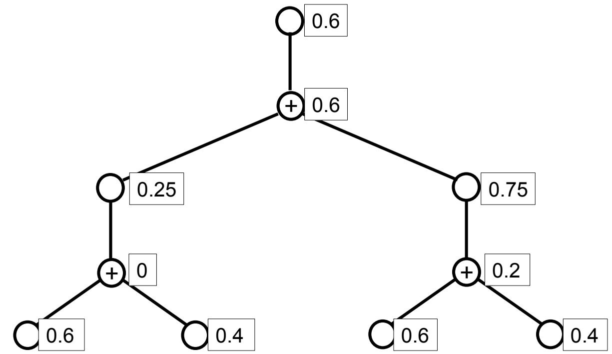
d	P(s)	P(¬s)
true	0.9	0.1
false	0.6	0.4

$P(\neg d|pee,c) = \alpha\, P(c|\neg d)\, P(pee)\, \Sigma_{p\in Poo} P(p)\, P(\neg d|pee,p)\, \Sigma_{s\in S}\, P(s|\neg d)$

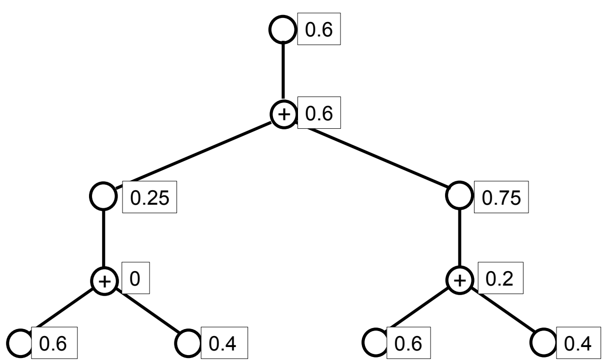
Answering Queries



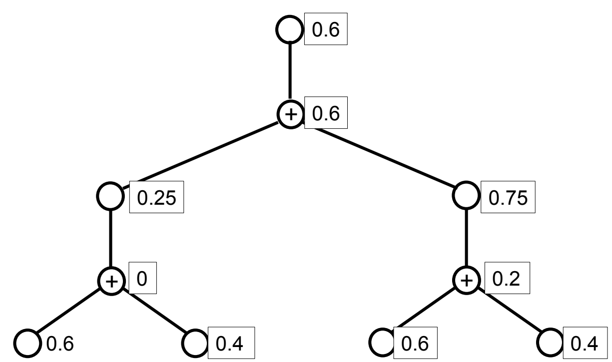
Answering Queries



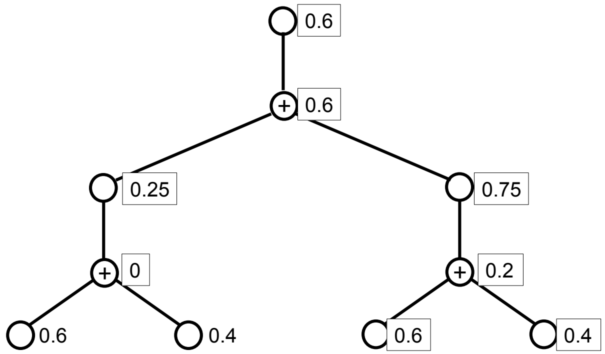
Answering Queries



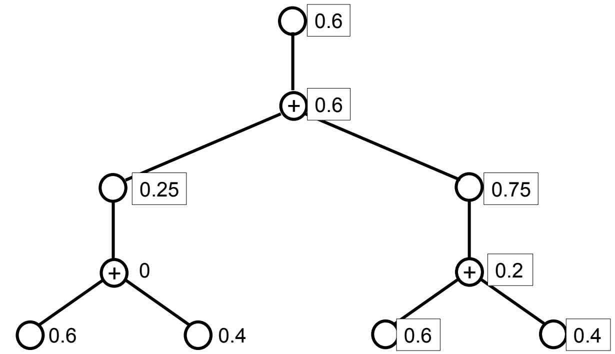
Answering Queries



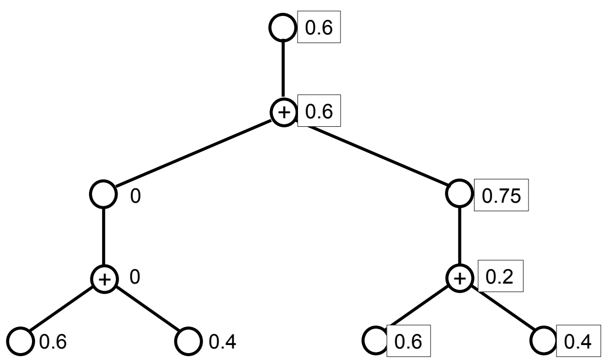
Answering Queries



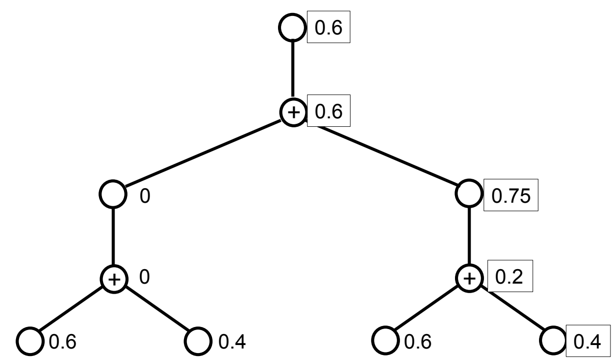
Answering Queries



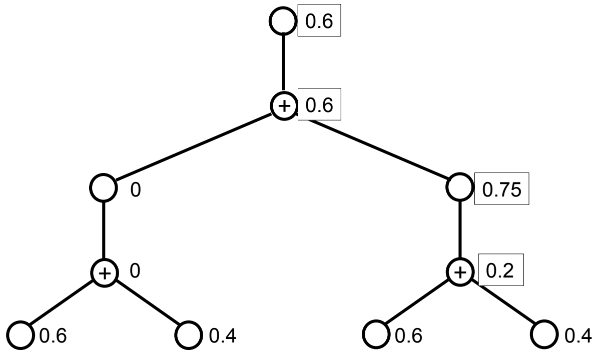
Answering Queries



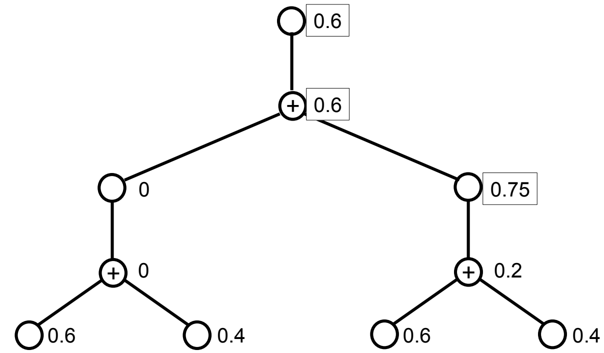
Answering Queries



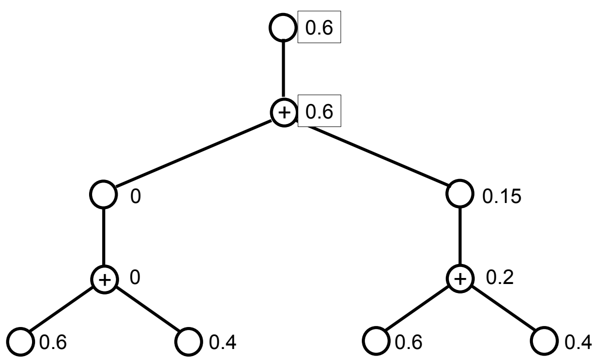
Answering Queries



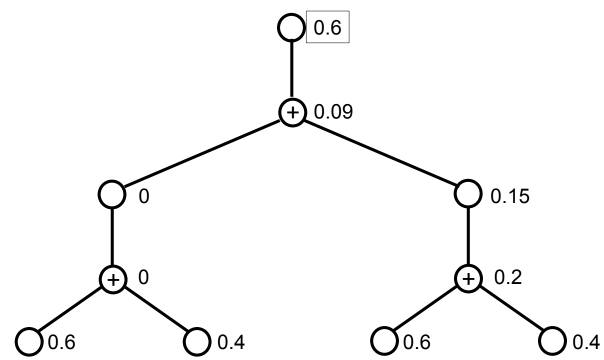
Answering Queries



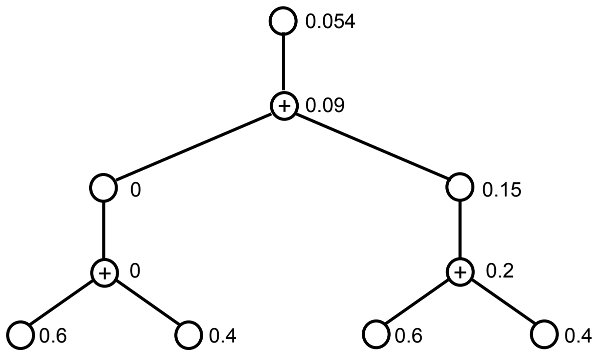
Answering Queries



Answering Queries



Answering Queries



Answering Queries

- Consequently, if we hear the baby cry and if we know that the baby has wet itself, then the probability distribution of “new diaper” is

$$P(D|pee,c) = \alpha <0.459,0.054> = <0.8947,0.1053>$$

- That is, with more than 89% the diaper needs changing!

Complexity of Exact Inference

- The previous algorithm needs time exponential in the number of random variables.
- We observe: it wastes a lot of time!
  - mostly by re-computing values several times
  - by computing values irrelevant to the query.

The Variable Elimination Algorithm

- We limit us to relevant computations only and avoid double work (same idea as in dynamic programming).
- We eliminate variables in anti-topological order of the Bayesian network.
- We ignore variables that are not ancestors of at least one evidence or query variable.

The Variable Elimination Algorithm

$P(D|pee,c) = \alpha\, P(pee)\, \Sigma_{p \in Poo}\, P(p)\, P(D|pee,p)\, P(c|D)\, \Sigma_{s \in S}\, P(s|D)$

The Variable Elimination Algorithm

$P(D|pee,c) = \alpha\, P(pee)\, \Sigma_{p \in Poo}\, P(p)\, P(D|pee,p)\, P(c|D)$

D	P(c|D)
true	0.9
false	0.6

D	Poo	P(D|pee,Poo)
true	true	1
true	false	0.8
false	true	0
false	false	0.2

Poo	P(Poo)
true	0.25
false	0.75

D	Poo	f1(D,Poo)
true	true	0.25
true	false	0.6
false	true	0
false	false	0.15

D	f2(D)
true	0.85
false	0.15

The Variable Elimination Algorithm

$P(D|pee,c) = \alpha\, P(pee)\, \Sigma_{p \in Poo}\, P(p)\, P(D|pee,p)\, P(c|D)$

D	P(c|D)
true	0.9
false	0.6

D	Poo	P(D|pee,Poo)
true	true	1
true	false	0.8
false	true	0
false	false	0.2

Poo	P(Poo)
true	0.25
false	0.75

P(pee)
0.6

D	P(c|D)
true	0.9
false	0.6

D	f2(D)
true	0.85
false	0.15

D	f3(D|pee,c)
true	0.459
false	0.054

Complexity of Inference

- Variable elimination on singly-connected Bayesian networks takes linear time (linear in the size of all conditional probability tables (CPT)).
- For general networks, exact inference is NP-hard (reduce to 3-SAT).

Non-Exact Inference: Rejection Sampling

- Another way of computing conditional inference is to draw values according to the CPT for each variable in topological ordering.
- Check all samples that match the evidence and use the relative frequencies of the query variable as approximation of the true conditional probability.
- More samples matching the evidence mean higher accuracy. However, in complex scenarios too many samples do not match the evidence.