# Bayesian Networks

# 4. Exact Inference / Variable Elimination

Lars Schmidt-Thieme

Information Systems and Machine Learning Lab (ISMLL)
Institute for Business Economics and Information Systems
& Institute for Computer Science
University of Hildesheim
http://www.ismll.uni-hildesheim.de





**1. Inference in Probabilistic Networks**

**2. Variable elimination**

## studfarm example

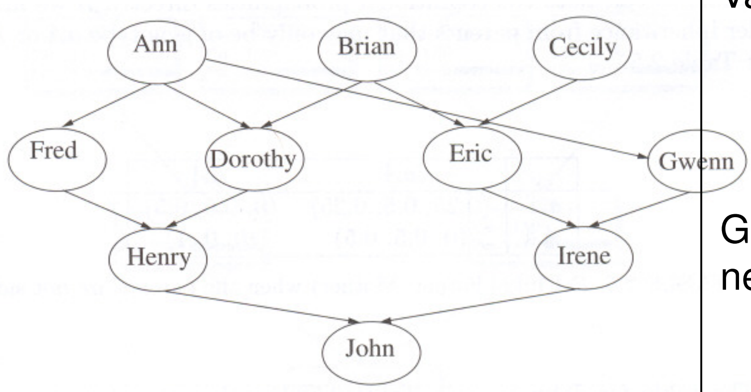



Figure 1: Genealogical structure for the horses in the studfarm example [Jen01, p. 47].

| | aa | aA | AA |
|---|---|---|---|
| aa | (1, 0, 0) | (0.5, 0.5, 0) | (0, 1, 0) |
| aA | (0.5, 0.5, 0) | (0.25, 0.5, 0.25) | (0, 0.5, 0.5) |
| AA | (0, 1, 0) | (0, 0.5, 0.5) | (0, 0, 1) |

Figure 2: p(Child | Father, Mother) for genetic inheritance. The numbers are the probabilities for (aa, aA, AA) [Jen01, p. 47].

Variable *disease* with three states:

pure (aa) carrier (aA) sick (AA)

Genalogic graph becomes bayesian network if

(i) each non-root vertex has conditional probability distribution

$$p(\text{child}|\text{father}, \text{mother})$$

as given in fig. 2,

(ii) each root vertex has probability distribution

$$p(aa) = .99, p(aA) = .01, p(AA) = .0$$





## studfarm example

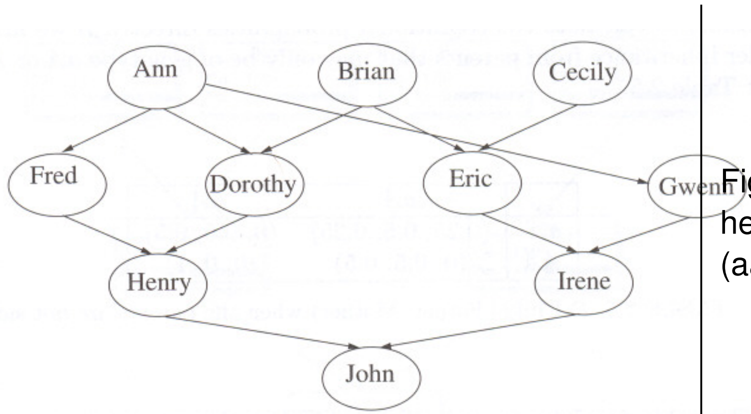



Figure 3: Genealogical structure for the horses in the studfarm example [Jen01, p. 47].

| | aa | aA | AA |
|---|---|---|---|
| aa | (1, 0, 0) | (0.5, 0.5, 0) | (0, 1, 0) |
| aA | (0.5, 0.5, 0) | (0.25, 0.5, 0.25) | (0, 0.5, 0.5) |
| AA | (0, 1, 0) | (0, 0.5, 0.5) | (0, 0, 1) |

Figure 4: p(Child | Father, Mother) for genetic inheritance. The numbers are the probabilities for (aa, aA, AA) [Jen01, p. 47].

| father | aa | | | aA | | | AA | | |
|---|---|---|---|---|---|---|---|---|---|
| mother | aa | aA | AA | aa | aA | AA | aa | aA | AA |
| aa | 1 | .5 | 0 | .5 | .25 | 0 | 0 | 0 | 0 |
| aA | 0 | .5 | 1 | .5 | .5 | .5 | 1 | .5 | 0 |
| AA | 0 | 0 | 0 | 0 | .25 | .5 | 0 | .5 | 1 |

| father | aa | | aA | |
|---|---|---|---|---|
| mother | aa | aA | aa | aA |
| aa | 1 | .5 | .5 | .25 |
| aA | 0 | .5 | .5 | .5 |
| AA | 0 | 0 | 0 | .25 |

| father | aa | | aA | |
|---|---|---|---|---|
| mother | aa | aA | aa | aA |
| aa | 1 | .5 | .5 | $\frac{1}{3}$ |
| aA | 0 | .5 | .5 | $\frac{2}{3}$ |

Figure 5: p(child | father, mother) in general (left), if father and mother cannot be sick (middle), and if child cannot be sick either (right).

## studfarm example / "forward inference"

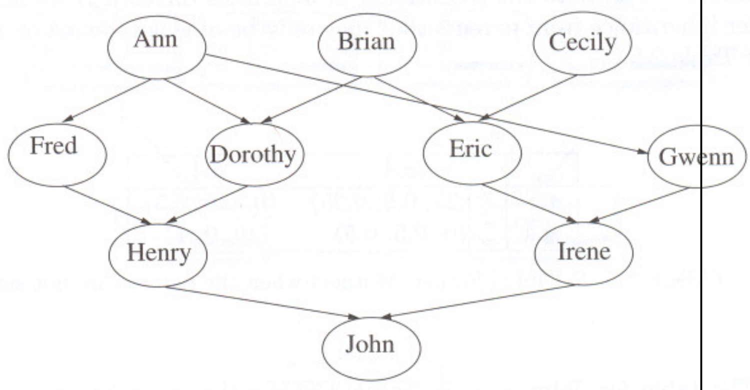



Figure 6: Genealogical structure for the horses in the studfarm example [Jen01, p. 47].

| father | aa | | aA | |
|---|---|---|---|---|
| mother | aa | aA | aa | aA |
| aa | 1 | .5 | .5 | .25 |
| aA | 0 | .5 | .5 | .5 |
| AA | 0 | 0 | 0 | .25 |

Figure 7: p(child | father, mother) if father and mother cannot be sick.

$$\begin{aligned} p(aa) = & && 0.99 \cdot 0.99 \\ & +2 \cdot \frac{1}{2} \cdot && 0.99 \cdot 0.01 \\ & +\frac{1}{4} \cdot && 0.01 \cdot 0.01 \\ & =0.990025 \end{aligned}$$

$$\begin{aligned} p(aA) = & +2 \cdot \frac{1}{2} \cdot && 0.99 \cdot 0.01 \\ & +\frac{1}{2} \cdot && 0.01 \cdot 0.01 \\ & =0.00995 \end{aligned}$$

$$\begin{aligned} p(AA) = & +\frac{1}{4} \cdot && 0.01 \cdot 0.01 \\ & =0.000025 \end{aligned}$$





## studfarm example / "forward inference"

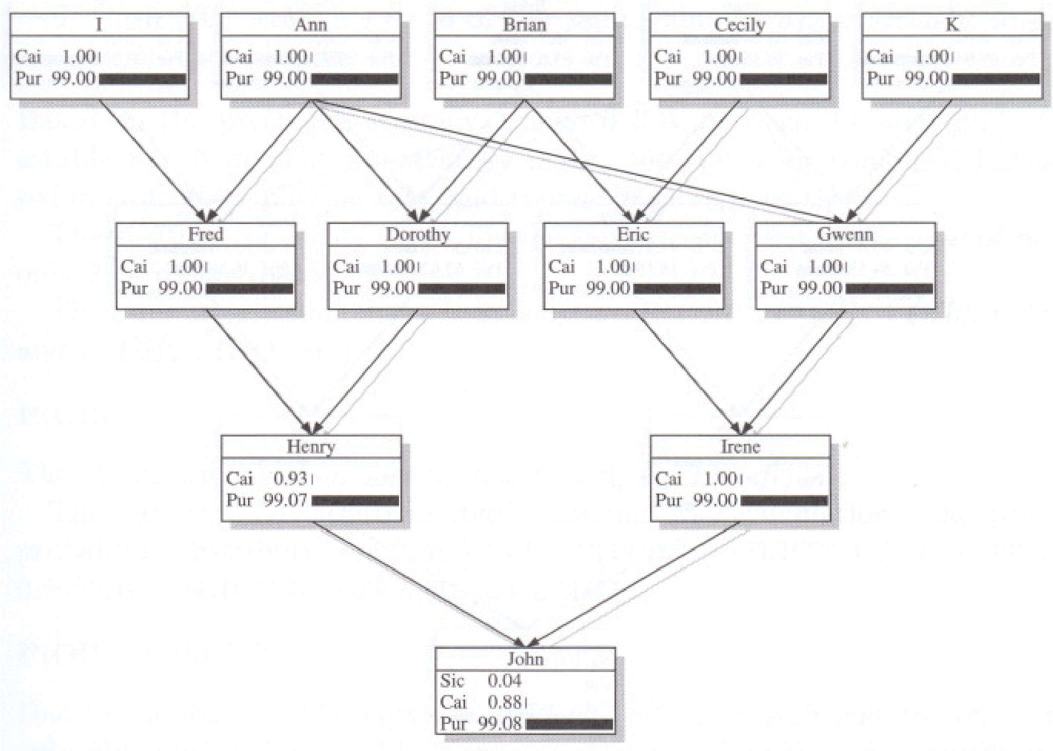



Figure 8: Probabilities without evidence. [Jen01, p. 49]

## studfarm example / "backward inference"

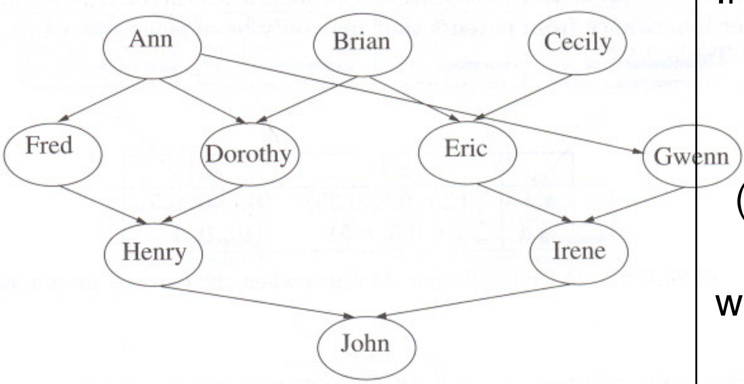



Figure 9: Genealogical structure for the horses in the studfarm example [Jen01, p. 47].

| father | aa | | aA | |
|---|---|---|---|---|
| mother | aa | aA | aa | aA |
| aa | 1 | .5 | .5 | .25 |
| aA | 0 | .5 | .5 | .5 |
| AA | 0 | 0 | 0 | .25 |

Figure 10: p(child | father, mother) if father and mother cannot be sick.

If we know, that

(i) all horses but John are not sick and

(ii) John is sick (AA),

we can infer that

(iii) Henry and Irene are carrier (aA) with $p = 1$.

If only Fred, Dorothy, Erik, and Gwen existed, we could further infer that for each of them

$$p(aa) = \frac{1}{3}, \quad p(aA) = \frac{2}{3}$$





## studfarm example / "backward inference"

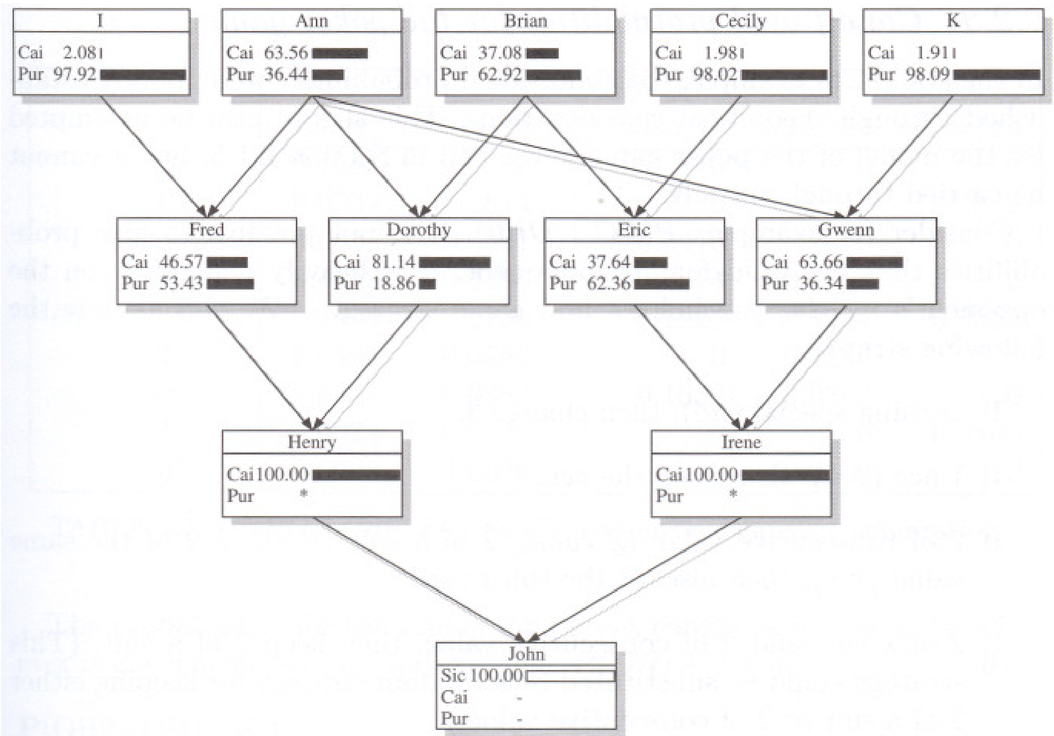



Figure 11: Probabilities given evidence that John is sick (AA). [Jen01, p. 49]

## Evidence

**Definition 1.** Let $V$ be a set of variables. The set

$$\mathcal{E} := \{E \subseteq \bigcup_{v \in V} \{v\} \times \mathrm{dom}(v) \,|\, \forall (v,c),(v,c') \in E : c = c'\}$$

is called **space of evidence of** $V$.
An element $E \in \mathcal{E}$ is called **evidence of** $V$. We call

$$\mathrm{dom}(E) := \{v \in V \,|\, \exists c \in \mathrm{dom}(v) : (v,c) \in E\}$$

the **set of evidential variables** and for each evidential variable $v \in \mathrm{dom}(E)$ we call the unique $E_v := c \in \mathrm{dom}(v)$ with $(v,c) \in E$ its **(evidential) value**.
Evidence $E$ corresponds to the probability distribution

Evidence is a setting of variables to specific values. "Fuzzy" or "uncertain evidence" that assigns probabilities to the different values of the variables, is not handled here.

$$\begin{aligned} \mathrm{epd}_E : \prod_{v \in \mathrm{dom}(E)} \mathrm{dom}(v) &\rightarrow \mathbb{R}_0^+ \\ (x)_{v \in \mathrm{dom}(E)} &\mapsto \begin{cases} 1, & \text{if } \forall v : (v,x) \in E \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$





## Evidence / example

**Example 1.** Let $V := \{A, B, C, D\}$ and

$$\begin{aligned} \mathrm{dom}(A) := \mathrm{dom}(B) &:= \{0,1\}, \\ \mathrm{dom}(C) &:= \{0,1,2\} \text{ and} \\ \mathrm{dom}(A) &:= \{0,1,2,3\}. \end{aligned}$$

Then

$$E := \{(A,1),(C,2)\}$$

is an evidence with the evidential variables $A$ and $C$. The evidential variable $A$ has value $1$, the variable $C$ value $2$.

The probability distribution corresponding to $E$ is

$$\mathrm{epd}_E(A = 1, C = 2) = 1$$

and

$$\mathrm{epd}_E(A = a, C = c) = 0$$

for all other values $a$ of $A$ or $c$ of $C$.

## Entering evidence

Let $V$ be a set of variables and $q$ be a potential on a subset of $V$. Let $E$ be evidence of $V$.

We call

$$\begin{aligned} q_E : \prod_{v \in \mathrm{dom}(q) \setminus \mathrm{dom}(E)} \mathrm{dom}(v) &\to \mathbb{R}_0^+ \\ (x)_{v \in \mathrm{dom}(q) \setminus \mathrm{dom}(E)} &\mapsto q(x, E) \end{aligned}$$

with

$$(x, E)(v) := \begin{cases} x_v, & \text{if } v \in \mathrm{dom}(q) \setminus \mathrm{dom}(E) \\ E_v, & \text{if } v \in \mathrm{dom}(E) \end{cases}$$

**the potential $q$ given evidence $E$.**

If $q$ is a JPD, then $q_E$ is the probability distribution on the non-evidential variables $\mathrm{dom}(q) \setminus \mathrm{dom}(E)$ for outcomes that conform to $E$ (i.e., have value $E_v$ for each variable $v \in \mathrm{dom}(E)$).

Warning: $q_E$ should not be confused with the conditional probability distribution $q^{|\,\mathrm{dom}(E)}$. In sloppy notation for $E = \{(v_1, c_1), \ldots, (v_n, c_n)\}$:

$$q_E = q(x, v_1 = c1, \ldots, v_n = c_n)$$

and

$$q^{|\,\mathrm{dom}(E)} = q(x \,|\, v_1, \ldots, v_n)$$





## Inferencing

Given a JPD $p$ on a set of variables $V$ and evidenve $E$ on $V$.

We distinguish three types of inference targets:

**(i) a single variable:** For a given variable $v \in V$ **infering $v$ based on $E$ w.r.t. $p$** means to compute

$$p(v|E) = \frac{p(v, E)}{p(E)} \sim p(v, E)$$

or (more exactly) $(p_E)^{\downarrow v|\emptyset}$.

**(ii) several variables separately:** For a given set of variables $W \subseteq V$ **infering $W$ separately based on $E$ w.r.t. $p$** means to compute

$$p(v|E) = \frac{p(v, E)}{p(E)} \sim p(v, E), \quad \forall v \in W$$

or $(p_E)^{\downarrow v|\emptyset}$

**(iii) joint distribution of several variabl**

For a given set of variables $W \subseteq V$ **infering the marginal $W$ based on $E$ w.r.t. $p$** means to compute

$$p(W|E) = \frac{p(W, E)}{p(E)} \sim p(W, E)$$

or $(p_E)^{\downarrow W|\emptyset}$

Normalizing is necessary, as $p_E$ in general is not a probability distribution, even if $p$ is.

## Inferencing / JPD as one large table

If $p$ is given as one large table, infering the marginal $W$ based on $E$ means

(i) select the subtable indexed by $E$,

(ii) aggregate to $W$, i.e., sum over all variables $V \setminus \mathrm{dom}(E) \setminus W$,

(iii) normalize.

| Pain | Y | | | | N | | | |
|---|---|---|---|---|---|---|---|---|
| Weightloss | Y | | N | | Y | | N | |
| Vomiting | Y | N | Y | N | Y | N | Y | N |
| Adeno Y | 220 | 220 | 25 | 25 | 95 | 95 | 10 | 10 |
| N | 4 | 9 | 5 | 12 | 31 | 76 | 50 | 113 |

Figure 12: JPD $p$ given as one large table.

| Pain | Y | | N | |
|---|---|---|---|---|
| Weightloss | Y | N | Y | N |
| Adeno Y | 220 | 25 | 95 | 10 |
| N | 4 | 5 | 31 | 50 |

Figure 13: Subtable for $E = \{(V, Y)\}$: distribution $p_E$ before normalization.

If we observe the evidence $V = Y$, then

$$p(\text{adeno}{=}Y|V = Y) = \sum_{w,q} p(\text{adeno}{=}Y, W = w, P = q|V = Y)$$
$$= \frac{220 + 25 + 95 + 10}{224 + 30 + 126 + 60} = \frac{350}{440} = 0.80$$





## Inferencing / JPD as product of potentials

If $p$ is given as product of potentials, i.e.,

$$p := (\prod_{q \in Q} q)^{|\emptyset}$$

the problem becomes more interesting.

**Naive approach:** we reduce the problem to inference w.r.t. $p$ as one large table by explicitly computing $p$ and then doing inference as on the former slide, actually computing

$$(p_E)^{\downarrow W|\emptyset} = (((\prod_{q \in Q} q)^{|\emptyset})_E)^{\downarrow W|\emptyset}$$

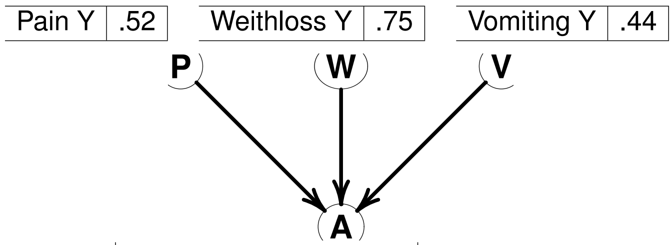
**Naive approach$_2$:** we

(i) enter evidence in the factors first, i.e., compute $q_E$, and then

(ii) compute $p_E$ as product of the $q_E$'s

$$(p_E)^{\downarrow W|\emptyset} = ((\prod_{q \in Q} q_E)^{\downarrow W|\emptyset})$$

## product of potentials / naive approach




| Pain | Y | | | | N | | | |
|---|---|---|---|---|---|---|---|---|
| Weightloss | Y | | N | | Y | | N | |
| Vomiting | Y | N | Y | N | Y | N | Y | N |
| Adeno Y | .982 | .961 | .833 | .676 | .754 | .556 | .167 | .081 |

Figure 14: Bayesian Network for adeno JPD.

| Pain | Y | | | | N | | | |
|---|---|---|---|---|---|---|---|---|
| Weightloss | Y | | N | | Y | | N | |
| Vomiting | Y | N | Y | N | Y | N | Y | N |
| Adeno Y | .169 | .210 | .048 | .049 | .119 | .112 | .009 | .005 |
| N | .003 | .009 | .010 | .024 | .039 | .090 | .044 | .062 |

Figure 15: JPD of Bayesian Network for adeno JPD.





## product of potentials / naive approach

| Pain | Y | | | | N | | | |
|---|---|---|---|---|---|---|---|---|
| Weightloss | Y | | N | | Y | | N | |
| Vomiting | Y | N | Y | N | Y | N | Y | N |
| Adeno Y | .169 | .210 | .048 | .049 | .119 | .112 | .009 | .005 |
| N | .003 | .009 | .010 | .024 | .039 | .090 | .044 | .062 |

Figure 16: JPD $p$ given as one large table.

| Pain | Y | | N | |
|---|---|---|---|---|
| Weightloss | Y | N | Y | N |
| Adeno Y | .169 | .048 | .119 | .009 |
| N | .003 | .010 | .039 | .044 |

Figure 17: Subtable for $E = \{(V, Y)\}$: distribution $p_E$ before normalization.

| | |
|---|---|
| Adeno Y | .345 |
| N | .096 |

Figure 18: Aggregate subtable for $E = \{(V, Y)\}$.

If we observe the evidence $V = Y$, then

$$p(\text{adeno}{=}Y|V = Y) = \sum_{w,q} p(\text{adeno}{=}Y, W = w, P = q|V = Y)$$

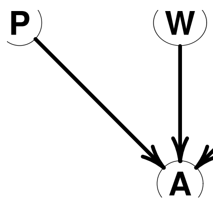
$$= \frac{.345}{.345 + .096} = 0.782$$

## product of potentials / naive approach$_2$

| | |
|---|---|
| Pain Y | .52 |

| | |
|---|---|
| Weithloss Y | .75 |

| | |
|---|---|
| Vomiting Y | ~~.44~~ 1.0 |




| Pain | Y | | | | N | | | |
|---|---|---|---|---|---|---|---|---|
| Weightloss | Y | | N | | Y | | N | |
| Vomiting | Y | N | Y | N | Y | N | Y | N |
| Adeno Y | .982 | .961 | .833 | .676 | .754 | .556 | .167 | .081 |

Figure 19: Bayesian Network for adeno JPD.

| Pain | Y | | | | N | | | |
|---|---|---|---|---|---|---|---|---|
| Weightloss | Y | | N | | Y | | N | |
| Vomiting | Y | N | Y | N | Y | N | Y | N |
| Adeno Y | .384 | 0 | .109 | 0 | .270 | 0 | .020 | 0 |
| N | .007 | 0 | .023 | 0 | .089 | 0 | .100 | 0 |

Figure 20: JPD of Bayesian Network for adeno JPD with evidenve $V = Y$ entered.





## Overview of inference methods [Guo and Hsu 2001]

(i) exact inference:

(a) Polytree algorithm
(b) conditioning
(c) clustering
(d) arc reversal
(e) variable elimination

(ii) approximate inference:

(a) stochastic sampling
(b) model simplification
(c) search-based
(d) loopy propagation

(iii) symbolic inference.

1. Inference in Probabilistic Networks

## 2. Variable elimination





## Aggregating products

Doing inference using the naive approach$_2$,

$$(p_E)^{\downarrow W|\emptyset} = ((\prod_{q \in Q} q_E)^{\downarrow W|\emptyset})$$

we compute a large table as product of $q_E$ and then aggregate to $W$.

Question: can we aggregate the factors and then multiply the aggregates?

$$(pq)^{\downarrow W} \stackrel{?}{=} p^{\downarrow W} q^{\downarrow W}$$

In general, this equation does not hold, as

$$(pq)^{\downarrow W}(x) = \sum_{y \in \prod_{X \in \mathrm{dom}(pq) \setminus W} \mathrm{dom}(X)} p(x,y)q(x,y)$$

but

$$(p^{\downarrow W} q^{\downarrow W})(x) = (\sum_{y \in \prod_{X \in \mathrm{dom}(p) \setminus W} \mathrm{dom}(X)} p(x,y)) \cdot (\sum_{y \in \prod_{X \in \mathrm{dom}(q) \setminus W} \mathrm{dom}(X)} q(x,y))$$

But it is true for $\mathrm{dom}(p) \cap \mathrm{dom}(q) \subseteq W$, i.e., if $p$ and $q$ have no common variables except those in $W$.

**Lemma 1.** *Let $p$ and $q$ be two potentials on a subset of variables $V$. Let $W \subseteq V$ a subset of the variables.*
*If $\mathrm{dom}(p) \cap \mathrm{dom}(q) \subseteq W$ then*

$$(pq)^{\downarrow W} = p^{\downarrow W} q^{\downarrow W}$$

## Variable elimination

We can make use of this observation for simplifying $(\prod_{q \in Q} q)^{\downarrow W}$:

(i) choose a variable $v \in V \setminus W$, clearly

$$(\prod_{q \in Q} q)^{\downarrow W} = ((\prod_{q \in Q} q)^{\downarrow cv})^{\downarrow W}$$

i.e., we can eliminate variable $v$ first,

(ii) let

$$R := \{q \in Q \,|\, v \in \mathrm{dom}(q)\}$$

be all potentials which's domain contains $v$ and

$$q' := \prod_{q \in R} q, \quad q_{\text{rest}} = \prod_{q \in Q \setminus R} q$$

(iii) Then

$$\mathrm{dom}(q') \cap \mathrm{dom}(q_{\text{rest}}) \subseteq V \setminus \{v\}$$

and thus

$$(\prod_{q \in Q} q)^{\downarrow W} = (q_{\text{rest}} \cdot q'^{\downarrow cv})^{\downarrow W}$$

i.e., we replace the potentials $R$ by

$$q'^{\downarrow cv}$$

After this replacement, the variable $v$ is eliminated from the potentials $Q' := Q \setminus R \cup \{q'^{\downarrow cv}\}$.





## Variable elimination

inference-varelim($Q$ : *set of potentials*, $W$ : *set of variables*) :
**while** $\bigcup_{q \in Q} \mathrm{dom}(q) \setminus W \neq \emptyset$ **do**
    choose $v \in \bigcup_{q \in Q} \mathrm{dom}(q) \setminus W$ arbitrarily
    $Q :=$ *eliminate-variable*$(Q, v)$
**od**
**return** $(\prod_{q \in Q} q)^{|\emptyset}$

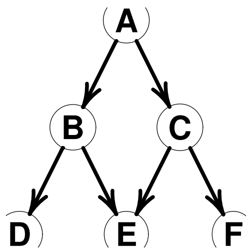
eliminate-variable($Q$ : *set of potentials*, $v$ : *variable*) :
$R := \{q \in Q \,|\, v \in \mathrm{dom}(q)\}$
$q' := (\prod_{q \in R} q)^{\downarrow cv}$
**return** $Q \setminus R \cup \{q'\}$

Also known as **bucket elimination**.

Useful if the set $W$ of variables to infer separately is small.

## example

**Example 2.** Let $(G, (p_v)_{v \in V})$ be the following Bayesian network




The conditional probabilities are

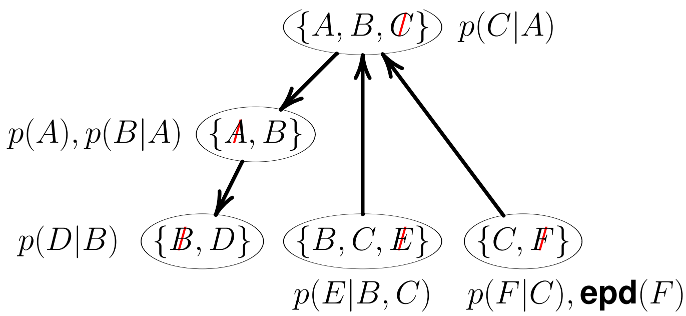
$$Q := \{p(A), p(B|A), p(C|A), p(D|B), \\ p(E|B,C), p(F|C)\}$$

We want to compute the marginal $p(D)$ given evidence on $F$. Thus we add $\text{epd}(F)$ to $Q$.

For the elimination sequence

$$F, E, C, A, B$$

the following steps have to be performed:





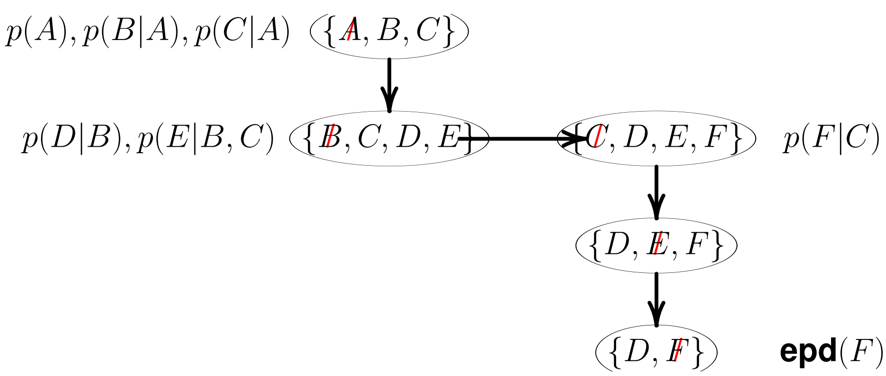



## example

For the elimination sequence

$$A, B, C, E, F$$

the following steps have to be performed:



# References

[Jen01] Finn V. Jensen. *Bayesian networks and decision graphs*. Springer, New York, 2001.