Review

- A probability space (Ω, P) consists of a sample space Ω and a probability measure P
- A set of outcomes $A \subseteq \Omega$ is called an event
- P should obey Kolmogorov's axioms:
  1. $P(A) \geq 0$ for all events A
  2. $P(\Omega) = 1$
  3. $P(A \cup B) = P(A) + P(B)$ for disjoint events A and B

Review

- A random variable X is a function from the sample space Ω to the domain $\text{dom}(X)$ of X
- A random variable has an associated density $P_X : \text{dom}(X) \to \mathbb{R}$; e.g. $P_{\text{flu}}(\text{yes}) = 0.1$
- From a joint density we can compute marginal and conditional densities
- Marginalization: $P(u) = \sum_v P(u, v)$
- The conditional probability of a A given B is

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

Review

- Random variable X is independent of random variable Y, written as $X \perp\!\!\!\perp Y$, if for all x and y

$$P(x \mid y) = P(x)$$

- Random variable X is conditionally independent of random variable Y given random variable Z, written as $X \perp\!\!\!\perp Y \mid Z$, if

$$P(x \mid y, z) = P(x \mid z)$$

whenever $P(y, z) > 0$. That is, knowledge of Y doesn't affect your belief in the value of X, given a value of Z.

Probabilistic inference

Example:

- Assume we have a joint density over the following five variables:
  - Temperature: $\mathsf{temp} \in \{\mathsf{high}, \mathsf{low}\}$
  - Fever: $\mathsf{fe} \in \{\mathsf{y}, \mathsf{n}\}$
  - Myalgia: $\mathsf{my} \in \{\mathsf{y}, \mathsf{n}\}$
  - Flu: $\mathsf{fl} \in \{\mathsf{y}, \mathsf{n}\}$
  - Pneumonia: $\mathsf{pn} \in \{\mathsf{y}, \mathsf{n}\}$

Probabilistic inference amounts to computing one or more (conditional) densities given (possibly empty) observations.

Conditioning and marginalization

- How to compute $P(\text{pn} \mid \text{temp=high})$ from the joint density $P(\text{temp, fe, my, fl, pn})$?
- Marginalization gives us:

$$P(\text{pn} \mid \text{temp=high}) \;=\; \sum_{\text{fe}} \sum_{\text{my}} \sum_{\text{fl}} P(\text{fe, my, fl, pn} \mid \text{temp=high})$$

- Conditioning gives us:

$$P(\text{fe, my, fl, pn} \mid \text{temp=high}) = \frac{P(\text{temp=high, fe, my, fl, pn})}{P(\text{temp=high})}$$

- Therefore:

$$P(\text{pn} \mid \text{temp=high}) = \frac{1}{Z} \sum_{\text{fe}} \sum_{\text{my}} \sum_{\text{fl}} P(\text{temp=high, fe, my, fl, pn})$$

Inference problem

$$P(\text{pn} \mid \text{temp=high}) = \frac{1}{Z} \sum_{\text{fe}} \sum_{\text{my}} \sum_{\text{fl}} P(\text{temp=high, fe, my, fl, pn})$$

- We don't need to compute Z. We just compute

$$\sum_{\text{fe}} \sum_{\text{my}} \sum_{\text{fl}} P(\text{temp=high, fe, my, fl, pn})$$

  and renormalize.
- We do need to compute the sums, which becomes expensive very fast (nested `for` loops)!

Representation problem

- In order to specify the joint density $P(\text{temp}, \text{fe}, \text{my}, \text{fl}, \text{pn})$ we need to estimate 31 probabilities given binary variables
- Probabilities can be estimated by means of knowledge engineering or by parameter learning
- This doesn't solve the problem
  - How does an expert estimate $P(\text{temp=low}, \text{fe=y}, \text{my=n}, \text{fl=y}, \text{pn=y})$?
  - Parameter learning requires huge databases containing multiple instances of each configuration
- Solution: conditional independence!

Chain rule revisited

- The chain rule allows us to write:

$$P(\text{temp}, \text{fe}, \text{my}, \text{fl}, \text{pn})$$
$$= P(\text{temp} \mid \text{fe}, \text{my}, \text{fl}, \text{pn}) P(\text{fe} \mid \text{my}, \text{fl}, \text{pn}) P(\text{my} \mid \text{fl}, \text{pn}) P(\text{fl} \mid \text{pn}) P(\text{pn})$$

- This requires $16 + 8 + 4 + 2 + 1 = 31$ probabilities
- We now make the following (conditional) independence assumptions:
  - $\text{fl} \perp\!\!\!\perp \text{pn}$
  - $\text{my} \perp\!\!\!\perp \{\text{temp}, \text{fe}, \text{pn}\} \mid \text{fl}$
  - $\text{temp} \perp\!\!\!\perp \{\text{my}, \text{fl}, \text{pn}\} \mid \text{fe}$
  - $\text{fe} \perp\!\!\!\perp \{\text{my}\} \mid \{\text{fl}, \text{pn}\}$

Chain rule revisited

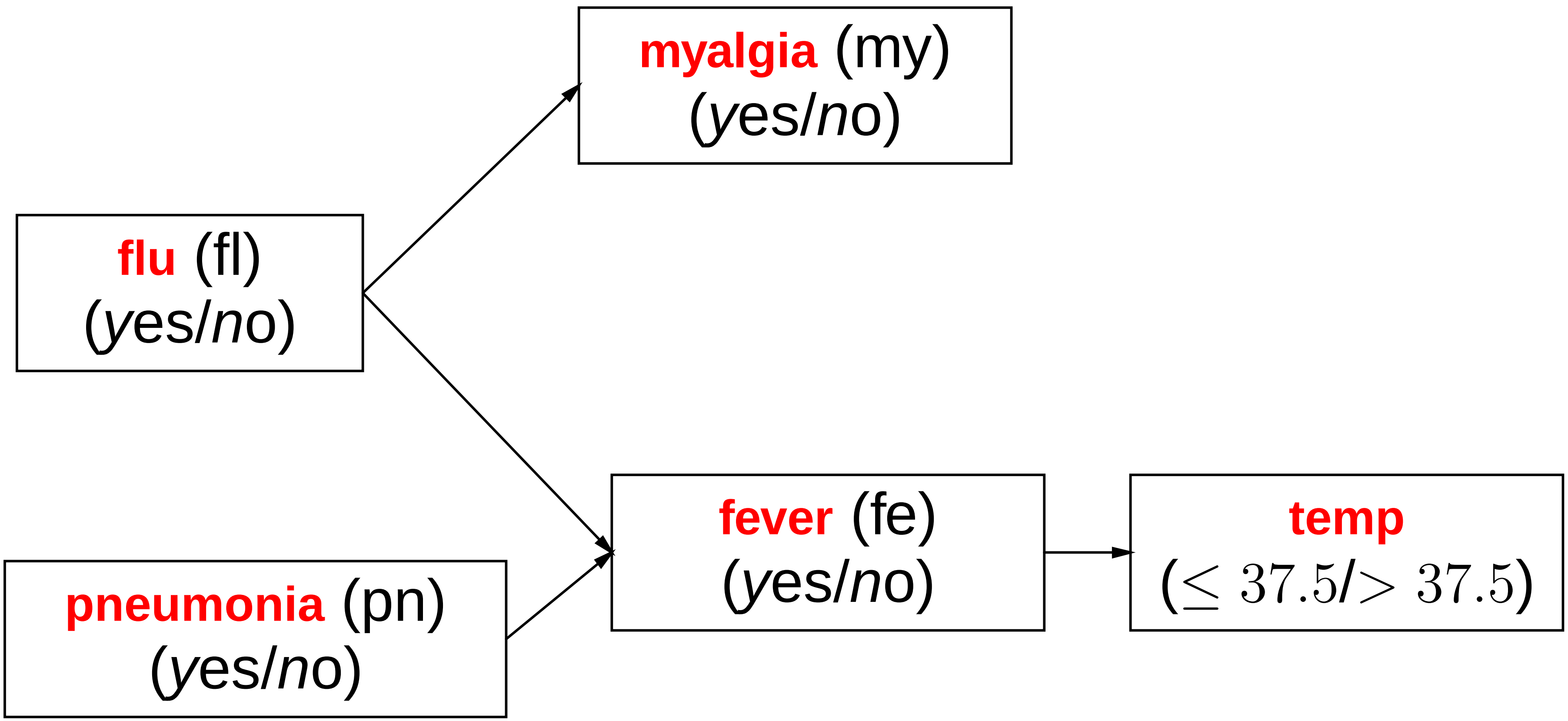
- By definition of conditional independence:

$$P(\text{temp}, \text{fe}, \text{my}, \text{fl}, \text{pn})$$
$$= P(\text{temp} \mid \text{fe}, \text{my}, \text{fl}, \text{pn}) P(\text{fe} \mid \text{my}, \text{fl}, \text{pn}) P(\text{my} \mid \text{fl}, \text{pn}) P(\text{fl} \mid \text{pn}) P(\text{pn})$$
$$= P(\text{temp} \mid \text{fe}) P(\text{fe} \mid \text{fl}, \text{pn}) P(\text{my} \mid \text{fl}) P(\text{fl}) P(\text{pn})$$

- This requires just $2 + 4 + 2 + 1 + 1 = 10$ instead of 31 probabilities!
- CI reduces the number of required probabilities and facilitates specification of the remaining probabilities:
  - $P(\text{my} \mid \text{fl})$: the probability of myalgia given that someone has flu
  - $P(\text{pn})$: the prior probability that a random person suffers from pneumonia

Bayesian networks

A Bayesian (belief) network is a convenient graphical representation of the independence structure of a joint density



Bayesian networks

- A Bayesian network consists of:
  - a directed acyclic graph with nodes labeled with random variables
  - a set of (conditional) densities for each variable given its parents
- Let $G = (V, E)$ be a directed acyclic graph (or DAG)
- Let P denote a joint density over a set of random variables $X = (X_v)_{v \in V}$ indexed by V.
- (G, P) is a Bayesian network if the joint probability distribution $P(X)$ factorizes according to

$$P(x) = \prod_{v \in V} P(x_v \mid x_{\mathsf{pa}(v)})$$

Bayesian networks

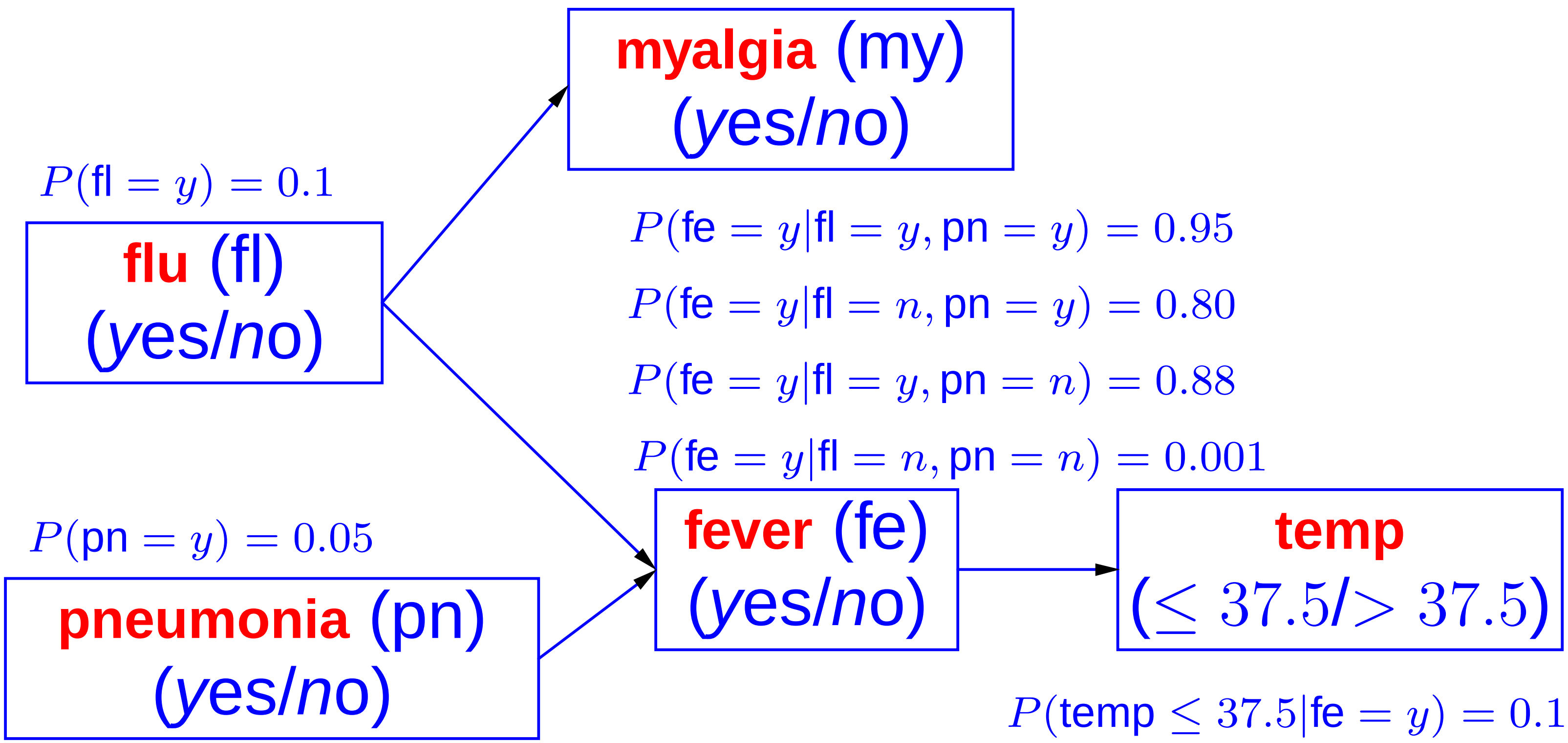
- Bayesian networks may consist of discrete or continuous random variables, or both
- A Bayesian network is a particular kind of probabilistic graphical model (dynamic Bayesian networks, Markov networks, chain graphs, factor graphs)
- Most statistical methods can be represented as graphical models (naive Bayes, factor analysis, gaussian mixture models, hidden Markov models, ...)
- The graphical structure is not just a convenient representation but also the basis for inference algorithms

Specification of probabilities

$$P(\text{temp}, \text{fe}, \text{my}, \text{fl}, \text{pn})$$

$P(\text{my} = y | \text{fl} = y) = 0.96$

$P(\text{my} = y | \text{fl} = n) = 0.20$



$P(\text{fl} = y) = 0.1$

$P(\text{fe} = y | \text{fl} = y, \text{pn} = y) = 0.95$

$P(\text{fe} = y | \text{fl} = n, \text{pn} = y) = 0.80$

$P(\text{fe} = y | \text{fl} = y, \text{pn} = n) = 0.88$

$P(\text{fe} = y | \text{fl} = n, \text{pn} = n) = 0.001$

$P(\text{pn} = y) = 0.05$

$P(\text{temp} \leq 37.5 | \text{fe} = y) = 0.1$

$P(\text{temp} \leq 37.5 | \text{fe} = n) = 0.99$

Bayesian network construction

To represent a domain in a Bayesian network, you need to consider:

- What are the relevant variables?
  - What would you like to find out?
  - What will you observe?
  - What other features make the model simpler (e.g. causal independence)?
- What values should the variables take (discrete, continuous)?
- What is the relationship between them? This should be expressed in terms of local influences.

Bayesian network construction

- A BN can be formally constructed as follows:
  1. choose an ordering of the variables;
  2. apply the chain rule; and
  3. use conditional independence assumptions to prune parents.
- The final structure depends on the variable ordering
- Another way to construct the network is to:
  - choose the parents of each node
  - heuristic: start at the variable(s) of interest and expand outwards
  - ensure that the resulting graph is acyclic.

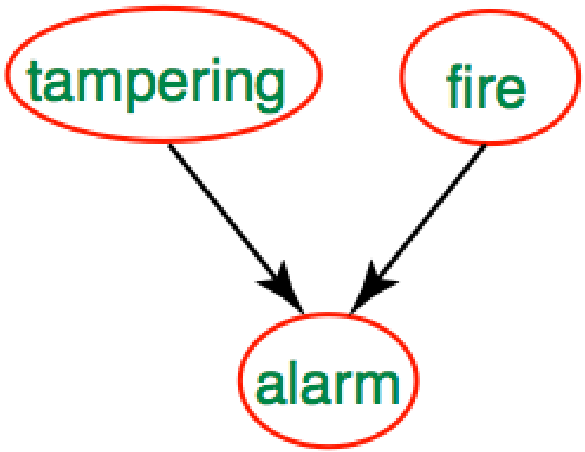
Causal networks

- Although BNs often model causal knowledge, they are not causal models!

$$X \rightarrow Y \equiv X \leftarrow Y \text{ since } P(X \mid Y)P(Y) = P(Y \mid X)P(X)$$

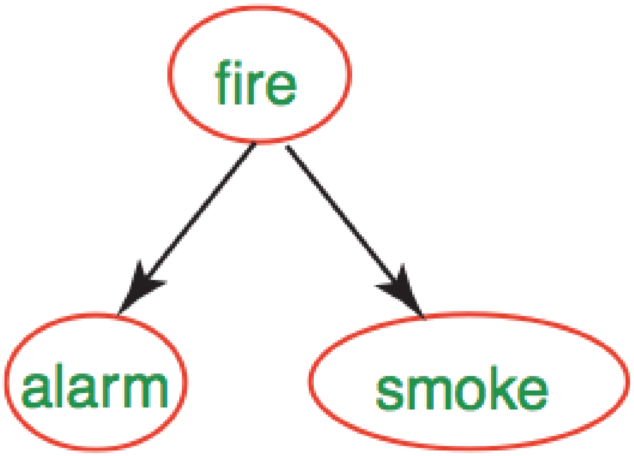
- A causal network is a Bayesian network with an explicit requirement that the relationships are causal:
  - If X is actively caused to be in a given state x then this is an action written as $do(X = x)$
  - If $do(X = x)$ then the joint density changes to the one of the network obtained by cutting the links from X's parents to X, and setting X to the caused value x
  - Using these semantics, one can predict the impact of external interventions from data obtained prior to intervention.

Common descendants



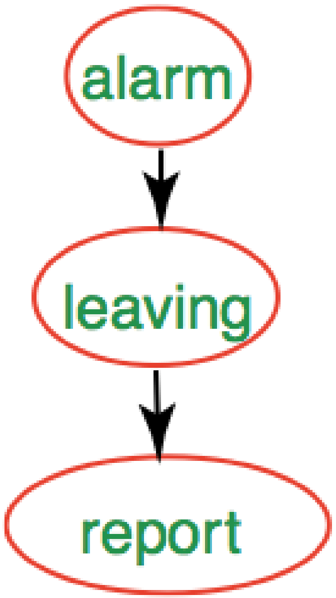
- tampering and fire are independent
- tampering and fire are dependent given alarm
- tampering can explain away fire

Common ancestors



- alarm and smoke are dependent
- alarm and spoke are independent given fire
- fire can explain alarm and smoke; learning about one can affect the other by changing your belief in fire

Chain



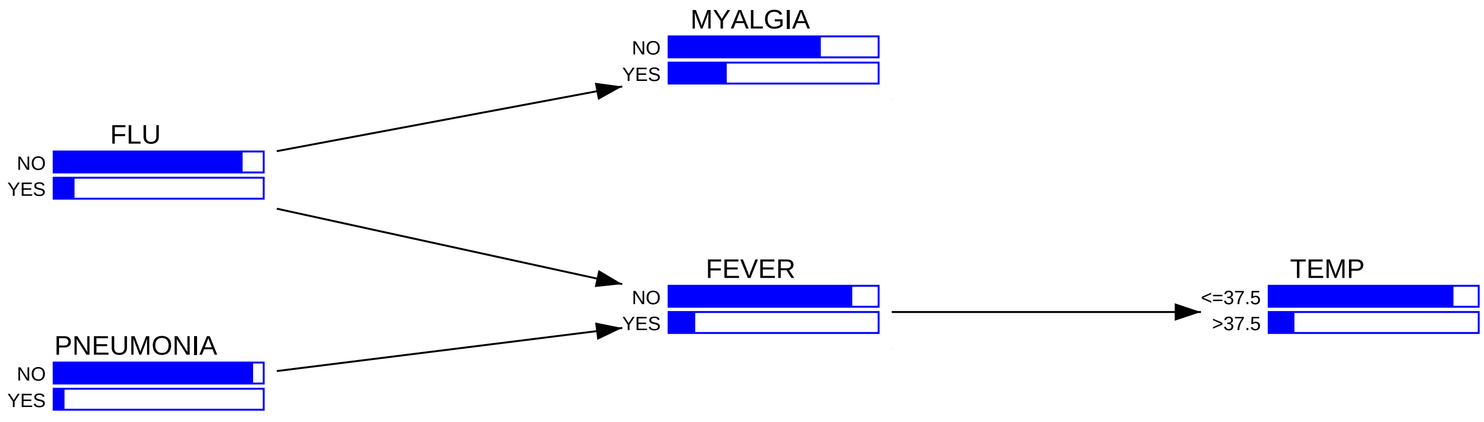
- alarm and report are dependent
- alarm and report are independent given leaving
- the only way alarm affects report is by affecting leaving

Testing for conditional independence

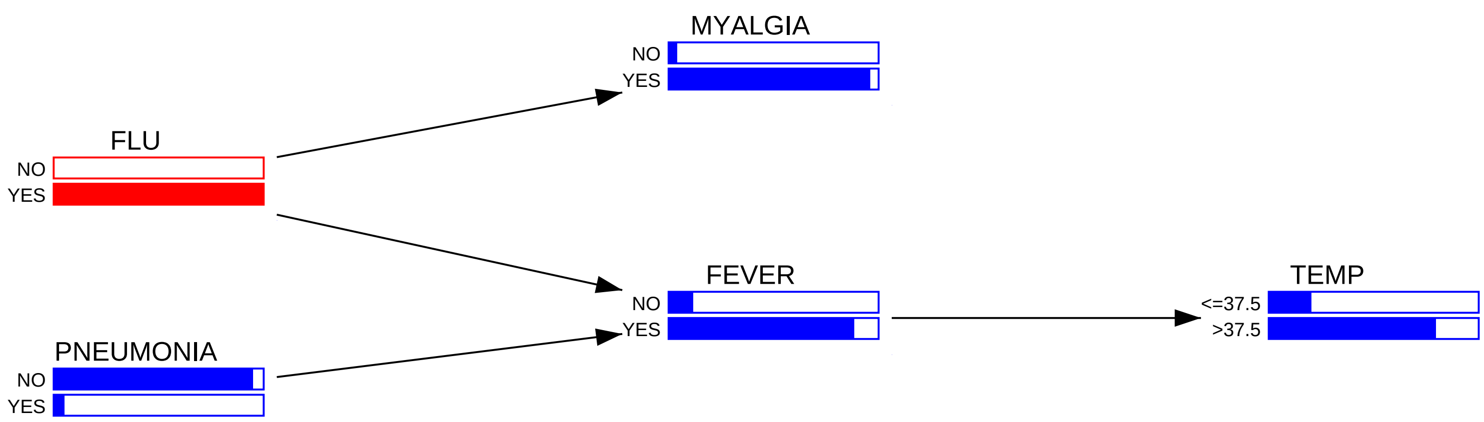
- Bayesian networks encode the independence properties of a joint density.
- If we enter evidence in a BN, the result is a conditional density that can have different independence properties.
- We can determine if a conditional independence $X \perp\!\!\!\perp Y \mid \{Z_1, \ldots, Z_k\}$ holds through the concept of d-separation
- X and Y are d-separated if there is no active path between them.

Inference: evidence propagation

- Nothing known:

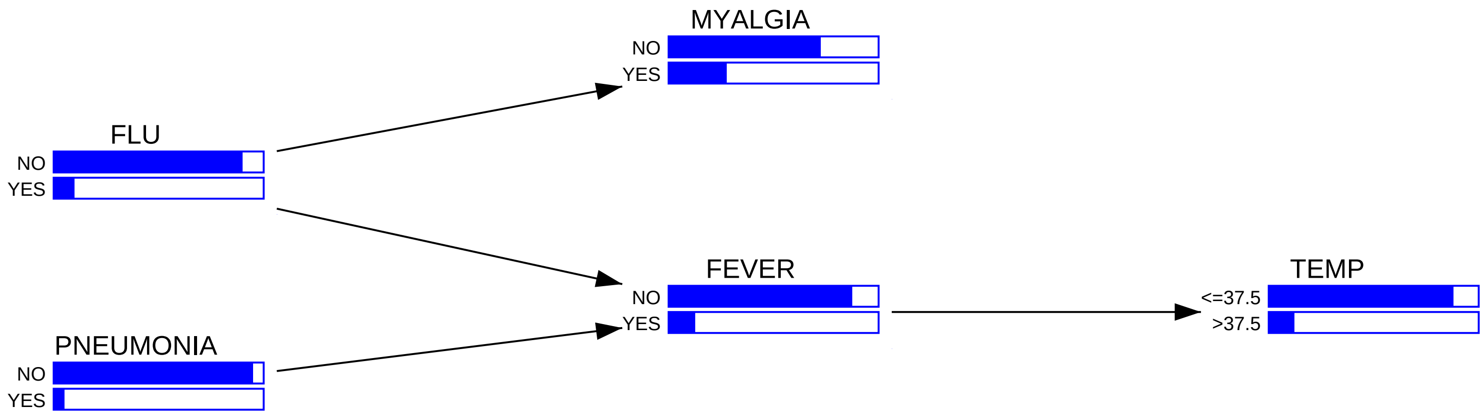


- Which symptoms belong to flu?

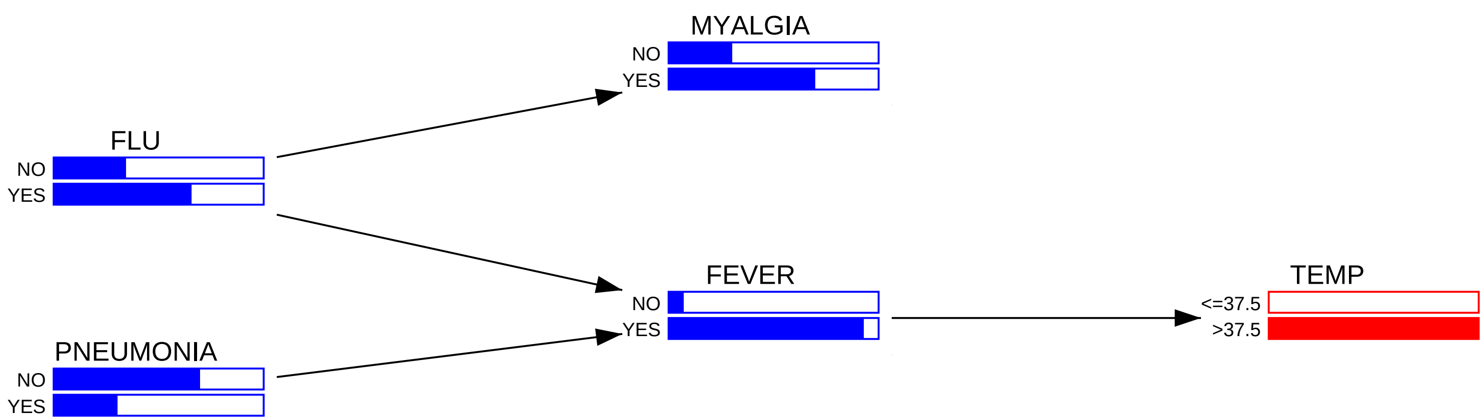


Inference: evidence propagation

- Nothing known:



- Temperature > 37.5 grades Celcius:



Efficient inference

- Conditional independence assumptions not only solve the representation problem but also make inference easier
- By plugging in the factorized density we obtain:

$$P(\text{pn} \mid \text{temp=high})$$

$$\propto \sum_{\text{fe}} \sum_{\text{my}} \sum_{\text{fl}} P(\text{temp=high}, \text{fe}, \text{my}, \text{fl}, \text{pn})$$

$$= \sum_{\text{fe}} \sum_{\text{my}} \sum_{\text{fl}} P(\text{temp=high} \mid \text{fe}) P(\text{fe} \mid \text{fl}, \text{pn}) P(\text{my} \mid \text{fl}) P(\text{fl}) P(\text{pn})$$

- Inference reduces to computing sums of products. An efficient way to do this is using variable elimination

Variable elimination

- How can we compute $ab + ac$ efficiently?

Variable elimination

- How can we compute $ab + ac$ efficiently?
- Distribute out a giving $a(b + c) \to$ 2 instead of 3 elementary operations

Variable elimination

- How can we compute $ab + ac$ efficiently?
- Distribute out a giving $a(b + c) \rightarrow$ 2 instead of 3 elementary operations
- Represent densities as factors or potential functions $P(\mathbf{x}, \mathbf{y} \mid \mathbf{z}) = f(\mathbf{x}, \mathbf{y}, \mathbf{z})$ with target variables $\mathbf{x}$, nuisance variables $\mathbf{y}$ and evidence $\mathbf{z}$

Variable elimination

- How can we compute $ab + ac$ efficiently?
- Distribute out a giving $a(b + c) \to$ 2 instead of 3 elementary operations
- Represent densities as factors or potential functions $P(\mathbf{x}, \mathbf{y} \mid \mathbf{z}) = f(\mathbf{x}, \mathbf{y}, \mathbf{z})$ with target variables $\mathbf{x}$, nuisance variables $\mathbf{y}$ and evidence $\mathbf{z}$
- How to efficiently compute

$$\sum_{y_1} \cdots \sum_{y_n} f(\mathbf{x}_1, \mathbf{y}_1, \mathbf{z}_1) \cdots f(\mathbf{x}_k, \mathbf{y}_k, \mathbf{z}_k)$$

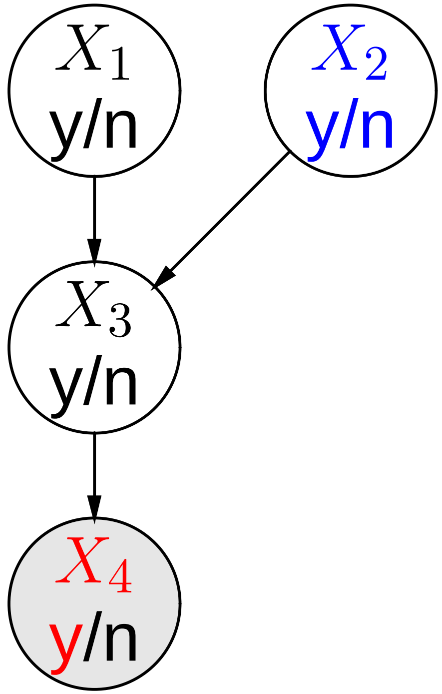
Variable elimination

- How can we compute $ab + ac$ efficiently?
- Distribute out a giving $a(b + c) \rightarrow$ 2 instead of 3 elementary operations
- Represent densities as factors or potential functions $P(\mathbf{x}, \mathbf{y} \mid \mathbf{z}) = f(\mathbf{x}, \mathbf{y}, \mathbf{z})$ with target variables $\mathbf{x}$, nuisance variables $\mathbf{y}$ and evidence $\mathbf{z}$
- How to efficiently compute

$$\sum_{y_1} \cdots \sum_{y_n} f(\mathbf{x}_1, \mathbf{y}_1, \mathbf{z}_1) \cdots f(\mathbf{x}_k, \mathbf{y}_k, \mathbf{z}_k)$$

  - Choose a variable Y_i to eliminate
  - Push in its sum as far as possible: factors $f(\mathbf{x}_j, \mathbf{y}_j, \mathbf{z}_j)$ left of the sum should not contain Y_i
  - Compute the sum, which gives a new factor

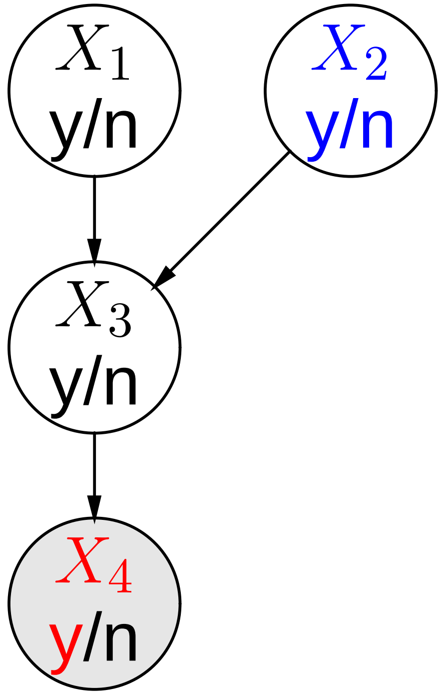
Example



$$P(x_4 \mid x_3) = 0.4$$
$$P(x_4 \mid \neg x_3) = 0.1$$
$$P(x_3 \mid x_1, x_2) = 0.3$$
$$P(x_3 \mid \neg x_1, x_2) = 0.5$$
$$P(x_3 \mid x_1, \neg x_2) = 0.7$$
$$P(x_3 \mid \neg x_1, \neg x_2) = 0.9$$
$$P(x_1) = 0.6$$
$$P(x_2) = 0.2$$

$$P(X_2 \mid x_4) = \frac{P(X_2, x_4)}{P(x_4)}$$

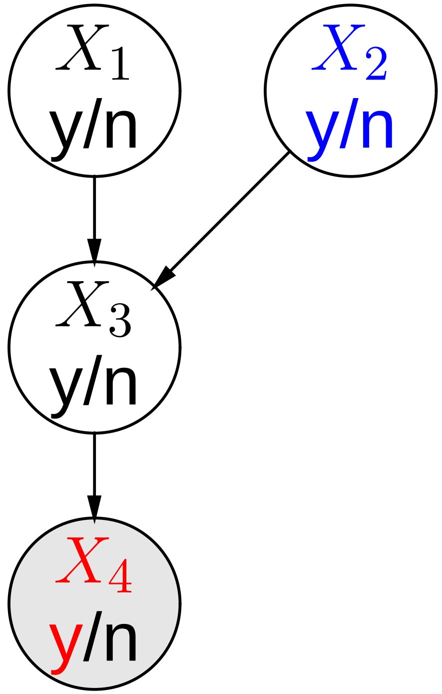
Decomposition



$$P(x_4 \mid x_3) = 0.4$$
$$P(x_4 \mid \neg x_3) = 0.1$$
$$P(x_3 \mid x_1, x_2) = 0.3$$
$$P(x_3 \mid \neg x_1, x_2) = 0.5$$
$$P(x_3 \mid x_1, \neg x_2) = 0.7$$
$$P(x_3 \mid \neg x_1, \neg x_2) = 0.9$$
$$P(x_1) = 0.6$$
$$P(x_2) = 0.2$$

$$P(X_2, x_4) = \sum_{X_3} \sum_{X_1} P(x_4|X_3)P(X_3|X_1, X_2)P(X_1)P(X_2)$$

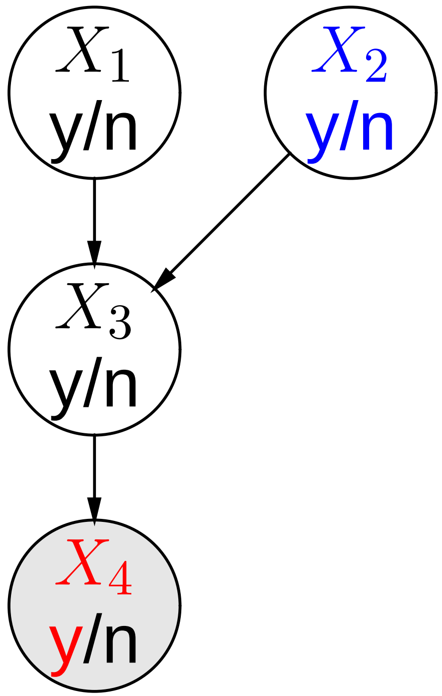
Factor representation



$$P(x_4 \mid x_3) = 0.4$$
$$P(x_4 \mid \neg x_3) = 0.1$$
$$P(x_3 \mid x_1, x_2) = 0.3$$
$$P(x_3 \mid \neg x_1, x_2) = 0.5$$
$$P(x_3 \mid x_1, \neg x_2) = 0.7$$
$$P(x_3 \mid \neg x_1, \neg x_2) = 0.9$$
$$P(x_1) = 0.6$$
$$P(x_2) = 0.2$$

$$P(X_2, x_4) = \sum_{X_3} \sum_{X_1} f_4(X_3, x_4) f_1(X_1, X_2, X_3) f_2(X_1) f_3(X_2)$$

Distribution



$$P(x_4 \mid x_3) = 0.4$$
$$P(x_4 \mid \neg x_3) = 0.1$$
$$P(x_3 \mid x_1, x_2) = 0.3$$
$$P(x_3 \mid \neg x_1, x_2) = 0.5$$
$$P(x_3 \mid x_1, \neg x_2) = 0.7$$
$$P(x_3 \mid \neg x_1, \neg x_2) = 0.9$$
$$P(x_1) = 0.6$$
$$P(x_2) = 0.2$$

$$P(X_2, x_4) = f_3(X_2) \sum_{X_3} f_4(X_3, x_4) \sum_{X_1} f_1(X_1, X_2, X_3) f_2(X_1)$$

Taking products

- Product of factors: $f_3(A,B,C) = f_1(A,B)f_2(B,C)$

f_1:

A	B	val
t	t	0.1
t	f	0.9
f	t	0.2
f	f	0.8

f_2:

B	C	val
t	t	0.3
t	f	0.7
f	t	0.6
f	f	0.4

$f_1 \times f_2$:

A	B	C	val
t	t	t	0.03
t	t	f	0.07
t	f	t	0.54
t	f	f	0.36
f	t	t	0.06
f	t	f	0.14
f	f	t	0.48
f	f	f	0.32

- Evidence $P(A = t, B)$: select corresponding rows in the tables

Taking sums

- Summing out variables: $f_3(A,C) = \sum_B f_3(A,B,C)$ (marginalization)

f_3:

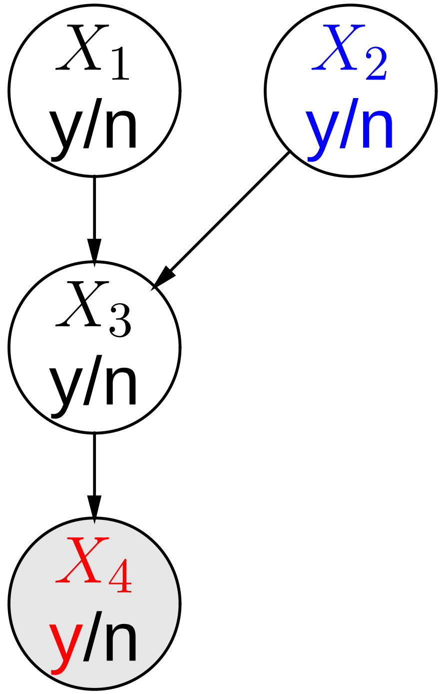
A	B	C	val
t	t	t	0.03
t	t	f	0.07
t	f	t	0.54
t	f	f	0.36
f	t	t	0.06
f	t	f	0.14
f	f	t	0.48
f	f	f	0.32

$\sum_B f_3$:

A	C	val
t	t	0.57
t	f	0.43
f	t	0.54
f	f	0.46

- Evidence $P(A = t, B, C)$: select corresponding rows in the tables

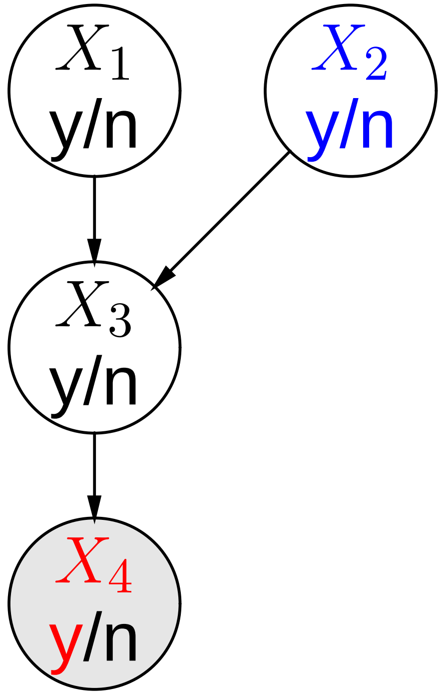
Example



$P(x_4 \mid x_3) = 0.4$
$P(x_4 \mid \neg x_3) = 0.1$
$P(x_3 \mid x_1, x_2) = 0.3$
$P(x_3 \mid \neg x_1, x_2) = 0.5$
$P(x_3 \mid x_1, \neg x_2) = 0.7$
$P(x_3 \mid \neg x_1, \neg x_2) = 0.9$
$P(x_1) = 0.6$
$P(x_2) = 0.2$

$$P(X_2, x_4) = f_3(X_2) \sum_{X_3} f_4(X_3, x_4) \sum_{X_1} f_5(X_1, X_2, X_3)$$

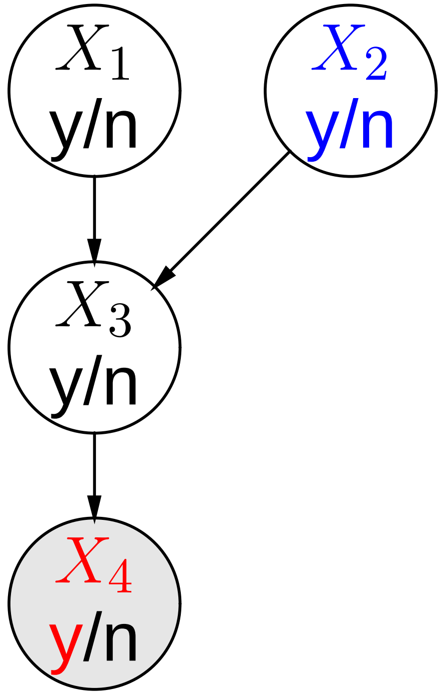
Example



$$P(x_4 \mid x_3) = 0.4$$
$$P(x_4 \mid \neg x_3) = 0.1$$
$$P(x_3 \mid x_1, x_2) = 0.3$$
$$P(x_3 \mid \neg x_1, x_2) = 0.5$$
$$P(x_3 \mid x_1, \neg x_2) = 0.7$$
$$P(x_3 \mid \neg x_1, \neg x_2) = 0.9$$
$$P(x_1) = 0.6$$
$$P(x_2) = 0.2$$

$$P(X_2, x_4) = f_3(X_2) \sum_{X_3} f_4(X_3, x_4) f_6(X_2, X_3)$$

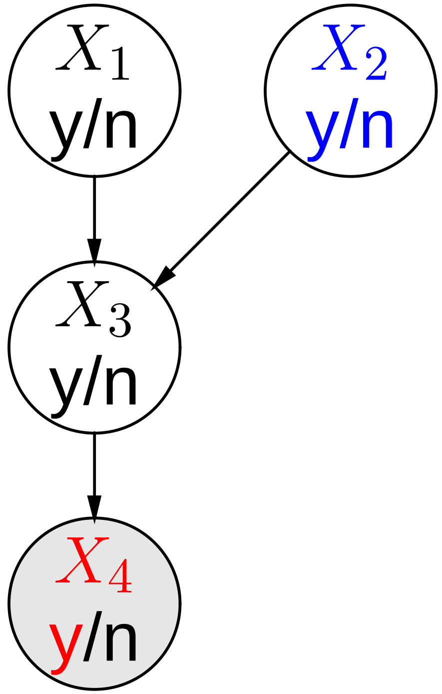
Example



$P(x_4 \mid x_3) = 0.4$
$P(x_4 \mid \neg x_3) = 0.1$
$P(x_3 \mid x_1, x_2) = 0.3$
$P(x_3 \mid \neg x_1, x_2) = 0.5$
$P(x_3 \mid x_1, \neg x_2) = 0.7$
$P(x_3 \mid \neg x_1, \neg x_2) = 0.9$
$P(x_1) = 0.6$
$P(x_2) = 0.2$

$$P(X_2, x_4) = f_3(X_2) \sum_{X_3} f_7(X_2, X_3, x_4)$$

Example

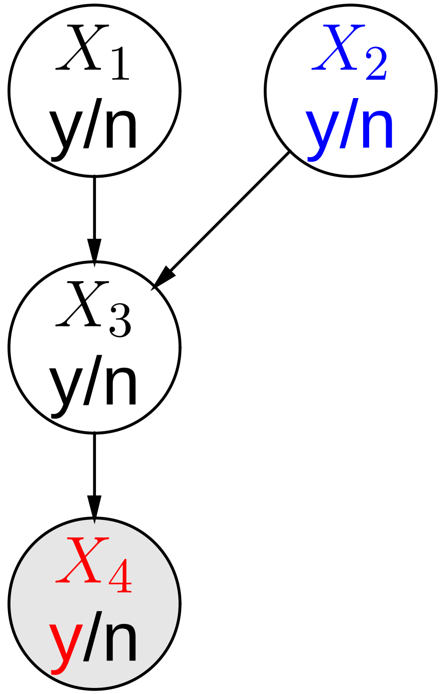


$$P(x_4 \mid x_3) = 0.4$$
$$P(x_4 \mid \neg x_3) = 0.1$$
$$P(x_3 \mid x_1, x_2) = 0.3$$
$$P(x_3 \mid \neg x_1, x_2) = 0.5$$
$$P(x_3 \mid x_1, \neg x_2) = 0.7$$
$$P(x_3 \mid \neg x_1, \neg x_2) = 0.9$$
$$P(x_1) = 0.6$$
$$P(x_2) = 0.2$$

$$P(X_2, x_4) = f_3(X_2) f_8(X_2, x_4)$$

- Compute $P(X_2 \mid x_4)$ by computing $P(x_2, x_4)$ for all $x_2 \in \text{dom}(X_2)$ and renormalizing.

Example



$$P(x_4 \mid x_3) = 0.4$$
$$P(x_4 \mid \neg x_3) = 0.1$$
$$P(x_3 \mid x_1, x_2) = 0.3$$
$$P(x_3 \mid \neg x_1, x_2) = 0.5$$
$$P(x_3 \mid x_1, \neg x_2) = 0.7$$
$$P(x_3 \mid \neg x_1, \neg x_2) = 0.9$$
$$P(x_1) = 0.6$$
$$P(x_2) = 0.2$$

How about:

$$P(x_3 \mid x_2)$$

Complete example

$$P(\text{pn}|\text{temp=high})$$

$$\propto \sum_{\text{fe}} \sum_{\text{my}} \sum_{\text{fl}} P(\text{temp=high}|\text{fe}) P(\text{fe}|\text{fl,pn}) P(\text{my}|\text{fl}) P(\text{fl}) P(\text{pn})$$

$$= \sum_{\text{fe}} \sum_{\text{fl}} P(\text{pn}) P(\text{temp=high}|\text{fe}) P(\text{fe}|\text{fl,pn}) P(\text{fl}) \sum_{\text{my}} P(\text{my}|\text{fl})$$

$$= \sum_{\text{fe}} \sum_{\text{fl}} P(\text{pn}) P(\text{temp=high}|\text{fe}) P(\text{fe}|\text{fl,pn}) P(\text{fl}) f_1(\text{fl})$$

$$= \sum_{\text{fe}} P(\text{pn}) P(\text{temp=high}|\text{fe}) \sum_{\text{fl}} P(\text{fe}|\text{fl,pn}) P(\text{fl}) f_1(\text{fl})$$

$$= \sum_{\text{fe}} P(\text{pn}) P(\text{temp=high}|\text{fe}) f_2(\text{fe,pn})$$

$$= P(\text{pn}) \sum_{\text{fe}} P(\text{temp=high}|\text{fe}) f_2(\text{fe,pn})$$

$$= P(\text{pn}) f_3(\text{pn})$$

$$= f_4(\text{pn})$$

Variable elimination summary

To compute $P(\mathbf{x}, \mathbf{y} \mid \mathbf{z})$:

- Construct a factor for each conditional probability.
- Set the observed variables $\mathbf{Z}$ to their observed values $\mathbf{z}$.
- Sum out each of the other variables $(Y_1, \ldots, Y_K)$ according to some elimination ordering.
- Multiply the remaining factors. Normalize by dividing the resulting factor $f(X)$ by $\sum_x f(x)$.

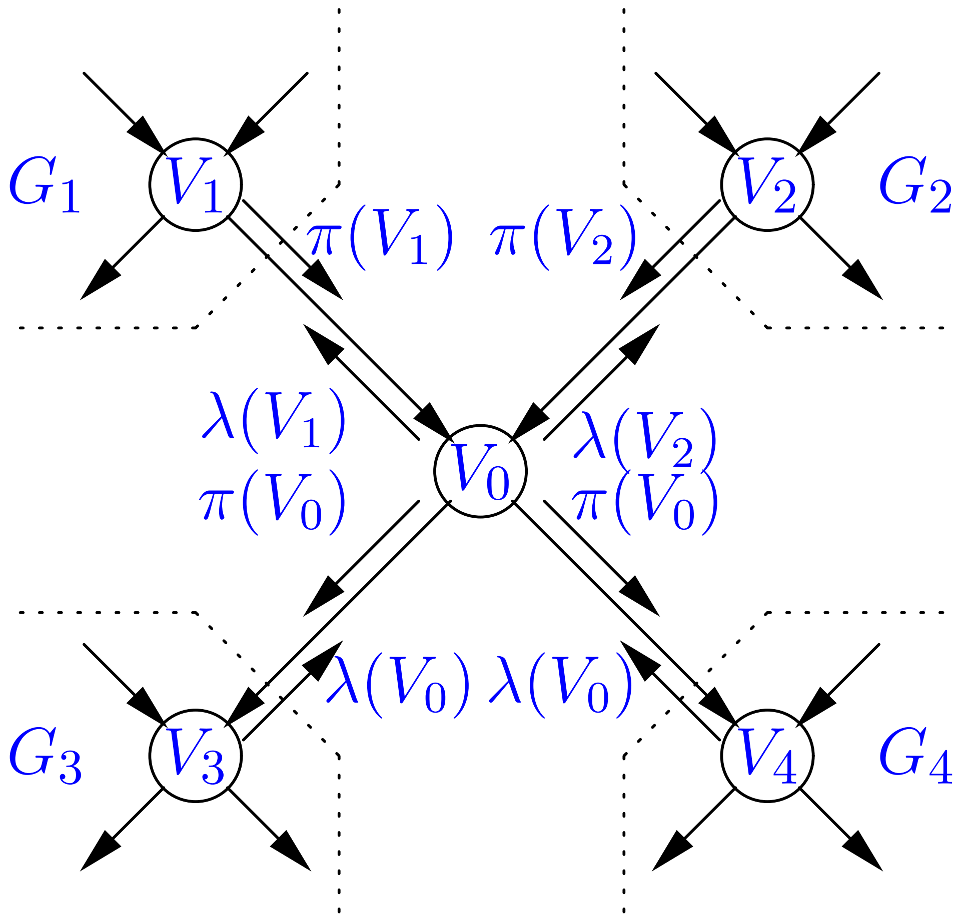
Variable elimination

Disadvantages:

- Requires global knowledge of network - can't update beliefs locally
- Is a sequential algorithm - hard to parallelize
- Doesn't show the updated belief of every variable in the network - only shows it for the hypothesis variable
- Takes exponential time and space for every query

Belief propagation

- Breakthrough algorithm due to Pearl (1988)



Belief propagation

- Object-oriented: nodes are objects that contain local information and carry out local computations
- Updating via message passing: arrows are communication channels
- Only works on polytrees: A directed graph with at most one undirected path between any two vertices.
- Loopy belief propagation: belief propagation in arbitrary graphs (approximate)
- Junction tree algorithm
  - transforms a directed acyclic graph into an undirected tree
  - belief propagation in an undirected graph

Software

- AILog http://artint.info/code/ailog/ailog_man.html
- Bayes net toolbox http://code.google.com/p/bnt/
- BayesBrain http://www.socsci.ru.nl/ marcelge/Site/Code.html
- Elvira http://www.ia.uned.es/ elvira/index-en.html
- Carmen http://www.cisiad.uned.es/carmen/acceso-codigo.html
- Genie/Smile http://genie.sis.pitt.edu/
- Infer.NET http://research.microsoft.com/en-us/um/cambridge/projects/infernet/
- Bayesbuilder http://www.snn.ru.nl/nijmegen/index.php ?option=com_content&view=article&id=89&Itemid=212
- Netica http://www.norsys.com/download.html
- Hugin http://www.hugin.com/productsservices/demo/hugin-lite

Probabilistic interpretation CF calculus?

- Rule-based uncertainty: $e \rightarrow h_x$
  - propagation from antecedent e to conclusion h (f_{prop})
  - combination of $\wedge$ and $\vee$ evidence in e ($f_{\wedge}$ and $f_{\vee}$)
  - co-concluding rules (f_{co}):

$$e_1 \rightarrow h_x$$
$$e_2 \rightarrow h_y$$

- Bayesian networks: joint probability distribution $P(X_1, \ldots, X_n)$ with marginalisation $\sum_Y P(Y, Z)$ and conditioning $P(Y \mid Z)$

Propagation

- f_{prop} (propagation):

$$e' \xrightarrow{\text{CF}(e,e')} e \xrightarrow{\text{CF}(h,e)} h$$

$$\text{CF}(h,e') = \text{CF}(h,e) \cdot \max\{0, \text{CF}(e,e')\}$$

- corresponding Bayesian network (with $P(e')$ extra):

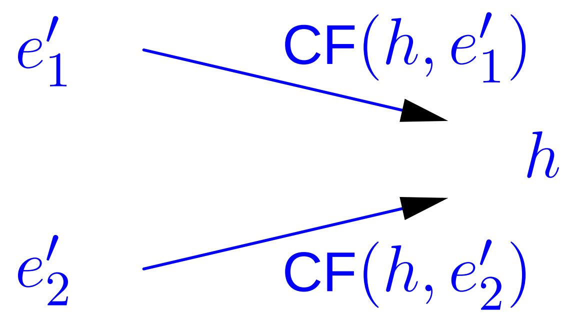
$$E' \xrightarrow{P(E \mid E')} E \xrightarrow{P(H \mid E)} H$$

$$P(h \mid e') = P(h \mid e)P(e \mid e') + P(h \mid \neg e)P(\neg e \mid e')$$

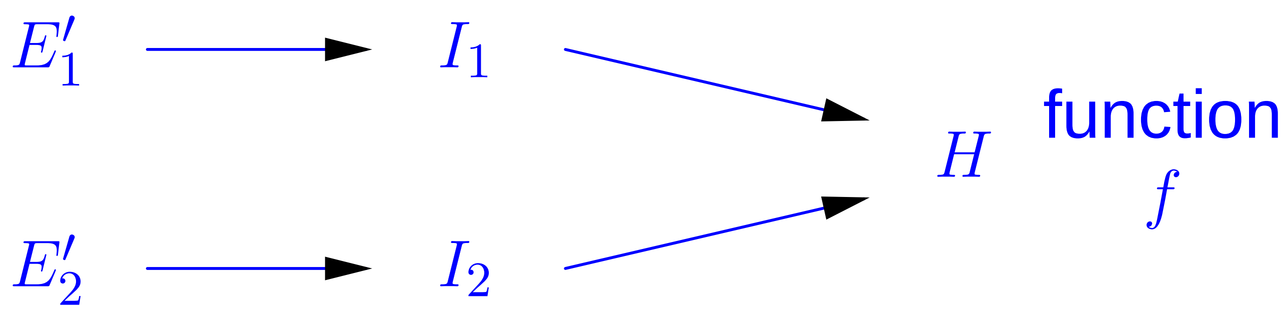
$\Rightarrow P(h \mid \neg e) = 0$ (assumption of CF-model)

Co-concluding

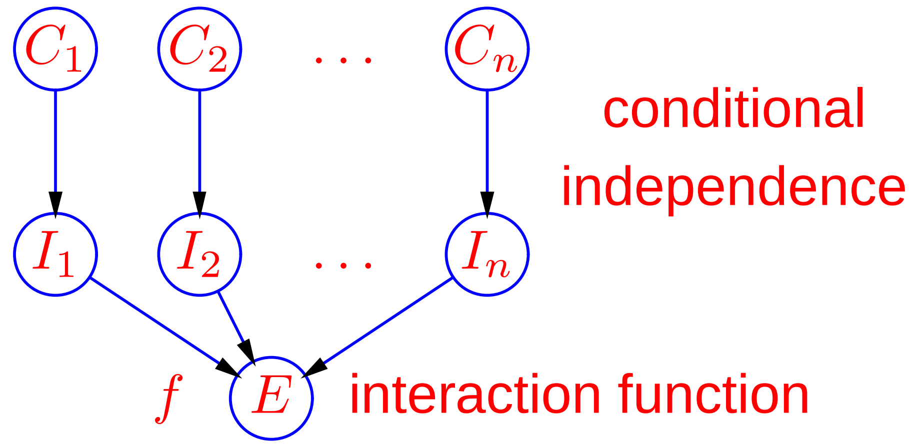
- f_{co} (co-concluding):



- idea: see this as uncertain deterministic interaction $\Rightarrow$ causal independence model



Causal Independence

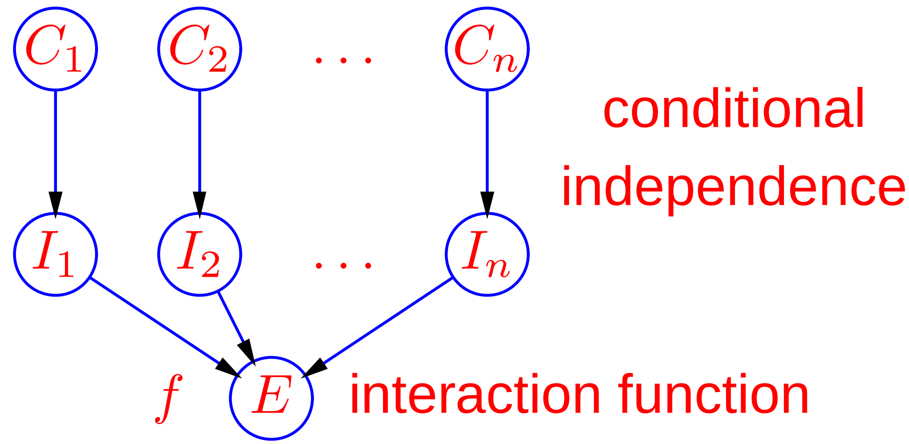


$$P(e \mid C_1, \ldots, C_n) = \sum_{I_1,\ldots,I_n} P(e \mid I_1, \ldots, I_n) \prod_{k=1}^{n} P(I_k \mid C_k)$$

$$= \sum_{f(I_1,\ldots,I_n)=e} \prod_{k=1}^{n} P(I_k \mid C_k)$$

Boolean functions: $P(E \mid I_1, \ldots, I_n) \in \{0, 1\}$ with $f(I_1, \ldots, I_n) = 1$ if $P(e \mid I_1, \ldots, I_n) = 1$

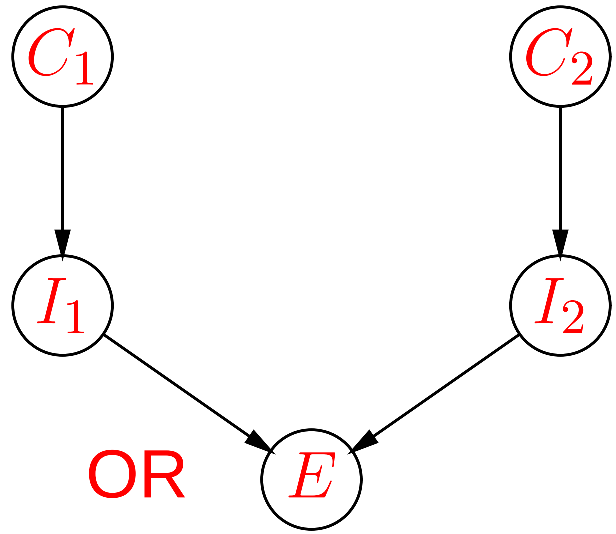
Causal Independence



$$P(e \mid C_1, \ldots, C_n) \;=\; \sum_{f(I_1,\ldots,I_n)=e} \prod_{k=1}^{n} P(I_k \mid C_k)$$

- Requires specification of one Boolean function and just n probabilities (assuming $P(i_k \mid \neg c_k) = 0$)
- Compare with 2^n probabilities for arbitrary $P(e \mid C_1, \ldots, C_n)$
- Simplifies BN construction/facilitates inference

Example: noisy OR



- Interactions between 'causes': logical OR
- Meaning: presence of the intermediate causes I_k produces effect e (i.e. $E = true$)

$$
\begin{aligned}
P(e|C_1, C_2) &= \sum_{I_1 \vee I_2 = e} \prod_{k=1,2} P(I_k \mid C_k) \\
&= P(i_1|C_1)P(i_2|C_2) + P(i_1|C_1)P(\neg i_2|C_2) \\
&\quad + P(i_2|C_2)P(\neg i_1|C_1)
\end{aligned}
$$

Noisy OR and f_{co}

- causal independence with logical OR (noisy OR):

$$
\begin{aligned}
P(e|C_1, C_2) &= P(i_1|C_1)P(i_2|C_2) + P(i_1|C_1)P(\neg i_2|C_2) \\
&\quad + P(i_2|C_2)P(\neg i_1|C_1) \\
&= P(i_1|C_1)(P(i_2|C_2) + P(\neg i_2|C_2)) \\
&\quad + P(i_2|C_2)P(\neg i_1|C_1) \\
&= P(i_1|C_1) + P(i_2|C_2)(1 - P(i_1|C_1))
\end{aligned}
$$

- f_{co}:

$$
\text{CF}(h, e'_1 \textbf{ co } e') = \text{CF}(h, e'_1) + \text{CF}(h, e'_2)(1 - \text{CF}(h, e'_1))
$$

for $\text{CF}(h, e'_1) \in [0, 1]$ and $\text{CF}(h, e'_2) \in [0, 1]$

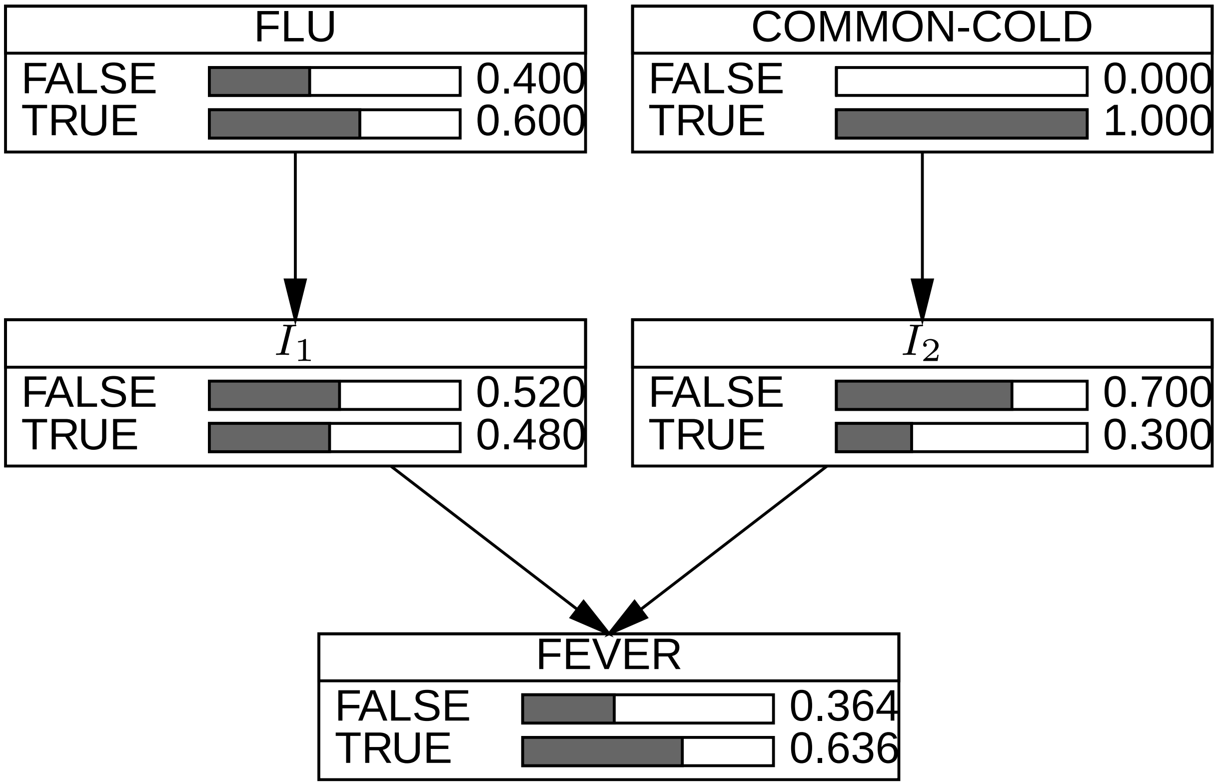
Example

- The consequences of ‘flu’ and ‘common cold’ on ‘fever’ are modelled by the variables I_1 and I_2:
  - $P(i_1 \mid \textit{flu}) = 0.8$, and
  - $P(i_2 \mid \textit{common-cold}) = 0.3$
- Furthermore, $P(i_k \mid c) = 0$, $k = 1, 2$, if $c \in \{\neg\textit{flu}, \neg\textit{common-cold}\}$
- Interaction between FLU and COMMON-COLD as noisy-OR:

$$P(\textit{fever} \mid I_1, I_2) = \begin{cases} 0 & \text{if } I_1 = \textit{false} \text{ and } I_2 = \textit{false} \\ 1 & \text{otherwise} \end{cases}$$

Result

- Bayesian network:



- Fragment CF model:

$$\text{CF}(\textit{fever}, e'_1 \textbf{ co } e'_2) = \text{CF}(\textit{fever}, e'_1) + \text{CF}(\textit{fever}, e'_2)(1 - \text{CF}(\textit{fever}, e'_1))$$
$$= 0.48 + 0.3(1 - 0.48) = 0.636$$

Conclusions

- Early rule-based (logical) approach to reasoning with uncertainty seemed attractive
- However, these had severe limitations (which ones?)
- Bayesian networks and other probabilistic graphical models are the state of the art for reasoning with uncertainty
- Earlier rule-based uncertainty reasoning can be mapped (partially) to specific Bayesian network structures

Outlook

- Decision making under uncertainty (Bayesian networks augmented with decision variables)
- Probabilistic logic: “Best of all worlds” (Leibniz)