Bayesian networks

AI 2025/2026

Content

Probabilities

The probability of a proposition is equal to the sum of the probabilities of worlds in which it holds.

$$P(\phi) = \sum_{w \in \phi} P(w), \quad \phi \text{ } \mathit{proposition}$$

Example: $P(Total = 11) = P((5,6)) + P((6,5)) = 1/36 + 1/36 = 1/18$.

Let's suppose $P(doubles) = 1/4$.

Unconditional probabilities (*prior*)

Conditional probabilities

Conditional probabilities (*posterior*) $P(A|B)$ is the fraction of possible worlds where B is true and then A is also true

- probability of A, given B

$$P(a|b) = \frac{P(a \wedge b)}{P(b)}$$

Example: $P(doubles|Die_1 = 5) = \frac{P(doubles \wedge Die_1 = 5)}{P(Die_1 = 5)}$.

The product rule: $P(a \wedge b) = P(a|b)P(b)$

Probabilities

- The probability distribution of a **discrete** random variable
  $P(Weather = sunny) = 0.6$
  $P(Weather = rain) = 0.1$
  $P(Weather = cloudy) = 0.29$
  $P(Weather = snow) = 0.01$

  $\mathbf{P}(Weather) = \langle 0.6, 0.1, 0.29, 0.01 \rangle$

- Probability density function to describe the probability distribution of a **continuous** random variable
  $P(NoonTemp = x) = Uniform_{[18C,26C]}(x)$ the noon temperature is uniformly distributed btw 18C and 26C

$$P(x) = lim_{dx \to 0} P(x \leq X \leq x + dx)/dx$$

- Joint probability distribution: the probabilities of all combinations of values of vars.
  $\mathbf{P}(Weather, Cavity)$: a 4 x 2 table
  Obs: instead of 8 ecuations, we could use
  $\mathbf{P}(Weather, Cavity) = \mathbf{P}(Weather|Cavity)\mathbf{P}(Cavity)$.

  Full joint probability distribution: the joint distribution for all of the random variables
  $Cavity, Toothache, Weather \rightarrow \mathbf{P}(Cavity, Toothache, Weather)$

Example: joint probability distribution

	toothache		¬toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Boolean vars. *Toothache*, *Cavity*, *Catch*; 3 binary vars: $2^3-1 = 7$ independent parameters;

For n boolean vars, the table has the size $O(2^n)$.

$$P(cavity \vee toothache) = 0.108+0.012+0.072+0.008+0.016+0.064 = 0.28$$

Content

Marginal distribution

Marginal distribution for a var / a subset of vars.

- Marginalization: sum the probabilities for each possible value of the **other variables**.

$$\mathbf{P}(\mathbf{Y}) = \sum_{\mathbf{z} \in \mathbf{Z}} \mathbf{P}(\mathbf{Y}, \mathbf{z})$$

Example: $\mathbf{P}(Cavity) = \sum_{\mathbf{z} \in \{Catch, Toothache\}} \mathbf{P}(Cavity, \mathbf{z})$

Marginal probability $P(cavity) = 0.108 + 0.012 + 0.072 + 0.008 = 0.2$

		toothache		¬toothache
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

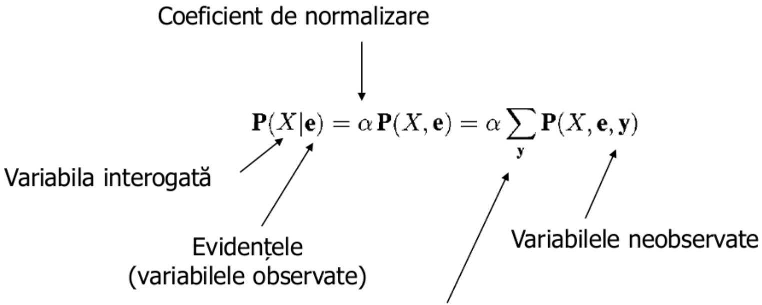
- Conditioning $\mathbf{P}(\mathbf{Y}) = \sum_{\mathbf{z}} \mathbf{P}(\mathbf{Y}|\mathbf{z})P(\mathbf{z})$

Probabilistic inference

Probabilistic inference: the computation of conditional probabilities, given certain observations.

Inference procedure: let's consider the var X, $\mathbf{E}$ the list of evidence variables, $\mathbf{e}$ the list of observed values, $\mathbf{Y}$ the remaining unobserved vars.

$$\mathbf{P}(X|\mathbf{e}) = \alpha\mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y}) \tag{1}$$



Conditional probabilities: example

	toothache		¬toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

$$P(cavity|toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)} = \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6$$

$P(\neg cavity|toothache) = 0.4$

$$\begin{aligned}\mathbf{P}(Cavity|toothache) &= \alpha \mathbf{P}(Cavity, toothache) \\ &= \alpha[\mathbf{P}(Cavity, toothache, catch) + \mathbf{P}(Cavity, toothache, \neg catch)] \\ &= \alpha[\langle 0.108, 0.016\rangle + \langle 0.012, 0.064\rangle] = \alpha\langle 0.12, 0.08\rangle = \langle 0.6, 0.4\rangle\end{aligned}$$

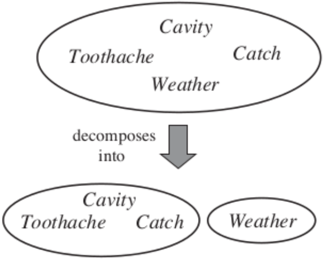
Obs: the term $1/P(toothache)$ const - normalization const for the prob. distribution $\mathbf{P}(Cavity|toothche)$

Content

Independence

The variables X and Y are independent: $\mathbf{P}(X|Y) = \mathbf{P}(X)$ or $\mathbf{P}(Y|X) = \mathbf{P}(Y)$ or $\mathbf{P}(X, Y) = \mathbf{P}(X)\mathbf{P}(Y)$.

Example: add the var *Weather*.



Less information needed to specify the joint probability distribution. The joint distribution can be *factored* into two distributions.

Independence: example

- Use the product rule $P(toothache, catch, cavity, cloudy) = P(cloudy|toothache, catch, cavity)P(toothache, catch, cavity)$
- Obs: dental problems do not influence the weather and vice versa.
  $P(cloudy|toothache, catch, cavity) = P(cloudy)$
  independence (marginal, absolute)
- Deduce
  $P(toothache, catch, cavity, cloudy) = P(cloudy)P(toothache, catch, cavity)$
  and
  $\mathbf{P}(Toothache, Catch, Cavity, Weather) = \mathbf{P}(Toothache, Catch, Cavity)\mathbf{P}(Weather)$

Content

Bayes' rule, '73

Using the product rule:

$$P(a \wedge b) = P(a|b)P(b)$$
$$P(a \wedge b) = P(b|a)P(a)$$

$\rightarrow$

$$P(b|a) = \frac{P(a|b)P(b)}{P(a)}$$

$$\mathbf{P}(Y|X) = \frac{\mathbf{P}(X|Y)\mathbf{P}(Y)}{\mathbf{P}(X)}$$

Bayes's rule

$$\mathbf{P}(H|E) = \frac{\mathbf{P}(E|H)\mathbf{P}(H)}{\mathbf{P}(E)}$$

- H hypothesis, E evidence (comes from observed data)
- $\mathbf{P}(H|E)$ a posteriori probability of the hypothesis, given the evidence (*posterior*)
- $\mathbf{P}(E|H)$ *likelihood*: the probability of observing the evidence, given the hypothesis
- $\mathbf{P}(H)$ *prior* probability (the degree of confidence in the hypothesis)

Bayes's rule

We know the evidence (the effect of an unknown cause), and we want to find the cause:

$$P(cause|effect) = \frac{P(effect|cause)P(cause)}{P(effect)}$$

Medical diagnosis: the doctor knows $P(symptoms|disease)$ and identifies the diagnostic $P(disease|symptoms)$.

Example: $P(s|m) = 0.7$ the meningitis causes a stiff neck in 70% of cases, $P(m) = 1/50000$, $P(s) = 0.01$.

$$P(m|s) = \frac{P(s|m)P(m)}{P(s)} = \frac{0.7 \cdot 1/50000}{0.01} = 0.0014$$

Bayes's rule

General form

$$\mathbf{P}(Y|X) = \alpha \mathbf{P}(X|Y)\mathbf{P}(Y)$$

α the normalization const

If we have **more than one evidence var**

- If we know the joint probability distribution $P(Cavity|toothache \wedge catch) = \alpha\langle 0.108, 0.016\rangle \approx \langle 0.871, 0.129\rangle$.
- Using Bayes' theorem:

$$\begin{aligned}\mathbf{P}(Cavity|toothache \wedge catch) &= \alpha \mathbf{P}(toothache \wedge catch|Cavity)\mathbf{P}(Cavity) \\ &= \alpha P(toothache|Cavity)P(catch|Cavity)P(Cavity)\end{aligned}$$

Conditional independence of *toothache* and *catch*, given *Cavity*

Bayes' theorem

Conditional independence of two variables X and Y given Z:
$\mathbf{P}(X, Y|Z) = \mathbf{P}(X|Z)\mathbf{P}(Y|Z)$.

Example: $\mathbf{P}(Toothache, Catch|Cavity) = \mathbf{P}(Toothache|Cavity)\mathbf{P}(Catch|Cavity)$

$$\begin{aligned}\mathbf{P}(Toothache, Catch, Cavity) &= \mathbf{P}(Toothache, Catch|Cavity)\mathbf{P}(Cavity)\\ &= \mathbf{P}(Toothache|Cavity)\mathbf{P}(Catch|Cavity)\mathbf{P}(Cavity)\end{aligned}$$

For n symptoms cond indep, given $Cavity$, the size of the representation grows linearly.

$\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause)\prod_i \mathbf{P}(Effect_i|Cause)$
($Naive\ Bayes$)

Independence and conditional independence

Example 1. Ion (A) and Maria (B) flip a coin 100 times. Everyone has a different coin.

- independent events
  $P(A|B) = P(A), P(B|A) = P(B)$
  The result of one experiment does not influence the result of the other experiment.

Example 2. Ion and Maria use the same coin

- if the coin is not correct, event A (John) may bring informations on event B (Mary)
- the events are not independent

Independence and conditional independece

Example 3. Ion and Maria live in different areas of the city and come to work by tram and car, respectively

- "John was late" and "Mary was late" may be considered independent
- if the tram drivers are on strike, then road traffic also increases; the events are conditionally independent

Content

Bayesian networks

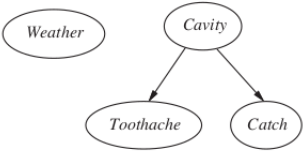
- **Probabilistic graphical models** that represent **dependencies** between variables
- Representation of information related to probabilistic events help us to perform **reasoning** efficiently
- Applications: systems for medical diagnosis, estimation of psychological characteristics from tests, environment modeling (polar bear populations), etc.

Bayesian network

Is an acyclic digraph

- each node corresponds to a random variable (event)
- an arc from X to Y (X is the parent of Y): a relationship X has a direct influence on Y
- each node X_i has a conditional probability distribution $\mathbf{P}(X_i|Parents(X_i))$ (the effect of parents on the node)

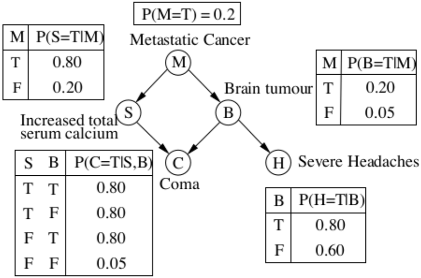
The network topology specifies conditional independence relationships:



Toothache and *Catch* are conditionally independent, given *Cavity*.

Bayesian networks: example

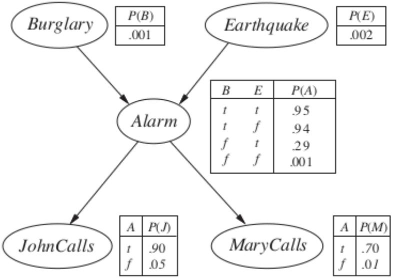
Metastatic cancer is a possible cause of brain tumors and is also an explanation for increased total serum calcium. Any of these could explain a patient going into a coma. Severe headache is also associated with brain tumors.



Obs: the conditional probability tables are given next.

Bayesian networks: example

An alarm system activates in the event of a burglary, but also in the event of an earthquake. Neighbors John and Mary call the owner at work if they hear the alarm.



Burglary and earthquakes influence the probability of triggering the alarm. The alarm influences the probability that John and Mary call.

We want to estimate the probability of a burglary based on who called.

Simple queries

A conjunction $P(X_1 = x_1 \land \ldots \land X_n = x_n)$

The assumption of the Bayesian network model is that a variable depends only on its parents:

$$P(x_1, \ldots x_n) = \prod_{i=1}^{n} P(x_i | parents(X_i)) \tag{2}$$

Example: the probability of the alarm to start when there was no burglary or earthquake and John and Mary called

$$\begin{aligned} P(j, m, a, \neg b, \neg e) &= P(j|a)P(m|a)P(a|\neg b \land \neg e)P(\neg b)P(\neg e) \\ &= 0.90 \cdot 0.70 \cdot 0.001 \cdot 0.999 \cdot 0.998 = 0.000628 \end{aligned}$$

Construction of a Bayesian network

We use the product rule to rewrite the joint probability distribution:

$$\begin{aligned} P(x_1, \ldots, x_n) &= P(x_n | x_{n-1}, \ldots, x_1) P(x_{n-1}, \ldots, x_1) \\ &= P(x_n | x_{n-1}, \ldots, x_1) P(x_{n-1} | x_{n-2}, \ldots, x_1) \ldots P(x_2 | x_1) P(x_1) \\ &= \prod_{i=1}^{n} P(x_i | x_{i-1}, \ldots, x_1) \end{aligned}$$

(**The chain rule**)

$$\mathbf{P}(X_i | X_{i-1}, \ldots, X_1) = \mathbf{P}(X_i | Parents(X_i)) \tag{3}$$

$\forall X_i$ network variable, provided that $Parents(X_i) \subseteq \{X_{i-1}, \ldots, X_1\}$.

The Bayesian network is a correct representation if each node is conditionally independent on its predecessors in the ordering, given its parents.

Construction of a Bayesian network

- Identify the set of vars $\{X_1, \ldots, X_n\}$. Sort the variables s.t. causes precede effects.
- For $i = 1, \ldots n$
  - choose from $X_1, \ldots, X_{i-1}$ a minimal set of parents s.t. equation (3) is satisfied
  - for each parent, insert an arc from it to X_i
  - add the conditional probability table $\mathbf{P}(X_i | Parents(X_i))$

Example:
$P(MaryCalls | JohnCalls, Alarm, Earthquake, Burglary) = P(MaryCalls | Alarm)$, so *Alarm* is the only parent of *MaryCalls*

Topological sort

- Topological sort of a digraph is a linear ordering of the nodes s.t. $\forall A \rightarrow B$, A appears before B.
- For a Bayesian network, topological sort ensures that parents will appear before children.
- If the graph is directeda and acyclic, there is at least one solution; if there are cycles, topological sorting is not possible.

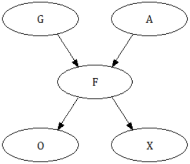
Topological sort: Kahn's algorithm

```
L ← Empty list that will contain the sorted elements
S ← Set of all nodes with no incoming edge

while S is not empty do
    remove a node n from S
    add n to L
    for each node m with an edge e from n to m do
        remove edge e from the graph
        if m has no other incoming edges then
            insert m into S

if graph has edges then
    return error   (graph has at least one cycle)
else
    return L   (a topologically sorted order)
```

Time complexity: $O(n+m)$, n vertices, m arcs



1. $L = \emptyset, S = \{G, A\}$
2. $L = \{G\}, S = \{A\}$
3. remove (GF)
   F cannot be added to S: $\exists(AF)$
4. $L = \{G, A\}, S = \emptyset$
5. remove (AF), $S = \{F\}$
6. $L = \{G, A, F\}, S = \emptyset$
7. remove (FO), $S = \{O\}$, ...
   $\rightarrow L = \{G, A, F, O, X\}$

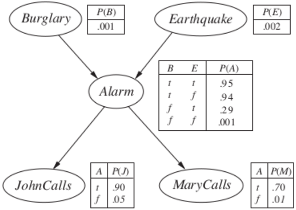
Bayesian networks: advantages

n variables, each influenced by at most k variables $\rightarrow 2^k$ (to specify a conditional probability table) $\rightarrow n2^k$
vs.
Joint distribution: 2^n

Example: n=30 vertices, $k = 5$ parents $\rightarrow$ 960 vs. 10^9.

Conditional independence

- Each variable is conditionally independent of all its non-descendants, given the parents.



JohnCalls is independent of *Burglary*, *Earthquake*, *MarryCalls*, given *Alarm*.

- *Markov blanket*: a node is conditionally independent of other nodes, given its parents, children, and parents of children

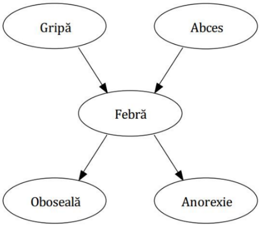
Burglary is independent of *JohnCalls* and *MaryCalls*, given *Alarm* and *Earthquake*

Content

Inference of marginal probabilities

- Compute the node probabilities, in the absence of evidence nodes
- For a node: compute the sum
  of the conditional probabilities of the possible combinations of parent values, multiplied by
  the parents' probabilities of having those values.

Inference of marginal probabilities: example



P(Gripă = Da)	P(Gripă = Nu)
0,1	0,9

P(Abces = Da)	P(Abces = Nu)
0,05	0,95

Gripă	Abces	P(Febră = Da)	P(Febră = Nu)
Da	Da	0,8	0,2
Da	Nu	0,7	0,3
Nu	Da	0,25	0,75
Nu	Nu	0,05	0,95

Febră	P(Oboseală = Da)	P(Oboseală = Nu)
Da	0,6	0,4
Nu	0,2	0,8

Febră	P(Anorexie = Da)	P(Anorexie = Nu)
Da	0,5	0,5
Nu	0,1	0,9

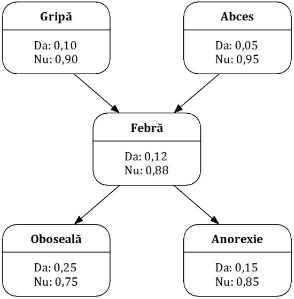
$$\begin{aligned} P(f) &= P(f|g,a)P(g)P(a) + P(f|g,\neg a)P(g)P(\neg a) \\ &+ P(f|\neg g,a)P(\neg g)P(a) + P(f|\neg g,\neg a)P(\neg g)P(\neg a) \\ &= 0.8 \cdot 0.1 \cdot 0.05 + 0.7 \cdot 0.1 \cdot 0.95 + 0.25 \cdot 0.9 \cdot 0.05 + 0.05 \cdot 0.9 \cdot 0.95 \\ &= 0.1245 \end{aligned}$$

$$P(\neg f) = 1 - P(f) = 0.8755$$

Example: Fatigue and Anorexia

$$
\begin{aligned}
P(o) &= P(o|f)P(f) + P(o|\neg f)P(\neg f) \\
&= 0.6 \cdot 0.1245 + 0.2 \cdot 0.8755 = 0.25 \\
P(\neg o) &= 1 - P(o) = 0.75
\end{aligned}
$$

$$
\begin{aligned}
P(x) &= P(x|f)P(f) + P(x|\neg f)P(\neg f) \\
&= 0.5 \cdot 0.1245 + 0.1 \cdot 0.8755 = 0.15 \\
P(\neg x) &= 1 - P(x) = 0.85
\end{aligned}
$$



Inference in bayesian networks

- Probabilistic inference system: compute the posterior probability distribution for a set of variables, given an event
- X variable, $\mathbf{E}$ the set of evidence variables $E_1, \ldots, E_m$, $\mathbf{e}$ event, $\mathbf{Y}$ non-evidence, $Y_1, \ldots Y_l$ hidden variables

  Posterior probability distribution $\mathbf{P}(X|\mathbf{e}) =?$

  Example: observe the event in which $JohnCalls = true$ and $MaryCalls = true$; then, the probability of a burglary: $\mathbf{P}(Burglary|JohnCalls = true, MaryCalls = true) = \langle 0.284, 0.716 \rangle$.

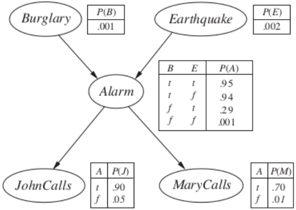
Inference by enumeration

$$\mathbf{P}(X|\mathbf{e}) = \alpha\mathbf{P}(X,\mathbf{e}) = \alpha\sum_{\mathbf{y}}\mathbf{P}(X,\mathbf{e},\mathbf{y}) \tag{1}$$

A Bayesian network provides a representation of the joint distribution. According to equation (2), the term $\mathbf{P}(X,\mathbf{e},\mathbf{y})$ can be written as a product of conditional probabilities.

Inference by enumeration: example

Query: What is the probability of a burglary when John and Mary call?
$P(Burglary|JohnCalls = true, MaryCalls = true)$



Inference by enumeration: example

$P(Burglary|JohnCalls = true, MaryCalls = true)$

- The hidden variables are *Earthquake* and *Alarm*.
  $\mathbf{P}(B|j,m) = \alpha\mathbf{P}(B,j,m) = \alpha\sum_e\sum_a\mathbf{P}(B,j,m,e,a)$

- For $Burglary = true$:
  According to (2):
  $P(b|j,m) = \alpha\sum_e\sum_a P(b)P(e)P(a|b,e)P(j|a)P(m|a)$

  The term $P(b)$ const, $P(e)$ it doesn't depend on a:

$$
\begin{aligned}
P(b|j,m) &= \alpha P(b)\sum_e P(e)\sum_a P(a|b,e)P(j|a)P(m|a)\\
&= \alpha P(b)\sum_e P(e)[P(a|b,e)P(j|a)P(m|a)+\\
&\quad P(\neg a|b,e)P(j|\neg a)P(m|\neg a)]\\
&= \ldots\\
&= \alpha \times 0.00059224
\end{aligned}
$$

Inference by enumeration: example

For *Burglary* = *false*, according to (2):

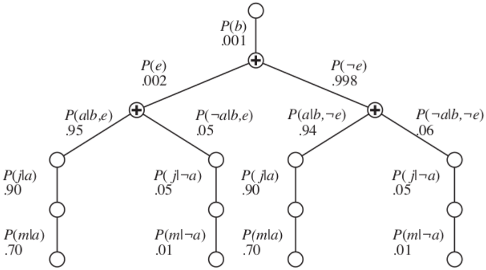
$$P(\neg b|j,m) = \alpha \sum_e \sum_a P(\neg b)P(e)P(a|\neg b,e)P(j|a)P(m|a)$$
$$= ...$$
$$= \alpha \times 0.0014919$$

$P(b|j,m) + P(\neg b|j,m) = 1 \rightarrow alpha = 479.8142$
$P(B|j,m) = \alpha\langle 0.00059224, 0.0014919\rangle \approx \langle 0.284, 0.716\rangle$.
The probability of a burglary: 28%.

Obs: to make the computation more efficient, it is recommended that the remaining nodes are first topologically sorted, i.e. parents to appear before children. In this case, it will be possible to decompose the sums more easily.

Example

The evaluation process:



Inference by enumeration

function ENUMERATION-ASK(X, $\mathbf{e}$, bn) **returns** a distribution over X
  inputs: X, the query variable
    $\mathbf{e}$, observed values for variables $\mathbf{E}$
    bn, a Bayes net with variables $\{X\} \cup \mathbf{E} \cup \mathbf{Y}$ /* $\mathbf{Y}$ = *hidden variables* */

  $\mathbf{Q}(X) \leftarrow$ a distribution over X, initially empty
  for each value x_i of X **do**
    $\mathbf{Q}(x_i) \leftarrow$ ENUMERATE-ALL(bn.VARS, $\mathbf{e}_{x_i}$)
      where $\mathbf{e}_{x_i}$ is $\mathbf{e}$ extended with $X = x_i$
  return NORMALIZE($\mathbf{Q}(X)$)

---

function ENUMERATE-ALL($vars$, $\mathbf{e}$) **returns** a real number
  if EMPTY?($vars$) **then return** 1.0
  $Y \leftarrow$ FIRST($vars$)
  if Y has value y in $\mathbf{e}$
    then return $P(y \mid parents(Y)) \times$ ENUMERATE-ALL(REST($vars$), $\mathbf{e}$)
    else return $\sum_y P(y \mid parents(Y)) \times$ ENUMERATE-ALL(REST($vars$), $\mathbf{e}_y$)
      where $\mathbf{e}_y$ is $\mathbf{e}$ extended with $Y = y$

The algorithm evaluates expression trees in DFS order. Space complexity: $O(n)$, n variables. Time complexity: $O(2^n)$ for a network with n variables *bool*.

Variable elimination algorithm

- Obs: $P(j|a)P(m|a)$ and $P(j|\neg a)P(m|\neg a)$ are computed twice, for each value of e.
- Idea: perform computations from right to left (from bottom to top) and save the results.

Variable elimination: example

$$P(B|j,m) = \alpha \underbrace{P(b)}_{f_1(B)} \sum_e \underbrace{P(e)}_{f_2(E)} \sum_a \underbrace{P(a|B,e)}_{f_3(A,B,E)} \underbrace{P(j|a)}_{f_4(A)} \underbrace{P(m|a)}_{f_5(A)}$$

- Each factor f_i is a matrix indexed by the argument variables:
  $f_4(A) = \begin{pmatrix} P(j|a) \\ P(j|\neg a) \end{pmatrix} = \begin{pmatrix} 0.90 \\ 0.05 \end{pmatrix}$, $f_5(A) = \begin{pmatrix} P(m|a) \\ P(m|\neg a) \end{pmatrix} = \begin{pmatrix} 0.70 \\ 0.01 \end{pmatrix}$.
  $f_3(A, B, E)$ is a 2x2x2 matrix.

$$P(B|j,m) = \alpha f_1(B) \times \sum_e f_2(E) \times \sum_a f_3(A,B,E) \times f_4(A) \times f_5(A)$$

Variable elimination: example

$$
\begin{aligned}
f_6(B,E) &= \sum_a f_3(A,B,E) \times f_4(A) \times f_5(A) \\
&= f_3(a,B,E) \times f_4(a) \times f_5(a) + f_3(\neg a,B,E) \times f_4(\neg a) \times f_5(\neg a)
\end{aligned}
$$

$$
\begin{aligned}
f_7(B) &= \sum_e f_2(E) \times f_6(B,E) \\
&= f_2(e) \times f_6(B,e) + f_2(\neg e) \times f_6(B,\neg e)
\end{aligned}
$$

$$P(B|j,m) = \alpha f_1(B) \times f_7(B)$$

Approximate inference

- Exact algorithms cannot be applied for complex networks with hundreds of nodes. Approximate inference increase computation speed.
- Random sampling algorithms for computing conditional probabilities: *Rejection sampling*, *Likelihood weighting*.
  - generate N samples from a distribution
  - compute an approximation $\hat{P}$ for the posterior probability (converges to the true probability P)
- *Prior sampling*

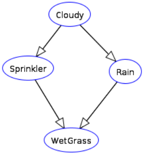
function PRIOR-SAMPLE(bn) **returns** an event sampled from the prior specified by bn
  inputs: bn, a Bayesian network specifying joint distribution $\mathbf{P}(X_1, \ldots, X_n)$

  $\mathbf{x} \leftarrow$ an event with n elements
  foreach variable X_i **in** $X_1, \ldots, X_n$ **do**
    $\mathbf{x}[i] \leftarrow$ a random sample from $\mathbf{P}(X_i \mid parents(X_i))$
  return x

The probability of generating a particular event:
$S_{PS}(x_1 \ldots x_n) = \prod_{i=1} P(x_i | parents(X_i)) = P(x_1 \ldots x_n)$.

Prior sampling: example

Consider the ordering $Cloudy, Sprinkler, Rain, WetGrass$.



Sample from $\mathbf{P}(Cloudy) =< 0.5, 0.5 >$ the value *true*.
Sample from $\mathbf{P}(Sprinkler|Cloudy = true) =< 0.1, 0.9 >$ the value *false*.
Sample from $\mathbf{P}(Rain|Cloudy = true) =< 0.8, 0.2 >$ the value *true*.
Sample from $\mathbf{P}(WetGrass|Sprinkler = false, Rain = true) =< 0.9, 0.1 >$ the value *true*.

$S_{PS}(t, f, t, t) = 0.5 \times 0.9 \times 0.8 \times 0.9 = 0.324 = P(t, f, t, t)$

$N_{PS}(x_1 \dots x_n)$ no. of samples generated for the event $x_1, \dots, x_n$.

$$\begin{aligned} lim_{N\to\infty}\hat{P}(x_1, \dots, x_n) &= lim_{N\to\infty}N_{PS}(x_1, \dots, x_n)/N \\ &= S_{PS}(x_1, \dots, x_n) \\ &= P(x_1 \dots x_n) \end{aligned}$$

If in 511 samples out of 1000 'Rain=true', then $\hat{P}(Rain = true) = 0.511$.

Rejection sampling

- generate samples from the prior distribution
- reject samples that do not fit the evidence
- estimate $\hat{P}(X = x|e)$

$$\hat{P}(X|e) = \alpha N_{PS}(X, e) = \frac{N_{PS}(X, e)}{N_{PS}(e)}$$

Example: estimate $\mathbf{P}(Rain|Sprinkler = true)$
Of the 100 samples generated, 73 have 'Sprinkler = false' (rejected) and 27 have 'Sprinkler = true'; of those 27, 8 have 'Rain = true' and 19 'Rain = false'.

$$P(Rain|Sprinkler = true) \approx NORMALIZE(< 8, 19 >) =< 0.296, 0.704 >$$

Rejection sampling

```
function REJECTION-SAMPLING(X, e, bn, N) returns an estimate of P(X|e)
  inputs: X, the query variable
          e, observed values for variables E
          bn, a Bayesian network
          N, the total number of samples to be generated
  local variables: N, a vector of counts for each value of X, initially zero

  for j = 1 to N do
     x ← PRIOR-SAMPLE(bn)
     if x is consistent with e then
        N[x] ← N[x]+1 where x is the value of X in x
  return NORMALIZE(N)
```

Disadvantage: many samples are rejected (the no. of consistent samples decreases exponentially with the no. of evidence vars)

Bibliography

- S. Russell, P. Norvig. *Artificial Intelligence: A Modern Approach*. Ch. 13. Quantifying Uncertainty; Ch. 14. Probabilistic Reasoning
- Belief and Decision Networks `https://aispace.org/bayes/`