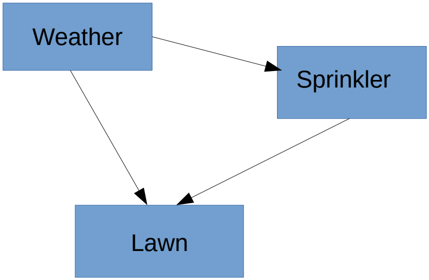
Probabilistic Knowledge Representation and Reasoning

sprinkler example based on Gilad Barkan's slides, slideshare

Chapter 14, AIMA, 3rd edition

Chapter 13, AIMA, 3rd edition – review on probability

Belief Networks



Belief Networks

- Probabilistic Graphical Model (PGM)
- Graphical (Directed Acyclic Graph) Model
- Nodes are variables (features):
  - random variable with a probability distribution
  - set of parameters/values/states. For example:
    - Weather = {sunny, cloudy, rainy};
    - Sprinkler={off,on};
    - Lawn={dry,wet}
    - Possible scenario:

$$Weather = rainy, Sprinkler = off, Lawn = wet$$

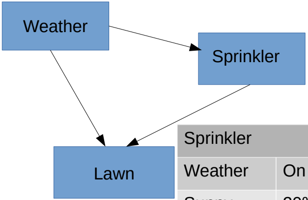
- Edges (links) represent relations between variables
- Edges **may** indicate causality (for example, "rainy weather" ou "sprinkler" **may be** the cause of "wet lawn").

Belief Networks

- Conditional Probability Table: used to store all beliefs related with the possible states of a node (variable)

Belief Networks

Weather	
Sunny	10%
Cloudy	30%
Rainy	60%



Sprinkler		
Weather	On	Off
Sunny	20%	80%
Cloudy	10%	90%
Rainy	0%	100%

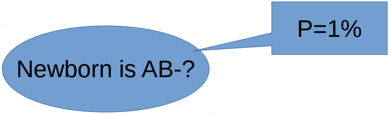
Lawn			
Weather	Sprinkler	Wet	Dry
Sunny	On	20%	80%
Cloudy	On	40%	60%
Rainy	On	100%	0%
Sunny	Off	0%	100%
Cloudy	Off	10%	90%
Rainy	Off	100%	0%

Conditional Probability: P(L=wet|W=sunny,S=on)

Bayesian Inference

- Once we have a model consisting of a **graph** and **TPCs**, we can answer queries like:
  - Given that it rained, would the lawn be wet? (trivial)
  - Given that the lawn is wet, what could be the reason?
    - "rainy weather"?
    - "sprinkler on"?

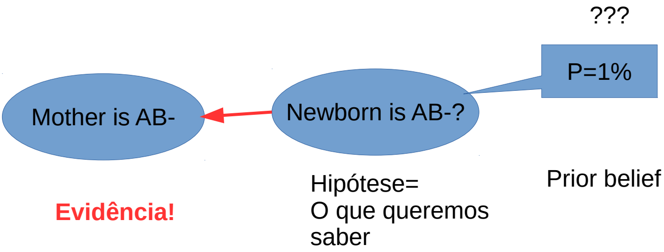
Bayes theorem in action!



Hipótese=
O que queremos
saber

Prior belief

Bayes theorem in action!

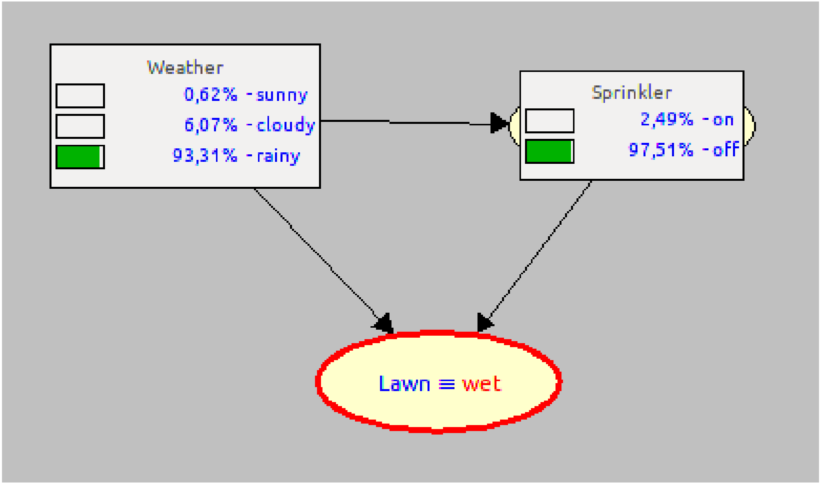


Sabendo a evidência, P pode deixar de ter valor 0.01
Nova probabilidade para P: a posteriori update belief

Bayesian Inference

- From our example, what can we answer using the Bayes theorem?
- Possible metric (Bayesian decision rule): Maximum a Posteriori Probability (MAP)
  - P(Weather=rainy|Lawn=wet) = 0.93; P(Sprinkler=on|Lawn=wet)=0.024
- In this case, the lawn is wet because it rained!

Bayesian Inference



Probabilities calculated with the samIam software

Bayesian Inference

Is the lawn wet because it rained?

$$P(Weather = rainy \mid Lawn = Wet) =$$

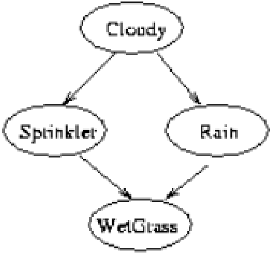
$$= \frac{P(Weather = rainy \wedge Lawn = Wet)}{P(Lawn = Wet)} =$$

$$= \frac{P(Weather = rainy \wedge Lawn = Wet \wedge Sprinkler)}{P(Lawn = Wet \wedge Weather \wedge Sprinkler)}$$

Bayesian Inference: simpler example with only boolean variables

P(C=F)	P(C=T)
0.5	0.5

C	P(S=F)	P(S=T)
F	0.5	0.5
T	0.9	0.1



C	P(R=F)	P(R=T)
F	0.8	0.2
T	0.2	0.8

S	R	P(W=F)	P(W=T)
F	F	1.0	0.0
T	F	0.1	0.9

Bayesian Inference: simpler example with only boolean variables

with True=1 and False=0, to simplify

$$P(S = 1 \mid W = 1) = \frac{P(S = 1, W = 1)}{P(W = 1)} = 0.43$$

$$\begin{aligned}
&P(W = 1) \\
&= \sum_{c=0}^{1} \sum_{r=0}^{1} \sum_{s=0}^{1} P(W = 1, C = c, R = r, S = s) \\
&= \sum_{c=0}^{1} \sum_{r=0}^{1} \sum_{s=0}^{1} P(W = 1 \mid S = s, R = r) \cdot P(S = s \mid C = c) \cdot P(R = r \mid C = c) \cdot P(C = c) \\
&= \sum_{c=0}^{1} P(C = c) \sum_{r=0}^{1} P(R = r \mid C = c) \sum_{s=0}^{1} P(S = s \mid C = c) \cdot P(W = 1 \mid S = s, R = r).
\end{aligned}$$

Bayesian Inference: simpler example with only boolean variables

$$
\begin{aligned}
& P(S=1 \mid W=1) \\
&= \frac{P(S=1, W=1)}{P(W=1)} \\
&= \frac{1}{P(W=1)} \sum_{r=0}^{1} \sum_{c=0}^{1} P(S=1, W=1, R=r, C=c) \\
&= \frac{1}{P(W=1)} \sum_{r=0}^{1} \sum_{c=0}^{1} P(W=1 \mid S=1, R=r) \cdot P(S=1 \mid C=c) \cdot P(R=r \mid C=c) \cdot P(C=c) \\
&= \frac{1}{P(W=1)} \sum_{c=0}^{1} P(C=c) P(S=1 \mid C=c) \sum_{r=0}^{1} P(R=r \mid C=c) \cdot P(W=1 \mid S=1, R=r)
\end{aligned}
$$

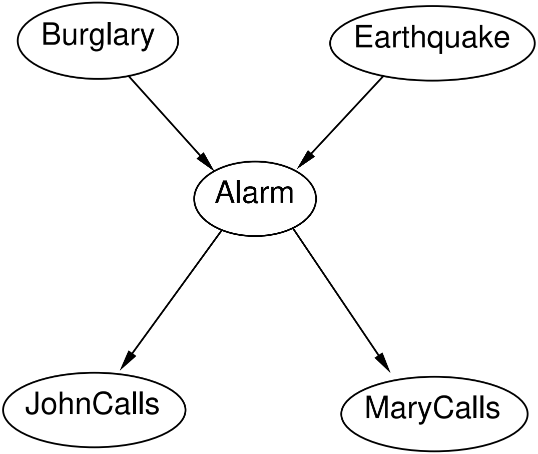
Probabilistic Reasoning Systems

- How to build reasoning systems that can use uncertainty?
- **Belief Networks** or Bayesian networks: graph with the following characteristics:
  - Nodes are random variables.
  - Directed edges are connections between the random variables.
  - Each node has a conditional probability table that probabilistically represents the effect of this nodes's parents.
  - The graph has no cycles (it is a DAG).
- relatively easy to define relations, but tricky to define probabilities.

Belief Networks

- Example: burglar alarm.
- Alarm sounds in two situations: burglary or earthquake.
- John e Mary are neighbours that call the owner's house when the alrm goes off.
- John calls whenever the alarm goes off and also when the telephone rings.
- Mary calls only when the alarm goes off, but sometimes she doesn't hear the alarm.
- Given the evidence of who called the house owner, what is the probability of a burglary?

Belief Networks

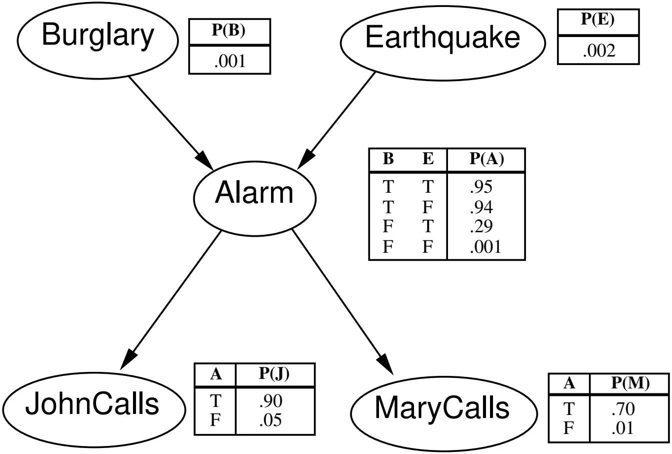


Belief Networks

- Network only represents direct connections, “causal”.
- Nothing is informed about Mary listening to loud music or to John confounding the telephone with the alarm sound.
- Conditional Probability Table:

Burglary	Earthquake	$\mathbf{P}(Alarm \mid Burglary, Earthquake)$
T	T	0.950 0.050
T	F	0.950 0.050
F	T	0.290 0.710
F	F	0.001 0.999

Belief Networks



Belief Networks

- Two ways of understanding:
  - representation of the joint probability distribution. Useful to **build** the network.
  - set of conditionally independent variables. Useful to design inference procedures.
- Representing the joint probability:
  - each table entry can be calculated through the information available in the network.
  - a generic table entry represents the probability of a conjunction of the random variable values: $P(X_1 = x_1 \wedge \ldots \wedge X_n = x_n)$.
    $P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid Parents(X_i))$

Belief Networks

- each table entry is represented by the product of the appropriate elements of the CPT (Conditional Probability Table).
- A CPT provides a decomposed representation of the joint distribution.
- Example:
  $P(J \wedge M \wedge A \wedge \neg B \wedge \neg E) = P(J \mid A)P(M \mid A)P(A \mid \neg B \wedge \neg E)P(\neg B)P(\neg E) = 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998 = 0.00062$

Belief Networks

- Method to build belief networks:
  - Network is built such that each node is conditionally independent of its predecessors, given the probability of its parents.
  - equation $P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid Parents(x_i))$ used to guide the knowledge engineer to build the network topology.
  - In order to build the network so that its structure is correct for the domain, choose suitable parents that guarantee that each node is conditionally independent on its parents.

Belief Networks

- In general: $\mathbf{P}(X_i \mid X_{i-1}, \ldots, X_1) = P(X_i \mid Parents(X_i))$, given that $Parents(X_i)$ is a subset of $\{x_{i-1}, \ldots, x_1\}$
- This condition holds as long as we label the nodes in a way that is consistent with the partial order that is implicit in the graph structure.
- For example: Mary calls: is not **directly** affected by burglary or earthquake. It is affected by its effect: the alarm goes off.
- John calls also does not have direct influence over Mary calls. In that case, we jave conditional independence: $\mathbf{P}(MaryCalls \mid JohnCalls, Alarm, Earthquake, Burglary) = \mathbf{P}(MaryCalls \mid Alarm)$

Belief Networks

- General procedure:
  1. Coose a subset X_i of relevante variables that describe the domain.
  2. Choose an order for those variables.
  3. While there are variables to be placed in the network:

     a) Take a var X_i and add a node to the network for that variable.

     b) Build the set $Parents(X_i)$ with a minimum set of nodes already in the net, such that the conditional independence constraint is satisfied.

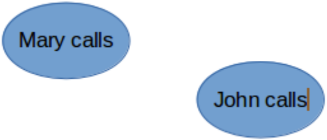
     c) Define the CPT for node X_i.

Belief Networks

- Procedure guarantees that network is acyclic.
- Network does not have redundant probability values.
- Guarantees that the axioms of probability theory are not violated.

Belief Networks

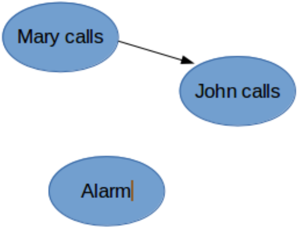
- Assuming the order: $M,\ J,\ A,\ B,\ E$



$$P(J \mid M) = P(J)?$$

Belief Networks

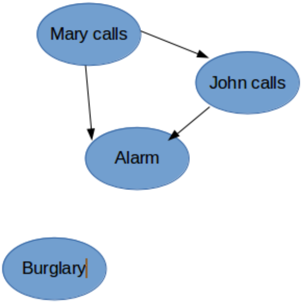
- Assuming the order: $M,\ J,\ A,\ B,\ E$



$P(J \mid M) = P(J)$? No
$P(A \mid J, M) = P(A \mid J)? P(A \mid J, M) = p(A)$?

Belief Networks

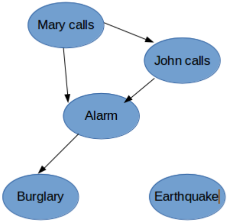
- Assuming the order: $M,\ J,\ A,\ B,\ E$



$P(J \mid M) = P(J)$? No
$P(A \mid J, M) = P(A \mid J)?P(A \mid J, M) = p(A)$? No
$P(B \mid A, J, M) = P(B \mid A)$?
$P(B \mid A, J, M) = P(B)$?

Belief Networks

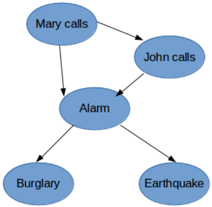
- Assuming the order: $M,\ J,\ A,\ B,\ E$



$P(J \mid M) = P(J)$? No
$P(A \mid J, M) = P(A \mid J)?P(A \mid J, M) = p(A)$? No
$P(B \mid A, J, M) = P(B \mid A)$? Yes
$P(B \mid A, J, M) = P(B)$? No
$P(E \mid B, A, J, M) = P(E \mid A)$?
$P(E \mid B, A, J, M) = P(E \mid A, B)$?

Belief Networks

- Assuming the order: $M,\ J,\ A,\ B,\ E$



$P(J \mid M) = P(J)$? No
$P(A \mid J, M) = P(A \mid J)?P(A \mid J, M) = p(A)$? No
$P(B \mid A, J, M) = P(B \mid A)$? Yes
$P(B \mid A, J, M) = P(B)$? No
$P(E \mid B, A, J, M) = P(E \mid A)$? No
$P(E \mid B, A, J, M) = P(E \mid A, B)$? Yes

Belief Networks

- Belief networks are more compact than the joint probability distribution table.
- **locally structured** systems: sparse with info distributed among the nodes.
- Polynomial growth.
- Assuming that most of the nodes are directly affected by at most k other variables (parents).
- Number of entries in each CPT: 2^k.
- To the complete network (n nodes): $n2^k$.

Belief Networks

- Concrete example: network with 20 nodes and at most 5 parents per node:
  - belief networks: 640 entries.
  - joint probability table: order of 10^6 entries.
- Number of edges: the more we have the greater the precision, but a big number of edges may increase the table sizes.
- General rule: add to the network: first the causal nodes and then their effects.

Belief Networks

- Problem: choose the conditional probabilities of a TPC.
- Relation between parents and children nodes can fit a canonical distribution. In that case, probs can be specified by names and aditional parameters.
- Simple example: deterministic nodes. Probs are the same as probs of their parents. Representation can be more compact.
- Non-deterministc nodes: noisy-OR.

Belief Networks

- Representation of probs:
  - If all parents are False, output node will be False with 100% probability.
  - If only one of the parents is True, output node will have value False with noisy parameter of that parent.
- Ex: $P(Fever \mid Cold) = 0.4$, $P(Fever \mid Flu) = 0.8$ and $P(Fever \mid Malaria) = 0.9$

Cold	Flu	Malaria	P(Fever)	P(¬ Fever)
F	F	F	0.0	1.0
F	F	T	0.9	0.1
F	T	F	0.8	0.2
F	T	T	0.98	$0.02 = 0.2 \times 0.1$
T	F	F	0.4	0.6
T	F	T	0.94	$0.06 = 0.6 \times 0.1$
T	T	F	0.88	$0.12 = 0.6 \times 0.2$
T	T	T	0.988	$0.012 = 0.6 \times 0.2 \times 0.1$

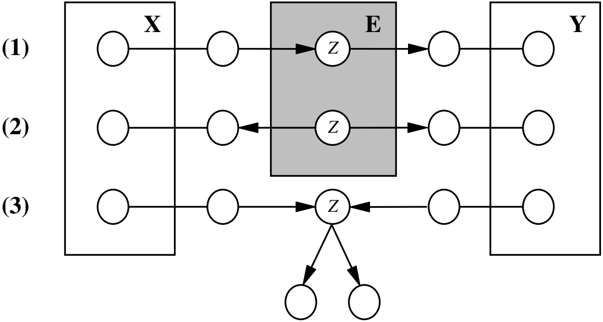
Belief Networks

- Need to answer: “Is there any set of nodes X independent of other set Y, given the evidence E?”
- Method: **direction-dependent separation** or **d-separation**.
- d-separation: *if every edge between X and Y is d-separated by E, then X and Y are conditionally independent, given E.* In that condition, we say that the edge is **blocked**.

Belief Networks

- An edge is blocked, given a set of nodes E, if there is a path such that one of the following conditions is satisfied:
  1. Variable Z is in E and Z has an arc incident and other not incident.
  2. Z is in E and Z has both arcs non incident.
  3. Nor Z nor any descendant of Z are in E, and both arcs are incident to Z.

Belief Networks



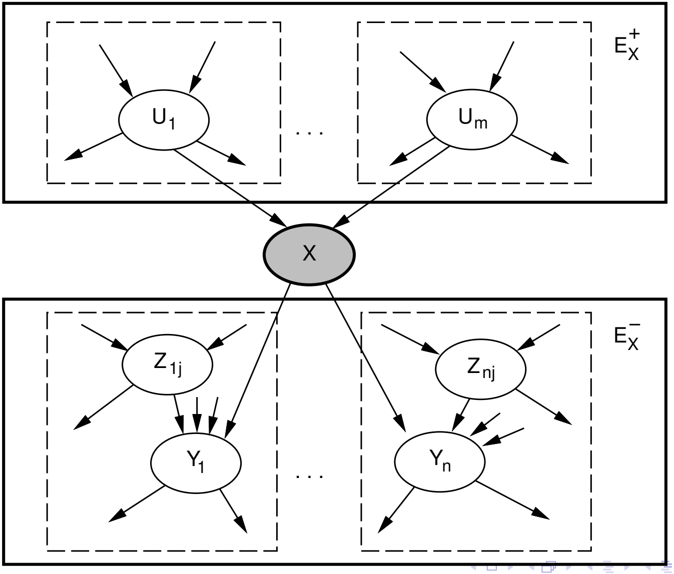
Belief Networks

- Objective: compute posterior probability distribution for a set of query variables, given the evidence: $P(Query \mid Evidence)$.
- In principle any node can be query or evidence. When learning classifiers, one variable is the query and some others are evidence.

Belief Networks

- **Diagnostic**: $P(Burglary \mid JohnCalls)$ (effetcs to causes).
- **Causal**: $P(JohnCalls \mid Burglary)$ (cause to effects).
- **Inter-causal**: $P(Burglary \mid Alarm \wedge Earthquake)$.
- **Mixed**: combination of one or more of the previous cases.
- Belief Networks can be used to:
  - make decisions.
  - decide which evidence variables to observe in order to obtain more useful information.
  - to perform "sensitivity analysis" and understand which aspects of the model have greater impact on the probability of the query variables.
  - Explain the results of the probabilistic inference to the user.

Belief Networks



Belief Networks

- Node X has parents U and children Y.
- Blocks are "singly connected": all blocks are disjunct and do not have links.
- X is the query variable.
- Objective: compute $P(X \mid E)$.
- Set of **causal support**: evidence variables "above" X that are connected through its parents.
- Set of **evidential support**: evidence variables that are "below" X and are connected through its children.
- $E_{U_i|X}$: evidence connected with all U_i nodes, *except* via the edge that goes through X.

Belief Networks

- General strategy:
  - Represent $\mathbf{P}(X \mid E)$ in terms of contributions E_X^+ e E_X^-.
  - Compute E_X^+ through its effects on X's parents. Obs: to compute the effects of X's parents can be done recursively.
  - idem to E_X^-
- Method: apply Bayes, and other methods such as simplifications (conditional independence).

Belief Networks

- $\mathbf{P}(X \mid E) = \mathbf{P}(X \mid E_X^-, E_X^+) = \frac{\mathbf{P}(E_X^- | X, E_X^+)\mathbf{P}(X | E_X^+)}{\mathbf{P}(E_X^- | E_X^+)}$
- As X d-separates E_X^+ from E_X^-, we can use conditional independence to simplify the first term's numerator. We can also use $\frac{1}{\mathbf{P}(E_X^- | E_X^+)}$ as a normalization constant: $\mathbf{P}(X \mid E) = \alpha \mathbf{P}(E_X^- \mid X)\mathbf{P}(X \mid E_X^+)$
- $\mathbf{P}(X \mid E_X^+) = \sum_u \mathbf{P}(X \mid \mathbf{u}, E_X^+) P(\mathbf{u} \mid E_X^+)$
- $\mathbf{P}(X \mid E_X^+) = \sum_u \mathbf{P}(X \mid \mathbf{u}) \prod_i \mathbf{P}(u_i \mid E_X^+)$
- $\mathbf{P}(X \mid E_X^+) = \sum_u \mathbf{P}(X \mid \mathbf{u}) \prod_i \mathbf{P}(u_i \mid E_{U_i | X})$

Belief Networks

- Let Z_i parents of Y_i and z_i a set of values for Z_i.
  $\mathbf{P}(E_X^- \mid X) = \prod_i \mathbf{P}(E_{Y_i|X} \mid X)$
  $\mathbf{P}(E_X^- \mid X) = \prod_i \sum_{y_i} \sum_{z_i} \mathbf{P}(E_{Y_i|X} \mid X, y_i, z_i)\mathbf{P}(y_i, z_i \mid X)$
- Decomposing $E_{Y_i|X}$ in two independent components $E_{Y_i}^+$ and $E_{Y_i|X}^-$
  $\mathbf{P}(E_X^- \mid X) = \prod_i \sum_{y_i} \sum_{z_i} \mathbf{P}(E_{Y_i}^- \mid X, y_i, z_i)\mathbf{P}(E_{Y_i|X}^+ \mid X, y_i, z_i)\mathbf{P}(y_i, z_i \mid X)$
  $\mathbf{P}(E_X^- \mid X) = \prod_i \sum_{y_i} \mathbf{P}(E_{Y_i}^- \mid y_i) \sum_{z_i} \mathbf{P}(E_{Y_i|X}^+ \mid z_i)\mathbf{P}(y_i, z_i \mid X)$

Belief Networks

- Applying Bayes to $\mathbf{P}(E^+_{Y_i|X} \mid z_i)$:

  $\mathbf{P}(X \mid E^-_X) = \prod_i \sum_{y_i} \mathbf{P}(E^-_{Y_i} \mid y_i) \sum_{z_i} \frac{P(z_i|E^+_{Y_i|X})P(E^+_{Y_i|X})}{P(z_i)} \mathbf{P}(y_i, z_i \mid X)$

  ...

  $\mathbf{P}(X \mid E^-_X) = \beta \prod_i \sum_{y_i} P(E^-_{Y_i} \mid y_i) \sum_{z_i} P(y_i \mid X, z_i) \prod_{z_i} P(z_{ij} \mid E_{Z_{ij}|Y_i})$
- $P(E^-_{Y_i} \mid y_i)$ is a recursive instance of $P(E^-_X \mid X)$.
- $P(y_i \mid X, z_i)$ is taken directly from the CPT of Y_i.
- $P(z_{ij} \mid E_{Z_{ij}|Y_i})$ is a recursive instance of $P(X \mid E)$.

Belief Networks

function BELIEF-NET-ASK(X) **returns** a probability distribution over the values of X
  inputs: X, a random variable

  SUPPORT-EXCEPT(X, *null*)

---

function SUPPORT-EXCEPT(X, V) **returns** $\mathbf{P}(X|E_{X\setminus V})$

  if EVIDENCE?(X) **then return** observed point distribution for X
  else
    calculate $\mathbf{P}(E^-_{X\setminus V}|X)$ = EVIDENCE-EXCEPT(X, V)
    $U \leftarrow$ PARENTS[X]
    if U is empty
      then return $\alpha\ \mathbf{P}(E^-_{X\setminus V}|X)\ \mathbf{P}(X)$
    else
      for each U_i **in** U
        calculate and store $\mathbf{P}(U_i|E_{U_i\setminus X})$ = SUPPORT-EXCEPT(U_i, X)
      return $\alpha\ \mathbf{P}(E^-_{X\setminus V}|X) \sum_{\mathbf{u}} \mathbf{P}(X|\mathbf{u}) \prod_i \mathbf{P}(U_i|E_{u_i\setminus X})$

---

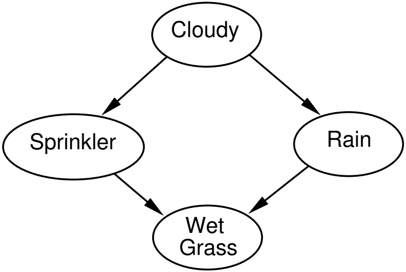
function EVIDENCE-EXCEPT(X, V) **returns** $\mathbf{P}(E^-_{X\setminus V}|X)$

  $\mathbf{Y} \leftarrow$ CHILDREN[X] $-$ V
  if Y is empty
    then return a uniform distribution
  else
    for each Y_i **in Y do**
      calculate $\mathbf{P}(E^-_{Y_i}|y_i)$ = EVIDENCE-EXCEPT(Y_i, null)
      $\mathbf{Z}_i \leftarrow$ PARENTS[Y_i] $-$ X
      for each Z_{ij} **in** $\mathbf{Z}_i$
        calculate $\mathbf{P}(Z_{ij}|E_{Z_{ij}\setminus Y_i})$ = SUPPORT-EXCEPT(Z_{ij}, Y_i)
    return $\beta \prod_i \sum_{y_i} P(E^-_{Y_i}|y_i) \sum_{\mathbf{z}_i} \mathbf{P}(y_i|X,\mathbf{z}_i) \prod_j P(z_{ij}|E_{Z_{ij}\setminus Y_i})$

Belief Networks

Multiply connected network: two nodes are connected through more than one path in the graph.

P(C) = .5



C	P(S)
T	.10
F	.50

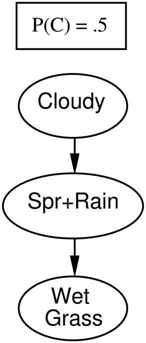
C	P(R)
T	.80
F	.20

S	R	P(W)
T	T	.99
T	F	.90
F	T	.90
F	F	.00

Belief Networks

- three classes of algorithms:
  - **Clustering**: transforms the network in a poly-tree probabilistically equivalent, but with different topology.
  - **Conditioning**: opposite of clustering, transforms the network in several poly-trees by instantiating different values for the random variables. Evaluates each poly-tree for each different instantiation.
  - **Stochastic Simulation** (logical sampling): computes an approximate probability through repeated simulations of the network, observing the frequency of relevant events.
  - In general: exact inference in belief networks is an NP-hard problem.

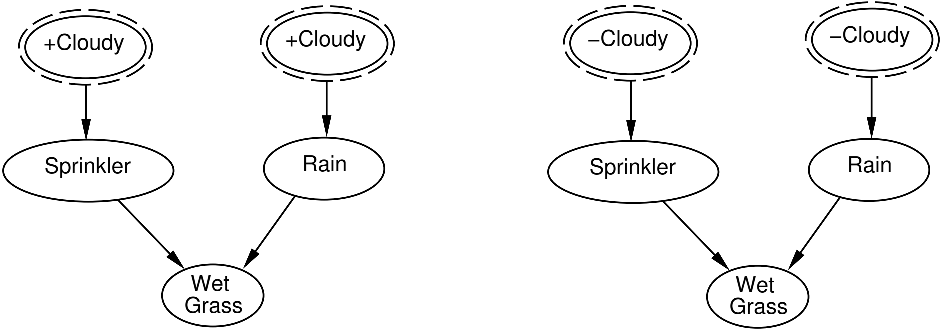
Belief Networks



C	P(S+R=x) TT	TF	FT	FF
T	.08	.02	.72	.18
F	.40	.10	.40	.10

S+R	P(W)
T T	.99
T F	.90
F T	.90
F F	.00

Belief Networks



Belief Networks

- Decide about what to talk.
- Decide about the vocabulary and random variables.
- Code the knowledge about dependencies among variables.
- Code a specific description of the problem.
- Query the system.