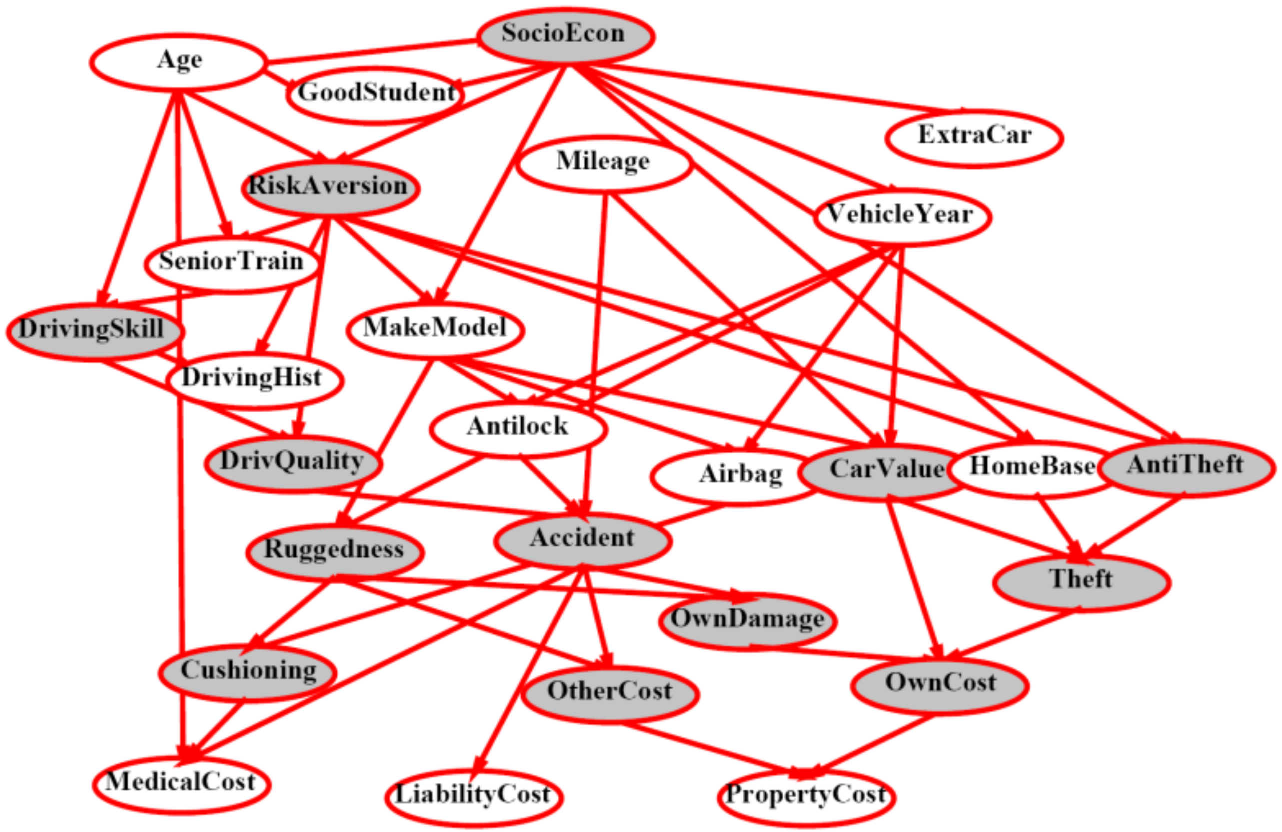
CSCI 3202: Intro to Artificial Intelligence

Lecture 21 & 22: Introduction to Bayesian Networks

Rachel Cox

Department of Computer Science



Uncertainty

Probabilistic reasoning gives us a framework for managing uncertain beliefs and knowledge.

In general:

- **Observed variables (evidence)**: Agent knows certain things about the state of the world (e.g. sensor readings or symptoms)
- **Unobserved variables**: Agent needs to reason about other aspects (e.g. where an object is or what disease is present)
- **Model**: Agent knows something about how the known variables relate to the unknown variables

Bayesian Reasoning

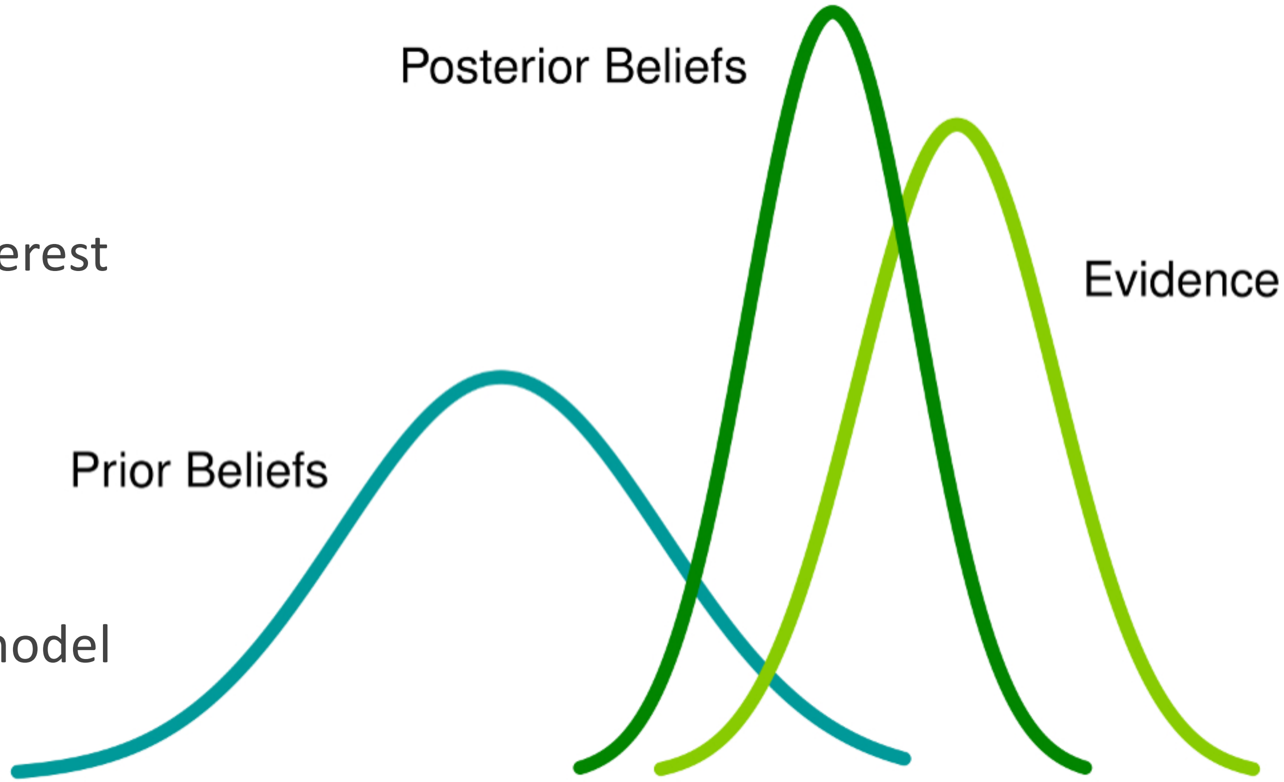
$$\underbrace{p(y|x)}_{\text{posterior}} = \frac{p(x|y)\,p(y)}{p(x)}$$

prior

evidence

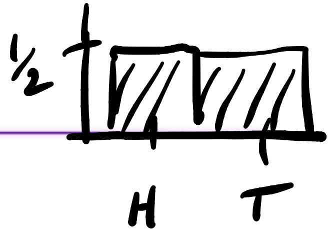
The whole Bayesian statistical framework can be summarized as follows:

- What is your process of interest?
- Get **data** (evidence) for your process of interest
- Build a **model** $\eta(\theta)$ of this process
  Depends on uncertain parameters θ
- Formalize your *a priori* knowledge of the model parameters θ in **prior distributions**, $P(\theta)$
- **Update** your prior knowledge using the match between your model and the data



Joint Distributions

coin flip :



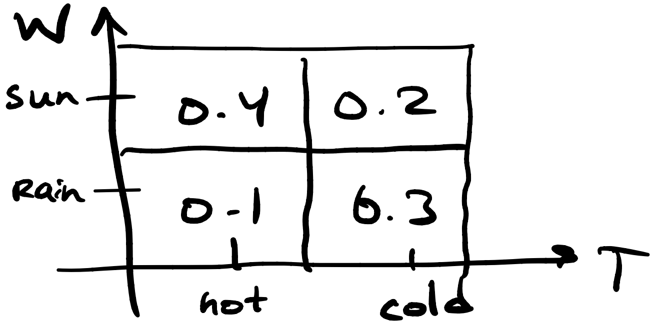
- A *joint distribution* over a set of random variables: $X_1, X_2, \ldots, X_n$ specifies a real number for each assignment (or *outcome*):

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3, \ldots, X_n = x_n) \text{ or equivalently: } P(x_1, x_2, \ldots, x_n)$$

- Must obey: $P(x_1, x_2, \ldots, x_n) \geq 0$ and
- $\boxed{\sum_{(x_1,x_2,\ldots,x_n)} P(x_1, x_2, \ldots, x_n) = 1}$
- all probabilities are between 0 and 1

$P(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3



- Size of distribution if n variables with domain sizes d?
  - For all but the smallest distributions, impractical to write out!

Probability Models

- A probability model is a joint distribution over a set of random variables

- Probability models:
  - (Random) variables with domains
  - Assignments are called *outcomes*
  - Joint distributions: say whether assignments (outcomes) are likely
  - *Normalized:* sum to 1.0
  - Ideally: only certain variables directly interact

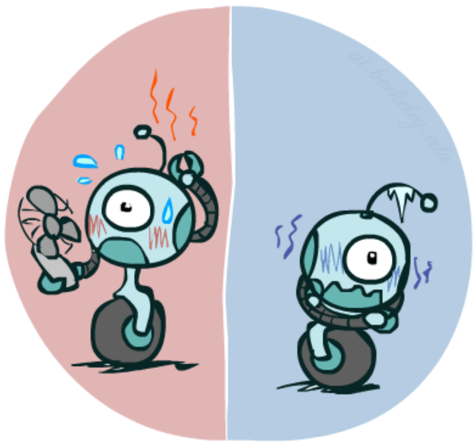
Distribution over T,W

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

Probability Distributions

- Associate a probability with each value

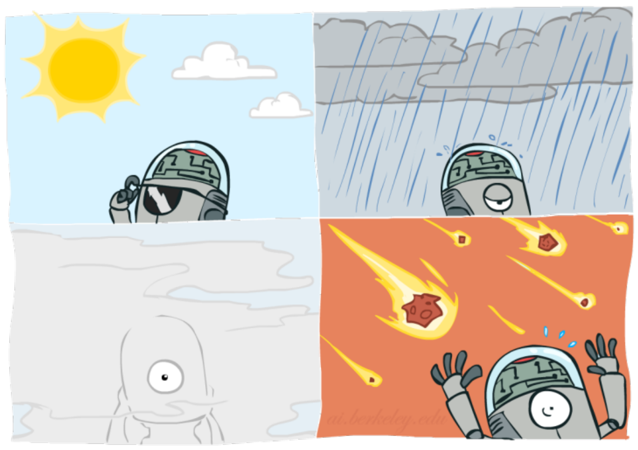
  - Temperature:



$$P(T)$$

T	P
hot	0.5
cold	0.5

  - Weather:



$$P(W)$$

W	P
sun	0.6
rain	0.1
fog	0.3
meteor	0.0

Marginal Distributions

- Marginal distributions are sub-tables which eliminate variables
- Marginalization (summing out): Combine collapsed rows by adding

$T = \{hot, cold\}$
$W = \{sun, rain\}$

$P(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$$P(t) = \sum_w P(t, w)$$

$P(hot) = P(hot, sun) + P(hot, rain)$
$P(cold) = P(cold, sun) + P(cold, rain)$

$P(T)$

T	P
hot	0.5
cold	0.5

$$P(w) = \sum_t P(t, w)$$

$P(W)$

W	P
sun	0.6
rain	0.4

$p(sun, cold) + p(sun, hot)$

e.g. $P(X_1 = x_1) = \sum_{x_2} P(X_1 = x_1, X_2 = x_2)$

Probability Models

Example:

- P(+x, +y) ? 0.2

- P(+x) ? 0.2 + 0.3 = 0.5

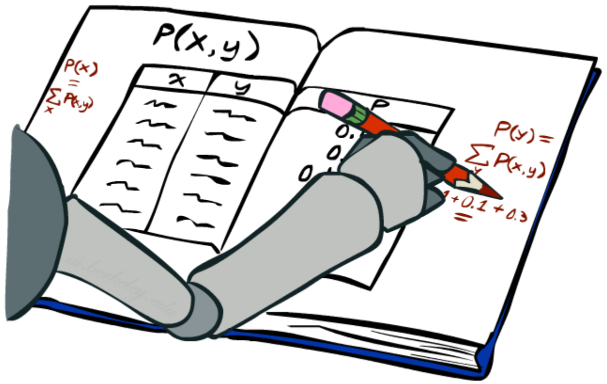
- P(-y OR +x) ? 0.6

$$P(X,Y)$$

X	Y	P
+x	+y	0.2
+x	-y	0.3
-x	+y	0.4
-x	-y	0.1

Marginal Distributions

exercise



$$P(X,Y)$$

X	Y	P
+x	+y	0.2
+x	-y	0.3
-x	+y	0.4
-x	-y	0.1

$$P(x) = \sum_y P(x,y)$$

$$P(y) = \sum_x P(x,y)$$

$$P(X)$$

X	P
+x	
-x	

$$P(Y)$$

Y	P
+y	
-y	

Conditional Probabilities

$P(X,Y)$

X	Y	P
+x	+y	0.2
+x	-y	0.3
-x	+y	0.4
-x	-y	0.1

- P(+x | +y) ?

$P(x \mid +y) \begin{bmatrix} 0.2 \\ 0.4 \end{bmatrix} \rightarrow$

$$P(+x \mid +y) = \frac{P(+x, +y)}{P(+y)} = \frac{0.2}{0.2+0.4} = 1/3$$

- P(-x | +y) ?

$$= \frac{P(-x, +y)}{P(+y)} = \frac{0.4}{0.2+0.4} = 2/3$$

- P(-y | +x) ?

$$= \frac{P(-y, +x)}{P(+x)} = \frac{0.3}{0.5} = .6$$

Conditional Distributions

- Conditional distributions are probability distributions over some variables given fixed values of others

Conditional Distributions

$P(W|T)$

$P(W|T = hot)$

W	P
sun	.8
rain	.2

$p(\text{sun} \mid T = \text{hot}) = \frac{p(\text{sun, hot})}{p(\text{hot})}$

$= \frac{0.4}{0.4 + 0.1}$

$=$

$P(W|T = cold)$

W	P
sun	
rain	

exercise

Joint Distribution

$P(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

Independence

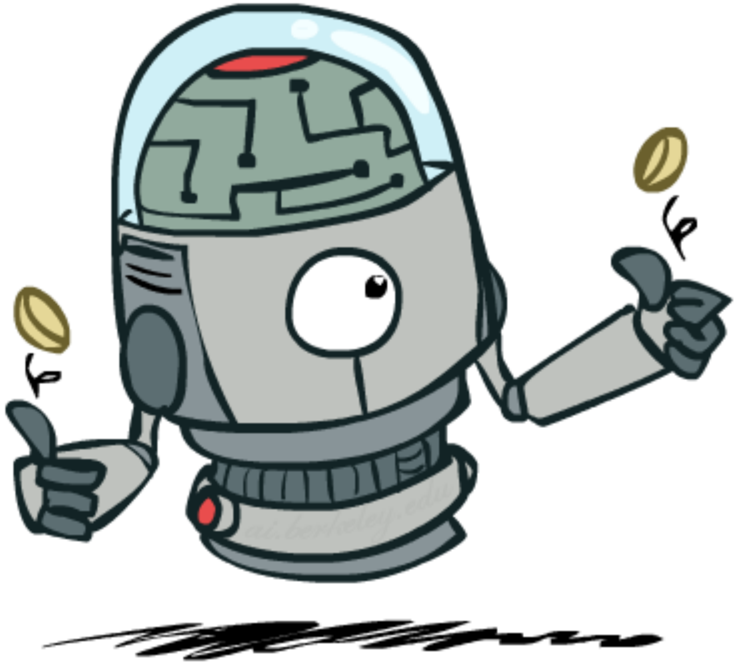
- Two variables are *independent* if:

$$\forall x, y : P(x, y) = P(x)P(y)$$

  - This says that their joint distribution *factors* into a product two simpler distributions
  - Another form:

$$\forall x, y : P(x|y) = P(x)$$

- Independence is a simplifying *modeling assumption*
  - *Empirical* joint distributions: at best "close" to independent
  - What could we assume for {Weather, Traffic, Cavity, Toothache}?



Independence

T, W not independent

If we have independence of T and W

$\Rightarrow p(T, W) = p(T) \cdot p(W)$

$P(T)$

T	P
hot	0.5
cold	0.5

$P_1(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$P(W)$

W	P
sun	0.6
rain	0.4

$P_2(T, W)$

T	W	P
hot	sun	0.3
hot	rain	0.2
cold	sun	0.3
cold	rain	0.2

$$P_1(T, W) = P(T) P(W)$$
$$= (0.5)(0.6)$$
$$= 0.3$$

This distribution assumes T, W are independent.

Conditional Independence

- Unconditional (absolute) independence very rare (why?)

- *Conditional independence* is our most basic and robust form of knowledge about uncertain environments.

- X is conditionally independent of Y given Z

$X \perp\!\!\!\perp Y \mid Z$

if and only if: $\forall x, y, z : P(x, y|z) = P(x|z)P(y|z)$

or, equivalently, if and only if $\forall x, y, z : P(x|z, y) = P(x|z)$ easier to compute

Bayesian Networks

Directed Acyclic Graph

The point of Bayes nets is to **represent full joint probability distributions**, *and*

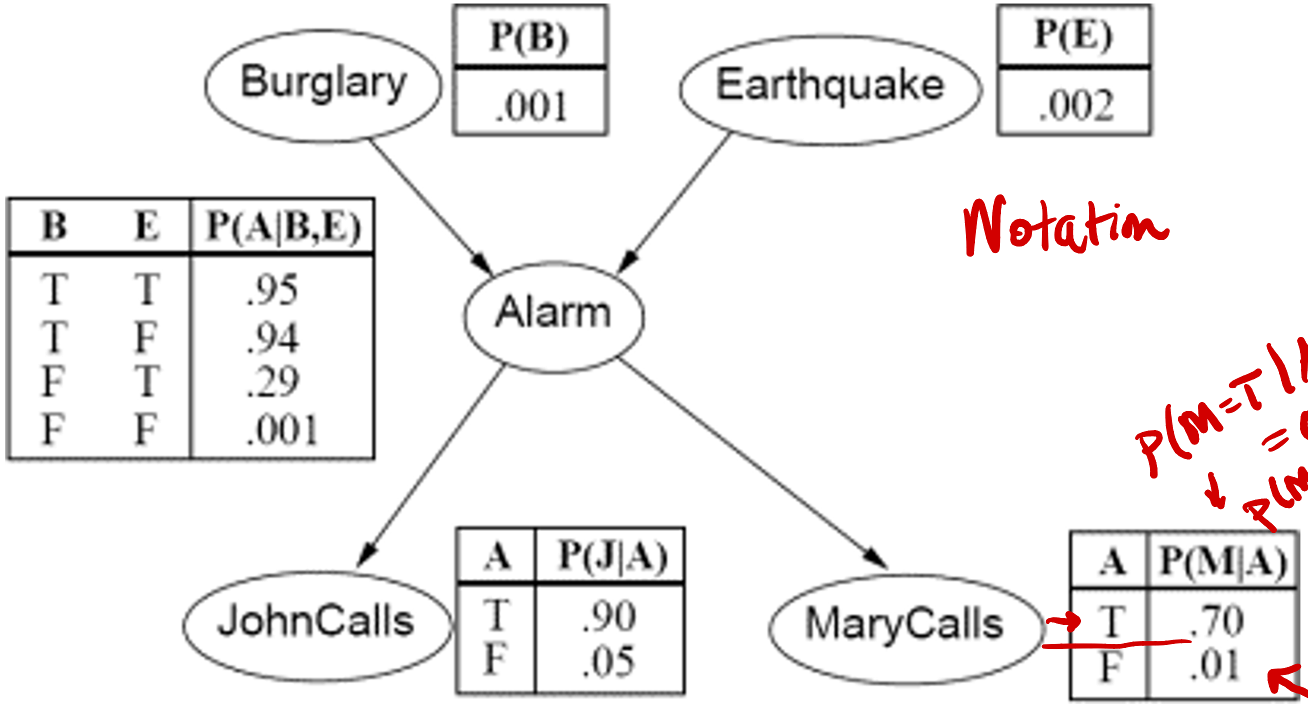
to encode an interrelated set of **conditional independence/probability statements**

chain rule

Full Joint Distribution

$P(B,E,A,J,M) = P(M \mid B,E,A,J)\,P(B,E,A,J)$

- Consists of **nodes** (events), and
- **conditional probability tables (CPTs)**, relating those events
- Describe how variables interact **locally**
- Chain together local interactions to estimate **global**, **indirect** interactions



P(B)
.001

P(E)
.002

B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

Notation

$P(M=T \mid A=T) = 0.7$

$P(M=F \mid A=T) = 0.3$

$P(M=T \mid A=F) = 0.01$

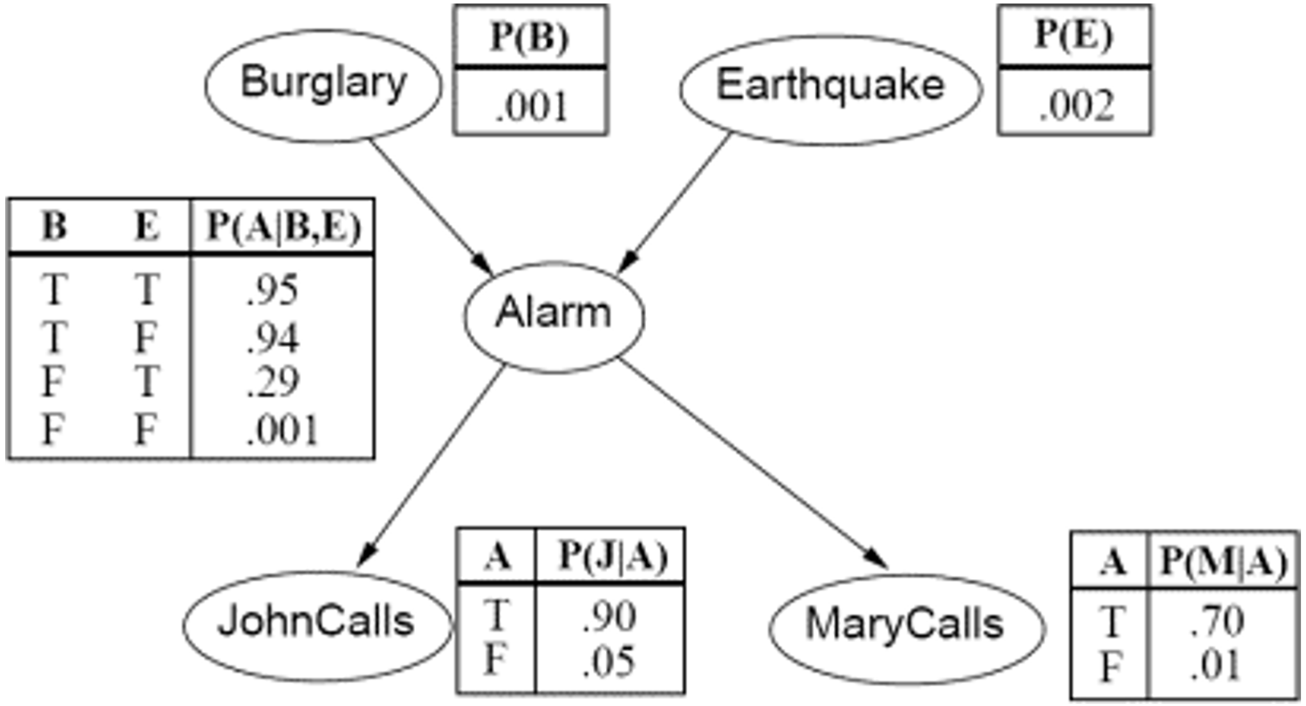
Bayesian Networks

The point of Bayes nets is to **represent full joint probability distributions**, *and*

to encode an interrelated set of **conditional independence/probability statements**

Notations: $P(A|B,E) = P(+A \mid +B, +E)$
$= P(A = \text{True} \mid B = T, E = T)$

- Consists of **nodes** (events), and
- **conditional probability tables (CPTs)**, relating those events
- Describe how variables interact **locally**
- Chain together local interactions to estimate **global**, **indirect** interactions



Bayesian Networks

The point of Bayes nets is to **represent full joint probability distributions**, *and*

to encode an interrelated set of **conditional independence/probability statements**

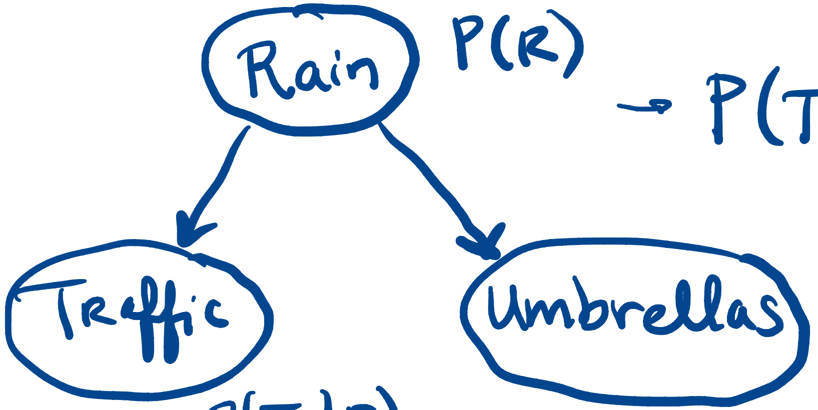
$T \perp\!\!\!\perp U \mid R$

__Example__: Represent the full joint distribution for P(Traffic, Rain, Umbrella)

Trivial decomposition: $P(T,R,U) = P(T \mid R,U)\,P(R,U)$

$= P(T \mid R,U)\,P(R \mid U)\,P(U)$

Conditional independence:



$P(R)$

$P(U \mid R)$

$P(T \mid R)$

$\rightarrow P(T,R,U) = P(T \mid R,U)\,P(R,U)$

$= \underline{P(T \mid R,U)}\,P(U \mid R)\,P(R)$ } Trivial decomposition

$= \underline{P(T \mid R)}\,P(U \mid R)\,P(R)$ } conditional independence

Bayesian Networks

The point of Bayes nets is to **represent full joint probability distributions**, *and*

to encode an interrelated set of **conditional independence/probability statements**

<u>Example (continued)</u>: Represent the full joint distribution for P(Traffic, Rain, Umbrella)

Trivial decomposition:

Conditional independence:

❖ **Bayes nets (graphical models) help us compactly express conditional independence assumptions/relationships.**

Bayesian Networks

Example: Coin flips

Make a Bayes net

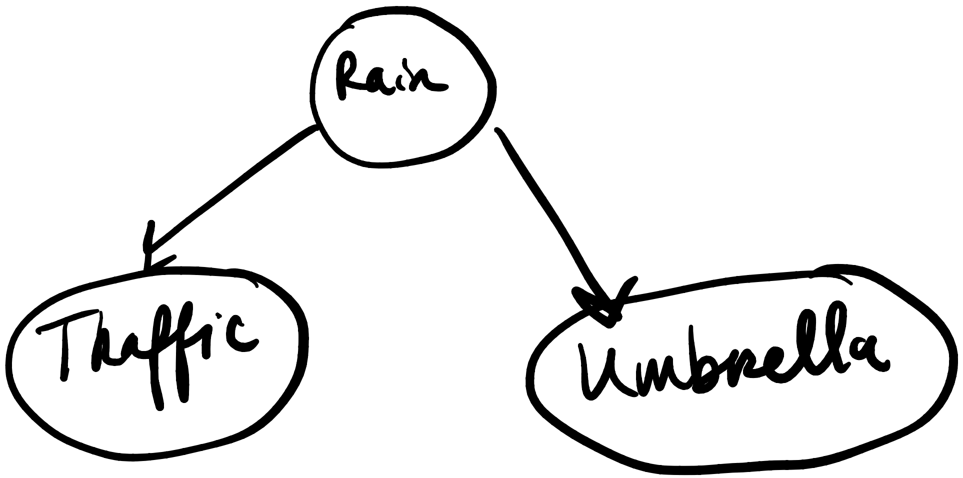
Suppose we flip a coin 3 times

$P(F_1 = H) = 1/2$

no edges indicates indepence of the nodes ,

Bayesian Networks

Example: ~~Coin flips~~ Traffic, Rain, Umbrella



This structure indicates that $T \perp\!\!\!\perp U \mid R$

Reminder:

X and *Y* are **independent** if $\forall x \, \forall y \; P(x, y) = P(x)\, P(y)$

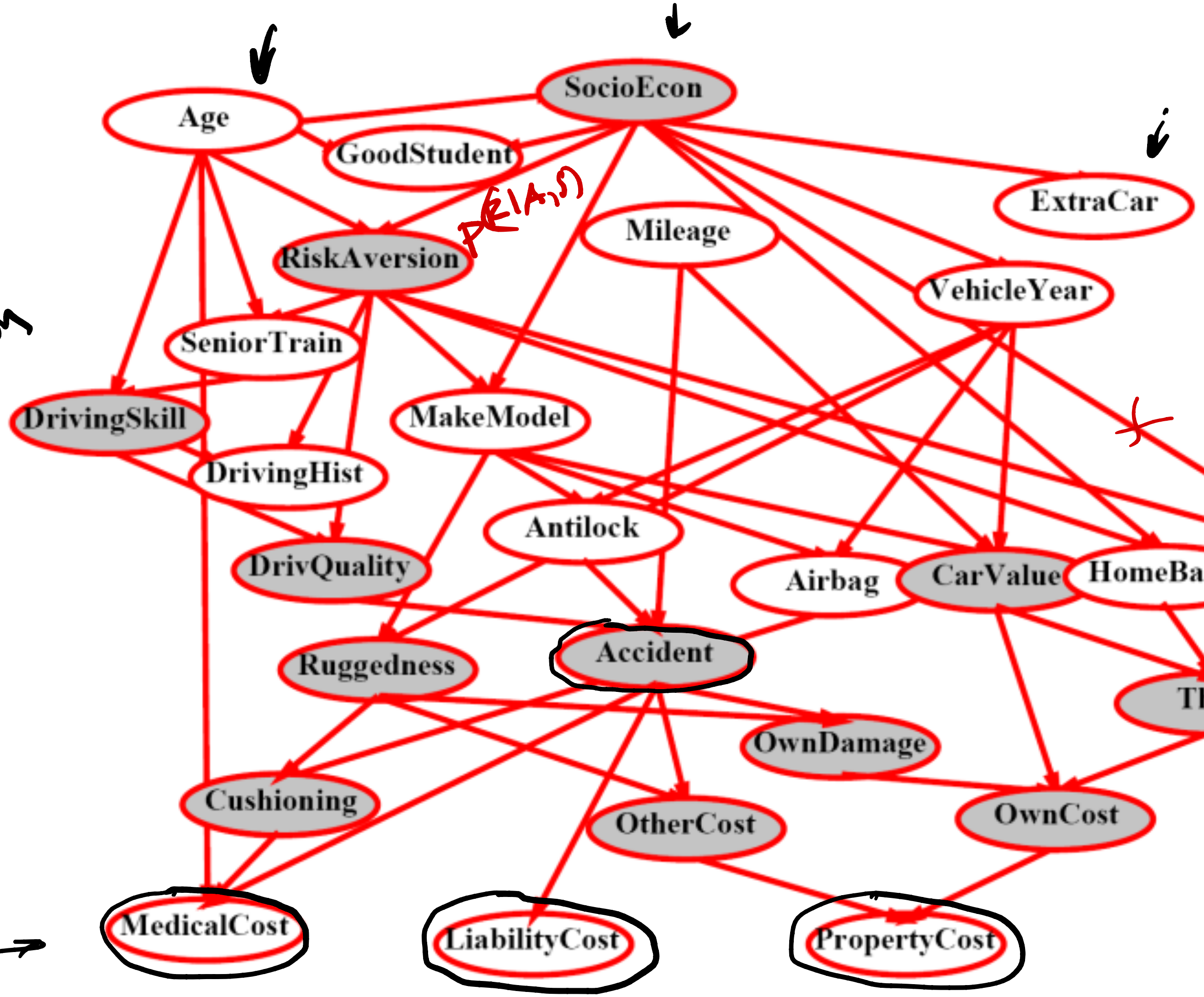
$$X \perp\!\!\!\perp Y$$

X and *Y* are **conditionally independent given *Z*** if $\forall x \, \forall y \, \forall z \; P(x, y \mid z) = P(x \mid z)\, P(y \mid z)$

$$X \perp\!\!\!\perp Y \mid Z$$

Graphical Model Notation

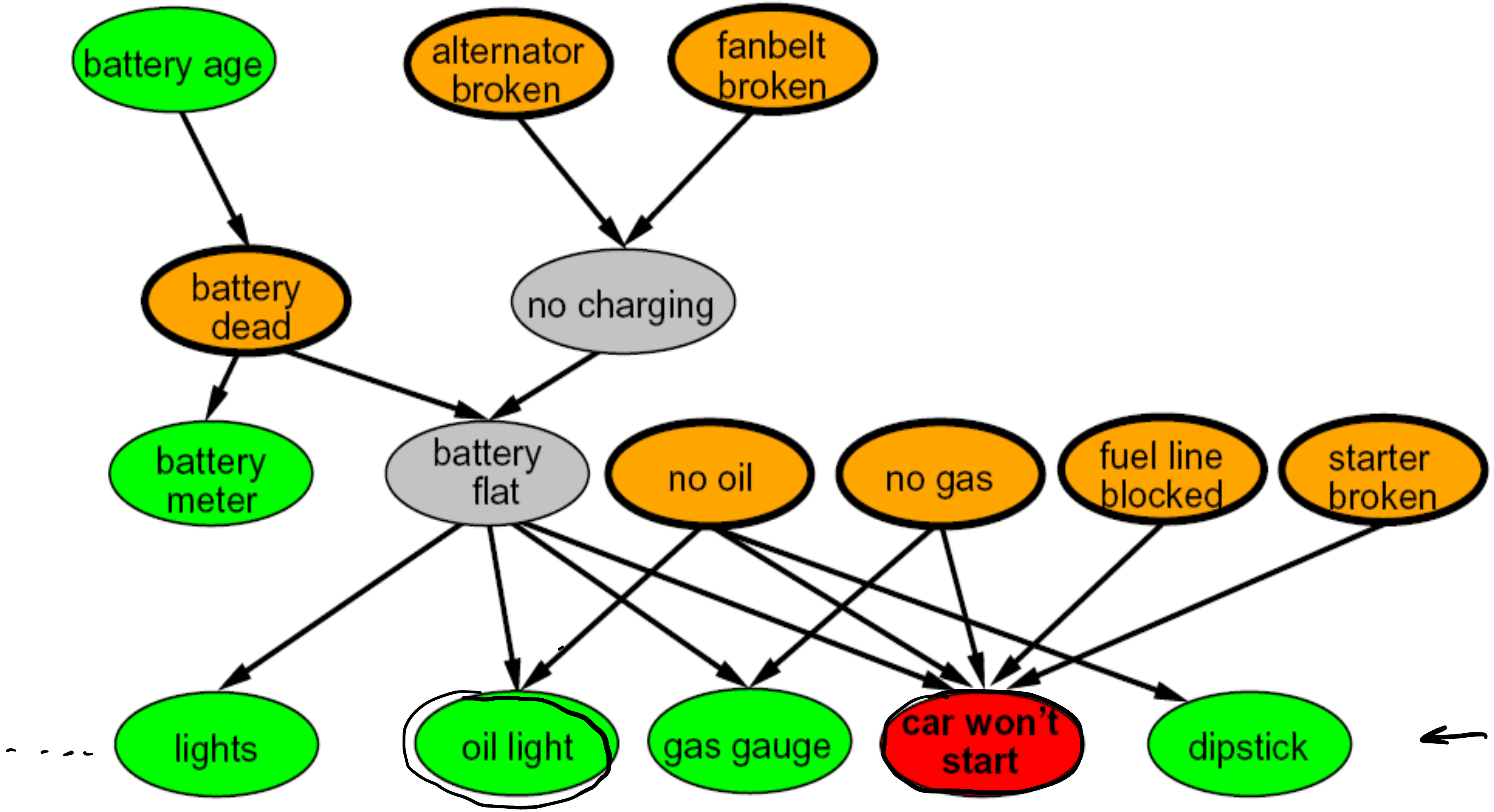
Downward Example of using a Bayesian network to make predictions



P(Antitheft)

Socioeconomic

Graphical Model Notation



Diagnostic Reading of a Bayes net

Bayesian Networks

Bayes nets implicitly encode joint distributions as a product of the local conditional distributions:

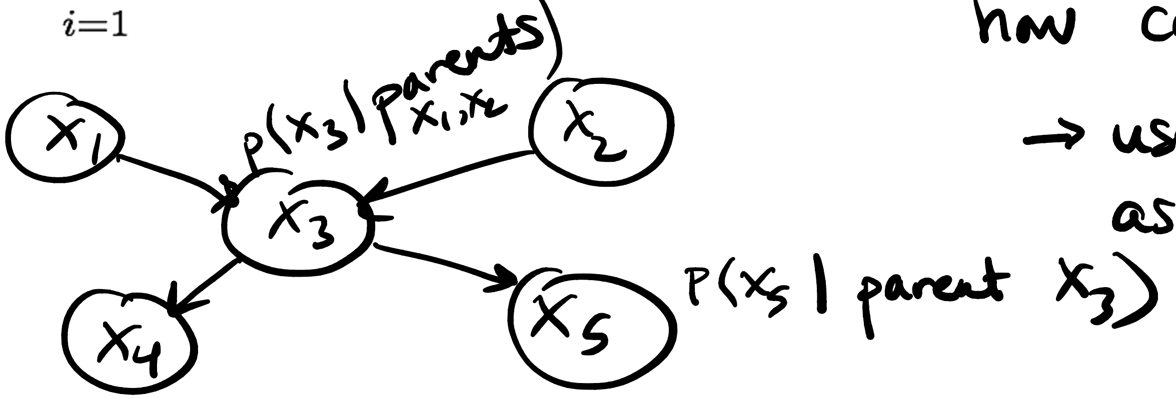
$P(x_1, x_2, \dots, x_n) = ?$

$$
\begin{aligned}
P(x_1, x_2, \dots, x_n) &= P(x_n \mid x_{n-1}, x_{n-2}, \dots, x_2, x_1) P(x_{n-1}, x_{n-2}, \dots, x_2, x_1) \\
&= P(x_n \mid x_{n-1}, x_{n-2}, \dots, x_2, x_1) P(x_{n-1} \mid x_{n-2}, \dots, x_2, x_1) P(x_{n-2}, \dots, x_2, x_1) \\
&= \dots \\
&= P(x_n \mid x_{n-1}, x_{n-2}, \dots, x_2, x_1) P(x_{n-1} \mid x_{n-2}, \dots, x_2, x_1) \dots P(x_3 \mid x_2, x_1) P(x_2 \mid x_1) P(x_1) \\
&= \prod_{i=1}^{n} P(x_i \mid x_{i-1}, x_{i-2}, \dots, x_2, x_1)
\end{aligned}
$$

how can we simplify?

→ use conditional independence assumptions

e.g.



Bayesian Networks

Bayes nets implicitly encode joint distributions as a product of the local conditional distributions:

$$P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid x_{i-1}, x_{i-2}, \ldots, x_2, x_1) = \ldots ?$$

Node ordering: write in such a way that

$$\text{parents}(X_i) \subseteq \{X_{i-1}, X_{i-2}, \ldots, X_2, X_1\}$$

$\rightarrow$

$$\prod_{i=1}^{n} P(x_i \mid x_{i-1}, x_{i-2}, \ldots, x_2, x_1) = \prod_{i=1}^{n} P(x_i \mid \text{parents}(X_i))$$

This statement:

$$P(X_i \mid X_{i-1}, X_{i-2}, \ldots, X_2, X_1) = P(X_i \mid \text{parents}(X_i))$$

is key: **each node is conditionally independent of its other predecessors, given its parents**

Bayesian Networks

Bayes nets implicitly encode joint distributions as a product of the local conditional distributions:

$$P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid \text{parents}(X_i))$$

↓ warmup

Example: Traffic

P(+*r*, ¬*t*) = ?

$P(R)$

+r	1/4
¬r	3/4

$P(T|R)$

+r →

+t	3/4
¬t	1/4

¬r →

+t	1/2
¬t	1/2

Next Time

- Bayesian Inference and Sampling