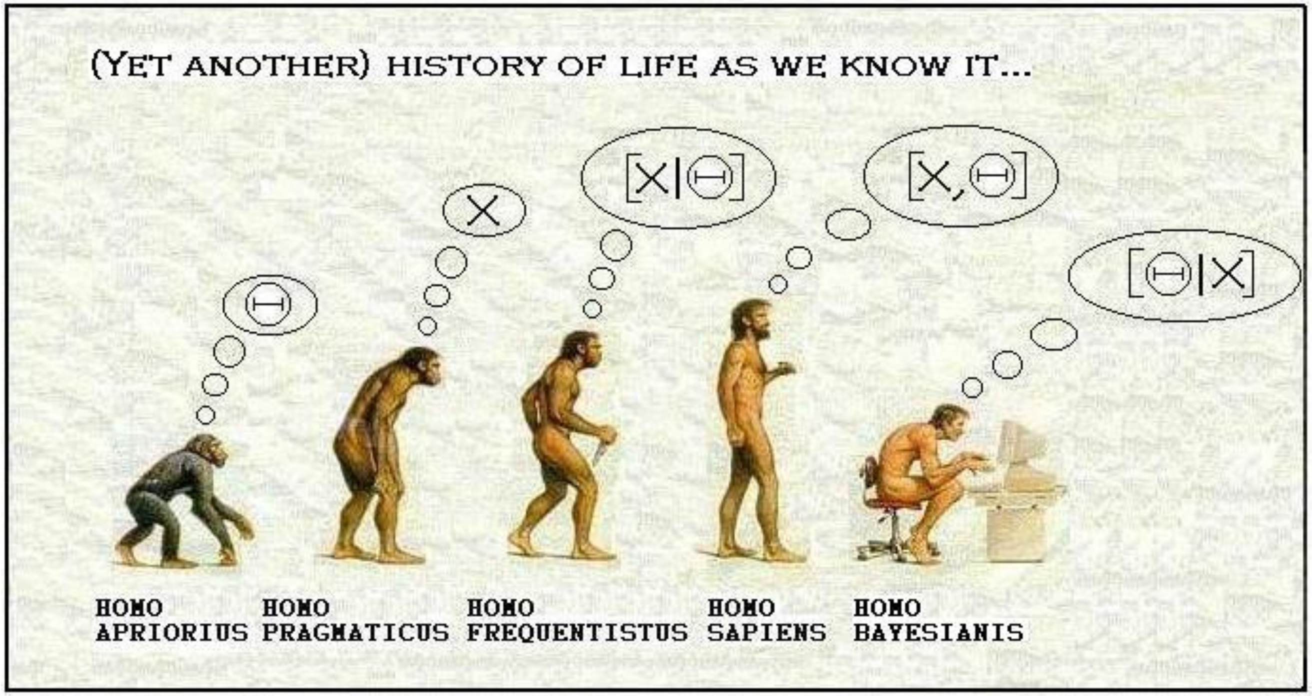
CSE 473

Lecture 21
(Chapter 14)

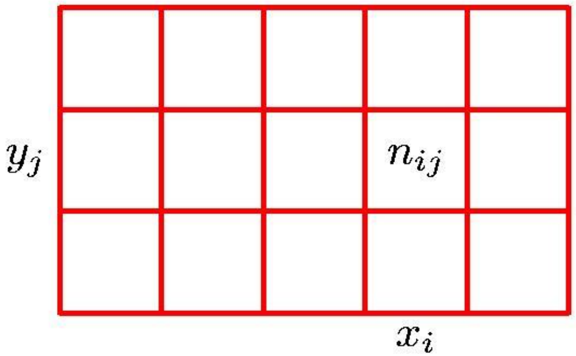
Bayesian Networks



(Courtesy Mike West)

Joint Probability

$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$

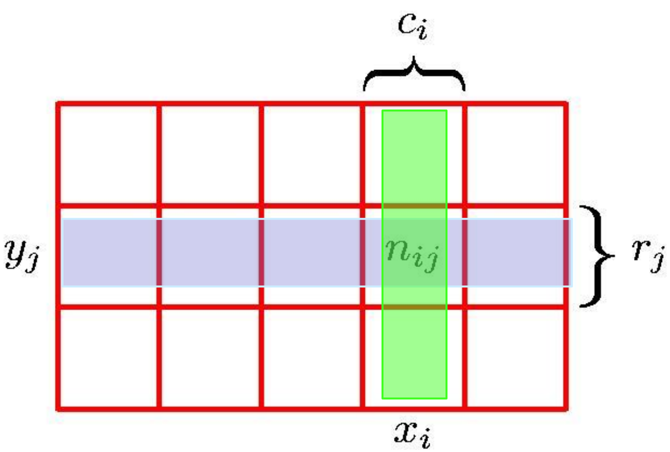


Total number of events = *N*

Marginal Probability

$$P(X = x_i) = \sum_j P(x_i, y_j) = \frac{c_i}{N}$$

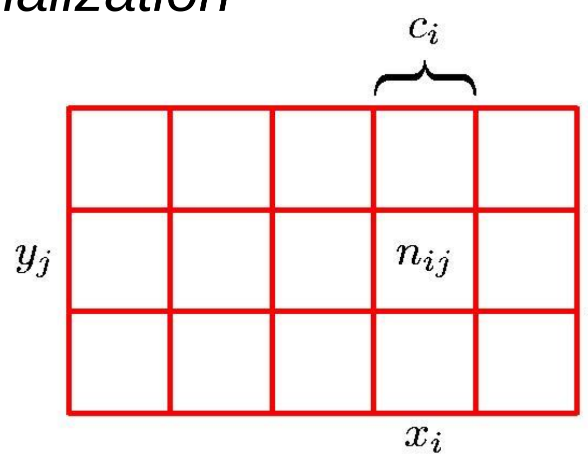
$$P(Y = y_j) = \sum_i P(x_i, y_i) = \frac{r_j}{N}$$



Summing out a variable is called *marginalization*

Conditional Probability

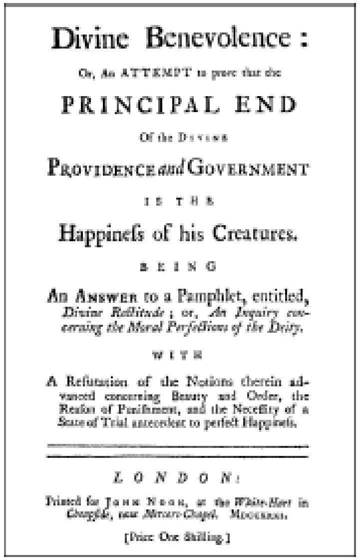
$$p(Y = y_j | X = x_i) = \frac{n_{ij}}{c_i}$$



Thomas Bayes

Reverand Thomas Bayes
Nonconformist minister
(1702-1761)

- Publications:
- *Divine Benevolence, or an Attempt to Prove That the Principal End of the Divine Providence and Government is the Happiness of His Creatures* (1731)
- *An Introduction to the Doctrine of Fluxions* (1736)
- *An Essay Towards Solving a Problem in the Doctrine of Chances* (1764)


Divine Benevolence:
Or, An ATTEMPT to prove that the
PRINCIPAL END
Of the DIVINE
PROVIDENCE and GOVERNMENT
IS THE
Happineſs of his Creatures.
BEING
An ANSWER to a Pamphlet, entitled, *Divine Rectitude; or, An Inquiry concerning the Moral Perfections of the Deity.*
WITH
A Refutation of the Notions therein advanced concerning Beauty and Order, the Reaſon of Puniſhment, and the Neceſſity of a State of Trial antecedent to perfect Happineſs.
LONDON:
Printed for JOHN NOON, at the *White-Hart* in *Cheapſide*, near *Mercers-Chapel*.
[Price One Shilling.]

Start with Definition of Conditional Probability

$$P(x \mid y) = \frac{P(x,y)}{P(y)}$$

$$P(y \mid x) = \frac{P(y,x)}{P(x)} = \frac{P(x,y)}{P(x)}$$

Therefore?

Bayes' Rule

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

$$i.e.$$

$$P(x \mid y) = \frac{P(y \mid x)\ P(x)}{P(y)}$$

What is this useful for?

$$P(Cause|Effect) = \frac{P(Effect|Cause)P(Cause)}{P(Effect)}$$

Bayes' rule is used to Compute *Diagnostic* Probability from *Causal* Probability

$$P(Cause|Effect) = \frac{P(Effect|Cause)P(Cause)}{P(Effect)}$$

E.g. let M be meningitis, S be stiff neck

P(M) = 0.0001,
P(S) = 0.1,
P(S|M)= 0.8 (note: these can be estimated from patients)

$$P(M \mid S) = \frac{P(S \mid M)P(M)}{P(S)} = \frac{0.8 \times 0.0001}{0.1} = 0.0008$$

Note: posterior probability of meningitis still very small!

(But chance of M did increase from 0.0001 to 0.0008 given stiff neck)

Normalization in Bayes' Rule

$$P(x \mid y) = \frac{P(y \mid x)\ P(x)}{P(y)} = \alpha\, P(y \mid x)\, P(x)$$

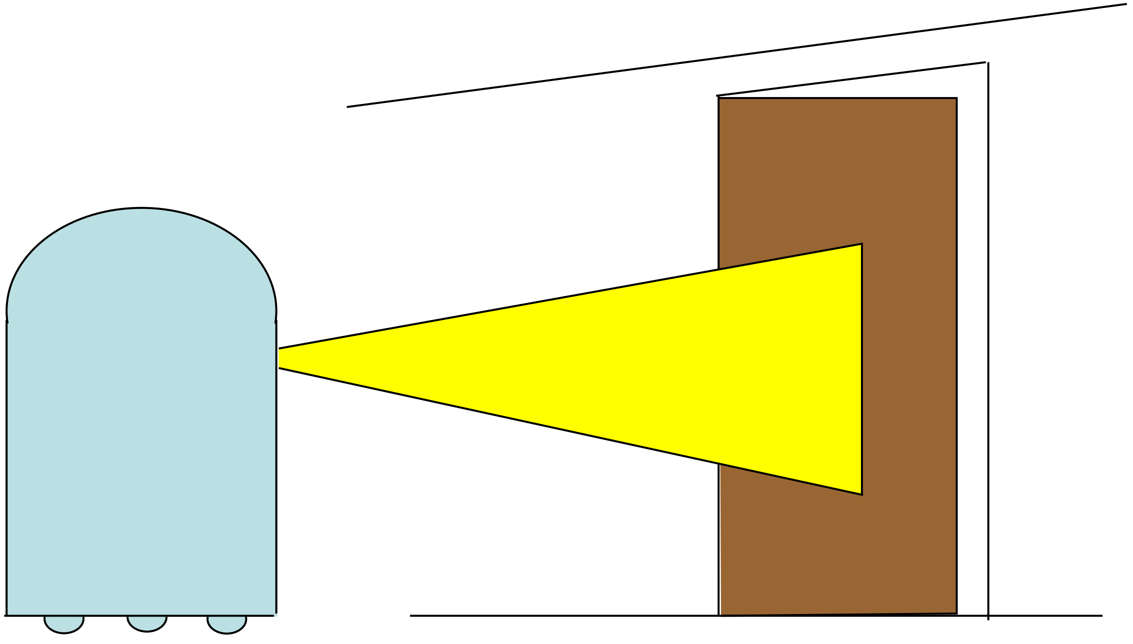
$$\alpha = \frac{1}{P(y)} =$$

α is called the normalization constant

(can be calculated by summing over numerator values)

Bayes Example 1: State Estimation

- Suppose a robot obtains measurement z
- What is $P(doorOpen|z)$?
- Use Bayes' rule!



Causal vs. Diagnostic Reasoning

- $P(open|z)$ is diagnostic.
- $P(z|open)$ is causal.
- Often causal knowledge is easier to obtain.

count frequencies!

- Bayes rule allows us to use causal knowledge to diagnose a situation:

$$P(open \mid z) = \frac{P(z \mid open)P(open)}{P(z)}$$

State Estimation Example

- *Suppose:* $P(z|open) = 0.6$ $P(z|\neg open) = 0.3$
- $P(open) = P(\neg open) = 0.5$
- $P(open|z) = ?$

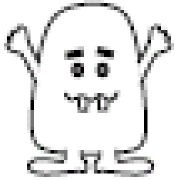
$$P(open \mid z) = \frac{P(z \mid open) P(open)}{P(z \mid open) p(open) + P(z \mid \neg open) p(\neg open)}$$

$$P(open \mid z) = \frac{0.6 \cdot 0.5}{0.6 \cdot 0.5 + 0.3 \cdot 0.5} = \frac{0.30}{0.45} = 0.67$$

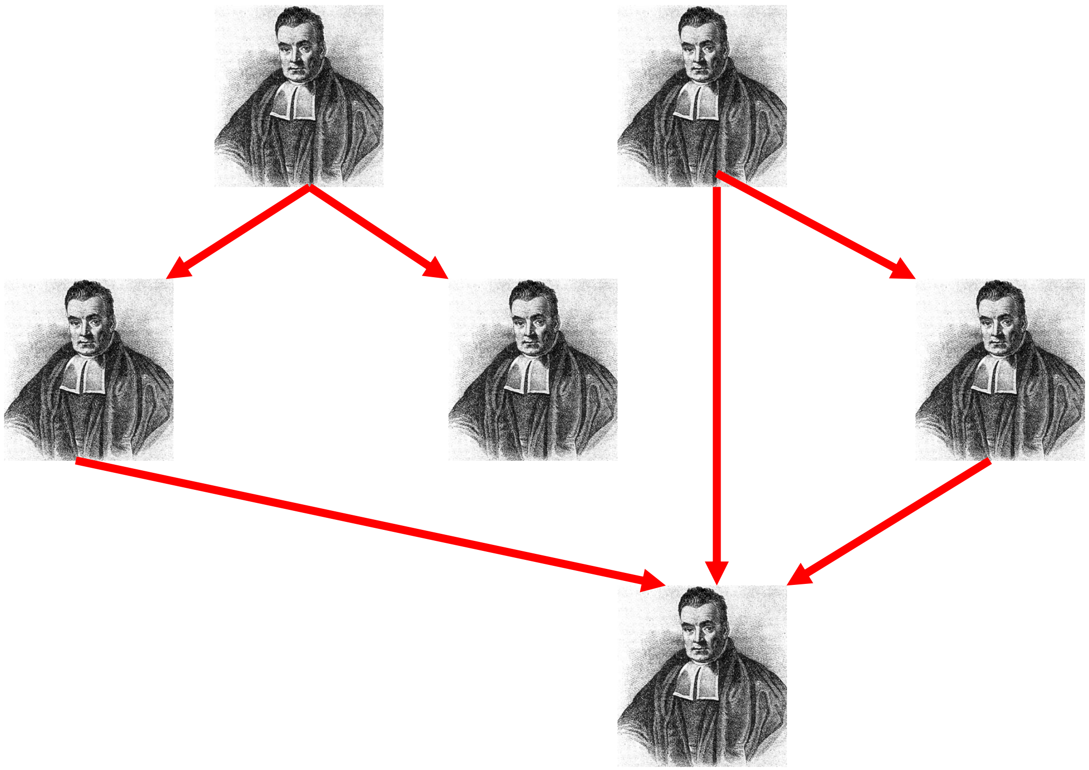
Measurement z raises the probability that the door is open from 0.5 to 0.67

Is there a general representation scheme for efficient probabilistic inference?

Yes!



Enter…Bayesian networks

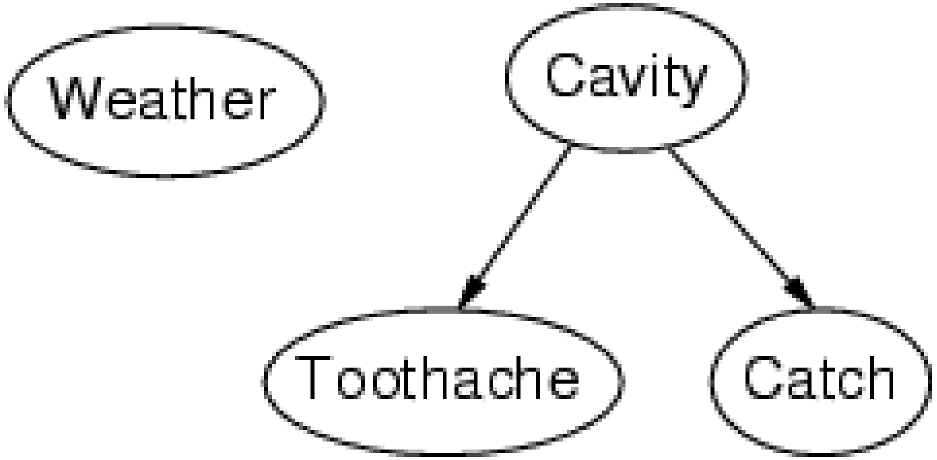


What are Bayesian networks?

- Simple, graphical notation for conditional independence assertions
  - Allows compact specification of full joint distributions

Example: Back at the Dentist's

- Topology of network encodes conditional independence assertions:



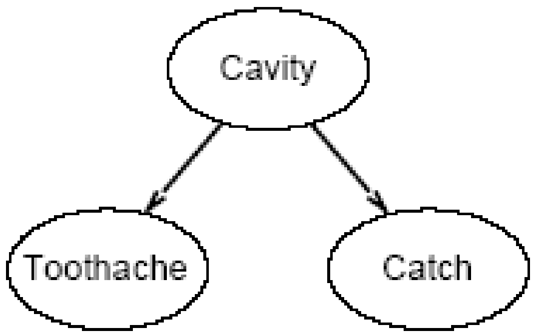
- *Weather* is independent of the other variables
- *Toothache* and *Catch* are <u>conditionally independent</u> of each other given *Cavity*

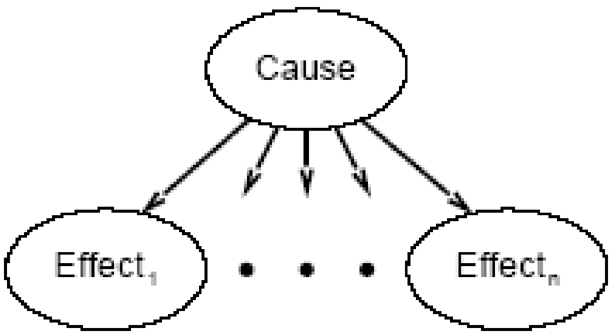
Conditional Independence and the "Naïve Bayes Model"

$$\mathbf{P}(Cavity|toothache \wedge catch)$$
$$= \alpha\, \mathbf{P}(toothache \wedge catch|Cavity)\mathbf{P}(Cavity)$$
$$= \alpha\, \mathbf{P}(toothache|Cavity)\mathbf{P}(catch|Cavity)\mathbf{P}(Cavity)$$

This is an example of a *naive Bayes* model:

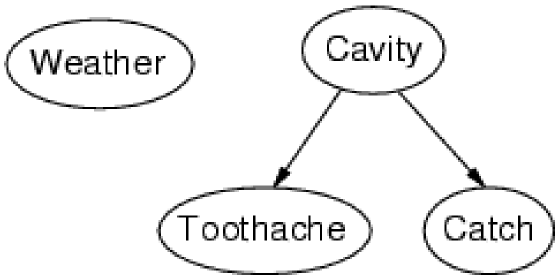
$$\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause)\Pi_i \mathbf{P}(Effect_i|Cause)$$





Total number of parameters is *linear* in n

Bayesian networks



- Syntax:
  - set of nodes, one per random variable
  - directed, acyclic graph (link ≈ "directly influences")
  - conditional distribution for each node given its parents:
    $P(X_i \mid Parents(X_i))$

- For discrete variables X_i, conditional distribution = conditional probability table (CPT) = probabilities for X_i given each combination of parent values

Example 2: Burglars and Earthquakes

- You are at a “Done with the AI class” party.
- Neighbor John calls to say your home alarm has gone off (but neighbor Mary doesn't).
- Sometimes your alarm is set off by minor earthquakes.
- Question: Is your home being burglarized?

Next Time

- Bayesian Networks for Burglary Detection and More!
- Inference Algorithms
  - Variable Elimination (VE)
- Hidden Markov Models
- To Do:
  - Project 3 due Sunday before midnight