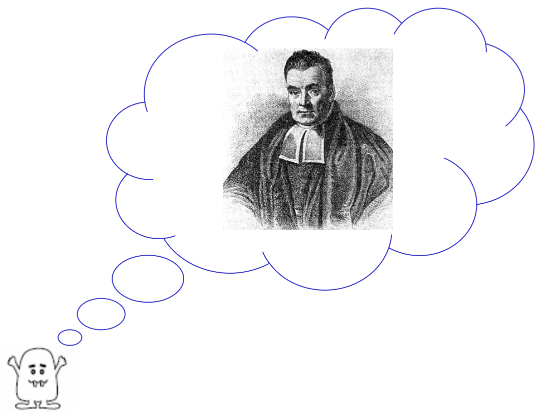
CSE 473

Chapter 13
Bayes' Rule & Bayesian Inference

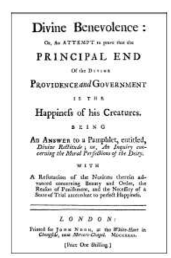


Thomas Bayes

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:

PRINCIPAL END

PROVIDENCE and GOVERNMENT

Happiness of his Creatures.

BEING

An Answer to a Pamphlet, entitled,

WITH

LONDON:

Recall: Conditional Probability

- P(x | y) is the probability of x given y
- Assumes that y is the only info known.
- Defined as:

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

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

Therefore?

Bayes' Rule

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

$$\Rightarrow$$

$$P(x \mid y) = \frac{P(y \mid x)\ P(x)}{P(y)} = \frac{\text{likelihood} \cdot \text{prior}}{\text{evidence}}$$

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

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

Note: posterior probability of meningitis still very small!

Normalization in Bayes' Rule

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

$$\alpha = P(y)^{-1} = \frac{1}{\sum_x P(y \mid x) P(x)}$$

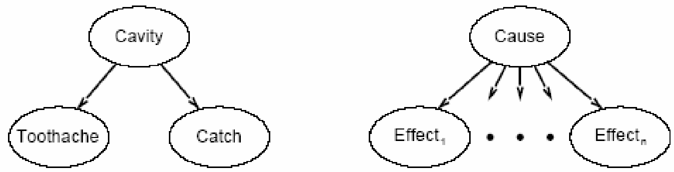
α is called the normalization constant

Cond. Independence and Naïve Bayes Model

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

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

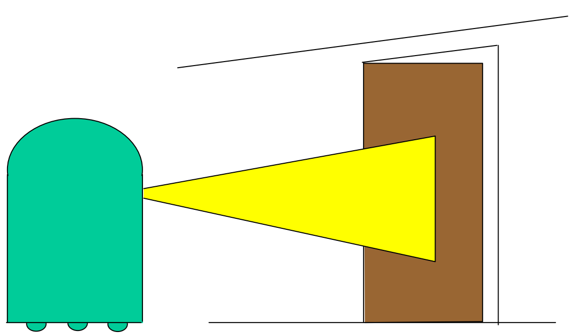
$$\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause)\Pi_i \mathbf{P}(Effect_i|Cause)$$



Total number of parameters is *linear* in n

Example 1: State Estimation

- Suppose a robot obtains measurement z
- What is *P(doorOpen|z)?*



Causal vs. Diagnostic Reasoning

- *P(open/z)* is diagnostic.
- *P(z/open)* is causal.
- Often causal knowledge is easier to obtain.
- Bayes rule allows us to use causal knowledge:

count frequencies!

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

State Estimation Example

- $P(z|open) = 0.6$ $\qquad$ $P(z|\neg open) = 0.3$
- $P(open) = P(\neg open) = 0.5$

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

$$P(open \mid z) = \frac{0.6 \cdot 0.5}{0.6 \cdot 0.5 + 0.3 \cdot 0.5} = \frac{2}{3} = 0.67$$

Measurement z raises the probability that the door is open.

Combining Evidence

- Suppose our robot obtains another observation z_2.
- How can we integrate this new information?
- More generally, how can we estimate $P(x| z_1...z_n)$?

Recursive Bayesian Updating

$$P(x \mid z_1,\ldots,z_n) = \frac{P(z_n \mid x, z_1,\ldots,z_{n-1})\, P(x \mid z_1,\ldots,z_{n-1})}{P(z_n \mid z_1,\ldots,z_{n-1})}$$

Markov assumption: z_n is independent of $z_1,\ldots,z_{n-1}$ if we know x.

$$P(x \mid z_1,\ldots,z_n) = \frac{P(z_n \mid x, z_1,\ldots,z_{n-1})\, P(x \mid z_1,\ldots,z_{n-1})}{P(z_n \mid z_1,\ldots,z_{n-1})}$$

$$= \frac{P(z_n \mid x)\, P(x \mid z_1,\ldots,z_{n-1})}{P(z_n \mid z_1,\ldots,z_{n-1})}$$

$$= \alpha\, P(z_n \mid x)\, P(x \mid z_1,\ldots,z_{n-1})$$

Recursive!

Incorporating a Second Measurement

- $P(z_2|open) = 0.5$ $P(z_2|\neg open) = 0.6$
- $P(open|z_1) = 2/3 = 0.67$

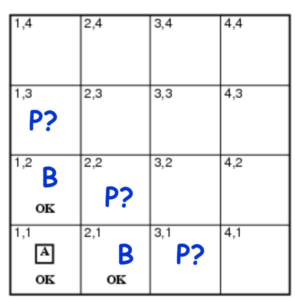
$$P(open \mid z_2, z_1) = \frac{P(z_2 \mid open)\, P(open \mid z_1)}{P(z_2 \mid open)\, P(open \mid z_1) + P(z_2 \mid \neg open)\, P(\neg open \mid z_1)}$$

$$= \frac{\frac{1}{2}\cdot\frac{2}{3}}{\frac{1}{2}\cdot\frac{2}{3}+\frac{3}{5}\cdot\frac{1}{3}} = \frac{5}{8} = 0.625$$

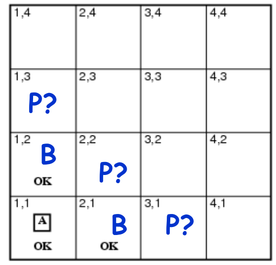
- z_2 lowers the probability that the door is open.

Example 2: Wumpus World

- Reduced wumpus world: only pits and breezes
- Squares adjacent to pits have breezes
- $p_{i,j}$ = square [i,j] contains a pit
- $b_{i,j}$ = there is a breeze in square [i,j]
- Probability of a square containing a pit = 0.2
- Known = [1,1], [2,1] and [1,2] contain no pit
  - known = $\neg p_{1,1} \wedge \neg p_{2,1} \wedge \neg p_{1,2}$
- Breeze in [1,2] and [2,1]
  - b = $\neg b_{1,1} \wedge b_{2,1} \wedge b_{1,2}$

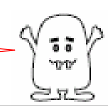


Is there a pit in [1,3], [2,2] or [3,1]?



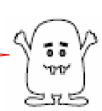
- Logical inference gives no definite answer ⇒ Must choose random action
- Can do better with probabilistic inference
- $\mathbf{P}(P_{1,3}|\text{known},b)$
  $= \alpha \sum_{\text{unknown}} \mathbf{P}(P_{1,3},\text{known},\text{unknown},b)$
  $= \alpha \sum_{\text{unknown}} \mathbf{P}(b|P_{1,3},\text{known},\text{unknown})\ \mathbf{P}(P_{1,3},\text{known},\text{unknown})$
- Can be simplified using:
  independence of pit occurrences across squares and
  conditional independence of breeze from other squares given neighboring squares
- We get (see text):
  $\mathbf{P}(P_{1,3}|\text{known},b) \approx \langle 0.31, 0.69 \rangle$
  $\mathbf{P}(P_{2,2}|\text{known},b) \approx \langle 0.86, 0.14 \rangle$

Avoid [2,2]!



These calculations seem laborious to do for each problem domain – is there a general representation scheme for probabilistic inference?

Yes!



Enter…Bayesian networks
(next time)