# Bayesian Modeling

Q550: Models in Cognitive Science

# Modeling Probabilistic Inference

- Probability was originally developed as a means for studying games of chance and making optimum gambling decisions
- Scientific formalizations of probability are a fairly recent development
- Jeffrey: Prior to the 17th C, probable meant *approvable*; a probable action or opinion was one that sensible people would hold (given the circumstances)
- More recently, we have used probability as a formal account of human decision making  (e.g., Gigerenzer et al., 1989)

# Bounded Rationality

- The assumption is that humans use bouned-rational decision strategies
  - We attempt to optimize our decisions under constraints imposed by human information processing limitations
  - Frequentist probability (Neyman, Venn, Fisher, etc.)

- For this, we make good use of **Bayes Theorem**
- Assume that human induction/learning/decision-making is based on a Bayes engine: “Belief is goverened by the laws of probability”

# Bayes Theorem

- Bayes Theorem is a way to update beliefs in light of new evidence (*a posteriori*)
- Humans are rational inference machines…we are testing hypotheses based on evidence from the environment
- Assume you are trying to infer the process that was responsible for generating observable data (**D**). **H** is the hypothesis about the true generating process.

- Assume you are trying to infer the process that was responsible for generating observable data (**D**). **H** is the hypothesis about the true generating process.

By Bayes Theorem:

$$P(H \mid D) = \frac{P(D \mid H)P(H)}{P(D)}$$

- P(H|D) = *posterior* (Hyp given Data)
- P(H) = *prior*
- P(H|D) = (Data given Hyp) *conditional* or *likelihood*
- P(D) = *marginal* or *normalizing constant* (ensures prob is normalized to sum one)

The posterior is proportional to the product of the prior probability and the likelihood

## An example (from Griffiths & Yuille, 2006):

- A box contains two coins: One that produces heads 50% of the time and one that produces heads 90% of the time
- You pick a coin and flip it 10 times; how do you modify your hypothesis based on the data?

Let $\theta$ = P(heads)

We need to identify the hypothesis space ***H***, the prior distribution, P(*H*), and the likelihood P(*D*|*H*)

$H_0$ is $\theta = 0.5$ and $H_1$ is $\theta = 0.9$.  $P(H_0) = P(H_1) = 0.5$

The probability of a particular sequence is binomial:

$$P(D \mid \theta) = \theta^{N_H}\left(1 - \theta^{N_T}\right)$$

With only two hypotheses, it is convenient to just work with the *posterior odds* (ratio of evidence for $H_0$ over $H_1$):

$$\frac{P(H_1 | D)}{P(H_0 | D)} = \frac{P(D | H_1)}{P(D | H_0)} \frac{P(H_1)}{P(H_0)}$$

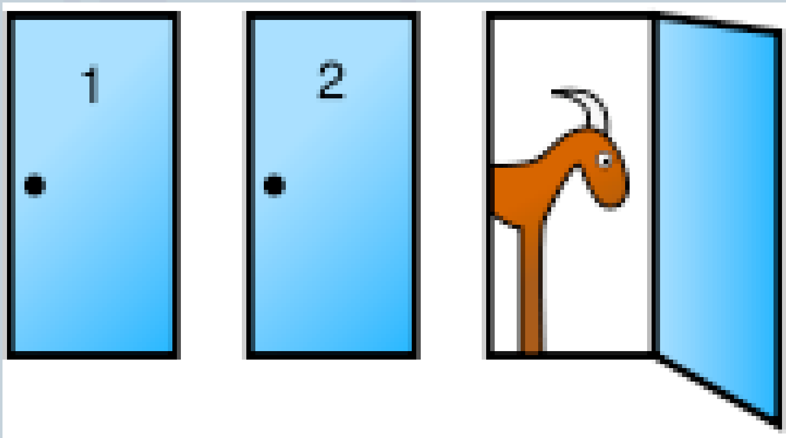
This is exactly as we did for model selection with the BIC

Sequence of data: HHHHHHHHHH = 357:1 in favor of $H_1$

Sequence of data: HHTHTHTTHT = 165:1 in favor of $H_0$

# Back to the Goat Game

- Recall our earlier "Goat Game" (aka, Monte Hall 3-door problem).




Hypothesis that car is behind a door: $H_1$, $H_2$, $H_3$

$$P(H_1) = P(H_2) = P(H_3) = 1/3$$

- Assume that we pick door #1
- Let D = “host opens door #2” Without prior knowledge, we would assign this event p = .5
- We don’t even need to see what’s behind door #2. If the car was there, he had to pick door 3. If the car was behind door #3, he had to pick door #2. Thus, given this evidence (*D*), we know:

$$P(D \mid H_2) = 0$$

$$P(D \mid H_3) = 1$$

$$P(D \mid H_1) = 1/2$$

$$P(D \mid H_2) = 0$$

$$P(D \mid H_3) = 1$$

$$P(D \mid H_1) = 1/2$$

Thus, you should always switch from the original door you picked:

$$P(H_1 \mid D) = \frac{P(D \mid H_1)P(H_1)}{P(D)} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}$$

$$P(H_2 \mid D) = \frac{P(D \mid H_2)P(H_2)}{P(D)} = \frac{0 \cdot \frac{1}{3}}{\frac{1}{2}} = 0$$

$$P(H_3 \mid D) = \frac{P(D \mid H_3)P(H_3)}{P(D)} = \frac{1 \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$

# Bounded Rationality

- Of course, most humans you'll meet think the stick and switch strategies are equally optimal
  - Thus, humans are **not** optimal Bayesian decision makers, but they still may be within the restrictions of *bounded rationality* (cf. Tversky & Kahneman's decision heuristics)
- Choosing between infinitely many hypotheses (e.g., language learning)…we get a probability density over a random variable, and can use the maximum or mean posterior from the distribution (see Griffiths & Yuille tutorial on probabilistic inference)
- Markov models and Bayes nets (graphical distributions)

# Deal or No Deal?

- Task: Create a Bayesian model that optimizes hypotheses about the amount in cases based on the behavior of the banker over trials…can you predict optimum choices?
- Distribution of payouts
- Synthetic RPS competition
- Who wants to be a millionaire w/ semantic models

# Examples of Bayesian Models:

**Tenenbaum, J. B. (1999). Bayesian modeling of human concept learning. *NIPS***

- Humans are able to learn efficiently from a small number of positive examples, and can successfully generalize without experiencing negative examples (e.g., word meanings, perceptual categories)
- Poverty of the stimulus problem
- By contrast, ML techniques require both positive and negative examples to generalize at all (and many examples of each to generalize successfully)

# Examples of Bayesian Models:

- Bayesian inference provides an excellent model of how humans induce and generalize structure in their environments
  - They succeed where many other models fail at learning flexible tasks as humans do.

**Other Examples:**

- Perfors, Tenenbaum, & Regier (2006)
- Anything by Dan Navarro
- The Topics model (Grif, Steyv, & Ten, in press, *PR*)
- Chater & Manning (2006)
- Kruschke (2006)
- For a good overview, see Tenenbaum, Griffiths, & Kemp (2006) *TICS*

## An application of modeling: Mike Mozer’s crazy adaptive house