What is Bayesian Inference?

Saturday, September 01, 2012 2:02 PM

- In Political Science, "Bayesian inference" encompasses many interrelated but distinct ideas
  - An epistemological viewpoint about the relationship between evidence and conclusions
  - A mathematical argument about the logically correct beliefs one should hold given prior information (information external to the data set) and information contained in a data set
  - A set of technical and computational tools for determining this belief distribution in complex situations

Bayes' Rule (*)

* statistical action

- We will examine each of these ideas in turn, focusing on the first two today

Bayesian vs. Frequentist epistemology

Saturday, September 01, 2012 3:04 PM

- Probably neither Bayesian nor frequentist ideas are developed enough to constitute a full epistemology or philosophy of science, but they do have starting points that are very different from one another

- Frequentists presume that there is a knowable state of the world that corresponds to some estimated parameter or quantity in a model

  bias

  $\beta \to \hat{\beta}$

  - uncertainty in estimates (presuming proper specification, etc.) comes from noise in the data generating process that obscures the signal
  - Probability distributions produced are probabilities of this estimate occurring given some state of the world and an (estimated) degree of noise in the world

    MLE: max

    $f(\text{data} \mid \hat{\beta})$

    - Ex: a left-tail *p*-value for an estimated OLS coefficient is $\Pr(\hat{\beta} | \beta = 0)$

    $\Pr(\hat{\beta} \mid \beta = 0)$

- Bayesians are more interested in the connection between evidence and a pattern of *beliefs* than in evidence and the *true state of the world*, which is unknown and possibly unknowable

  - Idea: given what I believed before this study, and the likelihood of what I saw in evidence given my beliefs, what should I believe after this study?

  - uncertainty in estimates can come from multiple places, including multiple states of the world, true noise/error, and incomplete evidence

Likelihood vs. Posterior

Saturday, September 01, 2012 4:26 PM

- Likelihood: the probability that a particular data set (or estimate from a data set) occurs given the state of the world

- e.g.: suppose we flip a coin *N* times, and the coin has an unknown *theta* probability of landing on heads. The likelihood that the coin lands on heads *k* times given *theta* is:
  - $\Pr(k|\theta) = \binom{N}{k}\theta^k(1-\theta)^{N-k}$ $\theta = pr(heads)$ ${}_NC_k$
  - Frequentists focus on this quantity; e.g., a "maximum likelihood" analysis would input N and k and then use numerical optimization to find the θ that made $\Pr(k|\theta)$ as large as possible $\hat{\theta}$ = =
- This is typically NOT the quantity that researchers/consumers of research are interested in

  - Wouldn't we want to know $\Pr(\theta|k)$, the probability that theta is some particular value given the number of heads?

  - e.g., in application we might want to know the probability that $\Pr(\theta > 0.5|k)$ and/or $\Pr(\theta < 0.5|k)$ -- whether the coin is fair

- In Bayesian terminology, beliefs are probability distributions over a set of states of the world (parameters, estimated quantities, hypotheses, etc.)

- The point of analysis is to determine the beliefs that a researcher should rationally hold after having seen the data set, so-called *posterior beliefs*

  - State of the world = θ

  - Discrete θ: belief is $\Pr(\theta = \theta_0|data)$

  - Continuous θ: belief that $\theta \in \Theta$ is $\int_\Theta f(\theta|data)$

- How do we do that?

Bayes' Rule

Saturday, September 01, 2012 4:50 PM

- There is a mathematical relationship between a likelihood and a posterior belief: Bayes' Rule

- Abstract variant: let A and B be distinct events. Then:

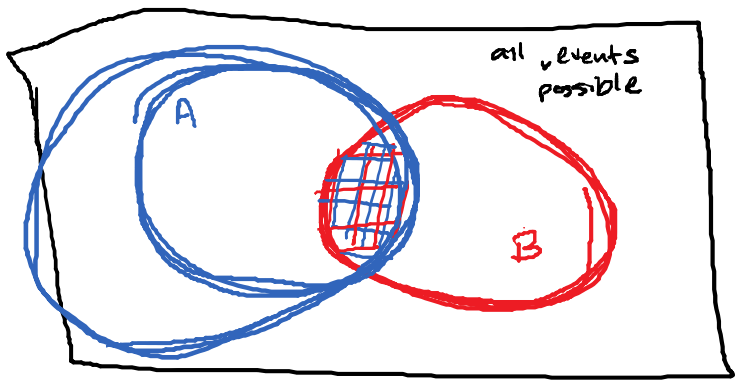
$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)}$$

This is easy to illustrate with a simple Venn diagram. It can also be proved if we accept definitionally that:

$$\Pr(A|B)\Pr(B) = \Pr(A \wedge B)$$

That is, the probability that A and B both occur is the probability that B occurs times the probability that A occurs given that B has already occurred.

$$pr(A|B) = \frac{pr(B|A)\cdot pr(A)}{pr(B)}$$



- This extends easily to multiple events:

$$\Pr(A_i|B) = \frac{\Pr(B|A_i)\Pr(A_i)}{\sum_j \Pr(B|A_j)\Pr(A_j)}$$

$A_i \in A$

$pr(B)$

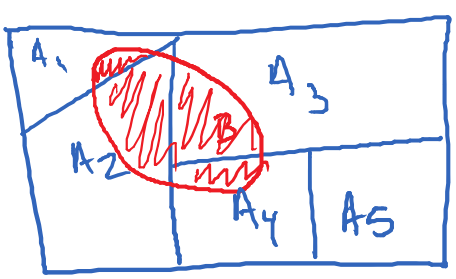
$$pr(A \wedge B) = pr(B|A)\cdot pr(A)$$
$$= pr(A|B)\cdot pr(B)$$

basic ideas of probability

- Or to continuous parameters:

$$f(\theta|B) = \frac{f(B|\theta)f(\theta)}{\int f(B|\theta)f(\theta)d\theta}$$

$= pr(B)$



$$pr(B|A)\cdot pr(A) = pr(A|B)\cdot pr(B)$$

Bayes' rule.

$$pr(B|A) = \frac{pr(A|B)\cdot pr(B)}{pr(A)}.$$

Three Parts of a Posterior

Saturday, September 01, 2012 5:16 PM

- In looking at Bayes' rule, there are three components to a Bayesian posterior belief distribution:

likelihood that the dataset occurs given θ

data

$$f(\theta|B) = \frac{f(B|\theta)f(\theta)}{\int f(B|\theta)f(\theta)d\theta}$$

prior belief about θ

normalizing constant

- The denominator is a normalizing constant (the same for every value of θ) and can be neglected or calculated numerically post hoc in some situations. This fact is reflected in the writing of:

$$f(\theta|B) \propto f(B|\theta)f(\theta)$$

Where $\propto$ means "is proportional to"

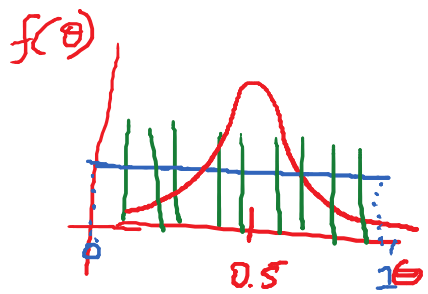
the posterior beliefs about θ are proportional to the likelihood and the prior

- Example: coin flip with *k* flips and *n* successes
  - Likelihood (less proportional constant)

$$\binom{N}{k}\theta^k(1-\theta)^{N-k}$$

  - Prior (in some cases without proportional constant)

$$f(\theta) = ?$$



  - Combine the two, and normalize to force the sum to be a proper density

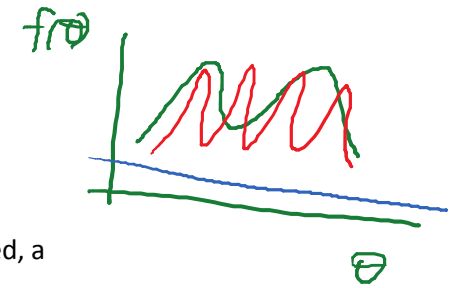
- Try it in R!

- Sensitivity analysis: how do our conclusions change under different priors?

- Priors can come from the literature!

Conjugate Priors

Saturday, September 01, 2012 7:25 PM



- Computational tools allow us to use exotic priors and likelihoods
- A strength of Bayesian analysis: flexibility
  - As long as the likelihood and prior functions can be computed, a posterior probability can be determined
- In some cases, however, the prior and likelihood are simple enough -- and fit together in such a way -- that the form of the posterior can be analytically known

Conjugate priors?

- Example: the binomial distribution is conjugate with the beta distribution
  - Binomial: $f(k|\theta) = \binom{N}{k}\theta^k(1-\theta)^{N-k}$
  - Beta distribution: $f(\theta|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$
  - Beta prior * binomial likelihood:
    $f(\theta|k,\alpha,\beta) \propto \binom{N}{k}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{k+\alpha-1}(1-\theta)^{N-k+\beta-1}$

    or $f(\theta|k,\alpha,\beta) \propto \theta^{k+\alpha-1}(1-\theta)^{N-k+\beta-1}$
  - Thus, α and β factor into the binomial likelihood as a number of "trials" that precede the observations of the instant sample data set
- Try it in R!

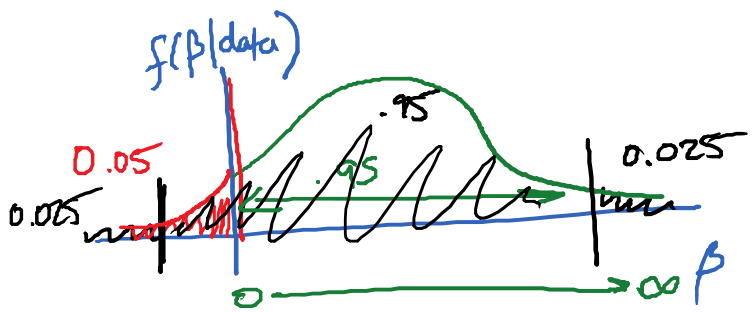
Hypothesis Testing: Credible Regions

Sunday, September 02, 2012 5:52 PM

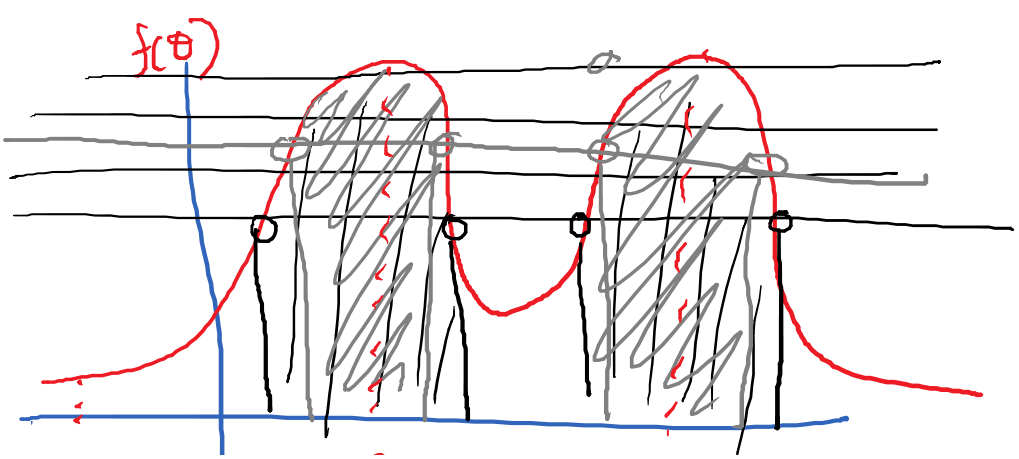
- Hypothesis testing: establish decision criteria for drawing conclusions about quantities of interest
- Focus directly on quantities of importance $\Pr(\theta \in \Theta | data)$ *
- Can import ideas from frequentist statistics
  - Example: conclude that some parameter β is greater than zero if $\int_{-\infty}^{0} f(\beta|data)d(\beta) < 0.05$
  - Integrals of the posterior density that cover $(1-\alpha)$ proportion of the density = "$(1-\alpha)$ credible regions"
  - If these integrals cover the highest-density regions, these are in some sense optimal credible regions

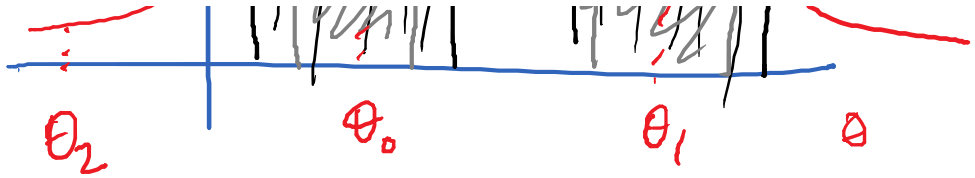
theory: $dy/dx > 0 \longrightarrow f(\text{data} \mid \beta = 0)$?

$\Pr(\beta > 0 \mid \text{data})$



90% CI




θ_2
θ_0
θ_1
θ

Back to Model Comparison: Bayes' Factors

Sunday, September 02, 2012 5:46 PM

- The Bayesian perspective on inference leads us to consider an alternative model comparison statistic
- Idea: we want to know whether $\Pr(model_0|data)$ is bigger than $\Pr(model_1|data)$
  - If so, accept model 0
  - If not, accept model 1
  - Might want to favor one model over another for various reasons
- Leads to the statement of the Bayes' Factor
  - $\frac{\Pr(M_1|y)}{\Pr(M_0|y)} * \frac{\Pr(M_0)}{\Pr(M_1)} = \frac{\Pr(y|M_1)}{\Pr(y|M_0)}$
  - A measure of the extent to which evidence has changed beliefs in favor of Model 1 over Model 0
  - "Proof":

$$\frac{\frac{pr(M_1|y)}{pr(M_0|y)}}{\frac{Pr(M_1)}{Pr(M_0)}}$$

posterior density ratios

prior density ratios

$$pr(y|M_1) = \frac{pr(M_1|y) \cdot pr(y)}{pr(M_1)}$$

$$pr(y|M_0) = \frac{pr(M_0|y) \cdot pr(y)}{pr(M_0)}.$$

$$\frac{pr(M_1|y)}{pr(M_0|y)} = \frac{\frac{pr(y|M_1) \cdot pr(M_1)}{\cancel{pr(y)}}}{\frac{pr(y|M_0) \cdot pr(M_0)}{\cancel{pr(y)}}}$$

$$= \frac{pr(y|M_1)}{pr(y|M_0)} \cdot \frac{pr(M_1)}{pr(M_0)}$$

$$\frac{pr(M_1|y)}{pr(M_0|y)} \cdot \frac{pr(M_0)}{pr(M_1)} = \frac{pr(y|M_1)}{pr(y|M_0)}$$

$$\ln\left(\frac{A}{B}\right) = \ln A - \ln B$$

$$\ln\left[\frac{pr(y|M_1)}{pr(y|M_0)}\right] = \ln\left(pr(y|M_1)\right) - \ln\left(pr(y|M_0)\right)$$