Intro to Bayesian Inference[1]

Tyler Ransom
Duke University

July 3, 2012

[1]Based on Greene Chapter 18

Outline

- Introduce Bayesian approach
- Compare/Contrast Bayesian vs. Frequentist approaches
- Discuss Random Coefficients

Bayesian Approach

- Consider a simple problem, viewed through the lens of both classical statistics and Bayesian statistics (adapted from Wikipedia):
- There are two bags of marbles. Bag #1 has 10 red and 10 blue. Bag #2 has 20 red and 10 blue.
- **Classical statistics**: What is the probability that a marble drawn at random from one of the bags is blue?
- **Bayesian statistics**: Given that the marble drawn is blue, what is the probability that it was drawn from Bag #1?

Bayesian Approach

- The classical approach would say something like "assume each bag can be drawn from with equal probability," and in the end we recover a probability of drawing a blue marble
- The Bayesian approach would say something like "assume each bag can be drawn from with equal probability," (a prior belief) and in the end, we recover a posterior probability that the draw came from Bag #1

Bayesian MLE

Recall Bayes' rule:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

We can rewrite this as

$$P(\text{parameters} \mid \text{data}) = \frac{P(\text{data} \mid \text{parameters})P(\text{parameters})}{P(\text{data})}$$

or, dividing out $P(\text{data})$ (which is uninformative for our purposes)

$$P(\text{parameters} \mid \text{data}) \propto P(\text{data} \mid \text{parameters})P(\text{parameters})$$

or

$$\text{posterior density} = \text{likelihood function} \times \text{prior density}$$

Bayesian vs. Classical Approach

- Bayesian proponents say that estimation is not deducing values of fixed parameters, but rather sharpening subjective beliefs about the state of the world
- But if this subjectivity is uncomfortable because a researcher prefers "black and white" truths, then classical method may be preferable
- Basically, each approach has pros and cons

Another Example

Another example to illustrate the Bayesian way of thinking: Consider the problem of predicting the temperature on a certain day of the year

- Classical statistics would possibly estimate an OLS model of temperature on a variety of covariates and then generate a prediction
- Bayesian statistics would create a prior distribution based on (e.g. the historical average of the temperatures on that day) and then estimate a posterior distribution

Pros and Cons of Bayesian Approach

Pros

- Can find efficient solutions to complex estimation problems (e.g. MCMC, which is quite popular)
- Doesn't tie the researcher to a single "black and white" result
- Recognizes that research process is sharpening of beliefs, not finding the one right answer

Cons

- More difficult to interpret results
- Requires more parameterization than classical MLE
- Posterior is influenced by the prior (so if your prior is bad, posterior may also be bad)

Random coefficients

- Random coefficients refers to a modeling technique where each individual is assumed to have a different β, e.g. $\beta_i \sim N(\beta, \Sigma)$
- Since the Bayesian approach assumes a distribution of the parameter vector, it is tempting to equate the posterior distribution with the distribution of random coefficients
- The two methods are quite different, and although both involve a distribution, they should not be confused

Random Coefficients vs. Unobserved Heterogeneity

When should I use cluster-corrected standard errors, and when should I use robust standard errors?

- We previously talked about heterogeneity in unobservables, which allows us to recover “correct” theta estimates
- Introducing heterogeneity into the parameter estimates is another method of allowing for “correct” estimates; it’s just harder to interpret when you get a different coefficient vector for each person
- Now need to make distributional assumptions on theta AND the error term

Random Coefficients: Example

This example comes from Train (pp. 137-138)

- Suppose the utility for individual i for choice j is $u_{ij} = X_{ij}\beta_i + \varepsilon_{ij}$
- The choice probability P_{ij} is now an integral over the (continuous) distribution of β:

$$P_{ij} = \int \left(\frac{\exp(X_{ij}\beta)}{\sum_k \exp(X_{ik}\beta)} \right) f(\beta) d\beta$$

- So our likelihood function is now

$$\ell = \sum_i \sum_j y_{ij} \ln \left(P_{ij} \right)$$

where the maximization goes over the parameters of the distribution (e.g. β and Σ)