Introduction to Bayesian Inference

(Really Laplacian Inference)

Introduction

- You (should have) already reviewed videos on:
  - Conditional probability.
  - Bayes theorem.
  - The Monte Hall problem.
- Distinction between Bayesian and classical (or frequentist) inference.
- Ingredients of Bayesian inference.
- Examples of Bayesian inference.
  - US penny flipping.
  - Correlation in bivariate normal distribution.
  - Bayesian linear regression model assuming normal priors.

The Buzz

"Now Bayesian statistics are rippling through everything from physics to cancer research, ecology to psychology. Enthusiasts say they are allowing scientists to solve problems that would have been considered impossible just 20 years ago. And lately, they have been thrust into an intense debate over the reliability of research results."

The New York Times, Sept. 14, 2014.

Why?

- General increase in the demand for data science.
  - Clinical trials.
  - Survey research.
  - Policy evaluation.
- Computational speed.
- Logical inconsistencies in the classical approach to inference.

A Brief History of Time

- Laplace, Legendre, de Morgan, and Gauss (1700’s through 1850’s) lay the foundations of probability and statistics.
  - Rev. Bayes’ essay published posthumously in 1760’s.
  - Laplace reformulates a mathematically-coherent version of inverse probability.
- Fisher, Neyman, and the Pearsons develop the classical interpretation of statistical inference (early 1900’s).
  - Fisher is quite hostile to Bayesian interpretation, calling it an “impenetrable jungle”.
  - As recently as 20 years ago, Bayesians were denounced as “unmarried marriage guidance counselors”.

The Distinction

- Classical approach to inference treats the unknown parameters (such as the β's in the linear model) as fixed things to be estimated.
  - Evaluate hypotheses and confidence intervals as if we had "repeated samples".
- Bayesian approach treats the β's probabilistically.
  - Prior probability: researcher's beliefs regarding the β's prior to the observation of any data.
  - Posterior probability: researcher's updated beliefs based on data.

Equivalent Statements

- Beliefs after seeing data ∝ Beliefs before seeing data * Chance we would see these particular data under different beliefs
- $P(\text{world}|\text{data}) \propto P(\text{world}) * P(\text{data}|\text{world})$
- $P(\beta|\text{data}) \propto P(\beta) * P(\text{data}|\beta)$
- Posterior Probability ∝ Prior Probability * Likelihood
- Posterior ∝ Prior * Likelihood

Basic Steps in Bayesian Inference

- Specify a probability model that includes some prior knowledge or belief about the β's. This is the prior probability (or "priors").
- Update the priors by conditioning them on observed data.
  - Posterior ∝ Prior * Likelihood
- Evaluate the fit of the model to the data.

Conceptual Experiment: Flipping a US Penny

Goal: Make inferences about whether the coin is “balanced”

Classical Approach

- Seek to test the null hypothesis that Pr(heads) = 0.5 by establishing a rejection region for the null hypothesis (or a t-stat in excess of ~2).
- Flip a penny 100 times.
- Calculate the sample proportion as: (number of heads)/100.
- Suppose the outcome is 53 heads: one would fail to reject the null because the t-stat is ~0.6.
- Suppose the outcome is 80 heads: one would reject the null because the t-stat is ~7.5.

Bayesian Approach

- Start with prior beliefs about the world, in this case the coin itself.
- Let’s make this practical and create our prior probability.
  - Flat (or uniform) priors.
  - Tails much more likely than heads.
  - Heads much more likely than tails.
  - Balanced coin.

Bayesian Approach (cont.)

- Instruct the computer to flip a coin 10 times and note the outcomes (the data or the likelihood).
  - We can generate the likelihood of the observed outcomes using a Binomial random variable.
- Based on the observed outcomes, I alter my prior probability to create a posterior probability.

Bayesian Approach (cont.)

- Examine how our prior probability interacts with the observed data from the coin flips to form our posterior probability.
- (NB: This example is easy because the posterior probability has a closed form that we can analyze visually.)

Let’s Do This in R

Flat Prior

- We start with a flat prior regarding the penny.
- We observe the data as represented in the likelihood.
- We develop our posterior by pulling probability mass away from where the likelihood is small and adding it to where the likelihood is large.

Strong Prior of Heads

- Suppose we believed that the coin was not balanced and was much more likely to land heads.
- Accordingly, we start with a strong prior regarding heads.
- Likelihood of observed outcomes tells us we're wrong.
- We update our posterior by moving away from strong prior of heads.

Strong Prior of Tails

- Suppose we believed that the coin was much more likely to land tails.
- Accordingly, we start with a strong prior regarding tails.
- Likelihood of observed outcomes tells us we're ***VERY*** wrong.
- We update our posterior by ***strongly*** moving away from strong prior of tails.

Prior that Penny Is Balanced

- Suppose we instead believe generally that the coin is balanced and form a normal prior centered around 0.5.
- Likelihood of observed outcomes leads us to update our posterior away from equally likely outcomes.

Distinctions: Classical v. Bayesian

- Probability is observed as the result of an "infinite" series of trials.
- Data are a random sample from a fixed DGP, and the parameters are fixed.

- Probability is the observer's degree of belief.
- Data are observed and therefore fixed by sample generated, and the parameters described probabilistically.

Distinctions: Classical v. Bayesian

- Point estimates and standard errors. 95% confidence intervals that cover the true parameter value in “19 out of 20 repeated samples.”
- Classical hypothesis testing using rejection regions (or t-stats).

- Summaries from posterior, such as averages and percentiles. Confidence internals are called “credible intervals”.
- Statements regarding parameters are probabilistic: Likelihood P(heads) >= .5.

Criticism of Bayesian Approach: Your Priors Are Subjective

- Yes, they are subjective but explicitly stated.
- Besides:
  - Posterior is dominated by observed data (the likelihood) and not by the prior.
  - Bayesian interpretation has no logical inconsistencies (unlike the classical approach).
  - Posterior can serve as a prior in another experiment with new data: a form of learning.

So What Do Bayesians Want?

- Formal estimation of a Bayesian model is the most difficult undertaking in Bayesian analysis.
- Bayesian approach is to summarize <u>the entire posterior distribution</u> in order to make probabilistic statements about it.
  – Achieved through Monte Carlo simulation (which I will show).
  – It’s the tough part.

A Note on Posterior Probabilities

- May be closed-form probability functions from which we can easily make probabilistic statements.
  - So-called conjugate posteriors that have the same form as the prior.
- They may, however, not have any closed-form, in which case we must use sampling.
  - Markov Chain Monte Carlo (MCMC) sampling using the algorithms of Gibbs and Metropolis-Hastings.
  - Example using bivariate normal to show you what's going on.

Simple Example of Bayesian Inference

- Examine the correlation of bivariate normal distribution.
  - [Wikipedia: Bivariate Normal Distribution]
  - Use Metropolis-Hastings MCMC to sample posterior.
- Key Bayesian elements.
  - Average of posterior.
  - 95% credible interval based on posterior.
  - Hypothesis test that correlation is >= 0.5.
- Think about the contrast with classical approach.

The Bayesian Linear Model

- We can use CAPM to estimate a Bayesian linear model.
- At the end of the day, we will have meaningful probabilistic statements about the world that can influence decision making.
- Highlights the issues that arise in frequentist hypothesis testing.
  - A 95% confidence interval will cover the "true parameter value" in 19 out of 20 repeated samples.