STAT J535: Introduction

David B. Hitchcock
E-Mail: hitchcock@stat.sc.edu

Spring 2012

Chapter 1: Introduction to Bayesian Data Analysis

- **Bayesian** statistical inference uses **Bayes' Law** (Bayes' Theorem) to combine prior information and sample data to make conclusions about a parameter of interest.
- Bayesian inference differs from classical inference in that it specifies a **probability distribution** for the parameter(s) of interest.
- Why use Bayesian methods? Some reasons:
  1. We wish to specifically incorporate previous knowledge we have about a parameter of interest.
  2. To logically update our knowledge about the parameter after observing sample data
  3. To make formal probability statements about the parameter of interest.
  4. To specify model assumptions and check model quality and sensitivity to these assumptions in a straightforward way.

- Why do people use classical methods?
  1. If the parameter(s) of interest is/are truly fixed (without the possibility of changing), as is possible in a highly controlled experiment
  2. If there is no prior information available about the parameter(s)
  3. If they prefer "cookbook"-type formulas with little input from the scientist/researcher
- Many reasons classical methods are more common than Bayesian methods are historical:
  1. Many methods were developed in the context of controlled experiments.
  2. Bayesian methods require a bit more mathematical formalism.
  3. Historically (but not now) realistic Bayesian analyses had been infeasible due to a lack of computing power.

Motivation for Bayesian Modeling

- Bayesians treat unobserved data and unknown parameters in similar ways.
- They describe each with a probability distribution.
- As their model, Bayesians specify:
  1. A **joint density function**, which describes the form of the distribution of the **full sample** of data (given the parameter values)
  2. A **prior distribution**, which describes the behavior of the parameter(s) **unconditional** on the data
- The prior could reflect:
  1. Uncertainty about a parameter that is actually fixed
     OR
  2. the variety of values that a truly stochastic parameter could take.

Exchangeability

- Bayesians usually assume the data values in the sample are **exchangeable**: that is, reordering the data values does not change the model.
- **Example**: In a social survey, respondents are asked whether they are generally happy. Let

$$Y_i = \begin{cases} 1 & \text{if respondent } i \text{ is happy} \\ 0 & \text{otherwise} \end{cases}$$

Exchangeability

- Consider the first 5 respondents. What are the probabilities of these 3 outcomes?

$$p(1,0,0,1,1) = ?$$
$$p(0,1,1,0,1) = ?$$
$$p(1,1,0,1,0) = ?$$

- If the data values are exchangeable, these three outcomes will have the same probability.

Exchangeability and iid

- **Theorem**: If the data are **independent** and **identically distributed** (iid), i.e., a random sample, and θ follows the distribution $p(\theta)$, then the data are exchangeable.
- **Proof**: Let $Y_1, \ldots, Y_n$ be iid given θ and let $\theta \sim p(\theta)$. Consider any permutation π of $\{1, \ldots, n\}$. Then for any $y_1, \ldots, y_n$:

Exchangeability and iid

$$\begin{aligned} p(y_1, \ldots, y_n) &= \int p(y_1, \ldots, y_n|\theta)p(\theta)\, d\theta \\ &= \int \left[\prod_{i=1}^{n} p(y_i|\theta)\right] p(\theta)\, d\theta \;\; \text{(since } Y_i \text{ iid)} \\ &= \int \left[\prod_{i=1}^{n} p(y_{\pi_i}|\theta)\right] p(\theta)\, d\theta \\ &\text{(since a product doesn't depend on order)} \\ &= p(y_{\pi_1}, \ldots, y_{\pi_n}). \end{aligned}$$

- $\Rightarrow Y_1, \ldots, Y_n$ are exchangeable.

Exchangeability and iid

- A famous theorem (de Finetti's Theorem) shows the converse is* **also** true:
  $Y_1, \ldots, Y_n$ are exchangeable for all n
  $\Rightarrow Y_1, \ldots, Y_n$ are iid given θ, $\theta \sim p(\theta)$.

* $=$ It is **usually** true: it's only approximate when sampling from a finite population without replacement.

Different Interpretations of Probability

1. **Frequentist** definition of the probability of an event: If we repeat an experiment a very large number of times, what is the proportion of times the event occurs?
    - **Problem**: For some situations, it is impossible to repeat (or even conceive of repeating) the experiment many times.
    - **Example**: The probability that President Obama is re-elected in 2012.
2. **Subjective probability**: Based on an individual's degree of belief that an event will occur.
    - **Example**: A bettor is willing to risk up to $200 betting that Obama will be re-elected, in order to win $100. The bettor's subjective $P[\text{Obama wins}]$ is $\frac{2}{3}$.
    - The Bayesian approach can naturally incorporate subjective probabilities about the parameter, where appropriate.

Some Probability Notation

- We denote events by letters such as $A, B, C, \ldots$
- The idea of **conditional probability** is crucial in Bayesian statistics:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- We denote random variables by letters such as X, Y, etc., taking on values denoted by x, y, etc.
- The space of all possible values of the r.v. is called its **support**.

Some Probability Notation

- We will deal with both **discrete** and **continuous** r.v.'s.
- In general, let $p(\cdot)$ denote the probability distribution (p.m.f. or p.d.f.) of a r.v.
- Thus $p(X)$ is the **marginal** distribution of X and $p(X, Y)$ is the **joint** distribution of X and Y.
- If X, Y independent, then $p(X, Y) = p(X)p(Y)$.

Some Probability Notation

- The expected value of any function $h(X)$ of X is:

$$E[h(X)] = \begin{cases} \sum_{x \in \mathcal{X}} h(x)p(x) & \text{if } X \text{ is discrete} \\ \int_{\mathcal{X}} h(x)p(x)\,dx & \text{if } X \text{ is continuous} \end{cases}$$

  where $\mathcal{X}$ denotes the support.
- Typically the distribution of X depends on some parameter(s), say θ, so in fact $p(X) = p(X|\theta)$.

Bayes' Law

In its simplest form, with two events A and B, Bayes' Law relates the conditional probabilities $P(A|B)$ and $P(B|A)$.
Recall

$$P(A|B) = \frac{P(A,B)}{P(B)}$$

and

$$P(B|A) = \frac{P(B,A)}{P(A)} = \frac{P(A,B)}{P(A)}$$

Hence $P(A,B) = P(A|B)P(B) = P(B|A)P(A)$

$$\Rightarrow P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Similarly,

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