STAT 638: Introduction to Applied Bayesian Methods

Instructor: Jeff Hart

What is Bayesian statisics?

Bayesian statistics is a way of doing statistical inference by means of *Bayes rule.* It provides a rational means of updating *prior* beliefs in light of information contained in data.

Some attractive features of Bayesian methods:

- Parameter estimates have *good statistical properties.*
- Straightforward methodology for *predicting* future and missing data.
- Cohesive, well-prescribed means of *estimating, selecting and validating models.*

Contrast Between Frequentist and Bayesian Statistics

- Frequentist statistics: Uncertainty about, for example, parameter estimates is quantified by investigating how estimates vary from one to the next *in repeated sampling from the same population.*

- Bayesian statistics: Uncertainty is quantified by determining how much *prior opinions about parameter values change in light of the observed data.*

To a Bayesian, data sets which might have been, but were not, observed are irrelevant to making inferences about the unknown parameters. *The only data set of any relevance is the one that was actually observed.*

In contrast, in the frequentist approach, data sets which might have been observed (but were not) are extremely relevant since they are the basis of determining measures of uncertainty.

You hear people talk about *turning the Bayesian crank.*

This means that, once a model is formulated, *one knows exactly how to proceed in a Bayesian analysis.*

This is a very appealing aspect of Bayesian methodology.

In contrast, frequentist statistics is more ad hoc, in that one often has to search for an optimal procedure (confidence interval or test) when a new problem presents itself.

The term *Bayesian* derives from Thomas Bayes (1702-1761), who was a British mathematician and Presbyterian minister. Bayes introduced *Bayes' theorem*, upon which all of Bayesian statistics rests.

Bayes' theorem Let $A_1, \ldots, A_m$ be mutually exclusive and exhaustive events. (Exhaustive means $A_1 \cup \cdots \cup A_m = \mathcal{S}$, where $\mathcal{S}$ is the sample space.) For any event B such that $P(B) > 0$,

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^m P(B|A_i)P(A_i)}, \ j = 1, \ldots, m.$$

Proof By the definition of conditional probability,

$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)}. \qquad (1)$$

Again by the definition of conditional probability,

$$P(A_j \cap B) = P(B|A_j)P(A_j). \tag{2}$$

Also,

$$\begin{aligned} P(B) &= P(B \cap \mathcal{S}) \\ &= P[B \cap (A_1 \cup \cdots \cup A_m)] \\ &= P[(A_1 \cap B) \cup \cdots \cup (A_m \cap B)] \\ &= \sum_{i=1}^{m} P(A_i \cap B) \\ &= \sum_{i=1}^{m} P(B|A_i)P(A_i). \end{aligned} \tag{3}$$

Substituting (2) and (3) into (1) proves the result.

Bayes' Theorem Applied to Statistical Models

Suppose we are to observe a vector of data $\boldsymbol{Y}$, a realization of which is denoted $\boldsymbol{y}$. $\boldsymbol{Y}$ has a probability distribution depending upon an unknown vector of parameters $\boldsymbol{\theta}$, which we wish to infer.

We'll denote the probability distribution of $\boldsymbol{Y}$ by $p(\boldsymbol{y}|\boldsymbol{\theta})$, emphasizing that this is the distribution of $\boldsymbol{Y}$ conditional on $\boldsymbol{\theta}$ being the true value of the unknown parameter vector.

The *prior* distribution of $\boldsymbol{\theta}$ is a probability distribution that represents the experimenter's beliefs about the unknown parameters prior to observing the data. This distribution will be denoted $p(\boldsymbol{\theta})$.

Suppose that $\boldsymbol{\theta}$ is a continuous vector and p is a probability density. Then p has the interpretation that

$$\int_A p(\boldsymbol{\theta})\, d\boldsymbol{\theta}$$

represents the experimenter's degree of belief that the true parameter lies somewhere in the region A.

Bayes' theorem applied to statistical model

$$\begin{aligned} p(\boldsymbol{\theta}|\boldsymbol{y}) &= \frac{p(\boldsymbol{y},\boldsymbol{\theta})}{m(\boldsymbol{y})} \\ &= \frac{p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int_\Theta p(\boldsymbol{y}|\boldsymbol{t})p(\boldsymbol{t})\, d\boldsymbol{t}}, \end{aligned}$$

where Θ is the parameter space, i.e., the set of all possible values for $\boldsymbol{\theta}$.

Summary of Terminology

p: The *prior* distribution.

$p(\boldsymbol{y}|\boldsymbol{\theta})$: The distribution of $\boldsymbol{Y}$ given $\boldsymbol{\theta}$. For given $\boldsymbol{y}$ and regarded as a function of $\boldsymbol{\theta}$, $p(\boldsymbol{y}|\boldsymbol{\theta})$ is also known as the *likelihood* function.

$p(\boldsymbol{\theta}|\boldsymbol{y})$: The *posterior* distribution.

m: The *marginal* distribution of $\boldsymbol{Y}$, or the *prior predictive* distribution.

The posterior distribution expresses the experimenter's updated beliefs about $\boldsymbol{\theta}$ in light of the observed data $\boldsymbol{y}$. The data may well change his opinions about the unknown parameters, and will usually sharpen them.

The marginal m is called the *prior predictive distribution* because (i) it is the unconditional distribution of $\boldsymbol{Y}$, and (ii) it may be used as an aid in predicting a value of $\boldsymbol{Y}$.

Let $\tilde{\boldsymbol{Y}}$ be a data vector that is yet to be observed. The *posterior predictive* distribution of $\tilde{\boldsymbol{Y}}$ given $\boldsymbol{Y} = \boldsymbol{y}$ is

$$m(\tilde{\boldsymbol{y}}|\boldsymbol{y}) = \frac{\int_{\Theta} p(\boldsymbol{y}, \tilde{\boldsymbol{y}}|\boldsymbol{\theta}) p(\boldsymbol{\theta})\, d\boldsymbol{\theta}}{m(\boldsymbol{y})}.$$

This distribution could be used to predict a value of $\tilde{\boldsymbol{Y}}$ given that $\boldsymbol{Y}$ is observed to be $\boldsymbol{y}$.

Subjectivity in Bayesian inference

A Bayesian analysis is *subjective* in that two different people may observe the same data $\boldsymbol{y}$ and yet arrive at different conclusions about $\boldsymbol{\theta}$. This can happen when the two people have different prior opinions about $\boldsymbol{\theta}$.

The subjectivity of the Bayesian paradigm is a source of controversy. A legitimate argument is that it seems “unscientific” for one’s personal prejudices to affect the conclusions of a scientific study.

Counters to the subjectivity criticism

- In “unscientific” situations, it seems natural that one’s conclusions will be affected by his/her prior opinions.

- When a large amount of data is available, the prior has little effect on the posterior, unless the prior is *extremely* sharp.

- A more objective approach in scientific situations is to use so-called *noninformative priors*. Such priors are meant to express ignorance about the unknown parameters. (We’ll discuss noninformative priors much more throughout the course.)