VectorBiTE Training 2018
Methods Workshop

Introduction to Bayesian Statistics

www.vectorbite.org

Learning Objectives

1. Understand the basic principles underlying Bayesian modeling methodology
2. Introduce how to use Bayesian inference for real-world problems
3. Introduce computation tools to perform inference for simple models in R (how to turn the Bayesian crank)
4. Appreciate the need for sensitivity analysis, model checking and comparison, and the potential dangers of Bayesian methods.

Recall: Bayes Theorem

Bayes Theorem allows us to relate the conditional probabilities of two events A and B:

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

What is Bayesian Inference?

In the Bayesian approach our probabilities numerically represent rational beliefs. Bayes rule provides a rational method for updating those beliefs in light of new information and incorporating/quantifying uncertainty in those beliefs.

Thus, Bayesian inference is an approach for understanding data inductively.

What is Bayesian Inference?

We can re-write Bayes rule in terms of our parameters, θ and our data, Y:

$$\Pr(\theta|Y) = \frac{\Pr(Y|\theta)\Pr(\theta)}{\Pr(Y)}$$

The LHS is the main quantity of interest in a Bayesian analysis, the posterior, denoted $f(\theta|Y)$:

$$\overbrace{f(\theta|Y)}^{\text{Posterior}} \propto \overbrace{\mathcal{L}(\theta; Y)}^{\text{Likelihood}} \times \overbrace{f(\theta)}^{\text{Prior}}$$

Bayesian methods provide

1. models for rational, quantitative learning
2. parameter estimates with good statistical properties
3. estimators that work for small and large sample sizes
4. parsimonious descriptions of data, predictions for missing data, and forecasts for future data
5. a coherent computational framework for model estimation, selection and validation

Classical vs Bayesian

The fundamental differences between classical and Bayesian methods is what is fixed and what is random in an analysis

Paradigm	Fixed	Random
Classical	param (θ)	data (Y)
Bayesian	data (Y)	param (θ)

Why/Why Not Bayesian Statistics?

Pros

1. If $f(\theta)$ & $\mathcal{L}(\theta; Y)$ represent a rational person's beliefs, then Bayes' rule is an optimal method of updating these beliefs given new info (Cox 1946, 1961; Savage 1954; 1972).
2. Provides more intuitive answers in terms of the probability that parameters have particular values.
3. In many complicated statistical problems there are no obvious non-Bayesian inference methods.

Cons

1. It can be hard to mathematically formulate prior beliefs (choice of $f(\theta)$ often ad hoc or for computational reasons)
2. Posterior distributions can be sensitive to prior choice.
3. Analyses can be computationally costly.

Steps to Making Inference

1. Research question
2. Data collection
3. Model $Y_i \approx f(X_i)$
4. Estimate the parameter in the model with uncertainty
5. Make inference

The difference between Classical and Bayesian lies in step 4: (C) uses maximum likelihood estimate (MLE), and (B) derives a posterior distribution.

Example: Estimating the probability of a rare event

Suppose we are interested in the prevalence of an infectious disease in a small city. A small random sample of 20 individuals will be checked for infection.

- Interest is in the fraction of infected individuals

$$\theta \in \Theta = [0, 1]$$

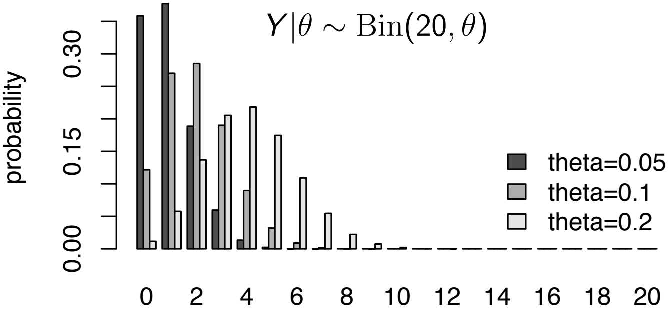
- The data records the number of infected individuals

$$y \in \mathcal{Y} = \{0, 1, \ldots, 20\}$$

Example: Likelihood/sampling model

Before the sample is obtained, the number of infected individuals is unknown.

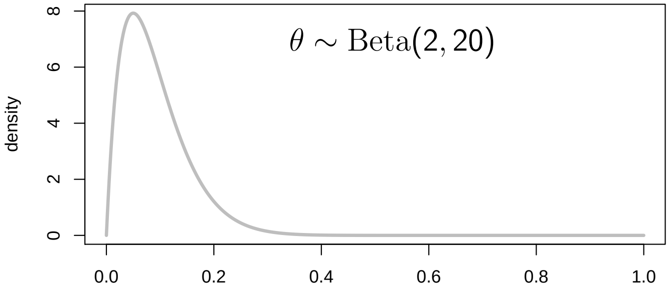
- Let Y denote this to-be-determined value
- If θ were known, a sensible sampling model is



Example: Prior

Other studies from various parts of the country indicate that the infection rate ranges from about 0.05 to 0.20, with an average prevalence of 0.1.

- Moment matching from a beta distribution (a convenient choice) gives the prior



Example: Posterior

The prior and sample model combination:

$$\begin{aligned} \theta &\sim \mathrm{Beta}(a, b) \\ Y|\theta &\sim \mathrm{Bin}(n, \theta) \end{aligned}$$

and an observed y (the data), leads to the posterior

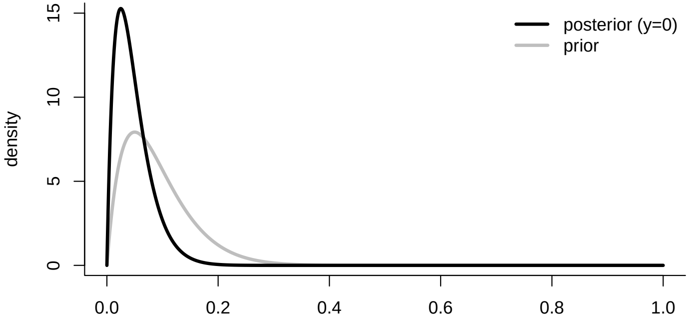
$$p(\theta|y) = \mathrm{Beta}(a + y, b + n - y)$$

Example: Posterior

For our case, we have $a = 2$, $b = 20$, $n = 20$.

If we don't find any infections, then $y = 0$ our posterior is:

$$p(\theta|y = 0) = \mathrm{Beta}(2, 40)$$



Example: Sensitivity Analysis

How influential is our prior?

The posterior expectation is

$$\mathrm{E}\{\theta|Y = y\} = \frac{n}{w+n}\bar{y} + \frac{w}{w+n}\theta_0$$

a weighted average of the sample mean and the prior expectation:

$$\theta_0 = \frac{a}{a+b} \quad \rightarrow \text{prior expectation (or guess)}$$

$$w = a + b \quad \rightarrow \text{prior confidence}$$

Example: A non-Bayesian approach

A standard estimate of a population proportion, θ is the sample mean $\bar{y} = y/n$, the fraction of infected people in the sample.

If $y = 0$, this gives zero, so reporting the sampling uncertainty is crucial (e.g., for reporting to health officials).

The most popular 95% confidence interval for a population proportion is the Wald Interval:

$$\bar{y} \pm 1.96\sqrt{\bar{y}(1-\bar{y})/n}.$$

This has the correct *asymptotic* coverage, but $y = 0$ is still problematic!

Exercise: is a treatment for cancer effective?

We have data on n cancer patients that have been given a particular treatment. Our outcome variable is whether or not it was "effective":

Patient	Effectiveness	Numerical Data
1	N	0
2	N	0
3	Y	1
$\vdots$	$\vdots$	$\vdots$

The appropriate sampling model for each patient is a Bernoilli:

$$Y_i|\theta \overset{iid}{\sim} f(Y|\theta) = \text{Bern}(\theta)$$

where θ is the success rate of the treatment. Based on this, write down the likelihood for the n patients. Then, assuming a $\mathrm{Beta}(a, b)$ prior for θ find the posterior distribution for $\theta|Y$. Does this look familiar?

Conjugate Bayesian Models

Some sets of priors/likelihoods/posteriors exhibit a special relationship called *conjugacy*. This happens if the posterior distribution has the same form as the prior. For instance, in our above Beta-Binomial/Bernoilli examples:

$$\begin{aligned} \theta &\sim \mathrm{Beta}(a, b) \\ Y|\theta &\sim \mathrm{Bin}(n, \theta) \\ \theta|Y &\sim \mathrm{Beta}(a*, b*) \end{aligned}$$

Are all posteriors in the same family as the priors? **No**
Conjugacy is a nice special property, but most of the time this isn't the case and getting an analytic form of the posterior distribution can be hard or impossible.

What do you do with a Posterior?