Single-parameter models: Binomial data

Applied Bayesian Statistics

Dr. Earvin Balderama

Department of Mathematics & Statistics
Loyola University Chicago

September 5, 2017
The binomial model

A basic task in statistics is to estimate a proportion using a series of independent trials.

- Conduct $n$ independent trials,
- each with success probability $\theta$,
- observe $Y =$ the total number of successes.

$\theta \in [0, 1]$ is the parameter we are trying to estimate. We would like to obtain:

- the posterior of $\theta$,
- a 95% interval,
- a test that $\theta$ equals some predetermined value $\theta_0$. 
Bayesian analysis - Likelihood

Given a success probability $\theta$, the distribution of $Y$ can be described as

$$Y \mid \theta \sim \text{Binomial}(n, \theta)$$

Thus, the likelihood function for $Y = y$ is

$$f(y \mid \theta) = \binom{n}{y} \theta^y (1 - \theta)^{n-y}$$

Note: this function is often viewed as a function of $\theta$, especially when performing MLE. In Bayesian statistics, although it is derived as a function of the data, it combines with the prior (a function of only $\theta$) to create a posterior that is also only a function of $\theta$. 
Bayesian analysis - Prior

$\theta$ is the parameter of interest, and is continuous between 0 and 1. Its prior distribution can then be

$$\theta \sim \text{Beta}(a, b)$$

Thus,

$$f(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1}(1 - \theta)^{b-1}$$
Bayesian analysis - Posterior

We can now derive the posterior distribution, which happens to be:

\[ \theta \mid Y \sim \text{Beta}(Y + a, n - Y + b) \]
One-parameter models

Bayesian analysis - Posterior

We can now derive the **posterior** distribution, which happens to be:

\[ \theta \mid Y \sim \text{Beta}(Y + a, n - Y + b) \]

- **Note:** \(a\) and \(b\) can be interpreted as the “prior number of successes and failures,” which may be useful for specifying the prior.
- What values of \(a\) and \(b\) to select if we have no information about \(\theta\) before collecting data?
- What if historical data/expert opinion indicates that \(\theta\) is likely between 0.6 and 0.8?
Shrinkage

- The prior mean was $E(\theta) = \frac{a}{a + b}$.
- The posterior mean is $\hat{\theta}_B = E(\theta \mid Y) = \frac{Y + a}{n + a + b}$.
- The posterior mean is between the sample proportion $\frac{Y}{n}$ and the prior mean $\frac{a}{a+b}$.

1. When is $\hat{\theta}_B$ close to the sample proportion $\frac{Y}{n}$?
2. When is $\hat{\theta}_B$ shrunk towards the prior mean $\frac{a}{a+b}$?
Summarizing the posterior

The posterior contains all the information about the parameter of interest.

- Plotting the posterior is a good idea.
- If the posterior is symmetric, then the posterior mean and standard deviation are good summaries.
- A 90% (or 95%) posterior credible set is also a good summary.
One-parameter models

One-sided hypothesis test

How can we test $H_0 : \theta \leq \theta_0$ versus $H_A : \theta > \theta_0$?

- Compute the posterior probability of the hypotheses:

$$\text{Prob}(H_0 \mid Y) = \text{Prob}(\theta \leq \theta_0 \mid Y)$$

- **Hint:** Use R! (\texttt{pbeta} function)
- Reject if ... ?
- How is this different than the $p$-value?
Two-sided hypothesis test

How can we test $H_0 : \theta = \theta_0$ versus $H_A : \theta \neq \theta_0$?

- Can we compute the posterior probability of the hypotheses?

\[
\text{Prob}(H_0 \mid Y) = \text{Prob}(\theta = \theta_0 \mid Y) = 0
\]

- Must use another approach:
  - Compute the Bayes’ factor (more on this later).
  - See if a 95% posterior interval includes $\theta_0$ (this is common).
Monte Carlo sampling

- Conjugacy leads to simple forms of the posterior distribution, which makes it mathematically easy to obtain posterior summaries.
- For harder problems, integration of the posterior may be very difficult or even impossible.
- Thus, we would have to resort to simulation.
  - **Monte Carlo method:** If we can sample from the posterior, then the empirical distribution will approximate the posterior.
  - Think of the posterior as the population, and we’re taking a (very large) random sample from the population. Then we can compute summaries of the sample, such as the mean and standard deviation.
Monte Carlo sampling

Let $\theta$ be the parameter of interest. Suppose we generate $S$ independent, random values of $\theta$ from the posterior distribution $f(\theta \mid Y)$, then the empirical distribution of the samples $\{\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(S)}\}$ would approximate $f(\theta \mid Y)$, with the approximation improving with increasing $S$.

Example

For the beta-binomial model, the `rbeta` function in R can generate samples, e.g.,

```r
> post <- rbeta(10000, 6, 4)
```
Monte Carlo sampling

We can then use the posterior samples \( \{\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(S)}\} \) to approximate summaries of the posterior.

Example

```r
> post <- rbeta(10000, 6, 4)
> mean(post)#posterior mean
> sd(post)#posterior standard deviation
> quantile(post, c(0.025,0.975))#95% credible interval
> mean(post > 0.5)#Probability that theta>0.5
```

Note: In this example, the functional form of the posterior was already known, so we were simply able to use a the built-in R function `rbeta`. This was just to illustrate the Monte Carlo method.
For harder problems, we must learn how to generate samples from a posterior of unfamiliar form.