STAT 535: Chapter 4: Balance, Sequentiality, and Subjectivity

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

Spring 2022

Example of Subjective Prior Specification: Bechtel Test

- Consider another example in our Beta-binomial framework.
- A movie passes the "Bechtel test" if
  - the movie has to have at least two women in it;
  - these two women talk to each other; and
  - they talk about something besides a man.
- What proportion of all movies pass the Bechtel test? We will try to estimate this unknown proportion π.

Three Different Subjective Priors

- What would you say is a prior guess for the value of π? Between which two values is π likely to fall?
- We could consider a pessimistic $Beta(5, 11)$ prior; an optimistic $Beta(14, 1)$ prior; and a $Beta(1, 1)$ prior that reflects a complete lack of prior knowledge.
- See the plots of these three priors, as well as our own subjective prior.

Informative vs. Vague Priors

- An **informative** prior distribution has lower variance, which reflects precise information about the parameter of interest.
- A **vague** (or diffuse) prior distribution has high variance, which reflects imprecise information about the parameter.
- A **flat** prior distribution implies that all possible values of the parameters are equally likely – this is completely noninformative.

How the Prior Affects the Posterior Inference

- Suppose a random sample of 20 movies yields $y = 9$ out of the 20 that pass the Bechtel test.
- Recall that the posterior here will be Beta$(\alpha + y, \beta + n - y)$.

Table: The prior and posterior models for π, with $y = 9$ and $n = 20$.

Analyst	Prior	Posterior
Pessimistic	$Beta(5, 11)$	$Beta(14, 22)$
Noninformative	$Beta(1, 1)$	$Beta(10, 12)$
Optimistic	$Beta(14, 1)$	$Beta(23, 12)$

The Three Different Posteriors

- See the plots of these posteriors.
- If the posterior mean is used as a point estimator of π, the pessimist would estimate π to be $14/(14+22) = 0.389$.
- The clueless person would estimate π to be $10/(10+12) = 0.455$.
- The optimist would estimate π to be $23/(23+12) = 0.657$.
- So the prior choice does have a substantial effect on the posterior estimate here.
- If the sample size had been larger than 20, the effect of the prior on the posterior would be weakened.
- See the R plots of the posterior for three different samples (each with $y/n \approx 0.46$) corresponding to the same Beta$(14, 1)$ prior.
- We see the effect of the data on the posterior is more substantial when the sample size is larger.

Bayesian Learning and Updating

- We can use the Bayesian approach to update our information about the parameter(s) of interest sequentially as new data become available.
- Suppose we formulate a prior for our parameter θ and observe a random sample $\boldsymbol{y}_1$.
- Then the posterior is

$$p(\theta|\boldsymbol{y}_1) \propto p(\theta)L(\theta|\boldsymbol{y}_1)$$

- Then we observe a new (independent) sample $\boldsymbol{y}_2$.

Bayesian Learning

- We can use our previous posterior as the **new prior** and derive a **new** posterior:

$$\begin{aligned} p(\theta|\boldsymbol{y}_1, \boldsymbol{y}_2) &\propto p(\theta|\boldsymbol{y}_1)L(\theta|\boldsymbol{y}_2) \\ &\propto p(\theta)L(\theta|\boldsymbol{y}_1)L(\theta|\boldsymbol{y}_2) \\ &= p(\theta)L(\theta|\boldsymbol{y}_1, \boldsymbol{y}_2) \\ &\text{(since } \boldsymbol{y}_1, \boldsymbol{y}_2 \text{ independent)} \end{aligned}$$

- Note this is the same posterior we would have obtained had $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ arrived at the same time!
- This "sequential updating" process can continue indefinitely in the Bayesian setup.

Example of Data Order Invariance

- When updating the posterior in this way, it does not matter in which order the data arrive:
- Consider the Bechtel test example with a pessimistic Beta$(5, 11)$ prior on π.
- If we observe $y = 9$ out of the $n = 20$ movies that pass the Bechtel test, then we know our posterior is Beta$(\alpha + y, \beta + n - y) \Rightarrow$ Beta$(14, 22)$.

Example of Data Order Invariance

- If we plan to gather more data, then we could use this posterior as the **prior** for our subsequent analysis.
- So consider a Beta$(14, 22)$ prior, and suppose we look at $n = 10$ more movies, where $y = 6$ of them pass the Bechtel test.
- Then our updated posterior for π is Beta$(14 + 6, 22 + 10 - 6)$ $=$ Beta$(20, 26)$.

Example of Data Order Invariance

- What if the $y = 6, n = 10$ sample had come first, followed by the $y = 9, n = 20$ sample?
- A Beta$(5, 11)$ prior with a $y = 6, n = 10$ sample yields a Beta$(5 + 6, 11 + 10 - 6) =$ Beta$(11, 15)$ posterior.
- Then an updated prior of Beta$(11, 15)$ with a $y = 9, n = 20$ sample yields an updated posterior for π of Beta$(11 + 9, 15 + 20 - 9) =$ Beta$(20, 26)$.
- So the eventual updated posterior is the same, regardless of the order that data came in.

Be Careful of the Support of Your Prior

- **Important:** The support of the posterior will always match the support of the prior.
- Suppose a severe pessimist put a Uniform$(0, 0.2)$ prior on π in the Bechtel test example.
- Then suppose she watched $n = 1000$ movies and $y = 900$ passed the Bechtel test (strong evidence that π is **large**!)
- Her posterior would still only consider values of π between 0 and 0.2, since the posterior's support matches the prior's support.
- The choice of the prior here isn't allowing the data to indicate to you that π is actually large.

Choose a Prior that Allows the Data to Have a Say

- The solution is to choose a prior that has support over the entire parameter space.
- You can still be pessimistic if you want: For example, a Beta$(10, 90)$ prior has prior mean of 0.1 and puts almost all its probability between 0 and 0.2 (see R plot)
- But with this prior, if she watched $n = 1000$ movies and $y = 900$ passed the Bechtel test, then the posterior would be Beta$(10 + 900, 90 + 1000 - 900)$ or Beta$(910, 190)$.
- The posterior mean would be $910/1100 \approx 0.827$.
- The extreme evidence in the data would now be allowed to overwhelm the pessimism in the prior (which is what **should** happen!).