STAT7630: Bayesian Statistics
Lecture Slides # 4

Chapter 4 - Balance and Sequentiality in Bayesian Analyses
(Balance, Sequentiality, and Subjectivity)

Elvan Ceyhan
Department of Mathematics & Statistics
Auburn University

Fall 2024,
Updated: September, 2024

Outline

Example of Subjective Prior Specification: Bechdel Test

- Let's examine another example within the Beta-Binomial framework.
  - **Bechdel Test: The Rules**
  - In Alison Bechdel's 1985 comic strip "The Rule," a movie passes the Bechdel test if:
    - It has at least two women.
    - These two women talk to each other.
    - The conversation is not about a man.
- Let's denote the proportion of movies passing the test as π, a random value between 0 and 1.

Question: Percentage of Movies Passing the Bechdel Test

- What percentage of recent movies pass the Bechdel test?
- Is it closer to 10%, 50%, 80%, or 100%?
- Our goal is to estimate the unknown proportion, π, of movies that pass the Bechdel test.

Three Friends' Prior Beliefs about π

- Three friends – the feminist, the clueless, and the optimist – have different prior models for π:
  - The feminist believes the majority of movies fail the test.
  - The clueless is uncertain about how many movies pass the test.
  - The optimist assumes most movies pass the test.

Outline

Introduction

Prior Specification

Posterior Inference - The Influence of Prior and Likelihood

Bayesian Learning

Three Different Subjective Priors: Beta Priors for π

- What would be your prior guess for the value of π?
- Between which two values is π likely to fall?
- We could consider the following priors:
  - A pessimistic Beta(5, 11) prior: The feminist uses a $\text{Beta}(5, 11)$ prior, placing more probability on $\pi < 0.5$,
  - An optimistic Beta(14, 1) prior: The optimist uses a $\text{Beta}(14, 1)$ prior, believing that most movies pass the test (π close to 1),
  - A Beta(1, 1) prior reflecting no prior knowledge: The clueless uses a $\text{Beta}(1, 1)$ prior (Uniform), placing equal weight on all values of π.
- Refer to the plots of these three priors, along with our own subjective prior.

Beta Prior Models: Visualization

- The clueless's prior: a flat line (Beta(1, 1)).
- The feminist's prior: a curve centered below 0.5 (Beta(5, 11)).
- The optimist's prior: a curve with mode near 1 (Beta(14, 1)).

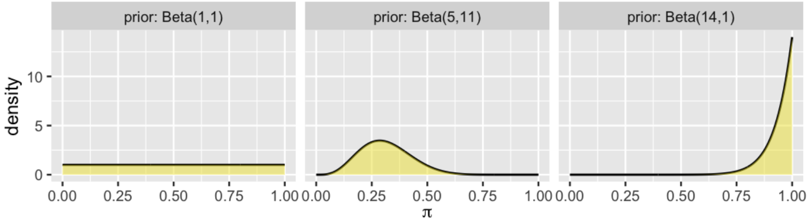


Figure 1: Beta Priors for the Clueless, Feminist, and Optimist

Informative vs. Vague Priors

- An **informative prior** has lower variance, reflecting precise information about the parameter of interest.
- A **vague** (or **diffuse**) **prior** has higher variance, indicating imprecise information about the parameter.
- A flat prior implies that all parameter values are equally likely, representing a completely noninformative prior.

Outline

How the Prior Affects the Posterior Inference

- A sample of $n = 20$ recent movies is taken.
- In this sample, $Y = 9$ movies pass the Bechdel test (45% pass rate).
- What impact will this data have on each friend's prior beliefs?
- Recall the posterior distribution (after observing data) will be Beta$(\alpha + y, \beta + n - y)$.
- where α and β are the parameters from the prior, and
- y is the number of successes (movies passing the test), and n is the total sample size.

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

Table 1: The prior and posterior models for π, with $y = 9$ and $n = 20$.

The Three Different Posteriors: Different Priors - Same Data

- See the plots of these posteriors.
- Using the posterior mean as a point estimator for π:
  - The feminist adjusts her beliefs but still leans pessimistic (mode ≈ 0.38). The pessimist estimates π as $\frac{14}{14+22} = 0.389$.
  - The clueless's posterior closely matches the observed data ($\pi = 0.45$). The clueless person estimates π as $\frac{10}{10+12} = 0.455$.
  - The optimist's prior still places weight on higher values of π, showing optimism. The optimist estimates π as $\frac{23}{23+12} = 0.657$.
- Thus, prior choice can have a significant impact on the posterior estimate.

Visualizing Posterior Models: Different Priors - Same Data

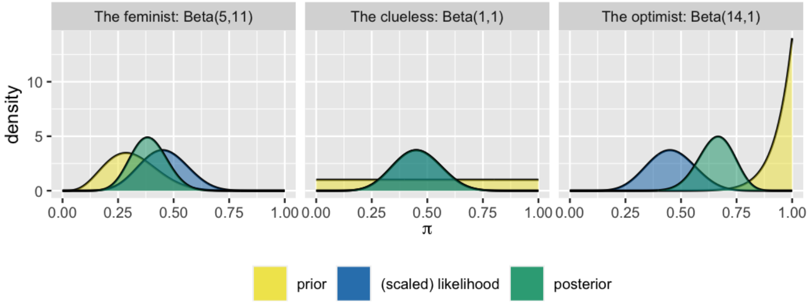


Figure 2: Posterior Models for Each Friend: Posterior models of π, constructed in light of the sample in which $Y = 9$ of $n = 20$ movies passed the Bechdel.

Different Data, Different Posteriors

If you're concerned by the fact that our three analysts have differing posterior understandings of π, don't forget the role that data plays in a Bayesian analysis.

Consider three new analysts – Morteza, Nadide, and Ursula – who all share the optimistic Beta(14,1) prior for π, but each have access to different data:

- Morteza reviews $n = 13$ movies from 1991, among which $Y = 6$ pass the Bechdel.
- Nadide reviews $n = 63$ movies from 2000, among which $Y = 29$ pass the Bechdel.
- Ursula reviews $n = 99$ movies from 2013, among which $Y = 46$ pass the Bechdel.

Though each observes a pass rate of roughly 46%, their sample sizes differ, and this will influence their posterior models.

Posterior Models and Data Influence

The three analysts' common prior (Beta(14,1)) and unique Binomial likelihood functions, reflecting their different data, lead to different posterior models:

Analyst	**Data**	**Posterior**
Morteza	$Y = 6$ of $n = 13$	Beta(20,8)
Nadide	$Y = 29$ of $n = 63$	Beta(43,35)
Ursula	$Y = 46$ of $n = 99$	Beta(60,54)

Notice that the larger the sample size, the more "insistent" the likelihood. Morteza's data, with $n = 13$, result in a wide likelihood, whereas Ursula's data, with $n = 99$, lead to a much narrower likelihood.

The more data, the more the posterior shifts toward the likelihood, reflecting the evidence provided by the sample.

The Three Different Posteriors: Same Prior - Different Data

- See the R plots of the posterior for three different samples (with $y/n \approx 0.46$) using the same Beta(14, 1) prior.
- The effect of the data on the posterior becomes more substantial as the sample size increases. (For example, here, if the sample size were larger than 20, the influence of the prior on the posterior would diminish.)

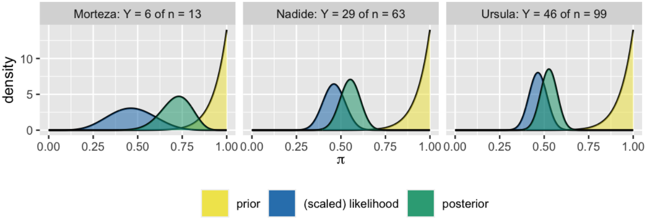


Figure 3: Posterior models of π, constructed from the same prior but different data, are plotted for each analyst.

Outline

Bayesian Learning and Updating

- The Bayesian approach allows us to update our knowledge of the parameter(s) sequentially as new data are collected.
- Start by formulating a prior for the parameter θ and observe an initial random sample $\mathbf{y}_1$.
- The posterior for θ after observing $\mathbf{y}_1$ is:

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

- Next, observe a new random sample $\mathbf{y}_2$ (independent of $\mathbf{y}_1$) to further update the posterior.

Bayesian Learning: Sequential Updating

- We can use the previous posterior as a new prior and update it to derive a new posterior:

$$p(\theta|\mathbf{y}_1, \mathbf{y}_2) \propto p(\theta|\mathbf{y}_1)L(\theta|\mathbf{y}_2) \propto p(\theta)L(\theta|\mathbf{y}_1)L(\theta|\mathbf{y}_2) \\ = p(\theta)L(\theta|\mathbf{y}_1, \mathbf{y}_2)$$

  (since $\mathbf{y}_1$ and $\mathbf{y}_2$ are exchangeable)

- This is the same posterior we would have obtained if $\mathbf{y}_1$ and $\mathbf{y}_2$ were observed together.
- This "sequential updating" process can continue indefinitely in a Bayesian framework.

Sequential Bayesian Analysis

- The posterior can be updated as new data arrives.
- Each new posterior serves as the prior for the next update.
- The final posterior is data order invariant.

Effect of Sample Size

- Larger sample sizes reduce the influence of the prior and increase the influence of data.
- As sample size increases, posteriors for all analysts will converge.

Example of Data Order Invariance

- When updating the posterior, the order in which the data arrive does not affect the result.
- Consider the Bechdel test example with a pessimistic Beta(5, 11) prior on π.
- After observing $y = 9$ out of $n = 20$ movies that pass the Bechdel test, the posterior is:

$$\text{Beta}(\alpha + y, \beta + n - y) \Rightarrow \text{Beta}(14, 22)$$

Example of Data Order Invariance

- If more data are collected, the current posterior can serve as the new prior for the next analysis.
- Suppose we start with a Beta(14, 22) prior and gather data from $n = 10$ additional movies, where $y = 6$ pass the Bechdel test.
- The updated posterior for π becomes:

$$\text{Beta}(14 + 6, 22 + 10 - 6) = \text{Beta}(20, 26)$$

Example of Data Order Invariance

- What if the $y = 6$, $n = 10$ sample had been observed first, followed by the $y = 9$, $n = 20$ sample?
- Starting with a Beta(5, 11) prior, the $y = 6$, $n = 10$ sample gives a posterior:

$$\text{Beta}(5 + 6, 11 + 10 - 6) = \text{Beta}(11, 15)$$

- Using this Beta(11, 15) posterior as the prior and observing the $y = 9$, $n = 20$ sample gives:

$$\text{Beta}(11 + 9, 15 + 20 - 9) = \text{Beta}(20, 26)$$

- Thus, the final posterior is the same, regardless of the order in which the data are received.

Avoiding Stubborn Priors: Be Careful of the Support of Your Prior

- Be cautious with priors that assign zero probability to certain outcomes.
- **Important:** The support of the posterior will always match the support of the prior. A stubborn prior will prevent learning from data.
- Consider a severe pessimist who places a Uniform(0, 0.2) prior on π in the Bechdel test example.
- Now, suppose she observes $n = 1000$ movies and $y = 900$ pass the Bechdel test (strong evidence that π is large).
- However, her posterior would still only consider values of π between 0 and 0.2, because the posterior's support must match the prior's support.
- This choice of prior prevents the data from indicating that π could actually be much larger.

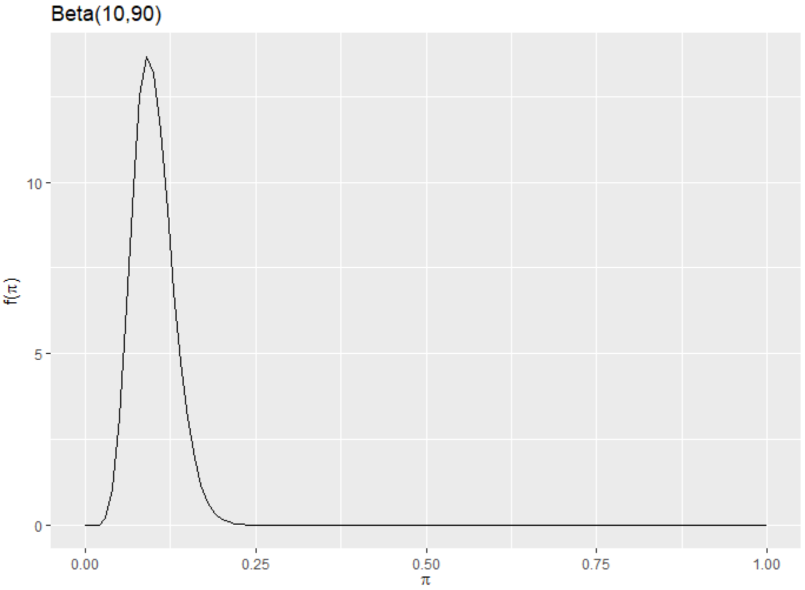
Choose a Prior that Allows the Data to Have a Say

- The solution is to select a prior with support over the entire parameter space.
- You can still be pessimistic, for example, a Beta(10, 90) prior has a mean of 0.1 and places most of its probability between 0 and 0.2 (see R plot).
- However, if $n = 1000$ movies are observed and $y = 900$ pass the Bechdel test, the updated posterior would be:

$$\text{Beta}(10 + 900, 90 + 1000 - 900) = \text{Beta}(910, 190)$$

- The posterior mean would then be $\frac{910}{1100} \approx 0.827$.
- This allows the strong evidence from the data to override the pessimistic prior, as it should!

Plot of Beta(10, 90) prior



Final Takeaways

- The balance between the prior and data depends on sample size and prior strength.
- In Bayesian analysis, data plays an increasingly important role as sample size grows.
- Sequential analysis allows continuous updating of beliefs as new data comes in.

Chapter Summary

- Posterior models reflect a balance between prior beliefs and observed data.
- The more informative the prior, the more influence it has on the posterior.
- The larger the dataset, the more influence the data has on the posterior.
- Sequential Bayesian analysis allows for updating as new data arrives, with data order invariance.