STAT 339: Statistical Theory

Bayesian Interval Estimation

Anthony Scotina

Bayesian Credible Intervals

Interval Estimates (Recap)

Frequentist Interval Estimation

> A $(1-\alpha)\times 100\%$ **confidence interval** is an interval $[\hat{\theta}_L, \hat{\theta}_U]$ such that
>
> $$P(\hat{\theta}_L \leq \theta \leq \hat{\theta}_U) = 1-\alpha,$$
>
> where $1-\alpha$ is the **confidence coefficient**.

- $[\hat{\theta}_L, \hat{\theta}_U] \longrightarrow$ **random** interval
- $\theta \longrightarrow$ **fixed**

Because θ is **fixed**, we do NOT interpret this interval as "the probability that θ is in the interval.

Interval Estimates (Recap)

Bayesian Interval Estimation

The parameter θ is *random variable* with a:

- **prior** distribution that reflects our prior beliefs about the variability of θ
- **posterior** distribution, $\theta \mid \mathbf{y}$, that reflects our *updated* understanding of θ after observing *data*.

Suppose θ has a posterior distribution $\theta \mid \mathbf{y}$ with posterior pdf $f(\theta \mid \mathbf{y})$. Then the probability that θ is in the interval (a, b) (given the *observed data*) is

$$P(a \leq \theta \leq b \mid \mathbf{y}) = \int_a^b f(\theta \mid \mathbf{y})\, d\theta.$$

If $P(a \leq \theta \leq b \mid \mathbf{y}) = 0.95$, then we say that (a, b) is a 95% **credible interval** for θ.

Animal Crossing!

Suppose a group of college students are interested in starting an *Animal Crossing* club.

- In order to estimate demand, the students want to provide an *interval estimate* for θ, the **proportion of students who play Animal Crossing**.

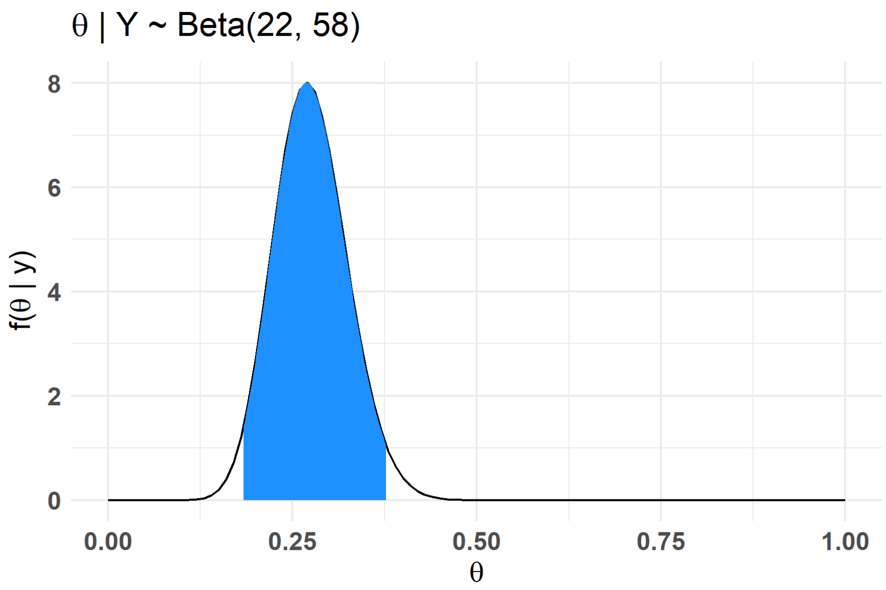
From a few weeks ago:

- **Prior**: $\theta \sim Beta(10, 40)$
- **Data**: $Y \mid \theta \sim Binomial(30, \theta)$, where we observe $Y = 12$
- **Posterior**: $\theta \mid Y \sim Beta(22, 58)$

Bayesian Credible Interval

Using the posterior $\theta \mid Y$, we can find a **95% credible interval** by finding the 2.5th and 97.5th **posterior percentiles**.

- These mark the *middle 95% of posterior plausible values* for θ.




```
c(qbeta(0.025, 22, 58), qbeta(0.975, 22, 58))
```

```
## [1] 0.1834550 0.3771967
```

Bayesian Credible Interval

```
c(qbeta(0.025, 22, 58), qbeta(0.975, 22, 58))
```

```
## [1] 0.1834550 0.3771967
```

There is a **95% posterior probability** that somewhere between 18.3% and 37.7% of college students play Animal Crossing.

- Posterior mean: $22/(22+58) = 0.275 \rightarrow$ 27.5%

Another way to think about this:

$$
\begin{aligned}
P(0.183 \leq \theta \leq 0.377 \mid Y = 12) &= \int_{0.183}^{0.377} f(\theta \mid y = 12)\, d\theta \\
&= \int_{0.183}^{0.377} \frac{\Gamma(22+58)}{\Gamma(22)\Gamma(58)} \theta^{22-1} (1-\theta)^{58-1}\, d\theta \\
&= 0.95
\end{aligned}
$$

Note: If we want to find, say, a **90% credible interval**, we just mark the *middle 90%* of the posterior distribution instead!

Comparison to Frequentist CI for p

Recall that a 95% **confidence interval** for p is given by

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

In our example...

- $\hat{p} = 12/30 = 0.4$
- $n = 30$

95% Confidence Interval: $(0.225, 0.575)$

95% Credible Interval: $(0.183, 0.377)$

- **Why so different?** 🤔

It has to do with our choice of **prior**!

- (Not to mention these intervals actually have *very* different meanings!)

Interpreting Credible Intervals

Unlike with *frequentist* **confidence intervals**, the *Bayesian* setup allows us to say that θ *is inside* $(0.183, 0.377)$ with some probability, **not 0 or 1**.

- Under the Bayesian framework, θ is a **random variable** with a **probability distribution**.

The 95% confidence interval of $(0.225, 0.575)$ is just one of the possible *realized values* of the *random interval*

$$\left(\hat{p} - 1.96\sqrt{\hat{p}(1-\hat{p})n}, \hat{p} + 1.96\sqrt{\hat{p}(1-\hat{p})n}\right)$$

- Under the frequentist framework, θ **does not move**! It is *fixed* and is inside $(0.225, 0.575)$ with probability **either 0 or 1**.

Bayesian Probability

Bayesians and frequentists also interpret probabilities differently, so it is important not to confuse **credible** (Bayesian) and **coverage** (frequentist) probability!

- **Credible** probability: Reflects the experimenter's *subjective beliefs*, which are expressed in the *prior* distribution and updated in the *posterior* distribution after observing **DATA**.
- **Coverage** probability: Represents a *long-run relative frequency* of identical trials; 95% of *realized* confidence intervals will cover θ.

[*Moose pic to fill space* 👇]

Cool Beans (yes, that one.)

Let Y_i, the number of people in front of you in line at **Cool Beans** on day i be distributed according to a Poisson distribution with parameter λ:

$$Y_i \mid \lambda \sim Poisson(\lambda)$$

- **Prior**: $\lambda \sim Gamma(11, 1)$
- **Data**: $n = 5$ days;
  $\mathbf{y} = (15, 12, 5, 8, 10)$

Find a 99% credible interval for λ.

Gamma-Exponential Credible Interval

Suppose we want to estimate the lifetime (in hours), θ, of a certain electrical component.

Consider the following:

- **Prior**: $\theta \sim Gamma(\alpha, \beta)$, where

$$f(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}$$

- **Likelihood**: $Y_1, Y_2, \ldots, Y_n \mid \theta \sim Exponential(\theta)$, where

$$f(y_i \mid \theta) = \theta e^{-\theta y_i}$$

Construct a 90% credible interval for θ and the mean of the exponential population, $\mu = 1/\theta$.

Credible Intervals for the Mean

The Normal-Normal Conjugacy

From Practice 5:

If:

- **Prior**: $\mu \sim N(\theta, \tau^2)$
- **Data**: $Y_i \mid \mu \sim N(\mu, \sigma^2)$ (μ is *unknown* but σ^2 is **known**)

then the *posterior distribution* is also Normally distributed:

$$\mu \mid \mathbf{y} \sim N\left(\theta \frac{\sigma^2}{n\tau^2 + \sigma^2} + \bar{y}\frac{n\tau^2}{n\tau^2 + \sigma^2}, \frac{\tau^2\sigma^2}{n\tau^2 + \sigma^2}\right)$$

Credible Intervals for a Normal Mean

Because $\mu \mid \mathbf{y}$ is **Normally** distributed (and hence, *symmetric*), we can use techniques similar to those used to derive *frequentist* CIs for a Normal mean!

Want: a and b such that

$$P(a \leq \mu \leq b \mid \mathbf{y}) = 1 - \alpha$$

Know:

- *Posterior Distribution*: **Normal**! 🔔
- *Posterior Mean*: $\mu_1 \equiv \theta \frac{\sigma^2}{n\tau^2 + \sigma^2} + \bar{y} \frac{n\tau^2}{n\tau^2 + \sigma^2}$
- *Posterior Variance*: $\sigma_1^2 \equiv \frac{\tau^2 \sigma^2}{n\tau^2 + \sigma^2}$

$$\implies P(\mu_1 - z_{\alpha/2}\sigma_1 < \mu < \mu_1 + z_{\alpha/2}\sigma_1 \mid \mathbf{y}) = 1 - \alpha$$

Credible Intervals for a Normal Mean

If:

- **Prior**: $\mu \sim N(\theta, \tau^2)$
- **Data**: $Y_1, Y_2, \ldots, Y_n \mid \mu \sim N(\mu, \sigma^2)$
    - μ is *unknown* and σ^2 is **known**

Then: A $(1-\alpha) \times 100\%$ **Bayesian credible interval** for μ is

$$\mu_1 \pm z_{\alpha/2}\sigma_1,$$

where μ_1 is the **posterior mean** (and the *Bayes estimator*) and σ_1^2 is the **posterior variance**.

The Stroop Test

The **Stroop Effect** describes the psychological phenomenon that occurs when the processing of one particular stimulus feature interferes with the simultaneous processing of a second stimulus feature.

A random sample of $n = 8$ study participants yielded the following *reaction times* (in seconds per hundred reactions):

- 95, 99, 106, 107, 107, 114, 120, 127

We'll assume $Y_i \mid \mu \sim N(\mu, 12^2)$

Based on prior studies, it is reasonable to assume that reaction times are *normally distributed* **with mean 100 and standard deviation 15**.

> Construct a 95% Bayesian credible interval for μ, the population mean reaction time for the Stroop Test.

Large-Sample Credible Intervals

Large-Sample Normal Approximation to the Posterior

Suppose we have **data** $Y_1, Y_2, \ldots, Y_n$ modeled from some (preferably named) distribution with parameter θ.

- For example, in the **Animal Crossing** example, $Y \mid \theta \sim Binomial(n, \theta)$.

If n is **large**, we can use the Normal distribution to approximate the posterior:

$$\theta \mid \mathbf{y} \sim (approx)\ Normal\left(\hat{\theta}_{MLE}, \frac{1}{I(\hat{\theta}_{MLE})}\right).$$

- $\hat{\theta}_{MLE}$ is the **MLE** for θ in the data model.
- $I(\hat{\theta}_{MLE}) = -\frac{d^2}{d\theta^2} \log L(\theta)\Big|_{\theta=\hat{\theta}_{MLE}}$ is the **Fisher information**.

Large-Sample Approximation

Suppose instead of surveying $n = 30$ students regarding their Animal Crossing preferences, we survey $n = 500$.

- $Y \mid \theta \sim Binomial(500, \theta)$
- **Note**: Because we're using a large sample approximation, it doesn't *really* matter what prior we use.
    - Though this is just for illustration! If you use a conjugate prior, there's no reason to do this...

Binomial MLE: $\hat{\theta}_{MLE} = Y/n$

- For large n, $\theta \mid Y \sim (approx)\ Normal\left(\hat{\theta}_{MLE}, \frac{1}{I(\hat{\theta}_{MLE})}\right)$.

Calculating Fisher Information

We just need to find $I(\hat{\theta}_{MLE}) = -\frac{d^2}{d\theta^2}\log L(\theta)\Big|_{\theta=\hat{\theta}_{MLE}}$

- $\log L(\theta) = \log\left[\binom{n}{y}\theta^y(1-\theta)^{n-y}\right] = \log\binom{n}{y} + y\log\theta + (n-y)\log(1-\theta)$
- $\frac{d}{d\theta}\log L(\theta) = \frac{y}{\theta} + \frac{n-y}{1-\theta}\cdot(-1)$
- $\frac{d^2}{d\theta^2}\log L(\theta) = (-1)\frac{y}{\theta^2} + (-1)\frac{n-y}{(1-\theta)^2}(-1)(-1)$

$\implies I(\hat{\theta}_{MLE}) = \frac{y}{\hat{\theta}^2_{MLE}} + \frac{n-y}{(1-\hat{\theta}_{MLE})^2}$

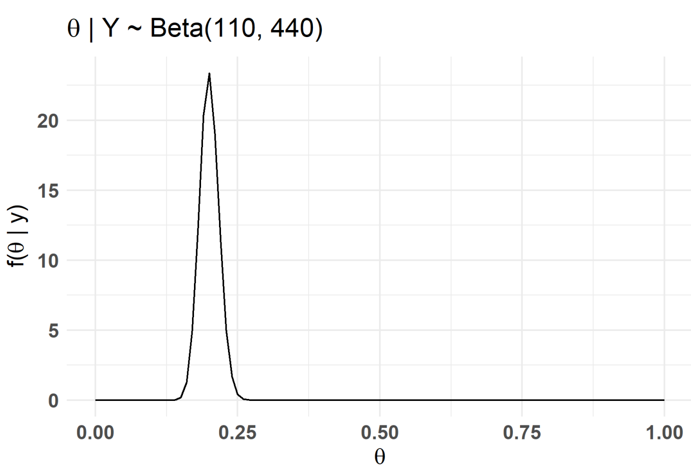
Comparing Approximation to Exact

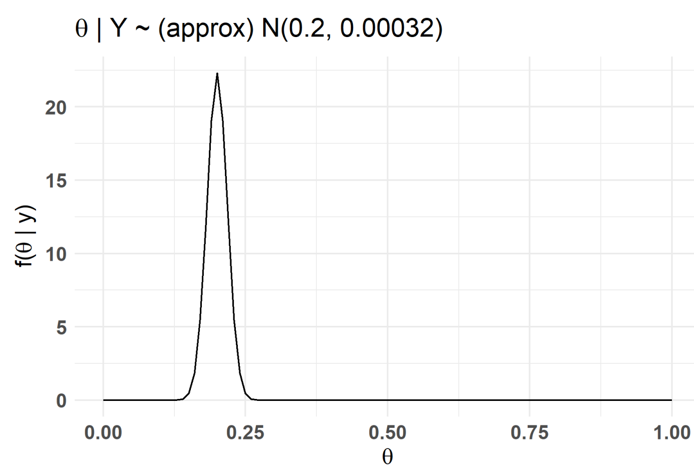
Assuming...

- $\theta \sim Beta(10, 40)$
- $Y \mid \theta \sim Binomial(n = 500, \theta)$ with observed $Y = 100$

Then the **exact** posterior is
$\theta \mid Y \sim Beta(10 + 100, 40 + 500 - 100) = Beta(110, 440)$.

Using the **large-sample Normal approximation**,
$\theta \mid Y \sim (approx)\ N(0.2, 0.00032)$.





Small Samples

This approximation doesn't work as well with small n. 😩😩😩

Assuming...

- $\theta \sim Beta(10, 40)$
- $Y \mid \theta \sim Binomial(n = 10, \theta)$ with observed $Y = 2$

Then the **exact** posterior is $\theta \mid Y \sim Beta(10 + 2, 40 + 10 - 2) = Beta(12, 48)$.

Using the **large-sample Normal approximation**, $\theta \mid Y \sim (approx)\ N(0.2, 0.016)$.