



# Homework 2

Brian Cervantes Alvarez

May 16, 2024

ST 559 Bayesian Statistics

## Problem 1

### Part A

Given that  $y_1, \dots, y_n | \theta \sim \text{iid Poisson}(\theta)$  and  $\theta \sim \text{Gamma}(a, b)$ , we can express the likelihood and prior as follows,

The likelihood of observing  $y_1, \dots, y_n$  given  $\theta$  for a Poisson distribution is given as,

$$L(\theta | y_1, \dots, y_n) = \prod_{i=1}^n \frac{e^{-\theta} \theta^{y_i}}{y_i!} = \frac{e^{-n\theta} \theta^{\sum_{i=1}^n y_i}}{\prod_{i=1}^n y_i!}$$

The prior distribution of  $\theta$  is,

$$p(\theta) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta}$$

Applying Bayes' theorem, the posterior distribution is,

$$\begin{aligned} p(\theta | y_1, \dots, y_n) &\propto L(\theta | y_1, \dots, y_n) \times p(\theta) = \frac{e^{-n\theta} \theta^{\sum_{i=1}^n y_i}}{\prod_{i=1}^n y_i!} \times \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} \\ &= \frac{b^a}{\Gamma(a) \prod_{i=1}^n y_i!} \theta^{(\sum_{i=1}^n y_i + a - 1)} e^{-(b+n)\theta} \end{aligned}$$

This simplifies to a Gamma distribution,

$$\theta | y_1, \dots, y_n \sim \text{Gamma} \left( \sum_{i=1}^n y_i + a, b + n \right)$$

Hence, our posterior distribution of  $\theta$  given  $y_1, \dots, y_n$  is a Gamma with parameters  $\sum_{i=1}^n y_i + a$  and  $b + n$ .



---

## Part B

---

The posterior predictive distribution for a new observation  $\tilde{y}$ , given  $y_1, \dots, y_n$ , is,

$$p(\tilde{y}|y_1, \dots, y_n) = \int_0^\infty p(\tilde{y}|\theta)p(\theta|y_1, \dots, y_n)d\theta$$

where  $p(\tilde{y}|\theta) = e^{-\theta}\theta^{\tilde{y}}/\tilde{y}!$ , and  $p(\theta|y_1, \dots, y_n)$  is  $\text{Gamma}(\sum_{i=1}^n y_i + a, b + n)$ .

The predictive distribution simplifies to a negative binomial distribution,

$$p(\tilde{y}|y_1, \dots, y_n) = \frac{\Gamma(\tilde{y} + \sum_{i=1}^n y_i + a)}{\tilde{y}!\Gamma(\sum_{i=1}^n y_i + a)} \left(\frac{b}{b+n}\right)^{a+\sum_{i=1}^n y_i} \left(\frac{n}{b+n}\right)^{\tilde{y}}$$

Therefore, the posterior predictive distribution of a new observation  $\tilde{y}$  given  $y_1, \dots, y_n$  is a negative binomial distribution with parameters  $\sum_{i=1}^n y_i + a$  and  $\frac{n}{b+n}$ .



---

## Problem 2

---

### Part A

---

For  $\theta_A$  and  $\theta_B$ , using their respective priors and the sums of tumor counts,

**Sum of Tumor Counts:**

$$\sum y_A = 117, \quad \sum y_B = 104$$

**Posterior for  $\theta_A$ :**

$$\theta_A | y_A \sim \text{Gamma}(237, 20)$$

- Mean:  $\mu_A = 11.85$
- Variance:  $\sigma_A^2 = 0.5925$

**Posterior for  $\theta_B$ :**

$$\theta_B | y_B \sim \text{Gamma}(116, 14)$$

- Mean:  $\mu_B = 8.29$
- Variance:  $\sigma_B^2 = 0.5918$

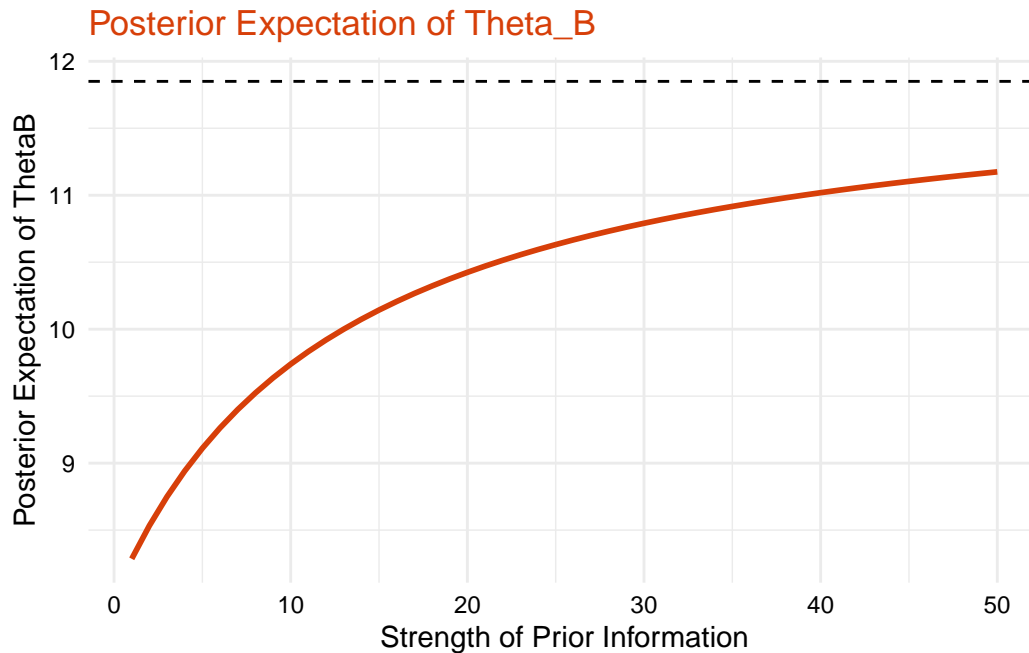
The 95% quantile-based Bayesian confidence intervals can be derived using R. I applied `qgamma(c(0.025, 0.975), shape = alphaA, rate = betaA)` and `qgamma(c(0.025, 0.975), shape = alphaB, rate = betaB)` to get the 95% Bayesian confidence intervals.

```
[1] "95% Confidence Interval for theta_A: [10.39, 13.41]"
```

```
[1] "95% Confidence Interval for theta_B: [6.85, 9.86]"
```

## Part B

We can plot the posterior expectation of  $\theta_B$  as a function of  $n_0$  in the prior  $\text{Gamma}(12n_0, n_0)$  using R with my favorite package `ggplot2`,



The posterior expectation helps us understand how much our previous knowledge (or assumptions) affects the results we get after seeing the actual data. By plotting and analyzing this, we can see how sensitive our results are to changes in what we previously thought. This is especially useful when our initial guesses or information might be a bit uncertain or not based on a lot of data. This way, we can check if our conclusions are reliable or if they're too dependent on shaky assumptions.