

# Assignment 1

Yufei Yin

## Exercise 1 (2 points):

Using the pdf of a Beta distribution, show that

$$\int_0^1 \binom{n}{y} \theta^y (1 - \theta)^{n-y} d\theta = \frac{1}{n+1}$$

### Beta Distribution

- $f(x; \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}$
- $\Gamma(\alpha) = (\alpha-1)!$
- $\int \text{Beta}(\theta|\alpha, \beta) = 1$

$\theta^y (1 - \theta)^{n-y}$  is the “kernel” of  $\text{Beta}(y+1, n-y+1)$ .

$$\binom{n}{y} = \frac{n!}{y!(n-y)!} = \frac{\Gamma(n+1)}{\Gamma(y+1)\Gamma(n-y+1)}$$

$$\text{Beta}(y+1, n-y+1) = \frac{\Gamma(y+1+n-y+1)}{\Gamma(y+1)\Gamma(n-y+1)} \theta^y (1-\theta)^{n-y} = \frac{\Gamma(n+2)}{\Gamma(y+1)\Gamma(n-y+1)} \theta^y (1-\theta)^{n-y}$$

$$\begin{aligned} & \int_0^1 \binom{n}{y} \theta^y (1 - \theta)^{n-y} d\theta \\ &= \int_0^1 \frac{\Gamma(n+1)}{\Gamma(y+1)\Gamma(n-y+1)} \theta^y (1 - \theta)^{n-y} d\theta \\ &= \int_0^1 \frac{\Gamma(n+2)/\Gamma(n+1)}{\Gamma(y+1)\Gamma(n-y+1)} \theta^y (1 - \theta)^{n-y} d\theta \\ &= \frac{1}{n+1} \int_0^1 \frac{\Gamma(n+2)}{\Gamma(y+1)\Gamma(n-y+1)} \theta^y (1 - \theta)^{n-y} d\theta \\ &= \frac{1}{n+1} \end{aligned}$$

## Exercise 2 (2 points):

Consider the Election Example from the lecture slides, specifically the case when more data are observed. Show that the following two approaches provide the same results:

1. We start with our original prior,  $\text{Beta}(5.5, 4.5)$ , update our prior based on the first set of data (10 likely voters are surveyed, 3 of whom say they will vote for candidate A), and use the resulting posterior as our new prior and update it based on the second set of data (an additional 20 likely voters are surveyed, 12 of whom say they will vote for candidate A).
2. We combine the two data sets and use them to update our original  $\text{Beta}(5.5, 4.5)$  prior distribution.

$$y|\theta \sim \text{Binomial}(n, \theta) \text{ with } \theta \sim \text{Beta}(\alpha, \beta)$$

$$\theta|y \sim \text{Beta}(\alpha + y, \beta + n - y)$$

1.

original prior distribution  $\text{Beta}(5.5, 4.5)$ ,  $n_1 = 10$   $y_1 = 3$

posterior distribution  $\text{Beta}(5.5+3, 4.5+10-3) = \text{Beta}(8.5, 11.5)$

updated prior distribution, new prior  $\text{Beta}(8.5, 11.5)$   $n_2 = 20$   $y_2 = 12$

posterior distribution  $\text{Beta}(8.5+12, 11.5+20-12) = \text{Beta}(20.5, 19.5)$

2.

original prior distribution  $\text{Beta}(5.5, 4.5)$ ,  $n = n_1 + n_2 = 10 + 20 = 30$   $y = y_1 + y_2 = 3 + 12 = 15$

posterior distribution  $\text{Beta}(5.5+15, 4.5+30-15) = \text{Beta}(20.5, 19.5)$

So these two approaches provides the same results.

## Exercise 3 (2 points) and 4 (4 points):

For Exercises 3 and 4, please do Chapter 2 Exercises 1 and 5 from Bayesian Data Analysis, 3rd Edition (Gelman et al.), which is linked to from the course Canvas page. Note that for the Chapter 2 Exercise 5 from BDA3, you may use your result from Exercise 1 assigned to you above, i.e.,

$$\int_0^1 \binom{n}{y} \theta^y (1-\theta)^{n-y} d\theta = \frac{1}{n+1}$$

### Chapter 2 Exercises 1

Posterior inference: suppose you have a  $\text{Beta}(4, 4)$  prior distribution on the probability  $\theta$  that a coin will yield a ‘head’ when spun in a specified manner. The coin is independently spun ten times, and ‘heads’ appear fewer than 3 times. You are not told how many heads were seen, only that the number is less than 3. Calculate your exact posterior density (up to a proportionality constant) for  $\theta$  and sketch it.

$$n = 10 \quad x < 3$$

$$x|\theta \sim \text{Binomial}(n, \theta)$$

#### Binomial( $n, \theta$ ) distribution

$$P(x|n, \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

#### prior distribution

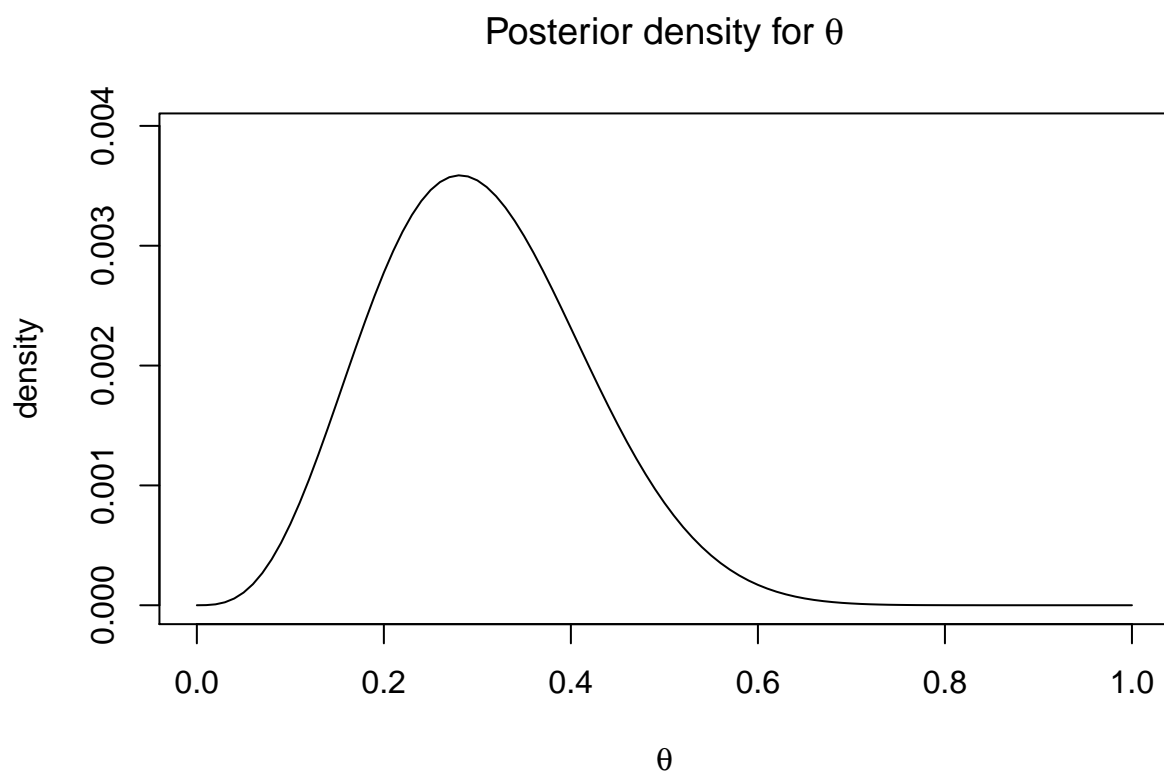
$$\theta \sim \frac{\Gamma(8)}{\Gamma(4)\Gamma(4)} \theta^3 (1-\theta)^3 \propto \theta^3 (1-\theta)^3$$

#### likelihood function ( $x < 3$ )

$$\begin{aligned} \sum_{i=0}^2 P(x=i|\theta) \\ &= \binom{10}{0} \theta^0 (1-\theta)^{10} + \binom{10}{1} \theta^1 (1-\theta)^9 + \binom{10}{2} \theta^2 (1-\theta)^8 \\ &= (1-\theta)^{10} + 10\theta(1-\theta)^9 + 45\theta^2(1-\theta)^8 \end{aligned}$$

#### posterior distribution

$$\begin{aligned} P(\theta|x) &\propto P(x < 3|\theta)P(\theta) \\ &\propto [(1-\theta)^{10} + 10\theta(1-\theta)^9 + 45\theta^2(1-\theta)^8] \times [\theta^3(1-\theta)^3] \\ &\propto \theta^3(1-\theta)^{13} + 10\theta^4(1-\theta)^{12} + 45\theta^5(1-\theta)^{11} \end{aligned}$$



## Chapter 2 Exercises 5

Posterior distribution as a compromise between prior information and data: let  $y$  be the number of heads in  $n$  spins of a coin, whose probability of heads is  $\theta$ .

- (a) If your prior distribution for  $\theta$  is uniform on the range  $[0, 1]$ , derive your prior predictive distribution for  $y$ ,

$$Pr(y = k) = \int_0^1 Pr(y = k|\theta)d\theta,$$

for each  $k = 0, 1, \dots, n$ .

$$\begin{aligned} Pr(y = k) &= \int_0^1 Pr(y = k|\theta)d\theta \\ &= \int_0^1 \binom{n}{k} \theta^k (1 - \theta)^{n-k} d\theta \\ &= \frac{1}{n+1} \end{aligned}$$

- (b) Suppose you assign a  $\text{Beta}(\alpha, \beta)$  prior distribution for  $\theta$ , and then you observe  $y$  heads out of  $n$  spins. Show algebraically that your posterior mean of  $\theta$  always lies between your prior mean,  $\frac{\alpha}{\alpha+\beta}$ , and the observed relative frequency of heads,  $\frac{y}{n}$ .

**prior distribution**

$$\text{Beta}(\alpha, \beta)$$

**posterior distribution**

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

**posterior mean**

$$E(\theta|y) = \frac{\alpha+y}{\alpha+y+\beta+n-y} = \frac{\alpha+y}{\alpha+\beta+n}$$

$E(\theta|y)$  is always lies between prior mean,  $\frac{\alpha}{\alpha+\beta}$ , and the observed relative frequency of heads,  $\frac{y}{n}$ .

**proof**

$$E(\theta|y) = \lambda\left(\frac{y}{n}\right) + (1 - \lambda)\left(\frac{\alpha}{\alpha+\beta}\right), \lambda \in (0, 1)$$

$$\frac{\alpha+y}{\alpha+\beta+n} = \lambda\left(\frac{y}{n} - \frac{\alpha}{\alpha+\beta}\right) + \frac{\alpha}{\alpha+\beta}$$

$$\frac{\alpha+y}{\alpha+\beta+n} - \frac{\alpha}{\alpha+\beta} = \lambda\left[\frac{(\alpha+\beta)y - \alpha n}{(\alpha+\beta)n}\right]$$

$$\frac{(\alpha+y)(\alpha+\beta) - \alpha(\alpha+\beta+n)}{(\alpha+\beta+n)(\alpha+\beta)} = \lambda\left[\frac{\alpha y + \beta y - \alpha n}{(\alpha+\beta)n}\right]$$

$$\frac{\alpha^2 + \alpha\beta + \alpha y + \beta y - \alpha^2 - \alpha\beta - \alpha n}{(\alpha+\beta+n)(\alpha+\beta)} = \lambda\left[\frac{\alpha y + \beta y - \alpha n}{(\alpha+\beta)n}\right]$$

$$\frac{\alpha y + \beta y - \alpha n}{(\alpha+\beta+n)(\alpha+\beta)} = \lambda\left[\frac{\alpha y + \beta y - \alpha n}{(\alpha+\beta)n}\right]$$

$$\lambda = \frac{\alpha y + \beta y - \alpha n}{(\alpha+\beta+n)(\alpha+\beta)} \times \frac{(\alpha+\beta)n}{\alpha y + \beta y - \alpha n}$$

$$\lambda = \frac{n}{\alpha+\beta+n}, \lambda \in (0, 1)$$

- (c) Show that, if the prior distribution on  $\theta$  is uniform, the posterior variance of  $\theta$  is always less than the prior variance.

**prior distribution**

Beta(1, 1)

**posterior distribution**

Beta(1+y, 1+n-y)

**prior variance**

$$\text{var}(\theta) = \frac{1}{12}$$

**posterior variance**

$$\text{var}(\theta|y) = \frac{(1+y)(1+n-y)}{(2+n)^2(3+n)}$$

**Proof**

$$\text{var}(\theta|y) = \frac{(1+y)(1+n-y)}{(2+n)^2(3+n)} < \text{var}(\theta) = \frac{1}{12}$$

$$\text{var}(\theta|y) = \frac{(1+y)(1+n-y)}{(2+n)^2(3+n)} = \frac{1+y}{2+n} \times \frac{1+n-y}{2+n} \times \frac{1}{3+n}$$

Because  $1 + y + 1 + n - y = 2 + n$ ,  $\frac{1+y}{2+n} \times \frac{1+n-y}{2+n} \leq \frac{1}{4}$ .

And  $\frac{1}{3+n} < \frac{1}{3}$ .

$$\text{Overall, } \text{var}(\theta|y) = \frac{(1+y)(1+n-y)}{(2+n)^2(3+n)} = \frac{1+y}{2+n} \times \frac{1+n-y}{2+n} \times \frac{1}{3+n} < \text{var}(\theta) = \frac{1}{12}$$

- (d) Give an example of a  $\text{Beta}(\alpha, \beta)$  prior distribution and data  $y, n$ , in which the posterior variance of  $\theta$  is higher than the prior variance.

The prior variance is larger than the posterior variance on average.

However, it's possible that  $\text{var}(\theta) < \text{var}(\theta|y)$

This is often evidence of model misspecification. The prior does not agree with the data!

For example,

**prior distribution**

$\text{Beta}(10, 1)$

$n = 10, y = 3$

**posterior distribution**

$\text{Beta}(10+3, 1+10-3) = \text{Beta}(13, 8)$

**prior variance**

$$\text{var}(\theta) = \frac{10 \times 1}{(10+1)^2(10+1+1)} = \frac{5}{726} \approx 6.887 \times 10^{-3}$$

**posterior variance**

$$\text{var}(\theta|y) = \frac{13 \times 8}{(13+8)^2(13+8+1)} = \frac{52}{4851} \approx 1.072 \times 10^{-2}$$

$$1.072 \times 10^{-2} > 6.887 \times 10^{-3}$$

In this example, the posterior variance of  $\theta$  is higher than the prior variance.