# STT 465 Homework 2

*Nate Davis*

*October 4, 2017*

## Problem 3.3

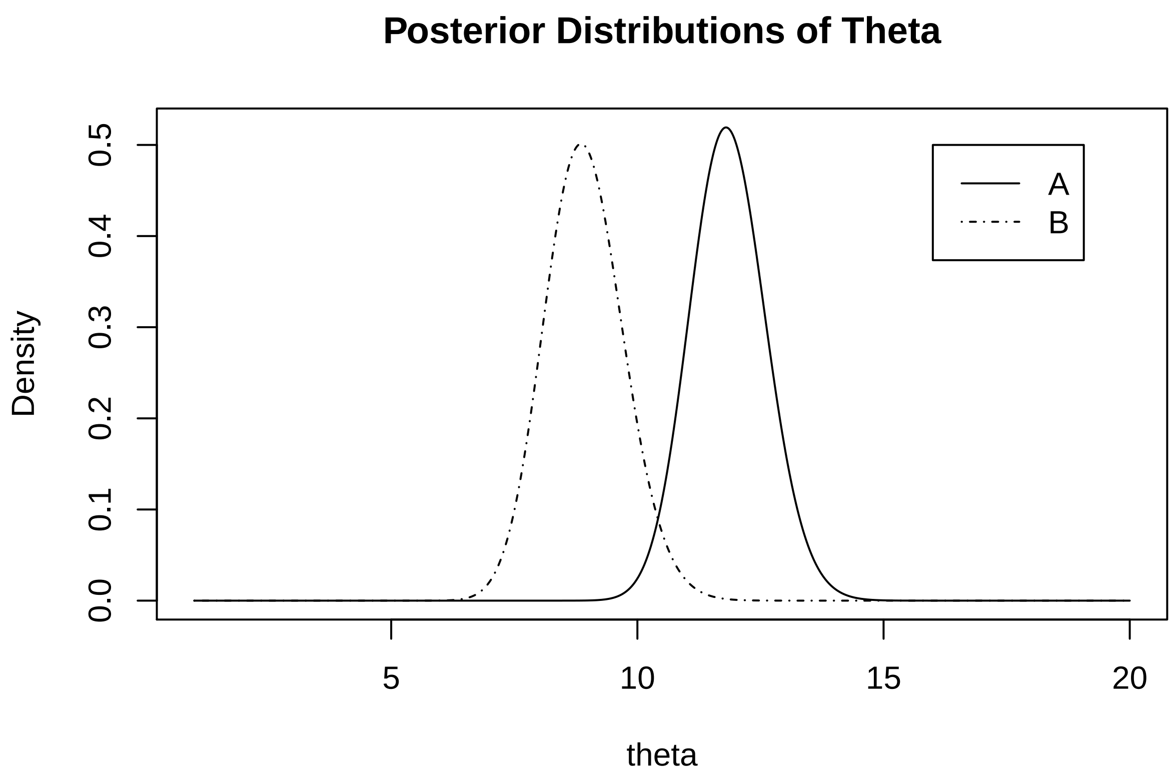
### Part a

First, we can establish our known parameters and data.

```r
# Data for y_a
y_a <- c(12, 9, 12, 14, 13, 13, 15, 8, 15, 6)

# Data for y_b
y_b <- c(11, 11, 10, 9, 9, 8, 7, 10, 6, 8, 8, 9, 7)

# Prior parameters for theta A ~ gamma(120, 10)
A_prior <- c(120,10)

# Prior parameters for theta B ~ gamma(12, 1)
B_prior <- c(12,1)

# Various values of theta
N <- seq(1,20,.01)
```

We know that a posterior distribution of $\theta$ with a gamma prior and Poisson likelihood follows a gamma distribution such that $(\theta|Y) \sim Gamma(\alpha_0 + \sum y, \beta_0 + n)$.

```r
# Posterior distribution of theta A
posterior_a <- dgamma(N, shape = A_prior[1] + sum(y_a), rate = A_prior[2] + length(y_a))

# Posterior distribution of theta B
posterior_b <- dgamma(N, shape = B_prior[1] + sum(y_b), rate = B_prior[2] + length(y_b))

# Plot the posterior distributions of A and B
plot(N, posterior_a, type = 'l', xlab = 'theta', ylab = 'Density', main =
         "Posterior Distributions of Theta")
lines(N, posterior_b, type = 'l', lty = 4)
legend(16, .5, legend = c('A','B'), lty = c(1,4))
```




```
# Posterior mean of A
(A_prior[1] + sum(y_a)) / (A_prior[2] + length(y_a))
```

```
## [1] 11.85
```

```
# Posterior variance of A
(A_prior[1] + sum(y_a)) / (A_prior[2] + length(y_a))^2
```

```
## [1] 0.5925
```

```
# 95% Conf Interval of Posterior mean A
qgamma( c(.025, .975), A_prior[1] + sum(y_a), rate = A_prior[2] + length(y_a))
```

```
## [1] 10.38924 13.40545
```

```
# Posterior mean of B
(B_prior[1] + sum(y_b)) / (B_prior[2] + length(y_b))
```

```
## [1] 8.928571
```

```
# Posterior variance of B
(B_prior[1] + sum(y_b)) / (B_prior[2] + length(y_b))^2
```

```
## [1] 0.6377551
```

```
# 95% Conf Interval of Posterior mean for B
qgamma( c(.025, .975), B_prior[1] + sum(y_b), rate = B_prior[2] + length(y_b))
```

```
## [1]  7.432064 10.560308
```

### Part b

```r
# Vector of set of values (1, 2, ... , 50)
n_0 <- 1:50

# Posterior expectations given values of n_0
expectations <- (12*n_0 + sum(y_b)) / (n_0 + length(y_b))

# Plot posterior expectations over values of n_0
plot(n_0, expectations, type = "l", xlab = "Values of n_0", ylab = "Expectation",
     main = "Expectations of Theta B")
```


We see above that the expectation slowly approaches the expectation of A as our $n_0$ gets larger. With $n_0 = 50$, our expectation is close to 11.85, the expectation of A. The exact value needed for the expectation of B to equal 11.85 is $n_0 \approx 273.67$.

### Part c

Because our beliefs state that the type B mice are related to type A mice, it doesn't make sense for the assumption that $P(\theta_A, \theta_B) = P(\theta_A) \times P(\theta_B)$ to hold. If B mice are related to A mice then we should be able to infer some information about B mice from A mice; we can intuit then that our prior beliefs about $\theta_A$ and $\theta_B$ are not independent of each other.

## Problem 3.7

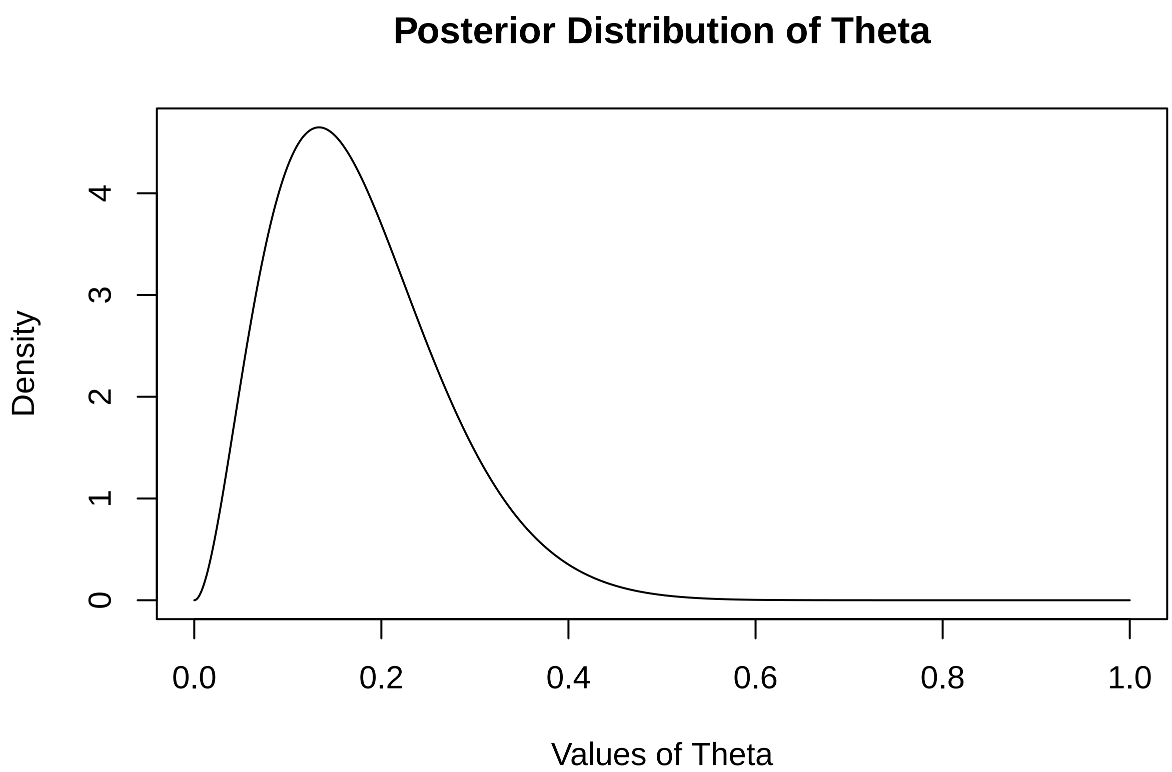
### Part a

We know that the prior distribution of $\theta$ follows a uniform distribution such that $\theta \sim U(0,1)$. However, the likelihood of the data given $\theta$, $P(Y|\theta)$ is unknown. We can use a binomial distribution for the likelihood.

$$
\begin{aligned}
P(\theta|Y) &\propto P(Y|\theta)P(\theta) \\
&\propto \binom{n}{y}\theta^y(1-\theta)^{n-y}\frac{1}{b-a+1} \\
&\propto \frac{1}{2}\binom{n}{y}\theta^y(1-\theta)^{n-y} \\
P(\theta|Y) &\propto \theta^y(1-\theta)^{n-y}
\end{aligned}
$$

This has the form of $\theta \sim Beta(y+1, n-y+1)$, thus our posterior distribution follows a beta distribution.

```
n1 <- 15
y1 <- 2

alpha <- y1 + 1
beta <- n1 - y1 + 1

thetas <- seq(0,1,0.001)

post <- dbeta(thetas, alpha, beta)

plot(thetas, post, type = "l", xlab = "Values of Theta", ylab = "Density",
     main = "Posterior Distribution of Theta")
```




```r
# Posterior mean
alpha / (alpha + beta)
```

```
## [1] 0.1764706
```

```r
# Posterior standard deviation
sqrt( (alpha*beta) / ((alpha+beta)^2 *(alpha+beta+1)) )
```

```
## [1] 0.08985443
```

```r
# Posterior mode
(alpha - 1) / (alpha + beta - 2)
```

```
## [1] 0.1333333
```

### Part b

#### i.

We know that if we are infering information about $Y_2$ from $Y_1$ then the two must not be independent of each: if they were independent, then $Y_1$ wouldn't give us any additional useful information for prediction. So $P(Y_2|Y_1) \neq P(Y_2)P(Y_1)$.

**ii. & iii.**

$$P(Y_2 = y_2|Y_1 = 2) = \int_0^1 P(Y_2 = y_2|\theta)P(\theta|Y_1 = 2)d\theta$$

$$= \int_0^1 \binom{n}{y_2}\theta^{y_2}(1-\theta)^{n-y_2}\frac{\theta^{Y_1}(1-\theta)^{n-Y_1}}{B(Y_1+1, n-Y_1+1)}d\theta$$

$$= \binom{n_2}{y_2}\frac{1}{B(3, n_1-1)}\int_0^1 \theta^{y_2}(1-\theta)^{n_2-y_2}\theta^2(1-\theta)^{n_1-2}$$

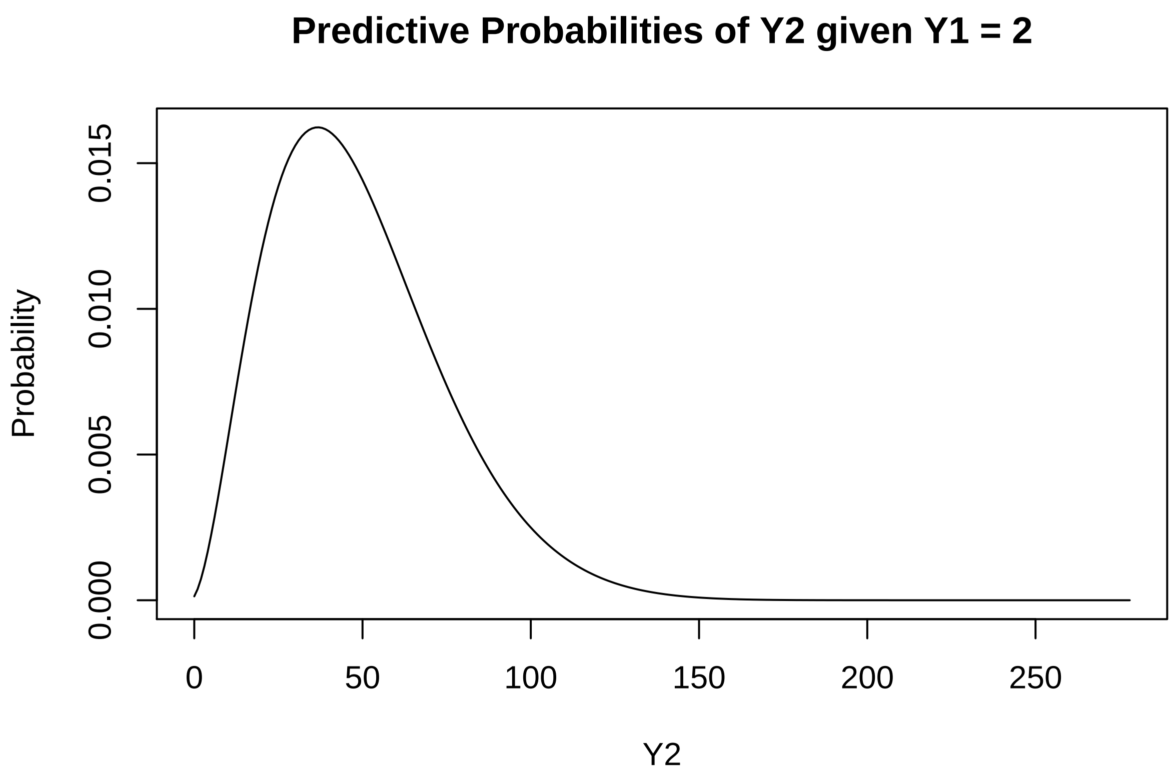
$$= \binom{n_2}{y_2}\frac{1}{B(3, n_1-1)}\int_0^1 \theta^{y_2+2}(1-\theta)^{n_2-y_2+n_1-2}$$

$$= \binom{n_2}{y_2}\frac{B(y_2+3, n_2-y_2+n_1-1)}{B(3, n_1-1)}$$

$$P(Y_2 = y_2|Y_1 = 2) = \binom{278}{y_2}\frac{B(y_2+3, 278-y_2+14)}{B(3, 14)}$$

## Part c

```
pred_dist <- function(y2){

    prob <- choose(278, y2) * ( (beta(y2 + 3, 278 - y2 + 14)) / (beta(3, 14)))

    return(prob)
}

y2_range <- 0:278

plot(y2_range, pred_dist(y2_range), type = "l", xlab = "Y2", ylab = "Probability",
     main = "Predictive Probabilities of Y2 given Y1 = 2")
```




The predictive distribution $P(Y_f|Y_1) \sim BetaBinomial(n, y_f + \alpha, n - y_f + \beta)$. This has well defined moments: if $X$ is beta-binomial then the $\mathbb{E}(X) = \frac{n\alpha}{\alpha+\beta}$ and $Var(X) = \frac{n\alpha\beta(\alpha+\beta+n)}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

$$(Y_2 = y_2|Y_1 = 2) \sim BetaBinomial(278, y_2 + 3, 278 - y_2 + 13)$$

$$\mathbb{E}(Y_2 = y_2|Y_1 = 2) = \frac{278 * 3}{3 + 14}$$

$$= 49.05882$$

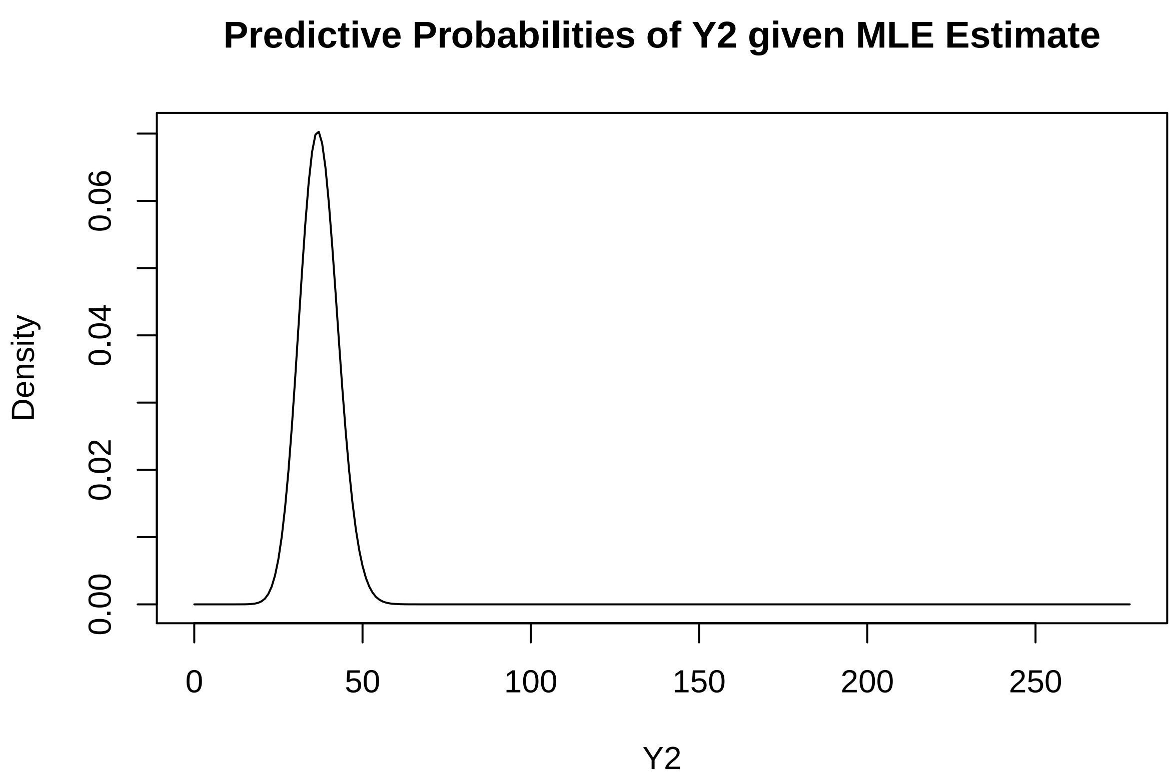
$$Var(Y_2 = y_2|Y_1 = 2) = \frac{278 * 3 * 14(3 + 14 + 278)}{(3 + 14)^2(3 + 14 + 1)}$$

$$= 662.1338$$

So the predictive distribution has a mean of 49.05882 and a standard deviation of 25.73196.

### Part d

```
# Plot binomial density with  ~ Bin(y2, 278, 2/15)
plot(y2_range, dbinom(y2_range, 278, 2/15), type = "l", xlab = "Y2", ylab = "Density",
     main = "Predictive Probabilities of Y2 given MLE Estimate")
```




```r
# Mean of the binomial distribution  -> np
278 * 2/15
```

```
## [1] 37.06667
```

```r
#Standard deviation of this binomial distribution  -> sqrt(np(1-p))
sqrt(278 * 2/15 * (1 - 2/15))
```

```
## [1] 5.667843
```

The predictive posterior distribution has more variance than that of the MLE. If you were very confident in the prior result then the MLE might be a good prediction. However, the predictive posterior distribution (beta-binomial) is a very robust predictive distribution centered around 50. Because of the small sample size in the previous sample, the predictive posterior distribution may be a better predictive model.

## Problem 3.9

### Part a

We need to derive the posterior distribution of $\theta$ to determine which class of conjugate priors is valid. We know that if $a$ is known and $Y$ follows a Galenshore distribution:

$$P(\theta|Y) \propto \frac{P(Y|\theta)P(\theta)}{P(Y)}$$

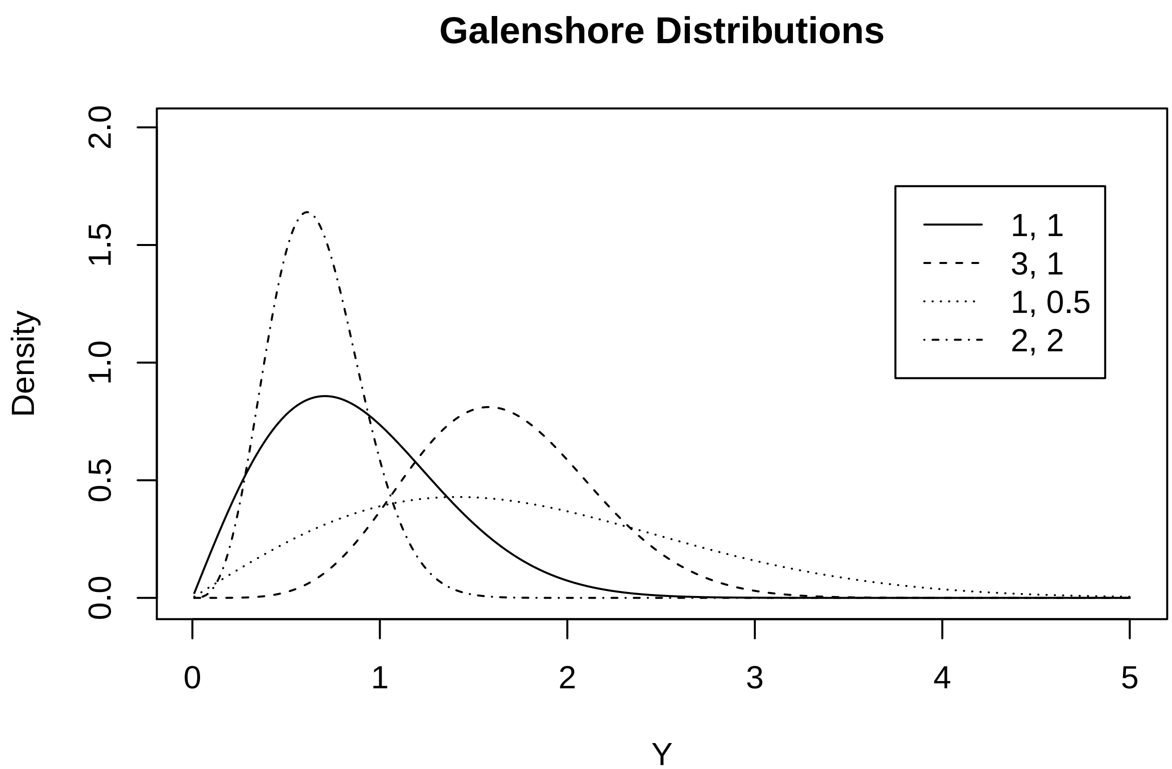
$$\propto P(Y|\theta)P(\theta)$$

$$\propto \frac{2}{\Gamma(a)}\theta^{2a}y^{2a-1}e^{-\theta^2 y^2}P(\theta)$$

$$P(\theta|Y) \propto \theta^{2a}e^{-\theta^2 y^2}P(\theta)$$

The prior $P(\theta)$ must have kernel that includes a term raised to the $2a-1$ and an exponential; the Galenshore distribution with prior parameters $a_0, \theta_0$ could be used with $\theta$ given that $a$ is known.

```
pgalenshore <- function(y, a, theta){

    p <- (2/gamma(a)) * theta^(2*a) * y^(2*a-1) * exp(-1*theta^2 * y^2)

    return(p)
}

ys <- seq(.01, 5, 0.01)

plot(ys, pgalenshore(ys, 1, 1), type =  "l", xlab = "Y", ylab = "Density",
     main = "Galenshore Distributions", ylim = c(-.01,2))
lines(ys, pgalenshore(ys, 3, 1), lty = 2)
lines(ys, pgalenshore(ys, 1, .5), lty = 3)
lines(ys, pgalenshore(ys, 2, 2), lty = 4)
legend(3.75, 1.75, legend = c('1, 1','3, 1', "1, 0.5", "2, 2"),
       lty = c(1,2,3,4))
```




### Part b

Our prior distribution follows Galenshore$(a_0, \theta_0)$ such that:

$$P(\theta) = \frac{2}{\Gamma(a_0)} \theta_0^{2a_0} \theta^{2a_0-1} e^{-\theta_0^2\theta^2}$$

So our prior distribution is as follows:

$$P(\theta|Y) \propto \theta^{2a} e^{-\theta^2 y^2} \frac{2}{\Gamma(a_0)} \theta_0^{2a_0} \theta^{2a_0-1} e^{-\theta_0^2\theta^2}$$

$$P(\theta|Y_1, ..., Y_n) \propto \theta^{2a_0-1} e^{-\theta_0^2\theta^2} \prod_{i=1}^{n} \theta^{2a} e^{-\theta^2 y_i^2}$$

$$\propto \theta^{2a_0-1} e^{-\theta_0^2\theta^2} \theta^{2an} e^{-\theta^2 \sum_{i=1}^n y_i^2}$$

$$\propto \theta^{2a_0-1+2an} e^{-\theta_0^2\theta^2 - \theta^2 \sum_{i=1}^n y_i^2}$$

$$P(\theta|Y_1, ..., Y_n) \propto \theta^{2(a_0+an)-1} e^{-\theta^2(\theta_0^2 + \sum_{i=1}^n y_i^2)}$$

So from here we see that the above has the kernel of a Galenshore distribution:

$$P(\theta|Y_1, ..., Y_n) \sim Galenshore(a_0 + an, \sqrt{\theta_0^2 + \sum_{i=1}^n y_i^2})$$

## Part c

$$\frac{P(\theta_a|Y_1,...,Y_n)}{P(\theta_b|Y_1,...,Y_n)} = \frac{\frac{2}{\Gamma(a_0+an)}(\theta_0^2 + \sum_{i=1}^n y_i^2)^{2(a_0+an)}\theta_a^{2(a_0+an)-1}e^{-\theta_a^2(\theta_0^2+\sum_{i=1}^n y_i^2)}}{\frac{2}{\Gamma(a_0+an)}(\theta_0^2 + \sum_{i=1}^n y_i^2)^{2(a_0+an)}\theta_b^{2(a_0+an)-1}e^{-\theta_b^2(\theta_0^2+\sum_{i=1}^n y_i^2)}}$$

$$= \frac{\theta_a^{2(a_0+an)-1}e^{-\theta_a^2(\theta_0^2+\sum_{i=1}^n y_i^2)}}{\theta_b^{2(a_0+an)-1}e^{-\theta_b^2(\theta_0^2+\sum_{i=1}^n y_i^2)}}$$

$$= (\frac{\theta_a^2}{\theta_b^2})^{2(a_0+an)-1}e^{(-\theta_a^2(\theta_0^2+\sum_{i=1}^n y_i^2))-(-\theta_b^2(\theta_0^2+\sum_{i=1}^n y_i^2))}$$

$$= (\frac{\theta_a^2}{\theta_b^2})^{2(a_0+an)-1}e^{-(\theta_0^2+\sum_{i=1}^n y_i^2)(\theta_a^2-\theta_b^2)}$$

$$\frac{P(\theta_a|Y_1,...,Y_n)}{P(\theta_b|Y_1,...,Y_n)} = (\frac{\theta_a^2}{\theta_b^2})^{2(a_0+an)-1}e^{-\theta_0^2(\theta_a^2-\theta_b^2)}e^{-\sum_{i=1}^n y_i^2(\theta_a^2-\theta_b^2)}$$

Thus $\sum_{i=1}^n y_i^2$ is a sufficient statistic for the posterior distribution of $\theta_a$ given data divided by that of $\theta_b$ as it is the only instance of $Y_1,...,Y_n$ and contains all information from the data.

## Part d

We know that if $X$ follows a Galenshore distribution such that $X \sim Galenshore(a,\theta)$ then the expectation $\mathbb{E}(X) = \frac{\Gamma(a+1/2)}{\theta\Gamma(a)}$. The posterior distribution of $\theta$ given some data, $P(\theta|Y_1,...,Y_n)$, follows a Galenshore distribution of the form $Galenshore(a_0+an, \sqrt{\theta_0^2 + \sum_{i=1}^n y_i^2})$. We can derive the expectation of the posterior knowing these properties:

$$\mathbb{E}(\theta|Y_1,...,Y_n) = \frac{\Gamma((a_0+an)+1/2)}{\Gamma(a_0+an)\sqrt{\theta_0^2+\sum_{i=1}^n y_i^2}}$$

## Part e

When $a$ is known, the predictive posterior distribution of $Y_f$, $P(Y_f|Y_1,...,Y_n)$ follows:

$$P(Y_f|Y_1,...,Y_n) = \int_0^\infty P(Y_f|\theta,Y_1,...,Y_n)P(\theta|Y_1,...,Y_n)d\theta$$

$$= \int_0^\infty P(Y_f|\theta)P(\theta|Y_1,...,Y_n)d\theta$$

$$= \int_0^\infty \frac{2}{\Gamma(a)}\theta^{2a}y_f^{2a-1}e^{-\theta^2 y_f^2}\frac{2}{\Gamma(a_0+an)}(\theta_0^2+\sum_{i=1}^n y_i^2)^{2(a_0+an)}\theta^{2(a_0+an)-1}e^{-\theta^2(\theta_0^2+\sum_{i=1}^n y_i^2)}d\theta$$

$$= \frac{2}{\Gamma(a)}y_f^{2a-1}\frac{2}{\Gamma(a_0+an)}(\theta_0^2+\sum_{i=1}^n y_i^2)^{2(a_0+an)}\int_0^\infty \theta^{2a}e^{-\theta^2 y_f^2}\theta^{2(a_0+an)-1}e^{-\theta^2(\theta_0^2+\sum_{i=1}^n y_i^2)}d\theta$$

$$= \frac{2}{\Gamma(a)}y_f^{2a-1}\frac{2}{\Gamma(a_0+an)}(\theta_0^2+\sum_{i=1}^n y_i^2)^{2(a_0+an)}\int_0^\infty \theta^{2a+2(a_0+an)-1}e^{-\theta^2 y_f^2-\theta^2(\theta_0^2+\sum_{i=1}^n y_i^2)}d\theta$$

$$P(Y_f|Y_1,...,Y_n) = \frac{2}{\Gamma(a)}y_f^{2a-1}\frac{2}{\Gamma(a_0+an)}(\theta_0^2+\sum_{i=1}^n y_i^2)^{2(a_0+an)}\int_0^\infty \theta^{2(a+a_0+an)-1}e^{-\theta^2(y_f^2+\theta_0^2+\sum_{i=1}^n y_i^2)}d\theta$$

The above kernel has the form of $Galenshore(a+a_0+an, \sqrt{y_f^2+\theta_0^2+\sum_{i=1}^n y_i^2})$ so with the addition of two terms, it integrates to 1. We use this to reduce $P(Y_f|Y_1,...,Y_n)$ below:

$$P(Y_f|Y_1,...,Y_n) = \frac{2}{\Gamma(a)}y_f^{2a-1}\frac{2}{\Gamma(a_0+an)}(\theta_0^2+\sum_{i=1}^n y_i^2)^{2(a_0+an)}\frac{\Gamma(a+a_0+an)}{2}(y_f^2+\theta_0^2+\sum_{i=1}^n y_i^2)^{a+a_0+an}$$

$$P(Y_f|Y_1,...,Y_n) = \frac{2\Gamma(a+a_0+an)}{\Gamma(a)\Gamma(a_0+an)}y_f^{2a-1}(\theta_0^2+\sum_{i=1}^n y_i^2)^{2(a_0+an)}(y_f^2+\theta_0^2+\sum_{i=1}^n y_i^2)^{a+a_0+an}$$

This reduces to value that has no unknown parameters (i.e. $\theta$).