Homework 7

Spring 2023

Jaiden Atterbury

Instructions

- This homework is due in Gradescope on Wednesday May 24 by midnight PST.
- Please answer the following questions in the order in which they are posed.
- Don't forget to knit the document frequently to make sure there are no compilation errors.
- When you are done, download the PDF file as instructed in section and submit it in Gradescope.

Please note: Hints have been given on several problems. This is to encourage you to problem solve on your own.

Exercises

1. (Binomial-Poisson hierarchical model) The number of eggs, X, laid by an insect is a random variable which is often taken to be $Pois(\lambda_0)$. That is, the marginal distribution of X is

$$P(X = x) = \frac{\lambda_0^x e^{-\lambda_0}}{x!},\ x = 0, 1, 2, \ldots$$

Furthermore, if we let Y = number of survivors, a common modeling assumption is that given that there are x eggs laid, Y is a binomial random variable. In other words:

$$P(Y = y|X = x) = Binom(x, \pi_0)$$

where π_0 is the unknown probability of survival.

a. Derive the joint PMF $f(x, y) = P(X = x, Y = y)$ showing your steps. Don't forget to state the possible values of x and y - these define the range of values for which the joint is non-zero.

Let X denote the number of eggs laid by a given insect. It follows that $X \sim Pois(\lambda_0)$. Hence the marginal distribution of X is $P(X = x) = f_1(x) = \frac{\lambda_0 e^{-\lambda_0}}{x!},\ x = 0, 1, 2, \ldots$. Furthermore, if we let Y denote the number of survivors out of the X eggs laid by the insect, then it follows that given there are x eggs, Y is a binomial random variable with conditional distribution $P(Y = y|X = x) = f(y|x) = \binom{x}{y}\pi_0^y(1 - \pi_0)^{x-y},\ y = 1, 2, 3, \ldots, x$. With that said, we are looking to derive the joint PMF $f(x, y) = P(X = x, Y = y)$.

To do this we will use Definition 14.3 which states that $f(y|x) = P(Y = y|X = x) = \frac{P(X=x,Y=y)}{P(X=x)} = \frac{f(x,y)}{f_1(x)}$. Hence, since we have the conditional distribution of Y given X, $f(y|x)$, and the marginal distribution of X, $f_1(x)$, it follows that we can simply manipulate the given Definition and find the form we need. It turns out

that this form is $f(x,y) = f(y|x)f_1(x)$. Hence we can write the join PMF $f(x,y) = P(X = x, Y = y)$ as follows

$$\begin{aligned} f(x,y) &= P(X = x, Y = y) \\ &= f(y|x)f_1(x) \\ &= \binom{x}{y}\pi_0^y(1-\pi_0)^{x-y} \cdot \frac{\lambda_0 e^{-\lambda_0}}{x!},\ x = 0,1,2,\ldots,\ y = 1,2,3,\ldots,x \end{aligned}$$

Thus, as showed above, the joint PMF $f(x,y) = P(X = x, Y = y) = \binom{x}{y}\pi_0^y(1-\pi_0)^{x-y} \cdot \frac{\lambda_0 e^{-\lambda_0}}{x!},\ x = 0,1,2,\ldots,\ y = 1,2,3,\ldots,x$.

b. Find $P(Y = y)$, the marginal PMF of Y. Is it a familiar distribution? State the values of the parameter(s) of the distribution.

- you will sum the joint distribution over x (think about the values you will sum over)
- you will make a change of variable $u = x - y$ in the summation

Our goal for this problem is to find $P(Y = y)$, the marginal PMF of Y. To do this we will use Definition 14.2 which states that $P(Y = y) = f_2(y) = \sum_x f(x,y)$. Thus, to find the marginal distribution we will need to sum over the joint distribution x, then we will need to make a change of variable $u = x - y$ in the summation. This process is shown below.

$$\begin{aligned} f_2(y) &= P(Y = y) \\ &= \sum_x f(x,y) \\ &= \sum_x f(y|x)f_1(x) \\ &= \sum_x \binom{x}{y}\pi_0^y(1-\pi_0)^{x-y} \cdot \frac{\lambda_0 e^{-\lambda_0}}{x!} \\ &= \sum_{x=y}^{\infty} \binom{x}{y}\pi_0^y(1-\pi_0)^{x-y} \cdot \frac{\lambda_0 e^{-\lambda_0}}{x!} \quad \text{(Since } f(y|x) \text{ is zero for } y > x\text{)} \\ &= \frac{x!}{y!(x-y)!}\pi_0^y(1-\pi_0)^{x-y} \cdot \frac{\lambda_0 e^{-\lambda_0}}{x!} \\ &= \frac{(\pi_0 \cdot \lambda_0)^y e^{-\lambda_0}}{y!} \sum_{x=y}^{\infty} \frac{(1-\pi_0)^{x-y}\lambda_0^{x-y}}{(x-y)!} \\ &= \frac{(\pi_0 \cdot \lambda_0)^y e^{-\lambda_0}}{y!} \sum_{u=0}^{\infty} \frac{(1-\pi_0)^{u}\lambda_0^{u}}{u!} \quad \text{(Let } u = x - y\text{)} \\ &= \frac{(\pi_0 \cdot \lambda_0)^y e^{-\lambda_0}}{y!} \sum_{u=0}^{\infty} \frac{(\lambda_0 - \lambda_0\pi_0)^{u}}{u!} \\ &= \frac{(\pi_0 \cdot \lambda_0)^y e^{-\lambda_0} e^{\lambda_0 - \lambda_0\pi_0}}{y!} \quad \text{(Since } \sum_{n=0}^{\infty} \frac{x^n}{n!} = e^x\text{)} \\ &= \frac{(\pi_0\lambda_0)^y e^{\pi_0\lambda_0}}{y!},\ y = 0,1,2,\ldots \end{aligned}$$

As can be seen above, $f_2(y) = \frac{(\pi_0\lambda_0)^y e^{\pi_0\lambda_0}}{y!},\ y = 0,1,2,\ldots$. Notice that this is a familiar distribution, namely $Y \sim Pois(\pi_0\lambda_0)$.

c. On the average, how many eggs will survive? That is, what is $E[Y]$ and why does the answer make sense intuitively?

Based on Lemma 8.1, if $Y \sim Pois(\lambda)$, then $E[Y] = \lambda$. Thus $E[Y] = \pi_0\lambda_0$. Hence, on the average, $\pi_0\lambda_0$ eggs will survive. This answer makes sense intuitively because on average, the number of eggs you'd expect to survive is the average number of eggs laid by this insect times the probability that any given egg survives. In particular, this is exactly what this expected value is calculating.

2. (Normal tolerance) Suppose we sample X from a Normal distribution with mean 0 and tolerance $\tau = \frac{1}{\sigma^2}$. In the Bayesian context, we treat τ as a random variable.

a. Write the PDF of X indexed by τ_0, a specific value for τ. We think of this as the conditional density of X given $\tau = \tau_0$. (Hint: write the usual Normal PDF but re-parametrized in terms of τ, not σ.)

Suppose we sample X values from a Normal distribution with mean 0 and tolerance $\tau = \frac{1}{\sigma^2}$. Unlike in the Frequentist approach, in the Bayesian context, we treat τ as a random variable. The goal for this part of the problem is to write the PDF of X indexed by τ_0, which is a specific value of τ. In particular, we think of this as the conditional density of X given $\tau = \tau_0$. To do this we will write the usual Normal PDF, but instead re-parametrized it in terms of τ, not σ. Based on Lemma 13.1, if $X \sim Norm(\mu, \sigma)$, then $f(x) = \frac{e^{-(x-\mu)^2/(2\sigma^2)}}{\sigma\sqrt{2\pi}}$, $-\infty < x < \infty$. In our case $\mu = 0$ and $\tau_0 = \frac{1}{\sigma_0^2} \implies \sigma_0^2 = \frac{1}{\tau_0} \implies \sigma_0 = \sqrt{\frac{1}{\tau_0}}$. We will now plug these value into the normal PDF and simplify.

$$\begin{aligned} f(x|\tau_0) &= Norm\left(0, \sqrt{\frac{1}{\tau_0}}\right) \\ &= \frac{e^{-(x-0)^2/(2\frac{1}{\tau_0})}}{\sqrt{\frac{1}{\tau_0}\sqrt{2\pi}}} \\ &= \frac{e^{-\tau_0 x^2/2}}{\sqrt{\frac{2\pi}{\tau_0}}} \\ &= \sqrt{\frac{\tau_0}{2\pi}} e^{-\tau_0 x^2/2}, -\infty < x < \infty \end{aligned}$$

Hence as computed above, the conditional density of X given $\tau = \tau_0$ is $f(x|\tau_0) = \sqrt{\frac{\tau_0}{2\pi}} e^{\frac{-\tau_0 x^2}{2}}, -\infty < x < \infty$.

b. Suppose we assume τ is a Gamma random variable, that is our prior distribution is $g(\tau_0) = Gamma(\alpha_0, \lambda_0)$[1] where α_0 is the shape and λ_0 is the rate parameter. Determine the form of the posterior distribution $h(\tau_0|x)$. Is it a familiar distribution? State the values for the parameters of the distribution.

Suppose we assume τ is a Gamma random variable, namely our prior distribution is $g(\tau_0) = Gamma(\alpha_0, \lambda_0)$, where α_0 is the shape and λ_0 is the rate parameter. Our goal for this problem is to find the posterior distribution $h(\tau_0|x)$. Since we know that our data X is $Norm\left(0, \sqrt{\frac{1}{\tau_0}}\right)$, and that our prior of τ_0 is $Gamma(\alpha_0, \lambda_0)$, we can use the fact that the posterior is proportional to the likelihood times the prior in order to find the posterior. Once we have this equation we will "recognize" the form of the kernel density and find the particular parameter values of the posterior.

First off, in part (a) we found that the likelihood function $f(x|\tau_0) = \sqrt{\frac{\tau_0}{2\pi}} e^{\frac{-\tau_0 x^2}{2}}$ which can be rewritten as $f(x|\tau_0) = \sqrt{\frac{1}{2\pi}} \tau_0^{\frac{1}{2}} e^{\frac{-\tau_0 x^2}{2}}$. Furthermore, by Definition 13.5 we know that a gamma random variable has the

[1] see Definition 13.5 in the NOTES

PDF $f(x) = \frac{\lambda^k}{\Gamma(k)} x^{k-1} e^{-\lambda x},\ 0 < x < \infty$, where k is the shape parameter and λ is the rate parameter. Thus in our case the prior distribution of τ_0 is $g(\tau_0) = \frac{\lambda_0^{\alpha_0}}{\Gamma(\alpha_0)} \tau_0^{\alpha_0 - 1} e^{\lambda_0 \tau_0}$.

Hence, after removing the constant terms from each of these PDFs, it follows that the posterior is proportional to $\tau_0^{\frac{1}{2}} e^{\frac{-\tau_0 x^2}{2}} \tau_0^{\alpha_0 - 1} e^{\lambda_0 \tau_0}$. This simplifies nicely to $\tau_0^{\alpha_0 - \frac{1}{2}} e^{-(\frac{1}{2}x^2 + \lambda_0)\tau_0}$. Thus, as can be seen from the previous expression, the posterior takes the form of a Gamma distribution. In particular it follows that the posterior, $h(\tau_0|x)$, has the PDF of $Gamma(\alpha_0 + \frac{1}{2}, \frac{1}{2}x^2 + \lambda_0)$.

3. (False Discovery) Are many medical discoveries actually Type 1 errors? In medical research, suppose that 10% of null hypotheses are actually false, and that when a null hypothesis is false, the chance of making a Type II error and failing to reject it (for example, due to insufficient sample size) is 0.55.

    a. Given that we reject a null hypothesis at level $\alpha = 0.05$, use Bayes' Rule to show that 50% of such studies are actually reporting Type I errors. (Please look up chapter 20 in the NOTES for definitions of Type 1 and Type 2 errors.)

    Hints:

    - Define events H: "null is false" and D: "do not reject the null".
    - You are given various probabilities. For example, $P(H) = 0.1$.
    - The level of significance α is the Type 1 error rate of the significance test. That is, it is the conditional probability $P(D^c|H^c)$.
    - You want to calculate $P(H^c|D^c)$. That is, of all the times when you reject the hypothesis, how often is the null actually true?

Suppose that in medical research, 10% of null hypotheses are false. Furthermore, suppose when a null hypothesis is false, the chance of making a Type II error and failing to reject the null is 0.55. Given that we reject a null hypothesis at level $\alpha = 0.05$, the goal of this problem is to use Bayes' Rule to show that 50% of such studies are actually reporting Type I errors.

To clearly solve this problem, we need to define some events. First off, we define the events H : the null hypothesis is false, and D : we don't reject the null hypothesis. Next we identify that the level of significance α is the Type 1 error rate of the significance test. This translates to the conditional probability $P(D^c|H^c)$, which we know equals 0.05 in our case. With that said, we want to calculate $P(H^c|D^c)$. That is, out of all the times when we reject the hypothesis, how often is the null actually true? By Bayes' Theorem, we can express $P(H^c|D^c)$, as $\frac{P(D^c|H^c) \cdot P(H^c)}{P(D^c)}$.

We will now do some manipulations of the previous probability statement in order to identify it in terms of probabilities we already know. Namely, we know $P(D^c)$ is the probability of the union of two disjoint events $P((D^c \cap H^c) \cup (D^c \cap H))$. By Definition 2.1.A3 we know that the probability of two disjoint events is simply the sum of the probabilities of the two events, thus we can rewrite this as $P(D^c \cap H^c) + P(D^c \cap H)$. Furthermore, using the chain rule of probability we can change the above statement to $P(D^c|H^c) \cdot P(H^c) + P(D^c|H) \cdot P(H)$. Hence our main probability statement becomes $\frac{P(D^c|H^c) \cdot P(H^c)}{P(D^c|H^c) \cdot P(H^c) + P(D^c|H) \cdot P(H)}$. However, we know what all of these probabilities equal, namely, $P(H) = 0.1$, $P(H^c) = 0.9$, $P(D^c|H^c) = 0.05$, and lastly $P(D^c|H) = 0.45$. Plugging these back into the previous probability statement we obtain $P(H^c|D^c) = \frac{0.15 \cdot 0.9}{0.05 \cdot 0.9 + 0.1 \cdot 0.45} = 0.5$. Hence our result has been shown.

    b. The probability you found in a. is called a *False Discovery Rate (FDR)*. Your calculation shows that even though we control the Type 1 error rate at 0.05, the FDR can be high.

Write a function in R to perform the FDR computation in part a. which takes as input: the probability of H, the Type 1 and Type 2 error rates, and calculates the FDR. Then run it for all combinations of the following inputs:

- $P(H)$: 0.05, 0.1, 0.2, 0.5
- Type 1 error rate: 0.0001, 0.005, 0.01, 0.05
- Type 2 error rate: 0.05, 0.2, 0.3, 0.5

  Make a table of the resulting FDRs, and write a few sentences summarizing what you observe. Conclude your summary with practical advice for the data scientist who is wondering how to choose their α level for a significance test. (You can manually supply all the combinations of inputs to your function, or check out `expand_grid` to generate a data frame with one row for each combination of the inputs. See `simulating-a-poisson-process.Rmd` from STAT 340 which I have pushed to your HW folder.)

The probability we found in part (a) is called the False Discovery Rate (FDR). Based on our calculations above, even when we control the Type I error rate to be 0.05, the FDR can still be high. Below we will write a function in R that performs the FDR computation from part (a). In particular, this function will take in the probability H, the Type I and Type II error rates, and calculate/return the FDR. After this function is created we will run it on for all combinations of the following inputs: $P(H)$: c(0.05, 0.1, 0.2, 0.5), Type I error rate: c(0.0001, 0.005, 0.01, 0.05), and Type II error rate: c(0.05, 0.2, 0.3, 0.5). After this we will make a table of the resulting FDRs in order to make conclusions about what we observe, and lastly make a concrete decision about which Type I error rate/α level the medical researchers should choose for a significance test.

Function and Table of the FDRs:

```
# Function to calculate the FDR given P(H), P(Type I), and P(Type II):
fdr_function <- function(h, type1, type2) {
  hc = 1 - h
  power = 1 - type2
  fdr = hc*type1 / (hc*type1 + h*power)
}

# Create the DataFrame of inputs for the function:
fdr_table <- expand_grid("P(H)" = c(0.05, 0.1, 0.2, 0.5),
                         "P(Type I)" = c(0.0001, 0.005, 0.01, 0.05),
                         "P(Type II)" = c(0.05, 0.2, 0.3, 0.5))

# Initialize and empty column for the FDRs in this DataFrame:
fdr_table["FDR"] = rep(0, 64)

# Calculate the FDR for each combination of probabilities:
for (index in 1:length(fdr_table$`P(H)`)) {
  fdr_table[index, 4] = fdr_function(fdr_table[index, 1],
                                     fdr_table[index, 2],
                                     fdr_table[index, 3])
}
```

Table 1: False discovery rate for varying inputs

	Type 1 error							
	0.001				0.005			
$P(H)$	Type II error				Type II error			
	0.05	0.2	0.3	0.5	0.05	0.2	0.3	0.5
0.05	0.002	0.0024	0.0027	0.0038	0.0909	0.1061	0.1195	0.1597
0.10	0.0009	0.0011	0.0013	0.0018	0.0452	0.0533	0.0604	0.0826
0.20	0.0004	0.0005	0.0006	0.0008	0.0206	0.0244	0.0278	0.0385
0.50	0.0001	0.0001	0.0001	0.0002	0.0052	0.0062	0.0071	0.0099
	Type 1 error							
	0.01				0.05			
$P(H)$	Type II error				Type II error			
	0.05	0.2	0.3	0.5	0.05	0.2	0.3	0.5
0.05	0.1667	0.1919	0.2135	0.2754	0.5	0.5429	0.5758	0.6552
0.10	0.0865	0.1011	0.1139	0.1525	0.3214	0.36	0.3913	0.4737
0.20	0.0404	0.0476	0.0541	0.0741	0.1739	0.2	0.2222	0.2857
0.50	0.0104	0.0123	0.0141	0.0196	0.05	0.0588	0.0667	0.0909

Summary of results:
As can be seen from the above table of different FDR values for different combinations of $P(H)$, $P(Type\ I)$, and $P(Type\ II)$, there were three main observations I made. First off, in general, as $P(H)$ increases, over all values of $P(Type\ I)$ and $P(Type\ II)$, the FDR decreases. This makes sense, because as the probability that the null is false increases, the less frequently we will make a mistake of rejecting the true null simply due to the fact that the null is true left often. Also, in general, as $P(Type\ I)$ increases, over all values of $P(H)$ and $P(Type\ II)$, the FDR increases. This makes sense, because as the probability that we reject the null given the null is true increases, the more often we will make a type 1 error, since this is of course by definition the type 1 error rate. Lastly, in general, as $P(Type\ II)$ decreases, over all values of $P(H)$ and $P(Type\ I)$, the FDR decreases. This make sense because as the power of our test increases (as $P(Type\ II)$ decreases), we will reject the null when it is truly false more often, thus our denominator will increase and the FDR will decrease.

Choosing α level:
To fully analyze which type 1 error rate the medical researchers should use we will find some summary statistics of the FDR grouped by the different type 1 errors.

```
# Find summary statistics of the FDR by different alpha levels:
fdr_table %>%
  group_by(`P(Type I)`) %>%
  summarise(mean = mean(FDR),
            sd = sd(FDR),
            min = min(FDR),
            max = max(FDR))
```

```
## # A tibble: 4 x 5
##   `P(Type I)`       mean         sd         min        max
##         <dbl>      <dbl>      <dbl>       <dbl>      <dbl>
## 1      0.0001 0.00117957 0.00108041 0.000105252 0.00378561
## 2      0.005  0.0535847  0.0463285  0.00523560  0.159664
## 3      0.01   0.0983854  0.0809324  0.0104167   0.275362
## 4      0.05   0.310528   0.198820   0.05        0.655172
```

As can be seen above, over our range of different type 1 errors, type 2 errors, and null being false rates, that as the type 1 error rate decreases, the average FDR value decreases, as well as the variability of said

FDR. However, as stated in Chapter 20 of the STAT 341 notes, as the type 1 error rate decreases, the type 2 error rate, in turn increases. Thus, since a type 1 error in a medical study could have severe consequences, I would recommend the medical researchers to use a type 1 error rate of 0.01 as this has a low to moderate FDR range, and requires the researchers to find something pretty significant before they can reject the null. I chose this alpha level since it is conservative, and it is a common choice in most medical studies.

4. Before a U.S. Presidential election, polls are taken in two swing states. The Republican candidate was preferred by 59 out of the 100 people sampled in state A and by 525 out of 1,000 sampled in state B.

a. If we can treat these polls as if the samples were randomly drawn from the population with a proportion π voting Republican, use a large sample Z test of $H_0 : \pi = 0.5$ versus $H_1 : \pi > 0.5$ to determine which state has greater evidence supporting a Republican victory. Show your work.

Suppose that before a U.S. Presidential election that polls are taken in two swing states. For the two states, the Republican candidate was preferred by 59 out of the 100 people sampled in state A and by 525 out of 1,000 sampled in state B. If we treat these polls as if the samples were randomly drawn from a population with a proportion π voting republican, then we can use a large sample Z test of $H_0 : \pi = 0.5$ versus $H_1 : \pi > 0.5$ in both states to determine which state has greater evidence supporting a Republican victory.

Z test for State A:
Since the data is quantitative and regarded as realizations from a Bernoulli distribution, and the data is treated as if the samples were randomly drawn from a population with a proportion π, and lastly the sample size of 100 is sufficiently large so that the sampling distribution of $\bar{X}$ is approximately normal, it turns out that our assumptions for the Z test are met.

In particular we will be testing the competing hypotheses $H_0 : \pi = 0.5$ versus $H_1 : \pi > 0.5$ at the 5% level of significance. Our tests statistic will be based on the sample proportion $\bar{X} = \frac{X}{n}$ where X is the number of people supporting the Republican candidate. In particular since the X_i's are Bernoulli it follows that $\bar{X} \approx Norm\left(\pi_0^{null}, \sqrt{\frac{\pi_0^{null}(1-\pi_0^{null})}{n}}\right)$, where n is the sample size and π_0^{null} is the specified null value. In our case $\pi_0^{null} = 0.5$ and $n = 100$, thus $\bar{X} \approx Norm\,(0.5, 0.05)$.

Since the alternative hypothesis is $H_1 : \pi > 0.5$, values in the upper tail will be in favor of the alternative. Our observed value of $\bar{X}$ is $\frac{59}{100} = 0.59$. We will now use the normal distribution to calculate our p-value in this case.

```
state_a_pval <- pnorm(q = 0.59, mean = 0.5, sd = 0.05, lower.tail = FALSE)
```

Thus as computed above the p-value is 0.0359303. Hence, since this p-value is lower than 0.05 we have significant evidence at the 5% level to reject the null hypothesis in favor of the alternative. Hence we have significant evidence that State A prefers the Republican candidate more than the other candidate.

Z test for State B:
Since the data is quantitative and regarded as realizations from a Bernoulli distribution, and the data is treated as if the samples were randomly drawn from a population with a proportion π, and lastly the sample size of 1000 is sufficiently large so that the sampling distribution of $\bar{X}$ is approximately normal, it turns out that our assumptions for the Z test are met.

In particular we will be testing the competing hypotheses $H_0 : \pi = 0.5$ versus $H_1 : \pi > 0.5$ at the 5% level of significance. Our tests statistic will be based on the sample proportion $\bar{X} = \frac{X}{n}$ where X is the number of people supporting the Republican candidate. In particular since the X_i's are Bernoulli it follows that $\bar{X} \approx Norm\left(\pi_0^{null}, \sqrt{\frac{\pi_0^{null}(1-\pi_0^{null})}{n}}\right)$, where n is the sample size and π_0^{null} is the specified null value. In our case $\pi_0^{null} = 0.5$ and $n = 1000$, thus $\bar{X} \approx Norm\left(0.5, \sqrt{0.00025}\right)$.

Since the alternative hypothesis is $H_1 : \pi > 0.5$, values in the upper tail will be in favor of the alternative. Our observed value of $\bar{X}$ is $\frac{525}{1000} = 0.525$. We will now use the normal distribution to calculate our p-value in this case.

```
state_b_pval <- pnorm(q = 0.525, mean = 0.5, sd = sqrt(0.00025), lower.tail = FALSE)
```

Thus as computed above the p-value is 0.0569231. Hence, since this p-value is greater than 0.05 we do not have significant evidence at the 5% level to reject the null hypothesis in favor of the alternative. In particular, we fail to reject the null hypothesis at the 5% level. Hence we can't say with any assurance that State A prefers the Republican candidate more than the other candidate.

Therefore, based on the Frequentist framework, since we rejected the null hypothesis at the 5% level for State A, and failed to reject the null hypothesis at the 5% level for State B, we have shown that State A has greater evidence supporting a Republican victory than State B does.

b. Conduct a Bayesian analysis to answer the question in part a. by finding in each case the posterior probability of the null hypothesis: $P(\pi \leq 0.5)$. Use a beta prior which has mean 0.5 and standard deviation 0.05. Explain any differences between conclusions. (This is an open ended question)

Using a beta prior which has mean 0.5 and standard deviation 0.05, in order to conduct a Bayesian analysis to answer the question in part (a) by finding in each case the posterior probability of the null hypothesis: $P(\pi \leq 0.5)$, we must first find the parameters of the prior Beta distribution which takes the form of $\pi_0 \sim Beta(\alpha_0, \beta_0)$.

First off, we know from Problem Section 7 the following three facts: If $X \sim Beta(\alpha, \beta)$, then $E[X] = \frac{\alpha}{\alpha+\beta}$, $Var[X] = \frac{\alpha}{\alpha+\beta} \cdot \frac{\beta}{\alpha+\beta} \cdot \frac{1}{1+\alpha+\beta}$, and lastly $\frac{\beta}{\alpha+\beta} = 1 - \frac{\alpha}{\alpha+\beta}$. Thus in our case, $E[\pi] = \frac{\alpha_0}{\alpha_0+\beta_0} = 0.5$, and $Var[\pi] = (E[\pi] - E[\pi]^2) \cdot \frac{1}{1+\alpha_0+\beta_0} = 0.0025$, which simplifies to $Var[\pi] = \frac{0.25}{1+\alpha_0+\beta_0} = 0.0025$.

If we solve $\frac{0.25}{1+\alpha_0+\beta_0} = 0.0025$ in terms of α_0, we obtain $\alpha_0 + \beta_0 + 1 = 100 \implies \alpha_0 = 99 - \beta_0$. Plugging this expression of α_0 into the equation $\frac{\alpha_0}{\alpha_0+\beta_0} = 0.5$, we obtain $\frac{\alpha_0}{\alpha_0+\beta_0} = 0.5$ which simplifies to $\frac{99-\beta_0}{99} = 0.5$. Solving for β_0 we obtain $\beta_0 = 49.5$. Lastly, plugging this back into our expression for α_0 we obtain $\alpha_0 = 99 - 49.5 = 49.5$.

Bayesian Calculation for State A:
Since $f(x|\pi_0) \sim Binom(100, \pi_0)$ and $g(\pi_0) \sim Beta(49.5, 49.5)$, then using Theorem 26.2 we can see that our posterior distribution takes the form of $h(\pi_0|x) \sim Beta(\alpha_0 + x, \beta_0 + n - x)$. In particular, our posterior takes the form $h(\pi_0|x) \sim Beta(49.5 + 59, 49.5 + 100 - 59) = Beta(108.5, 90.5)$. Hence, in order to find the posterior probability of the null hypothesis: $P(\pi \leq 0.5)$, we will use the pbeta function in r with our given parameter values and observed proportion calculate in part (a).

```
post_prob_a <- pbeta(q = 0.5, shape1 = 108.5, shape2 = 90.5)
```

Thus as computed above, the posterior probability of the null hypothesis for State A is: $P(\pi \leq 0.5)$ is 0.1003527.

Bayesian Calculation for State B:
Since $f(x|\pi_0) \sim Binom(1000, \pi_0)$ and $g(\pi_0) \sim Beta(49.5, 49.5)$, then using Theorem 26.2 we can see that our posterior distribution takes the form of $h(\pi_0|x) \sim Beta(\alpha_0 + x, \beta_0 + n - x)$. In particular, our posterior takes the form $h(\pi_0|x) \sim Beta(49.5 + 525, 49.5 + 1000 - 525) = Beta(574.5, 524.5)$. Hence, in order to find the posterior probability of the null hypothesis: $P(\pi \leq 0.5)$, we will use the pbeta function in r with our given parameter values and observed proportion calculate in part (a).

```r
post_prob_b <- pbeta(q = 0.5, shape1 = 574.5, shape2 = 524.5)
```

Thus as computed above, the posterior probability of the null hypothesis for State B is: $P(\pi \leq 0.5)$ is 0.0656398.

In conclusion, the Bayesian analysis states that State A has a higher chance of the null being true than does State B. This means that State A has a higher chance of being less in favor of the Republican candidate than does State B. Hence, State B has greater evidence supporting a Republican victory than State A does according to the Bayesian analysis. This is in direct contradiction of what we learned though the Frequentist approach. This is mainly due to the fact that Bayesian's see probability as the degree of belief and because State B has more data and still has a sample proportion higher than 0.5, the Bayesian point of view has less belief that State A will be able to match what State B has already done given 900 more sample values. Hence why State A has a higher probability that $\pi_0 < 0.5$/the null hypothesis is true.