Homework 7

Spring 2023

Tanner Huck

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 be a random variable that represents the number of eggs laid by an insect. We know that the marginal distribution of X is $P(X = x) = \frac{\lambda_0^x e^{-\lambda_0}}{x!}$. Let Y be the the number of eggs that survive, can can be denoted by, $P(Y = y|X = x) = Binom(x, \pi_0)$. We first want to derive the joint PMF $f(x, y) = P(X = x, Y = y)$.

First recall the definition of conditional probability, $P(A|B) = \frac{P(A \cap B)}{P(B)}$ which is equivalent to $P(A|B) \cdot P(B) = P(A \cap B)$, for the events A and B. Applying this to X and Y,

$$\begin{aligned} f(x, y) = P(X = x, Y = y) &= P(Y = y|X = x) \cdot P(X) \\ &\text{substituing in what we know about X and Y} \\ &= \binom{x}{y} \pi_0^y (1 - \pi_0)^{x-y} \cdot \frac{\lambda_0^x e^{-\lambda_0}}{x!} \end{aligned}$$

Thus the joint PMF $f(x,y) = P(X=x, Y=y) = \binom{x}{y}\pi_0^y(1-\pi_0)^{x-y}\cdot\frac{\lambda_0^x e^{-\lambda_0}}{x!}$. Where $x = 0,1,2,\ldots$ represents the number of eggs laid and $y = 0,1,2,\ldots,x$ represents the number of eggs that survive (capped by the number of eggs laid, 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

Now we want to find the marginal PMF of Y. Staring with the definition of a joint distribution,

$$f_2(y) = \sum_x f(x,y)$$

substituting in what we know about $f(x,y)$ from part a,

$$= \sum_{x=y}^{\infty}\binom{x}{y}\pi_0^y(1-\pi_0)^{x-y}\cdot\frac{\lambda_0^x e^{-\lambda_0}}{x!}$$

for $y \leq x$

$$= \pi_0^y e^{-\lambda_0}\sum_{x=y}^{\infty}\binom{x}{y}(1-\pi_0)^{x-y}\frac{\lambda_0^x}{x!}$$

expanding the choose function

$$= \pi_0^y e^{-\lambda_0}\sum_{x=y}^{\infty}\frac{x!}{y!(x-y)!}(1-\pi_0)^{x-y}\frac{\lambda_0^x}{x!}$$

$$= \frac{\pi_0^y e^{-\lambda_0}}{y!}\sum_{x=y}^{\infty}\frac{1}{(x-y)!}(1-\pi_0)^{x-y}\frac{\lambda_0^x}{1}$$

$$= \frac{\pi_0^y e^{-\lambda_0}\lambda_0^y}{y!}\sum_{x=y}^{\infty}\frac{(1-\pi_0)^{x-y}\lambda_0^{x-y}}{(x-y)!}$$

aplying the substitution $u = x - y$

$$= \frac{\pi_0^y e^{-\lambda_0}\lambda_0^y}{y!}\sum_{u=0}^{\infty}\frac{(1-\pi_0)^u\lambda_0^u}{(u)!}$$

note that lower bound of summation is because smallest value of u is 0, $x - y = 0 - 0 = 0$

$$= \frac{\pi_0^y e^{-\lambda_0}\lambda_0^y}{y!}\sum_{u=0}^{\infty}\frac{((1-\pi_0)\lambda_0)^u}{(u)!}$$

$$= \frac{\pi_0^y e^{-\lambda_0}\lambda_0^y}{y!}\sum_{u=0}^{\infty}\frac{(\lambda_0-\lambda_0\pi_0)^u}{(u)!}$$

by definition of exponential function

$$= \frac{\pi_0^y e^{-\lambda_0}\lambda_0^y}{y!}e^{(\lambda_0-\lambda_0\pi_0)}$$

$$= \frac{\pi_0^y e^{(-\lambda_0\pi_0)}\lambda_0^y}{y!}$$

for $y = 0,1,2,\ldots,$ x

Hence the marginal PMF of Y can be written as $\frac{\pi_0^y e^{(-\lambda_0\pi_0)}\lambda_0^y}{y!}$ which is the same as a $Y \sim Pois(\lambda_0\pi_0)$.

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

From part b we learned that the number of eggs that will survive can be denoted by $Y \sim Pois(\lambda_0\pi_0)$. From what we know about Poisson distributions, We know that $E[Y] = \lambda_0\pi_0$, thus on average, $\lambda_0\pi_0$ eggs will survive. This makes sense intuitively this value multiplies the probability of an egg surviving and the number of eggs laid together, giving us the expected number of surviving eggs.

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.)

Let X be a sample from $Norm(\mu, \sigma)$ where $\mu = 0$ and tolerance $\tau = \frac{1}{2\sigma^2}$ where τ is treated as random variable. First, we want to write the PDF of X indexed by τ_0, the specific value for τ. (Given $\tau = \tau_0$, we think of this as the conditional density of X).

Since we know $\tau = \frac{1}{2\sigma^2}$ it follows that $\sigma = \frac{1}{\sqrt{\tau}}$. Then from the definition of a normal distribution, $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{x\sigma^2}}$ Since we know $\mu = 0$ and $\sigma = \frac{1}{\sqrt{\tau}}$, we know that,

$$\begin{aligned}
f(x|\tau_0) &= \frac{1}{(\frac{1}{\sqrt{\tau}})\sqrt{2\pi}} e^{\frac{-(x-0)^2}{2(\frac{1}{\sqrt{\tau}})^2}} \\
&= \frac{e^{\frac{-(x)^2}{\frac{2}{\tau}}}}{(\frac{1}{\sqrt{\tau}})\sqrt{2\pi}} \\
&= e^{\frac{-(x)^2\tau}{2}} \cdot \sqrt{\frac{\tau}{2\pi}}
\end{aligned}$$

Hence $e^{\frac{-(x)^2\tau}{2}} \cdot \sqrt{\frac{\tau}{2\pi}}$ for $-\infty \leq x \leq \infty$.

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.

Now assume that τ is a Gamma random variable and the prior distribution is denoted by $g(\tau_0) = Gamma(\alpha_0, \lambda_0)$ where where α_0 is the shape and λ_0 is the rate parameter. We want to find the form of the posterior distribution $h(\tau_0|x)$.

Recall that the PDF of $Gamma(k, \lambda)$ is $f(x) = \frac{\lambda^k}{\Gamma(k)} x^{k-1} e^{-\lambda x}$. Thus in terms of τ we have, $g(\tau) =$

[1] see Definition 13.5 in the NOTES

$\frac{\lambda_0^{\alpha_0}}{\Gamma(\alpha_0)}\tau^{\alpha_0-1}e^{-\lambda\tau_0}$ for $0 \leq \tau < \infty$. Using this, we can write the posterior probability using Bayes Rule as,

$$
\begin{aligned}
h(\tau_0|x) &= \frac{P(X=x|\tau_0=\tau_0)g(\tau_0)}{\int_{\tau_0}^{\infty} P(X=x|\tau_0=\tau_0)g(\tau_0)d\tau_0} \\
&\text{From what we found in part a} \\
&= \frac{e^{\frac{-(x)^2\tau_0}{2}} \cdot \sqrt{\frac{\tau_0}{2\pi}} \times \frac{\lambda_0^{\alpha_0}}{\Gamma(\alpha_0)}\tau_0^{\alpha_0-1}e^{-\lambda\tau_0}}{\int_{\tau_0}^{\infty} e^{\frac{-(x)^2\tau_0}{2}} \cdot \sqrt{\frac{\tau_0}{2\pi}} \times \frac{\lambda_0^{\alpha_0}}{\Gamma(\alpha_0)}\tau_0^{\alpha_0-1}e^{-\lambda\tau_0}d\tau_0} \\
&= \frac{e^{-\tau_0(\frac{(x)^2}{2}+\lambda_0)} \cdot \frac{1}{\sqrt{2\pi}} \cdot \frac{\lambda_0^{\alpha_0}}{\Gamma(\alpha_0)} \cdot \tau_0^{\alpha_0-\frac{1}{2}}}{\int_{\tau_0}^{\infty} e^{-\tau_0(\frac{(x)^2}{2}+\lambda_0)} \cdot \frac{1}{\sqrt{2\pi}} \cdot \frac{\lambda_0^{\alpha_0}}{\Gamma(\alpha_0)} \cdot \tau_0^{\alpha_0-\frac{1}{2}}d\tau_0} \\
&= \frac{e^{-\tau_0(\frac{(x)^2}{2}+\lambda_0)} \cdot \tau_0^{\alpha_0-\frac{1}{2}}}{\int_{\tau_0}^{\infty} e^{-\tau_0(\frac{(x)^2}{2}+\lambda_0)} \cdot \tau_0^{\alpha_0-\frac{1}{2}}d\tau_0}
\end{aligned}
$$

Simplifying the denominator Recall that the integral of a gamma is equal to 1. More specifically, $Gamma(\alpha_0 + \frac{1}{2}, \frac{x^2}{2} + \lambda_0)$, which is very similar to our denominator, will integrate to 1. Hence,

$$
\begin{aligned}
\int_{\tau_0}^{\infty} e^{-\tau_0(\frac{(x)^2}{2}+\lambda_0)} \cdot \tau_0^{\alpha_0-\frac{1}{2}} \times \frac{(\frac{x^2}{2}+\lambda_0)^{\alpha_0+\frac{1}{2}}}{\Gamma(\alpha_0+\frac{1}{2})}d\tau_0 = 1& \\
&\text{note the last term can be pulled out of the integral} \\
&\text{and moved to the other side} \\
\int_{\tau_0}^{\infty} e^{-\tau_0(\frac{(x)^2}{2}+\lambda_0)} \cdot \tau_0^{\alpha_0-\frac{1}{2}}d\tau_0 = \frac{\Gamma(\alpha_0+\frac{1}{2})}{(\frac{x^2}{2}+\lambda_0)^{\alpha_0+\frac{1}{2}}}& \\
&\text{now we have a new term for the denominator}
\end{aligned}
$$

Hence substituting in this new value we get,

$$
\begin{aligned}
h(\tau_0|x) &= \frac{e^{-\tau_0(\frac{(x)^2}{2}+\lambda_0)} \cdot \tau_0^{\alpha_0-\frac{1}{2}}}{\frac{\Gamma(\alpha_0+\frac{1}{2})}{(\frac{x^2}{2}+\lambda_0)^{\alpha_0+\frac{1}{2}}}} \\
&\text{rewritting in a more familiar form} \\
&= \frac{(\frac{x^2}{2}+\lambda_0)^{\alpha_0+\frac{1}{2}}e^{-\tau_0(\frac{(x)^2}{2}+\lambda_0)} \cdot \tau_0^{\alpha_0-\frac{1}{2}}}{\Gamma(\alpha_0+\frac{1}{2})} \\
&= \frac{(\frac{x^2}{2}+\lambda_0)^{\alpha_0+\frac{1}{2}}}{\Gamma(\alpha_0+\frac{1}{2})}e^{-\tau_0(\frac{(x)^2}{2}+\lambda_0)} \cdot \tau_0^{\alpha_0-\frac{1}{2}}
\end{aligned}
$$

Our final result is $h(\tau_0|x) = \frac{(\frac{x^2}{2}+\lambda_0)^{\alpha_0+\frac{1}{2}}}{\Gamma(\alpha_0+\frac{1}{2})}e^{-\tau_0(\frac{(x)^2}{2}+\lambda_0)} \cdot \tau_0^{\alpha_0-\frac{1}{2}}$, we can see that this is a Gamma distribution. Thus the prior distribution is a Gamma distribution with $\alpha + \frac{1}{2}$ as the shape parameter and $\frac{x^2}{2} + \lambda_0$ as the rate parameter ($Gamma(\alpha + \frac{1}{2}, \frac{x^2}{2} + \lambda_0)$).

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?

In medical research, assume that 10% of null hypotheses are false and if so, the chance of a Type II error and failing to reject it is 0.55. We want to use Bayes rule to show that 50% of such studies are actually reporting Type I errors, given we reject the null hypothesis at an alpha level of 0.05.

In our context, let the event H represent when the null hypothesis is false and D is the event that we do not reject the null hypothesis. From our assumptions, we know that $P(H) = 0.10$, thus $P(H^c) = 1 - 0.10 = 0.90$ by definition of complements. From are assumptions we also have that $P(D^c|H^c) = 0.05$, the Type I error rate and $P(D|H) = 0.55$, the Type II error rate. From this we can calculate $P(D^c|H) = 1 - 0.55 = 0.45$. Then from the laws of total probability we can also calculate that $P(D^c) = P(D^c|H^c) \cdot P(D^c|H^c) + P(D^c|H) \cdot P(D^c|H) = (0.05)(0.9) + (0.45)(0.10) = 0.09$. Thus it follows that $P(D) = 1 - 0.09 = 0.91$. Now we have everything we need to find $P(H^c|D^c)$, the probability of the null hypothesis being true give that we reject the null hypothesis.

Using Bayes Rule, $P(H^c|D^c) = \frac{P(D^c|H^c)P(H^c)}{P(D^c)}$. Then plugging in our found values, we get $\frac{P(D^c|H^c)P(H^c)}{P(D^c)} = \frac{0.05 \cdot 0.9}{0.09} = 0.5$. Hence of all the times we reject the hypothesis, the null is actually true about 0.5 of the time.

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 (show your code) 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

  Fill in the table showing 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.)

```
fdr <- function(H_prob, type_1_err, type_2_err){
  fdr <- (type_1_err * (1- H_prob)) / (type_1_err * (1- H_prob) + (1-type_2_err) * H_prob)
}

fdr_table <- expand_grid("H_prob" = c(0.05, 0.1, 0.2, 0.5),
                         "type_1_err" = c(0.0001, 0.005, 0.01, 0.05),
                         "type_2_err" = c(0.05, 0.2, 0.3, 0.5))

fdr_table["FDR"] = rep(0, 64)
for (i in 1:64) {
```

```r
  fdr_table[i, 4] = fdr(fdr_table[i, 1],
                        fdr_table[i, 2],
                        fdr_table[i, 3])
}
```

Table 1: False Discovery Rate for varying inputs

	Type 1 error							
	0.0001				0.005			
	Type II error				Type II error			
$P(H)$	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	0.1667	0.1919	0.2135	0.2754	0.5	0.5429	0.5758	0.6552
0.20	9×10^{-4}	0.0011	0.0013	0.0018	0.0452	0.0533	0.0604	0.0826
0.50	0.0865	0.1011	0.1139	0.1525	0.3214	0.36	0.3913	0.4737

	Type 1 error							
	0.01				0.05			
	Type II error				Type II error			
$P(H)$	0.05	0.2	0.3	0.5	0.05	0.2	0.3	0.5
0.05	4×10^{-4}	5×10^{-4}	6×10^{-4}	8×10^{-4}	0.0206	0.0244	0.0278	0.0385
0.10	0.0404	0.0476	0.0541	0.0741	0.1739	0.2	0.2222	0.2857
0.20	10^{-4}	10^{-4}	10^{-4}	2×10^{-4}	0.0052	0.0062	0.0071	0.0099
0.50	0.0104	0.0123	0.0141	0.0196	0.05	0.0588	0.0667	0.0909

The above table shows us the false discovery rates for different probabilities of the null hypothesis being false and Type I and II error rates. We can see that as the probability of the null being false ($P(H)$) increases, so does the FDR for both type I and II error rates. We can also observe that as the Type I and II error rates increase, so does the FDR.

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.

In a poll, it was found that the Republican candidate was preferred by 59 out of the 100 people samples in State A and 525 out of 1000 people sampled in state B. Let us assume that the samples are randomly drawn from the population with a proportion of π voting Republican. Our goal is to use a large sample Z test of $H_0 : \pi = 0.5$ versus $H_1 : \pi > 0.5$ to check which state has greater evidence of supporting a Republican victory.

We can preform a Z-test for these samples because we have a sufficient size of 100 and 1000. We also know that from the samples we have $X_1, \ldots, X_100$ and $X_1, \ldots, X_1000$ observations that each represent the outcome of a Bernoulli random variable with probability π. Furthermore we know that $\bar{X} \sim Norm(\pi_0, \sqrt{\frac{\pi_0(1-\pi_0)}{n}})$ where π_0 is the probability of voting republican and n is the number of people from state A or B.

State A There are 100 people from State A, thus under the null hypothesis, $\bar{X} \sim Norm(0.5, \sqrt{\frac{0.5(1-0.5)}{100}}) = Norm(0.5, 0.05)$. Then to test our hypothesis, we need to find the p-value, $P(\bar{X} \geq \frac{59}{100})$. Using R,

```r
state_A_p_val <- pnorm(59/100, 0.5, 0.05, lower.tail=F)
round(state_A_p_val, 4)
```

```
## [1] 0.0359
```

State B Following similar steps as in State A, $\bar{X} \sim Norm(0.5, \sqrt{\frac{0.5(1-0.5)}{1000}}) \approx Norm(0.5, 0.0158)$, and the p-value is,

```
state_B_p_val <- pnorm(525/1000, 0.5, sqrt(0.5 * 0.5 / 1000), lower.tail=F)
round(state_B_p_val, 4)
```

```
## [1] 0.0569
```

Conclusion For State A we observe a P-value of about 0.0359, which is less than the significance level of 0.05. Whereas in State B, we observe a P-value of about 0.0569, which is above the alpha level. Thus in State A we have significant evidence against the null hypothesis and we can reject the null in favor of the alternative hypothesis, $H_1 : \pi > 0.5$. Thus in State A, we can say that the population prefers Republican. Controversially, state B dose not have significant evidence against the null hypothesis, and we cannot say that the population prefers Republican. Hence we know that State A has greater evidence supporting a Republican victory.

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)

Now we want to approach this problem through Bayesian analysis. We will find the posterior probability of the null hypothesis for each State using a beta prior with mean 0.5 and standard deviation 0.05.

First, we know that for some beta distribution $X \sim Beta(alpha_0, \beta_0)$ the expected value can be written as $E[X] = \frac{\alpha_0}{\alpha_0+\beta_0}$ and the variance is denoted by $Var[X] = \frac{\alpha_0}{\alpha_0+\beta_0} \cdot \frac{\beta_0}{\alpha_0+\beta_0} \cdot \frac{1}{\alpha_0+\beta_0+1}$. From our assumptions we know that, $0.5 = E[X]$ and $0.05 = Var[X]$. Thus solving the system of equations,

$$\begin{aligned}
0.5 &= \frac{\alpha_0}{\alpha_0 + \beta_0} \\
0.5\alpha_0 + 0.5\beta_0 &= \alpha_0 \\
\beta_0 &= \alpha_0
\end{aligned}$$

$$\begin{aligned}
0.05^2 &= \frac{\alpha_0}{\alpha_0 + \beta_0} \cdot \frac{\beta_0}{\alpha_0 + \beta_0} \cdot \frac{1}{\alpha_0 + \beta_0 + 1} \\
&\text{substituting in } \beta_0 = \alpha_0 \text{ and results found above} \\
0.05^2 &= 0.5 \cdot \frac{\beta_0}{2\beta_0} \cdot \frac{1}{2\beta_0 + 1} \\
\frac{0.05^2}{0.5} &= \frac{1}{2} \cdot \frac{1}{2\beta_0 + 1} \\
\frac{0.05^2}{0.5 \cdot 0.5} &= \frac{1}{2\beta_0 + 1} \\
2\beta_0 + 1 &= 100 \\
2\beta_0 &= 99 \\
\beta_0 &= 49.5 \\
\alpha_0 &= 49.5
\end{aligned}$$

Thus we know that the prior distribution can be denoted by $\pi \sim Beta(49.5, 49.5)$. We know that the posterior distribution of π will be $\pi \sim Beta(\alpha_0 + x, \beta_0 + n - x)$.

State A

For the sample of $n = 100$ and $x = 59/100$, we know $f(x|\pi_0) \sim Binom(100, \pi_0)$ where x is the number of people that vote republican. Thus with the prior distribution $\pi \sim Beta(49.5, 49.5)$, we can find the posterior probability to be $\pi|x \sim Beta(108.5, 90.5)$. Then using R to find $P(\pi \leq 0.5)$, the posterior probability of the null hypothesis,

```
state_A_p_val <- pbeta(0.5, 108.5, 90.5, lower.tail=T)
round(state_A_p_val, 4)
```

```
## [1] 0.1004
```

State B

Following a similar procedure as in State A, the sample in State B has parameters $n = 1000$ and $x = 525/1000$. Thus we can find the posterior distribution to be $\pi|x \sim Beta(574.5, 524.5)$. Thus using R to find $P(\pi \leq 0.5)$,

```
state_B_p_val <- pbeta(0.5, 574.5, 524.5, lower.tail=T)
round(state_B_p_val, 4)
```

```
## [1] 0.0656
```

Conclusion

We observe a P-value of about 0.1004 in State A and 0.0656 in State B. This tells us that State A has a higher probability of the null hypothesis being true. In other words, State A has a higher probability of there being no favoritism towards Rebublicans in comparison to State B.