Stat 251 Final Project

## [Stat 251 Final Project](http://127.0.0.1:18200/rmd_output/3/#content)

* [Introduction and Motivation](http://127.0.0.1:18200/rmd_output/3/#introduction-and-motivation)
* [Literature Review](http://127.0.0.1:18200/rmd_output/3/#literature-review)
* [Methods](http://127.0.0.1:18200/rmd_output/3/#methods)
* [Results](http://127.0.0.1:18200/rmd_output/3/#results)
* [Discussion](http://127.0.0.1:18200/rmd_output/3/#discussion)
* [Appendix](http://127.0.0.1:18200/rmd_output/3/#appendix)

### Cason Wight

## Introduction and Motivation

This project looks at the 2015 probability of a police shooting victim being male. It specifically looks at these proportions in Texas and California and the difference in the probability between Texan victims and Californian victims. It is known that males commit a disproportionate amount of crime, but I wish to know if police shootings happen disproportionately to men.

This report is organized as follows:  
1. Literature Review  
2. Methods  
3. Results  
4. Discussion  
5. Appendix

The results of this report, coupled with other demographic information, may explain if trends of bias or crime happen disproportionately among men. Although this report does not make any conclusions on bias or other underlying trends, it can help people in California and Texas know the proportion of shooting victims that are male. It can also explain any possible difference in this proportion between the two states.

## Literature Review

Gender disparity and discrimination of criminals from police officers is a topic that has been widely discussed by the public. Police shooting have also come under public scrutiny in recent years.  
Some studies have shown that male officers are more likely to be involved in shootings than female officers (see McElvain and Kposowa 2008). Worrall et al. (2018) suggest that black suspects are not the disproportionate targets of officers. Alpert, Dunham, and MacDonald (2004) found that officers are more likely to shoot proportional to the level of resistance that a suspect shows and that hispanic officers are less likely to shoot. According to White (2002), the single best predictor for a whether a shooting will occur or not is if the suspect is a gun-wielding man. White also found that somewhere around 95%-99% of police shooting victims are male. The literature on this topic is extensive, but this report will help show a specific comparison between two specific states. This report will analyze any significant difference between California and Texas in terms of male victim police shooting proportions, if one exists.

## Methods

This section deals mainly with justifying the distributions of choice, selecting priors for the distribution, and summarizing the different prior distributions.

The methods section contains the following elements:  
1. Explanation of data  
2. Justification of likelihood distribution  
3. Selection of priors  
4. Prior distributions on the desired probability  
5. Prior predictive distributions for number of males out of 20 new police shooting victims  
6. Prior distribution on the difference between the probabilities between the two states  
7. Plots summarizing the prior distributions

### Explanation of the Data

The data from this project can be found at <https://www.kaggle.com/fivethirtyeight/fivethirtyeight-police-killings-dataset/version/103>. The data is organized as a single observation binomial response. The observational unit in this case is a “State” (California or Texas). The response is the number of victims that are male, out of the total victims. I desire to know the posterior distribution of the probability of being male, among other things.

The data was originally obtained from theguardian.com, with some additional information added. Police shooting victim information is made pubicly available, which is how the data was obtained by theguardian.com.

The prior and posterior predictive distributions are calculated for binomially distributed data for both states, given 20 new victims. The estimated difference between the probabilities of the two states is calculated using Monte Carlo estimation.

### Justification of Likelihood Distribution

The data from kaggle is organized with each victim having his or her own row. Each row represents a victim of a police shooting. Two interesting covariates that Kaggle includes are the gender (male or female) and the State (Alabama, Alaska, etc) of each victim. We are given a set number of victims; the number of male victims in a given state is a random variable. This means that the number of male victims could follow a Bin(n,θ)Bin(n,θ), where nn is the total number of victims and θθ is the probability that any given victim is male. The data was reorganized so that State-level victim data (male or female) was obtained for 2015.

A binomial distribution is suitable with this data because the response for each victim is binary (male or female). Additionally, the total number of victims in each State works as a cap on how many male victims there were in that State. The binomial distribution also assumes that there is equal probability of being male for any given victim. This assumption may not hold in reality, but we will make this assumption in the report.

A good distribution for θθ is a Beta(a,b)Beta(a,b). The beta distribution has a support of [0,1][0,1], which matches the requirement for a probability. The beta is an intuitive and interpretable distribution for a probability. Also, the beta is a conjugate pair with the binomial, so the posterior could be easily updated with future years of information. The prior and posterior distribution for θθ will both be a beta, which is convenient for inference.

### Selection of Priors

According to unodc.org, around 80% of homicide victims are male. That information, coupled with the fact that most violent criminals are male, makes me think that most police shooting victims will be male. Additionally, the literature review validates that somewhere above 90% of police shooting victims are male. I am not too sure on a specific probability, so my prior for θθ is that θ∼Beta(10,1)θ∼Beta(10,1) for both Texans and Californians. The intuitive explanation of these parameters is that this is the equivalent of having seen 1 female and 10 males out of 11 total victims in the past.

### Prior Distributions

* Distributions of θθ
* Predictive distributions for Y∼Bin(20,θ)Y∼Bin(20,θ)
* Distribution of θ|TX−θ|CAθ|TX−θ|CA
* Plots
* Prior distribution on the difference between the probabilities between the two states

#### **Distributions of Probability**

θCA∼Beta(10,1)θCA∼Beta(10,1)

θTX∼Beta(10,1)θTX∼Beta(10,1)

π(θ|TX)=π(θ|CA)=Γ(a+b)Γ(a)Γ(b)θa−1(1−θ)b−1=Γ(11)Γ(10)Γ(1)θ9(1−θ)0=10θ9π(θ|TX)=π(θ|CA)=Γ(a+b)Γ(a)Γ(b)θa−1(1−θ)b−1=Γ(11)Γ(10)Γ(1)θ9(1−θ)0=10θ9

See figures 1 and 2 in the plots section for the plots of π(θ|TX)π(θ|TX) and π(θ|CA)π(θ|CA). Because the prior selections are the same for each state, the prior distributions for θθ are the same.

#### **Prior Predictives**

f(ynew|TX)=f(ynew|CA)=∫10f(ynew|θ)π(θ)dθf(ynew|TX)=f(ynew|CA)=∫01f(ynew|θ)π(θ)dθ

=(nnewynew)Γ(a+b)Γ(a+ynew)Γ(b+(nnew−ynew))Γ(a)Γ(b)Γ(a+b+nnew)=(nnewynew)Γ(a+b)Γ(a+ynew)Γ(b+(nnew−ynew))Γ(a)Γ(b)Γ(a+b+nnew)

=(20ynew)Γ(10+1)Γ(10+ynew)Γ(1+(20−ynew))Γ(10)Γ(1)Γ(10+1+20)=(20ynew)Γ(10+1)Γ(10+ynew)Γ(1+(20−ynew))Γ(10)Γ(1)Γ(10+1+20)

=(20ynew)Γ(11)Γ(10+ynew)Γ(1+(20−ynew))Γ(10)Γ(1)Γ(31)=(20ynew)Γ(11)Γ(10+ynew)Γ(1+(20−ynew))Γ(10)Γ(1)Γ(31)

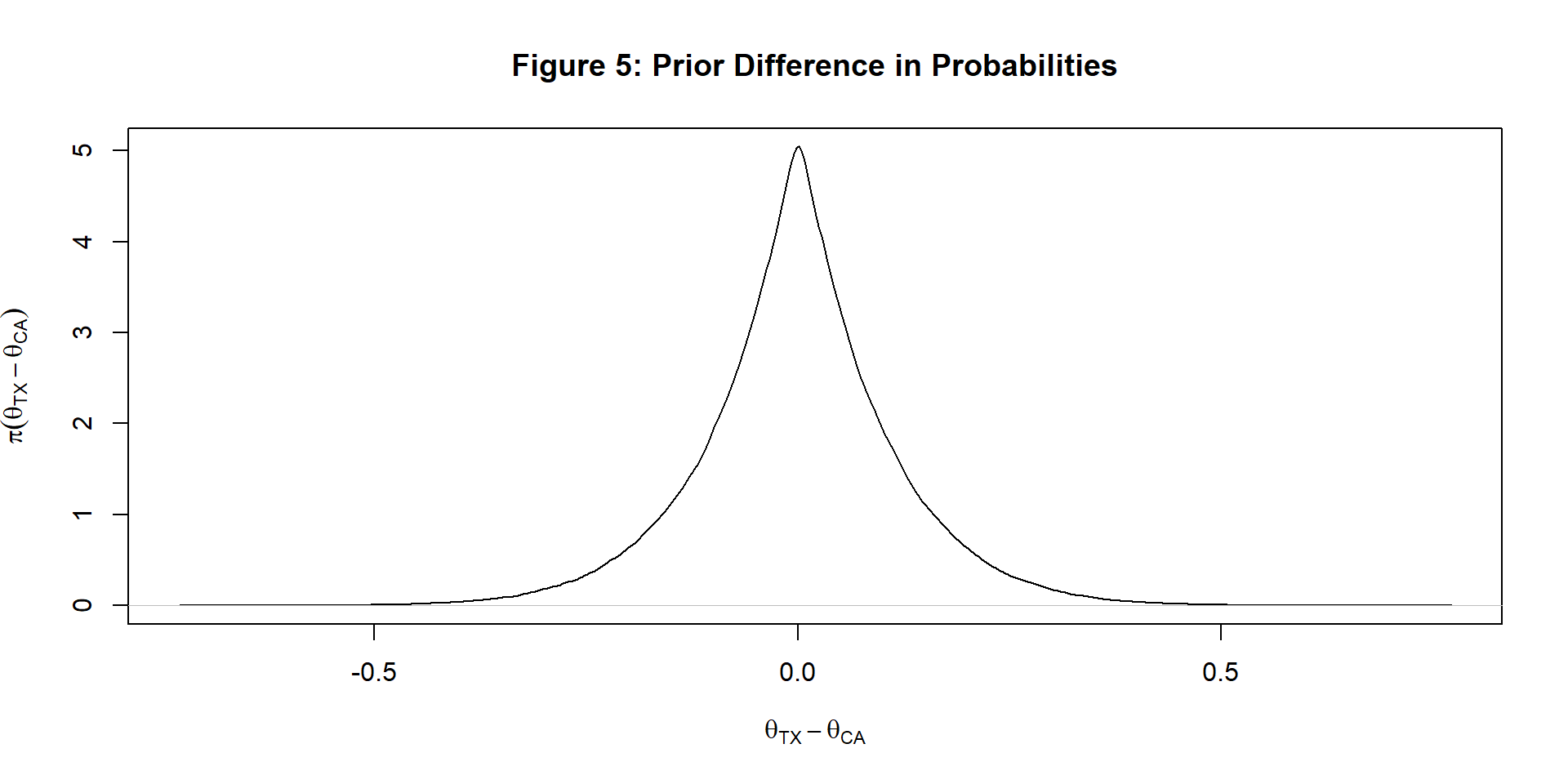
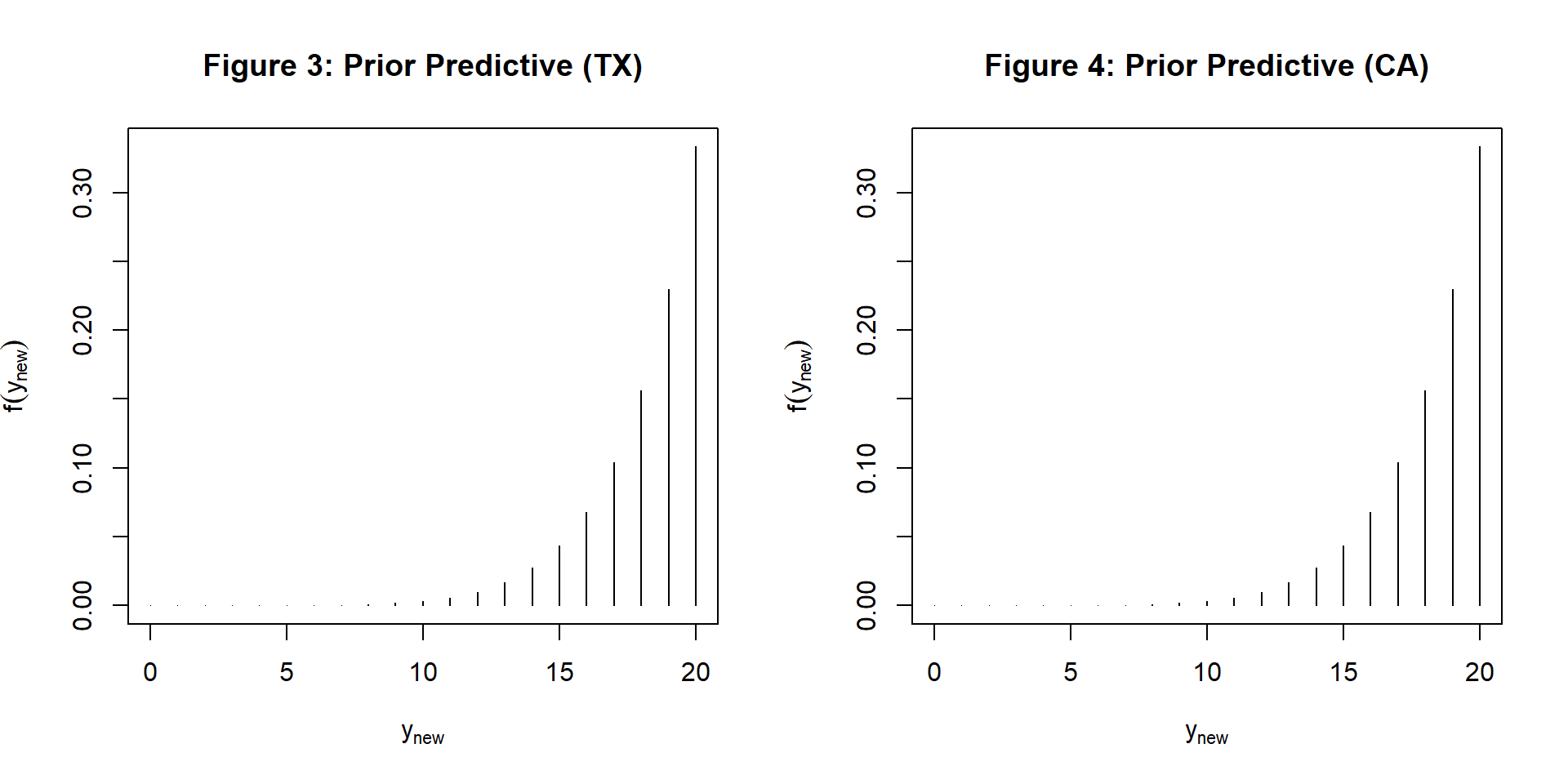
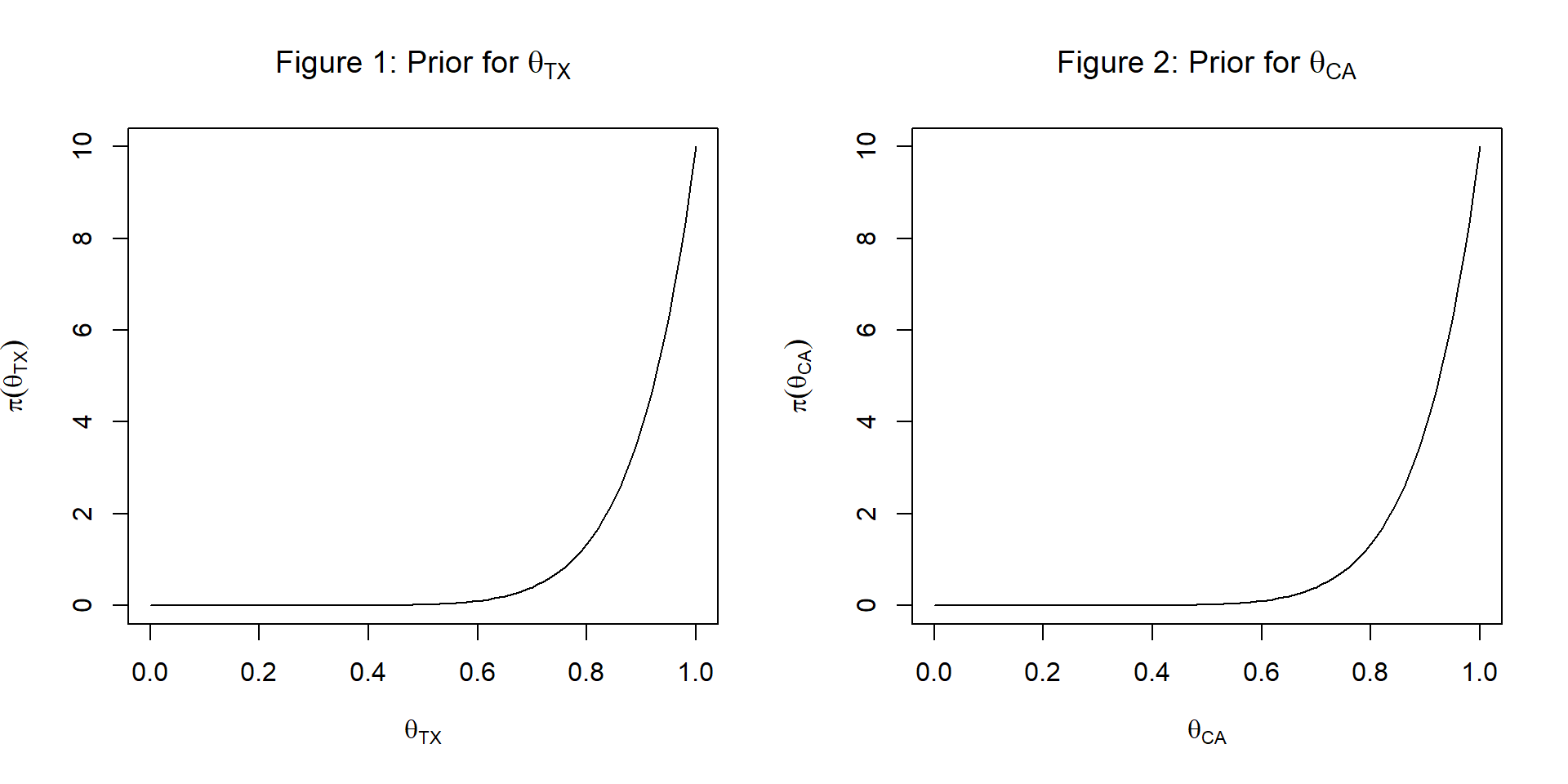
See figures 3 and 4 in the plots section for the plots of f(ynew|TX)f(ynew|TX) and f(ynew|TX)f(ynew|TX). The prior predictive distributions for the two states are the same because the prior selections are the same.

#### **Difference in Probabilities**

The prior distribution of the difference θCA−θTXθCA−θTX doesn’t have an analytical solution. See figure 5 in the plots section for a plot of the Monte Carlo estimated distribution of the difference.

#### **Plots**

The above described plots are depicted in this section. The posterior distributions will help us better understand any difference in probability of a victim being male between Texans and Californians.



The priors are not particularly helpful in this case. The posterior distribution (π(θTX−θCA|YTX,YCA)π(θTX−θCA|YTX,YCA)) will be more helpful in helping us see if there is any statistically significant difference between the probabilities of victims being male between the two States. The next section will go over how to obtain the posterior distribution, and any inferences that can be drawn from it (posterior credibility intervals for θTXθTX, θCAθCA, and θTX−θCAθTX−θCA; estimates for the three; and a statistical conclusion on if a significant difference exists).

## Results

This section contains an overview of the data, posterior inference, and associated plots. The content of this section is organized as follows: 1. Summary statistics for the data 2. Posterior distributions on the desired probability 3. Posterior predictive distributions for number of males out of 20 new police shooting victims 4. Posterior distribution on the difference between the probabilities between the two states 5. Plots summarizing the posterior distributions

### Summary Statistics

Table 1 below summarizes the number of male and female victims in each state. As shown, very few of the victims are female, which matches with the idea behind the selected priors. Out of 467 total police shooting victims in 2015, only 22 were females (4.71%).

Table 1: Police Shooting Victims by State

|  | **AK** | | | **AL** | | **AR** | | **AZ** | | **CA** | | **CO** | | **CT** | | **DC** | | **DE** | | **FL** | | **GA** | | **HI** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Female | 0 | | | 0 | | 0 | | 3 | | 5 | | 1 | | 0 | | 0 | | 0 | | 0 | | 2 | | 0 |
| Male | 2 | | | 8 | | 4 | | 22 | | 69 | | 11 | | 1 | | 1 | | 2 | | 29 | | 14 | | 4 |
| Total | 2 | | | 8 | | 4 | | 25 | | 74 | | 12 | | 1 | | 1 | | 2 | | 29 | | 16 | | 4 |
|  | | **IA** | **ID** | | **IL** | | **IN** | | **KS** | | **KY** | | **LA** | | **MA** | | **MD** | | **ME** | | **MI** | | **MN** | |
| Female | | 1 | 0 | | 0 | | 0 | | 1 | | 0 | | 0 | | 0 | | 0 | | 0 | | 0 | | 0 | |
| Male | | 1 | 4 | | 11 | | 8 | | 5 | | 7 | | 11 | | 5 | | 10 | | 1 | | 9 | | 6 | |
| Total | | 2 | 4 | | 11 | | 8 | | 6 | | 7 | | 11 | | 5 | | 10 | | 1 | | 9 | | 6 | |

|  | | **MO** | | **MS** | | **MT** | | **NC** | | **NE** | | **NH** | | **NJ** | | **NM** | | **NV** | | **NY** | | **OH** | | **OK** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Female | | 0 | | 0 | | 0 | | 2 | | 1 | | 0 | | 2 | | 0 | | 0 | | 0 | | 0 | | 0 |
| Male | | 10 | | 6 | | 2 | | 8 | | 5 | | 1 | | 9 | | 5 | | 3 | | 14 | | 10 | | 22 |
| Total | | 10 | | 6 | | 2 | | 10 | | 6 | | 1 | | 11 | | 5 | | 3 | | 14 | | 10 | | 22 |
|  | **OR** | | **PA** | | **SC** | | **TN** | | **TX** | | **UT** | | **VA** | | **WA** | | **WI** | | **WV** | | **WY** | | **Total** | |
| Female | 0 | | 0 | | 0 | | 0 | | 2 | | 0 | | 1 | | 0 | | 0 | | 0 | | 1 | | 22 | |
| Male | 8 | | 7 | | 9 | | 6 | | 44 | | 5 | | 8 | | 11 | | 5 | | 2 | | 0 | | 445 | |
| Total | 8 | | 7 | | 9 | | 6 | | 46 | | 5 | | 9 | | 11 | | 5 | | 2 | | 1 | | 467 | |

The estimated mean number of police shooting victims in a state is 9.94 (95% CI: 6.24 to 13.63). The estimated mean number of male police shooting victims in a state is 9.47 (95% CI: 5.99 to 12.95). The estimated mean number of female police shooting victims in a state is 0.47 (95% CI: 0.18 to 0.76). The two states with the most total victims are Texas and California (46 and 74 total victims, respectively). For this reason, and the large cultural difference between Texas and California, we are taking particular interest in inference for these two states. The most shootings occur in these politically and geographically differing States, so it may be useful to see if there is any trend between the two in terms of the gender of police shooting victims.

### Posterior

* Distributions of θθ
* Predictive distributions for Y∼Bin(20,θ)Y∼Bin(20,θ)
* Distribution of θ|TX−θ|CAθ|TX−θ|CA
* Plots

#### **Distributions of Probability**

θTX|YTX∼Beta(10+yTX,1+nTX−yTX)θTX|YTX∼Beta(10+yTX,1+nTX−yTX)

θTX|YTX∼Beta(10+44,1+46−44)θTX|YTX∼Beta(10+44,1+46−44)

θTX|YTX∼Beta(54,3)θTX|YTX∼Beta(54,3)

E(θTX)=aa+b=5454+3≈0.95E(θTX)=aa+b=5454+3≈0.95

The posterior estimate of the true proportion of Texan police shooting victims who are male is 0.95. With 95% probability, the true proportion of Texan police shooting victims who are male lies within the interval 0.88 to 0.99. See figure 6 in the plots section for a plot of the prior and posterior of θTXθTX, along with the 95% credibility interval and the mean.

θCA|YCA∼Beta(10+yCA,1+nCA−yCA)θCA|YCA∼Beta(10+yCA,1+nCA−yCA)

θCA|YCA∼Beta(10+69,1+74−69)θCA|YCA∼Beta(10+69,1+74−69)

θCA|YCA∼Beta(79,6)θCA|YCA∼Beta(79,6)

E(θCA)=aa+b=7979+6≈0.93E(θCA)=aa+b=7979+6≈0.93

The posterior estimate of the true proportion of Californian police shooting victims who are male is 0.93. With 95% probability, the true proportion of Californian police shooting victims who are male lies within the interval 0.87 to 0.97. See figure 7 in the plots section for a plot of the prior and posterior of θCAθCA, along with the 95% credibility interval and the mean.

See figure 8 in the plots section for a plot of the distributions of θTX|YTXθTX|YTX and θCA|YCAθCA|YCA together.

π(θ|YCA)=Γ(a⋆+b⋆)Γ(a⋆)Γ(b⋆)θa⋆−1(1−θ)b⋆−1=Γ(79+6)Γ(79)Γ(6)θ79−1(1−θ)6−1π(θ|YCA)=Γ(a⋆+b⋆)Γ(a⋆)Γ(b⋆)θa⋆−1(1−θ)b⋆−1=Γ(79+6)Γ(79)Γ(6)θ79−1(1−θ)6−1

=Γ(85)Γ(79)Γ(6)θ78(1−θ)5=Γ(85)Γ(79)Γ(6)θ78(1−θ)5

π(θ|YTX)=Γ(a⋆+b⋆)Γ(a⋆)Γ(b⋆)θa⋆−1(1−θ)b⋆−1=Γ(54+3)Γ(54)Γ(3)θ54−1(1−θ)3−1π(θ|YTX)=Γ(a⋆+b⋆)Γ(a⋆)Γ(b⋆)θa⋆−1(1−θ)b⋆−1=Γ(54+3)Γ(54)Γ(3)θ54−1(1−θ)3−1

=Γ(57)Γ(54)Γ(3)θ53(1−θ)2=Γ(57)Γ(54)Γ(3)θ53(1−θ)2

#### **Posterior Predictives**

f(ynew|yTX)=∫10f(ynew|θ)π(θ|yTX)dθf(ynew|yTX)=∫01f(ynew|θ)π(θ|yTX)dθ

=(nnewynew)Γ(a⋆+b⋆)Γ(a⋆+ynew)Γ(b⋆+nnew−ynew)Γ(a⋆)Γ(b⋆)Γ(a⋆+b⋆+nnew)=(nnewynew)Γ(a⋆+b⋆)Γ(a⋆+ynew)Γ(b⋆+nnew−ynew)Γ(a⋆)Γ(b⋆)Γ(a⋆+b⋆+nnew)

=(20ynew)Γ(54+3)Γ(54+ynew)Γ(3+20−ynew)Γ(54)Γ(3)Γ(54+3+20)=(20ynew)Γ(54+3)Γ(54+ynew)Γ(3+20−ynew)Γ(54)Γ(3)Γ(54+3+20)

=(20ynew)Γ(57)Γ(54+ynew)Γ(23−ynew)Γ(54)Γ(3)Γ(77)=(20ynew)Γ(57)Γ(54+ynew)Γ(23−ynew)Γ(54)Γ(3)Γ(77)

f(ynew|yCA)=∫10f(ynew|θ)π(θ|yCA)dθf(ynew|yCA)=∫01f(ynew|θ)π(θ|yCA)dθ

=(nnewynew)Γ(a⋆+b⋆)Γ(a⋆+ynew)Γ(b⋆+nnew−ynew)Γ(a⋆)Γ(b⋆)Γ(a⋆+b⋆+nnew)=(nnewynew)Γ(a⋆+b⋆)Γ(a⋆+ynew)Γ(b⋆+nnew−ynew)Γ(a⋆)Γ(b⋆)Γ(a⋆+b⋆+nnew)

=(20ynew)Γ(79+6)Γ(79+ynew)Γ(6+20−ynew)Γ(79)Γ(6)Γ(79+6+20)=(20ynew)Γ(79+6)Γ(79+ynew)Γ(6+20−ynew)Γ(79)Γ(6)Γ(79+6+20)

=(20ynew)Γ(85)Γ(79+ynew)Γ(26−ynew)Γ(79)Γ(6)Γ(105)=(20ynew)Γ(85)Γ(79+ynew)Γ(26−ynew)Γ(79)Γ(6)Γ(105)

See figure 9 in the plots section for a plot of the posterior predictive distributions for 20 new victims from both either Texas or California.

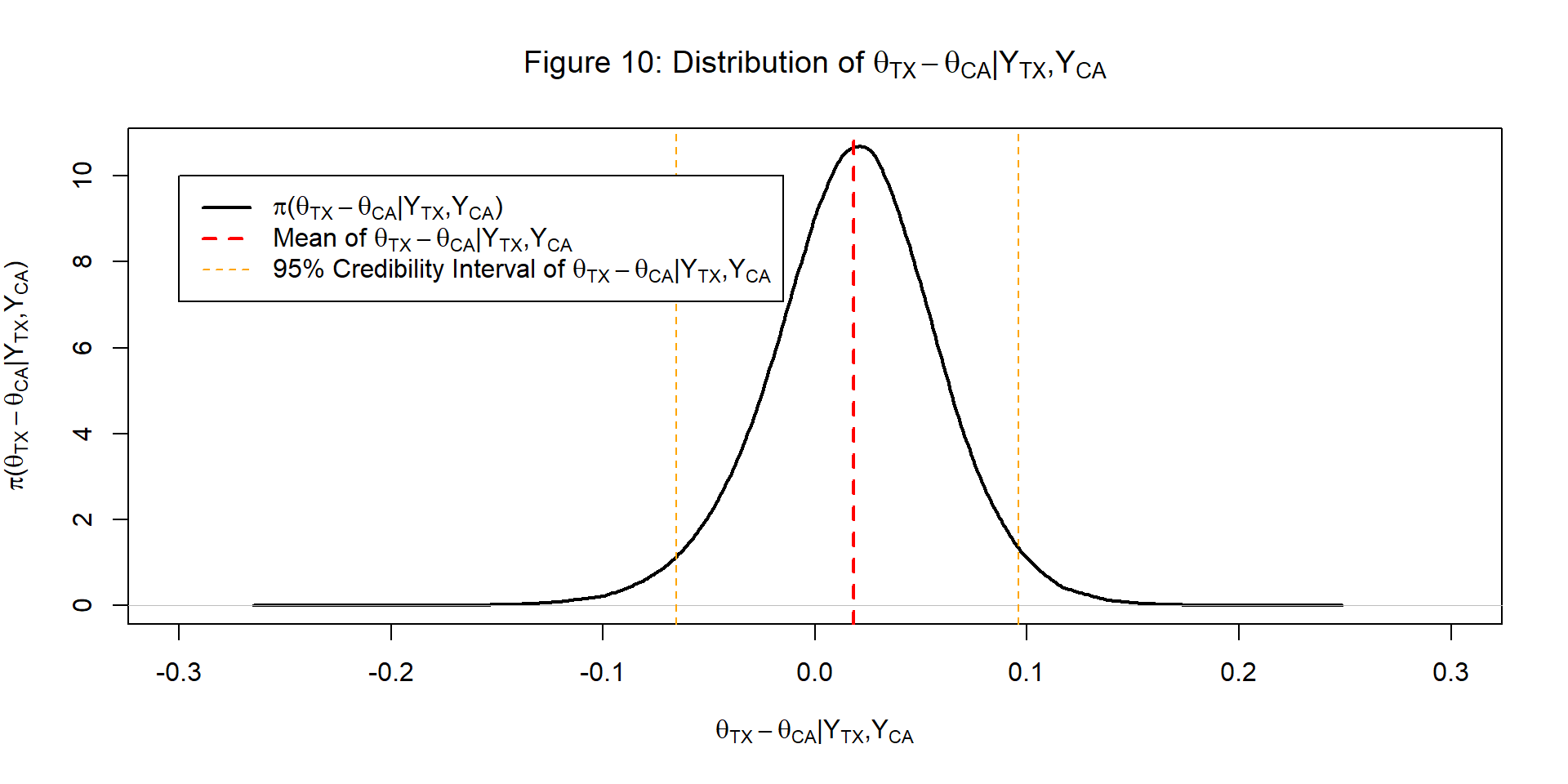
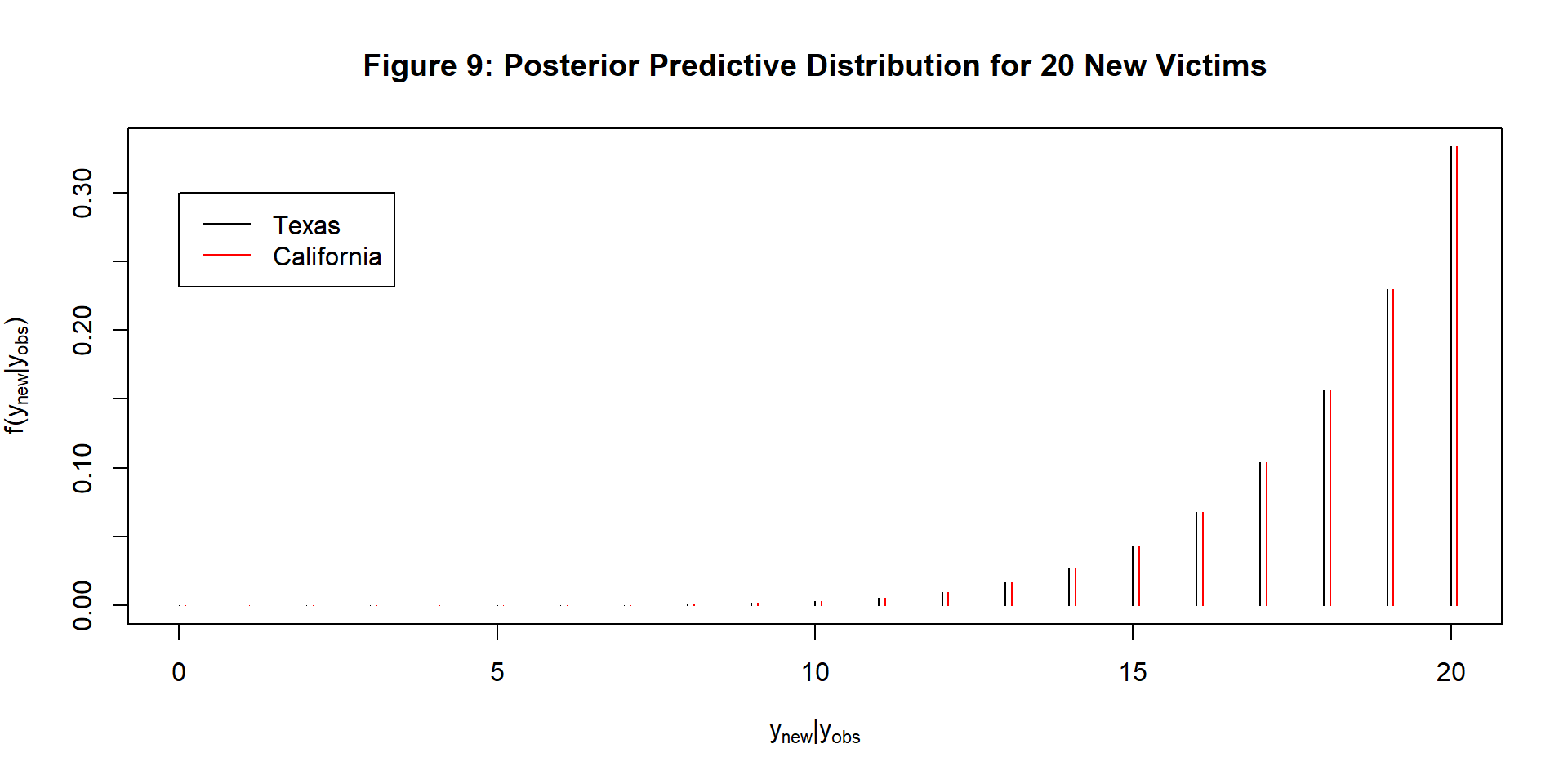
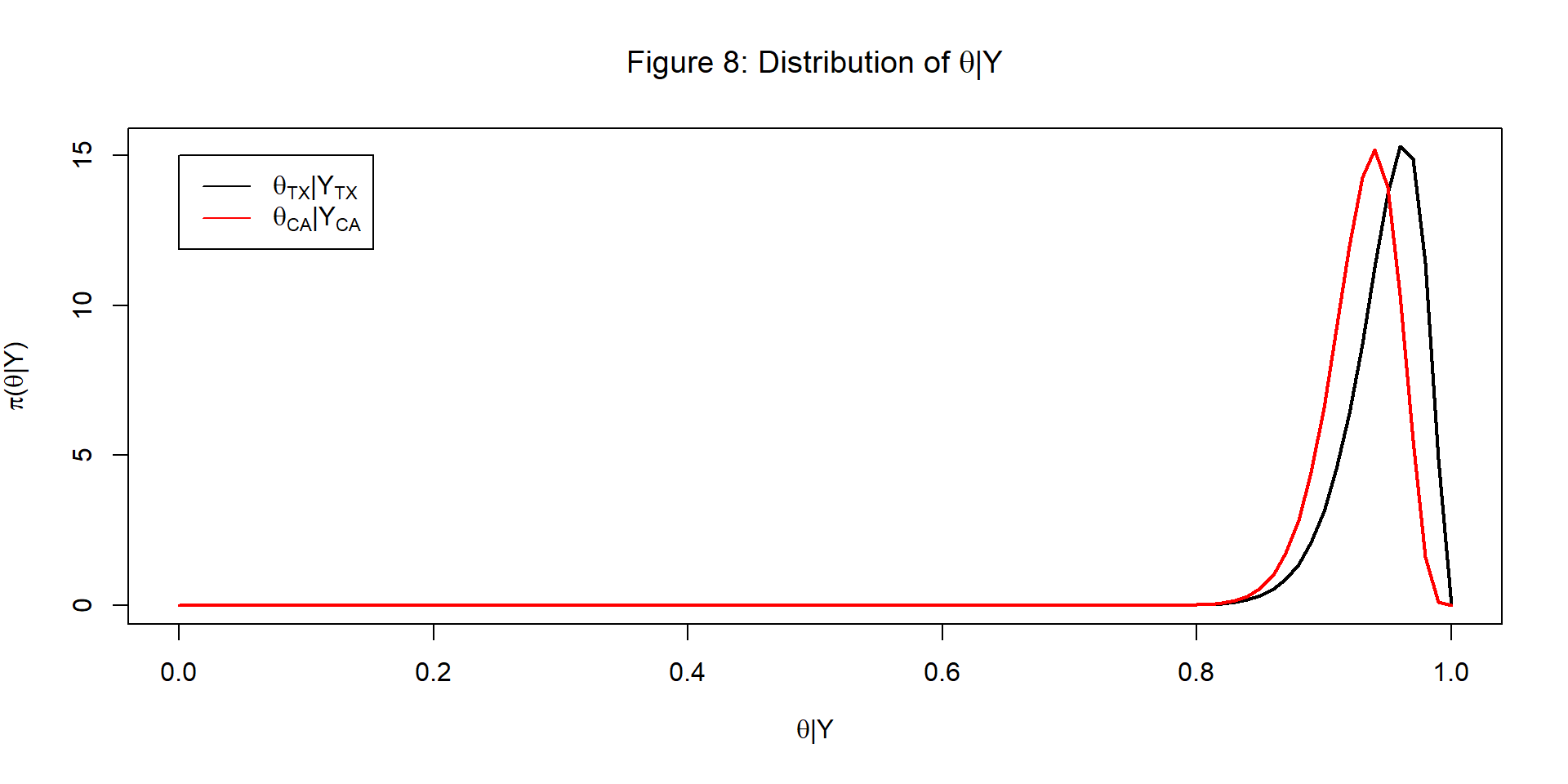
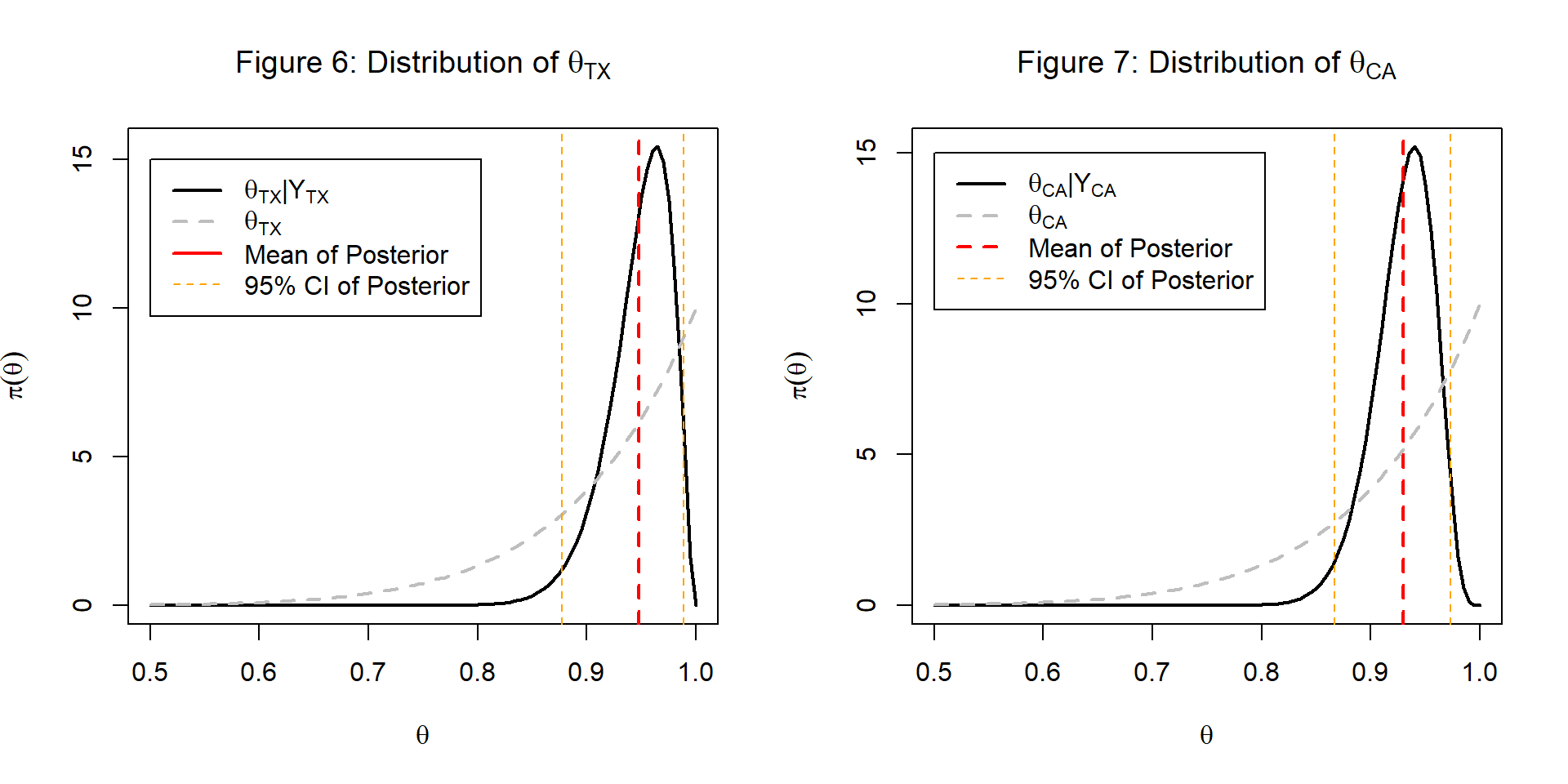
#### **Difference in Probabilities**

θCA−θTX|YCA,YTXθCA−θTX|YCA,YTX

The difference in probabilities θCA−θTX|YCA,YTXθCA−θTX|YCA,YTX doesn’t have an analytical solution. See figure 10 in the plots section below for a Monte Carlo estimate of the distribution, along with the 95% credibility interval and the estimated mean.

The posterior estimate of the true difference in proportions of police shooting victims who are male between Texans and Californians is 0.02. With 95% probability, the true difference in proportions of police shooting victims who are male between Texans and Californians lies within the interval -0.07 to 0.1. Because this interval contains 0, we can say that there is not sufficient evidence to say there is a difference between these two proportions.

#### **Plots**



## Discussion

The discussion contains the following:  
1. Summary of the results  
2. Shortcomings of the report  
3. Possible future work

### Summary of Results

The results of this report are not too surprising. In fact, they line up similarly to what the priors predicted. For example, the posterior estimate of the true difference in proportions of police shooting victims who are male between Texans and Californians is 0.02 (95% credibility interval: -0.07 to 0.1) and the prior predicted no difference. The credibility interval contains 0, which means that there is no statistically significant difference between the two proportions. This can give an initial conclusion to Texans and Californians given a police shooting victim in either state, the probability of being male is about the same (or there is no evidence to the contrary).

For both Texas and California, the prior estimated proportion of men was 1010+1≈0.911010+1≈0.91. The posterior estimates of the proportions of men for Texas and California were 5454+3≈0.955454+3≈0.95 and 7979+6≈0.937979+6≈0.93, respectively. Thus, the posterior distributions are not too different from the prior estimates.

### Shortcomings

An obvious shortcoming of this report is the lack of data. Given only data for 2015, these results are fairly limited. It would also be useful to perform an analysis with covariates in order to make inferences about police bias. Right now, the inferences on these proportions do not have a lot of practical use. A situational look on the different outcomes based on gender, after accounting for other factiors, would be a better approach.

### Future Work

For future work, there are two possible avenues:

One method is to perform this same analysis, updating it with new results for Texas and California each year. Looking back to previous years before 2015 would also enhance the results.

Another avenue is to look at a logistic regression model and see the effect of a victim being male or of being from a particular State, after accounting for various other covariates. This method may answer other questions of interest regarding bias or other trends with gender. O’brien and Dunson (2004) outline a Bayesian approach to logistic regression, which would be useful in this setting.

##### **Conclusion**

While the results of this project may not be particularly exciting, they could help people understand that roughly the same proportion of shooting victims are male, whether Californian or Texan.

## Appendix

### R Code

#### **Function for Predictive Distributions**

pred.dist <- **function**(yobs=0, nobs=0, a, b, ynew, nnew, type = 0){

astar <- yobs + a

bstar <- nobs-yobs+b

l.pos.pred <- **lgamma**(astar + bstar)-**lgamma**(astar) -

**lgamma**(bstar) + **lchoose**(nnew,ynew) +

**lgamma**(astar + ynew) + **lgamma**(bstar + nnew - ynew) -

**lgamma**(astar + bstar + nnew)

l.prior.pred <- **lgamma**(a + b)-**lgamma**(a) -

**lgamma**(b) + **lchoose**(nnew,ynew) +

**lgamma**(a + ynew) + **lgamma**(b + nnew - ynew) -

**lgamma**(a + b + nnew)

**if**(type == "Post"){

out <- **exp**(l.pos.pred)

} **else** **if** (type == "Prior") {

out <- **exp**(l.prior.pred)

} **else**{

out <- **cbind**(**exp**(l.pos.pred),**exp**(l.prior.pred))

**colnames**(out) <- **c**("Posterior Prediction", "Prior Prediction")

**rownames**(out) <- ynew

}

out

}

#### **Plots of Prior Distributions**

**par**(mfrow = **c**(1,2))

**curve**(**dbeta**(x,10,1), main = **expression**(**paste**("Figure 1: Prior for ",theta[TX])),

xlab = **expression**(theta[TX]),ylab = **expression**(**pi**(theta[TX])))

**curve**(**dbeta**(x,10,1), main = **expression**(**paste**("Figure 2: Prior for ",theta[CA])),

xlab = **expression**(theta[CA]),ylab = **expression**(**pi**(theta[CA])))

#### **Plots of Prior Predictive Distributions**

**plot**(0:20,**pred.dist**(a = 10, b = 1, ynew = 0:20, nnew = 20, type = "Prior"),type = "h",

main = "Figure 3: Prior Predictive (TX)",

ylab = **expression**(**f**(y[new])),xlab = **expression**(y[new]))

**plot**(0:20,**pred.dist**(a = 10, b = 1, ynew = 0:20, nnew = 20, type = "Prior"),type = "h",

main = "Figure 4: Prior Predictive (CA)",

ylab = **expression**(**f**(y[new])),xlab = **expression**(y[new]))

#### **Plot of Difference in Proportions Prior Distribution**

**par**(mfrow = **c**(1,1))

**plot**(**density**(**rbeta**(1000000,10,1)-**rbeta**(1000000,10,1)),

main = "Figure 5: Prior Difference in Probabilities",

xlab = **expression**(theta[TX]-theta[CA]),

ylab = **expression**(**pi**(theta[TX]-theta[CA])))

#### **Reading in and Formatting Data**

dats <- **read.csv**("police\_killings.csv", header = TRUE)

TX.dats <- dats[dats[,10]=="TX",**c**(3,10)]

CA.dats <- dats[dats[,10]=="CA",**c**(3,10)]

TX.males <- **sum**(TX.dats[,1]=="Male")

TX.deaths <- **nrow**(TX.dats)

CA.males <- **sum**(CA.dats[,1]=="Male")

CA.deaths <- **nrow**(CA.dats)

#### **Summary Statistics on Data**

**library**(kableExtra)

**names**(dats)[**c**(3,10)] <- **c**("Gender","State")

tab1 <- **addmargins**(**table**(dats[,**c**(3,10)]))

**rownames**(tab1)[3] <- "Total"

**colnames**(tab1)[**ncol**(tab1)] <- "Total"

**kable**(tab1[,1:12])

**kable**(tab1[,13:24])

**kable**(tab1[,25:36])

**kable**(tab1[,37:48])

#### **Posterior Estimates and Confidence Intervals**

# MC Distribution for TX

mean.tx <- (10 + TX.males)/(1 + TX.deaths - TX.males + 10 + TX.males)

ci.tx <- **qbeta**(**c**(.025,.975), 10 + TX.males,1 + TX.deaths - TX.males)

# MC Distribution for CA

mean.ca <- (10 + CA.males)/(1 + CA.deaths - CA.males + 10 + CA.males)

ci.ca <- **qbeta**(**c**(.025,.975), 10 + CA.males,1 + CA.deaths - CA.males)

# MC Distribution for the difference

sampled.diff <- **rbeta**(1000000,10+TX.males,1+TX.deaths-TX.males)-

**rbeta**(1000000,10+CA.males,1+CA.deaths-CA.males)

mean.diff <- **mean**(sampled.diff)

ci.diff <- **quantile**(sampled.diff, **c**(.025,.975))

#### **Plots of Posterior Distributions**

**par**(mfrow = **c**(1,2))

**curve**(**dbeta**(x,10+TX.males,1+TX.deaths-TX.males),

main = **expression**(**paste**("Figure 6: Distribution of ",theta[TX])), lwd = 2,

xlab = **expression**(theta),ylab = **expression**(**pi**(theta)), xlim = **c**(.5,1))

**curve**(**dbeta**(x,10,1),add = TRUE, col = "grey", lty = "dashed", lwd = 2)

**abline**(v = mean.tx, lty = "dashed", col = 2, lwd = 2)

**abline**(v = ci.tx, lty = "dashed", col = "orange")

**legend**(legend = **c**(**expression**(**paste**(theta[TX],"|",Y[TX])),

**expression**(theta[TX]), "Mean of Posterior",

"95% CI of Posterior"),

col = **c**("black","grey", "red", "orange"),

lty = **c**("solid","dashed"), x = .5, y = 15,

lwd = **c**(**rep**(2,3),1))

**curve**(**dbeta**(x,10+CA.males,1+CA.deaths-CA.males),

main = **expression**(**paste**("Figure 7: Distribution of ",theta[CA])),

xlab = **expression**(theta),ylab = **expression**(**pi**(theta)),

xlim = **c**(.5,1), lwd = 2)

**curve**(**dbeta**(x,10,1),add = TRUE, col = "grey", lty = "dashed", lwd = 2)

**abline**(v = mean.ca, lty = "dashed", col = 2, lwd = 2)

**abline**(v = ci.ca, lty = "dashed", col = "orange")

**legend**(legend = **c**(**expression**(**paste**(theta[CA],"|",Y[CA])),

**expression**(theta[CA]),

"Mean of Posterior", "95% CI of Posterior"),

col = **c**("black","grey", "red", "orange"),

lty = **c**("solid",**rep**("dashed",3)),

x = .5, y = 15, lwd = **c**(**rep**(2,3),1))

#### **Plot of Posterior Distributions on One Graph**

**par**(mfrow = **c**(1,1))

**curve**(**dbeta**(x,10+TX.males,1+TX.deaths-TX.males),

main = **expression**(**paste**("Figure 8: Distribution of ",theta,"|",Y)),

lwd = 2,

xlab = **expression**(**paste**(theta,"|",Y)),

ylab = **expression**(**paste**(pi,"(",theta,"|",Y,")")), xlim = **c**(0,1))

**curve**(**dbeta**(x,10+CA.males,1+CA.deaths-CA.males), add = TRUE, col = 2, lwd = 2)

**legend**(legend = **c**(**expression**(**paste**(theta[TX],"|",Y[TX])),

**expression**(**paste**(theta[CA],"|",Y[CA]))),

col = **c**("black","red"),lty = "solid", x = 0, y = 15)

#### **Plot of Posterior Predictive Distributions**

**plot**(0:20,**pred.dist**(yobs = TX.males,nobs = TX.deaths,

a = 10, b = 1, ynew = 0:20, nnew = 20,

type = "Prior"),type = "h",

main = "Figure 9: Posterior Predictive Distribution for 20 New Victims",

ylab = **expression**(**paste**(f,"(",y[new],"|",y[obs],")")),

xlab = **expression**(**paste**(y[new],"|",y[obs])))

**lines**(0:20+.1,**pred.dist**(yobs = CA.males,nobs = CA.deaths,

a = 10, b = 1, ynew = 0:20, nnew = 20,

type = "Prior"),type = "h", col = 2)

**legend**(legend = **c**("Texas", "California"), col = 1:2,

lty = "solid", x = 0, y = .3)

#### **Plot of Difference in Proportions Posterior Distribution**

**plot**(**density**(sampled.diff), xlim = **c**(-.3,.3),

main = **expression**(**paste**("Figure 10: Distribution of ",

theta[TX]-theta[CA],"|",Y[TX],",",Y[CA])),

xlab = **expression**(**paste**(theta[TX]-theta[CA],"|",Y[TX],",",Y[CA])),

ylab = **expression**(**paste**(pi,"(",theta[TX]-theta[CA],

"|",Y[TX],",",Y[CA],")")), lwd = 2)

**abline**(v = mean.diff, lty = "dashed", col = 2, lwd = 2)

**abline**(v = ci.diff, lty = "dashed", col = "orange")

**legend**(legend=**c**(**expression**(**paste**(pi,"(",theta[TX]-theta[CA],"|",

Y[TX],",",Y[CA],")")),

**expression**(**paste**("Mean of ",

theta[TX]-theta[CA],"|",Y[TX],",",Y[CA])),

**expression**(**paste**("95% Credibility Interval of ",

theta[TX]-theta[CA],"|",Y[TX],",",Y[CA]))),

col = **c**("black", "red", "orange"), lwd = **c**(2,2,1),

lty = **c**("solid",**rep**("dashed",2)), x = -.3, y = 10)

### Works Cited

Alpert, Geoffrey P, Roger G Dunham, and John M MacDonald. 2004. “Interactive Police-Citizen Encounters That Result in Force.” Police Quarterly 7 (4). Sage Publications Sage CA: Thousand Oaks, CA: 475–88.

McElvain, James P, and Augustine J Kposowa. 2008. “Police Officer Characteristics and the Likelihood of Using Deadly Force.” Criminal Justice and Behavior 35 (4). Sage Publications Sage CA: Los Angeles, CA: 505–21.

O’brien, Sean M, and David B Dunson. 2004. “Bayesian Multivariate Logistic Regression.” Biometrics 60 (3). Wiley Online Library: 739–46.

White, Michael D. 2002. “Identifying Situational Predictors of Police Shootings Using Multivariate Analysis.” Policing: An International Journal of Police Strategies & Management 25 (4). MCB UP Ltd: 726–51.

Worrall, John L, Stephen A Bishopp, Scott C Zinser, Andrew P Wheeler, and Scott W Phillips. 2018. “Exploring Bias in Police Shooting Decisions with Real Shoot/Don’t Shoot Cases.” Crime & Delinquency 64 (9). SAGE Publications Sage CA: Los Angeles, CA: 1171–92.