

Bayesian Analysis of Mass Shooting Fatalities

Christopher Danko

April 2020

1 Abstract

This paper investigates how states with different gun laws have different rates of mass shooting fatalities. First I use a Bayesian data analysis to demonstrate that there is a significant difference between states with strong gun laws and states with weak gun laws. Using data collected by the Gun Violence Archive for the years 2013-2019 provided to me by an associate, Devin Hughes, I used the information on fatalities to create a statistical experiment to determine whether or not states with strong gun laws have less mass shooting fatalities proportional to their population. A binomial test on each year demonstrates that in five out of eight years, states with strong gun laws have significantly less mass shooting fatalities. This result is consistent with expectations, as gun law advocates would expect states with strong gun laws to have less shooting fatalities per capita than states with weak gun laws. I also use a Monte Carlo Markov Chain model to simulate the posterior distribution of the probability that given a mass shooting fatality has occurred, the fatality occurs in a state with strong gun laws. These findings are consistent with expectations, as advocates of stronger gun laws would expect these laws to result in less lethal mass shootings. Gun laws are expected to reduce the number of high-capacity and high-fire rate firearms in circulation. This effect is expected to result in less total shots fired during a mass shooting, as potential shooters will be forced to resort to less effective firearms.

These results contradict expectations of ‘Good Guy With a Gun’ theorists, who argue that reducing access to firearms makes people more vulnerable to mass shooters, who will find ways to obtain lethal firearms regardless of laws. Under these expectations, stronger gun laws should either increase or have no significant effect on fatalities during mass shootings.

2 Introduction

The topic of mass murder, and mass shootings specifically, has been mostly ignored—even among criminologists—until fairly recently [6]. Searches of academic databases including JSTOR and Google Scholar did not yield any previous regression analyses of the effect of gun laws on mass shootings in the United States. Most scholarship around mass shootings has instead focused on the effects of public mass shootings on the public debate around gun control laws [8] and the purported links between mental illness and mass shootings [7].

The definition of mass shooting used here is the definition used by the Gun Violence Archive (GVA), where the data was obtained from. The GVA defines a mass shooting as an incident where 4 or more are injured or killed, not including the shooter. Noteworthy here is equal importance given to counting the number of injured in a shooting in addition to those killed in a mass shooting incident. This means that there are a number of mass shootings that had no fatalities.

A state is classified as having “strong” gun laws or “weak” gun laws based on the Giffords Gun Law Center grades [2]. A state with an A, B, or C rating is considered “strong” and a D or F rating would be considered weak. The cutoff of a C rating is chosen because there is a fairly significant difference between the gun laws in C states and those in D states. Take Florida, a C- rated state as of 2018. Florida has a mandatory waiting period of 3 days, a minimum firearm and

ammunition purchasing age of 21, ammunition restrictions ¹, and an extreme risk protection order law ² [1]. North Carolina, a D-rated state, has none of these laws. However, it does have a universal background check requirement for purchasing handguns [3]. This cutoff selection also has the nice property that it creates a fairly even distribution of population between the two groups. A table of states and their Giffords Gun Law Center Ratings is provided in Appendix A.

The relationship between stronger gun laws and the number of mass shootings is important to identify for policy-makers. If these laws demonstrably reduce the number of mass shootings that take place, that may be a justification for implementing them. If these policies do not reduce the number of mass shootings or other forms of gun violence, they may not be worth implementing.

This paper chooses to explore the fatality rate of mass shootings for three reasons. Theoretically, stricter gun laws—such as ones that completely ban AR-15s could have no effect on the number of mass shootings and one would conclude that stronger gun laws have no effect on mass shootings. However, mass shooters may be forced to use pistols or other low-capacity, low-fire rate firearms which would lead to fewer fatalities and injuries per shooting. This approach would capture this hypothetical effect, whereas the other approach would not. If stronger gun laws demonstrably reduce the number of casualties per mass shooting, that may be a justification for implementing them.

Second, as the Gun Violence Prevention data set shows, 2642 people have been killed and 9766 people have been injured since 2013. If stronger gun laws reduce the total number of people killed in mass shootings, that is a justification for implementing them.

¹Florida prohibits the manufacture, sale, delivery, and possession of ammunition types such as armor-piercing bullets, exploding shells, bolo shells, and flechette shells [4]

²Extreme risk protection order laws allow law enforcement officials to petition courts to obtain civil orders that prevent dangerous persons from accessing firearms for up to one year [5]

Finally, there is a clear impact of mass shootings on the public discourse around gun violence. As McGinty, et al. found, being exposed to stories about mass shootings made experiment participants more likely to support stronger firearm restrictions and develop more negative attitudes towards people with mental illnesses [8]. Mass shootings have clear psychological impacts on people that influence their political beliefs and attitudes. If stronger gun laws reduce the severity of events causing psychological stress, that is a justification for implementing them.

3 Data

The dataset I work with was compiled by the Gun Violence Archive, a project dedicated to academic research of gun violence. The data tracks mass shootings, defined as any shooting that causes four or more casualties. The data tracks the gun laws of a state, how many people were killed, how many people were injured, where the shooting occurred, whether the shooting was an act of terrorism or not, the location of the shooting, and a wide range of other qualitative variables.

To format the data in a way that was convenient to work with, it was transformed by adding several dummy variables indicating whether the state the shooting was a state with strong gun laws or not in the year of the shooting. The data was then subset by years so that the number of shooting fatalities per group could be compared. It was important to include multiple years in this analysis, as some states—such as Florida—changed gun laws significantly over the time period, making a binomial test over the average population invalid.

This paper also uses population data compiled by the U.S. Census for estimates of state populations.

4 Methods

This paper uses binomial tests to establish statistical difference between the two groups. The null hypothesis for this test is that the distribution of mass shooting fatalities should be equal to the distribution of population ³. If states with strong gun laws contain 50% of the population, we expect them to contain 50% of mass shooting fatalities.

The binomial test statistic is calculated by

$$\sum_{i=k}^N \binom{N}{i} p^i (1-p)^{N-i} \quad (1)$$

Where N is the total number of mass shooting fatalities in the particular year, k is the number of mass shooting fatalities that occurred in states with strong gun laws, and p is the ratio of the population of the states with strong gun laws to the population of all the states in the two groups.

This paper also uses Monte Carlo Markov Chains to sample from the posterior distributions of our statistical experiments conditional on set prior expectations. The algorithm is as follows:

Define $f(x)$ to be a function that is proportional to the desired probability distribution $P(x)$, which is our posterior probability distribution. .

Initialization: We choose an arbitrary point x_t to be the first sample and choose an arbitrary probability density $g(x|y)$ that suggests a candidate for the next sample value x , given the previous sample value y . The choice of distribution here is a Gaussian distribution centered at y , so that points closer to y are more likely to be visited next, making the sequence of samples into a random walk.

³Population estimates come from the U.S. Census Bureau. U.S. territories are excluded from all population ratio estimates

5 Findings

The results of the binomial tests by year are reported in table 9, as well as the sum of mass shooting fatalities in states with strong gun laws, the total sum of mass shooting fatalities, the population ratio, and the results of the binomial tests for each year.

Column 1 contains the number of mass shooting fatalities that occurred in states with strong gun laws. Column 2 contains the total number of mass shooting fatalities that happened in that year. Column 3 is simply column 1 divided by column 2. Column 4 contains the percentage of the total population contained in states with strong gun laws. The interpretation of p-values is the same as a one-sided t-test. Thus, in 6 out of 8 years, there were statistically significantly less shooting fatalities in states with strong gun laws than would be expected if fatalities were evenly distributed across the population.

Table 2 contains the results of a similar binomial test carried out where the statistical experiment is the count of how many mass shootings took place in states with strong gun laws.

In none of the years between 2013-2019 was there a statistically significantly less mass shootings in states with strong gun laws than would be expected if fatalities were evenly distributed across the population.

Now we use a Monte Carlo Markov Chain to take 10,000 samples from our statistical experiment for mass shooting fatalities in 2019 to estimate the probability that a given fatality happens in a state with strong gun scores. For our first chain, we use flat priors, imposing no expectations on the posterior distribution generated by the Markov Chain. This posterior distribution is plotted in figure 1. The black arrow denotes the population ratio, or the probability that we would expect our mean to be centered around if mass shooting fatalities were evenly distributed across population. Table 3 contains summary statistics

for this posterior distribution.

None of the 10,000 samples from the posterior distribution were at or above the population ratio

For the next MCMC analysis, we impose prior expectations using a beta distribution with a standard deviation of .05 and a mean at the population ratio.

The choice of standard deviation is due to the useful fact that

$$\sigma_{post}(X) = \frac{1}{\sqrt{N}} \quad (2)$$

Thus, choosing a standard deviation of .05 conditions our expectations as if we had observed a set of 400 shooting fatalities—almost the same size as the observed set—occurring in strong states roughly proportional to their population ratio.

To impose these prior expectations, we create a beta distribution with mean .5618 and standard deviation .05. A beta distribution is formed by two parameters, α and β . To calculate the parameters to achieve the desired mean and standard deviation, we use the following facts.

$$Mean(X) = \frac{\alpha}{\alpha + \beta} \quad (3)$$

$$\sigma^2(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} \quad (4)$$

We replace σ with $\frac{1}{\sqrt{N}}$ to obtain

$$\frac{1}{N} = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} \quad (5)$$

$$N = \frac{(\alpha + \beta)^2(\alpha + \beta + 1)}{\alpha\beta} \quad (6)$$

Equations 3 and 6 create a system of equations that can be solved for algebraically, the demonstration of which is contained in Appendix A.

Solving when $N = 400$ yields $\alpha = 54.76$ and $\beta = 42.71$. This distribution is heavily weighted at our expectation point, and curves off sharply to reflect our small standard deviation.

The resulting distribution function is graphed in figure 2 in blue, with uniform priors graphed in black as comparison. Note that the distribution mean is indeed the population ratio. Also note how the distribution sharply slopes off from the mean value.

Now we apply our prior distribution to our MCMC model to bias it towards believing that shooting fatalities are equally distributed across population. The posterior generated by this is plotted in figure 3. Every point of this estimated distribution is closer to the population ratio than we bias our expectations towards, which means that the priors are indeed regularizing.

Table 4 contains summary statistics for this posterior distribution. Even with strong priors introduced, there are no samples from this distribution greater than or equal to the population ratio. Note that the mean of this posterior distribution is much greater than the mean in table 3, reflecting how the samples are closed to the targeted ratio.

Now we want to generate a distribution that will place around 500 samples at or above the target value. In other words, we want to quantify the prior expectations one would have to hold in order to observe the statistical experiment and still conclude that there is no statistically significant difference for mass shootings fatalities between the two groups.

Using trial-and-error⁴, we estimate $\alpha = 2544.44$ and $\beta = 1984.32$ with a corre-

⁴We tested constant multiples of 460 for N to make the statistic easier to interpret in relation to the sample size.

sponding theoretical prior sample size of 18400 with a mean at the population ratio, or about 40 times the size of the observed sample.

The corresponding beta distribution is graphed in figure 4 in blue, with the beta distribution from figure 2 in green and flat priors graphed in black. This distribution is much more regularizing than the previous one. The distribution is more centered at the mean, and much less of the distribution is found farther away from the mean.

Now we apply this prior to our MCMC model to estimate the posterior distribution plotted in figure 5. Once again, every sample is closer to the target population ratio than in figure 3 and especially figure 1. This shows that this choice of beta distribution is more regularizing than the previous beta distribution. The population ratio itself falls just at the boundary between the second and third standard deviations of the distribution.

The summary statistics for this posterior distribution are presented in table 5. Notice that 563 samples, or 5.63% of the samples fell at or above the targeted population ratio. Thus, if one holds these priors, there is no statistically significant difference between states with strong gun laws and weak gun laws in 2019. These are extraordinarily strong priors, however, requiring almost three times as many prior mass shooting fatality observations than the seven years of data presented here.

Priors can do more than stress test our conclusion. In table 9, when taken separately, neither 2014 or 2018 demonstrate statistically significant difference between the count of fatalities and the population ratio. However, we can choose priors for these years based on the information contained in other years.

Take 2018. We choose priors based on the information contained in 2013-2017. We choose a sample size equal to the number of mass shooting fatalities in 2013-2017, with a target mean equal to the average distance between the

population ratio and fatality ratio. This prior is graphed in Figure 8

The Markov Chain Monte Carlo sample is plotted in Figure ?? . The summary statistics for this posterior are in 6. None of the samples fell at or above the population ratio. Thus, when our model is informed by the information in 2013-2017, we should still believe that there is a statistical difference between these two sets in 2018.

6 Conclusion

First, 9 strongly demonstrates that states with strong gun laws persistently have less mass shooting fatalities than states with weak gun laws. In five out of seven of the years tracked, states with strong gun laws had far less mass shooting fatalities than would be expected if gun laws were evenly spread across population.

Second, the resilience of this conclusion to strong priors is tested in the year 2019. 5 and 5 show that it takes extremely strong priors to conclude that there is no difference between the two groups of states. This quantifies this prior cutoff point and strongly indicates that there is a difference between the two populations, considering the extremity of priors required to conclude otherwise. Additionally, we demonstrate that by using informational priors from the other years, even years like 2018 are unlikely to have no statistical difference between fatality ratio and population ratio.

Finally, 2 suggests that there is no statistical difference between the two state groups in the number of total shootings.

The first two results points to a wide range of future research. This paper estimates the statistical difference between the two groups and tests its resilience; other papers should focus on the causal impact of strong gun laws on mass shooting fatalities.

The third result should also be examined with a more causal investigation. While this paper does not demonstrate a statistical difference between the two groups, there could be a number of variables correlated with group assignment and mass shootings that mask the effect of these gun laws.

7 References

References

- [1] Florida gun laws.
- [2] Giffords law center’s annual gun law scorecard.
- [3] North carolina gun laws.
- [4] Ammunition regulation in florida, Oct 2019.
- [5] Extreme risk protection orders in florida, Dec 2019.
- [6] J. Fox and M. DeLateur. Mass shootings in america: Newtown.
- [7] J. Knoll and G. Annas.
- [8] E. McGinty, D. Webster, M. Jarlenski, and C. Barry. New media framing of serious mental illness and gun violence in the united states, 1997-2012.

8 Appendix A

Solve the system of equations

$$\begin{aligned} \text{Mean}(X) &= \frac{\alpha}{\alpha + \beta} \\ N &= \frac{(\alpha + \beta)^2(\alpha + \beta + 1)}{\alpha\beta} \end{aligned}$$

for α and β where N and $\text{Mean}(X)$ are constants

Let $p = \text{Mean}(X)$

$$p = \frac{\alpha}{\alpha + \beta}$$

$$p(\alpha) + p(\beta) = \alpha$$

$$p(\beta) = (1 - p)\alpha$$

$$\beta = \frac{(1 - p)\alpha}{p}$$

$$\text{Define } \lambda = \frac{1 - p}{p}$$

$$\beta = \lambda\alpha \tag{7}$$

We substitute (7) into our original system to obtain

$$\begin{aligned} N &= \frac{(\lambda\alpha + \alpha)^2(\alpha + \lambda\alpha + 1)}{\lambda\alpha^2} \\ N &= \frac{((1 + \lambda)^2\alpha^2)((1 + \lambda)\alpha + 1)}{\lambda\alpha^2} \\ N &= \frac{(1 + \lambda)^3\alpha^3 + (1 + \lambda)^2\alpha^2}{\lambda\alpha^2} \\ N &= \frac{(1 + \lambda)^3\alpha + (1 + \lambda)^2}{\lambda} \\ \frac{\lambda N}{(1 + \lambda)^2} - \frac{1}{1 + \lambda} &= \alpha \end{aligned}$$

λ and N are constants

Solving for β then becomes trivial.

9 Tables and Figures

Summary of Binomial Tests					
Year	Subset Fatalities	Total Fatalities	Subset Fatality Ratio	Population Ratio	P-value
2019	205	464	0.4418103	0.5618297	2.361399e-07
2018	205	371	0.5525606	0.5454315	7.945654e-01
2017	128	437	0.2929062	0.4759656	9.078957e-15
2016	162	451	0.3592018	0.4863677	6.726561e-08
2015	126	368	0.3423913	0.4608653	4.894448e-06
2014	99	262	0.3778626	0.4677306	3.577042e-03
2013	128	289	0.4429066	0.4688328	4.093401e-01

Table 1: This table contains the results of all the binomial tests for whether states with strong gun laws have disproportionately less mass shooting fatalities than other states

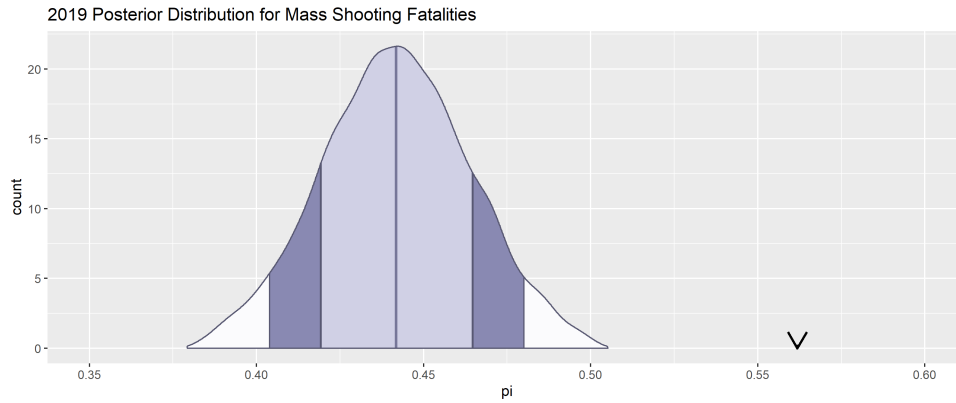


Figure 1: This figure shows the posterior distribution of 2019 mass shootings fatalities across the two state groups approximated by MCMC sampling with flat priors

Summary of Binomial Tests					
Year	Subset Shootings	Total Shootings	Subset Shooting Ratio	Population Ratio	P-value
2019	215	418	0.5143541	0.5618297	0.05446607
2018	189	337	0.5608309	0.5454315	0.58471992
2017	151	346	0.4364162	0.4759656	0.14630074
2016	182	382	0.4764398	0.4863677	0.72024979
2015	144	335	0.4298507	0.4608653	0.27306844
2014	129	269	0.4795539	0.4677306	0.71416632
2013	128	254	0.5039370	0.4688328	0.28517187

Table 2: This table contains the results of all the binomial tests for whether states with strong gun laws have disproportionately less mass shootings than other states

Summary of Posterior with Flat Priors		
Mean	Standard Deviation	Samples > Ratio
0.4418378	0.02300021	0

Table 3: This table contains summary statistics of the Markov Chain Monte Carlo model with flat priors

Summary of Posterior with Strong Priors Imposed		
Mean	Standard Deviation	Samples > Ratio
0.4627848	0.02098638	0

Table 4: This table contains summary statistics of the Markov Chain Monte Carlo model with strongly biased priors

Summary of Posterior with Strong Priors Imposed

Mean	Standard Deviation	Samples > Ratio
0.5506825	0.007066417	563

Table 5: This table contains summary statistics of the Markov Chain Monte Carlo model with extremely strong priors

Summary of Posterior with Strong Priors Imposed

Mean	Standard Deviation	Samples > Ratio
0.5506825	0.007066417	563

Table 6: This table contains summary statistics of the Markov Chain Monte Carlo model for 2018 with informed priors

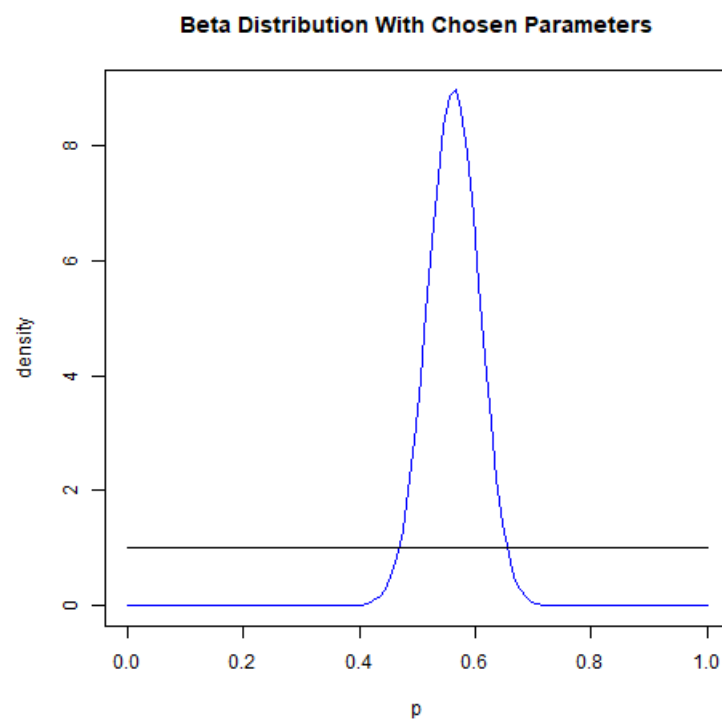


Figure 2: This figure demonstrates the beta distribution that is used to condition the MCMC model plotted in 3.

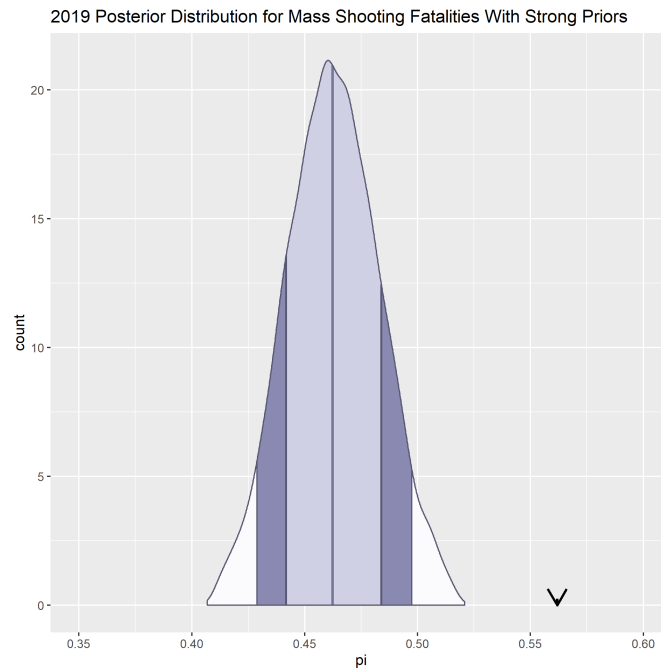


Figure 3: This figure shows the posterior distribution of 2019 mass shootings fatalities across the two state groups approximated by MCMC sampling with strong priors

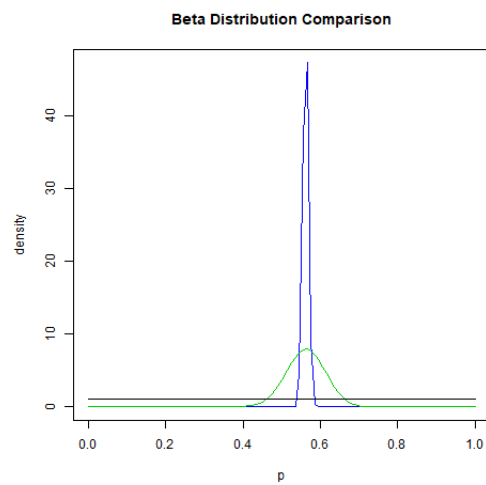


Figure 4: This figure demonstrates the beta distribution that conditions the MCMC model plotted in 5

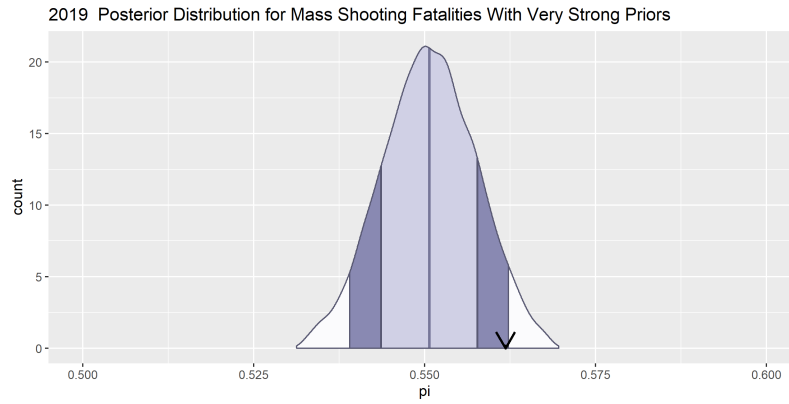


Figure 5: This figure shows the posterior distribution of 2019 mass shootings fatalities across the two state groups approximated by MCMC sampling with extremely strong priors

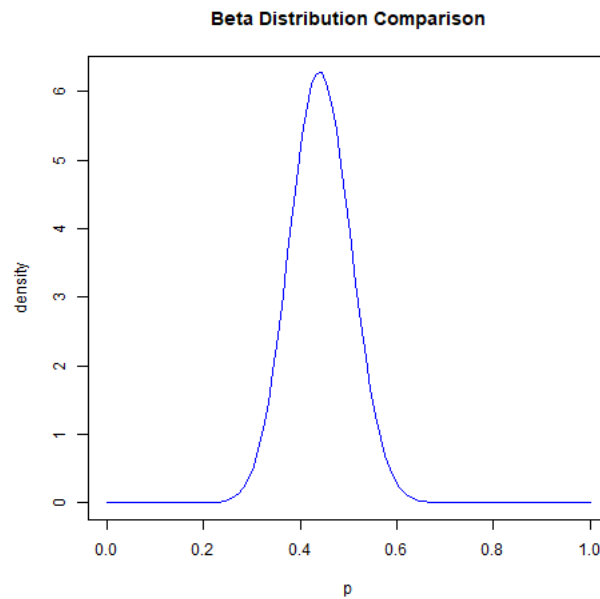


Figure 6: This figure shows the beta distribution created by the prior information contained in the 2013 fatality data

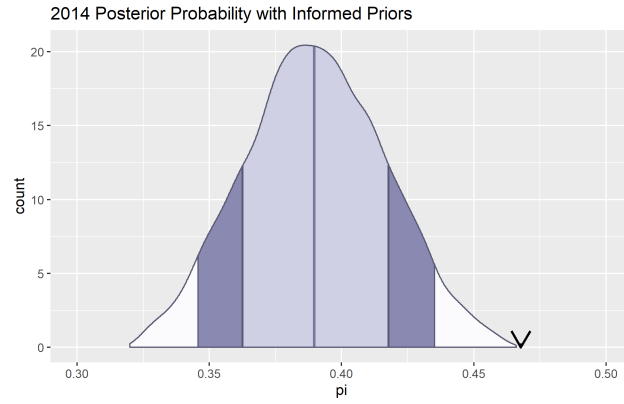


Figure 7: This figure shows the posterior distribution of 2014 mass shooting fatalities informed by the prior information in 2013

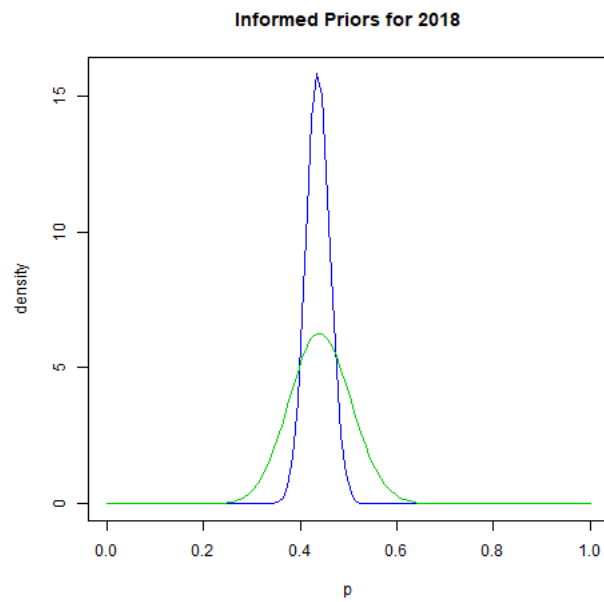


Figure 8: This figure shows the posterior distribution of 2018 mass shooting fatalities informed by the prior information in the first 5 years.