Shark Attack Analysis

AJ, Chase, Chloe

Shark Analysis

Introduction

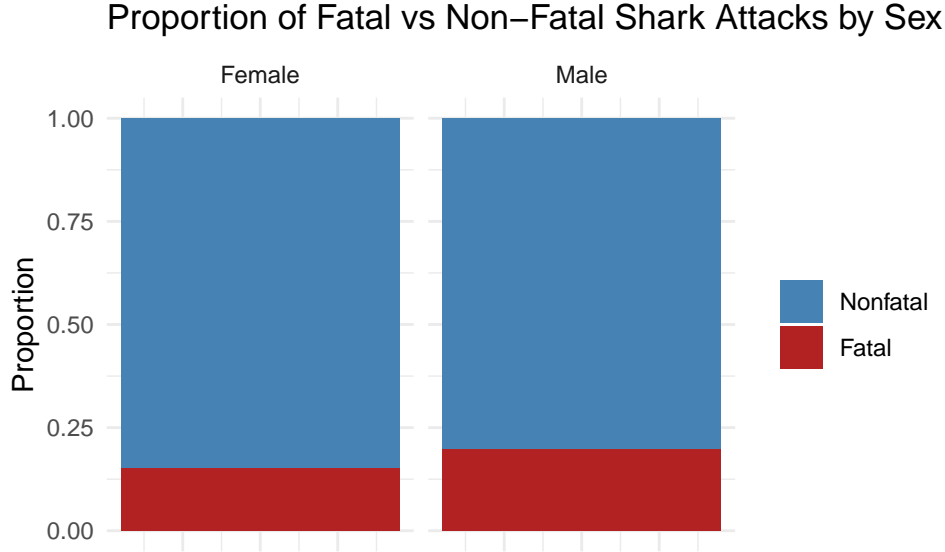Is there a difference between the fatality rate from shark attacks for men and women?

We chose this research question because we are curious about the difference in proportions of fatal shark attacks among men and among women.

Sharks have long been feared — especially since the movie *Jaws* came out in 1975. Are sharks really as dangerous as people think? After all, cows cause more annual deaths than sharks on average, and more commonplace things like cars are far more likely to kill you than a shark. We hope in this analysis to assuage fears (at least somewhat) concerning sharks.

Methods

Our data comes from Global Shark Attack File ("https://www.sharkattackfile.net/"). It is a compilation of shark attack information coming from many different sources. Each of our observations represent one occurrence of a shark attack. For each observation we have information about the sex of the victim and the fatality of the attack – whether the victim survived or was killed. Below is a graph that displays the differences in mortality rates for male and female victims.

```
        Fatal
Sex         No  Yes
  Female  601  108
  Male   3718  918
```

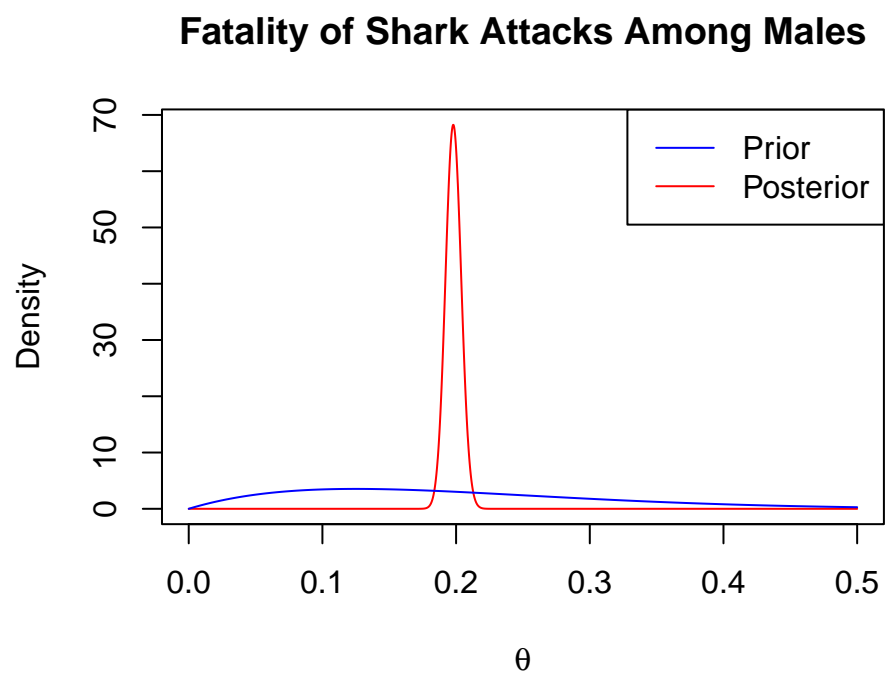## Proportion of Fatal vs Non–Fatal Shark Attacks by Sex

Our data lends itself to the binomial setting. We have a fixed number of events where each has a binary outcome, fatal or nonfatal. It is fairly reasonable to assume that our events are independent. It is unlikely that one shark attack has any sort of significant impact on another. The assumption of greatest concern in using a binomial likelihood is the constant fatality probability of each event. It is unlikely that the fatality of each shark attack is exactly the same. Ignoring this violation would normally result in underestimating the variability of our prediction for the probability of a shark attack being fatal. However, using beta-binomial Bayesian inference allows us to accurately convey our levels of uncertainty.

Our parameters of interest are θ_{female} and θ_{male}. These parameters represent the probability that a shark attack is fatal for each respective sex. With these parameters, we can use Monte Carlo simulation to generate a posterior distribution for $\theta_{female-male}$. With this distribution we can answer our question of whether one sex is more likely to survive a shark attack than another is. Based on the likelihood chosen and our prior beliefs about our shark attacks, we have chosen $\pi_{female} \sim beta(1,9)$ for our prior for females. We chose a similar $\pi_{male} \sim beta(2,8)$ prior for male fatality rates. We chose beta priors because they are conjugate with the binomial setting, and because they make it easy for us to encode our prior beliefs. In our priors we assume that about 1/10 shark attacks will be fatal to women and about 2/10 shark attacks will be fatal for men. Our priors encode a slight survival favoritism towards women because that is what we expect to occur in our data. This favoritism is very weak overall, allowing room for our data to inform our posterior distribution.
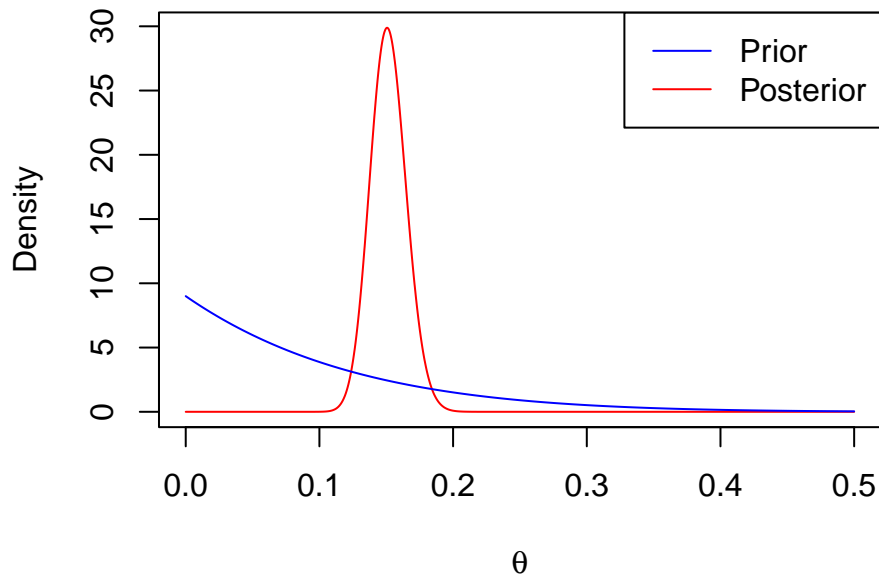
Results

First of all, in order to preform our analysis we obtain the posterior distributions for θ_{male}. The graph below displays the prior and posterior distribution for males.
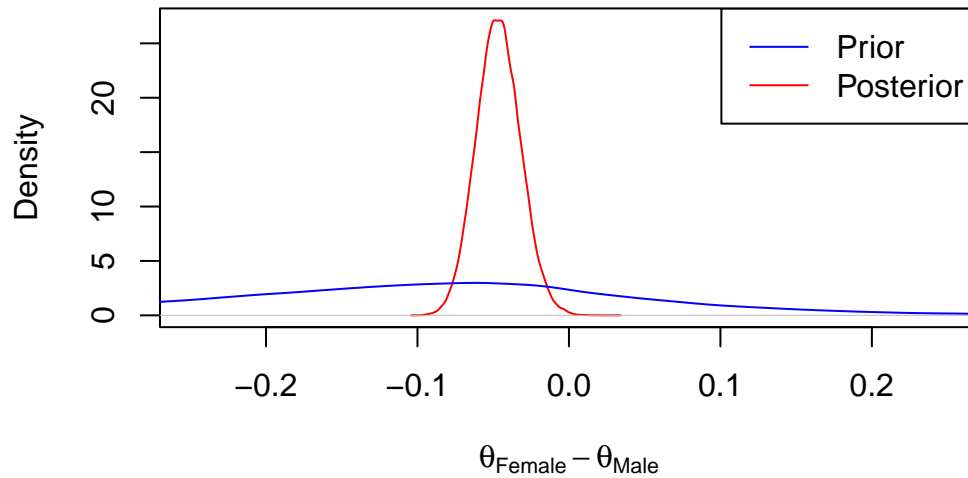
Fatality of Shark Attacks Among Males

We also obtain the posterior distribution for females and graph the prior and posterior for it.

Fatality of Shark Attacks Among Females

From each distribution we simulate 100,000 random draws and use them to generate a posterior distribution. We also plot the prior distribution using the same methods. From the 95% credible interval of the posterior, there is a 95% probability that the difference in proportions of fatal shark attacks for males and females is between -0.074 and -0.017. Therefore, using an alpha level of 0.05, we conclude that there is a statistically significant difference in the fatality rate from shark attacks for men and women, with women being less likely to die in an attack than men.

Difference in Fatality of Shark Attacks by Sex

Discussion

Appendix

```r
#set seed for reproducability
set.seed(2024)

#Read in data and clean Sex and Fatal Variables
#filter out data before 1900 as it is unreliable
sharks <- read_excel("sharks.xlsx") %>%
  select(1:14) %>%
  rename(Fatal = 'Fatal Y/N') %>%
  mutate(Year = as.numeric(Year),
         Sex = case_when(
           str_detect(tolower(Sex), "m") ~ "Male",
           str_detect(tolower(Sex), "f") ~ "Female",
           TRUE ~ NA_character_),
         Fatal = case_when(
           str_detect(tolower(Fatal), "n") ~ "No",
           str_detect(tolower(Fatal), "y") ~ "Yes",
           TRUE ~ NA_character_)) %>%
```

```r
  filter(Year > 1900) %>%
  select(Year, Sex, Fatal)

#generate posterior distribution of data for males
a_male <- 2
b_male <- 8
a_star_male <- nrow(sharks %>% filter(Sex == 'Male')
                    %>% filter(Fatal == "Yes")) + a_male
b_star_male <- nrow(sharks %>% filter(Sex == 'Male')
                    %>% filter(Fatal == "No")) + b_male

#graph the prior and posterior distribution for males
theta <- seq(0,.5,by = .0001)
plot(theta, dbeta(theta, a_star_male, b_star_male), type='l', col="red",
     xlab=expression(theta), ylab="Density",
     main="Fatality of Shark Attacks Among Males")
lines(theta, dbeta(theta, a_male, b_male), col="blue")
legend("topright", legend = c("Prior", "Posterior"),
       col = c("blue", "red"), lty = 1)

#generate posterior distribution for females
a_female <- 1
b_female <- 9
a_star_female <- nrow(sharks %>% filter(Sex == 'Female')
                      %>% filter(Fatal == "Yes")) + a_female
b_star_female <- nrow(sharks %>% filter(Sex == 'Female')
                      %>% filter(Fatal == "No")) + b_female

#graph prior and posterior distributions for females
theta <- seq(0,.5,by = .0001)
plot(theta, dbeta(theta, a_star_female, b_star_female), type='l', col="red",
     xlab=expression(theta), ylab="Density",
     main="Fatality of Shark Attacks Among Females")
lines(theta, dbeta(theta, a_female, b_female), col="blue")
legend("topright", legend = c("Prior", "Posterior"),
       col = c("blue", "red"), lty = 1)

#use Monte Carlo simulation to generate distribution for females minus males
diff_prior <- rbeta(100000, a_female, b_female) -
  rbeta(100000, a_male, b_male)
diff_post <- rbeta(100000, a_star_female, b_star_female) -
  rbeta(100000, a_star_male, b_star_male)
```

```r
#graph prior and posterior distribution for females minus males
plot(density(diff_post), xlab=expression(theta[Female]-theta[Male]),
     main="Difference in Fatality of Shark Attacks by Sex",
     col="red", xlim=c(-.25, .25))
lines(density(diff_prior), col="blue")
legend("topright", legend = c("Prior", "Posterior"),
       col = c("blue", "red"), lty = 1)

#95% credible interval for difference in fatalities between females and males
quantile(diff_post, c(.025, .975))
```