ST 540 – Applied Bayesian Analysis

Exam: Midterm 1 | Name: Chaitanya Rajeev | ID:200312570

Introduction

The data consists of information regarding no. of participants and no. of infected patients in two different groupings of the clinical trial of the Moderna vaccine developed against COVID-19. The first grouping consists of a Placebo group and a Vaccine Group. The second grouping consists of a white group and a non-white (communities of color) group. There are two parameters of interest: $\theta_{placebo}$(infection probability of the Placebo group) and $\theta_{vaccine}$(infection probability of the vaccine group). Another transformation parameter of interest is the efficacy of the vaccine, often defined as:

$$E = 1 - \frac{\theta_{vaccine}}{\theta_{placebo}}$$

Placebo Analysis

To start with, a placebo analysis of both race groups (combined) is conducted. Of the 14,073 patients given the Placebo treatment, 185 were infected with the virus. This data can be reasonably modelled with a Binomial distribution sample with 14,073 trials and 185 successes.

$$Y|\theta_{placebo} \sim Binomial(n, \theta_{placebo})$$

$Y \rightarrow no.of\ infected\ people\ (successes)\ in\ sample$
$n \rightarrow no.of\ participants\ (trials)$

'n' here is the number of trials. For an uninformative conjugate prior distribution of $\theta_{placebo}$, the Beta (1,1) distribution is chosen.

$$\theta_{placebo} \sim Beta(a = 1, b = 1)$$

The Beta-Binomial conjugate pair is often used for the estimation of a proportion parameter like theta where $\theta \in [0,1]$. The Binomial distribution assumption for the data is reasonable because the data can be compared to a series of trials and successful outcomes. Also, the number of infected

participants can only be discreet whole numbers which satisfies the requirements of the support. The Beta distribution for $\theta_{placebo}$ is assumed to have parameters a and b equal to 1. This is a valid assumption as it allows every possible value of theta to be equally likely and is thus uninformative. The pdf of this distribution resembles a continuous distribution between 0 and 1.

Derivation

In this section, we shall derive the posterior distribution of $\theta_{placebo}$ given the likelihood and the prior in the previous section.

$f(Y|\theta_{placebo}) = \binom{n}{Y}\theta_{placebo}^{Y}(1-\theta_{placebo})^{n-Y}$ (Binomial distribution pdf of the likelihood)

$\pi(\theta_{placebo}) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}$ (Beta distribution pdf of the prior)

$$P(\theta|Y) = \frac{f(Y|\theta).\pi(\theta)}{m(Y)}$$

$$= \frac{\left[\binom{n}{Y}\theta^{Y}(1-\theta)^{n-Y}\right].\left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}\right]}{m(Y)}$$

$$= \left[\binom{n}{Y}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\frac{1}{m(Y)}\right].\theta^{Y+a-1}(1-\theta)^{n-Y+b-1}$$

$$= C\theta^{Y+a-1}(1-\theta)^{n-Y+b-1} \quad \propto \quad \theta^{A-1}(1-\theta)^{B-1} \; where \; A = Y+a; B = n-Y+b$$

The posterior distribution is shown to be proportional to the kernel of a Beta (A, B) distribution.

$Hence, \; \theta|Y \sim Beta(Y+a, n-Y+b).$

Efficacy Analysis

In this section, a Bayesian analysis on the efficacy of the vaccine is presented. Both groups have a Beta (a = 1, b = 1) prior. Both Placebo and Vaccine data are assumed to come from a Binomial distribution. The Placebo data instance has 14,073 trials(participants) and 185 successes(infections). Similarly, the vaccine data instance has 14,134 trials(participants) and 11

successes(infections). As per the previous section, the posterior for $\theta_{placebo}$ can be derived as a Beta (185 + 1, 14073 – 185 + 1) = Beta (186,13889) distribution and that of $\theta_{vaccine}$, a Beta (11 + 1, 14134 – 11 + 1) = Beta (12,14124) distribution.

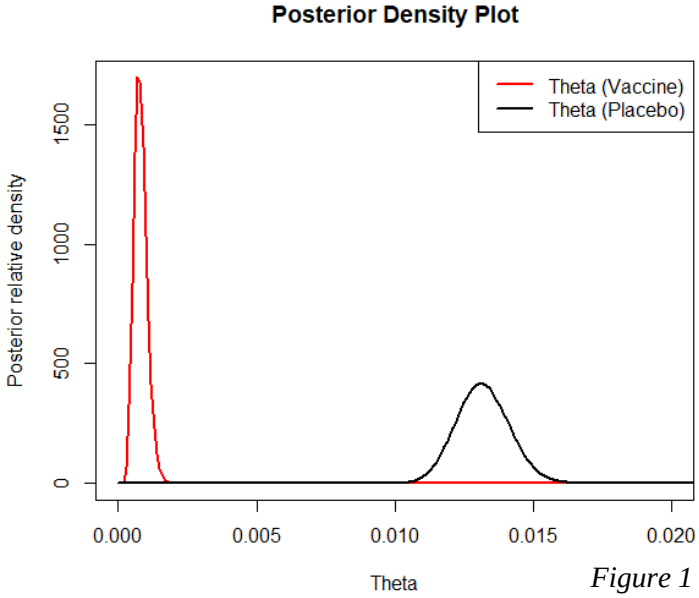

Figure 1

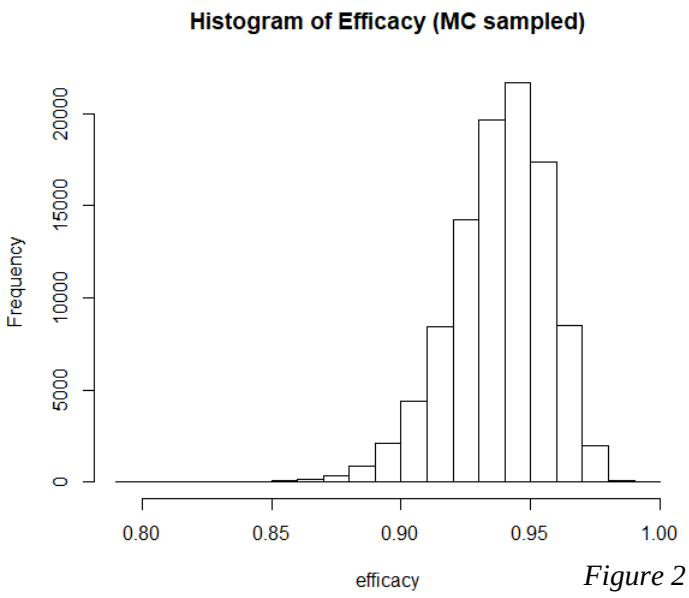

Figure 2

Figure 1 shows a plot of the Posterior relative densities for $\theta_{placebo}$ and $\theta_{vaccine}$. The Efficacy is calculated using 100,000 MC samples from both posterior distributions. Figure 2 shows a histogram for efficacy samples. Table 1 summarizes relevant statistics about the Efficacy.

Mean	Std. Dev	Quantile (2.5%)	Quantile (97.5%)
0.937	0.018	0.896	0.97

Table 1

The vaccine seems to have a mean Efficacy of 93.7% which is very high. Table 2 summarizes the results for Efficacy Analysis for different priors.

Prior	Efficacy			
	Mean	Std. Dev	Quantile (2.5%)	Quantile (97.5%)
Beta (1,1)	0.937	0.018	0.896	0.97
Beta (0.5,0.5)	0.938	0.018	0.895	0.97
Beta (2,2)	0.93	0.02	0.885	0.963
Beta (5,5)	0.915	0.021	0.867	0.952

Table 2

From the table, one can conclude that the results are stable and are not very sensitive to different priors. This could be because of the large difference in no. of trials and no. of successes in the data.

Hypothesis Test

Table 3

Prior	$P(E \geq 0.7)$
Beta (1,1)	1
Beta (0.5,0.5)	1
Beta (2,2)	1
Beta (5,5)	1

In this section, we test the hypothesis that Efficacy $E \geq 0.7$. The Bayesian hypothesis test yields 1 for a variety of priors as shown in Table 3. Hence, the null cannot be rejected, and we can be sure that the efficacy of the vaccine is at least 70%. The results are also not sensitive to the prior.

Subgroup Analysis

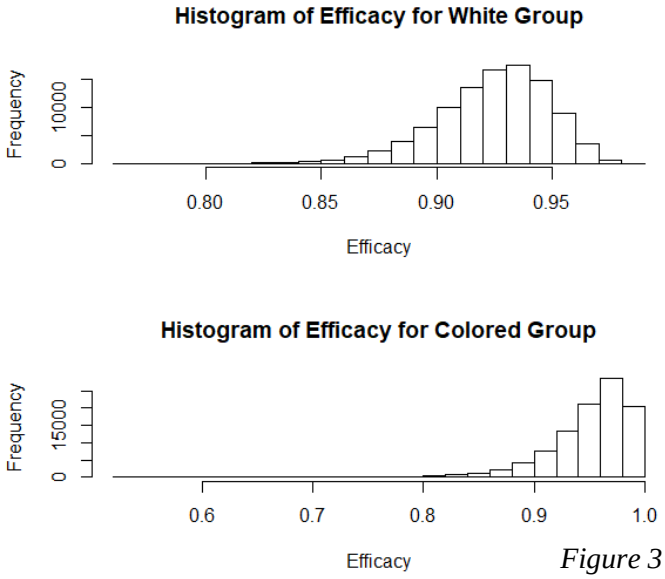

Figure 3

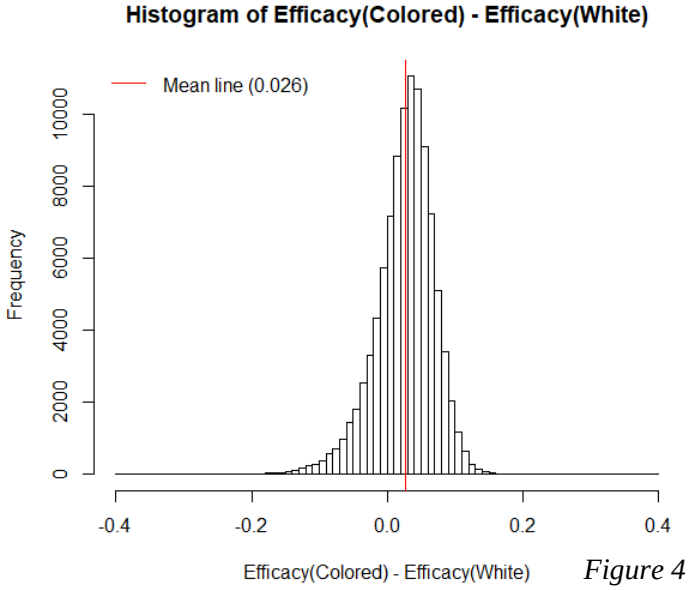

Figure 4

For the subgroup analysis, we compare the efficacies similarly computed for each the white group and the colored group. The relevant posterior distributions are summarized below (Beta(1,1) prior)

Group/Posterior	Placebo	Vaccine
White	$\theta_{placebo}\|Y \sim Beta(145,8773)$	$\theta_{vaccine}\|Y \sim Beta(11,9014)$
Colored	$\theta_{placebo}\|Y \sim Beta(42,5092)$	$\theta_{vaccine}\|Y \sim Beta(2,5088)$

The efficacies are again calculated using MC sampling from the posterior distributions. Figure 3 shows the histograms for efficacies of the two groups. In order to test whether the efficacies are different, we create a new transformation parameter for the difference between the two efficacies ($\nabla E = E_{Colored} - E_{White}$). Figure 4 shows a histogram for ∇E. We can create an equal-tailed 95% credible interval for ∇E. If the efficacies are indeed different, then the interval should not contain 0. The credible interval spans from -0.074 to 0.099. Hence, we can conclude that there is a lack of evidence of a significant difference between $E_{Colored}$ and E_{White}.

Appendix (R Code)

```
# defining inputs
P_All_part <- 14073
P_Wh_part <- 8916
P_Co_part <- 5132
P_All_inf <- 185
P_Wh_inf <- 144
P_Co_inf <- 41
V_All_part <- 14134
V_Wh_part <- 9023
V_Co_part <- 5088
V_All_inf <- 11
V_Wh_inf <- 10
V_Co_inf <- 1
# defining prior parameters
beta_a <- 1
beta_b <- 1
# Calculation of posterior densities for plotting
theta <- seq(0,0.2,0.0001)
posterior_plac <- dbeta(theta
                        ,P_All_inf + beta_a
                        ,P_All_part - P_All_inf + beta_b)
posterior_vacc <- dbeta(theta
                        ,V_All_inf + beta_a
                        ,V_All_part - V_All_inf + beta_b)
par(mfrow = c(1,1))
plot(theta,posterior_vacc, type = 'l',xlim =
c(0,0.02),col='red',lwd=2
     ,ylab = 'Posterior relative density'
     ,xlab = 'Theta', main = 'Posterior Density Plot')
lines(theta,posterior_plac, type = 'l',col='black',lwd=2)
legend("topright", c('Theta (Vaccine)','Theta (Placebo)')
       ,lty = 1, lwd = 2, col = c('red','black'))
# MC sampling from posterior to calculate Efficacy
theta_plac_mcmc <- rbeta(100000
                         ,P_All_inf + beta_a
                         ,P_All_part - P_All_inf + beta_b)
theta_vacc_mcmc <- rbeta(100000
                         ,V_All_inf + beta_a
                         ,V_All_part - V_All_inf + beta_b)
efficacy = 1 - (theta_vacc_mcmc/theta_plac_mcmc)
# getting efficacy histogram and summaries
hist(efficacy, main = 'Histogram of Efficacy (MC sampled)')
mean(efficacy)
sd(efficacy)
quantile(efficacy,c(0.025,0.975))
# testing efficacy is atleast 0.7
mean(efficacy >= 0.7)
```

```
# Calculating efficacies for White and Colored groups
theta_plac_mcmc_white <- rbeta(100000
                          ,P_Wh_inf + beta_a
                          ,P_Wh_part - P_Wh_inf + beta_b)
theta_vacc_mcmc_white <- rbeta(100000
                          ,V_Wh_inf + beta_a
                          ,V_Wh_part - V_Wh_inf + beta_b)
efficacy_white = 1 - (theta_vacc_mcmc_white/theta_plac_mcmc_white)

theta_plac_mcmc_color <- rbeta(100000
                               ,P_Co_inf + beta_a
                               ,P_Co_part - P_Co_inf + beta_b)
theta_vacc_mcmc_color <- rbeta(100000
                               ,V_Co_inf + beta_a
                               ,V_Co_part - V_Co_inf + beta_b)
efficacy_colored = 1 - (theta_vacc_mcmc_color/theta_plac_mcmc_color)

# plotting histograms for both efficacies
par(mfrow=c(2,1))
hist(efficacy_white,main = 'Histogram of Efficacy for White Group',xlab =
'Efficacy')
hist(efficacy_colored,main = 'Histogram of Efficacy for Colored
Group',xlab = 'Efficacy')

# plotting histogram for difference(delta_E) and calculated credible
interval
quantile(efficacy_colored - efficacy_white, c(0.025,0.975))
par(mfrow = c(1,1))
hist(efficacy_colored-efficacy_white,breaks = seq(-0.4,0.4,0.01)
     ,main = 'Histogram of Efficacy(Colored) - Efficacy(White)'
     ,xlab = 'Efficacy(Colored) - Efficacy(White)')
abline(v = mean(efficacy_colored - efficacy_white),col = 'red')
legend('topleft',c('Mean line (0.026)'),col = c('red'),lty = 1,bty = "n")
```