Worksheet 3

Foundations of Bayesian Methodology

Wenjie Tu

Spring Semester 2022

Exercise 1 (Individual project (Part 3A))

Table 1: ASAS20 responders at week 6

Group	n	Responders x (%)	Posterior Response rate	Difference (S-P)	95% CrI	PPS	MCse (PPS)
Secukinumab	23	14 (60.9%)	59.5%	34.7%	$11.5 - 56.4\%$	99.8%	?
Placebo	6	1 (16.7%)	24.5%				

Provide the estimate of the posterior probability of superiority (PPS) in Table 1 and its Monte Carlo standard error.

1.1

Given the posterior distribution $\text{Beta}(14.5, 10)$ of secukinumab group and the posterior distribution $\text{Beta}(12, 37)$ of placebo group, we would like to investigate by how much these two distributions differ.

We first run Monte Carlo simulation from the known posterior distributions of both groups. The simulation proceeds as follows:

- Generate a MC sample with a size of 1000000 from the posterior distribution of Secukinumab group
- Generate a MC sample with a size of 1000000 from the posterior distribution of Placebo group
- Take the difference in samples (i.e. response rate) between two groups
- Plot the distribution of the response rate difference
- Compute the mean and the standard error of the distribution of the response rate difference

We then construct a response rate difference (RRD) based on both MC samples simulated above:

$$RRD = \theta_\text{S} - \theta_\text{P}$$

- θ_S: sample response rate in Secukinumab group
- θ_P: sample response rate in Placebo group

```
## Set the seed for reproducible results
set.seed(44566)

## MC sample size
M <- 1000000

## Generate random numbers from posterior distribution in Secukinumab group
mc.secukin <- rbeta(M, shape1=14.5, shape2=10)

## Generate random numbers from posterior distribution in Placebo group
mc.placebo <- rbeta(M, shape1=12, shape2=37)
```

1.2

```r
## Load libraries
library(ggplot2)
```

```r
## Construct the response rate difference
mc.rrd <- mc.secukin - mc.placebo

## Generate a dataframe for histograms in ggplot
d.hist <- data.frame(
  ResponseRate = c(mc.secukin, mc.placebo, mc.rrd),
  Group = rep(c("Secukinumab", "Placebo", "RRD"), each=M)
)
```

```r
ggplot(d.hist, aes(x=ResponseRate, y=..density.., color=Group, fill=Group)) +
  geom_histogram(position="identity", bins=50, alpha=0.2) +
  labs(title="Histograms of MC samples",
       x="Response rate (difference)", y="Density") + theme_minimal()

ggplot(d.hist, aes(x=ResponseRate, color=Group, fill=Group)) +
  geom_density(alpha=0.2) +
  labs(title="Density plots of MC samples",
       x="Response rate (difference)", y="Density") + theme_minimal()
```


1.3
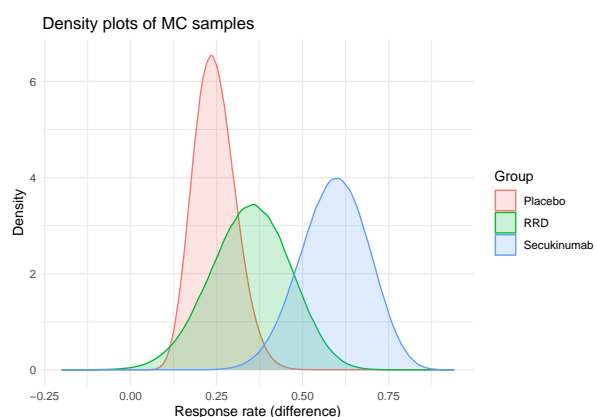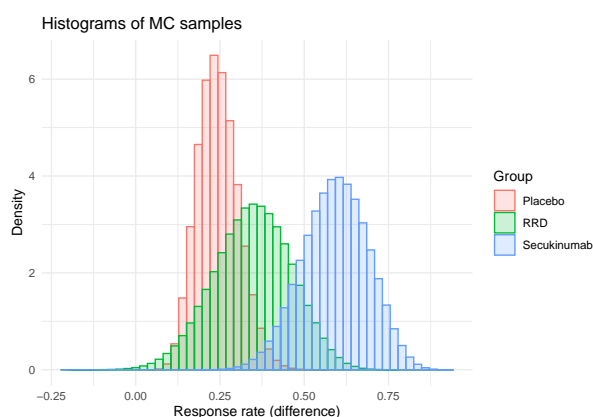
```r
## Compute 2.5%, 50%, 97.5% quantitles
rrd.qtl <- quantile(mc.rrd, probs=c(0.025, 0.5, 0.975)); rrd.qtl
```

```
##      2.5%       50%     97.5%
## 0.1156575 0.3495868 0.5627693
```

Interpretation:

When we draw a MC sample (size=1000000) from posterior distribution $\text{Beta}(14.5, 10)$ in Secukinumab group and a MC sample (size=1000000) from the posterior distribution $\text{Beta}(12, 37)$ in the placebo group, we can construct an empirical density of response rate difference (RRD) by simply taking the difference based on the two MC samples generated. All the following interpretations are based on the empirical density of RRD constructed above:

- The median 0.3495868 corresponds to the 50% probability of the sample density of RRD
- The equi-tailed credible interval $(0.1156575, 0.5627693)$ corresponds to the 95% probability of the sample density of RRD with equal tails (each tail corresponds to 2.5% probability)

2

1.4

PPS under RRD measure is defined as:

$$P(\theta_S - \theta_P > 0) \equiv P(RRD > 0)$$

```
## Posterior probability of superiority
rrd.pps <- mean(mc.rrd > 0); rrd.pps
```

```
## [1] 0.998305
```

The estimate of PPS is 99.83%, which is larger than the threshold of 95%.

1.5

MC standard error of PPS:

```
## MC standard error of PPS
rrd.se <- sqrt(var(mc.rrd > 0)/M); rrd.se
```

```
## [1] 4.113549e-05
```

```
d.rrd <- with(density(mc.rrd), data.frame(x, y))
ggplot(data=d.rrd, mapping=aes(x=x, y=y)) +
  geom_area(aes(x=ifelse(x>rrd.qtl[1] & x<rrd.qtl[3], x, 0), y=y),
            fill="lightblue", alpha=0.5) +
  geom_area(aes(x=ifelse(x<0, x, 0), y=y), fill="red", alpha=0.5) +
  geom_line(color="darkblue") + geom_vline(xintercept=0, color="red") +
  labs(title="Density of RRD samples with equi-tailed 95% CrI",
       x="Response rate difference", y="Density") +
  ylim(0, 4) + theme_minimal()
```
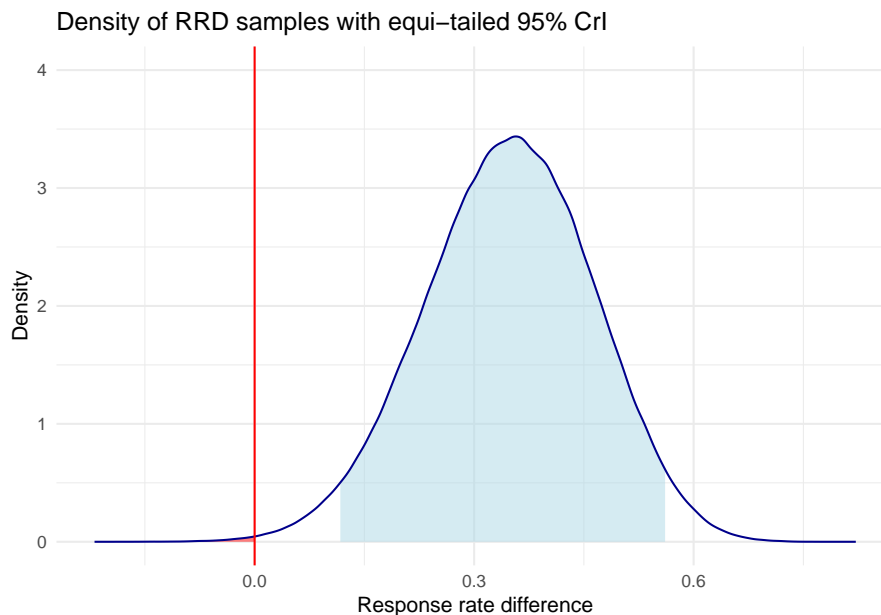

The figure above displays the density of difference of MC samples between Secukinumab and placebo groups with equi-tailed 95% credible interval. The red vertical line corresponds to the baseline (i.e. $RRD = 0$). The red shaded area corresponds to the probability of $RRD \leq 0$. The blue shaded area corresponds to the equi-tailed 95% credible probability. We can clearly see that the red vertical line lies outside the equi-tailed 95% credible interval. We can also get a numerical sense of how likely $RRD \leq 0$ is:

```
## P[RDD <= 0]
p.rrd <- mean(mc.rrd <= 0); p.rrd
```

```
## [1] 0.001695
```

Exercise 2 (Individual project (Part 3B))

Extend the response rate difference (RRD) argument outlined in Exercise 1 (Table 1) to response ratio
(RR) and odds ratio (OR).

Response ratio (RR)

- Construct the response ratio based on the MC samples
- Compute the median and the equi-tailed 95% credible interval
- Visualize the distribution of the response ratio
- Interpret the results

```
## Construct the response ratio
mc.rr <- mc.secukin/mc.placebo

## Compute 2.5%, 50%, 97.5% quantitles
rr.qtl <- quantile(mc.rr, probs=c(0.025, 0.5, 0.975)); rr.qtl
```

```
##     2.5%      50%    97.5%
## 1.361472 2.442151 4.614866
```

```
ggplot(data.frame(ResponseRatio=mc.rr), aes(x=ResponseRatio, y=..density..)) +
  geom_histogram(position="identity", bins=50, alpha=0.2, fill=5, color=5) +
  geom_density(color=5) +
  labs(title="Distribution of response ratio based on MC samples",
       x="Response Ratio", y="Density") +
  theme_minimal()

d.rr <- with(density(mc.rr), data.frame(x, y))
ggplot(data=d.rr, mapping=aes(x=x, y=y)) +
  geom_area(aes(x=ifelse(x>rr.qtl[1] & x<rr.qtl[3], x, 0), y=y),
            fill="lightblue", alpha=0.5) +
  geom_area(aes(x=ifelse(x<1, x, 0), y=y), fill="red", alpha=0.5) +
  geom_line(color="darkblue") + geom_vline(xintercept=1, color="red") +
  labs(title="Density of RR samples with equi-tailed 95% CrI",
       x="Response rate ratio", y="Density") +
  ylim(0, 0.6) + theme_minimal()
```
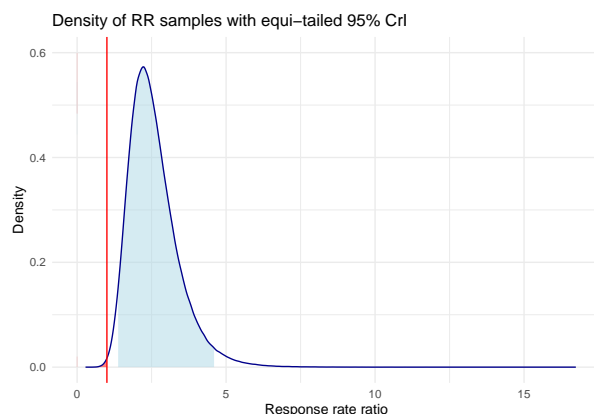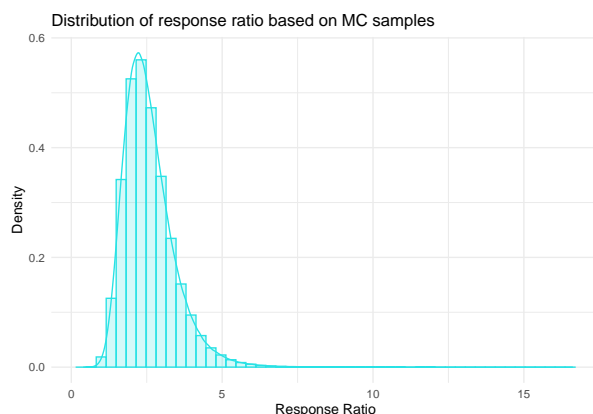
PPS under RR measure is defined as:

$$P\left(\frac{\theta_S}{\theta_P} > 1\right) \equiv P(RR > 0)$$

```
## Compute the PPS under response ratio measure
rr.pps <- mean(mc.rr > 1); rr.pps
```

```
## [1] 0.998305
```

```
## Compute the MC standard error
rr.se <- sqrt(var(mc.rr > 1)/M); rr.se
```

```
## [1] 4.113549e-05
```

Odds ratio (OR)

- Construct the odds ratio based on the MC samples
- Compute the median and the equi-tailed 95% credible interval
- Visualize the distribution of the odds ratio
- Interpret the results

```
## Construct the response ratio
mc.or <- (mc.secukin/(1-mc.secukin)) / (mc.placebo/(1-mc.placebo))

## Compute the MC standard error
or.qtl <- quantile(mc.or, probs=c(0.025, 0.5, 0.975)); or.qtl
```

```
##      2.5%       50%      97.5%
##   1.656347  4.638885 13.737391
```

```
ggplot(data.frame(OddsRatio=mc.or), aes(x=OddsRatio, y=..density..)) +
  geom_histogram(position="identity", bins=50, alpha=0.2, fill=6, color=6) +
  geom_density(color=6) +
  labs(title="Distribution of odds ratio based on MC samples",
       x="Odds Ratio", y="Density") +
  theme_minimal()

d.or <- with(density(mc.or), data.frame(x, y))
ggplot(data=d.or, mapping=aes(x=x, y=y)) +
  geom_area(aes(x=ifelse(x>or.qtl[1] & x<or.qtl[3], x, 0), y=y),
            fill="lightblue", alpha=0.5) +
  geom_area(aes(x=ifelse(x<1, x, 0), y=y), fill="red", alpha=0.5) +
  geom_line(color="darkblue") +
  geom_vline(xintercept=or.qtl[2], color="darkblue") +
  geom_vline(xintercept=1, color="red") +
  labs(title="Density of OR samples with equi-tailed 95% CrI",
       x="Response rate odds ratio", y="Density") +
  ylim(0, 0.2) + theme_minimal()
```

| Distribution of odds ratio based on MC samples | Density of OR samples with equi–tailed 95% CrI |
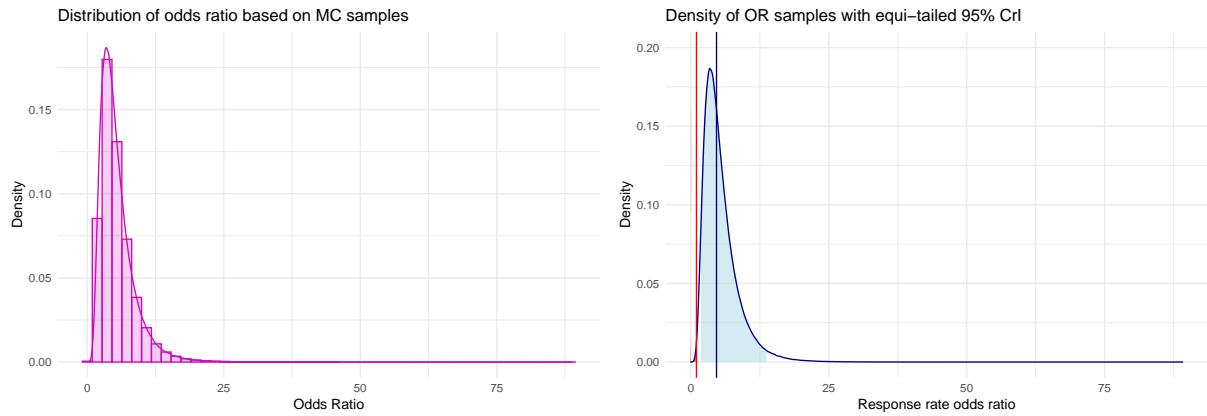
PPS under OR measure is defined as:

$$P\left(\frac{\theta_{\mathrm{S}}}{1-\theta_{\mathrm{S}}}\Big/\frac{\theta_{\mathrm{P}}}{1-\theta_{\mathrm{P}}} > 1\right) \equiv P(OR > 1)$$

```r
## Compute the PPS under odds ratio measure
or.pps <- mean(mc.or > 1); or.pps
```

```
## [1] 0.998305
```

```r
## Compute the MC standard error
or.se <- sqrt(var(mc.or > 1)/M); or.se
```

```
## [1] 4.113549e-05
```

Summary:

We construct three types of effect size in order to compute the posterior probability of superiority of Secukinumab (S) over Placebo (P):

$$\mathrm{PPS} = P(\theta_{\mathrm{S}} > \theta_{\mathrm{P}}) = \begin{cases} P(\mathrm{RRD} > 0) & \mathrm{RRD} = \theta_{\mathrm{S}} - \theta_{\mathrm{P}} \\ P(\mathrm{RR} > 1) & \mathrm{RR} = \frac{\theta_{\mathrm{S}}}{\theta_{\mathrm{P}}} \\ P(\mathrm{OR} > 1) & \mathrm{OR} = \frac{\theta_{\mathrm{S}}}{1-\theta_{\mathrm{S}}}\Big/\frac{\theta_{\mathrm{P}}}{1-\theta_{\mathrm{P}}} \end{cases}$$

```r
## Compute the means for RRD, RR, OR
rrd.m <- mean(mc.rrd)
rr.m <- mean(mc.rr)
or.m <- mean(mc.or)

## Print the results
cat(sprintf(
  "Mean of RDD: %.4f\nMean of RR: %.4f\nMean of OR: %.4f",
  rrd.m, rr.m, or.m
))
```

```
## Mean of RDD: 0.3469
## Mean of RR: 2.5823
## Mean of OR: 5.4187
```

```
## Generate a table for summary statistics
d.summary <- data.frame(
  rbind(rrd.qtl, rr.qtl, or.qtl),
  c(rrd.m, rr.m, or.m),
  c(rrd.pps, rr.pps, or.pps),
  c(rrd.se, rr.se, or.se)
)
colnames(d.summary) <- c("2.5% Quantile", "Median", "97.5% Quantile",
                         "Mean", "PPS", "MCse(PPS)")
rownames(d.summary) <- c("RRD", "RR", "OR")
knitr::kable(d.summary, align="c", caption="Summary statistics")
```

Table 2: Summary statistics

	2.5% Quantile	Median	97.5% Quantile	Mean	PPS	MCse(PPS)
RRD	0.1156575	0.3495868	0.5627693	0.3469135	0.998305	4.11e-05
RR	1.3614717	2.4421509	4.6148657	2.5822664	0.998305	4.11e-05
OR	1.6563465	4.6388852	13.7373911	5.4186874	0.998305	4.11e-05

Unsurprisingly we obtain the same PPS for three types of effect size (RDD, RR, OR) as these three in essence measure the same posterior probability (i.e. $P(\theta_\text{S} > \theta_\text{P})$).