Worksheet 4

Foundations of Bayesian Methodology

Wenjie Tu

Spring Semester 2022

Exercise 1 (Individual project (Part 4))

1.1

```
d.info <- data.frame(matrix(c(20, 0.6, 5, 0.25), nrow=2, byrow=TRUE))
colnames(d.info) <- c("Sample Size", "Response Rate")
rownames(d.info) <- c("Secukinumab", "Placebo")
knitr::kable(d.info, align=c("cc"), caption="Bayesian study design 4:1")
```

Table 1: Bayesian study design 4:1

	Sample Size	Response Rate
Secukinumab	20	0.60
Placebo	5	0.25

In the Secukinumab group:

- $P_{\mathrm{S}} = 0.6$, $n_{\mathrm{S}} = 20$
- $x_S^{(i)} \sim \mathrm{Bin}(n_{\mathrm{S}}, p_{\mathrm{S}})$
- $\hat{p}_{\mathrm{S}}^{(i)} = \dfrac{x_S^{(i)}}{n_{\mathrm{S}}}$

In the Placebo group:

- $P_{\mathrm{P}} = 0.25$, $n_{\mathrm{P}} = 5$
- $x_P^{(i)} \sim \mathrm{Bin}(n_{\mathrm{P}}, p_{\mathrm{P}})$
- $\hat{p}_{\mathrm{P}}^{(i)} = \dfrac{x_P^{(i)}}{n_{\mathrm{P}}}$

Construct the response rate difference between secukinumab and placebo):

$$d^{(i)} = \hat{p}_{\mathrm{S}}^{(i)} - \hat{p}_{\mathrm{P}}^{(i)}$$

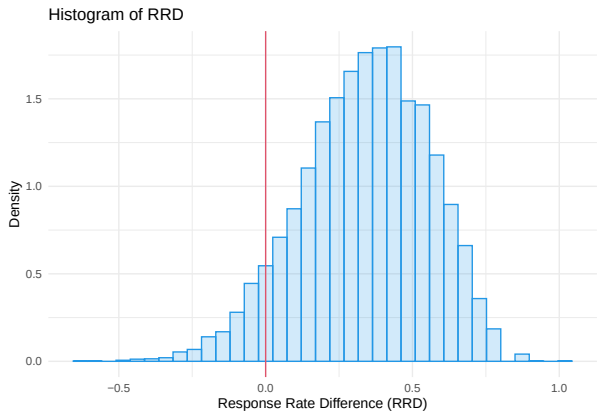
We are interested in the sample distribution of $d^{(i)}$ and would like to investigate the proof of concept (i.e. the probability that ASAS20 response rate on secukinumab is larger than that on placebo):

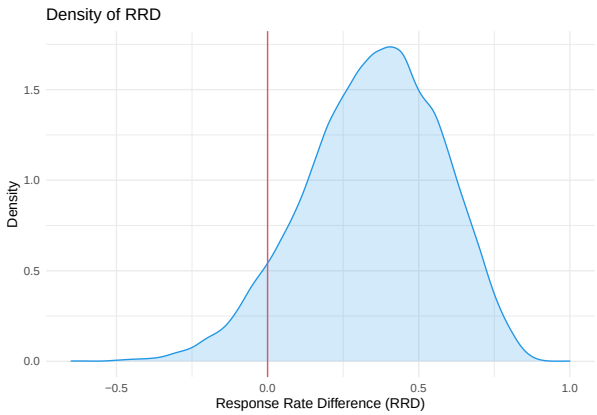
$$\mathrm{POC} = P[D > 0]$$

where D is the random variable and $d^{(i)}$ is the sample realizations

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

n.s <- 20        # sample size in Secukinumab
n.p <- 5         # sample size in Placebo

prob.s <- 0.6    # true response rate in Secukinumab
prob.p <- 0.25   # true response rate in Placebo


M <- 10000       # MC sample size
x.s <- rbinom(M, size=n.s, prob=prob.s)
x.p <- rbinom(M, size=n.p, prob=prob.p)

RR.s <- x.s / n.s      # MC sample response rate in Secukinumab
RR.p <- x.p / n.p      # MC sample response rate in Placebo
RRD <- RR.s - RR.p     # MC sample response rate difference
```

```r
library(ggplot2)

## Histogram
ggplot(data.frame(RRD=RRD), aes(x=RRD, y=..density..)) +
  geom_histogram(bins=35, color=4, fill=4, alpha=0.2) +
  geom_vline(xintercept=0, color=2) +
  labs(title="Histogram of RRD", x="Response Rate Difference (RRD)", y="Density") +
  theme_minimal()

## Density
ggplot(data.frame(RRD=RRD), aes(x=RRD, y=..density..)) +
  geom_density(color=4, fill=4, alpha=0.2) +
  geom_vline(xintercept=0, color=2) +
  labs(title="Density of RRD", x="Response Rate Difference (RRD)", y="Density") +
  theme_minimal()
```






```r
## Proof of concept (i.e. P[RRD > 0])
poc <- mean(RRD > 0); poc
```

```
## [1] 0.9146
```

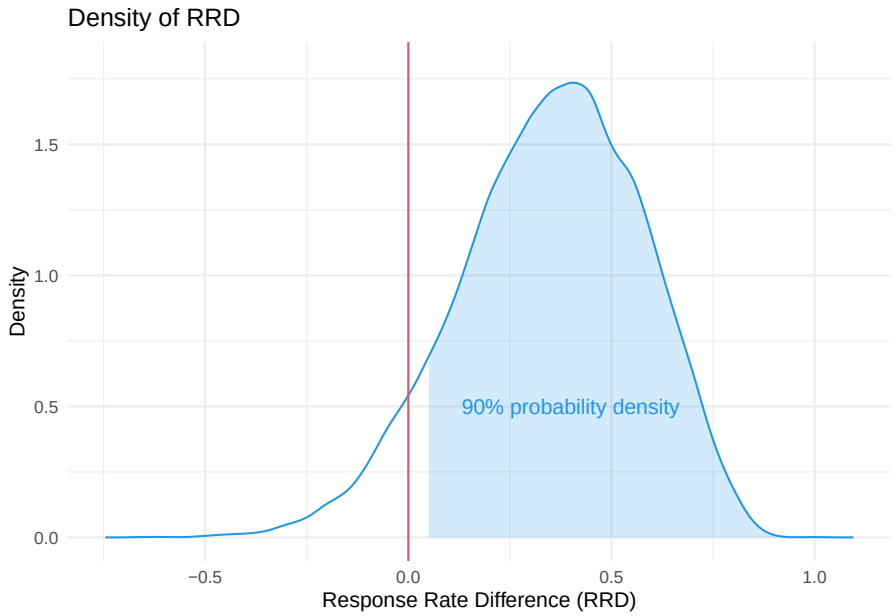
```r
## MC standard error
poc.se <- sqrt(var(RRD > 0)/M); poc.se
```

```
## [1] 0.0027949
```

```r
## 99% confidence interval
poc.CI99 <- c(poc - 3 * poc.se, poc + 3 * poc.se); poc.CI99
```

```
## [1] 0.9062153 0.9229847
```

```r
d.density <- with(density(RRD), data.frame(x, y))
ggplot(data=d.density, mapping=aes(x=x, y=y)) +
  geom_area(aes(x=ifelse(x>quantile(RRD, probs=0.1), x, 0), y=y), fill=4, alpha=0.2) +
  geom_line(color=4) + geom_vline(xintercept=0, color=2) + ylim(0, 1.8) +
  labs(title="Density of RRD", x="Response Rate Difference (RRD)", y="Density") +
  geom_text(aes(x=0.4, y=0.5, label="90% probability density"),
            color=4, check_overlap=TRUE) + theme_minimal()
```



1.2

```r
## Define a function for design analysis
## Params: M, size1, size2, prob1, prob2

design.analysis <- function(M,          # Monte Carlo sample size
                            size1,      # number of patients on secukinumab
                            size2,      # number of patients on placebo
                            prob1=0.6,  # response rate on secukinumab
                            prob2=0.25  # response rate on placebo
                            ) {
  ## MC number of responders on secukinumab
  mc1 <- rbinom(M, size=size1, prob=prob1)

  ## MC number of responders on placebo
  mc2 <- rbinom(M, size=size2, prob=prob2)

  RR1 <- mc1/size1  # MC response rate in Secukinumab
  RR2 <- mc2/size2  # MC response rate in Placebo
  RRD <- RR1 - RR2  # MC response rate difference

  ## Proof of concept
```

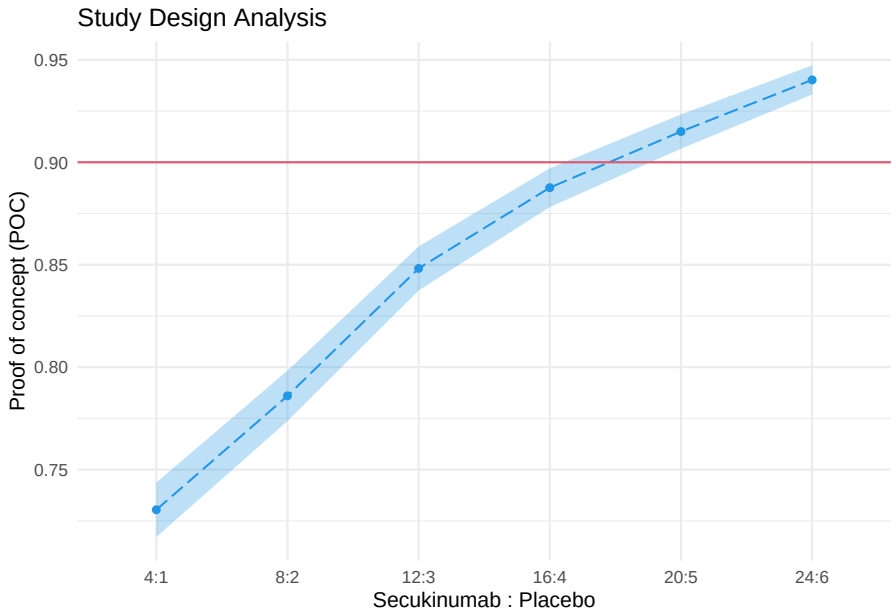
```
  poc <- mean(RRD > 0)

  ## MC standard error of POC
  poc.se <- sqrt(var(RRD > 0)/M)

  return(list(poc=poc, poc.se=poc.se))
}
```

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

df <- data.frame(matrix(nrow=6, ncol=5))
colnames(df) <- c("Design", "POC", "MCse", "lower99", "upper99")

for (i in 1:6) {
  results  <- design.analysis(M=10000, size1=i*4, size2=i, prob1=0.6, prob2=0.25)
  df[i, 1] <- sprintf("%.0f:%.0f", i*4, i)
  df[i, 2] <- results$poc
  df[i, 3] <- results$poc.se
  df[i, 4] <- results$poc - 3*results$poc.se
  df[i, 5] <- results$poc + 3*results$poc.se
}

knitr::kable(df, align=c("lccccc"), caption="POCs for different designs")
```

Table 2: POCs for different designs

Design	POC	MCse	lower99	upper99
4:1	0.7304	0.0044377	0.7170868	0.7437132
8:2	0.7860	0.0041015	0.7736956	0.7983044
12:3	0.8482	0.0035884	0.8374347	0.8589653
16:4	0.8876	0.0031587	0.8781238	0.8970762
20:5	0.9150	0.0027890	0.9066331	0.9233669
24:6	0.9402	0.0023713	0.9330862	0.9473138

```
lvls <- c("4:1", "8:2", "12:3", "16:4", "20:5", "24:6")
ggplot(df, aes(x=1:6, y=POC)) +
  geom_point(aes(x=factor(Design, levels=lvls)), color=4) +
  geom_line(color=4, linetype="longdash") +
  geom_ribbon(aes(ymin=lower99, ymax=upper99), fill=4, alpha=0.3) +
  geom_hline(yintercept=0.9, color=2) +
  labs(title="Study Design Analysis", x="Secukinumab : Placebo",
       y="Proof of concept (POC)") + theme_minimal()
```



Conditional on POC $> 90\%$, we see that the 4:1 study design based on 20 patients on secikinumab and 5 patients on placebo is the smallest number of patients for response rates (25% on placebo and 60% secukinumab)

1.3

Approach:

- From Baeten et al. (2013), there are 20 patients on Secukinumab with a true response rate of 60% and 5 patients on Placebo with a true response rate of 25%. Binomial distrbutions are used to model response rates in Secukinumab and Placebo groups.
- Given the data, Monte Carlo simulations are conducted on Secukinumab and Placebo separately to obtain the sample response rates for both groups.
- With the simulated data, the response rate difference (RRD) is constructed by taking the first difference between the sample response rate on Secukinumab and the sample response rate on Placebo. Hence, the sample distribution of RRD is obtained.
- The proof of concept is defined as the probability that the response rate on secukinumab is larger than that on Placebo. The question of interest can be easily translated into the form of RRD - POC $= P(\text{RRD} > 0)$.
- To show POC > 90, one simply has to show $P(\text{RRD} > 0) > 90\%$. With the sample distribution of RRD, $P(\text{RRD} > 0)$ (i.e. POC) can be easily computed using `mean(RRD > 0)`.
- Once the POC is computed, one can compare it with 90%. As demonstrated above, POC $= 91.46\% > 90\%$.