Final Report

STA421: Foundations of Bayesian Methodology

Wenje Tu

Zurich, May 2022

Contents

1 Classical analysis and sample size computation

In this chapter, we first explain the main goal of the study by Baeten et al. (2013), on which this report is based. We then provide preliminary analyses on the data using classical statistical methods. We next compute the optimal sample size to obtain a target power.

1.1 Classical analysis and sample size computation

1(a) Explain the main goal of the study

The main goal of the scientific article by Baeten et al. (2013) is to assess the efficacy and safety of the anti-IL-17A monoclonal antibody secukinumab in treating patients with active ankylosing spondylitis. A randomized double-blind proof-of-concept study was conducted and the randomization was done with a computer-generated block randomization list without a stratification process. The primary efficacy endpoint was the percentage of patients with a 20% response, and safety was assessed up to week 28.

1(b) Apply classical methods to the data

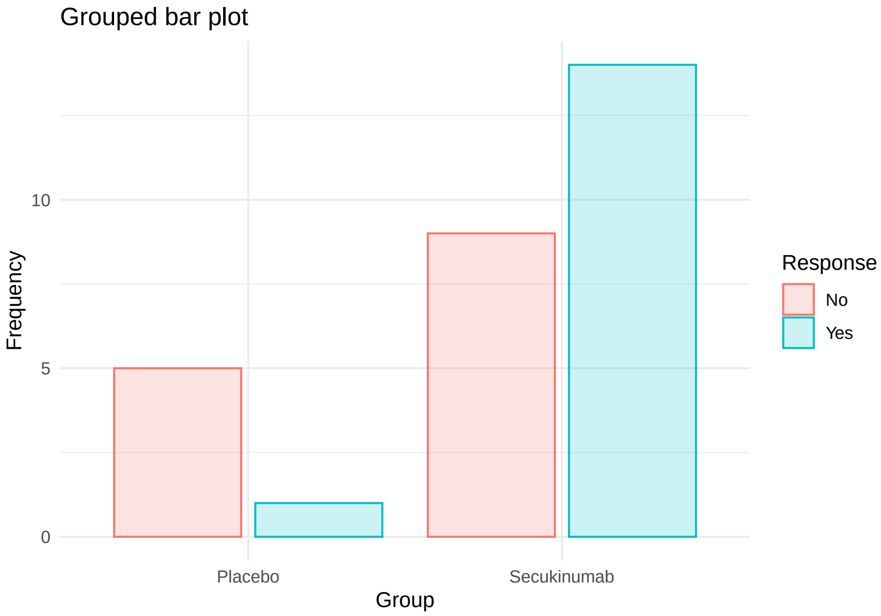
Table 1: Contingency table

	Non-responders	Responders	Total
Placebo	5	1	6
Secukinumab	9	14	23
Total	14	15	29

```r
d.baeten <- data.frame(Group = rep(c("Placebo", "Secukinumab"), c(6, 23)),
                       Response = rep(c("No", "Yes", "No", "Yes"), c(5, 1, 9, 14)))
table(d.baeten)
```

```
##              Response
## Group          No Yes
##   Placebo       5   1
##   Secukinumab   9  14
```

```r
## Data Visualization
library(ggplot2)
ggplot(d.baeten, aes(x=Group, fill=Response, color=Response)) +
  geom_bar(stat="count", position=position_dodge2(preserve="single"), alpha=0.2) +
  labs(title="Grouped bar plot", y="Frequency") + theme_minimal()
```



From Table 3 (the contingency table), we see that the sample size for some cells is small (i.e. below 5). As a rule of thumb, we use the Fisher's exact test here as the Chi-square test might be unreliable with small sample size.

Table 2: Chi-squared test vs Fisher's exact test

Criterion	**Chi-squared test**	**Fisher's exact test**
Minimal sample size	Large	Small
Accuracy	Approximate	Exact
Contingency table	Arbitrary dimension	Usually 2×2
Interpretation	Pearson residuals	Odds ratio

Source: *Testing Independence: Chi-Squared vs Fisher's Exact Test*

Fisher's Exact Test

```r
## Fisher's exact test
fisher.test(table(d.baeten), conf.level=0.95)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  table(d.baeten)
## p-value = 0.08008
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##    0.6570166 392.5294804
## sample estimates:
## odds ratio
##   7.259346
```

The odds ratio in our context represents that the odds that the patient responded in Secukinumab group, compared to the odds of the response in the Placebo group. We obtain a p-value of 0.08 and it can be interpreted that the odds ratio is equal to 1 with the probability of $(1 - 0.08)$ 92% and that the odds ratio is not equal to 1 with the probability of 8%. We set the confidence level at 95% and 0.08 is larger than 0.05. We can also see that 1 falls within the 95% confidence interval (0.657, 392.53). We thus cannot reject the null hypothesis that the true odds ratio is not equal to 1 with 95% confidence level.

Fitting Models with Disaggregate Data

```
d.baeten <- data.frame(group = rep(c(0, 1), c(6, 23)),
                       response = rep(c(0, 1, 0, 1), c(5, 1, 9, 14)))
table(d.baeten)
```

```
##      response
## group  0  1
##     0  5  1
##     1  9 14
```

Logistic regression:

$$log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_i$$

- y_i: 1 if the patient responded, 0 otherwise
- x_i: 1 if the patient was in Secukinumab group, 0 if in Placebo group
- $p_i = P(y_i = 1 \mid x_i) = \mathbb{E}[y_i \mid x_i] = \mu_i$

```
## Logistic regression
logistic.reg <- glm(response ~ group, data=d.baeten, family=binomial)
summary(logistic.reg)
```

```
##
## Call:
## glm(formula = response ~ group, family = binomial, data = d.baeten)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.3699  -1.3699   0.9964   0.9964   1.8930
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -1.609      1.095  -1.469   0.1418
## group          2.051      1.176   1.745   0.0811 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 40.168  on 28  degrees of freedom
## Residual deviance: 36.196  on 27  degrees of freedom
## AIC: 40.196
##
## Number of Fisher Scoring iterations: 4
```

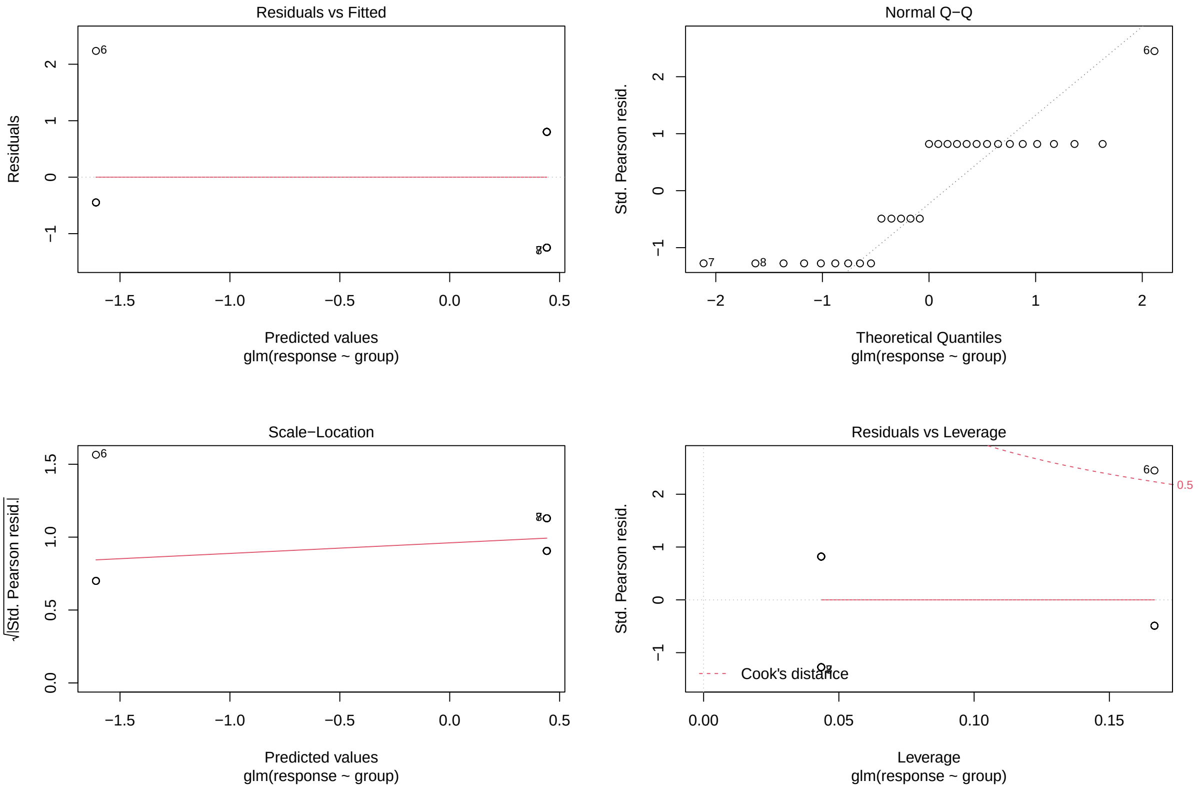
Global test: $H_0 : \beta_1 = 0$, $D(y, \hat{p}_{Null}) - D(y, \hat{p}_{Null}) \sim \chi^2$

```
1 - pchisq((logistic.reg$null.deviance - logistic.reg$deviance),
           df=(logistic.reg$df.null - logistic.reg$df.residual))
```

```
## [1] 0.04625658
```

The p-value is 0.046, hence the null hypothesis is reject. There is a significant effect of *group* on the odds of *response*. The coefficient can be interpreted as the treatment effect (Secukinumab vs. Placebo) on the log odds ratio of response.

```
## Residual analysis
plot(logistic.reg)
```



- There seems to be a violation of assumption of constant variance.
- There is a potential influential data point with index 6.

Poisson regression:

$$\log(\lambda_i) = \beta_0 + \beta_1 x_i$$

- y_i: 1 if the patient responded, 0 otherwise
- x_i: 1 if the patient was in Secukinumab group, 0 if in Placebo group
- $\lambda_i = \mathbb{E}[y_i \mid x_i]$

```
## Poisson regression
poisson.reg <- glm(response ~ group, data=d.baeten, family=poisson(link="log"))
summary(poisson.reg)
```

```
##
## Call:
## glm(formula = response ~ group, family = poisson(link = "log"),
##     data = d.baeten)
##
## Deviance Residuals:
```

```
##      Min        1Q    Median        3Q       Max
## -1.1034   -1.1034    0.4586    0.4586    1.3845
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.7918     0.9999  -1.792   0.0731 .
## group         1.2953     1.0350   1.252   0.2107
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 19.777  on 28  degrees of freedom
## Residual deviance: 17.484  on 27  degrees of freedom
## AIC: 51.484
##
## Number of Fisher Scoring iterations: 5
```

Does the Model Fit?

Null hypothesis: "The model fits well"

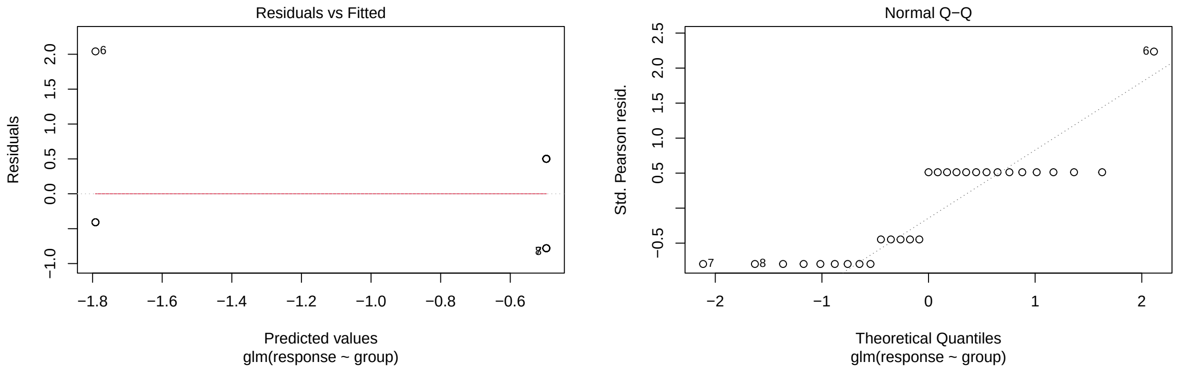
- Quick check: *residual deviance* $\gg$ *df*?
  - From the output, we see that $17.484 \gg 27$ does not hold so it seems that we do not have overdispersion issue.

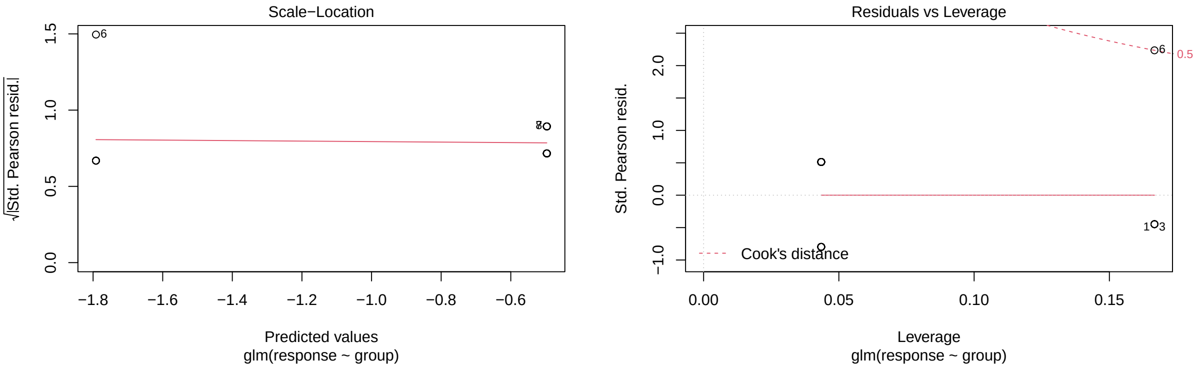
```
pchisq(17.484, 27, lower.tail=FALSE)
```

```
## [1] 0.9185002
```

- More exact test:
  - Null hypothesis: "The model fits well"
  - We obtain a p-value of 0.92 and we cannot reject the null hypothesis. The model fits OK.

```
## Residual analysis
plot(poisson.reg)
```





- Diagnostic plots look OK but we have too few data so these plots may be unreliable.

1(c) Compute the optimal sample size

- Design: 1:1 (i.e. equal size for both treatment and control groups)
- Response rate in Secukinumab group: 60%
- Response rate in Placebo group: 25%
- Power: $1 - \beta = 80\%$ (β is the Type II error probability)
- Significance level: α (α is the Type I error probability)

```
## Compute the optimal sample size for a 1:1 design
power.prop.test(power=0.8, p1=0.6, p2=0.25, sig.level=0.05)
```

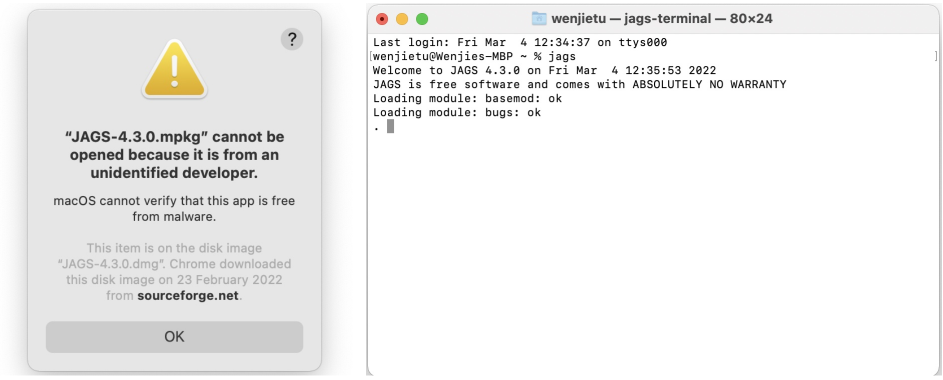
```
##
##      Two-sample comparison of proportions power calculation
##
##               n = 30.10887
##              p1 = 0.6
##              p2 = 0.25
##       sig.level = 0.05
##           power = 0.8
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

The above output can be read off that the optimal sample size for a 1:1 design for comparision of 60% (Secukinumab) and 25% (Placebo) with power 80% and significance level 5% is 60 (i.e. a sample size of 30 for each group).

1.2 Elicitation of your personal opinion

Without any further research from the Internet, my personal opinion about the height of adult Swiss females is that it follows a normal distribution with a mean of 170 cm and a standard deviation of 5 cm.

This guess has already been adjusted from the way of Bayesian thinking. If I had never been to Switzerland, my prior guess would have been based on the height of adult Chinese females. However, my prior gets updated and becomes a posterior after I come to Switzerland. From the course, we also learned that the posterior can serve as a prior again and it can be updated sequentially.



“JAGS-4.3.0.mpkg” cannot be opened because it is from an unidentified developer.

macOS cannot verify that this app is free from malware.

This item is on the disk image “JAGS-4.3.0.dmg”. Chrome downloaded this disk image on 23 February 2022 from **sourceforge.net**.

```
Last login: Fri Mar  4 12:34:37 on ttys000
wenjietu@Wenjies-MBP ~ % jags
Welcome to JAGS 4.3.0 on Fri Mar  4 12:35:53 2022
JAGS is free software and comes with ABSOLUTELY NO WARRANTY
Loading module: basemod: ok
Loading module: bugs: ok
.
```

Figure 1: Screenshots

1.3 Installation of programs

I have installed `R` and `RStudio` so there is no issue in this regard. Concerning the installation of JAGS, I got the error that *“JAGS-4.3.0.mpkg” cannot be opened because it is from an unidentified developer* (because I am a Mac user). I followed the Solution 4 given by this article and then the issue was resolved. With all the programs/softwares successfully installed, I simply ran the command `install.packages(c("rjags", "coda", "bayesmeta"))` in `R` console and all packages specified were successfully downloaded with a message that `also installing the dependency packages`.

```
## Install packages
install.packages(c("rjags", "coda", "bayesmeta"))
```

```
## Check whether the packages are successfully installed
library(rjags)
```

```
## Loading required package: coda

## Linked to JAGS 4.3.0

## Loaded modules: basemod,bugs
```

```
library(coda)
library(bayesmeta)
```

```
## Loading required package: forestplot

## Loading required package: grid

## Loading required package: magrittr

## Loading required package: checkmate

## Loading required package: metafor

## Loading required package: Matrix

##
## Loading the 'metafor' package (version 3.0-2). For an
## introduction to the package please type: help(metafor)
## Loading required package: mvtnorm

## Loading required package: numDeriv

##
## Attaching package: 'bayesmeta'

## The following object is masked from 'package:coda':
##
##     traceplot

## The following object is masked from 'package:stats':
##
##     convolve
```

2 Classical and Bayesian analyses and moment matching

In this chapter, we first provide classical and Bayesian analyses for Secukinumab and placebo separately. We next implement a moment-matching function that computes α and β shape parameters of a Beta distribution with known mean and variance.

2.1 Classical and Bayesian analyses

Table 3: ASAS20 responders at week 6

Group	n	Responders x (%)	Posterior response rate
Secukinumab	23	14 (60.9%)	59.5%
Placebo	6	1 (16.7%)	24.5%

1(a) Compute 95% CIs for the true probability of response

```
library(DescTools)

## Secukinumab group
BinomCI(x=14, n=23, conf.level=0.95, method="wilson")
```

```
##            est    lwr.ci    upr.ci
## [1,] 0.6086957 0.4078552 0.7784238
```

```
## Placebo group
BinomCI(x=1, n=6, conf.level=0.95, method="wilson")
```

```
##            est     lwr.ci    upr.ci
## [1,] 0.1666667 0.03005337 0.5635028
```

Interpretation of CIs:

- The 95% confidence interval for *Secukinumab* group is $(0.4079, 0.7784)$ and it means that for repeated samples from a binomial distribution with unknown parameter p, the 95% confidence interval will cover the true probability p in 95% of the cases.

- The 95% confidence interval for *Placebo* group is $(0.0301, 0.5635)$ and it means that for repeated samples from a binomial distribution with unknown parameter p, the 95% confidence interval will cover the true probability p in 95% of the cases.

Held and Sabanes Bove (2020, p116-119) demonstrate that the Wilson approach is superior to other approaches. Hence, Wilson method is used to compute the 95% CIs for the true probability of response in Secukinumab and in the placebo groups as it does not suffer from problems of overshoot and zero-width intervals that afflict the normal interval and it can also be safely used with small samples and skewed observations. To show the difference among CIs computed from other methods, the proportion test and the exact binomial test are conducted for comparison.

```
## Comparison

## Proportion test
prop.test(x=14, n=23, conf.level=0.95, correct=TRUE)
```

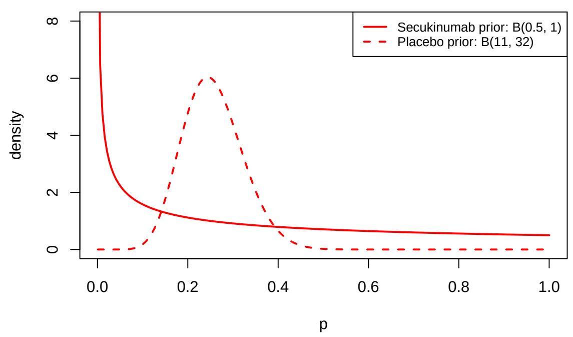
```
##
##  1-sample proportions test with continuity correction
##
## data:  14 out of 23, null probability 0.5
## X-squared = 0.69565, df = 1, p-value = 0.4042
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.3878251 0.7953232
## sample estimates:
##         p
## 0.6086957
```

```
## Exact binomial test
binom.test(x=14, n=23)
```

```
##
##  Exact binomial test
##
## data:  14 and 23
## number of successes = 14, number of trials = 23, p-value = 0.4049
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.3854190 0.8029236
## sample estimates:
## probability of success
##              0.6086957
```

1(b) Plot and summarize priors

```
## Define range
p <- seq(1e-3,1, length=200)

## Plot several Beta distributions
plot(p, dbeta(p, 0.5, 1), ylab="density", type="l", col="red", lwd=2,
     ylim=c(0, 8), main="Plot of priors for both Secukinumab and placebo")
lines(p, dbeta(p, 11, 32), col="red", lwd=2, lty=2)

## Add legend
legend("topright", col="red", lwd=2, lty=c(1, 2), cex=.8,
       legend=c("Secukinumab prior: B(0.5, 1)", "Placebo prior: B(11, 32)"))
```



Plot of priors for both Secukinumab and placebo

```r
## 2.5%, 50%, 97.5% quantiles of Beta(0.5, 1)
qbeta(c(0.025, 0.5, 0.975), 0.5, 1)
```

```
## [1] 0.000625 0.250000 0.950625
```

```r
## 2.5%, 50%, 97.5% quantiles of Beta(11, 32)
qbeta(c(0.025, 0.5, 0.975), 11, 32)
```

```
## [1] 0.1386101 0.2520003 0.3945024
```

We know that if $X \sim \text{Beta}(\alpha, \beta)$, then the expectation is given by:

$$\mathbb{E}X = \frac{\alpha}{\alpha + \beta}$$

We can apply this formula to obtain the prior means:

- Prior mean for $\text{Beta}(0.5, 1)$: 0.3334
- Prior mean for $\text{Beta}(11, 32)$: 0.2558

Table 4: Summary statistics of the prior distributions

		Mean	Median	Equi-tailed 95% interval
Group	Prior			
Secukinumab	$B(0.5, 1)$	0.3334	0.2500	$(0.0006, 0.9506)$
Placebo	$B(11, 32)$	0.2558	0.2520	$(0.1386, 0.3945)$

1(c) & 1(d) Plot and summarize posteriors

Prior distribution:

$$p \sim \text{Beta}(\alpha, \beta)$$

Posterior distribution:

$$p \mid y_1, \cdots, y_n \sim \text{Beta}(\alpha + n\bar{y}, \beta + n - n\bar{y})$$

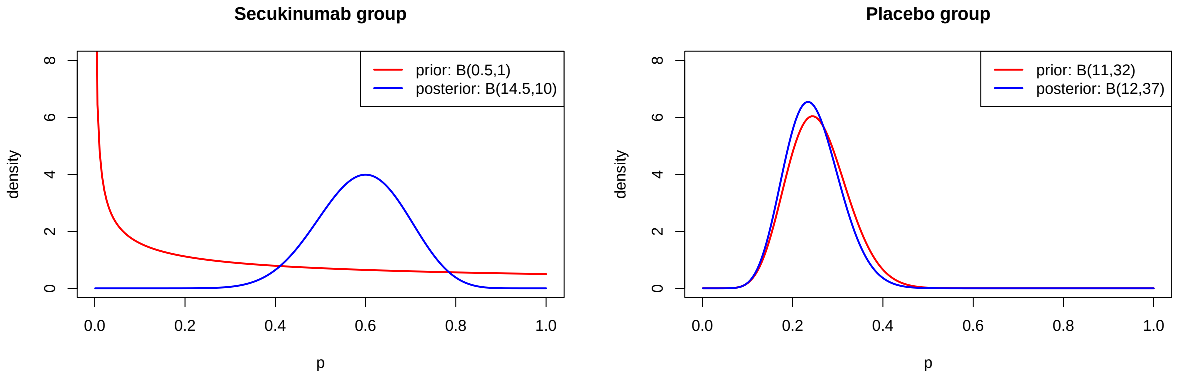
Since y_i is a binary variable and only takes 0 or 1, $\bar{y}$ corresponds to the proportion of the responders (i.e. response rate).

- $n\bar{y}$ corresponds to the number of responders (i.e. $\sum_{i=1}^{n} \mathbf{1}[y_i = 1]$)
- $n - n\bar{y}$ corresponds to the number of non-responders (i.e. $n - \sum_{i=1}^{n} \mathbf{1}[y_i = 1]$)

Table 5: Prior vs. Posterior

	Prior	# of Subjects	# of Responders	Posterior
Group	Beta(α, β)	n	x	Beta$(\alpha + x, \beta + n - x)$
Secukinumab	Beta$(0.5, 1)$	23	14	Beta$(14.5, 10)$
Placebo	Beta$(11, 32)$	6	1	Beta$(12, 37)$

```r
p <- seq(1e-3,1, length=200)

plot(p, dbeta(p, 0.5, 1), type="l", ylab="density", col="red",
     lwd=2, ylim=c(0,8), main="Secukinumab group")
lines(p, dbeta(p, 14.5,10), ylab="density", col="blue", lwd=2)

legend("topright", legend=c("prior: B(0.5,1)", "posterior: B(14.5,10)"),
       col=c("red", "blue"), lwd=2)

plot(p, dbeta(p, 11, 32), type="l", ylab="density", col="red",
     lwd=2, ylim=c(0, 8), main="Placebo group")
lines(p, dbeta(p, 12, 37), ylab="density", col="blue", lwd=2)

legend("topright", legend=c("prior: B(11,32)", "posterior: B(12,37)"),
       col=c("red", "blue"), lwd=2)
```



```r
# Secukinumab group
## 2.5%, 50%, 97.5% quantiles of Beta(14.5, 10)
qbeta(c(0.025, 0.5, 0.975), 14.5, 10)
```

```
## [1] 0.3958401 0.5943750 0.7736253
```

```r
# Placebo group
## 2.5%, 50%, 97.5% quantiles of Beta(12, 37)
qbeta(c(0.025, 0.5, 0.975), 12, 37)
```

```
## [1] 0.1363723 0.2414054 0.3731202
```

(e) Interpret 95% CrIs

Interpretation:

- In Secukinumab group, the posterior probability of the response rate lies between 0.3958 and 0.7736 with probability 95%, when a Beta$(0.5, 1)$ prior is assumed.

Table 6: Summary statistics of the posterior distributions

Group	Posterior	Mean	Median	95% CrI
Secukinumab	$B(14.5, 10)$	0.5918	0.5944	$(0.3958, 0.7736)$
Placebo	$B(12, 37)$	0.2449	0.2414	$(0.1364, 0.3731)$

- In Placebo group, the posterior probability of the response rate lies between 0.1364 and 0.3731 with probability 95%, when a Beta$(11, 32)$ prior is assumed.

2.2 Priori elicitation for Secukinumab group using moment matching

2(a) Derive moment matching formula for Beta distribution

Beta distribution:

$$X \sim \text{Beta}(\alpha, \beta)$$

Density function of Beta distribution:

$$f(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

Expectation of X:

$$\begin{aligned}
\mathbb{E}X &= \int_0^1 x f(x)\mathrm{d}x \\
&= \int_0^1 x \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}\mathrm{d}x \\
&= \frac{1}{B(\alpha, \beta)} \int_0^1 x^{\alpha}(1-x)^{\beta-1}\mathrm{d}x \\
&= \frac{B(\alpha+1, \beta)}{B(\alpha, \beta)} \\
&= \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \\
&= \frac{\alpha!(\beta-1)!}{(\alpha+\beta)!} \cdot \frac{(\alpha+\beta-1)!}{(\alpha-1)!(\beta-1)!} \\
&= \frac{\alpha}{\alpha+\beta}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[X^2] &= \int_0^1 x^2 f(x)\mathrm{d}x \\
&= \int_0^1 x^2 \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}\mathrm{d}x \\
&= \frac{1}{B(\alpha, \beta)} \int_0^1 x^{\alpha+1}(1-x)^{\beta-1}\mathrm{d}x \\
&= \frac{B(\alpha+2, \beta)}{B(\alpha, \beta)} \\
&= \frac{\Gamma(\alpha+2)\Gamma(\beta)}{\Gamma(\alpha+\beta+2)} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \\
&= \frac{(\alpha+1)!(\beta-1)!}{(\alpha+\beta+1)!} \cdot \frac{(\alpha+\beta-1)!}{(\alpha-1)!(\beta-1)!} \\
&= \frac{(\alpha+1)\alpha}{(\alpha+\beta+1)(\alpha+\beta)}
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}X &= \mathbb{E}\left[[X-\mathbb{E}X]^2\right] \\
&= \mathbb{E}[X^2] - [\mathbb{E}X]^2 \\
&= \frac{(\alpha+1)\alpha}{(\alpha+\beta+1)(\alpha+\beta)} - \left(\frac{\alpha}{\alpha+\beta}\right)^2 \\
&= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}
\end{aligned}$$

We know that:

$$\begin{cases} \mu = \frac{\alpha}{\alpha+\beta} \\ \sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \end{cases}$$

$$\mu = \frac{\alpha}{\alpha+\beta} \implies \beta = \frac{\alpha}{\mu} - \alpha$$

Plug $\beta = \frac{\alpha}{\mu} - \alpha$ into $\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$:

$$\begin{aligned}
\sigma^2 &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \\
&= \frac{\alpha\left(\frac{\alpha}{\mu}-\alpha\right)}{\left(\frac{\alpha}{\mu}\right)^2\left(\frac{\alpha}{\mu}+1\right)} \\
&= \frac{\mu^2(1-\mu)}{\alpha+\mu}
\end{aligned}$$

Rearranging above equation yields:

$$\begin{cases} \alpha = \left(\frac{1-\mu}{\sigma^2} - \frac{1}{\mu}\right)\mu^2 \\ \beta = \alpha\left(\frac{1}{\mu} - 1\right) \end{cases}$$

2(b) Implement moment matching formula in a function in R

Implementation:

- Input: sample mean and sample variance of a Beta distribution
- Output: α and β shape parameters of the Beta distribution

```
## Define a function
## Args: mean, var
estBetaParams <- function(mean, var) {
  alpha <- ((1 - mean) / var - 1 /mean) * mean ^ 2
  beta <- alpha * (1 / mean - 1)
  return(params = c(alpha=alpha, beta=beta))
}
```

2(c) Apply the function

Input:

- mean = 0.255814
- variance = 0.004326663

```
params <- estBetaParams(0.255814, 0.004326663); params
```

```
## alpha  beta
##    11    32
```

The resulting values of α and β parameters of the Beta distribution are 11 and 32 respectively.

3 Posterior probability of superiority

In this chapter, we conduct Monte Carlo simulations to estimate the posterior probability of superiority (PPS) and the Monte Carlo standard error based on the response rate difference (RRD). We next extend the response rate difference to response ratio (RR) and odds ratio (OR) to estimate PPS and corresponding Monte Carlo standard error.

3.1 Posterior probability of superiority based on difference

Table 7: ASAS20 responders at week 6

Group	n	Responders	Posterior	Difference	95% CrI	PPS	MCse (PPS)
		x (%)	Response rate	(S-P)			
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 7 and its Monte Carlo standard error.

1 Draw MC samples from posteriors for Secukinumab and placebo

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

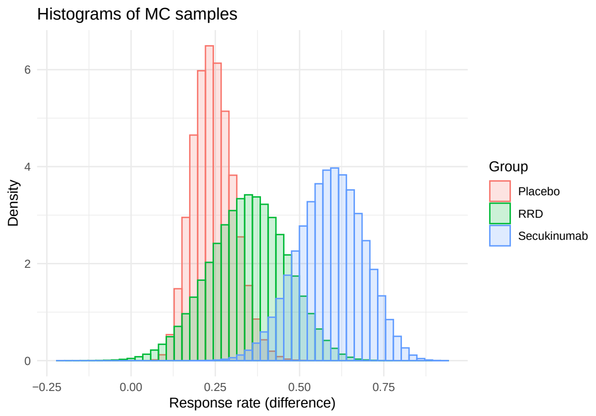
```
## Set the seed for reproducible results
set.seed(44566)

## 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)
```

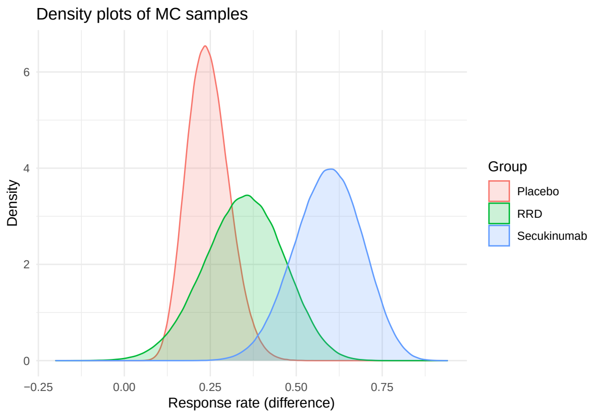
2 Compute and plot the response rate difference (RRD)

```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()
```





3 Compute and interpret the median and the 95% CrI of RDD

```r
## Compute 2.5%, 50%, 97.5% quantiles
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 Beta$(14.5, 10)$ in Secukinumab group and a MC sample (size=1000000) from the posterior distribution 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)

4 Compute the posterior probability of superiority

PPS under RRD measure is defined as:

$$P(\theta_{\mathrm{S}} - \theta_{\mathrm{P}} > 0) \equiv P(RRD > 0)$$

```r
## 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%.

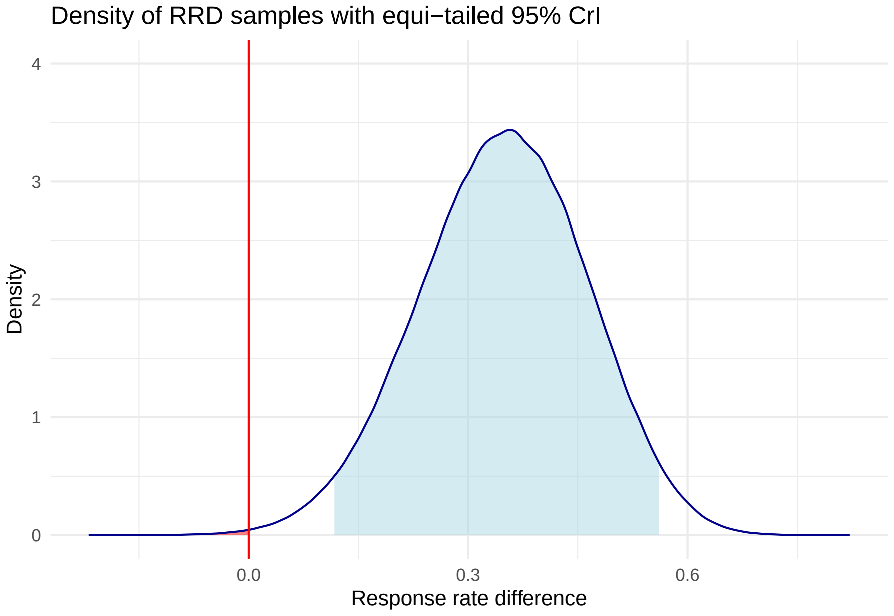
5 Compute the MC standard error

MC standard error of PPS:

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

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

```r
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
```

3.2 Extension of posterior probability of superiority to ratio and odds ratio

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

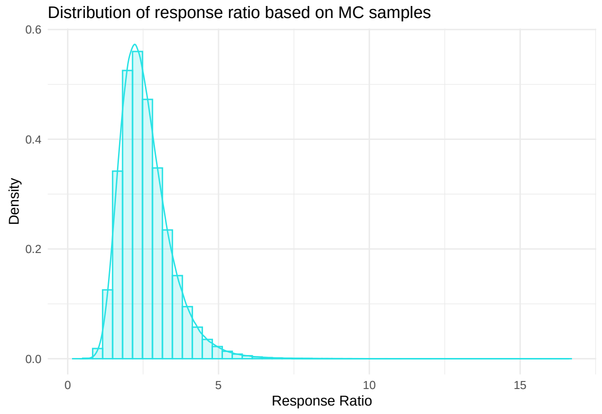
Response ratio (RR)

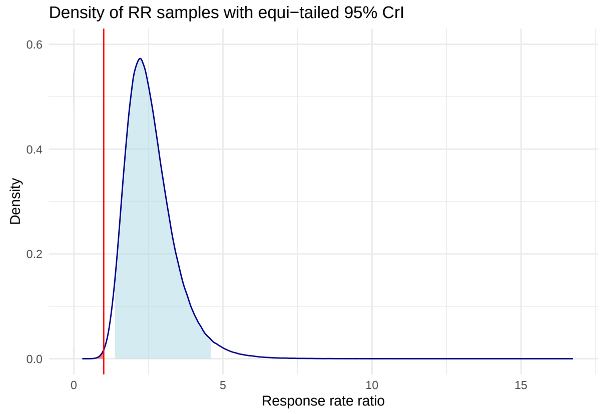
- Construct the response ratio based on the MC samples
- Compute the median and the equi-tailed 95% credible interval
- Visualize the distribution of the response ratio
- Interpret the results

```
## 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:

$$P\left(\frac{\theta_{\mathrm{S}}}{\theta_{\mathrm{P}}} > 1\right) \equiv P(RR > 0)$$

```
## 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
```

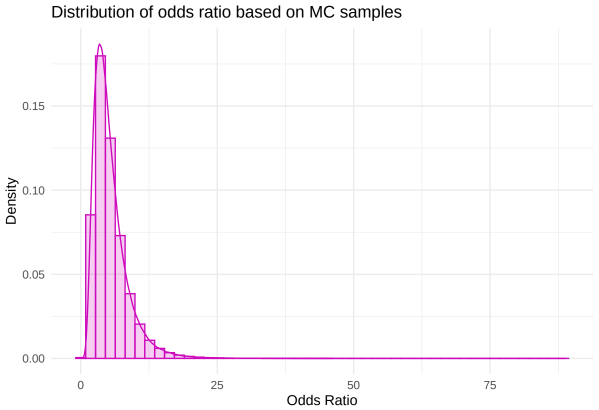
Odds ratio (OR)

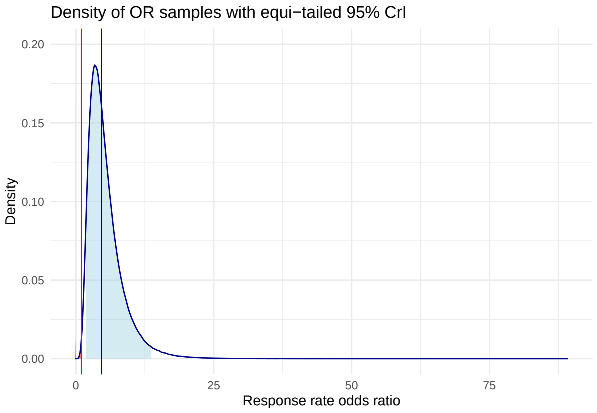
- Construct the odds ratio based on the MC samples
- Compute the median and the equi-tailed 95% credible interval
- Visualize the distribution of the odds ratio
- Interpret the results

```
## 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()
```





PPS under OR measure is defined as:

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

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

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

```
## 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):

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

```
## 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 8: 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. $P(\theta_\text{S} > \theta_\text{P})$).

4 Study design and sample size computation

In this chapter, we provide an interesting approach to plan a study. At this stage, there is no data but only hypotheses about the relation of the treatment an placebo groups. We introduce the proof of concept (POC) that requires the ASAS20 response rate on Secukinumab larger than that on placebo. With data from 20 patients on Secukinumab and 5 patients on placebo, Monte Carlo simulations are conducted to show that POC $>$ 90%, for true response rates of 25% on placebo and 60% on secukinumab.

4.1 Study design and proof of concept

1 Provide evidence for POC $> 90\%$ and compute its MC standard error

```
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 9: Bayesian study design 4:1

	Sample Size	Response Rate
Secukinumab	20	0.60
Placebo	5	0.25

In the Secukinumab group:

- $P_\text{S} = 0.6$, $n_\text{S} = 20$
- $x_S^{(i)} \sim \text{Bin}(n_\text{S}, p_\text{S})$

- $\hat{p}_{\mathrm{S}}^{(i)} = \frac{x_S^{(i)}}{n_{\mathrm{S}}}$

In the Placebo group:

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

Construct the response rate difference between secukinumab and placebo):

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

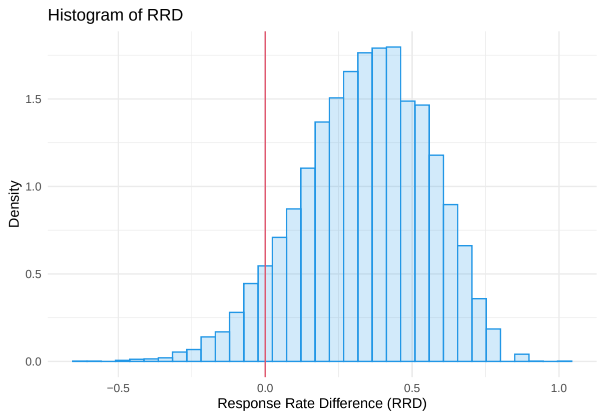
We are interested in the sample distribution of $d^{(i)}$ and would like to investigate the proof of concept (i.e. the probability that ASAS20 response rate on secukinumab is larger than that on placebo):

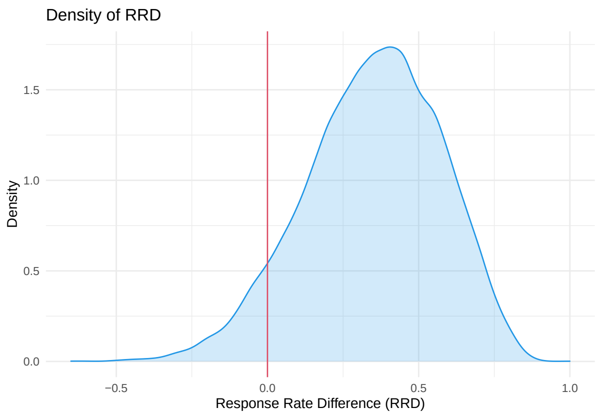
$$\mathrm{POC} = P[D > 0]$$

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()
```





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

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

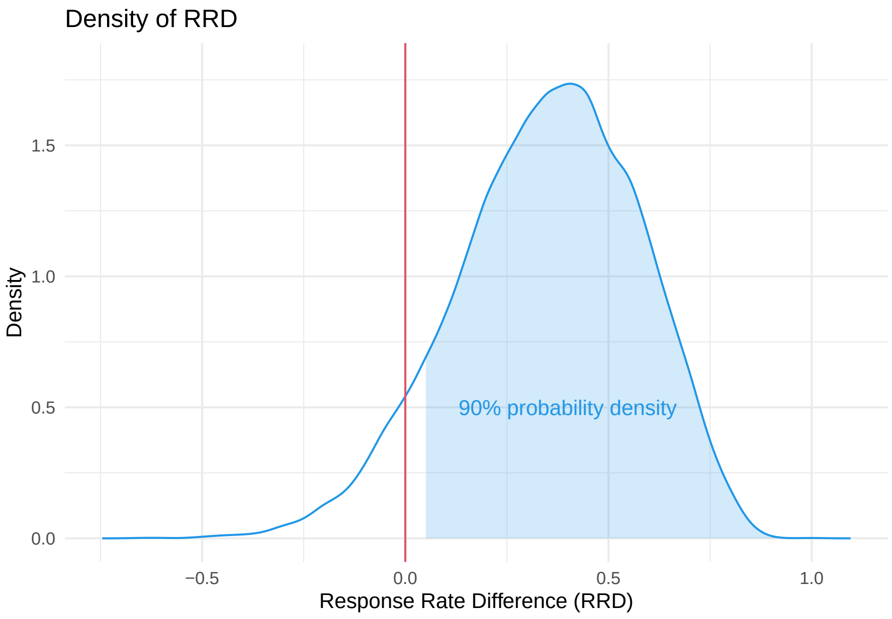
```
## MC standard error
poc.se <- sqrt(var(RRD > 0)/M); poc.se
```

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

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

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

```
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()
```



2 Compute the minimum number of patients required for the 4:1 design

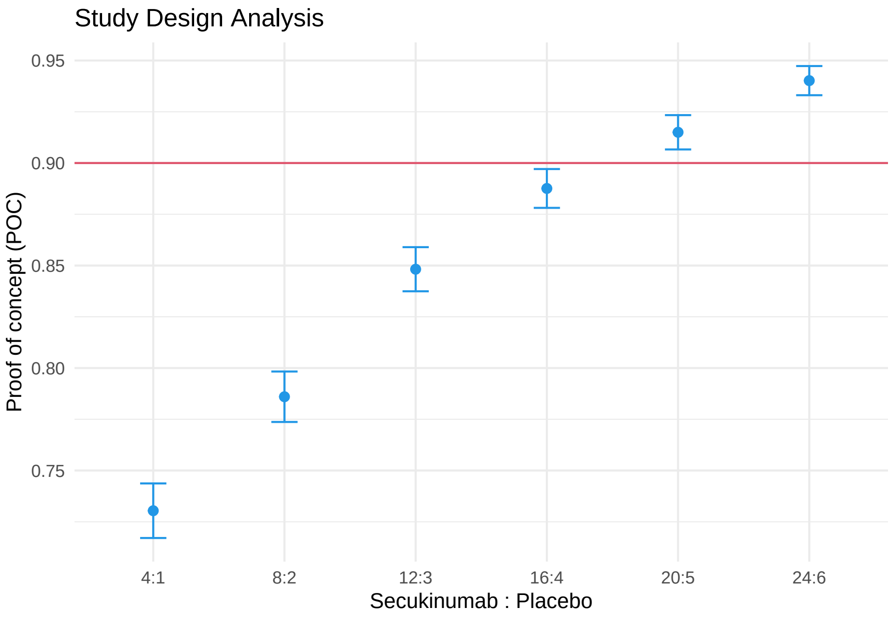
```
## Define a function for design analysis
## Params: M, size1, size2, prob1, prob2

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 10: 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

```r
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, size=2) +
  # geom_line(color=4, linetype="longdash") +
  # geom_ribbon(aes(ymin=lower99, ymax=upper99), fill=4, alpha=0.3) +
  geom_errorbar(aes(ymin=lower99, ymax=upper99), color=4, width=.2) +
  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)

3 Explain this approach to a client

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 $= 91.46\% > 90\%$.

5 Priori elicitation for placebo and treatment groups

In this chapter, we attempt to replicate the results from Baeten et al. (2013) according to the descriptive information in the paper. We first perform a random-effects meta-analysis of the 8 historical trials for placebo using JAGS to derive the posterior response rate for placebo. We next try to elicit the prior distribution for treatment according to the descriptive information in Baeten et al. (2013)[(Appendix)].

5.1 Elicitation of the prior distribution for placebo group

1. Apply the logit-transformation

$$X \sim \mathrm{Bin}(n, \pi)$$

$$\hat{\phi}_{\mathrm{ML}} = \mathrm{logit}(\hat{\pi}_{\mathrm{ML}}) = \log\left(\frac{\hat{\pi}_{\mathrm{ML}}}{1-\hat{\pi}_{\mathrm{ML}}}\right) = \log\left(\frac{x}{n-x}\right)$$

$$\mathrm{se}(\hat{\phi}_{\mathrm{ML}}) = \sqrt{\frac{1}{x} + \frac{1}{n-x}}$$

```r
pl_total <- c(107, 44, 51, 39, 139, 20, 78, 35)
pl_case <- c(23, 12, 19, 9, 39, 6, 9, 10)

## Logit-transformation
pl_y <- log(pl_case / (pl_total - pl_case))

## Standard error
pl_se <- sqrt(1/pl_case + 1/(pl_total - pl_case))

prec_s <- 1 / pl_se^2
```

2. Implement the code for a random effects meta-analysis in JAGS

```r
pl1_modelString <- "model{
  for (i in 1:length(y)) {
    y[i] ~ dnorm(theta[i], prec_s[i]);
    theta[i] ~ dnorm(mu, prec_tau);
  }

  theta_new ~ dnorm(mu, prec_tau); # predictive distribution for theta

  # predictive distribution at the probability scale
  p_new <- exp(theta_new) / (1 + exp(theta_new));

  mu ~ dnorm(0.0, 1.0E-4);
  prec_tau ~ dgamma(1.0E-3, 1.0E-3); # just our assumption
}"

writeLines(pl1_modelString, con="MetaAnalysis.txt")
```

Parametrization

$$\begin{aligned} y_i &\sim \mathrm{N}(\theta_i, 1/\tau_i^s) \\ \theta_i &\sim \mathrm{N}(\mu, 1/\tau) \end{aligned} \quad \text{for } i = 1, \cdots, N$$

$$\left.\begin{aligned}\mu &\sim \mathrm{N}(0, 100^2)\\ \tau &\sim \mathrm{G}(0.001, 0.001)\end{aligned}\right. \quad \text{priors for parameters}$$

Data required for JAGS:

$$y_i = \log\left(\frac{x_i}{n_i - x_i}\right)$$
$$\tau_i^s = 1\bigg/\left(\frac{1}{x_i} + \frac{1}{n_i - x_i}\right)$$

```
library(rjags)
library(coda)
library(ggplot2)
```

```
## Generate a data list for JAGS
dat.jags <- list(y=pl_y, prec_s=prec_s)

## Generate a list for initial values for JAGS
## Set initial values and seed for reproducible results
inits.jags <- list(
  list(mu=100, prec_tau=0.01, .RNG.name="base::Wichmann-Hill", .RNG.seed=314159),
  list(mu=10, prec_tau=0.05, .RNG.name="base::Marsaglia-Multicarry", .RNG.seed=159314),
  list(mu=-100, prec_tau=0.001, .RNG.name="base::Super-Duper", .RNG.seed=413159),
  list(mu=-10, prec_tau=0.1, .RNG.name="base::Mersenne-Twister", .RNG.seed=143915)
)

## Create a JAGS model object
model.jags <- jags.model(
  file = "MetaAnalysis.txt", # text file for JAGS
  data = dat.jags,           # the data list for JAGS
  inits = inits.jags,        # the list for initialization
  n.chains = 4,              # the number of parallel chains for the model
  n.adapt = 4000             # the number of iterations for adaption
)
```

```
## Compiling model graph
##    Resolving undeclared variables
##    Allocating nodes
## Graph information:
##    Observed stochastic nodes: 8
##    Unobserved stochastic nodes: 11
##    Total graph size: 34
##
## Initializing model
```

```
## Use the first 4000 iterations as burn-in iterations
update(model.jags, n.iter = 4000)

## Generate posterior samples
fit.jags.coda <- coda.samples(
  model = model.jags,            # JAGS model object
  variable.names = c("p_new"),   # variable to be monitored
  n.iter = 10000,                # number of iterations to monitor
  thin = 1                       # thinning interval for monitors
)
```

3. Explain what is being done in each line of the code

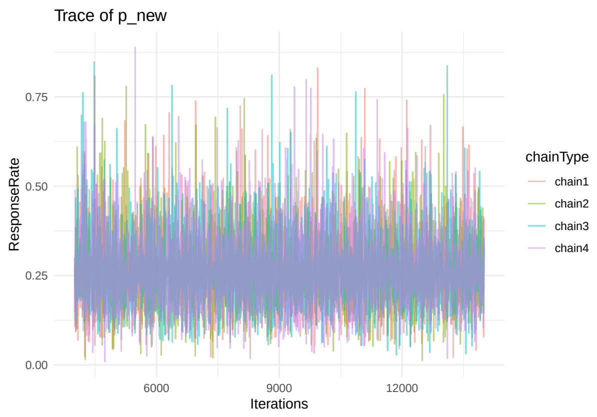
See the comments in the code chunks.

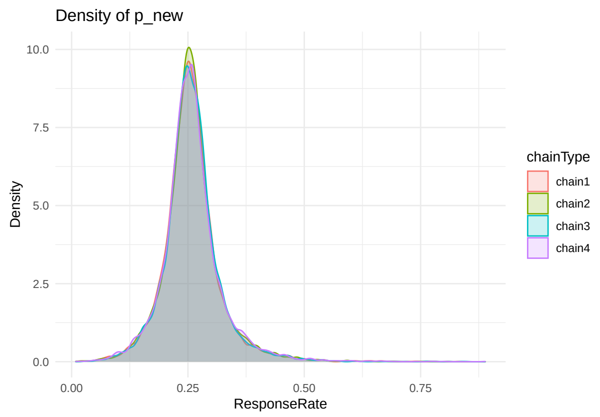
4. Summarize the posterior predictive distribution for p_new

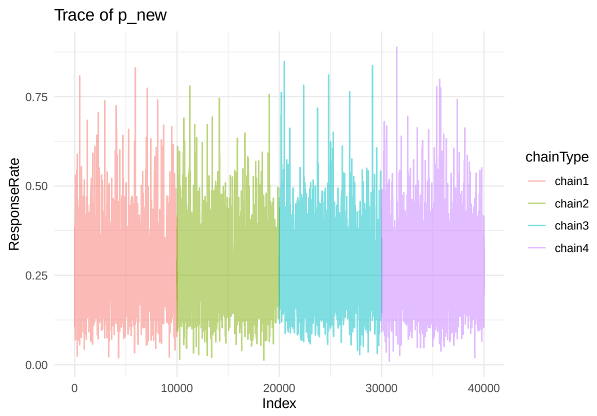
```
## Print the summary results of the posterior samples generated
summary(fit.jags.coda)
```

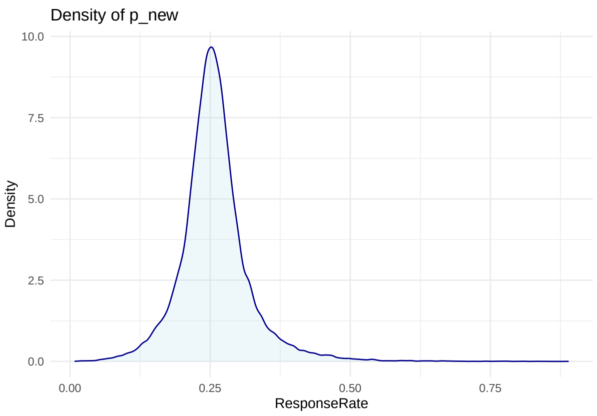
```
##
## Iterations = 4001:14000
## Thinning interval = 1
## Number of chains = 4
## Sample size per chain = 10000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##           Mean             SD       Naive SE Time-series SE
##      0.2585754      0.0633004      0.0003165      0.0004144
##
## 2. Quantiles for each variable:
##
##   2.5%    25%    50%    75%  97.5%
## 0.1416 0.2259 0.2543 0.2843 0.4069
```

```
## Concatenate 4 chains
m.fit.jags.coda <- as.matrix(fit.jags.coda)

d.mcmc <- data.frame(
  Iterations = rep(4001:14000, times=4),
  ResponseRate = m.fit.jags.coda[, "p_new"],
  chainType = rep(c("chain1", "chain2", "chain3", "chain4"), each=10000)
)

ggplot(d.mcmc, aes(x=Iterations, y=ResponseRate, color=chainType)) +
  geom_line(alpha=0.5) + labs(title="Trace of p_new") + theme_minimal()

ggplot(d.mcmc, aes(x=ResponseRate, y=..density.., fill=chainType, color=chainType)) +
  geom_density(alpha=0.2) + labs(title="Density of p_new", y="Density") + theme_minimal()

ggplot(d.mcmc, aes(x=1:nrow(d.mcmc), y=ResponseRate, color=chainType)) +
  geom_line(alpha=0.5) + labs(title="Trace of p_new", x="Index") + theme_minimal()

ggplot(d.mcmc, aes(x=ResponseRate, y=..density..)) +
  geom_density(alpha=0.2, color="darkblue", fill="lightblue") +
  labs(title="Density of p_new", y="Density") + theme_minimal()
```









```r
d.summary <- data.frame(t(rbind(
  colMeans(m.fit.jags.coda),
  apply(m.fit.jags.coda, 2, function(x) sd(x)),
  apply(m.fit.jags.coda, 2, function(x) var(x)),
  apply(m.fit.jags.coda, 2, function(x) quantile(x, probs=c(0.025, 0.5, 0.975)))
)))
colnames(d.summary) <- c("Mean", "SD", "Variance", "2.5%", "Median", "97.5%")
knitr::kable(d.summary, digits=5, align="c", caption="Summary statistics")
```

Table 11: Summary statistics

	Mean	SD	Variance	2.5%	Median	97.5%
p_new	0.25858	0.0633	0.00401	0.14158	0.25427	0.4069

5. Compare with the results obtained in Baeten’s study

```r
d.comparison <- data.frame(
  c(d.summary$Mean, 0.25),
  c(d.summary$`2.5%`, 0.13),
  c(d.summary$`97.5%`, 0.4)
)
colnames(d.comparison) <- c("Mean", "2.5%", "97.5%")
rownames(d.comparison) <- c("Meta-analysis", "Baeten's study")
knitr::kable(d.comparison, digits=3, align="c", caption="Comparision")
```

Table 12: Comparision

	Mean	2.5%	97.5%
Meta-analysis	0.259	0.142	0.407
Baeten’s study	0.250	0.130	0.400

The estimate from the meta-analysis and its 95% credible interval agree with those reported in Baeten et al. (2013)

6. Apply the function for moment matching

In Exercise 2 of Worksheet 2, we derived

$$\begin{cases} \alpha = \left(\frac{1-\mu}{\sigma^2} - \frac{1}{\mu}\right)\mu^2 \\ \beta = \alpha\left(\frac{1}{\mu} - 1\right) \end{cases}$$

```
## Moment matching function
estBetaParams <- function(mean, var) {
  alpha <- ((1 - mean) / var - 1 /mean) * mean ^ 2
  beta <- alpha * (1 / mean - 1)
  return(params = c(alpha=alpha, beta=beta))
}
```

7. Report the estimates

```
estBetaParams(mean=d.summary$Mean, var=d.summary$Variance)
```

```
##    alpha     beta
## 12.11310 34.73241
```

8. Explain the approach to a client

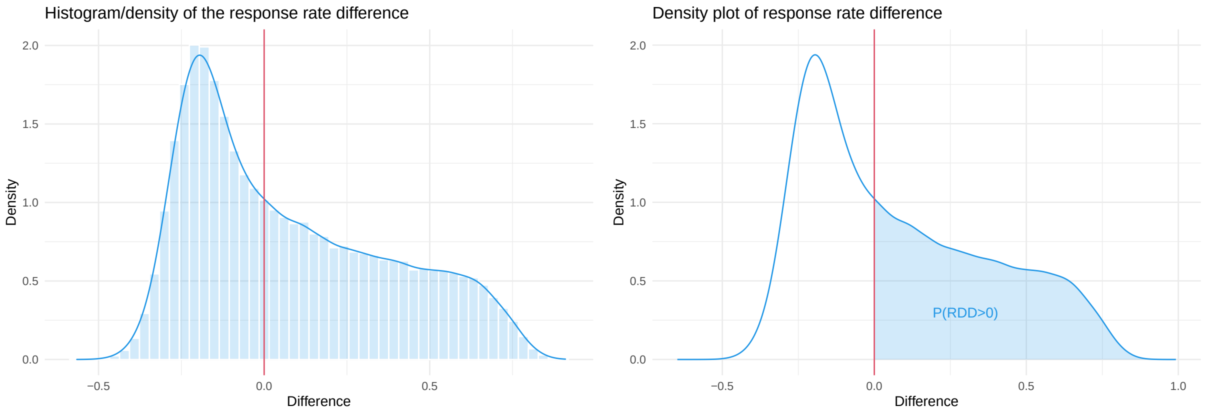
In this exercise, we attempt to replicate the results from Baeten's study and see whether the similar results can be obtained.

- We first gather the data from 8 historical studies
- We next apply logit-transformation to response rate for placebo group to obtain an approximately normal distribution of logit-transformed rates
- We then conduct a random-effects meta-analysis to derive the posterior predictive distribution of response rate
- We next summarize the posterior predictive distribution of response rates (i.e., mean and 95% credible interval)
- We then apply the function for moment matching to derive the parameters α and β of the Beta prior for the response rate in the placebo group.

5.2 Elicitation of the prior distribution for treatment group

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

## Monte Carlo sample size
M <- 100000

## Generate i.i.d. random samples
p.p <- rbeta(M, shape1=11, shape2=32)  # response rate in Placebo
p.t <- rbeta(M, shape1=0.5, shape2=1)  # response rate in Treatment

## Construct response rate difference (RRD)
rrd <- p.t - p.p
```

```
ggplot(data.frame(Difference=rrd), aes(x=Difference, y=..density..)) +
  geom_histogram(position="identity", bins=50, alpha=0.2, color="white", fill=4) +
  geom_density(alpha=0.2, color=4) + geom_vline(xintercept=0, color=2) + theme_minimal() +
  labs(title="Histogram/density of the response rate difference", y="Density")

ggplot(with(density(rrd), data.frame(x, y)), aes(x=x, y=y)) +
  geom_area(aes(x=ifelse(x>0, x, 0), y=y), fill=4, alpha=0.2) +
  geom_line(color=4) + geom_vline(xintercept=0, color=2) +
  labs(title="Density plot of response rate difference", x="Difference", y="Density") +
  geom_text(aes(x=0.3, y=0.3), label="P(RDD>0)", color=4, check_overlap=TRUE) +
  ylim(0, 2) + theme_minimal()
```



```
## Posterior probability of superiority of treatment over placebo
pps <- mean(rrd>0); pps
pps.se <- sqrt(var(rrd>0)/M); pps.se

## 99% confidence interval
c(pps-3*pps.se, pps+3*pps.se)
```

```
## [1] 0.50046
## [1] 0.001581146
## [1] 0.4957166 0.5052034
```

Beta$(0.5, 1)$ is used for treatment group and it has been shown that with such priors there is an approximately 50:50 chance that the response rate on active treatment would be greater than the response rate on placebo.

Approach:

- Generate a vector of i.i.d. random samples from Beta$(11, 32)$ as the response rates for the placebo group
- Generate a vector of i.i.d. random samples from Beta$(0.5, 1)$ as the response rates for the treatment group
- Construct the response rate difference between the response rates in the treatment group and the response rates in the placebo group
- Calculate the posterior probability of superiority and the corresponding Monte Carlo standard error