

HW 5

Problem 1

(1)

Since we have a uniform prior the posterior is

$$\begin{aligned} p(\theta|y) &\propto I_{[10^{-5}, 10^5]}(\theta) \prod_{i=1}^n \exp\left(-\frac{(y_i - \theta)^2}{2}\right) \\ &\propto I_{[10^{-5}, 10^5]}(\theta) \exp\left(-\frac{n\theta^2 + 2\sum_{i=1}^n y_i \theta}{2}\right) \\ &\propto I_{[10^{-5}, 10^5]}(\theta) \exp\left(-\frac{(\theta - \bar{y})^2}{2n^{-1}}\right). \end{aligned}$$

It's $N(\bar{y}, n^{-1})$ truncated in $[10^{-5}, 10^5]$. So the posterior predictive distribution is

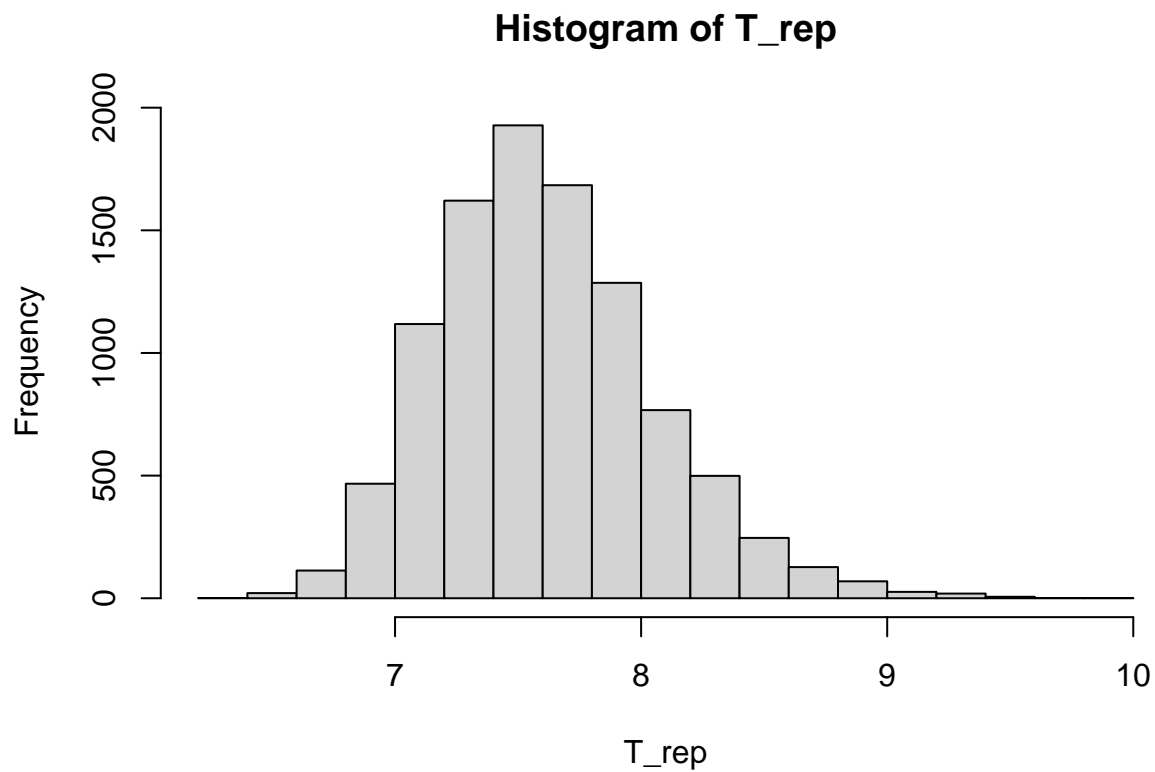
$$\begin{aligned} p(y^{rep}|y) &\propto \int p(y^{rep}|\theta)p(\theta|y) d\theta \\ &\propto \int \exp\left(-\frac{(y^{rep} - \theta)^2}{2}\right) \exp\left(-\frac{(\theta - \bar{y})^2}{2n^{-1}}\right) I_{[10^{-5}, 10^5]}(\theta) d\theta \end{aligned}$$

```
set.seed(0)

posterior <- rnorm(1e4, mean=5.1, sd=100^(-1/2))
posterior <-posterior[(posterior >= 1e-5 & posterior <= 1e5)]

T_rep <- c()
for (theta in posterior) {
  rep <- rnorm(100, mean=theta, sd=1)
  T_rep <- c(T_rep, max(rep))
}

hist(T_rep)
```



```
cat('p-value is', mean(T_rep >= 8.1))
```

```
## p-value is 0.1319
```

(2)

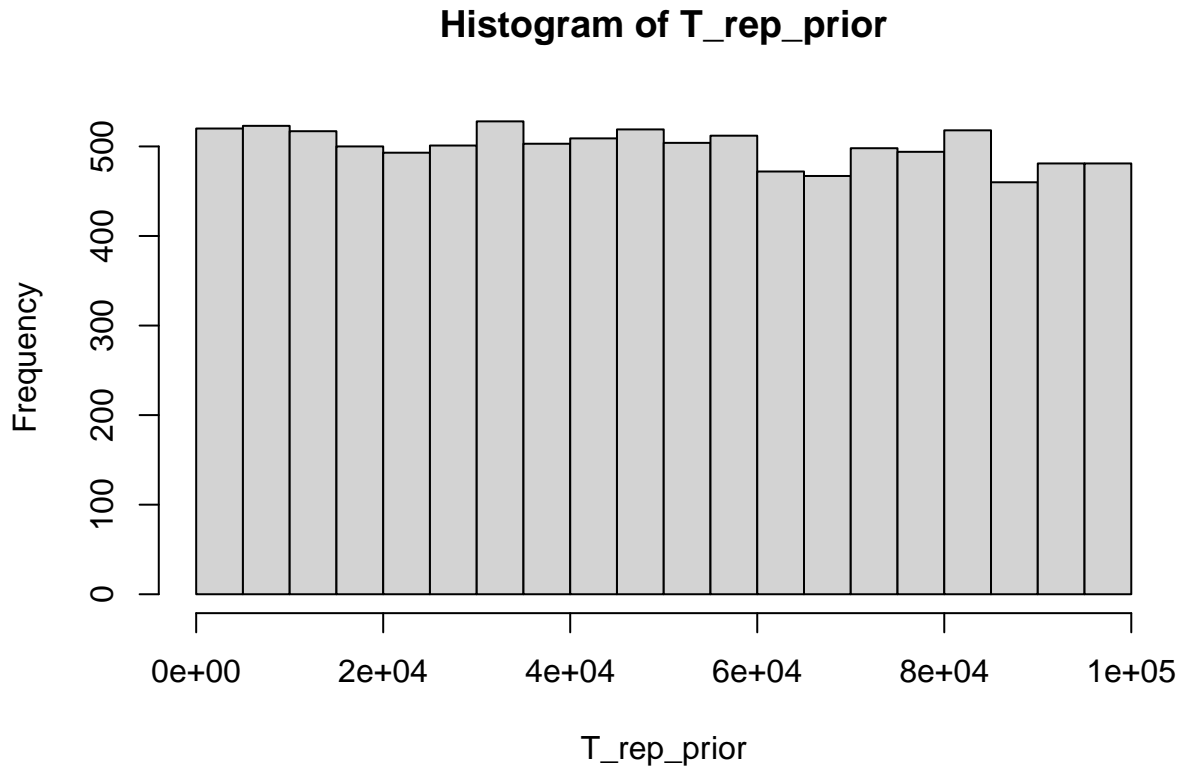
The prior predictive distribution is

$$\begin{aligned}
 p(y^{rep}) &\propto \int p(y^{rep}|\theta)p(\theta) d\theta \\
 &\propto \int \exp\left(-\frac{(y^{rep}-\theta)^2}{2}\right) I_{[10^{-5}, 10^5]}(\theta) d\theta \\
 &\propto P_{N(y^{rep}, 1)}(X \in [10^{-5}, 10^5]).
 \end{aligned}$$

```
prior <- runif(1e4, min=1e-5, max=1e5)

T_rep_prior <- c()
for (theta in prior) {
  rep <- rnorm(100, mean=theta, sd=1)
  T_rep_prior <- c(T_rep_prior, max(rep))
}

hist(T_rep_prior)
```



```
cat('p-value is', mean(T_rep_prior >= 8.1))
```

```
## p-value is 1
```

(3)

Obviously the prior does not contain any information of data y , and it's not very informative with pre-existing knowledge. So just using the prior we won't be able to replicate data y .

Problem 2

According to Chapter 5 the tumor rate θ has a Beta prior with parameters $\alpha = 1.4$ and $\beta = 8.6$ estimated using historical data. The posterior is Beta(5.4, 18.6).

Define test quantities as follows:

- The max tumor rate. $T_1(y) = \max(y/n)$.
- Measure the fluctuation. $T_2(y) = (n-1)^{-1} \sum (y_i/n_i - \overline{y_i/n_i})^2$.
- How many groups have 0 tumor rat. $T_3(y) = \sum I(y_i = 0)$.

```
n <- c(20,20,20,19,18,18,18,18,25,20,10,49,20,13,48,20,48,19,23,19,22,20,20,20,17,20,20,24,23,20,19,46,
y <- c(0, 0, 0, 0, 0, 0, 1, 1, 2, 2, 1, 5, 3, 2, 9, 4, 10, 4, 6, 5, 6, 0, 0, 0, 0, 1, 1, 2, 2, 2, 2, 5,
```

```
T1 <- max(y/n)
T2 <- var(y/n)
T3 <- sum(y/n == 0)
```

```
T1_rep <- c()
T2_rep <- c()
T3_rep <- c()
```

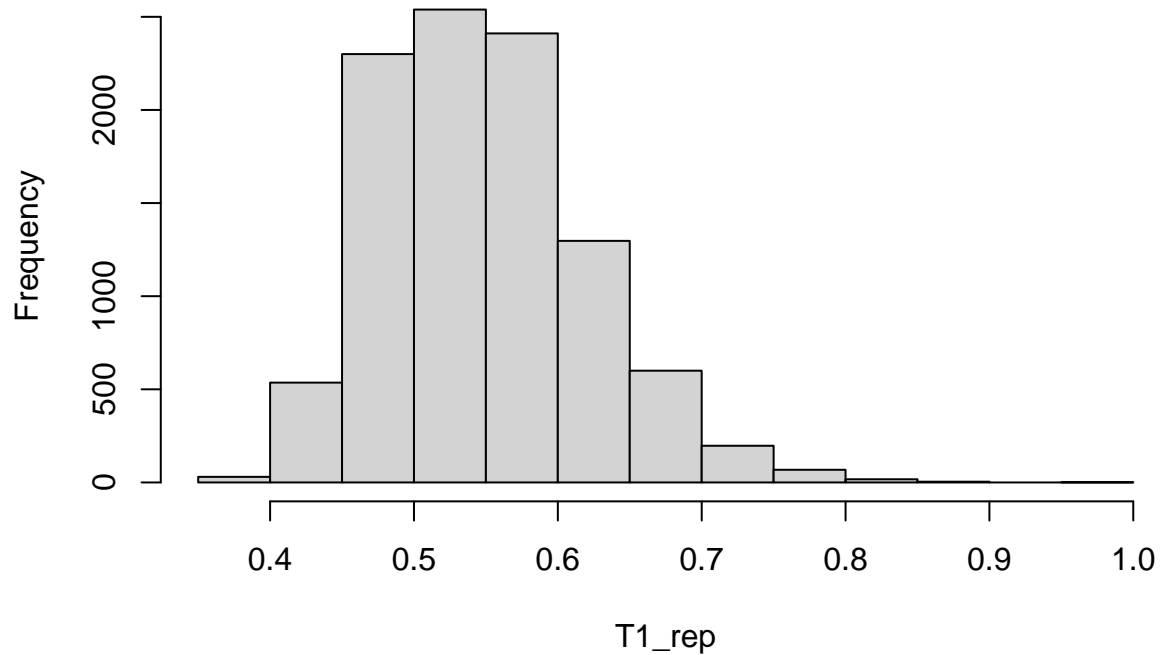
```

for (i in 1:1e4) {
  theta <- rbeta(71, 5.4, 18.6)
  rep <- c()
  for (j in 1:71) {
    rep <- c(rep, rbinom(1, size=n[j], theta[j])/n[j])
  }
  T1_rep <- c(T1_rep, max(rep))
  T2_rep <- c(T2_rep, var(rep))
  T3_rep <- c(T3_rep, sum(rep == 0))
}

hist(T1_rep)

```

Histogram of T1_rep



```

cat('p-value for T1 is', mean(T1_rep >= T1))

```

```

## p-value for T1 is 0.9998

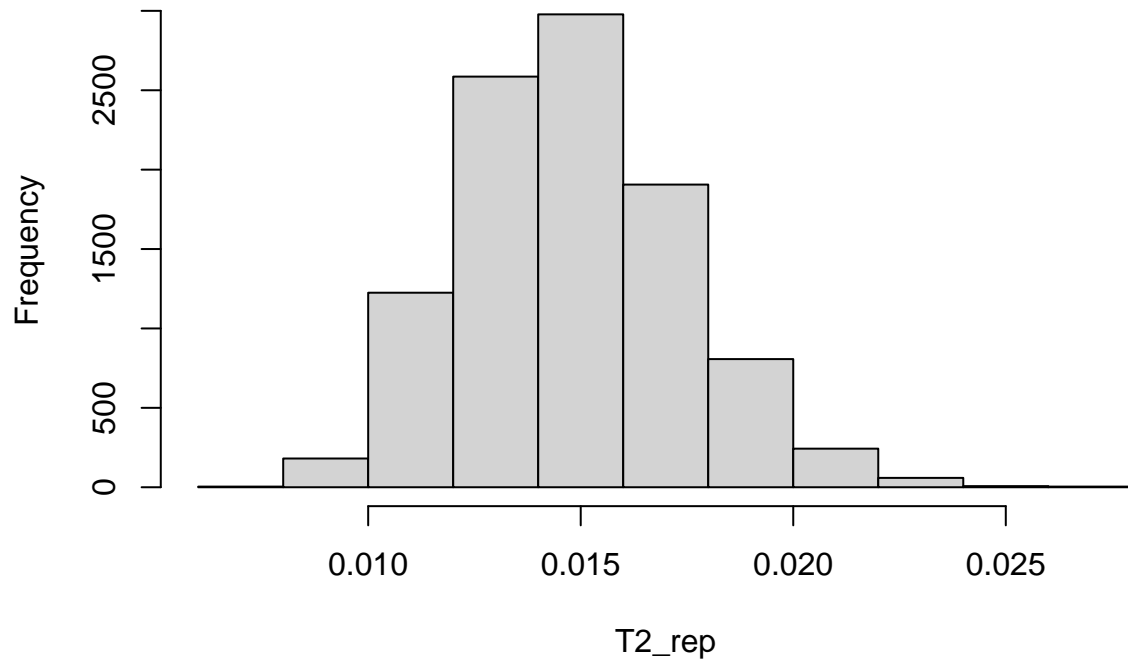
```

```

hist(T2_rep)

```

Histogram of T2_rep

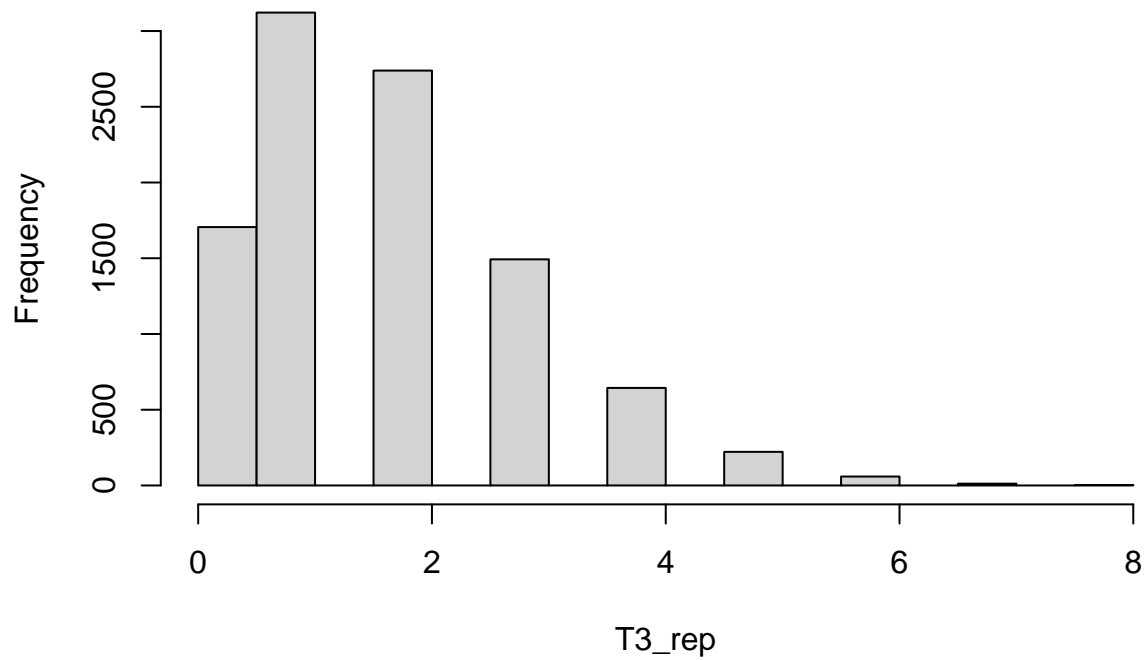


```
cat('p-value for T2 is', mean(T2_rep >= T2))
```

```
## p-value for T2 is 0.9447
```

```
hist(T3_rep)
```

Histogram of T3_rep



```
cat('p-value for T3 is', mean(T3_rep >= T3))
```

```
## p-value for T3 is 0
```