STAT 403 HW 3

Chongyi Xu

April 18, 2018

Question 1

a. Under a significant level $\alpha = 10\%$, can we reject the null hypothesis using t-test?

```r
t.test(chickwts[chickwts$feed=="linseed",1], mu = 240)
```

```
##
##  One Sample t-test
##
## data:  chickwts[chickwts$feed == "linseed", 1]
## t = -1.4092, df = 11, p-value = 0.1864
## alternative hypothesis: true mean is not equal to 240
## 95 percent confidence interval:
##  185.561 251.939
## sample estimates:
## mean of x
##    218.75
```

Using t-test, we can see that the p-value we got is ≈ 0.1864. Under $\alpha = 0.1$, we cannot reject the null hypothesis that chicken whose feed is linseed has an average weight 240.

b. What are the mean and standard deviation of this 12 chicken whose feed is linseed?

```r
dat <- chickwts[chickwts$feed=="linseed",]
print(paste('The mean of these chicken is: ', mean(dat$weight)))
```

```
## [1] "The mean of these chicken is:  218.75"
```

```r
print(paste('The standard deviation of these chicken is: ', sd(dat$weight)))
```

```
## [1] "The standard deviation of these chicken is:  52.2356983471857"
```

c. Assume the actual distribution of those chicken taking linseed as their feed follows from a normal distribution with mean 220 and standard deviation 52, i.e.,N(220,52^2). Use a size $N = 10,000$ Monte Carlo Simulation to find out the power of this t-test under significance level $\alpha = 10\%$?

```r
sample_size <- 12
N <- 10000
set.seed(403)

ts <- rep(NA, N)
for (i in 1:N) {
  sim_dat <- rnorm(sample_size, mean=220, sd=52)
  ts[i] <- t.test(sim_dat, mu=240)$p.value<=0.1
}
print(paste("The power of under alpha=10% is ", mean(ts)))
```

```
## [1] "The power of under alpha=10% is  0.3451"
```

d. What will the power be if the samplesize is $n = 12, 24, 36, 48, 60, 72, 84, 96$? Use a size $N = 10,000$ Monte Carlo Simulation for each case to find out. Moreover, use a figure to display the power versus sample size.

```r
n <- 12 * 1:8
powers <- rep(NA, length(n))
for (k in 1:length(n)) {
  sample_size <- n[k]
  N <- 10000

  ts <- rep(NA, N)
  for (i in 1:N) {
    sim_dat <- rnorm(sample_size, mean=220, sd=52)
    ts[i] <- t.test(sim_dat, mu=240)$p.value<=0.1
  }
  powers[k] <- mean(ts)
  print(paste("The power of under alpha=10% at n =", n[k], "is ", powers[k]))
}
```

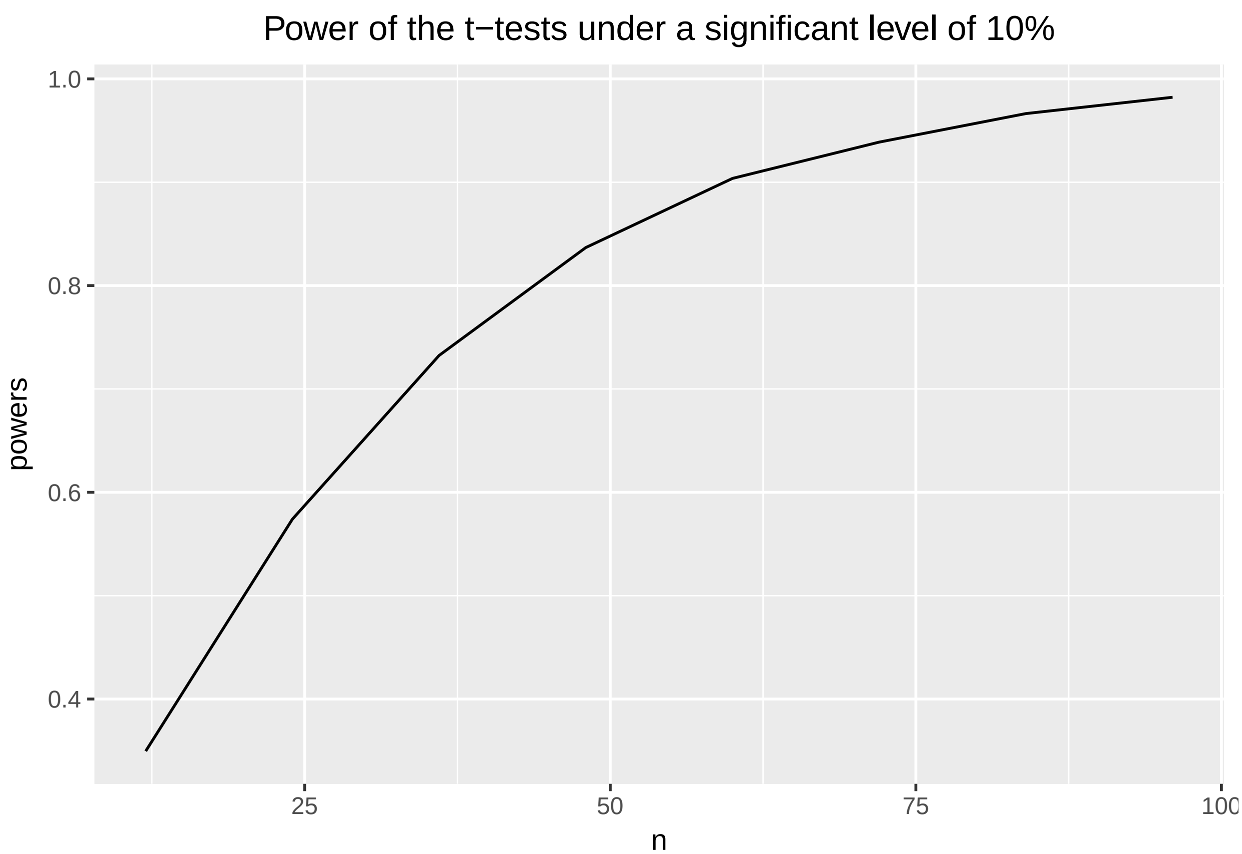
```
## [1] "The power of under alpha=10% at n = 12 is  0.3496"
## [1] "The power of under alpha=10% at n = 24 is  0.574"
## [1] "The power of under alpha=10% at n = 36 is  0.7324"
## [1] "The power of under alpha=10% at n = 48 is  0.8368"
## [1] "The power of under alpha=10% at n = 60 is  0.9037"
## [1] "The power of under alpha=10% at n = 72 is  0.9388"
## [1] "The power of under alpha=10% at n = 84 is  0.9664"
## [1] "The power of under alpha=10% at n = 96 is  0.9822"
```

Using a figure

```r
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.4.4
```

```r
ggplot() + geom_line(aes(x=n, y=powers)) +
  ggtitle('Power of the t-tests under a significant level of 10%') +
  theme(plot.title = element_text(hjust = 0.5))
```



e. Assume that when the sample size $n = 20$, estimating the power using a size $N = 10,000$ Monte Carlo Simulation has a Monte Carlo Error 0.005 (the variance caused by N). What will the Monte Carlo Error be when we increase N to $1,000,000$?

```
N <- 1000000
sample_size <- 20
ts <- rep(NA, N)
for (i in 1:N) {
  sim_dat <- rnorm(sample_size, mean=220, sd=52)
  ts[i] <- t.test(sim_dat, mu=240)$p.value<=0.1
}

sd(ts)/sqrt(N)
```

```
## [1] 0.0004999667
```

With N change to 1000000, the Monte Carlo Error will reduce to 0.0005.

Question 2

a. Use a size $N = 10000$ Monte Carlo Simulation to compute the mean of M_{50} and the standard deviation of M_{50}.

```
N <- 10000
sample_size <- 50

M50 <- rep(NA, N)
```

```r
for (k in 1:N) {
  M50[k] = min(runif(sample_size, min=0, max=1))
}

print(paste('The mean of M_50 is: ', mean(M50)))
```

```
## [1] "The mean of M_50 is:  0.0195291932990076"
```

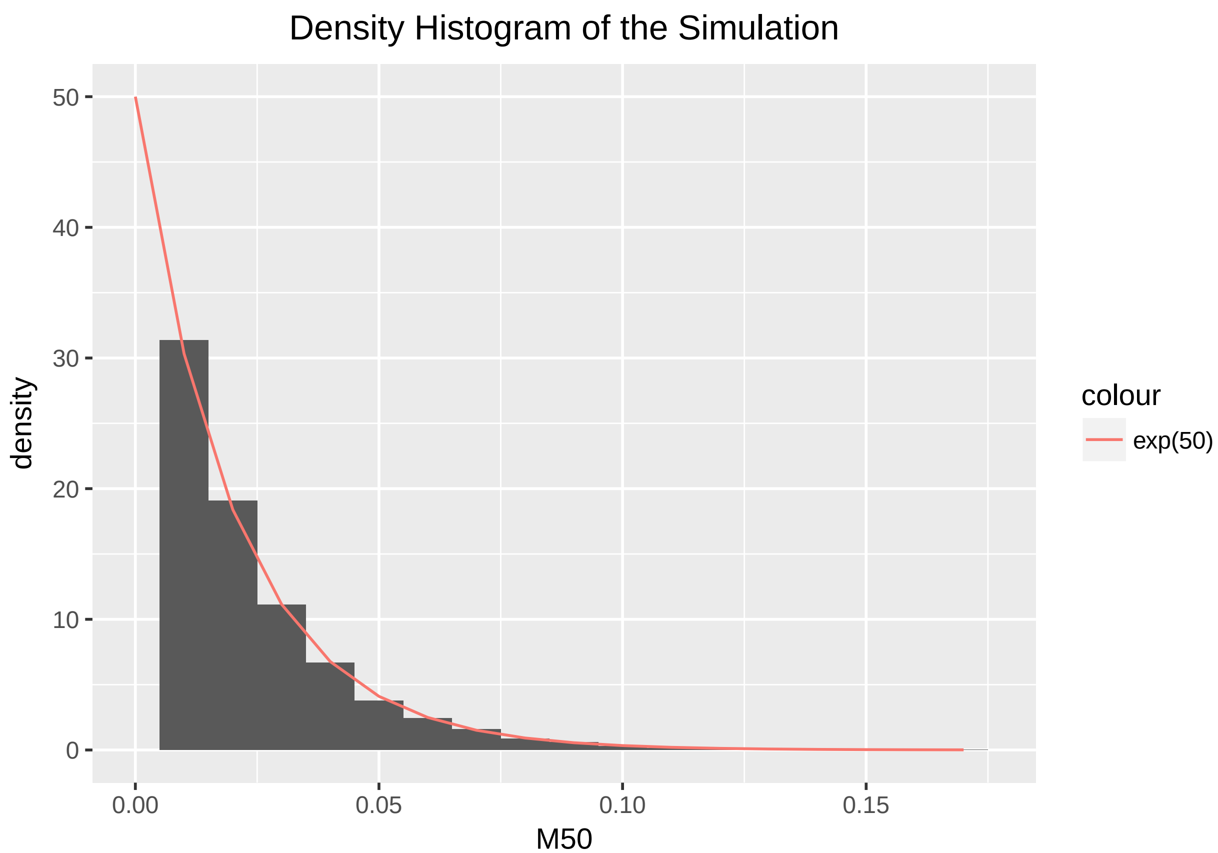
```r
print(paste('The standard deviation of M_50 is: ', sd(M50)))
```

```
## [1] "The standard deviation of M_50 is:  0.0191929943990882"
```

b It appears that the mean and standard deviation are roughly the same.This is a feature of the exponential distribution. Now use histogram to show that the distribution is roughly Exp(50), the exponential distribution with rate $\lambda = 50$. You need to attach the theoretical density curve of Exp(50) to this histogram.

```r
x <- seq(from=0, to=max(M50), by=0.01)
target <- dexp(x, rate=50)
ggplot() + geom_histogram(aes(M50, y=..density..), binwidth=0.01) +
  geom_line(aes(x, y=target, color='exp(50)')) +
  ggtitle('Density Histogram of the Simulation') +
  xlim(0, max(M50)) +
  theme(plot.title = element_text(hjust = 0.5))
```

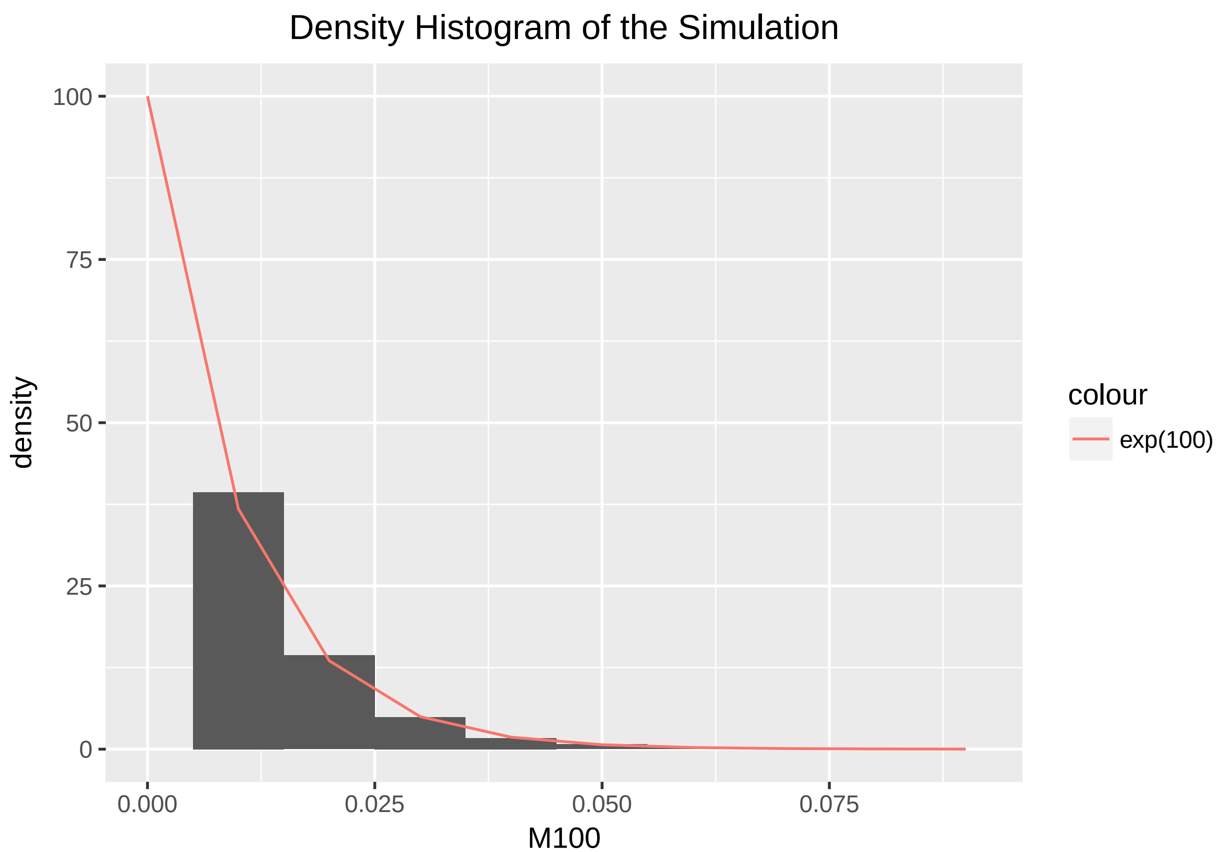
```
## Warning: Removed 1 rows containing missing values (geom_bar).
```



c. Now change the sample size 50 to 100 and repeat compute the histogram of $M_{100} = min\{U_1, \cdots, U_{100}\}$. Show the histogram and compare to the density curve of Exp(100).

```r
sample_size <- 100

M100 <- rep(NA, N)
for (k in 1:N) {
  M100[k] = min(runif(sample_size, min=0, max=1))
}
x <- seq(from=0, to=max(M100), by=0.01)
target <- dexp(x, rate=sample_size)
ggplot() + geom_histogram(aes(M100, y=..density..), binwidth=0.01) +
  geom_line(aes(x, y=target, color='exp(100)')) +
  ggtitle('Density Histogram of the Simulation') +
  xlim(0, max(M100)) +
  theme(plot.title = element_text(hjust = 0.5))
```



From the plot, we can see that the distribution still looks like Exp(100).

d. Show that for n IID random variables, $U_1, \cdots, U_n$~Uni[0,1], the random variable $M_n = minU_1, \cdots, U_n$ converges in distribution in the following sense:

$$n \cdot M_n \xrightarrow{\text{D}} \text{Exp}(1)$$

For n IID random variables, the probability

$$P(min\{U_1, \cdots, U_n\} \leq y) = 1 - P(\forall U \geq y) = 1 - (1 - F(y))^n$$

, where $F(y)$ is the CDF of Unif[0,1].

$$F(y) = \begin{cases} 0 & \text{for } y < 0 \\ \frac{y-0}{1-0} & \text{for } y \in [0,1) \\ 1 & \text{otherwise} \end{cases}$$

So

$$\begin{aligned} F_{n \cdot M_n}(y) &= P(n \cdot min\{U_1, \cdots, U_n\} \leq y) \\ &= 1 - (1 - F(y/n))^n \\ &= 1 - (1 - y/n)^n, 0 \leq y/n < 1 \end{aligned}$$

Consider $\lim_{n \to +\infty}(1 - y/n)^n$

$$\begin{aligned} e^{ln(1-\frac{y}{n})} &= e^{nln(1-\frac{y}{n})} \\ \lim_{n \to +\infty}(1 - \frac{y}{n})^n &= \lim_{n \to +\infty} e^{nln(1-y/n)} \\ &= e^{\lim \, nln(1-y/n)} \\ &= e^{\lim \, \frac{ln(1-y/n)}{1/n}} \\ &= e^{\lim \, \frac{\frac{y}{n^2}\frac{1}{1-y/n}}{-\frac{1}{n^2}}} \\ &= e^{\lim \, \frac{-y}{1-y/n}} \\ &= e^{-y} \end{aligned}$$

Therefore, we have $\lim_{n \to +\infty}(1 - y/n)^n \to e^{-y}$

So

$$\begin{aligned} F_{n \cdot M_n}(y) &= 1 - (1 - y/n)^n, 0 \leq y/n < 1 \\ &\to 1 - e^{-y} \\ &= F_{exp(1)}(y) \end{aligned}$$

In the other word, $n \cdot M_n \to$Exp(1). Q.E.D.