UNIVERSITY OF NEW SOUTH WALES
SCHOOL OF MATHEMATICS AND STATISTICS

MATH2901 Higher Theory of Statistics
Assignment One (Due on Wednesday 2pm, 23rd May)

Names (Print): KEEGAN GYOERY, JOHN LIDBETTER, EDWARD M°INNES, ALEX ROBINSON RUBY SMITH

I declare that this assessment item is my (our) own work, except where acknowledged, and has not been submitted for academic credit elsewhere, and acknowledge that the assessor of this item may, for the purpose of assessing this item:

- Reproduce this assessment item and provide a copy to another member of the University; and/or,

- Communicate a copy of this assessment item to a plagiarism checking service (which may then retain a copy of the assessment item on its database for the purpose of future plagiarism checking).

I certify that I have read and understood the University Rules in respect of Student Academic Misconduct.

Signed:

Date: 22/5/18

1

UNIVERSITY OF NEW SOUTH WALES

MATH 2901

HIGHER THEORY OF STATISTICS

_____

Assignment 2

_____

Keegan Gyoery (z5197058), John Lidbetter (z5160860),
Edward McInnes (z5162873), Alex Robinson (z5164884),
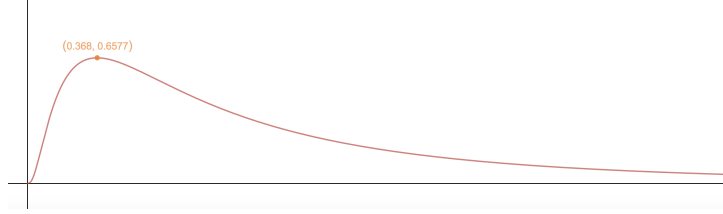Ruby Smith (z5113171)

May 22, 2018

1. X_1 and X_2 have the following density functions:

$$f_{X_1}(x) = \frac{1}{x\sqrt{2\pi}} e^{-(\ln x)^2/2} \qquad x > 0$$

$$f_{X_2}(x) = f_{X_1}(x)[1 + \sin(2\pi \ln x)] \qquad x > 0$$

(a) The graph for $f_{X_1}(x)$ is shown in the following figure.

(0.368, 0.6577)

The max point occurs at $\left(\frac{1}{e}, \sqrt{\frac{e}{2\pi}}\right)$. As x approaches 0, $\ln x$ approaches $-\infty$. Thus, $e^{-(\ln x)^2/2}$ approaches 0. As e^x dominates $\frac{1}{x}$, $f_{X_1}(x)$ approaches 0 as x approaches 0. As x approaches ∞, $\ln x$ approaches ∞. Thus, $e^{-(\ln x)^2/2}$ approaches 0. As e^x dominates $\frac{1}{x}$, $f_{X_1}(x)$ approaches 0 as x approaches ∞.

The graph for $f_{X_2}(x)$ is shown in the following figure.

(0.467, 1.2769)
(0.18, 1.0065)
(1.21, 0.6253)
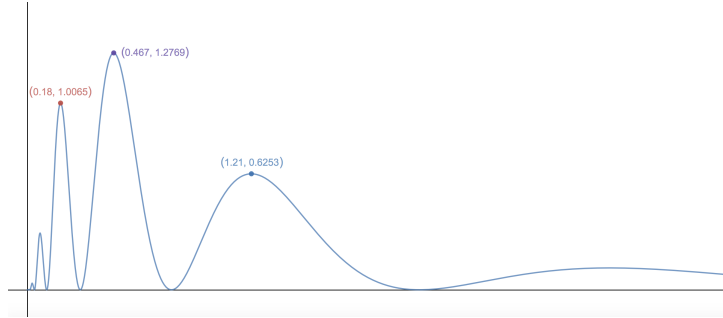
$f_{X_2}(x)$ behaves in the same manner as $f_{X_1}(x)$ when x approaches 0 or ∞, due to the same reasons, and e^x dominating $\sin x$. Furthermore, the multiplication by the sine function accounts for the sinusoidal-esque shape.

(b)

$$\mathbb{E}\big[X_1^r\big] = \int_{-\infty}^{\infty} x^r I_{(0,\infty)}(x) f_{X_1}(x) dx$$

$$= \int_0^{\infty} x^r f_{X_1}(x) dx$$

$$= \int_0^{\infty} \frac{x^r}{x\sqrt{2\pi}} e^{-(\ln x)^2/2} dx$$

$$= \int_0^{\infty} \frac{x^{r-1}}{\sqrt{2\pi}} e^{-(\ln x)^2/2} dx$$

Using the substitution $u = \ln x$, $x = e^u$, and $dx = e^u du$. The limits are now $-\infty$ and ∞.

$$= \int_{-\infty}^{\infty} \frac{(e^u)^{r-1}}{\sqrt{2\pi}} e^{-u^2/2} e^u du$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{u(r-1)} e^{(-u^2+2u)/2} du$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\left[-u^2+2u+2u(r-1)\right]/2} du$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{(-u^2+2ur)/2} du$$

$$= \int_{-\infty}^{\infty} e^{r^2/2} \frac{1}{\sqrt{2\pi}} e^{-(u^2-2ur+r^2)/2} du$$

$$= \int_{-\infty}^{\infty} e^{r^2/2} \frac{1}{\sqrt{2\pi}} e^{-(u-r)^2/2} du$$

Using the substitution $y = u - r$, $u = y + r$, and $du = dy$. The limits remain the same.

$$= e^{r^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

$$= e^{r^2/2}(1)$$

$$\therefore \mathbb{E}\big[X_1^r\big] = e^{r^2/2}$$

(c)

$$\mathbb{E}\big[X_2^r\big] = \int_{-\infty}^{\infty} x^r I_{(0,\infty)}(x) f_{X_2}(x) dx$$

$$= \int_0^{\infty} x^r f_{X_2}(x) dx$$

$$= \int_0^{\infty} x^r f_{X_1}(x)[1 + \sin(2\pi \ln x)] dx$$

$$= \int_0^{\infty} x^r f_{X_1}(x) + x^r f_{X_1}(x) \sin(2\pi \ln x) dx$$

$$= \int_0^{\infty} x^r f_{X_1}(x) dx + \int_0^{\infty} x^r f_{X_1}(x) \sin(2\pi \ln x) dx$$

$$= \mathbb{E}\big[X_1^r\big] + \int_0^{\infty} x^r f_{X_1}(x) \sin(2\pi \ln x) dx$$

$$\therefore \mathbb{E}\big[X_2^r\big] = \mathbb{E}\big[X_1^r\big] + \int_0^{\infty} x^r f_{X_1}(x) \sin(2\pi \ln x) dx$$

(d) Let $M = \int_0^\infty x^r f_{X_1}(x) \sin(2\pi \ln x) dx$

Using the substitution $u = \ln x$, $x = e^u$, and $dx = e^u du$. The limits are now $-\infty$ and ∞.

$$\therefore M = \int_{-\infty}^\infty \frac{(e^u)^{r-1}}{\sqrt{2\pi}} e^{-u^2/2} \sin(2\pi u) e^u du$$

$$= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{u(r-1)} e^{(-u^2+2u)/2} \sin(2\pi u) du$$

$$= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{\left[-u^2+2u+2u(r-1)\right]/2} \sin(2\pi u) du$$

$$= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{(-u^2+2ur)/2} \sin(2\pi u) du$$

$$= \int_{-\infty}^\infty e^{r^2/2} \frac{1}{\sqrt{2\pi}} e^{-(u^2-2ur+r^2)/2} \sin(2\pi u) du$$

$$= \int_{-\infty}^\infty e^{r^2/2} \frac{1}{\sqrt{2\pi}} e^{-(u-r)^2/2} \sin(2\pi u) du$$

Now, using integration by parts, we set

$$u' = e^{r^2/2} \frac{1}{\sqrt{2\pi}} e^{-(u-r)^2/2} \qquad u = e^{r^2/2}$$

$$v' = 2\pi \cos(2\pi u) \qquad v = \sin(2\pi u)$$

$$\therefore M = e^{r^2/2} \sin(2\pi u) \Big|_{-\infty}^\infty - \int_{-\infty}^\infty 2\pi e^{r^2/2} \cos(2\pi u) du$$

$$= \lim_{a\to\infty} \left[e^{r^2/2} \sin(2\pi u) \Big|_{-a}^a \right] - \lim_{a\to\infty} \left[\int_{-a}^a 2\pi e^{r^2/2} \cos(2\pi u) du \right]$$

$$= \lim_{a\to\infty} \left[e^{r^2/2} \sin(2\pi u) \Big|_{-a}^a \right] - \lim_{a\to\infty} \left[2\pi e^{r^2/2} \frac{1}{2\pi} \sin(2\pi u) \Big|_{-a}^a \right]$$

$$= e^{r^2/2} \lim_{a\to\infty} \left[\sin(2\pi u) \Big|_{-a}^a - \sin(2\pi u) \Big|_{-a}^a \right]$$

$$= e^{r^2/2} \lim_{a\to\infty} 0$$

$$= 0$$

$$\therefore M = 0$$

2. A random variable X is said to follow a Pareto(α, k) distribution if the density function of X is:

$$f_X(x) = \frac{\alpha k^\alpha}{x^{\alpha+1}} \qquad \alpha, k > 0 \text{ and } x > k$$

Suppose, for $n > 2$, we have a sequence of iid Pareto(α, k) random variables $X_1, \ldots, X_n$

(a) In order to compute the MLE for k and α, we derive the likelihood function, and the log-likelihood function.

$$L(\alpha, k) = \prod_{i=1}^{n} \frac{\alpha k^\alpha}{x_i^{\alpha+1}}$$

$$\therefore l(\alpha, k) = \ln \left(\prod_{i=1}^{n} \frac{\alpha k^\alpha}{x_i^{\alpha+1}} \right)$$

$$= \sum_{i=1}^{n} \ln \left(\frac{\alpha k^\alpha}{x_i^{\alpha+1}} \right)$$

$$= \sum_{i=1}^{n} \left[\ln \left(\alpha k^\alpha \right) - \ln \left(x_i^{\alpha+1} \right) \right]$$

$$= \sum_{i=1}^{n} \ln \left(\alpha k^\alpha \right) - \sum_{i=1}^{n} \ln \left(x_i^{\alpha+1} \right)$$

$$= n \ln \left(\alpha k^\alpha \right) - \sum_{i=1}^{n} \ln \left(x_i^{\alpha+1} \right)$$

$$= n \ln(\alpha) + n\alpha \ln(k) - (\alpha + 1) \sum_{i=1}^{n} \ln(x_i) \ldots (1)$$

Considering (1) as an equation in k, we note that (1) is increasing over all values of k, as $n, \alpha > 0$. Thus, $l(\alpha, k)$ is maximised when k takes its maximum value. As $k \leq x_i$, the maximum value that k can take is:

$$k = \min(x_i)$$
$$\therefore \widehat{k} = \min(X_i)$$

Thus, $\widehat{k}$ is the MLE for k.

Now, considering (1) as an equation in α, we have:

$$\therefore \frac{\partial l(\alpha, k)}{\partial \alpha} = \frac{n}{\alpha} + n\ln(k) - \sum_{i=1}^{n} \ln(x_i)$$

$$\frac{\partial l(\alpha, k)}{\partial \alpha} = 0$$

$$0 = \frac{n}{\alpha} + n\ln(k) - \sum_{i=1}^{n} \ln(x_i)$$

$$\frac{n}{\alpha} = \sum_{i=1}^{n} \ln(x_i) - n\ln(k)$$

$$\therefore \alpha = \frac{n}{\sum_{i=1}^{n} \ln(x_i) - n\ln(k)}$$

$$\therefore \widehat{\alpha} = \frac{n}{\sum_{i=1}^{n} \ln(X_i) - n\ln\left(\widehat{k}\right)}$$

$$\therefore \widehat{\alpha} = \frac{n}{\sum_{i=1}^{n} \left[\ln(X_i) - \ln(\min(X_i))\right]}$$

$$\frac{\partial^2 l(\alpha, k)}{\partial \alpha^2} = \frac{-n}{\alpha^2}$$

$$\therefore \frac{\partial^2 l(\alpha, k)}{\partial \alpha^2} < 0 \quad \forall \alpha$$

Thus, $\widehat{\alpha}$ maximises the log-likelihood function, and thus maximises the likelihood function, and therefore is the MLE for α.

(b) The MLE of k is $\widehat{k} = \min(X_i)$. In order to derive the distribution of $\widehat{k}$, we must consider the CDF of a minimum of a sequence of random variables. Let $Y = \min(X_i)$.

$$
\begin{aligned}
F_Y(y) &= \mathbb{P}(Y \le y) \\
&= \mathbb{P}(\min(X_i) \le y) \\
&= 1 - \mathbb{P}(\min(X_i) > y) \\
&= 1 - \left([1 - F_{X_1}(y)][1 - F_{X_2}(y)]\ldots[1 - F_{X_n}(y)]\right) \\
&= 1 - \left[1 - F_X(y)\right]^n \quad \text{as } X_i \; \forall i \text{ are iid} \\
\therefore F_{\widehat{k}}(x) &= 1 - \left[1 - F_X(x)\right]^n \\
&= 1 - \left[1 - \left[1 - \left(\frac{k}{x}\right)^\alpha\right]\right]^n \\
\therefore F_{\widehat{k}}(x) &= 1 - \left(\frac{k}{x}\right)^{n\alpha}
\end{aligned}
$$

Thus, $\widehat{k} \sim \text{Pareto}(n\alpha, k)$

(c) The Bias of $\widehat{k}$ is given by:

$$\text{Bias}\left(\widehat{k}\right) = \mathbb{E}\left(\widehat{k}\right) - k$$

$$= \frac{n\alpha k}{n\alpha - 1} - k \quad \text{for } n\alpha > 1$$

$$= \frac{n\alpha k}{n\alpha - 1} - \frac{(n\alpha - 1)k}{n\alpha - 1}$$

$$\text{Bias}\left(\widehat{k}\right) = \frac{k}{n\alpha - 1}$$

$$\therefore \text{Bias}\left(\widehat{k}\right) > 0 \quad \text{as } k > 0 \text{ and } n\alpha > 1$$

Thus, the MLE $\widehat{k}$ is a biased estimator for k. An unbiased estimator for k requires:

$$\left[\frac{n\alpha k}{n\alpha - 1}\right]C - k = 0 \quad \text{for } C \text{ some constant}$$

$$\left[\frac{n\alpha k}{n\alpha - 1}\right]C = k$$

$$\left[\frac{n\alpha}{n\alpha - 1}\right]C = 1$$

$$\therefore C = \left[\frac{n\alpha - 1}{n\alpha}\right]$$

Thus, an MLE for k that is unbiased is:

$$\widehat{k} = \left[\frac{n\alpha - 1}{n\alpha}\right]\min(X_i)$$

3. Let $X \sim \mathcal{N}(0,1)$ and $Y \sim \mathcal{N}(0,1)$ be two independent random variables, and define $Z = \min(X, Y)$. First, define the following results:

$$\mathbb{P}(X > -\sqrt{z}) = \mathbb{P}(X \leq \sqrt{z}) \dots (A) \quad \text{by symmetry}$$
$$\mathbb{P}(X \leq z) = \mathbb{P}(Y \leq z) \dots (B) \quad \text{as } X \text{ and } Y \text{ have the same distribution}$$

Note that for identity (B), any variation on the inequality still yields the same identity.

$$\mathbb{P}(Z \leq z) = \mathbb{P}(\min(X, Y) \leq z)$$
$$= 1 - \mathbb{P}(\min(X, Y) > z)$$
$$= 1 - \mathbb{P}(X > z, Y > z)$$
$$= 1 - \mathbb{P}(X > z)\mathbb{P}(Y > z) \dots (C) \quad \text{independence}$$

Now, considering Z^2

$$\mathbb{P}(Z^2 \le z) = \mathbb{P}(-\sqrt{z} \le Z \le \sqrt{z})$$
$$= \mathbb{P}(Z \le \sqrt{z}) - \mathbb{P}(Z \le -\sqrt{z})$$
$$= 1 - \mathbb{P}(X > \sqrt{z})\mathbb{P}(Y > \sqrt{z}) - [1 - \mathbb{P}(X > -\sqrt{z})\mathbb{P}(Y > -\sqrt{z})] \quad \text{from } (C)$$
$$= \mathbb{P}(X > -\sqrt{z})\mathbb{P}(Y > -\sqrt{z}) - \mathbb{P}(X > \sqrt{z})\mathbb{P}(Y > \sqrt{z})$$
$$= \mathbb{P}(X > -\sqrt{z})[1 - \mathbb{P}(Y \le -\sqrt{z})] - \mathbb{P}(X > \sqrt{z})[1 - \mathbb{P}(Y \le \sqrt{z})]$$
$$= \mathbb{P}(X > -\sqrt{z}) - \mathbb{P}(X > \sqrt{z}) - \mathbb{P}(X > -\sqrt{z})\mathbb{P}(Y \le -\sqrt{z}) + \mathbb{P}(X > \sqrt{z})\mathbb{P}(Y \le \sqrt{z})$$
$$= \mathbb{P}(X > -\sqrt{z}) - \mathbb{P}(X > \sqrt{z}) - \mathbb{P}(X \le \sqrt{z})\mathbb{P}(Y > \sqrt{z}) + \mathbb{P}(X > \sqrt{z})\mathbb{P}(Y \le \sqrt{z}) \quad \text{from } (A)$$
$$= \mathbb{P}(X > -\sqrt{z}) - \mathbb{P}(X > \sqrt{z}) - \mathbb{P}(Y \le \sqrt{z})\mathbb{P}(X > \sqrt{z}) + \mathbb{P}(X > \sqrt{z})\mathbb{P}(Y \le \sqrt{z}) \quad \text{from } (B)$$
$$\therefore \mathbb{P}(Z^2 \le z) = \mathbb{P}(X > -\sqrt{z}) - \mathbb{P}(X > \sqrt{z})$$
$$= 1 - \mathbb{P}(X \le -\sqrt{z}) - [1 - \mathbb{P}(X \le \sqrt{z})]$$
$$= \mathbb{P}(X \le \sqrt{z}) - \mathbb{P}(X \le -\sqrt{z})$$
$$= \mathbb{P}(-\sqrt{z} \le X \le \sqrt{z})$$
$$\therefore \mathbb{P}(Z^2 \le z) = \mathbb{P}(X^2 \le z)$$

Since $X \sim \mathcal{N}(0, 1)$, $X^2 \sim \chi_1^2$ and thus $Z^2 \sim \chi_1^2$
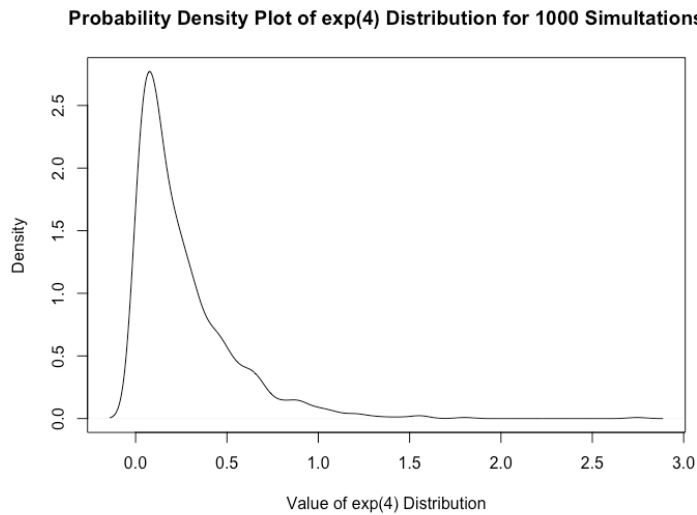
4. (a) Using the following code,

```
sim1=rexp(1000,4)
densim1=(density(sim1))
plot(densim1, xlab = "Value of exp(4) Distribution",
    main = "Probability Density Plot of exp(4) Distribution for 1000 Simultations")
```

we get the following figure.

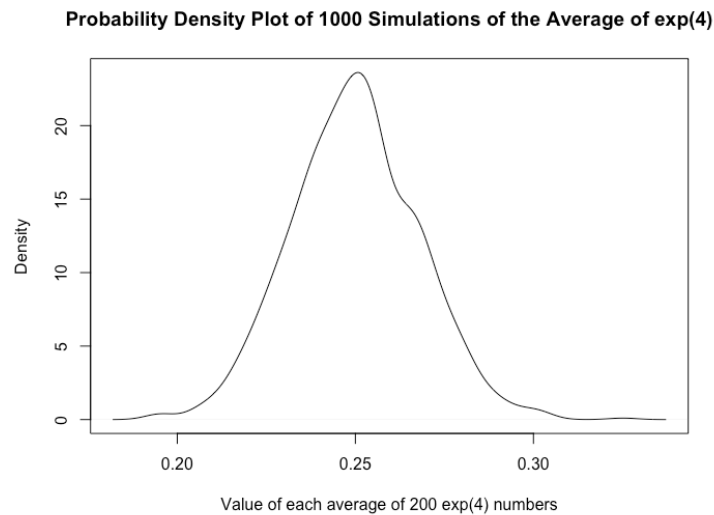Probability Density Plot of exp(4) Distribution for 1000 Simultations

The above figure's right tail is longer than the left tail, and the mean sits further to the right than the median, thus the data is right skewed.

(b) Using the following code,

```
sim2=replicate(1000,((1/200)*(sum(rexp(200,4)))))
densim2=density(sim2)
plot(densim2, xlab = "Value of each average of 200 exp(4) numbers",
    main = "Probability Density Plot of 1000 Simulations of the Average of exp(4)")
```
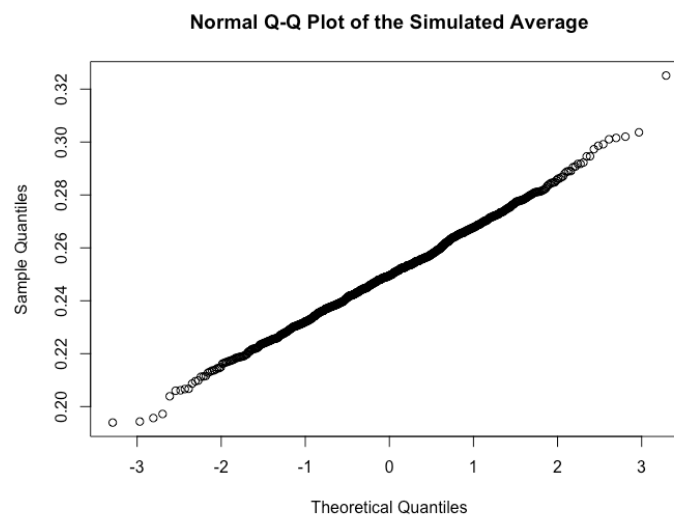
we get the following figure.

Probability Density Plot of 1000 Simulations of the Average of exp(4)

Using the following code,

```
qqnorm(sim2, main = ,"Normal Q-Q Plot of the Simulated Average")
```

we get the following figure.
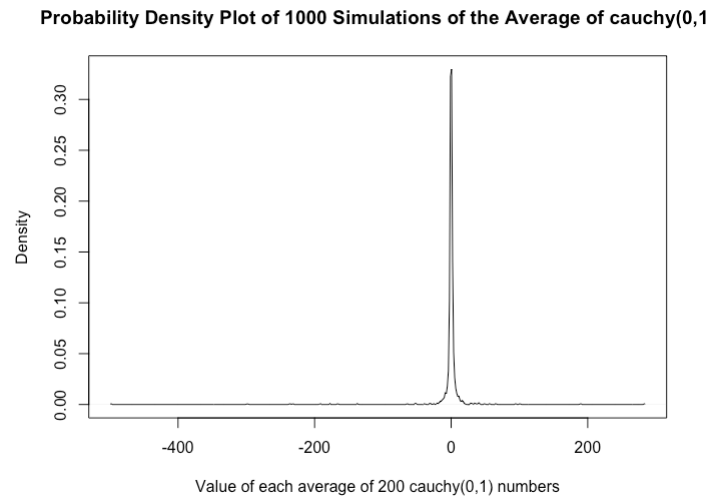
Normal Q-Q Plot of the Simulated Average

(c) The Probability Density Plot demonstrates that the distribution of averages of 200 simulations of the exp(4) distribution approximates a normal distribution. In addition, the quantile plot is approximately a straight line, and thus the distribution of the 200 simulations are a linear transform of the normal distribution. This indeed verifies the Central Limit Theorem.

(d) Using the following code,

```
sim3=replicate(1000,((1/200)*(sum(rcauchy(200,0,1)))))
densim3=density(sim3)
plot(densim3, xlab = "Value of each average of 200 cauchy(0,1) numbers",
    main = "Probability Density Plot of 1000 Simulations of the Average of cauchy(0,1)")
```
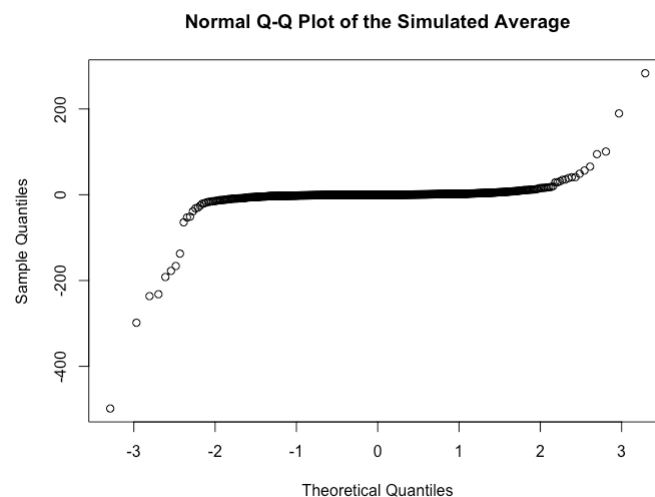
we get the following figure.

Probability Density Plot of 1000 Simulations of the Average of cauchy(0,1

Using the following code,

```
qqnorm(sim3, main = ,"Normal Q-Q Plot of the Simulated Average")
```

we get the following figure.

Normal Q-Q Plot of the Simulated Average

(e) The distribution of the mean of n independent identically distributed samples from a Cauchy Distribution has the same distribution as the original Cauchy Distribution, regardless of n. Moreover, the probability density plot clearly does not approximate a normal distribution due to its sharp peak and long tails. In addition, the quantile plot does not form a straight line, and thus, the samples from the Cauchy Distribution are not a linear transform of the normal distribution. Thus, the sample mean will never approximate a normal distribution, and in this way, the Central Limit Theorem is not verified.