

Homework 1

Exercise 1

Let $U \sim Unif(0, 1)$ and $X = g(u) = -c * \ln(u)$ given that $c > 0$

$$\begin{aligned} X = -c * \ln(U) &\implies \\ -\frac{X}{c} &= \ln(U) \implies \\ e^{-\frac{X}{c}} &= U \\ \therefore g^{-1}(u) &= e^{-\frac{u}{c}} \end{aligned} \tag{1}$$

Since $U \sim Unif(0, 1)$ the pdf is $f_U(u) = \frac{1}{1-0} = 1$ In order to get the the pdf of X, the following is done:

$$\begin{aligned} f(x) &= f_U(g^{-1}(x)) * |d/dx(g^{-1}(x))| = \\ &= 1 * |d/dx(e^{-x/c})| = \\ &= 1 * |-1/c * e^{-x/c}| = \\ &= 1/c * e^{-x/c} \sim exp(1/c) \end{aligned} \tag{2}$$

Exercise 2

Let $U_i \sim Unif(0, 1)$ where i is 1..n and $X = -\frac{1}{k} * \ln(\prod_{i=1}^n U_i)$

$$\begin{aligned} f(x) &= -\frac{1}{k} * \ln(\prod_{i=1}^n U_i) = \\ &= -\frac{1}{k} * \sum_{i=1}^n \ln(U_i) = \\ &= \sum_{i=1}^n -\frac{1}{k} * \ln(U_i) \end{aligned} \tag{3}$$

From Exercise one, we are given that $-\frac{1}{c}\ln(U) \sim exp(c)$ where $U \sim Unif(0, 1) \therefore -\frac{1}{k} * \ln(U_i) \sim exp(k)$ The sum of exponential distributions that are i.i.d have a gamma distribution such that $gamma(n, \lambda)$ where λ is the parameter passed from the exponential distribution and n is the number of random variables.

$$\therefore -\frac{1}{k} * \ln(\prod_{i=1}^n U_i) \sim gamma(n, k) \tag{4}$$

The proof of the summation of exponential random variables being the same as a gamma distribution can be proved by induction. let S_i be the sum of exponential distributions.

for $i = 1$, since $exp(\lambda) \equiv gamma(1, \lambda)$ the base case for induction is true.

Assume that S_i where $i = 1..n$ is distributed the following way:

$$S_{1..n} = gamma(n, \lambda) \sim \Gamma(n)^{-1} \lambda^n x^{n-1} e^{-\lambda * x} \tag{5}$$

The last step is $gamma(n + 1) \equiv S_{1..n} + S_1$

$$\begin{aligned}
f_Z(z) &= \int_0^z f_{S_{1..n}}(y) * f_{S_1}(z-y) dy = \\
&= \int_0^z \Gamma(n)^{-1} \lambda^n y^{n-1} e^{-\lambda y} * \lambda * e^{-(z-y)\lambda} dy = \\
&= \Gamma(n)^{-1} \lambda^{n+1} e^{-z\lambda} * \int_0^z y^{n-1} e^{-y\lambda+y\lambda} dy = \\
&= \Gamma(n)^{-1} \lambda^{n+1} e^{-z\lambda} * \frac{1}{n} * \int_0^z y^n dy = \\
&= \Gamma(n+1)^{-1} \lambda^{n+1} e^{-z\lambda} * z^{(n+1)-1} \sim \text{gamma}(n+1, \lambda)
\end{aligned} \tag{6}$$

\therefore The sum of iid random exponential variables follows a gamma distribution.

Exercise 3

Part A

$$\text{Let } z = 1 + e^{-(x-\mu)/\beta} \implies dz = -\frac{1}{\beta} e^{-(x-\mu)/\beta} * dx \implies dx = \frac{dz}{-\frac{1}{\beta} e^{-(x-\mu)/\beta}}$$

$$\begin{aligned}
F(x) &= \int_{-\infty}^x f(x) dx = \\
&= \int \frac{1}{\beta} \frac{e^{-(x-\mu)/\beta}}{(1 + e^{-(x-\mu)/\beta})^2} dx = \\
&= \int \frac{1}{\beta} \frac{e^{-(x-\mu)/\beta}}{z^2} * \frac{dz}{-\frac{1}{\beta} e^{-(x-\mu)/\beta}} = \\
&= \int -\frac{1}{z^2} dz = \\
\frac{1}{z} &= \Big|_{-\infty}^x \frac{1}{1 + e^{-(x-\mu)/\beta}} = \frac{1}{1 + e^{-(x-\mu)/\beta}}
\end{aligned} \tag{7}$$

To find the inverse, we do the following:

$$\begin{aligned}
Y &= \frac{1}{1 + e^{-(x-\mu)/\beta}} \implies \\
1 + e^{-(x-\mu)/\beta} &= \frac{1}{Y} \implies \\
e^{-(x-\mu)/\beta} &= \frac{1}{Y} - 1 \implies \\
-(x-\mu)/\beta &= \ln\left(\frac{1}{Y} - 1\right) \implies \\
-(x-\mu) &= \ln\left(\frac{1}{Y} - 1\right)\beta \implies \\
x &= \mu - \ln\left(\frac{1}{Y} - 1\right)\beta \\
\therefore F^{-1}(x) &= \mu - \ln\left(\frac{1}{x} - 1\right)\beta
\end{aligned} \tag{8}$$

```
u = runif(1e4)
```

```
inverse_logistic_cdf = function(x, mu = 0 , B = 1){
```

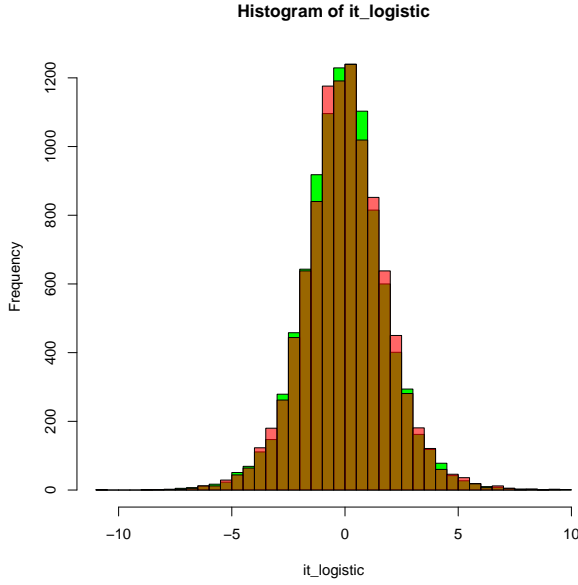
```

mu - log(1/x-1)*B
}

it_logistic = inverse_logistic_cdf(u)
r_logistic = rlogis(1e4)

hist(it_logistic, breaks = 50, col = "green")
hist(r_logistic, breaks = 50, col = "#FF000099", add = T)

```



Based on a comparison of the histograms of the sampled inverse function of the logistic distribution against the random generator of the logistic distribution, there seems to be a close relationship between the two. This indicates a similar behavior, which could lead to the conclusion that both functions act in the same way.

Part B

Let $\tan(z) = \frac{x-\mu}{\sigma} \implies dz * \sec^2(z) = \frac{1}{\sigma} * dx \implies dx = \sigma * \sec^2(z) * dz$

$$\begin{aligned}
 F(x) &= \int_{-\infty}^x f(x) dx = \\
 &= \int \frac{1}{\pi * \sigma} * \frac{1}{1 + (\frac{x-\mu}{\sigma})^2} dx = \\
 &= \int \frac{1}{\pi * \sigma} * \frac{1}{1 + \tan^2(z)} * \sigma * \sec^2(z) * dz = \\
 &= \frac{1}{\pi} \int 1 * dz = \\
 &= \frac{1}{\pi} \Big|_z = \\
 &= \frac{1}{\pi} \Big|_{-\infty}^x \arctan\left(\frac{x-\mu}{\sigma}\right) = \\
 &= \frac{1}{\pi} * \arctan\left(\frac{x-\mu}{\sigma}\right) + \frac{1}{2}
 \end{aligned} \tag{9}$$

To find the inverse, we do the following:

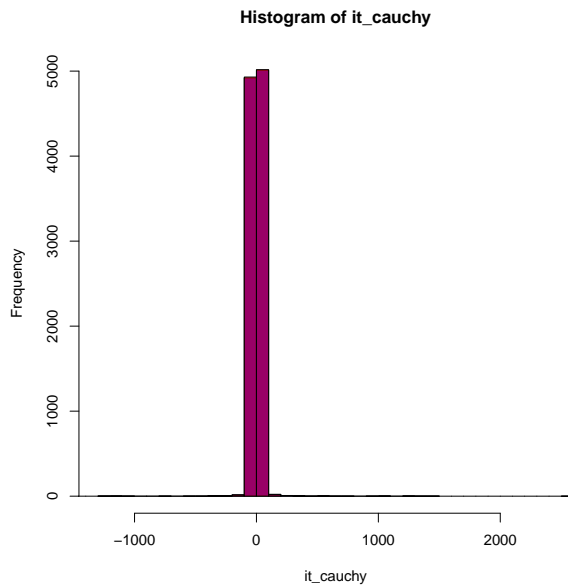
$$\begin{aligned}
 \frac{1}{\pi} * \arctan\left(\frac{x - \mu}{\sigma}\right) + \frac{1}{2} &= Y \implies \\
 \pi * \left(Y - \frac{1}{2}\right) &= \arctan\left(\frac{x - \mu}{\sigma}\right) \implies \tan\left(\pi * \left(Y - \frac{1}{2}\right)\right) = \frac{x - \mu}{\sigma} \implies \\
 \sigma * \tan\left(\pi * \left(Y - \frac{1}{2}\right)\right) + \mu &= x \\
 \therefore F^{-1}(x) &= \sigma * \tan\left(\pi * \left(x - \frac{1}{2}\right)\right) + \mu
 \end{aligned} \tag{10}$$

```

inverse_cauchy_cdf = function(x, mu = 0, sigma = 1){
  sigma*tan(pi*(x-0.5)) + mu
}
u = runif(1e4)
it_cauchy = inverse_cauchy_cdf(u)
r_cauchy = rcauchy(1e4)

hist(it_cauchy, breaks = 50, col = "blue")
hist(r_cauchy, breaks = 50, col = "#FF000099", add = T)

```



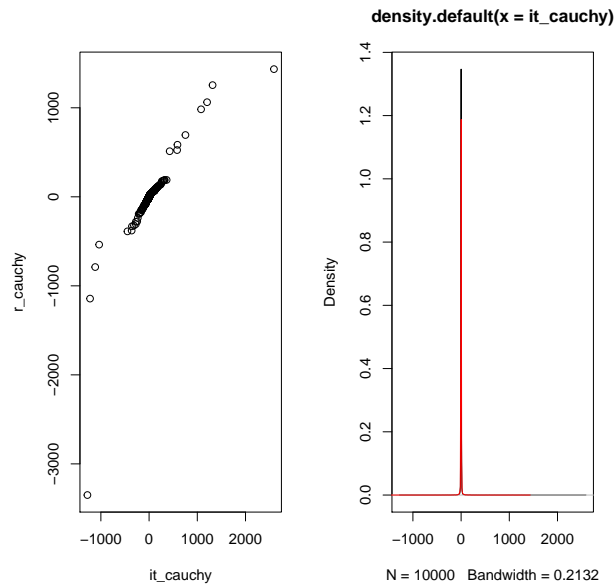
In both cases, the cauchy distribution seems to settle on numbers close to 0. However, the tails are harder to compare. Since there seems to be an overlap with the majority of points, it may be concluded that the inverse of the cauchy distribution act similar to the random generated sample of the cauchy distribution.

John Hotchkiss recommended looking at the density plot and qqplots between the two samples

```

par(mfrow=c(1,2))
qqplot(it_cauchy, r_cauchy)
plot(density(it_cauchy))
lines(density(r_cauchy), col = "red")

```



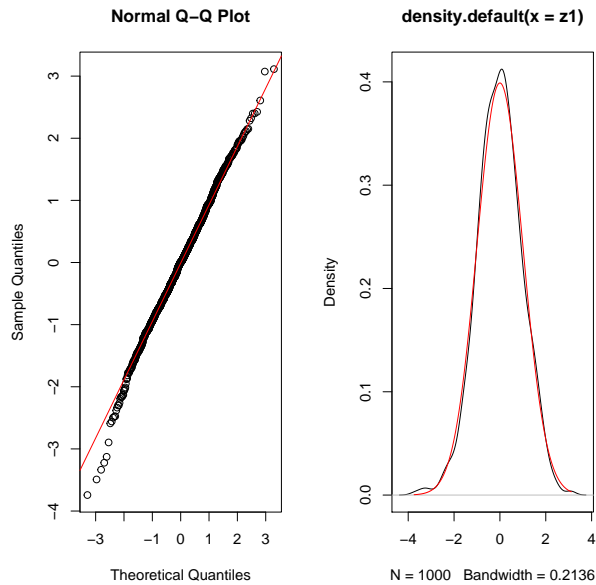
The qqplot doesn't show a clear relationship between the two, but the density plot shows a convergence at 0 for both densities

Exercise 4

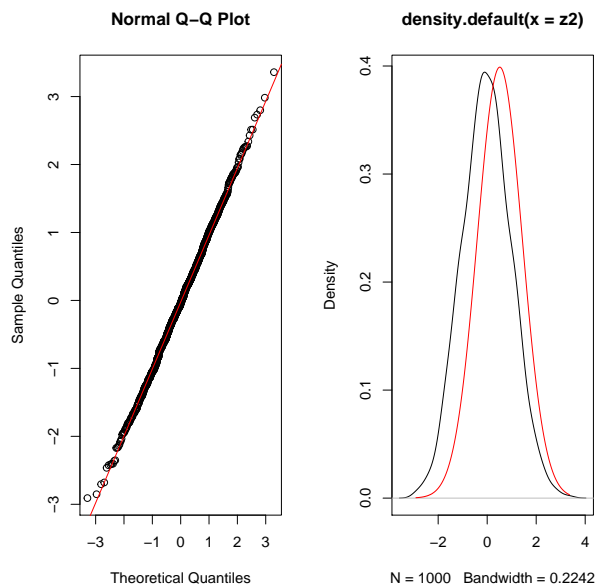
```
par(mfrow=c(1,2))
u1 = runif(1e3)
u2 = runif(1e3)

z1 = sqrt(-2*log(u1))*cos(2*pi*u2)
z2 = sqrt(-2*log(u1))*sin(2*pi*u2)

qqnorm(z1)
qqline(z1, col = "red")
plot(density(z1))
lines(seq(min(z1),max(z1), length = 1e4), dnorm(seq(min(z1),max(z1), length = 1e4)), col = "red")
```



```
qqnorm(z2)
qqline(z2, col = "red")
plot(density(z2))
lines(seq(min(z2),max(z2), length = 1e4), dnorm(seq(min(z1),max(z1), length = 1e4)), col = "red")
```



Both Z_1 and Z_2 follow a normal distribution based on a comparison of plots with the normal distribution's density along with the qqplots against the normal distribution.

Exercise 5

To Prove CLT works for a non-normal distribution, the beta distribution was sampled with $\alpha = 1$ and $\beta = 2$ such that $Samples \sim \text{beta}(1, 2)$

```

samples = matrix(nrow = 35, ncol = 1e3)
for(i in 1:1000){
  samples[,i] = rbeta(35, 1,2)
}

```

Part A

```

means = apply(samples, 2, mean)
summary(means)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.1911 0.3056 0.3342 0.3329 0.3608 0.4487

sd(means)

## [1] 0.0402071

```

From the definition of the beta distribution, $E(X) = \frac{\alpha}{\alpha+\beta} = \frac{1}{1+2} = \frac{1}{3}$. Based on the summary, the mean that is approximated is very close to the theoretical mean.

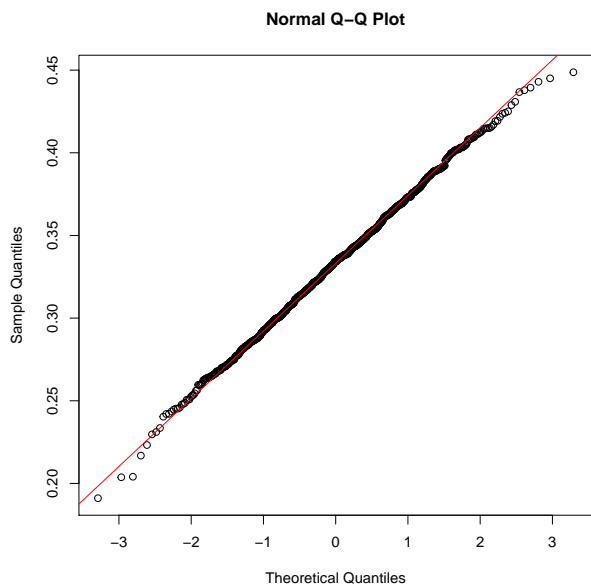
The variance is defined as $Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2 * (\alpha+\beta+1)}$. Since we are dealing with a sample of $n = 35$, the standard deviation should be $SD(X) = \sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2 * (\alpha+\beta+1)} / 35} = 0.0398$. The approximated value is also close to the theoretical standard deviation.

Part B

```

qqnorm(means)
qqline(means, col = "red")

```



Based off the qqplot, the distribution of means follows closely to the normal distribution; however, there are some noticeable movements away from the qqline at the tail ends.