Michael Leibert
Math 611
Homework 1

**1.** If $U \sim Uniform(0,1)$ and $c$ is a constant such that $c > 0$, show that the random variable $-c \cdot \log(U)$ is exponentially distributed with parameter c.

$$\text{Let } Z = -c \cdot \log(U) \implies -\frac{Z}{c} = \log(U) \implies \exp\left(-\frac{Z}{c}\right) = U$$

We can do a transformation to show that $Z$ has an exponential distribution.

$$F_Z(z) = P(Z \le z) = P(-c \cdot \log(U) \le z) = P\left(U \ge \exp\left(-\frac{z}{c}\right)\right) = 1 - \exp\left(-\frac{z}{c}\right)$$

$$\frac{d}{dZ} F_Z(z) = \frac{1}{c} \exp\left(-\frac{z}{c}\right) \qquad z > 0$$

Or...

$$f_Z(z) = f_U(u) \left| \frac{d}{dZ} \exp\left(-\frac{z}{c}\right) \right| = \frac{1}{1-0} \cdot \frac{1}{c} \cdot \exp\left(-\frac{z}{c}\right) = \frac{1}{c} \exp\left(-\frac{z}{c}\right) \qquad z > 0$$

**2.** If $U_1, U_2, \ldots, U_n$ are iid $Uniform(0,1)$ random variables, show that the random variable $X = -\frac{1}{k} \cdot \log\left(\prod_{i=1}^{n} U_i\right)$ has a Gamma distribution with parameters n and k, that is, $Gamma(n,k)$.

We can write $X$ as the following...

$$X = -\frac{1}{k} \log\left(\prod_{i=1}^{n} U_i\right) = -\frac{1}{k} \sum_{i=1}^{n} \log(U_i) = \sum_{i=1}^{n} -\frac{1}{k} \log(U_i)$$

And perform a transformation on $Z_i = -\frac{1}{k} \log(U_i)$, to show each $Z$ has an exponential distribution.

$$f_{Z_i}(z) = f_{U_i}(u) \left| \frac{d}{dZ_i} \exp(-kz) \right| = \frac{1}{1-0} \cdot k \cdot \exp(-kz) = k \exp(-kz)$$

Thus it is shown $Z \sim \text{Exp}\left(\frac{1}{k}\right)$. We now find the moment generating function of a single $Z_i$.

$$\begin{aligned}
M_Z(t) &= \int_0^\infty \exp(tz) \cdot k \cdot \exp(-kz) \, \mathrm{d}z \\
&= k \int_0^\infty \exp\Big[ - z(-t+k) \Big] \, \mathrm{d}z && u = z(k-t) \\
& && \mathrm{d}u = (k-t)\, \mathrm{d}z \\
&= k \int_0^\infty e^{-u} \frac{1}{k-t} \, \mathrm{d}u && \frac{1}{k-t} \mathrm{d}u = \mathrm{d}z \\
&= \frac{k}{k-t} \qquad t < k
\end{aligned}$$

The Mgf of the linear combination $V = \sum_{i=1}^{n} W_i$ is $M_V(t) = \prod_{i=1}^{n} M_W(t) = [M_W(t)]^n$.

The random variable $X$, which is the sum of $n$ exponential RVs, has the Mgf:

$$M_X(t) = \prod_{i=1}^{n} M_{Z_i}(t) = \prod_{i=1}^{n} \frac{k}{k-t} = \left(\frac{k}{k-t}\right)^n .$$

If $X \sim \mathit{Gamma}(n,k)$, the Mgf is:

$$\begin{aligned}
M_X(t) &= \int_0^\infty \exp(tx)\ \frac{k^n}{\Gamma(n)}\ x^{n-1}\ \exp(-kx)\ \mathrm{d}x \\
&= \frac{k^n}{\Gamma(n)} \int_0^\infty \exp\Big[-x\ (-t+k)\Big]\ x^{n-1}\ \mathrm{d}x \\
&= \frac{k^n}{\Gamma(n)} \int_0^\infty e^{-u}\ \cancel{\frac{1}{k-t}}\ u^{n-1}\ \left(\frac{1}{k-t}\right)^{\cancel{n-1}} \mathrm{d}u \\
&= \frac{k^n}{\Gamma(n)} \left(\frac{1}{k-t}\right)^n \underbrace{\int_0^\infty e^{-u}\ u^{n-1}\ \mathrm{d}u}_{\Gamma(n)} \\
&= \frac{1}{\cancel{\Gamma(n)}} \left(\frac{k}{k-t}\right)^n \cancel{\Gamma(n)} \\
&= \left(\frac{k}{k-t}\right)^n \qquad t < k.
\end{aligned}$$

$$\begin{aligned}
x &= \frac{u}{(k-t)} \\
u &= x(k-t) \\
\mathrm{d}u &= (k-t)\ \mathrm{d}x \\
\frac{1}{k-t}\ \mathrm{d}u &= \mathrm{d}x
\end{aligned}$$

Because the sum of $n$ exponential RVs and the $\mathit{Gamma}(n,k)$ have the same moment generating functions, it follows that they have the same probability distribution.

**3.** R&C 2.2 page 45. First derive the inverse CDFs ($\mu = 0, \beta = 1, \sigma = 1$):

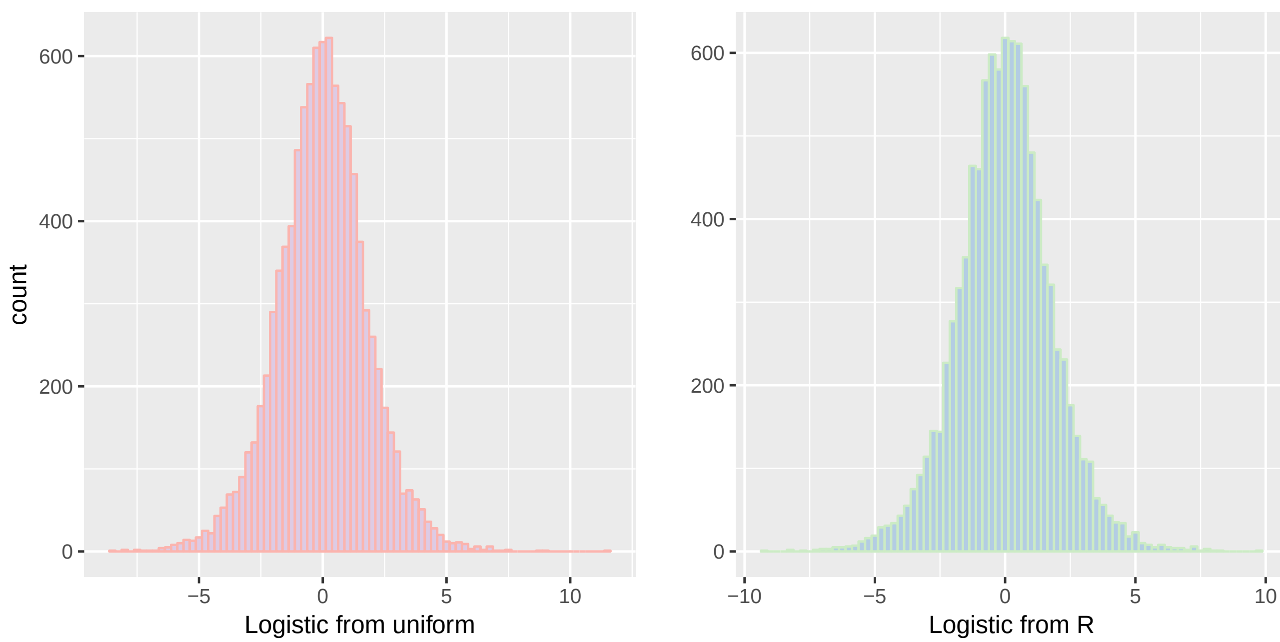
$$\begin{aligned}
\text{Logistic CDF: } F(x) &= \frac{1}{1+e^{-x}} \\
u &= \frac{1}{1+e^{-x}} \\
1 &= u\left(1+e^{-x}\right) \\
1 &= u + u \cdot e^{-x} \\
\frac{1-u}{u} &= e^{-x} \\
-\log\left(\frac{1-u}{u}\right) &= x
\end{aligned}$$

$$\begin{aligned}
\text{Cauchy CDF: } F(x) &= \frac{1}{2} + \frac{1}{\pi}\arctan(x) \\
u &= \frac{1}{2} + \frac{1}{\pi}\arctan(x) \\
u - \frac{1}{2} &= \frac{\arctan(x)}{\pi} \\
\pi u - \frac{\pi}{2} &= \arctan(x) \\
\tan\left(\pi u - \frac{\pi}{2}\right) &= x
\end{aligned}$$

```
## Loading required package:  ggplot2
```

```r
#3
n=10000

U<-runif(n)
X= -log((1-U)/U)
Y<-rlogis(n)
dat<-data.frame(X,Y)

p<-ggplot(dat, aes(X)) +  geom_histogram(col="#fbb4ae",fill="#decbe4",binwidth = .25) +
        labs(x="Logistic from uniform")
q<-ggplot(dat, aes(Y)) +  geom_histogram(col="#ccebc5",fill="#b3cde3",binwidth = .25) +
        labs(x="Logistic from R",y="")
```

```r
multiplot(p,q,cols=2)
```




```r
U<-runif(n)
X = tan( pi*U-pi/2 )
Y<-rcauchy(n)
dat<-data.frame(X,Y)
p<-ggplot(dat, aes(X)) +  geom_histogram(col="#3182bd",fill="#f7fcb9",  binwidth = .5) +
    labs(x="Cauchy from uniform") +  xlim(-10, 10)
q<-ggplot(dat, aes(Y)) +  geom_histogram(col="#636363",fill="#fdbb84",  binwidth = .5) +
    labs(x="Cauchy from R",y="") +  xlim(-10, 10)
```

```r
multiplot(p,q ,cols=2)

## Warning:  Removed 673 rows containing non-finite values (stat_bin).

## Warning:  Removed 664 rows containing non-finite values (stat_bin).
```

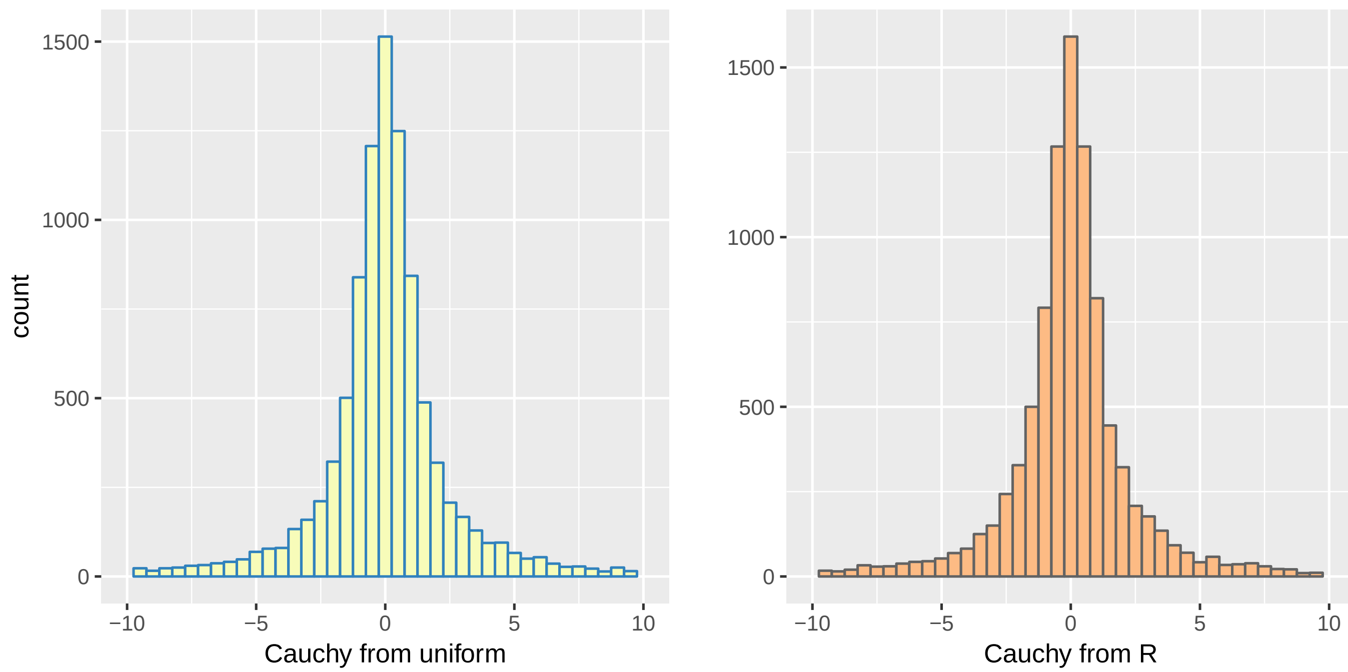
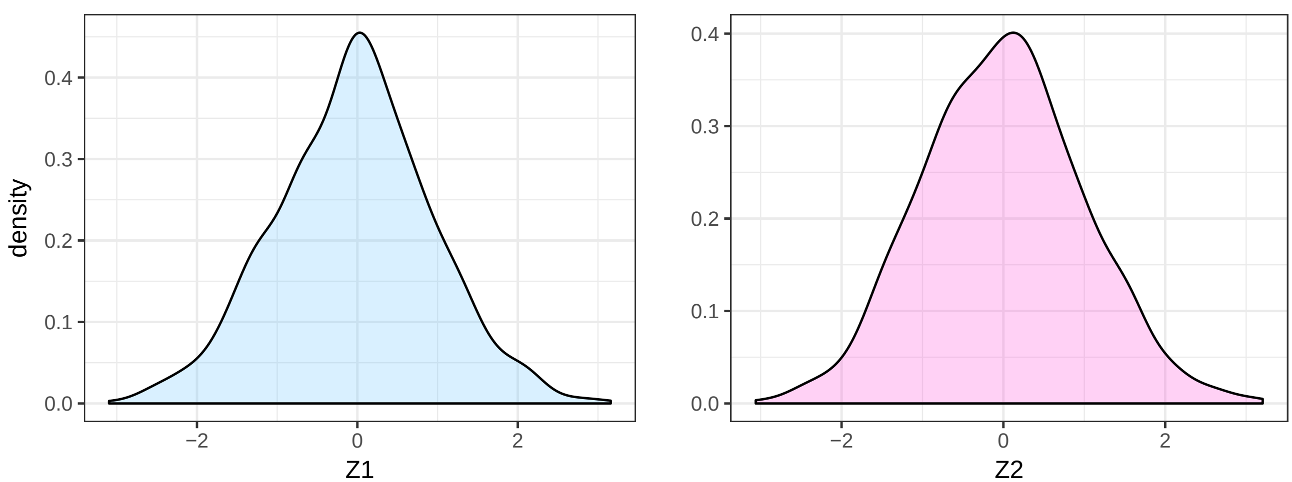



Figure 1: Each plot is missing 600+ values that were excluded for the sake of the visuals

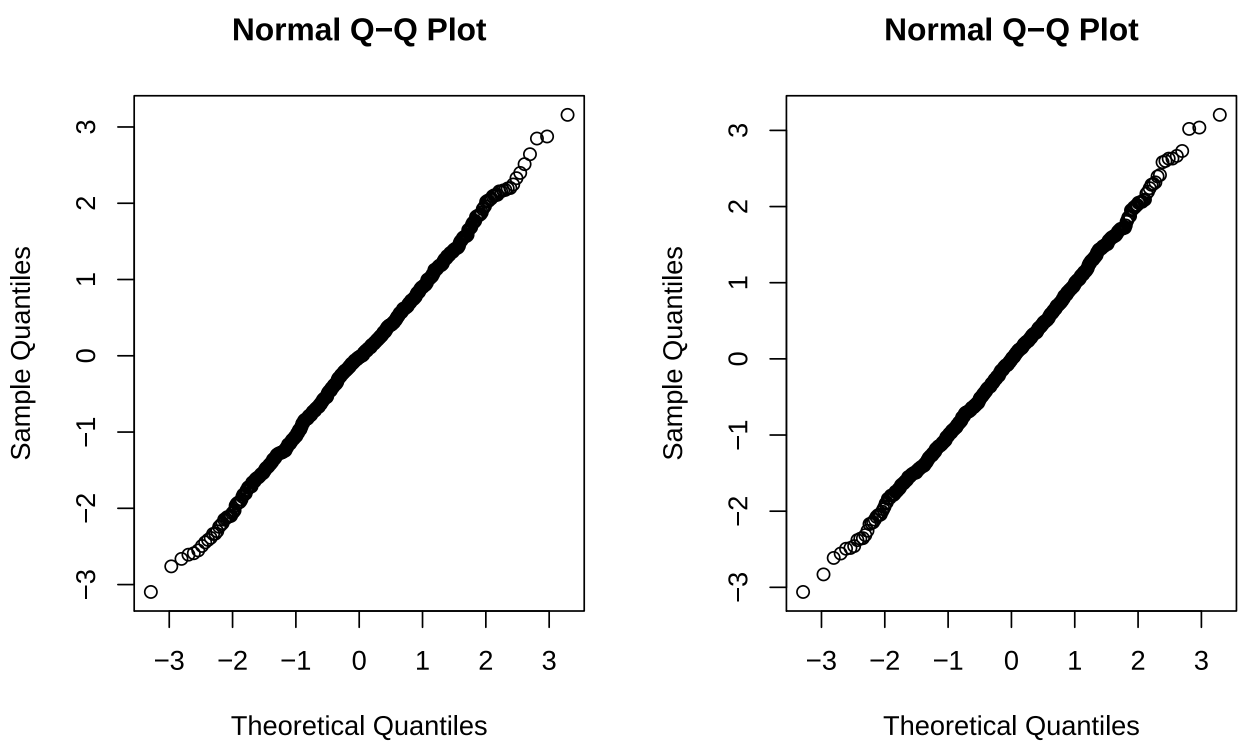
**4.** Use R to generate a pair of standard normal random variables $Z_1$ and $Z_2$ (of size $n = 1000$ each) using the Box-Muller transformation. For each variable construct a normal QQ-plot and a density plot.

```
#4 Box-Muller Transformation
n=1000
U<-runif(n);V<-runif(n)
Z1 <- sqrt( -2*log(U) ) * cos( 2*pi* V )
Z2 <- sqrt( -2*log(U) ) * sin( 2*pi* V )
BM<-data.frame(Z1 ,Z2)
p<-ggplot(BM, aes(Z1)) +  geom_density(alpha = 0.25  ,fill = "#69c6ff" )
q<-ggplot(BM, aes(Z2))  +  geom_density(alpha = 0.25   ,fill = "#ff4adb" )
```

```
multiplot(p+theme_bw(),q +theme_bw()+labs(y="") ,cols=2)
```



```r
par(mfrow=c(1,2))
qqnorm(Z1);qqnorm(Z2)
```




**5.** Use R and simulation to 'prove' the CLT: Pick a non-normal discrete or continuous probability model and generate 100 samples, each of size $n = 35$ from this distribution.

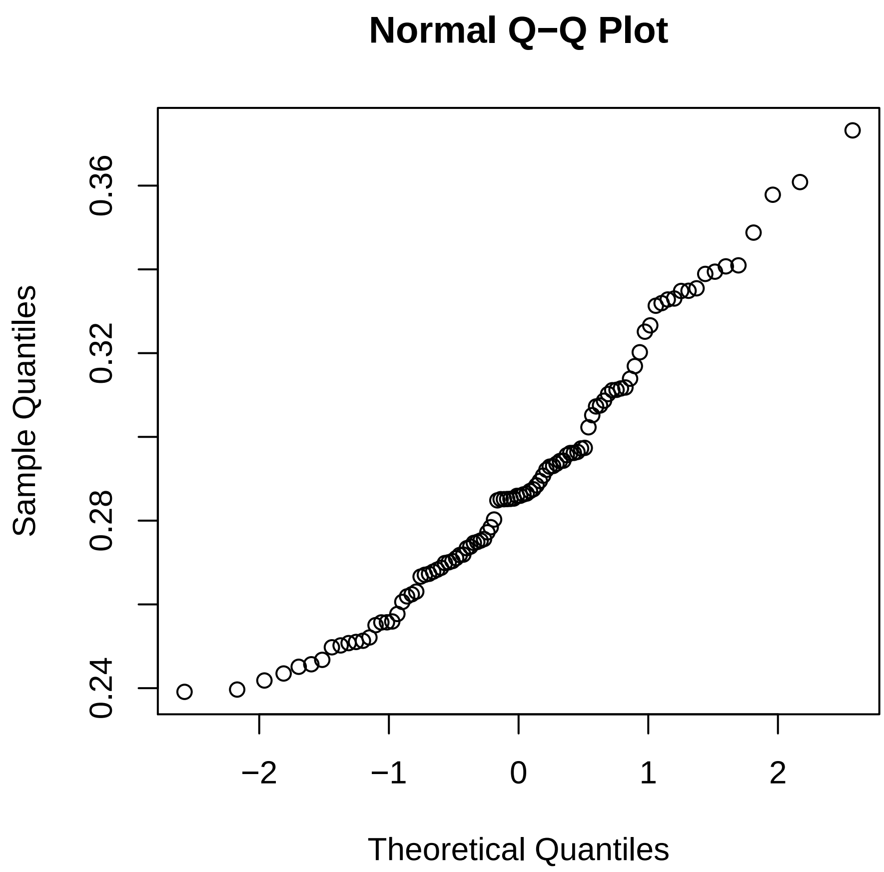
**a.** Derive the sampling distribution of the mean by showing its 5-number summary and standard deviation. Are the mean and standard deviation in agreement with what you would expect from the CLT?

```r
#5 CLT

n=35
CLT<-rep(NA,100)

for( i in 1:100) {      CLT[i]<-mean(   rbeta(n,2,5)    )       }
summary(CLT);sd(CLT) * ((n - 1) / n)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.2391  0.2677  0.2859  0.2891  0.3090  0.3732
## [1] 0.02972143

2/7;sqrt( ( 2*5 )/ ( ( (2+5)^2 )* ( 2+5+1 ) ) ) /sqrt(n)

## [1] 0.2857143
## [1] 0.02699746
```

The mean and standard deviation for a $Beta(2, 5)$ distribution are listed above, 0.2857143 and 0.02699746. They are remarkably close to the mean and standard deviation of the simulation. So yes, they are in agreement with what I would expect from the CLT.

**b.** Create a normal QQ-plot for the sampling distribution of the mean. What do you conclude?

```
qqnorm(CLT)
```




The dots almost line up on a 45° line, signaling that the sampling distribution of the mean is distributed approximately normal. If our sample size or iterations increased, the line would be even closer to 45°.