

The Central Limit Theorem

The Central Limit Theorem provides the explanation for the wide applicability of the normal distribution in significance testing, because it guarantees us that (providing certain assumptions are met) that statistics (like the mean) obtained by integrating (as in summing up for computing the mean) will turn out to be normally distributed, no matter what the distribution of the underlying data is.

The Central Limit Theorem is a mathematical demonstration that statistics (like the mean) that are based on summations or integrations of a set of observations are normally distributed, no matter what the distribution is of the data from which samples are being drawn. The theorem therefore assures us that the normal distribution is the appropriate reference distribution for a sample mean to be compared with, in order to judge the relative size of that sample mean.

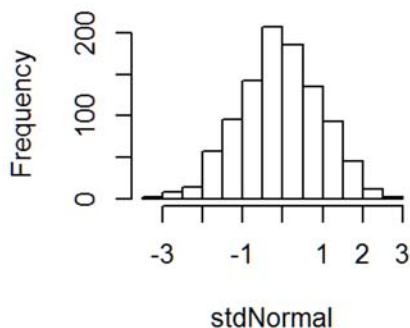
Empirical demonstration of the Central Limit Theorem

Generate 1000 random numbers from four different distributions, the standard normal, the normal, the log-normal and uniform distributions. Histograms of the random numbers will make it clear what the shape of the distributions are like

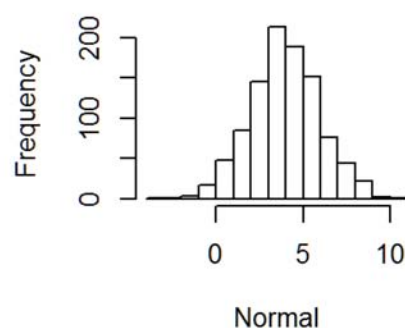
```
n <- 1000
stdNormal <- rnorm(n, mean=0, sd=1)
Normal <- rnorm(n, mean=4.0, sd=2.0)
logNormal <- rlnorm(n, meanlog=1.0, sdlog=0.5)
Uniform <- runif(n, min=0.0, max=1.0)
```

```
opar <- par(mfrow=c(2,2))
hist(stdNormal)
hist(Normal)
hist(logNormal)
hist(Uniform)
```

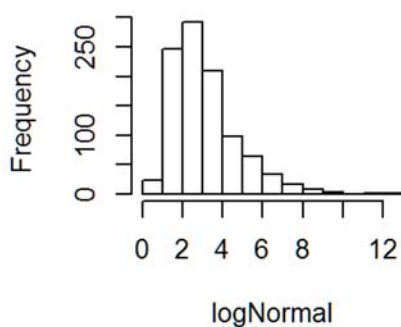
Histogram of stdNormal



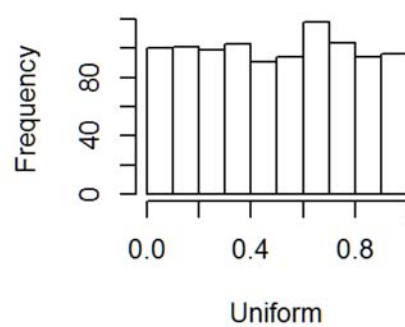
Histogram of Normal



Histogram of logNormal



Histogram of Uniform



```
par <- opar
```

The following script generates nsamp sample means from one of these distributions

```
# take repeated samples from each distribution and calculate and save means
# repeated sampling and calculation of means
nsamp <- 200 # number of samples
mean.stdNormal <- matrix(1:nsamp) # matrix to hold means
mean.Normal <- matrix(1:nsamp) # matrix to hold means
mean.logNormal <- matrix(1:nsamp) # matrix to hold means
mean.Uniform <- matrix(1:nsamp) # matrix to hold means

for (i in 1:nsamp) {
  samp <- sample(stdNormal, 30, replace=T)
  mean.stdNormal[i] <- mean(samp)

  samp <- sample(Normal, 30, replace=T)
  mean.Normal[i] <- mean(samp)

  samp <- sample(logNormal, 30, replace=T)
  mean.logNormal[i] <- mean(samp)

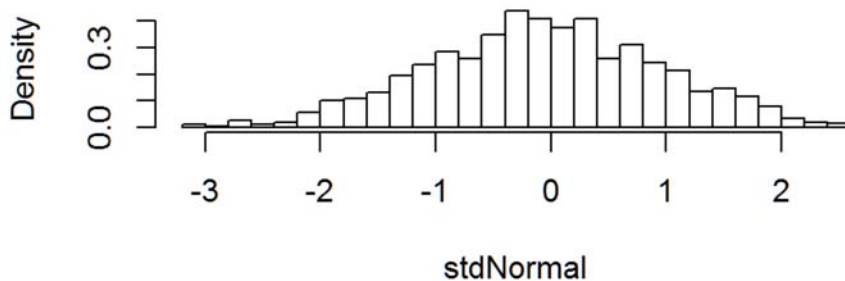
  samp <- sample(Uniform, 30, replace=T)
  mean.Uniform[i] <- mean(samp)
}
```

Histograms of the original data, as well as of the sample means can be obtained with the following script.

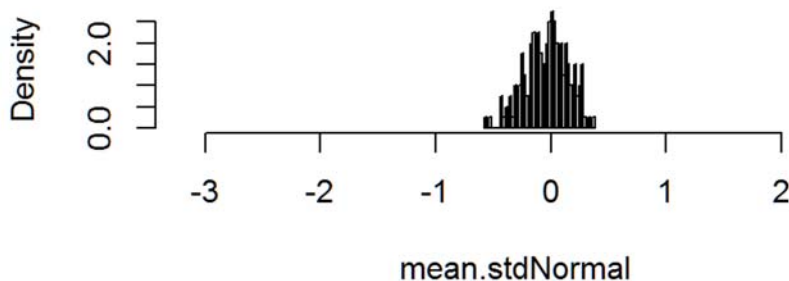
```
# histograms of data and of sample means
par(mfrow=c(2,1))

# standard Normal
xmax <- max(stdNormal)
xmin <- min(stdNormal)
hist(stdNormal, nclass=40, probability=T, xlim=c(xmin,xmax))
hist(mean.stdNormal, nclass=40, probability=T, xlim=c(xmin,xmax))
```

Histogram of stdNormal



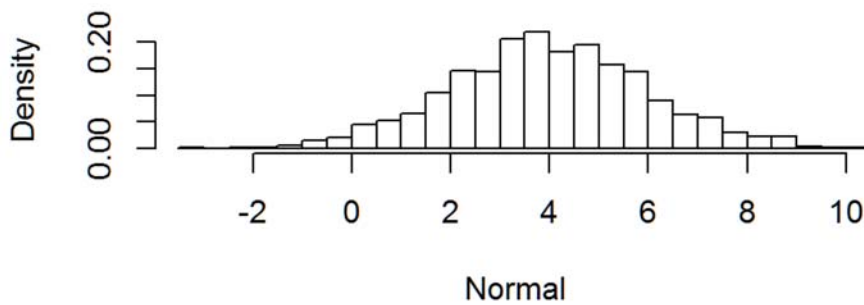
Histogram of mean.stdNormal



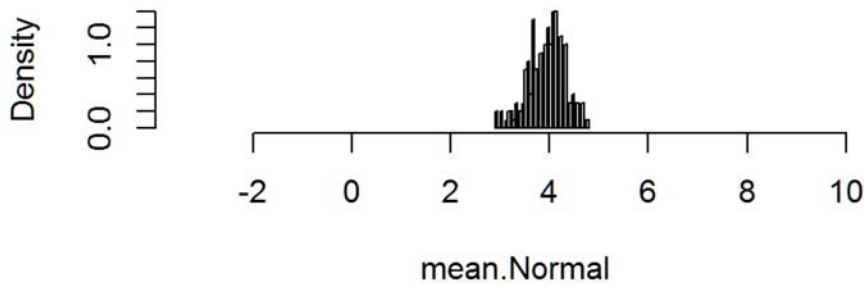
```
# Normal
xmax <- max(Normal)
```

```
xmin <- min(Normal)
hist(Normal, nclass=40, probability=T, xlim=c(xmin,xmax))
hist(mean.Normal, nclass=40, probability=T, xlim=c(xmin,xmax))
```

Histogram of Normal

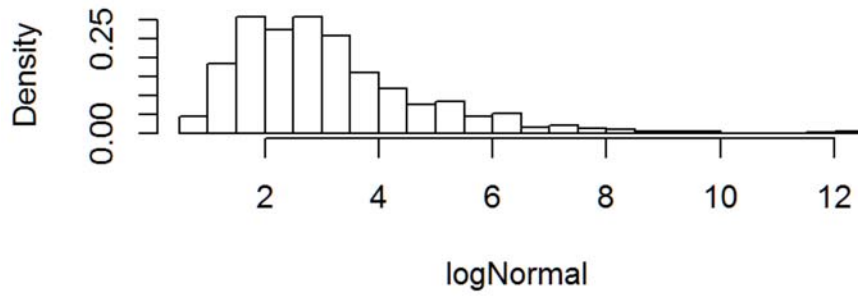


Histogram of mean.Normal

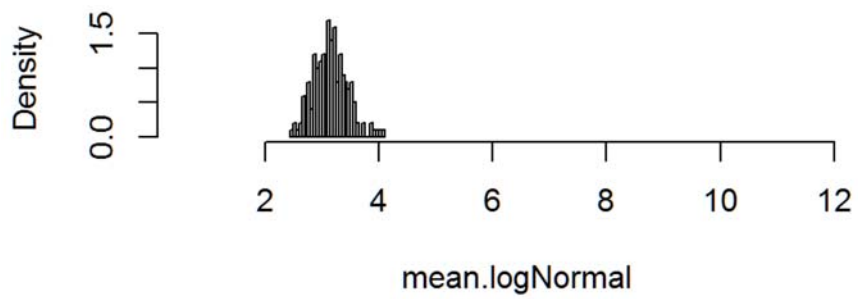


```
# log Normal
xmax <- max(logNormal)
xmin <- min(logNormal)
hist(logNormal, nclass=40, probability=T, xlim=c(xmin,xmax))
hist(mean.logNormal, nclass=40, probability=T, xlim=c(xmin,xmax))
```

Histogram of logNormal

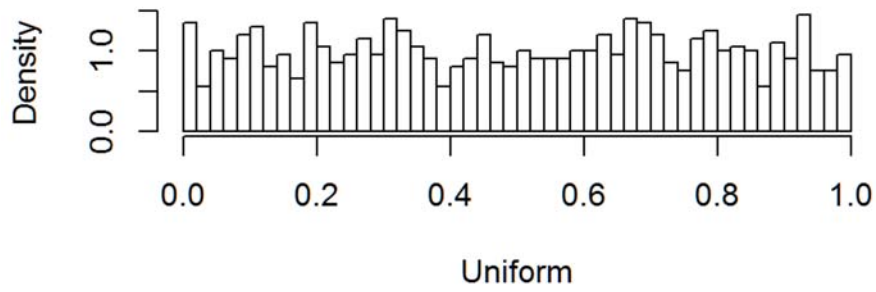


Histogram of mean.logNormal



```
# Uniform
xmax <- max(Uniform)
xmin <- min(Uniform)
hist(Uniform, nclass=40, probability=T, xlim=c(xmin,xmax))
hist(mean.Uniform, nclass=40, probability=T, xlim=c(xmin,xmax))
```

Histogram of Uniform



Histogram of mean.Uniform

