605_HW9.Rmd

Kumudini Bhave

April 5, 2017

COMMON DISTRIBUTIONS AND CENTRAL LIMIT THEOREM

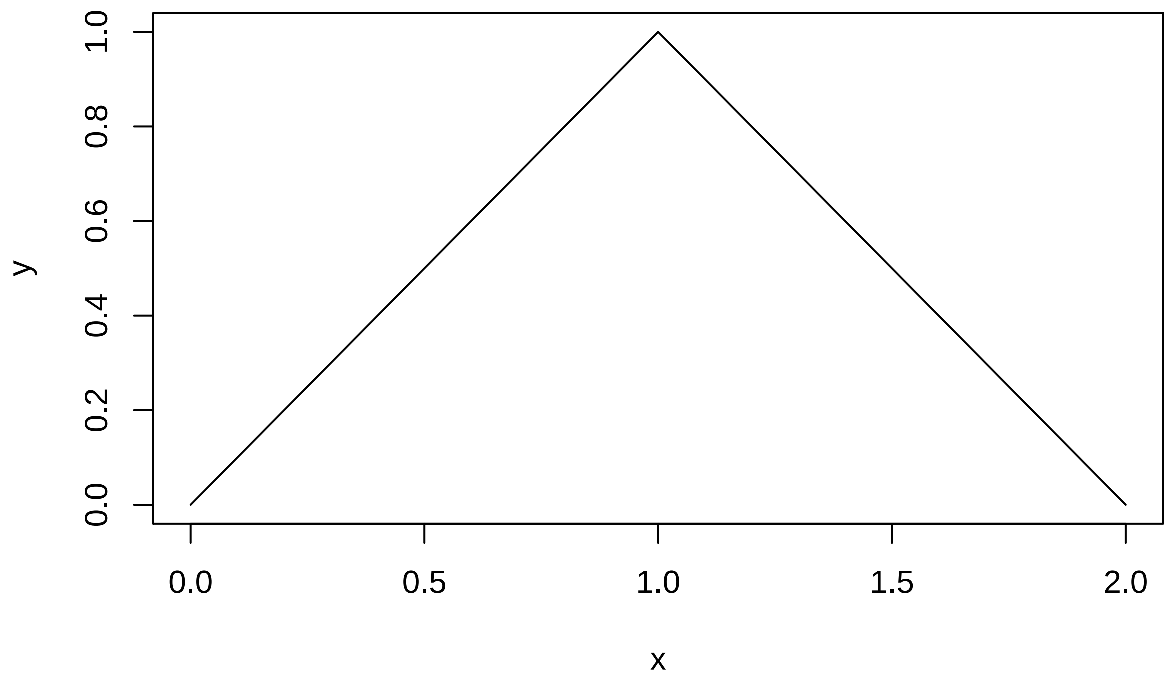
Problem Set 1 :

This week, we'll empirically verify Central Limit Theorem. We'll write code to run a small simulation on some distributions and verify that the results match what we expect from Central Limit Theorem. Please use R markdown to capture all your experiments and code. Please submit your Rmd file with your name as the filename. (1) First write a function that will produce a sample of random variable that is distributed as follows: f(x) = x; 0 <=x <= 1 f(x) = 2 - x; 1 < x <=2 That is, when your function is called, it will return a random variable between 0 and 2 that is distributed according to the above PDF. Please note that this is not the same as writing a function and sampling uniformly from it.

Solution :

```
knitr::opts_chunk$set(message = FALSE, echo = TRUE)

# Consider function f(x) = x; 0 <=x <= 1 Since, y1 = f(x1) <- x1

f_x1 <- function(x) {
    x
}

# Consider function f(x) = 2 - x; 1 < x <=2 Since, y2 = f(x2) <- 2 - x2

f_x2 <- function(x) {
    2 - x
}




# Plotting the set of equations

x1 <- seq(0, 1, 0.01)
y1 <- f_x1(x1)
x2 <- seq(1, 2, 0.01)
y2 <- f_x2(x2)

x <- c(x1, x2)
```

```r
y <- c(y1, y2)

plot(x, y, "l", main = NULL)
```




```r
f1x <- function(x) {
    if (x >= 0 && x < 1) {
        y <- x
        return(y)

    } else if (x > 1 && x <= 2) {
        y <- 2 - x
        return(y)
    }
}


f1x(1.02)
```

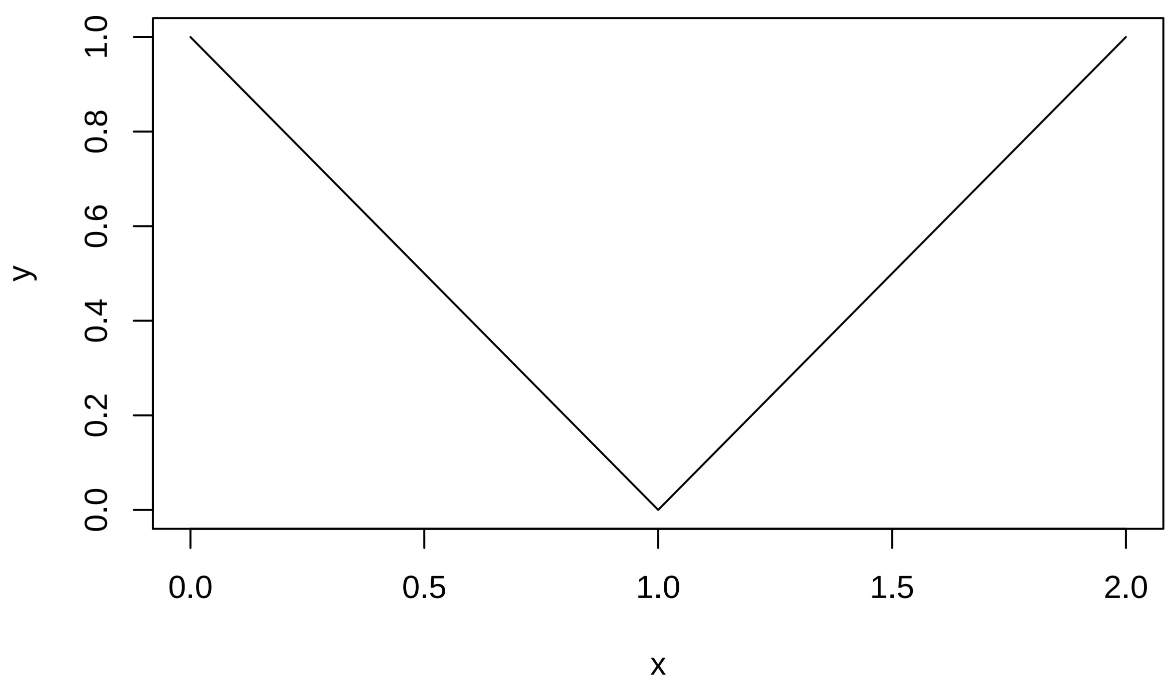
```
FALSE [1] 0.98
```

---

2 Now, write a function that will produce a sample of random variable that is distributed as follows: f(x) = 1 - x; 0 <= x <= 1 f(x) = x - 1; 1 < x <= 2

Solution :

```
knitr::opts_chunk$set(message = FALSE, echo = TRUE)


y1 <- 1 - x1
y2 <- x2 - 1


x <- c(x1, x2)
y <- c(y1, y2)

plot(x, y, "l", main = NULL)
```



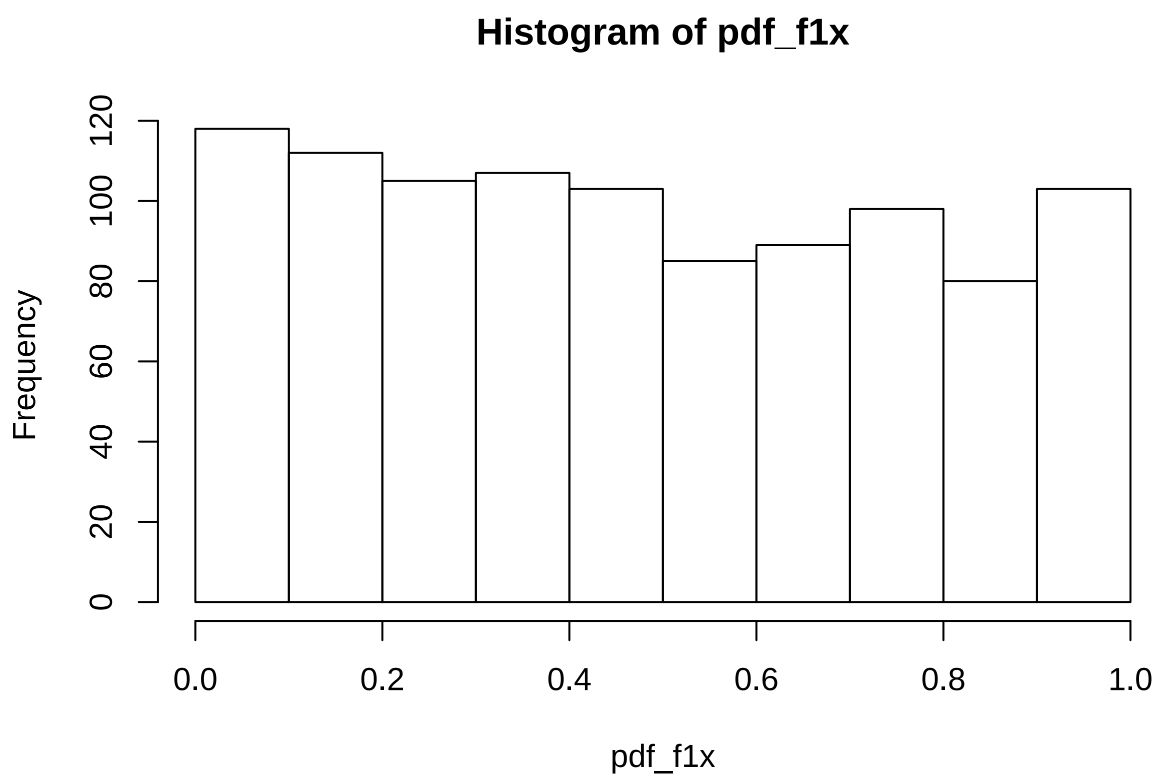

```
f2x <- function(x) {
    if (x >= 0 && x < 1) {
        y <- 1 - x
        return(y)

    } else if (x > 1 && x <= 2) {
        y <- x - 1
        return(y)
    }
```

```
}
```

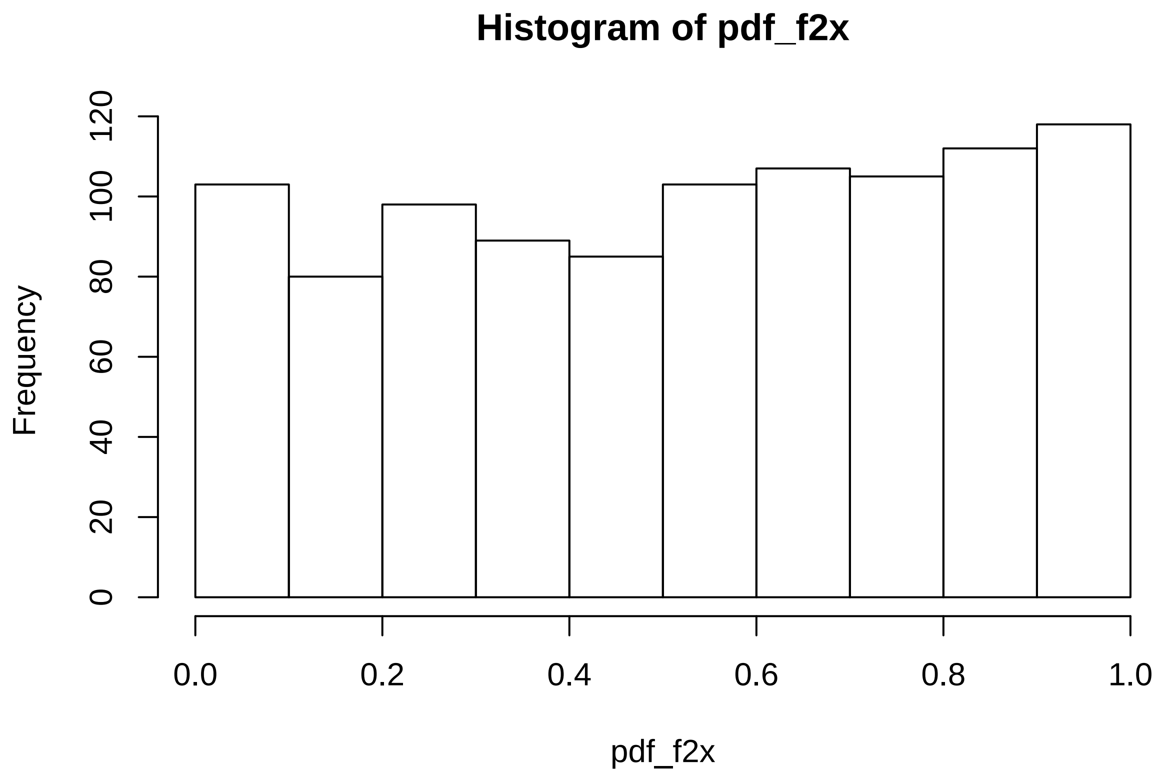
---

3 Draw 1000 samples (call your function 1000 times each) from each of the above two distributions and plot the resulting histograms. You should have one histogram for each PDF. See that it matches your understanding of these PDFs.

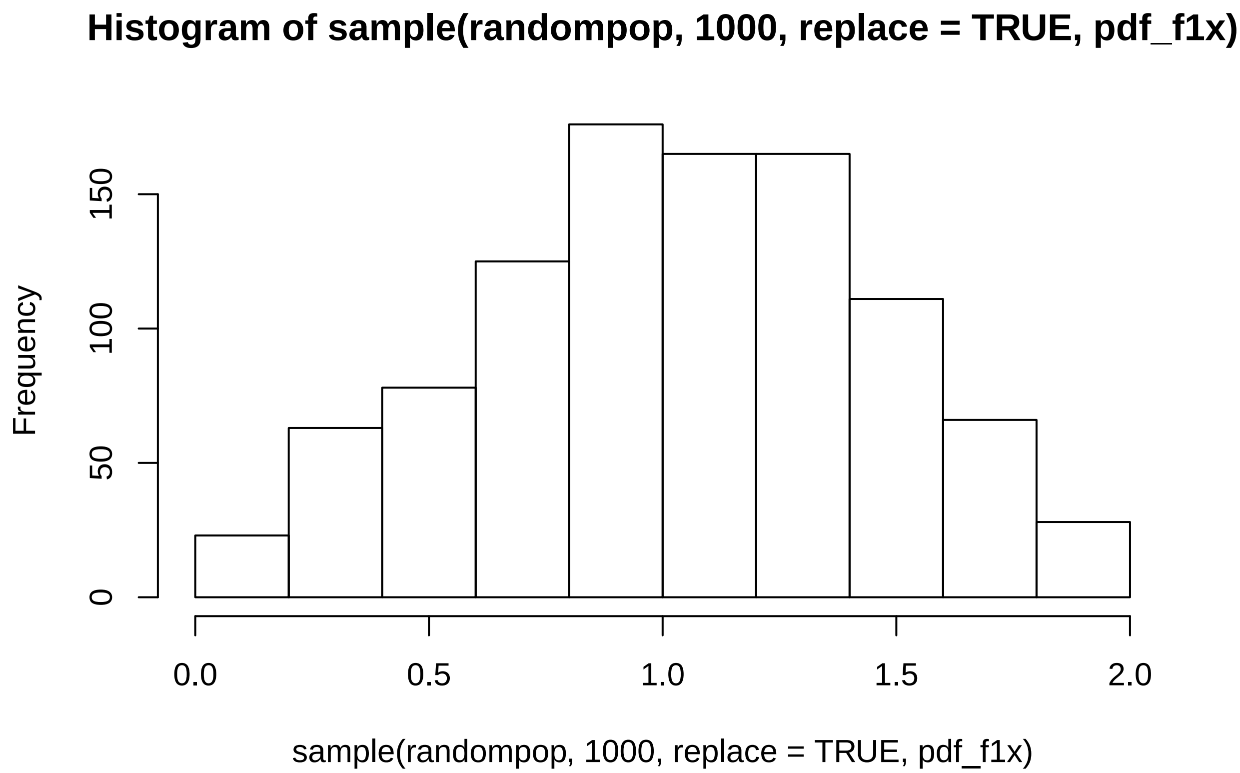
```
randompop <- runif(1000,0,2)

pdf_f1x <- sapply(randompop, f1x)

pdf_f2x <- sapply(randompop, f2x)

#histogram from pdf for first function
hist(pdf_f1x)
```



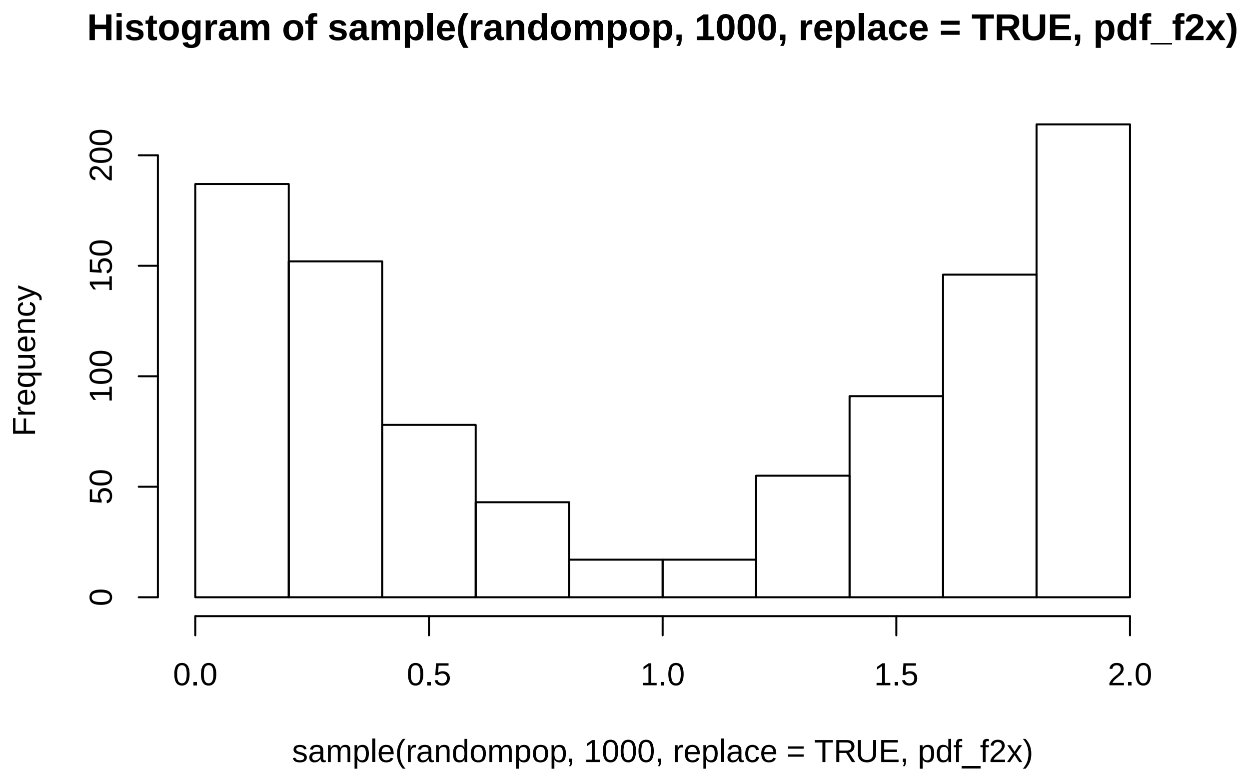

```
#histogram from pdf for second function
hist(pdf_f2x)
```




```r
#Plotting the probabilities does not yield a congruent plot

# The PDF probabilities do not match those in the the Part 1 and Part 2 functions plots

# hecne we sample from the functions using the probabilities
hist(sample(randompop,1000, replace=TRUE,pdf_f1x))
```




```r
hist(sample(randompop,1000, replace=TRUE,pdf_f2x))
```



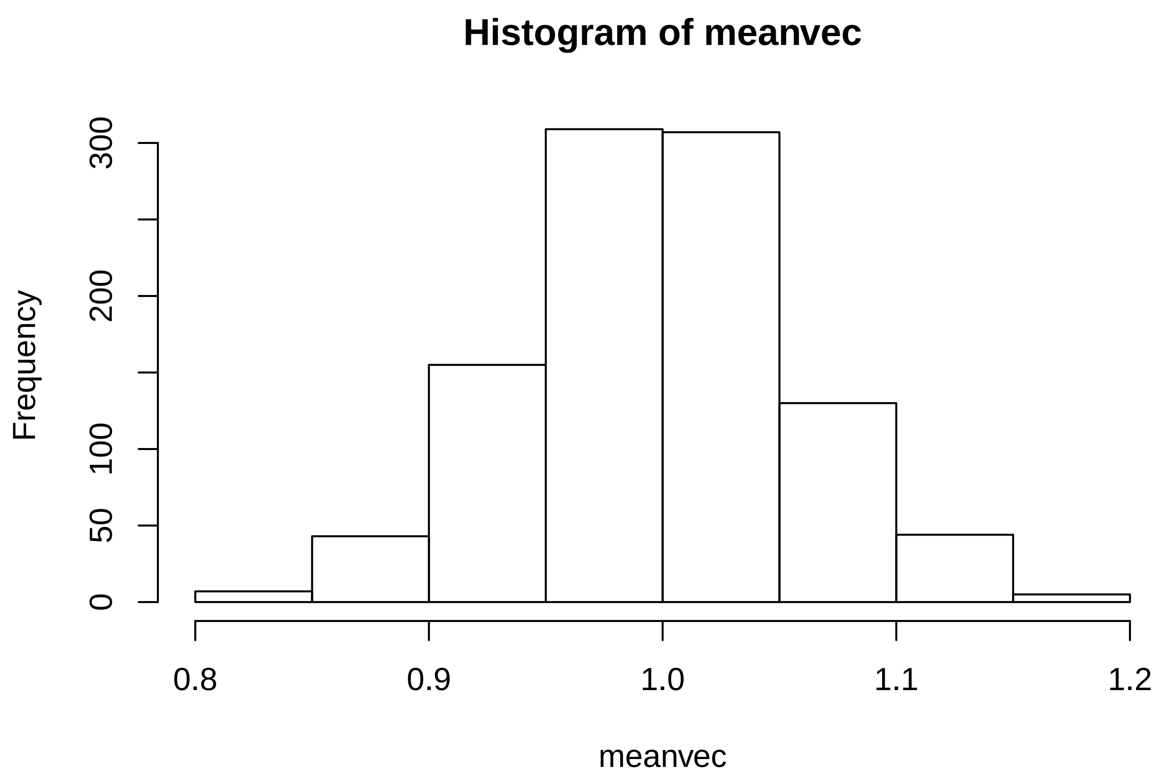

```
##########################################################
```

---

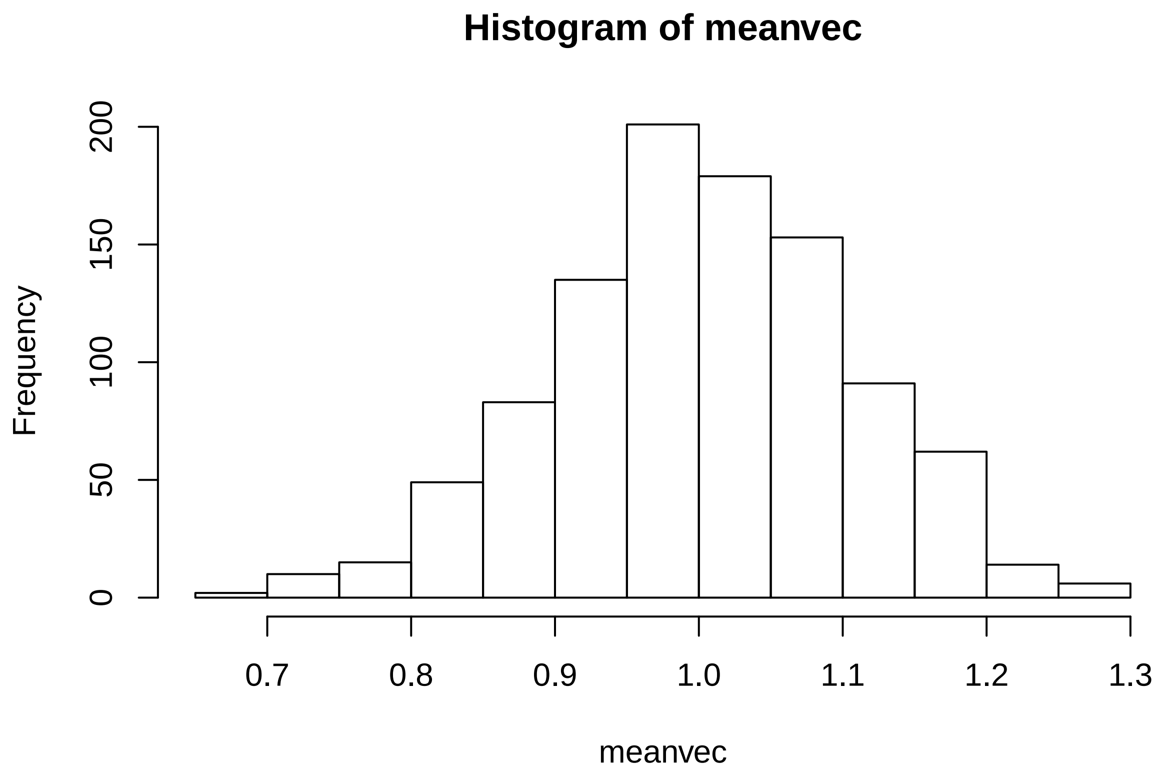
4 Now, write a program that will take a sample set size n as a parameter and the PDF as the second parameter, and perform 1000 iterations where it samples from the PDF, each time taking n samples and computes the mean of these n samples. It then plots a histogram of these 1000 means that it computes.

```
calc_CLT <- function(sampsize, pdf_Func )
{
     meanvec <- NULL
     # for thousand iterations
     for(i in 1 : 1000)
     {
          randomx <- runif(1000,0,2)

          randomy <- sapply(randomx, pdf_Func)

          # get the samples from the PDF
          samplevec <- sample(randomx, sampsize, replace=TRUE, randomy)


          # calculate the mean and store
          meanvec <- c(meanvec,mean(samplevec))
```

```r
    }

    return(hist(meanvec))

}


# Testing calc_CLT for the PDFs


calc_CLT(50, f1x) # for PDF function 1
```




```r
calc_CLT(50, f2x) # for PDF function 2
```

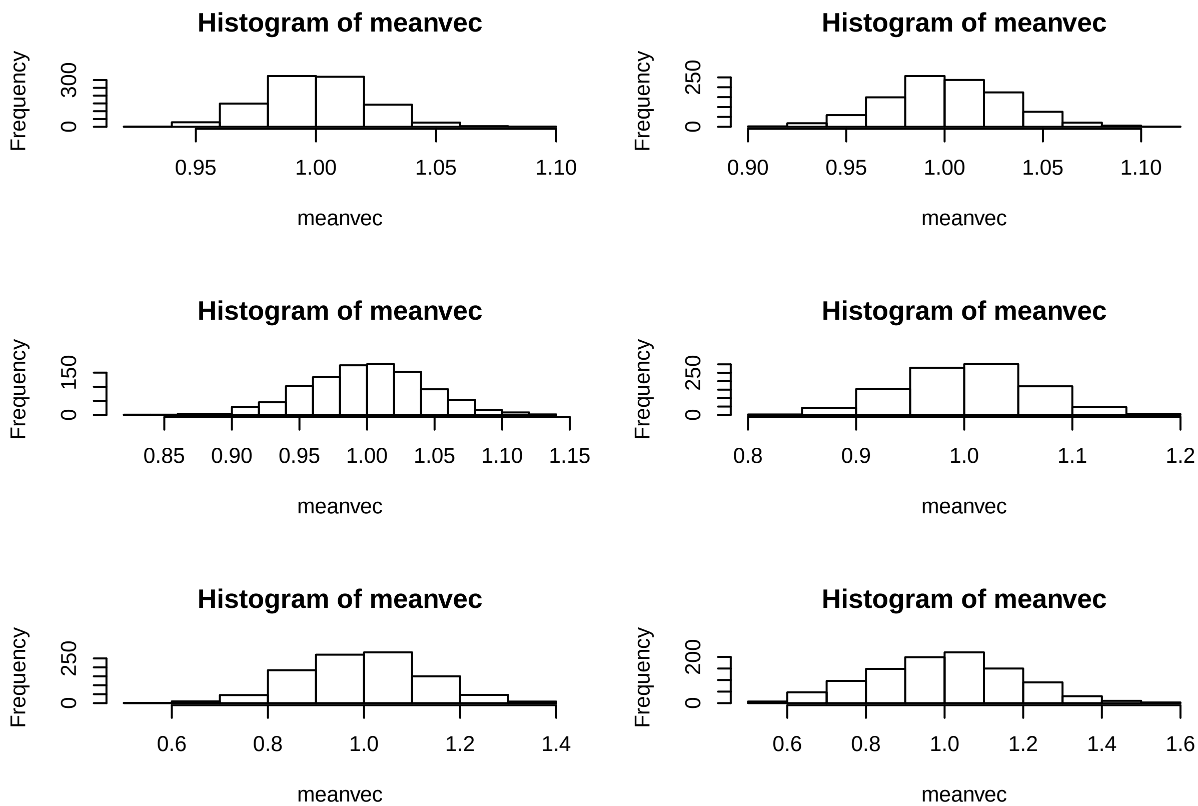



```
# We observe normal distribution for the sample mean of the population
```

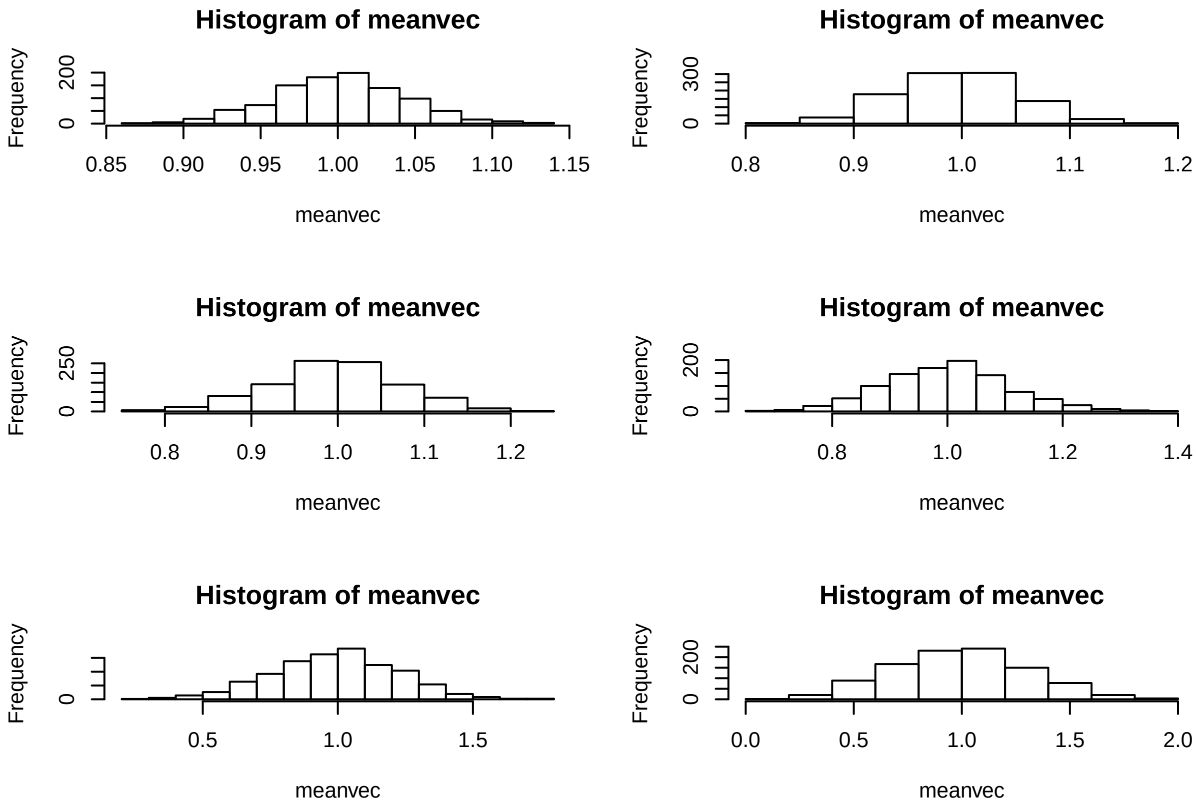
---

5 Verify that as you set n to something like 10 or 20, each of the two PDFs produce normally distributed mean of samples, empirically verifying the Central Limit Theorem. Please play around with various values of n and you'll see that even for reasonably small sample sizes such as 10, Central Limit Theorem holds.

```
# Verifying the Central Limit Theorem by calling the Probability Density Functions for different sample

# Verifying for the first PDF f1x()
# f(x) = x;      0 <=x <= 1
# f(x) = 2 - x;  1 < x <=2


par(mfrow = c(3,2))

calc_CLT(500, f1x)
calc_CLT(200, f1x)
calc_CLT(100, f1x)
calc_CLT(50, f1x)
calc_CLT(10, f1x)
calc_CLT(5, f1x)
```




```r
# Verifying for the first PDF f2x()
# f(x) = 1 - x; 0 <= x <= 1
# f(x) = x - 1; 1 < x <= 2

#par(mfrow = c(3,2))

calc_CLT(500, f2x)
calc_CLT(200, f2x)
calc_CLT(100, f2x)
calc_CLT(50, f2x)
calc_CLT(10, f2x)
calc_CLT(5, f2x)
```



We observe that the Central Limit Theorem holds good for any size of samples , and i.e. that the mean of the samples from the population will be approximately equal to the mean of the population. And the distributions of the sample means is a normal distribution