BLin_Assign9

Bin Lin

2017-4-2

1. First write a function that will produce a sample of random variable that is distributed as follows:

$$f(x) = x, \quad 0 \le x \le 1$$

$$f(x) = 2 - x, \quad 1 \le x \le 2$$

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

```
random_variable1 <- function()
{
  x <- runif(1, 0, 2)
  if (x <= 1)
  {
    return (x)
  }
  else
  {
    return (2 - x)
  }
}

random_variable1()
```

```
## [1] 0.9233496
```

2. Now, write a function that will produce a sample of random variable that is distributed as follows:

$$f(x) = 1 - x, \quad 0 \le x \le 1$$

$$f(x) = x - 1, \quad 1 \le x \le 2$$

```
random_variable2 <- function()
{
  x <- runif(1, 0, 2)
  if (x <= 1)
  {
    return (1 - x)
  }
  else
  {
    return (x - 1)
  }
}

random_variable2()
```

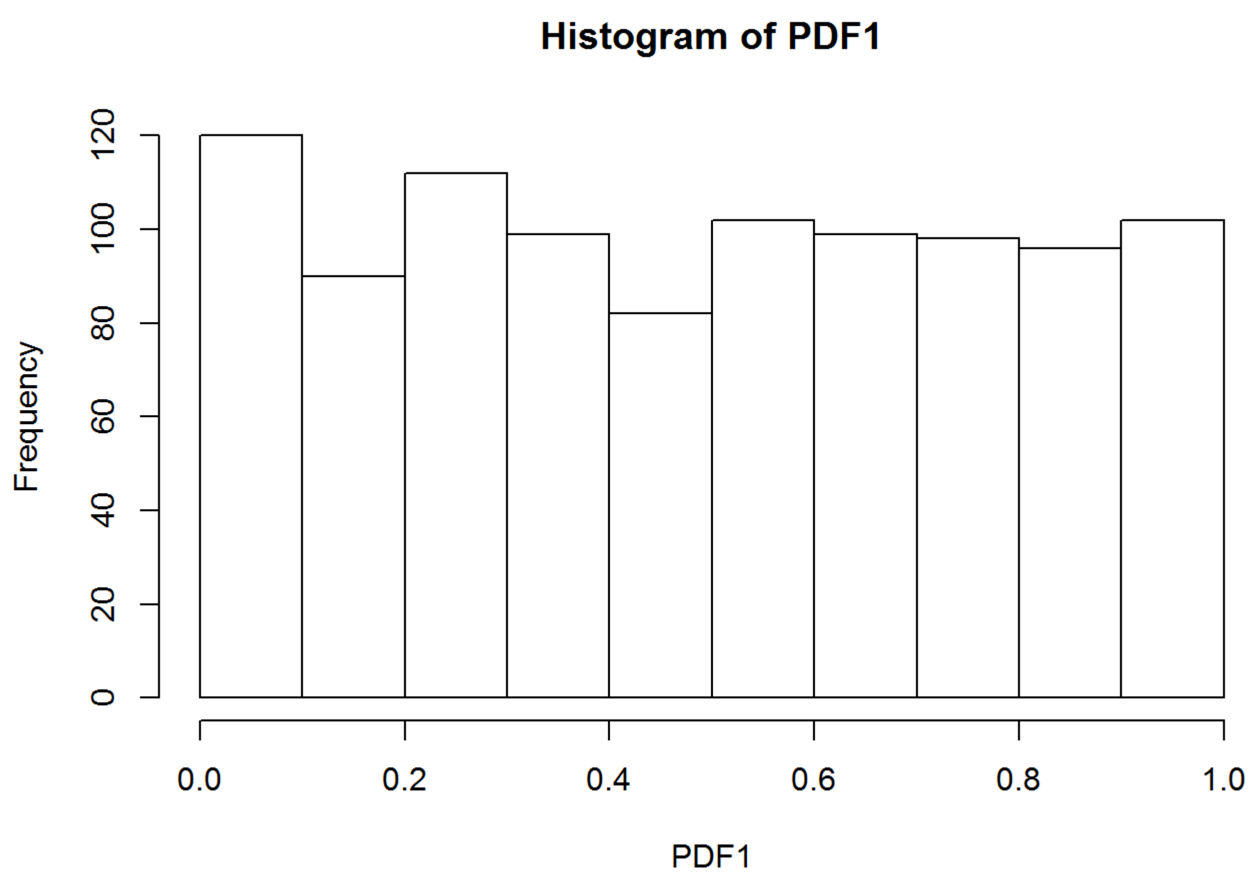
```
## [1] 0.718818
```

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.

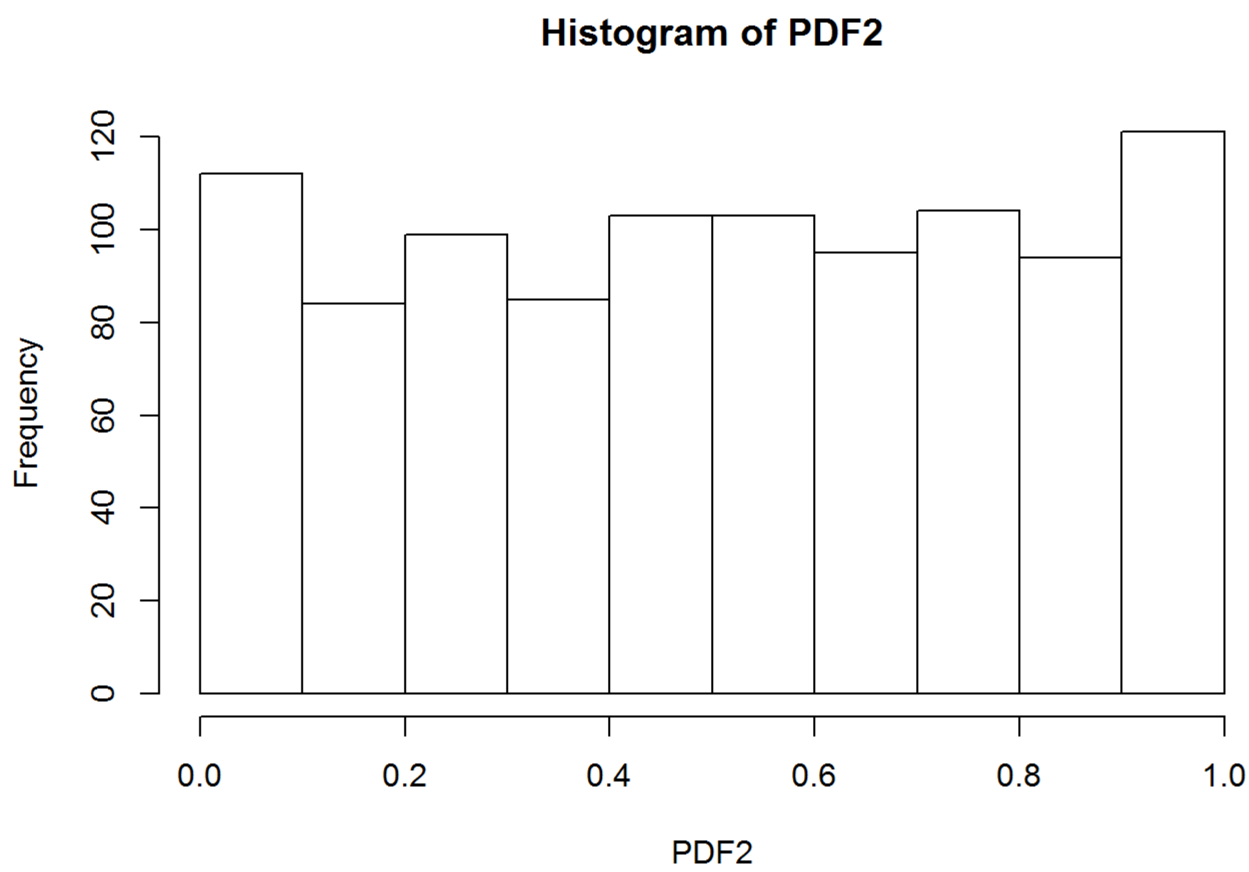
```
PDF1 <- vector()
PDF2 <- vector()

for (i in 1:1000)
{
  PDF1 <- append(PDF1, random_variable1())
  PDF2 <- append(PDF2, random_variable2())
}

hist(PDF1)
```




```
hist(PDF2)
```

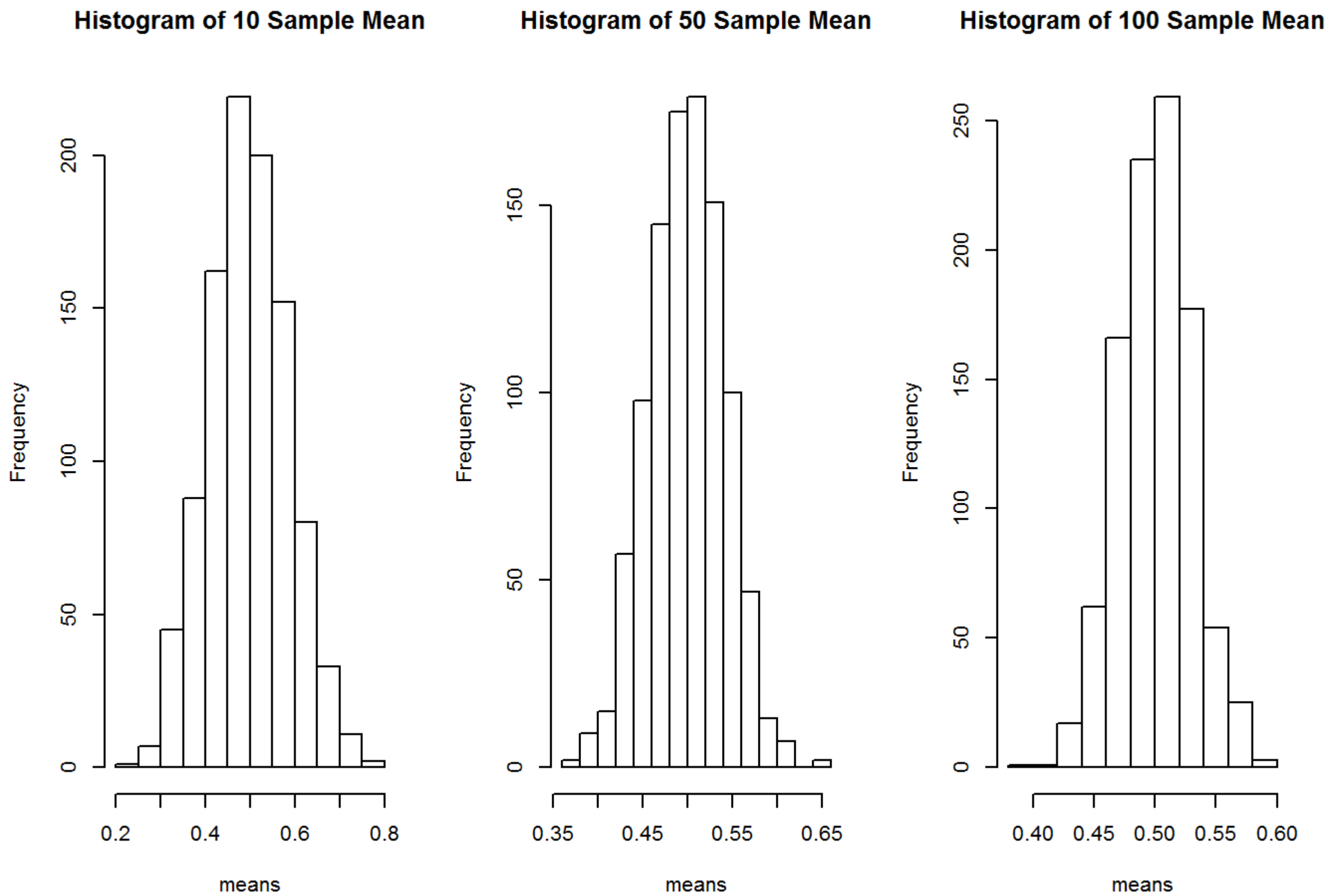


4. Now, write a program that will take a sample set size n as a parameter and the PDF as the second parameter, and perform 1000 iterations where it samples from the PDF, each time taking n samples and computes the mean of these n samples. It then plots a histogram of these 1000 means that it computes.

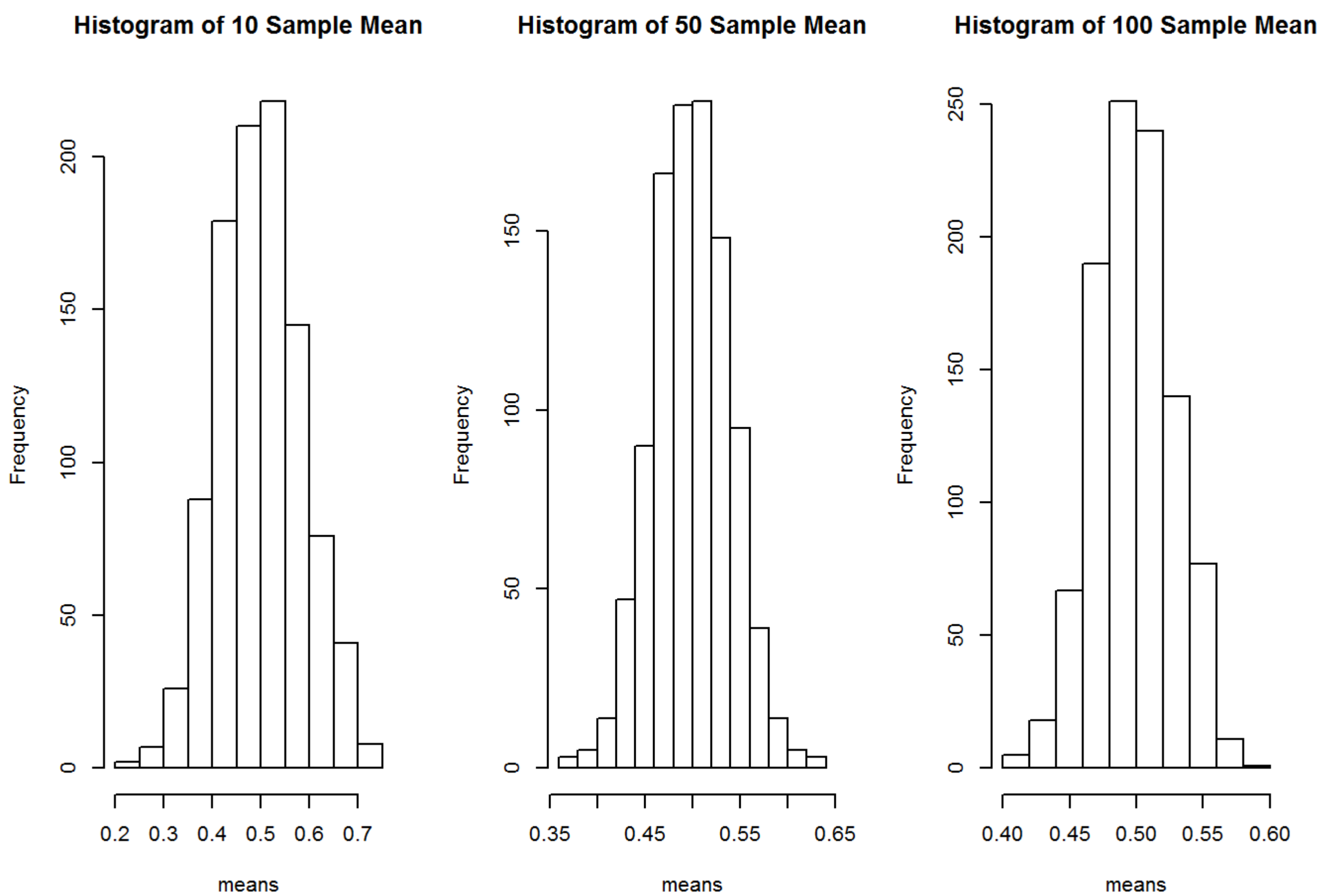
```
sampling_distribution <- function(n, fx)
{
  means <- numeric(1000)
  for (i in 1:1000)
  {
    PDF <- vector()
    for (j in 1:n)
    {
      PDF <- append(PDF, fx())
    }
    means[i] <- mean(PDF)
  }
  hist(means, main = (paste("Histogram of",  toString(n), "Sample Mean")))
}
```

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.

```
par(mfrow = c(1,3))
sampling_distribution(10, random_variable1)
sampling_distribution(50, random_variable1)
sampling_distribution(100, random_variable1)
```




```
sampling_distribution(10, random_variable2)
sampling_distribution(50, random_variable2)
sampling_distribution(100, random_variable2)
```


Histogram of 10 Sample Mean
Frequency
0
50
100
150
200
0.2 0.3 0.4 0.5 0.6 0.7
means
Histogram of 50 Sample Mean
Frequency
0
50
100
150
0.35 0.45 0.55 0.65
means
Histogram of 100 Sample Mean
Frequency
0
50
100
150
200
250
0.40 0.45 0.50 0.55 0.60
means