

CFerrari_Assignment9

Charley Ferrari

Friday, March 27, 2015

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:

$$f(x) = x, 0 \leq x \leq 1$$
$$f(x) = 2, -x1 < x \leq 2$$

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. In the online session this week, I'll cover Sampling techniques. You will find it useful when you do the assignment for this week. In addition, as usual, there are one-liners in R that will give you samples from a function. We'll cover both of these approaches in the online session.

```
library(ggplot2)

invcdf1 <- function(y){
  if (y >= 0 && y <= 2){
    ret <- ifelse(y < 0.5, sqrt(2*y), 2-sqrt(2*(1-y)))
  }
}

pdfcreator <- function(invcdf1){
  return(sapply(runif(1000), invcdf1))
}

sampleeq1 <- pdfcreator(invcdf1)
```

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

$$f(x) = 1 - x, 0 \leq x \leq 1$$
$$f(x) = x - 1, 1 < x \leq 2$$

```
#sqrt(2*x-4)+2

#2 - sqrt(4-2*y)

invcdf2 <- function(y){
```

```

if(y >= 0 && y <= 2){
  ret <- ifelse(y < 0.5, 1-sqrt(1-2*y), 1+sqrt(2*y-1))
}
}

```

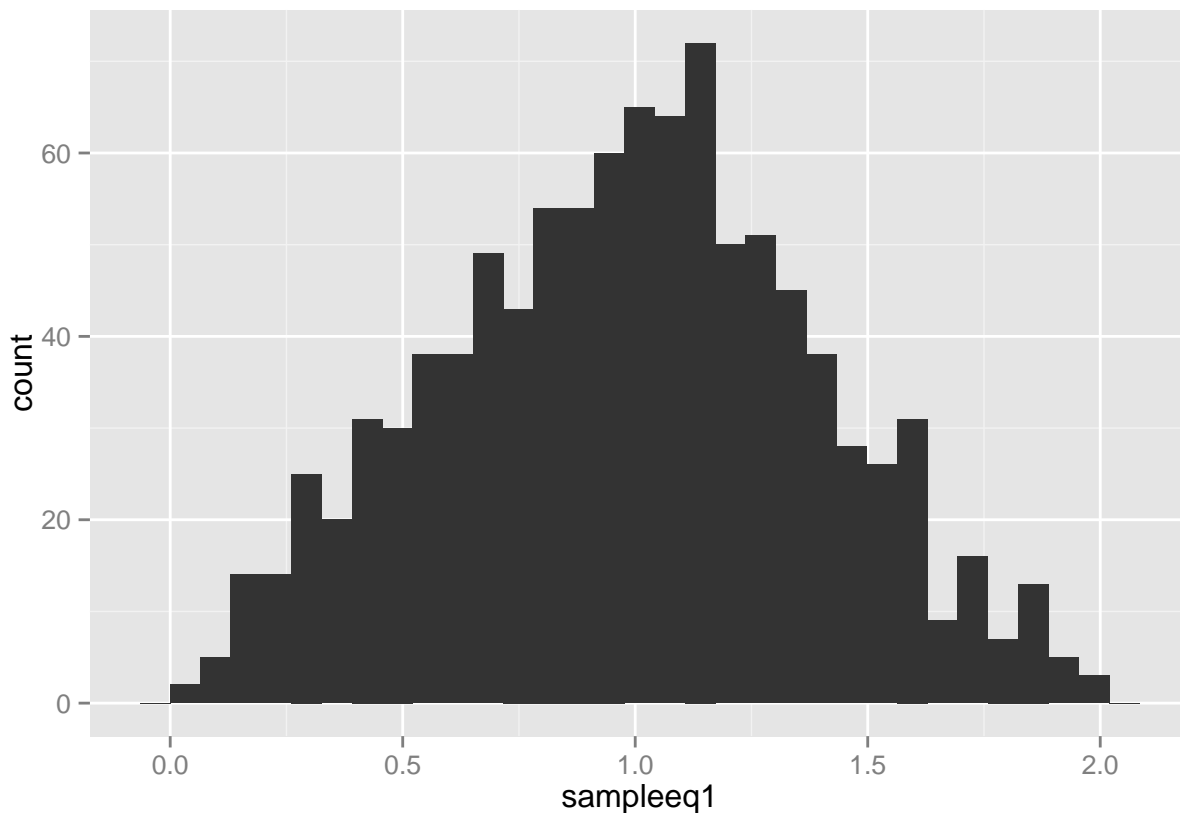
```
sampleeq2 <- pdfcreator(invcd2)
```

- (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.

The function call is done above, below are the qplots

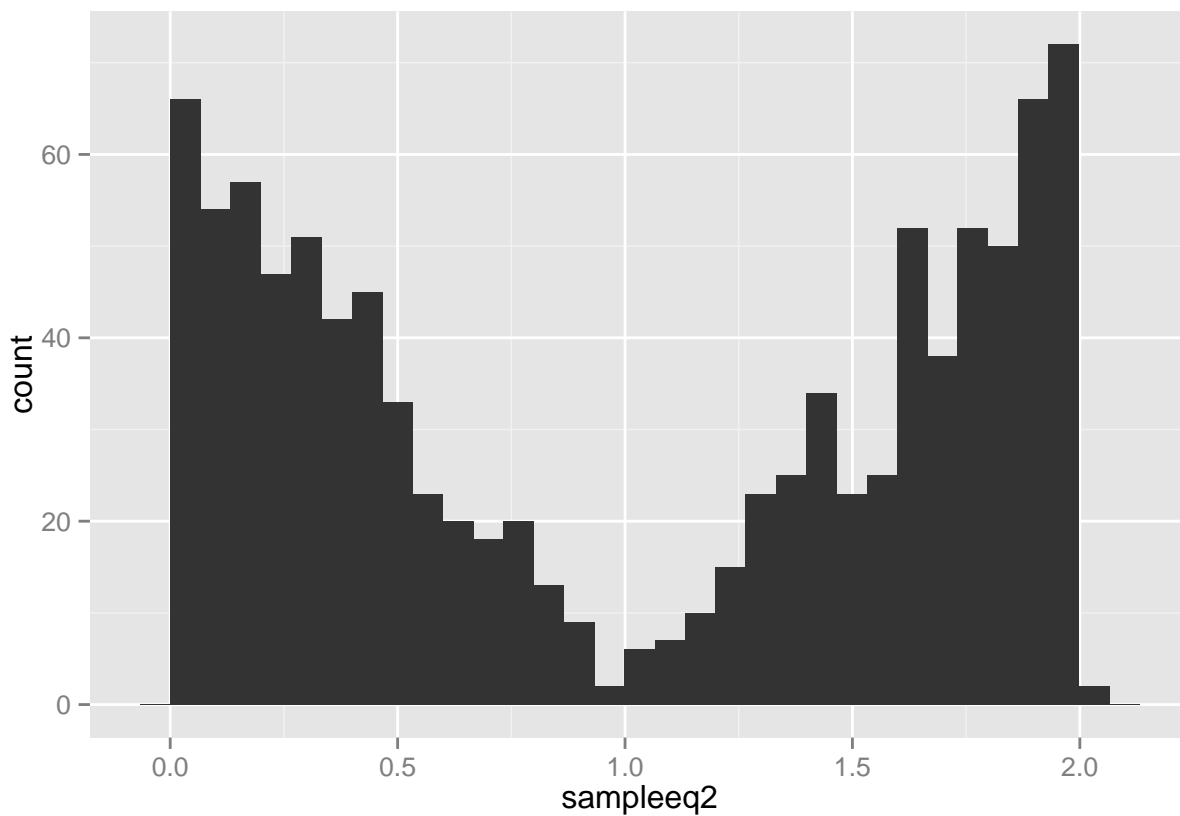
```
qplot(sampleeq1)
```

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```



```
qplot(sampleeq2)
```

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```

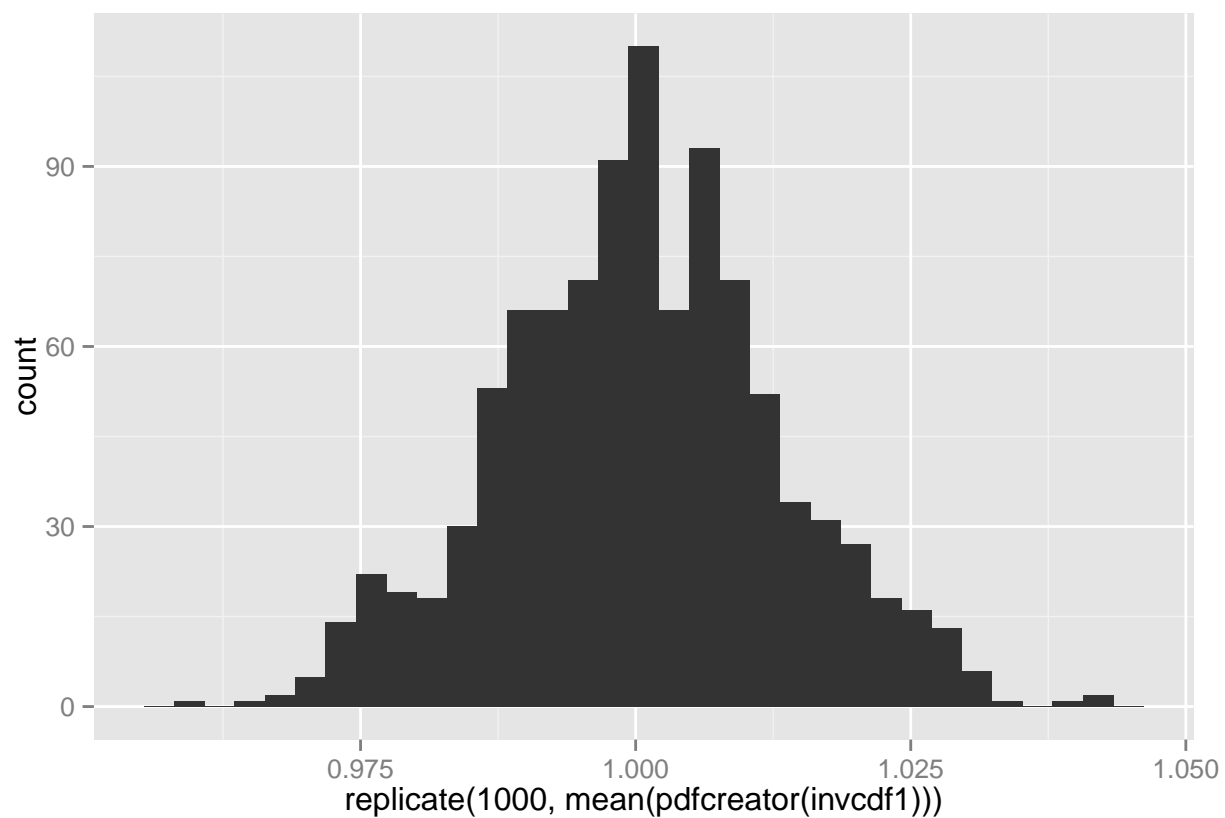


- (4) Now, write a program that will take a sample set size n as a parameter and the PDF as the second parameter, and perform 1000 iterations where it samples from the PDF, each time taking n samples and computes the mean of these n samples. It then plots a histogram of these 1000 means that it computes

```
qplot(replicate(1000,mean(pdfcreator(invcdf1))))
```

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```

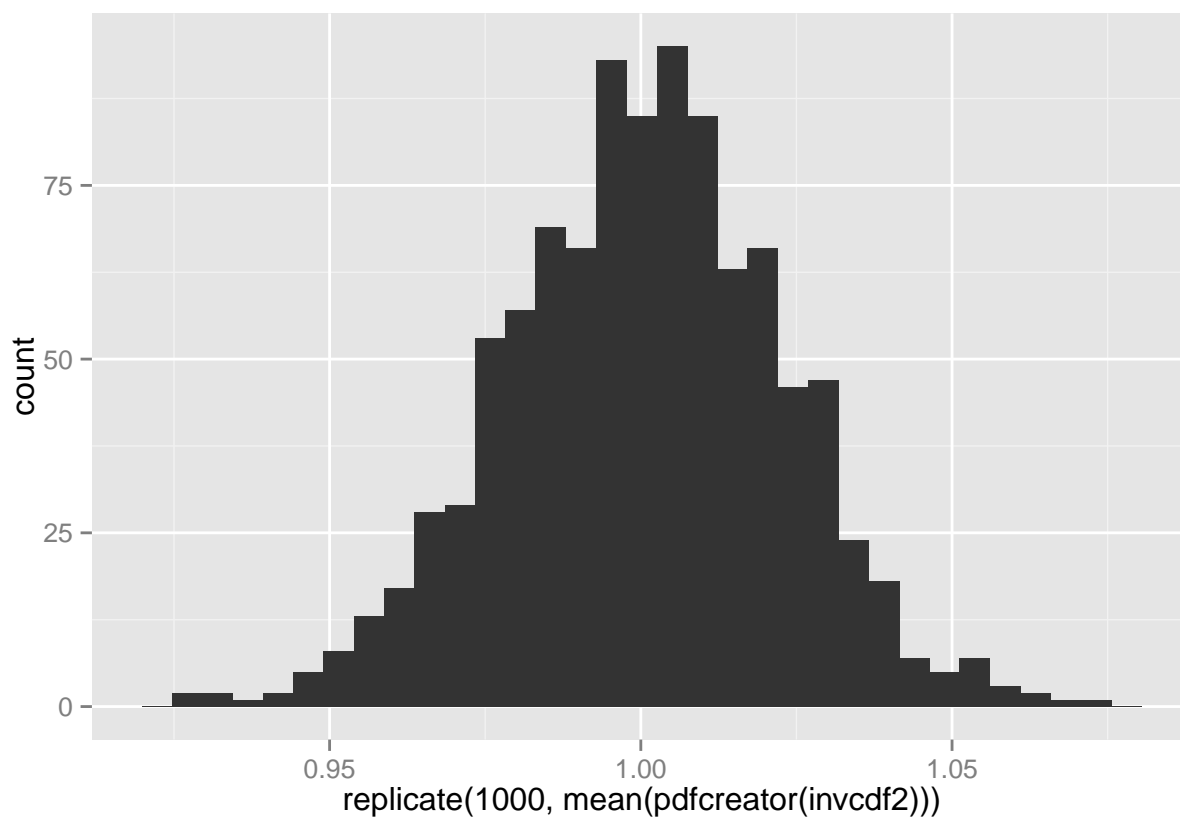
```
## Warning: position_stack requires constant width: output may be incorrect
```



```
qplot(replicate(1000,mean(pdfcreator(invcd2))))
```

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```

```
## Warning: position_stack requires constant width: output may be incorrect
```



- (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.