Week 7 Lab problems

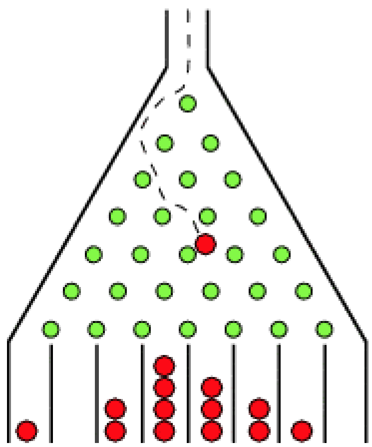
EEB 429
Bhaskar Kumawat

Submit all your answers in a single .Rmd file (along with the knitted html).

A. What is a simulation?

Simulation: An imitation of a process (generally on the computer).

Example process: The *quincunx*



The process generates a *distribution* of beads at the bottom of the structure. This distribution is a result of the stochastic process of the beads making their way from the top to the bottom. Surprisingly, the quincunx process gives us a ***normal distribution***.

To find the structure of this distribution, we can do one of three things:

1. **Simulate:** Simulate the process itself and see what it gives us. On a computer, you would code up a physics simulation that models the entire process of the beads making their way down the board.

2. **Sample:** Derive the structure of the distribution mathematically, and then draw random numbers from this distribution. In the case of the quincunx, a mathematician would derive the analytical form of the distribution (in this case a normal distribution). Then, you can draw random samples from this distribution and plot a histogram of the results.

3. **Plot pdf**: Just plot the analytical form of the distribution directly. In this case, it would be akin to plotting the pdf of the normal distribution.

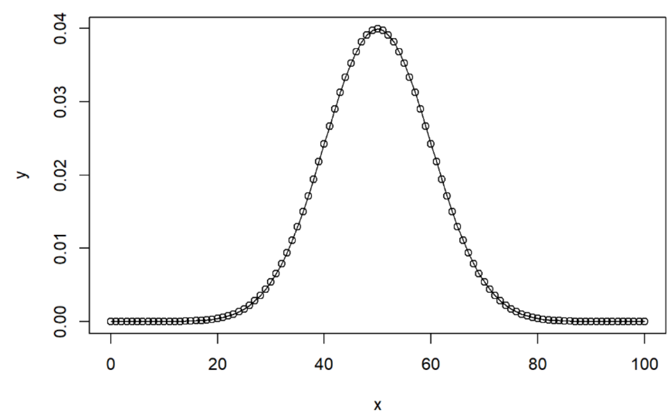
The next pages show three plots generated using the three different ways of finding the structure of a Normal distribution (along with the code). Match each of those plots to one of the three methods described above.

(Hint 1: You can look at the help pages of the functions used to generate the plots.)

(Hint 2: A normal distribution can be obtained by drawing a large sample from any distribution and taking the mean of the sample. The means obtained this way will be normally distributed!)

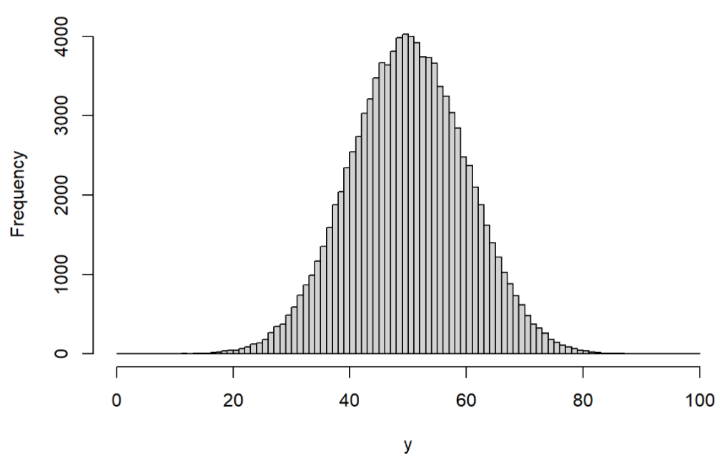
A

```r
x <- seq(0,100,1)
y <- dnorm(x,mean=50,sd=10)

plot(x,y)
lines(x,y)
```



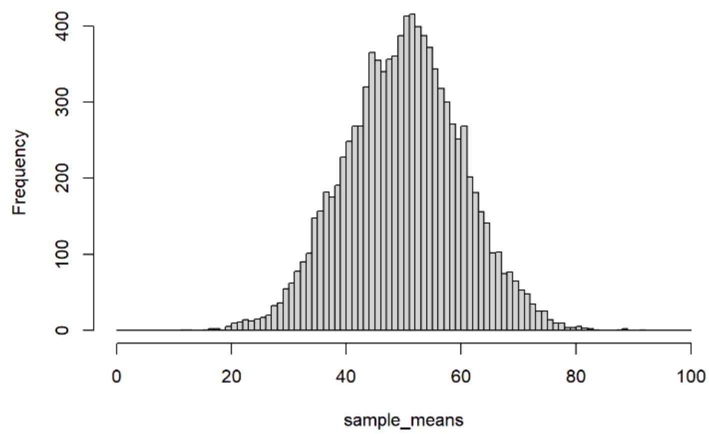
B

```r
n <- 100000
y <- rnorm(n,mean=50,sd=10)


hist(y,breaks=seq(0,100,1),main=NULL)
```



C

```
n_samples <- 10000

sample_means <- rep(0,n_samples)

for(i in 1:n_samples){
  sample <- runif(1000,-500,600)
  smean <- mean(sample)
  sample_means[i] <- smean
}

hist(sample_means,breaks=seq(0,100,1),main=NULL)
```



For full credit submit your matchings (eg 4-K, 5-M etc.) as a part of the Rmd/html file.

B. Sample mean and sample variance

The sample mean and sample variance for a set of values $\{x_1, x_2, x_3,..., x_N\}$ is defined as,

$$\mu' = \sum_{i=1}^{N} x_i / N$$

$$V' = \sum_{i=1}^{N} (x_i - \mu')^2 / (N - 1)$$

1. Write two functions in R: A function that calculates the mean (call it s_mean), and a function that calculates the variance from a given vector of values (call it s_var). You should NOT use the mean or variance function in R to do this. Instead, directly use the formulae above.

2. Now, write a function that,

    a. Takes in a vector of integer values,

    b. For each value in the vector (call the value n),

i. Samples n numbers from a normal distribution with mean 20 and variance 9. (See help for function rnorm)

ii. Uses your functions s_mean and s_var to calculate the the mean and variance of this sample of values

c. Plots all the input values of n against the corresponding sample means. (along with a horizontal line at the *real* mean value of 20)

d. Plots all the input values of n against the corresponding sample variances. (along with a horizontal line at the *real* variance value of 9)

3. Using the above function, find out:

   a. Approximately how many samples do you need to be within 1% of the real mean?

   b. Approximately how many samples do you need to be within 1% of the real variance?

   (Hint: You'll have to play around with the functions and eyeball the graphs to figure this one out)

For full credit, submit your functions for part 1 and 2, example plots for part 2, reasoning for part 3, and plots showing that your derived number of samples actually gives the real mean and variance to within 1% accuracy.