MATH 4753 Laboratory 8

Sampling Distributions and the CLT

In this lab we will take what we have learnt already concerning sampling distributions and include ideas pertaining to the central limit theorem (CLT). We will cover the following:

1. Sample from any distribution using r ----(), example rbinom(), rpois(), runif(), etc (see ?distributions in R)
2. How to create a statistic, usually the *sum* or the *mean*.
3. How to store the statistic.
4. How to repeat the procedure for a designated number of iterations.
5. When finished learn how to create a histogram of the statistic with other graphs.

The method for doing this will be to use a ready-made R script, adapt it and re-run it for the problems given below.

# Objectives

In this lab you will learn how to:

1. Create a sample from one population.
2. Create the sample mean or sum.
3. Create sampling distributions and appropriate graphs for these particular statistics.
4. Apply the CLT and discern its limitations.

Iter

n

n

n

n

### Theory: Suppose that are independent random variables taken from some distribution (not necessarily normal) and could be continuous or discrete.

***Then the expected value of the mean and sum can be calculated as follows***

### The variance of T and can also be found

***The CLT says that if n is large then with good approximation and .***

### Tasks

Please complete the script file Lab 8.R and then use it to create an RMD file. This is then to be knit into html and pdf files

**Note: All plots you are asked to make should be created through Rmd.**

* Task 1
  + Make a folder LAB8
  + Download the file “lab8.r”
  + Place this file with the others in LAB8.
  + Start Rstudio
  + Open “lab8.r” from within Rstudio.
  + Go to the “session” menu within Rstudio and “set working directory” to where the source files are located.
  + Issue the function getwd() and copy the output here.
* Task 2
  + Create your own R file and record the R code you used to complete the lab.
  + Create a sample of size n=10 from a uniform distribution that has lower limit 0 and upper limit 5 by using runif(10,0,5). Record the results here.
  + Give the mean and variance of the uniform for the case where a=0, b=5, i.e.
  + Use the sample you made to calculate . How do they compare to the population parameters?
  + Use the above theory to write down the mean and variance of the distribution of:
    - The sum
    - The mean
  + Below I have given the simple function myclt()
    - myclt=function(n,iter){
    - y=runif(n\*iter,0,5) # A
    - data=matrix(y,nr=n,nc=iter,byrow=TRUE) #B
    - sm=apply(data,2,sum) #C
    - hist(sm)
    - sm
    - }
    - w=myclt(n=10,iter=10000) #D
    - Explain what the following lines do
      * A
      * B
      * C
      * D
  + Record the plot made when D is executed
  + Using the object w, find sample estimates (you can use mean() and var())
  + Change the code in myclt() so that it produces a histogram of the sample means and releases a vector of sample means. **Hint:** You only need to change the last three lines beginning with line C.
  + Use this changed function and execute line D as the first step to calculate the estimates record these and the plot that the function makes.
* Task 3
  + We will now make a more sophisticated graph using mycltu().
  + Examine the function and the comments which explain what the code does.
    - w=apply(data,2,mean), how does the apply function use the 2?
    - How many terms are in w, when mycltu(n=20,iter=100000) is called?
    - curve(dnorm(x,mean=(a+b)/2,
    - sd=(b-a)/(sqrt(12\*n))),add=TRUE,col="Red",lty=2,lwd=3): Explain why sd takes the formula as shown in the function.
  + Record the plots using the following parameters and options
    - n=1, iter=10000, a=0,b=10
    - n=2, iter=10000, a=0,b=10
    - n=3,iter=10000,a=0,b=10
    - n=5,iter=10000,a=0,b=10
    - n=10, iter=10000,a=0,b=10
    - n=30,iter=10000,a=0,b=10
  + What do you conclude?
* Task 4
  + We will now make samples from a binomial distribution using mycltb().
  + Make graphs for the following parameters and options
    - n=4,iter=10000,p=0.3
    - n=5,iter=10000,p=0.3
    - n=10,iter=10000,p=0.3
    - n=20,iter=10000,p=0.3
  + Do the same, except use p=0.7
  + Do the same again this time with p=0.5
  + What do you conclude?
* Task 5
  + This task will need to be recorded
  + This time we will make use of the Poisson distribution using mycltp().
  + Make graphs for the following parameters and options
    - n=2, iter=10000,lambda=4
    - n=3,iter=10000,lambda=4
    - n=5,iter=10000,lambda=4
    - n=10,iter=10000,lambda=4
    - n=20,iter=10000,lambda=4
  + Do the same for lambda=10.
  + Now record all of TASK 5 with BBFLASHBACK
  + Place the .fbr file into the lab 8 dropbox

################### LAB FINISHES HERE ###############################

* Task 6:: Extra for experts!
  + Repeat task 5 but only after you have changed appropriate code to make mycltp() produce the sampling distribution for the sum rather than the mean.
  + Verify (by using the function) that the sampling distribution of the sum is approximately normal when n is large.