Assignment 1 - Introduction and R programming

Name: Amine Natik
Date Date: January 25, 2018

1. Let us first recall the weak law of large numbers (WLLN),

> Let $\{X_n\}_{n\geq 1}$ be a sequence of independent and identically distributed random variables, such that $\mathbb{E}[|X_1|] < \infty$ then :
>
> $$\frac{1}{n}(X_1 + \cdots + X_n) \to \mathbb{E}[X_1] \text{ in probability, as } n \to \infty$$

We write create a sample of size n from three different distributions and then calculate the mean of the sample to prove the weak law of large numbers. We choose to sample from, **Poisson**(λ), **Exponential**(λ) and **Chi squared**(p). Lets recall the properties of each one of these distributions,

	pdf / pmf	mean	variance
Poisson(λ)	$\frac{e^{-\lambda}\lambda^x}{x!}$	λ	λ
Exponential(λ)	$\lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Chi squared(p)	$\frac{1}{\Gamma(\frac{p}{2})2^{\frac{p}{2}}} x^{\frac{p}{2}-1} e^{-\frac{x}{2}}$	p	$2p$

(a) We will write a function called $WLLN(n, dist, para)$ that takes as inputs, n the sample size, $dist$ is a string of characters which is the distribution we sample from, here $distr$ can be $poisson$, $exponential$ or $chisquared$ and $para$ which is the parameter of the distribution $dist$.

```
WLLN=function(n=10, dist="poisson", para=1){
  if(dist=="poisson"){
    if (para>=0){
      X=rpois(n,para)
    }
    else{
      print("Please enter a nonnegative parameter!")
    }}
  if(dist=="exponential"){
    if(para>0){
      X=rexp(n,para)
    }
  else{
    print("Please enter a positive parameter!")
  }}
  if(dist=="chisquared"){
    if((para==round(para))&&(para>0)){
      X=rchisq(n,para)
    }
    else{
      print("Please enter a positive integer parameter!")
    }}
  return (list(mean(X), n, dist))
}
```

(b) We will now show that the mean of the sample is very close to the mean of the distribution when the sample size n is very big.

- *Poisson case :* Here we choose the parameter $\lambda = 2$ hence the mean is 2

```
# The parameter is 2
> para=2
```

```
# The sample size is 1000
> n=1000
> WLLN(n,"poisson",para)
[[1]]
[1] 1.933
[[2]]
[1] 1000
[[3]]
[1] "poisson"

# The sample size is 10000
> n=10000
> WLLN(n,"poisson",para)
[[1]]
[1] 1.9919
[[2]]
[1] 10000
[[3]]
[1] "poisson"

# The sample size is 100000
> n=100000
> WLLN(n,"poisson",para)
[[1]]
[1] 2.00125
[[2]]
[1] 1e+05
[[3]]
[1] "poisson"
```

The sample means are 1.933, 1.9919 and 2.00125, all are very close to 2.

- *Exponential case :* We choose $\lambda = 2$ then the mean is 0.5

```
# The parameter is 2
> para=2

# The sample size is 1000
> n=1000
> WLLN(n,"exponential",para)
[[1]]
[1] 0.5040239
[[2]]
[1] 1000
[[3]]
[1] "exponential"

# The sample size is 10000
> n=10000
> WLLN(n,"exponential",para)
[[1]]
[1] 0.5115946
[[2]]
[1] 10000
[[3]]
[1] "exponential"

# The sample size is 100000
```

```
> n=100000
> WLLN(n,"exponential",para)
[[1]]
[1] 0.500396
[[2]]
[1] 1e+05
[[3]]
[1] "exponential"
```

The sample means are 0.5040239, 0.5115946 and 0.500396, all are very close to 0.5.

- *Chi squared :* We choose $p = 2$ then the mean is 2

```
# The parameter is 2
> para=2

# The sample size is 1000
> n=1000
> WLLN(n,"chisquared",para)
[[1]]
[1] 1.958581
[[2]]
[1] 1000
[[3]]
[1] "chisquared"

# The sample size is 10000
> n=10000
> WLLN(n,"chisquared",para)
[[1]]
[1] 1.982737
[[2]]
[1] 10000
[[3]]
[1] "chisquared"

# The sample size is 1000
> n=100000
> WLLN(n,"chisquared",para)
[[1]]
[1] 1.998634
[[2]]
[1] 1e+05
[[3]]
[1] "chisquared"
```

The sample means are 1.958581, 1.982737 and 1.998634, all are very close to 2.

Conclusion: In all the previous cases we see that the sample mean is very close to the distribution mean when n is very big.

2. Recall first the Central Limit Theorem (CLT)

Let $\{X_n\}_{n\geq 1}$ be a sequence of independent and identically distributed random variables having mean μ and finite nonzero variance σ^2. Define the random variable Z_n by :

$$Z_n = \frac{1}{\sigma\sqrt{n}}\left(\frac{1}{n}(X_1 + \cdots + X_n) - \mu\right) \text{ for all } n \geq 1$$

Then $\{Z_n\}_{n\geq 1}$ converges in distribution to a standard normal distribution $\mathcal{N}(0,1)$ as $n \to \infty$

In order to show the CLT, we choose an integer n very large, and then calculate Z_n as before, where the X_i's are sampled from a given distribution, we repeat this process to have a random sample $\tilde{Z} = \left(\tilde{Z}_n^{(1)}, \ldots, \tilde{Z}_n^{(m)}\right)$ of size m from the distribution of Z_n, we then plot the histogram of $\tilde{Z}$, it should be similar to the histogram of the standard normal distribution.

(a) As explained before, we create a function $CLT(n, m, dist, para)$, where m is the size of $\tilde{Z}$, n, $dist$ and $para$ are defined as before (see question 1), the function should return the sample $\tilde{Z}$.

```
CLT=function(m=10,n=10,dist="poisson",para=1){
  Z=c()
  for (i in 1:m){
    if(dist=="poisson"){
      if (para>=0){
        X=rpois(n,para)
        mean=para
        sd=sqrt(para)
      }
      else{
        print("Please enter a nonnegative parameter!")
      }}
    if(dist=="exponential"){
      if(para>0){
        X=rexp(n,para)
        mean=1/para
        sd=1/para
      }
      else{
        print("Please enter a positive parameter!")
      }}
    if(dist=="chisquared"){
      if((para==round(para))&&(para>0)){
        X=rchisq(n,para)
        mean=para
        sd=sqrt(2*para)
      }
      else{
        print("Please enter a positive integer parameter!")
      }}
    Z=c(Z,(sqrt(n)*(mean(X)-mean)/sd))
  }
  return (Z)
}
```

(b) Here we take the sample size of $\tilde{Z}$, $m = 1000$.

- *Poisson case :* (parameter $\lambda = 2$),

```
> # We first use the following command to plot
> # all the nine histograms in the same figure.
> par(mfrow=c(3,3))
> # The parameter is 2
> para=2
> m=1000
> # 1st simulation
> n=1000
> Z=CLT(m,n,"poisson",para)
> hist(Z,prob=T,main="Poisson (n=1000)")
> # The following command add
```

```
> # in the same figure, the graph
> # of a standard normal distribution.
> curve(dnorm(x),from=-4,to=4,add=T)
> # 2nd simulation
> n=10000
> Z=CLT(m,n,"poisson",para)
> hist(Z,prob=T,main="Poisson (n=10000)")
> curve(dnorm(x),from=-4,to=4,add=T)
> # 3rd simulation
> n=100000
> Z=CLT(m,n,"poisson",para)
> hist(Z,prob=T,main="Poisson (n=10000)")
> curve(dnorm(x),from=-4,to=4,add=T)
```

- *Exponential case :* (parameter $\lambda = 2$),

```
> # The parameter is 2
> para=2
> m=1000
> # 1st simulation
> n=1000
> Z=CLT(m,n,"exponential",para)
> hist(Z,prob=T,main="Exponential (n=1000)")
> curve(dnorm(x),from=-4,to=4,add=T)
> # 2nd simulation
> n=10000
> Z=CLT(m,n,"exponential",para)
> hist(Z,prob=T,main="Exponential (n=10000)")
> curve(dnorm(x),from=-4,to=4,add=T)
> # 3rd simulation
> n=100000
> Z=CLT(m,n,"exponential",para)
> hist(Z,prob=T,main="Exponential (n=100000)")
> curve(dnorm(x),from=-4,to=4,add=T)
```

- *Chi Squared case :* (parameter $p = 2$),

```
> # The parameter is 2
> para=2
> m=1000
> # 1st simulation
> n=1000
> Z=CLT(m,n,"chisquared",para)
> hist(Z,prob=T,main="ChiSquared (n=1000)")
> curve(dnorm(x),from=-4,to=4,add=T)
> # 2nd simulation
> n=10000
> Z=CLT(m,n,"chisquared",para)
> hist(Z,prob=T,main="ChiSquared (n=10000)")
> curve(dnorm(x),from=-4,to=4,add=T)
> # 3rd simulation
> n=100000
> Z=CLT(m,n,"chisquared",para)
> hist(Z,prob=T,main="ChiSquared (n=100000)")
> curve(dnorm(x),from=-4,to=4,add=T)
```

The figure below gives us the histograms given by the above simulations,

Figure 1: CLT simulations.

Conclusion: In all the previous figures we see that the histogram is very close to the curve of the standard normal distribution, which demonstrates the CLT.

3. We define the pascal(n) function as:

```
pascal=function(n){
  T=list(c(1),c(1,1))
  for (i in 3:n){
    T[[i]]=rep(1,i)
      for (j in 2:(i-1)){
        T[[i]][j]=T[[i-1]][j]+T[[i-1]][j-1]
      }
  }
  return (T)
}
```

Now we plot the value of *pascal*(6),

```
> pascal(6)
[[1]]
[1] 1

[[2]]
[1] 1 1

[[3]]
[1] 1 2 1

[[4]]
[1] 1 3 3 1

[[5]]
[1] 1 4 6 4 1

[[6]]
[1]  1  5 10 10  5  1
```

4. The function is:

```
Generations=function(){
  # Use this commande to have all the nine histograms,
  par(mfrow=c(3,3))
  # The initial population is,
  population=rnorm(200,160,20)
  # elements from 1 to 100 are Males,
  # elements from 101 to 200 are Females.
  #Create the list of Males
  Males=list()
  # Create the list of Females
  Females=list()
  ##
  # The list Males will contain
  # all the males of all the generations,
  # same thing for the list Females.
  for (i in 1:9){ # i the index of the generation
```

```
    # define vectors of 100 elements:
    Males[[i]]=vector(length =100)
    Females[[i]]=vector(length =100)
    for(j in 1:100){
      # take a random male :
      m=sample(population[1:100],1)
      # take a random female :
      f=sample(population[101:200],1)
      # Define the two offspring:
      Males[[i]][j]=(m+f)/2
      Females[[i]][j]=(m+f)/2
    }
    # The population of i-th generation is:
    population=c(Males[[i]],Females[[i]])
    # We then plot the histogram :
    hist(population,main=paste("population",i))
    abline(v=mean(population),col="red")
  }
}
```

We call this function to have the plots in Figure 2.

```
> Generations()
```

Figure 2: Histograms of all the generations.

Conclusion: we see that the height of all the generations are concentrated around the red line (the sample mean).