Rapport du TP 3

1. La méthode des moments

Q1 : pour répondre à cette question, nous allons utiliser la fonction *runifa()* proposée qui va nous générer un échantillon issu de la loi $U_{[0,a]}$ où a est un paramètre inconnu que nous allons estimer par la méthode des moments. On sait que l'espérance (le moment d'ordre 1) d'une variable aléatoire X qui suit la loi $U_{[0,a]}$ est donnée par :

$$E(X) = \frac{a}{2}$$

Par conséquent :

$$a = 2E(X)$$

Comme on sait que la moyenne empirique $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ est un estimateur de $E(X)$ alors l'estimateur de a en utilisant le moment d'ordre 1 est donné comme suit :

$$\hat{a} = 2\bar{X}$$

Par conséquent, pour chaque échantillon de taille n on peut calculer une réalisation de $\hat{a}$, et c'est la fonction *estim() :*

```
> estim <- function(x) {
+   return(2*mean(x))
+ }
```

Q2 : on utilise la fonction *replicate()* pour estimer a avec des échantillons différents comme proposé :

```
> a <- replicate(1000 , estim(runifa(100)))
```

A ce niveau nous avons 1000 estimations différentes du paramètre a stocké dans la variable a, nous pouvons calculer la moyenne de la variable a pour avoir une idée sur la valeur du paramètre a :

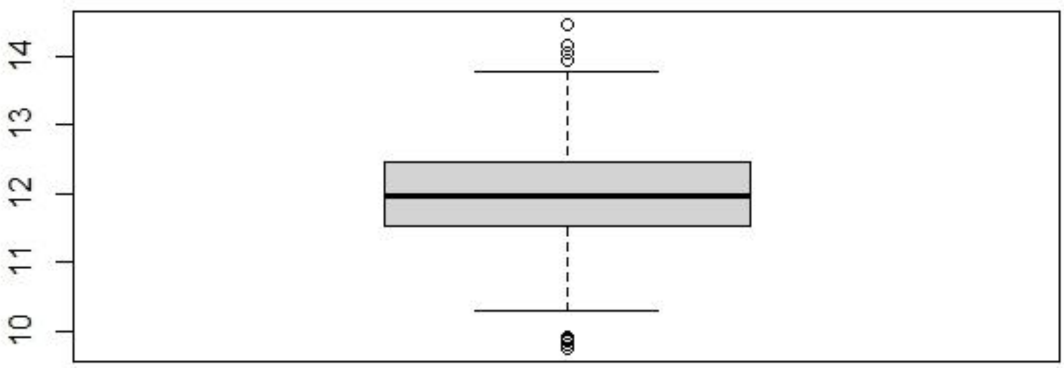
```
> mean(a)
[1] 11.98956
```

A ce niveau le paramètre a est probablement égale à 12.

Q3 : la moyenne est un bon moyen de résumer le vecteur stocké dans la variable a. Mais une boite à moustache donne plus d'informations :

```
> boxplot(a)
```

On obtient alors :



Ce qui confirme ce que nous avons obtenu en utilisant la moyenne dans Q2.

Q4 : nous savons que les moments d'ordre k sont donné par :

$$E(X^k) = \frac{a^k}{k+1}$$

Par conséquent, à partir d'un moment de n'importe quel ordre $k \in \mathbb{N}$ nous pouvons calculer la valeur de a :

$$a = \sqrt[k]{(k+1)E(X^k)}$$

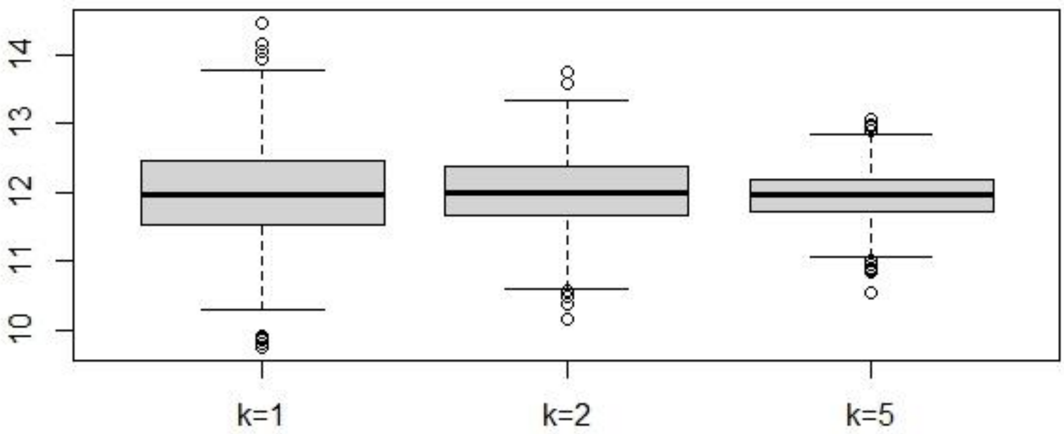
Comme on sait que $\widehat{m}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ est un estimateur de $E(X^k)$ alors l'estimateur de a en utilisant le moment d'ordre k est donné comme suit :

$$\hat{a} = \sqrt[k]{(k+1)\widehat{m}_k}$$

Par conséquent, pour chaque échantillon de taille n on peut calculer une réalisation de $\hat{a}$ par rapport à un moment d'ordre k, et c'est la fonction *estimk() :*

```
> estimk <- function(x , k) {
+   return(((k+1) * mean(x^k))^(1/k))
+ }
> a2 <- replicate(1000 , estimk(runifa(100) , 2))
> a5 <- replicate(1000 , estimk(runifa(100) , 5))
> boxplot(a,a2,a5 , names = c("k=1" , "k=2" , "k=5"))
```

Les boites à moustache dans le cas $k = 1, k = 2$ et $k = 5$ sont données dans la figure suivante :



On remarque que plus l'ordre k du moment est grand plus l'estimateur est précis.

2. Le théorème central limite

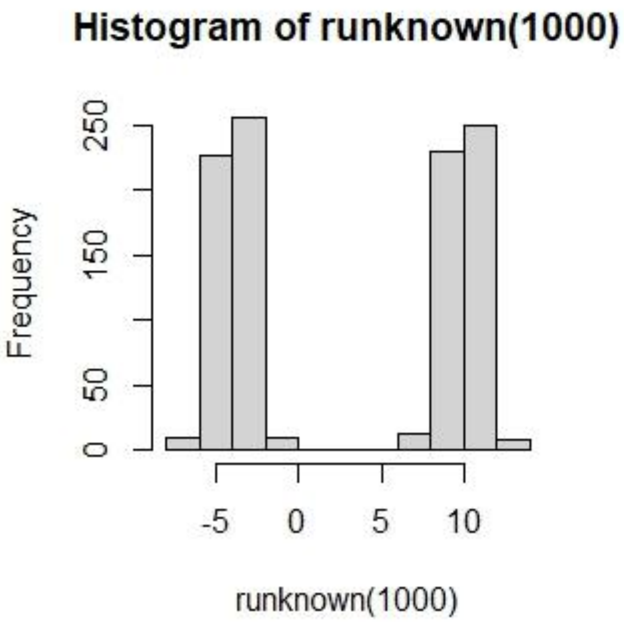
Q5 : pour vérifier l’espérance et l’écart type de la loi inconnue il faudra générer un échantillon de cette loi et puis on utilise les fonctions *mean()* et *sd()* :

```
> mean(runknown(100))
[1] 3.401173
> sd(runknown(100))
[1] 7.053212
> sqrt(50)
[1] 7.071068
```

Q6 : pour avoir un histogramme de cette loi il nous faut un échantillon sur qui nous allons appliquer la fonction *hist()* comme suit :

```
> hist(runknown(1000))
```

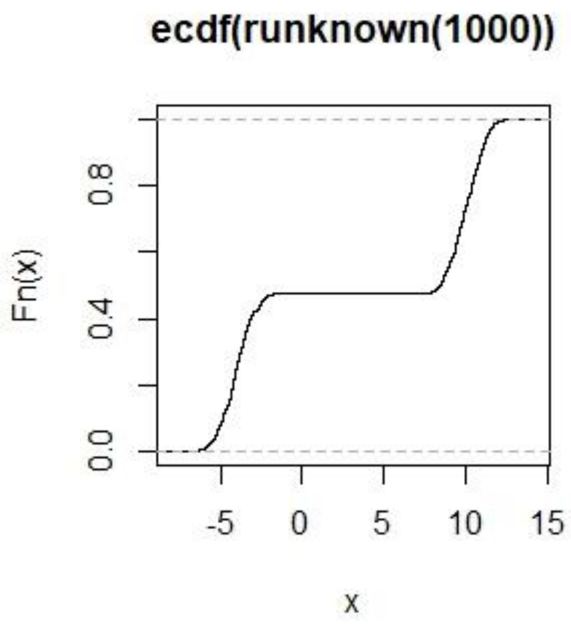
Ce qui nous donne la figure suivante :



Q7 : nous avons la fonction *ecdf()* qui donne la fonction de répartition empirique d’une loi à partir d’un échantillon de cette loi et donc il suffit de la dessiner avec la fonction *plot()* :

```
> plot(ecdf(runknown(1000)))
```

Ce qui va donner le graphe suivant :



On voit bien que ce n'est pas le graphe d'une fonction de répartition d'une loi connue encore moins de la loi normale centrée réduite.

Q8 : à partir d'un échantillon quelconque de la loi inconnue on peut calculer une réalisation de la variable T donnée par :

$$T = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

Le but est de vérifier expérimentalement qu'à partir d'une loi quelconque de X la variable T suit forcément une loi normale centrée réduite pour n tendant vers l'infini. Pour avoir une réalisation de T il suffit de générer un échantillon de la loi inconnu et de considérer que μ =3 et $\sigma = \sqrt{50}$:

```
> (mean(runknown(1000))-3)/(sqrt(50)/sqrt(1000))
[1] -0.8649685
```

Puis, on utilise encore une fois la fonction *replicate()* pour générer un échantillon de taille 1000 de la variable aléatoire T :

```
> T <- replicate(1000 , (mean(runknown(1000))-3)/(sqrt(50)/sqrt(1000)))
```

Q9 : il ne reste plus qu'à comparer la fonction de répartition empirique obtenue à partir de l'échantillon issu de la variable T avec la fonction de répartition de la loi normale centrée réduite :

```
> plot(ecdf(T))
> curve(pnorm(x , 0 , 1) , add = TRUE , col = "red")
```

Ce qui donne au final deux graphes presque similaires ce qui confirme le théorème central limite :