CY-Tech	Statistiques Corrigé TD 1	Mathématiques
2020 - 2021		ING2-GI

1 Exercice 1

1. X = Temps d'attente, en général modélisé par une loi exponentielle, donc $X \sim Exp(\theta)$.
2. Pour une loi exponentielle, on sait que $E(X) = \dfrac{1}{\theta}$

   Or on sait que $E(X)$ peut être estimé par son estimateur usuel, la moyenne empirique,

   $\overline{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ où X_i est le temps d'attente d'un étudiant et $n = 30$ ici.

   On en déduit donc une valeur approchée pour θ avec $\hat{\theta} = \dfrac{1}{\overline{x}} = \dfrac{1}{9}$
3. On suppose maintenant que $X \sim Exp\left(\dfrac{1}{9}\right)$, d'où

$P(X > 15) = 1 - F_X\ (15) = e^{-15\times\theta} = e^{-15/9} \simeq 0.20$

Il y a 20% de chance qu'il attende plus de 15 min avant d'être servi.

Remarque : Pour la loi exponentielle, on a aussi $V(X) = \dfrac{1}{\theta^2}$. On aurait pu utiliser l'écart-type empirique s^* pour estimer θ. On aurait obtenu $\hat{\theta} = \dfrac{1}{s^*}$, ce qui n'est pas la même chose.

On sait cependant que la moyenne est un meilleur estimateur de θ.

2 Exercice 2

$$f(x) = \begin{cases} \dfrac{4x^3}{\theta^4} & \text{si } x \in [0, \theta] \\ 0 & \text{sinon} \end{cases}$$

1. $E(X) = \displaystyle\int_{-\infty}^{+\infty} x\ f(x)\ dx = ... = \dfrac{4\theta}{5}$,

   $E(X^2) = \displaystyle\int_{-\infty}^{+\infty} x^2\ f(x)\ dx = ... = \dfrac{4\theta^2}{6}$, d'où $Var(X) = E(X) - (E(X))^2 = \dfrac{2\theta^2}{75}$.
2. $E(T_n) = a_n\ E(X_1 + X_2 + ... + X_n) = a_n\ n\ E(X) = n\ a_n \dfrac{4\theta}{5}$,

   Pour qu'il n'y ait pas de biais, il faut avoir $E(T_n) = \theta$.

   Il faut prendre $a_n = \dfrac{5}{4n}$, c'est à dire : $T_n = \dfrac{5}{4n}(X_1 + X_2 + ... + X_n) = \dfrac{5}{4}\overline{X}$.

   Le risque quadratique est alors :

   $R_\theta(T_n) = E\left((T_n - \theta)^2\right) = Var(T_n) = \left(\dfrac{5}{4}\right)^2 Var(\overline{X}) = \left(\dfrac{5}{4}\right)^2 \dfrac{V(X)}{n} = \dfrac{\theta^2}{24n}$

3 Exercice 3

La v.a. étudiée ici est $X = 1$ si la personne est atteinte, $= 0$ sinon. Elle suit donc une loi de Bernouilli : $X \sim \mathcal{B}(p)$. En particulier $E(X) = p$ et $Var(X) = p(1-p)$.

1. La fréquence (moyenne) empirique est toujours un estimateur sans biais de l'espérance. $\overline{X}_1$ et $\overline{X}_2$ sont donc sans biais.

   Les deux échantillons sont indépendants donc les deux fréquences (moyennes) empiriques aussi.

2. $E(\overline{X}_1) = E(\overline{X}_2) = p \quad \Longrightarrow E(T_3) = E\left(\frac{\overline{X_1}+\overline{X_2}}{2}\right) = p,$

   et de la même manière $E(T_4) = E\left(\frac{2\overline{X_1}+3\overline{X_2}}{5}\right) = p$. Les deux sont sans biais.

3. Nous savons que : $Var(\overline{X}) = \frac{Var(X)}{n} = \frac{p(1-p)}{n}$. Nous en déduisons :

   $Var(T_1) = \frac{p(1-p)}{40}$,

   $Var(T_2) = \frac{p(1-p)}{60}$,

   $Var(T_3) = \frac{Var(T_1)+Var(T_2)}{2^2} = \frac{5p(1-p)}{480}$

   $Var(T_4) = \frac{4Var(T_1)+9Var(T_2)}{25} = \frac{p(1-p)}{100}$

4. T_4 est celui qui a la plus petite variance, et donc le plus efficace.
   Avec les données des échantillons, la meilleure estimation est celle donnée par T_4 :
   $\hat{p} = \frac{2\overline{x_1}+3\overline{x_2}}{5} = \frac{2\frac{25}{40}+3\frac{38}{60}}{5} = 0.63$
   Il faut noter que T_4 revient à regrouper les 2 échantillons en un seul de taille 100 et à en calculer la fréquence (moyenne) empirique.

4 Exercice 4

$$f(x) = \begin{cases} \frac{2}{\theta}\left(1-\frac{x}{\theta}\right) & \text{si } 0 \le x \le \theta \\ 0 & \text{sinon} \end{cases} \qquad \text{et} \quad T_n = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

1. $E(T_n) = E(X) = \frac{\theta}{3}$, ce qui en fait un estimateur biaisé. Pour enlever ce biais, il suffit de multiplier T_n par 3.

   Dans la suite $\quad T_n = 3\overline{X} = \frac{3}{n}\sum_{i=1}^{n} X_i$.

2. La loi (faible) des grands nombres garantit la convergence en probabilité de $\overline{X}$ vers $E(X) = \frac{\theta}{3}$.
   Par conséquent T_n converge en probabilité vers θ.

3. Encore une fois, $Var(\overline{X}) = \frac{Var(X)}{n} = \frac{\theta^2}{18n}$. On en déduit :

   $R_\theta(T_n) = Var(T_n) = Var(3\overline{X}) = \frac{\theta^2}{2n}$ tend bien vers 0 lorsque n tend vers l'infini.
   L'estimateur est bien convergent.

5 Exercice 5

1. **(a)** $E(T_1) = E(\overline{X}) = E(X) = \theta$, de même $E(T_2) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) = E\left(X^2\right)$.

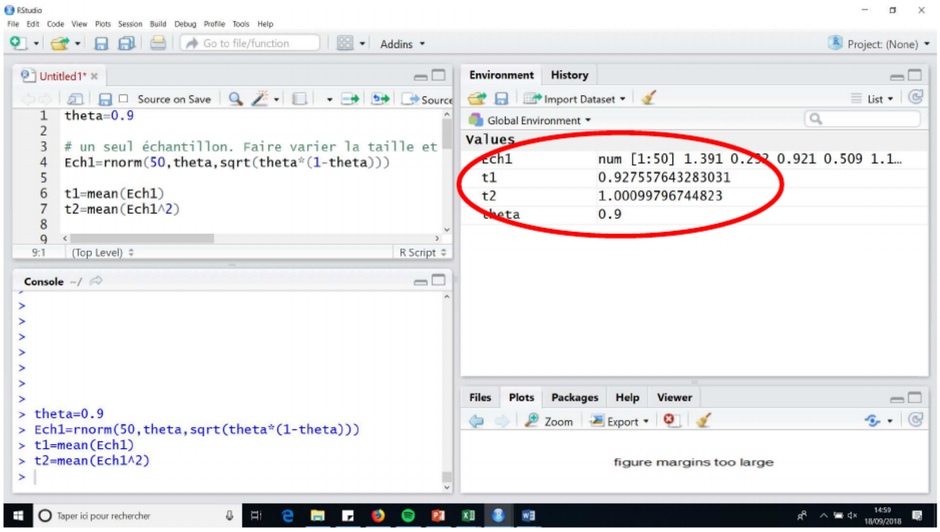
   Or $E\left(X^2\right) = Var(X) + (E(X))^2 = \theta(1-\theta) + \theta^2 = \theta$.
   Les deux estimateurs sont sans biais.
   (b) De plus la loi des grands nombres assure la convergence de T_1 vers $E(X) = \theta$ et celle de T_2 vers $E\left(X^2\right) = \theta$ également.
   (c) $Var(T_1) = \frac{Var(X)}{n} = \frac{\theta(1-\theta)}{n}, \qquad Var(T_2) = \frac{Var\left(X^2\right)}{n} = \frac{2\theta^2\left(1-\theta^2\right)}{n}$.
   $\frac{Var(T_2)}{Var(T_1)} = \frac{2\theta^2\left(1-\theta^2\right)}{\theta(1-\theta)} = 2\theta(1+\theta)$ qu'il convient de comparer à 1.

Il faut donc étudier le signe de $g(\theta) = 2\theta(1+\theta) - 1 = 2\theta^2 + 2\theta - 1$ qui a pour racines $\theta_1 = \dfrac{-1-\sqrt{3}}{2} < 0$ et $\theta_2 = \dfrac{-1+\sqrt{3}}{2} \simeq 0.37$

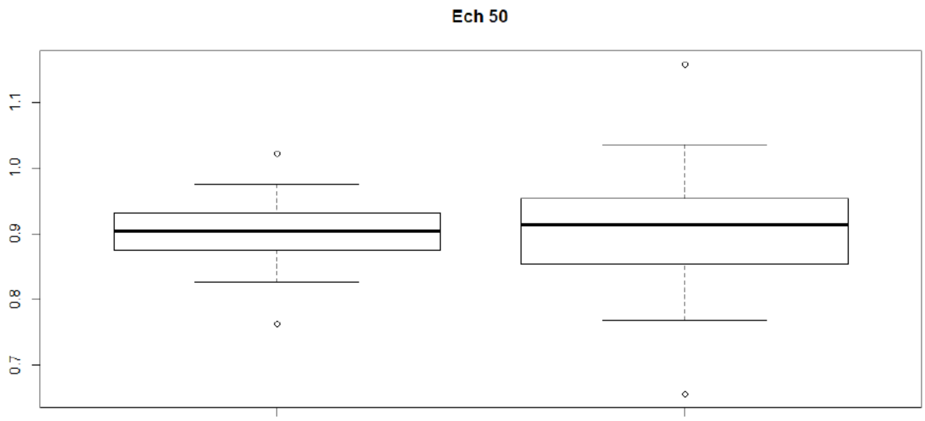
Sur $]0;\theta_2]$, $g \leq 0$, T_2 a une plus petite variance et est donc meilleur.

Sur $]\theta_2;1[$, $g > 0$, T_1 a une plus petite variance et est donc meilleur.

2. (a)
```
theta=0.9
Ech1=rnorm(50,theta,sqrt(theta*(1-theta)))
t1=mean(Ech1)
t2=mean(Ech1^2)
```

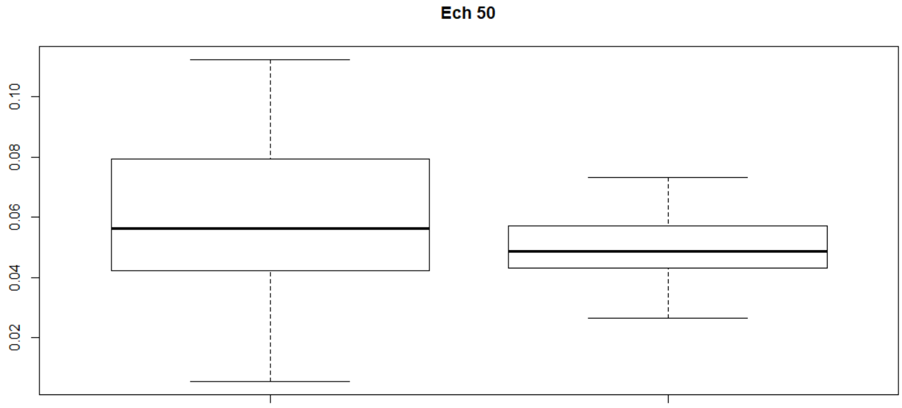


(b)
```
Tirages.50=rnorm(1500,theta,sqrt(theta*(1-theta)))
Ech.50=matrix(Tirages.50,nrow=30,ncol=50)
t1.50=apply(Ech.50,1,mean)
t2.50=apply(Ech.50^2,1,mean)
boxplot(t1.50,t2.50,ylim=range(t1.50,t2.50),main="Ech 50")
```



On note une estimation médiane à quasiment 0,9 pour les deux estimateurs. En revanche l'estimateur 2 présente plus de fluctuations que le 1.

(c) Avec $\theta = 0.05$, on obtient le résultat inverse comme prévu par la partie théorique.



(d) Avec $n = 5000$, la différence entre les deux estimateurs s'estompe car ils convergent tous les deux en moyenne quadratique.