Rapport sur une étude de trois estimateurs

LO Ousmane

2017/12/06

Sommaire

1- Introduction :

Ce rapport illustre une étude menée afin de comparer la performance de trois estimateurs de θ que sont la **moyenne empirique** $\hat{\theta}_1$, la **médiane empirique** $\hat{\theta}_2$ et la **moyenne empirique tronquée** $\hat{\theta}_3$ par des simulations Monte-Carlo. $\hat{\theta}$ est le paramètre de translation du modèle de translation d'une loi de Student.

Comparaison des trois estimateurs $\hat{\theta}_1 = \bar{X_n}, \hat{\theta}_2 = x_{\frac{1}{2}}$ et $\hat{\theta}_3 = X_{tronq(\gamma)}$ en fixant la taille d'échantillon **n**, le degré de liberté **q**, le paramètre θ et le nombre **K** d'échantillons. C'est-à-dire que l'on ne genere que **K** d'échantillions de taille **n** à simuler. θ est le paramètre de translation du modèle de translation d'une loi de Student.

```
## [1] "Le risque quadratique pour la moyenne empirique est :"
```

```
## [1] 0.0318122
```

```
## [1] "Le risque quadratique pour la mediane empirique  est :"
```

```
## [1] 0.03427264
```

```
## [1] "Le risque quadratique pour la moyenne tronquée est :"
```

```
## [1] 0.02748426
```

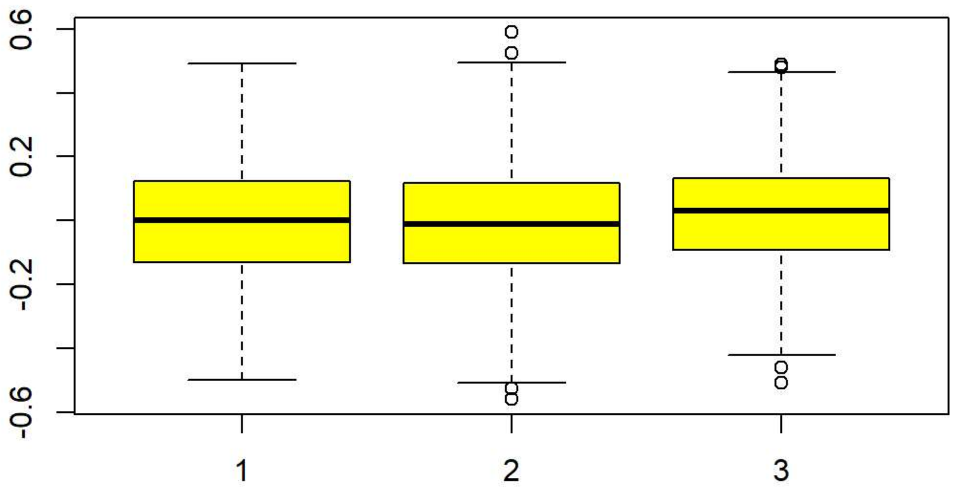

Figure 1: Les bloxplots des trois estimateurs de theta.

3- Interprétation des boxplots :

En fixant tous les paramètres (**n**, **q**, θ, **K**), On constate que $\hat{\theta}_3$ a le risque quadratique le plus proche de zéro “0”. De plus, la Figure 1 nous montre qu’il a le moins de valeurs aberrantes. Donc $\hat{\theta}_3$ est le meilleur estimateur de **theta**. Néanmoins, la conclusion que nous avons donnée est insuffisante quand tous les paramètres sont fixés. Donc il est nécessaire de faire varier un paramètre en fixant les autres et ensuite essayer de voir si cette variation influence sur la performance des trois estimateurs.

4- L’influence des valeurs n,q** et θ sur la performance des trois estimateurs :**

Pour évaluer si les valeurs de **q**,θ et **n** ont une influence sur la performance des trois estimateurs notre travail consistera à :

4.1– Fixer **q** et θ et varier **n**;

4.2– Fixer **n** et θ et varier **q**;

4.3– Fixer **n** et **q** et de faire varier θ.

En suite, dans chaque cas on compare les estimateurs.

4.1 Etudions d’abord le premier cas c’est à dire celui de faire varier **n**:

```
## Warning in matrix(c(1:3, 3, 1), 1, 3, byrow = TRUE): la longueur des
## données [5] n'est pas un diviseur ni un multiple du nombre de colonnes [3]
```

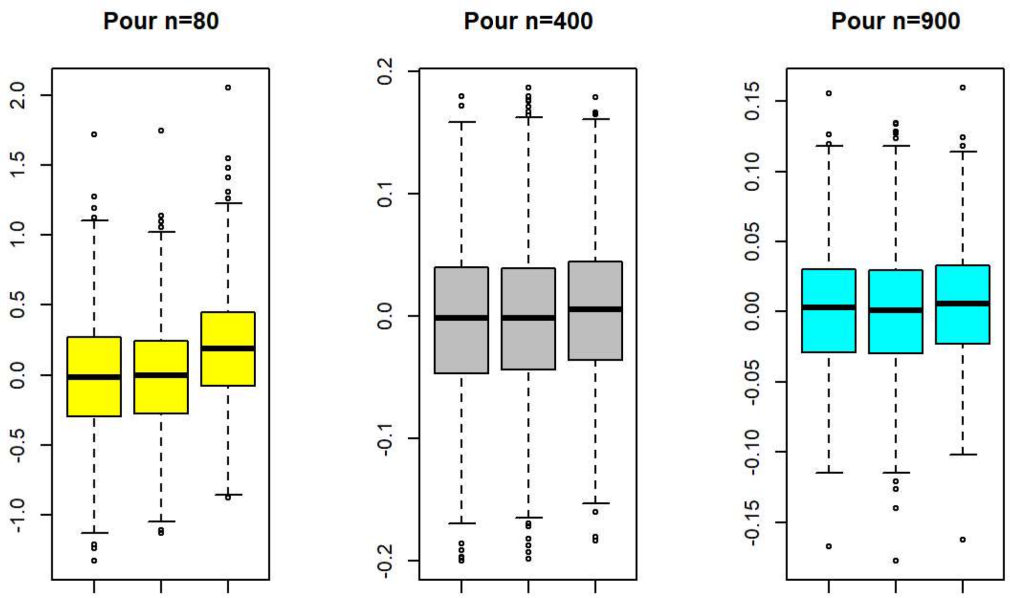

Figure 2: Les bloxplots des trois estimateurs de théta pour les différents n.

Voici respectivement les risques quadratiques des trois estimateurs que sont $\hat{\theta}_1 = \bar{X}_n$, $\hat{\theta}_2 = \ x_{\frac{1}{2}}$ et $\hat{\theta}_3 = X_{tronq(\gamma)}$ pour n=80.

```
## [1] 0.1639429 0.1497875 0.1889656
```

Voici respectivement $\hat{\theta}_1 = \bar{X}_n$, $\hat{\theta}_2 = \ x_{\frac{1}{2}}$ et $\hat{\theta}_3 = X_{tronq(\gamma)}$ et la moyenne tronquée pour n=400.

```
## [1] 0.003930837 0.004158181 0.003458261
```

Voici respectivement les risques quadratiques des trois estimateurs $\hat{\theta}_1 = \bar{X}_n$, $\hat{\theta}_2 = \ x_{\frac{1}{2}}$ et $\hat{\theta}_3 = X_{tronq(\gamma)}$ pour n=900.

```
## [1] 0.001876391 0.001992030 0.001693658
```

Interprétation :

Dans Figure 2, les trois estimateurs ont presque la meme performance,

Ici, on voit que plus **n** est grand plus la moyenne empirique tronquée a un risque quadratique plus petit que cels des deux autres. Donc quand **n** est grand, elle reste aussi le meilleur estimateur de θ comparée au deux autres.

4.2 Fixons maintenant **n** et θ et faisons varier **q**.

```
## Warning in matrix(c(1:4, 4, 1), 1, 4, byrow = TRUE): la longueur des
## données [6] n'est pas un diviseur ni un multiple du nombre de colonnes [4]
```

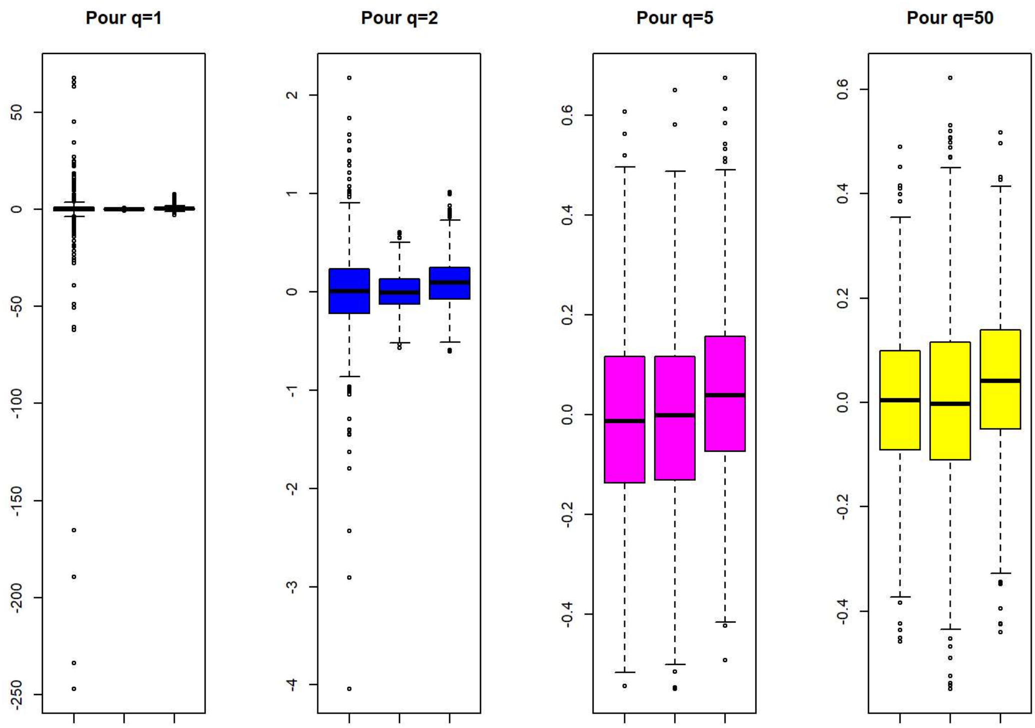

Figure 3: Les bloxplots des trois estimateurs de théta pour les différents valeurs de q.

Voici respectivement les risques quadratiques des trois estimateurs que sont $\hat{\theta}_1 = \bar{X_n}, \hat{\theta}_2 = \ x_{\frac{1}{2}}$ et $\hat{\theta}_3 = X_{tronq(\gamma)}$ pour q=1.

```
## [1] 226.39558715   0.05093019   0.87653931
```

Voici respectivement les risques quadratiques des trois estimateurs $\hat{\theta}_1 = \bar{X_n}, \hat{\theta}_2 = \ x_{\frac{1}{2}}$ et $\hat{\theta}_3 = X_{tronq(\gamma)}$ pour q=2.

```
## [1] 0.18415504 0.03735494 0.07282372
```

Voici respectivement les risques quadratiques des trois estimateurs $\hat{\theta}_1 = \bar{X_n}, \hat{\theta}_2 = \ x_{\frac{1}{2}}$ et $\hat{\theta}_3 = X_{tronq(\gamma)}$ pour q=5.

```
## [1] 0.03369405 0.03439695 0.03149183
```

Voici respectivement les risques quadratiques des trois estimateurs que sont $\hat{\theta}_1 = \bar{X_n}, \hat{\theta}_2 = \ x_{\frac{1}{2}}$ et $\hat{\theta}_3 = X_{tronq(\gamma)}$ pour q=50.

```
## [1] 0.02143247 0.03095711 0.02339673
```

Interprétation :

Pour Figure 3, on constate aussi que le fait de rendre **q** encore plus grand augmente la performance de la moyenne empirique tronquée. De ce fait, elle reste encore le meilleur estimateur.

On fixe d'abord **n**=50,θ=0.5 , **k**=1000 quand **q**=1 la loi Student coincide avec la loi de cauchy qui est une loi à queues lourdes et qui n'est pas integrable. Quand **q**>2 et tend vers l'infini la loi converge vers la loi normale standard qui est une loi a queues légères. Les resultats des risuques quadratiques et le boxplot montre que les 3 estimateurs comportent différemment en fonction de la valeur de **q**. C'est-à-dire que **q** peut inluencer la performance des estimateurs. Quand **q**=1 l'estimateur $\hat{\theta}_1$ est très mauvais avec risque quadratiques 7.253982e+03,et $\hat{\theta}_3$ l'est aussi avec risque quadratique 2.383461e+0, donc dans ce cas $\hat{\theta}_2$ qui est l'estimateur de la mediane empirique est le meilleur. Dans le cas **q**=5, $\hat{\theta}_1$ est une petite avance sur les deux autres pour des lois a qeues légères. Quand le q augmente, $\hat{\theta}_3$ a une tres bonne performance dans tous les cas. On peut dire que $\hat{\theta}_3$ est la version robuste de la moyenne adaptée aux lois a qeueus lourdes. De plus d'apres les valeurs des risques quadratiques, on observe que $\hat{\theta}_1$ est tres proche des vraies valeurs.

4.3 Maintenant on fait varier θ pour étudier le comportement des trois estimateurs. En fait, quand on a une loi de translation, theoriquement la variation de la valeur de theta n'influence pas sur les parformances des estimateurs. On vérifie si les estimateurs sont bien dans ce cas en pratique

```
## Warning in matrix(c(1:3, 3, 1), 1, 3, byrow = TRUE): la longueur des
## données [5] n'est pas un diviseur ni un multiple du nombre de colonnes [3]
```

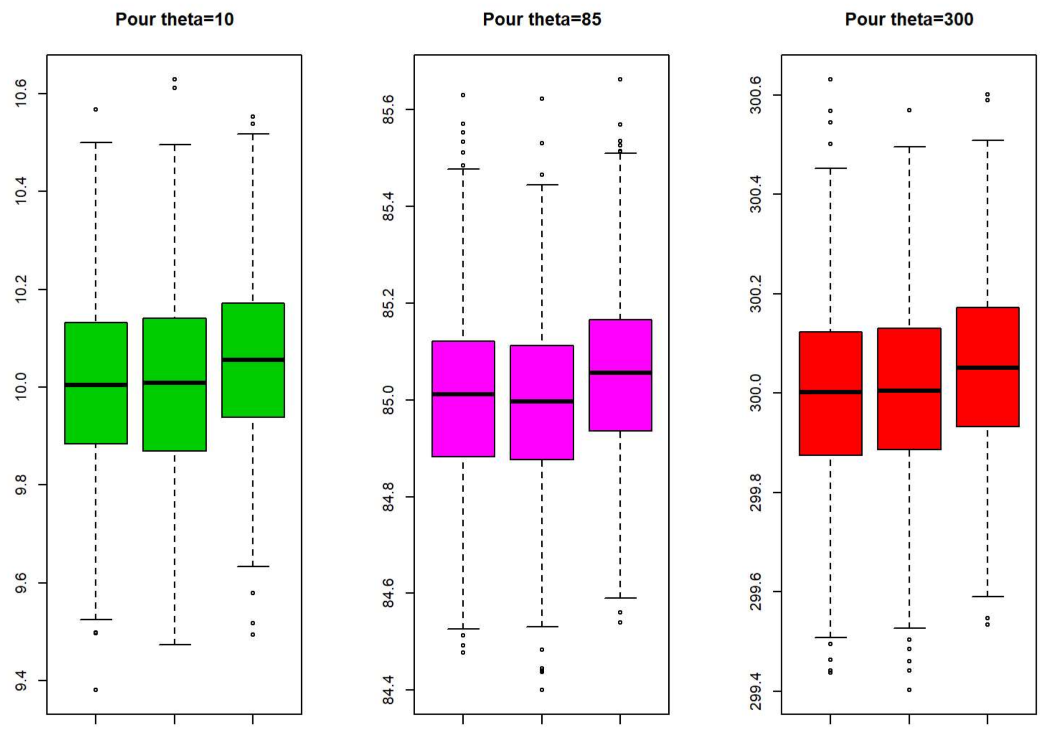

Figure 4: Les bloxplots des trois estimateurs de théta pour theta=50.

Voici respectivement les risques quadratiques des trois estimateurs que sont $\hat{\theta}_1 = \bar{X}_n, \hat{\theta}_2 = x_{\frac{1}{2}}$ et $\hat{\theta}_3 = X_{tronq(\gamma)}$ pour theta=10.

```
## [1] 100.0975 100.0788 101.1217
```

Voici respectivement les risques quadratiques des trois estimateurs $\hat{\theta}_1 = \bar{X}_n, \hat{\theta}_2 = x_{\frac{1}{2}}$ et $\hat{\theta}_3 = X_{tronq(\gamma)}$ pour theta=85.

```
## [1] 7225.563 7223.774 7233.931
```

Voici respectivement les risques quadratiques des trois estimateurs $\hat{\theta}_1 = \bar{X}_n, \hat{\theta}_2 = x_{\frac{1}{2}}$ et $\hat{\theta}_3 = X_{tronq(\gamma)}$ pour theta=300.

```
## [1] 89999.75 90003.96 90031.34
```

Interprétation :

Dans Figure 4 En pratique on a bien montre que le fait de faire varier θ n'a prèsque pas d'influence sur la performance des trois estimateurs donc ne modifierait pas l'hypothèse en théorie.

5- Conclusion :

En somme, d'aprés les variations des paramètres, on peut en déduire qu'il exsite une influence sur la performance des estimateurs par rapport à la valeur de **n** et **q**. C'est-à-dire que le fait de fixer simplement les paramètres ne nous donne pas une idée exacte sur l'effet des trois estimateurs. En variant **n** et **q**, sur le Figure 2et Figure 3, c'est $\hat{\theta}_1$ a une meilleur performance, alors qu'en comparant les riques quadratiques $\hat{\theta}_3$ est meilleur dans tous les cas.