TP 2 : Loi des grands nombres, théorème central limite

Ce TP n°2 continue de nous familiariser avec la simulation sous R, et aborde aussi l'utilisation d'estimateurs du point de vue pratique.

Exercice 1 : Loi des grands nombres et Théorème central limite Dans cet exercice, nous cherchons à illustrer la loi des grands nombres (LGN) et le Théorème central limite (TCL).

1. Rappeler ces deux théorèmes ainsi que leurs hypothèses.

Prenons tout d'abord comme exemple une suite de n variables aléatoires i.i.d. (c'est-à-dire indépendantes et de même loi) de loi exponentielle de paramètre $\lambda = 4$.

2. Quelle est la moyenne théorique (espérance) μ de la loi considérée ? Simuler $n = 8000$ telles variables et représenter

$$Y_k := \frac{1}{k} \sum_{i=1}^{k} X_i$$

en fonction de k pour k allant de 0 à n. Superposer une droite d'ordonnée égale à μ. Que remarque-t-on ? *Indication : on pourra utiliser la fonction* `abline`$(a, b, col = \text{"..."})$ *qui permet de tracer la droite* $y = a + bx$ *par-dessus un graphique.*

3. Quelle est la variance σ^2 de notre loi ? Cette fois-ci, on va simuler $N = 100$ échantillons de $n = 100$ variables, notées $(X_{i,j})_{1 \leq i \leq N, 1 \leq j \leq n}$. Pour tout $1 \leq i \leq N$ on pose

$$Z_i := \sqrt{n} \frac{\frac{1}{n} \sum_{j=1}^{n} X_{i,j} - \mu}{\sigma}.$$

Représenter à présent l'histogramme des Z_i (avec l'option `freq=F` et un bon réglage de `breaks`). Quelle courbe peut-on raisonnablement superposer à cet histogramme à cet histogramme (fonction `curve` (avec option `add=T,col="red"`)) ? Le faire et observer. Refaire éventuellement l'expérience avec N et n plus grands. Qu'observe-t-on ?

A présent, on suppose que la loi des X_n est une loi à densité $f(x)$ vérifiant

$$f(x) := \begin{cases} \frac{C}{x^2} & \text{si } x \geq 1, \\ 0 & \text{sinon.} \end{cases}$$

4. Quelle doit être la constante C pour que f soit bien une fonction de densité de probabilité ?

5. Calculer la fonction de répartition F associée à f, et en déduire une façon de simuler $n = 8000$ variables indépendantes de loi de densité f *en utilisant que des variables de loi uniforme sur* $[0, 1]$.

Ensuite, représenter

$$Y_k := \frac{1}{k} \sum_{i=1}^{k} X_i$$

en fonction de k pour k allant de 0 à n. Augmenter n. Semble-t-il y avoir convergence ? Expliquer.

Enfin, on suppose que la loi des X_n est une loi à densité $g(x)$ vérifiant

$$g(x) := \begin{cases} \frac{D}{x^3} & \text{si } x \geq 1, \\ 0 & \text{sinon.} \end{cases}$$

6. Quelle doit être la constante D pour que g soit bien une fonction de densité de probabilité ?

7. Calculer la fonction de répartition G associée à g, et simuler comme précédemment $n = 8000$ variables indépendantes de loi de densité g *en utilisant que des variables de loi uniforme sur* $[0, 1]$. La loi des grands nombres s'applique-t-elle ici ? Justifier.

8. Que vaut l'espérance μ de la loi ? On cherche pour finir à illustrer le théorème central limite. On suit la même stratégie que pour la question 3. : on va simuler $N = 10000$ échantillons de $n = 10000$ variables, notées $(X_{i,j})_{1\leq i\leq N, 1\leq j\leq n}$. Pour tout $1 \leq i \leq N$, on pose

$$Z_i := \sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{n} X_{i,j} - \mu\right).$$

Représenter l'histogramme des Z_i. Que se passe-t-il et pourquoi ?

9. (optionnel) Refaire les questions 2. et 3. avec des lois de Poisson, de Bernouilli, uniformes.