ENSAE
2021/2022

TP d'Introduction au langage R

Vocabulaire Quand on dira *Avec au plus n boucles*, pour $n \in \mathbb{N}^*$, il s'agira d'utiliser au plus n boucles de type `while`, `for`, ou une fonction de la famille `apply`. Sans boucle signifira sans utiliser une boucle de type `while`, ni `for`, ni une fonction de la famille `apply`.

Exercice 1 - Programmation efficace. Les questions ci-dessous sont indépendantes.

- Ecrice **sans boucle** une fonction `zeta(n, s)` qui renvoie la valeur de
$$\sum_{k=1}^{n} \frac{1}{k^s}.$$
- Écrire **sans boucle** une fonction `ppositive(x)` qui prend en argument un vecteur $x = (x_i)_{1 \leq i \leq n}$ numérique et renvoie un vecteur où la partie positive de x a été prise élément par élément, c'est-à-dire
$$(\max(x_i, 0))_{1 \leq i \leq n}.$$
- Écrire **sans boucle** une fonction `deriv1(f, a, b, h)` qui prend une fonction f en argument, deux nombres a et b et renvoie l'approximation de la dérivée par $(f(x+h) - f(x))/h$ pour x dans $[a, b]$ avec un pas de h, c'est à dire
$$\left(\frac{f(a+(i+1)h) - f(a+ih)}{h}\right)_{i \leq 0,\ (1+i)h \leq b-a}$$

Ercice 2 - Importation simple de données.

1. Importer correctement les données : `https://www.nicolasbaradel.fr/R/donnees/N_n.txt`, on mettra dans l'import la vérification du type.
2. Remplacer les `NA` par des 0.
3. Transformez vos données en une matrice.
4. Calculer la moyenne par ligne, sans faire de boucle.

Exercice 3 - Simulation et Monte Carlo. Soit $(X_k)_{k \geq 1}$ une suite de variables aléatoires réelles i.i.d. et N une variable aléatoire indépendante à valeurs dans $\mathbb{N}$. Soit :
$$S := \sum_{k=1}^{N} X_k.$$

L'objectif de cet exercice est d'être capable de simuler plusieurs copies de la variable aléatoire S afin d'en calculer des grandeurs statistiques par Monte Carlo. Nous allons donc simuler une suite $(S_i)_{1 \leq i \leq n}$ avec $n \geq 1$ fixé. En introduisant l'indice i de la simulation, on réécrit S sous la forme :
$$S_i := \sum_{k=1}^{N_i} X_k^i, \quad 1 \leq i \leq n$$

On fixe la loi des N_i qui suivent tous une loi de Poisson de paramètre $\lambda = 5$. Les sinistres X_k suivront une loi log-normale de paramètres `meanlog =` $\mu = 11$ et `sdlog =` $\sigma = 2$ à laquelle on ajoute 10^5 (il 'agit donc de la somme de la loi log-normale et de 10^5).

Afin d'avoir un code efficace, on propose dans un premier temps de ranger les simulations dans une matrice de la façon suivante. Si $n = 4$ et que, **par exemple**, on obtient $N_1 = 2$; $N_2 = 5$; $N_3 = 0$ et $N_4 = 7$, on a la matrice :

$$\begin{pmatrix} X_1^1 & X_2^1 & 0 & 0 & 0 & 0 & 0 \\ X_1^2 & X_2^2 & X_3^2 & X_4^2 & X_5^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ X_1^4 & X_2^4 & X_3^4 & X_4^4 & X_5^4 & X_6^4 & X_7^4 \end{pmatrix}.$$

Ainsi, chaque ligne i de la matrice représente les X_k^i allant de 1 à N_i. Le nombre de colonnes de la matrice est le nombre maximum parmi les simulations $(N_i)_{1 \leq i \leq n}$ et la matrice est complétée par des 0.

Écriture du code:

1. Fixer un nombre de simulations $n = 15$ et simuler le vecteur $(N_i)_{1 \leq i \leq n}$ selon la loi donnée ;
2. Construire la matrice et la remplir. *Avec au plus une boucle.*
3. *(Difficile)* Construire la matrice du point 2. et la remplir mais sans utiliser aucune boucle.
4. En utilisant la matrice, calculer le vecteur $(S_i)_{1 \leq i \leq n}$ sans aucune boucle.
5. On avait choisi $n = 15$ pour mieux tester, passer maintenant à $n = 10^5$ et en déduire une estimation du quantile à 99.5% de la variable aléatoire S.