UEVE 2017-2018
EC322 Probabilités

Projet: simulation d'une variable aléatoire

Remarques. **a) Ce projet est facultatif et peut être fait avec le logiciel R ou le langage C. Il est noté sur** 20 **et n'est en aucun cas pénalisant par rapport à la note de Devoir Surveillé (DS) puisqu'on comptera le** $\max($**Note Projet**$,$ **Note DS**$)$. **b) Il est fortement recommandé que ce soit fait en binôme même s'il est autorisé de le faire seul. Il peut aussi être fait en groupe de** n **étudiants,** $n \geq 3$. **Dans ce cas il sera noté sur** $20 - n$. **Si deux copies rendues séparément présentent des ressemblances la note finale des deux groupes sera de**

$$\min(\mathrm{NFG1}, \mathrm{NFG2}) - S$$

où NFG1 et NFG2 sont les notes finales des deux groupes et S **une note comprise entre** 1 **et** 20 **qui dépend du (et augmente avec le) degré de similitude des deux copies.**

ENONCÉ. On se donne une variable aléatoire X à valeurs dans $\{-2, -1, 0, 1\}$ avec

$$\mathbb{P}(X = -2) = \frac{1}{4}, \quad \mathbb{P}(X = -1) = \frac{1}{6}, \quad \mathbb{P}(X = 0) = \frac{1}{3}, \quad \mathbb{P}(X = 1) = \frac{1}{4}.$$

$\rightsquigarrow$ ***Simulation de réalisations de*** X. L'objectif de cette partie est de simuler n réalisations indépendantes $X_1(\omega), \ldots, X_n(\omega)$ (autrement dit, une suite de n valeurs de $X(\omega)$ prises indépendamment les unes des autres, on l'appelle aussi un échantillon de taille n de la loi de la variable aléatoire X) de la variable aléatoire X en utilisant le langage **C** ou le logiciel **R**, où le nombre de simulations n sera fixé. Par exemple, si *SimulLoiDiscrete(n, . . .)* est la fonction **C** qui simule n réalisations indépendantes de la variable aléatoire X, l'appelle à la fonction *SimulLoiDiscrete(10, . . .)* génèrera de façon aléatoire 10 valeurs appartenant à l'ensemble $\{-2, -1, 0, 1\}$ des valeurs prises par la variable aléatoire X. On peut par exemple avoir la suite de valeurs

```
-1  0  0  -2  - 2  0  1  -1  0  1.
```

La simulation d'une variable aléatoire est basée sur le principe suivant. Soit X une variable aléatoire de fonction de répartition $F(x) = \mathbb{P}(X \leq x)$, $x \in \mathbb{R}$. On suppose que l'on sait simuler des réalisations d'une variable aléatoire U de loi uniforme sur $]0, 1[$: $U \sim \mathcal{U}(]0, 1[)$. Pour simuler des réalisations de la variable aléatoire X on utilise le résultat suivant: si F^{-1} est la fonction quantile de X définie par

$$F^{-1}(u) := \inf\{x \in \mathbb{R} : \quad F(x) \geq u\}, \qquad \text{pour tout } u \in]0, 1[$$

alors X et $F^{-1}(U)$ ont la même loi de probabilité: on note $X \stackrel{\mathcal{L}}{=} F^{-1}(U)$. Nous allons utiliser ce dernier résultat pour simuler des réalisations de X.

Posons $p_1 = \frac{1}{4}$, $p_2 = \frac{1}{6}$, $p_3 = \frac{1}{3}$ et $p_4 = \frac{1}{4}$ les probabilités que X prenne les valeurs $a_1 = -2, a_2 = -1, a_3 = 0, a_4 = 1$, respectivement. La fonction de répartition F de X est

$$F(x) = \begin{cases} 0 & \text{si } x < -2 \\ 1/4 & \text{si } x \in [-2, -1[\\ 5/12 & \text{si } x \in [-1, 0[\\ 9/12 & \text{si } x \in [0, 1[\\ 1 & \text{si } x \geq 1 \end{cases}$$

et la fonction quantile F^{-1} est donnée par

$$F^{-1}(u) = \begin{cases} a_1 & \text{si } u \in]0, 1/4] \\ a_2 & \text{si } u \in]1/4, 5/12] \\ a_3 & \text{si } u \in]5/12, 9/12] \\ a_4 & \text{si } u \in]9/12, 1[. \end{cases}$$

Ainsi, pour simuler une réalisation x de X on procède comme suit:

- ***on simule d'abord une réalisation*** u ***de la loi uniforme sur*** $]0, 1[$**.**
- ***Si*** $u \in]0, 1/4]$ ***alors on pose*** $x = a_1$***, si*** $u \in]1/4, 5/12]$ ***alors on pose*** $x = a_2$***, si*** $u \in]5/12, 9/12]$ ***alors on pose*** $x = a_3$ ***et enfin, si*** $u \in]9/12, 1]$ ***alors on pose*** $x = a_4$**.**

On répètera cette dernière action n fois si on veut simuler un échantillon de taille n de X.

$\rightsquigarrow$ *Indications*. Vous pouvez utiliser "l'entête" ci-dessous (à compléter éventuellement) pour votre programme C.

```
#define n 10
#define p1 1.0/4
#define p2 1.0/6
#define p3 1.0/3
#define p4 1.0/4

#define a1 -2
#define a2 -1
#define a3 0
#define a4 1

#include<stdio.h>
#include<math.h>
#include<stdlib.h>
#include<time.h>
#include <ctype.h>
#include<string.h>
```

1. Ecrivez une fonction *SimulLoiDiscrete* telle que l'appelle à la fonction

   SimulLoiDiscrete(n, a, a2, a3, a4, p1, p2, p3, p4, p5, ValSim)

   ne renvoie rien mais remplit le tableau *ValSim* des N valeurs simulées de la variable aléatoire X. Le prototype de la fonction ressemblera a

   *void SimulLoiDiscrete(int n, . . .,double * Valsim)*

2. Appelez la fonction *SimulLoiDiscrete(n, . . .)* pour $n = 10^3$, $n = 10^4$, $n = 10^5$ et pour chaque valeur de n comparer les fréquences d'apparition de chacune des valeurs du tableau *ValSim* et les comparer aux probabilités p_i, $i = 1, 2, 3, 4$. Commenter.

3. Représentez graphiquement les fréquences d'apparition de chacune des valeurs du tableau *ValSim* en fonction de n, pour $n = 1, 2, 3, 4, \dots, 1000$. Pour cela, vous pourriez récupérer le tableau des fréquences dans un fichier texte et faire le graphique avec un logiciel de votre choix.

Autres indications. Le tableau *ValSim* dans la fonction *main* pourra être défini comme suit

*ValSim = (double *) malloc(n *sizeof(double));*

Dans la fonction `main`, on doit rajouter l'instruction ci-après avant d'appeler la fonction *SimulLoiDiscrete*:

srand((unsigned)time(NULL));

Pour simuler une réalisation de la loi uniforme dans la fonction *SimulLoiDiscrete* utilisez l'instruction *u = rand()/(RAND_MAX + 1.0);*

$\rightsquigarrow$ ***Approximation de*** $\mathbb{E}(X)$ ***par la loi des grands nombres***. La loi des grands nombres s'énonce comme suit: si $X_1, \dots, X_n$ est une suite de N réalisations indépendantes de X alors, avec probabilité 1, la moyenne empirique $\bar{X}_n = (X_1 + \dots + X_n)/n$ converge vers $\mathbb{E}(X)$ lorsque N tend vers l'infini.

4. Calculer $\mathbb{E}(X)$, puis (en posant $f_n(i)$ la fréquence d'apparition de i, $i \in \{-2, -1, 0, 1\}$)

   $$\bar{X}_n(\omega) = \frac{X_1(\omega) + \dots + X_n(\omega)}{n} = (-2)\times f_n(-1)+(-1)\times f_n(-1)+0\times f_n(0)+1\times f_n(1) \tag{1}$$

   en utilisant les fréquences obtenues dans la question 2. pour $n = 10^3$, $n = 10^4$ et $n = 10^5$. Comparez ces valeurs de $X_n(\omega)$ à $\mathbb{E}(X)$.

5. Représentez sur le même graphique la constante $\mathbb{E}(X)$ et la fonction $n \mapsto \bar{X}_n$, pour $n =$ 1, 2, 3, 4, . . ., 1000. Pour cela, vous pourriez récupérer le tableau (pour n variant) des moyennes empiriques $\bar{X}_n$ dans un fichier texte et faire le graphique avec un logiciel de votre choix.

$\rightsquigarrow$ ***Théorème de la limite centrale***. Le théorème central limite s'énonce comme suit: soit $X_1, \ldots, X_n$ est une suite de n réalisations indépendantes de X, si la $\mathrm{Var}(X) = \sigma^2$ est finie alors la variable aléatoire

$$Z_n = \frac{\bar{X}_n - \mathbb{E}(X)}{\sqrt{\mathrm{Var}(\bar{X}_n)}} = \sqrt{n}\,\frac{\bar{X}_n - \mathbb{E}(X)}{\sigma} \tag{2}$$

converge (en loi) vers une variable aléatoire Normale centrée réduite $Z \sim \mathcal{N}(0,1)$, lorsque $n \to +\infty$.

Maintenant, on sait que pour n fixé, une réalisation $Z_n(\omega^1) = Z_n^1$ de la variable aléatoire Z_n est obtenue en simulant un échantillon de taille n de la loi de X et en calculant la quantité (2). On pose $n_1 = 20$ et $n_2 = 30$.

6. Simulez un échantillon $Z_n(\omega^1) = Z_n^1, \ldots, Z_n(\omega^m) = Z_n^m$ de taille $m = 3$ de la loi de Z_n, pour $n = n_1$, puis pour $n = n_2$. Pour cela, vous devez calculer explicitement $\mathrm{Var}(X)$.

7. Simulez un échantillon $Z_n(\omega^1) = Z_n^1, \ldots, Z_n(\omega^m) = Z_n^m$ de taille $m = 5 \times 10^5$ de la loi de Z_n, pour $n = n_1$, puis pour $n = n_2$. Récupérez les résultats (pour $n_1 = 20$ et pour $n_2 = 30$) obtenus dans deux fichiers textes séparés.

8. En utilisant le logiciel de votre choix, tracez, dans deux figures séparées, l'histogramme des fréquences pour l'échantillon $Z_n^1, \ldots, Z_n^m$ (on pourra choisir 30 classes). Dans chacune des deux figures précédentes, tracez la densité f de la loi Normale centrée réduite définie par
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \qquad x \in \mathbb{R}.$$

9. Commentez les graphiques de la question 8 en vertu du théorème central limite.

Remarque. Pour ceux qui souhaitent le faire avec le logiciel **R** vous trouverez dans la page de M. Chiquet
http://julien.cremeriefamily.info/index.html une introduction à **R**.