TP6 : Généralités sur la fonction `rand` et simulations de v.a.r. uniformes

Pré-requis : je vous invite à consulter les chapitres de cours correspondants sur ma page (support informatique). Pour ce TP, on pourra en particulier se reporter à la section « Tracés d'histogrammes » du CH 8.

▶ Dans votre dossier `Info_2a`, créez le dossier `TP_6`.

I. Avant-propos

I.1. Sur la fonction `rand`

La fonction `rand` est un générateur de nombres **pseudo-aléatoires**.
Il vérifie notamment les propriétés suivantes :

- si $X(\omega)$ désigne le résultat de l'appel `rand()`, alors X est une v.a.r. de loi uniforme sur $[0, 1]$.
- si $X_1(\omega), \dots, X_N(\omega)$ désignent les résultats de N appels successifs de `rand()`, alors les variables aléatoires $X_1, \dots, X_N$ sont mutuellement indépendantes.

> En mathématiques, on travaille souvent avec des variables aléatoires, mais en informatique, on simule **une réalisation** de ces variables aléatoires (ce qui correspond à la notion d'observation en statistique).

I.2. Sur la loi (faible) des grands nombres pour une v.a.r. discrète

La loi (faible) des grands nombres[1] est un résultat clé en statistique inférentielle.
Illustrons brièvement ce résultat. On considère une population $\mathscr{P}$.

- Les membres de $\mathscr{P}$ appartiennent à l'une des catégories suivantes :

  1. enfant, *2.* adolescent, *3.* adulte.

- On considère l'expérience aléatoire consistant à tirer au sort une personne de la population.
  On note alors X la v.a.r. donnant le numéro de la catégorie de la personne choisie.
  Ainsi, $\mathbb{P}([X = 3])$ donne la proportion d'adultes dans la population.
- Afin d'obtenir des informations sur la population, on va effectuer des tirages successifs d'un individu.
  Ces tirages se font avec remise ce qui permet d'en assurer l'**indépendance**. On note alors X_i la v.a.r. qui donne la catégorie de la $i^{\text{ème}}$ personne choisie.
  Évidemment, X_i suit la même loi que X (on a notamment $\mathbb{P}([X_i = 3]) = \mathbb{P}([X = 3])$).
  On définit ce qu'on appelle un **échantillon** $(X_1, \dots, X_N)$ de la v.a.r. X.

Le théorème stipule que l'on peut approcher $\mathbb{E}(X)$ à l'aide d'un échantillon $(X_1, \dots, X_N)$.
Pour ce faire, on observe la valeur $(x_1, \dots, x_N)$ de cet échantillon.

> Lorsque N est grand, la moyenne des valeurs observées $\dfrac{1}{N} \sum\limits_{k=1}^{N} x_k$ devient proche de $\mathbb{E}(X)$.

[1] Ce résultat sera vu dans le chapitre « Convergences et approximations ».

Remarque

- On ne développe pas ici le formalisme de la loi (faible) des grands nombres. L'idée est plutôt de comprendre qu'il existe un théorème qui correspond à l'intuition mathématique suivante.

Afin d'approcher la moyenne probabiliste $\mathbb{E}(X)$ d'une variable aléatoire X, il suffit de prendre la moyenne arithmétique d'un **grand nombre** d'observations de X.

- Par « grand nombre », il faut évidemment comprendre qu'il y a une notion de limite sous-jacente. Plus le nombre d'observations est grand, plus on se rapproche du résultat théorique $\mathbb{E}(X)$.
  On obtiendrait la valeur exacte de $\mathbb{E}(X)$ si l'on savait faire une infinité d'observations.

▶ Revenons à notre exemple. On considère maintenant la v.a.r. Z suivante.

$$Z : \begin{array}{ccl} \Omega & \to & \mathbb{R} \\ \omega & \mapsto & \begin{cases} 1 & \text{si } X(\omega) = 3 \\ 0 & \text{si } X(\omega) \neq 3 \end{cases} \end{array}$$

Démontrer que Z est une v.a.r. discrète et déterminer son espérance.

▶ En déduire une méthode permettant d'obtenir une valeur approchée de la quantité $\mathbb{P}([X = 3])$.

De manière plus générale, lorsqu'on effectue un N-tirage, on note l'effectif (*i.e.* le nombre d'individus) de chaque catégorie. La loi faible permet d'affirmer que si N est grand, le diagramme des fréquences des observations $\left(\dfrac{\text{effectif de chaque catégorie}}{\text{nombre d'observations}}\right)$ fournit un diagramme approché de la distribution de probabilité $(\mathbb{P}([X = 1]), \mathbb{P}([X = 2]), \mathbb{P}([X = 3]))$.

I.3. Sur la loi (faible) des grands nombres pour une v.a.r. à densité

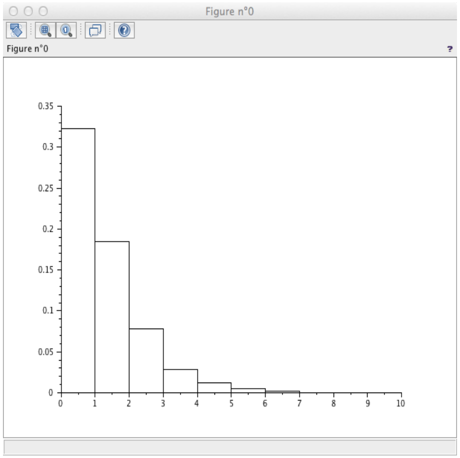
Dans l'épreuve `EDHEC 2017`, on s'intéressait à la loi de Gumbel et à la manière de la simuler informatiquement. Par définition, on dit qu'une v.a.r. X suit une loi de Gumbel si sa fonction de répartition F_X est donnée par :

$$\forall x \in \mathbb{R}, \ F_X(x) = \mathrm{e}^{-\mathrm{e}^{-x}}$$

On démontrait dans cet énoncé que si on dispose d'une v.a.r. V telle que $V \hookrightarrow \mathcal{E}(1)$ alors la v.a.r. $W = -\ln(V)$ suit une loi de Gumbel.

On considérait alors le script et l'histogramme résultat suivant :

```
V = grand(1,10000,'exp',1)
W = -log(V)
s = linspace(0,10,11)
histplot(s,W)
```



L'énoncé donnait l'explication suivante :

Ce script simule 10000 *variables indépendantes, regroupe les valeurs renvoyées en 10 classes qui sont les intervalles* $[0,1]$, $]1,2]$, $]2,3]$, $\ldots$, $]9,10]$ *et trace l'histogramme correspondant (la largeur de chaque rectangle est égale à* 1 *et leur hauteur est proportionnelle à l'effectif de chaque classe).*

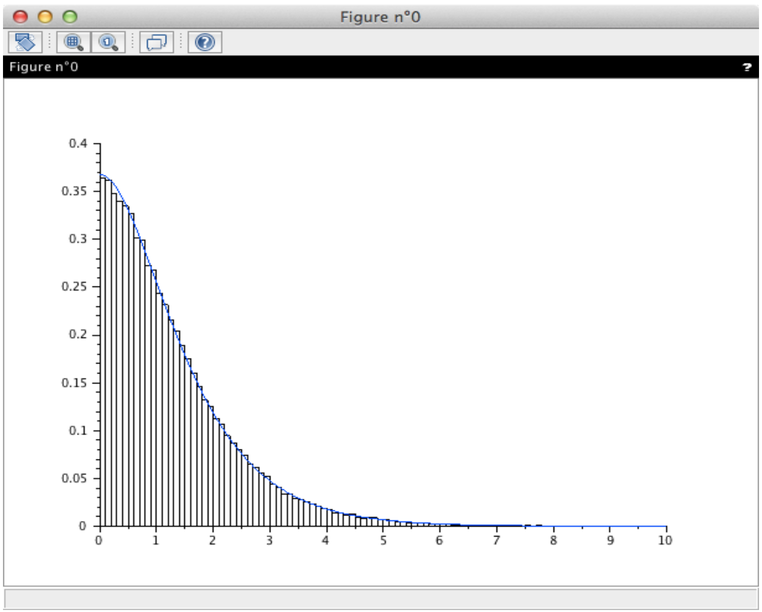
► Que représente chaque barre de l'histogramme ?
Détailler chaque ligne de la simulation en faisant le lien avec la loi (faible) des grands nombres.

Par défaut, un histogramme est normalisé : en sommant les aires de chaque barre de l'histogramme, on obtient 1. C'est donc bien l'aire de chaque barre qui est importante et pas la hauteur !

On considère maintenant le programme suivant.

```
V = grand(1,100000,'exp',1) // plus d'observations
W = -log(V)
s = 0:0.1:10 // des classes plus petites
histplot(s,W)
plot(s, exp(-exp(-s)) .* exp(-s))
```

► Qu'observe-t-on ? Expliquer le résultat obtenu.



II. Simulation de v.a.r. suivant une loi uniforme continue

II.1. Simulation de v.a.r. suivant une loi uniforme sur $[0, 1]$

- Évaluer `help rand` et prendre connaissance des informations notamment celles présentes dans la section `Générer des nombres aléatoires`.
- Évaluer `rand(2,6)`. Qu'obtient-on ?

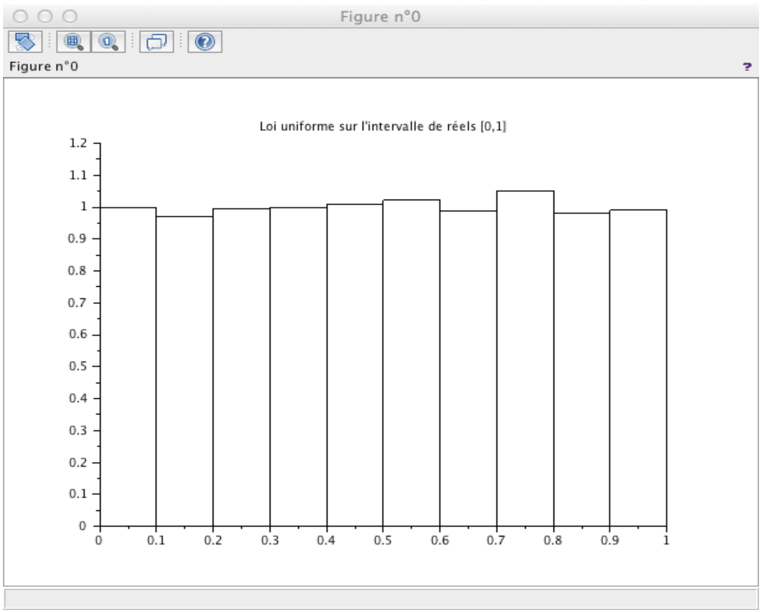
- Comparer ce résultat avec votre voisin. Commenter brièvement.

On peut se reporter, pour plus de détails au TP de l'an dernier.

On considère le programme `uniforme_continue.sce` suivant.

```
1 clf()
2 U = rand(1,10000) // correspond à 10000 appels successifs de rand()
3 histplot(0:0.1:1,U)
4 xtitle("Loi uniforme sur [0,1] : distribution approchée")
```

- Recopier ce fichier dans l'éditeur **SciNotes** et l'exécuter (touche `F5`).
  Donner une allure de la figure obtenue et l'expliquer brièvement.



- Remplacer la ligne 3 par `histplot(0:0.1:1, U, normalization=%f)`.
  À quoi sert de désactiver l'option `normalization` ?
  Noter le nombre d'éléments du vecteur `U` appartenant à la première classe.

- L'histogramme précédent contient 10 barres. Repérer le paramètre permettant de définir ces barres et modifiez-le de sorte à afficher 20 barres. Noter la ligne modifiée.

II.2. Loi uniforme sur $[a, b]$

On dispose du programme `uniforme_cont_2_7.sce` suivant.

```
clf()
U = 2+(7-2)*rand(1,10000) // correspond à 10000 appels successifs
histplot(2:0.5:7,U)
xtitle("Loi uniforme sur [2,7] : distribution approchée")
```

- Copier ce fichier dans l'éditeur **SciNotes** et l'exécuter (touche `F5`).
  Expliquer brièvement le résultat obtenu.

- En vous inspirant du résultat précédent, écrire un programme qui :
  - × demande à l'utilisateur d'entrer au clavier deux réels `a` et `b` et un entier `N`,
  - × affiche l'histogramme des fréquences d'un échantillon de taille `N` de v.a.r. suivant la loi uniforme sur l'intervalle `[a, b]` simulée par la fonction `rand`.
    Le titre de la figure devra être `Loi uniforme sur [`*va*`,`*vb*`] : distribution approchée` où *va* et *vb* désignent les valeurs entrées au clavier par l'utilisateur.

III. Simulation de v.a.r. suivant une loi uniforme discrète

III.1. Loi uniforme sur $[\![n1, n2]\!]$

On considère la fonction suivante.

```
function y = unifDiscrete(n1,n2)
    y = n1 + floor((n2-n1+1) ⋆ rand())
endfunction
```

- Que réalise cette fonction ?

- Écrire un programme qui :
  - × demande à l'utilisateur d'entrer au clavier deux entiers `n1` et `n2` et un entier `N`,
  - × stocke `N` appels successifs à la fonction `unifDiscrete` dans un vecteur ligne `U`,
  - × affiche le diagramme des fréquences d'un échantillon de taille `N` de v.a.r. suivant la loi uniforme sur l'intervalle $[\![$`n1, n2`$]\!]$ simulée par la fonction `rand`.
    Le titre de la figure devra être `Loi uniforme sur [|`$vn1$`,`$vn2$`|] : distribution approchée` où $vn1$ et $vn2$ désignent les valeurs entrées au clavier par l'utilisateur.

- Évaluer dans la console `X = tabul(U)` puis `Y = tabul(U,"i")`. À quoi servent ces commandes ?

- Évaluer alors `clf(); Y = tabul(U,"i"); bar(Y(:,1), Y(:,2))`.
  Expliquer le fonctionnement de la fonction `bar` en détaillant le contenu de `Y(:,1)` et `Y(:,2)`.

- Commenter l'intérêt de l'option `width` en évaluant `clf(); bar(Y(:,1), Y(:,2), width=0.1)`.