TP 1 : Simulations de lois

On utilisera le logiciel **Python**.

Soit $\mathcal{L}$ une loi, soit F sa fonction de répartition. L'inverse généralisé de la fonction de répartition F est la fonction F_g^{-1} définie, sur $]0,1[$, par

$$F_g^{-1}(x) = \inf\{t \in \mathbb{R} : F(t) > x\}.$$

C'est la fonction réciproque de F lorsque F est continue et strictement croissante.

Si $U \sim \mathcal{U}([0,1])$, alors $X = F_g^{-1}(U)$ suit la loi $\mathcal{L}$.

On peut utiliser ce fait pour simuler des échantillons de certaines variables aléatoires, en partant d'une réalisation d'un échantillon $(U_1, \ldots, U_n)$, de loi uniforme sur $[0,1]$, donnée par l'instruction `rand`. C'est possible lorsqu'on peut calculer et inverser la fonction de répartition de la loi en question. D'autres méthodes existent, pour la loi normale par exemple, ou la loi de Dirichlet.

La section 1 donne la preuve (sous forme d'exercices) des propriétés énoncées ci-dessus. La section 2 constitue le TP proprement dit, dans lequel il est toutefois nécessaire de se munir de papier et d'un crayon pour certaines étapes.

1 Inverse généralisé de la fonction de répartition

Soit F une fonction définie sur $\mathbb{R}$, à valeurs dans $[0,1]$, croissante, continue à droite en tout point et telle que

$$\lim_{-\infty} F(t) = 0, \quad \lim_{+\infty} F(t) = 1.$$

On pose

$$\forall x \in]0,1[, \quad F_g^{-1}(x) = \inf\{t \in \mathbb{R} : F(t) > x\}.$$

Ces exercices montrent que cette fonction F_g^{-1} est bien définie, en donnent les propriétés et une application.

Exercice 1 On considère une fonction F ayant les propriétés ci-dessus et l'on veut définir F_g^{-1}.

1. Pour $x \in]0,1[$, on pose $A_x = \{t \in \mathbb{R} : F(t) > x\}$. Démontrer que A_x n'est pas vide et qu'il est minoré. En déduire que sa borne inférieure est bien définie et que c'est un réel (pas $\pm\infty$).
2. Pour $x \in]0,1[$, démontrer que
$$]F_g^{-1}(x), +\infty[\subset A_x \subset [F_g^{-1}(x), +\infty[.$$
3. Pour $x \in]0,1[$, démontrer que $x < F(t)$ implique que $F_g^{-1}(x) \leq t$.
4. Pour $x \in]0,1[$, démontrer que $F_g^{-1}(x) \leq t$ implique que $x \leq F(t)$.

Exercice 2

1. Soit $([0,1], \mathcal{B}([0,1]), \lambda)$, où $\mathcal{B}([0,1])$ est la tribu borélienne sur $[0,1]$ et λ, la mesure de Lebesgue. Soit U la fonction définie sur $[0,1]$ par : $\forall x \in [0,1]$, $U(x) = x$. Vérifier que $U \sim \mathcal{U}([0,1])$.
2. On considère maintenant une fonction F ayant les propriétés ci-dessus et son inverse généralisée F_g^{-1}.

(a) Soit U une variable aléatoire de loi $\mathcal{U}([0,1])$, ce qui existe vu la question précédente. On pose $X = F_g^{-1}(U)$. Démontrer que X a pour fonction de répartition la fonction F. On pourra utiliser l'exercice précédent.

(b) En déduire qu'une fonction ayant les propriétés de F est la fonction de répartition d'une variable aléatoire.

(c) Application : Trouver une fonction g telle que $g(U)$ suive une loi exponentielle de paramètre $\lambda > 0$.

2 Simulations de lois

Exercice 3 Lois uniformes

Vous pouvez faire cet exercice en ligne de commande ou faire une fonction ayant pour arguments la longueur de l'échantillon et les bornes de l'intervalle.

1. Lire l'aide des instructions rand, plot, subplot, hist, cumsum, shape.
2. Simuler une réalisation d'une variable aléatoire de loi $\mathcal{U}([0,1])$ en utilisant l'instruction rand.
3. Simuler une réalisation d'une variable aléatoire de loi $\mathcal{U}([-5,3])$. On pourra penser qu'un changement de variable affine permet d'envoyer l'intervalle $[0,1]$ sur l'intervalle $[-5,3]$.
4. Simuler un **échantillon** de taille $N = 100$ de la loi $\mathcal{U}([0,1])$, $X1 = (X_1(\omega), \ldots, X_N(\omega))$, c'est-à-dire simuler N variables aléatoires indépendantes de même loi $\mathcal{U}([0,1])$, prises en un point ω de Ω.
5. Comment obtenir le vecteur $Z1$ de coordonnées
$$(X_1(\omega), \frac{1}{2}(X_1(\omega) + X_2(\omega)), \ldots, \frac{1}{N}\sum_{k=1}^{N} X_k(\omega)) \ ?$$
Ce vecteur contient les moyennes empiriques successives. *On essaiera, dans cette question comme ailleurs, de faire le moins de boucles possible.*
6. Simuler un échantillon de taille $N = 100$ de la loi $\mathcal{U}([3,6])$, $X2 = (X_1(\omega), \ldots, X_N(\omega))$.
7. Calculer $Z2$ (des moyennes empiriques pour $X2$).
8. Tracer $Z1, Z2$. Vers quoi convergent, apparemment, les suites $(Z1(n))_n, (Z2(n))_n, ?$
9. Sauvez votre figure.

Exercice 4 Simulation de lois. Moyenne empirique. Tracés

Soit $\mathcal{L}$ la loi de Bernoulli de paramètre $p \in]0,1[$. Soit N un entier.

1. Ecrire une fonction, d'arguments entrants N et p, qui fait les opérations suivantes :
   (a) Simule un échantillon de taille N de la loi $\mathcal{L}$, $X = (X_1(\omega), \ldots, X_N(\omega))$.
   (b) Trace la trajectoire de X (la courbe affine par morceaux joignant les points $(k, X_k(\omega))$).
   (c) Calcule $Z = (\overline{X_1}(\omega), \ldots, \overline{X_N}(\omega))$, où $\overline{X_k}(\omega) = \dfrac{1}{k}\displaystyle\sum_{1}^{k} X_s(\omega)$.
   (d) Trace l'histogramme correspondant à Z, la trajectoire de Z. Quelle propriété illustre-t-on ainsi ?
2. Mêmes questions avec une loi exponentielle de paramètre $\lambda > 0$.

On peut tracer la trajectoire et l'histogramme sur des figures successives ou dans deux “cases” d'une même figure. Légende et titre sont aussi laissés à votre choix

Exercice 5 Loi de Cauchy

Soit f la fonction définie sur $\mathbb{R}$ par
$$f(x) = \frac{1}{\pi}\frac{1}{1+x^2}.$$

1. Questions à faire chez vous mais à lire ici :
   (a) Prouver que f est bien une densité.
   (b) Si X est une variable aléatoire de densité f, vérifier que X n'admet pas de moment d'ordre 1 (ni d'ordre 2).
2. Ecrire une fonction, d'arguments entrants N et p, qui fait les opérations suivantes :
   (a) Simule un échantillon de taille N de la loi de Cauchy $X = (X_1(\omega), \dots, X_N(\omega))$.
   (b) Trace la trajectoire de X (la courbe affine par morceaux joignant les points $(k, X_k(\omega))$).
   (c) Calcule $Z = (\overline{X_1}(\omega), \dots, \overline{X_N}(\omega))$, où $\overline{X_k}(\omega) = \dfrac{1}{k}\displaystyle\sum_1^k X_s(\omega)$.
   (d) Trace l'histogramme correspondant à Z, la trajectoire de Z. Que constate-t-on ? Comment proposez-vous de l'expliquer ?

Exercice 6 Simulation de lois discrètes à valeurs dans $\mathbb{N}$.

1. Lire l'aide des instructions `floor`, `round`, `ceil`, `mod`, `divmod`.
2. Grâce à ces instructions, simuler une loi uniforme $\mathcal{U}(10)$, c'est-à-dire la loi d'une variable aléatoire prenant les valeurs $1, 2 \dots, 9, 10$ avec la probabilité $1/10$.
3. Ecrire une fonction renvoyant un échantillon de taille N, suivant une loi binomiale $b(n, p)$, avec $n \in \mathbb{N}^*$ et $p \in]0, 1[$. Les arguments de la fonction doivent être N, n, p. (*Indication :* comment obtient-on une v.a. binomiale à partir de v.a. de Bernoulli ?)

Exercice 7 Loi binomiale négative

Soit une expérience $\mathcal{E}$ dont le succès est décrit par une loi de Bernoulli $\mathcal{B}(p), 0 < p < 1$. On répète cette expérience jusqu'à ce qu'elle donne $r(\in \mathbb{N}^*)$ succès et l'on note le nombre d'essais N qui a été nécessaire, en supposant les expériences successives indépendantes. Par exemple, un représentant doit placer r contrats, un par foyer au plus, avant d'arrêter ses démarchages.
La v.a. N suit une loi appelée binomiale négative de paramètres r et p. Les valeurs prises par N sont $r, r+1, \dots$ (les entiers $\geq r$) et, si $k \geq 0$,

$$P(N = r + k) = p^r (1-p)^k \binom{r+k-1}{r-1}.$$

On peut démontrer que N suit la même loi que la somme de r variables indépendantes $X_1, \dots, X_r$, suivant chacune une loi $\mathcal{G}^*(p)$ c'est-à-dire une loi géométrique à valeurs dans $\mathbb{N}^*$.

1. Si $X \sim \mathcal{G}^*(p)$, rappeler $P(X = k)$ pour $k \in \mathbb{N}^*$ et calculer $S_n = \sum_{k=1}^n P(X = k)$.
2. On considère que $[0, 1]$ est subdivisé en intervalles successifs de longueur $P(X = k)$, délimités par les S_n. Soit $u \in]0, 1[$. Déterminer n tel que $u \in [S_n, S_{n+1}[$.
3. Faire une fonction qui renvoie un échantillon de taille N suivant une loi $\mathcal{G}^*(p)$, les arguments entrants sont N, p.
4. Simuler un échantillon suivant une loi binomiale négative.

Exercice 8 Loi de Pareto

Soit $p > 1$, soit la densité de Pareto

$$f_p(x) = \frac{p-1}{x^p} \mathbf{1}_{[1,+\infty)}(x).$$

1. Pourquoi p est-il supposé > 1 ?
2. Déterminer à la main la fonction de répartition F_p d'une v.a. X suivant la loi de Pareto de paramètre p et son inverse généralisée G_p.
3. Écrire un programme ou une fonction (`simpareto`) qui effectue les actions suivantes :

(a) simule un échantillon $(X_1, \dots, X_N)$ de N v.a. suivant la loi de Pareto de paramètre p.

(b) trace la trajectoire de $(X_1, \dots, X_N)$,

(c) calcule les moyennes empiriques, c.a.d. le vecteur $(Y_1, \dots, Y_N)$, où $Y_k = \frac{1}{k}\sum_{s=1}^{k} X_s$ et en trace la trajectoire.

Ce programme ou cette fonction doit avoir pour données les nombres p et N.

4. Lancer plusieurs fois ce programme avec $p = 3/2$ puis $p = 3$ et des N de votre choix. Remarquez-vous une différence et si oui, comment l'expliquez-vous ?

Exercice 9 Simulation de lois, maximum

Soit $(X_1, \dots, X_n)$ une famille de n v.a.i.i.d.. On note F_X la fonction de répartition commune à toutes les X_i et l'on pose

$$Y = \max(X_1, \dots, X_n).$$

1. Déterminer la fonction de répartition F_Y de Y en fonction de F_X. Justifiez. (Ce n'est pas une question de programmation.)
2. Exprimer l'inverse généralisée de F_Y en fonction de celle de F_X. (Ce n'est pas une question de programmation.)
3. Ici, les variables X suivent une loi exponentielle de paramètre $\lambda > 0$. Si $Y = \max(X_1, \dots, X_n)$, on dira que Y suit la loi $\mathcal{L}_n$.

   (a) Rappeler l'inverse généralisée de la fonction de répartition d'une v.a. de loi $\mathcal{E}(\lambda)$.

   (b) Faire une fonction maxexp, ayant pour données l'entier $n > 1$, λ et l'entier $T \in \mathbb{N}^*$ (taille de l'échantillon que l'on souhaite fabriquer), qui effectue les actions suivantes :
   — simule un échantillon $(Y_1, \dots, Y_T)$ de v.a. de loi $\mathcal{L}_n$;
   — en calcule les moyennes empiriques, c.a.d. le vecteur $(\bar{Y}_1, \dots, \bar{Y}_T)$, où $\bar{Y}_k = \frac{1}{k}\sum_{s=1}^{k} Y_s$;
   — trace la trajectoire des moyennes empiriques ;
   — renvoie comme résultat res la dernière moyenne empirique $\bar{Y}_T$.

   (c) Choisissez $n > 1$, λ et T assez grand puis lancez plusieurs fois ce programme.
   — Précisez $n =$, $\lambda =$, $T =$.
   — Que remarque-t-on ? Quels res obtenez-vous ?
   — Quel résultat fondamental cette simulation illustre-t-elle ?

Exercice 10 Loi de Dirichlet

Soit $(X_1, \dots X_n)$ une famille de n v.a.i.i.d. de loi uniforme $\mathcal{U}([a,b])$, définies sur un espace $(\Omega, \mathcal{T}, P)$. On définit la v.a. $X_<$, à valeurs dans $\mathbb{R}^n$, par

$$X_<(\omega) = (X_{i_1}(\omega), X_{i_2}(\omega), \dots, X_{i_n}(\omega)),$$

où l'on a choisi les indices de sorte que $X_{i_1}(\omega) \leq X_{i_2}(\omega) \leq X_{i_n}(\omega))$. On classe les coordonnées de X par ordre croissant.

On peut prouver que $X_<$ suit une loi de densité

$$(x_1, \dots, x_n) \mapsto \frac{n!}{(b-a)^n} \mathbf{1}_{a < x_i < \cdots < x_n < b}.$$

Cette loi est la loi de Dirichlet $D_n([a,b])$. C'est la loi que suit un échantillon uniforme quand il a été réordonné.

Simuler un échantillon de taille N d'une loi de Dirichlet $D_n([a,b])$. On pourra utiliser l'instruction sort.