

Feuille 2 : simulations stochastiques

Remarque. Les exercices ou questions "étoilés" * sont généralement un peu plus difficiles ou simplement moins importants. Ils ne seront pas nécessairement traités en séance TD.

Exercice 1. Sur le théorème d'inversion

Le théorème d'inversion utilise la notion d'inverse généralisé F_g^{-1} d'une fonction de répartition $F : \mathbb{R} \rightarrow [0, 1]$ (encore appelée fonction quantile) d'une v.a. $X : \Omega \rightarrow \mathbb{R}$:

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

† Le but de cet exercice est de montrer que l'inverse généralisé permet en théorie de simuler n'importe quelle loi à partir de la loi uniforme $\mathcal{U}_{]0,1[}$ (Théorème d'inversion).

1. Si F est continue sur $\mathbb{R}$ et strictement croissante sur $X(\Omega)$ lorsque $X(\Omega)$ est un intervalle ouvert de $\mathbb{R}$ (borné ou non), montrer que $F_g^{-1}(y) = F^{-1}(y)$, $\forall y \in]0, 1[$.
(Indication. Montrer que $\mathcal{F} := \{x \in \mathbb{R} : F(x) \geq y\} = [F^{-1}(y), +\infty[$ par double inclusion)
2. Montrer que l'inverse généralisé possède la propriété de monotonie suivante.

$$\forall x \in \mathbb{R}, \forall y \in]0, 1[, \quad F(x) \geq y \Leftrightarrow x \geq F_g^{-1}(y) \quad (1)$$

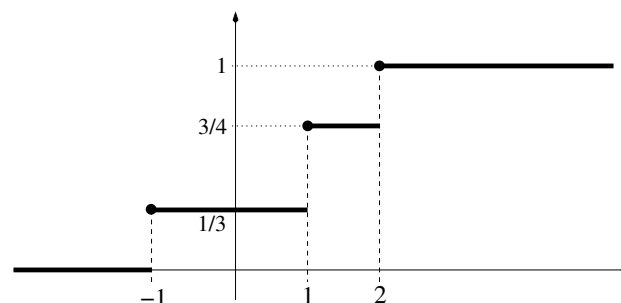
On rappelle que la fonction de répartition F est croissante sur $\mathbb{R}$ et qu'elle est continue à droite i.e. $\lim_{\substack{\epsilon \rightarrow 0 \\ \epsilon > 0}} F(x + \epsilon) = F(x)$.

3. (*) En déduire que l'inverse généralisé est croissant.
4. Soit U une v.a. de loi uniforme sur $]0, 1[$ i.e. $U \sim \mathcal{U}_{]0,1[}$. Montrer que la v.a. réelle

$$V = F_g^{-1}(U)$$

a pour fonction de répartition F i.e. V a même loi que X . En déduire une méthode pour engendrer des réalisations $x_1, \dots, x_n$ d'une v.a. X dont on connaît l'inverse généralisé.

5. Soit la fonction de répartition F suivante :



- (a) Quelle est la loi correspondante ?
- (b) Calculer l'inverse généralisé F_g^{-1} .
- (c) Donner la méthode pour engendrer des réalisations de la loi représentée par F .
- (d) (*) Mêmes questions pour une loi discrète finie quelconque.

†. Rappel. Pour un ensemble $\mathcal{F} \subset \mathbb{R}$, l'infimum $m = \inf \mathcal{F}$ est défini par

1. $\forall x \in \mathcal{F}, m \leq x$
2. $\forall x \in \mathbb{R}$ tel que $m < x$ alors $\exists z \in \mathcal{F}$ tel que $m \leq z < x$

6. En utilisant cette méthode :

- (a) Comment simuler la loi de Cauchy de paramètres $a > 0$ et $x_0 \in \mathbb{R}$ dont la densité et la fonction de répartition valent ($x \in \mathbb{R}$) :

$$f(x) = \frac{1}{\pi} \left[\frac{a}{(x - x_0)^2 + a^2} \right], \quad F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{x - x_0}{a}\right) \quad (2)$$

- (b) Comment simuler la loi de Rayleigh de paramètre 1, notée $\mathcal{R}(1)$ dont la densité et la fonction de répartition valent ($r \in \mathbb{R}$) :

$$f_R(r) = re^{-\frac{1}{2}r^2} \mathbb{1}_{\mathbb{R}^+}(r), \quad F(r) = (1 - e^{-\frac{1}{2}r^2}) \mathbb{1}_{\mathbb{R}^+}(r) \quad (3)$$

Rmq : dans le cours, on a vu que les générateurs de $\mathcal{U}_{[0,1]}$ ne sortent jamais 1 et souvent jamais 0. Dans ce dernier cas, en appelant `rand()` la fonction nous donnant des réalisations indépendantes de $\mathcal{U}_{[0,1]}$, il est possible de remplacer `1-rand()` par `rand`, évitant ainsi une soustraction. Pourquoi ?

Exercice 2*. Génération de $\mathcal{N}(0, 1)$ par la transformation de Box-Muller (1958)

On rappelle (cf. MAP B) que si $Y \sim \mathcal{N}(\mu, \sigma^2)$ alors $X := (Y - \mu)/\sigma \sim \mathcal{N}(0, 1)$ (et inversement). Il suffit ainsi de se focaliser sur la génération de nombres aléatoires suivant $\mathcal{N}(0, 1)$.

Pour cet exercice, nous aurons besoin du résultat suivant, qui est le théorème de transfert pour les couples de v.a. à densité. Si pour toute fonction $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$ bornée et mesurable, on a :

$$E(\varphi(X, Y)) = \iint_{\mathbb{R}^2} \varphi(x, y) f(x, y) dx dy$$

alors f est la densité du couple (X, Y) et réciproquement.

Soit le couple aléatoire (R, Θ) formée par deux v.a. indépendantes avec $R \sim \mathcal{R}(1)$ (loi de Rayleigh, cf exercice précédent) et $\Theta \sim \mathcal{U}_{[0, 2\pi]}$:

1. Quelle est la densité de (R, Θ) ?
2. On pose $(X, Y) := h(R, \Theta) = (R \cos(\Theta), R \sin(\Theta))$. Quelle est la densité de (X, Y) ? Aide : utiliser la méthode générale $E(\varphi(X, Y)) = E((\varphi \circ h)(R, \Theta))$ et exprimer cette dernière expression à l'aide de la formule de transfert. Dédurre du résultat obtenu que $X \perp Y$ et que $X, Y \sim \mathcal{N}(0, 1)$.
3. Comment simuler les v.a. R et Θ à l'aide de deux v.a. U_1 et U_2 indépendantes de loi $\mathcal{U}_{[0,1]}$? En déduire une méthode qui, à partir de deux nombres sortis par `rand` (c'est à dire deux réalisations indépendantes de $\mathcal{U}_{[0,1]}$), permet d'obtenir deux réalisations indépendantes de $\mathcal{N}(0, 1)$.

Remarque : il existe une autre version plus efficace (appelée “forme polaire”) qui ne requiert pas l'utilisation des fonctions `cos` et `sin` au prix de consommer un peu plus de nombres issus de $\mathcal{U}_{[0,1]}$. Néanmoins l'algorithme le plus efficace pour simuler $\mathcal{N}(0, 1)$ est basé sur la méthode “ziggurat” maintenant couramment utilisée dans les bibliothèques.

Exercice 3. Générateurs de nombres aléatoires LCG

Soit la suite s_n engendrée par le LCG $s_{n+1} = as_n + c \pmod m$.

1. Ecrire un petit script python pour calculer[‡] la suite (s_n) . Déterminer la période pour le générateur ($a = 5, c = 1, m = 12$) avec $x_0 = 0$.
2. Pour le générateur ($a = 13, c = 1, m = 24$), montrer que la période est maximale. Vérifier avec votre script.
3. On choisit $m = 2^e$, $e \geq 2$ entier. A partir du théorème de Hull-Dobell, montrer que si c est impair et $a \equiv 1 \pmod 4$ alors la période du générateur est maximale.

‡. L'opérateur modulo (reste de la division euclidienne) est donné en python par `%`.