

2. Méthodes de simulation de variables réelles

Soit X une variable aléatoire réelle de loi $\mathbb{P}_X$. La fonction de répartition

$$F_X(t) = \mathbb{P}_X[(-\infty, t]] = \mathbb{P}[X \leq t]$$

caractérise la loi de X . En effet, elle détermine les valeurs de $\mathbb{P}_X$ sur la classe stable par intersection des intervalles $(-\infty, t]$, et la tribu engendrée par ces intervalles est la tribu des boréliens, donc par le lemme de classe monotone, F_X détermine $\mathbb{P}_X$. Remarquons par ailleurs que :

- (1) La fonction de répartition est croissante de $\mathbb{R}$ vers $[0, 1]$.
- (2) Cette fonction est continue à droite en tout point : $\lim_{s \rightarrow t, s > t} F_X(s) = F_X(t)$.
- (3) On a $\lim_{t \rightarrow -\infty} F_X(t) = 0$ et $\lim_{t \rightarrow +\infty} F_X(t) = 1$.

Nous verrons plus loin que toute fonction avec ces trois propriétés est la fonction de répartition d'une variable aléatoire réelle.

1. Simulation par inversion.

Soit $F : \mathbb{R} \rightarrow [0, 1]$ une fonction croissante et continue à droite, de limites 0 et 1 en $\pm\infty$. On appelle *inverse généralisée* de F la fonction $F^{-1} : (0, 1) \rightarrow \mathbb{R} \sqcup \{\pm\infty\}$ définie par

$$F^{-1}(u) = \inf(\{t \mid F(t) \geq u\}),$$

avec par convention $\inf(\emptyset) = +\infty$. Si la fonction F est continue strictement croissante, alors F^{-1} est simplement la bijection réciproque de F .

- (1) Pour $0 < u < 1$, notons

$$A_u = \{t \mid F(t) \geq u\}.$$

Montrer que $A_u \neq \emptyset$ et que $A_u \neq \mathbb{R}$. Si $s \in A_u$ et $t > s$, montrer que $t \in A_u$. En déduire que $F^{-1}(u) \in \mathbb{R}$ et que $A_u = [F^{-1}(u), +\infty)$.

- (2) Soit U une variable de loi uniforme sur $[0, 1]$. Calculer la fonction de répartition de la variable aléatoire $F^{-1}(U)$. En déduire que les trois propriétés listées au début du chapitre caractérisent les fonctions $F : \mathbb{R} \rightarrow [0, 1]$ qui sont des fonctions de répartition.
- (3) Soit X une variable aléatoire à valeur réelle et F_X sa fonction de répartition. Pour U uniforme sur $[0, 1]$, quelle est la loi de $F_X^{-1}(U)$? En déduire une méthode pour simuler une variable aléatoire qui a la même loi que X .
- (4) Retrouver la méthode de simulation des variables exponentielles de paramètre λ .
- (5) La *loi de Cauchy* est la loi à densité

$$f(x) = \frac{1}{\pi(1+x^2)} dx$$

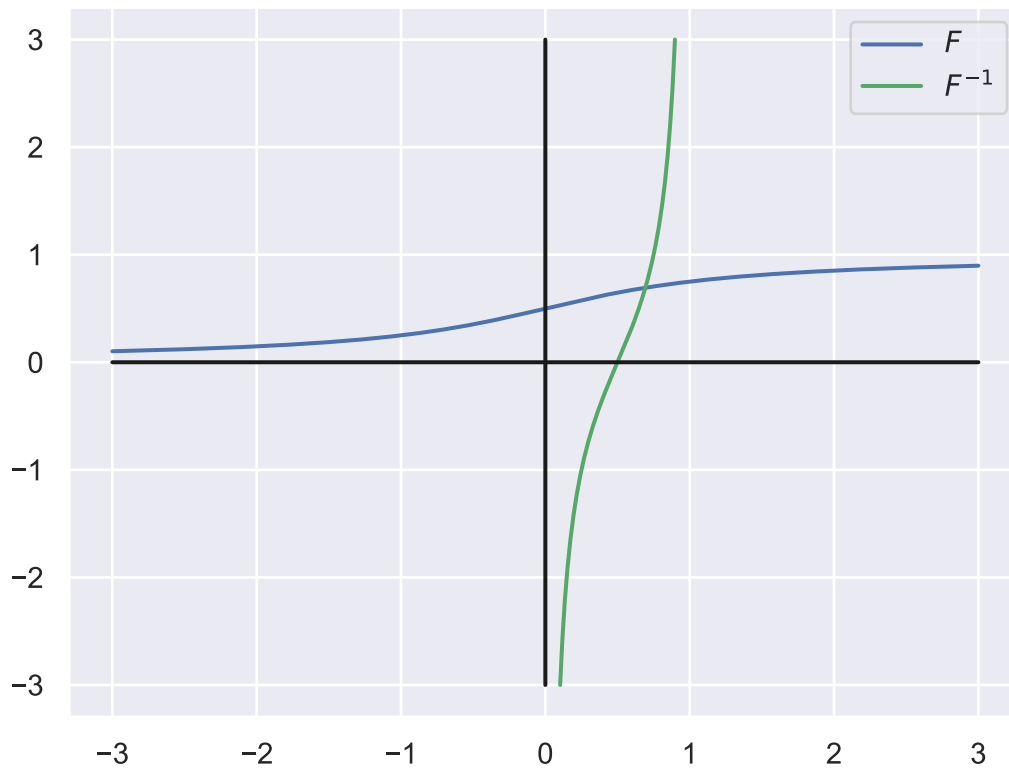


FIG. 2.1. Inverse d'une fonction de répartition.

sur $\mathbb{R}$. Calculer la fonction de répartition correspondante $F(x)$, puis son inverse généralisée $F^{-1}(x)$. Écrire un programme qui simule des variables aléatoires suivant la loi de Cauchy.

- (6) Trouver m tel que si X suit la loi de Cauchy, alors $\mathbb{P}[|X| \geq m] \leq 0.05$. Dessiner sur un même graphique la fonction de répartition empirique d'un échantillon de $N = 10, 100, 1000$ variables de Cauchy indépendantes, et la fonction $F(x)$ calculée à la question précédente.

La méthode de *simulation par inversion* permet ainsi d'engendrer n'importe quelle variable aléatoire réelle dont la fonction de répartition est calculable, ainsi que son inverse. Pour traiter les cas où ceci n'est pas possible, on combinera cette méthode avec la méthode dite du *rejet*.

2. Méthode de rejet pour des variables à support compact.

Soit X une variable aléatoire dont la loi à un support compact inclus dans un intervalle $[a, b]$, et a une densité bornée $f : [a, b] \rightarrow [0, K]$.

- (1) Écrire un programme `uniforme(a, b)` qui simule une variable de loi $\text{Unif}([a, b])$, et donc de densité

$$g(x) = \frac{1_{a \leq x \leq b}}{b - a}.$$

Écrire ensuite un programme `point_uniforme(a, b, K)` qui renvoie un couple (X, Y) avec X et Y indépendants, $X \sim \text{Unif}([a, b])$ et $Y \sim \text{Unif}([0, K])$.

- (2) On considère le cas où $[a, b] = [1, 2]$ et $f(x) = C(x^2 - 1) 1_{(1 \leq x \leq 2)}$. Calculer C pour que f soit une densité de probabilité, puis K pour que f prenne ses valeurs dans $[0, K]$.
- (3) Écrire un programme `nuage_points(N, f, a, b, K)` qui prend en argument une densité $f : [a, b] \rightarrow [0, K]$, tire au hasard N points indépendants $(X_1, Y_1), \dots, (X_N, Y_N)$ dans le rectangle $[a, b] \times [0, K]$, et les dessine, en mettant en bleu les points sous la courbe $y = f(x)$, et en vert les points au-dessus de la courbe. On pourra utiliser la commande `ax.scatter(X, Y, c=list_of_colors)` pour dessiner un nuage de points d'abscisses contenues dans le vecteur X , et d'ordonnées contenues dans le vecteur Y . Utiliser ce programme avec $N = 200$ et la fonction f de la question précédente.

Dans le contexte de la question précédente, on note :

$$M = \inf(\{i \geq 1 \mid f(X_i) > Y_i\})$$

et on dit que (X_M, Y_M) est le *point accepté*.

- (4) Quelle est la loi de M ? Calculer ensuite

$$\mathbb{P}[M = n \text{ et } X_M \leq t]$$

pour un paramètre $t \in [a, b]$. En déduire que M et X_M sont indépendants, et que X_M a la loi de densité f .

- (5) Écrire un programme `nuage_points_accept(N, f, a, b, K)` qui simule N points acceptés, les dessine dans le plan (ainsi que le graphe de la fonction f), et renvoie le vecteur formé des N abscisses de ces points. Tester ce programme avec la même fonction f que précédemment et $N = 1000$. Dessiner sur un même graphique la fonction de répartition empirique du vecteur obtenu, et la fonction de répartition théorique F correspondant à f .

On a ainsi une méthode générale de simulation pour toute loi à densité bornée et à support compact. La troisième section étend ce principe, en retirant ces deux hypothèses sur la densité.

3. Méthode de rejet dans le cas général.

Soient f et g deux densités de probabilité sur $\mathbb{R}$, telles que $f(x) \leq \frac{1}{p} g(x)$ pour un certain $p \in (0, 1)$ et tout x dans $\mathbb{R}$. On tire au hasard $X_1, X_2, \dots$ indépendants de densité g (par exemple avec la méthode d'inversion), puis $Y_1, Y_2, \dots$ définis par :

$$Y_i = \frac{1}{p} g(X_i) U_i,$$

où les U_i sont variables aléatoires indépendantes entre elles et indépendantes des variables X_i , de loi uniforme sur $[0, 1]$. Remarquons que la construction de la section précédente correspondait au cas où $g(x) = \frac{1_{a \leq x \leq b}}{b-a}$ et $p = \frac{1}{K(b-a)}$.

On pose comme précédemment :

$$M = \inf(\{i \geq 1 \mid f(X_i) > Y_i\}).$$

- (1) En toute généralité, montrer que M suit une loi géométrique de paramètre p , et que X_M est indépendant de M , et a la loi de densité f .
- (2) Écrire un programme `nuage_points_2(N, f, g, p, gen_g)` qui :
- prend en paramètre un entier N , deux densités f et g , un paramètre $p \in (0, 1)$ tel que $f(x) \leq \frac{1}{p} g(x)$, et un générateur de variables aléatoires sous la densité g .

- tire au hasard N points $(X_1, Y_1), \dots, (X_N, Y_N)$ et les dessine dans le plan, en mettant en bleu les points sous la courbe $y = f(x)$ et en vert les points au-dessus de cette courbe. On mettra également sur le dessin les graphes des fonctions $f(x)$ et $\frac{1}{p}g(x)$.
 - renvoie le vecteur formé des abscisses des points (X_i, Y_i) acceptés, c'est-à-dire tels que $f(X_i) > Y_i$.
- (3) On pose $g(x) = \frac{1}{2} e^{-|x|}$. Montrer que si $E \sim \text{Exp}(1)$ et S est une variable indépendante de E , à valeurs dans $\{\pm 1\}$ et avec $\mathbb{P}[S = 1] = \mathbb{P}[S = -1] = \frac{1}{2}$, alors $X = ES$ suit la loi de densité g .
- (4) Soit

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

la densité de la *loi normale (ou gaussienne)* $N(0, 1)$. Trouver un paramètre p tel que $f(x) \leq \frac{1}{p}g(x)$, et utiliser le programme `nuage_points_2` avec ces fonctions, ce paramètre p , un générateur adéquat pour la loi g , et $N = 500$.

- (5) Modifier le programme `nuage_points_2` pour que le paramètre N donne le nombre de points acceptés du résultat. Dessiner sur un même graphique la fonction de répartition empirique du vecteur obtenu (toujours avec $N = 500$ et les fonctions f et g ci-dessus), et la fonction de répartition théorique de la gaussienne, qui est donnée par la commande `scs.norm.cdf` (appliquée à un *array numpy* de valeurs).

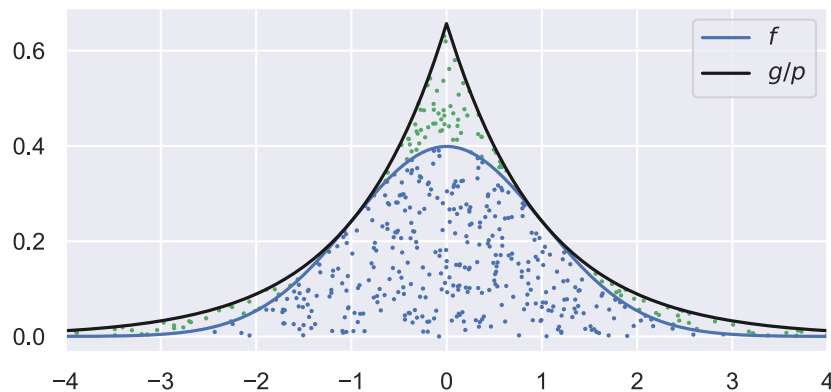


FIG. 2.2. Méthode de rejet : on tire des abscisses aléatoires de loi à densité g , des ordonnées aléatoires $Y_i = \frac{1}{p}g(X_i)U_i$, et on accepte les abscisses des points placés sous la courbe $y = f(x)$.