

TP2 : Autour de la fonction de répartition

4 Fonction de répartition

Soit (Ω, \mathcal{F}) un espace mesurable et \mathbb{P} une mesure de probabilité sur (Ω, \mathcal{F}) . On munit \mathbb{R} de la tribu des boréliens $\mathcal{B}(\mathbb{R})$. Une variable aléatoire réelle X est alors une application mesurable $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Dans ce cas, on appelle loi de la variable aléatoire la mesure image de \mathbb{P} par l'application X :

$$\forall A \in \mathcal{B}(\mathbb{R}), \quad P^X(A) = \mathbb{P}(X \in A) = \mathbb{P}(\{\omega : X(\omega) \in A\})$$

Définition 1. On appelle fonction de répartition de la variable aléatoire X l'application $F_X : \mathbb{R} \rightarrow [0; 1]$ définie par

$$\forall x \in \mathbb{R}, \quad F_X(x) = P^X(]-\infty, x]).$$

Le théorème suivant donne les propriétés fondamentales de ces fonctions.

Théorème 1. 1. La fonction de répartition F_X est croissante, de limites 0 en $-\infty$ et 1 en $+\infty$. Elle est continue à droite en tout point.

2. La fonction F_X ne dépend que de la loi de X , et elle détermine cette loi (c'est-à-dire que $F_X = F_Y$ implique que X et Y ont même loi).

Exercice 6. —

- Familiarisez-vous avec la fonction `np.sort(A)` qui retourne un tableau qui contient les entrées de `A` triées dans l'ordre croissant. Alternativement, on peut trier sur place (en modifiant donc `A`) en appelant `A.sort()` (`A` doit alors être un `array numpy` ou une liste).
- Familiarisez-vous avec la fonction `plt.step(X, Y)` pour représenter des fonctions constantes par morceaux.
 - Quelle est la fonction de répartition d'une variable X de loi de Bernoulli $\mathcal{B}(p)$? Utiliser la fonction `plt.step` pour la représenter.
 - Quelle est la fonction de répartition d'une variable d'une variable Z de loi uniforme sur l'ensemble d'entiers $\{1, 2, 3, 4, 5, 6\}$? Utiliser la fonction `plt.step` pour la représenter.
 - Quelle est la fonction de répartition d'une variable d'une variable Z de loi uniforme sur $[0; 1]$? Utiliser cette fois-ci la fonction `plt.plot` pour la représenter.
- Soit $n \in \mathbb{N}^*$ et X une variable uniforme sur un ensemble $\{x_1 < x_2 < \dots < x_n\}$ (c'est-à-dire que pour tout i , $\mathbb{P}(X = x_i) = \frac{1}{n}$). Que vaut $F_X(t)$ pour t dans l'intervalle $[x_i; x_{i+1}[$? Montrer que pour tout t ,

$$F_X(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{x_i \leq t}.$$

- Soit (X_1, \dots, X_n) un échantillon de variables aléatoires de même loi (c'est-à-dire que les X_i sont indépendants et de même loi). On cherche à tracer sa fonction de répartition empirique définie par

$$\widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X_i \leq x}.$$

Soient $(X^{(1)}, \dots, X^{(n)})$ les valeurs de l'échantillon dans l'ordre croissant. Montrer que \widehat{F}_n est une fonction constante par morceaux. Quels sont ses points de discontinuité ? Montrer que $\widehat{F}_n(x) = \frac{k}{n}$ pour $x \in [X^{(k)}, X^{(k+1)}[$, $k \in \{0, \dots, n\}$.

5. Écrire une fonction `fonction_repartition_empirique` qui prend comme argument un échantillon $X = (X_1, \dots, X_n)$, et qui trace sa fonction de répartition empirique en utilisant la fonction `plt.step`.
6. Tester la fonction avec des échantillons (X_1, \dots, X_n) où les X_i sont indépendants de loi uniforme sur $[0; 1]$, et avec des tailles variées (tester pour $n = 5$, $n = 15$ et $n = 100$). Qu'observez-vous ?
7. Superposer à ce graphe le graphe de la fonction de répartition de la loi uniforme sur $[0; 1]$. Qu'observez-vous ?

5 Méthode de simulation par inversion de la fonction de répartition

Définition 2. Si $F : \mathbb{R} \rightarrow [0; 1]$, on appelle inverse généralisée de F la fonction $F^{-1} :]0; 1[\rightarrow [-\infty, \infty]$ définie par :

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

avec la convention que $\inf \emptyset = +\infty$.

Exercice 7. — Soit $F : \mathbb{R} \rightarrow [0; 1]$ une fonction croissante, de limites 0 en $-\infty$ et 1 en $+\infty$, et continue à droite.

1. Pour $0 < u < 1$, on note

$$A_u = \{x : F(x) \geq u\}.$$

- (a) Montrer que A_u est non vide
 - (b) Montrer que A_u n'est pas \mathbb{R} .
 - (c) Montrer que si s est dans A_u et $t > s$, alors t est dans A_u .
 - (d) Montrer que A_u est de la forme $[a; +\infty[$.
 - (e) Montrer $F^{-1}(u)$ est dans \mathbb{R} et $A_u = [F^{-1}(u); \infty[$.
2. Soit U une variable de loi uniforme sur $[0; 1]$. Calculer la fonction de répartition de la variable aléatoire $F^{-1}(U)$.
 3. En déduire que F est la fonction de répartition d'une variable aléatoire si et seulement si $F : \mathbb{R} \rightarrow [0; 1]$ une fonction croissante, de limites 0 en $-\infty$ et 1 en $+\infty$, et continue à droite.
 4. 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 .

Exercice 8. — On rappelle que la loi exponentielle $\mathcal{E}(\lambda)$ de paramètre λ est la loi de densité $f(x) = \lambda e^{-\lambda x} \mathbf{1}_{]0, \infty[}(x)$.

1. Quelle est la fonction de répartition F_X d'une variable aléatoire X qui a cette loi ?
2. Que vaut alors F_X^{-1} ?
3. Écrire un programme qui simule une variable aléatoire de loi exponentielle de paramètre λ , à l'aide d'une variable aléatoire de loi uniforme sur $[0; 1]$.
4. Trouver m_λ tel que si X a pour loi $\mathcal{E}(\lambda)$, alors $\mathbb{P}(X > m_\lambda) = 0.05$.

- Tracer sur un même graphe et pour plusieurs tailles d'échantillons la fonction de répartition empirique d'un échantillon de variable aléatoire exponentielle de paramètre 1, et la fonction de répartition de la loi exponentielle. On prendra garde à représenter seulement la courbe sur un intervalle où la variable a de bonnes chances d'être présente, par exemple $[0; m_\lambda]$.

Exercice 9. — La loi de Cauchy est la loi de densité $f(x) = \frac{1}{\pi(1+x^2)}$ sur \mathbb{R} .

- Quelle est la fonction de répartition F_X d'une variable aléatoire X qui a cette loi ?
- Que vaut alors F_X^{-1} ?
- Simuler une variable aléatoire de loi de Cauchy, à l'aide d'une variable aléatoire de loi uniforme sur $[0; 1]$.
- Trouver m tel que si X a la loi de Cauchy, alors $\mathbb{P}(|X| > m) = 0.05$.
- Tracer pour plusieurs tailles d'échantillons la fonction de répartition empirique d'un échantillon de variables de loi de Cauchy, et la fonction de répartition de la loi de Cauchy. On prendra garde à représenter seulement la courbe sur un intervalle où la variable a de bonnes chances d'être présente, par exemple $[-m; m]$.

6 Vitesse de convergence dans Glivenko–Cantelli

Exercice 10. — On suppose que X_1, X_2, \dots est une suite de variables aléatoires i.i.d. de loi uniforme sur $[0; 1]$. D'après le théorème de Glivenko–Cantelli, la fonction de répartition empirique \widehat{F}_n converge presque sûrement uniformément vers la fonction $F(x) = x$ quand $n \rightarrow \infty$. On souhaite illustrer cette convergence.

- Montrer que

$$\|\widehat{F}_n - F\|_\infty = \max \left(\max_{k=1, \dots, n} \left\{ \frac{k}{n} - X^{(k)} \right\}, \max_{k=1, \dots, n} \left\{ X^{(k)} - \frac{k-1}{n} \right\} \right).$$

- Écrire une fonction `distance_loi_uniforme` qui calcule cette quantité pour un échantillon donné.
- Tracer $\|\widehat{F}_n - F\|_\infty$ en fonction de n pour $n = 1, \dots, N = 5000$. Superposer une ligne horizontale correspondant à l'axe des abscisses. Qu'observez-vous? Sans le théorème de Glivenko–Cantelli, peut-on affirmer après un seul tracé avec une certaine certitude que $\|\widehat{F}_n - F\|_\infty \rightarrow 0$ presque sûrement quand $n \rightarrow \infty$?
- On cherche à estimer la vitesse de convergence de $\mathbb{E}[\|\widehat{F}_n - F\|_\infty]$ vers 0. On stipule que cette convergence est polynomiale en n :

$$\mathbb{E}[\|\widehat{F}_n - F\|_\infty] \sim Cn^{-\rho},$$

pour un certain $\rho > 0$. Par conséquent, on prend $n \in \{2^4, 2^5, \dots, 2^{13}\}$. Pour chacune de ces valeurs de n , estimer $\mathbb{E}[\|\widehat{F}_n - F\|_\infty]$ en prenant la moyenne de 100 réalisations indépendantes de $\|\widehat{F}_n - F\|_\infty$. Puis, estimer C et ρ avec `np.polyfit`. Énoncer une conjecture concernant la valeur de ρ .

- Faire un tracé log-log de ces moyennes en fonction de n , et superposer un tracé log-log de $\widehat{C}n^{-\rho}$, où \widehat{C} est l'estimation de C ; pour ρ on utilisera sa valeur conjecturée.