TD et TP n° 3 : Convergences de variables aléatoires

Remarque : sur cette feuille, on pourra utiliser toutes les fonctionnalités de `scipy.stats`*, y compris pour la simulation de variables aléatoires.*

Exercice 1. *(Illustration de convergence p.s., TP seulement)*

1. Tirer un échantillon $(X_1, \dots, X_N)$ de v.a. de loi uniforme sur $[0,1]$, de taille $N = 10\,000$. On pose $S_k = X_1 + \ldots + X_k$ la somme partielle. Tracer S_k/k en fonction de k et en fonction de $\log_{10} k$ pour $k \in \{1, \dots, N\}$ (faire deux tracés, pour le deuxième on utilisera la commande `plt.xscale("log")`). On pourra utiliser la fonction `np.cumsum` pour le calcul de S_k. Ajouter sur ces graphiques la ligne horizontale $y = E[X_1]$ (avec `plt.axhline`).
2. Superposer $M = 5$ échantillons. Commenter.
3. Reprendre la question 1 avec X_1 variable aléatoire de Cauchy. Que constate-t-on ? Expliquer.

Exercice 2. *(Urne de Polya)* L'urne de Polya élémentaire est définie comme ceci : à l'instant initial $n = 0$, l'urne comprend deux boules, une de couleur blanche et une de couleur noire. Pour connaître la composition de l'urne à l'instant $n+1$, on tire une boule dans l'urne à l'instant n uniformément au hasard, et on la remet dans l'urne avec une autre boule de la même couleur à l'instant $n+1$. On s'intéresse alors à la proportion des boules blanches dans l'urne en temps long. On note X_n le nombre de boules blanches dans l'urne à l'instant n.

1. (TD) Montrer par récurrence que $P(X_n = j) = \frac{1}{n+1}, \quad j \in \{1, \dots, n+1\}$.
2. (TD) En déduire une convergence en loi de la proportion du nombre de boules blanches quand $n \to \infty$.
3. (TP) Écrire une fonction `Polya` qui prend comme argument le nombre d'étapes N à effectuer et renvoie une réalisation de la suite $(X_0, \dots, X_N)$.
4. (TP) Superposer quelques trajectoires de la proportion du nombre de boules blanches dans l'urne jusqu'à l'instant $N = 1000$. Que suggèrent ces trajectoires ? Y a-t-il convergence p.s. ?
5. (TP) Ecrire une fonction `Polya2` qui généralise la fonction `Polya` en permettant de pouvoir choisir une condition initiale générale avec a boules blanches et b boules noires. Refaire la question 4.
6. (TP) Etudier par simulation la convergence en loi de la proportion du nombre de boules blanches. Pour cela, on pourra créer $M = 1000$ réalisations de X_{25}, X_{50} et X_{100}, et superposer les 3 fonctions de répartitions empiriques.

Exercice 3. *(Fonction de répartition empirique)* Soit $X_1, X_2, \dots$ une suite de v.a. iid de loi uniforme sur $[0,1]$. On définit pour tout $n \in \mathbb{N}^*$ la fonction de répartition empirique de l'échantillon $(X_1, \dots, X_n)$ par

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

Soient $(X^{(1)}, \dots, X^{(n)})$ les valeurs de l'échantillon dans l'ordre croissant. On pose $X^{(0)} = 0$ et $X^{(n+1)} = 1$.

1. (TD) Montrer que $\hat{G}_n(x) = k/n$ pour $x \in [X^{(k)}, X^{(k+1)}[$, $k \in \{0, \dots, n\}$.
2. (TD) En déduire que

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

3. (TD) Rappeler le comportement limite de $d_K(\hat{G}_n, G)$ quand $n \to \infty$ (énoncer un théorème).
4. (TP) Ecrire une fonction `DistanceToUnif` qui calcule $\|\hat{G}_n - G\|_\infty$ pour un échantillon $(X_1, \dots, X_n)$ donné.
5. (TP) Tracer cinq réalisations de la suite $(\|\hat{G}_n - G\|_\infty)_{n=1,\dots,N}$, avec $N = 5000$. Superposer une ligne horizontale correspondant à l'axe des abscisses. Commenter.
6. (TP) On définit $Z_n = n^{1/2} \|\hat{G}_n - G\|_\infty$. Illustrer la convergence en loi de Z_n quand $n \to \infty$. Pour cela, on pourra créer $M = 1000$ réalisations de Z_{25}, Z_{50} et Z_{100}, et superposer les 3 fonctions de répartitions empiriques.

Exercice 4. (*Bonus*) Le but de cet exercice est d'illustrer la convergence en loi de sommes de variables aléatoires iid convenablement recentrées, en calculant explicitement la loi et en la comparant à la loi limite.

1. (TD) Soient $X_1, X_2, \ldots$ des v.a. iid de loi exponentielle de paramètre 1. Rappeler la loi de $S_n = X_1 + \cdots + X_n$. Trouver des suites $(a_n)_{n\geq 1}$ et $(b_n)_{n\geq 1}$ telles que $Z_n = (S_n - b_n)/a_n$ converge en loi (vers une loi non-triviale) et énoncer le théorème qui justifie cette convergence.
2. (TP) Superposer sur le même graphe les densités de Z_n pour $n = 10, 40, 160$, ainsi que celle de la loi limite (on pourra utiliser des fonctions dans `scipy.stats`). Faire de même pour la fonction de répartition. Commenter.
3. (TP) Calculer numériquement la distance de Kolmogorov d_n entre la loi de Z_n et celle de la loi limite. Tracer d_n en fonction de n pour $n = 1, \ldots, 200$. Commenter. La vitesse de convergence a-t-elle l'air exponentielle ou polynomiale ?
4. (TP) Faire un tracé log-log de d_n en fonction de n (avec `plt.loglog` ou en combinant `plt.plot` avec `plt.xscale("log")` et `plt.yscale("log")`) et vérifier l'hypothèse émise en réponse à la dernière question. En utilisant les valeurs de d_{100} et d_{200}, estimer un équivalent asymptotique de la vitesse de convergence. Connaissez-vous le théorème justifiant cette vitesse de convergence ?