INF 12 - Statistiques.

Dans ce TP, on étudie certaines fonctions des séries statistiques avec Python. On aura besoin des bibliothèques `numpy` et `matplotlib`. Les séries statistiques seront représentées sous formes de vecteurs ou de matrices.

1. LES COMMANDES À CONNAÎTRE

1. la commande `np.sum` somme tous les éléments du tableau
2. les commandes `np.mean` et `np.max` donnent respectivement le minimum et le maximum d'un tableau
3. la commande `np.min` donne la moyenne d'un tableau
4. la commande `np.median` donne la médiane d'un tableau
5. la commande `np.var` donne la variance d'un tableau
6. la commande `np.std` donne l'écart-type d'un tableau (c'est la racine carrée de la variance
7. la commande `np.cumsum` crée de tableau des sommes partielles en parcourant le tableau.

Remarque 1.1 — Chacune de ces commande peut prendre un argument supplémentaire `0` ou `1` qui permet d'appliquer chacune de ces commandes sur les lignes ou sur les colonnes. On tapera alors, par exemple, `np.mean(M,0)` ou `np.mean(M,1)` Essayez sur une matrice carrée.

Exercice 1 Essayez ces fonctions sur des matrices et vecteurs de votre choix.

Exercice 2 Pour différentes valeurs de N, simuler N binomiales de paramètres de votre choix. Calculer les moyennes et variances empiriques obtenues. Que remarque-t-on ?

2. FRÉQUENCES ET HISTOGRAMMES

On rappelle que dans une série statistique, la fréquence d'une valeur est donnée par l'effectif des entrées qui ont cette valeur divisé par l'effectif total. On compte aussi les fréquences cumulées : on additionne toutes les fréquences de toutes les valeurs inférieures ou égales à une valeur donnée.

Exercice 3 Pour différentes valeurs de p : dans un tableau, simuler un échantillon de lois $B(10, p)$. Représenter le vecteurs des fréquences et des fréquences cumulées.

La commande `plt.hist` crée l'histogramme représentant la série de donnée utilisée en entrée. Pensez bien à taper `plt.show()` si vous voulez l'afficher.

Exercice 4 A partir de la commande précédente, comment tracer des histogrammes pour les simulations de l'exercice précédent.

3. PREMIÈRE APPROCHE DE LA LOI NORMALE

Exercice 5 A l'aide de la méthode des rectangle, quelle valeur conjecturez-vous pour l'intégrale $\int_{-\infty}^{+\infty} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx$.

Exercice 6 Pour différentes valeurs de $n \in \mathbf{N}$ et $p \in [0, 1]$:

1. Construire une échantillon de
$$\frac{X - np}{\sqrt{np(1-p)}}$$
où X suit une loi binomiale $B(n, p)$.
2. Représenter l'histogramme de l'échantillon.
3. Comparer l'histogramme avec la courbe de la fonction dont on a calculé l'intégrale à la question précédente.
4. Que remarque-t-on ?

La loi de probabilité à densité donnée par cette fonction s'appelle la loi normale centrée réduite. Elle apparait naturellement quand on itère une même loi un nombre de fois tendant vers l'infini! c'est ce qu'on appelle le **Théorème de la limite centrale** ou **Théorème Central Limite**.

4. UN EXERCICE DE PROBABILITÉS TYPE CONCOURS

On considère une urne avec a boules blanches et b boules noires. A chaque étape, on tire une boule : si elle est blanche on arrête l'expérience, sinon on rajoute r boules noires, on remet la boule tirée dans l'urne, et on recommence. On note X la variable aléatoire qui compte le nombre de tirages effectués.

1. Écrire une fonction qui prend en entrée a, b, r et renvoie le résultat de l'expérience X.
2. Écrire une fonction qui prend en entrée a, b, r, N des entiers, réalise N fois l'expérience et renvoie la moyenne des résultats obtenus.
3. On prend $a = b = 1, \mathrm{N} = 1000$. Tracer l'évolution de la moyenne en fonction de r (pour r dans l'intervalle de votre choix).