TP 2

On rappelle que la commande `help` X permet d'obtenir une aide détaillée sur la commande X.

Exercice 1 Loi des grands nombres et théorème central limite

Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a. i.i.d. de loi géométrique $\mathcal{G}(\frac{1}{\lambda})$, où λ est les deux derniers chiffres de votre numéro d'étudiant (si votre numéro est 12027045, alors $\lambda = 45$, si votre numéro se termine par deux zéros comme pour 12027500, vous prenez les deux nombres suivants soit donc dans ce cas $\lambda = 75$).

On rappelle que ici, pour être conforme avec Python, X suit une loi géométrique de paramètre $p = 1/\lambda$ si X est une variable aléatoire sur $\mathbb{N}^*$ vérifiant, pour $k \geqslant 1$,

$$P(X = k) = p(1-p)^{k-1}.$$

Posons

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{et} \quad Z_n = \sqrt{n}\frac{\bar{X}_n - m}{\sigma}$$

où $m = E(X_1)$ et $\sigma^2 = Var(X_1)$.

1. **Loi forte des grands nombres.**
   (a) Simuler 10 réalisations indépendantes $(\bar{x}_1^{(i)}, \cdots, \bar{x}_{1000}^{(i)})$ de $(\bar{X}_1, \cdots, \bar{X}_{1000})$, pour $i = 1, \cdots, 10$ et représenter ces suites sur le même graphique.
   (b) Donner les valeurs de $\bar{x}_{1000}^{(i)}$, $i = 1, \cdots, 10$.
   (c) Donner la valeur de m (ou pourra regarder sur wikipedia ou dans un livre) et comparer avec les 10 valeurs obtenues $(\bar{x}_{1000}^{(i)})_{i\in\{1,\cdots,10\}}$. Quel type de convergence illustre-t-on ici ?
2. **Théorème Central limite.**
   (a) Calculer, ou trouver sur wikipedia, la valeur de σ.
   (b) Simuler 1000 réalisations indépendantes $z_1^{(i)}, z_3^{(i)}, z_{100}^{(i)}$ de Z_1, Z_3, Z_{100}, pour $i = 1, \cdots, 1000$, et représenter l'histogramme illustrant les lois respectives de Z_1, Z_3, Z_{100} sur 3 graphiques côte à côte. On pourra utiliser

```
fig,ax = plt.subplots(1,3)
ax[0].hist(...)
ax[1].hist(...)
...
```

   (c) Quel type de convergence illustre-t-on ici ? Commenter
3. **Théorème Central limite, suite.**
   (a) Simuler 1000 réalisations indépendantes $z_2^{(i)}, z_{10}^{(i)}, z_{100}^{(i)}$ de Z_2, Z_{10}, Z_{100}, pour $i = 1, \cdots, 1000$, et représenter les fonctions de répartition empirique des lois respectives de Z_2, Z_{10}, Z_{100} sur le même graphique.
   On rappelle que si on a $(X_1, \cdots, X_n)$, n valeurs d'un échantillon, la fonction de répartition empirique est la fonction

$$t \mapsto \frac{1}{n}\sum_{i=1}^{n} 1_{]-\infty,t]}(X_i)$$

   On pourra pour cela adapter la ligne suivante :

```
plt.plot(np.sort(z),np.arange(len(z))/len(z))
```

   (b) Quel type de convergence illustre-t-on ici ? Commenter
   (c) Simuler 10 réalisations indépendantes $(z_1^{(i)}, \cdots, z_{100}^{(i)})$ de $(Z_1, \cdots, Z_{100})$, pour $i = 1, \cdots, 10$ et représenter ces suites sur le même graphique. Qu'observe-t-on (ou n'observe t'on pas) ? Commenter.

Exercice 2 Intervalle de confiance, Méthodes de Monte-Carlo

1. Intervalle de confiance.
   Soit (X_i) une suite de variables aléatoires indépendantes et de même loi, de carré intégrable. On note m son espérance et v sa variance, et pour tout $n \geqslant 1$, $S_n = \sum_{i=1}^{n} X_i$ et $\overline{X_n} = S_n/n$.
   On se donne également une variable aléatoire Z de loi normale $\mathcal{N}(0, 1)$.

(a) Rappeler les limites de $\overline{X_n}$ et de $Z_n = \sqrt{\frac{n}{v}}\left(\overline{X_n} - m\right)$, en précisant les modes de convergence.

(b) Montrer que, pour tout $\alpha \in]0,1[$, il existe un unique réel u_α tel que $\mathbf{P}(Z \in [-u_\alpha, u_\alpha]) = 1 - \alpha$.

(c) Soit $\alpha \in]0,1[$ fixé. Quelle est la limite de $\mathbf{P}(Z_n \in [-u_\alpha, u_\alpha])$ lorsque n tend vers $+\infty$?

(d) On appelle « intervalle de confiance de niveau $(1-\alpha)$ de m » l'intervalle $I_{n,\alpha} = \left[\overline{X_n} - u_\alpha\sqrt{\frac{v}{n}}\,,\ \overline{X_n} + u_\alpha\sqrt{\frac{v}{n}}\right]$. Pour n grand, que peut-on dire de la probabilité de l'événement $\{m \in I_{n,\alpha}\}$.

(e) Dans la pratique, on estime la variance v par la variance empirique

$$\overline{V_n} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X_n})^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X_n}^2\right).$$

Justifier la convergence de $\overline{V_n}$ vers v.

(f) On suppose pour cette question et la suivante que les (X_i) sont de loi de Bernoulli de même paramètre p. Quelles sont les valeurs de m et v ?

(g) Lire l'aide de la fonction `stats.norm.ppf`. Écrire une fonction python renvoyant la valeur de u_α.

(h) Choisir une valeur de p dans l'intervalle $[0.2, 0.8]$. Pour $n = 30$, et $\alpha \in \{0.01, 0.05, 0.1\}$, générer un échantillon de taille $N = 10^5$ d'intervalles de confiance $I_{n,\alpha}$ et dénombrer combien d'entre eux contiennent p.

Le principe de la méthode de Monte-Carlo est d'utiliser la loi des grands nombres et le théorème central limite pour estimer l'intégrale d'une fonction sur un intervalle. Nous allons voir deux façons de la mettre en oeuvre.

2. **Méthode 1 : on lance des points**.

Soit $f : x \mapsto \sqrt{1-x^2}$ définie sur l'intervalle $[0,1]$. On note I son intégrale et A le domaine $\{(x,y), 0 \leqslant y \leqslant f(x)\}$. On considère également une suite de couples aléatoires (U_n, V_n), formé de variables aléatoires de loi uniforme sur $[0,1]$ et toutes indépendantes. Les premières questions sont à résoudre sur feuille !

(a) Justifier que I est égal à $\pi/4$.

(b) Quelle est la probabilité de l'événement $\{(U_1, V_1) \in A\}$?

(c) On note N_n et p_n les variables aléatoires suivantes :

$$N_n = \text{card}\{i \leqslant n, (U_i, V_i) \in A\} \quad \text{et} \quad p_n = \frac{N_n}{n}.$$

Quelle est la loi de N_n ? Quelle est la limite de p_n ?

(d) On revient au notebook, créer une fonction renvoyant les valeurs de f.

(e) On choisit $n = 1000$. Déterminer un intervalle de confiance de I de niveau $1 - \alpha = 95\%$, c'est-à-dire un intervalle J_n tel que $\mathbf{P}(I \in J_n) = 0.95$.

(f) Observer comment cet intervalle évolue quand on fait varier le niveau de confiance et la taille de l'échantillon.

(g) A nouveau pour $n = 1000$, générer $N = 10^5$ intervalles de confiance de I de même niveau de confiance et déterminer quelle proportion d'entre eux contiennent $\pi/4$.

3. **Méthode 2 : on estime une espérance**

On considère ici une suite (U_i) de variables aléatoires indépendantes et de loi uniforme sur $[0,1]$ et on note $X_i = f(U_i)$.

(a) Calculer $\mathbb{E}(X_i)$ et $\text{var}(X_i)$.

(b) Générer un échantillon suivant la loi de (X_1) de taille $n = 10^3$. Calculer sa moyenne et sa variance empiriques.

(c) Déterminer un intervalle de confiance de $\mathbb{E}(X_i)$ de niveau 95%.

(d) On choisit $n = 1000$. Générer $N = 10^5$ intervalles de confiance de I de même niveau de confiance et déterminer quelle proportion d'entre eux contiennent $\pi/4$.

4. **Méthode 3 : on réduit la variance**.

On note g la fonction définie sur $[0,1]$ par $g(x) = (f(x) + f(1-x))/2$. Toujours pour U une variable aléatoire de loi uniforme sur $[0,1]$, calculer $\mathbb{E}(g(U))$.

Reprendre la méthode ci-dessus pour déterminer des intervalles de confiance de I.

5. Comparer les différentes méthodes de Monte-Carlo, en prenant en compte la largeur des intervalles de confiance obtenus pour les mêmes valeurs de n et α, ainsi que la facilité à les mettre en oeuvre.