Probabilités - Statistiques
TP 2

1. **Théorème central limite**
Le but de cett exercice est de vérifier numériquement le résultat du théorème central limite dans le cas d'une variable aléatoire Y suivant une loi de Cauchy (i.e. de densité $f_Y(t) = \frac{1}{\pi}\frac{1}{1+t^2}$), une loi exponentielle ou une loi de Bernouilli.
   (a) On commence par considérer une suite de variables indépendantes X_i, $i \in \mathbb{N}$, de loi uniforme sur $[-1, 1]$. Écrire une fonction qui prend en argument un entier N et représente graphiquement les réalisations de la suite des variables aléatoires $S_n := \frac{1}{n}(X_1 + \cdots + X_n)$ pour $n = 1, \dots, N$. Vers quelle valeur cette suite semble-t-elle converger ? Expliquer pourquoi.
   (b) On considère maintenant les variables aléatoires $Y_i = \tan(\frac{\pi}{2}X_i)$, $i \in \mathbb{N}$. Montrer théoriquement que les Y_i sont indépendantes, et déterminer leur loi. Écrire une fonction similaire à celle de la question précédente, qui représente graphiquement les réalisations de la suite des $T_n := \frac{1}{n}(Y_1 + \cdots + Y_n)$ pour $n = 1, \dots, N$. Vers quelle valeur cette suite semble-t-elle converger ? Expliquer pourquoi.
   (c) Calculer la moyenne μ et l'écart-type σ des Y_i. Afin d'illustrer le Théorème central limite, on va chercher à observer que la suite de variables aléatoires $\frac{(T_n-\mu)}{\sigma/\sqrt{n}}$ converge en loi vers $\mathcal{N}(0, 1)$ lorsque n devient grand. Écrire une fonction qui prend en argument n et simule la variable aléatoire $\frac{(T_n-\mu)}{\sigma/\sqrt{n}}$. Pour $n = 50$, simuler 1000 réalisations de $\frac{(T_n-\mu)}{\sigma/\sqrt{n}}$, tracer l'histogramme associé et le comparer à la densité de la loi normale. Même question pour $n = 200$, 500 et 1000. Les résultats observés sont-ils cohérents ?
   (d) Même question qu'en (c) en prenant cette fois des variables aléatoires Y_i suivant une loi exponentielle de paramètre 2. À partir de quelle valeur de n la convergence vous semble-t-elle satisfaisante ?
   (e) Toujours la même question en prenant cette fois des variables aléatoires Y_i suivant une loi de Bernouilli de paramètre $\frac{1}{2}$. À partir de quelle valeur de n la convergence vous semble-t-elle satisfaisante ? Comparer aux valeurs obtenues précédemment.

2. **Battage de cartes par le *riffle shuffle***
On considère un jeu de n cartes numérotées de 0 à $n-1$. Au départ, elles sont ordonnées par ordre croissant et on cherche à les battre (càd appliquer une permutation aléatoire de $\{0, \dots, n-1\}$) de façon à ce que les cartes soient réparties de la façon la plus aléatoire possible. La solution optimale consisterait à appliquer une permutation aléatoire choisie de façon uniforme, mais cela est bien sûr hors de portée, le choix d'une telle permutation devant se faire parmi les $n!$ permutations possibles, ce qui est beaucoup trop si n vaut, par exemple, 52. On opère en conséquence à un battage "artisanal" (mais réglementé lorsque l'on se trouve dans un cercle de jeu public). Nous allons ici étudier un exemple de battage, le riffle shuffle : nous proposerons un modèle probabiliste pour celui-ci et étudierons numériquement le caractère uniforme de la répartition des cartes qu'il produit. Nous chercherons notamment le nombre minimum de riffle shuffles à réaliser pour que le paquet de cartes soit suffisamment mélangé.
Le riffle shuffle se décompose en deux étapes que l'on modélise comme suit.
   — *Étape 1.* On coupe le jeu en deux paquets de taille k et $n-k$: les k premières cartes du paquet et les $n-k$ suivantes, k étant choisi de manière aléatoire suivant une loi Binomiale de paramètres n et $1/2$.
   — *Étape 2.* On entrelace alors les deux paquets de la manière suivante :
      — si on a n_1 cartes dans le paquet 1 et n_2 cartes dans le paquet 2, on choisit d'abord une carte : la première carte du paquet 1 avec la probabilité $\frac{n_1}{n_1+n_2}$ ou la première carte du paquet 2 avec la probabilité $\frac{n_2}{n_1+n_2}$;

— on continue l'expérience, en actualisant n_1 et n_2 à chaque étape, jusqu'à épuisement des deux paquets.

(a) Écrire une fonction `RS` qui s'applique à un objet de type `list`, appelé `J` (il sera pratique de supposer que `J` ne contient que des valeurs 2 à 2 distinctes), qui renvoie un objet de type `list`, appelé `R`, correspondant à un riffle shuffle appliqué à `J` (`R` contiendra donc les mêmes éléments que `J`, mais dans un autre ordre).

(b) Écrire une fonction `Trajectoire` qui s'applique à un objet de type `list`, appelé `J0` et à un `int` appelé `T`, qui renvoie un objet de type `list`, de longueur `T+1`, dont le premier élément est `J0` et dont les éléments se déduisent successivement les uns des autres par application de `RS`.

(c) On a donc une fonction qui, à un ordre initial, que l'on choisira `J0=range(n)`, pour $n = 52$, associe la suite J_0, $J_1 = \mathtt{RS}(J_0), \ldots, J_{t-1} = \mathtt{RS}(J_t), \ldots, J_T$, de ses riffle shuffle successifs. On cherche à mesurer comment la loi de J_t se rapproche de la loi uniforme sur l'ensemble des permutations de $\{0, \ldots, n-1\}$ lorsque t augmente. Si la loi de J_t était la loi uniforme sur l'ensemble des permutations de $\{0, \ldots, n-1\}$, alors la loi de chaque coordonnée $J_t(k)$ serait la loi uniforme sur $\{0, \ldots, n-1\}$, càd que l'on aurait, pour tous k, i,

$$P(J_t(k) = i) = \frac{1}{n}.$$

On va alors mesurer le caractère uniforme de la distribution de J_t avec l'indicateur

$$I(J_t) := \max\{|nP(J_t(k) = i) - 1|, (k, i) \in \{0, \ldots, n-1\}^2\}.$$

Calculer, pour $n = 52$, $T = 10$, une estimation $\hat{I}(J_t)$ de $I(J_t)$ pour $t \in \{0, \ldots, T\}$, afficher le graphe de $\hat{I}(J_t)$ en fonction de t et donner le t minimal permettant d'avoir $I(J_t) < 1$ (si un tel t est trouvé).
Indication : Pour estimer $P(J_t(k) = i)$, on simulera un grand nombre de trajectoires et on pourra utiliser, par exemple, la fonction `np.bincount`.

(d) Pour J_{unif} une permutation aléatoire uniforme de `J0=range(n)`, on a $I(J_{unif}) = 0$. Néanmoins, ce à quoi il est légitime de comparer les estimations $\hat{I}(J_t)$ n'est pas $I(J_{unif})$ mais l'estimation $\hat{I}(J_{unif})$ obtenue avec un échantillon de même taille de permutations aléatoires uniformes indépendantes. Dans le même programme, calculer cette estimation $\hat{I}(J_{unif})$ et faire apparaître, sur le graphique précédent, une ligne horizontale d'ordonnée $\hat{I}(J_{unif})$.
Indication : Pour simuler des permutations aléatoires uniformes indépendantes, on pourra utiliser `np.random.shuffle`.