

1. Vous ne pouvez pas vous connecter à votre session personnelle : à la place, vous vous connecterez à la session **mathens** avec le mot de passe **mathens**.
2. À la fin du partiel, vous enregistrerez votre feuille de calcul Jupyter dans le dossier **remise_copie** présent sur le bureau. Il faut impérativement inclure votre nom dans le nom des fichiers (donc, sauvegarder un fichier nommé par exemple **NOM.ipynb**). Par sécurité, il est également conseillé de sauvegarder une exportation **NOM.pdf** de la feuille Jupyter.
3. La réponse aux questions théoriques doit être préférentiellement sur les copies papier.
4. On rappelle les commandes d'import des bibliothèques adéquates :

```
import numpy as np
import scipy.stats as scs
import matplotlib.pyplot as plt
import numpy.random as random
```

Problème. Le processus des restaurants chinois est une suite aléatoire $(c_n)_{n \geq 1}$, où chaque $c_n = (c_{n,1}, c_{n,2}, \dots, c_{n,\ell})$ est une suite finie de nombres entiers positifs tels que $c_{n,1} + c_{n,2} + \dots + c_{n,\ell} = n$. Chaque élément c_n représente une façon de répartir n personnes en des petits groupes de taille $c_{n,1}, c_{n,2}, \dots, c_{n,\ell}$ (les personnes qui s'installent à des tables différentes dans un restaurant). Si c_n est connu, on construit c_{n+1} aléatoirement comme suit :

- Avec probabilité $\frac{1}{n+1}$, la nouvelle personne qui rentre dans le restaurant s'installe seule à une nouvelle table. Il y a donc un nouveau groupe de taille 1 :

$$c_{n+1} = (c_{n,1}, c_{n,2}, \dots, c_{n,\ell}, 1) \quad \text{avec probabilité } \frac{1}{n+1}.$$

- Sinon, pour tout groupe $c_{n,k}$ déjà existant, la nouvelle personne se rajoute au groupe $c_{n,k}$ avec probabilité $\frac{c_{n,k}}{n+1}$. On a donc :

$$c_{n+1} = (c_{n,1}, \dots, c_{n,k} + 1, \dots, c_{n,\ell}) \quad \text{avec probabilité } \frac{c_{n,k}}{n+1}.$$

On suppose que les choix aléatoires à chaque temps n sont indépendants. On part bien sûr au temps $n = 1$ avec $c_1 = (1)$.

1. Écrire un programme **random_group(n)** qui construit jusqu'au temps n la suite c_n , et qui renvoie c_n (sous forme d'une liste Python). Produire une suite aléatoire c_{1000} .
2. Le nombre de groupes $\ell = \ell_n$ dans c_n est une suite croissante aléatoire. Montrer que :

$$\ell_n = \sum_{k=1}^n \text{Ber}\left(\frac{1}{k}\right),$$

le terme de droite étant une somme de variables de Bernoulli indépendantes de paramètres $1, \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{n}$.

3. Montrer que $\lim_{n \rightarrow \infty} \frac{\mathbb{E}[\ell_n]}{\log n} = 1$, et que $\lim_{n \rightarrow \infty} \frac{\text{var}(\ell_n)}{\log n} = 1$.
4. Écrire un programme `nb_group(n)` qui dessine la fonction $k \mapsto \frac{\ell_k}{\log(k+1)}$ jusqu'au temps n .
5. Exécuter la fonction de la question précédente avec n grand, par exemple $n = 1000000$ (c'est possible si la fonction de la question précédente est programmée intelligemment). Que peut-on conjecturer concernant le comportement de la variable aléatoire ℓ_n lorsque n tend vers l'infini ?
6. On s'intéresse au nombre P_n de personnes qui au temps n sont seules dans un groupe. Autrement dit :

$$P_n = \text{card}(\{k \in \llbracket 1, \ell_n \rrbracket \mid c_{n,k} = 1\}).$$

Avec $n = 10000$ et un échantillon de taille $N = 1000$, dessiner l'histogramme empirique de P_n , et comparer cet histogramme à celui d'une variable aléatoire P qui suit une loi Poisson(1) :

$$\mathbb{P}[P = k] = \frac{1}{e k!} \quad \text{pour tout } k \geq 0.$$

On pourra utiliser `L.count(x)` pour compter le nombre d'occurrences de x dans une liste.

7. Comment interpréter l'histogramme de la question précédente, concernant le comportement de la variable aléatoire P_n lorsque n tend vers l'infini ?
8. On s'intéresse maintenant au nombre $c_{n,1}$ de personnes dans le premier groupe. Avec $n \in \{100, 1000, 10000\}$, tracer la fonction de répartition empirique de $\frac{c_{n,1}}{n}$, qui est la proportion de personnes dans le premier groupe. On pourra comme dans la question 6. utiliser des échantillons de taille $N = 1000$.
9. Commenter ces fonctions de répartition empirique. Que semblent-elles indiquer concernant le comportement de la variable $U_n = \frac{c_{n,1}}{n}$ lorsque n tend vers l'infini ?
10. Soit k un entier entre 1 et n . Montrer que, si $c_{n,1} = k$, alors il existe des entiers $a_2 < a_3 < \dots < a_k \leq n$ tels que :

$$c_{l,1} = j \quad \text{pour tout } l \in \llbracket a_j, a_{j+1} - 1 \rrbracket$$

si $j \in \llbracket 1, k \rrbracket$ (avec par convention $a_1 = 1$ et $a_{k+1} - 1 = n$). Étant donnée une telle suite de paramètres, quelle est la probabilité pour obtenir $c_{n,1} = k$ et cette suite $(a_2, a_3, \dots, a_k)$ lors de la construction de c_n ?

11. Dédurre de la question précédente la loi de $c_{n,1}$ pour tout temps $n \geq 1$, et une preuve du comportement conjecturé à la question 9 pour les variables U_n .
12. Soit $\sigma \in \mathfrak{S}(n)$ une permutation de taille n , c'est-à-dire une bijection $\sigma : \llbracket 1, n \rrbracket \rightarrow \llbracket 1, n \rrbracket$. On note $d_{n,1}$ la taille du cycle C_1 de σ contenant 1 ; $d_{n,2}$ la taille du cycle C_2 de σ contenant le plus petit entier qui n'est pas dans C_1 ; etc. jusqu'à $d_{n,\ell}$, la taille du cycle C_ℓ de σ contenant le plus petit entier qui n'est pas dans $C_1 \sqcup C_2 \sqcup \dots \sqcup C_{\ell-1}$. On choisit σ aléatoirement et uniformément parmi les $n!$ permutations de taille n . Montrer que la suite aléatoire $(d_{n,1}, \dots, d_{n,\ell})$, qui est un élément aléatoire dans l'ensemble $\mathfrak{C}(n)$ des suites d'entiers de somme égale à n , a la même distribution que $(c_{n,1}, \dots, c_{n,\ell})$:

$$\forall c \in \mathfrak{C}(n), \quad \mathbb{P}[(d_{n,1}, \dots, d_{n,\ell}) = c] = \mathbb{P}[(c_{n,1}, \dots, c_{n,\ell}) = c].$$