

**Exercice 1** **Correction de code** **6 points**

Lors de la mise au point de cet examen, Il semblerait qu'une dizaine d'erreurs se soient cachées dans le code ci-dessous. Saurez vous les retrouver ?

```
# importez les bibliotheques necessaires
import matplotlib as plt
import numpy as np
import scipy.io as sio

# chargez les donnees
mat = sio.loadmat("median_m8.mat")
X = mat["X"]
n, p = len(X)

# Centrez et reduisez les donnees du tableau X
m = np.mean(X)
s = np.std(X)
Xn = X - m / s**2
# Verifiez que vos calculs des moyennes et des ecart types sont justes
np.mean(Xn), np.std(Xn)

# Calculez les axes principaux
eigvc, eigval = np.linalg.eig(Xn @ Xn/n)

# Quelle est leur importance en pourcentage d information ?
plt.bar(np.arange(n), eigval, .5)

# Representez les variables dans le plan des deux premiers facteurs principaux.
Vn = eigvc@np.diag(np.sqrt(eigval))
plt.plot(Vn[:, 0], Vn[:, 1], '*')
```

**Exercice 2** **Le résultat aide** **4 points**

On se donne une fonction coût  $J$  définie par:

$$J(\lambda) = \sum_{i=1}^n \log(\lambda \exp^{-\lambda x_i}).$$

1. Donnez l'expression de  $\lambda$  qui minimise le coût  $J$ .
2. Application numérique : calculez ce  $\lambda$  qui minimise le coût  $J$  pour  $n = 10$  et

|       |       |       |       |       |       |       |       |       |          |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|
| $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ | $x_9$ | $x_{10}$ |
| 0.338 | 0.939 | 0.198 | 0.539 | 0.839 | 0.269 | 0.324 | 0.524 | 0.612 | 0.418    |

**Exercice 3** **Assez (P) de musique** **10 points**

Le but de cet exercice est d'analyser les activités et tendances culturelles dans différents pays à partir des données sur les musique enregistrées en 1998. Les variables disponibles sont :

- Vente : le montant en dolars par habitant,
- MusPopNat : le pourcentage de musique pop nationale enregistrée (en %)
- MusPopInt : le pourcentage de musique pop internationale enregistrée (en %)
- MusClas : le pourcentage de musique classique enregistrée (en %)
- TxEnrPirates : le pourcentage de musique pirate enregistrée (en %)
- TxImp : le pourcentage de musique importée enregistrée (en %)
- LectCD : le nombre de lecteurs de CD pour 100 Hab

Commencez par lire les données disponibles en exécutant les instructions suivantes :

```
X = np.loadtxt("data_median.csv", delimiter=",")
variables = np.loadtxt("variables_median.csv", delimiter=",", dtype=str)
pays = np.loadtxt("pays_median.csv", delimiter=",", dtype=str)
```

1. Combien y a t'il de pays et de variables dans ces données ?
2. Étude de la variable **vente**
  - a) tracez la fonction de répartition empirique de la variable vente et commentez là,
  - b) dessinez la boîte à moustache de la variable vente et commentez là.
3. Les trois variables **MusPopNat**, **MusPopInt** et **MusClas** sont liées puisqu'il s'agit de pourcentages et que la somme de ces trois variable est toujours égale à 100.
  - a) Quel est l'effet de ce lien sur les résultats de l'ACP ?
  - b) Quelle serait votre recommandation pour traiter des trois variables ?
4. Calculez les axes factoriels associés au tableau  $X$
5. Quel est le nombre minimal d'axes factoriels qu'il faut considérer pour représenter au moins 50% de l'information contenue dans les variables disponibles ?
6. Les deux premières composantes principales.
  - a) Visualisez sur les deux premières composantes principales.
  - b) Proposez une interprétation de ces axes ?
7. Comment interpréter les axes en terme de variables ?
  - a) calculez les corrélations entre variables et les composantes principales.
  - b) Visualisez les 2 premiers axes factoriels et commentez les.
8. A la fin du corrigé de l'examen, vous avez trouvé ce bout de programme python et le résultat de son exécution. Que pouvez vous en déduire ? Par exemple sur la Hongrie et la Suisse.

```
Vn = V*np.diag(np.sqrt(vals)) / np.sqrt(n)
U = Xn@V

plt.figure(1, figsize=(20, 10))
plt.subplot(1,2,1)
a = np.linspace(0, 2*np.pi, 100)
plt.plot(Vn[:, 2], Vn[:, 3], '*')
for i in range(p):
    plt.text(Vn[i, 2], Vn[i, 3], variables[i])
plt.plot(np.cos(a), np.sin(a), '-')
plt.subplot(1,2,2)
plt.plot(U[:, 2], U[:, 3], 'o')
for i in range(n):
    plt.text(U[i, 2], U[i, 3], pays[i])
```

