## Questions de cours

1. Énoncer la caractérisation de la projection orthogonale sur un sous-espace vectoriel $F$ de $\mathbb{R}^n$.
2. Énoncer la loi faible des grands nombres.

## Exercice 1

Dans tout le problème, $n$ désigne un entier naturel non nul.
On se place dans $\mathbb{R}^n$ muni du produit scalaire canonique, on rappelle que si $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ et $y = (y_1, \dots, y_n) \in \mathbb{R}^n$ sont deux vecteurs de $\mathbb{R}^n$ alors leur produit scalaire canonique noté $\langle x, y\rangle$ est défini par :

$$\langle x, y\rangle = \sum_{i=1}^{n} x_i y_i.$$

On confondra les réels et les matrices carrées de $\mathcal{M}_{1,1}(\mathbb{R})$. Si $A$ est une matrice, sa transposée est notée $A^\top$. Si $E$ est un sous-espace vectoriel de $\mathbb{R}^n$, son supplémentaire orthogonal est noté $E^\perp$.
On note $S_n^+(\mathbb{R})$ l'ensemble des matrices réelles symétriques dont le spectre est inclus dans $\mathbb{R}_+$, autrement dit :

$$S_n^+(\mathbb{R}) = \{A \in \mathcal{M}_n(\mathbb{R}),\ A^\top = A \text{ et } \operatorname{Sp}(A) \subset \mathbb{R}_+\}.$$

L'objectif de ce problème est de démontrer l'appartenance à $S_n^+(\mathbb{R})$ de certaines matrices.
La partie A est consacrée à l'étude d'exemples en petites dimensions. Dans la partie B, on fait le lien avec la notion de produit scalaire dans $\mathbb{R}^n$, et dans la partie C, on fait le lien avec la notion de covariance. Ces deux dernières parties ne sont pas indépendantes.

### Partie A : Quelques exemples

1. Préciser, en justifiant, parmi les matrices ci-dessous lesquelles appartiennent à $S_2^+(\mathbb{R})$ :

$$A = \begin{pmatrix} 0 & 2 \\ -3 & 5 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad D = \begin{pmatrix} 2 & \sqrt{3} \\ \sqrt{3} & 4 \end{pmatrix}.$$

2. On se place dans $\mathcal{M}_3(\mathbb{R})$ et on considère l'ensemble $\mathcal{M}$ défini par :

$$\mathcal{M} = \{M(a,b),\ (a,b) \in \mathbb{R}^2\} \quad \text{où} \quad M(a,b) = \begin{pmatrix} a & b & 0 \\ b & a & b \\ 0 & b & a \end{pmatrix}.$$

   a. Montrer que $\mathcal{M}$ est un sous-espace vectoriel engendré par une famille judicieusement choisie. Justifier que cette famille est libre et en déduire la dimension de $\mathcal{M}$.

   b. Soit $(a,b) \in \mathbb{R}^2$. Déterminer deux vecteurs propres de $M(a,b)$ sous la forme $\begin{pmatrix} 1 \\ y \\ 1 \end{pmatrix}$ (avec $y \in \mathbb{R}$) puis un troisième vecteur propre orthogonal aux deux précédents.

   c. En déduire un réel $r \geq 0$ et une matrice $P$ inversible telle que $P^\top = P^{-1}$ et :

$$\forall (a,b) \in \mathbb{R}^2, \quad M(a,b) = P \begin{pmatrix} a & 0 & 0 \\ 0 & a + br & 0 \\ 0 & 0 & a - br \end{pmatrix} P^\top.$$

   d. Déterminer une condition nécessaire et suffisante portant sur $a$ et $b$ pour que $M(a,b)$ appartienne à $S_3^+(\mathbb{R})$.

## Partie B : Matrice de Gram

Soit une famille $(e_1, \dots, e_n)$ de $n$ vecteurs de $\mathbb{R}^n$, on note $E = \mathrm{Vect}(e_1, \dots, e_n)$ et on appelle matrice de Gram de cette famille la matrice $G$ définie par :

$$G = \begin{pmatrix} \langle e_1, e_1\rangle & \langle e_1, e_2\rangle & \dots & \langle e_1, e_n\rangle \\ \langle e_2, e_1\rangle & \langle e_2, e_2\rangle & \dots & \langle e_2, e_n\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle e_n, e_1\rangle & \langle e_n, e_2\rangle & \dots & \langle e_n, e_n\rangle \end{pmatrix}.$$

3. a. Écrire en Python une fonction `prod_scal` qui prend en argument deux vecteurs `u` et `v` représentés par des listes (ou des arrays unidimensionnels) et qui renvoie leur produit scalaire (on ne vérifiera pas que celui-ci existe).
   b. Écrire en Python une fonction `matG` qui prend en argument une liste `e` de vecteurs $[e_1, \dots, e_n]$ et qui renvoie un array représentant la matrice $G$ définie ci-dessus.
4. a. Montrer que pour tout $x \in \mathbb{R}^n$ :
$$x \in E^\perp \Leftrightarrow (\forall i \in [\![1, n]\!],\ \langle x, e_i\rangle = 0).$$
   b. En déduire que pour tout $(x_1, \dots, x_n) \in \mathbb{R}^n$ :
$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathrm{Ker}(G) \Leftrightarrow \sum_{i=1}^n x_i e_i \in E^\perp.$$
   c. Conclure que la famille $(e_1, \dots, e_n)$ est libre si et seulement si $G$ est une matrice inversible.
5. Dans cette question, on cherche à prouver que $G \in S_n^+(\mathbb{R})$ et on pose $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathcal{M}_{n,1}(\mathbb{R})$.
   a. Calculer le produit matriciel $GX$ et montrer que $X^\top G X = \langle x', x'\rangle$ où $x'$ est un vecteur de $\mathbb{R}^n$ à préciser.
   b. Soit $\lambda \in \mathrm{Sp}(G)$ et $X$ un vecteur propre de $G$ associé à cette valeur propre. En calculant de deux façons $X^\top G X$, montrer que $\lambda \in \mathbb{R}_+$ et conclure.

## Partie C : Matrice de covariance

Soient $(X_1, \dots, X_n)$ $n$ variables aléatoires réelles définies sur un même espace probabilisé $(\Omega, \mathcal{F}, P)$ telles que pour tout $(i, j) \in [\![1, n]\!]^2$, la covariance de $X_i$ et $X_j$, notée $\mathrm{Cov}(X_i, X_j)$, existe.
On appelle matrice de covariance de $(X_1, \dots, X_n)$ la matrice $\Sigma$ définie par :

$$\Sigma = \begin{pmatrix} \mathrm{Cov}(X_1, X_1) & \mathrm{Cov}(X_1, X_2) & \dots & \mathrm{Cov}(X_1, X_n) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Cov}(X_2, X_2) & \dots & \mathrm{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_n, X_1) & \mathrm{Cov}(X_n, X_2) & \dots & \mathrm{Cov}(X_n, X_n) \end{pmatrix}.$$

6. a. Montrer que :
$$\forall (i, j, k) \in [\![1, n]\!]^3, \forall x \in \mathbb{R} \quad \mathrm{Cov}(X_i, X_j + xX_k) = \mathrm{Cov}(X_i, X_j) + x\,\mathrm{Cov}(X_i, X_k).$$
   b. En déduire que :
$$\forall i \in [\![1, n]\!], \forall (x_1, \dots, x_n) \in \mathbb{R}^n, \quad \mathrm{Cov}(X_i, \sum_{j=1}^n x_j X_j) = \sum_{j=1}^n x_j\,\mathrm{Cov}(X_i, X_j).$$
7. a. Justifier que $\Sigma$ est une matrice symétrique réelle.
   b. En s'inspirant de la question 5 de la partie précédente, montrer que $\Sigma \in S_n^+(\mathbb{R})$.

### Exercice 2

Ce problème est relatif à l'étude de variables aléatoires réelles qui suivent une loi en lien avec les lois géométriques, binomiales et normales. Dans la partie A, on étudie deux premières variables aléatoires réelles. Cette étude, généralisée en partie B, permet de montrer que les variables aléatoires étudiées suivent toutes la même loi. Enfin, en partie C, on établit des liens entre cette loi, la loi binomiale et la loi normale. Ces trois parties ne sont pas indépendantes.

Dans tout le problème $k$ désigne une entier naturel, $n$ un entier naturel non nul et $p$ un réel appartenant à $]0,1[$. On note enfin $q = 1 - p$.

## Partie A : Autour de la loi géométrique

Une équipe de géologues travaille en zone sismique et installe un sismomètre au sommet d'un volcan. Chaque jour, si ce sismomètre détecte une onde sismique, ce dernier envoie une alerte par satellite à un camp de base situé au pied du volcan. On note $X_1$ la variable aléatoire réelle correspondant au nombre de jours sans alerte avant une première alerte, par exemple si cette première alerte survient au troisième jour après l'installation du sismomètre alors $X_1$ prend la valeur 2. On admet que ce sismomètre envoie au plus une alerte par jour, et que la probabilité que survienne une alerte est constante et égale à $p$, indépendante des résultats obtenus les jours précédents.

1. a. Déterminer l'ensemble des valeurs prises par $X_1$ noté $X_1(\Omega)$, et pour tout $k \in X_1(\Omega)$, calculer $\mathbb{P}(X_1 = k)$.
   b. En déduire que $1 + X_1$ suit la loi géométrique de paramètre $p$ ce qu'on notera dorénavant $X_1 \hookrightarrow \mathcal{G}^{-1}(p)$.
2. Ce sismomètre étant extrêmement sensible, les géologues décident de mener des analyses complémentaires à partir d'une seconde alerte. Aussi on note $X_2$ la variable aléatoire réelle correspondant au nombre de jours sans alerte avant cette seconde alerte, depuis l'installation du sismomètre. Par exemple, si une première alerte survient au troisième jour après l'installation de ce sismomètre et une seconde au septième jour alors $X_2$ prend la valeur 5.
   a. Pour tout $(j,k) \in \mathbb{N}^2$, calculer $\mathbb{P}(X_2 = k | X_1 = j)$.
   b. En déduire que :
$$\forall k \in \mathbb{N}, \mathbb{P}(X_2 = k) = (k+1)q^k p^2$$

## Partie B : Généralisation de la situation précédente

On généralise l'étude précédente et pour tout $n \in \mathbb{N}^*$, on note $X_n$ la variable aléatoire réelle correspondant au nombre de jours sans alerte avant une $n$-ième alerte.

3. Écrire une fonction en Python prenant en argument un entier naturel $n$ non nul et un flottant $p \in ]0,1[$ qui simule et renvoie une réalisation de la variable aléatoire $X_n$.
4. a. Démontrer par récurrence sur $k \in \mathbb{N}$ que :
$$\forall k \in \mathbb{N}, \binom{n+k}{k} = \sum_{j=0}^{k} \binom{n+j-1}{j}.$$
   b. Démontrer par récurrence sur $n \in \mathbb{N}^*$ que :
$$\forall k \in \mathbb{N}, \mathbb{P}(X_n = k) = \binom{n+k-1}{k} q^k p^n.$$
   On le notera dorénavant $X_n \hookrightarrow \mathcal{B}^{-1}(n,p)$.
5. Soit $(X_1^{(i)})_{i \in [\![1,n]\!]}$ une famille de $n$ variables aléatoires réelles mutuellement indépendantes telle que :
$$\forall i \in [\![1,n]\!], X_1^{(i)} \hookrightarrow \mathcal{G}^{-1}(p).$$
   On cherche à démontrer par récurrence sur $n$ que :
$$\sum_{i=1}^{n} X_1^{(i)} \hookrightarrow \mathcal{B}^{-1}(n,p).$$

a. Vérifier que le résultat est vrai pour $n = 1$ et après avoir posé l'hypothèse de récurrence $(H_n)$, justifier que les variables aléatoires $\sum_{k=1}^{n} X_1^{(k)}$ et $X_1^{(n+1)}$ sont indépendantes.

b. Conclure.

c. Si $X_n$ et $X_m$ (avec $m \in \mathbb{N}^*$) sont deux variables aléatoires réelles indépendantes telles que $X_n \hookrightarrow \mathcal{B}^{-1}(n,p)$ et $X_m \hookrightarrow \mathcal{B}^{-1}(m,p)$, que dire (sans démonstration) de $X_n + X_m$ ?

## Partie C : Une loi binomiale et une loi normale

6. Pour tout $m \in \mathbb{N}^*$, on note $Y_m$ la variable aléatoire réelle correspondant au nombre de jours avec alerte pendant les $m$ premiers jours.

   a. Justifier que $Y_{n+k}$ suit une loi binomiale dont on précisera les paramètres.

   b. Démontrer que les événements $[X_n > k]$ et $[Y_{n+k} < n]$ sont égaux.

7. Soient $(a,b) \in \mathbb{R}^2$ tel que $a < b$ et $Z_n$ une variable aléatoire réelle suivant une loi normale de paramètres la moyenne de $X_n$ et la variance de $X_n$.

   a. Justifier que :

$$Z_n \hookrightarrow \mathcal{N}\left(\frac{nq}{p}, \frac{nq}{p^2}\right)$$

   b. Énoncer précisément le théorème de la limite centrale (en définissant toutes les notations) et en déduire que les deux suites ci-dessous convergent vers une même limite $\ell$ qu'on donnera sous forme d'une intégrale qu'on ne cherchera pas à calculer :

$$\left(\mathbb{P}\left(a < \frac{\frac{X_n}{n} - \frac{q}{p}}{\frac{\sqrt{q}}{p\sqrt{n}}} < b\right)\right)_{n\in\mathbb{N}^*}, \left(\mathbb{P}\left(a < \frac{\frac{Z_n}{n} - \frac{q}{p}}{\frac{\sqrt{q}}{p\sqrt{n}}} < b\right)\right)_{n\in\mathbb{N}^*}.$$

* *
*