



PRÉPASIMPLE

prepasimple.fr

EDHEC

Mathématiques

2000

Corrigé de mathématiques approfondies

Mathématiques approfondies — Prépa ECG

Annales de concours · Corrigé

Exercice 1

1) Probabilité conditionnelle de panne.

Question adaptée au programme actuel. La continuité de f sur la demi-droite positive a été ajoutée : sans régularité de f au point t , l'équivalent de la question b n'est pas garanti. La densité est nulle pour les réels strictement négatifs ; sa valeur en zéro est choisie par continuité à droite.

Pour tout $t \geq 0$, on a $F(t) < 1$. En effet, f est continue et strictement positive sur $]0, +\infty[$, donc son intégrale sur $[t+1, t+2]$ est strictement positive. Ainsi $P(X > t) > 0$, ce qui autorise toutes les probabilités conditionnelles et divisions qui suivent. Par ailleurs, une variable à densité ne prend aucune valeur donnée avec une probabilité positive.

- a. L'événement « fonctionne encore à l'instant t » est $\{X > t\}$. Sous cette condition, une panne avant $t+h$ correspond à $\{t < X \leq t+h\}$. La définition de la probabilité conditionnelle donne

$$p(t, h) = \frac{P(t < X \leq t+h)}{P(X > t)} = \frac{F(t+h) - F(t)}{1 - F(t)}.$$

Les choix d'inégalités strictes ou larges aux bornes ne modifient pas cette probabilité.

- b. Comme f est continue en $t > 0$, la fonction de répartition est dérivable en t et $F'(t) = f(t)$. Par définition de la dérivée,

$$F(t+h) - F(t) = f(t)h + o(h) \quad (h \rightarrow 0^+).$$

On divise par le nombre strictement positif $1 - F(t)$. Puisque $f(t) > 0$, le coefficient principal est non nul ; on obtient bien

$$p(t, h) \underset{h \rightarrow 0^+}{\sim} \frac{f(t)}{1 - F(t)} h = \lambda_X(t)h.$$

Le taux de panne mesure donc la probabilité de panne par unité de temps, sur une courte durée, conditionnellement à la survie jusqu'à t .

2) Détermination de la loi à partir du taux de panne.

Question adaptée au programme actuel. La continuité est demandée sur $[0, +\infty[$ et non sur toute la droite réelle. Une densité exponentielle usuelle n'est pas continue en zéro sur toute la droite ; cette précision rend la question b cohérente avec les hypothèses.

- a.** La variable étant positive et à densité, $F(0) = 0$ et $F(t) = 0$ pour $t < 0$. Sur $[0, t]$, le dénominateur $1 - F(u)$ est strictement positif. La continuité de f sur cette demi-droite assure celle de λ_X sur le segment ; l'intégrale demandée existe.

Pour $u > 0$, la dérivation du logarithme donne

$$\frac{d}{du} [-\ln(1 - F(u))] = \frac{F'(u)}{1 - F(u)} = \lambda_X(u).$$

Le théorème fondamental du calcul intégral, avec la continuité à droite en zéro, fournit

$$\int_0^t \lambda_X(u) du = -\ln(1 - F(t)) + \ln(1 - F(0)) = -\ln(1 - F(t)).$$

On exponentie puis on isole $F(t)$:

$$F(t) = 1 - \exp \left(- \int_0^t \lambda_X(u) du \right) \quad (t \geq 0).$$

Avec $F(t) = 0$ pour $t < 0$, cette formule détermine toute la fonction de répartition, donc la loi de X .

- b.** Si X suit la loi exponentielle de paramètre $c > 0$, on peut choisir $f(t) = ce^{-ct}$ pour $t \geq 0$, et $f(t) = 0$ pour $t < 0$. Pour $t \geq 0$, $1 - F(t) = e^{-ct}$, d'où

$$\lambda_X(t) = \frac{ce^{-ct}}{e^{-ct}} = c.$$

Réciproquement, si le taux de panne est constant, égal à c , alors $c > 0$, car $\lambda_X(t) > 0$ pour $t > 0$. La question a donne

$$F(t) = 1 - e^{-ct} \quad (t \geq 0), \quad F(t) = 0 \quad (t < 0).$$

Il s'agit exactement de la fonction de répartition de la loi exponentielle de paramètre c . Les lois exponentielles sont donc les seules lois du cadre de l'exercice dont le taux de panne est constant.

3) Taux de panne $\lambda_X(t) = t^3$.

La formule précédente donne, pour $t \geq 0$,

$$P(X > t) = \exp \left(- \int_0^t u^3 du \right) = e^{-t^4/4}.$$

La fonction de répartition correspondante est $F(t) = 1 - e^{-t^4/4}$ pour $t \geq 0$, et 0 sinon. Elle possède la densité $f(t) = t^3 e^{-t^4/4}$ pour $t \geq 0$, nulle pour $t < 0$. Cette densité satisfait les hypothèses : elle est continue sur la demi-droite positive, strictement positive pour $t > 0$ et son intégrale vaut $F(+\infty) - F(0) = 1$.

a. Survivre plus d'un an signifie $X > 1$. Ainsi

$$P(X > 1) = e^{-1/4} \approx 0,7788.$$

b. La formulation «encore plus de deux ans» désigne deux années supplémentaires après l'âge de un an. On cherche donc $P(X > 3 \mid X > 1)$. Comme $\{X > 3\} \subset \{X > 1\}$ et $P(X > 1) > 0$,

$$P(X > 3 \mid X > 1) = \frac{P(X > 3)}{P(X > 1)} = \frac{e^{-81/4}}{e^{-1/4}} = e^{-20} \approx 2,061 \times 10^{-9}.$$

Précision locale : le PDF dit seulement « survive plus de deux ans », sans expliciter si les deux ans sont une durée restante ou un âge total. L'énoncé adapté précise ici la durée restante. Si l'on demandait plutôt la survie au-delà de l'âge de deux ans, la réponse serait $P(X > 2 \mid X > 1) = e^{-15/4} \approx 0,02352$.

Exercice 2

On travaille avec $x \in [0, 1[$. Pour alléger les expressions, on note

$$H_k = \sum_{p=1}^k \frac{1}{p}$$

pour $k \geq 1$.

1) Encadrement des sommes harmoniques.

a. Sur le segment $[p, p+1]$, avec $p \geq 1$, la fonction $t \mapsto 1/t$ est continue et décroissante.

On a donc

$$\frac{1}{p+1} \leq \frac{1}{t} \leq \frac{1}{p}.$$

En intégrant sur un segment de longueur 1, la croissance de l'intégrale donne

$$\frac{1}{p+1} \leq \int_p^{p+1} \frac{dt}{t} \leq \frac{1}{p}.$$

b. Pour $k = 1$, $H_1 - \ln 1 = 1$, donc l'encadrement est vérifié. Supposons $k \geq 2$. On somme les inégalités de a pour $p = 1, \dots, k-1$ et on utilise la relation de Chasles :

$$\sum_{p=1}^{k-1} \frac{1}{p+1} \leq \int_1^k \frac{dt}{t} \leq \sum_{p=1}^{k-1} \frac{1}{p}.$$

Ainsi $H_k - 1 \leq \ln k \leq H_k - 1/k$, ce qui donne

$$\frac{1}{k} \leq H_k - \ln k \leq 1.$$

En particulier, pour tout $k \geq 1$, $0 \leq H_k - \ln k \leq 1$, comme demandé.

2) Somme de la série de terme général x^n/n .

a. On part de l'identité géométrique finie, valable pour $t \in [0, x]$ puisque $x < 1$:

$$\sum_{p=1}^n t^{p-1} = \frac{1-t^n}{1-t} = \frac{1}{1-t} - \frac{t^n}{1-t}.$$

Tous les termes sont continus sur le segment $[0, x]$. En intégrant cette égalité, la linéarité pour une somme finie donne

$$\sum_{p=1}^n \frac{x^p}{p} = \int_0^x \frac{dt}{1-t} - \int_0^x \frac{t^n}{1-t} dt = -\ln(1-x) - \int_0^x \frac{t^n}{1-t} dt.$$

Pour $x = 0$, les deux membres valent zéro. Aucun échange entre une intégrale et une série infinie n'a été utilisé.

b. Si $x = 0$, tous les termes de la série sont nuls, donc sa somme est $0 = -\ln 1$. Fixons désormais $x \in]0, 1[$. Sur $[0, x]$, on a $1/(1-t) \leq 1/(1-x)$, donc

$$0 \leq \int_0^x \frac{t^n}{1-t} dt \leq \frac{1}{1-x} \int_0^x t^n dt = \frac{x^{n+1}}{(n+1)(1-x)}.$$

La borne supérieure tend vers zéro, car $0 < x < 1$. La formule de a montre alors que les sommes partielles convergent vers $-\ln(1-x)$. Par conséquent,

$$\sum_{n=1}^{+\infty} \frac{x^n}{n} = -\ln(1-x) \quad (x \in [0, 1[).$$

3) Convergence de la série de terme général $\ln(n)x^n$.

a. Fixons $x \in]0, 1[$ et considérons $n \geq 2$, afin que l'argument du logarithme soit strictement positif. On écrit

$$\ln(n^2 \ln(n)x^n) = 2 \ln n + \ln(\ln n) + n \ln x.$$

Les croissances comparées donnent $\ln n/n \rightarrow 0$. De plus, pour n assez grand, $0 \leq \ln(\ln n) \leq \ln n$, donc $\ln(\ln n)/n \rightarrow 0$. Ainsi

$$\frac{1}{n} \ln(n^2 \ln(n)x^n) \rightarrow \ln x < 0.$$

Pour n assez grand, ce quotient est inférieur à $\frac{1}{2} \ln x < 0$; le logarithme étudié est alors inférieur à $\frac{n}{2} \ln x$, qui tend vers $-\infty$. Donc

$$\lim_{n \rightarrow +\infty} \ln(n^2 \ln(n)x^n) = -\infty.$$

La précision $n \geq 2$ ajoutée à l'énoncé évite le logarithme de zéro pour $n = 1$; elle ne modifie pas la limite.

b. Pour $x \in]0, 1[$, on exponentie le résultat de a :

$$n^2 \ln(n)x^n \rightarrow 0.$$

Autrement dit, $\ln(n)x^n = o(1/n^2)$. Les termes étant positifs ou nuls, ils sont, pour n assez grand, majorés par $1/n^2$. La série de Riemann

$$\sum_{n \geq 1} 1/n^2$$

converge ; la comparaison assure donc la convergence de

$$\sum_{n \geq 1} \ln(n)x^n.$$

Pour $x = 0$, tous les termes sont nuls. La convergence est ainsi établie pour tout $x \in [0, 1[$, et $S(x)$ est bien défini.

4) Équivalent de $S(x)$ au voisinage de 1^- .

a. D'après 1 b, $0 \leq H_k - \ln k \leq 1$. Comme $x^k \geq 0$, la multiplication par x^k conserve les inégalités. On somme ensuite pour $k = 1, \dots, n$:

$$0 \leq \sum_{k=1}^n (H_k - \ln k)x^k \leq \sum_{k=1}^n x^k.$$

Or

$$\sum_{k=1}^n H_k x^k = \sum_{k=1}^n \sum_{p=1}^k x^k / p$$

et

$$\sum_{k=1}^n \ln(k)x^k = S_n(x).$$

Cela donne l'encadrement demandé.

- b.** La double somme est finie : on peut regrouper ses termes selon l'indice p sans condition de convergence. Les indices satisfont $1 \leq p \leq k \leq n$, donc

$$\sum_{k=1}^n \sum_{p=1}^k \frac{x^k}{p} = \sum_{p=1}^n \frac{1}{p} \sum_{k=p}^n x^k.$$

La somme géométrique finie vaut

$$\sum_{k=p}^n x^k = (x^p - x^{n+1})/(1 - x),$$

y compris pour $x = 0$. Par conséquent,

$$\sum_{k=1}^n \sum_{p=1}^k \frac{x^k}{p} = \frac{1}{1-x} \left(\sum_{p=1}^n \frac{x^p}{p} - x^{n+1} H_n \right).$$

D'autre part,

$$\sum_{k=1}^n x^k = x(1 - x^n)/(1 - x) \leq x/(1 - x).$$

En remplaçant dans a, on obtient

$$0 \leq \frac{1}{1-x} \left(\sum_{p=1}^n \frac{x^p}{p} - x^{n+1} H_n \right) - S_n(x) \leq \frac{x}{1-x}.$$

c. Méthode 1 — Utiliser l'encadrement harmonique.

Pour $x = 0$, l'expression est nulle. Pour $x \in]0, 1[$, la question 1 b donne

$$0 \leq x^{n+1}H_n \leq x^{n+1}(1 + \ln n).$$

Posons $c = -\ln x > 0$. Alors $x^{n+1} = e^{-c(n+1)}$. Pour n assez grand, $\ln n \leq n$, et les croissances comparées assurent que $ne^{-cn} \rightarrow 0$. Ainsi $x^{n+1}(1 + \ln n) \rightarrow 0$. Par encadrement, $x^{n+1}H_n \rightarrow 0$.

Méthode 2 — Une majoration plus simple suffit.

Chaque terme $1/p$ étant inférieur ou égal à 1, on a $0 \leq H_n \leq n$. Si $0 < x < 1$, les croissances comparées donnent $nx^{n+1} \rightarrow 0$, d'où

$$0 \leq x^{n+1}H_n \leq nx^{n+1} \rightarrow 0.$$

Le cas $x = 0$ reste immédiat car l'expression est identiquement nulle. La première méthode réutilise l'estimation fine de l'exercice ; la seconde montre qu'une borne grossière suffit pour cette limite à x fixé.

d. Fixons d'abord $x \in]0, 1[$. Dans l'encadrement de b, les questions 2 b, 3 b et 4 c permettent de faire tendre n vers $+\infty$. Il vient

$$0 \leq -\frac{\ln(1-x)}{1-x} - S(x) \leq \frac{x}{1-x}.$$

Posons $L(x) = -\ln(1-x)/(1-x)$. Pour $0 < x < 1$, $L(x) > 0$. On divise l'encadrement par ce nombre :

$$0 \leq 1 - \frac{S(x)}{L(x)} \leq \frac{x}{-\ln(1-x)}.$$

Lorsque $x \rightarrow 1^-$, le numérateur de la borne supérieure tend vers 1 et son dénominateur vers $+\infty$. Le théorème d'encadrement donne $S(x)/L(x) \rightarrow 1$. Ainsi

$$S(x) \underset{x \rightarrow 1^-}{\sim} -\frac{\ln(1-x)}{1-x}.$$

On a successivement pris une limite en n à x fixé, puis une limite en x . Aucun échange de ces limites n'est nécessaire.

Exercice 3

On note D l'événement « la personne est d'accord avec l'affirmation A » et C l'événement « la carte tirée porte le numéro 1 ». On a $P(D) = p$, $P(C) = 1/20$ et C est indépendant de D .

- 1) Le système $(C, \overline{C})$ est complet, avec des probabilités non nulles. Si C est réalisé, la réponse est vraie exactement lorsque D l'est ; sinon, la réponse est vraie exactement lorsque $\overline{D}$ l'est. Les probabilités totales donnent

$$\theta = P(V) = P(C)P(V | C) + P(\overline{C})P(V | \overline{C}) = \frac{p}{20} + \frac{19}{20}(1-p).$$

Ainsi

$$\theta = \frac{19}{20} - \frac{9}{10}p, \quad p = \frac{19 - 20\theta}{18} = \frac{19}{18} - \frac{10}{9}\theta.$$

Puisque $0 \leq p \leq 1$, les valeurs possibles de θ forment le segment $[1/20, 19/20]$. En particulier, $0 < \theta < 1$ dans tout le modèle, ce qui sera utile pour le théorème central limite.

2) Cas $p = 17/18$.

a. En remplaçant p dans l'expression précédente,

$$\theta = \frac{19}{20} - \frac{9}{10} \frac{17}{18} = \frac{19}{20} - \frac{17}{20} = \frac{1}{10}.$$

b. On cherche $P(D | V)$. L'intersection $D \cap V$ est exactement $D \cap C$: une personne qui approuve A ne répond vrai que si sa carte porte le numéro 1. Par indépendance,

$$P(D \cap V) = P(D \cap C) = \frac{p}{20} = \frac{17}{360}.$$

Comme $P(V) = 1/10 > 0$, on peut conditionner par V :

$$P(D | V) = \frac{P(D \cap V)}{P(V)} = \frac{17/360}{1/10} = \frac{17}{36}.$$

3) Estimation de θ .

Question adaptée au programme actuel. L'approximation par des tirages avec remise est justifiée par une population grande devant la taille de l'échantillon, et non par une grande taille d'échantillon à elle seule. Les indépendances entre les tirages de personnes et de cartes sont explicitées. Les calculs suivants sont effectués dans ce modèle avec remise.

a. Pour $i \in \{1, \dots, n\}$, notons Y_i l'indicatrice de la réponse « vrai » de la personne numéro i . Les Y_i sont indépendantes, de même loi de Bernoulli de paramètre θ , et

$$S_n = \sum_{i=1}^n Y_i.$$

La somme de n variables de Bernoulli indépendantes de même paramètre suit une loi

binomiale :

$$S_n \sim \mathcal{B}(n, \theta), \quad P(S_n = k) = \binom{n}{k} \theta^k (1 - \theta)^{n-k} \quad (0 \leq k \leq n).$$

Ces variables ont un support fini, donc leurs espérances et variances existent. Par linéarité de l'espérance et par indépendance pour la variance,

$$E(S_n) = n\theta, \quad V(S_n) = n\theta(1 - \theta).$$

b. Posons $T_n = S_n/n$. C'est une statistique calculable à partir des réponses, qui ne dépend pas du paramètre inconnu. On a

$$E(T_n) = \frac{1}{n}E(S_n) = \theta, \quad V(T_n) = \frac{1}{n^2}V(S_n) = \frac{\theta(1 - \theta)}{n}.$$

L'égalité $E(T_n) = \theta$ montre l'absence de biais. Pour établir la convergence en probabilité, fixons $\varepsilon > 0$ et appliquons Bienaymé-Tchebychev :

$$P(|T_n - \theta| \geq \varepsilon) \leq \frac{V(T_n)}{\varepsilon^2} = \frac{\theta(1 - \theta)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2} \rightarrow 0.$$

Ainsi $T_n \rightarrow \theta$ en probabilité. L'estimateur T_n est donc sans biais et convergent.

4) Échantillon de taille 100, avec 23 réponses vraies.

a. La valeur observée de T_{100} est 23/100. En utilisant la relation affine entre les paramètres, on obtient

$$\hat{\theta} = 0,23, \quad \hat{p} = \frac{19 - 20 \times 0,23}{18} = \frac{14,4}{18} = 0,8.$$

Les deux estimations appartiennent aux ensembles de valeurs possibles de leurs paramètres.

- b.** Question adaptée au programme actuel. L'énoncé précise que l'intervalle obtenu par approximation normale est asymptotique, de niveau au moins 95%, et guide vers la majoration de la variance. Le théorème central limite ne garantit pas une couverture exacte à taille 100.

Méthode 1 — Théorème central limite et majoration de la variance.

Les variables Y_i sont indépendantes et de même loi, d'espérance θ et de variance finie strictement positive $\theta(1 - \theta)$. Le théorème central limite s'applique donc :

$$Z_n = \frac{\sqrt{n}(T_n - \theta)}{\sqrt{\theta(1 - \theta)}} \xrightarrow{\mathcal{L}} Z, \quad Z \sim \mathcal{N}(0, 1).$$

Pour chaque paramètre fixé θ , le nombre $c_\theta = 1,96/(2\sqrt{\theta(1 - \theta)})$ est constant et vérifie $c_\theta \geq 1,96$. La continuité de la fonction de répartition normale aux deux bornes donne

$$P\left(|T_n - \theta| \leq \frac{1,96}{2\sqrt{n}}\right) = P(|Z_n| \leq c_\theta) \longrightarrow 2\Phi(c_\theta) - 1 \geq 2\Phi(1,96) - 1 = 0,95.$$

Ainsi l'intervalle aléatoire

$$I_{\theta,n} = \left[T_n - \frac{1,96}{2\sqrt{n}}, T_n + \frac{1,96}{2\sqrt{n}}\right] \cap \left[\frac{1}{20}, \frac{19}{20}\right]$$

a une couverture asymptotique d'au moins 95%, pour tout θ du modèle. L'intersection avec l'ensemble des paramètres possibles ne modifie pas la couverture : le vrai paramètre appartient toujours à cet ensemble.

Pour $n = 100$ et $T_{100} = 0,23$, la demi-largeur vaut $1,96/20 = 0,098$. L'intersection ne change pas les bornes, et l'intervalle observé pour θ est

$$I_\theta = [0,132 ; 0,328].$$

La fonction $g(\theta) = (19 - 20\theta)/18$ est strictement décroissante. On inverse donc l'ordre des bornes pour obtenir l'intervalle de p :

$$I_p = \left[\frac{19 - 20 \times 0,328}{18} ; \frac{19 - 20 \times 0,132}{18} \right] = \left[0,8 - \frac{49}{450} ; 0,8 + \frac{49}{450} \right] \approx [0,6911 ; 0,9089].$$

L'événement $\{\theta \in I_{\theta,n}\}$ équivaut à l'événement d'appartenance de p à l'intervalle transformé, donc les deux intervalles ont la même couverture. Ces intervalles sont asymptotiques ; aucune substitution d'une variance estimée dans une convergence en loi n'est utilisée.

Méthode 2 — Une garantie d'au moins 95% à toute taille d'échantillon.

On peut également utiliser directement Bienaymé-Tchebychev. Comme $V(T_n) \leq 1/(4n)$, si $\delta_n = \sqrt{5/n}$, alors

$$P(|T_n - \theta| > \delta_n) \leq \frac{1}{4n\delta_n^2} = \frac{1}{20} = 0,05.$$

L'intervalle $[T_n - \delta_n, T_n + \delta_n] \cap [1/20, 19/20]$ a donc une couverture d'au moins 95% pour tout n dans le modèle binomial, et en particulier une couverture asymptotique au moins égale à ce niveau. Pour les données observées, $\delta_{100} = \sqrt{5}/10$, d'où

$$J_\theta = \left[\frac{1}{20} ; 0,23 + \frac{\sqrt{5}}{10} \right] \approx [0,05 ; 0,4536].$$

En transformant les bornes par g , on obtient

$$J_p = \left[0,8 - \frac{\sqrt{5}}{9} ; 1 \right] \approx [0,5515 ; 1].$$

La méthode normale donne un intervalle plus étroit avec une garantie asymptotique. La méthode de Bienaymé-Tchebychev donne un intervalle plus large avec une garantie valable à taille finie dans le modèle retenu.

Problème

Tous les espaces et toutes les matrices sont réels. La borne maximale $n = 4$, annoncée sans démonstration dans le préambule, est admise comme dans le sujet initial. Les questions démontrent les propriétés demandées et construisent explicitement une famille de quatre vecteurs.

Partie 1

On identifie les colonnes de $\mathcal{M}_{n,1}(\mathbb{R})$ aux vecteurs de $\mathbb{R}^n$. On note $\mathbf{1} = {}^t(1, \dots, 1)$ et $E_1, \dots, E_n$ les vecteurs de la base canonique de $\mathbb{R}^n$, pour les distinguer des vecteurs e_1, e_2, e_3 de $\mathbb{R}^3$.

1) Diagonalisation de J .

a. Méthode 1 — Théorème spectral réel.

La matrice J est réelle et symétrique, puisque ${}^tJ = J$. Le théorème spectral réel affirme qu'une telle matrice possède une base orthonormée de vecteurs propres. Elle est donc diagonalisable sur $\mathbb{R}$.

Méthode 2 — Construire une base propre.

Pour tout $X = {}^t(x_1, \dots, x_n)$, la multiplication matricielle donne

$$JX = \left(\sum_{i=1}^n x_i \right) \mathbf{1}.$$

Il s'ensuit que $J\mathbf{1} = n\mathbf{1}$ et que $JX = 0$ exactement lorsque

$$\sum_i x_i = 0.$$

Posons

$$H = \{X : \sum_i x_i = 0\}.$$

Les vecteurs $E_1 - E_n, \dots, E_{n-1} - E_n$ forment une base de H : ils sont libres, et tout $X \in H$ s'écrit

$$\sum_{i=1}^{n-1} x_i (E_i - E_n).$$

Comme $\mathbf{1} \notin H$, la famille

$$(\mathbf{1}, E_1 - E_n, \dots, E_{n-1} - E_n)$$

est une base de $\mathbb{R}^n$, entièrement formée de vecteurs propres de J . Cela prouve aussi sa diagonalisation. La première méthode s'appuie sur la symétrie ; la seconde décrit explicitement les sous-espaces propres qui serviront à l'étude de M_a .

b. Pour tous i, j , le coefficient (i, j) de J^2 vaut

$$\sum_{k=1}^n 1 \times 1 = n.$$

Donc

$$J^2 = nJ.$$

Si ρ est une valeur propre, il existe $X \neq 0$ tel que $JX = \rho X$. L'identité précédente appliquée à X donne $\rho^2 X = n\rho X$, donc $\rho(\rho - n) = 0$. Les seules valeurs propres possibles sont 0 et n .

Elles sont toutes deux effectivement présentes : $J\mathbf{1} = n\mathbf{1}$, et $J(E_1 - E_2) = 0$ avec $E_1 - E_2 \neq 0$, puisque $n \geq 2$. Ainsi

$$\text{Sp}(J) = \{0, n\}.$$

Les espaces propres sont H pour 0, de dimension $n - 1$, et $\text{Vect}(\mathbf{1})$ pour n , de dimension 1.

2) Éléments propres et inversibilité de M_a .

a. On utilise la base propre de J , comme demandé. D'après la définition des coefficients,

$$M_a = (1 - a)I + aJ.$$

Sur $\text{Vect}(\mathbf{1})$, J agit par multiplication par n , donc M_a agit par multiplication par $1 - a + an = 1 + (n - 1)a$. Sur H , J est nul, donc M_a agit par multiplication par $1 - a$. Dans la base propre précédente, la matrice de M_a est ainsi

$$\text{diag}(1 + (n - 1)a, 1 - a, \dots, 1 - a).$$

Si $a \neq 0$, les deux nombres sont distincts, car leur différence est $na \neq 0$. Ils sont les deux valeurs propres de M_a , associées respectivement à $\text{Vect}(\mathbf{1})$ et à H .

Si $a = 0$, alors $M_0 = I$ et sa seule valeur propre est 1, avec tout $\mathbb{R}^n$ pour espace propre.

Rectification locale : le PDF demande « les deux valeurs propres » sans restriction sur a . Pour $a = 0$, il n'y en a qu'une ; l'énoncé adapté distingue ce cas.

b. Méthode 1 — Utiliser la matrice diagonale.

Une matrice est inversible si et seulement si sa matrice dans une autre base l'est. Une matrice diagonale est inversible exactement lorsque tous ses coefficients diagonaux sont non nuls. La question a donc

$$M_a \text{ inversible} \iff 1 - a \neq 0 \text{ et } 1 + (n - 1)a \neq 0.$$

Puisque $n - 1 > 0$, cela équivaut à

$$a \neq 1 \quad \text{et} \quad a \neq -\frac{1}{n - 1}.$$

Méthode 2 — Résoudre directement le système homogène.

Pour $X = {}^t(x_1, \dots, x_n)$, posons

$$s = \sum_i x_i.$$

Les équations $M_a X = 0$ sont $(1 - a)x_i + as = 0$, pour tout i . En les sommant, on obtient $[1 + (n - 1)a]s = 0$.

Si $1 + (n - 1)a \neq 0$ et $a \neq 1$, alors $s = 0$, puis $x_i = 0$ pour tout i . Le noyau de M_a est réduit à zéro ; comme il s'agit d'un endomorphisme d'un espace de dimension finie, il est inversible.

Si $a = 1$, $M_a = J$ et $E_1 - E_2$ est un vecteur non nul du noyau. Si $a = -1/(n - 1)$, on a $M_a \mathbf{1} = 0$ avec $\mathbf{1} \neq 0$. Dans chacun de ces deux cas, la matrice n'est pas inversible. Cette résolution directe retrouve le critère sans diagonalisation ; la première méthode réutilise immédiatement la question a.

Partie 2

1) Relation linéaire entre les vecteurs.

La restriction $n \geq 2$ est explicitée dans cette question afin que $-1/(n - 1)$ et les résultats de la partie 1 soient utilisables. Le cas $n = 1$ sera traité séparément en question 2.

a. Supposons

$$\sum_{k=1}^n \lambda_k u_k = 0.$$

Pour chaque $i \in \{1, \dots, n\}$, on prend le produit scalaire avec u_i . Par bilinéarité,

$$0 = \sum_{k=1}^n \lambda_k (u_i | u_k) = \lambda_i + \alpha \sum_{k \neq i} \lambda_k.$$

On a utilisé $\|u_i\|^2 = 1$ et $(u_i | u_k) = \alpha$ pour $k \neq i$, la symétrie du produit scalaire étendant la propriété aux deux ordres d'indices. Le dernier membre est la composante numéro i de $M_\alpha \Lambda$, où $\Lambda = {}^t(\lambda_1, \dots, \lambda_n)$. Ainsi

$$M_\alpha \Lambda = 0.$$

- b.** Si $\alpha \neq 1$ et $\alpha \neq -1/(n-1)$, la partie 1 montre que M_α est inversible. La question a entraîne donc $\Lambda = 0$. Toute relation linéaire entre les u_i est triviale : la famille est libre.

Une famille libre dans $\mathbb{R}^3$, de dimension 3, contient au plus trois vecteurs. Donc $n \leq 3$. Cette borne est atteinte : la base canonique (e_1, e_2, e_3) est une famille de trois vecteurs unitaires distincts, de produits scalaires mutuels nuls, avec $\alpha = 0 \neq 1$ et $0 \neq -1/2$. La valeur maximale dans ce cas est donc 3.

2) Cas $\alpha = 1$.

- a.** Pour deux vecteurs u_i, u_j , l'inégalité de Cauchy-Schwarz s'écrit

$$|(u_i | u_j)| \leq \|u_i\| \|u_j\| = 1.$$

Il y a égalité si et seulement si ces deux vecteurs sont colinéaires. Ici ils sont non nuls car unitaires ; on peut donc écrire $u_j = cu_i$ en cas d'égalité. Les normes imposent $|c| = 1$. Lorsque $(u_i | u_j) = 1$, on a en outre $(u_i | cu_i) = c = 1$, donc $u_j = u_i$.

- b.** Méthode 1 — Utiliser le cas d'égalité.

S'il existait au moins deux vecteurs, on pourrait choisir $i \neq j$. Leur produit scalaire vaudrait 1 ; la question a imposerait $u_i = u_j$, en contradiction avec la distinction deux à deux. Ainsi $n \leq 1$. La famille étant non vide, $n = 1$. Pour un seul vecteur unitaire, la condition sur les couples $i < j$ est vide et ne pose aucune difficulté.

Méthode 2 — Calculer la distance entre deux vecteurs.

Pour $i \neq j$, la bilinéarité donne

$$\|u_i - u_j\|^2 = \|u_i\|^2 + \|u_j\|^2 - 2(u_i | u_j) = 1 + 1 - 2 = 0.$$

La norme est nulle uniquement pour le vecteur nul, donc $u_i = u_j$. On retrouve la contradiction si $n \geq 2$, et donc $n = 1$. Cette seconde méthode utilise directement le lien entre distance et produit scalaire; la première réemploie la question imposée sur Cauchy-Schwarz.

3) Famille de quatre vecteurs.

L'existence d'une telle famille est une hypothèse admise dans cette question; elle sera vérifiée par une construction en partie 3.

- a.** Comme les quatre vecteurs sont distincts, la question 2 exclut $\alpha = 1$. Si, de plus, $\alpha \neq -1/3$, la question 1 b, appliquée avec $n = 4$, imposerait $n \leq 3$. Cela contredirait l'hypothèse. On a donc nécessairement

$$\alpha = -\frac{1}{3}.$$

- b.** Supposons $c_1 u_1 + c_2 u_2 + c_3 u_3 = 0$. En prenant successivement le produit scalaire avec u_1, u_2, u_3 , on obtient $M_{-1/3}^{(3)} C = 0$, où $M_{-1/3}^{(3)}$ désigne la matrice de la partie 1 de taille 3 et $C = {}^t(c_1, c_2, c_3)$.

Pour cette taille, les coefficients diagonaux de sa diagonalisation sont

$$1 - \left(-\frac{1}{3}\right) = \frac{4}{3} \quad \text{et} \quad 1 + 2 \left(-\frac{1}{3}\right) = \frac{1}{3}.$$

Ils sont tous non nuls : la matrice est inversible. Ainsi $C = 0$, et (u_1, u_2, u_3) est libre. Une famille libre de trois vecteurs dans $\mathbb{R}^3$ est une base de cet espace.

c. Méthode 1 — Déterminer les coordonnées par produits scalaires.

Puisque (u_1, u_2, u_3) est une base, il existe un unique triplet (c_1, c_2, c_3) tel que $u_4 = c_1 u_1 + c_2 u_2 + c_3 u_3$. En prenant le produit scalaire avec chacun des trois vecteurs de base, on obtient

$$M_{-1/3}^{(3)} C = -\frac{1}{3} \mathbf{1}.$$

Chaque ligne de $M_{-1/3}^{(3)}$ a pour somme $1 - 1/3 - 1/3 = 1/3$, donc $M_{-1/3}^{(3)} \mathbf{1} = (1/3) \mathbf{1}$. Par conséquent $C = -\mathbf{1}$ satisfait le système. Son unicité découle de l'inversibilité établie en b. Les coordonnées demandées sont donc

$$(-1, -1, -1), \quad u_4 = -u_1 - u_2 - u_3.$$

Méthode 2 — Utiliser la norme de la somme des quatre vecteurs.

Posons $s = u_1 + u_2 + u_3 + u_4$. La bilinéarité et la symétrie donnent

$$\|s\|^2 = \sum_{i=1}^4 \|u_i\|^2 + 2 \sum_{1 \leq i < j \leq 4} (u_i | u_j) = 4 + 2 \times 6 \times \left(-\frac{1}{3}\right) = 0.$$

Donc $s = 0$ et $u_4 = -u_1 - u_2 - u_3$. La question b assure que cette écriture donne bien les coordonnées dans une base : $(-1, -1, -1)$. La première méthode est générale pour calculer des coordonnées dans une base non orthonormée ; la seconde exploite la symétrie particulière des quatre vecteurs.

Partie 3

1) La famille (e_1, e_2, e_3) convient. La base canonique est orthonormée : chaque vecteur a une norme égale à 1 et les produits scalaires de deux vecteurs différents valent 0. Les vecteurs sont distincts, donc toutes les conditions sont satisfaites pour $n = 3$ et $\alpha = 0$.

2) Construction d'une famille de quatre vecteurs.

a. Dans la base canonique orthonormée,

$$v_1 = (1, 0, 0), \quad v_2 = \left(-\frac{1}{2}, \frac{\sqrt{3}}{2}, 0\right), \quad v_3 = \left(-\frac{1}{2}, -\frac{\sqrt{3}}{2}, 0\right).$$

Leurs normes au carré sont 1, $1/4 + 3/4 = 1$ et $1/4 + 3/4 = 1$. Ils sont distincts : les premières coordonnées distinguent v_1 des deux autres, et les secondes coordonnées distinguent v_2 de v_3 .

Leurs produits scalaires sont

$$(v_1 | v_2) = -\frac{1}{2}, \quad (v_1 | v_3) = -\frac{1}{2}, \quad (v_2 | v_3) = \frac{1}{4} - \frac{3}{4} = -\frac{1}{2}.$$

La famille (v_1, v_2, v_3) satisfait donc toutes les conditions avec $\alpha = -1/2$.

b. Posons $w_i = \lambda v_i + \mu e_3$ pour $i = 1, 2, 3$. On procède par analyse puis synthèse, afin de vérifier toutes les conditions.

Analyse. Si (e_3, w_1, w_2, w_3) est une famille solution de cardinal 4, la partie 2 impose $\alpha = -1/3$. Comme chaque v_i est orthogonal à e_3 ,

$$(e_3 | w_i) = \mu.$$

Il faut donc $\mu = -1/3$. Puisque $\|v_i\| = \|e_3\| = 1$ et $(v_i | e_3) = 0$, on a $\|w_i\|^2 = \lambda^2 + \mu^2$. La condition de norme unitaire impose

$$\lambda^2 + \frac{1}{9} = 1, \quad \lambda^2 = \frac{8}{9}, \quad \lambda = \frac{2\sqrt{2}}{3} \text{ ou } \lambda = -\frac{2\sqrt{2}}{3}.$$

Synthèse. Pour chacun de ces deux choix de λ et pour $\mu = -1/3$, les vecteurs w_i sont unitaires. Pour $i \neq j$, la question a et les orthogonalités donnent

$$(w_i | w_j) = \lambda^2(v_i | v_j) + \mu^2 = -\frac{1}{2} \frac{8}{9} + \frac{1}{9} = -\frac{1}{3}.$$

On a également $(e_3 | w_i) = -1/3$. Puisque $\lambda \neq 0$ et que les v_i sont distincts, les w_i sont distincts deux à deux. De plus, aucun w_i n'est égal à e_3 , car son produit scalaire avec e_3 vaut $-1/3$, et non 1.

Toutes les conditions sont donc vérifiées. Les deux couples possibles sont

$$(\lambda, \mu) = \left(\frac{2\sqrt{2}}{3}, -\frac{1}{3} \right) \quad \text{ou} \quad (\lambda, \mu) = \left(-\frac{2\sqrt{2}}{3}, -\frac{1}{3} \right).$$

Par exemple, le premier couple fournit la famille explicite

$$\left((0, 0, 1), \left(\frac{2\sqrt{2}}{3}, 0, -\frac{1}{3} \right), \left(-\frac{\sqrt{2}}{3}, \frac{\sqrt{6}}{3}, -\frac{1}{3} \right), \left(-\frac{\sqrt{2}}{3}, -\frac{\sqrt{6}}{3}, -\frac{1}{3} \right) \right).$$

Son existence réalise le cardinal maximal 4 annoncé dans le préambule, la maximalité elle-même restant admise conformément à l'énoncé.

Relevé des adaptations

- Exercice 1, 1 et 1 b : ajout de la continuité de la densité sur $[0, +\infty[$ afin de justifier l'équivalent par $F'(t) = f(t)$; la nullité est précisée pour les réels strictement négatifs afin de choisir la valeur en zéro par continuité à droite.
- Exercice 1, 2 et 2 b : continuité de la densité précisée sur la demi-droite positive, ce qui permet de traiter les lois exponentielles sans imposer une continuité globale impossible pour leur densité usuelle en zéro.
- Exercice 1, 3 b : ajout du mot « encore » pour préciser que les deux ans sont une durée de survie supplémentaire ; le corrigé donne aussi la réponse correspondant à une survie au-delà de l'âge total de deux ans.
- Exercice 2, 3 a : précision $n \geq 2$ pour que $\ln(n^2 \ln(n)x^n)$ soit défini ; la limite est inchangée.
- Exercice 3, 1 et 3 : indépendance du tirage des cartes explicitée ; remplacement de l'hypothèse « n assez grand » par une population assez grande devant n pour justifier le modèle avec remise, avec indépendance entre les réponses.
- Exercice 3, 4 b : précision d'un niveau asymptotique au moins égal à 95% et indication $\theta(1 - \theta) \leq 1/4$. Le corrigé justifie l'intervalle sans remplacer la variance inconnue dans une convergence en loi et propose aussi un intervalle garanti à taille finie par Bienaymé-Tchebychev.
- Problème, partie 1, 2 a : distinction du cas $a = 0$, où $M_a = I$ ne possède qu'une valeur propre et non deux.
- Problème, partie 2, 1 : précision $n \geq 2$ pour éviter le quotient $-1/(n - 1)$ lorsque $n = 1$; ce dernier cas est conservé dans la question 2.