



PRÉPASIMPLE

prepasimple.fr

HEC — 2014

Corrigé de mathématiques approfondies

Mathématiques approfondies — Prépa ECG

Annales de concours · Corrigé

Corrigé

Rappels et conventions utilisés dans tout le corrigé.

On note u la colonne de $\mathcal{M}_{k,1}(\mathbb{R})$ dont tous les coefficients valent 1 (notation de la partie I, réutilisée partout sauf en 8, où l'énoncé appelle aussi u une colonne quelconque). Pour $x \in \mathcal{M}_{k,1}(\mathbb{R})$,

$${}^tux = \sum_{i=1}^k x_i = \langle u, x \rangle,$$

et $\langle x, y \rangle = {}^txy$.

Une matrice Q de $\mathcal{M}_k(\mathbb{R})$ est orthogonale si ${}^tQQ = I_k$; elle est alors inversible d'inverse tQ . C'est le cas d'une matrice dont les colonnes forment une base orthonormale de $\mathcal{M}_{k,1}(\mathbb{R})$.

Théorème spectral : toute matrice symétrique réelle A est diagonalisable dans une base orthonormale de vecteurs propres, c'est-à-dire qu'il existe P orthogonale et D diagonale réelle telles que $A = PD{}^tP$; les sous-espaces propres de A sont deux à deux orthogonaux.

Covariance de variables discrètes admettant une variance : elle existe, elle est bilinéaire et symétrique, $\text{Cov}(Z, Z) = V(Z)$, $\text{Cov}(Z, c) = 0$ pour une constante c , et $\text{Cov}(Z, Z') = 0$ si Z et Z' sont indépendantes. Une variable discrète Z admettant une variance vérifie $V(Z) = 0$ si et seulement si $P([Z = E(Z)]) = 1$. Si $Z \geq 0$ est discrète et $E(Z) = 0$, alors $P([Z = 0]) = 1$ (chaque valeur $z > 0$ vérifie $zP([Z = z]) \leq E(Z) = 0$).

Formule de l'espérance totale (cadre discret) : si (A_n) est un système complet d'événements de probabilités non nulles et si Z est une variable discrète admettant une espérance, alors Z admet une espérance pour chaque P_{A_n} et

$$E(Z) = \sum_n P(A_n)E(Z|A_n),$$

la série convergeant absolument.

Lois usuelles : si $Z \hookrightarrow \mathcal{B}(n, \theta)$, $E(Z) = n\theta$ et $V(Z) = n\theta(1 - \theta)$; si $G \hookrightarrow \mathcal{G}(\theta)$ avec $0 < \theta \leq 1$ (loi géométrique sur $\mathbb{N}^*$, $P([G = m]) = (1 - \theta)^{m-1}\theta$), $E(G) = \frac{1}{\theta}$ et $V(G) = \frac{1 - \theta}{\theta^2}$. Une somme de n variables de Bernoulli de même paramètre θ mutuellement indépendantes suit la loi $\mathcal{B}(n, \theta)$.

Séries géométriques dérivées : pour $|x| < 1$,

$$\sum_{i=1}^{+\infty} ix^{i-1} = \frac{1}{(1-x)^2}$$

et

$$\sum_{i=2}^{+\infty} i(i-1)x^{i-2} = \frac{2}{(1-x)^3}.$$

Ces formules restent vraies pour $x = 0$ avec la convention $0^0 = 1$.

Partie I. Lois généralisées de Bernoulli

1. a) Méthode : calcul direct du produit d'une colonne par une ligne.

Le coefficient d'indice (i, j) de $M = a^t u$ est $a_i \times 1 = a_i$. Ainsi

$$M = \begin{pmatrix} a_1 & a_1 & \cdots & a_1 \\ a_2 & a_2 & \cdots & a_2 \\ \vdots & \vdots & & \vdots \\ a_k & a_k & \cdots & a_k \end{pmatrix}.$$

Toutes les colonnes de M sont égales à a , qui est non nulle : l'espace engendré par les colonnes est $\text{Vect}(a)$, de dimension 1.

Conclusion

M est la matrice dont toutes les colonnes valent a , et $\text{rg}(M) = 1$.

b) Méthode : associativité du produit matriciel.

$$Ma = a ({}^t u a) = a \left(\sum_{i=1}^k a_i \right) = \alpha a.$$

Comme $a \neq 0$, a est un vecteur propre de M associé à α .

Conclusion

$Ma = \alpha a$, donc α est valeur propre de M (de vecteur propre a).

c) Méthode : polynôme annulateur.

$M^2 = a ({}^t u a) {}^t u = \alpha a {}^t u = \alpha M$. Le polynôme $X^2 - \alpha X = X(X - \alpha)$ annule M ; toute valeur propre de M en est racine. Concrètement, si $Mx = \mu x$ avec $x \neq 0$, alors $M^2 x = \mu^2 x$ et $\alpha Mx = \alpha \mu x$, donc $\mu(\mu - \alpha)x = 0$ et $\mu \in \{0, \alpha\}$.

De plus 0 est effectivement valeur propre, car $\text{rg}(M) = 1 < k$ (puisque $k \geq 2$), donc M n'est pas inversible; et α est valeur propre d'après b).

Conclusion

$M^2 = \alpha M$ et les valeurs propres de M sont exactement 0 et α (ces deux valeurs étant confondues si $\alpha = 0$).

d) Méthode : somme des dimensions des sous-espaces propres.

Pour $x \in \mathcal{M}_{k,1}(\mathbb{R})$,

$$Mx = a ({}^t u x) = \left(\sum_{i=1}^k x_i \right) a.$$

Comme $a \neq 0$,

$$\text{Ker } M = \left\{ x ; \sum_{i=1}^k x_i = 0 \right\},$$

de dimension $k - 1$ (théorème du rang, avec $\text{rg}(M) = 1$).

Si $\alpha \neq 0$: les valeurs propres sont 0 et α , distinctes. Le sous-espace propre associé à 0 est $\text{Ker } M$, de dimension $k - 1$, et le sous-espace propre associé à α contient $a \neq 0$, donc est de dimension au moins 1. La somme des dimensions des sous-espaces propres vaut au moins k , donc exactement k : M est diagonalisable.

Si $\alpha = 0$: la seule valeur propre de M est 0. Si M était diagonalisable, elle serait semblable à la matrice nulle, donc nulle, ce qui contredit $\text{rg}(M) = 1$. Autre formulation : le seul sous-espace propre est $\text{Ker } M$, de dimension $k - 1 < k$.

Conclusion

M est diagonalisable si et seulement si $\alpha \neq 0$.

e) Méthode 1 — Valeurs propres. $I_k - M$ est inversible si et seulement si $(I_k - M)x = 0$ n'a que la solution nulle, c'est-à-dire si et seulement si 1 n'est pas valeur propre de M . D'après c), les valeurs propres de M sont 0 et α ; comme $1 \neq 0$, la condition équivaut à $\alpha \neq 1$.

Méthode 2 — Inverse explicite. Si $\alpha \neq 1$, posons $N = I_k + \frac{1}{1-\alpha}M$. Avec $M^2 = \alpha M$:

$$(I_k - M)N = I_k + \frac{1}{1-\alpha}M - M - \frac{\alpha}{1-\alpha}M = I_k + \left(\frac{1-\alpha}{1-\alpha} - 1\right)M = I_k,$$

donc $I_k - M$ est inversible d'inverse N . Si $\alpha = 1$, $(I_k - M)a = a - a = 0$ avec $a \neq 0$, donc $I_k - M$ n'est pas inversible.

Comparaison : la méthode 1 est immédiate grâce à c) ; la méthode 2 fournit en plus l'inverse.

Conclusion

$$I_k - M \text{ est inversible si et seulement si } \alpha \neq 1, \text{ et alors } (I_k - M)^{-1} = I_k + \frac{1}{1 - \alpha} M.$$

f) Méthode : caractérisation $M^2 = M$ des projecteurs, puis orthogonalité du noyau et de l'image.

Si $\alpha = 1$, c) donne $M^2 = M$: l'endomorphisme de $\mathbb{R}^k$ canoniquement associé à M est un projecteur. Son image est l'espace engendré par les colonnes, $\text{Vect}(a)$ (question a)), et son noyau est

$$\text{Ker } M = \left\{ x ; \sum_{i=1}^k x_i = 0 \right\}$$

(question d)), c'est-à-dire l'orthogonal de $\text{Vect}(u)$, puisque

$$\sum x_i = \langle u, x \rangle.$$

Méthode 1 — Noyau orthogonal à l'image. Le projecteur est orthogonal si et seulement si son noyau et son image sont orthogonaux, c'est-à-dire si et seulement si $a \in (\text{Vect}(u)^\perp)^\perp = \text{Vect}(u)$. Cela s'écrit $a = cu$ avec c réel, et la condition

$$\sum a_i = 1$$

impose $kc = 1$, soit $a = \frac{1}{k}u$.

Méthode 2 — Symétrie de la matrice. La base canonique étant orthonormale, un projecteur est orthogonal si et seulement si sa matrice dans cette base est symétrique. Or le coefficient (i, j) de M est a_i et le coefficient (j, i) est a_j : M est symétrique si et seulement si tous les a_i sont égaux, donc égaux à $\frac{1}{k}$ puisque leur somme vaut 1.

Comparaison : la méthode 1 utilise la définition géométrique ; la méthode 2 est un calcul immédiat sur la matrice.

Conclusion

pour $\alpha = 1$, M est la matrice du projecteur sur $\text{Vect}(a)$ parallèlement à l'hyperplan

$$\left\{ x ; \sum_{i=1}^k x_i = 0 \right\} ;$$

ce projecteur est orthogonal si et seulement si $a = \frac{1}{k}u$, c'est-à-dire si M est la matrice dont tous les coefficients valent $\frac{1}{k}$.

2. a) Méthode : description presque sûre des valeurs de X .

Si $X = e_i$, la i -ème composante de X vaut 1 : $[X = e_i] \subset [X_i = 1]$.

Les événements $[X = e_1], \dots, [X = e_k]$ sont deux à deux incompatibles et

$$\sum_{j=1}^k P([X = e_j]) = \sum p_j = 1.$$

Donc l'événement $D = \bigcup_{j=1}^k [X = e_j]$ est de probabilité 1 et son contraire $\overline{D}$ est négligeable. Sur D , X est égal à l'un des e_j , dont la seule composante égale à 1 est la j -ème ; si de plus $X_i = 1$, alors $j = i$. Ainsi $[X_i = 1] \cap D = [X = e_i]$, et $[X_i = 1]$ ne diffère de $[X = e_i]$ que par une partie de l'événement négligeable $\overline{D}$.

On en déduit $P([X_i = 1]) = P([X = e_i]) = p_i$. De même, $\bigcup_{j \neq i} [X = e_j] \subset [X_i = 0]$, donc

$P([X_i = 0]) \geq 1 - p_i$, d'où $P([X_i = 0]) = 1 - p_i$: X_i prend presque sûrement ses valeurs dans $\{0, 1\}$ et suit la loi de Bernoulli $\mathcal{B}(p_i)$, d'espérance p_i .

Conclusion

$[X = e_i] \subset [X_i = 1]$, ces deux événements étant égaux à un événement négligeable près (égaux si l'on convient que X ne prend que les valeurs $e_1, \dots, e_k$) ; $X_i \hookrightarrow \mathcal{B}(p_i)$ et $\mathcal{E}(X) = p$.

b) Méthode : valeur de $X_1 + X_2$ sur chacun des événements $[X = e_j]$.

Sur $[X = e_1] \cup [X = e_2]$, $X_1 + X_2 = 1$; sur $[X = e_j]$ avec $j \geq 3$, $X_1 + X_2 = 0$. Ces événements recouvrent D , de probabilité 1. Donc $X_1 + X_2$ prend presque sûrement les valeurs 0 et 1, avec $P([X_1 + X_2 = 1]) = p_1 + p_2$.

Conclusion

$X_1 + X_2 \hookrightarrow \mathcal{B}(p_1 + p_2)$.

c) Méthode 1 — Formule de König-Huygens. Aucun e_j n'a deux composantes non nulles, donc $X_1 X_2 = 0$ sur D , c'est-à-dire presque sûrement, et $E(X_1 X_2) = 0$. Par suite $\text{Cov}(X_1, X_2) = E(X_1 X_2) - E(X_1)E(X_2) = -p_1 p_2$.

Méthode 2 — Variance d'une somme. $V(X_1 + X_2) = V(X_1) + V(X_2) + 2\text{Cov}(X_1, X_2)$.
D'après a) et b) :

$$(p_1 + p_2)(1 - p_1 - p_2) = p_1(1 - p_1) + p_2(1 - p_2) + 2\text{Cov}(X_1, X_2).$$

Le membre de gauche vaut $p_1 + p_2 - p_1^2 - 2p_1 p_2 - p_2^2$, d'où $2\text{Cov}(X_1, X_2) = -2p_1 p_2$.

Comparaison : la méthode 1 est la plus directe ; la méthode 2 réutilise b).

Conclusion

$$\text{Cov}(X_1, X_2) = -p_1 p_2.$$

d) Méthode : généralisation de c) à un couple quelconque d'indices.

Le raisonnement de c) s'applique à tout couple (i, j) avec $i \neq j : X_i X_j = 0$ presque sûrement, donc $\text{Cov}(X_i, X_j) = -p_i p_j$. Et $V(X_i) = p_i(1 - p_i)$ par a).

Conclusion

$$\mathcal{V}(X) = \begin{pmatrix} p_1(1 - p_1) & -p_1 p_2 & \cdots & -p_1 p_k \\ -p_2 p_1 & p_2(1 - p_2) & \cdots & -p_2 p_k \\ \vdots & \vdots & \ddots & \vdots \\ -p_k p_1 & -p_k p_2 & \cdots & p_k(1 - p_k) \end{pmatrix},$$

c'est-à-dire $\mathcal{V}(X) = \text{Diag}(p) - p {}^t p$.

3. a) Méthode : calcul matriciel.

${}^t u \text{Diag}(p)$ est la ligne de coefficients $1 \times p_j = p_j$, c'est-à-dire ${}^t p$. Donc

$$(I_k - M(p)) \text{Diag}(p) = \text{Diag}(p) - p ({}^t u \text{Diag}(p)) = \text{Diag}(p) - p {}^t p,$$

dont le coefficient (i, j) est $p_i - p_i^2$ si $i = j$ et $-p_i p_j$ si $i \neq j$.

Conclusion

d'après 2. d), $\mathcal{V}(X) = (I_k - M(p)) \text{Diag}(p)$.

b) Méthode : rang d'un projecteur et invariance du rang par multiplication par une matrice inversible.

On applique 1. avec $a = p$, non nulle, et

$$\alpha = \sum p_i = 1.$$

D'après 1. f), $M(p)$ est la matrice d'un projecteur d'image $\text{Vect}(p)$ et de noyau l'hyperplan

$$\{x ; \sum x_i = 0\}.$$

Alors $(I_k - M(p))^2 = I_k - 2M(p) + M(p)^2 = I_k - M(p)$: $I_k - M(p)$ est aussi un projecteur, dont l'image est $\text{Ker } M(p)$ (si $M(p)x = 0$, $(I_k - M(p))x = x$; et $M(p)(I_k - M(p)) = 0$). Donc $\text{rg}(I_k - M(p)) = \dim \text{Ker } M(p) = k - 1$.

Si tous les p_i sont non nuls, $\text{Diag}(p)$ est inversible ; multiplier une matrice à droite par une matrice inversible ne change pas son rang.

Conclusion

si $p_1, \dots, p_k$ sont tous non nuls, $\text{rg}(\mathcal{V}(X)) = \text{rg}(I_k - M(p)) = k - 1$.

c) Méthode : changement de base par une matrice de permutation.

Soit Q la matrice de $\mathcal{M}_k(\mathbb{R})$ dont la j -ème colonne est $e_{\sigma(j)}$, c'est-à-dire $Q_{i,j} = 1$ si $i = \sigma(j)$ et $Q_{i,j} = 0$ sinon. Ses colonnes forment une base orthonormale (la base canonique réordonnée), donc ${}^tQQ = I_k$ et $Q^{-1} = {}^tQ$.

Calculs :

la j -ème composante de tQp est

$$\sum_i Q_{i,j} p_i = p_{\sigma(j)},$$

donc ${}^tQp = p_\sigma$ et ${}^t p Q = {}^t p_\sigma$;

le coefficient (j, l) de ${}^tQ \text{Diag}(p)Q$ est

$$\sum_i Q_{i,j} p_i Q_{i,l};$$

il vaut $p_{\sigma(j)}$ si $l = j$ et 0 sinon (car $\sigma(j) \neq \sigma(l)$), donc ${}^tQ \text{Diag}(p)Q = \text{Diag}(p_\sigma)$.

Avec 2. d) :

$${}^tQ \mathcal{V}(X)Q = {}^tQ \text{Diag}(p)Q - ({}^tQ p)({}^t p Q) = \text{Diag}(p_\sigma) - p_\sigma {}^t p_\sigma.$$

Le calcul de 3. a), valable pour toute colonne, donne

$$\text{Diag}(p_\sigma) - p_\sigma {}^t p_\sigma = (I_k - p_\sigma {}^t u) \text{Diag}(p_\sigma).$$

$$\text{Ainsi } \mathcal{V}(X) = Q (I_k - p_\sigma {}^t u) \text{Diag}(p_\sigma) Q^{-1}.$$

Conclusion

$\mathcal{V}(X)$ est semblable à $(I_k - p_\sigma {}^t u) \text{Diag}(p_\sigma)$.

d) Méthode : se ramener par c) au cas où les composantes non nulles sont les premières, puis appliquer b) à un bloc.

Notons m le nombre d'indices i tels que $p_i \neq 0$; comme

$$\sum p_i = 1,$$

$m \geq 1$. Choisissons une permutation σ telle que $p_{\sigma(1)}, \dots, p_{\sigma(m)}$ soient non nuls et $p_{\sigma(m+1)}, \dots, p_{\sigma(k)}$ nuls. Notons p' la colonne de $\mathcal{M}_{m,1}(\mathbb{R})$ de composantes $p_{\sigma(1)}, \dots, p_{\sigma(m)}$: elles sont strictement positives et de somme 1.

La matrice $A = \text{Diag}(p_\sigma) - p_\sigma {}^t p_\sigma$ a pour coefficient (j, l) le réel $p_{\sigma(j)} \delta_{j,l} - p_{\sigma(j)} p_{\sigma(l)}$, nul

dès que $j > m$ ou $l > m$. Donc $A = \begin{pmatrix} A' & 0 \\ 0 & 0 \end{pmatrix}$ avec $A' = \text{Diag}(p') - p' {}^t p' \in \mathcal{M}_m(\mathbb{R})$, et

$$\text{rg}(A) = \text{rg}(A').$$

Si $m \geq 2$: le calcul de b) (qui n'utilise que

$$\sum p'_i = 1$$

et $p'_i \neq 0$, avec m à la place de k) donne $\text{rg}(A') = m - 1$.

Si $m = 1$: $p' = (1)$ et $A' = 1 - 1 = 0$, de rang $0 = m - 1$. Dans ce cas X est presque sûrement constant.

Deux matrices semblables ont le même rang, donc $\text{rg}(\mathcal{V}(X)) = \text{rg}(A)$ d'après c).

Conclusion

le rang de $\mathcal{V}(X)$ est $m - 1$, où m est le nombre d'indices i tels que $p_i \neq 0$.

Partie II. Tirages avec remise dans une population stratifiée

Dans cette partie, tous les p_i sont strictement positifs, donc aussi strictement inférieurs à 1 (car $k \geq 2$ et

$$p_i = 1 - \sum_{j \neq i} p_j$$

). D'après 2. a), pour tout n , $X^{(n)}$ prend presque sûrement ses valeurs dans $\{e_1, \dots, e_k\}$. Les événements portant sur des tirages différents sont indépendants : par exemple, pour des parties $B_1, \dots, B_n$ de $\mathbb{R}^k$, en prenant pour φ_j l'indicatrice de B_j , on obtient

$$P \left(\bigcap_{j=1}^n [X^{(j)} \in B_j] \right) = \prod_{j=1}^n P([X^{(j)} \in B_j])$$

(une fonction φ_j constante égale à 1 permet de ne pas imposer de condition au tirage j).

4. a) Méthode : double inclusion.

Chaque $S_i^{(n)}$ est une somme de n variables à valeurs presque sûrement dans $\{0, 1\}$, donc $S_i^{(n)} \in \llbracket 0, n \rrbracket$ presque sûrement. De plus

$$\sum_{i=1}^k S_i^{(n)} = \sum_{j=1}^n \sum_{i=1}^k X_i^{(j)} = n$$

presque sûrement, car chaque $X^{(j)}$ est presque sûrement un e_l , dont la somme des composantes vaut 1. Donc si $P([S^{(n)} = s]) > 0$, alors s a des composantes entières naturelles de somme n .

Réciproquement, soit s une colonne à composantes $s_1, \dots, s_k$ dans $\mathbb{N}$ de somme n . Considérons l'événement C : « les s_1 premiers tirages donnent la catégorie 1, les s_2 suivants la catégorie 2, ..., les s_k derniers la catégorie k ». C'est une intersection $\bigcap_{j=1}^n [X^{(j)} = e_{c_j}]$ où c_j est la catégorie imposée au tirage j ; par indépendance,

$$P(C) = \prod_{j=1}^n p_{c_j} = \prod_{i=1}^k p_i^{s_i} > 0.$$

Comme $C \subset [S^{(n)} = s]$, on a $P([S^{(n)} = s]) > 0$.

Conclusion

$$N_n = \left\{ s \in \mathcal{M}_{k,1}(\mathbb{R}) ; \forall i, s_i \in \mathbb{N} \text{ et } \sum_{i=1}^k s_i = n \right\}.$$

b) Méthode : sommes de variables de Bernoulli indépendantes, puis recherche d'un contre-exemple à l'indépendance.

Lois. D'après 2. a), les variables $X_1^{(j)}$ ($1 \leq j \leq n$) suivent la loi $\mathcal{B}(p_1)$; elles sont mutuellement indépendantes (prendre $\varphi_j(x) = x_1$). Donc $S_1^{(n)} \hookrightarrow \mathcal{B}(n, p_1)$. De même, d'après 2. b), les variables $X_1^{(j)} + X_2^{(j)}$ suivent la loi $\mathcal{B}(p_1 + p_2)$ et sont mutuellement indépendantes (prendre $\varphi_j(x) = x_1 + x_2$), donc $S_1^{(n)} + S_2^{(n)} \hookrightarrow \mathcal{B}(n, p_1 + p_2)$.

Indépendance. Cas $k = 2$: $p_1 + p_2 = 1$ et $S_1^{(n)} + S_2^{(n)} = n$ presque sûrement (question a)). Une variable presque sûrement constante est indépendante de toute variable : pour tous x, y , $P([S_1^{(n)} = x] \cap [S_1^{(n)} + S_2^{(n)} = y])$ vaut $P([S_1^{(n)} = x])$ si $y = n$ et 0 sinon, ce qui est bien $P([S_1^{(n)} = x])P([S_1^{(n)} + S_2^{(n)} = y])$.

Cas $k \geq 3$: $p_3 > 0$, donc $0 < 1 - p_1 - p_2 < 1$. On a $P([S_1^{(n)} = n]) = p_1^n > 0$ et $P([S_1^{(n)} + S_2^{(n)} = 0]) = (1 - p_1 - p_2)^n > 0$, alors que $[S_1^{(n)} = n] \cap [S_1^{(n)} + S_2^{(n)} = 0]$ est impossible ($S_2^{(n)} \geq 0$ presque sûrement et $n \geq 1$). Le produit des probabilités est non nul alors que la probabilité de l'intersection est nulle.

Conclusion

$S_1^{(n)} \hookrightarrow \mathcal{B}(n, p_1)$ et $S_1^{(n)} + S_2^{(n)} \hookrightarrow \mathcal{B}(n, p_1 + p_2)$; ces deux variables sont indépendantes si $k = 2$ (la seconde est alors constante égale à n) et ne sont pas indépendantes si $k \geq 3$.

c) Méthode : bilinéarité de la covariance et indépendance des tirages.

Les $S_i^{(n)}$ admettent une variance (lois binomiales). Pour $(i, l) \in \llbracket 1, k \rrbracket^2$:

$$\text{Cov}(S_i^{(n)}, S_l^{(n)}) = \sum_{j=1}^n \sum_{j'=1}^n \text{Cov}(X_i^{(j)}, X_l^{(j')}).$$

Si $j \neq j'$, $X_i^{(j)}$ et $X_l^{(j')}$ sont indépendantes (prendre $\varphi_j(x) = x_i$, $\varphi_{j'}(x) = x_l$ et les autres fonctions constantes), donc leur covariance est nulle. Il reste

$$\sum_{j=1}^n \text{Cov}(X_i^{(j)}, X_l^{(j)}).$$

Chaque $X^{(j)}$ suit la loi $\mathcal{B}_k(p)$, donc d'après 2. d) $\text{Cov}(X_i^{(j)}, X_l^{(j)})$ est le coefficient (i, l) de $\mathcal{V}(X^{(1)})$, qui ne dépend que de p . Ce calcul vaut aussi pour $i = l$, où la covariance est une variance.

Conclusion

$$\mathcal{V}(S^{(n)}) = n \mathcal{V}(X^{(1)}) = n (\text{Diag}(p) - p^t p).$$

5. a) Méthode : comparaison de séries à termes positifs.

Notons $W(\Omega) = \{w_m\}$ l'ensemble au plus dénombrable des valeurs de W . Comme W admet une variance, W^2 admet une espérance, c'est-à-dire que la série

$$\sum_m w_m^2 P([W = w_m])$$

converge (somme finie si $W(\Omega)$ est fini). Pour tout m :

$$0 \leq w_m^2 P_H([W = w_m]) = w_m^2 \frac{P([W = w_m] \cap H)}{P(H)} \leq \frac{1}{P(H)} w_m^2 P([W = w_m]).$$

Par comparaison de séries à termes positifs,

$$\sum_m w_m^2 P_H([W = w_m])$$

converge. De même, avec $|w_m| \leq 1 + w_m^2$, W admet une espérance pour P_H . Le même raisonnement vaut pour $P_{\bar{H}}$, puisque $0 < P(\bar{H}) < 1$.

Conclusion

$$E(W^2|H) \text{ existe (ainsi que } E(W|H), E(W|\bar{H}) \text{ et } E(W^2|\bar{H})).$$

b) Méthode : formule de l'espérance totale, puis minoration de $E(W^2|\bar{H})$.

Notons $h = P(H) \in]0, 1[$, $m_1 = E(W|H)$ et $m_2 = E(W|\bar{H})$. La formule de l'espérance totale pour le système complet $(H, \bar{H})$, appliquée à W et à W^2 , donne :

$$E(W) = hm_1 + (1 - h)m_2 \text{ et } E(W^2) = hE(W^2|H) + (1 - h)E(W^2|\bar{H}).$$

La variance de W pour la probabilité $P_{\overline{H}}$ est positive : $E(W^2|\overline{H}) \geq m_2^2$. Donc

$$\begin{aligned} V(W) - hV(W|H) &= hE(W^2|H) \\ &+ (1-h)E(W^2|\overline{H}) - (hm_1 + (1-h)m_2)^2 - hE(W^2|H) + hm_1^2 \\ &\geq hm_1^2 + (1-h)m_2^2 - (hm_1 + (1-h)m_2)^2. \end{aligned}$$

En développant :

$$\begin{aligned} hm_1^2 + (1-h)m_2^2 - h^2m_1^2 - 2h(1-h)m_1m_2 - (1-h)^2m_2^2 &= h(1-h)m_1^2 + h(1-h)m_2^2 \\ - 2h(1-h)m_1m_2 &= h(1-h)(m_1 - m_2)^2 \geq 0. \end{aligned}$$

Conclusion

$V(W) \geq P(H) V(W|H)$ (plus précisément

$$V(W) - P(H)V(W|H) \geq P(H)P(\overline{H}) (E(W|H) - E(W|\overline{H}))^2$$

).

6. a) Méthode : continuité monotone décroissante.

Pour $n \in \mathbb{N}^*$, $[T_i > n] = \bigcap_{j=1}^n [X_i^{(j)} = 0]$, donc par indépendance $P([T_i > n]) = (1 - p_i)^n$. L'événement « T_i est infini » est $\bigcap_{n \geq 1} [T_i > n]$, intersection d'une suite décroissante d'événements. Par continuité monotone, sa probabilité est $\lim_{n \rightarrow +\infty} (1 - p_i)^n = 0$, car $0 \leq 1 - p_i < 1$.

Pour $n \geq 1$:

$$P([T_i = n]) = P([T_i > n-1]) - P([T_i > n]) = (1 - p_i)^{n-1} - (1 - p_i)^n = (1 - p_i)^{n-1} p_i$$

(avec $P([T_i > 0]) = 1$).

Conclusion

T_i est presque sûrement fini et $T_i \hookrightarrow \mathcal{G}(p_i)$; en particulier $E(T_i) = \frac{1}{p_i}$ et $V(T_i) = \frac{1 - p_i}{p_i^2}$.

b) Méthode : traduire H_k en termes des tirages, puis utiliser l'indépendance des tirages.

Posons $G_k = \bigcap_{j=1}^{k-1} [X^{(j)} = e_j]$ («le tirage j donne la catégorie j pour $j = 1, \dots, k-1$ »). Sur G_k , pour $i \leq k-1$, le tirage i donne la catégorie i et les tirages précédents donnent d'autres catégories, donc $T_i = i : G_k \subset H_k$. Réciproquement, sur H_k , $X_i^{(i)} = 1$ pour $i \leq k-1$, donc $X^{(i)} = e_i$ hors d'un événement négligeable (question 2. a)). Ainsi H_k et G_k diffèrent d'un événement négligeable, et par indépendance des tirages :

$$P(H_k) = P(G_k) = \prod_{i=1}^{k-1} p_i.$$

Loi conditionnelle. Sur G_k , aucun des tirages $1, \dots, k-1$ ne donne la catégorie k . Pour $m \in \mathbb{N}^*$, l'événement $H_k \cap [T_k - (k-1) = m]$ coïncide donc, à un événement négligeable près, avec

$$G_k \cap \bigcap_{j=k}^{k+m-2} [X_k^{(j)} = 0] \cap [X_k^{(k+m-1)} = 1],$$

dont la probabilité vaut, par indépendance des tirages, $P(G_k)(1 - p_k)^{m-1}p_k$. En divisant par $P(H_k) = P(G_k) > 0$:

$$P_{H_k}([T_k - (k-1) = m]) = (1 - p_k)^{m-1}p_k.$$

Conséquences. Pour la probabilité P_{H_k} , $T_k = (k-1) + G$ où G suit la loi $\mathcal{G}(p_k)$. Par linéarité de l'espérance et invariance de la variance par translation (pour la probabilité P_{H_k}) :

$$E(T_k | H_k) = k-1 + \frac{1}{p_k} \text{ et } V(T_k | H_k) = \frac{1 - p_k}{p_k^2}.$$

Conclusion

$$P(H_k) = \prod_{i=1}^{k-1} p_i;$$

sachant H_k , $T_k - (k - 1)$ suit la loi géométrique $\mathcal{G}(p_k)$; $E(T_k|H_k) = k - 1 + \frac{1}{p_k}$ et $V(T_k|H_k) = \frac{1 - p_k}{p_k^2}$.

c) Méthode : appliquer 5. b) à

$$W = \sum_{i=1}^k v_i T_i$$

et $H = H_k$.

Hypothèses de 5. b). W est discrète (car T prend un nombre au plus dénombrable de valeurs) et admet une variance, comme combinaison linéaire de variables admettant une variance (lois géométriques). D'autre part

$$P(H_k) = \prod_{i=1}^{k-1} p_i > 0$$

et $P(H_k) \leq p_1 < 1$.

Calcul de $V(W|H_k)$. Sur H_k , $T_i = i$ pour $i \leq k - 1$, donc $W = c + v_k T_k$ avec

$$c = \sum_{i=1}^{k-1} i v_i,$$

et cette égalité a lieu P_{H_k} -presque sûrement. Pour la probabilité P_{H_k} , la variance vérifie les règles usuelles : $V(W|H_k) = v_k^2 V(T_k|H_k) = v_k^2 \frac{1 - p_k}{p_k^2}$.

D'après 5. b), $V(W) \geq P(H_k) V(W|H_k)$.

Conclusion

$$V\left(\sum_{i=1}^k v_i T_i\right) \geq \frac{v_k^2(1-p_k)}{p_k^2} \times \prod_{i=1}^{k-1} p_i.$$

d) Méthode : même raisonnement après renumérotation des catégories autres que j .

Soit $j \in \llbracket 1, k \rrbracket$ et $i_1 < i_2 < \dots < i_{k-1}$ les éléments de $\llbracket 1, k \rrbracket \setminus \{j\}$. On pose $H = \bigcap_{l=1}^{k-1} [T_{i_l} = l] : \text{«le tirage } l \text{ donne la catégorie } i_l \text{ pour } l = 1, \dots, k-1 \text{»}$. Exactement comme en b), à un événement négligeable près, $H = \bigcap_{l=1}^{k-1} [X^{(l)} = e_{i_l}]$, donc

$$P(H) = \prod_{i \neq j} p_i,$$

qui est dans $]0, 1[$; aucun des tirages $1, \dots, k-1$ ne donne la catégorie j sur H , donc sachant H , $T_j - (k-1)$ suit la loi $\mathcal{G}(p_j)$ et $V(T_j|H) = \frac{1-p_j}{p_j^2}$.

Sur H ,

$$W = \sum_{l=1}^{k-1} l v_{i_l} + v_j T_j,$$

donc $V(W|H) = v_j^2 \frac{1-p_j}{p_j^2}$, et 5. b) donne $V(W) \geq P(H)V(W|H)$.

Conclusion

pour tout $j \in \llbracket 1, k \rrbracket$,

$$V\left(\sum_{i=1}^k v_i T_i\right) \geq \frac{v_j^2(1-p_j)}{p_j^2} \times \prod_{i \neq j} p_i.$$

Partie III. Support et rang stochastiques d'un vecteur aléatoire

Remarque préliminaire. Y étant discret, pour toute partie B de $\mathcal{M}_{k,1}(\mathbb{R})$, $[Y - \mathcal{E}(Y) \in B]$ est une réunion au plus dénombrable d'événements $[Y = y]$, donc un événement; et toute fonction de Y est une variable aléatoire discrète.

7. a) Méthode : propriété de la borne inférieure dans $\mathbb{N}$.

$P([Y - \mathcal{E}(Y) \in \mathcal{M}_{k,1}(\mathbb{R})]) = P(\Omega) = 1$, donc $\mathcal{M}_{k,1}(\mathbb{R}) \in \mathcal{S}(Y)$ et $\mathcal{S}(Y)$ n'est pas vide. L'ensemble des dimensions des éléments de $\mathcal{S}(Y)$ est donc une partie non vide de $\mathbb{N}$ (incluse dans $\llbracket 0, k \rrbracket$); toute partie non vide de $\mathbb{N}$ admet un plus petit élément.

Conclusion

$R_s(Y)$ existe et $0 \leq R_s(Y) \leq k$.

b) Méthode : le seul sous-espace de dimension 0 est $\{0\}$.

$R_s(Y) = 0$ si et seulement si $\{0\} \in \mathcal{S}(Y)$, c'est-à-dire si et seulement si $P([Y = \mathcal{E}(Y)]) = 1$. Cette condition équivaut à : pour tout i , $P([Y_i = E(Y_i)]) = 1$ (une intersection finie d'événements presque sûrs est presque sûre, et réciproquement $[Y = \mathcal{E}(Y)] \subset [Y_i = E(Y_i)]$, c'est-à-dire à : pour tout i , $V(Y_i) = 0$. Dans ce cas toutes les covariances sont nulles aussi (covariance avec une variable presque sûrement constante).

Conclusion

$R_s(Y) = 0$ si et seulement si Y est presque sûrement constant (égal à $\mathcal{E}(Y)$), ce qui équivaut à $V(Y_1) = \dots = V(Y_k) = 0$, ou encore à $\mathcal{V}(Y) = 0$.

c) Méthode : passage au complémentaire.

$F_1 \cap F_2$ est un sous-espace vectoriel de $\mathcal{M}_{k,1}(\mathbb{R})$ et

$$[Y - \mathcal{E}(Y) \in F_1 \cap F_2] = [Y - \mathcal{E}(Y) \in F_1] \cap [Y - \mathcal{E}(Y) \in F_2].$$

Son contraire est la réunion des contraires, de probabilités nulles, donc

$$P([Y - \mathcal{E}(Y) \in F_1 \cap F_2]) \geq 1 - 0 - 0 = 1.$$

Conclusion

$$F_1 \cap F_2 \in \mathcal{S}(Y).$$

d) Méthode : existence par définition du minimum, unicité par c).

Existence : par définition de $R_s(Y)$ comme plus petit élément, il existe $F \in \mathcal{S}(Y)$ de dimension $R_s(Y)$.

Unicité : soit F et F' deux éléments de $\mathcal{S}(Y)$ de dimension $R_s(Y)$. D'après c), $F \cap F' \in \mathcal{S}(Y)$, donc $\dim(F \cap F') \geq R_s(Y)$ par minimalité; comme $F \cap F' \subset F$, $\dim(F \cap F') \leq \dim F = R_s(Y)$. Ainsi $F \cap F'$ est un sous-espace de F de même dimension que F , donc $F \cap F' = F$. De même $F \cap F' = F'$, d'où $F = F'$.

Conclusion

il existe un unique support vectoriel de dimension $R_s(Y)$. (Le même argument montre qu'il est inclus dans tout support vectoriel de Y .)

8. Dans cette question, u désigne la colonne quelconque de l'énoncé.

a) Méthode : bilinéarité de la covariance.

Chaque Y_i admet une variance; une combinaison linéaire de variables discrètes admettant une variance admet une variance (on a

$$\left(\sum u_i Y_i\right)^2 \leq k \sum u_i^2 Y_i^2,$$

ou on utilise l'existence des covariances). Par bilinéarité :

$$V \left(\sum_{i=1}^k u_i Y_i \right) = \text{Cov} \left(\sum_i u_i Y_i, \sum_j u_j Y_j \right) = \sum_{i=1}^k \sum_{j=1}^k u_i u_j \text{Cov}(Y_i, Y_j),$$

en notant $\text{Cov}(Y_i, Y_i) = V(Y_i)$. On reconnaît ${}^t u \mathcal{V}(Y) u$.

Conclusion

$$\sum_{i=1}^k u_i Y_i$$

admet une variance et

$$V \left(\sum_{i=1}^k u_i Y_i \right) = {}^t u \mathcal{V}(Y) u.$$

b) Méthode : théorème spectral pour l'existence, dimensions des sous-espaces propres pour l'unicité.

Existence. $\mathcal{V}(Y)$ est symétrique réelle. D'après le théorème spectral, il existe une base orthonormale $(\varepsilon_1, \dots, \varepsilon_k)$ de vecteurs propres : $\mathcal{V}(Y)\varepsilon_i = \mu_i \varepsilon_i$. Pour chaque i , d'après a) appliqué à $u = \varepsilon_i$:

$$\mu_i = \mu_i \|\varepsilon_i\|^2 = {}^t \varepsilon_i \mathcal{V}(Y) \varepsilon_i = V \left(\sum_j (\varepsilon_i)_j Y_j \right) \geq 0.$$

Quitte à renuméroter les ε_i , on peut supposer $\mu_1 \geq \mu_2 \geq \dots \geq \mu_k \geq 0$. En notant P la matrice orthogonale de colonnes $\varepsilon_1, \dots, \varepsilon_k$ et $\lambda = (\mu_1, \dots, \mu_k)$, on a $\mathcal{V}(Y) = P \text{Diag}(\lambda) {}^t P = P \text{Diag}(\lambda) P^{-1}$.

Unicité. Si $\mathcal{V}(Y)$ est semblable à $\text{Diag}(\lambda)$ et à $\text{Diag}(\lambda')$, avec λ et λ' rangés dans l'ordre décroissant, alors pour tout réel μ , la dimension de $\text{Ker}(\mathcal{V}(Y) - \mu I_k)$ est égale à celle de

$\text{Ker}(\text{Diag}(\lambda) - \mu I_k)$, c'est-à-dire au nombre d'indices i tels que $\lambda_i = \mu$ (deux matrices semblables représentent le même endomorphisme dans deux bases). Ainsi chaque réel figure le même nombre de fois dans λ et dans λ' ; ces deux listes étant rangées dans l'ordre décroissant, elles sont égales.

Conclusion

il existe un unique $\lambda = (\lambda_1, \dots, \lambda_k)$ avec $\lambda_1 \geq \dots \geq \lambda_k \geq 0$ tel que $\mathcal{V}(Y)$ soit semblable à $\text{Diag}(\lambda)$: ce sont les valeurs propres de $\mathcal{V}(Y)$ comptées avec la dimension de leur sous-espace propre. Dans toute la suite, $(\varepsilon_1, \dots, \varepsilon_k)$ désigne une base orthonormale de vecteurs propres telle que $\mathcal{V}(Y)\varepsilon_i = \lambda_i \varepsilon_i$.

c) Méthode : calcul dans une base orthonormale de vecteurs propres.

Par linéarité de l'espérance,

$$E(\|Y - \mathcal{E}(Y)\|^2) = \sum_{i=1}^k E((Y_i - E(Y_i))^2) = \sum_{i=1}^k V(Y_i),$$

somme des coefficients diagonaux de $\mathcal{V}(Y)$.

Avec $\mathcal{V}(Y) = P \text{Diag}(\lambda) {}^t P$, le coefficient diagonal d'indice i vaut

$$\sum_{j=1}^k P_{i,j}^2 \lambda_j.$$

Donc

$$\sum_{i=1}^k V(Y_i) = \sum_{j=1}^k \lambda_j \sum_{i=1}^k P_{i,j}^2 = \sum_{j=1}^k \lambda_j \|\varepsilon_j\|^2 = \sum_{j=1}^k \lambda_j.$$

Conclusion

$$E(\|Y - \mathcal{E}(Y)\|^2) = \sum_{i=1}^k \lambda_i.$$

9. a) Méthode : projection orthogonale sur F et théorème de Pythagore.

Fixons ω et posons $z = Y(\omega) - \mathcal{E}(Y) \in \mathcal{M}_{k,1}(\mathbb{R})$. L'ensemble $\{\|z - x\|^2; x \in F\}$ est une partie non vide de $\mathbb{R}$ minorée par 0 : sa borne inférieure $Q_F(\omega)$ existe.

Soit

$$\pi(z) = \sum_{j=1}^q \langle z, f^{(j)} \rangle f^{(j)}$$

le projeté orthogonal de z sur F ; $z - \pi(z)$ est orthogonal à F . Pour $x \in F$, $\pi(z) - x \in F$, donc par Pythagore

$$\|z - x\|^2 = \|z - \pi(z)\|^2 + \|\pi(z) - x\|^2 \geq \|z - \pi(z)\|^2,$$

avec égalité pour $x = \pi(z)$. Ainsi $Q_F(\omega) = \|z - \pi(z)\|^2$ (la borne inférieure est un minimum). Enfin, toujours par Pythagore et parce que $(f^{(j)})$ est orthonormale :

$$\|z\|^2 = \|z - \pi(z)\|^2 + \|\pi(z)\|^2 = Q_F(\omega) + \sum_{j=1}^q \langle z, f^{(j)} \rangle^2.$$

Conclusion

$Q_F(\omega)$ existe, c'est le carré de la distance de $Y(\omega) - \mathcal{E}(Y)$ à F , et

$$\|Y(\omega) - \mathcal{E}(Y)\|^2 = Q_F(\omega) + \sum_{j=1}^q \langle Y(\omega) - \mathcal{E}(Y), f^{(j)} \rangle^2.$$

b) Méthode : linéarité de l'espérance et question 8.

Q_F est une fonction de Y , donc une variable aléatoire discrète, positive, majorée par $\|Y - \mathcal{E}(Y)\|^2$ qui admet une espérance (8. c)) : Q_F admet une espérance. Pour $f \in \mathcal{M}_{k,1}(\mathbb{R})$ de composantes $f_1, \dots, f_k$, posons

$$Z_f = \sum_{i=1}^k f_i Y_i;$$

alors

$$E(Z_f) = \sum f_i E(Y_i)$$

et

$$\langle Y - \mathcal{E}(Y), f \rangle = \sum_{i=1}^k f_i (Y_i - E(Y_i)) = Z_f - E(Z_f),$$

donc $E(\langle Y - \mathcal{E}(Y), f \rangle^2) = V(Z_f) = {}^t f \mathcal{V}(Y) f$ d'après 8. a). En prenant l'espérance dans l'égalité de a) et avec 8. c) :

$$\sum_{i=1}^k \lambda_i = E(Q_F) + \sum_{j=1}^q {}^t f^{(j)} \mathcal{V}(Y) f^{(j)}.$$

Conclusion

$$E(Q_F) = \sum_{i=1}^k \lambda_i - \sum_{j=1}^q {}^t f^{(j)} \mathcal{V}(Y) f^{(j)}.$$

c) Méthode : cas $q = k$.

Si $F = \mathcal{M}_{k,1}(\mathbb{R})$, alors $Y(\omega) - \mathcal{E}(Y) \in F$ pour tout ω , donc $Q_F = 0$ et $E(Q_F) = 0$.

Conclusion

pour toute base orthonormale $(f^{(1)}, \dots, f^{(k)})$ de $\mathcal{M}_{k,1}(\mathbb{R})$,

$$\sum_{j=1}^k {}^t f^{(j)} \mathcal{V}(Y) f^{(j)} = \sum_{i=1}^k \lambda_i :$$

cette somme ne dépend pas de la base orthonormale choisie (c'est la somme des coefficients diagonaux de la matrice de l'endomorphisme associé dans cette base).

10. a) Méthode : décomposition dans la base orthonormale de vecteurs propres.

Écrivons

$$f = \sum_{i=1}^k c_i \varepsilon_i;$$

comme la base est orthonormale,

$$\sum c_i^2 = \|f\|^2 = 1.$$

Alors

$$\mathcal{V}(Y)f = \sum_i c_i \lambda_i \varepsilon_i$$

et

$${}^t f \mathcal{V}(Y) f = \langle f, \mathcal{V}(Y) f \rangle = \sum_{i=1}^k \lambda_i c_i^2 \leq \lambda_1 \sum_{i=1}^k c_i^2 = \lambda_1.$$

Conclusion

pour tout f de norme 1, ${}^t f \mathcal{V}(Y) f \leq \lambda_1$, avec égalité pour $f = \varepsilon_1$.

b) Méthode : 9. b) avec $q = 1$, puis 10. a).

Une droite vectorielle F s'écrit $F = \text{Vect}(f)$ avec $\|f\| = 1$ (f est alors une base orthonormale de F). D'après 9. b),

$$E(Q_F) = \sum_{i=1}^k \lambda_i - {}^t f \mathcal{V}(Y) f \geq \sum_{i=1}^k \lambda_i - \lambda_1,$$

avec égalité pour $F = \text{Vect}(\varepsilon_1)$.

Conclusion

lorsque F décrit l'ensemble des droites vectorielles, la borne inférieure de $E(Q_F)$ est $\lambda_2 + \lambda_3 + \dots + \lambda_k$; c'est un minimum, atteint pour la droite engendrée par un vecteur propre associé à la plus grande valeur propre λ_1 (et exactement pour les droites $\text{Vect}(f)$, $\|f\| = 1$, telles que ${}^t f \mathcal{V}(Y) f = \lambda_1$).

c) Méthode : écriture de $\mathcal{V}(Y)$ à l'aide de 2. d), puis recherche directe des sous-espaces propres.

Avec $p_i = \frac{1}{k}$, la question 2. d) donne $\mathcal{V}(Y) = \frac{1}{k} I_k - \frac{1}{k^2} u {}^t u$, où u est la colonne dont tous les coefficients valent 1. Pour $x \in \mathcal{M}_{k,1}(\mathbb{R})$,

$$\mathcal{V}(Y)x = \frac{1}{k}x - \frac{1}{k^2} \left(\sum_i x_i \right) u.$$

Si

$$\sum_i x_i = 0$$

(hyperplan $u^\perp$, de dimension $k - 1$), $\mathcal{V}(Y)x = \frac{1}{k}x$. Et $\mathcal{V}(Y)u = \frac{1}{k}u - \frac{k}{k^2}u = 0$. Les sous-espaces $u^\perp$ et $\text{Vect}(u)$, de dimensions $k - 1$ et 1 , sont contenus dans des sous-espaces propres et leurs dimensions ont pour somme k .

Valeurs propres : $\frac{1}{k}$, de sous-espace propre l'hyperplan

$$\{x ; \sum x_i = 0\}$$

(dimension $k - 1$), et 0 , de sous-espace propre $\text{Vect}(u)$. Donc $\lambda_1 = \dots = \lambda_{k-1} = \frac{1}{k}$ et $\lambda_k = 0$ (ce qui est cohérent avec 3. b) : rang $k - 1$).

Borne : d'après b), elle vaut $\lambda_2 + \dots + \lambda_k = (k - 2)\frac{1}{k} = \frac{k - 2}{k}$.

Droites qui l'atteignent : pour $\|f\| = 1$,

$${}^t f \mathcal{V}(Y) f = \frac{1}{k} \|f\|^2 - \frac{1}{k^2} \left(\sum_i f_i \right)^2 = \frac{1}{k} - \frac{1}{k^2} \left(\sum_i f_i \right)^2.$$

Ainsi

$$E(Q_{\text{Vect}(f)}) = \frac{k-1}{k} - \frac{1}{k} + \frac{1}{k^2} \left(\sum_i f_i \right)^2 = \frac{k-2}{k} + \frac{1}{k^2} \left(\sum_i f_i \right)^2,$$

et la borne est atteinte si et seulement si

$$\sum_i f_i = 0.$$

Conclusion

les valeurs propres de $\mathcal{V}(Y)$ sont $\frac{1}{k}$ (sous-espace propre de dimension $k - 1$) et 0 (sous-espace propre $\text{Vect}(u)$) ; la borne inférieure de $E(Q_F)$ sur les droites vaut $\frac{k-2}{k}$ et elle est atteinte exactement pour les droites contenues dans l'hyperplan

$$\left\{ x ; \sum_{i=1}^k x_i = 0 \right\}$$

(pour $k = 2$: une seule droite, $\text{Vect}(e_1 - e_2)$, et la borne vaut 0).

11. Notations. D'après 8. b), $\mathcal{V}(Y)$ est semblable à $\text{Diag}(\lambda)$, donc son rang r est le nombre d'indices i tels que $\lambda_i \neq 0$: $\lambda_1 \geq \dots \geq \lambda_r > 0 = \lambda_{r+1} = \dots = \lambda_k$. Pour chaque valeur propre μ , le sous-espace propre associé contient les ε_i tels que $\lambda_i = \mu$, dont le nombre est sa dimension (8. b)) : il est donc engendré par ces ε_i . Par suite $F_0 = \text{Vect}(\varepsilon_1, \dots, \varepsilon_r)$, de dimension r , avec $(\varepsilon_1, \dots, \varepsilon_r)$ base orthonormale de F_0 .

a) Méthode : 9. b) avec la base orthonormale $(\varepsilon_1, \dots, \varepsilon_r)$, puis espérance nulle d'une variable positive.

Comme $1 \leq r \leq k$, 9. b) s'applique avec $q = r$:

$$E(Q_{F_0}) = \sum_{i=1}^k \lambda_i - \sum_{j=1}^r {}^t \varepsilon_j \mathcal{V}(Y) \varepsilon_j = \sum_{i=1}^k \lambda_i - \sum_{j=1}^r \lambda_j = \sum_{i=r+1}^k \lambda_i = 0.$$

Q_{F_0} est une variable discrète positive d'espérance nulle, donc $P([Q_{F_0} = 0]) = 1$. Or d'après 9. a), $Q_{F_0}(\omega)$ est le carré de la distance de $Y(\omega) - \mathcal{E}(Y)$ à F_0 , nul si et seulement si $Y(\omega) - \mathcal{E}(Y) \in F_0$. Donc $P([Y - \mathcal{E}(Y) \in F_0]) = 1$.

Conclusion

$E(Q_{F_0}) = 0$ et F_0 est un support vectoriel de Y .

b) Méthode : supplémentaire orthogonal de F dans F_0 .

F est un sous-espace strict de F_0 , donc $\dim F < r$. Le sous-espace $G = F_0 \cap F^\perp$ vérifie $F_0 = F \oplus G$ (pour $x \in F_0$, $x = \pi_F(x) + (x - \pi_F(x))$ avec $\pi_F(x) \in F \subset F_0$, donc $x - \pi_F(x) \in F_0 \cap F^\perp$; et $F \cap F^\perp = \{0\}$). Donc $\dim G = r - \dim F \geq 1$: G contient un vecteur non nul, qu'on divise par sa norme.

Conclusion

il existe $f^{(r)} \in F_0$, orthogonal à F , de norme 1.

c) Méthode : décomposition de $f^{(r)}$ sur $(\varepsilon_1, \dots, \varepsilon_r)$, puis 9. b) appliqué à F et à $F \oplus \text{Vect}(f^{(r)})$.

Écrivons

$$f^{(r)} = \sum_{i=1}^r c_i \varepsilon_i$$

avec

$$\sum c_i^2 = 1.$$

Comme en 10. a) :

$${}^t f^{(r)} \mathcal{V}(Y) f^{(r)} = \sum_{i=1}^r \lambda_i c_i^2 \geq \lambda_r \sum_{i=1}^r c_i^2 = \lambda_r > 0.$$

Notons $q = \dim F$.

Si $q \geq 1$: soit $(f^{(1)}, \dots, f^{(q)})$ une base orthonormale de F . Comme $f^{(r)}$ est unitaire et orthogonal à F , $(f^{(1)}, \dots, f^{(q)}, f^{(r)})$ est une base orthonormale de $F' = F \oplus \text{Vect}(f^{(r)})$, de dimension $q + 1 \leq r \leq k$. En appliquant 9. b) à F et à F' :

$$E(Q_F) = \sum_{i=1}^k \lambda_i - \sum_{j=1}^q {}^t f^{(j)} \mathcal{V}(Y) f^{(j)} = E(Q_{F'}) + {}^t f^{(r)} \mathcal{V}(Y) f^{(r)} \geq {}^t f^{(r)} \mathcal{V}(Y) f^{(r)} > 0,$$

car $Q_{F'} \geq 0$.

Si $q = 0$, c'est-à-dire $F = \{0\}$ (cas non couvert par la question 9, où $q \geq 1$) :

$Q_F = \|Y - \mathcal{E}(Y)\|^2$ et, d'après 8. c),

$$E(Q_F) = \sum_{i=1}^k \lambda_i \geq \lambda_1 > 0.$$

Conclusion

${}^t f^{(r)} \mathcal{V}(Y) f^{(r)} \geq \lambda_r > 0$ et $E(Q_F) > 0$, donc $E(Q_F) \neq 0$, pour tout sous-espace F strictement inclus dans F_0 .

d) Méthode : double inégalité, à l'aide de a), de 7. c) et de c).

D'après a), $F_0 \in \mathcal{S}(Y)$ et $\dim F_0 = r$, donc $R_s(Y) \leq r$.

Supposons $R_s(Y) < r$ et notons F_s le support stochastique de Y . D'après 7. c),

$G = F_s \cap F_0$ est un support vectoriel de Y ; il est inclus dans F_0 et

$\dim G \leq \dim F_s < r = \dim F_0$, donc $G \neq F_0$. Comme G est un support vectoriel,

$Y - \mathcal{E}(Y) \in G$ presque sûrement, donc $Q_G = 0$ presque sûrement et $E(Q_G) = 0$, ce qui contredit c). Donc $R_s(Y) \geq r$.

Conclusion

$R_s(Y) = r = \text{rg}(\mathcal{V}(Y))$, et le support stochastique de Y est F_0 (unique support de dimension r , d'après 7. d)).

12. a) Méthode : montrer que $\mathcal{V}(T)$ est inversible grâce à 6. d), puis appliquer 11. d).

T est un vecteur aléatoire discret dont les composantes suivent des lois géométriques, donc admettent une espérance et une variance : la partie III s'applique à $Y = T$.

Soit $v \in \mathcal{M}_{k,1}(\mathbb{R})$ tel que $\mathcal{V}(T)v = 0$. Alors, d'après 8. a),

$$V \left(\sum_i v_i T_i \right) = {}^t v \mathcal{V}(T) v = 0.$$

D'après 6. d), pour tout j :

$$0 \geq \frac{v_j^2(1-p_j)}{p_j^2} \times \prod_{i \neq j} p_i,$$

et comme $0 < p_j < 1$ et $p_i > 0$, on obtient $v_j = 0$. Donc $\text{Ker } \mathcal{V}(T) = \{0\}$ et le rang de $\mathcal{V}(T)$ est k , non nul. D'après 11. d), $R_s(T) = k$.

Conclusion

$R_s(T) = k$; aucun hyperplan ne contient presque sûrement $T - \mathcal{E}(T)$.

b) Question adaptée au programme actuel (rectification d'une erreur d'énoncé) :

l'original affirme l'égalité « pour tout $i \in \mathbb{N}^*$ ». Or, lorsque $k = 2$, l'événement

$[T_1 = i] \cap [T_2 > i]$ est de probabilité nulle pour $i \geq 2$ (calcul ci-dessous) et

l'espérance conditionnelle n'est pas définie. L'énoncé adapté restreint l'égalité aux i pour lesquels cet événement est de probabilité non nulle, c'est-à-dire à tout $i \in \mathbb{N}^*$ si $k \geq 3$ et à $i = 1$ si $k = 2$.

Méthode : traduction de l'événement en termes des tirages, puis loi conditionnelle de $T_2 - i$.

Notons $C_i = [T_1 = i] \cap [T_2 > i]$ et $\rho = 1 - p_1 - p_2 \in [0, 1[$ ($\rho = 0$ si et seulement si $k = 2$). Sur $[T_1 = i]$, le tirage i donne la catégorie 1, donc pas la catégorie 2; ainsi C_i est « aucun des tirages $1, \dots, i-1$ ne donne la catégorie 1 ou 2, et le tirage i donne la catégorie 1 ». À un événement négligeable près :

$$C_i = \bigcap_{j=1}^{i-1} [X_1^{(j)} + X_2^{(j)} = 0] \cap [X^{(i)} = e_1], \text{ donc } P(C_i) = \rho^{i-1} p_1 \text{ (avec } 0^0 = 1),$$

par indépendance des tirages et 2. b). Supposons $P(C_i) > 0$. Pour $m \in \mathbb{N}^*$, l'événement $C_i \cap [T_2 = i + m]$ s'écrit $C_i \cap \bigcap_{j=i+1}^{i+m-1} [X_2^{(j)} = 0] \cap [X_2^{(i+m)} = 1]$; par indépendance des tirages, sa probabilité est $P(C_i)(1 - p_2)^{m-1} p_2$. Donc, pour la probabilité $P_{C_i}, T_2 - i$ suit la loi $\mathcal{G}(p_2)$ et $E(T_2|C_i) = i + \frac{1}{p_2}$.

Sur $C_i, T_1 T_2 = i T_2$, donc par linéarité $E(T_1 T_2|C_i) = i E(T_2|C_i)$.

Conclusion

$$\text{pour tout } i \in \mathbb{N}^* \text{ tel que } P([T_1 = i] \cap [T_2 > i]) > 0, \quad E(T_1 T_2 | [T_1 = i] \cap [T_2 > i]) = i \left(i + \frac{1}{p_2} \right).$$

c) Méthode : formule de l'espérance totale sur le système des événements C_i et

$$C'_i = [T_2 = i] \cap [T_1 > i], \text{ puis séries géométriques dérivées.}$$

Existence. $0 \leq T_1 T_2 \leq \frac{1}{2}(T_1^2 + T_2^2)$ et T_1^2, T_2^2 admettent une espérance : $T_1 T_2$ (discrète) admet une espérance.

Système d'événements. Un même tirage ne peut donner à la fois les catégories 1 et 2, donc $[T_1 = T_2]$ est négligeable; avec 6. a), presque sûrement T_1 et T_2 sont finis et distincts. Les événements C_i et C'_i ($i \in \mathbb{N}^*$) sont deux à deux incompatibles et leur réunion est de probabilité 1 : sur $[T_1 < T_2]$ on est dans C_{T_1} , sur $[T_2 < T_1]$ dans C'_{T_2} . Par symétrie des rôles des catégories 1 et 2, $P(C'_i) = \rho^{i-1} p_2$ et, si $P(C'_i) > 0$,

$E(T_1 T_2 | C'_i) = i \left(i + \frac{1}{p_1} \right)$. On retient ceux de ces événements qui sont de probabilité non nulle (tous si $k \geq 3$; C_1 et C'_1 seulement si $k = 2$) et on adjoint à C'_1 l'événement négligeable restant, ce qui ne modifie ni sa probabilité ni les espérances conditionnelles : on obtient un système complet d'événements de probabilités non nulles.

Calcul. Les termes d'indice i tels que $P(C_i) = 0$ étant nuls avec la convention $0^0 = 1$, la formule de l'espérance totale donne dans tous les cas :

$$\begin{aligned} E(T_1 T_2) &= \sum_{i=1}^{+\infty} \rho^{i-1} p_1 i \left(i + \frac{1}{p_2} \right) + \sum_{i=1}^{+\infty} \rho^{i-1} p_2 i \left(i + \frac{1}{p_1} \right) = (p_1 + p_2) \sum_{i=1}^{+\infty} i^2 \rho^{i-1} \\ &+ \left(\frac{p_1}{p_2} + \frac{p_2}{p_1} \right) \sum_{i=1}^{+\infty} i \rho^{i-1}. \end{aligned}$$

Posons $s = p_1 + p_2 = 1 - \rho \in]0, 1]$. D'après les séries géométriques dérivées :

$$\sum_{i=1}^{+\infty} i \rho^{i-1} = \frac{1}{s^2}$$

et

$$\sum_{i=1}^{+\infty} i^2 \rho^{i-1} = \sum_{i=1}^{+\infty} i(i-1) \rho^{i-1} + \sum_{i=1}^{+\infty} i \rho^{i-1} = \frac{2\rho}{s^3} + \frac{1}{s^2} = \frac{1+\rho}{s^3} = \frac{2-s}{s^3}.$$

Avec $\frac{p_1}{p_2} + \frac{p_2}{p_1} = \frac{p_1^2 + p_2^2}{p_1 p_2} = \frac{s^2 - 2p_1 p_2}{p_1 p_2} :$

$$E(T_1 T_2) = \frac{2-s}{s^2} + \frac{s^2 - 2p_1 p_2}{p_1 p_2 s^2} = \frac{2-s}{s^2} + \frac{1}{p_1 p_2} - \frac{2}{s^2} = \frac{1}{p_1 p_2} - \frac{1}{s}.$$

Vérification pour $k = 2$ ($s = 1, \rho = 0$) : seuls les termes $i = 1$ subsistent et

$$E(T_1 T_2) = p_1 \left(1 + \frac{1}{p_2}\right) + p_2 \left(1 + \frac{1}{p_1}\right) = 1 + \frac{p_1^2 + p_2^2}{p_1 p_2} = \frac{1}{p_1 p_2} - 1,$$

ce qui est bien la formule.

Conclusion

$$E(T_1 T_2) = \frac{1}{p_1 p_2} - \frac{1}{p_1 + p_2}.$$

d) Méthode : reconnaître $\Pi = \mathcal{V}(T)$.

Les catégories jouant des rôles symétriques, 12. b) et 12. c) valent pour tout couple (i, j) d'indices distincts : $E(T_i T_j) = \frac{1}{p_i p_j} - \frac{1}{p_i + p_j}$. Donc, avec 6. a) :

$$\text{Cov}(T_i, T_j) = E(T_i T_j) - E(T_i)E(T_j) = \frac{1}{p_i p_j} - \frac{1}{p_i + p_j} - \frac{1}{p_i p_j} = -\frac{1}{p_i + p_j}$$

$$\text{et } V(T_i) = \frac{1 - p_i}{p_i^2}.$$

Ainsi $\Pi = \mathcal{V}(T)$. D'après 12. a), $\mathcal{V}(T)$ est de rang k (son noyau est nul).

Conclusion

$\Pi = \mathcal{V}(T)$ est inversible. (Plus précisément, d'après 6. d) et 8. a), ${}^t v \Pi v > 0$ pour tout $v \neq 0$.)

Relevé des adaptations

12. b) L'original énonce l'égalité $E(T_1 T_2 | [T_1 = i] \cap [T_2 > i]) = i \left(i + \frac{1}{p_2}\right)$ « pour tout $i \in \mathbb{N}^*$ ». Erreur d'énoncé : pour $k = 2$, $P([T_1 = i] \cap [T_2 > i]) = (1 - p_1 - p_2)^{i-1} p_1 = 0$ dès que $i \geq 2$, et l'espérance conditionnelle n'est pas définie. L'énoncé adapté porte

«pour tout $i \in \mathbb{N}^*$ tel que $P([T_1 = i] \cap [T_2 > i]) > 0$ ». Justification : rectifier l'erreur sans changer le contenu de la question (pour $k \geq 3$ elle est inchangée). Le corrigé de 12. c) en tient compte (les événements négligeables sont écartés du système complet, et le cas $k = 2$ est vérifié à part).

Les autres questions sont conformes au référentiel ECG mathématiques approfondies et conservées telles quelles ; les méthodes ont été choisies dans le programme (théorème spectral et dimensions des sous-espaces propres en 8. b), sans polynôme caractéristique ni déterminant ; formule de l'espérance totale dans le cadre discret en 5. b) et 12. c)).