

74.07B

SESSION 2007

Filière BCPST

MATHÉMATIQUES

Epreuve commune aux ENS de Paris, Lyon et Cachan

Durée : 4 heures

L'usage de calculatrice est interdit.

Ce sujet concerne la coexistence de m formes alléliques distinctes d'un même gène ($m \geq 2$) dans une population diploïde.

*Il comporte cinq parties de difficulté **croissante** qui ne sont pas complètement indépendantes : toutes les équations **numérotées** sont susceptibles d'être utilisées à un stade ultérieur du sujet. Toutefois, à condition de lire l'introduction de chacune des parties, les candidats pourront aborder celles-ci dans l'ordre qui leur conviendra.*

Il est recommandé de lire les énoncés attentivement et patiemment. Il est demandé de veiller au soin de la présentation, ainsi qu'à la rigueur et à la concision des raisonnements.

Notations

L'ensemble des nombres réels est noté \mathbb{R} et celui des nombres réels strictement positifs \mathbb{R}_+^* . L'espace vectoriel des matrices réelles à m lignes et k colonnes est noté $\mathfrak{M}_{m,k}(\mathbb{R})$.

On désigne par $\mathbf{1}$ le vecteur colonne dont toutes les composantes sont égales à 1. Ainsi la somme des composantes d'un vecteur colonne $X \in \mathfrak{M}_{m,1}(\mathbb{R})$ peut s'écrire $\sum_{i=1}^m X_i = {}^t\mathbf{1}X = {}^tX\mathbf{1}$, où tX désigne la *transposée* de X (vecteur ligne).

Pour tout $B \in \mathfrak{M}_{m,k}(\mathbb{R})$, $B \geq \mathbf{0}$ signifie que **tous** les coefficients de B sont positifs ou nuls. On appelle *distribution* tout vecteur $X \in \mathfrak{M}_{m,1}(\mathbb{R})$ tel que $X \geq 0$ et ${}^t\mathbf{1}X = 1$.

Pour toute matrice carrée $B \in \mathfrak{M}_{m,m}(\mathbb{R})$ et tout nombre réel λ , on note $d_B(\lambda)$ la *dimension* du sous-espace vectoriel $\text{Ker}(B - \lambda I)$, où I désigne la *matrice-identité* :

$$d_B(\lambda) = \dim \text{Ker}(B - \lambda I).$$

Ainsi λ est *valeur propre* de B si et seulement si $d_B(\lambda) \geq 1$. On dit que $d_B(\lambda)$ est la *multiplicité* de la valeur propre λ .

Le *spectre* de B , c'est-à-dire l'ensemble de ses valeurs propres, est noté $\text{Sp}(B)$. Afin de conserver l'information sur la multiplicité de chaque valeur propre, on utilisera la notation $\text{Sp}^*(B)$ pour le vecteur contenant chaque valeur propre λ **répétée** $d_B(\lambda)$ fois. Ainsi $\text{Sp}(\mathbf{1}) = \{1\}$, tandis que $\text{Sp}^*(\mathbf{1}) = (1, \dots, 1)$.

Enfin pour tout sous-ensemble U de \mathbb{R} , $D_B(U)$ désignera le nombre de valeurs propres de B appartenant à U et *comptées avec leur multiplicité* :

$$D_B(U) = \text{Card}(\text{Sp}^*(B) \cap U) = \sum_{\lambda \in U} d_B(\lambda),$$

où les termes non nuls de la dernière somme sont en nombre fini.

Introduction

Dans une population diploïde panmictique, on s'intéresse à un gène existant sous la forme de m *allèles* $\mathcal{A}_1, \dots, \mathcal{A}_m$, avec $m \geq 2$. On désigne par p_i la **proportion de gamètes** qui portent l'allèle \mathcal{A}_i . En particulier, le vecteur p ayant $(p_i)_{1 \leq i \leq m}$ pour composantes est une distribution, appelée *distribution allélique*.

On suppose également que tous les organismes issus d'un zygote $\mathcal{A}_i\mathcal{A}_j$ ont le même **taux de reproduction**, noté $a_{ij} \geq 0$. Ainsi, il sera toujours supposé que la matrice carrée $A = (a_{ij})_{1 \leq i, j \leq m} \in \mathfrak{M}_{m,m}(\mathbb{R})$ **est non nulle et symétrique**.

La dynamique des distributions alléliques d'une génération sur l'autre est donnée par (1), où p' désigne la distribution allélique à la *génération qui suit* :

$$p'_i = V^{-1}p_i(Ap)_i = V^{-1}p_i \sum_{j=1}^m a_{ij}p_j \quad \forall i \in \{1, \dots, m\}, \quad (1)$$

$$\text{où } V = V(p) = {}^t p A p = \sum_{i=1}^m \sum_{j=1}^m a_{ij} p_i p_j.$$

Remarquer que p' est bien une distribution (qui n'est définie que si $V \neq 0$). On appellera $V(p)$ la **viabilité globale** de la population et $V_i(p) = (Ap)_i$ la **viabilité marginale** de l'allèle \mathcal{A}_i .

Première partie : Préliminaires

On dira qu'une distribution allélique P est un **équilibre** si $P' = P$. Si de plus $P_i \neq 0$ pour tout i , on parlera d'**équilibre non dégénéré**.

1. Montrer que si P est un équilibre non dégénéré, alors tous les allèles ont la même viabilité marginale, égale à la viabilité globale de la population :

$$V_i(P) = V(P) \quad \forall i \in \{1, \dots, m\}. \quad (2)$$

2. Soient P et Q deux équilibres non dégénérés. Montrer que

$$V(P) = V(Q). \quad (3)$$

Désormais on désignera par V_e la viabilité globale de la population en tout équilibre non dégénéré (lorsqu'il en existe au moins un).

Deuxième partie : Unicité de l'équilibre

On supposera dans cette partie qu'il existe un équilibre non dégénéré P . Il s'agit d'établir un critère nécessaire et suffisant sur la matrice A pour que P soit l'**unique équilibre non dégénéré**.

Soit $H = \{X \in \mathfrak{M}_{m,1}(\mathbb{R}) : \sum X_i = 0\}$ le sous-espace vectoriel de \mathbb{R}^m d'équation $X_1 + \dots + X_m = 0$. Précisément, il s'agit dans la suite de démontrer l'équivalence :

$$P \text{ est le seul équilibre non dégénéré} \Leftrightarrow H \cap \text{Ker}(A) = \{0\}. \quad (4)$$

1. Soit Q un autre équilibre non dégénéré.
 - a) Montrer que $P - Q \in \text{Ker}(A)$.
 - b) En déduire un des deux sens de l'équivalence (4).
2. On suppose dans cette question qu'il existe un vecteur non nul $v \in H \cap \text{Ker}(A)$.

a) Soit

$$c = \min_{i:v_i \neq 0} \frac{P_i}{|v_i|}.$$

Montrer que c est bien défini et que $c > 0$. Montrer que pour tout $\varepsilon \in]-c; c[$, les composantes du vecteur $Q^{(\varepsilon)} = P + \varepsilon v$ sont toutes dans $]0; 1[$, et que leur somme vaut 1.

b) Conclure.

3. Soit

$$A = \begin{pmatrix} 2 & 1 & 3 \\ 1 & 2 & 0 \\ 3 & 0 & 6 \end{pmatrix}$$

a) Caractériser $H \cap \text{Ker}(A)$.

b) Exhiber deux équilibres non dégénérés distincts et calculer V_e .

Troisième partie : Un principe fondamental de la sélection naturelle

Il s'agit ici de démontrer que la viabilité globale de la population est une fonction **croissante** du temps. Il faudra parfois se servir, sans la démontrer, de l'inégalité de convexité suivante :

$$\sum_{i=1}^m B_i X_i^\gamma \geq \left(\sum_{i=1}^m B_i X_i \right)^\gamma,$$

qui est valide pour toute puissance réelle $\gamma \geq 1$, pour tout $X \geq 0$, et pour toute *distribution* B .

1. Montrer que la viabilité globale à la génération qui suit vaut

$$V' = V^{-2} \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m a_{ij} a_{ik} p_i p_j p_k V_j.$$

2. a) Établir l'inégalité suivante

$$\frac{a+b}{2} \geq \sqrt{ab} \quad \forall a, b \geq 0.$$

b) Montrer que

$$V' \geq V^{-2} \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m a_{ij} a_{ik} p_i p_j p_k \sqrt{V_j V_k}.$$

3. Montrer que

$$V' \geq V^{-2} \left(\sum_{i=1}^m p_i \sum_{j=1}^m a_{ij} p_j \sqrt{V_j} \right)^2.$$

4. a) Montrer que

$$V' \geq V^{-2} \left(\sum_{i=1}^m p_i V_i^{3/2} \right)^2.$$

b) Conclure.

Quatrième partie : Stabilité des équilibres non dégénérés

Soit P un équilibre ($P' = P$). Si partant de toute distribution allélique *sensiblement différente* de P , les distributions alléliques futures de la population convergent vers P , l'équilibre P sera dit **stable**. Au vu de la partie précédente, un équilibre P est stable si et seulement si $V(P) = {}^tPAP$ est un *maximum local strict* de viabilité.

Dans cette partie, on suppose qu'il existe un équilibre non dégénéré P et l'on cherche à déterminer à quelle condition $V(P)$ est un **maximum local** de viabilité. L'objectif est de prouver les deux équivalences suivantes :

$$V(P) \text{ est un maximum local de viabilité} \quad (5)$$

$$\Updownarrow$$

$$V_e \text{ est un maximum global de viabilité} \quad (6)$$

$$\Updownarrow$$

$$\alpha = 1,$$

où

$$\alpha = D_A(\mathbb{R}_+^*).$$

1. a) Montrer que $A = RD^tR$ où R est une matrice dont on donnera les caractéristiques, et D est une matrice diagonale dont on qualifiera les éléments.

b) En déduire que $\alpha \geq 1$.

Indication. Considérer les quantités du type tXAX .

2. On désigne par $\|\cdot\|$ la norme euclidienne sur \mathbb{R}^m , définie par $\|X\| = \sqrt{{}^tXX}$.

a) En écrivant toute distribution allélique S sous la forme $S = P + X$, montrer que tPAP est un maximum local de viabilité ssi

$$\exists \varepsilon > 0 \text{ tel que } \forall X \in \mathbb{R}^m, \text{ si } \|X\| < \varepsilon \text{ et } {}^tIX = 0, \text{ alors } {}^tXAX \leq 0.$$

b) Montrer que cette dernière équation est équivalente à

$$\forall X \in \mathbb{R}^m, \text{ si } {}^tIX = 0, \text{ alors } {}^tXAX \leq 0.$$

c) Soit $Q = {}^tRP$. Prouver (5) et montrer que V_e est un maximum global de viabilité ssi

$$\forall Y \in \mathbb{R}^m, \text{ si } {}^tQDY = 0, \text{ alors } {}^tYDY \leq 0.$$

3. Déduire des questions précédentes que si V_e est un maximum de viabilité, alors $\alpha = 1$.

4. Réciproquement, on suppose dans cette question que $\alpha = 1$. Sans perte de généralité, on écrira $\text{Sp}^*(A) = (\lambda_1, \dots, \lambda_m)$, avec $\lambda_1 > 0$.

a) Montrer que $\lambda_1 Q_1^2 > \sum_{k=2}^m |\lambda_k| Q_k^2$.

b) Soit Y tel que ${}^t Q D Y = 0$. Démontrer l'inégalité suivante

$$\lambda_1^2 Q_1^2 Y_1^2 \leq \left(\sum_{k=2}^m |\lambda_k| Q_k^2 \right) \left(\sum_{j=2}^m |\lambda_j| Y_j^2 \right).$$

c) Conclure.

5. On souhaite ici appliquer les équations numérotées (4) et (6) au cas particulier où tous les hétérozygotes ont le même taux de reproduction égal à 1 et tous les homozygotes ont le même taux de reproduction égal à $1 - s$, avec $s \in]0; 1[$.

a) Expliciter la matrice A et montrer que $-s$ et $m - s$ sont des valeurs propres de A dont on précisera la multiplicité.

b) Montrer qu'il existe au plus un équilibre non dégénéré.

c) Montrer qu'il existe un équilibre non dégénéré réalisant un maximum de viabilité V_e dont on précisera la valeur.

Cinquième partie : Extinction d'allèles

Le but de cette partie est d'étendre le résultat de la partie précédente en répondant à la question suivante : si A a plus d'une valeur propre strictement positive, **combien d'allèles en proportions non nulles** peut compter un équilibre stable (alors dégénéré) ?

On dit d'une matrice B obtenue à partir de A après lui avoir ôté q lignes et q colonnes indicées par un **même** sous-ensemble $\{i_1, \dots, i_q\}$ de $\{1, \dots, m\}$, qu'elle est une **sous-matrice principale** de A d'ordre $m - q$.

A Un théorème de Cauchy

Dans toute cette sous-partie, B désigne une sous-matrice principale de A d'ordre $m - 1$. Soit D la matrice diagonale d'ordre $m - 1$ dont les coefficients sont exactement les éléments de $\text{Sp}^* B = (\mu_1, \dots, \mu_{m-1})$.

1. Montrer qu'il existe une matrice réelle inversible V d'ordre m , un vecteur colonne z à $m - 1$ composantes réelles, un nombre réel a , tels que

$$V^{-1} A V = \left(\begin{array}{c|c} D & z \\ \hline \frac{t_z}{a} & a \end{array} \right).$$

Jusqu'à la fin de cette sous-partie, on notera \mathcal{I} l'ensemble des indices correspondant aux composantes **non nulles** de z , \mathcal{M} l'ensemble des éléments de $\text{Sp}^* B$ indicés au moins une fois par

\mathcal{I} , et \mathcal{J} l'ensemble de tous les entiers indiquant un élément de \mathcal{M} :

$$\begin{aligned}\mathcal{I} &= \{i \leq m-1 : z_i \neq 0\} \\ \mathcal{J} &= \{j \leq m-1 : \exists i \in \mathcal{I}, \mu_i = \mu_j\} \\ \mathcal{M} &= \{\mu_i : i \in \mathcal{I}\} = \{\mu_j : j \in \mathcal{J}\}.\end{aligned}$$

2. Montrer que pour tout réel λ et tout vecteur X ,

$$X \in \text{Ker}(V^{-1}(A - \lambda I)V) \Leftrightarrow \begin{cases} (\mu_i - \lambda)X_i + z_i X_m = 0 & \forall i \leq m-1 \\ \sum_{j=1}^{m-1} z_j X_j + (a - \lambda)X_m = 0. \end{cases}$$

3. Soit $\lambda \in \mathcal{M}$. Supposons que $\mathcal{K} = \{i \leq m-1 : \lambda = \mu_i\}$ s'écrive $\mathcal{K} = \{1, \dots, k\}$.

a) Montrer que $\text{Ker}(V^{-1}(A - \lambda I)V)$ est aussi le noyau d'une application linéaire à valeurs dans \mathbb{R} , à préciser, du sous-espace vectoriel E_k de \mathbb{R}^m , où $E_k = \{X \in \mathbb{R}^m : \forall i \geq k+1, X_i = 0\}$.

b) En déduire, grâce au théorème du rang, que $d_A(\lambda) = d_B(\lambda) - 1$.

4. Pour tout $x \notin \mathcal{M}$, on définit

$$f(x) = \sum_{j \in \mathcal{J}} \frac{z_j^2}{x - \mu_j} + a - x.$$

Soit $\Omega = \{x \notin \mathcal{M} : f(x) = 0\}$ l'ensemble des racines de f . En cherchant dans chacun des deux cas une base de $\text{Ker}(V^{-1}(A - \lambda I)V)$, établir les deux implications suivantes :

$$\forall \lambda \notin \mathcal{M} \quad \begin{cases} \lambda \in \Omega \Rightarrow d_A(\lambda) = d_B(\lambda) + 1, \\ \lambda \notin \Omega \Rightarrow d_A(\lambda) = d_B(\lambda). \end{cases}$$

5. Soit n le nombre d'éléments distincts de \mathcal{M} , et $\theta_1 < \dots < \theta_n$ ces éléments.

a) Donner le domaine de définition de f .

b) Soit $]a; b[$ un intervalle quelconque de \mathbb{R} intersectant \mathcal{M} en k de ses n points. Montrer graphiquement que f s'annule au plus $k+1$ fois sur cet intervalle.

c) Prouver le théorème de Cauchy : pour tout intervalle $]a; b[$ de \mathbb{R} ,

$$D_A(]a; b[) \leq D_B(]a; b[) + 1.$$

B Nombre d'allèles survivants

6. Soit B une sous-matrice principale de A d'ordre $m - q$. Montrer que

$$D_B(\mathbb{R}_+^*) \geq D_A(\mathbb{R}_+^*) - q.$$

7. En utilisant les équations numérotées (5) et (6), déduire de la question précédente que si A a r valeurs propres strictement positives comptées avec leur multiplicité, un équilibre ne peut réaliser un maximum local de viabilité que si au moins $r - 1$ allèles y sont en proportions nulles.

