Gènes sur les chromosomes sexuels

©2002 Vincent Guirardel (copyleft LDL : Licence pour Documents Libres).

Source: chromosomes-sexuels.tex.

Version imprimable: chromosomes-sexuels.pdf

Algèbre linéaire. DEUG deuxième année. Angle pédagogique : *À quoi ça sert.*

Objectifs et commentaires. *Etude d'un système dynamique linéaire issu de la génétique. Cet exercice permet de voir les notions de vecteur propre et de valeur propre dans un contexte différent, et motivant.*

Remarque. *L'exercice est probablement trop long pour être traité en entier en TD. On peut le donner en devoir à la maison ou n'en traiter qu'une partie. La première partie est intéressante pour motiver les notions de vecteur propre et valeur propre* avant *leur introduction en cours (ou à défaut comme premier exercice d'une feuille de TD sur ces notions). La troisième partie est une itération d'une application affine. Les calculs de vecteurs propres sont (un peu) plus délicats. Toujours pour la troisième partie, l'interprétation de la valeur limite en fonction des taux de sélection et de mutations ne sont pas évidentes.*

Question 1. Problème introductif : gènes sur les chromosomes sexuels, effet de la reproduction seule

Les femmes ont deux chromosomes X et les hommes ont un chromosome X et un chromosome Y. Certains gènes sont situés sur le chromosome X. C'est le cas par exemple pour un gène G lié à une forme de daltonisme. Ce gène G possède deux versions (on dit deux *allèles* en biologie), une qu'on appellera S (sain) et l'autre qu'on appellera M (malade) qui est à l'origine du daltonisme. En fait, ce gène est recessif, ce qui signifie que seules les femmes qui ont deux fois la version M du gène seront daltoniennes. Les hommes eux n'ont qu'une copie du gène G et sont daltoniens si ils ont la version M. Le but du problème est d'étudier la propagation de ce gène.

Soit H_0 la proportion des gènes G des hommes qui sont en version M à la génération 0. Soit F_0 la proportion des gènes G des femmes qui sont en version M à la génération 0. Pour rendre les choses plus concrètes, on peut supposer que dans l'etat initial, le gène version M n'est présent que chez les hommes, et disons avec une proportion $H_0 = 2\%$

a. Décrire l'évolution des proportions de gènes G chez les hommes et les femmes d'une génération à la suivante. *Indication :* un homme reçoit forcément son chromosome X de sa mère. Une femme reçoit un chromosome X de chaque parent.

b. Exprimer les équations d'évolution de H_n et F_n matriciellement. En déduire une expression du vecteur $\begin{pmatrix} H_n \\ F_n \end{pmatrix}$ en fonction de $\begin{pmatrix} H_0 \\ F_0 \end{pmatrix}$ et d'une puissance d'une matrice A.

c. Des conditions initiales pour lesquelles les calculs sont faciles. — Supposons que la population initiale, décrite par $v_0 = \begin{pmatrix} H_0 \\ F_0 \end{pmatrix}$ satisfasse $A.v_0 = \lambda v_0$ pour un certain réel λ.[1] Vérifier que $\begin{pmatrix} H_n \\ F_n \end{pmatrix}$ est alors facile à calculer.

d. Chercher les $\lambda \in \mathbb{R}$ tels que l'équation d'inconnue v $A.v = \lambda v$ admette une autre solution que le vecteur nul. Quels sont les vecteurs v correspondants ? Ces vecteurs ont-ils une interprétation biologiques ?

---

[1] On dit dans ce cas que v_0 est vecteur propre de A (si $v_0 \neq 0$).

e. Principe de superposition (linéarité) : comment résoudre le problème à partir des cas des cas où les calculs sont faciles. — Soit $v_0 = \begin{pmatrix} 2/100 \\ 0 \end{pmatrix}$. Écrire v_0 comme une somme de deux vecteurs propres v_1, v_2 (dont les images par A^n sont faciles à calculer). En déduire $\begin{pmatrix} H_n \\ F_n \end{pmatrix}$.

f. En déduire la limite de $\begin{pmatrix} H_n \\ F_n \end{pmatrix}$ quand $n \to \infty$ Que peut-on constater, à la limite, pour les proportions de gène M chez les hommes et les femmes ? Au bout de combien de génération ceci est-il vrai à 1% près ? À 10^{-6} près ?

Remarques. En résumé, le cheminement est le suivant : on cherche le plus possible de vecteurs propres pour lesquels l'image par A^n est facile à calculer. On espère en trouver assez pour que toute condition initiale (ou au moins, celle qui nous interesse) puisse s'exprimer comme superposition (combinaison linéaire) de vecteurs propres. Si c'est le cas, on a gagné.

Le fait que *tout* vecteur peut s'exprimer comme combinaison linéaire de vecteurs propres signifie qu'il existe une base formée de vecteurs propres. On dit dans ce cas que A est diagonalisable.

Question 2. Si on introduit la sélection naturelle

Il est plus commode de repenser la question suivante en terme d'opérateur (opérateur linéaire = endomorphisme). On s'intéresse toujours à la répartition du gène M dans la population masculine et féminine. On appelle donc *état de la population* le vecteur $v = \begin{pmatrix} H \\ F \end{pmatrix}$. L'espace des états de la population est donc ici $V = \mathbb{R}^2$. [2]

[2] Même si les états pertinents biologiquement ont leurs deux coordonnées positives, on a vu l'utilité de considérer des vecteurs a priori non pertinents biologiquement : pour parler de vecteur propre, il faut se trouver dans un espace vectoriel.

Dans la question précédente, on a étudié l'effet de l'itération de *l'opérateur reproduction.* Cet opérateur prend en entrée l'état de la population avant reproduction, et sort l'état de la population après reproduction. Dans le modèle donné au dessus, R est une application linéaire de V dans V donnée par $R(\begin{pmatrix} H \\ F \end{pmatrix}) = \begin{pmatrix} F \\ \frac{1}{2}(H+F) \end{pmatrix}$.

On veut de même modéliser la sélection naturelle par un *opérateur de sélection* S. On part du principe que les hommes sont beaucoup plus atteints de daltonisme que les femmes (car M est recessif, donc pour qu'une femme soit malade, il faut qu'elle ait deux gènes M), et on fait l'hypothèse que les femmes ne sont jamais malades, et donc que la sélection naturelle ne s'applique pas sur elles. Le nombre d'hommes malades est proportionnel à H, et on peut supposer que la sélection naturelle fait passer d'une proportion H de gènes M à une proportion σH où σ (comme Sélection) est un réel entre 0 et 1, c'est un paramètre du modèle.

a. Que vaut σ pour une sélection naturelle très forte (maladie très handicapante) ? Et pour une maladie peu handicapante ?

b. Donner la matrice de l'opérateur de sélection S.

c. On suppose qu'à chaque génération, il y a d'abord reproduction puis sélection. Donner la matrice de l'opérateur qui fait passer d'une génération n à la suivante.

d. En s'inspirant de ce qui a été fait dans la question précédente, que peut-on dire du comportement de l'état de la population quand n tend vers l'infini ?

e. On a choisi de faire agir la reproduction avant la sélection. Qu'aurait donné le modèle si on avait fait le contraire ?

Question 3. Et des mutations

On introduit maintenant un nouvel opérateur M de mutation. On note μ la probabilité qu'un gène malade mute en gène sain et μ' la probabilité pour qu'une mutation inverse se produise (on suppose que ces probabilités de mutations sont les mêmes chez les hommes et chez les femmes).

a. Calculer l'opérateur de mutation. Que constatez-vous ? Pouvez-vous donner sa matrice ? Ecrire $M(v)$ sous la forme $N(v) + v_0$.

b. Supposons qu'à chaque génération, il y ait d'abord mutation, puis reproduction, puis sélection naturelle. Quel est l'opérateur qui fait passer d'une génération à la suivante ? Écrire cet opérateur sous la forme $A(v) + w_0$.

On est ainsi conduit à l'étude de l'itération d'une application affine $v \mapsto A(v) + w_0$. Un *truc* permet de se ramener à l'étude d'un opérateur linéaire : il s'agit d'introduire artificiellement une coordonnée supplémentaire z, telle qu'on retrouve l'opérateur précédent en faisant $z = 1$, mais de sorte que le nouvel opérateur soit linéaire (et non plus affine). Ici, on considèrera $M' : \begin{pmatrix} H \\ F \\ z \end{pmatrix} \mapsto \begin{pmatrix} A \begin{pmatrix} H \\ F \end{pmatrix} + z w_0 \\ z \end{pmatrix}$.

[3]

c. A partir de l'étude des vecteurs propres et valeurs propres de M', donner le comportement de la population lorsque n tend vers l'infini.

---

[3] On peut aussi faire un changement de variables affine bien choisi.