

## THÈME 3

### SOUS-THÈME 3-1

# ÉQUILIBRE DE HARDY-WEINBERG

#### Mots-clés

Fréquences, probabilités, allèles, génotypes, Hardy-Weinberg.

#### Références au programme

Les populations évoluent au cours du temps. Des modèles mathématiques probabilistes et des outils statistiques permettent d'étudier les mécanismes évolutifs impliqués.

#### Savoirs

Le modèle mathématique de Hardy-Weinberg utilise la théorie des probabilités pour décrire le phénomène aléatoire de transmission des allèles dans une population. En assimilant les probabilités à des fréquences pour des effectifs de grande taille (loi des grands nombres), le modèle prédit que la structure génétique d'une population de grand effectif est stable d'une génération à l'autre sous certaines conditions (absence de migration, de mutation et de sélection). Cette stabilité théorique est connue sous le nom d'équilibre de Hardy-Weinberg.

#### Savoir-faire

Pour la transmission de deux allèles dans le cadre du modèle de Hardy-Weinberg, établir les relations entre les probabilités des génotypes d'une génération et celles de la génération précédente.

Produire une démonstration mathématique ou un calcul sur tableur ou un programme en Python pour prouver ou constater que les probabilités des génotypes sont constantes à partir de la seconde génération (modèle de Hardy-Weinberg).

#### Notions mathématiques travaillées

- Suites
- Relations de récurrence
- Calculs de fréquences et de probabilités
- Principe de l'estimation d'une probabilité, ou d'une proportion dans une population, par une fréquence observée sur un échantillon.
- Arbre pondéré
- Fonction polynôme du second degré

## Histoire, enjeux, débats

À la fin du XIX<sup>e</sup> siècle, la redécouverte des travaux de Gregor Mendel suscite quelques doutes. En particulier, certains se demandent pourquoi le phénotype défini par un allèle récessif ne disparaît pas au cours des générations. Le généticien Reginald Punnett, professeur à l'université de Cambridge pose la question à son collègue mathématicien Godfrey Harold Hardy. Hardy propose sa solution dans un article intitulé *Mendelian proportions in a mixed population* publié dans la revue *Science* en 1908. Plusieurs décennies plus tard, il apparaît que la loi de Hardy avait été découverte la même année par un médecin allemand Wilhelm Weinberg. Son article, dans lequel il expose la même loi de stabilité que Hardy, était paru dans un journal scientifique peu connu et n'avait pas été remarqué. On a finalement donné le nom de loi de Hardy-Weinberg à la loi de stabilité des génotypes découverte de manière indépendante par ces deux scientifiques.

## Les mathématiques et l'évolution de la composition génétique

### Allèles et génotypes

Les unités d'informations génétiques qui déterminent les caractères héréditaires sont de petites portions d'une séquence d'ADN appelées gènes. Chaque gène occupe une place déterminée sur les chromosomes. Si la cellule est diploïde, c'est-à-dire si elle possède un nombre pair de chromosomes, ceux-ci sont réunis en paires dites homologues. Chaque gène est présent en deux exemplaires, un exemplaire sur chacun des deux chromosomes homologues. Chacun des deux exemplaires d'un gène s'appelle un allèle.

Dans ce qui suit, on se limite au cas de deux types d'allèles, notés A et a, d'un même gène. Lors de la conception, un organisme reçoit deux allèles (un de chacun de ses parents). Ces deux allèles constituent son génotype qui peut alors être AA, Aa ou aa. Les génotypes AA et aa sont dits homozygotes et Aa est dit hétérozygote.

Pour simplifier l'étude, on fait ici les hypothèses suivantes :

- la population est de grande taille, ce qui permet d'assimiler les fréquences observées à des probabilités, en vertu de la loi des grands nombres ;
- au sein de la population, le choix du partenaire sexuel se fait au hasard. C'est l'hypothèse de « panmixie » ;
- il n'y a pas de migration : aucune copie allélique n'est apportée de l'extérieur ;
- il n'y a pas de mutation ;
- il n'y a pas de sélection ;
- les générations sont séparées (pas d'union possible entre des individus de générations différentes).

### Évolution du génotype et des allèles d'une génération à la suivante

On se propose d'étudier conjointement l'évolution des proportions des trois génotypes et des deux allèles. On note  $p_n$  la proportion de génotypes AA de la génération  $n$ ,  $q_n$  la proportion de génotypes Aa et  $r_n$  celle des génotypes aa. Pour tout entier naturel  $n$ , les trois nombres  $p_n, q_n, r_n$  sont donc compris entre 0 et 1 et vérifient  $p_n + q_n + r_n = 1$ .

Il est mathématiquement plus simple de considérer dans un premier temps, non pas l'ensemble des génotypes, mais l'ensemble des allèles (le pool allélique) et la proportion de chacune des allèles A et a dans cet ensemble de référence à chaque génération.

On note  $A_n$  (respectivement  $a_n$ ) la proportion d'allèles A (respectivement d'allèles a) de la génération  $n$ .

Pour tout naturel, les deux nombres  $A_n$  et  $a_n$  sont donc compris entre 0 et 1 et vérifient  $a_n + A_n = 1$ .

La transmission des allèles de la génération  $n$  à la génération  $n + 1$  peut être représentée par un tableau.

|                                     |    |    |
|-------------------------------------|----|----|
| Allèle de la mère<br>Allèle du père | A  | a  |
| A                                   | AA | Aa |
| a                                   | Aa | aa |

Le génotype AA à la génération  $n + 1$  ne pouvant provenir que de deux allèles A et A de la génération  $n$ , il vient :  $p_{n+1} = A_n \times A_n = A_n^2$

Le principe multiplicatif se fonde à la fois sur l'assimilation des fréquences à des probabilités et sur l'indépendance des croisements : la probabilité du génotype AA à la génération  $n + 1$  est donc le produit des probabilités de chaque allèle A à la génération  $n$ .

Pour la même raison,  $r_{n+1} = a_n^2$ .

$$q_{n+1} = 2A_n a_n$$

D'où les relations de récurrence, valables pour tout entier naturel  $n$  :

$$p_{n+1} = A_n^2$$

$$q_{n+1} = 2A_n a_n$$

$$r_{n+1} = a_n^2$$

Chaque génotype AA correspond à deux allèles A, chaque génotype Aa correspond à un allèle A et un allèle a et chaque génotype aa correspond à deux allèles a.

Si on note  $N$  le nombre total de génotypes à la génération  $n$ , alors le nombre total d'allèles à la génération  $n$  est  $2N$ , chaque génotype correspondant à deux allèles.

Le nombre total d'allèles A à la génération  $n$  est égal à  $2 \times p_n \times N + q_n \times N$ . D'où la proportion  $A_n$  d'allèles A parmi la totalité des  $2N$  allèles :

$$A_n = \frac{2 \times p_n \times N + q_n \times N}{2N} = p_n + \frac{1}{2} q_n$$

En raisonnant de même pour  $a_n$  ou en remarquant que  $a_n = 1 - A_n = p_n + q_n + r_n - \left(p_n + \frac{1}{2} q_n\right)$ , on a finalement :

$$A_n = p_n + \frac{1}{2} q_n \quad a_n = r_n + \frac{1}{2} q_n$$

D'où les relations entre les proportions génotypiques de la génération  $n$  et de la génération  $n + 1$  :

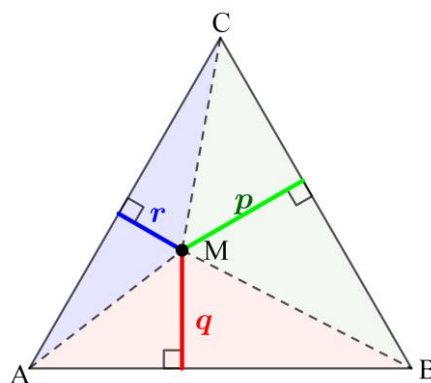
$$p_{n+1} = \left(p_n + \frac{1}{2}q_n\right)^2$$

$$q_{n+1} = 2\left(p_n + \frac{1}{2}q_n\right)\left(r_n + \frac{1}{2}q_n\right)$$

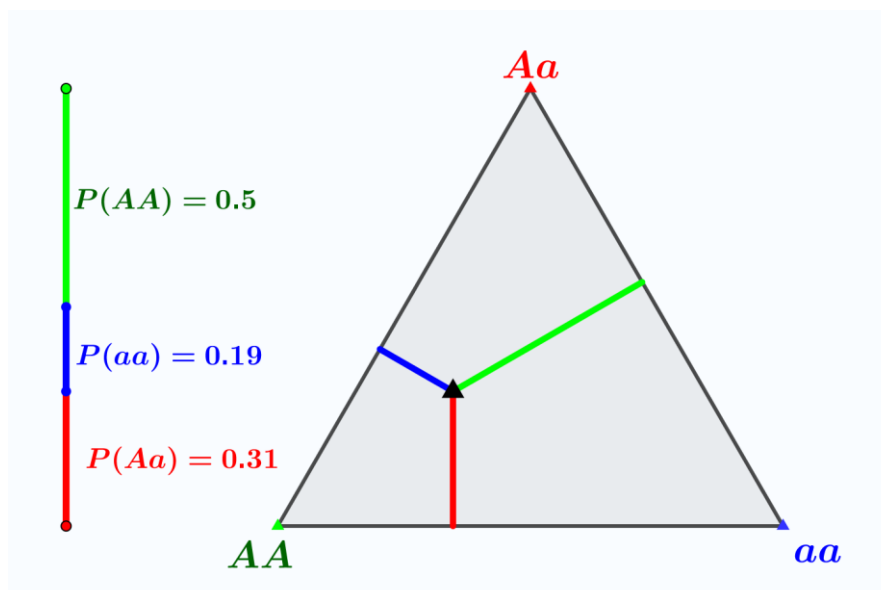
$$r_{n+1} = \left(r_n + \frac{1}{2}q_n\right)^2$$

### Représentation géométrique : diagramme de De Finetti

On commence par démontrer que trois nombres  $p, q, r$  positifs de somme égale à 1 peuvent être représentés par les trois distances d'un point intérieur à un triangle équilatéral à chacun des côtés de ce triangle.



L'aire de chacun des triangles  $AMB, BMC, CMA$  est proportionnelle à leur hauteur et la somme de ces aires, égale à celle du triangle  $ABC$ , reste constante quand le point  $M$  varie à l'intérieur du triangle. La somme  $p + q + r$ , d'ailleurs égale à la mesure de la hauteur du triangle équilatéral  $ABC$  (théorème de Viviani) est donc constante. À un facteur d'homogénéité près, on peut considérer que cette somme vaut 1. L'appliquette GeoGebra intitulée « [Hardy-Weinberg & diagramme de De Finetti](#) » permet de visualiser toutes les probabilités possibles pour les trois génotypes  $AA, Aa, aa$  en déplaçant le petit triangle noir à l'intérieur du triangle gris (frontière comprise) de sommets  $Aa, aa, AA$  :



## Constat de la stabilité des fréquences au tableur

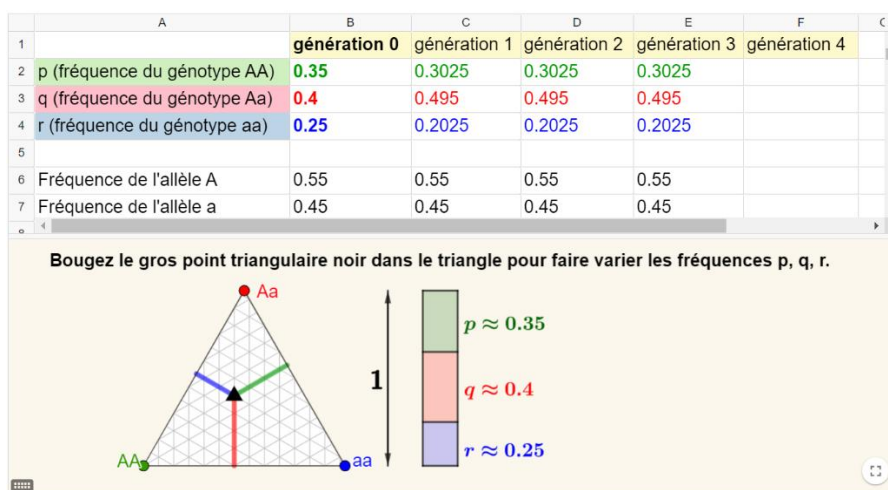
Après avoir saisi les formules au tableur, on obtient par recopie, les premiers termes des suites  $(p_n)$ ,  $(q_n)$ ,  $(r_n)$ ,  $(A_n)$ ,  $(a_n)$ .

|   | A | B     | C        | D     | E            | F            |
|---|---|-------|----------|-------|--------------|--------------|
| 1 | n | $p_n$ | $q_n$    | $r_n$ | $A_n$        | $a_n$        |
| 2 | 0 | 0,7   | 0,2      | 0,1   | =B2 + C2 / 2 | =D2 + C2 / 2 |
| 3 | 1 | =E2^2 | =2*E2*F2 | =F2^2 | =B3 + C3 / 2 | =D3 + C3 / 2 |
| 4 |   |       |          |       |              |              |

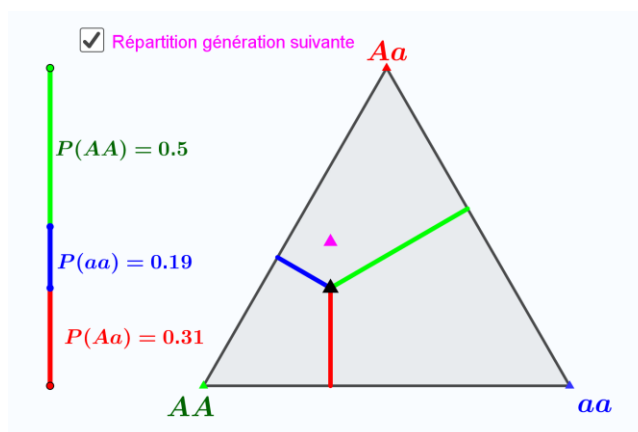
|   | A | B     | C     | D     | E     | F     |
|---|---|-------|-------|-------|-------|-------|
| 1 | n | $p_n$ | $q_n$ | $r_n$ | $A_n$ | $a_n$ |
| 2 | 0 | 0,7   | 0,2   | 0,1   | 0,8   | 0,2   |
| 3 | 1 | 0,64  | 0,32  | 0,04  | 0,8   | 0,2   |
| 4 | 2 | 0,64  | 0,32  | 0,04  | 0,8   | 0,2   |
| 5 | 3 | 0,64  | 0,32  | 0,04  | 0,8   | 0,2   |
| 6 | 4 | 0,64  | 0,32  | 0,04  | 0,8   | 0,2   |
| 7 | 5 | 0,64  | 0,32  | 0,04  | 0,8   | 0,2   |

Si on modifie les valeurs de  $p_0$  et  $q_0$ , on constate que les fréquences alléliques  $A_n$  et  $a_n$  sont constantes à partir de  $n = 0$  (première génération) et que les fréquences génotypiques  $p_n$ ,  $q_n$  et  $r_n$  sont constantes à partir de  $n = 1$  (deuxième génération).

L'appliquette GeoGebra intitulée « [Hardy-Weinberg – Tableur](#) » permet aussi d'observer cette stabilité des fréquences.



## Visualisation de la stabilité sur un diagramme de De Finetti



L'appliquette GeoGebra intitulée « [Hardy-Weinberg & diagramme de De Finetti](#) » permet de visualiser :

- le point triangulaire noir ( $M_0$ ) correspondant aux fréquences alléliques ( $p_0, q_0, r_0$ ) à la génération 0 ;
- le point triangulaire rose ( $M_1$ ) correspondant aux fréquences alléliques ( $p_1, q_1, r_1$ ) à la génération 1 si on coche la case « répartition génération suivante ».

On peut faire plusieurs observations.

- Si on déplace manuellement le point initial  $M_0$ , jusqu'à le superposer au point  $M_1$ , on constate que le point suivant,  $M_2$ , coïncide aussi avec  $M_1$ . Cela illustre la stabilité des fréquences génotypiques à partir du rang 1 ( $M_1 = M_2 = M_3 = \dots$ ).
- Pour toute position initiale du point  $M_0$ , le point  $M_1$  est à la verticale de  $M_0$ . Cela peut s'expliquer par le fait que les fréquences alléliques sont constantes :  $A_0 = A_1$ ;  $a_0 = a_1$
- En activant la trace, on peut visualiser le lieu du point  $M_1$ , qui est un arc de parabole correspondant à tous les équilibres de Hardy-Weinberg.

## Démonstration mathématique de la stabilité de Hardy Weinberg

Soit  $n$  un entier naturel. On a :

$$A_{n+1} = p_{n+1} + \frac{1}{2}q_{n+1} = A_n^2 + A_n a_n = A_n(A_n + a_n) = A_n \text{ puisque } A_n + a_n = 1.$$

Ceci pour tout entier naturel  $n$ , ainsi la suite  $(A_n)$  est constante, et donc la suite  $(a_n)$  aussi puisque pour tout  $n$  on a  $a_n = 1 - A_n$ .

On a ainsi démontré la stabilité des fréquences alléliques dès la première génération.

Puisque  $p_{n+1} = A_n^2$  et que tous les  $A_n$  sont égaux entre eux à partir de  $n = 0$ , on en déduit que tous les  $p_n$  sont aussi égaux entre eux à partir de  $n = 1$ .

De la même manière, tous les  $q_n$  sont aussi égaux entre eux à partir de  $n = 1$  et tous les  $r_n$  sont aussi égaux entre eux à partir de  $n = 1$ .

On a ainsi démontré la stabilité des fréquences génotypiques à partir de la deuxième génération.

## Représentation des fréquences génotypiques en fonction de la fréquence initiale de l'allèle A

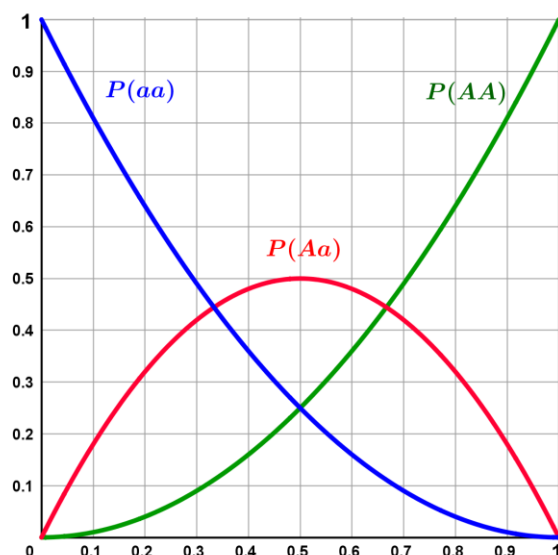
On note  $A_0 = x$  la fréquence de l'allèle A à la génération 0.

On a  $p_1 = x^2$  et, pour tout entier naturel supérieur ou égal à 1,  $p_n = p_1 = x^2$ .

$q_1 = 2x(1 - x)$  et, pour tout entier naturel supérieur ou égal à 1,  $q_n = q_1 = 2x(1 - x)$ .

$r_n = (1 - x)^2$  et, pour tout entier naturel supérieur ou égal à 1,  $r_n = r_1 = (1 - x)^2$ .

Les courbes représentatives des fonctions  $x \mapsto x^2$ ,  $x \mapsto x(1 - x)$  et  $x \mapsto (1 - x)^2$  représentent les fréquences des génotypes AA, Aa et aa en fonction de la fréquence de l'allèle A à la génération 0.



L'interprétation de ces courbes montre :

- d'une part que la fréquence maximale des hétérozygotes (génotype Aa) est atteinte lorsque la fréquence  $x$  de l'allèle A vaut  $\frac{1}{2}$  ;
- la parabole verte est située sous la parabole rouge pour de petites valeurs de  $x$ . Ainsi, lorsque l'un des allèles est rare (par exemple un allèle récessif responsable d'une maladie rare), la majorité des sujets porteurs de cet allèle est hétérozygote (donc non atteinte par la maladie).

#### Intentions pédagogiques

L'étude mathématique de ce sous-thème permet de modéliser un phénomène génétique dépendant du hasard. Le recours aux proportions permet d'éviter d'utiliser les probabilités conditionnelles.

## Propositions d'activités

Les activités proposées permettent de manipuler les relations entre les différentes fréquences alléliques et génotypiques.

### Activité 1

La phénylcétonurie est une maladie génétique rare. Elle est due à un défaut porté par un allèle a récessif. La maladie apparaît uniquement à l'état homozygote aa, mais un sujet non malade hétérozygote peut être porteur du gène défectueux. La probabilité d'apparition du défaut sur un allèle est de  $\frac{1}{100}$ .

1. Calculer la proportion de sujets atteints par la phénylcétonurie.
2. Calculer la proportion de sujets porteurs de l'allèle défectueux, mais non atteints par la phénylcétonurie.

## Activité 2

La mucoviscidose est une maladie qui frappe 1 enfant sur 2500. L'étude de sa transmission a montré qu'elle est due à l'état homozygote (aa) d'un certain gène. Les individus hétérozygotes (Aa) sont sains et il est même impossible de détecter chez eux l'allèle pathogène.

1. Quelle est la fréquence de l'allèle pathogène ?
2. Quelle est la fréquence des sujets porteurs sains d'un allèle a ?

## Liens avec le programme de la spécialité « mathématiques » de première

Les élèves qui ont suivi en classe de première l'enseignement de spécialité mathématiques peuvent utiliser la probabilité d'événements indépendants, par exemple pour retrouver la relation :

$$A_n = p_n + \frac{1}{2}q_n.$$

En effet, la probabilité  $A_n$  d'obtenir un allèle A à la génération  $n$  est égale à la somme :

- de la probabilité d'obtenir le génotype AA à la génération  $n$  et que ce génotype fournisse l'allèle A, que l'on peut noter  $P((AA)_n \cap (AA \rightarrow A))$  ;
- de la probabilité d'obtenir le génotype Aa à la génération  $n$  et que ce génotype fournisse l'allèle A, que l'on peut noter  $P((Aa)_n \cap (Aa \rightarrow A))$ .

Or  $P((AA)_n \cap (AA \rightarrow A)) = P((AA)_n) \times P(AA \rightarrow A) = p_n \times 1 = p_n$ .

De même,  $P((Aa)_n \cap (Aa \rightarrow A)) = P((Aa)_n) \times P(Aa \rightarrow A) = q_n \times \frac{1}{2}$ .

D'où le résultat.

## Bibliographie et sitographie

- Le modèle de Hardy-Weinberg face à la réalité. IREM Paris 13
- Nicolas Bacaër, *Histoires de mathématiques et de populations*, Éditions Cassini, collection Le sel et le fer.
- C. Cannings, A. W. F. Edwards, *Natural selection and the De Finetti diagram*.