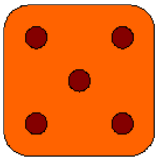
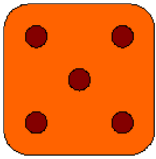


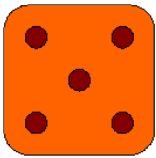
## V. GÉNÉTIQUE ET PROBABILITÉS



Un peu d'histoire



Les maladies génétiques



Loi de Hardy – Weinberg

## UN PEU D'HISTOIRE



Depuis ses débuts, la génétique entretient avec la statistique des rapports étroits.

La découverte de Gregor Mendel et les enseignements qu'il sut en tirer reste un modèle de nos jours. Il reste le premier à avoir utilisé le raisonnement statistique pour étayer sa théorie corpusculaire de l'hérédité et réfuter la conception erronée du « mélange des sangs » en vogue précédemment.

Rappelons brièvement l'historique de la découverte de Mendel au travers de ses études portant sur les pois réalisées au milieu du 19<sup>ième</sup> siècle.

Il s'intéressa en particulier à deux caractères l'aspect lisse ou ridé et la couleur jaune ou verte de la graine.

En croisant des lignées pures de pois ayant des graines rondes avec des lignées pures de pois ayant des graines ridées il remarqua que les pois hybrides ainsi obtenus

avaient tous des graines rondes. Il en déduisit que le caractère « rond » était dominant sur le caractère « ridé » qualifié lui, de récessif. En croisant les hybrides entre eux, Mendel observa à nouveau le caractère récessif « ridé », chez une partie des descendants.

En croisant des lignées pures de pois ayant des graines jaunes et rondes avec des lignées pures de pois ayant des graines vertes et ridées, il nota que les pois hybrides ainsi obtenus étaient à graines jaunes et rondes. Mendel en conclut que la couleur jaune devait être dominante sur la couleur verte (comme la forme ronde l'était sur la forme ridée)

En croisant ces hybrides à graines jaunes et rondes entre eux, il observa la réapparition des caractères récessifs, c'est-à-dire qu'une partie des descendants étaient à graines vertes et ridées.

Le tableau ci-dessous donne les résultats observés par Mendel :

Aspect des graines	Jaunes Rondes	Jaunes Ridées	Vertes Rondes	Vertes Ridées
Effectifs	315	101	108	32

Les conclusions de Mendel ont été depuis pleinement confirmées par la découverte des gènes portés par les chromosomes.

Dans le cas des pois le caractère couleur est codé par un gène<sup>1</sup> présentant deux formes allèles J et j correspondant aux couleurs jaune et vert. Le jaune est dominant, le vert récessif. Le caractère rond ou ridé, est porté par un autre gène à deux allèles R (rond dominant) et r (ridé récessif). Si on croise deux individus dont le génotype est JjRr, on obtient 16 génotypes équiprobables comme le montre le tableau ci-dessous :

<sup>1</sup> Voir en annexe 1 un petit glossaire des termes utilisés en génétique

		Parent 1				
		R		r		
		J	j	J	j	
Parent 2	R	J	jaune	jaune	jaune	jaune
		j	jaune	verte	jaune	verte
	r	J	jaune	jaune	jaune et ridée	jaune et ridée
		j	jaune	verte	jaune et ridée	verte et ridée

Les descendants seront jaunes et ronds dans 9 cas sur 16, jaunes et ridés dans 3 cas sur 16, verts et ronds dans 3 cas sur 16 finalement verts et ridés dans 1 cas sur 16. Dans le cas du croisement de 556 pois réalisé par Mendel les effectifs théoriques devraient être ceux du tableau ci-dessous :

Aspect des graines	Jaunes Rondes	Jaunes Ridées	Vertes Rondes	Vertes Ridées
Effectifs théoriques	312.75	104.25	104.25	34.75

Comparons des effectifs observés et des effectifs théoriques à l'aide d'un test du Khi deux<sup>2</sup> pour lequel les calculs sont détaillés dans le tableau ci-dessous :

Aspect des graines	Jaunes Rondes	Jaunes Ridées	Vertes Rondes	Vertes Ridées	Totaux
Effectifs observés	315	101	108	32	556
Effectifs théoriques	312,75	104,25	104,25	34,75	556
Écart	2,25	-3,25	3,75	-2,75	0
Carrés	5,0625	10,5625	14,0625	7,5625	37,25
Carrés/théoriques	0,016187	0,101319	0,134892	0,217626	0,470024
				<b>d<sup>2</sup>=</b>	<b>0,470024</b>

Pour pouvoir rejeter l'hypothèse d'adéquation  $d^2$  devrait dépasser 7,82 (valeur lue dans la table du Chi deux pour 3 degrés de liberté). La valeur observée est très inférieure à ce seuil.

En fait, elle est même inférieure au premier décile ce qui signifie que, dans l'hypothèse où le modèle proposé est le bon, un tel résultat expérimental avait, a priori, une probabilité inférieure à 1/10 de se produire. Ce résultat, presque « trop beau pour être vrai », a valu à Mendel d'être soupçonné d'avoir manipulé ses chiffres pour qu'ils confirment plus nettement sa théorie.

<sup>2</sup> Le principe du test d'adéquation du Chi deux est présenté dans le chapitre 6.

## L'EXEMPLE DES MALADIES GENETIQUES

### Maladies génétiques causées par un gène récessif

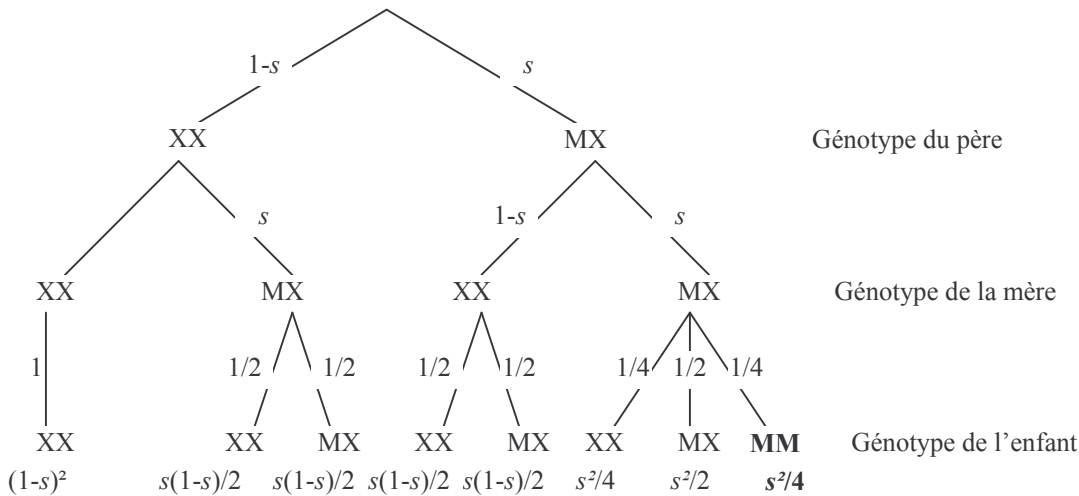
Considérons une maladie génétique humaine due à la présence d'un gène spécifique que nous noterons M réparti indépendamment du sexe. Supposons que seuls les porteurs de la combinaison homozygote MM développent la maladie. Les porteurs de la combinaison MX (où X désigne un allèle autre que M) sont des porteurs sains.

Supposons en outre que la gravité de la maladie est telle que les personnes atteintes ne peuvent avoir d'enfants. La maladie ne pourra apparaître que lorsque les deux parents sont hétérozygotes de type MX.

		Père	
		M	X
Mère	M	MM (malade)	MX (porteur sain)
	X	MX (porteur sain)	XX (non porteur)

Un premier problème est de calculer la proportion  $s$  de porteurs sains dans la population, connaissant la proportion  $f$  de malades.

Si tous les couples de parents sont équiprobables on peut construire l'arbre ci-dessous :



On lit en particulier que le génotype MM a pour probabilité  $f = \frac{s^2}{4}$  et donc que  $s = 2\sqrt{f}$ .

La mucoviscidose est une maladie correspondant approximativement au modèle ci-dessus.

En France environ 1 enfant sur 2000 en est atteint, c'est-à-dire  $f = \frac{1}{2000}$ . Il en résulte que

$s = 2\sqrt{\frac{1}{2000}} \approx 0.0224$  c'est-à-dire que les porteurs sains du gène responsable représentent 2.24% de la population totale.

Dans le cas où les personnes malades peuvent avoir des enfants et en supposant encore que tous les couples de parents sont équiprobables, il faudrait ajouter une troisième branche MM aux deux premiers niveaux de l'arbre ci-dessus. Résumons plutôt les situations possibles et leurs probabilités dans un tableau :

		Génotype du père		
		MM	MX	XX
Génotype de la mère	MM	$f^2$ 1	$fs$ 1/2	$f(1-f-s)$ 0
	MX	$fs$ 1/2	$s^2$ 1/4	$s(1-f-s)$ 0
	XX	$f(1-f-s)$ 0	$s(1-f-s)$ 0	$(1-f-s)^2$ 0

Dans chaque cellule de ce tableau on lit la probabilité de la combinaison de génotype des parents correspondante (première ligne), ainsi que la probabilité que l'enfant soit malade sachant que l'on est dans cette combinaison de génotype des parents.

La probabilité qu'un enfant soit malade dans ce cas est :  $f^2 + fs + \frac{s^2}{4}$ .

On doit donc avoir :

$$f^2 + fs + \frac{s^2}{4} = f$$

que l'on peut écrire sous la forme habituelle d'une équation du second degré d'inconnue  $s$  :

$$\frac{s^2}{4} + fs + f^2 - f = 0$$

dont la résolution donne :

$$\Delta = f^2 - 4 \times \frac{1}{4} (f^2 - f) = f \quad \text{et} \quad s = \frac{-f + \sqrt{f}}{\frac{1}{2}} = 2(\sqrt{f} - f) \quad (\text{seule solution positive}).$$

L'hémochromatose génétique est une maladie correspondant à ce mode de transmission. Elle provoque une accumulation progressive du fer dans l'organisme. Il existe un traitement simple, efficace et peu coûteux (saignées) d'où l'importance d'un dépistage efficace. La fréquence de cette maladie est de l'ordre de 5/1000.

La proportion de porteurs sains est donc :  $s = 2(\sqrt{f} - f) = 2(\sqrt{0.005} - 0.005) \approx 0.131$ .

La maladie étant assez rare, l'utilisation de la formule du paragraphe précédent donnerait  $s = 2\sqrt{f} = 2\sqrt{0.005} \approx 0.141$  soit un résultat encore assez proche de la bonne valeur.

## DANS NOS CLASSES

### Activité sur la loi de Hardy-Weinberg en terminale S

Ce texte a été donné à une classe de Terminale S dans le cadre d'un TD. La partie I a été traitée correctement en une heure environ et la partie II, donnée à faire à la maison de manière facultative, n'a été abordée que par quelques élèves de bon niveau.

Dans les cas simples un gène est formé de 2 allèles pouvant prendre les états A et a et un individu peut donc avoir l'un des trois génotypes suivants : AA, Aa, aa.

Un enfant lors de la conception hérite d'un gène de chacun de ses parents, chacun d'eux étant choisi au hasard. Ainsi si le père est du type AA et la mère de type Aa, les enfants peuvent être du type AA ou Aa.

I. On considère une population dont les proportions des génotypes sont données par le tableau suivant :

AA	Aa	aa
$p_0=0,4$	$q_0=0,3$	$r_0=0,3$

1. Si on considère que les couples se forment au hasard quant aux 3 génotypes considérés, calculer la probabilité pour qu'un enfant ait

- Deux parents de type AA
- Un parent de type AA et un parent de type Aa
- Deux parents de type Aa.

2. Quelle est la probabilité pour que l'enfant soit de type AA sachant que :

- les deux parents sont de type AA ?
- un des deux parents est de type AA et l'autre du type Aa ?
- les deux parents sont de type Aa ?

3. *a.* Calculer la probabilité  $p_1$  pour que l'enfant soit de type AA.

*b.* Calculer de même la probabilité  $r_1$  pour que l'enfant soit de type aa.

*c.* En déduire la probabilité  $q_1$  pour que l'enfant soit de type Aa.

4. Calculer de même les probabilités  $p_2$ ,  $q_2$  et  $r_2$  pour qu'un enfant de seconde génération soit de type AA, Aa et aa. Que remarque-t-on ?

### II. Généralisation

$p_0$ ,  $q_0$  et  $r_0$  sont maintenant des nombres positifs de somme égale à 1.

1. *a.* Montrer que la probabilité pour qu'un enfant soit de type AA est  $p_1 = \left(p_0 + \frac{1}{2}q_0\right)^2$ .

*b.* Sans calcul, donner une expression de  $r_1$ .

2. Soit  $\alpha = p_0 - r_0$ .

a. Montrer que  $p_1 = \frac{(1 + \alpha)^2}{4}$ .

b. En déduire les expressions de  $r_1$  puis de  $q_1$  en fonction de  $\alpha$ . Que peut-on dire de  $p_1 - r_1$  ?

3. Donner les expressions de  $p_2$ ,  $q_2$  et  $r_2$  en fonction de  $\alpha$ .

4. On a étudié les génotypes de 100 enfants choisis au hasard parmi les enfants de la première génération puis les génotypes de 100 enfants choisis au hasard parmi les enfants de la seconde génération. Les fréquences d'enfants de type AA, Aa, aa étant

respectivement  $(\hat{p}_1, \hat{q}_1, \hat{r}_1)$  et  $(\hat{p}_2, \hat{q}_2, \hat{r}_2)$  on remarque que :  $\frac{\hat{q}_1^2}{\hat{p}_1 \hat{r}_1} \approx 3,91$  et que

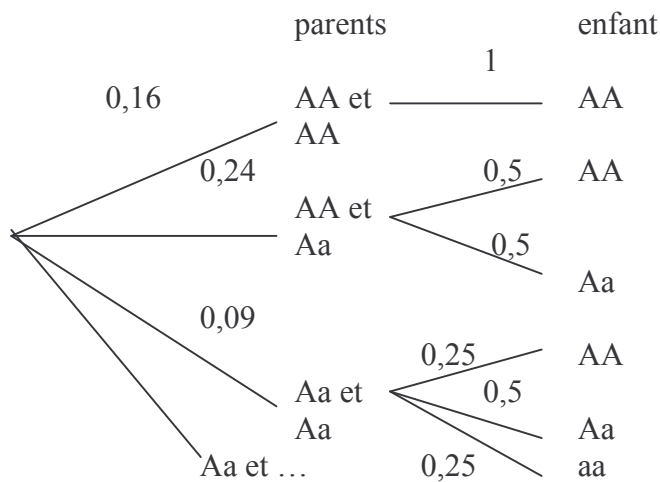
$\frac{\hat{q}_2^2}{\hat{p}_2 \hat{r}_2} \approx 3,95$ . Qu'en pensez-vous ?

**Corrigé**

**I.1.** La probabilité pour qu'un enfant ait ses deux parents de type :

- AA est  $0,4 \times 0,4 = 0,16$  car les deux événements sont indépendants
- AA et Aa est  $2 \times 0,4 \times 0,3 = 0,24$
- Aa est  $0,3 \times 0,3 = 0,09$

**2.** L'enfant ne peut être de type AA si un des parents est de type aa. Ainsi il est possible de représenter la situation à l'aide de l'arbre suivant :



**3. a.** On utilise ici la formule des probabilités totales :

$$p_1 = 0,16 \times 1 + 0,24 \times 0,5 + 0,09 \times 0,25$$

$$= 0,3025.$$

**b.** A l'aide d'un arbre du même type :

$$r_1 = 0,09 \times 1 + 0,18 \times 0,5 + 0,09 \times 0,25$$

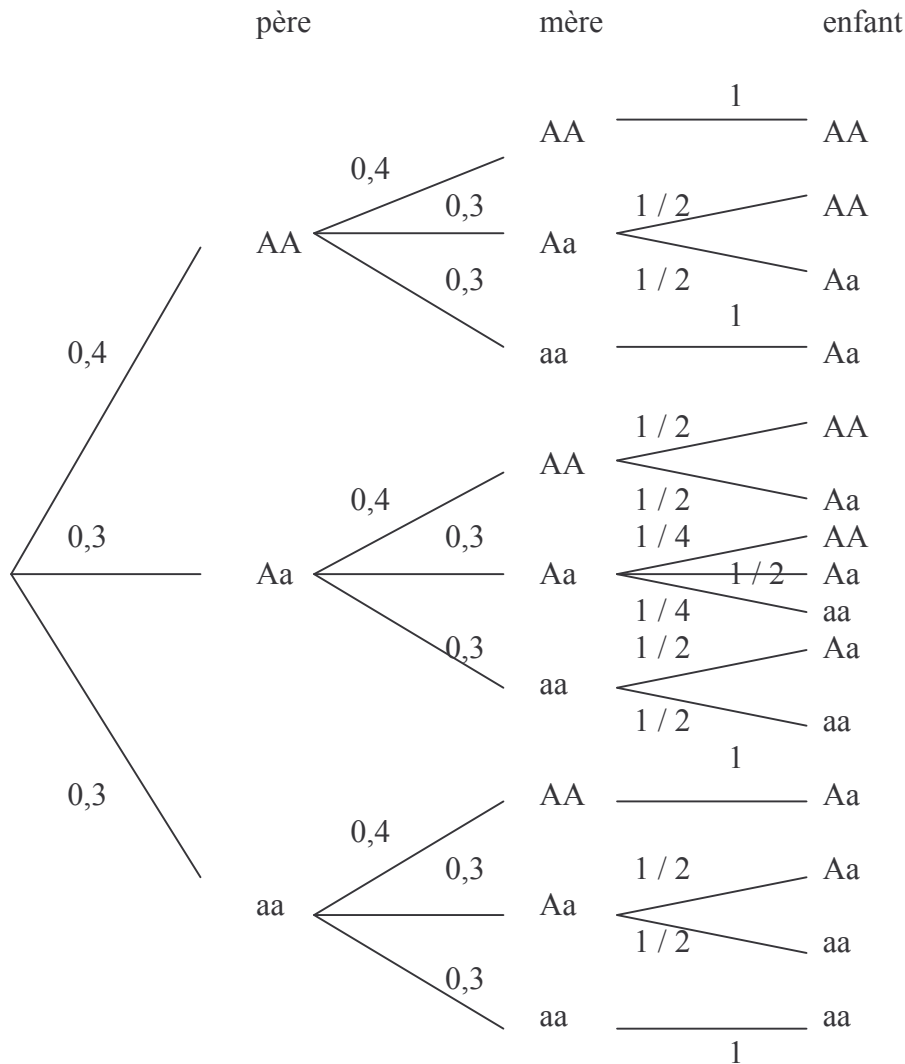
$$= 0,2025.$$

**c.** D'où  $q_1$  :

$$q_1 = 1 - p_1 - r_1 = 0,495.$$



Remarquons qu'un arbre complet permet de répondre directement aux questions 1., 2. et 3. et de retrouver ainsi  $p_1$ ,  $q_1$  et  $r_1$ .



4. Grâce à des arbres du même type on a comme dans la question 3. :

$$p_2 = p_1^2 \times 1 + 2 p_1 q_1 \times 0,5 + q_1^2 \times 0,25 = 0,3025,$$

$r_2 = r_1^2 \times 1 + 2 r_1 q_1 \times 0,5 + q_1^2 \times 0,25 = 0,2025$  et par conséquent  $q_2 = 0,495$ . On remarque que, dès la génération 1, il y a stabilisation des probabilités.

II. 1. a. On obtient de la même façon l'identité  $p_1 = \left( p_0 + \frac{1}{2} q_0 \right)^2$  à partir de la formule des probabilités totales :  $p_1 = p_0^2 \times 1 + 2 p_0 q_0 \times 0,5 + q_0^2 \times 0,25$ .

b. Il est intéressant de remarquer ou de faire remarquer que  $p_0$  et  $r_0$  jouent un rôle symétrique. Ainsi :  $r_1 = \left( r_0 + \frac{1}{2} q_0 \right)^2$ .

2. a. On sait que  $q_0 = 1 - p_0 - r_0$ . Si on remplace dans  $p_1 = \left( p_0 + \frac{1}{2} q_0 \right)^2$  on obtient :

$$p_1 = \left( p_0 + \frac{1}{2} (1 - p_0 - r_0) \right)^2 = \left( \frac{1}{2} + \frac{1}{2} p_0 - \frac{1}{2} r_0 \right)^2 = \frac{(1 + \alpha)^2}{4}.$$

**b.** On peut encore remarquer que  $p_0$  et  $r_0$  jouent un rôle symétrique et que  $r_0 - p_0 = -\alpha$ .

Ainsi :  $r_1 = \frac{(1-\alpha)^2}{4}$  et dans ces conditions :

$$q_1 = 1 - p_1 - r_1 = 1 - \frac{(1+\alpha)^2}{4} - \frac{(1-\alpha)^2}{4} = \frac{1-\alpha^2}{2}.$$

**3.** Comme  $p_1 - r_1 = \frac{(1+\alpha)^2}{4} - \frac{(1-\alpha)^2}{4} = \alpha$ , les calculs précédents sont encore valables pour l'indice suivant et on a :  $p_2 = p_1$ ,  $q_2 = q_1$ ,  $r_2 = r_1$ .

**4.** Formulée comme cela, cette question n'a trouvée aucune réponse, aussi bien auprès des élèves de TS qu'auprès des stagiaires de notre formation. Seul a été mis en évidence le fait que :

$\frac{\hat{q}_1^2}{\hat{p}_1 \hat{r}_1} \approx \frac{\hat{q}_2^2}{\hat{p}_2 \hat{r}_2}$ , ce qui est normal lorsqu'il y a stabilisation des fréquences dès la première génération. En fait il est normal de trouver des valeurs proches de 4 puisque :

$$\frac{q_1^2}{p_1 r_1} = \frac{\left(\frac{1-\alpha^2}{2}\right)^2}{\frac{(1+\alpha)^2}{4} \frac{(1-\alpha)^2}{4}} = 4. \text{ On aurait dû demander d'abord de calculer } \frac{q_1^2}{p_1 r_1} \text{ puis d'en}$$

tirer ensuite une conclusion.

Notons enfin cette question peut être abordée par des élèves n'ayant fait que la partie I.

### En conclusion

Ce modèle a été mis en évidence de façon indépendante au début du XX<sup>e</sup> siècle par Hardy, mathématicien et Weinberg, médecin. Cet équilibre des probabilités dès la première génération suppose réalisées les conditions suivantes : la population est panmictique (les couples se forment au hasard au sens des gènes et leurs gamètes se rencontrent au hasard), il ne doit avoir ni sélection, ni mutation ni migration et enfin il ne doit pas avoir non plus de croisement entre générations successives.

On peut remarquer que si on mélange tous les allèles des individus de la population initiale on obtient alors un ensemble contenant une proportion  $p_0 + \frac{1}{2}q_0$  d'allèles A et une proportion  $r_0 + \frac{1}{2}q_0$  d'allèles a. Si l'ensemble est de taille suffisante, les tirages au hasard de deux allèles pour former un gène peuvent être considérés comme indépendants et on retrouve ainsi dans ces conditions les probabilités  $p_1$  et  $r_1$  de la question II 1.

Par ailleurs si A est l'allèle dominant et si aa est le gène d'une maladie, les individus de type aa sont malades et les individus de type Aa sont des porteurs sains.

Si la population est grande et homogène on doit observer des valeurs  $p_k$ ,  $q_k$  et  $r_k$  constantes de générations en générations.

**Pour les professeurs uniquement**

Et s'il n'y a pas de stabilisation ? Et bien il y a plusieurs causes possibles : consanguinité, migration d'une partie de la population, mortalité accrue des porteurs de l'allèle aa avant l'âge de la procréation. Reste à savoir comment tester l'hypothèse de la stabilité des différentes proportions. La première idée est d'utiliser le  $\chi^2$ .

On suppose qu'on a analysé 939 génotypes d'un échantillon d'une classe d'âge donnée. Le tableau suivant résume les résultats obtenus :

AA	Aa	aa	Total
322	410	207	939

Peut-on dire au vu de ces données si les proportions de génotypes seront identiques à la génération suivante ? Ceci suppose que les proportions ne dépendent pas du sexe et que l'homme et la femme donnent un allèle au hasard et de façon indépendante.

On note  $\hat{p}_0$ ,  $\hat{q}_0$  et  $\hat{r}_0$  les fréquences d'obtention d'individus de type respectif AA, Aa et aa dans un échantillon de taille N.

$\hat{p}_0$ ,  $\hat{q}_0$  et  $\hat{r}_0$  étant des estimateurs de  $p_0$ ,  $q_0$  et  $r_0$  on détermine  $p_1$ ,  $q_1$  et  $r_1$  en utilisant les formules trouvées dans la partie II. :

$$p_1 = \left( \hat{p}_0 + \frac{1}{2} \hat{q}_0 \right)^2, \quad r_1 = \left( \hat{r}_0 + \frac{1}{2} \hat{q}_0 \right)^2 \quad \text{et} \quad q_1 = 1 - p_1 - r_1.$$

Sous hypothèse de stabilité  $p_0$ ,  $q_0$  et  $r_0$  doivent être égaux respectivement à  $p_1$ ,  $q_1$  et  $r_1$  et la quantité  $d^2 = \frac{(\hat{p}_0 - p_1)^2}{p_1} + \frac{(\hat{q}_0 - q_1)^2}{q_1} + \frac{(\hat{r}_0 - r_1)^2}{r_1}$  doit être faible.

On démontre que  $d^2 = \frac{(\hat{p}_0 \hat{r}_0 - \frac{1}{4} \hat{q}_0^2)^2}{(\hat{p}_0 + \frac{1}{2} \hat{q}_0)^2 (\hat{r}_0 + \frac{1}{2} \hat{q}_0)^2}$  (1) et que N  $d^2$  suit asymptotiquement

une loi du  $\chi^2$  à 1 degré de liberté. On peut alors par une table ou par simulation déterminer un seuil, par exemple 10% et décider de rejeter l'hypothèse de stabilité si N  $d^2$  dépasse ce seuil. Ici :

$$\hat{p}_0 = \frac{322}{939} \quad \hat{q}_0 = \frac{410}{939} \quad \hat{r}_0 = \frac{207}{939}$$

et on obtient :

$$N d^2 = 939 \frac{(322 \times 207 - \frac{1}{4} 410^2)^2}{(322 + \frac{1}{2} 410)^2 \times (207 + \frac{1}{2} 410)^2} \approx 12,08.$$

Par simulation<sup>3</sup> on obtient un neuvième décile de 2,7 environ et on peut ainsi rejeter l'hypothèse d'une stabilité des proportions avec un risque d'erreur inférieur à 10%.

<sup>3</sup> On trouvera un programme de simulation en ligne sur : <http://perso.wanadoo.fr/jpq/proba/hardy-weinberg/hwkh2.htm>

### Une autre façon d'envisager le problème

On sait que sous hypothèse de stabilité  $\frac{q_k^2}{p_k r_k} = 4$  pour tout  $k \neq 0$ , (II. 4.) et donc la valeur

de  $\Delta = \frac{\hat{q}_0^2}{\hat{p}_0 \hat{r}_0}$  doit être proche de 4 si N est grand.

La simulation<sup>4</sup> permet d'approcher sous cette hypothèse la distribution de cette variable aléatoire.

Si le risque choisi est  $r$  on définit deux valeurs  $a$  et  $b$  de telle façon que  $r/2$  valeurs de cette distribution soient inférieures à  $a$  et que  $r/2$  valeurs soient supérieures à  $b$ . Ainsi le rejet de l'hypothèse de situation stable dès que  $\Delta$  est à l'extérieur de l'intervalle  $[a; b]$  se fera avec un risque d'erreur inférieur à  $r$ . Si par exemple 5% des valeurs simulées sont inférieures à  $a$  et si 5% des valeurs simulées sont supérieures à  $b$  alors le risque de rejet à tort sera de 10%. Le test utilisé ici est encore appelé test bilatéral.

### Loi de Hardy-Weinberg et test d'indépendance

Dire que les proportions des génotypes sont constantes de générations en générations revient à supposer que ces proportions sont identiques chez l'homme et la femme, ceux-ci donnant lors de la conception un allèle au hasard et de façon indépendante. Les données du paragraphe précédent doivent être compatibles avec le modèle suivant :

Le père donne l'allèle La mère donne l'allèle	A	a
	A	$\frac{1}{2}q$
a	$\frac{1}{2}q$	$r$

où  $p, q, r$  sont les proportions des génotypes AA, Aa, aa.

Sous cette hypothèse le tableau des données pourra être représenté sous la forme :

Le père donne l'allèle La mère donne l'allèle	A	a	totaux
	A	322	205
a	205	207	412
totaux	527	412	939

De façon plus générale si les effectifs sont représentés par le tableau suivant :

<sup>4</sup> Simulation disponible sur  
<http://perso.wanadoo.fr/jpq/proba/hardy-weinberg/hw.htm> et  
<http://perso.wanadoo.fr/jpq/proba/hardy-weinberg/hw1.htm>

Le père donne l'allèle	A	a	totaux
La mère donne l'allèle			
A	a	b	n
A	c	d	N-n
totaux	m	N-m	N

Avec  $b = c$  et donc  $m = n$  alors on démontre que  $z = N \frac{(ad - bc)^2}{nm(N-n)(N-m)}$  suit

approximativement pour N assez grand une loi du khi2 à 1 degré de liberté (voir document d'accompagnement – Annexe, probabilités et statistique Séries ES et S page 148 et 149).

Or on peut remarquer que z est égal à la quantité  $N d^2$  du paragraphe précédent et que par conséquent le premier test rencontré dans ce document n'est rien d'autre qu'un test d'indépendance.

**Démonstration de l'égalité (1)**  $d^2 = \frac{(\hat{p}_0 \hat{r}_0 - \frac{1}{4} \hat{q}_0^2)^2}{(\hat{p}_0 + \frac{1}{2} \hat{q}_0)^2 (\hat{r}_0 + \frac{1}{2} \hat{q}_0)^2}$  :

$d^2$  est définie par l'égalité  $d^2 = \frac{(\hat{p}_0 - p_1)^2}{p_1} + \frac{(\hat{q}_0 - q_1)^2}{q_1} + \frac{(\hat{r}_0 - r_1)^2}{r_1}$ .

Si on note  $\alpha$  la différence  $\hat{p}_0 - \hat{r}_0$  alors d'après le **II**. du TD :

$\alpha = \hat{p}_0 - \hat{r}_0 = p_1 - r_1$ . On a donc  $\hat{p}_0 - p_1 = \hat{r}_0 - r_1$  et par suite  $\hat{q}_0 - q_1 = -2(\hat{p}_0 - p_1)$ . Ainsi  $d^2$  s'écrit encore :

$$\begin{aligned} d^2 &= (\hat{p}_0 - p_1)^2 \left( \frac{1}{p_1} + \frac{4}{q_1} + \frac{1}{r_1} \right) \\ &= (\hat{p}_0 - p_1)^2 \left( \frac{q_1 r_1 + 4 p_1 r_1 + p_1 q_1}{p_1 q_1 r_1} \right) \\ &= (\hat{p}_0 - p_1)^2 \left( \frac{q_1 (1 - q_1) + 4 p_1 r_1}{p_1 q_1 r_1} \right) \\ &= (\hat{p}_0 - p_1)^2 \left( \frac{1}{p_1 r_1} \right) \end{aligned}$$

car  $q_1^2 = 4 p_1 r_1$ .

Or  $p_1 = (\hat{p}_0 + \frac{1}{2} \hat{q}_0)^2$  et  $r_1 = (\hat{r}_0 + \frac{1}{2} \hat{q}_0)^2$ . Par suite  $\hat{p}_0 - p_1 = \hat{p}_0 - \hat{p}_0^2 - \hat{p}_0 \hat{q}_0 - \frac{1}{4} \hat{q}_0^2$

et on a donc bien

$$\frac{(\hat{p}_0 - p_1)^2}{p_1 r_1} = \frac{(\hat{p}_0 \hat{r}_0 - \frac{1}{4} \hat{q}_0^2)^2}{(\hat{p}_0 + \frac{1}{2} \hat{q}_0)^2 (\hat{r}_0 + \frac{1}{2} \hat{q}_0)^2}$$

# ANNEXE 1 : PETIT GLOSSAIRE DE GENETIQUE<sup>5</sup>

**ADN :**

Acide désoxyribonucléique. Macromolécule biologique constitutive des chromosomes qui est le support moléculaire de l'information génétique et de l'hérédité.

**Allèle :**

Variante, état d'un gène. Tout gène peut avoir plusieurs allèles (au moins deux) qui déterminent souvent l'apparition de caractères héréditaires différents.

**Caractère :**

Particularité physique ou biologique d'un organisme commandée par un ou plusieurs gènes. L'ensemble des caractères constitue le phénotype.

**Chromosome :**

Au moment des divisions cellulaires, l'ADN se compacte sur lui-même. Les morceaux d'ADN du génome deviennent visibles au microscope : ce sont les chromosomes, qui correspondent donc à un état provisoire de l'ADN. Chaque chromosome possède un bras court appelé p et un bras long appelé q, reliés au niveau du centromère.

**Gène :**

Élément du patrimoine génétique correspondant à une portion d'ADN sur un chromosome et participant à l'apparition d'un caractère.

**Hérédité :**

Transmission des caractères génétiques d'une génération aux suivantes.

**Hétérozygote :**

Présence de deux allèles différents (un allèle dominant et un allèle récessif ou deux allèles codominants) pour un gène donné sur les chromosomes homologues.

**Homozygote :**

Présence de deux allèles identiques pour un gène donné sur les chromosomes homologues.

**Locus :**

Emplacement d'un gène sur un chromosome. On le désigne par le n° du chromosome puis sa position sur le bras chromosomique.

**Phénotype :**

Ensemble des caractères observables résultant de l'expression du génotype.

**Récessif (allèle) :**

Version d'un gène qui ne s'exprime pas en présence de l'allèle dominant. Cet allèle ne s'exprime qu'à l'état homozygote.

---

<sup>5</sup> Pour plus de précision on peut consulter les sites :

<http://www.science-generation.com/biotlexi.php>

<http://gerard.chevrier.m2osw.com/gcartable/genetique/genlexique.htm>