A propos du Travail Dirigé sur la Loi de Hardy-Weinberg

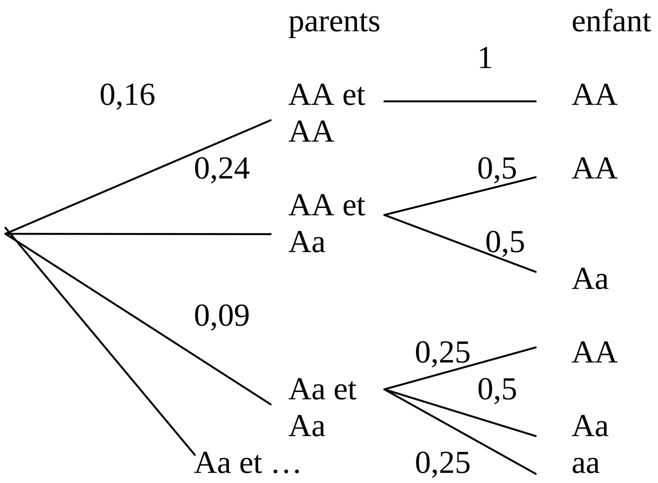
Corrigé :

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.

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 $2 \times 0{,}4$ x $0{,}3 = 0{,}24$
- 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 :

$p_1 = 0{,}16 \times 1 + 0{,}24 \times 0{,}5 + 0{,}09 \times 0{,}25$
$\quad = 0{,}3025.$

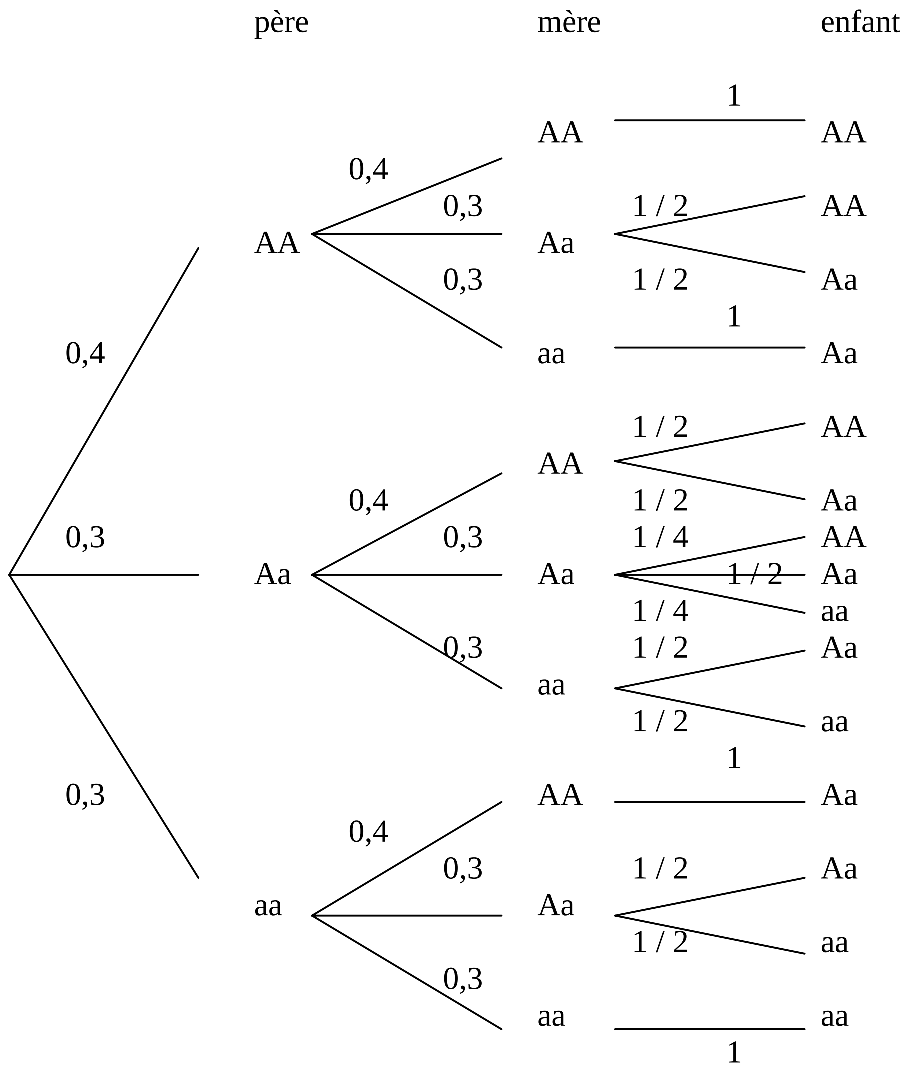
b. A l’aide d’un arbre du même type :

$r_1 = 0{,}09 \times 1 + 0{,}18 \times 0{,}5 + 0{,}09 \times 0{,}25 = 0{,}2025.$

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 $p_1 = \left(p_0 + \dfrac{1}{2}q_0\right)^2$ 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 : $r_1 = \left(r_0 + \dfrac{1}{2}q_0\right)^2$.

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

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

b. On peut encore remarquer que p_0 et r_0 jouent un rôle symétrique et que $r_0 - p_0 = -\alpha$. Ainsi : $r_1 = \dfrac{(1-\alpha)^2}{4}$ 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 $p_1 - r_1 = \dfrac{(1+\alpha)^2}{4} - \dfrac{(1-\alpha)^2}{4} = \alpha$, 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 :

$\dfrac{\hat{q}_1{}^2}{\hat{p}_1\hat{r}_1} \approx \dfrac{\hat{q}_2{}^2}{\hat{p}_2\hat{r}_2}$, 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 :

$\dfrac{q_1{}^2}{p_1r_1} = \dfrac{\left(\dfrac{1-\alpha^2}{2}\right)^2}{\dfrac{(1+\alpha)^2}{4}\dfrac{(1-\alpha)^2}{4}} = 4$. On aurait dû demander d'abord de calculer $\dfrac{q_1{}^2}{p_1r_1}$ 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^{e} siècle par Hardy, mathématicien et Weinberg, médecin. Cet équilibre des probabilités dès la première génération suppose réalisées les conditions suivantes : les proportions des génotypes sont les mêmes chez les hommes et les femmes, 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 un proportion $p_0 + \dfrac{1}{2}q_0$ d'allèles A et une proportion $r_0 + \dfrac{1}{2}q_0$ 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 χ^2.

On suppose qu'on a analysé 193 génotypes d'un échantillon d'une classe d'âge donnée. Le tableau suivant résume les résultats obtenus :

AA	Aa	aa	Total
64	82	47	193

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.** :

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

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 $d^2 = \dfrac{(\hat{p}_0\hat{r}_0 - \frac{1}{4}\hat{q}_0{}^2)^2}{(\hat{p}_0 + \frac{1}{2}\hat{q}_0)^2(\hat{r}_0 + \frac{1}{2}\hat{q}_0)^2}$ (1) et que N d^2 suit asymptotiquement une loi du χ^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 :

$\hat{p}_0 = \frac{64}{193}$, $\hat{q}_0 = \frac{82}{193}$, $\hat{r}_0 = \frac{47}{193}$ et on obtient :

$$\mathrm{N}\, d^2 = 193\frac{(64\times 47 - \frac{1}{4}82^2)^2}{(64 + \frac{1}{2}82)^2 \times (47 + \frac{1}{2}82)^2} \approx 3{,}98\,.$$

Par simulation (http://perso.wanadoo.fr/jpq/proba/hardy-weinberg/hwkhi2.htm) 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%.

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, (**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. Ici $\Delta \approx 2{,}235$.

La simulation présente à http://perso.wanadoo.fr/jpq/proba/hardy-weinberg/hw.htm ou à http://perso.wanadoo.fr/jpq/proba/hardy-weinberg/hw1.htm permet de simuler sous cette hypothèse la distribution de cette variable aléatoire.

Si le risque choisi est r alors on définit deux valeurs a et b de telle façon que $\frac{r}{2}\text{N}$ valeurs de cette distribution soient inférieures à a et que $\frac{r}{2}\text{N}$ valeurs soient supérieures à b. Ainsi le rejet de l'hypothèse de situation stable dès que Δ 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	p	$\frac{1}{2}q$
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	527
A	205	207	412
Totaux	527	412	939

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

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

avec $b = c$ et donc $m = n$.

On démontre que $z = \text{N}\frac{(ad-bc)^2}{n\,m(\text{N}-n)(\text{N}-m)}$ suit approximativement pour N assez grand une loi du χ^2 à 1 degré de liberté (exemple similaire dans le 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) $d^2 = \dfrac{(\hat{p}_0\hat{r}_0 - \frac{1}{4}\hat{q}_0^{\,2})^2}{(\hat{p}_0 + \frac{1}{2}\hat{q}_0)^2(\hat{r}_0 + \frac{1}{2}\hat{q}_0)^2}$ **:**

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

Si on note α 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 :

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

car $q_1^{\,2} = 4p_1 r_1$.

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

CQFD