Bulletin de l'APMEP n°350 - Décembre 1985

quelques modèles probabilistes pour l'étude de l'hérédité

(une réflexion interdisciplinaire à propos des terminales scientifiques)

par N. Roussignol (professeur de mathématiques)
D. Slansky (professeur de biologie)
Lycée Marcellin-Berthelot, Saint-Maur

N.D.L.R. - A la demande expresse des auteurs nous publions ce texte tel que l'a publié le bulletin de l'A.P.B.G. à l'intention des naturalistes. Pour un mathématicien qui modélise une situation expérimentale, il convient d'expliciter les hypothèses qui fondent le modèle puis les conclusions qu'on peut en tirer.

Ainsi au § B (monohybridisme) on suppose l'équiprobabilité sur l'espace produit Ω et on en déduit l'indépendance des allèles transmis par le père et la mère. Si au § C (dihybridisme) on suppose l'indépendance de la transmission des allèles a et b eux-mêmes équiprobables, on en déduit l'équiprobabilité des couples.

A. Pourquoi ce travail

Il a débuté à la suite d'une discussion que nous avons eue en novembre 1983.

Prof. de biologie : "Nos élèves ont eu l'air dépassés lorsque j'ai utilisé un produit de probabilités ; pourriez-vous leur préciser de quoi il s'agit ?".

Prof. de mathématiques : "Ce n'est pas surprenant car il n'y a plus de probabilités en première et nous n'en avons pas encore fait en Terminale. De plus, avec le nouveau programme de TC, nous n'avons plus à parler d'indépendance et c'est de cela qu'il s'agit. Par contre les TD ont encore l'indépendance probabiliste au programme."

Nous avons alors travaillé le sujet ensemble et l'étude suivante (une partie de B et C) a été exposée en cours de mathématiques aux élèves à la suite du chapitre, fort réduit maintenant, de probabilités. Cela a procuré un exemple parmi d'autres pour introduire la définition théorique d'indépendance :

deux événements A et B sont indépendants si p(A et B) = p(A) × p(B).

B. Le monohybridisme

1) Les génotypes

Un être vivant à $2n$ chromosomes (homme, animal, plante) reçoit de chacun de ses parents un gène. Ce gène peut se présenter sous deux états appelés les allèles du gène. Un être vivant est dit de race pure pour un gène s'il a reçu de ses parents le même allèle de ce gène.

Supposons qu'un gène ait deux allèles a et a'.

Supposons que l'on opère un croisement d'un individu de race pure pour l'allèle a, et d'un individu de sexe opposé de race pure pour l'allèle a'. La première génération F_1 issue d'un tel couple sera constituée des deux allèles du gène car chacun reçoit un allèle a d'un parent et un allèle a' de l'autre parent. Les individus de cette génération sont appelés hybrides. Ils ont chacun un allèle a et un allèle a'. En agriculture, ils sont très utilisés pour avoir des variétés possèdant des qualités intéressantes. Chez les grainetiers, on trouve des hybrides variés (choux, tomates, carottes,...).

La deuxième génération F_2 obtenue en croisant des individus de F_1 sera plus variée. Nous supposons que le père transmet de façon équiprobable son allèle a ou son allèle a' :

$$p(\text{père donne a}) = \frac{1}{2}\ ;\ p(\text{père donne a}') = \frac{1}{2}.$$

De même nous supposons :

$$p(\text{mère donne a}) = \frac{1}{2}\ ;\ p(\text{mère donne a}') = \frac{1}{2}$$

Lorsque un individu de cette deuxième génération reçoit :

a de sa mère et a de son père ; nous le notons $\frac{a}{a}$;*

a de sa mère et a' de son père ; nous le notons $\frac{a}{a'}$;

a' de sa mère et a de son père ; nous le notons $\frac{a'}{a}$;

a' de sa mère et a' de son père ; nous le notons $\frac{a'}{a'}$.

Si nous supposons que cette répartition se fait de façon complètement aléatoire, nous dirons que ces quatre possibilités ont la même probabilité de se produire.

Ensemble des possibilités : $\Omega = \{\frac{a}{a}, \frac{a}{a'}, \frac{a'}{a}, \frac{a'}{a'}\}$

* Cette notation n'est pas toujours celle utilisée en sciences naturelles et peut-être pas toujours avec le même sens. On pourrait aussi bien noter avec des couples (a,a), (a',a), (a,a'), (a',a') mais ce serait moins commode pour le dihybridisme. On note en haut l'allèle de la mère et en bas celle du père.

Probabilité de chaque possibilité :

$p(\frac{a}{a}) = \frac{1}{4}$; $p(\frac{a}{a'}) = \frac{1}{4}$; $p(\frac{a'}{a}) = \frac{1}{4}$; $p(\frac{a'}{a'}) = \frac{1}{4}$.

Remarquons que :

p("père donne a" et "mère donne a") $= p(\frac{a}{a}) = \frac{1}{4}$

$=$ p(père donne a) $\times$ p(mère donne a)

De même on montrerait que :

p("père donne a" et "mère donne a'") =
p(père donne a) $\times$ p(mère donne a'). etc.

2) Les phénotypes

Lorsque l'on observe les individus de F_2, on ne voit pas leurs gènes mais un caractère visible du à ces gènes.

1ère hypothèse : a domine a', c'est-à-dire que les individus $\frac{a}{a'}$ et $\frac{a'}{a}$ ont la même apparence que les individus $\frac{a}{a}$. Nous noterons [a] leur apparence, appelé phénotype.

Univers des apparences possibles : $\Omega' = \{[a], [a']\}$

Probabilité de chaque apparence :

$$p([a]) = p(\frac{a}{a}) + p(\frac{a}{a'}) + p(\frac{a'}{a}) = \frac{3}{4}$$

$$p([a']) = p(\frac{a'}{a'}) = \frac{1}{4}$$

Une expérience de Mendel relevant de ce modèle :

Parents de race pure : pois à graines lisses $\times$ pois à graines ridées
Génération F_1 : toutes les graines sont lisses
Génération F_2 : 5474 graines lisses et 1850 graines ridées soit environ 74,7 % de graines lisses et 25,3 % de graines ridées.

2ème hypothèse : il n'y a pas de dominance de a sur a' ou de a' sur a.

Les individus $\frac{a}{a'}$ et $\frac{a'}{a}$ ont une apparence intermédiaire notée [aa'].

Univers des apparences possibles : $\Omega' = \{[a], [aa'], [a']\}$

Probabilité de chaque apparence :

$$p([a]) = p(\frac{a}{a}) = \frac{1}{4}$$

$$p([aa']) = p(\frac{a}{a'}) + p(\frac{a'}{a}) = \frac{1}{2}$$

$$p([a']) = p(\frac{a'}{a'}) = \frac{1}{4}$$

Une expérience relevant de ce modèle :

Parents de race pure : belles de nuit à fleurs rouges × belles de nuit à fleurs blanches.
Génération F_1 : toutes les fleurs sont roses.
Génération F_2 : 72 plantes à fleurs rouges, 145 plantes à fleurs roses et 70 plantes à fleurs blanches soit environ 21,1 % à fleurs rouges, 50,5 % à fleurs roses et 24,4 % à fleurs blanches.

Conclusion : De nombreux êtres vivants ont des caractères pouvant être expliqués par ces modèles. Cependant les bactéries ont un "chromosome" formé d'une seule molécule d'A.D.N. circulaire. De nombreux végétaux (certaines algues, certains champignons, les mousses) n'ont que *n* chromosomes.

Il existe encore d'autres types de modèles mais nous n'en parlerons pas dans le cadre du programme de Terminale.

C. Le dihybridisme

1) Les génotypes

On s'intéresse maintenant non plus à un, mais à deux caractères, portés par deux gènes distincts, l'un d'allèle a et a′, l'autre d'allèles b et b′, ces deux gènes étant supposés vérifier séparément les hypothèses faites au B.

On croise un individu de race pure pour les allèles a et b avec un individu de sexe opposé de race pure pour les allèles a′ et b′. On croise $(\frac{a}{a}, \frac{b}{b})$ avec $(\frac{a'}{a'}, \frac{b'}{b'})$. La première génération F_1 est constituée entièrement d'individus disposant des deux allèles de chaque gène, a et a′ d'une part, b et b′ d'autre part. La génération F_2 est obtenue en croisant des individus de F_1. L'ensemble des possibilités est :

$$\Omega = \{(\frac{a}{a}, \frac{b}{b}), (\frac{a}{a}, \frac{b}{b'}), (\frac{a}{a}, \frac{b'}{b}), (\frac{a}{a}, \frac{b'}{b'}), (\frac{a}{a'}, \frac{b}{b}), \ldots\}$$

Il y a donc 4×4, c'est-à-dire 16 génotypes possibles à cette génération. Rappelons que lorsque l'on note $\frac{a}{a'}$, a est fourni par la mère et a′ par le père.

Nous supposons maintenant que la transmission des allèles a ou a′ par les parents n'a pas d'influence sur la transmission des allèles b ou b′. Cette hypothèse rajoutée à celles faites au B nous permet de dire que les 16 génotypes ont la même probabilité de se produire, soit $\frac{1}{16}$.

Exemple : $p(\frac{a}{a}, \frac{b}{b}) = \frac{1}{16}$, $p(\frac{a}{a'}, \frac{b}{b'}) = \frac{1}{16}$, etc.

Remarquons alors que : $p(\frac{a}{a} \text{ et } \frac{b}{b'}) = \frac{1}{16} = \frac{1}{4}\,\frac{1}{4} = p(\frac{a}{a}) \times p(\frac{b}{b'})$.

On montrerait des formules analogues pour les autres génotypes.

Au paragraphe B comme ci-dessus nous avons abouti à des formules du genre

$$p(E \text{ et } F) = p(E) \times p(F)$$

où E et F représentent deux événements. Ces formules résultent à chaque fois de la non influence de E sur F et de F sur E. Des situations expérimentales dans divers domaines ont abouti à ce genre de formule et c'est pourquoi, en théorie des probabilités, on a la définition suivante :
"Deux événements E et F sont indépendants si $p(E \text{ et } F) = p(E) \times p(F)$".*

2) Les phénotypes

L'indépendance des génotypes nous laisse penser qu'il y a aussi indépendance au niveau des phénotypes. C'est ce que nous allons voir plus loin.

1ère hypothèse : a domine a' et b domine b'

Le tableau suivant va nous permettre de compter les possibilités génétiques aboutissant à une même apparence (ou phénotype). Dans une direction est indiqué le patrimoine héréditaire relatif à un gène, dans l'autre celui relatif à l'autre gène. Nous signalons aux enseignants de biologie que ce tableau n'est pas le même que celui rencontré dans des cas analogues dans les ouvrages de biologie.

On lit le phénotype dans le tableau.

	$\frac{b}{b}$	$\frac{b}{b'}$	$\frac{b'}{b}$	$\frac{b'}{b'}$
$\frac{a}{a}$	[a,b]	[a,b]	[a,b]	[a,b']
$\frac{a}{a'}$	[a,b]	[a,b]	[a,b]	[a,b']
$\frac{a'}{a}$	[a,b]	[a,b]	[a,b]	[a,b']
$\frac{a'}{a'}$	[a',b]	[a',b]	[a',b]	[a',b']

* Il y a une autre définition de l'indépendance, mais nous n'en parlons pas ici.

On trouve : $p([a,b]) = \frac{9}{16} = 0{,}563$ car on obtient [a,b] pour 9 génotypes sur 16.

$$p([a,b']) = \frac{3}{16} = 0{,}188$$

$$p([a',b]) = \frac{3}{16} = 0{,}188$$

$$p([a',b']) = \frac{1}{16} = 0{,}063$$

Il y a quatre phénotypes ici.

Remarque : Si l'on se réfère au monohybridisme pour chaque caractère considéré seul, on voit que :

$$p([a]) \times p([b]) = \frac{3}{4} \times \frac{3}{4} = \frac{9}{16} = p([a,b])$$

$$p([a]) \times p([b']) = \frac{3}{4} \times \frac{1}{4} = \frac{3}{16} = p([a,b'])$$

$$p([a']) \times p([b]) = \frac{1}{4} \times \frac{3}{4} = \frac{3}{16} = p([a',b])$$

$$p([a']) \times p([b']) = \frac{1}{4} \times \frac{1}{4} = \frac{1}{16} = p([a',b'])$$

Il y a donc bien indépendance probabiliste au niveau des phénotypes.

Autre remarques :

$$p([a,.]) = p([a,b]) + p([a,b']) = \frac{9}{16} + \frac{3}{16} = \frac{12}{16} = \frac{3}{4}$$

$$p([a',.]) = p([a',b]) + p([a',b']) = \frac{3}{16} + \frac{1}{16} = \frac{4}{16} = \frac{1}{4}$$

On retrouve les probabilités des phénotypes a et a' lorsqu'on étudie un seul caractère. Ce genre de calcul est utile dans les exercices pour se ramener à un seul caractère.

Une expérience relevant de ce modèle :

Parents de race pure pour chaque caractère :
cobayes noirs à poils courts × cobayes blancs à poils longs.

Génération F_1 : les hydrides sont tous noirs à poils courts

Génération F_2 : 1500 cobayes noirs à poils courts soit 57,0 %
486 cobayes noirs à poils longs soit 18,5 %
483 cobayes blancs à poils courts soit 18,4 %
162 cobayes blancs à poils longs soit 6,2 %

2ème hypothèse : a domine a' et il n'y a pas de dominance entre b et b'

Tableau donnant les phénotypes avec les mêmes conventions que ci-dessus.

	$\frac{b}{b}$	$\frac{b}{b'}$	$\frac{b'}{b}$	$\frac{b'}{b'}$
$\frac{a}{a}$	[a,b]	[a,bb']	[a,bb']	[a,b']
$\frac{a}{a'}$	[a,b]	[a,bb']	[a,bb']	[a,b']
$\frac{a'}{a}$	[a,b]	[a,bb']	[a,bb']	[a,b']
$\frac{a'}{a'}$	[a',b]	[a',bb']	[a',bb']	[a',b']

$p([a,b]) = \frac{3}{16} = 0{,}188$ d'autre part $p([a]) \times p([b]) = \frac{3}{4} \times \frac{1}{4}$

$p([a,bb']) = \frac{6}{16} = 0{,}375$ d'autre part $p([a] \times p([bb']) = \frac{3}{4} \times \frac{2}{4}$

$p([a,b']) = \frac{3}{16} = 0{,}188$ d'autre part $p([a] \times p([b']) = \frac{3}{4} \times \frac{1}{4}$

$p([a',b]) = \frac{1}{16} = 0{,}063$ d'autre part $p([a']) \times p([b]) = \frac{1}{4} \times \frac{1}{4}$

$p([a',bb']) = \frac{2}{16} = 0{,}125$ d'autre part $p([a']) \times p([bb']) = \frac{1}{4} \times \frac{2}{4}$

$p([a',b']) = \frac{1}{16} = 0{,}063$ d'autre part $p([a']) \times p([b']) = \frac{1}{4} \times \frac{1}{4}$

On retrouve la même propriété qu'avec la première hypothèse.

Remarquons qu'ici il y a six phénotypes.

Remarquons là encore que :

$p([a,.]) = \frac{3}{16} + \frac{6}{16} + \frac{3}{16} = \frac{3}{4}$; $p([.,b]) = \frac{3}{16} + \frac{1}{16} = \frac{1}{4}$

$p([a',.]) = \frac{1}{16} + \frac{2}{16} + \frac{1}{16} = \frac{1}{4}$; $p([.,bb']) = \frac{6}{16} + \frac{2}{16} = \frac{1}{2}$

$p([.,b']) = \frac{3}{16} + \frac{1}{16} = \frac{1}{4}$

Nous retrouvons séparément les proportions de chaque phénotype pour chacun des caractères. C'est un élément pour reconnaître dans quelle situation on est.

Une expérience illustrant cette situation : avec une variété de mûflier.

Parents : plantes à feuilles glabres et à fleurs rouges × plantes à feuilles velues et à fleurs bleues.

Génération F_1 : tous les hybrides sont à feuilles *velues* et à fleurs *mauves*

Génération F_2 : 305 plantes à feuilles velues et fleurs rouges soit 18,9 %
607 plantes à feuilles velues et fleurs mauves soit 37,6 %
303 plantes à feuilles velues et fleurs bleues soit 18,8 %
98 plantes à feuilles glabres et fleurs rouges soit 6,1 %
200 plantes à feuilles glabres et fleurs mauves soit 12,4 %
103 plantes à feuilles glabres et fleurs bleues soit 6,3 %

3ème hypothèse : ***il n'y a pas de dominance entre a et a' et il n'y a pas de dominance entre b et b'***

Tableau des phénotypes. Il y a cette fois-ci neuf phénotypes.

	$\frac{b}{b}$	$\frac{b}{b'}$	$\frac{b'}{b}$	$\frac{b'}{b'}$
$\frac{a}{a}$	[a,b]	[a,bb']	[a,bb']	[a,b']
$\frac{a}{a'}$	[aa',b]	[aa',bb']	[aa',bb']	[aa',b']
$\frac{a'}{a}$	[aa',b]	[aa',bb']	[aa',bb']	[aa',b']
$\frac{a'}{a'}$	[a',b]	[a',bb']	[a',bb']	[a',b']

$p([a,b]) = \frac{1}{16} = 0{,}063$ et $p([a]) \times p([b]) = \frac{1}{4} \times \frac{1}{4}$

$p([a,bb']) = \frac{2}{16} = 0{,}125$ et $p([a]) \times p([bb']) = \frac{1}{4} \times \frac{2}{4}$

$p([a,b']) = \frac{1}{16} = 0{,}063$ et $p([a]) \times p([b']) = \frac{1}{4} \times \frac{1}{4}$

$p([aa',b]) = \frac{2}{16} = 0{,}125$ et $p([aa']) \times p([b]) = \frac{2}{4} \times \frac{1}{4}$

$p([aa',bb']) = \frac{4}{16} = 0{,}250$ et etc...

$p([aa',b']) = \frac{2}{16} = 0{,}125$

$p([a',b]) = \frac{1}{16} = 0{,}063$

$p([a',bb']) = \frac{2}{16} = 0{,}125$

$p([a',b']) = \frac{1}{16} = 0{,}063$

On retrouve toujours la même propriété.

Et on peut, là encore, calculer $p([a,.])$, $p([aa',.])$, $p([a',.])$, $p([.,b])$, $p([.,bb'])$, $p([.,b'])$.

Une expérience conforme à ce modèle : une sorte de muflier.

Parents : plantes à fleurs rouges et feuilles larges × plantes à fleurs blanches et feuilles étroites.

Génération F_1 : tous les hybrides sont à fleurs *roses* et feuilles *moyennes*.

Génération F_2 : 185 plantes à fleurs rouges et feuilles larges soit 6,2 %
374 plantes à fleurs rouges et feuilles moyennes soit 12,5 %
187 plantes à fleurs rouges et feuilles étroites soit 6,3 %
374 plantes à fleurs roses et feuilles larges soit 12,5 %
744 plantes à fleurs roses et feuilles moyennes soit 24,9 %
376 plantes à fleurs roses et feuilles étroites soit 12,6 %
186 plantes à fleurs blanches et feuilles larges soit 6,2 %
375 plantes à fleurs blanches et feuilles moyennes soit 12,5 %
188 plantes à fleurs blanches et feuilles étroites soit 6,3 %.

Remarque pour les mathématiciens :

Tous les calculs développés à partir des tableaux relatifs aux phénotypes aboutissent à l'indépendance probabiliste des phénotypes. Ils ne sont en fait qu'une application du théorème :

"Soit X et Y deux variables aléatoires indépendantes et f et g deux fonctions définies sur ces variables, alors f(X) et g(Y) sont également des variables aléatoires indépendantes."

D. Une réflexion sur ce qui précède

Les modèles ci-dessus ne permettent pas de décrire tous les phénomènes d'hérédité. Après avoir décrit les cas simples, l'enseignant de biologie présente certaines expériences où aucun des modèles ci-dessus ne convient.

Les raisons physiologiques qui expliquent un processus différent peuvent être de plusieurs types.

Il se peut que les deux gènes soit portés par le même chromosome. On serait tenté de croire que dans ce cas les deux gènes sont totalement liés, mais en fait ils ne le sont que partiellement car il se produit au cours de la meiose des crossing-over (recombinaison de morceaux de chromosomes) en proportion constante pour un même gène, mais variable suivant les gènes.

Il se peut encore que plusieurs gènes portés par des chromosomes différents agissent sur un caractère (par exemple plusieurs gènes peuvent donner la couleur noire à une drosophyle).

Si, en analysant les résultats d'une expérience de dihybridisme on peut montrer que les proportions de chaque type d'individu de la génération F_2 vérifient un des modèles ci-dessus, on pourra conclure que l'on n'est pas dans ces cas singuliers.

Les trois lois dites de Mendel signifient exactement que des gènes vérifient les hypothèses au B et au C . Ce sont les cas simples qui ont été historiquement les premiers étudiés.

E. En guise de conclusion

Prof. de biologie : "C'est beau tout cela, mais on perd de vue que notre démarche est expérimentale. Nous commençons par présenter les expériences, puis nous donnons une explication à l'aide des chromosomes."

Prof. de mathématiques : "Bien sûr, mais lorsque vous voulez faire une étude qualitative vous êtes obligés de tenir compte des connaissances mathématiques des élèves. S'ils ont déjà réfléchi à la notion d'indépendance et connaissent les calculs correspondants, vous n'avez qu'à faire comprendre les phénomènes biologiques qui la traduisent et dérouler des calculs rapides. Par contre s'ils n'en ont jamais entendu parler, vous êtes obligée de détailler avec les tableaux."

Prof. de biologie : "Nous en reparlerons à la prochaine rentrée."

Prof. de mathématiques : "Si la notion d'indépendance était remise au programme de mathématiques en TC, vous pourriez en biologie exploiter au mieux les connaissances que les élèves doivent acquérir au cours de génétique, à condition que le cours de mathématique ait été fait avant."

Bibliographie

— *Contremanuel de statistiques et probabilités* par M. PELTIER, N. ROUCHE, M. MANDERNICK. Editions Vie Ouvrière.

— *Basic concepts of probability and statistics* par J.L. HODGES, E.L. LEHMANN. Chap. 4. Holden-Day.

— *L'enseignement des probabilités et de la statistique* par A. ENGEL. Vol. 2, chap. 7. CEDIC.

— *Mathématiques du hasard dans les lycées.* Revue "Recherches pédagogiques" INRP n° 104, p. 31 à 42.

— IREM de Grenoble. Groupe proba. Nov. 77, n° 2, p. 31.