

MODELISATION ITINERANTE : DU SAGUENAY A NOS CLASSES

Nathalie Erard, Catherine Fauvé, Carole Hébert et Mireille Vuong

**Master de didactique des mathématiques
Module de modélisation – Mai 2009
Université Paris VII**

SOMMAIRE

I.	<u>NOTRE DEMARCHE</u>	page 2
II.	<u>BREF HISTORIQUE</u>	page 3
III.	<u>DESCRIPTION RAPIDE DE L'ATAXIE SPASTIQUE</u>	page 7
IV.	<u>QUELQUES EXEMPLES DE MODELISATION</u>	page 9
	1) MODELE DE HARDY – WEINBERG	page 9
	2) STABILISATION GENETIQUE AVEC LA THEORIE DES GRAPHS	page 14
	3) MODELISATION DE DERIVE GENETIQUE	page 20
V.	<u>ANNEXES</u>	page 32
VI.	<u>ACTIVITES EN CLASSE</u>	page 39
	1) POUR LA CLASSE DE 3 ^{ème}	page 41
	2) POUR LA CLASSE DE TERMINALE S	page 51
	3) POUR LA CLASSE DE TERMINALE ES	page 54
VII.	<u>CONCLUSION</u>	page 66
VIII.	<u>REMERCIEMENTS</u>	page 67
IX.	<u>REFERENCES/ BIBLIOGRAPHIE</u>	page 68

I. NOTRE DEMARCHE

Nous sommes un groupe de 4 professeurs de mathématiques, deux travaillant en collège et deux en lycée.

Nous étions intéressées par une problématique en lien avec la biologie. Lors de recherches internet à partir du site du collège de France, nous avons été interpellées par une étude faite sur une population spécifique de la région du Saguenay Lac- Saint-Jean au Québec (CNRS info, mars 1999, « Comportements démographiques et maladies héréditaires rares », Frédéric Austerlitz, Evelyne Heyer).



Cette population, constituée des descendants de pionniers du XVII^{ème} siècle comporte une fréquence très élevée de certaines maladies héréditaires très rares ou inexistantes dans les autres populations d'origine européenne.

Nous avons pris contact avec Evelyne Heyer (professeur au Muséum d'histoire naturelle, directeur adjoint du département Hommes, Natures, Sociétés et dirige une équipe de recherche en génétique sur des populations humaines au musée de l'Homme) qui a accepté de nous envoyer ses publications concernant le taux de porteurs et de malades atteints d'ataxie spastique.

Nous avons voulu comparer les résultats obtenus avec différents types de modélisation.

A partir de ce moment-là, nous avons été amenées à nous intéresser à l'histoire de la génétique.

II. BREF HISTORIQUE

Toute théorie nouvelle est une rupture dans la conception du monde. Il en fut ainsi de la théorie darwinienne de l'évolution et de la théorie mendélienne de l'hérédité. La génétique des populations est ce qui a permis la synthèse entre ces deux théories.

En 1859, Charles Darwin, médecin et naturaliste anglais, dévoilait au monde ses découvertes dans l'ouvrage « Origine des espèces par la sélection naturelle ».

Avant Charles Darwin, l'évolution était considérée comme un processus individuel d'adaptation sous les contraintes de l'environnement.

Il fut le premier des naturalistes à concevoir la sélection naturelle comme une force active de l'évolution des espèces.

Au cours d'un grand voyage, en 1832, en Uruguay, Charles Darwin découvrit des petits rongeurs aveugles qui vivaient sous terre comme des taupes : le tuco-tuco (animal qui semble avoir disparu de nos jours).

Darwin en conclut que c'était là une manifestation concrète de la théorie de l'évolution : la rétractation graduelle d'organes par défaut d'utilisation.

De même, en 1835 aux Galápagos, il trouva des pinsons aux becs dissemblables, qui après analyses, s'avérèrent appartenir à la même famille.

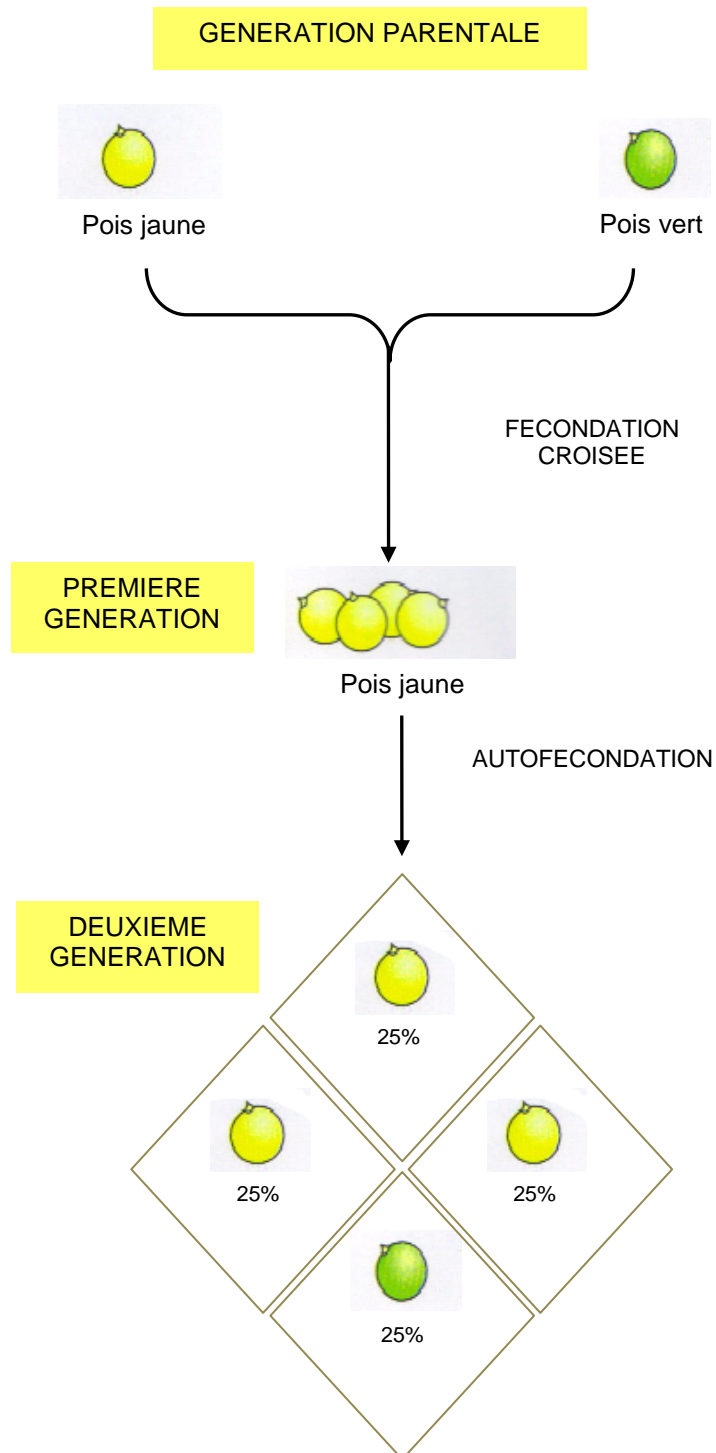
Il en conclut que les êtres vivants évoluent non seulement à travers les siècles, mais aussi en fonction de leur environnement.

Pour Darwin, l'évolution est un processus de tri par sélection naturelle, au sein d'une population entre les individus qui sont le mieux adaptés à ces contraintes et qui, de fait, laissent le plus de descendants.

Cette fécondité différentielle induit une modification de la diversité génétique des populations vers une augmentation de la fréquence des types les mieux adaptés et aboutit, après une longue période, à une transformation de l'espèce.

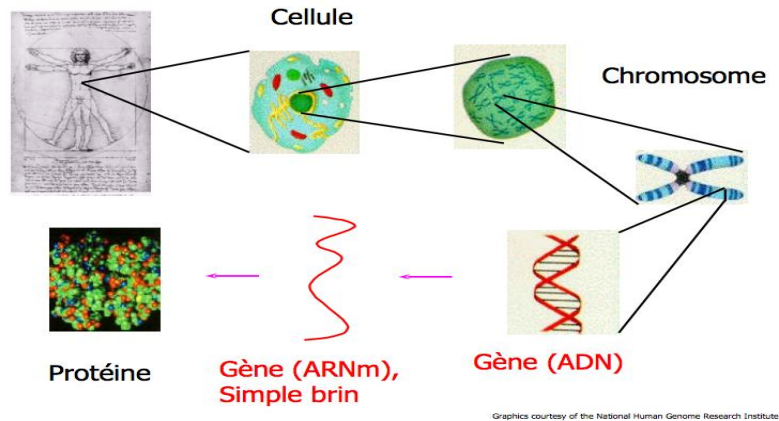
La survie du plus apte est donc le moteur de l'évolution.

En 1866, Gregor Mendel a déchiffré les lois de l'hérédité en étudiant la transmission d'une poignée de caractères distincts et indépendants chez les petits pois. Il a ainsi pu révéler la nature discontinue de l'hérédité.



Il faudra attendre le début du XX^{ème} siècle, avec les travaux de Thomas H. Morgan sur les drosophiles, pour établir la théorie chromosomique de l'hérédité.

A partir de 1941, la biologie moléculaire par l'étude approfondie de la structure de l'ADN a montré qu'un même gène peut exister sous plusieurs versions appelées allèles.

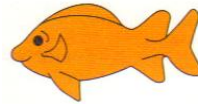
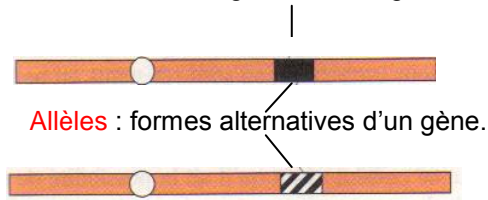


Le modèle ci-dessous illustre les différents états possibles d'un gène possédant deux allèles A et a.

Gène : unité fonctionnelle d'hérédité, correspondant habituellement à un segment d'ADN codant des gènes d'un organisme.

Génome : l'ensemble des gènes d'un organisme.

Locus : site d'un gène dans le génome.



Sauvage : type de gène retrouvé normalement.

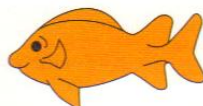
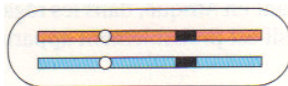


Mutant : différent du sauvage en raison d'un changement génétique (mutation).

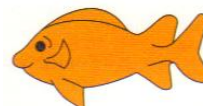
Génotype : l'ensemble d'allèles spécifiques formant le génome d'un individu

Phénotype : caractères visibles d'un individu

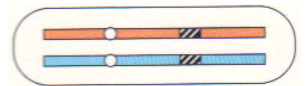
Homozygote A/A



Hétérozygote a/A



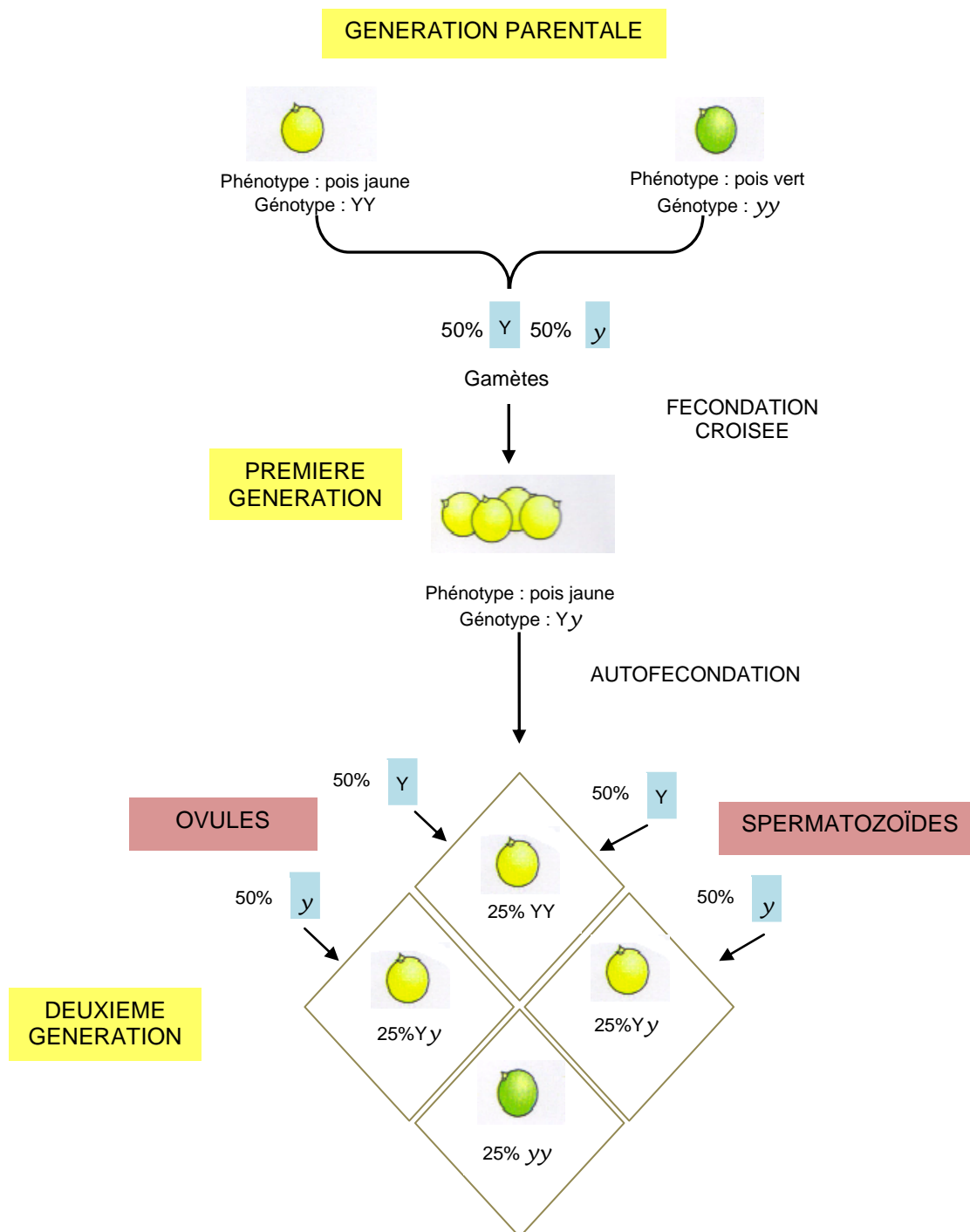
Homozygote a/a



L'allèle A est **dominant** (par rapport à a) ; l'allèle a est **récessif** (par rapport à A).

Dans l'exemple ci-dessus, le phénotype de l'hétérozygote a/A est le même que celui de l'homozygote A/A, même si leur capital génétique diffère.

C'est à la même époque que Hardy, physicien et probabiliste anglais et Weinberg, physiologiste allemand aboutissent indépendamment à la création d'un modèle. Celui-ci, appelé, modèle de Hardy - Weinberg permet de valider la théorie de Mendel, en la sortant du laboratoire « hors d'un milieu contrôlé par l'homme » et d'abandonner le concept d'hérédité par mélange des Darwiniens.



Le processus de production de cellules sexuelles appelées gamètes (ovocytes ou spermatozoïdes) se nomme la méiose. Les gamètes ne contiennent qu'un allèle de chaque gène, ce sont des cellules dites haploïdes, obtenues à partir de la redistribution aléatoire des chromosomes.

Lors de la fécondation, il y a reconstitution d'une cellule diploïde par la réunion de deux gamètes qui apportent chacune un exemplaire de chaque gène sous une forme allélique donnée.

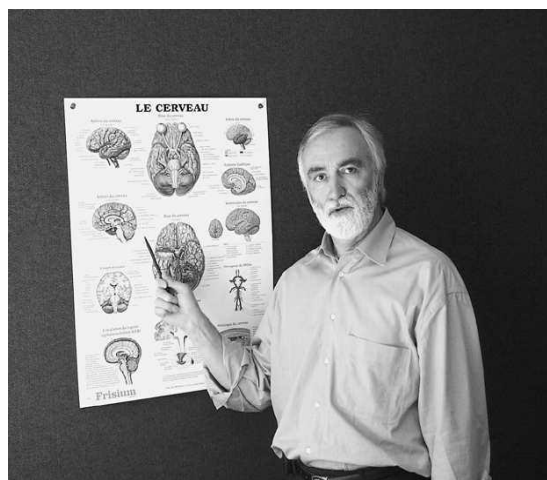
L'étude de ces processus a permis de fonder la génétique des populations dont l'un des objectifs était de quantifier et de modéliser l'évolution génétique des espèces à travers l'évolution de la fréquence des allèles des gènes, sous l'effet de sélection en faveur de certains allèles.

Pour revenir à notre sujet, nous cherchons à comprendre pour quelles raisons la fréquence des porteurs de l'ataxie spastique au Saguenay Lac - Saint - Jean, est passée en seulement douze générations de 1 sur 5000 à 1 sur 20.

III. DESCRIPTION RAPIDE DE L'ATAXIE SPASTIQUE

C'est vers les années 1970 que cette forme d'ataxie a été diagnostiquée par une équipe de chercheurs québécois dont faisait partie le professeur Bouchard.

Ces malades avaient des symptômes différents de ceux rencontrés jusque là et étaient tous originaires de Charlevoix ou du Saguenay : dès la petite enfance, l'ataxie provoque des problèmes de coordination, et l'apprentissage de la marche en est un peu retardée. Un manque d'équilibre entraîne des chutes fréquentes, une raideur est constatée dans les jambes ainsi que des problèmes de coordination dans les mouvements des bras, d'où des difficultés pour écrire, difficultés également d'élocution, faiblesse musculaire ensuite des bras et des jambes, mais les capacités intellectuelles ne sont pas affectées.



Très souvent vers 30 ans ces patients ont besoin d'une canne pour se déplacer et vers 50 ans ils sont en fauteuil roulant. L'espérance de vie est suivant les articles, identique aux personnes non malades ou légèrement inférieure.

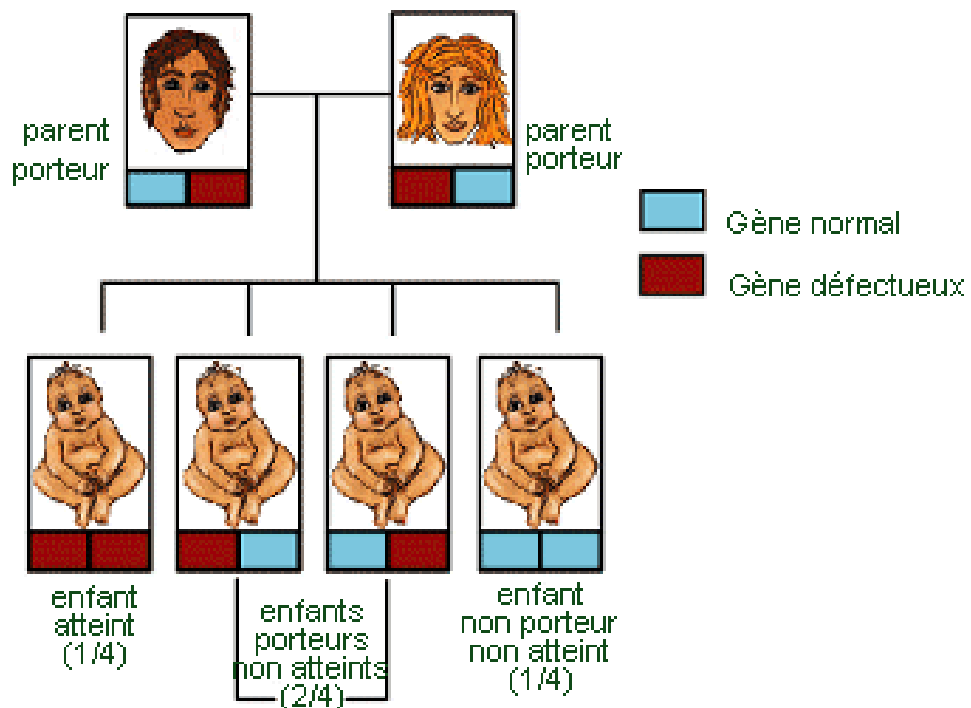
C'est en 2000 que le gène responsable de cette maladie a été découvert. Il est situé sur le 13^{ème} chromosome : « Le gène muté qui cause la maladie n'était pas connu chez l'humain, signale Jean-Pierre Bouchard. On connaissait son existence chez la souris et il semble s'être très bien conservé entre la souris et l'homme ».

Ce gène synthétise une protéine, la sachsine, nommée à partir de la contraction du nom anglais de la maladie, SACS (Spastic Ataxia of Charlevoix-Saguenay). La sachsine possède des caractéristiques particulières qui portent à croire qu'elle interviendrait dans le pliage des protéines, une opération qui détermine leur configuration tridimensionnelle. Les deux mutations identifiées jusqu'à présent dans ce gène produisent des protéines tronquées non fonctionnelles. On ignore si c'est la protéine elle-même ou son impact sur d'autres protéines qui cause la maladie, signale le chercheur.

Cette maladie est à caractère récessif, seule une personne possédant les deux allèles a est malade. Une personne qui ne possède qu'un seul allèle est dite « porteuse » mais non malade.

Les maladies récessives sont transmises par les deux parents. Lorsque deux parents sont porteurs d'une copie du gène défectueux (allèle a), les probabilités de transmettre cette maladie récessive sont les suivantes :

- Une probabilité sur quatre (25%) d'avoir un enfant atteint.
- Deux probabilités sur quatre (50%) d'avoir un enfant qui est porteur du gène mais qui n'est pas atteint de la maladie.
- Une probabilité sur quatre (25%) d'avoir un enfant qui n'est pas atteint de la maladie et qui n'est pas porteur du gène défectueux.



Dans ces régions 1 individu sur 22 est porteur du gène, et 200 personnes environ sont atteintes de la maladie, soit 1 personne sur 1000 environ.

IV. QUELQUES EXEMPLES DE MODELISATION

1) MODELE DE HARDY - WEINBERG

Ce modèle est le modèle de base de la génétique, il est très souvent vérifié au sein de populations naturelles.

Description du modèle :

a) En considérant les couples et le cycle vital de la reproduction :

Dans le cas d'un gène di-allélique, d'allèles A et a, la composition génétique d'une population d'organismes à sexe séparé est constituée des trois génotypes possibles :

A / A de fréquence u, A/a de fréquence v et a/a de fréquence w avec $u + v + w = 1$.

La fréquence p des allèles A est alors égale à $p = u + \frac{v}{2}$ et la fréquence q des allèles a vaut $q = w + \frac{v}{2}$ avec $p + q = 1$

On considérera que la fréquence des génotypes est identique pour chacun des deux sexes et que les couples se forment au hasard (condition 1 de panmixie).

Il a été démontré que, en une génération de panmixie, les fréquences entre les sexes sont égalisées même si elles diffèrent à la génération précédente.

On peut alors construire le tableau des fréquences des couples ou mariages possibles :

MALE	FEMELLE		
	AA de fréquence u	Aa de fréquence v	aa de fréquence w
AA de fréquence u	u^2	uv	uw
Aa de fréquence v	vu	v^2	vw
aa de fréquence w	wu	wv	w^2

Afin de pouvoir considérer que les fréquences des couples sont égales à leurs probabilités, on considérera que la population est de taille infinie pour pouvoir y appliquer la loi des grands nombres (condition 2).

On pourra alors étudier la probabilité des génotypes des descendants issus de ces différents mariages :

Type de mariages	Probabilité du mariage	Probabilité du génotype AA	Probabilité du génotype Aa	Probabilité du génotype aa
AA x AA	u^2	u^2	0	0
AAxAa ou AaxAA	$2 \times u \times v$	$u \times v$	$u \times v$	0
AAxaa ou aaxAA	$2 \times u \times w$	0	$2 \times u \times w$	0
AaxaA	v^2	$\frac{v^2}{4}$	$\frac{v^2}{2}$	$\frac{v^2}{4}$
Aaxaa ou aaxaA	$2 \times v \times w$	0	$v \times w$	$v \times w$
aaxaa	w^2	0	0	w^2

On considérera l'absence de mutation d'un allèle dans la gamétogenèse c'est-à-dire AA ne peut pas donner Aa (condition 3 du modèle de Hardy - Weinberg).

On considérera que les deux types de gamètes présents lors de la fécondation sont théoriquement dans des proportions égales.

Lors de la fécondation, ces proportions pourraient avoir été modifiées si un mécanisme de sélection gamétique avait joué en faveur de l'un.

On admet l'absence de sélection gamétique (condition 4 du modèle de Hardy - Weinberg).

Les proportions des génotypes chez les descendants lors de la fécondation ne sont maintenues au stade final de reproducteur adulte que **s'il y a absence de sélection zygotique lors de l'embryogenèse ou de la croissance (condition 5 du modèle de Hardy - Weinberg).**

Les trois génotypes sont donc supposés avoir la même espérance de vie et ne présentent, du moins pour le gène considéré, aucune mortalité différentielle.

Pour la génération suivante, ces conditions étant remplies, on calcule la probabilité de chaque génotype en sommant les probabilités de chaque colonne.

Pour ce faire, il faut d'abord considérer que les couples sont également fertiles et qu'il n'y a pas eu d'apport extérieur de population qui aurait pu de chance de présenter la même constitution génétique.

Ce sont les conditions d'absence de sélection en termes de fertilité différentielle et d'absence de migrations (condition 6 du modèle de Hardy - Weinberg).

On obtient alors :

$$\text{Fréquence AA à la génération 1 : } u_1 = u^2 + u \times v + \frac{v^2}{4} = \left(u + \frac{v}{2}\right)^2 = p^2$$

$$\text{Fréquence Aa à la génération 1 : } v_1 = u \times v + 2 \times u \times w + \frac{v^2}{2} + v \times w = 2 \left(u + \frac{v}{2}\right) \left(w + \frac{v}{2}\right) = 2pq$$

$$\text{Fréquence aa à la génération 1 : } w_1 = \frac{v^2}{4} + vw + w^2 = \left(\frac{v}{2} + w\right)^2 = q^2$$

La relation ainsi établie entre les fréquences alléliques et les fréquences génotypiques s'appellent « la relation de Hardy - Weinberg » ou la relation panmictique car elle découle directement de l'hypothèse de panmixie.

On calcule aussi la fréquence des deux allèles à la génération 1 :

$$\text{Fréquence de l'allèle A : } p_1 = u_1 + \frac{v_1}{2} = p^2 + p \times q = p(p + q) = p$$

$$\text{Fréquence de l'allèle a : } q_1 = w_1 + \frac{v_1}{2} = q^2 + p \times q = q(p + q) = q$$

On observe qu'à la génération 1, la fréquence des allèles n'a pas changé.

En dressant le même tableau pour passer de la génération 1 à la génération 2, on aurait :

$$\text{Fréquence AA à la génération 2 : } u_2 = p_1^2 = p^2 = u_1$$

$$\text{Fréquence Aa à la génération 2 : } v_2 = 2p_1q_1 = 2pq = v_1$$

$$\text{Fréquence aa à la génération 2 : } w_2 = q_1^2 = q^2 = w_1$$

Conclusion :

Les fréquences génotypiques et alléliques d'un gène autosomique di-alléliques sont rigoureusement invariantes d'une génération à l'autre.

Après une génération, les fréquences des différents types de gènes atteignent une répartition stationnaire. Cette stabilité s'appelle l'équilibre de Hardy - Weinberg.

Le modèle de Hardy - Weinberg, tel qu'il est défini, permet de conclure qu'une population panmictique, de grande taille, sans mutations, ni sélection, ni migration maintient son polymorphisme génétique en l'état.

L'évolution n'est alors possible que si des forces sélectives viennent influencer les caractères existants d'une population.

La relation de Hardy - Weinberg va permettre d'estimer les fréquences alléliques et génotypiques dans l'étude d'un gène présentant des allèles soit dominants ou récessifs. Ces calculs vont permettre de quantifier la diversité génétique en général et chez l'homme d'aborder les problèmes de diagnostic, de dépistage ou de santé publique.

Le développement des modèles de génétique des populations consistera la plupart du temps à partir de ce modèle de Hardy - Weinberg, pour voir comment la constitution génétique d'une population peut être modifiée, vers quelle limite et à quelle vitesse, si telle ou telle des conditions de cet équilibre n'est pas respectée.

Rappel des 6 conditions du modèle de Hardy - Weinberg :

1. La panmixie c'est à dire que les couples reproducteurs se forment au hasard.
2. Effectif infini de la population : afin de pouvoir considérer que les fréquences des couples ou des génotypes sont égales à leurs probabilités respectives, du fait de la loi des grands nombres.
3. Il n'y a pas de mutation lors de la production des gamètes. Des parents A1/A1 et A1/A1 ne pourront pas donner des descendants A1/A2.
4. Absence de sélection gamétique lors de la fécondation. Les deux types de gamètes formés par des hétérozygotes restent dans les proportions 50/50.
5. Absence de sélection zygotique : Les trois génotypes sont supposés avoir la même espérance de vie et ne présentent, du moins pour le gène considéré, aucune mortalité différentielle.
6. Pour calculer les fréquences des génotypes chez les adultes reproducteurs de la génération suivante, il faut considérer qu'il y a absence de sélection en terme de fertilité différentielle de ces couples et aussi absence de migrations (ni entrant, ni sortant).

Dans la population du lac de Saguenay Saint Jean, on observe un changement de fréquences ce qui montre que cette population s'écarte du modèle de Hardy - Weinberg.

On peut en conclure qu'au moins une des conditions n'est pas remplie, en particulier l'effectif infini de la population.

b) En considérant une urne gamétique :

Ceci permet alors de considérer non plus les couples reproducteurs mais une urne gamétique.

Tout se passe comme si les couples mettaient en vrac leurs gamètes dans une urne de spermatozoïdes pour les males et dans une urne d'ovules pour les femelles et que tout descendant est issu de l'union de deux gamètes tirées au hasard dans chacune des deux urnes.

Cette situation est une réalité biologique pour les espèces végétales et de nombreuses espèces animales aquatiques.

Si les fréquences alléliques sont les mêmes dans les deux sexes, les urnes sont identiques.

Les 6 conditions énoncées ci-dessus doivent être respectées ce qui conduit à la même conclusion :

Si la fréquence de l'allèle A, notée $f(A)$, est égale à p dans chaque urne et la fréquence de l'allèle a, notée $f(a)$, est égale à q dans chaque urne, alors chez les descendants :

$$f(A/A) = p^2, f(A/a) = 2pq \text{ et } f(a/a) = q^2.$$

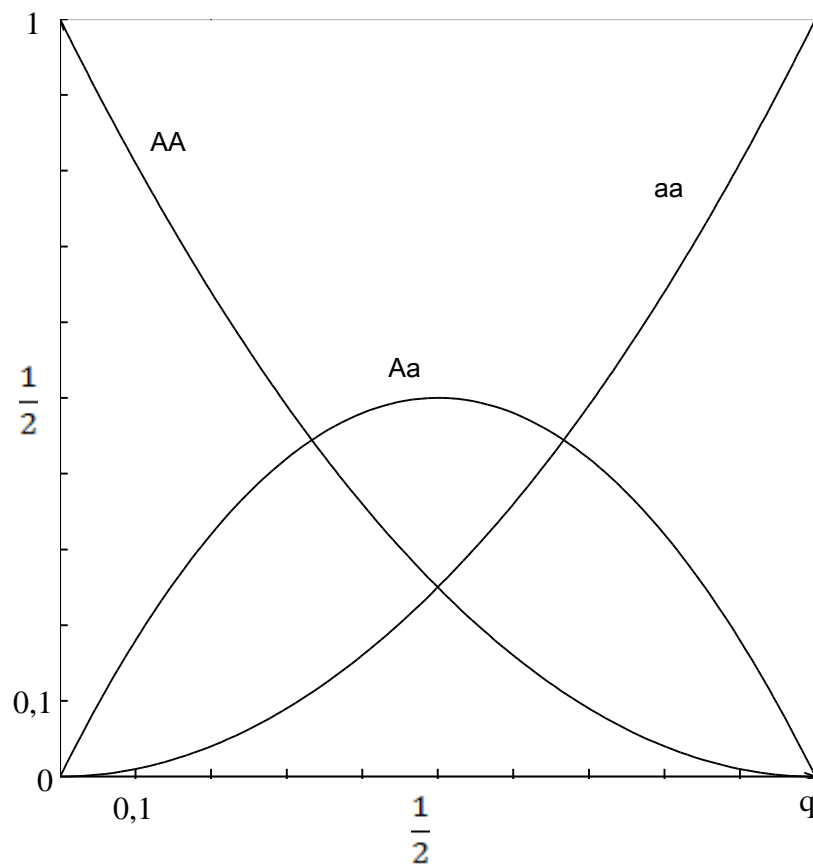
On retrouve ici les valeurs trouvées au dessus.

Le principe de l'urne gamétique est très souvent utilisé dans les modèles en génétique.

c) Conséquence :

En régime panmictique, la fréquence des hétérozygotes est souvent supérieure à celle des homozygotes.

En effet, on peut tracer les courbes des fréquences des différents génotypes en fonction de la valeur de q (fréquence de l'allèle a), portée en abscisse. La fréquence du génotype AA correspond à la courbe d'équation $y = (1-q)^2$, celle du génotype aa à la courbe d'équation $y = q^2$, celle du génotype Aa à la courbe d'équation $y = 2q(1-q)$.



Prenons q , la fréquence de l'allèle a égale à 0,1, le nombre d'homozygotes aa est très faible.

L'allèle a est surtout présent dans les hétérozygotes Aa .

Quand un allèle est très rare (q ou $p = 1 - q$ très proche de zéro), la plupart des exemplaires de ces allèles sont présents dans les individus hétérozygotes et très peu dans les homozygotes.

Cette situation n'a pas les mêmes conséquences selon que l'allèle est récessif ou dominant.

L'effet de l'allèle récessif n'est perceptible que chez les homozygotes aa et sera donc rare, par contre si l'allèle est dominant, son effet se rencontrera aussi chez les hétérozygotes.

La fréquence des homozygotes quels qu'ils soient sera même le plus souvent considérée comme négligeable au regard de la fréquence des hétérozygotes.

Par ailleurs, la fréquence maximale du génotype des hétérozygotes est égale à 50 %, ce qui correspond à la situation où les deux fréquences alléliques sont égales.

d) Test statistique de vérification et généralisation de ce modèle :

Un test statistique utilisant la distribution d'une variable de χ^2 permet de valider l'hypothèse qu'une population donnée répond effectivement aux conditions d'équilibre de Hardy - Weinberg.

Grâce à la relation de Hardy - Weinberg, à partir des fréquences phénotypiques observées, on calculera les fréquences alléliques ainsi que les fréquences des porteurs sains d'une maladie (hétérozygotes en cas de gènes récessifs).

En annexe, on trouvera des exemples concrets sur l'albinisme (gène récessif), sur des maladies à gènes dominants ainsi qu'un calcul portant sur les groupes sanguins A, B et O où les gènes A et B sont dominants et le gène O récessif.

Le modèle de Hardy - Weinberg a été généralisé à un gène pluri-allélique et à de nombreuses autres situations, ce qui en fait le modèle de référence.

Il existe plusieurs écarts au modèle de Hardy - Weinberg indiqués en annexe.

2) STABILISATION GENETIQUE AVEC LA THEORIE DES GRAPHS

Nous avons ensuite étudié l'évolution génétique de cette population, avec la théorie des graphes. Dans cette théorie, la population n'a pas besoin d'être très grande, il suffit qu'elle reste de taille identique d'une génération à la génération suivante.

a) Cas simplifié :

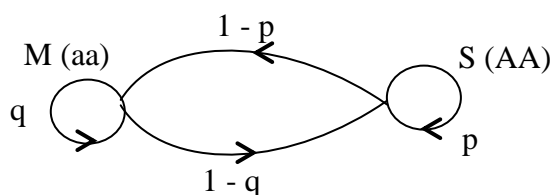
Il n'existe que deux états malade M et sain S.

Soit p la probabilité que d'une génération à l'autre, une personne saine reste saine et donc $1 - p$ la probabilité qu'elle devienne malade.

De même, soit q la probabilité que d'une génération à la suivante, une personne reste malade et donc $1 - q$ la probabilité qu'elle passe de l'état malade à l'état sain.

L'état à la génération $n + 1$ dépend de l'état à la génération n et de ces probabilités p et q , donc on a une chaîne de Markov.

Schéma (que l'on appelle graphe) :



Soit $(a_0; b_0)$ l'état initial, c'est-à-dire que a_0 est la fréquence de gens malades, au départ et b_0 la fréquence de personnes saines (on a $a_0 + b_0 = 1$).

A la génération 1 on aura :

$$\begin{aligned} a_1 &= (1-p)b_0 + qa_0 \\ b_1 &= pb_0 + (1-q)a_0 \end{aligned}$$

Avec des matrices on aura :

$$\begin{pmatrix} a_1 & b_1 \end{pmatrix} = \begin{pmatrix} a_0 & b_0 \end{pmatrix} \begin{bmatrix} q & 1-q \\ 1-p & p \end{bmatrix}$$

Cette dernière matrice est la matrice de passage de l'état de la génération n à la génération suivante $n + 1$.

Montrons qu'il y a toujours un état stable c'est-à-dire au bout d'un certain temps les pourcentages des états malade et sain ne changeront plus.

Si P est la matrice de passage, soit E l'état qui serait stable, donc on devrait avoir :

$E = E \cdot P$, ce qui équivaut à $f(E) = E$ si f est la fonction associée à la matrice de passage, ou $f(E) = \lambda E$, avec $\lambda = 1$, donc $\lambda = 1$ doit être une valeur propre.

Pour trouver les valeurs propres, on cherche les racines du polynôme caractéristique (c'est-à-dire le déterminant de la matrice de $P - \lambda I$, où I est la matrice identité).

$$P - \lambda I = \begin{bmatrix} q - \lambda & 1 - q \\ 1 - p & p - \lambda \end{bmatrix} \text{ et le polynôme caractéristique est :}$$

$$\det(P - \lambda I) = (q - \lambda)(p - \lambda) - (1 - q)(1 - p) = \lambda^2 - \lambda(p + q) - 1 + p + q$$

On vérifie que si $\lambda = 1$ on obtient bien 0, donc 1 est une valeur propre et il existe bien un état stable.

Cet état stable ne dépend pas des pourcentages initiaux d'individus malades ou sains, mais uniquement des valeurs de p et q .

Cherchons cet état stable :

Soit $(a \ b)$ l'état stable.

Si on le multiplie par la matrice de passage, on doit retrouver $(a \ b)$.

$$\begin{bmatrix} a & b \end{bmatrix} = \begin{bmatrix} a & b \end{bmatrix} \times \begin{bmatrix} q & 1-q \\ 1-p & p \end{bmatrix}$$

$$\text{soit } \begin{cases} a = a(q) + b(1-p) \\ b = a(1-q) + b(p) \end{cases} \Leftrightarrow \begin{cases} 0 = a(q-1) + b(1-p) \\ 0 = a(1-q) + b(p-1) \end{cases} \quad \text{or, } a + b = 1 \text{ d'où } \begin{cases} a = \frac{1-p}{2-p-q} \\ b = \frac{1-q}{2-p-q} \end{cases}$$

Conclusion :

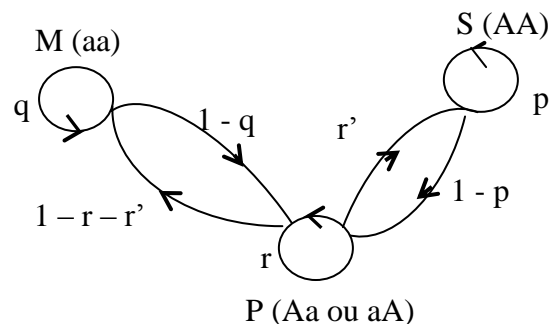
Lorsqu'il n'y a que deux états, Malade ou Sain, il y a un état stable qui ne dépend pas de l'état initial, mais uniquement des probabilités de passage d'un état à l'autre, d'une génération à la suivante ;

Par exemple, dans deux populations dont les proportions initiales de personnes malades et saines sont totalement différentes : population A avec 5% de malades (M) et 95% de sains (S), puis une population B avec 30% de malades (M) et 70 % de sains (S), on obtiendra un même état stable si les pourcentages de passages $M \rightarrow S$ et $S \rightarrow M$ d'une génération à la suivante sont les mêmes dans les populations A et B !

b) Cas d'un gène récessif : l'ataxie spastique

On considère 3 états : malade (M, aa), sain (S, AA) et porteur (P, aA ou Aa).

On construit le graphe :



On remarque que d'une génération à la suivante on ne peut pas passer de Sain à Malade, ni de Malade à Sain, donc les probabilités qui correspondent sur les arêtes sont nulles (on ne trace donc pas ces arêtes).

Soit p la probabilité de rester sain, donc $1 - p$ la probabilité de passer de Sain à Porteur.

Soit q la probabilité de rester Malade, donc $1 - q$ la probabilité de passer de Malade à Porteur.

Soit r la probabilité de rester porteur, soit r' la probabilité de passer de Porteur à Sain.

Donc $1 - r - r'$ est la probabilité de passer de Porteur à Malade.

On forme les suites m_n, s_n, p_n qui indiquent les pourcentages de personnes respectivement malades, saines, porteuses, à la génération n .

On a :

$$\begin{aligned} m_1 &= qm_0 + (1-r-r')p_0 \\ s_1 &= ps_0 + r'p_0 \\ p_1 &= rp_0 + (1-p)s_0 + (1-q)m_0 \end{aligned} \quad \text{d'où} \quad \begin{bmatrix} m_1 & s_1 & p_1 \end{bmatrix} = \begin{bmatrix} m_0 & s_0 & p_0 \end{bmatrix} \times \begin{bmatrix} q & 0 & 1-q \\ 0 & p & 1-p \\ 1-r-r' & r' & r \end{bmatrix}$$

Cherchons si $\lambda = 1$ est une valeur propre de cette matrice de passage :

$$\text{Det} (M - \lambda I) = (q - \lambda)[(p - \lambda)(r - \lambda) - r'(1 - p)] + (1 - r - r')[-(p - \lambda)(1 - q)].$$

$$\text{Si } \lambda = 1, \text{ on obtient } (q - 1)(pr - p - r + 1 - r' + r'p) + (1 - r - r')(-p + pq + 1 - q)$$

En développant on obtient 0, donc il existe bien un état stable.

Cherchons cet état stable :

$$[a \quad b \quad c] = [a \quad b \quad c] \times M \text{ d'où le système}$$

$$\begin{cases} a = aq + c(1 - r - r') \\ b = bp + cr' \\ c = a(1 - q) + b(1 - p) + cr \\ a + b + c = 1 \end{cases}$$

D'où :

$$\begin{cases} a + b + c = 1 \\ a(q - 1) + c(1 - r - r') = 0 \\ b = \frac{cr'}{1 - p} \\ a(1 - q) + b(1 - p) + c(r - 1) = 0 \end{cases} \quad \begin{cases} a + b + c = 1 \\ a = \frac{1 - r - r'}{1 - q} c \\ b = \frac{cr'}{1 - p} \\ a(1 - q) + b(1 - p) + c(r - 1) = 0 \end{cases}$$

avec les trois premières équations on obtient :

$$\frac{1 - r - r'}{1 - q} c + \frac{r'}{1 - p} c + c = 1 \Leftrightarrow (1 - p)(1 - r - r')c + (1 - q)r'c + (1 - p)(1 - q)c = (1 - p)(1 - q)$$

D'où

$$c = \frac{(1 - p)(1 - q)}{(1 - p)(1 - r - r') + (1 - q)r' + (1 - p)(1 - q)}$$

$$a = \frac{(1 - p)(1 - r - r')}{(1 - p)(1 - r - r') + (1 - q)r' + (1 - p)(1 - q)}$$

$$b = \frac{(1 - q)r'}{(1 - p)(1 - r - r') + (1 - q)r' + (1 - p)(1 - q)}$$

Quelques simulations d'états stables sur Excel :

p	q	r	r'	a	b	c
0,95	0,8	0,5	0,3	0,125	0,75	0,125
0,9	0,7	0,5	0,3	0,14285714	0,64285714	0,21428571
0,9	0,6	0,5	0,3	0,11111111	0,66666667	0,22222222
0,9	0,6	0,7	0,3	3,4694E-17	0,75	0,25
0,8	0,6	0,7	0,2	0,11111111	0,44444444	0,44444444

(Choisir r et r' tels que r + r' inférieur à 1 pour que 1 - r - r' soit une probabilité).
Nous allons chercher à adapter cette modélisation aux fréquences actuelles de l'Ataxie spastique de Saguenay lac- Saint-Jean.

Pour cela nous devons calculer les poids des arêtes du graphe.

Sachant que 1 personne sur 1000 est atteinte, et 1 sur 22 est porteuse, la fréquence des personnes saines est $u = 0,954$, celle des porteurs est $0,045$ et celle des malades est $0,001$.

Considérons le tableau des probabilités de mariages, suivant :

	AA $u=0.954$	Aa $v=0.045$	aa $w=0.001$
AA $u=0.954$	u^2	uv	uw
Aa $v=0.045$	uv	v^2	vw
aa $w=0.001$	uw	vw	w^2

Poids de l'arête de M vers P :

Pour qu'une personne M ait son descendant P, il faut qu'elle rencontre soit un P, soit un S.

La probabilité qu'elle rencontre un P est vw , mais dans ce cas, il y a 1 chance sur 2 pour que le descendant soit porteur et 1 chance sur 2 pour qu'il soit malade d'où $\frac{vw}{2}$.

La probabilité qu'elle rencontre une personne saine est uw et dans ce cas le descendant est porteur.

Le poids de l'arête de M vers P est donc $uw + \frac{vw}{2} \approx 9.765 \times 10^{-4}$.

Poids de l'arête de M vers M est donc $1 - 9.765 \times 10^{-4} = 0.9990235$.

Poids de l'arête de S vers P :

Pour que le descendant d'une personne saine soit porteur, il faut que cette personne saine rencontre ou un porteur (probabilité uv) et dans ce cas il y a une chance sur deux pour que le descendant soit porteur, ou une personne malade (probabilité uw) et dans ce cas le descendant est porteur.

Le poids de cette arête est donc $\frac{uv}{2} + uw \approx 0.022419$.

Le poids de l'arête de S vers S est donc $1 - 0.022419 = 0.977581$

Poids de l'arête de P vers S :

Pour qu'une personne porteuse ait un descendant sain, il faut qu'elle rencontre une personne saine (probabilité uv) il y a alors une chance sur deux pour que le descendant soit sain, ou une personne porteuse (probabilité v^2).

Il y a alors une chance sur 4 pour que le descendant soit sain, d'où le poids de cette

$$\text{arête : } \frac{uv}{2} + \frac{v^2}{4} \approx 0,02197125 .$$

Poids de l'arête de P vers M: de même on montre que le poids est

$$\frac{v^2}{4} + \frac{vw}{2} \approx 5.2875 \times 10^{-4} .$$

Le poids de la boucle de P vers P est alors 0.9775.

Ce qui permet de compléter le tableau des simulations précédent par la dernière ligne :

p	q	r	r'	a	b	c
0,95	0,8	0,5	0,3	0,125	0,75	0,125
0,9	0,7	0,5	0,3	0,14285714	0,64285714	0,21428571
0,9	0,6	0,5	0,3	0,11111111	0,66666667	0,22222222
0,9	0,6	0,7	0,3	3,4694E-17	0,75	0,25
0,8	0,6	0,7	0,2	0,11111111	0,44444444	0,44444444
0,977581	0,9990235	0,9775	0,02197125	0,21474284	0,38866826	0,3965889

L'état stable donnerait une fréquence de $a = 0,215$ de personnes malades, de $b = 0,389$ de personnes saines et de $c = 0,397$ de personnes porteuses.

Ceci correspond à une très grande augmentation de la fréquence des personnes malades ce qui, nous l'espérons ne correspond pas à la réalité !

Au bout de combien de générations cet état stable serait-il atteint ?

De proche en proche avec une calculatrice nous cherchons n pour obtenir :

$$[m_0 \quad s_0 \quad p_0] \times \begin{bmatrix} q & 0 & 1-q \\ 0 & p & 1-p \\ 1-r-r' & r' & r \end{bmatrix}^n = [a \quad b \quad c]$$

$$[0.001 \quad 0.954 \quad 0.045] \times \begin{bmatrix} 0.9990235 & 0 & 9.765 \times 10^{-4} \\ 0 & 0.977581 & 0.022419 \\ 5.2875 \times 10^{-4} & 0.02197125 & 0.9775 \end{bmatrix}^n = [0.215 \quad 0.387 \quad 0.396]$$

Avec $n = 500$

on obtient $[0.0965 \quad 0.4488 \quad 0.4545]$

Avec $n = 1000$

on obtient $[0.151 \quad 0.421 \quad 0.428]$

Avec $n = 2000$

on obtient $[0.1962 \quad 0.398 \quad 0.4056]$

Avec $n = 3000$

on obtient $[0.2093 \quad 0.3914 \quad 0.3992]$

Et avec $n = 3500$ on obtient $[0.2118 \quad 0.3901 \quad 0.398]$, ce qui est très proche de l'état stable.

Une génération étant d'environ 25 ans, cet état stable serait atteint au bout de 17 500 ans environ !

Conclusion :

Dans les hypothèses utilisées pour la théorie des graphes, nous avons supposé que la population globale composée de malades, de porteurs et de sain reste constante au cours des générations, ce qui n'est pas applicable dans la réalité. Il y a donc aussi modification des différents pourcentages de passage d'une catégorie à une autre. Dans cette maladie, la théorie des graphes ne reflète pas la réalité observée.

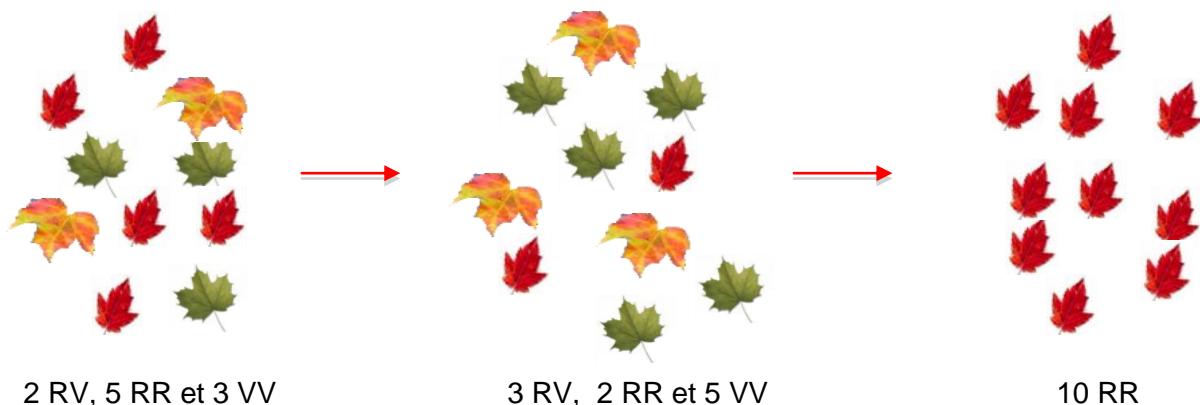
3) MODELISATION DE DERIVE GENETIQUE

L'hypothèse de population de taille infinie dans les conditions du modèle de Hardy- Weinberg peut dans certain cas être remise en cause. Ce modèle ne s'applique plus alors.

Deux causes peuvent provoquer une réduction de taille de population :

- une catastrophe (feu, chasse ou pêche intensive, inondation, épidémie). C'est ce qu'on appelle **l'effet d'étranglement** ;
- naissance d'une petite population par migration de quelques individus vers un nouvel habitat. C'est **l'effet fondateur** qui nous intéresse au Saguenay.

La limitation de l'effectif d'une population conduit à un phénomène appelé **dérive génétique** parce qu'il est caractérisé par une fluctuation aléatoire des fréquences alléliques d'une génération à l'autre. Ce phénomène conduit avec le temps à une réduction progressive de la diversité génétique au niveau allélique.



La fréquence des allèles R et V change aléatoirement à chaque génération, il en est de même pour la fréquence des génotypes RR, RV, VV.

1) Introduction

On considère l'ataxie spastique représentée par un gène possédant deux allèles *a* et *A*.

Soit une population de $\frac{N}{2}$ individus. Le gène de l'ataxie spastique est représenté par deux allèles ***a* et *A***. Tout se passe, dans cette population de taille $\frac{N}{2}$, comme si chaque individu mettait dans une urne leurs deux d'allèles : *aA* ou *AA* ou *aa*. La génération suivante est issue d'un tirage aléatoire, avec remise, de *N* allèles de cette urne.

Hypothèse 1 :

A chaque génération *t*, les $\frac{N}{2}$ individus c'est à dire leurs *N* allèles disparaissent et donnent naissance à *N* allèles ($\frac{N}{2}$ individus).

Hypothèse 2 :

Le tirage s'effectue aléatoirement, il n'y a donc pas de sélection.

Les fréquences des allèles ***a* et *A*** peuvent différer considérablement d'un tirage à un autre et diffèrent aussi de leur probabilité de tirage initiale selon le phénomène de dérive génétique.

Imaginons à la génération *t*, une urne contenant *N* allèles : ***A* et *a***.

On choisit dans cette urne **au hasard avec remise**, *N* allèles. On constitue alors la génération suivante *t* + 1.

On s'intéresse au nombre d'allèles ***a*** à la génération *t*.

On définit la variable aléatoire X_t comptant le nombre d'allèles ***a*** à la génération *t*.

X_t prend ses valeurs dans $\{0;1;2;...;N\}$.

Soit $X_t = i$.

A la génération suivante *t* + 1, la probabilité de choisir un allèle ***a*** est de $\frac{i}{N}$, noté $P(a / X_t = i) = \frac{i}{N}$

Le choix d'un allèle peut être considéré comme une épreuve de Bernoulli de paramètre $\frac{i}{N}$. La variable aléatoire X_{t+1} / X_t suit donc une loi binomiale de

paramètre $\left(\frac{i}{N}; N\right)$.

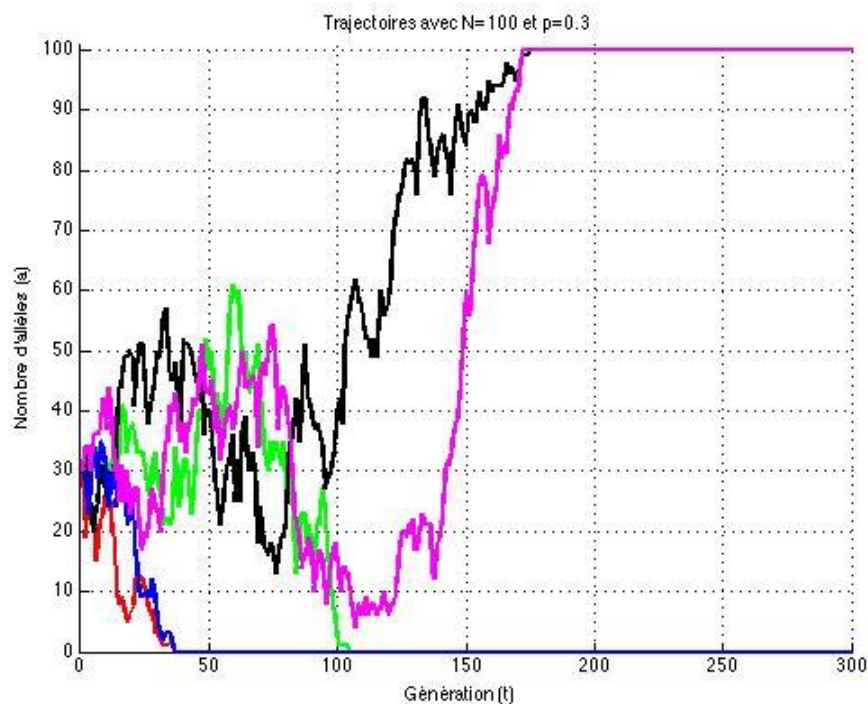
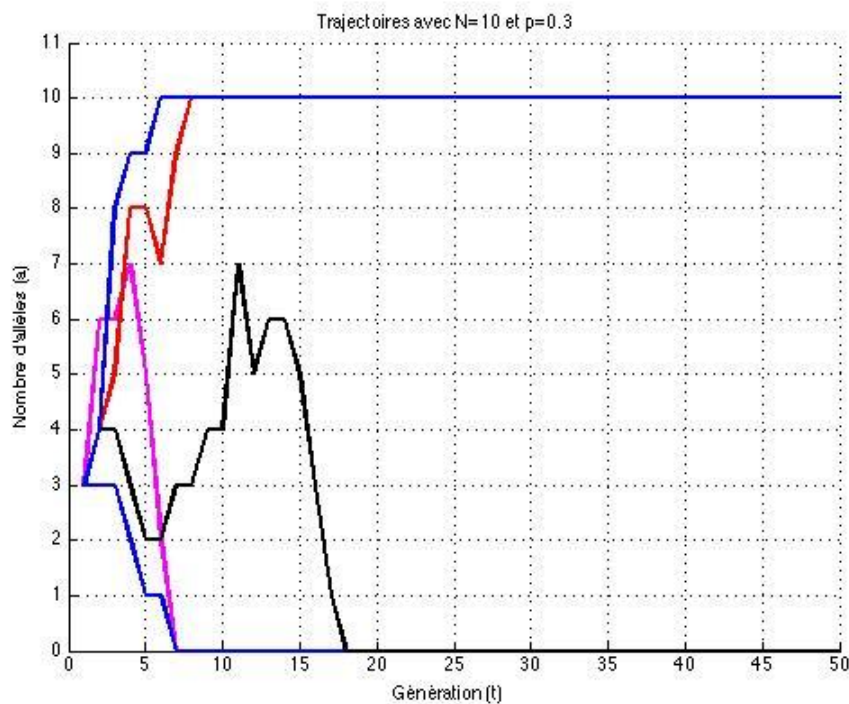
Donc $P(X_{t+1} = j / X_t = i) = \binom{N}{j} \left(\frac{i}{N}\right)^j \left(1 - \frac{i}{N}\right)^{N-j}$.

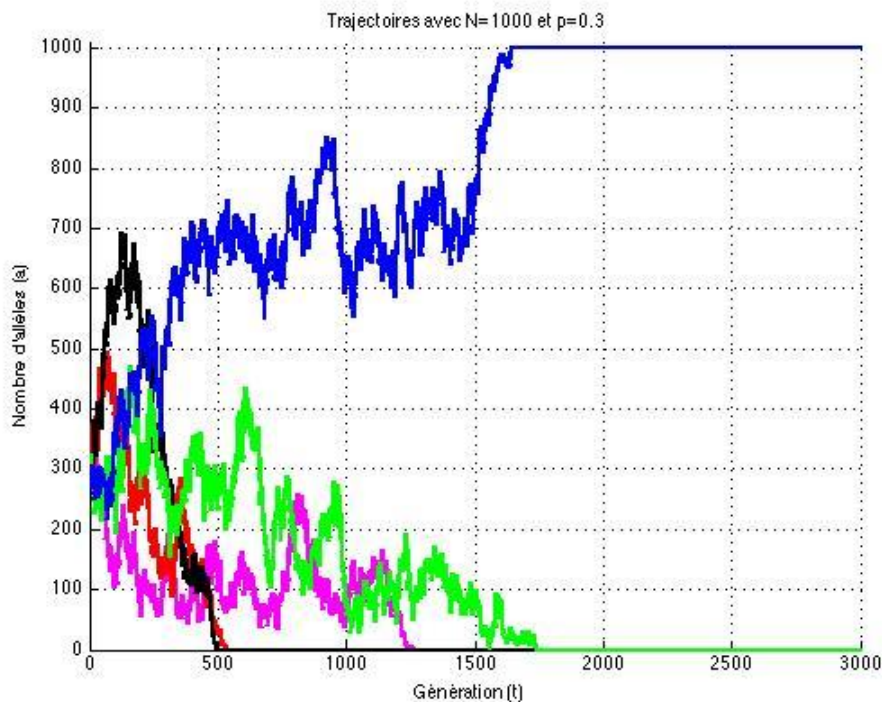
2) Simulation avec Matlab

Les figures ci-dessous illustrent le fait suivant :

Trois populations au départ identiques dans leur composition allélique (proportion p d'allèles **a** égale à 0,3) évoluent de façon différente en raison de la dérive génétique.

Chaque point d'une trajectoire correspond au nombre d'allèles **a** obtenus par le tirage aléatoire avec remise de N allèles ($N = 10$, $N = 100$ ou $N = 1\,000$) dans l'urne pour laquelle la proportion d'allèle **a** est 0,3.





Nous remarquons que dans toutes ces simulations au bout d'un certain nombre de générations, il y a soit une disparition de l'allèle **a**, soit une disparition de l'allèle **A**.
Ce sont les états absorbants.

3) Modèle de Wright-Fisher

Nous avons vu que le choix d'un allèle peut être considéré comme une épreuve de Bernoulli de paramètre $\frac{i}{N}$. La variable aléatoire X_{t+1} / X_t suit donc une loi binomiale de

paramètre $\left(\frac{i}{N}; N\right)$. Donc $P(X_{t+1} = j / X_t = i) = \binom{N}{j} \left(\frac{i}{N}\right)^j \left(1 - \frac{i}{N}\right)^{N-j}$.

Les valeurs prises par X_{t+1} dépendent de X_t et ne dépendent pas directement de $X_0; X_1; X_2; \dots; X_{t-1}$. On dit que $(X_t; t \in \mathbb{N})$ suit une chaîne de Markov.

On définit une matrice, appelée matrice de transition. Elle est formée par les valeurs

des probabilités de $X_{t+1} = j / X_t = i$ c'est à dire par $P(X_{t+1} = j / X_t = i) = \binom{N}{j} \left(\frac{i}{N}\right)^j \left(1 - \frac{i}{N}\right)^{N-j}$ que nous noterons $P(i; j)$.

Chaque élément $P(i; j)$ de la matrice représente la probabilité d'obtenir j allèles **a à la génération $t + 1$ sachant qu'il y avait i allèles **a** à la génération t .**

Exemple : pour $N = 4$ et $P(X_0 = 1) = \frac{1}{4}$

$$M = \begin{bmatrix} P(0;0) & P(0;1) & P(0;2) & \dots & \dots \\ P(1;0) & \dots & \dots & \dots & \dots \\ P(2;0) & \dots & P(2;2) & \dots & \dots \\ P(3;0) & \dots & \dots & \dots & \dots \\ P(4;0) & \dots & \dots & \dots & P(4;4) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0,3164 & 0,4219 & 0,2109 & 0,0469 & 0,0039 \\ 0,0625 & 0,25 & 0,375 & 0,25 & 0,0625 \\ 0,0039 & 0,0469 & 0,2109 & 0,4219 & 0,3164 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

0,0469 représente la probabilité d'obtenir 3 allèles **a** sachant que l'on avait à la génération précédente 1 seul allèle **a**. ($P(1;3) = 0,0469$).

Attention la ligne 1 de la matrice correspond à 0 allèle **a** dans l'urne.

On admet : $P(X_t = j / X_0 = i) = P^t(i, j)$.

Nous avons donc besoin de déterminer l'élément de la matrice M^t correspondant à **i** allèles **a** à la génération initiale ($X_0 = i$) et **j** allèles **a** à la génération **t** ($X_t = j$).

Pour déterminer M^t , nous devons chercher à diagonaliser cette matrice.

La matrice de transition **M** vérifie $M \times V = V \times D \Leftrightarrow M = V \times D \times V^{-1}$

On en déduit alors que $M^t = V \times D^t \times V^{-1}$.

Cette matrice **M** possède cinq valeurs propres et cinq vecteurs propres (déterminés avec matlab) :

- la matrice **V** des vecteurs propres est :

$$V = \begin{bmatrix} 0 & 0 & 0 & 0,7303 & 0 \\ -0,5145 & 0,7071 & -0,4402 & 0,5477 & 0,1826 \\ -0,686 & 0 & 0,7826 & 0,3651 & 0,3651 \\ -0,5145 & -0,7071 & -0,4402 & 0,1826 & 0,5477 \\ 0 & 0 & 0 & 0 & 0,7303 \end{bmatrix}$$

- la matrice diagonale **D** est :

$$D = \begin{bmatrix} 0,75 & 0 & 0 & 0 & 0 \\ 0 & 0,375 & 0 & 0 & 0 \\ 0 & 0 & 0,937 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

On peut alors déterminer M^t :

$$M^{10} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0,7248 & 0,0161 & 0,0181 & 0,0161 & 0,2249 \\ 0,4665 & 0,0215 & 0,0241 & 0,0215 & 0,4665 \\ 0,2249 & 0,0161 & 0,0181 & 0,0161 & 0,3164 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad M^{33} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0,75 & 0 & 0 & 0 & 0,25 \\ 0,5 & 0 & 0 & 0 & 0,5 \\ 0,25 & 0 & 0 & 0 & 0,75 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

0,75 représente la probabilité d'obtenir 0 allèle **a** à la génération 33 sachant qu'il y avait un allèle **a** à la génération initiale.

0,5 représente la probabilité d'obtenir 4 allèles **a** à la génération 33 sachant qu'il y avait deux allèles **a** à la génération initiale.

On remarque que dans cette dernière matrice, il y a deux états à la 33^{ème} génération :

- soit le nombre d'allèles **a** est 0
- soit le nombre d'allèles **a** est 4.

Ce sont les états absorbants. On peut remarquer que dès la 10^{ème} génération les probabilités $P(i,0)$ et $P(i,4)$ sont proches de celles obtenues à la 33^{ème} génération.

Dans le modèle de Wright-Fisher, il a été démontré que la chaîne de Markov $(X_t; t \in \mathbb{N})$ atteint en un temps fini un des deux états absorbants.

C'est à dire qu'au bout d'un certain nombre de générations, tous les allèles seront **a** ou tous les allèles seront **A**.

On ne peut déterminer la première génération pour laquelle l'état absorbant est atteint, mais on connaît deux résultats :

- Lorsque l'état absorbant est atteint, la probabilité d'atteindre l'état absorbant N est égale à $p = \frac{i}{N}$.

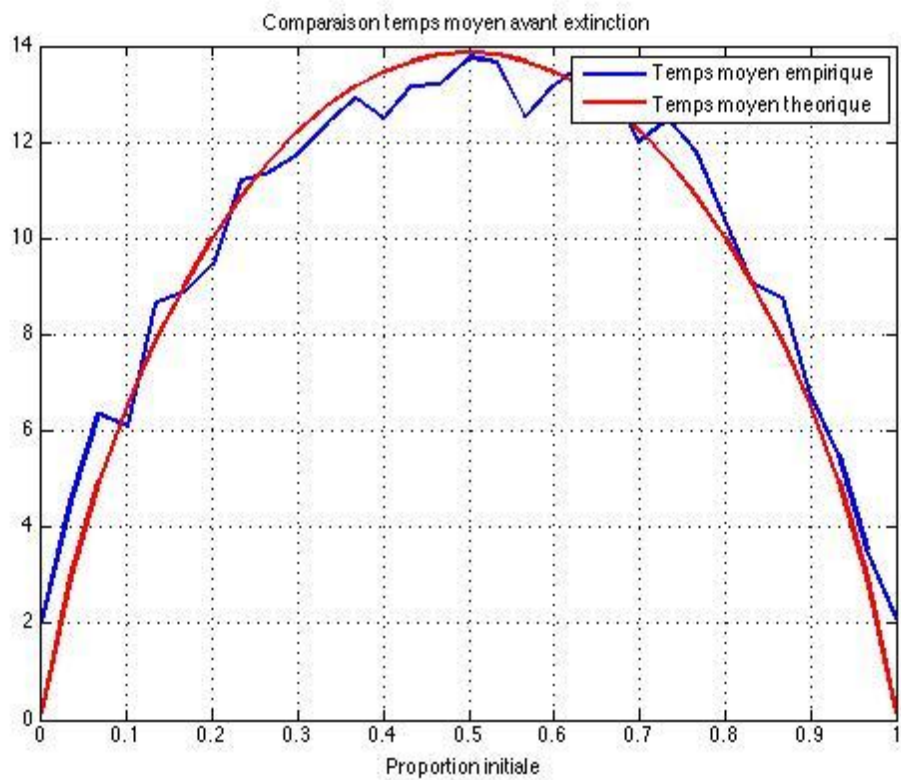
- Le temps moyen où l'on atteint un des états absorbant est approximé par la formule : $T_{\text{moyen}} \approx 2N[x \ln(x) + (1-x) \ln(1-x)]$ avec $x = \frac{i}{N}$ ($x \in [0;1]$)

T_{moyen} est obtenu en calculant la moyenne des premières générations pour atteindre l'état absorbant sur 300 simulations.

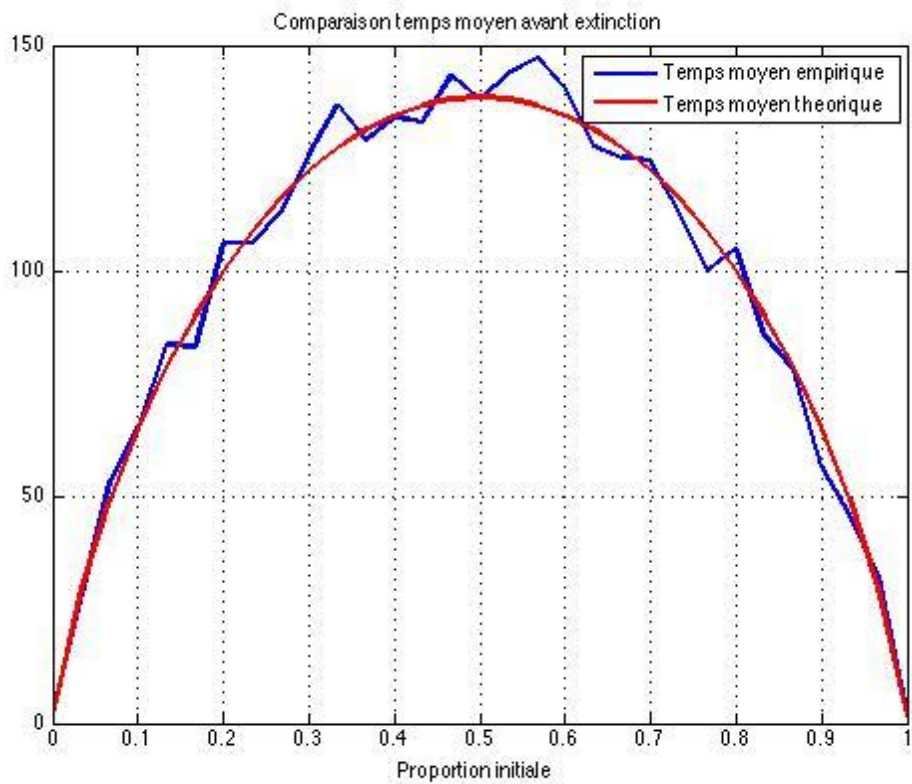
Les graphiques suivants comparent le temps moyen de fixation, obtenu en fonction de la probabilité initiale :

- soit par la théorie
- soit par la simulation de 300 trajectoires

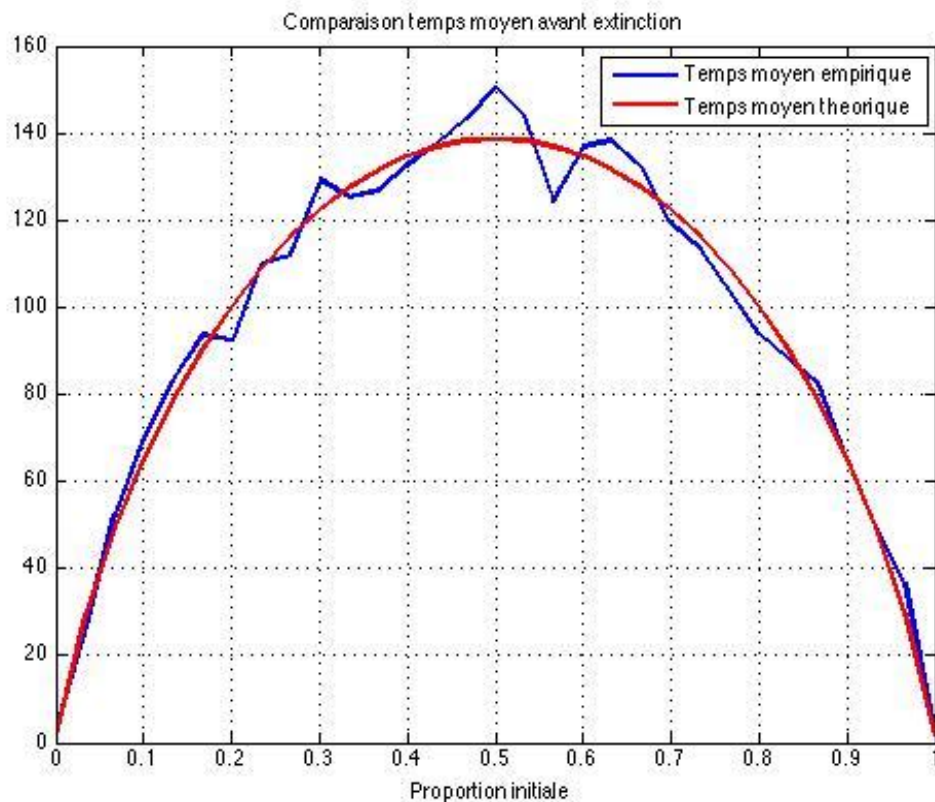
$N = 10$



$N = 100$



$$N = 1\,000$$



Retour au Québec :

Si nous supposons que 1000 individus se sont installés dans la région du Lac - Saint - Jean dont 1 était atteint d'ataxie spastique, le temps moyen de fixation de l'un des gènes obtenu par le calcul est proche de 17 générations.

Dans le cas de l'ataxie spastique de cette région, la proportion de porteurs est passée de $\frac{1}{5000}$ à $\frac{1}{20}$ en 12 générations.

D'après le modèle de Hardy Weinberg, la fréquence allélique serait au bout de 12 générations de $\frac{1}{1000} + \frac{1}{2} \times \frac{1}{20} = \frac{26}{1000} = 0,026$. Alors qu'au départ elle était de

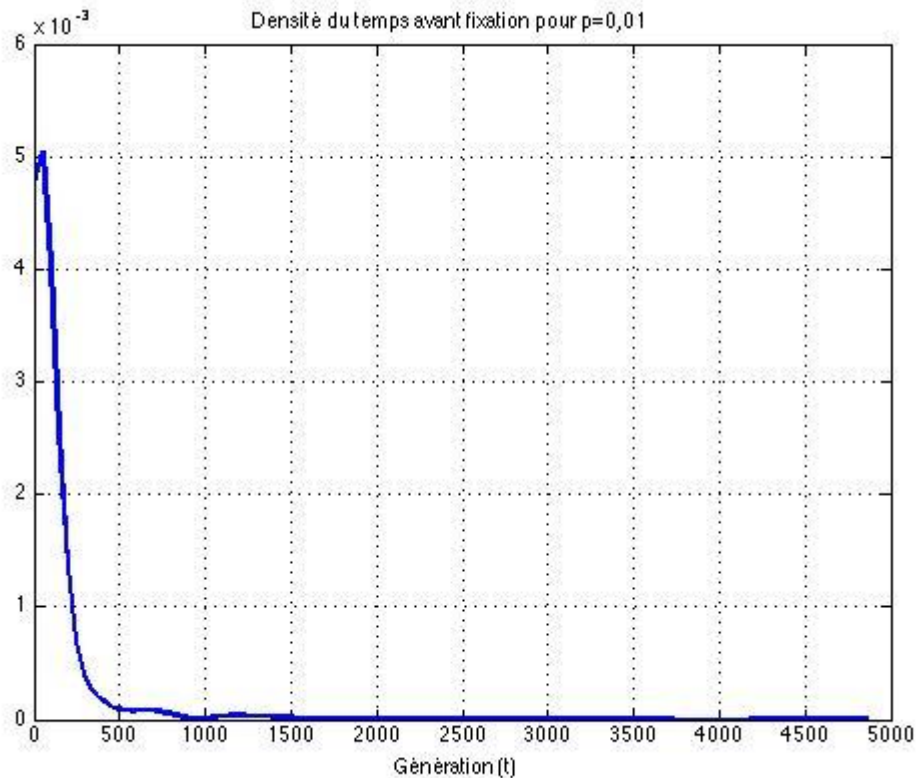
$\frac{1}{1000} + \frac{1}{2} \times \frac{1}{5000} = \frac{11}{10000} = 0,0011$. ($\frac{1}{1000}$ représente le taux de malades ; $\frac{1}{5000}$ et $\frac{1}{20}$ représentent le taux de porteurs).

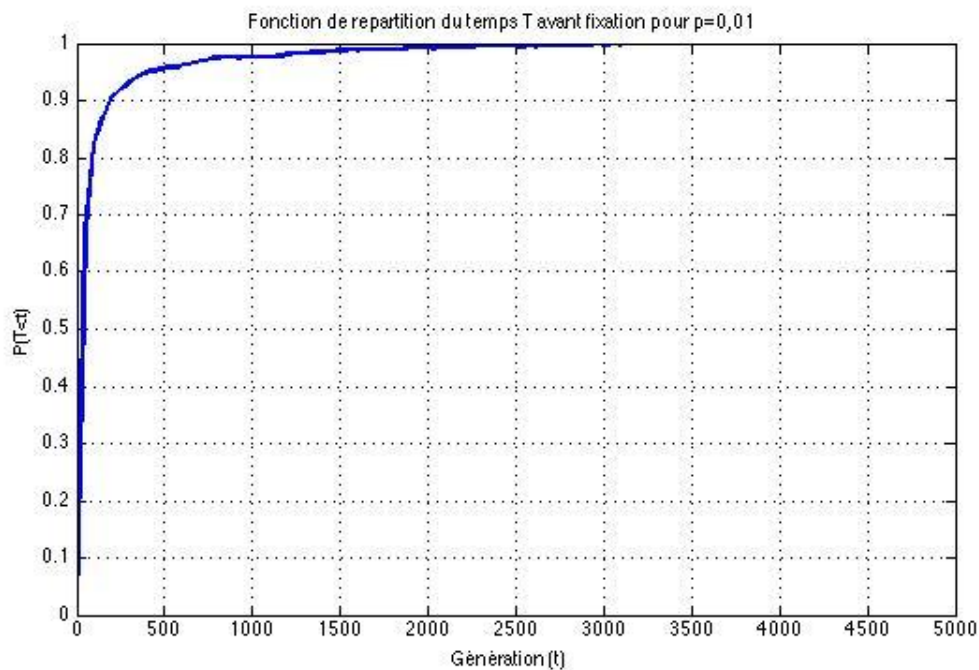
Quelle est la probabilité que le temps moyen de fixation d'un allèle soit inférieur à 12 ? Autrement dit qu'elle est la probabilité qu'il reste des allèles **a** et **A** à la 12^{ème} génération ?

Nous avons simulé 1000 trajectoires suivant la chaîne de Markov avec comme probabilité initiale $p = 0,01$.

On peut remarquer qu'un temps de fixation très grand n'est pas un événement rare.

La densité de probabilité du temps avant fixation est représentée ci-dessous :





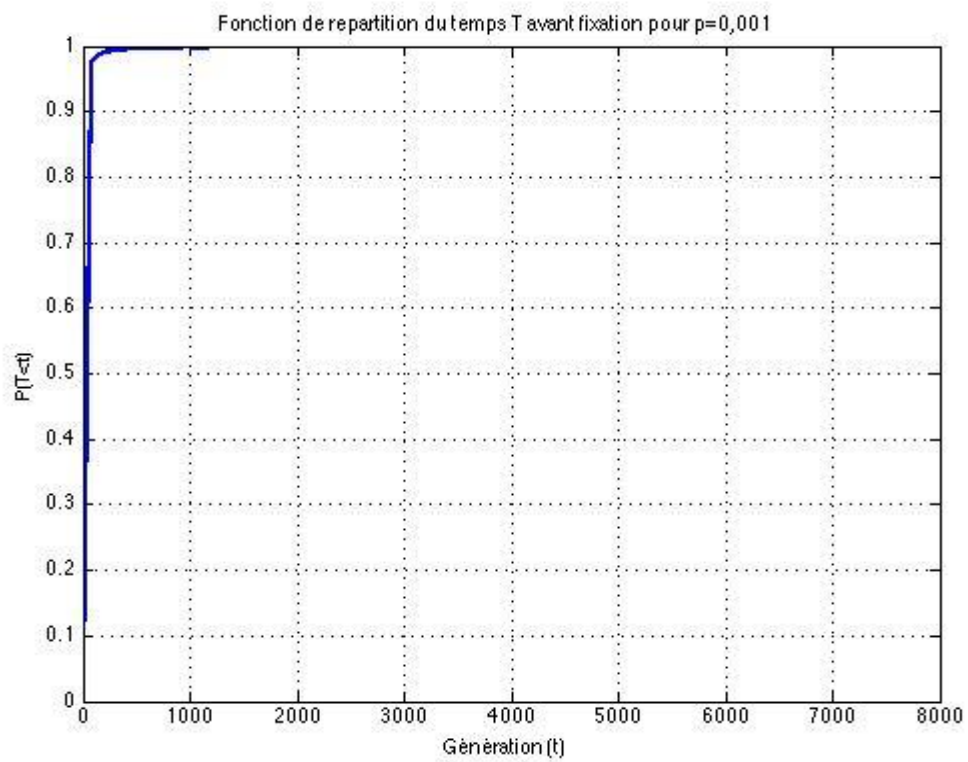
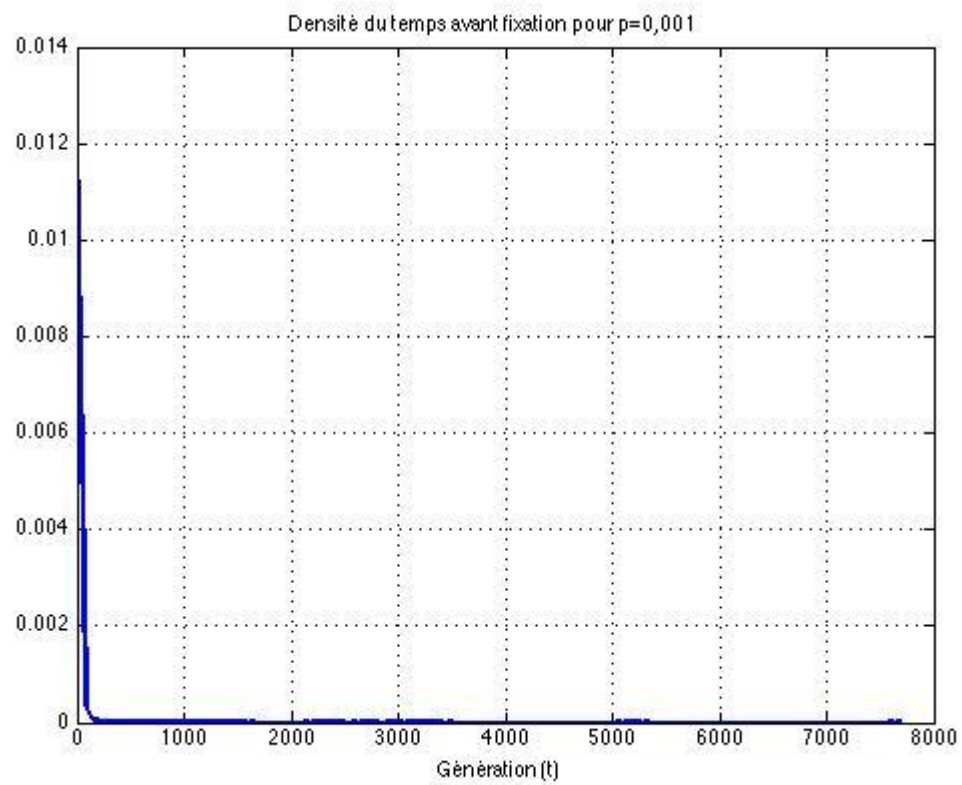
Pour être encore plus proche de la proportion d'allèles a dans la région du Saguenay, nous avons simulé 10 000 trajectoires suivant la chaîne de Markov avec comme probabilité initiale $p = 0,0011$.

Nous avons classé ces 10 000 temps de fixation dans des classes de largeur 76 générations :

Classes (génération)	$[0; 76[$	$[76; 152[$	$[152; 228[$	$[228; 304[$	$[304; 380[$	$[380; 456[$
Fréquence (%)	96,24	0,166	0,69	0,33	0,18	0,1	...	0	0	0,01

On peut encore remarquer qu'un temps de fixation très grand est un événement rare mais qui a un impact important sur la distribution.

Les courbes représentant la densité et la fonction de répartition du temps avant fixation d'un des deux allèles pour $N = 1000$ et $p = 0,0011$ sont données ci dessous :



Conclusion :

Le temps de fixation de 17 générations calculé précédemment ne contredit pas la fréquence actuelle (1/20) de porteurs de l'ataxie.

La fonction de répartition du temps de fixation d'un allèle montre que le pourcentage de trajectoires ayant un temps de fixation inférieur ou égal à 12 est d'environ 40 %. (Le pourcentage de trajectoires ayant un temps de fixation inférieur ou égal à 76 est d'environ 96 %).

Le modèle de Wright-Fisher explique la dérive génétique observée dans la région du Saguenay.

Ce modèle est assez loin de la réalité mais on peut l'améliorer pour qu'il soit un peu plus réaliste en modifiant les probabilités $P(i;j)$ dans la matrice de transition M

- soit en admettant la possibilité de mutation d'un allèle **a** en **A** et/ou la mutation d'un allèle **A** en allèle **a**.

- soit en admettant la possibilité de migration, c'est à dire que des individus extérieurs viennent changer la population.

- soit par des causes environnementales ou comportementales.

Les articles d'Evelyn Heyer montrent que les comportements des familles qui se sont installées dans la région du Saguenay sont à l'origine de ces bouleversements de fréquence allélique. Son travail montre comment la transmission d'une génération à la suivante de comportements démographiques explique la forte fréquence des maladies héréditaires rares observées dans la région du Saguenay-lac-Saint-Jean.

Elle introduit un paramètre : la distribution du « nombre d'enfants utiles ». Un enfant utile est un enfant qui se reproduit au sein de la population dont il est issu. Ce nombre d'enfants utiles d'un couple est corrélé au nombre d'enfants utiles qu'ont eu les parents de ce couple. Cette corrélation c est de 0,2 alors que la corrélation du nombre d'enfants (utiles ou non) n'est que de 0,07. Le facteur c joue un rôle d'accélérateur dans le phénomène de dérive génétique (CNRS Info mars 1999).

Extrait du livre « La mathématisation du réel » de G. Israel p 310.

« L'accord entre théorie mathématique et réalité empirique est assez bon dans les situations à la limite. Le système des interactions entre les sujets y est tellement schématique et raréfié qu'il se confond presque avec un processus mécanique. Dès que des caractéristiques spécifiquement biologiques interviennent des difficultés sérieuses apparaissent.

De plus, les écosystèmes et les processus d'interaction entre sujets biologiques comportent un aspect historique que l'on ne peut ni éliminer ni négliger et ils sont aussi soumis à des influences économiques et sociales. Ils dépendent de choix subjectifs difficiles à formaliser en termes mathématiques et plus encore en termes mécaniques.

Une des principales sources des difficultés vient probablement de la réduction du temps historique dans lequel se déploie le processus, au temps mathématique abstrait.

Cette réduction fait violence aux caractéristiques empiriques du phénomène de façon beaucoup plus forte que quand on utilise la notion de temps mathématique dans les situations à la limite ou dans les expériences de laboratoire. »

V. ANNEXES

1) Applications concrètes du modèle de Hardy Weinberg :

La population étudiée doit être dans les conditions du modèle de Hardy Weinberg et l'échantillon choisi doit être représentatif de cette population.

Grâce à la relation de Hardy - Weinberg, à partir des fréquences phénotypiques observées, on calculera les fréquences alléliques ainsi que les fréquences des porteurs sains de la maladie.

a) Maladie à gène récessif : exemple de l'albinos.

Chez l'homme, le gène a de l'albinisme est récessif par rapport au gène A de la pigmentation normale de la peau. Les individus albinos sont donc homozygotes aa. Des recherches ont montré que la fréquence de l'albinisme, observé à la naissance, est de 1 pour 10 000 personnes.

On cherche à connaître la fréquence des hétérozygotes Aa, porteurs sains du caractère albinisme.

Selon le modèle de Hardy Weinberg (bien qu'ici le caractère aléatoire des mariages ou panmixie soit contestable), on peut dresser le tableau suivant :

	Pigmenté		Albinos
Génotypes	AA	Aa	aa
	u+v = 9 999/10 000		w = 1/10 000
Fréquences génotypiques si Hardy Weinberg	p^2	$2pq$	q^2

Ceci permet de calculer $q = \sqrt{w}$, d'après la relation $p + q = 1$, on obtient alors $p = 1 - q$ et $2pq = 2(1 - q)q \approx 2q$ si q est faible.

On calcule le nombre de porteurs sains pour un atteint : $\frac{2pq}{q^2} = \frac{2(1 - q)}{q}$.

Dans cet exemple, la fréquence de l'allèle responsable de l'albinisme est égale à 1% ($q = 1\%$) et la fréquence des hétérozygotes, non albinos mais porteurs de l'albinisme est égal à 2%. Dans une population de 10 000 individus, on rencontre donc un seul albinos, mais le nombre de porteurs sains pour un atteint est de 200, soit un habitant sur 50.

Bien que, pour la plupart des maladies héréditaires qui affectent l'espérance de vie ou la fertilité et donc la fécondité, la condition d'absence de sélection ne soit pas valide, les généticiens considèrent qu'il est possible d'appliquer le modèle de Hardy - Weinberg entre les fréquences des allèles chez les parents et les fréquences génotypiques chez leurs enfants, avant que la sélection ne fasse son effet.

b) Maladies à gène dominant :

Dans ce cas, le phénotype sain est l'homozygote aa tandis que le phénotype atteint est AA ou Aa.

Les maladies relevant de ce type de gènes sont, par exemple, les hypercholestérolémies familiales, la neurofibromatose type I, la dystrophie myotonique...

Les conditions d'application du modèle de Hardy Weinberg étant respectées, la fréquence F des individus atteints est alors égale à $p^2 + 2pq$, ce qui revient à dire que la fréquence complémentaire $(1-F)$ des individus sains est égale à q^2 .

Connaissant F, la fréquence des individus atteints, et sachant que $p+q=1$, on peut en déduire directement :

- la fréquence de l'allèle a : $q = \sqrt{1-F}$
- la fréquence de l'allèle A dominant, responsable de la maladie :
 $p = 1 - \sqrt{1-F}$

On peut considérer que si la maladie est rare, p est proche de zéro donc :

$$F = p^2 + 2p(1-p) = 2p - p^2 \approx 2p.$$

On peut donc considérer que $p \approx \frac{1}{2}F$.

On montre ici que la maladie est surtout portée par les hétérozygotes toujours plus nombreux et qu'une mutation d'un allèle vers un allèle dominant est un phénomène grave car il génère beaucoup plus de malades que la mutation d'un allèle vers un allèle récessif.

C'est pourquoi, même si les mutations vers des allèles malades dominants sont 1 000 fois plus rares que des mutations vers des allèles malades récessifs, les maladies à gènes dominants sont plutôt moins rares que les maladies à gènes récessifs.

2) Modèle de Wright-Fisher et mutation

Supposons que l'allèle A mute en a avec une probabilité u et que l'allèle a mute en A avec une probabilité v .

Alors la variable aléatoire X_{t+1} / X_t suit une loi binomiale de paramètre $(p_{i,t}; N)$

$$\text{où } p_{j,t+1} = (1-u)p_{i,t} + v(1-p_{i,t}).$$

$(X_t; t \in \mathbb{N})$ suit une chaîne de Markov. Sa matrice de transition est formée par les valeurs

$$\text{des probabilités de } X_{t+1} = j / X_t = i \text{ c'est à dire par : } P(i; j) = \binom{N}{j} (p_i)^j (1-p_i)^{N-j}.$$

En remarquant que, pour toutes valeurs de i comprises entre 0 et N :

$$E(X_{t+1} = j / X_t = i) = (1-u)i + v(N-i).$$

On montre que la proportion du nombre d'allèle A tend vers $\frac{v}{u+v}$.

3) Programmes Matlab pour la simulation du modèle de Wright Fisher

```
clear all
% close all

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Matrice de transition a pour une population de taille n
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
n=10

a=zeros(n+1,n+1);

for i=0:n
a(i+1,:) = binopdf((0:n),n,i/n);
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Diagonalisation
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
[V,D] = eig(a);
D

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Vecteurs propres
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

Vstat1=V(:,n+1);
a*Vstat1;
D(n+1,n+1)*Vstat1;

Vstat2=V(:,n+1);
a*Vstat2;
D(n+1,n+1)*Vstat2;

vh = rand(100,1);
vh = vh/sum(vh);
```

```

%%%%%%%%%%
% Puissance de a
%%%%%%%%%%
for i=1:1:40
a^i
display('$$$')
Vstat1;
pause
end

```

```

%%%%%%%%%%
% simulation de la dérive génétique
%%%%%%%%%%
N=100;
P = [];
P(1)=0.3;
NbAll(1) = P(1)*N;
for i=1:300
    NbAll(i+1) = binornd(N,P(i));
    P(i+1) = NbAll(i+1)/N;
end

```

```

figure(1)
hold on
plot(NbAll,'w','LineWidth',2)
axis([0 300 0 20])
title('Trajectoires avec N=100 et p=0.3')
xlabel('Génération (t)')
ylabel('Nombre d\'allèles (a)')
grid

```

```

%%%%%%%%%%
% Calcul du temps de fixation dans un modèle de dérive
% génétique de type Fisher
%%%%%%%%%%
% clear all
close all

N=1000;% taille de la population d'allèles

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Calcul de la moyenne empirique du temps de fixation en
%fonction de P1
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

NbMc = 10000;% Nombre de trajectoires
NbPointsP1 = 100;
tabp1 = 0.001 + (0:NbPointsP1)*(0.999-0.001)/NbPointsP1;

k=0;
for p1 = tabp1
    P(1) = p1;
    k = k+1;
    for j=1:NbMc
        i=1;
        u = -1;
        while (u ~= 0 & u ~= N)
            NbAll(i) = binornd(N,P(i));
            u = NbAll(i);
            P(i+1) = NbAll(i)/N;
            i=i+1;
        end
        ind(j)=i;
    end
    MoyenneT(k) = mean(ind);
    SigmaT(k) = sqrt(cov(ind));
    progressbar(k/length(tabp1))
end

x=tabp1;
MoyenneTtheo = -2*N*(x.*log(x)+(1-x).*log(1-x));

figure
plot(tabp1,MoyenneT,'LineWidth',2)
hold on
plot(tabp1,MoyenneTtheo,'r','LineWidth',2)
grid
legend('Temps moyen empirique','Temps moyen theorique')
title('Comparaison temps moyen avant fixation')
xlabel('Proportion initiale')

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Calcul densité et fonction de répartition du temps moyen
%pour p1
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

```

p1 = 0.001;
P(1) = p1;
for j=1:NbMc
    i=1;
    u = -1;
    while (u ~= 0 & u ~= N)
        NbAll(i) = binornd(N,P(i));
        u = NbAll(i);
        P(i+1) = NbAll(i)/N;
        i=i+1;
    end
    ind(j)=i;
end

tab_points = [0:(max(ind)-min(ind))/100:max(ind)+50];
data = ind;

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% calcul de la densité
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

```

densite = dens1D_non_adapt(tab_points,data,1);

% normalisation de la densite
dx = [tab_points(2)-tab_points(1)];
Int = dx*sum(densite);
densite = densite/Int;

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% calcul de la fonction de répartition
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
k = 0;
FR = [];
for t = tab_points
    k = k+1;
    FR(k) = sum((ind <= t))/NbMc;
end

figure
plot(tab_points,FR,'LineWidth',2)
grid
title('Fonction de repartition du temps T avant fixation
pour p=0,001')
xlabel('Génération (t)')
ylabel('P(T<t)')

figure
plot(tab_points,densite,'b','LineWidth',2)
grid
title('Densité du temps avant fixation pour p=0,001')
xlabel('Génération (t)')

yy = hist(data);

```

VI. **ACTIVITES EN CLASSE**

La modélisation est au programme aussi bien au collège qu'au lycée.

Extraits des Programmes du collège (bulletin officiel n° 6 du 28 août 2008).

« A travers la résolution de problèmes, la modélisation de quelques situations et l'apprentissage progressif de la démonstration, les élèves prennent conscience petit à petit de ce qu'est une véritable activité mathématiques ... »

« La notion de probabilité est utilisée pour modéliser des situations simples de la vie courante. »

« Parmi les situations qui ne relèvent pas de la proportionnalité, certaines sont cependant modélisables par une fonction ... »

Extraits des Programmes du lycée

En seconde (projet pour la rentrée 2009)

« L'objectif de ce programme est de former les élèves à la démarche scientifique sous toutes ses formes pour les rendre capables de :

- pratiquer une activité expérimentale ou algorithmique ;
- modéliser et s'engager dans une activité de recherche ;
- conduire un raisonnement, une démonstration ;
- faire une analyse critique ;
- pratiquer une lecture active de l'information (critique, traitement), en privilégiant les changements de registre (graphique, numérique, algébrique, géométrique) ;
- utiliser les outils logiciels (ordinateur ou calculatrice) adaptés à la résolution d'un problème ;
- communiquer à l'écrit et à l'oral »

En 1^{ère} ES, option mathématiques

Un chapitre de géométrie vise à étendre à l'espace les acquis antérieurs dans le plan: calculs et illustrations graphiques seront menés simultanément et prépareront le terrain à des modélisations ultérieures.

En TPE

Le thème : modèle, modélisation était proposé en filière S cette année.

En Terminale ES, spécialité

Une ouverture sur la théorie des graphes

« Ce choix est cohérent tant avec le programme de la classe antérieure qu'avec les exigences de formation ultérieure : on trouve en effet ici quelques applications intéressantes du calcul matriciel développé dans l'option de première ES; par ailleurs, les problèmes résolus constituent une première approche, volontairement modeste, de situations complexes (d'ordonnancement, d'optimisation de flux, de recherche de fichiers informatiques, d'études de migrations de populations...) auxquelles de nombreux élèves seront par la suite confrontés, notamment en gestion ou en informatique. Ce thème sensibilise naturellement à l'algorithmique et, en montrant la puissance de la théorie des graphes pour la modélisation, permet un autre regard mathématique sur diverses situations.

L'objectif est de savoir modéliser des situations par des graphes et d'identifier en terme de propriétés de graphes la question à résoudre. »

En Terminale S

« Les formations supérieures sont naturellement diverses et offrent aux mathématiques une place variable. Dans certaines filières, elles seront une matière centrale et pour toutes un outil de modélisation et de calcul. La réussite des jeunes étudiants dépendra donc crucialement de leur maîtrise des sciences mathématiques ; pour les préparer, le programme prend en compte les évolutions de la discipline et différentes demandes qui sont l'expression des besoins mathématiques croissants de notre société.

L'initiation au calcul intégral et à la problématique des équations différentielles : la présence de ces dernières, bien que modeste dans le libellé du programme, est fondamentale pour amener à la compréhension de la puissance des mathématiques pour la modélisation; un travail conjoint avec les autres disciplines favorisera cet objectif. »

Voici quelques activités proposées à nos élèves :

1) POUR LA CLASSE DE 3^{ème}

Etude des gènes A, B et O, responsables des phénotypes dominants et récessifs des groupes sanguins.

Cette activité s'adresse à des élèves de 3^{ème} et s'appuie sur les connaissances acquises aux cours de Sciences et Vie de la Terre.

Nous nous proposons de l'enrichir par l'analyse de données statistiques réelles (prises dans Génétiques des populations de Jean-Louis Serre).

FICHE ELEVE

ETUDE DES ALLELES A, B ET O DU GENE LE GROUPE SANGUIN.

Objectif : Identifier le nombre de « porteurs » des allèles A et O parmi des individus de groupe sanguin [A] et le nombre de « porteurs » des allèles B et O parmi des individus de groupe sanguin [B].

1) Questions préliminaires

Nous considérons 100 cellules reproductrices dont 45 sont porteurs de l'allèle A, 13 sont porteurs de l'allèle B et 42 sont porteurs de l'allèle O, sur le chromosome n°9.

- Calculer les fréquences des cellules reproductrices, « porteurs respectifs » des allèles A, B et O.
- Ces cellules reproductrices vont s'unir au hasard, il n'y a ni mutation, ni sélection, selon le modèle de messieurs Hardy et Weinberg.
Compléter le tableau ci-dessous en calculant la fréquence de chaque union possible :

	A	B	O
A	AA	AB	AO
B	BA	BB	BO
O	OA	OB	OO

- A l'aide du tableau, calculer la fréquence de chaque couple d'allèles possible :

	AA	AB	AO	BB	BO	OO
fréquences

Comparer la fréquence du couple d'allèles AO à celle du couple d'allèles AA.

Comparer la fréquence du couple d'allèles BO à celle du couple d'allèles BB.

- Calculer la fréquence de chaque groupe sanguin :

- Cas général

Nous considérons les fréquences alléliques théoriques suivantes :
p pour l'allèle A, q pour l'allèle B et r pour l'allèle O avec $p+q+r=1$.

Le modèle mathématique établi par Messieurs Hardy et Weinberg suppose que les couples se forment au hasard et qu'il n'y a ni mutation, ni sélection d'allèles.

Ces hypothèses permettent de compléter le tableau théorique suivant :

	A fréquence p	B fréquence q	O fréquence r
A fréquence p	AA	AB	AO
B fréquence q	BA	BB	BO
O fréquence r	OA	OB	OO

On a relevé les effectifs de chaque groupe sanguin dans un échantillon aléatoire de 1 000 individus.

Compléter le tableau suivant à l'aide du tableau précédent :

Groupe sanguin	Paires d'allèles	Fréquences	Effectifs observés
[A]	AA et AO		460
[B]	BB et BO		140
[AB]	AB		50
[O]	OO		350
Total			1 000

On note $f_{[A] + [O]}$ la somme des fréquences des groupes sanguins [A] et [O] et $f_{[B] + [O]}$ la somme des fréquences des groupes sanguins [B] et [O].

Compléter, puis factoriser : $f_{[A] + [O]} = \dots = \dots$

Sachant que $p + q + r = 1$, exprimer $f_{[A] + [O]}$ en fonction de q.

De même, compléter, puis factoriser : $f_{[B] + [O]} = \dots = \dots$

Sachant que $p + q + r = 1$, exprimer $f_{[B] + [O]}$ en fonction de p.

En déduire les valeurs numériques de p et q, afin de compléter le tableau suivant :

Groupe sanguin	Paires d'allèles	Effectifs	Effectifs observés
[A]	AA		} 460
	AO		
[B]	BB		} 140
	BO		
[AB]	AB		50
[O]	OO		350

Quel écart maximum pouvez vous observer entre les effectifs observés et ceux obtenus pour chaque groupe sanguin, par type d'allèles, par la loi de Hardy Weinberg ?

Quel est le pourcentage de porteurs du couple d'allèles AO dans le groupe sanguin [A] ?

Quel est le pourcentage de porteurs BO dans le groupe sanguin [B] ?

A partir de ces résultats, quelle remarque pouvez-vous formuler ?

FICHE CORRECTION

ETUDE DES ALLELES A, B ET O DU GENE LE GROUPE SANGUIN.

Objectif : Identifier le nombre de « porteurs » des allèles A et O parmi des individus de groupe sanguin [A] et le nombre de « porteurs » des allèles B et O parmi des individus de groupe sanguin [B].

Questions préliminaires

Nous considérons 100 cellules reproductrices dont 45 sont porteurs de l'allèle A, 13 sont porteurs de l'allèle B et 42 sont porteurs de l'allèle O, sur le chromosome n°9.

Calculer les fréquences des cellules reproductrices, « porteurs respectifs » des allèles A, B et O.

La fréquence des cellules reproductrices, « porteurs » de l'allèle A est 0,45, celle de l'allèle B est 0,13 et celle de l'allèle O est de 0,42.

Ces cellules reproductrices vont s'unir au hasard, il n'y a ni mutation, ni sélection, selon le modèle de messieurs Hardy et Weinberg.

Compléter le tableau ci-dessous en calculant la fréquence de chaque union possible :

	A	B	O
A	AA $0,45 \times 0,45 = 0,2025$	AB $0,45 \times 0,13 = 0,0585$	AO $0,45 \times 0,42 = 0,189$
B	BA $0,45 \times 0,13 = 0,0585$	BB $0,13 \times 0,13 = 0,0169$	BO $0,13 \times 0,42 = 0,0546$
O	OA $0,45 \times 0,42 = 0,189$	OB $0,42 \times 0,13 = 0,0546$	OO $0,42 \times 0,42 = 0,1764$

A l'aide du tableau, calculer la fréquence de chaque couple d'allèles possible :

	AA	AB	AO	BB	BO	OO
f	0,2025	$0,0585 \times 2 = 0,117$	$0,189 \times 2 = 0,378$	0,0169	$0,0546 \times 2 = 0,1092$	0,1746

Comparer la fréquence du couple d'allèles AO à celle du couple d'allèles AA.

La fréquence du couple AO est supérieure à celle de AA.

Comparer la fréquence du couple d'allèles BO à celle du couple d'allèles BB.

La fréquence du couple BO est six fois supérieure à celle du couple BB.

Calculer la fréquence de chaque groupe sanguin :

La fréquence du groupe sanguin [A] est la somme de la fréquence du couple AA et du couple AO : $0,2025 + 0,378 = 0,5544$.

La fréquence du groupe sanguin [B] est la somme de la fréquence du couple BB et du couple BO : $0,0169 + 0,1092 = 0,1261$.

La fréquence du groupe sanguin [AB] est la fréquence du couple AB : 0,117.

La fréquence du groupe sanguin [O] est la fréquence du couple OO : 0,1746.

Cas général

Nous considérons les fréquences alléliques théoriques suivantes :
p pour l'allèle A, q pour l'allèle B et r pour l'allèle O avec $p+q+r=1$.

Le modèle mathématique établi par Messieurs Hardy et Weinberg suppose que les couples se forment au hasard et qu'il n'y a ni mutation, ni sélection d'allèles.

Ces hypothèses permettent de compléter le tableau théorique suivant :

	A fréquence p	B fréquence q	O fréquence r
A fréquence p	AA p^2	AB pq	AO pr
B fréquence q	AB qp	BB q^2	BO qr
O fréquence r	AO rp	OB rq	OO r^2

On a relevé les effectifs de chaque groupe sanguin dans un échantillon aléatoire de 1 000 individus.

Compléter le tableau suivant à l'aide du tableau précédent :

Groupe sanguin	Paires d'allèles	Fréquences	Effectifs observés
[A]	AA et AO	p^2+2pr	460
[B]	BB et BO	q^2+2qr	140
[AB]	AB	$2pq$	50
[O]	OO	r^2	350
Total		1	1 000

On note $f_{[A]+[O]}$ la somme des fréquences des groupes sanguins [A] et [O] et $f_{[B]+[O]}$ la somme des fréquences des groupes sanguins [B] et [O].

Compléter, puis factoriser : $f_{[A]+[O]} = p^2 + 2pr + r^2 = (p+r)^2$

Sachant que $p + q + r = 1$, exprimer $f_{[A]+[O]}$ en fonction de q : $(1-q)^2$

De même, compléter, puis factoriser : $f_{[B]+[O]} = q^2 + 2qr + r^2 = (q+r)^2$

Sachant que $p + q + r = 1$, exprimer $f_{[B]+[O]}$ en fonction de p : $(1-p)^2$

En déduire les valeurs numériques de p et q, afin de compléter le tableau suivant : $q = 1 - \sqrt{f_{[A]+[O]}}$ et $p = 1 - \sqrt{f_{[B]+[O]}}$

Groupe sanguin	Paires d'allèles	Effectifs	Effectifs observés
[A]	AA	90	} 460
	AO	360	
[B]	BB	10	} 140
	BO	120	
[AB]	AB	60	50
[O]	OO	360	350

Quel écart maximum pouvez vous observer entre les effectifs observés et ceux obtenus pour chaque groupe sanguin, par type d'allèles, par la loi de Hardy Weinberg ?

L'écart pour chaque groupe sanguin est de 10 individus soit en perte soit en excès.

Quel est le pourcentage de porteurs du couple d'allèles AO dans le groupe sanguin [A] ?

Il y a 80% de porteurs d'allèles AO dans le groupe sanguin [A].

Quel est le pourcentage de porteurs BO dans le groupe sanguin [B] ?

Il y a 92 % de porteurs d'allèles BO.

A partir de ces résultats, quelle remarque pouvez-vous formuler ?

Le nombre de porteurs d'allèles AO et d'allèles BO est très supérieur aux porteurs d'allèles AA ou BB. Dans les générations suivantes, le groupe sanguin [O] va se perpétuer bien que l'allèle O soit récessif.

Commentaires

Cette activité a été proposée à une classe de 28 élèves très hétérogène, vivante et pleine de bonne volonté. Elle s'est déroulée en trois étapes :

La première a eu lieu en classe un vendredi et a permis de travailler les questions préliminaires qui n'ont pas soulevé de problème.

La deuxième consistait en un travail à effectuer à la maison.

La troisième étape est la correction en classe le lundi suivant.

Peu d'élèves avaient fait leur travail (c'était un week-end très ensoleillé de l'Ascension...) et nous avons donc travaillé cette partie ensemble.

Factorisation avec identité remarquable, résolution d'équation et passage de fréquence à effectif ont posées quelques problèmes à des élèves en difficulté, mais, dans l'ensemble, la partie mathématique correspondait aux révisions du brevet blanc passé peu de temps auparavant.

Les notations ont un peu dérouté les élèves.

En revanche, les questions concernant « l'analyse » des résultats ont été bien réussies, ce changement de cadre n'a pas gêné les élèves, ils ont répondu rapidement et correctement.

En fait, il aurait fallu que cette activité ait lieu plus tôt dans l'année, au moment où le professeur de SVT enseigne cette partie du programme.

Ma collègue a « validé » cette activité et m'a proposé de la faire ensemble l'année prochaine. Cela sera plus enrichissant pour les élèves (et moi !).

ETUDE DES ALLELES A, B ET O DU GENE LE GROUPE SANGUIN.

Objectif : Identifier le nombre de « porteurs » des allèles A et O parmi des individus de groupe sanguin [A] et le nombre de « porteurs » des allèles B et O parmi des individus de groupe sanguin [B].

1) Questions préliminaires

Nous considérons 100 cellules reproductrices dont 45 sont porteurs de l'allèle A, 13 sont porteurs de l'allèle B et 42 sont porteurs de l'allèle O, sur le chromosome n°9.

- a) Calculer les fréquences des cellules reproductrices, « porteurs respectifs » des allèles A, B et O.

$$f_A = \frac{45}{100} = 0,45.$$

$$f_B = \frac{13}{100} = 0,13.$$

$$f_O = \frac{42}{100} = 0,42.$$

- b) Ces cellules reproductrices vont s'unir au hasard, il n'y a ni mutation, ni sélection, selon le modèle de messieurs Hardy et Weinberg.

Compléter le tableau ci-dessous en calculant la fréquence de chaque union possible :

	A	B	O
A	AA 0,2025	AB 0,117	AO 0,378
B	BA 0,117	BB 0,0169	BO 0,1092
O	OA 0,378	OB 0,1092	OO 0,1764

- c) A l'aide du tableau, calculer la fréquence de chaque couple d'allèles possible :

	AA	AB	AO	BB	BO	OO
fréquences	0,2025	0,117	0,378	0,0169	0,1092	0,1764

- Comparer la fréquence du couple d'allèles AO à celle du couple d'allèles AA.
- Comparer la fréquence du couple d'allèles BO à celle du couple d'allèles BB.

- d) Calculer la fréquence de chaque groupe sanguin :

$$f_A = AA + AO = 0,5805$$

$$f_B = BB + BO = 0,1261$$

$$f_O = BO + OO = 0,2856$$

$$f_{AB} = 0,117 =$$

2) Cas général

Nous considérons les fréquences alléliques théoriques suivantes :
p pour l'allèle A, q pour l'allèle B et r pour l'allèle O avec $p+q+r=1$.

Le modèle mathématique établi par Messieurs Hardy et Weinberg suppose que les couples se forment au hasard et qu'il n'y a ni mutation, ni sélection d'allèles.

Elève 1

2

Ces hypothèses permettent de compléter le tableau théorique suivant :

	A fréquence p	B fréquence q	O fréquence r
A fréquence p	AA fréquence p^2	AB fréq. pq	AO pr
B fréquence q	BA qp	BB q^2	BO qr
O fréquence r	OA rp	OB rq	OO r^2

- a) On a relevé les effectifs de chaque groupe sanguin dans un échantillon aléatoire de 1 000 individus.

Compléter le tableau suivant à l'aide du tableau précédent :

Groupe sanguin	Paires d'allèles	Fréquences	Effectifs observés
[A]	AA et AO	$p^2 + 2pr$	460
[B]	BB et BO	$q^2 + 2qr$	140
[AB]	AB	$2pq$	50
[O]	OO	r^2	350
Total		1	1 000

- b) On note $f_{[A] + [O]}$ la somme des fréquence des groupes sanguins [A] et [O] et $f_{[B] + [O]}$ la somme des fréquence des groupes sanguins [B] et [O].

Compléter, puis factoriser : $f_{[A] + [O]} = p^2 + 2pr + r^2 = (p+r)^2$

Sachant que $p + q + r = 1$, exprimer $f_{[A] + [O]}$ en fonction de q : $(1-q)^2$

De même, compléter, puis factoriser : $f_{[B] + [O]} = q^2 + 2qr + r^2 = (q+r)^2$

Sachant que $p + q + r = 1$, exprimer $f_{[B] + [O]}$ en fonction de p : $(1-p)^2$

- c) En déduire les valeurs numériques de p et q, afin de compléter le tableau suivant :

Groupe sanguin	Paires d'allèles	Effectifs	Effectifs observés
[A]	AA	90	460
	AO	360	
[B]	BB	10	140
	BO	120	
[AB]	AB	60	50
[O]	OO	360	350

$$p = \frac{90}{1000} = 0,09$$

$$q = 0,3$$

$$q = 1 - \sqrt{p_{[A] + [O]}}$$

$$p = 1 - \sqrt{p_{[B] + [O]}}$$

$$r = 1 - (p + q)$$

- Quel écart maximum pouvez vous observer entre les effectifs observés et ceux obtenus pour chaque groupe sanguin, par type d'allèles, par la loi de Hardy Weinberg ?

10

- Quel est le pourcentage de porteurs du couple d'allèles AO dans le groupe sanguin [A] ?

$$\frac{360}{460} = 0,8 \rightarrow 80\%$$

- Quel est le pourcentage de porteurs BO dans le groupe sanguin [B] ?

$$\frac{120}{140} \approx 85,7\%$$

- A partir de ces résultats, quelle remarque pouvez-vous formuler ?

ETUDE DES ALLELES A, B ET O DU GENE LE GROUPE SANGUIN.

Objectif : Identifier le nombre de « porteurs » des allèles A et O parmi des individus de groupe sanguin [A] et le nombre de « porteurs » des allèles B et O parmi des individus de groupe sanguin [B].

1) Questions préliminaires

Nous considérons 100 cellules reproductrices dont 45 sont porteurs de l'allèle A, 13 sont porteurs de l'allèle B et 42 sont porteurs de l'allèle O, sur le chromosome n°9.

- a) Calculer les fréquences des cellules reproductrices, « porteurs respectifs » des allèles A, B et O.

$$\text{fréquence de A} = \frac{45}{100} = 0,45 \quad \text{fréquence de B} = \frac{13}{100} = 0,13 \quad \text{fréquence de O} = \frac{42}{100} = 0,42$$

- b) Ces cellules reproductrices vont s'unir au hasard, il n'y a ni mutation, ni sélection, selon le modèle de messieurs Hardy et Weinberg.

Compléter le tableau ci-dessous en calculant la fréquence de chaque union possible :

	A	B	O
A	AA 0,2025	AB 0,0585	AO 0,189
B	BA 0,0585	BB 0,0169	BO 0,0546
O	OA 0,189	OB 0,0546	OO 0,1764

- c) A l'aide du tableau, calculer la fréquence de chaque couple d'allèles possible :

	AA	AB	AO	BB	BO	OO
fréquences	0,2025	0,117	0,378	0,0169	0,1092	0,1764

- Comparer la fréquence du couple d'allèles AO à celle du couple d'allèles AA.

fréq AO est supérieur à fréq AA

- Comparer la fréquence du couple d'allèles BO à celle du couple d'allèles BB.

fréq BO est 6 fois supérieur aux fréq BB

- d) Calculer la fréquence de chaque groupe sanguin :

$$\begin{aligned} \text{groupe A} & \quad AA + AO = 0,2025 + 0,378 = 0,5805 \\ \text{" B} & \quad BB + BO = 0,0169 + 0,1092 = 0,1261 \\ \text{" O} & \quad OO = 0,1764 \\ \text{" AB} & \quad AB = 0,117 \end{aligned}$$

2) Cas général

Nous considérons les fréquences alléliques théoriques suivantes :
p pour l'allèle A, q pour l'allèle B et r pour l'allèle O avec $p+q+r=1$.

Le modèle mathématique établi par Messieurs Hardy et Weinberg suppose que les couples se forment au hasard et qu'il n'y a ni mutation, ni sélection d'allèles.

Ces hypothèses permettent de compléter le tableau théorique suivant :

	A fréquence p	B fréquence q	O fréquence r
A fréquence p	AA p^2	AB $p \cdot q$	AO $p \cdot r$
B fréquence q	BA $q \cdot p$	BB q^2	BO $q \cdot r$
O fréquence r	OA $r \cdot p$	OB $r \cdot q$	OO r^2

- a) On a relevé les effectifs de chaque groupe sanguin dans un échantillon aléatoire de 1 000 individus.

Compléter le tableau suivant à l'aide du tableau précédent :

Groupe sanguin	Paires d'allèles	Fréquences	Effectifs observés
[A]	AA et AO	$p^2 + 2pr$	460
[B]	BB et BO	$q^2 + 2qr$	140
[AB]	AB	$2pq$	50
[O]	OO	r^2	350
Total		1	1 000

- b) On note $f_{[A] + [O]}$ la somme des fréquences des groupes sanguins [A] et [O] et $f_{[B] + [O]}$ la somme des fréquences des groupes sanguins [B] et [O].

Compléter, puis factoriser : $f_{[A] + [O]} = p^2 + 2pr + r^2 = (p+r)^2$

Sachant que $p + q + r = 1$, exprimer $f_{[A] + [O]}$ en fonction de q : $(1-q)^2$

De même, compléter, puis factoriser : $f_{[B] + [O]} = q^2 + 2qr + r^2 = (q+r)^2$

Sachant que $p + q + r = 1$, exprimer $f_{[B] + [O]}$ en fonction de p : $(1-p)^2$

- c) En déduire les valeurs numériques de p et q, afin de compléter le tableau suivant :

Groupe sanguin	Paires d'allèles	Fréquences	Effectifs observés
[A]	AA	0,36	} 460
	AO	0,24	
[B]	BB	0,09	} 140
	BO	0,11	
[AB]	AB	0,10	50
[O]	OO	0,35	350

- Quel écart maximum pouvez vous observer entre les effectifs observés et ceux obtenus pour chaque groupe sanguin, par type d'allèles, par la loi de Hardy Weinberg ?

Il est de dix ou en moins

- Quel est le pourcentage de porteurs du couple d'allèles AO dans le groupe sanguin [A] ?

Il est d'environ 80 %

- Quel est le pourcentage de porteurs BO dans le groupe sanguin [B] ?

Il est d'environ 82 %

- A partir de ces résultats, quelle remarque pouvez-vous formuler ?

que les AO et les BO sont majoritaires.

2) POUR LA CLASSE DE TERMINALE S

FICHE ELEVE

Situation : Prenons un croisement entre deux populations de souris, une blanche (homozygote g/g) et une grise (hétérozygote G/g), l'allèle G étant dominant et produisant des souris grises.

Au lieu de travailler avec les génotypes, on peut ne prendre en compte que les allèles g et G .

A chaque génération, **les descendants sont croisés entre eux** (c'est à dire qu'ils ne se croisent pas avec les individus des générations précédentes), ce processus se répète, et au bout d'un nombre variable de générations, l'un des deux allèles sera complètement et définitivement éliminé, et tous les individus seront porteurs du même allèle.

Modélisation :

Soit une population de 10 souris (20 allèles) dont la fréquence d'allèles g est égale à 0,4. On note $N=20$ et $p=0,4$.

On s'intéresse au nombre d'allèles g obtenus.

- A. Quelle expérience mathématique pouvez-vous faire afin d'étudier la répartition des allèles G et g à la génération suivante ?

Aide éventuelle :

Pour simuler la génération suivante, on effectue un tirage au hasard avec remise de 20 allèles.

1) Etudier la loi de probabilité de la variable aléatoire X égale au nombre d'allèles g .

2) Calculer $E(X)$ et $\sigma(X)$.

Nous vous proposons de simuler cette expérience avec un tableur.

Aide éventuelle :

N=	20
p=	0,4
Génération	1
Fréquences	0,4
	0
	0
	1
	0

Quelle est la fréquence d'allèles g obtenue dans cette nouvelle population ?

B. Comment programmer votre feuille de calcul pour obtenir la fréquence de l'allèle g au bout de 40 générations ?

Aide éventuelle :

N=	20				
p=	0,4				
Génération	1	2	3	4	5
Fréquences	0,4	0,5	0,7	0,9	0,9
	0	1	1	1	1
	0	1	1	1	1
	1	1	1	1	1
	0	1	1	1	1
	1	0	1	1	1
	1	1	1	1	1
	0	0	1	0	0

- 1) Quelle fréquence obtenez-vous à la 40^{ème} génération ?
- 2) Tracer la courbe des fréquences en fonction du nombre de générations.?
- 3) A l'aide de la touche F9, simuler d'autres tirages.
 Obtenez-vous les mêmes fréquences à la 40^{ème} génération ?
 Vos simulations correspondent-elles à la situation décrite ?

Commentaires

L'activité a été proposée à un groupe d'élèves volontaires issus d'une classe de terminale S spécialité SVT. Ce sont les professeurs de SVT et de mathématiques qui sont venus en parler à la classe.

Les élèves ont dans l'ensemble été très surpris par la situation et n'ont pas repéré dans un premier temps les mathématiques sous jacentes.

Le professeur de mathématiques aidé de celui de SVT a dû insister sur le choix aléatoire d'allèles pour qu'ils pensent aux probabilités. L'image de l'urne allélique est venue alors assez rapidement mais le tirage avec remise leur a été expliqué par la collègue de SVT en faisant référence à la transmission des gènes des parents à leurs enfants.

Les élèves ont ensuite pensé spontanément à la loi binomiale de paramètres $N = 20$ et $p = 0,4$.

Deux difficultés sont apparues lors de la simulation sur tableur :

modéliser le choix d'un allèle g par le chiffre 1 et celui de 0 pour G.

comment utiliser la fonction alea.

Une élève a eu l'idée de taper « $\text{ent}(\text{alea} \times 10) + 1$ » mais ne voyait pas comment faire afficher 0 ou 1. Le professeur leur a parlé de la fonction « si...alors...sinon ».

Ils ont ensuite continué leur travail et calculé la fréquence de l'allèle g dans la génération nouvellement constituée.

Il a fallu encore quelques interventions du professeur pour les aider à terminer leur feuille de calcul et obtenir la courbe représentant la fréquence de l'allèle g en fonction du nombre de générations.

Le professeur de SVT a conclu en expliquant et en donnant des exemples de dérive génétique.

Les élèves ont été très intéressés, ils ont trouvé enrichissant la collaboration de deux enseignants de disciplines différentes et ont comparé les approches de chacun.

3) POUR LA CLASSE DE TERMINALE ES spécialité mathématiques

FICHE ELEVE

Premier exercice

Dans une population, dont le nombre total de personnes reste constant, on suppose qu'il y a des personnes malades, notées M et d'autres non malades donc saines notées S.

Chaque année, le pourcentage de passage de « Malade » à « Malade » est $p = 0,8$ et le pourcentage de passage de « Sain » à « Sain » est $q = 0,9$.

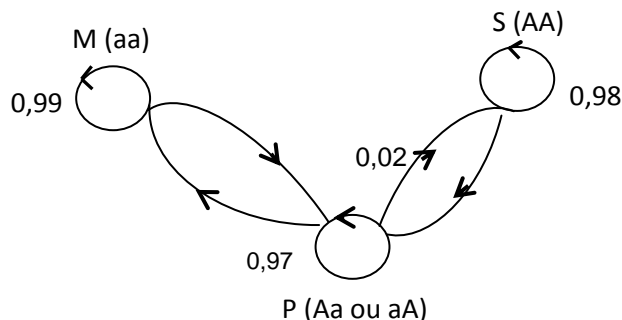
- 1) Construire le graphe complet qui correspond à cette situation.
(Vous déterminerez le pourcentage des personnes qui passent de l'état malade à l'état sain, puis celui des personnes qui passent de l'état sain à l'état malade).
- 2) Donner la matrice de ce graphe en précisant le sens concret des termes.
Soit $(a_0 \ b_0)$ l'état initial dans lequel a_0 indique le nombre de personnes malades et b_0 le nombre de personnes saines, l'année 1990.
On notera $(a_n \ b_n)$ l'état de cette population, l'année 1990 + n.
- 3) Déterminer les états en 1991, puis 1992, puis 1993 lorsque $a_0 = 100$ et $b_0 = 2000$.
- 4) On admet qu'il existe un état stable (a b). Déterminer cet état stable.

Deuxième exercice

On considère une maladie génétique ayant un caractère récessif :

une personne malade, notée M, aura le gène aa, une personne non malade, notée S, aura le gène AA, et une personne porteuse, notée P, aura le gène aA (ou Aa), une personne porteuse n'est pas malade mais peut transmettre le gène de la maladie.

On considère que d'une génération à la génération suivante, l'effectif de la population totale reste fixe et d'une génération à la génération suivante les pourcentages de passage d'un état à un autre sont donnés par le graphe suivant :



2) Compléter ce graphe.

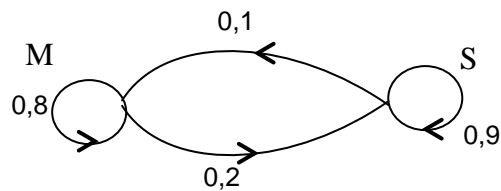
On notera $[m_n \quad s_n \quad p_n]$ la matrice ligne qui indique le nombre de malades, de sains, de porteurs, à la génération n .

L'année 2000, que l'on considère comme génération 0, il y a 1 personne sur 1000 de malades, soit 250 personnes malades, 1 sur 20 porteuses.

- 3) Calculer la population totale de cette région l'année 2000, en déduire le nombre de personnes porteuses et le nombre de personnes saines.
- 4) En utilisant un produit de matrices, donner la situation à la génération 1, puis à la génération 2.
- 5) On admet qu'il existe pour cette population un état stable, déterminer cet état stable.
- 6) En utilisant votre calculatrice et par essais successifs, essayez de déterminer au bout de combien de générations on obtiendrait cet état stable.

FICHE CORRECTION

Premier exercice



- La matrice de passage est : $\begin{bmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{bmatrix}$.

0.8 est le pourcentage de passage de M à M, 0.2 est le pourcentage de passage de M vers S, 0.1 le pourcentage de passage de S vers M et 0.9 le pourcentage de passage de S vers S.

- L'état l'année 1991 est

$$\begin{bmatrix} a_1 & b_1 \end{bmatrix} = \begin{bmatrix} a_0 & b_0 \end{bmatrix} \times \begin{bmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{bmatrix} = \begin{bmatrix} 100 & 2000 \end{bmatrix} \times \begin{bmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{bmatrix} = \begin{bmatrix} 280 & 1820 \end{bmatrix}.$$

L'état l'année 1992 est :

$$\begin{bmatrix} a_2 & b_2 \end{bmatrix} = \begin{bmatrix} a_1 & b_1 \end{bmatrix} \times \begin{bmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{bmatrix} = \begin{bmatrix} 280 & 1820 \end{bmatrix} \times \begin{bmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{bmatrix} = \begin{bmatrix} 406 & 1694 \end{bmatrix}$$

Et l'année 1993 : $\begin{bmatrix} a_3 & b_3 \end{bmatrix} = \begin{bmatrix} a_2 & b_2 \end{bmatrix} \times \begin{bmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{bmatrix} = \begin{bmatrix} 494.2 & 1605.8 \end{bmatrix}.$

- Si P est la matrice de passage, l'état stable $\begin{bmatrix} a & b \end{bmatrix}$ doit vérifier :

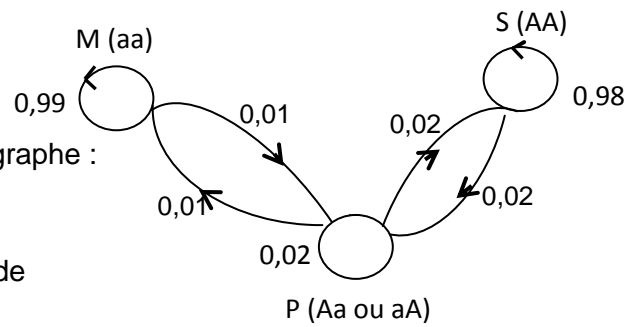
$$\begin{bmatrix} a & b \end{bmatrix} = \begin{bmatrix} a & b \end{bmatrix} \times P \Leftrightarrow \begin{bmatrix} a & b \end{bmatrix} = \begin{bmatrix} a & b \end{bmatrix} \times \begin{bmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{bmatrix} \text{ ou encore } \begin{cases} a = 0.8a + 0.1b \\ b = 0.2a + 0.9b \\ a + b = 2100 \end{cases}$$

car la population totale reste 2100.

On obtient alors $b=1400$ et $a=700$, ce qui signifie que lorsque l'état stable est atteint, il y a 1400 personnes Saines et 700 personnes Malades.

Deuxième exercice

- 1) La somme des poids des arêtes qui partent d'un même sommet doit être égale à 1, ce qui permet de compléter le graphe :



- 2) Puisque une personne sur 1000 est malade et qu'il y a 250 malades, il y a donc dans cette population 250 000 personnes en tout. Comme 1 personne sur 20 est porteuse de la maladie, il y a $250\,000/20=12\,500$ personnes porteuses. Le nombre de personnes saines est donc : $250\,000 - 12\,500 - 250 = 237\,250$.

- 3) A partir du graphe, on cherche la matrice de passage d'une génération à la génération suivante : Soit A cette matrice. On a

$$A = \begin{bmatrix} 0.99 & 0 & 0.01 \\ 0 & 0.98 & 0.02 \\ 0.01 & 0.02 & 0.97 \end{bmatrix} \text{ (on respecte l'ordre M, S, P) ;}$$

Si l'état initial est E_0 , on aura $E_1 = E_0 \times A$, de même $E_2 = E_1 \times A = E_0 \times A^2$.

On obtient

$$E_1 = [372.5 \quad 232755 \quad 16872.5] \text{ et } E_2 = [537.5 \quad 228437.35 \quad 21025.15].$$

(remarque : vérifier que la somme est toujours 250 000).

- 4) L'état stable E est celui qui doit vérifier

$$E = E \times A \Leftrightarrow [m \quad s \quad p] = [m \quad s \quad p] \times \begin{bmatrix} 0.99 & 0 & 0.01 \\ 0 & 0.98 & 0.02 \\ 0.01 & 0.02 & 0.97 \end{bmatrix} \quad m + p + s = 250000$$

- 5) D'où le système
$$\begin{cases} 0.99m + 0.01p = m \\ 0.98s + 0.02p = s \\ 0.01m + 0.02s + 0.97p = p \\ m + p + s = 250000 \end{cases} \Leftrightarrow m = p = s = \frac{250000}{3} \approx 83333.33.$$

Remarque : cet état stable ne dépend pas de l'état initial mais uniquement de la matrice de passage. Si les pourcentages de passage d'un état à un autre étaient différents, l'état stable ne donnerait pas forcément trois groupes de même taille !

Si le pourcentage de Porteur à Porteur est 0.96 au lieu de 0.97, et celui de Porteur à Sain devient 0.03 au lieu de 0.02, les autres valeurs restant identiques, le nouvel état stable serait $m=p=71\,428.57$ et $s=107\,142.86$, il y aurait d'avantage de personnes Saines.

Commentaires

Classe de terminale ES, spécialité mathématiques, 17 élèves. 2 heures consécutives le lundi après midi.

Les graphes probabilistes ont été expliqués la séance précédente, avec des exemples et un exercice était à finir pour ce jour.

Séance du lundi 18/05/2009 :

Correction de l'exercice qui était à chercher ce qui permet de faire un résumé des connaissances nécessaires aux trois élèves qui avaient été absents. Autre exercice sur les graphes probabilistes et correction.

Le texte des deux exercices est distribué.

Consigne :

Les exercices, bien rédigés, sont à faire sur feuille, certaines feuilles seront ramassées, mais ne seront pas notées. Les élèves peuvent demander de l'aide au professeur et aux proches voisins.

Exercice 1 :

Temps mis par les élèves : entre 30 minutes et 50 minutes.

Plusieurs élèves demandent s'il faut justifier les poids des arêtes inscrits sur le graphe.

Lorsque l'on demande la « matrice du graphe » est-ce la même chose que la matrice de passage ?

Que signifie « préciser le sens concret » ?

Est-ce qu'il existe une maladie pour laquelle certaines personnes restent malade d'une année sur l'autre et d'autres guérissent ?

Doute de certains élèves : l'état en 1991 correspond-il à a_1b_1 ?

Faut-il arrondir les résultats, et comment les arrondir ? aura-t-on toujours des résultats entiers ?

Pour la recherche de l'état stable, beaucoup de difficultés rencontrées par les élèves, soit ils se trompent dans les calculs, soit ils ont oublié la condition traduisant la population totale, soit ils ont écrit $a+b=1$. Aucun élève n'y est parvenu seul, même les deux meilleurs de la classe. (Mais dans les exercices rencontrés auparavant, a et b étaient des probabilités donc on avait alors $a+b=1$).

Un élève ne comprend pas que « multiplier b par $\frac{1}{2}$ c'est la même chose que diviser b par 2 ».

Un élève a écrit la matrice de passage avec comme 1^{er} terme de sain à sain, donc ses calculs à la fin sont faux car il n'a pas pensé à inverser a et b. pour éviter ce type d'erreur, on peut préciser dans le texte, ou donner le terme p_{11} de la matrice de passage.

Exercice 2 :

Un seul élève aura fini les deux exercices en 1 heure ; beaucoup n'ont fait dans le deuxième exercice que les questions 1 et 2 ainsi que la matrice de passage.

- a) Plusieurs élèves n'arrivent pas à calculer la population totale : le professeur doit les aider en rappelant la proportionnalité.
- b) Une élève pose la question : « le 1/20 est 1 sur 20 malades ». Le professeur doit réexpliquer ce qu'est un « porteur ».
- c) Pour les 2 ou trois élèves qui cherchent l'état stable, certains ont encore écrit $a+b+c=1$.

Remarques générales :

* les élèves n'avaient pas encore bien assimilés le sens des calculs faits, il faudrait donc insister sur ce point.

* ils ont mieux travaillé que les autres jours car leur travail pouvait être ramassé, et cependant ils pouvaient recevoir de l'aide par l'enseignant.

* une dizaine d'élèves sur les 17 ont beaucoup de mal dans les calculs, exemple : $a - 0.2a = ?$ ou $0.1b = 0.2a \Leftrightarrow ?$

* le texte doit être très précis pour les élèves.

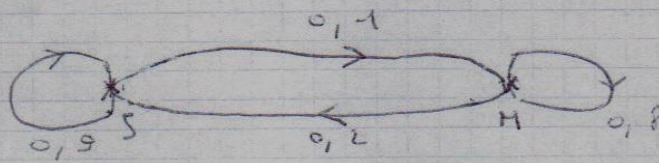
Conclusion :

Il est indispensable de donner à la fois du travail d'entraînement aux calculs, très souvent et ramassé, et du travail rédigé dans lequel les élèves sont obligés de trouver le sens de leurs calculs. Même en classe, en aidant les élèves, on peut demander un travail à rendre. Seul problème : avons-nous le temps de le faire ?...

Elève qui avait été absent le cours précédent -

Exercice 1

1)



2)

	M	S
M	0,8	0,2
S	0,1	0,9

- 0,8 est la probabilité de passage de "Malade" à "Malade"
- 0,2 est la probabilité de passage de "Malade" à "Sain"
- 0,1 est la probabilité de passage de "Sain" à "Malade"
- 0,9 est la probabilité de passage de "Sain" à "Sain".

3)

C-6 2

Année 1990 + 1.

$$\begin{aligned}
 (a_1 \quad b_1) &= (a_0 \quad b_0) \times M \\
 &= (100 \quad 2000) \times \begin{pmatrix} 0,8 & 0,2 \\ 0,1 & 0,9 \end{pmatrix} \\
 &= \begin{pmatrix} 100 \times 0,8 + 2000 \times 0,1 & 100 \times 0,2 + 2000 \times 0,9 \end{pmatrix} \\
 &= \begin{pmatrix} 80 + 200 & 20 + 1800 \end{pmatrix} \\
 &= \begin{pmatrix} 280 & 2000 \end{pmatrix}
 \end{aligned}$$

Année 1990 + 2.

$$\begin{aligned}
 (a_2 \quad b_2) &= (a_0 \quad b_0) \times M^2 \\
 &= (100 \quad 2000) \times \begin{pmatrix} 0,8 & 0,2 \\ 0,1 & 0,9 \end{pmatrix}^2 \\
 &= \begin{pmatrix} 406 & 1694 \end{pmatrix}
 \end{aligned}$$

Année 1990 + 3

$$\begin{aligned}
 (a_3 \quad b_3) &= (a_0 \quad b_0) \times M^3 \\
 &= (100 \quad 2000) \times \begin{pmatrix} 0,8 & 0,2 \\ 0,1 & 0,9 \end{pmatrix}^3 \\
 &= \begin{pmatrix} 494,8 & 1605,8 \end{pmatrix}
 \end{aligned}$$

$$4) (a \quad b) = (a \quad b) \times M \quad (a \quad b) \times \begin{pmatrix} 0,8 & 0,2 \\ 0,1 & 0,9 \end{pmatrix}$$

$$\begin{cases} a = 0,8a + 0,1b \\ b = 0,2a + 0,9b \\ a + b = 2100 \end{cases} \quad \left. \begin{array}{l} \text{l'équation} \\ \text{est} \\ \text{"bloquée"} \end{array} \right\} \quad \begin{array}{l} a - 0,8a = \\ a(1 - 0,8) \end{array}$$

fait par
le
professeur

$$\begin{cases} 0,2a = 0,1b \\ 0,1b = 0,2a \\ a + b = 2100 \end{cases} \quad \begin{cases} a = \frac{0,1}{0,2} b \\ a + b = 2100 \end{cases} \quad \begin{cases} a = \frac{1}{2} b \\ a + b = 2100 \end{cases}$$

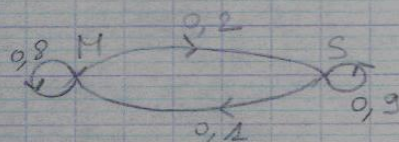
$$\begin{cases} a = \frac{1}{2} b \\ \frac{1}{2} b + b = 2100 \end{cases} \quad \begin{cases} a = \frac{1}{2} b \\ \frac{3}{2} b = 2100 \end{cases} \quad \begin{cases} a = 700 \\ b = 2100 \times \frac{2}{3} = 1400 \end{cases}$$

N'a pas eu le temps de faire le 2ème exercice
sauf le graphique

mathématiques spé. Lundi 18 mai

Premier exercice

1.



La somme des probabilités notées sur les arêtes qui partent d'un même noeud doit valoir 1.

Donc, on en déduit les probabilités manquantes sur le graphe.

2.

$$P = \begin{array}{c|cc} & M & S \\ \hline M & 0,8 & 0,2 \\ S & 0,1 & 0,9 \end{array}$$

• Le pourcentage des personnes qui passent de l'état "malade" à "malade" est de 80%.

• Le pourcentage des personnes qui passent de l'état "sain" à "sain" est de 90%.

• La probabilité des personnes qui passent de l'état "malade" à "sain" est de 0,2.

• Celle des personnes qui passent de l'état "sain" à "malade" est de 0,1.

3. Afin de déterminer la matrice $[a_1 \ b_1]$, $[a_2 \ b_2]$ et $[a_3 \ b_3]$ il faut calculer :

$$\begin{aligned} [a_1 \ b_1] &= [a_0 \ b_0] \times P^1 \\ &= [100 \ 2000] \times \begin{bmatrix} 0,8 & 0,2 \\ 0,1 & 0,9 \end{bmatrix} \\ &= [280 \ 1820] \end{aligned}$$

En 1991, il y a 280 personnes malades et 1820 saines.

$$[a_2 \ b_2] = [100 \ 2000] \times \begin{bmatrix} 0,8 & 0,2 \\ 0,1 & 0,9 \end{bmatrix}^2$$

$$= [406 \ 1694]$$

En 1992, il y a 406 personnes malades et 1694 saines.

$$[a_3 \ b_3] = [a_0 \ b_0] \times \begin{bmatrix} 0,8 & 0,2 \\ 0,1 & 0,9 \end{bmatrix}^3$$

$$= [494,2 \ 1605,8]$$

En 1993, il y a 494 personnes malades et 1605 saines.

4. $[a \ b] = [a \ b] \times \begin{bmatrix} 0,8 & 0,2 \\ 0,1 & 0,9 \end{bmatrix}$

$$\begin{cases} 0,8a + 0,1b = a \\ 0,2a + 0,9b = b \\ a + b = 2100 \end{cases}$$

très bien

$$\begin{cases} -0,2a = -0,1b \\ -0,1b + 0,2a = 0 \\ a = 2100 - b \end{cases}$$

$$\begin{cases} -0,2(2100 - b) + 0,1b = 0 \\ 0,2a - 0,1(2100 - a) = 0 \\ b = 2100 - a \end{cases}$$

$$\begin{cases} -420 + 0,2b + 0,1b = 0 \\ 0,2a - 210 + 0,1a = 0 \end{cases}$$

$$\begin{cases} b = \frac{420}{0,3} = 1400 \\ a = \frac{210}{0,3} = 700 \end{cases}$$

$$[700 \ 1400]$$

donner une phrase
pour conclure.

Deuxième exercice

1. Sur graphe
La somme des probabilités sur les arêtes qui partent d'un même nœud doit valoir 1.

$$1 = 0,99 + 0,01$$

$$1 = 0,98 + 0,02$$

2. On sait que $\frac{1}{1000}$ sont malades et

que ces personnes sont 250.

$$\frac{1}{1000} \times x = 250$$

$$x = 250\,000$$

Il y a donc 250 000 personnes au total.

On sait que $\frac{1}{20}$ personnes est porteuse, donc :

$$\frac{1}{20} \times 250\,000 = 12\,500$$

Il y a donc 12 500 personnes porteuses.

$$250\,000 - (250 + 12\,500) = 237\,250$$

Il y a donc 237 250 personnes saines.

- 3.

$P =$

	M	S	P
π	0,99	0	0,01
S	0	0,98	0,02
p	0,01	0,02	0,97

a_0 représente les personnes malades.
 b_0 représente les personnes porteuses
 c_0 représente les personnes saines.

$$a = 250$$

$$b = 12\,500$$

$$c = 237\,250$$

malade

Vous avez
infecté s et p

le texte dit →

m

s

p

$[a_1$

b_1

$c_1]$

$= [250$

$12\,500$

$237\,250]$

\times

$\begin{bmatrix} 0,99 & 0 & 0,91 \\ 0 & 0,98 & 0,02 \\ 0,01 & 0,02 & 0,97 \end{bmatrix}$

\times

$\begin{bmatrix} 0,99 & 0 & 0,91 \\ 0 & 0,98 & 0,02 \\ 0,01 & 0,02 & 0,97 \end{bmatrix}$

\times

$\begin{bmatrix} 0,99 & 0 & 0,91 \\ 0 & 0,98 & 0,02 \\ 0,01 & 0,02 & 0,97 \end{bmatrix}$

\times

$\begin{bmatrix} 0,99 & 0 & 0,91 \\ 0 & 0,98 & 0,02 \\ 0,01 & 0,02 & 0,97 \end{bmatrix}$

\times

$\begin{bmatrix} 0,99 & 0 & 0,91 \\ 0 & 0,98 & 0,02 \\ 0,01 & 0,02 & 0,97 \end{bmatrix}$

erreurs

malade

sain

porteur

$= [2620$

16995

$230385]$

\times

$\begin{bmatrix} 0,99 & 0 & 0,91 \\ 0 & 0,98 & 0,02 \\ 0,01 & 0,02 & 0,97 \end{bmatrix}$

\times

$\begin{bmatrix} 0,99 & 0 & 0,91 \\ 0 & 0,98 & 0,02 \\ 0,01 & 0,02 & 0,97 \end{bmatrix}$

\times

$\begin{bmatrix} 0,99 & 0 & 0,91 \\ 0 & 0,98 & 0,02 \\ 0,01 & 0,02 & 0,97 \end{bmatrix}$

\times

$\begin{bmatrix} 0,99 & 0 & 0,91 \\ 0 & 0,98 & 0,02 \\ 0,01 & 0,02 & 0,97 \end{bmatrix}$

\times

$\begin{bmatrix} 0,99 & 0 & 0,91 \\ 0 & 0,98 & 0,02 \\ 0,01 & 0,02 & 0,97 \end{bmatrix}$

$[a_2$

b_2

$c_2]$

$= [250$

$12\,500$

$237\,250]$

\times

P^2

\times

$\begin{bmatrix} 0,99 & 0 & 0,91 \\ 0 & 0,98 & 0,02 \\ 0,01 & 0,02 & 0,97 \end{bmatrix}$

\times

$\begin{bmatrix} 0,99 & 0 & 0,91 \\ 0 & 0,98 & 0,02 \\ 0,01 & 0,02 & 0,97 \end{bmatrix}$

\times

$\begin{bmatrix} 0,99 & 0 & 0,91 \\ 0 & 0,98 & 0,02 \\ 0,01 & 0,02 & 0,97 \end{bmatrix}$

$= [4897,65$

$21262,8$

$223839,55]$

\times

$\begin{bmatrix} 0,99 & 0 & 0,91 \\ 0 & 0,98 & 0,02 \\ 0,01 & 0,02 & 0,97 \end{bmatrix}$

\times

$\begin{bmatrix} 0,99 & 0 & 0,91 \\ 0 & 0,98 & 0,02 \\ 0,01 & 0,02 & 0,97 \end{bmatrix}$

\times

$\begin{bmatrix} 0,99 & 0 & 0,91 \\ 0 & 0,98 & 0,02 \\ 0,01 & 0,02 & 0,97 \end{bmatrix}$

\times

$\begin{bmatrix} 0,99 & 0 & 0,91 \\ 0 & 0,98 & 0,02 \\ 0,01 & 0,02 & 0,97 \end{bmatrix}$

\times

$\begin{bmatrix} 0,99 & 0 & 0,91 \\ 0 & 0,98 & 0,02 \\ 0,01 & 0,02 & 0,97 \end{bmatrix}$

$[a_3$

b_3

$c_3]$

$= [250$

$12\,500$

$237\,250]$

\times

P^3

\times

$\begin{bmatrix} 0,99 & 0 & 0,91 \\ 0 & 0,98 & 0,02 \\ 0,01 & 0,02 & 0,97 \end{bmatrix}$

\times

$\begin{bmatrix} 0,99 & 0 & 0,91 \\ 0 & 0,98 & 0,02 \\ 0,01 & 0,02 & 0,97 \end{bmatrix}$

\times

$\begin{bmatrix} 0,99 & 0 & 0,91 \\ 0 & 0,98 & 0,02 \\ 0,01 & 0,02 & 0,97 \end{bmatrix}$

$= [7087,069$

$25314,335$

$217598,596]$

\times

$\begin{bmatrix} 0,99 & 0 & 0,91 \\ 0 & 0,98 & 0,02 \\ 0,01 & 0,02 & 0,97 \end{bmatrix}$

\times

$\begin{bmatrix} 0,99 & 0 & 0,91 \\ 0 & 0,98 & 0,02 \\ 0,01 & 0,02 & 0,97 \end{bmatrix}$

\times

$\begin{bmatrix} 0,99 & 0 & 0,91 \\ 0 & 0,98 & 0,02 \\ 0,01 & 0,02 & 0,97 \end{bmatrix}$

\times

$\begin{bmatrix} 0,99 & 0 & 0,91 \\ 0 & 0,98 & 0,02 \\ 0,01 & 0,02 & 0,97 \end{bmatrix}$

\times

$\begin{bmatrix} 0,99 & 0 & 0,91 \\ 0 & 0,98 & 0,02 \\ 0,01 & 0,02 & 0,97 \end{bmatrix}$

non

demandé

dans la matrice de passage, l'ordre
 était m s p donc vous devez
 respecter cet ordre aussi dans vos calculs

VII. CONCLUSION

Lorsque nous avons débuté cette unité sur la modélisation, nous étions perplexes et nous ne savions pas vraiment ce que devions réaliser.

Et puis, petit à petit, le projet a pris forme, les idées sont venues, et après de nombreuses recherches, questionnements, mises en commun et conseils de nos enseignants, nous avons réalisé ce document.

Et voici le temps du bilan !

Un bilan positif à tout point de vue, aussi bien personnel que professionnel.

Personnel, car nous avons eu le plaisir de travailler en groupe, alors que nous ne nous connaissions pas réellement. Tout d'abord il a fallu s'écouter pour s'orienter ensemble vers un projet de biologie, étudier les propositions de chacune, s'y intéresser et rechercher l'unité du groupe dans le respect des goûts et des savoirs de chacune ! Une fois les compétences biologiques débroussaillées, la distribution des sous-parties d'étude s'est faite de façon très naturelle et nous avons eu à cœur de partager les unes avec les autres le fruit de notre travail plus personnel, dans le respect des objectifs fixés par nos enseignants.

Nous avons pu nous enrichir sur le plan des connaissances dans un domaine, la génétique, que nous ne connaissions pas et que n'aurions sans doute pas abordé sans ce travail.

Nous avons développé une curiosité sur les mathématiques utilisées dans situations quotidiennes.

D'un point de vue professionnel, cette expérience nous a permis de vivre la richesse du travail en groupe, si difficile à mettre en œuvre dans nos établissements scolaires mais si fructueuse pour chacun !

Nous en sommes convaincues et nous aurons peut-être plus d'audace pour le provoquer à l'avenir ou tout du moins saisir la moindre occasion de pratiquer le travail à plusieurs.

Travailler sur la modélisation nous a permis de faire des mathématiques différentes de celles que nous enseignons habituellement, de découvrir et/ou approfondir la connaissance de logiciels tels que matlab.

Nous comprenons mieux les difficultés des élèves face à la recherche d'un sujet de modélisation, en TPE par exemple.

Ce travail nous a permis de repérer des exercices à traiter sous forme de modélisation plutôt que sous forme d'exercices guidés, de faire passer une notion mathématique en la reliant à une situation modélisable et par là d'avoir une lecture plus pertinente des thèmes de convergences proposés au niveau collège.

Nous avons pris conscience que nous aimions les mathématiques pour les mathématiques, peu importait l'application possible, or nous avons découvert avec ce travail de modélisation, que les applications peuvent être aussi intéressantes mathématiquement.

Les mathématiques enseignées sont le plus souvent tournées sur elles-mêmes, sans relation avec les autres disciplines et d'ailleurs combien de collègues prennent connaissance du contenu des autres matières que la leur ?

Cependant nous avons ressenti aussi une certaine frustration par rapport aux mathématiques et aux Sciences de la Vie et de la Terre lorsque nous n'avons pu comprendre certains articles et que nous avons ressenti l'énorme distance qu'il existe entre chercheurs, et nous, professeurs du secondaire.

Nous sommes convaincues que ce type de travail ne peut que rejaillir sur notre enseignement, et être bénéfique pour les élèves. En effet, il permet d'avoir une autre vision plus large des mathématiques, de développer la curiosité et la motivation des élèves grâce à un travail interdisciplinaire permettant de donner de la réalité aux mathématiques. La démarche de la modélisation constitue un changement de cadre, à la fois source de difficultés et de motivation, autre que les habituels passages géométrie/algèbre. Ce changement permet de montrer aux élèves combien la maîtrise de l'outil mathématique est exploitable au service de la compréhension et de la résolution de problèmes concrets, dans des domaines divers.

Peut-être serons-nous, enfin, davantage capables de répondre à l'inévitable question des élèves : « à quoi ça sert les maths ? ».

Mais il est évident que ce seul travail n'est pas suffisant pour aborder sereinement une démarche de modélisation en classe.

Pour permettre une véritable ouverture d'esprit, l'acquisition de connaissances dans diverses disciplines afin de travailler avec les enseignants d'autres matières, il est souhaitable que le travail sur la modélisation soit non seulement pris en compte au sein de la formation initiale des enseignants, mais aussi, bien sûr en formation continue, aussi bien par le biais du PAF, que par celui des Conseillers Pédagogiques.

A nous aussi d'insuffler l'envie de travailler sur la modélisation et de susciter l'enthousiasme auprès de nos collègues !

VIII. REMERCIEMENTS

Nous remercions :

Michèle Artigue, Guy Rumelhard et tous les intervenants pour leur encadrement, leur disponibilité, leurs conseils et leur enthousiasme communicatif,

Evelyne Heyer pour l'envoi des documents concernant ses recherches.

IX. REFERENCES / BIBLIOGRAPHIE

Articles

- Articles de Frédéric Austerlitz et Evelyne Heyer :
 - dans "The National Academic of Science", décembre 1998, vol 95, 5140-15144.
 - dans " Genetic Epdidemioloy", 1999, vol 16, 2-14.
 - dans "Human Journal of Huamn Genetics", 2 000, vol 8, 980-985.
- Revue GEO n°361 mars 2009.
- Epreuve de modélisation – Agrégation externe de Mathématiques – Session 2008 - Université de Rennes I.

Livres

- "Génétique des populations" de Jean-Louis Serre aux éditions Dunod.
- « La mathématisation du réel » de Giorgio Israel aux éditions du Seuil.
- Biomathématiques Pharma 1 par Bénazeth, Boniface, Demarquilly, Lasserre, Lemdani et Nicolis chez Masson.
- Sciences de la Vie et de la Terre terminale S spécialité chez Didier.
- Sciences de la Vie et de la Terre terminale S obligatoire chez Bordas.

Sites Internet

- « comportements démographiques et maladies héréditaires ».
<http://www.cnrs.org/Cnrspresse/n372a2.htm>
- définition de la pathologie d'Ataxie spastique récessive de Charlevoix-Saguenay.
<http://www.geocities.com/gudes111/definition.htm?20097>
- « le gène de l'ataxie de Charlevoix-Saguenay identifié » et le professeur Bouchard.
<http://www.scom.ulaval.ca/au.fil.des.evenements/2000/04.27/ataxie.html>
- Un article de la Revue Accromath sur une application des chaines de Markov :
« Récidivisme » de M. Laforest et A. Ross (École Polytechnique de Montréal et Cégep de Lévis-Lauzon) dans le n° 3 Hiver Printemps Année 2008.
<http://accromath.uqam.ca>

Biographies



➤ Ronald Fisher

Professor Sir Ronald Aymler Fisher, F.R.S. (17 février 1890 — 29 juillet 1962) était un biologiste et statisticien britannique, « le plus grand des successeurs de Darwin »[1] et l'homme qui a - « presque d'une seule main - fondé les statistiques modernes » [2]. Ses travaux sur les statistiques lui valurent la médaille Darwin en 1948, la médaille Copley en 1955 et la médaille d'argent Darwin-Wallace en 1958.

Dans le domaine des statistiques, il a introduit de nombreux concepts clés tels que le maximum de vraisemblance, l'information de Fisher et l'analyse de la variance (ou ANOVA). Il est considéré comme un grand précurseur de Shannon.

Il est également un des fondateurs de la génétique moderne et un grand continuateur de Darwin, en particulier grâce à son utilisation des méthodes statistiques, incontournables dans la génétique des populations. Il a ainsi contribué à la formalisation mathématique du principe de sélection naturelle (Wikipedia).



➤ Sewall Wright

Green Sewall Wright (Décembre 21, 1889 - Mars 3, 1988) était un Américain généticien connu pour son travail d'influence sur la théorie de l'évolution et aussi pour ses travaux sur la voie d'analyse. Avec RA Fisher et JBS Haldane, il a été l'un des fondateurs de la théorie de la génétique des populations. Il est le découvreur du coefficient de consanguinité et de méthodes de calcul dans les pedigrees. Il a étendu ce travail à la population, le calcul du montant de la consanguinité des membres des populations à la suite aléatoire de la dérive

génétique, et il Fisher et pionnier des méthodes de calcul de la répartition des fréquences des gènes entre les populations à la suite de l'interaction de la sélection naturelle, mutation, de migration et de dérive génétique. Les travaux de Fisher, Wright, Haldane et théoriques de la génétique des populations a été une étape importante dans le développement de la synthèse moderne de l'évolution de la génétique avec l'évolution. Wright a également fait d'importantes contributions à la génétique et de mammifères génétique biochimique (Wikipedia).



➤ **Godfrey Harold Hardy**

Mathématicien anglais, né à Granleigh (1877), dans le Surrey, et mort à Cambridge (1947). Godfrey Harold Hardy fit ses études au Trinity College de Cambridge, où enseigna de 1906 à 1919. En 1908, il découvre, en il même temps que le physicien W. Weinberg, mais indépendamment de lui, la loi de Hardy-Weinberg, qui décrit l'équilibre génétique au sein d'une population et qui aura une très grande importance pour l'étude des facteurs Rhésus dans les groupes sanguins. (Encyclopedie Universalis)

1) **Wilhelm Weinberg**

Médecin allemand (1862-1937), il consacra l'essentiel de son temps à s'occuper des pauvres. Il publia 160 articles scientifiques en langue allemande qui n'attirèrent guère l'attention des généticiens – le plus souvent anglophones. Mais quelques années plus tard, ses articles devinrent des références. En 1908, il découvre ce qu'on appelle aujourd'hui la loi de Hardy Weinberg. (Cerveau-Psycho N°33)