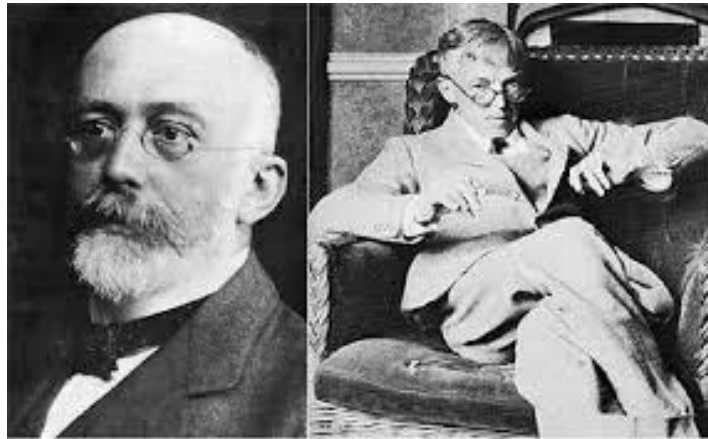


# Le modèle de Hardy-Weinberg face à la réalité

## Projet scientifique du cycle terminal du lycée



Origines du projet et description du modèle	p.2
I. Notions du programme abordées et compétences développées	p.3
II. Le devoir maison de mathématiques	p.4
III. Vers la simulation	p.7
1. Prise en main du langage Python	p.8
2. Hardy Weinberg et populations de taille finie	p.9
a. Le TP des tortues	p.10
b. Conception de l'algorithme	p.11
c. Résultats et analyses	p.12
IV. Annexes	p.16



## Origines du projet

Le modèle de Hardy-Weinberg (1908) est une théorie de génétique des populations qui postule qu'au sein d'une "population idéale", les fréquences alléliques et génotypiques sont stables au cours des générations.

Pour simplifier, on considère qu'un gène peut prendre deux formes (ou allèles) A et a.

Chaque gène se trouvant en deux exemplaires, un individu peut donc présenter l'un des trois génotypes suivants : AA, Aa ou aa.

Un enfant hérite d'un allèle de chacun de ses parents, chaque allèle étant choisi au hasard. Ainsi, si par exemple le père est de type AA et la mère de type Aa, alors les enfants peuvent être du type AA ou Aa.

On entend par population idéale, le fait que dans cette population les couples se forment au hasard quant à leurs génotypes, que l'environnement est stable et que la population est suffisamment grande pour être considérée comme infinie.

L'évolution des proportions des différents génotypes et allèles dans la population est donc, selon ce modèle, constante au cours des générations.

Mathématiquement, cette propriété peut être justifiée assez facilement à l'aide des probabilités conditionnelles abordées en terminale. C'est dans ce cadre que nombre de professeurs de mathématiques proposent l'étude de ce modèle en application, et éventuellement en interdisciplinarité.

Ce projet est né en 2012, alors que mes collègues de SVT (B. Hare, F. Lauvray) me sollicitaient pour simuler sur tableur les notions de dérive génétique et de sélection naturelle rencontrées dans un TP de terminale S, TP dit "des tortues". Il y est question de mettre en évidence l'évolution aléatoire de la fréquence des allèles dans une population au cours des générations, ainsi que la pression de l'environnement sur cette fréquence.

Parallèlement, en mathématiques, nous faisons l'étude du modèle de Hardy-Weinberg en exercice d'application, mais sans faire le lien avec les notions du TP.

Avec une collègue de mathématiques (E. Rocher) et d'autres collègues de SVT (M. Farina, A. Juras) nous avons travaillé, d'une part sur la simulation du TP à l'aide du tableur, et d'autre part sur la confrontation du modèle de H. W. à la réalité. En particulier, un questionnaire est apparu sur le fait que les résultats obtenus dans le TP sur la notion de dérive génétique semblaient en contradiction avec le modèle.

Il y a deux ans, s'est constitué le groupe de travail Math-SVT de l'IREM Paris Nord où j'entrepris de simuler le modèle en langage Python, langage suggéré par les nouvelles directives des programmes et qui se révéla très adapté à ce type de simulation.

C'est en élargissant les simulations Python dans le contexte du TP des tortues que j'ai pu faire le lien entre ces notions en apparence contradiction.

Quoi de mieux que de faire vivre cette expérience aux élèves dans nos classes et montrer que la simulation est une composante essentielle de l'activité mathématique.

La première année, nous avons développé ce projet au lycée P. Picasso de Fontenay sous-bois avec A. Juras et au collège de la Pléiade de Sevran avec A. Boccaccini, R. Estavoyer et P. Coelho. Cette année nous avons reconduit le projet avec V. Halleux au lycée P. Picasso et la même équipe au collège de la Pléiade.

A noter enfin que ce projet sera reconduit l'an prochain en terminale S sous cette forme pour la dernière fois. Cependant, après quelques adaptations, il pourra tout à fait être repris les années suivantes dans le cadre de l'enseignement scientifique de la voie générale du lycée.

JL Marcia

## I. Notions du programme abordées

### MATHEMATIQUES

**Loi des grands nombres**

**Loi binomiale**

**Conditionnement, indépendance**

- Arbre pondéré
- Formule des probabilités totales

**Intervalle de confiance**

**Suites numériques – Raisonnement par récurrence**

**Algorithmique** : langage de programmation Python

- Instructions élémentaires d'entrée et de sortie
- Boucles bornées et non bornées
- Instructions conditionnelles
- Fonctions

### SVT

**Notions à construire** : “ Sous l'effet de la pression du milieu, de la concurrence entre êtres vivants et du hasard, la diversité des populations change au cours des générations. L'évolution est la transformation des populations qui résulte de ces différences de survie et du nombre de descendants. ”

**Capacités à travailler** :

- Pratiquer une démarche scientifique (observer, questionner, formuler une hypothèse, expérimenter, raisonner avec rigueur, modéliser).
- Comprendre le lien entre les phénomènes naturels et le langage mathématique

### Compétences développées

**Chercher** : Mettre en œuvre une recherche de façon autonome. Observer des supports de représentation variés pour conjecturer des propriétés

**Modéliser** : Description d'un modèle, simulations

**Représenter** : Représentation d'un modèle, arbre de probabilité, représentations graphiques

**Raisonner** : Confrontation du modèle à la réalité, preuve de la stabilisation allélique et génotypique

**Calculer** : Calcul de probabilités, calcul de fréquences, intervalle de confiance

**Communiquer** : Devoir maison, compte rendu des simulations

## II. Le devoir maison de mathématiques

Le modèle théorique de H.W. est présenté et développé à travers un devoir maison de mathématiques qui est donné aux élèves après avoir abordé le chapitre relatif aux probabilités conditionnelles.

Il y a deux façons de mettre en place le modèle.

La première, plus naturelle pour les élèves, fait l'objet de la version 1 du devoir ([annexe 1](#)). Cette version a été construite à partir d'un exercice proposé dans le livre (p. 373) *Hyperbole de terminale S* (programme 2012) chez Nathan.

Dans un premier temps, on met en place la stabilisation génotypique au cours des générations. Le raisonnement repose sur la rencontre au hasard de deux individus de la population et à partir de leurs génotypes possibles (Fig 1), la formule des probabilités totales donne la distribution génotypique chez l'enfant.

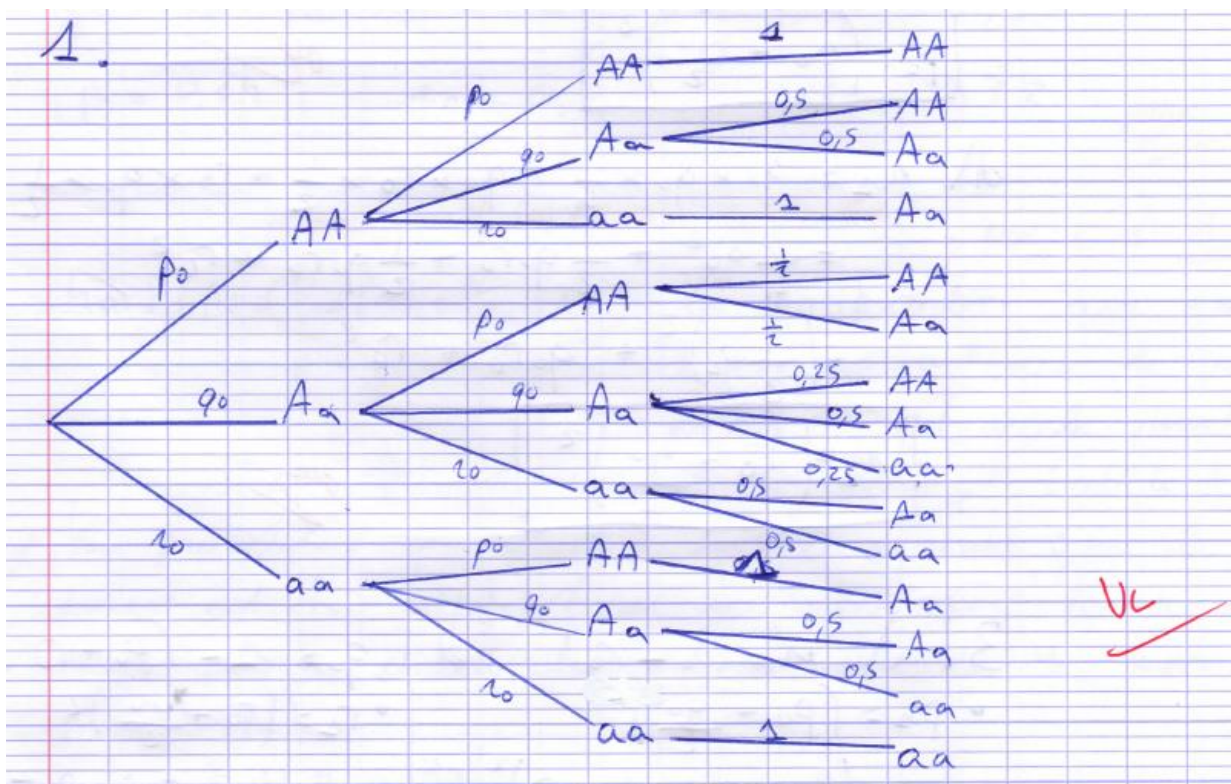


Fig. 1 : Arbre de Lydia

De cette nouvelle génération, on réitère le raisonnement pour obtenir la distribution génotypique de la génération suivante et on constate qu'elle est identique à la précédente.

Reste ensuite à faire le lien entre distribution génotypique et distribution allélique pour conclure également à sa stabilisation au cours des générations.

Le calcul algébrique y tient une place importante, ce qui est l'occasion de faire progresser les élèves dans ce domaine même si on peut reprocher à cette version la lourdeur des calculs.

L'utilisation du tableur (Fig 2), montre aux élèves son importance dans la recherche en l'occurrence pour faire apparaître des propriétés et émettre des conjectures.



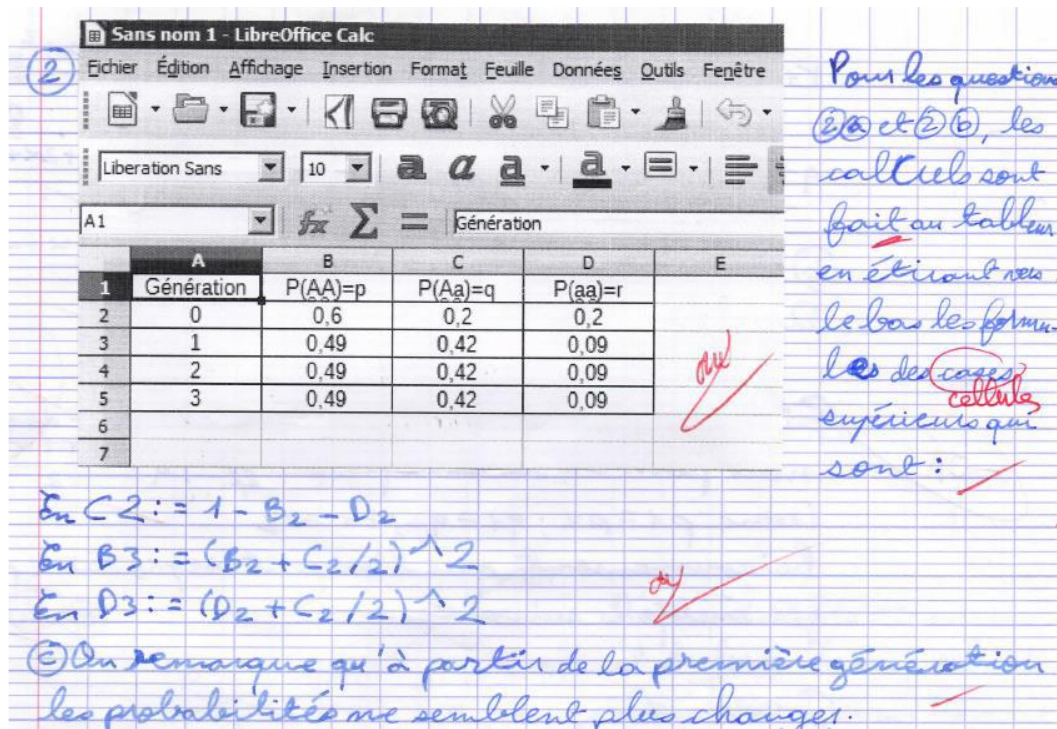


Fig. 2 : Simulation de Mohamed

La fin du devoir est consacrée à la confrontation du modèle théorique à la réalité, ce qui constituera une étape essentielle de notre projet, à savoir la mise en place des hypothèses de travail permettant de simuler et d'affiner le modèle.

Pour répondre à cette dernière partie, les élèves ne disposent pas encore des notions du programme de SVT de terminale. Ils peuvent cependant s'appuyer sur les connaissances de collège et de la classe de seconde sur le thème de l'évolution des populations. Cela signifie que les notions de populations, sélection naturelle et dérive génétique ont déjà été vues.

L'analyse des réponses des élèves peut se faire en trois axes :

- (Copie 1 et 2) Pour formuler des hypothèses sur les mécanismes faisant varier les fréquences alléliques, les élèves doivent sortir du raisonnement purement mathématique pour replacer ces variables dans un contexte biologique : l'évolution des populations. Cette contextualisation montre une mauvaise appropriation du vocabulaire, en particulier le terme "population" souvent synonyme pour beaucoup d'élève de "population humaine".
- (Copie 2) Les deux tiers des réponses des élèves révèlent qu'ils utilisent une recherche sur internet pour répondre à cette question. On trouve ainsi des noms de mécanismes tels que "sélection naturelle", "dérive génétique" ou encore "dépression consanguine". Si certains termes sont sélectionnés par les élèves car faisant appel à leur souvenir des classes antérieures, aucun d'eux n'est véritablement expliqué et mis en lien avec l'évolution des fréquences alléliques ou génotypiques.
- (Copie 2, 3 et 4) Dans un tiers des copies, les élèves ont le réflexe de revenir à l'introduction du DM donnant les conditions initiales du modèle afin de les rediscuter. Même si cette analyse est souvent non aboutie, elle montre que les élèves savent faire preuve d'esprit critique.

### Copie 1

c) Il se peut qu'il y ~~ait~~<sup>ait</sup> des écarts car dans le modèle, la population est considérée comme infinie mais en réalité même si nous sommes 7 milliards, on ne rencontre pas 7 milliards d'êtres humains dans sa vie. Donc la population

### Copie 2

c) La première hypothèse consiste en le fait que l'on considère cette étude sur une population suffisamment grande pour être considérée comme infinie, or la population humaine n'est pas infinie.  
Seconde hypothèse, plusieurs effets qui peuvent expliquer cet écart tel que :  
( la dépression de la consanguinité ou encore la dérive génétique.

Qui jouent sur quoi ?

### Copie 3

c) Il s'agit d'un modèle théorique, cependant dans la pratique, une population n'est pas infinie et les couples se forment majoritairement au sein d'une même population. Des fréquences alléliques peuvent varier selon la région du monde. - le modèle est trop simple, il y a donc forcément des écarts selon la population étudiée. Dès qu'un couple se forme les fréquences changent. des mutations génétiques peuvent aussi expliquer cet écart



(C) On a remarqué dans l'exercice II que lorsque  $q_0$  ne changeait pas,  $p_1$  et  $p_2$  restaient les mêmes. Ainsi on peut mettre d'hypothèse que ces résultats ne seraient pas les mêmes si  $q_0$  change. Celui-ci peut changer pour diverses raisons :

- environnement pas stable, et donc...
- population pas assez grande pour être supposée infinie, il y a donc un déséquilibre -

Corrigé de la version 1 ([annexe 2](#))

La version 2 ([annexe 3](#)) du devoir a l'avantage de supprimer la lourdeur des calculs algébriques évoqués précédemment. C'est d'ailleurs la démarche adoptée en général dans le supérieur lors de l'étude théorique du modèle en génétique des populations. Mais le raisonnement concernant la stabilisation allélique peut dérouter voire être source de confusion. En effet, on met en place la distribution génotypique de l'enfant à partir de la rencontre au hasard de deux allèles dans la population. Cette approche s'affranchit des individus et considère la reproduction comme une "rencontre de deux allèles issus de la population". Cela se justifie par la notion d'*indépendance d'événements*, indispensable ici et qui était simplement sous-entendue dans la version 1 du devoir mais qui, en y regardant de plus près, apparaissait dans le résultat simplifié des formules établies.

En effet, dans cette version 1 on établit la formule :  $p_1 = \left(p_0 + \frac{q_0}{2}\right)^2$ , qui traduit le fait que

$P(\text{enfant AA}) = P(A) \times P(A)$ , en d'autres termes, la probabilité que le génotype de l'enfant soit AA est égale au produit des probabilités que chaque allèle "choisi" soit A. Ce qui prouve l'indépendance et valide le raisonnement.

D'autre part, la version 2 privilégie la notion de suites numériques et leur donne ainsi toute légitimité aux yeux des élèves.

Enfin, l'utilisation du tableur et la confrontation du modèle théorique à la réalité y sont développées de la même manière.

Corrigé de la version 2 ([annexe 4](#))

### III. Vers la simulation

La dernière question posée dans le devoir de mathématiques motive l'emploi de la simulation. En effet les élèves auront émis des hypothèses (au sens expérimental) qui expliqueraient pourquoi les observations effectuées sur certaines populations ne mènent pas à cette stabilisation génotypique et allélique du modèle. Comment la simulation peut-elle nous permettre d'améliorer le modèle ?

En particulier l'hypothèse (au sens mathématique) "de population infinie" qui n'est évidemment pas vérifiée dans la réalité suffirait-elle à expliquer cet écart entre modèle et réalité ? Cette question est ambiguë pour les élèves, car le concept d'infini mathématique n'est naturellement pas acquis. Il est compris selon l'usage du commun : "ce qui est trop grand pour être compté, comme les étoiles dans le ciel ou les grains de sable dans le désert."

La question est donc du point de vue des mathématiques : à partir de quel effectif de population et sur quelle période d'observation peut-on considérer que cet écart n'est plus significatif ? Ou alors, en dessous de quels seuils considère-t-on cet écart significatif ?

Bien évidemment, d'autres hypothèses du modèle théorique pourront être modifiées pour coller au plus près de la réalité, comme par exemple la pression de l'environnement. Répondre à ces questions par la simulation nécessite une importante phase de modélisation et de conception algorithmique qui n'est pas encore à la portée de tous nos élèves. Il faut trouver le juste milieu entre autonomie laissée aux élèves et apports du professeur pour y parvenir "honnêtement". Voici la démarche adoptée et développée lors des séances d'Accompagnement Personnalisé (6 à 8 séances ou plus...) mais aussi sur le temps personnel des élèves. En effet le temps d'acquisition du langage et de conception des programmes est très variable selon les élèves et il est important de leurs laisser du temps entre les séances d'AP consacrées à ce thème.

## 1. Prise en main du langage Python

Cette prise en main doit se faire le plus tôt possible dans l'année, avant même la donnée du devoir maison, ce qui permet aux élèves d'être familiarisés avec le langage pour la phase de simulation à proprement dite.

La séquence 1 ([annexe 5](#)) distribuée aux élèves en début d'année propose l'installation sur clé USB d'une version portable de Python, "EDUPYTHON", ce qui rendra son utilisation très souple que ce soit dans l'établissement mais aussi au domicile de l'élève.

L'acquisition du langage reste une question importante. Il n'est pas question ici de fournir aux élèves une formation complète sur le langage Python. La méthode choisie à la fois simple et efficace leurs permet d'acquérir les bases nécessaires pour programmer rapidement dans ce langage.

En particulier, on propose dans les séquences 2 et 3 ([liens ci-dessous](#)), deux programmes écrits en langage Python et commentés ligne à ligne. Ceci permet de comprendre l'algorithme sous-jacent tout en commençant à se familiariser avec la syntaxe du langage. La présence du professeur est ici indispensable et c'est pendant les séances d'AP que sont étudiés ces programmes. On demande ensuite aux élèves de les modifier en se référant aux documents d'aide ([doc1](#) + [doc2](#)) conçus par J. Berhouet et A. Baraquin et réunissant les principales instructions du langage. Ce travail peut être commencé en AP et finalisé à la maison afin que chaque élève puisse aller à son rythme.

La séquence 2 ([annexe 6](#)) propose un programme renvoyant l'évolution sous forme de graphique, de la fréquence d'apparition de deux événements au fur et à mesure des tirages. L'urne contient quatre "0" et deux "1", les tirages sont faits avec remise et les événements considérés sont l'apparition du "0" et l'apparition du "1".

Dans un premier temps, les élèves analysent et exécutent ce programme (ce qui peut déjà prendre un certain temps). On leur demande de commenter le graphique obtenu ([Fig 3](#)). Dans cette question, le lien attendu avec la notion de probabilité desdits événements n'est pas une évidence pour tous. En particulier *la loi des grands nombres* reste encore "un peu" floue chez de nombreux élèves.

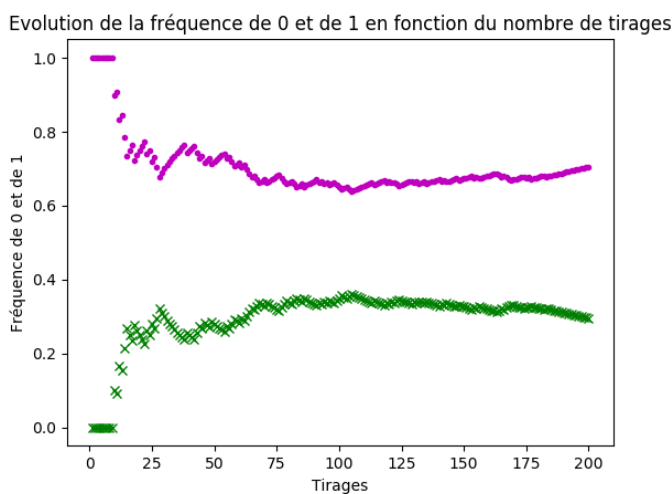


Fig. 3

Ensuite, il s'agit pour eux de réinvestir cette notion en modifiant le programme afin qu'il renvoie une approximation de la probabilité d'obtenir six "0" sur dix tirages.



Il s'agit donc de répéter un grand nombre de fois le programme précédent mais en fixant le nombre de tirages à 10, de supprimer les instructions relatives au graphique et de comptabiliser le nombre d'apparitions de l'évènement six "0". Cet exercice permet, non sans mal, de familiariser les élèves à la syntaxe du langage Python.

La fréquence renvoyée par le programme ainsi modifié, donne alors une approximation de la probabilité cherchée, ou plus exactement permet d'estimer cette probabilité par le calcul d'un intervalle de confiance à 95 %. Il n'est pas inutile de rappeler ici aux élèves les conditions d'utilisation vues en seconde d'un tel intervalle, à savoir  $n \geq 25$  et  $0,2 \leq f \leq 0,8$ , conditions qui sont ici remplies.

On termine la séquence par la détermination de cette probabilité en espérant que les élèves reconnaissent *la loi binomiale*, ce qui permet aussi de vérifier son appartenance à l'intervalle de confiance proposé.

Corrigé de la séquence 2 ([annexe 7](#))

La séquence 3 ([annexe 8](#)) propose un programme renvoyant la liste des tirages sans remise effectués dans l'urne précédente jusqu'à la vider. L'analyse de ce programme ne pose pas de difficulté mais la modification demandée nécessite une technique algorithmique peu utilisée par les élèves jusqu'ici, à savoir l'utilisation d'une variable tampon. En effet, il s'agit d'évaluer la probabilité d'obtenir la liste des tirages égale à l'urne initiale, urne initiale qui doit être "sauvegardée" puisque vouée à être vidée.

Après exécution du programme, on obtient un intervalle de confiance à 95 % de cette probabilité mais ici, il faut noter que les conditions vues plus haut ne sont pas remplies, en particulier la fréquence obtenue est trop faible (inférieure à 0,2). Il faut donc faire accepter aux élèves, quitte à anticiper sur la notion d'estimation du programme de terminale abordée en général en fin d'année, que le choix de  $n$  suffisamment grand palliera le problème.

Ensuite, la détermination de la probabilité nécessite une technique de dénombrement peu maîtrisée par nos élèves, en particulier l'utilisation des combinaisons, mais quelques élèves ont su s'en passer et ont proposé le bon résultat.

Enfin, on vérifie là encore l'appartenance de cette probabilité à l'intervalle de confiance proposé.

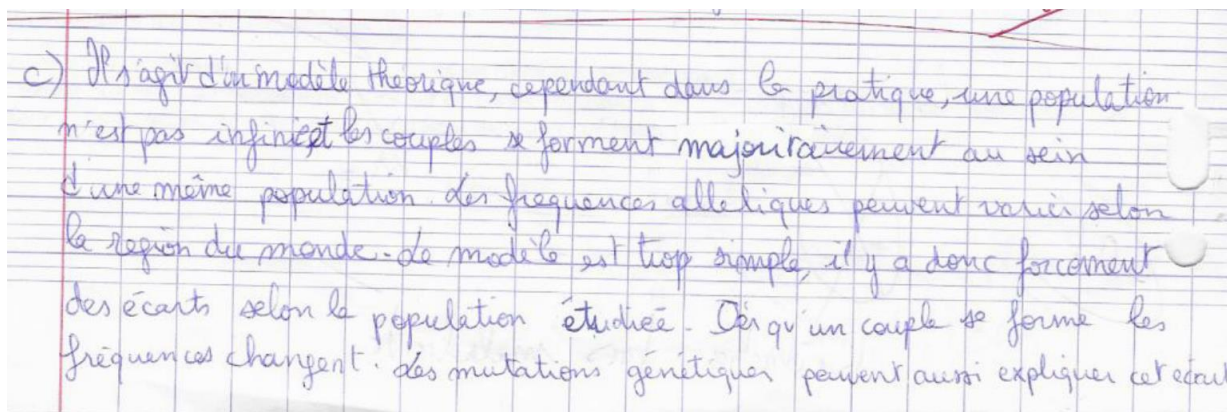
Corrigé de la séquence 3 ([annexe 9](#))

Ces séquences sont fondamentales pour les notions de mathématiques figurant au programme du lycée. Sur le plan algorithmique, on renforce les notions de boucles bornées et non bornées, et chose importante pour notre projet, on renforce la technique de détermination de la fréquence d'un évènement liée à l'évaluation d'une probabilité.

## 2. Hardy Weinberg et populations de taille finie

Les hypothèses émises par les élèves (*Fig. 4*) qui pourraient expliquer l'écart entre le modèle de H. W. et la réalité sont le point de départ de la simulation. Reste à vérifier leurs effets sur l'évolution des fréquences alléliques.

Une chose délicate est ici de faire accepter aux élèves que l'on ne peut pas considérer ces hypothèses simultanément car on ne saurait plus laquelle est à l'origine de tel ou tel effet. D'autre part, cela compliquerait la conception de l'algorithme.



c) Il s'agit d'un modèle théorique, cependant dans la pratique, une population n'est pas infinie et les couples se forment majoritairement au sein d'une même population. Des fréquences alléliques peuvent varier selon la région du monde. Ce modèle est trop simple, il y a donc forcément des écarts selon la population étudiée. Dès qu'un couple se forme les fréquences changent. Des mutations génétiques peuvent aussi expliquer cet écart.

Fig. 4 : Hypothèses d'Hugo

L'hypothèse la plus évidente pour eux a été d'évoquer le caractère fini de la population, ce qui revient à dire que le choix des couples d'individus est en réalité fait sans remise. Nous nous sommes donc consacrés dans un premier temps à cet aspect.

Si on pose a priori la question aux élèves : " Quel effet aura le choix sans remise sur l'évolution des fréquences alléliques ? ", la réponse implacable est bien souvent que les fréquences ne se stabiliseront pas ! Mais ils ont plus de mal à préciser leur pensée, et certains vont jusqu'à imaginer que le hasard compenserait finalement les effets et que globalement les fréquences alléliques se stabiliseraient de la même manière que dans le modèle théorique. Il est donc temps de passer à la simulation ...

### a. Le TP des tortues

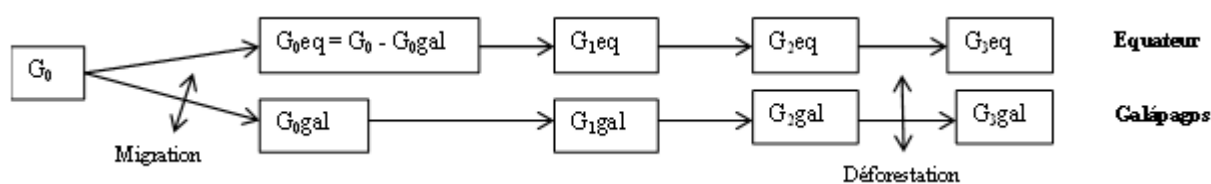
Le principe du **TP de SVT sur les tortues** ([annexe 10](#)) est de modéliser, à partir d'un tableur Excel ([annexe 11](#)), l'évolution sur trois générations d'une petite population de tortues dont l'environnement varie.

Le scénario de départ est volontairement simple :

On travaille sur un caractère observable, la couleur de la carapace (phénotype) contrôlé par un gène existant sous deux allèles vert (v) et marron (m) avec " m dominant ".

On choisit de faire migrer quelques individus de la population initiale de façon à créer deux populations de tortues de taille différente dont on suivra l'évolution sur 3 générations.

L'environnement (et donc le taux de survie) est considéré comme stable jusqu'à la troisième génération en Equateur. Par contre aux îles Galápagos, entre la 2<sup>ème</sup> et la 3<sup>ème</sup> génération, on introduit une modification environnementale (déforestation) de façon à modéliser la sélection naturelle. En effet, les tortues de phénotype [vert], c'est-à-dire de génotype VV, moins bien cachées en l'absence de végétation, sont plus soumises à la prédation.

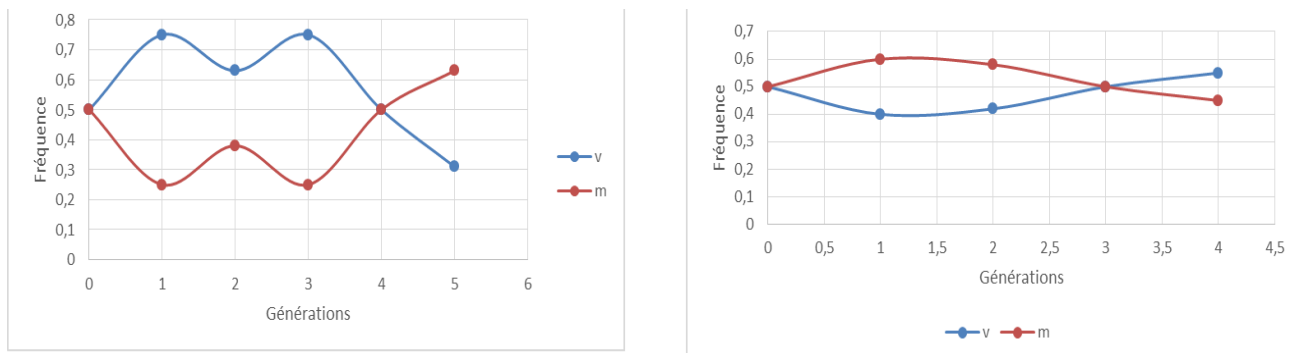


L'évolution des tortues est matérialisée en remplissant le tableur Excel préconçu et en réalisant des tirages au sort pour former les couples et les descendants (4 descendants par couple). Cette phase de modélisation guidée est importante pour mettre en évidence les outils, les mécanismes et les éléments nécessaires à la programmation :

- Pour appréhender l'échelle d'étude des mécanismes de l'évolution, il faut travailler sur une population et sur plusieurs générations (échelles spatiale et temporelle qui sont souvent des obstacles pour les élèves qui ont le réflexe de se placer à l'échelle de l'individu).
- Pour appréhender les mécanismes aléatoires mis en jeu, la fonction " tirage aléatoire " est un outil intéressant. Par la suite, les élèves font le lien entre tirage au sort des allèles des descendants et brassage génétique.
- Pour mesurer l'évolution, il faut un paramètre mesurable : la notion de fréquence allélique.

Des exemples de résultats :

*Evolution de la fréquence des allèles m et v d'une population de tortues au cours des générations aux Galápagos (gauche) et en Equateur (droite)*



En fin de TP, chaque groupe met ses résultats à disposition de la classe pour faciliter la comparaison et donc les conclusions. La comparaison des résultats de la population équateur pour plusieurs groupes permet de montrer l'évolution aléatoire du paramètre fréquence en lien avec les tirages au sort, c'est la dérive génétique. En comparant les populations de tailles différentes (équateur et Galápagos) sur les deux premières générations, on peut observer que la rapidité de l'évolution est fonction de la taille de la population. Enfin en comparant la population Galápagos à partir de la 3<sup>ème</sup> génération entre plusieurs groupes, on voit que le changement d'environnement induit une évolution orientée (diminution pour tous les groupes de la fréquence de l'allèle vert), c'est la sélection naturelle.

Dans un deuxième temps, une réflexion sur la qualité de la modélisation en elle-même va être riche en apprentissage. De manière générale, trois générations sont insuffisantes mais le TP a déjà duré deux heures. Certains groupes ne modéliseront pas la sélection naturelle car de façon aléatoire, le phénotype [vert], c'est-à-dire le génotype VV n'est pas présent.

Par ailleurs, les élèves proposent alors d'eux même l'idée d'automatiser le remplissage du fichier pour gagner du temps et accéder à une modélisation plus complète avec plus d'individus et plus de générations. La transition vers la conception d'un algorithme est alors logique pour eux.

### **b. Conception de l'algorithme**

Le TP des tortues constitue pour les élèves un cas concret illustrant à la fois les mécanismes permettant de passer à la génération suivante mais aussi les facteurs liés à l'environnement.

La séquence 4 ([annexe 12](#)) pose les questions propres au modèle et fixe les paramètres et objets qui constitueront les éléments clés de notre simulation.

Les élèves sont guidés mais la conception algorithmique de chaque étape leur est laissée entièrement.

Sur le plan pratique, et pour s'adapter à l'effectif des classes et à l'équipement en ordinateurs des salles de classe, on constitue des groupes de 2 ou 3 élèves. Comme en TPE, il leur est demandé de tenir à jour un carnet de bord numérique et de le présenter au professeur à chaque séance.

La progression est très variable d'un groupe à l'autre mais c'est ici qu'il est important de veiller à ce que chaque groupe puisse aller à son rythme. Il faut donc espacer les séances d'AP afin que chacun puisse poursuivre le travail en dehors de la classe.

Il faut encourager les groupes qui rencontrent des obstacles en leur apportant l'aide nécessaire mais sans donner la solution. En particulier, on peut leur montrer que chaque étape de l'algorithme peut être testée en "la faisant tourner", ce qui procure une certaine satisfaction, permet de repérer les erreurs et motive pour continuer.

Avec les groupes plus à l'aise, on peut exiger plus de rigueur dans la conception de l'algorithme et imposer des versions différentes intégrant tel ou tel paramètre. On peut aussi leur montrer comment intégrer des fonctions de façon à rendre les programmes plus lisibles et plus simples à concevoir.

Bien évidemment, il est difficile d'empêcher les groupes de communiquer entre eux mais globalement les élèves ont joué le jeu et ont produit des programmes très personnels.

La première question posée dans la séquence 4 est relative à la modélisation d'une population d'individus caractérisés par leur génotype.

En biologie, le génotype d'un individu (pour un gène donné) est caractérisé par une paire d'allèles.



Mais pour des raisons pratiques, dans notre simulation nous modélisons ce génotype par un couple d'allèles. En mathématiques, la notion de couple impose un ordre. Il faut donc faire comprendre aux élèves que ce choix n'aura aucune incidence sur la fiabilité des résultats.

De même, l'ensemble des individus d'une population est modélisé par une liste d'individus.

Ensuite, cette première question sous-entend d'autres questions, telles que : " faut-il considérer une population sexuée ? ", " comment fixer les fréquences alléliques de la première génération ? ", " combien d'individus ? ".

Tout ceci provoque des débats très intéressants au sein et entre les groupes et montre aux élèves l'importance de cette phase dans la simulation.

C'est aussi l'occasion de leur expliquer qu'il faut commencer par simplifier le modèle et l'affiner dans un second temps si cela est nécessaire.

Globalement les élèves ont fixé leur population initiale à l'identique de celle du TP des tortues, c'est-à-dire 4 individus mâles et 4 individus femelles (Programme élève : [annexe 13](#)) et ont su facilement en déterminer les fréquences alléliques.

La création de la génération suivante a posé plus de problème mais grâce au travail effectué dans les séquences 2 et 3 et à la simulation sur tableur, les élèves ont pu faire les liens et parvenir tant bien que mal au résultat. Là encore la plupart a fait le choix de 4 descendants par couple, comme dans le TP, ce qui entraîne une croissance exponentielle de la population, mais à ce stade ils n'ont pas pris conscience des conséquences en termes de temps d'exécution.

Il restait à boucler le processus, chose simple en écriture mais qui a soulevé beaucoup de questions. Les élèves ont bien pensé à la boucle " pour " mais ont mis beaucoup de temps à trouver la solution, à savoir faire jouer le rôle de population initiale à la population suivante.

Finalement, au bout de 4 séances quelques groupes sont parvenus à faire tourner leur programme et obtenir les premiers graphiques avec une joie et une satisfaction que vous pouvez imaginer (carnet de bord : [annexe 14](#)). Malheureusement, cela a créé une démotivation chez certains groupes et il a fallu suggérer la possibilité d'une collaboration entre groupes afin que chacun puisse avancer.

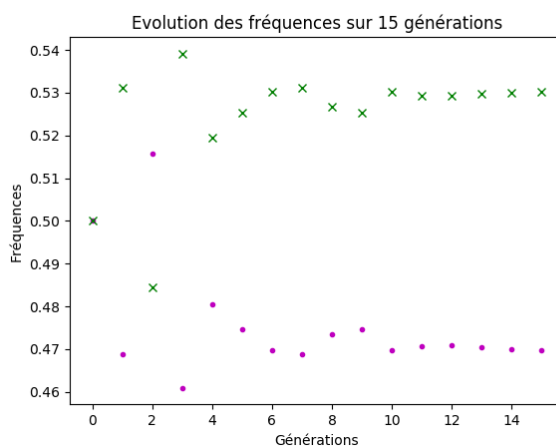
Ceci a permis de passer rapidement à l'exploitation et à l'analyse des résultats.

### c. Résultats et analyse

Dans un premier temps, on suggère aux élèves de tester leur programme en comparant le graphique obtenu avec celui du tableur réalisé lors du TP des tortues. Sur les 5 premières générations le résultat est identique, à savoir une large fluctuation des fréquences alléliques.

Mais sur plus de générations (*Fig. 5*), et c'est ici que les élèves prennent conscience de l'efficacité de leur programme, on peut noter une stabilisation des fréquences assez rapidement.

On peut noter aussi que ceux qui ont voulu voir ce qui se passe au-delà de la 15<sup>ème</sup> génération ont été surpris par la lenteur de l'exécution et cela a été l'occasion de parler du temps d'exécution du programme par rapport à la croissance exponentielle de l'effectif de la population.



*Fig. 5 : Modèle exponentiel – 4 couples d'individus en génération initiale*

Une première conjecture s'impose. Il semblerait qu'au-delà de la 10<sup>ème</sup> génération ( $8 \times 2^{10} \approx 8000$  individus) les fréquences alléliques se stabilisent. Ceci serait-il directement lié au fait qu'à ce stade, la taille de la population est suffisante pour que le modèle de H. W. soit applicable ?

Pour le vérifier, il faut passer à un modèle de croissance nulle, c'est-à-dire à effectif constant.

La réalisation de ce modèle est très simple mais le choix des paramètres de taille de population et de durée d'observation est plus difficile à maîtriser.

En effet, un effectif trop petit entraîne une grande fluctuation des fréquences et à plus ou moins court terme une fixation de ces fréquences à 0 ou 1 (Fig. 6 et Fig. 7).

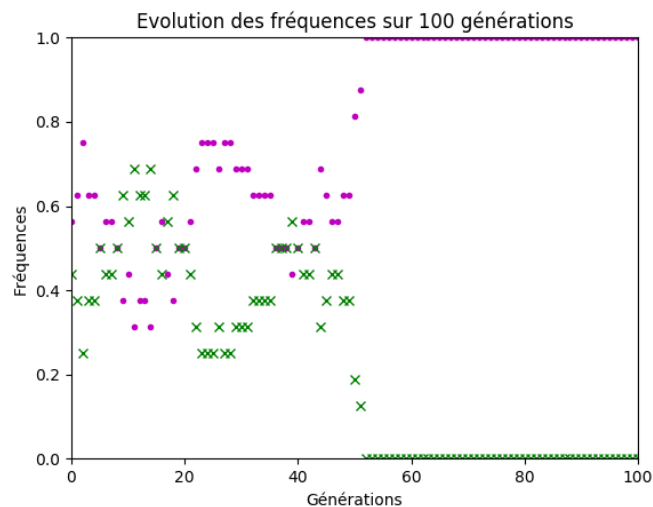
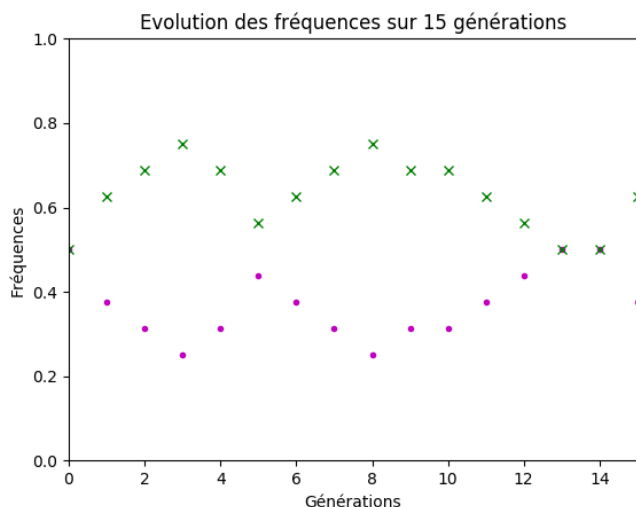


Fig. 6, Fig. 7 : Modèle constant - 4 couples d'individus en génération initiale

Même avec un effectif conséquent d'environ 1000 individus, les fréquences globalement stables sur une centaine de générations finissent par diverger et se fixer également sur un temps d'observation suffisamment long (Fig. 8 et Fig. 9).

On peut penser, que quelle que soit la taille de la population, cette divergence et l'extinction d'un des allèles finira par arriver à condition d'attendre suffisamment longtemps. Mais cela reste à démontrer !

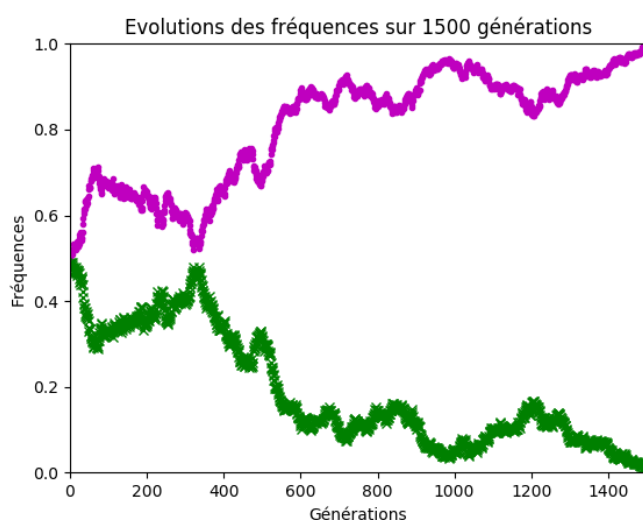
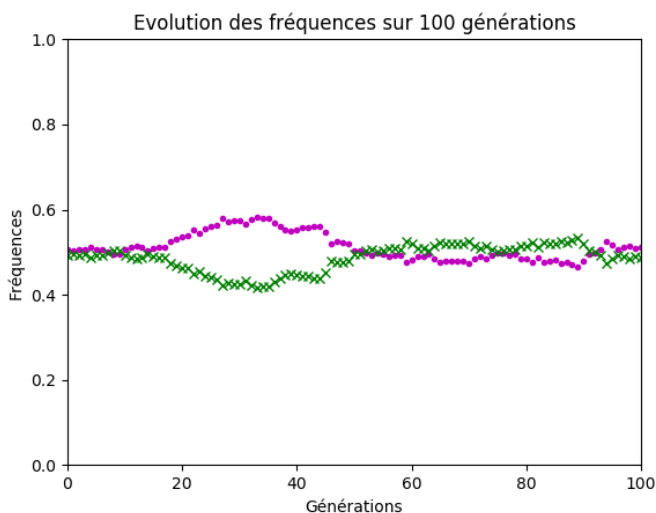


Fig. 8, Fig. 9 : Modèle constant - 500 couples d'individus en génération initiale

Si on passe à 8 000 individus et un temps d'observation raisonnable de 200 générations (Fig. 10 et Fig. 11), la conjecture émise semble se vérifier, y compris lorsque les fréquences alléliques initiales sont distinctes.

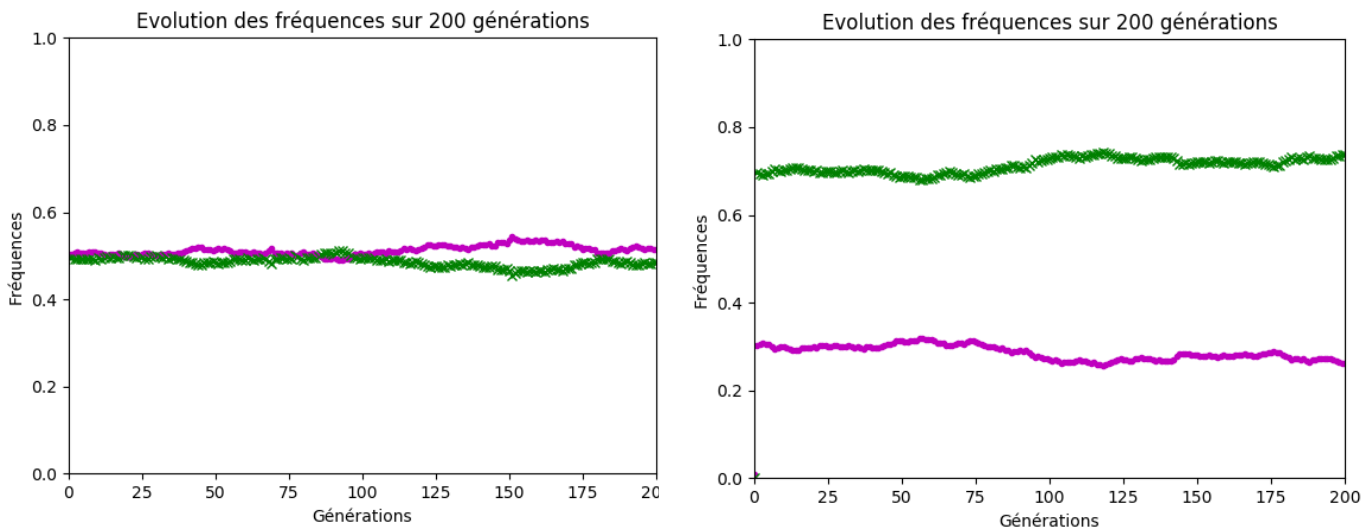


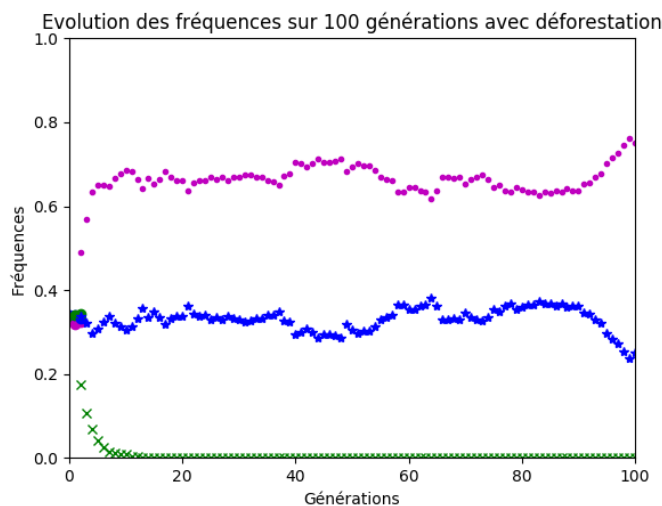
Fig. 10, Fig. 11 : Modèle constant - 4000 couples d'individus en génération initiale

Dans ce dernier cas et pour fixer les fréquences initiales à des valeurs distinctes, on peut intégrer au programme une fonction qui, à une fréquence  $p$  et une taille  $n$  données, renvoie une population de taille  $n$  où la fréquence de l'allèle choisi est environ  $p$ . Sauf si la population est créée "à la main" en respectant la fréquence choisie, cela nécessite la création d'une variable aléatoire  $r$  suivant la loi uniforme sur  $[0 ; 1]$  et une instruction conditionnelle renvoyant l'allèle choisi si  $r \leq p$  (ou l'autre allèle sinon). Là encore, cette technique n'est pas maîtrisée par les élèves et il faudra peut-être leur souffler la solution.

Finalement on peut accepter le modèle de H.W. sur des effectifs de population suffisants mais aussi sur un temps d'observation adapté. Dans les autres cas on parlera de dérive génétique.

Malheureusement, ce travail d'analyse n'a été fait que par très peu de groupes, faute de temps. Mais les groupes qui y sont parvenus, même partiellement, ont pu exposer aux autres leurs résultats et leur analyse.

Pour ce qui est de la sélection naturelle, on peut intégrer au programme la déforestation du TP des tortues où les individus dont le génotype contient l'allèle discriminant meurent. On rajoute ici un troisième allèle et on observe que la fréquence de l'allèle en question se fixe alors très vite à 0 au cours des générations (Fig. 12).





*Fig. 12 : Modèle constant 3 allèles - 500 couples d'individus en génération initiale et avec déforestation*

Plus généralement, on peut envisager d'affiner encore le modèle. On pourrait considérer un taux de reproduction variable suivant les individus, dépendant par exemple du génotype. Ou encore, que la rencontre ne soit plus le fruit du hasard mais dépende aussi du génotype des individus. Ce travail pourrait se faire dans le cadre d'un atelier de mathématiques ouvert aux intéressés.

***Fin de l'article***

## IV. Annexes

### Annexe 1 : version 1 du devoir

#### La loi de Hardy-Weinberg

Dans les cas simples, un gène peut prendre deux formes (ou allèles) A et a. Chaque gène se trouvant en deux exemplaires, un individu peut donc présenter l'un des trois génotypes suivants : AA, Aa ou aa.

On considère une population dont les proportions respectives de ces génotypes sont  $p_0$ ,  $q_0$  et  $r_0$  (avec  $p_0 + q_0 + r_0 = 1$ ).

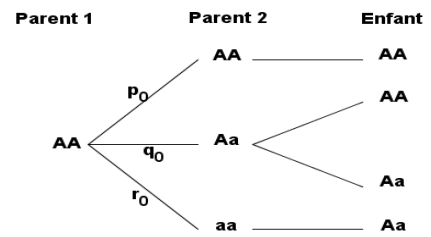
Un enfant hérite d'un allèle de chacun de ses parents, chaque allèle étant choisi au hasard. Ainsi, si le père est de type AA et la mère de type Aa, alors les enfants peuvent être du type AA ou Aa.

On admettra que, dans ce modèle théorique les couples se forment au hasard quant à leurs génotypes, que l'environnement est stable et que la population est suffisamment grande pour être considérée comme infinie.

On étudie l'évolution des proportions des différents génotypes dans la population à chaque génération.

#### 1. Première génération

- On suppose que le génotype d'un des parents est AA. Compléter l'arbre pondéré ci-contre, puis calculer la probabilité que le génotype de l'enfant soit AA puis qu'il soit Aa.
- On suppose maintenant que le génotype d'un des parents est Aa. A l'aide d'un nouvel arbre pondéré, calculer la probabilité que le génotype de l'enfant soit AA ; Aa ; aa.
- On note  $p_1$  la probabilité qu'un enfant de la première génération ait le génotype AA. Démontrer que  $p_1 = \left(p_0 + \frac{q_0}{2}\right)^2$ .
- Démontrer que la probabilité  $r_1$  qu'un enfant de la première génération ait le génotype aa est :  $r_1 = \left(r_0 + \frac{q_0}{2}\right)^2$ .



#### 2. Etablir une conjecture

On donne les proportions  $p_0 = 0,6$  et  $r_0 = 0,2$ . Pour cette partie, on pourra utiliser un tableur comme présenté ci-contre ; dans ce cas, on joindra à la copie la feuille de calculs imprimée, ainsi que les formules saisies dans les cellules. A défaut de tableur, on écrira les calculs effectués sur la copie.

	A	B	C	D
1	Génération	P(AA) = p	P(Aa) = q	P(aa) = r
2	0	0,6		0,2
3	1			
4	2			
5	3			

- Calculer  $q_0$ , puis  $p_1$ ,  $q_1$  et  $r_1$ .
- Calculer de même les proportions des différents génotypes aux générations suivantes. On notera  $p_n$ ,  $q_n$  et  $r_n$  les proportions respectives des différents génotypes à la n-ième génération,  $n \geq 1$ .
- Que remarque-t-on ?
- Modifier les proportions initiales  $p_0$  et  $r_0$ . Proposer une conjecture à partir de ces observations.

#### 3. Cas général

On pose  $\alpha = p_0 - r_0$ .

- Démontrer que  $p_1 = \frac{(1+\alpha)^2}{4}$ . Donner également l'expression de  $r_1$  puis de  $q_1$  en fonction de  $\alpha$ .
- Montrer que  $p_1 - r_1 = \alpha$ .
- Etablir les expressions de  $p_2$ ,  $q_2$  et  $r_2$  en fonction de  $\alpha$ .
- Que peut-on conclure pour les proportions  $p_n$ ,  $q_n$  et  $r_n$  des différents génotypes à la n-ième génération ( $n \geq 1$ ) ?  
Faire une application numérique avec  $p_0 = 0,7$ ,  $q_0 = 0,2$  et  $r_0 = 0,1$ .

#### 4. Pour aller plus loin

Dans la pratique, pour suivre l'évolution d'une population, on s'intéresse aux fréquences alléliques au cours des générations.

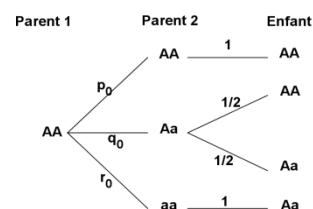
On conserve les hypothèses du modèle et on note  $A_n$  et  $a_n$  les fréquences respectives des allèles A et a dans la population au cours de la n-ième génération.

- Montrer que pour tout entier naturel n,  $A_n = p_n + \frac{q_n}{2}$  et  $a_n = r_n + \frac{q_n}{2}$ .
- Conclure sur l'évolution des fréquences alléliques au cours des générations.
- Il se trouve que les observations effectuées sur certaines populations ne correspondent pas toujours aux résultats de ce modèle. Emettre des hypothèses qui pourraient expliquer cet écart.

## Annexe 2 : corrigé de la version 1

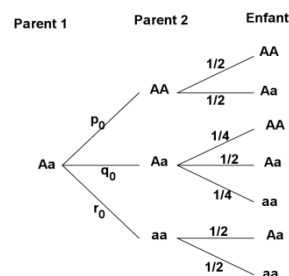
### 1. Première génération

a) D'après l'arbre obtenu ci-contre, lorsqu'un des parents est de génotype AA, la probabilité que l'enfant soit AA est obtenue en appliquant la formule des probabilités totales :  $P_{\text{parent1 AA}}(\text{enfant AA}) = p_0 + \frac{1}{2}q_0$ . La probabilité qu'il soit Aa est :  $P_{\text{parent1 AA}}(\text{enfant Aa}) = \frac{1}{2}q_0 + r_0$ .



b) On utilise également un arbre pondéré et on obtient, de la même façon, que, lorsqu'un des parents est de génotype Aa, la probabilité que l'enfant soit AA est  $P_{\text{parent1 Aa}}(\text{enfant AA}) = \frac{1}{2}p_0 + \frac{1}{4}q_0$ .

Celle qu'il soit Aa est  $P_{\text{parent1 Aa}}(\text{enfant Aa}) = \frac{1}{2}p_0 + \frac{1}{2}q_0 + \frac{1}{2}r_0$ . Enfin, celle qu'il soit de génotype aa est  $P_{\text{parent1 Aa}}(\text{enfant aa}) = \frac{1}{4}q_0 + \frac{1}{2}r_0$ .



c) On note  $p_1$  la probabilité qu'un enfant de la première génération soit de génotype AA. D'après la formule des probabilités totales, on a :

$p_1 = P_{\text{parent1 AA}}(\text{enfant AA}) \times P(\text{parent1 AA}) + P_{\text{parent1 Aa}}(\text{enfant AA}) \times P(\text{parent1 Aa})$ . En utilisant les résultats précédents, on obtient :

$$p_1 = (p_0 + \frac{1}{2}q_0)p_0 + (\frac{1}{2}p_0 + \frac{1}{4}q_0)q_0 = p_0^2 + p_0q_0 + \frac{1}{4}q_0^2.$$

$$\text{On a bien alors } p_1 = \left(p_0 + \frac{q_0}{2}\right)^2.$$

d) Un raisonnement similaire (basé sur le rôle symétrique des génotypes AA et aa) conduit à

$$r_1 = \left(r_0 + \frac{q_0}{2}\right)^2.$$

### 2. Conjecture

Pour la conjecture sur tableur, on saisit les formules suivantes :

- En C2 :  $=1-B2-D2$ .
- En B3 :  $=(B2+C2/2)^2$ .
- En D3 :  $=(D2+C2/2)^2$ .

	A	B	C	D
1	Génération	P(AA) = p	P(Aa) = q	P(aa) = r
2	0	0,6	0,2	0,2
3	1	0,49	0,42	0,09
4	2	0,49	0,42	0,09
5	3	0,49	0,42	0,09

Ensuite, on colle toutes ces formules vers le bas.

L'observation des résultats laisse penser que dès la première génération, la répartition des différents génotypes AA, Aa et aa dans la population ne change plus.

### 3. Cas général

$\alpha = p_0 - r_0$  ; on sait de plus que  $p_0 + q_0 + r_0 = 1$ , ce qui équivaut à  $p_0 + q_0 = 1 - r_0$ .

a)  $p_1 = \left(p_0 + \frac{q_0}{2}\right)^2 = \frac{(2p_0+q_0)^2}{4} = \frac{(p_0+p_0+q_0)^2}{4} = \frac{(p_0+1-r_0)^2}{4}$ , soit  $p_1 = \frac{(1+\alpha)^2}{4}$ . On obtient de façon similaire :  $r_1 = \frac{(1-\alpha)^2}{4}$  et  $q_1 = 1 - p_1 - r_1 = \frac{1-\alpha^2}{2}$ .

b) De là,  $p_1 - r_1 = \frac{(1+\alpha)^2}{4} - \frac{(1-\alpha)^2}{4} = \frac{1+2\alpha+\alpha^2-(1-2\alpha+\alpha^2)}{4} = \frac{4\alpha}{4}$ . On a bien :  $p_1 - r_1 = \alpha$ .

c) Puisque  $p_1 - r_1 = \alpha$ , le raisonnement tenu dans la question a) nous amène à :  $p_2 = \frac{(1+\alpha)^2}{4}$  ;  $q_2 = \frac{1-\alpha^2}{2}$  et  $r_2 = \frac{(1-\alpha)^2}{4}$ . On a donc  $p_2 = p_1$  ;  $q_2 = q_1$  et  $r_2 = r_1$ . De là, on a à nouveau  $p_2 - r_2 = \alpha$ .

d) On peut réitérer le raisonnement pour toutes les générations suivantes.

Avec  $p_0 = 0,7$ ,  $q_0 = 0,2$  et  $r_0 = 0,1$ , on obtient, pour la première génération :  $p_1 = 0,64$ ,  $q_1 = 0,32$  et  $r_1 = 0,04$  et donc pour tout entier  $n \geq 1$ , on aura toujours  $p_n = 0,64$ ,  $q_n = 0,32$  et  $r_n = 0,04$ .

### 4. Pour aller plus loin

Dans la pratique, pour suivre l'évolution d'une population, on s'intéresse aux fréquences alléliques au cours des générations.

On conserve les hypothèses du modèle et on note  $A_n$  et  $a_n$  les fréquences respectives des allèles A et a dans la population au cours de la n-ième génération.

a) On choisit au hasard un allèle parmi ceux de la population à la génération n et on note : A l'évènement « l'allèle est du type A », AA l'évènement « l'allèle provient d'un individu de génotype AA » et Aa l'évènement : « l'allèle provient d'un individu de génotype Aa ».

$$\text{On a : } A_n = P(A) = P_{AA}(A) \times P(AA) + P_{Aa}(A) \times P(Aa) = 1 \times p_n + \frac{1}{2} \times q_n, \text{ soit } A_n = p_n + \frac{1}{2}q_n.$$

$$\text{De même, on obtient : } a_n = r_n + \frac{1}{2}q_n.$$

b) Etant donné que les suites p, q, r sont constantes à partir du rang 1, on conclut d'après les relations ci-dessus que les fréquences alléliques sont constantes à partir de la génération 1 (on peut même vérifier qu'elles le sont à partir de la génération 0).

c) Le modèle d'Hardy-Weinberg n'a pas de réalité biologique car le modèle raisonne avec une population infinie d'individus qui peuvent tous se reproduire et où l'ensemble de leurs gamètes est utilisé dans la reproduction. Or les populations ne sont pas infinies, elles sont plus ou moins isolées géographiquement les unes des autres, le choix du conjoint est donc limité.

Par ailleurs, seule une infime minorité des gamètes produits est utilisée dans la reproduction.



**La loi de Hardy-Weinberg**

Dans les cas simples, un gène peut prendre deux formes (ou allèles) A et a. Chaque gène se trouvant en deux exemplaires, un individu peut donc présenter l'un des trois génotypes suivants : A//A, A//a ou a//a.

Un enfant hérite d'un allèle de chacun de ses parents, chaque allèle étant choisi au hasard. Ainsi, si l'un des parents est de type A//A et l'autre de type A//a, alors l'enfant peut être de type A//A ou A//a.

On admettra que, dans ce modèle théorique, les couples ont la même capacité de reproduction et qu'ils se forment au hasard indépendamment de leurs génotypes, que l'environnement est stable et que la population est suffisamment grande pour être considérée comme infinie.

Pour tout entier naturel n, on note  $G_n$  la population à la génération n et,

- $p_n, q_n$  et  $r_n$  les proportions respectives des génotypes A//A, A//a et a//a dans  $G_n$  avec  $p_n + q_n + r_n = 1$ .
- $A_n$  et  $a_n$  les proportions respectives des allèles A et a dans  $G_n$  avec  $A_n + a_n = 1$ .

On se propose d'étudier l'évolution de ces différentes proportions au cours des générations.

**1. Lien entre les proportions dans  $G_n$**

Soit n un entier naturel. On choisit au hasard un allèle parmi ceux de la population  $G_n$ .

On note :

A l'évènement : « l'allèle est du type A » ; a l'évènement : « l'allèle est du type a ».

AA l'évènement : « l'allèle provient d'un individu de génotype A//A » ; Aa l'évènement : « l'allèle provient d'un individu de génotype A//a » ; aa l'évènement : « l'allèle provient d'un individu de génotype a//a ».

On a alors pour tout entier naturel n,  $P(A) = A_n, P(a) = a_n, P(AA) = p_n, P(Aa) = q_n$  et  $P(aa) = r_n$ .

A l'aide de la formule des probabilités totales, montrer que pour tout entier naturel n,  $A_n = p_n + \frac{q_n}{2}$  et  $a_n = r_n + \frac{q_n}{2}$ .

On pourra faire un arbre faisant intervenir les évènements A, a, AA, Aa et aa.

**2. D'une génération à la suivante**

Soit n un entier naturel. On considère que la population  $G_{n+1}$  est constituée uniquement des enfants issus des couples formés dans  $G_n$ .

On choisit au hasard un enfant dans la population  $G_{n+1}$ .

- a. Expliquer pourquoi on peut considérer le génotype de l'enfant comme le résultat de la rencontre au hasard de deux allèles issus de  $G_n$ .
- b. En déduire que pour tout entier naturel n,  $p_{n+1} = A_n^2, r_{n+1} = a_n^2$  puis que  $q_{n+1} = 2a_n A_n$ .

**3. Etude des suites**

- a. Dans cette question on pose :  $p_0 = 0,7 ; q_0 = 0,2$  et donc  $r_0 = 0,1$ .

A l'aide d'un tableur, créer une feuille de calcul comme présentée ci-contre et calculer les 6 premiers termes des suites p, q, r, A et a.

On joindra à la copie la feuille de calculs imprimée, ainsi que les formules saisies dans les cellules. A défaut de tableur, on écrira les calculs effectués sur la copie.

- b. Reprendre les calculs précédents en choisissant d'autres valeurs initiales de  $p_0$  et  $q_0$ . Emettre une conjecture sur le comportement de chacune de ces suites.
- c. Montrer que pour tout entier naturel n,  $A_{n+1} = A_n$  et  $a_{n+1} = a_n$ .
- d. En déduire le comportement des suites A et a puis celui des suites p, q et r.
- e. Enoncer une loi génétique caractérisant ce modèle de population.

	A	B	C	D	E	F
1	Génération	$p_n$	$q_n$	$r_n$	$A_n$	$a_n$
2	0	0,7	0,2	0,1		
3	1					
4	2					
5	3					
6	4					
7	5					

**4. Confrontation du modèle à la réalité**

- a. Il se trouve que les observations effectuées sur certaines populations ne correspondent pas toujours aux résultats de ce modèle. Emettre des hypothèses qui pourraient expliquer cet écart.
- b. Quelle démarche nous permettrait d'améliorer ce modèle ?

## Annexe 4 : corrigé de la version 2 du devoir

### **La loi de Hardy-Weinberg**

On admettra que, dans ce modèle théorique les couples se forment au hasard et indépendamment de leurs génotypes, que l'environnement est stable et que la population est suffisamment grande pour être considérée comme infinie.

Pour tout entier naturel  $n$ , on note  $G_n$  la population à la génération  $n$  et,

- $p_n$ ,  $q_n$  et  $r_n$  les proportions respectives des génotypes  $A//A$ ,  $A//a$  et  $a//a$  dans  $G_n$  avec  $p_n + q_n + r_n = 1$ .
- $A_n$  et  $a_n$  les proportions respectives des allèles  $A$  et  $a$  dans  $G_n$  avec  $A_n + a_n = 1$ .

Etude de l'évolution des différentes proportions au cours des générations.

#### 1. Lien entre les proportions dans $G_n$

Soit  $n$  un entier naturel. On choisit au hasard un allèle parmi ceux de la population  $G_n$ .

On note :

$A$  l'évènement : « l'allèle est du type  $A$  » ;  $a$  l'évènement : « l'allèle est du type  $a$  ».

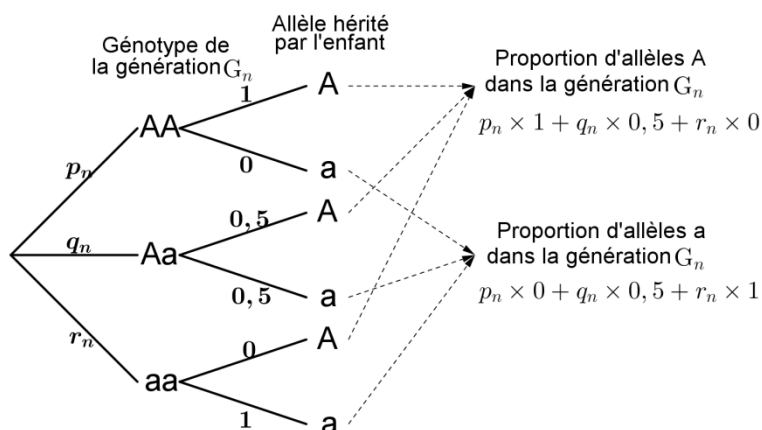
$AA$  l'évènement : « l'allèle provient d'un individu de génotype  $A//A$  » ;  $Aa$  l'évènement : « l'allèle provient d'un individu de génotype  $A//a$  » ;  $aa$  l'évènement : « l'allèle provient d'un individu de génotype  $a//a$  ».

On a alors pour tout entier naturel  $n$ ,  $P(A) = A_n$ ,  $P(a) = a_n$ ,  $P(AA) = p_n$ ,  $P(Aa) = q_n$  et  $P(aa) = r_n$ .

Soit  $n \in \mathbb{N}$ .

Les allèles sont transmis à l'enfant au hasard donc :  $P_{AA}(A) = P_{aa}(a) = 1$  ;  $P_{Aa}(A) = P_{Aa}(a) = \frac{1}{2}$  ;  $P_{aa}(A) = P_{AA}(a) = 0$ .

D'où l'arbre ci-dessous :



Dans la partition  $\{AA ; Aa ; aa\}$  on applique alors la formule des probabilités totales :

$$P(A) = P(A \cap AA) + P(A \cap Aa) + P(A \cap aa) = P_{AA}(A) \times P(AA) + P_{Aa}(A) \times P(Aa) + P_{aa}(A) \times P(aa) = 1 \times p_n + \frac{1}{2} q_n + 0 \times r_n,$$

$$\text{d'où } A_n = p_n + \frac{q_n}{2}.$$

$$\text{et, } P(a) = P(a \cap AA) + P(a \cap Aa) + P(a \cap aa) = P_{AA}(a) \times P(AA) + P_{Aa}(a) \times P(Aa) + P_{aa}(a) \times P(aa) = 0 \times p_n + \frac{1}{2} q_n + 1 \times r_n,$$

$$\text{d'où } a_n = r_n + \frac{q_n}{2}.$$

## 2. D'une génération à la suivante

Soit  $n$  un entier naturel. On considère que la population  $G_{n+1}$  est constituée uniquement des enfants issus des couples formés dans  $G_n$ .

On choisit au hasard un enfant dans la population  $G_{n+1}$ .

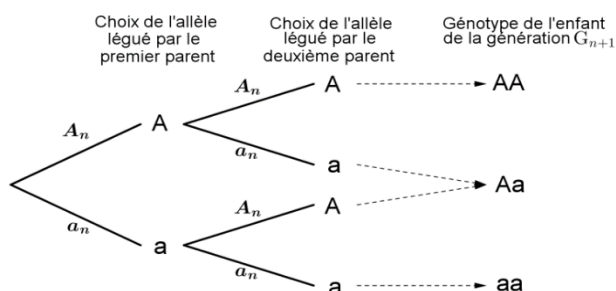
a. D'une part, les couples se forment au hasard et indépendamment de leurs génotypes. D'autre part les allèles sont transmis à l'enfant au hasard. On peut donc considérer que le génotype de l'enfant est uniquement déterminé par la rencontre au hasard de deux allèles issus de  $G_n$ .

b. La population est suffisamment grande pour être considérée comme infinie, alors d'après **a.**, on déduit que pour tout entier naturel  $n$ , l'évènement AA (génotype de l'enfant dans  $G_{n+1}$ ) correspond au résultat de deux tirages successifs avec remise d'un allèle de la population  $G_n$ .

Ainsi,  $p_{n+1} = A_n \times A_n = A_n^2$ , de même,  $r_{n+1} = a_n \times a_n = a_n^2$ .

et,  $q_{n+1} = 1 - p_{n+1} - r_{n+1} = (A_n + a_n) - A_n^2 - a_n^2 = A_n - A_n^2 + a_n - a_n^2 = A_n(1 - A_n) + a_n(1 - a_n) = A_n a_n + a_n A_n = 2a_n A_n$ .

Soit  $p_{n+1} = A_n^2$ ,  $r_{n+1} = a_n^2$ ,  $q_{n+1} = 2a_n A_n$ . Ces résultats peuvent aussi s'établir à l'aide de l'arbre suivant :



## 3. Etude des suites

a. On prend :  $p_0 = 0,7$  ;  $q_0 = 0,2$  et donc  $r_0 = 0,1$ .

Après avoir saisi les formules figurant dans le tableau ci-contre,

on obtient par recopie les 6 premiers termes des suites  $p$ ,  $q$ ,  $r$ ,  $A$  et  $a$ .

	A	B	C	D	E	F
1 Génération	pn	qn		rn	An	an
2 0	0,7	0,2		0,1	=B2+C2/2	=D2+C2/2
3 1	=E2^2	=2*E2*F2		=F2^2	=B3+C3/2	=D3+C3/2

b. Après avoir refait les calculs avec d'autres valeurs de  $p_0$  et  $q_0$ , on constate que les fréquences alléliques sont stables à partir de  $G_0$  et que les fréquences génotypiques sont stables à partir de  $G_1$ .

c. Soit  $n \in \mathbb{N}$ . D'après **1.**,  $A_{n+1} = p_{n+1} + \frac{1}{2}q_{n+1}$ , ce qui donne d'après

**2.**,  
 $A_{n+1} = A_n^2 + a_n A_n = A_n(A_n + a_n) = A_n \times 1$ , d'où,  $A_{n+1} = A_n$ .  
 de même on obtient,  $a_{n+1} = a_n$ .

d. On déduit de la question précédente que les suites  $A$  et  $a$  sont constantes à partir du rang 0 et d'après **2. b.**, les suites  $p$ ,  $q$ ,  $r$  le sont à partir du rang 1.

e. Dans une population vérifiant les hypothèses décrites en début d'énoncé, les fréquences alléliques et les fréquences génotypiques restent stables de génération en génération.

	A	B	C	D	E	F
1 Génération	pn	qn		rn	An	an
2 0	0,7	0,2		0,1	0,8	0,2
3 1	0,64	0,32		0,04	0,8	0,2
4 2	0,64	0,32		0,04	0,8	0,2
5 3	0,64	0,32		0,04	0,8	0,2
6 4	0,64	0,32		0,04	0,8	0,2
7 5	0,64	0,32		0,04	0,8	0,2

## 4. Confrontation du modèle à la réalité

a. Le modèle d'Hardy-Weinberg n'a pas de réalité biologique car il impose une population infinie d'individus qui peuvent tous se reproduire et où l'ensemble de leurs gamètes est utilisé dans la reproduction. Or les populations ne sont pas infinies, elles sont plus ou moins isolées géographiquement les unes des autres, le choix du conjoint est donc limité.

Par ailleurs, seule une infime minorité des gamètes produits est utilisée dans la reproduction.

Enfin le modèle exclut tout facteur externe à la population, et ne prend pas en compte les interactions avec son environnement.

C'est ce qui explique que les observations effectuées sur certaines populations ne correspondent pas toujours aux résultats de ce modèle.

b. La simulation va nous permettre d'améliorer le modèle théorique car nous pouvons intégrer des paramètres permettant de coller au plus près à la réalité.

## Annexe 5 : séquence 1

### Prise en main du langage Python

#### **Etape 1**

Connectez-vous sur le site d'EduPython à l'adresse : <https://edupython.tuxfamily.org/>. Cet environnement est pratique car il propose une distribution portable sous Python 3.

Installez la version la plus récente sur une clé USB, clé que vous conserverez toute l'année (et plus si affinité !) car c'est sur cet espace que vous développerez vos programmes. Si vous rencontrez des difficultés, voici un tutoriel vidéo préparé par Mr Berhouet : <https://www.youtube.com/watch?v=MyAYmcSmhT4&feature=youtu.be>

Récupérez le document PDF réalisé par Agnès Baraquin ainsi que les fiches élèves qui seront précieux pour une bonne prise en main du langage.

#### **Etape 2**

Je vous propose d'étudier deux programmes écrits et commentés en langage Python. Dirigez-vous vers votre espace de travail sur le réseau du lycée ; copiez les deux programmes du dossier " *Zone des énoncés* " et collez-les dans un dossier personnel de votre clé USB.

#### **Etape 3**

Nous étudierons ensemble ces deux programmes et vous pourrez les faire tourner pour vous familiariser avec l'éditeur PyScripteur. Vous effectuerez ensuite les modifications demandées.

#### **Etape 4**

Vous me rendrez ce travail en déposant vos fichiers (avec votre nom dans le titre) dans la " *zone de dépose des travaux* " du réseau du lycée.

*Bon courage*



## Annexe 6 : séquence 2

### **Exemple 1 - Tirages avec remise dans une urne**

Voici un programme Python commenté, simulant n tirages avec remise dans une urne.

On représente graphiquement l'évolution de la fréquence d'apparition de deux évènements au fur et à mesure des tirages.

#### Programme Python

#### Commentaires

<pre>from lycee import * from random import *  Urne=[0,1,0,0,0,1]  Tirages=[]  n=int(input("nombre de tirages"))  for i in range(1,n+1):     b=choice(Urne)     Tirages.append(b)     N_0=Tirages.count(0)     N_1=Tirages.count(1)     repere.plot(i,N_0/i,'m.')     repere.plot(i,N_1/i,'gx')  repere.show()</pre>	<pre># Importation des bibliothèques lycée et random  # Urne est une liste. # La liste Tirages est vide. # Saisie d'un entier affecté à la variable n. # (le texte "nombre de tirages" est affiché)  # i est un entier allant de 1 à n.  # On choisit au hasard un élément de Urne et on # l'affecte à la variable b. # On ajoute l'élément b à Tirages.  # On compte le nombre d'apparitions de l'élément # 0 dans Tirages. # On compte le nombre d'apparitions de l'élément # 1 dans Tirages. # On affiche le point de coordonnées(i;N_0/i) dans # le repère avec la marque . de couleur mauve. # On affiche le point de coordonnées (i;N_1/i) # dans le repère avec la marque x de couleur # verte. On remarque qu'il n'y a pas de end ; c'est l'indentation (décalage) du bloc qui le remplace.  # Affichage du repère</pre>
--	--

1. Commenter le graphique obtenu après exécution du programme. On copiera ce graphique sur le compte rendu.
2. Modifier ce programme afin d'évaluer la probabilité de tirer six 0 sur 10 tirages.
3. Quelle loi permet de déterminer la probabilité de l'évènement précédent ? Déterminer cette probabilité.

## Annexe 7 : Corrigé de la séquence 2

```
# Tirages avec remise dans une urne fixée
# On affiche les tirages au fur et à mesure
# On évalue la probabilité de tirer six 0 avec 10 tirages
```

### Programme Python

```
from lycee import *
from random import *

Urne=[0,1,0,0,0,1]

c=0

for j in range(1,10001):
    Tirages=[]

    for i in range(1,11):
        b=choice(Urne)
        Tirages.append(b)

    N_0=Tirages.count(0)

    print(Tirages)

    if N_0==6:
        c=c+1

print(c/1000)
```

### Commentaires

```
# Importation des bibliothèques lycée et random

# il s'agit d'un compteur initialisé à 0
# on répète 10000 fois 10 tirages dans l'urne

# on effectue les 10 tirages

# on compte le nombre de 0 dans Tirages

# on affiche la liste Tirages

# on ajoute 1 à chaque fois que l'évènement considéré
est réalisé

# fréquence de l'évènement considéré
```

L'exécution du programme renvoie une fréquence d'environ 0,23.

Plus précisément, si la valeur renvoyée est de 0,2315 alors on sait qu'un intervalle de confiance à 95 % de la probabilité cherchée est  $\left[ 0,2315 - \frac{1}{\sqrt{10000}} ; 0,2315 + \frac{1}{\sqrt{10000}} \right]$ , soit  $[0,2215 ; 0,2415]$ .

Le calcul exact de cette valeur peut être déterminé à l'aide de la loi binomiale.

En effet, les 10 tirages étant réalisés avec remise, on est en présence d'un schéma de Bernoulli à 10 répétitions où le succès est l'évènement "obtenir 0" de probabilité  $2/3$ . La variable X comptant le nombre de succès suit alors la loi binomiale de paramètres  $n = 10$  et  $p = 2/3$ .

Ainsi,  $P(X = 6) \approx 0,2276$  à  $10^{-4}$  près.

## Annexe 8 : séquence 3

### **Exemple 2 - Tirages sans remise dans une urne**

Voici un programme Python commenté, simulant des tirages sans remise dans une urne jusqu'à la vider.

#### Programme Python

```
from lycee import *
from random import *

Urne=[0,1,0,0,0,1]

print(Urne)

Tirages=[]

while Urne != []:

    b=choice(Urne)

    Urne.remove(b)

    Tirages.append(b)

print(Tirages)
```

#### Commentaires

```
# Importation des bibliothèques lycée et random

# Urne est une liste

# Affichage de Urne

# La liste Tirages est vide

# Tant que Urne n'est pas vide

# On choisit au hasard un élément de Urne et on l'affecte à la variable b

# On retire l'élément b de Urne

# On ajoute l'élément b à Tirages

# On affiche Tirages
```

1. Exécuter ce programme.
2. Modifier ce programme afin d'évaluer la probabilité de l'évènement :  
« Tirages = Urne initiale ».
3. Déterminer cette probabilité.

### Annexe 9 : Corrigé de la séquence 3

# Tirages sans remise dans une urne fixée jusqu'à la vider  
# On évalue la probabilité d'obtenir l'évènement " Tirages = Urne initiale "

#### Programme Python

```
from lycee import *
from random import *

c = 0

for j in range(1,10001):
    Urne = [0,1,0,0,0,1]
    Urne_0 = [0,1,0,0,0,1]
    Tirages = []

    while Urne != []:
        b = choice(Urne)
        Urne.remove(b)
        Tirages.append(b)

    if Tirages == Urne_0:
        c = c + 1

print(c/10000)
```

#### Commentaires

```
# Importation des bibliothèques lycée et random

# Il s'agit d'un compteur initialisé à 0

# On répète 10 000 fois le vidage de l'urne

# Urne_0 est une urne tampon ; il faut garder une trace de l'urne
initiale. !! Ne pas écrire Urne_0 = Urne !!

# On ajoute 1 au compteur à chaque fois que l'évènement
" Tirages = Urne initiale " est réalisé

# Fréquence de l'évènement " Tirages = Urne initiale "
```

L'exécution du programme renvoie une fréquence d'environ 0,07.

Plus précisément, si la valeur renvoyée est de 0,0689 alors on sait qu'un intervalle de confiance à 95 % de la fréquence théorique est  $\left[0,0689 - \frac{1}{\sqrt{10000}} ; 0,0689 + \frac{1}{\sqrt{10000}}\right]$ , soit  $[0,0589 ; 0,0789]$ .

Le calcul exact de cette valeur peut être déterminé en dénombrant le nombre de façons de placer les deux " 1 " dans les six places possibles ; il y en a  $\binom{6}{2} = \frac{6 \times 5}{2} = 15$ .

La probabilité cherchée est donc de 1/15, soit environ 0,0667 à  $10^{-4}$  près.



## Annexe 10 : TP des tortues



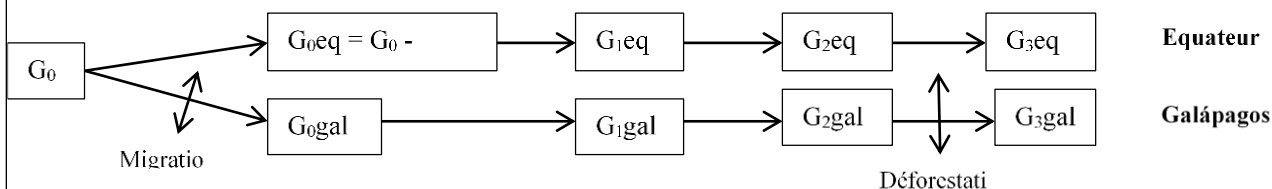
### TP 5 : Evolution génétique des populations au cours du temps

**Situation initiale :** En Equateur vit une population de tortues (appelée  $G_0$ ) comportant 4 mâles et 4 femelles dont les carapaces peuvent être vertes ou marron. Dans un milieu équatorial forestier, comme en Equateur (ou aux îles Galápagos), l'ensemble des tortues peuvent se camoufler (plantes vertes, feuilles mortes) et donc y survivre et se reproduire.

La couleur des carapaces est sous la dépendance d'un gène sous la forme de 2 allèles v,m  
L'allèle m est toujours dominant.

Un glissement de terrain prend au piège deux tortues de  $G_0$  (un mâle et une femelle) qui sont emportés sur un radeau (formant alors la population  $G_{0gal}$ ). Le radeau et les 2 tortues dérivent sur la mer jusqu'aux îles Galápagos où elles se reproduisent formant les tortues de la génération  $G_{1gal}$  qui sera à l'origine de la génération  $G_{2gal}$  puis  $G_{3gal}$ . Pendant ce même temps, les autres couples de la population initiale ( $G_{0eq} = G_0 - G_{0gal}$ ) se reproduisent en Equateur durant le même temps soit 3 générations :  $G_{1eq}$ ,  $G_{2eq}$ ,  $G_{3eq}$

Cependant, l'île des Galápagos subit, à la génération  $G_2$ , une déforestation orchestrée par l'homme. Les tortues de carapaces vertes ne sont donc plus camouflées dans cet environnement et sont donc repérées par les prédateurs et mangées. Les survivantes ( $G_2'gal$ ) se reproduisent et forment les générations  $G_3gal$ .



#### **Problème : Comment suivre et expliquer l'évolution de la population de tortues au cours du temps ?**

**Consigne :** Vous êtes généticiens des populations, spécialistes des tortues. Le ministère de l'écologie et de la biodiversité vous charge d'établir l'évolution des fréquences alléliques et phénotypiques des tortues en Equateur ou aux Galápagos durant votre mission de 5 ans (soit durant 4 générations de tortues).

En utilisant les supports à disposition, modéliser l'évolution de la population de tortues au cours du temps et en prenant en compte les différents évènements.

#### **Production**

Les résultats des fréquences alléliques de chaque population (équateur et Galápagos) pour chaque génération seront reportés dans un tableau Excel puis un graphique Excel de l'évolution des fréquences alléliques des 2 populations de tortues sur les 3 générations est réalisé. En comparant vos résultats à celui d'un autre groupe, rédigez un rapport au ministre expliquant l'origine des variations alléliques observées.

#### **Modélisation de l'évolution des deux populations de tortues sur 3 générations**

##### Protocole de modélisation de la reproduction sexuée des tortues

Chaque parent a 2 allèles mais n'en transmet qu'un seul à chaque descendant, de manière aléatoire.

- Utiliser le dé pour réaliser un tirage au sort de l'allèle transmis par la mère ( $n^{\circ}$  impair du dé = 1<sup>er</sup> allèle transmis ;  $n^{\circ}$  pair = 2<sup>ème</sup> allèle transmis) pour chaque descendant
- Utiliser le dé pour réaliser un tirage au sort de l'allèle transmis par le père ( $n^{\circ}$  impair du dé = 1<sup>er</sup> allèle transmis ;  $n^{\circ}$  pair = 2<sup>ème</sup> allèle transmis) pour chaque descendant
- Notez les allèles possédés par chaque descendant du couple de parent (et leur phénotype)

# Annexe 11 : Fichier Excel

**Consignes:**  
 Chaque couple a 4 descendants.  
 Le tirage des couples est un tirage aléatoire (fonction Alea).  
 Les génotypes des descendants des couples se font avec un tirage aléatoire d'un allèle par parent (fonction Alea).  
 Reporter les résultats des tirages dans les cases correspondantes, à la main: numéro du couple ou allèle choisi.  
 Les fréquences des différents allèles, à chaque génération, sont calculées à la main.

**Choix du couple 8**  
 Réaliser successivement 2 tirages (F9)  
 Recommencer le tirage si le même numéro est tiré 2 fois

**Reproduction sexuée**  
 choix d'un allèle par parent  
 Réaliser successivement 2 tirages (F9)  
 (1 tirage par parent)  
 si 1: 1er allèle (à gauche),  
 si 2: 2ème allèle (à droite)

**départ d'un couple sur les Galapagos**

**les 3 autres couples restent en Equateur**

**Choix des 3 couples 2**  
 Réaliser successivement 6 tirages (F9)  
 Recommencer le tirage si le même numéro est tiré 2 fois

**Reproduction sexuée**  
 choix d'un allèle par parent  
 Réaliser successivement 2 tirages (F9)  
 (1 tirage par parent)  
 si 1: 1er allèle (à gauche),  
 si 2: 2ème allèle (à droite)

**Choix des deux couples 1**  
 Réaliser successivement 4 tirages (F9)  
 Recommencer le tirage si le même numéro est tiré 2 fois

**Reproduction sexuée**  
 choix d'un allèle par parent  
 Réaliser successivement 2 tirages (F9)  
 (1 tirage par parent)  
 si 1: 1er allèle (à gauche),  
 si 2: 2ème allèle (à droite)

**Choix des 6 couples 9**  
 Réaliser successivement 12 tirages (F9)  
 Recommencer le tirage si le même numéro est tiré 2 fois

**Reproduction sexuée**  
 choix d'un allèle par parent  
 Réaliser successivement 2 tirages (F9)  
 (1 tirage par parent)  
 si 1: 1er allèle (à gauche),  
 si 2: 2ème allèle (à droite)

**déforestation entraînant la mort des tortues de phénotype [vert]**

**Choix des 3 couples 1**  
 Réaliser successivement 6 tirages (F9)  
 Recommencer le tirage si le même numéro est tiré 2 fois ou si l'individu a été tué par la déforestation

**Reproduction sexuée**  
 choix d'un allèle par parent  
 Réaliser successivement 2 tirages (F9)  
 (1 tirage par parent)  
 si 1: 1er allèle (à gauche),  
 si 2: 2ème allèle (à droite)

**Choix des 12 couples 8**  
 Réaliser successivement 24 tirages (F9)  
 Recommencer le tirage si le même numéro est tiré 2 fois

**Reproduction sexuée**  
 choix d'un allèle par parent  
 Réaliser successivement 2 tirages (F9)  
 (1 tirage par parent)  
 si 1: 1er allèle (à gauche),  
 si 2: 2ème allèle (à droite)

**Production attendue: Réaliser un graphique des évolutions des 3 fréquences alléliques (v, m, b) en fonction des générations pour la population des Galapagos**

temps	v	m
G0	0	0,5
G0gal	1	0,75
G1gal	2	0,63
G2gal	3	0,75
G2gal	4	0,5
G3gal	5	0,31

temps	v	m
G0	0	0,5
G0eq	1	0,4
G1eq	2	0,42
G2eq	3	0,5
G3eq	4	0,55

D	v	m
G0	0	0,5
G0eq	1	0,4
G1eq	2	0,62
G2eq	3	0,6
G3eq	4	0,5

**Production attendue: Réaliser un graphique des évolutions des 3 fréquences alléliques (v, m, b) en fonction des générations pour la population de l'équateur**

**Evolution des fréquences alléliques de la population galapagos**

**Evolution des fréquences alléliques de la population équateur**

**Evolution des fréquences alléliques de la population équateur sim2**

## Annexe 12 : séquence 4

### Hardy-Weinberg et populations de taille finie

Nous allons étudier dans cette séquence, grâce à la simulation, l'évolution au cours des générations des distributions alléliques dans des populations de taille finie et avec éventuellement plus de deux allèles.

Les conditions du modèle de Hardy-Weinberg énoncées dans le devoir seront respectées sauf celle portant sur la taille de la population et éventuellement d'autres à préciser.

#### Protocole et procédure de la simulation :

Comment simuler une population d'individus caractérisés par leur génotype ?

Comment déterminer la distribution allélique dans cette population ?

Comment former les générations suivantes ?

Comment représenter l'évolution des distributions alléliques au cours des générations ?

Les étapes suivantes vont vous aider à répondre à ces questions ; vous pourrez ensuite élaborer votre programme en Python. Il est important de garder une trace de votre travail en tenant à jour un cahier de bord numérique qui devra être visé à chaque séance par votre professeur.

*Bon courage,*

#### Etape 1 : Simuler une population d'individus caractérisés par leur génotype

Nous avons vu dans les séquences 2 et 3 que Python gère facilement les listes.

Or une population est une liste d'individus et chaque individu est caractérisé par une liste de 2 (ou 3) allèles. Exemple :

[" m ", " v "] ou [" v ", " m "] pour un individu de génotype m/v.

[[" m ", " m "], [" m ", " v "], [" v ", " v "]] pour une population formée de 3 individus.

Vous pouvez ainsi créer un programme (ou une fonction) renvoyant une population initiale constituée de n individus générés au hasard ou non. Vous pouvez éventuellement imposer la distribution allélique dans cette population.

#### Etape 2 : Déterminer la distribution allélique dans cette population

Un outil très utile du langage Python est la possibilité d'utiliser l'instruction "for" avec une variable parcourant une liste. Exemple :

```
Liste = [[0,1],[1,1],[1,0],[1,0]]
```

```
.....
```

```
for I in Liste :
```

```
.....
```

```
" A compléter afin de compter le nombre de 0 et de 1 figurant dans Liste. "
```

Créer un programme (ou une fonction) renvoyant la proportion de chaque allèle dans la population.

#### Etape 3 : Génération suivante

On peut choisir (ou pas) de considérer une population qui réunit une population mâle et une population femelle de même effectif.

On considère que la génération suivante est constituée uniquement des enfants issus des couples formés à la génération précédente.

Reste à définir le nombre de descendants par couple, c'est-à-dire la capacité de reproduction supposée constante dans la population.

Créer un programme (ou une fonction) qui à partir d'une population initiale, renvoie la liste des enfants (mâles et femelles) issus de tous les couples tirés au hasard et sans remise (jusqu'à vider la population initiale).

#### Etape 4 : Boucle

Répéter l'étape précédente pour couvrir plusieurs générations et afficher le graphique donnant l'évolution des distributions alléliques.

### Annexe 13 : Programme élève

Lydia-Lucie-  
harlotte-Adonis

#### Programme Python

```
from lycee import *
from random import *

G0m = ["m", "v"], ["v", "m"], ["m", "m"], ["m", "m"] #population initiale
G0f=["v", "m"], ["v", "v"], ["v", "v"], ["v", "m"]
G0 = G0m + G0f

N_M = 0 #détermination des fréquences
N_V = 0
for I in G0:
    N_M= I.count("m")+N_M
    N_V= I.count("v")+N_V
M = (N_M/16)
V = (N_V/16)
repere.plot(0,M, 'm. ')
repere.plot(0,V, 'gx')

G1=[] #population suivante

for k in range (1,16):
    G1m=[]
    G1f=[]
    while G0f != []:
        tm=choice(G0m) #choisir male dans G0m
        G0m.remove(tm) #enlever male choisi de G0m
        tf=choice(G0f) #choisir femelle dans G0f
        G0f.remove(tf) #enlever femelle choisi de G0f
        am=choice(tm)
        af=choice(tf)

        G1m.append([am,af]) #création population suivante
        am=choice(tm)
        af=choice(tf)
        G1m.append([am,af])
        am=choice(tm)
        af=choice(tf)
        G1f.append([am,af])
        am=choice(tm)
        af=choice(tf)
        G1f.append([am,af])

    G0m=G1m;G0f=G1f
    G1 = G1f+G1m

N_M = 0 #fréquences population suivante
N_V = 0
for I in G1:
    N_M= I.count("m")+N_M
    N_V= I.count("v")+N_V
M = (N_M/ (2*len(G1)))
V = (N_V/ (2*len(G1)))

print(len(G1)) #affichage et graphique
print(M,V)
repere.plot(k,M, 'm. ')
repere.plot(k,V, 'gx')
repere.show ()
```



## Annexe 14 : Carnet de bord (groupe Lydia-Adonis-Charlotte-Lucie)

### Carnet de bord AP SVT /Maths :

4/12 :

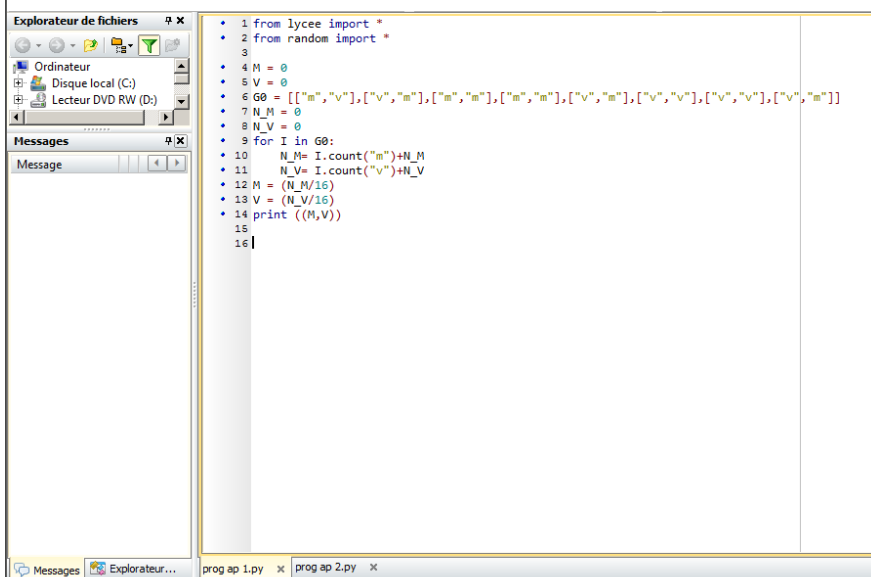
Chaque individu est une liste de deux allèles m et/ou v.

On prend huit individus de départ : [m;v], [v;m], [m;m], [m;m], [v;m], [v;v], [v;v], [v;m].

On appelle  $G_0$  la liste qui contient les individus de la génération de départ.

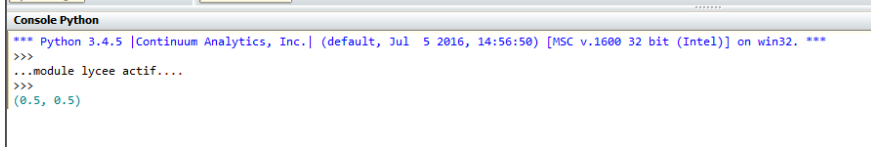
On considère que chaque couple a 4 descendants et qu'il y a 50 % de mâle et de femelle à chaque génération.

On considère également que les individus de la génération précédente décèdent après avoir eu leurs descendants.

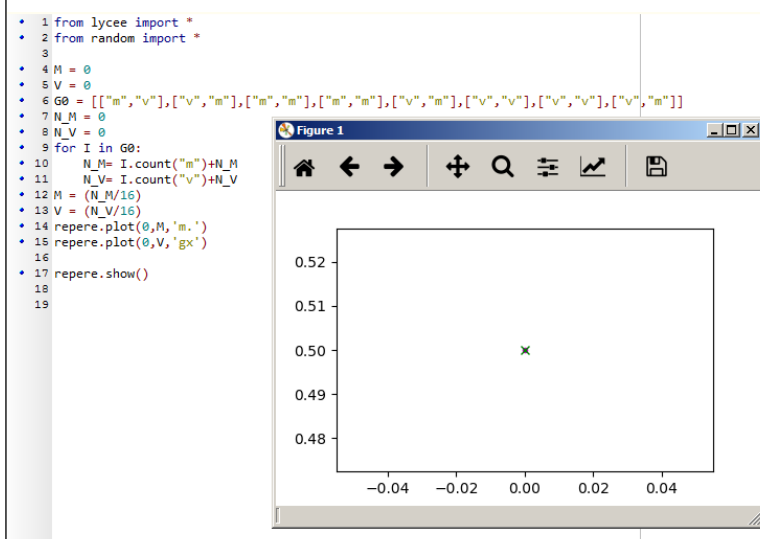


```
1 from lycee import *
2 from random import *
3
4 M = 0
5 V = 0
6 G0 = [['m','v'],['v','m'],['m','m'],['m','m'],['v','m'],['v','v'],['v','v'],['v','m']]
7 N_M = 0
8 N_V = 0
9 for I in G0:
10     N_M= I.count("m")+N_M
11     N_V= I.count("v")+N_V
12 M = (N_M/16)
13 V = (N_V/16)
14 print ((M,V))
15
16|
```

Programme pour avoir la fréquence des allèles m et v à la génération  $G_0$



```
*** Python 3.4.5 [Continuum Analytics, Inc.] (default, Jul 5 2016, 14:56:50) [MSC v.1600 32 bit (Intel)] on win32. ***
>>>
...module lycee actif...
>>>
(0.5, 0.5)
```



programme qui donne le graphique des fréquences de m et v à la génération  $G_0$

Création du premier couple : on divise  $G_0$  en deux listes de 4 individus,  $G_{0m}$  (mâle) et  $G_{0f}$  (femelle).

## Carnet de bord AP SVT /Maths : (suite)

```
1 from lycee import *
2 from random import *
3
4 G1=[]
5
6 G0m = [['m','v'],['v','m'],['m','m'],['m','m']]
7 G0f=[['v','m'],['v','v'],['v','v'],['v','m']]
8 G0 = G0m + G0f
9 N_M = 0
10 N_V = 0
11 for I in G0:
12     N_M= I.count("m")+N_M
13     N_V= I.count("v")+N_V
14 M = (N_M/16)
15 V = (N_V/16)
16 repere.plot(0,M,'m.')
17 repere.plot(0,V,'gx')
18
19
20
21 G1m=[]
22 G1f=[]
23 G1 = G1m + G1f
24 m=[]
25 f=[]
26 tm=choice(G0m) #choisir male dans G0m
27 G0m.remove(tm) #enlever male choisi de G0m
28 tf=choice(G0f) #choisir femelle dans G0f
29 G0f.remove(tf) #enlever femelle choisi de G0f
30 for k in range (1,5):
31     am=choice(tm)
32     af=choice(tf)
33     G1.append([am,af])
34 print(G1)
35
```

On a créer les 4 enfants du 1<sup>er</sup> couple .

le 8/01 :

```
1 from lycee import *
2 from random import *
3
4 G1=[]
5
6 G0m = [['m','v'],['v','m'],['m','m'],['m','m']]
7 G0f=[['v','m'],['v','v'],['v','v'],['v','m']]
8 G0 = G0m + G0f
9 N_M = 0
10 N_V = 0
11 for I in G0:
12     N_M= I.count("m")+N_M
13     N_V= I.count("v")+N_V
14 M = (N_M/16)
15 V = (N_V/16)
16 repere.plot(0,M,'m.')
17 repere.plot(0,V,'gx')
18
19
20
21 G1m=[]
22 while G0f != []:
23     tm=choice(G0m) #choisir male dans G0m
24     G0m.remove(tm) #enlever male choisi de G0m
25     tf=choice(G0f) #choisir femelle dans G0f
26     G0f.remove(tf) #enlever femelle choisi de G0f
27     for k in range (1,5):
28         am=choice(tm)
29         af=choice(tf)
30         G1.append([am,af])
31 print(G1)
32
```

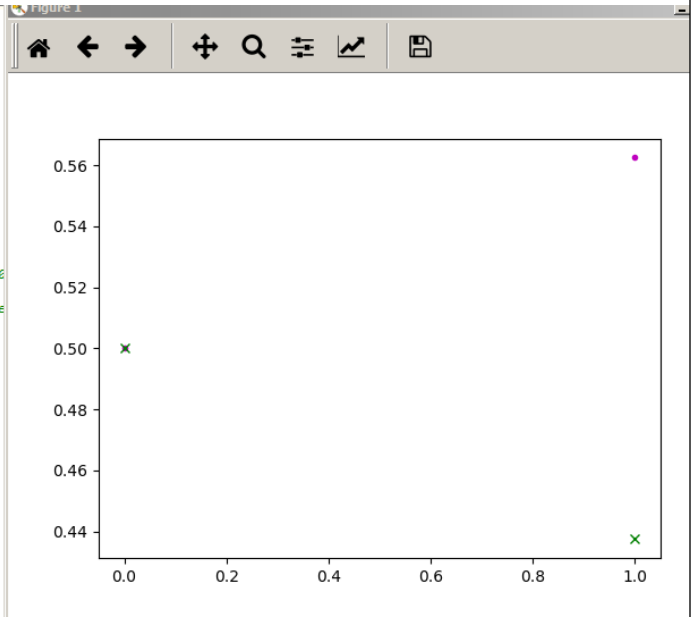
File "E:\TermS1\AP maths\prog ap 3.py", line 23, in <module>  
tm=choice(G0m)  
File "E:\EduPython\app\lib\random.py", line 255, in choice  
raise IndexError('Cannot choose from an empty sequence')  
IndexError: Cannot choose from an empty sequence  
>>> [['m', 'v'], ['v', 'v'], ['v', 'v'], ['v', 'v'], ['m', 'v'], ['m', 'v'], ['m', 'v'], ['m', 'v'], ['v', 'm'], ['m', 'm'], ['m', 'm'], ['m', 'v'], ['m', 'm'], ['m', 'v']]  
>>>

On a créé la génération 1 (G1) constituée des 16 enfants.

```

• 9 N_M = 0
• 10 N_V = 0
• 11 for I in G0:
• 12     N_M= I.count("m")+N_M
• 13     N_V= I.count("v")+N_V
• 14 M = (N_M/16)
• 15 V = (N_V/16)
• 16 repere.plot(0,M,'m.')
• 17 repere.plot(0,V,'gx')
• 18
• 19
• 20
• 21 G1=[]
• 22 while G0f != []:
• 23     tm=choice(G0m)           #choisir male dans G0m
• 24     G0m.remove(tm)          #enlever male choisi de G0
• 25     tf=choice(G0f)           #choisir femelle dans G0f
• 26     G0f.remove(tf)          #enlever femelle choisi de G0f
• 27     for k in range (1,5):
• 28         am=choice(tm)
• 29         af=choice(tf)
• 30         G1.append([am,af])
• 31 print(G1)
• 32
• 33
• 34 N_M = 0
• 35 N_V = 0
• 36 for I in G1:
• 37     N_M= I.count("m")+N_M
• 38     N_V= I.count("v")+N_V
• 39 M = (N_M/(2*len(G1)))
• 40 V = (N_V/(2*len(G1)))
• 41 repere.plot(1,M,'m.')
• 42 repere.plot(1,V,'gx')
• 43 repere.show ()
• 44

```



Graphique des fréquences alléliques pour les générations 0 et 1.

le 29/01 :

Il nous restait à boucler le programme, nous avons bien divisé la génération suivante (G1) en deux groupes : femelle et mâle.

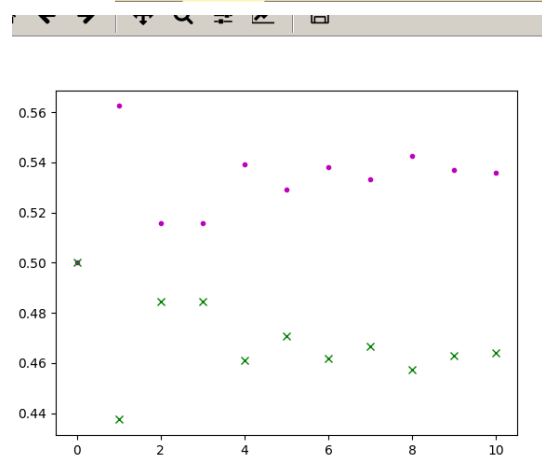
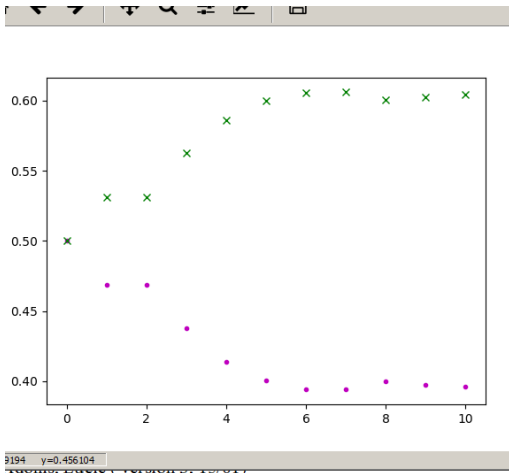
Nous avons bouclé le programme, il tourne pour 10 générations.

Ça marche !!!

```

• 1 from Joyce import *
• 2 from random import *
• 3 G1=[]
• 4 G0m = [{"m": "v"}, {"m": "m"}, {"m": "v"}, {"m": "m"}, {"m": "v"}, {"m": "m"}, {"m": "v"}, {"m": "m"}, {"m": "v"}, {"m": "m"}, {"m": "v"}, {"m": "m"}, {"m": "v"}, {"m": "m"}, {"m": "v"}, {"m": "m"}]
• 5 G0f = G0m * G0f
• 6 N_M = 0
• 7 N_V = 0
• 8 for I in G0:
• 9     N_M= I.count("m")+N_M
• 10    N_V= I.count("v")+N_V
• 11 M = (N_M/16)
• 12 V = (N_V/16)
• 13 repere.plot(0,M,'m.')
• 14 repere.plot(0,V,'gx')
• 15
• 16
• 17
• 18 for k in range (1,11):
• 19     G1=[]
• 20     while G0f != []:
• 21         tm=choice(G0m)           #choisir male dans G0m
• 22         G0m.remove(tm)          #enlever male choisi de G0m
• 23         tf=choice(G0f)           #choisir femelle dans G0f
• 24         G0f.remove(tf)          #enlever femelle choisi de G0f
• 25         am=choice(tm)
• 26         af=choice(tf)
• 27         G1.append([am,af])
• 28         am=choice(tm)
• 29         af=choice(tf)
• 30         G1.append([am,af])
• 31         am=choice(tm)
• 32         af=choice(tf)
• 33         G1.append([am,af])
• 34         am=choice(tm)
• 35         af=choice(tf)
• 36         G1.append([am,af])
• 37         am=choice(tm)
• 38         af=choice(tf)
• 39         G1.append([am,af])
• 40 G0m=G1m;G0f=G1f
• 41 G1 = G1*G1m
• 42 N_M = 0
• 43 N_V = 0
• 44 for I in G1:
• 45     N_M= I.count("m")+N_M
• 46     N_V= I.count("v")+N_V
• 47 M = (N_M/(2*len(G1)))
• 48 V = (N_V/(2*len(G1)))
• 49 repere.plot(1,M,'m.')
• 50 repere.plot(1,V,'gx')
• 51 repere.show ()
• 52

```



Quelques graphiques que nous avons obtenues.

Lydia, Charlotte, Adonis, Lucie ( version 3, 29/01)