

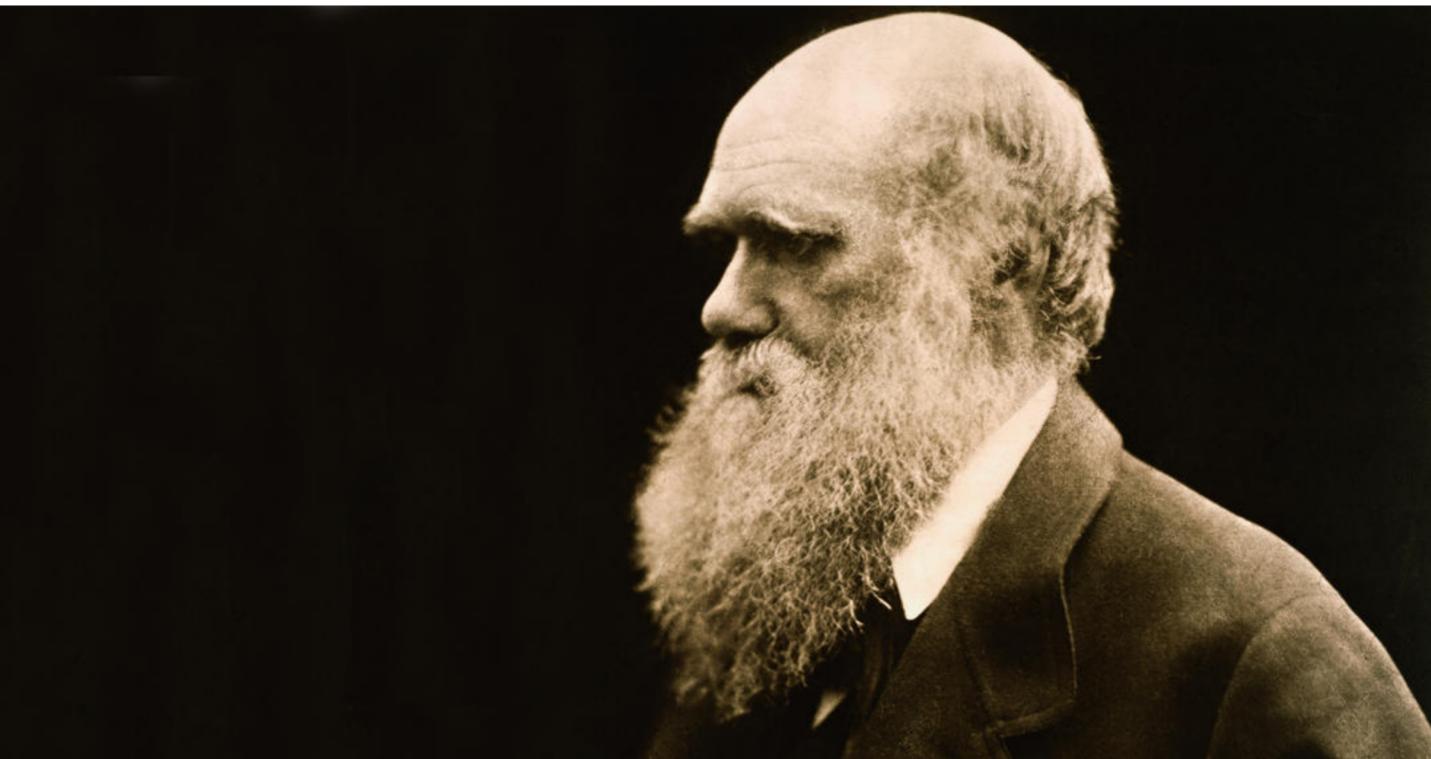
Modélisation Mathématique des effets de la sélection naturelle

Achille Baucher
Ibtissam Lachhab

Younes Belkada
Igor Rouzine

Laurent Dang-Vu
(Encadrant du LCQB)

12 juillet 2021



1 Sommaire

Table des matières

1	Sommaire	2
2	Introduction	3
3	Présentation	3
3.1	Situation	3
3.2	Notations	3
3.3	Outils utilisés	3
4	Travaux préliminaires	3
5	Cas sauvage : Absence de mutations	4
5.1	Linéarité de l'équation	4
5.2	Détermination de la fréquence f de n_1 dans N	5
5.3	Conclusion, et remarques	6
6	Cas de mutations régulières	9
6.1	Présentation et premières équations	9
6.1.1	Sous-cas numéro 1 : facteur de mutation plus important	10
6.2	Sous cas numéro 2 : sélection plus importante	11
6.2.1	Etude disjointe de l'évolution de la fréquence	12
6.2.2	L'état d'équilibre	12
6.2.3	Début	13
6.2.4	Demi temps	14
6.3	Conclusions et graphique	14
7	Cas avec deux mutations	16
7.1	Présentation	16
7.2	Premières équations pour l'évolution des populations	16
7.3	Détermination des fréquences en fonction de S_{ij} à l'état stationnaire	18
7.3.1	Cas général	18
7.3.2	Allèles mutés, combinaisons et épistases	18
7.3.3	Les individus de n_{11} sont les plus avantagés	19
7.3.4	Les populations n_{01} et n_{10} sont les plus avantagées	22
7.3.5	Conclusion sur les populations à l'état stable	23
8	L'auto-stop génétique	24
8.1	Présentation	24
8.2	Exemple	24
8.3	Conclusion :	25
9	Conclusion générale :	26

2 Introduction

L'objet de l'étude est d'observer l'évolution temporelle quantitative de plusieurs populations en compétition dans un environnement limité. Nous complexifierons le modèle au fur et à mesure, en partant d'un cas simple pour introduire ensuite des mutations régulières.

3 Présentation

3.1 Situation

Considérons ainsi deux populations disjointes n_0 et n_1 , mais vivant dans le même milieu. n_0 représentant la population saine (ie dépourvue de la mutation sur laquelle on se concentre) et n_1 représente la population mutée. Dans ce premier cas, les individus de n_1 auront un petit peu plus de chance de se reproduire que les autres. Nous étudierons alors comment évolue la démographie de ces deux populations.

3.2 Notations

Les notations utilisées :

n_0 : Population sauvage, non mutée

n_1 : Population mutée, plus performante en reproduction

N : Population totale = $n_1 + n_0$ constante

K_0 : Proportion des individus de n_0 qui se reproduiront par unité de temps

K_1 : Proportion des individus de n_1 qui se reproduiront par unité de temps

d_0 : Proportion des individus de n_0 qui mourront par unité de temps

d_1 : Proportion des individus de n_1 qui mourront par unité de temps

$S > 0$ tel que $K_1 = (1 + S) K_0$: coefficient de compétitivité

3.3 Outils utilisés

Durant notre étude nous utiliserons principalement l'analyse différentielle pour la modélisation et les calculs et MatLab pour les graphiques.

4 Travaux préliminaires

Nous souhaitons donc étudier l'évolution des deux populations en fonction du temps. Pour ce faire, nous pouvons constater que l'effectif d'une population isolée dépend du nombre de naissances et du nombre de morts. Pour étudier l'évolution d'une population lorsque le temps varie de manière continue, on use d'équations différentielles. Commençons par constituer les dérivées de n_0 et n_1 . Par définition de K_0 (c.f. K_1), on sait que $K_0 * n_0$ (c.f. $K_1 * n_1$) est la quantité d'individus de n_0 (c.f. n_1) qui naissent par unité de temps et $d_0 * n_0$ (c.f. $d_1 * n_1$) la quantité d'individus qui meurt dans ce même intervalle de temps. Ainsi l'évolution des populations est donnée par les équations différentielles suivantes :

$$\frac{dn_0}{dt} = K_0 n_0 - d_0 n_0 \quad (1)$$

$$\frac{dn_1}{dt} = K_1 n_1 - d_1 n_1 \quad (2)$$

Ceci modélise bien la situation ; la variation de n_0 et de n_1 en fonction du temps correspond au gain (naissance) moins les pertes (décès). Nous allons consacrer une partie de notre étude en supposant ces équations vraies sans prendre en compte pour l'instant les autres paramètres qui compliqueront le problème.

Dans ce problème, nous imposons qu'un individu de n_0 a autant de chance de mourir qu'un individu de n_1 , les différences se feront sur les avantages de reproduction. On a donc $d_0 = d_1$. Pour simplifier nos équations, nous introduisons alors les notations suivantes : $k_0 = \frac{K_0}{d_0}$, $k_1 = \frac{K_1}{d_1} = \frac{K_1}{d_0}$ et $dt' = dt * d_0$, qui correspond à une génération. En divisant par n_0 dans (1) et (2), on a donc :

$$\boxed{\frac{dn_0}{dt'} = (k_0 - 1)n_0} \quad (3)$$

$$\boxed{\frac{dn_1}{dt'} = (k_1 - 1)n_1} \quad (4)$$

On prendra pour la suite $dt = dt'$.

5 Cas sauvage : Absence de mutations

Pour cette première étude, nous ne ferons pas intervenir de mutations. Nous n'observerons donc que les effets de la compétition dans un milieu limité, car N , la population totale, est une constante. Nous allons considérer que les individus de les individus de n_1 ont un avantage en terme de reproduction sur ceux de n_0 :

$$k_1 = (1 + S) * k_0 \text{ avec } S > 0$$

Donc d'après (4) :

$$\boxed{\frac{dn_1}{dt'} = (k_0(1 + S) - 1)n_1} \quad (5)$$

5.1 Linéarité de l'équation

Pour nous aider à résoudre l'équation différentielle (5), il faut d'abord déterminer si elle est linéaire, c'est-à-dire si k_0 (et donc k_1) sont des constantes. Supposons que k_0 ne dépende pas du temps. On aurait donc :

$$\frac{dn_0}{dt'} = (k_0 - 1)n_0 \implies n_0 = A_0 * e^{(k_0 - 1)*t} \quad (6)$$

$$\frac{dn_1}{dt'} = (k_0(1 + S) - 1)n_1 = A_1 * e^{(k_0(1 + S) - 1)*t} \quad (7)$$

Avec A_0 et A_1 constantes réelles positives car n_0 et n_1 sont des nombres positifs.

Vérifions alors si $N = n_0 + n_1$ reste bien constant, selon la valeur de k_0 :

-Soit $k_0 < \frac{1}{1+S}$: On a alors $k_0 < 1$ et $k_0(1 + S) - 1 < 0$ donc on remarque dans (6) et (7) que n_0 et n_1 sont décroissants donc que N aussi.

-Soit $k_0 = \frac{1}{1+S}$: On a alors $k_0 < 1$ et $k_0(1 + S) - 1 = 0$ et on en déduit que n_0 décroît et que n_1 est constant, donc que N décroît.

-Soit $k_0 > \frac{1}{1+S}$: On a alors $k_0(1 + S) - 1 > 0$ donc n_1 tend vers l'infini et N aussi.

On peut donc observer que dans tous les cas N n'est pas constant. Or pour introduire des mécanismes de compétition dans un système limité beaucoup plus intéressant à étudier, la population totale N doit être constante. On en déduit donc par contraposée que k_0 doit dépendre du temps. Ce qui nous amène à des équations différentielles non linéaires.

5.2 Détermination de la fréquence f de n_1 dans N

On définit f_0 et f , respectivement les fréquences de la population n_0 et de la population n_1 telles que :

$$f_0 = \frac{n_0}{N} \quad (8)$$

$$f = \frac{n_1}{N} = \frac{N - n_0}{N} \quad (9)$$

Nous cherchons ici à déterminer l'expression de f . En introduisant la variable $\epsilon = k_0 - 1$ on obtient pour les dérivées de n_0 , n_1 et N :

$$\begin{aligned} \frac{dn_0}{dt} &= (k_0 - 1)n_0 = \epsilon * n_0 \\ \frac{dn_1}{dt} &= (k_0 - 1 + k_0 * S) * n_1 = (\epsilon + k_0 * S) * n_0 \\ \frac{dN}{dt} &= \frac{dn_1}{dt} + \frac{dn_0}{dt} = \epsilon n_0 + (\epsilon + k_0 * S) * n_1 \end{aligned}$$

On peut alors calculer la dérivée de f_1 :

$$\begin{aligned} \frac{df}{dt} &= \frac{d}{dt} \left(\frac{n_1}{N} \right) \\ &= \frac{N dn_1}{dt} - \frac{n_1 dN}{dt} \\ &= \frac{N(\epsilon + k_0 * S)n_1 - n_1(\epsilon n_0 + (\epsilon + k_0 * S)n_1)}{N^2} \\ &= \frac{n_1(\epsilon + k_0 * S)(n_0 + n_1) - \epsilon n_0 - (\epsilon + k_0 * S)n_1}{N} \\ &= f \frac{(\epsilon + k_0 * S)n_0 - \epsilon n_0}{N} \\ &= f \frac{k_0 * S * n_0}{N} \\ &= f(1 - f)k_0 * S \end{aligned}$$

On sait que la population ne change pas trop vite et donc que le nombre de décès est proche de celui du nombre de mort. On a donc $K_0 \approx d_0$ et donc $k_0 \approx 1$:

$$\frac{df}{dt} = f(1 - f)S$$

(10)

Essayons alors d'intégrer l'équation (10). On a :

$$\frac{df}{f(1 - f)} = S dt$$

Avec la méthode de la décomposition en éléments simples on obtient :

$$\begin{aligned}
\frac{df}{f} + \frac{df}{1-f} &= Sdt \\
\ln f - \ln(1-f) &= St + A \text{ (avec } A \text{ constante réelle)} \\
\ln \frac{f}{1-f} &= St + A \\
\frac{f}{1-f} &= e^{St+A} \\
f &= (1-f)e^{St+A} \\
f(t) &= \frac{e^{St+A}}{1+e^{St+A}}
\end{aligned}$$

On introduit une autre constante $A' = e^A$ ce qui donne :

$$f(t) = \frac{A'e^{St}}{1+A'e^{St}} \quad (11)$$

Déterminons à présent A' en fonction des conditions initiales $f(t=0) = f_0$:

$$f_0 = \frac{A'}{1+A'}$$

On obtient donc :

$$A' = \frac{f_0}{1-f_0} \quad (12)$$

Et finalement, en remplaçant (12) dans (11) on obtient la fréquence en fonction du temps et des conditions initiales :

$$f(t) = \frac{f_0 e^{S*t}}{1-f_0 + f_0 e^{S*t}}$$

(13)

5.3 Conclusion, et remarques

On choisit ici de prendre une faible population de n_1 initiale par rapport à n_0 . Nous verrons si et de quelle manière n_1 prend le dessus sur n_0 . Prévoyons d'abord analytiquement certains résultats du modèle grâce aux équations (10) et (13) :

Etats stationnaires : On voit avec l'équation (10) que la dérivée de f s'annule dans deux cas précis : $f = 0$ et $f = 1$. C'est très cohérent : s'il n'y a plus aucun membre de n_1 (c'est-à-dire si $f = 0$), alors ils ne peuvent pas se reproduire et f restera nul. Même chose pour la population de n_0 si $f = 1$.

Début : Au début, lorsque la fréquence f est très faible, on pourra négliger le terme en $(1-f)$ de l'équation (10) et l'approximer à 1 :

$$1-f \approx 1 \implies \frac{df}{dt} \approx f * S$$

Ce qui implique une forme exponentielle positive de f dans un premier temps, c'est-à-dire une croissance de n_1 de plus en plus rapide.

Demi temps : Voyons à quel moment la population n_1 doit atteindre la moitié de la population totale d'après l'équation (13) :

$$\begin{aligned}
 \frac{f_0 e^{S t_{1/2}}}{1 - f_0 + f_0 e^{S t_{1/2}}} &= 0.5 \\
 f_0 e^{S t_{1/2}} &= 0.5 * (1 - f_0 + f_0 e^{S t_{1/2}}) \\
 e^{S t_{1/2}} &= \frac{1 - f_0}{f_0} \\
 S t_{1/2} &= \ln \frac{1 - f_0}{f_0} \\
 t_{1/2} &= \frac{1}{S} \ln \frac{1 - f_0}{f_0}
 \end{aligned}$$

Ce temps est donc inversement proportionnel à S . Plus S augmente, plus la population n_1 est rapide à atteindre la moitié de la population totale.

Etat final : Si f s'approche de 1, alors on voit que le terme en $1 - f$ devient proche de 0, et que donc la dérivée de f s'approche de zero (et s'annule si $f = 1$). Cependant, elle reste positive donc on peut en conclure que f continuera à augmenter jusqu'à tendre vers 1, état stable.

Graphe : On trace alors le graphe de la f en fonction du temps avec la formule (13), pour plusieurs valeurs de S et pour f_0 arbitrairement choisi à $\frac{1}{100}$.

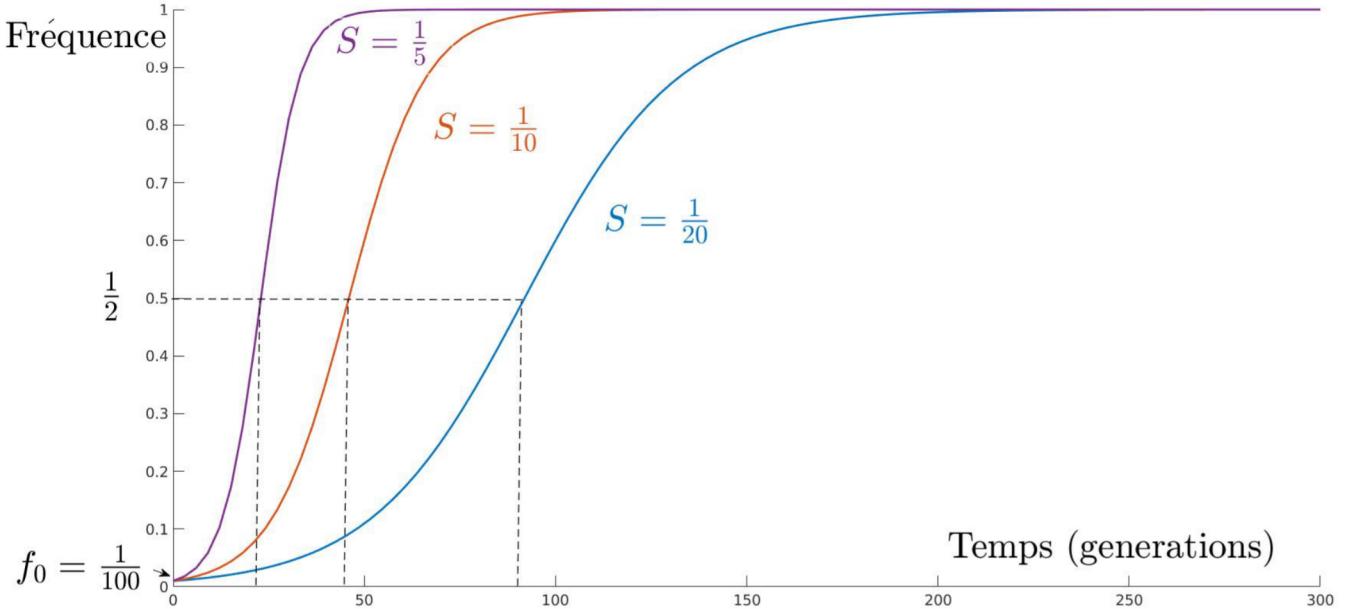


FIGURE 1 – Evolution de la fréquence de la population n_1 pour plusieurs valeurs de S , dans mutations ($\mu = 0$)

On observe précisément ce que nous avions pu prévoir analytiquement. La fréquence de la population favorisée évolue de manière exponentielle au début et ce jusqu'à ce qu'elle atteigne la moitié. Elle commence alors à croître plus lentement, évoluant selon une exponentielle négative et finira par tendre vers 1.

On peut donc en conclure qu'en introduisant un tout petit pourcentage d'individus présentant une mutation favorable à leur reproduction, ces derniers finiront par remplacer entièrement la population saine. A condition bien sûr que l'on se trouve dans un contexte de compétition, induit ici par la taille fixe de la population totale. On remarque que ce mécanisme fonctionne peu importe la puissance de l'avantage dont jouissent les individus mutés : celui-ci (le coefficient S) influe seulement sur la vitesse à laquelle les individus sains seront remplacés.

6 Cas de mutations régulières

6.1 Présentation et premières équations

Dans ce cas, nous nous intéressons à une population sauvage, n_0 , qui sera seule initialement. Puis une mutation va se produire et les individus mutés auront alors des avantages de reproduction. Il serait intéressant de voir comment évolue la démographie de la population sauvage et celle de la population mutée au cours du temps.

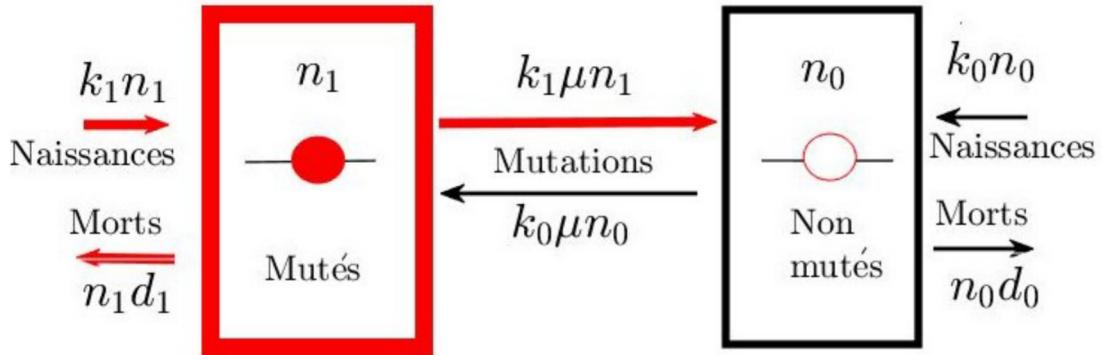
Nous allons introduire un coefficient de mutation μ : c'est le pourcentage de nouveaux-nés qui subiront une mutation.

μ est analogue à une très faible probabilité, il est donc supérieur à 0 et petit devant 1.

$$\mu > 0 \text{ et } \mu \ll 1$$

Pour simplifier le modèle, nous considérerons que la mutation ne se produit que sur un seul allèle, lui-même binaire. Tout individu muté de la population 1 devient un individu de la population 0, car sa mutation le rend identique aux individus de la population 0 et inversement.

Ainsi, on multiplie μ par le nombre de nouveaux nés par génération $n_0 * K_0$ (cf $n_1 * K_1$) et on obtient le nombre d'individus de n_0 (cf n_1) mutés en n_1 (cf n_0) par génération. La situation est résumée sur ce schéma :



Les flèches entrantes représentent la naissance ou l'apparition de nouveaux individus et les flèches sortantes représentent le décès ou la disparition d'individus au sein d'une population. Ainsi, en s'appuyant sur le schéma nous en déduisons les équations suivante :

$$\begin{aligned} \frac{dn_0}{dt} &= (K_0 - d_0)n_0 + K_1 * \mu * n_1 - K_0 * \mu * n_0 \\ \iff \frac{dn_0}{dt} &= (k_0 - 1)n_0 + \mu * (k_1 * n_1 - k_0 * n_0) \end{aligned} \quad (14)$$

$$\begin{aligned} \frac{dn_1}{dt} &= (K_1 - d_1)n_1 + K_0 * \mu * n_0 - K_1 * \mu * n_1 \\ \iff \frac{dn_1}{dt} &= (k_1 - 1)n_1 + \mu * (k_0 * n_0 - k_1 * n_1) \end{aligned} \quad (15)$$

On se concentre dorénavant sur la fréquence $f = \frac{n_1}{N}$. On sait d'après l'équation (10) que :

$$\frac{df}{dt} = \frac{1}{N} \frac{dn_1}{dt} = Sf(1 - f)$$

dans le cas précédent. En s'appuyant sur la faible valeur de μ et donc sur la faible modification de la fréquence par les termes $\mu * (k_0 * n_0 - k_1 * n_1)$, on ajoute les termes trouvés en (15) :

$$\begin{aligned}\frac{df}{dt} &= Sf(1-f) + \mu * \frac{k_0 * n_0 - k_1 * n_1}{N} \\ &\quad + \mu * \frac{k_0 * (N-n_1) - k_1 * n_1}{N} \\ &\quad + \mu * \left(k_0 - \frac{(k_0+k_1)n_1}{N}\right) \\ &\quad + \mu * (k_0 - (k_0+k_1)f)\end{aligned}$$

La reproduction étant lente, on prendra $k_0 \approx k_1 \approx 1$

On obtient ainsi :

$$\frac{df}{dt} = Sf(1-f) + (1-2f)\mu$$

(16)

6.1.1 Sous-cas numéro 1 : facteur de mutation plus important

Dans ce premier cas, nous allons supposer que le facteur de mutation est beaucoup plus important que celui de la sélection naturelle, on aura donc :

$$S \ll \mu$$

On peut donc négliger le terme en S dans l'équation (16), ce qui donne :

$$\frac{df}{dt} = (1-2f)\mu \quad (17)$$

Cherchons alors une forme intégrée de f :

$$\begin{aligned}\frac{df}{dt} &= (1-2f)\mu \\ \frac{df}{(1-2f)} &= \mu dt \\ -\frac{1}{2} \ln(1-2f) &= \mu t + A \text{ (avec } A \text{ constante)} \\ 1-2f &= e^{-2\mu t-2A} \\ f &= \frac{1}{2} - \frac{1}{2}e^{-2\mu t-2A}\end{aligned}$$

On pose $A' = e^{-2A}$ et on a finalement :

$$f = \frac{1}{2}(1 - A'e^{-2\mu t})$$

On pose la condition initiale $f(0) = 0$ comme dit plus tôt :

$$\begin{aligned}\frac{1}{2}(1 - A') &= 0 \\ A' &= 1\end{aligned}$$

Donc :

$$f = \frac{1}{2}(1 - e^{-2\mu t}) \quad (18)$$

Conclusions et remarques :

Etat stationnaire : On voit dans l'équation (17) que l'état stationnaire est atteint pour $\frac{1}{2}$. Ce qui signifie que la moitié de la population est atteinte et donc que les mutations s'effectuent de la même manière les deux côtés : il y a plus de variations.

Début : On remarque dans l'équation (17) qu'au début, lorsque $f \approx 0$, on a : $\frac{df}{dt} \approx \mu$. On peut en déduire que f augmente linéairement à μ .

Etat final : Lorsque f s'approche de $\frac{1}{2}$, la dérivée s'approche de zero mais reste positive. On en déduit que f augmentera en tendant vers $\frac{1}{2}$.

On obtient donc la courbe suivante d'après l'équation (19), en choisissant plusieurs petites valeurs de μ :

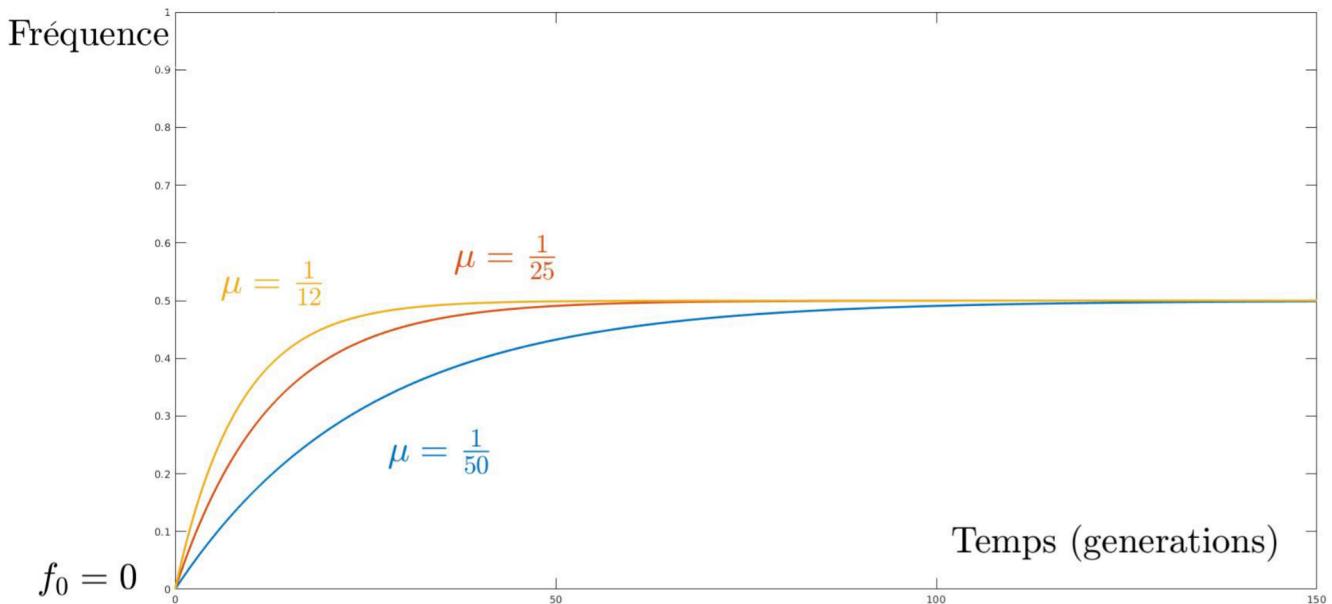


FIGURE 2 – Evolution de la fréquence de la population n_1 pour plusieurs valeurs de μ et une sélection négligeable

On observe bien ce qu'on avait analysé. Ce cas n'est toutefois pas très intéressant car il néglige les dynamiques de compétitions apportées par le facteur de sélection naturelle.

6.2 Sous cas numéro 2 : sélection plus importante

Dans ce deuxième cas, nous allons supposer que le facteur de sélection naturelle est beaucoup plus important que celui de la mutation , on aura donc :

$$S \gg \mu$$

Si on néglige alors le terme $(1 - 2 * f) * \mu$ dans l'équation (16) et on se retrouve dans le cas sans mutations étudié dans la partie précédente, avec :

$$f(t) = \frac{f_0 e^{S*t}}{1 - f_0 + f_0 * e^{(S*t)}}$$

Cependant, si μ n'est pas complètement nul, on observera un autre phénomène beaucoup plus intéressant.

6.2.1 Etude disjointe de l'évolution de la fréquence

Considérons à présent que le facteur de sélection naturelle est bien plus important que celui de la mutation, mais qu'on ne le néglige pas pour autant. Afin d'éviter de se confronter directement à la résolution de l'équation (16), nous procéderons par disjonction, en étudiant chacun des états de f séparément.

6.2.2 L'état d'équilibre

Déterminons et étudions d'abord les états stables de f à l'aide de l'équation (16) :

$$\begin{aligned} \frac{df}{dt} &= 0 \\ Sf(1-f) + (1-2f)\mu &= 0 \\ Sf(1-f) &= (2f-1)\mu \\ f(1-f) &= (2f-1)\frac{\mu}{S} \end{aligned}$$

On sait de plus que :

$$S \gg \mu \implies \frac{\mu}{S} \approx 0 \implies f(1-f) \approx 0$$

On en déduit qu'il y a état stationnaire autour de 1 et/ou de 0.

Cas f proche de 1 : Etudions d'abord la possibilité d'un état stationnaire à $f \approx 1$:

On a :

$$\begin{aligned} 1 * (1-f) &= (2-1) * \frac{\mu}{S} \\ f &= 1 - \frac{\mu}{S} \end{aligned}$$

Vérifions la cohérence de notre résultat. Si $f = 1 - \frac{\mu}{S}$ alors :

$$\begin{aligned} \frac{df}{dt} &= S\left(1 - \frac{\mu}{S}\right)\left(\frac{\mu}{S}\right) + (2\frac{\mu}{S} - 1)\mu \\ &= S\left(\frac{\mu}{S}\right)^2 < \left(\frac{\mu}{S}\right)^2 \end{aligned}$$

Ce qui est suffisamment proche de zero et positif.

Ainsi on trouve l'état stable autour de $f \approx 1$ qui est $f = 1 - \frac{\mu}{s}$.

Cas f proche de 0 : Etudions d'abord la possibilité d'un état stationnaire à $f \approx 0$:

$$\begin{aligned}\frac{df}{dt} &= Sf(1 - 0) + (1 - 2 * 0)\mu \\ \frac{df}{dt} &= Sf - \mu \\ f &= -\frac{\mu}{S}\end{aligned}$$

Ce qui est impossible car f est une proportion. Donc l'état stationnaire est impossible à $f \approx 0$.

Conclusion : Ainsi, nous pouvons en déduire par un raisonnement par l'absurde que le seul état stable de f est atteint en :

$$f = 1 - \frac{\mu}{s} \quad (19)$$

Comme la dérivée de f est positive autour de cette valeur, on peut en conclure que f tend en augmentant vers $1 - \frac{\mu}{s}$, ce qui nous donne l'état final.

6.2.3 Début

Essayons d'étudier la tendance de f juste après l'état initial. On rappelle l'expression de la dérivée de f :

$$\frac{df}{dt} = Sf(1 - f) + (1 - 2f)\mu$$

On remarque que à $f = 0$, état initial, on a le terme de gauche qui disparait et donc :

$$\frac{df}{dt} = \mu$$

Ce qui donne une tendance linéaire de coefficient directeur μ pour $f = 0$ et donc pour f très proche de zero car sa dérivée est continue. Essayons d'étudier jusqu'à quand cette tendance se vérifie, c'est-à-dire jusqu'à quand le terme de gauche est très petit par rapport au terme de droite. On va donc déterminer la valeur de f pour laquelle les deux termes sont égaux :

$$Sf(1 - f) = (1 - 2f)\mu$$

Comme on sait que cela ne se vérifie que lorsque f est très proche de zero, on peut directement écrire :

$$Sf = \mu \iff f = \frac{\mu}{S}$$

Ainsi, avant cette valeur, f gardera une courbe plutôt linéaire :

$$f(t) = \mu t$$

Cette valeur sera donc atteinte en $t \approx \frac{1}{S}$.

C'est donc un laps de temps très court. Après, le terme de gauche deviendra vite assez fort et la fonction se changera en exponentielle positive comme nous l'avons vu dans les cas précédents.

Conclusion : Ainsi, la tendance de f sera tout d'abord linéaire et proche de $f(t) = \mu t$ jusqu'aux environs de $t = \frac{1}{S}$. On est ensuite bientôt ramené au cas précédent, qui commence donc à $f \approx \frac{\mu}{S}$. On peut donc reprendre la formule (10) pour $t > \frac{1}{S}$:

$$f(t) = \frac{f_0 e^{S*t'}}{1 - f_0 + f_0 e^{S*t'}}$$

avec $f_0 \approx \frac{\mu}{S}$ et $t' = t - \frac{1}{S}$.

6.2.4 Demi temps

Voyons à quel moment les individus de n_1 ont atteint la moitié de la population. Lorsqu'on a $f \approx \frac{1}{2}$, on étudie alors les deux termes de l'équation (16) :

$$Sf(1-f) \approx \frac{S}{4} \text{ et } (1-2f)\mu \approx 0 * \mu$$

On en déduit que le premier terme est totalement prépondérant, d'autant plus que $S \gg \mu$. Nous pouvons donc négliger $(1-2f)\mu$ pour nous ramener à un cas déjà connu avec $\frac{df}{dt} = Sf(1-f)$. Celui-ci a été résolu et nous avions trouvé pour formule :

$$t_{1/2} = \ln \frac{1 - f_0}{f_0} * \frac{1}{S}$$

Nous prenons ici $f_0 = \frac{\mu}{S}$ comme dit précédemment et on a donc :

$$t_{1/2} = \ln \left(\frac{S}{\mu} - 1 \right) * \frac{1}{S}$$

Comme $S > \mu$, cela donne bien un temps positif. On peut remarquer qu'il réduit selon S car le terme à l'intérieur du logarithme est moins puissant. Il décroît aussi en fonction de μ . On peut en conclure que plus la sélection naturelle est importante et plus les mutations sont fréquentes, plus les individus de n_1 seront rapides à atteindre la moitié de la population, ce qui est très vraisemblable.

Avec des données biologiques pour certains virus, on prend $S = \frac{1}{3}$ et $\mu = \frac{1}{10^5}$, et on obtient un $t_{1/2}$ autour de 32. L'unité du temps est ici le temps qu'il faut à un virus pour mourir.

6.3 Conclusions et graphique

On a donc une courbe de f linéaire au tout début, les seuls individus de n_1 qui apparaissent sont ceux qui mutent de n_0 vers n_1 . Puis, très vite à partir d'un temps inverse au coefficient de sélection naturelle, les individus de n_1 vont se reproduire. On aura alors, et ce pendant la majorité de l'évolution, une courbe de la même tendance que dans le cas sans mutations. Celles-ci deviennent marginales devant la rapidité de la reproduction de la population n_1 . Puis, après avoir dépassé la moitié de la population totale, f va croître de moins en moins vite, jusqu'à tendre vers $1 - \frac{\mu}{S}$. La raison pour laquelle elle ne remplace pas totalement la population est qu'il reste toujours quelques mutations de n_1 vers n_0 .

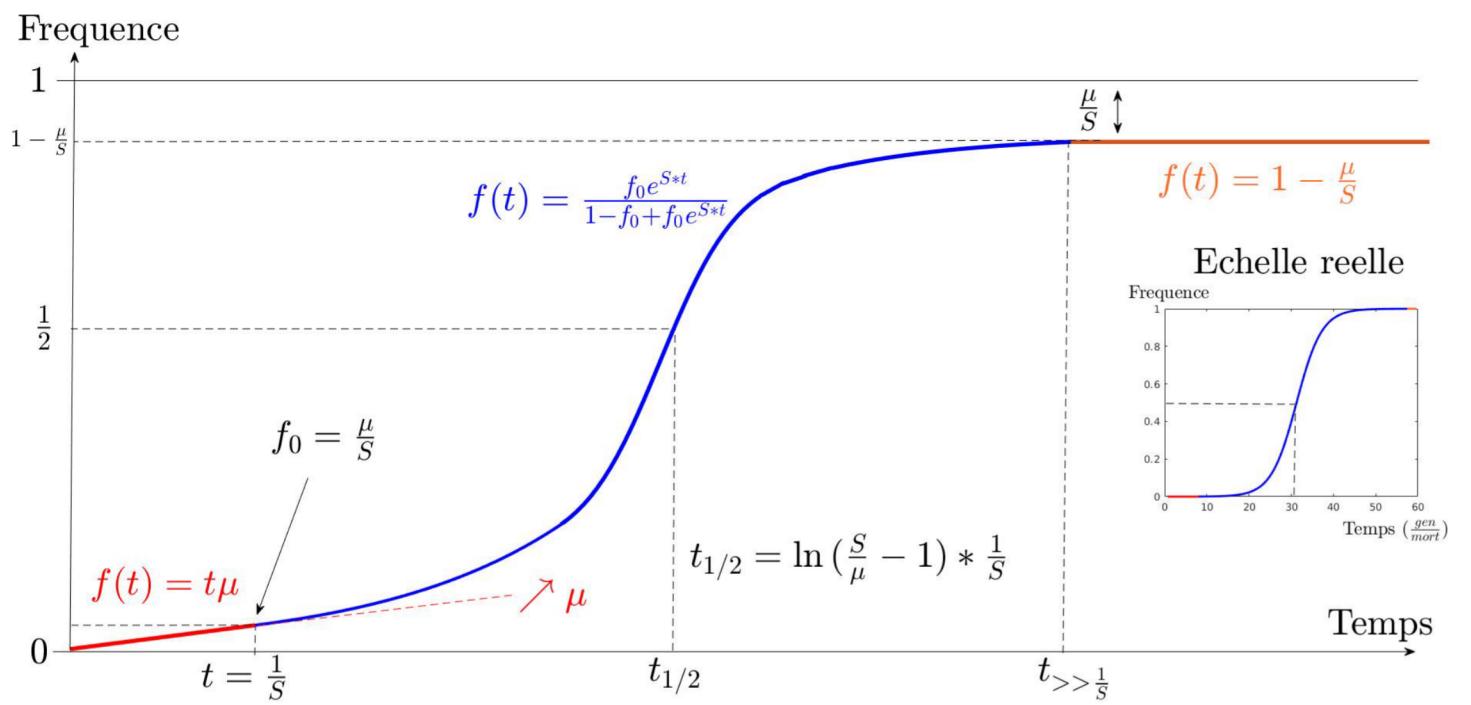


FIGURE 3 – Représentation schématique des différents régimes de la courbe

7 Cas avec deux mutations

7.1 Présentation

On étudie désormais une population qui peut subir une mutation de deux de ses gènes, ce qui donne au total 4 allèles et donc 4 génotypes et 4 populations différentes :

1. La population 00 des individus n'ayant subi aucune mutation
2. La population 11 des individus ayant subi les deux mutations
3. La population 10 des individus ayant subi la mutation d'un gène
4. La population 01 des individus ayant subi la mutation de l'autre gène

Nous supposerons d'une part que les chances de mourir de chaque population est la même :

$$d_{01} = d_{10} = d_{11} = d_{00} = 1$$

Et donc que la différentiation et la compétition se feront sur les chances de reproduction :

$$k_{ij} = k(1 + S_{ij})$$

avec k_{ij} la proportion des individus de ij qui se reproduiront par unité de temps et S_{ij} l'avantage de reproduction dont bénéficient la population ij , pour reprendre les notations précédentes. On supposera comme précédemment que $k \approx 1$

On supposera aussi, comme dans le cas précédent, que :

$$\mu \ll S$$

$$\mu^2 \ll \mu$$

$$S\mu \approx 0$$

$$\mu^2 \approx 0$$

Car ce sont les ordres de grandeurs observés par les biologistes.

7.2 Premières équations pour l'évolution des populations

Commençons par déterminer l'expression de l'évolution de la population n_{11} .

En l'espace de dt générations, la population 11 gagne quatre types d'individus :

-Les individus de n_{11} qui ne mutent d'aucun gène. Cette probabilité est égale à celle qu'un des gènes ne mutent pas au carré :

$$(1 - \mu)^2 = 1 - \mu^2 - 2\mu \approx 1 - 2\mu$$

-Les individus de n_{01} qui ont muté du premier gène. C'est la probabilité que le premier gène mute mais pas l'autre :

$$\mu * (1 - \mu) = \mu - \mu^2 \approx \mu$$

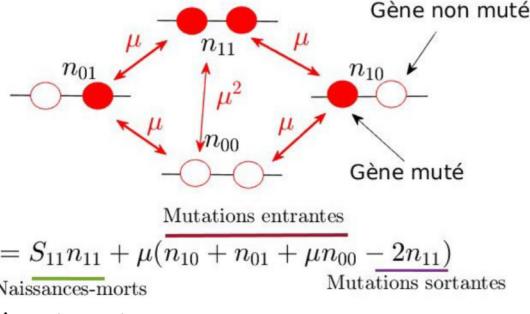
-Les individus de n_{10} qui ont muté du deuxième gène. Inversement :

$$(1 - \mu) * \mu = \mu - \mu^2 \approx \mu$$

-Les individus de n_{00} qui ont muté des deux gènes :

$$\mu * \mu = \mu^2$$

Probabilités de mutation



De plus, elle perd ceux qui sont morts :

$$n_{11} * d_{11} = n_{11}$$

On multiplie alors chaque probabilité de mutation par le nombre d'individus qui naissent par unité de temps dans la population correspondante, c'est à dire $n_{ij} * k_{ij}$, et en additionnant tous les gains et pertes de n_{11} par unité de temps on obtient alors sa différentielle :

$$\begin{aligned} dn_{11} = & + [(1 - 2\mu)k_{11}n_{11}]dt \text{ (Les non mutants)} \\ & + [n_{00}k_{00}\mu^2]dt \text{ (Les mutés de } n_{00}) \\ & + [n_{10}k_{10}\mu]dt \text{ (Les mutés de } n_{10}) \\ & + [n_{01}k_{01}\mu]dt \text{ (Les mutés de } n_{01}) \\ & - [n_{11}]dt \text{ (Les morts)} \end{aligned}$$

Et donc sa dérivée :

$$\frac{dn_{11}}{dt} = (1 - 2\mu)k_{11}n_{11} - n_{11} + [n_{00}k_{00}\mu^2 + n_{10}k_{10}\mu + n_{01}k_{01}\mu]$$

Précisons le calcul en remplaçant k_{ij} par $k(1 + S_{ij})$:

$$\begin{aligned} \frac{dn_{11}}{dt} &= (1 - 2\mu)(1 + S_{11})n_{11} - n_{11} + \mu[n_{00}k_{00}\mu + n_{10}k_{10} + n_{01}k_{01}] \\ &= (S_{11} - 2\mu - 2\mu S_{11})n_{11} + \mu[n_{00}(1 + S_{00})\mu + n_{10}(1 + S_{10}) + n_{01}(1 + S_{01})] \\ &= (S_{11} - 2\mu - 2\mu S_{11})n_{11} + n_{00}\mu^2 + n_{00}S_{00}\mu^2 + \mu n_{10} + \mu n_{10}S_{10} + \mu n_{01} + \mu n_{01}S_{01} \end{aligned}$$

En négligeant les bon termes, on obtient finalement :

$$\frac{dn_{11}}{dt} = s_{11}n_{11} + \mu(n_{10} + n_{01} - 2n_{11})$$

(20)

Au final, en suivant le même raisonnement et en changeant les indices pour les trois autres cas, on obtient les expressions suivantes :

$$\begin{aligned} \frac{dn_{10}}{dt} &= s_{10}n_{10} + \mu(n_{11} + n_{00} - 2n_{10}) \\ \frac{dn_{01}}{dt} &= s_{01}n_{01} + \mu(n_{00} + n_{11} - 2n_{01}) \\ \frac{dn_{00}}{dt} &= s_{00}n_{00} + \mu(n_{10} + n_{01} - 2n_{00}) \end{aligned}$$

7.3 Détermination des fréquences en fonction de S_{ij} à l'état stationnaire

7.3.1 Cas général

Considérons à présent une situation où nous avons atteint un état stable. Chaque population est donc en état stationnaire :

$$\frac{dn_{ij}}{dt} = 0 \quad (21)$$

Observons ce que cela implique pour la première population n_{11} . On a d'après l'équation (20) :

$$\begin{aligned} 0 &= S_{11}n_{11} + \mu(n_{10} + n_{01} - 2n_{11}) \\ (2\mu - S_{11})n_{11} &= \mu(n_{10} + n_{01}) \\ n_{11} &= \mu \frac{n_{10} + n_{01}}{2\mu - S_{11}} \end{aligned}$$

Donc :

$$f_{11} = \frac{\mu}{(2\mu - S_{11})}(f_{10} + f_{01}) \quad (22)$$

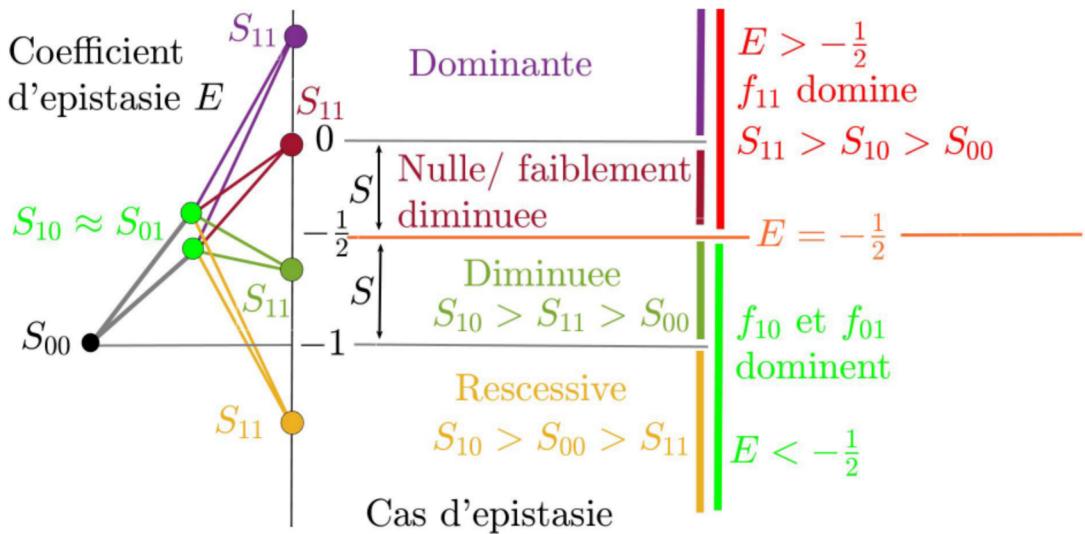
De la même façon, nous appliquons cette méthode aux autres populations et on obtient :

$$\begin{aligned} f_{10} &= \frac{\mu}{2\mu - S_{10}}(f_{00} + f_{11}) \\ f_{01} &= \frac{\mu}{2\mu - S_{01}}(f_{00} + f_{11}) \\ f_{00} &= \frac{\mu}{(2\mu - S_{00})}(f_{10} + f_{01}) \end{aligned}$$

7.3.2 Allèles mutés, combinaisons et épistases

Introduisons maintenant des conditions plus spécifiques sur les avantages que fournissent ou non les mutations de chaque allèles.

Dans notre cas, nous considérerons que chaque mutation d'un des deux allèles engendrent des avantages en terme de sélection naturelle. Ces mutations pourront alors se combiner et interagir pour présenter des caractéristiques particulières : c'est le phénomène **d'épistasie**. Si la combinaison est complémentaire, c'est-à-dire si elle représente un avantage bien plus grand que les deux mutations isolées, on parle alors **d'épistasie dominante / magnitude epistasis**. Au contraire, les deux mutations peuvent se gêner. Si la combinaison engendrée est moins bien que celle d'une mutation simple mais toujours meilleure que celle sans mutations on parle **d'épistasie diminuée / sign epistasis**, et si elle est pire **d'épistasie récessive / recessive epistasis**.



Nous considérerons dans la suite que chaque mutation est porteuse d'un avantage équivalent :

$$S_{10} = S_{01}$$

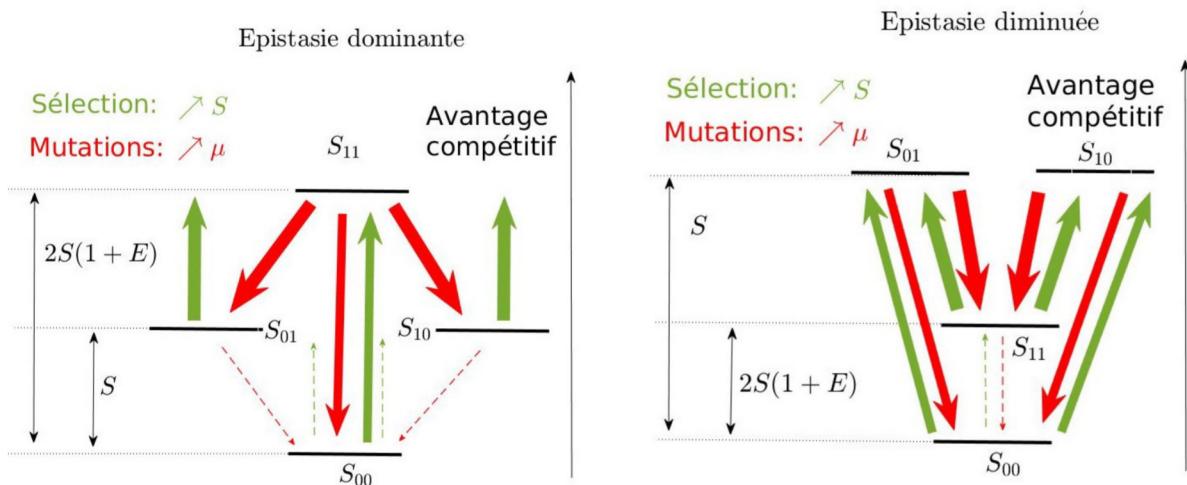
Introduisons une nouvelle variable S qui sera la différence de compétitivité entre la population n_{00} et n_{10} :

$$S = S_{10} - S_{00}$$

Dans notre modèle, cette différence est indépendante des contextes et elle sera donc constante.

On a le coefficient d'epistasis E tel que :

$$\begin{aligned} S_{11} - S_{00} &= (S_{01} - S_{00} + S_{10} - S_{00})(1 + E) \\ &= (2S_{01} - 2S_{00})(1 + E) \\ &= \boxed{2S(1 + E)} \end{aligned}$$



7.3.3 Les individus de n_{11} sont les plus avantagés

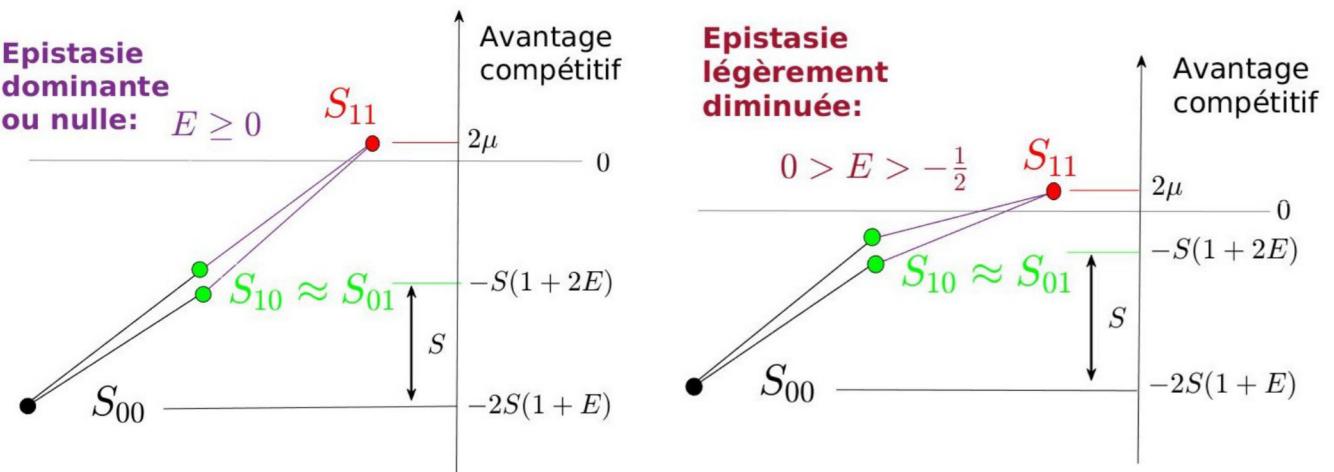
Considérons premièrement que la population n_{11} reste avantagée par rapport aux autres, c'est-à-dire lorsque $E > -0,5$ (voir schémas ci-dessous). Les équations seront valables pour une

épistasie dominante, nulle ou légèrement diminuée. Les deux mutations sont porteuses d'avantages compétitifs, on a donc la population n_{11} la plus avantageée et n_{00} la défavorisée :

$$S_{11} > S_{10}, S_{01} > S_{00}$$

Nous avons vu dans lors de nos études précédentes que si une population possédait un léger avantage de ce type, alors elle tendrait à remplacer totalement les autres sur le long terme. Nous en déduisons des suppositions d'ordres de grandeur pour les populations aux états stationnaires :

$$\begin{aligned} n_{11} &>> n_{10}, n_{01} >> n_{00} \\ \iff f_{11} &>> f_{10}, f_{01} >> f_{00} \end{aligned}$$



Explication du schéma : Les cas d'épistasie que nous étudions premièrement sont ceux dans lesquels n_{11} est dominant et donc dispose d'un avantage compétitif S_{11} supérieur aux autres.

- Ordonnée : L'avantage compétitif dont dispose une population d'individus.
- Abscisse : Les différentes populations existantes.
- Valeurs en ordonnées : On voit dans le paragraphe suivant que $S_{11} \approx 2\mu$, ce qui est très proche de zéro par rapport à S . On en déduit ainsi le positionnement de zéro dans le schéma, et on peut conclure sur les valeurs de S_{00} et de S_{10} :

$$\begin{aligned} S_{11} - S_{00} &= 2S(1+E) \iff S_{00} = -2S(1+E) \\ S_{10} = S_{00} + S &= -2S(1+E) + S \iff S_{10} = -S(1+2E) \end{aligned}$$

Un résultat intéressant : Les avantages de compétitivité pour S_{01} et pour S_{00} ne dépendent donc que de leur différence avec la population dominante s_{11} , parce que la compétition ne se joue quasiment qu'avec elle.

Self-consistance : Observons si nos suppositions sont cohérentes et quels résultats plus précis nous pouvons obtenir avec. On reprend successivement les équations trouvées précédemment pour chacune des populations.

Pour la population dominante n_{11} :

$$f_{11} = \frac{\mu}{(2\mu - S_{11})} (f_{10} + f_{01})$$

On voit que si la fréquence f_{11} est très supérieure aux autres, alors on doit avoir :

$$\begin{aligned} f_{11} &>> (f_{10} + f_{01}) > 0 \\ \iff \frac{\mu}{(2\mu - S_{11})}(f_{10} + f_{01}) &>> (f_{10} + f_{01}) > 0 \\ \iff \frac{\mu}{(2\mu - S_{11})} &>> 1 > 0 \\ \iff \mu &>> 2\mu - S_{11} > 0 \end{aligned}$$

Ce qui ne peut être vrai que lorsque $S_{11} = 2\mu - \epsilon$ avec $0 < \epsilon \ll \mu$, ce qui signifie que S_{11} est très proche de 2μ . On a donc un coefficient S_{11} positif et faible dans le cas où la population n_{11} dominante.

$$S_{11} \approx 2\mu$$

C'est très cohérent puisqu'il représente un avantage par rapport aux autres, qui ont un coefficient négatif. Et il est proche de 2μ puisqu'à l'état stationnaire cet avantage de reproduction sert juste à compenser la perte des individus qui mutent vers les autres populations (voir l'équation (20)). On peut donc prendre $f_{11} \approx 1$ pour le reste de cette analyse.

Pour les populations intermédiaires n_{10} et n_{01} :

$$\begin{aligned} f_{01} &= \frac{\mu}{2\mu - S_{01}}(f_{00} + f_{11}) \ll 1 \\ 0 &< \frac{\mu}{2\mu - S_{01}} \ll 1 \\ 0 &< \mu \ll 2\mu - S_{01} \end{aligned}$$

avec μ positif. On en déduit donc que $|S_{01}| >> \mu$ et que $S_{01} < 0$ ce qui va dans le sens de ce qui était énoncé précédemment. On peut donc négliger le terme 2μ dans le dénominateur de la formule et on obtient :

$$f_{01} = \frac{\mu}{-S_{01}} \iff f_{01} = \frac{\mu}{S(1+2E)} \quad (23)$$

et, par symétrie :

$$f_{10} = \frac{\mu}{-S_{10}} \iff f_{10} = \frac{\mu}{S(1+2E)}$$

Pour la population la plus faible n_{00} :

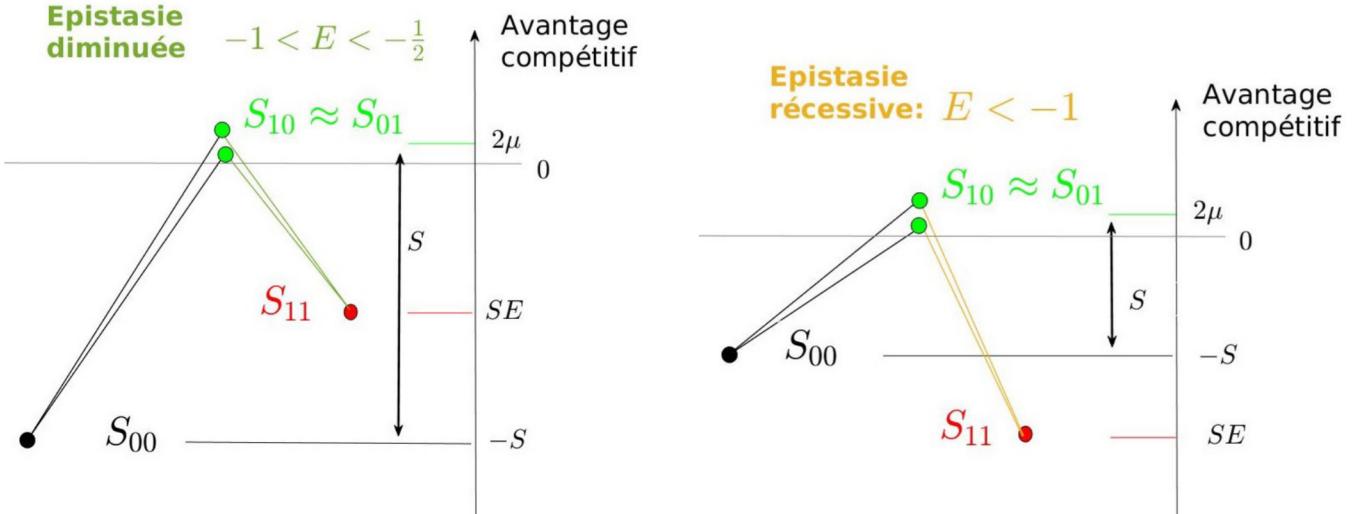
$$\begin{aligned} f_{00} &= \frac{\mu}{(2\mu - S_{00})}(f_{10} + f_{01}) \\ &= \frac{\mu}{-S_{00}} \left(\frac{\mu}{-S_{10}} + \frac{\mu}{-S_{10}} \right) \\ &= \frac{\mu^2}{S_{00}} \left(\frac{1}{S_{10}} + \frac{1}{S_{10}} \right) \\ f_{00} &= \frac{\mu}{2S(1+E)} \left(\frac{2\mu}{S(1+2E)} \right) \\ &= \frac{\mu^2}{S^2} \frac{1}{(1+2E)(1+E)} \end{aligned}$$

On voit donc que pour les populations dominées, la fréquence à l'état stationnaire diminue en fonction de leurs désavantages et augmente en fonction de la fréquence des mutations, qui viennent surtout de la population majoritaire, ici n_{11} . On a, de plus, $f_{00} \ll f_{10}, f_{01}$ car $\mu^2 \ll \mu$.

Conclusion : Nos suppositions sont bien cohérente avec nos résultats.

7.3.4 Les populations n_{01} et n_{10} sont les plus avantageées

Ceci comprend les cas où $E < -\frac{1}{2}$, donc une épistasie très diminuée ou récessive.



Explication du schéma : Les cas d'épistasie que nous étudions deuxièmement sont ceux dans lesquels n_{10} et n_{01} sont dominants et donc disposent d'un avantage compétitif S_{10} et S_{01} supérieur aux autres.

- Pour les valeurs : On voit dans le paragraphe suivant que $S_{01} = S_{10} \approx 2\mu$, ce qui est très proche de zero par rapport à S . On en déduit ainsi le positionnement de zero dans le schéma, et on peut conclure sur les valeurs de S_{00} et de S_{11} :

$$\begin{aligned} S_{10} - S_{00} &= S \iff S_{00} = -S \\ S_{11} = S_{00} + 2S(1+E) &= -S + S(1+E) + S \iff S_{11} = SE \end{aligned}$$

Comme précédemment, les avantages de compétitivité pour S_{11} et pour S_{00} ne dépendent que de leur différence avec les populations dominantes S_{10} et S_{01} , parce que la compétition ne se joue quasiment qu'avec elles.

Pour les populations dominantes n_{01} et n_{10} :

On peut là aussi déduire de notre expérience que les populations n_{01} et n_{10} domineront largement les autres. On a donc :

$$f_{10} + f_{01} \approx 1$$

Comme on suppose que $S_{10} = S_{01}$, on en déduit par symétrie que :

$$f_{01} = f_{10} \approx \frac{1}{2}$$

En reprenant la formule trouvée précédemment dans le cas général :

$$f_{10} = \frac{\mu}{(2\mu - S_{10})}(f_{00} + f_{11}) \approx \frac{1}{2}$$

En observant que $f_{11} + f_{00}$ est très petit, et en raisonnant de la même manière que dans le cas précédent on en déduit que $S_{10} \approx 2\mu$ et $S_{10} > 2\mu$. D'où le placement du zero dans le schéma ci-dessus

Pour n_{00} :

$$\begin{aligned} f_{00} &= \frac{\mu}{(2\mu - S_{00})}(f_{10} + f_{01}) \\ &\approx \frac{\mu}{(2\mu - S_{00})} \end{aligned}$$

Et comme on sait (voir schéma) que $S_{00} \gg \mu$, on peut négliger μ dans l'équation. Et en remplaçant aussi S_{00} par $-S$ comme trouvé plus haut, on obtient :

$$f_{00} = \frac{\mu}{S}$$

Pour n_{11} : De même, on a :

$$\begin{aligned} f_{11} &= \frac{\mu}{(2\mu - S_{11})}(f_{10} + f_{01}) \\ &\approx \frac{\mu}{(2\mu - S_{00})} \\ &\approx \frac{\mu}{-(1+2E)S} \end{aligned}$$

On en déduit alors des valeurs plus précises pour les populations dominantes :

$$f_{10} + f_{01} = 1 - \mu \left[\frac{1}{S} - \frac{1}{S(1+E)} \right]$$

7.3.5 Conclusion sur les populations à l'état stable

On a à présent tous les éléments pour déduire la variation des fréquences des populations en fonction du coefficient d'épiastie E .

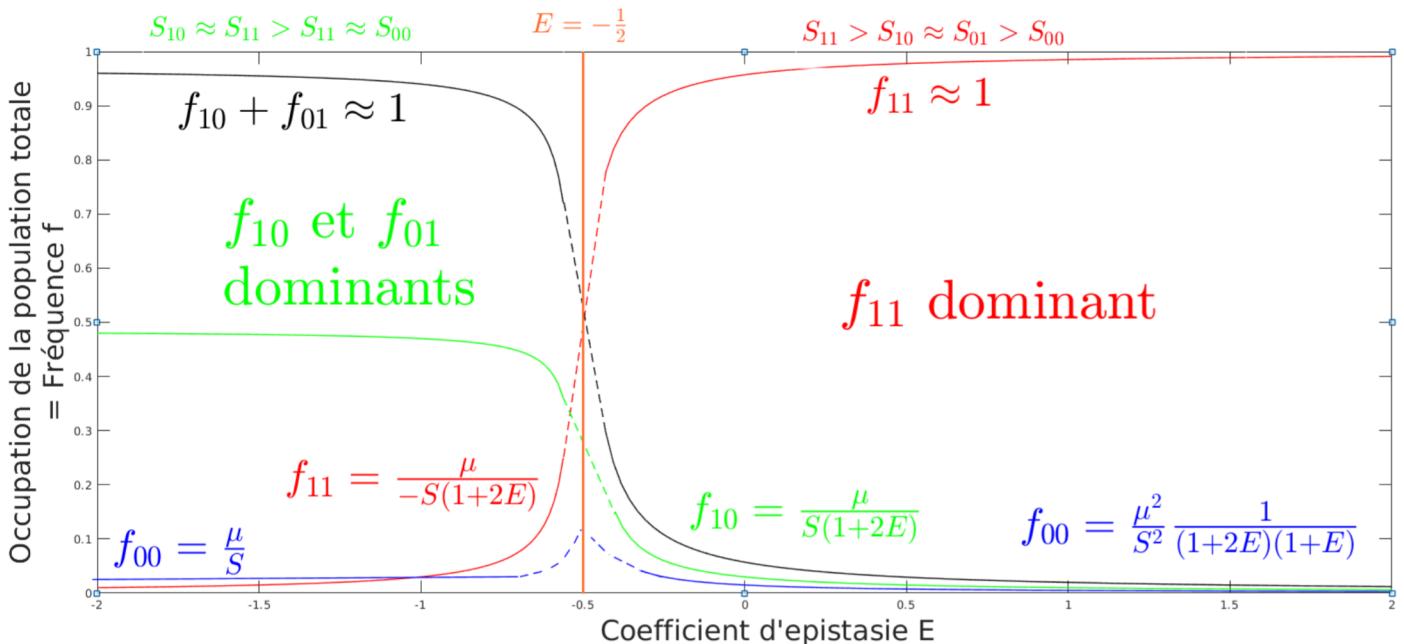


FIGURE 4 – Fréquences à l'état stationnaire des populations en fonction du coefficient d'épiastie

Explication du graphique :

-Abscisse : Coefficient d'épistasie E tel que $S_{11} - S_{00} = 2S(1 + E)$, en orange la valeur remarquable $E = -\frac{1}{2}$. Du côté droit les formules des hypothèses $f_{11} >> f_{10}, f_{01} >> f_{00}$ sont appliquées, tandis que du côté gauche c'est $f_{10}, f_{01} >> f_{11} >> f_{00}$

-Ordonnée : Pourcentage de la population totale.

-Valeurs particulières : Nous avons choisi S et μ arbitrairement (bien que satisfaisant les hypothèses $S >> \mu > 0$) pour que le graphique soit lisible.

- **Côté droit :** $E > -\frac{1}{2}$

De ce côté, la population f_{11} , qui possède un avantage de compétitivité, domine complètement les autres, et cet écart s'accroît avec le coefficient d'épistasie. Pour les deux autres c'est l'inverse. On peut remarquer par ailleurs que f_{00} décroît avec le carré de E , ce qui signifie qu'il est encore plus sensible à sa variation. En effet, voir schéma.

- **Autour de $E = \frac{1}{2}$** : Les différences de compétitivités entre les populations n_{11} , n_{10} et n_{01} étant très faibles, il ne peut pas y avoir de domination écrasante d'une population. C'est pourquoi les hypothèses de notre étude ne se vérifient pas et on trouve des valeurs limites pour chacune des formules. Pour avoir une continuité sur notre graphique, nous avons tracé en pointillé les jointures entre les limites des différentes formules. Les valeurs observées sont d'ailleurs assez compatibles, les populations dominantes étant du même ordre de grandeur dans cet intervalle.

- **Côté gauche :** $E < \frac{1}{2}$

Ici, ce sont les populations n_{10} et n_{01} qui dominent les autres, car elles disposent du meilleur avantage de reproduction. On voit bien que f_{11} se réduit avec le coefficient d'épistasie E , car il devient de plus en plus faible par rapport aux populations dominantes. On remarque que f_{00} , lui, reste constant. En effet, son écart de compétitivité avec les populations dominantes est fixé à S et ne varie pas.

8 L'auto-stop génétique

8.1 Présentation

Nous allons dans cette partie introduire un autre type de facteur qui influe sur la fréquence des populations : l'auto-stop génétique (hitchhiking en anglais).

On prend toujours deux gènes qui peuvent chacun subir une mutation. Dans notre cas, ces deux gènes sont très proches l'un de l'autre. Cependant, on ne considère plus l'état stable mais on s'intéresse plutôt à l'évolution dynamique des populations.

L'auto-stop génétique est le fait que la fréquence d'un allèle, qui ne représente pas forcément un avantage pour un individu, peut être modifiée par un autre allèle très proche de ce premier, qui lui apporte une plus grande compétitivité.

8.2 Exemple

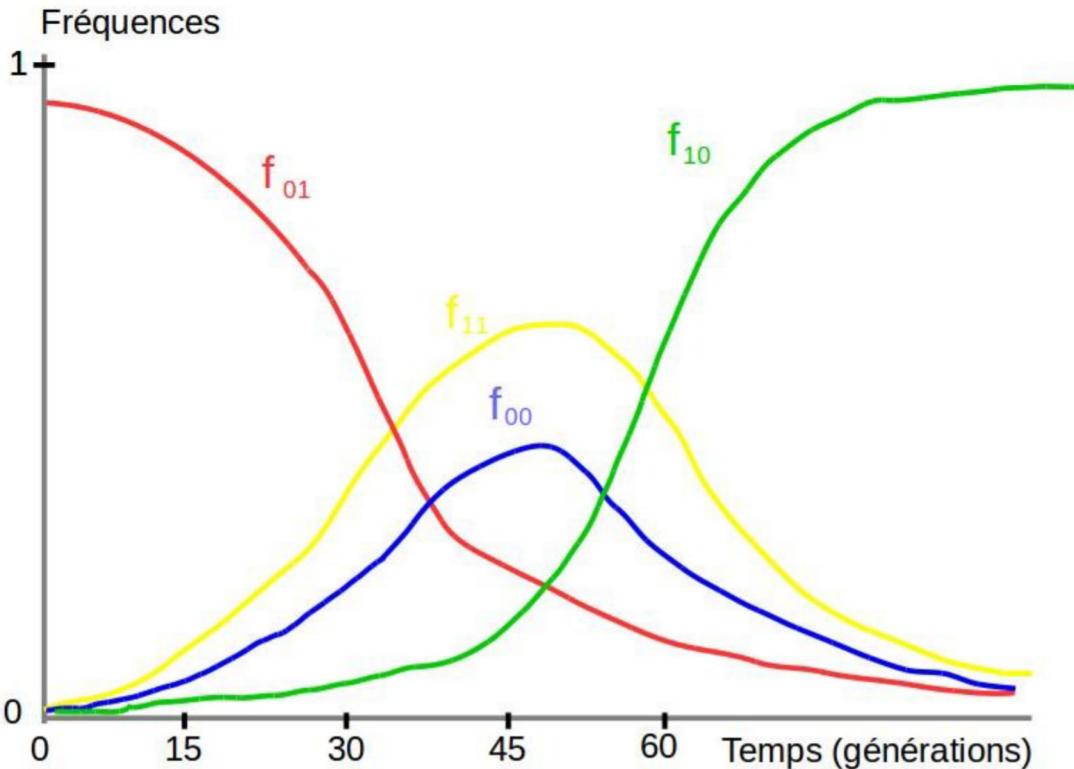
Prenons l'exemple suivant. Une mutation, celle du premier gène, présente un avantage, et on aura donc $S_{10} > S_{00}$. En revanche, la mutation du deuxième gène est un inconvénient : $S_{01} < S_{00}$. Cependant, l'avantage apporté par le premier étant plus important que l'inconvénient imposé par le deuxième, la double mutation représente une force : $S_{11} > S_{00}$. On a donc finalement : $S_{10} > S_{11} > S_{00} > S_{01}$. Les deux gènes sont très proches l'un de l'autre. Etudions

la fréquence d'apparition des mutations des deux gènes.

Conditions initiales : La population n_{01} est dominante au début : $n_{01} >> n_{10}, n_{01}, n_{00}$. On peut imaginer par exemple que le second gène représentait précédemment un avantage significatif mais qu'un brusque changement dans l'environnement des populations fait basculer la valeur des coefficient de compétitivité vers celles que nous avons définies ci-dessus. Nous nous plaçons au moment de ce basculement.

Nous allons tracer intuitivement la courbe représentatives des fréquences f_{00}, f_{01}, f_{10} et f_{11} .

- f_{01} décroît toujours comme sa valeur sélective est la plus basse des quatre.
- f_{10} met assez longtemps à acquérir une quantité suffisante d'individus pour se reproduire car elle a peu de chance de recevoir des mutants de la part de la population f_{01} . Toutefois, lorsqu'elle en aura assez, elle va se reproduire plus vite que toutes les autres et rapidement être totalement dominante.
- f_{00} et f_{11} vont recevoir les mutants de f_{01} assez vite car il n'y a besoin que d'un seul gène muté. Comme ils sont plus avantageux que f_{01} elles auront tendance à croître assez vite dès qu'elles seront assez nombreux pour se reproduire. Cependant, lorsque les individus f_{10} vont commencer à apparaître, ces deux populations vont subir leur concurrence et commenceront à décroître pour être finalement dominées.



8.3 Conclusion :

L'auto-stop génétique est donc un autre facteur d'évolution tout aussi intéressant que les autres, qui permet de donner plus de complexité dans l'évolution des populations en engendrant des phénomènes non monotones. La rareté d'un gène peut donc être rattrapé par ce phénomène d'auto-stop génétique comme nous avons pu le voir avec notre cas de figure.

9 Conclusion générale :

Nous avons donc réussi à modéliser de façon mathématique l'évolution de populations en prenant en nous appuyant sur des hypothèses proches de ce que l'on peut voir dans la théorie Darwinienne. A travers notre étude nous avons pu montrer les différents types d'évolution et de mutations qui peuvent exister pour des brassages génétiques simples. Bien qu'aucune de nos analyses n'aient concerné des cas concrèts, nous avons pu découvrir et décortiquer les principaux mécanismes des dynamiques de la sélection naturelle. La principale compétence que nous avons du développer était l'interprétation mathématique de nos modèles et de nos résultats. Pour chaque variable, inégalité, limite, nous devions en deviner sa signification et évaluer sa pertinence. Ainsi, cette étude nous a laissé entrevoir la complexité de l'évolution des populations dans la réalité et la façon de procéder pour les comprendre.