

Cet article est rédigé par des élèves. Il peut comporter des oublis ou des imperfections, autant que possible signalés par nos relecteurs dans les notes d'édition.

Maladie génétique et héréditaire

Atelier Math en Jeans

FAURE Olivier, VILLANI Neven, BERGER Clémentine, FERNEX Romain
(2nde, Lycée Blaise Pascal, Orsay)

Année 2015-2016

Enseignants : Hélène Cochard, Emmanuel Roblet, Dennis Julliot, Didier Missenard, Nicolas Segarra.

Chercheurs : Nicolas Burq, Tiago Jardim da Fonseca (Université Paris-Sud).

Table des matières

1	<u>Présentation du sujet</u>	2
2	<u>Introduction et conjectures</u>	2
2.1	<u>Définir les caractéristiques d'une population</u>	2
2.2	<u>Définir la maladie</u>	3
2.3	<u>Descendance</u>	4
2.4	<u>Propagation de la maladie dans le temps</u>	6
2.5	<u>Scénario apocalyptique</u>	9
3	<u>Conclusion</u>	11
4	Notes d'édition	12

1 Présentation du sujet

Nous avons axé notre sujet sur l'étude de la propagation des maladies génétiques dans une population quelconque et nous avons choisi d'ancrer notre sujet dans la réalité en prenant une maladie génétique réelle : la mucoviscidose.

2 Introduction et conjectures

La mucoviscidose est une maladie génétique qui touche en moyenne 1 enfant sur 3500. Celui-ci naît avec une déficience pulmonaire qui entraîne une mort prématurée de l'individu.

Nous nous sommes donc demandés comment la mucoviscidose peut-elle être si fréquente étant donné qu'elle entraîne une mort précoce et comment va évoluer la maladie dans les prochaines années.

Afin de répondre à ces questions, nous utiliserons le domaine de la génétique des populations qui est un domaine entre les mathématiques (probabilités) et la génétique.

On conjecture que si la maladie est si fréquente dans la population, c'est que le nombre de personnes capables de transmettre la maladie est important. On admettra que les individus malades sont stériles. Ils ne peuvent pas transmettre leurs gènes. Le brassage génétique donnera donc un désavantage aux individus malades. On conjecture donc que la maladie risque de toucher de moins en moins de personnes au fil des générations.

2.1 Définir les caractéristiques d'une population

Les populations que nous utiliserons auront les caractéristiques suivantes :

- individus diploïdes (les individus possèdent deux exemplaires de chaque gène)
- individus hermaphrodites
- chaque individu descend de parents tirés aléatoirement dans la génération précédente
- les générations sont non-chevauchantes, c'est-à-dire qu'elles sont distinctes dans le temps et qu'il n'y a aucune interaction entre elles.

Ce modèle implique de nombreux écarts à la réalité mais il doit être suffisamment simple pour être traité analytiquement ou numériquement. Il a cependant déjà été proposé par deux scientifiques dans les années 1930 : Sewall Wright (biologiste) et Ronald Fisher (généticien et statisticien).



Sewall Wright



Ronald Fisher

2.2 Définir la maladie

La maladie en question est définie par les critères suivants [1] :

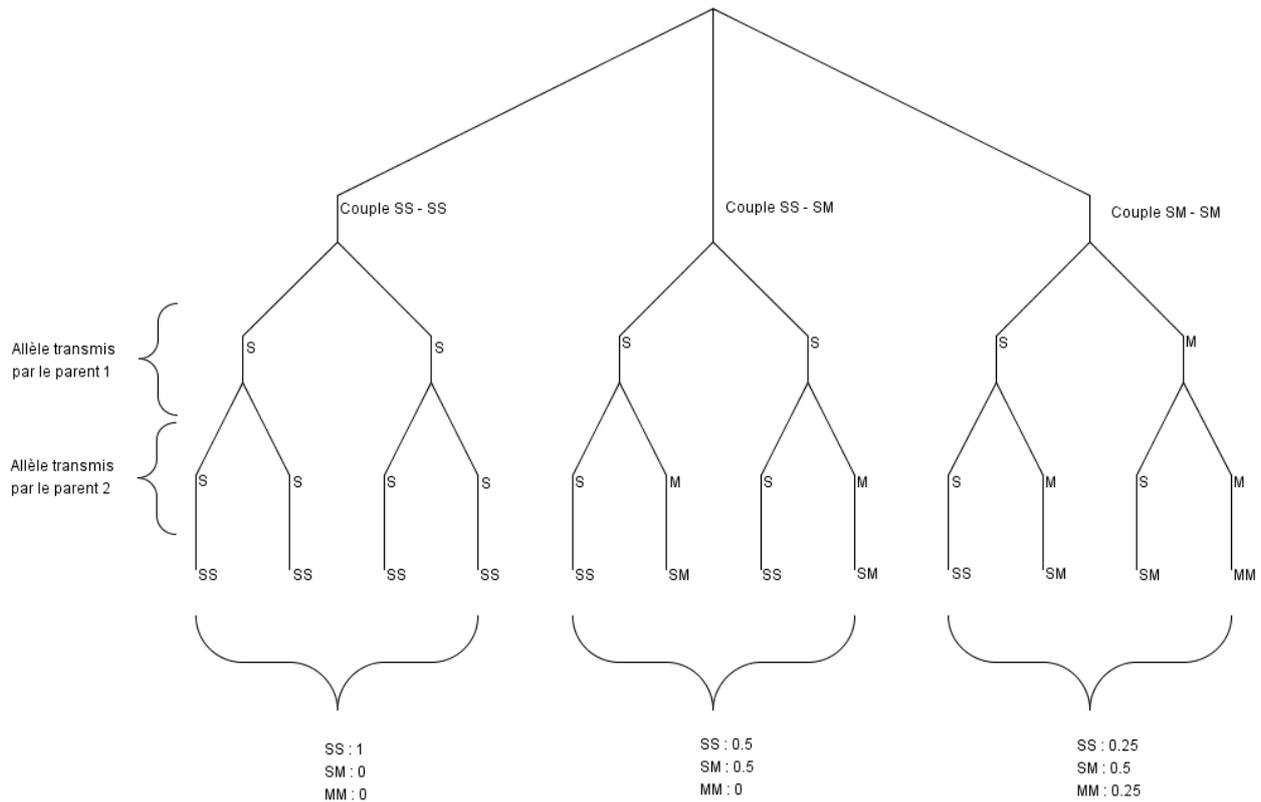
- monogénique, c'est-à-dire qu'il n'y a qu'un gène et deux versions de celui-ci définies par deux allèles que nous noterons S et M. S étant l'allèle « sain » et M étant l'allèle « malade », une mutation du gène.
- récessive, c'est-à-dire qu'un individu est malade uniquement s'il possède deux allèles M. En effet on observe trois cas possible :

allèles de l'individu	conséquence
MM	individu malade, stérile
SS	individu sain
SM	individu porteur sain

On a ici considéré qu'un individu malade est stérile car la mucoviscidose entraîne une mort en bas âge, souvent avant que l'individu ait eu le temps de se reproduire.

2.3 Descendance

Nous allons désormais nous intéresser à la descendance des différents couples, on la représente selon un arbre de probabilités.



Arbre représentant la généalogie des individus au niveau allélique

Voici ces mêmes résultats sous forme de tableau.

Parent 1	Parent 2	Couple fertile?	SS	SM	MM
SS	SS	OUI	1	∅	∅
SS	SM	OUI	0,5	0,5	∅
SS	MM	NON	∅	∅	∅
SM	SM	OUI	0,25	0,5	0,25
SM	MM	NON	∅	∅	∅
MM	MM	NON	∅	∅	∅

On remarque ainsi que le seul moyen d'avoir un individu malade est que ses deux parents soient porteurs sains.

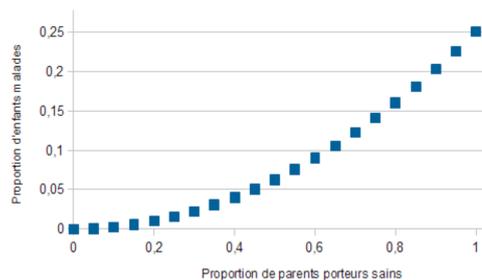
Nous nous intéressons à présent aux lois qui lient les générations les unes aux autres : le nombre d'enfants malades d'une génération n ne dépend que du nombre de parents porteurs sains de la génération précédente.

On en déduit donc la formule suivante [2] :

$$(\text{fréquence MM à la génération } n) = (\text{fréquence de MS à la génération } n - 1)^2 * \frac{25}{100}$$

Comme un individu MM ne peut qu'être le résultat de l'union d'un couple de deux porteurs sains, la probabilité d'avoir un individu malade revient à calculer la probabilité d'avoir un couple de porteurs sains à la génération précédente ici (*fréquence de MS à la génération $N-1$*)² ainsi que la probabilité que ce couple ait une descendance malade (ici le 0,25)

Afin de mettre cette relation en évidence, on a choisi de tracer une courbe représentative de cette formule en fonction des proportions des porteurs sains.



Graphique représentant la proportion d'individus malades de la génération n en fonction de la proportion de porteurs sains de la génération $n-1$.

À partir de la loi établie précédemment, on va maintenant calculer la proportion théorique de porteurs sains français capables de transmettre la mucoviscidose.

En notant SM la fréquence de MS à la génération $n - 1$, on a :

$$\begin{aligned} \frac{1}{3500} &= (SM)^2 * \frac{25}{100} \\ \Leftrightarrow (SM)^2 &= \frac{1}{3500} * \frac{100}{25} \\ \Leftrightarrow SM &= \sqrt{\frac{1}{3500} * \frac{100}{25}} \approx 3,4 \cdot 10^{-2} \end{aligned}$$

Soit environ une personne sur trente. [3]

2.4 Propagation de la maladie dans le temps

Dans le but d'étudier la propagation de la maladie dans le temps nous avons choisi d'utiliser des simulations sur des populations de tailles différentes, nous permettant ainsi de collecter des données et d'en extraire des généralités.

1. Petite échelle

Nous avons dans un premier temps réalisé des simulations de la propagation de la maladie dans des petites populations : 25 individus.

Voici un extrait du programme écrit en *Python* qui nous a permis de faire ces simulations : [4]

```
1 from random import randint
2
3 #Initialisation
4 taillePop, freqM = 25, 0.5
5 G, SS, MM, MS = 0, 0, 0, 0
6 popP = popE = [{"_"}*2 for i in range(taillePop)]
7 for i in range(taillePop):
8     if randint(0,10)<=freqM*10:
9         popP[i][0]="S"
10        popP[i][1]="M"
11    else:
12        popP[i][0]="S"
13        popP[i][1]="S"
14
15 #Trouver un parent fertile
16 def parent(tab):
17     found=False
18     while not(found):
19         ind=randint(0,taillePop-1)
20         if tab[ind][0]=="S" or tab[ind][1]=="S":
21             found=True
22             return(ind)
23
24 #Comptage
25 def count(tab):
26     mm, ms, ss = 0, 0, 0
27     for k in range(taillePop):
28         if tab[k][0]=="S" and tab[k][1]=="S":
29             ss+=1
30         elif tab[k][0]=="M" and tab[k][1]=="M":
```

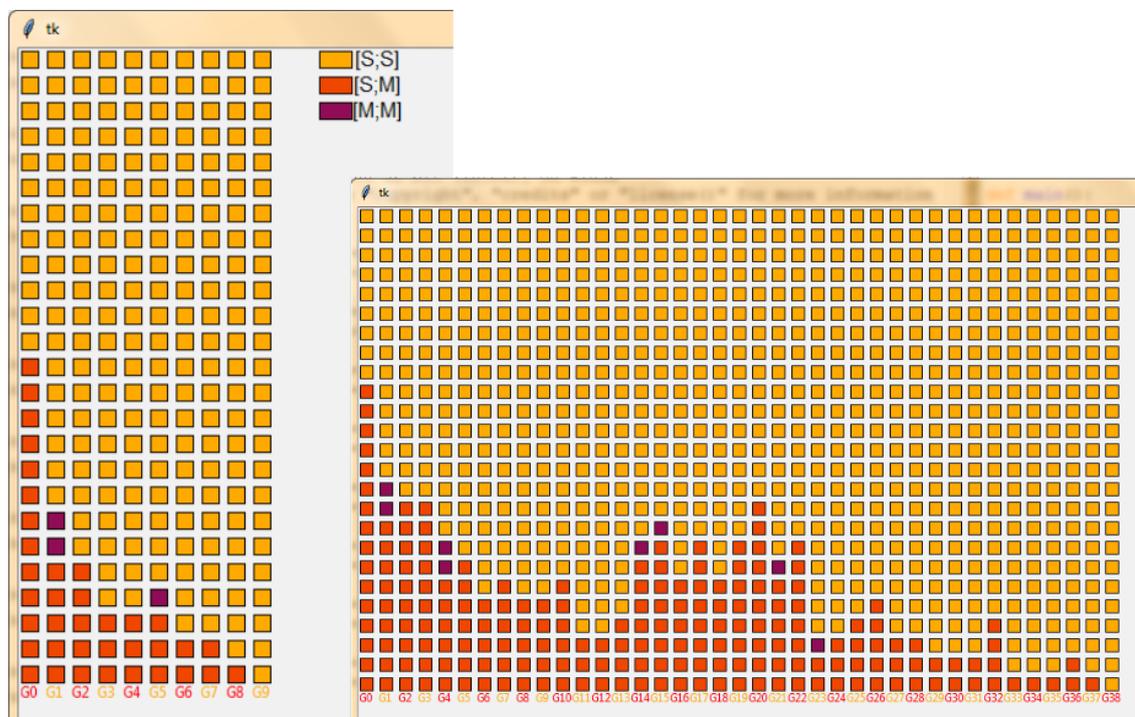
```

31         mm+=1
32     else :
33         ms+=1
34     return (ss ,mm,ms)
35
36
37 SS, MM, MS=count (popP)
38 print (G, " " ,SS, " " ,MM " " ,MS)
39
40 #Reproduction
41 while (["S" ,"M"] in popP or ["M" ,"S"] in popP) and G<40:
42     G+=1
43     for j in range (taillePop): # determine la population des enfants
44         p1=parent (popP)
45         p2=parent (popP)
46         popE[j][0]=popP[p1][randint (0 ,1)]
47         popE[j][1]=popP[p2][randint (0 ,1)]
48
49     for k in range (taillePop): # les enfants deviennent les parents
50         for l in range (2):
51             popP[k][l]=popE[k][l]
52             popE[k][l]=" "
53
54     SS,MM,MS=count (popP)
55     print (G, " " ,SS, " " ,MM " " ,MS)

```

Ce programme crée une population de 25 individus qui à la génération 0 ont chacun une chance sur deux d'être porteur sain (les autres sont sains) les transitions d'une génération à la suivante (les enfants deviennent alors parents).

Voici les résultats que nous avons obtenus :



Chaque colonne correspond à une génération et chaque carré à un individu : jaune s'il est sain, rouge s'il est porteur sain, et mauve s'il est malade.

On observe que les résultats sont très différents mais qu'ils aboutissent tous à l'élimination de l'allèle M. En effet, nous avons fait beaucoup d'autres simulations et avons à chaque fois observé l'élimination de l'allèle M au bout d'un certain nombre de générations.

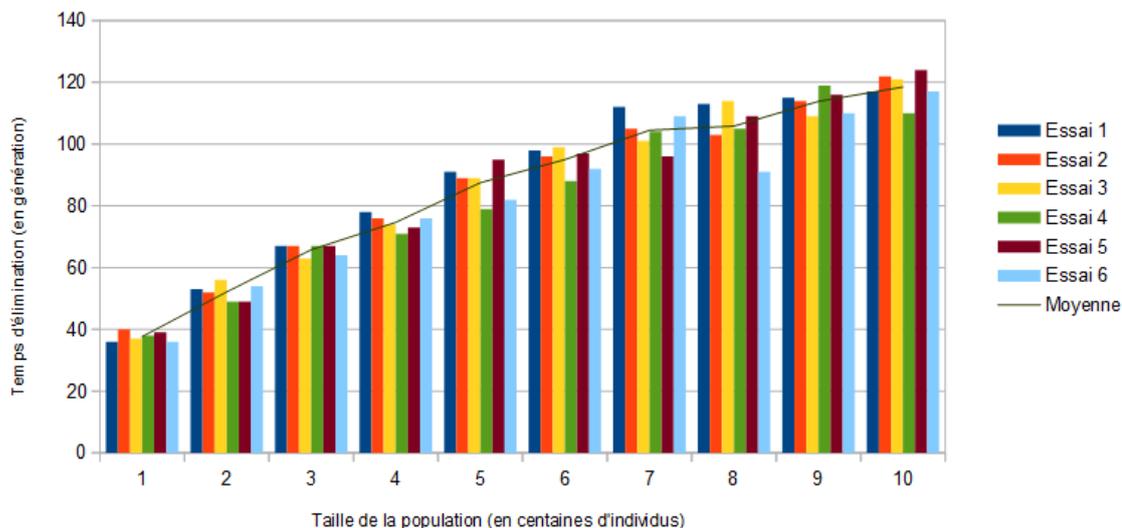
2. Grande échelle

Nous avons ensuite voulu nous intéresser à des populations de tailles plus conséquentes, cependant, ne disposant pas de moyens informatiques suffisamment puissants pour nous permettre de tester la simulation sur des populations de tailles comparables à la population mondiale actuelle, nous avons réalisé des simulations sur des populations allant de 100 à 1000 individus.

Nous avons, suite aux simulations que nous avons effectuées, représenté graphiquement la courbe des moyennes des temps d'élimination de l'allèle M en fonction de la taille de la population. Il s'agit des moyennes des résultats obtenus sur un échantillon de 6000 populations de tailles variables (100 à 1000 individus), regroupant 1 000 000 d'individus au total.

Voici la courbe obtenue :

Diagramme du temps moyen d'élimination de l'allèle M dans une population en fonction de la taille de celle-ci

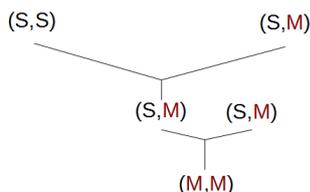


On observe que plus la population est grande, plus le temps d'élimination de l'allèle M est long. Pour une population de 1000 individus, il faut en moyenne 120 générations pour que l'allèle M disparaisse et en estimant qu'une génération équivaut à 20 ans environ, cela représente au moins 2400 ans !

Outre ces informations, on a remarqué à chaque fois que l'allèle M était éliminé au bout d'un certain temps.

2.5 Scénario apocalyptique

Bien que nous ayons toujours observé que l'allèle M est éliminé au bout d'un certain nombre de générations, nous avons tout de même voulu savoir quel serait le temps minimal pour qu'une population soit entièrement malade. En effet cela est théoriquement possible car si la population n'est constituée que de porteurs sains, ceux-ci peuvent engendrer des enfants malades, comme le montre cet arbre généalogique d'une personne malade :



Nous étudierons une population de 100 individus dont 10% sont porteurs sains.

Pour que la maladie se propage le plus rapidement dans la populations, il faut que le nombre de porteurs sains soit maximal, or pour avoir des porteurs sains, il faut former le plus possible de couples constitués d'un individu porteur sain et d'un individu sain. [5]

On a donc représenté l'évolution de cette population par ce tableau :

génération	Individus (M,S)	Individus (S,S)	Individus (M,M)	Couples (S,S)/(S,S)	Couples (S,M)/(S,S)	Couples (S,M)/(S,M)
1	10	90	∅	40	10	∅
2	20	80	∅	30	20	∅
3	40	60	∅	10	40	∅
4	80	20	∅	∅	20	30
5	100	∅	∅	∅	∅	50
6	∅	∅	100	∅	∅	∅

Il faut donc au minimum 6 générations pour que la maladie envahisse totalement cette population.

Nous avons calculé la probabilité qu'un tel scénario se produise.

Voici le calcul :

$$\begin{aligned}
 & 0.5^{20} * 0.5^{40} * 0.5^{80} * [0.5^{40} + 0.5^{60}] * 0.25^{100} \\
 & = 0.5^{140} * [0.5^{40} + 0.5^{60}] * 0.25^{100} \\
 & \simeq 4 * 10^{-115}
 \end{aligned}$$

Les « 0.5 » sont la probabilité pour un couple constitué d'un individu porteur sain et d'un individu sain d'avoir un enfant porteur sain. Les puissances sont les nombres de couples concernés par la situation (la probabilité). Mais étant donné qu'ils ont deux enfants, il s'agit de : *nombre de couples * 2*. Les crochets viennent du fait qu'à la génération 4, le nombre de porteurs sains est supérieur au nombre d'individus sains ; ainsi il faut aussi prendre en compte la probabilité pour un couple de deux porteurs sains d'avoir un enfant porteur sain. Enfin le « 0.25 » est la probabilité pour les couples formés de deux individus porteurs sains d'avoir un enfant malade.

Cependant nous n'avons pas pris en compte la probabilité qu'à chaque génération, tous les porteurs sains s'accouplent avec des individus sains si les couples sont formés aléatoirement. Ainsi, ce résultat est en réalité plus faible encore.

3 Conclusion

Quelle que soit la population étudiée, on a toujours constaté la disparition de la maladie, cependant pour la population actuelle de 7 milliards d'individus, le temps d'élimination de cet allèle n'est pas observable à l'échelle humaine car on a observé que pour des populations de centaines d'individus, les temps moyens d'élimination étaient déjà considérables. De plus, la proportion de porteurs sains actuelle dans la population mondiale est non-négligeable (bien qu'elle reste largement inférieure aux 10 % utilisés dans la population du tableau précédent), ce qui accroît le temps d'élimination théorique.

Enfin les individus malades sont aujourd'hui capables de se reproduire grâce à une assistance médicale.

4 Notes d'édition

[1] Les auteurs précisent ici un certain nombre d'hypothèses sur la maladie, qui seront ensuite utilisées pour faire des simulations numériques. Il faut préciser que les auteurs supposent aussi implicitement que l'effectif de la population étudiée est constant au fil des générations.

[2] Cette égalité ne met pas en jeu des fréquences, mais des probabilités. Par ailleurs, le terme au carré est la probabilité qu'un individu non malade soit porteur sain (puisqu'on suppose que les malades n'ont pas de descendance)

[3] On suppose connu expérimentalement que la maladie touche une personne sur 3500 dans la population française.

[4] Pour comprendre le modèle étudié, il faut regarder précisément le programme.

- une population est une suite de 25 individus, chacun étant caractérisé par un des critères SS, SM, MM.
- On construit au départ (lignes 3-13), par un tirage aléatoire, une population de personnes qui sont soit saines, soit porteurs sains (avec une probabilité de 50%).
- La fonction parent (lignes 15-22) tire aléatoirement une personne fertile dans la population.
- La partie reproduction (lignes 43-47) tire successivement 25 couples dans la population parents (avec répétition, et éventuellement les deux parents sont une même personne). Pour chacun des ces couples, on tire aléatoirement les allèles d'un enfant issu de ce couple. On obtient ainsi 25 enfants.
- On passe à la génération suivante (ligne 49-52), les enfants de la génération n devenant les parents de la génération $n + 1$.
- On répète cette reproduction sur au plus 40 générations.

[5] Cette dernière section laisse le lecteur sceptique. On abandonne le modèle probabiliste précédent. On part ici d'une population de 100 personnes ayant initialement 10% de porteurs sains, et on imagine un scénario qui est décrit par le tableau. Dans la population, il se forme à chaque génération 50 couples qui ont chacun deux enfants. On regarde le cas où à chaque fois que c'est possible les couples sont composés d'un porteur sain et d'une personne saine, et que dans ce cas, les deux enfants sont des porteurs sains.

Le calcul de probabilité qui suit ne tient compte que de la probabilité de naissance des enfants, et pas du tout de la probabilité que les couples se constituent de la façon qui a été supposée.