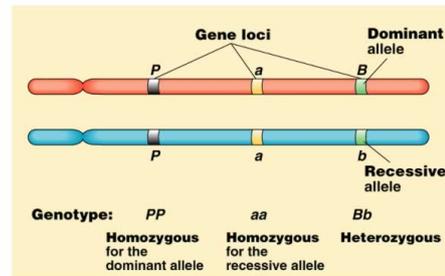


Projet 4 — *Protection de la diversité génétique*

Le gène est l'unité de l'hérédité génétique. Il est caractérisé par son emplacement précis - ou locus - sur un chromosome. Un exemple bien connu est le gène de la couleur des yeux ou du groupe sanguin chez les êtres humains. Un gène correspond également à une séquence de nucléotides, c'est-à-dire à des fragments unitaires d'ADN. Un allèle est une version d'un gène. Ainsi la couleur des yeux chez les humains dépend des allèles portés par l'individu sur le gène concerné. La diversité génétique correspond à l'existence d'allèles différents d'un même gène. Dans les cellules diploïdes (n paires de chromosomes), c'est-à-dire dans les cellules des êtres humains et de la plupart des animaux, chaque individu a deux allèles - correspondant à une paire de chromosomes - pour chaque gène et un individu reçoit, au hasard, un des deux allèles de chacun de ses parents, correspondant à un des deux chromosomes de ses parents. Ces allèles peuvent être identiques ou non. Prenons l'exemple du groupe sanguin avec, par exemple, la situation suivante : un individu A/B a hérité au hasard l'allèle A d'un des parents qui était A/O et l'allèle B de l'autre parent qui était B/O . Un individu dont un gène présente des allèles identiques sur les deux chromosomes est dit homozygote pour ce gène. Dans le cas contraire, on dit que l'individu est hétérozygote pour ce gène.

Dans ce dernier cas, il y a deux possibilités pour le phénotype (état d'un caractère observable) : un des allèles est dominant et l'autre est récessif ou ce n'est pas le cas. Si plus de deux allèles correspondent à un même gène, le gène est polyallélique. La diversité génétique est un des facteurs permettant aux espèces de s'adapter à l'évolution de leur environnement tout particulièrement dans le cadre du changement climatique et des changements globaux. De nombreux articles



publiés dans la littérature de la conservation biologique et consacrés aux programmes de conservation des espèces traitent de problèmes d'optimisation de la diversité génétique. Un type de problèmes concerne la gestion de la reproduction de populations en captivité, de façon à maintenir le plus possible la diversité génétique. Nous présentons ci-dessous un problème de ce type.

Maintien de la diversité allélique dans la descendance d'une population

Nous présentons ici un problème classique rencontré dans les programmes de conservation génétique. Il consiste à déterminer la contribution (nombre de progénitures) optimale de chaque individu de façon à minimiser la perte d'allèles dans la population engendrée. On considère une population formée de P individus, I_1, I_2, \dots, I_p , appartenant à une espèce donnée et un ensemble de M gènes polyalléliques à prendre en compte, g_1, g_2, \dots, g_M . Pour chaque gène g_i , les allèles possibles sont $a_{i_1}, a_{i_2}, \dots, a_{i_{t(i)}}$. On se place dans le cas de cellules diploïdes (être humain et la plupart des animaux) où, pour chaque gène, chaque individu a deux allèles, chaque allèle correspondant à un chromosome de la paire considérée. Ces allèles peuvent être identiques ou non. Chaque individu reçoit au hasard un des deux allèles de chacun de ses parents (voir la Figure 1 ci-dessous). On cherche à déterminer le nombre optimal de progénitures de chaque individu de façon à minimiser l'espérance mathématique du nombre d'allèles disparus dans la génération suivante tout en conservant une population de taille constante.

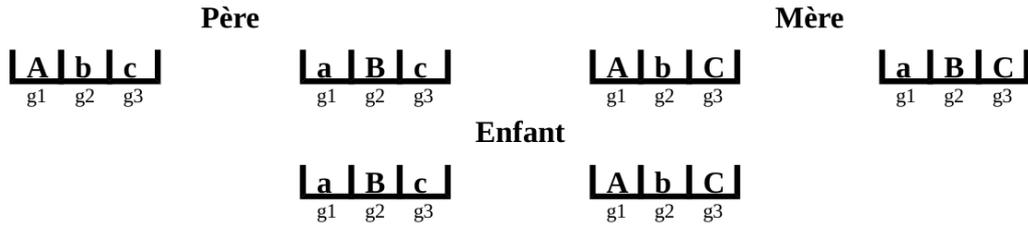
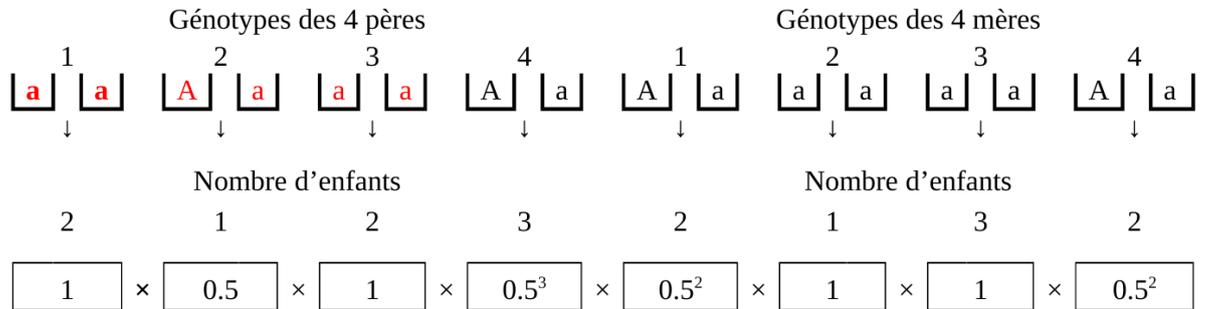


FIGURE 1 – Transmission de 3 gènes bi-alléliques $g_1(A/a)$, $g_2(B/b)$ et $g_3(C/c)$ sur une paire de chromosomes.

Exemple de calcul de la probabilité de disparition d'un allèle donné (sur un locus donné)



Probabilité de disparition de l'allèle $A = 0.5^8$

Ex. couples : (1,1) → 2 (2,2) → 1 (3,3) → 2 (4,3) → 1 (4,4) → 2

Ex. couples : (1,1) → 1 (1,2) → 1 (2,1) → 1 (3,4) → 2 (4,3) → 3

Approximation de la fonction logarithmique par une fonction linéaire par morceaux

Considérons la contrainte :

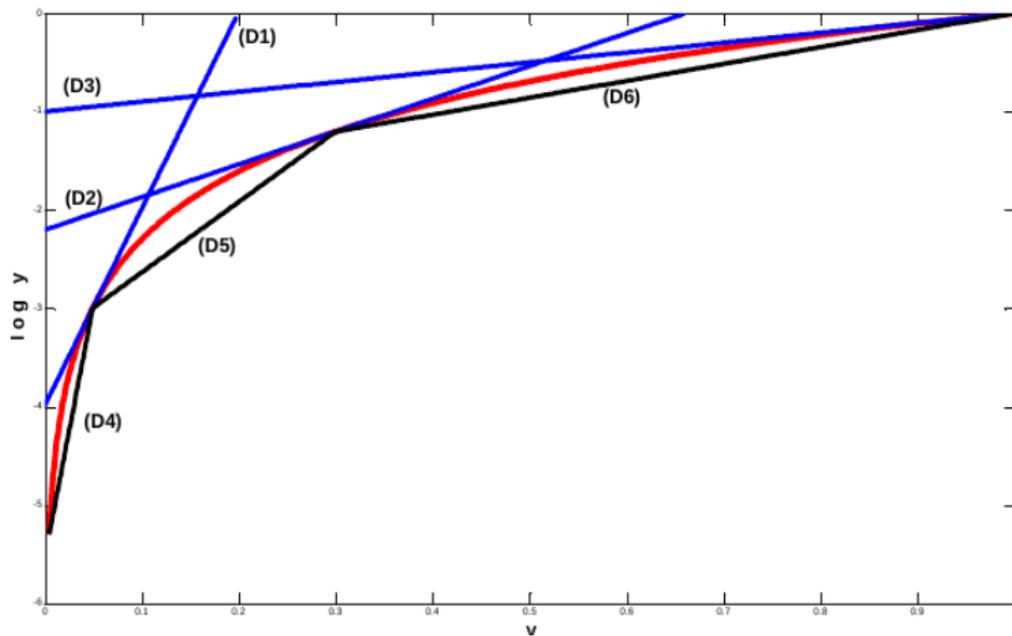
$$\log(y) \geq f(x_1, x_2, \dots, x_n) \quad (1)$$

où $0 < y \leq 1$ et $x \in \mathbb{R}^n$.

L'ensemble de contraintes (C) constitue une **relaxation** de la contrainte (1) :

$$(C) : \log(\theta_r) + \frac{1}{\theta_r}(y - \theta_r) \geq f(x_1, x_2, \dots, x_n) \quad (r \in H)$$

où $H = \{1, 2, \dots, h\}$ et θ est un vecteur donné de \mathbb{R}^h tel que $0 < \theta_1 < \theta_2 < \dots < \theta_h = 1$.



Deux approximations de $\log(y)$ ($0 < y \leq 1$), par une fonction linéaire par morceaux.
 Les droites $(D1)$, $(D2)$ et $(D3)$ sont les contraintes C .
 Les droites $(D4)$, $(D5)$ et $(D6)$ une approximation inférieure.

Travail demandé

- i) Modéliser le problème de la détermination du nombre optimal de progénitures de chaque individu de façon à minimiser l'espérance mathématique du nombre d'allèles disparus dans la génération suivante tout en conservant une population de taille constante.
- ii) Associer au modèle précédent un programme linéaire en variables mixtes permettant de trouver simultanément une solution admissible du problème et une borne inférieure de la valeur de la solution optimale. Idée : approximer la fonction logarithmique par une fonction linéaire par morceaux.
- iii) Résoudre l'instance présentée ci-dessous en limitant le nombre de progénitures de chaque individu à 3 puis à 2.
- iv) Pour chacune des 2 instances, donner le temps de calcul, le nombre de noeuds développés dans l'arbre de recherche, la probabilité de disparition de chaque allèle, l'espérance du nombre d'allèles disparus et la valeur de la borne inférieure.
- v) Étudier le comportement du programme linéaire en fonction de la taille des instances et du nombre de morceaux considérés dans la fonction linéaire par morceaux approxinant la fonction logarithmique, en engendrant aléatoirement des jeux d'essai.

Description des instances et solutions

Instance

Nombre d'individus dans la population : 8 (4 mâles / 4 femelles)

Nombre de paires de chromosomes (par individu) considérées : 1

Nombre de locus considérés par chromosome : 5

Nombre d'allèles par gène : 2

M Â L E S	1	a B c d E	A B C D e
	2	a b C D E	a B c d E
	3	a b c D E	a B C d e
	4	a B C d e	a B c d e
F E M E L L E S	1	a B C d E	A B C d e
	2	A B c d e	a B c D E
	3	A B c D E	A B c D E
	4	a B c d e	a B c D E

Solutions

Solution optimale avec la contrainte $x_j \leq 3 \quad \forall j$								Solution optimale avec la contrainte $x_j = 2 \quad \forall j$							
x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
2	3	3	0	3	1	1	3	2	2	2	2	2	2	2	2
Espérance du nombre d'allèles perdus 0.015869*								Espérance du nombre d'allèles perdus 0.062500*							
Borne inférieure 0.015588*								Borne inférieure 0.062433*							
Probabilité de disparition des allèles A, a, B, C, c, D, d, E = 0 Probabilité de disparition de l'allèle b = 0.015625, de l'allèle e = 0.000244141								Probabilité de disparition des allèles A, a, B, C, c, D, d, E, e = 0 Probabilité de disparition de l'allèle b = 0.062500 (=0.5⁴)							

Résultats obtenus avec $\theta_r = \theta_1^{\frac{h-r}{h-1}}$ où $\theta_1 = 0.001$ et $h = 50$.

Remarque : En prenant $\theta_r = \theta_1^{\frac{h-r}{h-1}}$ où $\theta_1 = 0.0001$ et $h = 50$, on obtient, pour l'instance où $x_j \leq 3$ pour tout j , une solution admissible de valeur 0.015625 et une borne inférieure de valeur 0.015621. La seule probabilité non nulle de disparition d'un allèle correspond à l'allèle b et est égale à 0.015625.