

ANALYSE DU COMPORTEMENT ET DE LA FRÉQUENCE DES GÈNES A EFFETS VISIBLES DANS DES FRATRIES DE GERMAINS ET DE DEMI-GERMAINS

G. LEFORT, L. OLLIVIER * et P. SELLIER *

*Laboratoire de Recherches de la Chaire de Mathématiques,
Institut national agronomique, Paris-Grignon,
16, rue Claude-Bernard,
75231 Paris Cedex 05*

** Station de Génétique quantitative et appliquée,
Centre national de Recherches zootechniques, I. N. R. A.,
78350 Jouy en Josas*

RÉSUMÉ

Cet article décrit une méthode qui permet, dans le cas d'une anomalie héréditaire due à un gène autosomal récessif, d'estimer simultanément sa fréquence q et sa pénétrance w , connaissant le nombre de descendants anormaux observés sur une série de descendance paternelles, chaque descendance paternelle pouvant elle-même être subdivisée en un nombre variable de familles de germains. On suppose qu'on ne dispose d'aucune information sur les parents eux-mêmes, si ce n'est une éventuelle sélection entre la naissance et la mise à la reproduction : ainsi, par exemple, l'élimination complète des anormaux, dans le cas d'une anomalie létale.

On écrit les espérances, en fonction de q et de w , des 3 variables aléatoires X_0 , nombre total de descendants anormaux, D_0 , nombre total de familles de germains contenant au moins un anormal et S_0 , nombre de pères ayant au moins un descendant anormal, et on égale ces espérances aux valeurs observées (méthode des moments). En regroupant les équations relatives à D_0 et S_0 , on obtient un système de 2 équations qui peut être résolu par itération pour calculer \hat{q} et \hat{w} . Le calcul des covariances des estimateurs requiert d'approcher ceux-ci par des fonctions linéaires de X_0 , D_0 et S_0 , obtenues en dérivant les 2 équations précédentes par rapport à q et w , et d'établir la matrice des covariances de X_0 , D_0 et S_0 . La méthode a été programmée sur ordinateur et un exemple numérique, sur le Porc, a été traité.

INTRODUCTION

Dans le cas d'une anomalie héréditaire due à un gène autosomal récessif, il est possible, connaissant le nombre de descendants anormaux d'une série de pères, d'estimer simultanément la fréquence du gène ainsi que sa pénétrance chez l'homo-

zygote (c'est-à-dire la probabilité que l'homozygote récessif manifeste la tare). La méthode a été présentée par LAUVERGNE et LEFORT (1973) pour le cas d'anomalies rencontrées dans l'espèce bovine, à partir d'observations portant sur les descendants de taureaux mis à l'épreuve sur leur descendance ; les descendants d'un même taureau sont alors tous issus de mères différentes et donc demi-germains. Lorsque la descendance de chaque père est subdivisée en famille de germains, comme c'est le cas par exemple chez le Porc où plusieurs individus de la même portée peuvent être observés, la méthode précédente doit être modifiée. L'objet de cet article est d'étendre la méthode à ce cas plus général.

MODÈLE GÉNÉTIQUE ET STRUCTURE DES OBSERVATIONS

La tare est supposée monofactorielle, autosomale, récessive.

A la naissance, les fréquences des trois génotypes AA , Aa et aa dans la population sont supposées conformes à la loi de Hardy-Weinberg (ceci suppose la panmixie et l'égalité fécondité des accouplements). La fréquence du gène (a) responsable de l'anomalie est q ($p = 1 - q$) et sa pénétrance chez l'homozygote est w . Dans cet article les pénétrances sont supposées égales dans les deux sexes, mais la généralisation à des pénétrances différentes ne présente pas de difficultés. L'étude suppose par ailleurs que la pénétrance est d'origine non génétique, mais elle reste approximativement valable tant que la pénétrance est faiblement héritable.

A la mise à la reproduction, les trois génotypes AA , Aa et aa ont des probabilités λ , 2μ et ν respectivement chez les parents ($\lambda + 2\mu + \nu = 1$). λ , μ , ν sont des fonctions de q et de la sélection appliquée aux trois génotypes entre la naissance et la mise à la reproduction. Par exemple, en l'absence de toute sélection, $\lambda = p^2$, $\mu = pq$ et $\nu = q^2$. Lorsque les porteurs de l'anomalie sont écartés de la reproduction, on a :

$$\begin{aligned}\lambda &= p^2/(1 - q^2w), \\ \mu &= pq/(1 - q^2w), \\ \nu &= q^2(1 - w)/(1 - q^2w).\end{aligned}$$

Enfin, si une proportion α des porteurs de l'anomalie est écartée de la reproduction, on a :

$$\begin{aligned}\lambda &= p^2/(1 - \alpha q^2w), \\ \mu &= pq/(1 - \alpha q^2w), \\ \nu &= q^2(1 - \alpha w)/(1 - \alpha q^2w).\end{aligned}$$

Dans la suite, nous supposons α connu. En pratique, on prendra souvent $\alpha = 0$ ou $\alpha = 1$.

Les mâles utilisés sont tirés au hasard dans la population de manière que la probabilité qu'un individu appartienne à l'échantillon soit la même pour tous. Il en est de même pour les femelles accouplées à chacun de ces mâles. La taille de l'échantillon est supposée faible par rapport à celle de la population et donc on peut admettre que les géniteurs ne sont pas apparentés. Tous les descendants nés des accouplements réalisés sont observés.

MODÈLE PROBABILISTE

On désigne par i ($i = 1, 2, \dots, l_0$) l'indice du père, ij l'indice d'une famille de germains ($j = 1, 2, \dots, m_i$) et k l'indice d'un descendant dans une famille ($k = 1, 2, \dots, n_{ij}$). Dans la suite nous adopterons la convention selon laquelle, si une somme est effectuée sur un indice, cet indice est remplacé par un signe $+$. Ainsi, $\sum_j n_{ij} = n_{i+}$.

On notera :

n_0 le nombre total n_{++} des descendants,

m_0 le nombre total m_+ des mères,

l_0 le nombre de pères.

Les aléatoires étudiées sont :

X_{ijk} indicatrice de l'événement « le descendant ijk est anormal »,

D_{ij} indicatrice de l'événement « la famille ij contient au moins un anormal »,

S_i indicatrice de l'événement « le père i a au moins un descendant anormal »,

X_{+++} noté ultérieurement X_0 , est le nombre total de descendants anormaux,

D_{++} noté ultérieurement D_0 , est le nombre total de familles de germains contenant au moins un anormal,

S_+ noté ultérieurement S_0 , est le nombre total de pères ayant au moins un descendant anormal.

Les valeurs prises par ces aléatoires (valeurs observées) sont désignées par les minuscules correspondantes.

Les lois de ces aléatoires sont définies comme suit :

— connaissant i et j et donc les génotypes parentaux et la taille n_{ij} de la famille, les aléatoires X_{ijk} sont indépendantes et la loi de la somme X_{ij+} est binomiale ; la probabilité d'obtenir un anormal est $w, w/2, w/4$ ou 0 selon que le couple ij est $aa \times aa, Aa \times aa, Aa \times Aa$ ou d'un autre type ;

— connaissant le génotype du père i , les sommes X_{i+} sont indépendantes (on a supposé les mères non apparentées) ;

— les sommes X_{i++} sont également indépendantes (on a supposé les pères non apparentés) ;

— les aléatoires D_{ij} et S_i sont des fonctions certaines de X_{ijk} , ce qui détermine leur loi :

$$D_{ij} = 0 \quad \text{si et seulement si} \quad X_{ij+} = 0,$$

$$S_i = 0 \quad \text{si et seulement si} \quad X_{i++} = 0$$

$$\text{ou si et seulement si} \quad D_{i+} = 0.$$

Les expressions des espérances, variances et covariances de toutes ces aléatoires sont données en Annexe.

ESTIMATIONS

La méthode d'estimation consiste à écrire l'espérance des deux variables aléatoires X_0 et $D_0 + S_0$ en fonction des paramètres à estimer et à égaliser ces espérances aux valeurs observées, soit respectivement x_0 et $d_0 + s_0$. C'est une application de la *méthode des moments*.

Le système d'équations à résoudre est donc :

$$(I) \quad \begin{cases} x_0 = E(X_0) = n_0(\mu + \nu)^2 w \\ d_0 + s_0 = E(D_0 + S_0) = m_0(I - \lambda)^2 + l_0(I - \lambda) - \varphi_{++} - \psi_+ \end{cases}$$

où φ_{++} et ψ_+ sont des sommes de fonctions de q , w , des n_{ij} et des m_i , dont les expressions sont données en Annexe.

L'étude de l'aléatoire $D_0 + S_0$ (le regroupement des résultats relatifs aux géniteurs mâles et femelles) est justifiée parce que, dans les expressions de $E(D_0)$ et $E(S_0)$, les termes prépondérants :

$$m_0(I - \lambda)^2 \quad \text{et} \quad l_0(I - \lambda)$$

dépendent tout deux du même paramètre λ . Traiter séparément $E(D_0)$ et $E(S_0)$ donne donc deux estimations différentes de λ sans apporter d'information supplémentaire sur les autres paramètres.

Le système (I) peut être résolu par itération. Partant d'un couple q_0 , w_0 , le système :

$$\begin{aligned} d_0 + s_0 &= m_0(I - \lambda_1)^2 + l_0(I - \lambda_1) - \varphi_{++}(q_0, w_0) - \psi_+(q_0, w_0) \\ x_0 &= n_0(\mu_1 + \nu_1)^2 w_1 \end{aligned}$$

où on a posé :

$$\begin{aligned} \lambda_1 &= p_1^2 / (I - \alpha q_1^2 w_0) \\ \mu_1 &= p_1 q_1 / (I - \alpha q_1^2 w_0) \\ \nu_1 &= q_1^2 (I - \alpha w_0) / (I - \alpha q_1^2 w_0) \end{aligned}$$

permet de calculer λ_1 (d'où p_1 et q_1) dans la première équation, puis w_1 dans la seconde.

Partant de q_1 , w_1 , on obtient q_2 , w_2 et ainsi de suite. D'autres méthodes de résolution numérique, par exemple celle de Newton, peuvent aussi être utilisées.

Pour le calcul des variances et covariances des estimateurs Q et W , on remplace les fonctions de ces aléatoires par leurs différentielles en q et w ; en désignant par $f(q, w)$ l'espérance de X_0 et par $g(q, w)$ celle de $D_0 + S_0$, on aura donc :

$$\begin{aligned} f(Q, W) &\simeq f(q, w) + \frac{\partial f}{\partial q}(q, w)(Q - q) + \frac{\partial f}{\partial w}(q, w)(W - w) \\ g(Q, W) &\simeq g(q, w) + \frac{\partial g}{\partial q}(q, w)(Q - q) + \frac{\partial g}{\partial w}(q, w)(W - w) \end{aligned}$$

Par exemple, si $\alpha = 0$, $\mu + \nu = q$ et donc :

$$f(Q, W) = n_0 Q^2 W \simeq n_0 q^2 w + 2 n_0 q w (Q - q) + n_0 q^2 (W - w)$$

Les estimateurs définis par :

$$\begin{aligned} X_0 &= f(Q, W) \\ D_0 + S_0 &= g(Q, W) \end{aligned}$$

sont alors approchés par des fonctions linéaires de X_0 et $D_0 + S_0$ dont les variances et covariances se calculent aisément : la matrice V des covariances (asymptotiques) des estimateurs Q et W s'écrit en effet :

$$V = T^{-1} \Gamma T'^{-1}$$

où Γ est la matrice des covariances de X_0 et $D_0 + S_0$

et T est la matrice des coefficients de $Q - q$ et $W - w$ dans les différentielles de f et g . On a donc :

$$\Gamma = \begin{pmatrix} V(X_0) & \text{Cov}(X_0, D_0) + \text{Cov}(X_0, S_0) \\ \text{Cov}(X_0, D_0) + \text{Cov}(X_0, S_0) & V(D_0) + V(S_0) + 2 \text{Cov}(D_0, S_0) \end{pmatrix}$$

et

$$T = \begin{pmatrix} \frac{\partial f}{\partial q} & \frac{\partial f}{\partial w} \\ \frac{\partial g}{\partial q} & \frac{\partial g}{\partial w} \end{pmatrix}$$

Les éléments de Γ et T sont donnés en Annexe.

Les valeurs exactes des paramètres q et w étant inconnues, elles sont remplacées, pour le calcul numérique, par leurs estimations \hat{q} et \hat{w} .

EXEMPLE NUMÉRIQUE

Nous avons appliqué la méthode qui vient d'être décrite ici en détail à des observations relatives au syndrome d'hyperthermie maligne chez le porc de *Piétrain* (voir OLLIVIER *et al.*, 1975). Les données figurent au tableau I.

Le système (1) à résoudre pour obtenir \hat{q} et \hat{w} est simplifié du fait qu'on suppose l'absence de sélection chez les parents ($\alpha = 0$) :

$$\begin{cases} 39 = 138 q^2 w \\ 21 = 25 (1 - p^2)^2 + 7 (1 - p^2) - \varphi_{++} - \psi_+ \end{cases}$$

Les valeurs initiales choisies pour q_0 et w_0 sont $w_0 = 1$ et, en utilisant la première équation, $q_0 = \sqrt{39/138} = 0,53$. La convergence est rapide et après 5 itérations la précision obtenue est supérieure à 10^{-3} . On obtient :

$$\hat{q} = 0,64 \quad \hat{w} = 0,69$$

En remplaçant q et w par les valeurs ci-dessus on obtient la matrice Δ des covariances des aléatoires X_0 , D_0 et S_0 , et Γ qui en dérive :

$$\Delta = \begin{pmatrix} 122,30 & 28,85 & 5,93 \\ 28,85 & 10,77 & 2,36 \\ 5,93 & 2,36 & 1,02 \end{pmatrix} \quad \Gamma = \begin{pmatrix} 122,30 & 34,78 \\ 34,78 & 16,50 \end{pmatrix}$$

$$T = \begin{pmatrix} 121,88 & 56,52 \\ 39,53 & 7,95 \end{pmatrix}$$

On en déduit la matrice des covariances de Q et W :

$$V = \begin{pmatrix} 0,018 & -0,025 \\ -0,025 & 0,063 \end{pmatrix}$$

et les écarts-types et le coefficient de corrélation :

$$\sigma_q = 0,13 \quad \sigma_w = 0,25 \quad \rho(q, w) = -0,75$$

On remarque que Q et W ont une corrélation négative très forte, ce qui s'explique par le fait que le produit qw intervient fréquemment dans les équations. Par ailleurs, la précision sur w est très mauvaise et la valeur $w = 1$ ne peut pas être exclue.

TABLEAU I

Répartition par famille des sujets manifestant le syndrome d'hyperthermie maligne

Père i	Mère j	Descendance			
		n_{ij}	X_{ij+}	D_{ij}	S_i
1	1	7	0	0	} 0
	2	7	0	0	
	3	5	0	0	
	4	4	0	0	
	5	7	0	0	
2	1	8	0	0	0
3	1	9	4	1	} 1
	2	2	2	1	
	3	9	5	1	
	4	4	4	1	
4	1	9	1	1	} 1
	2	4	1	1	
	3	7	3	1	
5	1	1	0	0	} 1
	2	4	1	1	
6	1	7	1	1	} 1
	2	3	3	1	
	3	10	0	0	
	4	6	1	1	
	5	2	1	1	
	6	4	1	1	
7	1	7	6	1	} 1
	2	4	0	0	
	3	4	2	1	
	4	4	3	1	
TOTAUX	m_0	n_0	X_0	D_0	S_0
	25	138	39	16	5

(Symboles expliqués dans le texte).

DISCUSSION

L'analyse qui vient d'être présentée permet d'étendre aux espèces animales domestiques polytoques (Mouton, Porc, Lapin), ainsi qu'à la Poule, une méthode mise au point initialement pour l'étude des populations bovines. Ce genre de démarche, analogue à celle suivie dans l'étude des maladies génétiques chez l'Homme, présente surtout de l'intérêt lorsque l'expérimentation directe est impossible (comme chez l'Homme) ou trop coûteuse (comme chez les Bovins). Pour les espèces domestiques de petite taille, l'expérimentation est en général possible et elle permet de préciser le déterminisme exact du phénomène observé. La méthode que nous venons de présenter peut être utilisée, dans le cas où un déterminisme monofactoriel récessif est établi, pour l'estimation de la fréquence génique et de la pénétrance. Remarquons cependant que le résultat d'accouplements entre individus porteurs de l'anomalie apporte une information plus précise sur la pénétrance.

L'avantage de la méthode d'estimation exposée réside dans la simplicité du calcul des estimations. Les difficultés dues à l'existence de corrélations entre germains et entre demi-germains sont réduites au minimum dans le calcul des espérances. Elles commencent à apparaître, par contre, dans le calcul des variances et covariances.

D'autres méthodes d'estimation sont possibles, parmi lesquelles celle du maximum de vraisemblance, qui est la plus couramment utilisée; elle a été adaptée, en particulier, à l'étude des populations humaines par MORTON (1965), et a permis de prendre en considération des hypothèses génétiques complexes (voir MORTON *et al.*, 1971). L'application au cas des populations animales domestiques paraît cependant difficile du fait de l'imbrication des familles de germains à l'intérieur des descendance paternelles.

Le calcul des variances et covariances utilise des résultats « asymptotiques » ; l'approximation faite est d'autant plus satisfaisante que la variabilité relative des aléatoires est faible ; il est donc nécessaire que les effectifs expérimentaux soient assez grands. Dans l'exemple traité, on est à la limite de ce qui est acceptable. La même remarque vaudrait également pour le maximum de vraisemblance.

Remarquons enfin que les solutions données par LAUVERGNE et LEFORT (1973), qui supposent $n_{ij} = 1$, s'appliquent au cas où l'anomalie a une faible incidence dans la population (q^2w faible). Cette restriction est ici levée, mais au prix de calculs plus complexes. La méthode a été programmée en FORTRAN et en APL, et les programmes peuvent être obtenus sur demande.

Reçu pour publication en novembre 1975.

SUMMARY

AN ANALYSIS OF THE BEHAVIOUR AND FREQUENCY OF GENES WITH
VISIBLE EFFECTS IN FULL-AND HALF-SIBSHIPS

A method is described which, in the case of a hereditary abnormality due to an autosomal recessive gene, gives a simultaneous estimation of its frequency q and of its penetrance w , knowing the number of abnormal observed on a series of sire-progenies, each sire-progeny being itself

eventually subdivided into a variable number of full-sib families. No information is supposed to be available on the parents themselves, but the method assumes that a known selection has been applied to them between birth and reproduction, as for instance complete selection against the abnormality, which is the case for a lethal character.

The method is based on writing, as functions of q and w , the expectations of 3 random variables, X_0 , total number of abnormals, D_0 , total number of full-sib families including at least one abnormal, and S_0 , number of sires having at least one abnormal offspring, and on equating these expectations to the observed values (method of moments). By grouping the equations for D_0 and S_0 , a system of 2 equations is obtained which can be solved by iteration to yield the 2 estimations \hat{q} and \hat{w} . In order to obtain the covariances of the estimators, these are approximated by linear functions of X_0 , D_0 and S_0 (taking the derivatives of the 2 above equations with respect to q and w) and the variance-covariance matrix of X_0 , D_0 and S_0 is established. A computer program has been written and a numerical application, on pig data, is given.

RÉFÉRENCES BIBLIOGRAPHIQUES

- LAUVERGNE J. J., LEFORT G., 1973. Nouvelle méthode pour analyser le comportement et la fréquence des gènes récessifs à effets visibles dans les populations bovines. *C. R. Acad. Sci. Paris, Série D*, **277**, 2793-2796.
- MORTON N. E., 1965. Models and evidence in human population genetics, in GEERTS S. J., *Genetics Today*, **3**, 935-951. Pergamon Press, Oxford. *Proc. XI Int. Congr. Genetics*, The Hague, 1963.
- MORTON N. E., YEE S., LEW R., 1971. Complex segregation analysis. *Amer. J. hum. Genet.*, **23**, 602-611.
- OLLIVIER L., SELLIER P., MONIN G., 1975. Déterminisme génétique du syndrome d'hyperthermie maligne chez le porc de Piétrain. *Ann. Génét. Sé. anim.*, **7**, 159-165.

ANNEXE

I. — Calcul des moments des aléatoires observables

L'étude qui va suivre utilise la loi conditionnelle des aléatoires, connaissant les effectifs n_{ij} des familles (la répartition des tailles de famille étant supposée ne pas dépendre des génotypes parentaux). Une espérance est calculée pour chaque type d'accouplement (espérance conditionnelle) et l'espérance recherchée s'obtient en pondérant les espérances conditionnelles par les probabilités des divers types d'accouplements.

1. Moments d'ordre 1 (moyennes).

Espérances de S_0 et D_0 :

$$E(S_i) = P(S_i = 1) = 1 - P(D_{i+} = 0)$$

Connaissant le génotype du père i , les aléatoires D_{ij} sont indépendantes et :

$$P(D_{i+} = 0) = \prod_1^{m_i} P(D_{ij} = 0)$$

On a les situations suivantes :

Génotype du père	Probabilité	$P(D_{ij} = 1)$	$P(D_{i+} = 0)$
<i>AA</i>	λ	0	1
<i>Aa</i>	2μ	$h_{ij} = 1 - \lambda - 2\mu(1 - w/4)^{n_{ij}} - \nu(1 - w/2)^{n_{ij}}$	$\prod_j (1 - h_{ij})$
<i>aa</i>	ν	$r_{ij} = 1 - \lambda - 2\mu(1 - w/2)^{n_{ij}} - \nu(1 - w)^{n_{ij}}$	$\prod_j (1 - r_{ij})$

D'où :

$$E(D_{ij}) = P(D_{ij} = 1) = 2\mu h_{ij} + \nu r_{ij} = (1 - \lambda)^2 - \varphi_{ij}$$

avec :

$$\varphi_{ij} = \nu^2(1 - w)^{n_{ij}} + 4\mu\nu(1 - w/2)^{n_{ij}} + 4\mu^2(1 - w/4)^{n_{ij}}$$

et

$$P(D_{i+} = 0) = \lambda + \psi_i$$

avec

$$\psi_i = 2\mu \prod_j (1 - h_{ij}) + \nu \prod_j (1 - r_{ij}).$$

On en déduit :

$$E(S_i) = 1 - P(D_{i+} = 0) = 1 - \lambda - \psi_i$$

et

$$E(S_0) = \sum_i E(S_i) = l_0(1 - \lambda) - \psi_+$$

$$E(D_0) = \sum_{ij} E(D_{ij}) = m_0(1 - \lambda)^2 - \varphi_{++}$$

Espérance de X_0 :

La fréquence du gène responsable de l'anomalie étant $\mu + \nu$ chez les parents, on en déduit :

$$E(X_{ijk}) = P(X_{ijk} = 1) = w(\mu + \nu)^2$$

et

$$E(X_0) = \sum_{ijk} E(X_{ijk}) = n_0 w (\mu + \nu)^2$$

2. *Moments d'ordre 2 (variances et covariances).*

Si les descendants considérés n'avaient aucun lien de parenté entre eux, les aléatoires X_{ijk} seraient indépendantes et de même les D_{ij} et les S_i . Les expressions des variances et covariances de X_0 , D_0 et S_0 seraient alors simples. En fait, l'existence de germains et de demi-germains dans l'échantillon complique les formules.

Variance de S_0 :

La variance de l'indicatrice S_i est :

$$E(S_i) [1 - E(S_i)]$$

Les indicatrices S_i sont indépendantes. D'où :

$$V(S_0) = \sum_i V(S_i) = \sum_i (\lambda + \psi_i) (\mathbf{I} - \lambda - \psi_i)$$

Covariance de S_0 et D_0 :

$$S_i = \mathbf{I} \Leftrightarrow D_{i+} \neq 0$$

D'où :

$$\begin{aligned} S_i D_{i+} &= D_{i+} \quad \text{et} \quad E(S_i D_{i+}) = E(D_{i+}) \\ \text{Cov}(S_i, D_{i+}) &= E(S_i D_{i+}) - E(S_i) E(D_{i+}) \\ &= E(D_{i+}) [\mathbf{I} - E(S_i)] \end{aligned}$$

Les couples d'aléatoires $(S_i; D_{i+})$ sont indépendants, d'où :

$$\begin{aligned} \text{Cov}(S_0, D_0) &= \sum_i \text{Cov}(S_i, D_{i+}) \\ &= \sum_i [m_i (\mathbf{I} - \lambda)^2 - \varphi_{i+}] (\lambda + \psi_i) \end{aligned}$$

Covariance de S_0 et X_0 :

Comme précédemment, $S_i = \mathbf{I} \Leftrightarrow X_{i++} \neq 0$.

Les couples d'aléatoires $(S_i; X_{i++})$ sont indépendants, d'où :

$$\begin{aligned} \text{Cov}(S_0, X_0) &= \sum_i \text{Cov}(S_i, X_{i++}) \\ &= w(\mu + \nu)^2 (n_0 \lambda + \sum_i n_{i+} \psi_i) \end{aligned}$$

Variance de D_0 :

Conditionnellement au génotype du père, les aléatoires D_{ij} et $D_{i'j'}$ ($j' \neq j$) sont indépendantes ; d'où :

$$E(D_{ij} D_{i'j'} / \text{père}) = E(D_{ij} / \text{père}) E(D_{i'j'} / \text{père})$$

D'où :

$$E(D_{ij} D_{i'j'}) = 2\mu h_{ij} h_{i'j'} + \nu r_{ij} r_{i'j'}$$

$$\begin{aligned} \text{et} \quad \text{Cov}(D_{ij}, D_{i'j'}) &= E(D_{ij} D_{i'j'}) - E(D_{ij}) E(D_{i'j'}) \\ &= 2\mu h_{ij} h_{i'j'} + \nu r_{ij} r_{i'j'} - [(I - \lambda)^2 - \varphi_{ij}] [(I - \lambda)^2 - \varphi_{i'j'}] \end{aligned}$$

On en déduit la variance de D_{i+}

$$V(D_{i+}) = \sum_j \left[V(D_{ij}) + \sum_{j' \neq j} \text{Cov}(D_{ij}, D_{i'j'}) \right]$$

et en raison de l'indépendance des aléatoires D_{i+}

$$V(D_0) = \sum_i V(D_{i+})$$

L'aléatoire D_{ij} est une indicatrice et sa variance est donc :

$$V(D_{ij}) = [(I - \lambda)^2 - \varphi_{ij}] [I - (I - \lambda)^2 + \varphi_{ij}]$$

D'où :

$$V(D_0) = m_0(\mathbf{I} - \lambda)^2 - \varphi_{++} - \sum [m_i(\mathbf{I} - \lambda)^2 - \varphi_{i+}]^2 \\ + 2\mu \sum_i (h_{i+}^2 - \sum_j h_{ij}^2) + \nu \sum_i (r_{i+}^2 - \sum_j r_{ij}^2).$$

Covariance de D_0 et X_0 :

$$D_{ij} = \mathbf{I} \Leftrightarrow X_{ij+} \neq 0$$

D'où :

$$D_{ij}X_{ij+} = X_{ij+} \quad \text{et} \quad E(D_{ij}X_{ij+}) = E(X_{ij+}) = n_{ij}w(\mu + \nu)^2 \\ \text{Cov}(D_{ij}, X_{ij+}) = E(X_{ij+}) [\mathbf{I} - E(D_{ij})] = n_{ij}w(\mu + \nu)^2 [\mathbf{I} - (\mathbf{I} - \lambda)^2 + \varphi_{ij}]$$

Conditionnellement au génotype du père, il y a indépendance de D_{ij} et $X_{ij'+}$ si $j' \neq j$; d'où :

$$E(D_{ij}X_{ij'+}/\text{père}) = E(D_{ij}/\text{père}) E(X_{ij'+}/\text{père}).$$

Sachant que :

$$E(X_{ij+}/\text{père } Aa) = n_{ij}(\mu + \nu)w/2 \quad \text{et} \quad E(X_{ij+}/\text{père } aa) = n_{ij}(\mu + \nu)w$$

on a :

$$E(D_{ij}X_{ij'+}) = 2\mu h_{ij}n_{ij'}(\mu + \nu)w/2 + \nu r_{ij}n_{ij'}(\mu + \nu)w$$

et

$$\text{Cov}(D_{ij}, X_{ij'+}) = n_{ij'}w(\mu + \nu) \{ \mu h_{ij} + \nu r_{ij} - (\mu + \nu) [(\mathbf{I} - \lambda)^2 - \varphi_{ij}] \}$$

On en déduit :

$$\text{Cov}(D_{i+}, X_{i++}) = \sum_j [\text{Cov}(D_{ij}, X_{ij+}) + \sum_{j' \neq j} \text{Cov}(D_{ij}, X_{ij'+})] \\ = \sum_j \{ n_{ij}w(\mu + \nu)^2 + (n_{i+} - n_{ij}) w(\mu + \nu) (\mu h_{ij} + \nu r_{ij}) \\ - n_{ij}w(\mu + \nu)^2 [(\mathbf{I} - \lambda)^2 - \varphi_{ij}] \} \\ = n_{i+}w(\mu + \nu)^2 [\mathbf{I} - m_i(\mathbf{I} - \lambda)^2 + \varphi_{i+}] \\ + w(\mu + \nu) \left[n_{i+}(\mu h_{i+} + \nu r_{i+}) - \sum_j n_{ij}(\mu h_{ij} + \nu r_{ij}) \right]$$

D'où, puisque les couples $(D_{i+}; X_{i++})$ sont indépendants :

$$\text{Cov}(D_0, X_0) = \sum_i \text{Cov}(D_{i+}, X_{i++}) \\ = w(\mu + \nu)^2 \left\{ n_0 - \sum_i n_{i+} [m_i(\mathbf{I} - \lambda)^2 - \varphi_{i+}] \right\} \\ + w(\mu + \nu) \left\{ \sum_i n_{i+}(\mu h_{i+} + \nu r_{i+}) - \sum_{ij} n_{ij}(\mu h_{ij} + \nu r_{ij}) \right\}$$

Variance de X_0 :

La variance de X_0 est la somme des variances des X_{ijk} et des covariances entre X_{ijk} et $X_{ijk'}$ ($k' \neq k$) et entre X_{ijk} et $X_{ij'k}$ ($j' \neq j$). Il s'agit donc de covariances entre germains et entre demi-germains; celles-ci sont, en régime de panmixie, des fonctions

simples de la variance génétique additive σ_A^2 et de la variance de dominance σ_D^2 de l'aléatoire X_{ijk} , soit $\sigma_A^2/4$ pour la covariance entre demi-germains et $(\sigma_A^2/2 + \sigma_D^2/4)$ pour la covariance entre germains. Ces expressions des covariances restent approximativement valables si les fréquences génotypiques parentales ne s'écartent pas trop de l'équilibre de Hardy-Weinberg. Les variances σ_A^2 et σ_D^2 s'expriment en fonction des fréquences géniques p' et q' chez les descendants (égales à p et q en l'absence de sélection chez les parents et à $\lambda + \mu$ et $\mu + \nu$ respectivement dans le cas contraire) :

$$\sigma_A^2 = 2p'q'^2w^2 \quad \text{et} \quad \sigma_D^2 = p'^2q'^2w^2$$

on a donc :

$$V(X_0) = \sum_{ijk} V(X_{ijk}) + \sum_{ij} n_{ij}(n_{ij} - 1) (p'q'^3w^2 + p'^2q'^2w^2/4) \\ + \left(\sum_i n_{i+}^2 - \sum_{ij} n_{ij}^2 \right) (p'q'^3w^2/2)$$

Comme $V(X_{ijk}) = q'^2w(1 - q'^2w)$, on obtient :

$$V(X_0) = n_0q'^2w [1 - w(1 + q')^2/4] + \sum_{ij} n_{ij}^2q'^2(1 - q'^2)w^2/4 + \sum_i n_{i+}^2p'q'^3w^2/2$$

II. — Calcul des dérivées partielles des espérances de S_0 , D_0 , X_0

L'algorithme de calcul est le suivant :

1. Calcul des dérivées partielles des fréquences génotypiques parentales :

$$\frac{\partial \lambda}{\partial q} = -2p(1 - \alpha qw)/(1 - \alpha q^2w)^2 \quad \frac{\partial \lambda}{\partial w} = \alpha p^2q^2/(1 - \alpha q^2w)^2 \\ \frac{\partial \mu}{\partial q} = [p - q(1 - \alpha qw)]/(1 - \alpha q^2w)^2 \quad \frac{\partial \mu}{\partial w} = \alpha pq^3/(1 - \alpha q^2w)^2 \\ \frac{\partial \nu}{\partial q} = 2q(1 - \alpha w)/(1 - \alpha q^2w)^2 \quad \frac{\partial \nu}{\partial w} = \alpha q^2(1 - q^2)/(1 - \alpha q^2w)^2$$

2. Calcul des dérivées partielles des h_{ij} et r_{ij} :

$$\frac{\partial h_{ij}}{\partial q} = -\frac{\partial \lambda}{\partial q} - 2\frac{\partial \mu}{\partial q} (1 - w/4)^{n_{ij}} - \frac{\partial \nu}{\partial q} (1 - w/2)^{n_{ij}} \\ \frac{\partial r_{ij}}{\partial q} = -\frac{\partial \lambda}{\partial q} - 2\frac{\partial \mu}{\partial q} (1 - w/2)^{n_{ij}} - \frac{\partial \nu}{\partial q} (1 - w)^{n_{ij}} \\ \frac{\partial h_{ij}}{\partial w} = -\frac{\partial \lambda}{\partial w} - 2\frac{\partial \mu}{\partial w} (1 - w/4)^{n_{ij}} - \frac{\partial \nu}{\partial w} (1 - w/2)^{n_{ij}} \\ + \frac{n_{ij}}{2} [\mu(1 - w/4)^{n_{ij}-1} + \nu(1 - w/2)^{n_{ij}-1}] \\ \frac{\partial r_{ij}}{\partial w} = -\frac{\partial \lambda}{\partial w} - 2\frac{\partial \mu}{\partial w} (1 - w/2)^{n_{ij}} - \frac{\partial \nu}{\partial w} (1 - w)^{n_{ij}} \\ + n_{ij}[\mu(1 - w/2)^{n_{ij}-1} + \nu(1 - w)^{n_{ij}-1}].$$

3. Les dérivées de φ_{ij} se déduisent des dérivées précédentes. Pour le calcul des dérivées des produits figurant dans ψ_i , on utilise les dérivées logarithmiques :

$$\frac{\partial}{\partial q} \prod_j (1 - h_{ij}) / \prod_j (1 - h_{ij}) = -\sum_j \left[\frac{\partial h_{ij}}{\partial q} / (1 - h_{ij}) \right]$$

4. Les dérivées partielles de :

$$E(X_0) = f(q, w) = n_0 w (\mu + v)^2$$

$$E(D_0 + S_0) = g(q, w) = m_0 (1 - \lambda)^2 + l_0 (1 - \lambda) - \varphi_{++} - \psi_+$$

se déduisent simplement des résultats précédents.