

Commentaires sur une proposition d'extension du modèle génétique additif classique à des situations non-panmictiques

C. CHEVALET

*I.N.R.A., Laboratoire de Génétique cellulaire, Centre de Recherches de Toulouse
Auzeville, B.P. 12, F - 31320 Castanet-Tolosan*

Résumé

On précise quelques résultats obtenus par LANGLOIS (*Ann. Génét. Sél. anim.*, **13**, 151-163, 1981) sur l'expression des ressemblances entre apparentés en présence d'homogamie et de corrélation entre les effets du génotype et du milieu. On énonce ensuite les règles générales de calcul qui ont été utilisées implicitement, on en donne quelques exemples et on discute leur intérêt et les difficultés qui peuvent surgir.

La proposition, présentée par LANGLOIS (1981), appelle quelques commentaires que, je pense, en souligneront l'intérêt. Certains résultats, d'une part, méritent d'être précisés, et les procédés de calcul utilisés, d'autre part, gagnent à être exposés et discutés avec plus de détail en raison de leur portée générale.

L'intérêt et la signification des résultats

L'objet de l'article est d'introduire, dans l'interprétation de la ressemblance entre apparentés, deux paramètres en général supposés nuls : la corrélation r entre les phénotypes de conjoints, et une corrélation s entre génotype et effet du milieu. Il faut souligner immédiatement une différence importante entre ces deux paramètres : le premier est directement mesurable, le second ne l'est pas et n'est introduit que pour tenter d'obtenir un meilleur ajustement des données au modèle.

Les résultats obtenus par LANGLOIS sont, essentiellement :

1) Pour $s = 0$, $r \neq 0$, les résultats de FISHER (1918) et de WRIGHT (1921) sont retrouvés, ils sont étendus à la corrélation entre demi-frères en tenant compte d'une corrélation quelconque entre les conjoints du parent commun.

2) Pour $s \neq 0$, les formes théoriques des covariances *phénotypiques* sont identiques à celles obtenues avec $s = 0$, quand on y remplace l'expression h^2 par :

$h^2 = (h + es)^2$. Donc, en ce qui concerne les corrélations phénotypiques, le calcul avec un paramètre s non nul n'apporte rien. Autrement dit, si pour tous les individus observés ou à naître dans la population, on admet l'existence d'une même corrélation s entre génotype et environnement, alors le paramètre estimable et permettant la prédiction statistique des phénotypes est la quantité $h^2 = (h + es)^2$, dans laquelle h et s ne sont pas séparables.

La distinction entre ces paramètres h et h' apparaît seulement si l'on considère les covariances ou les régressions entre génotypes et phénotypes, par exemple si l'on veut estimer la valeur génétique d'un père (A_p) d'après la valeur phénotypique (D) d'un descendant :

$$b_{A_p/D} = \frac{1}{2} h h' (1 + r h^2) \text{Var}(P) / \text{Var}(D)$$

Cette formule a un caractère prédictif intéressant seulement si le père en question doit être utilisé dans un autre système de production, où sa valeur génétique A_p est un prédicteur de la valeur phénotypique de futurs descendants. Et cette formule est utilisable seulement si l'héritabilité « pure », h^2 , a pu être estimée, par exemple à partir d'un échantillon d'animaux élevés de telle sorte que pour eux la corrélation s soit nulle.

En conclusion, l'introduction dans les calculs d'une corrélation non nulle entre génotype et milieu se justifie uniquement dans les situations où plusieurs systèmes d'élevage, caractérisés par des valeurs distinctes de cette corrélation, coexistent, et où l'on doit évaluer la valeur d'élevage d'un individu dans un système d'après les performances réalisées par ses apparentés dans l'autre système.

Les règles de calcul des corrélations

La technique de calcul des corrélations constitue à mon avis l'intérêt majeur de cet article, parce qu'elle permet de travailler directement au niveau statistique, sans ré-écrire le modèle génétique multi-factoriel sous jacent, et sans faire appel à la théorie des coefficients de piste. Il me semble utile de formaliser un peu plus précisément la méthode, pour montrer qu'elle peut être d'un emploi assez général, et pour signaler les difficultés de son application dans certaines situations.

Les règles et hypothèses peuvent s'énoncer ainsi :

- 1) On suppose la linéarité des régressions.
- 2) Le modèle d'hérédité quantitative est défini par un certain nombre de paramètres (des coefficients de corrélation) qui expriment certains effets (homogamie, liaison entre génotype et milieu, hérédité mendélienne, milieu commun, etc.). Les coefficients de corrélation qui ne sont pas définis par un paramètre du modèle sont déduits par application de la règle suivante, et s'expriment en fonction des seuls paramètres du modèle.
- 3) Les corrélations non spécifiées par la définition d'un paramètre sont calculées selon le principe suivant : quand une nouvelle variable du modèle est introduite, les

nouvelles corrélations (non définies par un paramètre) sont choisies de telle sorte que la matrice des coefficients de corrélation ait un déterminant maximum. Ce principe peut s'énoncer aussi : quand on fixe les variables qui ont avec la nouvelle variable introduite des corrélations fixées par le modèle, cette nouvelle variable devient (conditionnellement) indépendante des autres variables non fixées. L'équivalence entre ces deux énoncés peut être vérifiée facilement en montrant que la solution donnée par l'hypothèse d'indépendance conditionnelle assure le maximum du déterminant quand on y a fixé les corrélations postulées par le modèle. En toute rigueur, l'hypothèse de linéarité n'est pas nécessaire pour utiliser le critère du déterminant, qui présente aussi l'avantage de pouvoir se dispenser d'un diagramme. Néanmoins les calculs de maximisation du déterminant sont généralement plus lourds, et même si l'hypothèse de linéarité paraît mal justifiée, elle est un moyen de conduire les calculs plus aisément.

Ces règles sont appliquées tout au long des calculs présentés dans l'article, elles conduisent à des résultats qui dépendent de l'ordre dans lequel les nouvelles variables sont introduites. L'ordre adopté, dans l'article de LANGLOIS, ne pose pas de problème, mais certaines situations peuvent conduire à des indéterminations. Quelques principes de bon sens s'imposent pour le choix de l'ordre : les individus sont introduits dans un ordre compatible avec les généalogies considérées (les parents sont introduits avant les enfants). Un individu étant caractérisé par plusieurs variables, il paraît logique d'introduire en premier lieu celles dont les corrélations avec les variables précédentes sont définies par un paramètre du modèle : pour un conjoint, le phénotype (corrélation r), pour un descendant le génotype A_d :

$$E(A_d/A_p, A_m) = \frac{1}{2} A_p + \frac{1}{2} A_m$$

pour un enfant adoptif on peut penser introduire d'abord le milieu. Cette règle n'est pas d'une application toujours évidente, quand plusieurs variables attachées à un même individu sont liées aux variables précédentes : introduction d'un frère en présence d'un effet de milieu commun dans une fratrie ; choix d'un conjoint apparenté en présence d'homogamie. Le choix de l'ordre d'introduction doit alors s'appuyer sur des considérations ou des hypothèses supplémentaires. L'introduction simultanée de plusieurs nouvelles variables est possible tout en respectant le principe d'indépendance conditionnelle ou de déterminant maximum, mais elle conduit à des systèmes d'équations algébriques non linéaires dont les solutions sont les corrélations à déduire. Ces équations peuvent avoir plusieurs solutions, dont on ne peut en général pas donner d'expression explicite.

Exemples de calculs

Pratiquement, la conduite des calculs peut s'appuyer sur un diagramme où sont représentées toutes les variables du système, mais où l'on ne relie que les couples de variables dont la corrélation est définie par un paramètre du modèle. Les schémas 1 et 2 de LANGLOIS peuvent être représentés par un seul diagramme (figure 1), où l'on ne porte que la corrélation r entre les phénotypes P et M des conjoints, les

coefficients de corrélation partielle 1/2 des lois de Mendel, la corrélation s entre génotype et milieu (fixée à 0 dans le schéma 1 et à une valeur quelconque dans le schéma 2), et les corrélations h et e entre le phénotype et ses composantes génétique et environnementale ($h^2 + 2hes + e^2 = 1$). Chaque parent est représenté par le triplet (phénotype - génotype - milieu) ; les deux conjoints sont liés par la corrélation phénotypique r (qui est fixée aussi, et nulle, dans le cas de panmixie) ; si le milieu du descendant n'est pas directement dépendant des valeurs parentales, le phénotype D du descendant ne dépend des variables parentales que par le génotype A_d . L'ordre de calcul est donc : P, A_p , R_p , M, A_m , R_m , A_d , R_d , D. Entre les trois premières variables, la matrice de corrélations est par définition :

$$\begin{vmatrix} 1 & h + es & hs + e \\ h + es & 1 & s \\ hs + e & s & 1 \end{vmatrix}$$

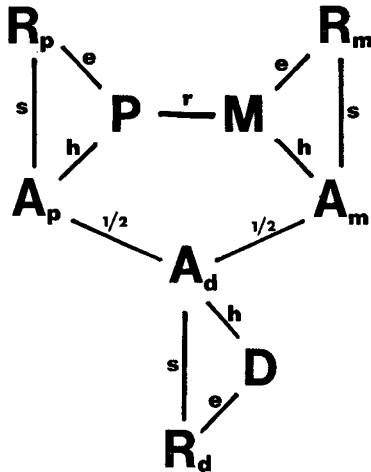


FIG. 1

Diagramme des corrélations fixées entre un individu et ses deux parents, en présence d'homogamie et de corrélation entre génotype et milieu

Diagram of the fixed correlations between parents and offspring, with assortative mating and genotype-environment correlation

La corrélation entre P et M est r , par définition ; les corrélations de M avec A_p et R_p , soient x et y sont déterminées par l'application de la troisième règle. Pour utiliser le critère du déterminant maximum, il faut prendre soin de considérer un ensemble non redondant de variables : ci-dessus, les trois variables sont liées par l'identité :

$$R_p = P - A_p$$

et le déterminant de la matrice de corrélation est nul ; le critère du déterminant doit être appliqué à la suite : P, A_p , M, A_m , A_d , R_d . L'équation en x s'écrit :

$$\frac{d}{dx} \begin{vmatrix} 1 & h + es & r \\ h + es & 1 & x \\ r & x & 1 \end{vmatrix} = 0$$

soit :

$$-2 \begin{vmatrix} 1 & r \\ h + es & x \end{vmatrix} = -2(x - r(h + es)) = 0$$

dont la solution est :

$$x = r(h + es)$$

On déduit ensuite la valeur de y en remarquant que la corrélation de M avec $P = A_p + R_p$ doit être égale à r , soit :

$$xh + ye = r$$

L'application du critère équivalent d'indépendance conditionnelle est plus souple, et n'est pas limitée à des ensembles sans redondance. En supposant les variables centrées, on écrira ici :

$$\begin{aligned} E(MA_p) &= E(E(MA_p/P)) \\ &= E(E(M/P) \cdot E(A_p/P)) \end{aligned}$$

en appliquant successivement le théorème des espérances conditionnées et l'hypothèse d'indépendance conditionnelle. L'hypothèse de linéarité des régressions conduit alors à :

$$E(M/P) = rP(\text{Var}(M))^{1/2} (\text{Var}(P))^{-1/2}$$

et :

$$E(A_p/P) = (h + es) P(\text{Var}(A_p))^{1/2} (\text{Var}(P))^{-1/2}$$

d'où :

$$E(MA_p) = r(h + es) (\text{Var}(M))^{1/2} (\text{Var}(A_p))^{-1/2}$$

et donc :

$$x = r(h + es)$$

Le calcul de y se conduit de la même façon, il peut être allégé en posant $\text{Var}(P) = \text{Var}(M) = 1$:

$$\begin{aligned} E(MR_p) &= E(E(MR_p/P)) \\ &= E(E(M/P) \cdot E(R_p/P)) \\ &= E(rP \cdot e(hs + e)P) \\ &= re(hs + e) \end{aligned}$$

d'où :

$$y = r(hs + e)$$

En introduisant ensuite A_m , on détermine de la même façon la corrélation avec P (égale à x par symétrie) ; la corrélation entre A_m et A_p demande un conditionnement par le couple (P, M) de variables, on écrit :

$$\begin{aligned} E(A_p A_m) &= E(E(A_p A_m / P, M)) \\ &= E(E(A_p / P, M) \cdot E(A_m / P, M)) \end{aligned}$$

La valeur précédemment obtenue pour x conduit aux résultats :

$$E(A_p / P, M) = E(A_p / P) \quad \text{et} \quad E(A_m / P, M) = E(A_m / M)$$

puis à :

$$E(A_p A_m) = r(h + es)^2 h^2$$

c'est-à-dire à un coefficient de corrélation de :

$$r(h + es)^2$$

Enfin, l'introduction d'un descendant permet d'illustrer la prise en compte d'une hérédité génétique additive. Le principe d'indépendance conditionnelle n'est pas ici une hypothèse, mais l'expression statistique des lois de Mendel : quand les valeurs génétiques A_p et A_m des parents sont fixées, la valeur génétique A_d d'un descendant est indépendante de toute autre variable, et son espérance, conditionnée par A_p et A_m

est : $\frac{1}{2} A_p + \frac{1}{2} A_m$. L'usage de cette loi ne pose pas de problème quand les calculs

sont faits par la méthode des espérances conditionnelles ; en revanche, si l'on emploie

le critère du déterminant maximum, il faut noter que les coefficients $\frac{1}{2}$ sont des

coefficients de régression partiels, il faut les remplacer dans la matrice de corrélation

par les coefficients de corrélation simples, qui sont égaux à $\frac{1}{2}(1 + \rho)$; où ρ est le coefficient de corrélation entre A_p et A_m .

On obtient ainsi :

$$\begin{aligned} E(A_d P) &= E(E(A_d P / A_p, A_m)) \\ &= E\left(\left(\frac{1}{2} A_p + \frac{1}{2} A_m\right) \cdot E(P / A_p, A_m)\right) \end{aligned}$$

La régression double de P sur A_p et A_m se calcule selon les règles usuelles en tenant compte de la corrélation x entre P et A_m , et de la corrélation entre A_p et A_m précédemment calculée. Le calcul conduit à un coefficient de corrélation de :

$$\frac{1}{2} (1 + r) (h + es)$$

entre les variables A_d et P .

Les difficultés se présentent quand on introduit plusieurs variables simultanément, parce qu'elles lient un même individu à plusieurs variables déjà introduites. Je prends comme exemple le choix d'un conjoint apparenté, en présence d'homogamie. Le diagramme est celui de la figure 2, on va ignorer pour simplifier la corrélation entre génotype et milieu, soit : $s = 0$. (L'adoption d'une autre valeur de s ne complique ni ne simplifie le problème, il suffirait dans la suite de remplacer h par : $h' = h + es$.) Pour le couple (P, A_p) , la matrice de corrélations est :

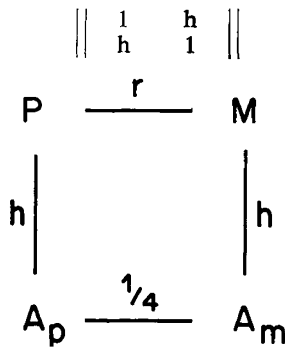


FIG. 2

Diagramme des corrélations fixées chez un couple de demi-germains, en présence d'homogamie

Diagram of the fixed correlations between half-sib mates, with assortative mating

Pour le conjoint, on a le choix entre M et A_m comme première variable à introduire ; en supposant que les deux individus sont demi-germains (corrélation de $1/4$ entre A_p et A_m), on obtient par application de la règle de calcul, selon le cas :

a) Introduction de M puis de A_m

La matrice de corrélations entre P, A_p et M s'écrit :

$$\begin{array}{|c|c|c|c|} \hline 1 & h & r & \\ \hline h & 1 & x & \\ \hline r & x & 1 & \\ \hline \end{array}$$

et le calcul conduit à la même valeur que précédemment :

$$x = rh$$

Désignant par z la corrélation recherchée entre P et A_m , la matrice entre les quatre variables est :

$$\begin{array}{|c|c|c|c|c|} \hline 1 & h & r & z & \\ \hline h & 1 & rh & \frac{1}{4} & \\ \hline r & rh & 1 & h & \\ \hline z & \frac{1}{4} & h & 1 & \\ \hline \end{array}$$

ce qui conduit à la solution :

$$z = \frac{\frac{1}{4} h (1 - r^2) - rh (1 - h^2)}{1 - r^2 h^2}$$

b) Introduction de A_m puis de M

Des calculs analogues conduisent successivement aux valeurs :

$$z' = \frac{1}{4} h$$

$$x' = \frac{\frac{1}{4} h (1 - h^2) + \frac{15}{16} rh}{1 - \frac{h^2}{16}}$$

Ces deux ensembles de solutions ne coïncident pas ; de plus la symétrie du diagramme est rompue, elle doit se traduire par l'existence d'une seule corrélation, x'' , entre P et A_m et entre M et A_p . Cela conduit naturellement à introduire ensemble les deux variables décrivant le conjoint, et à chercher la valeur x'' qui maximise le déterminant suivant :

$$\begin{vmatrix} 1 & h & r & x'' \\ h & 1 & x'' & \frac{1}{4} \\ r & x'' & 1 & h \\ x'' & \frac{1}{4} & h & 1 \end{vmatrix}$$

Les valeurs qui maximisent ce déterminant sont solutions de :

$$-4 \begin{vmatrix} h & r & x'' \\ 1 & x'' & \frac{1}{4} \\ x'' & 1 & h \end{vmatrix} = 0$$

qui conduit à l'équation :

$$(x'')^3 - x'' \left(1 + h^2 + \frac{r}{4} \right) + h \left(\frac{1}{4} + r \right) = 0$$

Pour r positif, cette équation a une racine négative inférieure à (-1) , sans signification ici ; selon les valeurs de r et de h , il y a une ou deux racines positives et inférieures à 1. Il y a donc toujours une seule valeur, la plus petite racine positive de l'équation, qui correspond à un maximum local du déterminant. Si l'on conduit le calcul selon le principe d'indépendance conditionnelle, en calculant l'espérance du produit ($A_m P$) en fixant les variables M et A_p , on obtient la même équation en x'' , le choix de la bonne racine parmi les trois possibles peut se faire en retenant celle qui rend la matrice définie positive.

Une difficulté analogue se présente si, dans le diagramme de la figure 1, on introduit un frère ($A_{q'}$, $R_{q'}$) du descendant déjà considéré et si l'on postule dans le modèle l'existence d'une corrélation c entre les effets de milieu au sein d'une même fratrie. Le problème ne se pose que si la corrélation s est un paramètre du modèle et est non nul (sinon les deux variables $A_{q'}$ et $R_{q'}$ peuvent être introduites dans un ordre indifférent).

Ces obstacles à une détermination explicite, par la voie algébrique, de certaines corrélations déduites ne sont pas nécessairement une gêne pour l'utilisation pratique de ce type de modélisation. En effet, d'une part, le calcul numérique des corrélations déduites à partir des paramètres du modèle n'est pas un problème difficile à résoudre. D'autre part l'estimation des paramètres du modèle peut être envisagée si l'on renforce l'hypothèse de linéarité des régressions par celle de normalité conjointe des distributions, car on peut alors utiliser l'hypothèse d'indépendance conditionnelle pour déterminer la vraisemblance d'un ensemble d'observations en fonction des seuls paramètres du modèle.

Reçu pour publication en octobre 1981.

Summary

Comments about proposals for the extension of the conventional additive genetic model to non-panmictic situations

Some results obtained by LANGLOIS (*Ann. Génét. Sél. anim.*, **13**, 151-163, 1981) are discussed, which deal with the expression of correlation between relatives when assortative mating and/or correlation between genotype and environment are taken into account. This note is mainly devoted to the exposition of some mathematical rules which are extensively used in the discussed paper, and which seem to be of quite general interest. Some examples of calculations are given, and some potential difficulties are outlined.

Références bibliographiques

- FISHER R.A., 1918. The correlation between relatives on the supposition of Mendelian inheritance. *Trans. r. Soc. Edinburgh*, **52**, 399-433.
- LANGLOIS B., 1981. Proposition d'extension du modèle génétique additif classique à des situations non panmictiques. *Ann. Génét. Sél. anim.*, **13**, 151-163.
- WRIGHT S., 1921. Systems of mating. III. Assortative mating based on somatic resemblance. *Genetics*, **6**, 144-161.