Ann. Génét. Séle. anim., 1976, **8** (2), 181-206.

ESTIMATION DES COMPOSANTES DE LA VARIANCE PHÉNOTYPIQUE DANS UNE POPULATION CONSANGUINE*

I. — Élaboration du modèle

C. CHEVALET

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

RÉSUMÉ

L'expression théorique des corrélations entre apparentés quelconques permet l'écriture d'un modèle statistique général, pour estimer les composantes de la variance phénotypique dans une population consanguine. A moins de supposer le caractère additif, il n'y a pas de solution statistique convenable sans une structuration précise du plan d'accouplements. La condition mathématique pour l'existence de résumés exhaustifs et indépendants en vue de l'estimation des effets fixes et des composantes de variance, est que les matrices de variances et covariances relatives à deux familles possibles de paramètres soient commutatives. Les modèles vérifiant cette condition sont dits stables. Pour un caractère quelconque, le schéma d'analyse de variance construit selon la hiérarchie père-mère-descendant peut être rendu partiellement stable en établissant des familles de père où les coefficients d'identité entre descendants ne dépendent que de l'indice du père ; pour un caractère additif le modèle est toujours stable mais dans tous les cas les calculs numériques ne sont guère possibles que si les conditions précédentes sont réalisées. Les estimateurs du maximum de vraisemblance, et les estimateurs quadratiques justes à norme minimale sont envisagés.

INTRODUCTION

Cette étude constitue le troisième volet d'une démarche méthodologique qui vise à fournir un outil d'analyse génétique des caractères quantitatifs dans les populations consanguines. Cette démarche s'appuie sur les travaux théoriques de Gillois (1964), qui a établi notamment les expressions générales des variances et covariances phénotypiques d'individus apparentés quelconques, en admettant des interactions de dominance et l'existence d'effets maternels génétiques.

(*) Ce travail fait partie d'une thèse de Doctorat ès-sciences naturelles, soutenue à l'Université de Paris-Sud (Orsay), le 6 mai 1975.

Les possibilités d'une analyse statistique fondée sur ces résultats sont présentées ici, elles s'appuient aussi sur deux études précédentes qui complètent les travaux de GILLOIS. Dans la première (CHEVALET, 1971 *a*), nous avons exposé une méthode de calcul automatique des coefficients d'identité utilisable à partir des fichiers généalogiques constitués dans les élevages. Dans la seconde (CHEVALET, 1971 *b*), nous avons précisé les conditions dans lesquelles l'analyse de GILLOIS pouvait aboutir à une expression des corrélations génétiques entre individus d'une même lignée consanguine. Les formules qu'il a données supposent en effet, pour entreprendre une analyse statistique, que plusieurs lignées présentant la même généalogie soient constituées à partir d'une même population de base, ce qui représente en général une impossibilité pratique dans des populations d'animaux domestiques.

Enfin nous présenterons dans une étude ultérieure une illustration de l'ensemble de cette démarche, qui précisera les problèmes pratiques liés au calcul numérique, et à l'évaluation de la précision statistique des estimations, dans un contexte zootechnique.

En Génétique quantitative, les gènes intervenant dans l'expression d'un caractère sont généralement inconnus, et le mode de l'hérédité est caractérisé par une analyse statistique. Celle-ci repose usuellement sur l'écriture d'un modèle logique linéaire, dans lequel le phénotype d'un individu se décompose en une somme d'effets liés à des causes distinctes : le type d'élevage, le numéro de portée, l'effet aléatoire de l'environnement, l'effet direct du père, l'effet direct de la mère, l'interaction entre ces effets, etc. Cependant, l'interprétation génétique des effets introduits, et la justification des propriétés statistiques imposées au modèle requièrent des hypothèses restrictives sur la structure de la population étudiée. La définition d'une variable aléatoire « effet maternel direct », qui prenne pour chaque mère des valeurs indépendantes, ne concorde avec la théorie génétique que si les mères des individus observés sont tirées au hasard dans une population panmictique infinie ; l'hypothèse de normalité d'une variable d'interaction n'a guère de sens biologique.

Pour ces raisons, il ne nous a pas paru logique, pour entreprendre l'analyse statistique d'un caractère quantitatif dans une population consanguine, de poser le modèle linéaire habituel, et d'y introduire ensuite les modifications nécessaires pour tenir compte des nouveaux types de corrélation induits par la consanguinité.

Au contraire, nous nous sommes appuyés sur les formules théoriques de la Génétique quantitative où seules apparaissent les variables aléatoires génotypiques. Nous écrivons ensuite un modèle statistique pour estimer les différentes composantes de variance introduites par le modèle génétique, sans faire appel à de nouvelles hypothèses. Ainsi la condition de normalité, quand nous la supposons respectée, concerne seulement la variable aléatoire phénotypique, ce qui est justifié dans le cadre du modèle classique de FISHER où l'on admet que les gènes contribuent chacun pour une petite part à la variabilité totale du caractère quantitatif.

Après avoir écrit la transposition statistique du modèle génétique, où nous admettons *a priori* l'existence d'interactions de dominance, nous avons recherché dans quelle mesure les méthodes générales d'estimation, mises au point pour le modèle linéaire d'analyse de variance, pouvaient être transposées au cas étudié ici. Les estimations fondées sur le principe du maximum de vraisemblance, et les estimations quadratiques justes à variance minimum des composantes de la variance ont été envisagées, en s'attachant particulièrement aux cas qui ne font pas appel

à des méthodes numériques trop lourdes. Ainsi suggérons-nous une modification du schéma hiérarchique d'analyse de variance pour l'étude des populations d'effectif génétique restreint où les accouplements peuvent être maîtrisés.

I. — CONSTRUCTION DU MODÈLE STATISTIQUE

A. — *Modèle génétique*

On suppose que le caractère quantitatif est le fait d'un effet aléatoire d'environnement e, commun à tous les individus, d'espérance E (e) et de variance var (e) et de la somme des effets de L couples de gènes, situés en L loci, ne présentant pas d'interactions épistatiques mais pouvant présenter des interactions de dominance. Pour l'individus (i), le phénotype s'écrit :

$$P_i = e_i + \sum_{x=1}^{x=L} (X_i^x + X'^x_i + D_i^x)$$

où X_i^x, X'^x_i sont les effets additifs des gènes au locus x et D_i^x le résidu de dominance au locus x. On note $Y_i^x = X_i^x + X'^x_i$, et $Z_i^x = Y_i^x + D_i^x$. On supposera de plus que la dispersion des effets des différents loci est petite : les formules approchées des variances et covariances peuvent être utilisées, et la fraction génétique du phénotype suit une loi proche de la loi normale.

Les moments d'ordres 1 et 2 des phénotypes sont alors :

$$E(P_i) = E(e) + (1 - f_i)E_p(Z) + f_i E_c(D) + \sigma \cdot n_i.$$

$$\text{var}\,(P_i) = \text{var}\,(e) + (1 + f_i)\,\text{var}_p\,(Y) + (1 - f_i)\,\text{var}_p\,(D) + 4 f_i\,\text{cov}_c\,(XD) + f_i\,\text{var}_c\,(D) + \sigma' n'_i.$$

$$\text{cov}\,(P_i P_j) = 2\varphi_{ij}\,\text{var}_p\,(Y) + (\delta_9^{ij} + \delta_{12}^{ij})\,\text{var}_p\,(D) + (4\delta_1^{ij} + \delta_2^{ij} + \delta_3^{ij} + \delta_4^{ij} + \delta_5^{ij})\,\text{cov}_c\,(XD) + \delta_1^{ij}\,\text{var}_c\,(D) + \sigma'' \cdot n_{ij}.$$

En adoptant les notations de GILLOIS (1964).

n_i, n_i', n_{ij} sont des réalisations de variables aléatoires de loi proche de N(0, 1) ; σ, σ' et σ'' sont des quantités petites en l'absence de gènes majeurs (CHEVALET, 1971).

f_i, φ_{ij}, δ_k^{ij} sont respectivement : le coefficient de consanguinité de l'individu (i), le coefficient de parenté du couple (i, j), au sens de MALÉCOT (1948) et le k^{ie} coefficient d'identité du couple (i, j) (GILLOIS, 1964).

B. — *Modèle statistique*

Le modèle statistique déduit de ces formules est le suivant :

Étant donné N individus d'une population, les phénotypes observés p_1, p_2,..., p_N sont des réalisations uniques de N variables aléatoires P_1, P_2,..., P_N dont les moments d'ordres 1 et 2 s'écrivent :

$$\begin{aligned} E(P_i) &= u + f_i v \\ \text{var}\,(P_i) &= \theta_1 + d_2^{ii}\theta_2 + d_3^{ii}\theta_3 + d_4^{ii}\theta_4 + d_5^{ii}\theta_5 \\ \text{cov}\,(P_i P_j) &= d_2^{ij}\theta_2 + d_3^{ij}\theta_3 + d_4^{ij}\theta_4 + d_5^{ij}\theta_5 \end{aligned}$$

En notations matricelles :

$$E(P) = uJ + v\,F.$$

$$\text{cov}\,(PP') = \sum_{k=1}^{5} d_k \theta_k$$

où d_1^{ij}, d_2^{ij},..., d_5^{ij} sont les fonctions de coefficients d'identité introduites dans les formules génétiques, f_i est toujours le coefficient de consanguinité u, v, θ_1 à θ_5 sont les paramètres que l'on se propose d'estimer.

On supposera de plus, pour traiter de l'efficacité des estimateurs que les N variables P_i suivent une loi multinormale : cette hypothèse est approximativement vérifiée dans le modèle génétique initial.

A partir de modèles génétiques plus complexes, on peut faire la même démarche, et construire un modèle statistique qui formalise avec des concepts mathématiques simples une structure aussi proche que possible de la structure génétique vraie. Tous les modèles statistiques de ce type, établis à partir d'une étude théorique fondée sur la relation d'identité présentent la même forme générale.

Un tel modèle statistique introduit les différentes composantes qui apparaissent dans l'analyse génétique. Mais les formules génétiques de référence sont approchées, et il convient pour chacun des termes, de discuter du degré d'approximation. En effet l'écart-type de l'erreur relative faite en adoptant comme multiplicateur d'une composante θ_k (telle que var_c (D)) le coefficient d'identité correspondant $d(\delta_1$ pour var_c (D)) est :

$$((1 - d)(1 + R^2)/dL)^{1/2}$$

où L est le nombre de loci contribuant au caractère, et R^2 la dispersion des effets de ces loci relativement à la composante θ_k : l'approximation est satisfaisante si cet écart-type est petit devant l'unité, donc si la quantité $(Ld)^{1/2}$ est grande. Un terme génétique pour lequel cette condition n'est pas satisfaite ne doit pas être pris en compte dans le modèle statistique. La variabilité génétique associée se répartit entre les autres composantes estimées.

C. — *Remarque*

Si l'on envisage la possibilité de répliquer M fois une généalogie donnée, on obtient M groupes équivalents de N observations :

$$\begin{array}{llll} p_1^1 & p_2^1 & \cdots\cdots\cdots\cdots\cdots & p_N^1 \\ p_1^2 & p_2^2 & \cdots\cdots\cdots\cdots\cdots & p_N^2 \\ p_1^M & p_2^M & \cdots\cdots\cdots\cdots\cdots & p_N^M \end{array}$$

Il faut considérer chaque N-vecteur

$$p_1^l \; p_2^l \cdots\cdots\cdots\cdots\cdots\cdots p_N^l$$

comme une réalisation du vecteur aléatoire d'ordre N

$$P_1 \; P_2 \cdots\cdots\cdots\cdots\cdots\cdots P_N$$

où les aléatoires P_i ont ensemble une loi caractérisée par les formules exactes de GILLOIS.

La quantité théorique cov (P_iP_j) se rapporte à la covariance expérimentale des séries de mesures p_i^l et p_j^l :

$$\text{cov}\,(P_iP_j) = E\left(\frac{M-1}{1}\sum_l p_i^l p_j^l - \frac{M(M-1)}{1}\left(\sum_l P_i^l\right)\left(\sum_m p_j^m\right)\right)$$

mais ne peut être rapportée à une moyenne de produits d'observations tirés d'une même réplication même si les couples correspondants présentent les mêmes fonctions d'identité. On a affaire à un modèle statistique différent de celui qui est étudié dans la suite.

II. — MÉTHODES GÉNÉRALES D'ESTIMATION

Le modèle statistique défini en I-B n'a pas été étudié en tant que tel, dans toute sa généralité. Cependant diverses études concernent des modèles statistiques qui présentent des caractères communs avec lui, et donnent des méthodes numériques utilisables, quoique les hypothèses de base diffèrent (SEARLE, 1971). En particulier les estimations du maximum de vraisemblance dans l'hypothèse de normalité (HARTLEY et RAO, 1967 ; SEARLE, 1970), et des estimations quadratiques à variance ou norme minimale (RAO,1971), ont été envisagées pour un modèle général d'analyse de variance à effets fixes et aléatoires, dont la matrice de variances et covariances présente la même structure que celle du modèle étudié ici. Ces deux points de vue sont abordés.

A. — *Définitions préliminaires*

A_1 *Changement des paramètres à estimer.*

Les propriétés de la matrice V de variances et covariances dépendent des rapports entre les composantes θ_k, et non de leurs valeurs absolues. Choisissant arbitrairement l'une d'elles, soit θ_1, on posera :

$$t = \theta_1 \; ; \; r_k = \theta_k/\theta_1 \; ; \; t > 0$$

En prenant pour θ_1 une composante correspondant à un paramètre génétique positif, comme la variance des effets additifs des gènes, ou la variance des effets du milieu.

θ désignera le vecteur des composantes θ_k, r le vecteur des rapports r_k. On notera de façons équivalentes :

$$V(\theta) = V(t, r) = \sum_k d_k\theta_k = t\sum_k d_k r_k = tT(r).$$

A_2 *Transformation canonique des observations, stabilité.*

Pour la valeur vraie r_0 des paramètres, la matrice $V(t, r_0)$ est définie positive, et le demeure par continuité dans un voisinage de r_0. Dans ce voisinage la matrice $V(\theta)$ peut être diagonalisée : soient $w_i(\theta)$ les valeurs propres, de multiplicités $n_i (i = 1; s)$, soit $U(r)$ une matrice orthogonale dont les lignes $U'_{ij} (i = 1, s ; j = 1, n_i)$ sont des

vecteurs propres normés, deux à deux orthogonaux, associés aux valeurs propres $w_i(\theta)$. La matrice diagonale W (θ) de ces valeurs propres s'écrit :

$$W(\theta) = U(r) \cdot V(\theta) \cdot U'(r) \text{ (*)}.$$

Soit Y = UP le vecteur des variables transformées,

$$Y_{ij} = U'_{ij}P \text{ l'une d'elles.}$$

Y_{ij} est une variable aléatoire d'espérance :

$$E(Y_{ij}) = uU'_{ij}J + vU'_{ij}F = ua_{ij} + vb_{ij}$$

Si le calcul est fait pour *la valeur vraie* r_0, les aléatoires Y_{ij} sont non corrélées et de variances :

$$\text{var } (Y_{ij}) = w_i(\theta_0)$$

En général, la transformation U dépend du paramètre r, et les variables transformées Y_i sont corrélées :

$$\text{cov } (YY') = U(r)U'(r_0)W(\theta_0)U(r_0)U'(r)$$

et la condition nécessaire et suffisante pour que cette matrice soit diagonale est que le produit des matrices $V(\theta_0)$ et $V(\theta)$ soit commutatif. La transformation U est alors indépendante de r et associée à une base orthonomée de vecteurs propres communs aux matrices $V(\theta)$ et $V(\theta_0)$.

Si cette condition de commutativité est vérifiée pour toute valeur θ différente de θ_0 on dira que le modèle est *stable.*

Un modèle stable, réduit à sa forme canonique, s'énonce : les phénotypes transformés observés $y_{ij} = U'_{ij}\, p$ sont des réalisations des N variables aléatoires Y_{ij} dont les moments d'ordres un et deux s'écrivent :

$$\begin{aligned} E(Y_{ij}) &= ua_{ij} + vb_{ij} \\ \text{var } (Y_{ij}) &= \sum_k c_{ik}\theta_k \qquad \text{pour tout } j = 1, n_i \\ \text{cov } (Y_{ij}Y_{i'j'}) &= 0 \qquad \text{pour } (i, j) \neq (i', j'). \end{aligned}$$

Les quantités a_{ij}, b_{ij}, c_{ik} étant intrinsèques au modèle initial.

B. — *Estimations du maximum de vraisemblance*

On suppose ici les phénotypes distribués selon la loi normale à N variables. Si $|\ V\ |$ désigne le déterminant de la matrice V, la vraisemblance d'une observation p de l'aléatoire P est :

$$L(p/u, v, \theta) = (2\pi)^{-N/2}\ |\ V(\theta)\ |^{-1/2} \exp\ (-(p' - uJ' - vF')V(\theta)^{-1}(p - uJ - vF)/2)$$

B_1 *Moments du premier ordre.*

Les équations du maximum de vraisemblance relatives aux paramètres u et v sont :

$$\begin{aligned} \delta \text{ Log } L/\delta u &= p'V^{-1}J - u \cdot J'V^{-1}J - v \cdot J'V^{-1}F = 0 \\ \delta \text{ Log } L/\delta v &= p'V^{-1}F - u \cdot J'V^{-1}F - v \cdot F'V^{-1}F = 0 \end{aligned}$$

où $V = V(\theta)$. Les estimateurs $\hat{u}(p/r)$ et $\hat{v}(p/r)$ qui en résultent dépendent du paramètre r adopté, et non de t ; ils sont justes quel que soit r.

(*) U' désigne la matrice transposée de la matrice U.

B_2 *Moments d'ordre deux.*

Le système d'équation :

$$\frac{1}{L}\frac{\delta L}{\delta \theta_k} = 0$$

se ramène à un système de 5 équations algébriques homogènes de degré N-1 aux 5 inconnues θ_k, qui ne peut être résolu explicitement. Une résolution numérique est nécessaire. Au contraire, si le paramètre r est connu ($r = r_0$), l'équation du maximum de vraisemblance relative à t est :

$$\delta \log L/\delta t = -N/(2t) + (p' - uJ' - vF')T(r_0)^{-1}(p - uJ - vF)/(2t^2) = 0$$

qui donne l'estimation :

$$\hat{t}(p/u, v, r_0) = (p' - uJ' - vF')T(r_0)^{-1}(p - uJ - vF)/N$$

Si u et v sont connus, $\hat{t}$ est juste et $N\hat{t}/t_0$ suit la loi χ^2 (N). Sinon les trois équations :

$$\delta \log L/\delta u = \delta \log L/\delta v = \delta \log L/\delta t = 0$$

donnent immédiatement l'estimation simultanée de u, v, t (RAO, 1965), en remplaçant $p - uJ - vF$ par la quantité $q = p - \hat{u}J - \hat{v}F$.

Si le paramètre r n'est pas connu, la fonction de vraisemblance L, prend sa valeur maximale quand u, v, t et r sont liés par les trois équations précédentes. L, est alors une fonction de r seulement, qui s'écrit :

$$L_1(p/r) = L(p/\hat{u}(p/r), \hat{v}(p/r), \hat{t}(p/\hat{u}, \hat{v}, r), r) = (2\pi e)^{-N/2}(q'T(r)^{-1}q/N)^{-N/2} \mid T(r) \mid^{-1/2}.$$

La solution des équations du maximum de vraisemblance est donnée par la valeur $\hat{r}$ qui maximise la fonction L_1, et par les valeurs $\hat{u}(p/\hat{r})$, $\hat{v}(p/\hat{r})$, $\hat{t}(p/\hat{u}, \hat{v}, \hat{r})$ pour les paramètres u, v, t, respectivement.

HARTLEY et RAO (1967) ont proposé une méthode itérative pour calculer les solutions $\hat{r}_k$ qui maximisent cette fonction. Cette résolution peut se faire aussi directement par la méthode des relaxations successives. Dans ces deux cas la difficulté numérique majeure est d'inverser la matrice $T(r)$, à moins que la stabilité ne conduise à des calculs très simples : dans ce cas $T(r)$ est diagonale.

B_3 *Convergence des estimations.*

Dans les schémas usuels d'estimation, on a N observations (p_1, p_2, ..., p_N) qui sont N réalisations de N aléatoires (P_1, P_2, ..., P_N) indépendantes et équidistribuées. La vraisemblance $L_N(p_1, p_2, ..., p_N/\theta)$ est alors le produit des vraisemblances $L_i(p_i/\theta)$, et

$$\log L_N = \sum_{i=1}^{N} \log L_i$$

On démontre la convergence de l'estimation du maximum de vraisemblance en s'appuyant sur la loi des grands nombres. Dans le schéma étudié ici, on ne peut se ramener à des variables indépendantes que si le schéma est stable, mais les variables ne sont pas équidistribuées. De plus l'introduction d'une nouvelle observation modifie l'ensemble de ces variables et leurs distributions. On ne peut donc définir de convergence au sens usuel ; il est nécessaire de considérer une suite S_1, S_2, ..., S_k, ...

de schémas expérimentaux, impliquant N_1, N_2, ..., N_k, observations. Si $\hat{\theta}(k)$ est l'estimation du paramètre θ sur le schéma S_k, on dit que $\theta(k)$ tend en probabilité vers θ si, $\eta > 0$ étant fixé, la probabilité

$$Pr(\| \hat{\theta}(k) - \theta \| > \eta)$$

tend vers 0 quand k tend vers l'infini. HARTLEY et RAO (1967) ont donné des conditions suffisantes pour qu'une telle convergence ait lieu, pour le modèle général d'analyse de variance suivant :

$$p = X\alpha + U_1 b_1 + \ldots + U_c b_c + e$$

où X est une matrice $(n \times k)$ de nombres connus, $k < n$
U_i est une matrice $(n \times m_i)$ de nombres connus, $m_i < n$
α est un vecteur $(k \times l)$ de constantes inconnues
b_i est un vecteur $(m_i \times l)$ de variables indépendantes de loi $N(0, r_i t)$
e est un vecteur $(n \times l)$ de variables indépendantes de loi $N(0,t)$.

La matrice des variances et covariances s'écrit :

$$\text{cov}(PP') = t(I + r_1 U_1 U_1' + r_2 U_2 U_2' + \cdots + r_c U_c U_c') = V(t, r)$$

Cependant le théorème de convergence s'appuie implicitement sur l'existence d'une transformation orthogonale unique ramenant les deux matrices $V(t, r)$ et $V(t, r_0)$ à une forme diagonale : cela nécessite la commutativité du produit de ces matrices pour tout $r \neq r_0$ et par conséquent la commutativité entre les matrices $U_i U_i'$:

$$U_i U_i' \, U_j U_j' = U_j U_j' \, U_i U_i' \qquad \text{pour } i \neq j.$$

Cette propriété ne résulte pas des hypothèses faites sur les matrices U_i, ainsi que le montre l'exemple suivant :

$$\begin{bmatrix} P_1 \\ P_2 \\ P_3 \\ P_4 \\ P_5 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \alpha + \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} b_{11} \\ b_{12} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} b_{21} \\ b_{22} \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{bmatrix}$$

L'hypothèse supplémentaire de stabilité des modèles doit donc être faite pour que la démonstration de convergence soit valable.

On se limitera dans la suite à la détermination d'intervalles de confiance, approchés, relatifs à l'estimation d'un ou plusieurs paramètres pour un schéma expérimental S donné, stable, et ne faisant intervenir que deux composantes de la variance. Dans les autres cas le principe du maximum de vraisemblance conduit à des calculs très lourds, et à des estimations dont la précision ne peut être quantifiée pour un schéma expérimental donné.

C. — *Estimations quadratiques justes*

Un estimateur quadratique d'une composante θ_k est une variable aléatoire de la forme :

$$\hat{\theta}_k = P' A_k P$$

où A_k est une matrice carrée symétrique (N × N). Son espérance est :

$$E(\hat{\theta}_k) = (uv)(JF)'A_k(JF)(uv)' + \text{trace } A_kV(\theta_o).$$

En se limitant aux estimateurs « invariants », tels que :

$$A_k(JF) = o,$$

l'espérance se ramène à

$$E(\hat{\theta}_k) = \text{trace } A_kV(\theta_c) = \sum_l \theta_{ol} \text{ trace } A_kd_l$$

L'estimateur est juste si :

$$\begin{aligned} \text{trace } A_kd_l &= 1 \text{ pour } l = k \\ &= o \text{ pour } l \neq k \end{aligned}$$

Si les coefficients de kurtosis des distributions sont nuls, la variance de l'estimateur invariant juste est :

$$\text{var } (\hat{\theta}_k) = 2 \text{ trace } A_kV(\theta_o)A_kV(\theta_o).$$

D'une façon générale si $\bar{\theta}$ désigne une valeur arbitraire du paramètre θ telle que $V(\bar{\theta})$ soit définie positive, la forme bilinéaire :

$$\psi(A, B) = \text{trace } AV(\bar{\theta})BV(\bar{\theta})$$

définit une norme dans l'espace des matrices carrées symétriques (N × N), et induit donc une norme dans l'espace des estimateurs quadratiques invariants. (RAO, 1971 *a*, *b*). Dans le cas du modèle général d'analyse de variance énoncé en II — B_3, (RAO, 1971 *a*) a établi l'existence d'un estimateur à norme minimale et donné le moyen de sa détermination numérique. Ces méthodes peuvent être transposées sans difficulté au modèle statistique étudié ici :

Un paramètre $\bar{\theta}$ étant choisi, la fonction :

$$A_k \longrightarrow \text{trace } A_kV(\bar{\theta})A_kV(\bar{\theta}),$$

définie sur l'ensemble des estimateurs invariants justes de la composante θ_k, prend sa valeur minimum en $A_k(\bar{\theta})$, qui définit l'estimateur à norme $(\bar{\theta})$ minimum. Cette norme, cependant, ne représente une entité statistique que si $\bar{\theta} = \theta_o$ (c'est alors la variance si les distributions sont normales). L'intérêt de cet estimateur dépend donc des deux critères suivants, au point θ_o.

1° la valeur absolue de la norme minimum basée sur θ_o :

$$\text{trace } A_k(\theta_o)V(\theta_o)A_k(\theta_o)V(\theta_o),$$

qui caractérise l'efficacité théorique du schéma expérimental :

2° le comportement de la fonction :

$$\bar{\theta} \longrightarrow \text{trace } A_k(\bar{\theta})V(\theta_o)A_k(\bar{\theta})V(\theta_o)$$

quand $\bar{\theta}$ parcourt un voisinage de θ_o :

si cette fonction admet de grandes variations quand $\bar{\theta}$ s'écarte légèrement de θ_o, la minimisation de la norme relative à $\bar{\theta}$ perd tout intérêt pratique.

III. — MODÈLES STABLES

La propriété fondamentale des modèles stables est l'existence de résumés exhaustifs pour l'estimation des effets fixes u et v, et des composantes θ_k, qui sont respectivement des formes linéaires et quadratiques des observations P_i (dans l'hypothèse de normalité).

On suppose réalisée la transformation par U, c'est-à-dire le modèle réduit à sa forme canonique : les observations transformées $y_{ij} = U'_{ij}p$ sont des réalisations des variables aléatoires Y_{ij} indépendantes et de lois $N(ua_{ij} + vb_{ij}, w_i(\theta))$ où $w_i(\theta)$ est une fonction linéaire des composantes θ_k.

A. — *Expression de la vraisemblance, exhaustivité*

La vraisemblance de l'épreuve y s'écrit :

$$L(y/u, v, \theta) = (2\pi)^{-N/2} \prod_i (w_i(\theta))^{-n_i/2} \exp \{ - \sum_i \sum_j (y_{ij} - ua_{ij} - vb_{ij})^2/(2w_i(\theta) \}$$

Elle est le produit des vraisemblances $L_i(y_{i.}/u, v, \theta)$ attachées aux épreuves partielles

$$y_{i.} = \{y_{ij}, j = 1, n_i\}$$

Si la matrice des coefficients :

$$\begin{bmatrix} a_{i1}\ a_{i2} \cdots \cdots \cdots \cdots \cdots \cdots a_{in_i} \\ b_{i1}\ b_{j2} \cdots \cdots \cdots \cdots \cdots \cdots b_{in_i} \end{bmatrix}$$

est de rang 2, les observations $y_{ij}(j = 1, n_i)$ procurent des estimations partielles de u et v, par le principe du maximum de vraisemblance. Les estimateurs u_i et v_i sont donnés par le système :

$$\begin{cases} \sum_j a_{ij}Y_{ij} = u_i \sum_j a_{ij}^2 + v_i \sum_j \Sigma a_{ij}b_{ij} \\ \sum_i b_{ij}Y_{ij} = u_i \sum_i a_{ij}b_{ij} + v_i \sum_i b_{ij}^2 \end{cases}$$

En notant :

$$AA_i = \sum_j a_{ij}^2 \ ; \ AB_i = \sum_j a_{ij}b_{ij} \ ; \ BB_i = \sum_j b_{ij}^2, \text{ et : } D_i = AA_i BB_i - (AB_i)^2,$$

$$X_i = \sum_i Y_{ij}^2 - (AA_i u_i^2 + 2AB_i u_i v_i + BB_i v_i^2), \text{ et :}$$

$$Q_i = AA_i(u_i - u)^2 + 2\ AB_i(u_i - u)(v_i - v) + B_i(v_i - v)^2,$$

on a l'identité algébrique suivante :

$$\sum_j (Y_{ij} - ua_{ij} - vb_{ij})^2 \equiv X_i + Q_i \qquad (1)$$

Il résulte alors de la définition des quantités u_i et v_i, et de l'indépendance des aléatoires $(Y_{ij} - ua_{ij} - vb_{ij})$ de même loi N (o, $w_i(\theta)$), que les formes quadratiques $Q_i/w_i(\theta)$ et $X_i/w_i(\theta)$ sont des aléatoires indépendantes de lois respectives χ^2 (2) et $\chi^2(n_i - 2)$.

En effet, la forme quadratique $Q_i/w_i(\theta)$ est définie par la matrice inverse de la matrice de variances et de covariance des formes linéaires $u_i - u$ et $v_i - v$: elle suit donc une loi χ^2 (2), et peut s'écrire comme la somme de deux carrés de variables N (0,1) indépendantes :

$$\frac{Q_i}{w_i(\theta)} = \frac{z^2_{i1}}{w_i(\theta)} + \frac{z^2_{i2}}{w_i(\theta)} \qquad (2)$$

z_{i1} et z_{i2} sont deux combinaisons linéaires en u_i et v_i qui peuvent s'écrire :

$$z_{i1} = \sum_j Z_{i1j}(Y_{ij} - ua_{ij} - vb_{ij})$$
$$z_{i2} = \sum_j Z_{i2j}(Y_{ij} - ua_{ij} - vb_{ij})$$

Z_{i1} et Z_{i2} étant deux vecteurs $(n_i \times 1)$ normés et orthogonaux dont les composantes Z_{i1j} et Z_{i2j} sont fonctions seulement des coefficients a_{ij} et b_{ij}. Ces deux vecteurs peuvent être complétés par $n_i - 2$ vecteurs $Z_{ik}(k > 2)$ pour former une base orthogonale de dimension n_i. Les n_i aléatoires z_{ik} construites comme z_{i1} et z_{i2} sont encore indépendantes et de même loi $N(0, w_i(\theta))$. Compte tenu de la relation d'orthogonalité :

$$\sum_j (Y_{ij} - ua_{ij} - vb_{ij})^2 = \sum_j z^2_{ij},$$

et des relations (1) et (2), il vient :

$$X_i = \sum_{j>2} z^2_{ji}.$$

La distribution conjointe de X_i et Q_i en résulte, et la loi conjointe des variables aléatoires u_i, v_i, et X_i a pour densité de probabilité :

$$(2w_i(\theta))^{-1}D_i^{1/2} \exp\tfrac{1}{i}\left(-\frac{Q_i}{2w_i(\theta)}\right) \cdot \Gamma(k_i)(2w_i(\theta))^{-k_i} \exp\left(-\frac{X_i}{2w_i(\theta)}\right)X_i^{k_i-1}$$
$$= \pi^{k_i}\Gamma(k_i)D_i^{1/2}X_i^{k_i-1} \cdot L_i(Y_{i.}/u, v, \theta)$$

en posant :

$$k_i = \frac{n_i}{2} - 1.$$

L'ensemble des statistiques $(u_i, v_i, X_i, i = 1, I)$ constituent donc des résumés exhaustifs pour l'estimation des paramètres u, v, et θ.

Si la matrice des coefficients (a_{ij}, b_{ij}) est de rang 1, on aura de même une décomposition de la forme :

$$\chi^2(n_i) = \chi^2(n_i - 1) + \chi^2(1) \qquad (1)$$

Si le rang est 0 ($a_{ij} = b_{ij} = 0$ pour $j = 1, n_i$), il est immédiat que $X_i = \sum_j Y^2_{ij}$ suit une loi $w_i(\theta) \cdot \chi^2(n_i)$.

De façon générale, on désignera par m_i le nombre de degrés de liberté de la variable X_i : m_i est égal à $n_i - 1$ ou à $n_i - 2$.

B. — *Estimations par le maximum de vraisemblance*

La stabilité du modèle apporte toujours des simplifications des calculs numériques, mais ne permet pas de donner des solutions explicites aux équations du maximum de vraisemblance. Cependant l'efficacité des estimations peut être déterminée,

pour un schéma expérimental donné, si la matrice des covariances entre phénotypes P_i a la forme suivante :

$$V(t, r) = t(rI + \varnothing)$$

C'est le cas si les rapports θ_k/θ_2 entre les composantes génétiques de la variance sont connus, et donc, en particulier, si le caractère génétique est additif ($\theta_k = 0$ pour $k > 2$).

Aucune condition d'équilibre des données n'est requis.

Une modification de la méthode 3 d'HENDERSON (1953) a été proposée pour un modèle de ce type, qui ne fait intervenir que deux composantes de variance (CUNNINGHAM et HENDERSON, 1968 ; THOMPSON, 1969) en utilisant une estimation préalable du rapport r et en recourant à une itération.

B_1. *Caractère génétique additif : estimation de l'hérédité.*

Un seul rapport $r = r_1 = \theta_1/\theta_2$ est inconnu ; il se rattache à l'héritabilité h^2 du caractère par :

$$h^2 = 1/(1 + r)$$

La matrice $\varnothing$ est la matrice des doubles coefficients de parenté entre les individus observés dans l'expérience, c'est une matrice définie positive ; soient $(c_i, n_i ; i = 1, s)$ ses valeurs propres et leurs multiplicités, U la matrice orthogonale de passage vers une base de vecteurs propres de $\varnothing$, et donc de V ; $Y_{ij} = U'_{ij}P$ les phénotypes transformés.

Le caractère étant supposé additif, on peut poser :

$$v = 0$$

et les aléatoires Y_{ij} sont indépendantes et de lois

$$N(u_0 a_{ij}, t_0(r_0 + c_i))$$

où u_0, t_0, r_0 désignent les valeurs vraies des paramètres.

On suppose les valeurs propres rangées dans l'ordre tel que :

$$AA_i > 0 \text{ pour } i \leqslant s_1,\ AA_i = 0 \text{ pour } i > s_1\ (s_1 \leqslant s)$$

Les résumés exhaustifs, et indépendants, sont :

$$u_i = AA_i^{-1} \sum_j a_{ij} Y_{ij}, \text{ de loi } N\ (u_0,\ AA_i^{-1} t_0(r_0 + c_i))\ (i \leqslant s_1)$$

$$X_i = \sum_j Y_{ij}^2 - AA u_i^2, \text{ de loi } t_0(r_0 + c_i) \cdot \chi^2(m_i)$$

en posant $m_i = n_i - 1$ pour $i \leqslant s_1$, $m_i = n_i$ pour $i > s_1$, $M = \sum_i m_i$

La vraisemblance devient :

$$\text{Log } L(Y/u, t, r) = -\frac{N}{2} \log (2\pi t) - \sum_i (n_i/2) \log (r + c_i)$$
$$- \sum_{i \leqslant s_1} (X_i + AA_i(u_i - u)^2)/(2t(r + c_i)) - \sum_{i > s_1} X_i/(2t(r + c_i))$$

et l'estimateur conjoint de (u, t, r) qui maximise cette vraisemblance est donnée par le système :

$$\begin{cases} \hat{r} \text{ maximise la fonction :} \\ r \longrightarrow L_1(Y, r) = (2\pi e)^{-N/2} \prod_i (r + c_i)^{-n_i/2} (T(Y, r))^{-N/2} \\ \hat{u} = A(Y, \hat{r}) \\ \hat{t} = T(Y, \hat{r}) \end{cases}$$

où les fonctions $A(Y, r)$ et $T(Y, r)$ sont définies par :

$$A(Y, r) = \left(\sum_{i \leqslant s_1} u_i AA_i (r + c_i)^{-1} \right) \Big/ \left(\sum_{i \leqslant s_1} AA_i (r + c_i)^{-1} \right)$$

$$T(Y, r) = \frac{1}{N} \left\{ \sum_{i \leqslant s_1} (X_i + AA_i (u_i - A(Y, r))^2) / (r + c_i) + \sum_{i > s_1} X_i / (r + c_i) \right\}.$$

Les lois de $A(Y, r)$ et de $T(Y, r)$, conditionnées par une valeur donnée de r peuvent être explicitées :

$A(Y, r)$ est normale :

$$N \left\{ u_0, \left(\sum_{i \leqslant s_1} AA_i (r_0 + c_i)(r + c_i)^{-2} \right) \cdot \left(\sum_{i \leqslant s_1} AA_i (r + c_i)^{-1} \right)^{-2} \right\}$$

$T(Y, r)$ est une forme quadratique, formée de la somme pondérée de s variables χ^2 indépendantes, et d'une forme quadratique indépendante des variables χ^2. En effet :

$$t_0^{-1} N T(Y, r) = \sum_i \frac{X_i}{t_0 (r_0 + c_i)} \cdot \frac{r_0 + c_i}{r + c_i} + \sum_{i \leqslant s_1} \frac{AA_i (u_i - A(Y, r))^2}{t_0 (r_0 + c_i)} \frac{r_0 + c_i}{r + c_i}$$

Dans la première somme les variables $Z_i = X_i / (t_0 (r_0 + c_i))$ sont de lois $\chi^2(m_i)$ indépendantes ; et la deuxième somme, indépendante des variables Z_i puisque fonction seulement des u_i a une loi $\chi^2(s_1 - 1)$ si et seulement si $r = r_0$: dans ce cas $NT(Y, r)/t_0$ a la loi $\chi^2(N - 1)$.

La loi de $\hat{r}$ ne peut être explicitée, mais on peut construire des intervalles de confiance approchés en se fondant seulement sur les statistiques X_i : on prend pour nouvel estimateur $\ddot{r}$ de r la valeur qui minimise la fonction :

$$r \longrightarrow L_2(X, r) = \left(\prod_i (r + c_i)^{m_i/M} \right) \cdot \left(\sum_i X_i / (r + c_i) \right)$$

qui est, à une constante près, la fonction $L_1(X, r)$ élevée à la puissance $(- 2/M)$.

Intervalle de confiance de l'estimation $\ddot{r}$.

La probabilité que la fonction $L_2(X, r)$ admette un minimum en un point $\ddot{r}$ de l'intervalle $]r_0 - \eta, r_0 + \varepsilon[$ est supérieure à la probabilité que les deux événements suivants soient réalisés simultanément, puisque la fonction L_2 est continue en r (pourvu que $r > - \inf c_i$) :

$$E_- = \{ L_2(X, r_0 - \eta) > L_2(X, r_0) \}$$
$$E_+ = \{ L_2(X, r_0 + \varepsilon) > L_2(X, r_0) \}$$

soit :

$$Pr \{ r \in]r_0 - \eta, r_0 + \varepsilon[\} > Pr \{ E_- \cap E_+ \} > Pr \{ E_- \} + Pr \{ E_+ \} - 1$$

Un événement $E_r = \{ L_{r_2}(X, r) > L_{r_2}(X, r_0) \}$ est défini relativement aux aléatoires X_i, il s'écrit :

$$E_r = \left\{ \sum_i Z_i K_i(r, r_0) > 0 \right\}$$

où les aléatoires :

$$Z_i = X_i/(t_0(r_0 + c_i))$$

suivent des lois $\chi^2(m_i)$ indépendantes, et où :

$$K_i(r, r_0) = \frac{r_0 + c_i}{r + c_i} \prod_h (r + c_k)^{m_k/M} - \prod_k (r_0 + c_k)^{m_k/M}.$$

En considérant la deuxième fonction caractéristique, on montre que si les quantités :

$$k^{-1}2^{k-1}\left(\sum_i m_i K_i^k(r, r_0)\right)\left(\sum_i 2m_i K_i^2(r, r_0)\right)^{-k/2}, \; k \geqslant 3$$

sont très petites devant 1, la variable aléatoire

$$\sum_i Z_i K_i(r, r_0)$$

suit une loi proche de la loi normale ayant mêmes espérance et variance. Il en résulte une valeur approchée de $Pr(E_r)$, qui est fonction de l'expression :

$$Z(r, r_0) = \left(\sum_i m_i K_i(r, r_0)\right)\left(\sum_i 2m_i K_i^2(r, r_0)\right)^{-1/2}$$

dont on peut tabuler les valeurs pour un schéma donné.

Le comportement de cette fonction au voisinage de r_0 donne une première indication sur l'efficacité de l'estimation :

$$Z(r, r_0) \sim | r - r_0 | \left(\frac{1}{8}\sum_i m_i\left(\frac{1}{r_0 + c_i} - \frac{1}{M}\sum_k \frac{m_k}{r_0 + c_k}\right)^2\right)^{1/2} \qquad (3)$$

Loi de l'estimateur de t conditionnée par une valeur r.

Quand on se réduit à l'information contenue dans X_i l'estimateur de t est :

$$\bar{t} = T'(X, r) = \frac{1}{M}\sum_i \frac{X_i}{r + c_i}$$

Les premiers moments sont :

$$E(T'(X, r)) = \frac{1}{M} t_0 \left(\sum_i m_i \frac{r_0 + c_i}{r + c_i}\right) = e(r)$$

$$\text{var } (T'(X, r)) = \frac{2}{M^2} t_0^2 \left(\sum_i m_i \frac{(r_0 + c_i)^2}{(r + c_i)^2}\right) = (v(r))^2$$

Pour $r = r_0$, MT'/t_0 suit une loi χ^2 (M) ; pour des valeurs peu éloignées de r_0, on peut admettre que T' suit une loi assez proche d'une loi normale pour considérer que les probabilités

$$Pr(| T'(X, r) - e(r) | < K v(r))$$

sont minorées, uniformément en r, pour une fonction P (K). $T'(X, r)$ est une somme de variables χ^2, pondérée par des coefficients positifs : sa loi peut être aussi approchée par celle d'une loi χ^2 (GRAYBILL *et al.*, 1956).

Domaine de confiance de l'estimation ($\ddot{\theta}_2 = \ddot{t}$; $\ddot{\theta}_2 = \ddot{r}\tilde{t}$).

Soit D le domaine (fig. 1) limité par les rayons $r_- = r_0 - \eta$ et $r_+ = r_0 + \varepsilon$, d'une part, et par les deux courbes $D_\pm$ ($\theta_2 = e(r) \pm Kv(r)$; $\theta_1 = r\theta_2$), qui coupent un rayon r aux points $A_+(r)$ et $A_-(r)$, d'autre part. On a :

$$P_r((\ddot{\theta}_1, \ddot{\theta}_2) \in D) = \int_{r_-}^{r_+} P_r((\ddot{\theta}_1, \ddot{\theta}_2) \in [A_-, A_+]/\ddot{r} = r) \cdot dP_r(\ddot{r} = r)$$

$$> \int_{r_-}^{r_+} P(K) \cdot dP_r(\ddot{r} = r) \geqslant P(K) \cdot P_r(E_+ \cap E_-).$$

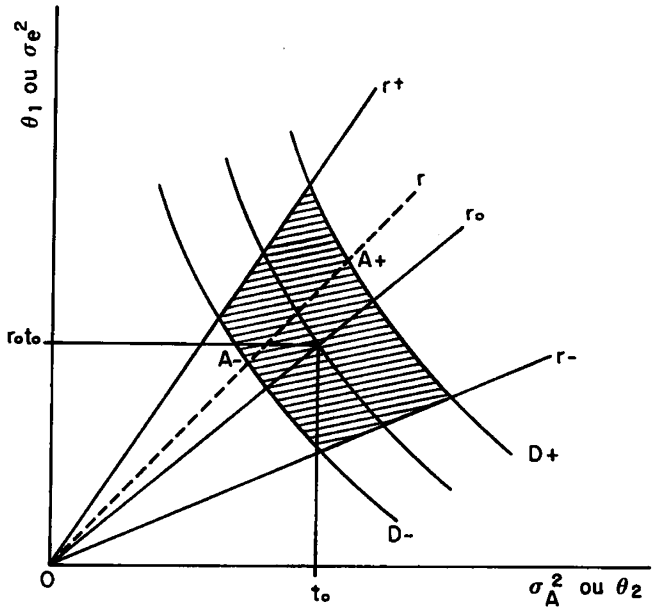


FIG. 1. — *Domaine de confiance D* (hachuré) *de l'estimation par le maximum de vraisemblance des composantes de la variance d'un caractère additif*

Approximate confidence field for the estimation of variance components of an additive trait (maximum likelihood estimator)

B_2 *Caractère génétique quelconque.*

Il existe alors plusieurs rapports r_k inconnus, et la conditions suffisante pour que la fonction $L_2(X, r)$ (où r est maintenant un vecteur) admette un minimum dans une boule ouverte $\mathring{B}$ de centre r_0, une norme étant définie, est que tous les événements E_r soient réalisés simultanément quand r parcourt la frontière $\dot{B}$ de $\mathring{B}$:

$$P_r(\ddot{r} \in \mathring{B}) \geqslant P_r\left(\bigcap_{r \in \dot{B}} E_r\right)$$

Il faudrait, pour utiliser ce critère, déterminer la probabilité d'une intersection infinie d'événements, c'est-à-dire connaître la loi de la variable aléatoire :

$$\inf_{r \in \dot{B}} L_2(X, r)$$

En général une simulation est nécessaire : il est alors plus judicieux de simuler directement la loi de l'estimation $\tilde{r}$ qui ne demande pas plus de calculs numériques.

Enfin des conditions de régularité de la fonction aléatoire $L_{\tilde{r}}(X, r)$ doivent être vérifiées pour que la variable aléatoire précédente soit définie et que sa simulation ait un sens.

Pour un schéma stable donné, tout critère d'efficacité des estimations du maximum de vraisemblance exige donc de longs calculs. Des résultats asymptotiques de convergence en probabilité pourraient être établis en considérant une suite de schémas stables et en indiquant des conditions sur les multiplicités des valeurs propres. Leur intérêt est plus théorique que pratique.

C. — *Estimations quadratiques justes à norme minimale*

La construction d'estimateurs quadratiques à norme minimale ne repose sur aucune hypothèse concernant les distributions. Les statistiques X_i, u_i, v_i ne peuvent donc plus être considérées comme des résumés exhaustifs mais elles possèdent encore la propriété remarquable de fournir des formes quadratiques orthogonales par rapport à toute norme définie par un vecteur $\bar{\theta}$ arbitraire. Ce sont les formes X_i et Q_i définies au paragraphe III. — A dont l'orthogonalité est immédiate quand on les exprime en fonction des aléatoires z_{ij} ; en effet :

$$Q_i = z_{i1}^2 + z_{i2}^2$$

$$X_i = \sum_{j=3}^{j=n_i} z_{ij}^2$$

et la norme est définie, par rapport aux aléatoires z_{ij}, par la matrice diagonale d'éléments $w_i(\bar{\theta})^{-1}$.

La procédure suivante de construction des estimateurs quadratiques justes, invariants, à norme minimale, en résulte :

1° choisir une famille $\bar{\theta}$ de paramètres pour définir la norme.

2° supposer les effets fixes u et v connus, et déduire les équations linéaires en $\bar{\theta}_k$ de la minimisation de la somme de carrés suivante :

$$\sum_i \frac{(X_i - m_i \sum_k c_{ik}\theta_k)^2}{m_i w_i^2(\bar{\theta})} + \sum_j \frac{(Q_j(u, v) - 2 \sum_k c_{ik}\theta_k)^2}{2w_j^2(\bar{\theta})}$$

Ces équations s'écrivent :

$$L_k(\theta) = M_k(X_i, u_j, v_j, u, v)$$

où L_k est une forme linéaire en θ_l et M_k une fonction de la forme :

$$M = M^{(1)}(X_i, u_i^2, u_i v_i, v^2) + uM^{(2)}(u_i, v_i) + vM^{(3)}(u_i, v_i) + u^2M^{(4)} + uvM^{(5)} + v^2M^{(6)}.$$

$M^{(1)}$, $M^{(2)}$, $M^{(3)}$ sont des fonctions linéaires des arguments ;
$M^{(4)}$, $M^{(5)}$ et $M^{(6)}$ sont des scalaires.

3° remplacer u et v par leurs estimateurs linéaires justes $\hat{u}$ et $\hat{v}$, à norme minimale. Les seconds membres :

$$M_k(X_i, u_j, v_j, \hat{u}, \hat{v})$$

sont alors des formes quadratiques des observations Y_{ij}.

4° recalculer les espérances des quantités $M_k(X_i, u_j, v_j, \hat{u}, \hat{v})$ ce sont de nouvelles formes linéaires $L'_k(\theta_l)$ en θ_l, indépendantes de u et v

5° la solution du système d'équations linéaires :

$$L'_k(\theta_l) = M_k(X_i, u_j, v_j, \hat{u}, \hat{v})$$

coïncide avec les estimateurs quadratiques justes à norme minimale définis par RAO.

Cette procédure est généralement plus rapide que la méthode générale de RAO si le modèle se présente sous sa forme canonique. L'identité des estimateurs obtenus par les deux méthodes peut être vérifiée en faisant le calcul explicite dans les deux cas.

Si l'on ne prend en compte que l'information contenue dans les résumés X_i, les estimateurs sont obtenus directement par minimisation de la somme :

$$\sum_i \frac{(X_i - m_i \sum_k c_{ik}\theta_k)^2}{m_i w_i^2(\bar{\theta})}$$

Les équations sont :

$$\sum_j \theta_j \sum_i \frac{m_i c_{ik} c_{ij}}{w_i^2(\bar{\theta})} = \sum_i \frac{c_{ik} X_i}{w_i^2(\bar{\theta})}$$

IV. — MODIFICATIONS DES SCHÉMAS EXPÉRIMENTAUX EN GÉNÉTIQUE ANIMALE

Les résultats précédents montrent que seuls les modèles stables sont susceptibles de fournir des estimations dont la précision soit connue sans recourir à des expériences de simulation. Il est donc nécessaire d'analyser les possibilités pratiques de construire des schémas expérimentaux stables. Sauf dans le cas où les composantes génétiques ont des rapports connus (§ III. — B), les conditions à remplir se posent au niveau de la structure du pedigree de la population étudiée. On envisage spécialement ici les conditions supplémentaires qu'il faut adjoindre au schéma hiérarchique d'analyse de variance, généralement utilisé en génétique animale.

A. — *Les hypothèses du schéma hiérarchique*

Les phénotypes mesurés P_{mlk} sont les phénotypes des sdn produits issus de s pères accouplés chacun à d mères qui donnent chacune n produits. Les s pères et les sd mères sont issus d'une population panmictique infinie et sont non apparentés. Les espérances et les covariances des phénotypes sont alors :

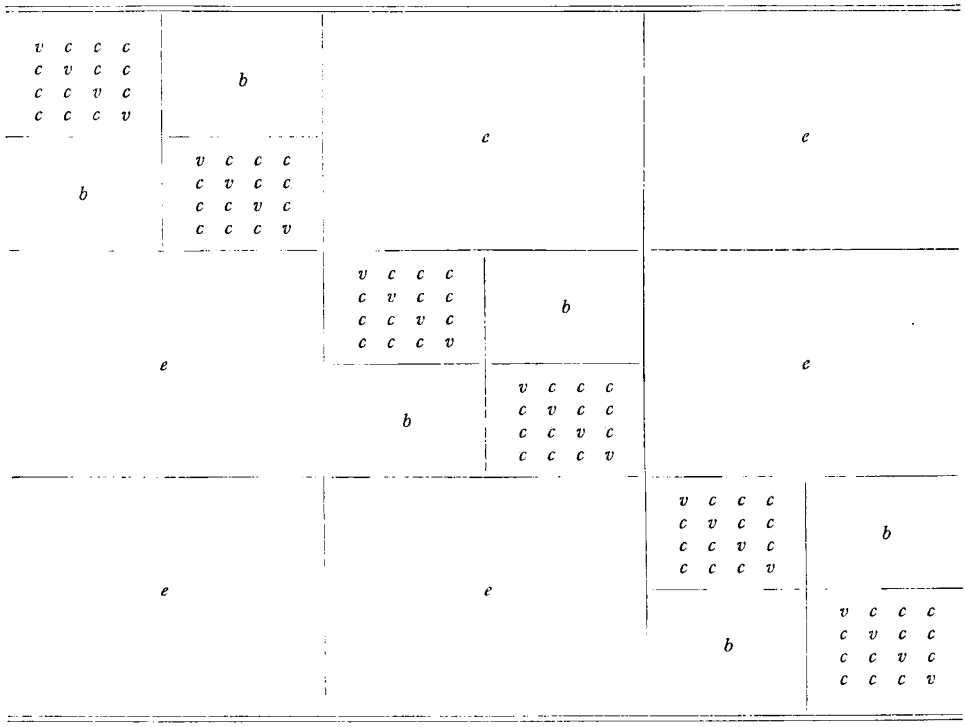
$$\begin{aligned}
E(P_{mlk}) &= u \\
\text{var}\,(P_{mlk}) &= \theta_1 + \theta_2 + \theta_3 = v \\
\text{cov}\,(P_{mlk}P_{mlk'}) &= 1/2 \cdot \theta_2 + 1/4 \cdot \theta_3 = c \\
\text{cov}\,(P_{mlk}P_{ml'k'}) &= 1/4 \cdot \theta_2 = b \\
\text{cov}\,(P_{mlk}P_{m'l'k'}) &= 0 = e
\end{aligned} \qquad (4)$$

Le tableau 1 représente la matrice V(v, c, b, e) correspondante, et le tableau 2 une matrice de vecteurs propres orthogonaux.

TABLEAU I

Matrice des variances et covariances V(v, c, b, e) *pour* $s = 3$, $d = 2$, $n = 4$

Variance-covariance matrix V(v, c, b, e) *for the hierarchical classification. Values of parameters are* : $s = 3$, $d = 2$, $n = 4$



En notant $\overline{P_{mi.}}$, $\overline{P_{m..}}$, $\overline{P_{...}}$ les moyennes des phénotypes par rapport aux indices omis, les valeurs propres, leurs multiplicités et les sommes de carrés X_i sont celles du tableau 3 d'analyse de variance. L'existence des trois sommes de carrés, orthogonales, et définies indépendamment des paramètres v, c, b et e est assurée par l'existence des vecteurs propres eux-mêmes indépendants de ces paramètres : il est aisé de vérifier la commutativité de deux matrices du type V (v, c, b, e). Inversement l'utilisation pratique de cette analyse de variance classique n'est justifiée théoriquement que si la matrice V de variances et covariances admet les mêmes vecteurs propres associés à des valeurs propres ayant les mêmes multiplicités, et cela entraîne que V a exactement la forme V (v, c, b, e).

Rao (1965) présente un modèle généralisé d'analyse de variance qui distingue deux types de covariances entre produits issus de pères distincts :

$$\text{cov}\,(P_{mlk}P_{m'lk'}) = e'$$
$$\text{cov}\,(P_{mlk}P_{m'l'k'}) = e'' \qquad l \neq l'$$

TABLEAU 2

Vecteurs propres lignes, orthogonaux mais non normés, de la matrice V(v, c, b, e) *du tableau 1 ; valeurs propres associées et leurs multiplicités*

Orthogonal eigen-vectors of the matrix V *of table 1 ; eigen-values and their multiplicities*

−1 1 −1 −1 2 −1 −1 −1 3 −1 1 −1 −1 2 −1 −1 −1 3 −1 1 −1 −1 2 −1 −1 −1 3 −1 1 −1 −1 2 −1 −1 −1 3 −1 1 −1 −1 2 −1 −1 −1 3	$v - c$	$sd(n-1)$
−1 −1 −1 −1 1 1 1 1 −1 −1 −1 −1 1 1 1 1 −1 −1 −1 −1 1 1 1 1	$v + (n-1)c - nb$	$s(d-1)$
−1 −1 −1 −1 −1 −1 −1 −1 1 1 1 1 1 1 1 1 −1 −1 −1 −1 −1 −1 −1 −1 −1 −1 −1 −1 −1 −1 −1 −1 2 2 2 2 2 2 2 2	$v + (n-1)c + (d-1)nb - dne$	$s - 1$
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	$v + (n-1)c + (d-1)nb + (s-1)dne$	1

Les matrices V ayant cette structure sont commutatives, et RAO donne les expressions des sommes de carrés X_i orthogonales.

Ces modèles statistiques, qui ne font intervenir que quatre ou cinq valeurs distinctes de covariances entre apparentés, ne sont pas adaptés à l'analyse des populations animales consanguines. D'une part la structure hiérarchique des accouplements ne garantit pas l'égalité des covariances entre apparentés du même type : la matrice de covariance ne possède plus la forme $V(v, c, b, e)$ qui assure l'orthogonalité des sommes des carrés de l'analyse de variance. D'autre part, si l'on tient compte des coefficients de parenté et d'identité calculés d'après les généalogies, il est impossible de construire un plan d'accouplements qui conduise à l'un des modèles statistiques précédents sans limiter considérablement le nombre de reproducteurs.

TABLEAU 3

Analyse de variance relative à la matrice du tableau 1

Analysis of variance table for the matrix V *of table 1*

w_i	m_i	X_i
$v - c$	$sd(n-1)$	$\Sigma\Sigma\Sigma_{mlk} (P_{mlk} - \overline{P_{ml.}}) = X_1$
$v + (n-1)c - nb$	$s(d-1)$	$n \Sigma\Sigma_{ml} \overline{(P_{ml.} - P_{m..})^2} = X_2$
$v + (n-1)c + (d-1)nb - dne$	$s-1$	$nd \Sigma_{m} \overline{(P_{m..} - P_{...})^2} = X_3$

Les souches sélectionnées, et maintenues en troupeaux fermés, des élevages avicoles illustrent cette difficulté. Les calculs de coefficients de parenté ont été faits dans une telle souche où les généalogies remontaient à quinze générations. A chaque génération, s coqs ($s = 20$ et 40 dans les deux dernières générations) sont accouplés chacun à d poules ($d = 14$) donnant chacune n descendants contrôlés. L'analyse classique distingue trois types de corrélations entre apparentés, caractérisés par les coefficients de parenté φ :

$$\text{entre pleins-frères : } \varphi = \frac{c}{2} = \frac{1}{4}$$

$$\text{entre demi-frères : } \varphi = \frac{b}{2} = \frac{1}{8}$$

$$\text{entre individus de pères distincts : } \varphi = \frac{e}{2} = 0$$

Or le calcul montre que pour la dernière génération, ces trois types de coefficients de parenté prennent leurs valeurs dans des domaines assez larges et qui se recouvrent :

$$\text{entre pleins-frères : } 0{,}43 < \varphi = \frac{c}{2} < 0{,}55$$

$$\text{entre demi-frères : } 0{,}33 < \varphi = \frac{b}{2} < 0{,}47$$

$$\text{entre individus de pères distincts : } 0{,}24 < \varphi = \frac{e}{2} < 0{,}40$$

(CHEVALET, 1974).

Dans ces conditions les sommes X_i de l'analyse de variance n'ont pas pour espérances les quantités $m_i \cdot w_i$ du tableau 3 (ces quantités ne sont d'ailleurs pas définies), et ne sont pas orthogonales. L'interprétation génétique des résultats est donc sujette à caution, en particulier l'estimation de l'héritabilité par cette méthode.

Par ailleurs, il est difficile, dans une population où les coefficients de parenté varient continûment de 0,24 à 0,55 de définir un plan d'accouplements tel que dans la génération suivante les coefficients de parenté ne puissent prendre que quelques valeurs distinctes.

Une analyse cohérente de la variabilité des populations consanguines exige donc simultanément la modification des procédés statistiques d'analyse de variance et la prise en compte systématique des coefficients de parenté et d'identité dans la conception des plans d'accouplements.

B. — *Modifications du schéma d'analyse de variance*

1. *Modification d'Hinkelmann.*

Dans le cadre du schéma génétique étudié par MALÉCOT (1939), où les $s + sd$ reproducteurs peuvent être apparentés quand ils ne sont pas accouplés, HINKELMANN (1971) a calculé les espérances des trois sommes X_i de l'analyse de variance en tenant compte des coefficients de parenté. Les estimations des composantes θ_1, θ_2 et θ_3 qui en résultent sont justes, mais les sommes X_i ne sont en général pas orthogonales. La détermination numérique de leurs variances, ou de leurs normes, exigerait à peu près autant de calculs que le calcul direct des estimateurs à norme minimale.

2. *Conditions de stabilité partielle.*

Des conditions moins strictes que les relations (4) permettent d'utiliser le plan hiérarchique d'accouplements, dont l'intérêt zootechnique est évident. En effet il existe toujours des vecteurs propres indépendants des composantes inconnues dès qu'il existe des familles de pleins-frères : la matrice des covariances entre individus de la famille (m, l) a toujours la forme :

$$\begin{bmatrix} v & c & c & c \\ c & v & c & c \\ c & c & v & c \\ c & c & c & v \end{bmatrix} \tag{5}$$

et la covariance

$$\operatorname{cov}\,(P_{mlk}, P_{m'l'k'}),\ (m', l') \neq (m, l),$$

ne dépend pas du représentant (k) de la famille (m, l).

Les sommes

$$X_{1,m,l} = \sum_k (P_{mlk} - \overline{P_{ml.}})^2$$

sont donc orthogonales et d'espérances $(n - l)(v - c)$ où les nombres n, v et c peuvent dépendre des indices (m, l). Pour associer à des vecteurs propres indépendants des composantes des sommes du type :

$$X_{2,m} = n \sum_l \overline{(P_{ml.} - P_{m..})^2}$$

il est nécessaire d'imposer les conditions suivantes :

1° Le modèle est équilibré à l'intérieur de chacune des familles de père. Le nombre de femelles accouplées à un mâle peut varier, mais le nombre de produits issus d'une mère doit être seulement fonction de l'indice du père :

$$n_{ml} = n_m$$

2° La matrice des covariances entre moyennes de familles de mères a la même forme que la matrice (5). Il faut pour cela :

$$\begin{aligned} \text{var } (P_{mlk}) &= v_m \\ \text{cov } (P_{mlk}, P_{mlk'}) &= c_m \\ \text{cov } (P_{mlk}, P_{ml'k'}) &= b_m \\ \text{cov } (P_{mlk}, P_{m'l'k'}) &= e_{m,m'} \end{aligned} \qquad (6)$$

Dans ces conditions on peut dresser un tableau d'analyse de variance partielle (tabl. 4), qui comporte $2s$ sommes de carrés orthogonales associées à $\sum_m d_m n_m - s$ vecteurs propres indépendants des composantes inconnues.

TABLEAU 4

Analyse de variance partielle relative à la matrice définie par les relations (6)

Partial analysis of variance table for a matrix defined by equations (6) *in the text*

w_i	m_i	$X_{i.\,m}$
$v_m - c_m$	$d_m(n_m - 1)$	$\sum\sum_{lk} (P_{mlk} - \overline{P_{ml.}})^2 = X_{1.m}$ s sommes de ce type
$v_m + (n_m - 1)c_m - n_m b_m$	$d_m - 1$	$n_m \sum_l \overline{(P_{ml.} - P_{m..})^2} = X_{2.m}$ s sommes de ce type

Pour compléter l'analyse il faut prendre en compte la matrice de variances et covariances relatives aux quantités $(d_m n_m)^{-1/2} P_{m..}$:

$$\begin{aligned} \text{var } ((d_m n_m)^{-1/2} P_{m..}) &= v_m + (n_m - 1)C_m + n_m(d_m - 1)b_m \\ \text{cov } ((d_m n_m)^{-1/2} P_{m..}, (d_{m'} n_{m'})^{-1/2} P_{m'..}) &= (d_m n_m d_{m'} n_{m'})^{1/2} e_{mm'} \end{aligned}$$

En général le modèle partiel relatif à ces aléatoires n'est pas stable. Son analyse par les méthodes générales d'estimation est nécessaire car la variabilité correspondante est essentiellement d'origine génétique alors que la variabilité des phénotypes P_{mlk} autour des moyennes $\overline{P}_{ml.}$ est principalement due aux effets du milieu, notamment si l'héritabilité du caractère est faible.

DISCUSSION

L'étude précédente révèle les difficultés de l'analyse statistique dans les populations consanguines. Dans une population où les relations de parenté n'ont pas été contrôlées et maîtrisées systématiquement, les méthodes usuelles d'estimation sont peu précises et biaisées notamment si le caractère est soumis à des interactions de dominance : avec un coefficient de consanguinité moyen de 0,33 une analyse de variance non corrigée donnera une estimation au moins deux fois trop faible du rapport de la variance de dominance à la variance génétique additive, et par conséquent surestimera la fraction génétique additive de la variance totale phénotypique. D'autre part, des hypothèses précises doivent être formulées, sur la nature du caractère quantitatif, ou sur la structure de la population pour que les estimations aient une précision connue et soient d'un calcul aisé.

Les estimateurs des effets fixes et des composantes de variance s'expriment en fonction de résumés exhaustifs des observations, définis indépendamment des paramètres inconnus, si et seulement si le modèle statistique est « stable » : la matrice des variances et covariances entre les observations s'écrivant (§ I. — B) :

$$\text{cov}\,(PP') = \theta_1 d_1 + \theta_2 d_2 + \cdots + \theta_k d_k.$$

il faut supposer la commutativité des produits de matrices $d_i d_j$. Cette condition, nécessaire pour expliciter les propriétés des estimateurs, est très rigoureuse. Cependant elle est souvent admise dans les études théoriques : elle est respectée dans le modèle général d'analyse de variance traité par RAO (1965), dans le cas équilibré ; elle est implicite dans l'étude de HARTLEY et RAO (1967) pour démontrer la convergence de l'estimation du maximum de vraisemblance et donc pour justifier l'importance des calculs numériques requis par cette méthode ; elle apparaît aussi dans un exemple du calcul des estimations quadratiques justes à variance minimale dans le cas d'un modèle déséquilibré à un niveau de classification, traité par LAMOTTE (1973).

Le seul cas, en Génétique animale, où le modèle est « stable » est celui où le caractère est purement additif. La propriété disparaît si le caractère est soumis à une influence maternelle, ou plus généralement s'il existe des effets aléatoires d'environnement spécifiques, à un troupeau, à une étable, à une famille, etc. Même si le modèle est satisfaisant, on se heurte au problème numérique de diagonaliser une matrice dont l'ordre est égal au nombre de fratries dans la population analysée (donc au nombre de mères dans le cas d'un schéma hiérarchique d'accouplements). Dans les cas généraux le modèle n'est pas stable, il faut recourir à des méthodes générales d'estimation, qui reposent sur le calcul de l'inverse d'une matrice d'ordre

égal au nombre de fratries. Outre leur coût élevé, de tels calculs numériques sont sujets à des erreurs d'arrondi, même avec de puissants calculateurs (DURAND, 1961).

Il nous paraît donc nécessaire, autant pour des raisons statistiques théoriques que pour des raisons pratiques de calcul numérique, d'envisager une structuration précise du plan d'expérience. Les conditions énoncées au paragraphe IV B. — 2 relatives au schéma hiérarchique d'accouplements satisfont à cette exigence en rendant le modèle partiellement « stable », dans toutes les hypothèses génétiques, et en limitant l'ordre de la matrice à inverser ou à diagonaliser au nombre des pères. Ces conditions portent sur les valeurs des coefficients d'identité et sur l'équilibre des données. L'égalité des coefficients d'identité, entre demi-frères issus d'un même père, et entre deux produits quelconques issus de deux pères différents semble être assez facile à réaliser dans des populations où le nombre des reproducteurs est petit, pourvu que la souche ne soit pas conduite en effectif limité depuis de nombreuses générations. En revanche, la condition d'équilibre des données à l'intérieur de chaque famille de père est trop exigeante, sauf peut-être pour des espèces peu prolifiques comme les bovins où les veaux jumeaux sont rares. Le non-respect de l'équilibre interdit de considérer que le modèle est partiellement stable : le recours aux méthodes numériques générales d'estimation est alors nécessaire. Dans les cas faiblement déséquilibrés, on peut adapter les programmes de calcul, écrits pour le cas équilibré, de façon à ce que les estimations quadratiques demeurent non biaisées. L'approximation résulte du fait que l'on considère alors comme orthogonales des sommes de carrés qui ne le sont plus. De même le calcul de la vraisemblance est faussé.

Par ailleurs, les tests d'hypothèse sont difficiles à mettre en œuvre. Dans le cas le plus favorable des modèles stables on peut envisager un test du rapport des vraisemblances. En se limitant à l'information contenue dans les résumés quadratiques X_i (§ III. — A), avec :

$$E(X_i) = m_i \sum_{k=1}^{K} c_{ik}\theta_k \qquad (i = 1, s),$$

le test d'une hypothèse H_o (où par exemple on suppose $\theta_1 = 0$) par rapport à l'hypothèse contraire $H_1(\theta_1 \neq 0)$ s'écrit :

$$-2 \log r = \frac{\min_{\theta_2..\theta_K} \sum_i \left[X_i/(\sum_2^K c_{ik}\theta_k) + m_i \log (\sum_2^K c_{ik}\theta_k) \right]}{\min_{\theta_1,\ \theta_2..\theta_K} \sum_i \left[X_i/(\sum_1^K c_{ik}\theta_k) + m_i \log (\sum_1^K c_{ik}\theta_k) \right]}$$

Sauf dans des cas très particuliers (par exemple si la matrice des coefficients c_{ik} est carrée, de rang égal à son ordre, et si de plus deux des sommes du type $\Sigma_2^K c_{ik}\theta_k$ coïncident), cette expression ne peut pas être explicitée en fonction des quantités X_i. D'une façon générale, seuls des résultats asymptotiques sur la distribution du test peuvent être établis (HARTLEY et RAO, 1967).

Tous les modèles de Génétique quantitative, fondés sur la notion d'identité des gènes, conduisent à une même classe de modèles statistiques caractérisés par l'expression de la matrice des variances et covariances, rappelée au début de cette discussion. On peut ainsi rendre compte dans les populations consanguines d'effets

maternels génotypiques (GILLOIS, 1964), d'interactions épistatiques (GALLAIS, 1970), et même des caractères quantitatifs dans les populations d'individus tétraploïdes (BOUFFETTE, 1966). Par ailleurs nous disposons de programmes de calcul des coefficients d'identité (CHEVALET, 1971). Les modèles fondés sur la relation d'identité supposent cependant l'absence de toute homogamie et de toute sélection : si ces phénomènes ne peuvent être exclus, les formules que nous utilisons doivent être considérées comme des approximations. Cependant la sélection pour un caractère quantitatif n'induit, au niveau d'un gène particulier, qu'une faible pression de sélection ; et l'on peut admettre que, pendant quelques générations au moins, l'évolution de la variabilité génétique est due principalement au phénomène de dérive génétique dont on tient compte *a priori* en calculant les coefficients d'identité.

En se fondant sur ces modèles génétiques, on peut envisager l'estimation de termes d'interaction qui ont une grande importance zootechnique, en particulier l'interaction entre les effets direct et indirect maternels. On est cependant amené à introduire un grand nombre de paramètres à estimer, et une structuration du plan expérimental est nécessaire, afin de pouvoir éliminer ou regrouper certaines quantités. L'écriture d'un modèle très général, tenant compte d'interactions de dominance et d'un effet maternel génotypique conduit ainsi à quatorze composantes de variance dans une population consanguine. Il n'est pas possible, par une analyse de variance, de choisir entre plusieurs hypothèses simplificatrices pour décrire le caractère ; des expériences complémentaires qui confrontent prévision et résultat expérimental sont nécessaires : ainsi le modèle additif est-il confirmé par des expériences de sélection sur plusieurs générations quand l'héritabilité estimée par analyse de variance coïncide avec l'héritabilité réalisée. Une telle confirmation ne prouve pas l'hypothèse génétique d'additivité, mais justifie une description statistique du caractère quantitatif. De la même façon une méthode d'estimation de composantes liées à des effets non-additifs peut être un préalable à l'élaboration d'une meilleure transcription statistique de l'hérédité quantitative, sans pour autant apporter de preuve catégorique à l'appui d'une hypothèse génétique.

Reçu pour publication en mars 1976.

SUMMARY

ESTIMATION
OF PHENOTYPIC VARIANCE COMPONENTS FROM AN INBRED POPULATION
I. — ELABORATION OF THE MODEL

From the theoretical expressions of genetic variances and covariances among related individuals, a general statistical model may be derived for the estimation of variance components from data collected in an inbred flock. The general expression of the variance, covariance matrix of phenotypes is : $V(\theta) = \sum_k d_k \theta_k$, where (d_k) are matrices of kinship, and identity coefficients, and (θ_k) are the scalar components of variance to be estimated. The condition for existence of a set of sufficient and independent statistics is that, for any two sets θ and θ' of parameters, the identity : $V(\theta)V(\theta') = V(\theta')V(\theta)$ holds. Therefore, except in the case of additive genic contributions, there is no general and efficient estimation. In all other cases a precise structuration of the mating scheme is necessary to obtain, at least, a partial reduction of information into independent statistics. Conditions are defined about the usual hierarchical

design, that allow some simplifications in the numerical computations, and give efficient estimations. In these conditions are involved the values taken by identity coefficients, the distribution of these values within and between families ; a partial equilibrium of data is also needed. Even in the additive case, such conditions are required, for the computations to be tractable. Uses of maximum likelihood estimators, and of minimum variance quadratic unbiased estimators are investigated and discussed.

RÉFÉRENCES BIBLIOGRAPHIQUES

BOUFFETTE J., 1966. *Expression de la covariance génotypique chez les trétraploïdes.* Thèse 3e cycle, Fac. Sciences, Lyon.

CUNNINGHAM E. P., HENDERSON C. R., 1968. An iterative procedure for estimating fixed effects and variance components in mixed model situations. *Biometrics*, **24**, 13-25.

DURAND E., 1961. *Solutions numériques des équations algébriques.* Tome 2, Masson et Cie, Paris.

CHEVALET C., 1971 *a*. Calcul automatique des coefficients d'identité. *Ann. Génét. Sél. anim.*, **3**, 449-462.

CHEVALET C., 1971 *b*. Calcul *a priori*, intra- et inter- populations des variances et covariances génotypiques entre apparentés quelconques. *Ann. Génét. Sél. anim.*, **3**, 463-477.

GALLAIS A., 1970. Covariances entre apparentés quelconques avec linkage et épistasie. I : Expression générale. *Ann. Génét. Sél. anim.*, **2**, 281-310.

GILLOIS M., 1964. *La relation d'identité en génétique.* Thèse, Fac. Sciences, Paris, 294 p.

GRAYBILL F. R., MARTIN F., GODFREY G., 1956. Confidence intervals for variance ratios specifying genetic heritability. *Biometrics*, **12**, 99-109.

HARTLEY H., O., RAO J. N. K., 1967. Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika*, **54**, 93-108.

HENDERSON C. R., 1953. Estimation of variance and covariance components. *Biometrics*, **7**, 226-252.

HINKELMANN K., 1969. Estimation of heritability from experiments with related dams. *Biometrics*, **25**, 755-766.

HINKELMANN K., 1971. Estimation of heritability from experiments with inbred and related individuals. *Biometrics*, **27**, 183-190.

LAMOTTE L. R., 1973. Quadratic estimation of variance components. *Biometrics*, **29**, 311-330.

MALÉCOT G., 1939. *Théorie mathématique de l'hérédité mendélienne généralisée.* Thèse Fac. Sciences, Paris. Republiée in : MALÉCOT G., 1966. *Probabilité et hérédité.* I.N.E.D., cahier n° 47, P.U.F., Paris.

RAO C. R., 1965. *Linear statistical inference and its applications.* Wiley, New York, 522 p.

RAO C. R., 1971 *a*. Estimation of variance and covariance components. MINQUE theory. *J. of multivariate Analysis*, **1**, 257-275.

RAO C. R., 1971 *b*. Minimum variance quadratic unbiased estimation of variance components. *J. of multivariate Analysis*, **1**, 445-456.

SEARLE S. R., 1970. Large sample variances of maximum likelihood estimations of variance components using unbalanced data. *Biometrics*, **26**, 505-524.

SEARLE S. R., 1971. Topics in variance components estimation. *Biometrics*, **27**, 1-76.

THOMPSON R., 1969. Iterative estimation of variance components for non-orthogonal data. *Biometrics*, **25**, 767-773.