Estimation bayesienne empirique de parametres relatifs a n caracteres binaires polygeniques

Jean Louis J. L. Foulley, S. Im, Daniel Gianola, I. Hoschele

Génét. Sél. Evol., 1987, **19** (2), 197-224

Empirical Bayes estimation of parameters for n polygenic binary traits

J.L. FOULLEY*, ***, S. IM **, D. GIANOLA *** and Ina HÖSCHELE ****

* *I.N.R.A., Station de Génétique quantitative et appliquée, Centre de Recherches Zootechniques, F 78350 Jouy-en-Josas*

** *I.N.R.A., Laboratoire de Biométrie, Centre de Recherches de Toulouse, B.P. 27, F 31326 Castanet-Tolosan Cedex.*

*** *Department of Animal Sciences, University of Illinois Urbana, Illinois 61801, U.S.A.*

**** *Universität Hohenheim, Institut 470, Haustiergenetik, D-7000 Stuttgart 70, R.F.A.*

Summary

The conditional probability of an observation in a subpopulation i (a combination of levels of explanatory variables) falling into one of 2^n mutually exclusive and exhaustive categories is modelled using a normal integral in n-dimensions. The mean of subpopulation i is written as a linear combination of an unknown vector $\boldsymbol{\theta}$ which can include « fixed » effects (e.g., nuisance environmental effects, genetic group effects) and « random » effects such as additive genetic value or producing ability. Conditionally on $\boldsymbol{\theta}$, the normal integral depends on an unknown matrix $\mathbf{R}$ comprising residual correlations in a multivariate standard normal conceptual scale. The random variables in $\boldsymbol{\theta}$ have a dispersion matrix $\mathbf{G} \otimes \mathbf{A}$, where usually $\mathbf{A}$ is a known matrix of additive genetic relationships, and $\mathbf{G}$ is a matrix of unknown genetic variances and covariances. It is assumed *a priori* that $\boldsymbol{\theta}$ follows a multivariate normal distribution $f(\boldsymbol{\theta} \mid \mathbf{G})$, which does not depend on $\mathbf{R}$, and the likelihood function is taken as product multinomial. The point estimator of $\boldsymbol{\theta}$ is the mode of the posterior distribution $f(\boldsymbol{\theta} \mid \mathbf{Y}, \mathbf{G} = \mathbf{G}^*, \mathbf{R} = \mathbf{R}^*)$ where $\mathbf{Y}$ is data, and $\mathbf{G}^*$ and $\mathbf{R}^*$ are the components of the mode of the marginal posterior distribution $f(\mathbf{G}, \mathbf{R} \mid \mathbf{Y})$ using « flat » priors for $\mathbf{G}$ and $\mathbf{R}$. The matrices $\mathbf{G}^*$ and $\mathbf{R}^*$ correspond to the marginal maximum likelihood estimators of the corresponding matrices. The point estimator of $\boldsymbol{\theta}$ is of the empirical Bayes types. Overall, computations involve solving 3 non-linear systems in $\boldsymbol{\theta}$, $\mathbf{G}$ and $\mathbf{R}$. $\mathbf{G}^*$ can be computed with an expectation-maximization type algorithm ; an estimator of $\mathbf{R}^*$ is suggested, and this is related to results published elsewhere on maximum likelihood estimation in contingency tables. Problems discussed include non-linearity, size of the system to be solved, rate of convergence, approximations made and the possible use of informative priors for the dispersion parameters.

Key words : Multiple trait evaluation, all-or-none traits, categorial variates, Bayesian methods.

Résumé

Estimation bayésienne empirique de paramètres relatifs à n caractères binaires polygéniques

La probabilité conditionnelle qu'une observation d'une sous-population donnée (combinaison de niveaux de facteurs) se trouve dans l'une des 2^n catégories possibles de réponse (exclusives et exhaustives) est modélisée par une intégrale normale à n-dimensions. La moyenne de la i^e sous-

population s'écrit comme une combinaison linéaire d'un vecteur $\boldsymbol{\theta}$ de paramètres inconnus qui peuvent comprendre des effets « fixes » (effets de milieu parasites, effets de groupe génétique) et des effets aléatoires (valeur génétique additive ou aptitude à la production). Sachant $\boldsymbol{\theta}$, l'intégrale normale dépend d'une matrice inconnue $\mathbf{R}$ fonction des corrélations résiduelles entre les n variables normales sous-jacentes standardisées. Les effets aléatoires de $\boldsymbol{\theta}$ présentent une matrice de dispersion de la forme $\mathbf{G} \otimes \mathbf{A}$ où $\mathbf{A}$ est généralement une matrice connue de parenté et $\mathbf{G}$ une matrice inconnue de variances et covariances génétiques. On suppose qu'*a priori* $\boldsymbol{\theta}$ suit une loi multinormale de densité $f(\boldsymbol{\theta} \mid \mathbf{G})$ qui ne dépend pas de $\mathbf{R}$. La vraisemblance s'exprime alors comme un produit de multinomiales. L'estimateur de position de $\boldsymbol{\theta}$ est défini comme le mode de la distribution *a posteriori* $f(\boldsymbol{\theta} \mid \mathbf{Y}, \mathbf{G} = \mathbf{G}^*, \mathbf{R} = \mathbf{R}^*)$ où $\mathbf{Y}$ est le vecteur des données, $\mathbf{G}^*$ et $\mathbf{R}^*$ sont les composantes du mode de la distribution marginale $f(\mathbf{G}, \mathbf{R} \mid \mathbf{Y})$ avec des *a priori* uniformes pour $\mathbf{G}$ et $\mathbf{R}$. $\mathbf{G}^*$ et $\mathbf{R}^*$ correspondent alors aux estimateurs du maximum de vraisemblance marginale et $\boldsymbol{\theta}$ à un estimateur de type bayésien empirique. Les calculs impliquent la résolution de 3 systèmes non-linéaires en $\boldsymbol{\theta}$, $\mathbf{G}$ et $\mathbf{R}$. $\mathbf{G}^*$ se calcule selon un algorithme de type E.M. Une approximation de $\mathbf{R}^*$ est suggérée en relation avec des résultats antérieurs publiés à propos d'une estimation du maximum de vraisemblance pour les tables de contingence. Divers problèmes sont abordés en discussion tels que la non-linéarité, la taille du système à résoudre, la vitesse de convergence, le degré d'approximation et l'emploi possible d'*a priori* informatifs pour les paramètres de dispersion.

Mots clés : Evaluation multidimensionnelle, caractères tout-ou-rien, variables discrètes, méthodes bayésiennes.

I. Introduction

Several new procedures of sire evaluation for discrete characters postulate an underlying normal distribution which is made discrete via a set of thresholds (GIANOLA & FOULLEY, 1982, 1983 ; FOULLEY & GIANOLA, 1984 ; HARVILLE & MEE, 1984 ; GILMOUR *et al.*, 1985). In the method of GIANOLA & FOULLEY, the records in a sample are allocated to sub-populations consisting of one or more individuals ; the mean of each sub-population is a linear combination of an unknown vector $\boldsymbol{\theta}$. The link between these means and the discrete observations is provided by a multivariate normal integral with an argument dependent on location and dispersion parameters (HÖSCHELE *et al.*, 1986). Inferences about $\boldsymbol{\theta}$ are made using Bayesian procedures which readily accommodate « fixed » effects (nuisance environmental parameters, genetic group means) and « random » effects such as the breeding values of animals to be evaluated. As in the case of genetic evaluation by best linear unbiased prediction (HENDERSON, 1973), the estimators and predictors are obtained from the posterior distribution of $\boldsymbol{\theta}$, conditionally on the intervening dispersion parameters, e.g., heritabilities, genetic and residual correlations. The objective of this paper is to further generalize the methods for discrete variables by considering the situation where the values of the dispersion parameters are not known. In particular, we present a solution based upon replacing these parameters by point estimates obtained from their marginal posterior distribution (O'HAGAN, 1980 ; GIANOLA *et al.*, 1986). The procedure provides estimates of the components of the dispersion structure and predictors of linear combination of $\boldsymbol{\theta}$ which can be viewed as of the empirical. Bayes type. We consider the situation of n jointly distributed binary variates as described by HÖSCHELE *et al.* (1986). The multivariate empirical Bayes approach discussed can be viewed as a generalization of univariate results of HARVILLE & MEE (1984). The paper includes sections on theory, computing algorithms and a numerical application.

II. The model

A. *The data*

The records can be arranged into an $s \times 2^n$ contingency table **Y** where the rows ($j = 1, 2, \ldots, s$) represent sub-populations and the columns ($k = 1, 2, \ldots, 2^n$) are categories of response ; category k is designated by an n-bit digit with a 0 or a 1 for attributes coded [0] or [1], respectively, in trait i ($i = 1, 2, \ldots n$). Symbolically, one can write

$$\mathbf{Y}' = [\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_j, \ldots, \mathbf{Y}_s] \qquad [1]$$

where $\mathbf{Y}_j$ is a $2^n \times 1$ column vector such that

$$\mathbf{Y}_j = \sum_r \mathbf{Y}_{jr} \; ; \; r = 1, \ldots, n_{j+} \qquad [2]$$

and $\mathbf{Y}_{jr}$ is a $2^n \times 1$ column vector having a 1 in the category of response and 0's elsewhere. The marginal totals $n_{1+}, n_{2+}, \ldots, n_{j+}, \ldots, n_{s+}$ of each row of **Y** are assumed fixed by sampling and non-null.

B. *The threshold model*

The model used to analyze this joint distribution of discrete variables assumes the existence of underlying variables rendered discrete by a set of abrupt thresholds. This concept, introduced by WRIGHT (1934), has been used by several authors (ROBERTSON & LERNER, 1949 ; DEMPSTER & LERNER, 1950 ; FALCONER, 1965 ; THOMPSON, 1972 ; CURNOW & SMITH, 1975). The probability that observation o of sub-population j responds in category k depends on values taken by n underlying variates ($l_1, l_2, \ldots, l_n$) in relation to fixed thresholds ($\tau_1, \tau_2, \ldots, \tau_n$). The underlying variates are written as

$$l_{ijo} = \eta_{ij} + \varepsilon_{ijo} \qquad \begin{matrix} i = 1, 2, \ldots, n \\ j = 1, 2, \ldots, s \\ o = 1, 2, \ldots, n_{j+} \end{matrix} \qquad [3]$$

where η_{ij} is a location parameter and ε_{ijo} is a residual. Along the lines of a polygenic inheritance model, it is assumed that the residuals follow the multivariate normal distribution :

$$\begin{bmatrix} \varepsilon_{ijo} \\ \\ \varepsilon_{i'jo} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{\varepsilon_i} & & r_{ii'} \sigma_{\varepsilon_i} \sigma_{\varepsilon_{i'}} \\ \\ r_{ii'} \sigma_{\varepsilon_i} \sigma_{\varepsilon_{i'}} & & \sigma^2_{\varepsilon_{i'}} \end{bmatrix} \right) \qquad [4]$$

where $r_{ii'}$ is a residual correlation and σ_{ε_i} is the residual standard deviation of underlying variate i. Further, it is assumed that Cov ($\varepsilon_{ijo}, \varepsilon_{i'j'o'}$) = 0 unless $o = o'$ and $j = j'$.

Conditionally on the parameters η_{ij}, the probability that on observation in sub-population j responds in category of response k can be written as

$$P_{j, [w_1^{[k]}, \ldots, w_n^{[k]}]} = \text{Prob} \left\{ \bigcap_{i=1}^{n} (-1)^{w_i^{[k]}} l_{ijo} < (-1)^{w_i^{[k]}} \tau_i \mid \eta_{ij} \right\} \qquad [5]$$

where $w_1^{[k]},\ldots, w_n^{[k]}$ is an n-bit digit indicating the category of response, with $w_i^{[k]} = 0$ or 1, depending on whether attribute coded [0] or [1] in trait i is observed. HÖSCHELE *et al.* (1986) showed that [5] is equal to

$$P_{jk} = \Phi_n(\boldsymbol{\mu}_j^{[k]} ; \mathbf{R}^{[k]}) \qquad [6]$$

where for simplicity the n-bit digit is replaced by k, Φ_n is the n-dimensional normal distribution function,

$$\boldsymbol{\mu}_j^{[k]} = \{ (-1)^{w_i^{[k]}} \mu_{ij} \} , \; i = 1, \ldots, n$$

is an n × 1 row vector, and

$$\mu_{ij} = (\tau_i - \eta_{ij})/\sigma_{e_i}$$

is the distance between the threshold for the *ith* conceptual variate and the location parameter η_{ij} expressed in units of residual standard deviation. Finally, the matrix $\mathbf{R}^{[k]}$ is a matrix of functions of residual correlations with typical element

$$r_{ii'}^{[k]} = (-1)^{w_i^{[k]} - w_{i'}^{[k]}} r_{ii'}.$$

C. *Sources of variation*

Because of the assumption of multivariate normality, it is reasonable to adopt linear models to describe the underlying variates so we write

$$\boldsymbol{\mu}_i = \mathbf{X}_i\boldsymbol{\beta}_i + \mathbf{Z}_i\mathbf{u}_i \qquad i = 1, \ldots, n \qquad [7]$$

where $\boldsymbol{\mu}_i$ is an s × 1 column vector of elements $\mu_{ij}(j = 1, \ldots, s)$, $\mathbf{X}_i$ ($\mathbf{Z}_i$) is a known incidence matrix of order s × p_i (s × q), $\boldsymbol{\beta}_i$ is a vector of « fixed » effects and $\mathbf{u}_i$ is a vector of « random » effects. In animal breeding, the $\boldsymbol{\beta}$'s often are nuisance environmental parameters (herd, year, season, age of dam) or effects of genetic populations (lines, generations, groups). The $\mathbf{u}$'s can represent breeding values, producing abilities or, typically, transmitting abilities of sires. Model [7] can be put more compactly as

$$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} = \mathbf{L}\,\boldsymbol{\theta} \qquad [8]$$

with

$$\boldsymbol{\mu}' = [\boldsymbol{\mu}'_1, \ldots, \boldsymbol{\mu}'_i, \ldots, \boldsymbol{\mu}'_n]$$
$$\boldsymbol{\beta}' = [\boldsymbol{\beta}'_1, \ldots, \boldsymbol{\beta}'_i, \ldots, \boldsymbol{\beta}'_n]$$
$$\mathbf{u}' = [\mathbf{u}'_1, \ldots, \mathbf{u}'_i, \ldots, \mathbf{u}'_n]$$
$$\mathbf{X} = \oplus\, \mathbf{X}_i \; ; \; \mathbf{Z} = \oplus\, \mathbf{Z}_i$$
$$\mathbf{L} = [\mathbf{X}, \mathbf{Z}] \; ; \; \boldsymbol{\theta}' = [\boldsymbol{\beta}', \mathbf{u}']$$

D. *Conditional distribution of the records*

Given $\boldsymbol{\theta}$ and $\mathbf{R}$, the vectors $\mathbf{Y}_j$ are conditionally independent following the multinomial distribution

$$f(\mathbf{Y}_j \mid \boldsymbol{\theta}, \mathbf{R}) \propto \prod_k P_{jk}^{n_{jk}} \qquad [9]$$

where the P_{jk}'s are the multivariate normal integrals in [6].

III. Methods of inference

As in other studies dealing with genetic evaluation of animals (RONNINGEN, 1971 ; DEMPFLE, 1977 ; LEFORT, 1980 ; GIANOLA & FOULLEY, 1983 ; GIANOLA & FERNANDO, 1986), a Bayesian procedure is adopted here. Because of the assumption of polygenic inheritance used to justify [3], it is reasonable to assume, *a priori*, that $\mathbf{u}$ in [8] follows the multivariate normal distribution

$$\mathbf{u} \mid \boldsymbol{\Sigma}_u \sim N(\mathbf{0}, \boldsymbol{\Sigma}_u) \qquad [10]$$

With $\mathbf{u}$ partitioned as in the equations subsequent to [8], one can write

$$\boldsymbol{\Sigma}_u = \mathbf{G} \otimes \mathbf{A} \qquad [11]$$

where

- $\mathbf{G}$ is an $n \times n$ matrix of « u » variance and covariance components ; in many applications, $\mathbf{G}$ is a matrix of genetic variances and covariances, and
- $\mathbf{A}$ is a $q \times q$ symmetric matrix with elements equal to twice Malecot's coefficients of parentage.

It is assumed *a priori* that $\boldsymbol{\beta}$ follows a uniform distribution so as to reflect complete ignorance about this vector (BOX & TIAO, 1973) ; this corresponds to a « fixed » $\boldsymbol{\beta}$ vector in a frequentist analysis. Further, we assume *a priori* that $\boldsymbol{\beta}$ and $\mathbf{u}$ are independent so

$$f(\boldsymbol{\theta} \mid \boldsymbol{\Sigma}) = f(\mathbf{u} \mid \boldsymbol{\Sigma}_u) \cdot f(\boldsymbol{\beta}) \propto f(\mathbf{u} \mid \boldsymbol{\Sigma}_u) \qquad [12]$$

Let now $\mathbf{g}$ be a column vector containing the $n(n+1)/2$ variances and covariances in $\mathbf{G}$, and $\mathbf{r}$ be the column vector containing the $n(n-1)/2$ residual correlations in $\mathbf{R}$. Further, let $\boldsymbol{\gamma}' = [\mathbf{g}', \mathbf{r}']$ represent all non-trivial dispersion parameters. The joint posterior distribution of all unknowns, i.e., $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$, can be written using [9] and [12] as

$$f(\boldsymbol{\theta}, \boldsymbol{\gamma} \mid \mathbf{Y}) \propto f(\mathbf{Y} \mid \boldsymbol{\theta}, r) \cdot f(\mathbf{u} \mid \mathbf{g}) \cdot f(\mathbf{g}, \mathbf{r}) \qquad [13]$$

where $f(\mathbf{g}, \mathbf{r})$ is the joint prior density of the dispersion parameters.

From the viewpoint of genetic evaluation of animals, the parameters of interest are in $\boldsymbol{\theta}$ (and sometimes only in $\mathbf{u}$), in which case $\boldsymbol{\gamma}$ should be considered as a nuisance vector. For example, sires are usually evaluated from estimated linear combinations of $\boldsymbol{\beta}$ and $\mathbf{u}$ (HENDERSON, 1973) ; if a quadratic loss function is employed, the corresponding Bayesian estimator is the posterior mean of the appropriate linear combination. Further, if k out of m candidates are to be selected, ranking individuals using the posterior mean maximizes expected genetic progress (GOFFINET & ELSEN, 1984 ; GIANOLA & FERNANDO, 1986).

The calculation of $E(\boldsymbol{\theta} \mid \mathbf{Y})$ involves integrating $\boldsymbol{\gamma}$ out of [13] but this, in general, is extremely difficult if not impossible to do. Hence, it is necessary to consider alternative estimators. One possibility would be to consider modal estimators of $\boldsymbol{\theta}$, i.e., the values that have maximum density given a specified posterior distribution. Several distributions and modes can be considered : 1) the $\boldsymbol{\theta}$ mode of $f(\boldsymbol{\theta} \mid \mathbf{Y})$, which is difficult to obtain for the reasons mentioned above ; 2) the $\boldsymbol{\theta}$ mode of the joint posterior density [13] ; or 3) the mode of $f(\boldsymbol{\theta} \mid \mathbf{Y}, \boldsymbol{\gamma} = \boldsymbol{\gamma}^*)$, where $\boldsymbol{\gamma}^*$ is some value of $\boldsymbol{\gamma}$. In principle, these approaches lead to different point estimates of $\boldsymbol{\theta}$.

Procedure (2) corresponds to the estimators described by LINDLEY & SMITH (1972) for the multivariate normal case. Because in many instances this procedure leads to trivial point estimates (HARVILLE, 1977 ; THOMPSON, 1980), we do not consider it any further. In this paper, we adopt procedure (3) with $\boldsymbol{\gamma}^*$ being the mode of the marginal posterior distribution $f(\boldsymbol{\gamma} \mid \mathbf{Y})$. This is based on O'HAGAN (1980) who stated : « ...one should (a) estimate variance components by the mode of their marginal distribution after integrating out the other parameters, then (b) estimate the remaining parameters by the mode of their conditional distribution given that the variance parameters have the values obtained in (a) ». When a uniform prior distribution is adopted for $\boldsymbol{\gamma}$, the mode $\boldsymbol{\gamma}^*$ is the marginal maximum likelihood estimator (MALECOT, 1947) found by maximizing $f(\mathbf{Y} \mid \boldsymbol{\gamma})$ with respect to the dispersion parameters. Under multivariate normality, this corresponds to the restricted maximum likelihood estimator of $\boldsymbol{\gamma}$ (HARVILLE, 1974). Further, the point estimator of $\boldsymbol{\theta}$ so obtained can be viewed as belonging to the class of empirical Bayes estimators (CASELLA, 1985). This mode of reasoning has also been employed by other workers in multivariate normal (GIANOLA *et al.*, 1986) and discrete (HARVILLE & MEE, 1984 ; STIRATELLI *et al.*, 1984) settings. Finally, the mode $\boldsymbol{\theta}^*$ of the joint posterior distribution $f(\boldsymbol{\theta} \mid \mathbf{Y}, \boldsymbol{\gamma}^*)$ can be viewed as an approximation to the mode of $f(\boldsymbol{\theta} \mid \mathbf{Y})$ (BOX & TIAO, 1973). This can be established by writing

$$f(\boldsymbol{\theta} \mid \mathbf{Y}) = \int_{R_\gamma} f(\boldsymbol{\theta} \mid \mathbf{Y}, \boldsymbol{\gamma}) \cdot f(\boldsymbol{\gamma} \mid \mathbf{Y}) d\boldsymbol{\gamma} \qquad [14A]$$

from which it follows that $f(\boldsymbol{\theta} \mid \mathbf{Y}) = E[f(\boldsymbol{\theta} \mid \mathbf{Y}, \boldsymbol{\gamma})]$, where the expectation is taken with respect to $f(\boldsymbol{\gamma} \mid \mathbf{Y})$. If this distribution is symmetrical or quasi-symmetrical about its mode $\boldsymbol{\gamma}^*$, it follows that $f(\boldsymbol{\theta} \mid \mathbf{Y}) \simeq f(\boldsymbol{\theta} \mid \mathbf{Y}, \boldsymbol{\gamma}^*)$. Equivalently, the approximation can be justified using the first-order expansion

$$f(\boldsymbol{\theta} \mid \mathbf{Y}, \boldsymbol{\gamma}) \simeq f(\boldsymbol{\theta} \mid \mathbf{Y}, \boldsymbol{\gamma}^*) + (\boldsymbol{\gamma} - \boldsymbol{\gamma}^*)' \left. \frac{[\delta f(\boldsymbol{\theta} \mid \mathbf{Y}, \boldsymbol{\gamma})]}{\delta \boldsymbol{\gamma}} \right| \boldsymbol{\gamma} = \boldsymbol{\gamma}^* \qquad [14B]$$

and then taking expectation with respect to $f(\boldsymbol{\gamma} \mid \mathbf{Y})$. The second term vanishes only if $E(\boldsymbol{\gamma} \mid \mathbf{Y}) = \boldsymbol{\gamma}^*$; this holds when the posterior distribution of $\boldsymbol{\gamma}$ is symmetric, or to first order approximation, when the mode is close to the mean.

IV. Estimation of location parameters

As pointed out earlier, the point estimator of $\boldsymbol{\theta}$ is the statistic $\boldsymbol{\theta}^*$ $(\boldsymbol{\gamma}^*)$ such that

$$ln\,[f(\boldsymbol{\theta}^*(\boldsymbol{\gamma}^*) \mid \mathbf{Y}, \boldsymbol{\gamma}^*] = \underset{\boldsymbol{\theta}}{\text{Max}}\; ln\; f(\boldsymbol{\theta} \mid \mathbf{Y}, \boldsymbol{\gamma} = \boldsymbol{\gamma}^*) \qquad [15]$$

where $\boldsymbol{\gamma}^*$ is defined by

$$ln\; f(\boldsymbol{\gamma}^* \mid \mathbf{Y}) = \underset{\boldsymbol{\gamma}}{\text{Max}}\; ln\; f(\boldsymbol{\gamma} \mid \mathbf{Y}) \qquad [16]$$

Using [9] and [12], one can write

$$f(\boldsymbol{\theta} \mid \mathbf{Y}, \boldsymbol{\gamma}) \propto f(\mathbf{Y} \mid \boldsymbol{\theta}, \mathbf{r}) \cdot f(\mathbf{u} \mid \mathbf{g}) \qquad [17]$$

where $\mathbf{r}$ and $\mathbf{g}$ are the components of $\boldsymbol{\gamma}$. Because the likelihood is product multinomial and the prior distribution is multivariate normal, one can write the log of [17] as

$$L(\boldsymbol{\theta}) = \sum_j \sum_k n_{jk}\, lnP_{jk} - \frac{1}{2}\, \mathbf{u}'\, (\mathbf{G}^{-1} \otimes \mathbf{A}^{-1})\mathbf{u} + \text{constant} \qquad [18]$$

Maximization of [18] with respect to $\boldsymbol{\theta}$ can be done via the Newton-Raphson algorithm, and HÖSCHELE *et al.* (1986) have shown that this involves iteration with equations

$$\begin{bmatrix} \mathbf{X}_1'\mathbf{W}_{11}^{[t-1]}\mathbf{X}_1 & \mathbf{X}_1'\mathbf{W}_{11}^{[t-1]}\mathbf{Z}_1 & \mathbf{X}_1'\mathbf{W}_{1n}^{[t-1]}\mathbf{X}_n & \mathbf{X}_1'\mathbf{W}_{1n}^{[t-1]}\mathbf{Z}_n \\ \mathbf{Z}_1'\mathbf{W}_{11}^{[t-1]}\mathbf{X}_1 & \mathbf{Z}_1'\mathbf{W}_{11}^{[t-1]}\mathbf{Z}_1 + \boldsymbol{\Sigma}_u^{11} & \mathbf{Z}_1'\mathbf{W}_{1n}^{[t-1]}\mathbf{X}_n & \mathbf{Z}_1'\mathbf{W}_{1n}^{[t-1]}\mathbf{Z}_n + \boldsymbol{\Sigma}_u^{1n} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{X}_n'\mathbf{W}_{n1}^{[t-1]}\mathbf{X}_1 & \mathbf{X}_n'\mathbf{W}_{n1}^{[t-1]}\mathbf{Z}_1 & \mathbf{X}_n'\mathbf{W}_{nn}^{[t-1]}\mathbf{X}_n & \mathbf{X}_n'\mathbf{W}_{nn}^{[t-1]}\mathbf{Z}_n \\ \mathbf{Z}_n'\mathbf{W}_{n1}^{[t-1]}\mathbf{X}_1 & \mathbf{Z}_n'\mathbf{W}_{n1}^{[t-1]}\mathbf{Z}_1 + \boldsymbol{\Sigma}_u^{n1} & \mathbf{Z}_n'\mathbf{W}_{nn}^{[t-1]}\mathbf{X}_n & \mathbf{Z}_n'\mathbf{W}_{nn}^{[t-1]}\mathbf{Z}_n + \boldsymbol{\Sigma}_u^{nn} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1^{[t]} \\ \mathbf{u}_1^{[t]} \\ \vdots \\ \boldsymbol{\beta}_n^{[t]} \\ \mathbf{u}_n^{[t]} \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{X}_1'\mathbf{W}_{11}^{[t-1]}\,\mathbf{y}_1^{[t-1]} + \sum_{i \neq 1} & \mathbf{X}_1'\mathbf{W}_{1i}^{[t-1]}\,\mathbf{y}_i^{[t-1]} \\ \mathbf{Z}_1'\mathbf{W}_{11}^{[t-1]}\,\mathbf{y}_1^{[t-1]} + \sum_{i \neq 1} & \mathbf{Z}_1'\mathbf{W}_{1i}^{[t-1]}\,\mathbf{y}_i^{[t-1]} \\ \vdots & \vdots \\ \mathbf{X}_n'\mathbf{W}_{nn}^{[t-1]}\mathbf{y}_n^{[t-1]} + \sum_{i \neq n} & \mathbf{X}_n'\mathbf{W}_{ni}^{[t-1]}\mathbf{y}_i^{[t-1]} \\ \mathbf{Z}_n'\mathbf{W}_{nn}^{[t-1]}\mathbf{y}_n^{[t-1]} + \sum_{i \neq n} & \mathbf{Z}_n'\mathbf{W}_{ni}^{[t-1]}\mathbf{y}_i^{[t-1]} \end{bmatrix} \qquad [19]$$

where t is iterate number and

$$\begin{bmatrix} \mathbf{y}_1^{[t]} \\ \vdots \\ \mathbf{y}_i^{[t]} \\ \vdots \\ \mathbf{y}_n^{[t]} \end{bmatrix} \begin{bmatrix} \mathbf{X}_1\boldsymbol{\beta}_1^{[t]} + \mathbf{Z}_1\mathbf{u}_1^{[t]} \\ \vdots \\ \mathbf{X}_i\boldsymbol{\beta}_i^{[t]} + \mathbf{Z}_i\mathbf{u}_i^{[t]} \\ \vdots \\ \mathbf{X}_n\boldsymbol{\beta}_n^{[t]} + \mathbf{Z}_n\mathbf{u}_n^{[t]} \end{bmatrix} + \begin{bmatrix} \mathbf{W}_{11}^{[t]} \ldots \mathbf{W}_{1n}^{[t]} \\ \vdots \quad \vdots \\ \mathbf{W}_{i1}^{[t]} \ldots \mathbf{W}_{in}^{[t]} \\ \vdots \quad \vdots \\ \mathbf{W}_{n1}^{[t]} \ldots \mathbf{W}_{nn}^{[t]} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{v}_1^{[t]} \\ \vdots \\ \mathbf{v}_i^{[t]} \\ \vdots \\ \mathbf{v}_n^{[t]} \end{bmatrix} \qquad [20]$$

are « working » vectors. In [19] and [20] above, the $\mathbf{W}_{ii'}$ arrays are diagonal matrices and the $\mathbf{v}_i$'s ares $s \times 1$ vectors ; formulae to calculate elements of these matrices and vectors are given by HÖSCHELE *et al.* (1986). Further, the $\boldsymbol{\Sigma}_u^{ii'}$ sub-matrices are appropriate blocks of $\boldsymbol{\Sigma}_u^{-1}$ (evaluated at $\mathbf{g}$)*. The parallel between (19) and the multiple-trait mixed model equations (HENDERSON & QUAAS, 1976) is remarkable.

The matrix of second derivatives of the log-posterior in [18] with respect to $\boldsymbol{\theta}$ is the negative of the coefficient matrix in [19]. This Hessian matrix is negative definitive provided the matrices $\mathbf{G}$ and $\mathbf{R}$ defined earlier and evaluated at $\boldsymbol{\gamma}^*$ are positive definite ; this is shown in Annex A. Therefore, the Newton-Raphson algorithm converges to a unique maximum of the log-posterior density if it exists (DAHLQUIST & BJORCK, 1974 ; EVERITT, 1984). Computations involve a double-iterative scheme with [19] and with the equations used to calculate $\boldsymbol{\gamma}^*$. We return to this in a later section of this article.

It is useful to point out that the matrix

$$C(\boldsymbol{\gamma}) = -\left\{\frac{\delta^2 L(\boldsymbol{\theta})}{\delta\boldsymbol{\theta}\,\delta\boldsymbol{\theta}'}\right\}^{-1}\Bigg|\ \boldsymbol{\theta} = \boldsymbol{\theta}^*(\boldsymbol{\gamma}) \qquad [21]$$

evaluated at the modal value θ^* (γ^*) gives an expression for the asymptotic covariance matrix of the posterior distribution of θ (Cox & HINKLEY, 1975, p. 400 ; BERGER, 1985, p. 224).

V. Estimation of genetic variances and covariances

Let

$$P(\mathbf{g}, \mathbf{r}, \mathbf{Y}) = \frac{\delta}{\delta \mathbf{g}} ln\, f(\mathbf{g}, \mathbf{r} \mid \mathbf{Y}) \qquad [22]$$

Calculating [22] requires first the integration of the joint posterior distribution of all unknowns with respect to θ, that is

$$P(\mathbf{g}, \mathbf{r}, \mathbf{Y}) = \frac{\delta}{\delta \mathbf{g}} ln \int_{R_\beta} \int_{R_u} f(\mathbf{g}, \mathbf{r}, \boldsymbol{\beta}, \mathbf{u} \mid \mathbf{Y})\, d\boldsymbol{\beta}\, d\mathbf{u} \qquad [23]$$

It is shown in Annex B that irrespective of the form of the density involved in [23], the above integration leads to the expression :

$$P(\mathbf{g}, \mathbf{r}, \mathbf{Y}) = E_c \left[\frac{\delta}{\delta \mathbf{g}} ln\, f(\mathbf{u} \mid \mathbf{g}) \right] + \frac{\delta}{\delta \mathbf{g}} ln\, f(\mathbf{g}) \qquad [24]$$

where E_c indicates expectation taken with respect to the conditional distribution $f(\mathbf{u} \mid \mathbf{Y}, \mathbf{g}, \mathbf{r})$, and $f(\mathbf{g})$ is the prior density of the vector of genetic variances and covariances. To satisfy (16), we need to set $P(\mathbf{g}, \mathbf{r}, \mathbf{Y}) = 0$, which leads to a nonlinear system in $\mathbf{g}$. An important simplification arises if E_c in [24] is evaluated at $\mathbf{g}^{[t]}$, a vector representing the genetic (co)variances at iteration t. Then,

$$\int_{R_u} \left[\frac{\delta}{\delta \mathbf{g}} ln\, f(\mathbf{u} \mid \mathbf{g}) \right] f(\mathbf{u} \mid \mathbf{Y}, \mathbf{g}^{[t]}, \mathbf{r})\, d\mathbf{u} = \frac{\delta}{\delta \mathbf{g}} \int_{R_u} ln\, f(\mathbf{u} \mid \mathbf{g})\, f(\mathbf{u} \mid \mathbf{Y}, \mathbf{g}^{[t]}, \mathbf{r})\, d\mathbf{u} \qquad [25]$$

Hence, at iteration t,

$$E_c^{[t]} \left\{ \frac{\delta}{\delta \mathbf{g}} ln\, f(\mathbf{u} \mid \mathbf{g}) \right\} = \frac{\delta}{\delta \mathbf{g}} E_c^{[t]} \{ ln\, f(\mathbf{u} \mid \mathbf{g}) \} \qquad [26]$$

Collecting [22], [24] and [26], it follows that at iteration t

$$\frac{\delta}{\delta \mathbf{g}} ln\, f(\mathbf{g}, \mathbf{r} \mid \mathbf{Y}) = \frac{\delta}{\delta \mathbf{g}} E_c \{ ln\, f(\mathbf{u} \mid \mathbf{g}) \} + \frac{\delta}{\delta \mathbf{g}} ln\, f(\mathbf{g}) \qquad [27]$$

The above result implies that whenever a flat prior is used for $\mathbf{g}$, maximization of the joint posterior distribution of all variances and covariances with respect to $\mathbf{g}$ can be done by maximizing $E_c^{[t]} \{ ln\, f(\mathbf{u} \mid \mathbf{g}) \}$ at each iterate. More general situations, e.g., using informative prior distributions for $\mathbf{g}$, are dealt with in the discussion section of this paper.

From [10] and [11]

$$f(\mathbf{u} \mid \mathbf{g}) = (2\pi)^{-nq/2} \mid \boldsymbol{\Sigma}_u \mid^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \mathbf{u}' \boldsymbol{\Sigma}_u^{-1} \mathbf{u}\right) \qquad [28]$$

with $| \Sigma_u | = | G |^q . | A |^n$ (ANDERSON, 1984, p. 600), and $\Sigma_u^{-1} = G^{-1} \otimes A^{-1}$. Now,

$$E_c \{ln\ f(u \mid g)\} = \text{constant} - (q/2)\ ln\ | G | - \frac{1}{2} \{ E_c (u' \Sigma_u^{-1} u)\}$$

$$= \text{constant} - (q/2)\ ln\ | G | - \frac{1}{2} \text{tr} \{G^{-1} E_c (D)\} \quad [30]$$

where $D = \{u_i' A^{-1} u_{i'}\}$ $(i = 1, ..., n, i' = 1, ..., n)$ is an $n \times n$ matrix. Using Lemma 3.2.2 of ANDERSON (1984, p. 62), expression [30] is maximum at

$$G^* = E_c (D)/q \quad [31]$$

with the typical element of G^* being

$$g_{ii'} = E_c (u_i' A^{-1} u_{i'})/q \quad [32]$$

and this holding at each iteration. Under multivariate normality, the above formulae lead to the iterative algorithm

$$g_{ii'}^{[t+1]} = \{\hat{u}_i' A^{-1} \hat{u}_{i'} + \text{tr} (A^{-1} C_{ii'})\}^{[t]}/q \quad [33]$$

where $\hat{u}_i = E (u_i \mid Y, \gamma)$ and $C_{ii'} = \text{Cov} (u_i, u_{i'} \mid Y, \gamma)$. This is precisely the expectation-maximization algorithm (DEMPSTER *et al.*, 1977) applied to a multiple trait setting (HENDERSON, 1984). In the multivariate discrete problem addressed in this paper, it is not possible to evaluate [32] explicitly. Hence, as suggested by other authors (HARVILLE & MEE, 1984 ; STIRATELLI *et al.*, 1984), we replace $\hat{u}_i^{[t]}$ in [33] by $u^* (\gamma = \gamma^{[t]})$, the mode of $f(u \mid Y, \gamma))$ evaluated at $\gamma^{[t]} = (g'^{[t]}, r')'$ and $C_{ii'}^{[t]}$ by $C_{ii'} (\gamma = \gamma^{[t]})$. With these approximations, [33] generalizes the results for a univariate threshold model presented by HARVILLE & MEE (1984). As pointed out earlier, [33] holds for the case where a flat prior distribution is used for g.

As shown in Annex C, if X in [8] is a full-column rank matrix (this is not restrictive because a reparameterization to full rank always exists) and if $G^{[t]}$ is positive-definite, then $G^{[t+1]}$ calculated with [33] is also positive-definite. This property is important in the construction of predictors of breeding values as pointed out by HILL & THOMPSON (1978) and FOULLEY & OLLIVIER (1986). Finally, equation [16] is satisfied at

$$g_{ii'}^* = [u_i^{*'} (\gamma^*) A^{-1} u_{i'}^* (\gamma^*) + \text{tr} (A^{-1} C_{ii'} (\gamma^*)]/q \quad [34]$$

This procedure is general and can be applied to models with several sets of random effects (FOULLEY *et al.*, 1986).

VI. Estimation of residual correlations

Define

$$Q (g, r, Y) = \frac{\delta}{\delta r} ln\ f (g, r, Y) \quad [35]$$

Using a reasoning similar to the one employed in the preceding section it can be shown (Annex B) that the *ith* element of the vector in [35] takes the form

$$Q_i (g, r, Y) = \left[\frac{\delta}{\delta r_i} ln\ f (Y \mid \theta, \gamma)\right] - \frac{1}{2} \text{tr} \left[\frac{\delta M (\gamma)}{\delta r_i} \cdot C (\gamma)\right]_{\theta = \theta^*} + \frac{\delta}{\delta r_i} ln\ f (r) \quad [36]$$

where $\mathbf{M}(\gamma)$ is the coefficient matrix in [19] excluding the contributions from the prior distribution, i.e., without the Σ_u^{ij} sub-matrices.

In many applications, the form of the prior distribution of $\mathbf{r}$ is not very important as the residual correlations can be well estimated from the body of data used in the analysis. In this study, we adopted a uniform prior distribution for $\mathbf{r}$ so the last term of [36] vanishes. The first term represents the contribution of the likelihood function evaluated at $\boldsymbol{\theta}^*$, the mode of $f(\boldsymbol{\theta} \mid \mathbf{Y}, \gamma)$. The second term stems from a local integration (in the neighborhood of $\boldsymbol{\theta}^*$) to second order with respect to $\boldsymbol{\theta}$. Because calculating the second term involves complex computations, we consider at this point only the first term. This implies that we search for $\mathbf{r}^*$ such that

$$ln\, f(\mathbf{Y} \mid \boldsymbol{\theta}, \mathbf{r}^*)_{\boldsymbol{\theta} = \boldsymbol{\theta}^*} = \underset{\mathbf{r}}{\text{Max}}\, ln\, f(\mathbf{Y} \mid \boldsymbol{\theta}, \mathbf{r})\Big|_{\boldsymbol{\theta} = \boldsymbol{\theta}^*} \qquad [37]$$

which can be viewed as a modification of estimation by maximum likelihood (TALLIS, 1962 ; THOMPSON, 1972 ; ANDERSON & PEMBERTON, 1985).

From [9], the log-likelihood viewed as a function or $\mathbf{r}$ can be written as

$$L(\mathbf{r}) = \sum_j \sum_k n_{jk}\, ln\, P^*_{jk} \qquad [38]$$

where P^*_{jk} is as in [6] with $\mu_j^{*[k]}$ replacing $\mu_j^{[k]}$. From now on, we do not use the *'s on the P's and μ's to simplify notation. Maximization of [38] can be achieved using Fisher's scoring algorithm :

$$-E\left\{\frac{\delta^2 L(\mathbf{r})}{\delta \mathbf{r}\, \delta \mathbf{r}'}\right\}_{\mathbf{r} = \mathbf{r}^{[t-1]}} \Delta \mathbf{r}[t] = \left\{\frac{\delta L(\mathbf{r})}{\delta \mathbf{r}}\right\}_{\mathbf{r} = \mathbf{r}^{[t-1]}} \qquad [39]$$

where $\Delta \mathbf{r}^{[t]} = \mathbf{r}^{[t]} - \mathbf{r}^{[t-1]}$, $\mathbf{r}^{[t]}$ is a solution at iterate t, and the expectation is taken with respect to $f(\mathbf{Y} \mid \boldsymbol{\theta}, \mathbf{r})$. Using a result of PLACKETT (1954), one can write from [6]

$$\frac{\delta P_{jk}}{\delta r_{ef}} = (-1)^{w_e^{[k]} - w_f^{[k]}}\, \phi_2(\mu_{ej}, \mu_{fj} ; r_{ef})\, \Phi_{n-2}[\mathbf{h}^{[k]}_{(\cdot ef)j} ; \mathbf{R}^{[k]}_{\cdot ef}] \qquad [40]$$

where :

- r_{ef} is the residual correlation between traits e and f ;
- ϕ_2 is a bivariate standard normal density ;
- Φ_{n-2} is the multivariate normal distribution function of order $n - 2$;
- $\mathbf{h}^{[k]}_{(\cdot ef)j} = \{h^{[k]}_{(d \cdot ef)j}\}$ for every d different than e and f is an $(n - 2) \times 1$ vector, with $h^{[k]}_{(d \cdot ef)j} = (-1)^{w_d^{[k]}} h_{(d \cdot ef)j}$

  and $h_{(d \cdot ef)j}$ is the third element of the row vector

  $$[\mu_{fj}, \mu_{ej}, \mu_{dj}] \cdot \mathbf{T}^{-1} ;$$

- $\mathbf{T}$ is the 3×3 upper triangular matrix of the Cholesky decomposition $\mathbf{T}'\mathbf{T}$ of the residual correlation matrix between traits f, e, d and taken in that order ;
- $\mathbf{R}^{[k]}_{\cdot ef}$ is a correlation matrix of order $n - 2$, with typical element

  $$r^{[k]}_{cd \cdot ef} = (-1)^{w_c^{[k]} - w_d^{[k]}} \cdot r_{cd \cdot ef}$$

  and $r_{cd \cdot ef}$ is the partial residual correlation between c and d with e and f fixed.

Applying [40], one can write

$$\frac{\delta L\,(\mathbf{r})}{\delta r_{ef}} = \sum_{j=1}^{s} \phi_2\,(\mu_{ej},\, \mu_{fj}\,;\, r_{ef}) \sum_{k=1}^{2^n} (-1)^{w_e^{[k]} - w_f^{[k]}} \frac{n_{jk}}{P_{jk}} \cdot \Phi_{n-2}\,[\mathbf{h}_{(\cdot ef)j}\,;\, \mathbf{R}^{[k]}_{\cdot ef}] \qquad [41]$$

The second derivatives of the log-likelihood can be obtained using Fisher's information measure for a multinomial distribution. This yields

$$-\,E\left\{\frac{\delta^2 L\,(\mathbf{r}}{\delta r_{ef}\delta r_{gh}}\right\} = \sum_{j=1}^{s} n_{j+}\,\phi_2\,(\mu_{ej},\, \mu_{fj}\,;\, r_{ef})\;\phi_2\,(\mu_{gj},\, \mu_{hj}\,;\, r_{gh}) \times$$

$$\sum_{k=1}^{2^n} (-1)^{w_e^{[k]} - w_f^{[k]} + w_g^{[k]} - w_h^{[k]}}\, P_{jk}^{-1}\;\Phi_{n-2}\,\{\mathbf{h}^{[k]}_{\cdot ef)j}\,;\, \mathbf{R}^{[k]}_{\cdot ef}\}\;\Phi_{n-2}\,\{\mathbf{h}^{[k]}_{(\cdot gh)j}\,;\, \mathbf{R}^{[k]}_{\cdot gh}\} \qquad [42]$$

In the case of two binary traits, [41] and [42] reduce to the formulae given by TALLIS (1962).

VII. Computing algorithm

Satisfying [15] and [16] involves computations with a system composed of equations [19], [33] and [39]. Because the three sub-systems are non-linear in the unknowns, an iterative solution is needed. The algorithm considered in this paper can be described as follows :

i) given $\boldsymbol{\gamma}^{[t]}$, calculate $\boldsymbol{\theta}^{[t]} = \boldsymbol{\theta}\,(\boldsymbol{\gamma}^{[t]})$ using equations [19] ;

ii) apply [39] to compute $\mathbf{r}^{[t+1]} = \mathbf{r}\,(\boldsymbol{\theta}^{[t]}, \mathbf{g}^{[t]}, \mathbf{r}^{[t]})$;

iii) compute $\mathbf{g}^{[t+1]}$ using [33] as a function of $\boldsymbol{\theta}^{[t]}$ and of $\mathbf{C}\,(\boldsymbol{\gamma}^{[t]})$.

Because [19] and [39] require « internal » iteration, steps (i) and (ii) can be combined into a single, more rapid process. Let the iterate for $\boldsymbol{\theta}$ or $\mathbf{r}$ be denoted by two indexes (t_1, t_2), where t_1 indicates the number of iterations carried out for $\mathbf{g}$, and t_2 denotes the iterate number for $\boldsymbol{\theta}$ and $\mathbf{r}$, *intra* t_1. With this notation, the modified algorithm becomes :

i) From $\mathbf{g}^{[t_1]}$ and $\mathbf{r}^{[t_1,\, t_2-1]}$, calculate $\boldsymbol{\theta}^{[t_1,\, t_2]}$ with a single iteration of [19], so

$$\boldsymbol{\theta}^{[t_1,\, t_2]} = \boldsymbol{\theta}_1\,(\mathbf{g}^{[t_1]}, \mathbf{r}^{[t_1,\, t_2-1]}) \qquad [43]$$

Having [43], compute with a single iterate of [39]

$$\mathbf{r}^{[t_1,\, t_2]} = \mathbf{r}_1\,(\boldsymbol{\theta}^{[t_1,\, t_2]}, \mathbf{g}^{[t_1]}, \mathbf{r}^{[t_1,\, t_2-1]}) \qquad [44]$$

and perform a new iteration on $\boldsymbol{\theta}$ so

$$\boldsymbol{\theta}^{[t_1,\, t_2+1]} = \boldsymbol{\theta}_1\,(\mathbf{g}^{[t_1]}, \mathbf{r}^{[t_1,\, t_2]}) \qquad [45]$$

The process stops at iterate $\bar{t}_2$ when

$$\{\Delta'_{\boldsymbol{\theta}}{}^{[t_1,\, t_2]}\,\Delta_{\boldsymbol{\theta}}{}^{[t_1,\, t_2]}/\dim\,(\boldsymbol{\theta})\}^{1/2} \leqslant \varepsilon_2$$

where ε_2 is an arbitrarilly small positive number.

ii) The second step pertains to the calculation of $\mathbf{g}^{[t_1+1]}$ as a function of $\boldsymbol{\theta}^{[t_1,\, \bar{t}_2]}$ and of the corresponding inverse of the coefficient matrix, that is

$$\mathbf{g}^{[t_1+1]} = \mathbf{g}\,\{\boldsymbol{\theta}^{[t_1,\, \bar{t}_2]},\; \mathbf{C}\,(\mathbf{g}^{[t_1]}, \boldsymbol{\theta}^{[t_1,\, \bar{t}_2]}, \mathbf{r}^{[t_1,\, \bar{t}_2]})\} \qquad [46]$$

At this point, we return to (i) of the modified algorithm and calculate [43] with $t_1 + 1$ as the index for $\mathbf{g}$, and $\mathbf{r}^{[t_1+1,\,0]} = \mathbf{r}^{[t_1,\,\bar{t}_2]}$ as the « new » residual correlation. The overall process stops at « main » iterate $t_1 = \bar{t}_1$ when

$$\{\boldsymbol{\Delta}'^{[t_1,\,\bar{t}_2]}_{\boldsymbol{\theta}}\,\boldsymbol{\Delta}^{[t_1,\,\bar{t}_2]}_{\boldsymbol{\theta}}/\dim(\boldsymbol{\theta})\}^{\frac{1}{2}} \leq \varepsilon_1 \qquad [47]$$

For example, as suggested by HARVILLE & MEE (1984) one could choose not to iterate on step (i), by calculating a single pass $\boldsymbol{\theta}^{[t_1,\,1]}$, $\mathbf{r}^{[t_1,\,1]}$ so as to save time. In fact, for values of ϵ_2 of the order of 10^{-3}, the number of iterations required for $\mathbf{g}$ can be reduced considerably. In this study, we opted to calculate the first iterates for $\mathbf{g}$ using $\boldsymbol{\theta}$ values close to the mode of $f(\boldsymbol{\theta} \mid \mathbf{Y}, \mathbf{g}^{[t_1]}, \mathbf{r}^{[t_1,\,\bar{t}_2]})$. Because in the examples so far examined $\mathbf{r}$ seems to converge rapidly, it would be possible to stop calculating the residual correlations early during iteration, or to revise their values only periodically during the process. It should be mentioned that the general properties of convergence, e.g., convergence to a unique global maximum, are not known to hold for this modified algorithm.

VIII. Numerical application

A. *The records*

In order to illustrate the procedures, an example involving 30 bulls progeny tested for calving difficulty (trait A) and perinatal mortality (trait B), is considered ; it is assumed that average progeny group size is 100 calves, 50 males and 50 females. Calving difficulty is scored as an « all or none » trait ([0] : unassisted or mildly assisted births ; [1] : mechanical assistance or caesarean section). Perinatal mortality classes are [0] for live, viable calves, or [1] for calves that are born dead or of doubtful viability. The data were obtained by simulating 2 conceptual underlying variates l_A and l_B, corresponding to traits A and B, respectively, following the bivariate normal distribution described below. The assignment ot categories of responses was by reference to fixed thresholds τ_A and τ_B :

$$\begin{aligned}
(l_A \leq \tau_A) \cap (l_B \leq \tau_B) &\Rightarrow [00]\\
(l_A \leq \tau_A) \cap (l_B > \tau_B) &\Rightarrow [01]\\
(l_A > \tau_A) \cap (l_B \leq \tau_B) &\Rightarrow [10]\\
(l_A > \tau_A) \cap (l_B > \tau_B) &\Rightarrow [11]
\end{aligned}$$

For reasons of simplicity, the thresholds were set at 0, and the parameterization was on $\eta_{ij} = \tau - \mu_{ij}$, with τ null (see section **II, B**). This implies that factors increasing the η's would increase the probability of response in the categories coded as [1]. The linear model employed to describe the underlying variates was

$$l^X_{mnop} = T^X_m + S^X_n + u^X_o + e^X_{mnop},\ X = A, B \qquad [48]$$

T_m : effect of calving season m ($m = 1, 2$)

S_n : effect of sex of calf n ($n = 1, 2$ for males and females, respectively)

u_o : effect of sire o ($o = 1, 2, \ldots, 30$)

e_{mnop} : residual effect.

A reparameterization to full-rank was achieved by putting $\beta_1 = T_1 + S_2$, $\beta_2 = T_2 + S_2$ and $\beta_3 = S_1 - S_2$. Based on HÖSCHELE *et al.* (1986), the values chosen for the β's in the parameterization of the η's were

A : $\beta_1 = -1.05$ $\beta_2 = -1.42$ $\beta_3 = .77$

B : $\beta_1 = -.60$ $\beta_2 = -1.36$ $\beta_3 = .20$

For example, the value for calving difficulty of $\beta_3 = .77$, represents the difference in liability between male and female calves. Because the difference is positive, linear combinations of liability variates pertaining to males would have a higher probability of difficult calving than those of females. The sire and residual effects were sampled from independent bivariate normal distributions with null means and respective covariance matrices :

$$\mathbf{G} = \begin{bmatrix} h_A^2/(4 - h_A^2) & r_g h_A h_B/[(4 - h_A^2)(4 - h_B^2)]^{\frac{1}{2}} \\ \text{Symmetric} & h_B^2/(4 - h_B^2) \end{bmatrix}$$

with $h_A^2 = .35$ and $h_B^2 = .05$ as heritabilities and $r_g = 0$ as genetic correlation, and

$$\mathbf{R} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

with $\rho = .35$ as residual correlation.

The 3 000 calves were assigned to the sires at random ; each sire had an equal probability of having the calf included in the corresponding progeny group. The distribution of progeny across seasons of calving was also at random, with probabilities .30 and .70 for seasons 1 and 2, respectively, and independently of the assignment of records to sexes. The sample so obtained is presented in table 1 ; progeny group sizes ranged between 76 and 116. Marginal and joint raw frequencies for the two traits are in table 2 for each of the season × sex combinations. The overall proportions of unassisted births and of calves having good viability were about .79 and .82, respectively. It is interesting to observe that the data in table 2 suggest a slight interaction between season of birth and sex, especially for the response in category [11] (difficult birth, poor vigor). For example, proportions of calves born in season 1 having a difficult birth and poor vigor were .17 and .06 for males and females, respectively ; corresponding proportions in season 2 were .06 and .02. This is purely due to nonlinearity as the model used to simulate the data additive. Linear analyses of 0-1 scores sometimes require including interactions which, biologically speaking, may be non-existent. Statistically, this leads to the specification of highly non-parsimonious models (McCULLAGH & NELDER, 1983), with an unnecessarily large number of parameters. These « interactions » can be more marked at higher incidence. The data in table 2 indicate an association between the two traits ; correlations of .43 of .35 were calculated with formulae of Yule and Pearson, respectively (KENDALL & STUART, 1973, p. 539 ; RUTLEDGE, 1977, p. 395). This was expected as the residual correlation was $\rho = .35$ in the simulation of the data.

TABLE 1

Distribution of records [a] by sire of calf, season of calving and sex of calf.

Sire	No. of progeny	Season 1								Season 2							
		Males				Females				Males				Females			
		[00]	[01]	[10]	[11]	[00]	[01]	[10]	[11]	[00]	[01]	[10]	[11]	[00]	[01]	[10]	[11]
1	92	4	2	4	3	10	5	2	1	20	2	10	1	24	1	1	2
2	99	6	3	4	2	7	0	2	1	23	6	7	5	26	3	3	1
3	101	12	5	3	1	7	4	1	0	32	3	7	0	21	3	1	1
4	104	6	3	5	2	13	3	2	3	18	3	13	3	28	1	0	1
5	107	6	2	5	2	11	7	1	1	30	4	5	2	28	3	0	0
6	111	10	5	0	1	12	2	0	0	34	3	9	1	30	2	2	0
7	100	3	0	9	3	11	5	3	0	24	2	8	3	25	1	3	0
8	92	7	3	3	4	6	4	3	0	27	0	3	2	28	1	1	0
9	104	7	3	1	5	12	2	0	1	20	2	9	1	36	2	3	0
10	98	12	1	5	4	7	1	3	3	14	1	18	1	28	0	0	0
11	95	5	2	5	5	2	6	4	0	19	3	8	2	25	5	3	1
12	76	7	2	1	2	8	1	1	2	19	0	6	4	18	2	3	0
13	95	5	2	2	4	9	2	3	0	23	2	10	2	28	0	2	1
14	95	8	4	3	3	10	2	0	1	17	4	1	0	35	6	1	0
15	116	8	3	0	1	19	5	0	0	27	6	4	0	38	5	0	0
16	99	7	4	2	3	5	5	1	1	17	5	10	4	32	0	3	0
17	110	2	2	3	2	7	1	3	2	15	5	14	6	29	4	10	5
18	104	3	4	1	2	8	7	0	1	27	5	9	5	28	2	1	1
19	81	5	3	4	1	3	3	2	1	16	4	6	4	25	3	1	0
20	108	8	2	4	2	16	2	1	1	19	1	5	1	37	7	1	1
21	107	8	2	3	3	16	1	3	1	23	2	9	1	28	4	2	1
22	104	10	0	3	3	10	1	7	1	17	1	9	1	34	2	4	1
23	92	7	4	6	1	7	4	1	3	14	0	9	3	24	4	3	2
24	112	6	2	1	4	10	5	0	2	27	6	6	0	33	5	3	2
25	97	7	1	2	0	7	10	1	0	25	4	3	3	29	3	1	1
26	96	3	1	5	6	2	4	0	0	21	0	18	1	14	4	15	2
27	98	9	2	2	5	5	3	3	1	28	3	4	3	28	0	1	1
28	92	8	2	1	1	8	2	0	0	24	3	3	1	32	5	2	0
29	101	6	2	4	0	9	0	2	0	32	1	4	1	37	3	0	0
30	114	12	3	0	1	13	4	0	0	28	3	2	1	39	6	1	1

(a) [00] : « easy » birth, « good » vigor ; [01] : « easy » birth, poor vigor ; [10] : « difficult » birth, good vigor ; [11] : « difficult » birth, « poor » vigor.

B. *Results*

The data were analyzed with [48], the model used to simulate the records but the analysis was carried out on the μ_{ij} metric (τ-η_{ij}). Computations were carried out as described in **VII** using APL in an IBM PC-XT/370 micro computer with an 8087 co-processor. The first iterate for $\boldsymbol{\theta}$ was obtained solving univariate mixed model equations applied to 0-1 data. The multivariate normal integrals required thereafter were calculated using DUTT's algorithm with 10 or 4 positive roots of Hermite polynomials for one or two dimensions, respectively (DUCROCQ & COLLEAU, 1986). The final solutions

TABLE 2

Raw joint and marginal frequencies, by combination of fixed effects.

Subclass	Joint frequencies					Marginal frequencies	
	No.	[00]	[01]	[10]	[11]	[0 −]	[− 0]
Season 1 × males	448	.4621	.1652	.2031	.1696	.6272	.6652
Season 1 × females	447	.6040	.2260	.1096	.0604	.8300	.7136
Season 2 × males	1 055	.6445	.0796	.2171	.0588	.7242	.8616
Season 2 × females	1 050	.8257	.0829	.0676	.0238	.9086	.8933
Overall	3 000	.6747	.1153	.1467	.0633	.7900	.8213

(a) [00] : « easy » birth, « good » vigor ; [01] : « easy » birth, « poor » vigor ; [10] : « difficult » birth, « good » vigor ; [11] « difficult » birth, « poor » vigor.

TABLE 3

Estimates of location parameters and of their posterior precision[a].

Parameter [b]	Calving difficulty	Perinatal mortality
$t_1 + s_2$	1.0329 ± 0.0873	0.5687± 0.0585
$t_2 + s_2$	1.3722 ± 0.0823	1.2529 ± 0.0532
$s_1 - s_2$	− 0.7263 ± 0.0568	− 0.1528 ± 0.0554
u_1	− 0.1726 ± 0.1484	− 0.0014 ± 0.1060
u_2	− 0.1298 ± 0.1448	− 0.0844 ± 0.1030
u_3	0.2765 ± 0.1547	0.0402 ± 0.1043
u_4	− 0.2055 ± 0.1418	0.0065 ± 0.1031
u_5	0.2095 ± 0.1530	− 0.0166 ± 0.1022
u_6	0.3267 ± 0.1551	0.1058 ± 0.1045
u_7	− 0.2504 ± 0.1430	0.0813 ± 0.1055
u_8	0.1188 ± 0.1576	0.0760 ± 0.1080
u_9	0.0120 ± 0.1497	0.0500 ± 0.1051
u_{10}	− 0.3545 ± 0.1416	0.1506 ± 0.1076
u_{11}	− 0.2658 ± 0.1440	− 0.1355 ± 0.1029
u_{12}	− 0.1272 ± 0.1586	0.0167 ± 0.1106
u_{13}	− 0.1516 ± 0.1477	0.0749 ± 0.1075
u_{14}	0.3878 ± 0.1716	− 0.0447 ± 0.1046
u_{15}	0.6672 ± 0.1789	0.0182 ± 0.1013
u_{16}	− 0.1071 ± 0.1461	− 0.0838 ± 0.1034
u_{17}	− 0.6465 ± 0.1332	− 0.1958 ± 0.0995
u_{18}	0.0510 ± 0.1484	− 0.1605 ± 0.1013
u_{19}	− 0.0786 ± 0.1568	− 0.1027 ± 0.1070
u_{20}	0.1307 ± 0.1536	0.0413 ± 0.1033
u_{21}	− 0.0592 ± 0.1445	0.0813 ± 0.1040
u_{22}	− 0.2973 ± 0.1411	0.1620 ± 0.1069
u_{23}	− 0.2973 ± 0.1452	− 0.0783 ± 0.1042
u_{24}	0.1014 ± 0.1478	− 0.1177 ± 0.1002
u_{25}	0.2850 ± 0.1629	− 0.0867 ± 0.1035
u_{26}	− 0.7619 ± 0.1369	− 0.0647 ± 0.1048
u_{27}	0.0399 ± 0.1504	0.0101 ± 0.1053
u_{28}	0.3885 ± 0.1731	0.0294 ± 0.1074
u_{29}	0.3147 ± 0.1642	0.2033 ± 0.1096
u_{30}	0.5955 ± 0.1742	0.0251 ± 0.1022

(a) « Standard erros » calcutated from the elements of the inverse of the coefficient matrix in (19) ; (b) The higher the value of the parameters, the higher the probability of response in the category coded as [0].

TABLE 4

Estimates of genetic and residual dispersion parameters.

Stopping rule [a]	Set of starting values	n[c]	ρ	$\sigma^{2(d)}_{u_A}$	$\sigma^{(d)}_{u_{AB}}$	$\sigma^{2(d)}_{u_B}$
10^{-3}	1	25	.2834	12.7874	.9435	1.9065
	2	24	.2835	12.7910	.9670	2.1327
	3	25	.2835	12.7898	.9569	2.1177
10^{-4}	1	35	.2834	12.7904	.9636	2.0005
	2	34	.2834	12.7905	.9635	2.0248
	3	35	.2834	12.7905	.9643	2.0235
10^{-6}	1	55	.2834	12.7905	.9641	2.0125
	2	54	.2834	12.7905	.9641	2.0128
	3	54	.2834	12.7905	.9641	2.0128

(a) As in [47]

(b) 1 : $h^2_A = h^2_B = 0.05$, $r_g = \rho = 0$

2 : $h^2_A = h^2_B = 0.50$, $r_g = \rho = 0$

3 : $h^2_A = h^2_B = 1$, $r_g = -0.9$, $\rho = 0.9$

(c) Number of iterates

(d) Entries for sire variances and covariances should be multiplied by 10^{-2}.

for the components of $\boldsymbol{\theta}$ are shown in table 3 and those corresponding to the components of **g** and **r** are in table 4. The estimates of fixed effects agreed well with the values used in simulating the data (except, of course, for the change in sign). For example β^A_1 was estimated at 1.03 and the « true » value was − 1.05. Likewise, the estimate of β^B_3 was − .15 as opposed to .20. The transmitting abilities were also reasonably well predicted as suggested by the values of the correlations between « true » and predicted values which were .94 and .64 for calving ease and perinatal mortalily, respectively. In a balanced layout with known mean and 100 progeny per sire, the expected values of these correlations under normality would have been .95 and .75, respectively. In view of the lack of balance, the presence of unknown fixed effects in the model, and the intrinsic non-linearity of the problem, the agreement between these two sets of correlations can be considered satisfactory.

As shown in table 4, the iterative process converged almost to the same solution irrespective of the values employed to start iteration ; three markedly different starting sets were used and these are described in a footnote to table 4. The estimates of sire variances and covariances were $\hat{\sigma}^2_{u_A} = 12.79 \times 10^{-2}$, $\hat{\sigma}^2_{u_B} = 2.01 \times 10^{-2}$, and $\hat{\sigma}_{u_{AB}} = .96 \times 10^{-2}$. The estimated genetic correlation was .19 (r_g was 0 in the simulation), and the estimates of heritability in the underlying scale were .45 and .08 for calving difficulty and perinatal mortality, respectively ; the corresponding « true » heritabilities were .35 and .05, respectively. The residual correlation stabilized at .2834 ($\rho = .35$ in the simulation) after 5 iterations. For stopping values ranging between 10^{-3} and 10^{-6} and with the tests applied to the θ-values, between 25 and 55 iterations were required to attain « convergence ». In this example, the number of iterates required did not depend on the starting values used. However, calculations conducted with a smaller exampl

sires and 20 progeny per sire) suggested that the number of iterates can strongly depend, although in a seemingly unpredictable manner, on the values used to begin iteration. In this smaller example and for a stopping value of 10^{-4}, 56, 153 and 105 iterations using sets 1,2 and 3 in table 4, respectively, were needed. The estimated parameters were $\hat{h}_A^2 = .40, \hat{h}_B^2 = .17, \hat{r}_g = -.82$, and $\hat{\rho} = .37$. This indicates that the algorithm can be very slow to converge when progeny group sizes are small. This is not surprising because of the relationship between the expressions employed and the E-M algorithm, as discussed earlier. Research on numerical aspects of the procedure is warranted.

IX. Discussion

This article describes a further contribution to the solution of the problem of genetic evaluation with multiple binary responses along the lines of methods developed by GIANOLA & FOULLEY (1983), FOULLEY *et al.* (1983), FOULLEY & GIANOLA (1984), HARVILLE & MEE (1984) and HÖSCHELE *et al.* (1986). Several points such as the analogy with multivariate generalized linear models, the justification for multiple trait analyses, the calculation of genetic evaluations on the probability scale, and the numerical aspects of solving a large non-linear system on $\boldsymbol{\theta}$ have been already discussed by HÖSCHELE *et al.* (1986), so they will not be dealt with here.

In the context of the present paper, three aspects merit discussion as they may limit the usefulness of the results presented. The first issue relates to the consequence of ignoring the second terms of [36] in the estimation of the residual correlations. While this may be unsatisfactory from a theoretical viewpoint, it can be conjectured that the consequences will be small when the method is applied to the large data sets that frequently arise in animal breeding applications. In fact, when this term is included, the estimator can be interpreted as marginal maximum likelihood ; when it is ignored, the procedure is closely related to maximum likelihood (ML). Because estimates of residual variances and covariances obtained by these two methods using multiple trait mixed models often differ little, it is reasonable to speculate that the same would hold in the non-linear domain.

The second aspect is the approximation of the mean vector and covariance matrix of the distribution $\mathbf{u} \mid \mathbf{Y}, \boldsymbol{\gamma}$ by the $\mathbf{u}$-component of the mode of the density $f(\boldsymbol{\theta} \mid \mathbf{Y}, \boldsymbol{\gamma})$ and by the matrix $\mathbf{C}(\boldsymbol{\gamma})$, which is the inverse of the coefficient matrix in [19]. This approximation, also made by HARVILLE & MEE (1984) and by STIRATELLI *et al.* (1984), could be critical. In the context of sire evaluation, for example, this approximation might be crude if progeny group sizes are small. This can cause bias in the estimates of G. GILMOUR *et al.* (1985) conducted a univariate analysis using the procedure described here and in HARVILLE & MEE (1984), and found that the intra-class correlation was under-estimated when family sizes were less or equal than 8. This potential problem merits further study.

The third point concerns the slow convergence of the algorithm used to estimate $\mathbf{G}$ (see formulae [33] and [46]). These expressions, related to the EM algorithm (DEMPSTER *et al.*, 1977), are very slow to converge, particularly when the eigenvalues of $\mathbf{G}$ are small (THOMPSON, 1979). Techniques used to accelerate convergence in the case of normal variables (THOMPSON & CAMERON, 1986) might be useful here. Another possibility would be to develop algorithms based on second derivatives of $f(\boldsymbol{\gamma} \mid \mathbf{Y})$

with respect to **g**, or to extend the techniques described by SMITH & GRASER (1986) to the discrete domain. It would be useful to develop procedures yielding at least approximations to the posterior dispersion matrix of **g**. For example, LOUIS (1982) has addressed this problem in the context of the EM algorithm.

Because precise estimation of genetic variances and covariances requires an extensive amount of data, in instances in which little data is available it may be useful to incorporate prior information about **G** in the estimation procedure. For example, this prior information could stem from previous data sets pertinent to the problem. The form of [30] suggests using an inverted Wishart distribution as an informative conjugate prior (CHEN, 1979) The density is then

$$f(\mathbf{G} \mid \boldsymbol{\Omega}, \nu) \mid \mathbf{G} \mid^{-\frac{1}{2}(\nu+n+1)} \exp[-\nu \text{ tr } (\boldsymbol{\Omega}\, \mathbf{G}^{-1})] \qquad [49]$$

where :

- $\boldsymbol{\Omega}$ is an $n \times n$ known matrix interpreted as a location parameter of the prior distribution such that $E(\mathbf{G}^{-1} \mid \boldsymbol{\Omega}, \nu) = \boldsymbol{\Omega}^{-1}$, and
- ν is an integer interpreted as degrees of freedom or as a measure of « degree of belief » in $\boldsymbol{\Omega}$.

When $\nu = 0$, [49] becomes $\mid \mathbf{G} \mid^{-\frac{1}{2}(n+1)}$ which is a non-informative prior distribution for **G**.

In general, the new estimator (**G****) obtained using the informative prior [49] can be written as

$$\mathbf{G}^{**} = (q\mathbf{G}^{*} + \nu\boldsymbol{\Omega})/(q + \nu + n + 1) \qquad [50]$$

where **G*** is the marginal maximum likelihood estimator of **G**. Expression [50] can be viewed as a weighted average of **G*** and $\boldsymbol{\Omega}$. This estimator is not invariant under transformations. For example, if one is interested in making inferences about $\boldsymbol{\Lambda} = \mathbf{G}^{-1}$, which is reasonable in view of the form of equations [19], one would obtain $\boldsymbol{\Lambda}^{*} = (\mathbf{G}^{*})^{-1}$ as marginal maximum likelihood estimator of $\boldsymbol{\Lambda}$. However, the estimator based on [49] is

$$\boldsymbol{\Lambda}^{**} = (q\mathbf{G}^{*} + \nu\boldsymbol{\Omega})^{-1}\,(q + \nu - n - 1)$$

which is not the inverse of [50]. Use of reference priors (BERNARDO, 1979) would be worth investigating.

The methodology described in this paper consists of basing inferences on $\boldsymbol{\theta}$ on the conditional distribution $f(\boldsymbol{\theta} \mid \mathbf{Y}, \boldsymbol{\gamma}^{*})$, where $\boldsymbol{\gamma}^{*}$ is the mode of $f(\boldsymbol{\gamma} \mid \mathbf{Y})$. This is along the lines suggested by O'HAGAN (1980) and GIANOLA *et al.* (1986). However, there are alternatives. As pointed out by BROEMELING (1985, p. 144), the mixed model can be viewed as having two levels of parameters. The first or « primary » level includes the location parameters $\boldsymbol{\beta}$ ans **u** and the vector of residual correlations **r**. The « secondary » level comprises the elements of **g**, or **u**-components of variance and covariance ; these are regarded in Bayesian inference as « hyper-parameters » linked to the prior distribution of **u**. If the hyper-parameters are known, the prior distribution of **u** is completely specified, and inferences are based on $f(\boldsymbol{\beta}, \mathbf{u}, \mathbf{r} \mid \mathbf{Y}, \mathbf{g})$. Alternatively, as done in empirical Bayes estimation, one could base inferences on $f(\boldsymbol{\beta}, \mathbf{u}, \mathbf{r} \mid \mathbf{Y}, \mathbf{g} = \tilde{\mathbf{g}})$, where $\tilde{\mathbf{g}}$ is the maximum of $f(\mathbf{g} \mid \mathbf{Y})$, a marginal posterior distribution based on a flat prior for **g**. It is shown in Annex D via the method of « cyclic ascent » (ZANGWILL, 1969 ; OBERHOFER & KMENTA, 1974), that $\tilde{\boldsymbol{\beta}}$ and $\tilde{\mathbf{u}}$, the components of the mode of $f(\boldsymbol{\beta}, \mathbf{u}, \mathbf{r} \mid \mathbf{Y}, \mathbf{g} = \tilde{\mathbf{g}})$ correspond to the mode of $f(\boldsymbol{\beta}, \mathbf{u} \mid \mathbf{Y}, \tilde{\mathbf{g}}, \tilde{\mathbf{r}})$ where $\tilde{\mathbf{r}}$ is

the maximum with respect to $\mathbf{r}$ of the function $f(\mathbf{Y} \mid \tilde{\boldsymbol{\beta}}, \tilde{\mathbf{u}}, \tilde{\mathbf{g}}, \mathbf{r}).f(\mathbf{r})$. With a flat prior for $\mathbf{r}$, the estimates so obtained for $\boldsymbol{\beta}$, $\mathbf{u}$, and $\mathbf{r}$ have the same form of those presented in the article when the residual correlations are estimated by an ML-type procedure (see Section VI). The difference resides in conditioning on $\mathbf{g} = \tilde{\mathbf{g}}$ rather on $\mathbf{g} = \mathbf{g}^*$, where $\mathbf{g}^*$ is the $\mathbf{g}$-component of the mode of $f(\boldsymbol{\gamma} \mid \mathbf{Y})$. This illustrates at least one variation of the theme, and that there may be alternative approximations to $E(\boldsymbol{\theta} \mid \mathbf{Y})$. From a theoretical point of view, it would be desirable to completely marginalize the posterior distribution of $\mathbf{u}$ by integrating out all « nuisance » parameters, i.e., the fixed effects $\boldsymbol{\beta}$ and all the dispersion parameters $\boldsymbol{\gamma}$. This type of inference has been discussed by HARVILLE (1985), and by GIANOLA *et al.* (1986) in animal breeding settings.

Received May 29, 1986.
Accepted November 17, 1986.

Acknowledgements

Part of this research was conducted while J.L. FOULLEY was a George A. MILLER Visiting Scholar at the University of Illinois. He acknowledges the support of the Direction des Productions animales and Direction des Relations internationales, I.N.R.A. D. GIANOLA wishes to acknowledge the support of the Illinois Agriculture Experiment Station, and of Grant U.S.-805-84 from BARD-The United States-Israel Binational Agricultural Research and Development Fund. Thanks are also extended to Dr. C. CHEVALET (I.N.R.A., Toulouse) and to one anonymous referee for very valuable comments which helped to improve the manuscript.

References

ANDERSON T.W., 1984. *An introduction to multivariate statistical analysis.* 675 pp., John Wiley & Sons, New York.

ANDERSON J.A., PEMBERTON J.D., 1985. The grouped continuous model for multivariate ordered categorical variates and covariate adjustment. *Biometrics,* **41**, 875-885.

BERGER J.O., 1985. *Statistical decision theory and Bayesian analysis.* 2nd ed., 617 pp., Springer-Verlag, New York.

BERNARDO J.M., 1979. Reference posterior distribution for Bayesian inference. *J. R. Statist. Soc.* (B), 113-147 (with discussion).

BOX G.E.P., TIAO G.C., 1973. *Bayesian inference in statistical analysis.* 585 pp., Addison-Wesley, Reading.

BROEMELING L.D., 1985. *Bayesian analysis of linear models.* 454 pp., Marcel Dekker, New York.

CASELLA G., 1985. An introduction to empirical Bayes data analysis. *Am. Stat.,* **39**, 83-87.

CHEN C.F., 1979. Bayesian inference for a normal dispersion matrix and its application to stochastic multiple regression analysis. *J. R. Statist. Soc.* (B), **41**, 235-248.

COX D.R., HINKLEY D.V., 1974. *Theoretical Statistics.* 511 pp., Chapman & Hall, London.

CRAMER H., 1974. *Mathematical methods of statistics.* 421 pp., Princeton University Press, Princeton.

CURNOW R.N., SMITH C., 1975. Multifactorial models for familial diseases in man. *J. R. Statist. Soc.* (A), **138**, 131-156.

DAHLQUIST G., BJORCK A., 1974. *Numerical methods.* 574 pp., Prentice Hall, Englewood Cliffs.

DEMPFLE L., 1977. Relation entre BLUP et estimateurs Bayésiens. *Ann. Génét. Sél. Anim.,* **9**, 27-32.

DEMPSTER E.R., LERNER I.M., 1950. Heritability of threshold characters. *Genetics*, **35**, 212-236.

DEMPSTER A.P., LAIRD N.M., RUBIN D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Statist. Soc.* (B), **39**, 1-38.

DUCROCQ V., COLLEAU J.J., 1986. Interest in quantitative genetics of Dutt's and Deak's methods for numerical computation of multivariate normal probability integrals. *Génét. Sél. Evol.*, **18**, 447-474.

EVERITT B.S., 1984. *An introduction to latent variate models.* 107 pp., Chapman & Hall, London.

FALCONER D.S., 1965. The inheritance of liability to certain diseases estimated from the incidence among relatives. *Ann. Hum. Genet.*, **29**, 51-76.

FOULLEY J.L., GIANOLA D., THOMPSON R., 1983. Prediction of genetic merit from data on binary and quantitative variates with an application to calving difficulty, birth weight and pelvic opening. *Génét. Sél. Evol.*, **15**, 401-424.

FOULLEY J.L., GIANOLA D., 1984. Estimation of genetic merit from bivariate « all-or-none » responses. *Génét. Sél. Evol.*, **16**, 285-306.

FOULLEY J.L., OLLIVIER L., 1986. Criteria of coherence for the parameters used to construct a selection index. *Proc. 35th Annual National Breeders' Roundtable, May 1-2, 1986, St-Louis, Missouri.* 17 pp., mimeo.

FOULLEY J.L., GIANOLA D., IM S., 1986. Genetic evaluation of traits distributed as Poisson-binomial with reference to reproductive characters. *Theor. Appl. Genet.* (accepted).

GIANOLA D., FOULLEY J.L., 1982. Non-linear prediction of latent genetic liability with binary expression : an empirical Bayes approach. *2nd World Congress on genetics applied to livestock production,* Madrid, 4-8 October 1982, **7**, 293-303, Editorial Garsi, Madrid.

GIANOLA D., FOULLEY J.L., 1983. Sire evaluation for ordered categorical data with a threshold model. *Génét. Sél. Evol.*, **15**, 201-223.

GIANOLA D., FERNANDO R.L., 1986. Bayesian methods in animal breeding theory. *J. Anim. Sci.*, **63**, 217-244.

GIANOLA D., FOULLEY J.L., FERNANDO R.L., 1986. Prediction of breeding values when variances are not known. *Génét. Sél. Evol.*, **18**, 485-498.

GILMOUR A.R., ANDERSON R.D., RAE A.L., 1985. The analysis of binomial data by a generalized linear mixed model. *Biometrika,* **72**, 593-599.

GOFFINET B., ELSEN J.M., 1984. Critère optimal de sélection : quelques résultats généraux. *Génét. Sél. Evol.*, **16**, 307-318.

HARVILLE D.A., 1974. Bayesian inference for variance components using only error contrats. *Biometrika*, **61**, 383-385.

HARVILLE D.A., 1977. Maximum likelihood approaches to variance component estimation and to related problems. *J. Am. Statist. Assoc.*, **72**, 320-340.

HARVILLE D.A., 1985. Decomposition of prediction error. *J. Am. Statist. Assoc.*, **80**, 132-138.

HARVILLE D.A., MEE R.W., 1984. A mixed model procedure for analyzing ordered categorical data. *Biometrics,* **40**, 393-408.

HENDERSON C.R., 1973. Sire evaluation and genetic trends. *Proceedings of the Animal Breeding and Genetics Symposium in honor of Dr. J.L. Lush.* American Society of Animal Science and American Dairy Science Association, 10-41. Champaign, Illinois.

HENDERSON C.R., 1984. *Application of linear models in animal breeding.* 462 pp., University of Guelph, Guelph.

HENDERSON C.R., QUAAS R.L., 1976. Multiple trait evaluation using relatives' records. *J. Anim. Sci.*, **43**, 1188-1197.

HILL W.G., THOMPSON R., 1978. Probabilities of non-positive definite between-group or genetic covariance matrices. *Biometrics,* **34**, 429-439.

HOGG R.V., CRAIG A.T., 1968. *Introduction to Mathematical Statistics.* 415 pp., MacMillan Publishing Co., New York.

HÖSCHELE Ina, FOULLEY J.L., COLLEAU J.J., GIANOLA D., 1986. Genetic evaluation for multiple binary responses. *Génét. Sél. Evol.*, **18**, 299-320.

KENDALL M.G., STUART A., 1973. *The Advanced Theory of Statistics.* Vol. 2, 676 pp., Charles Griffin & Co, London.

LEFORT G., 1980. Le modèle de base de la sélection, justifications et limites. *In* : LEGAY J.M. *et al.* (éd.), *Biométrie et Génétique*, **4**, 1-14. Société française de biométrie. I.N.R.A., Département de Biométrie.

LINDLEY D.V., SMITH A.F.M., 1972. Bayes estimates for the linear model. *J. R. Statist. Soc.* (B), **34**, 1-41.

LOUIS T.A., 1982. Finding the observed information matrix when using the EM algorithm. *J. R. Statist. Soc.* (B), **44**, 226-233.

MALECOT G., 1947. Les critères statistiques et la subjectivité de la connaissance scientifique. *Ann. Univ. Lyon*, **10**, 43-74.

MCCULLAGH P., NELDER J.A., 1983. *Generalized linear models.* 263 pp., Chapman & Hall, London.

OBERHOFER W., KMENTA J., 1974. A general procedure for obtaining maximum likelihood estimates in generalized regression models. *Econometrica*, **42**, 579-590.

O'HAGAN A., 1980. Likelihood, sufficiency and ancilarity : reply to the discussion. *In :* BERNARDO J.M., DE GROOT M.H., LINDLEY D.V., SMITH A.F.M., (eds.), *Bayesian Statistics. Proceedings of the First International Meeting held in Valencia (Spain), May, 8-June 2, 1975*, 185-203. University Press. Valencia.

PLACKETT R.L., 1954. A reduction formula for normal multivariate integrals. *Biometrika*, **41**, 351-360.

ROBERTSON A., LERNER I.M., 1949. The heritability of all-or-none traits : viability of poultry. *Genetics*, **34**, 395-411.

RONNINGEN K., 1971. Some properties of the selection index derived by « Henderson's Mixed Model Method ». *Z. Tierz. Zuchtbiol.*, **88**, 186-193.

RUTLEDGE J.J., 1977. Repeatability of threshold traits. *Biometrics*, **33**, 395-399.

SMITH S.P., GRASER H.V., 1986. Estimating variance components in a class of mixed models by restricted maximum likelihood. *J. Dairy Sci.*, **69**, 1156-1165.

STIRATELLI R., LAIRD N., WARE J.H., 1984. Random effects models for serial observations with binary response. *Biometrics*, **40**, 961-971.

TALLIS G.M., 1962. The maximum likelihood estimation of correlation from contingency tables. *Biometrics*, **18**, 342-353.

THOMPSON R., 1972. The maximum likelihood approach to the estimate of liability. *Ann. Hum. Genet.*, **36**, 221-231.

THOMPSON R., 1979. Sire evaluation. *Biometrics*, **35**, 339-353.

THOMPSON R., 1980. Maximum likelihood estimation of variance components. *Math. Operationforsch. Statist.*, **11**, 545-561.

THOMPSON R., CAMERON N.D., 1986. Estimation of variances and covariances. *In* : DICKERSON G.E., JOHNSON R.K. (eds.), *3rd World Congress on Genetics Applied to Livestock Production, Lincoln, 17-22 July 1986*, **11**, 371-381, Agricultural Communications, Lincoln.

WRIGHT S., 1934. An analysis of variability in number of digits in an inbred strain of Guinea pigs. *Genetics*, **19**, 506-536.

ZANGWILL W.I., 1969. *Nonlinear programming : a unified approach.* 356 pp., Prentice-Hall, Englewood Cliffs.

Annex A

Positive definiteness of the expected value of the negative of the matrix of second derivatives of the log-posterior density with respect to θ

We consider the positive definiteness of the matrix

$$-\underset{\mathbf{Y}|\boldsymbol{\theta}}{\mathrm{E}}\frac{[\delta^2 \mathrm{L}(\boldsymbol{\theta})]}{\delta\boldsymbol{\theta}\,\delta\boldsymbol{\theta}'}$$

which we refer to as the « information » matrix. Let $\mathbf{r}$ be the vector of residual correlations, and

$$\boldsymbol{\mu}_{ij} = [\mu_{ij}, \mu_{2j}, \ldots \mu_{nj}]'$$

Define the « information » matrix in subpopulation j as

$$\mathbf{J}_j = \mathbf{J}_j(\boldsymbol{\mu}_j, \mathbf{r}) = \begin{bmatrix} J_{j11} & J_{j12} \\ J_{j21} & J_{j22} \end{bmatrix} \qquad j = 1, 2, \ldots, s \qquad [A1]$$

where

$$J_{j11} = n_{j+} \sum_{k=1}^{2^n} \frac{\delta P_{jk}}{\delta \boldsymbol{\mu}_j} \cdot \frac{\delta P_{jk}}{\delta \boldsymbol{\mu}_j'} / P_{jk}$$

$$J_{j12} = n_{j+} \sum_{k} \frac{\delta P_{jk}}{\delta \boldsymbol{\mu}_j} \cdot \frac{\delta P_{jk}}{\delta \mathbf{r}'} / P_{jk}$$

$$J_{j21} = J_{j12}'$$

and

$$J_{j22} = n_{j+} \sum_{k} \frac{\delta P_{jk}}{\delta \mathbf{r}} \cdot \frac{\delta P_{jk}}{\delta \mathbf{r}'} / P_{jk}$$

Thus $\mathbf{J}_j$ is a matrix of order $n + n(n-1)/2 = n(n+1)/2$. Letting

$$\boldsymbol{\psi}_j = [\mathbf{u}_j', \mathbf{r}']'$$

$$\mathbf{p}_j' = \{P_{jk}\}$$

and $\dfrac{\delta \mathbf{p}_j'}{\delta \boldsymbol{\psi}_j}$

be the Jacobian of the transformation $\boldsymbol{\psi}_j \rightarrow \mathbf{p}_j$. Assume this transformation has rank $\frac{1}{2} n(n+1)$. Then

$$\mathbf{J}_j = n_{j+} \cdot \frac{\delta \mathbf{p}_j'}{\delta \boldsymbol{\psi}_j} \cdot \mathbf{D} \cdot \frac{\delta \mathbf{p}_j}{\delta \boldsymbol{\psi}_j}, \qquad [A2]$$

where $\mathbf{D}$ is a $2^n \times 2^n$ diagonal matrix with typical element P_{jk}^{-1}. Using Theorem A.1.1 of ANDERSON (1984, p. 583), it follows automatically that $\mathbf{J}_j$ is positive definite.

The information matrix for all sub-populations can be written as

$$\mathbf{J} = \mathbf{J}(\boldsymbol{\mu}, \mathbf{r}) = \begin{bmatrix} \mathbf{J}_{11} & \mathbf{J}_{12} \\ \mathbf{J}_{21} & \mathbf{J}_{22} \end{bmatrix} \qquad [A3]$$

where

$\mathbf{J}_{11} = \oplus \mathbf{J}_{j11}$, $j = 1, \ldots, s$,

and $\oplus$ is a direct-sum operator,

$\mathbf{J}_{21} = \mathbf{J}'_{12} = [\mathbf{J}'_{112}, \ldots, \mathbf{J}'_{j12}, \ldots, \mathbf{J}'_{s12}]$,

and

$$\mathbf{J}_{22} = \sum_{j=1}^{s} \mathbf{J}_{j22}$$

From ANDERSON (1984, p. 594)

$$|\mathbf{J}| = |\mathbf{J}_{11}| \cdot |\mathbf{J}_{22} - \mathbf{J}_{21}\mathbf{J}_{11}^{-1}\mathbf{J}_{12}| = |\mathbf{J}_{11}| \cdot |\sum_{j} (\mathbf{J}_{j22} - \mathbf{J}_{j21}\mathbf{J}_{j11}^{-1}\mathbf{J}_{j12})| \qquad [A4.1]$$

and

$$|\mathbf{J}_{11}| = \prod_{j} |\mathbf{J}_{j11}| > 0 \qquad [A4.2]$$

Now, the matrices $\mathbf{J}_j$ in [A1] and $\mathbf{J}$ in [A3] can be considered as variance-covariance matrices of corresponding multivariate normal distributions, and the matrix inside the second determinant in [A4.1] would be the variance-covariance matrix of a conditional distribution. This implies that this latter matrix is positive definite. Because from [A4.2] $|\mathbf{J}_{11}|$ is positive, the product of the two determinants in [A4.1] is also positive. It follows that $\mathbf{J}$ is positive definite.

The « information » matrix for $\boldsymbol{\beta}$, $\mathbf{u}$, $\mathbf{r}$ can be written as

$$\mathbf{J}^* = \mathbf{J}(\boldsymbol{\beta}, \mathbf{r}, \mathbf{u}) = \mathbf{K}' \mathbf{J} \mathbf{K} \qquad [A5]$$

where

$$\mathbf{K} = \begin{bmatrix} \mathbf{X} & \mathbf{0} & \mathbf{Z} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \end{bmatrix}$$

Observe that the number of rows in $\mathbf{K}$ is larger than the number of columns. Using again ANDERSON (1984, p. 583), it follows that $\mathbf{J}^*$ is positive-definite when $\mathbf{K}$ has full-column rank ; otherwise, is positive semi-definite.

Define

$L_1 = L(\mathbf{Y} \mid \boldsymbol{\beta}, \mathbf{r}, \mathbf{u})$: logarithm of the likelihood function,

$L_2 = L_2(\mathbf{u} \mid \boldsymbol{\Sigma}_u)$: logarithm of the prior distribution of $\mathbf{u}$,

$L = L_1 + L_2$

$\boldsymbol{\xi} = [\boldsymbol{\beta}, \mathbf{r}, \mathbf{u}]$.

Then

$$\frac{\delta^2 L}{\delta\xi\,\delta\xi'} = \frac{\delta^2 L_1}{\delta\xi\,\delta\xi'} + \frac{\delta^2 L_2}{\delta\xi\,\delta\xi'} \qquad [A6.1]$$

with

$$\frac{\delta^2 L_1}{\delta\xi\,\delta\xi'} \simeq -\,\mathbf{J}^* \qquad [A6.2]$$

and

$$\frac{\delta^2 L_2}{\delta\xi\,\delta\xi'} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\,\boldsymbol{\Sigma}_u^{-1} \end{bmatrix} \qquad [A6.3]$$

For any vector $\mathbf{v}' = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3]$, we have

$$\mathbf{v}' \left\{ \frac{-\,\delta^2 L}{\delta\xi\,\delta\xi'} \right\} \mathbf{v} = \mathbf{v}'\,\mathbf{J}^*\mathbf{v} + \mathbf{v}_3\,\boldsymbol{\Sigma}_u^{-1}\,\mathbf{v}_3 \qquad [A7]$$

Clearly, none of the two terms of [A7] can be negative. Hence [A7] is null only if both terms are null. If $\mathbf{G}$ is positive definite, $\boldsymbol{\Sigma}_u^{-1}$ is also positive definite so for [A7] being null, $\mathbf{v}_3$ must be null. This implies that the first term of [A7] has the form :

$$\mathbf{v}'\,\mathbf{J}^*\,\mathbf{v} = [\mathbf{v}'_1,\ \mathbf{v}'_2] \cdot \begin{bmatrix} \mathbf{X}' & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \cdot \mathbf{J} \cdot \begin{bmatrix} \mathbf{X} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{bmatrix} \qquad [A8]$$

Now, [A8] is null only if $\mathbf{X}\mathbf{v}_1 = \mathbf{0}$ and $\mathbf{v}_2 = \mathbf{0}$ because $\mathbf{J}$ is positive-definite. Further if $\mathbf{X}$ has full-column rank, $\mathbf{X}\mathbf{v}_1$ is null only if $\mathbf{v}_1 = \mathbf{0}$. Thus, [A7] is null only if $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$ are all null, which implies that

$$\left\{ \frac{-\,\delta^2 L}{\delta\xi\,\delta\xi'} \right\}$$

is positive definite. This property also applies to the particular case $\boldsymbol{\xi} = [\boldsymbol{\beta}, \mathbf{u}]$ provided, as before, that $\mathbf{X}$ has full-column rank and $\mathbf{G}$ is positive-definite. Finally, it should be mentioned that $\mathbf{R}$ is implicitly positive-definite because otherwise, the probabilities P_{jk} would be ill-defined.

Annex B

Components of the mode of the joint posterior distribution of the residual and genetic dispersion parameters

Genetic components

Let

$$\mathbf{P}(\mathbf{g}, \mathbf{r}, \mathbf{Y}) = \frac{\delta}{\delta \mathbf{g}} ln \int_{R_u} \int_{R_\beta} f(\mathbf{g}, \mathbf{r}, \mathbf{u}, \boldsymbol{\beta} \mid \mathbf{Y})\, \mathbf{du}\, \mathbf{d\beta} \qquad [B1]$$

$$= [f(\mathbf{g}, \mathbf{r} \mid \mathbf{Y})]^{-1} \frac{\delta}{\delta \mathbf{g}} \int_{R_u} \int_{R_\beta} f(\mathbf{g}, \mathbf{r}, \mathbf{u}, \boldsymbol{\beta} \mid \mathbf{Y})\, \mathbf{du}\, \mathbf{d\beta}$$

$$= [f(\mathbf{g}, \mathbf{r} \mid \mathbf{Y})]^{-1} \int_{R_u} \int_{R_\beta} \left[\frac{\delta}{\delta \mathbf{g}} ln\, f(\mathbf{g}, \mathbf{r}, \mathbf{u}, \boldsymbol{\beta} \mid \mathbf{Y})\right] \cdot f(\mathbf{g}, \mathbf{r}, \mathbf{u}, \boldsymbol{\beta} \mid \mathbf{Y})\, \mathbf{du}\, \mathbf{d\beta}$$

Now [B2]

$$f(\mathbf{g}, \mathbf{r}, \mathbf{u}, \boldsymbol{\beta} \mid \mathbf{Y}) \propto f(\mathbf{Y} \mid \mathbf{g}, \mathbf{r}, \mathbf{u}, \boldsymbol{\beta}) \cdot f(\mathbf{u}, \boldsymbol{\beta} \mid \mathbf{g}, \mathbf{r}) \cdot f(\mathbf{g}, \mathbf{r}),$$

Remembering that : i) the likelihood function does not depend on **g**, ii) **β** is *a priori* independent of the dispersion parameters and of **u**, iii) the prior distribution of **u** depends only on **g**, and iv) taking **g** and **r** independent *a priori*, the preceding expression becomes :

$$f(\mathbf{g}, \mathbf{r}, \mathbf{u}, \boldsymbol{\beta} \mid \mathbf{Y}) \propto f(\mathbf{Y} \mid \mathbf{r}, \mathbf{u}, \boldsymbol{\beta}) \cdot f(\mathbf{u} \mid \mathbf{g}) \cdot f(\boldsymbol{\beta}) . f(\mathbf{g}) \cdot f(\mathbf{r}) \qquad [B3]$$

Thus

$$\frac{\delta}{\delta \mathbf{g}} ln\, f(\mathbf{g}, \mathbf{r}, \mathbf{u}, \boldsymbol{\beta} \mid \mathbf{Y}) = \frac{\delta}{\delta \mathbf{g}} [ln\, f(\mathbf{u} \mid \mathbf{g})] + \frac{\delta}{\delta \mathbf{g}} ln\, f(\mathbf{g}) \qquad [B4]$$

gives the first term under the integral sign in [B2]. The second term can be written as

$$f(\mathbf{g}, \mathbf{r}, \mathbf{u}, \boldsymbol{\beta} \mid \mathbf{Y}) = f(\mathbf{u}, \boldsymbol{\beta} \mid \mathbf{g}, \mathbf{r}, \mathbf{Y}) \cdot f(\mathbf{g}, \mathbf{r} \mid \mathbf{Y}) \qquad [B5]$$

Inserting [B4] and [B5] in [B2]

$$\mathbf{P}(\mathbf{g}, \mathbf{r}, \mathbf{Y}) = \int_{R_u} \int_{R_\beta} \left\{\frac{\delta}{\delta \mathbf{g}} ln\, f(\mathbf{u} \mid \mathbf{g}) + \frac{\delta}{\delta \mathbf{g}} ln\, f(\mathbf{g})\right\} f(\mathbf{u}, \boldsymbol{\beta} \mid \mathbf{g}, \mathbf{r}, \mathbf{Y})\, \mathbf{du}\, \mathbf{d\beta}$$

$$= \int_{R_u} \left[\frac{\delta}{\delta \mathbf{g}} ln\, f(\mathbf{u} \mid \mathbf{g})\right] \cdot f(\mathbf{u} \mid \mathbf{g}, \mathbf{r}, \mathbf{Y})\, \mathbf{du} + \frac{\delta}{\delta \mathbf{g}} ln\, f(\mathbf{g})$$

$$= \underset{c}{E} \left[\frac{\delta}{\delta \mathbf{g}} ln\, f(\mathbf{u} \mid \mathbf{g})\right] + \frac{\delta}{\delta \mathbf{g}} ln\, f(\mathbf{g}) \qquad [B6]$$

where $\underset{c}{E}$ is as defined in the main text.

Residual components

The same reasoning applies when searching for the **r** component of the mode of $f(\mathbf{g}, \mathbf{r} \mid \mathbf{Y})$. Let

$$\mathbf{Q}(\mathbf{g}, \mathbf{r}, \mathbf{Y}) = \frac{\delta}{\delta \mathbf{r}} ln \int_{R_u} \int_{R_\beta} f(\mathbf{g}, \mathbf{r}, \mathbf{u}, \boldsymbol{\beta} \mid \mathbf{Y})\, \mathbf{du}\, \mathbf{d\beta} \qquad [B7]$$

In view of [B3]

$$\frac{\delta}{\delta \mathbf{r}} ln\, f(\mathbf{g},\ \mathbf{r},\ \mathbf{u},\ \boldsymbol{\beta} \mid \mathbf{Y}) = \frac{\delta}{\delta \mathbf{r}} ln\, f(\mathbf{Y} \mid \mathbf{r},\ \mathbf{u},\ \boldsymbol{\beta}) + \frac{\delta}{\delta \mathbf{r}} ln\, f(\mathbf{r}) \qquad [B8]$$

By analogy with [B6]

$$\mathbf{Q}(\mathbf{g},\ \mathbf{r},\ \mathbf{Y}) = \underset{c}{E}\left[\frac{\delta}{\delta \mathbf{r}} ln\, f(\mathbf{Y} \mid \mathbf{r},\ \mathbf{u},\ \boldsymbol{\beta})\right] + \frac{\delta}{\delta \mathbf{r}} ln\, f(\mathbf{r}) \qquad [B9]$$

where, again, the expectation is taken with respect to the distribution $f(\mathbf{u},\ \boldsymbol{\beta} \mid \mathbf{g},\ \mathbf{r},\ \mathbf{Y})$. The first term is the above equation can be expanded by a Taylor series about $\hat{\boldsymbol{\theta}}$, the mean of the distribution $f(\mathbf{u},\ \boldsymbol{\beta} \mid \mathbf{g},\ \mathbf{r},\ \mathbf{Y})$, to obtain as term i of $\mathbf{Q}$:

$$[\mathbf{Q}(\mathbf{g},\ \mathbf{r},\ \mathbf{Y})]_i \simeq \left\{\frac{\delta}{\delta r_i} ln\, f(\mathbf{Y} \mid \mathbf{r},\ \mathbf{u},\ \boldsymbol{\beta})\right\}_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}} + \underset{c}{E}\left\{\frac{\delta^2}{\delta r_i\, \delta \boldsymbol{\theta}'} ln\, f(\mathbf{Y} \mid \mathbf{r},\ \mathbf{u},\ \boldsymbol{\beta})\right\}_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$$

$$+ \frac{1}{2} \underset{c}{E}\left\{(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'\left[\frac{\delta^3}{\delta r_i\, \delta \boldsymbol{\theta}\, \delta \boldsymbol{\theta}'} ln\, f(\mathbf{Y} \mid \mathbf{r},\ \mathbf{u},\ \boldsymbol{\beta})\right]_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\right\} + \frac{\delta}{\delta r_i} f(\mathbf{r}) \qquad [B10]$$

Letting

$$\mathbf{M}(\boldsymbol{\gamma}) = \frac{-\delta^2}{\delta \boldsymbol{\theta}\, \delta \boldsymbol{\theta}'} ln\, f(\mathbf{Y} \mid \boldsymbol{\beta},\ \mathbf{u},\ \mathbf{r})$$

which is the coefficient matrix in [19] without Σ_u^{-1}, [B10] finally becomes

$$[\mathbf{Q}(\mathbf{g},\ \mathbf{r},\ \mathbf{Y})]_i = \left\{\frac{\delta}{\delta r_i} ln\, f(\mathbf{Y} \mid \boldsymbol{\beta},\ \mathbf{u},\ \mathbf{r})\right\}_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}} - \frac{1}{2} \text{tr}\left\{\frac{\delta\, \mathbf{M}(\boldsymbol{\gamma})}{\delta r_i} \cdot \mathbf{C}(\boldsymbol{\gamma})\right\}_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}} + \frac{\delta}{\delta r_i} ln\, f(\mathbf{r}) \qquad [B11]$$

where $\mathbf{C}(\boldsymbol{\gamma})$ is as in [21]. Because calculating $\hat{\boldsymbol{\theta}}$ is impossible in the discrete case considered in this paper, we suggest to approximate the posterior mean by the posterior mode $\boldsymbol{\theta}^*$ and to calculate [B11] accordingly.

Annex C

Positive-definiteness of the matrix $\mathbf{G}^{[t+1]}$

From [32], at iteration t + 1 we have

$$\mathbf{G}^{[t+1]} = E_c^{[t]} (\mathbf{U}' \mathbf{A}^{-1} \mathbf{U})/q \qquad [C1]$$

where

$$\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_i, \ldots, \mathbf{u}_n]$$

is a q × n matrix obtained by rearranging the elements of **u** in [8], with candidates for selection in rows and traits in columns ; $E_c^{[t]}$ indicates expectation with respect to the distribution f (**u** | $\mathbf{G}^{[t]}$, **R**, **Y**).

We first prove that if $\mathbf{A}^{-1}$ is positive-definite (which is true except in very special situations such as when there are identical twins in the data set), then $\mathbf{G}^{[t+1]}$ is at least positive semi-definite. Consider the quadratic forms

$$\mathbf{x}' \mathbf{U}' \mathbf{A}^{-1} \mathbf{U}\mathbf{x} \text{ and } E_c^{[t]} (\mathbf{x}' \mathbf{U}' \mathbf{A}^{-1} \mathbf{U}\mathbf{x}) \qquad [C2]$$

where **x** is a non-null vector. These are both ≥ 0 ; note that the second one is a weighted average of non-negative terms so it cannot be negative. Now, from [C1]

$$E_c^{[t]} (\mathbf{x}' \mathbf{U}' \mathbf{A}^{-1} \mathbf{U}\mathbf{x}) = q \cdot \mathbf{x}' \mathbf{G}^{[t+1]} \mathbf{x}$$

so

$$\mathbf{x}' \mathbf{G}^{[t+1]} \mathbf{x} \geq 0 \qquad [C3]$$

It is shown next that [C3] is positive by reduction to the absurd. Suppose that there is a non-null **x** such that $\mathbf{x}' \mathbf{G}^{[t+1]} \mathbf{x} = 0$. If this is true, both quadratic forms in [C2] are null. Theorem 6 in Hogg & Craig (1968, p. 47) states that for every positive constant *c* (which can be arbitrarily small)

$$\text{Prob} \{\mathbf{x}' \mathbf{U}' \mathbf{A}^{-1} \mathbf{U}\mathbf{x} \geq c\} \leq \frac{E (\mathbf{x}' \mathbf{U}' \mathbf{A}^{-1} \mathbf{U}\mathbf{x})}{c}$$

so $\text{Prob}^{[t]} (\mathbf{x}' \mathbf{U}' \mathbf{A}^{-1} \mathbf{U}\mathbf{x} = 0) = 1$. Because $\mathbf{A}^{-1}$ is positive definite, this implies that

$$\text{Prob} \{\mathbf{U}\mathbf{x} = \mathbf{0}\} = 1 \qquad [C4]$$

Because **x** is non-null, [C4] would imply that the distribution with density f (**u** | $\mathbf{G}^{[t]}$, **R**, **Y**) is not « regular » (Cramer, 1974, p. 298), or equivalently, that the posterior variance-covariance matrix $\mathbf{C}_{uu}$, which is a partition of

$$\mathbf{C} = \left\{ - \frac{\delta^2}{\delta\boldsymbol{\theta}\, \delta\boldsymbol{\theta}'} ln\, f (\boldsymbol{\theta} \mid \mathbf{G}^{[t]}, \mathbf{R}, \mathbf{Y}) \right\}^{-1},$$

is singular. However, if **X** has full-column rank and $\mathbf{G}^{[t]}$ is positive-definite, $\mathbf{C}^{-1}$ is positive-definite (see Annex A), as well as **C** and $\mathbf{C}_{uu}$. This contradicts the preceding conclusion so the case that a non-null **x** exists such that $\mathbf{x}' \mathbf{G}^{[t+1]} \mathbf{x} = 0$ is excluded. Therefore, $\mathbf{G}^{[t+1]}$ is positive definite.

Annex D

Components of the mode of the density $f(\boldsymbol{\theta}, \mathbf{r} \mid \tilde{\mathbf{g}}, \mathbf{Y})$

From standard distribution theory

$$f(\boldsymbol{\theta}, \mathbf{r} \mid \mathbf{g}, \mathbf{Y}) = f(\boldsymbol{\theta}, \mathbf{g}, \mathbf{r}, \mathbf{Y})/f(\mathbf{g}, \mathbf{Y})$$
$$= f(\mathbf{Y} \mid \boldsymbol{\theta}, \mathbf{g}, \mathbf{r}) \cdot f(\boldsymbol{\theta} \mid \mathbf{g}, \mathbf{r}) \cdot f(\mathbf{g} \mid \mathbf{r}) \cdot f(\mathbf{r})/f(\mathbf{g}, \mathbf{Y}) \quad [D1]$$

From [12], $f(\boldsymbol{\theta} \mid \mathbf{g}, \mathbf{r}) \simeq f(\mathbf{u} \mid \mathbf{g})$. Reasoning conditionally to a value $\tilde{\mathbf{g}}$ of $\mathbf{g}$, one can write [D1] as

$$f(\boldsymbol{\theta}, \mathbf{r} \mid \tilde{\mathbf{g}}, \mathbf{Y}) \propto f(\mathbf{Y} \mid \boldsymbol{\theta}, \tilde{\mathbf{g}}, \mathbf{r}) \cdot f(\mathbf{u} \mid \tilde{\mathbf{g}}) \cdot f(\mathbf{r}) \quad [D2]$$

The $\boldsymbol{\theta}$ and $\mathbf{r}$ components of the mode of [D2] can be obtained by maximizing this density via the method of « cyclic ascent » (ZANGWILL, 1969 ; OBERHOFER & KMENTA, 1974), as described below.

Let $\boldsymbol{\theta}^{[t-1]}$ and $\mathbf{r}^{[t-1]}$ the values of these parameters at iterate $t-1$ of this method. Values at the following iteration, can be obtained in two steps :

i) find

$$\boldsymbol{\theta}^{[t]} : \underset{\boldsymbol{\theta}}{\text{Max}}\, f(\boldsymbol{\theta}, \mathbf{r}^{[t-1]} \mid \tilde{\mathbf{g}}, \mathbf{Y})$$
$$= \underset{\boldsymbol{\theta}}{\text{Max}} \{f(\mathbf{Y} \mid \boldsymbol{\theta}, \tilde{\mathbf{g}}, \mathbf{r}^{[t-1]}\} \cdot f(\mathbf{u} \mid \tilde{\mathbf{g}})\}$$

which follows from [D2]. Observe that the product of the two densities in the above expression is the posterior distribution of $\boldsymbol{\theta}$ when the dispersion parameters are $\tilde{\mathbf{g}}$ and $\mathbf{r}^{[t-1]}$. Thus

$$\boldsymbol{\theta}^{[t]} : \underset{\boldsymbol{\theta}}{\text{Max}}\, f(\boldsymbol{\theta} \mid \tilde{\mathbf{g}}, \mathbf{r}^{[t-1]}, \mathbf{Y}) \quad [D3]$$

ii) Setting $\boldsymbol{\theta}$ to the value calculated in [D3], calculate $\mathbf{r}^{[t]}$ as

$$\mathbf{r}^{[t]} : \underset{\mathbf{r}}{\text{Max}}\, f(\boldsymbol{\theta}^{[t]}, \mathbf{r} \mid \tilde{\mathbf{g}}, \mathbf{Y})$$
$$= \underset{\mathbf{r}}{\text{Max}} \{f(\mathbf{Y} \mid \boldsymbol{\theta}^{[t]}, \tilde{\mathbf{g}}, \mathbf{r}) \cdot f(\mathbf{r})\} \quad [D4]$$

from [D2].