



**HAL**  
open science

## Précision et information dans les modèles linéaires d'évaluation génétique.

Denis Laloë

► **To cite this version:**

Denis Laloë. Précision et information dans les modèles linéaires d'évaluation génétique.. *Genetics Selection Evolution*, 1993, 25 (6), pp.557-576. hal-02699431

**HAL Id: hal-02699431**

**<https://hal.inrae.fr/hal-02699431>**

Submitted on 1 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Precision and information in linear models of genetic evaluation

D Laloë

*Institut National de la Recherche Agronomique,  
Station de Génétique Quantitative et Appliquée,  
Centre de Recherches de Jouy-en-Josas, 78352 Jouy-en-Josas Cedex, France*

(Received 14 September 1992; accepted 5 August 1993)

**Summary** – Some criteria for measuring the overall precision of a genetic evaluation using linear mixed-model methodology are presented. They are derived *via* an extension of the coefficient of determination to linear combinations of estimates and *via* the use of the Kullback information. A parallel is drawn between inestimability of fixed-effects contrasts and the zero coefficient of determination for contrasts of random effects. The procedure is illustrated with 2 minor hypothetical examples of genetic evaluation based on an animal model and on a sire model.

**genetic evaluation / Kullback information / precision / mixed linear model / disconnectedness**

**Résumé** – Précision et information dans les modèles linéaires d'évaluation génétique. Des critères de précision globale d'une évaluation génétique utilisant la méthodologie du modèle linéaire mixte sont présentés. Leur dérivation utilise une extension du coefficient de détermination à des combinaisons linéaires d'estimées, ainsi que l'information de Kullback. Un parallèle entre inestimabilité de contrastes pour les effets fixés et existence de contrastes à coefficient de détermination nul pour les effets aléatoires est établi. La procédure est illustrée par 2 petits exemples fictifs, un modèle animal et un modèle père.

**évaluation génétique / précision / information de Kullback / modèle linéaire mixte / disconnexion**

## INTRODUCTION

The accuracy of predicted breeding values is commonly assessed by the so-called coefficient of determination (CD), *ie* the squared correlation between the true and estimated genetic values. This measures the amount of information that contributes to the prediction of breeding values, and was first used in the context of selection

indices, where it was easily computed because the environmental effects were supposed to be known exactly, and information was of the same type for every evaluated animal. This theory was based upon a strong assumption: the genetic levels among environmental factor levels were identical. Should this assumption not hold, the comparisons between animals would be valid only for animals raised in the same environment. The evaluation was then usually restricted to, for instance, intra-herd selection. Consequently, the breeder's interest was mainly concentrated on individual CDs.

BLUP (Best linear unbiased predictor), which uses a simultaneous estimation of the environmental and genetic effects and the whole pedigree information of the analysed animals, does not require this assumption and allows genetic evaluations at a population level. The comparisons between animals become meaningful whatever their environments. Since the aim of the breeder is to compare animals in order to select the best, these comparisons are even more important than the individual values. On the other hand, the predicted values supplied by BLUP are not independent and individual CDs are no longer sufficient to look at the precision of comparisons.

Precision depends mainly on: i) the amount of information, *ie* the number of observations that can be related to an animal; and ii) the structure of the design: an unbalanced design leads to less precise predictors than a balanced one.

The same goes for precision investigation, which can be done in 2 different ways:

- studying the structure of the design, and especially the genetic ties between environmental factor levels and the problem of disconnectedness in genetic effects. However, as explained in detail by Foulley *et al* (1990, 1992), complete disconnectedness can never occur in random effects. Foulley *et al* suggest some methods to quantify the non-orthogonality of the design, called the degree of disconnectedness.
- studying some criteria of precision, applicable to any comparison of animals, as well to an entire design.

The aim of this paper is to follow the second approach by extending the concept of the individual CD. This extended CD is shown to be close to a specific measure of information, the Kullback information, and is used to study a disconnectedness-like concept, which could be applicable to random effects. The procedure is illustrated with 2 minor hypothetical examples, an animal model and a sire model.

## BLUP AND CDs: AN OVERVIEW

Let us consider a mixed model with a single random factor (and the residual effect):

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e} \quad [1]$$

where  $\mathbf{b}$  is the fixed effect vector,  $\mathbf{X}$  the pertaining incidence matrix,  $\mathbf{u}$  the random effect vector,  $\mathbf{Z}$  the pertaining incidence matrix, and  $\mathbf{e}$  the residual vector.

The random factors are normally distributed with the following first and second moments:

$$E(\mathbf{u}', \mathbf{e}') = (\mathbf{0}', \mathbf{0}')$$

$$\text{Vr} \begin{pmatrix} \mathbf{u} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} \mathbf{A}\sigma_a^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}\sigma_e^2 \end{pmatrix}$$

The ratio  $\lambda = \sigma_e^2/\sigma_a^2$  is assumed to be exactly known and  $\mathbf{A}$  is assumed to be non-singular, *ie* in the particular case of genetic evaluations, there are no monozygotic twins in the population.

### Mixed model equations

BLUE (Best linear unbiased estimator) of  $\mathbf{b}$  and BLUP of  $\mathbf{u}$  are solutions of the following equation system (Henderson, 1984):

$$\begin{pmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \lambda\mathbf{A}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{b}^\circ \\ \hat{\mathbf{u}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{pmatrix}$$

or, with  $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ ,

$$(\mathbf{Z}'\mathbf{M}\mathbf{Z} + \lambda\mathbf{A}^{-1})\hat{\mathbf{u}} = \mathbf{Z}'\mathbf{M}\mathbf{y}$$

$\mathbf{M}$  is a projector, orthogonal to the vector subspace spanned by  $\mathbf{X}$  columns:

$$\mathbf{M}\mathbf{X} = \mathbf{0} \quad [2]$$

or, if  $\mathbf{x}$  is a linear combination of  $\mathbf{X}$  columns,

$$\mathbf{M}\mathbf{x} = \mathbf{0} \quad [3]$$

### Precision of the estimates, CD

The prediction error variance matrix of  $\mathbf{u}$  is (Henderson, 1984):

$$\text{var}(\mathbf{u}|\hat{\mathbf{u}}) = \text{var}(\hat{\mathbf{u}} - \mathbf{u}) = (\mathbf{Z}'\mathbf{M}\mathbf{Z} + \lambda\mathbf{A}^{-1})^{-1}\sigma_e^2 = \mathbf{\Omega}_1\sigma_e^2$$

The CD of an animal  $i$  is a function of the ratio of the variance of  $\mathbf{u}_i$  knowing the results of the experiment ( $\text{var}(\mathbf{u}_i|\hat{\mathbf{u}}_i)$ ) to the variance of  $\mathbf{u}_i$  before the experiment ( $\text{var}(\mathbf{u}_i)$ ):

$$\begin{aligned} \text{CD}(i) &= 1 - \{\text{var}(\mathbf{u}_i|\hat{\mathbf{u}}_i)/\text{var}(\mathbf{u}_i)\} = 1 - \lambda\mathbf{\Omega}_1(i,i) \\ \text{or} \quad \text{CD}(i) &= 1 - \mathbf{\Omega}(i,i) \end{aligned} \quad [4]$$

where  $\mathbf{\Omega} = \lambda\mathbf{\Omega}_1$ .

This CD equals the squared correlation coefficient between  $\hat{\mathbf{u}}_i$  and  $\mathbf{u}_i$ , and measures the amount of information supplied by the data that has contributed to the prediction of  $\mathbf{u}_i$ .

### Generalization of the CD

An obvious way of examining the precision of comparisons between individuals is to study the corresponding contrasts: the comparison between 2 individuals  $i$  and  $j$

will be related to the contrast  $u_i - u_j$ ; the comparison between 2 sets of individuals will be related to the contrast between both sets, *ie* the average difference of both sets of estimates. Contrasts are particular linear combinations  $\mathbf{x}'\mathbf{u}$ , where  $\mathbf{x}$  is a vector whose elements sum to 0. The precision of any comparison will be evaluated by a precision criterion concerning a linear combination of estimates.

The CD of a linear combination  $\mathbf{u}'\mathbf{u}$  will be a function of the ratio of the variance of  $\mathbf{x}'\mathbf{u}$  after the experiment to the variance of  $\mathbf{x}'\mathbf{u}$  before the experiment, *ie*:

$$CD(\mathbf{x}) = 1 - \{\text{var}(\mathbf{x}'\mathbf{u}|\hat{\mathbf{u}})/\text{var}(\mathbf{x}'\mathbf{u})\}$$

$$CD(\mathbf{x}) = 1 - \mathbf{x}'\boldsymbol{\Omega}_1\mathbf{x}\sigma_e^2/\mathbf{x}'\mathbf{A}\mathbf{x}\sigma_a^2$$

$$CD(\mathbf{x}) = 1 - \mathbf{x}'\boldsymbol{\Omega}\mathbf{x}/\mathbf{x}'\mathbf{A}\mathbf{x}$$

or

$$CD(\mathbf{x}) = \mathbf{x}'(\mathbf{A} - \boldsymbol{\Omega})\mathbf{x}/\mathbf{x}'\mathbf{A}\mathbf{x} \tag{5}$$

The CD of an individual is a particular form of this formula. In an individual CD,  $CD(\mathbf{x}) = 0$  implies that  $\mathbf{x}'\hat{\mathbf{u}} = 0$ .

All the CDs, of both individuals and linear combinations, are then ratios of quadratic forms  $\mathbf{x}'(\mathbf{A} - \boldsymbol{\Omega})\mathbf{x}/\mathbf{x}'\mathbf{A}\mathbf{x}$ . Because quadratic forms associated with a matrix are related to the eigenvalues of the matrix the above ratios of quadratic forms can be related to the generalized eigenvalue problem (Golub and Van Loan, 1983):

$$[(\mathbf{A} - \boldsymbol{\Omega}) - \mu\mathbf{A}]\boldsymbol{\beta} = 0 \tag{6}$$

$$\Leftrightarrow [\boldsymbol{\Omega} - (1 - \mu)\mathbf{A}]\boldsymbol{\beta} = 0 \tag{7}$$

As in the standard eigenvalue problem, the vectors  $\boldsymbol{\beta}$  and the scalars  $\mu$ , the solutions of [6], are called eigenvectors and eigenvalues, respectively.

The solutions  $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_n)$  and  $(\mu_1, \mu_2, \dots, \mu_n)$  of [6], sorted in ascending order, are such that, for  $i$  different from  $j$ :

$$\boldsymbol{\beta}'_i\mathbf{A}\boldsymbol{\beta}_j = 0 \text{ (A - orthogonality)}, \tag{8}$$

$$\boldsymbol{\beta}'_i\boldsymbol{\Omega}\boldsymbol{\beta}_j = 0 \text{ (\boldsymbol{\Omega} - orthogonality)}, \tag{9}$$

$$\boldsymbol{\beta}'_i\mathbf{A}\boldsymbol{\beta}_i = 1, \tag{10}$$

$$\mu_i = \boldsymbol{\beta}'_i(\mathbf{A} - \boldsymbol{\Omega})\boldsymbol{\beta}_i/\boldsymbol{\beta}'_i\mathbf{A}\boldsymbol{\beta}_i = CD(\boldsymbol{\beta}_i)$$

For any non-null vector  $\mathbf{x}$ ,  $\mu_1 \leq CD(\mathbf{x}) \leq \mu_n$  [11]

Studying the magnitude of the ratios of quadratic forms then amounts to the study of the magnitude of these eigenvalues. The occurrence of the null eigenvalue will be particularly interesting to study, because the CDs of the corresponding eigenvectors are null.

Since  $\mathbf{A}$  is positive definite, a lower triangular and non-singular matrix  $\mathbf{L}$  exists such that  $\mathbf{A} = \mathbf{L}\mathbf{L}'$ . Hence:

$$[7] \quad \Leftrightarrow [\mathbf{L}^{-1}\boldsymbol{\Omega} - (1 - \mu)\mathbf{L}^{-1}\mathbf{A}]\boldsymbol{\beta} = 0$$

$$\Leftrightarrow \mathbf{L}^{-1}\boldsymbol{\Omega}\boldsymbol{\beta} = (1 - \mu)\mathbf{L}^{-1}\mathbf{A}\boldsymbol{\beta} = (1 - \mu)\mathbf{L}'\boldsymbol{\beta}$$

$$\begin{aligned}
&\Leftrightarrow \mathbf{L}^{-1}\boldsymbol{\Omega}\mathbf{L}'^{-1}\boldsymbol{\alpha} = (1 - \mu)\boldsymbol{\alpha} \\
&\Leftrightarrow [\mathbf{L}^{-1}\boldsymbol{\Omega}\mathbf{L}'^{-1} - (1 - \mu)\mathbf{I}]\boldsymbol{\alpha} = 0 \\
&\Leftrightarrow [\boldsymbol{\Theta} - \mu\mathbf{I}]\boldsymbol{\alpha} = 0
\end{aligned} \tag{12}$$

where  $\boldsymbol{\alpha} = \mathbf{L}'\boldsymbol{\beta}$  and  $\boldsymbol{\Theta} = \mathbf{I} - \mathbf{L}^{-1}\boldsymbol{\Omega}\mathbf{L}'^{-1}$ .

Equations [6] and [12] have the same eigenvalues. For convenience, we will use [6] when studying the eigenvectors, and [12] when studying the eigenvalues.

### *Dispersion of the CDs of linear combinations*

Since

$$\mathbf{I} - \boldsymbol{\Theta} = \lambda\mathbf{L}^{-1}(\mathbf{Z}'\mathbf{M}\mathbf{Z} + \lambda\mathbf{A}^{-1})^{-1}(\mathbf{L}')^{-1}$$

and:  $(\mathbf{I} - \boldsymbol{\Theta})^{-1} = \lambda^{-1}(\mathbf{L}'(\mathbf{Z}'\mathbf{M}\mathbf{Z}) + \lambda\mathbf{L}'\mathbf{A}^{-1}\mathbf{L}) = \lambda^{-1}\mathbf{L}'(\mathbf{Z}'\mathbf{M}\mathbf{Z})\mathbf{L} + \mathbf{I}$

or:  $\mathbf{I} - \boldsymbol{\Theta} = (\lambda^{-1}\mathbf{L}'(\mathbf{Z}'\mathbf{M}\mathbf{Z})\mathbf{L} + \mathbf{I})^{-1}$

$\boldsymbol{\Theta}$  can be written as:

$$\boldsymbol{\Theta} = \mathbf{I} - (\lambda^{-1}\mathbf{L}'(\mathbf{Z}'\mathbf{M}\mathbf{Z})\mathbf{L} + \mathbf{I})^{-1} \tag{13}$$

Some remarks are worth mentioning at this stage:

- $\boldsymbol{\Theta}$  and  $\mathbf{L}'(\mathbf{Z}'\mathbf{M}\mathbf{Z})\mathbf{L}$  have the same set of eigenvectors, since  $\boldsymbol{\Theta}$  is a linear function of  $\mathbf{I}$  and the inverse of a linear function of  $\mathbf{I}$  and  $\mathbf{L}'(\mathbf{Z}'\mathbf{M}\mathbf{Z})\mathbf{L}$ .
- The CDs can be verified to be between 0 and 1: if, for a given eigenvector, the eigenvalue of  $\mathbf{L}'\mathbf{Z}'\mathbf{M}\mathbf{Z}\mathbf{L}$  is  $\eta$ , then the respective eigenvalue of  $\boldsymbol{\Theta}$  is  $\mu$ , such that:

$$\mu = 1 - [1/(1 + \lambda^{-1}\eta)] \tag{14}$$

Since  $\eta \geq 0$ , we have:  $0 \leq \mu < 1$ .

- $\boldsymbol{\Theta}$  and  $\mathbf{Z}'\mathbf{M}\mathbf{Z}$  have the same rank.  $\boldsymbol{\Theta}$  and  $\mathbf{L}'\mathbf{Z}'\mathbf{M}\mathbf{Z}\mathbf{L}$  have the same eigenvectors, and, from [14], a null eigenvalue of  $\boldsymbol{\Theta}$  corresponds to a null eigenvalue of  $\mathbf{L}'\mathbf{Z}'\mathbf{M}\mathbf{Z}\mathbf{L}$ . Both matrices then have the same rank, and, since  $\mathbf{L}$  and  $\mathbf{L}'$  are non-singular,  $\boldsymbol{\Theta}$ , and  $\mathbf{Z}'\mathbf{M}\mathbf{Z}$  have the same rank.

### *Overall precision criteria*

The location interval [11] of the CDs can lead to some average criteria, like the arithmetic ( $\rho_1$ ) and the geometric ( $\rho_2$ ) means of the eigenvalues. Since the rank of  $\boldsymbol{\Theta}$  is equal to the rank of  $\mathbf{Z}'\mathbf{M}\mathbf{Z}$ , which is less than  $n$ , there is always a null eigenvalue. Thus, the geometric mean of the eigenvalues is null and meaningless. We will then restrict our interest to the  $(n - 1)$  greatest eigenvalues of  $\boldsymbol{\Theta}$ . If the  $\mu_i$ , eigenvalues of  $\boldsymbol{\Theta}$ , are sorted in ascending order, we have:

$$\rho_1 = \text{tr}(\boldsymbol{\Theta})/(n - 1) \tag{15}$$

$$\rho_2 = \left[ \prod_{i=2}^n \mu_i \right]^{\frac{1}{n-1}} \tag{16}$$

**Relationship with selection index theory**

These eigenvalues and associated criteria can be related to selection index theory. Consider a simple balanced sire model, including a single fixed effect (the mean) and a sire effect ( $n$  sires and  $t$  progeny per sire). It can be shown (see *Appendix I*) that the eigenvalues of [6] are:

- 0 with multiplicity 1. The corresponding eigenvector is proportional to  $\mathbf{1}$ ;
- $t/(t + \lambda)$  with multiplicity  $(n - 1)$ . The corresponding eigenvectors  $\mathbf{b}$  are contrasts between sires.

The CD of any between-sires comparison (for instance, the CD of a comparison between a particular sire and the others) is equal to the CD of a sire that would be obtained in the context of the selection index theory. This could have been expected, since considering such comparisons relaxes the uncertainty about the mean. The  $(n - 1)$  greatest eigenvalues of [6] are the same, and we get:  $\rho_1 = \rho_2 = t/(t + \lambda)$ .

**Information supplied by experiment**

Another way to look at the overall precision is to evaluate the amount of precision supplied by the experiment, by calculating the mean of a specific measure of information, the Kullback information (Kullback, 1968; 1983). This measure was introduced in animal breeding theory by Foulley *et al* (1990, 1992), in order to derive the so-called degree of disconnectedness.

**Kullback information**

The Kullback information (Kullback, 1968; 1983) can be used to measure the discrepancy between 2 continuous probability distributions  $p$  and  $q$ , noted  $\mathbf{I}(p:q)$ . This varies from 0 to infinity, and equals:

$$\mathbf{I}(p:q) = \int \log \frac{p(t)}{q(t)} p(t) dt \tag{17}$$

A value of 0 exhibits a total identity between both distributions.

If  $p$  and  $q$  are  $N_n(\xi_1, \Sigma_1)$  and  $N_n(\xi_2, \Sigma_2)$ , respectively, then:

$$\mathbf{I}(p:q) = 1/2 \log \left[ \det(\Sigma_1^{-1} \Sigma_2) + \text{tr} \left( \Sigma_1 \left( \Sigma_2^{-1} - \Sigma_1^{-1} \right) \right) + \text{tr} \left( \Sigma_2^{-1} (\xi_1 - \xi_2)(\xi_1 - \xi_2)' \right) \right] \tag{18}$$

This measure can be used to calculate the information supplied by an experiment, by comparing the probability distribution conditional on the results of this experiment with the initial probability distribution (Kullback, 1968). In our context, the initial probability distribution is the distribution  $f(\mathbf{u})$  of  $\mathbf{u}$ , and the conditional distribution is the distribution  $g(\mathbf{u}|\hat{\mathbf{u}})$  of  $\mathbf{u}$  conditional on  $\mathbf{X}, \mathbf{Z}, \mathbf{A}$  and  $\mathbf{y}$ , *ie* knowing  $\hat{\mathbf{u}}$ . The information depends on a particular  $\mathbf{y}$ , and then on a particular  $\hat{\mathbf{u}}$ . We will restrict our interest to the mean information, given  $\mathbf{X}, \mathbf{Z}$ , and  $\mathbf{A}$ , *ie* the information given the data design:

$$\mathbf{I} = \mathbf{E} \hat{\mathbf{u}} (\mathbf{I} | g(\mathbf{u}|\hat{\mathbf{u}}) : f(\mathbf{u})) \tag{19}$$

$\mathbf{I}$  is equal to the Kullback information between the joint distribution of  $\mathbf{u}$  and  $\hat{\mathbf{u}}$  and the product of the marginal distributions of  $\mathbf{u}$  and  $\hat{\mathbf{u}}$  (cf, *Appendix II*). After some algebra (cf, *Appendix II*):

$$\mathbf{I} = -1/2[\log(\det(\mathbf{I} - \Theta))] \quad [20]$$

or:

$$\mathbf{I} = -1/2 \sum_{i=1}^n \log(1 - \mu_i) \quad [21]$$

where the  $\mu_i$ 's are the eigenvalues of  $\Theta$ . Since the smallest eigenvalue  $\mu_1$  is null, we have:

$$\mathbf{I} = -1/2 \sum_{i=2}^n \log(1 - \mu_i) \quad [22]$$

### *Information for a linear combination*

The distributions of linear combinations  $\mathbf{x}'\mathbf{u}$  and  $\mathbf{x}'\mathbf{u}|\hat{\mathbf{u}}$  are:

$$\begin{aligned} \mathbf{x}'\mathbf{u} &\rightarrow N(0, \mathbf{x}'\mathbf{A}\mathbf{x}\sigma_a^2) \\ \mathbf{x}'\mathbf{u}|\hat{\mathbf{u}} &\rightarrow N(\mathbf{x}'\hat{\mathbf{u}}, \mathbf{x}'\mathbf{\Omega}\mathbf{x}\sigma_e^2) \end{aligned}$$

By the algebra in *Appendix II*, we then get the Kullback information between these 2 distributions, denoted  $\mathbf{I}_x$ :

$$\mathbf{I}_x = -1/2 \log \det(\mathbf{x}'\mathbf{\Omega}\mathbf{x}\sigma_e^2 / \mathbf{x}'\mathbf{A}\mathbf{x}\sigma_a^2)$$

Then we get:

$$\mathbf{I}_x = -1/2 \log(1 - \text{CD}(\mathbf{x})) \quad [23]$$

and:

$$\text{CD}(\mathbf{x}) = 1 - \exp(-2\mathbf{I}_x) \quad [24]$$

The CD is then a simple function of the information. The information for a linear combination of  $\mathbf{u}$  increases with  $\text{CD}(\mathbf{x})$ ; it is null when  $\text{CD}(\mathbf{u})$  is null, and tends to infinity when  $\text{CD}(\mathbf{u})$  tends to 1.

### *Mean CD corresponding to the mean information*

We can derive another overall criterion by writing [22] as:

$$\mathbf{I} = \sum_{i=2}^n \mathbf{I}_{\beta_i} \quad [25]$$

where the  $\beta_i$ 's are the eigenvectors corresponding to the positive eigenvalues of [6]. The total information is the sum of the information for the  $\beta_i$ 's. These vectors are independent under both distributions of  $\mathbf{u}$  and  $\hat{\mathbf{u}}|\mathbf{u}$ ; this result could have been



expected since Kullback information is additive for independent events. We can define  $\iota$ , equal to  $\mathbf{I}/(n - 1)$ , as the average information for a contrast. The mean CD we can deduce from this is:

$$\rho_3 = 1 - \exp(-2\iota) \tag{26}$$

or: 
$$\rho_3 = 1 - \{\det(\mathbf{I} - \Theta)\}^{1/n-1} \tag{27}$$

Let us note that, in the example studied above (*Relationship with selection index theory*),  $\rho_3 = t/(t + \lambda)$ .

**DISCONNECTED DATA**

In the extreme case, unbalanced data for a fixed-effect model, results in disconnectedness. Disconnectedness decreases the rank of the coefficient matrix and, since this rank is the number of independent estimable contrasts, leads to the inestimability of some independent contrasts (Chakrabarti, 1963; Foulley *et al*, 1990). Disconnectedness is often defined by these consequences. Such a definition implies that disconnectedness never occurs for random effects, since their contrasts are always estimable. However, the data design is the same whether the effect is fixed or random (we will refer to this kind of design as a disconnected design). Even for a random effect, a disconnected design can have important consequences on the CDs of contrasts and matrix ranks.

Linear estimable functions in a fixed model can be characterized in terms of eigenvectors (see Graybill 1961, p 237 , Theorem 11.9). Considering model (I) and treating  $\mathbf{u}$  as fixed, the linear estimable functions are linear combinations of the non-null eigenvectors of  $\mathbf{Z}'\mathbf{M}\mathbf{Z}$ . In the following, we will derive a similar characterization for random effects by examining the incidence of the design on the eigenvalues and the eigenvectors of the generalized eigenvalue problem [6]. Since we will consider  $\mathbf{u}$  as either a fixed or random effect, we will denote  $\hat{\mathbf{u}}$  the predictor of  $\mathbf{u}$  when it is treated as random, and  $\hat{\mathbf{u}}$  the estimator of  $\mathbf{u}$  when it is treated as fixed.

***Relationship between  $\mathbf{Z}'\mathbf{M}\mathbf{Z}$  and [6]***

A relationship can be found between eigenvectors of  $\mathbf{Z}'\mathbf{M}\mathbf{Z}$ , which are related to the null eigenvalues, and eigenvectors of [6] which also correspond to the null eigenvalues (Foulley *et al*, 1990):

$$\text{CD}(\mathbf{x}) = 0 \Leftrightarrow (\mathbf{\Omega} - \mathbf{A})(\mathbf{x}) = 0 \Leftrightarrow (\mathbf{Z}'\mathbf{M}\mathbf{Z})(\mathbf{A}\mathbf{x}) = 0 \tag{28}$$

or, symmetrically,

$$\mathbf{Z}'\mathbf{M}\mathbf{Z}\mathbf{v} = 0 \Leftrightarrow (\mathbf{\Omega} - \mathbf{A})(\mathbf{A}^{-1}\mathbf{v}) = 0 \Leftrightarrow \text{CD}(\mathbf{A}^{-1}\mathbf{v}) = 0 \tag{29}$$

or: 
$$\mathbf{Z}'\mathbf{M}\mathbf{Z}\mathbf{v} = 0 \Leftrightarrow \mathbf{v}'\mathbf{A}^{-1}\hat{\mathbf{u}} = 0 \tag{30}$$

These equations lead to a system of built-in constraints similar to the system of constraints that have to be set in order to let a fixed-effects model be of full rank.

If  $\mathbf{Z}'\mathbf{M}\mathbf{Z}\mathbf{v} = 0$ , the corresponding constraint for  $\mathbf{u}$  treated as fixed will be  $\mathbf{v}'\hat{\mathbf{u}} = 0$ . For  $\mathbf{u}$  treated as random, we will have  $\mathbf{v}'\mathbf{A}^{-1}\hat{\mathbf{u}} = 0$ :

$$\mathbf{v}'\hat{\mathbf{u}} = 0 \Leftrightarrow \mathbf{v}'\mathbf{A}^{-1}\hat{\mathbf{u}} = 0 \quad [31]$$

More generally, to a system of constraints for a fixed effect,  $\mathbf{C}\hat{\mathbf{u}} = 0$ , corresponds a system of constraints for a random effect  $\mathbf{C}^*\hat{\mathbf{u}} = 0$ , where  $\mathbf{C}^* = \mathbf{C}\mathbf{A}^{-1}$ :

$$\mathbf{C}\hat{\mathbf{u}} = 0 \Leftrightarrow \mathbf{C}\mathbf{A}^{-1}\hat{\mathbf{u}} = \mathbf{C}^*\hat{\mathbf{u}} = 0 \quad [32]$$

$\mathbf{C}$  and  $\mathbf{C}^*$  have the same rank and the same number of independent constraints, whether  $\mathbf{u}$  is fixed or random.

Relationship [31] holds for  $\mathbf{V} = \mathbf{1}$ .  $\mathbf{Z}\mathbf{1}$  is the vector of the row sums of  $\mathbf{Z}$  and is therefore equal to  $\mathbf{1}$ ,  $\mathbf{1}$  is a linear combination of columns of  $\mathbf{X}$  and  $\mathbf{M}\mathbf{1}$  is equal to 0 by applying [3]. Then  $\mathbf{Z}'\mathbf{M}\mathbf{Z}\mathbf{1} = 0$ , and:

$$(\mathbf{\Omega} - \mathbf{A})(\mathbf{A}^{-1}\mathbf{1}) = 0 \quad [33]$$

and we get the well-known equality (eg, Foulley *et al*, 1990):

$$\mathbf{1}'\mathbf{A}^{-1}\hat{\mathbf{u}} = 0 \quad [34]$$

corresponding to the fixed-effect constraint:

$$\mathbf{1}'\hat{\mathbf{u}} = 0 \quad [35]$$

If the design is connected, the only constraint to set for a fixed  $\mathbf{u}$  is [35], and then the corresponding constraint for a random  $\hat{\mathbf{u}}$  is [34]. All the eigenvectors of  $\mathbf{Z}'\mathbf{M}\mathbf{Z}$  corresponding to a non-null eigenvalue are orthogonal to  $\mathbf{1}$  and the sum of their elements is null. These eigenvectors then correspond to contrasts.

Similarly, all the eigenvectors  $\mathbf{\beta}$  of [6] associated with eigenvalues different from 1 are  $\mathbf{A}$ -orthogonal to  $\mathbf{A}^{-1}\mathbf{1}$ , ie are such that  $\mathbf{\beta}'\mathbf{A}\mathbf{A}^{-1}\mathbf{1} = 0 = \mathbf{\beta}'\mathbf{1}$ . These eigenvectors then also correspond to contrasts. Consequently, all the non-null eigenvalues of  $\mathbf{\Theta}$  are CD of contrasts. In order to study the influence of design disconnectedness, we can then restrict our interest to the set of contrasts.

### ***Disconnectedness, inestimability and information supply***

If  $\mathbf{u}$  is treated as fixed and if the design is disconnected,  $\text{rank}(\mathbf{Z}'\mathbf{M}\mathbf{Z}) = r < n - 1$ . These are  $r$  positive eigenvalues and  $r$  corresponding eigenvectors that are linear estimable contrasts. Since the set of estimable contrasts is a vector space, every contrast that is a linear combination of these eigenvectors is estimable, and at most  $r$  independent contrasts are estimable. However, every contrast that cannot be expressed as a linear combination of these eigenvectors is not estimable. Then, non-estimable contrasts can be sums of estimable and non-estimable contrasts.

When  $\mathbf{u}$  is random, for the above design we have:

$$\text{rank}(\mathbf{\Theta}) = \text{rank}(\mathbf{Z}'\mathbf{M}\mathbf{Z}) = r < n - 1$$

It can easily be shown from [28] that the set of vectors with a null CD, or without information supply, is a vector space. Its dimension equals the multiplicity of the

null eigenvalue of  $\Theta$ , that is  $n - r$ . As  $\mathbf{1}$  belongs to this space, the subspace of contrasts without information supply is a  $(n - r - 1)$ -dimensional space. There are at most  $(n - r - 1)$  independent contrasts that have no information supply. Every contrast without information supply is then a linear combination of these  $(n - r - 1)$  contrasts. However, the CD of every contrast that cannot be expressed as a linear combination of these vectors is positive. In contrast to the fixed-effects case, in which a sum of a non-estimable contrast and of an estimable contrast is not estimable, a contrast with a positive CD can be sum of a contrast with a positive CD and a contrast with a null CD.

If we define disconnectedness in terms of information supply by the experiment rather than contrast inestimability, we can extend this concept to random-effects factors. Whether the effects are fixed or random, there is a disconnection, provided that for at least 1 contrast, no information is supplied by the experiment. However, the fixed-effects case is more restrictive, since there are more independent contrasts with positive CD in the random-effects case than independent estimable contrasts in the fixed-effects case. An example will be presented in the numerical applications.

**Interpretation of  $\rho_1$ ,  $\rho_2$  and  $\rho_3$**

The 3 criteria,  $\rho_1$ ,  $\rho_2$  and  $\rho_3$ , are functions of  $\mu_i$ , the eigenvalues of  $\Theta$ . If they are sorted in ascending order, we have:

$$\rho_1 = \sum_{i=1}^n \mu_i / (n - 1) \tag{36}$$

$$\rho_2 = \left[ \prod_{i=2}^n \mu_i \right]^{\frac{1}{n-1}} \tag{37}$$

$$\rho_3 = 1 - \left[ \prod_{i=1}^n (1 - \mu_i) \right]^{\frac{1}{n-1}} \tag{38}$$

The  $\mu_i$  vary from 0 to 1, as do the criteria. They are equal when all the eigenvalues are equal. Otherwise, we have the following inequalities:

$$\rho_3 \geq \rho_1 \geq \rho_2$$

The dispersion of the eigenvalues and therefore the dispersion of the criteria reflect the design unbalancedness (Chakrabarti, 1963).

$\rho_2$  is more sensitive to low eigenvalues. A null value leads to a null  $\rho_2$ , which indicates that there exists at least 1 contrast without information supply and that the design is disconnected.  $\rho_3$  is sensitive to values of eigenvalues close to 1. If a  $\mu_i$  equals 1, then so does  $\rho_3$ .

### Subpopulation of animals

These criteria are the averaged values of CD, which can include all the evaluated animals. They can be easily restricted to a particular set of  $q$  interesting animals, by working with the submatrices of  $\mathbf{A}$  and  $\mathbf{\Omega}$  pertaining to these animals,  $\mathbf{A}^*$  and  $\mathbf{\Omega}^*$ , respectively. If this set does not include all the animals with performance, the eigenvectors with positive CD are no longer contrasts, and use of [6] leads to overall criteria with a slightly different interpretation: they are no longer averaged values of the CD of contrasts, but averaged values of the CD of all possible linear combinations of the genetic values.

Equation [6] can be modified in order to force the eigenvectors to be contrasts (Darroch and Mosimann, 1985), and then becomes:

$$[\mathbf{T}(\mathbf{A}^* - \mathbf{\Omega}^*)\mathbf{T}' - \nu\mathbf{A}^*]\boldsymbol{\gamma} = 0 \quad [39]$$

where  $\mathbf{T} = \mathbf{I} - \mathbf{1}\mathbf{1}'\mathbf{A}^{*-1}(\mathbf{1}'\mathbf{A}^{*-1}\mathbf{1})^{-1}$

The smaller eigenvalue of [39] is null. Furthermore, the eigenvalues of [39] are between the eigenvalues of [6]:

$$1 > \mu_q \geq \nu_q \geq \mu_{q-1} \geq \nu_{q-1} \geq \dots \geq \mu_1 \geq \nu_1 = 0 \quad [40]$$

Use of the  $q-1$  largest eigenvalues of [39] yields overall criteria that are averaged CD of contrasts. Such a procedure is used in the second numerical example.

Let us note that from [40] using the eigenvalues of [6] instead of [39] would lead to good approximations of these criteria when  $n$  is large. Moreover, if there is a disconnection, this approximate procedure leads to a null value of  $\rho_2$ . In which case,  $0 = \nu_2 \geq \lambda_1$ , and then  $\lambda_1 = 0$ .

Let us note that models including pedigree animals, *ie* without performance, are trivially disconnected. For each pedigree animal, there is a null eigenvalue of  $\Theta$ .

## NUMERICAL APPLICATIONS

### *An animal model example*

#### Data

A hypothetical animal model example with 12 animals (5 with performances) is presented here. The model consists of a herd effect and an animal genetic effect. The heritability was 0.5. Data and pedigree structure have been presented in table I.

#### Results

The rank of  $\Theta$  is 2. Considering  $\mathbf{u}$  as fixed, 10 constraints are needed in order to let the model have full rank. Seven animals have no performance and so must be set to 0, and we also have to set 3 other constraints (1 per herd). Then,  $\text{rank}(\Theta) = \text{rank}(\mathbf{Z}'\mathbf{M}\mathbf{Z}) = 12 - 10 = 2$ . There are only 2 independent contrasts:  $u_1 - u_2$  and  $u_4 - u_5$ .

The complete system of constraints for a fixed  $\mathbf{u}$  is  $\mathbf{C}\hat{\mathbf{u}} = 0$ , where:

$$\mathbf{C} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The first 3 rows of  $\mathbf{C}$  express the within-herd constraints; the other rows are the trivial constraints about the pedigree animals without performance. The corresponding built-in system of constraints for a random  $\mathbf{u}$  from [32], is  $\mathbf{C}^*\hat{\mathbf{u}} = \mathbf{C}\mathbf{A}^{-1}\hat{\mathbf{u}} = 0$  with:

$$\mathbf{C}^* = \begin{bmatrix} 2 & 2 & 0 & 0 & 0 & -1 & -1 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 2 & 2 & 0 & -1 & -1 & -1 & 0 & 0 & -1 \\ -1 & 0 & -1 & 0 & 0 & 2 & 0 & 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & -1 & 0 & -1 & 0 & 0 & 2 & 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1.5 & 0 & 0 & 0 & 0.5 \\ -1 & 0 & 0 & -1 & 0 & 0.5 & 0.5 & 0 & 2 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0 & 1.5 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0 & 1.5 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0.5 & 0 & 0 & 0 & 1.5 \end{bmatrix}$$

**Table I.** Data and pedigree structure.

<i>Animal</i>	<i>Sire</i>	<i>Dam</i>	<i>Herd</i>
1	6	9	1
2	7	10	1
3	6	11	2
4	7	9	3
5	8	12	3
6	—	—	—
7	—	—	—
8	—	—	—
9	—	—	—
10	—	—	—
11	—	—	—
12	—	—	—

The last 7 rows of  $C^*$  are the mixed-model equations about the pedigree animals without performance. Two of the eigenvalues of  $\Theta$  are 0.5; the others are null.  $\rho_1$ ,  $\rho_2$  and  $\rho_3$  are equal to 0.083, 0 and 0.109, respectively. The precisions of the individual and between-animal comparisons are presented in table II.

**Table II.** Individual and between-animal comparison precision\*.

<i>Animal</i>	1	2	3	4	5	6	7	8	9	10	11	12
1	0.266	0.500	0.104	0.260	0.320	0.078	0.289	0.195	0.078	0.289	0.133	0.195
2		0.266	0.203	0.260	0.320	0.289	0.078	0.195	0.289	0.078	0.133	0.195
3			0.016	0.133	0.133	0.016	0.102	0.039	0.039	0.070	0.016	0.039
4				0.250	0.500	0.156	0.125	0.287	0.125	0.156	0.125	0.281
5					0.250	0.156	0.312	0.062	0.312	0.156	0.125	0.062
6						0.062	0.156	0.062	0.031	0.125	0.031	0.062
7							0.125	0.156	0.125	0.031	0.062	0.156
8								0.062	0.156	0.062	0.031	0.000
9									0.125	0.156	0.062	0.156
10										0.062	0.031	0.062
11											0.000	0.031
12												0.062

\* On-diagonals are individual CDs:  $CD(i) = 1 - \lambda\Omega_1(i,i)/A(i,i)$ ; off-diagonals are  $CD(x_{i,j})$ ,  $x_{i,j}$  being the contrast between animals  $i$  and  $j$ .  $CD(x_{i,j}) = 1 - \{\lambda(\Omega_1(i,i) + \Omega_1(j,j) - 2\Omega_1(i,j))/(A(i,i) + A(j,j) - 2A(i,j))\}$ .

There are 11 independent contrasts with non-null CDs on the first row of table II, while there are only 2 independent estimable contrasts in the fixed effects case. (This illustrates the discussion in the section *Disconnectedness, inestimability and information supply*, the number of independent contrasts with positive CDs is

greater than the rank of  $\Theta$ .) Table II shows that comparison CDs are usually low. The most precise comparisons are those between recorded animals in the same herd (1-2, 4-5). Similarly, for animals with no performance, the most precise comparisons are those between the animals with progeny recorder in the same herd (6-7, 9-10). The least precise comparisons are for the triplets, "animal-sire-dam", where the relationship is important.  $CD(x_{8,12}) = 0$ , and concerns 2 mates evaluated from the performance of the same progeny. No other information indicates whether there is an assortative mating. However,  $CD(x_{7,8})$  is quite high (equal to 0.125) compared with other matings (CDs equal to 0.031). Apart their common progeny, each has another progeny (1 and 2), raised in the same herd, and the other matings have just 1 product, or another progeny, but raised in different herds. The effect of the design can be seen here in the precision of the comparison.

**Application to a sire model**

Let us study a hypothetical model containing the fixed effect year (5 years) and 11 sires (2 tested sires per year and a reference sire used over 5 years) according to table III. Within each year, the number of progeny of the first sire was  $n_1$  and the number of progeny of the second sire was  $n_2$ . The reference sire had  $m$  progeny per year and was unrelated to the tested sires. All the tested sires were related by a relationship coefficient  $\gamma$ . The heritability is noted  $h^2$ .

**Table III.** Structure of the design.

	Sires	1	2	3	4	5	6	7	8	9	10	11
Years												
A		$n_1$	$n_2$									$m$
B				$n_1$	$n_2$							$m$
C						$n_1$	$n_2$					$m$
D								$n_1$	$n_2$			$m$
E										$n_1$	$n_2$	$m$

The values of different criteria, the individual CDs, and the peculiar contrasts CDs ( $CD_{1-2}$ : comparison of 2 sires born in the same year ( $u_1 - u_2$ );  $CD_{1-3}$ : comparison of 2 sires born in different years ( $u_1 - u_3$ );  $CD_y$ : comparison of the genetic levels of 2 years ( $u_1 + u_2 - u_3 - u_4$ ) were computed according to different values of  $m$ ,  $h^2$ ,  $\gamma$  and the unbalancedness of the design. The evaluation of the reference sire is not interesting and the overall criteria were computed from the submatrices pertaining to the 10 tested sires, according to [39]. These results are given in tables IVa-d. The comparisons between sires born in the same year are the most precise, and the comparisons between genetic levels of the years are the least precise.

All the precisions decrease with unbalancedness (table IVa), especially  $CD_{1-2}$  (40% of the decrease between balancedness ( $n_1 = 25$ ) and extreme unbalancedness ( $n_1 = 5$ )), while  $CD_y$  remains about the same. Correlatively,  $\rho_3$  is more sensitive to unbalancedness than  $\rho_2$  (39 and 29% of the decrease between balancedness and

**Table IVa.** Variation of precision with respect to the unbalancedness of the design\*.

$n_1$	$n_2$	$CD_1$	$CD_2$	$\rho_1$	$\rho_2$	$\rho_3$	$CD_{1-2}$	$CD_{1-3}$	$CD_y$
25	25	0.289	0.289	0.339	0.248	0.375	0.535	0.315	0.095
40	10	0.273	0.216	0.282	0.216	0.303	0.434	0.296	0.090
45	5	0.239	0.147	0.218	0.175	0.227	0.319	0.258	0.090

\* $h^2 = 0.2$ ;  $m = 5$ ;  $\gamma = 0.125$ .

extreme unbalancedness, respectively);  $\rho_3$  is more sensitive to changes in high values of CDs.

The comparisons between genetic levels of years are the most affected by variations of  $m$  (table IVb);  $CD_y$  goes from 0 ( $m = 0$ ), *ie* disconnection, to 0.161 ( $m = 10$ ).  $CD_{1-2}$  is the same whatever  $m$ , which could have been expected: a reference sire does not affect the comparisons of within-year sires. Since the low CDs increase with  $m$ ,  $\rho_2$  is more sensitive to variations in  $m$  (27% of variation for  $\rho_2$  when  $m$  goes from 10 to 4, compared with 10 and 7% for  $\rho_1$  and  $\rho_3$ , respectively).

**Table IVb.** Variation of precision with respect to the number of progeny of the reference sire ( $m$ )\*.

$m$	$CD_1$	$\rho_1$	$\rho_2$	$\rho_3$	$CD_{1-2}$	$CD_{1-3}$	$CD_y$
0	0.234	0.268	0	0.318	0.535	0.268	0
1	0.251	0.307	0.130	0.353	0.535	0.279	0.022
2	0.263	0.316	0.173	0.359	0.535	0.289	0.042
3	0.273	0.324	0.204	0.365	0.535	0.298	0.061
4	0.282	0.332	0.228	0.370	0.535	0.307	0.079
5	0.289	0.339	0.248	0.375	0.535	0.315	0.095
10	0.318	0.369	0.314	0.396	0.535	0.348	0.161

\* $h^2 = 0.2$ ;  $n_1 = n_2 = 25$ ;  $\gamma = 0.125$ .

The precision decreases when the relationship between sires increases (table IVc). When the sires are unrelated,  $CD_{1-2}$  is equal to the individual CDs found in selection index theory. Precision increases with heritability (table IVd).

**Table IVc.** Variation of precision with respect to the relationship between sire ( $\gamma$ )\*.

$\gamma$	$CD_1$	$\rho_1$	$\rho_2$	$\rho_3$	$CD_{1-2}$	$CD_{1-3}$	$CD_y$
0	0.332	0.363	0.270	0.404	0.568	0.338	0.107
0.125	0.289	0.339	0.248	0.375	0.535	0.315	0.095
0.250	0.260	0.313	0.223	0.343	0.497	0.290	0.082
0.500	0.237	0.246	0.167	0.264	0.397	0.227	0.056

\* $h^2 = 0.2$ ;  $n_1 = n_2 = 25$ ;  $m = 5$ .



**Table IVd.** Variation of precision with respect to the heritability ( $h^2$ )\*.

$h^2$	$CD_1$	$\rho_1$	$\rho_2$	$\rho_3$	$CD_{1-2}$	$CD_{1-3}$	$CD_y$
0.1	0.190	0.221	0.148	0.236	0.359	0.204	0.049
0.2	0.289	0.339	0.248	0.375	0.535	0.315	0.095
0.3	0.354	0.417	0.324	0.469	0.639	0.389	0.139
0.4	0.401	0.474	0.386	0.539	0.709	0.445	0.181

\* $\gamma = 0.125$ ;  $n_1 = n_2 = 25$ ;  $m = 5$ .

These results are relatively trivial. But beyond them we can see how to adapt a precision study according to the aim of the experiment. For a selection experiment, the precision of comparison between genetic levels of different years (*eg*,  $CD_y$ ) is to be maximized. Since the precision of this type of comparison is low,  $\rho_2$ , is more sensitive to low precision variations and would be an interesting parameter. On the contrary, the first aim of a routine evaluation is to compare animals to each other; contrast precision like  $CD_{1-2}$  or  $CD_{1-3}$  must be examined. Since these precisions are relatively high, variations of  $\rho_3$  should be examined.

This example is, of course, oversimplified and the above remarks need to be refined by more realistic studies.

## DISCUSSION AND CONCLUSION

We have assumed throughout this paper that the variance ratio  $\lambda$  was known, which is never the case. Leaving aside the uncertainty about  $\lambda$  leads one to an underestimation of  $\text{var}(\mathbf{u}|\hat{\mathbf{u}})$  (Harville and Carriquiry, 1992), and, then, to an overestimation of the precision.

Nevertheless, even if disconnectedness never occurs in the strictest sense for random effects, its effect is not negligible. It leads to contrasts that are surely estimable, but whose values are null. The concept of estimability is, in the framework of random effects, over-optimistic, and should be replaced by the more realistic notion of information supplied by the data.

This notion is related to the CD, where information is supplied by data for a contrast when its CD is positive. The CD of a contrast is a precision criterion for a comparison between animals, and can be interpreted in the same way as the CD of individuals. Its use allows the validation of particular comparisons. These can be used, for instance, in genetic progress studies to look at the precision of comparison between animals born in different years. They could also be used in cluster analysis, in order to build groups of animals that are comparable to each other, as in Foulley *et al* (1990).

The overall criteria evaluate the precision level of a set of animals. This set can include all the analysed animals, or a particular group of animals, which allows the comparison of designs. A parallel can be drawn between our criteria and optimal design theory criteria (Coursol, 1980; Steinberg and Hunter, 1984): maximization of  $\rho_1$  and **A**-optimality (maximization of the trace of the coefficient matrix); and maximization of  $\rho_2$  and **D**-optimality (maximization of the determinant of the

coefficient matrix). The optimal design research methods could then be adapted to the context of genetic evaluation, with, however, one important restriction: the relative impossibility for the breeder to act on a design, which he can often modify only by some incitement to use more artificial insemination (AI). This is done for French beef cattle, within the framework of the natural service bull progeny test (Foulley and Sapa, 1982; Laloë *et al*, 1992). More recently, for beef cattle evaluation from field data with an animal model, the rule of publication of bull genetic values have been set, based on a minimal use of AI in bulls within the herds (INRA, Nouvel Institut de l'Élevage, 1992). The rules have been set relatively empirically. A study based upon our criteria could lead to optimal rules, combining minimal precision and a maximal number of published bulls.

Use of such criteria becomes impossible as soon as the analysis involves more than 1 000 animals. Approximations or simplifications similar to those presented in Foulley *et al* (1992), considering models consisting only of environmental effects and phantom groups, could be found.

A method presented by Boichard *et al* (1992) yields a reasonably accurate approximation of  $\text{tr}(\mathbf{A}^{-1}\mathbf{\Omega}_1)$  for animal models with 1 class of fixed effects and 1 class of random effects which can be used for large data sets. In their examples, the bias in percent of the true value of the trace was less than 4%. Since  $\rho_1$  is a simple function of  $\text{tr}(\mathbf{A}^{-1}\mathbf{\Omega}_1)$ , ( $\rho_1 = \{n - \lambda \text{tr}(\mathbf{A}^{-1}\mathbf{\Omega}_1)\}/(n - 1)$ ) this method can be used in order to approximate  $\rho_1$  for this special kind of model, even for large data sets.

Another approach would be to approximate these matrices by size-reduced matrices. These matrices would be built from parameters such as the respective distribution of natural service and AI sires across the herds, as well as the number of performances per herd. Criteria would be computed from these matrices. This approach will be used in the context of French beef cattle evaluation from field data.

## REFERENCES

- Boichard D, Schaeffer LR, Lee AJ (1992) Approximate restricted maximum likelihood and approximate prediction error variance of the Mendelian sampling effect. *Genet Sel Evol* 24, 331-343
- Chakrabarti C (1963) On the C matrix in design for experiments. *J Indian Stat Assoc* 1, 8-23
- Coursol J (1980) Techniques statistiques des modèles linéaires, I. *Aspects théoriques* CIMPA
- Darroch JN, Mosimann JE (1985) Canonical and principal components of shape. *Biometrika* 72, 241-252
- Foulley JL, Sapa J (1982) The French evaluation program for natural beef bulls using AI sire progeny as herd ties. *Br Cattle Breed Club Winter Conf*, Jan 1982, Cambr UK
- Foulley JL, Bouix J, Goffinet B, Elsen JM (1990) Connectedness in genetic evaluation. *Advances in Statistical Methods for Genetic Improvement of Livestock* (D Gianola, K Hammond, eds) Springer, Heidelberg, 302-337

- Foulley JL, Hanocq E, Boichard D (1992) A criterion for measuring the degree of connectedness in linear models of genetic evaluation. *Genet Sel Evol* 24, 315-330
- Golub GH, Van Loan CF (1983) *Matrix Computations*. Johns Hopkins Univ Press, MD, USA
- Graybill FA (1961) *An Introduction to Linear Statistical Models, vol I*. McGraw-Hill, NY
- Graybill FA (1983) *Matrices with Applications in Statistics*. Wadsworth, Belmont
- Harville DA, Carriquiry AL (1992) Classical and Bayesian prediction as applied to an unbalanced mixed linear model. *Biometrics* 48, 987-1003
- Henderson CR (1973) Sire evaluation and genetic trends, *Proceedings of the animal breeding and genetics symposium* Aug 1972, 10-41. American Society of Animal Science, Champaign, Il, USA
- Henderson CR (1984) *Applications of linear models in animal breeding*, University of Guelph, Guelph, Ontario, Canada
- INRA-NIE (1991) Evaluations génétiques des taureaux entre troupeaux par la méthode "BLUP modèle animal". Recommandations pratiques pour la connexion. *Races bovines allaitantes*, Oct 1991
- Kullback S (1968) *Information Theory and Statistics*. Dover, NY
- Kullback S (1983) Kullback information *Encyclopedia of Statistical Sciences* (S Kotz, NL Johnson, eds) John Wiley and Sons, NY 4, 421-425
- Laloë D, Renand G, Sapa J, Ménissier F (1992) Use of relationship matrix in the evaluation of natural service Limousin bulls. *Genet Sel Evol* 24, 137-145
- Steinberg DM, Hunter WG (1984) Experimental design: review and comment. *Technometrics* 26, 71-97

## APPENDIX

### *Precision in a balanced-sire model*

The model includes a fixed effect (the mean) and the sire effect. The  $n$  sires are unrelated. Each sire has  $t$  progeny. The coefficient matrix is as follows (for convenience we will write the matrices with  $n = 3$ )

$$\begin{aligned}
 & \begin{bmatrix} tn & t & t & t \\ t & t + \lambda & 0 & 0 \\ t & 0 & t + \lambda & 0 \\ t & 0 & 0 & t + \lambda \end{bmatrix} \\
 \mathbf{Z}'\mathbf{M}\mathbf{Z} + \lambda\mathbf{I} = & \begin{bmatrix} \lambda + \frac{t(n-1)}{n} & -\frac{t}{n} & -\frac{t}{n} \\ -\frac{t}{n} & \lambda + \frac{t(n-1)}{n} & -\frac{t}{n} \\ -\frac{t}{n} & -\frac{t}{n} & \lambda + \frac{t(n-1)}{n} \end{bmatrix}
 \end{aligned}$$

This matrix has a special pattern of the type:

$$\begin{bmatrix} a & b & b \\ b & a & b \\ b & b & a \end{bmatrix}$$

It can be shown (eg, Graybill, 1982, theorem 8.5.2, p 206) that such a matrix has 2 eigenvalues:

-  $a + (n - 1)b$ , with multiplicity 1, and the corresponding eigenvector is proportional to 1;

-  $a - b$  with multiplicity  $(n - 1)$ , and the corresponding eigenvectors  $\mathbf{B}'$  are contrasts ( $\mathbf{B}'\mathbf{1} = 0$ ).

The eigenvalues of  $\mathbf{Z}'\mathbf{M}\mathbf{Z} + \lambda\mathbf{I}$  are  $\lambda$  and  $t + \lambda$ . The eigenvalues of  $\mathbf{\Omega}_1 = (\mathbf{Z}'\mathbf{M}\mathbf{Z} + \lambda\mathbf{I})^{-1}$  are  $1/\lambda$  and  $1/(t + \lambda)$ . Hence, the eigenvalues of  $\mathbf{\Omega} = \lambda\mathbf{\Omega}_1$  are respectively 1 and  $\lambda/(t + \lambda)$ .

Since  $\mathbf{A} = \mathbf{I}$ , [7] is reduced to:

$$[\mathbf{\Omega} - (1 - \mu)\mathbf{I}]\mathbf{\beta} = 0,$$

and we get:

-  $\mu_1 = 0$ , the corresponding eigenvector is proportional to 1;

-  $\mu_2 = \dots = \mu_n = 1 - \{\lambda/(t + \lambda)\} = t/(t + \lambda)$ . The  $(n - 1)$  corresponding eigenvectors are contrasts, and span the  $(n - 1)$ -dimensional vector space of the contrasts. The CD of any contrast is then equal to  $t/(t + \lambda)$ .

## APPENDIX II

Given  $\mathbf{X}$ ,  $\mathbf{Z}$ , and  $\mathbf{A}$ , the distributions of  $\mathbf{u}$ ,  $\mathbf{u}$  given  $\hat{\mathbf{u}}$ , and  $\hat{\mathbf{u}}$  are (Henderson, 1973; 1984):

$$\begin{aligned} f(\mathbf{u}) &: N_n(0, \mathbf{A}\sigma_a^2) \\ g(\mathbf{u}|\hat{\mathbf{u}}) &: N_n(\hat{\mathbf{u}}, \mathbf{\Omega}_1\sigma_e^2) \\ p(\hat{\mathbf{u}}) &: N_n(0, \mathbf{\Psi}) \end{aligned}$$

where  $\mathbf{\Psi} = \mathbf{A}\sigma_a^2 - \mathbf{\Omega}_1\sigma_e^2$ ,

The joint distribution of  $\mathbf{u}$  and  $\hat{\mathbf{u}}$  is:

$$\begin{aligned} q(\mathbf{u}, \hat{\mathbf{u}}) &: N_{2n}(0, \mathbf{\Sigma}_1) \\ \text{where } \mathbf{\Sigma}_1 &= \begin{pmatrix} \mathbf{A}\sigma_a^2 & \mathbf{\Psi} \\ \mathbf{\Psi} & \mathbf{\Psi} \end{pmatrix} \end{aligned}$$

and the product of the marginal distributions is:

$$f(\mathbf{u}) \otimes p(\mathbf{u}) : N_{2n}(0, \mathbf{\Sigma}_2)$$

where  $\mathbf{\Sigma}_2 = \begin{pmatrix} \mathbf{A}\sigma_a^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Psi} \end{pmatrix}$

We have, from [17]:

$$\begin{aligned}
 \mathbf{I} &= E_{\hat{\mathbf{u}}}(\mathbf{I}[g(\mathbf{u}|\hat{\mathbf{u}}) : f(\mathbf{u})]) = E_{\hat{\mathbf{u}}} \left( \int \log \frac{g(\mathbf{u}|\hat{\mathbf{u}})}{f(\mathbf{u})} g(\mathbf{u}|\hat{\mathbf{u}}) d\mathbf{u} \right) \\
 &= \int \int \log \frac{g(\mathbf{u}|\hat{\mathbf{u}})}{f(\mathbf{u})} g(\mathbf{u}|\hat{\mathbf{u}}) p(\hat{\mathbf{u}}) d\mathbf{u} d\hat{\mathbf{u}} \\
 &= \int \int \log \frac{g(\mathbf{u}|\hat{\mathbf{u}}) p(\hat{\mathbf{u}})}{f(\mathbf{u}) p(\hat{\mathbf{u}})} \mathbf{g}(\mathbf{u}|\hat{\mathbf{u}}) p(\hat{\mathbf{u}}) d\mathbf{u} d\hat{\mathbf{u}} \\
 &= \int \int \log \frac{q(\mathbf{u}, \hat{\mathbf{u}})}{f(\mathbf{u}) p(\hat{\mathbf{u}})} q(\mathbf{u}, \hat{\mathbf{u}}) d\mathbf{u} d\hat{\mathbf{u}} \\
 &= \mathbf{I}[q(\mathbf{u}, \hat{\mathbf{u}}) : f(\mathbf{u}) \otimes p(\hat{\mathbf{u}})]
 \end{aligned}$$

We then apply [18]. Since:

$$\begin{aligned}
 \text{i) } \det(\Sigma_1^{-1} \Sigma_2) &= \left[ \text{deg}(\Sigma_2^{-1} \Sigma_1) \right]^{-1} \\
 &= \left| \begin{pmatrix} \mathbf{A}^{-1} \sigma_a^{-2} & \mathbf{0} \\ \mathbf{0} & \Psi^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{A} \sigma_a^2 & \Psi^{-1} \\ \Psi^{-1} & \Psi^{-1} \end{pmatrix} \right|^{-1} = \left| \begin{pmatrix} \mathbf{I} & \Psi \mathbf{A}^{-1} \sigma_a^{-2} \\ \mathbf{I} & \mathbf{I} \end{pmatrix} \right|^{-1} \\
 &= \left| \begin{pmatrix} \mathbf{I} & \mathbf{I} - \lambda \mathbf{A}^{-1} \Omega_1 \\ \mathbf{I} & \mathbf{I} \end{pmatrix} \right|^{-1} \\
 &= \left| \mathbf{I} - (\mathbf{I} - \lambda \mathbf{A}^{-1} \Omega_1) \right|^{-1} = \det(\mathbf{A} \Omega^{-1})
 \end{aligned}$$

$$\text{ii) } \text{tr}(\Sigma_1(\Sigma_2^{-1} - \Sigma_1^{-1})) = \text{tr}(\Sigma_2^{-1} \Sigma_1) - \text{tr}(\Sigma_1^{-1} \Sigma_1) = 2n - 2n = 0$$

iii) the expectations of both distributions are null, we get:

$$\begin{aligned}
 \mathbf{I} &= 1/2 \left[ \log \left( \det \Omega^{-1} \mathbf{A} \right) \right] \\
 \mathbf{I} &= 1/2 \left[ \log \left( \det [\mathbf{I} - \Theta]^{-1} \right) \right] = -1/2 [\log(\det(\mathbf{I} - \Theta))]
 \end{aligned}$$

or:

$$\mathbf{I} = -1/2 \sum_{i=1}^n \log(1 - \mu_i)$$

where  $\mu_i$  are eigenvalues of  $\Theta$  or [6].