Estimation of Genetic Parameters

Estimation of Genetic Parameters: principles

Introduction

The estimation of genetic parameters is an important issue in animal breeding. First of all, estimating additive genetic and possible non-additive genetic variances contributes to a better understanding of the genetic mechanism. Secondly, estimates of genetic and phenotypic variances and covariances are essential for the prediction of breeding values (selection index and BLUP) and for the prediction of the expected genetic response of selection programmes. Parameters that are of interest are heritability, genetic and phenotypic correlation and repeatability, and those are computed as functions of the variance components.

Estimation of heritability is based on methods that determine resemblance between genetically related animals. Roughly, there are two methods that can be used. 1) the resemblance between parents and offspring. If we plot the observations on offspring against the values of their parents (either sires, or dams, or their average), we can perform Offspring-Parent regression. The slope of the regression line though is plot reflects how much of the phenotypic differences that we find in parents are retrieved in their offspring. The expected value of the regression line is $b_{OP} = 0.5h^2$. (or h^2 when regression is on midparent mean) . Offspring-parent regression is not often used in practice. It requires data on 2 generations, and uses only this data. It is also not able to utilize genetic relationships among parents. However, the method is robust against selection of parents.

2) The estimation of variance components (within and between family components). If the variation within families is large relative to differences between families, the trait must be lowly heritable. Variance components are attributed to specific effects. For example, the (paternal) half-sib variance is due to differences between sires. The variance component represents the sire variance, which is a quarter of the additive genetic variance.

Estimation of variance components is easier to generalise, and this method is generally used to estimate genetic parameters. This chapter will therefore mostly deal with variance component estimation.

In analysing data, we are promptly faced with variances. With each set of data we assume a (mixed) model that explains the observations. In this, we make a distinction between **fixed effects**, that determine the level (expected means) of observations, and **random effects** that determine variance. A model at least exists of one fixed (mean) and one random effect (residual error variance). If observations also are influenced by a genetic contribution of the animals, then a genetic variance component exists as well. In that situation, we have two components contributing to the total variance of the observations: a genetic and a residual variance component. If we calculate responses in breeding programs, we make use of those parameters. In predicting breeding values, a model can be applied which use both fixed and random effects and Best Linear Unbiased Prediction (**BLUP**). Variances and covariances are assumed to be known.

When to estimate variance components?

In general, the estimation of variances and covariances has to be based on a sufficient amount of data. Depending on the data structure and the circumstances during measuring, estimations can be based on some hundreds (selection experiments) or more than 10,000 observations (field recorded data). It is obvious that we are not interested in estimating variance components from every data set. The information in literature is in many cases even better than estimations based on a small data set. In general, we have to estimate variance if:

- we are interested in a new trait, from which no parameters are available;
- variances and covariances might have changed over time
- considerable changes have occurred in a population e.g. due to recent importations.

Mostly it is assumed that variances and covariances, and especially the ratio of both of them (like heritability, correlation), are based on particular biological rules, which do not rapidly change over time. However, it is well known that the genetic variance changes as consequence of selection. Changes are especially expected in situations with short generation intervals, high selection intensities or high degrees of inbreeding or in a situation in which a trait is determined by only a few genes. Secondly, the circumstances under which measurements are taken can change. If conditions are

getting more uniform over time, the environmental variance decreases, and consequently the heritability increases. Thirdly, the biological interpretation of a trait can change as consequence of a changed environment; feed intake under limited feeding is not the same as feed intake under ad-lib feeding. In conclusion, there are sufficient reasons for regular estimation of (co-)variance components.

Variance components

What are variance components? Measures of extend of differences or variability is indicated with variance. Variance is always related to a particular effect, that has an impact on observations. When we want to compute the variance on n observations (vector **y**), then an estimator for the variance is,

$$var(y) = \sum_{i=1}^{n} (y_i - y)^2 / (n - 1)$$

The statistical model that describes those observations is,

$$y_i = \mu + e_i$$

An estimator for μ is the average of y. The differences between an observation and μ, (y_i -y.), are the random deviations as consequence of the residual (or error-) effect (e_i). In this situation, the variance of y is equal to the variance of only the random component in the model (var(y) = var(e), is the residual variance). In the numerator, the estimator of var(e) contains the **Sum of Squares** that can be ascribed to the residuals. The **expectation** of the sum of squares is equal to the multiplication of a coefficient times the variance component. In this situation it is equal to the degrees of freedom, which remain for the residual effect. Therefore, the variance is an average of the squared differences as consequence of the concerning effect.

In a situation with more random effects, we are able to estimate more variance components. For this, we first have to quantify the contribution of each random effect. Afterwards we can compute the sum of squares for each of them. The test and estimation procedure widely used, is **ANOVA**. In balanced data, it is rather simple to estimate variance components, by setting the "Mean Squares" equal to their expectations. Those expectations are linear functions of the variance components.

As an example, we can take a simple model with one main sire effect (a_i).

$$y_{ij} = \mu + a_i + e_{ij}$$

Assume **N** observations, with **s** sires, with N/s=**n** is the number daughters per sire. Then, the **ANOVA** table is as follows,

Source	df	Sum of Squares	Mean Squares	EMS
Mean	1	SSM	SSM	
Sires	s-1	SSA	SSA/(s-1)	$n\sigma_s^2 + \sigma_e^2$
Error	N-s	SSE	SSE/(N-s)	σ_e^2
Total	N	SST		

where

$$SST = \sum_{i=1}^{s}\sum_{j=1}^{n} y_{ij}^2 \quad 1$$

y_{ij} is an observation on the j^{th} daughter of the i^{th} sire. The total sum of squares (SST) is therefor the sum of each of the observations squared.

$$SSM = N * \bar{y}_{..}^2$$

The mean sum of squares is therefore N times the means squared.

$$SSA = n \sum_{i=1}^{s} (\bar{y}_{i.} - \bar{y}_{..})^2$$

The sum of squares due to a particular effect (e.g. the sire effect) is therefore the sum over all observations of the estimated (sire) effect in each observation squared (in balanced data this is the difference between the progeny group mean of a sire and the overall mean).

$$SSE = \sum_{i=1}^{s} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{i.})^2$$

The sum of squares due to the residual (error) is the sum over all observations of the residual effect in each observation squared (this is the difference between the observation and its group mean).

From the ANOVA table we can calculate estimates of variance components as

$$\hat{\sigma}_e^2 = SSE / (N - s)$$

and $$\hat{\sigma}_s^2 = [(SSA / (s-1)) - \hat{\sigma}_e^2] / n$$

Notice that the sum of squares for the main effect (SSA) is the sum of all the squared estimates of a_i, because in a balanced data set the estimate of a_i is equal to $(y_{i.} - y_{..})$. In a **balanced data**, it is rather simple to form the expectations for each sum of squares, because the number of observations per class of a is constant (n).
Originally, in **unbalanced data**, the same technique was applied, using for each class a weighted average. Henderson (1953) developed analogue techniques for unbalanced data. Because of the use of vector notation those techniques became popular for use in computer programmes, like Harvey and SAS. In essence techniques are the same as in balanced data, using an ANOVA table with the sum of squares for the different effects and their expectations.

Methods used for estimation of genetic parameters

The methods of **Henderson** use **Least Squares equations** and variance components are estimated of certain quadratics (usually differences between quadratics) and their expectations. The variance of the estimates is not minimized (i.e. the estimation is not the most accurate) because sums of squares and expectations are not dependent on the variance-covariance structure of the data but rather on LS equations. Estimates of variances are unbiased but can fall outside the parameter space (e.g. they can be negative). Estimates are also not unique because, when there are several random effects, sums of squares due to random effects can be computed in several ways, i.e. corrected for several combinations of other effects.

ML (**M**aximum **L**ikelihood)-estimators maximize the likelihood of the parameters given the **density functions** and the **data**. Estimates are not unbiased but they have smaller variance then the unbiased estimators.
REML (**Re**stricted **ML**) estimators maximize the likelihood of the parameters **after correcting for the fixed effects** (formally: in the space orthogonal to the fixed effects). In ML methods the loss in degrees of freedom due to correction for fixed effects is not taken into account. In REML this loss in degrees of freedom is accounted for. Different quadratic forms are calculated **based on the mixed model equations**

In most algorithms to obtain REML estimates, iteration is used. This process starts with a certain set of variance components and stops when the set of variance components which results in the highest likelihood is found. REML estimators are within the parameters space by definition but therefore they are biased. There are several algorithms to compute REML and in practice some algorithms give even negative estimates (therefore formally not REML).

Choice of the best method to estimated variances is not obvious. One could choose for unbiasedness but in the practice of estimating variances accuracy (minimal variance) is usually preferred. It is also important to notice that unbiased methods use least squares equations and therefore can not correct for selection in animal breeding data, e.g. through the relationships matrix or by using correlated traits. In animal breeding, data used to estimate variance components frequently originates from selection experiments or livestock improvement schemes, which involve continuous culling of animals on the basis of their performance or breeding values. In that case, ANOVA estimators, which assume that data are randomly sampled, tend to be subject to selection bias. Under certain conditions (RE)ML will account for selection, because it makes use of the mixed model equations. This very important feature has made REML the method of choice for most animal breeding applications.

Genetic parameters

Variance components provide us with genetic parameters such as

Heritability (h^2) = VA / VP = additive genetic variance / total phenotypic variance

or 4* sire variance / phenotypic variance

In an *animal model* we fit the additive genetic effect of the animal and the variance of this term gives us the additive genetic variance. In a model where we fit the effect of sire, we estimate *sire variance* and this needs to be multiplied by 4 to get additive genetic variance.

Genetic Corelation = r_g = Cov (A1,A2)/ sqrt(VA1*VA2)
= genetic covariance divided by the product of genetic SD.

Repeatability = (VA + Vep) / VP
= sum of VA and Permanent Environm. Variance divided by Phenotypic Variance

Models of analysis

Developments in variance component estimation specific to animal breeding have been closely linked with advances in the genetic evaluation of animals by Best Linear Unbiased Prediction. Early REML applications were generally limited to models largely equivalent to those in corresponding ANOVA type analysis, considering one random effect only and estimating genetic variances from paternal half sib covariances (so-called sire model).

Recently Animal Model (AM) has come to dominate genetic evaluation schemes, allowing information on all known relationships between animals to be incorporated in the analysis. With the introduction of the AM, expanded models that are more accurate were described, e.g. models with maternal, permanent environmental, cytoplasmic or dominance effects or effects at QTL. These effects are fitted as additional random effects. Maximum likelihood based methods appeared to be most flexible to accommodate such models. In terms of (RE)ML estimation of variance components has changed thinking from the expectation of mean squares and the interpretation of observational components of variance in genetic terms (e.g. variance between and within half sib families) to a more direct approach of calculating a likelihood of a data vector for a given model with a given set of parameters, and maximizing this likelihood. Such models indeed can be complicated with several random effects and covariances amongst the levels of each random effect to be specified e.g. additive genetic or dominance relationships.

There is obviously an advantage in using (RE)ML methods that are more flexible in handling animal breeding data on several (overlapping) generations (and possibly

several random effects). However, the use of such methods has a danger in the sense that we need not to think explicitly anymore about **data structure**. To estimate, as an example, additive genetic variance, we need to have a data set that contains a certain family structure that allows us to separate differences between families from differences within families. Or in other words, we need to separate genetic and residual variance. ANOVA methods require more explicit knowledge about such structure, since the data has to be ordered according to family structures (e.g. by half sib groups). Such ordering is not necessary in Likelihood Estimation. Some REML packages may even allow estimation based on data that have single records per animal and no family structure. Obviously, such data does not allow estimation of heritability.

In these notes we assume a mixed model with one random effect only. This could be either a sire effect or an animal effect. Different derivations and algorithms are easier to follow for models with one random effect only, but we will bear in mind that most methods, and particularly (restricted) maximum likelihood has been extended and applied to more complicated models.

An simple example

To get some feel for why it is useful to calculate sums of squares in the construction and testing of statistical models for prediction, consider the following example.
Suppose we have 4 observations and a one-way classification with 2 levels (A and B).
Calculate the sum of squares for the total, the mean, the model and the residual.
Residual 1 refers to a model where only the mean is fitted and residual 2 to a model where also the class effect is fitted.
Calculate sums of squares 'by hand' based on the numbers the column.

	class	Observation	Mean	Residual 1	Predicted Y	Residual 2
	A1	8				
	A1	9				
	A2	11				
	A2	12				
Sum of Squares						

Compare SS calculations with expressions for SST, SSM, SSA and SSE on page 107.

Methods to estimate genetic parameters

Henderson's method 3

Genetic parameters have been estimated for many years using analysis of variance (ANOVA) or analogous methods. The ANOVA method has been popular because standard software like SAS provides such estimates.

In general, these methods require that individuals can be assigned to groups with the same degree of relationship for all members. Family structures considered most often are paternal half-sib groups or full-sib groups. In the case of paternal half-sib group all offspring of one sire are treated as one group and offspring of different sires are allocated to different groups.

Using ANOVA, the covariance among members of a family or group of relatives is usually determined as the variance component between groups. For example, in case of a sire model, the variance between sires σ_s^2 and variance within sires σ_e^2 . As shown earlier, the sire variance $\sigma_s^2 = ¼\sigma_a^2$ while the variance within sires is $.75\sigma_a^2+\sigma_e^2$.

Calculating the variance between groups, involves partitioning the sum of squared observations (SS) due to different sources of variation in the model of analysis, groups of relatives being one of them, and equating the corresponding mean squares. Mean squares are derived as the SS divided by the associated degrees of freedom, to their expectations. The same principle applies for multivariate analyses but considering sums of cross-products between traits instead of SS. For balanced data, the partial SSs are orthogonal and their expected values are simple linear combinations of the variance components between groups so that calculations are straightforward, even for multiple cross-classifications, and estimators are unique.

Data arising from animal genetics are usually not balanced but methods analogous to the ANOVA have been developed for unbalanced data. In particular, Henderson's (1953) method 3 of 'fitting constants' has found extensive use. This approach replaces the Sums of squares (SS) in the balanced ANOVA by quadratic forms involving the least squares solutions of effects for which variances are to be estimated. Its widespread application was greatly aided by the availability of a 'general' least-squares computer program tailored towards applications commonly arising in animal breeding (Harvey, 1977). Henderson method 3 is also implemented in the statistical package SAS.

Consider a mixed linear model for one trait, represented by

$$\mathbf{Y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$$

with **y**, **b**, **u** and **e** representing the vectors of observations, fixed effects, random effects (e.g. sire) and residual errors, respectively, and **X** and **Z** the corresponding design matrices. Assume all levels of **u** pertain to the same source of variation, for example sires, and that $V(\mathbf{u})= \sigma_u^2 \mathbf{I}$, $V(\mathbf{e})= \sigma_e^2 \mathbf{I}$ and $cov(\mathbf{u},\mathbf{e}')=0$.

The Least Squares equations are:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'Y \end{bmatrix}$$

Absorbing the fixed effects reduces the equations to

$$\mathbf{Z'MZ}\hat{\mathbf{u}} = \mathbf{Z'My}$$

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

When the inverse of (**X'X**) does not exist, a generalized inverse can be used in its place.

Method 3 estimates of variance components are then:

$$\hat{\sigma}_e^2 = \frac{\left(y'y - \hat{u}'Z'y - \hat{b}'X'y\right)}{\left(N - r(X) - r(Z) + 1\right)}$$

$$\hat{\sigma}_u^2 = \frac{\left(\hat{u}'Z'My - (r(Z) - 1)\hat{\sigma}_e^2\right)}{tr(Z'MZ)}$$

with r(**X**) and r(**Z**) denoting the column rank of **X** and **Z**, respectively, N the number of observations, and tr the trace operator. In this method any covariances between levels of **u** (i.e. relations between sires) are ignored. An extension of method 3 to account for relationships between **u** has been considered by Sørensen and Kennedy (1986).

{An analogy with the earlier ANOVA methods can be seen as follows:

The expression **y'y** is a vector notation for 'total sum of squares'. Expressions like $\hat{b}'X'y$ (solution multiplied by right hand side) can also be written as $y'X(X'X)^{-1}X'y$ since $\hat{b} = (X'X)^{-1}X'y$ (we ignore now random effects). Since X'y contains the class totals, and X'X contains the number of observations per class, the expressions $y'X(X'X)^{-1}X'y$ gives the sum of the class totals squared, divided by the number of observations per class. This is exactly the Sum of squares due to the *b*-effect. The expression $y'y - \hat{u}'Z'y - \hat{b}'X'y$ is therefore equal to the residual sum of squares.}

Restricted Maximum Likelihood

General interest in Maximum Likelihood estimators of variance components has been propelled by their desirable statistical properties: they are consistent, asymptotically normal and efficient. Harville (1977) has given an extensive review of ML estimation. Furthermore, the ML framework provides a great deal of flexibility, allowing for designs and models for analysis which cannot be accommodated by ANOVA type of estimators. Initial interest in ML, to estimate both genetic parameters and fixed effects, was stimulated by concern about bias due to selection. A number of simulation studies have illustrated that selection can be accounted for by REML (Sorensen and Kennedy, 1984; Van der Werf and De Boer, 1990) when the complete mixed model is used with all genetic relationships and all data used for selection included.

Restricted Maximum Likelihood is a ML method that accounts for the loss of degrees of freedom due to fitting fixed effects. Patterson and Thompson (1971) formally described REML. The procedure requires that **y** have a multivariate normal distribution although various authors have indicated that ML or REML estimators may be an appropriate choice even if normality does not hold (Meyer, 1990).

Over the last decade, extensive research effort has been directed towards the development of specialized and efficient algorithms for particular classes of models. These procedures will be discussed on the following pages. Before starting with that, let me give a brief introduction. In ML and REML the aim is to find the set of parameters which maximizes the likelihood of the data. The likelihood of the data for a given model can be written as a function. From calculus we know that we can find the maximum of a function by taking the first derivative and set that equal to zero. Solving

that would result in the desired parameters (assuming that we did not find the minimum, this can be checked using second derivatives). The first and second derivatives of the likelihood function are complicated formulas. Different algorithms have been developed which try to circumvent this problem. An overview of different methods is given by Meyer (1990).

Principle of Maximum Likelihood

Suppose we have a variable y with mean μ and standard deviation σ. The normal distribution of this variable can be represented as $y = N(\mu, \sigma^2)$. A mathematical representation of a density function for a normally distributed variable is

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(y-\mu)^2}{\sigma^2}}$$

This is called the **Probability Density Function** (PDF) of y.

A function for a multidimensional normal distribution $\mathbf{y}=N(\mathbf{Xb}, \mathbf{V})$ is

$$f(y) = \frac{1}{2\pi^{\frac{1}{2}N}|V|^{\frac{1}{2}}} e^{-\frac{1}{2}(y-Xb)'V^{-1}(y-Xb)}$$

where N is the length of **y** and $|\mathbf{V}|$ is the determinant of **V**. The function f(**y**) is called a density function of **y**. The function gives the probability of finding a certain y given the parameters. The parameters are the means in **Xb** ("*location parameters*") and the variances in **V** ("*dispersion parameters*"). However, this function can also be used the other way around: if we have observed data, it gives us the probability of having such data for certain parameter values. Therefore, the probability density function can be used as a likelihood function as well. When the data **y** is known, f(**y**) is a likelihood function and this function can be maximized in the parameters, i.e. we want to find the parameters for which f(**y**) has the highest value. Instead of maximizing f(**y**) we can also maximize the ^{e}log of f(y); L(**b,V** | **X, y**), which is the log likelihood function:

$$L(b,V|X,y) = -\tfrac{1}{2}N\log(2\pi) - \tfrac{1}{2}\log(V) - \tfrac{1}{2}(y - Xb)'V^{-1}(Y - Xb) \qquad [1]$$

This function gives the likelihood of the unknown parameters **b** and **V** given the observed data **y** and the design matrix **X**. The matrix **V** depends on the variance components we are interested in. V has susesually a known design (e.g. genetic relationships) but is proportional to unknown parameter values, e.g.

V = ZAZ'σ_a^2 + I σ_e^2. The maximum likelihood estimates of the parameters are obtained by maximizing the likelihood function.

In Restricted Maximum Likelihood as suggested by Patterson and Thompson (1971) the likelihood function of the data is maximized '*in the space of error contrasts*'. In other words, the density function is maximized after correcting all observations first for the fixed effects

Methods available to get REML estimates can be divided in the following groups:

1) Methods using first derivatives of the likelihood function.
2) Methods using first and second derivatives of the likelihood function.
3) Derivative free methods.

For models with more random factors it is more difficult to find the maximum and it is also more difficult to construct derivatives. In categories 1 and 2, the derivatives can be calculated exact but in most methods approximations are used.

REML using derivatives

Methods which use both first and second derivatives, i.e. geometrically speaking information on slope and curvature of the function, have been found to converge quickest (Meyer, 1989). However, even for simple models, calculation of actual or expected second derivatives was initially computationally highly demanding if not prohibitive. Therefore, initially many REML applications were based on the so-called Expectation-Maximization (EM) algorithm. This requires, implicitly, first derivatives of the likelihood to be evaluated. The resulting estimators then have the form of quadratics in the vector of random effects solutions, obtained by BLUP for the assumed values of variances to be estimated, which are equated to their expectations. For the mixed model equations:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + \hat{\alpha}\,\mathrm{A}^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'Y \end{bmatrix}$$

Note that $\hat{\alpha}$ is a function of the variance parameters that need to be estimated. Therefore, initially a prior (starting) value of α is used. The REML estimates of variance components using the EM algorithm can be obtained as:

$$\sigma_a^2 = \left[\hat{a}'\mathrm{A}^{-1}a + tr(\mathrm{A}^{-1}\mathrm{C})\sigma_e^2 \right] / q$$

$$\sigma_e^2 = \left[y'y - \hat{b}'X'y - \hat{a}'Z'y \right] / \left(N - r(\mathrm{X})\right)$$

where N is the number of observations, q is the number of random genetic effect levels and **C** the part of the inverse of the mixed model equations that corresponds with the random effects. The model can contain animal effects and, and **a** would denote the vector of additive genetic effects.

The EM algorithms have the property of always yielding positive estimates as long as prior values (values which are used to start the calculations) are positive (Harville, 1977). The EM algorithm is not very difficult to program, because all elements which needed can be derived from the mixed model equations. What is needed for each round of iteration is the solutions to the mixed model equations and the trace of the inverse of the random part of the coefficient matrix. This last element is computationally the most difficult part. Iterative methods can be used to obtain estimates for the fixed and random effect but the EM algorithm requires the direct inverse of a matrix of size equal to the number of levels of the random effects, in each round of iteration. This imposes restrictions on the kind of analyses feasible, especially for multivariate analyses.

The EM algorithm is an iterative procedure to get estimates. One starts the process with solving the equations for a given (prior) value of the variance components. These values are used in estimating the effects of the model (α depends on the assumed levels of the variance components). This results in a new value for the variance components and the corresponding value of α. In an iterative process, the old values and the new value of the next iteration round are becoming more and more similar, and ultimately converge (when the difference is very small) to a solution

Derivative Free REML (DFREML)

In the development of algorithms to compute REML an approach that did not make use of derivatives proved to be particularly successful to compute variance components from an animal model. This approach is called a derivative free approach, and was first introduced by Smith and Graser (1986) and Graser et al. (1987). The maximum is found by comparing likelihood values of different parameter values.

The likelihood function from [1] can be re-written. First it is written after eliminating ('correcting for') the fixed effects. This is called the Restricted Maximum Likelihood. Secondly, it is re-written in terms of elements that relate to the mixed model equations:

$$\log L = -\frac{1}{2}\left[const + q\log\sigma_a^2 + N\log\sigma_e^2 + \mathrm{y'Py} + \log|\mathrm{W}| + \log|\mathrm{A}|\right]$$

where **W** is the coefficient matrix of the mixed model equations. The log |**A**| is a constant which does not depend on the parameters of interest (genetic relations between animals are constant for a given data set) and does not have to be evaluated. The matrix **P** is quite complicated, but Smith and Graser (1986) and Graser *et al.* (1987) showed that **y'Py** represents the sum of squares of residuals. With the log determinant of the coefficient matrix (log |**W**|) it can be evaluated simultaneously by augmenting **W** by the vector of right hand sides and the total SS (**y'y**) and absorbing all rows and columns into the latter. The augmented mixed model array is:

$$\begin{bmatrix} \mathrm{y'y} & \mathrm{y'X} & \mathrm{y'Z} \\ \mathrm{X'y} & \mathrm{X'X} & \mathrm{X'Z} \\ \mathrm{Z'y} & \mathrm{Z'X} & \mathrm{Z'Z} + \alpha\mathrm{A}^{-1} \end{bmatrix}$$

Absorption, which is also referred to as Gaussian Elimination, is used to calculate the quantities **y'Py** and log |**W**|. The residual variance can be estimated as **y'Py**/(N-r(X)) so that log L can be maximized with respect to one parameter only, the variance ratio α, estimating subsequently σ_a^2/α.

This principle has been extended to models including additional random effects, such as environmental effect due to litters or a maternal genetic effect, and to multivariate analyses (Meyer, 1989).

The derivative free algorithm has been applied in the DFREML programmes that are written and distributed by Karin Meyer. These programs can be used for uni- and multivariate analysis and for models with several random effects.
Groeneveld (1991) presented a second package for estimating variance components using a derivative free approach. This programme is distributed under the name VCE and has its setup is similar to that of PEST.
A more robust and efficient algorithm analysis is 'Average Information REML', now applied by the DFREML package. A very powerful program for parameter estimation is the ASREML package (Gilmour et al., 1996).

REML using the Average Information algorithm.

First we discuss more formally first and second derivatives of the likelihood function. Then, the mechanism of an AI algorithm will be presented.

The partial derivative with respect to the vector of fixed effects, b, in equation 2 can be calculated by using a general result for matrix derivatives (Lynch and Walsh, 1998),

$$\frac{\partial\left[(y-Xb)'V^{-1}(y-Xb)\right]}{\partial b} = -2X'V^{-1}(y-Xb) \qquad [3]$$

It is noted that from equation 1 and 3,

$$\frac{\partial L(Xb,\sigma_i^2)}{\partial b} = X'V^{-1}(y-Xb) \qquad [4]$$

The partial derivatives of ln |V| in equation 1 with respect to the variance of random effects, σ_i^2 (e.g. i = a and e) can be obtained from matrix theory (Searle, 1982)

$$\frac{\partial}{\partial\sigma_i^2}\ln|V| = tr\left[V^{-1}\frac{\partial V}{\partial\sigma_i^2}\right] \qquad [5\ a]$$

Differentiating V^{-1} with respect to σ_i^2 results in

$$\frac{\partial V^{-1}}{\partial \sigma_i^2} = -V^{-1}\frac{\partial V}{\partial \sigma_i^2}V^{-1} \qquad [5\text{ b}]$$

where, tr (trace) is the sum of the diagonal elements of a square matrix.

From equation 5a, the first derivatives of L from equation 1 with respect to the variance components can be obtained as,

$$\frac{\partial L(Xb,\sigma_i^2)}{\partial \sigma_i^2} = -\frac{1}{2}tr(V^{-1}V_i) + \frac{1}{2}(y - X\hat{b})'V^{-1}V_iV^{-1}(y - X\hat{b}) \qquad [6\text{ a}]$$

and from equation 5b, the second derivative is,

$$\frac{\partial^2 L(Xb,\sigma_i^2)}{\partial \sigma_i^2 \partial \sigma_j^2} = \frac{1}{2}tr(V^{-1}V_iV^{-1}V_j) - (y - X\hat{b})'V^{-1}V_iV^{-1}V_jV^{-1}(y - X\hat{b}) \qquad [6\text{ b}]$$

where, $\frac{\partial V}{\partial \sigma_i^2}$ is simplified as V_i.

Suppose V is defined as ZAZ' σ_a^2 + I σ_e^2, then the following equation evaluates the expression in equation 6 a and 6 b.

$$\frac{\partial V}{\partial \sigma_i^2} = V_i = \begin{cases} ZAZ & (i = a) \\ I & (i = e) \end{cases} \qquad [7]$$

As Lynch and Walsh (1998) stated, the ML estimators are obtained by making the first derivative of L (equation 6a) equal to zero and solving, thus equation 6 a gives

$$tr(V^{-1}V_i) = (y - X\hat{b})'V^{-1}V_iV^{-1}(y - X\hat{b}) = y'\hat{P}V_i\hat{P}y \qquad [8]$$

where $P = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}$

From equation 7 and 8, ML equations are

$$tr(V^{-1}(ZAZ')) = y'\hat{P}(ZAZ')\hat{P}y \qquad \text{for } \sigma_a^2 \qquad [9\text{ a}]$$

$tr(V^{-1}) = y'\hat{P}\hat{P}y$ for σ_e^2 [9 b]

Similarly, but removing the fixed effects from the model, REML equations can be derived from ML equations. In REML we use a transformation matrix K such that,

KX = 0 [10]

Multiplying the K with the mixed linear model (y = Xb + Za + e) replaces

y by Ky ~ N (0, KVK')

X by KX = 0

Z by KZ

V by KVK' [11]

Applying equation 11 to equation 9a or 9b yields,

$tr((KVK')^{-1}KV_iK') = y'K'(KVK')^{-1}KV_iK'(KVK')^{-1}Ky$ [12]

Searle *et al*. (1992) proved the following equation.

P = K' (KVK')$^{-1}$ K [13]

Therefore, equation 12 can be

$tr(PV_i) = y'\hat{P}V_i\hat{P}y$ [14]

Equation 6a and 6b can be transformed as,

$$\frac{\partial L}{\partial \sigma_i^2} = -\frac{1}{2}tr(PV_i) + \frac{1}{2}y'PV_iPy \qquad [15\ a]$$

$$\frac{\partial^2 L}{\partial \sigma_i^2 \partial \sigma_j^2} = \frac{1}{2}tr(PV_iPV_j) - y'PV_iPV_jPy \qquad [15\ b]$$

From equation 14, REML equations are

$tr(P(ZAZ')) = y'\hat{P}(ZAZ')\hat{P}y$ for σ_a^2 [16 a]

$tr(P) = y'\hat{P}\hat{P}y$ for σ_e^2 [16 b]

The Average Information Algorithm for REML estimation

Various techniques for solving ML / REML equations have been introduced (e.g. the Newton-Rhapson algorithm, Fisher's scoring method and DF algorithm).

In this section, the Newton-Rhapson algorithm and Fisher's scoring method are firstly described, and the Hessian matrix and the Fisher information matrix are derived. This may help to understand the property of the AI algorithm and the AI matrix. And then, the method for estimating the elements of AI matrix is described which is key process for the AI algorithm.

Average Information from the Hessian and the Fischer information matrix

The Newton-Raphson algorithm obtains the REML estimate using the following equation (Lynch and Walsh, 1998).

$$\Theta^{(k+1)} = \Theta^{(k)} - (H^{(k)})^{-1} \frac{\partial L}{\partial \Theta}\Big|\Theta^{(k)} \qquad [17]$$

where Θ is a vector of parameters, k is k th iteration, $\frac{\partial L}{\partial \Theta}$ is a column vector of the first derivatives of the log likelihood function with respect to each parameter, and H is the Hessian matrix elements of which are the second derivatives of the log likelihood function with respect to the variance components. From equation 7 and 15 b, the Hessian matrix for variance components (residual variance and additive genetic variance) is,

$$\mathrm{H} = \frac{\partial^2 L}{\partial \sigma_i^2 \partial \sigma_j^2} = \frac{1}{2}\begin{bmatrix} tr(PP) - 2y'PPPy & tr(PA^*P) - 2y'PA^*PPy \\ tr(PA^*P) - 2y'PA^*PPy & tr(PA^*PA^*) - y'PA^*PA^*Py \end{bmatrix} \qquad [18]$$

where $A^* = ZAZ'$

In Fisher's scoring method, the inverse of the Hessian matrix in equation 18 is replaced by its expected value (Lynch and Walsh, 1998).

$$\Theta^{(k+1)} = \Theta^{(k)} + (F^{(k)})^{-1} \frac{\partial L}{\partial \Theta} \Big| \Theta^{(k)} \qquad [19]$$

where F is the Fisher information matrix.

$$\text{F} = -\text{E}\left(\frac{\partial^2 L}{\partial \sigma_i^2 \partial \sigma_j^2}\right) = \frac{1}{2}\begin{bmatrix} tr(PP) & tr(PA^*P) \\ tr(PA^*P) & tr(PA^*PA^*) \end{bmatrix} \qquad [20]$$

The average information from the Hessian and Fisher's information matrix is,

$$\text{AI} = (-\text{H} + \text{F}) / 2 = \frac{1}{2}\begin{bmatrix} y'PPPy & y'PA^*PPy \\ y'PA^*PPy & y'PA^*PA^*Py \end{bmatrix} \qquad [21]$$

From equation 17 and 19,

$$\Theta^{(k+1)} = \Theta^{(k)} + (AI^{(k)})^{-1} \frac{\partial L}{\partial \Theta} \Big| \Theta^{(k)} \qquad [22]$$

The method for calculating the elements of the AI matrix requires term like $V_i Py$, which are referred to as working variates, $y(\sigma_i^2)$ (Gilmour *et al*., 1995).

$$y(\sigma_i^2) = V_i Py$$

According to Johnson and Thompson (1995), the working variate for additive genetic variance and residuals are expressed as,

$$y(\sigma_a^2) = \text{A*Py} = \frac{1}{\sigma_a^2} Z\hat{a} \qquad [23\text{ a}]$$

$$y(\sigma_e^2) = \text{Py} = \frac{1}{\sigma_e^2} \hat{e} \qquad [23\text{ b}]$$

where $\hat{a}$ is vector of solutions from the mixed model equation (MME) for the mixed linear model, and $\hat{e} = y - X\hat{b} - Z\hat{a}$.

According to Johnson and Thompson (1995), the elements of the AI matrix can be calculated as a vector product of the working variates from the MME in which y is replaced by the working variates. For example, consider the element, y'PPPy in the AI matrix.

y'PPPy = $y(\sigma_e^2)$ ' P $y(\sigma_e^2)$ [24]

The term, y'PPPy, can be obtained by multiplying the transpose of the column vector of $y(\sigma_e^2)$ [23 b] by the column vector of residuals, P $y(\sigma_e^2)$. All the elements in the AI matrix can be calculated in the same manner.

Alternatively, Gilmour *et al*. (1995) calculated the AI matrix using the Gaussian elimination of the M matrix.

$$M = \begin{bmatrix} y'R^{-1}y & y'R^{-1}X & y'R^{-1}Z \\ X'R^{-1}y & X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}y & Z'R^{-1}X & Z'R^{-1}Z + A^{-1}\sigma_a^2 \end{bmatrix} \quad [25]$$

After performing Gaussian elimination, it is well known that the element of the first row and first column in the M matrix, M* (1,1), equals y'Py (Graser *et al*., 1987). If y is replaced by the working variate for σ_e^2 ($y(\sigma_e^2)$ from equation 23 b), then M* (1,1) after Gaussian elimination equals y'PPPy in the AI matrix. If y is replaced by the working variate for σ_a^2 ($y(\sigma_a^2)$ from equation 23 a), then M* (1,1) after Gaussian elimination equals y'PA*PA*Py in the AI matrix. For a cross product (e.g. y'PA*PPy in the AI matrix), the M matrix will be formed as,

$$\mathrm{M} = \begin{bmatrix} y(\sigma_a^2)'R^{-1}y(\sigma_e^2) & y(\sigma_a^2)'R^{-1}X & y(\sigma_a^2)'R^{-1}Z \\ X'R^{-1}y(\sigma_e^2) & X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}y(\sigma_e^2) & Z'R^{-1}X & Z'R^{-1}Z + A^{-1}\sigma_a^2 \end{bmatrix}$$

After performing Gaussian elimination of M, M* (1,1) equals y'PA*PPy in the AI matrix.

All elements in the AI matrix can be calculated in the same manner.

After establishing the AI matrix, $\frac{\partial L}{\partial \Theta}$ is required for equation 22.

$$\frac{\partial L}{\partial \sigma_i^2} = -\frac{1}{2}tr(PV_i) + \frac{1}{2}y'PV_iPy \qquad \text{(from equation 15 a)}$$

$y'PV_iPy$ can be calculated from equation 23 a and b.

$$\frac{\partial L}{\partial \sigma_u^2} = -\frac{1}{2}(tr(PA^*) - y'PA^*Py) = -\frac{1}{2}\left[\frac{N_a}{\sigma_a^2} - \frac{tr(A^{-1}C^{aa})}{\sigma_a^4} - \left(\frac{\hat{e}}{\sigma_e^2}\right)'\left(\frac{Z\hat{a}}{\sigma_a^2}\right)\right] \quad [26\text{ a}]$$

$$\frac{\partial L}{\partial \sigma_e^2} = -\frac{1}{2}(tr(P) - y'PPy) = -\frac{1}{2}\left[\frac{N - r(X)}{\sigma_e^2} - \left(N_a - \frac{tr(A^{-1}C^{aa})}{\sigma_a^2}\right)\frac{1}{\sigma_e^2} - \left(\frac{\hat{e}'\hat{e}}{\sigma_e^4}\right)\right] \quad [26\text{ b}]$$

The procedure of the AI algorithm

Efficient computational procedures for the AI algorithm for univariate case were described in several studies (Johnson and Thompson, 1995; Gilmour *et al*., 1995).

1. Construction of mixed model equation (MME) or matrix M [25]

2. Calculating log likelihood in current stage of iteration

The log likelihood from equation 2 can be calculated by the following equation (Meyer, 1989).

$$L = -\frac{1}{2}\left[(N - rank(X) - n)\log\hat{\sigma}_e^2 + \log|C| + \log|A| + n\log\hat{\sigma}_u^2 + y'Py\right]$$

where n is the number of animals, | | is the determinant of the matrices.

4. Estimating $\hat{b}$, $\hat{u}$ and $\hat{q}$

For the calculation of the working variates in equation 23, the fixed effects ($\hat{b}$) and additive random effects ($\hat{u}$ and $\hat{q}$) are obtained from the MME. Alternatively, using intermediate terms formed during Gaussian elimination, $\hat{b}$, $\hat{u}$ and $\hat{q}$ are efficiently obtained (see Gilmour *et al*., 1995). Given the effects ($\hat{b}$, $\hat{u}$ and $\hat{q}$), residuals can be obtained (i.e. $\hat{e} = y - X\hat{b} - Z_1\hat{u} - Z_2\hat{q}$).

5. AI matrix using [21], [23], [24] and [25]

6. First derivatives from [26]

7. Update from [22]

8. Convergence

References

Gilmour. A. R, Thompson. R and Cullis. B. R. (1995) Average information REML: An efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics.* 51: 1440-1450.

Gilmour, A.R. R Thompson, B.R. Cullis. and S. Welham 1996. ASREML – Biometrics Bulletin 3, NSW Agriculture 90pp.

Groeneveld, E., 1991. Simultaneous REML estimation of 60 covariance components in an animal model with missing values using the simplex algorithm. Paper presented at 42nd annual meeting of the EAAP, Berlin.

Harvey, W.R, 1977. Users Guide for LSML76. Mixed model least squares and maximum likelihood computer program. U.S. depart. of Agric., ARS.

Harville, D.A., 1977. Maximum Likelihood approaches to variance component estimation and to related problems. J. Amer. Stat. Ass. 72:320-338.

Henderson, C.R., 1953. Estimation of variance and covariance components Biometrics 9:226-252.

Henderson, C.R., 1975a. Multiple Trait sire evaluation using the relationship matrix. J. Dairy Sci. 59:769

Johnson. D. L and Thompson. R. (1995) Restricted Maximum Likelihood Estimation of variance components for univariate animal models using sparse matrix techniques and average information. *J. Dairy. Sci.* 78: 449-456.

Lange, K., Westlake, J., and Spence, M.A. (1976) Extensions to pedigree analysis. III. Variance components by the scoring method. *Annals of Human Genetics* 39: 485-491.

Lynch. M, Walsh. B. (1998) Variance component estimation. In: *Genetics and analysis of quantitative traits.* Eds. Lynch. M and Walsh. B. Sinauer Associaties, Inc. Sunderland, Massachusetts, U.S.A. pp 779-803.

Meyer K. (1989) Restricted maximum likelihood to estimate variance components for animals models with several random effects using a derivative-free algorithm. *Genet. Sel. Evol.* 21: 317-340.

Meyer, K., 1990. Present status of knowledge about statistical procedures and algorithms to estimate variance and covariance components. 4th World Congress on Genet. Applied to Livest. Prod. Edinburgh. p 403.

Patterson, H.D. and R. Thompson, 1971. Recovery of inter-block information when block sizes are unequal. Biometrika 58:545-554.

Searle, S.R., Casella, G., and McCulloch, C.E. (1992) *Variance components*. John Wiley and Sons, NY.

Smith S.P., H.-U. Graser, 1986. Estimating variance components in a class of mixed models by restricted maximum likelihood. J. Dairy Sci 69: 1156-1165.

Sorensen, D.A. and B.W. Kennedy, 1984. Estimation of genetic variance from selected and unselected populations. J. Anim. Sci. 59:1213.

Van der Werf, J.H.J. and I.J.M. de Boer, 1990. Estimation of additive genetic variance when base populations are selected. J. Anim. Sci. 65: