

Genotypic covariance matrices and their inverses for models allowing dominance and inbreeding

SP Smith*, A Mäki-Tanila**

*Animal Genetics and Breeding Unit, University of New England, Armidale,
NSW 2351, Australia*

(Received 11 June 1988; accepted 3 July 1989)

Summary – Dominance models are parameterized under conditions of inbreeding. The properties of an infinitesimal dominance model are reconsidered. It is shown that mixed-model methodology is justifiable as normality assumptions can be met. Tabular methods for calculating genotypic covariances among inbred relatives are described. These methods employ 5 parameters required to accommodate additivity, dominance and inbreeding. Rules for calculating inverse genotypic covariance matrices are presented. These inverse matrices can be used directly to set up the mixed-model equations. The mixed-model methodology allowing for dominance and inbreeding provides a powerful framework to better explain and utilize the observed variation in quantitative traits.

dominance / inbreeding / infinitesimal models / inverse / mixed model / recursion

Résumé – Matrices de covariances génotypiques et leurs inverses dans les modèles incluant dominance et consanguinité. Les modèles de génétique quantitative incluant la dominance sont considérés dans des conditions de consanguinité. Après une discussion des propriétés du modèle infinitésimal, on montre que la méthodologie des modèles mixtes peut être appliquée à cette situation, dans la mesure où les hypothèses de normalité peuvent être satisfaites. On décrit des méthodes tabulaires pour calculer les covariances génotypiques parmi des apparentés consanguins, dont l'emploi nécessite l'introduction de 5 composantes de variances. On présente les règles du calcul direct de l'inverse de ces matrices de covariances génotypiques, connaissant la généalogie, et ces 5 composantes de la variance. Ces matrices inverse peuvent être utilisées directement pour établir les équations du modèle mixte. La méthodologie du modèle mixte, prenant en compte les interactions de dominance et la consanguinité, fournit un cadre pour une meilleure explication et une meilleure utilisation de la variabilité des caractères quantitatifs.

dominance / consanguinité / modèle infinitésimal / inverse / modèle mixte / algorithme récursif

* Permanent address: Department of Animal Breeding, Agricultural Research Centre (MTTK), 31600 Jokioinen, Finland.

** Correspondence and reprints

INTRODUCTION

The mixed linear model has enjoyed widespread acceptance in animal breeding. Most applications have been restricted to models which depict additive gene action. However, there is also concern with non-additive effects within and between breeds and crosses (eg, Hill, 1969; Kinghorn, 1987; Mäki-Tanila and Kennedy, 1986). Henderson (1985) provided a statistical framework for modelling additive and non-additive genetic effects when there is no inbreeding. With inbreeding, the mixed model allows statistical analysis, however, considerable developmental work remains. Inbreeding complicates covariance structures (Harris, 1964). Moreover, inbreeding depression is a manifestation of interactions like dominance and epistasis. Models which include only additive effects and covariates for inbreeding (eg, Hudson and Van Vleck, 1984) are rough approximations.

The proper treatment of inbreeding and dominance involves 6 genetic parameters (Gillois, 1964; Harris, 1964). These parameters define the first and second moments of genotypic values in the absence of epistasis. A genetic analysis is possible by repetitive sampling of lines derived from one population through a fixed pedigree (eg, Chevalet and Gillois, 1977). However, we should like to perform an analysis where the pedigrees are realized with selection and/or random mating. This could be done if an infinitesimal model was feasible and we could apply normal theory and the mixed model. Furthermore, it would be useful to build covariance matrices and inverse structures easily, to enable use of Henderson's (1973) mixed-model equations. This paper shows how to justify and implement these activities. It is an extension of Smith's (1984) attempt to generalize models with dominance and inbreeding.

DOMINANCE MODELS

Finite loci

In this section we introduce the 6 genetic parameters needed to model additivity, dominance and inbreeding depression. These parameters are functions of gene frequency (p_i for the i^{th} allele) in much the same way that heritability depends on gene frequency for purely additive traits.

First, consider the genotypic effect, g_{ij} for 1 locus represented by

$$g_{ij} = \mu + a_i + a_j + d_{ij} \quad (1)$$

where μ is the mean, a_i and a_j are the additive effects for the i^{th} and j^{th} allele, and d_{ij} is the corresponding dominance deviation. Equation (1) represents a system of $r(r+1)/2$ equations in $r+1+r(r+1)/2$ unknowns (ie, μ, a_i, a_j, d_{ij}) where r is the number of alleles. To uniquely determine μ, a_i and d_{ij} requires additional $r+1$ constraints given as:

$$\sum_i p_i a_i = \sum_i p_i d_{ij} = \sum_j p_j d_{ij} = \sum_{ij} p_i p_j d_{ij} = 0 \quad (2)$$

These constraints are derived from effectual definitions applied to populations in Hardy-Weinberg equilibrium.

It follows that in populations undergoing random mating, the additive variance is:

$$\tilde{\sigma}_a^2 = \sum_i p_i a_i^2$$

and the dominance variance is:

$$\tilde{\sigma}_d^2 = \sum_{ij} p_i p_j d_{ij}^2$$

To accommodate inbreeding requires 3 additional parameters: (i) the complete inbreeding depression:

$$\tilde{u}_\delta = \sum_i p_i d_{ii}$$

(ii) the dominance variance among homozygotes:

$$\tilde{\sigma}_\delta^2 = \sum_i p_i d_{ii}^2 - u_\delta^2$$

and (iii) the covariance between additive and dominance effects among homozygotes:

$$\tilde{\sigma}_{a\delta} = \sum_i p_i a_i d_{ii}$$

It is convenient to work with the parameter $\tilde{\delta}^2 = \tilde{\sigma}_\delta^2 + \tilde{u}_\delta^2$ which is a second moment. The symbol “ $\tilde{}$ ” is a reminder that the associated parameter refers to 1 locus.

When there are n loci, then the parameters of interest, say \mathbf{v} , are the single locus terms $\tilde{\mathbf{v}} = \{\tilde{\sigma}_a^2, \tilde{\sigma}_d^2, \tilde{u}_\delta, \tilde{u}_\delta^2, \tilde{\delta}^2$ or $\tilde{\sigma}_\delta^2, \tilde{\sigma}_{a\delta}\}$ “summed” over loci. All parameters in $\mathbf{v} = \{\sigma_a^2, \sigma_d^2, u_\delta, u_\delta^2, \delta^2$ or $\sigma_\delta^2, \sigma_{a\delta}\}$ are formal sums. The column vector of inbreeding depression (u_δ) which is defined as a list of \tilde{u}_δ for loci $1, 2, \dots, n$, is also very useful. Among the parameters, we have the dependancies $u_\delta^2 = u'_\delta u_\delta$ and $\sigma_\delta^2 = \delta^2 - u_\delta^2$.

The parameters \mathbf{v} describe a hypothetical population of infinite size undergoing random mating and inlinkage equilibrium. This population is sometimes referred to as the base population, but we find this usage misleading. In the spirit of Bulmer (1971), let us introduce segregation effects defined as deviations from mid-parent values. In fact, both additive and dominance effects have mid-parents values, as will be seen later. Now we can define \mathbf{v} as parameters that determine the stochastic properties of segregation effects for an observed sample of animals from a known pedigree. Whether or not these segregation effects are representative of some ancestral population (perhaps several generations old) is, of course, questionable. Indeed, ancestral effects associated with a sample of animals can be treated as fixed (Graser *et al*, 1987) and, hence, segregation effects and estimates of \mathbf{v} can be far removed from the ancestral base. This interpretation is robust under selection, with the added assumption that linkage disequilibrium in one generation influences the next generation only through the mid-parent values. Our assumption need only be approximately correct over a few generations (perhaps far removed from the base). It is important to point out that these views are definitional and no method of estimating \mathbf{v} (free of selection bias) has been proposed as yet.

The disruptive forces of genetic drift on our usage of \mathbf{v} are probably of negligible importance; a small population is just another repetition of a fixed pedigree sampled from the base population.

Infinite loci

It is feasible to define an infinitesimal model with dominance (Fisher, 1918). When there is directional dominance, we might observe $|u_\delta|$ going to infinity or σ_δ^2 and σ_d^2 going to zero (Robertson and Hill, 1983). However, it is our belief that this problem is characteristic of particular infinitesimal models, not all infinitesimal models. To show this, we have constructed a counter example.

Because $\sigma_a^2, \sigma_d^2, \sigma_\delta^2, \sigma_{a\delta}$ and u_δ are formal sums, it is necessary (but not sufficient) for the contributions from single loci to be of the order n^{-1} where n is the number of loci; *ie*, if the limit of \mathbf{v} is finite. Whereas, it might seem reasonable to require location effects like \tilde{u}_δ to approach 0 at a rate of $n^{-1/2}$, this is not necessary and it may result in infinite inbreeding depression.

Now let us imagine an infinite number of loci, each with 4 possible alleles, that could be sampled with equal likelihood. Assuming that the dominance deviations for each locus are as given in Table I, these deviations are consistent with constraints (2). In this example it is possible to use any additive effects also consistent with (2), where $\tilde{\sigma}_a^2$ is proportional to n^{-1} . For a particular locus, the inbreeding depression and dominance variances are: $\tilde{u}_\delta = -1/(2n)$; $\tilde{\sigma}_d^2 = 1/(4n^2) + 3/(8n)$; $\tilde{\sigma}_\delta^2 = 1/(4n^2) + 1/(2n)$. Summed over n loci these become: $u_\delta = -1/2$; $\sigma_d^2 = 1/(4n) + 3/8$; $\sigma_\delta^2 = 1/(4n) + 1/2$. Letting n drift to infinity gives the following non-trivial parameters: $u_\delta = -1/2$; $\sigma_d^2 = 3/8$; $\sigma_\delta^2 = 1/2$. This provides our counter example. There does not seem to be an analogous example involving only two alleles. However, the biallelic situation is uninteresting because it implies a singularity: $\tilde{\sigma}_{a\delta}^2 = \tilde{\sigma}_a^2 \tilde{\sigma}_\delta^2$.

Table I. Dominance deviations for four alleles; 1, 2, 3, 4.

	1	2	3	4
1	$-n^{-1}$	$n^{-1/2}$	$-n^{-1/2}$	n^{-1}
2	$n^{-1/2}$	$-n^{-1/2}$	0	0
3	$-n^{-1/2}$	0	$n^{-1/2}$	0
4	n^{-1}	0	0	$-n^{-1}$

The above demonstration may seem artificial because it is spoiled by global changes in gene frequency (WG Hill, 1988, personal communication). However, we can construct other more elaborate counter examples. For instance, let loci vary in their contribution to the parameters. Let there be infinite loci indexed $1, 2, \dots, n$, where $\sigma_d^2, \sigma_\delta^2 \neq 0$, and there is no directional dominance; *ie*, $u_\delta = 0$. Among the partial sum of n loci, we can take approximately $n^{1/2}$ indexed $1, 4, \dots, k^2$,

where $k^2 \leq n < (k+1)^2$. By redefining the contributions from single loci to $\tilde{u}_\delta = -n^{-1/2} + \tilde{u}_\delta$ and $\tilde{\sigma}_d^2, \tilde{\sigma}_\delta^2 \leq n^{-1/2}$ (perhaps by altering gene frequency), we notice that $u_\delta = -1$, and σ_d^2 and σ_δ^2 are non-zero at the limit when n goes to infinity. We can create yet another subsequence with indices $2, 5, \dots, k^2 + 1$, where $k^2 + 1 \leq n < (k+1)^2 + 1$. Taking the previous alterations and assuming for the latter sequence $\tilde{u}_\delta = 1/2n^{-1/2} + \tilde{u}_\delta$, the limit value of u_δ is $-1/2$. It is possible to select a finite number of subsequences and make similar alterations. Each of these sequences becomes infinitely long as n approaches infinity. Hence, infinitesimal models, where $0 < |u_\delta| < \infty$, $\sigma_d^2 \neq 0$ and $\sigma_\delta^2 \neq 0$, are feasible.

With an infinitesimal model, \mathbf{v} is a function of summary statistics that involves gene frequency. Individual gene frequencies have little or no effect on \mathbf{v} . Moreover, genotypic effects summed over loci follow a normal distribution. This implies that selection and genetic drift can be accommodated by the mixed model, as suggested by Bayesian arguments (*eg*, Gianola and Fernando, 1986). In particular, the assumption about the influence of linkage disequilibrium, discussed earlier, is valid under the infinitesimal model.

The real issue is not whether $|u_\delta|$ is infinite or dominance variances are zero, but whether normality and linearity are appropriate assumptions given a finite number of loci. If $|u_\delta|$ is estimated from real data, it will be found to be infinite, although it may be very large. Furthermore, if dominance variances are found to be non-zero, and if many loci are involved, then it would seem that a contrived infinitesimal model (like the ones above) is appropriate. Normal approximations are adequate under most realistic models for genetic variation; there being a small number of major loci and a large number of minor loci (Robertson, 1967). However, with a very small number of loci, these approximations become less adequate with each additional generation of selection.

GENOTYPIC COVARIANCE STRUCTURES

Harris (1964) developed recursion formulae for evaluating the identity coefficients needed to determine covariances among inbred relatives. In a later paper, Cockerham (1971) elaborated on these methods. Using zygotic networks, Gillois (1964) also devised a scheme to evaluate identity coefficients, and Nadot and Vaysseix (1973) published an algorithm for implementing Gillois's procedure.

In this paper, tabular methods for evaluating second moments are presented. These techniques allow the exact evaluation of genotypic covariances without calculating individual identity coefficients. The first class of methods are conceptually easy and are modelled after the genomic table described by Smith and Allaire (1985). The second class (those based on compression) are conceptually more difficult, but perhaps numerically more feasible.

Methods based on gametes

Each animal in a pedigree receives 1 genomic half or gamete from each of its parents. Thus, every animal has 2 genomic halves and the total number of such halves is $r = 2s$, where s is the number of animals.

Let \mathbf{a}_i be a column vector of additive effects, such that the ℓ^{th} element of \mathbf{a}_i equals the additive contribution of the ℓ^{th} locus in the i^{th} gamete. If there are n loci, then \mathbf{a}_i has length n . Under an infinitesimal model, \mathbf{a}_i is infinitely long. Define \mathbf{d}_{ij} as a vector of dominance deviations, typical of the union of gametes i and j . The ℓ^{th} element of \mathbf{d}_{ij} equals the dominance contribution of the ℓ^{th} locus. If i and j are genomic halves from different animals, the vector \mathbf{d}_{ij} depicts the dominance deviations for a phantom animal.

Like animals, gametes have a pedigree; genomic halves in one animal form a parental pair for producing gametes. Let us assume the gametes are ordered such that $i > j$, if gamete i is a descendant of gamete j . Furthermore, let us assume $i > j$ implies that gamete j is a base population gamete if i is. Next, imagine the ordered sequence:

$$\mathbf{I}, \mathbf{a}_1, \mathbf{d}_{11}, \mathbf{a}_2, \mathbf{d}_{21}, \mathbf{d}_{22}, \mathbf{a}_3, \mathbf{d}_{31}, \mathbf{d}_{32}, \mathbf{d}_{33}, \dots, \mathbf{d}_{rr}$$

where \mathbf{I} is an identity matrix of order n . This is a very long list comprising of $(r+1)(r+2)/2$ arrays. Fortunately, we need only select a much smaller subsequence, $\mathbf{G} = \{\mathbf{I}, \mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_p\}$ from this list; *ie*, the arrays that are actually needed for recursive calculations. An algorithm for extracting \mathbf{G} is presented in Appendix A.

The elements of \mathbf{G} are used as row and column headings in a table depicting the second moments $E\{\mathbf{G}'\mathbf{G}\}$ which is represented by:

\mathbf{I}	$E\{\mathbf{g}_1\}$	$E\{\mathbf{g}_2\}$		$E\{\mathbf{g}_p\}$
$E\{\mathbf{g}'_1\}$	$E\{\mathbf{g}'_1\mathbf{g}_1\}$	$E\{\mathbf{g}'_1\mathbf{g}_2\}$	\dots	$E\{\mathbf{g}'_1\mathbf{g}_p\}$
$E\{\mathbf{g}'_2\}$	$E\{\mathbf{g}'_2\mathbf{g}_1\}$	$E\{\mathbf{g}'_2\mathbf{g}_2\}$		$E\{\mathbf{g}'_2\mathbf{g}_p\}$
	\vdots		\ddots	
$E\{\mathbf{g}'_p\}$	$E\{\mathbf{g}'_p\mathbf{g}_1\}$	$E\{\mathbf{g}'_p\mathbf{g}_2\}$		$E\{\mathbf{g}'_p\mathbf{g}_p\}$

This table is referred to as the extended genomic table (*cf* Smith and Allaire, 1985), and is denoted by \mathbf{E} .

Elements of \mathbf{E} are computed by recursion. Starting with the first row, elements are evaluated from left to right. When the first row is completed, the first column is filled in using symmetry. The remaining elements in the second row are evaluated from left to right and the second column is then filled in using symmetry. This process is continued for each additional row and column. The recursions used to compute \mathbf{E} are listed below, where θ is defined as the index set of all base gametes, $i > j$, k, m , $k > m$, and parent gametes of i are x and y . The proofs of these formulae are due to properties involving sums of expectations and conditional expectations. For example:

$$E\{\mathbf{g}'_v \mathbf{g}_w\} = \sum_{\ell} E\{\mathbf{g}_{v\ell} \mathbf{g}_{w\ell}\} = 1/2 \sum_{\ell} E\{\mathbf{g}_{v\ell} \mathbf{g}_{w\ell} | i \equiv x\} + 1/2 \sum_{\ell} E\{\mathbf{g}_{v\ell} \mathbf{g}_{w\ell} | i \equiv y\}$$

where $i \equiv x$ or y represents the event that the ℓ^{th} locus of gamete i is identical by descent to that x or y , respectively. The product $\mathbf{g}'_v \mathbf{g}_w$ is intended to involve the gametes i, j, k and m (v and w are used to identify the associated columns of \mathbf{G}).

(i) First n rows:

$$E\{\mathbf{I}\} = \mathbf{I}; E\{\mathbf{a}_i\} = 0 \quad (\text{all values of } i)$$

$$E\{\mathbf{d}_{ij}\} = \begin{cases} 0 & (i \neq j; i \in \theta) \\ u_{\delta} & (i = j \in \theta) \\ 1/2E\{\mathbf{d}_{xj}\} + 1/2E\{\mathbf{d}_{yj}\} & (i \neq j; i \notin \theta) \\ 1/2E\{\mathbf{d}_{xx}\} + 1/2E\{\mathbf{d}_{yy}\} & (i = j \notin \theta) \end{cases}$$

(ii) Subsequent rows:

(a) Additive and additive

$$E\{\mathbf{a}'_i \mathbf{a}_j\} = \begin{cases} 0 & (i \neq j; i \in \theta) \\ \sigma_a^2 & (i = j \in \theta) \\ 1/2E\{\mathbf{a}'_x \mathbf{a}_j\} + 1/2E\{\mathbf{a}'_y \mathbf{a}_j\} & (i \neq j; i \notin \theta) \\ 1/2E\{\mathbf{a}'_x \mathbf{a}_x\} + 1/2E\{\mathbf{a}'_y \mathbf{a}_y\} & (i = j \notin \theta) \end{cases}$$

(b) Additive and dominance

$$E\{\mathbf{a}'_i \mathbf{d}_{km}\} = \begin{cases} 0 & (i \neq k \text{ or } i \neq m; i \in \theta) \\ \sigma_{a\delta} & (i = k = m \in \theta) \\ 1/2E\{\mathbf{a}'_x \mathbf{d}_{km}\} + 1/2E\{\mathbf{a}'_y \mathbf{d}_{km}\} & (i \neq k, i \notin \theta) \\ 1/2E\{\mathbf{a}'_x \mathbf{d}_{xm}\} + 1/2E\{\mathbf{a}'_y \mathbf{d}_{ym}\} & (i = k; i \neq m; i \notin \theta) \\ 1/2E\{\mathbf{a}'_x \mathbf{d}_{xx}\} + 1/2E\{\mathbf{a}'_y \mathbf{d}_{yy}\} & (i = k = m \notin \theta) \end{cases}$$

$$E\{\mathbf{d}'_{ij} \mathbf{a}_k\} = \begin{cases} 0 & (i \neq k \text{ or } j \neq k; i \in \theta) \\ 1/2E\{\mathbf{d}'_{xj} \mathbf{a}_k\} + 1/2E\{\mathbf{d}'_{yj} \mathbf{a}_k\} & (i \neq j; i \neq k; i \notin \theta) \\ 1/2E\{\mathbf{d}'_{xx} \mathbf{a}_k\} + 1/2E\{\mathbf{d}'_{yy} \mathbf{a}_k\} & (i = j; i \neq k; i \notin \theta) \end{cases}$$

(c) Dominance and dominance

$$E\{\mathbf{d}'_{ij}\mathbf{d}_{km}\} = \begin{cases} 0 & ((i \neq j \text{ or } k \neq m) \& \\ & (i \neq k \text{ or } j \neq m); i\varepsilon\theta) \\ u_\delta^2 & (i = j \neq k = m; i\varepsilon\theta) \\ \sigma_a^2 & (i \neq j; k \neq m; i = k; \\ & j = m; i\varepsilon\theta) \\ \delta^2 & (i = j = k = m\varepsilon\theta) \\ 1/2E\{\mathbf{d}'_{xj}\mathbf{d}_{km}\} + 1/2E\{\mathbf{d}'_{yj}\mathbf{d}_{km}\} & (i \neq j; i \neq k; i \neq \theta) \\ 1/2E\{\mathbf{d}'_{xx}\mathbf{d}_{km}\} + 1/2E\{\mathbf{d}'_{yy}\mathbf{d}_{km}\} & (i = j; i \neq k; i \neq \theta) \\ 1/2E\{\mathbf{d}'_{xj}\mathbf{d}_{xm}\} + 1/2E\{\mathbf{d}'_{yj}\mathbf{d}_{ym}\} & (i \neq j; i = k; i \neq m; i \neq \theta) \\ 1/2E\{\mathbf{d}'_{xx}\mathbf{d}_{xx}\} + 1/2E\{\mathbf{d}'_{yy}\mathbf{d}_{yy}\} & (i = j = k = m; i \neq \theta) \end{cases}$$

the recursive formulae in (c) appears in Smith (1984).

When $i\varepsilon\theta$, the above recursions are initialized assuming that gametes are sampled at random from a single population. For this case, we have additional simplification for all values of i :

$$E\{\mathbf{d}_{ii}\} = u_\delta; E\{\mathbf{a}'_i\mathbf{a}_i\} = \sigma_a^2; E\{\mathbf{a}'_i\mathbf{d}_{ii}\} = \sigma_{a\delta}; E\{\mathbf{d}'_{ii}\mathbf{d}_{ii}\} = \delta^2$$

Now that the recursive structure of \mathbf{E} has been shown, it is possible to describe the algorithm of Appendix A. Define $f(v)$ as the youngest gamete associated with the v^{th} column of $\tilde{\mathbf{G}}$, say \mathbf{g}_v . The matrix or list $\tilde{\mathbf{G}}$ is said to be closed under gametic recursion if the terms used to expand any \mathbf{g}_v by parent gametes of $f(v)$ are also of $\tilde{\mathbf{G}}$. More formally, $(\mathbf{g}_v \mid f(v) = x)$ and $(\mathbf{g}_v \mid f(v) = y)$ are columns of $\tilde{\mathbf{G}}$ when $f(v) \neq \theta$, and has parent gametes x and y . The algorithm in Appendix A is called a depth-first search and it produces sets of vectors closed under recursion. Any element needed to evaluate any recursion can always be found in $\tilde{\mathbf{E}}$. The algorithm of Appendix A can also be used to define the subsequence $\tilde{\mathbf{G}}$ introduced below.

It is possible to combine additive and dominance effects into genotypic effects, say $\tilde{\mathbf{g}}_{ij} = \mathbf{a}_i + \mathbf{a}_j + \mathbf{d}_{ij}$, and use these as row and column headings of a new table. The headings are ordered as some subsequence, say $\tilde{\mathbf{G}}$, of

$$\mathbf{I}, \tilde{\mathbf{g}}_{11}, \tilde{\mathbf{g}}_{21}, \tilde{\mathbf{g}}_{22}, \tilde{\mathbf{g}}_{31}, \tilde{\mathbf{g}}_{32}, \tilde{\mathbf{g}}_{33}, \dots, \tilde{\mathbf{g}}_{rr}$$

The recursions for $E\{\tilde{\mathbf{G}}'\tilde{\mathbf{G}}\}$ are exactly as they are for $E\{\mathbf{d}_{ij}\}$ and $E\{\mathbf{d}'_{ij}\mathbf{d}_{km}\}$, except that initializations (when $i\varepsilon\theta$) are different:

$$E\{\tilde{\mathbf{g}}_{ij}\} = \begin{cases} 0 & (i \neq j) \\ u_\delta & (i = j) \end{cases}$$

$$E\{\tilde{\mathbf{g}}'_{ij}\tilde{\mathbf{g}}_{km}\} = \begin{cases} 0 & (\text{none of } i, j, k, m \text{ are equal}) \\ u_\delta^2 & (i = j \neq k = m) \\ \sigma_a^2 & (i \neq j; k \neq m; i = k \text{ or } j = k \text{ or } j = m) \\ 2\sigma_a^2 + \sigma_{ad} & (3 \text{ of } i, j, k, m \text{ are equal}) \\ 2\sigma_a^2 + \sigma_d^2 & (i = k \neq j = m) \\ 4\sigma_a^2 + 4\sigma_{ad} + \delta^2 & (i = j = k = m) \end{cases}$$

After building a matrix of second moments, the (co)variance matrix (for genetic effects summed over loci) is obtained by absorbing the first n rows. The resulting array is a function of u_δ only through $u_\delta^2 = \mathbf{u}'_\delta \mathbf{u}_\delta$. The vw^{th} element of the absorbed array \mathbf{E} is:

$$\sum_t [E\{g_{vt}g_{wt}\} - E\{g_{vt}\}E\{g_{wt}\}]$$

which reflects the assumption that genotypes are additive over independently segregating loci. This assumption can be relaxed, as linkage disequilibrium can sometimes be accommodated via conditional (*ie*, Bayesian) analyses.

In practice, we never evaluate the entire array \mathbf{E} or $E\{\tilde{\mathbf{G}}'\tilde{\mathbf{G}}\}$. In particular, the first n rows and columns can be represented implicitly by one row and column: rows of:

$$[E\{\mathbf{g}_1\}, E\{\mathbf{g}_2\}, \dots, E\{\mathbf{g}_p\}]$$

are simple multiple of each other. Our purpose is to show structural properties that allow inversion rules. Nevertheless, the above recursions are helpful in evaluating particular moments; *eg*, those needed to compute the inverse. This can be accomplished by adapting Tier's (1990) recursive pedigree algorithm: one calculates only needed moments and avoids redundant calculation. We may add to our recursions, shortcuts for particular degenerate cases:

$$(i) E\{\mathbf{d}'_{ij}\mathbf{d}_{km}\} = \lambda[E\{\mathbf{a}'_i\mathbf{a}_k\}E\{\mathbf{a}'_j\mathbf{a}_m\} + E\{\mathbf{a}'_i\mathbf{a}_m\}E\{\mathbf{a}'_j\mathbf{a}_k\}]$$

$$\text{if } E\{\mathbf{a}'_i\mathbf{a}_j\} = E\{\mathbf{a}'_k\mathbf{a}_m\} = 0$$

$$(ii) E\{\mathbf{d}'_{ij}\mathbf{d}_{km}\} = \lambda E\{\mathbf{a}'_i\mathbf{a}_m\}E\{\mathbf{a}'_j\mathbf{a}_k\} + \tau E\{\mathbf{a}'_i\mathbf{a}_j\}E\{\mathbf{a}'_k\mathbf{a}_m\}$$

$$\text{if } E\{\mathbf{a}'_i\mathbf{a}_k\} = E\{\mathbf{a}'_j\mathbf{a}_m\} = 0$$

$$(iii) E\{\mathbf{d}'_{ij}\mathbf{d}_{km}\} = \lambda E\{\mathbf{a}'_i\mathbf{a}_m\}E\{\mathbf{a}'_j\mathbf{a}_k\}$$

$$\text{if } E\{\mathbf{a}'_i\mathbf{a}_j\} = E\{\mathbf{a}'_k\mathbf{a}_m\} = 0$$

where $\lambda = \sigma_d^2 \sigma_a^{-4}$ and $\tau = u_\delta^2 \sigma_a^{-4}$.

These remarkable results do not depend on $i \geq j$, k, m or $k \geq m$. They are due to the principle of conditional independence and to the rule that probabilities are additive for mutually exclusive events. The first rule appeared in Mäki-Tanila and Kennedy (1986). It is similar to a rule in Crow and Kimura (1970, p 134) based on additive relationship, although rule (i) is more robust under inbreeding. We also have the following more obvious rules:

$$(iv) E\{\mathbf{a}'_i\mathbf{d}_{ij}\} = E\{\mathbf{a}'_i\mathbf{d}_{jj}\} = \eta E\{\mathbf{a}'_i\mathbf{a}_j\}$$

$$(v) E\{\mathbf{a}'_i \mathbf{d}_{jk}\} = 0, \text{ if } E\{\mathbf{a}'_i \mathbf{a}_j\}, E\{\mathbf{a}'_i \mathbf{a}_k\} \text{ or } E\{\mathbf{a}'_j \mathbf{a}_k\} = 0$$

$$(vi) E\{\mathbf{d}'_{ii} \mathbf{d}_{ij}\} = \gamma E\{\mathbf{a}'_i \mathbf{a}_j\}$$

$$(vii) E\{\mathbf{d}'_{ij} \mathbf{d}_{ij}\} = \gamma E\{\mathbf{a}'_i \mathbf{a}_j\} + \alpha [\sigma_d^{-2} - E\{\mathbf{a}'_i \mathbf{a}_j\}]$$

$$(viii) E\{\mathbf{d}'_{ij} \mathbf{d}_{ik}\} = \begin{cases} \alpha E\{\mathbf{a}'_j \mathbf{a}_k\}, & \text{if } E\{\mathbf{a}'_i \mathbf{a}_j\} \text{ or } E\{\mathbf{a}'_i \mathbf{a}_k\} = 0 \\ 0, & \text{if } E\{\mathbf{a}'_j \mathbf{a}_k\} = 0 \end{cases}$$

$$(ix) E\{\mathbf{d}'_{ii} \mathbf{d}_{jk}\} = \begin{cases} \rho E\{\mathbf{a}'_j \mathbf{a}_k\}, & \text{if } E\{\mathbf{a}'_i \mathbf{a}_j\} \text{ or } E\{\mathbf{a}'_i \mathbf{a}_k\} = 0 \\ 0, & \text{if } E\{\mathbf{a}'_j \mathbf{a}_k\} = 0 \end{cases}$$

where $\eta = \sigma_{ad} \sigma_a^{-2}$, $\gamma = \delta^2 \sigma_a^{-2}$, $\alpha = \sigma_d^2 \sigma_a^{-2}$ and $\rho = u_\delta^2 \sigma_a^{-2}$.

Because \mathbf{E} , excluding the first n rows and columns, is at most of the order $r^2/2$ by $r^2/2$, where r is the number of gametes, one might incorrectly conclude that proposed calculations are prohibitive (of the order $r^4/8$) and of no practical value. Recursive algorithms, like the depth-first search in Appendix A, can be surprisingly fast. The value of $r^4/8$ should be regarded as an upper boundary that protects the algorithm from combinatorial explosion – the kind of explosion that might occur when enumerating genetic pathways in a pathological pedigree.

Compressed tables

The genomic table given by Smith and Allaire (1985) can be compressed. We may add together elements in 2 by 2 blocks corresponding to animals, and multiply this matrix by 1/2, to give the numerator relationship matrix. Compression of \mathbf{E} is also feasible, and has already been demonstrated above for a case involving $\tilde{\mathbf{G}}$. In general, \mathbf{E} is compressed by combining columns of \mathbf{G} to create a new matrix \mathbf{C} . To be useful, \mathbf{C} should be smaller than \mathbf{G} and contain pertinent effects.

It is possible to devise recursive methods for evaluating $E\{\mathbf{C}'\mathbf{C}\}$, when \mathbf{C} is not closed under recursion. However, methods become more meticulous. For example, since the vector of additive merits for animals is not closed under gametic recursion, we need to add inbreeding coefficients to the diagonals when calculating the numerator relationship matrix.

Whereas, when compression is defined as the addition of all \mathbf{G} columns, it is possible to do this stochastically, as Harris (1964) has done. For example, Harris, by preferring a zygotic analysis over a gametic analysis, devised a scheme where entities were created by a random sampling of genes from existing genotypes.

Compression is an important area and it needs to be developed further. Some concepts will be illustrated later by an example.

INVERSE STRUCTURES

General rule

Conditions under which \mathbf{E}^{-1} exists are clarified in the next section. For now, let us assume that the inverse exists.

Matrix \mathbf{E} contains second moments and not (co)variances as required by the mixed model equations. However, deleting the first n rows and columns of \mathbf{E}^{-1} gives precisely an inverse matrix of (co)variances. The extended genomic table is characterized by blocks along the diagonal. By inspecting labels attached to vectors in \mathbf{G} , it is seen that they come in groups. For example, the group associated with gamete i is a subsequence of $\mathbf{a}_i, \mathbf{d}_{1i}, \mathbf{d}_{2i}, \dots, \mathbf{d}_{ii}$. Likewise, when considering $\mathbf{E} = E\{\mathbf{G}'\mathbf{G}\}$ we find blocks along the diagonal associated with gametes. Recursions above the diagonal blocks are functions of column indices and not of row indices.

Now consider a submatrix \mathbf{A}_k where

$$\mathbf{E} = \begin{bmatrix} \mathbf{A}_k & \mathbf{A}_k \mathbf{L} \\ \mathbf{L}' \mathbf{A}_k & \mathbf{B} \end{bmatrix}$$

for some \mathbf{L} and \mathbf{A}_k contains the first $k+1$ blocks. The matrix \mathbf{L} is a simple matrix defined by column indices. If $k=1$ note that:

$$\mathbf{A}_1 = \begin{bmatrix} \mathbf{A}_0 & \mathbf{A}_0 \mathbf{L}_0 \\ \mathbf{L}'_0 \mathbf{A}_0 & \mathbf{B}_0 \end{bmatrix}$$

for some \mathbf{L}_0 , where \mathbf{A}_0 corresponds to the base assignments, and \mathbf{B}_0 is the second block.

Let us assume that \mathbf{A}_0^{-1} is given (perhaps without the first n rows and columns) and note that:

$$\mathbf{A}_1^{-1} = \begin{bmatrix} \mathbf{A}_0^{-1} + \mathbf{L}_0(\mathbf{B}_0 - \mathbf{L}'_0 \mathbf{A}_0 \mathbf{L}_0)^{-1} \mathbf{L}'_0 & -\mathbf{L}_0(\mathbf{B}_0 - \mathbf{L}'_0 \mathbf{A}_0 \mathbf{L}_0)^{-1} \\ -(\mathbf{B}_0 - \mathbf{L}'_0 \mathbf{A}_0 \mathbf{L}_0)^{-1} \mathbf{L}'_0 & (\mathbf{B}_0 - \mathbf{L}'_0 \mathbf{A}_0 \mathbf{L}_0)^{-1} \end{bmatrix}$$

With $(\mathbf{B}_0 - \mathbf{L}'_0 \mathbf{A}_0 \mathbf{L}_0)^{-1}$ evaluated, we find that \mathbf{A}_1^{-1} is a simple function of \mathbf{A}_0^{-1} . Given \mathbf{A}_0^{-1} , it is possible to compute \mathbf{A}_2^{-1} , where

$$\mathbf{A}_2 = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_1 \mathbf{L}_1 \\ \mathbf{L}'_1 \mathbf{A}_1 & \mathbf{B}_1 \end{bmatrix}$$

and \mathbf{B}_1 is the third block. In general, given \mathbf{A}_k^{-1} we can evaluate \mathbf{A}_{k+1}^{-1} , where

$$\mathbf{A}_{k+1} = \begin{bmatrix} \mathbf{A}_k & \mathbf{A}_k \mathbf{L}_k \\ \mathbf{L}'_k \mathbf{A}_k & \mathbf{B}_k \end{bmatrix}$$

and \mathbf{B}_k is block $k+2$. The general inversion formula is:

$$\mathbf{A}_{k+1}^{-1} = \begin{bmatrix} \mathbf{A}_k^{-1} + \mathbf{L}_k(\mathbf{B}_k - \mathbf{L}'_k \mathbf{A}_k \mathbf{L}_k)^{-1} \mathbf{L}'_k & -\mathbf{L}_k(\mathbf{B}_k - \mathbf{L}'_k \mathbf{A}_k \mathbf{L}_k)^{-1} \\ -(\mathbf{B}_k - \mathbf{L}'_k \mathbf{A}_k \mathbf{L}_k)^{-1} & (\mathbf{B}_k - \mathbf{L}'_k \mathbf{A}_k \mathbf{L}_k)^{-1} \end{bmatrix} \quad (3)$$

To evaluate \mathbf{E}^{-1} , apply this rule recursively starting with $k=0$.

It is hoped that $\mathbf{B}_k - \mathbf{L}'_k \mathbf{A}_k \mathbf{L}_k$ will be sufficiently small or sufficiently sparse so that its inversion is feasible (eg, Tier and Smith, 1989). For evaluating \mathbf{E}^{-1} , the worst scenario is that the order of $\mathbf{B}_k - \mathbf{L}'_k \mathbf{A}_k \mathbf{L}_k$ is $r+1$. However, this occurrence is unlikely. Note that Henderson's (1975) rule for calculating the inverse numerator relationship matrix is a special case of (3), where $\mathbf{B}_k - \mathbf{L}'_k \mathbf{A}_k \mathbf{L}_k$ is always a scalar.

There are some notable simplifications when \mathbf{E}^{-1} is to be evaluated. First, evaluation of \mathbf{A}_0^{-1} is best done by absorbing the first n rows and then deleting the first n rows and columns. The resulting matrix is some permutation of a block diagonal matrix involving 2 by 2 matrices:

$$\begin{bmatrix} \sigma_a^2 & \sigma_{a\delta} \\ \sigma_{a\delta} & \sigma_\delta^2 \end{bmatrix}$$

and 1 by 1 matrices σ_a^2 and σ_δ^2 . This is a trivial matrix to invert.

Second, $\mathbf{B}_k - \mathbf{L}'_k \mathbf{A}_k \mathbf{L}_k$ has a peculiar structure that can be identified by examining the recursive definition in Section IIIA. If block \mathbf{B}_k corresponds to gamete i which has parent gametes x and y , then \mathbf{L}_k is a matrix that "picks" appropriate terms from \mathbf{A}_k that involve x and y . Moreover, \mathbf{B}_k is also defined by terms that involve x and y . Assume that the column headings for \mathbf{B}_k are:

$$\mathbf{F}_i = \{\mathbf{a}_i, \mathbf{d}_{ij_1}, \mathbf{d}_{ij_2}, \dots, \mathbf{d}_{ij_m}\}$$

It might be that $i = j_m$. Now define the column headings

$$\mathbf{H} = \{\mathbf{H}_x, \mathbf{H}_y\}$$

where $\mathbf{H}_x = (\mathbf{F}_i \mid i = x) = \{\mathbf{a}_x, \mathbf{d}_{xj_1}, \mathbf{d}_{xj_2}, \dots, \mathbf{d}_{xj_m}\}$

and $\mathbf{H}_y = (\mathbf{F}_i \mid i = y) = \{\mathbf{a}_y, \mathbf{d}_{yj_1}, \mathbf{d}_{yj_2}, \dots, \mathbf{d}_{yj_m}\}$

In the definition of \mathbf{H}_x and \mathbf{H}_y , it is understood that $\mathbf{d}_{xj_m} = \mathbf{d}_{xx}$ and $\mathbf{d}_{yj_m} = \mathbf{d}_{yy}$, if $i = j_m$. Select elements from \mathbf{A}_k and build the matrix,

$$\mathbf{M}_k = E\{\mathbf{H}'\mathbf{H}\} = \begin{bmatrix} \mathbf{M}_{xx} & \mathbf{M}_{xy} \\ \mathbf{M}_{yx} & \mathbf{M}_{yy} \end{bmatrix} \quad (4)$$

where $\mathbf{M}_{xx} = E\{\mathbf{H}'_x \mathbf{H}_x\}$, $\mathbf{M}_{xy} = \mathbf{M}'_{yx} = E\{\mathbf{H}'_x \mathbf{H}_y\}$, and $\mathbf{M}_{yy} = E\{\mathbf{H}'_y \mathbf{H}_y\}$.

A direct application of the recursions gives:

$$\mathbf{B}_k = E\{\mathbf{F}'_i \mathbf{F}_i\} = 1/2E\{\mathbf{F}'_i \mathbf{F}_i \mid i = x\} + 1/2E\{\mathbf{F}'_i \mathbf{F}_i \mid i = y\} = 1/2(\mathbf{M}_{xx} + \mathbf{M}_{yy})$$

Futhermore, as \mathbf{L}_k is a matrix that "picks" terms under headings \mathbf{H}_x and \mathbf{H}_y :

$$\mathbf{L}'_k \mathbf{A}_k \mathbf{L}_k = 1/4(\mathbf{I}, \mathbf{I})\mathbf{M}_k(\mathbf{I}, \mathbf{I})' = 1/4(\mathbf{M}_{xx} + \mathbf{M}_{xy} + \mathbf{M}_{yx} + \mathbf{M}_{yy})$$

and thus:

$$\mathbf{B}_k - \mathbf{L}'_k \mathbf{A}_k \mathbf{L}_k = 1/4(\mathbf{M}_{xx} - \mathbf{M}_{xy} - \mathbf{M}_{yx} + \mathbf{M}_{yy}) \quad (5)$$

Equation (5) can also be derived if $\mathbf{B}_k - \mathbf{L}'_k \mathbf{A}_k \mathbf{L}_k$ is recognized as the (co)variance matrix for the segregation effects due to recombination of gametes x and y in the formation of gamete i . The mid-parent values of \mathbf{F}_i are the column vectors of $1/2\mathbf{H}_x + 1/2\mathbf{H}_y$. As the segregation effects, $\mathbf{S} = \mathbf{F}_i - 1/2\mathbf{H}_x - 1/2\mathbf{H}_y$, have a mean of zero, the (co)variance matrix is:

$$\begin{aligned} E\{\mathbf{S}'\mathbf{S}\} &= 1/2E\{\mathbf{S}'\mathbf{S} \mid i = x\} + 1/2E\{\mathbf{S}'\mathbf{S} \mid i = y\} \\ &= 1/2E\{\tilde{\mathbf{S}}'\tilde{\mathbf{S}}\} + 1/2E\{(-\tilde{\mathbf{S}})'(-\tilde{\mathbf{S}})\} = E\{\tilde{\mathbf{S}}'\tilde{\mathbf{S}}\} \end{aligned}$$

where $\tilde{\mathbf{S}} = 1/2\mathbf{H}_x - 1/2\mathbf{H}_y$. Evaluating $E\{\tilde{\mathbf{S}}'\tilde{\mathbf{S}}\}$ gives eqn(5).

Finally, in rule (3) $\mathbf{L}_k(\mathbf{B}_k - \mathbf{L}'_k \mathbf{A}_k \mathbf{L}_k)^{-1} \mathbf{L}'_k$, $-\mathbf{L}_k(\mathbf{B}_k - \mathbf{L}'_k \mathbf{A}_k \mathbf{L}_k)^{-1}$ and $(\mathbf{B}_k - \mathbf{L}'_k \mathbf{A}_k \mathbf{L}_k)^{-1}$ are contributions added under \mathbf{H} by \mathbf{H} headings, \mathbf{H} by \mathbf{F}_i headings, and \mathbf{F}_i by \mathbf{F}_i headings, respectively.

Existence of inverses

When there is no inbreeding, \mathbf{E}^{-1} can be shown to exist. First, we present the following Lemma:

Lemma 1: In the absence of inbreeding, there exists a matrix $\overline{\mathbf{M}}_k$, which is a submatrix of \mathbf{A}_k , and there exists a matrix $\overline{\mathbf{X}}_k$ of full column rank such that:

$$\mathbf{B}_k - \mathbf{L}'_k \mathbf{A}_k \mathbf{L}_k = \overline{\mathbf{X}}'_k \overline{\mathbf{M}}_k \overline{\mathbf{X}}_k$$

where \mathbf{B}_k , \mathbf{L}_k and \mathbf{A}_k are associated with \mathbf{E} .

Proof: Because equ(5) is given, we only prove that such $\overline{\mathbf{M}}_k$ and $\overline{\mathbf{X}}_k$ can be found where $\overline{\mathbf{X}}'_k \overline{\mathbf{M}}_k \overline{\mathbf{X}}_k = 1/4(\mathbf{M}_{xx} - \mathbf{M}_{xy} - \mathbf{M}_{yx} + \mathbf{M}_{yy})$. The matrix \mathbf{M}_k , defined by equ(4), will be a submatrix of \mathbf{A}_k if there are no indices $j_m = i$, $j_v = x$ and $j_w = y$ used in the definition of \mathbf{F}_i . The algorithm presented in Appendix A will not create indices $j_m = i$, $j_v = x$ and $j_w = y$ when there is no inbreeding. For this case, we can take $\overline{\mathbf{M}}_k = \mathbf{M}_k$, and $\overline{\mathbf{X}}'_k = 1/2\{\mathbf{I}, -\mathbf{I}\}$, where the identity matrix \mathbf{I} has order $m + 1$. Matrix $\overline{\mathbf{X}}_k$ has full column rank (Q.E.D).

Theorem 1: If \mathbf{E} is constructed by applying the recursion rules to some finite and non-inbred pedigree, then \mathbf{E}^{-1} exists, provided:

$$\sigma_a^2 > 0, \sigma_d^2 > 0.$$

Proof: The matrix \mathbf{A}_0 is non-singular when the condition of the theorem holds. Now assume that \mathbf{A}_k^{-1} exists, then by the inversion rule (3) \mathbf{A}_{k+1}^{-1} exists if $(\mathbf{B}_k - \mathbf{L}'_k \mathbf{A}_k \mathbf{L}_k)^{-1}$ exists. By the above Lemma, $\mathbf{B}_k - \mathbf{L}'_k \mathbf{A}_k \mathbf{L}_k = \overline{\mathbf{X}}'_k \overline{\mathbf{M}}_k \overline{\mathbf{X}}_k$. Because $\overline{\mathbf{M}}_k$ is a submatrix of \mathbf{A}_k , it is non-singular. Therefore, $(\overline{\mathbf{X}}'_k \overline{\mathbf{M}}_k \overline{\mathbf{X}}_k)^{-1}$ exists because $\overline{\mathbf{X}}_k$ has full column rank. We conclude that the existence of \mathbf{A}_k^{-1} implies the existence of \mathbf{A}_{k+1}^{-1} . As \mathbf{A}_0^{-1} exists, the theorem follows by mathematical induction (Q.E.D).

The reader might think that the concurrence of identical twins would contradict Theorem 1. However, this is not the case, as the theory assumes that gametes are distinct and can be ordered using indices. Thus, for identical twins, the recursive formulae presented earlier are incomplete. This is not a practical problem, as identical gametes can be represented only once in \mathbf{G} .

Henderson (1985) considered a non-inbred population and studied a dominance relationship matrix \mathbf{D} . He used \mathbf{D}^{-1} in many formulae without proof of its existence. However, as \mathbf{D} is a submatrix of \mathbf{E} , the theorem implies that \mathbf{D}^{-1} exists.

When there is inbreeding, the algorithm in Appendix A will produce labels like \mathbf{d}_{ii} , \mathbf{d}_{ix} and \mathbf{d}_{iy} , where $i \neq \theta$ and has parent gametes x and y . In general, \mathbf{E} is singular because of the dependence:

$$\mathbf{d}_{ii} = \mathbf{d}_{ix} + \mathbf{d}_{iy} - \mathbf{d}_{xy} \quad (6)$$

Even with inbreeding, 4 labels like those in eqn(6) need not occur together as headings of \mathbf{E} . However, it is possible to merge the identity (6) with the recursion and remove \mathbf{d}_{ii} for $i \neq \theta$ from the sequence \mathbf{G} to get $\overline{\mathbf{G}}$ – see Appendix A. The vector $\overline{\mathbf{G}}$ is closed under recursion if eqn(6) is used to redefine \mathbf{d}_{ii} for $i \neq \theta$. The matrix of second moments $\overline{\mathbf{E}} = E\{\overline{\mathbf{G}}'\overline{\mathbf{G}}\}$ has no row and column heading \mathbf{d}_{ii} for $i \neq \theta$ and it is non-singular. To prove this, we introduce Lemma 2 for $\overline{\mathbf{E}}$. First, let us imagine that $\mathbf{B}_k - \mathbf{L}'_k \mathbf{A}_k \mathbf{L}_k$ corresponds to a particular absorbed block of $\overline{\mathbf{E}}$.

Lemma 2: There exists a matrix $\overline{\mathbf{M}}_k$, which is a submatrix of \mathbf{A}_k , and there exists a matrix $\overline{\mathbf{X}}_k$ of full-column rank, such that $\mathbf{B}_k - \mathbf{L}'_k \mathbf{A}_k \mathbf{L}_k = \overline{\mathbf{X}}'_k \overline{\mathbf{M}}_k \overline{\mathbf{X}}_k$, where \mathbf{B}_k , \mathbf{L}_k and \mathbf{A}_k are associated with $\overline{\mathbf{E}}$.

Proof: Because the pedigree is inbred, we expect indices $j_v = x$ and $j_w = y$ (x, y parents of i) in the definition of \mathbf{F}_i – otherwise, follow the proof of Lemma 1. By construction, there is no index $j_m = i$. This implies that there will be vectors \mathbf{d}_{xy} , $\mathbf{d}_{xx} = \mathbf{d}_{xx_x} + \mathbf{d}_{xx_y} - \mathbf{d}_{x_x x_y}$ and $\mathbf{d}_{yy} = \mathbf{d}_{yy_x} + \mathbf{d}_{yy_y} - \mathbf{d}_{y_x y_y}$ in \mathbf{H} , where x_x, x_y and y_x, y_y are parent gametes of x and y , respectively. With no loss in generality, switch the vectors \mathbf{a}_i with \mathbf{d}_{ix} , and \mathbf{d}_{iy} with \mathbf{d}_{ij_m} in \mathbf{F}_i . So as to maintain consistency with the definition $\mathbf{H} = \{\mathbf{H}_x = (\mathbf{F}_i | i = x), \mathbf{H}_y = (\mathbf{F}_i | i = y)\}$, columns of \mathbf{H} may be altered by switching:

$$\mathbf{a}_x \leftrightarrow \mathbf{d}_{xx}, \mathbf{d}_{xy} \leftrightarrow \mathbf{d}_{xj_m} \text{ in } \mathbf{H}_x$$

$$\mathbf{a}_y \leftrightarrow \mathbf{d}_{yx}, \mathbf{d}_{yy} \leftrightarrow \mathbf{d}_{yj_m} \text{ in } \mathbf{H}_y$$

These permutations preserve the identity

$$\mathbf{B}_k - \mathbf{L}'_k \mathbf{A}_k \mathbf{L}_k = \mathbf{X}'_k \mathbf{M}_k \mathbf{X}_k$$

where $\mathbf{X}_k = 1/2\{\mathbf{I}, -\mathbf{I}\}$ and $\mathbf{M}_k = E\{\mathbf{H}'\mathbf{H}\}$ if we further stipulate that selected columns of $\overline{\mathbf{G}}$ have also been rearranged. Note that columns and rows $m + 1$ and $m + 2$ in \mathbf{M}_k are redundant, as they are both represented by $\mathbf{d}_{xy} = \mathbf{d}_{yx}$. Therefore, we may delete row and column $m + 2$ to create $\widetilde{\mathbf{M}}_k$ (delete column $m + 2$ in \mathbf{H}) and note that

$$\widetilde{\mathbf{X}}'_k \widetilde{\mathbf{M}}_k \widetilde{\mathbf{X}}_k = \mathbf{X}'_k \mathbf{M}_k \mathbf{X}_k$$

where

$$\widetilde{\mathbf{X}}'_k = 1/2 \begin{pmatrix} 1 & 0 & -1 & 0 & 0 \\ 0 & \mathbf{I} & 0 & -\mathbf{I} & 0 \\ 0 & 0 & 1 & 0 & -1 \end{pmatrix}$$

and \mathbf{I} has order $m - 2$. Clearly, $\widetilde{\mathbf{X}}_k$ is of full-column rank and is a candidate for $\overline{\mathbf{X}}_k$. When x and y are base gametes (ie, x_x, x_y, y_x and y_y are unknown), we are finished as we can take $\overline{\mathbf{M}}_k = \widetilde{\mathbf{M}}_k$, which is a submatrix of \mathbf{A}_k . Unfortunately, when at least one of the zygotes (x_x, x_y) and (y_x, y_y) are known, $\overline{\mathbf{M}}_k$ is not a submatrix of \mathbf{A}_k because of the composite vectors \mathbf{d}_{xx} and/or \mathbf{d}_{yy} . We assume that both (x_x, y_y) and (y_x, y_x) are known in the remainder of the proof (when only 1 zygote is known, the argument can be modified slightly).

It might be that \mathbf{d}_{xx} and \mathbf{d}_{xy} exist already between columns 2 and m in the redefined \mathbf{H} . Likewise, \mathbf{d}_{yx} and \mathbf{d}_{yy} may exist beyond position m . If any of the labels xx , xy , yx and yy do not exist, we introduce them as dominance vectors in \mathbf{H} . Further, we may redefine the first and last columns to represent labels xxy and yxy , respectively. This gives us a modified matrix

$$\tilde{\mathbf{H}} = \{\mathbf{d}_{xxy}, \dots, \mathbf{d}_{xx}, \dots, \mathbf{d}_{xy}, \dots, \mathbf{d}_{yx}, \dots, \mathbf{d}_{yy}, \dots, \mathbf{d}_{yxy}\}.$$

Because $\bar{\mathbf{G}}$ is closed under recursion, all columns of $\tilde{\mathbf{H}}$ are in $\bar{\mathbf{G}}$. Moreover, all columns of $\tilde{\mathbf{H}}$ are unique; $xxy \neq yxy$, because animals cannot mate with themselves. We conclude that the matrix

$$\bar{\mathbf{M}}_k = E\{\tilde{\mathbf{H}}'\tilde{\mathbf{H}}\}$$

is a submatrix of \mathbf{A}_k . Now define the matrix

$$\bar{\mathbf{X}}'_k = 1/2 \begin{pmatrix} -1 & \mathbf{m}' & -1 & 0 & 0 \\ 0 & \mathbf{M} & 0 & -\mathbf{F} & 0 \\ 0 & 0 & 1 & -\mathbf{f}' & 1 \end{pmatrix}$$

where \mathbf{m} and \mathbf{f} are null vectors, except for ones at positions corresponding to xx , xy , yx and yy ; \mathbf{M} and \mathbf{F} are like identity matrices, except that rows corresponding to xx , xy , yx and yy would be deleted if the corresponding columns were introduced into $\tilde{\mathbf{H}}$. Now $\bar{\mathbf{X}}_k$ has full-column rank. Since

$$\bar{\mathbf{X}}'_k \bar{\mathbf{M}}_k \bar{\mathbf{X}}_k = \tilde{\mathbf{X}}'_k \tilde{\mathbf{M}}_k \tilde{\mathbf{X}}_k = \mathbf{X}'_k \mathbf{M}_k \mathbf{X}_k,$$

the Lemma is proved (Q.E.D).

Theorem 2: If $\bar{\mathbf{E}}$ is constructed by applying recursion modified by eqn(6) to some finite pedigree, then $\bar{\mathbf{E}}^{-1}$ exists, provided

$$\sigma_a^2 > 0, \sigma_d^2 > 0, \sigma_a^2 \sigma_d^2 - \sigma_{ad}^2 > 0, \sigma_d^2 > 0$$

Proof: With the first n rows absorbed, \mathbf{A}_0 is non-singular when the condition of the theorem holds. Therefore, \mathbf{A}_0 is non-singular when the first n rows are intact. The remainder of the proof is identical to that of Theorem 1, except that reference is made to Lemma 2, rather than Lemma 1 (Q.E.D).

We should expect singularities when inbreeding coefficients approach unity, as occurs with the numerator relationship matrix. Indeed, the matrix $\tilde{\mathbf{S}} = 1/2\mathbf{H}_x - 1/2\mathbf{H}_y$ becomes a null matrix when gametes x and y are identically equal. As $E\{\tilde{\mathbf{S}}'\tilde{\mathbf{S}}\}$ equals eqn(5), the absorbed block for gamete i is a matrix of zeros and $\bar{\mathbf{E}}$ is singular. Our construction of $\bar{\mathbf{E}}$ assumes that the base population is a random set of unrelated gametes which have been sampled from some infinite population; that is even though the base population is finite, inbreeding in the base does not exist. Under diploidy, inbreeding coefficients cannot be unity with finite pedigrees.

Feasibility

It is beyond the scope of the present paper to demonstrate the calculation of $\bar{\mathbf{E}}^{-1}$ for a real population. However, we have applied the algorithm in Appendix A to a

pedigree borrowed from a beef cattle selection experiment conducted at the Agricultural Research Centre, Trangie, NSW (PF Parnell (1988), personal communication). The pedigree involved 1 122 animals with records and 625 ancestors without records. The average number of generations on the female side was 9 and the maximum was 16. On the male side, the average was 8 generations, with a maximum of 13. There were only 55 base gametes, and the average inbreeding of the population was 0.1 (the maximum was 0.26).

The order of \mathbf{E} and $\bar{\mathbf{E}}$ was about 190 000. Matrix \mathbf{A}_0 was of the order 1 595. However, excluding \mathbf{A}_0 , the maximum block sizes were 321 and 323, respectively. The distribution of block sizes is presented in Table II. Most of the blocks were of the order 2 and 3. Given the absorbed blocks, and ignoring possible singularities, matrices \mathbf{E}^{-1} and $\bar{\mathbf{E}}^{-1}$ can be evaluated. The computations are not trivial, but are feasible.

Table II. Distribution of diagonals blocks of \mathbf{E} and $\bar{\mathbf{E}}$, not including \mathbf{A}_0 .

<i>Order</i>	\mathbf{E}	$\bar{\mathbf{E}}$
2-3	1248	1248
4-7	292	292
8-15	263	263
16-27	224	224
28-50	225	219
51-75	327	325
76-125	500	497
126-175	405	409
176-225	137	141
226-250	15	17
251-350	23	24

There are other approaches to modelling dominance that do not involve \mathbf{E}^{-1} or $\bar{\mathbf{E}}^{-1}$. For example, we might extract from \mathbf{E} only those elements that involve animals or animals with records. If the extracted matrix is put into a matrix $\bar{\mathbf{D}}$, then we could attempt to evaluate $\bar{\mathbf{D}}^{-1}$ using sparse matrix absorption (Tier and Smith, 1989). Then $\bar{\mathbf{D}}^{-1}$ could be used in the mixed-model equations. Alternatively, $\bar{\mathbf{D}}$ could be used directly in the same way that Henderson (1985) used \mathbf{D} . Because \mathbf{E} is an enlarged matrix with unrealized combinations of gametes (the so-called phantom animals), alternatives to \mathbf{E}^{-1} are attractive. However, breeding schemes that utilize dominance variation are bound to require the prediction of dominance effects that correspond to untried gamete pairs (say, among gametes of parents that contribute genes to the next generation). Further research is needed to evaluate all competing methods.

ILLUSTRATION

In this section we demonstrate our methods, using the simple pedigree displayed in Fig. 1. This pedigree involves 4 animals or 8 genomic halves. For simplicity, we assume $\sigma_d^2 = \sigma_\delta^2 = 1$; $u_\delta^2 = \sigma_a^2 = 0$. Given that gametes pairs 12, 35, 64 and 78 represent dominance vectors for animals and are to be included in \mathbf{G} , the algorithm of Appendix A creates 14 additional pairs – ignoring first array \mathbf{I} and additive vectors. The matrix \mathbf{G} is given as row and column labels of Table III. Second moments of Table III are derived by applying the recursion of formulae. For example, the element $E\{\mathbf{d}'_{21}\mathbf{d}_{51}\}$ was calculated as $1/2E\{\mathbf{d}'_{21}\mathbf{d}_{11}\} + 1/2E\{\mathbf{d}'_{21}\mathbf{d}_{21}\} = 0 + 1/2 = 1/2$.

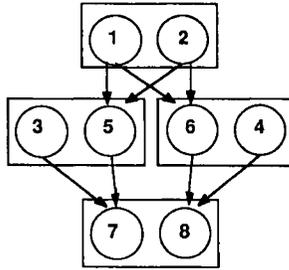


Fig 1. Pedigree depicting genetic relationships among genomic halves of 4 animals.

To evaluate \mathbf{E}^{-1} requires the determination of $(\mathbf{A}_i - \mathbf{L}'_i\mathbf{B}_i\mathbf{L}_i)^{-1}$ for $i = 0, 1, 2, 3$. The absorbed blocks can be evaluated recursively, but for now, the reader may obtain these by applying Gaussian elimination directly to Table III. The inverse blocks are:

$$\begin{pmatrix} 1/2 & -1/4 & 0 & 0 \\ -1/4 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1/2 \end{pmatrix}^{-1} = \begin{pmatrix} 8/3 & 4/3 & 0 & 0 \\ 4/3 & 8/3 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}, \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, 2$$

The matrices \mathbf{L}'_i for $i = 0, 1, 2, 3$, are displayed below the diagonal in Table IV. In accordance with formulae (3), the elements of \mathbf{E}^{-1} are given above the diagonal.

We can now compress \mathbf{E} by defining $\mathbf{d}_b = \mathbf{d}_{11} + \mathbf{d}_{22}$ (with labels, this is notated as $b: 11 + 22$), $\mathbf{d}_c = \mathbf{d}_{31} + \mathbf{d}_{32}$, $\mathbf{d}_d = \mathbf{d}_{41} + \mathbf{d}_{42}$, $\mathbf{d}_f = \mathbf{d}_{51} + \mathbf{d}_{52}$, $\mathbf{d}_l = \mathbf{d}_{74} + \mathbf{d}_{76}$. These definitions, as well as other symbolic definitions, are given as row labels for the compressed matrix of Table V. As an exercise, the compressed table can be evaluated by inspecting Table III and adding together appropriate elements. For example, $E\{\mathbf{d}'_f\mathbf{d}_l\}$ can be computed by adding together 4 elements of Table III:

$$E\{\mathbf{d}'_{51}\mathbf{d}_{74}\} + E\{\mathbf{d}'_{51}\mathbf{d}_{76}\} + E\{\mathbf{d}'_{52}\mathbf{d}_{74}\} + E\{\mathbf{d}'_{52}\mathbf{d}_{76}\} = 0 + 1/4 + 0 + 1/4 = 1/2$$

Table III. The matrix **E** for the pedigree of Fig. 1. Headings ij correspond to effect d_{ij} . Genetic parameters assumed to be $\sigma_d^2 = \delta^2 = 1$; $w_d^2 = \sigma_a^2 = 0$.

	11	21	22	31	32	41	42	43	51	52	53	54	63	64	65	74	76	87
11	1	0	0	0	0	0	0	0	1/2	0	0	0	0	0	1/4	0	1/8	1/16
21	0	1	0	0	0	0	0	0	1/2	1/2	0	0	0	0	1/2	0	1/4	1/8
22	0	0	1	0	0	0	0	0	0	1/2	0	0	0	0	1/4	0	1/8	1/16
31	0	0	0	1	0	0	0	0	0	0	1/2	0	1/2	0	0	0	1/4	1/8
32	0	0	0	0	1	0	0	0	0	0	1/2	0	1/2	0	0	0	1/4	1/8
41	0	0	0	0	0	1	0	0	0	0	0	1/2	0	1/2	0	1/4	0	1/8
42	0	0	0	0	0	0	1	0	0	0	0	1/2	0	1/2	0	1/4	0	1/8
43	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1/2	0	1/4
51	1/2	1/2	0	0	0	0	0	0	1	0	0	0	0	0	1/2	0	1/4	1/8
52	0	1/2	1/2	0	0	0	0	0	0	1	0	0	0	0	1/2	0	1/4	1/8
53	0	0	0	1/2	1/2	0	0	0	0	0	1	0	1/2	0	0	0	1/4	1/8
54	0	0	0	0	0	1/2	1/2	0	0	0	0	1	0	1/2	0	1/2	0	1/4
63	0	0	0	1/2	1/2	0	0	0	0	0	1/2	0	1	0	0	0	1/2	1/4
64	0	0	0	0	0	1/2	1/2	0	0	0	0	1/2	0	0	0	1/4	0	1/8
65	1/4	1/2	1/4	0	0	0	0	0	1/2	1/2	0	0	0	0	1	0	1/2	1/4
74	0	0	0	0	0	1/4	1/4	1/2	0	0	0	1/2	0	1/4	0	1	0	1/2
76	1/8	1/4	1/8	1/4	1/4	0	0	0	1/4	1/4	1/4	0	1/2	0	1/2	0	1	1/2
87	1/16	1/8	1/16	1/8	1/8	1/8	1/8	1/4	1/8	1/8	1/8	1/4	1/4	1/8	1/4	1/2	1/2	1

Table IV. Matrices L'_i for $i = 0, 1, 2, 3$ (below diagonal) and E^{-1} (above diagonal)

	11	21	22	31	32	41	42	43	51	52	53	54	63	64	65	74	76	87
11	5/3	1	1/3	0	0	0	0	0	-4/3	-2/3	0	0	0	0	0	0	0	0
21		3	1	0	0	0	0	0	-2	-2	0	0	0	0	0	0	0	0
22			5/3	0	0	0	0	0	-2/3	-4/3	0	0	0	0	0	0	0	0
31				2	1	0	0	0	0	0	-1	0	-1	0	0	0	0	0
32					2	0	0	0	0	0	-1	0	-1	0	0	0	0	0
41						2	1	0	0	0	0	-1	0	-1	0	0	0	0
42							2	0	0	0	0	-1	0	-1	0	0	0	0
43								3/2	0	0	0	1/2	0	0	0	-1	0	0
51	1/2	1/2	0	0	0	0	0	0	19/6	11/6	0	0	0	0	-1	0	0	0
52	0	1/2	1/2	0	0	0	0	0	0	19/6	0	0	0	0	-1	0	0	0
53	0	0	0	1/2	1/2	0	0	0	0	0	2	0	0	0	0	0	0	0
54	0	0	0	0	0	1/2	1/2	0	0	0	0	5/2	0	0	0	-1	0	0
63	0	0	0	1/2	1/2	0	0	0	0	0	0	0	5/2	0	1/2	0	-1	0
64	0	0	0	0	0	1/2	1/2	0	0	0	0	0	0	2	0	0	0	0
65	0	0	0	0	0	0	0	0	1/2	1/2	0	0	0	0	5/2	0	-1	0
74	0	0	0	0	0	0	0	1/2	0	0	0	1/2	0	0	0	5/2	1/2	-1
76	0	0	0	0	0	0	0	0	0	0	0	0	1/2	0	1/2	5/2	5/2	-1
87	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1/2	1/2	2

The real challenge is to find simplifying recursions that allow evaluation of compressed tables. In order to apply inversion formulae (3), recursions to the right of blocks along the diagonal should be maintained exclusively as a function of column index, as in the present example. These recursions are given implicitly below the diagonal in Table VI as matrices \mathbf{L}_i for $i = 0, 1, 2, 3$. For example, the element $E\{\mathbf{d}'_f \mathbf{d}_l\}$ can also be obtained by adding

$$1/2E\{\mathbf{d}'_f \mathbf{d}_e\} + 1/2E\{\mathbf{d}'_f \mathbf{d}_k\} + 1/2E\{\mathbf{d}'_f \mathbf{d}_i\} + \{\mathbf{d}'_f \mathbf{d}_k\} = 0 + 0 + 0 + 1/2 = 1/2$$

The blocks $(\mathbf{A}_i - \mathbf{L}'_i \mathbf{B}_i \mathbf{L}_i)^{-1}$ for $i = 0, 1, 2, 3$ are:

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}, 1, 2$$

Inverse elements of the compressed matrix are given above the diagonal in Table VI.

DISCUSSION

Although we have not emphasized the evaluation of various identity coefficients, gametic recursion of such coefficients is feasible. Moreover, gametic recursions, like those used to define \mathbf{E} , are very easy to understand and may provide an approach useful in teaching, whilst there may be more complicated (but perhaps numerically more efficient) alternatives. But gametic recursion need not be numerically inefficient and is certainly not subject to combinatorial explosion. Depth-first searches, like the algorithm of Appendix A, can be used to implement gametic recursion in an efficient way. For example, to evaluate inbreeding coefficients via gametic recursion, one would first conceive of a symmetric table with $r^2/2$ elements. However, $r^2/2$ operations are not required to evaluate inbreeding coefficients via gametic recursion. All that is needed is to identify labels ij of the dominance vectors \mathbf{d}_{ij} in \mathbf{G} . By moving down a list of such labels, it is possible to evaluate the coefficients using one work vector. We need only modify this scenario for general identity coefficients.

The simplest explanation of the two commonly found and complementary phenomena (heterosis and inbreeding depression) is that there is dominance of alleles at many loci. Therefore, dominance should be regarded as an essential feature of genetic models for loci affecting quantitative traits. Characteristics closely connected with fitness, such as reproduction, would mostly have only non-additive variation left, as the additive has been exhausted (Robertson, 1955). There are also suggestions that dominance of alleles that maintain normal enzyme activity, is a universal biochemical property (Kacser and Burns, 1981).

It is clear that additivity, dominance and inbreeding can be modelled by applying mixed-model methodology. The ramifications of such a development are far reaching. We list possible applications:

Table V. The compressed matrix. Definitions are given on the left column.

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>	<i>i</i>	<i>j</i>	<i>k</i>	<i>l</i>	<i>m</i>
<i>a</i> : 12	1	0	0	0	0	1	0	0	0	0	1/2	1/4	1/8
<i>b</i> : 11 + 22	0	2	0	0	0	1	0	0	0	0	1/2	1/4	1/8
<i>c</i> : 31 + 32	0	0	2	0	0	0	1	0	1	0	0	1/2	1/4
<i>d</i> : 41 + 42	0	0	0	2	0	0	0	1	0	1	0	1/2	1/4
<i>e</i> : 43	0	0	0	0	1	0	0	0	0	0	0	1/2	1/4
<i>f</i> : 51 + 52	1	1	0	0	0	2	0	0	0	0	1	1/2	1/4
<i>g</i> : 53	0	0	1	0	0	0	1	0	1/2	0	0	1/4	1/8
<i>h</i> : 54	0	0	0	1	0	0	0	1	0	1/2	0	1/2	1/4
<i>i</i> : 63	0	0	1	0	0	0	1/2	0	1	0	0	1/2	1/4
<i>j</i> : 64	0	0	0	1	0	0	0	1/2	0	1	0	1/4	1/8
<i>k</i> : 65	1/2	1/2	0	0	0	1	0	0	0	0	1	1/2	1/4
<i>l</i> : 74 + 76	1/4	1/4	1/2	1/2	1/2	1/2	1/4	1/2	1/2	1/4	1/2	2	1
<i>m</i> : 87	1/8	1/8	1/4	1/4	1/4	1/4	1/8	1/4	1/4	1/8	1/4	1	1

Table VI. Matrices L'_1 for $i = 0, 1, 2, 3$ (below diagonal), and C^{-1} (above diagonal), where C is given by Table 6.

	a	b	c	d	e	f	g	h	i	j	k	l	m
$a : 12$	3	1	0	0	0	-2	0	0	0	0	0	0	0
$b : 11 + 22$		1	0	0	0	-1	0	0	0	0	0	0	0
$c : 31 + 32$			$3/2$	0	0	0	-1	0	-1	0	0	0	0
$d : 41 + 42$				$3/2$	0	0	0	-1	0	-1	0	0	0
$e : 43$					$5/4$	0	0	$1/4$	$1/4$	0	$1/4$	- $1/2$	0
$f : 51 + 52$	1	$1/2$	0	0	0	$5/2$	0	0	0	0	-1	0	0
$g : 53$	0	0	$1/2$	0	0	0	2	0	0	0	0	0	0
$h : 54$	0	0	0	$1/2$	0	0	0	$9/4$	$1/4$	0	$1/4$	- $1/2$	0
$i : 63$	0	0	$1/2$	0	0	0	0	0	$9/4$	0	$1/4$	- $1/2$	0
$j : 64$	0	0	0	$1/2$	0	0	0	0	0	2	0	0	0
$k : 65$	0	0	0	0	0	$1/2$	0	0	0	0	$9/4$	- $1/2$	0
$l : 74 + 76$	0	0	0	0	$1/2$	0	0	$1/2$	$1/2$	0	$1/2$	$3/2$	-1
$m : 87$	0	0	0	0	0	0	0	0	0	0	0	$1/2$	2

(i) Determination of optimum and dynamic mating structures which capitalize on additive and dominance variance while providing for inbreeding. Some of these breeding strategies can be studied by the use of moment-generating matrices for regular mating systems. Cockerham (1971) has given an example of such a transition matrix for full-sib mating. Possible application areas are:

(a) mate selection (Jansen and Wilton, 1985; Smith and Allaire, 1985);

(b) group selection (Jansen, 1985; Smith and Hammond, 1987) when the selection of a random mating gene pool is created by a finite number of parents. The objective of group selection is to improve both additive merit and the average specific combining ability of genes in the pool;

(c) crossbreeding plans to utilize between breed additive and heterotic effects (Kingham, 1987);

(d) selection of clones and animals created by futuristic techniques.

(ii) To allow dominance variance to be partitioned, and thus, remove some of the confounding that would otherwise corrupt statistical analyses.

(iii) To improve our understanding of traits which are likely to show considerable amounts of non-additive genetic variation. This understanding can be advanced by simulation studies, *eg*, concerning selection bias and finite loci. Other studies may involve simulation of infinitesimal models with dominance *via* sampling from normal distributions.

The theory we present here is still very underdeveloped. New methods are needed for the estimation of genetic parameters, as past approaches have met with very limited success (*eg*, Gallais, 1977). It is our belief that an extension of the derivative-free algorithm of Graser *et al.* (1987) is possible, and consequently, genetic parameters could be estimated by restricted maximum likelihood. Research is needed in the area of computing strategies (*eg*, compression, inversion). Further, models which allow different parameters for different populations would be useful in crossbreeding studies. A multivariate extension of our work would also be desirable.

The forgotten papers of Harris (1964) and Gillois (1964) have been resurrected. While the methods are arguably complicated, they are, however, feasible. Whereas, classical quantitative genetics is unable to fully utilize information on dominance and inbreeding in predictions, the appropriate mixed model under a wide range of assumptions, including selection and environmental noise, does that.

ACKNOWLEDGEMENTS

The authors thank Keith Hammond for his encouragement. This work was supported by the Australian Meat and Livestock Research and Development Corporation, the Australian Pig Research Council, and the Finnish Agricultural Research Council.

REFERENCES

- Bulmer MG (1971) The effect of selection on genetic variability. *Am Nat* 105, 201-211
- Chevalet CL, Gillois M (1977) Estimation of genotypic variance components with dominance in small consanguineous populations, *Proc Int Conf on Quantitative Genetics*. (Pollak E, Kempthorne O, Bailey TB Jr, eds). Iowa State University Press, Ames 271-296
- Cockerham CC (1971) Higher order probability functions of identity of alleles by descent. *Genetics* 69, 235-246
- Crow JF, Kimura M (1970) *An Introduction to Population Genetics Theory*. Harper and Row, New York
- Fisher RA (1918) The correlation between relatives on the supposition of Mendelian inheritance. *Trans R Soc Edinburgh* 52, 399-433
- Gallais A (1977) An experimental check of quantitative genetics on an autotetraploid plant, *Medicago sativa* L, with special reference to the identity by descent relationship. *Proc Int Conf Quantitative Genetics*, (Pollak E, Kempthorne O, Bailey TB Jr, eds) Iowa State University Press, Ames, pp 519-540
- Gianola D, Fernando RL (1986) Bayesian methods in animal breeding theory. *J Anim Sci* 63, 217-244
- Gillois M (1964) La relation d'identité en génétique. Thèse Fac Sci Paris
- Graser HU, Smith SP, Tier B (1987) A derivative-free approach for estimating variance components in animal models by restricted maximum likelihood. *J Anim Sci* 64, 1362-1370
- Harris DL (1964) Genotypic covariances between inbred relatives. *Genetics* 50, 1319-1348
- Henderson CR (1973) Sire evaluation and genetic trends. *Proc. of the Animal Breeding and Genetic Symp. in Honor of Dr Jay L Lush, ASAS-ADSA, Champaign*, 10-41
- Henderson CR (1975) Rapid method for computing the inverse of the relationship matrix. *J Dairy Sci* 58, 1727-1730
- Henderson CR (1985) Best linear unbiased prediction of non-additive merits in non-inbred populations. *J Anim Sci* 60, 111-117
- Hill WG (1969) The rate of selection advance for non-additive loci. *Genet Res* 13, 165-173
- Hudson GFS, Van Vleck LD (1984) Effect of inbreeding on milk and fat production, stayability, and calving interval of registered Ayrshire cattle in the Northeastern United States. *J Dairy Sci* 67, 171-179
- Jansen GB (1985) Selection and mating strategies to improve quadratic merit. PhD Thesis, The University of Guelph
- Jansen GB, Wilton JW (1985) Selecting mating pairs with linear programming techniques. *J Dairy Sci* 68, 1302-1305

- Kacser H, Burns JA (1981) The molecular basis of dominance. *Genetics* 97, 639-666
- Kinghorn B (1987) On computing strategies for mate allocation. *Z Tierzüchtg Züchtgsbiol* 104, 12-22
- Mäki-Tanila A, Kennedy BW (1986) Mixed model methodology under genetic models with a small number of additive and non-additive loci. *Proc 3rd World Congr Genetics Applied Livestock Prod XII*, 443-448
- Nadot R, Vaysseix G (1973) Apparement et identité. Algorithme de calcul des coefficients d'identité. *Biometrics* 29, 347-359
- Robertson A (1955) Selection in animals: synthesis *Cold Spring Harbor Symp. Quant Biol* 20, 225-229
- Robertson A (1967) *The Nature of Quantitative Variation. Heritage from Mendel* (Brink RA, Styles ED, eds) University of Wisconsin Press, Madison, 265-280
- Robertson A, Hill WG (1983) Population and quantitative genetics of many linked loci in finite populations. *Proc. R. Soc. London B* 219, 253-264
- Smith SP (1984) Dominance relationship matrix and inverse for an inbred population. Unpublished mimeo, Dept of Dairy Sci, Ohio State University, Columbus
- Smith SP, Allaire FR (1985) Efficient selection rules to increase non-linear merit: application in mate selection. *Genet Sel Evol* 17, 387-406
- Smith SP, Hammond K (1987) Portfolio theory, utility theory and mate selection. *Genet Sel Evol* 19, 321-336
- Tier B (1990) Computing inbreeding coefficients quickly. *Genet Sel Evol* (in press)
- Tier B, Smith SP (1990) Use of sparse matrix absorption in animal breeding. *Genet Sel Evol* 21, 457-466

APPENDIX A

Depth-first search for extracting gamete pairs

The following account deals with the extraction of labels ij of the dominance vectors \mathbf{d}_{ij} of \mathbf{G} . To identify the additive vectors, simply require both \mathbf{a}_i and \mathbf{a}_j to be included in the list if \mathbf{d}_{ij} is. All gamete indices are assumed to have been transformed, such that base gametes come first and ancestors precede their descendants.

1. Definitions

(a) FLAG(I,J) is an indicator variable. Upon completion of the search, FLAG(I,J) is true when pair IJ has been selected, and false otherwise. That is, FLAG(I,J) provides the gametic pair associated with the dominance vector \mathbf{d}_{ij} , using the convention that $I=i$ and $J=j$. FLAG can be a very large array because each of the IJ entries need only occupy one bit of computer memory.

(b) PEDIGR(1,I), PEDIGR(2,I) are parent gametes of gamete I.

(c) WORK(1,I), WORK(2,I) is a work space indexed by I.

2. Initializations

- (a) the pedigree file is read and PEDIGR is built;
- (b) all values of FLAG are set to false;
- (c) WORK is initialized by reading a list of LEN animals (say, those animals that are associated with phenotypic measurement) and adding gamete pairs to WORK. The index for the maternal gamete of animal I is placed in WORK(1,I). The index for the paternal gamete is placed in WORK(2,I). There are LEN initializations of WORK(1,I), WORK(2,I), $I = 1, 2, \dots, \text{LEN}$.

3. Depth-first search

- (a) If $\text{LEN} = 0$, then stop as the search is completed;
- (b) Set $I = \text{MAX}(\text{WORK}(1,\text{LEN}), \text{WORK}(2,\text{LEN}))$
and $J = \text{MIN}(\text{WORK}(1,\text{LEN}), \text{WORK}(2,\text{LEN}))$. The gamete pair represented by I and J (formally stored in position LEN of WORK) is now set for evaluation;
- (c) Set $\text{LEN} = \text{LEN} - 1$. As the gamete pair IJ is set for evaluation, the list of gamete pairs left for consideration is shortened by one.
- (d) If $\text{FLAG}(I,J) = \text{true}$, go to (a). In this case, the gamete pair IJ has already been selected and no further action is necessary;
- (e) Set $\text{FLAG}(I,J) = \text{true}$. This action causes future evaluations of the gamete pair IJ to stop at 3(d). At this point, indices I and J can be written to an exterior file, thus avoiding a full scan of FLAG after the depth-first search;
- (f) If $I \leq$ the largest base index, go to (a). In this case, the pair IJ belongs to the base population and the search for ancestors of I and J ($J < I$) can stop. The algorithm returns to 3(a) to process the next pair in WORK. If I is not a base gamete, the algorithm proceeds with steps 3(c) and 3(b). These steps add ancestral gamete pairs to WORK for future evaluation;
- (g) If $I=J$, then set
 - $\text{LEN} = \text{LEN} + 1$
 - $\text{WORK}(1,\text{LEN}) = \text{PEDIGR}(1,I)$
 - $\text{WORK}(2,\text{LEN}) = \text{PEDIGR}(1,I)$
 - $\text{LEN} = \text{LEN} + 1$
 - $\text{WORK}(1,\text{LEN}) = \text{PEDIGR}(2,I)$
 - $\text{WORK}(2,\text{LEN}) = \text{PEDIGR}(2,I)$
- (h) If $I > J$, then set
 - $\text{LEN} = \text{LEN} + 1$
 - $\text{WORK}(1,\text{LEN}) = \text{PEDIGR}(1,I)$
 - $\text{WORK}(2,\text{LEN}) = J$
 - $\text{LEN} = \text{LEN} + 1$
 - $\text{WORK}(1,\text{LEN}) = \text{PEDIGR}(2,I)$
 - $\text{WORK}(2,\text{LEN}) = J$
- (i) go to (a)

4. Modifications

To enumerate vectors \mathbf{d}_{ij} of $\overline{\mathbf{G}}$, it is necessary to prevent the writing of I and J to an exterior file at step 3(e), if $I = J$ and $I >$ the largest base index, and replace 3(g) with

If $I = J$, then set

```
LEN = LEN + 1
WORK(1,LEN) = PEDIGR(1,I)
WORK(2,LEN) = J
LEN = LEN + 1
WORK(1,LEN) = PEDIGR(2,I)
WORK(2,LEN) = J
LEN = LEN + 1
WORK(1,LEN) = PEDIGR(1,I)
WORK(2,LEN) = PEDIGR(2,I)
```