Two-locus theory in quantitative genetics

B. S. Weir and *C. Clark Cockerham*
DEPARTMENT OF STATISTICS
NORTH CAROLINA STATE UNIVERSITY, RALEIGH, NC 27607

INTRODUCTION

The consideration of the effects of two or more loci on quantitative traits goes back at least as far as Fisher's 1918 paper. Fisher introduced the concept of epistasis whereby genes at different loci act in a dependent fashion. Apart from epistasis there are the complications of linkage and linkage disequilibrium for pairs of loci, and a comprehensive theory should also take account of inbreeding. It is only recently (Gallais, 1974) that all these factors have been considered simultaneously, although not in full generality.

Between 1918 and 1974 there were a series of papers dealing with some of these factors and evaluating some or all of the mean and variance of a quantitative trait or the covariance (or correlation) between the genotypic values of relatives. Reviews of the literature up to that time have been provided by Kempthorne (1955, 1957) and Gallais (1974). Some of the key references are marked with asterisks in the Bibliography.

Paper No. 5120 of the Journal Series of the North Carolina Agricultural Experiment Station, Raleigh, North Carolina. This investigation was supported in part by NIH Research Grant No. GM 11546 from the National Institute of General Medical Sciences.

Proceedings of the International Conference on QUANTITATIVE GENETICS
© Iowa State University Press, Ames, Iowa 1977

One of the main difficulties with two-locus work in quantitative genetics is that expressions quickly become unwieldy. A systematic procedure is needed to avoid missing terms and such procedures are heavily dependent on the set of parameters chosen to reflect the effects of inbreeding, linkage and linkage disequilibrium. Parameters can be chosen to allow means, variances and covariances to be expressed as linear combinations of additive, dominance and epistatic effects in a reference population.

For genes at one locus, a complete set of parameters was given by Gillois (1964) and Harris (1964), and methods for calculating an equivalent set were given by Cockerham (1971). Denniston (1967) and Gallais (1970) introduced a set of two-locus parameters which allowed covariances to be found, although those of Denniston were restricted to non-inbred relatives. Both of these sets of measures were for eight genes, four at each of two loci.

When we introduced our four-gene descent measures (Cockerham and Weir, 1973) we remarked that they could be used to provide means and variances for a two-locus quantitative model of gene effects. Such expressions are derived here, and the problem of determining the covariance of relatives is discussed.

While the expressions, especially that for the mean, do offer some insight into the effects of the various dependencies, they also illustrate our belief that exact treatments of completely general situations are unlikely to be of great use in quantitative genetics. An adequate approximate treatment is needed.

ONE-LOCUS THEORY

A brief review of one-locus theory will be of considerable help in introducing our approach for the two-locus case. For completely independent genes of course, one-locus theory can be extended directly to multi-locus situations.

For locus A the genotypic value of individuals with genotype A_iA_k will be written as G_k^i and is the sum of a mean, additive effects for each of the two A genes and a dominance effect for the interaction of these genes.

$$G^i_k = \mu_0 + a_i + a_k + d^a_{ik} \quad , \quad G^i_k = G^k_i \quad .$$

The superscript a on the dominance term emphasizes that the A locus is being considered.

The mean and effects may be defined for any population, although with difficulty in some cases. We use an infinite random-mating reference population which has alleles A_i with frequency p_i. Then the usual least squares values for the mean and effects are

$$\mu_0 = \sum_{i,k} p_i p_k G^i_k = G^{\cdot}_{\cdot} \quad , \quad a_i = \sum_k p_k G^i_k - G^{\cdot}_{\cdot} = G^i_{\cdot} - G^{\cdot}_{\cdot}$$

$$d^a_{ik} = G^i_k - G^i_{\cdot} - G^{\cdot}_k + G^{\cdot}_{\cdot}$$

so that $$\sum_i p_i a_i = 0, \quad \sum_i p_i d^a_{ik} = 0, \quad i \neq k \; . \qquad (1)$$

Means

In any generation after the initial one, the mean genotypic value is obtained as the expectation of the value of any member of that generation. Expectation, written as $\mathcal{E}$, refers to the averaging over all possible replicate populations of the same size and mating system. We write the mean as μ_F to indicate that such populations may be inbred. If a random member of the population has genotype A_iA_k, then

$$\mu_F = \mathcal{E}\, G^i_k = \mathcal{E}(\mu_0 + a_i + a_k + d^a_{ik}) = \mu_0 + \mathcal{E}\, d^a_{ik} \; .$$

The additive effects have zero expectation from relation (1), but for the dominance effects we need to take account of the fact that genes A_i and A_k may be identical by descent. With probability $1 - F_1$ the two genes are not identical by descent and so have descended from genes on distinct initial gametes. The probability that a random pair of initial gametes carry genes A_i, A_k is taken to be $\theta^i_k = p_ip_k$. With probability F_1 [illegible] carries genes which are identical by descent, a [illegible]cended from a gene on one initial gamete. Th[illegible]ne is A_i with probability $\theta^i_{\cdot} = p_i$ so that

$$\mathcal{E}\, d^a_{ik} = (1 - F_1) \sum_{i,k} p_i p_k \, d^a_{ik} + F_1 \sum_i p_i d^a_{ii} \; .$$

From (1) the first sum here is zero, and we recover the well-known result

$$\mu_F = \mu_0 + F_1 \sum_i p_i d^a_{ii} \; . \tag{2}$$

If the mean of a completely inbred population, $F_1 = 1$, is written as μ_1 then (2) can be expressed as

$$\mu_F = \mu_0 + F_1 \; (\mu_1 - \mu_0) \; ,$$

as given by Wright (1951). With inbreeding then, the change in the population mean is a linear function of the inbreeding coefficient and involves dominance effects.

By using the inbreeding coefficient we were able to express a property, the mean, of a population in terms of properties of an initial population. The inbreeding coefficient summarizes the progress of the population since the formation of the initial population, while the frequencies θ^i_k summarize the structure of this initial population. We assumed here that $\theta^i_k = p_i p_k$, with the p_i the same as in the reference population for which effects are defined. This assumption is not necessary and we have previously (Cockerham and Weir, 1973) discussed other situations. Specific sets of initial gametes or even specific initial individuals may be accommodated. For specific gametes θ^i_k is modified, and for specific individuals $1 - F_1$ is extended to distinguish the cases of two gametes between or within individuals.

Another aspect of initial populations deserves mention. We are often interested in the cross population $\mathcal{F}_1$ from parental populations Π_I and Π_{II}. This situation has been considered by several authors previously. Stuber and Cockerham (1966) used the terms "dually defined" and "uniquely defined" to refer to the two cases when gene effects are defined for genes from Π_1 and Π_2 separately or for the combined set respectively.

In the dual case G^i_k now refers to individuals receiving a gamete carrying A_i from Π_I and a gamete carrying a_k from Π_{II}, and we write

$$G^i_k = \mu_C + a^i + a_k + d^{ai}_k \quad .$$

We must also distinguish between p^i and p_k, the frequencies of A_i and A_k in populations Π_I and Π_{II} respectively. If $\mathcal{F}_1$ is formed by the random union of gametes from the two parental populations, then the least squares values for the mean and effects are

$$\mu_C = \sum_{i,k} p^i p_k G^i_k = G^{\cdot}_{\cdot} \quad , \quad a^i = \sum_k p_k G^i_k - G^{\cdot}_{\cdot} = G^i_{\cdot} - G^{\cdot}_{\cdot}$$

$$a_k = \sum_i p^i G^i_k - G^{\cdot}_{\cdot} = G^{\cdot}_k - G^{\cdot}_{\cdot} \quad , \quad d^{ai}_k = G^i_k - G^i_{\cdot} - G^{\cdot}_k + G^{\cdot}_{\cdot} \quad .$$

The possible relations between μ_C, the mean of $\mathcal{F}_1$, and the means of Π_I, Π_{II} have been discussed by Kempthorne (1957). The dual formulation shows that even though the same alleles may be contained in both populations, they have different effects unless the two parental populations are identical.

For the $\mathcal{F}_2$ and subsequent generations, when inbreeding is a possibility, we employ the unique definitions. Care must be taken of course to account for any relatedness of populations Π_I and Π_{II}.

Variances

In formulating variances we need the expectations of squares of effects. It is helpful to write out the linear model for G^i_k as headings in a two-way table, and identify the classes with different expectations.

	μ_0	a_i	a_k	d^a_{ik}
μ_0	1	2	2	3
a_i	2	4	5	6
a_k	2	5	4	6
d^a_{ik}	3	6	6	7

There are seven classes, the first of which is just μ_0^2 and the second of which has zero expectation . The expectations of four of the remaining five classes depend on the identity by descent of the two genes A_i, A_k. These five expectations are now shown.

Class	Expectations
3	$\mu_0 F_1 \sum_i p_i d^a_{ii}$
4	$\sum_i p_i a_i^2$
5	$F_1 \sum_i p_i a_i^2 + (1 - F_1) \sum_{i,k} p_i p_k a_i a_k$
6	$F_1 \sum_i p_i a_i d^a_{ii} + (1 - F_1) \sum_{i,k} p_i p_k a_i d^a_{ik}$
7	$F_1 \sum_i p_i (d^a_{ii})^2 + (1 - F_1) \sum_{i,k} p_i p_k (d^a_{ik})^2$

Recognizing that the second sums for classes 5 and 6 are zero, and defining the familiar additive and dominance variance components as

$$\sigma_A^2 = 2\sigma_a^2 = 2\sum_i p_i a_i^2 \ , \quad \sigma_D^2 = \sum_{i,k} p_i p_k (d^a_{ik})^2$$

we can now find $\mathcal{E}(G^i_k)^2$ by collecting terms over classes 1-7. If we then subtract μ_F^2 from (2) we have the variance:

$$\sigma^2_{G_F} = (1 + F_1)\,\sigma_A^2 + (1 - F_1)\,\sigma_D^2 + 4F_1 \sum_i p_i a_i d^a_{ii}$$

$$+ F_1 \sum_i p_i (d^a_{ii})^2 - (F_1)^2 \Big(\sum_i p_i d^a_{ii}\Big)^2 . \quad (3)$$

Setting $F_1 = 1$ or 0 gives the variances $\sigma^2_{G_1}$ or $\sigma^2_{G_0}$ for completely inbred or non-inbred populations respectively and (3) can be rewritten in the form given by Wright (1951):

$$\sigma^2_{G_F} = (1 - F_1)\,\sigma^2_{G_0} + F_1\,\sigma^2_{G_1} + F_1(1 - F_1)(\mu_1 - \mu_0)^2 \ .$$

The form of equation (3) is helpful in showing how both additive and dominance effects are involved in the way inbreeding changes population variances. Without dominance, complete inbreeding will increase genetic variance to twice the non-inbred value, but with dominance nothing can be said. It would be possible for example for inbreeding to decrease variance.

Covariance Between Relatives

To find the covariance between the genotypic values G_k^i of individual X: A_iA_k and G_ℓ^j of individual Y: A_jA_ℓ we have six classes of expectations:

	μ_0	a_i	a_k	d_{ik}^a
μ_0	1	2	2	3
a_j	2	4	4	5
a_ℓ	2	4	4	5
$d_{j\ell}^a$	3	5	5	6

As in the variance situation, the first class has expectation μ_0^2, the second has zero expectation and the third has expectation $\mu_0 F_1 \sum_i p_i d_{ii}^a$. The remaining three classes require account to be taken of the identity relations between two, three or four of the four genes within X and Y. A complete set of four-gene measures was given by Gillois (1964) and Harris (1964), and we use the equivalent set of Cockerham (1971). Recall that the mean and variance required only the two-gene measure F_1.

For individuals W and Z with genes A_1A_2 and A_3A_4 respectively at the A locus, there are fifteen possible identity states. Generally we do not wish to distinguish between the two genes within an individual and so we can work with just nine measures. One measure, not needed for the expectations, is for the case where no subset of the four genes are identical by descent. Two further measures are the inbreeding coefficients F_{1_W}, F_{1_Z} for individuals W, Z. The remaining six measures are now defined in terms of the genes which are identical by descent in each case. An equivalence sign ≡ denotes identity by descent, and nothing is implied about the identity of genes not connected by an equivalence sign.

Identity Measure	Definition
$4\theta_{WZ}$	$\text{prob}(A_1 \equiv A_3) + \text{prob}(A_1 \equiv A_4) + \text{prob}(A_2 \equiv A_3) + \text{prob}(A_2 \equiv A_4)$
$2\gamma_{\ddot{W}Z}$	$\text{prob}(A_1 \equiv A_2 \equiv A_3) + \text{prob}(A_1 \equiv A_2 \equiv A_4)$
$2\gamma_{W\ddot{Z}}$	$\text{prob}(A_1 \equiv A_3 \equiv A_4) + \text{prob}(A_2 \equiv A_3 \equiv A_4)$
$\delta_{\ddot{W}.\ddot{Z}}$	$\text{prob}(A_1 \equiv A_2 \text{ and } A_3 \equiv A_4)$
$2\delta_{\ddot{W}+\ddot{Z}}$	$\text{prob}(A_1 \equiv A_3 \text{ and } A_2 \equiv A_4) + \text{prob}(A_1 \equiv A_4 \text{ and } A_2 \equiv A_3)$
$\delta_{\ddot{W}\ddot{Z}}$	$\text{prob}(A_1 \equiv A_2 \equiv A_3 \equiv A_4)$

We recognize θ_{WZ} as the coancestry coefficient of individuals W and Z.

These identity measures are now applied directly to give the expectations for the remaining three classes needed in the covariance of G_k^i and G_ℓ^j. Only non-zero terms are shown.

Class	Expectations
4	$4\theta_{XY} \sum_i p_i a_i^2$
5	$2(\gamma_{\ddot{X}Y} + \gamma_{X\ddot{Y}}) \sum_i p_i a_i d_{ii}^a$
6	$\delta_{\ddot{X}.\ddot{Y}} (\sum_i p_i d_{ii}^a)^2 + \delta_{\ddot{X}\ddot{Y}} \sum_i p_i (d_{ii}^a)^2$ $+ 2\delta_{\ddot{X}+\ddot{Y}} \sum_{i,k} p_i p_k (d_{ik}^a)^2$

Collecting terms over classes 1-6, and subtracting the product $\mathcal{E}\, G_k^i\, \mathcal{E}\, G_\ell^j$ leads to the desired covariance

$$\mathcal{C}_{XY} = 2\theta_{XY}\, \sigma_A^2 + 2\delta_{\ddot{X}+\ddot{Y}}\, \sigma_D^2 + 2(\gamma_{\ddot{X}Y} + \gamma_{X\ddot{Y}}) \sum_i p_i a_i d_{ii}^a$$

$$+ \delta_{\ddot{X}\ddot{Y}} \sum_i p_i (d_{ii}^a)^2 + (\delta_{\ddot{X}.\ddot{Y}} - F_{1_X} F_{1_Y})(\sum_i p_i d_{ii}^a)^2 . \qquad (4)$$

As a check on this result we see that it reduces to (3) when X and Y are the same individual. The reduction uses

$2\theta_{XX} = 1 + F_{1X}$, $2\delta_{\ddot{X}+\ddot{X}} = 1 - F_{1X}$, $\gamma_{\ddot{X}X} = F_{1X}$, $\delta_{\ddot{X}\ddot{X}} = F_{1X}$, $\delta_{\ddot{X}.\ddot{X}} = 0$.

If there is no dominance, equation (4) shows that the covariance of relatives

$$\mathcal{C}_{XY} = 2\theta_{XY}\,\sigma_A^2$$

depends explicitly only on the degree of relatedness of the relatives. If neither relative is inbred

$$\mathcal{C}_{XY} = 2\theta_{XY}\,\sigma_A^2 + 2\delta_{\ddot{X}+\ddot{Y}}\,\sigma_D^2$$

involves only the additive and dominance variance components. In other situations, however, the use of covariances in estimating variance components must take proper account of the three sums involving d_{ii}^a terms.

There is great simplification for the case of only two alleles at a locus. Explicit expressions for the effects and variance components are

$$a_1 = \frac{-p_2}{p_1}\,a_2 = p_2[p_1(G_1^1 - G_2^1) + p_2(G_2^1 - G_2^2)]$$

$$\frac{p_1}{p_2}\,d_1^1 = \frac{p_2}{p_1}\,d_2^2 = -d_2^1 = p_1p_2(G_1^1 + G_2^2 - 2G_2^1)$$

$$\sigma_A^2 = 2p_1p_2[p_1(G_1^1 - G_2^1) + p_2(G_2^1 - G_2^2)]^2,$$

$$\sigma_D^2 = p_1^2p_2^2(G_1^1 + G_2^2 - 2G_2^1)^2\ .$$

For two equally frequent alleles, $p_1 = p_2 = 0.5$, equations (3) and (4) reduce to

$$\sigma_G^2 = (1 + F_1)[\sigma_A^2 + (1 - F_1)\,\sigma_D^2]$$

$$\mathcal{C}_{XY} = 2\theta_{XY}\sigma_A^2 + (2\delta_{\ddot{X}+\ddot{Y}} + \delta_{\ddot{X}\ddot{Y}} + \delta_{\ddot{X}.\ddot{Y}} - F_{1X}F_{1Y})\sigma_D^2\ .$$

Two general comments are in order at this point. In the first place, just as discussed for the mean, the treatment of variances and covariances can be extended to include other types of initial population. Equations (3), (4) were based on the simple assumption that the initial population was constituted from the random pairing of gametes drawn at random from the infinite random-mating reference population. Secondly, the expressions for means, variances and covariances involve various identity measures. Procedures are well established (e.g. Cockerham, 1971) for the evaluation of these measures in any generation of an inbreeding system.

TWO-LOCUS THEORY

We have seen that one-locus theory is concerned with the dependencies between the actions or frequencies of allelic genes caused by dominance or inbreeding. To these features we must add epistasis, linkage and linkage disequilibrium when two-locus models are constructed.

The general two-locus model for the genotypic value of individuals with genotype A_iB_j/A_kB_ℓ formed by the union of gametes A_iB_j and A_kB_ℓ is written as

$$G^{ij}_{k\ell} = \mu_0 + a_i + a_k + b_j + b_\ell + ab_{ij} + ab_{i\ell} + ab_{kj} + ab_{k\ell} + d^a_{ik} + d^b_{j\ell} + ad^b_{ij\ell} + ad^b_{kj\ell} + bd^a_{jik} + bd^a_{\ell ik} + d^ad^b_{ikj\ell}$$

where the effects are defined for an infinite random-mating reference population in linkage equilibrium. Additive or dominance effects at the B locus are written as b or d^b respectively. The epistatic effects are written as ab (additive by additive), ad^b or bd^a (additive by dominance), and d^ad^b (dominance by dominance). Least squares values of the effects may be found as in the one-locus case.

Means and variances are sought for a population initiated from the random union of gametes A_iB_j in frequency $(p_iq_j + \mathcal{D}_{ij})$, where p_i and q_j are allelic frequencies and $\mathcal{D}_{ij}$ the initial linkage disequilibria. When expectations are taken of $G^{ij}_{k\ell}$ or $(G^{ij}_{k\ell})^2$, we need to take notice of which genes have descended from the same initial gamete so that the effects of such initial linkage disequilibria may be incorporated. We use our two-locus descent measures for this purpose.

A pair of genes are said to be equivalent by descent if they both descend from genes on one initial gamete. Allelic genes which are equivalent by descent are also identical by descent. For a random individual with genes A_1B_1/A_2B_2 in the population of interest there are six pairs of genes, and fifteen possible arrangements on gametes of the initial genes from which A_1, B_1, A_2, B_2 are descended. The number of arrangements reduces to nine when each locus is assumed to be equally inbred. The probabilities, or descent measures, of these nine arrangements sum to one and we choose to work with an equivalent set of eight summary measures. These eight measures, and the corresponding set of equivalent genes, are as follows.

Descent Measure	Equivalent Genes	Descent Measure	Equivalent Genes
F_1	$A_1 \equiv A_2$ or $B_1 \equiv B_2$	F_{11}	$A_1 \equiv A_2$ and $B_1 \equiv B_2$
F^1	$A_1 \equiv B_1$ or $A_2 \equiv B_2$	F^{11}	$A_1 \equiv B_1$ and $A_2 \equiv B_2$
${}_1F$	$A_1 \equiv B_2$ or $A_2 \equiv B_1$	${}_{11}F$	$A_1 \equiv B_2$ and $A_2 \equiv B_1$
${}_1F_1^1$	$A_1 \equiv B_1 \equiv B_2$ or $A_2 \equiv B_1 \equiv B_2$ or $A_1 \equiv A_2 \equiv B_1$ or $A_1 \equiv A_2 \equiv B_2$	F_{11}^{11}	$A_1 \equiv B_1 \equiv A_2 \equiv B_2$

In each of these definitions, nothing is implied about the equivalence by descent of genes not mentioned. The measures hold for an arbitrary number of alleles at a locus and allow the treatment of populations descending from a random sample or a specific set of initial gametes. They can be modified to accommodate specific initial individuals. We have also established general procedures for evaluating the measures in any situation, and have given detailed results for selfing, sib mating, mixed self and random mating, and random mating in finite monoecious populations (Cockerham and Weir, 1973, Weir and Cockerham, 1973, 1974).

Means

The descent measures allow genes on a random member of a population to be related to genes on initial gametes. This treatment is the essence of the expectation process. Frequencies of initial gametes are needed and the notation used is as follows. Two initial gametes carry genes A_iB_j and A_kB_ℓ respectively with probability $\theta^{ij}_{k\ell}$. Three or four initial gametes carry genes A_iB_j, A_k, B_ℓ or A_i, B_j, A_k, B_ℓ with probabilities $\theta^{ij}_{k|\ell}$ or $\theta^{i|j}_{k|\ell}$ respectively. Appropriate sums, denoted by dots, give frequencies for one, two or three genes and these initial frequencies were discussed by Cockerham and Weir (1973).

The expectations of the mean, the additive and dominance effects were given in the one-locus treatment. For the expectations of the remaining effects in $G^{ij}_{k\ell}$ we make use of the fact that effects where any one of the genes is independent of the others have zero expectations (Gallais, 1974). Specifically, each of the following sums is zero:

$$\sum_i p_i ab_{ij}\ , \quad \sum_j q_j ab_{ij},\ \sum_i p_i ad^b_{ij\ell},\ \sum_j q_j ad^b_{ij\ell},\ \sum_i p_i ad^b_{ijj}\ ,$$
$$\sum_i p_i d^a d^b_{ikj\ell},\ \sum_j q_j d^a d^b_{ikj\ell},\ \sum_i p_i d^a d^b_{ikjj},\ \sum_j q_j d^a d^b_{iij\ell}\ .$$

Terms leading only to such sums are omitted in the following list of expectations.

<u>Effect</u>	<u>Expectation</u>
additive by additive	
ab_{ij}	$F^1 \sum_{i,j} \theta^{ij} ab_{ij} = F^1 \sum_{i,j} \mathcal{D}_{ij} ab_{ij}$
$ab_{i\ell}$	${}_1F \sum_{i,j} \theta^{ij}_{\cdot\cdot} ab_{ij} = {}_1F \sum_{i,j} \mathcal{D}_{ij} ab_{ij}$
additive by dominance	
$ad^b_{ij\ell}$	${}_1F^1_1 \sum_{i,j} \theta^{ij}_{\cdot\cdot} ad^b_{ijj} = {}_1F^1_1 \sum_{i,j} \mathcal{D}_{ij} ad^b_{ijj}$
bd^a_{jik}	${}_1F^1_1 \sum_{i,j} \theta^{ij}_{\cdot\cdot} bd^a_{jii} = {}_1F^1_1 \sum_{i,j} \mathcal{D}_{ij} bd^a_{jii}$
dominance by dominance	
$d^a d^b_{ikj\ell}$	$F^{11}_{11} \sum_{i,j} \theta^{ij}_{\cdot\cdot} d^a d^b_{iijj}$ $+ (F_{11} - F^{11}_{11}) \sum_{i,j} \theta^{i\cdot}_{\cdot j} d^a d^b_{iijj}$

$$+ (F^{11} - F_{11}^{11}) \sum_{i,j,k,\ell} \theta_{k\ell}^{ij} d^a d^b_{ikj\ell}$$
$$+ ({}_{11}F - F_{11}^{11}) \sum_{i,j,k,\ell} \theta_{kj}^{i\ell} d^a d^b_{ikj\ell}$$
$$= F_{11} \sum_{i,j} p_i q_j d^a d^b_{iijj}$$
$$+ F_{11}^{11} \sum_{i,j} \mathcal{D}_{ij} d^a d^b_{iijj}$$
$$+ (F^{11} + {}_{11}F - 2F_{11}^{11}) \sum_{i,j,k,\ell} \mathcal{D}_{ij} \mathcal{D}_{k\ell} d^a d^b_{ikj\ell}.$$

Collecting the expectations of all sixteen terms in $\mathcal{E}\, G_{k\ell}^{ij}$ gives

$$\mu_F = \mu_0 + F_1 (\sum_i p_i d^a_{ii} + \sum_j q_j d^b_{jj}) + F_{11} \sum_{i,j} p_i q_j d^a d^b_{iijj}$$
$$+ 2(F^1 + {}_1F) \sum_{i,j} ab_{ij} \mathcal{D}_{ij} + 2\,{}_1F_1^1 \sum_{i,j} (ad^b_{ijj} + bd^a_{jii}) \mathcal{D}_{ij}$$
$$+ F_{11}^{11} \sum_{i,j} d^a d^b_{iijj} \mathcal{D}_{ij} + (F^{11} + {}_{11}F - 2F_{11}^{11}) \sum_{i,j,k,\ell} d^a d^b_{ikj\ell} \mathcal{D}_{ij} \mathcal{D}_{k\ell}. \tag{5}$$

This result is equivalent to that given by Gallais (1974).

For the non-inbred initial population, $F^1 = F^{11} = 1$ and all other descent measures are zero so that the mean becomes

$$\mu_I = \mu_0 + 2 \sum_{i,j} ab_{ij} \mathcal{D}_{ij} + \sum_{i,j,k,\ell} d^a d^b_{ikj\ell} \mathcal{D}_{ij} \mathcal{D}_{k\ell}$$

which reflects the effects of initial linkage disequilibrium. For completely inbred populations we know (Cockerham and Weir, 1973) that

$$F_1 = F_{11} = 1, \quad F^1 = {}_1F = {}_1F_1^1 = F^{11} = {}_{11}F = F_{11}^{11}$$

and the common value of the last six measures is written as $F^1_{(\infty)}$. Equation (5) then gives

$$\mu_1 = \mu_0 + \sum_i p_i d^a_{ii} + \sum_j q_j d^b_{jj} + \sum_{i,j} p_i q_j d^a d^b_{iijj}$$
$$+ F^1_{(\infty)} \sum_{i,j} (4ab_{ij} + 2ad^b_{ijj} + 2bd^a_{jii} + d^a d^b_{iijj}) \mathcal{D}_{ij} .$$

Evidently there is no longer a simple relation between μ_0, μ_1 and μ_F.

In the absence of linkage disequilibrium, the mean is seen to involve only dominance and dominance by dominance effects. The result for this special case was given by van Aarde (1974), who then equated F_{11} and F_1^2 in the case of no linkage to duplicate the result of Kempthorne (1957). In general however, we know that $F_1 \geq F_{11} \geq (F_1)^2$ with the last equality holding in the case of no linkage only when all matings are specified and all members of a generation have the same pedigree. Previously (Weir and Cockerham, 1969) we discussed the identity disequilibrium

$$\eta_{11} = F_{11} - (F_1)^2$$

and showed that, although it is quite small, it is still non-zero for no linkage in finite random-mating populations.

Variances

The first step in the evaluation of $\mathcal{E}(G^{ij}_{k\ell})^2$ is to identify the distinct classes of expectations of squares and products of effects, as was done in the one-locus situation. Recognizing symmetry between loci, there are 34 classes, the first 7 of which were treated in that previous case. The classes are now listed and the expectation process illustrated for classes 23 and 24.

Class	Typical element	No. of terms in class	Class	Typical element	No. of terms in class
1	μ_0^2	1	7	$(d^a_{ik})^2$	2
2	$\mu_0 a_i$	8	8	$\mu_0 ab_{ij}$	8
3	$\mu_0 d^a_{ik}$	4	9	$\mu_0 ad^b_{ij\ell}$	8
4	a_i^2	4	10	$\mu_0 d^a d^b_{ikj\ell}$	2
5	$a_i a_k$	4	11	$a_i b_j$	8
6	$a_i d^a_{ik}$	8	12	$a_i ab_{ij}$	16

Class	Typical element	No. of terms in class	Class	Typical element	No. of terms in class
13	$a_i ab_{kj}$	16	24	$ab_{ij}ad^b_{kj\ell}$	16
14	ab^2_{ij}	4	25	$d^a_{ik}ad^b_{ij\ell}$	8
15	$ab_{ij}ab_{ik}$	8	26	$d^a_{ik}bd^a_{jik}$	8
16	$ab_{ij}ab_{k\ell}$	4	27	$(ad^b_{ij\ell})^2$	4
17	$a_i d^b_{j\ell}$	8	28	$ad^b_{ij\ell}ad^b_{kj\ell}$	4
18	$ab_{ij}d^a_{ik}$	16	29	$ad^b_{ij\ell}bd^a_{jik}$	8
19	$d^a_{ik}d^b_{j\ell}$	2	30	$a_i d^a d^b_{ikj\ell}$	8
20	$a_i ad^b_{ij\ell}$	8	31	$ab_{ij}d^a d^b_{ikj\ell}$	8
21	$a_i ad^b_{kj\ell}$	8	32	$d^a_{ik}d^a d^b_{ikj\ell}$	4
22	$a_i bd^a_{jik}$	16	33	$ad^b_{ij\ell}d^a d^b_{ikj\ell}$	8
23	$ab_{ij}ad^b_{ij\ell}$	16	34	$(d^a d^b_{ikj\ell})^2$	1

$$\begin{aligned}\delta(ab_{ij}ad^b_{ij\ell}) &= {}_1F^1_1 \sum_{i,j} \theta^{ij}_{..}ab_{ij}ad^b_{ijj} + (F_1 - {}_1F^1_1) \sum_{i,j} \theta^{i.}_{.j}ab_{ij}ad^b_{ijj} \\ &\quad + (F^1 + {}_1F - 2{}_1F^1_1)/2 \sum_{i,j,\ell} \theta^{ij}_{.\ell}ab_{i\ell}ad^b_{ij\ell}\end{aligned}$$

$$\begin{aligned}\delta(ab_{ij}ad^b_{kj\ell}) &= F^{11}_{11} \sum_{i,j} \theta^{ij}_{..}ab_{ij}ad^b_{ijj} + (F_{11} - F^{11}_{11}) \sum_{i,j} \theta^{i.}_{.j}ab_{ij}ad^b_{ijj} \\ &\quad + ({}_1F^1_1 - F^{11}_{11}) \sum_{i,j,\ell} \theta^{ij}_{.\ell}ab_{i\ell}ad^b_{ij\ell} \\ &\quad + (F^{11} - F^{11}_{11}) \sum_{i,j,k,\ell} \theta^{ij}_{k\ell}ab_{ij}ad^b_{kj\ell} \\ &\quad + ({}_{11}F - F^{11}_{11}) \sum_{i,j,k,\ell} \theta^{i\ell}_{kj}ab_{i\ell}ad^b_{kj\ell}.\end{aligned}$$

Simplification of the sum of all these expectations follows the introduction of the following familiar variance components:

$$\sigma_A^2 = 2\sigma_a^2 + 2\sigma_b^2 = 2\sum_i p_i a_i^2 + 2\sum_j q_j b_j^2, \quad \sigma_{AA}^2 = 4\sigma_{ab}^2 = 4\sum_{i,j} p_i q_j ab_{ij}^2$$

$$\sigma_D^2 = \sigma_{d^a}^2 + \sigma_{d^b}^2 = \sum_{i,k} p_i p_k (d_{ik}^a)^2 + \sum_{j,\ell} q_j q_\ell (d_{j\ell}^b)^2$$

$$\sigma_{AD}^2 = 2\sigma_{ad^b}^2 + 2\sigma_{bd^a}^2 = 2\sum_{i,j,\ell} p_i q_j q_\ell (ad_{ij\ell}^b)^2 + 2\sum_{i,j,k} p_i q_j p_k (bd_{jik}^a)^2$$

$$\sigma_{DD}^2 = \sigma_{d^a d^b}^2 = \sum_{i,j,k,\ell} p_i q_j p_k q_\ell (d^a d_{ikj\ell}^b)^2$$

When the square of the mean is subtracted from $\mathcal{E}(G_{k\ell}^{ij})^2$ we obtain the general variance formula displayed in equation (6) on the next two pages. Some features of this unwieldy result can be noted.

The coefficients of the five variance components in the first line of (6) almost show the kind of pattern often suggested in the literature. The coefficients of the epistatic components differ from the products of the corresponding additive and dominance components by identity disequilibrium:

$$\begin{aligned}\sigma_{G_F}^2 &= (1 + F_1)\,\sigma_A^2 + (1 - F_1)\,\sigma_D^2 + (1 + F_1)^2\,\sigma_{AA}^2 \\ &\quad + (1 + F_1)(1 - F_1)\,\sigma_{AD}^2 + (1 - F_1)^2\,\sigma_{DD}^2 \\ &\quad + \eta_{11}\,(\sigma_{AA}^2 - \sigma_{AD}^2 + \sigma_{DD}^2) + \ldots\end{aligned}$$

The identity disequilibrium again features in the variance of genes which are unlinked, for which there is no linkage disequilibrium, and for which there are no epistatic effects. We find from equation (6) that

$$\begin{aligned}\sigma_{G_F}^2 &= \sigma_{G_F}^2\ (\text{locus A only}) + \sigma_{G_F}^2\ (\text{locus B only}) \\ &\quad + 2\eta_{11} \sum_i p_i d_{ii}^a \sum_i q_i d_{jj}^b\ .\end{aligned}$$

The first two, one-locus, variances are as given in equation (3) and the last term may be regarded as the covariance between the effects at the two loci. There is still a dependence between the loci, caused by the inbreeding system, depending on dominance at each locus and the amount of identity disequilibrium.

$$
\begin{aligned}
\sigma^2_{G_F} = {} & (1+F_1)\sigma^2_A + (1-F_1)\sigma^2_D + (1+2F_1+F_{11})\sigma^2_{AA} + (1-F_{11})\sigma^2_{AD} \\
& + (1-2F_1+F_{11})\sigma^2_{DD} + F_1[\sum_i p_i(4a_i + d^a_{ii})d^a_{ii} + \sum_j q_j(4b_j + d^b_{jj})d^b_{jj}] \\
& + 2F_{11}(\sum_i p_i d^a_{ii})(\sum_j q_j d^b_{jj}) + 2(F_1+F_{11})\sum_{i,j} p_i q_j[(ad^b_{ijj})^2 \\
& + (bd^a_{jii})^2 + 4ab_{ij}(ad^b_{ijj} + bd^a_{jii})] + 4F_{11}\sum_{i,j} p_i q_j(d^a_{ii}ad^b_{ijj} \\
& + d^b_{jj}bd^a_{jii} + 2ad^b_{ijj}bd^a_{jii}) + F_{11}\sum_{i,j} p_i q_j[4(a_i+b_j) + 8ab_{ij} \\
& + 2(d^a_{ii} + d^b_{jj}) + 4(ad^b_{ijj} + bd^a_{jii}) + d^a d^b_{iijj}]d^a d^b_{iijj} \\
& + 4(F_1+F_{11})\sum_{i,j} p_i q_j(a_i ad^b_{ijj} + b_j bd^a_{jii}) \\
& + (F_1-F_{11})\sum_{i,j,k} p_i q_j p_k(2d^a_{ik} + 4bd^a_{jik} + d^a d^b_{ikjj})d^a d^b_{ikjj} \\
& + (F_1-F_{11})\sum_{i,j,\ell} p_i q_j q_\ell(2d^b_{j\ell} + 4ad^b_{ij\ell} + d^a d^b_{iij\ell})d^a d^b_{iij\ell} \\
& + 4(F^1 + {}_1F)\sum_{i,j} a_i b_j \mathcal{D}_{ij} + 4(F^1 + {}_1F + 2\,{}_1F^1_1)\sum_{i,j}(a_i+b_j)ab_{ij}\mathcal{D}_{ij} \\
& + 2(F^1 + {}_1F + 4\,{}_1F^1_1 + 2F^{11}_{11})\sum_{i,j}(ab_{ij})^2\mathcal{D}_{ij} + 2F^{11}_{11}\sum_{i,j} d^a_{ii}d^b_{jj}\mathcal{D}_{ij} \\
& + 4\,{}_1F^1_1\sum_{i,j}[(a_i + 2ab_{ij})d^b_{jj} + (b_j + 2ab_{ij})d^a_{ii}]\mathcal{D}_{ij} \\
& + 2({}_1F^1_1 + F^{11}_{11})\sum_{i,j}[(2b_j + 4ab_{ij} + bd^a_{jii})bd^a_{jii} \\
& + (2a_i + 4ab_{ij} + ad^b_{ijj})ad^b_{ijj}]\mathcal{D}_{ij} \\
& + 4\,{}_1F^1_1\sum_{i,j}[(2a_i + d^a_{ii})bd^a_{jii} + (2b_j + d^b_{jj})ad^b_{ijj}]\mathcal{D}_{ij} \\
& + 2(F^1 + {}_1F - 2\,{}_1F^1_1)\sum_{i,j,k} p_k[(2a_k + d^a_{ik})bd^a_{jik} + 2ab_{kj}d^a_{ik}]\mathcal{D}_{ij} \\
& + 2(F^1 + {}_1F - 2\,{}_1F^1_1)\sum_{i,j,\ell} q_\ell[(2b_\ell + d^b_{j\ell})ad^b_{ij\ell} + 2ab_{i\ell}d^b_{j\ell}]\mathcal{D}_{ij}
\end{aligned}
$$

$$
\begin{aligned}
&+ 8({}_1F_1^1 - F_{11}^{11}) \sum_{i,j,k} p_k(d_{ik}^a + bd_{jik}^a)ad_{kjj}^b\mathcal{D}_{ij} \\
&+ 8({}_1F_1^1 - F_{11}^{11}) \sum_{i,j,\ell} q_\ell(d_{j\ell}^b + ad_{ij\ell}^b)bd_{\ell ii}^a\mathcal{D}_{ij} \\
&+ 4F_{11}^{11} \sum_{i,j} (d_{ii}^a ad_{ijj}^b + 2ad_{ijj}^b bd_{jii}^a + d_{jj}^b bd_{jii}^a)\mathcal{D}_{ij} \\
&+ 2(F^1 + {}_1F - 2F_{11}^{11}) \sum_{i,j,k} p_k(2ab_{kj} + bd_{jik}^a)bd_{jik}^a\mathcal{D}_{ij} \\
&+ 2(F^1 + {}_1F - 2F_{11}^{11}) \sum_{i,j,\ell} q_\ell(2ab_{i\ell} + ad_{ij\ell}^b)ad_{ij\ell}^b\mathcal{D}_{ij} \\
&+ F_{11}^{11} \sum_{i,j} [4(a_i + b_j) + 8ab_{ij} + 2(d_{ii}^a + d_{jj}^b) + 4(ad_{ijj}^b + bd_{jii}^a) \\
&+ d^ad_{iijj}^b]d^ad_{iijj}^b\mathcal{D}_{ij} \\
&+ 2({}_1F_1^1 - F_{11}^{11}) \sum_{i,j,k} p_k(2a_k + 4ab_{kj} + d_{ik}^a + 2ad_{kjj}^b + 2bd_{jik}^a \\
&+ d^ad_{ikjj}^b)d^ad_{ikjj}^b\mathcal{D}_{ij} \\
&+ 2({}_1F_1^1 - F_{11}^{11}) \sum_{i,j,\ell} q_\ell(2b_\ell + 4ab_{i\ell} + d_{j\ell}^b + 2bd_{\ell ii}^a + 2ad_{ij\ell}^b \\
&+ d^ad_{iij\ell}^b)d^ad_{iij\ell}^b\mathcal{D}_{ij} \\
&+ 2(4F_{11}^{11} - F^{11} - {}_{11}F - 4\,{}_1F_1^1 + F^1 + {}_1F) \sum_{i,j,k,\ell} p_kq_\ell(4ab_{k\ell} \\
&+ 2ad_{kj\ell}^b + 2bd_{\ell ik}^a + d^ad_{ikj\ell}^b)d^ad_{ikj\ell}^b\mathcal{D}_{ij} \\
&+ (F^{11} + {}_{11}F - 2F_{11}^{11}) \sum_{i,j,k,\ell} [4a_iad_{kj\ell}^b + 4b_jbd_{\ell ik}^a + 4ab_{ij}ab_{k\ell} \\
&+ 8ab_{ij}(ad_{kj\ell}^b + bd_{\ell ik}^a) + 4d_{ik}^aad_{ij\ell}^b + 4d_{j\ell}^bbd_{jik}^a + 8ad_{ij\ell}^bbd_{jik}^a \\
&+ 2ad_{ij\ell}^bad_{kj\ell}^b + 2bd_{jik}^abd_{\ell ik}^a + (4a_i + 4b_j + 8ab_{ij} + 2d_{ik}^a + 2d_{j\ell}^b \\
&+ 2d_{ik}^ad_{j\ell}^b + 4ad_{ij\ell}^b + 4bd_{jik}^a + d^ad_{ikj\ell}^b)d^ad_{ikj\ell}^b]\mathcal{D}_{ij}\mathcal{D}_{k\ell} \\
&- (\mu_F - \mu_0)^2 \; . \qquad (6)
\end{aligned}
$$

Covariances

Expectations of terms in the product $G_{k\ell}^{ij}\ G_{pq}^{mn}$ for relatives X: A_iB_j/A_kB_ℓ and Y: A_mB_n/A_pB_q require eight-gene descent measures. Such measures would give the probabilities of all the ways in which eight genes shared by two individuals could be located on one to eight initial gametes. In the one-locus case we found many more four-gene measures than there were two-gene measures, so in this two-locus case we expect many more eight-gene measures than there are four-gene measures. There is great reduction in the number of eight-gene measures however when symmetry between the loci is assumed and if the two individuals X, Y are in the same generation of a regular pedigree or monoecious random-mating system. This latter case allows the arrangement of the four gametes among individuals X, Y to be ignored.

Regardless of the number of measures, however, we know that the covariance of $G_{k\ell}^{ij}$ and G_{pq}^{mn} will contain exactly the same terms involving effects, linkage disequilibria and gene frequencies as does the variance expression (6). The difference will be in the descent measure coefficients. The variance has coefficients which are measures defined for genes in one individual while the covariance has coefficients which are measures defined for genes in two individuals. The same relationship is apparent between equations (3) and (4).

We have not identified or evaluated the measures for two individuals. Gallais (1974) has identified one set, but gave no general indication of how to evaluate components of the set. By restricting epistasis to additive by additive he could list an expression for covariance. (This expression contains two omissions: $E\beta_{i_1 i_1}\xi_{\overline{j_1 k_2}}$, $E\beta_{i_1 * j_1}\xi_{j_1 k_2 *}$ in his notation). The terms he gives reduce to the appropriate terms in equation (6) when the two individuals are the same.

The task is made considerably easier if linkage equilibrium is assumed. It is then not necessary to consider the equivalence by descent of non-allelic genes, as the initial frequency of any set of such genes is just the product of the corresponding allelic frequencies regardless of on how many gametes these initial non-allelic genes are located. We can work entirely with identity measures. For one individual there are 2 identity measures, F_1 , F_{11}. For two individuals there are 9 identity measures at each locus, and so 81 at two loci. However we do not need the measure for complete non-identity, and we generally assume that each locus is equally inbred so that we need to work with 44 identity measures. In this case of linkage equilibrium, the two-locus variance has 33 terms, the coefficients of which are combinations of 2 identity measures and the two-locus covariance will have the same 33 terms but with coefficients that are combinations of 44 identity measures.

Denniston (1967, 1975) and van Aarde (1963, 1975) both made the further assumption that individuals are not inbred. Identity by descent is still possible for genes in different individuals (relatives) however. The two-locus variance and covariance in such cases of linkage equilibrium and no inbreeding involve only the five variance components. Both van Aarde and Denniston discussed the evaluation of the various measures of "between individual" identity by descent.

DISCUSSION

We have sought to demonstrate the use of our two-locus descent measures in finding two-locus means, variances and covariances. Both positive and negative features emerge.

On the positive side we believe that the measures, for means and variances, provide coefficients that are capable of immediate interpretation. The two-locus inbreeding coefficient F_{11} for example is an obvious extension of the one-locus coefficient F_1. The expressions also explicitly demonstrate the effects of linkage disequilibrium on means and variances, and initial allelic frequencies enter explicitly. The effects of linkage and inbreeding are accounted for by the descent measures. A variety of assumptions about the initial population can be made. Of some importance we feel is that procedures have been established (Cockerham and Weir, 1973), and illustrated (Weir and Cockerham, 1973, 1974), for evaluating these measures in any situation. The effects of mating system, linkage and time are thus accommodated and demonstrated.

Balancing these factors are two negative ones. We do not have a detailed theory of eight-gene measures, and hesitate to establish what would necessarily be a complex machinery. The complexity of the expressions, such as equation (6), is the second negative feature. As it stands, the result is of little use. The most promising approach would seem to be to find simple approximations for descent measures, maybe in terms of the one-locus inbreeding coefficient as has been suggested recently by Gillois (this conference).

BIBLIOGRAPHY

*Cockerham, C. C. (1954). An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. Genetics 39:859-882.

*Cockerham, C. C. (1956). Effects of linkage on the covariances between relatives. Genetics 41:138-141.

Cockerham, C. C. (1971). Higher order probability functions of identity of alleles by descent. Genetics 69: 235-246.

Cockerham, C. C., Weir, B. S. (1973). Descent measures for two loci with some applications. Theor. Pop. Biol. 4: 300-330.

*Denniston, C. (1967). Probability and genetic relationship. Unpublished thesis, University of Wisconsin, Madison.

*Denniston, C. (1975). Probability and genetic relationship: two loci. Annals of Human Genetics 39:89-104.

*Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. Trans. Roy. Soc. Edinburgh 52:399-433.

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

*Gallais, A. (1974). Covariances between arbitrary relatives with linkage and epistasis in the case of linkage disequilibrium. Biometrics 30:429-446.

Gillois, M. (1964). La relation d'identité en génétique. Thèse Fac. Sci., Universite de Paris, Paris.

*Gillois, M. (1966). Note sur la variance et la covariance génotypiques entre apparantés. Ann. Inst. Henri Poincaré 2:349-352.

*Harris, D. L. (1964). Genotypic covariances between inbred relatives. Genetics 50:1319-1348.

*Kempthorne, O. (1954). The correlation between relatives in a random mating population. Proc. Roy. Soc. London B 143:103-113.

*Kempthorne, O. (1955). The theoretical values of correlations between relatives in random mating populations. Genetics 40:153-167.

*Kempthorne, O. (1957). An introduction to genetic statistics. Wiley, New York.

Malécot, G. (1948). Les mathématiques de l'hérédité. Masson, Paris.

*Schnell, F. W. (1963). The covariance between relatives in the presence of linkage. In Statistical Genetics and Plant Breeding. Natl. Acad. Sci., Natl. Res. Council. Publication 982:468-483.

Stuber, C. W., Cockerham, C. C. (1966). Gene effects and variances in hybrid populations. Genetics 54:1279-1286.

*van Aarde, I. M. R. (1963). Covariances of relatives in random mating populations with linkage. Unpublished thesis, Iowa State University, Ames.

*van Aarde, I. M. R. (1974). The effect of linkage on the mean value of inbreds derived from a random mating population. Genetics 78:1245-1249.

*van Aarde, I. M. R. (1975). The covariance of relatives derived from a random mating population. Theor. Pop. Biol. 8:166-183.

Weir, B. S., Cockerham, C. C. (1969). Group inbreeding with two linked loci. Genetics 63:711-742.

Weir, B. S., Cockerham, C. C. (1973). Mixed self and random mating at two loci. Genetical Research 21:247-262.

Weir, B. S., Cockerham, C. C. (1974). Behavior of pairs of loci in finite monoecious populations. Theor. Pop. Biol. 6:323-354.

*Wright, S. (1935). The analysis of variance and the correlations between relatives with respect to deviations from an optimum. J. Genetics 30:243-256.

Wright, S. (1951). The genetical structure of populations. Annals of Eugenics 15:323-354.