J. Genet., Vol. 67, No. 1, April 1988, pp. 53–61. © Printed in India.

Extending the definition of additive genetic variance to more than one gene—a viewpoint

V ARUNACHALAM
Division of Genetics, Indian Agricultural Research Institute, New Delhi 110 012, India

MS received 24 February 1988

Abstract. Following the Fisherian approach, the expression for additive genetic variance is derived in a single gene system through a regression equation in two variables which are used to obtain the additive and dominance variances. The approach is extended to two genes with restricted linkage and inbreeding. It was brought out that additive genetic variance defined essentially for one gene does not extend *per se* to multi-gene systems.

Keywords. Genetic variance; additive variance; linkage; inbreeding; epistasis.

1. Introduction

Additive genetic variance is a vital component in any genetic system not only because it is the heritable part of total genetic variance but it is one of the main factors in choosing proper mating designs and selection schemes to accelerate improvement per unit time in successive filial generations. Fisher (1941) defined additive genetic variance; in particular, he devised two parameters, 'average excess' (α) and 'average effect' (α) of gene substitution to partition the genotypic variance in a single gene system into additive and dominance variances. From then on, the topic of partitioning total genetic variance into its components has been dealt with extensively, both in single (Li 1955; Falconer 1964; Crow and Kimura 1970; Mather and Jinks 1971) and multi-gene (Crow and Kimura 1970; Weir and Cockerham 1977; Ewens 1979) systems under most general assumptions on inbreeding and linkage. This paper does not aim therefore to deal with that topic again. But it does aim to obtain by a method using the Fisherian approach, not reported so far, additive variance in a single-gene system and in a two-gene system with restricted inbreeding and linkage. The objective is to elucidate, using this approach, that additive variance, as defined originally by Fisher (1941) for a single-gene system, does not extend *pari passu* to a two-gene system. The procedure will first be demonstrated for a single-gene and then be extended to a two-gene system. The difficulties in extending the definition of additive genetic variance from one to two genes will be highlighted.

2. Method

2.1 *One gene*

We use symbolic algebra for representing values of a quantitative character and various genetic effects as described in Arunachalam and Owen (1971) (to be

referred to henceforth as A&O). Let AA, Aa and aa be the genotypes, p, the frequency of gene A, $q = 1-p$, that of a gene. Let the genotypic values of a character be also represented by AA, Aa and aa.

If we define $S = \mathrm{A}p + \mathrm{a}q$ in symbolic notation, $\mathrm{A} = S + q(\mathrm{A}-\mathrm{a})$ and $\mathrm{AA} = [S + q(\mathrm{A}-\mathrm{a})]^2 = S_1 + 2qL + q^2Q$. Similarly, $\mathrm{Aa} = S_1 + (q-p)L - pqQ$ and $\mathrm{aa} = S_1 - 2pL + p^2Q$ where

$$S_1 = S^2,$$

$$L = \tfrac{1}{2}(\mathrm{d}S_1/\mathrm{d}p)$$
$$= (\mathrm{A}-\mathrm{a})(\mathrm{A}p + \mathrm{a}q) = p(\mathrm{AA}-\mathrm{Aa}) + q(\mathrm{Aa}-\mathrm{aa})$$
$$= \text{additive effect, and}$$

$$Q = \tfrac{1}{2}(\mathrm{d}^2S_1/\mathrm{d}p^2) = (\mathrm{A}-\mathrm{a})^2 = \text{dominance effect.}$$

Let θ (Hayman 1954) be a variable taking the values 1, 0 and -1 for AA, Aa and aa, respectively, and $\phi = 1-\theta^2$. A&O constructed a variable, $\omega = (p-q)\theta + \phi$ and showed that θ and ω could be used to compute the additive and dominance variances. We note that $E(\theta) = p-q$, $E(\omega) = p^2+q^2$ and $E(\phi) = 2pq$. If we measure the variables from their respective means and use X to denote the value of a quantitative character, then those for the three genotypes can be generalised in the form of a regression equation given by

$$X = S_1 + (q-p+\theta)L + \tfrac{1}{2}[q^2+p^2+(q-p)\theta-\phi]Q. \qquad (1)$$

It has further been shown by A & O that the regression coefficient of X on $\theta = L$ and that of X on $\omega = -\frac{1}{2}Q$. They were defined as the 'average effects' of gene substitution with respect to the variables θ and ω. Thus, under random mating, additive variance $= L^2\sigma_\theta^2 = 2pqL^2$ and dominance variance $= \frac{1}{4}Q^2\sigma_\omega^2 = p^2q^2Q^2$, so that the genotypic variance $= 2pqL^2 + p^2q^2Q^2$.

Now the genotypic frequencies under inbreeding are given by,

$$\text{freq } (AA) = p^2\mathrm{f}' + p\mathrm{f}$$
$$\text{freq } (Aa) = 2pq\mathrm{f}'$$

and

$$\text{freq } (aa) = q^2\mathrm{f}' + p\mathrm{f} \qquad (2)$$

where $\mathrm{f}' = 1-\mathrm{f}$ and f = the inbreeding coefficient.

Rewriting (1), we get

$$X = M + [L + \tfrac{1}{2}(q-p)Q]\theta - \tfrac{1}{2}Q\phi, \qquad (3)$$

where

$$M = S^2 + (q-p)L + \tfrac{1}{2}(q^2+p^2)Q, \quad \text{which on simplification,}$$
$$= \tfrac{1}{2}(AA + aa)$$
$$= \text{mean value of the two homozygotes in the population.}$$

When the frequencies given by (2) hold, the value of $E(\theta)$ does not change and $= (p-q)$. Thus θ can still be used to obtain the additive component of variance. We note,

$$E(\theta\phi) = 0,$$

$$E(\theta)\,E(\phi) = 2pq\mathrm{f}'(p-q),$$

$$C(\theta,\phi) = \text{covariance of } \theta \text{ and } \phi$$
$$= 2pq\mathrm{f}'(q-p),$$

$$\sigma_\theta^2 = 2pq(1+\mathrm{f}),$$

$$\sigma_\phi^2 = 2pq\mathrm{f}'(1-2pq\mathrm{f}').$$

We now find a variable ω in terms of θ and ϕ such that θ and ω are uncorrelated.

Let $\omega = \phi + B\theta$

$$C(\theta,\omega) = C(\theta,\phi) + B\sigma_\theta^2 = 0, \text{ if } B = -\,C(\theta,\phi)/\sigma_\theta^2,$$

i.e., $B = (p-q)\mathrm{f}'\mathrm{f}''$, where $\mathrm{f}'' = 1/(1+\mathrm{f})$.

Thus

$$\omega = \phi + (p-q)\mathrm{f}'\mathrm{f}''\theta \tag{4}$$

is uncorrelated with θ.

Substituting for ϕ in (3), we get

$$X = M + [L + (q-p)\mathrm{ff}''Q]\theta - \tfrac{1}{2}Q\omega. \tag{5}$$

From (4), $\phi = \omega + (q-p)\mathrm{f}'\mathrm{f}''\theta$, so that

$$\sigma_\phi^2 = C(\phi,\omega) + (q-p)\mathrm{f}'\mathrm{f}''C(\theta,\phi).$$

Also from (4),

$$\sigma_\omega^2 = C(\phi,\omega), \text{ since } C(\theta,\omega) = 0.$$

Hence

$$\sigma_\phi^2 = \sigma_\omega^2 + 2pq(q-p)^2(1-\mathrm{f})^2/(1+\mathrm{f}).$$

On simplification,

$$\sigma_\omega^2 = 4pq\mathrm{f}'[(1+\mathrm{f})pq + \mathrm{ff}''(1-4pq)].$$

From (5),

$$C(X,\theta) = [L + (q-p)\mathrm{ff}''Q]\sigma_\theta^2,$$

and

$$\alpha_\theta = \text{average effect of substitution of } A \text{ for } a \text{ with respect to the variable } \theta,$$
$$= C(X,\theta)/\sigma_\theta^2 = L + (q-p)\mathrm{ff}''Q.$$

$$V_A, \text{ additive variance} = \alpha_\theta^2\sigma_\theta^2$$
$$= 2pq(1+\mathrm{f})[L + (q-p)\mathrm{ff}''Q]^2.$$

From (2), $C(X,\omega) = -\tfrac{1}{2}Q\sigma_\omega^2$, since $C(\theta,\omega) = 0$.

Hence, $\alpha_\omega = C(X,\omega)/\sigma_\omega^2 = -\tfrac{1}{2}Q$.

$$V_D, \text{ dominance variance} = \alpha_\omega^2 \sigma_\omega^2$$
$$= pqf'Q^2[pq(1+f)+ff''(1-4pq)]$$

The expressions for V_A and V_D have earlier been derived in a slightly different form by Crow and Kimura (1970). However, the regression and symbolic algebra approach employed here provide an elegant alternative.

Thus the genotypic variance under inbreeding,

$$V_T = V_A + V_D.$$

It is interesting to note that, if the X values of AA, Aa and aa are denoted by d, h and $-d$, following Mather and Jinks (1971),

$$L = d+(q-p)h, \qquad Q = -2h, \text{ and } \alpha = d+h(q-p)f'f''.$$

2.2 *Two genes*

2.2a *Inbreeding coefficient:* Following the classical definition of the inbreeding coefficient as the correlation between uniting gametes, one can, unlike in the single-gene case, define a set of correlation coefficients, f_{ij}, between uniting gametes $(i,j = 1,4)$. However, for the objectives of this paper, it is enough to confine our attention to the case of identity of both genes in a gamete by descent. Further, it is useful to establish correspondence in results between one- and two-gene systems by assuming $f_{ij} = f$.

For the purpose of partitioning the total genetic variance, it is convenient to write the frequencies of the 9 genotypes (ignoring position effects) in the form of a 3×3 matrix:

	A_2A_2	A_2a_2	a_2a_2	Marginal
A_1A_1	$(1-f)P_1^2+fP_1$	$2(1-f)P_1P_3$	$(1-f)P_3^2+fP_3$	$(1-f)p_1^2+fP_1$
A_1a_1	$2(1-f)P_1P_4$	$2(1-f)(P_1P_2+P_3P_4)$	$2(1-f)P_2P_3$	$2(1-f)p_1q_1$
a_1a_1	$(1-f)P_4^2+fP_4$	$2(1-f)P_2P_4$	$(1-f)P_2^2+fP_2$	$(1-f)q_1^2+fq_1$
Marginal	$(1-f)p_2^2+fp_2$	$2(1-f)p_2q_2$	$(1-f)q_2^2+fq_2$	1

where P_i $(i = 1,4)$ are the frequencies of the gametes, A_1A_2, a_1a_2, A_1a_2 and a_1A_2, respectively, p_1, gene frequency of A_1, and p_2, gene frequency of A_2.

We now extend the use of the variables θ and ω to two genes. Let us consider the case of two independent genes $(D \equiv 0)$. Following A&O, we define as in the one-gene case, variables θ_i, ω_i corresponding to the genes A_i $(i = 1,2)$ so that $E(\theta_i) = p_i - q_i$; $E(\omega_i) = p_i^2 + q_i^2$. Then any quantitative character X can be represented as a regression equation in the variables θ and ω.

We measure the variables from their means and as in the one-gene case, set $\omega_i = \phi - f'f''(q_i - p_i)\theta_i$, so that $E(\theta_i) = E(\omega_i) = 0$. When $f = D = 0$, the variables θ_1, θ_2, ω_1, ω_2, $\theta_1\theta_2$, $\theta_1\omega_2$, $\theta_2\omega_1$ and $\omega_1\omega_2$ are mutually uncorrelated and have means = zero. A character X can then be represented as a perfect regression equation in those variables so that

$$Y = X - E(X) = \sum_{i=1}^{8} \alpha_i \xi_i,$$

where

$$\begin{aligned} \xi_1 &= \theta_1, & \xi_5 &= \theta_1\theta_2, \\ \xi_2 &= \theta_2, & \xi_6 &= \theta_1\omega_2, \\ \xi_3 &= \omega_1, & \xi_7 &= \theta_2\omega_1, \qquad \text{and} \\ \xi_4 &= \omega_2, & \xi_8 &= \omega_1\omega_2, \end{aligned}$$

can be used to calculate the various components of variance.

When $f \neq 0$, ω_1 and ω_2 are correlated and $E(\omega_1\omega_2) = 4p_1q_1p_2q_2ff'$. The covariance matrix, $[C]$ of the variables, ξ_i is then non-diagonal, and is given in table 1.

The expressions, $\mathrm{Cov}(X, \xi_i)$ can easily be obtained using symbolic algebra.

Following A & O, the set of average excesses and average effects can be obtained and the total variance partitioned. But the variables, ξ_i are correlated. The components of variance are therefore modified by the contribution of the covariances between correlated variables, as will be shown now.

Consider, for instance, the case $D = f = 0$. Here the variables are mutually uncorrelated.

It is important to note that, in this case, the covariance matrix of the variables is *diagonal*, so that the "average excesses",

$$\boldsymbol{\alpha} = \mathbf{Z}C^{-1} \tag{6}$$

where,

$$\begin{aligned} \boldsymbol{\alpha} &= [\alpha_1, \alpha_2, \ldots \alpha_8] \\ \mathbf{Z} &= [\mathrm{Cov}(X, \xi_1), \ldots \mathrm{Cov}(X, \xi_8)] \\ c^{ii} &= 1/\mathrm{Var}(\xi_i), \; c^{ij} = 0, \qquad (i, j = 1, 8). \end{aligned}$$

The eight components of variance are then given by

$$\begin{aligned} \mathrm{Var}(X_i) &= \alpha_i \,\mathrm{Cov}(X, \xi_i) \\ &= [\mathrm{Cov}(X, \xi_i)]^2/\mathrm{Var}(\xi_i) \\ &= \alpha_i a_i \,\mathrm{Var}(\xi_i), \qquad (i = 1, \; 8), \end{aligned}$$

where $\mathrm{Var}(X_1) = \mathrm{Var}(\text{Add. } 1)$ etc., and $\alpha_i = a_i$. The set of variables ξ_i was called basic set by A & O (table 1) (Add.–additive).

When they considered partitioning the total genetic variance in a two-gene system with $D \neq 0$, they could not construct such a set of mutually uncorrelated variables. They defined another set of variables, ψ_i, such that their covariance matrix entailed a number of zero elements, if not entirely diagonal. As a consequence, it was difficult to explain the average effects in the Fisherian sense (see § 3.8 of A & O).

When we now consider inbreeding, such difficulties are encountered even when $D = 0$, since ξ_3 and ξ_4 are correlated. However, we keep the basic set of variables, ξ_i, and use the covariance matrix of ξ_i's which is now non-diagonal.

$$\alpha_i = c^{i1}\,\mathrm{Cov}(X, \xi_1) + c^{i2}\,\mathrm{Cov}(X, \xi_2) + \ldots + c^{i8}\,\mathrm{Cov}(X, \xi_8), \quad (i = 1, 8).$$

Table 1. Covariance matrix of the basic set of variables, ξ_i, when $D = 0$, $f \neq 0$ in a two-gene system.

	ξ_1	ξ_2	ξ_3	ξ_4	ξ_5	ξ_6	ξ_7	ξ_8
ξ_1	$2s_1k$	0	0	0	0	$-4s_{12}ff'$	0	$8td_1$
ξ_2		$2s_2k$	0	0	0	0	$-4s_{12}ff'$	$8td_2$
ξ_3			$4ns_1r_1$	$4s_{12}ff'$	0	$8td_1$	0	$8tf''[(1+f)r_3 - f'u_1]$
ξ_4				$4ns_2r_2$	0	0	$8td_2$	$8tf''[(1+f)r_2 - f'u_2]$
ξ_5					$4s_{12}(1+3f)$	$-8td_2$	$-8td_1$	$16td_{12}f''$
ξ_6						$8ns_{12}f''(s_2zf'^2+2f)$	$16td_{12}f''$	$f''^3[1-(2+f)r_4]$
ξ_7							$8ns_{12}f''(s_1zf'^2+2f)$	$f''^3[1-(2+f)r_3]$
ξ_8								$16Es_{12}f'f''^4$

The following abbreviations are used in the table:

$k = 1/f''$ $\quad n = f'f''$ $\quad z = 1-2f-f^2$

$s_1 = p_1q_1$ $\quad r_1 = f + p_1q_1f'^2$ $\quad t = s_{12}ff'f''$

$s_2 = p_2q_2$ $\quad r_2 = f + p_2q_2f'^2$ $\quad C_1 = 3f^3 + 15f^2 + 17f + 1$

$s_{12} = s_1s_2$ $\quad r_3 = f + 2p_1q_1f'^2$ $\quad C_2 = f^3 + 7f^2 + 7f - 3$

$d_1 = q_1 - p_1$ $\quad r_4 = f + 2p_2q_2f'^2$ $\quad E = s_{12}f'^4C_1 + ff'^2C_2(s_1+s_2) + 4f^4 + ff'^3$

$d_2 = q_2 - p_2$ $\quad u_1 = 1 - 2p_1q_1f'$

$d_{12} = d_1d_2$ $\quad u_2 = 1 - 2p_2q_2f'$

The additive component 1 of the total genetic variance is then given by,

$$\begin{aligned}\text{Var (Add.1)} &= \alpha_1 \text{ Cov}(X, \xi_1) \\ &= c^{11}[\text{Cov}(X, \xi_1)]^2 + c^{12} \text{ Cov}(X, \xi_1) \text{ Cov}(X, \xi_2) + \ldots \\ &\quad + c^{18} \text{ Cov}(X, \xi_1) \text{ Cov}(X, \xi_8). \end{aligned} \tag{7}$$

We note that, were we successful in locating an ideal set of mutually uncorrelated ξ_i's, we would have, Var (Add.1) $= c^{11}$ [Cov $(X, \xi_1)]^2$ since c^{ij} would then $= 0$.

When we consider $D \neq 0$, f $\neq 0$, there are two possibilities (i) to keep the basic set of variables, ξ_i, and (ii) to use A&O's corrections for D and use the set of variables, ψ_i.

All these cases were analysed in a numerical example (table 2) and both the above methods were used for the case, $D = 0{\cdot}1$, f $= 0$ for illustration (cases III and IV). In case V, A&O's corrections for D were used.

3. Discussion

The salient feature of the Fisherian approach in a single-gene system is that the regression variables θ and ω are independent providing the values of additive and dominance variance both under random mating and inbreeding. Such an extension was feasible in a two-gene system only under random mating with independent genes ($D \equiv$ f $\equiv 0$). In other cases, it was not possible to find a set of regression variables which are mutually uncorrelated. Considering the expression (7) for Var (Add.1), it would be evident that, when the sum of the second to eighth terms becomes negative and its absolute value exceeds the value of the first term, it is possible to get a negative estimate for Var (Add.1) (table 2; see also Ewens 1979). Table 2 column *a* gives the components of total genetic variance as defined in this paper. Column *b* gives the 'ideal' values of the components, were the variables, ξ_i, mutually uncorrelated. It is found, as expected, that the latter values, though always positive, are biased and do not add to the actual total variance given under column *a*. When $D = 0{\cdot}1$, they overestimate, in our example, many components and total variance. A comparison of values given under *a* and *b* would provide an idea as to what extent the mutual correlations among ξ_i's account for the differences.

The only alternative is then to derive expressions for additive genetic variance of any one gene adjusted for the association of that gene with others. In other words, adjustment is done in a hierarchical manner like additive variance (gene 1), additive variance (gene 2) adjusted for its association with gene 1, additive variance (gene 3) adjusted for its association with genes 1 and 2, and so on. This approach in essence has been followed by many workers. In this process, the elegance of defining, in general, additive genetic variance (gene j) as equal to 'average excess' × 'average effect' × variance of the variable defining additive variance of gene j, is lost.

The results reported here and in A&O clearly conclude that, in genetic systems governed by two (or more) genes, arbitrary inbreeding and linkage, (a) it is not possible to express a quantitative character X as an exact regression equation in such variables that account for the various components of the total genetic variance, (b) the definition of additive (and, likewise, other) components of

Table 2. Specific examples of partitioning total genetic variance.

	$p_1 = 0{\cdot}6$;	$p_2 = 0{\cdot}3$	
		X values	
	A_2A_2	A_2a_2	a_2a_2
A_1A_1	15	5	30
A_1a_1	10	25	2
a_1a_1	18	7	20

Components	I	II		III		IV*		V	
	D = f = 0	D = 0; f = 0·5		D = 0·1; f = 0 (basic set)		D = 0·1; f = 0		D = 0·1; f = 0·5	
		a	*b*	*a*	*b*	*a*	*b*	*a*	*b*
Add.1	3·446	6·742	6·153	− 1·961	1·124	1·124	1·124	1·194	0·197
Add.2	0·141	8·458	7·542	− 0·349	0·864	− 0·527	1·078	6·524	4·410
Dom.1	3·460	15·661	11·388	3·150	2·867	2·867	2·867	13·228	7·942
Dom.2	0·143	3·951	1·946	0·324	0·735	0·451	0·765	1·725	0·402
$A \times A$	15·259	9·770	9·493	10·503	7·194	13·532	20·826	9·237	11·944
$A \times D$	7·642	3·171	4·549	8·730	10·049	9·848	10·610	4·234	9·188
$D \times A$	45·777	19·719	27·103	47·329	55·967	41·188	80·248	21·199	41·786
$D \times D$	46·294	29·570	21·085	41·832	56·555	40·625	123·882	28·672	20·673
Total	122·162	97·042	89·259	109·108	135·355	109·108	241·400	86·013	96·542

a–Components defined by regression equation (see text); *b*–$[\mathrm{Cov}(X, \xi_i)]^2/\mathrm{Var}(\xi_i)$; *–after removing the correlation between ω_1 and ω_2 by defining $\omega_1' = \omega_1$; $\omega_2' = \omega_2 - [s_2 f(1+f)/r_1]\omega_1$, (cf: A & O).

variance do differ in essence from those extended from a random mating single-gene system, (c) these components of variance do contain contributions due to their association with other components and (d) it is not possible to separate and estimate them. Since covariance of relatives is expressed ultimately in terms of the components of genotypic variance, in practice, it is a moot question to ask what would be the ideal choice of mating and selection schemes in situations governed by more than one gene. The results usually obtained in practical situations by conceiving quantitative characters as being governed by a number of additive and independent genes, as a close approximation to reality, can, in the light of these results, be really far from it.

Acknowledgements

The author expresses his gratitude to Dr A R G Owen, formerly of University of Toronto, Canada, for providing ideas for this paper, Drs N Takahata and M Kimura of the National Institute of Genetics, Japan, for critical comments and Dr A Bandyopadhyay of the National Academy of Agricultural Research Management, Hyderabad, for his immense help in the preparation of this paper.

References

Arunachalam V and Owen A R G 1971 *Polymorphisms with linked loci* (London: Chapman and Hall)

Crow J F and Kimura M 1970 *An introduction to population genetics theory* (New York: Harper and Row)

Ewens W J 1979 *Mathematical population genetics* (New York: Springer-Verlag)

Falconer D S 1964 *Introduction to quantitative genetics* (Edinburgh: Oliver and Boyd)

Fisher R A 1941 Average excess and average effect of a gene substitution. *Ann. Eugen.* 11: 53–63

Hayman B I 1954 The theory and analysis of diallel crosses. *Genetics* 39: 789–809

Li C C 1955 *Population genetics* (Chicago: University Press)

Mather K and Jinks J L 1971 *Biometrical genetics* (London: Chapman and Hall)

Weir B S and Cockerham C C 1977 Two locus theory in quantitative genetics. In *Proc. Int. Conf. Quant. Genet.* (eds) E Pollak, O Kempthorne and T B Bailer Jr. (Iowa: Iowa State University Press) p. 247