

PARTITIONING OF VARIANCE UNDER FULL-SIB MATING

BY M. RAJAGOPALAN

Indian Council of Agricultural Research, New Delhi

WHEN related individuals are mated the resulting progeny is said to be inbred. In dioecious plants and higher animals full brother-sister mating ensures the closest form of inbreeding. Under any system of inbreeding the whole population divides itself into different inbred lines. We will be concerned with the inherited make-up of the individuals and the variability present among them within these lines.

Mather (1949) described the method of partitioning the heritable portion of the continuous variation into fixable and unfixable parts, the former depending on "the difference in average character expression associated with the two homozygotes for each of the gene pairs involved" and the latter arising from "the difference between the expression of heterozygotes and the average of the two corresponding homozygotes". In his terminology these two components are denoted by D and H respectively. Bateman and Mather (1951) obtained the coefficients of D and H in the expression for variance of progeny in the case of selfing for successive generations and for different measures of variation.

For various systems of inbreeding (both with and without the presence of linkage) a general method of obtaining these coefficients was described by Nelder (1952). Following this procedure of Nelder, Hayman (1953) obtained these coefficients for full-sib mating started by a cross between two true breeding lines. Nelder's procedure is given briefly below.

"Individuals of generation S_k which have a common ancestral pair in S_l ($l < k$) but not in S_{l+1} form an l -group. V_{ls_k} is that portion of the heritable variation of generation S_k which is given by the variance of the means of $(k - l + 1)$ groups taken round the means of $(k - l)$ groups and averaged over the latter. Hence V_{ls_k} can be expressed as the difference of the total variance of $(k - l + 1)$ group means and $(k - l)$ group means in S_k . The variance of $(k - l)$ group means is computed by using the mean of S_{k-1} as variate while frequencies are those of S_l ."

In this paper we consider starting with a random mating population, the make-up of the population obtained by continued full-sib mating in n -generations, the partitioning of the heritable portion of the variation within inbred lines into additive genetic and dominance variations and further into the parts contributed by the variations between non-sibs, half-sibs and full-sibs among the progeny.

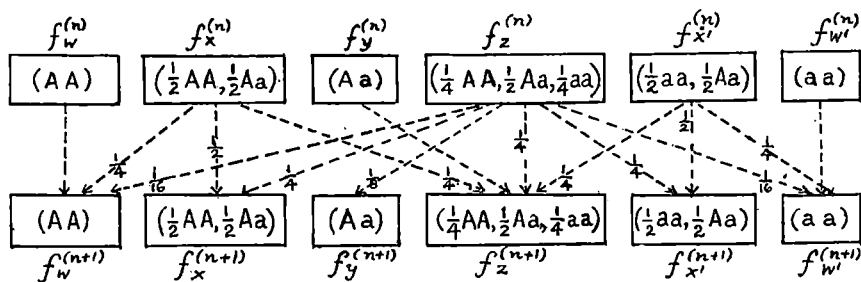
2. Consider two alleles A and a of a gene and the possible genotypes AA , Aa and aa . Let the frequencies of these genotypes in the population be ϕ_2 , ϕ_1 and ϕ_0 respectively, where $\phi_2 + \phi_1 + \phi_0 = 1$. Assuming that both the sexes are equally frequent the following pattern as given in Table I, will result by mating individuals at random. There will be six mating types and corresponding to each there will be a full-sib family in the first generation.

TABLE I

Frequency	Mating type	Full-sib family of the first generation	Symbol of family type
ϕ_2^2	$AA \times AA$	(AA)	W
$2\phi_2\phi_1$	$AA \times Aa$	$(\frac{1}{2}AA, \frac{1}{2}Aa)$	X
$2\phi_2\phi_0$	$AA \times aa$	(Aa)	Y
ϕ_1^2	$Aa \times Aa$	$(\frac{1}{4}AA, \frac{1}{2}Aa, \frac{1}{4}aa)$	Z
$2\phi_0\phi_1$	$aa \times Aa$	$(\frac{1}{2}aa, \frac{1}{2}Aa)$	X'
ϕ_0^2	$aa \times aa$	(aa)	W'

Let the different types of full-sib families be denoted by W , X , Y , ..., etc., as shown in the table. Now, within each of the first generation families full-sib mating is continued in successive generations. Within each of these full-sibships the mating is at random and both the sexes are equally frequent.

3. Let $f_w^{(n)}$, $f_x^{(n)}$, $f_y^{(n)}$, ..., etc., denote the frequencies of the types of families W , X , Y , ..., etc., respectively in the n -th generation. Following the procedure described by Haldane (1937) we obtain the relationship of the frequencies in the n -th generation to those in the $(n+1)$ -th generation by means of the diagram given below:—



We obtain,

$$\left. \begin{aligned}
 f_w^{(n+1)} &= f_w^{(n)} + \frac{1}{4} f_x^{(n)} + \frac{1}{16} f_z^{(n)} \\
 f_x^{(n+1)} &= \frac{1}{2} f_x^{(n)} + \frac{1}{4} f_z^{(n)} \\
 f_y^{(n+1)} &= \frac{1}{8} f_z^{(n)} \\
 f_z^{(n+1)} &= \frac{1}{4} f_x^{(n)} + f_y^{(n)} + \frac{1}{4} f_z^{(n)} + \frac{1}{4} f_{x'}^{(n)} \\
 f_{x'}^{(n+1)} &= \frac{1}{4} f_z^{(n)} + \frac{1}{2} f_{x'}^{(n)} \\
 f_{w'}^{(n+1)} &= \frac{1}{8} f_z^{(n)} + \frac{1}{4} f_{x'}^{(n)} + f_{w'}^{(n)}
 \end{aligned} \right\} (3.1)$$

It can be seen that the middle four equations are independent of $f_w^{(n)}$ and $f_{w'}^{(n)}$. It is also obvious that the family types W and W' generate progeny which are all homozygous and the variation within these families is zero. Therefore, we will not consider the recurrence relations, etc., for these.

4. We now proceed to obtain the recurrence relations for $f_z^{(n)}$ and $f_{x'}^{(n)}$.

From (3.1)

$$\begin{aligned}
 f_z^{(n+1)} &= \frac{1}{4} (f_x^{(n)} + f_{x'}^{(n)}) + f_y^{(n)} + \frac{1}{4} f_z^{(n)} \\
 &= \frac{1}{4} (f_x^{(n)} + f_{x'}^{(n)}) + \frac{1}{8} f_z^{(n-1)} + \frac{1}{4} f_z^{(n)}.
 \end{aligned}$$

From second and fifth equations of (3.1), we obtain

$$f_x^{(n)} + f_{x'}^{(n)} = \frac{1}{2} (f_x^{(n-1)} + f_{x'}^{(n-1)}) + \frac{1}{2} f_z^{(n-1)}.$$

Therefore,

$$\begin{aligned}
 f_z^{(n+1)} &= \frac{1}{8} (f_x^{(n-1)} + f_{x'}^{(n-1)}) + \frac{1}{4} f_z^{(n-1)} + \frac{1}{4} f_z^{(n)} \\
 &= \frac{1}{2} (f_z^{(n)} - \frac{1}{8} f_z^{(n-2)}) + \frac{1}{4} f_z^{(n-1)} + \frac{1}{4} f_z^{(n-1)} + \frac{1}{4} f_z^{(n)}.
 \end{aligned}$$

Hence,

$$f_z^{(n+1)} - \frac{3}{4}f_z^{(n)} - \frac{1}{8}f_z^{(n-1)} + \frac{1}{16}f_z^{(n-2)} = 0 \quad (4.1)$$

It is seen from (4.1) that the solution for $f_z^{(n)}$ is given by

$$f_z^{(n)} = L_1\mu_1^{(n)} + L_2\mu_2^{(n)} + L_3\mu_3^{(n)} \quad (4.2)$$

where μ_1, μ_2 and μ_3 are the roots of the cubic $\mu^3 - \frac{3}{4}\mu^2 - \frac{1}{8}\mu + \frac{1}{16} = 0$ and L_1, L_2 and L_3 are arbitrary constants to be determined from the initial conditions. The roots of the cubic are $\frac{1}{4}, \frac{1}{4}(1 + \sqrt{5})$ and $\frac{1}{4}(1 - \sqrt{5})$. By symmetry a similar recurrence relation and solution hold for $f_y^{(n)}$ also.

By substituting $(f_x^{(n+1)} - \frac{1}{2}f_x^{(n)})$ in the recurrence relation for $f_z^{(n)}$ we obtain the recurrence relation for $f_x^{(n)}$. It is given as,

$$f_x^{(n+3)} - \frac{5}{4}f_x^{(n+2)} + \frac{1}{4}f_x^{(n+1)} + \frac{1}{8}f_x^{(n)} - \frac{1}{32}f_x^{(n-1)} = 0 \quad (4.3)$$

and the solution is given by

$$f_x^{(n)} = K_1\lambda^n + K_2\mu_1^n + K_3\mu_2^n + K_4\mu_3^n \quad (4.4)$$

where λ, μ_1, μ_2 and μ_3 are the roots of the quartic $\lambda^4 - \frac{5}{4}\lambda^3 + \frac{1}{4}\lambda^2 + \frac{1}{8}\lambda - \frac{1}{32} = 0$ and K_1, K_2, K_3 and K_4 are arbitrary constants. The roots of the quartic are $\frac{1}{2}, \frac{1}{4}, \frac{1}{4}(1 + \sqrt{5})$ and $\frac{1}{4}(1 - \sqrt{5})$.

As the full-sib mating is continued within each of the first generation families mentioned above, we will be concerned with the frequencies of the different types of full-sib families in the n -th generation within each of these first generation families. Let $f_j^{(n)}$ denote the frequency of the j -th type of family in the n -th generation within the i -th type of first generation family, where $i, j = (X, Y, X', Z)$. The expressions for these frequencies in terms of the roots of the equations were obtained from 4.2 and 4.4 by making use of the initial conditions (namely, the frequencies up to four generations) and are given in the Table II.

5. So far we have discussed about the frequencies of different types of families in the n -th generation and now we proceed to obtain the variation between the progeny within families resulting from full-sib mating successively for n generations. This total variance will be further divided into the variations contributed by non-sibs, half-sibs and full-sibs among the progeny. These separate causes will also be expressed in terms of the additive genetic and dominance variations.

In obtaining these expressions we make use of the lemma described by Comstock and Robinson (1948).

TABLE II

Type of first generation family	Type of family in the n -th generation			
	X	Y	Z	X'
	Frequency			
X	$\lambda^n + \frac{1}{5} \mu_1^{n+1} + \frac{2}{5} (\mu_2^{n+1} + \mu_3^{n+1})$	$-\frac{2}{5} \mu_1^n + \frac{1}{5} \times (\mu_2^{n-1} + \mu_3^{n-1})$	$-\frac{4}{5} \mu_1^n + \frac{2}{5} \times (\mu_2^n + \mu_3^n)$	$-\lambda^n + \frac{1}{5} \mu_1^{n+1} + \frac{2}{5} (\mu_2^{n+1} + \mu_3^{n+1})$
Y	$-\frac{4}{5} \mu_1^{n-1} + \frac{2}{5} (\mu_2^{n-1} + \mu_3^{n-1})$	$\frac{8}{5} \mu_1^n - \frac{2}{5} \times (\mu_2^n + \mu_3^n) + \frac{2}{5} (\mu_2^{n-1} + \mu_3^{n-1})$	$\frac{1}{5} \mu_1^n + \frac{8}{5} \times (\mu_2^n + \mu_3^n) - \frac{4}{5} (\mu_2^{n-1} + \mu_3^{n-1})$	$-\frac{4}{5} \mu_1^{n-1} + \frac{2}{5} \times (\mu_2^{n-1} + \mu_3^{n-1})$
Z	$-\frac{4}{5} \mu_1^n + \frac{2}{5} (\mu_2^n + \mu_3^n)$	$\frac{2}{5} \mu_1^n + \frac{1}{5} \times (\mu_2^n + \mu_3^n) - \frac{1}{5} (\mu_2^{n-1} + \mu_3^{n-1})$	$\frac{4}{5} \mu_1^n + \frac{2}{5} \times (\mu_2^n + \mu_3^n)$	$-\frac{4}{5} \mu_1^n + \frac{2}{5} \times (\mu_2^n + \mu_3^n)$
X'	$-\lambda^n + \frac{1}{5} \mu_1^{n+1} + \frac{2}{5} (\mu_2^{n+1} + \mu_3^{n+1})$	$-\frac{2}{5} \mu_1^n + \frac{1}{5} \times (\mu_2^{n-1} + \mu_3^{n-1})$	$-\frac{4}{5} \mu_1^n + \frac{2}{5} \times (\mu_2^n + \mu_3^n)$	$\lambda^n + \frac{1}{5} \mu_1^{n+1} + \frac{2}{5} (\mu_2^{n+1} + \mu_3^{n+1})$

Lemma.—Let π_2, π_1 and π_0 be the frequencies of the genotypes AA, Aa and aa respectively and let the deviations of the phenotypes of the same from the mean phenotype of AA and aa be $d, h, -d$ respectively. Then the additive genetic variation is given by

$$\sigma_a^2 = \theta \left(d - \frac{\phi h}{\theta} \right)^2 \tag{5.1}$$

and the dominance variation is given by

$$\sigma_d^2 = \left(\psi - \frac{\phi^2}{\theta} \right) h^2 \tag{5.2}$$

where

$$\theta = (\pi_2 + \pi_0) - (\pi_2 - \pi_0)^2$$

$$\phi = \pi_1 (\pi_2 - \pi_0)$$

and

$$\psi = \pi_1 (\pi_2 + \pi_0)$$

Noting that the additive genetic variance is that portion of the variance of genetic effects explained by regression on the number of A genes in the genotype and the variance due to dominance as the variation of deviations of genetic effects from the regression, the proof of the lemma follows immediately.

In populations we generally encounter, for example, poultry, sheep, etc., the family classification is a hierarchical one where a male is mated to a set of females and each mating producing a certain number of offspring. The variation between the progeny thus obtained can be expressed as that between non-sibs, half-sibs and full-sibs among them. The first part, namely, the variation between non-sib families is obtained by the variation between the progeny means of males, the second part between the progeny means of females allotted to each male and pooling over all males and the third part from the variation between the progeny resulting from each mating pair and pooling over all such mating pairs.

In Table III, the frequency of genotypes of the progeny obtained by mating at random within each type of family, the additive genetic and dominance variations within each and further the variation between the progeny expressed in terms of the variation between non-sibs, half-sibs and full-sibs among them are presented.

TABLE III

Type of family	Frequency of genotypes in progeny			σ_a^2	σ_d^2	Between non-sibs	Between half-sibs	Between full-sibs
	AA	Aa	aa					
X	$\frac{9}{16}$	$\frac{6}{16}$	$\frac{1}{16}$	$\frac{3}{8} \left(d - \frac{h}{2} \right)^2$	$\frac{9}{64} h^2$	$\frac{1}{16} \left(d - \frac{h}{2} \right)^2$	$\frac{1}{16} \left(d - \frac{h}{2} \right)^2 + \frac{h^2}{64}$	$\frac{1}{4} \left(d - \frac{h}{2} \right)^2 + \frac{h^2}{8}$
Y	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2} d^2$	$\frac{1}{4} h^2$	$\frac{1}{2} d^2 + \frac{1}{4} h^2$
Z	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2} d^2$	$\frac{1}{4} h^2$	$\frac{1}{8} d^2$	$\frac{1}{8} d^2 + \frac{1}{16} h^2$	$\frac{1}{4} d^2 + \frac{3}{16} h^2$
X'	$\frac{1}{16}$	$\frac{6}{16}$	$\frac{9}{16}$	$\frac{3}{8} \left(d + \frac{h}{2} \right)^2$	$\frac{9}{64} h^2$	$\frac{1}{16} \left(d + \frac{h}{2} \right)^2$	$\frac{1}{16} \left(d + \frac{h}{2} \right)^2 + \frac{h^2}{64}$	$\frac{1}{4} \left(d + \frac{h}{2} \right)^2 + \frac{h^2}{8}$

For any first generation family i ($= X, Y, Z$ and X'), the different parts of variation among the progeny after n successive generations of

full-sib mating are obtained by multiplying the corresponding expressions for each type of family with the appropriate frequencies, namely, ${}_i f_x^{(n)}$, ${}_i f_y^{(n)}$, ${}_i f_z^{(n)}$ and ${}_i f_x'^{(n)}$ and adding them up.

The additive genetic variation and the dominance variance for the i -th first generation family are as given below:—

$$\begin{aligned} {}_i \sigma_a^{2(n)} &= \left\{ \frac{3}{8} ({}_i f_x^{(n)} + {}_i f_x'^{(n)}) + \frac{1}{2} ({}_i f_y^{(n)} + {}_i f_z^{(n)}) \right\} d^2 \\ &\quad + \frac{3}{8} \{ {}_i f_x^{(n)} - {}_i f_x'^{(n)} \} dh + \frac{3}{8} ({}_i f_x^{(n)} + {}_i f_x'^{(n)}) h^2 \\ {}_i \sigma_d^{2(n)} &= \left\{ \frac{9}{8} ({}_i f_x^{(n)} + {}_i f_x'^{(n)}) + \frac{1}{4} ({}_i f_y^{(n)} + {}_i f_z^{(n)}) \right\} h^2. \end{aligned}$$

The variation between the progeny expressed in terms of the constituent parts for the i -th first generation family are as follows:—

Variation between	Expression for the variance
Non-sibs	$\frac{1}{6} {}_i \sigma_a^{2(n)} - \frac{1}{12} \epsilon_i^{(n)} d^2$
Half-sibs	$\frac{1}{6} {}_i \sigma_a^{2(n)} - \frac{1}{12} \epsilon_i^{(n)} d^2 + \frac{1}{9} {}_i \sigma_d^{2(n)} - \frac{1}{36} \eta_i^{(n)} h^2$
Full-sibs	$\frac{2}{3} {}_i \sigma_a^{2(n)} + \frac{1}{6} \epsilon_i^{(n)} d^2 + \frac{8}{9} {}_i \sigma_d^{2(n)} + \frac{1}{36} \eta_i^{(n)} h^2$
where,	$\epsilon_i^{(n)} = {}_i f_y^{(n)} - \frac{1}{2} {}_i f_z^{(n)}$
and	$\eta_i^{(n)} = {}_i f_y^{(n)} - \frac{5}{4} {}_i f_z^{(n)}$.

These expressions for the entire population are obtained by multiplying the corresponding expressions for each of the first generation families X, Y, Z and X' with their respective frequencies $2\phi_2 \phi_1$, $2\phi_2 \phi_0$, ϕ_1^2 and $2\phi_0 \phi_1$ and adding them up.

6. Let us denote by $\sigma_a^{2(n)}$ and $\sigma_d^{2(n)}$ the additive genetic and dominance variations for the population.

Noting that $({}_i f_x^{(n)} + {}_i f_x'^{(n)})$, ${}_i f_y^{(n)}$ and ${}_i f_z^{(n)}$ have separately the same recurrence relation it follows that the recurrence relation for $\sigma_a^{2(n)}$ is of the form,

$$\sigma_a^{2(n+2)} - \frac{3}{4} \sigma_a^{2(n+1)} - \frac{1}{8} \sigma_a^{2(n)} + \frac{1}{16} \sigma_a^{2(n-1)} = 0.$$

It follows when dominance is not present, i.e., $h = 0$, $\sigma_a^{2(n)}$ has the same recurrence relation as given above.

It can be seen that the coefficient of d^2 in the expression for additive genetic variation given by

$$\sum_i \left\{ \frac{3}{8} ({}_i f_x^{(n)} + {}_i f_x'^{(n)}) + \frac{1}{2} ({}_i f_y^{(n)} + {}_i f_z^{(n)}) \right\}$$

where the summation extends for all the first generation families, is nothing but the frequency of heterozygotes in the population and can be put in the form $2qp(1 - F^{(n+1)})$ where $F^{(n+1)}$ is the inbreeding coefficient for the progeny obtained by full-sib mating for n successive generations and q and p are the frequencies of A and a respectively in the population. From this relation it follows that the recurrence relation for the inbreeding coefficient is of the same form and is given by the well-known result

$$F^{(n+2)} - \frac{3}{4}F^{(n+1)} - \frac{1}{8}F^{(n)} + \frac{1}{16}F^{(n-1)} = 0.$$

For a population mating at random the additive genetic variance is given by $2pqd^2$. Let this be denoted by σ_0^2 . Then the additive genetic variance within inbred lines after n generations of full-sib mating, when there is no dominance is given by

$$\sigma_a^{2(n)} = (1 - F^{(n+1)})\sigma_0^2.$$

7. The above expressions are derived for a single factor. When the character concerned is governed by several independent factors and epistatic variation is taken to be absent, the corresponding expressions hold except for individual values being replaced by summation over the factors involved, as for example $S(d^2)$ for d^2 .

8. DISCUSSION

For non-inbred populations under random mating the different components of genetic variance are as follows:—

Between non-sibs	= $\frac{1}{4} \times$ additive genetic variance
Between half-sibs	= $\frac{1}{4} \times$ additive genetic variance + $\frac{1}{4} \times$ dominance variance
Between full-sibs	= $\frac{1}{2} \times$ additive genetic variance + $\frac{3}{4} \times$ dominance variance.

In the case of full-sib mating, however, it can be seen from the results in Section 5, that the different components contain in addition, the terms involving d^2 and h^2 which render the utilization of the sib-analysis for estimation of components of variance difficult. It is seen that the estimation is possible only when (i) $\epsilon^{(n)}$ vanishes or (ii) there is no dominance. The first can be shown to be true only when $n = 1$, viz., if full-sib mating is practised in the first generation only and is followed by random mating.

In the present study both linkage and epistacy have been ignored. Extension to linked factors can be made on the same lines given above.

A preliminary investigation seems to show that for the case when epistacy is present it may be possible to elicit information only when special types of designs for mating are adopted such as diallel matings.

In this paper, only one of the important systems of inbreeding, viz., full-sib mating has been tackled. Other important systems such as parent-offspring mating and half-sib mating, etc., are being investigated.

9. SUMMARY

The expressions for the additive genetic and dominance variations and the variation contributed by non-sibs, half-sibs and full-sibs among the progeny to the total variation between the progeny within inbred lines obtained by continued full-sib mating for n generations are derived.

10. My grateful thanks are due to Mr. V. N. Amble, Statistician, Indian Council of Agricultural Research, for his valuable suggestions and to Dr. V. G. Panse, Statistical Adviser, Indian Council of Agricultural Research, for the encouragement he has given me.

REFERENCES

1. Haldane, J. B. S. .. "Some theoretical results of continued brother-sister mating," *Journal of Genetics*, 1937, **34**, 265-74.
2. Comstock, R. E. and Robinson, H. F. "The components of genetic variance in populations of biparental progenies and their use in estimating the average degree of dominance," *Biometrics*, 1948, **4** (3).
3. Mather, K. .. *Biometrical Genetics*, Methuen & Co., London, 1949.
4. Bateman, A. J. and Mather, K. ... "The progress of inbreeding in barley," *Heredity*, 1951, **5**, 324.
5. Nelder, J. A. "Some genotypic frequencies and variance components occurring in biometrical genetics," *Ibid.*, 1952, **6**, 387-94.
6. Haakman, B. I. .. "Components of variation under sib-mating," *Ibid.*, 1953, **6**.