ESTIMATION OF THE EFFECTIVE NUMBER OF SEGREGATING FACTORS IN QUANTITATIVE INHERITANCE

By T. R. Puri

Indian Council of Agricultural Research

1.1. One of the most important problems in plant or animal breeding is the estimation of the limit to the genetic advance that can be achieved and the speed of advance under a particular type of breeding and with a given degree of selection. Both these are known to depend upon the number of segregating factors which govern the inheritance of the particular character under study. Hence the importance of the estimation of the number of factors.

Most of the characters of economic importance are polygenic and since in these characters the factors cannot be followed individually we have to use biometrical quantities relating to the population in successive generations in order to estimate the number of factors indirectly. The way of dealing with the problem is to make some assumptions regarding the distribution of the magnitudes of the effects of the factors or that of the magnitude of their dominance effects and to deduce formulæ connecting the number of factors with such biometrical constants as can be estimated from measurements on samples of successive generations. The validity of the assumptions will have to be verified by drawing such inferences from them as can be tested by means of experimental evidence. Wright (1934), Goodwin (1944) and others have worked out estimates of the number of factors on the assumption that the magnitude of the effects of all the factors is the same and further that allelomorphs are distributed isodirectionally. Mathur (1952) has found other estimates replacing the assumption of equality by that of symmetry in distribution. But his estimates are given by the roots of certain equations between which however he has not been able to discriminate. Panse (1940) derived an estimate assuming the equality of a certain function of magnitude of effects and of dominance effects. In the present investigation an attempt has been made (i) to discriminate between the roots given by Mathur and (ii) to find out a new estimate which assumes only the symmetry of the effects about their own mean.

1.2. It must be mentioned at the outset that all the estimates given so far, including the one given in this paper, ignore the presence of linkage between the factors. We have no means of separating the effects of variation of all the pairs of genes which are segregating, so that a change in some of them may be partly swamped by lack of it in others. The extent to which we can push the analysis of polygenic system into the ultimate units which genetics has taught us to recognise, must thus be limited by the conditions under which our observations are made. Thus in an investigation, the ultimate units, which we can hope to reach, are at best a combination of genes. These units may be thought to be so constituted that no recombination occurs within them, while it does occur between these units with 50% frequency. We shall therefore only be estimating such groups of linked genes which may be called effective number of factors governing the inheritance of the particular character. The attempt is worthwhile in so far as the number of such groups is not likely to alter in the course of a few generations and the estimation will therefore lead to an approximate inference regarding the limit and speed of advance under selection which holds for the present.

1.3. The estimate of the number of factors used by Wright (1934), Charles and Goodwin (1943) is $K_1 = A^2/D$, where A is half the difference between parental means, being equal to $S(d_a)$; where d_a and $-d_a$ are the average effects of the genotypes AA and aa on the magnitude of the character in question measured as a deviation from their mean value, and D is equal to $S(d_a^2)$. This process of estimation assumes that

(i) The effects of all the factors are equal, *i.e.*, $d_a = d_b = \cdots\cdots = d_k = d$ and

(ii) The positive and negative allelomorphs are distributed isodirectionally in the parents.

The departure from any of these assumptions leads to an under-estimation of the number of effective factors.

A similar estimate using h increments instead of d increments can be used, where h is the magnitude of dominance effect. It is given by $K_1' = B^2/H$, where $B = S(h_a)$ is the departure of F_1 means from the mid-parental value and $H = S(h_a^2)$. This estimate, like the previous one, assumes the equality of h's and that all h's reinforce one another. This will also be an under-estimate when either of the assumptions is violated. Further, in selfed progenies the variance of estimate of H is larger than the variance of the estimate of D; and therefore the previous estimate is likely to give better estimates than this one.

1.4. Panse (1940) has given another estimate which overcomes the difficulty of under-estimation due to incomplete concentration or the incomplete reinforcement of the effective factors. This is given by the ratio which the square of the heritable portion of the mean variance of all the selfed progenies bears to the heritable portion of the variance of the variances of the families in F_3. This is $K_2 = ({}_H\bar{V}_{F_3})^2 / {}_H V_{V_{F_3}}$, where ${}_H\bar{V}_{F_3}$ is the heritable portion of the mean variance of all the F_3 families and ${}_H V_{V_{F_3}}$ is the heritable portion of the variance of the variances of F_3.

This estimate assumes the equality

$$(\tfrac{1}{2}d_a^2 + \tfrac{1}{4}h_a^2) = (\tfrac{1}{2}d_b^2 + \tfrac{1}{4}h_b^2) = \ldots\ldots\ldots = (\tfrac{1}{2}d^2 + \tfrac{1}{4}h^2).$$

The estimate K_2 is superior to K_1 in as much as it is not subject to reduction due to incomplete concentration or incomplete reinforcement of the effective factors. But the variance of the quantity $(\tfrac{1}{2}d_a^2 + \tfrac{1}{4}h_a^2)$ is expected to be much more than of either d_a or h_a, if the assumption of equality is not satisfied. This deflates the value of K_2 much more than K_1 and that is why we sometimes get K_2 less than K_1.

1.5. Mathur has replaced the assumption of equality of the magnitude of effects of different factors or their dominance effects by a less stringent one that is of the symmetry of the distribution of the magnitudes of the effects or the dominance effects and has found out two more estimates.

If h denotes mean dominance of the effective factors and V_β is the variance of h's and the distribution of the dominance effects is assumed to be symmetrical, B, H and G given by Σh, Σh^2 and Σh^3 respectively reduce to

$$B = Kh \quad \text{(i)}$$

$$H = Kh^2 + KV_\beta \quad \text{(ii)}$$

$$G = Kh^3 + 3KhV_\beta \quad \text{(iii)}$$

and the first estimate is given by the roots of

$$K_3 = \frac{3}{2}\,\frac{BH}{G}\left\{1 \pm \left(1 - \frac{8BG}{9H^2}\right)^{\frac{1}{2}}\right\} \quad \text{(iv)}$$

If d is the mean and V_α the variance of the distribution of d's and if we assume the symmetry in the distribution d's and also that there is no kurtosis in the distribution, *i.e.*, $\mu_{3\alpha} = 0$ and $\mu_{4\alpha} = 3V_\alpha^2$; then A, D and Δ, given by Σd, Σd^2, and Σd^4 will reduce to

$$A = Kd \tag{v}$$

$$D = Kd^2 + KV_\alpha \tag{vi}$$

$$\Delta = K(d^4 + 6d^2V_\alpha + 3V_\alpha^2) \tag{vii}$$

and the second estimate is given by one of the roots of

$$\Delta K^3 - 3D^2K^2 + 2A^4 = 0 \tag{viii}$$

Equation (iv) gives two values of K_3 which are both positive and Equation (viii) gives three roots of K_4 two of them being positive and one negative. Only one root is to be taken in any particular case as giving the estimate of the effective number of factors. The choice of one solution or the other depends on the intrinsic genetic make-up of the character under study. Such a discrimination had remained to be attempted. In the present paper a function of the known quantities has been found out whose values show, which of the roots gives the real value of the effective number of segregating factors.

2.1. K_3 is given by one of the roots of Equation (iv). We want such a functional relation of quantities which can be estimated from experimental data as will discriminate between the roots.

Evidently one of the roots is greater than $3BH/2G$ and the other is smaller than it, so that the smaller root is to be taken when the effective number of factors is less than $3BH/2G$.

Substituting the values of B, H and G as given in (i), (ii) and (iii) in $3BH/2G$ we find that the smaller root will give the estimate of the effective number of factors when $h^2 > 3V_\beta$. Now h and V_β cannot be estimated directly and so we want to find out such a functional relation which is positive only when $h^2 > 3V_\beta$ and is negative otherwise. For this we consider

$$\text{“}10G^2 - 9HR\text{”}$$

where

$$R = s(h_a^4)$$

Under the assumption that there is no kurtosis in the distribution of h's, R reduces to

$$K(h^2 + 6h^2V_\beta + 3V_\beta^2) \tag{ix}$$

Substituting the values of G, H and R and simplifying we get the value of the functional relation as

$$K^2(h^2 - 3V_\beta)(h^4 + 9V_\beta^2).$$

The sign of this depends upon $(h^2 - 3V_\beta)$, the sign of the other factor being positive. If the value of the functional relation $(10G^2 - 9HR)$ turns out to be positive, the lower root should be taken and if it is negative the upper root should be taken. If the value is zero, $h^2 = 3V_\beta$ and in this case the two roots of K_3 are equal. The quantities H, G and R can be estimated from the estimated statistical relations by the method of least squares.

This requires the use of fourth degree statistics and the equations leading to their estimation have been given by Mathur (1952).

2·2. Turning next to the estimate K_4, its value is given by the Equation (viii) which is a cubic equation, the sign of the constant term being positive. Therefore one of the roots must be negative in which we are not interested. The other two roots are positive as we can easily see from the Descarte's Rule of Signs. So we want to discriminate between these two positive roots.

It can be shown that of the two positive roots one is less than $2D^2/\Delta$ and the other is greater than $2D^2/\Delta$ because the function is positive for $K = 0$ and $K = \infty$ and for $K = 2D^2/\Delta$ after substituting the values of Δ, D and A the function turns out to be equal to

$$-(x^6 - 12x^4 + 4x^3 + 21x^2 + 12x + 2)$$

where

$$x = \frac{d^2}{V_a}$$

or

$$= -[x - (\sqrt{2} + 1)]^2 [x^4 + 2(\sqrt{2} + 1)x^3 + (6\sqrt{2} - 3)x^2 + 4(2 - \sqrt{2})x + 2(3 - 2\sqrt{2})]$$

which is clearly negative.

Now we will see when we have to take the upper root and when to take the lower one. Evidently if $2D^2/\Delta$ is greater than the effective number of factors, we shall take the lower root. Substituting the values of D and Δ and after a little simplification we find that the lower root should be taken when $d^2 > (\sqrt{2} + 1) V_a$.

As before we want to find out a functional relation which is positive when $d^2 > (\sqrt{2} + 1) V_a$ and is negative otherwise. For this we consider

"$\sqrt{2}\,\Delta^2 - ZD$"

where

$$Z = \Sigma (d_a^6)$$

With the assumption $\mu_6 = 15\mu_2^3$, the usual relation in the case of normality Z reduces to

$$K(d^6 + 15d^4 V_\alpha + 45d^2 V_\alpha^2 + 15V_\alpha^3)$$

Substituting the values of Z, Δ and D and substituting X for d^2/V_α we find after simplification that the functional relation will be positive if

$$(\sqrt{2} - 1)x^4 + 4x^3(3\sqrt{2} - 4) + 6x^2(7\sqrt{2} - 10) \\ - 12x(5 - 3\sqrt{2}) - 3(5 - 3\sqrt{2}) \text{ is positive.}$$

The expression has only one change in sign and therefore cannot have more than one positive root if equated to zero. That one root can be verified to be equal to $(\sqrt{2} + 1)$. So the expression will change sign at $x = (\sqrt{2} + 1)$ and it cannot change sign for any other value of X. We can easily see that it is negative for all values of X which are less than $(\sqrt{2} + 1)$ and is positive for all values of X which are greater than $(\sqrt{2} + 1)$.

So we conclude that the functional relation

$$(ZD - \sqrt{2}\,\Delta^2)$$

will be positive when

$$x = \frac{d^2}{V_\alpha} > \sqrt{2} + 1$$

i.e., when

$$d^2 > (\sqrt{2} + 1)\,V_\alpha.$$

Hence if $(ZD - \sqrt{2}\,\Delta^2)$ is positive, we take the lower root and if it is negative, we take the upper root. Its vanishing is equivalent to both the positive roots being equal in which case the question of discrimination does not arise.

2·3. Another approximate method of discriminating between the roots can be derived as follows which can be used only when both estimates K_1 and K_2 are obtained.

It has been shown by Mather (1949) that

$$V_\alpha = \frac{K_1 - K_2}{rK_2 - K_1}$$

where

$$r = \frac{V_x}{V_\alpha}, \quad \text{and} \quad V_x = V\left(\tfrac{1}{2}d^2 + \tfrac{1}{4}h^2\right)$$

and also that

$$K = \frac{K_1 K_2 (r-1)}{rK_2 - K_1}$$

We get the value of V_α/d^2, where $d^2 = D/K$. r is unknown but its value can be taken as 4 as reported by Mather when V_α is small. Hence we can know if $d^2 > (\sqrt{2}+1)\, V_\alpha$ or not and decide whether to take the upper root or the lower root for K_4.

2.4. In 2.1 and 2.2 we have found functional relations which discriminate between the roots of K_3 and K_4. The parametric values of these relations are not known and they can only be estimated from the sample data. As such the values assigned to them are subject to error. In the next section we shall find the variance of the estimates of these functional relations.

3.1. The relationship discriminating between the roots of K_3 is " $10G^2 - 9HR$". The variance of this (Kendall, 1948) by the theory of large sample approximation is given by

$$\begin{aligned} V(10G^2 - 9HR) = {} & 400G^2 V_G + 81R^2 V_H + 81H^2 V_R \\ & + 162HR \text{ cov. } (HR) - 360\, GR \text{ cov. } (GH) \\ & - 360GH \text{ cov. } (GR). \end{aligned}$$

On the same lines we find that the variance of the relation discriminating the roots of K_4 is given by

$$\begin{aligned} V(\sqrt{2}\,\Delta^2 - ZD) = {} & 8\Delta^2 V_\Delta + D^2 V_Z + Z^2 V_D + 2DZ \text{ cov. } (DZ) \\ & - 4\sqrt{2}\Delta Z \text{ cov. } (D\Delta) - 4\sqrt{2}\,\Delta D \text{ cov. } (Z\Delta) \end{aligned}$$

The variances of these functional relations involve the variances and covariances of the statistics G, H, R, etc., which can be obtained by an extension of the method of least squares indicated by Mather.

Having found the functional relations and their variances, we can apply a test of significance and find out whether the value of this function is significantly positive or negative.

4.1. In finding K_3 and K_4, use of the first moment has been made which is equivalent to the assumption of complete concentration of genes. It has been shown by Mathur that the lower root is still lowered and the upper root is increased if this assumption is not satisfied. The estimates are therefore biased. These estimates are superior to K_1 and K_1' inasmuch as instead of assuming the equality of d's or h's symmetry of their effects is only assumed. To K_2 these estimates are superior inasmuch as unlike K_2 these do not assume the equality of $(\frac{1}{2}d^2 + \frac{1}{4}h^2)$'s, but K_2 does not have the drawback of assuming complete concentration of the allelomorphs in the parents as it is based on the second and fourth moments only.

Following the same lines as that for K_3 and K_4 another estimate may be found out which is not subject to either under-estimation or over-estimation due to the lack of complete concentration, as in this case the first moment is not utilized. Besides, the equation whose roots give the effective-number of factors turns out to be cubic whose two roots are imaginary and only one root is real, which is positive, and therefore the irksome question of discrimination does not arise.

Proceeding similarly as in the case of K_3 and K_4 we can put

$$h_i = h + \beta_i,\ i = 1 \cdots K$$

where

$$\sum_{i=1}^{k} \beta_i = 0.$$

Assuming the symmetry in the distribution of β's and also that there is no kurtosis in the said distribution, *i.e.*,

$$\mu_{3\beta} = 0 \text{ and } \mu_{4\beta} = 3V_\beta^2 \quad \text{and}$$

taking the second, third and fourth moments we get the Equations (ii), (iii) and (ix) as giving the values of H, G and R. Eliminating h and V_β from (ii), (iii) and (ix) and after simplification we get

$$4R^3K^3 + 2G^2(G^2 - 12HR)K^2 + 9H^3(8G^2 - 3HR)K - 27H^6 = 0 \qquad \text{(x)}$$

This is a cubic in K giving three values, one of which must be positive as the last term is negative. Now we shall examine the other two roots.

4.3. The Equation (x) can be written as

$$p_0x^3 + 3p_1x^2 + 3p_2x + p_3 = 0,$$

where

$$p_0 = 4R^3$$

$$p_1 = \tfrac{2}{3}G^2(G^2 - 12HR)$$

$$p_2 = 3H^3(8G^2 - 3HR) \text{ and } p_3 = -27H^6.$$

The equation can be reduced to $x^3 + 3ax + b = 0$ by increasing its roots by p_1/p_0 where

$$a = p_0 p_2 - p_1^2$$
$$= 12H^3R^3 (8G^2 - 3HR) - \{\tfrac{2}{3}G^2 (G^2 - 12HR)\}^2 \qquad \text{(i)}$$
$$b = p_3 p_0^2 - 3p_0 p_1 p_2 + 2p_1^3$$
$$= -432H^6R^6 - 24R^3H^3G^2 (8G^4 - 99G^2HR + 36H^2R^2) + 16/27G^6 (G^6 - 36G^4HR + 432G^2H^2R^2 - 1728 H^3R^3) \qquad \text{(ii)}$$

For all the roots to be real, a and $(4a^3 + b^2)$ should both be negative (Burnside and Panton, 1935).

Z is positive and can be shown to be less than unity by substituting the values of G, H and R.

For a to be negative, it can be shown that

$$Z^4 - 24Z^3 + 144Z^2 - 216Z + 81$$

should be positive which will be so if

$$Z < \cdot 6$$

since the other possibility, *viz.*,

$$Z > 1 \cdot 42$$

is not admissible.

For $(4a^3 + b^2)$ to be negative, after substituting the values of a and b and simplifying, we see that

$$(-36450Z^5 + 918540Z^4 - 6303480 \cdot 75Z^3 + 16966746Z^2 - 17891847\, Z + 6377292)$$

should be negative.

By differentiating this we see that it is a decreasing function of Z for its permissible values and is positive for the highest value, *i.e.*, 0·6. Therefore it is positive for all the permissible values of Z and hence (X) cannot have all its roots real. As the imaginary roots occur in pairs, two of its roots must be imaginary and hence it has only one real positive root, which gives us the effective number of factors.

4.4. This estimate denoted by K_5 is superior to K_1 and K_1' in the sense that it is not subject to reduction by lack of complete concentration and secondly it does not assume the equality of h's but only assumes the symmetry of their distribution which is a less stringent condition.

It is superior to K_2 for not assuming the equality of $(\frac{1}{2}d^2 + \frac{1}{4}h^2)$'s.

To K_3 and K_4, it is superior for not being affected by incomplete concentration. From computational point of view it is preferable to K_3 and K_4 in not requiring any study of functional relations to discriminate between the roots.

We shall obtain the variance of K_5 and examine the efficiencies of the different estimates relatively to one another in the next section.

5.1. *Variance of* K_5.—Let the estimates of K_5, G, R and H be $K_5(1+k)$, $R(1+r)$, $G(1+g)$ and $H(1+h)$ respectively where the expected values of k, r, g and h are zero. Substituting these values in (x) we get

$$\begin{aligned} 4R^3K_5^3(1+r)^3(1+k)^3 &+ 2K_5^2G^2(1+k)^2(1+g)^2 \\ &\{G^2(1+g)^2 - 12HR(1+h)(1+r)\} \\ &+ 9H^3K_5(1+h)^3(1+k)\{8G^2(1+g)^2 \\ &- 3HR(1+h)(1+r) - 27H^6(1+h)^6\} = 0 \quad \text{(xi)} \end{aligned}$$

Subtracting (x) from (xi), omitting the infinitesimals of second and higher order, squaring and taking the expectation, we get

$$\begin{aligned} V(K_5) = (12R^3K_5^2 &+ 4G^4K_5 - 48G^2HRK_5 + 72H^3G^2 \\ &- 27H^4R)^{-2}[9(8K_5^2G^2H - 4K_5^3R^2 + 9H^4K_5)^2 V_R \\ &+ 64(6GHRK_5^2 - K_5^2G^3 - 18H^3K_5G)^2 V_G \\ &+ 36(4K_5^2G^2R - 36H^2G^2K_5 + 18H^3RK_5 \\ &+ 27H^5)^2 V_H + 48(8K_5^2G^2H - 4K_5^3R^2 \\ &+ 9H^4K_5)(6HRGK_5^2 - K_5^2G^3 - 18H^3K_5G) \\ &\times \text{cov.}(RG) + 36(8K_5^2G^2H - 4K_5^3R^2 \\ &+ 9H^4K_5)(4K_5^2G^2R - 36H^2G^2K_5 + 18H^3RK_5 \\ &+ 27H^5)\,\text{cov.}(RH) + 96(6HRGK_5^2 - K_5^2G^3 \\ &- 18H^3K_5G)(4K_5^2G^2R - 36H^2G^2K_5 + 18H^3RK_5 \\ &+ 27H^5)\,\text{cov.}(GH)]. \end{aligned}$$

If we now take KR as equal to H^2 and $G^2 = HR$, which will be so, if the differences of the heterozygotes from the mid-parental values for all the effective factors are equal, we get

$$\begin{aligned} V(K_5) = K_5^2 \Big[9\frac{V_R}{R^2} &+ 64\frac{V_G}{G^2} + 36\frac{V_H}{H^2} \\ &- 48\frac{\text{cov.}(RG)}{RG} + 36\frac{\text{cov.}(HR)}{HR} - 96\frac{\text{cov.}(GH)}{GH} \Big] \end{aligned}$$

5.2. The large sample approximation to the variances of the different estimates are given by the following expressions:

$$V(K_1) = K_1^2 \left[\frac{V_D}{D^2} + 4\frac{V_A}{A^2} - 4\frac{\text{cov.}\,(AD)}{AD}\right]$$

$$V(K_1') = K_1'^2 \left[\frac{V_H}{H^2} + 4\frac{V_B}{B^2} - 4\frac{\text{cov.}\,BH}{BH}\right]$$

$$V(K_2) = K_2^2 \left[\frac{V_D + \frac{1}{4}V_H + \text{cov.}\,DH}{(\frac{1}{2}D + \frac{1}{4}H)^2}\right.$$

$$+ \frac{\frac{1}{16}V_\Delta + \frac{1}{256}V_R + \frac{1}{16}V_J + \frac{1}{32}\text{cov.}\,(\Delta R) + \frac{1}{8}\text{cov.}\,(\Delta J) + \frac{1}{32}\text{cov.}\,(JR)}{\left(\frac{1}{4}\Delta + \frac{1}{16}R + \frac{1}{4}J\right)^2}$$

$$\left. - \frac{\frac{1}{2}\text{cov.}\,(D\Delta) + \frac{1}{8}\text{cov.}\,(DR) + \frac{1}{2}\text{cov.}\,(DJ) + \frac{1}{4}\text{cov.}\,(H\Delta) + \frac{1}{16}\text{cov.}\,(HR) + \frac{1}{4}\text{cov.}\,(HJ)}{(\frac{1}{2}D + \frac{1}{4}H)\left(\frac{1}{4}\Delta + \frac{1}{16}R + \frac{1}{4}J\right)}\right]$$

where

$$J = \Sigma\,(d^2h^2).$$

The expressions for the variances of K_3, K_4 and K_5 become very cumbersome and if we make the assumption that the magnitude of all the effective factors or the differences of the heterozygotes from the mid-parental value are equal, they reduce to

$$V(K_3) = K_3^2 \left[\frac{V_G}{G^2} + 9\frac{V_B}{B^2} + 9\frac{V_H}{H^2} + 6\frac{\text{cov.}\,GB}{GB} - 6\frac{\text{cov.}\,GH}{GH} - 18\frac{\text{cov.}\,BH}{BH}\right]$$

$$V(K_4) = K_4^2 \left[\frac{4V_D}{D^2} + \frac{1}{9}\frac{V_\Delta}{\Delta^2} - \frac{4}{3}\frac{\text{cov.}\,D\Delta}{D\Delta} + \frac{64}{9}\frac{V_A}{A^2} - \frac{32}{9}\frac{\text{cov.}\,AD}{AD} + \frac{16}{9}\frac{\text{cov.}\,A\Delta}{A\Delta}\right]$$

$$V(K_5) = K_5^2 \left[9\frac{V_R}{R^2} + 64\frac{V_G}{G^2} + 36\frac{V_H}{H^2} - 48\frac{\text{cov.}\,RG}{RG} + 36\frac{\text{cov.}\,HR}{HR} - 96\frac{\text{cov.}\,GH}{GH}\right]$$

Now we shall use some simple approximations for these variances in order to compare their relative magnitudes. V_A/A^2 and V_B/B^2 are expected to be very small as compared to V_D/D^2 and V_H/H^2 respectively and can therefore be ignored. Further V_G/G^2, V_R/R^2, etc., are also expected to be negligible as compared to V_H/H^2 and similarly neglecting V_Δ/Δ^2 in comparison to V_D/D^2 and with the further-assumption that D and H are estimated with the relative precisions in the ratio of $M:1$ *i.e.*, $V_D/D^2 = 1/m^2\ V_H/H^2$, the variances of the different estimates reduce to

$$V(K_1) = K_1^2 \frac{V_D}{D^2}$$

$$V(K_1') = m^2 K_1'^2 \frac{V_H}{H^2}$$

$V(k_2)$ will lie between

$$4K_2^2 \frac{V_D}{D^2} \quad \text{and} \quad 4m^2K_2^2 \frac{V_D}{D^2}$$

$$V(K_3) = 9m^2 K_3^2 \frac{V_D}{D^2}$$

$$V(K_4) = 9K_4^2 \frac{V_D}{D^2}$$

$$V(K_5) = 36m^2K_5^2 \frac{V_D}{D^2}.$$

If we take $m = 1$, which is expected to be so when the data from the back-crosses is used it is seen that the standard error of K_5 is six times the standard error of K_1, or K_1', three times as much as the standard error of K_2 and K_4 and twice as much as the standard error of K_3. This estimate K_5 is not biased due to the lack of complete concentration like K_1, K_1', K_3 and K_4. The extent of under-estimation in K_1 and K_1' due to lack of complete concentration may be judged by the table given on next page.

Similar figures will give the under-estimation in K_1' if there is lack of reinforcement in h's.

It is clear from the table that the slightest disturbance in the concentration of the genes has a substantial effect on lowering the estimate of the number of factors. K_2 is also biased as it gives us an under-estimate when the assumption of the equality of $(\frac{1}{2}d^2 + \frac{1}{4}h^2)$'s is not satisfied. This estimate is also not very general as it can be

Lack of concentration (proportion of unfavourable genes in the favourable parent) %	Under-estimation of K_1 %
5	19
10	36
15	51
20	64
30	84
40	96

obtained only in cases where selfing is possible and F_3's are obtained. It is therefore desirable that the estimate K_5, though it has a higher variance, is used wherever possible as this is the only estimate which is unbiased.

An estimate similar to K_5 using d increments can be obtained. Second, fourth and sixth moments in that case will have to be used as $\Sigma (d^3)$ and $\Sigma (d^5)$ cannot be estimated from the third degree and fifth degree statistics.

Summary

The estimation of the effective number of factors has its importance because of the fact that the limit to the genetic advance that can be achieved and the speed of advance under a particular type of breeding and with a given degree of selection are known to depend upon it. Due to its importance the problem has been drawing the attention of different authors at different times.

Wright (1934), Goodwin (1944) and others worked out an estimate assuming the magnitude of the effects of all the factors to be the same and further that the allelomorphs are distributed isodirectionally. The departure from any of these assumptions leads to an under-estimation of the number of effective factors. Mathur (1952) has found other estimates replacing the assumption of equality of the effects by symmetry. His estimates are given by the roots of certain equations between which he has not been able to discriminate. Panse (1940) derived an estimate assuming the equality of a certain function of magnitude of

effects. In the present investigation the functional relations $(10G^2-9HR)$ and $(ZD-\sqrt{2}\Delta^2)$ have been found out, the signs of which discriminate between the roots of K_3 and K_4. The variances of these functional relations have also been worked out. Another unbiased estimate which assumes only the symmetry of the effects about their mean has been found out and is given by the real root of the equation

$$4R^3K^3 + 2G^2(G^2 - 12HR)K^2 + 9H^3(8G^2 - 3HR)K - 27H^6 = 0$$

the other two roots being imaginary. The variances of the different estimates have been found and compared under certain assumptions.

Acknowledgments

I am thankful to Dr. V. G. Panse and Shri V. N. Amble for their guidance and help to carry out this work.

References

1. Burnside, W. S. and Panton, A. W. *Theory of Equations*, 1935, **1** (Hodges, Figgis and Co., Dublin).
2. Kendall, M. G. .. *The Advanced Theory of Statistics*, 1948, **1** (Charles Griffin and Co., Ltd.).
3. Mather, K. .. *Biometrical Genetics*, 1949 (Mathuen and Co. Ltd.).
4. Mathur, P. N. .. "The estimation of the effective number of factors responsible for the inheritance of quantitative characters (Unpublished *Thesis*, Diploma in Agriculture and Animal Husbandry Statistics, I.C.A.R.).
5. Panse, V. G. .. *A Statistical Study of Quantitative Inheritance*, *Annals of Eugenics*, 1940, **10**, 76–105.