

Journal of the Indian Society of Agricultural Statistics

Vol. XVII]

1965

[No. 1

ESTIMATION OF GENETIC VARIABILITY IN NATURAL POPULATION OF SELF-FERTILIZED CROPS

By T. P. ABRAHAM

Institute of Agricultural Research Statistics

SELECTION from existing natural population is one of the important means of plant improvement, especially with self-fertilized crops such as rice. The success of such selection depends on the amount of genetic variability present in the population and it is therefore important to assess the extent of genetic variability present in different populations before resorting to selection and breeding. There are two well-known methods of estimating genetic fraction of the gross variability. These are use of (1) the regression of progeny mean on parental value and (2) from the analysis of variance of replicated progeny-row trials. These two methods are considered in detail by Hutchinson and Panse (1937), Panse (1940), Panse and Bokil (1948). However, the relative efficiencies of these two procedures have not been considered by these authors. In the present investigation, the different estimation procedure have been compared. An alternative estimate is also proposed.

I. METHODS OF ESTIMATING GENETIC VARIABILITY

The experimental procedure consists in taking measurements of the character under consideration on a number of plants taken at random from the population and growing the progenies of a subsample of these plants in replicated progeny-row trials. If x_i denotes the observed value for the i -th plant, y_i , the corresponding progeny mean, g_i , its genetic value and l_i , the environmental modification, then the basic model is:

$$x_i = g_i + e_i; \quad y_i = g_i + e_i' \quad (1)$$

where e_i' is the environmental component associated with progeny means.

Assuming no interaction between genotype and environment within the ranges under consideration we have

$$\begin{aligned} V(x) &= V(g) + V(e) \\ V(y) &= V(g) + V(e') \end{aligned} \quad (2)$$

$V(x)$ denoting the variance of x , etc.

For convenience, we may denote

$$\begin{aligned} V(x) &= \sigma_x^2; \quad V(y) = \sigma_y^2; \quad V(g) = \sigma_g^2 \\ V(e) &= \sigma_e^2; \quad V(e') = \sigma_{e'}^2. \end{aligned}$$

If H is the heritability of individual plants in the original population by definition $H = \sigma_g^2 / (\sigma_g^2 + \sigma_e^2)$ and if H' is the heritability for progeny mean then $H' = \sigma_g^2 / (\sigma_g^2 + \sigma_{e'}^2)$.

(i) *Estimation of σ_g^2 by Regression Technique:*

If N plants taken at random from the population are measured of which the progenies of ' n ' plants taken at random are grown, cov. (x, y) ($= A_1$ say) for the n , pairs of parent progeny means is easily seen to be an estimate of σ_g^2 . The observations on the additional $N - n = n'$ plants of the parental generation can be utilized to improve this estimate. If $s_x'^2$ denotes the variance of the N parents and b , the regression of the progeny mean on the parental value calculated from the n , pairs of parent progeny means, an estimate of σ_g^2 , is given by

$$bs_x'^2 (= A_2 \text{ say}).$$

It is easy to derive asymptotic variances of these estimates. Assuming that covariance of b and $s_x'^2$ is zero which is true for large samples we have

$$V(A_2) = \beta^2 V(s_x'^2) + \sigma_x^4 V(b) \quad (3)$$

If ρ is the correlation between x and y in the population, we have

$$V(A_2) = \beta^2 \frac{2\sigma_x^4}{N} + \frac{1}{n} \sigma_x^2 \sigma_y^2 (1 - \rho^2) \quad (4)$$

since

$$V(b) = \frac{1}{n} \frac{\sigma_y^2}{\sigma_x^2} (1 - \rho^2).$$

On replacing ρ^2 and β^2 in term of σ_g^2 , $\sigma_e'^2$ and σ_e^2 we have

$$V(A_2) = \sigma_g^4 \left[\frac{2}{N} - \frac{1}{n} + \frac{\left(1 + \frac{\sigma_e^2}{\sigma_g^2}\right) \left(1 + \frac{\sigma_e'^2}{\sigma_g^2}\right)}{n} \right]. \tag{5}$$

Since

$$\rho^2 = \frac{\sigma_g^4}{(\sigma_g^2 + \sigma_e^2)(\sigma_g^2 + \sigma_e'^2)}.$$

Replacing σ_e^2 and $\sigma_e'^2$ in terms of H and H'

$$V(A_2) = \sigma_g^4 \left[\frac{2}{N} + \frac{1}{n} \left(\frac{1}{HH'} - 1 \right) \right] \tag{6}$$

Putting $N = n$ we can easily get

$$V(A_1) = \frac{\sigma_g^4}{n} \left[1 + \frac{(\sigma_g^2 + \sigma_e^2)(\sigma_g^2 + \sigma_e'^2)}{\sigma_g^4} \right] \tag{7}$$

$$= \frac{\sigma_g^4}{n} \left[1 + \frac{1}{HH'} \right] \tag{8}$$

(ii) *Estimation of σ_g^2 from Replicated Progeny Row Trials*

Let the progenies of 'n' plants be grown replicated r times and the analysis of variance be carried out.

TABLE I

Analysis of variance of a replicated progeny-row trial

Source of variation	d.f.	M.S.	E(M.S.)
Blocks	$(r-1)$		
Progenies	$(n-1)$	P	$r(\hat{\sigma}_g^2 + \sigma_e'^2)$
Error	$(n-1)(r-1)$	s^2	$r\sigma_e^2$

Estimate of σ_g^2 is therefore given by

$$B = \hat{\sigma}_g^2 = \frac{P - s^2}{r} \tag{9}$$

$$V(B) = \frac{2}{r^2} \left[\frac{E^2(P)}{(n-1)} + \frac{E^2(S^2)}{(n-1)(r-1)} \right]$$

$$= \frac{2}{(n-1)} \left[(\sigma_g^2 + \sigma_e^2)^2 + \frac{1}{(r-1)} \sigma_e'^4 \right] \quad (10)$$

$$= \frac{2}{(n-1)} \sigma_g^4 \left[\frac{1}{H'^2} + \frac{1}{(r-1)} \left(\frac{1}{H'} - 1 \right)^2 \right] \quad (11)$$

(iii) *Estimation of σ_g^2 Using Maximum Likelihood Method*

It is clear that in estimating genetic variance either by covariance technique or by progeny mean square method a part of the data is not utilised. When covariance is used, the information on environmental variance supplied by the individual replications of the progenies is not utilized, while in using progeny-row mean square method, the information on parental values is not utilised. Estimate based on the full data can be obtained by the maximum likelihood method. For this, we may assume that the progeny mean and parental values are distributed as a bivariate normal distribution. If x , as before, denote the parental value, y , the progeny mean, σ_x^2 and σ_y^2 the variance of parental and progeny means respectively, μ , the population mean same for parents and progenies, ρ , the correlation between parents and progenies the likelihood function L is given by

$$L = \frac{\text{constant}}{(1 - \rho^2)^{n/2} \sigma_x^{(n+n')} \sigma_y^n (\sigma^2)^{(n-1)(r-1)/2}}$$

$$\times \exp \left[-\frac{1}{2(1 - \rho^2)} \sum_n \left\{ \frac{(x - \mu)^2}{\sigma_x^2} - \frac{2\rho(x - \mu)(y - \mu)}{\sigma_x \sigma_y} + \frac{(y - \mu)^2}{\sigma_y^2} \right\} \right] \left[\exp. -\frac{1}{2} \sum_{n'} \frac{(x - \mu)^2}{\sigma_x^2} \right]$$

$$\times \left[\exp. \frac{(n-1)(r-1)S^2}{2\sigma^2} \right] (S^2)^{(n-1)(r-1)/2}$$

$$= L_1 \times L_2 \times L_3$$

where

$$\sigma_x^2 = \sigma_g^2 + \sigma_e^2; \quad \sigma_y^2 = \sigma_g^2 + \sigma_e'^2$$

$$\rho = \frac{\sigma_g^4}{\sigma_x^2 \sigma_y^2} \quad \text{and} \quad \sigma^2 = r(\sigma_y^2 - \sigma_g^2).$$

As σ_x^2 and σ_y^2 are linear functions of σ_e^2 , σ_g^2 and σ_o^2 we can easily show that

$$\hat{\sigma}_x^2 = \hat{\sigma}_g^2 + \hat{\sigma}_e^2$$

and

$$\hat{\sigma}_y^2 = \hat{\sigma}_g^2 + \hat{\sigma}_e^2$$

where $\hat{\sigma}^2$ is the maximum likelihood estimate of σ^2 . So it is immaterial whether we estimate σ_x^2 , σ_y^2 , or σ_e^2 , σ_g^2 .

The likelihood equations are

$$\frac{\partial \log L}{\partial \sigma_x^2} = A + BS_1 + CS_2 + DP + WS_3 = 0 \tag{13}$$

where

$$A = -\frac{n \sigma_y^2}{2 \Sigma} - \frac{n'}{2 \sigma_x^2}; \quad B = \frac{\sigma_y^4}{2 \Sigma^2}; \quad C = \frac{\sigma_g^4}{2 \Sigma^2};$$

$$D = -\frac{\sigma_g^2 \sigma_y^2}{\Sigma^2}; \quad W = \frac{1}{2 \sigma_x^4}; \quad \Sigma = \sigma_x^2 \sigma_y^2 - \sigma_g^4;$$

$$S_1 = \sum_n (x - \hat{\mu})^2; \quad S_2 = \sum_n (y - \hat{\mu})^2;$$

$$P = \sum_n (x - \hat{\mu})(y - \hat{\mu}); \quad S_3 = \sum_{n'} (x - \hat{\mu})^2;$$

and

$$\hat{\mu} = \frac{(m + n') \bar{x}_{n+n'} + n \bar{y}_n}{2n + n'}$$

$$\frac{\partial \log L}{\partial \sigma_y^2} = E + CS_1 + GS_2 + HP + TS^2 = 0 \tag{14}$$

where

$$E = -\frac{n \sigma_x^2}{2 \Sigma} - \frac{(n-1)(r-1)}{2(\sigma_y^2 - \sigma_g^2)}; \quad C = \frac{\sigma_g^4}{2 \Sigma^2};$$

$$G = \frac{\sigma_x^4}{2 \Sigma^2}; \quad H = -\frac{\sigma_g^2 \sigma_x^2}{\Sigma^2}; \quad T = \frac{(n-1)(r-1)}{2r(\sigma_y^2 - \sigma_g^2)}$$

and

s^2 = error mean square in Table I for the replicated progeny row trial.

$$\frac{\partial \log L}{\partial \sigma_g^2} = I + JS_1 + KS_2 + LP - TS^2 \tag{15}$$

Where

$$I = \frac{n\sigma_g^2}{\Sigma} + \frac{(n-1)(r-1)}{2(\sigma_y^2 - \sigma_g^2)}; \quad J = -\frac{\sigma_y^2\sigma_g^2}{\Sigma^2};$$

$$K = -\frac{\sigma_x^2\sigma_g^2}{\Sigma^2} \quad \text{and} \quad L = \frac{1}{\Sigma^2}(\sigma_x^2\sigma_y^2 + \sigma_g^4).$$

The information matrix I is given below:

$$I = \left(\left(E \left\{ -\frac{\partial^2 \log L}{\partial \theta_i \partial \theta_j} \right\} \right) \right) \quad i, j = 1, 2, 3$$

$\theta_1 = \sigma_x^2, \theta_2 = \sigma_y^2$ and $\theta_3 = \sigma_g^2$

$$I = \begin{bmatrix} \frac{n\sigma_y^4}{2\Sigma} + \frac{n'}{2\sigma_x^4} & & \\ \frac{n\sigma_g^4}{2\Sigma^2} & \frac{n\sigma_x^4}{2\Sigma^2} + \frac{d}{2(\sigma_y^2 - \sigma_g^2)^2} & \\ -\frac{n\sigma_y^2\sigma_g^2}{\Sigma^2} & -\frac{n\sigma_x^2\sigma_g^2}{\Sigma^2} - \frac{d}{2(\sigma_y^2 - \sigma_g^2)^2} & \frac{n}{\Sigma^2}(\sigma_x^2\sigma_y^2 + \sigma_g^4) + \frac{d}{2}(\sigma_y^2 - \sigma_g^2)^2 \end{bmatrix}$$

where

$$d = (n-1)(r-1).$$

To solve the likelihood Equation (13), (14) and (15) let us have $\sigma_{x_0}^2$, $\sigma_{y_0}^2$ and $\sigma_{g_0}^2$ as initial values.

We can improve upon these values by the amount $\delta\sigma_{x_0}^2$, $\delta\sigma_{y_0}^2$ and $\delta\sigma_{g_0}^2$ where these increments are given by the following simultaneous equation in three unknowns.

Denoting

$$\Delta_0 = (\delta\sigma_{x_0}^2, \delta\sigma_{y_0}^2, \delta\sigma_{g_0}^2)$$

and

$$t = \left(\frac{\partial \log L}{\partial \sigma_{x_0}^2}, \frac{\partial \log L}{\partial \sigma_{y_0}^2}, \frac{\partial \log L}{\partial \sigma_{g_0}^2} \right)$$

these equations are

$$I_0 \Delta_0' = t'.$$

Thus proceeding we can get the solution of the Equations (13), (14) and (15), *i.e.*, the maximum likelihood estimates of σ_x^2 , σ_y^2 and σ_g^2 to

the desired degree of precision. The sample variance-covariance matrix of the estimates are given by I^{-1} .

Let M be the maximum likelihood estimate of σ_v^2 ; then

$$V(M) = \frac{\frac{n}{4\Sigma^2} \left(\frac{n(\sigma_x^2\sigma_y^2 + \sigma_v^4)}{\Sigma} + n' + \frac{d\sigma_y^4}{(\sigma_y^2 - \sigma_v^2)^2} \right) + \frac{n'}{4} \frac{d}{(\sigma_y^2 - \sigma_v^2)^2} \cdot \frac{1}{\sigma_x^4}}{\frac{n^2}{4\Sigma^3} \left\{ n+n'+d \left(1 + \frac{\Sigma}{2(\sigma_y^2 - \sigma_v^2)^2} \right) \right\} + \frac{nm'd}{\delta\Sigma^2(\sigma_y^2 - \sigma_v^2)^2} \cdot \frac{1}{\sigma_x^4} \{ (\sigma_y^2 - 2\sigma_v^2)^2 + 2\Sigma \}}$$
(16)

$1/V(M)$ can be considered to be the amount of information contained in the data on the genetic parameter σ_v^2 .

The estimates A_1 and A_2 can be obtained easily from the likelihood function as shown below:

(1) Considering the likelihood function L_1 only we get

$$\hat{\sigma}_x^2 = \frac{S_1}{n}; \quad \hat{\sigma}_y^2 = \frac{S_2}{n} \quad \text{and} \quad \sigma_v^2 = \frac{P}{n} = A_1$$

$$V(\hat{\sigma}_v^2) = \frac{1}{n} (\sigma_x^2\sigma_y^2 + \sigma_v^4) = \frac{\sigma_v^4}{n} \left(1 + \frac{1}{HH'} \right) = V(A_1).$$

(2) Considering the likelihood function L_1 and L_2 we have

$$\hat{\sigma}_x^2 = \frac{S_1 + S_3}{n + n'}; \quad \sigma_y^2 = \frac{1}{n} \left[S_2 - \frac{P^2}{S_1^2} \frac{(n'S_1 - nS_3)}{n + n'} \right];$$

$$\hat{\sigma}_v^2 = \frac{P}{(n + n')} \frac{(S_1 + S_3)}{S_1} = A_2;$$

$$V(\hat{\sigma}_v^2) = \frac{n' + \frac{n}{\Sigma} (\sigma_x^2\sigma_y^2 + \sigma_v^4)}{\frac{n}{\Sigma} (n + n')} = \sigma_v^4 \left[\frac{2}{N} + \frac{1}{n} \left(\frac{1}{HH'} - 1 \right) \right]$$

$$= V(A_2).$$

Thus we see the estimates A_1 and A_2 , earlier discussed, are maximum likelihood estimates if the information on the experimental error in replicated progeny-row trials is not taken into account.

(iv) *An Alternative to Maximum Likelihood Estimate*

Since the maximum likelihood equations cannot be explicitly solved easily we may consider alternative estimates. One such estimator

is the best (minimum variance) linear combination of the estimates A_1 and B .

$$G^{(1)} = \frac{w_1 A_1 + w_2 B}{w_1 + w_2}$$

where

$$w_1 = V(B) - \text{cov.} (A_1, B)$$

and

$$w_2 = V(A_1) - \text{cov.} (A_1, B).$$

It is proved below that the above combined estimate has the same asymptotic efficiency as the likelihood estimate.

Let $M^{(1)}$ be the maximum likelihood estimate that can be obtained from the likelihood functions L_1 and L_3 (*i.e.*, not considering the n additional observations on the parents).

We can easily derive that

$$V(M^{(1)}) = \frac{2}{n} \left\{ \frac{n(\sigma_x^2 \sigma_y^2 + \sigma_g^4) + \frac{d\sigma_y^4 \Sigma}{(\sigma_y^2 - \sigma_g^2)^2}}{2n + d \left(2 + \frac{\Sigma}{(\sigma_y^2 - \sigma_g^2)^2} \right)} \right\} \quad (17)$$

$$V(G^{(1)}) = \frac{V(A_1) V(B) - \text{cov.}^2(A_1, B)}{V(A_1) + V(B) - 2 \text{cov.}(A_1, B)}$$

It can be shown that

$$\text{Cov.}(A_1, B) = \frac{2}{n} \sigma_g^2 \sigma_y^2. \quad (18)$$

Substituting for the variances and covariance we have

$$V(G^{(1)}) = \frac{\frac{2}{n} (\sigma_x^2 \sigma_y^2 + \sigma_g^4) \left(\frac{\sigma_y^4}{(n-1)} + \frac{(\sigma_y^2 - \sigma_g^2)^2}{d} \right) - \frac{4}{n^2} \sigma_y^4 \sigma_y^4}{2 \left[\frac{\sigma_y^4}{(n-1)} + \frac{(\sigma_y^2 - \sigma_g^2)^2}{d} \right] + \frac{1}{n} (\sigma_x^2 \sigma_y^2 + \sigma_g^4) - \frac{4}{n} \sigma_g^2 \sigma_y^2} \quad (19)$$

(v) *Theorem:* $V(G^{(1)}) = V(M^{(1)})$ to the first order of approximation.

Proof: Replacing $(n-1)$ by n in the formula (19) we have

$$V(G^{(1)}) = \frac{2}{n} \left[\frac{\frac{n}{d} (\sigma_x^2 \sigma_y^2 + \sigma_g^4) (\sigma_y^2 - \sigma_g^2)^2 + (\sigma_x^2 \sigma_y^2 - \sigma_g^4) \sigma_y^4}{\frac{2n}{d} (\sigma_y^2 - \sigma_g^2) + 2 (\sigma_y^2 - \sigma_g^2)^2 + (\sigma_x^2 \sigma_y^2 - \sigma_g^4)} \right]$$

Multiplying numerator and denominator by

$$\frac{d}{(\sigma_y^2 - \sigma_g^2)^2}$$

$$= \frac{2}{n} \left[\frac{n(\sigma_x^2 \sigma_y^2 + \sigma_g^4) + \frac{d}{(\sigma_y^2 - \sigma_g^2)} \sigma_y^4 (\sigma_x^2 \sigma_y^2 - \sigma_g^4)}{2n + d \left\{ 2 + \frac{(\sigma_x^2 \sigma_y^2 - \sigma_g^4)}{(\sigma_y^2 - \sigma_g^2)^2} \right\}} \right]$$

$$= V(M^{(1)}) \quad \text{(From 17)}$$

Similarly we can define another best (minimum variance) linear estimate $G^{(2)}$ making use of the additional observations on the parent namely

$$G^{(2)} = \frac{w_1 A_2 + w_2 B}{w_1 + w_2}$$

with

$$w_1 = V(B) - \text{cov.}(A_2, B)$$

and

$$w_2 = V(A_2) - \text{cov.}(A_2, B).$$

$V(A_2)$, $V(B)$ are given by formulae 5 and (10) and we can easily derive that:

$$\text{Cov.}(A_2, B) = \frac{2}{N} \sigma_g^2 \sigma_y^2 + \frac{2n' \sigma_g^2 \Sigma}{Nn \sigma_x^2}.$$

Thus we have

$$V(G^{(2)}) = \frac{\frac{2}{n^2} (\sigma_y^4 + \sigma_x^2 \sigma_y^2) \left\{ \sigma_y^4 + \frac{n}{d} (\sigma_y^2 - \sigma_g^2)^2 \right\} - \frac{4}{n^2} \sigma_g^4 \sigma_y^4 - \frac{4n'}{Nn^2} \sigma_g^4 \left[\sigma_y^4 + \frac{n}{d} (\sigma_y^2 - \sigma_g^2)^2 \right] - \frac{4n'^2 \sigma_g^{12}}{N^2 n^2 \sigma_x^4} - \frac{8n' \sigma_g^8 \sigma_y^2}{Nn^2 \sigma_x^2}}{\frac{2}{n} \left\{ \sigma_y^4 + \frac{n}{d} (\sigma_y^2 - \sigma_g^2)^2 \right\} + \frac{1}{n} (\sigma_g^4 + \sigma_x^2 \sigma_y^2) - \frac{4}{n} \sigma_g^2 \sigma_y^2 - \frac{2n'}{Nn} \sigma_g^4 - \frac{4n' \sigma_g^6}{Nn \sigma_x^2}}$$

(20)

II. COMPARISON OF DIFFERENT ESTIMATES WITHIN THE RANGE OF HERITABILITIES ORDINARILY OF INTEREST

We have defined earlier H and H' the heritability of individual plants in the original population and that of progeny means.

$$H = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}; \quad H' = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{e'}^2}.$$

The heritability of progeny means will be higher on account of the smaller contribution of environmental factors to the gross variability of progeny means. The magnitude $\sigma_{e'}^2$ depends on the number of plants grown per progeny and the layout adopted, etc. Heritability depends, apart from the contributions of environmental variation, on the genetic variance of the population.

The minimum coefficient of variation for genetic variability to achieve worthwhile improvement by selection can be taken as approximately 10%.

At the Central Rice Research Institute and elsewhere it has been found that the environmental variation of single plant yields grown under transplanted conditions is about 40% for rice crop. If the progeny size is M , the environmental variance between progeny means will not be $1/M$ times the single plant variation on account of the plot size being larger than single plots, but will be reduced to a lesser extent depending upon the replication and number of plants per plot. With increase in plot size the coefficient of variation (C.V.) generally decreases according to the law ax^{-g} where 'x' is plot size and 'g' and 'a' are constants.

The coefficient of variation of progeny means

$$= \frac{\text{C.V. (single plants)}}{\left(\frac{M}{r}\right)^g \sqrt{r}}.$$

For paddy, from uniformity trial data collected at the central Rice Research Institute, the value of g is found nearly 0.25. For this value of g , the standard error per cent. of progeny mean

$$= \frac{\text{C.V. (single plants)}}{(Mr)^{\frac{1}{4}}}.$$

If about 100 plants are grown for each progeny putting 5 replications of 20 plants each the standard error per cent. will be approximately 8% when single plant variation is 40%. The s.e.% actually

obtained in progeny trials of similar size agrees with this value. For this value of environmental variance and taking C.V. for genetic variability as about 10% the value of HH' will be about 0.03. For some of the material on cotton examined by Panse (1940) the heritabilities are much higher. In general, the value of HH' will seldom exceed 0.2.

(i) *Gain due to additional observations on parents*

From formulæ (6) and (7), we have

$$V(A_1) = \frac{\sigma_g^4}{n} \left(\frac{1}{HH'} + 1 \right)$$

and

$$V(A_2) = \sigma_g^4 \left\{ \frac{2}{N} + \frac{1}{n} \left(\frac{1}{HH'} - 1 \right) \right\}.$$

Additional information on σ_g^2 obtained from the measurements on the $(N - n)$ plants,

$$= \frac{\frac{1}{n} \left(\frac{1}{HH'} + 1 \right)}{\frac{2}{N} + \frac{1}{n} \left(\frac{1}{HH'} - 1 \right)} - 1,$$

which simplifies to

$$\frac{1}{\frac{1}{2} \left(1 + \frac{n}{n'} \right) \left(1 + \frac{1}{HH'} \right)}$$

where $n' = N - n$.

Formula (21) shows that the gain in information depends on the proportion of additional observations n' to the number of progenies grown n , and the heritabilities H and H' . If the heritabilities are high the additional information for given sample sizes will be large. The limit to gain in information is obtained when $n' \rightarrow \infty$ i.e., when $(n/n') \rightarrow 0$. This gives the limiting information as $2HH'/1 - HH'$. Ordinarily HH' will seldom exceed 0.2 giving the maximum gain in information as 50%. In the majority of cases the maximum gain will be much less, as HH' is likely to be of the order of .05 in which case the upper limit to the gain in information will be about 1.5%. It is therefore obvious that ordinarily there will be little advantage in additional measurements on parental generation.

(ii) *Relative Efficiency of Progeny Mean Square, Covariance and Their Combined Estimates*

Using the variance formulæ (8) and (11) for the covariance estimate A_1 , and progeny mean square estimate B respectively, we have the relative efficiency of progeny mean square estimate compared to covariance estimate as

$$2 \frac{H}{H'} \left[1 + \frac{1-H'}{r-1} \right] \simeq \frac{H'}{2H} \text{ (approximately).} \quad (22)$$

The efficiency of the combined estimate $G^{(1)}$ is given by

$$\frac{V(A_1)}{V(G^{(1)})} = 1 + \frac{\left[1 - \frac{\text{cov.}(A_1, B)}{V(A_1)} \right]}{\frac{V(B)}{V(A_1)} - \left\{ \frac{\text{cov.}(A_1, B)}{V(A_1)} \right\}^2} \quad (23)$$

From (7), (10) and (18) we have $V(A_1)$, $V(B)$ and $\text{cov.}(A_1, B)$. On substituting these values in (23) we have relative efficiency of the combined estimate as

$$1 + \frac{\left(1 - \frac{2H}{1+HH'} \right)^2}{\frac{2H}{1+HH'} \left[\frac{r}{r-1} \times \frac{1}{H} - \frac{2}{r-1} + \frac{H'}{r-1} \right] - \left(\frac{2H}{1+HH'} \right)^2} \\ \simeq 1 + \frac{1-4H}{\frac{2H}{r-1} \left[\frac{r}{H'} - 2 \right]} \text{ when } H \text{ and } H' \text{ are small.} \quad (24)$$

It will be seen that the relative efficiencies are not dependent directly on the sample size, but depends on the number of replications and heritabilities of parent and progeny means. Table II gives the relative efficiencies of progeny mean square and combined estimates compared to the covariance estimate for some selected values of H and H' .

The progeny mean square estimate is more efficient than covariance estimate, when the heritability of progeny means is more than about two and half times the heritability of individual plants. The efficiency of progeny mean square estimate is increased when the heritability of progeny means is increased, as is to be expected. Increased number of replications also increases the efficiency of mean square estimate, even when the values of H and H' remain the same. The combined

TABLE II

Relative efficiencies of progeny mean square and combined estimates over covariance estimate

(Top figures refer to progeny, mean square estimate and bottom figures to combined estimate)

<i>H'</i> \ <i>H</i>	.05	.08	.10	.20	.15	.05	.08	.10	.15	.20
.10	0.85	0.54	0.80	0.50
	1.71	1.39	1.65	1.36
.15	1.32	0.83	0.66	0.13	0.77	0.61
	2.09	1.60	1.44	2.00	1.55	1.41
.20	1.79	1.13	0.90	0.61	..	1.66	1.05	0.84	0.56	..
	2.48	1.82	1.60	1.32	..	2.37	1.76	1.56	1.30	..
.30	2.77	1.79	1.41	0.95	0.72	2.62	1.65	1.33	0.90	0.68
	3.31	2.30	1.96	1.52	1.31	3.19	2.23	1.91	1.49	1.29

Note.— $H' \geq H$. Therefore efficiencies are not given for values $H' < H$.

estimate is far superior to either of the other estimates, its efficiency being generally more than 30% over the more efficient of the progeny mean square or covariance estimates. We have proved earlier (I, V) that the combined estimate $G^{(1)}$ has the same asymptotic efficiency as the likelihood estimate M . However, the full gain in information will not be obtained due to inaccuracies in the estimates of the weights, but with large samples, as will ordinarily be the case, the loss in information on this account is not likely to be so high as to offset the advantage of combining the estimates.

- (iv) *Comparison of combined estimate with maximum likelihood estimated, both estimates based on full data including the additional observations on the parent plants.*

From the formulæ (16), (17) and (20) derived earlier we have

$$V(G^{(1)}) = \frac{2}{n} \sigma_e^4 \frac{K_1}{K_2}$$

$$V(M') = \frac{2}{n} \sigma_g^4 \frac{K_1 + L_1 t}{K_2 + L_2 t}$$

$$V(G^{(2)}) = \frac{2}{n} \sigma_g^4 \left[\frac{K_1 + B_1 t + A_1 t^2}{K_2 + B_2 t + A_2 t^2} \right]$$

where

$$A_1 = K_1 + P_1 + P_2; \quad B_1 = 2K_1 + P_1; \quad A_2 = K_2 + P_3;$$

$$B_2 = 2K_2 + M_3; \quad t = n'/n;$$

$$K_1 = \left(1 + \frac{1}{HH'}\right) \left\{ \frac{1}{(H')^2} + \frac{(1-H')^2}{(r-1)H'^2} \right\} - \frac{2}{H'^2}$$

$$K_2 = 2 \left[\frac{1}{H'^2} + \frac{(1-H')^2}{(r-1)H'^2} \right] + \left(\frac{1}{HH'} + 1 \right) - \frac{4}{H'}$$

$$L_1 = \left(\frac{1}{H'} - 1 \right)^2 \frac{1}{r-1} \left(\frac{1}{HH'} - 1 \right) + H^2 \left(\frac{1}{HH'} - 1 \right)^3$$

$$L_2 = \left(\frac{2}{r-1} \right) \left(\frac{1}{H'} - 1 \right)^2 + \left(\frac{1}{HH'} - 1 \right) H^2$$

$$\times \left[\left(\frac{1}{H} - 2 \right)^2 - 2 \left(\frac{1}{HH'} - 1 \right) \right]$$

$$P_1 = -2 \left[\frac{1}{H'^2} + \frac{1}{r-1} \left(\frac{1}{H'} - 1 \right)^2 + \frac{2H}{H'} \right]$$

$$P_2 = -2H^2; \quad P_3 = -2(1 + 2H).$$

In Table III, we have given the relative efficiency of combined estimate $G^{(2)}$ as compared to the maximum likelihood estimate M for different values of H and H' .

Thus we find from Table III, that the combined estimate is as good as the maximum likelihood estimate over a wide range of heritability for the progeny means. Even for higher values of H , the combined estimate is practically of the same efficiency as the maximum likelihood estimate. Earlier we found (II, i) that there is no advantage in having additional measurement on parental generations. Thus the combined estimate $G^{(1)}$ defined earlier is good enough for all practical purposes.

The minimum number of progenies to be raised for assigned levels of accuracies for the estimates of genetic variance have been worked out in Table IV.

TABLE III

% Relative efficiency of the combined estimate G^2 compared to maximum likelihood estimate M

$H' \backslash H$	·05	·10	·20
·20	99·9	99·5	..
·30	99·9	99·4	98·9
·40	99·8	99·2	98·6

TABLE IV

Minimum number of progenies (n) required to be grown for estimation of genetic variability using combined estimate $G^{(1)}$ with assigned levels of accuracy for different values of H and H'

H'	$H = 0·05$			$H = 0·1$			$H = 0·2$		
	S.E.%								
	20	30	40	20	30	40	20	30	40
0·3	531	236	133	449	199	112	341	151	85
0·4	313	139	78	248	111	62	234	104	58
0·5	203	90	51	191	85	48	170	75	42

Note.—It is assumed that 4 replications will be laid out.

It is seen from Table IV that within the range of heritabilities commonly found, a large number of progenies are to be grown to estimate genetic variability with a reasonable degree of accuracy.

III. EXAMPLE

For illustration of the method of calculating the combined estimate, let us consider the data from a replicated progeny-row trial on rice

variety K-60 conducted in Kashmir in 1962-63 under the Rice improvement scheme. There were 45 progenies tried in a replicated randomised block progeny-row trial with 4 replications and single row plots of 30 plants each.

Let x denote the parental value and y , the progeny mean. Estimates of variances and covariances were obtained as

$$\sigma_x^2 = 26.9503 \text{ gm.}^2/\text{plot.}$$

$$\sigma_y^2 = 4.0213 \quad ,,$$

$$\sigma_{xy} = 0.1586 \quad ,,$$

The analysis of variance of progeny fields is given below:

Source of variation	Degrees of freedom	Mean square gm. ² /plot	<i>F.</i>
Blocks	3	45.2376	
Progenies	44	15.7713	1.26
Error	132	12.4837	
Total	179		

The F value fails to be significant at 5%. However, for purpose of illustration, the calculations are carried out on the data for estimation of genetic variability.

Progeny mean square estimate B is given by

$$\frac{1}{4}(15.7713 - 12.4837) = 0.8219.$$

Variance,

$$V(B) = \frac{2}{4 \times 4} \left[\frac{(15.77)^2}{44} + \frac{(12.48)^2}{132} \right] = 0.8542.$$

Covariance estimate A_1 is given by

$$\sigma_{xy} = 0.1586.$$

Variance of the covariance estimate $V(A_1)$

$$= \frac{1}{4 \times 8} [26.9503 \times 4.0213 + (0.1586)^2] = 2.4689.$$

Covariance $(A_1, B) = 2 \times 4.0213 \times 0.1586 = 1.2473$.

$$w_1 = V(B) - \text{cov.}(A_1, B) = -0.3931.$$

$$w_2 = V(A_1) - \text{cov.}(A_1, B) = 1.1616.$$

Combined estimate $G^{(1)}$

$$= \frac{w_1 A_1 + w_2 B}{w_1 + w_2} = \frac{-0.3931 \times 0.1586 + 1.1616 \times 0.8219}{-0.3931 + 1.1616}$$

Variance of the combined estimate

$$\begin{aligned} & \frac{V(A_1) + V(B) - \text{cov.}^2(A_1, B)}{V(A_1) + V(B) - 2 \text{cov.}(A_1, B)} \\ &= \frac{2.4089 + 0.8542 - (1.2473)^2}{2.4089 + 0.8542 - 2.4946} \\ &= \frac{0.5019}{0.7685} = 0.6531. \end{aligned}$$

Efficiency of the combined estimate compared to progeny mean square estimate is given by $0.8542/0.6531 = 1.31$.

IV. SUMMARY

Selection from existing natural populations is one of the most widely adopted techniques in the improvement of varieties of self-fertilized crops such as rice. The content of genetic variability in such populations is estimated by growing the progenies of a sample of plants taken at random from the population and using either the regression of progeny mean on parental values by the analysis of variance. In the present paper the relative efficiency of these two procedures has been investigated. An estimate combining the above two estimates has been proposed and the gain in efficiency has been shown to be generally above 30% over the other estimates. The maximum likelihood estimate has been investigated and it is shown that the combined estimate is asymptotically as efficient as the maximum likelihood estimate. The minimum number of progenies to be grown for estimating genetic variability with different levels of accuracy have been calculated. It is shown that ordinarily about 200-300 progenies each of about 100 plants are to be raised for estimating the genetic variability with a moderate level of accuracy in rice crop.

V. ACKNOWLEDGEMENTS

I am grateful to Dr. V. G. Panse, Statistical Adviser, Institute of Agricultural Research Statistics, for his encouragement and to Shri R. Jagannathan, Institute of Agricultural Research Statistics, for assisting me in some of the calculations.

REFERENCES

1. Hutchinson, J. B. and Panse, V. G. "The design of field tests of plant breeding material," *Ind. J. Agric. Sci.*, 1937, 7, 531-64.
2. Panse, V. G. .. "Inheritance of quantitative characters and plant breeding," *J. Gen.*, 1940, 40, 283-302.
3. ——— and Bokil, S. D. .. "Estimation of genetic variability in plants," *J. Ind. Soc. Agric. Stat.*, 1948, 1, 80-90.