

# IMPROVEMENT THROUGH SELECTION AT SUCCESSIVE STAGES

BY J. P. JAIN\* AND V. N. AMBLE

*Institute of Agricultural Research Statistics, New Delhi*

## 1. INTRODUCTION

PLANT and animal breeders are often faced with the problem of finding the optimal values of intensities of selection at various stages, the optimal values considered being those which result in maximizing the annual genetic advance. With dairy cows under selection for milk yield, for example, each successive lactation provides new data on the milk yielding capacity of the animal. These lactation yields would form the basis for the successive stages of selection, selection at the  $r$ -th stage being made on the evidence of yield in the latest lactation combined with the available information on previous lactation yields. A usual feature in selection problems is that we cannot assess directly the genetic value of character which we wish to improve. In the present case selection is aimed at securing cows with superior genotypic value for milk yield but it has to be based on the observed or phenotypic values in successive lactations. The basic approach, *viz.*, of improving some character  $y$  which is not directly measurable, by means of indirect selection that is made from a group of tests or measurements  $x_1, x_2, \dots, x_r$  at successive stages, has been discussed by Cochran (1951). He has also derived the general form for gain in  $y$  expected after two-stage selection for the case when the variates  $y, x_1$  and  $x_2$  follow a multivariate normal distribution. This paper deals with the extension of the expression for selection gain to  $r$  stages. Approximate working methods for dealing with multiple stage selection are also given.

## 2. REVIEW OF LITERATURE

Smith (1936) has discussed the genetic gain for one stage selection assuming linear relationship between the genotypic and phenotypic values. Another form of the same expression has been presented by

---

\* Paper relates to work carried out by the former author as a part of thesis for Diploma in Agricultural Statistics awarded by the Institute of Agricultural Research Statistics.

Hazel (1943) while Panse (1946) has discussed its application taking several characters simultaneously with particular reference to selection in poultry.

Sieben (1954) and Keuls and Sieben (1955) who have discussed a similar problem with reference to plant selection follow a different scheme of selection. After arbitrarily partitioning the whole population of varieties into a few 'good' ones (high yielders), a few 'bad' ones (low yielders) and a large number of intermediate varieties whose yields are such that it is immaterial whether they are retained or rejected, the rule of selection is based on the consideration which aims at minimising the probability of rejection of good varieties and of selection of bad ones.

All these authors, however, are concerned primarily with one stage selection. Dickerson and Hazel (1944) have gone one stage further. The application of the formula which they have given is however restricted in that the values of proportions retained after second culling among those retained after first, must not be either much larger or smaller than 0.5. For these restricted values, the exact value of the selection differential expected after second culling does not differ appreciably from that expected from a normal distribution.

Finney (1957) has advanced a theory for two-stage selection programme and discussed the implications of its extension to  $r$  stage. This assumes that selection at stage  $r$  would be based solely on the evidence of yields in that stage. Consequently this would mean sacrificing information on the previous  $(r - 1)$  records which might have been usefully utilized in accelerating the pace of genetic gain. Nevertheless the results are of interest in that they provide a lower limit to the gains that may accrue from different rules of selection. Finney has remarked that although in theory the methods used for computing the consequences of two-stage selection could be extended to any number of stages, the complexity of the formulæ and the limits of accuracy of various mathematical tables that are employed make this impracticable even for  $r = 3$ .

### 3. APPROACH TO THE PROBLEM

The line of approach followed in this paper is the same as suggested by Cochran (1951) which may be summarised as follows:

Assume the variates  $y, x_1, x_2, \dots, x_r$  to follow a distribution whose frequency function is known. If the regression  $\eta(x)$  of  $y$  on the  $x$ 's

exists, he has shown firstly that  $\eta(x)$  is the best selection index, *i.e.*, the regressions  $\eta_1(x)$  of  $y$  on  $x_1$ ;  $\eta_2(x)$  of  $y$  on  $x_1$  and  $x_2$ , etc., will constitute the optimum selection indices at different stages of selection. If the proportions selected  $\alpha_1, \alpha_2, \dots, \alpha_r$  at different stages have been decided in advance, the units at the first stage will be selected whenever  $\eta_1 \geq k_1$ , where  $k_1$  is the truncation point corresponding to the frequency of selection  $\alpha_1$ ; the units selected at the second stage will be those for which  $\eta_2 \geq k_2$ , where given  $k_1, k_2$  is the truncation point corresponding to the frequency of selection  $\alpha_1 \alpha_2$ . The same argument will be true for further stages of selection. Secondly the gain in  $y$  is a linear function of gains in  $\eta$ 's.

#### 4. SELECTION IN $r$ STAGES

Hereafter it is assumed that the variates  $y, x_1, x_2, \dots, x_r$  follow a multivariate normal distribution. For convenience we may take deviations of all variates  $y, \eta_1, \eta_2, \dots$  and  $\eta_r$  from their respective means and effect a scaling transformation so that in the population all the variates have zero means and unit variances. Let the parameters  $\rho_1, \rho_2, \dots, \rho_r$  denote the simple correlations between  $y$  and  $\eta_1, y$  and  $\eta_2$  and so on and  $\rho_{ij}$  the correlation between  $\eta_i$  and  $\eta_j$ . For fixed  $\alpha_1, \alpha_2, \dots, \alpha_r$  the points of truncation  $k_1, k_2, \dots, k_r$  will be given by the following  $r$  equations:

$$\alpha_1 = \frac{1}{\sqrt{2\pi}} \int_{k_1}^{\infty} e^{-\eta_1^2/2} d\eta_1; \tag{1}$$

$$\alpha_1 \alpha_2 = \frac{1}{2\pi \sqrt{1 - \rho_{12}^2}} \int_{k_1}^{\infty} d\eta_1 \int_{k_2}^{\infty} e^{-\left[\frac{1}{2(1 - \rho_{12}^2)}\right] [\eta_1^2 - 2\rho_{12}\eta_1\eta_2 + \eta_2^2]} d\eta_2; \tag{2}$$

.....  
 .....

Finally

$$\alpha_1 \alpha_2 \dots \alpha_r = \frac{1}{(2\pi)^{r/2} \Delta_r^{1/2}} \int_{k_1}^{\infty} d\eta_1 \int_{k_2}^{\infty} d\eta_2 \dots \int_{k_r}^{\infty} e^{-\frac{1}{2} \sum_{i=1}^r \sum_{j=1}^r A_{ij} \eta_i \eta_j} d\eta_r \tag{3}$$

where  $\Delta_r$  is the variance-covariance matrix of order  $r$

i.e.,

$$\Delta_r = |a_{ij}|_r = \begin{vmatrix} 1 & \rho_{12} & \rho_{13} & \dots & \rho_{1r} \\ \rho_{21} & 1 & \rho_{23} & \dots & \rho_{2r} \\ \rho_{31} & \rho_{32} & 1 & \dots & \rho_{3r} \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \rho_{r1} & \rho_{r2} & \rho_{r3} & \dots & 1 \end{vmatrix}$$

and

$$A_{ij} = \frac{\text{cofactor of } a_{ij} \text{ in } |a_{ij}|_r}{\Delta_r} = \frac{B_{ij}}{\Delta_r}$$

and  $a_1, (a_1 a_2), \dots$  and  $(a_1 a_2 \dots a_r)$  are the proportions retained at first, second,  $\dots$  and  $r$ -th stages of selection respectively from the unselected units.

If  $f(y, x_1, x_2, \dots, x_r)$  is the joint frequency function of the variates  $y, x_1, x_2, \dots, x_r$ , the gain in  $y$  due to selection over  $r$  stages will be

$$G(y) = \frac{1}{a_1 a_2 \dots a_r} \int_{-\infty}^{\infty} dy \int_{\eta_1 \geq k_1} \int_{\eta_2 \geq k_2} \dots \int_{\eta_r \geq k_r} y f(y, x_1, x_2, \dots, x_r) \times dx_1 dx_2 \dots dx_r.$$

This may be shown to be equal to

$$\frac{1}{a_1 a_2 \dots a_r} \int_{\eta_1 \geq k_1} \int_{\eta_2 \geq k_2} \dots \int_{\eta_r \geq k_r} \eta(x_1, x_2, \dots, x_r) \times f_1(x_1, x_2, \dots, x_r) dx_1 dx_2 \dots dx_r$$

where  $f_1(x_1, x_2, \dots, x_r)$  is the joint frequency function of the  $x$ 's.

$\eta$ 's being linear functions of  $x$ 's, the gain in  $y$  due to selection on  $\eta_1$ , followed by selection on  $\eta_2$ , etc., will be a linear function of the gains in  $\eta_1, \eta_2, \dots$  and  $\eta_r$ .

If

$$y = \beta_1 \eta_1 + \beta_2 \eta_2 + \dots + \beta_r \eta_r + e$$

where  $e$  is the component of error which is independently normally distributed with zero mean and  $(\beta_1 \eta_1 + \beta_2 \eta_2 + \dots + \beta_r \eta_r)$  is the multiple regression of  $y$  on the  $\eta$ 's in the unselected population, then the expected value of  $y$ , the expectation being taken over the selected

part of the universe, will be the gain in  $y$  as the variates measured from their respective means and can be written as

$$G(y) = \beta_1 G(\eta_1) + \beta_2 G(\eta_2) + \dots + \beta_r G(\eta_r) \tag{4}$$

We, therefore, need only find  $G(\eta_1), G(\eta_2), \dots$  and  $G(\eta_r)$ .  
For that consider

$$\begin{aligned} Q_{(r)} &= a_1 a_2 \dots a_r G \left[ \frac{\partial g}{\partial \eta_r} \right] \\ &= \frac{1}{(2\pi)^{r/2} \Delta_r^{1/2}} \int_{k_1}^{\infty} d\eta_1 \int_{k_2}^{\infty} d\eta_2 \dots \int_{k_r}^{\infty} \left[ \frac{\partial g}{\partial \eta_r} \right] e^{-\frac{[g]}{\Delta_r}} d\eta_r \end{aligned}$$

where

$$g = \frac{1}{2} \sum_{i=1}^r \sum_{j=1}^r B_{ij} \eta_i \eta_j$$

Integrating with respect to  $\eta_r$

$$Q_{(r)} = \frac{\Delta_r^{1/2}}{(2\pi)^{r/2}} \int_{k_1}^{\infty} d\eta_1 \int_{k_2}^{\infty} d\eta_2 \dots \int_{k_{r-1}}^{\infty} e^{-\frac{1}{2\Delta_r} \left[ \sum_{i=1}^{r-1} \sum_{j=1}^{r-1} B_{ij} \eta_i \eta_j + 2k_r \sum_{i=1}^{r-1} B_{ir} \eta_i + B_{rr} k_r^2 \right]} d\eta_{r-1}$$

Transforming the variates  $\eta_i$ 's to  $u_i$ 's where

$$u_i = \eta_i - a_{ir} k_r \quad (i = 1, 2, \dots, r-1)$$

the above integral reduces to

$$\begin{aligned} Q_{(r)} &= a_1 a_2 \dots a_r G \left[ \sum_{i=1}^r B_{ir} \eta_i \right] \\ &= \frac{\Delta_r^{1/2}}{(2\pi)^{r/2}} \int_{D_1} du_1 \int_{D_2} du_2 \dots \int_{D_{r-1}} e^{-\frac{1}{2\Delta_r} \left[ \sum_{i=1}^{r-1} \sum_{j=1}^{r-1} B_{ij} u_i u_j + \Delta_r k_r^2 \right]} du_{r-1} \\ &= \left( \frac{e^{-\frac{1}{2} k_r^2}}{\sqrt{2\pi}} \right) \frac{\Delta_r^{1/2}}{(2\pi)^{(r-1)/2}} \int_{D_1} du_1 \int_{D_2} du_2 \dots \int_{D_{r-1}} e^{-\frac{1}{2\Delta_r} \sum_{i=1}^{r-1} \sum_{j=1}^{r-1} B_{ij} u_i u_j} du_{r-1} \end{aligned} \tag{5}$$

where

$$D_i \text{ ranges from } k_i - a_{ir} k_r \text{ to } \infty \quad (i = 1, 2, \dots, r-1)$$

At this stage we define a new variance-covariance matrix,  $\Delta'_{r-1}$  of order  $r-1$  formed by all the possible combinations of the first order

partial correlation coefficients of variates ranging from 1 to  $r-1$  keeping  $r$ -th variate constant, viz.,

$$\Delta'_{r-1} = |a_{ij}'|_{r-1} = \begin{vmatrix} 1 & \rho_{12,r} & \rho_{13,r} & \dots & \rho_{1,r-1,r} \\ \rho_{21,r} & 1 & \rho_{23,r} & \dots & \rho_{2,r-1,r} \\ \rho_{31,r} & \rho_{32,r} & 1 & \dots & \rho_{3,r-1,r} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \rho_{r-1,1,r} & \rho_{r-1,2,r} & \rho_{r-1,3,r} & \dots & 1 \end{vmatrix}$$

and

$$B_{ij}' = \text{cofactor of } a_{ij}' \text{ in } |a_{ij}'|_{r-1} = \Delta'_{r-1} A_{ij}'.$$

Now effecting the following transformation

$$u_i = \sqrt{1 - \rho_{ir}^2} t_i \quad (i = 1, 2, \dots, r-1)$$

and noting

$$(i) \Delta'_{r-1} = \frac{\Delta_r}{\prod_{i=1}^{r-1} (1 - \rho_{ir}^2)}$$

$$(ii) B_{ii}' = \frac{B_{ii}}{\prod_j (1 - \rho_{jr}^2)} \quad \left. \begin{array}{l} i, j = 1, 2, \dots, r-1 \\ i \neq j \end{array} \right\}$$

$$(iii) B_{ij}' = B_{ij} \sqrt{\frac{B_{ii}' B_{jj}'}{B_{ii} B_{jj}}}$$

$$(iv) \frac{\Delta_r}{\Delta'_{r-1}} \frac{B_{ii}'}{B_{ii}} = (1 - \rho_{ir}^2) \quad (i = 1, 2, \dots, r-1)$$

The integral  $Q_{(r)}$  reduces to

$$\begin{aligned} & \alpha_1 \alpha_2 \dots \alpha_r G \left[ \sum_{i=1}^r B_{ir} \eta_i \right] \\ & = \Delta_r \left( \frac{e^{-\frac{1}{2} k r^2}}{\sqrt{2\pi}} \right) \left( \frac{1}{(2\pi)^{(r-1)/2} \Delta'_{r-1}{}^{1/2}} \right) \\ & \quad \times \int_{D_1} dt_1 \int_{D_2} dt_2 \dots \int_{D_{r-1}} e^{-\frac{1}{2} \sum_{i=1}^{r-1} \sum_{j=1}^{r-1} A_{ij}' t_i t_j} dt_{r-1} \end{aligned}$$

where

$$D_i' \text{ ranges from } \frac{k_i - a_{ir}k_r}{\sqrt{1 - \rho_{ir}^2}} \text{ to } \infty \text{ (} i = 1, 2, \dots, r - 1 \text{).}$$

Thus,

$$Q_{(r)} = \Delta_r Z_{(k_r)} I_{12\dots r-1}$$

where  $Z$  denotes the ordinate of the univariate normal curve and  $I$  the incomplete volume of the  $(r - 1)$  variate normal surface respectively.

Similarly

$$Q_{(r-1)} = a_1 a_2 \dots a_r G \left[ \sum_{i=1}^r B_{ir-1} \eta_i \right] = \Delta_r Z_{(k_{r-1})} I_{12\dots (r-2)r}$$

.....  
 .....

$$Q_{(1)} = a_1 a_2 \dots a_r G \left[ \sum_{i=1}^r B_{i1} \eta_i \right] = \Delta_r Z_{(k_1)} I_{23\dots r}$$

Solving these equations for  $G(\eta_1), G(\eta_2), \dots$ . We get

$$a_1 a_2 \dots a_r \begin{bmatrix} G(\eta_1) \\ G(\eta_2) \\ \vdots \\ G(\eta_r) \end{bmatrix} = [a_{ij}]_r \begin{bmatrix} (Z_{(k_1)} I_{23\dots r}) \\ (Z_{(k_2)} I_{13\dots r}) \\ \vdots \\ (Z_{(k_r)} I_{12\dots r-1}) \end{bmatrix}$$

Substituting the values of  $G(\eta)$ 's in (4), we have

$$\begin{aligned} a_1 a_2 \dots a_r G(y) &= (\beta_1 + \rho_{12}\beta_2 + \rho_{13}\beta_3 + \dots + \rho_{1r}\beta_r) Z_{(k_1)} I_{23\dots r} \\ &+ (\rho_{21}\beta_1 + \beta_2 + \rho_{23}\beta_3 + \dots + \rho_{2r}\beta_r) Z_{(k_2)} I_{13\dots r} \\ &+ \dots \\ &+ (\rho_{r1}\beta_1 + \rho_{r2}\beta_2 + \rho_{r3}\beta_3 + \dots + \beta_r) Z_{(k_r)} I_{12\dots r-1} \end{aligned}$$

But by definition

$$\begin{aligned} \rho_i &= \text{COV}[y\eta_i] = \text{COV}[(\beta_1\eta_1 + \beta_2\eta_2 + \dots + \beta_r\eta_r)\eta_i] \\ &= \rho_{i1}\beta_1 + \rho_{i2}\beta_2 + \dots + \rho_{ir}\beta_r \\ &[i = 1, 2, \dots, r \text{ and } \rho_{ii} = 1] \end{aligned}$$

Therefore

$$G(y) = \frac{\rho_1 Z_{(k_1)} I_{23\dots r} + \rho_2 Z_{(k_2)} I_{13\dots r} + \dots + \rho_r Z_{(k_r)} I_{12\dots r-1}}{\alpha_1 \alpha_2 \dots \alpha_r} \quad (6)$$

If  $y$  does not have unit standard deviation the only change needed is to multiply the right-hand side of (6) by genetic standard deviation  $\sigma(y)$ .  $G(y) \times ch_1$ , where  $c$  is the coefficient of variation and  $h_1$  is the heritability gives the genetic advance as a percentage of the mean of the unselected population.

##### 5. APPLICATION TO A SELECTION PROBLEM IN ANIMAL BREEDING

The application of the foregoing formula may be illustrated with reference to dairy cattle under selection for improvement in their level of production. Let  $N$  be the number of cows of the initial generation,  $F_0$ , completing their first lactation. Out of these  $N$  cows, a fraction  $\alpha_1$  having the highest yields is selected the rest being discarded. From among those  $\alpha_1 N$  cows which complete the second lactation, a fraction  $\alpha_2$  having the highest total yield in the first two lactations is selected, the remainder being culled. A fraction  $\alpha_3$  is selected from  $\alpha_1 \alpha_2 N$  cows on the basis of their first three lactation records and the remainder is discarded and similarly selection being made for further stages of selection.

It is not advisable to retain even the better animals excepting the few outstanding ones, if any, for more than four lactations for the case when the herd strength from generation to generation is envisaged to be more or less constant. This is so because the culling of all the cows after only one or two lactations would mean a continuous reduction in the herd strength, and on the other hand retaining selected cows for a larger number of lactations would mean an increase in the generation interval and a corresponding decrease in the rate of genetic improvement per year.

The problem, in case of one-stage selection programme while envisaging a constant female strength from generation to generation, amounts to one of considering the optimal value of intensity of selection subject to the condition

$$pN + p\alpha_1 N = N$$

or

$$\alpha_1 = \frac{1-p}{p} \quad (7a)$$



where for every  $N$  cows bred,  $pN$  is the number of their daughters expected to complete their first lactation.

This equation admits solution of  $a_1$  for  $p \geq 0.5$ . But in dairy cows such a high rate of reproduction is not possible. In most cases, the value of  $p$  is expected to be in the neighbourhood of 0.4. For this value of  $p$ , equation (7 a) gives an impossible value of 1.5 for  $a_1$ . This means that a permissible value of  $a_1$  cannot be determined unless restriction of raising the same number of females in each generation is waived. In that case there will be a continuous decrease in the herd strength from generation to generation of the order of 1/5 to 3/5 times the previous generation number.

For two-stage selection, equation corresponding to (7 a) takes the form

$$a_1 + a_1 a_2 = \frac{1-p}{p} \quad (7 b)$$

Although this equation is solvable for  $a_1$  and  $a_2$ , the contribution to the expected percentage genetic advance will be quite low as compared to that under three-stage selection. This can be seen from Table I, *vide* sets 2, 3 and 7.

Keeping the above considerations in view the problem then reduces to one of considering the optimal values of intensities of selection at three different stages subject to the restriction that the same number of females are raised in each generation, *i.e.*,

$$a_1 + a_1 a_2 + a_1 a_2 a_3 = \frac{1-p}{p} \quad (7 c)$$

As already mentioned at the outset the principle of adopting the optimum procedure of selection is to maximise the genetic advance per year. This is given by the ratio of average genetic advance to the average generation interval. If the frequencies of selection are  $a_1, a_1 a_2$  and  $a_1 a_2 a_3$ , the average genetic advance  $A$  is given by

$$\frac{N a_1 G_1'(y) + N a_1 a_2 G_2'(y) + N a_1 a_2 a_3 G_3'(y)}{N + N a_1 + N a_1 a_2 + N a_1 a_2 a_3}$$

where  $G_1'(y)$ ,  $G_2'(y)$  and  $G_3'(y)$  are the amounts of genetic advance expected after first, second and third selection respectively. These are given by

$$G_1'(y) = \frac{\rho_1 Z_{(k_1)}}{\alpha_1} \sigma_1(y)$$

$$G_2'(y) = \frac{\rho_1 Z_{(k_1)} I_2 + \rho_2 Z_{(k_2)} I_1}{a_1 a_2} \sigma_2'(y)$$

$$G_3'(y) = \frac{\rho_1 Z_{(k_1)} I_{23} + \rho_2 Z_{(k_2)} I_{13} + \rho_3 Z_{(k_3)} I_{12}}{a_1 a_2 a_3} \sigma_3'(y)$$

where

$\sigma_1^2(y)$  is the variance of  $y$  in the unselected population,

$\sigma_2^2(y)$  is the variance of  $y$  among the units retained after first selection,

and

$\sigma_3^2(y)$  is the variance of  $y$  among the units retained after second selection.

Further, if  $g_1, g_2, g_3$  and  $g_4$  denote respectively the average ages of the parents when their first, second, third and fourth progenies are born, the average generation interval  $I$  will be given by the expression

$$\frac{Ng_1 + Na_1g_2 + Na_1a_2g_3 + Na_1a_2a_3g_4}{N + Na_1 + Na_1a_2 + Na_1a_2a_3}$$

Thus the problem reduces to one of finding such values of  $a_1, a_2$  and  $a_3$  satisfying equation (7c) as will maximise  $A/I$ , i.e.,

$$\frac{Na_1 G_1'(y) + Na_1a_2 G_2'(y) + Na_1a_2a_3 G_3'(y)}{Ng_1 + Na_1g_2 + Na_1a_2g_3 + Na_1a_2a_3g_4} \quad (8)$$

It is difficult to obtain a general solution to this problem. The only feasible means of arriving at approximate solution is an empirical approach of considering a range of sets of values of  $a_1, a_2$  and  $a_3$  satisfying equation (7c) and examining the values which maximize the average genetic gain per year. This approach is in itself fraught with considerable difficulty of numerical computation as will be seen from the following section.

## 6. SELECTION IN THREE STAGES

It is not possible to utilize the general formula derived in Section 4 for the case of more than three stages of selection since the tables of multivariate normal integrals is limited to trivariates only. We may consider the case of three variables in more specific detail.

In this case the general formula reduces to

$$G(y) = \frac{\rho_1 Z_{(k_1)} I_{23} + \rho_2 Z_{(k_2)} I_{13} + \rho_3 Z_{(k_3)} I_{12}}{a_1 a_2 a_3} \quad (9)$$

where  $k_1$ ,  $k_2$  and  $k_3$  will be found from the normal tables satisfying the following three equations:

$$a_1 = \frac{1}{\sqrt{2\pi}} \int_{k_1}^{\infty} e^{-\eta_1^2/2} d\eta_1;$$

$$a_1 a_2 = \frac{1}{2\pi \sqrt{1 - \rho_{12}^2}} \int_{k_1}^{\infty} d\eta_1 \int_{k_2}^{\infty} e^{-\left[\frac{1}{2(1 - \rho_{12}^2)}\right] [\eta_1^2 - 2\rho_{12}\eta_1\eta_2 + \eta_2^2]} d\eta_2;$$

$$a_1 a_2 a_3 = \frac{1}{(2\pi)^{3/2} \Delta^{1/2}} \int_{k_1}^{\infty} d\eta_1 \int_{k_2}^{\infty} d\eta_2 \int_{k_3}^{\infty} e^{-\frac{1}{2} \sum_{i=1}^3 \sum_{j=1}^3 A_{ij} \eta_i \eta_j} d\eta_3$$

where

$$A_{11} = \frac{1 - \rho_{23}^2}{\Delta}, \quad A_{22} = \frac{1 - \rho_{13}^2}{\Delta}, \quad A_{33} = \frac{1 - \rho_{12}^2}{\Delta}$$

$$A_{12} = \frac{\rho_{13}\rho_{23} - \rho_{12}}{\Delta}, \quad A_{13} = \frac{\rho_{12}\rho_{23} - \rho_{13}}{\Delta}, \quad A_{23} = \frac{\rho_{12}\rho_{13} - \rho_{23}}{\Delta}$$

and

$$\Delta = 1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 + 2\rho_{12}\rho_{13}\rho_{23}$$

$Z(k_1)$ ,  $Z(k_2)$  and  $Z(k_3)$  are the ordinates of the univariate normal curve corresponding to  $k_1$ ,  $k_2$  and  $k_3$ ; and  $I$ 's the incomplete volumes of the bivariate normal surface, where

$$I_{12} = I\left(\frac{k_1 - k_3\rho_{13}}{\sqrt{1 - \rho_{13}^2}}, \frac{k_2 - k_3\rho_{23}}{\sqrt{1 - \rho_{23}^2}}, \rho_{12.3}\right)$$

$$I_{23} = I\left(\frac{k_2 - k_1\rho_{21}}{\sqrt{1 - \rho_{12}^2}}, \frac{k_3 - k_1\rho_{31}}{\sqrt{1 - \rho_{13}^2}}, \rho_{23.1}\right)$$

and

$$I_{31} = I\left(\frac{k_3 - k_2\rho_{32}}{\sqrt{1 - \rho_{23}^2}}, \frac{k_1 - k_2\rho_{12}}{\sqrt{1 - \rho_{12}^2}}, \rho_{31.2}\right).$$

$k_1$  will be obtained from the univariate normal tables (Pearson, 1931) corresponding to  $a_1$ .

The value of  $k_2$  can be got from the bivariate normal tables given by Pearson (1931). But the use of these tables involves considerable amount of interpolation work. Tables computed by Owen (1956) overcomes this difficulty to some extent. However, the method of S. C. Das (1956) which consists in reducing bivariate integral to a single

integral which is then to be evaluated numerically seems more suitable for fixing the truncation point  $k_2$ .

The value of  $k_3$  can be fixed with the help of  $T$ -function tabulated by Owen (1956) and  $S$ -function tabulated by Steck (1958) coupled with univariate normal tables. However these tables which are better suited for evaluating the volume of the trivariate distribution given the range of integration are not very helpful for the reverse procedure of reading  $k_3$ .

Another general approach to the problem is by means of the tetrachoric series which has been generalized by M. G. Kendall (1941). From a theoretical point of view this solves the problem but in practice, since the tetrachoric series converges very slowly for large  $\rho_{ij}$  it is of little use.

The method of Plackett (1954) which expresses the trivariate integral as a sum of lower dimensional normal integrals and an integral which is to be evaluated by numerical integration too is not suited to our problem.

The procedure given by S. C. Das (1956) which consists in reducing the trivariate normal integral to a single integral which is then to be evaluated numerically, meets this situation. But this method is also limited in its scope since for it implies that the correlations  $\rho_{12}$ ,  $\rho_{13}$  and  $\rho_{23}$  should be such that their joint product is positive and each one is numerically greater than the product of the other two. For problems of selection in dairy cattle breeding  $\rho_{13}$  is always equal to the product of the other two as has been shown in a later section. The method is thus ruled out.

In the words of Peter Ihm (1959), who was concerned with the evaluation of multivariate normal integral, the most satisfying general method seems to be the Monte Carlo method by use of an electronic computer.

In special cases where the units under selection are all retained at one of the stages, the problem is much simplified. It reduces to two-stage selection scheme. A problem of this nature has been exemplified in the next section.

## 7. NUMERICAL ILLUSTRATION

Reverting to the problem outlined in Section 5, before the average genetic advance per year given by formula (8) for different sets of values

of  $\alpha_1$ ,  $\alpha_2$  and  $\alpha_3$  is found the estimates of different parameters appearing therein have to be obtained.

Let  $y$  denote the genotypic value of lactation yield of a cow and  $x_1$ ,  $x_2$  and  $x_3$  the phenotypic values for the first, second and third lactation yields respectively. These may be expressed as

$$x_1 = y + e_p + e_1$$

$$x_2 = y + e_p + e_2$$

$$x_3 = y + e_p + e_3$$

where  $e_p$  is the environmental error considered constant over different lactations and  $e$ 's are the errors due to environmental factors varying from lactation to lactation.

The variates  $y$ ,  $e_p$ , and  $e$ 's are assumed to be independently distributed. We may also assume

$$\sigma_{e_1}^2 = \sigma_{e_2}^2 = \sigma_{e_3}^2.$$

In that case  $\sigma_{x_1}^2 = \sigma_{x_2}^2 = \sigma_{x_3}^2 = \sigma_P^2$  (say).

By the theory of least squares,  $\eta_1$ ,  $\eta_2$  and  $\eta_3$  can then be shown to be equal to

$$\eta_1 = h_1^2 x_1$$

$$\eta_2 = \frac{h_1^2}{1+R} (x_1 + x_2)$$

$$\eta_3 = \frac{h_1^2}{1+2R} (x_1 + x_2 + x_3)$$

where  $h_1^2$  the coefficient of heritability and  $R$  the coefficient of repeatability are defined respectively as

$$\frac{\sigma_{(y)}^2}{\sigma_P^2} \quad \text{and} \quad \frac{\sigma_{(y)}^2 + \sigma_{e_p}^2}{(\sigma_{(y)}^2 + \sigma_{e_p}^2) + \sigma_e^2}$$

$\rho_1$ ,  $\rho_2$  and  $\rho_3$  can be shown easily to be equal to

$$\rho_1 = h_1$$

$$\rho_2 = h_1 \sqrt{\frac{2}{1+R}}$$

$$\rho_3 = h_1 \sqrt{\frac{3}{1+2R}}$$

Then  $\rho_{12}$ , the correlation between  $\eta_1$  and  $\eta_2$ , becomes  $\sqrt{(1+R)/2}$ . Likewise  $\rho_{13}$  and  $\rho_{23}$  can be shown to have values

$$\rho_{13} = \sqrt{\frac{1+2R}{3}}$$

$$\rho_{23} = \sqrt{\frac{2(1+2R)}{3(1+R)}}$$

(It can be seen that  $\rho_{13} = \rho_{12} \rho_{23}$ )

For computation we may assume the values of  $h_1^2$ ,  $R$ ,  $c$ , the coefficient of variation, and  $p$  as 0.3, 0.7, 40 and 0.4 respectively. These are close to the values obtained in the course of extensive studies on breeding data of herds of Indian cattle at live-stock farms. Assuming further, the values of age at first calving, age at first service and the calving interval as  $3\frac{1}{2}$  years,  $3\frac{3}{4}$  years and  $1\frac{1}{2}$  years respectively and that the cows are served by fresh set of bulls each time, in that case,  $g$ 's take the values  $g_1 = 4$  years,  $g_2 = 4\frac{3}{4}$  years,  $g_3 = 5\frac{1}{2}$  years and  $g_4 = 6\frac{1}{4}$  years.

The genetic variance  $\sigma^2(y)$  decreases with successive stages of selection, the magnitude of which depends upon the intensity of selection. However for computational convenience  $\sigma^2(y)$  has been assumed to remain unaltered under various stages of selection.

Eight sets of values of  $a_1$ ,  $a_2$  and  $a_3$  satisfying equation (7 c) have been considered in Table I. These sets cover the entire range of values of  $a$ 's, viz.,  $0.5 \leq a_1 \leq 1$ ,  $0.25 \leq a_2 \leq 1$ ,  $0 \leq a_3 \leq 1$ . It will be seen from the table that for cases of three-stage selection at least one of the  $a$ 's has been assigned its maximum limiting value, viz., unity thus fixing the corresponding point of truncation as  $-\infty$ . After fixing one of the  $k$ 's in the manner described above, the other two points could now easily be found by following S. C. Das's method for two variates referred to earlier. The other sets containing odd values of  $a_1$ ,  $a_2$  and  $a_3$  have been omitted as in those cases the fixation of  $k_3$  would have involved very heavy computation (*vide* Section 6).

Further computations are self-explanatory and can be followed easily step by step. Finally the expected average percentage genetic gain per year in  $y$  has been calculated from the formula (8) which after simplification reduces to:

$$\frac{\rho_1 Z_{(k_1)}(1 + I_2 + I_{23}) + \rho_2 Z_{(k_2)}(I_1 + I_{13}) + \rho_3 Z_{(k_3)} I_{12}}{g_1 + a_1 g_2 + a_1 a_2 g_3 + a_1 a_2 a_3 g_4} \times h_1 c$$

TABLE I

Expected average percentage genetic advance per year

Sl. No.	Proportions retained			Points of truncation			Normal ordinates at the points of truncation		
	$a_1$	$a_2$	$a_3$	$k_1$	$k_2$	$k_3$	$Z(k_1)$	$Z(k_2)$	$Z(k_3)$
1	2	3	4	5	6	7	8	9	10
1	1	0.25	1	$-\infty$	0.674490	$-\infty$	0	0.317776	0
2	1	0.50	..	$-\infty$	0	..	0	0.398942	..
3	0.75	1	..	-0.674500	$-\infty$	..	0.317776	0	..
4	0.50	1	1	0	$-\infty$	$-\infty$	0.398942	0	0
5	1	0.36	0.36	$-\infty$	0.368459	1.175000	0	0.374118	0.200040
6	0.58	1	0.58	-0.201900	$-\infty$	0.400000	0.390894	0	0.368269
7	0.82	0.82	..	-0.915400	-0.481700	..	0.262400	0.355237	..
8	0.65	0.65	1	-0.385300	0.166200	$-\infty$	0.370399	0.393470	0

$\frac{k_1 - \rho_{12}k_2}{\sqrt{1 - \rho_{12}^2}}$ = $L$ (say)	$\frac{k_2 - \rho_{12}k_1}{\sqrt{1 - \rho_{12}^2}}$ = $M$ (say)	$I_1 = \frac{1}{\sqrt{2\pi}} \int_L^{\infty} e^{-t^2/2} dt$	$I_2 = \frac{1}{\sqrt{2\pi}} \int_M^{\infty} e^{-t^2/2} dt$	$\frac{k_1 - k_3\rho_{13}}{\sqrt{1 - \rho_{13}^2}}$ = $K_1$ (say)	$\frac{k_2 - k_3\rho_{23}}{\sqrt{1 - \rho_{23}^2}}$ = $K_2$ (say)	$\frac{k_2 - k_1\rho_{21}}{\sqrt{1 - \rho_{21}^2}}$ = $K_3$ (say)	$\frac{k_3 - k_1\rho_{31}}{\sqrt{1 - \rho_{31}^2}}$ = $K_4$ (say)	$\frac{k_3 - k_2\rho_{32}}{\sqrt{1 - \rho_{32}^2}}$ = $K_5$ (say)	$\frac{k_1 - k_2\rho_{12}}{\sqrt{1 - \rho_{12}^2}}$ = $K_6$ (say)
11	12	13	14	15	16	17	18	19	20
$-\infty$	$-\infty$	1	0	$-\infty$	$+\infty$	$+\infty$	$\infty - \infty$	$-\infty$	$-\infty$
$-\infty$	$+\infty$	1	0	..	..	..	..	..	..
$+\infty$	$-\infty$	0	1	..	..	..	..	..	..
$+\infty$	$-\infty$	0	1	$+\infty$	$\infty - \infty$	$-\infty$	$-\infty$	$-\infty$	$+\infty$
$-\infty$	$+\infty$	1	0	$-\infty$	-3.222500	$+\infty$	$+\infty$	3.411300	$-\infty$
$+\infty$	$-\infty$	0	1	-1.251200	$-\infty$	$-\infty$	1.298000	$+\infty$	$+\infty$
-1.216900	0.935300	0.888178	0.174818	..	..	..	..	..	..
-1.390400	1.346300	0.917796	0.089105	$+\infty$	$+\infty$	1.346300	$-\infty$	$-\infty$	-1.390469

$\rho_1 = 0.547723$ ;  $\rho_2 = 0.594089$ ;  $\rho_3 = 0.612373$ .

TABLE I—(Contd.)

$I_{12}=I(K_1, K_2; \rho_{12.3})$	$I_{23}=I(K_3, K_4; \rho_{23.1})$	$I_{31}=I(K_5, K_6; \rho_{13.2})$	$\rho_1 Z_{(k_2)} (1+I_2+I_{23})$	$\rho_2 Z_{(k_2)} (I_1+I_{13})$	$\rho_3 Z_{(k_3)} I_{12}$
21	22	23	24	25	26
0	0	1	0	0.377574	0
..	..	..	0	0.237007	..
0	..	..	0.348106	0	..
0	1	0	0.655530	0	0
1	0	0	0	0.222259	0.122499
0.894569	0.097145	0	0.449003	0	0.201741
..	..	..	0.168847	0.187443	..
0	0.089105	0.917807	0.239030	0.429083	0

$\rho_1 Z_{(k_2)} (1+I_2+I_{23})$ $+ \rho_2 Z_{(k_2)} (I_1+I_{13})$ $+ \rho_3 Z_{(k_3)} I_{12}$ (24) + (25) + (26)	$g_1+g_2a_1$ $+g_3a_1a_2$ $+g_4a_1a_2a_3$	Percentage genetic advance per year		Bounds to the percentage genetic advance per year			
		$\frac{(27)}{(28)} \times 40 h_1$	Rank	Upper bound		Lower bound	
				$\left( \frac{h_1 Z_1' + h_2 Z_2' + h_3 Z_3'}{g_1+g_2a_1+g_3a_1a_2+g_4a_1a_2a_3} \right) \times 40 h_1$		$\left( \frac{h_1 (Z_1' + Z_2' + Z_3')}{g_1+g_2a_1+g_3a_1a_2+g_4a_1a_2a_3} \right) \times 40 h_1$	
27	28	29	30	31		32	33
0.377574	11.6875	0.71	4	0.72		0.65	4
0.237007	11.5000	0.45	8	0.45		0.42	8
0.348106	11.6875	0.65	7	0.68		0.65	5
0.655530	12.2500	1.17	3	1.25		1.17	1
0.344768	11.5400	0.65	6	0.67		0.61	7
0.650744	12.0474	1.18	2	1.22		1.14	3
0.356290	11.5931	0.67	5	0.68		0.64	6
0.668113	12.0518	1.22	1	1.23		1.15	2

$\rho_{13}=0.894427; \rho_{23}=0.970143.$

$\rho_{12.3}=0.500000, \rho_{23.1}=0.840000; \rho_{31.2}=0$



From Table I, column 29, it is seen that the scheme number 8, *i.e.*,  $a_1 = 0.65 = a_2$  and  $a_3 = 1$  is the best set to adopt for selection programme as this results in maximum average percentage genetic advance out of all the eight different schemes considered here.

### 8. BOUNDS TO GENETIC ADVANCE

It will be observed that the use of the foregoing formula even for three-stage selection programme involves very cumbersome integrals which are not easy to evaluate. Beyond stage three, we require multi-variate normal tables for fixing the values of truncation points. These are not available at present. To overcome these difficulties, various empirical approximations were tried and two simple methods giving respectively the upper and lower bounds to the genetic advance were secured. In the first of these approximations, the amount of genetic advance expected at different successive stages of selection is taken to be

$$G_1''(y) = h_1 \frac{Z_1'}{a_1} \sigma_1(y)$$

$$G_2''(y) = h_2 \frac{Z_2'}{a_1 a_2} \sigma_2(y)$$

$$G_3''(y) = h_3 \frac{Z_3'}{a_1 a_2 a_3} \sigma_3(y)$$

.....  
 .....  

$$G_r''(y) = h_r \frac{Z_r'}{a_1 a_2 \dots a_r} \sigma_r(y)$$

where

$Z_1'$  = the normal ordinate corresponding to  $a_1$ ,

$Z_2'$  = the normal ordinate corresponding to  $a_1 a_2$ ,

.....  
 $Z_r'$  = the normal ordinate corresponding to  $a_1 a_2 \dots a_r$ ,

$h_1^2$  = the coefficient of heritability based on first lactation records,

$h_2^2$  = the coefficient of heritability based on first two lactation records =  $2h_1^2/(1+R)$ ,

.....

$h_r^2$  = the coefficient of heritability based on first  $r$  lactation records =  $rh_1^2/[1+(r-1)R]$ ,

$c$  = coefficient of variation of the lactation yield,

$R$  = the coefficient of repeatability,

$\sigma_1^2(y)$ ,  $\sigma_2^2(y)$ , etc., have already been defined in Section 5.

Ignoring as before the reduction in genetic variance from stage to stage the corresponding form for average percentage genetic advance in this case reduces to:

Average percentage genetic advance per year

$$\bar{G}_r''(y) = \frac{h_1 Z_1' + h_2 Z_2' + \dots + h_r Z_r'}{g_1 + g_2 a_1 + g_3 a_1 a_2 + \dots + g_{r+1} a_1 a_2 \dots a_r} \times h_1 c. \quad (10)$$

It will be seen that the expression for  $\bar{G}_r''(y)$  strictly holds when selection is practised on the basis of first  $r$  records of all cows and a proportion ( $a_1 a_2 \dots a_r$ ) of the best cows from the original population is retained while affecting selection at the  $r$ -th stage. But in practice the selection will be based on a more limited information inasmuch as the earlier cullings would have already been made on the basis of fewer lactation records and as such the advance is likely to be smaller. However, the approximation is of interest as it provides an upper bound to the gain that may accrue from selection on the completion of successive lactations.

If on the other hand, cognisance is not taken of the information provided by the previous ( $r-1$ ) records, the corresponding expression would give a lower bound to the selection gain. The expression for the lower bound is

$$G_r''(y) = \frac{h_1 (Z_1' + Z_2' + \dots + Z_r')}{g_1 + g_2 a_1 + g_3 a_1 a_2 + \dots + g_{r+1} a_1 a_2 \dots a_r} \times h_1 c.$$

The values of the bounds for the cases worked out in Section 7 have been given in columns 31 and 32 of Table I. The two sets of values clearly enclose the true values got by the exact procedure and at the same time reflect the closeness in results achieved by these two methods for different sets of values of  $a_1$ ,  $a_2$  and  $a_3$ .

An attempt was made to simplify the task of locating the optimum of  $a$ 's based on the principle of maximising the average genetic gain per year. The line of attack was as follows:

Consider first the former approximation.

Let  $t_1, t_2 \dots$  and  $t_r$  be the points of truncation of the normal curve corresponding to the frequencies of selection  $a_1, (a_1 a_2) \dots$  and  $(a_1 a_2 \dots a_r)$  respectively. If  $\phi_r$  denotes the average percentage genetic advance per year, the expression  $\phi_r$  in terms of  $t$ 's can be written as

$$\phi_r = \frac{\frac{h_1}{\sqrt{2\pi}} e^{-t_1^2/2} + \frac{h_2}{\sqrt{2\pi}} e^{-t_2^2/2} + \dots + \frac{h_r}{\sqrt{2\pi}} e^{-t_r^2/2}}{g_1 + g_2 \frac{1}{\sqrt{2\pi}} \int_{t_1}^{\infty} e^{-t^2/2} dt + g_3 \frac{1}{\sqrt{2\pi}} \int_{t_2}^{\infty} e^{-t^2/2} dt + \dots + g_{r+1} \frac{1}{\sqrt{2\pi}} \int_{t_r}^{\infty} e^{-t^2/2} dt} \times h_1 c.$$

The  $a$ 's or by implication  $t$ 's are to be so chosen that  $\phi_r$  is maximum subject to the restraint corresponding to (7), viz., that the strength of adult females stock remains constant from generation to generation, i.e.

$$a_1 + a_1 a_2 + \dots + a_1 a_2 \dots a_r = \frac{1-p}{p}$$

which in terms of  $t$ 's is equivalent to

$$\frac{1}{\sqrt{2\pi}} \int_{t_1}^{\infty} e^{-t^2/2} dt + \frac{1}{\sqrt{2\pi}} \int_{t_2}^{\infty} e^{-t^2/2} dt + \dots + \frac{1}{\sqrt{2\pi}} \int_{t_r}^{\infty} e^{-t^2/2} dt - \frac{1-p}{p} = K \text{ (say).}$$

Maximising  $\phi_r$  subject to the above condition is the same thing as maximising  $\phi_r + \lambda K$  where  $\lambda$  is the Lagrange's multiplier.

Differentiating partially with respect to  $t_1, t_2, \dots, t$  and  $\lambda$  equating them to zero respectively, we obtain

$$g_2 \phi_r - (h_1 c) h_1 t_1 - \lambda \left[ g_1 + \frac{g_2}{\sqrt{2\pi}} \int_{t_1}^{\infty} e^{-t^2/2} dt + \dots + \frac{g_{r+1}}{\sqrt{2\pi}} \int_{t_r}^{\infty} e^{-t^2/2} dt \right] = 0$$

$$g_3 \phi_r - (h_1 c) h_2 t_2 - \lambda \left[ g_1 + \frac{g_2}{\sqrt{2\pi}} \int_{t_1}^{\infty} e^{-t^2/2} dt + \dots + \frac{g_{r+1}}{\sqrt{2\pi}} \int_{t_r}^{\infty} e^{-t^2/2} dt \right] = 0$$

.....  
 .....

$$g_{r+1} \phi_r - (h_1 c) h_r t_r - \lambda \left[ g_1 + \frac{g_2}{\sqrt{2\pi}} \int_{t_1}^{\infty} e^{-t^2/2} dt + \dots + \frac{g_{r+1}}{\sqrt{2\pi}} \int_{t_r}^{\infty} e^{-t^2/2} dt \right] = 0$$

and

$$\frac{1}{\sqrt{2\pi}} \int_{t_1}^{\infty} e^{-t^2/2} dt + \frac{1}{\sqrt{2\pi}} \int_{t_2}^{\infty} e^{-t^2/2} dt + \dots + \frac{1}{\sqrt{2\pi}} \int_{t_r}^{\infty} e^{-t^2/2} dt - \frac{1-p}{p} = 0.$$

These  $(r+1)$  equations are to be solved for  $t_1, \dots, t_r$  and  $\lambda$  in terms of known parameters, viz.,  $h$ 's,  $g$ 's,  $p$  and  $c$ . An exact solution of  $t$ 's is difficult to attempt.

If instead second approximation is used the only change in the above  $(r+1)$  equations would be to replace different  $h$ 's by  $h_1$  and the difficulties in obtaining a solution remain as before.

In the absence of getting the optimum set algebraically the only recourse which can be had is to locate it empirically by trying out different likely sets of selection intensities. Looking at the values of upper and lower bounds of percentage annual genetic gain for different sets of values of  $\alpha$ 's given in Table I, sets numbered 4, 6 and 8 are possibly close contenders. The annual genetic gain for the non-degenerate cases near about these have been worked out in Table II.

TABLE II  
*Bounds to the percentage genetic advance per year*

Sl. No.	Proportions retained			Percentage annual genetic gain	
	$\alpha_1$	$\alpha_2$	$\alpha_3$	Upper bound	Lower bound
1	0.550	0.900	0.920	1.2528	1.1734
2	0.600	0.850	0.760	1.2391	1.1606
3	0.600	0.800	0.875	1.2436	1.1645
4	0.650	0.800	0.640	1.2061	1.1296
5	0.650	0.750	0.750	1.2179	1.1401
6	0.700	0.750	0.600	1.1592	1.0847
7	0.800	0.500	0.750	1.0991	1.0247
8	0.750	0.750	0.330	1.0604	0.9944
9	0.700	0.700	0.630	1.1789	1.1033
10	0.650	0.650	1.000	1.2272	1.1482
11	0.500	1.000	1.000	1.2516	1.1724

The results in both the tables show that the two approximate methods lead to the same ranking of alternative procedures of selection in regard to genetic gain per year. It will also be further seen that this ranking is not very different (in the degenerate cases studied) from that according to the exact procedure. Further the two extreme values of percentage genetic advance per year for the same set are quite close to each other in as much as the width of the interval is hardly about 6 per cent of the lower value on an average. Thus it may be concluded that in practical situations the second approximate procedure may be adopted in locating the optimum set of values for selection procedure and bounds to the optimum genetic gain per year obtained by using both the approximate procedures.

## 9. SUMMARY

An expression for the gain in genetic advance for multivariate normal populations under successive stages of selection has been derived. The problem of three-stage selection has been dealt with in details with particular reference to animal breeding and an example has been furnished to illustrate the procedure.

Alternative approximate methods for estimating genetic advance which are easily amenable to calculations have been given.

## 10. REFERENCES

1. Cochran, W. G. .. "Improvement by means of selection," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, 1951, 449-70.
2. Das, S. C. .. "The numerical evaluation of a class of integrals, II," *Proc. Cambridge Phil. Soc.*, 1956, 52, 442-48.
3. Dickerson, G. E. and Hazel, L. N. "Effectiveness of selection on progeny performance as a supplement to earlier culling in livestock," *Jour. Agr. Res.*, 1944, 69, 459-76.
4. Finney, D. J. .. "Statistical problems of plant selection," *Proceedings of the 30th Session of the International Statistical Institute*, 1957.
5. Hazel, L. N. .. "The genetic basis for constructing selection indices," *Genetics*, 1943, 28, 476-90.
6. Ihm, Peter .. "Numerical evaluation of certain multivariate normal integrals," *Sankhya*, 1959, 21, 363-66.

7. Kendall, M. G. .. "Proof of relations connected with the tetra-choric series and its generalisation," *Biometrika*, 1941, **32**, 196-98.
8. Keuls, M. and Sieben, J. W. "Two statistical problems in plant selection," *Euphytica*, 1955, **4**, 34-44.
9. Owen, D. B. .. "Tables for computing bivariate normal probabilities," *Ann. Math. Stat.*, 1956, **27**, 1075-1090.
10. Panse, V. G. .. "An application of the discriminant function for selection in poultry," *J. Genetics*, 1946, **47**, 242-48.
11. Pearson, Karl .. *Tables for Statisticians and Biometricians—Part II*. Cambridge University Press, 1931.
12. Plackett, R. L. .. "A reduction formula for normal multivariate integrals," *Biometrika*, 1954, **41**, 351-60.
13. Sieben, J. W. .. "Selectie bij de plantenveredeling op grond van proefveld resultaten," *Statistica Neerlandica*, 1954, **8**, 179-89.
14. Smith, H. F. .. "A discriminant function for plant selection," *Ann. of Eugenics*, 1936, **7**, 240-50.
15. Steck, G. P. .. "A table for computing trivariate normal probabilities," *Ann. of Math. Stat.*, 1958, **29**, 780-800.