Theor. Appl. Genet. 50, 153-161 (1977)

Theory for the Number of Genes Affecting Quantitative Characters*

I. Estimation of and Variance of the Estimation of Gene Number for Quantitative Traits Controlled by Additive Genes Having Equal Effect

Y.C. Park**
Department of Statistics, Florida State Univeristy, Tallahassee, Florida (U.S.A.)

Summary. A general expression for gene number estimation which encompasses the conventional formula was derived. It provides a basis for gene number estimation from the data of recurrent selection experiments that are not of sufficient duration to measure total response to selection.

Gene number estimates are considerably more reliable when heritability is high. The effect of heritability on sampling variance is particularly important when gene number is large.

Generally the most effective ways of decreasing the variance of a gene number estimate will be 1) to increase the number of generations in a primary selection program, 2) to increase the number of generations in the two way selection program and 3) to increase population size.

Key words: Gene Number - Estimation of Gene Number - Variance of the Estimation of Gene Number

Introduction

The number of genes involved in the inheritance of quantitative characters influences the limits of progress from recurrent selection. Falconer (1960) stated that, with a given amount of initial variation, a small number of genes will produce less response than a large number and if a given amount of variation is produced by few genes their effects are greater than if many genes were involved.

Gene number estimates must be obtained indirectly since the genes affecting a quantitative character cannot be observed directly. Methods used included 1) the identification of genetic effects related to specific segments of chromosomes (Thoday and Boam 1959; Thoday 1961; Speickett and Thoday 1966; Robertson 1967), 2) comparison of theoretical and observed distributions when simultaneously considering different generations, e.g., pure line parents, F_1, F_2 and backcrosses (Powers 1934, 1963; Gates 1963), and 3) inference from phenotypic means and genetic variances using materials in which the phenotypic means provide information concerning genetic extremes (Castle 1921; Wright 1934, 1952, 1968; Falconer 1960; Roberts 1966; Hill and Robertson 1966; Comstock 1969; Mather and Jinks 1971). The last method usually involves long term selection experiments or pure lines with genetic extremes and F_1 and F_2 from the pure lines. Estimates by this method are rarely available for economic species because of the genetic extremes and the long generation intervals.

We estimate gene number from recurrent selection experiments in a population obtained by crossing any pair of inbred lines. The proposed estimation procedure does not require that total response to selection be determined. The variance of the estimate is derived and factors affecting its size are discussed.

A Basis for Gene Number Estimation When Gene Effects Are Equal and Additive

When there is no epistasis and linkage disequilibrium, the average contribution of genotypes from all segregating loci ($\bar{y}$) and the additive genetic variance (σ_g^2)

* From a thesis submitted by the author in partial fulfillment of the requirements for the Ph.D. degree. Received March, 1975. Work supported by Public Health Service Grant GM 16074, by the Minnesota Agricultural Experiment Station and by National Institutes of Environmental Health Sciences Grant No. 5T32 ES07011-02.

** Former Research Assistant, Genetics and Cell Biology, University of Minnesota; Currently Postdoctoral Fellow in Environmental Health Measurement and Statistics.

are (Comstock and Robinson 1948):

$$\bar{y} = \sum_{i=1}^{n} (2q_i - 1)u_i + \sum_{i=1}^{n} 2q_i(1 - q_i)a_i u_i$$

$$\sigma_g^2 = \sum_{i=1}^{n} 2q_i(1 - q_i)[1 + (1 - 2q_i)a_i]^2 u_i^2 ,$$

where q_i is the frequency of a favorable allele at the ith locus, u_i is onehalf the difference in effect between the two homozygotes, a_i is a measure of dominance at the ith locus and n is the number of segregating loci. Considering the special situation in which there is no dominance, all u's are equal, and all q's are equal, the two quantities reduce to:

$\bar{y} = n(2q - 1)u$ and

$$\sigma_g^2 = n\,2q(1 - q)u^2 .$$

Taking derivatives and removing q:

$$n = -\bar{y}\,\frac{d\bar{y}}{d\sigma_g^2} .$$

Integrating the differential equation

$$\int_{\sigma_{gb}^2}^{\sigma_{gc}^2} n\,d\sigma_g^2 = \int_{\bar{y}_b}^{\bar{y}_c} -\bar{y}\,d\bar{y}$$

results in a difference equation

$$n = \frac{\bar{y}_c^2 - \bar{y}_b^2}{2(\sigma_{gb}^2 - \sigma_{gc}^2)} , \qquad \ldots\ldots\ldots\ldots (1)$$

where b and c identify any two generations in a recurrent selection program with $c > b$.

When $q_b = 0.5$, and $q_c = 1.0$ or 0.0, the above formula reduces:

$$n = \frac{R^2}{2\sigma_{go}^2} ,$$

where R represents the maximum response attainable by selection upward (or downward) and σ_{go}^2 represents the initial additive genetic variance when gene frequency is 0.5 at all segregating locis. This is the formula described by Comstock (1969). When D represents the range of values between extremes of up and down lines in a selection experiment, one half of D is equivalent to R. Hence

$$n = \frac{(D/2)^2}{2\sigma_{go}^2}$$

$$= \frac{D^2}{8\sigma_{go}^2} ,$$

which is the formula described by Falconer (1960). Since D is equivalent to the difference between the extreme plus and minus types and σ_{go}^2 can be estimated by $\sigma_{F2}^2 - \sigma_{F1}^2$, $\sigma_{F2}^2 - \sigma_p^2$, and $\sigma_{F2}^2 - \sigma_E^2$, the above formula can be expressed as follows:

$$n = \frac{D^2}{8(\sigma_{F2}^2 - \sigma_{F1}^2)} \qquad \text{(Castle 1921),}$$

$$= \frac{D^2}{8(\sigma_{F2}^2 - \sigma_p^2)} \qquad \text{(Wright 1934), or}$$

$$= \frac{D^2}{8(\sigma_{F_2}^2 - \sigma_E^2)} \qquad \text{(Wright 1968),}$$

where σ_{F2}^2, σ_{F1}^2 and σ_p^2 represent the total variances among individuals in F_2, F_1 and the parental strains respectively. σ_E^2 is the variance among individuals due to environmental effects.

Equation (1) can be considered as a general form which encompasses the formula described by Castle (1921), Comstock (1969), Falconer (1960), Wright (1934, 1952, 1968) etc.

Variance of the Gene Number Estimate

Variance of a quantity (Z) that is a function of two random variables (U, V) is approximately (Johnson and Kotz 1969):

$$\sigma_Z^2 = \left(\frac{\delta F}{\delta U}\right)^2 \sigma_U^2 + \left(\frac{\delta F}{\delta V}\right)^2 \sigma_V^2 + 2\left(\frac{\delta F}{\delta U}\right)\left(\frac{\delta F}{\delta V}\right)\sigma_{UV} \cdots\cdots \qquad (2)$$

The estimate of gene number from equation (1) can be considered as a function of two variables

$$n = \frac{U}{2V},$$

where $U = \bar{y}_c^2 - \bar{y}_b^2$ and

$$V = \sigma_{gb}^2 - \sigma_{gc}^2 .$$

Using the partial derivatives in equation (2),

$$\sigma_{\hat{n}}^2 = \left(\frac{1}{2V}\right)^2 \sigma_{\hat{U}}^2 + \left(\frac{U}{2V^2}\right)^2 \sigma_{\hat{V}}^2 - \left(\frac{U}{2V^3}\right) \sigma_{\hat{U}\hat{V}} \cdots\cdots, \quad (3)$$

where $\hat{}$ indicates an estimate.

In practice $Y = \bar{y}_c^2 - \bar{y}_b^2 = (\bar{y}_c + \bar{y}_b)(\bar{y}_c - \bar{y}_b)$ from one way selection. The primary selection program can be upward or downward. Let $U = XG$ where $X = \bar{y}_c + \bar{y}_b$ and $G = \bar{y}_c - \bar{y}_b$. Applying the partial derivatives to equation (2) once more we obtain

$$\sigma_{\hat{U}}^2 = G^2 \sigma_{\hat{X}}^2 + X^2 \sigma_{\hat{G}}^2 + 2XG\sigma_{\hat{X}\hat{G}} \cdots\cdots\cdots \quad (4)$$

V is the difference of the additive genetic variances between the two generations b and c. The most efficient method of estimating the additive genetic variance is by the product of the realized heritability and the phenotypic variance estimates (Hill 1972). Assuming the phenotypic variance will be the same in generations b and c, we can write

$$V = (R_b - R_c)\sigma_p^2 ,$$

where R is the realized heritability and σ_p^2 is the phenotypic variance. Let

$$S = R_b - R_c \quad \text{and}$$

$$Q = \sigma_p^2 .$$

Then

$$\sigma_{\hat{V}}^2 = Q^2 \sigma_{\hat{S}}^2 + S^2 \sigma_{\hat{Q}}^2 + 2SQ\sigma_{\hat{S}\hat{Q}} \cdots\cdots\cdots \quad (5)$$

It is shown below that both covariances $\sigma_{\hat{U}\hat{V}}$ and $\sigma_{\hat{S}\hat{Q}}$ are zero. Substituting equations (4) and (5) into (3) we obtain

$$\sigma_{\hat{n}}^2 = \left(\frac{1}{2V}\right)^2 [G^2\sigma_{\hat{X}}^2 + X^2\sigma_{\hat{G}}^2 + 2XG\sigma_{\hat{X}\hat{G}}] + \left(-\frac{U}{2V}\right)^2 [Q^2\sigma_{\hat{S}}^2 + S^2\sigma_{\hat{Q}}^2] \cdots\cdots \quad (6)$$

The real values of U, V, G, X, S and Q given the assumptions basic to equation (1) will be

$$U = \bar{y}_c^2 - \bar{y}_b^2$$

$$= 4n^2u^2(q_c - q_b)(q_c + q_b - 1),$$

$$V = \sigma_{gb}^2 - \sigma_{gc}^2$$

$$= n2u^2(q_c - q_b)(q_c + q_b - 1),$$

$$X = \bar{y}_c + \bar{y}_b$$

$$= 2nu(q_c + q_b - 1),$$

$$G = \bar{y}_c - \bar{y}_b$$

$$= 2nu(q_c - q_b) \quad \text{and}$$

$$S = (\sigma_{gb}^2 - \sigma_{gc}^2)/\sigma_p^2 = v/\sigma_p^2$$

$$= 2nu^2(q_c - q_b)(q_c + q_b - 1)/\sigma_p^2 .$$

Substituting these in equation (6),

$$\sigma_{\hat{n}}^2 = \left[\frac{1}{2u(q_c + q_b - 1)}\right]^2 \sigma_{\hat{X}}^2 + \left[\frac{1}{2u(q_c - q_b)}\right]^2 \sigma_{\hat{G}}^2 + \left[\frac{1}{2u^2(q_c - q_b)(q_c + q_b - 1)}\right] \sigma_{\hat{X}\hat{G}} + \left[\frac{\sigma_p^2}{2u^2(q_c - q_b)(q_c + q_b - 1)}\right]^2 \sigma_{\hat{S}}^2 + \left[\frac{n}{\sigma_p^2}\right]^2 \sigma_{\hat{Q}}^2 . \quad (7)$$

X and G can be estimated in terms of the population means for generations 0, b and c. In genral

$$\overline{P} = \overline{Y} + e ,$$

where $\overline{P}$ is the population's phenotypic mean, $\overline{Y}$ is the true population genotypic mean and e is the deviation of $\overline{P}$ from $\overline{Y}$. Also

$$E(\overline{P}) = \overline{Y} = \alpha + \bar{y}$$

because $E(e) = 0$. It follows that

$$E(\overline{P}_b) = \alpha + \bar{y}_b$$

$$E(\overline{P}_c) = \alpha + \bar{y}_c \text{ and}$$

$$E(P_0) = \alpha + \bar{y}_0 ,$$

where $\bar{y}_0 = n(2q_0 - 1)u = 0$, since $q_0 = 1/2$. Therefore, $E(\overline{P}_b + \overline{P}_c - 2\overline{P}_0) = \bar{y}_b + \bar{y}_c = X$ and $E(\overline{P}_c - \overline{P}_b) = \bar{y}_c - \bar{y}_a = G$, so X and G can be estimated as:

$$\hat{X} = \overline{P}_b + P_c - 2\overline{P}_0 \text{ and}$$

$$\hat{G} = \overline{P}_c - \overline{P}_b .$$

If there are no covariances among $\overline{P}_0$, $\overline{P}_b$ and $\overline{P}_c$

$$\sigma^2_{\hat{X}} = V(\overline{P}_b) + V(\overline{P}_c) + 4V(\overline{P}_0) , \quad (8)$$

$$\sigma^2_{\hat{G}} = V(\overline{P}_b) + V(\overline{P}_c) \text{ and} \quad (9)$$

$$\sigma_{\hat{X}\hat{G}} = V(\overline{P}_c) - V(\overline{P}_b) , \quad (10)$$

when $b \neq 0$. In the special case where $b = 0$,

$$\hat{X} = \hat{G} = \overline{P}_c - \overline{P}_0 \text{ and}$$

$$\sigma^2_{\hat{X}} = \sigma^2_{\hat{G}} = \sigma_{\hat{X}\hat{G}} = V(\overline{P}_c) + V(\overline{P}_b) . \quad (11)$$

Consider now the variance of the estimate of $S = R_b - R_c$. The standard procedure for estimating realized heritability is to divide and estimate of genetic response to selection by the total selection differential employed in achieving the response. The two way selection program for the estimation of the additive genetic variances at generations b and c is more practical than the one way program for two reasons. First, more selection can be accomplished in a given period of time. Second, the populations are compared to obtain the required estimate of genetic response, become available at the same time and can be compared in the same macro-environment (Comstock and Moll 1963). The advantage is that macro-environment effects are thereby eliminated as a source of error in the estimate of genetic response. Let us assume that, starting from generation c, two-way selection is practiced for two generations. Let

$\overline{P}_{ch}$ = the observed phenotypic mean of the population resulting from the two generations of selection for high performance,

and $\overline{P}_{cl}$ = the observed phenotypic mean of the population resulting from two generations of selection for low performance.

Then $\overline{P}_{ch} - \overline{P}_{cl}$ estimates the total genetic response. Hence

$$\overline{P}_{ch} - \overline{P}_{cl} = \bar{y}_{ch} - \bar{y}_{cl} + e_{ch} - e_{cl} , \quad (12)$$

where $\bar{y}_{ch}$ is the coded genotypic mean for the second (final) set of selected parents in the population selected for high performance and $\bar{y}_{cl}$ is the mean of parents selected for low performance.

The following subdivision of $\bar{y}_{ch} - \bar{y}_{cl}$ will be useful:

$$\bar{y}_{ch2} - \bar{y}_{c12} = H2 + H1 - L1 - L2,$$

where $H1 = \bar{y}_{ch1} - \bar{y}_c$, $L2 = \bar{y}_{c12} - \bar{y}_{c11}$, etc. Because no covariances are to be exprected among H2, H1, L1 and L2, the variance of $\bar{y}_{ch2} - \bar{y}_{c12}$ is

$$V(H2) + V(H1) + V(L1) + V(L2) .$$

The variances of single generation responses [V(H1), V(L1), etc.] around their exprected values are due to deviations of genotypic values from regression on phenotype in the selected trait. Given only additive gene effects (the situation being considered)

$$\sigma^2_{y \cdot p} = \sigma^2_y - \left(\frac{\sigma_{yp}}{\sigma^2_p}\right)^2 \sigma^2_p = \sigma^2_g(1 - h^2)$$

where h^2 is heritability of the selected trait and because $\sigma_{yp} = \sigma^2_y = \sigma^2_g$. It follows that the variance around the expectation for the average of N selected individuals will be

$$\sigma^2_g(1 - h^2)/N .$$

However, given m individuals selected to be male parents and f selected to be female parents, the variance of the effective mean deviation will be

$$\frac{1}{4}\left[\frac{\sigma^2_g(1 - h^2)}{m} + \frac{\sigma^2_g(1 - h^2)}{f}\right] = \frac{m + f}{4mf}(1 - h)\sigma^2_g .$$

For two generations of two way selection

$$V(\bar{y}_{ch} - \bar{y}_{cl}) = 4\sigma_g^2(1 - h^2)/N$$

if numbers selected are the same in both populations in both generations and the change in σ_g^2 between generations is assumed to be trivial. In general, given x generations of two way selection

$$V(\bar{y}_{ch} - \bar{y}_{cl}) = 2x\sigma_g^2(1 - h^2)/N \ , \qquad (13)$$

where $N = 4mf/(m+f)$ when selected male parents are distinct from selected female parents. The number of generations (x) should be small enough so the assumption that changes in σ_g^2 will be trivial is justified.

The quantity symbolized as "e" in equation (12) has both a genetic and a nongenetic component. However, if we assume that each parent has the same number of offspring as every other parent of the same sex

$$V(e) = \frac{B}{f} + \frac{W}{f \cdot k} \ , \qquad (14)$$

where w is the total variance among individuals within full-sib families, B is the non genetic variance among full-sib families, f is number of female parents and k is number of offpsing per female parent. The total variance of the estimate of genetic response is obtained by combining (13) and (14)

$$V(\bar{P}_{ch} - \bar{P}_{cl}) = \frac{2x\sigma_g^2(1 - h^2)}{N} + 2\left[\frac{B}{f} + \frac{W}{fk}\right] . \qquad (15)$$

Then since $S = R_b - R_c$ where $R_c = (\bar{R}_{ch} - \bar{R}_{cl})/D_c$ and D_c is the total of the selection differentials in the two way selection

$$\sigma_{\hat{s}}^2 = \frac{1}{D_b^2}\left[\frac{2x}{N}(1 - h_b^2)\sigma_{gb}^2 + \frac{2}{f}\left(B + \frac{W_b}{k}\right)\right]$$

$$+ \frac{1}{D_c^2}\left[\frac{2x}{N}(1 - h_c^2)\sigma_{gc}^2 + \frac{2}{f}\left(B + \frac{W_c}{k}\right)\right] . \qquad (16)$$

Here N, x, f and k are assumed equal in the two way selection initiated in generations b and c.

The most natural source of an estimate of $Q(Q = \sigma_p^2)$ is from data two-way selection proposed as the source of realized heritability estimates. Assuming a normal distribution, the sampling variance of the estimate of a variance is (Fisher 1958) twice the square of the variance divided by the number of degrees of freedom for the estimate.

Hence, if all data from the two-way selection programs were combined

$$\sigma_{\hat{Q}}^2 \doteq \frac{2\sigma_p^4}{4x(fk - 1)} \doteq \frac{\sigma_p^4}{2xfk} \ .$$

Not all degrees of freedom have meaning for the components of total variance derived from genetic variance among parents or from family environmental effects. However, since the within family component will ordinarily be the major protion of the variance, (16) is considered a satisfactory approximation.

Data from which $\hat{X}$ and $\hat{G}$ are computed are distinct from the data used to compute $\hat{S}$ and $\hat{Q}$, $\sigma_{\hat{U}\hat{V}}$ = zero. Though the data from which $\hat{Q}$ is computed is part of that used for $\hat{S}$, there is no covariance $\sigma_{\hat{S}\hat{Q}}$, because $\hat{S}$ is a function of phenotypic means and $\hat{Q}$ is pooled variance within groups.

Consideration of equations (7), (8), (9), (10) and (11) indicates that a final expression for $\sigma_{\hat{n}}^2$ will be different for the case when $b \neq 0$ than for the case when $b = 0$. At the same time it is clear from equation (7) that the size of $(q_c - q_b)$ has a major effect on the variance of $\hat{n}$; the coefficients of $\sigma_{\hat{G}}^2$ and $\sigma_{\hat{S}}^2$ are inversely proportional to the square of $(q_c - q_b)$. Given any fixed number of generations in the primary selection program, $(q_c - q_b)$ will be a maximum when $b = 0$. For these reasons attention will be focused on the $b = 0$ case. Using equations (11), (16) and (17) to make substitutions in (7) and remembering that $q_b = 1/2$ when $b = 0$, we obtain

$$\sigma_{\hat{n}}^2 = \left[\frac{1}{u(q_c - \frac{1}{2})}\right]^2 [V(\bar{P}_c) + V(\bar{P}_0)]$$

$$+ \left[\frac{\sigma_p^2}{2u^2(q_c - \frac{1}{2})^2}\right]\frac{2x}{N}\left[\frac{(1-h_0^2)\sigma_{g0}^2}{D_0^2} + \frac{(1-h_c^2)\sigma_{gc}^2}{D_c^2}\right]$$

$$+ \left[\frac{\sigma_p^2}{2u^2(q_c - \frac{1}{2})^2}\right]^2 \frac{2}{f}\left[\frac{B + W_{0/k}}{D_0^2} + \frac{B + W_{c/k}}{D_c^2}\right]$$

$$+ \frac{n^2}{2xfk} . \qquad (18)$$

Assuming the total selection differential is the same in the two-way selection programs conducted for estimation of R_b and R_c, then

$$D_0 = 2xi\sigma_{p0} \text{ and } D_c = 2xi\sigma_{pc} ,$$

where i is the average selection differential (in standard deviations) per population per generation. Now (18) can be rewritten as follows:

$$\sigma^2_{\hat{n}} = I + II + III + IV , \qquad (19)$$

where

$$I = \left[\frac{1}{u(q_c - \frac{1}{2})}\right]^2 [V(\overline{P}_c) + V(\overline{P}_0)] ,$$

$$II = \left[\frac{\sigma^2_p}{2u^2(q_c - \frac{1}{2})^2}\right] \frac{1}{2xi^2N} [(1-h^2_0)h^2_0 + (1-h^2_c)h^2_c] ,$$

$$III = \left[\frac{\sigma^2_p}{2u^2(q_c - \frac{1}{2})^2}\right]^2 \frac{1}{2fx^2i^2} \left[\frac{B + W_{0/k}}{\sigma^2_{p0}} + \frac{B + W_{c/k}}{\sigma^2_{pc}}\right] \text{ and}$$

$$IV = \frac{n^2}{2xfk} .$$

This subdivision of $\sigma^2_{\hat{n}}$ is made because evaluation of the components of $\sigma^2_{\hat{n}}$ may provide a basis for reduction of $\sigma^2_{\hat{n}}$ by changes in experimental design.

A portion of the parameters assumed as the basis for evaluation of $\sigma^2_{\hat{n}}$ will be components of the phenotypic variance among individuals in the base population (generation zero). Let

$$\sigma^2_p = M + F + W , \qquad (20)$$

where F is the variance among individuals due to different female parents, M is the genetic variance due to different male parents, and W is the variance among full-sibs. F can be subdivided into a nongenetic portion (defined earlier as B) and a portion due to gentic variation among female parents. W can also be divided into genetic and nongenetic portions. Given only additive gene effects, the assumption being made, $M = 1/4\,\sigma^2_g$, the genetic portion of F is equal to M and the genetic portion of W is equal to 2M. Thus, for this case

$$\sigma^2_p = 4M + B + E , \qquad (21)$$

where E is the nongenetic portion of W, and

$$h^2 = 4M/\sigma^2_p . \qquad (22)$$

From equation (20) it is apparent that if a population consists of the individuals, k in each of f full-sib families with f/m families sired by each of m male parents,

$$V(\overline{P}) = \frac{M}{m} + \frac{F}{f} + \frac{W}{fk} . \qquad (23)$$

q_c will depend on parameters n, the number of segregating genes, and c, the number of generations of selection in the selection phase. The expected change of gene frequency from one generation to the next can be approximated as follows (Falconer 1960):

$$E(\Delta q) = \frac{D}{\sigma^2_p} q(1 - q)u .$$

Since $E(\Delta q)$ is the expected change per generation, we have

$$\frac{dq}{dt} = \frac{D}{\sigma^2_p} q(1 - q)u .$$

The solution of this differential equation is

$$\int_{q_0}^{q_c} \frac{dq}{q(1-q)} = \frac{D}{\sigma^2_p} u \int_0^c dt ,$$

where q_0 and q_c are frequencies at generation 0 and c, respectively. Assuming

$q_0 = 1/2$, we obtain

$$\ln \frac{q_c}{1 - q_c} = \frac{Duc}{\sigma^2_p} \qquad (24)$$

and that the expected gene frequency at generation c is

$$q_c = \frac{e^z}{1+e^z}, \qquad (25)$$

where $z = \frac{Duc}{\sigma_p^2}$.

In this instance $D = I\sigma_p$ is the selection differential per generation and $z = \frac{Iuc}{\sigma_p}$. Given the magnitude of σ_{g0}^2 or M_0, u can be evaluated as $u = \sqrt{2\sigma_{g0}^2/n}$, because $\sigma_{g0}^2 = 2q(1-q_0)\,u^2n = u^2n/2$, when $q_0 = 1/2$, and gene effects are additive and all have the same u value.

Quantitative Evaluations

Equations (19) through (25) can be employed to obtain quantitative information concerning $\sigma_{\hat{n}}^2$. Results are shown in Table 1 for the following set of parameters:

I(selection differential in standard units in the primary selection program) = 1.2;

i(selection differential in standard units in the two-way selection program) = 2.0;

M = 5, D = 15, B = 10, W = 80, h^2 = 0.2 or

M = 10, D = 20, B = 10, W = 70, h^2 = 0.4;

f(number of female parents in the two-way selection programs) = 4m (where m is the number of sires);

k(number of offspring per dam in the two-way selection programs) = 10;

x(number of generations in the two-way selection programs) = 1, 2, 3, 4;

n(number of segregating genes) = 10, 50, 200;

and c(number of generation in the primary selection program) = 5, 10, 20.

Values of f and k were taken to be the same in generations 0 and c of the primary selection program and in the two-way selection programs, but they

Table 1. Variance of the gene number estimate ($\sigma_{\hat{n}}^2$) and the number of sires (m) required to estimate gene number (n) at various levels of coefficient variation ($\sqrt{\sigma_{\hat{n}/n}^2}$) with given heritability (h^2) and number of generations in the primary (c) and two way (x) selection program

		n = 10				n = 50				n = 200			
			m to make $\sqrt{\sigma_{\hat{n}/n}^2}$ =				m to make $\sqrt{\sigma_{\hat{n}/n}^2}$ =				m to make $\sqrt{\sigma_{\hat{n}/n}^2}$ =		
c	x	$m\cdot\sigma_{\hat{n}}^2$	0.1	0.2	0.4	$m\cdot\sigma_{\hat{n}}^2$	0.1	0.2	0.4	$m\cdot\sigma_{\hat{n}}^2$	0.1	0.2	0.4
h^2 = 0.4													
5	1	95	95	24	6	25,997	1,040	260	65	5,806,305	14,516	3,629	907
	2	57	57	14	4	10,944	438	109	27	2,383,672	5,959	1,490	372
	3	46	46	12	3	6,974	279	70	17	1,476,749	3,689	922	230
	4	42	42	11	3	5,187	207	52	13	1,065,467	2,664	666	166
10	1	28	28	7	2	2,269	91	23	6	407,570	1,019	255	64
	2	19	19	5	1	1,088	44	11	3	167,240	418	105	26
	3	16	16	4	1	733	29	7	2	103,916	260	65	16
	4	15	15	4	1	575	23	6	1	75,400	189	47	12
20	1	23	23	6	1	612	24	6	2	36,918	92	23	6
	2	16	16	4	1	343	14	3	1	15,281	38	10	2
	3	15	15	4	1	274	11	3	1	9,639	24	6	2
	4	14	14	4	1	246	10	2	1	7,112	18	4	1
h^2 = 0.2													
5	1	703	703	176	44	319,689	12,788	3,197	799	73,453,020	181,633	45,908	11,477
	2	306	306	77	19	123,210	4,928	1,232	308	28,014,298	70,036	17,509	4,377
	3	209	209	52	13	74,067	2,963	741	185	16,745,473	41,864	10,466	2,616
	4	167	167	42	10	52,561	2,102	526	131	11,838,077	29,595	7,399	1,850
10	1	117	117	29	7	24,345	974	243	61	4,812,388	12,031	3,008	752
	2	57	57	14	4	9,253	370	93	23	1,838,233	4,596	1,149	287
	3	43	43	11	3	5,538	222	55	14	1,000,622	2,752	688	172
	4	37	37	9	2	3,928	157	39	10	779,398	1,948	487	122
20	1	51	51	13	3	3,046	122	30	8	385,428	964	241	60
	2	29	29	7	2	1,119	45	11	3	147,339	368	92	23
	3	21	21	5	1	663	27	7	2	88,507	221	55	14
	4	19	19	5	1	469	19	5	1	62,941	157	39	10

Table 2. Coefficient of variation of the gene number estimate when the number of sires (m) is 50 with given levels of heritability (h^2) and generation number in the primary (c) and two way (x) selection program

		$h^2 = 0.4$			$h^2 = 0.2$		
c	x	n = 10	n = 50	n = 200	n = 10	n = 50	n = 200
10	2	0.062	0.093	0.289	0.107	0.272	0.96
	3	0.056	0.076	0.228	0.093	0.210	0.71
	4	0.055	0.068	0.194	0.086	0.177	0.62
20	2	0.056	0.052	0.087	0.076	0.095	0.27
	3	0.055	0.047	0.069	0.065	0.073	0.21
	4	0.053	0.044	0.060	0.062	0.061	0.18

could be different. The number of male parents (m) per generation was not specified and results are presented in terms of m.

Table 2 shows the coefficient of variation of $\hat{n}$ when m, the number of sires, is 50 for all combinations of c = 10, 20; x = 2, 3, 4; h^2 = 0.4, 0.2 and n = 10, 50, 200.

Let m = 50, x = 3 and c = 20. Table 2 shows the coefficients of variation are different, dependging on heritability and gene number. Using the coefficients of variation in Table 2 and assuming a normal distribution for $\hat{n}$ and no bias in the estimates, the following probability statements can be made. If the gene number (n) is 10 and h^2 = 0.4, probabilities of the estimate will be

$P(8.9 < \hat{n} < 11.1) = 0.95$ and

$P(\hat{n} > 11.7) = 0.001$.

When $h^2 = 0.2$

$P(8.7 < \hat{n} < 11.3) = 0.95$ and

$P(\hat{n} > 12.0) = 0.001$.

If n = 50 and $h^2 = 0.4$

$P(45.5 < \hat{n} < 54.5) = 0.95$,

$P(\hat{n} > 57.3) = 0.001$ and

$P(\hat{n} < 42.7) = 0.001$.

If $h^2 = 0.2$

$P(42.8 < \hat{n} < 57.2) = 0.95$,

$P(\hat{n} > 61.3) = 0.001$ and

$P(\hat{n} < 38.7) = 0.001$.

If n = 200 and $h^2 = 0.4$

$P(173.0 < \hat{n} < 227.0) = 0.95$ and

$P(\hat{n} < 157.4) = 0.001$.

If $h^2 = 0.2$

$P(117.7 < \hat{n} < 282.3) = 0.95$ and

$P(\hat{n} < 70.2) = 0.001$.

These probabilities together with Tables 1 and 2 indicate that estimates are considerably more reliable when heritability is at least 0.4 than when it is low (h^2 = 0.2). The effect of heritability on sampling variance is particularly important when the gene number is large. The probability statements indicate that when estimates are obtained from a relatively large experiment conducted through a reasonable time span, sampling errors would not prevent distinguishing among small (in the region of 10), intermediate (in the region of 50) or large (200 or more) numbers of genes.

Generally the most effective way of decreasing the variance of gene number estimates is to increase the number of generations (c) in primary selection, since terms II and III of the variance of $\hat{n}$ are decreased in proportion to the fourth power of the change of gene frequency. Increasing the number of generations (x) in the two-way selection program would have similar effects since term III is decreased in proportion to x^2 and term II in proportion to x. However, two-way selection should not be continued beyond the generations of linear responses. Table 1 shows gains are small from increasing x beyond four. Table 1 indicates all portions of the variance $\hat{n}$ are inversley proportional to the number of sires and therefore to population size. Term II decreases with increased number of sires in the two-way selection phase and the others vary inversely with sire number in primary selection.

If heritability of a quantitative character is low and the gene number is large, many generations of primary selection are required to obtain an estimate of gene number with reasonable precision. However, with two-way selection, estimates of heritability V_0

can be obtained early before the number of generations is critical. If this indicates considerable reduction in the additive genetic variance, the final two-way selection phase can be initiated and the program brought to termination with some confidence that the sampling variance of the estimate obtained will be reasonable. On the other hand if little change in additive genetic variance is indicated, the primary selection phase should be continued over a longer period.

Acknowledgment

I am grateful to Professor R.E. Comstock for his helpful advice during the course of this work.

Literature

Castle, W.E.: An improved method of estimating the number of genetic factors concerned in cases of blending inheritance. Science 54, 223 (1921)

Comstock, R.E.: Number of genes affecting growth in mice. Genetic Lectures (Oregon State University) 1, 137-148 (1969)

Comstock, R.E.; Moll, R.H.: Genotype-environment interactions. Statistical Genetics and Plant Breeding. NAS-NRC p. 982 (1963)

Comstock, R.E.; Robinson, H.F.: The components of genetic variance in populations of biparental progenies and their use in estimating the average degree of dominance. Biometrics 4, 254-266 (1948)

Falconer, D.S.: Introduction to quantitative genetics. The Ronald Press Company 1960

Gates, C.E.: Discussion: Some considerations in variance component and partitioning methods of genetic analysis. Statistical Genetics and Plant Breeding. NAS-NRC p. 982 (1963)

Hill, W.G.: Estimation of realized heritability from selection experiments. I. Divergent selection. Biometrics 28, 747-765 (1972)

Hill, W.G.; Robertson, A.: The effect of linkage on limits to artificial selection. Genetical Research 8, 269-294 (1966)

Johnson, N.L.; Kotz, S.: Discrete distributions. Boston: Houghton-Mifflin 1969

Mather, K.; Jinks, J.L.: Biometrical genetics. Cornell University Press 1971

Powers, LeRoy: The Nature and interaction of genes differentiating habit of growth in a cross between varieties of *Triticum vulgare*. J. Agr. Res. 49, 573-605 (1934)

Powers, LeRoy: The partitioning method of genetic analysis and some aspects of its application to plant breeding. Statistical Genetics and Plant Breeding. NAS-NRC pp. 280-318 (1963)

Roberts, R.C.: The limits to artificial selection for body weight in the mouse. I. The limits attained in earlier experiments. Genet. Res. 8, 347-360 (1966)

Robertson, A.: The nature of quantitative genetic variation. In: Heritage from Mendel (ed. Brink, R.A.) pp. 265-280. University of Wis. Press 1967

Speickett, S.G.; Thoday, J.M.: Regular responses to selection: 3 Interaction between located polygenes. Genet. Res. 7, 96-121 (1966)

Thoday, J.M.: Location of polygenes. Nature 191, 368-370 (1961)

Thoday, J.M.; Boam, T.B.: Effects of disruptive selection. II. Polymorphism and divergence without isolation. Heredity 13, 205-208 (1959)

Wright, S.: The results of crosses between inbred stains of guinea pigs differing in number of digits. Genetics 19, 537-551 (1934)

Wright, S.: The genetics of quantitative variability. Quantitative Inheritance (eds. Reeve, E.C.R.; Waddington, C.H.) London: H.M.S.O. 1952

Wright, S.: Evolution and the Genetics of populations. Genetics and Biometric Foundations, Vol. 1. The University of Chicago Press 1968

Received April 15, 1977
Communicated by R.W. Allard

Dr. Y.C. Park
Department of Statistics
Florida State University
Tallahassee, Florida 32306
(U.S.A.)