Theor. Appl. Genet. 50, 163-172 (1977)

Theory for the Number of Genes Affecting Quantitative Characters*
II. Biases from Drift, Dominance, Inequality of Gene Effects, Linkage Disequilibrium and Epistasis

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

Summary. Bias in the estimation of gene number from a population of finite size will be reduced when the number of generations and the selection differential in the primary selection program are increased. A procedure to minimize bias due to dominance is proposed. Bias due to linkage disequilibrium is decreased by long generations of selection and small change of linkage disequilibrium. Bias due to inequality of gene effects was studied assuming that gene effects follow a half normal distribution. The epistatic bias arising from multiplicative gene effects disappears when the logarithms of the original measurements are considered even though the gene number estimate is unsatisfactory without transformation.

Key words: Biases - Drift - Inequality of Gene Effects - Linkage Disequilibrium - Epistasis

Introduction

Conventional methods of Castle (1921), Wright (1934, 1952, 1968), Falconer (1960), Comstock (1969), Park (1977), et al. for estimating gene number are based on some assumptions, i.e., additivity and equality of gene effects. Hence, it is possible to imagine that the estimations of gene number are biased for some quantitative characters in which the assumptions cannot be justified.

The procedure proposed permits estimating gene number from the measures of response and additive genetic variance provided by recurrent selection experiments in which initial frequencies of alleles at segregating loci are 0.5, but not long enough to measure maximum responses. The allele frequencies of 0.5 at segregating loci can be obtained by crossing a pair of inbred lines. This paper evaluates estimation biases of the proposed procedure resulting from random drift, variations in gene effects, linkage disequilibrium, variations of dominance, sampling variance, and epistasis.

* 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 5 T32 ES07011-02.

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

Genetic Drift

When the population size is finite there will be sampling variation of the gene pool in each generation. Let us consider the bias from this random genetic drift in estimating gene number.

Assuming additive and equal gene effects and linkage equilibrium, the coded genotypic mean and additive genetic variance from n segregating loci with the favorable allele frequency q and one-half the difference in effect between the two homozygotes u are as follows (Comstock and Robinson 1948):

$$\bar{y} = nu(2\bar{q} - 1)$$

$$\sigma_q^2 = 2u^2\left[n\bar{q} - n\bar{q}^2 - \sum_i (q_i - \bar{q})^2\right],$$

where $\bar{q}$ represents the real average of q over all loci.

A general formula for estimating gene number was described as follows (Park 1977):

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

where subscripts c and b specify points in time and $c > b$. Substituting the coded genotypic means and additive genetic variances with proper subscripts in equation (1), we get gene number concerning genetic drift.

$$n_1 = n\left|\frac{1}{1+x}\right|, \qquad (2)$$

where $x = \frac{1}{2}\,\frac{\sum(\bar{q}_{ci} - \bar{q}_c)^2 - \sum(q_{bi} - \bar{q}_b)^2}{(\bar{q}_c - \bar{q}_b)(\bar{q}_c + \bar{q}_b - 1)}$. The expected value of the segregating gene number is

$$En_1 = nE\left|\frac{1}{1+x}\right|$$

$$\approx n\left|\frac{1}{1+Ex}\right|.$$

$$Ex \approx \frac{1}{n}\,\frac{E\sum(\bar{q}_{ci} - \bar{q}_c)^2 - E\sum(q_{bi} - \bar{q}_b)^2}{E(\bar{q}_c - \bar{q}_b)(\bar{q}_c + \bar{q}_b - 1)}$$

because the expectation of a fraction is not exactly equal to the expectation of the numerator divided by the expectation of the denominator, unless the latter is a constant.

$$E(x) \approx \left(\frac{n-1}{n}\right)\left|\frac{V(q_c) - V(q_b)}{(\hat{q}_c - \frac{1}{2})^2}\right| \qquad (3)$$

Here the term $V(\bar{q}_c)$ has been ignored in the denominator of (3). It can be shown that $V(\bar{q}_c)$ will be trivial if $n \geqslant 10$.

Since an expression for the variance of the gene frequency due to random drift in the presence of selection was not found in the literature, the following approximation was derived (Park 1977):

$$V(q_t) = \sum_{i=1}^{t} \frac{\bar{q}_i(1-\bar{q}_i)}{2N}\left(1 - \frac{1}{2N}\right)^{t-i}, \qquad (4)$$

where $\bar{q}_i$ symbolizes the expectation of q in the i^{th} generation, and t specifies any time of generation.

This differs from using $\bar{q}$ in expressions for $\bar{y}$ and σ_g^2. As noted elsewhere, in those expressions $\bar{q}$ symbolizes the real average of q over the n effective loci of the recurrent selection population.

Applying equation (4), we get

$$V(q_b) = \frac{\bar{q}_0(1-\bar{q}_0)}{2N} \quad \text{when } b = 0 \text{ and}$$

$$V(q_c) = \sum_{i=1}^{c+1} \frac{\bar{q}_i(1-\bar{q}_i)}{2N}\left(1 - \frac{1}{2N}\right)^{c-i+1}.$$

Since the range of t is from b + 1 to c + 1 in the primary selection program (Park 1977), attention is drawn to the fact that as symbols are used in the above expression, b = 0 refers to the $\overline{F}_2$ generation, but i = 0 refers to the F_1.

Bias is inversely related to the square of change in gene frequency due to selection (equations 2, 3); hence, bias will be reduced by increasing $\bar{q}_c$ in upward selection. $\bar{q}_c$ is increased not only when the number of generations in the primary selection program (c) is increased, but also when the selection coefficient (s) is increased; s will be increased when heritability (h^2) is increased, the segregating gene number (n) is decreased or the selection differential in the primary selection program (k) is increased. Bias also is directly related to $V(q_c)$; hence, bias will be reduced when N is increased (equation 4).

To obtain information on actual size of bias, E(x) and $E(n_1)$ were computed for the following set of pa-

Table 1. Expected gene number concerning genetic drift (En_1) and expected levels of bias ($E(1/(1+x))$) from random drift with given segregating loci (n), heritability (h^2) and number of sires (m)

n	h^2	m	c	$E(n_1)$	$E(1/(1+x))$
10	0.2	10	10	8.9	0.893
			20	9.1	0.908
		20	10	9.4	0.941
			20	9.5	0.953
	0.4	10	10	9.3	0.933
			20	9.5	0.947
		20	10	9.6	0.963
			20	9.7	0.968
50	0.2	10	10	33	0.658
			20	38.5	0.771
		20	10	39.5	0.786
			20	43	0.862
	0.4	10	10	39	0.781
			20	42.5	0.853
		20	10	43.5	0.872
			20	45.5	0.912
200	0.2	10	10	65	0.325
			20	96	0.478
		20	10	96	0.481
			20	126	0.629
	0.4	10	10	98	0.493
			20	130	0.654
		20	10	130	0.650
			20	156	0.776

rameters n = 10, 50, 200, h^2 = 0.2, 0.4, c = 5, 10, 20, k = 1.2, f = 4m and m = 10, 20, where f and m are the numbers of female and male parents, respectively.

As anticipated earlier (on the basis of equations 2, 3, and 4) bias is decreased by increases in population size, heritability, or number of generations of selection (c) and by decreases in the value of n. Table 1 shows that the ranges of bias are 3 ~ 11% when n = 10, 9-34% when n = 50, and 22-67% when n = 200.

Dominance

Since the basic model assumes that gene effects are unchanging quantities and the same for every locus, and that gene frequencies at all loci are the same, we get

$$\bar{y} = nu(2q - 1) + nau\ 2q(1 - q)$$

if it is also assumed that the dominance level (a) is an unchanging quantity and the same for every locus.

Given the above assumptions, the additive genetic variance can be expressed as

$$\sigma_g^2 = nu^2\, 2q(1 - q)[1 + (1 - 2q)a]^2 .$$

In the case where initial values of $\bar{y}$ and σ_g^2 are those of the F_2 generation, $\bar{y}_c$ is estimated as $\bar{p}_c - \bar{p}_0$, and $\bar{y}_b$ is taken to be zero, and

$$\sigma_{gb}^2 = \frac{n}{2} u^2 \quad \text{and}$$

$$\bar{y}_c = nu\left[(2q_c - 1) + 2q_c(1 - q_c)a - \frac{a}{2}\right] .$$

Substituting in equation (1) we obtain the gene number concerning effect of dominance (n_2) as

$$n_2 = n \cdot \theta_1 ,$$

$$\text{where } \theta_1 = \frac{1}{4}\left[\frac{\{2[(2q_c - 1) + 2q_c(1 - q_c)a] - a\}^2}{1 - 4q_c(1 - q_c)[1 + (1 - 2q_c)a]^2}\right] .$$

Values of θ_1 are listed in Table 2.

They show that positive dominance causes downward bias, which is very large when dominance is high or $(q_c - 1/2)$ is small, because the length of the primary selection program is too short. When dominance is negative the situation is even less satisfactory. Both the negative and the high positive values of θ_1 result from the fact that when a is negative additive genetic variance increases as q increases from one half until its maximum is reached.

Table 2 shows that when there is dominance, reasonable estimates of n are unlikely. An alternative procedure is possible if both F_1 and F_2 data are collected. Let the corresponding phenotypic means by symbolized as $\bar{F}_1$ and $\bar{F}_2$. Since

$$\bar{p}_c = \alpha + (2q_c - 1)\, nu + 2q_c(1 - q_c)\, nau$$

$$\bar{F}_2 = \alpha + \frac{nau}{2} \quad \text{and}$$

$$\bar{F}_1 = \alpha + nau ,$$

where α is a constant.

We can eliminate -nau/2 from the $\bar{y}_c$ by using $\bar{p}_c - 2\bar{F}_2 + \bar{F}_1$ as the estimate of $\bar{y}_c$. Then

$$n_2 = n \cdot \theta_2 , \text{ where}$$

$$\theta_2 = \frac{[(2q_c - 1) + 2q_c(1 - q_c)a]^2}{[1 - 4q_c(1 - q_c)[1 + (1 - 2q_c)a]^2]} .$$

Values of θ_2 are listed in Table 3.

Table 2. Numerical values of θ_1 which indicate levels of bias from presence of dominance (a)

	a					
q_c	-1.0	-0.6	-0.2	0.2	0.6	1.0
0.6	-0.13	-0.22	-1.09	0.33	0.14	0.08
0.8	-0.93	-2.71	2.06	0.63	0.33	0.20
1.0	2.25	1.69	1.21	0.81	0.49	0.25

Table 3. Numerical values of θ_2 which indicate levels of bias from presence of dominance (a)

	a					
q_c	-1.0	-0.6	-0.2	0.2	0.6	1.0
0.6	-0.21	-0.04	-0.28	0.76	0.93	1.20
0.8	-0.12	-0.91	1.46	0.87	0.85	0.94
1.0	1.00	1.00	1.00	1.00	1.00	1.00

Table 4. Bias in gene number estimation from dominance when dominance levels are variable

Dominance Levels (Frequency)			Range of c	Range of $E(q_c)$	$E(n_2)$	θ_2
0.2(0.2)	0.6(0.6)	1.0(0.2)	15	0.81 - 0.90	45.1	0.86
0.2(0.4)	0.6(0.4)	1.0(0.2)	8	0.72 - 0.76	42.8	0.86
-0.3(0.2)	0.1(0.6)	0.5(0.2)	8	0.75 - 0.79	47.6	0.95
-0.3(0.3)	0.1(0.4)	0.5(0.3)	8	0.75 - 0.79	48.9	0.98
-0.4(0.2)	0(0.6)	0.4(0.2)	8	0.75 - 0.80	50.7	1.01
-0.5(0.3)	-0.1(0.4)	0.3(0.3)	15	0.88 - 0.95	50.1	1.00

Table 3 indicates that if $\bar{p}_c - 2\overline{F}_2 + \overline{F}_1$ is used as the estimate of $\bar{y}_c$, large biases will not result if all a-values are positive when they are not zero. It also shows that large biases can result if most a-values are negative and the primary selection program is not continued until the frequencies of favorable alleles are very close to 1.0. However, assuming the presence of dominance, it is realistic to suppose that dominance levels are not the same for every locus, and that gene frequencies after selection through c generations will vary because of inequality of dominance levels.

To evaluate the effect of dominance on gene number estimates when level of dominance varies, we need to determine the expected gene frequencies after some specified number of generations of selection.

The expected change of gene frequency per generation can be expressed as follows (Falconer 1960):

$$E(\Delta q) = \frac{k}{\sigma_p} q(1 - q)\,[1 + (1 - 2q)a]u\ .$$

We can obtain the expected gene frequency at generation c, given specific values of a and u, by recurrent application of the above equation. For example, considering a situation in which $h^2 = 0.4$, n = 50, k = 1.2, and c = 8, with 20 % loci of dominance level 0.2, 60 % loci of dominance level 0.6, and 20 % loci of dominance level 1.0, the foolowing table is obtained.

Level of dominance	No. of loci	σ_{g0i}	σ_{g8i}	$\bar{y}_{8i}$
0.2	10	$5u^2$	$2.885u^2$	6.001u
0.6	30	$15u^2$	$1.928u^2$	26.006u
1.0	10	$5u^2$	$1.241u^2$	8.454u
Total	50	$25u^2$	$6.054u^2$	40.461u

For the case of c = 8

$$n_2 = \frac{(40.461u)^2}{2(25u^2 - 6.054u^2)}$$

$$= 43.2.$$

Hence $\theta_2 = 0.86$. Continuing the calculations gives the results shown in Table 4.

Table 4 indicates the approximate bias from symmetrical distributions of a-values in three ranges. None of the θ_2 values differs greatly from 1.0 and most of them are close to 1.0. It is not surprising that when all a's are positive θ_2 is close to its average for the a's involved, as shown in Table 3.

The results obtained for distributions involving both positive and negative a's were not foreseen so easily and are particularly encouraging. Taking all of the Table 4 results together and assuming that the a-values in any real situation would vary, it seems safe to conclude that, given any degree of positive heterosis or a small amount of negative heterosis caused by dominance, bias due to dominance would be quite small.

Linkage Disequilibrium

When there is linkage disequilibrium, but no epistasis, the additive genetic variance will be as follows

$$\sigma_g^2 = n2q(1-q)u^2 + \sum_i \sum_{j \neq i} 2(pt - rs)_{ij} u^2\ ,$$

where p is the frequency of gametes with the favorable allele at both the i^{th} and j^{th} locus, t is the frequency of gametes with the unfavorable allele at both loci

and r and s are the frequencies of the two repulsion type gametes.

In the absence of epistasis and position effects, the genetic mean will not be affected by linkage disequilibrium. Hence

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

Substituting the above relations to equation (1) yields gene number concerning linkage disequilibrium (n_3).

$$n_3 = n\left[\frac{1}{1 + \dfrac{\sum_i \sum_{j \neq i} (p_0 t_0 - r_0 s_0)_{ij} - \sum_i \sum_{j \neq i} (p_c t_c - r_c s_c)_{ij}}{n(q_c - \frac{1}{2})^2}}\right]$$

This shows that bias from linkage disequilibrium is decreased by long generations of selection and small changes of linkage disequilibrium.

Evidence from design III experiments (Comstock and Robinson 1952) on the magnitude of $\sum_i \sum_{j \neq i} (p_0 t_0 - r_0 s_0) u_i u_j$ is available for a few populations. Moll et al. (1964) studied seven traits in two populations of maize. They found coupling significantly in excess for one trait in one population and repulsion significantly in excess for another trait in both populations. The quantity $\sum \sum (p_0 t_0 - r_0 s_0) u_i u_j$ was not significant in eleven other cases. Data by Enfield et al. (1969) for pupa weight in *Tribolium* gave no evidence of excess coupling or repulsion linkage.

Variation in Gene Effects

When we posit that the u's vary from locus to locus, estimation of gene number will be biased due to inequality itself. The inequality of the u's will cause inequality of gene frequency as a secondary effect. Gene frequency is affected by drift as well as selection, but in this section we neglect drift effect, which means we assume large population size.

If one writes the expressions for $\bar{y}$ and σ_g^2 when a = 0 and the u's vary, one has

$$\bar{y} = \sum_i (2q_i - 1)u_i \quad \text{and}$$

$$\sigma_g^2 = \sum 2q_i(1 - q_i)u_i^2 .$$

Substitution into equation (1) gives a formula for gene number as a function of variable gene effects (n_4).

$$n_4 = \frac{\left|\sum (2q_{ci} - 1)u_i\right|^2}{2\left|\frac{1}{2}\sum u_i^2 - 2\sum q_{ci}u_i^2 + 2\sum q_{ci}^2 u_i^2\right|}$$

Now consider the special case in which all the q_{ci}'s are 1.0 because of i) large N in the primary selection program and ii) selection until no further response.

$$n_4 = n\left[\frac{1}{1 + \sigma_u^2/\bar{u}^2}\right],$$

as reported by Comstock (1969). It shows the unequal u's will be the source of downward bias in $\hat{n}$ and that the amount of bias is related to the coefficient of variation among the u's.

Comstock (1969) suggested that the loci with very small effects might be unimportant and that it is best not to include them in counting pertinent loci. One might inquire then whether the inherent bias due to variable u's would cause the expectation of the estimate to be reasonably close to the number of loci important enough in their effects to deserve being counted. He thinks that there can be no unique answer because so much depends on the actual distribution of the u's and the view concerning the size of effect that makes a locus worth counting.

Since the experimental evidences, such as Thoday (1961), Wehrhahn and Allard (1965), and Robertson (1967), suggest that genes with large effects are rarer than those with progressively smaller effects, the upper half of a normal distribution of gene effects may be a good approximation for some real distributions of gene effects in order to study the level of bias for some quantitative characters.

Supposing the u's are distributed as the upper half of a normal distribution, Comstock (1969) showed that $\sigma_u^2/\bar{u}^2$ would be about 0.56. Hence, the expectation of the gene number estimate would be 0.64n. Sixty-four percent of the loci, those with the

Table 5. Additive genetic variance (σ_g^2) and coded genotypic mean ($\bar{y}$) from a half normal distribution of gene effects (u)

Class Boundaries in σ's	u	Rel freq.	Abs. freq.	σ^2_{g0i}	q_{ci}	σ^2_{gci}	y_{ci}
0 - 0.4	0.2	0.311	31	0.62	0.574	0.6064	0.917
0.4 - 0.8	0.6	0.265	27	4.86	0.710	4.0050	6.795
0.8 - 1.2	1.0	0.194	19	9.50	0.816	5.7036	12.011
1.2 - 1.6	1.4	0.121	12	11.76	0.890	4.6223	13.088
1.6 - 2.0	1.8	0.064	6	9.72	0.936	2.3306	9.417
2.0 - 2.4	2.2	0.029	3	7.26	0.964	1.0168	6.120
2.4 - 2.8	2.6	0.011	1	3.38	0.980	0.2696	2.494
2.8 - 3.2	3.0	0.005	1	4.50	0.989	0.2014	2.932
		1.000	100	51.60		18.7557	53.774

largest effects, would account for about 90 percent of $\sum_i u_i$.

For the case in which the q_c's have not reached 1.0, but in which N is large enough to allow drift to be ignored, the differences in the u's will cause variable gene frequencies from locus to locus as a secondary effect. Consider again the half normal distribution of u's. Approximations of the joint effect of variable u's and the consequent variation in q_c can be computed from a discontinuous approximation to the assumed distribution obtained with the aid of a table of the normal distribution. For illustration, we will subdivide into classes that are 0.4σ wide and assume that each u in a given class has a value that is at the center of the class. Considering a given situation in which n = 100, $h^2 = 0.2$, c = 20, and k = 1.2, gives table 5.

Here σ^2_{g0i} is the contribution of genes in the i^{th} class to additive genetic variance in generation zero, q_{ci} is the expected frequency of genes in the i^{th} class in generation c, σ^2_{gci} has the same meaning as σ^2_{g0i} except for generation c, and $\bar{y}_{ci}$ is the contribution of genes in the i^{th} class to $\bar{y}_c$.

Values of q_{ci} were computed using the relation $q_{ci} = e^{ku_i}/(1 + e^{ku_i})$, (Park 1977), where $\sigma^2_p = \sigma^2_{g0}/h^2 = \frac{51.6}{0.2} = 258$ and $K = \frac{kc}{\sigma_p} = \frac{1.2 \times 20}{\sqrt{258}} = 1.49$.

Having obtained q_{ci} values, $\bar{y}_{ci}$ can be computed for each class. Finally,

$$n_4 = \frac{(53.774)^2}{2[51.60 - 18.7557)}$$

$$= 44.02 .$$

Table 6. Gene number (n_4) and level of bias ($1 - \alpha$) from a half normal distribution of gene effects

n	h^2	c	n_4	$n_4 = \alpha n$
200	0.4	20	90.80	0.454n
		10	77.50	0.388n
	0.2	20	80.75	0.404n
		10	72.70	0.364n
50	0.4	20	26.94	0.539n
		10	22.68	0.454n
	0.2	20	24.80	0.496n
		10	20.78	0.416n
10	0.4	20	6.50	0.650n
		10	5.93	0.593n
	0.2	20	6.26	0.626n
		10	5.53	0.553n

Table 6 shows the amounts and directions of bias in various cases by the same method of calculation. The table shows that when n_4 is obtained with equation (1), n_4 is about half the actual gene number in the case of a half normal distribution of u's. The biases indicated all exceed the expected when selection continues until all favorable alleles are homozygous.

Biases were greater when c was less and h^2 was less, i.e., when conditions were such that q_c was less. n_4 varied around 0.5n, ranging from 0.36n~ 0.65n.

Sampling Variance

Equation (1) will yield biased estimates of n even if all underlying assumptions are correct, because when a parameter, say α, is estimated with error, the ex-

pectation of $f(\hat{\alpha}) \neq f(\alpha)$ if $f(\alpha)$ is a nonlinear function of α.

Consider equation (1) for the case when generation b is the F_2 generation. The $\bar{y}_b = 0$ and equation (1) reduces to

$$n = \frac{Z^2}{D},$$

where $Z = \bar{y}_c$ and $D = 2(\sigma^2_{gb} - \sigma^2_{gc})$. Then

$$\hat{n} = \frac{\hat{Z}^2}{\hat{D}},$$

where $\hat{}$ indicates an estimate.

The bias arising from sampling errors is given by the ratio

$$E\left(\frac{\hat{Z}^2}{\hat{D}}\right) \Big/ \frac{Z^2}{D} = \frac{D}{Z^2} E\left(\frac{\hat{Z}^2}{\hat{D}}\right), \quad \text{where}$$

$$E\left(\frac{\hat{Z}^2}{\hat{D}}\right) = D(\hat{Z}^2)\, E\left(\frac{1}{\hat{D}}\right)$$

because with the proposed estimation procedures there will be no sampling covariance between estimates of $\bar{y}_c$ and $(\sigma^2_{gb} - \sigma^2_{gc})$.

To evaluate $E(1/\hat{D})$ we take the expectation of the first three terms of Taylor's expansion and obtain

$$\frac{D}{Z^2} E\left(\frac{\hat{Z}^2}{\hat{D}}\right) = \left[1 + \frac{\sigma^2_{\hat{Z}}}{Z^2}\right]\left[1 + \frac{\sigma^2_{\hat{D}}}{D^2}\right].$$

We see that the bias from sampling variance is upward. Turning to the magnitude of the bias, it is apparent that the information provided by the experiment must be sufficient so that $\sigma_{\hat{Z}}/Z$ and $\sigma_{\hat{D}}/Z$ do not exceed 0.25 if a reasonably precise estimate of n is to be obtained. At this practical limit we see that in an otherwise adequate experiment the bias from sampling variance would be no greater than 13%.

Epistasis

The biases that may arise from epistasis present a particularly difficult problem because we do not know the real nature of gene interaction. Models most discussed in the literature for understanding gene interactions include multiplicative, optimum number, complementary, and duplicative models (Horner et al 1955; NaPuket 1967; Wright 1935, 1968). The possibility of two or more types of gene interactions being present in the same genetic system probably has not been emphasized as much as it deserves to be.

Consider the multiplicative model that specifies the value of a genotype as

$$Y = \alpha\, e^{n1}\, b^{n2},$$

where n1 is the number of heterozygous loci, n2 is the number of loci homozygous for the favorable allele, $b > 1.0$, $b \geq e \geq 1.0$ and α is a constant reflecting both the measurement scale and the value of the genotype that is homozygous for the least favorable allele at each locus heterozygous in the F_1.

Assuming equal genotypic effects, same gene frequency, and a population in which Hardy-Weinberg genotype frequencies can reasonably be assumed, the genotypic mean is

$$\bar{Y} = \alpha[bq^2 + e2q(1-q) + (1-q)^2]n, \quad \text{where}$$

$$a = \frac{2e - b - 1}{b - 1},$$

$$u = \frac{\alpha}{2}(b-1)[bq^2 + 2eq(1-q) + (1-q)^2]^{n-1}$$

and the additive genetic variance is

$$\sigma^2_g = \alpha^2 n2q(1-q)\left|1 + (1-2q)\left(\frac{2e-b-1}{b-1}\right)\right|^2 \frac{1}{4}(b-1)^2 \times$$

$$[bq^2 + e2q(1-q) + (1-q)^2]^{2(n-1)}.$$

Because both $\bar{Y}$ and σ^2_g are functions of b, e, and n, a particular value of the ratio, $\sqrt{\sigma^2_g}/\bar{Y}$, requires particular values of b and e, if n and the level of dominance (a) are specified. Let $\sqrt{\sigma^2_g}/\bar{Y}$ be referred to as the genetic coefficient of variation and symbolized as GCV. Then note that

$$GCV = \sqrt{\sigma^2_g}/\bar{Y} = CV\sqrt{h^2},$$

where CV = phenotypic coefficient of variation. The following table shows GCV values associated with

Table 7. Values of n_5 which are gene numbers concerning multiplicative gene effects associated with different gene frequencies achieved by selection when there is no dominance ($e = (b+1)/2$), GCV = 0.089 and n = 10, 50 or 200

	n = 10			n = 50			n = 200		
q	$\overline{Y}$	σ_g^2	n_5	$\overline{Y}$	σ_g^2	n_5	$\overline{Y}$	σ_g^2	n_5
0	1.000	0.000000	7.0	1.000	0.000000	22.3	1.000	0.000000	43.6
0.1	1.086	0.003597	6.0	1.197	0.004147	18.1	1.433	0.005965	37.2
0.3	1.279	0.011262	3.7	1.711	0.019497	9.8	2.936	0.058028	20.6
0.5	1.502	0.017903		2.440	0.046534		6.001	0.286529	
0.7	1.759	0.019991	-15.8	3.471	0.077984	- 16.9	12.234	0.993248	- 27.5
0.9	2.055	0.011338	23.3	4.925	0.066358	-155.8	24.879	1.747855	- 121.9
1.0	2.220	0.000000	14.4	5.861	0.000000	125.8	35.445	0.000000	1513.8

specific combinations of CV and h^2 spanning the normal ranges of those parameters.

	h^2		
CV	0.2	0.4	0.6
0.1	0.045	0.063	0.077
0.2	0.089	0.126	0.155
0.3	0.134	0.190	0.232

If dominance level, number of genes, and gene frequency are specified, the value of GCV associated with a specific value of b can be calculated. Thus, by iteration, reasonable values for b (and for e) can be determined. As examples, given no dominance, $e = (b + 1)/2$ (Horner et al. 1955) and n = 10, 50 or 200, values of b (and e) required to make GCV = 0.045 in the F_2 follow:

n	b	e
10	1.041	1.0205
50	1.0182	1.0091
200	1.0090	1.0045

With this procedure to obtain the necessary values of b and e, we are able to investigate bias in gene number (n_5).

Table 7 shows results for no dominance, $e = (b + 1)/2$, and GCV = 0.089 which corresponds to any of the following paired values of h^2 and CV.

h^2	CV
0.2	0.200
0.3	0.162
0.4	0.141
0.5	0.115

The table shows that estimates of n based on equation (1) will almost always be unsatisfactory when gene effects are multiplicative. Negative values of n_5 occur when the additive genetic variance is increased by changes in gene frequency due to selection. In Table 7 with n = 50 and no dominance, the additive genetic variance is larger when $0.5 < q < 0.9$ than it is when q = 0.5. When $(\sigma^2_{gb} - \sigma^2_{gc})$ first becomes positive (as at q = 0.95 in this instance), its small size causes n_5 to be extremely large.

Complete dominance (e = b) and no genic dominance ($b = e^2$) showed similar biases.

The demonstrated biases given multiplicative gene effects strongly suggest that the procedure being considered will also be unsatisfactory in the presence of other types of epistasis.

Discussion

The sources of bias for gene number estimates have been studied individually but in the end their joint effects must be established. However, at the present level of investigation, there are two situations in which the procedure may be expected to work reasonably well:

a) when there is no epistasis,
b) when gene effects are multiplicative and logarithms of original measures are used.

It follows that procedures for discriminating among the following genetic situations are required:

a) No epistasis,
b) Multiplicative gene effects,
c) Epistasis other than multiplicative.

Whether such discrimination would always or ever be possible is a question needing further investigation. However, there are four criteria based on generation means (Mather 1949) that should theoretically agree in the absence of epistasis, even though any or all of dominance, drift, linkage disequilibrium, and inequality of gene effects are present.

$$\overline{F}_2 = \frac{\overline{F}_1}{2} + \frac{\overline{P}_1 + \overline{P}_2}{4},$$

$$\overline{B}_1 = \frac{(\overline{F}_1 + \overline{P}_1)}{2},$$

$$\overline{B}_2 = \frac{(\overline{F}_2 + \overline{P}_2)}{2} \quad \text{and}$$

$$\overline{F}_2 = \frac{(\overline{B}_1 + \overline{B}_2)}{2},$$

where P indicates a homozygous parent and B indicates a first backcross. If these equalities are not observed when satisfactory data are available, absence of epistasis cannot be assumed. Unfortunately, satisfactory data cannot always be obtained. The greatest obstacle may be the difficulty of assuming a common environment when the time pattern of development varies among the various test generations (Sentz et al. 1954).

In view of the problems encountered if there is negative dominance ($a < 0$) at most loci and selection in the primary phase is upward, the difference between the generation means of offspring and parents ($\overline{F}_1 - \overline{P}$) should always be estimated anf if it is large relative to $\overline{P}$ it should determine the direction of primary phase selection.

Given $\overline{F}_1 - \overline{P} > 0$, select upward.

Given $\overline{F}_1 - \overline{P} < 0$, select downward.

Biased from drift and from sampling variance differ from other biases because they can be controlled. Bias from drift can be reduced as much as desired by increasing the effective population size in the primary selection program. Bias from sampling variance can be reduced by doing the things (i.e. longer generation (c) in the primary selection program) required to decrease variances of estimates of $\bar{y}_c$ and $(\sigma^2_{gb} - \sigma^2_{gc})$.

Assuming absence of epistasis of a form other than multiplicative, major questions are "What are the circumstances in which the new estimator will provide satisfactory estimates?" and "How much time is saved compared with that required by the conventional methods that require complete fixation?"

Whether estimates are satisfactory can be judged by their sampling variance and the amount of bias.

Part I. of this paper indicated that if heritability is 0.2 and n = 200, a very large project would be required. Even assuming c = 20, number of sires (m) must be about 40-50 to obtain a coefficient of variation of $\hat{n}$ as small as 0.2. To reduce that to 0.1 would require many more sires (about 200). Table 1 shows that bias from drift would be about 35% given m = 20. Hence, in this situation it would seem desirable to use $m \geqslant 30$ in the primary selection phase to insure that the amount of bias would not be excessive. This estimation procedure would be very costly in cases involving large organisms with long life cycles.

When h = 0.2 but n is as small as 50, requirements are not so extreme. Assuming values of m specified in the case above where n = 200, about the same precision of n estimates is obtained with c reduced from 20 to 10 and the bias would be less. Alternatively, if c is 20, considerably smaller values of m yield satisfactory values of the coefficient of variation and of bias.

When heritability is greater (e.g., 0.4) requirements for satisfactory precision and bias are much less. Length of the primary selection program and/or the values of m can be greatly reduced from those required when heritability is 0.2.

After initial heritability has been determined, an investigator may proceed with confidence if heritability is large or decide not to proceed if heritability is too low (say less than 0.2). He may also be faced with possibilities of satisfactory results if n is moderate in size, or unsatisfactory results if n is large.

To evaluate the amount of time saved by the proposed procedure, one needs information on the time required by older methods that involve selecting until there is no further response. Equation (25) from Part I. of this paper can be used to obtain the approximate allele frequencies after c generations of selection. The value given by the equation approaches 1.0 asymptotically and therefore a value less than 1.0 must be used. For the present purpose, 0.98 will be

used because in the case of additive genes it represents 96 % of all possible response and only an 8 % downward bias in the estimate obtained, assuming complete fixation.

The numbers of generations obtained in four specific situations follow.

$h^2 = 0.2$		$h^2 = 0.4$	
$n = 50$	$n = 200$	$n = 50$	$n = 200$
35	70	25	50

Here the selection differential has been assumed to be 1.2 as before. For $c = 10 \sim 20$, from 20 to 70 % of the time required by the older methods would be saved.

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. Genetics Lectures (Oregon State University) 1, 137-148 (1969)

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)

Comstock, R.E.; Robinson, H.F.: Estimation of average dominance of genes. In: Heterosis, pp. 494-516. Ames, Iowa: Iowa State University Press 1952

Enfield, F.D.; Comstock, R.E.; Goodwill, R.; Braskerud, O.: Selection for pupa weight in *Tribolium castaneum*. II. Linkage and level of dominance. Genetics 62, 849-857 (1969)

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

Horner, T.W.; Comstock, R.E.; Robinson, H.F.: Nonallelic gene interactions and the interpretation of quantitative genetic data. North Carolina Agr. Exp. Sta. Tech. Bull. 118 (1955)

Mather, K.: Biometrical genetics. Dover Publications, Inc. 1949

Moll, R.H.; Lindsay, M.F.; Robinson, H.F.: Estimates of genetic variances and level of dominance in maize. Genetics 49, 411-423 (1964)

NaPuket, S.R.: Estimation of number of genes controlling quantitative traits. Ph.D. Thesis. Univ. of Minn. (1967)

Park, Y.C.: Variance of gene frequencies from recurrent selection in finite populations. Genetics 85, 339-346 (1977)

Park, Y.C.: Theory for the number of genes affecting quantitative characters I. Estimation of and variance of the estimate of gene number for quantitative traits controlled by additive genes having equal effects. Theor. Appl. Genet. 50, 153-161 (1977)

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 Wisconsin Press 1967

Sentz, J.C.; Robinson, H.F.; Comstock, R.E.: Relations between heterozygosis and performance in maize. Agron J. 46, 514-520 (1954)

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

Wehrhahn, C.; Allard, R.W.: The detection and measurement of the effects of individual genes involved in the inheritance of a quantitative character in wheat. Genetics 51, 109-119 (1965)

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

Wright, S.: The analysis of variance and the correlations between relatives with respect to deviations from an optimum. J. Genetics 30, 243-256 (1935)

Wright, S.: The genetics of quantitative variability. In: 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.)