Change of variance

Hossein Jorjani
Last updated: November 2007

In the absence of selection *(Finite locus model)*

Starting from the basic QG model for the phenotypic value, genetic value and environmental deviation (and following Bulmer's (1971) notation), we have:

[1] $P = G + E$

And for the variances we have:

[2] $V_P = V_G + V_E$

Please notice that in Eq. 2 genotype-environment correlation and genotype-environment interaction have been assumed to be zero.

The genetic variance can be further partitioned into two parts for the additive value and dominance deviation:

[3] $V_G = V_A + V_D$

Please notice that is Eq. 3 epistatic interaction has been assumed to be zero.

Under the polygenic model (a less restrict version of the infinitesimal model) the total genotypic value (G) is:

[4] $G = \sum_{i=1}^{N} g_i$

Where g_i are the genotypic value of locus *i*. The genetic variance is then:

[5] $V_G = Var(G) = Var(\sum_{i=1}^{N} g_i) = \sum_{i=1}^{N} Var(g_i) + \sum_{i \neq `j}^{N} Cov(g_i, g_j)$

Please notice that the genetic variance has now been partitioned into two parts:

[6] $\sum_{i=1}^{N} Var(g_i)$

[7] $\sum_{i \neq `j}^{N} Cov(g_i, g_j)$

The first part (Eq. 6) is the within locus component of the genetic variance and therefore contains even the dominance variance. This part may also (loosely speaking) be called the genic variance.

In Falconer and Mackay's book (1996) the second part (Eq. 7) has been assumed to be zero in many places, especially in Chapter 8, where the genetic variance is discussed. In F&M (1996) terminology Eq. 6 is the sum of additive and dominance variance, i.e.

[8] $\sum_{i=1}^{N} Var(g_i) = V_A + V_D = \sum 2pq[a + d(q - p)]^2 + \sum (2pqd)^2$

Please notice that when there is a summation over many loci, there should be a subscripts for *a*, *d*, *p* and *q*. However, for the sake of clarity I have not used them in eq. 8.

In the presence of selection *(Finite locus model)*

The second part of Eq. 5 (which is Eq. 7) is the across loci component of the genetic variance and reflects gametic phase (linkage) disequilibrium. Under random mating and without selection Eq. 7 is equal to zero. However, under selection (especially stabilizing selection) joint genotype frequencies of different loci (two or more loci considered jointly) are different from their expected equilibrium values. In other words, under selection the Mendelian law of random assortment of genes is violated and advantageous genotypes of different loci are accompanying each other more than is expected from Hardy-Weinberg ratios. This is the gametic phase (linkage) disequilibrium. In the context of reduction of variance because of selection, it is also called "the Bulmer effect".

In the absence of selection *(Infinitesimal model)*

Now let's see what happen under selection:

Consider the following model depicting regression of offspring phenotype (Y) on both parents' phenotypes (P_1 and P_2):

[9] $Y = a + b_1 P_1 + b_2 P_2 + e$

Assuming symmetry of regression of offspring on female and male parents, i.e. $b_1 = b_2$, we have

[10] $Y = a + b(P_1 + P_2) + e$

Covariance of offspring phenotype (Y) on both parents' phenotypes (P_1 and P_2), assuming the phenotypic variance in the parental and offspring generations are equal, is:

[11] $Cov(Y, P_1 + P_2) = Cov(Y, P_1) + Cov(Y, P_2)$

From Falconer & Mackay (1996) recall that:

[12] $Cov_{OP} = ½ h^2 V_P = Cov_{Y,P_1}$

Therefore,

[13] $Cov(Y, P_1 + P_2) = Cov(Y, P_1) + Cov(Y, P_2) = 2(½ h^2 V_P)$

Please also notice that:

[14] $Var(P_1 + P_2) = 2V_P$

Therefore,

[15] $b = \frac{Cov(Y, P_1 + P_2)}{Var(P_1 + P_2)} = \frac{h^2 V_P}{2V_P} = ½ h^2$

Now, we can define the phenotypic variance as:

[16] $Var(Y) = b^2 Var(P_1 + P_2) + Var(e)$

[17] $Var(Y) = b^2 \left[Var(P_1) + Var(P_2)\right] + Var(e)$

[18] $Var(Y) = V_P = 2b^2 V_P + Var(e)$

[19] $Var(e) = V_P - 2b^2 V_P = (1 - 2b^2) V_P = (1 - 2(¼ h^4)) V_P = (1 - ½ h^4) V_P$

In the presence of selection (Infinitesimal model)

In order to understand the Bulmer effect we should consider the following paradox (Bulmer, 1971):

"This paper arose from the following apparently paradoxical consideration. If a metric character is subjected to stabilizing selection, which reduces the phenotypic variance in each generation by weeding out extreme deviations from the norm, one would expect the genetic variance to be reduced proportionately in each generation and ultimately to be eliminated altogether. But Crow and Kimura (1970) show that the change in the variance under selection decreases as the number of loci involved increases and must eventually tend to zero when the number of loci becomes effectively infinite. Also, most metric characters investigated in natural and domesticated populations possess much genetic variability, and are likely subject to stabilizing selection (Falconer 1960 [=Falconer and Mackay 1996]). The intuitively obvious argument that stabilizing selection will lead to the rapid elimination of genetic variability must therefore be wrong. Why?"

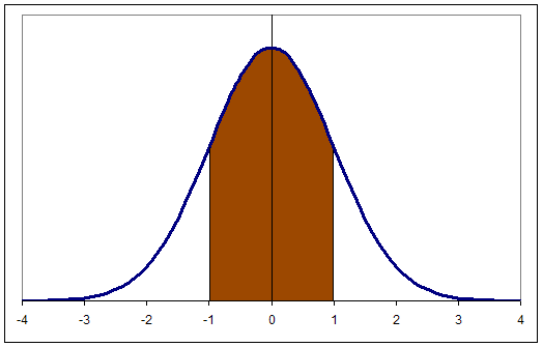


Figure 1- Schematic representation of the stabilizing selection.

From Eq. 18 recall that:

[18] $Var(Y) = 2b^2 V_P + Var(e)$

Under stabilizing selection the offspring generation's variance $Var(Y^*)$ would be less than in the unselected population, because as can be seen in Figure 1, selected parents' phenotypic variance would be less than the unselected population's phenotypic variance. Let's say that the selected parents' phenotypic variance is $V_P^{'}$, i.e.:

[20] $V_P^{'} = V_P + d_{V_P}$

where d_{V_P} is the change in the variance because of selection of parents. Then,

[21] $Var(Y^*) = 2b^2 (V_P + d_{V_P}) + Var(e)$

Now, after using a little algebra and if we replace b for $½h^2$ from Eq. 15 and $Var(e)$ for $(1-½h^4)V_P$ from Eq. 19, then we get:

[22] $Var(Y^*) = 2b^2 V_P + 2b^2 d_{V_P} + (1-½h^4)V_P$

[22.a] $Var(Y^*) = 2(¼h^4)V_P + 2(¼h^4) d_{V_P} + V_P - ½h^4 V_P$

[22.b] $Var(Y^*) = ½h^4 V_P + ½h^4 d_{V_P} + V_P - ½h^4 V_P$

[23] $Var(Y^*) = V_P + ½h^4 d_{V_P}$

Recall that d_{V_P} is a negative quantity, therefore the phenotypic variance decreases by $½h^4 d_{V_P}$ after one generation of selection. Again, that is the Bulmer effect.

Items to be added
Add Bulmer (1976)
Correspondence between Bulmer (1971, 1976) and Falconer & Mackay (1996) equations

Breeding value for variance:
Kolmodin et al. (2002, 2003, 2004)
Hill (2002), Hill & Zhang (2004)
San Crisobal-Gaudy et al. (1998)
Sorensen & Waagepetersen (2003)

Assumptions of the Hardy-Weinberg Model (HWM)

Hossein Jorjani
Last updated: September 2007

Falconer & Mackay (1996) Introduction to Quantitative Genetics

Explicit

- Large population
- Random mating
- No selection, migration or mutation

Implicit

- Diploid organism
- Sexual reproduction
- Normal segregation
- Reciprocal equality of mating gametes (i.e. $A_1A_2 = A_1A_2$)
- Equal allele frequency in males and females
- Locus is not sex-linked
- Large population (negligible sampling variance)
- Random mating
- No selection (equal fertility, equal viability)
- No migration

Hartl & Clark (1989) Principals of population genetics

- Large population
- Random mating
- Diploid organism with sexual reproduction
- ***Non-overlappning generations***
- Migration and mutation negligible
- Natural selection does not affect the gene under consideration

Crow & Kimura (1970) An introduction to population genetics theory

- Large population
- Random mating
- Same allele frequency in males and females

The population will approach equilibrium gradually if

- Genes are sex-linked
- Allele frequencies are different in the two sexes
- Asexual and sexual reproduction is possible
- Generations are overlapping

Lynch & Walsh (1998) Genetics and analysis of quantitative traits

- Infinite population size
- Random mating
- Autosomal locus uninfluenced by selection and mutation
- *Discrete generations*
- No migration

Ayala & Kiger (1980) Modern Genetics

- Absence of mutation, migration, drift and selection for allele frequencies to remain constant
- Random mating for genotype frequencies to be simply related to allele frequencies
- Autosomal locus, not sex-linked for equilibrium to be reached immediately

Cavalli-Sforoza & Bodmer (1971) The genetics of human populations

- Random mating
- Infinite population size
- Diploid organism
- Negligible mutation rates
- No selection
- ***Genotypes can be distinguished unequivocally***
- Normal segregation
- Discrete generations

Is REML biased?

Hossein Jorjani
Last updated: November 2001

ML is biased for two reasons:

1. Estimates are constrained to lie within the range of the parameters, and for variance components they cannot be less than zero;

2. It assumes that fixed effects are known, i.e. the loss of degrees of freedom due to fitting these effects is ignored.

REML takes care of (2) by maximizing only the part of the likelihood that is independent of the fixed effects.

How?
This is done by replacing the data by linear functions of them, 'error contrasts', with an expectation of zero.

What?
It is equal to adjustment of the observations for the generalized least-square estimates of the fixed effect.

Properties of ML

I. Accounts for selection (only under certain conditions. This means all information that have contributed to the selection decisions must be included in the analysis= all individuals + all traits' measurements[1]).
II. ML estimates are consistent, i.e. they are asymptotically normal and efficient.

Assumptions and requirements:

Distribution of the data is known. Generally, the data is assumed to be distributed as multivariate normal. This leads to two properties:

a- Regression of trait B on trait A is unbiased if A is selected
b- Conditional variance of B given A, ($V_{(B|A)}$), is not affected by selection of A.

What happened to the (1) above?

[1] Of course, this is still under debate! For example look at Sorensen et al (2001).

Mutual Contributions of Quantitative Genetics and Molecular Genetics

Hossein Jorjani
Last updated: September 1999

Falconer & Mackay, 1997
Selection in one form or another is the means whereby all improvement of domesticated animals and plants has been made.

Selection means breeding from the 'BEST' individuals, whatever 'BEST' may be.

$R = h^2S$

$R_L = h^2 S / L = h^2 i \sigma_P / L = h\, i\sigma_A / L$

where
R_L= Response to selection per unit of time;
h^2= Heritability of the trait;
h= Accuracy of selection;
S= Selection differential;
i= Standardized selection differential or intensity of selection;
σ_P= Phenotypic standard deviation;
σ_A= Additive genetic standard deviation;
L= Unit of time (commonly expressed as generation interval).

The genetic change depends on:	To increase the genetic change:
The population variation (V_A);	Generate new variation (a);
The heritability (h^2);	Use uniform environment (b);
The accuracy of selection (r_{IA});	Devise statistical methods that make better use of records (c);
	Measure the physical attribute more appropriately (d);
	Use physiological criteria of the genetic merit (e);
	Identify genes or markers for the trait under selection (f);
The selection intensity (i);	Increase population size (g);
The generation interval (L).	Lower the age of parents at the time of reproduction (h).

Skepticism about quantitative genetics theory

1. Individual selection has successfully been used for thousands of years in plants and animals;
2. Usage of sib selection goes back at least to Hebrew laws and progeny testing at least to 2000 years ago;
3. Discrepancies of observed and predicted responses;
4. Inconsistency of reports on genetic variance reduction;
5. Biological phenomena which do not follow assumptions and theories

Claims

Kempthorne (1977): Without knowledge of quantitative genetics one can embark on practicing selection and do so successfully. Simple Mendelian genetics and complicated quantitative genetics are only some tools for argumentation of statistical procedures. Without them experiments should be only larger than they are at the moment.

Lerner (1958): Mendelian inheritance, viewed comprehensively, is a *statistical* phenomenon, and evolution, in nature or under man's control, is a process depending on a subtile balance of forces which can largely be described in *probability* terms.

Nordskog (1977): we are, in fact, really in dark as to how much genetics has really contributed to the improvement that we observe … the majority of theories are unverified … Lush's combination selection index has *never* been proved by experimental test.

Lewontin (1977): Two general sorts of knowledge are required to provide a better understanding of mechanisms underlying quantitative characters;

> (*i*) Structure of the genome; that is number of loci, their linkage relations, their allelic number and frequency distribution and the mutational and recombinational sources of new variation;
>
> > Contribution of Molecular biology:
> > Detection of QTL is the answer to loci number;
> > Gene mapping can be employed to detect linkage groups;
> > Detection of variation at the DNA level in transgenic animals and their crosses provides enough information on allelic number and frequency distribution and also mutational or recombinational sources of variation.
>
> (*ii*) Relations between gene and organism; that is how gene action, in particular environments, is translated into phenotype.
>
> > Contribution of Quantitative Genetics:
> > Analysis of the data from part (*i*) with the statistical methods developed in the framework of quantitative genetics: Genetically engineered combination of genes at the most precise levels are needed to be examined in different environments and the obtained data analyzed in order to have a reasonable picture of genotype by environment interaction (norm of reaction of genotypes in different environments).

To mate or not to mate, that is not the question.

Hossein Jorjani
Last updated: September 1997

Lars-Erik Liljedahl (Personal communication, 1995)

- Mating
  - Random
  - Non-random
    - Inbreeding
    - Outbreeding
    - Assortative mating

Falconer & Mackay (1996)

- Mating
  - Random
  - Non-random
    - Inbreeding
    - Assortative mating

Hartl & Clark (1989)

- Mating
  - Random
  - Assortative

Pirchner (1983)

- Mating
  - Random
  - Assortative
    - Positive
      - Genetic = Inbreeding
      - Phenotypic = Positive assortative mating
    - Negative
      - Genetic = Outbreeding
      - Phenotypic = Negative assortative mating

Spiess (1977)

- Broadly: Mating
  - Random
  - Assortative
    - Grouping of mates by common ancestral relationship = Inbreeding
    - . . .
- Ordinarily: Mating
  - Without grouping = Random
  - With grouping
    - By genotype via common ancestor = Inbreeding
    - By phenotypic criteria = Assortative

Jorjani (1995)

Process(es) at work	Observed pattern
Random mating	r_A = Average $r_P = 0.0$
Assortative mating	r_A = Average $r_P \neq 0.0$
Inbreeding / outbreeding	$r_A \neq$ Average $r_P = 0.0$
Inbreeding / outbreeding Assortative mating	$r_A \neq$ Average $r_P \neq 0.0$

Realized heritability (h_r^2)

Hossein Jorjani
Last updated, November 2002

Does the realized heritability (h_r^2) take into account the inbreeding depression?

The realized heritability (h_r^2) actually measures the heritability in the base population (h_o^2)!

However, the h_r^2 **underestimates** h_o^2 because:

1. There is a reduction of variance and consequently a reduction of selection response after the first generation of selection;
2. Systematic environmental effects are neglected;
3. Inbreeding depression leads to an underestimation of the h_o^2, unless we use control population(s);
4. Changes due to random genetic drift affect generation mean and the h_r^2, unless we use replicated selection lines.

Is this for real?

Selection Differential

Hossein Jorjani
Last updated: December 1995

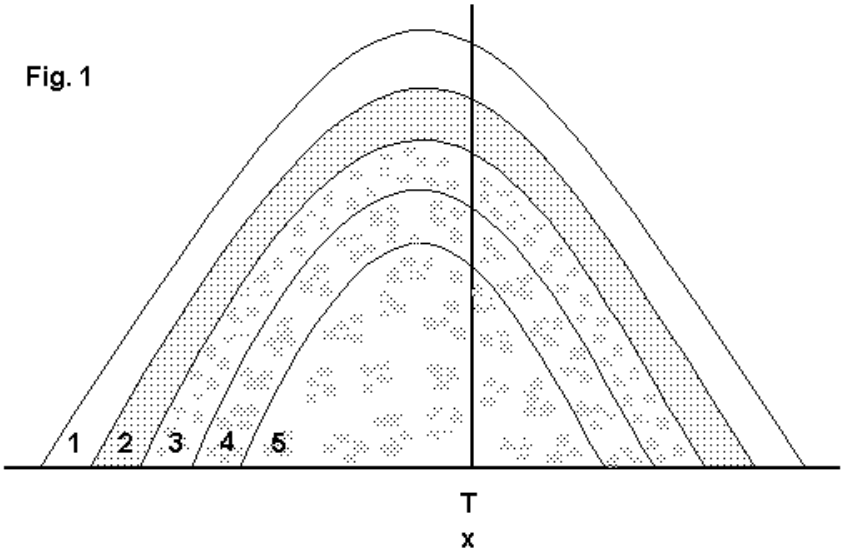


Figure 1: Schematic representation of the various populations with increasing structural complexity considered in theoretical and applied quantitative genetic studies.

Population 1: A theoretical population comprising of an infinitely large number of uncorrelated observations (individuals) with a normal distribution.

Population 2: The same as Population 1 but with finite size.

Population 3: The same as Population 2 but members of the population are grouped in families of the varying size.

Population 4: The same as Population 3 but after some generations in which combination of selection and finite population size has caused inbreeding so that some families share one or more common ancestors. Matings are assumed to be random.

Population 5: the same as Population 4 but matings are assortative.

A theoretical normal population, consisting of infinite number of independent observation points (schematically represented as Population 1 in Figure 1), can be described by the normal probability density function:

$$f(x) = z = \frac{1}{\sigma\sqrt{(2\pi)}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{(A1)}$$

where
z = the height of the ordinate at the observation point;
x = the deviation of the observation point from the population mean;
μ = the population mean; and
σ^2 = the variance of the observation points (population).

The cumulative distribution function for such a curve (population) is given by

$$F(x) = \frac{1}{\sigma\sqrt{(2\pi)}} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \quad \text{(A2)}$$

In such a population with size n the expected order statistics (EOS) of the j^{th} top ranking individual (x_j) is:

$$E(x_{j|n}) = \frac{n!}{(n-j)!\,(j-1)!} \int_{-\infty}^{\infty} x\,[‰ - G(x)]^{j-1}\,[‰ + G(x)]^{n-j}\,f(x)\,dx \quad \text{(A3)}$$

where

$$G(x) = \frac{1}{\sigma\sqrt{(2\pi)}} \int_{0}^{x} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \quad \text{(A4)}$$

If and when observations are independent, as is assumed in a theoretical normal distribution, the mean of the top ranking EOS, i, is obtained by simple averaging

$$i = \frac{1}{j}\sum_{1}^{j} E(x_{j|n}) \quad \text{(A5)}$$

This quantity, defined as *standardized selection differential* or *intensity of selection* in animal breeding context (Falconer, 1989), is the average superiority of the selected individuals and can also be obtained by:

$$i = z/p \quad \text{(A6)}$$

where z is the height of the ordinate at the smallest observation point of the j top ranking individuals, or as is known in the animal breeding context, the point of truncation (T), and p is the proportion of the population selected comprised of j top ranking individuals (the area under the normal curve). If a portion of the population is selected by the method of truncation, p is the proportion lying beyond the point of truncation (T) whose deviation from the population mean is designated as x. This area can be obtained from:

$$F(T \le x \le \infty) = p = 1 - \int_{-\infty}^{T} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx \qquad \text{(A7)}$$

Because of the difficulties associated with application of these equations it is more convenient to work with tabulated results of the numerical integrations of these equations or some approximations of them.

When *p* or *x* is known

Assuming the population under consideration is large (e.g. several thousand individuals) and if the proportion selected, *p*, or the deviation of the truncation point from the population mean, *x*, is known, suitable tables could be used to find out corresponding values for *x* and *i*. Falconer (1965) has provided such a table, a summarized version of which can be found in Falconer (1989) as Appendix Table A or similarly in Tables 1 and 2 in Pearson and Hartley (1972).

Because repeated reference to such tables is burdensome and, furthermore, these tables cannot be used in computer programs, it is helpful to use some approximations for these equations. If $p \le 0.5$, then according to Abramowitz and Stugen (1964) *T* can be calculated as:

$$T = x = t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3} + \varepsilon \qquad \text{(A8)}$$

in which

$$t = \sqrt{\ln \frac{1}{p^2}} \qquad \text{(A9)}$$

where c_0=2.515517 d_1=1.432788
c_1=0.802853 d_2=0.189269
c_2=0.010328 d_3=0.001308 and $|\varepsilon| < 4.5 \times 10^{-4}$.

Substitution of *T* (=*x*) from Eq. A8 into Eq. A1 provides *z* which can be used to obtain *i* from Eq. A6. For $p>0.5$, $x_p = -x_{(1-p)}$ and $i_p = ((1-p)/p)\, i_{(1-p)}$.

When *T* is known

If *T* is known, the following approximation can be used to calculate the proportion selected:

$$F(T \le x \le \infty) = p = Z^* (b_1 t + b_2 t^2 + b_3 t^3 + b_4 t^4 + b_5 t^5) + \varepsilon(T) \qquad \text{(A10)}$$

in which

$$t = \frac{1}{1 + wT} \qquad \text{(A11)}$$

where Z^* can be obtained from Eq. A1 by replacing *x* for *T* and equating μ=0 and σ=1,

w= 0.2316419 b_3= 1.781477937
b_1= 0.319381530 b_4=-1.821255978
b_2=-0.356563782 b_5= 1.330274429, and $|\varepsilon| < 7.5 \times 10^{-8}$.

When population is small

Harter (1961) under the usual assumptions of infiniteness of the population, using numerical integration, has provided table of the EOS of the j^{th} top ranking individual for a sample of size n, where n=2, ..., 400 and j=1, ..., 200, where applicable. Similar, but less extensive, tables have been provided by Pearson and Hartley (1972) in their Table 9. Averages of the EOS values from Harter's table are presented in Becker (1984) and summarized in Falconer's (1989) Appendix Table B. equivalently Table XX of Fisher and Yates (1963) can be used for the same purpose.

Considering the finite size in Population 2 the EOS of the j^{th} top ranking individuals should differ from that of Population 1. Borrows (1972) has presented the following equation for adjustment for finiteness of the populations:

$$i^* = i - \frac{(N - n)}{2n\,(N + 1)\,i} \qquad \text{(A12)}$$

where
i^* = the adjusted i;
i = the average of the EOS for an infinite population, *e.g.* Becker's (1984) Table 2;
N =the total number of observations; and
n = the number of top ranking observations.

Borrows' results can also be written as:

$$i^* = i - (1 - p) / [2p\,(N + 1)\,i] \qquad \text{(A13)}$$

where p, the proportion selected, is n/N.

If the population is assumed to be large random sample of a normal population, Falconer's (1989) Appendix Table A, Pearson and Hartley's (1972) Tables 1 and 2, or equivalently Equations A1-A11 can be used in A12 instead of Becker's (1984) Table 2.

When observations are correlated

The populations considered so far were assumed to consist of individual observations whose values are independent of each other. In animal breeding, however, every individual is genetically related to a part of the population, *i.e.* its family members. Rawlings (1976) and Hill (1976) have considered the EOS in animal breeding populations where observations are correlated, i.e. Population 3 in Figure 1.

Hill (1976) concluded that as the intraclass correlation of the group members, t, approaches 1 the mean EOS of the highest ranking individuals, i, is reduced below its infinite population size value. For individual selection and large population sizes the reduction is negligible, but if family mean is used as the criteria of selection or if the population size is small, the reduction in i may be substantial. Hill (1976) also provided the i values for s=2, ..., 20 and n=2, ..., 8 when p is up to 0.50.

Approximation formulas for small populations of related individuals have been proposed by Rawlings (1976):

$$i^* = [1 - \frac{(n-1)\,t}{n\,s - 1}]i \qquad \text{(A14)}$$

and Hill (1976):

$$i^* = i - (1 - p) / \{ 2ip [ns (1 - t) + st + 1]\} \qquad \text{(A15)}$$

Meuwissen (1991) has also presented an approximation that accounts for family structure, which produced better results than (A15) under conditions of an adult MOET breeding scheme.

REFERENCES

Abramowitz, M., Stegun, L. A. 1964. Handbook of mathematical functions. Dover Publications, Inc., New York, 1046 pp.

Becker, W. A. 1984. Manual of Quantitative Genetics (4th ed). Academic Enterprises, Pullman, Wash., USA., 188 pp.

Burrows, P. M. 1972. Expected selection differentials for directional selection. Biometrics, 28, 1091-119.

Falconer, D. S. 1965. The inheritance of liability to certain diseases, estimated from the incidence among relatives. Ann. Hum. Genet., 29, 51-76.

Falconer, D. S. 1989. Introduction to Quantitative Genetics. Longman, Harlow, England, 438 pp.

Fisher, R.A. & Yates, F. 1963. Statistical Tables. Longman, London and New York.

Harter, H. L. 1961. Expected values of normal order statistics. Biometrika, 48, 151-165.

Hill, W.G. 1976. Order statistics of correlated variables and implications in genetic selection programmes. Biometrics, 32, 889-902.

Meuwissen, T. H. E. 1991. Reduction of selection differentials in finite populations with a nested full-half sib family structure. Biometrics, 47, 195-203.

Pearson, E. S., & Hartley, H. O. 1972. Biometrika Tables for Statisticians. Cambridge University Press. London, 385 pp.

Rawlings, J. O. 1976. Order statistics for a special class of unequally correlated multinormal variates. Biometrics, 32, 875-887.

Selection Response

(Breeders' formula)

Hossein Jorjani

Last updated: November 1997

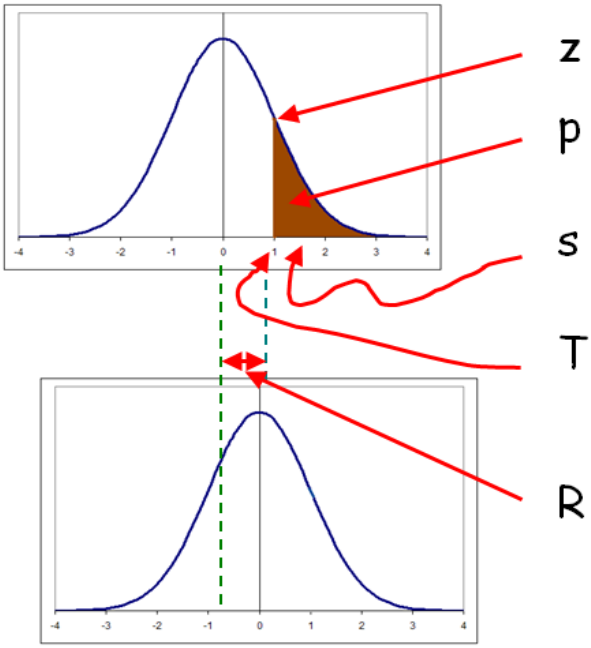
$R = h^2 S$

$R = \sigma^2_A/\sigma^2_P S$

$R = h^2 i \sigma_P$

$R = h i \sigma_A$

$R = h \sigma_A z/p$



or

$R_L = h^2 S/L$

$R_L = \sigma^2_A/\sigma^2_P S\ 1/L$

$R_L = h^2 i \sigma_P/L$

$R_L = h i \sigma_A / L$

$R_L = h \sigma_A z/p\ 1/L$

where

R	response to selection per generation;
R_L	response to selection per unit of time;
h^2	heritability of the trait; and $h^2 = \sigma^2_A/\sigma^2_P$
h	accuracy of selection (not always!); and $h = \sigma_A/\sigma_P$
S	selection differential; and $S = i\,\sigma_P = \sigma_P\ z/P$
i	standardized selection differential (intensity of selection); and $i = S/\sigma_P = z/P$
σ^2_P	phenotypic variance;
σ^2_A	additive genetic variance;
z	height of the ordinate at the point of truncation in a normal Gaussian curve describing the trait under selection;
p	proportion selected; and
L	unit of time (*e.g.* generation interval).

Some definitions[1]

Hossein Jorjani
Last update, September 2001

Finite Polygenic Model
The FPM assumes biallelic loci, additive gene action, constancy of additive effects and allele frequencies across loci, and fits polygenic number rather than genotypes at individual loci. Du & Hoeschele (2000)

Finite Locus Model
The FLM assumes biallelic loci but fits genotypes at individual loci, and allows for non-additive gene action and variable gene effects across loci. Allele frequencies are still held constant at 0.5, but could be estimated in the FLMs. Du & Hoeschele (2000)

An alternative to the infinitesimal model is a finite locus model. Initially it was assumed that the trait was controlled by a specified number of genes all with equal effect and gene frequencies of 0.5. Goddard (2001)

The Infinitesimal Model
This model assumes that there are a very large number of genes each with a very small, additive effect and ignores mutation. Goddard (2001)

Extended Infinitesimal Model
Effects other than breeding value can be added to the infinitesimal model, e.g. maternal, dominance, epistatic, cytoplasmic or imprinted effects. Goddard (2001)

[1] References are not to the original sources, but to a recent publication.

Impact of molecular genetic data on choice of genetic models

Hossein Jorjani
Last updated: November 2004

The entire field of quantitative genetics and much of the field of population genetics deal with continuous variation at the phenotypic level. The most important genetic model used to study the continuous variation, the infinitesimal model (Fisher, 1918; Wright, 1921), maintains that continuous variation follows the Mendelian rules of transmission genetics. However, there are two main differences between the simple single locus model and the infinitesimal model. Under the infinitesimal model it is assumed that a) a very large number of loci, possibly infinite number of loci, control the character/trait under consideration; and b) the effect of each locus on the phenotypic variance is very small, possibly infinitesimally small. Combining these two properties of the infinitesimal together enables us to invoke central limit theorem and consequently the normal distribution for genetic and environmental values and their sum, i.e. phenotypic values.

In the light of results obtained in the field of molecular genetics, and especially published results of microarray and QTL studies, it is quite clear that the infinitesimal model is not a correct model. A brief description of the evidence for and against the infinitesimal model is as follows. Needless to say that many of the examples mentioned below are taken from animal genetics, which hitherto has been the subject of my research.

Evidence in favor of the infinitesimal model

The Infinitesimal model was initially a theoretical construct which stemmed from mathematical treatment of traits with continuous distribution (Fisher, 1918; Wright, 1921; see also Crow & Kimura, 1970; Bulmer, 1971, 1985; Turelli & Barton, 1994). Adherence to it also was because of its mathematical tractability rather than a belief in its physical correctness (Hill, 1994). However, there are at least three lines of evidence that suggest it actually is not far from being true.

Biochemical evidence

From a biochemist's perspective virtually all enzymes affect almost all traits. According to the 'metabolic control theory' (e.g. Kacser & Burns, 1979, 1981; Hofmeyr & Cornish-Bowden, 1991; Bagheri & Wagner, 2004) flux of any enzymatic chain is affected by all enzymes in that chain. However, effect of each enzyme follows a hyperbolic function. Following a suggestion by Hartl et al. (1985), it has also been shown that the control coefficient exerted by any enzyme is subject to change under selection (Dean et al., 1986; Dykhuizen et al., 1987). Consequently, effects of some enzymes may be so small in certain generations that their effect might not be measurable in small sample sizes. Kacser (1989) went so far as to claim that in the long run any quantitative trait may be under control of a very large number of loci, say 5000!

Comparison of per trait mutation rate and per locus mutation rate

Close examination of natural and laboratory populations reveals that per trait rate of mutation is between 10^3-10^4 times larger than per locus rate of mutation (e.g. Houle et al., 1992; Keighley & Hill, 1992; Santiago et al., 1992; Mackay et al., 1994). By comparison of these two rates one can easily deduce that the number loci affecting a trait may be as high as 1000-10000.

Microarray data

Results of the various genome projects put the total number of genes between 30000 and 50000. The question is how many of them affect each trait. We know from old 2-dimentional gel electrophoresis studies (e.g. Bullfield, 1982) that each cell line may contain several thousands of proteins. Data from modern microarray studies are pouring in and it may be too soon to form an opinion on the number of loci affecting a trait. However, and for example, in a study of seven brain tumors Watson et al. (2001; see also Shannon et al., 2002) found differential expression for 196 genes out of 1013 genes studied.

Extrapolation of these numbers to total number of genes expressed in any cell line, and then to total number of cell lines affecting any quantitative trait easily puts the number of loci to thousands for any trait of continuous distribution.

Evidence against the infinitesimal model

When it comes to the other assumption of the infinitesimal model, i.e. the effects of all loci on variance being small, there are two lines of evidence that this assumption is not true.

Observations from long term selection experiments

In many long term selection experiments one can find evidence that certain alleles have a large effect on the variance. In Yoo's famous long term selection for bristle number in Drosophila (Yoo, 1980) recessive lethals were responsible for some jumps in selection response. Evidence for loci of large effect sometimes manifest itself in form of selection plateau (or even selection limit). One example is Falconer's experiment in mice (Falconer, 1973) in which a selection plateau was observed after 27 generations of selection.

QTL studies

Of course there are many difficulties to detect QTLs (for some discussion on this see for example: Otto & Jones, 2001; Hayes & Goddard, 2001; Barton & Keightly, 2002) and therefore, may be, it is prudent to assume that the number of reported QTLs in any study is the minimum number of existing QTLs for that trait and that population. In any case, studies aiming at finding QTLs provide the best evidence for loci having different sizes of effects. First of all, usually very few QTLs are detected. In swine, Andesson-Eklund et al. (1998) found only 1-2 QTLs for each of the 24 traits studied. In two recent studies national cattle populations were screened for QTLs affecting economically important traits. Olsen et al. (2002) found 2-3 QTLs for each of the five milk production traits in the Norwegian cattle population. In Germany, Kuhn et al. (2003) found 2-4 QTLs for each of 8 functional traits in their study. Even when a larger number of loci are found, it turns out to be so that a fraction of them are responsible for a large fraction of variance. Goldman et al. (1993, 1994) found 18-26 QTL's for four traits in maize and among these 4-7 QTLs were responsible for 43%-84% of the variation. In the above mentioned study by Andersson-Eklund et al. (1998) the 1-2 QTLs were responsible for 9%-16% of the variation.

Synthesis

Combining all the evidence for and against the infinitesimal model we must come to the conclusion that *potentially* a large number of loci are responsible for expression of any trait with continuous distribution. In the long term, it is possible that the "average" effect of all loci, across a large number of generations and environments, are small and probably more or less equal. However, in the short term and in any specific population, only a fraction of them have discernible and measurable effect on the trait of interest.

Conventional methods of detecting major genes, QTLs, and even differentially expressed genes in cases where more than two states are considered, rely heavily on statistical methods that, directly or indirectly, rely on the infinitesimal model and certain statistical distributions that may not be entirely appropriate. Therefore, conventional methods of detecting major genes, QTLs and microarrays need to be modified to accommodate the facts put forward by recent molecular genetic findings. This is an exciting line of research to pursue.

References

Andersson-Eklund, L., Marklund, L., Lundström, K., Haley, C. S., Andersson, K., Hansson, I., Moller, M., & Andersson, L. 1998. J. Anim. Sci. 76: 694-700.

Bagheri, H. C., & Wagner, G. P. 2004. Genetics, 168: 1713-1735.

Barton, N. H. & Keightley, P. D. 2002. Nature Review 3: 11-21.

Bulmer, M. G. 1971. Am. Nat., 105, 201-211.

Bulmer, M. G. 1985. The Mathematical Theory of Quantitative Genetics. Oxford University Press.

Crow, J. F. & Kimura, M. 1970. An Introduction to Population Genetics Theory, Harper and Row.
Dean, A. M., Dykhuizen, D. E. & Hartl, D. L. 1986. Genet. Res. 48, 1-8.
Dykhuizen, D. E. Dean, A. M. & Hartl, D. L. 1987. Genetics 115, 25-31.
Falconer, D. S. 1973. Gen. res., Camb. 22: 291-321.
Fisher, R. A. 1918. Trans. Roy. Soc. Edin., 52, 399-433.
Goldman, I. L., Rocheford, T. R. & Dudley, J. W. 1993. Theor. Appl. Genet. 87: 217-224.
Goldman, I. L., Rocheford, T. R. & Dudley, J. W. 1994. Crop Sci. 34: 908-915.
Hartl, D. L., Dykhuizen, D. E. & Dean, A. M. 1985. Genetics 111: 655-674.
Hayes, B. & Goddard, M. E. 2001. Genet. Sel. Evol. 33: 209-229.
Hill, W. G. 1994. Proc. of the 5th World Cong. on Genet. Appl. to Livest. Prod., University of Guelph, 19: 125-126.
Hofmeyr, J. H.-S. & Cornish-Bowden, A. 1991. Eur. J. Biochem., 200, 223-236.
Houle, D., Hoffmaster, D. K., Assimacopoulos, S. & Charlesworth, B. 1992. Nature 359: 58-60.
Kacser, H. & Burns, J. A. 1979. Biochem. Soc. Trans., 7, 1149-1160.
Kacser, H. & Burns, J. A. 1981. Genetics 97: 639-674.
Kacser, H. 1989. pp 219-226 In: Hill, W. G. & Mackay, T. F. C. (eds) Evolution and animal Breeding. CAB International, Oxon,
Keightley, P. & Hill, W. G. 1992. Genetics 131: 693-700.
Kuhn, Ch., Bennewitz, J., Reinsch, N., Xu, N., Thomsen, H., Looft, C., Brockmann, G. A., Schwerin, M., Weimann, C., Hienleder, S., Erhadt, G., Medjugorac, I., Förster, M., Brenig, B., Reinhardt, F., Reents, R., Russ, I., Averdunk, G., Blumel, J. & Kalm, E. 2003. J. Dairy Sci. 86: 360-368.
Mackay, T.F.C., Fry, J.D., Lyman, R.F. and Nuzhdin, S.V. 1994. Genetics 136:937-951.
Olsen, H. G., Gomez-Raya, L., Våge, D. I., Olsaker, I., Klungland, H., Svendsen, M., Ådnoy, T., Sabry, A., Klemetsdal, G., Schulman, N., Krämer, W., Thaller, G., Ronningen, K. & Lien, S. 2002. J. Dairy Sci. 85: 3124-3130.
Otto, S. & Jones, C. D. 2000. Genetics, 156: 2093-2107.
Santiago, E., Albornoz, J. Dominguez, A. Toro, M. A. & Lopez-Fanjul, C. 1992. Genetics 132: 771-781.
Shannon, W., Watson, M. A., Perry, A., & Rich, K. 2002. Genetic Epidemiology 23: 87-96.
Turelli, M., & Barton, N. H. 1994. Genetics 138: 913-941.
Watson, M. A., Perry, A., Budhjara, V., Hicks, C. Shannon, W. & Rich, K. 2001. Cancer Res 61: 1825-1829.
Yoo, B. H. 1980. Genet. Res. 35: 1-17.