Prediction of response to long-term selection considering skewness of distribution and curvilinear relationship between traits caused by previous selection

A. Nishida, H. Ogasawara , T. Konari, Y. Ohtomo and H. Shinohara

Graduate School of Agricultural Science, Faculty of Agriculture, Tohoku University,
981-8555 Aoba-ku, Sendai-shi, Japan,. E mail: anishida@bios.tohoku.ac.jp
(2004 EAAP Annual Meeting in Bled, Slovenia)

Introduction

Prediction of response to repeated selection in livestock population is necessary to choose the most efficient selection program considering the successive changes in genetic characteristics of the population from one generation to the next (Nishida and Abe, 1980a). A theory of “continuous distribution model” for the prediction of the response to repeated selection was summarized from its basic idea to the estimation of the gamete’s genetic distribution.

The basic idea of the prediction theory

Assuming phenotypic value; *P*, is sum of the breeding value; *G*, and the environmental effect; *E*, in a coordinates with the axes of *G* and *E*, *P* axis is in the middle of *G* and *E* axis as given in Figure 1. On the coordinates, truncation selection based on *P* means to take a part of joint distribution of *G* and *E* cut by the plane stands perpendicularly on the truncation line which meets at right angles with *P* axis. The marginal distribution of *G* for the selected part is $f(G) = \int_s F(G,E)dE$, where $F(G,E)$ is the joint distribution and *s* means the restriction in range of integral to the selected portion. Thus the mean, the variance (Nishida and Hayashi, 1980; Robertson, 1977) and the skewness of the marginal distribution are as follows.

$\overline{G} = \int_{gl}^{gu} G \cdot f(G)dG/p$, $\sigma^2{}_G = \int_{gl}^{gu} (G-\overline{G})^2 f(G)dG/p$, $SK_G = \int_{gl}^{gu} (G-\overline{G})^3 f(G)dG/p\sigma^3{}_G$ where p is the proportion of selection, *gl* and *gu* are the lower and upper limit of the distribution.

If it is assumed for the simplicity that the distribution of breeding values of selected animals does not change in the process of reproduction, the bivariate distribution of *G* and *E* for the next generation can be obtained by random coupling of these two variables. Then the repeated use of the basic idea given in Fig. 1 and the equations shown above enables us to predict long-term selection response (Fi.2).

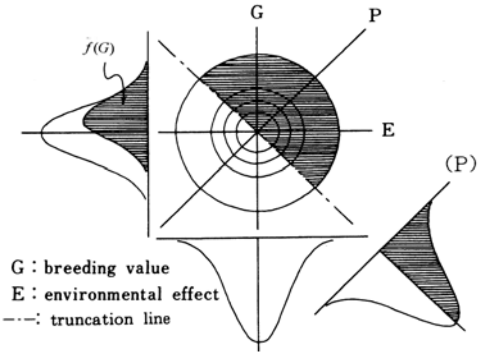

Figure 1. Basic idea to predict response to selection based on phenotypic value

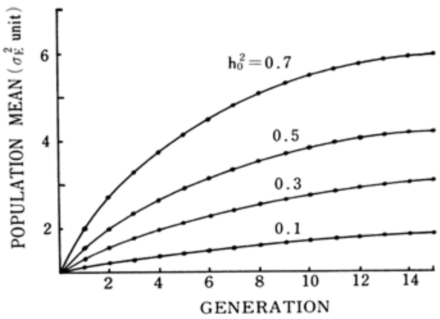

Figure 2. Repose curves to long-term selection based on phenotypic value ($h_0{}^2$: initial heritability)

The predicted changes in the distributions are shown in Fig. 3 and 4.

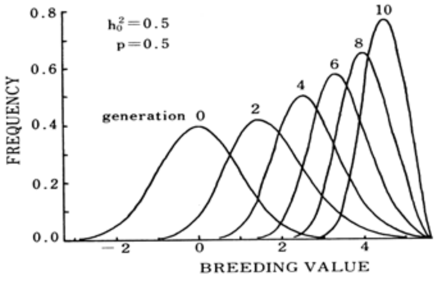

Figure 3. Change in the distribution

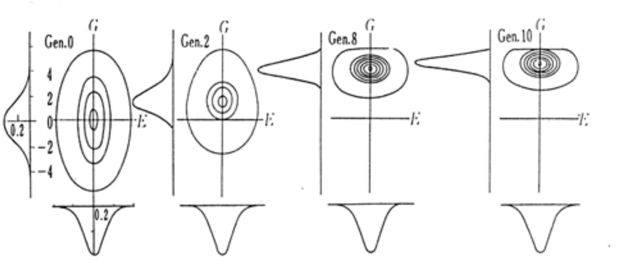

Figure 4. Effects of selection on the joint distribution

of *G* **of *G* and *E***

Using the procedure mentioned above, it is shown (Nishida and Abe, 1980b) that asymmetrical response to divergent selection can be attributed to the skeweness of distribution of *G* in the base population (Fig. 5).

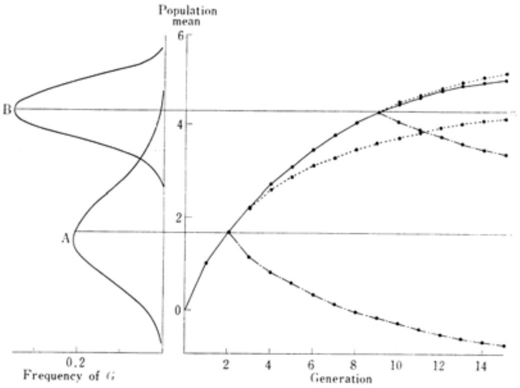

Figure 5. Asymmetrical response to divergent selection caused by the skewness of distribution of *G* in the base population (dotted line: mirror image of the response to downward selection)

The effect of repeated selection on genetic correlation

Hereafter, the phenotype, breeding value and environmental effect of the i^{th} trait will be denoted as P_i, G_i and E_i, respectively. Simple augmentation of the two-dimensional coordinates system mentioned above into the three-dimensional system with the axes of G_1, E_1 and G_2 (Fig. 6) enables us to predict the effect of repeated selection for P_1 on the genetic correlation between G_1 and G_2 using the next formulae (Nishida *et al.*, 1983).

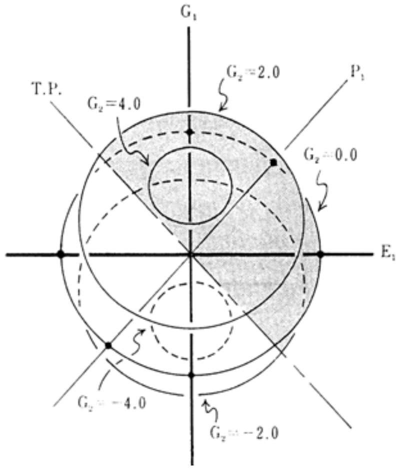

Figure 6. Prediction of the effect of selection for P_1 on the joint distribution of G_1 and G_2 (G_2 axis meets at right angles with G_1,E_1 plane at the origin)

$$f_{12}(G_1,G_2)=\int_s F(G_1,G_2,E_1)dE_1 \quad , \quad f_1(G_1)=\int_s f_{12}(G_1,G_2)dG_2 \quad , \quad f_2(G_2)=\int_s f_{12}(G_1,G_2)dG_1 \quad ,$$

$$r_{G_1G_2}=\iint_s (G_1-\overline{G}_1)(G_2-\overline{G}_2)f_{12}(G_1,G_2)dG_1dG_2 \Big/ \sqrt{\int_s (G_1-\overline{G}_1)^2 f_1(G_1)dG_1\cdot\int_s (G_2-\overline{G}_2)^2 f_2(G_2)dG_2}$$

Some of the results of the prediction were given in Figure 7, 8 and 9. The initial genetic correlation 0.5 decreased toward zero with generation of selection. The negative initial genetic correlation also approached zero with generation of selection.

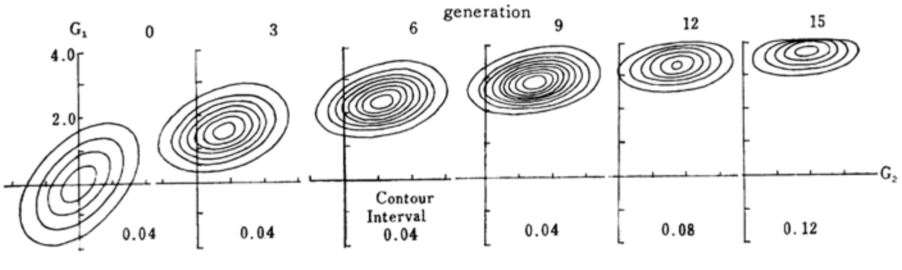

Figure 7. Change in contour line of joint distribution of G_1 and G_2 caused by the selection based on P_1 (initial heritabilities: 0.5, initial genetic correlation: 0.5)

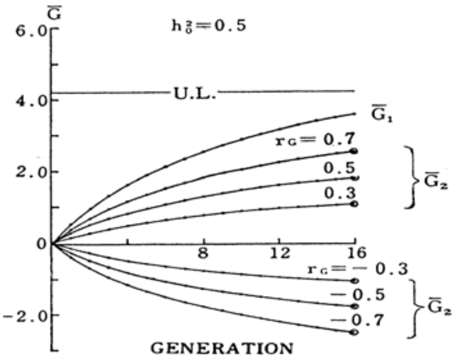

Figure 8. Direct response of G_1 and the correlated response of G_2 to the selection based on P_1 (initial heritability : 0.5.)

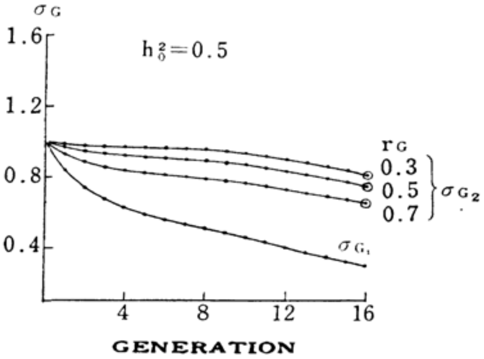

Figure 9. Direct and indirect effect of selection for P_1 on σ_{G_1} and σ_{G_2}

The genetic regression of correlated trait G_2 on the selected trait G_1 was not affected by the selection for P_1 since the truncation plane was parallel with the G_2 axis while the genetic correlation approached zero with generation of selection. But the genetic regression of G_1 on G_2 increased its curvilinearity with generation of selection as illustrated in Figure 10.

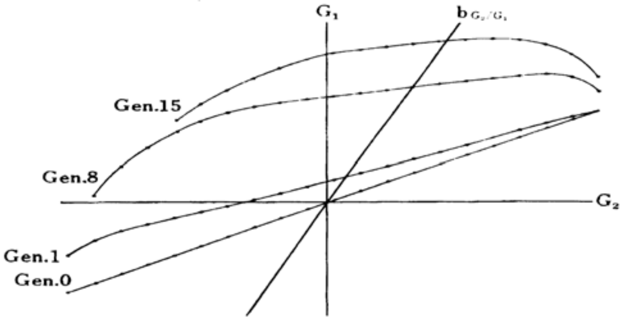

Figure 10. Change in genetic regression of G_1 on G_2 with generation of selection for P_1

The effect of repeated index selection on the joint distribution of G

To improve G_1 and G_2 simultaneously using the selection index ; $I = b_1 \cdot P_1 + b_2 \cdot P_2$, the marginal distribution of G_1 and G_2 ; $f_{12}(G_1, G_2)$, for the selected part is obtained as follows (Nishida *et al.*, 1984).

$$f_{12}(G_1,G_2) = \iint_s F(G_1,G_2,E_1,E_2)dE_1dE_2 \ , \quad f_1(G_1) = f_{12}(G_1,G_2)dG_2 \ , \quad f_2(G_2) = f_{12}(G_1,G_2)dG_2$$

where f_{12} is the joint distribution of G_1 and G_2 for the selected part and s limits the range of integral to the selected part. As illustrated in Figure 11, the genetic correlation between G_1 and G_2 decreased with generation of selection.

GNERATION 0 6 12

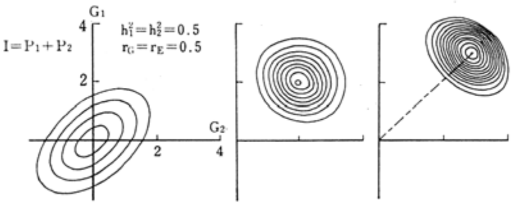

Figure 11-1. Effect of repeated index selection on distribution of G_1 and G_2

GENERATION 0 6 12

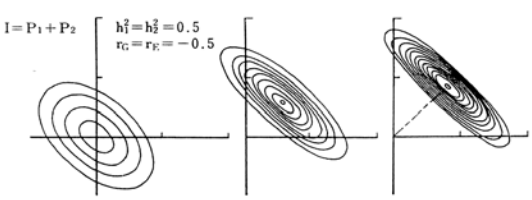

Figure 11-2.

GENERATION 0 6 12

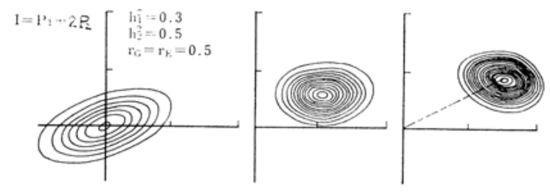

Figure 11-3.

GENERATION 0 6 12

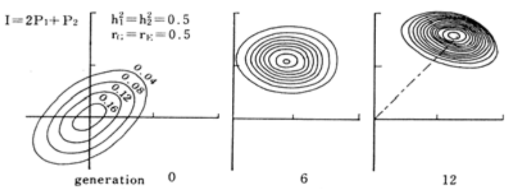

Figure 11-4.

This time the truncation plane was not parallel with G_1 and G_2 axes. Then the genetic regression of G_1 on G_2 and of G_2 on G_1 both changed with generation of selection (Figure12).

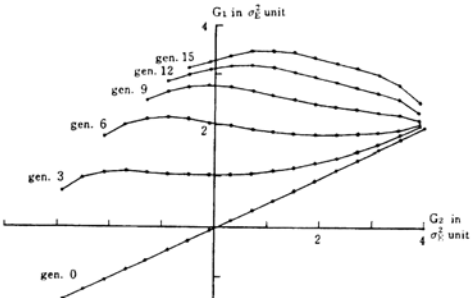

Figure 12. Change in the genetic regression with generation (I=P_1+P_2,h^2_1=0.5, h^2_2=0.5,r_G=r_E=0.5)

Estimating genetic distribution of gametes using β-distribution

It was assumed for the simplicity that the distribution of breeding values of selected animals does not change in the reproduction process. But reproduction partially restores the decreased genetic variance and weaken the skewness of distributions caused by selection. Thus the effects of selection given above were somewhat overestimated.

Takeda *et al.* (1988) have shown that response to repeated selection can be predicted by “gene model” assuming number of loci concerned, number of alleles in each locus, effect of genes and the gene frequencies. In the “gene model”, the gene flow is followed along selection, segregation of genes in gamete genesis and reproduction of the next generation by mating. The genetic distribution of gametes can be accurately predicted by the “gene model” based on the assumptions about genetic characteristics mentioned above. However, in many practical cases, it is difficult to assume these genetic characteristics based on reality.

Thus “continuous distribution model” is necessary for the prediction of response to repeated selection based only on the distribution of phenotypic and breeding values in base population. But predicting genetic distribution of gametes is difficult in the “continuous distribution model”. Then a procedure was developed for predicting the distribution of genetic values of gametes using the distribution of the breeding values of selected animals and β-distribution function in the “continuous distribution model”.

β-distribution function. β-distribution has only two parameters and they are easily estimable. Then it was investigated whether β-distribution fits well to the skewed distribution of genetic values of gametes. The data following skewed distributions were generated by the numerical selection experiments in the “gene model”.

The β-distribution function is written as

$$f_\beta(G) = G^{(p-1)}(1-G)^{(q-1)} / B(p,q) \text{ where } B(p,q) = \int_0^1 u^{(p-1)}(1-u)^{(q-1)}du. \cdot$$

The variable G represent the additive genetic values of gametes divided by their maximum values, $0 \le G \le 1$ and $p,q \ge 0$. Assuming that skewed distribution of G follows β-distribution, the parameters, p and q, are easily estimable with the mean, $\overline{G}$, and the variance, $\sigma_G{}^2$, as

$$p = \overline{G}(\overline{G}(1-\overline{G})/\sigma_G{}^2 - 1) \text{ and } q = (1-\overline{G})(\overline{G}(1-\overline{G})/\sigma_G{}^2 - 1).$$

In early generations of selection for the trait with low or medium initial heritability, highly significant goodness of fit of β-distribution to the skewed distribution was confirmed.

Gamete genesis. Konari *et al.* (1998) has fitted curves of empirical formulae to the close relationships between the additive genetic variance of selected animals and the genetic variance of the gametes, which were produced by the selected animals in many different cases of their numerical selection experiment with the “gene model”. Figure 13 is an example of the fitted curve of the empirical formula.

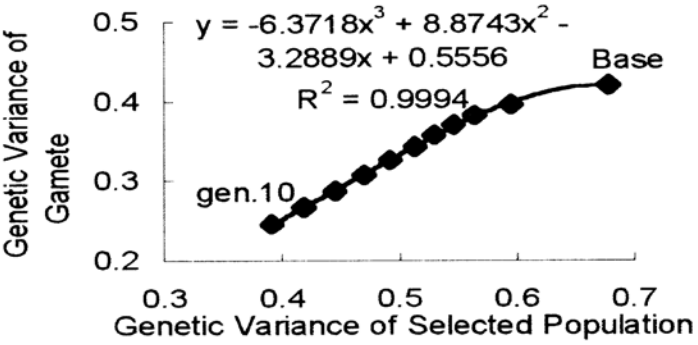

Figure 13. Close relationship between genetic variance of gamete and selected animals

Using the distribution of breeding values of selected animals, the curve and β-distribution function, the genetic distribution of gametes was predicted. The distribution of gametes' genetic values predicted by the fitted β-distribution was compared with the realized distribution in the numerical selection experiment with the "gene model". The goodness of fit of β-distributions was tested with chi–square distribution.

Quite high goodness of fit of β-distribution to the weakly skewed distributions of gametes obtained from numerical repeated selection experiment by the "gene model". Figure 14 illustrates one of the well-fitted β-distribution functions to the skewed distribution of gamete's genetic values.

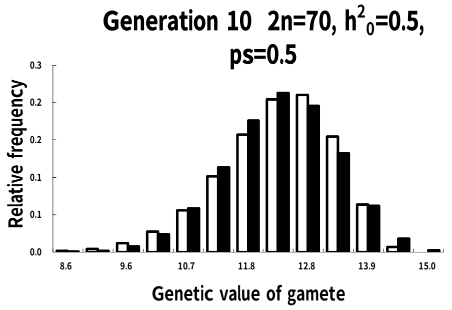

Figure 14. β-distribution (white bar) fitted to the data (black bar)

Using β-distribution, it is possible to predict the effect of reproduction on the genetic properties of selected population and to predict more accurately the response to repeated selection. But the goodness of fit was relatively poor or even fitting was impossible to the distributions with strong skewness. Other distribution functions, which have the higher goodness of fit to the strongly skewed distributions and a small number of easily estimable parameters have to be searched in future.

The characteristics of the continuous distribution model compare with the current theories

Current theories to predict response to repeated selection are based on normal distribution of data with constant variance and based on linear relationships between traits and between breeding values and phenotypic values. However, it has been shown that genetic variance curvilinearly decreases with generation of selection, and skewness of distribution caused by truncation selection brings curvilinear relationships between variables.

Further, it is becoming clear recently that the region in genome called intron plays an important role to control the expression of exon. The clarifying role of intron suggests that the result of QTL analysis should be taken not only as a point of peak in LOD score or F value which possibly means the position of exon but regions which significantly affects the quantitative trait through interacting exon and intron. Based on these facts, a continuous distribution model instead of discrete gene model, for the prediction of response to long-term single trait selection, was developed by improving the previous model.

In current methods to estimate genetic and phenotypic parameters and to predict long-term selection response as VCE, REML, Gibbs sampling and random regression do not adequately cope with skewed distribution, change in genetic variance and curvilinear relationships between variables. On the other hand, the continuous distribution model can treat with these problems correctly.

REFERENCES

Konari, K., Nishida, A. and Yamagishi, T. (1999) *J. Anim. Breed. Genet.* **116**:217-224;
Nishida, A. and Abe, T. (1974) *Can. J. Genet. Cytol.* 16:3-10;
Nishida, A. and Abe, T. (1980a) *Jpn. J. Zootech. Sci.* **51**:485-494;
Nishida, A. and Abe, T. (1980b) *Jpn. J. Zootech. Sci.* **51**:495-500;
Nishida, A. and Hayashi, T. (1980) *Jpn. J. Zootech. Sci.* **51**:745-7474;
Nishida, A., Hayashi, T. and Nagamine, Y. (1983) *Jpn. J. Zootech. Sci.* **54**:239-244;
Nishida, A., Hayashi, T. and Nagamine, Y. (1984) *Jpn. J. Zootech. Sci.* **55**:350-355;
Robertson, A. (1977) *Z. Tierzüchtg. Züchtgsbiol.*94:131:135;
Takeda, H., Nishida, A. and Takebe, A. (1988) *Jap. J. Zootech. Sci.* **59**: 554〜559;