

Optimal selection from families

D. LINDGREN^{1,*}, R. P. WEI¹ & L. BONDESSON²

¹Department of Forest Genetics and Plant Physiology and ²Department of Biometry and Forest Management, Swedish University of Agricultural Sciences, S — 901 83 UMEÅ, Sweden

A method of finding the optimal selected proportions within large individual families is derived. The method identifies family contributions which maximize genetic gain at a given diversity and selected proportion (or rather suggests an optimum combination of these three entities). The population considered is a number of large unrelated families with normal within-family variation. The optimally selected proportion of members from a family is dependent on the average breeding value of the family, the average selected proportion, the diversity, the heritability and the intraclass correlation for the family type. A numerical example is given.

Keywords: effective population number, genetic diversity, response to selection.

Introduction

Methods to maximize gain by selection at a given selected proportion are well known. However, the goal of breeding can be described as maximizing genetic gain while considering diversity. Diversity can be described as effective number (N_e). For deploying clones for clonal forestry or seed orchards, an analytical solution has been found which in an optimal way combines a high genetic gain with a high effective number (Lindgren *et al.*, 1989). For selection of individuals in a population with genetic relationships, no analytical solution has yet been found. Efforts have been made to find the contributions of parents to the next generation which will maximize effective number (e.g. Toro & Nieto, 1984). In optimization, however, it is insufficient to consider parental influence. Lindgren (1986) attempted considering within-family selection using a simplified estimate of the within-family selection gain, but the importance of this simplification is unknown.

The aim of this study is to maximize genetic gain by selecting from a population with a family structure while keeping the effective family number and selected proportion constant.

Model, assumptions and analytical derivations

To avoid complex arguments, we have sometimes chosen to make the prerequisites more constraining

than necessary. A population with a family structure is considered. The population comprises a number (N) of unrelated families of equal size. The average breeding value of members of family j is known to be x_j . The reference point is chosen so that $\sum x_j = 0$. Within-family deviations are distributed with zero mean and variance σ^2 .

From the i th family the proportion w_j is selected. The total selected proportion $W = \sum w_j / N$ of the initial population. We set $p_j = w_j / (NW)$.

The average performance of the part selected from family j is $i(w_j)\sigma$. The function i is a standardized average, referred to as selection intensity. The genetic gain within a family is $i(w_j)h^2\sigma$, where h^2 is the heritability of within-family deviations. The breeding value of the selections from family j will be $x_j + C*i(w_j)$, where $C = h^2\sigma$. The average breeding value (genetic gain, G) of the selected population is obtained by summing the family contributions,

$$G = \sum p_j x_j + C * \sum p_j * i(w_j) \\ = (\sum w_j x_j + C * \sum w_j * i(w_j)) / (NW). \quad (1)$$

A high value of G is desirable.

Genetic diversity can be measured by effective family number, which relates to the genetic structure of the parent population. We define effective family number, N_e , by

$$N_e = (\sum p_j)^2 / \sum p_j^2 = (\sum w_j)^2 / \sum w_j^2 = 1 / \sum p_j^2. \quad (2)$$

This expression has been used by Robertson (1961), Toro & Nieto (1984), Kang & Namkoong (1988), and others as a measure of the effective population number

*Correspondence.

for populations consisting of full-sibs. As it may conflict with other definitions of effective population number, we prefer to define it as effective family number, and regard it as a measure of diversity.

Selection has two important consequences, which can be expressed in two values relating the selected population to the initial: (i) genetic gain; and (ii) effective family number.

The problem is to maximize G under the constraint of constant N_e and W .

It is evident that truncation selection is the most efficient way of applying within family selection in this context. We would like to find the set of w_j 's maximizing G for given N_e and W .

We assume that the distribution within families is normal and that families are large enough to be considered as infinite. These assumptions make it possible to make simplifications. The interval for the w_j 's will be open ($0 < w_j < 1$). Even in the best families black sheep may appear, and upstarts may appear in bad families. Thus, the optimum w_j will never be exactly zero or one.

We proceed in analogy with Lindgren *et al.* (1989).

To maximize G under the constraints

$$\sum w_j = WN \tag{3}$$

$$\sum w_j^2 = N^2 W^2 / N_e \tag{4}$$

$$0 < w_j < 1 \text{ for each } j, \tag{5}$$

we use the Lagrange multiplier technique (see e.g. Luenberger, 1984). Let

$$L = G - \lambda_1(\sum w_j - WN) - \lambda_2(\sum w_j^2 - N^2 W^2 / N_e), \tag{6}$$

where λ_1 and λ_2 are the multipliers. If (w_1, \dots, w_n) is a local maximum point for G (and the w_j 's are not all equal), then

$$0 = \partial L / \partial w_j = \partial G / \partial w_j - \lambda_1 - 2\lambda_2 w_j = 0. \tag{7}$$

Using (1) for G

$$NW \partial G / \partial w_j = x_j + C * \partial [w_j * i(w_j)] / \partial w_j. \tag{8}$$

The probability density of the standardized normal distribution is $\phi(x)$, and its integral above t is $Q(t)$. The function $i(w)$ is then of the form $\phi[t(w)]/w$. The expression $t(w) = Q^{-1}(w)$ is the truncation point, t , above which the standardized normal distribution has mass w .

The latter part of eqn (8) can be simplified, as it can be verified that $\partial [w * i(w)] / \partial w = t(w)$.

Condition (7) becomes

$$x_j + C * t(w_j) - \lambda_1 - 2\lambda_2 w_j = 0 \tag{9}$$

or equivalently,

$$w_j = Q((\lambda_1 + 2\lambda_2 w_j - x_j) / C). \tag{10}$$

It is obvious from eqn (10) that only w_j 's compatible with condition (5) can appear as solutions. There will be no problems at the border of the domain, as w_j cannot attain the values 0 or 1 for a maximum. (However the global maximum of G will occur when $w_j = 1$ for the bottom families). If $\lambda_2 > 0$, there cannot be multiple solutions. Since the global maximum point is in the interior of the domain at least for a number of cases we have studied, the solution will maximize G .

Numerical methods

There is no known analytical way to get an explicit expression of w_j from eqn (10), but numerical solutions can be found by iterative methods.

If λ_1 and λ_2 are known, w_j 's can be calculated for each family by solving (10). The problem is formulated as finding the value where a function, f , becomes zero,

$$f(w_j) = w_j - Q((\lambda_1 + 2\lambda_2 w_j - x_j) / C) = 0. \tag{11}$$

To solve this problem, we have used quadratic interpolation. For each iteration the function $f(w_j)$ was approximated by a parabola passing three points. If starting values were chosen unwisely, the iterative process might not converge or might converge towards a false value. Some combinations of λ_1 and λ_2 are incompatible.

An example is shown in Table 1 as a demonstration. We assumed full-sib families and total phenotypic variance = 1. Individual heritability was set to 0.25. For full-sibs this corresponds to $h^2 = 0.1428$, $\sigma = 0.9354$

Table 1 Optimally selected proportions for 10 full sib families as a function of their average breeding values (x_j). The entries λ_1 and λ_2 are chosen to give the tabulated values of N_e and W

	λ_1	0.4093	0.0542	-0.4200
	λ_2	0.1178	0.9858	3.0050
x_j				
0.6205	0.6611	0.3191	0.1924	
0.3693	0.2369	0.2137	0.1540	
0.2395	0.0792	0.1611	0.1343	
0.1366	0.0190	0.1211	0.1189	
0.0445	0.0031	0.0872	0.1052	
-0.0445	0.0003	0.0570	0.0920	
-0.1366	0.0000	0.0303	0.0786	
-0.2395	0.0000	0.0096	0.0639	
-0.3693	0.0000	0.0007	0.0459	
-0.6205	0.0000	0.0000	0.0149	
N_e	2	5	8	
W	0.1000	0.1000	0.1000	
G	0.6371	0.5224	0.3973	

and $C=0.1336$. We used normal order statistics to get 10 family means from a normal population with zero mean and the intraclass correlation as variance. The proportions presented in Table 1 are optimal in the sense that given the particular W and N_e , there are no other proportions which produce a higher G .

Discussion

The derived method may be applied to several breeding problems. This paper will not explore these possibilities.

Some of the assumptions made can be relaxed without essential changes in the derived method. It seems possible to use predictions instead of known values for family values, provided interpretations of some parameters involved are modified and some constraints are introduced. Families need not be infinite, just large.

If the families are replaced by clones ($C=0$), no gain by within-family selection is made, and eqn (9) simplifies to $x_j - \lambda_1 - 2\lambda_2 w_j = 0$; from which $w_j = (x_j - \lambda_1) / 2\lambda_2$. Thus the optimum proportions are linearly related to the breeding values. This special case has been dealt with by Lindgren *et al.* (1989). The present, more complex, case is technically simpler to prove because it is evident that the border constraints are fulfilled. The simpler algorithm is probably a rather good approximation of the present, more complex case (Lindgren, 1986) and can probably be used for generating starting values for the iterative procedure to avoid convergence problems.

Here the problem has been formulated as maximizing gain at preset given diversity and selected proportion. The solution is also the solution to the problem of maximizing diversity at preset gain and selected proportion or the problem of maximizing selected proportion to get preset gain and diversity. Thus, it is justified to call the solution optimum.

The method gives a range of solutions rather than a single solution. A single unique optimum solution can be found when the benefits of gain and effective family number are expressed on the same scale. Thus, if the scaling factor K is known, there is a single optimum solution maximizing $G + KN_e$.

Acknowledgements

Grants supporting this work have been received from The Swedish Council for Forestry and Agricultural Research and '1959 års fond'. We are indebted to Professor W. G. Hill for discussions and valuable comments on the manuscript. Nigel Rollison has given linguistic advice on the manuscript.

References

- KANG, H. AND NAMKOONG, G. 1988. Inbreeding effective population size under some artificial selection schemes. I. Linear distribution of breeding values. *Theor. Appl. Genet.*, **75**, 333–339.
- LINDGREN, D. 1986. How should breeders respond to breeding values? *Proceedings from IUFRO working parties on breeding theory, progeny testing and seed orchards*, 13–17 October 1986, Williamsburgh, Virginia, pp. 361–371.
- LINDGREN, D., LIBBY, W. J. AND BONDESSON, F. L. 1989. Deployment to plantations and proportions of clones with special emphasis on maximizing gain at a constant diversity. *Theor. Appl. Genet.*, **77**, 825–831.
- LUENBERGER, D. G. 1984. *Linear and Nonlinear Programming* 2nd edn. Addison-Wesley, Reading, Massachusetts, p. 372.
- ROBERTSSON, A. 1961. Inbreeding in artificial selection programmes. *Genet. Res. Camb.*, **23**, 189–194.
- TORO, M. A. AND NIETO, B. M. 1984. A simple method for increasing the response to artificial selection. *Genet. Res. Camb.*, **44**, 347–349.