Original article

Prediction of annual genetic gain and improvement lag between populations

JM Elsen

*INRA, Station d'Amélioration Génétique des Animaux,
BP 27, 31326 Castanet Tolosan, France*

Summary – An approach for computing the expected genetic gain and the improvement lag between subpopulations, based on matrix algebra, is proposed. This is a generalization of the classical Rendel and Robertson (1950) formula, whose main feature is a comparison of successive generation mean values. A simple example is given.

selection response / genetic gain / gene flow

Résumé – **Prédiction du progrès génétique annuel et du décalage génétique entre sous-populations.** *Une approche du calcul de l'espérance du progrès génétique et du décalage entre sous-populations, basée sur l'algèbre matricielle, est proposée. Il s'agit d'une généralisation de la formule classique de Rendel et Robertson (1950), dont la caractéristique principale est de comparer les valeurs moyennes des générations successives. Un exemple simple est donné.*

réponse à la sélection / gain génétique / flux de gènes

INTRODUCTION

The formula of Rendel and Robertson (1950) for estimating the annual genetic gain is well suited to closed homogeneous populations. It may be used directly when there is only one type of breeding animal per sex. In other cases, such as progeny test designs where known and tested males are both reproducing, the formula has to be adapted (Lindhé, 1968). Bichard (1971) showed how to process a hierarchical population and how to estimate the improvement lag between subpopulations. These methods are based on comparisons between the mean additive genetic values of successive generations. More recently, iterative methods (Hill, 1974; Elsen and Mocquot, 1974; Elsen 1980; Ducrocq and Quaas, 1988) have been developed in

order to take account of the year by year change of genetic values. They are well fitted for the description of hierarchical populations.

In the present paper, we propose a new method for estimating the genetic progress and improvement lag between subpopulations, which can be applied to these heterogeneous populations. Like the method of Rendel and Robertson (1950), our procedure is based on a comparison of successive generations and is, thus, of a simpler formulation than iterative methods.

METHODS

Description of the population

We consider only stable populations where the selection policy (selection pressure, organisation of matings) and structure are constant.

The population is subdivided into groups of breeding animals. Let X_i be the mean genetic value of the ith group. The model gives the value of the vector $\mathbf{X}$ of X_i given the mean values at the previous generation, Y_i.

The groups are defined in the following way. Two individuals belong to the same group, i:
 − if they are of the same sex;
 − if the probabilities that their respective parents of the same sex belong to the same group $j(p_{ij})$ are equal;
 − if they have equal probabilities of being parents of individuals belonging to the same group of the next generation.

Generations are defined here in a relative way: let us consider the population at a given time. All individuals belonging to group i at this moment constitute a generation of this group. By definition their parents which belong to group j are from the previous (parental) generation of group j relative to i. With this definition, this parental generation of group j does not comprise the same individuals if considering their offspring from group i or from group i'. Its mean genetic value will be noted Y_{ji} for group i.

Computations of the mean genetic values X_i and Y_{ji} are made by considering the individuals, at birth, prior to any selection.

The principle of the method is to write relationships between groups of one generation and groups of the previous one. For group i, we have:

$$X_i = \sum_{j \in \mathcal{M}} \frac{1}{2} \, p_{ij}(Y_{ji} + \Delta_{ji}) + \sum_{j \in \mathcal{F}} \frac{1}{2} \, p_{ij}(Y_{ji} + \Delta_{ji}) \qquad [1]$$

where $\mathcal{M}$ and $\mathcal{F}$ are male and female breeding animals respectively, and where Δ_{ji} is the deviation between the mean value of group j at birth, Y_{ji}, and the value of those individuals from this group which will actually be parents of group i.

On the other hand, due to the genetic progress (ΔG per year), we have:

$$Y_{ji} = X_j - L_{ji}\Delta G \qquad [2]$$

where L_{ji} is the generation interval between group j and group i.

Solution

Pooling relations [1] and [2], we get

$$X_i = \sum_j a_{ij}(X_j - L_{ji}\Delta G + \Delta_{ji}) \qquad [3]$$

with $a_{ij} = \frac{1}{2}p_{ij}$.

Or, in matrix notation,

$$\mathbf{X} = \mathbf{AX} + \mathbf{H} \qquad [4]$$

where $\mathbf{A}$ is the matrix of a_{ij} and $\mathbf{H}$ the vector of $\sum_j a_{ij}(\Delta_{ji} - L_{ji}\Delta G)$, which we

shall write $\mathbf{H} = \mathbf{\Delta} - \mathbf{L}\Delta G$, $\mathbf{\Delta}$ and $\mathbf{L}$ being the vectors of $\sum_j a_{ij}\Delta_{ji}$ and $\sum_j a_{ij}L_{ji}$,

respectively.

Case of a single population

In this case matrix $\mathbf{A}$ is stochastic. Indeed, the events "sire (or dam) of an individual of group i belongs to group j", defined over the different groups j of the population, form a complete system of events, and

$$\sum_{j\in\mathcal{M}} p_{ij} = 1 \text{ and } \sum_{j\in\mathcal{F}} p_{ij} = 1, \text{thus } \sum_j a_{ij} = 1 \text{ with } 0 \leqslant a_{ij} \leqslant \frac{1}{2}.$$

The largest eigenvalue of $\mathbf{A}$ is 1. Let $\mathbf{V}$ (transpose $\mathbf{V}^T$) be the eigenvector corresponding to 1. This vector $\mathbf{V}$ may easily be found by substitution since $\mathbf{V}^T\mathbf{A} = \mathbf{V}^T$ (note that to simplify this substitution one of the elements of $\mathbf{V}$ may be fixed to 1).

Knowing $\mathbf{V}$, the annual genetic gain is easily deduced, using:

$$\mathbf{V}^T\mathbf{X} = \mathbf{V}^T\mathbf{AX} + \mathbf{V}^T\mathbf{H}$$

Thus, $\mathbf{V}^T\mathbf{H} = \mathbf{0}$

or

$$\Delta G = \frac{\sum_i V_i(\sum_j a_{ij}\,\Delta_{ji})}{\sum_i V_i(\sum_j a_{ij}\,L_{ji})} = \frac{\mathbf{V}^T\mathbf{\Delta}}{\mathbf{V}^T\mathbf{L}}$$

Case of a composite population

The general equation is still of type [4], but here we have:

$$\mathbf{X} = \begin{pmatrix} {}_1\mathbf{X} \\ {}_2\mathbf{X} \\ \vdots \\ {}_n\mathbf{X} \end{pmatrix} = \begin{pmatrix} {}_{11}\mathbf{A} & {}_{12}\mathbf{A} & \cdots & {}_{1n}\mathbf{A} \\ {}_{21}\mathbf{A} & {}_{22}\mathbf{A} & \cdots & {}_{2n}\mathbf{A} \\ \vdots & \vdots & \ddots & \vdots \\ {}_{n1}\mathbf{A} & {}_{n2}\mathbf{A} & \cdots & {}_{nn}\mathbf{A} \end{pmatrix} \mathbf{X} + \begin{pmatrix} {}_1\mathbf{H} \\ {}_2\mathbf{H} \\ \vdots \\ {}_n\mathbf{H} \end{pmatrix} \qquad [5]$$

where ${}_h\mathbf{X}$, ${}_{hh'}\mathbf{A}$ and ${}_h\mathbf{H}$ are vectors and matrices specific to subpopulations h and h'. In particular, an annual genetic gain ΔG_h, specific to population h, is found for each ${}_h\mathbf{H}$.

As a whole, the matrix $\mathbf{A}$ is still stochastic, but each of its ${}_{hh}\mathbf{A}$ elements is not necessarily of this type.

Thus, we have:

$$\mathbf{A} = \mathbf{U}\mathbf{V}^T(\mathbf{V}^T\mathbf{U})^{-1} + \mathbf{T} = \mathbf{R} + \mathbf{T}$$

where $\mathbf{U}$ is the column eigenvector (made of 1's) and $\mathbf{V}^T$ the row eigenvector corresponding to the eigenvalue 1. The $\mathbf{T}$ matrix is such that $\mathbf{RT} = \mathbf{TR} = \mathbf{0}$.

Thus, $\mathbf{X} = \mathbf{RX} + \mathbf{TX} + \mathbf{H}$, or $(\mathbf{I} - \mathbf{T})\mathbf{X} = \mathbf{RX} + \mathbf{H}$.

Using $(\mathbf{I} - \mathbf{T})^{-1} = \mathbf{I} + \mathbf{T} + \mathbf{T}^2 + \ldots$, we get $\mathbf{X} = \mathbf{RX} + (\mathbf{I} - \mathbf{T})^{-1}\mathbf{H}$

The lag between 2 groups k and k' is given by $\mathbf{g}^T\mathbf{X}$, where $\mathbf{g}^T$ is a row vector with all its elements zero, except the elements corresponding to the groups $k(g_k = 1)$ and $k'(g_{k'} = -1)$. The lag between 2 subpopulations, which could be defined by the difference in mean values of productive animals (milking cows, slaughtered lambs...), will most often be given by the lag between 2 groups belonging respectively to these 2 subpopulations and defined on an equivalent basis. Nevertheless, one can imagine that in some instances the level of a subpopulation may be characterized by a weighted sum $\sum_k g_k X_k$.

Thus, as $\mathbf{RX}$ is a vector all of whose elements are equal, $\mathbf{g}^T\mathbf{RX} = 0$ and the lag E is:

$$E = \mathbf{g}^T\mathbf{X} = \mathbf{g}^T(\mathbf{I} - \mathbf{T})^{-1}\mathbf{H} = \mathbf{g}^T(\mathbf{I} - \mathbf{A} + \mathbf{U}\mathbf{V}^T(\mathbf{V}^T\mathbf{U})^{-1})^{-1}\mathbf{H}$$

or

$$E = \mathbf{g}^T\mathbf{M}^{-1}\mathbf{H} = \mathbf{w}^T\mathbf{H} = \mathbf{w}^T\Delta - \mathbf{w}^T\mathbf{L}\Delta G$$

EXAMPLE

Model

Let a population comprise a nucleus and a base. In the nucleus, as in the base, the selection is on maternal performance only. Good females (selection pressures $\Delta_{F_1 F_1}$ in the nucleus, $\Delta_{F_2 F_2}$ in the base) are dams of young females and natural service males. A fraction d_1 of the nucleus female replacement is made through artificial

insemination. The AI sires are sons of elite dams (selection pressure $\Delta_{F_1 A_1}$) and AI sires.

Among the sires used in the base, a fraction d_2 was born in the nucleus, given the diffusion of genetic gain. These males are chosen from among those born from artificial insemination, with a selection pressure of $\Delta_{F_1 A_2}$ on maternal value.

The mean values of the breeding animals will be denoted:

- for the nucleus:

$$_1\mathbf{X} = \begin{pmatrix} X_{F_1} \\ X_{A_1} \\ X_{B_1} \end{pmatrix} \quad \begin{array}{l} \text{for females} \\ \text{for AI males} \\ \text{for natural service males} \end{array}$$

- for the base:

$$_2\mathbf{X} = \begin{pmatrix} X_{F_2} \\ X_{A_2} \\ X_{B_2} \end{pmatrix} \quad \begin{array}{l} \text{for females} \\ \text{for males born in the nucleus} \\ \text{for males born in the base} \end{array}$$

Noting that $\Delta_{F_1 F_1} = \Delta_{F_1 B_1}$ and $\Delta_{F_2 F_2} = \Delta_{F_2 B_2}$, equation [4] is

$$\begin{pmatrix} X_{F_1} \\ X_{A_1} \\ X_{B_1} \\ X_{F_2} \\ X_{A_2} \\ X_{B_2} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{d_1}{2} & \frac{1-d_1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{d_1}{2} & \frac{1-d_1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{d_2}{2} & \frac{1-d_2}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{d_2}{2} & \frac{1-d_2}{2} \end{pmatrix} \begin{pmatrix} X_{F_1} \\ X_{A_1} \\ X_{B_1} \\ X_{F_2} \\ X_{A_2} \\ X_{B_2} \end{pmatrix}$$

$$+ \begin{pmatrix} \frac{\Delta_{F_1 F_1}}{2} - (L_{F_1 F_1} + d_1 L_{A_1 F_1} + (1-d_1)L_{B_1 F_1})\frac{\Delta G}{2} \\ \frac{\Delta_{F_1 A_1}}{2} - (L_{F_1 A_1} + L_{A_1 A_1})\frac{\Delta G}{2} \\ \frac{\Delta_{F_1 F_1}}{2} - (L_{F_1 B_1} + d_1 L_{A_1 B_1} + (1-d_1)L_{B_1 B_1})\frac{\Delta G}{2} \\ \frac{\Delta_{F_2 F_2}}{2} - (L_{F_2 F_2} + d_2 L_{A_2 F_2} + (1-d_2)L_{B_2 F_2})\frac{\Delta G}{2} \\ \frac{\Delta_{F_1 A_2}}{2} - (L_{F_1 A_2} + L_{A_1 A_2})\frac{\Delta G}{2} \\ \frac{\Delta_{F_2 F_2}}{2} - (L_{F_2 B_2} + d_2 L_{A_2 B_2} + (1-d_2)L_{B_2 B_2})\frac{\Delta G}{2} \end{pmatrix}$$

The eigenvector $_1\mathbf{V}$ of the submatrix $_{11}\mathbf{A}$ may be written, fixing its first element to 1,

$$_1\mathbf{V}^T = (1, \frac{2d_1}{1+d_1}, \frac{1-d_1}{1+d_1})$$

The annual genetic gain becomes $\Delta G = \dfrac{_1\mathbf{V}_1^T\boldsymbol{\Delta}}{_1\mathbf{V}_1^T\mathbf{L}}$, with

$$_1\mathbf{V}_1^T\boldsymbol{\Delta} = \frac{d_1}{1+d_1}\,\Delta_{F_1A_1} + \frac{1}{1+d_1}\,\Delta_{F_1F_1}$$

$$_1\mathbf{V}_1^T\mathbf{L} = \frac{1}{2}(L_{F_1F_1} + d_1 L_{A_1F_1} + (1-d_1)L_{B_1F_1}) + \frac{d_1}{1+d_1}(L_{F_1A_1} + L_{A_1A_1})$$

$$+\frac{1-d_1}{2(1+d_1)}(L_{F_1B_1} + d_1 L_{A_1B_1} + (1-d_1)L_{B_1B_1})$$

Noting that the eigenvector $\mathbf{V}$ of the matrix $\mathbf{A}$ is $(1\mathbf{V}^T, \mathbf{0})$, we find that

$2(1+d_1)M =$

$$\begin{pmatrix}
2(d_1+1) & d_1(1-d_1) & d_1(d_1-1) & 0 & 0 & 0 \\
0 & 1+3d_1 & 1-d_1 & 0 & 0 & 0 \\
0 & d_1(1-d_1) & d_1^2+d_1+2 & 0 & 0 & 0 \\
1+d_1 & 2d_1 & 1-d_1 & 1+d_1 & -d_2(1+d_1) & (1+d_1)(d_2-1) \\
0 & d_1-1 & 1-d_1 & 0 & 2(1+d_1) & 0 \\
1+d_1 & 2d_1 & 1-d_1 & -(1+d_1) & -d_2(1+d_1) & (1+d_1)(1+d_2)
\end{pmatrix}$$

The lag between the 2 subpopulations is given by $E = \mathbf{g}^T\mathbf{M}^{-1}\mathbf{H}$. Taking $\mathbf{g}^T = (1,0,0,-1,0,0)$, we get:

$$\mathbf{w}^T = \mathbf{g}^T\mathbf{M}^{-1} = \left(\frac{1+d_2}{d_2},\ \frac{d_2+2}{d_2}\,\frac{d_1}{1+d_1},\ \frac{1+d_2}{d_2}\,\frac{1-d_1}{1+d_1},\ -\frac{1+d_2}{d_2},\ -1,\ -\frac{1-d_2}{d_2}\right)$$

E is easily deduced.

Numerical application

We consider the simple situation where all the dam-progeny generation lengths $(L_{F_1F_1}, L_{F_1A_1}, L_{F_1B_1}, L_{F_1A_2}, L_{F_2F_2}, L_{F_2B_2})$ are 5 years and sire-progeny generation lengths $(L_{A_1F_1}, L_{B_1F_1}, L_{A_1A_1}, L_{A_1B_1}, L_{B_1B_1}, L_{A_2F_2}, L_{B_2F_2}, L_{A_1A_2}, L_{A_2B_2}, L_{B_2B_2})$ are 3 years.

It is also assumed that the females are not recorded in the base ($\Delta_{F_2F_2} = \Delta_{F_2B_2} = 0$), and that, in the nucleus, the dam-daughter are the best 50%, the dam–AI sire are the best 10% and the dam–natural mating sire, the next 20%. Given a common accuracy $h = 0.5$ for the dam, the selection differentials, in standard deviation units, are given by:

$$\Delta_{F_1F_1} = i(0.5)h = 0.4$$

$$\Delta_{F_1A_1} = i(0.1)h = 0.878$$

$$\Delta_{F_1B_1} = \frac{0.3i(0.3) - 0.1i(0.1)}{0.2}\,h = 0.425$$

where i is the selection intensity function, assuming the trait normally distributed. With these assumptions, we find

$$\Delta G = \frac{0.878d_1 + 0.4}{8(1 + d_1)}$$

$$\mathbf{H} = \begin{pmatrix} 0.2 & - (5+3)) & \Delta G/2 \\ 0.439 & - (5+3) & \Delta G/2 \\ 0.2 & - (5+3) & \Delta G/2 \\ 0.0 & - (5+3) & \Delta G/2 \\ 0.2125 & - (5+3) & \Delta G/2 \\ 0.0 & - (5+3) & \Delta G/2 \end{pmatrix}$$

The lag $E = \mathbf{w}^T \mathbf{H}$ is then

$$E = 0.4 \, \frac{2 + 2d_1 + 2d_1 d_2 + d_2}{2d_2(1 + d_1)^2} + 0.878 \, \frac{d_1}{d_2(1 + d_1)} - 0.2125$$

The figure 1 shows the behaviour of the genetic gain and the improvement lag with varying fractions d_1 and d_2.
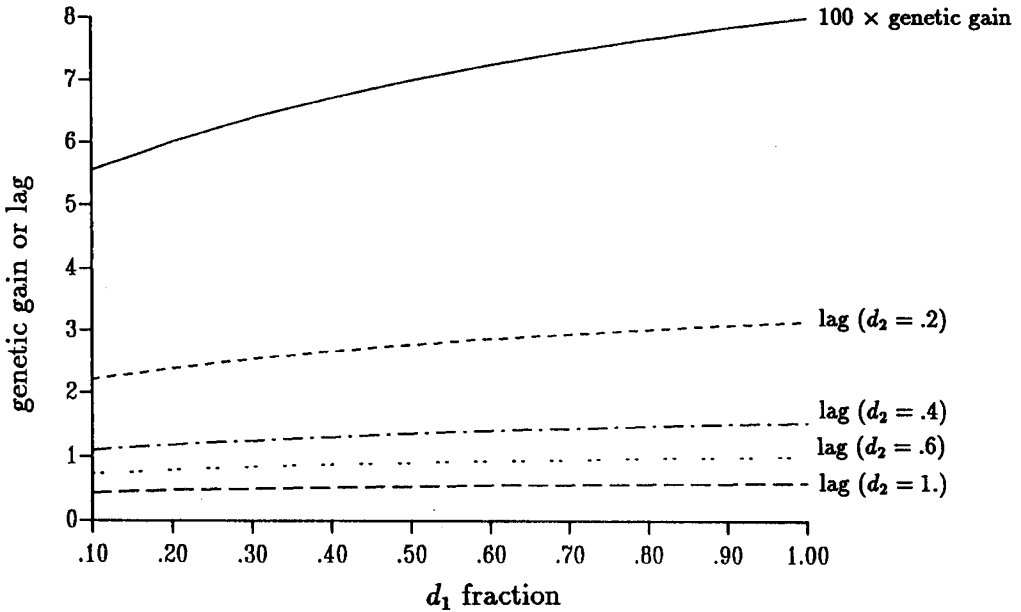

Fig 1. Genetic gain and improvement lag depending on the fractions d_1 and d_2 (in standard deviation unit).

CONCLUSION

The main difficulty of the method is the definition of groups. A particular population may be analysed in different ways. The smaller the number of groups, the more easily the eigenvector V and the inverse matrix M^{-1} will be found, but the more difficult will be the correct writing of matrix A and vectors Δ and L. There are 2 extreme cases: the first is one in which only 2 groups are considered, in keeping with the classical demonstration of the formula of Rendel and Robertson (1950); the second is one in which individuals of the same group have the same age, similar to the model of Hill (1974) and Elsen and Mocquot (1974).

Finally, it should be emphasized that the preceding results are only asymptotic and need constant selection pressure and population structure in the long run.

ACKNOWLEDGMENT

REFERENCES

Bichard M (1971) Dissemination of genetic improvement through a livestock industry. *Anim Prod* 13, 401-411

Ducrocq V, Quaas RL (1988) Prediction of genetic response to truncation selection across generations. *J Dairy Sci* 71, 2543-2553

Elsen JM, Mocquot JC (1974) *Recherches pour une Rationalisation Technique des Schémas de Sélection des Bovins et des Ovins. Bull Tec Dép Génét Anim No 17*, INRA, Paris

Elsen JM (1980) Diffusion du progrès génétique dans les populations avec générations imbriquées: quelques propriétés d'un modèle de prévision. *Ann Génét Sél Anim* 12, 49-80

Hill WG (1974) Prediction and evaluation of response to selection with overlapping generations. *Anim Prod* 18, 117-140

Lindhé B (1968) Model simulation of AI breeding within a dual purpose breed of cattle. *Acta Agric Scand* 18, 33-41

Rendel JM, Robertson A (1950) Estimation of genetic gain in milk yield by selection in a closed herd of dairy cattle. *J Genet* 50, 1-8