A COMPARISON OF ESTIMATORS OF SELECTION RESPONSE IN SIMULATED SHEEP POPULATIONS

P. Berg, J. Lauridsen, J. Jensen

Abstract

Stochastic simulation of alternative sheep breeding schemes are described. A set of programs based on a modular approach in mixed programming languages, simplifying extensions and modificartions of modules to study a wide range of breeding strategies, are presented.

Estimators of response to selection in simulated sheep populations are compared. The estimators are regression of breeding values on year of birth, average response between firsts and last year and a curvilinear regression of breeding values on year of birth. Variation of response to selection can be estimated from a single replicate. Average response had a slightly lower sampling variance than the linear trend. In small populations response to selection might decrease over time.

Key words: Stochastic stimulation; software; regression; sampling variance

Introduction

Stochastic simulation is often chosen to study complex systems, such as animal breeding schemes, where only approximate inference can be made. Stochastic simulation are generally implemented to a specific application. Hence standard software for simulation of animal breeding schemes is generally less available than software to analyse data (i.e. BLUP and REML software). In this paper the strategy used for implementing a flexible set of programs/tools to simulate animal breeding schemes is presented.

In simulated data the true breeding values are known, and can be used to estimate annual response to selection. However several estimators of annual response to selection are available, such as regression of breeding values on year of birth, average response between first and last year simulated. Another possibility is that response to selection under an infinitesimal model (Bulmer

1980) decreases with time, as the genetic variance decreases due to inbreeding. These estimators are compared and their sampling variance over repeated sampling compared to their expected sampling variance.

Material and Methods

2.1. Simulation programs

To facilitate development and maintenance of the programs, an animal breeding sheme was described as a set of separate events and these were implemented in separate modules. The events considered was
- Generate base population
- Generate mating combinations
- Generate offspring breeding values and observations
- Compute individual inbreeding coefficients
- Compute selection criterion (prediction of breeding values)
- Selection of breeders
- Compute summary statistics and store information

Breeding values of offspring are simulated as

$$ai = \tfrac{1}{2}(a_s + a_d) + (1 - \tfrac{1}{4}k)(1 - \tfrac{1}{2}(F_s + F_d))*L\,'r_1 \tag{1}$$

where
a_i, a_s, a_d is a vector of breeding values of animal i and its sire s and dam d (O if unknown)
F_s, F_d is coefficients of inbreeding of the sire and dam (0 if unknown)
L is a cholesky decomposition of the genetic (co) variance matrix, $G = L'L$, where L is an upper triangular matrix.
r_1, r_2 is vectors of random deviates $\sim N(O,I)$
k is the number of known parents (0 in base population, 2 elsewhere)

Individual inbreeding coefficients are computed using the algorithm proposed by Meuwissen & Lu (1992).

Observations are simulated as

$$y_i = \mu + a_i + Z_d y_d + C'r_2 \tag{2}$$

where
y_i, y_d is a vector of observations of animal i and its dam d.
μ is a vector of means

Z_d is a matrix relating pseudo observations on the dam to observations on the offspring, to mimic maternal genetic and permanent environmental effects.

C is a cholesky decomposition of the residual (co) variance matrix, R= C'C, where C is an upper triangular matrix.

All possible observations on an animal is simulated, and observations discarded if the are not realised, due to culling.

Selection is performed within groups or across groups. Mating can either random, assortativ or random but with avoidance of matings between close relatives.

To simplify extensions and the development of the modules all information is stored in a simple database composed of two tables, an animal table (containing parentage information, time of birth, status, breeding values and individual inbreeding coefficients) and an observation table (containing obser-vations and the time for their are realisation). Any module only assumes knowledge about the variables in these tables that is used in the specific module.

The chosen strategy allows for modules being implemented in different programming languages. Currently SAS and Fortran90 is used.

Sheep breeding programs

For a further description of the sheep breeding schemes simulated see Lauridsen et al. (1997). Additionally a simple breeding scheme with one herd and all traits measured was simulated with either 20, 200 or 1200 females. These populations were simulated for 20 years, and response determined from the last 15 years. Traits recorded was littersize, gain 0 to 2 months, gain 2 to 4 months, ultrasonic measurements and carcass grade. The selection criterion, economic merit, was based on littersize, gain 0 to 2 months (direct and maternal), gain 2 to 4 months (direct and maternal and carcass grade.

Estimators of response to selection

It was decided to discard the first 5 generations before computing response to selection, based on the correlation between predicted and true breeding values and generation intervals. These indicated that a "steady state" was attained, such that any effects of the base population was negligible after this time and that information on relatives had accumulated.

The model underlying the estimation of response to selection was

$$a = X_j \beta_{ij} + e_i \qquad (3)$$

where a is a vector of breeding values, Xj is a design matrix relating β_{ij} to a, and e is defined as either $e_1 \sim N(0, I\sigma^2)$ or $e_2 \sim N(0, A\sigma^2)$. σ^2 is the genetic variance and A the numerator relationship matrix.

From each replicate β_{ij} and its variance was estimated as

$$\hat{\beta}_{1j} = (X_j^T X_j)^{-1} X_j^T a \tag{4a}$$

$$\hat{\beta}_{2j} = (X_j^T A^{-1} X_j)^{-1} X_j^T A^{-1} a \tag{4b}$$

and

$$V(\hat{\beta}_{1j}) = (X_j^T X_j)^{-1} X_j^T A X_j (X_j^T X_j)^{-1} \sigma^2 \tag{5a}$$

$$V(\hat{\beta}_{2j}) = (X_j^T A^{-1} X_j)^{-1} \sigma^2 \tag{5b}$$

X_j was defined so that β_{ij} contains

$\beta_{.1}$ An intercept and a linear regression of breeding values on year of birth for animals born after generation 5.

$\beta_{.2}$ The average response from generation 6 to 20. X is a row vector with elements (- 14 * n_{20})/($n_6 + n_{20}$) for animals born in year 6 and (14* n_6)/($n_6 + n_{20}$) for animals born in year 20, where nx is the number of animals born in year x.

$\beta_{.3}$ An intercept and a linear and quadratic regression of breeding values on year of birth.

$\beta_{1.}$ Is a Least Squares (LS) estimator and $\beta_{2.}$ is a Generalised Least Squares (GLS) estimator.

The sampling variance estimated within replicates was compared to the realized sampling variance across replicates.

Results and discussion

The least squares estimators are shown in Table 1, with population sizes of 20, 200 and 1200 ewes.

GLS vs. LS.

As expected GLS (4b) estimates the mean of the distribution of breeding values, and accordingly gives a trend of zero. Conditional on the parents, the new generation doesn't deviate from the parental average. The term $X^T A^{-1} a$ in (4b) is a linear function of mendelian sampling terms, with expectation zero.

458

However computing the sampling variance of the LS estimator (4a) requires A, the numerator relationship matrix. In the simulated population individual inbreeding coefficients are computed and thus X^TAX can be computed as X^TL^TDLX (Henderson 1976), where D is a diagonal matrix with elements that are a simple function of parental inbreeding coefficients and L, a upper triangular matrix , can be computed one row at a time.

Linear trend vs. Average response

The variance of average response (β_2) was slightly lower than the variance of the linear response (β_1). Both gave an unbiased estimate of the sampling variance. Thus the variation in response to selection can be estimated from a single replicate, a desirable situation in computer intensive simulations with large populations, Animal Model evaluations and simulation over many years as described by Lauridsen et al. (1997). This also allows for prediction of the number of replicates needed to obtain a given power in the comparison between alternative breeding schemes.

Curvilinear response to selection

None of the populations simulated in Table 1 showed a significant curvilinear response to selection. However it is seen from Table 1, that the decrease in response over time increases with a decreasing population size. A significant curvilinear response has been estimated in a large population of 1200 ewes divided into 60 isolated herds (Lauridsen et al. 1997). In this case to selection decreased significantly with time from 6,16 kr/lamb/year in year 6 to 3,02 kr/lamb/year in year 20. Response should be presented as response at a given time or as the average response in a specified time interval.

Table 1. - RESPONSE TO SELECTION ESTIMATED AS LINEAR TREND AFTER GENERATION 5, AVERAGE RESPONSE BETWEEN YEARS 6 AND 20 AND CURVILINEAR RESPONSE AFTER GENERATION 5. MEAN OF LS ESTIMATES ($\overline{R}$) , SAMPLING STANDARD DEVIATION OF ESTIMATES (STD (R)) AND THE SQUARE OF THE MEAN VARIANCE (STD) FROM 10 REPLICATES. THE RESPONSE FROM A CURVILINEAR REGRESSION WAS PRESENTED AS THE FIRST DERIVATIVE OF THE REGRESSION EQUATION WITH RESPECT TO TIME.

	20 ewes			200 ewes			1200 ewes		
	R	Std(R)	Std	R	Std(R)	Std	R	Std(R)	Std
Linear	8.25	2.25	2.34	14.27	0.99	1.08	14.59	0.56	0.49
Average	8.10	2.33	2.26	14.00	1.02	1.03	14.58	0.46	0.46
Quadratic	12.10	7.66	8.99	14.72	3.73	3.76	13.24	1.52	1.66
	-.29*t	0.56	0.64	-.03*t	0.26	0.28	+.10*t	0.10	0.12

Conclusion

The variance of response to selection can be accurately estimated within replicates.

The average response to selection has a slightly lower sampling variance than a linear regression. In small populations simulated under an infinitesimal model, response to selection decreases with time.

REFERENCES

1. B u l m e r , M. G. (1980): The mathematical theory of quantitative genetics. Claredon Press, Oxford. Henderson, C.R. 1976. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. Biometrics 32:69:83.

2. L u r i d s e n , J., P. B e r g , J. J e n s e n (1997): Comparision of alternative sheep breeding schemes using stochastic simulation. 48 th Annual Meeting of EAAP, Session GCS2.4 (thid meeting).

3. M e u w i s s e n , T. H. E., Z. Lu (1992): Computing inbreeeding coefficients in large populations. Genet. Sel.Evol.24:305-313.

USPOREDBA PROCJENITELJA REAGIRANJA NA SELEKCIJU U SIMULIRANIM POPULACIJAMA OVACA

Sažetak

Opisana je pretpostavljena stohastična simulacija alternativnih programa uzgoja ovaca. Prikazan je niz programa što se temelje na modularnom pristupu u raznim jezicima programiranja, pojednostavljenim proširenjima i modifikacijama modula radi proučavanja čitavog niza uzgojnih strategija. Uspoređuju se procjenitelji reagiranja na selekciju u simuliranim populacijama ovaca. Procjenitelji su regresija uzgojnih vrijednosti na godinu rođenja, prosječno reagiranje između prve i zadnje godine, te linearna regresija uzgojnih vrijednosti na godinu rođenja. Varijacija reagiranja na selekciju može se procijeniti iz jednog ponavljanja. Prosječno reagiranje na selekciju može se s vremenom smanjiti.

Ključne riječi: Pretpostavljena stohastična simulacija, software, regresija, varijanca uzorka.