Theor. Appl. Genet. 51, 35-39 (1977)

Measuring the Precision of Genetic Parameters by a Simulation Technique

D.D. Rodda, L.R. Schaeffer, K. Mullen and G.W. Friars
University of Guelph, Guelph, Ontario (Canada)

Summary. Approximate standard errors of genetic parameter estimates were obtained using a simulation technique and approximation formulae for a simple statistical model. The similarity of the corresponding estimates of standard errors from the two methods indicated that the simulation technique may be useful for estimating the precision of genetic parameter estimates for complex models or unbalanced population structures where approximation formulae do not apply. The method of generating simulation populations in the computer is outlined, and a technique of setting approximate confidence limits to heritability estimates is described.

Key words: Parameters - Genetics - Precision - Simulation

Introduction

Several attempts have been made to provide approximate formulae for the standard errors of heritability and genetic correlation estimates derived from variance and covariance components (e.g. Dickerson 1960; Robertson 1959; Grossman and Norton 1974) for a few limited hierarchical statistical models, sometimes with very restrictive assumptions. However, there are no general formulae that are applicable to the many possible models or combinations of circumstances.

A simulation technique was used by Van Vleck and Henderson (1961) and Shook and Barr (1968) as an empirical check on the approximation formulae of Reeve (1955) and Tallis (1959) for the sampling variance of genetic correlation estimates. In this study, simulation was carried out to check the approximate standard errors from formulae of genetic parameter estimates of egg composition traits in a study by Roda, Friars, Gavora and Merritt (1977) from 998 progeny distributed over 229 dams and 56 sires. The experiment was designed to obtain 5 progeny from each of the 4 sire-dam families, but in all cases a minimum of 17 progeny per sire was maintained by the addition of a fifth family, if necessary. Thus, the data were very close to being equal number subclasses. A computer was used to generate many simulated populations of observations from the parameters estimated from the original experimental observations. Subsequently, parameter estimates were calculated from the simulated populations, their distribution was examined, and standard errors calculated.

Experimental Method

Estimation of heritabilities and genetic correlations

Sire, dam and progeny variance components were estimated by Henderson's Method 1 (1953) for egg composition and egg production traits using the nested model:

$$y_{ijk} = \mu + s_i + d_{ij} + p_{ijk}$$

where y_{ijk} is the observation on the k^{th} progeny within the j^{th} dam within the i^{th} sire;
μ is a constant common to all observations;
s_i, d_{ij} and p_{ijk} refer to random effects of sires, dams and progeny with zero means and variances σ_s^2, d_d^2, and σ_p^2 respectively.

To estimate the covariance components between traits A and B, a new trait, trait C was formed by summing values of traits A and B for each observation. Variance components for trait C were estimated, and covariance components between traits A and B

were obtained from the identity $Cov_{AB} = (\sigma^2_C - \sigma^2_A - \sigma^2_B)/2$ were σ^2_A, σ^2_B and σ^2_C are variance components for traits, A, B and C respectively. Sire half-sib and full-sib heritabilities and genetic correlations were estimated from these variance and covariance component estimates.

The approximate standard errors of the heritability estimates were estimated using the formulae reported by Dickerson (1960) as simplified by Becker (1967):

$$S.E.(h^2_{SHS}) \simeq \frac{\sqrt[4]{Var(\hat{\sigma}^2_s)}}{\hat{\sigma}^2_s + \hat{\sigma}^2_d + \hat{\sigma}^2_p}$$

$$S.E.(h^2_{FS}) \simeq \frac{\sqrt[2]{Var(\hat{\sigma}^2_s) + Var(\hat{\sigma}^2_d) + 2\,Cov(\hat{\sigma}^2_s, \hat{\sigma}^2_d)}}{\hat{\sigma}^2_s + \hat{\sigma}^2_d + \hat{\sigma}^2_p}$$

where SHS denotes sire half-sib and FS denotes full-sib.

The sampling variances and covariances of the variance component estimates in the numerators of the above formulae were estimated from the exact formulae given by Searle (1971), under the assumption that the effects in the model are normally distributed, and are exact only for specified values of the variances which are unknown.

The formula of Robertson (1959) was used to estimate the standard errors of genetic correlation estimates. This formula assumes equal family sizes, and equal and known heritability values of the traits.

Formation of simulated populations

The assumptions required to form simulated populations were as follows:

(a) the random effects in the model were normally distributed;

(b) the estimates of variances and covariances due to sires, dams and progeny derived from the experimental data were the true population parameters.

The method of obtaining a multinormally distributed observation vector of q traits described by Hocking and Smith (1967) is outlined below.

(a) Form a $q \times q$ symmetric matrix $\underset{\sim}{V}$ of variances and covariances for each random effect in the model (i.e. $\underset{\sim}{V}_s$, $\underset{\sim}{V}_d$, $\underset{\sim}{V}_p$).

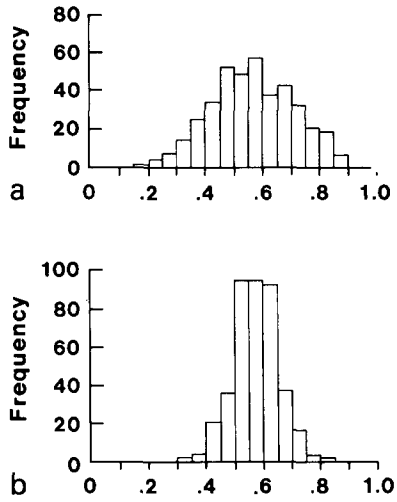

Fig. 1. Frequency distributions of (a) sire-half-sib and (b) full-sib hertability estimates for albumen percentage-solids from 400 simulated populations

(b) For each $\underset{\sim}{V}$ matrix, form a corresponding $q \times q$ lower triangular $\underset{\sim}{T}$ matrix, such that $\underset{\sim}{T}\underset{\sim}{T}' = \underset{\sim}{V}$. Let t_{ij}, v_{ij} represent elements of $\underset{\sim}{T}$ and $\underset{\sim}{V}$ respectively, then

$$t_{gg} = \left(v_{gg} - \sum_{k=1}^{g-1} t^2_{gk} \right)^{1/2}$$

$$t_{gj}(g > j) = \left(v_{gj} - \sum_{k=1}^{j-1} t_{jk} t_{gk} \right) / t_{jj}$$

$(g = 1, \ldots, q, (j = 1, \ldots, g))$.

(c) Using a random normal deviate generator, form a vector $\underset{\sim}{w}$ of q random normal deviates with $E(\underset{\sim}{w}) = \underset{\sim}{0}$ and $Var(\underset{\sim}{w}) = \underset{\sim}{I}$, (i.e. zero mean and unit variance).

(d) Postmultiply $\underset{\sim}{T}$ by $\underset{\sim}{w}$ to form $\underset{\sim}{u}$, which is a multinormally distributed vector of q effects with $E(\underset{\sim}{u}) = \underset{\sim}{0}$ and $Var(\underset{\sim}{u}) = \underset{\sim}{V}$. The vector $\underset{\sim}{u}$ is now a vector of simulated effects found in the model. The data from the original experiment were sorted progeny within dams and dams within sires. Then, taking each progeny in order, a vector $\underset{\sim}{u}_p$ of effects was generated for each progeny, a vector $\underset{\sim}{u}_d$ for each new dam and a vector $\underset{\sim}{u}_g$ for each new sire, and the three vectors summed to form the observation vector for q traits, just as in the statistical model:

$$y_{ijk} = \mu + s_i + d_{ij} + p_{ijk}.$$

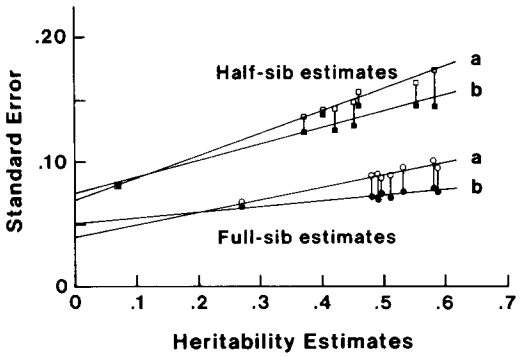

Fig.2. Relationship between heritability estimates and their standard errors using (a) approximate formulae and (b) the simulation technique
Standard errors for each trait are joined by a vertical line

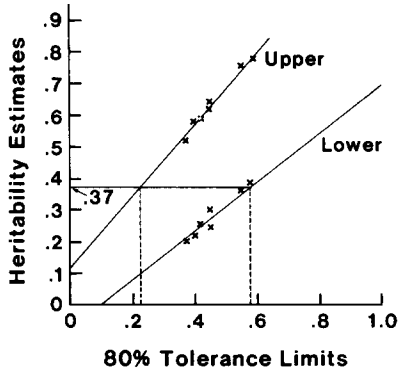

Fig.3. Graph for obtaining 80 % Tolerance Limits for heritability estimates for the specific population structure and method of estimation used to derive estimates

Without loss of generality, μ was set equal to zero. The design for the simulated observations was therefore identical to that in the original data.

When the simulated population was complete, genetic parameters were estimated by the same statistical analysis as the original experimental data. The expectations of sums of squares were identical for each simulated sample, and so it was unnecessary to recalculate them for each sample. This procedure was repeated 200 or 400 times, providing a frequency distribution of estimates for each genetic parameter (example, Fig.1), and standard deviations were estimated. The points beyond which the upper and lower 10 % of the distribution of estimates fell are known as estimates of the 80 % tolerance limits for the population.

Interpretation of results from the simulation technique

The variance of heritability estimates is found from the distribution of $\hat{h}^2$, an estimate of h^2, which depends on true (unknown) values of h^2, the number of sires, dams, and progeny, and the method of estimation. For a given design pattern of sires, dams and progeny and a method of estimation, the simulation method provides an approximation to the true sample variance of $\hat{h}^2$ for each fixed h^2. The standard deviation (s.e.$_s$) of the distribution is an estimate of the true standard error of $\hat{h}^2$.

If the assumed value of h^2 used to simulate the populations was close to the unknown true value of h^2 for that trait, then s.e.$_s$ is likely to be close to the true standard error of $\hat{h}^2$. The assumption is made that small differences between the assumed and true value of h^2 do not greatly change s.e.$_s$. A similar assumption is made for genetic correlation estimates.

Results

Heritability estimates and standard errors

The regression of s.e.$_s$ on h^2 was small for half-sib heritability estimates and almost zero for full-sib estimates, suggesting that the values for s.e.$_s$ were good approximations of the standard errors of estimates of the unknown heritabilities. The standard error (s.e.$_f$) calculated from the modified Dickerson's formula were slightly larger, (Fig.2) but were otherwise quite consistent with those from the simulation study.

The 80 % tolerance limits of the half-sib heritability estimates were plotted (Fig.3) against the parameter values on which the simulations were based. Regression lines were fitted to upper and lower limits.

When an estimate of heritability is available, 80 % confidence limits of the true heritability can be found by constructing a horizontal line through the value of the estimate on the vertical axis; this line will intersect the two regressions at points which may be read

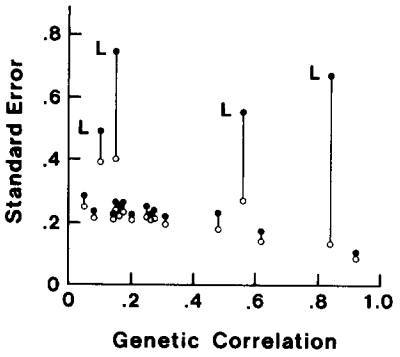

Fig. 4. Standard errors of sire-half-sib genetic correlation estimates from approximate formulae (white circles) and from simulation (black circles) Estimates involving egg laying intensity are marked with an "L"

off on the horizontal axis. These two points are the approximate 80 % confidence limits for the estimate (Mood et al. 1974) if the regressions are correctly assumed to be linear. For example, the confidence limits on the heritability estimate of 0.37 for egg weight were 0.22 and 0.58. Better estimates of the slopes of the regression lines would have been obtained had some "artificial" traits been simulated with h^2 ranging from 0.00 to 1.00 in steps of 0.10. The graph in Fig. 3 can only be used for the particular population size, structure, and method of estimation from which it was derived. That is, a design with 56 sires, each mated to 4 or 5 dams giving rise to 4 or 5 progeny per mating for a total of 998 observations.

Genetic correlations

The standard errors of sire half-sib genetic correlation estimates calculated from the approximate formula of Robertson (1959) and from the simulation study showed good agreement (Fig. 4) with four obvious exceptions. These exceptions apply to correlations involving egg laying intensity which had a low heritability (0.07). Those four cases seriously violated two of the assumptions on which the formula was based: (i) the heritabilities were far from equal 0.07 v 0.55; 0.07 v 0.45; 0.07 v 0.40; 0.07 v. 0.58); and (ii) the estimate of the heritability of laying intensity had a comparatively large standard error ($s.e._s = 0.07$) relative to the heritability value. In the remainder of the cases, the estimates were not greatly dissimilar (lowest 0.37; highest 0.58) and were fairly precisely estimated (standard errors of estimates were around 0.14). In all cases the assumption of equal family size was not greatly violated.

Fig. 4 suggests a dependence between the value of the correlation and its standard error. An overestimate of the true genetic correlation would lead to an underestimate of the true standard error and vice versa.

Discussion

The approximation formulae gave good estimates of the standard errors of heritability estimates because it can be shown that the numerator terms in the formulae are the actual standard errors of the estimates of additive genetic variance. The approximation formula assumes the denominator is a constant, but the the correct value of the standard error of heritabilities depends on the sampling variation of the denomiator and its correlation with the numerator. The formulae make no other assumptions except that the effects are normally distributed. For genetic correlations, the simulation technique produced estimates of standard errors which were quite close to the values obtained by the formula of Robertson (1959) when the assumptions were not seriously violated. The close agreement between standard error estimates by the approximation formulae and simulation procedure could be due to the nearly equal subclass numbers in this study. Since the approximation formulae assume equal subclass numbers, the simulation procedure might be expected to yield better estimates of standard errors than the formulae under drastically unequal subclass number situations.

The simulation technique appears to be a useful and easy method of finding approximate standard errors of heritabilities and genetic correlations, and can also be used to obtain approximate confidence limits around heritability estimates. Its great potential lies in that it is not limited to particular statistical models, nor to conditions of equal family sizes. For mixed models, the fixed effects can be assigned arbitrary values, the easiest are zero, and the simulated data can be analyzed exactly as the original data. If the distribution of effects is not normal, then

samples may be drawn from the appropriate distribution if it is known. Although the simulation technique may be laborius in terms of computer program development, it offers a general solution of the problem of estimating the sampling variation of genetic parameter estimates.

Acknowledgement

The authors wishes to thank Dr. C.R. Henderson for suggesting the simulation technique, the Animal Research Institute, Agriculture Canada, Ottawa for supplying some of the data used, and the Ontario Ministry of Agriculture and Food for financial support.

Literature

Becker, W.A.: A manual of procedures in quantitative genetics. 2nd ed. 130 pp. Pullman: Washington State Univ. press 1967

Dickerson, G.E.: Techniques for research in quantitative animal genetics. In: Techniques and Procedures in Animal Production Research. American Society of Animal Production, pp. 56-105 (1960)

Grossman, M.; Norton, H.W.: Simplication of the sampling variance of the correlation coefficients. Theor. Appl. Genet. 44, 332 (1974)

Henderson, C.R.: Estimation of variance and covariance components. Biometrics 9, 226-252 (1953)

Hocking, R.R.; Smith, W.B.: Generation of random samples from a Wishart distribution. Technical Report No. 6, Institute of Statistics, Texax A & M University, College Station, Texas (1967)

Mood, A.M.; Graybill, F.A.; Boes, D.C.: Introduction to the theory of statistics. New York: McGraw Hill 1974

Reeve, E.C.R.: The variance of the genetic correlation coefficient. Biometrics 11, 357-374 (1955)

Robertson, A.: The sampling variance of the genetic correlation coefficient. Biometrics 15, 469-485 (1959)

Rodda, D.D.: Quantitative genetics of egg composition. Ph.D. thesis, University of Guelph, Guelph, Ontario, Canada (1975)

Rodda, D.D.; Friars, G.W.; Gavora, J.S.; Merritt, E.S.: Quantitative genetics of egg composition. Br. Poult. Sci. (In press) (1977)

Searle, S.R.: Linear Models. New York: John Wiley 1971

Shook, G.E.; Barr, G.R.: Empirical sampling variances of genetic correlations computed from sire components of variance and covariance. J. Dairy Sci. 61, 959 (abstract) (1968)

Tallis, G.M.: Sampling errors of genetic correlation coefficients calculated from the analyses of variance and covariance. Aust. J. Statist. 1, 35-43 (1959)

Van Vleck, L.D.; Henderson, C.R.: Empirical sampling estimates of genetic correlations. Biometrics 17, 359-371 (1961)

Received April 15, 1977
Communicated by L.D. Van Vleck

Dr. D.D. Rodda
Scorrier, Redruth
Cornwall, TR15 5BT
United Kingdom

Dr. L.R. Schaeffer and Dr. G.W. Friars
Department of Animal and Poultry Science
Dr. K. Mullen
Department of Mathematics and Statistics
University of Guelph