Bias of maximum likelihood estimator of intraclass correlation

C. S. Wang[1], **B. S. Yandell**[2] **and J. J. Rutledge**[1]

[1] Department of Meat and Animal Science, 1675 Observatory Dr., University of Wisconsin, Madison, WI 53706, USA
[2] Departments of Horticulture and of Statistics, 1210 West Dayton St., University of Wisconsin, Madison, WI 53706, USA

Summary. A bias correction was derived for the maximum likelihood estimator (MLE) of the intraclass correlation. The bias consisted of two parts: a correction from MLE to the analysis of variance estimator (ANOVA) and the bias of ANOVA. The total possible bias was always negative and depended upon both the degree of correlation and the design size and balance. The first part of the bias was an exact algebraic expression from MLE to ANOVA, and the corrected estimator by this part was ANOVA. It was also shown that the first correction term was equivalent to Fisher's reciprocal bias correction on his Z scores. The total possible bias of MLE was large for small and moderate samples. Relative biases were larger for small parametric values and vice versa. To ensure a relative bias less than 10% assuming an intraclass correlation of 0.025, which is not unusual in most of the animal genetic studies, the total number of observations (N) should be not less than 500. From a design point of view, minimum bias occurred at $n=2$, the minimum family size possible, under N fixed.

Key words: Intraclass correlation – Maximum likelihood estimator – Bias

Introduction

A frequently used tool in human quantitative genetics and animal breeding is the intraclass correlation coefficient. In these disciplines, the intraclass correlation is often used to measure the degree of resemblance among family members. Thus, it is proportional to heritability (by one-quarter if derived from half-sib families, or one-half if derived from full-sib families), under certain assumptions. Heritability is a key parameter for prediction of selection gains. The square root of heritability measures accuracy of selection (Falconer 1982). A common method of estimating intraclass correlation is analysis of variance (ANOVA) using a one-way random effects model. Because of the availability of inexpensive and friendly computing, some computer-intensive statistical methods, such as maximum likelihood estimation (ML), have caught the attention of researchers in recent years. The maximum likelihood estimator of the intraclass correlation was derived by Rosner et al. (1977) for balanced data and by Donner and Koval (1980) for unbalanced data. Pearson's pairwise product-moment correlation, averaged over all possible pairs of observations that can be constructed within a family (Fisher 1925), is the MLE in the balanced case (Donner and Koval 1980). The MLE has desirable asymptotic properties including the asymptotic unbiasedness, a feature shared by all maximum likelihood estimators. However, the MLE is biased for small and moderate samples. This bias aspect of the MLE has not been adequately addressed with its increasing popularity. Fisher gave a bias correction for the product-moment correlation for the balanced case (Fisher 1925). However, his correction has either been treated lightly, or else ignored or misinterpreted (e.g., Weinberg and Patel 1981). It is the purpose of this note to show the extent of possible bias of the MLE for balanced data and to elucidate the nature of Fisher's bias correction. Numerical calculations will be given to demonstrate the extent of bias for certain design situations.

ANOVA and ML estimators of the intraclass correlation

Consider the ANOVA estimator under the one-way random effects ANOVA model:

$$y_{ij}=\mu+a_i+e_{ij}\,(i=1,\ldots,s,\text{ and }j=1,\ldots,n), \tag{1}$$

Table 1. Analysis of variance of model 1

Source	df	SS	MS	EMS
Between families	$s-1$	SSA	MSA	$\sigma_e^2 + n\sigma_a^2$
Within families	$s(n-1)$	SSE	MSE	σ_e^2

where $a_i \sim$ iid $(0, \sigma_a^2)$, $e_{ij} \sim$ iid $(0, \sigma_e^2)$, $\sigma_y^2 = \sigma_a^2 + \sigma_e^2$, and $E(y_{ij}) = \mu$. Usually a_i and e_{ij} are assumed to follow a normal distribution, but it is not necessary.

In the argument that follows, we shall refer to a_i as families without loss of generality. Let j and k index different members of family i; then the population intraclass correlation is defined as:

$$\varrho(y_{ij}, y_{ik}) = \mathrm{Cov}(y_{ij}, y_{ik})/(\sigma_{y_{ij}} \sigma_{y_{ik}}) = \sigma_a^2/(\sigma_a^2 + \sigma_e^2). \quad (2)$$

The analysis of variance of Eq. 1 is summarized in Table 1 (Snedecor and Cochran 1980, Sect. 13.5; and Fisher 1925, Sect. VII).

An ANOVA estimator or a sample version of ϱ is

$$\hat{\varrho}_A = \frac{\hat{\sigma}_a^2}{\hat{\sigma}_a^2 + \hat{\sigma}_e^2} = \frac{\mathrm{MSA} - \mathrm{MSE}}{\mathrm{MSA} + (n-1)\,\mathrm{MSE}}, \quad (3)$$

where $\hat{\sigma}_e^2 = \mathrm{MSE}$ and $\hat{\sigma}_a^2 = (\mathrm{MSA} - \mathrm{MSE})/n$.

For the MLE, an alternative model is considered. The assumptions of the model are: data y_{ij} are all distributed about a common mean with the same variance σ_y^2, and y_{ij} in the same family have a common correlation coefficient, ϱ, called the intraclass correlation coefficient (Snedecor and Cochran 1980). This alternative form of the one-way random model is also called a common correlation model (CCM) (Donner and Koval 1980). By further assuming multivariate normality of data, Rosner et al. (1977) derived the maximum likelihood estimator ($\hat{\varrho}_{CML}$) for balanced data by directly maximizing the likelihood function with respect to ϱ:

$$\hat{\varrho}_{ML} = \sum_{i=1}^{s} \sum_{j=1}^{n} \sum_{k=1}^{n} (y_{ij} - \bar{y})(y_{ik} - \bar{y})/\{N(n-1)\hat{\sigma}_y^2\}, \quad \text{for } j \neq k, \quad (4)$$

where

$$\bar{y} = \sum_{i=1}^{s} \sum_{j=1}^{n} y_{ij}/N,$$

$$\hat{\sigma}_y^2 = \sum_{i=1}^{s} \sum_{j=1}^{n} (y_{ij} - \bar{y})^2/N \quad \text{and} \quad N = s\,n.$$

Donner and Koval (1980) noted that $\hat{\varrho}_{ML} = \hat{\varrho}_P$, the pairwise intraclass correlation which is defined as the Pearson product-moment correlation averaged over all possible pairs of observations that can be constructed within families, i.e., a family of size n, could form $2\binom{n}{2}$ pairs (Fisher 1925).

Possible bias of MLE

We shall first consider bias of the ANOVA estimator, since it will be used as an additional bias for MLE. An approximate bias formula for balanced data was given by Ponzoni and James (1978):

$$E(\hat{\varrho}_A - \varrho) = \frac{-2(1-\varrho)\{\varrho + (1-\varrho)/n\}\{\varrho + (1-\varrho)/(sn)\}}{s-1}. \quad (5)$$

Notice that bias of $\hat{\varrho}_A$ is always downward. Numerical studies showed good agreement between Eq. 5 and the exact bias
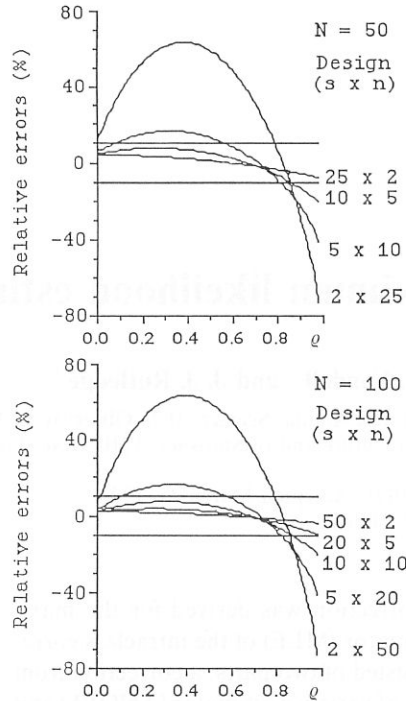
Fig. 1. Relative errors (%) of Ponzoni and James' (1978) bias formula. Relative error was the deviation of Eq. 5 from the exact integrated bias divided by the exact bias. The design label refers to number of families (s) and family size (n), respectively, and $N = sn$

calculated (Fig. 1) via the numerical integrations, using the exact sample distribution of the ANOVA estimator derived by Donner and Koval (1983). For the designs considered ($N = 50$ and 100), the relative errors of bias formula were well within the $\pm 10\%$ band for $\varrho < 0.8$ and $s > 5$ designs. Relative error was the deviation of Eq. 5 from the exact integrated bias divided by the exact bias.

Now we examine the relationship between ANOVA and MLE. Thus, the possible bias of MLE can be easily assembled. As shown in the appendix, the MLE can be written in terms of mean squares as:

$$\hat{\varrho}_{ML} = \frac{\mathrm{MSA}(s-1)/s - \mathrm{MSE}}{\mathrm{MSA}(s-1)/s + (n-1)\,\mathrm{MSE}}. \quad (6)$$

The difference between MLE and ANOVA is:

$$\Delta\hat{\varrho} = \hat{\varrho}_{ML} - \hat{\varrho}_A$$
$$= -\frac{n\,\mathrm{MSA}\,\mathrm{MSE}}{[\mathrm{MSA} + (n-1)\,\mathrm{MSE}][(s-1)\,\mathrm{MSA} + s(n-1)\,\mathrm{MSE}]}. \quad (7)$$

This difference is never positive. If one wishes, $\Delta\hat{\varrho}$ can also be expressed in terms of $\hat{\varrho}_{ML}$ and $\hat{\varrho}_A$ as:

$$\Delta\hat{\varrho} = -\frac{(1 - \hat{\varrho}_{ML})(1 + (n-1)\hat{\varrho}_{ML})}{1 + n(s-1) + (n-1)\hat{\varrho}_{ML}}, \quad (8)$$

and

$$\Delta\hat{\varrho} = -\frac{(1 - \hat{\varrho}_A)(1 + (n-1)\hat{\varrho}_A)}{N - 1 - (n-1)\hat{\varrho}_A}, \quad \text{respectively.} \quad (9)$$

The possible bias of MLE, Bias $(\hat{\varrho}_{ML})$, is simply the sum of $\Delta\hat{\varrho}$ and Eq. 5, or Bias$(\hat{\varrho}_{ML}) = \Delta\hat{\varrho} + (5)$.

Results

It was noted earlier that Fisher's $\hat{\varrho}_P = \hat{\varrho}_{ML}$. The bias given by Fisher was in terms of Z scores. Recall that (Fisher 1925):

$$Z = \frac{1}{2} \ln \frac{1+(n-1)\,\hat{\varrho}_{ML}}{1-\hat{\varrho}_{ML}}, \qquad \hat{\varrho}_{ML} = \frac{e^{2Z}-1}{e^{2Z}+(n-1)},$$

and $\text{Bias}(Z) = -\dfrac{1}{2} \ln \dfrac{s}{s-1}$.

The bias-corrected Z, $Z^* = Z - \text{Bias}(Z)$. Consequently, the bias-corrected estimator found by substituting Z^* in the place of Z is

$$\hat{\varrho}^* = \frac{\hat{\varrho}_{ML}\,(N-1)+1}{\hat{\varrho}_{ML}\,(n-1)+n(s-1)+1}. \qquad (10)$$

The difference between $\hat{\varrho}_{ML}$ and $\hat{\varrho}^*$ is

$$\Delta\hat{\varrho}^* = \hat{\varrho}_{ML} - \hat{\varrho}^* = -\frac{(1-\hat{\varrho}_{ML})(1+(n-1)\,\hat{\varrho}_{ML})}{1+n(s-1)+(n-1)\,\hat{\varrho}_{ML}}. \qquad (11)$$

This is the same form as that of Eq. 8. That is to say that the Fisher's bias correction for MLE is not an unbiased correction. Instead, the corrected estimator is the ANOVA estimator, which has an additional bias of Eq. 5.

The bias depends on both correlation and design. The bias of the MLE is always negative and is composed of two parts: one is an exact algebraic correction term from MLE to ANOVA and the other is the additional bias inherent in the ANOVA estimator.

To investigate the extent of the bias, we conducted a number of numerical calculations using the sum of Eqs. 9 and 5. Notice that the first term, Eq. 9, was a function of the ANOVA estimator rather than the parametric value. Treating the ANOVA as the parametric value during the calculations might introduce some biases. However, these introduced biases were negligible since the bias of ANOVA was relatively small. The results are summarized in Figs. 2–4.

Figure 2 gives the biases and relative biases of $\hat{\varrho}_{ML}$ as functions of ϱ under different layouts. Relative biases
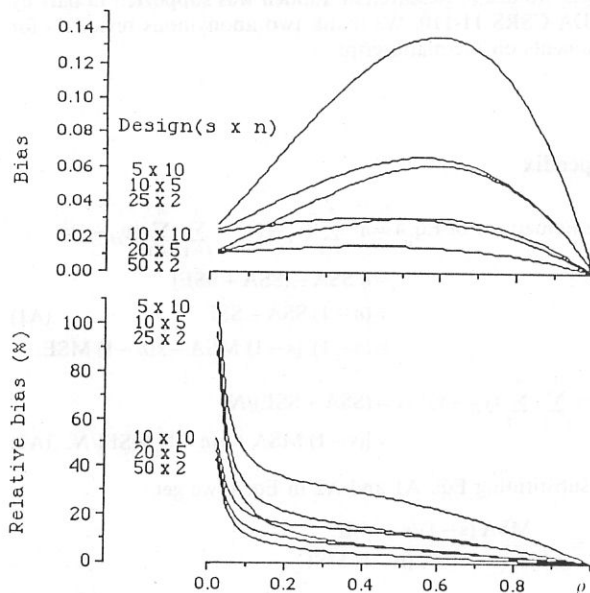
Fig. 2. Bias and relative bias (%) of $\hat{\varrho}_{ML}$ under different layouts. Relative bias was defined as bias over parametric value. The design label refers to number of families (s) and family size (n), respectively, and $N=sn$
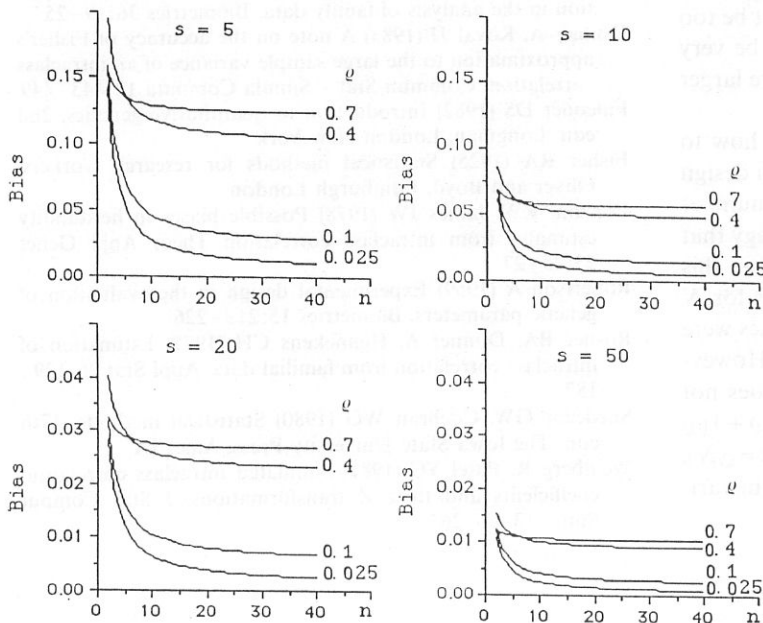
Fig. 3. Bias of $\hat{\varrho}_{ML}$ as a function of family size (n)

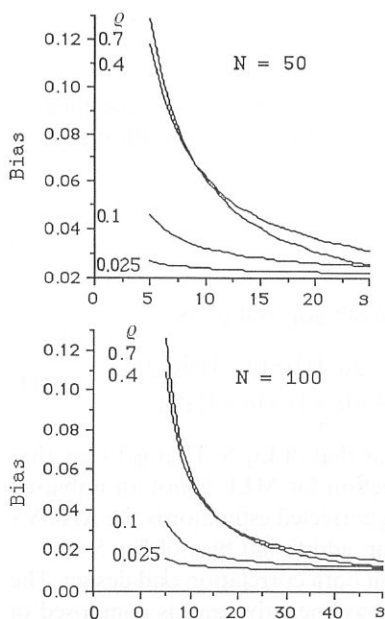**Fig. 4.** Bias of $\hat{\varrho}_{ML}$ as a function of s (number of families) under N (total number of observations) fixed

(RB) were defined as biases over parametric values. Six layouts were used; for example, 20×5 meant a design of 20 families, each having five offspring. Biases were large (Fig. 2), usually not less than 20% of relative biases, for the designs considered. Secondly, there was a maximum for a given design. Maxima varied from 0.425 to 0.60 for different layouts. Furthermore, pictures of relative biases were different from those of biases. RBs decreased unimodally with ϱ. RBs of small ϱ ($\varrho \leq 0.1$) were substantially larger than that of large ϱ.

Figure 3 shows biases of $\hat{\varrho}_{ML}$ as a function of n, given s fixed. For all the cases considered, there was one clear message that the number of families, s, should not be too small (not less than 20); otherwise, biases could be very large. Consider varying ϱ's: biases of large ϱ's were larger than those of small ϱ's, and vice versa for RBs.

Finally, it is always interesting to determine how to obtain smaller biases under fixed resources from a design point of view. In other words, given N, the total number of observations, what is the 'best' allocation strategy that would give the smallest bias possible? To answer this question, biases and RBs were plotted against s for N fixed (Fig. 4). It was very clear that minimum biases were attained at $n = 2$, the smallest family size possible. However, this optimal condition with respect to bias does not coincide with Robertson's optimal condition, $n = (\varrho + 1)/\varrho$ or $s = \varrho N/(\varrho + 1)$, or approximately $n = 1/\varrho$ (or $s = \varrho N$), when ϱ is very small, in terms of yielding minimum variance under N fixed (Robertson 1959).

Overall, to yield a RB not greater than 10%, roughly 500 data points (N) are needed for $\varrho = 0.025$, 250 for $\varrho = 0.1$, and 100 for $\varrho = 0.5$. In most animal genetic studies, ϱ is usually not greater than 0.1. Therefore, the bias of $\hat{\varrho}_{ML}$ would be large for small and moderate samples.

Appendix

The numerator of Eq. $4 = n^2 \sum_{i=1}^{s} (\bar{y}_i - \bar{y})^2 - \sum_{i=1}^{s} \sum_{j=1}^{n} (y_{ij} - \bar{y})^2$

$$= n \, \text{SSA} - (\text{SSA} + \text{SSE})$$
$$= (n-1) \, \text{SSA} - \text{SSE} \qquad (A1)$$
$$= (n-1)(s-1) \, \text{MSA} - s(n-1) \, \text{MSE}.$$

$$\hat{\sigma}_y^2 = \sum_{i=1}^{s} \sum_{j=1}^{n} (y_{ij} - \bar{y})^2/N = (\text{SSA} + \text{SSE})/N$$
$$= [(s-1) \, \text{MSA} + s(n-1) \, \text{MSE}]/N. \qquad (A2)$$

By substituting Eqs. A1 and A2 in Eq. 4, we get

$$\hat{\varrho}_{ML} = \frac{\text{MSA}(s-1)/s - \text{MSE}}{\text{MSA}(s-1)/s + (n-1)\,\text{MSE}}.$$

References

Donner A, Koval JJ (1980) The estimation of intraclass correlation in the analysis of family data. Biometrics 36:19–25

Donner A, Koval JJ (1983) A note on the accuracy of Fisher's approximation to the large sample variance of an intraclass correlation. Commun Stat – Simula Computa 12:443–449

Falconer DS (1982) Introduction to quantitative genetics, 2nd edn. Longman, London New York

Fisher RA (1925) Statistical methods for research workers. Oliver and Boyd, Edinburgh London

Ponzoni RW, James JW (1978) Possible biases in heritability estimates from intraclass correlation Theor Appl Genet 53:25–27

Robertson A (1959) Experimental design in the evaluation of genetic parameters. Biometrics 15:219–226

Rosner BA, Donner A, Hennekens CH (1977) Estimation of intraclass correlation from familial data. Appl Stat 26:179–187

Snedecor GW, Cochran WG (1980) Statistical methods, 17th edn. The Iowa State University Press, Ames/IA

Weinberg R, Patel YC (1981) Simulated intraclass correlation coefficients and their Z transformations. J Stat Comput Simul 13:13–26