ESTIMATION OF GENETIC COVARIANCE FROM JOINT OFFSPRING-PARENT AND SIB-SIB STATISTICS

DANIEL GIANOLA
Department of Animal Science,
University of Illinois, Urbana 61801

Manuscript received March 16, 1979
Revised copy received May 17, 1979

ABSTRACT

Formulae for calculating the variances and covariances of estimates of genetic variance or covariance from offspring-parent and sib covariance data are presented. The model consists of a one-way classification of families with unequal numbers of offspring, where normally distributed traits $x$ (a parental measure), $y$ and $z$ (offspring measures) are recorded. Procedures for pooling offspring-parent and sib covariance estimators are discussed. An application of the results to estimating genetic variance is presented, and the offspring-parent and sib covariance estimators and a pooled statistic are compared in terms of exact and asymptotic formulae. Implications for experimental design and allocation of experimental resources are also reported.

OPTIMAL decisions for selection to improve quantitative characters in plant or animal breeding depend on the availability of reliable estimates of genetic and phenotypic variances and covariances (HENDERSON 1963; HARRIS 1964; VANDEPITTE and HAZEL 1977). Most methods of estimating genetic parameters consist of equating a function of the observations, a covariance between relatives, to its expectation in terms of a genetic model, and then proceeding to fit the parameters of this model by ad hoc or well defined procedures such as maximum likelihood or least-squares. Depending on the mating design, a number of covariances among relatives may arise, and the investigator is faced with the problem of interpreting differing estimates of the same genetic parameter and their relative agreement or disagreement. It is difficult to do so in a judicious manner without considering the correlation structure between the estimates. In addition, it would be desirable to develop a single estimator combining the available information from all covariances between relatives in an “optimal” manner, in some sense.

EISEN (1967) and HILL and NICHOLAS (1974) were apparently the first to address the question of pooling the information from all sources arising in specific mating designs into a single “best” estimator. Inter alia, HILL and NICHOLAS (1974) derived an approximate expression for the correlation between the offspring-parent and the full-sib estimators of heritability and proposed a maximum-likelihood estimator of heritability. As it happens, in maximum-likelihood estimation of variance components (SEARLE 1970), the estimators could not be
obtained explicitly, but the authors derived, under normality, its asymptotic variance. They assumed additive genetic effects as the only source of correlation among relatives and their approximate formulae hold true only for the balanced case, i.e., where family sizes are constant throughout the experimental region. THOMPSON (1976) later extended the maximum likelihood procedure to the case of selection and assortative mating.

In this paper, we consider an experimental situation in which a number of parents have an arbitrary number of offspring. For the case where there is neither selective nor assortative mating, an exact expression is presented for the correlation between a sib-sib estimator of the covariance between traits \( y \) and \( z \) and the estimator of the offspring-parent covariance between traits \( x \) and \( y \). The only assumption is that \( x \), \( y \) and \( z \) are normally distributed, but the two covariances between relatives may have different expectations; for the case where a single estimator is desired, pooling procedures are presented. Developments are then applied to the estimation of genetic variance and, for simplicity, the balanced case is used. In this context the following were considered: (a) correlation between offspring-parent and half-sib estimators of genetic variance; (b) variances of the estimators; (c) large sample approximations; and (d) allocation of experimental resources. It should be noted that the estimation of genetic variances and covariances is considered in this paper. In so doing, the problem of dealing with the distribution of ratios is circumvented and, therefore, exact results can be obtained. In fact, in most practical applications of selection theory, such as index selection and best linear unbiased prediction (HENDERSON 1963, 1973), parameterizations in terms of heritability and genetic correlations may not be to the advantage of the breeder since only an approximate distribution theory is available.

**THEORY**

Consider a vector of parental records \( x \), with elements \( x_i, i = 1 \ldots s \), such that \( x \sim N(\mu_x, \mathbf{I} \sigma_x^2) \), and the vectors of progeny records \( y \sim N(\mu_y, \mathbf{V} \sigma_y^2) \) and \( z \sim N(\mu_z, \mathbf{V} \sigma_z^2) \). Further, assume that \( y_{ij} \) and \( z_{ij} \), which are individual elements of \( y \) and \( z \), respectively, can be described by the model:

\[
\begin{align*}
y_{ij} &= \mu_y + s_i + e_{ij} \\
z_{ij} &= \mu_z + s_i' + e_{ij}'
\end{align*}
\]

with \( i = 1 \ldots s, j = 1 \ldots n_i, N = \sum_n n_i, \sigma_y^2 = \sigma_g^2 + \sigma_e^2, \sigma_z^2 = \sigma_g^2 + \sigma_e^2 \), \( \text{Cov}(s_i, s_i') = \psi_{yy}, \text{Cov}(e_{ij}, e_{ij'}) = \psi_{ee} \); any other covariances are assumed to be zero. \( \mathbf{V} \) is a block diagonal matrix with elements \( v_{jk,lm} = 1 \) when \( j = l \) and \( k = m \), \( v_{jk,lm} = t_i = \sigma_g^2/\sigma_e^2 \) (\( i=y,z \)) when \( j = l \) and \( k \neq m \), and 0 otherwise.

The offspring-parent covariance between \( x \) and \( y \) is estimated by computing the usual repeat-record crossproduct

\[
c = \frac{1}{k} \sum_i \sum_j (x_i - \bar{x})(y_{ij} - \bar{y}) = \frac{1}{k} \sum_i (x_i - \bar{x})(y_i, -n_i \bar{y}) = \frac{1}{k} \mathbf{x}^* \mathbf{y}^*
\]

where the dot indicates a summation over the subscript, \( \bar{x} = \sum_i x_i / s, \bar{y} = \sum_j y_{ij} / N \).
and $k = N - \sum n_i^2/N$. In (2), $x^*$ is an $s \times 1$ vector with elements $x_i^* = x_i - \bar{x}$, and $y^*$ is an $s \times 1$ vector with elements $y_i^* = y_i - \bar{y}$. It follows that $x^* \sim N(O, C)$ where $C$ has elements $c_{ii} = \left(\frac{s-1}{s}\right)\sigma^2_i$ for $i = 1 \ldots s$ and $c_{ij} = -\sigma^2_i/s$ for $i \neq j$. Likewise $y^* \sim N(O, V^*)$, where $V^*$ has elements

$$

v_{ii}^* = \frac{n_i \sigma^2_i}{N} \{N(1-t_i) + n_i [t_y(N - 2n_i + \frac{\sum n_i^2}{N} + 1) - 1]\} \quad i = 1 \ldots s

$$

and

$$

v_{ij}^* = -\frac{n_i n_j}{N} \sigma^2_i \{1 + t_y(n_i + n_j - \sum n_i^2/N - 1)\}, \quad i \neq j .

$$

Further, let Cov($x^*, y^*$) = $K$, where $k_{ii} = \psi_{yy} n_i \left(\frac{N-n_i}{N}\right)$, $i = 1 \ldots s$, and $k_{ij} = -n_i n_j / \psi_{yy} / N$; here, $\psi_{yy}$ denotes the offspring-parent covariance, and $E(c) = \psi_{yy}$.

In the estimation of sib-sib covariances, the following bilinear forms are considered:

$$

T_0 = z'y, \quad T_\sigma = (z_1, \ldots, z_s) (y_1/n_1, \ldots, y_s/n_s)^',

$$

and $T_\mu = \frac{1}{N} z'Jy$, where $J$ is an $n \times n$ matrix of one's. An unbiased estimator of $\psi_{yz}$ is given by:

$$

\hat{\psi}_{yz} = \frac{1}{k} (w_1 T_\sigma + w_2 T_0 + w_3 T_\mu)

$$

where $w_1 = (N-1)/(N-s)$, $w_2 = (1-s)/(N-s)$, and $w_3 = -1$.

**Sampling variance of the estimator of offspring-parent covariance (EOPC)**

Given that $x^*$ and $y^*$ are normally distributed with covariance matrix $K$, the variance of the bilinear form $c = x^*y^*/k$ is derived as per SEARLE (1971).

$$

\text{Var}(c) = \frac{1}{k^2} \left[ \text{tr}(K^2) + \text{tr}(V^*C) \right]

$$

where tr denotes the trace operation. The following expression is obtained:

$$

\text{Var}(c) = \frac{\psi_{yz}}{k^2 N^2} [\sum_i n_i^2 (N-n_i)^2 + \sum_i \sum_{i \neq j} n_i^2 n_j^2] + \frac{\sigma^2_i}{k^2 s} \left[ (s-1) \sum_i v_{ii}^* - \sum_{i \neq j} \sum_i v_{ij}^* \right]

$$

**Sampling variance of the estimator of sib-sib covariance (ESSC)**

Following SEARLE (1956), it can be shown that:

$$

\text{Var}(\hat{\psi}_{yz}) = \frac{(s-1)^2}{k} \left\{ \frac{(N-1) (\psi_{yz}^{E2} + e'e')}{N(s-1)} + \frac{Nk(2 \psi_{yz}^{E2} \psi_{yz} + e's + es')}{N(s-1)^2} \right\}

$$

$$

+ \frac{N^2 \sum_i n_i^2 + (\sum_i n_i^2)^2 - 2N \sum_i n_i^3}{N^2 (s-1)^2} \left\{ \psi_{yz}^{E2} + ss' \right\}

$$

where $e = \sigma^2_y$, $e' = \sigma^2_z$, $s = \sigma^2_y$, $s' = \sigma^2_z$. 


Covariance between EOPC and ESSC

The covariance between the estimator of the offspring-parent covariance and the estimator of the sib-sib covariance is given by

\[
\text{Cov}(c, \hat{\Psi}_{yz}) = \frac{1}{K^2} \left[ \nu_1 \text{Cov}(x^*y^*, T_s) + \nu_2 \text{Cov}(x^*y^*, T_o) + \nu_3 \text{Cov}(x^*y^*, T_{\mu}) \right]
\]  

(9)

We first consider the bilinear forms \( \text{Cov}(x^*y^*, T_s) = \text{Cov}(x^*y^*, z') \), where \( z = (z_1, \ldots, z_s) \) and \( y' = (y_1/n_1, \ldots, y_e/n_e) \). Since \( z \) and \( y' \) are linear combinations of \( z \) and \( y \), respectively, it can be assumed that they follow a normal distribution. Therefore,

\[
\text{Cov}(x^*y^*, z') = \text{tr} (C_{y^s} C_{x^s}) + \text{tr} (C_{y^s} C_{z^s})
\]  

(10)

where the \( C \)'s denote the covariance matrices of the vectors indicated in the subscripts. Individual elements of the matrices in (10) above are the following:

\[
\begin{align*}
        c_{y^s,z_i^s} &= n_i (N-n_i) (n_i \Psi_{yz} + \Psi_{yy})/N \\
   c_{y^s,z_j^s} &= -n_i n_j (n_i \Psi_{yz} + \Psi_{yy})/N \\
   c_{y^s,y_i^s} &= \Psi_{yz} (s-1)/s \\
   c_{y^s,y_j^s} &= -\Psi_{yz}/s \\
   c_{y^s,y_i'j} &= \sigma_y^2 \left[ 1 + (n_i-1) t_y \right] \\
   c_{y^s,y_i'j} &= -\sigma_y^2 n_i \left[ 1 + (n_j-1) t_y \right]/N \\
   c_{y^s,z_i^s} &= \Psi_{zz} n_i (s-1)/s \\
   c_{y^s,z_j^s} &= -\Psi_{zz} n_i/s
         \end{align*}
\]

After carrying out the computations indicated in equation (10),

\[
\text{Cov}(x^*y^*, T_s) = \frac{1}{Ns} \left[ (s-1) \sum_i \sum_{i'j} (N-n_i) \theta_i + \sum_{i'j} n_i n_j \theta_{ij} \right] ,
\]  

(11)

where \( \theta_i = (n_i \Psi_{yz} + \Psi_{yy}) + \sigma_y^2 \Psi_{xx} \left[ 1 + (n_i-1) t_y \right] \).

Next, consider

\[
\text{Cov}(x^*y^*, T_o) = \text{Cov}(x^*y^*, z'y) = \text{tr} (C_{y^s} C_{x^s} + C_{y^s} C_{z^s})
\]  

(12)

with the \( C \) matrices having elements:

\[
\begin{align*}
        c_{y^s,z_i^s} &= (N-n_i) (n_i \Psi_{yz} + \Psi_{yy})/N \\
   c_{y^s,z_j^s} &= -n_i n_j (n_i \Psi_{yz} + \Psi_{yy})/N \\
   c_{y^s,y_i^s} &= \Psi_{yz} (s-1)/s \\
   c_{y^s,y_j^s} &= -\Psi_{yz}/s \\
   c_{y^s,y_i'j} &= \sigma_y^2 \left[ 1 + (n_i-1) t_y \right] \\
   c_{y^s,y_i'j} &= -\sigma_y^2 n_i \left[ 1 + (n_j-1) t_y \right]/N \\
   c_{y^s,z_i^s} &= \Psi_{xx} (s-1)/s \\
   c_{y^s,z_j^s} &= -\Psi_{xx}/s
         \end{align*}
\]

It is then direct to show that:

\[
\text{Cov}(x^*y^*, T_o) = \text{Cov}(x^*y^*, T_o) .
\]  

(13)
Third, consider

\[
\text{Cov}(x^*y^*, T_{\mu}) = \text{Cov}(x^*y^*, \frac{1}{N} z' J y) = \frac{1}{N} \text{tr} \left( C_{y^*} J C_{x^*} + C_{y^*} J C_{x^*} \right). \tag{14}
\]

Individual elements of the covariance matrices involved in equation (14) have already been presented. Thus,

\[
\text{Cov}(x^*y^*, T_{\mu}) = \frac{1}{N^2 s} \left\{ \sum_{i} (sn_i-N)n_i(N-n_i) \theta_i - \sum_{i \neq j} (sn_i-N)n_in_j \theta_j \right\}. \tag{15}
\]

In view of (9), (13) and of \(\omega_1+\omega_2=1\), it follows then that

\[
\text{Cov}(c, \hat{\Psi}_{yz}) = \frac{1}{k^2} \left[ \text{Cov}(x^*y^*, T_s) - \text{Cov}(x^*y^*, T_{\mu}) \right], \tag{16}
\]

or, more explicitly

\[
\text{Cov}(c, \hat{\Psi}_{yz}) = \frac{1}{k^2 N^2 s} \left[ \sum_{i} n_i(N-n_i)^2 \theta_i + \sum_{i \neq j} n_i n_j \theta_j \right]. \tag{17}
\]

From equations (7), (8) and (17), it is possible to form an expression for the exact correlation between the estimator of offspring-parent covariance and the estimator of sib-sib covariance. This correlation is a function of many parameters and, whenever dominance, epistasis, environmental correlations and maternal effects enter into the picture, it cannot be given a simple interpretation. However, an investigator may wish to insert parameter estimates and proceed to estimate such a correlation in the context of a specific data set. In the latter situation, he would be in a better position to assess whether or not estimates of offspring-parent and sib-sib covariances from the same data set—as opposed to independent estimates—"agree" or "disagree" when in fact they are expected to do so.

**Pooling of estimators**

If functions of the records can be written as linear functions of a vector of genetic parameters, then it is possible to find a best linear unbiased estimator of these parameters. The solution to this problem for any vector of estimated covariances is given by the generalized least-squares estimator (Aitken 1935) and an example in the case of heritability estimation is presented by Hill and Nicholas (1974).

Suppose that two unbiased estimators of the same genetic parameter are available, \(E_1\) and \(E_2\). Then, the generalized least squares solution is given by

\[
E_p = \frac{\text{Var}(E_2) - \text{Cov}(E_1, E_2)}{\text{Var}(E_1 - E_2)} E_1 + \frac{\text{Var}(E_1) - \text{Cov}(E_1, E_2)}{\text{Var}(E_2 - E_1)} E_2 \tag{18}
\]

with \(\text{Var}(E_p) = \left[ \text{Var}(E_1) \cdot \text{Var}(E_2) - \text{Cov}^2(E_1, E_2) \right] / \text{Var}(E_1 - E_2)\). \tag{19}

In general, only parameter estimates are available to calculate (18) and (19); hence, poor estimates may lead to considerable departure from optimality. It is easy to see that \(E_p\) in (18) is unbiased for the parameter since the sum of the weights is equal to one. If there are more than two estimators of the same parameter \(\hat{\Psi}\), say \(E_1, E_2, \ldots, E_m\) such that \(E(E_i) = \hat{\Psi}\) and \(\text{Var}(E) = C\), where \(E\) is the vector of estimates, then
\[ E_p = \left( \sum_i \sum_j c_{ij}^* E_i \right) / \left( \sum_i \sum_j c_{ij}^* \right), \]  

(20)

where \( c_{ij}^* \) is an element of \( C^{-1} \). Likewise,

\[ \text{Var}(E_p) = \left( \sum_i \sum_j c_{ij}^* \right)^{-3} \]

(21)

Finally, if \( E(E_i) = k_i; \Psi \), where \( k_i \) is a \( p \times 1 \) vector of constants and \( \Psi \) is \( p \times 1 \) vector of parameters such that \( p < m \):

\[ \hat{\Psi} = (K'C^{-1}K)^{-1}K'C^{-1}E \text{ and } \text{Var}(\hat{\Psi}) = (K'C^{-1}K)^{-1} \]

where \( K' = (k_1, k_2, \ldots, k_m) \) is a \( p \times m \) matrix. This is, of course, the case where the \( E_i \)'s may have different expectations, and \( \hat{\Psi} \) is then the generalized least-squares estimator. If \( C \) is estimated from the data, \( \hat{\Psi} \) should be obtained iteratively.

**APPLICATION TO THE ESTIMATION OF GENETIC VARIANCE**

This is the case where the same trait is scored in the offspring and in the parents, so that \( x, y \) and \( z \) index the same random variable. For simplicity, the balanced case (\( n \) progeny per family, \( N = ns \)) is used to illustrate the principles developed in the previous section and half-sib families are considered.

From equation (7), it can be shown that:

\[ \text{Var}(c) = \frac{\Psi^2}{s-1} + \frac{\sigma^4}{s-1} \left[ \frac{1 + (n-1)t}{n} \right], \]

(22)

where \( \Psi \) is the offspring-parent covariance. In the absence of maternal genetic effects and of offspring-parent environmental correlations, \( \Psi = h^2 \sigma^2/2 \), where \( h^2 \) is the heritability of \( y \). Since \( c \) is an unbiased estimator of \( \Psi \), a first estimator of genetic variance is given by \( E_1 = 2c \). Further, if the covariance among half sibs is entirely due to additive effects, \( t = h^2/4 \) and:

\[ \text{Var}(c) = \frac{\sigma^4}{s-1} \left[ h^4/4 + \frac{1 + (n-1)h^2/4}{n} \right]. \]

(23)

If \( s \) and \( n \) are sufficiently large such that \( s/(s-1) \approx 1 \), \( n/(n-1) \approx 1 \), and terms of order \( s^{-1} \) and \( n^{-1} \) can be ignored relative to one, it follows that

\[ \text{Var}(E_1) \approx h^2(1 + h^2) \sigma^4/s. \]

(24)

Equation (8) reduces to

\[ \text{Var}(\hat{\Psi}_{yy}) = \frac{2\sigma^4}{n(s-1)} \left[ \frac{(1-t)^2}{n-1} + 2t(1-t) + nt^2 \right]. \]

(25)

If the covariance among half-sibs is entirely due to additive genetic effects, \( \Psi_{yy} = h^2 \sigma^2/4 \), and an additional unbiased estimator of the genetic variance is given by \( E_2 = 4\hat{\Psi}_{yy} \). With large \( n \) and \( s \),

\[ \text{Var}(E_2) \approx 2h^4\sigma^4/s. \]

(26)
Finally, equation (17) becomes:

\[ \text{Cov}(c, \hat{\Psi}_{yy}) = \frac{2\psi\sigma^2}{s-1} \frac{[1 + (n-1)t]}{n} \]

\[ = \frac{h^2\sigma^4}{s-1} \frac{[1 + (n-1)h^2/4]}{n}, \]  \hspace{1cm} (27)

and for large \( n \) and \( s \),

\[ \text{Cov}(E_1, E_2) = \text{Cov}(2c, 4\hat{\Psi}_{yy}) = 2h^2\sigma^4/s \]  \hspace{1cm} (28)

**Correlation between \( E_1 \) and \( E_2 \)**

Table 1 presents values of the correlation between \( E_1 \) and \( E_2 \), based on exact expressions (equations 23, 25, 27) and on large sample approximations (equations 24, 26, 28), as a function of heritability. It is clear that the correlation between offspring-parent and half-sib estimators of genetic variance is a sizeable one, except when the heritability is near zero. The correlation between \( E_1 \) and \( E_2 \) also depends on the total size of the data set, which would be \( T = s(n+1) \), and on the allocation of experimental resources to parents and offspring (mating design). In this respect, \( r_{E_1,E_2} \) seems to be more sensitive to the mating design in smaller experiments, except at high heritability levels. In very large samples, \( r_{E_1,E_2} = [2h^2/(1+h^2)]^{1/2} \); this formula provides an upper limit for the value of the correlation and approximates well its exact value, particularly when heritability is high.

It should be concluded that, in general, offspring-parent and half-sib estimators of genetic variance are not statistically independent; therefore, pooling procedures that ignore this dependency structure may be grossly inappropriate.

**Variances of the estimators**

It is possible to address analytically the question of the relative efficiency of \( E_1 \) and \( E_2 \) by finding values of \( h^2 \), \( n \) and \( s \) that satisfy \( \text{Var}(E_1) = \text{Var}(E_2) \). In so

---

**TABLE 1**

<table>
<thead>
<tr>
<th>( h^2 )</th>
<th>( s=20, n=60 )</th>
<th>( s=60, n=20 )</th>
<th>( s=60, n=100 )</th>
<th>( s=100, n=60 )</th>
<th>LSA*</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>0.1</td>
<td>0.34</td>
<td>0.25</td>
<td>0.36</td>
<td>0.37</td>
<td>0.43</td>
</tr>
<tr>
<td>0.2</td>
<td>0.51</td>
<td>0.43</td>
<td>0.54</td>
<td>0.51</td>
<td>0.58</td>
</tr>
<tr>
<td>0.3</td>
<td>0.63</td>
<td>0.56</td>
<td>0.65</td>
<td>0.63</td>
<td>0.68</td>
</tr>
<tr>
<td>0.4</td>
<td>0.72</td>
<td>0.66</td>
<td>0.73</td>
<td>0.72</td>
<td>0.76</td>
</tr>
<tr>
<td>0.5</td>
<td>0.79</td>
<td>0.73</td>
<td>0.80</td>
<td>0.79</td>
<td>0.82</td>
</tr>
<tr>
<td>0.6</td>
<td>0.84</td>
<td>0.79</td>
<td>0.85</td>
<td>0.84</td>
<td>0.87</td>
</tr>
<tr>
<td>0.7</td>
<td>0.89</td>
<td>0.85</td>
<td>0.90</td>
<td>0.89</td>
<td>0.91</td>
</tr>
<tr>
<td>0.8</td>
<td>0.93</td>
<td>0.89</td>
<td>0.93</td>
<td>0.93</td>
<td>0.94</td>
</tr>
<tr>
<td>0.9</td>
<td>0.96</td>
<td>0.93</td>
<td>0.96</td>
<td>0.96</td>
<td>0.97</td>
</tr>
<tr>
<td>1.0</td>
<td>0.99</td>
<td>0.96</td>
<td>0.99</td>
<td>0.99</td>
<td>1.00</td>
</tr>
</tbody>
</table>

* Large sample approximation: \( r = [2h^2/(1+h^2)]^{1/2} \).
Table 2 presents the exact values of $n$ at which $\text{Var}(E_i) = \text{Var}(E_s)$ for several heritability levels. For heritability ranging between 0.05 and 0.50, the point of isoefficiency ($n^*$) varies between 9.63 and 19.34 and then increases considerably up to a very large number when $h^2=1.0$. In addition, the following conditions are true at each $h^2$ value:

\[(29) \quad \text{Var}(E_s) > \text{Var}(E_i) \quad \text{when} \quad n < n^* \]
\[(30) \quad \text{Var}(E_s) < \text{Var}(E_i) \quad \text{when} \quad n > n^* \]

In dairy cattle breeding, with artificial insemination, sire families are generally very large and the heritability of most production related traits ranges between 0.20 and 0.50; therefore, it is likely that in most data sets, estimates of genetic variance from half-sib statistics would be expected to have smaller sampling variances than their offspring-parent counterparts. In experiments involving mice and rats, on the other hand, $n$ is generally smaller than $n^*$ and offspring-parent estimators tend to be more efficient.

Another point of theoretical interest is to what extent the variance of the pooled estimator $E_p$ differs from those of $E_i$ and $E_s$. From equation (19), it is possible to find that

\[\text{Var}(E_i) = \text{Var}(E_p) \quad \text{if and only if} \quad \text{Var}(E_i) = \text{Cov}_{E_i,E_s} .\]

From equations (26) and (28) it is clear that this holds true for the half-sib estimator of genetic variance when $n$ and $s$ are large. It follows that asymptotically, $E_s$ tends to "dominate" in $E_p$ and, therefore, $E_s$ provides an excellent approximation to the best linear unbiased estimator of genetic variance developed in equations (18), (19), (20) and (21). Table 3 illustrates, for some selected "small sample" situations, the extent to which $E_p$ yields an estimator with smaller variance than either $E_i$ or $E_s$. In an experiment where $s=10$ and $n=20$, 

\[\begin{array}{cccc}
0.05 & 9.63 & 0.05 & 21.43 \\
0.10 & 10.31 & 0.60 & 24.02 \\
0.15 & 11.05 & 0.65 & 27.31 \\
0.20 & 11.86 & 0.70 & 31.68 \\
0.25 & 12.76 & 0.75 & 37.73 \\
0.30 & 13.75 & 0.80 & 46.80 \\
0.35 & 14.86 & 0.85 & 61.85 \\
0.40 & 16.14 & 0.90 & 91.90 \\
0.45 & 17.61 & 0.95 & 181.95 \\
0.50 & 19.34 & 1.00 & \infty \\
\end{array}\]
## TABLE 3

Comparison between the variance of offspring-parent (\(E_1\)), half-sib (\(E_h\)) and pooled (\(E_p\)) estimators of genetic variance for selected values of heritability (\(h^2\)), number of families (\(s\)) and family size (\(n\)).

<table>
<thead>
<tr>
<th>(h^2)</th>
<th>(V(E_1))</th>
<th>(V(E_h))</th>
<th>(S=10, n=20)</th>
<th>(V(E_p))/min (V(E_1))</th>
<th>(V(E_h))</th>
<th>(V(E_p))/min (V(E_1))</th>
<th>(S=20, n=10)</th>
<th>(V(E_p))/min (V(E_1))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>3.4</td>
<td>2.0</td>
<td>1.6</td>
<td>0.80</td>
<td>2.6</td>
<td>2.7</td>
<td>1.6</td>
<td>0.62</td>
</tr>
<tr>
<td>0.3</td>
<td>6.4</td>
<td>5.3</td>
<td>4.5</td>
<td>0.85</td>
<td>4.0</td>
<td>4.9</td>
<td>3.2</td>
<td>0.80</td>
</tr>
<tr>
<td>0.5</td>
<td>10.3</td>
<td>10.2</td>
<td>8.9</td>
<td>0.87</td>
<td>5.8</td>
<td>7.8</td>
<td>5.4</td>
<td>0.93</td>
</tr>
<tr>
<td>0.7</td>
<td>15.1</td>
<td>16.7</td>
<td>14.5</td>
<td>0.96</td>
<td>8.0</td>
<td>11.3</td>
<td>8.0</td>
<td>1.00</td>
</tr>
<tr>
<td>0.9</td>
<td>20.7</td>
<td>24.8</td>
<td>20.7</td>
<td>1.00</td>
<td>15.5</td>
<td>20.7</td>
<td>1.00</td>
<td>0.98</td>
</tr>
<tr>
<td>(s=40, n=80)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>4.1</td>
<td>1.1</td>
<td>1.1</td>
<td>1.00</td>
<td>2.6</td>
<td>1.0</td>
<td>0.9</td>
<td>0.90</td>
</tr>
<tr>
<td>0.3</td>
<td>11.2</td>
<td>6.2</td>
<td>6.1</td>
<td>0.98</td>
<td>6.1</td>
<td>3.9</td>
<td>3.7</td>
<td>0.95</td>
</tr>
<tr>
<td>0.5</td>
<td>20.4</td>
<td>15.2</td>
<td>15.0</td>
<td>0.99</td>
<td>10.6</td>
<td>8.7</td>
<td>8.4</td>
<td>0.97</td>
</tr>
<tr>
<td>0.7</td>
<td>31.6</td>
<td>28.2</td>
<td>27.8</td>
<td>0.99</td>
<td>16.1</td>
<td>15.5</td>
<td>14.8</td>
<td>0.95</td>
</tr>
<tr>
<td>0.9</td>
<td>44.8</td>
<td>45.2</td>
<td>44.2</td>
<td>0.99</td>
<td>22.6</td>
<td>24.2</td>
<td>22.6</td>
<td>0.93</td>
</tr>
</tbody>
</table>

* Entries should be multiplied by \(\sigma^4 \times 10^{-2}\).
† Entries should be multiplied by \(\sigma^4 \times 10^{-3}\).

## TABLE 4

Ratios between exact variances and their large sample approximations for selected values of heritability (\(h^2\)), number of families (\(s\)) and family size (\(n\)).

<table>
<thead>
<tr>
<th>(h^2)</th>
<th>(s)</th>
<th>(E_s)</th>
<th>(E_h)</th>
<th>(E_p)</th>
<th>(E_s)</th>
<th>(E_h)</th>
<th>(E_p)</th>
<th>(E_s)</th>
<th>(E_h)</th>
<th>(E_p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>10</td>
<td>5.4</td>
<td>5.1</td>
<td>28.6</td>
<td>16.8</td>
<td>4.0</td>
<td>3.7</td>
<td>14.9</td>
<td>10.6</td>
<td>3.3</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>5.1</td>
<td>4.7</td>
<td>26.4</td>
<td>15.5</td>
<td>3.7</td>
<td>3.4</td>
<td>13.8</td>
<td>9.8</td>
<td>3.0</td>
</tr>
<tr>
<td>0.50</td>
<td>10</td>
<td>1.9</td>
<td>1.6</td>
<td>3.3</td>
<td>2.3</td>
<td>1.6</td>
<td>1.5</td>
<td>2.4</td>
<td>2.0</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>1.7</td>
<td>1.5</td>
<td>3.0</td>
<td>2.1</td>
<td>1.5</td>
<td>1.3</td>
<td>2.2</td>
<td>1.8</td>
<td>1.4</td>
</tr>
<tr>
<td>0.90</td>
<td>10</td>
<td>1.5</td>
<td>1.3</td>
<td>2.0</td>
<td>1.4</td>
<td>1.4</td>
<td>1.2</td>
<td>1.7</td>
<td>1.3</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>1.4</td>
<td>1.2</td>
<td>1.9</td>
<td>1.3</td>
<td>1.3</td>
<td>1.1</td>
<td>1.6</td>
<td>1.2</td>
<td>1.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(h^2)</th>
<th>(s)</th>
<th>(E_s)</th>
<th>(E_h)</th>
<th>(E_p)</th>
<th>(E_s)</th>
<th>(E_h)</th>
<th>(E_p)</th>
<th>(E_s)</th>
<th>(E_h)</th>
<th>(E_p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>60</td>
<td>2.0</td>
<td>1.9</td>
<td>4.0</td>
<td>3.6</td>
<td>1.7</td>
<td>1.6</td>
<td>2.8</td>
<td>2.6</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>2.0</td>
<td>1.9</td>
<td>4.0</td>
<td>3.6</td>
<td>1.7</td>
<td>1.6</td>
<td>2.8</td>
<td>2.6</td>
<td>1.4</td>
</tr>
<tr>
<td>0.50</td>
<td>60</td>
<td>1.2</td>
<td>1.1</td>
<td>1.4</td>
<td>1.3</td>
<td>1.1</td>
<td>1.1</td>
<td>1.3</td>
<td>1.2</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>1.2</td>
<td>1.1</td>
<td>1.4</td>
<td>1.3</td>
<td>1.1</td>
<td>1.1</td>
<td>1.3</td>
<td>1.2</td>
<td>1.1</td>
</tr>
<tr>
<td>0.90</td>
<td>60</td>
<td>1.1</td>
<td>1.1</td>
<td>1.2</td>
<td>1.2</td>
<td>1.1</td>
<td>1.1</td>
<td>1.2</td>
<td>1.1</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>1.1</td>
<td>1.1</td>
<td>1.2</td>
<td>1.2</td>
<td>1.1</td>
<td>1.1</td>
<td>1.2</td>
<td>1.1</td>
<td>1.1</td>
</tr>
</tbody>
</table>

* Entries are values of equation (27)/equation (28).
† Entries are values of equation (23)/equation (24).
‡ Entries are values of equation (25)/equation (26).
§ Ratio of the exact variance of the pooled estimator to its large sample approximation.
or vice versa, the variance of $E_p$ is about 10 to 40% smaller than the variance of $E_1$ or $E_2$ (whichever was minimum), except at heritability levels above 0.60, where there would be little advantage in using $E_p$. In a larger experiment ($s=40$, $n=80$, or vice versa), the variance reduction effect of $E_p$ would be at most of 10%.

Adequacy of large sample approximations

It is of interest to examine the consequences of using the large sample approximations (equations 24, 26 and 28) in "small" data sets. This is illustrated in Table 4 for selected values of $s$, $n$ and $h^2$. For all experimental circumstances illustrated in Table 4, large sample approximations considerably underestimate the exact values of the covariance ($E_o$) between the offspring-parent and half-sib estimators of genetic variance, the variance of the offspring-parent estimator $[\text{Var}(E_o)]$, the variance of the half-sib estimator $[\text{Var}(E_s)]$ and the variance of the pooled estimator $[\text{Var}(E_p)]$. The extent of the underestimation is, of course, inversely related to $n$ and $s$ and is much more drastic at lower heritability levels. For example, when $h^2=0.10$, $s=10$ and $n=10$, the large sample approximation of $\text{Var}(E_o)$ is only about 1/17 of its exact value. On the other hand, when $h^2=0.90$, $s=10$, $n=10$, the large sample approximation is about 13/17 of its exact value. Also, $\text{Var}(E_s)$ appeared to be less affected by departures from large sample assumptions than $\text{Var}(E_o)$ or $\text{Var}(E_p)$.

It should be concluded that large sample approximations of the variances of estimators and of the covariance between the two estimators and of the covariance between the two estimators of genetic variance should not be used in most designed animal breeding experiments involving a few hundreds or thousands of observations. On the other hand, in data sets from field records, *e.g.*, dairy

### TABLE 5

*Effect of alternative allocations of number of families ($s$) and family size ($n$) on the variance of a pooled estimator of genetic variance ($E_p$) for $N=ns=1000$*  

<table>
<thead>
<tr>
<th>$s$</th>
<th>$n$</th>
<th>$h^2=0.1$</th>
<th>$h^2=0.5$</th>
<th>$h^2=0.9$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>500</td>
<td>23.21</td>
<td>513.90</td>
<td>1641.09</td>
</tr>
<tr>
<td>4</td>
<td>250</td>
<td>8.87</td>
<td>175.87</td>
<td>553.32</td>
</tr>
<tr>
<td>5</td>
<td>200</td>
<td>7.09</td>
<td>133.60</td>
<td>417.17</td>
</tr>
<tr>
<td>8</td>
<td>125</td>
<td>4.84</td>
<td>79.21</td>
<td>241.77</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>4.20</td>
<td>63.06</td>
<td>189.60</td>
</tr>
<tr>
<td>20</td>
<td>50</td>
<td>3.12</td>
<td>33.19</td>
<td>92.73</td>
</tr>
<tr>
<td>25</td>
<td>40</td>
<td>2.96</td>
<td>27.52</td>
<td>74.30</td>
</tr>
<tr>
<td>40</td>
<td>25</td>
<td>2.83</td>
<td>19.10</td>
<td>47.02</td>
</tr>
<tr>
<td>50</td>
<td>20</td>
<td>2.84</td>
<td>16.27</td>
<td>37.98</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>3.05</td>
<td>10.39</td>
<td>19.99</td>
</tr>
<tr>
<td>125</td>
<td>8</td>
<td>3.16</td>
<td>9.13</td>
<td>16.43</td>
</tr>
<tr>
<td>200</td>
<td>5</td>
<td>3.40</td>
<td>7.14</td>
<td>11.21</td>
</tr>
<tr>
<td>250</td>
<td>4</td>
<td>3.51</td>
<td>6.45</td>
<td>9.52</td>
</tr>
<tr>
<td>500</td>
<td>2</td>
<td>3.81</td>
<td>5.00</td>
<td>6.30</td>
</tr>
</tbody>
</table>

* Entries should be multiplied by $\sigma^4 \times 10^{-3}$. 
cattle data with hundreds of sires and thousands of lactations, large-sample formulae would be expected to approximate reasonably well the exact values.

Allocation of experimental resources

Finally, it is possible, although tedious, to obtain values of \( n \) and \( s \) that minimize the variance of \( E_p \) subject to a number of constraints, i.e., fixed number of experimental units, fixed cost functions, etc. We do not address this specifically in the present paper, but illustrate the principles by means of an example in which there is a constraint in the number of progeny that can be scored, \( N = ns \).

Suppose we wish to know the optimal combination of \( n \) and \( s \) that minimizes \( \text{Var}(E_p) \) for \( N = 1000 \). All possible combinations of feasible values of \( n \) and \( s \) that satisfy the previous condition and the \( \text{Var}(E_p) \) at each of three heritability levels are presented in Table 5. When \( h^2 = 0.1 \), the optimal allocation would be \( s = 40 \) and \( n = 25 \), but when \( h^2 = 0.5 \) or 0.9, \( \text{Var}(E_p) \) is minimized by maximizing the number of half-sib families, that is when \( s = 500 \) and \( n = 2 \). In small samples, it is difficult to interpret the outcome of the optimal allocation since \( \text{Var}(E_p) \) depends on \( \text{Var}(E_1) \), \( \text{Var}(E_2) \) and their covariance, and the relative importance of these changes (within a fixed \( N \)) with \( h^2 \), \( n \) and \( s \). On the other hand, since in large samples the variance of \( E_2 \) approximates well the variance of \( E_p \), optimal allocations would be attained by maximizing the number of families (see equation 26).

LITERATURE CITED


Thompson, R., 1976 Design of experiments to estimate heritability when observations are available on parents and offspring. Biometrics 32: 283–304.


Corresponding editor: J. F. Kidwell