Ann. Hum. Genet., Lond. (1980), **43**, 265–284
Printed in Great Britain

Estimating genetic correlations

By CEDRIC A. B. SMITH
Galton Laboratory, University College, London

1. INTRODUCTION

It is approximately 100 years (1872) since Francis Galton first thought of the idea of 'reversion towards ancestral type' in connexion with the genetics of the sizes of sweet pea seeds, and explained this idea later in a paper read to the Royal Institution (Galton, 1877). In his Presidential Address to the Anthropological Section of the British Association (Galton, 1885*a*) he applied the idea to human stature, but then called it 'regression' towards the mean. In his autobiography (Galton, 1908) he tells that he was worried for a time by the fact that if children appeared to regress towards the mean of their parents, it was also statistically true that parents regressed towards the mean of their children. However, inspiration eventually came, in that Galton discovered the bivariate normal distribution (with assistance from a Cambridge mathematician, Hamilton Dickson), and saw that this apparently paradoxical result was a necessary consequence of the form of the distribution. In 1885 he was considering the further problem of expressing in the most appropriate manner the resemblances between characters, such as stature, cubit, head length, middle-finger length, etc., which had different ranges of variability. He saw that this could be done by first standardizing each character in terms of its 'probable errors', and then finding the regression of one on the other. In essence, this is the modern idea of what Galton called 'co-relation, or correlation of structure' (Galton, 1885*b*), (although we would now use standard errors rather than probable errors), and the first co-relations were those calculated between various anthropometric measures.

It was immediately evident to Karl Pearson (see Pearson, 1930) that Galton had produced a measure of resemblance between characters which did not depend on any specific assumption of causation, and thereby extended the applicability of statistical method in science, and Pearson (1895, 1896) refined the measure into the modern correlation coefficient. However, as Galton considered co-relation as a particular case of regression, he denoted it by the letter r, as the initial letter of 'reversion' and 'regression', and this has continued as the standard custom ever since (except in so far as Fisher (1925) changed it to the corresponding Greek letter ρ when considering the population value rather than the sample estimate). Further historical details of these matters are to be found in Forrest (1974) and Pearson (1930).

The ideas of regression, correlation, and multivariate normality have therefore arisen very naturally from the study of genetics [and, incidentally, largely from human genetics, though this may be a historical accident (see, for example, Pearson & Lee, 1903)] and remain major elements of modern quantitative genetics. Since the time of Galton there have been many advances in the theory of genetic correlation. The most notable ones use Fisher's (1918) classic paper, showing how correlations are influenced by dominance, epistasis, linkage and assortative mating and those of Wright (1921, 1922, 1923) introducing path coefficients, of Fisher (1925) showing the relationship between intraclass correlation and components of variance, and of Li & Sacks (1954) greatly simplifying the theoretical background.

Unfortunately the practical problem of estimating correlations efficiently has advanced less rapidly than the theory. In its original definition, correlation relates to pairs (x, y) of measured characters. For example, x = stature and y = arm length of an individual. Each individual then gives one pair (x, y) of values, and in any sample the obvious procedure is to estimate in the standard way the respective variances v_{xx}, v_{xy} of x and y, the covariance v_{xy}, and then find the regression $b = v_{xy}/v_{xx}$ or correlation $r = v_{xy}/\surd(v_{xx}v_{xy})$. In a monogamous society we can similarly estimate, for example, the correlation between x = husband's stature and y = wife's stature. But when we come to estimate mother–daughter correlation we have the complication of varying family size. In a family with, say, 6 daughters, we can form 6 (x, y) pairs, where in each pair x = the mother's stature and y = a different daughter's stature. But then in this case the same value of x will occur in 6 different pairs, whereas in a family with only 2 daughters the mother's stature will occur as x in only 2 pairs (x, y). It is not clear that it is 'fair' that one value of x should be counted 6 times over. More to the point, it is questionable whether this gives the most accurate estimate of the correlation. The situation becomes even more dramatic with an intraclass correlation, such as that between sisters. We can form all possible pairs of values (x, y), where x = one sister's stature and y = another sister's stature, and calculate the correlation between all such pairs by the standard procedure. But a family of 6 sisters then gives $6 \times 5 = 30$ such pairs, and a family of 2 gives only 2 pairs. It is evident to common sense, even without detailed mathematical analysis, that a family of 6 sisters does not give anything like 15 times the information given by a family of 2.

This kind of problem has long been known. Estimation of intraclass correlation is effectively the same problem as estimation of variance components, as noticed by Fisher (1925), and this gives an alternative approach. An estimate of variance components, using analysis of variance, was given by Brownlee (1949), and a more thorough investigation of the problem was undertaken by Fieller & Smith (1951).

An obvious approach to the problem is to consider the special situation in which all observed characters follow a joint multivariate normal distribution, in which the correlation between any two depends only on the degree of relationship. Thus if (x, y) denote a pair of values in 2 sisters, we suppose that the correlation ρ_{SS} between them has the same value, independently of which pair of sisters we take. A similar remark is assumed to apply to ρ_{MO} = mother–daughter correlation, and for other degrees of genetic relationship. Using the standard formula for a multivariate normal distribution, the likelihood of any observed sample can then be expressed as a function of the means, variances, and correlations (or equivalently, in terms of variance and covariance components). In principle, efficient estimation of these parameters is then quite straightforward, using the method of maximum likelihood. In practice, the computations become very heavy. They have been studied in various papers, including especially Hartley & Rao (1957), Cunningham & Henderson (1968), Nelder (1968), Lindley & Smith (1972), Rao (1971*a*, *b*, 1972), Patterson & Thompson (1971, 1974), Hemmerle & Hartley (1973), Corbeil & Searle (1976), Thompson (1977), Dempster, Laird & Rubin (1977), Schaeffer, Wilton and Thompson (1978). A good general account and bibliography is given by Harville (1977). However, a geneticist wishing to make use of these papers to analyse actual data will not find them too easy going.

It is the purpose of this paper to present a rather simpler method of tackling the practical problem of estimating correlations and components of variance, mostly using the normal model described above. It is certainly not claimed that this method is entirely new, indeed, several remarks made below will be found also in the papers quoted above. The computations may not be as rapid as in the other methods. However, the scheme presented here is a unified one, covering most cases arising in practice, and using only straightforward rules of procedure. (Even so, except in very simple cases, it is reasonably practicable only with the use of a computer.)

One further remark seems in order. Various devices are used by statisticians to distinguish between population and sample values, e.g. Fisher's use of Greek and Latin letters, ρ = population correlation, r = sample estimate, or Fisher's use of a circumflex, $\hat{\rho}$ = maximum likelihood estimate of ρ. The present paper is on estimation, hence usually a symbol will denote an estimate. To use a circumflex consistently would be logically correct, but typographically rather a burden and probably not too helpful to the reader. Instead, I will in most cases rely on context to make the distinction. For example, the symbol 'B' will be used to denote (i) the population within-family component of variance, (ii) a sample estimate of this, or (iii) the corresponding dummy variable in the likelihood function. In principle, this is a serious breach of good behaviour by using one symbol with many meanings; in practice I hope that no confusion will arise here (no more than it does in everyday language, where context is an essential clue to meaning).

2. THE MATHEMATICAL MODEL

Consider first the situation in which we have one measured character x (e.g. stature), and we wish to find a sister–sister correlation.

In a family F_f we will find a 'sistership' of n_f sisters, numbered from 1 to n_f. The observed value of sister number i in family F_f will be denoted by x_{fi}.

We make the following assumptions:

(α) The population under study is large enough to be treated as infinite.

(β) In any one given family F_f, the observed values x_{fi} of the n_f sisters are n_f random values independently drawn from some distribution, whose mean we denote by μ_f and whose variance by v_f. Thus μ_f is the 'true mean of sistership f' and v_f the 'true variance within sistership f'. This assumption is quite often plausible, but it could break down if there is a maternal age effect, or if different sisters were brought up in different environments, e.g. by a change with time in the economic status of the family.

(γ) The distribution within a sistership is statistically independent of its size n_f; that is, large families do not produce different distributions from small ones. In particular, μ_f and v_f are independent of n_f. This assumption is again often plausible, but not invariably so.

(δ) The sisterships themselves, number k in all, can be considered as a random independent selection of k sisterships from the whole population. Thus the μ_f will be a selection of k independent values from some distribution, with mean denoted by μ (= the general mean of the population) and variance B (= the 'component of variance between sisterships'). The v_f will also have a distribution. Its mean, A, is the (average) component of variance within sisterships.

Besides this, the distribution of the v_f will have some variance, and there will be a covariance between the μ_f and v_f. Most investigators in population genetics have ignored the variance of the v_f, the covariance of the μ_f and v_f, and moments of higher orders, such as skewness and kurtosis, and we will do so here. But note that ignoring these could be ignoring important genetic information, especially should a character be influenced mainly by few alleles at few loci.

(ϵ) In the normal model, already referred to, all distributions are supposed (multivariate) normal, and all the variances v_f within families are equal, having common value A. Hence the variance of v_f and covariance of μ_f with v_f are both zero, and higher moments are irrelevant, being determined by the means, variances and covariances. This model will apply if the character is determined additively by genotypes at a large number of loci. It is conceivable that several quantitative characters, such as stature, approximate reasonably well to the normal model.

Consider the value x_{fi} of the ith sister in family f. Write $\delta_{fi} = x_{fi} - \mu_f$, deviation from the true mean μ_f of that sistership. Within the sistership, the expectation

$$\mathscr{E}(\delta_{fi} \mid f) = 0 \tag{2.1}$$

by definition, hence (since μ_f is a constant when f is given) $\mathscr{E}(\mu_f \delta_{fi} \mid f) = 0$. Averaging over all families, we have by the relation $\mathscr{E}(x) = \mathscr{E}_f(\mathscr{E}(x \mid f))$,

$$\mathscr{E}(\delta_{fi}) = \mathscr{E}(\mu_f \delta_{fi}) = 0, \tag{2.2}$$

(here these are expected values in the whole population) and hence

$$\operatorname{cov}(\mu_f, \delta_{fi}) = \mathscr{E}(\mu_f \delta_{fi}) - \mathscr{E}(\mu_f)\mathscr{E}(\delta_{fi}) = 0. \tag{2.3}$$

Also,

$$\operatorname{var}(\mu_f) = B \text{ (by definition)}, \tag{2.4}$$

$$\begin{aligned} \operatorname{var}(\delta_{fi}) &= \mathscr{E}(\delta_{fi}^2) \\ &= \mathscr{E}_f(\mathscr{E}(\delta_{fi}^2 \mid f)) \\ &= \mathscr{E}_f(\operatorname{var}\delta_{fi}) \quad \text{(using (2.1))} \\ &= \mathscr{E}_f(v_f) \\ &= A \quad \text{(by definition).} \end{aligned} \tag{2.5}$$

Since by definition

$$x_{fi} = \mu_f + \delta_{fi}, \tag{2.6}$$

we have, using (3), (4) and (5), that the population variance is

$$\operatorname{var}(x_{fi}) = \operatorname{var}(\mu_f) + \operatorname{var}(\delta_{fi}) = A + B.$$

Furthermore, if x_{fi}, x_{fj} $(i \neq j)$ are the values of the sisters, their covariance is, on using (2.3),

$$\begin{aligned} \operatorname{cov}(x_{fi}, x_{fj}) &= \operatorname{cov}(\mu_f + \delta_{fj}, \mu_f + \delta_{fj}) \\ &= \operatorname{cov}(\mu_f, \mu_f) + \operatorname{cov}(\delta_{fi}, \delta_{fj}). \end{aligned}$$

But $\operatorname{cov}(\mu_f, \mu_f) = \operatorname{var}\mu_f = B$, by definition. And using (2.2)

$$\begin{aligned} \operatorname{cov}(\delta_{fi}, \delta_{fj}) &= \mathscr{E}(\delta_{fi}, \delta_{fj}) \\ &= \mathscr{E}_f \mathscr{E}(\delta_{fi}, \delta_{fj} \mid f) \\ &= 0, \end{aligned} \tag{2.7}$$

since, given f, δ_{fi} and δ_{fj} are independent random variables with mean 0. Hence the covariance of sisters is

$$\operatorname{cov}(x_{fi}, x_{fj}) = B. \tag{2.8}$$

The correlation between sisters is thus

$$\rho_{SS} = \frac{\operatorname{cov}(x_{fi}, x_{fj})}{\sqrt{[\operatorname{var}(x_{fi}) \operatorname{var}(x_{fj})]}} = \frac{B}{A+B}. \tag{2.9}$$

Notice that we do *not* assume here either that the distributions are normal or that the variances

v_f are the same within the different sibships. In fact, we only need assumptions (α), (β) and (δ) above. Thus the finding of the correlation ρ_{SS} is effectively equivalent to finding the components of variance, A within families and B between families.

Since the theoretical treatment by Fisher (1918), Li & Sacks (1954) and others directly relates to the correlations, it would seem that it is natural and appropriate to estimate correlations in this way, at least in the first place, before going on to consider other aspects such as the variance of the v_f or other higher moments.

3. WITHIN FAMILIES VARIANCE A

As is well known, the obvious estimate of A is the 'within families mean square' in the analysis of variance. Within each family f we calculate a total $T_f = \Sigma_i x_{fi}$, sample mean $\bar{x}_f = T_f/n_f$, sum of squares of deviations, or 'deviance',

$$\Delta_f = \Sigma_i (x_{fi} - \bar{x}_f)^2,$$

and degrees of freedom $\nu_f = n_f - 1$. We then sum over families to get the total deviance and total degrees of freedom

$$\Delta = \sum_f \Delta_f; \quad \nu = \sum_f \nu_f. \tag{3.1}$$

The mean square within is then by definition

$$msq_A = \Delta/\nu, \tag{3.2}$$

and provided that assumption (γ) holds, as well as (α), (β), (δ), it is a consistent estimate of A.

One at-first-sight pedantic comment is in order. If there are no sisters in a family, i.e. $n_f = 0$, the formulas given above break down. The obvious course is to ignore such families, i.e. set $\Delta_f = 0$, $\nu_f = 0$ (instead of $\nu_f = n_f - 1 = -1$). This is trivial when we are considering only sisterships, since families with no sisters would then not be recorded. But when we come to sister–brother correlations there will be recorded families with some brothers but no sisters (or vice versa) and they *must* be counted (especially in a computer program) as giving zero deviance and zero degrees of freedom where there are no observed values.

4. POPULATION MEAN μ

The obvious estimate for the population mean μ is a mean of the separate observed family means $\bar{x}_f$. However, with varying family size, the $\bar{x}_f$ will not be of equal accuracy, and hence it seems preferable to use a weighted mean. If we give to $\bar{x}_f$ some (non-negative) weight ω_f, we would have the estimated population mean

$$\mu = (\Sigma\, \omega_f \bar{x}_f)/(\Sigma\, \omega_f). \tag{4.1}$$

(Note: since μ here is an estimate, and not the population value, we should according to the usual convention distinguish it as $\hat{\mu}$, or m, or in some such way. But, as already explained, we here rely on context to make the distinction.)

The question arises of what values to take for the weights ω_f. If the family f is empty, $n_f = 0$, we *must* give it zero weight, $\omega_f = 0$; the mean $\bar{x}_f$ is then indeterminate, but can conveniently also be set equal to 0. Otherwise, if we choose any (fixed) positive values for ω_f, we will get an estimate

μ of the population mean which is consistent, and unbiased (in the sense that its expected value is the true population mean). It is well known that the most accurate weighted mean is that in which the weight ω_f is chosen as the reciprocal of the sampling variance of $\bar{x}_f$, easily found to be

$$\operatorname{var}(\bar{x}_f) = v_f/n_f + B. \tag{4.2}$$

We discuss later the choice of a suitable value of B. The estimation of v_f in this formula presents more serious problems. If the situation is such that we can plausibly assume that the true variances v_f within sibships are all equal, with common value A, then we will have

$$\operatorname{var}(\bar{x}_f) = A/n_f + B, \tag{4.3}$$

and hence the appropriate 'semiweight' ω_f is

$$\omega_f = (A/n_f + B)^{-1}. \tag{4.4}$$

We have already found an estimate (msq_A) for A. This weighting is due to Cochran (1954) and independently to Lee Crump (see Smith, 1957). Unfortunately, if v_f varies from family to family, we cannot reliably estimate it in the usual genetic situation as Δ_f/v_f, the variance within the individual family f, because families are too small for such an estimate to be reliable (but the situation could be different in organisms with large families). It seems a reasonable compromise to use the formula (4.4) in any case for the weight ω_f, even though with varying ω_f it is not strictly optimal.

5. VARIANCE COMPONENT BETWEEN CLASSES, B

Smith (1957) suggested the use of a 'weighted deviance between classes' as a means of estimating B. With some modification, the argument proceeds as follows. Let

$$\xi_f = \bar{x}_f - \mu \tag{5.1}$$

be the deviation of the sample mean $\bar{x}_f$ of family f from the *estimated* population mean μ. Assign to each family a weight w_f (zero for empty sisterships). Form the 'weighted deviance'

$$\Delta_B = \Sigma w_f \xi_f^2. \tag{5.2}$$

In a *large* sample the variance of the estimated μ is small. Since ξ_r has zero expectation, we have approximately

$$\begin{aligned} \mathscr{E}(\xi_f^2) &= \mathscr{E}(\operatorname{var}\xi_f) \\ &\simeq \mathscr{E}(\operatorname{var}\bar{x}_f) \\ &= \mathscr{E}(v_f/n_f + B) \\ &= A/n_f + B, \end{aligned} \tag{5.3}$$

and hence

$$\mathscr{E}(\Delta_B) \simeq A\,\Sigma(w_f n_f^{-1}) + B\Sigma w_f. \tag{5.4}$$

More exactly, setting $W = \Sigma w_f$ and k = no. of sisterships,

$$\mathscr{E}(\Delta_B) = A\Sigma[n_f^{-1}(1 - w_f/W)^2] + B[k - 2 + k\Sigma w_f^2/W^2], \tag{5.5}$$

but in large samples the difference between this and (5.4) is usually not important. Smith (1957) suggested that this led to a method of estimating μ and B. We replace the expectation $\mathscr{E}(\Delta_B)$

on the left-hand side of (5.4) or (5.5) by its actual value Δ_B, which will not be greatly different, getting an equation

$$\Delta_B = A\,\Sigma(w_f n_f^{-1}) + B\Sigma w_f. \tag{5.6}$$

Smith (1957) also suggested (in the paragraph following equation (30)) that it would be optimal to choose the w_f to be the same semi-weights ω_f as defined in (4.4) for estimating the mean μ. But as Alan Robertson pointed out (in a personal communication) the argument for optimality is fallacious. However,this does not prevent us setting $w_f = \omega_f$ if we so wish, and it is not an unreasonable choice. Assuming that we have already estimated A, the equations (4.1), (4.4), (5.1), (5.2) and (5.6) together can then be solved in principle to give the values of ω_f, u, ξ_f, Δ_B and B. Unfortunately, these equations are troublesome to solve directly. But their solution can be found iteratively as follows. Begin by guessing some provisional value for B; e.g. $B = A$ will usually suffice. (**). Use (4.4) to find ω_f, (5.1) to find ξ_f, (5.2) to find Δ_B (with $w_f = \omega_f$), and then solve (5.6) to get a new and improved estimate of B. Go back to (**) and repeat the procedure with this new value of B as 'provisional' value, to get a further improved value, and so on iteratively until the provisional and improved values agree sufficiently well. For further details see Smith (1957).

In essence, what we are doing here is as follows. We consider some expression, F, involving observed measurements (and possibly also parameters), and find its expected value, E, which will involve the unknown parameters. We then set $F = E$, providing us with one equation connecting the parameters. By repeating this device if necessary, we end up with as many equations as there are parameters to estimate, and then solve these. We might reasonably expect that this would lead to consistent and nearly unbiased estimators.

However, there is one complication. In the equation

$$\Delta_B = A\,\Sigma(w_f n_f^{-1}) + B\,\Sigma(w_f) \tag{5.7}$$

the right-hand side (E) is the expectation of the left-hand side (F) when the weights w_f are held constant. But in the semi-weighted method we allow the values of the $w_f = \omega_f$ to depend on the parameters A and B being estimated, so this is no longer quite obvious. To deal with this kind of situation in general terms, suppose we have h unknown parameters, θ_1 to θ_h, forming a (column) vector $\boldsymbol{\theta}$. Let $\mathbf{x}$ denote the measurements in the sample (of size n), and $\mathbf{w}$ a set of quantities, for the moment arbitrary, used in the estimation procedure. Choose a set of h functions $F_i(\mathbf{x}, \mathbf{w}, \boldsymbol{\theta})$, forming a vector $\mathbf{F}(\mathbf{x}, \mathbf{w}, \boldsymbol{\theta})$, and let $\mathbf{E}(\mathbf{w}, \boldsymbol{\theta})$ be the expectation of $\mathbf{F}(\mathbf{x}, \mathbf{w}, \boldsymbol{\theta})$ for fixed $\mathbf{w}$ and fixed $\boldsymbol{\theta}$. We will not attempt to present the argument in completely rigorous form, which could lead to considerable complication, but we will note that it is important that all functions used should be expansible in Taylor series up to at least the quadratic terms. But that will almost always hold in practical applications.

We will denote the actual population value of $\boldsymbol{\theta}$ by $\boldsymbol{\theta}_0$, and the expected value of $\mathbf{x}$ (when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$) by $\mathbf{x}_0$. Thus $\delta\mathbf{x} = \mathbf{x} - \mathbf{x}_0$ will usually be of order of magnitude $n^{-\frac{1}{2}}$. (The 'usually' means that, here and subsequently, we suppose excluded from the argument 'exceptional' samples with large deviations from expectation, which occur only with low probability.) Let $\mathbf{w}_0$ be some value, to be defined more precisely later, such that $\delta\mathbf{w} = \mathbf{w} - \mathbf{w}_0$ is 'usually' anticipated to be of order of magnitude $n^{-\frac{1}{2}}$, and let $\boldsymbol{\theta}$ be such that $\delta\boldsymbol{\theta} = \boldsymbol{\theta} - \boldsymbol{\theta}_0$ is also $O(n^{-\frac{1}{2}})$. Then, neglecting quantities of order n^{-1}, we have

$$\mathbf{F}(\mathbf{x}, \mathbf{w}, \boldsymbol{\theta}) = \mathbf{F}(\mathbf{x}_0, \mathbf{w}_0, \boldsymbol{\theta}_0) + \mathbf{a}\delta\mathbf{x} + \mathbf{b}\delta\mathbf{w} + \mathbf{c}\delta\boldsymbol{\theta}, \tag{5.8}$$

where $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ are matrices of partial derivatives. The matrix $\mathbf{c}$ is square, and will be assumed non-singular. Let $\mathbf{E}(\mathbf{w}, \boldsymbol{\theta})$ denote the expected value of $\mathbf{F}(\mathbf{x}, \mathbf{w}, \boldsymbol{\theta})$ calculated on the assumption that $\mathbf{w}$ is fixed and that the true value of $\boldsymbol{\theta}$ is in fact $\boldsymbol{\theta}_0$, i.e. $\delta\boldsymbol{\theta} = \mathbf{0}$. Note that we are not asserting either that $\mathbf{w}$ is fixed or that $\boldsymbol{\theta} = \boldsymbol{\theta}_0$; we are merely saying that $\mathbf{E}(\mathbf{w}, \boldsymbol{\theta})$ is a function of $\mathbf{w}$ and $\boldsymbol{\theta}$ obtained by taking expectations as if these assumptions were true. From (5.8) we therefore have

$$\mathbf{E}(\mathbf{w}, \boldsymbol{\theta}) = \mathbf{F}(\mathbf{x}_0, \mathbf{w}_0, \boldsymbol{\theta}_0) + \mathbf{b}\delta\mathbf{w}. \tag{5.9}$$

If, therefore we take our estimation equations to be

$$\mathbf{F}(\mathbf{x}, \mathbf{w}, \boldsymbol{\theta}) = \mathbf{E}(\mathbf{w}, \boldsymbol{\theta}), \tag{5.10}$$

then, from (5.8) and (5.9), these become $\mathbf{a}\delta\mathbf{x} + \mathbf{c}\delta\boldsymbol{\theta} = \mathbf{0}$, or

$$\boldsymbol{\theta} = \boldsymbol{\theta}_0 + \delta\boldsymbol{\theta} = \boldsymbol{\theta}_0 - \mathbf{c}^{-1}\mathbf{a}\delta\mathbf{x}.$$

Since $\mathscr{E}(\delta\mathbf{x}) = \mathbf{0}$, we have $\mathscr{E}(\boldsymbol{\theta}) = \boldsymbol{\theta}_0$, i.e. $\boldsymbol{\theta}$ is an unbiased estimator, on neglecting terms of order n^{-1}. In deriving this we have assumed that $\delta\mathbf{w}$ is of order $n^{-\frac{1}{2}}$. This is certainly true if $\mathbf{w}$ is fixed. But it is also true if we set $\mathbf{w}$ = some function $\boldsymbol{\omega}(\mathbf{x}, \boldsymbol{\theta})$ of $\mathbf{x}$ and $\boldsymbol{\theta}$, and correspondingly

$$\mathbf{w}_0 = \boldsymbol{\omega}(\mathbf{x}_0, \boldsymbol{\theta}_0).$$

This guarantees that estimates obtained in this way, including virtually all the estimates considered in this paper, will be unbiased when quantities of order n^{-1} are neglected, and, if necessary, rare 'exceptional' samples excluded. We call this the 'unbiasedness principle'.

The fact that the maximum-likelihood equations for estimating components of variance are of this type has been noticed by various authors, but they do not all seem to have drawn the conclusion that the equations are therefore robust, in the sense of giving nearly *unbiased (though* inefficient) estimates even when the distributions are not normal.

6. MAXIMUM-LIKELIHOOD ESTIMATION OF NORMAL INTRACLASS CORRELATION

The method used above of estimating the within family component of variance A as the mean square within families is easily seen to be less than optimal, even when all within family true variances v_f are equal to A. For the family mean $\bar{x}_f$ has variance $(An_f^{-1} + B) = u_f$ (say) and this implies that (when the n_f vary from family to family) the values of the $\bar{x}_f$ by themselves would give information about the values of both B and A, even if we were not told anything about variation within families. An optimal method of estimation should take this into account.

The 'normal' model, with all within family variances v_f equal to A and all distributions normal, will often approximate closely to real life. It has the advantage that estimates can be found using maximum likelihood. In fact, the method turns out to be not much more complicated than the semi-weighted method.

In a large sample $\xi_f = \bar{x}_f - u$ has expected mean 0 and variance nearly equal to $u_f = A_f^{-1}n + B$, ignoring a small correction due to the variability of μ. Hence

$$\mathscr{E}(\xi_f^2) \simeq An_f^{-1} + B = u_f = \omega_f^{-1}. \tag{6.1}$$

Consider the weighted sums

$$S_A = \Sigma\,\omega_f^2 n_f^{-1} \xi_f^2 + \alpha^2\Delta,$$
$$S_B = \Sigma\,\omega_f^2 \xi_f^2, \tag{6.2}$$

where for the moment α is a constant. Then we have, nearly enough,

$$\left.\begin{aligned} \mathscr{E}(S_A) &= A(\Sigma\,\omega_f^2 n_f^{-2} + \alpha^2\nu) + B\,\Sigma\,\omega_f^2 n_f^{-1} \\ \mathscr{E}(S_B) &= A\,\Sigma\,\omega_f^2 n_f^{-1} + \qquad B\,\Sigma\,\omega_f^2 \end{aligned}\right\}. \tag{6.3}$$

If we replace the expectations on the left-hand side by the actual values of S_A, S_B, this gives two linear equations to solve for the values of A, B: we can write them as

$$\mathbf{S} = \begin{bmatrix} S_A \\ S_B \end{bmatrix} = \mathbf{T}\begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} \Sigma\,\omega_f^2 n_f^{-2} + \alpha^2\nu & \Sigma\,\omega_f^2 n_f^{-1} \\ \Sigma\,\omega_f^2 n_f^{-1} & \Sigma\,\omega_f^2 \end{bmatrix}\begin{bmatrix} A \\ B \end{bmatrix}, \tag{6.4}$$

where $\mathbf{T}$ is a 2×2 symmetric matrix. If $\mathbf{T}$ was constant, it would follow that the solutions

$$\begin{bmatrix} A \\ B \end{bmatrix} = \mathbf{T}^{-1}\mathbf{S} = \mathbf{tS} \tag{6.5}$$

say, where $\mathbf{t} = \mathbf{T}^{-1}$, would be consistent unbiased estimates of the true values of A, B. However, this is not quite true, because ω_f depends on A and B, and we will now put

$$\alpha = A^{-1}, \tag{6.6}$$

also dependent on A. But the equations (4.1), (4.4), (5.1), (6.2), (6.5) and (6.6) give a set of simultaneous equations which we can solve to find μ, A, B, ω_f and ξ_f. This is not difficult to do iteratively. We first begin with 'provisional' values for A, B, for these we could conveniently take A = the mean square within families and $B = A$. (**). Then we set $\alpha = A^{-1}$, find ω_f from (4.4), μ from (4.1), ξ_f from (5.1), S_A, S_B from (6.2), $\mathbf{T}$ from (6.4), $\mathbf{T}^{-1} = \mathbf{t}$ by matrix inversion and new improved values of A, B from (6.5). We now take these improved values as preliminary values, repeat from (**), getting still better values, and so on iteratively until the provisional and improved values agree sufficiently closely. The final values can then be shown to be the maximum likelihood estimates under the normal model. In addition, the error variance of μ is Ω^{-1}, where $\Omega = \Sigma\,\omega_f$, and hence its standard error is $\Omega^{-\frac{1}{2}}$. The error variance matrix of the vector $\begin{bmatrix} A \\ B \end{bmatrix}$ is

$$\operatorname{var}\begin{bmatrix} A \\ B \end{bmatrix} = 2\mathbf{t}. \tag{6.7}$$

Hence the standard errors of A, B are respectively $\sqrt{(2t_{11})}$ and $\sqrt{(2t_{22})}$.

Because of the way in which these estimates have been found, the 'unbiasedness principle' shows that they are robust in the sense that μ, A, B will be consistent and nearly unbiased estimates of the corresponding population values as long as the assumptions (α), (β), (γ), (δ) of Section 2 hold good, even if the distributions are not normal. They will no longer be optimal estimates, but they can usually be expected to be reasonable ones. The formulas for the standard errors given above will only apply to the normal case, but they can be expected to indicate roughly the values of the S.E.'s in other cases.

7. EXAMPLES OF ESTIMATION OF INTRACLASS CORRELATION

For illustrative purposes, a small sample of 14 families has been chosen from a large mass of Polish dermatoglyphic family data, kindly made available by Dr D. Z. Loesch. The families consisted of sisters and brothers, together with their fathers and mothers. (We could have used the

Table 1. *Dermatoglyphic family data: values of pattern intensity on fingers*

Fam. no.	(1) Sisters' values				(2) Brothers' values				(3)	(4)
f		n_{f1}	$\bar{x}_{f1}$	Δ_{f11}		n_{f2}	$\bar{x}_{f2}$	Δ_{f22}	Mother	Father
1	2	1	2·0	0	2	1	2·0	0	2	3
2	—	0	—	—	2, 3	2	2·5	0·5	2	3
3	2, 2	2	2·0	0·0	2	1	2·0	0	2	3
4	2, 2, 2, 2	4	2·0	0·0	2	1	2·0	0	2	4
5	—	0	—	—	6, 6	2	6·0	0·0	6	7
6	4	1	4·0	0	3, 3	2	3·0	0·0	4	3
7	2, 2, 3	3	2·33	0·67	6, 3, 5, 4	4	4·5	5·0	4	3
8	2, 4, 7, 4, 4	5	4·2	12·8	7, 8	2	7·5	0·5	3	7
9	5	1	5·0	0	6	1	6·0	0	5	5
10	4	1	4·0	0	5, 4	2	4·5	0·5	5	4
11	—	0	—	—	5, 3, 4, 4	4	4·0	2·0	5	6
12	—	0	—	—	2, 4	2	3·0	2·0	2	4
13	4, 3, 3	3	3·33	0·67	3	1	3·0	0	6	3
14	2, 2, 2	3	2·0	0·0	—	0	—	—	2	3
Total	71	24	30·87	14·14	102	25	50·0	10·5	50	55
Mean	2·96	n_1	3·09	1·01*	4·08	n_2	3·85	0·88*	3·57	3·93

* Mean square within = (total deviance Δ_{ii})/(total d.f. ν_i).

terms 'daughters' rather than 'sisters', but it is customary to speak of 'sister–sister' rather than 'daughter–daughter' correlation. Similar remarks apply to 'brothers' or 'sons'; indeed, statements made below for one sex will be taken to imply similar ones for the other, when appropriate.) The data is presented in Table 1. Where there are no sisters in a particular family, a dash '—' is used to show that the mean $\bar{x}_{f1}$ and deviance Δ_{f11} are indeterminate. Where there is only 1 sister in a family, the deviance Δ_{f11} is necessarily exactly 0; elsewhere the means and deviances are given in the table to 1 or 2 decimal places. The actual calculations which follow were done to at least 4 significant figures, and so may not agree exactly with the sometimes rounded-off values in Table 1. This 4 figure accuracy is well beyond practical needs (2 or 3 significant figures would be ample), but was used here to see how quickly the iterations approached the final answer.

The suffixes 1, 2, 3, 4 mean respectively 'sisters' (or daughters), 'brothers', 'mothers', and 'fathers'. We begin by estimating the sister–sister correlation. Because no other position in the family is involved, we drop the suffix '1' in this calculation, as superfluous. We can conveniently delete from the data all 'empty' sisterships containing no sisters; there remain $k = 10$ non-empty sisterships. The mean square within sisterships, msq_A, is found by dividing the total deviance $\Delta = \Sigma\,\Delta_f = 14{\cdot}14$, see Table 1, by the total degrees of freedom,

$$\begin{aligned}\nu &= \text{total no. of sisters} - \text{no. of non-empty sisterships}\\ &= \Sigma\, n_f - k = 24 - 10 = 14,\end{aligned}$$

getting $msq_A = \Delta/\nu = 1{\cdot}01$. This serves as a preliminary estimate of A.

We take the (unweighted) mean of sistership means $\bar{x}_f$ as a preliminary estimate of μ, the population mean:

$$\mu \simeq \Sigma\,\bar{x}_f/k = 30{\cdot}87/10 = 3{\cdot}087.$$

To find a preliminary estimate of B, we could if we wished start off by assuming $B = A = 1{\cdot}01$

this is only a very rough value to start the iterative process going. But apart from its crudeness, this estimate has the disadvantage that in some situations (which we come to later) there may be no 'A' value to use for 'B'. A better, more universally applicable method is as follows. We calculate Δ_B, the sum of squares of deviations of sample means $\bar{x}_f$ from the estimated population mean μ, that is

$$\Delta_B = \Sigma(\bar{x}_f - \mu)^2 = 11{\cdot}92.$$

We set this equal to $\Sigma(n_f^{-1})A + kB$. We readily find $\Sigma(n_f^{-1})$, summed over non-empty sisterships, to be 5·950. We already have $k = 10$, and an assumed value 1·010 for A, so we have to solve the equation

$$11{\cdot}92 = 5{\cdot}950A + 10B,$$

giving $B = 0{\cdot}592$.

We improve these estimates iteratively, as follows.

(**) For each family, calculate a weight $\omega_f = (B + A/n_f)^{-1}$ (so that $\omega_1 = 0{\cdot}6246$, $\omega_2 = 0{\cdot}9120$, etc.). Calculate

$$\Sigma(\omega_f) = 9{\cdot}0863,$$
$$\Sigma(\omega_f \bar{x}_f) = 27{\cdot}1116,$$

whence we get a new and better estimate of the population mean,

$$\mu = \Sigma(\omega_f \bar{x}_f)/\Sigma(\omega_f) = 27{\cdot}1116/9{\cdot}0863 = 2{\cdot}9839.$$

We have already found $\nu = 14$ and $\Delta = 14{\cdot}124$ to be the respective total degrees of freedom and deviance within sisterships. We now calculate

$$\nu/A^2 = 14/1{\cdot}009^2 = 13{\cdot}75,$$
$$\Delta/A^2 = 14{\cdot}124/1{\cdot}009^2 = 13{\cdot}87.$$

For each family f we calculate the deviation $\xi_f = \bar{x}_f - \mu$ of its (sample) mean from the estimated population mean; for example, $\xi_1 = 2{\cdot}0 - 2{\cdot}9838 = -0{\cdot}9838$, $\xi_2 = -0{\cdot}9838$, etc. Then find (always excluding empty sisterships)

$$\begin{aligned} T_{11} &= \Sigma(\omega_f^2 n_f^{-2}) + \nu/A^2 = 16{\cdot}0564, \\ T_{12} &= T_{21} = \Sigma(\omega_f^2 n_f^{-1}) = 3{\cdot}8049, \\ T_{22} &= \Sigma(\omega_f^2) = 8{\cdot}8635, \\ S_1 &= S_A = \Sigma(\omega_f^2 n_f^{-1} \xi_f^2) + \Delta/A^2 = 18{\cdot}4360, \\ S_2 &= S_B = \Sigma(\omega_f^2 \xi_f^2) = 9{\cdot}0355. \end{aligned}$$

By a simple matrix inversion, $\mathbf{t} = \mathbf{T}^{-1}$ is

$$\begin{bmatrix} 16{\cdot}0564 & 3{\cdot}8049 \\ 3{\cdot}8049 & 8{\cdot}8635 \end{bmatrix}^{-1} = \begin{bmatrix} 0{\cdot}0693 & -0{\cdot}0298 \\ -0{\cdot}0298 & 0{\cdot}1256 \end{bmatrix},$$

so that the improved estimates of A, B are given by

$$\begin{bmatrix} A \\ B \end{bmatrix} = \mathbf{tS} = \begin{bmatrix} 0{\cdot}0693 & -0{\cdot}0298 \\ -0{\cdot}0298 & 0{\cdot}1256 \end{bmatrix} \begin{bmatrix} 18{\cdot}436 \\ 9{\cdot}036 \end{bmatrix} = \begin{bmatrix} 1{\cdot}0093 \\ 0{\cdot}5861 \end{bmatrix}.$$

Table 2. *Estimates of μ, A, B for brothers*

	Estimates of		
Iteration	μ	A	B
0	3·846	0·875	2·327
1	3·8980	0·8872	2·2523
2	3·8999	0·8882	2·2482
3	3·9000	0·8886	2·2478
$\geqslant 4$	3·9001	0·8886	2·2478
S.E. (approx)	0·47	0·36	0·79

We go back to (**) and repeat the whole process, using the new improved estimates of A and B. This leads to the still better estimates

$$\mu = 2{\cdot}9834, \quad A = 1{\cdot}0092, \quad B = 0{\cdot}5862.$$

A and B only differ from the previous estimates by 1 in the fourth decimal place, and hence are clearly the final estimates correct to 4 places. Thus, it seems that the first improvement had already provided estimates correct to almost 4 decimal places (in the sense of agreeing with the final values to this accuracy), though we could hardly have known that without calculating the second improved estimates.

Although dermatoglyphic distributions are usually distinctly non-normal, standard errors calculated as if we had a large sample from a normal distribution may be a useful indication of the approximate accuracy of our estimates. The error variance of μ is $1/\Sigma\,\omega_f = 1/9{\cdot}0863 = 0{\cdot}110$, whence its standard error is $\sqrt{0{\cdot}110} = 0{\cdot}33$. The error variance of A is $2t_{11} = 0{\cdot}1386$, whence S.E. $(A) = \sqrt{0{\cdot}1386} = 0{\cdot}37$. Similarly, the error variance of B is at $2t_{22}$, giving standard error $0{\cdot}50$, and the error covariance is $2t_{12}$. Note that, in the normal case, μ is statistically independent of A and B.

The estimated sister–sister correlation is

$$\rho = B/(A+B) = 0{\cdot}586/(1{\cdot}009+0{\cdot}586) = 0{\cdot}37.$$

The standard error of ρ is given by the formula (applicable, really, only to large samples)

$$\text{S.E.}\,(\rho) = (\rho^2/B^2)\sqrt{[B^2 \operatorname{var} A + A^2 \operatorname{var} B - 2AB \operatorname{cov}(A,B)]},$$

where $\operatorname{var} A$ means the error variance of A $(= 0{\cdot}1386)$, and so on. This gives S.E. $(\rho) = 0{\cdot}22$.

The results of a similar calculation on brothers are summarized in Table 2. The convergence is again extremely rapid, although not quite so rapid as with sisters. After only one improvement, the estimates differ from their final values by less than 1 % of their standard errors, which is more than ample accuracy.

8. CORRELATIONS BETWEEN RELATIVES OF DIFFERENT TYPES

The technique described above readily extends to correlations between different classes of relatives, such as sister–brother, mother–daughter, and so on. Let p denote a 'position' in a family, among the filial generation $p = 1$ meaning 'sister', $p = 2$ 'brother', and in the parental generation $p = 3$ 'mother' and $p = 4$ 'father'.

We suppose that there is one character, x, measured on all members of the family: thus x_{fpi} means the measurement of individual i of position p in family f (and x_{211} = measure of first sister in family number 2). We make assumptions analogous to (α), (β), (γ), (δ) of Section 2. Within any one family f and position p the character is distributed with true mean μ_{fp} and true variance v_{fpp}. The mean, among all families f, of μ_{fp} is μ_p, the population mean of individuals in position p; and the mean of the v_{fpp} is A_{pp}, the within-families component of variance. The covariance among different families of μ_{fp} and μ_{fq} is the between families covariance (or variance if $p = q$).

In a sample of individuals, suppose there are n_{fp} individuals in family f and position p. The values x_{fpi} of these individuals will have sample mean $\bar{x}_{fp}$, deviance Δ_{fpp} with degrees of freedom ν_{fpp}, except that when $n_{fp} = 0$ we set $\bar{x}_{fp}$, Δ_{fpp} and ν_{fpp} all equal to zero.

Note that in this context, there is no within-family covariance v_{fpq} for $p \neq q$, and hence no A_{pq} and no Δ_{fpq}. However, with a view to later applications, it is convenient to bring in the symbols v_{fpq}, A_{pq}, Δ_{fpq} and ν_{fpq} but to take them all to be zero. (The reader can simply omit these symbols). With this convention the true covariance between individuals of positions p and q, with $p \neq q$, is $A_{pq}+B_{pq}$, the (true) variance of individuals of position p is $A_{pp}+B_{pp}$, and hence the correlation between positions p and q is (for $p \neq q$)

$$\rho_{pq} = \frac{A_{pq}+B_{pq}}{\sqrt{[(A_{pp}+B_{pp})(A_{pq}+B_{qq})]}}. \tag{8.1}$$

Let $\Delta_{pp} = \Sigma_f \Delta_{fpp}$, $\nu_{pp} = \Sigma_f \nu_{fpp}$. Then Δ_{pp}/ν_{pp}, the within family mean square, is a suitable preliminary estimate of A_{pp}. Let

$$\Delta_{Bpq} = \sum_f (\bar{x}_{fp}-\mu_p)(\bar{x}_{fq}-\mu_q),$$

summed over all k_{pq} families for which $n_{fp} \neq 0 \neq n_{fq}$, and let $R_p = \Sigma(n_{fp}^{-1})$, summed over all f for which $n_{fp} \neq 0$. Then $(\Delta_{Bpp}-R_p A_{pp})/k_{pp}$ is a suitable preliminary estimate of B_{pp}, and Δ_{Bpq}/k_{pq} is a preliminary estimate of B_{pq} when $p \neq q$.

We estimate the population means μ_p as a weighted mean of the family sample means $\bar{x}_{fp}$. Let the matrix $\mathbf{u}_f$ have elements

$$u_{fpq} = A_{pq}s_{pq}n_{fp}^{-1}+B_{pq}, \tag{8.2}$$

where $s_{pp} = 1$ and $s_{pq} = 0$ when $p \neq q$. Then the inverse matrix $\boldsymbol{\omega}_f = \mathbf{u}_f^{-1}$ is the 'semi-weight' matrix for family f. Set $\boldsymbol{\Omega} = \Sigma_f \boldsymbol{\omega}_f$, and $\boldsymbol{\Psi} = \boldsymbol{\Omega}^{-1}$. Then the appropriate weighted mean estimate for the population mean is

$$\boldsymbol{\mu} = \boldsymbol{\Psi} \sum_f \boldsymbol{\omega}_f \bar{\mathbf{x}}_f, \tag{8.3}$$

in the sense that the component μ_f is the estimate of the mean character in individuals of position p.

An apparent difficulty arises when $n_{fp} = 0$ (as can frequently happen in practice) when the matrix $\mathbf{u}_f$ will have an infinite diagonal element u_{fpp}. But $\mathbf{u}_f$ will have a finite inverse, $\boldsymbol{\omega}_f$, found as follows:

(i) for each infinite diagonal element u_{fpp}, delete the corresponding row and column (row p, column p) from the matrix $\mathbf{u}_f$. When all such deletions have been performed, let $\mathbf{u}_f^*$ be the resulting matrix, which in general will be non-singular,

(ii) invert $\mathbf{u}_f^*$ to give $\boldsymbol{\omega}_f^* = (\mathbf{u}_f^*)^{-1}$,

(iii) whenever the pth row and column have been deleted from $\mathbf{u}_f$, put back into $\boldsymbol{\omega}_f^*$ a pth row and column consisting entirely of zero elements. Thus (using ? to denote an indeterminate or arbitrary element),

$$\mathbf{u}_f = \begin{bmatrix} 2 & ? & 3 & ? \\ ? & \infty & ? & ? \\ 3 & ? & 5 & ? \\ ? & ? & ? & \infty \end{bmatrix} \Rightarrow \mathbf{u}_f^* = \begin{bmatrix} 2 & 3 \\ 3 & 5 \end{bmatrix}$$

$$\Rightarrow \boldsymbol{\omega}_f^* = \begin{bmatrix} 5 & -3 \\ -3 & 2 \end{bmatrix} \Rightarrow \boldsymbol{\omega}_f = \begin{bmatrix} 5 & 0 & -3 & 0 \\ 0 & 0 & 0 & 0 \\ -3 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Having found $\boldsymbol{\omega}_f$, the rest of the calculation proceeds as above. In programming this for a computer, we are unable to set $u_{fpp} = \infty$, there being no such number in computer code. But the problem is not too difficult to overcome. We have only to note that the rows and columns which are to be deleted, and subsequently reinserted as zero rows and columns, are marked by $n_{fp} = 0$.

An additional complication can arise here. In our example, we have supposed that the 'positions' $p = 1, 2, 3, 4$ in the family were respectively sisters, brothers, mothers, fathers. Now we could suppose that sisters and mothers, both being female, would have the same population mean, $\mu_1 = \mu_3$, and similarly for brothers and fathers, $\mu_2 = \mu_4$. This would not be plausible for a character such as height, which would be influenced by a change in environmental conditions between successive generations, and probably also has appreciable selective value (even though it is difficult to demonstrate that experimentally). But it could be true (nearly enough) for some dermatoglyphic characters. We can easily modify the calculation to take this into account. Note that the equation (8.3) is the equivalent matrix form, of the equation $\boldsymbol{\Omega}\boldsymbol{\mu} = \sum_f \boldsymbol{\omega}_f \bar{\mathbf{x}}_f$. Written out in full this is a set of linear equations for μ_p

$$\begin{aligned} \Omega_{11}\mu_1 + \Omega_{12}\mu_2 + \Omega_{13}\mu_3 + \dots &= \sum_f \sum_q \omega_{f1q}\bar{x}_{fq}, \\ \Omega_{21}\mu_1 + \Omega_{22}\mu_2 + \Omega_{23}\mu_3 + \dots &= \sum_f \sum_q \omega_{f2q}\bar{x}_{fq}, \\ \dots \qquad \dots \qquad \dots &= \quad \dots \end{aligned} \qquad (8.4$$

What we now do is to replace μ_3 by μ_1, and at the same time add together the first and thir equations. Similarly we replace μ_4 by μ_2, and add together the second and fourth equations More formally, we proceed as follows. Because $\mu_1 = \mu_3$, we do not have two separate parameter μ_1, μ_3, to estimate, but only one, which we will call M_1. Similarly we replace μ_2, μ_4 by a single parameter M_2.

Define the symbol m_{ip} to have the value 1 when the population mean value μ_p of individual of position p is equal to the parameter M_i, and to be zero otherwise. (Thus in our case, $m_{11} = m_{13} =$ showing that sisters (1) and mothers (3) are assumed to have the same true mean, and $m_{22} = m_{24} = 1$.)

Find

$$\begin{aligned} X_i &= \sum_{f,p,q} m_{ip}\,\omega_{fpq}\,\bar{x}_{fq}, \\ \Omega_{ij} &= \sum_{f,p,q} m_{ip}\,m_{jq}\,\omega_{fpq}, \\ \mathbf{\Psi} &= \mathbf{\Omega}^{-1}, \\ \mathbf{M} &= \mathbf{\Psi}\mathbf{X}. \end{aligned} \tag{8.5}$$

Then the components M_i of $\mathbf{M}$ are improved estimates of the parameters. Each parameter M_i represents the assumed value of one or more means μ_p, and hence in finding an improved value of M_i we find at the same time an improved value of μ_p. For the purpose of computer programming, this can be done by setting

$$\mu_p = \sum_i m_{ip} M_i.$$

Similar considerations arise in the estimation of the variance components A_{pq} and B_{pq}; in general, there will be more of these than parameters to be estimated. For one thing, we necessarily have $A_{qp} = A_{pq}$ and $B_{qp} = B_{pq}$. For another, we have seen that in the situation at present being considered, $A_{pq} = 0$ when $p \neq q$. Furthermore, since each family has only one mother, there is no within family variation for mothers, so $A_{33} = 0$, and similarly for fathers $A_{44} = 0$.

We will therefore rename the separate variance and covariance components which are to be estimated as $C_1, C_2, \ldots$. Each C_i will represent either the value of one or more within family variance or covariance components A_{pq}, in which case we set $a_{ipq} = 1$, or else the value of one or more between family variance or covariance components B_{pq}, in which case we set $b_{ipq} = 1$. Otherwise, we set a_{ipq} and b_{ipq} to zero. As before we will write

$$\xi_{fp} = \bar{x}_{fp} - \mu_p,$$

the deviation of the mean of position p in family f from the corresponding estimated population mean. We will denote the matrix inverse to $\mathbf{A}$ by $\boldsymbol{\alpha}$. And, to avoid infinite values, it will be useful to introduce the symbol

$$r_{fp} = n_{fp}^{-1} \quad (\text{when } n_{fp} \neq 0),\ 0\ (\text{when } n_{fp} = 0). \tag{8.6}$$

If the parameter C_i represents a within family component, we set

$$S_i = \sum_{p,q} (a_{ipq} [\sum_{f,P,Q} \omega_{fPp}\,\omega_{fqQ}\,r_{fP}\,\xi_{fP}\,\xi_{fQ} + \sum_{P,Q} \Delta_{PQ}\,\alpha_{Pp}\,\alpha_{qQ}]). \tag{8.7}$$

If C_i represents a between family component, set

$$S_i = \sum_{p,q} (b_{ipq} \sum_{f,P,Q} \omega_{fPp}\,\omega_{fqQ}\,\xi_{fP}\,\xi_{fQ}). \tag{8.8}$$

Since, for any given i, either the right-hand side of (8.7) or the right-hand side of (8.8) is zero, we could formally define S_i for all i as the sum of the expressions on the right-hand sides of equations (8.7), (8.8). This could simplify computer programming. In the same way, (at most) one of the following four expressions will be non-zero, depending on whether C_i and C_j represent within family or between family components. We define T_{ij} to be this non-zero value, or, equivalently, as the sum of the four expressions:

$$\begin{aligned} &\sum_{p,q}\sum_{P,Q} (a_{ipq}\,a_{jPQ} [\sum_f r_{fp}\,r_{fP}\,\omega_{fPp}\,\omega_{fqQ} + \nu_p\,\alpha_{Pp}\,\alpha_{qQ}]), \\ &\sum_{p,q}\sum_{P,Q} (a_{ipq}\,b_{jPQ} [\sum_f r_{fp}\,\omega_{fPp}\,\omega_{fqQ}]), \\ &\sum_{p,q}\sum_{P,Q} (b_{ipq}\,a_{jPQ} [\sum_f r_{fP}\,\omega_{fPp}\,\omega_{fqQ}]), \\ &\sum_{p,q}\sum_{P.Q} (b_{ipq}\,b_{jPQ} [\sum_f \omega_{fPp}\,\omega_{fqQ}]). \end{aligned} \tag{8.9}$$

The T_{ij} form a square symmetric matrix. Let $\mathbf{t} = \mathbf{T}^{-1}$ be its inverse. Then

$$\mathbf{C} = \mathbf{tS} \tag{8.10}$$

is a vector of improved estimates of the C_i. These, of course, immediately give improved estimates of the A_{pq} and B_{pq}, which can be written

$$A_{pq} = \sum_i a_{ipq} C_i, \quad B_{pq} = \sum_i b_{ipq} C_i. \tag{8.11}$$

As before, the process can be repeated to get still better estimates. We go back to (8.2) to define a new matrix $\mathbf{u}_f$, and hence $\boldsymbol{\omega}_f = \mathbf{u}_f^{-1}$. We recalculate $\mathbf{X}$, $\boldsymbol{\Omega}$, $\boldsymbol{\Psi}$, and $\mathbf{M}$ from (8.5), and hence $\boldsymbol{\mu}$. We then find $\mathbf{S}$ from (8.7) and (8.8), $\mathbf{T}$ from (8.9), and new $\mathbf{C}$ from (8.10). The iteration is halted when provisional and improved values agree sufficiently closely.

Under assumptions of a large sample and normal distribution, the error variance matrix of $\mathbf{M}$ is $\boldsymbol{\Psi}$, i.e. the standard error of M_i is $\sqrt{\Psi_{ii}}$. The $\mathbf{C}$ parameter estimates are statistically independent of the $\mathbf{M}$, and the error variance matrix of the $\mathbf{C}$ is $2\mathbf{t}$.

This procedure is again justified by the unbiasedness principle. In any family f, taking $\boldsymbol{\mu}$ to be fixed,

$$\mathscr{E}(\xi_{fp}\xi_{fq}) = B_{pq} + r_{fp} A_{pq} = u_{fpq}. \tag{8.12}$$

By using the fact that, by definition, $\mathbf{u}_f = \boldsymbol{\omega}_f^{-1}$, it is not difficult to show that $\mathscr{E}(\mathbf{S}) = \mathbf{TC}$. Hence, by setting $\mathbf{S} = \mathbf{TC}$, or $\mathbf{C} = \mathbf{T}^{-1}\mathbf{S} = \mathbf{tS}$, we get nearly unbiased estimates of the $\mathbf{C}$. The method of Section 6 is a particular case. Later we sketch a proof that in the normal case these are maximum-likelihood estimates.

Unfortunately, this method can lead in practice to the inversion of large matrices. Applied to brother–sister correlation, it needs the inversion of 5×5 matrices $\mathbf{T}$. Applied to the simultaneous estimation of components of variance and covariance for sisters, brothers, mothers, fathers, it would give rise to 12×12 $\mathbf{T}$ matrices, which, however, would still be practical on a computer. For more complicated problems, it would seem wise not to try to estimate everything simultaneously, even at the cost of a slight loss of efficiency. There is also the point that we might reasonably require the variance of mothers to be the same as that of sisters, leading to a relation $B_{33} = A_{11} + B_{11}$. I intend to show how to deal with such a problem in a further paper.

9. CROSS-CORRELATIONS

In addition to correlations between, say, sister's height and brother's height, there are 'cross-correlations' between sister's height and brother's forearm length. These can readily be covered by our scheme, by using the following formal device. We call 'sister's height', 'sister's forearm length', 'brother's height', 'brother's forearm length' different 'positions' in the family, say positions numbers 1, 2, 3, 4. It is as if in imagination we replaced each sister by a pair of identical twins, indistinguishable except for order of birth. On the first-born twin we measure only the height, on the second only the arm-length, so that 'first-born sister' and 'second-born sister' do occupy different positions in the family, and are represented by different measurements. But we assume that on each individual all relevant measurements are taken; no real sister has her height measured but not her arm length. Otherwise complications arise. We now proceed exactly as before, with one slight modification. Because the measures of sister's height (position 1) and arm length (position 2) are in fact measured on the same individual, they will be associated in pairs, so there will now be a codeviance Δ_{f12} *within* families, and a corresponding non-zero covariance component A_{12}. The formulas given above allow for these.

In principle, this would enable us to estimate simultaneously correlations and cross-correlations between any number of characters in any number of 'positions' in a family. But the size of matrices to be inverted rises very rapidly indeed with the number of characters and positions used.

10. JUSTIFICATION OF THE METHOD

We now sketch a proof that the method described above does give the maximum-likelihood estimate, or 'evaluate', in the case of a multivariate normal distribution.

We will consider here the case without cross-correlations, i.e. with $A_{pq} = 0$ when $p \neq q$; the case with some $A_{pq} \neq 0$ follows in a similar way, with suitable modifications.

Let L denote the support function, that is, the natural logarithm of the likelihood. This will be the sum of the support functions of the separate families,

$$L = \Sigma L_f. \tag{10.1}$$

Now it is a property of the normal distribution that it factorizes into two independent parts. One 'between families' part represents the distribution of the sample means of families. Since the $\bar{\mathbf{x}}_f$ have population mean $\boldsymbol{\mu}$ and variance matrix $\mathbf{u}_f$, as defined above, the formula for the normal distribution shows that this gives a contribution to the support (with $\boldsymbol{\omega}_f = \mathbf{u}_f^{-1}$)

$$L_{fB} = -\tfrac{1}{2}[\ln\det\mathbf{u}_f + (\bar{\mathbf{x}}_f - \boldsymbol{\mu})^T\,\boldsymbol{\omega}_f(\bar{\mathbf{x}}_f - \boldsymbol{\mu})]. \tag{10.2}$$

The second part of the support relates to variability within families. Let $\boldsymbol{\alpha} = \mathbf{A}^{-1}$. When $\mathbf{A}$ is diagonal, so also is $\boldsymbol{\alpha}$, with diagonal elements $\alpha_{pp} = A_{pp}^{-1}$. By suitable rotation of axes it is not difficult to show that the contribution to the support is

$$L_{fA} = \sum_p (\Delta_{fpp}\alpha_{pp} + \nu_{fp}\ln A_{pp}). \tag{10.3}$$

(When $\mathbf{A}$ has off-diagonal elements, L_{fA} takes a slightly more complicated form. But note that we suppose that the sampling is such that $\nu_{fp} = \nu_{fq}$ whenever $A_{pq} \neq 0$.) The total support is accordingly

$$L = \sum_f L_{fA} + \sum_f L_{fB} = \Sigma[\Delta_{pp}\alpha_{pp} + \nu_p A_{pp}] - \tfrac{1}{2}\Sigma[\ln\det\mathbf{u}_f + \sum_{p,q}\ \omega_{fpq}(\bar{x}_{fp} - \mu_p)(\bar{x}_{fq} - \mu_q)]. \tag{10.4}$$

The efficient scores are $U_i^M = \partial L/\partial M_i$, $U_j^C = \partial L/\partial C_j$, considered as functions of the M_i, C_j, and the maximum likelihood estimates are solutions of the equations

$$U_i^m(\mathbf{M},\mathbf{C}) = 0, \quad U_j^C(\mathbf{M},\mathbf{C}) = 0. \tag{10.5}$$

Now we have immediately from the definition

$$\begin{aligned} \partial u_p/\partial M_i &= m_{ip}, \\ \partial A_{pq}/\partial C_i &= a_{ipq}, \\ \partial B_{pq}/\partial C_i &= b_{ipq}. \end{aligned}$$

Since $u_{fpq} = B_{pq} + n_{fp}^{-1}A_{pq}$, we have

$$\begin{aligned} \partial u_{fpq}/\partial B_{PQ} &= 1 \quad \text{when} \quad p = P, q = Q, \quad \text{otherwise } 0; \\ \partial u_{fpq}/\partial A_{PQ} &= n_{fp}^{-1} \quad \text{when} \quad p = P, \quad q = Q, \quad \text{otherwise } 0. \end{aligned}$$

If $\mathbf{u}$, $\boldsymbol{\omega}$ are any two inverse matrices, the relation $\mathbf{u}\boldsymbol{\omega} = \mathbf{I}$ gives $d\mathbf{u}\,.\,\boldsymbol{\omega}+\mathbf{u}\,d\boldsymbol{\omega} = \mathbf{0}$, whence

$$\partial\omega_{pq}/\partial u_{PQ} = -\omega_{pP}\omega_{Qq}.$$

Also an expansion of $\det\mathbf{u}$ by a row (or column) shows that $\partial(\det\mathbf{u})/\partial u_{pq}$ is the cofactor of u_{pq}, whence

$$\partial(\ln\det\mathbf{u})/\partial u_{pq} = \omega_{qp}. \tag{10.6}$$

We can thus find U_i^M, U_j^C by straightforward differentiation. We can easily find (see equation (8.5))

$$\begin{aligned} U_i^m &= \sum_{f,p,q} m_{ip}\omega_{fpq}(\bar{x}_{fq}-\mu_q) \\ &= \sum_{f,p,q} m_{ip}\omega_{fpq}\bar{x}_{fq} - \sum_{f,p,q,j} m_{ip}m_{fq}\omega_{fpq}M_j \\ &= X_i - \sum \Omega_{ij}M_j. \end{aligned} \tag{10.7}$$

Hence setting $U_i^M = 0$ for all i amounts to setting $\mathbf{X} = \boldsymbol{\Omega}\mathbf{M}$, or $\mathbf{M} = \boldsymbol{\Omega}^{-1}\mathbf{X} = \boldsymbol{\Psi}\mathbf{X}$, in accordance with (8.5).

Analogously we find, after some algebraic manipulation, that

$$U_j^C = -\tfrac{1}{2}\big(\sum_k T_{jk}C_k - S_j\big), \tag{10.8}$$

valid whether C_j represents a within family or between family component. Hence to set all $U_j^C = 0$ amounts to setting $\mathbf{TC} = \mathbf{S}$, or $\mathbf{C} = \mathbf{T}^{-1}\mathbf{S} = \mathbf{tS}$, in agreement with (8.10).

The information matrix $\mathbf{J}$ can as usual be found as the matrix of the negative second derivative of the support, i.e. first derivatives of the scores $\mathbf{U}$. From (10.8) we readily find that

$$\begin{aligned} J_{ih}^{MM} &= -\partial U_i^M/\partial M_h \\ &= \sum m_{ip}m_{hq}\omega_{fpq} \\ &= \Omega_{ih}. \end{aligned} \tag{10.9}$$

For the mixed second derivatives of the form $J_{ij}^{MC} = -\partial U_i^m/\partial C_j$, we see from the top line of (10.8) that this must be

$$J_{ij}^{MC} = -\sum_{f,p,q} m_{ip}(\partial\omega_{fpq}/\partial C_j)(\bar{x}_{fq}-\mu_q),$$

i.e. a linear function of the $(\bar{x}_{fq}-\mu_q)$. But $(\bar{x}_{fq}-\mu_q)$ has expectation zero, hence J_{ij}^{MC} also has expectation zero, and can be put equal to 0 accurately enough. Thus we see that the $\mathbf{J}$ matrix partitions in the form

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}^{MM} & \mathbf{o} \\ \mathbf{o} & \mathbf{J}^{CC} \end{bmatrix},$$

where $\mathbf{J}^{MM} = \boldsymbol{\Omega}$.

The error variance matrix of M is accordingly $(\mathbf{J}^{MM})^{-1} = \boldsymbol{\Omega}^{-1} = \boldsymbol{\Psi}$, as previously asserted.

When we come to evaluate $J_{jk}^{CC} = -\partial u_j^C/\partial C_k$, we obtain rather complicated algebraic expressions involving products $\xi_{fP}\xi_{fQ}$ arising from the S_j term. However, $\xi_{fP}\xi_{fQ}$ has expectation u_{fPQ}. If we replace the $\xi_{fQ}\xi_{fQ}$ by these expectations, it will be found that J_{jk}^{CC} reduces to $\tfrac{1}{2}T_{jk}$. Hence the error variance matrix for the $\mathbf{C}$ is $(\mathbf{J}^{CC})^{-1} = (\tfrac{1}{2}\mathbf{T})^{-1} = 2\mathbf{t}$, as previously stated.

SUMMARY

A method is proposed for simultaneous estimation of genetical variance and covariance components for quantitative characters and hence correlations and cross-correlations between relatives. It proceeds by successive approximations, which are not difficult to compute. Judging by examples, they appear to converge rapidly in practice. When distributions are normal and homoscedastic, the method gives the maximum-likelihood estimates and their error variances and covariances. It can be readily modified to take into account assumptions such as that means and variances are equal in fathers and sons. But the amount of calculations required goes up very rapidly with the number of characters and relatives considered simultaneously.

I have to thank Dr Abdülbari Bener for interest and help in this problem, and for tracing other work in this field.

REFERENCES

BROWNLEE, K. A. (1949). *Industrial Experimentation.* 4th ed. London: Her Majesty's Stationery Office.

COCHRAN, W. G. (1954). The combination of estimates from different experiments. *Biometrics* **10**, 101–129.

CORBEIL, R. R. & SEARLE, S. R. (1976). Restricted maximum likelihood estimation of variance components in the mixed model. *Technometrics* **18**, 31–38.

CUNNINGHAM, E. P. & HENDERSON, C. R. (1968). An iterative procedure for estimating fixed effects and variance components in mixed model situations. *Biometrics* **24**, 13–25.

DEMPSTER, A. P., LAIRD, N. M. & RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society* B **39**, 1–38.

FIELLER, E. C. & SMITH, C. A. B. (1951). Note on the analysis of variance and intraclass correlation. *Annals of Eugenics* **16**, 97–104.

FISHER, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh* **52**, 399–433.

FISHER, R. A. (1925). *Statistical Methods for Research Workers*, 1st ed. Edinburgh: Oliver and Boyd.

FORREST, D. W. (1974). *Francis Galton: the Life and Work of a Victorian Genius.* London: Paul Elek. New York: Taplinger.

GALTON, F. (1877). Typical laws of heredity. *Proceedings of the Royal Institution* **8**, 282–301.

GALTON, F. (1885*a*). Presidential Address, Section H, Anthropology. *British Association Reports* **55**, 1206–1214.

GALTON, F. (1885*b*). Co-relations and their measurement, chiefly from anthropometric data. *Proceedings of the Royal Society* **45**, 135–145.

GALTON, F. (1908). *Memories of My Life.* London: Methuen.

HARTLEY, H. O. & RAO, J. N. K. (1957). Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika* **54**, 93–108.

HARVILLE, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* **72**, 320–338.

HEMMERLE, W. J. & HARTLEY, H. O. (1973). Computing maximum likelihood estimates for the mixed ADV model using the W transformation. *Technometrics* **18**, 819–831.

LI, C. C. & SACKS, L. (1954). The derivation of the joint distribution and correlation between relatives by the use of stochastic matrices. *Biometrics*, **10**, 347–360.

LINDLEY, D. V. & SMITH, A. F. M. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society* B 1–18.

NELDER, J. A. (1968). The combination of information in generally balanced designs. *Journal of the Royal Statistical Society* B **30**, 303–311.

PATTERSON, H. D. & THOMPSON, R. (1971). Recovery of interblock information when block sizes are unequal. *Biometrika* **58**, 545–554.

PATTERSON, H. D. & THOMPSON, R. (1974). Maximum likelihood estimation of components of variance. *Proceedings of the 8th International Biometric Conference*, pp. 197–207.

PEARSON, K. (1895). Regression, heredity and panmixia. *Philosophical Transactions of the Royal Society* A **187**, 253–318.

PEARSON, K. (1896). Mathematical contributions to the theory of evolution. *Philosophical Transactions of the Royal Society* A **187**, 279–283.

PEARSON, K. (1930). *Life, Letters and Labours of Francis Galton*, vol IIIA. Cambridge: Cambridge University Press.

PEARSON, K. & LEE, A. (1903). On the laws of inheritance in man. I. *Biometrika* **2**, 357–462.

RAO, C. R. (1971*a*). Estimation of variance and covariance components – MINQUE theory. *Journal of Multivariate Analysis* **1**, 257–275.

RAO, C. R. (1971*b*). Minimum variance quadratic unbiased estimation of variance components. *Journal of Multivariate Analysis* **1**, 445–456.

RAO, C. R. (1972). Estimation of variance and covariance components in linear models. *Journal of the American Statistical Association* **67**, 112–115.

SCHAEFFER, L. R., WILTON, J. W. & THOMPSON, R. (1978). Simultaneous estimation of variance and covariance components from multi-trait mixed model equations. *Biometrics* **34**, 199–208.

SMITH, C. A. B. (1957). On the estimation of intraclass correlation. *Annals of Human Genetics* **21**, 363–373.

THOMPSON, R. (1977). Estimation of quantitative genetic parameters. *Proceedings of the International Conference on Quantitative Genetics*, 639–657. Ames: Iowa State University Press.

WRIGHT, S. (1921). Correlation and causation. *Journal of Agricultural Research* **20**, 557–585.

WRIGHT, S. (1922). Coefficients of inbreeding and relationship. *American Naturalist* **56**, 330–338.

WRIGHT, S. (1923). The theory of path coefficients. *Genetics* **8**, 239–255.