Ann. Inst. Statist. Math.
34 (1982), Part A, 505-515

ASYMPTOTIC PROPERTIES OF ESTIMATORS OF INTERCLASS CORRELATION FROM FAMILIAL DATA

SADANORI KONISHI

(Received Feb. 24, 1982; revised July 29, 1982)

Summary

Asymptotic properties of several estimators of interclass correlation from familial data are examined in the case of a variable number of siblings per family. After showing that the usual sib-mean estimator is not consistent, a modified sib-mean estimator is proposed. Asymptotic distributions of estimators are derived and a test procedure is provided for a certain testing problem concerning interclass correlation. Several estimators are compared in the various mean number of siblings per family, using asymptotic mean square errors.

1. Introduction

An important problem in the analysis of familial data is to estimate the degree of resemblance between a parent and siblings. Several estimators, the pairwise, sib-mean, random-sib and ensemble estimators, have been proposed for a parent-child correlation or interclass correlation. Rosner, Donner and Hennekens [8] have compared these estimators through Monte Carlo simulation, and showed that the pairwise and ensemble estimators are far superior to the sib-mean and random-sib estimators in terms of mean square errors. Furthermore, by the same approach, Rosner [6] has compared the pairwise and ensemble estimators with the maximum likelihood estimate obtained by using the Newton-Raphson methods. The maximum likelihood estimate can not be expressed in closed form.

In this paper asymptotic properties of these estimators are examined in the case of a variable number of siblings per family. It can be seen that the sib-mean estimator is not a consistent estimator, whereas the other estimators are consistent. A modified sib-mean estimator is proposed and compares favourably with previous estimators. Further, we

Key words and phrases: Interclass correlation, familial data, consistent estimator, asymptotic distribution.

derive asymptotic distributions of estimators and consider the testing problem concerning interclass correlations when one has familial data. Several estimators for interclass correlation are compared for various values of the mean number of siblings per family, using asymptotic mean square errors.

2. Model and estimates

2.1. *Familial data*

Suppose we have a random sample of N families each consisting of mother, in general, parent and her siblings. Let $x_\alpha=(x_{1\alpha}, x_{2\alpha},\cdots, x_{k_\alpha+1,\alpha})'$ be the observed values of the αth family concerning a certain character where $x_{1\alpha}$ is the mother's score and $x_{2\alpha},\cdots, x_{k_\alpha+1,\alpha}$ are the scores of her k_α siblings. That is, we consider a situation that there are a variable number of siblings per family. Assume that x_α have a $(k_\alpha+1)$-variate normal distribution with mean vector $(\mu_m, \mu_s,\cdots, \mu_s)'$ and covariance matrix

$$\Sigma_\alpha=\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{12}' & \Sigma_{22}\end{pmatrix} \tag{2.1}$$

where $\Sigma_{11}=\sigma_m^2$, $\Sigma_{12}=(\rho_{ms}\sigma_m\sigma_s,\cdots, \rho_{ms}\sigma_m\sigma_s)$ and $\Sigma_{22}=\sigma_s^2\{(1-\rho_{ss})I+\rho_{ss}ee'\}$ with I the identity matrix of order k_α and $e=(1,\cdots,1)'$ the k_α dimensional vector. We assume that the parameters do not depend upon sibship size and that sibships have a limited size and $k_\alpha\geqq 1$. On the basis of observations from N families, we wish to estimate the interclass correlation ρ_{ms}. Estimator is used to assess the degree of resemblance between a parent and his or her siblings.

2.2. *Estimators of interclass correlation*

Estimators discussed in the present paper are as follows:

(1) *Pairwise estimator*. The pairwise estimator is obtained by pairing each mother's score with the k_α sibling scores and assuming such pairs to be independent. This estimator is of the form

$$\hat\rho_{ms,p}=\sum_{\alpha=1}^N\sum_{i=2}^{k_\alpha+1}(x_{1\alpha}-\tilde x_m)(x_{i\alpha}-\tilde x_s)\Big/\left\{\sum_{\alpha=1}^N k_\alpha(x_{1\alpha}-\tilde x_m)^2\sum_{\alpha=1}^N\sum_{i=2}^{k_\alpha+1}(x_{i\alpha}-\tilde x_s)^2\right\}^{1/2}$$

where $\tilde x_m=\sum_{\alpha=1}^N k_\alpha x_{1\alpha}\Big/\sum_{\alpha=1}^N k_\alpha$ and $\tilde x_s=\sum_{\alpha=1}^N\sum_{i=2}^{k_\alpha+1} x_{i\alpha}\Big/\sum_{\alpha=1}^N k_\alpha$. As discussed in the Appendix, the pairwise estimator $\hat\rho_{ms,p}$ is a consistent estimator.

(2) *Random-sib estimator*. The random-sib estimator, say $\hat\rho_{ms,r}$, is obtained by choosing one sibling randomly from each family and computing the ordinary product-moment correlation based on samples paired with the mother of that family.

(3) *Ensemble estimator.* Rosner, Donner and Hennekens [8] proposed the estimator

$$\hat{\rho}_{ms,e}=\sum_{\alpha=1}^{N}(x_{1\alpha}-\bar{x}_m)(\bar{x}_{s\alpha}-\bar{x}_s)\bigg/\left[\sum_{\alpha=1}^{N}(x_{1\alpha}-\bar{x}_m)^2\left\{\frac{n}{N}\sum_{\alpha=1}^{N}\sum_{i=2}^{k_\alpha+1}(x_{i\alpha}-\bar{x}_{s\alpha})^2/k_\alpha+\sum_{\alpha=1}^{N}(\bar{x}_{s\alpha}-\bar{x}_s)^2\right\}\right]^{1/2}$$

where $n=N-1$, $\bar{x}_m=\frac{1}{N}\sum_{\alpha=1}^{N}x_{1\alpha}$, $\bar{x}_{s\alpha}=\frac{1}{k_\alpha}\sum_{i=2}^{k_\alpha+1}x_{i\alpha}$ and $\bar{x}_s=\frac{1}{N}\sum_{\alpha=1}^{N}\bar{x}_{s\alpha}$. This estimator was obtained by computing the expected value of the random-sib estimator $\hat{\rho}_{ms,r}$ approximately over all possible choices of random sibs from each family. Following the same line of approach as discussed in Appendix, we may find that $\hat{\rho}_{ms,e}$ is a consistent estimator for ρ_{ms}.

(4) *Sib-mean estimator.* The sib-mean estimator is obtained by pairing each mother's score with the mean of her k_α siblings. So the estimator has the following form:

$$\hat{\rho}_{ms,s}=\sum_{\alpha=1}^{N}(x_{1\alpha}-\bar{x}_m)(\bar{x}_{s\alpha}-\bar{x}_s)\bigg/\left\{\sum_{\alpha=1}^{N}(x_{1\alpha}-\bar{x}_m)^2\sum_{\alpha=1}^{N}(\bar{x}_{s\alpha}-\bar{x}_s)^2\right\}^{1/2}.$$

The sib-mean estimator seems, intuitively, to be more effective than the pairwise estimator which is obtained by treating the pairs of each mother's score and her k_α siblings' scores as if they are independent. We may, however, find that the expectation of $\hat{\rho}_{ms,s}$ is, noting that $N^{-1}\sum_\alpha(1/k_\alpha)=O(1)$, asymptotically

$$\mathrm{E}\,[\hat{\rho}_{ms,s}]=\left\{\frac{1}{N}\sum_{\alpha=1}^{N}\frac{1}{k_\alpha}+\left(1-\frac{1}{N}\sum_{\alpha=1}^{N}\frac{1}{k_\alpha}\right)\rho_{ss}\right\}^{-1/2}\rho_{ms}.$$

The sib-mean estimator is therefore not a consistent estimator for ρ_{ms}.

(5) *Modified sib-mean estimator.* We modify the sib-mean estimator as follows:

$$\hat{\rho}_{ms,cs}=\{c_1+(1-c_1)\rho_{ss}\}^{1/2}\hat{\rho}_{ms,s} \tag{2.2}$$

where $c_1=N^{-1}\sum_\alpha(1/k_\alpha)$. This is a consistent estimator for ρ_{ms}. The intraclass correlation ρ_{ss} contained in (2.2) is in general unknown and has to be estimated from familial data. Several estimators have been proposed for the intraclass correlation. An obvious estimator for ρ_{ss} is given by

$$\hat{\rho}_{ss}=\left\{\sum_{\alpha=1}^{N}\sum_{i\neq j}^{k_\alpha+1}(x_{i\alpha}-\hat{\mu}_s)(x_{j\alpha}-\hat{\mu}_s)\bigg/\sum_{\alpha=1}^{N}k_\alpha(k_\alpha-1)\right\}\bigg/\left\{\sum_{\alpha=1}^{N}\sum_{i=2}^{k_\alpha+1}(x_{i\alpha}-\hat{\mu}_s)^2\bigg/\sum_{\alpha=1}^{N}k_\alpha\right\} \tag{2.3}$$

where $\hat{\mu}_s=\sum_{\alpha=1}^{N}\sum_{i=2}^{k_\alpha+1} x_{i\alpha}\Big/\sum_{\alpha=1}^{N} k_\alpha$.

The usual sib-mean estimator can be markedly improved in terms of mean square error by using the modification (2.2) with $\hat{\rho}_{ss}$. If all families are of the same size, (2.3) gives the maximum likelihood estimate for ρ_{ss} and the pairwise estimator is equivalent to the maximum likelihood estimate for ρ_{ms} (Rosner, Donner and Hennekens [8]). In the case of a variable number of siblings per family, the maximum likelihood estimate does not exist in closed form. Recently, an iterative algorithm for calculating the maximum likelihood estimate was given by Mak and Ng [5].

3. Asymptotic results

3.1. *Moments and distributions*

As discussed in the Appendix, if we expand each estimator around ρ_{ms} and calculate expected values, then we can obtain the asymptotic moments of the estimators. The results for the first moments of the pairwise, ensemble and modified sib-mean estimators around ρ_{ms} are, respectively, given as follows:

$$\begin{aligned} \mathrm{E}\,[\hat{\rho}_{ms,p}-\rho_{ms}]=&\frac{1}{n}\rho_{ms}\Big\{a_1+(\textstyle\sum k_\alpha^2/\sum k_\alpha)a_2\\ &+\frac{1}{2}(1-\rho_{ss})(1-\textstyle\sum k_\alpha^2/\sum k_\alpha)\Big\}\{n\textstyle\sum k_\alpha/\sum_{\alpha\neq\beta} k_\alpha k_\beta\}\ , \end{aligned}$$

$$\mathrm{E}\,[\hat{\rho}_{ms,e}-\rho_{ms}]=\frac{1}{n}\rho_{ms}(c_1a_1+a_2)\ , \tag{3.1}$$

$$\begin{aligned} \mathrm{E}\,[\hat{\rho}_{ms,cs}-\rho_{ms}]=&\frac{1}{n}\rho_{ms}\Big[-\frac{13}{8}+\frac{1}{2}c_2^{-1}\rho_{ms}^2+\frac{3}{8}c_2^{-2}\{3\rho_{ss}^2+6\rho_{ss}(1-\rho_{ss})c_1\\ &+(1-\rho_{ss})^2(c_1^2+2N^{-1}\textstyle\sum k_\alpha^{-2})\}\\ &+\frac{1}{2}c_2^{-1}(1-c_1)\rho_{ss}(a_3+a_4+a_5)-\frac{1}{8}c_2^{-2}(1-c_1)^2\rho_{ss}^2a_6\Big]\ , \end{aligned}$$

where $c_1=N^{-1}\sum(1/k_\alpha)$, $c_2=N^{-1}\sum(1/k_\alpha)+\{1-N^{-1}\sum(1/k_\alpha)\}\rho_{ss}$ and

$$a_1=-\frac{1}{4}+\rho_{ss}-\frac{3}{4}\rho_{ss}^2\ ,\qquad a_2=-\frac{1}{4}-\rho_{ss}+\frac{3}{4}\rho_{ss}^2+\frac{1}{2}\rho_{ms}^2\ ,$$

$$\begin{aligned} a_3=&2(\textstyle\sum k_\alpha/n-1)(1-\rho_{ss})\{n/\sum k_\alpha+\rho_{ss}^{-1}n/\sum k_\alpha(k_\alpha-1)\}+\rho_{ms}^2(1-\rho_{ss}^{-1})\\ &+c_2^{-1}\Big[(n/\textstyle\sum k_\alpha)(1-\rho_{ss})(\rho_{ss}+c_2)+\rho_{ss}\Big(\rho_{ss}-\frac{3}{2}\Big)\\ &+\{n/\textstyle\sum k_\alpha(k_\alpha-1)\}\Big\{c_1(\rho_{ss}+\rho_{ss}^{-1}-2)-3\rho_{ss}-\rho_{ss}^{-1}+4+(\sum k_\alpha/n) \end{aligned}$$

$$\times\left(\frac{3}{2}\rho_{ss}-2\right)+\frac{1}{2}(\sum k_\alpha^2/n)\rho_{ss}\Big\}\Big]\ ,$$

$$\begin{aligned}(3.2)\qquad a_4=&-2(n/\sum k_\alpha)[1+\{\sum k_\alpha(k_\alpha-1)(k_\alpha-2)/\sum k_\alpha(k_\alpha-1)\}\rho_{ss}\\&-\{\sum k_\alpha(k_\alpha-1)/\sum k_\alpha\}\rho_{ss}^2]\ ,\\ a_5=&(n/\sum k_\alpha)[1-\rho_{ss}^{-1}+\{\sum k_\alpha(k_\alpha-1)/\sum k_\alpha\}(1+\rho_{ss})\\&-2\{\sum k_\alpha(k_\alpha-1)^2/\sum k_\alpha(k_\alpha-1)\}]\ ,\\ a_6=&2(n/\sum k_\alpha)[\{\sum k_\alpha(k_\alpha-1)/\sum k_\alpha\}\rho_{ss}^2\\&-2\{\sum k_\alpha(k_\alpha-1)(k_\alpha-2)/\sum k_\alpha(k_\alpha-1)\}\rho_{ss}-3]\\&+2\{n/\sum k_\alpha(k_\alpha-1)\}[\rho_{ss}^{-2}+2\{\sum k_\alpha(k_\alpha-1)(k_\alpha-2)/\sum k_\alpha(k_\alpha-1)\}\rho_{ss}^{-1}\\&+\{\sum k_\alpha(k_\alpha-1)(k_\alpha^2-3k_\alpha+3)/\sum k_\alpha(k_\alpha-1)\}]\ .\end{aligned}$$

Here $\sum$ stands for $\sum_{\alpha=1}^{N}$. Because of a number of siblings being finite, it may be noted that in (3.2) $(n/\sum k_\alpha)$, $(n\sum k_\alpha/\sum_{\alpha\neq\beta} k_\alpha k_\beta)$, $\{\sum k_\alpha(k_\alpha-1)/\sum k_\alpha\}$, and so forth, are bounded as n tends to infinity.

The second moments of $\hat{\rho}_{ms,p}$, $\hat{\rho}_{ms,e}$ and $\hat{\rho}_{ms,cs}$ around ρ_{ms} are asymptotically

$$(3.3)\qquad \mathrm{E}\,[(\hat{\rho}_{ms,p}-\rho_{ms})^2]=\frac{1}{n}\rho_{ms}^2\{a_7+(\sum k_\alpha^2/\sum k_\alpha)a_8\}(n\sum k_\alpha/\sum_{\alpha\neq\beta} k_\alpha k_\beta)\ ,$$

$$(3.4)\qquad \mathrm{E}\,[(\hat{\rho}_{ms,e}-\rho_{ms})^2]=\frac{1}{n}\rho_{ms}^2(c_1a_7+a_8)\ ,$$

$$\begin{aligned}(3.5)\qquad \mathrm{E}\,[(\hat{\rho}_{ms,cs}-\rho_{ms})^2]=\frac{1}{n}\rho_{ms}^2\Big[&-\frac{11}{4}+c_2^{-1}\rho_{ms}^2+c_2\rho_{ms}^{-2}\\&+\frac{1}{4}c_2^{-2}\{3\rho_{ss}^2+6c_1\rho_{ss}(1-\rho_{ss})+(c_1^2+2N^{-1}\sum k_\alpha^{-2})\\&\times(1-\rho_{ss})^2\}+(1-c_1)c_2^{-1}\rho_{ss}\\&\times\Big\{a_3+\frac{1}{4}(1-c_1)c_2^{-1}\rho_{ss}a_6\Big\}\Big]\ ,\end{aligned}$$

where

$$(3.6)\qquad \begin{aligned}a_7&=-\frac{3}{2}+2\rho_{ss}-\frac{1}{2}\rho_{ss}^2+(1-\rho_{ss})\rho_{ms}^{-2}\ ,\\ a_8&=-\frac{1}{2}-2\rho_{ss}+\frac{1}{2}\rho_{ss}^2+\rho_{ms}^2+\rho_{ss}\rho_{ms}^{-2}\ .\end{aligned}$$

From the results derived above, it follows that the pairwise, ensemble and modified sib-mean estimators are asymptotically normally distributed with means ρ_{ms} and variances (3.3), (3.4) and (3.5), respectively. The accuracy of these approximations may be improved by the use of the results in (3.1) which give the approximate biases in the estimators. For $k_\alpha = k$, $\alpha = 1, \cdots, N$, the maximum likelihood estimate for ρ_{ms} derived by Rosner et al. [8] is approximately normal with mean ρ_{ms} and variance $n^{-1}\rho_{ms}^2(a_7/k + a_8)$ (see also Elston [1]).

3.2. *Tests for interclass correlation*

In practice it is of interest to test the hypothesis H_0: $\rho_{ms} = 0$ against H_1: $\rho_{ms} > 0$. For this testing problem under the familial data set-up, we consider test statistics of the following form;

$$\hat{\rho}_{ms,\cdot} / \{\widehat{AV}(\hat{\rho}_{ms,\cdot})\}^{1/2}$$

where $\widehat{AV}(\hat{\rho}_{ms,\cdot})$ is a consistent estimate of $AV(\hat{\rho}_{ms,\cdot})$, the asymptotic variance of each estimator $\hat{\rho}_{ms,\cdot}$. For example, when $\rho_{ms} = 0$, it follows from (3.3) that the asymptotic variance of the pairwise estimator $\hat{\rho}_{ms,p}$ reduces to

$$\frac{1}{n}\left\{1 + \rho_{ss}\left(\frac{\sum k_\alpha^2}{\sum k_\alpha} - 1\right)\right\} \frac{n \sum k_\alpha}{\sum_{\alpha \neq \beta} k_\alpha k_\beta} .$$

The unknown parameter ρ_{ss} is estimated by a consistent estimator $\hat{\rho}_{ss}$ given by (2.3). So H_0 is rejected if $\hat{\rho}_{ms,p} / \{\widehat{AV}(\hat{\rho}_{ms,p})\}^{1/2} > z_{1-\alpha}$, where $z_{1-\alpha}$ is $100(1-\alpha)$ percentile point of a standard normal distribution.

For a one-to-one and continuously differentiable function f, an asymptotic distribution of $f(\hat{\rho}_{ms,\cdot})$ for each estimator $\hat{\rho}_{ms,\cdot}$ is

$$\{f(\hat{\rho}_{ms,\cdot}) - f(\rho_{ms})\} / [\{AV(\hat{\rho}_{ms,\cdot})\}^{1/2} f'(\rho_{ms})]$$

where $AV(\hat{\rho}_{ms,\cdot})$ is given by (3.3), (3.4) or (3.5). It can be seen from the expression of asymptotic variance that the usual test procedure based on Fisher's [2] z-transformation for the estimator is not appropriate for testing problem concerning interclass correlation. In general, the variance stabilizing transformation for each estimator does not seem to be expressible as an elementary function even for a fixed ρ_{ss}. The relation between variance stabilizing and normalizing transformations of certain statistics was shown by Konishi [3], [4].

4. Numerical results

For various values of the mean number of siblings per family we compare the estimators for ρ_{ms} discussed in Section 2, using the asymptotic mean square errors (3.3), (3.4) and (3.5). For the random-sib estimator, we use the well-known asymptotic result that

(4.1) $$E[(\hat{\rho}_{ms,r}-\rho_{ms})^2]=\frac{1}{n}(1-\rho_{ms}^2) .$$

As in Rosner et al. [7], [8], the sibship sizes k_α are randomly generated from a negative binomial distribution $NB(k, p)$; $\{(x+k-1)!p^k(1-p)^x\}/\{(k-1)!x!\}$. In Table 1 the asymptotic results (3.3), (3.4) and (4.1) are examined by comparing with the simulation results due to Rosner et al. ([8], p. 183, Table 1), although slightly different random numbers for siblings seem to be used. In the present paper the mean number of siblings generated from a negative binomial distribution $NB(2.84, 0.517)$ truncated so that $1\leqq x\leqq 15$ is 3.18. In Table 1 the values in blackets are the mean square errors obtained from Monte Carlo simulation (Rosner et al. [8]). The notations PA, EN, RS and MS used in Tables 1, 2 and 3 refer to the pairwise, ensemble, random-sib and modified sib-mean estimators, respectively.

Table 1. Comparison between asymptotic mean square errors and simulation results

ρ_{ms}/ρ_{ss}		.0	.1	.3	.5	.8
	PA	.0065(.0060)	.0086(.0079)	.0128(.0120)	.0170(.0162)	.0234(.0232)
.0	EN	.0094(.0093)	.0105(.0104)	.0127(.0125)	.0149(.0148)	.0182(.0183)
	RS	.0204(.0208)	.0204(.0210)	.0204(.0212)	.0204(.0215)	.0204(.0218)
	PA		.0083(.0081)	.0125(.0122)	.0166(.0164)	.0229(.0232)
.1	EN		.0102(.0103)	.0124(.0125)	.0146(.0147)	.0178(.0181)
	RS		.0200(.0204)	.0200(.0207)	.0200(.0211)	.0200(.0215)
	PA		.0063(.0070)	.0099(.0107)	.0135(.0146)	.0190(.0206)
.3	EN		.0083(.0089)	.0101(.0107)	.0120(.0128)	.0149(.0158)
	RS		.0169(.0167)	.0169(.0174)	.0169(.0180)	.0169(.0185)
	PA			.0057(.0069)	.0083(.0100)	.0125(.0148)
.5	EN			.0064(.0069)	.0077(.0086)	.0099(.0111)
	RS			.0115(.0115)	.0115(.0123)	.0115(.0130)
	PA					.0023(.0033)
.8	EN					.0020(.0024)
	RS					.0026(.0030)

The fit of approximations is quite good. Further, we find that the comparison of estimators based on asymptotic mean square errors gives the same result as that on the simulation.

Tables 2 and 3 compare asymptotic mean square errors ($\times 10^4$) of the estimators in the various mean number of siblings, for which MNS stands for the mean of sibship sizes k_α ($\alpha=1,\cdots,50$) generated from a negative binomial distribution $NB(k, p)$. The values in parentheses are asymptotic mean square errors of the modified sib-mean estimator hav-

Table 2. Comparison of asymptotic mean square errors ($\times 10^4$) of estimators in various mean number of siblings for $N=50$

MNS		$\rho_{ss}=0.1$	$\rho_{ss}=0.3$	$\rho_{ss}=0.5$	$\rho_{ss}=0.7$	$\rho_{ss}=0.9$
			$\rho_{ms}=0.1$			
2.22	PA	111	154	197	240	283
$NB(3, 0.7)$	EN	140	153	167	180	193
	MS	140(139)	153(153)	167(166)	180(180)	193(193)
3.30	PA	79	118	157	195	234
$NB(9, 0.7)$	EN	97	120	143	166	189
	MS	97(96)	120(119)	143(142)	166(165)	189(188)
5.74	PA	56	101	145	190	235
$NB(5, 0.5)$	EN	62	93	123	154	185
	MS	62(61)	93(91)	124(122)	154(153)	185(185)
8.08	PA	45	88	131	175	218
$NB(9, 0.5)$	EN	54	86	119	151	184
	MS	54(52)	86(85)	119(118)	151(151)	184(184)
	RS	200	200	200	200	200
			$\rho_{ms}=0.3$			
2.22	PA	87	123	161	199	237
$NB(3, 0.7)$	EN	116	127	139	151	163
	MS	114(111)	127(123)	140(136)	152(149)	164(162)
3.30	PA	60	93	127	161	197
$NB(9, 0.7)$	EN	78	98	117	138	158
	MS	77(69)	98(90)	118(112)	139(135)	160(157)
5.74	PA	40	78	117	156	197
$NB(5, 0.5)$	EN	48	74	100	127	155
	MS	46(35)	75(63)	102(93)	130(123)	157(154)
8.08	PA	31	68	105	144	183
$NB(9, 0.5)$	EN	40	68	96	125	154
	MS	41(30)	69(57)	97(88)	126(120)	155(153)
	RS	169	169	169	169	169

ing a true value of ρ_{ss}. We select the combinations of (ρ_{ms}, ρ_{ss}) such that $\rho_{ms}^2<\rho_{ss}$ so as to satisfy the condition that the covariance matrix Σ_a given by (2.1) to be positive definite. We note from (4.1) that asymptotic mean square errors of the random-sib estimator depend only upon the values of ρ_{ms}.

The tables show that for the lower mean sibship sizes the pairwise estimator is superior to the other estimators for $\rho_{ms}<0.3$ and $\rho_{ss}<0.3$. For $\rho_{ss}\geqq 0.3$, the ensemble and modified sib-mean estimators are superior to the pairwise estimator. For the higher values of ρ_{ss}, the pairwise estimator is not so good, but as the mean number of siblings increases,

Table 3. Comparison of asymptotic mean square errors ($\times 10^4$) of estimators in various mean number of siblings for $N=50$

MNS		$\rho_{ms}=0.5$				$\rho_{ms}=0.7$		
		$\rho_{ss}=0.3$	$\rho_{ss}=0.5$	$\rho_{ss}=0.7$	$\rho_{ss}=0.9$	$\rho_{ss}=0.5$	$\rho_{ss}=0.7$	$\rho_{ss}=0.9$
2.22 $NB(3, 0.7)$	PA	74	100	128	159	39	52	70
	EN	83	92	100	110	40	44	50
	MS	82(74)	93(85)	103(96)	113(109)	43(31)	50(39)	56(48)
3.30 $NB(9, 0.7)$	PA	54	78	104	131	29	42	58
	EN	61	75	90	106	31	38	47
	MS	62(45)	78(63)	93(83)	109(104)	36(17)	44(29)	54(44)
5.74 $NB(5, 0.5)$	PA	43	70	99	131	24	39	57
	EN	43	61	82	103	23	33	46
	MS	47(22)	68(45)	88(72)	109(100)	35(6)	45(21)	56(42)
8.08 $NB(9, 0.5)$	PA	36	62	91	121	21	35	53
	EN	38	58	80	103	21	31	45
	MS	42(19)	62(42)	83(70)	106(99)	29(6)	39(20)	51(41)
	RS	115	115	115	115	53	53	53

its asymptotic mean square error is closer to those of the ensemble and modified sib-mean estimators. We compared the modified sib-mean estimator having $\hat{\rho}_{ss}$ estimated by (2.3) with that having a true value of ρ_{ss}. These two types of situation yield the differences between asymptotic mean square errors of $\hat{\rho}_{ms,cs}$ and those in parentheses. Taking another estimator for ρ_{ss} may improve the asymptotic mean square error of $\hat{\rho}_{ms,cs}$ for the higher values of ρ_{ms}. The random-sib estimator is not effective, because of the loss of information which arises from choosing only one siblings per family. Similar comparisons were made for $N=100$, 200 and we found the results described above to be unchanged.

Appendix. Derivation of asymptotic moments

The derivation of the asymptotic moments of the estimators is outlined in the case of the pairwise estimator $\hat{\rho}_{ms,p}$. Under the assumption that a sample is drawn from a $(k_\alpha+1)$-variate normal distribution with mean vector $(\mu_m, \mu_s, \cdots, \mu_s)'$ and covariance matrix Σ_α given by (2.1), we have

$$\mathrm{E}\left[\sum_{\alpha=1}^{N} k_\alpha (x_{1\alpha}-\tilde{x}_m)^2\right]=N_k\sigma_m^2 ,$$

$$\mathrm{E}\left[\sum_{\alpha=1}^{N}\sum_{i=2}^{k_\alpha+1}(x_{i\alpha}-\tilde{x}_s)^2\right]=(\sum k_\alpha-1)\sigma_s^2-\{\sum k_\alpha(k_\alpha-1)/\sum k_\alpha\}\rho_{ss}\sigma_s^2 ,$$

$$\mathrm{E}\left[\sum_{\alpha=1}^{N}\sum_{i=2}^{k_\alpha+1}(x_{1\alpha}-\tilde{x}_m)(x_{i\alpha}-\tilde{x}_s)\right]=N_k\sigma_{ms}\ ,$$

where $N_k=\sum_{\alpha\neq\beta}^{N} k_\alpha k_\beta/\sum k_\alpha$. Let

$$K_m=\sum_\alpha k_\alpha(x_{1\alpha}-\tilde{x}_m)^2/N_k\ ,\qquad K_s=\sum_\alpha\sum_{i=2}^{k_\alpha+1}(x_{i\alpha}-\tilde{x}_s)^2/N_k\ ,$$

$$K_{ms}=\sum_\alpha\sum_{i=2}^{k_\alpha+1}(x_{1\alpha}-\tilde{x}_m)(x_{i\alpha}-\tilde{x}_s)/N_k\ .$$

Noting that $\sum k_\alpha/N_k=1+(\sum k_\alpha^2/\sum k_\alpha)/N_k$ and $\sum k_\alpha^2/\sum k_\alpha=O(1)$, the statistics K_m, K_s and K_{ms} converge to σ_m^2, σ_s^2 and σ_{ms}, respectively, as N tends to infinity, that is, as $N_k\to+\infty$. So it is possible to write the pairwise estimator as

$$\text{(A.1)}\qquad \hat{\rho}_{ms,p}=\rho_{ms}\left(1+\frac{K_{ms}-\sigma_{ms}}{\sigma_{ms}}\right)\left\{\left(1+\frac{K_m-\sigma_m^2}{\sigma_m^2}\right)\left(1+\frac{K_s-\sigma_s^2}{\sigma_s^2}\right)\right\}^{-1/2}.$$

Expanding (A.1) gives

$$\hat{\rho}_{ms,p}=\rho_{ms}\left\{1+\left(K_{ms}^*-\frac{1}{2}K_m^*-\frac{1}{2}K_s^*\right)+\left(\frac{3}{8}K_m^{*2}+\frac{3}{8}K_s^{*2}\right.\right.$$
$$\left.\left.+\frac{1}{4}K_m^*K_s^*-\frac{1}{2}K_{ms}^*K_s^*-\frac{1}{2}K_{ms}^*K_m^*\right)+\text{higher order terms}\right\}\ ,$$

where $K_{ms}^*=(K_{ms}-\sigma_{ms})/\sigma_{ms}$, $K_m^*=(K_m-\sigma_m^2)/\sigma_m^2$ and $K_s^*=(K_s-\sigma_s^2)/\sigma_s^2$. Calculating each expectation of $\hat{\rho}_{ms,p}-\rho_{ms}$ and $(\hat{\rho}_{ms,p}-\rho_{ms})^2$, we can obtain the final results given in Subsection 3.1. Calculations are laborious, because of the complication of varying family size.

For the ensemble estimator $\hat{\rho}_{ms,e}$, we may find that

$$\mathrm{E}\left[\sum_{\alpha=1}^{N}(x_{1\alpha}-\bar{x}_m)(\bar{x}_{s\alpha}-\bar{x}_s)\right]=n\sigma_{ms}\ ,\qquad \mathrm{E}\left[\sum_{\alpha=1}^{N}(x_{1\alpha}-\bar{x}_m)^2\right]=n\sigma_m^2\ ,$$

$$\mathrm{E}\left[\frac{n}{N}\sum_{\alpha=1}^{N}\sum_{i=2}^{k_\alpha+1}(x_{i\alpha}-\bar{x}_{s\alpha})^2/k_\alpha+\sum_{\alpha=1}^{N}(\bar{x}_{s\alpha}-\bar{x}_s)^2\right]=n\sigma_s^2\ .$$

By an argument similar to that discussed above, we obtain the expansion of $\hat{\rho}_{ms,e}$. Calculating each expectation of the resulting formula, we have the results given in Subsection 3.1.

The modified sib-mean estimator $\hat{\rho}_{ms,cs}$ contains the estimator $\hat{\rho}_{ss}$ given by (2.3). Expectations of the numerator and denominator of $\hat{\rho}_{ss}$, which we put K_{ss} and K_{0s}, respectively, are

$$\mathrm{E}\,[K_{ss}]=\rho_{ss}\sigma_s^2-\sigma_s^2/\sum k_\alpha-\rho_{ss}\sigma_s^2\{2\sum k_\alpha(k_\alpha-1)^2/\sum k_\alpha(k_\alpha-1)$$
$$-\sum k_\alpha(k_\alpha-1)/\sum k_\alpha\}/\sum k_\alpha\ ,$$

$$E[K_{0s}]=\sigma_s^2-\sigma_s^2\{1+\rho_{ss}\sum k_\alpha(k_\alpha-1)/\sum k_\alpha\}/\sum k_\alpha \,.$$

Hence we can expand $\hat{\rho}_{ss}$ around ρ_{ss} in the form of

$$\hat{\rho}_{ss}=\rho_{ss}\{1+(K_{ss}^*-K_{0s}^*)+(K_{0s}^{*2}-K_{ss}^*K_{0s}^*)+\cdots\}$$

where $K_{ss}^*=(K_{ss}-\rho_{ss}\sigma_s^2)/\rho_{ss}\sigma_s^2$ and $K_{0s}^*=(K_{0s}-\sigma_s^2)/\sigma_s^2$.

Noting that the leading term of the Taylor series expansion of $\{c_1+(1-c_1)\hat{\rho}_{ss}\}^{1/2}$ is $\{c_1+(1-c_1)\rho_{ss}\}^{1/2}$, we can similarly expand $\{c_1+(1-c_1)\cdot\rho_{ss}\}^{1/2}\hat{\rho}_{ms,s}$. Combining these results yields the expansion of $\hat{\rho}_{ms,cs}$, from which we have, after a calculation of each expectation, the final results given in Subsection 3.1. Details are omitted, because of pressure on space.

Acknowledgement

The author would like to thank the referee for valuable comments and suggestions and Mrs. Sato for carrying out the computer analyses.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] Elston, R. C. (1975). On the correlation between correlations, *Biometrika*, **62**, 133-140.
[2] Fisher, R. A. (1921). On the "probable error" of a coefficient of correlation deduced from a small sample, *Metron*, **1**, 1-32.
[3] Konishi, S. (1978). An approximation to the distribution of the sample correlation coefficient, *Biometrika*, **65**, 654-656.
[4] Konishi, S. (1981). Normalizing transformations of some statistics in multivariate analysis, *Biometrika*, **68**, 647-651.
[5] Mak, T. K. and Ng, K. W. (1981). Analysis of familial data: Linear-model approach, *Biometrika*, **68**, 457-461.
[6] Rosner, B. (1979). Maximum likelihood estimation of interclass correlations, *Biometrika*, **66**, 533-538.
[7] Rosner, B. and Donner, A. (1979). Significance testing of interclass correlations from familial data, *Biometrics*, **35**, 461-471.
[8] Rosner, B., Donner, A. and Hennekens, C. H. (1977). Estimation of interclass correlation from familial data, *Appl. Statist.*, **26**, 179-187.