

A simple estimator for correlation in incomplete data

W. Albers*, M.F. Teulings

Department of Applied Mathematics, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

Received October 1995; revised February 1996

Abstract

It seems fair that the usual sample correlation coefficient should allow improvement when additional samples from the marginals are available. However, some intuitive attempts fail badly. Using control variates, a simple method is presented which asymptotically achieves the optimal improvement.

Keywords: Missing data; Control variates; Asymptotic relative efficiency

Mathematics subject classification: 62F12; 62H20

1. Introduction

We consider the situation of a bivariate normal distribution: $(X_1, X_2) \cong N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. Suppose we have a sample of size n on (X_1, X_2) , and in addition samples of size m_i on $X_i, i = 1, 2$. In other words, the total sample size may be $n + m_1 + m_2$, but we are handicapped by missing data with respect to both X_1 and X_2 . It is in this situation that we are interested in estimating $\rho = \rho(X_1, X_2)$.

Intuitively, it is clear that one should be able to do better here than with just the complete data subset of size n : the additional separate observations on X_i should help. Of course, this intuition is correct. Nevertheless, as we shall point out by, e.g., referring to Johnson and Kotz (1972), one is easily led astray in trying to incorporate this additional information and may actually wind up by doing worse rather than better.

To make quite clear what goes wrong in such efforts and how this should be remedied, we shall begin by considering in Section 2 the complete data case with known means and variances. Note that this case fits in the framework above by letting $m_i \rightarrow \infty, i = 1, 2$. The key notion is that the sample variances act as control variates for the sample covariance. Hence they should not be replaced simply by the actual variances even if these are known. Such a replacement would actually lead to a worse estimator, rather than to a better one. Optimal application of the control variate approach does lead to the desired improvement, however.

In Section 3 we return to the incomplete data case and demonstrate how the results from the previous section can be extended to the situation. The simple estimator proposed here is linked to the existing literature

* Corresponding author.

in Section 4. Moreover, some simulations are discussed, as well as a problem we recently encountered and which serves as an illustration for the use of methods like the one outlined here.

2. The complete data case

Here we just have a complete sample of size n on (X_1, X_2) from the previous section. Then (cf. e.g. Johnson and Kotz 1972, p. 104) the MLE of ρ is given by

$$\tilde{R} = \tilde{S}_{12} / (\tilde{S}_1 \tilde{S}_2), \quad (1)$$

where $\tilde{S}_i^2 = (n-1)^{-1} \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2$, $\bar{X}_i = n^{-1} \sum_{j=1}^n X_{ij}$, $i = 1, 2$ and $\tilde{S}_{12} = (n-1)^{-1} \sum_{j=1}^n (X_{1j} - \bar{X}_1)(X_{2j} - \bar{X}_2)$.

As is well-known (cf. e.g. Serfling 1980, p. 125), $n^{1/2}(\tilde{R} - \rho)$ is asymptotically normal with mean zero and variance $\tau^2 = \lim_{n \rightarrow \infty} (n \text{ var } \tilde{R})$. Comparison of \tilde{R} to possible competitors in case of known means and variances will be based on this variance τ^2 of the limiting distribution. This enables us to simplify matters in the following sense. It is straightforward to verify that τ^2 remains the same to first order if we replace the sample means \bar{X}_i by μ_i , $i = 1, 2$, in (1). Hence, the effect of the means being either known or unknown, is negligible. Consequently, in the sequel we shall simply assume μ_i to be known. Without loss of generality, these known values can both be assumed to be zero. Hence instead of \tilde{R} from (1) we can take as our starting point the more simple

$$R = S_{12} / (S_1 S_2), \quad (2)$$

where $S_i^2 = n^{-1} \sum_{j=1}^n X_{ij}^2$ and $S_{12} = n^{-1} \sum_{j=1}^n X_{1j} X_{2j}$.

For the variances, matters are less straightforward: it now makes a difference whether these are known or not. In fact, this precisely motivates the consideration of possible alternatives for R . Nevertheless, in analyzing the behavior of these estimators, we can always, without loss of generality, assume that $\sigma_i^2 = 1$, $i = 1, 2$. Hence, we shall simply let $(X_1, X_2) \cong N(0, 0, 1, 1, \rho)$ and, depending on the situation, assume the fact that $\sigma_i^2 = 1$ to be known or not.

First consider the standard case where the σ_i^2 are unknown and R is the MLE for ρ (cf. again Johnson and Kotz 1972, p. 104), Letting $V = S_{12} - \rho$ and $U_i = S_i^2 - 1$, $i = 1, 2$, we obtain that

$$R = (\rho + V) \prod_{i=1}^2 (1 + U_i)^{-1/2} \doteq \rho + V - \frac{1}{2} \rho \sum_{i=1}^2 U_i, \quad (3)$$

where “ \doteq ” stands for equality to first order. It is easy to verify that

$$nEV^2 = 1 + \rho^2, \quad nEVU_i = 2\rho, \quad nE \left(\sum_{i=1}^2 U_i \right)^2 = 4(1 + \rho^2). \quad (4)$$

This in its turn implies that $\tau^2 \doteq n \text{ var } (R) \doteq n \text{ MSE } (R) \doteq n \{ EV^2 - \rho \sum_{i=1}^2 EVU_i + \frac{1}{4} \rho^2 E(\sum_{i=1}^2 U_i)^2 \} = (1 - \rho^2)^2$, which agrees with Johnson and Kotz (1972, p. 104).

Suppose now, the fact that $\sigma_i^2 = 1$ is known. At first sight it may seem indicated in this situation to replace R from (2) by

$$R(0) = S_{12} / (\sigma_1 \sigma_2) = S_{12}. \quad (5)$$

But from (3) and (4), it is immediate that the variance of the limiting distribution of $n^{1/2}(R(0) - \rho)$ equals $1 + \rho^2$. This is considerably larger, instead of smaller, than $(1 - \rho^2)^2 = \tau^2$.

The explanation for this phenomenon lies in the fact that the S_i act as control variates, and as such are more useful than the known values σ_i . Briefly, let T_i be unbiased estimators for $\theta_i, i = 1, 2$, with $\hat{\rho} = \rho(T_1, T_2) \neq 0$. If θ_2 is known, T_1 can be improved upon by $T_3 = T_1 + \gamma(T_2 - \theta_2)$ as an estimator for σ_1 by choosing the coefficient γ of the control variate in the right way. In fact, it is straightforward that for $\gamma = -\hat{\rho}\sigma(T_1)/\sigma(T_2)$ we achieve the minimal value $\sigma^2(T_3) = \sigma^2(T_1)(1 - \hat{\rho}^2) < \sigma^2(T_1)$.

Application of this technique here leads to the proposal:

$$R(\gamma) = S_{12} / \left\{ \sigma_1 \sigma_2 \prod_{i=1}^2 (S_i^2 / \sigma_i^2)^{\gamma/2} \right\}. \tag{6}$$

Note that $R(1) = R$ from (2), the sole possible choice for unknown σ_1 , whereas $R(0)$ in (6) indeed agrees with $R(0)$ from (5). Using the fact that the known values of σ_i are both equal to one, $R(\gamma)$ reduces to $S_{12} / \{ \prod_{i=1}^2 S_i^2 \}^{\gamma/2}$, which in analogy to (3) satisfies $R(\gamma) \doteq \rho + V - \frac{1}{2}\gamma\rho \sum_{i=1}^2 U_i$. Together with (4), this immediately leads to an expression for the variance of the limiting distribution of $n^{1/2}(R(\gamma) - \rho)$, which is easily seen to be minimized for

$$\gamma^* = \frac{2}{1 + \rho^2}, \quad \text{with } n \text{ var } (R(\gamma^*)) \doteq \frac{(1 - \rho^2)^2}{1 + \rho^2}. \tag{7}$$

As $1 \leq \gamma^* \leq 2$, it is seen that the known σ_i are not used to replace the S_i , but rather to add more weight to these control variates. Indeed the choice $R(0)$ literally goes into the wrong direction. To apply the method, we need to replace γ^* from (7) by its estimated counterpart $\hat{\gamma}^* = 2/(1 + R^2(0))$. It is again straightforward to verify that this change is negligible as far as the asymptotic variance is concerned.

From (7) it follows that

$$ARE(R(\hat{\gamma}^*), R) = 1 + \rho^2, \tag{8}$$

which shows that the improvement can be quite worthwhile. To put the result in (8) in its proper perspective, it is desirable to compare the performance of $R(\gamma^*)$ to that of an efficient estimator such as the MLE. In fact, Johnson and Kotz (1972) on p. 105 discuss the MLE for this case of known parameters for the marginal distributions. However, it turns out to be rather complicated, requiring the solution of a cubic equation. Therefore, Johnson and Kotz also give a first-order version which they advocate for many practical purposes. Now observe that the estimator corresponding to $\hat{\gamma}^*$ mentioned above satisfies

$$R(\hat{\gamma}^*) \doteq R(0) \left\{ 1 - (\hat{\gamma}^*/2) \left(\sum_{i=1}^2 (S_i^2 - 1) \right) \right\} = R(0) + \frac{R(0)}{(1 + R^2(0))} (2 - S_1^2 - S_2^2). \tag{9}$$

But this latter expression precisely agrees with the aforementioned first-order version of the MLE from Johnson and Kotz. Hence, $R(\hat{\gamma}^*)$ not only improves on $R = R(1)$ as shown in (8), but moreover it is asymptotically efficient.

A final remark concerns the following. Application of MLEs automatically ensures that the estimators of ρ do not fall outside the parameter space $[-1, 1]$. If desired, a similar restriction can be imposed on first-order versions like $R(\hat{\gamma}^*)$ or the one in (9).

3. Missing observations

After considering the limiting case in the previous section, we now address the general case, where the variances are not assumed to be known, but we have samples of size m_i on $X_i, i = 1, 2$, in addition to the complete sample of size n on (X_1, X_2) . Just as in Section 2, the feeling arises that R from (2) should

be improved upon, this time by utilizing the information for the $(m_1 + m_2)$ additional single observations. However, in this general case, even more care is needed in such attempts. For example, Johnson and Kotz (1972, p. 108) mention a possibility, but more or less discard it again by pointing out that its possible increase of accuracy must be judged against the possibility of obtaining values outside $[-1, 1]$.

We shall use the following setup and notation. Suppose we have X_{ij} for $i = 1, 2$, $j = 1, \dots, n$ and in addition for $i = 1$, $j = n + 1, \dots, n + m_1$ and $i = 2$, $j = n + m_1 + 1, \dots, n + m_1 + m_2$. Let

$$\sum_j = \sum_{j=1}^2, \quad \sum_j^{(1)} = \sum_{j=n+1}^{n+m_1}, \quad \sum_j^{(2)} = \sum_{j=n+m_1+1}^{n+m_1+m_2}, \quad \lambda_i = \frac{n}{n+m_i}, \quad i = 1, 2. \quad (10)$$

For our purpose we are typically interested in constant λ_i (apart from rounding effects). Hence, we shall suppose that m_i , $i = 1, 2$, and n tend to infinity at a common rate. From Section 2 it is clear that we should not use the additional observations to simply replace S_i , $i = 1, 2$ by new sample variances based on $(n + m_i)$ observations. For larger values of m_i/n , the resulting estimator would behave more and more like $R(0)$, which was shown to be an extremely bad choice.

Fortunately, Section 2 also suggests which is the right direction to take: we should mimic the behavior of $R(\gamma)$ from (6). In addition to $S_i^2 = n^{-1} \sum_{j=1}^n X_{ij}^2$, let $(S_i')^2 = m_i^{-1} \sum_j^{(i)} X_{ij}^2$ and $U_i = S_i^2 - 1$, $U_i' = (S_i')^2 - 1$, $i = 1, 2$. We replace (6) by

$$R(\gamma_1, \gamma_2) = S_{12} / \{(S_1')^{(1-\gamma_1)} S_1^{\gamma_1} (S_2')^{(1-\gamma_2)} S_2^{\gamma_2}\}. \quad (11)$$

By analogy to (3) we obtain that this $R(\gamma_1, \gamma_2) \doteq \rho + V - \frac{1}{2}\rho(\sum_{i=1}^2 \{(1 - \gamma_i)U_i' + \gamma_i U_i\})$. To evaluate the variance of the limiting distribution of $n^{1/2}(R(\gamma_1, \gamma_2) - \rho)$, some results are needed in addition to (4). To begin with, we observe that $EVU_i' = EU_i U_i' = EU_1' U_2' = 0$. Moreover, $nEU_i'^2 = 2$, $nEU_1 U_2 = 2\rho^2$ and $E(U_i')^2 = 2/m_i = 2\lambda_i/(1 - \lambda_i)$. Together with (11), this shows that $n \text{ var } R(\gamma_1, \gamma_2) \doteq$

$$(1 + \rho^2) - 2\rho^2 \sum_{i=1}^2 \gamma_i + \rho^2 \left\{ \gamma_1 \gamma_2 \rho^2 + \frac{1}{2} \sum_{i=1}^2 (\gamma_i^2 + (1 - \gamma_i)^2 \lambda_i / (1 - \lambda_i)) \right\}. \quad (12)$$

The expression in (12) is easily seen to be minimized for

$$\gamma_i^* = \frac{2 - \lambda_i - \rho^2(1 - \lambda_i)(2 - \lambda_{3-i})}{1 - \rho^4(1 - \lambda_1)(1 - \lambda_2)}, \quad (13)$$

$i = 1, 2$, whereas the corresponding minimal value satisfies

$$n \text{ var } R(\gamma_1^*, \gamma_2^*) \doteq \frac{(1 - \rho^2)^2}{1 - \rho^4(1 - \lambda_1)(1 - \lambda_2)} \left\{ 1 - \frac{1}{2}\rho^2(2 - \lambda_1 - \lambda_2) \right\}. \quad (14)$$

Note that for the special case of equal numbers of additional observations, (13) and (14) simplify to

$$\gamma^* = \frac{2 - \lambda}{1 + \rho^2(1 - \lambda)}, \quad n \text{ var } R(\gamma^*, \gamma^*) = \frac{(1 - \rho^2)^2}{1 + \rho^2(1 - \lambda)}, \quad (15)$$

with $\lambda_1 = \lambda_2 = \lambda$ and $\gamma_1^* = \gamma_2^* = \gamma^*$. Also observe that the variance expression(s) in (14) and (15) produce the values $\tau^2 = (1 - \rho^2)^2$ and $(1 - \rho^2)^2 / (1 + \rho^2)$ for $\lambda_i = 1$ and $\lambda_i = 0$, respectively. Hence, the result in (14) indeed fills the gap between the two extremes considered in Section 2. If only a small fraction of additional observations is available, the λ_i are close to 1, the γ_i^* from (13) are near one, and the variance of the limiting distribution is close to τ^2 . Likewise, for relatively large numbers of additional observations, the λ_i are close to 0 and we approach the result for the asymptotic variance from (7).

To make the estimator applicable in practice, we replace γ_i^* from (13) by $\hat{\gamma}_i^*$ by substituting R for ρ . The difference caused is again negligible and it follows that (8) generalizes to

$$ARE(R(\hat{\gamma}_1^*, \hat{\gamma}_2^*), R) = \frac{1 - \rho^4(1 - \lambda_1)(1 - \lambda_2)}{1 - \frac{1}{2}\rho^2(2 - \lambda_1 - \lambda_2)}, \tag{16}$$

which reduces to $ARE(R(\hat{\gamma}^*, \hat{\gamma}^*), R) = 1 + \rho^2(1 - \lambda)$ for $\lambda = \lambda_1 = \lambda_2$. This surprisingly simple relation adequately makes clear the gain to be achieved by the optimal use of additional information.

4. Concluding remarks

In searching the literature for related approaches, we came across Hocking and Smith (1968) (see also Smith, 1968). Here initial estimators are based on the complete data, after which these estimators are improved step-wise by taking the subsequent groups of incomplete observations into account. At each step, the improvement is achieved by adding a suitable multiple of the difference between the estimator of a parameter based on the complete data, and the estimator of the same parameter based on the group of incomplete data considered in this step. As $R(\gamma_1, \gamma_2)$ from (11) satisfies $R(\gamma_1, \gamma_2) \doteq R(1, 1)(1 - \frac{1}{2} \sum_{i=1}^2 (1 - \gamma_i)\{(S_i')^2 - S_i^2\})$, a strong resemblance to the method suggested by Hocking and Smith (1968) suggests itself. Indeed, explicit evaluation reveals that both methods are in fact equivalent to first order and, hence, in particular, produce the same efficiency figures. As the actual evaluation is tedious but straightforward, we refer to Teulings (1994) for details. In comparing the two methods, we note that the method we propose here is much easier to implement and to understand.

A second topic we briefly mention is a small simulation study, performed to obtain an impression of how the first-order approximations work out for finite sample sizes. Again we refer to Teulings (1994) for full details. Of course, we now work again with \tilde{R} , \tilde{S}_i^2 and \tilde{S}_{12} from (1), rather than with R , S_i^2 and S_{12} from (2). Likewise, we apply $(\tilde{S}_i')^2 = (m_i - 1)^{-1} \sum_j^{(i)} (X_{ij} - \bar{X}_i')^2$, where $\bar{X}_i' = m_i^{-1} \sum_j^{(i)} X_{ij}$.

In Table 1 we present results for $\rho = 0.70$, $n = 20$ and $n = 100$ complete observations, and equal numbers $m_1 = m_2 = m$ of additional observations, running from $m = 10$ to $m = 2000$. Using 1000 simulations on each run, we obtain estimates $\hat{E}\tilde{R}$, $\widehat{\text{var}} \tilde{R}$, $\hat{E}\tilde{R}(\hat{\gamma}^*, \hat{\gamma}^*)$ and $\widehat{\text{var}} \tilde{R}(\hat{\gamma}^*, \hat{\gamma}^*)$ (where $\tilde{R}(\hat{\gamma}^*, \hat{\gamma}^*)$ is obtained from $R(\hat{\gamma}^*, \hat{\gamma}^*)$ replacing S_i by \tilde{S}_i and S_i' by \tilde{S}_i' , $i = 1, 2$). Another reason for performing the simulations is to compare the present estimator to the one mentioned above, which was introduced by Hocking and Smith (1968). Hence in Table 1 we also list the values $\hat{E}\tilde{R}_{H\&S}$ and $\widehat{\text{var}} \tilde{R}_{H\&S}$, obtained for this estimator. Finally, the values based on the asymptotics to which we want to compare are $AV_1 = n^{-1}\tau^2 = (1 - \rho^2)^2/n$ and $AV_2 = (1 - \rho^2)^2\{1 - \frac{1}{2}\rho^2(2 - \lambda_1 - \lambda_2)\}/\{(1 - \rho^4(1 - \lambda_1)(1 - \lambda_2))n\}$ (cf. (14)), respectively. These are given as well: AV_1 for comparison to $\widehat{\text{var}} \tilde{R}$ and AV_2 for both $\widehat{\text{var}} \tilde{R}(\hat{\gamma}^*, \hat{\gamma}^*)$ and $\widehat{\text{var}} \tilde{R}_{H\&S}$.

From Table 1 we observe that the value $n = 20$ is still quite small: the values of $\widehat{\text{var}} \tilde{R}$ do not get very close to AV_1 . As a consequence, there also remains a bit of a gap between $\widehat{\text{var}} \tilde{R}(\hat{\gamma}^*, \hat{\gamma}^*)$ and $\widehat{\text{var}} \tilde{R}_{H\&S}$ on the one hand, and AV_2 on the other. For $n = 100$, the picture is much better: the values of $\widehat{\text{var}} \tilde{R}$ are quite close to $AV_1 = 0.0026$. Moreover, the behavior of the other two variances is now adequately approximated by AV_2 . Also note that the figures for $n = 100$, nicely illustrate the result on the ARE in (16), according to which the efficiency values AV_1/AV_2 in this example are $1 + 0.49m/(100 + m)$. Finally, observe that there is little difference in performance between $\tilde{R}(\hat{\gamma}^*, \hat{\gamma}^*)$ and $\tilde{R}_{H\&S}$. Hence, the simulations suggest no preference for either approach. Thus, we uphold the point of view expressed above, according to which the present method has the advantage of being more simple, both with respect to computing and to understanding it.

Table 1
Simulation results

m	$\hat{E}\tilde{R}$	$\widehat{\text{var}} \tilde{R}$	$\hat{E}\tilde{R}(\hat{\gamma}^*, \hat{\gamma}^*)$	$\widehat{\text{var}} \tilde{R}(\hat{\gamma}^*, \hat{\gamma}^*)$	$\hat{E}\tilde{R}_{H\&S}$	$\widehat{\text{var}} \tilde{R}_{H\&S}$	AV_2
$n = 20, AV_1 = 0.013$							
10	0.693	0.018	0.689	0.016	0.692	0.016	0.011
20	0.692	0.016	0.699	0.014	0.694	0.014	0.010
50	0.694	0.015	0.709	0.012	0.695	0.012	0.0096
100	0.698	0.016	0.715	0.011	0.697	0.012	0.0092
200	0.682	0.017	0.711	0.012	0.687	0.013	0.0090
500	0.694	0.015	0.720	0.0098	0.693	0.011	0.0088
1000	0.691	0.017	0.716	0.012	0.689	0.013	0.0088
2000	0.693	0.016	0.717	0.011	0.691	0.012	0.0087
$n = 100, AV_1 = 0.0026$							
10	0.697	0.0028	0.697	0.0027	0.700	0.0027	0.0025
20	0.701	0.0028	0.699	0.0023	0.701	0.0026	0.0024
50	0.700	0.0028	0.701	0.0023	0.701	0.0023	0.0022
100	0.698	0.0031	0.702	0.0025	0.701	0.0025	0.0021
200	0.700	0.0027	0.704	0.0020	0.701	0.0020	0.0020
500	0.700	0.0027	0.705	0.0020	0.701	0.0020	0.0019
1000	0.700	0.0027	0.707	0.0018	0.702	0.0019	0.0018
2000	0.702	0.0027	0.706	0.0019	0.702	0.0019	0.0018

The final topic we raise here concerns possible applications. One could argue of course that estimating correlation is so common that examples are superfluous. Nevertheless, it may be advisable to illustrate that missing observations in the sense considered here – not just a small fraction of randomly missing observations, but a considerable subsample on which only the other variable is present – occur in practice in a natural way. In fact, we recently came across precisely this type of situation in connection with the problem of setting tests limits.

In large-scale production processes (e.g. ICs in semiconductor industry), all parts are compared to a specification limit s . Due to measurement error, one is forced to accept only those products which pass a slightly more strict test limit t . Selecting t is a delicate matter: if t is too close to s , the outgoing quality may be too low, but if t is too conservative, the producer faces an unnecessary loss of yield. In Albers et al. (1994a, b), accurate methods for setting test limits have been derived. In many applications, however, the situation is more complicated: instead of direct measurements, be it with a measurement error, only so-called correlated measurements are available. The idea is that direct measurements are often expensive or even destructive and that it is therefore preferable to measure a strongly related characteristic which is more easily accessible. Some examples occurring in literature concern voltage at an internal point of an electronic device, together with voltage at some surface point, or the fracture stress of a material, together with the distribution of a component in the material that can be observed by X ray (see Tang, 1987).

Let X_1 and X_2 denote the direct and the correlated measurement, respectively. Then a sample from (X_1, X_2) enables us to estimate means, variances and correlation. Using this, accurate test limits can be derived for this situation as well (see Albers et al., 1995). Quite often, however, the test sample of size n from (X_1, X_2) will form part of a considerably larger production sample of size N from X_2 : throughout production as a routine the simple correlated measurements are taken. To be able to set test limits, a subsample has to be selected for which the direct measurement is obtained as well.

Note that this situation indeed provides the kind of example we are looking for: we have a complete data set of size n , with $m_1 = 0$ additional observations for X_1 and $m_2 = N - n$ additional ones on X_2 . Hence, $\lambda_1 = 1$, $\lambda_2 = n/N$ and $\gamma_1^* = 1$, $\gamma_2^* = 1 + (1 - \rho^2)(1 - \lambda_2)$, while (16) reduces to $ARE(R(1, \gamma_2^*), R) = 1 / \{1 - \frac{1}{2}\rho^2(1 - \lambda_2)\}$.

References

- Albers, W., W.C.M. Kallenberg and G.D. Otten (1994a), Accurate test limits with estimated parameters, *Technometrics* **36**, 92–101.
- Albers, W., W.C.M. Kallenberg and G.D. Otten (1994b), Setting test limits under prescribed consumer loss, *Metrika* **41**, 163–181.
- Albers, W., G.R.J. Arts and W.C.M. Kallenberg (1995), Test limits using correlated measurements, Technical Report 1260, University of Twente.
- Hocking, R.R. and W.B. Smith (1968), Estimation of parameters in the multivariate normal distribution with missing observations, *J. Amer. Statist. Ass.* **63**, 159–173.
- Johnson, N.L. and S. Kotz (1972), *Distributions in Statistics: Continuous Multivariate distributions* (Wiley, New York).
- Serfling, R.J. (1980), *Approximation Theorem of Mathematical Statistics* (Wiley, New York).
- Smith, W.B. (1968), Bivariate samples with missing values II, *Technometrics* **10**, 867–868.
- Tang, K. (1987), Economic design of a one-sided screening procedure using a correlated variable, *Technometrics* **29**, 477–485.
- Teulings, M.F. (1994), Estimation of the correlation coefficient in a bivariate normal sample with missing values, Master's Thesis, University of Twente.