Hacettepe Journal of Mathematics and Statistics
Volume 42 (2) (2013), 189 – 197

MULTIVARIATE ESTIMATION FROM "TWO VARIABLES AT A TIME" OBSERVATIONS

Christopher S. Withers* and Saralees Nadarajah†

Received 10 : 01 : 2012 : Accepted 24 : 04 : 2012

Abstract

Suppose that we wish to estimate the mean $\boldsymbol{\mu}$ and the covariance $\mathbf{C}$ of a random p-vector $\mathbf{X}$ with $p > 2$, but we can only sample from the vector $\mathbf{X}$ two of its p components at a time. We give both nonparametric estimates and the maximum likelihood estimates (MLEs) under normality, and their covariances.

Keywords: Maximum likelihood estimation; Multivariate normal; Nonparametric estimation.

2000 AMS Classification: 62F10

1. Introduction and summary

Suppose $\mathbf{X}$ is a p-variate random vector with $p > 2$. There are many situations in the sciences and engineering, where one can deal with only two of the components of $\mathbf{X}$ at a time. Some examples are:

(1) In some electrical engineering problems, one has only two ports for communication at any one time (Davidovitz, 1995; Premoli and Storace, 2004; Zhao *et al.*, 2008);
(2) Much research in political science involves analyzing just one or two variables at a time (McNabb, 2004);
(3) Most experimental designs measures only one or two variables at a time;
(4) Health research depends on laboratory tests on animals, in which only one or two variables at a time are tested;
(5) In Physics, some gas laws relate only two variables at a time.

*Applied Mathematics Group, Industrial Research Limited, Lower Hutt, NEW ZEALAND email kit.withers@gmail.com

†School of Mathematics, University of Manchester, Manchester M13 9PL, UK. E-mail: mbbsssn2@manchester.ac.uk

In each of these situations, the practitioner wants to know how the population mean and population covariance of $\mathbf{X}$ can be estimated using the limited data. For example, the practitioner may want to estimate the mean and covariance of signals received at all ports of a communication network when data are available only from two of the ports at a time.

The aim of this note is to present estimation procedures for data of the above kind. We estimate both the population mean and population covariance, that is

$$\boldsymbol{\mu} = E\mathbf{X}, \ \mathbf{C} = E\ (\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})',$$

assuming that $\mathbf{C}$ is positive definite. Two estimation procedures are presented: 1) by assuming normality (normal maximum likelihood estimation procedure); 2) by assuming no specific distribution (nonparametric estimation procedure).

We take a sample, of size n_{ij} say, from $\binom{X_i}{X_j}$, say $\binom{X_{i1}}{X_{j1}} \cdots \binom{X_{im}}{X_{jm}}$ with $m = n_{ij}$, for all possible pairs of rows $1 \leq i < j \leq p$. In addition, we may sometimes have a sample of size $n_{ii} \geq 0$ from X_i alone, say $X_{i1}, \ldots, X_{im}$ with $m = n_{ii}$. So, the total number of observations of row i is

$$n_i = \sum_{j=1}^{p} n_{ij},$$

where $n_{ji} = n_{ij}$, and the total sample size is

$$n = \sum_{1 \leq i \leq j \leq p} n_{ij}.$$

Set $n_0 = \min_{1 \leq i < j \leq p} n_{ij}$.

The results of this note are organized as follows. In Section 2, we give empirical-type nonparametric estimates $\widetilde{\boldsymbol{\mu}}$, $\widetilde{\mathbf{C}}$ and their covariance. In Section 3, we assume normality and give the MLEs $\widehat{\boldsymbol{\mu}}$, $\widehat{\mathbf{C}}$ and their covariances. An example comparing the efficiencies of the two estimates by simulation and by analytical means is given in Section 4. Finally, some future work are noted in Section 5.

2. Nonparametric estimates

The natural estimates are

$$\widetilde{\mu}_i = n_i^{-1} \sum_{j=1}^{p} \sum_{k=1}^{n_{ij}} X_{ik},$$

$$\widetilde{C}_{ii} = n_i^{-1} \sum_{j=1}^{p} \sum_{k=1}^{n_{ij}} \left(X_{ik} - \widetilde{\mu}_i\right)^2,$$

$$\widetilde{C}_{ij} = n_{ij}^{-1} \sum_{k=1}^{n_{ij}} \left(X_{ik} - \widetilde{\mu}_i\right)\left(X_{jk} - \widetilde{\mu}_j\right)$$

for $i \neq j$. One would not estimate C_{ii} by $n_i^{-1} \sum_{j=1}^{p} \sum_{k=1}^{n_{ij}} (X_{ik}^2 - \widetilde{\mu}_i^2)$ since this may be negative with probability $O(\exp(-n_0\lambda_1))$ (Anderson, 2003), where $\lambda_1 > 0$. The estimate $\widetilde{\boldsymbol{\mu}}$ is unbiased and

$$\text{var}\ \widetilde{\mu}_i = n_i^{-1} C_{ii}, \quad \text{covar}\,(\widetilde{\mu}_i, \widetilde{\mu}_j) = n_i^{-1} n_j^{-1} n_{ij} C_{ij} \tag{2.1}$$

for $i \neq j$. So,

$$\text{corr}\ \left(\widetilde{\mu}_i, \widetilde{\mu}_j\right) = \rho_{ij} n_{ij} / \left(n_i n_j\right)^{1/2},$$

where $\rho_{ij} = \text{corr}\ (X_i, X_j)$. Note that $\widetilde{\mathbf{C}}$ may have negative eigenvalues with probability $O(\exp(-n\lambda_2))$, where $\lambda_2 > 0$. To avoid this one can if desired replace $\widetilde{\mathbf{C}} = \mathbf{H}\boldsymbol{\Lambda}\mathbf{H}'$

say, where $\mathbf{HH}' = \mathbf{I}_p$ and $\mathbf{\Lambda} = \mathrm{diag}(\lambda_i)$ by $\widetilde{\mathbf{C}}_+ = \mathbf{H\Lambda}_+\mathbf{H}'$, where $\mathbf{\Lambda}_+ = \mathrm{diag}(\lambda_{i+})$ and $\lambda_{i+} = \max(\lambda_i, 0)$. This increases the "energy" of the system on average by only $O(\exp(-n_0\lambda_2))$. Set

$$\mu_{ij\cdots} = E\ (X-\mu)_i (X-\mu)_j \cdots .$$

So, $\mu_{ij} = C_{ij}$. Using the influence functions method one can show that the covariances of $\widetilde{\boldsymbol{\mu}}$, $\widetilde{\mathbf{C}}$ are given by

$$\mathrm{covar}\left(\widetilde{\mu}_i, \widetilde{C}_{ij}\right) \approx n_i^{-1}\mu_{iij},$$

$$\mathrm{covar}\left(\widetilde{\mu}_i, \widetilde{C}_{jj}\right) \approx n_i^{-1}n_j^{-1}n_{ij}\mu_{ijj},$$

$$\mathrm{covar}\left(\widetilde{\mu}_i, \widetilde{C}_{jk}\right) = 0,$$

$$\mathrm{var}\ \widetilde{C}_{ii} \approx n_i^{-1}\left(\mu_{iiii} - \mu_{ii}^2\right), \tag{2.2}$$

$$\mathrm{covar}\left(\widetilde{C}_{ii}, \widetilde{C}_{ij}\right) \approx n_i^{-1}\left(\mu_{iiij} - \mu_{ii}\mu_{ij}\right), \tag{2.3}$$

$$\mathrm{covar}\left(\widetilde{C}_{ii}, \widetilde{C}_{jj}\right) \approx n_i^{-1}n_j^{-1}n_{ij}\left(\mu_{iijj} - \mu_{ii}\mu_{jj}\right), \tag{2.4}$$

$$0 = \mathrm{covar}\left(\widetilde{C}_{ii}, \widetilde{C}_{jk}\right) = \mathrm{covar}\left(\widetilde{C}_{ii}, \widetilde{C}_{jl}\right) = \mathrm{covar}\left(\widetilde{C}_{ij}, \widetilde{C}_{kl}\right) \approx \mathrm{covar}\left(\widetilde{C}_{ij}, \widetilde{C}_{ik}\right).$$

This is because for $F_{ij} = \mathcal{L}(X_i, X_j)$, $\widetilde{\boldsymbol{\mu}}$ and $\widetilde{\mathbf{C}}$ have influence functions (partial first functional derivatives)

$$\mu_i\left(x, F_{ij}\right) = n_i^{-1}n_{ij}\left(x_i - \mu_i\right),$$
$$C_{ij}\left(x, F_{ij}\right) = \left(x_i - \mu_i\right)\left(x_j - \mu_j\right) - C_{ij},$$
$$C_{ii}\left(x, F_{ij}\right) = n_i^{-1}n_{ij}\left\{\left(x_i - \mu_i\right)^2 - C_{ii}\right\}$$

and by equation (3.4) of Withers (1988), for $\mathbf{T}(\{F_{ij}\})$, $k \times 1$,

$$\mathrm{covar}\ \mathbf{T}\left(\left\{\widehat{F}_{ij}\right\}\right) \approx \sum_{ij} n_{ij}^{-1} \int \mathbf{T}\left(x, F_{ij}\right)\mathbf{T}\left(x, F_{ij}\right)' dF_{ij}(x).$$

The other combinations are obtained by replacing the indices i, j, k, l by $1, \ldots, p$.

3. Normal data

Here, we suppose that $\mathbf{X}$ is normal. So, $\mu_{ijk} = 0$ making $\widetilde{\boldsymbol{\mu}}$, $\widetilde{\mathbf{C}}$ asymptotically independent, and by (2.2)-(2.4),

$$\mathrm{var}\ \widetilde{C}_{ii} \approx 2n_i^{-1}C_{ii}^2,$$
$$\mathrm{covar}\left(\widetilde{C}_{ii}, \widetilde{C}_{ij}\right) \approx 2n_i^{-1}C_{ii}C_{ij},$$
$$\mathrm{covar}\left(\widetilde{C}_{ii}, \widetilde{C}_{jj}\right) \approx 2n_i^{-1}n_j^{-1}C_{ij}^2.$$

However, the MLEs $\widehat{\boldsymbol{\mu}}$, $\widehat{\mathbf{C}}$ will in general have smaller mean square error than these. Their disadvantage is that they must be found iteratively as we now express them as implicit solutions of equations of the form

$$\boldsymbol{\mu} = \mathbf{f}(\mathbf{C}),\ \mathbf{C} = \mathbf{g}(\boldsymbol{\mu}, \mathbf{C}). \tag{3.1}$$

For $i \neq j$, $\binom{X_i}{X_j}$ has covariance

$$\mathbf{C}_{(ij)} = \begin{pmatrix} C_{ii} & C_{ij} \\ C_{ij} & C_{jj} \end{pmatrix}$$

with inverse

$$\mathbf{C}_{(ij)}^{-1} = d_{ij}^{-1} \begin{pmatrix} C_{jj} & -C_{ij} \\ -C_{ij} & C_{ii} \end{pmatrix} = \left(C_{ij}^{ab} \right)$$

say, for $a, b = i, j$, where $d_{ij} = \det \mathbf{C}_{(ij)} = C_{ii}C_{jj} - C_{ij}^2$. Set $Y_{ak} = X_{ak} - \mu_a$, $\delta_{ab} = 1$ if $a = b$ and $\delta_{ab} = 0$ if $a \neq b$. The log likelihood of the sample of size n_{ij} is

$$L_{ij} = -n_{ij} \log 2\pi - 2^{-1} n_{ij} \log d_{ij} - 2^{-1} \sum_{k=1}^{n_{ij}} \begin{pmatrix} Y_{ik} \\ Y_{jk} \end{pmatrix}' \mathbf{C}_{(ij)}^{-1} \begin{pmatrix} Y_{ik} \\ Y_{jk} \end{pmatrix}$$

for $i \neq j$ and

$$L_{ii} = -2^{-1} n_{ii} \log 2\pi - 2^{-1} \log C_{ii} - 2^{-1} \sum_{k=1}^{n_{ii}} Y_{ik}^2 / C_{ii}.$$

The log likelihood of the combined sample is

$$L = \sum_{1 \leq i \leq j \leq p} L_{ij}.$$

This is maximized by μ_i when

$$0 = \partial L / \partial \mu_i = \gamma_i - \sum_{j=1}^{p} \beta_{ij} \mu_j,$$

where

$$\begin{aligned} \gamma_i &= \sum_{j \neq i} d_{ij}^{-1} \sum_{k=1}^{n_{ij}} (X_{ik} C_{jj} - X_{jk} C_{ij}) + C_{ii}^{-1} \sum_{k=1}^{n_{ii}} X_{ik}, \\ \beta_{ii} &= \sum_{j \neq i} n_{ij} d_{ij}^{-1} C_{jj} + n_{ii} C_{ii}^{-1}, \\ \beta_{ij} &= -n_{ij} d_{ij}^{-1} C_{ij} \end{aligned}$$

for $i \neq j$. So, the MLE of $\boldsymbol{\mu}$ given $\mathbf{C}$ is

$$\widehat{\boldsymbol{\mu}}(\mathbf{C}) = \boldsymbol{\beta}^{-1} \boldsymbol{\gamma}$$

for $\det \boldsymbol{\beta} \neq 0$. This gives the first of the two equations of (3.1). For given $\mathbf{C}$, $\widehat{\boldsymbol{\mu}}(\mathbf{C})$ is unbiased, but $\widehat{\boldsymbol{\mu}}(\widehat{\mathbf{C}})$ may be biased since the usual MLE $\widehat{\mathbf{C}}$ of $\mathbf{C}$ is biased (Anderson, 2003). Since

$$\partial / \partial x \log |\mathbf{C}| = \text{trace } \mathbf{C}^{-1} \partial \mathbf{C} / \partial x$$

and

$$\partial \mathbf{C}_{(ij)}^{-1} / \partial C_{ii} = d_{ij}^{-2} \begin{pmatrix} -C_{jj}^2 & C_{ij} C_{jj} \\ C_{ij} C_{jj} & -C_{ij}^2 \end{pmatrix}$$

for $i \neq j$, L is maximized by C_{ii} when

$$0 = \partial L / \partial C_{ii} = 2^{-1} C_{ii}^{-2} \sum_{k=1}^{n_{ii}} D_{iik} + 2^{-1} \sum_{j \neq i} d_{ij}^{-2} \sum_{k=1}^{n_{ij}} \left(C_{jj}^2 D_{iik} - 2 C_{jj} C_{ij} D_{ijk} + C_{ij}^2 D_{jjk} \right),$$

where $D_{ijk} = Y_{ik} Y_{jk} - C_{ij}$ for sample n_{ii} or n_{ij} as appropriate. This can be written as

$$C_{ii} = N_i(\boldsymbol{\mu}, \mathbf{C}) / D_i(\boldsymbol{\mu}, \mathbf{C}) = g_{ii}(\boldsymbol{\mu}, \mathbf{C}) \tag{3.2}$$

say, where

$$D_i(\boldsymbol{\mu}, \mathbf{C}) = \sum_{j\neq i} n_{ij} d_{ij}^{-2} C_{jj}^2,$$

$$N_i(\boldsymbol{\mu}, \mathbf{C}) = n_{ii} C_{ii}^{-2}\{\left(Y_i^2\right)_{ii} - C_{ii}\}$$

$$+\sum_{j\neq i} n_{ij} d_{ij}^{-2}\left[C_{jj}^2\left(Y_i^2\right)_{ij} - 2C_{jj}C_{ij}\{(Y_iY_j)_{ij} - C_{ij}\} + C_{ij}^2\{\left(Y_j^2\right)_{ij} - C_{jj}\}\right],$$

$$(Y_aY_b)_{ij} = n_{ij}^{-1}\sum_{k=1}^{n_{ij}} Y_{ak}Y_{bk}.$$

Since $\partial C_{(ij)}^{ij}/\partial C_{ij} = -d_{ij}^{-2}(C_{ii}C_{jj} + C_{ij}^2)$, L is maximized by C_{ij} when

$$0 = \partial L/\partial C_{ij} = -d_{ij}^{-2}\sum_{k=1}^{n_{ij}}\{C_{ij}C_{jj}D_{iik} - \left(C_{ii}C_{jj} + C_{ij}^2\right)D_{ijk} + C_{ii}C_{ij}D_{jjk}\},$$

which we can write as

$$C_{ij}^2 + A_{ij}C_{ij} + C_{ii}C_{jj} = 0,$$

where $A_{ij} = \{d_{ij} - (Y_i^2)_{ij}C_{jj} - (Y_j^2)_{ij}C_{ii}\}/(Y_iY_j)_{ij}$. Since $(Y_aY_b)_{ij} \to C_{ab}$ almost surely as $n_{ij} \to \infty$, the consistent root is

$$C_{ij} = -A_{ij}/2 - \left(A_{ij}^2/4 - C_{ii}C_{jj}\right)^{1/2} = g_{ij}(\boldsymbol{\mu}, \mathbf{C}) \tag{3.3}$$

say. We see that (3.2), (3.3) give the second equation in (3.1). These can be solved by iteration: $\boldsymbol{\mu}_{(i+1)} = \mathbf{f}(\mathbf{C}_{(i)})$ and $\mathbf{C}_{(i+1)} = \mathbf{g}(\boldsymbol{\mu}_{(i)}, \mathbf{C}_{(i)})$, starting from $\boldsymbol{\mu}_{(0)} = \widetilde{\boldsymbol{\mu}}$ and $\mathbf{C}_{(0)} = \widetilde{\mathbf{C}}$ of Section 2.

It is not clear if the right hand side of (3.3) can be non-real (complex), or if this is ruled out by $|\rho_{ij}| \leq 1$, $|\widehat{\rho}_{ij}| \leq 1$, where $\rho_{ij} = \text{corr}(X_i, X_j)$. If it can be complex, then this will occur with probability $O(\exp(-\lambda n_{ij}))$, where $\lambda > 0$. If the right hand side of (3.3) is complex, it should be replaced by $-A_{ij}/2$.

These "normal MLEs" $(\widehat{\boldsymbol{\mu}}, \widehat{\mathbf{C}})$ will still be consistent as $n_0 \to \infty$ if $\mathbf{X}$ is not normal.

As $n_0 \to \infty$,

$$\widehat{\boldsymbol{\theta}} \dot{\sim} \mathcal{N}_q\left(\boldsymbol{\theta}, \mathbf{I}_n(\boldsymbol{\theta})^{-1}\right),$$

where $\boldsymbol{\theta}' = (\boldsymbol{\mu}', \text{vech } \mathbf{C}')$, vech $\mathbf{C} = (C_{11}, \ldots, C_{1p}, C_{22}, \ldots, C_{2p}, \ldots, C_{pp})$ is $r \times 1$, $r = p(p+1)/2$, $q = r + p$, $\widehat{\boldsymbol{\theta}}$ is the corresponding MLE, and $\mathbf{I}_n(\boldsymbol{\theta})$ is Fisher's information matrix given by $\mathbf{I}_n(\boldsymbol{\theta}) = E\ \partial L/\partial\boldsymbol{\theta}\partial L/\partial\boldsymbol{\theta}'$. The elements of $\partial L/\partial\boldsymbol{\theta}$ were given above. So,

$$\mathbf{I}_n(\boldsymbol{\theta}) = \begin{bmatrix} \langle\boldsymbol{\mu}\boldsymbol{\mu}'\rangle & \langle\boldsymbol{\mu}\text{ vech }\mathbf{C}'\rangle \\ \langle\boldsymbol{\mu}'\text{ vech }\mathbf{C}\rangle & \langle\text{vech }\mathbf{C}\text{ vech }\mathbf{C}'\rangle \end{bmatrix},$$

where, for example,

$$\langle\boldsymbol{\mu}\boldsymbol{\mu}'\rangle = E\ \partial L/\partial\boldsymbol{\mu}\ \partial L/\partial\boldsymbol{\mu}' = E\ (\boldsymbol{\gamma} - \boldsymbol{\beta}\boldsymbol{\mu})(\boldsymbol{\gamma} - \boldsymbol{\beta}\boldsymbol{\mu})'$$

with ii element

$$\sum_{j\neq i} n_{ij} d_{ij}^{-1} C_{jj} + n_{ii} C_{ii}^{-1}$$

and ij element $-n_{ij}d_{ij}^{-1}C_{ij}$ for $i \neq j$. Also $\langle\boldsymbol{\mu}\text{ vech }\mathbf{C}'\rangle = \mathbf{0}$ since it is a sum of third central moments. So, $\widehat{\boldsymbol{\mu}}$, $\widehat{\mathbf{C}}$ are asymptotically independent, and

$$\widehat{\boldsymbol{\mu}} \dot{\sim} \mathcal{N}_p\left(\boldsymbol{\mu}, \langle\boldsymbol{\mu}\boldsymbol{\mu}'\rangle^{-1}\right),\ \text{vech } \widehat{\mathbf{C}} \dot{\sim} \mathcal{N}_r\left(\text{ vech } \mathbf{C}, \langle\text{ vech } \mathbf{C}\text{ vech } \mathbf{C}'\rangle^{-1}\right). \tag{3.4}$$

Note that if $\{n_{ii}\}$ are all positive, the diagonal of $\langle \boldsymbol{\mu}\boldsymbol{\mu}' \rangle$ is increased, increasing its eigenvalues; c.f. ridge regression. By construction $\mathbf{I}_n(\boldsymbol{\theta})$ is positive-definite. The elements of $\langle$ vech $\mathbf{C}$ vech $\mathbf{C}' \rangle$ needed for (3.4) are

$$\begin{aligned}
\langle C_{ii}^2 \rangle &= 2^{-1} \sum_{j \neq i} n_{ij} d_{ij}^{-2} C_{jj}^2 + 2^{-1} n_{ii} C_{ii}^{-2}, \\
\langle C_{ii} C_{ij} \rangle &= -n_{ij} d_{ij}^{-2} C_{ij} C_{jj}, \\
\langle C_{ii} C_{jj} \rangle &= 2^{-1} n_{ij} d_{ij}^{-4} C_{ij}^2 \left(C_{ii} C_{jj} d_{ij}^2 + C_{ij}^4 \right), \\
\langle C_{ii} C_{jk} \rangle &= \langle C_{ij} C_{ik} \rangle = \langle C_{ij} C_{kl} \rangle = 0.
\end{aligned}$$

3.1. Example. Suppose that $p = 3$, so vech $\mathbf{C} = (C_{11}, C_{12}, C_{13}, C_{22}, C_{23}, C_{33})'$. Then

$$\mathbf{I}_n \langle \text{ vech } \mathbf{C} \text{ vech } \mathbf{C}' \rangle = \begin{bmatrix}
\langle C_{11}^2 \rangle & \langle C_{11} C_{12} \rangle & \langle C_{11} C_{13} \rangle & \langle C_{11} C_{22} \rangle & 0 & \langle C_{11} C_{33} \rangle \\
 & \langle C_{12}^2 \rangle & 0 & \langle C_{12} C_{22} \rangle & 0 & 0 \\
 & & \langle C_{13}^2 \rangle & 0 & 0 & \langle C_{13} C_{33} \rangle \\
 & & & \langle C_{22}^2 \rangle & \langle C_{22} C_{23} \rangle & \langle C_{22} C_{33} \rangle \\
 & & & & \langle C_{23}^2 \rangle & \langle C_{23} C_{33} \rangle \\
 & & & & & \langle C_{33}^2 \rangle
\end{bmatrix}.$$

The "ridge regression" effect if $n_{11} n_{22} n_{33} \neq 0$ is confined to $\langle C_{ii}^2 \rangle$, that is only half of the diagonal elements are increased.

4. Comparison example

Suppose that $C_{ii} \equiv 1$, $C_{ij} \equiv \rho$ for $i \neq j$, $n_{ii} \equiv n_{11}$, $n_{ij} \equiv m$ for $i \neq j$. So, $n_i = n_{11} + mr$, where $r = p - 1$. By (2.1) the ij element of covar$(\widetilde{\boldsymbol{\mu}})$ is n_1^{-1} for $i = j$ and $n_1^{-2} m\rho$ for $i \neq j$. The asymptotic covariance of $\widehat{\boldsymbol{\mu}}$ has diagonal elements $(1-c)/(V_{11} - V_{12})$ and off diagonals $-c/(V_{11} - V_{12})$, where $\mathbf{V} = \langle \boldsymbol{\mu}\boldsymbol{\mu}' \rangle$, that is $V_{11} - V_{12} = n_{11} + m(r + \rho)/(1 - \rho^2)$ and $c = -m\rho(1-\rho)^{-1} \{ n_{11}(1+\rho) + mr \}^{-1}$. So, var $\widehat{\mu}_1 \approx (1-c)/(V_{11} - V_{12})$. So, $\widetilde{\mu}_1$ has asymptotic efficiency about

$$\begin{aligned}
\eta &= \text{var } \widehat{\mu}_1 / \text{var } \widetilde{\mu}_1 = n_1 (1-c) \left(V_{11} - V_{12}\right)^{-1} \\
&= 1 - n_{11}^{-1} m\rho (2 + r\rho)(1 - \rho^2)^{-1} + O\left(n_{11}^{-2}\right) \qquad (4.1)
\end{aligned}$$

for fixed m. So, for large n_{11} and moderate m, $\widetilde{\mu}_1$ has high efficiency. For $m/n_{11} = 0.1$, the efficiency in (4.1) is plotted in Figure 4.1. The range of ρ is restricted by

$$0 < \begin{vmatrix} 1 & & \rho \\ & \ddots & \\ & & 1 \end{vmatrix} = (1-\rho)^{n-1} (1 + \rho p - \rho)$$

that is $\rho > -1/(p-1)$. It is clear that (4.1) is a decreasing function of both p and ρ. So, the efficiency can be considered high if p is small and/or ρ is small. At $\rho = -(p-1)^{-1}$, the efficiency is $1 - n_{11}^{-1} m\rho(1-\rho^2)^{-1}$ which is also a decreasing function of ρ.

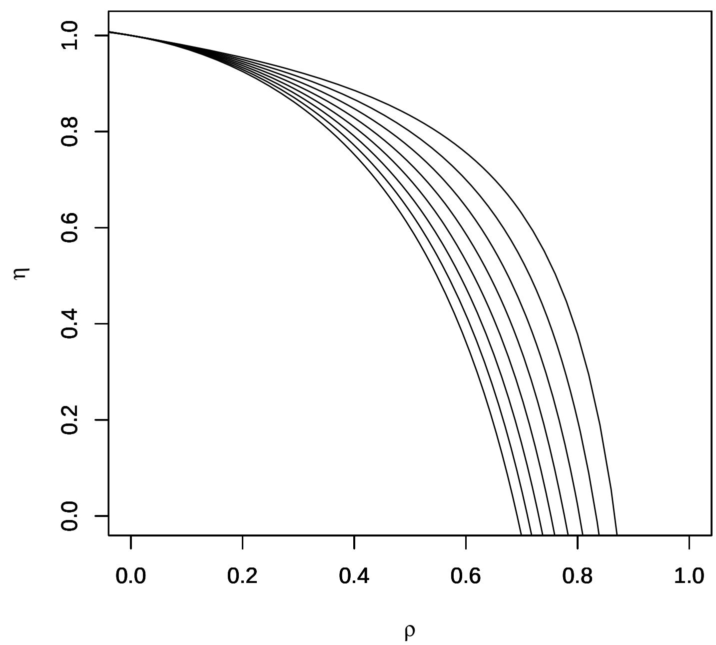
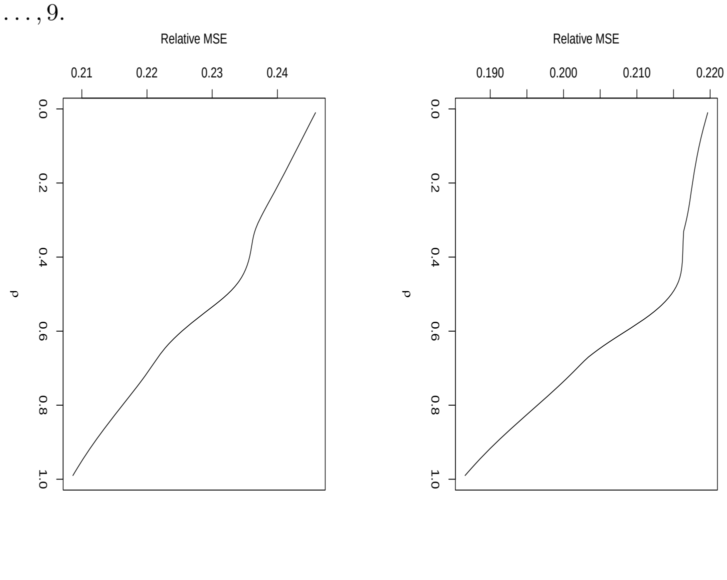


Figure 4.1 The efficiency, (4.1), versus ρ when n_{ij} are all equal, $C_{ii} \equiv 1$ and $C_{ij} \equiv \rho$ for $i \neq j$. The curves from the top to bottom correspond to the increasing values $p = 2, 3, \ldots, 9$.



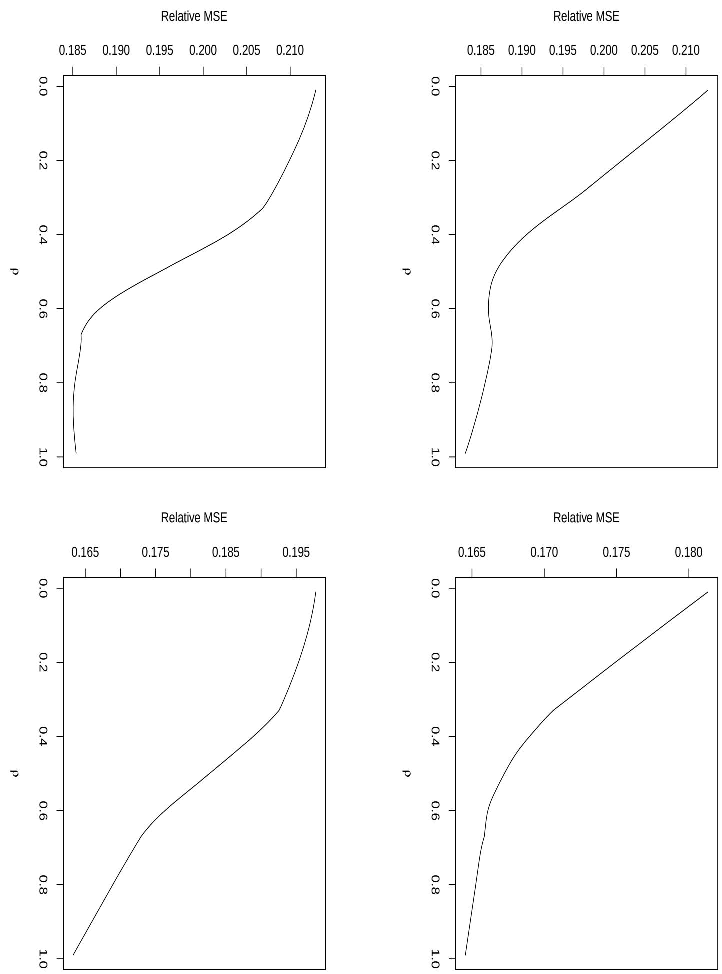


Figure 4.2 Relative mean squared error versus ρ for $m/n_{11} = 0.2$ (top left), $m/n_{11} = 0.4$ (top right), $m/n_{11} = 0.6$ (middle left), $m/n_{11} = 0.8$ (middle right), $m/n_{11} = 1$ (bottom left) and $m/n_{11} = 2$ (bottom right).

The above calculations are analytical and asymptotic. For better understanding about how $\widehat{\mu}_1$ and $\widetilde{\mu}_1$ compare, we now perform a simulation study. We use the following scheme:

(1) Suppose $(X_1, X_2, \ldots, X_p)$ has the p-variate normal distribution with zero means and covariance given by $C_{ii} \equiv 1$, $C_{ij} \equiv \rho$ for $i \neq j$;
(2) Simulate a sample of size m from $\binom{X_i}{X_j}$ for all $1 \leq i < j \leq p$;
(3) Simulate a sample of size n_{11} from X_i for all $1 \leq i \leq p$;
(4) Using the data in steps 2 and 3, estimate $\widehat{\mu}_1$ and $\widetilde{\mu}_1$;
(5) Repeat steps 2 to 4 ten thousand times;
(6) Compute the mean squared errors of $\widehat{\mu}_1$ and $\widetilde{\mu}_1$;
(7) Compute the relative mean squared error as the mean squared error of $\widehat{\mu}_1$ divided by that of $\widetilde{\mu}_1$.

We executed this scheme for $p = 3$, $\rho = 0.01, 0.02, \ldots, 0.99$ and $m/n_{11} = 0.2, 0.4, 0.6, 0.8, 1, 2$.

The relative mean squared errors are plotted in Figure 4.2. The actual values plotted are the lowess (Cleveland, 1979, 1981) smoothed versions versus ρ for $\rho = 0.01, 0.02, \ldots, 0.99$. While lowess smoothing, we used the default options. These are: a smoothing span of 2/3, three 'robustifying' iterations and the speed of computations determined by 0.01th of the range of the ρ values.

The following observations can be drawn from Figure 4.2:

(1) the relative mean squared error decreases with respect to ρ for all m/n_{11} even for $m/n_{11} > 1$;
(2) the relative mean squared errors are largest for $m/n_{11} = 0.2$;
(3) the relative mean squared errors are smallest for $m/n_{11} = 2$.

In the simulations, we have limited $p = 3$. But similar results hold for higher p. We have not presented them in order to avoid repetitive discussion.

5. Future work

The work of this note can be extended in several ways. One way is to consider estimation of $\boldsymbol{\mu}$ and $\mathbf{C}$ of when samples are available only from q of the components of $\mathbf{X}$, where $q < p$. Another work is to study properties of $\widetilde{\mathbf{C}}_+ = \mathbf{H}\boldsymbol{\Lambda}_+\mathbf{H}'$. Yet another work is to perform simulations and analytical calculations as in Section 4 to compare the efficiencies of $\widetilde{C}_{ij}$ and $\widehat{C}_{ij}$.

Acknowledgments

The authors would like to thank the Editor and the referee for careful reading and for their comments which greatly improved the paper.

References

[1] Anderson, T. W., *An Introduction to Multivariate Statistical Analysis*, third edition. John Wiley and Sons, New York, 2003.
[2] Cleveland, W. S., *Robust locally weighted regression and smoothing scatterplots*, Journal of the American Statistical Association, **74**, 829-836, 1979.
[3] Cleveland, W. S., *LOWESS: A program for smoothing scatterplots by robust locally weighted regression*, The American Statistician, **35**, 54, 1981.
[4] Davidovitz, M., *Reconstruction of the S-matrix for a 3-port using measurements at only two ports*, IEEE Microwave and Guided Wave Letters, **5**, 349-350, 1995.
[5] McNabb, D. E., *Research Methods for Political Science: Quantitative and Qualitative Methods*, M. E. Sharpe Publishers, London, 2004.
[6] Premoli, A. and Storace, M.,*Two-port ideal power transferitors: a unified introduction to ideal transformer and gyrator*, IEEE Transactions on Circuits and Systems II: Express Briefs, **51**, 426-429, 2004.
[7] Withers, C. S., *Nonparametric confidence intervals for functions of several distributions*, Annals of the Institute of Statistical Mathematics, **40**, 727-746, 1988.
[8] Zhao, Y., Yu, S. and Yang, Z., *Evolutionary mechanism of port clusters based on the theory of dissipative structure*, In: Proceedings of the 2008 International Conference on Intelligent Computation Technology and Automation, volume 1, pp. 52-55, 2008.