Electronic Journal of Statistics
Vol. 4 (2010) 161–171
ISSN: 1935-7524
DOI: 10.1214/10-EJS562

Stabilizing the asymptotic covariance of an estimate

Christopher Withers

Applied Mathematics Group
Industrial Research Limited
Lower Hutt, New Zealand

and

Saralees Nadarajah

School of Mathematics
University of Manchester
Manchester, United Kingdom
e-mail: mbbsssn2@manchester.ac.uk

Abstract: Suppose $n^{1/2}(\widehat{\theta}_n - \theta) \to \mathcal{N}_p(0, V(\theta))$ as $n \to \infty$ for some estimate $\widehat{\theta}_n$ of θ in R^p. If $p = 1$ and $g(\theta) = \int_0^\theta V(x)^{-1/2} dx$, it is well known that $n^{1/2}(g(\widehat{\theta}_n) - g(\theta)) \to \mathcal{N}(0,1)$ as $n \to \infty$, the distribution often being less skew so that inference based on the approximation $n^{1/2}(g(\widehat{\theta}_n) - g(\theta)) \sim \mathcal{N}(0,1)$ should be more accurate than inference based on the approximation $V(\widehat{\theta}_n)^{-1/2} n^{1/2}(\widehat{\theta}_n - \theta) \sim \mathcal{N}(0,1)$. If $p > 1$ there is generally no such one to one transformation $g(\cdot)$. We consider three different types of stabilization of $V(\theta)$. We also consider the problem of finding $g(\cdot)$ so that the components of $g(\widehat{\theta}_n)$ are asymptotically independent.

AMS 2000 subject classifications: Primary 62G08, 62G20.
Keywords and phrases: Asymptotic, confidence regions, distribution-free, estimates, influence function, multivariate normal, stabilizing.

Received January 2010.

Contents

1. Introduction and summary

Suppose we wish to make an inference (such as constructing a consistent test or confidence region (CR) for θ) on a parameter θ in R^p on the basis of knowing that for some estimate $\widehat{\theta}_n$, $n^{1/2}(\widehat{\theta}_n - \theta)$ is approximately (written $\dot{\sim}$) $\mathcal{N}_n(0, V(\theta, \psi))$ when n, the sample size, is large; where $\psi \in R^s$ is a nuisance parameter; $V(\cdot, \cdot)$ is a given function such that $V(\theta, \psi) > 0$ (positive definite) and $V(\cdot, \cdot)$ is continuous at (θ, ψ). By Slutsky's Theorem if $\widehat{\psi}_n \stackrel{p}{\to} \psi$ then as $n \to \infty$, $V(\widehat{\theta}_n, \widehat{\psi}_n) \stackrel{p}{\to} V(\theta, \psi)$ so that we may make our inference on

$$V\left(\widehat{\theta}_n, \widehat{\psi}_n\right)^{-1/2} n^{1/2} \left(\widehat{\theta} - \theta\right) \dot{\sim} \mathcal{N}_p\left(0, I_p\right). \tag{1.1}$$

On the other hand if we choose a transformation $g(\cdot) : R^p \to R^p$ which is one to one and continuous in a neighborhood of θ, then

$$n^{1/2} \left(g(\widehat{\theta}_n) - g(\theta)\right) \dot{\sim} \mathcal{N}_p\left(0, V_g(\theta, \psi)\right), \tag{1.2}$$

where

$$V_g(\theta, \psi) = \partial g(\theta)/\partial \theta' V(\theta, \psi) \partial g(\theta)'/\partial \theta \tag{1.3}$$

and $\partial g(\theta)/\partial \theta'$ is the $p \times p$ matrix with (i, j) element $\partial g_i(\theta)/\partial \theta_j$. So, we could base our inference on

$$V_g\left(\widehat{\theta}_n, \widehat{\psi}_n\right)^{-1/2} n^{1/2} \left(g(\widehat{\theta}) - g(\theta)\right) \dot{\sim} \mathcal{N}_p\left(0, I_p\right).$$

When $V(\theta, \psi)$ does not depend on ψ we denote it by $V(\theta)$, and likewise for V_g.

If $p = 1$ and $V(\theta, \psi) = V(\theta)$, then it is generally recommended that one choose

$$g(\theta) = \int_0^\theta V(x)^{-1/2} dx \tag{1.4}$$

since this yields $V_g(\theta) \equiv 1$ so that there is no error from estimating $V_g(\theta)$. This transformation also generally reduces skewness and so improves the approximation. So, for example, an asymptotically $1 - \alpha$ level CR for θ is

$$-S_{n\alpha} \le g\left(\widehat{\theta}_n\right) - g(\theta) \le S_{n\alpha}$$

provided $n^{1/2} S_{n\alpha} \to \Phi^{-1}(1 - \alpha/2)$ as $n \to \infty$.

Example 1.1. *For $\widehat{\theta}$ the sample variance from a normal population with variance θ, $V(\theta) = 2\theta^2$ so that (1.4) yields $g(\theta) = 2^{-1/2} \log \theta$. (With $\{S_{n\alpha}\}$ chosen appropriately, the above CR for $g(\theta)$ reduces to the usual exact CR for the variance of a normal population.)*

This transformation reduces $\gamma_1 = \mu_3 \mu_2^{-3/2}$ from $2^{3/2} n^{-1/2}(1 + o(1))$ for $\widehat{\theta}_n$ to $-2^{1/2} n^{-1/2}(1 + o(1))$ for $g(\widehat{\theta}_n)$). □

Example 1.2. *Let $\widehat{\theta}_n$ be the sample correlation from a bivariate normal population with correlation θ. Then $V(\theta) = (1 - \theta^2)^2$ for which (1.4) gives*

$$g(\theta) = 2^{-1} \log \frac{1 + \theta}{1 - \theta}.$$

This reduces γ_1 from $-6\theta n^{-1/2}(1+o(1))$ for $\widehat{\theta}_n$ to $\theta^3 n^{-3/2}(1+o(1))$ for $g(\widehat{\theta}_n)$. (See for instance pages 212, 216 of Hotelling [5] and equation (16.78) of Stuart and Ord [7]). □

In Example 1.2, the variance stabilizing transformation also reduces the skewness from $O(n^{-1/2})$ to $O(n^{-3/2})$. Withers and Nadarajah [8] show how to choose a transformation which will effect such a reduction. In general, however, it will not be the same as the variance stabilizer: for Example 1.1 above it would be $g(\theta) = \theta^3$.

For $p > 1$ one cannot hope to find a one to one $g(\cdot)$ such that

$$n^{1/2}\left(g(\widehat{\theta}_n) - g(\theta)\right) \dot{\sim} \mathcal{N}_p\left(0, I_p\right)$$

even when $V(\theta, \psi) = V(\theta)$ – except for special types of $V(\theta)$ diagonal – since we cannot impose $p+p(p-1)/2$ constraints with just p functions $g_1, \ldots, g_p$. For some exceptions see Holland [4]. However, for problems such as testing $\theta = \theta_0$, a given value, or the corresponding parameter from a second sample), or constructing a p-dimensional CR (confidence region) for θ, one may try instead to find a one to one $g(\cdot)$ such that either
(i) the diagonal of $V_g(\theta)$ is constant, say

$$V_g(\theta)_{ii} = 1$$

for $1 \leq i \leq p$.
or (ii) such that the diagonal of $V_g(\theta)^{-1}$ is constant, say

$$V_g(\theta)^{ii} = 1$$

for $1 \leq i \leq p$, where $(V^{ij}) = V^{-1}$.
Problem (ii) appears to be more relevant than Problem (i) for obtaining an approximate confidence region for θ. Problem (i) is considered in §3. In §5 we illustrate that Problem (ii) often has no solution.

In many cases one will be constrained to pick $g(\cdot)$ in the form

$$g_i(\theta) = g_{(i)}\left(\theta_i\right) \tag{1.5}$$

for $1 \leq i \leq p$, where each $g_{(i)} : R \to R$ is one to one – typically either because separate CRs are required by the client for each θ_i (despite their being dependent) or because a transformation such as in (i) or (ii) may give a peculiarly shaped CR in R^p that is difficult to visualize. So, we have
Problem (iii): how can one choose $g(\cdot)$ of type (1.5) so that $\{V_g(\theta)_{ii},\ 1 \leq i \leq p\}$ are 'as stable as possible'? Since for $g(\cdot)$ of type (1.5)

$$V_g(\theta)_{ii} = \dot{g}_{(i)}\left(\theta_i\right)^2 V(\theta)_{ii},\ 1 \leq i \leq p$$

one cannot hope to correct for variation from $\{\theta_j,\ j \neq 1\}$ in $V(\theta)_{ii}$ using such a transformation. Consider for instance

Example 1.3. *Suppose F is bivariate normal with arbitrary mean and covariance $\begin{pmatrix} \theta_2 & \theta_1 \\ \theta_1 & \theta_3 \end{pmatrix}$ and sample covariance $\begin{pmatrix} \widehat{\theta}_2 & \widehat{\theta}_1 \\ \widehat{\theta}_1 & \widehat{\theta}_3 \end{pmatrix}$. By (1.2) this has*

$$V(\theta) = \begin{pmatrix} \theta_1^2 + 2\theta_2\theta_3 & 2\theta_1\theta_2 & 2\theta_1\theta_3 \\ . & 2\theta_2^2 & 2\theta_1^2 \\ . & . & 2\theta_3^2 \end{pmatrix}. \tag{1.6}$$

So, under (1.5), $V_g(\theta)_{11} = \dot{g}_{(1)}(\theta_1)^2(\theta_1^2 + 2\theta_2\theta_3)$, so that there is no obvious way to proceed. One might decide to stabilize the dominant term, viz. θ_1^2 if $\widehat{\theta}_1^2 \gg 2\widehat{\theta}_2\widehat{\theta}_3$ with $g_{(1)}(\theta_1) = \log \theta_1$, or $2\theta_2\theta_3$ if $\widehat{\theta}_1 \ll 2\widehat{\theta}_2\widehat{\theta}_3$ with $g_{(1)}(\theta_1) = \theta_1$. So, there appears to be no satisfactory answer to Problem (iii) and we shall not consider it further.

Quite a different problem is (iv) that of choosing a one to one $g(\cdot) : R^p \to R^p$ such that $g_1(\widehat{\theta}_n), \ldots, g_p(\widehat{\theta}_n)$ are asymptotically independent:

$$V_g(\theta)_{ij} = 0$$

for $i \neq j$. This problem is considered in §4. Clearly, there will generally be no solution if $p > 3$, and in fact even for $p = 2$ there is often no solution. In this case, one compromise is to look for $g(\cdot)$ such that the off-diagonal elements are "small". The use of this sort of reparameterization has been argued for by Gillis and Ratkowsky [3]. In fact, Ratkowsky [6] advocates that the client be urged to switch his interest in $\theta_1, \ldots, \theta_p$ to concern about $\phi_1, \ldots, \phi_p$, where $\phi = g(\theta)$. Such orthogonal reparameterization has also proved useful in simplifying computational procedures. For example, the classical use of orthogonal polynomials in calculating regression parameters amounts to an orthogonal reparameterization.

When Problem (iv) has no solution – or as a first step to finding a solution of problem (iv) – we have

Problem (v): find a one to one $g(\cdot) : R^p \to R^p$ so that $g(\widehat{\theta}_n)_1$ is independent (asymptotically) of the other components; i.e.

$$V_g(\theta)_{ij} = 0$$

for $2 \leq j \leq p$. Problem (v) is also considered in §4.

So far we have assumed $V(\theta)$ is a known function. However, finding it for a particular estimate $\widehat{\theta}$ may be sometimes extremely difficult by classical means. In §2 we give a brief account of how $V(\theta)$ may be found very simply.

2. Finding the asymptotic covariance

Consider a random sample $Z_1, \ldots, Z_n$ of observations from R^s with unknown distribution F and empirical distribution F_n. Let $\theta(\cdot)$ be a p-dimensional functional on the set of distributions on R^s. Suppose we decide to estimate $\theta(F)$ by $\widehat{\theta}_n = \theta_n(F_n)$, where $\theta_n(\cdot)$ is a functional such that

$$n^{1/2}\left(\theta_n(F) - \theta(F)\right) \to 0$$

as $n \to \infty$. For example, the unbiased estimate of $\Sigma(F) = \mathrm{covar}_F(Z_1)$ is $\Sigma_n(F_n)$, where

$$\Sigma_n(F) = \left(1 - n^{-1}\right)^{-1} \Sigma(F).$$

Under regularity conditions, (I), say,

$$A_\theta(F)^{-1/2} n^{1/2} \left(\theta_n(F_n) - \theta(F)\right) \to N\left(0, I_p\right) \tag{2.1}$$

as $n \to \infty$, where

$$A_\theta(F) = \int I_{\theta,F}(z) I_{\theta,F}(z)' dF(z)$$

and

$$I_{\theta,F}(z) = \lim_{\varepsilon \downarrow 0} \left(\theta\left(\overline{1-\varepsilon} F + \varepsilon\delta_z\right) - \theta(F)\right)/\varepsilon,$$

– the *influence function* of $\theta(\cdot)$. Here, δ_z is the distribution putting mass 1 at z. (Under slightly weaker conditions we may replace A_θ by A_{θ_n} in (2.1)). For example, when $\theta_n(F) = \theta(F) = \int z dF(z)$, we may take (I): $A_\theta(F)$ finite. Non-parametric confidence regions based on (2.1) can then be computed.

Now suppose that we may assume (say by the physical nature of the problem) that F has a known parametric form F_{θ_0}, ψ, where (θ_0, ψ) are unknown, θ_0 lies in R^p, and ψ in R^q. Then $\theta_n(F_n)$ is a consistent estimate of θ_0 provided $\theta(F_{\theta_0,\psi}) \equiv \theta_0$, and $V(\theta_0) = A_\theta(F_{\theta_0,\psi})$.

For $p = 1$ we saw that $V_g(\theta) = 1$ for the transformation (1.4). If $V_n(\theta)$ approximates the covariance of $\theta_n(F_n)$ more closely than $V(\theta)$, one might prefer to use instead

$$g_n(\theta) = \int_0^\theta V_n(x)^{-1/2} dx. \tag{2.2}$$

So, in Example 1.2, $N^{1/2}(\widehat{\theta}_n - \theta)$ has variance $(1-\theta^2)^2(1 + 11\theta^2/2N + O(N^{-2}))$, where $N = n - 1$, by equation (16.74) of Stuart and Ord [7], for which (2.2) gives

$$g_n(\theta) = g(\theta) - (4N)^{-1}\left(3g(\theta) + \theta\right),$$

where $g(\theta) = 1/2 \log\{(1+\theta)/(1-\theta)\}$.

3. Stabilizing the diagonal of the covariance

In this section, we illustrate Problems (i) and (ii) for the case, where we wish to estimate $\theta(F) = \{\mu(F), \Sigma(F)\}$ or a component of it such as $\mu(F)$, where $\mu(F) = E_F Z_1 = \int z dF(z)$, and $\Sigma(F) = \mathrm{covar}_F(Z_1)$ with $\widehat{\theta}_n = \theta_n(F_n) = \{\mu(F_n), \Sigma_n(F_n)\}$ and

$$n^{1/2}\left(\Sigma_n(F) - \Sigma(F)\right) \to 0$$

as $n \to \infty$. By equation (1.4) of Withers and Nadarajah [8], $\theta(\cdot)$ has influence function

$$I_{\theta,F}(z) = \left\{ I_{\mu,F}(z), I_{\Sigma,F}(z) \right\},$$

where

$$I_{\mu,F}(z) = z - \mu(F), \;\; I_{\Sigma,F}(z) = (z - \mu(F))\,(z - \mu(F))' - \Sigma(F).$$

So, $n^{1/2}(\widehat{\theta}_n - \theta)$ has asymptotic covariance $A_\theta(F) = \{A_{\theta_i\theta_j}(F)\}$ given by

$$A_{\mu_i,\mu_j}(F) = \Sigma_{ij}(F), \tag{3.1}$$

$$A_{\mu_k,\Sigma_{ij}}(F) = \mu_{ijk}(F), \tag{3.2}$$

where

$$\mu_{ijk}(F) = \int (z_i - \mu_i(F))\,(z_j - \mu_j(F))\,(z_k - \mu_k(F))\,dF(z)$$

and

$$A_{\Sigma_{ij},\Sigma_{kl}}(F) = \mu_{ijkl}(F) - \Sigma_{ij}(F)\Sigma_{kl}(F), \tag{3.3}$$

where $\mu_{ijkl}(F)$ is defined analogously. (Here, we have extended our notation in an obvious manner since θ is not a vector.)

So, for any parametric distribution, F_θ, for which the third central moments vanish, $\mu_{ijk}(F_\theta) \equiv 0$, $\widehat{\mu}_n$ and $\widehat{\Sigma}_n$ are asymptotically independent.

Example 3.1. *Consider $F(z) = N_s(\mu, \Sigma)$. So, $\theta(F) = \theta$, where $\theta = \{\mu, \Sigma\}$. (So, the M.L.E. of θ is $\theta(F_n)$ while an unbiased estimate of θ is $\theta_n(F_n)$, where $\theta_n(F) = \{\mu(F), \frac{n}{n-1}\Sigma(F)\}$.) By (3.1)–(3.3), $V(\theta)$ is given by*

$$V_{\mu,\mu}(\theta) = \Sigma, \;\; V_{\mu,\Sigma}(\theta) = 0,$$
$$V_{\Sigma_{ij},\Sigma_{kl}}(\theta) = \Sigma_{ik}\Sigma_{jl} + \Sigma_{il}\Sigma_{jk},$$

which does not depend on μ. $\square$

In general, when $\theta = \binom{\theta_1}{\theta 2}$ and $V(\theta)$ does not depend on θ_1 then for $V_g(\theta)$ to satisfy Problem (i) or (ii), (1.3) requires that $g(\theta)$ is independent of θ_1, so that $g(\cdot)$ is not one to one and there is no solution. So, in Example 3.1 Problem (i), like Problem (iv), has no solution when θ includes any component of μ.

Does Problem (i) have a solution for $\theta = \Sigma$, $\psi = \mu$? First consider two subsets of Σ. As noted for the case $p = 1$ a solution to Problems (i) and (iv) for $\theta = (\Sigma_{11}, \ldots, \Sigma_{pp})'$ is (1.5) with $g_i(\cdot) \equiv p(\cdot)$, where

$$p(x) = 2^{-1/2} \log x. \tag{3.4}$$

In this case,

$$V_g(\theta) = \left(2\rho_{ij}^2\right),$$

where

$$\rho_{ij} = \mathrm{corr}\left((Z_1)_i , (Z_1)_j \right) = \Sigma_{ij} \left(\Sigma_{ii} \Sigma_{jj} \right)^{-1/2}.$$

Now consider Problem (i) as in Example 1.3 for

$$\theta = \left(\Sigma_{12}, \Sigma_{11}, \Sigma_{22} \right)'. \tag{3.5}$$

Note that $V(\theta)$ is given by (1.6). To find a solution one may proceed sequentially. One can first stabilize those $V(\theta)_{ii}$ which are functions of θ_i alone, using (1.4). By (1.3) this has the effect of dividing the corresponding elements of the ith row and column of $V(\theta)$ by $V(\theta)_{ii}^{1/2}$, where the (i, i) element is counted twice. That is, transforming to

$$\begin{aligned} x(\theta) &= \left(\theta_1, p\left(\theta_2\right), p\left(\theta_3\right) \right)', \\ V_x(\theta) &= \begin{pmatrix} \theta_1^2 + \theta_2\theta_3 & 2^{1/2}\theta_1 & 2^{1/2}\theta_1 \\ . & 1 & \theta_1^2\theta_2^{-1}\theta_3^{-1} \\ . & . & 1 \end{pmatrix}. \end{aligned}$$

As a second step we might look for a transformation of type $u(\theta) = y(x) = y(x(\theta)) = (a(\theta_1)b(\theta_2)c(\theta_3), x_2, x_3)'$ such that $V_u(\theta)_{11} = 1$. If these steps are done in the reverse order the result will be the same: first we try for $g(\theta) = (u(\theta), \theta_2, \theta_3)'$ such that $V_g(\theta)_{11}$ depends on $u = u(\theta)$ alone, where $u(\theta)$ has form $\theta_1 bc$, and $b = b(\theta_2)$, $c = c(\theta_3)$. Since $g(\theta)_i = \theta_i$, $i = 2, 3$, $V_g(\theta)_{ij} = V(\theta)_{ij}$, $2 \le i$, $j \le 3$. In particular, $V(\theta)_{22}$, $V(\theta)_{33}$ are not altered. (Recall these are in the required form for the use of (1.4).)

By (1.3), we obtain

$$\begin{aligned} V_g(\theta)_{11} \quad = \quad & \theta_1^2 b^2 c^2 + \theta_2 b^2 \theta_3 c^2 + \theta_1^2 \dot{b}^2 2\theta_2^2 c^2 + 2\theta_1^2 b^2 \dot{c}^2 \theta_2^2 + 4\theta_1^2 \theta_2 b\dot{b} c^2 \\ & + 4\theta_1^2 b^2 c\dot{c} + 4\theta_1^4 \dot{b} b \dot{c} c, \end{aligned}$$

which is a quartic in θ_1 and hence in u. In particular, we require that $4\dot{b}b^{-3}\dot{c}c^{-3}$, the coefficient of u^4, be constant. This is achieved by $b = \theta_2^{-1/2}$, $c = \theta_3^{-1/2}$, (which gives $u(\theta) = \rho_{12}$.)

Checking the other terms, we find that this works; $V_g(\theta)_{11}$ is a function of ρ_{12} alone. We obtain from (1.3), for

$$\theta = \begin{pmatrix} \rho_{12} \\ \Sigma_{11} \\ \Sigma_{22} \end{pmatrix}, \quad V(\theta) = \begin{pmatrix} \left(1 - \theta_1^2\right)^2 & \left(\theta_1 - \theta_1^3\right)\theta_2 & \left(\theta_1 - \theta_1^3\right)\theta_3 \\ . & 2\theta_2^2 & 2\theta_1^2\theta_2\theta_3 \\ . & . & 2\theta_3^2 \end{pmatrix}.$$

So, applying (1.4) to diagonal terms we obtain as a solution to Problem (i): for

$$\theta = \begin{pmatrix} z\left(\rho_{12}\right) \\ p\left(\Sigma_{11}\right) \\ p\left(\Sigma_{22}\right) \end{pmatrix}, \quad V(\theta) = \begin{pmatrix} 1 & 2^{-1/2}\rho_{12} & 2^{-1/2}\rho_{12} \\ . & 1 & \rho_{12} \\ . & . & 1 \end{pmatrix}. \tag{3.6}$$

Here, $z(\cdot)$ is Fisher's z-transformation

$$z(\rho) = 2^{-1} \log \frac{1+\rho}{1-\rho},$$

and $p(\cdot)$ is given by (3.4). So, an answer to Problem (i) for $\theta = \Sigma$ is

$$g(\theta) = \{z(\rho_{ij}), p(\Sigma_{ii})\} = \{q_{ij}, \lambda_i\}$$

say, for which by (1.3), $V_g(\theta : \theta_i, \theta_j) = V_g(\theta)_{ij}$ is given by

$$\begin{aligned}
V_g(\theta : \lambda_i, \lambda_j) &= \rho_{ij}^2, \\
V_g(\theta : q_{ij}, q_{kl}) &= \left(1-\rho_{ij}^2\right)^{-1} \left(1-\rho_{kl}^2\right)^{-1} \Big\{ \rho_{ki}\rho_{lj} + \rho_{kj}\rho_{li} - \rho_{kl}\rho_{ki}\rho_{kj} \\
&\quad - \rho_{kl}\rho_{li}\rho_{lj} - \rho_{ij}\rho_{ki}\rho_{li} \\
&\quad - \rho_{ij}\rho_{kj}\rho_{lj} + 1/2\rho_{ij}\rho_{kl}(\rho_{ki}^2 + \rho_{li}^2 + \rho_{kj}^2 + \rho_{lj}^2) \Big\}, \\
V_g(\theta : q_{ij}, \lambda_k) &= \left(1-\rho_{ij}^2\right)^{-1} 2^{-1/2} \Big\{ 2\rho_{ik}\rho_{jk} - \rho_{ij}\left(\rho_{ik}^2 + \rho_{jk}^2\right) \Big\}.
\end{aligned}$$

4. Obtaining asymptotic independence

Here, we consider Problems (iv) and (v). Let us again proceed stepwise, at the first step transforming to $g(\theta)' = (u(\theta), \theta_2, \theta_3, \ldots)$ and seeking to make $u(\widehat{\theta})$ asymptotically independent of $(\widehat{\theta}_2, \widehat{\theta}_3, \ldots)$; i.e. we seek $u(\cdot)$ so that $V_g(\theta)_{12} = \cdots = V_g(\theta)_{1p} = 0$. This will solve Problem (v) and reduce the dimension of Problem (iv) from p to $p-1$. We have

$$V_g(\theta) = \begin{pmatrix} \sum u_i u_j V_{ij} & \sum u_i V_{i2} & \cdots & \sum u_i V_{ip} \\ . & V_{22} & \cdots & V_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ . & V_{p2} & \cdots & V_{pp} \end{pmatrix},$$

where $V_{ij} = V_{ij}(\theta)$ and $u_i = \partial u(\theta)/\partial \theta_i$. The method of obtaining the general solution of

$$\sum_{i=1}^{p} u_i V_{ij} = 0 \tag{4.1}$$

for a fixed j, is known as the method of characteristic curves, and is as follows, (c.f. pages 28-30 of Courant and Hilbert [2]).

Solve $d\theta_i/ds = V_{ij}(\theta)$, $1 \le i \le p$ for functions $\theta_1(s), \ldots, \theta_p(s)$, with p arbitrary coefficients. Eliminate s to get $p-1$ equations of the form $c_i = \phi_i(\theta)$, $1 \le i \le p-1$, where c_i is an arbitrary constant, and ϕ_i a completely specified function. Then the general solution of (4.1) is $u_j(\theta) = w(\phi_1(\theta), \ldots, \phi_{p-1}(\theta))$, where $w(\cdot)$ is a suitably differentiable but otherwise arbitrary function. So, in order to solve Problem (v) we have first to find the solutions $u_j(\theta)$ to (4.1) for

$j = 2, \ldots, p$ in terms of $p-1$ arbitrary functions, and then to see if we can choose our arbitrary functions so that $u_2(\theta) = \cdots = u_p(\theta)$.

Generally one can simplify the problem taking suitable linear combinations of (4.1) and solving the partial differential equations so obtained.

Consider again

Example 4.1. *Consider $F(z) = N_2(\mu, \Sigma)$ with θ given by (3.5). Solving the problem for θ is the same as solving it for $g(\theta)$, if g is one-to-one, so we may as well start from the form θ given by (3.6), since it has a fairly simple form for $V(\theta)$. Set $\rho = \rho_{12}$. Then (4.1) with $j = 2, 3$ is just*

$$u_1 2^{-1/2}\rho + u_2 + u_3\rho^2 = 0 = u_1 2^{-1/2}\rho + u_2\rho^2 + u_3,$$

which simplifies, assuming $|\rho| \neq 1$, to

$$u_2 = u_3$$

and

$$u_1 = -\psi(\theta_1)\, u_2,$$

where $\psi(\theta_1) = 2^{1/2}(\rho + \rho^{-1})$, $\rho = \tanh(\theta_1)$.

By the method of characteristic curves the solution is $u(\theta) = f(\theta_2 - \theta_3 - b(\theta_1))$, where $f(\cdot)$ is differentiable but otherwise arbitrary, and $\dot{b}(\theta_1) = \psi(\theta_1)$, i.e. $b(\theta_1) = 2^{1/2} \log \sinh |2\theta_1|$. $\square$

In particular, taking $f(x) = 2^{1/2}x$ gives $u(\theta) = 2^{1/2}(\theta_2 + \theta_3) - \log\sinh|2\theta_1|$. That is $u(\widehat{\theta})$ is asymptotically independent of $(\widehat{\theta}_2, \widehat{\theta}_3)$. Transforming back to Σ we obtain:

Theorem 4.1. *Set $\rho = \rho_{12}$. Let*

$$M_{12}(\Sigma) = \Sigma_{11}\Sigma_{22}\Sigma_{12}^{-1} - \Sigma_{12} = (\Sigma_{11}\Sigma_{22})^{1/2}\left(\rho^{-1} - \rho\right). \tag{4.2}$$

Then $M_{12}(\widehat{\Sigma})$ is asymptotically independent of $(\widehat{\Sigma}_{11}, \widehat{\Sigma}_{22})$.

An application of (1.3) to (1.1) yields $n\mathrm{var} M_{12}(\widehat{\Sigma}) \to A_{M_{12}}(\Sigma) = d(\rho^2)\Sigma_{11}\Sigma_{22}$ and $n\mathrm{var}(\widehat{\rho}^{-1} - \widehat{\rho}) \to d(\rho^2)$ as $n \to \infty$, where $d(x) = x - 4 - x^{-1} + x^{-2}$. So, by the conditional properties of the normal distribution (for example, page 28 of Anderson [1]), we have conditional convergence:

Corollary 4.1. *As $n \to \infty$,*

$$n^{1/2}\left(M_{12}\left(\widehat{\Sigma}\right) - M_{12}(\Sigma)\right)\Big|\left(\widehat{\Sigma}_{11}, \widehat{\Sigma}_{22}\right) \to \mathcal{N}\left(0, A_{M_{12}}(\Sigma)\right),$$

$$n^{1/2}\left(\widehat{\rho}^{-1} - \widehat{\rho} - \rho^{-1} + \rho\right)\Big|\left(\widehat{\Sigma}_{11}, \widehat{\Sigma}_{22}\right) \to \mathcal{N}\left(0, d\left(\rho^2\right)\right).$$

This solves Problem (v) for θ given by (3.5) and reduces Problem (iv) to finding a one to one $g(\cdot) : R^2 \to R^2$ so that $g_1(\widehat{\Sigma}_{11}, \widehat{\Sigma}_{22})$ is asymptotically independent of $g_2(\widehat{\Sigma}_{11}, \widehat{\Sigma}_{22})$. For simplicity transform $\theta = (\Sigma_{11}, \Sigma_{22})'$ to

$$\theta = \left(p\left(\Sigma_{11}\right), p\left(\Sigma_{22}\right)\right)' \text{ for which } V(\theta) = \begin{pmatrix} 1 & \rho_{12}^2 \\ . & 1 \end{pmatrix}, \tag{4.3}$$

where we may regard ρ_{12} as a nuisance parameter. In this form we see Problem (iv) has no solution for θ in (4.3), and hence Problem (iv) has no solution for $\theta = (\Sigma_{12}, \Sigma_{11}, \Sigma_{22})'$ – since transforming (4.3) suitably will spoil the independence obtained in (4.2).

5. Stabilizing the diagonal of the inverse covariance

Here, we give an example, where Problem (ii) has no solution. Because of the labor in computing inverses we take the case $p = 3$. Again suppose F is bivariate normal and $\theta' = (\Sigma_{12}, \Sigma_{11}, \Sigma_{22})$. To solve Problem (ii) for θ is the same as solving it for $g(\theta)$, where g is one-to-one. So, let us take θ as in (3.6), and set $\rho = \rho_{12}$. Then

$$V(\theta)^{-1} = \begin{pmatrix} 1+\rho^2 & -\rho^2/2 & -\rho^2/2 \\ . & b(\rho) & c(\rho) \\ . & . & b(\rho) \end{pmatrix},$$

where $b(\rho) = 1/2 + 1/2(1-\rho^2)^{-1}$ and $c(\rho) = 1/2 - 1/2(1-\rho^2)^{-1}$.

Since $V(\theta)$ depends only on ρ, let us try a transformation of the form $g(\theta)' = (a_1(\rho),\ a_2(\rho)\theta_2,\ a_3(\rho)\theta_3)$. Computing the coefficient of $V_g(\theta)^{11} = (V_g(\theta)^{-1})_{11}$ one finds that $|V_g(\theta)|.V_g(\theta)^{11}$ is linear in $(1, \theta_2, \theta_3)$ while $|V_g(\theta)|$ is quadratic in $(1, \theta_2, \theta_3)$. So, for $V_g(\theta)^{11}$ to be independent of θ_2, we require $0 =$ coefficient of θ_2^2 in $|V_g(\theta)| = A_1^2 A_2^2 A_3^2 (1/2\rho^2 - 1)$, where $A_i(\theta_1) = a_i(\rho)$. So, A_2, A_3 are constant, say 1. If we choose $A_1 = (1+\rho^2)^{1/2}$ we obtain $V_g(\theta)^{11} = 1$, but $V_g(\theta)^{ii} = b(\rho)$, $i = 2, 3$. So, it seems that in this example we can obtain either $V_g(\theta)^{11} = 1$ or $V_g(\theta)^{22} = V_g(\theta)^{33} = 1$, but not both. So, it appears that Problem (ii) often has no solution.

References

[1] Anderson, T. W. (1958). *An Introduction to Multivariate Analysis*. John Wiley and Sons, New York. MR0091588

[2] Courant, R. and Hilbert, D. (1962). *Methods of Mathematical Physics*, volume 2. John Wiley and Sons, New York. MR0140802

[3] Gillis P. R. and Ratkowsky, D. A. (1978). The behaviour of the parameters of various yield-density relationships. *Biometrics* **34** 191–198.

[4] Holland, P. W. (1973). Covariance stabilising transformations. *Annals of Statistics* **1** 84–92. MR0331589

[5] Hotelling, H. (1953). New light on the correlation coefficient and its transforms. *Journal of the Royal Statistical Society* B **15** 193–232. MR0060794

[6] Ratkowsky, D. A. (1978). The behaviour of estimators of the parameter of various sigmoidal growth relationships. Presented at the fourth Australian Statistical Conference, Canberra.

[7] Stuart, A. and Ord, J. K. (1987). *Kendall's Advanced Theory of Statistics*, volume 1, fifth edition. Griffin, London.

[8] Withers, C. S. and Nadarajah, S. (2010). Reduction of bias and skewness with applications to second order accuracy. *Technical Report*, Applied Mathematics Group, Industrial Research Ltd., Lower Hutt, New Zealand.