High-dimensional analysis and estimation in general multivariate linear models

Dietrich von Rosen[1,2]

[1] Department of Energy and Technology, Swedish University of Agricultural Sciences

[2] Mathematical Statistics, Linköping University, Sweden

Outline

We present a new approach of estimating the parameters describing the mean structure in the Growth Curve model when the number of variables, p, compared with the number of observations, n, is large.

What can be performed?

- Test hypothesis (one-dimensional quantity)
- Estimate functions of parameters (including subsets)

  (spectral density, Wigner's semicircle law, random matrix theory, free probability, functional data analysis)

Background: Multivariate Linear Models

MANOVA:

$$\boldsymbol{X} \sim N_{p,n}(\boldsymbol{\mu C}, \boldsymbol{\Sigma}, \boldsymbol{I}) \qquad \text{(independent columns)}$$

$$\boldsymbol{X} : p \times n, \quad \boldsymbol{\mu} : p \times q, \quad \boldsymbol{C} : q \times n, \quad \boldsymbol{\Sigma} : p \times p$$

Growth Curve model:

$$\boldsymbol{X} \sim N_{p,n}(\underline{\underline{\boldsymbol{A}}}\boldsymbol{BC}, \boldsymbol{\Sigma}, \boldsymbol{I})$$

$$\boldsymbol{X} : p \times n, \quad \boldsymbol{A} : p \times q, \quad \boldsymbol{B} : q \times k \quad \boldsymbol{C} : q \times n, \quad \boldsymbol{\Sigma} : p \times p$$

Fixed size of mean parameter space.

Background: Growth Curve model

Sufficient statistics for the Growth Curve model are

$$\boldsymbol{S} = \boldsymbol{X}(\boldsymbol{I} - \boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-}\boldsymbol{C})\boldsymbol{X}', \qquad \boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-}\boldsymbol{C}.$$

Due to the normality assumption, i.e. since the distribution is symmetric around the mean, in order to estimate the mean parameters it is natural to consider

$$\begin{aligned}
&\tfrac{1}{p}tr\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{X} - \boldsymbol{A}\boldsymbol{B}\boldsymbol{C})(\boldsymbol{X} - \boldsymbol{A}\boldsymbol{B}\boldsymbol{C})'\}\\
&= \tfrac{1}{p}tr\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-}\boldsymbol{C} - \boldsymbol{A}\boldsymbol{B}\boldsymbol{C})(\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-}\boldsymbol{C} - \boldsymbol{A}\boldsymbol{B}\boldsymbol{C})'\}\\
&\qquad + \tfrac{1}{p}tr\{\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\}.
\end{aligned}$$

The factor $1/p$ is used to handle the increase in size of $tr(\bullet)$ when $p \to \infty$, i.e. the trace functions have been averaged.

Background: Growth Curve model

$$L(\boldsymbol{B},\boldsymbol{\Sigma}) \approx |\boldsymbol{\Sigma}|^{-n/2}\exp(\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{ABC})(\boldsymbol{X}-\boldsymbol{ABC})')$$

$$\begin{aligned}\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{ABC})\boldsymbol{C}' &= \boldsymbol{0}\\ n\boldsymbol{\Sigma} &= (\boldsymbol{X}-\boldsymbol{ABC})(\boldsymbol{X}-\boldsymbol{ABC})'\end{aligned}$$

MANOVA

$$\begin{aligned}\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{BC})\boldsymbol{C}' &= \boldsymbol{0}\\ n\boldsymbol{\Sigma} &= (\boldsymbol{X}-\boldsymbol{BC})(\boldsymbol{X}-\boldsymbol{BC})'\end{aligned}$$

Background: $\boldsymbol{X} = \boldsymbol{ABC} + \boldsymbol{E}$

Estimators in the Growth Curve model (MANOVA)

- Known $\boldsymbol{\Sigma}$, p.d.:

$$\boldsymbol{A}\widehat{\boldsymbol{B}}\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-}\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-}\boldsymbol{C}$$

$$(\widehat{\boldsymbol{B}}\boldsymbol{C} = \boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-}\boldsymbol{C})$$

- Unknown $\boldsymbol{\Sigma}$, p.d.:

$$\boldsymbol{A}\widehat{\boldsymbol{B}}\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A})^{-}\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-}\boldsymbol{C},$$

$$(\widehat{\boldsymbol{B}}\boldsymbol{C} = \boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-}\boldsymbol{C})$$

where

$$\boldsymbol{S} = \boldsymbol{X}(\boldsymbol{I} - \boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-}\boldsymbol{C})\boldsymbol{X}'.$$

Background: $\boldsymbol{X} = \boldsymbol{ABC} + \boldsymbol{E}$

$$n\boldsymbol{\Sigma} = \boldsymbol{S} + (\boldsymbol{I} - \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A})^{-}\boldsymbol{A}'\boldsymbol{S}^{-1})\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1}\boldsymbol{C}$$
$$\times \boldsymbol{X}'(\boldsymbol{I} - \boldsymbol{S}^{-1}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A})^{-}\boldsymbol{A}')$$

MANOVA

$$n\boldsymbol{\Sigma} = \boldsymbol{S}$$

Extended Growth Curve model

$$\boldsymbol{X} = \sum_{i}^{m} \boldsymbol{A}_i\boldsymbol{B}_i\boldsymbol{C}_i + \boldsymbol{E}, \qquad \mathcal{C}(\boldsymbol{C}'_m) \subseteq \mathcal{C}(\boldsymbol{C}'_{m-1}) \subseteq \cdots \subseteq \mathcal{C}(\boldsymbol{C}'_1)$$

Asymptotics $p/n \to c$

$$\begin{aligned}\boldsymbol{T}_1 &= \tfrac{1}{p} tr\{\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\}\\ \boldsymbol{T}_2 &= \tfrac{1}{p} tr\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-}\boldsymbol{C}-\boldsymbol{A}\boldsymbol{B}\boldsymbol{C})(\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-}\boldsymbol{C}-\boldsymbol{A}\boldsymbol{B}\boldsymbol{C})'\},\end{aligned}$$

In high-dimensional analysis, one often considers $\frac{1}{p}tr(\boldsymbol{S})$ or $\frac{1}{p}tr(\boldsymbol{S}^2)$ (e.g. see Ledoit & Wolf, 2002 or Srivastava, 2005) but in this case the asymptotics depends on $\boldsymbol{\Sigma}$.

Asymptotics $p/n \longrightarrow c$

$\boldsymbol{T}_1$ is chi-square distributed with n' degrees of freedom. Hence, the characteristic function $\varphi_{\boldsymbol{T}_1}(t)$ equals

$$\varphi_{\boldsymbol{T}_1}(t) = (1 - i\,t\tfrac{2}{p})^{-pn'/2},$$

where i is the imaginary unit. If taking the logarithm of the characteristic function and expanding it as a power series in p and n, it follows that

$$\begin{aligned}
\ln \varphi_{\boldsymbol{T}_1}(t) &= -pn'/2 \ln(1 - i\,t\tfrac{2}{p}) = \frac{pn'}{2} \sum_{j=1}^{\infty} \left(\tfrac{2}{p}\right)^j \tfrac{1}{j} i^k\, t^j \\
&= i\,tn' - \frac{n'p}{2}\frac{2^2}{p^2}\frac{1}{2}t^2 + \frac{n'p}{2}\frac{2^3}{p^3} i^3 \frac{1}{3} t^3 + \cdots \\
&\approx i\,tn' - \frac{n'p}{2}\frac{2^2}{p^2}\frac{1}{2}t^2.
\end{aligned}$$

Asymptotics $p/n \to c$

This implies that under $\frac{p}{n}$-asymptotics

$$\frac{\frac{1}{p}tr\{\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\} - n'}{\sqrt{\frac{n'}{p}}} \overset{a}{\sim} N(\mathbf{0}, 2),$$

where $\overset{a}{\sim}$ means ”asymptotically distributed as”.

Asymptotics $p/n \to c$

Represent $\boldsymbol{T}_2$ as $\boldsymbol{T}_2 = \frac{1}{p} tr\{\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\boldsymbol{V}'\}$, where

$$\boldsymbol{V} = \boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-}\boldsymbol{C} - \boldsymbol{A}\boldsymbol{B}\boldsymbol{C}$$

with $\boldsymbol{V}\boldsymbol{V}' \sim W_p(\boldsymbol{\Sigma}, r)$, $r = r(\boldsymbol{C})$. In this case the number of degrees of freedom of the distribution is fixed.
The logarithm of the characteristic function of $\sqrt{p}\,\boldsymbol{T}_2$ equals

$$\ln \varphi_{\sqrt{p}\boldsymbol{T}_2}(t) = -\frac{rp}{2}\ln(1 - i\,t\frac{2}{\sqrt{p}}).$$

Asymptotics $p/n \to c$

Thus,

$$\begin{aligned}\ln \varphi_{\sqrt{p}\boldsymbol{T}_2}(t) &= -\tfrac{rp}{2}\ln(1 - i\,t\tfrac{2}{\sqrt{p}}) = \tfrac{rp}{2}\sum_{j=1}^{\infty} p^{-\frac{j}{2}} 2^j \tfrac{1}{j}\, i^j\, t^j \\ &= i\,tr\sqrt{p} - rt^2 + i^3\, t^3 rp^{-\frac{1}{2}}\tfrac{1}{3} + \cdots\end{aligned}$$

and

$$\frac{\frac{1}{\sqrt{p}} tr\{\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\boldsymbol{V}'\} - r\sqrt{p}}{\sqrt{r}} \overset{\text{a}}{\sim} N(\mathbf{0}, 2).$$

Asymptotics $p/n \to c$

The following results which will serve as a starting point have been verified:

Under $\frac{p}{n}$-asymptotics T_1 converges to $N(0,2)$,

and for any n and $p \to \infty$, $\sqrt{p}T_2$ also converges to $N(0,2)$.

Since $\boldsymbol{S}$ and $\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-}\boldsymbol{C}$ are sufficient statistics, we may note that $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ include the relevant information for estimating the mean parameters of the Growth Curve model. Thus, based on $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ an asymptotic likelihood approach may be presented.

Estimation

From the previous section, it follows that an asymptotic likelihood based on $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ is proportional to

$$exp\{-\frac{1}{4}(pn'(\frac{1}{pn'}tr\{\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\}-1)^2)\}exp\{-\frac{1}{4}(pr(\frac{1}{pr}tr\{\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\boldsymbol{V}'\}-1)^2)\}.$$

Following the likelihood principle this function needs to be maximized. Since $\boldsymbol{\Sigma}$ is assumed to be of full rank and unstructured, and $\boldsymbol{S}$ may be singular if $\frac{p}{n} \to c > 1$ it is impossible to get appropriate estimators for all elements of $\boldsymbol{\Sigma}$ and $\boldsymbol{B}$. However, we are only interested in the estimation of $\boldsymbol{B}$ and its variance. Therefore, we will investigate the two terms separately, and suggest an approach similar to the restricted maximum likelihood method.

Estimation

Let us start with the first term, i.e.

$$\left(\frac{1}{pn'} tr\{\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\} - 1\right)^2.$$

By choosing

$$\widehat{\boldsymbol{\Sigma}}^{-1} = \max(p, n')\boldsymbol{S}^{-}$$

the above expression equals 0, where $\boldsymbol{S}^{-}$ denotes an arbitrary g-inverse of $\boldsymbol{S}$.

Estimation

The main drawback with this estimator is that it is not unique. However, since we are dealing with estimation it is natural to suppose that $\mathcal{C}(\boldsymbol{S}^-) = \mathcal{C}(\boldsymbol{S})$ which implies that $r(\boldsymbol{S}^-) = r(\boldsymbol{S})$. The latter condition implies that $\boldsymbol{S}^-$ is a reflexive g-inverse, i.e. $\boldsymbol{S}^-\boldsymbol{S}\boldsymbol{S}^- = \boldsymbol{S}^-$ holds besides the defining condition $\boldsymbol{S}\boldsymbol{S}^-\boldsymbol{S} = \boldsymbol{S}$. If $\boldsymbol{S}^-$ is not a reflexive g-inverse, then $r(\boldsymbol{S}) < r(\boldsymbol{S}^-)$ and therefore we can estimate more elements in $\boldsymbol{\Sigma}^{-1}$ than in $\boldsymbol{\Sigma}$ which does not make sense. Furthermore, if $\mathcal{C}(\boldsymbol{S}^-) = \mathcal{C}(\boldsymbol{S})$ then, $r(\boldsymbol{S}^-\boldsymbol{S} - \boldsymbol{S}\boldsymbol{S}^-) = r(\boldsymbol{S}(\boldsymbol{S}^-\boldsymbol{S} - \boldsymbol{S}\boldsymbol{S}^-)\boldsymbol{S}) = 0$. Thus, $\mathcal{C}(\boldsymbol{S}^-) = \mathcal{C}(\boldsymbol{S})$ implies that $\boldsymbol{S}^-$ is the unique Moore-Penrose g-inverse which will be denoted $\boldsymbol{S}^+$.

Estimation

In the next we replace $\boldsymbol{\Sigma}^{-1}$ by $\max(p,n')\boldsymbol{S}^{+}$ in the second exponent and thus have to minimize

$$\left(\frac{\max(p,n')}{pr}tr\{\boldsymbol{S}^{+}\boldsymbol{V}\boldsymbol{V}'\}-1\right)^{2}.$$

Differentiating this expression with respect to $\boldsymbol{B}$ we get the equation

$$\left(\frac{\max(p,n')}{pr}tr\boldsymbol{S}^{+}\boldsymbol{V}\boldsymbol{V}'-1\right)\boldsymbol{A}'\boldsymbol{S}^{+}(\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-}\boldsymbol{C}-\boldsymbol{A}\boldsymbol{B}\boldsymbol{C})\boldsymbol{C}'=\mathbf{0}.$$

With probability 1, the expression $(\frac{n'}{pr}tr\boldsymbol{S}^{+}\boldsymbol{V}\boldsymbol{V}'-1)$ differs from 0, and thus the following linear equation in $\boldsymbol{B}$ emerges:

$$\boldsymbol{A}'\boldsymbol{S}^{+}(\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-}\boldsymbol{C}-\boldsymbol{A}\boldsymbol{B}\boldsymbol{C})\boldsymbol{C}'=\mathbf{0}.$$

Estimation

This equation is consistent if the column space relation $\mathcal{C}(\boldsymbol{A}'\boldsymbol{S}^{+}) = \mathcal{C}(\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})$ holds, which is true since $\boldsymbol{S}^{+}$ is p.s.d. Hence,

$$\widehat{\boldsymbol{B}} = (\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{-}\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-} + (\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{o}\boldsymbol{Z}_1 + \boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A}\boldsymbol{Z}_2\boldsymbol{C}^{o\prime},$$

where $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ are arbitrary matrices, and $(\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{o}$ and $\boldsymbol{C}^{o}$ are any arbitrary matrices spanning the orthogonal complement to $\mathcal{C}(\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})$ and $\mathcal{C}(\boldsymbol{C})$, respectively.

Estimation

From here we obtain the following result:

The estimator $\widehat{\boldsymbol{B}}$, given above, is unique and with probability 1 equals

$$\widehat{\boldsymbol{B}} = (\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1},$$

if and only if $r(\boldsymbol{A}) = q < \min(p, n')$, $r(\boldsymbol{C}) = k$ and $\mathcal{C}(\boldsymbol{A}) \cap \mathcal{C}(\boldsymbol{S})^{\perp} = \{\mathbf{0}\}$.

If $\boldsymbol{S}$ is of full rank, i.e. $(p \leq n')$, $\widehat{\boldsymbol{B}}$ is identical to the maximum likelihood estimator.

Properties

Since $\boldsymbol{XC}'$ and $\boldsymbol{S}$ are independently distributed

$$\begin{aligned} E[\widehat{\boldsymbol{B}}] &= E[(\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{S}^{+}]E[\boldsymbol{XC}'(\boldsymbol{CC}')^{-1}] \\ &= E[(\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{S}^{+}]\boldsymbol{AB} = \boldsymbol{B}. \end{aligned}$$

The dispersion matrix

$$D[\widehat{\boldsymbol{B}}] = E[vec(\widehat{\boldsymbol{B}} - \boldsymbol{B})vec'(\widehat{\boldsymbol{B}} - \boldsymbol{B})],$$

where $vec(\cdot)$ is the usual vec-operator, is much more complicated to obtain.

Properties

Since $D[\boldsymbol{X}] = \boldsymbol{I} \otimes \boldsymbol{\Sigma}$,

$$D[\widehat{\boldsymbol{B}}] = (\boldsymbol{C}\boldsymbol{C}')^{-1} \otimes E[(\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{\Sigma}\boldsymbol{S}^{+}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{-1}]$$

has to be considered. If $p > n'$, it follows that if the denominator in the next expression is larger than 0, then

$$D[\widehat{\boldsymbol{B}}] = (\boldsymbol{C}\boldsymbol{C}')^{-1} \otimes (\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1} \frac{(p-q-1)(p-1)}{(n'-q-1)(p-n'+q-1)}.$$

Note that if $(\boldsymbol{C}\boldsymbol{C}')^{-1} \to \mathbf{0}$ then $D[\widehat{\boldsymbol{B}}] \to \mathbf{0}$, and if $(n' - q - 1)$ or $(p - n' + q - 1)$ are small, $D[\widehat{\boldsymbol{B}}]$ is large. It also follows that if n is much smaller than p, the dispersion $D[\widehat{\boldsymbol{B}}]$ will be large if not $(\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1}$ is small.

Properties

It follows that an unbiased estimator of $D[\widehat{\boldsymbol{B}}]$ is given by

$$\widehat{D[\widehat{\boldsymbol{B}}]} = (\boldsymbol{C}\boldsymbol{C}')^{-1} \otimes (\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{-1} \frac{(p-1)}{(p-n'+q)(p-n'+q-1)}.$$

If $p \leq n'$,

$$\begin{aligned} D[\widehat{\boldsymbol{B}}] &= (\boldsymbol{C}\boldsymbol{C}')^{-1} \otimes (\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1} \frac{n'-1}{n'-p+q-1}, \\ \widehat{D[\widehat{\boldsymbol{B}}]} &= (\boldsymbol{C}\boldsymbol{C}')^{-1} \otimes (\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A})^{-1} \frac{(n'-1)}{(n'-p+q)(n'-p+q-1)}. \end{aligned}$$

If $p = n'$ both variances are equal.

$E[(\boldsymbol{A}'\boldsymbol{S}^+\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{S}^+\boldsymbol{\Sigma}\boldsymbol{S}^+\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}^+\boldsymbol{A})^{-1}]$

$$E[(\boldsymbol{A}'\boldsymbol{S}^+\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{S}^+\boldsymbol{\Sigma}\boldsymbol{S}^+\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}^+\boldsymbol{A})^{-1}]$$

is obtained. First the expectation is presented in a canonical form. There exist always an orthogonal matrix $\boldsymbol{\Gamma}$ and a non-singular matrix $\boldsymbol{L}$ such that

$$\boldsymbol{A}' = \boldsymbol{L}(\boldsymbol{I}_q : \mathbf{0})\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{\frac{1}{2}}.$$

Moreover, let

$$\boldsymbol{U} = \boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{S}\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Gamma}' \sim W_p(\boldsymbol{I}_p, n'),$$

and thus

$E[(\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{\Sigma}\boldsymbol{S}^{+}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{-1}]$

$$
\begin{aligned}
&E[(\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{\Sigma}\boldsymbol{S}^{+}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{-1}] \\
&\quad = (\boldsymbol{L}')^{-1}E\Big[\Big((\boldsymbol{I}_q : \mathbf{0})\boldsymbol{U}^{+}\left(\begin{smallmatrix}\boldsymbol{I}_q \\ \mathbf{0}\end{smallmatrix}\right)\Big)^{-1}(\boldsymbol{I}_q : \mathbf{0})\boldsymbol{U}^{+}\boldsymbol{U}^{+}\left(\begin{smallmatrix}\boldsymbol{I}_q \\ \mathbf{0}\end{smallmatrix}\right) \\
&\qquad \times \Big((\boldsymbol{I}_q : \mathbf{0})\boldsymbol{U}^{+}\left(\begin{smallmatrix}\boldsymbol{I}_q \\ \mathbf{0}\end{smallmatrix}\right)\Big)^{-1}\Big]\boldsymbol{L}^{-1}.
\end{aligned}
$$

$E[(\boldsymbol{A}'\boldsymbol{S}^+\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{S}^+\boldsymbol{\Sigma}\boldsymbol{S}^+\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}^+\boldsymbol{A})^{-1}]$

Suppose now that $p > n'$. We need to know the following partitioned Moore-Penrose:

$$\boldsymbol{U}^+ = \begin{pmatrix} \boldsymbol{U}_{11} & \boldsymbol{U}_{12} \\ \boldsymbol{U}_{21} & \boldsymbol{U}_{22} \end{pmatrix}^+, \qquad \begin{pmatrix} n' \times n' & n' \times (p-n') \\ (p-n') \times n' & (p-n') \times (p-n') \end{pmatrix}.$$

However, because of Wishartness, $\boldsymbol{U} = \boldsymbol{Y}\boldsymbol{Y}'$, $\boldsymbol{Y} = (\boldsymbol{Y}_1' : \boldsymbol{Y}_2')'$, $\boldsymbol{Y} \sim N_{p,n'}(\mathbf{0}, \boldsymbol{I}_p, \boldsymbol{I}_{n'})$ and $\boldsymbol{Y}_1 \sim N_{n',n'}(\mathbf{0}, \boldsymbol{I}_{n'}, \boldsymbol{I}_{n'})$, where it is assumed that $r(\boldsymbol{Y}_1) = n'$. Then,

$$\boldsymbol{U}^+ = \binom{\boldsymbol{Y}_1}{\boldsymbol{Y}_2} (\boldsymbol{Y}_1'\boldsymbol{Y}_1 + \boldsymbol{Y}_2'\boldsymbol{Y}_2)^{-1}(\boldsymbol{Y}_1'\boldsymbol{Y}_1 + \boldsymbol{Y}_2'\boldsymbol{Y}_2)^{-1}(\boldsymbol{Y}_1' \ \boldsymbol{Y}_2').$$

$E[(\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{\Sigma}\boldsymbol{S}^{+}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{-1}]$

Put

$$\boldsymbol{H} = \boldsymbol{Y}_1'\boldsymbol{Y}_1 + \boldsymbol{Y}_2'\boldsymbol{Y}_2 \sim W_{n'}(\boldsymbol{I}_{n'}, p).$$

Thus,

$$\begin{aligned}(\boldsymbol{L}')^{-1}E\Bigg[&\Big((\boldsymbol{I}_q : \mathbf{0})\boldsymbol{Y}_1\boldsymbol{H}^{-1/2}\boldsymbol{H}^{-1}\boldsymbol{H}^{-1/2}\boldsymbol{Y}_1'(\boldsymbol{I}_q : \mathbf{0})'\Big)^{-1}\\ &\times(\boldsymbol{I}_q : \mathbf{0})\boldsymbol{Y}_1\boldsymbol{H}^{-1/2}\boldsymbol{H}^{-1}\boldsymbol{H}^{-1}\boldsymbol{H}^{-1/2}\boldsymbol{Y}_1'(\boldsymbol{I}_q : \mathbf{0})'\\ &\times\Big((\boldsymbol{I}_q : \mathbf{0})\boldsymbol{Y}_1\boldsymbol{H}^{-1/2}\boldsymbol{H}^{-1}\boldsymbol{H}^{-1/2}\boldsymbol{Y}_1'(\boldsymbol{I}_q : \mathbf{0})'\Big)^{-1}\Bigg]\boldsymbol{L}^{-1}.\end{aligned}$$

$E[(\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{\Sigma}\boldsymbol{S}^{+}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{-1}]$

One can show (e.g. see the proof of Theorem 2.4.10 in Kollo & von Rosen, 2005) that $\boldsymbol{Y}_1\boldsymbol{H}^{-1/2}$ is independent of $\boldsymbol{H}$. Furthermore, there exist a non-singular $\boldsymbol{L}_1$ and an orthogonal matrix $\boldsymbol{\Gamma}_1$ such that

$$(\boldsymbol{I}_q : \mathbf{0})\boldsymbol{Y}_1^{1/2}\boldsymbol{H}^{-1/2} = \boldsymbol{L}_1(\boldsymbol{I}_q : \mathbf{0})\boldsymbol{\Gamma}_1$$

and partition $\boldsymbol{H}$ ($\boldsymbol{H}^{-1}$) as

$$\boldsymbol{H} = \begin{pmatrix} \boldsymbol{H}_{11} & \boldsymbol{H}_{12} \\ \boldsymbol{H}_{21} & \boldsymbol{H}_{22} \end{pmatrix}, \qquad \boldsymbol{H}^{-1} = \begin{pmatrix} \boldsymbol{H}^{11} & \boldsymbol{H}^{12} \\ \boldsymbol{H}^{21} & \boldsymbol{H}^{22} \end{pmatrix},$$

$$\begin{pmatrix} q \times q & q \times (n'-q) \\ (n'-q) \times q & (n'-q) \times (n'-q) \end{pmatrix}.$$

$E[(\boldsymbol{A}'\boldsymbol{S}^+\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{S}^+\boldsymbol{\Sigma}\boldsymbol{S}^+\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}^+\boldsymbol{A})^{-1}]$

Therefore,

$$(\boldsymbol{L}')^{-1}E[(\boldsymbol{L}_1')^{-1}E[(\boldsymbol{H}^{11})^{-1}(\boldsymbol{H}^{11}:\boldsymbol{H}^{12})(\boldsymbol{H}^{11}:\boldsymbol{H}^{12})'(\boldsymbol{H}^{11})^{-1}]\boldsymbol{L}_1^{-1}]\boldsymbol{L}^{-1}$$
$$=(\boldsymbol{L}')^{-1}E[(\boldsymbol{L}_1')^{-1}(\boldsymbol{I}+E[\boldsymbol{H}_{12}\boldsymbol{H}_{22}^{-1}\boldsymbol{H}_{22}^{-1}\boldsymbol{H}_{21}])(\boldsymbol{L}_1)^{-1}](\boldsymbol{L})^{-1}.$$

However, since $\boldsymbol{H}$ is Wishart distributed (see e.g. Kollo & von Rosen, 2005; p. 413)

$$\boldsymbol{I}+E[\boldsymbol{H}_{12}\boldsymbol{H}_{22}^{-1}\boldsymbol{H}_{22}^{-1}\boldsymbol{H}_{21}]=\frac{p-1}{p-n'+q-1}\boldsymbol{I}.$$

$E[(\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{\Sigma}\boldsymbol{S}^{+}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{-1}]$

Furthermore, put

$$\boldsymbol{G} = \boldsymbol{Y}_{11}(\boldsymbol{Y}'_{11}\boldsymbol{Y}_{11} + \boldsymbol{W})^{-1}\boldsymbol{Y}'_{11},$$

where $\boldsymbol{Y}_1 = (\boldsymbol{Y}'_{11} : \boldsymbol{Y}'_{12})'$, and $\boldsymbol{W} = \boldsymbol{Y}'_{12}\boldsymbol{Y}_{12} + \boldsymbol{Y}'_{2}\boldsymbol{Y}_{2} \sim W_{n'}(\boldsymbol{I}, p-q)$, $\boldsymbol{Y}_{11} \sim N_{q,n'}(\mathbf{0}, \boldsymbol{I}_q, \boldsymbol{I}_{n'})$.
Then,

$$(\boldsymbol{L}_1\boldsymbol{L}'_1)^{-1} \quad = \quad \boldsymbol{G}^{-1}$$

$E[(\boldsymbol{A}'\boldsymbol{S}^+\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{S}^+\boldsymbol{\Sigma}\boldsymbol{S}^+\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}^+\boldsymbol{A})^{-1}]$

It follows (Kollo & von Rosen, 2005; Theorem 2.4.10) that the density of $\boldsymbol{G}$ equals

$$f_{\boldsymbol{G}}(\boldsymbol{G}) = c_0|\boldsymbol{G}|^{\frac{n'-q-1}{2}}|\boldsymbol{I}_{n'} - \boldsymbol{G}|^{\frac{p-q-n'-1}{2}},$$

where c_0 is a known constant. The aim is to derive $E[\boldsymbol{G}^{-1}]$. Let $\frac{d}{d\boldsymbol{G}}$ be the matrix derivative defined in Kollo & von Rosen (2005, formula (1.4.48). Then, among others, $\frac{d\boldsymbol{G}}{d\boldsymbol{G}} = \frac{1}{2}(\boldsymbol{I} + \boldsymbol{K}_{n,n})$ where $\boldsymbol{K}_{n,n}$ is the commutation matrix.

$E[(\boldsymbol{A}'\boldsymbol{S}^+\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{S}^+\boldsymbol{\Sigma}\boldsymbol{S}^+\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}^+\boldsymbol{A})^{-1}]$

The basic trick when obtaining $E[\boldsymbol{G}^{-1}]$ is to use the multivariate integration by parts formula

$$\mathbf{0} = \int_{\boldsymbol{G}>0} \frac{d}{d\boldsymbol{G}}(c|\boldsymbol{G}|^{\frac{n'-q-1}{2}}|\boldsymbol{I}_{n'} - \boldsymbol{G}|^{\frac{p-q-n'-1}{2}})d\boldsymbol{G}$$

which is equivalent to

$$\mathbf{0} = \tfrac{1}{2}(n'-q-1)\frac{d\boldsymbol{G}}{d\boldsymbol{G}}E[\mathrm{vec}\boldsymbol{G}^{-1}] - \tfrac{1}{2}(p-q-n'-1)\frac{d\boldsymbol{G}}{d\boldsymbol{G}}E[\mathrm{vec}(\boldsymbol{I}_{n'} - \boldsymbol{G})^{-1}].$$

Thus,

$$E[\boldsymbol{G}^{-1}] = E[(\boldsymbol{I} - \boldsymbol{G})^{-1}]\frac{p-q-n'-1}{n'-q-1}.$$

$E[(\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{\Sigma}\boldsymbol{S}^{+}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{-1}]$

However,

$$E[(\boldsymbol{I}-\boldsymbol{G})^{-1}] = \boldsymbol{I}_q + E[\boldsymbol{Y}_{11}\boldsymbol{W}^{-1}\boldsymbol{Y}_{11}'] = \boldsymbol{I}_q \frac{p-q-1}{p-q-n'-1}$$

and hence

$$E[\boldsymbol{G}^{-1}] = \boldsymbol{I}_q \frac{p-q-1}{n'-q-1}.$$

Properties

It can be shows that $\widehat{\boldsymbol{B}}$ is asymptotically equivalent to

$$\widetilde{\boldsymbol{B}} = (\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1},$$

i.e. is asymptotic normally distributed. Consider the following difference

$$\widehat{\boldsymbol{B}} - \widetilde{\boldsymbol{B}} = (\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{S}^{+}(\boldsymbol{I} - \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{\Sigma}^{-1})\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1}.$$

It follows that $E[\widehat{\boldsymbol{B}} - \widetilde{\boldsymbol{B}}] = \mathbf{0}$, and it can be shown that $D[\widehat{\boldsymbol{B}} - \widetilde{\boldsymbol{B}}] \to \mathbf{0}$ when $n, p \to \infty$.

Properties

The results show that if $n \to \infty$ or the $\frac{p}{n}$-asymptotics holds the estimator of the mean parameter, proposed by the approach of this paper, behaves in the same way, i.e the large number of dispersion parameters does not seriously influence the estimator of $\boldsymbol{B}$. The critical point is when $\frac{p}{n} \to 1$.

Simulation

In order to illustrate the derived results a small simulation study has been performed. Data was generated according to $\boldsymbol{X} = \boldsymbol{ABC} + \boldsymbol{E}$, where ($\mathbf{1}_a$ is a vector of a ones) the matrix $\boldsymbol{C}$ has the following structure

$$\boldsymbol{C} = \begin{pmatrix} \mathbf{1}'_{n_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1}'_{n_2} \end{pmatrix},$$

which corresponds to two different treatment groups. Moreover, let $\boldsymbol{a}'_1 = (1, 2, \ldots, p) * 0.7$, $\boldsymbol{a}'_2 = (1, 2^2, \ldots, p^2) * 0.01$ and $\boldsymbol{A} = (\mathbf{1}_p, \boldsymbol{a}_1, \boldsymbol{a}_2)$. Thus, $\boldsymbol{B}$: 3×2. The matrix $\boldsymbol{\Sigma} = \boldsymbol{QQ}'$ is generated via standard normal elements in $\boldsymbol{Q}$ and $\boldsymbol{E}$ is generated by $N_{p,n}(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{I}_n)$.

Simulation

In the simulation it was either supposed that
$p = 250$ and (n_1, n_2) equals $(20, 40)$, $(30, 60)$, $(40, 80)$, $(50, 100)$, $(60, 120)$, $(70, 140)$, $(80, 160)$,
or
$(n_1, n_2) = (10, 20)$ and $p = 50, 100, 150, 200, 250, 350$. The results of the simulations are reported in the next tables.

Simulation

Table 1. Based on 100 simulations averaged estimates of $\boldsymbol{B} = (b_{ij})$ are presented, where $N = n_1 + n_2$.

		b_{11}	b_{12}	b_{21}	b_{22}	b_{31}	b_{32}
True values		1	3	2	2	7	2
N	p	Estimates					
60	250	0.91	3.00	2.01	1.88	7.03	1.99
90	250	1.13	2.93	2.04	1.97	6.99	2.02
120	250	0.99	3.02	1.99	2.02	7.00	2.00
150	250	0.91	3.05	1.97	1.99	7.01	1.99
180	250	1.02	3.01	1.99	1.98	7.00	2.01
210	250	0.98	3.01	2.00	2.00	7.00	2.00
240	250	0.99	3.01	2.00	2.03	6.99	2.01

Simulation

Table 1 cont.. Based on 100 simulations averaged estimates of $\boldsymbol{B} = (b_{ij})$ are presented, where $N = n_1 + n_2$.

		b_{11}	b_{12}	b_{21}	b_{22}	b_{31}	b_{32}
True values		1	3	2	2	7	2
N	p	Estimates					
30	50	0.90	3.33	0.95	2.17	6.59	3.08
30	100	1.01	2.81	2.34	1.92	7.07	1.98
30	150	1.12	2.87	2.14	1.94	6.96	2.05
30	200	1.07	2.95	2.03	2.21	6.82	2.12
30	250	0.94	2.99	2.01	2.07	7.00	1.98
30	350	1.04	2.93	2.03	1.82	7.06	1.98

Simulation

From Table 1 we see that except the case $N = 30$, $p = 50$ the estimators work excellent. In the next we present the estimated standard deviation (squared root of the estimated variance) for $\widehat{\boldsymbol{B}}$.

Simulation

Table 2. Based on 100 simulations averaged square roots s_{ij} of the variance estimates for $\widehat{\boldsymbol{B}} = (\hat{b}_{ij})$ in Table 1 are presented.

N	p	s_{11}	s_{12}	s_{21}	s_{22}	s_{31}	s_{32}
60	250	0.75	0.39	0.22	0.53	0.27	0.15
90	250	0.48	0.25	0.14	0.34	0.18	0.10
120	250	0.36	0.19	0.10	0.25	0.13	0.07
150	250	0.28	0.15	0.08	0.19	0.11	0.06
180	250	0.23	0.13	0.07	0.16	0.09	0.05
210	250	0.19	0.11	0.06	0.14	0.07	0.04
240	250	0.16	0.08	0.04	0.11	0.06	0.03

Simulation

Table 2. Based on 100 simulations averaged square roots s_{ij} of the variance estimates for $\widehat{\boldsymbol{B}} = (\hat{b}_{ij})$ in Table 1 are presented.

N	p	s_{11}	s_{12}	s_{21}	s_{22}	s_{31}	s_{32}
30	50	0.50	1.34	3.69	0.70	1.90	5.22
30	100	0.99	1.27	1.78	0.70	0.90	1.26
30	150	1.23	1.12	1.04	0.87	0.79	0.74
30	200	1.41	0.94	0.64	0.99	0.66	0.45
30	250	1.58	0.82	0.46	1.11	0.58	0.32
30	350	1.66	0.71	0.32	1.17	0.50	0.23

Simulation

From Table 2 one may observe that for small N the variance estimator as expected is rather poor.

Concluding remarks

In this paper we have tried to systemize estimation in a multivariate linear model belonging to the curved exponential family when many nuisance parameters exist. In order to evaluate the estimators there are four quantities involved: $(\boldsymbol{C}\boldsymbol{C}')^{-1}$, $(\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{-1}$, p and n'. Now we know how to estimate $\boldsymbol{B}$ as well as obtain its variance irrespective if $p > n'$ or $p \leq n'$. In particular, one should be careful when p/n is close to 1 but the performance of the estimators are heavily connected to $\boldsymbol{C}$ and $\boldsymbol{A}$. In earlier approaches in high-dimensional analysis $\boldsymbol{\Sigma}^{-1}$ has also been replaced by $\boldsymbol{S}^{+}$ but in this paper it is the first time moment calculations stating the effect of $\boldsymbol{S}^{+}$ are explicitly given.

Concluding remarks

Indeed, our starting point could have been

$$\widehat{\boldsymbol{B}} = (\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{S}^{+}\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1}$$

considering this estimator as a plug-in estimator. However, we prefer to use the likelihood approach via the asymptotics of the sufficient statistics $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ presented in Section 2.