Multilinear Models

Multivariate linear model:

$$y_i = B x_i + \Sigma^{1/2} z_i, \quad z_1 \ldots z_n \sim \text{iid } N(0, I_p)$$

$B \in \mathbb{R}^{p \times q}$, $\Sigma \in \mathcal{S}_+^p$ unknown, $x_1 \ldots x_n \subset \mathbb{R}^q$ observed

$$\Rightarrow \quad y_i \sim N(B x_i, \Sigma) \quad y_1 \ldots y_n \text{ indep.}$$

Matrix form:

$$Y = \begin{pmatrix} y_1^T & - \\ \vdots & \\ y_n^T & - \end{pmatrix} = \begin{pmatrix} x_1^T B^T & - \\ \vdots & \\ x_n^T B^T & - \end{pmatrix} + \begin{pmatrix} z_1^T \Sigma^{1/2} \\ \vdots \\ z_n^T \Sigma^{1/2} \end{pmatrix}$$

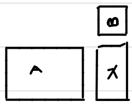
$$= \begin{pmatrix} x_1^T & - \\ & \\ x_n^T & - \end{pmatrix} B^T + \begin{pmatrix} z_1^T & - \\ \vdots & \\ z_n^T & - \end{pmatrix} \Sigma^{1/2}$$

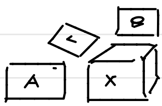
$$Y = X B^T + Z \Sigma^{1/2}$$

Model Extensions: Additive $\quad Y = A W^T + X B^T + \Psi^{1/2} E + Z \Sigma^{1/2}$

Multiplicative $\quad Y = A X B^T + \Psi^{1/2} Z \Sigma^{1/2}$

Reps : $\quad Y_i = A X_i B^T + \Psi^{1/2} Z_i \Sigma^{1/2} \quad i = 1 \ldots n$ independently

Extending to tensors: Matrix $\quad E(Y) = A X B^T =$ 

Tensor 

Vec & Kronecker product

$$Y = \underset{n\times q}{X}\,\underset{q\times p}{B} + Z\Sigma^{1/2}$$

(note writing as XB, not XB^T)

Let
$$y = \text{vec}(Y) = \begin{bmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{n1} \\ y_{12} \\ \vdots \\ y_{n2} \\ \vdots \\ y_{1p} \\ \vdots \\ y_{np} \end{bmatrix} \begin{array}{l} \left.\begin{array}{l} \\ \\ \\ \end{array}\right\} \text{1st col} \\ \left.\begin{array}{l} \\ \\ \end{array}\right\} \text{2nd col} \\ \vdots \\ \left.\begin{array}{l} \\ \\ \end{array}\right\} \text{pth col} \end{array}$$

$$\text{vec}(XB) = \begin{bmatrix} x_1^T b_1 \\ x_2^T b_1 \\ \vdots \\ x_n^T b_1 \\ \vdots \\ x_1^T b_p \\ \vdots \\ x_n^T b_p \end{bmatrix} = \begin{bmatrix} \overbrace{\boxed{\begin{matrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{matrix}}}^{X!} & \begin{matrix} 0 \\ 0 \\ \\ 0 \end{matrix} & \begin{matrix} 0 \\ \cdots \\ \\ 0 \end{matrix} & \begin{matrix} 0 \\ \\ \\ \cdots \end{matrix} \\ 0 & \boxed{\begin{matrix} x_1^T \\ x_2^T \\ x_n^T \end{matrix}} & \begin{matrix} 0 \\ 0 \\ 0 \end{matrix} & \cdots \\ \cdots & \cdots & \cdots & \boxed{\begin{matrix} x_1^T \\ \vdots \\ x_n^T \end{matrix}} \end{bmatrix} \begin{bmatrix} b_{11} \\ \vdots \\ b_{q1} \\ \vdots \\ b_{1p} \\ \vdots \\ b_{qp} \end{bmatrix}$$

$$= (I \otimes X)\,\text{vec}(B)$$

Def (Kronecker Product) for $A \in \mathbb{R}^{n_1 \times p_1}$, $B \in \mathbb{R}^{n_2 \times p_2}$ then

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & & a_{1p_1}B \\ a_{21}B & & \ddots & \\ & & & a_{n_1p_1}B \end{bmatrix} \in \mathbb{R}^{n_1 n_2 \times p_1 p_2}$$

Thm (vec-Kronecker Identity)

$$\text{vec}(AXB^T) = (B \otimes A)\,x, \quad x = \text{vec}(X)$$

Tricks: $\quad \text{vec}(AB) = \text{vec}(ABI) = (I \otimes A)\,\text{vec}(B)$

$= \text{vec}(IAB) = (B^T \otimes I)\,\text{vec}(A)$

Demo: OLS MV Regression estimator:

find B to minimize $\|Y - XB^T\|^2$ $\qquad b = \text{vec}(B^T)$

$\|\underset{np\times 1}{y} - (I \otimes X)b\|^2$

$= y^T y - 2b^T(I\otimes X)^T y + b^T(I\otimes X)^T(I\otimes X)b$

Identity: $(A\otimes B)^T = (A^T \otimes B^T)$

$(A\otimes B)(C\otimes D) = (AC \otimes BD)$

$= y^T y - b^T(I\otimes X^T)y + b^T(I\otimes X^T X)b$

Minimize in b: $\quad \frac{\partial}{\partial b}$ " $= -2(I\otimes X^T)y + 2(I\otimes X^T X)b$

$\hat{b} = (I\otimes X^T X)^{-1}(I\otimes X^T)y$

Identity $(A\otimes B)^{-1} = (A^{-1}\otimes B^{-1})$

$= (I \otimes (X^T X)^{-1})(I\otimes X^T)\,y$

$= (I\otimes (X^T X)^{-1}X^T)\,y$

$\hat{b} = (I\otimes (X^T X)^{-1}X^T)y \Rightarrow \underset{q\times p}{\hat{B}^T} = \underset{q\times q}{(X^T X)^{-1}}\,\underset{q\times n}{X^T}\,\underset{n\times p}{Y}$

$\underset{p\times q}{\hat{B}} = Y^T X (X^T X)^{-1}$

Exercise: Find MLE of B for case $y_i \sim N_p(Bx_i, \Sigma)$, Σ known.

i.e. minimize $\|(Y - XB^T)\Sigma^{-1/2}\|^2$

Kronecker Covariance

Let $y_1 \ldots y_n \sim \text{iid } N_p(0, \Sigma)$

Then $y_i = \Sigma^{1/2} z_i \qquad z_1, \ldots, z_n \sim \text{iid } N(0, I)$

$Y = Z \Sigma^{1/2} \qquad \{z_{ij}\} \sim \text{iid } N(0,1)$

Covariance of Y: Intuitively, $\text{Cov}(y_{ij}, y_{ij'}) = \sigma_{jj'}$

$\text{Cov}(y_{ij}, y_{i'j'}) = 0$

Cov of $\text{vec}(Y)$: $\text{vec}(Y) = y = (\Sigma^{1/2} \otimes I) z \quad , \; z = \text{vec}(Z)$

$E[y] = 0$, so

$$
\begin{aligned}
\text{Cov}(y) &= E[y y^T] \\
&= E[(\Sigma^{1/2} \otimes I) z z^T (\Sigma^{1/2} \otimes I)] \\
&= (\Sigma^{1/2} \otimes I) E(z z^T) (\Sigma^{1/2} \otimes I) \\
&= (\Sigma^{1/2} \otimes I)(\Sigma^{1/2} \otimes I) \\
&= \underline{\underline{\Sigma \otimes I}}
\end{aligned}
$$

row covariance

column covariance

$$
\Rightarrow \text{cov}\begin{pmatrix} y_{11} \\ \vdots \\ y_{n1} \\ y_{12} \\ \vdots \\ y_{n2} \\ \vdots \\ y_{1p} \\ \vdots \\ y_{np} \end{pmatrix} = \Sigma \otimes I = \begin{bmatrix} \sigma_{11} I & \sigma_{12} I & \\ \sigma_{12} I & \sigma_{22} I & \\ & & \ddots \end{bmatrix}
$$

Row + Column covariance:

$$Y = \Psi^{1/2} Z \Sigma^{1/2} \quad , \quad \Psi \in \mathcal{S}_n^+ \; , \quad \Sigma \in \mathcal{S}_p^+$$

What is $\text{Cov}(y_{ij}, y_{i'j'})$?

Let $y = \text{vec}(Y)$, so $\quad y = (\Sigma^{1/2} \otimes \Psi^{1/2}) z$

$$\begin{aligned} C(y) = E(yy^T) &= (\Sigma^{1/2} \otimes \Psi^{1/2}) E(zz^T)(\Sigma^{1/2} \otimes \Psi^{1/2}) \\ &= (\Sigma^{1/2} \otimes \Psi^{1/2})(\Sigma^{1/2} \otimes \Psi^{1/2}) \\ &= \Sigma \otimes \Psi \\ &= \begin{bmatrix} \sigma_{11}\Psi & \sigma_{12}\Psi & & \\ \sigma_{21}\Psi & & & \\ & & \ddots & \\ & & & \sigma_{pp}\Psi \end{bmatrix} \end{aligned}$$

$\rightarrow \text{Cov}(y_{ij}, y_{i'j'}) = \psi_{ii'} \sigma_{jj'}$

Ψ represents covariance of the rows
Σ represents covariance of the columns

Kronecker / Separable / Multilinear Covariance model

If Z is normal, sometimes called "Matrix normal model" (Dawid 19..) but this isn't a great name.

<u>Note</u> (Nonidentifiability) $\quad \Psi \otimes \Sigma = (\Psi/a) \otimes (a\Sigma)$

So scales of Ψ, Σ not separately identifiable.

Estimation: First consider Moments, MLE, then Bayes

Observe $Y \sim N_{n \times p}(0, \Sigma \otimes \Psi)$. Goal: estimate Σ, Ψ or $\Sigma \otimes \Psi$.

Moment estimators: "Typical" estimator of Σ is $Y^T Y / n$

Recall $Y \overset{\mathcal{L}}{=} \Psi^{1/2} Z \Sigma^{1/2}$, $\quad E[\, Y^T Y \,] = E[\, \Sigma^{1/2} Z^T \Psi Z \Sigma^{1/2} \,]$

$$= \Sigma^{1/2} E[Z^T \Psi Z] \Sigma^{1/2}$$

Exercise: Show $E[Z^T \Psi Z] = I \times \mathrm{tr}(\Psi)$

So $\quad E[\, Y^T Y \,] = \Sigma^{1/2} (I \cdot \mathrm{tr}(\Psi)) \Sigma^{1/2} = \Sigma \times \mathrm{tr}(\Psi)$

scale

Exercise: Show $E[YY^T] = \Psi \times \mathrm{tr}(\Sigma)$

So! $\quad Y^T Y$ is an unbiased est of Σ (up to scale)
$\quad YY^T$ " Ψ "

However: $\quad YY^T \in \mathbb{R}^{n \times n}$, but $\mathrm{rank}(YY^T) = p$!

Related to this, $\quad \log p(Y \mid \Sigma, \Psi)$ is unbounded in (Σ, Ψ)
(unless $n = p$)

$\Rightarrow$ No MLE
Need a proper prior to do Bayes inference.

<u>Replications</u>: Suppose $Y_1 \ldots Y_r \sim \text{iid } N_{n \times p}(0, \Sigma \otimes \Psi)$

What is the likelihood?

$y_1 \ldots y_r \sim \text{iid } N_{np}(0, \Sigma \otimes \Psi) \Rightarrow$

$$p(y_1 \ldots y_r) \propto |\Sigma \otimes \Psi|^{-r/2} \exp\{-\tfrac{1}{2} \sum_{i=1}^{r} y_i^T (\Sigma \otimes \Psi)^{-1} y_i)\}$$

<u>Identity</u>: $|\underset{p \times p}{A} \otimes \underset{n \times n}{B}| = |A|^n |B|^p$

Also,
$$\begin{aligned} y_i^T (\Sigma \otimes \Psi)^{-1} y_i &= y_i^T (\Sigma^{-1/2} \otimes \Psi^{-1/2})(\Sigma^{-1/2} \otimes \Psi^{-1/2}) y_i \\ &= \| (\Sigma^{-1/2} \otimes \Psi^{-1/2}) y_i \|^2 \\ &= \| \Psi^{-1/2} Y \Sigma^{-1/2} \|^2 \\ &= tr((\Psi^{-1/2} Y \Sigma^{-1/2})(\Sigma^{-1/2} Y^T \Psi^{-1/2})) \\ &= tr(\Psi^{-1/2} Y \Sigma^{-1} Y \Psi^{-1/2}) \\ &= tr(\Psi^{-1} Y_i \Sigma^{-1} Y_i^T) \end{aligned}$$

$$p(y_1 \ldots y_r | \Sigma, \Psi) \propto |\Sigma|^{-\frac{nr}{2}} |\Psi|^{-\frac{pr}{2}} \, etr(-\tfrac{1}{2} \sum_{i=1}^{r} \Psi^{-1} Y_i \Sigma^{-1} Y_i^T)$$

<u>Iterative estimation</u>:

$$p(Y_1 \ldots Y_n | \Sigma, \Psi) \underset{\Sigma}{\propto} |\Sigma|^{-\frac{nr}{2}} etr(-\Sigma^{-1} S_\Sigma) , \quad S_\Sigma = \sum_{i=1}^{r} Y_i^T \Psi^{-1} Y_i$$

$\Longrightarrow$ optimizer in Σ is $\tilde{\Sigma} = S_\Sigma / nr$, total # of rows

$\Longrightarrow$ lik is prop. to inverse-Wishart dist in Σ

$$p(Y_1 \ldots Y_n | \Sigma, \Psi) \underset{\Psi}{\propto} |\Psi|^{-\frac{pr}{2}} etr(-\Psi^{-1} S_\Psi / 2), \quad S_\Psi = \sum_{i=1}^{r} Y_i \Sigma^{-1} Y_i^T$$

$\Longrightarrow$ optimizer in Ψ is $\tilde{\Psi} = S_\Psi / pr$, total # of cols

$\Rightarrow$ lik is prop to inv. wish dist in Ψ.

Comments:

* MLE via Block coordinate descent ("flip flop algorithm")

$$\underbrace{\tilde{\Psi}}_{n \times n} = \frac{1}{pr} \sum_{i=1}^{r} \underbrace{Y_i \Sigma^{-1} Y_i^T}_{\text{rank } p}$$

is generally of rank pr

will need $pr \geq n$ for $\tilde{\Psi}$ to have an inverse

$r \geq n/p$ (needed to update Σ).

However: $r = n/p$ is not generally enough to guarantee lik. is bounded

$r \geq n+p$ is, but results for $\frac{n}{p} < r < n+p$ not available

(see, e.g. Roś et al. 2016 JMVA)

* Bayes: No sample size limitations if proper priors are used.

* Reduced models: y_{ij} = observation of variable j at time i

$$\Rightarrow \text{maybe restrict } \Psi = \Psi(\theta) = \tau^2 \begin{bmatrix} 1 & \rho & \rho^2 & \cdots \\ \rho & 1 & \rho & \\ \rho^2 & \rho & 1 & \\ \vdots & & & \end{bmatrix}$$

In this case, MLEs may exist, but estimation procedure is a bit more complicated.