Nonparametric covariance estimation for longitudinal data via tensor product smoothing

Tayler Blake [1] Dr. Yoonkyung Lee [2]

[1] Information Control Company

[2] The Ohio State University, Department of Statistics

July 26, 2017

The data:

$$Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{iM_i})', \qquad i = 1, \ldots, N$$

associated with measurement times

$$t_1 < t_2 < \cdots < t_{M_i}.$$

Goal: estimate

$$Cov\,(Y) = \Sigma$$

The flaming hoops:

- Covariance matrices (and their estimates) should be positive definite.
  - Constrained optimization is a headache.
- The $\{t_{ij}\}$ may be suboptimal.
  - Observation times may not fall on a regular grid, may vary across subjects.
- More dimensions, more problems (maybe.)
  - Sample covariance matrix falls apart when m is large.

The flaming hoops:

- Covariance matrices (and their estimates) should be positive definite.
  - Constrained optimization is a headache.
- The $\{t_{ij}\}$ may be suboptimal.
  - Observation times may not fall on a regular grid, may vary across subjects.
- More dimensions, more problems (maybe.)
  - Sample covariance matrix falls apart when m is large.

The flaming hoops:

- Covariance matrices (and their estimates) should be positive definite.
  - Constrained optimization is a headache.
- The $\{t_{ij}\}$ may be suboptimal.
  - Observation times may not fall on a regular grid, may vary across subjects.
- More dimensions, more problems (maybe.)
  - Sample covariance matrix falls apart when m is large.

The flaming hoops:

- Covariance matrices (and their estimates) should be positive definite. A cute little reparameterization $\Longrightarrow$ unconstrained estimation, meaningful interpretation
- The $\{t_{ij}\}$ may be messy.
  Frame covariance estimation as function estimation.
- More dimensions, more problems (maybe.)

Figure: Regulate like Nate Dogg.

Covariance dress-up: the modified Cholesky decomposition

$$Y = (Y_1, \ldots, Y_M)' \sim \mathcal{N}(0, \Sigma).$$

For any positive definite Σ, we can find T which diagonalizes Σ:

$$D = T\Sigma T', \quad T = \begin{bmatrix} 1 & 0 & \ldots & & \\ -\phi_{21} & 1 & & & \\ -\phi_{31} & -\phi_{32} & 1 & & \\ \vdots & & & \ddots & \\ -\phi_{M1} & -\phi_{M2} & \ldots & -\phi_{M,M-1} & 1 \end{bmatrix}$$

Now, for the cutest part:

Okay, really:

Regress Y_j on $Y_{(1:j-1)} = (Y_1, \ldots, Y_{j-1})'$:

$$y_j = \begin{cases} e_1 & j = 1, \\ \sum_{k=1}^{j-1} \phi_{jk} y_k + \sigma_j e_j & j = 2, \ldots, M \end{cases} \tag{1}$$

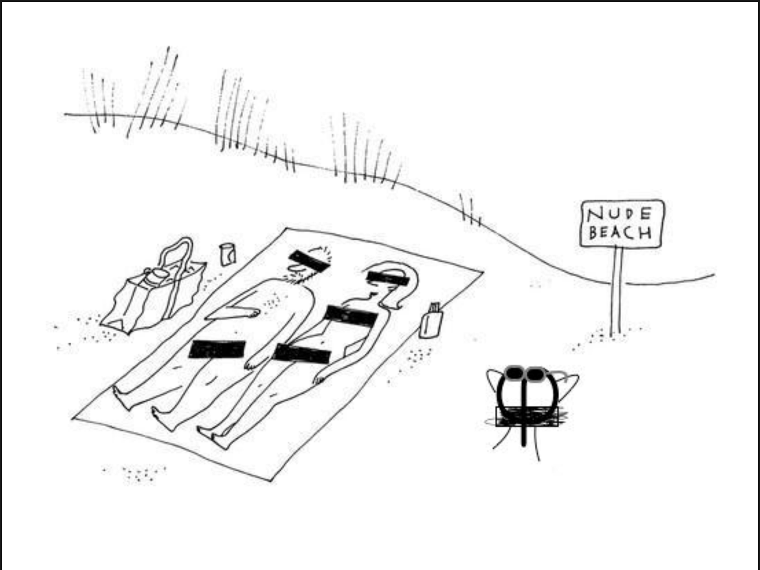
In matrix form:

$$e = TY, \tag{2}$$

and taking covariances on both sides:

$$D = \sigma_e^2 \mathrm{diag}\left(\sigma_1^2, \ldots, \sigma_M^2\right) = T \Sigma T'. \tag{3}$$

No constraints on the ϕ_{jk}s!



The regression model tool box is a deep, luxurious toolbox.

$$Y_j \longrightarrow Y(t_j) \qquad e_j \longrightarrow e(t_j)$$

$$\phi_{jk} \longrightarrow \phi(t_j, t_k) \qquad \sigma_j^2 \longrightarrow \sigma^2(t_j)$$

$$y(t_j) = \sum_{k=1}^{j-1} \phi(t_j, t_k)\, y(t_k) + \sigma(t_j)\, e(t_j), \tag{4}$$

where

$$e(s) \sim \mathcal{WN}\left(0, \sigma_e^2\right)$$

Regularization of $\phi(s,t)$ is more intuitive if we transform the s-t axis.

$$l = s - t$$
$$m = \frac{1}{2}(s+t)$$

Reparameterize ϕ:

$$\phi(s,t) = \phi^*(l,m) = \phi^*\left(s-t, \frac{1}{2}(s+t)\right)$$

Take $\hat{\phi}^*$ to be the minimizer of

$$-2L_\phi(\phi, y_1, \dots, y_N) = \sum_{i=1}^{n} \sum_{j=2}^{m_i} \sigma_{ij}^{-2} \left(y_{ij} - \sum_{k=1}^{j-1} \phi(t_{ij}, t_{ik})\, y_{ik} \right)^2$$

Smooth ANOVA models

Decompose

$$\phi^*(l,m) = \mu + \phi_1(l) + \phi_2(m) + \phi_{12}(l,m), \tag{6}$$

so Model 4 becomes

$$y(t_j) = \sum_{k=1}^{j-1} \Big[\mu + \phi_1(l_{jk}) + \phi_2(m_{jk}) \\ + \phi_{12}(l_{jk}, m_{jk})\Big] y(t_k) + \sigma(t_j)\, e(t_j)$$

Approximate ϕ_1, ϕ_2, ϕ_{12} with B-splines.

$$
\begin{aligned}
\phi^*(l,m) &= B\theta \\
&= \phi_1(l) + \phi_2(m) + \phi_{12}(l,m) \\
&= \sum_{c=1}^{c_l} B_c(l)\,\theta_{lc} B_l\theta_l + \sum_{c'=1}^{c_m} B_{c'}(m)\,\theta_{mc'} \\
&\qquad + \sum_{c=1}^{c_l}\sum_{c'=1}^{c_m} B_c(l)\,B_{c'}(m)\,\theta_{cc'}
\end{aligned} \tag{7}
$$

where $\theta \equiv (\mu, \theta_l, \theta_m, \theta_{lm})'$ and

$$
\begin{aligned}
B &\equiv [\, 1_p \mid B_l \mid B_m \mid B_{lm} \,] \\
&= [\, 1_p \mid B_l \mid B_m \mid B_m \,\Box\, B_l]
\end{aligned}
$$

PS-ANOVA Penalty

Find θ minimizing

$$\ell(\theta,\lambda) = (Y - WB\theta)' D^{-1} (Y - WB\theta) + \theta' P \theta$$
$$= \sum_{i=1}^{n} \sum_{j=2}^{m_i} \sigma_{ij}^{-2} \left(y_{ij} - \sum_{k=1}^{j-1} \phi(t_{ij}, t_{ik}) y_{ik} \right)^2 \tag{8}$$
$$+ \lambda_l ||D_l \theta_l||^2 + \lambda_m ||D_m \theta_m||^2 + \lambda_{lm} ||D_{lm} \theta_{lm}||^2$$

$$P = \begin{bmatrix} 0 & & & \\ & \underbrace{\lambda_l D'_{d_l} D_{d_l}}_{P_l} & & \\ & & \underbrace{\lambda_l D'_{d_l} D_{d_l}}_{P_m} & \\ & & & \underbrace{\tau_m D'_{d_m} D_{d_m} \otimes I_{c_l} + \tau_l I_{c_m} \otimes D'_{d_l} D_{d_l}} \end{bmatrix}$$

Mixed model representation

Find orthogonal transformation $\mathrm{Q} = \left[\ \mathrm{Q}_0 \,\middle|\, \mathrm{Q}_1\ \right]$; map

$$\begin{aligned} BQ &= \left[\ B\mathrm{Q}_0 \,\middle|\, B\mathrm{Q}_1\ \right] \\ &= \left[\ X \,\middle|\, Z\ \right] \end{aligned}$$

such that

$$B\theta = X\beta + Z\alpha$$

Mixed model representation

Model 8 becomes

$$Y = W\left(X\beta + Z\alpha\right) + e \tag{9}$$

$$\alpha \sim \mathcal{N}\left(0, G\right), \quad e \sim \mathcal{N}\left(0, \sigma_e^2 D\right)$$

For $d_l = d_m = 2$,

$$X = \left[\ 1 \mid 1\square l \mid m\square 1 \mid m\square l\ \right]$$
$$Z = \left[\ 1\square Z_l \mid Z_m\square 1 \mid m\square Z_l \mid Z_m\square l \mid Z_m\square Z_l\ \right]$$

Nested B-spline bases

$$\phi_{12}(l,m) = \sum_{r=1}^{d_m-1} m^r g_{lr}(l) + \sum_{r'=1}^{d_l-1} l^{r'} g_{mr'}(m) + h(l,m),$$

For $d_l = d_m = 2$,

$$\phi_{12}(l,m) = g_1(l)\; m + l\; g_2(m) + \; h(l,m)$$

$$B = \left[\, 1 \mid B_1 \mid B_2 \mid B_3 \mid B_4 \mid B_5 \,\right], \tag{10}$$

where

$$B_3 = m \,\square\, B_1, \qquad B_4 = B_2 \square\, l, \qquad B_5 = B_2 \,\square\, B_1$$

$$P = \text{blockdiag}\left(0,\; P_1,\; P_2,\; P_3,\; P_4,\; P_5\right),$$

Smoothing parameter selection = variance component estimation

$G = \sigma_e^2 F^{-1}$ where

$$F = \begin{bmatrix} F_l & & \\ & F_m & \\ & & F_{lm} \end{bmatrix} = \begin{bmatrix} \lambda_l \tilde{\Delta}_l & & \\ & \lambda_m \tilde{\Delta}_m & \\ & & F_{lm} \end{bmatrix},$$

$$F_{lm} = \begin{bmatrix} \tau_l \tilde{\Delta}_l & & \\ & \tau_m \tilde{\Delta}_m & \\ & & \tau_m \tilde{\Delta}_m \otimes I_{c_l - d_l} + I_{c_m - d_m} \otimes \tau_l \tilde{\Delta}_l \end{bmatrix}$$

Now the variance components have simpler structure.

Let $\alpha_k \sim \mathcal{N}(0, G_k)$,

$$
\begin{aligned}
Z &= \left[\ Z_1 \mid Z_2 \mid Z_3 \mid Z_4 \mid Z_5\ \right] \\
&= \left[\ \underbrace{1 \square Z_l}_{\alpha_1} \mid \underbrace{Z_m \square 1}_{\alpha_2} \mid \underbrace{m \square Z_l}_{\alpha_3} \mid \underbrace{Z_m \square l}_{\alpha_4} \mid \underbrace{Z_m \square Z_l}_{\alpha_5}\ \right]
\end{aligned}
$$

Take $Z_k^* = Z_k F_k^{-1}$, then

$$
\begin{aligned}
G &= \text{blockdiag}\left(G_1,\ G_2,\ G_3,\ G_4,\ G_5\right) \\
&= \text{blockdiag}\left(\tau_1^2 I_{c_l-2},\ \tau_2^2 I_{c_m-2},\ \tau_3^2 I_{c_l-2},\ \tau_4^2 I_{c_m-2},\ \tau_5^2 I_{(c_l-2)(c_m-2)}\right)
\end{aligned}
$$

REML for σ_e^2 and τ_k^2

Take $\hat{\sigma}_e^2$, $\hat{\tau}_k^2$ to minimize

$$-2\mathcal{L} = n \log \sigma_e + \sum_{j=1}^{5} \left(k_j \log \tau_j + \log \sigma_e + \frac{\alpha_j' \alpha_j}{\tau_j^2} \right) + \sigma_e^{-2} \left(Y - W \left(X\beta - Z\alpha \right) \right)' \left(Y - W \left(X\beta - Z\alpha \right) \right)$$

$$\hat{\tau}_k^2 = \frac{\hat{\alpha}_k' \hat{\alpha}_k}{\mathrm{ED}_j},$$

$$\hat{\sigma}_e^2 = \frac{\left(Y - W \left(X\hat{\beta} - Z\hat{\alpha} \right) \right)' \left(Y - W \left(X\hat{\beta} - Z\hat{\alpha} \right) \right)}{N - \mathrm{ED}}$$