

Nonparametric covariance estimation for longitudinal data via tensor product smoothing

Tayler A. Blake

July 6, 2017

The data:

$$Y_i = (Y_{i1}, Y_{i2}, \dots, Y_{im})', \quad i = 1, \dots, N$$

associated with measurement times

$$t_1 < t_2 < \dots < t_m.$$

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 \implies unconstrained estimation, meaningful interpretation
- ▶ The $\{t_{ij}\}$ may be messy.
Frame covariance estimation as function estimation.
- ▶ More dimensions, more problems (maybe.)

The flaming hoops:

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

The flaming hoops:

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

The flaming hoops:

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

The flaming hoops:

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

The flaming hoops:

- ▶ Covariance matrices (and their estimates) should be positive definite. A cute little reparameterization \implies 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

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

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

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

Now, for the cutest part:



Okay, really:

Imagine regressing Y_j on its predecessors:

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

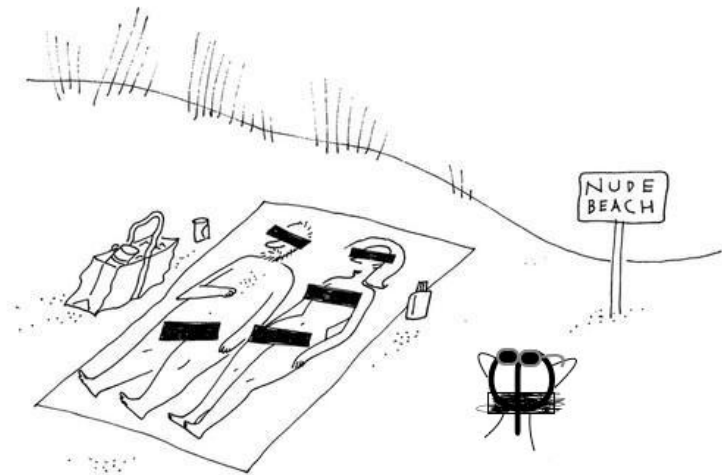
In matrix form:

$$e = TY, \quad (2)$$

and taking covariances on both sides:

$$D = \text{diag}(\sigma_1^2, \dots, \sigma_M^2) = T\Sigma T'. \quad (3)$$

No constraints on the ϕ_{jk} s!



The regression model tool box: a deep treasure chest of luxury.

Model Y_j, e_j as

$$\begin{aligned} Y_j &= Y(t_j), \quad e_j = e(t_j), \\ e(s) &\sim \mathcal{WN}(0, 1), \end{aligned}$$

Swap the standard regression model 1 for a varying coefficient model:

$$\phi_{jk} = \phi(t_j, t_k),$$

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

Penalized maximum likelihood estimation

1. Fix $\sigma_{ij}^2 = \sigma_{ij0}^2$, $i = 1, \dots, N$, $j = 1, \dots, M$.
2. Find $\phi_0 = \arg \min_{\phi} -2L_{\phi}(\phi, y_1, \dots, y_N) + \lambda J(\phi)$
3. Fix $\phi = \phi_0$.
4. Find $\sigma_0^2 = \arg \min_{\sigma^2} -2L_{\sigma}^2(\sigma^2, y_1, \dots, y_N) + \lambda J(\sigma^2)$

$$-2L_{\phi}(\phi, y_1, \dots, y_N) = \sum_{i=1}^N \sum_{j=2}^{m_i} \sigma_{ij0}^{-2} \left(y_{ij} - \sum_{k=1}^{j-1} \phi(t_{ij}, t_{ik}) y_{ik} \right)^2$$

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

Rotate the input axes:

$$\begin{aligned}l &= s - t \\ m &= \frac{1}{2} (s + t) .\end{aligned}$$

Then ϕ becomes

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

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

$$-2L + \lambda J(\phi^*)$$

Smooth ANOVA models

Decompose

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

so Model 4 becomes

$$y(t_j) = \sum_{k=1}^{j-1} \left[\mu + \phi_1(l_{jk}) + \phi_2(m_{jk}) + \phi_{12}(l_{jk}, m_{jk}) \right] y(t_k) + \sigma(t_j) e(t_j) \quad (7)$$

We can use B-splines to construct the model basis.

$$\begin{aligned}\phi_1(l) &= \sum_{c=1}^{c_l} B_c(l; q_l) \theta_{lc}, \\ \phi_2(m) &= \sum_{c'=1}^{c_m} B_{c'}(m; q_m) \theta_{mc'},\end{aligned}\tag{8}$$

$$\phi_{12}(l, m) = \sum_{c=1}^{c_l} \sum_{c'=1}^{c_m} B_c(l; q_l) B_{c'}(m; q_m) \theta_{cc'}\tag{9}$$

PS-ANOVA model basis

In matrix notation, Model 7 becomes

$$E[Y|W] = WB\theta,$$

where W is the matrix of covariates holding the past values of Y , and B is the B -spline regression basis:

$$B = [1_p \mid B_l \mid B_m \mid B_{lm}] \quad (10)$$

where

$$\begin{aligned} B_{lm} &= B_m \square B_l \\ &\equiv (B_m \otimes 1'_{c_l}) \odot (1'_{c_m} \otimes B_l) . \end{aligned}$$

Difference penalty had to regulate.

For $f(x) = \sum_{i=1}^p B_i(x) \theta_i$, approximate

$$\begin{aligned} \int_0^1 (f''(x))^2 dx &= \int_0^1 \left\{ \sum_{i=1}^p B_i''(x) \theta_i \right\}^2 dx \\ &= k_1 \sum_i (\Delta^2 \theta_i)^2 + k_2, \end{aligned} \tag{11}$$

by

$$\|D_2 \theta\|^2, \quad D_2 \theta = (\Delta^2 \theta_1, \dots, \Delta^2 \theta_{p-2})'$$

In general,

approximate $\int_0^1 (f^{(d)})^2 dx$ with $\|D_d \theta\|^2$

PS-ANOVA Penalty

Estimate B -spline coefficients by minimizing

$$(Y - WB\theta)'(Y - WB\theta) + \theta'P\theta$$

where

$$P = \text{blockdiag}(0, P_l, P_m, P_{lm}), \quad (12)$$

$$P_l = \lambda_l D'_{d_l} D_{d_l}$$

$$P_m = \lambda_m D'_{d_m} D_{d_m}$$

$$P_{12} = \tau_l D'_{d_l} D_{d_l} \otimes I_{c_m} + \tau_m I_{c_l} \otimes D'_{d_m} D_{d_m}$$

Mixed model representation

In matrix notation, Model 7 is given by

$$E[Y|W] = WB\theta,$$

where W is the matrix of covariates holding the past values of Y , and B is the B -spline regression basis:

$$B = [1_p \mid B_l \mid B_m \mid B_{lm}] \quad (13)$$

$$B_{lm} = (B_m \otimes 1'_{cl}) \odot (1'_{cm} \otimes B_l),$$

and

$$\theta = (\mu, \theta_l, \theta_l, \theta_{lm})'$$

Mixed model representation

Find transformation M to map

$$B \longrightarrow [X \mid Z], \quad \theta \longrightarrow (\beta, \alpha)'$$

such that

$$BM = [X \mid Z], \quad B\theta = X\beta + Z\alpha$$

and

$$\begin{aligned} Y &= W(X\beta + Z\alpha) + e, \\ \alpha &\sim \mathcal{N}(0, G), \quad e \sim \mathcal{N}(0, D) \end{aligned} \tag{14}$$

Mixed model representation

$$G \longrightarrow F^{-1}, \quad \theta' P \theta \longrightarrow \alpha' F \alpha$$

$$F = \text{blockdiag} \left(\underbrace{F_l}_{\lambda_l \tilde{\Delta}_l}, \underbrace{F_m}_{\lambda_m \tilde{\Delta}_m}, F_{lm} \right),$$

$$\left[\begin{array}{ccc} \lambda_l \tilde{\Delta}_l & & \\ & \lambda_m \tilde{\Delta}_m & \\ & & \left[\begin{array}{ccc} \tau_l \tilde{\Delta}_l & & \\ & \tau_m \tilde{\Delta}_m & \\ & & \tau_m \tilde{\Delta}_m \otimes I_{c_l-d_l} + I_{c_m-d_m} \otimes \tau_l \tilde{\Delta}_l \end{array} \right] \end{array} \right]$$

Decomposition of ϕ^* for $d_l = d_m = 2$

	$\{1\}$	$\{m\}$	$\{B_{j'}(m)\}$
$\{1\}$	$\{1\}$	$\{m\}$	$\{B_{j'}(m)\}$
$\{l\}$	$\{l\}$	$l \times m$	$l \times \{B_{j'}(m)\}$
$\{B_j(l)\}$	$\{B_j(l)\}$	$m \times \{B_j(l)\}$	$\{B_j(l) B_{j'}(m)\}$

Nested PS-ANOVA

Re-express ϕ_{12} :

$$\phi_{12}(l, m) = g_1(l) \left[\sum_{r=1}^{d_m-1} m^r \right] + \left[\sum_{r=1}^{d_l-1} l^r \right] g_2(m) + h(l, m),$$

For $d_l = d_m = 2$,

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

with basis:

$$B = [1_p \mid B_1 \mid B_2 \mid B_3 \mid B_4 \mid B_5], \quad (15)$$

where

$$B_3 = m \square B_1$$

$$B_5 = B_2 \square B_1$$

$$B_4 = B_2 \square l$$

Nested PS-ANOVA

$$P = \text{blockdiag}(0, P_1, P_2, P_3, P_4, P_5), \quad (16)$$

Nested PS-ANOVA

Re-express ϕ_{12} :

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

For $d_l = d_m = 2$,

$$\phi_{12}(l, m) = g_{l1}(l) \ m + l \ g_{m1}(m) + h(l, m)$$

with basis:

$$B = [1_p \mid B_1 \mid B_2 \mid B_3 \mid B_4 \mid B_5], \quad (17)$$

where

$$B_3 = m \square B_1$$

$$B_5 = B_2 \square B_1$$

$$B_4 = B_2 \square l$$

Nested PS-ANOVA

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

Remove the redundant columns, and give each penalized component its own random effect:

$$B$$