Nonparametric covariance estimation for longitudinal data via tensor product smoothing

Tayler A. Blake

July 3, 2017

The data:

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

associated with measurement times

$$t_1 < t_2 < \cdots < t_m.$$

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.)

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:

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

The matrix T is the *Cholesy factor* of the precision matrix.

Now, for the cutest part:

Okay, really:

Imagine regressing Y_j on its predecessors:

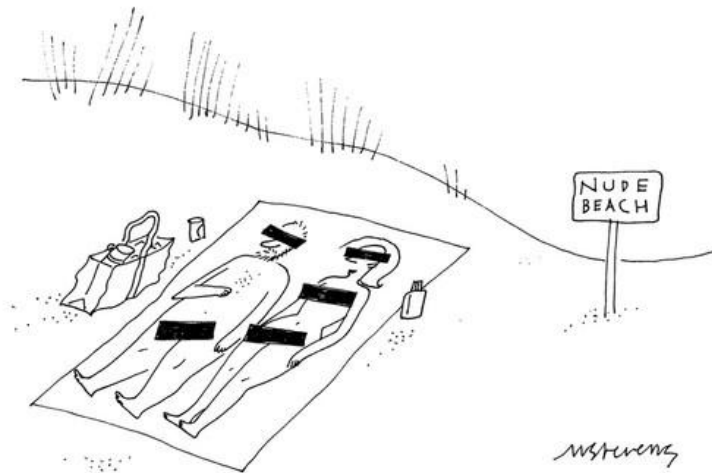$$Y_j = \begin{cases} e_1 & j = 1, \\ \sum\limits_{k=1}^{j-1} \phi_{jk} Y_k + \sigma_j e_j & j = 2, \ldots, M \end{cases} \tag{2}$$

In matrix form:

$$e = TY, \tag{3}$$

and taking covariances on both sides:

$$D = diag\left(\sigma_1^2, \ldots, \sigma_M^2\right) = T\Sigma T'. \tag{4}$$

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

Model Y_j, e_j as

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

Swap the standard regression model 2 for a varying coefficient model:

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

$$y(t_j) = \sum_{k=1}^{j-1} \phi(t_j, t_k) y(t_k) + \sigma(t_j) e(t_j) \qquad (5)$$

The $\{\phi_{jk}\}$ are called *generalized autoregressive parameters*.
The $\{\sigma_j^2\}$ are called the *innovation variances*.

(Iterated) penalized maximum likelihood estimation

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

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

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

Rotate the input axes:

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

Then ϕ becomes

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

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

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

Smooth ANOVA models

Decompose

$$\phi^*(l, m) = \mu + \phi_1(l) + \phi_2(m) + \phi_{12}(l, m), \qquad (7)$$

so Model 5 becomes

$$y(t_j) = \sum_{k=1}^{j-1} \left[\mu + \phi_1(l_{jk}) + \phi_2(m_{jk}) + \phi_{12}(l_{jk}, m_{jk}) \right] y(t_k) + \sigma(t_j)\, \epsilon(t_j) \qquad (8)$$

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

Represent the main effects as

$$\phi_1\left(l\right) = \sum_{c=1}^{c_l} B_c\left(l; q_l\right)\theta_{lc}, \qquad (9)$$

$$\phi_2\left(m\right) = \sum_{c'=1}^{c_m} B_{c'}\left(m; q_m\right)\theta_{mc'}, \qquad (10)$$

and the interaction term by the tensor product of the marginal bases 9 and 10:

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

PS-ANOVA model basis

In matrix notation, Model 8 becomes

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

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

$$B = \left[\, 1_p \mid B_l \mid B_m \mid B_{lm} \,\right] \qquad (11)$$

$$B_{lm}\left(B_m \otimes 1'_{c_l}\right) \odot \left(1'_{c_m} \otimes B_l\right),$$

and

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

Difference penalty had to regulate.

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

$$\int_0^1 (f''(x))^2 \, dx = \int_0^1 \left\{ \sum_{i=1}^{p} B_i''(x)\,\theta_i \right\}^2 dx \qquad (12)$$

$$= k_1 \sum_i (\Delta^2 \theta_i)^2 + k_2,$$

by

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

In general, $\boxed{\text{approximate } \int\limits_0^1 \left(f^{(d)} \right)^2 dx \text{ with } ||D_d\theta||^2}$

PS-ANOVA Penalty

Estimate B-spline coefficients by minimizing

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

where

$$P = \text{blockdiag}\left(0, \ P_1, \ P_2, \ P_{12}\right), \tag{13}$$

$$P_i = \lambda_i D'_{d_i} D_{d_i}$$
$$P_{12} = \lambda_3 D'_{d_1} D_{d_1} \otimes I_{c_m} + \lambda_4 I_{c_l} \otimes D'_{d_2} D_{d_2}$$

Mixed model representation

In matrix notation, Model 8 is given by

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

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

$$B = \left[\; 1_p \mid B_l \mid B_m \mid B_{lm} \;\right] \tag{14}$$

$$B_{lm}\left(B_m \otimes 1'_{c_l}\right) \odot \left(1'_{c_m} \otimes B_l\right),$$

and

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

Mixed model representation

Find transformation M to map

$$B \to \left[\; X \mid Z \; \right], \qquad \theta \to (\beta, \alpha)'$$

such that

$$BM = \left[\; X \mid Z \; \right], \qquad B\theta = X\beta + Z\alpha$$

Applying M to basis 14 and penalty 13, Model 8 becomes

$$Y = W\left(X\beta + Z\alpha\right) + e,$$
$$\alpha \sim \mathcal{N}\left(0, G\right), \quad e \sim \mathcal{N}\left(0, D\right) \tag{15}$$

The $\{\lambda_j\}$ become

$$\lambda_j = \frac{\sigma^2}{\tau_j^2}$$

where $\mathrm{diag}\left(G\right) = \left\{\tau_j^2\right\}$.

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}:

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

For $d_l = d_m = 2$,

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

with basis:

$$B = \left[\ 1_p \mid B_1 \mid B_2 \mid B_3 \mid B_4 \mid B_5 \ \right], \tag{16}$$

where

$$B_3 = \ m \ \Box \ B_1 \qquad\qquad B_5 = \ B_2 \ \Box \ B_1$$
$$B_4 = \ B_2 \ \Box \ l$$

Nested PS-ANOVA