Nonparametric covariance estimation for longitudinal data via tensor product smoothing

Tayler A. Blake

June 28, 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.)

Figure: Regulate like Nate Dogg.

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.

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.

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.

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.

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:

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

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:

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

In matrix form:

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

and taking covariances on both sides:

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

No constraints on the ϕ_{jk}s!

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:

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

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

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

(Iterated) penalized maximum likelihood estimation

1. Fix $\sigma^2_{ij} = \sigma^2_{ij0}$, $i = 1, \ldots, N$,$j = 1, \ldots, M$.
2. Find $\phi_0 = \underset{\phi}{arg\, min} - 2L_\phi\left(\phi, y_1, \ldots, y_N\right) + \lambda J\left(\phi\right)$
3. Fix $\phi = \phi_0$.
4. Find $\sigma^2_0 = \underset{\sigma^2}{arg\, min} - 2L^2_\sigma\left(\sigma^2, y_1, \ldots, y_N\right) + \lambda J\left(\sigma^2\right)$

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

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

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

so Model 5 becomes

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

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

Represent the main effects as

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

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

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

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

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:

$$B = \left[\ 1_p \mid B_l \mid B_m \mid B_{lm}\ \right] \tag{11}$$

where

$$B_{lm}\left(B_m \otimes 1'_{c_l}\right) \odot \left(1'_{c_m} \otimes B_l\right). \tag{12}$$