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

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

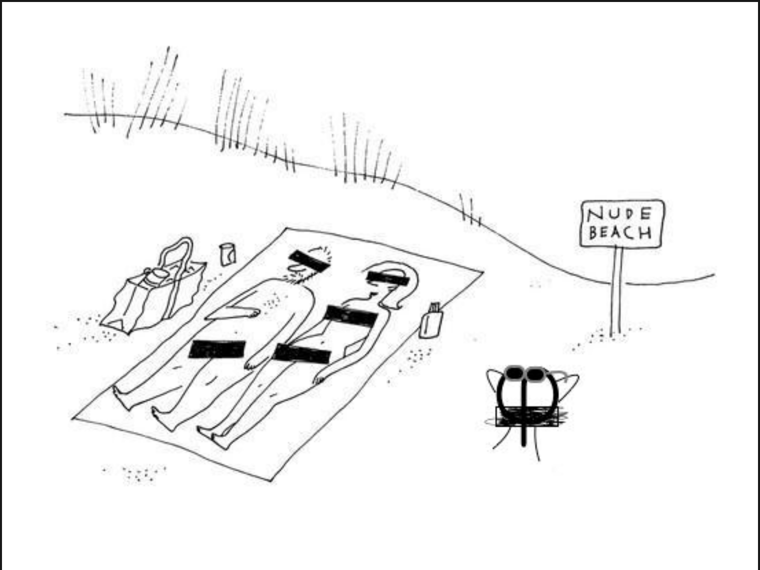
where $e_j \sim N\left(0, \sigma^2\right)$. In matrix form:

$$e = TY$$

Taking covariances on both sides:

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

No constraints on the ϕ_{jk}s!



Accompanying the regression model is a deep, luxurious toolbox.

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

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

where

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

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

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

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

The negative log likelihood can be written

$$\begin{aligned} -2L &= \sum_{i=1}^{n}\sum_{j=2}^{m_i} \sigma_{ij}^{-2}\left(y_{ij} - \sum_{k=1}^{j-1}\phi(t_{ij}, t_{ik})\, y_{ik}\right)^2 \\ &= \sum_{i=1}^{n}\sum_{j=2}^{m_i} \sigma_{ij}^{-2}\left(y_{ij} - \sum_{k=1}^{j-1}\phi^*(l_{ijk}, m_{ijk})\, y_{ik}\right)^2 \end{aligned}$$

Smooth via a tensor product B-spline basis

Equip l and m with

$$B_1(l), \ldots, B_K(l),$$
$$B_1(m), \ldots, B_L(m)$$

to build

$$T_{jk}(l, m) = B_j(l)\, B_k(m)$$

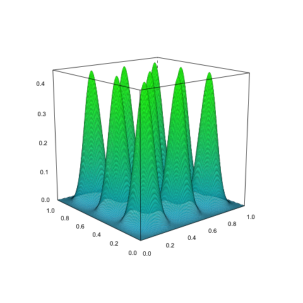


Figure: A "thinned" tensor product basis

$$\phi^*(l, m) = \sum_{i=1}^{K} \sum_{j=1}^{L} \theta_{ij} B_i(l)\, B_j(m)$$

Recruit the difference penalty to regulate.

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

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

by

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

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

Append the difference penalties to the log likelihood:

Find θ minimizing

$$\mathcal{Q} = \sum_{i=1}^{n} \sum_{j=2}^{m_i} \sigma_{ij}^{-2} \left[y_{ij} - \sum_{k=1}^{j-1} \left(\sum_{i=1}^{K} \sum_{j=1}^{L} \theta_{ij} B_i \left(l_{ijk} \right) B_j \left(m_{ijk} \right) \right) y_{ik} \right]^2 + \lambda_l \sum_k \left| D_d \theta_{k\cdot} \right|^2 + \lambda_m \sum_l \left| D_{d_m} \theta_{\cdot l} \right|^2$$

Differencing in two dimensions

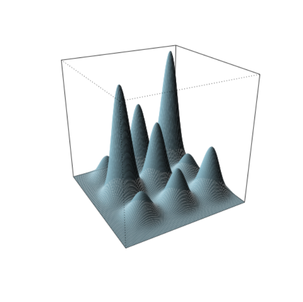


Figure: Strong row penalty

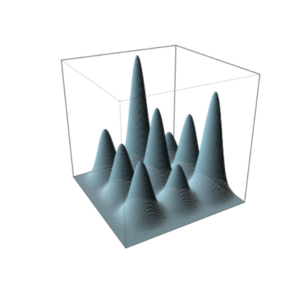


Figure: Strong column penalty

"Unfold" the two dimensional surface:

via the *kronecker product*:

$$
\begin{aligned}
B &= B_l \Box B_m \\
&= (B_m \otimes 1'_K) \odot (1'_L \otimes B_l) \\
P_l &= I_L \otimes D'_{d_l} D_{d_l} \\
P_m &= D'_{d_m} D_{d_m} \otimes I_K
\end{aligned}
$$

to write

$$
\mathrm{vec}\left\{\phi^*\left(l, m\right)\right\} = B\theta
$$

Then the penalized likelihood has convenient form:

$$Q = (Y - WB\theta)' \, D^{-1} \, (Y - WB\theta) + \lambda_l \theta' P_l \theta + \lambda_m \theta' P_m \theta$$

$$\hat{\theta}_\lambda = \left[(WB)' \, D^{-1} WB + \lambda_l P_l + \lambda_m P_m \right]^{-1} (WB)' \, Y$$

Degrees of freedom can be approximated as in the usual smoothing case:

$$\begin{aligned} \mathrm{ED} &= \mathrm{tr}\left[H_\lambda \right] \\ &= \mathrm{tr}\left[\left[(WB)' \, D^{-1} WB + \lambda_l P_l + \lambda_m P_m \right]^{-1} (WB)' \, WB \right] \end{aligned}$$

AIC, GCV, CV accessible for model diagnostics.

Simulations: $\Sigma = 0.3^2 I$

$N = 30,\ M = 20,\ d_l = d_m = 0$

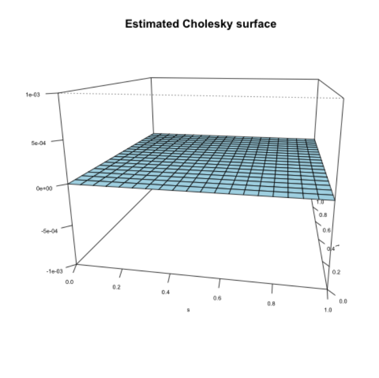


Figure: Estimated T

Simulations: $\Sigma = 0.3^2 I$
$N = 30,\ M = 20,\ d_l = d_m = 0$

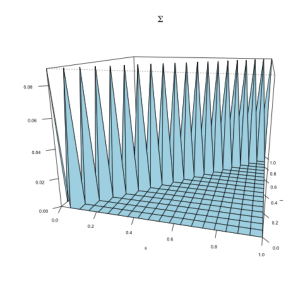


Figure: True Σ

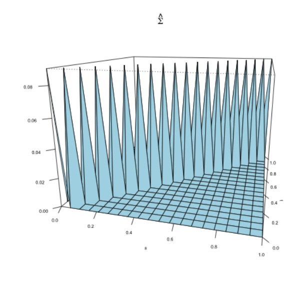


Figure: $\hat{\Omega}^{-1}$

Simulations: $\phi(s,t) = s - \frac{1}{2}$, $\sigma^2 = 0.3^2$ $N = 30$, $M = 20$, $d_l = d_m = 2$

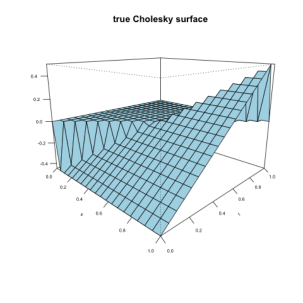


Figure: True T

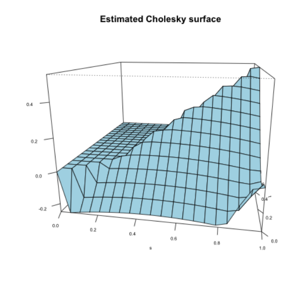


Figure: Estimated T

Simulations: $\phi(s,t) = s - \frac{1}{2}$, $\sigma^2 = 0.3^2$ $N = 30$, $M = 20$, $d_l = d_m = 2$

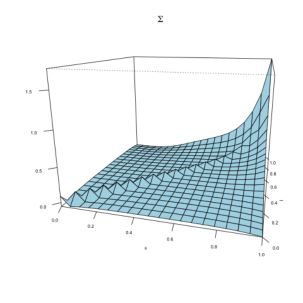


Figure: Σ

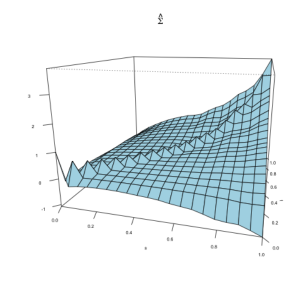


Figure: $\hat{\Omega}^{-1}$

Simulations: $\Sigma = 0.7J + 0.3I$

$N = 30,\ M = 20,\ d_l = 2,\ d_m = 1$

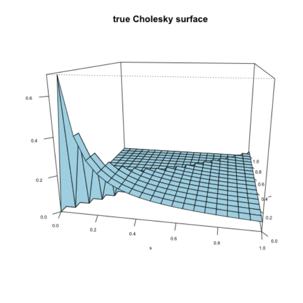


Figure: True T

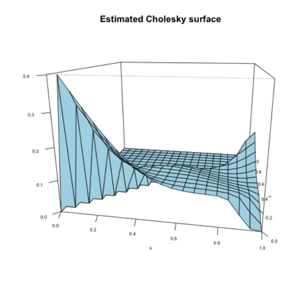


Figure: Estimated T

Simulations: $\Sigma = 0.7J + 0.3I$
$N = 30,\ M = 20,\ d_l = 2,\ d_m = 1$

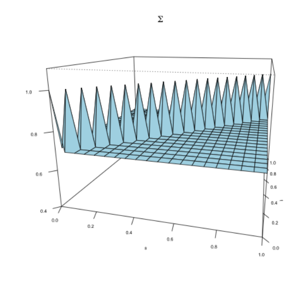


Figure: Σ

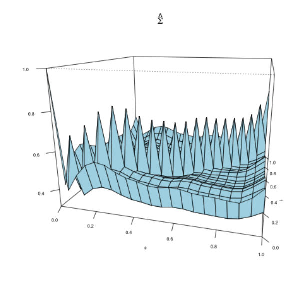


Figure: $\hat{\Omega}^{-1}$

Recap: from the top

- We propose a general framework for unconstrained covariance estimation.
- Flexibility permits imposing various types of regularization with ease.
- Penalty specification is crucial for performance.

What's next?

- "designer" penalties - impose desirable shape constraints (decay in l)
- Additive, ANOVA models for ϕ^*
- P-spline model reparameterized as mixture models

Thank you!