

Bivariate Thin-plate Splines Models for Nonparametric Covariance Estimation with Longitudinal Data

Tayler A. Blake*

Yoonkyung Lee†

September 15, 2017

The theoretical foundations of the thin-plate spline was laid in the seminal work of ?. For a bivariate function $f(x_1, x_2)$, the usual thin-plate spline functional ($d = m = 2$) is given by

$$J_2(f) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (f_{x_1 x_1}^2 + f_{x_1 x_2}^2 + f_{x_2 x_2}^2) dx_1 dx_2 \quad (1)$$

and in general,

For $d = 2$, define the inner product of functions f and g as follows:

$$\langle f, g \rangle = \sum_{\alpha_1 + \alpha_2 = m} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{\partial^m f}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2}} \right) \left(\frac{\partial^m g}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2}} \right) dx_1 dx_2. \quad (2)$$

We suppose that $f \in \mathcal{X}$, the space of functions with partial derivatives of total order m belong to $\mathcal{L}_2(E^2)$. We endow \mathcal{X} with seminorm $J_m^2(f)$; for such \mathcal{X} to be a reproducing kernel Hilbert space, i.e. for the evaluation functionals to be bounded in \mathcal{X} , if it necessary and sufficient that $2m > d$. For $d = 2$, we require $m > 1$.

The data model for a random vector $y_i = (y_{i1}, \dots, y_{i, M_i})'$ is given by

$$y_{ij} = \sum_{k < j} \phi^*(v_{ijk}) y_{ik} + \sigma(v_{ijk}) e_{ij} \quad (3)$$

where $v_{ijk} = (t_{ij} - t_{ik}, \frac{1}{2}(t_{ij} + t_{ik})) = (l_{ijk}, m_{ijk})$. We assume that $\phi^* \in \mathcal{X}$ and $e_{ij} \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$. If we have a random sample of observed vectors y_1, \dots, y_N available for estimating ϕ^* , then we take ϕ^* to be the minimizer of

$$Q_\lambda(\phi^*) = \sum_{i=1}^N \sum_{j=2}^{n_i} \sigma_{ij}^{-2} \left(y_{ij} - \sum_{k < j} \phi^*(v_{ijk}) y_{ik} \right)^2 + \lambda J_m^2(\phi^*) \quad (4)$$

*The Ohio State University, 1958 Neil Avenue, Columbus, OH 43201

†The Ohio State University, 1958 Neil Avenue, Columbus, OH 43201

where $\sigma_{ij}^2 = \sigma^2(t_{ij})$. The null space of the penalty functional $J_m^2(\phi^*)$, denoted \mathcal{H}_0 , corresponds to the $d_0 = \binom{2+m-1}{2}$ -dimensional space spanned by the polynomials in two variables of total degree $< m$. For example, for $d = m = 2$, we have that $d_0 = 3$, and the null space of J_2^2 is spanned by η_1, η_2 , and η_3 where

$$\eta_1(v) = 1, \quad \eta_2(v) = l, \quad \eta_3(v) = m.$$

In general, we let $\eta_1, \dots, \eta_{d_0}$ denote the d_0 monomials of total degree less than m .

? showed that if the $\{v_{ijk}\}$ are such that the least squares regression of $\{y_{ijk}\}$ on $\eta_1, \dots, \eta_{d_0}$ is unique, then there exists a unique minimizer of 4, ϕ_λ^* , which has the form

$$\phi_\lambda^*(v) = \sum_{\nu=0}^{d_0} d_\nu \eta_\nu(v) + \sum_{v_i \in \mathcal{V}} c_i E_m(v, v_i) \quad (5)$$

where \mathcal{V} denotes the set of unique within-subject pairs of observed $\{v_{ijk}\}$. E_m is a Green's function of the m -iterated Laplacian. Let

$$E_m(\tau) = \begin{cases} \theta_{m,d} |\tau|^{2m-d} \log |\tau| & 2m-d \text{ even} \\ \theta_{m,d} |\tau|^{2m-d} & 2m-d \text{ odd} \end{cases} \quad (6)$$

$$\theta_{md} = \begin{cases} \frac{(-1)^{\frac{d}{2}+1+m}}{2^{2m-1} \pi^{\frac{d}{2}} (m-1)! (m-\frac{d}{2})!} & 2m-d \text{ even} \\ \frac{\Gamma(\frac{d}{2}-m)}{2^{2m} \pi^{\frac{d}{2}} (m-1)!} & 2m-d \text{ odd} \end{cases} \quad (7)$$

Defining $|v - v_i| = \left[(l - l_i)^2 + (m - m_i)^2 \right]^{1/2}$, then we can write

$$E_m(v, \tilde{v}) = E_m(|v - \tilde{v}|)$$

Formally, we have that

$$\Delta^m E_m(\cdot, v_i) = \delta_{v_i},$$

so

$$\Delta^m \phi_\lambda^*(v) = 0 \text{ for } v \neq v_i, \quad i = 1, \dots, n$$

where $n = |\mathcal{V}|$.

The kernel E_m is not positive definite, but rather *conditionally positive definite*....

Stack the N observed response vectors y_1, \dots, y_N less their first element y_{i1} into a single vector Y of dimension $n_y = \left(\sum_i M_i \right) - N$. Let B denote the $n \times d_0$ matrix with i - ν^{th} element $\eta_\nu(v_i)$, which we assume has full column rank; let K denote the $n \times n$ kernel matrix with i - j^{th} element

$E_m(v_i, v_j)$, and let D denote the $n_y \times n_y$ diagonal matrix of innovation variances σ_{ijk}^2 . The ϕ^* minimizing 4 corresponds to the coefficient vectors c, d minimizing

$$\begin{aligned} Q_\lambda(c, d) &= -\ell(Y|c, d) + \lambda J_m^2(\phi^*) \\ &= (Y - W(Bd + Kc))' D^{-1} (Y - W(Bd + Kc)) + \lambda c' Kc \end{aligned} \quad (8)$$

where W is the matrix of autoregressive covariates constructed so that 4 and 8 are equivalent.

Differentiating Q_λ with respect to c and d and setting equal to zero, we have that

$$\begin{aligned} \frac{\partial Q_\lambda}{\partial c} &= KW'D^{-1} [W(Bd + Kc) - Y] + \lambda Kc = 0 \\ \iff W'D^{-1}W [Bd + Kc] + \lambda c &= W'D^{-1}Y \end{aligned} \quad (9)$$

$$\begin{aligned} \frac{\partial Q_\lambda}{\partial d} &= B'W'D^{-1} [W(Bd + Kc) - Y] = 0 \\ \iff -\lambda B'c &= 0 \end{aligned}$$

So, the coefficients satisfy the normal equations

$$Y = W \left[Bd + \left(K + \lambda (W'D^{-1}W)^{-1} \right) c \right] \quad (10)$$

$$B'c = 0 \quad (11)$$

Let

$$\begin{aligned} \tilde{K} &= (W'D^{-1}W) K (W'D^{-1}W) \\ \tilde{c} &= (W'D^{-1}W)^{-1} c \\ \tilde{B} &= (W'D^{-1}W) B \\ \tilde{d} &= d \\ \tilde{Y} &= W'D^{-1}Y \end{aligned}$$

then, the system defined by 10 and 11 may be written

$$\tilde{Y} = \tilde{B}\tilde{d} + \left(\tilde{K} + \lambda (W' D^{-1} W) \right) \tilde{c} \quad (12)$$

$$\tilde{B}'\tilde{c} = 0 \quad (13)$$

Using the QR decomposition of \tilde{B} , we may write

$$\tilde{B} = \tilde{Q}\tilde{R} = [\tilde{Q}_1 \quad \tilde{Q}_2] \begin{bmatrix} \tilde{R} \\ 0 \end{bmatrix} = \tilde{Q}_1\tilde{R}$$

where \tilde{Q} is an orthogonal matrix; \tilde{Q}_1 has dimension $n \times d_0$, and \tilde{Q}_2 has dimension $n \times (n - d_0)$. Since $\tilde{B}'\tilde{c} = 0$, \tilde{c} must belong to the subspace spanned by the columns of \tilde{Q}_2 , so

$$\tilde{c} = \tilde{Q}_2\gamma$$

for some $\gamma \in \mathbb{R}^{n-d_0}$. Premultiplying 12 by \tilde{Q}_2' , it follows that

$$\tilde{c} = \tilde{Q}_2 \left[\tilde{Q}_2' \left(\tilde{K} + \lambda (W' D^{-1} W) \right) \tilde{Q}_2 \right]^{-1} \tilde{Q}_2' \tilde{Y} \quad (14)$$

Using $\tilde{B} = \tilde{Q}_1\tilde{R}$, we can write

$$\tilde{d} = \tilde{R}^{-1}\tilde{Q}_1' \left[\tilde{Y} - \left(\tilde{K} + \lambda (W' D^{-1} W) \right) \tilde{c} \right] \quad (15)$$

1 Estimating the smoothing parameter

1.1 Cross Validation

Let $\phi^{*[kl]}$ be the minimizer of

$$\sum_{\substack{i,j \\ (i,j) \neq (k,l)}} \sigma_{ij}^{-2} \left(y_{ij} - \sum_{j' < j} \phi^* (v_{ijj'}) y_{ij'} \right)^2 + \lambda J_m^2(\phi^*). \quad (16)$$

The *ordinary cross validation function* $V_0(\lambda)$ is given by

$$\sum_{i=1}^N \sum_{j=2}^{n_i} \sigma_{ij}^{-2} \left(y_{ij} - \hat{y}_{ij}^{[ij]} \right)^2 \quad (17)$$

where $\hat{y}_{ij}^{[ij]} = \sum_{k < j} \phi^{[ij]*} (v_{ijk}) y_{ik}$. The value of λ minimizing $V_0(\lambda)$ is the OCV estimate.

Indexing the y_{ij} using a single integer $k = 1, \dots, n_y$, when the innovation variances are known, it can be shown that $V_0(\lambda)$ can be written

$$V_0(\lambda) = \sum_{k=1}^{n_y} (\sigma_k^{-1} (y_k - \hat{y}_k))^2 / (1 - a_{kk}(\lambda))^2 \quad (18)$$

where $\{a_{kk}(\lambda)\}$ are the diagonal elements of the smoothing matrix $A(\lambda)$ which satisfies

$$\hat{Y} = A(\lambda) Y.$$

The *generalized cross validation function* $V(\lambda)$ is obtained by replacing a_{kk} by

$$\bar{a}(\lambda) = n_y^{-1} \sum_{j=1}^{n_y} a_{jj}(\lambda) = n_y^{-1} \text{tr} A(\lambda).$$

The GCV function is defined as

$$\begin{aligned} V(\lambda) &= \sum_{k=1}^{n_y} (\sigma_k^{-1} (y_k - \hat{y}_k))^2 / (1 - \bar{a}(\lambda))^2 \\ &= \frac{\|D^{-1/2} (I - A(\lambda))\|^2}{[\text{tr} (I - A(\lambda))]^2} \end{aligned} \quad (19)$$

Since $\tilde{A}(\lambda) \tilde{Y} = \tilde{K} \tilde{c} + \tilde{B} \tilde{d}$, from 12, we can derive a simple expression for $I - \tilde{A}(\lambda)$:

$$\begin{aligned} (I - \tilde{A}(\lambda)) Y &= \lambda (W' D^{-1} W) \tilde{c} \\ &= \lambda (W' D^{-1} W) \tilde{Q}_2 \left[\tilde{Q}_2' (\tilde{K} + \lambda (W' D^{-1} W)) \tilde{Q}_2 \right]^{-1} \tilde{Q}_2' \tilde{Y}, \end{aligned} \quad (20)$$

so

$$I - \tilde{A}(\lambda) = \lambda (W' D^{-1} W) \tilde{Q}_2 \left[\tilde{Q}_2' (\tilde{K} + \lambda (W' D^{-1} W)) \tilde{Q}_2 \right]^{-1} \tilde{Q}_2'.$$

Then the GCV criterion can be written

$$V(\lambda) = \frac{n_y^{-1} \tilde{Y}' \tilde{Q}_2' \left[\tilde{Q}_2' (\tilde{K} + \lambda M) \tilde{Q}_2 \right]^{-1} \tilde{Q}_2' M^2 \tilde{Q}_2 \left[\tilde{Q}_2' (\tilde{K} + \lambda M) \tilde{Q}_2 \right]^{-1} \tilde{Q}_2' \tilde{Y}}{\left[n_y^{-1} \text{tr} M \tilde{Q}_2 \left[\tilde{Q}_2' (\tilde{K} + \lambda M) \tilde{Q}_2 \right]^{-1} \tilde{Q}_2' \right]^2} \quad (21)$$

where $M = W' D^{-1} W$.

1.2 Unbiased Risk Estimate

$$M(\lambda) = \frac{(D^{-1/2}Y)'(I - A(\lambda))(D^{-1/2}Y)}{[\det^+(I - A(\lambda))]^{1/(n-d_0)}}$$

where $\det^+(\cdot)$ denotes the product of the non-zero eigenvalues.

1.3 Generalized Maximum Likelihood

$$U(\lambda) = n_y^{-1} \|(I - A(\lambda))D^{-1/2}Y\|^2 + 2\text{tr}A(\lambda)$$