Technical Supplement for 'Spatial kriging for replicated temporal point processes'

Daniel Gervini
Department of Mathematical Sciences
University of Wisconsin–Milwaukee

June 30, 2021

Contents

1 Mean and covariance estimation

1.1 Nonparametric estimators

The nonparametric mean and covariance estimators are derived from Campbell's formula (Baddeley, 2007, Theorem 2.2; Møller and Waagepetersen, 2004, Proposition 4.1). Let X be a point process with count function $N(B)$ and define the intensity measure $\nu(B) = E\{N(B)\}$ for every bounded set B. Suppose there exists a non-negative function $\lambda(t)$, called intensity function, such that $\nu(B) = \int_B \lambda(t)dt$. Then Campbell's formula says that for any measurable function $f(t)$,

$$E\left\{\int f\ dN\right\} = \int f\ d\nu,$$

or equivalently,

$$E\left\{\sum_{t\in X} f(t)\right\} = \int f(t)\lambda(t)\ dt.$$

Then, for a doubly stochastic Poisson process (X, Λ), we have

$$\begin{aligned} E\left\{\sum_{t\in X} f(t)\right\} &= E\left[E\left\{\sum_{t\in X} f(t) \mid \Lambda\right\}\right] = E\left[\int f(t)\Lambda(t)\ dt\right] \\ &= \int f(t)E\{\Lambda(t)\}\ dt = \int f(t)\mu(t)\ dt, \end{aligned}$$

where $\mu(t) = E\{\Lambda(t)\}$.

For the second moments, given A and B bounded sets in $\mathbb{R}$, define $N_{[2]}(A\times B) = N(A)N(B) - N(A\cap B)$ and $\nu_{[2]}(A\times B) = E\left\{N_{[2]}(A\times B)\right\}$. If there exists a non-negative $\lambda_{[2]}(t,t')$ such that $\nu_{[2]}(A\times B) = \iint_{A\times B} \lambda_{[2]}(t,t')\ dt\ dt'$, Campbell's formula says that for any measurable function $f(t,t')$ we have

$$E\left\{\iint f\ dN_{[2]}\right\} = \iint f\ d\nu_{[2]},$$

which comes down to

$$E\left\{\sum_{t\in X}\ \sum_{t'\in X, t'\neq t} f(t,t')\right\} = \iint f(t,t')\lambda_{[2]}(t,t')\ dt\ dt'.$$

For the specific case of a Poisson process with intensity $\lambda(t)$, we can write

$$\begin{aligned} N(A)N(B) &= N(A)N(B\cap A^c) + N(A)N(B\cap A) \\ &= N(A)N(B\cap A^c) + N(B\cap A)^2 + N(A\cap B^c)N(B\cap A) \end{aligned}$$

and then

$$\begin{aligned} E\left\{N_{[2]}(A\times B)\right\} &= (\int_A \lambda)(\int_{B\cap A^c} \lambda) + (\int_{B\cap A} \lambda) + (\int_{B\cap A} \lambda)^2 \\ &\quad + (\int_{A\cap B^c} \lambda)(\int_{B\cap A} \lambda) - (\int_{A\cap B} \lambda) \\ &= (\int_A \lambda)(\int_{B\cap A^c} \lambda) + (\int_{B\cap A} \lambda)^2 + (\int_{A\cap B^c} \lambda)(\int_{B\cap A} \lambda) \\ &= (\int_A \lambda)(\int_B \lambda) \\ &= \iint_{A\times B} \lambda_{[2]}(t,t')\ dt\ dt' \end{aligned}$$

with $\lambda_{[2]}(t,t') = \lambda(t)\lambda(t')$, so that

$$E\left\{\sum_{t\in X}\sum_{t'\in X, t'\neq t} f(t,t')\right\} = \iint f(t,t')\lambda(t)\lambda(t')\ dt\ dt'.$$

For a doubly stochastic Poisson process (X,Λ), using iterated expectations as before we get

$$E\left\{\sum_{t\in X}\sum_{t'\in X, t'\neq t} f(t,t')\right\} = \iint f(t,t')R(t,t')\ dt\ dt'$$

with $R(t,t') = E\left\{\Lambda(t)\Lambda(t')\right\}$.

For two independent Poisson processes X_1 and X_2 with respective intensity functions $\lambda_1(t)$ and $\lambda_2(t)$, the respective counts $N_1(A)$ and $N_2(B)$ are independent regardless of A and B, therefore it is not necessary to "subtract the intersection"; we can work directly with the product count $N(A\times B) = N_1(A)N_2(B)$ and the product measure $\nu(A\times B) = E\{N_1(A)N_2(B)\} = E\{N_1(A)\}E\{N_2(B)\}$. Campbell's formula

$$E\left\{\iint f\ dN\right\} = \iint f\ d\nu$$

then comes down to

$$E\left\{\sum_{t\in X_1}\sum_{t'\in X_2} f(t,t')\right\} = \iint f(t,t')\lambda_1(t)\lambda_2(t')\ dt\ dt'.$$

For a multivariate doubly stochastic Poisson process $(X_1, X_2, \Lambda_1, \Lambda_2)$ we then have, using iterated expectations, that

$$E\left\{\sum_{t\in X_1}\sum_{t'\in X_2} f(t,t')\right\} = \iint f(t,t')R(t,t')\ dt\ dt',$$

with $R(t,t') = E\left\{\Lambda_1(t)\Lambda_2(t')\right\}$.

1.2 Penalty functions and closed forms

The estimator of $\mathbf{B}$ is defined as

$$\hat{\mathbf{B}} = \arg\min_{\mathbf{B}} \sum_{j=1}^{d} \|\hat{\mathbf{a}}_j - \mathbf{B}\boldsymbol{\gamma}(\mathbf{s}_j)\|^2 + \xi_B P_1(\mathbf{B}),$$

where $P_1(\mathbf{B})$ is a roughness penalty function and ξ_B is a smoothing parameter. We assume the function $\mathbf{f}(\mathbf{s}) = \mathbf{B}\boldsymbol{\gamma}(\mathbf{s})$ is smooth in $\mathbf{s}$. As roughness measure of a bivariate function $f(s^1, s^2)$ we take $\iint(\sum_{1\le i,j\le 2} f_{ij}^2)ds^1 ds^2$, where $f_{ij} = \partial^2 f/\partial s^i \partial s^j$, so $P_1(\mathbf{B}) = \sum_{k=1}^{p} \iint(\sum_{1\le i,j\le 2} f_{ij}^{(k)2})ds^1 ds^2$. Since $f^{(k)}(\mathbf{s}) = \mathbf{b}_k^T\boldsymbol{\gamma}(\mathbf{s})$, where $\mathbf{b}_k^T$ is the k-th row of $\mathbf{B}$, we have $f_{ij}^{(k)}(\mathbf{s}) = \mathbf{b}_k^T\boldsymbol{\gamma}_{ij}(\mathbf{s})$, where the derivatives in $\boldsymbol{\gamma}_{ij}(\mathbf{s})$ are taken in a componentwise manner. Then

$$\begin{aligned}\sum_{1\le i,j\le 2} \{f_{ij}^{(k)}(\mathbf{s})\}^2 &= \sum_{1\le i,j\le 2} \mathbf{b}_k^T\boldsymbol{\gamma}_{ij}(\mathbf{s})\boldsymbol{\gamma}_{ij}(\mathbf{s})^T\mathbf{b}_k \\ &= \mathbf{b}_k^T\left\{\boldsymbol{\gamma}_{11}(\mathbf{s})\boldsymbol{\gamma}_{11}(\mathbf{s})^T + 2\boldsymbol{\gamma}_{12}(\mathbf{s})\boldsymbol{\gamma}_{12}(\mathbf{s})^T + \boldsymbol{\gamma}_{22}(\mathbf{s})\boldsymbol{\gamma}_{22}(\mathbf{s})^T\right\}\mathbf{b}_k,\end{aligned}$$

so, if we define

$$\mathbf{J} = \iint \left\{\boldsymbol{\gamma}_{11}(\mathbf{s})\boldsymbol{\gamma}_{11}(\mathbf{s})^T + 2\boldsymbol{\gamma}_{12}(\mathbf{s})\boldsymbol{\gamma}_{12}(\mathbf{s})^T + \boldsymbol{\gamma}_{22}(\mathbf{s})\boldsymbol{\gamma}_{22}(\mathbf{s})^T\right\}\, d\mathbf{s},$$

we have

$$\begin{aligned}P_1(\mathbf{B}) &= \sum_{k=1}^{p} \mathbf{b}_k^T\mathbf{J}\mathbf{b}_k \\ &= \operatorname{tr}\left(\sum_{k=1}^{p} \mathbf{b}_k\mathbf{b}_k^T\mathbf{J}\right) \\ &= \operatorname{tr}\left(\mathbf{B}^T\mathbf{B}\mathbf{J}\right).\end{aligned}$$

Note that we can write

$$\operatorname{tr}\left(\mathbf{B}^T\mathbf{B}\mathbf{J}\right) = \operatorname{vec}(\mathbf{B})^T(\mathbf{J}\otimes\mathbf{I}_p)\operatorname{vec}(\mathbf{B}).$$

Denoting by $\|\cdot\|_F$ the Frobenius matrix norm, we can write

$$\sum_{j=1}^{d} \|\hat{\mathbf{a}}_j - \mathbf{B}\boldsymbol{\gamma}(\mathbf{s}_j)\|^2 =$$

$$\begin{aligned}
&= \left\|\mathbf{A} - \mathbf{B}\boldsymbol{\Gamma}^T\right\|_F^2 \\
&= \operatorname{tr}\left\{(\mathbf{A} - \mathbf{B}\boldsymbol{\Gamma}^T)^T(\mathbf{A} - \mathbf{B}\boldsymbol{\Gamma}^T)\right\} \\
&= \operatorname{tr}\left(\mathbf{A}^T\mathbf{A}\right) - \operatorname{tr}\left(\boldsymbol{\Gamma}\mathbf{B}^T\mathbf{A}\right) - \operatorname{tr}\left(\mathbf{A}^T\mathbf{B}\boldsymbol{\Gamma}^T\right) + \operatorname{tr}\left(\boldsymbol{\Gamma}\mathbf{B}^T\mathbf{B}\boldsymbol{\Gamma}^T\right) \\
&= \operatorname{tr}\left(\mathbf{A}^T\mathbf{A}\right) - 2\operatorname{vec}(\mathbf{B})^T\operatorname{vec}(\mathbf{A}\boldsymbol{\Gamma}) + \operatorname{vec}(\mathbf{B})^T(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma} \otimes \mathbf{I}_p)\operatorname{vec}(\mathbf{B}).
\end{aligned}$$

Then

$$\begin{aligned}
\hat{\mathbf{B}} &= \arg\min_{\mathbf{B}}[-2\operatorname{vec}(\mathbf{B})^T\operatorname{vec}(\mathbf{A}\boldsymbol{\Gamma}) \\
&\quad + \operatorname{vec}(\mathbf{B})^T\left\{(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma} \otimes \mathbf{I}_p) + \xi_B(\mathbf{J} \otimes \mathbf{I}_p)\right\}\operatorname{vec}(\mathbf{B})],
\end{aligned}$$

which is given in closed form by

$$\begin{aligned}
\operatorname{vec}(\hat{\mathbf{B}}) &= \left\{(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}+\xi_B\mathbf{J}) \otimes \mathbf{I}_p\right\}^{-1}\operatorname{vec}(\mathbf{A}\boldsymbol{\Gamma}) \\
&= \left\{(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}+\xi_B\mathbf{J})^{-1} \otimes \mathbf{I}_p\right\}\operatorname{vec}(\mathbf{A}\boldsymbol{\Gamma}) \\
&= \operatorname{vec}\left\{\mathbf{A}\boldsymbol{\Gamma}(\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}+\xi_B\mathbf{J})^{-1}\right\},
\end{aligned}$$

so $\hat{\mathbf{B}}$ is as claimed.

The estimator of $\mathbf{C}$ is defined as

$$\hat{\mathbf{C}} = \arg\min_{\mathbf{C}} \sum_{j=1}^{d} \sum_{\substack{k=1 \\ k \neq j}}^{d} \left\{\hat{\Sigma}_{jk} - \boldsymbol{\gamma}(\mathbf{s}_j)^T\mathbf{C}\boldsymbol{\gamma}(\mathbf{s}_k)\right\}^2 + \xi_C P_2(\mathbf{C}),$$

where, as before, $P_2(\mathbf{C})$ is a roughness penalty function and ξ_C is a smoothing parameter. We take as roughness measure of a function $f(s^1, s^2, s^3, s^4)$ the usual $\iint(\sum_{1\leq i,j,k,l\leq 2} f_{ijkl}^2)ds^1ds^2ds^3ds^4$. Then for $f(\mathbf{s}, \mathbf{s}') = \boldsymbol{\gamma}(\mathbf{s})^T\mathbf{C}\boldsymbol{\gamma}(\mathbf{s}')$ we have $f_{ijkl}(\mathbf{s}, \mathbf{s}') =$

$\boldsymbol{\gamma}_{ij}(\mathbf{s})^T\mathbf{C}\boldsymbol{\gamma}_{kl}(\mathbf{s}')$, so

$$\begin{aligned}
\sum_{1\le i,j,k,l\le 2} f_{ijkl}^2 &= \sum_{1\le i,j,k,l\le 2} \boldsymbol{\gamma}_{ij}(\mathbf{s})^T\mathbf{C}\boldsymbol{\gamma}_{kl}(\mathbf{s}')\boldsymbol{\gamma}_{kl}(\mathbf{s}')^T\mathbf{C}^T\boldsymbol{\gamma}_{ij}(\mathbf{s}) \\
&= \sum_{1\le i,j,k,l\le 2} \operatorname{tr}\left\{\mathbf{C}\boldsymbol{\gamma}_{kl}(\mathbf{s}')\boldsymbol{\gamma}_{kl}(\mathbf{s}')^T\mathbf{C}^T\boldsymbol{\gamma}_{ij}(\mathbf{s})\boldsymbol{\gamma}_{ij}(\mathbf{s})^T\right\} \\
&= \operatorname{tr}\left\{\mathbf{C}\sum_{1\le k,l\le 2}\boldsymbol{\gamma}_{kl}(\mathbf{s}')\boldsymbol{\gamma}_{kl}(\mathbf{s}')^T\mathbf{C}^T\sum_{1\le i,j\le 2}\boldsymbol{\gamma}_{ij}(\mathbf{s})\boldsymbol{\gamma}_{ij}(\mathbf{s})^T\right\}.
\end{aligned}$$

Then

$$\begin{aligned}
P_2(\mathbf{C}) &= \iint\Big(\sum_{1\le i,j,k,l\le 2} f_{ijkl}^2\Big)\, d\mathbf{s}\, d\mathbf{s}' \\
&= \operatorname{tr}\left(\mathbf{C}\mathbf{J}\mathbf{C}^T\mathbf{J}\right) \\
&= \operatorname{vec}(\mathbf{C})^T(\mathbf{J}\otimes\mathbf{J})\operatorname{vec}(\mathbf{C})
\end{aligned}$$

with $\mathbf{J}$ as before. On the other hand we can write

$$\sum_{j=1}^{d}\sum_{\substack{k=1\\k\neq j}}^{d}\left\{\hat{\Sigma}_{jk}-\boldsymbol{\gamma}(\mathbf{s}_j)^T\mathbf{C}\boldsymbol{\gamma}(\mathbf{s}_k)\right\}^2 =$$

$$\begin{aligned}
&= \|\hat{\boldsymbol{\Sigma}}-\boldsymbol{\Gamma}\mathbf{C}\boldsymbol{\Gamma}^T\|_F^2-\|\mathrm{d}(\hat{\boldsymbol{\Sigma}})-\mathrm{d}(\boldsymbol{\Gamma}\mathbf{C}\boldsymbol{\Gamma}^T)\|^2 \\
&= \|\operatorname{vec}(\hat{\boldsymbol{\Sigma}})-\operatorname{vec}(\boldsymbol{\Gamma}\mathbf{C}\boldsymbol{\Gamma}^T)\|^2-\|\mathrm{d}(\hat{\boldsymbol{\Sigma}})-\mathrm{d}(\boldsymbol{\Gamma}\mathbf{C}\boldsymbol{\Gamma}^T)\|^2 \\
&= \|\operatorname{vec}(\hat{\boldsymbol{\Sigma}})-(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})\operatorname{vec}(\mathbf{C})\|^2-\|\mathrm{d}(\hat{\boldsymbol{\Sigma}})-\mathrm{d}(\boldsymbol{\Gamma}\mathbf{C}\boldsymbol{\Gamma}^T)\|^2,
\end{aligned}$$

where $\mathrm{d}(\cdot)$ denotes the diagonal vector. The vector $\mathrm{d}(\boldsymbol{\Gamma}\mathbf{C}\boldsymbol{\Gamma}^T)$ has elements

$$\begin{aligned}
\boldsymbol{\gamma}(\mathbf{s}_j)^T\mathbf{C}\boldsymbol{\gamma}(\mathbf{s}_j) &= (\boldsymbol{\gamma}(\mathbf{s}_j)^T\otimes\boldsymbol{\gamma}(\mathbf{s}_j)^T)\operatorname{vec}(\mathbf{C}) \\
&= (\mathbf{e}_j^T\otimes\mathbf{e}_j^T)(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})\operatorname{vec}(\mathbf{C}),
\end{aligned}$$

where $\mathbf{e}_j$ is the j-th canonical vector in $\mathbb{R}^d$. Then

$$\mathrm{d}(\boldsymbol{\Gamma}\mathbf{C}\boldsymbol{\Gamma}^T)=\mathbf{E}(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})\operatorname{vec}(\mathbf{C})$$

with $\mathbf{E}$ the matrix with rows $\mathbf{e}_j^T \otimes \mathbf{e}_j^T$. Therefore

$$\begin{aligned}\|\operatorname{vec}(\hat{\mathbf{\Sigma}}) - (\mathbf{\Gamma}\otimes\mathbf{\Gamma})\operatorname{vec}(\mathbf{C})\|^2 &= \|\operatorname{vec}(\hat{\mathbf{\Sigma}})\|^2 - 2\operatorname{vec}(\hat{\mathbf{\Sigma}})^T(\mathbf{\Gamma}\otimes\mathbf{\Gamma})\operatorname{vec}(\mathbf{C}) \\ &\quad + \operatorname{vec}(\mathbf{C})^T(\mathbf{\Gamma}\otimes\mathbf{\Gamma})^T(\mathbf{\Gamma}\otimes\mathbf{\Gamma})\operatorname{vec}(\mathbf{C})\end{aligned}$$

and

$$\begin{aligned}\|\mathrm{d}(\hat{\mathbf{\Sigma}}) - \mathrm{d}(\mathbf{\Gamma}\mathbf{C}\mathbf{\Gamma}^T)\|^2 &= \|\mathrm{d}(\hat{\mathbf{\Sigma}})\|^2 - 2\mathrm{d}(\hat{\mathbf{\Sigma}})^T\mathbf{E}(\mathbf{\Gamma}\otimes\mathbf{\Gamma})\operatorname{vec}(\mathbf{C}) \\ &\quad + \operatorname{vec}(\mathbf{C})^T(\mathbf{\Gamma}\otimes\mathbf{\Gamma})^T\mathbf{E}^T\mathbf{E}(\mathbf{\Gamma}\otimes\mathbf{\Gamma})\operatorname{vec}(\mathbf{C}).\end{aligned}$$

Then $\hat{\mathbf{C}}$ minimizes the function

$$\begin{aligned}&-2\{\operatorname{vec}(\hat{\mathbf{\Sigma}})^T + \mathrm{d}(\hat{\mathbf{\Sigma}})^T\mathbf{E}\}(\mathbf{\Gamma}\otimes\mathbf{\Gamma})\operatorname{vec}(\mathbf{C}) \\ &+ \operatorname{vec}(\mathbf{C})^T\{(\mathbf{\Gamma}\otimes\mathbf{\Gamma})^T(\mathbf{I}-\mathbf{E}^T\mathbf{E})(\mathbf{\Gamma}\otimes\mathbf{\Gamma}) + \xi_C(\mathbf{J}\otimes\mathbf{J})\}\operatorname{vec}(\mathbf{C})\end{aligned}$$

and the solution has the explicit form

$$\operatorname{vec}(\hat{\mathbf{C}}) = \mathbf{\Omega}^{-1}(\mathbf{\Gamma}^T\otimes\mathbf{\Gamma}^T)\{\operatorname{vec}(\hat{\mathbf{\Sigma}}) + \mathbf{E}^T\mathrm{d}(\hat{\mathbf{\Sigma}})\}$$

with $\mathbf{\Omega} = \big\{(\mathbf{\Gamma}^T\otimes\mathbf{\Gamma}^T)(\mathbf{I}-\mathbf{E}^T\mathbf{E})(\mathbf{\Gamma}\otimes\mathbf{\Gamma}) + \xi_C(\mathbf{J}\otimes\mathbf{J})\big\}$. Note that $\mathbf{E}^T\mathbf{E} = \sum_{j=1}^d \mathbf{e}_j\mathbf{e}_j^T \otimes \mathbf{e}_j\mathbf{e}_j^T$, as claimed, and

$$\begin{aligned}\mathbf{E}^T\mathrm{d}(\hat{\mathbf{\Sigma}}) &= \sum_{j=1}^d(\mathbf{e}_j\otimes\mathbf{e}_j)\hat{\mathbf{\Sigma}}_{jj} \\ &= \operatorname{vec}(\operatorname{diag}\hat{\mathbf{\Sigma}}),\end{aligned}$$

where $\operatorname{diag}\hat{\mathbf{\Sigma}}$ is the diagonal matrix with elements $\hat{\mathbf{\Sigma}}_{jj}$'s on the diagonal. Therefore

$$\operatorname{vec}(\hat{\mathbf{\Sigma}}) + \mathbf{E}^T\mathrm{d}(\hat{\mathbf{\Sigma}}) = \operatorname{vec}(\hat{\mathbf{\Sigma}} - \operatorname{diag}\hat{\mathbf{\Sigma}}),$$

as claimed.

1.3 Computational implementation

The estimator

$$\hat{\mathbf{B}} = \mathbf{A}\mathbf{\Gamma}(\mathbf{\Gamma}^T\mathbf{\Gamma}+\xi_B\mathbf{J})^{-1}$$

can be computed as-is, since $\mathbf{\Gamma}^T\mathbf{\Gamma}+\xi_B\mathbf{J}$ is just a $q\times q$ matrix and q is usually not too large. The same applies to the degrees of freedom $\mathrm{df}_B = \mathrm{tr}\left\{(\mathbf{\Gamma}^T\mathbf{\Gamma}+\xi_B\mathbf{J})^{-1}\mathbf{\Gamma}^T\mathbf{\Gamma}\right\}$.

On the other hand, the estimator

$$\mathrm{vec}(\hat{\mathbf{C}}) = \mathbf{\Omega}^{-1}(\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T)\,\mathrm{vec}(\hat{\mathbf{\Sigma}} - \mathrm{diag}\,\hat{\mathbf{\Sigma}})$$

involves the $q^2\times q^2$ matrix $\mathbf{\Omega}$, as do the degrees of freedom $\mathrm{df}_C = \mathrm{tr}\left\{\mathbf{\Omega}^{-1}(\mathbf{\Gamma}^T\mathbf{\Gamma} \otimes \mathbf{\Gamma}^T\mathbf{\Gamma})\right\}$. These computations have to be done more carefully for speed's sake.

Let $\mathbf{\Gamma} = \mathbf{U}\mathbf{S}\mathbf{V}^T$ be the singular value decomposition of $\mathbf{\Gamma}$, with $\mathbf{U}$ $d \times d$, $\mathbf{S}$ $d \times d$, and $\mathbf{V}$ $q \times d$. Usually $d < q$, so $\mathbf{S}$ is invertible. We have

$$\mathbf{\Omega} = \left\{(\mathbf{VS} \otimes \mathbf{VS})(\mathbf{U}^T \otimes \mathbf{U}^T)(\mathbf{I} - \mathbf{E}^T\mathbf{E})(\mathbf{U} \otimes \mathbf{U})(\mathbf{S}\mathbf{V}^T \otimes \mathbf{S}\mathbf{V}^T) + \xi_C(\mathbf{J} \otimes \mathbf{J})\right\}$$

where

$$(\mathbf{U}^T \otimes \mathbf{U}^T)(\mathbf{I} - \mathbf{E}^T\mathbf{E})(\mathbf{U} \otimes \mathbf{U}) = \mathbf{I}_{d^2} - (\mathbf{U}^T \otimes \mathbf{U}^T)\mathbf{E}^T\mathbf{E}(\mathbf{U} \otimes \mathbf{U})$$

and

$$\begin{aligned}(\mathbf{U}^T \otimes \mathbf{U}^T)\mathbf{E}^T\mathbf{E}(\mathbf{U} \otimes \mathbf{U}) &= \sum_{j=1}^{d} \mathbf{u}_{j\cdot}^T\mathbf{u}_{j\cdot} \otimes \mathbf{u}_{j\cdot}^T\mathbf{u}_{j\cdot} \\ &= \sum_{j=1}^{d}(\mathbf{u}_{j\cdot}^T \otimes \mathbf{u}_{j\cdot}^T)(\mathbf{u}_{j\cdot} \otimes \mathbf{u}_{j\cdot}),\end{aligned}$$

where $\mathbf{u}_{j\cdot}$ denotes the j-th row of $\mathbf{U}$. Since the vectors $\left\{\mathbf{u}_{j\cdot}^T \otimes \mathbf{u}_{j\cdot}^T\right\}$ are orthonormal in $\mathbb{R}^{d^2}$, we have that

$$\sum_{j=1}^{d}(\mathbf{u}_{j\cdot}^T \otimes \mathbf{u}_{j\cdot}^T)(\mathbf{u}_{j\cdot} \otimes \mathbf{u}_{j\cdot}) = \mathbf{P},$$

a projection matrix of rank d. Then

$$\begin{aligned}\mathbf{\Omega} &= \left\{(\mathbf{VS} \otimes \mathbf{VS})(\mathbf{I}_{d^2} - \mathbf{P})(\mathbf{S}\mathbf{V}^T \otimes \mathbf{S}\mathbf{V}^T) + \xi_C(\mathbf{J} \otimes \mathbf{J})\right\} \\ &= (\mathbf{VS} \otimes \mathbf{VS})\left\{\mathbf{I}_{d^2} - \mathbf{P} + \xi_C(\mathbf{S}^{-1}\mathbf{V}^T\mathbf{J}\mathbf{V}\mathbf{S}^{-1} \otimes \mathbf{S}^{-1}\mathbf{V}^T\mathbf{J}\mathbf{V}\mathbf{S}^{-1})\right\}(\mathbf{S}\mathbf{V}^T \otimes \mathbf{S}\mathbf{V}^T).\end{aligned}$$

Now let $\mathbf{S}^{-1}\mathbf{V}^T\mathbf{J}\mathbf{V}\mathbf{S}^{-1} = \mathbf{W}\mathbf{F}\mathbf{W}^T$ be the spectral decomposition. Then

$$\mathbf{I}_{d^2} - \mathbf{P} + \xi_C(\mathbf{S}^{-1}\mathbf{V}^T\mathbf{J}\mathbf{V}\mathbf{S}^{-1} \otimes \mathbf{S}^{-1}\mathbf{V}^T\mathbf{J}\mathbf{V}\mathbf{S}^{-1}) =$$

$$\begin{aligned} &= \mathbf{I}_{d^2} - \mathbf{P} + \xi_C(\mathbf{W}\mathbf{F}\mathbf{W}^T \otimes \mathbf{W}\mathbf{F}\mathbf{W}^T) \\ &= (\mathbf{W} \otimes \mathbf{W}) \left\{ \mathbf{I}_{d^2} - \tilde{\mathbf{P}} + \xi_C(\mathbf{F} \otimes \mathbf{F}) \right\} (\mathbf{W}^T \otimes \mathbf{W}^T) \end{aligned}$$

with

$$\begin{aligned} \tilde{\mathbf{P}} &= (\mathbf{W}^T \otimes \mathbf{W}^T)\mathbf{P}(\mathbf{W} \otimes \mathbf{W}) \\ &= (\mathbf{W}^T\mathbf{U}^T \otimes \mathbf{W}^T\mathbf{U}^T)\mathbf{E}^T\mathbf{E}(\mathbf{U}\mathbf{W} \otimes \mathbf{U}\mathbf{W}), \end{aligned}$$

so $\tilde{\mathbf{P}}$ is the projection matrix associated with $\left\{(\mathbf{uw})_{j\cdot}^T \otimes (\mathbf{uw})_{j\cdot}^T\right\}$. Then $\mathbf{\Omega}$ comes down to

$$\mathbf{\Omega} = (\mathbf{V}\mathbf{S}\mathbf{W} \otimes \mathbf{V}\mathbf{S}\mathbf{W}) \left\{ \mathbf{I}_{d^2} - \tilde{\mathbf{P}} + \xi_C(\mathbf{F} \otimes \mathbf{F}) \right\} (\mathbf{W}^T\mathbf{S}\mathbf{V}^T \otimes \mathbf{W}^T\mathbf{S}\mathbf{V}^T)$$

and its inverse is

$$\mathbf{\Omega}^{-1} = (\mathbf{V}\mathbf{S}^{-1}\mathbf{W} \otimes \mathbf{V}\mathbf{S}^{-1}\mathbf{W}) \left\{ \mathbf{I}_{d^2} - \tilde{\mathbf{P}} + \xi_C(\mathbf{F} \otimes \mathbf{F}) \right\}^{-1} (\mathbf{W}^T\mathbf{S}^{-1}\mathbf{V}^T \otimes \mathbf{W}^T\mathbf{S}^{-1}\mathbf{V}^T).$$

The inverse of $\mathbf{I}_{d^2} - \tilde{\mathbf{P}} + \xi_C(\mathbf{F} \otimes \mathbf{F})$ is approximated as follows. Let $\mathbf{\Delta} = \mathbf{I}_{d^2} + \xi_C(\mathbf{F} \otimes \mathbf{F})$, which is a diagonal matrix with elements $\mathbf{1} + \xi_C(\mathbf{f} \otimes \mathbf{f})$, all greater than 1 since $\xi_C > 0$ and $f_j > 0$. Then

$$\begin{aligned} (\mathbf{\Delta} - \tilde{\mathbf{P}})^{-1} &= \sum_{k=0}^{\infty} \left(\mathbf{\Delta}^{-1}\tilde{\mathbf{P}} \right)^k \mathbf{\Delta}^{-1} \\ &\approx \mathbf{\Delta}^{-1} + \mathbf{\Delta}^{-1}\tilde{\mathbf{P}}\mathbf{\Delta}^{-1}. \end{aligned}$$

Then, if $\tilde{\mathbf{\Sigma}} = \hat{\mathbf{\Sigma}} - \operatorname{diag}\hat{\mathbf{\Sigma}}$, we have

$$
\begin{aligned}
\operatorname{vec}(\hat{\mathbf{C}}) &= \mathbf{\Omega}^{-1}(\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T)\operatorname{vec}(\tilde{\mathbf{\Sigma}}) \\
&= (\mathbf{V}\mathbf{S}^{-1}\mathbf{W} \otimes \mathbf{V}\mathbf{S}^{-1}\mathbf{W})\left(\mathbf{\Delta} - \tilde{\mathbf{P}}\right)^{-1}(\mathbf{W}^T\mathbf{S}^{-1}\mathbf{V}^T \otimes \mathbf{W}^T\mathbf{S}^{-1}\mathbf{V}^T) \\
&\quad \times(\mathbf{V}\mathbf{S} \otimes \mathbf{V}\mathbf{S})(\mathbf{U}^T \otimes \mathbf{U}^T)\operatorname{vec}(\tilde{\mathbf{\Sigma}}) \\
&= (\mathbf{V}\mathbf{S}^{-1}\mathbf{W} \otimes \mathbf{V}\mathbf{S}^{-1}\mathbf{W})\left(\mathbf{\Delta} - \tilde{\mathbf{P}}\right)^{-1}(\mathbf{W}^T\mathbf{U}^T \otimes \mathbf{W}^T\mathbf{U}^T)\operatorname{vec}(\tilde{\mathbf{\Sigma}}) \\
&= (\mathbf{V}\mathbf{S}^{-1}\mathbf{W} \otimes \mathbf{V}\mathbf{S}^{-1}\mathbf{W})\left(\mathbf{\Delta} - \tilde{\mathbf{P}}\right)^{-1}\operatorname{vec}(\mathbf{W}^T\mathbf{U}^T\tilde{\mathbf{\Sigma}}\mathbf{U}\mathbf{W})
\end{aligned}
$$

so that

$$
\hat{\mathbf{C}} = \mathbf{V}\mathbf{S}^{-1}\mathbf{W}\operatorname{unvec}\{(\mathbf{\Delta} - \tilde{\mathbf{P}})^{-1}\operatorname{vec}(\mathbf{W}^T\mathbf{U}^T\tilde{\mathbf{\Sigma}}\mathbf{U}\mathbf{W})\}\mathbf{W}^T\mathbf{S}^{-1}\mathbf{V}^T.
$$

The hat matrix is

$$
\begin{aligned}
\mathbf{H}_C &= (\mathbf{\Gamma} \otimes \mathbf{\Gamma})\mathbf{\Omega}^{-1}(\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T) \\
&= (\mathbf{U}\mathbf{W} \otimes \mathbf{U}\mathbf{W})\left(\mathbf{\Delta} - \tilde{\mathbf{P}}\right)^{-1}(\mathbf{W}^T\mathbf{U}^T \otimes \mathbf{W}^T\mathbf{U}^T)
\end{aligned}
$$

so that

$$
\begin{aligned}
\mathrm{df}_C &= \operatorname{tr}(\mathbf{H}_C) \\
&= \operatorname{tr}\{\left(\mathbf{\Delta} - \tilde{\mathbf{P}}\right)^{-1}\} \\
&\approx \operatorname{tr}(\mathbf{\Delta}^{-1} + \mathbf{\Delta}^{-1}\tilde{\mathbf{P}}\mathbf{\Delta}^{-1}).
\end{aligned}
$$

To evaluate $(\mathbf{\Delta} - \tilde{\mathbf{P}})^{-1}\operatorname{vec}(\mathbf{A})$, with $\mathbf{A} = \mathbf{W}^T\mathbf{U}^T\tilde{\mathbf{\Sigma}}\mathbf{U}\mathbf{W}$, note that

$$
\begin{aligned}
(\mathbf{\Delta} - \tilde{\mathbf{P}})^{-1}\operatorname{vec}(\mathbf{A}) &\approx \mathbf{\Delta}^{-1}\operatorname{vec}(\mathbf{A}) + \mathbf{\Delta}^{-1}\tilde{\mathbf{P}}\mathbf{\Delta}^{-1}\operatorname{vec}(\mathbf{A}) \\
&= \operatorname{vec}(\mathbf{D}_1 \odot \mathbf{A}) + \mathbf{\Delta}^{-1}\tilde{\mathbf{P}}\operatorname{vec}(\mathbf{D}_1 \odot \mathbf{A}),
\end{aligned}
$$

where $\mathbf{D}_1$ is the $d \times d$ reshape of the diagonal of $\mathbf{\Delta}^{-1}$. Then, to evaluate $\tilde{\mathbf{P}}\operatorname{vec}(\mathbf{M})$, with $\mathbf{M} = \mathbf{D}_1 \odot \mathbf{A}$, we have

$$\begin{aligned}
\tilde{\mathbf{P}} \operatorname{vec}(\mathbf{M}) &= (\mathbf{W}^T\mathbf{U}^T \otimes \mathbf{W}^T\mathbf{U}^T)\{\sum_{j=1}^{d} \mathbf{e}_j\mathbf{e}_j^T \otimes \mathbf{e}_j\mathbf{e}_j^T\}(\mathbf{U}\mathbf{W} \otimes \mathbf{U}\mathbf{W}) \operatorname{vec}(\mathbf{M}) \\
&= \sum_{j=1}^{d} \operatorname{vec}\left(\mathbf{W}^T\mathbf{U}^T\mathbf{e}_j\mathbf{e}_j^T\mathbf{U}\mathbf{W}\mathbf{M}\mathbf{W}^T\mathbf{U}^T\mathbf{e}_j\mathbf{e}_j^T\mathbf{U}\mathbf{W}\right),
\end{aligned}$$

so that

$$\mathbf{\Delta}^{-1}\tilde{\mathbf{P}} \operatorname{vec}(\mathbf{D}_1 \odot \mathbf{A}) = \operatorname{vec}\{\mathbf{D}_1 \odot \sum_{j=1}^{d} (\mathbf{U}\mathbf{W})^T \mathbf{e}_j\mathbf{e}_j^T (\mathbf{U}\mathbf{W})(\mathbf{D_1} \odot \mathbf{A})(\mathbf{U}\mathbf{W})^T\mathbf{e}_j\mathbf{e}_j^T(\mathbf{U}\mathbf{W})\}$$

and then

$$\begin{aligned}
&\operatorname{unvec}\{(\mathbf{\Delta} - \tilde{\mathbf{P}})^{-1} \operatorname{vec}(\mathbf{A})\} \\
&= (\mathbf{D_1} \odot \mathbf{A}) + \mathbf{D}_1 \odot \sum_{j=1}^{d} (\mathbf{U}\mathbf{W})^T \mathbf{e}_j\mathbf{e}_j^T (\mathbf{U}\mathbf{W})(\mathbf{D_1} \odot \mathbf{A})(\mathbf{U}\mathbf{W})^T\mathbf{e}_j\mathbf{e}_j^T(\mathbf{U}\mathbf{W}).
\end{aligned}$$

For $\mathrm{df}_C = \operatorname{tr}(\mathbf{\Delta}^{-1} + \mathbf{\Delta}^{-1}\tilde{\mathbf{P}}\mathbf{\Delta}^{-1})$, note that

$$\operatorname{tr}(\mathbf{\Delta}^{-1}) = \operatorname{sum}(\mathbf{D_1}(:))$$

in Matlab, and

$$\operatorname{tr}(\mathbf{\Delta}^{-1}\tilde{\mathbf{P}}\mathbf{\Delta}^{-1}) = \sum_{k=1}^{d^2} \mathbf{\Delta}_{kk}^{-2}\tilde{\mathbf{P}}_{kk}.$$

To obtain the diagonal of $\tilde{\mathbf{P}}$, note that for a vector $\mathbf{v}$ we have $\mathrm{d}(\mathbf{v}\mathbf{v}^T \otimes \mathbf{v}\mathbf{v}^T) = \mathbf{v}^{\odot 2} \otimes \mathbf{v}^{\odot 2}$, so

$$\mathrm{d}(\tilde{\mathbf{P}}) = \sum_{j=1}^{d} \{(\mathbf{uw})_{j\cdot}^T\}^{\odot 2} \otimes \{(\mathbf{uw})_{j\cdot}^T\}^{\odot 2}.$$

2 Asymptotics

2.1 Proof of Theorem 1

For each $j = 1, \dots, d$, the estimator $\hat{\mu}_j(t)$ can be expressed as

$$\hat{\mu}_j(t) = \boldsymbol{\beta}(t)^T \mathbf{G}^{-1} \bar{\mathbf{Y}},$$

with

$$\mathbf{Y}_i = \sum_{u \in X_i^j} \boldsymbol{\beta}(u).$$

Then

$$E\left\{\hat{\mu}_j(t) - \mu_j(t)\right\}^2 = \operatorname{var} \hat{\mu}_j(t) + \left\{E\hat{\mu}_j(t) - \mu_j(t)\right\}^2$$

and

$$E\left\|\hat{\mu}_j - \mu_j\right\|^2 = \int \operatorname{var} \hat{\mu}_j(t)\ dt + \left\|E\hat{\mu}_j - \mu_j\right\|^2.$$

For the first term, the variance term, we have

$$\operatorname{var} \hat{\mu}_j(t) = \boldsymbol{\beta}(t)^T \mathbf{G}^{-1} \left\{\frac{1}{n} \operatorname{cov}(\mathbf{Y})\right\} \mathbf{G}^{-1} \boldsymbol{\beta}(t)$$

and then

$$\int \operatorname{var} \hat{\mu}_j(t)\ dt = \frac{1}{n} \operatorname{tr}\left\{\mathbf{G}^{-1} \operatorname{cov}(\mathbf{Y})\right\}.$$

The covariance matrix $\operatorname{cov}(\mathbf{Y})$ can be obtained using Campbell's formulas, as in Section 1.1. Since

$$\begin{aligned} E\left(\mathbf{Y}\mathbf{Y}^T\right) &= E\left\{\sum_{u \in X^j} \boldsymbol{\beta}(u) \sum_{v \in X^j} \boldsymbol{\beta}(v)^T\right\} \\ &= E\left\{\sum_{u \in X^j} \sum_{v \in X^j, v \neq u} \boldsymbol{\beta}(u)\boldsymbol{\beta}(v)^T\right\} + E\left\{\sum_{u \in X^j} \boldsymbol{\beta}(u)\boldsymbol{\beta}(u)^T\right\} \\ &= \iint \boldsymbol{\beta}(u)\boldsymbol{\beta}(v)^T R_{jj}(u, v)\ du\ dv + \int \boldsymbol{\beta}(u)\boldsymbol{\beta}(u)^T \mu_j(u)\ du \end{aligned}$$

and

$$
\begin{aligned}
E(\mathbf{Y}) &= E\left\{\sum_{u \in X^j} \boldsymbol{\beta}(u)\right\} \\
&= \int \boldsymbol{\beta}(u)\mu_j(u)\ du,
\end{aligned}
$$

we have

$$
\operatorname{cov}(\mathbf{Y}) = \iint \boldsymbol{\beta}(u)\boldsymbol{\beta}(v)^T \rho_{jj}(u,v)\ du\ dv + \int \boldsymbol{\beta}(u)\boldsymbol{\beta}(u)^T \mu_j(u)\ du.
$$

Then, if we define

$$
\mathbf{M}_1 = \iint \boldsymbol{\beta}(u)\boldsymbol{\beta}(v)^T \rho_{jj}(u,v)\ du\ dv
$$

and

$$
\mathbf{M}_2 = \int \boldsymbol{\beta}(u)\boldsymbol{\beta}(u)^T \mu_j(u)\ du,
$$

we have

$$
\int \operatorname{var} \hat{\mu}_j(t)\ dt = \frac{1}{n}\{\operatorname{tr}\left(\mathbf{G}^{-1}\mathbf{M}_1\right) + \operatorname{tr}\left(\mathbf{G}^{-1}\mathbf{M}_2\right)\}.
$$

From Theorem 6.1 in Agarwal and Studden (1980) we get

$$
\operatorname{tr}\left(\mathbf{G}^{-1}\mathbf{M}_2\right) = O(k).
$$

From Lemma 6.5 in Zhou et al. (1998) we have

$$
\operatorname{tr}\left(\mathbf{G}^{-1}\mathbf{M}_1\right) \leq \lambda_{\max}(\mathbf{G}^{-1}) \operatorname{tr}(\mathbf{M}_1),
$$

where $\lambda_{\max}$ denotes the maximum eigenvalue. By Lemma 6.2 in Zhou et al. (1998) or Lemma 6.5 in Agarwal and Studden (1980) we get

$$
\lambda_{\max}(\mathbf{G}^{-1}) = O(k), \tag{1}
$$

and, since the B-spline basis functions $\beta_l(t)$ satisfy $0 \leq \beta_l(t) \leq 1$ and $\sum_{l=1}^{p} \beta_l(t) = 1$

for all t, we have

$$\begin{aligned}
\operatorname{tr}(\mathbf{M}_1) &= \iint \{\boldsymbol{\beta}(v)^T\boldsymbol{\beta}(u)\}\rho_{jj}(u,v)\ du\ dv \\
&\leq \max_{(u,v)}\{\rho_{jj}(u,v)\} \iint \{\sum_{l=1}^{p}\beta_l(v)\beta_l(u)\}\ du\ dv \\
&\leq \max_{(u,v)}\{\rho_{jj}(u,v)\}(b-a)\int \{\sum_{l=1}^{p}\beta_l(u)\}\ du \\
&= \max_{(u,v)}\{\rho_{jj}(u,v)\}(b-a)^2.
\end{aligned}$$

Then

$$\operatorname{tr}\left(\mathbf{G}^{-1}\mathbf{M}_1\right) = O(k)$$

and we have, for the variance term, that

$$\int \operatorname{var}\hat{\mu}_j(t)\ dt = \frac{1}{n}O\left(k\right),$$

as claimed. (The chain of inequalities used to prove that $\operatorname{tr}(\mathbf{G}^{-1}\mathbf{M}_1) = O(k)$ could also have been used to prove that $\operatorname{tr}(\mathbf{G}^{-1}\mathbf{M}_2) = O(k)$, without resorting to Theorem 6.1 in Agarwal and Studden (1980).)

For the second term, the bias term, observe that

$$\begin{aligned}
E\hat{\mu}_j(t) &= \boldsymbol{\beta}(t)^T\mathbf{G}^{-1}E(\mathbf{Y}) \\
&= \boldsymbol{\beta}(t)^T\mathbf{G}^{-1}\int \boldsymbol{\beta}(u)\mu_j(u)\ du \\
&= P_{\mathscr{B}}\mu_j(t),
\end{aligned}$$

where $P_{\mathscr{B}}\mu_j$ denotes the L^2-projection of the function μ_j on the space $\mathscr{B}$ spanned by $\boldsymbol{\beta}(t)$. Then, by Theorem 6.27 in Schumaker (2007), we have

$$\left\|E\hat{\mu}_j - \mu_j\right\| \leq C_1\omega_r(\mu_j;\max_{1\leq i\leq k+1}|\tau_i - \tau_{i-1}|)_2,$$

where C_1 is a constant that depends only on r and $\omega_r(f;t)_2$ is the modulus of continuity, which, by Theorem 2.59 in Schumaker (2007), satisfies

$$\omega_r(f;t)_2 \leq t^r \left\|D^r f\right\|.$$

Then, since $\max_{1\le i\le k}|\tau_i - \tau_{i-1}| = O(1/k)$, we have

$$\left\|E\hat{\mu}_j - \mu_j\right\|^2 = O\left(\frac{1}{k^{2r}}\right),$$

as claimed. (The same result can be obtained from Lemma 6.7 in Agarwal and Studden (1980).)

For a function $g(k) = \alpha k/n + \beta k^{-2r}$, $g'(k) = \alpha/n + \beta(-2r)k^{-2r-1}$ and then the minimum k^* solves

$$\frac{\alpha}{n} - \beta 2rk^{-2r-1} = 0 \iff \frac{\alpha}{n}k^{2r+1} - \beta 2r = 0 \iff k = cn^{1/(2r+1)},$$

so $k^* = O(n^{1/(2r+1)})$ and

$$g(k^*) = \frac{\alpha}{n}cn^{1/(2r+1)} + \beta c^{-2r}n^{-2r/(2r+1)} = (\alpha c + \beta c^{-2r})n^{-2r/(2r+1)}.$$

2.2 Proof of Theorem 2

We do the proof for $\hat{R}_{jk}(t,t')$ with $j \neq k$; the proof for $\hat{R}_{jj}(t,t')$ is analogous. We can write

$$\begin{aligned}\hat{R}_{jk}(t,t') &= \boldsymbol{\beta}(t)^T\mathbf{G}^{-1}\bar{\mathbf{Y}}\mathbf{G}^{-1}\boldsymbol{\beta}(t') \\ &= \{\boldsymbol{\beta}(t')^T \otimes \boldsymbol{\beta}(t)^T\}(\mathbf{G}^{-1}\otimes\mathbf{G}^{-1})\,\mathrm{vec}\,\bar{\mathbf{Y}},\end{aligned}$$

with

$$\mathbf{Y}_i = \sum_{u\in X_i^j}\sum_{v\in X_i^k}\boldsymbol{\beta}(u)\boldsymbol{\beta}(v)^T.$$

Then

$$E\left\{\hat{R}_{jk}(t,t') - R_{jk}(t,t')\right\}^2 = \mathrm{var}\,\hat{R}_{jk}(t,t') + \left\{E\hat{R}_{jk}(t,t') - R_{jk}(t,t')\right\}^2,$$

so

$$E\left\|\hat{R}_{jk}(t,t') - R_{jk}(t,t')\right\|^2 = \iint \mathrm{var}\,\hat{R}_{jk}(t,t')\ dt\ dt' + \left\|E\hat{R}_{jk} - R_{jk}\right\|^2.$$

For the first term, the variance, we have

$$\mathrm{var}\,\hat{R}_{jk}(t,t') =$$

$$= \{\boldsymbol{\beta}(t')^T \otimes \boldsymbol{\beta}(t)^T\}(\mathbf{G}^{-1} \otimes \mathbf{G}^{-1})\{\frac{1}{n}\operatorname{cov}(\operatorname{vec}\mathbf{Y})\}(\mathbf{G}^{-1} \otimes \mathbf{G}^{-1})\{\boldsymbol{\beta}(t')^T \otimes \boldsymbol{\beta}(t)^T\},$$

so

$$\iint \operatorname{var} \hat{R}_{jk}(t,t')\ dt\ dt' = \operatorname{tr}\left\{(\mathbf{G}^{-1} \otimes \mathbf{G}^{-1})\frac{1}{n}\operatorname{cov}(\operatorname{vec}\mathbf{Y})\right\}.$$

Now, $\operatorname{cov}(\operatorname{vec}\mathbf{Y}) = E(\operatorname{vec}\mathbf{Y})(\operatorname{vec}\mathbf{Y})^T - (E\operatorname{vec}\mathbf{Y})(E\operatorname{vec}\mathbf{Y})^T$, where

$$\operatorname{vec}\mathbf{Y} = \sum_{u\in X^j}\sum_{v\in X^k}\boldsymbol{\beta}(v)\otimes\boldsymbol{\beta}(u)$$

and

$$\begin{aligned}(\operatorname{vec}\mathbf{Y})(\operatorname{vec}\mathbf{Y})^T &= \{\sum_{u\in X^j}\sum_{v\in X^k}\boldsymbol{\beta}(v)\otimes\boldsymbol{\beta}(u)\}\{\sum_{u'\in X^j}\sum_{v'\in X^k}\boldsymbol{\beta}(v')\otimes\boldsymbol{\beta}(u')\}^T \\ &= \sum_{v\in X^k}\sum_{v'\in X^k}\boldsymbol{\beta}(v)\boldsymbol{\beta}(v')^T \otimes \sum_{u\in X^j}\sum_{u'\in X^j}\boldsymbol{\beta}(u)\boldsymbol{\beta}(u')^T.\end{aligned}$$

Using iterated expectations, the conditional independence of (X^j, X^k) given (Λ^j, Λ^k), and Campbell's formulas, we get

$$E\operatorname{vec}\mathbf{Y} = \iint\{\boldsymbol{\beta}(v)\otimes\boldsymbol{\beta}(u)\}R_{jk}(u,v)\ du\ dv$$

and

$$E(\operatorname{vec}\mathbf{Y})(\operatorname{vec}\mathbf{Y})^T =$$

$$= E\left[E\left\{\sum_{v\in X^k}\sum_{v'\in X^k}\boldsymbol{\beta}(v)\boldsymbol{\beta}(v')^T \mid \Lambda^k\right\}\otimes E\left\{\sum_{u\in X^j}\sum_{u'\in X^j}\boldsymbol{\beta}(u)\boldsymbol{\beta}(u')^T \mid \Lambda^j\right\}\right].$$

Since

$$\begin{aligned}E\left\{\sum_{v\in X^k}\sum_{v'\in X^k}\boldsymbol{\beta}(v)\boldsymbol{\beta}(v')^T \mid \Lambda^k\right\} &= \iint \boldsymbol{\beta}(v)\boldsymbol{\beta}(v')^T\Lambda^k(v)\Lambda^k(v')\ dv\ dv' + \\ &\quad \int \boldsymbol{\beta}(v)\boldsymbol{\beta}(v)^T\Lambda^k(v)\ dv\end{aligned}$$

and similarly for the other factor, after distributing the four terms and taking the outer expectations we get

$$E(\operatorname{vec}\mathbf{Y})(\operatorname{vec}\mathbf{Y})^T =$$

$$\iiiint \{\boldsymbol{\beta}(v)\boldsymbol{\beta}(v')^T \otimes \boldsymbol{\beta}(u)\boldsymbol{\beta}(u')^T\} R^{[4]}_{jjkk}(u,u',v,v')\ du\ du'\ dv\ dv' +$$
$$\iiint \{\boldsymbol{\beta}(v)\boldsymbol{\beta}(v)^T \otimes \boldsymbol{\beta}(u)\boldsymbol{\beta}(u')^T\} R^{[3]}_{jjk}(u,u',v)\ du\ du'\ dv +$$
$$\iiint \{\boldsymbol{\beta}(v)\boldsymbol{\beta}(v')^T \otimes \boldsymbol{\beta}(u)\boldsymbol{\beta}(u)^T\} R^{[3]}_{jkk}(u,v,v')\ du\ dv\ dv' +$$
$$\iint \{\boldsymbol{\beta}(v)\boldsymbol{\beta}(v)^T \otimes \boldsymbol{\beta}(u)\boldsymbol{\beta}(u)^T\} R_{jk}(u,v)\ du\ dv,$$

where

$$\begin{aligned} R^{[4]}_{jjkk}(u,u',v,v') &= E\{\Lambda^j(u)\Lambda^j(u')\Lambda^k(v)\Lambda^k(v')\} \\ R^{[3]}_{jjk}(u,u',v) &= E\{\Lambda^j(u)\Lambda^j(u')\Lambda^k(v)\} \\ R^{[3]}_{jkk}(u,v,v') &= E\{\Lambda^j(u)\Lambda^k(v)\Lambda^k(v')\}. \end{aligned}$$

Then, if we define $\rho^{[4]}_{jjkk}(u,u',v,v') = R^{[4]}_{jjkk}(u,u',v,v') - R_{jk}(u,v)R_{jk}(u',v')$, we have

$$\mathrm{cov}(\mathrm{vec}\,\mathbf{Y}) = \mathbf{M}_1 + \mathbf{M}_2 + \mathbf{M}_3 + \mathbf{M}_4,$$

with

$$\begin{aligned} \mathbf{M}_1 &= \iiiint \{\boldsymbol{\beta}(v)\boldsymbol{\beta}(v')^T \otimes \boldsymbol{\beta}(u)\boldsymbol{\beta}(u')^T\} \rho^{[4]}_{jjkk}(u,u',v,v')\ du\ du'\ dv\ dv' \\ \mathbf{M}_2 &= \iiint \{\boldsymbol{\beta}(v)\boldsymbol{\beta}(v)^T \otimes \boldsymbol{\beta}(u)\boldsymbol{\beta}(u')^T\} R^{[3]}_{jjk}(u,u',v)\ du\ du'\ dv \\ \mathbf{M}_3 &= \iiint \{\boldsymbol{\beta}(v)\boldsymbol{\beta}(v')^T \otimes \boldsymbol{\beta}(u)\boldsymbol{\beta}(u)^T\} R^{[3]}_{jkk}(u,v,v')\ du\ dv\ dv' \\ \mathbf{M}_4 &= \iint \{\boldsymbol{\beta}(v)\boldsymbol{\beta}(v)^T \otimes \boldsymbol{\beta}(u)\boldsymbol{\beta}(u)^T\} R_{jk}(u,v)\ du\ dv, \end{aligned}$$

all of them symmetric nonnegative definite matrices. Then

$$\iint \mathrm{var}\,\hat{R}_{jk}(t,t')\ dt\ dt' \leq \lambda_{\max}(\mathbf{G}^{-1} \otimes \mathbf{G}^{-1}) \frac{1}{n}\,\mathrm{tr}(\mathbf{M}_1 + \mathbf{M}_2 + \mathbf{M}_3 + \mathbf{M}_4).$$

As in the proof of Theorem 1 above, we use the fact that $0 \leq \beta_l(t) \leq 1$ and

$\sum_{l=1}^{p} \beta_l(t) = 1$ for all t to obtain the inequalities

$$
\begin{aligned}
\mathrm{tr}(\mathbf{M}_1) &= \iiiint \{\boldsymbol{\beta}(v')^T \boldsymbol{\beta}(v)\}\{\boldsymbol{\beta}(u')^T \boldsymbol{\beta}(u)\} \rho^{[4]}_{jjkk}(u,u',v,v')\ du\ du'\ dv\ dv' \\
&\leq \max \rho^{[4]}_{jjkk} \iiiint \{\sum_{l=1}^{p} \beta_l(v)\}\{\sum_{l=1}^{p} \beta_l(u)\}\ du\ du'\ dv\ dv' \\
&= (b-a)^4 \max \rho^{[4]}_{jjkk},
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{tr}(\mathbf{M}_2) &= \iiint \{\boldsymbol{\beta}(v)^T \boldsymbol{\beta}(v)\}\{\boldsymbol{\beta}(u')^T \boldsymbol{\beta}(u)\} R^{[3]}_{jjk}(u,u',v)\ du\ du'\ dv \\
&\leq \max R^{[3]}_{jjk} \iiint \{\sum_{l=1}^{p} \beta_l(v)\}\{\sum_{l=1}^{p} \beta_l(u)\}\ du\ du'\ dv \\
&= (b-a)^3 \max R^{[3]}_{jjk},
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{tr}(\mathbf{M}_3) &= \iiint \{\boldsymbol{\beta}(v')^T \boldsymbol{\beta}(v)\}\{\boldsymbol{\beta}(u)^T \boldsymbol{\beta}(u)\} R^{[3]}_{jkk}(u,v,v')\ du\ dv\ dv' \\
&\leq \max R^{[3]}_{jkk} \iiint \{\sum_{l=1}^{p} \beta_l(v)\}\{\sum_{l=1}^{p} \beta_l(u)\}\ du\ dv\ dv' \\
&= (b-a)^3 \max R^{[3]}_{jkk},
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{tr}(\mathbf{M}_4) &= \iint \{\boldsymbol{\beta}(v)^T \boldsymbol{\beta}(v)\}\{\boldsymbol{\beta}(u)^T \boldsymbol{\beta}(u)\} R_{jk}(u,v)\ du\ dv \\
&\leq \max R_{jk} \iint \{\sum_{l=1}^{p} \beta_l(v)\}\{\sum_{l=1}^{p} \beta_l(u)\}\ du\ dv \\
&= (b-a)^2 \max R_{jk}.
\end{aligned}
$$

On the other hand,

$$
\lambda_{\max}(\mathbf{G}^{-1} \otimes \mathbf{G}^{-1}) = \lambda_{\max}(\mathbf{G}^{-1})^2 = O(k)^2
$$

from (1). Then

$$\iint \operatorname{var} \hat{R}_{jk}(t,t')\ dt\ dt' = \frac{1}{n} O\left(k^2\right).$$

Regarding the bias term, we have

$$\begin{aligned} E\hat{R}_{jk}(t,t') &= \{\boldsymbol{\beta}(t')^T \otimes \boldsymbol{\beta}(t)^T\}(\mathbf{G}^{-1} \otimes \mathbf{G}^{-1}) E \operatorname{vec} \mathbf{Y} \\ &= \{\boldsymbol{\beta}(t')^T \otimes \boldsymbol{\beta}(t)^T\}(\mathbf{G}^{-1} \otimes \mathbf{G}^{-1}) \iint \{\boldsymbol{\beta}(v) \otimes \boldsymbol{\beta}(u)\} R_{jk}(u,v)\ du\ dv, \end{aligned}$$

which is the L_2 orthogonal projection of the bivariate function R_{jk} on the tensor-product B-spline space spanned by $\boldsymbol{\beta}(t') \otimes \boldsymbol{\beta}(t)$. Then, by Theorem 12.8 of Schumaker (2007) we have

$$\left\| R_{jk} - E\hat{R}_{jk} \right\| \leq C(\delta^r \left\| D_1^r R_{jk} \right\| + \delta^r \left\| D_2^r R_{jk} \right\|)$$

where C is a constant that depends only on r and $\delta = \max_{1 \leq i \leq k+1} |\tau_i - \tau_{i-1}|$. Since $\delta = O(1/k)$, we have

$$\left\| R_{jk} - E\hat{R}_{jk} \right\|^2 = O\left(\frac{1}{k^{2r}}\right).$$

Putting both variance and bias terms together we get that

$$E \left\| \hat{R}_{jk}(t,t') - R_{jk}(t,t') \right\|^2 = \frac{1}{n} O\left(k^2\right) + O\left(\frac{1}{k^{2r}}\right),$$

as claimed. To find the optimal rate k^*, let $g(k) = \alpha k^2/n + \beta k^{-2r}$. Then $g'(k) = \alpha 2k/n - \beta 2r k^{-2r-1}$, so that k^* solves

$$\frac{\alpha 2k}{n} - \beta 2r k^{-2r-1} = 0 \iff \frac{\alpha 2k^{2r+2}}{n} - \beta 2r = 0 \iff k = cn^{1/(2r+2)}.$$

Then $k^* = O(n^{1/(2r+2)})$ and

$$g(k^*) = \alpha c^2 n^{\frac{2}{2r+2}-1} + \beta c^{-2r} n^{-\frac{2r}{2r+2}} = (\alpha c^2 + \beta c^{-2r}) n^{-\frac{2r}{2r+2}},$$

so $g(k^*) = O(n^{-2r/(2r+2)})$ as claimed.

3 Simulations

3.1 Mean and covariance functions

The latent processes $\Lambda(t, \mathbf{s}_j)$'s were generated following the log-Gaussian model

$$\Lambda(t, \mathbf{s}_j) = \exp\{\nu(t) + U_j \phi(t)\} \tag{2}$$

for $t \in [0, 1]$, with $\nu(t) = \sin(\pi t) + \ln 20$ and $\phi(t) = \sqrt{2}\sin(\pi t)$. The U_j's were defined as

$$U_j = g(\mathbf{s}_j)W + E_j, \tag{3}$$

with $g(\mathbf{s}) = 1/(1 + \|\mathbf{s}\|)$ for Model 1 and $g(\mathbf{s}) = 1$ for Model 2, $W \sim N(0, 0.072)$ and $E_j \sim N(0, 0.018)$. The E_j's were independent among themselves and of W.

If follows from (2) that $\Lambda(t, \mathbf{s}_j) \sim LN(\nu(t), \sigma_j^2 \phi^2(t))$, where

$$\begin{aligned} \sigma_j^2 &= \operatorname{var}(U_j) \\ &= g^2(\mathbf{s}_j)\sigma_W^2 + \sigma_E^2. \end{aligned}$$

Then

$$\mu(t, \mathbf{s}_j) = \exp\left\{\nu(t) + \sigma_j^2 \phi^2(t)/2\right\}.$$

Similarly, $\Lambda^2(t, \mathbf{s}_j) \sim LN(2\nu(t), 4\sigma_j^2 \phi^2(t))$, so

$$\operatorname{var}\{\Lambda(t, \mathbf{s}_j)\} = \exp\left\{2\nu(t) + 2\sigma_j^2 \phi^2(t)\right\} - \exp\left\{2\nu(t) + \sigma_j^2 \phi^2(t)\right\}.$$

Finally, for $\mathbf{s}_j \neq \mathbf{s}_k$,

$$\Lambda(t, \mathbf{s}_j)\Lambda(t, \mathbf{s}_k) = \exp\{2\nu(t) + (U_j + U_k)\phi(t)\}$$

with

$$U_j + U_k = \{g(\mathbf{s}_j) + g(\mathbf{s}_k)\}W + E_j + E_k,$$

so $\Lambda(t, \mathbf{s}_j)\Lambda(t, \mathbf{s}_k) \sim LN(2\nu(t), \sigma_{jk}^2 \phi^2(t))$ where

$$\sigma_{jk}^2 = \{g(\mathbf{s}_j) + g(\mathbf{s}_k)\}^2 \sigma_W^2 + 2\sigma_E^2.$$

Grid	n	Model 1 $\mathbf{M}$	$\mathbf{m}_0$	$\boldsymbol{\Sigma}$	$\boldsymbol{\sigma}_0$	SPE	Model 2 $\mathbf{M}$	$\mathbf{m}_0$	$\boldsymbol{\Sigma}$	$\boldsymbol{\sigma}_0$	SPE
(i)	50	.0049	.027	.044	.34	.80	.0046	.0037	.024	.25	.48
	100	.0030	.022	.044	.34	.51	.0034	.0025	.018	.24	.27
	200	.0056	.018	.035	.31	.51	.0081	.0080	.016	.24	.41
	400	.0046	.020	.015	.34	.61	.0046	.0042	.011	.25	.61
(ii)	50	.012	.0043	.050	.12	.070	.0035	.0023	.029	.022	.045
	100	.010	.0047	.021	.08	.049	.0070	.0070	.037	.035	.031
	200	.003	.0131	.016	.10	.040	.0033	.0025	.024	.029	.018
	400	.006	.0085	.012	.08	.034	.0018	.0015	.013	.013	.018
(iii)	50	.0122	.012	.046	.14	.23	.0122	.0119	.032	.022	.18
	100	.0037	.024	.025	.11	.21	.0039	.0030	.052	.048	.13
	200	.0058	.028	.024	.13	.19	.0031	.0026	.015	.012	.09
	400	.0016	.022	.020	.12	.17	.0022	.0018	.011	.014	.06

Table 1: Simulation Results. Relative biases of parameter estimators.

Then

$$\begin{aligned} \operatorname{cov}\{\Lambda(t,\mathbf{s}_j),\Lambda(t,\mathbf{s}_k)\} &= \exp\left\{2\nu(t)+\sigma_{jk}^2\phi^2(t)/2\right\} \\ &\quad -\exp\left\{2\nu(t)+(\sigma_j^2+\sigma_k^2)\phi^2(t)/2\right\}. \end{aligned}$$

3.2 Additional simulation results

The following tables show relative biases (Table 1) and standard deviations (Table 2) of the estimators.

4 References

Agarwal, G.G., and Studden, W.J. (1980). Asymptotic integrated mean square error using least squares and bias minimizing splines. *The Annals of Statistics* **8** 1307–1325.

Baddeley, A. (2007). Spatial point processes and their applications. In *Stochastic Geometry, Lecture Notes in Mathematics* **1892** 1–75. Springer, New York.

Grid	n	Model 1 $\mathbf{M}$	$\mathbf{m}_0$	$\mathbf{\Sigma}$	$\boldsymbol{\sigma}_0$	SPE	Model 2 $\mathbf{M}$	$\mathbf{m}_0$	$\mathbf{\Sigma}$	$\boldsymbol{\sigma}_0$	SPE
(i)	50	.077	.069	.41	.27	.58	.110	.105	.41	.34	.47
	100	.057	.052	.29	.20	.37	.070	.066	.28	.24	.26
	200	.041	.039	.22	.14	.19	.047	.044	.19	.16	.35
	400	.026	.024	.14	.10	.22	.041	.040	.12	.10	.32
(ii)	50	.088	.080	.35	.25	.042	.102	.097	.43	.39	.037
	100	.064	.060	.27	.21	.031	.073	.069	.25	.22	.027
	200	.043	.039	.18	.13	.016	.048	.045	.19	.17	.015
	400	.031	.028	.13	.10	.012	.039	.037	.13	.11	.015
(iii)	50	.075	.067	.38	.23	.057	.109	.104	.38	.36	.066
	100	.053	.048	.29	.19	.037	.071	.067	.25	.22	.044
	200	.036	.032	.20	.12	.037	.053	.050	.18	.15	.033
	400	.027	.025	.15	.10	.026	.033	.031	.13	.11	.021

Table 2: Simulation Results. Relative standard deviations of parameter estimators.

Møller, J., and Waagepetersen, R.P. (2004). *Statistical Inference and Simulation for Spatial Point Processes.* Chapman and Hall/CRC, Boca Raton.

Schumaker, L.L. (2007). *Spline Functions: Basic Theory. Third Edition.* Cambridge University Press, Cambridge, UK.

Zhou, S., Shen, X., and Wolfe, D.A. (1998). Local asymptotics for regression splines and confidence region. *The Annals of Statistics* **26** 1760–1782.