Smooth Scalar-on-Image Regression via
Spatial Bayesian Variable Selection: Supplement A

By: Jeff Goldsmith, Lei Huang, Ciprian Crainiceanu

A Theoretical Results

Recall our previous notation that δ_l is the neighborhood consisting of all image locations sharing a face (but not a corner) with location l, and that $d_l = |\delta_l|$ where $|\cdot|$ denotes the cardinality of a set.

Theorem 1. *If there exists at least one location l for which $\gamma_l = 0$, then $p(\boldsymbol{\beta} \mid \boldsymbol{\gamma})$ is proper.*

Proof. Define sets $L = \{l : \boldsymbol{\gamma}_l = 1\}$ and $L^C = \{l : \boldsymbol{\gamma}_l = 0\}$ that partition image locations in predictive and non-predictive locations, and assume that L^C is non-empty. Let $\boldsymbol{\beta}_L$ and $\boldsymbol{\beta}_{L^C}$ be the coefficient image at the predictive and non-predictive locations, respectively; also let $\underset{\sim}{\boldsymbol{\beta}}_L$ be the regression coefficients of predictive locations expressed as a vector.

For $l \in L^C$, $[\beta_l | \gamma_l = 0] \sim \delta(0)$ where $\delta(0)$ is the Dirac delta function. For $l \in L$, $[\beta_l | \boldsymbol{\beta}_{-l}, \boldsymbol{\gamma}_{-l}] \sim \mathrm{N}\left[\frac{\sum_{l' \in \delta_l} \beta_{l'}\gamma_{l'}}{d_l}, \sigma^2_{\boldsymbol{\beta}}/d_l\right]$. Note that $\frac{\sum_{l' \in \delta_l} \beta_{l'}\gamma_{l'}}{d_l} = \frac{\sum_{l' \in (\delta_l \cap L)} \beta_{l'}}{d_l}$ so that the distribution depends only on elements of $\boldsymbol{\beta}_L$. Defining $b_{ll'} = \frac{1}{d_l}$, we have $\frac{\sum_{l' \in (\delta_l \cap L)} \beta_{l'}}{d_l} = \sum_{l' \in (\delta_l \cap L)} b_{ll'}\beta_{l'}$. Following Brook's lemma (Brook, 1964), this prior specification for $[\beta_l | \boldsymbol{\beta}_{-l}, \boldsymbol{\gamma}_{-l}]$ results in the joint density $f(\underset{\sim}{\boldsymbol{\beta}}_L) \propto \exp\left[\frac{-1}{2}\underset{\sim}{\boldsymbol{\beta}}_L^T D^{-1}(I - B)\underset{\sim}{\boldsymbol{\beta}}_L\right]$ where D and B are square matrices of size $|L| \times |L|$ whose rows and columns correspond to the location ordering of $\underset{\sim}{\boldsymbol{\beta}}_L$. D is a diagonal matrix with entries $D_{ll} = \sigma^2_{\boldsymbol{\beta}}/d_l$ and B has elements $B_{ll'} = b_{ll'}$ defined above and is zero elsewhere.

Notice $b_{ll'}d_l = b_{l'l}d_{l'}$, so $D^{-1}(I - B)$ is a symmetric matrix. Next, we show this matrix is positive definite. In the following, the notation $l < l'$ will indicate that the location l precedes location l' in the ordering of the vector $\underset{\sim}{\boldsymbol{\beta}}_L$. For an arbitrary nonzero vector x of length $|L|$

$$
\begin{aligned}
x^T D^{-1}(I-B)x &= \sum_{l\in L} \frac{d_l}{\sigma^2_{\boldsymbol{\beta}}} x_l^2 - 2\sum_{l<l'} \frac{d_l}{\sigma^2_{\boldsymbol{\beta}}} b_{ll'} x_l x_{l'} \\
&= \sum_{l\in L} \frac{d_l}{\sigma^2_{\boldsymbol{\beta}}} x_l^2 + \sum_{l<l'} \frac{d_l}{\sigma^2_{\boldsymbol{\beta}}} b_{ll'} (x_l^2 - 2x_l x_{l'} + x_l^2) - \sum_{l<l'} \frac{d_l}{\sigma^2_{\boldsymbol{\beta}}} b_{ll'} (x_l^2 + x_{l'}^2) \\
&= \sum_{l\in L} \frac{d_l}{\sigma^2_{\boldsymbol{\beta}}} x_l^2 + \sum_{l<l'} \frac{d_l}{\sigma^2_{\boldsymbol{\beta}}} b_{ll'} (x_l - x_{l'})^2 - \sum_{l\neq l'} \frac{d_l}{\sigma^2_{\boldsymbol{\beta}}} b_{ll'} x_l^2 \\
&= \sum_{l\in L} \left[\frac{d_l}{\sigma^2_{\boldsymbol{\beta}}} \left(1 - \sum_{\{l':l'\neq l\}} b_{ll'} \right) x_l^2 \right] + \sum_{l<l'} \frac{d_l}{\sigma^2_{\boldsymbol{\beta}}} b_{ll'} (x_l - x_l')^2. \qquad (10)
\end{aligned}
$$

Because $b_{ll'}$ is non-negative, we have $\sum_{l<l'} \frac{d_l}{\sigma^2_{\boldsymbol{\beta}}} b_{ll'}(x_l - x_l')^2 \geq 0$. Moreover, for every l the term $\sum_{\{l':l'\neq l\}} b_{ll'} = \frac{1}{d_l} \cdot |\{l' : l' \in \delta_l \cap L\}| \leq 1$. Because we assume L^C is non-empty, there exists at least one location l such that the preceding inequality is strict. Therefore there is at least one location l for which

$$
\sum_{l\in L} \left[\frac{d_l}{\sigma^2_{\boldsymbol{\beta}}} \left(1 - \sum_{\{l':l'\neq l\}} b_{ll'} \right) x_l^2 \right] > 0
$$

and thus $x^T D^{-1}(I-B)x > 0$. So, the matrix $D^{-1}(I-B)$ is symmetric and positive definite, and $[\underset{\sim}{\boldsymbol{\beta}}_L]$ has a proper joint distribution. Finally, we note $f(\boldsymbol{\beta}|\boldsymbol{\gamma}) = f(\underset{\sim}{\boldsymbol{\beta}}_L)\prod_{l\in L^C}\delta(0)$, which is the product of proper densities. $\square$