Objective Bayesian Analysis of Spatially Correlated Data

J. Berger, V. De Oliveira, and B. Sansó

Duke University and Universidad Simón Bolívar

Outline

- The spatial model
- Common objective priors
- The Jeffreys prior
- The reference prior
- Choosing among spatial models
- Conclusions and generalizations

The Spatial Model

$Z(\mathbf{s})$ is a random process observed at $\mathbf{s}_i \in \Re^k, i = 1, \ldots, n$,

$$Z(\mathbf{s}) = \sum_{j=1}^{p} \beta_j f_j(\mathbf{s}) + \varepsilon(\mathbf{s})$$

where $f_j(\mathbf{s})$ are known regression functions, β_j are unknown coefficients and ε is a Gaussian process with zero mean and

$$cov(\varepsilon(\mathbf{s}), \varepsilon(\mathbf{u})) = cov(Z(\mathbf{s}), Z(\mathbf{u})) = \sigma^2 K_\vartheta(\|\mathbf{s} - \mathbf{u}\|)$$

Example: $K_\vartheta(\|\mathbf{s} - \mathbf{u}\|) = e^{-\|\mathbf{s}-\mathbf{u}\|/\vartheta}, \;\; \vartheta > 0.$

Likelihood Function:

Given data $\mathbf{z} = (z_1, \ldots, z_n)'$, $z_i = Z(\mathbf{s}_i)$,

$$L(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\vartheta}; \mathbf{z}) = \frac{|\Sigma_\vartheta|^{-1/2}}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{(\mathbf{z} - X\boldsymbol{\beta})'\Sigma_\vartheta^{-1}(\mathbf{z} - X\boldsymbol{\beta})}{2\sigma^2} \right)$$

where X $(n \times p)$ is defined by $X_{ij} = f_j(\mathbf{s}_i)$,

and Σ_ϑ $(n \times n)$ is defined by $\Sigma_{\vartheta,ij} = K_\vartheta(\|\mathbf{s}_i - \mathbf{s}_j\|)$.

Example: $\Sigma_\vartheta = \left(e^{-d_{ij}/\vartheta} \right)$, $\quad d_{ij} = \|\mathbf{s}_i - \mathbf{s}_j\|$, $\; D = (d_{ij})$.

Common Objective Priors

Natural objective priors are of the form

$$\pi(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\vartheta}) = \frac{\pi(\boldsymbol{\vartheta})}{\sigma^2}.$$

How should $\pi(\boldsymbol{\vartheta})$ be chosen?

- **Uniform:** $\pi(\boldsymbol{\vartheta}) = 1$ (Handcock and Stein, 1993).
  Problem: The resulting posterior is improper.

- **Truncated uniform:** $\pi(\vartheta) = \frac{1}{M}$ on $0 < \vartheta < M$.
  Problem: M must be a 'moderate' subjective choice; for 'large' M, posterior mass will pile up near M.

- **Vague proper priors:** Should not be used because
  - same problem: mass will pile up at extremes;
  - this may not be detectable, whereas with improper priors any impropriety in the posterior often is revealed through the computation;
  - even sensitivity studies can fail to pick up the problem.
- **Transformation of ϑ**: When $\vartheta \in R_1$, transform ϑ to $(0,1)$ by $\psi = h(\vartheta)$, and then choose $\pi(\psi) = 1$.
  But this is equivalent to choosing $\pi(\vartheta) = |h^{'}(\vartheta)|$.

The Jeffreys Prior

Computation shows that the Jeffreys prior is $\pi^J(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\vartheta}) = \pi^J(\boldsymbol{\vartheta}) \,/\, \sigma^2$, where

$$\pi^J(\boldsymbol{\vartheta}) = [\sum_{i=1}^{n}(u_i(\boldsymbol{\vartheta}) - \overline{u}(\boldsymbol{\vartheta}))^2]^{\frac{1}{2}},$$

$u_1, \ldots, u_n$ being the eigenvalues of $U_\vartheta = (\frac{\partial}{\partial\vartheta}\Sigma_\vartheta)\Sigma_\vartheta^{-1}$ and $\overline{u}$ being their average.

- Posterior is proper if X does not contain a constant column.
- Posterior is improper if X contains a constant column.

The Reference Prior

Step 1. Use two groups, $\boldsymbol{\beta}$ as the nuisance parameters and $(\sigma^2, \boldsymbol{\vartheta})$ as the parameters of interest, and write

$$\pi^R(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\vartheta}) = \pi^R(\boldsymbol{\beta}|\sigma^2, \boldsymbol{\vartheta})\pi^R(\sigma^2, \boldsymbol{\vartheta}).$$

Step 2. Use the conditional reference (Jeffreys) prior

$$\pi^R(\boldsymbol{\beta}|\sigma^2, \boldsymbol{\vartheta}) = 1.$$

Step 3. Integrate out $\boldsymbol{\beta}$, obtaining the *marginal model*

$$L^I(\sigma^2, \boldsymbol{\vartheta}; \mathbf{z}) = \int_{R^p} L(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\vartheta}; \mathbf{z}) \pi^R(\boldsymbol{\beta} | \sigma^2, \boldsymbol{\vartheta}) d\boldsymbol{\beta}$$

$$\propto \frac{|X' \Sigma_\vartheta^{-1} X|^{-1/2}}{\sigma^{(n-p)} |\Sigma_\vartheta|^{1/2}} \exp\{-\frac{1}{2\sigma^2} (\mathbf{z} - X\hat{\boldsymbol{\beta}}_\vartheta)' \Sigma_\vartheta^{-1} (\mathbf{z} - X\hat{\boldsymbol{\beta}}_\vartheta)\},$$

where $\hat{\boldsymbol{\beta}}_\vartheta = (X'\Sigma_\vartheta^{-1} X)^{-1} X' \Sigma_\vartheta^{-1} \mathbf{z}$.

Step 4. Compute Jeffreys prior for this marginal model:

$$\pi^R(\sigma^2, \boldsymbol{\vartheta}) \propto \frac{1}{\sigma^2} \{tr[W_\vartheta] - \frac{1}{n-p} (tr[W_\vartheta])^2\}^{\frac{1}{2}},$$

where $W_\vartheta = \frac{\partial}{\partial \vartheta} \Sigma_\vartheta^{-1} \Sigma_\vartheta^{-1} (I - X(X'\Sigma_\vartheta^{-1} X)^{-1} X' \Sigma_\vartheta^{-1})$.

Step 5. After technical details (operating on compacts, taking asymptotic limits, algebra), the reference prior is

$$\pi^R(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\vartheta}) = \frac{1}{\sigma^2}\,\pi^R(\boldsymbol{\vartheta}) = \frac{1}{\sigma^2}\,[\sum_{i=1}^{n-p}(w_i(\boldsymbol{\vartheta}) - \overline{w}(\boldsymbol{\vartheta}))^2]^{\frac{1}{2}},$$

where $w_1, \ldots, w_{n-p}$ are the eigenvalues of W_ϑ and $\overline{w}$ is their average.

Properties and comments:

- The reference prior always yields a proper posterior for suitable $K_\vartheta(\|\mathbf{s} - \mathbf{u}\|)$.

- The reference prior algorithm given in (say) Berger and Bernardo (1992), based on asymptotic marginalization, does not yield a proper posterior.

- $\pi^R(\vartheta)$ is a proper density.

- There are numerical issues in computing W_ϑ; otherwise computation with the prior is easy.

Needed conditions on the covariance matrix:

- $\Sigma_\vartheta = \mathbf{1}\mathbf{1}' + \varepsilon_\vartheta B + \varepsilon_\vartheta B \Lambda_\vartheta^{(1)}$, as $\vartheta \to \infty$,
  where $\mathbf{1} = (1, \ldots, 1)'$, B is a fixed nonsingular matrix and $\vartheta^{-1}\sqrt{\varepsilon_\vartheta}$ and $\vartheta^{-1}\Lambda_\vartheta^{(1)}$ are integrable at infinity.
- $\frac{\partial}{\partial\vartheta}\Sigma_\vartheta = c\vartheta^{-1}\varepsilon_\vartheta B(I + \Lambda_\vartheta^{(2)})$,
  where $\vartheta^{-1}\Lambda_\vartheta^{(2)}$ is integrable at infinity.

Example: $\Sigma_\vartheta = \left(e^{-d_{ij}/\vartheta}\right) = \mathbf{1}\mathbf{1}' + \vartheta^{-1}D + O(\vartheta^{-2})$,
$\frac{\partial}{\partial\vartheta}\Sigma_\vartheta = -\vartheta^{-2}D + O(\vartheta^{-3})$.
Other examples: Spherical, power exponential (known shape parameter), rational quadratic (known shape), and Matérn (known order).

Behavior of the likelihood function:

Integrating out $(\boldsymbol{\beta}, \sigma^2)$ yields the integrated likelihood

$$L^I(\vartheta) \propto |\Sigma_\vartheta|^{-\frac{1}{2}} |X' \Sigma_\vartheta^{-1} X|^{-\frac{1}{2}} \tilde{q}_\vartheta^{-\frac{1}{2}(n-p)}$$

where $\tilde{q}_\vartheta = (\mathbf{z} - X\hat{\boldsymbol{\beta}}_\vartheta)' \Sigma_\vartheta^{-1} (\mathbf{z} - X\hat{\boldsymbol{\beta}}_\vartheta)$.

Example: For the exponential covariance function, $\lim_{\vartheta \to 0} L^I(\vartheta) = c_1 > 0$ and

$$L^I(\vartheta) = \begin{cases} O(1) & \text{if } \mathbf{1} \text{ is a column of } X \\ O(\vartheta^{-\frac{1}{2}}) & \text{if } \mathbf{1} \text{ is not a column of } X \end{cases}, \text{ as } \vartheta \to \infty.$$

Choosing Among Spatial Models

For comparing two covariance kernels, $K^1_{\vartheta_1}$ and $K^2_{\vartheta_2}$, one can simply compute the Bayes factor using the reference priors. This is justified because:

- The (conditional) reference priors, $\pi^R(\vartheta_1)$ and $\pi^R(\vartheta_2)$, are proper so the marginal models formed by integrating out ϑ are proper models.
- The marginal models are location-scale models in $(\boldsymbol{\beta}, \sigma^2)$, so one can use the reference prior in these variables (Berger, Pericchi, and Varshavsky, 1998).

Conclusions and Generalizations

- Standard noninformative priors cannot be used for spatial models.
- The Jeffreys prior works only if there is no constant term in the model.
- The reference prior based on asymptotic mixing fails.
- The reference prior based on exact mixing works.
- Reference priors can be used directly for Bayes factor comparisons of spatial covariance kernels.

- The results generalize to any Gaussian problem s.t.

$$\Sigma_\vartheta = \mathbf{1}\mathbf{1}' + \varepsilon_\vartheta B + \varepsilon_\vartheta B \Lambda_\vartheta^{(1)}.$$

  This includes many time series and mixed model problems.