Online Supplement: Computational Issues

Shengde Liang, Bradley P. Carlin and Alan E. Gelfand

Abstract

In this supplement to the paper by Liang et al. (2009), we provide full details of the Monte Carlo algorithms needed to approximate the complex point process likelihoods in the paper. In particular, we flesh out the details of our knot-based predictive process approximation, and give general guidelines for how the knots should be selected in any given application.

The likelihoods in the main paper require us to integrate the intensity function over D, which is a stochastic integral with regard to the $w_k(\mathbf{s})$ processes and is unavailable in closed form. Hence, we need to approximate these random variables.

Recalling the discussion below (3), let $\mathcal{C} = \{\mathbf{s}_j^*,\, j = 1, 2, \cdots, T\}$ be a generic set of locations in the domain that permits a sufficiently accurate approximation of the integral in the likelihood by

$$\int_D \int_{\mathcal{V}} \lambda_k(\mathbf{s}, \mathbf{v}) d\mathbf{v} d\mathbf{s} \approx \mathcal{F}_k(\mathbf{s}_1^*, \mathbf{s}_2^*, \cdots, \mathbf{s}_T^*),$$

where $\mathcal{F}_k(\cdot)$ is some mark-specific function and the integral over $\mathcal{V}$ has already been computed analytically. For instance, if $\mathcal{C}$ is a uniformly drawn random set in the region D, $\mathcal{F}_k$ becomes the Monte Carlo sum,

$$\mathcal{F}_k(\mathbf{s}_1^*, \mathbf{s}_2{}^*, \cdots, \mathbf{s}_T^*) = \int_{\mathcal{V}} \frac{|D|}{T} \sum_j \lambda_k(\mathbf{s}_j^*, \mathbf{v}) d\mathbf{v}\ .$$

Note the integral with respect to $\mathbf{v}$ can be done analytically regardless of whether $\mathbf{v}$ is discrete or continuous, as discussed previously. However, this results in a computational difficulty: the likelihood will require us to work with an $(n + T)$-dimensional multivariate normal distribution, where T could be very large. This will in turn require inverses and determinants of very large covariance matrices.

This so-called "big n problem" is receiving significant attention along several different paths. The

first seeks approximations for the spatial process using kernel convolutions, moving averages or basis functions (e.g., Wikle and Cressie, 1999; Higdon, 2002; Ver Hoef et al., 2004; Xia and Gelfand, 2006; Paciorek and Schervish, 2006, Banerjee et al., 2008). Essentially, the process $w(\mathbf{s})$ is replaced by an approximation $\tilde{w}(\mathbf{s})$ that represents the realizations in a lower-dimensional subspace. A second approach seeks instead to approximate the likelihood either by working in the spectral domain of the spatial process and avoiding the matrix computations (Stein, 1999; Fuentes, 2007; Paciorek, 2007) or by forming a product of appropriate conditional distributions to approximate the likelihood (e.g. Vecchia, 1988; Jones and Zhang, 1997; Stein et al., 2004). Still another approach is to replace the process (random field) model by a *Markov* random field (Cressie, 1993), or else approximate the random field model by a Markov random field (Rue and Tjelmeland, 2002; Rue and Held, 2006).

Here we will employ the predictive process method advocated by Banerjee et al. (2008). With multiple types of outcomes, if we assume that $\mathbf{w}(s) = (w_1(\mathbf{s}), \ldots, w_K(\mathbf{s}))'$ is a multivariate Gaussian process $MVGP(\mathbf{0}, \Gamma_{\mathbf{w}}(\cdot, \cdot, \boldsymbol{\theta}))$, then the original process is actually a multivariate log Gaussian Cox process. Let $S^* = \{\mathbf{s}_1^*, \ldots, \mathbf{s}_m^*\}$ be an arbitrary set of knots, then $\mathbf{w}^* = [w_k(\mathbf{s}_j^*)]_{k,j} \sim MVN(\mathbf{0}, \Gamma^*(\boldsymbol{\theta}))$ is an realization of $\mathbf{w}(\mathbf{s})$ over S^*, with $\Gamma^*(\boldsymbol{\theta}) = [\Gamma(\mathbf{s}_i^*, \mathbf{s}_j^*, \boldsymbol{\theta})]$ being the $mK \times mK$ corresponding covariance matrix. For an arbitrary site $\mathbf{s}_0$, spatial interpolation ("kriging") defines a spatial process $\tilde{\mathbf{w}}(\mathbf{s}_0) = E(\mathbf{w}(\mathbf{s}_0)|\mathbf{w}^*)$, called the *predictive process* that is derived from the *parent process* $\mathbf{w}(\mathbf{s})$. The realizations of $\tilde{\mathbf{w}}(\mathbf{s})$ are in fact optimal predictions conditional upon a realization of $\mathbf{w}(\mathbf{s})$ over S^* (see Banerjee et al., 2008, and references therein).

Replacing the original process $\mathbf{w}(\mathbf{s})$ by $\tilde{\mathbf{w}}(\mathbf{s})$, we approximate (8) by

$$\prod_k \left\{ \tilde{\mathcal{F}}_k(\mathbf{t}_1, \mathbf{t}_2, \cdots, \mathbf{t}_J) \right) \prod_{i=1}^{n_k} r(\mathbf{s}_{ki}) \exp(\beta_{0k} + \mathbf{z}(\mathbf{s}_{ki})'\boldsymbol{\beta}_k + \mathbf{v}_{ki}'\boldsymbol{\alpha}_k + (\mathbf{v}_{ki} \otimes \mathbf{z}(\mathbf{s}_{ki}))'\boldsymbol{\gamma}_k + \tilde{w}_k(\mathbf{s}_{ki})),$$

where $\tilde{\mathcal{F}}$ is obtained by replacing $\mathbf{w}$ in $\mathcal{F}$ by $\tilde{\mathbf{w}}$.

Combined with priors for regression parameters and spatial residual processes, we have a fully

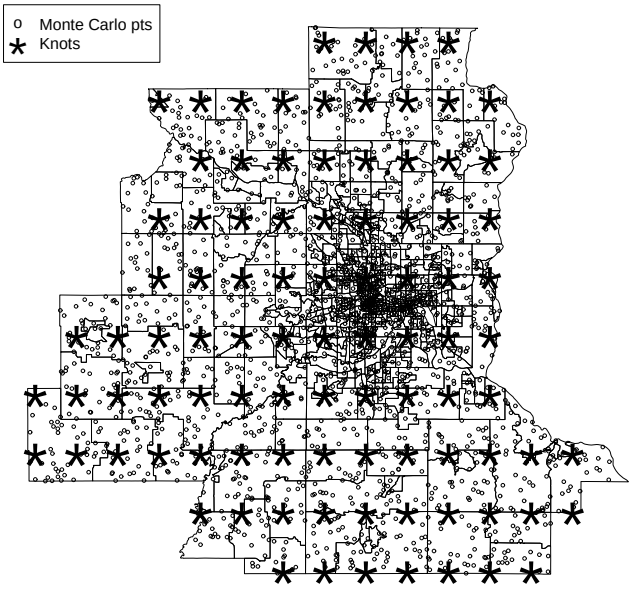

Figure 1: Monte Carlo points used for the integral approximation and predictive process knots.

specified Bayesian model. We obtain posterior samples under this model using a straightforward Markov chain Monte Carlo (MCMC) algorithm.

To implement a predictive process approximation for our dataset, we randomly generate points $\mathbf{s}_j^*$ in each census tract, proportional to its area, resulting in $J = 2129$ points in total with between 1 and 41 points in each census tract. To apply the predictive process approximation, we assume that $\mathbf{w}(\mathbf{s}) = (w_1(\mathbf{s}), w_2(\mathbf{s}))$ is a bivariate Gaussian process $MVGP(\mathbf{0}, \Gamma_{\mathbf{w}}(\cdot,\cdot,\boldsymbol{\theta}))$ with $\Gamma_{\mathbf{w}}(\cdot,\cdot,\boldsymbol{\theta})) = \varrho(\cdot,\phi)\Lambda$, a separable cross-covariance specification where $\Lambda = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$ and $\varrho(\cdot,\phi)$ is the univariate exponential correlation function. Hence ρ denotes dependence between the spatial residuals for $k = 1, 2$ in (5). Conditional on the coefficients, it measures the association between realizations of the two intensity surfaces at any location. We choose a regular grid with 93 knots $\mathbf{s}_j^*$ spread evenly across the region. Figure 1 shows the $J = 2129$ points and $m = 93$ grid knots.

Additional references

Cressie, N.A.C. (1993). *Statistics for Spatial Data*, 2nd ed. New York: Wiley.

Fuentes, M. (2007). Approximate likelihood for large irregularly spaced spatial data. *J. Amer. Statist. Assoc.*, **102**, 321–331.

Higdon, D. (2002). Space and space time modeling using process convolutions. In *Quantitative Methods for Current Environmental Issues*, C. Anderson, V. Barnett, P.C. Chatwin, and A.H. El-Shaarawi, eds., London: Springer-Verlag, pp. 37–56.

Jones, R.H. and Zhang, Y. (1997). Models for continuous stationary space-time processes. In *Modelling Longitudinal and Spatially Correlated Data: Methods, Applications, and Future Directions*, Lecture Notes in Statistics **122**, eds. T.G. Gregoire et al., Berlin: Springer-Verlag, pp. 289–298.

Paciorek, C.J. (2007). Computational techniques for spatial logistic regression with large datasets. To appear *Computational Statistics and Data Analysis*.

Paciorek, C.J. and Schervish, M.J. (2006). Spatial modelling using a New class of nonstationary covariance functions. *Environmetrics*, **17**, 483–506.

Rue, H. and Held, L. (2006). *Gaussian Markov Random Fields: Theory and Applications*. Boca Raton, FL: Chapman and Hall/CRC Press.

Rue, H. and Tjelmeland, H. (2002). Fitting Gaussian Markov random fields to Gaussian fields. *Scandinavian Journal of Statistics*, **29**, 31–49.

Stein, M.L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. New York: Springer.

Stein, M.L., Chi, Z. and Welty, L.J. (2004) Approximating likelihoods for large spatial datasets. *Journal of the Royal Statistical Society, Series B*, **66**, 275-296.

Vecchia, A.V. (1988). Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society Series B*, **50**, 297–312.

Ver Hoef, J. M., Cressie, N.A.C., and Barry, R. P. (2004). Flexible spatial models based on the fast Fourier transform (FFT) for cokriging. *J. Comp. Graph. Statist.*, **13**, 265–282.

Wikle, C. and Cressie, N. (1999). A dimension-reduced approach to space-time Kalman filtering. *Biometrika*, **86**, 815–829.

Xia, G. and Gelfand, A.E. (2006). Stationary process approximation for the analysis of large spatial datasets. Technical report, Department of Statistical Science, Duke University.