spmodel: Spatial Modeling in R – Supplementary Material

A Preprint

Michael Dumelle *
United States
Environmental Protection Agency
200 SW 35th St, Corvallis, OR, 97333
Dumelle.Michael@epa.gov

Matt Higham
Department of Math, Computer Science, and Statistics
St. Lawrence University
23 Romoda Drive, Canton, NY, 13617
mhigham@stlawu.edu

Jay M. Ver Hoef
National Oceanic and Atmospheric Administration
Alaska Fisheries Science Center
Marine Mammal Laboratory, Seattle, WA, 98115
jay.verhoef@noaa.gov

April 23, 2022

Abstract

Enter the text of your abstract here.

Keywords Spatial covariance · Linear Model · Autoregressive model

1 Estimation

1.1 Likelihood-based Estimation

Minus twice a profiled Gaussian log-likelihood, denoted $-2l(\boldsymbol{\theta}|\mathbf{y})$ is given by

$$-2l(\boldsymbol{\theta}|\mathbf{y}) = \ln|\boldsymbol{\Sigma}| + (\mathbf{y} - \mathbf{X}\tilde{\beta})^\intercal \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\beta}) + n \ln 2\pi, \tag{1}$$

where $\tilde{\boldsymbol{\beta}} = (\mathbf{X}^\intercal \boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^\intercal \boldsymbol{\Sigma}^{-1}\mathbf{y}$. Minimizing Equation 1 yields $\hat{\boldsymbol{\theta}}_{ml}$, the maximum likelihood estimates for $\boldsymbol{\theta}$. Then a closed for solution exists for $\hat{\boldsymbol{\beta}}_{ml}$, the maximum likelihood estimates for $\boldsymbol{\beta}$: $\hat{\boldsymbol{\beta}}_{ml} = \tilde{\boldsymbol{\beta}}_{ml}$, where $\tilde{\boldsymbol{\beta}}_{ml}$ is $\tilde{\boldsymbol{\beta}}$ evaluated at $\hat{\boldsymbol{\theta}}_{ml}$. Unfortunately $\hat{\boldsymbol{\theta}}_{ml}$ can be badly biased for $\boldsymbol{\theta}$ (especially for small sample sizes), which impacts the estimation of $\boldsymbol{\beta}$ (Patterson and Thompson 1971). This bias occurs due to the simultaneous estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ To reduce this bias, restricted maximum likelihood estimation (REML) emerged (Patterson and Thompson 1971; Harville 1977; Wolfinger, Tobias, and Sall 1994). It can be shown that integrating $\boldsymbol{\beta}$ out of a Gaussian likelihood yields the restricted Gaussian likelihood used in REML estimation. Minus twice a restricted Gaussian log-likelihood, denoted $-2l_R(\boldsymbol{\theta}|\mathbf{y})$ is given by

$$-2l_R(\boldsymbol{\theta}|\mathbf{y}) = -2l(\boldsymbol{\theta}|\mathbf{y}) + \ln|\mathbf{X}^\intercal \boldsymbol{\Sigma}^{-1}\mathbf{X}| - p \ln 2\pi. \tag{2}$$

Minimizing Equation 2 yields $\hat{\boldsymbol{\theta}}_{reml}$, the restricted maximum likelihood estimates for $\boldsymbol{\theta}$. Then a closed for solution exists for $\hat{\boldsymbol{\beta}}_{reml}$, the restricted maximum likelihood estimates for $\boldsymbol{\beta}$: $\hat{\boldsymbol{\beta}}_{reml} = \tilde{\boldsymbol{\beta}}_{reml}$, where $\tilde{\boldsymbol{\beta}}_{reml}$ is $\tilde{\boldsymbol{\beta}}$ evaluated at $\hat{\boldsymbol{\theta}}_{reml}$.

When all variance parameters are unknown, the overall variance, σ^2, can be profiled out of Equation 1 and Equation 2. This reduces the number of parameters requiring optimization by one, which can dramatically

*Corresponding Author

reduce estimation time. For example, profiling σ^2 out of Equation 1 yields

$$-2l^*(\boldsymbol{\theta}^*|\mathbf{y}) = \ln|\boldsymbol{\Sigma}^*| + n\ln[(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})^\intercal \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})] + n + n\ln 2\pi/n. \tag{3}$$

After finding $\hat{\boldsymbol{\theta}}^*_{ml}$ a closed form solution for $\hat{\sigma}^2_{ml}$ exists: $\hat{\sigma}^2_{ml} = [(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})^\intercal \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})]/n$. Then $\hat{\boldsymbol{\theta}}^*_{ml}$ is combined with $\hat{\sigma}^2_{ml}$ to yield $\hat{\boldsymbol{\theta}}_{ml}$ and subsequently $\hat{\boldsymbol{\beta}}_{ml}$

Next describe REML adjustments

1.2 Semivariogram-based Estimation

1.2.1 Weighted Least Squares

(Cressie 1985, 1993)

1.2.2 Composite Likelihood

(Curriero and Lele 1999)

2 Hypothesis Testing

2.1 The General Linear Hypothesis Test

2.2 Contrasts

3 Random Effects

3.1 BLUPs

References

Cressie, Noel. 1985. "Fitting Variogram Models by Weighted Least Squares." *Journal of the International Association for Mathematical Geology* 17 (5): 563–86.

———. 1993. *Statistics for Spatial Data*. John Wiley & Sons.

Curriero, Frank C, and Subhash Lele. 1999. "A Composite Likelihood Approach to Semivariogram Estimation." *Journal of Agricultural, Biological, and Environmental Statistics*, 9–28.

Harville, David A. 1977. "Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems." *Journal of the American Statistical Association* 72 (358): 320–38.

Patterson, H Desmond, and Robin Thompson. 1971. "Recovery of Inter-Block Information When Block Sizes Are Unequal." *Biometrika* 58 (3): 545–54.

Wolfinger, Russ, Randy Tobias, and John Sall. 1994. "Computing Gaussian Likelihoods and Their Derivatives for General Linear Mixed Models." *SIAM Journal on Scientific Computing* 15 (6): 1294–1310.