Bayesian Empirical Likelihood Spatial Model applying Leroux Structure
Farzana Jahan and Kerrie Mengersen
Queensland University of Technology

Introduction

- Bayesian Empirical Likelihood (BEL) methods have been applied to spatial data analysis for small area estimation (SAE) by extending the Fay-Herriot (FH) model [1] in recent times.
- The present study is an attempt to extend a BEL spatial model utilising the Leroux prior [2].
- Bayesian Parametric Hierarchical Spatial Model-
  - Assumes a parametric distribution of data in small areas, for example, for counts of disease incidence, \( Y_i \sim \text{Poisson}(\mu_i E_i) \).
  - Model \( \mu_i \) using covariate effect (fixed effect) and a spatial random effects \( \mu_i = \log(E_i) + x_i' \beta + \psi_i \).
- Bayesian Semiparametric Spatial Model-
  - Does not assume any parametric distribution for the data.
  - Calculates Empirical likelihood weights from the data and replace the data likelihood part in Bayesian hierarchical spatial models framework.

Bayesian Empirical Likelihood (BEL)

- Empirical Likelihood (EL) [3] can be used in place of a density and, when multiplied by prior can yield the posterior distribution of interest [4].
- For data points \( y_1, y_2, \ldots, y_n \) from some unknown distribution \( F \), if the parameter of interest \( \theta(F) \) can be determined by the estimating equation \( f(y_i, \theta) \), then the empirical likelihood ratio function can be defined as:
  \[
  W_n(\theta) = 2 \sum_{i=1}^{n} \log \left( 1 + \lambda f(y_i, \theta) \right)
  \]
  where the vector \( \lambda \) satisfies
  \[
  \sum_{i=1}^{n} \frac{f(y_i, \theta)}{1 + \lambda f(y_i, \theta)} = 0.
  \]

BEL for Spatial Analysis in literature

- The FH model for small area estimation can be written as:
  \[
  Y_i = \mu_i + \epsilon_i,
  \mu_i = X_i' \beta + \psi_i
  \]
- \( Y_i \) is a design unbiased estimate of \( \mu_i, \epsilon_i \) is an unstructured error component, \( X_i \) is the vector of covariate information for area \( i \), \( \beta \) is the vector of fixed covariate effects and \( \psi_i \) is a vector of spatially referenced random effects.
- For estimation, BEL estimating equations were used as [5, 6] are:
  \[
  \sum_{i=1}^{n} w_i(y_i - \mu_i) = 0
  \]
- EL weights for the data have to satisfy the constraint:
  \[
  W_n = \{ \sum_{i=1}^{n} w_i = 1; w_i > 0 \forall i; \sum_{i=1}^{n} w_i \mu_i(y_i, \mu_i) = 0 \forall j \}
  \]

Leroux structure Prior

- The spatial random effect \( \psi_i \) under Leroux specification [2] is distributed as:
  \[
  \psi_i \sim \text{MVN}(0, \sigma^2 D)
  \]
- The proposed generalised inverse of \( D, D^* \) as in [2]:
  \[
  D^* = (1 - \rho) D + \rho R
  \]
- \( R \) is the intrinsic Autoregressive matrix with diagonal elements, \( R_{ii} \), denoting number of neighbours of \( i \)th area and off diagonal elements taking value 1 if area \( i \) and \( j \) are adjacent, otherwise 0.
- \( \rho \in [0, 1] \) is the parameter denoting spatial dependence.

Bayesian Semiparametric Leroux Empirical Likelihood Model

- The existing semiparametric spatial BEL model by specifying the prior of \( \psi_i \) as a Leroux structure prior as:
  \[
  \psi_i \sim \text{MVN}(0, \sigma^2 D)
  \]
  where \( \sigma^2 \) is the overall variance , \( D \) is a singular covariance matrix.

Sampling from proposed BSLEL model

An MCMC sampling algorithm applying a random walk Metropolis Hasting (MH) and EL estimating equations has been developed for drawing the posterior samples for the parameters of interest.

1. Obtaining starting values: Using Maximum Empirical Likelihood Estimate (MEL) of \( \beta \). Empirical likelihood weights \( W_i \) for the data is calculated by considering \( \psi = \theta \) in \( \mu_i = \log(E_i) + x_i' \beta + \psi_i \).

2. Sampling spatial random effects, \( \psi_i \): To sample \( \psi_i \) we use MCMC by blocks for generating EL weights \( W_i \), performing a Metropolis-Hastings step with the following posterior density ratio if the constraint (11) is satisfied,

3. Sampling fixed effects, \( \beta \): Using a multivariate normal proposal, \( \beta \sim \text{MVN}(\hat{\beta}_{BSHEL}, V) \). with these proposal covariance tuned on the basis of the path chains, \( \beta \) is sampled using a random-walk MH step with posterior density ratio [6],

4. Sampling precision parameter \( \tau \). as will be sampled with a Gaussian proposal as \( \tau^\ast \sim \mathcal{N}(\tau, \Sigma_{\tau}) \), a random walk MH sample will be drawn with posterior density ratio

5. The steps 4-5 are repeated until convergence.

Application to Simulated Cancer Incidence Data

- The response variable \( Y_i \) denotes the number of lung cancer incidence amongst females in \( i \)th area, \( i = 1, 2, \ldots, 2147 \).
- The simulated cancer incidence data (Cancer Council Queensland) is expected to represent the original incidence in small areas across Australia.
- The covariate \( X_i \) is 1, if \( i \)th area is a major city and 0 if \( i \)th area is not major city (regional or remote) [7].
- Fixed value for the spatial dependence parameter \( \rho \) is used ( \( \rho = 0.99 \)).

Table 1: Mean squared prediction error (MSPE) for five leave-one-out validation for parametric and semiparametric models

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MSPE</td>
<td>10.71916</td>
<td>10.93718</td>
<td>12.937</td>
<td>19.485</td>
</tr>
</tbody>
</table>

Raw and fitted Standard Incidence Ratios for Lung Cancer (females)

<table>
<thead>
<tr>
<th>Raw SIR</th>
<th>Modelled SIR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Conclusions

- The proposed BSLEL performed similarly with BSHEL model, but provided smaller MSPE, i.e., better prediction.
- To reduce the computational time, more effective algorithms may be developed using Hamiltonian Monte Carlo [8].
- More specific covariate information to improve the prediction will be considered, rather than using a single binary covariate.
- The spatial dependence parameter \( \rho \) will be estimated simultaneously in the MCMC.

Key References