Bayesian Geostatistics

Sudipto Banerjee

Spring, 2021

Department of Biostatistics, Fielding School of Public Health, University of California, Los Angeles.

Bayesian spatial random effect models

- Continuous data:

$$
\begin{gathered}
y(s_i)\,|\,\mu(s_i),\ \tau^2 \overset{ind}{\sim} N(\mu(s_i),\tau^2)\ ;\quad i=1,2,\ldots,n\ ;\\
\mu(s_i)=\beta_0+\beta_1 x_1(s_i)+\beta_2 x_2(s_i)+\cdots+\beta_p x_p(s_i)+w(s_i)\ ;\\
\beta_j \overset{ind}{\sim} N(0,\sigma^2_\beta)\ ;\quad j=0,1,\ldots,p\ ;\\
w=(w(s_1),w(s_2),\ldots,w(s_n))^\top \sim N(0,\sigma^2 R_w(\phi))\ ;\\
1/\tau^2 \sim \mathrm{Gamma}(a_\tau,b_\tau)\ ;\quad 1/\sigma^2 \sim \mathrm{Gamma}(a_\sigma,b_\sigma)\ ;\\
\phi \sim \mathrm{Unif}(a_\phi,b_\phi)\ .
\end{gathered}
$$

- $R_w(\phi)$ is $n\times n$ spatial correlation matrix.

Bayesian spatial random effect models

- Count data:

$$
\begin{aligned}
y(s_i) &\sim Poi(\lambda(s_i)) \; ; \quad i = 1, 2, \ldots, n \; ; \\
\log \lambda(s_i) &= \beta_0 + \beta_1 x_1(s_i) + \beta_2 x_2(s_i) + \cdots + \beta_p x_p(s_i) + w(s_i) \; ; \\
\beta_j &\overset{ind}{\sim} N(0, \sigma_\beta^2) \; ; \quad j = 0, 1, \ldots, p \; ; \quad w \sim N(0, \sigma^2 R_w) \; ; \\
& 1/\sigma^2 \sim \mathrm{Gamma}(a_\sigma, b_\sigma) \; ; \quad \phi \sim \mathrm{Unif}(a_\phi, b_\phi) \; .
\end{aligned}
$$

- $R_w(\phi)$ is $n \times n$ spatial correlation matrix.

Bayesian spatial random effect models

- Binary data:

$$
\begin{aligned}
y(s_i) \sim Ber(p(s_i))\ ; \quad i = 1, 2, \ldots, n\ ; \\
\log\left(\frac{p(s_i)}{1 - p(s_i)}\right) = \beta_0 + \beta_1 x_1(s_i) + \beta_2 x_2(s_i) + \cdots + \beta_p x_p(s_i) + w(s_i)\ ; \\
\beta_j \overset{ind}{\sim} N(0, \sigma_\beta^2)\ ; \quad j = 0, 1, \ldots, p\ ; \quad w \sim N(0, \sigma^2 R_w)\ ; \\
1/\sigma^2 \sim \text{Gamma}(a_\sigma, b_\sigma)\ ; \quad \phi \sim \text{Unif}(a_\phi, b_\phi)\ .
\end{aligned}
$$

- $R_w(\phi)$ is $n \times n$ spatial correlation matrix.

Spatial random effects: Gaussian process

- We say that $w(s) \sim GP(0, \sigma^2 \rho(\cdot))$:

$$w = (w(s_1), w(s_2), \ldots, w(s_n))^\top \sim N(0, \sigma^2 R_w) \ ;$$

- R_w is $n \times n$ spatial correlation matrix:

$$R_w[i,j] = \rho(s_i, s_j) \ .$$

- The correlation function is parametrized to capture strength of association as a function of distance. Practical choice (works well for a variety of situations):

$$\rho(s_i, s_j) = \exp(-\phi \|s_i - s_j\|) \ .$$

Bayesian inference for continuous spatial data

- Step-I: Estimate parameters (MCMC) by sampling from

$$[\beta, w, \tau^2, \sigma^2, \phi \,|\, y, X]$$

- Step-II: Estimate the latent process $w(s_0)$ at new location s_0 by sampling from

$$[w(s_0) \,|\, w, \sigma^2, \phi]$$

for each sampled value of w, σ^2 and ϕ obtained in Step-I.

- Step III: Obtain posterior samples of $\mu(s_0)$:

$$\mu(s_0) = \beta_0 + \beta_1 x_1(s_0) + \beta_2 x_2(s_0) + \cdots + \beta_p x_p(s_0) + w(s_0)\ .$$

- Step-IV: Predict $y(s_0)$ by drawing its value from

$$N(\mu(s_0), \tau^2)$$

for each sampled $\mu(s_0)$ (from Step-III) and τ^2 (from Step-I).

Bayesian inference for spatial count data

- Step-I: Estimate parameters (MCMC) by sampling from

$$[\beta, w, \sigma^2, \phi \,|\, y, X]$$

- Step-II: Estimate the latent process $w(s_0)$ at new location s_0 by drawing from

$$[w(s_0) \,|\, w, \sigma^2, \phi]$$

  for each sampled value of w, σ^2 and ϕ obtained in Step-I.
- Step III: Obtain posterior samples of $\lambda(s_0)$:

$$\lambda(s_0) = \exp\left(\beta_0 + \beta_1 x_1(s_0) + \beta_2 x_2(s_0) + \cdots + \beta_p x_p(s_0) + w(s_0)\right) \; .$$

- Step-IV: Predict $y(s_0)$ by drawing its value from

$$Poi(\lambda(s_0))$$

  for each sampled $\lambda(s_0)$ in Step-III.

Bayesian inference for binary count data

- Step-I: Estimate parameters (MCMC) by sampling from

$$[\beta, w, \sigma^2, \phi \mid y, X]$$

- Step-II: Estimate the latent process $w(s_0)$ at new location s_0 by drawing from

$$[w(s_0) \mid w, \sigma^2, \phi]$$

  for each sampled value of w, σ^2 and ϕ obtained in Step-I.
- Step III: Obtain posterior samples of $p(s_0)$:

  $$p(s_0) = \text{logit}^{-1}\left(\beta_0 + \beta_1 x_1(s_0) + \beta_2 x_2(s_0) + \cdots + \beta_p x_p(s_0) + w(s_0)\right) .$$

- Step-IV: Predict $y(s_0)$ by drawing its value from

$$Ber(p(s_0))$$

  for each sampled $p(s_0)$ in Step-III.

Setting Priors

- For regression slopes we customarily assign non-informative priors.
- For the variance component σ^2 (partial sill), we customarily choose an Inverse-Gamma (or, equivalently, Gamma prior for $1/\sigma^2$)—the shape parameter is taken to be 2 and the scale parameter is chosen so that the prior mean is equal to the scale parameter. This value can be set from an exploratory variogram analysis. Strategy for τ^2 (nugget) is similar.
- For the range parameter ϕ, we usually set it so that the effective range (distance where spatial correlation drops to 0.05) is between some small number and does not exceed about 50% of the maximum inter-site distance. For example, with the exponential correlation function we solve $\rho(d; \phi) = 0.05$ and see that $\phi \approx 3/d$, where d is the effective spatial range. We bound $d \in (d_{\text{min}}, d_{\text{max}})$ and this suggests $\phi \sim \text{Unif}(3/d_{\text{max}}, 3/d_{\text{max}})$.