Bayesian inference for spatial GP models

Abhi Datta

February 22, 2018

Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland

Review of last lecture

- Bayesian Principles – Bayes theorem, posterior inference, credible intervals
- Bayesian Linear model
- Conjugate Normal-Inverse Gamma priors for (β, σ^2)
- Sampling based inference – Monte Carlo Integration
- Composition sampling – Scope and limitations

Bayesian inference for spatial linear model

- $y(s) = x(s)'\beta + w(s) + \epsilon(s)$, $w(s) \sim GP(0, C(\cdot,\cdot\,|\,\phi))$, $\epsilon \stackrel{\text{iid}}{\sim} N(0, \tau^2)$
- For n locations, unmarginalized model: $y \sim N(X\beta + w, \tau^2 I)$, $w \sim N(0, \sigma^2 R(\phi))$
- Marginalized model: $y \sim N(X\beta, \sigma^2 R(\phi) + \tau^2 I)$
- Assume ϕ is known, $\sigma^2 \sim IG(a_\sigma, b_\sigma)$, $\tau^2 \sim IG(a_\tau, b_\tau)$ and $\beta \sim N(\mu, V)$
- Composition sampling does not help with either of the models
- How to do Bayesian inference ?

Unmarginalized model

- Likelihood: $N(y \,|\, X\beta + w, \tau^2 I) \times N(w \,|\, 0, \sigma^2 R(\phi) \times N(\beta \,|\, \mu, V) \times IG(\sigma^2 \,|\, a_\sigma, b_\sigma) \times IG(\tau^2 \,|\, a_\tau, b_\tau)$

Unmarginalized model

- Likelihood: $N(y \mid X\beta + w, \tau^2 I) \times N(w \mid 0, \sigma^2 R(\phi) \times N(\beta \mid \mu, V) \times IG(\sigma^2 \mid a_\sigma, b_\sigma) \times IG(\tau^2 \mid a_\tau, b_\tau)$
- Observe that
    - $\beta \mid \sigma^2, \tau^2, w, y \sim N(\mu^*, V^*)$
    - $w \mid \sigma^2, \tau^2, \beta, y \sim N(m, C^*)$
    - $\sigma^2 \mid \beta, \tau^2, w, y \sim IG(a_\sigma^*, b_\sigma^*)$
    - $\tau^2 \mid \beta, \sigma^2, w, y \sim IG(a_\tau^*, b_\tau^*)$
- Can we use these nice *full conditionals* to obtain posterior inference?

Unmarginalized model

- Likelihood: $N(y \mid X\beta + w, \tau^2 I) \times N(w \mid 0, \sigma^2 R(\phi) \times N(\beta \mid \mu, V) \times IG(\sigma^2 \mid a_\sigma, b_\sigma) \times IG(\tau^2 \mid a_\tau, b_\tau)$
- Observe that
  - $\beta \mid \sigma^2, \tau^2, w, y \sim N(\mu^*, V^*)$
  - $w \mid \sigma^2, \tau^2, \beta, y \sim N(m, C^*)$
  - $\sigma^2 \mid \beta, \tau^2, w, y \sim IG(a_\sigma^*, b_\sigma^*)$
  - $\tau^2 \mid \beta, \sigma^2, w, y \sim IG(a_\tau^*, b_\tau^*)$
- Can we use these nice *full conditionals* to obtain posterior inference?
- Yes! Via Gibbs sampling

Gibbs sampling

- Suppose that $\theta = (\theta_1, \theta_2)$ and we seek the posterior distribution $p(\theta_1, \theta_2 \mid y)$.

- For many interesting hierarchical models, we have access to *full conditional distributions* $p(\theta_1 \mid \theta_2, y)$ and $p(\theta_2 \mid \theta_1, y)$.

- The *Gibbs sampler* proposes the following sampling scheme. Set starting values $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)})$ For $j = 1, \ldots, M$
    - Draw $\theta_1^{(j)} \sim p(\theta_1 \mid \theta_2^{(j-1)}, y)$
    - Draw $\theta_2^{(j)} \sim p(\theta_2 \mid \theta_1^{(j)}, y)$

- This constructs a *Markov Chain* and, after an initial "burn-in" period when the chains are trying to find their way, $\{\theta_1^{(j)}, \theta_2^{(j)}\}_{j=M_0+1}^{M}$ will be *Markov Chain Monte Carlo (MCMC)* samples from $p(\theta_1, \theta_2 \mid y)$, where M_0 is the burn-in period..

Gibbs sampling

- More generally, if $\theta = (\theta_1, \ldots, \theta_p)$ are the parameters in our model, we provide a set of initial values $\theta^{(0)} = (\theta_1^{(0)}, \ldots, \theta_p^{(0)})$ and then performs the j-th iteration, say for $j = 1, \ldots, M$, by updating successively from the *full conditional* distributions:

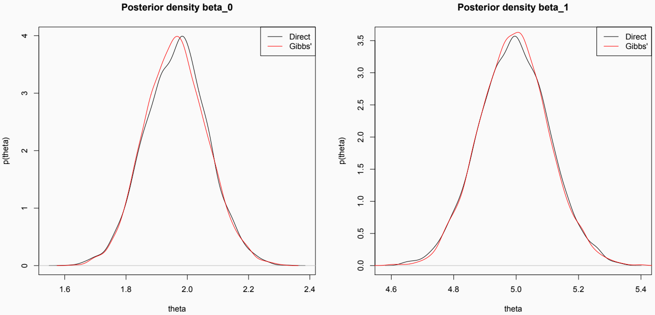
$$
\begin{array}{l}
\theta_1^{(j)} \sim p(\theta_1^{(j)} \mid \theta_2^{(j-1)}, \ldots, \theta_p^{(j-1)}, y) \\
\theta_2^{(j)} \sim p(\theta_2 \mid \theta_1^{(j)}, \theta_3^{(j-1)}, \ldots, \theta_p^{(j-1)}, y) \\
\ldots \\
\text{(the generic } k^{th} \text{ element)} \\
\theta_k^{(j)} \sim p(\theta_k | \theta_1^{(j)}, \ldots, \theta_{k-1}^{(j)}, \theta_{k+1}^{(j-1)}, \ldots, \theta_p^{(j-1)}, y) \\
\ldots \\
\theta_p^{(j)} \sim p(\theta_p \mid \theta_1^{(j)}, \ldots, \theta_{p-1}^{(j)}, y)
\end{array}
$$

Example

- $Y_i \overset{ind}{\sim} N(\beta_0 + \beta_1 X_i, 1)$ for $i = 1, \ldots, n$ where $\beta_0 = 2$, $\beta_1 = 5$ (both unknown) and $n = 100$
- We assume independent priors $\beta_0 \sim N(0, \gamma_0)$ and $\beta_1 \sim N(0, \gamma_1)$ where $\gamma_0 = 100$ and $\gamma_1 = 10$
- Gibbs sampling (Gelfand and Smith, 1990):
  - $\beta_0 \,|\, \beta_1, Y \sim N(1'(Y - \beta_1 X)/(n + 1/\gamma_0), 1/(n + 1/\gamma_0))$
  - $\beta_1 \,|\, \beta_0, Y \sim N(X'(Y - \beta_0 1)/(\sum_{i=1}^n X_i^2 + 1/\gamma_1), 1/(\sum_{i=1}^n X_i^2 + 1/\gamma_1))$
- Direct approach: $(\beta_0, \beta_1 \mid Y) \sim N\left(V_\beta (1, X)' Y, V_\beta\right)$ where $V_\beta = ((1, X)'(1, X) + diag(1/\gamma_0, 1/\gamma_1))^{-1}$

Example

- $Y_i \overset{ind}{\sim} N(\beta_0 + \beta_1 X_i, 1)$ for $i = 1, \ldots, n$ where $\beta_0 = 2$, $\beta_1 = 5$ (both unknown) and $n = 100$
- We assume independent priors $\beta_0 \sim N(0, \gamma_0)$ and $\beta_1 \sim N(0, \gamma_1)$ where $\gamma_0 = 100$ and $\gamma_1 = 10$

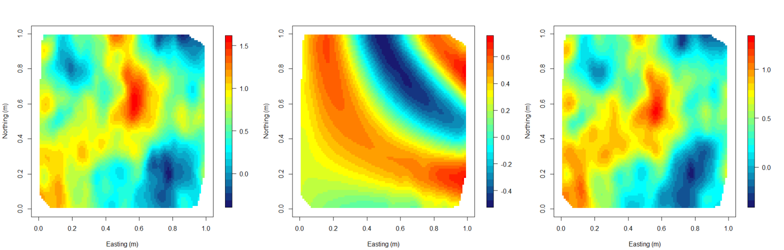


Block Gibbs update

- Recall for the spatial model we had full conditionals for vectors β and w
- Let $\theta = (\theta_1, \theta_2, \ldots, \theta_k) = (\eta_1', \eta_2', \ldots, \eta_m')$ where η_j are blocks of θ_i's
- One can use the Gibbs updates for the blocks η_j's instead of using the individual updates for θ_i
- In many models, the block full conditionals are easier to obtain, substantially reduces computation and improves rate of convergence

Data analysis

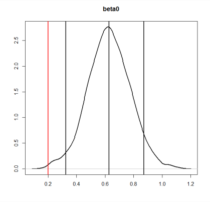
- Dataset 3 from Lecture 1
- True model: $y(s) \sim N(0.2 - 0.3x(s) + w(s), 0.01)$,
  $w(s) \sim GP,\ Cov(w(s_i), w(s_j)) = 0.25 * exp(-2||s_i - s_j||)$



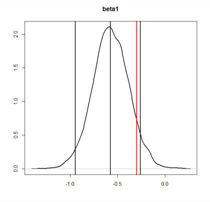
$y(s)$ $x(s)$ $w(s)$

Parameter posteriors

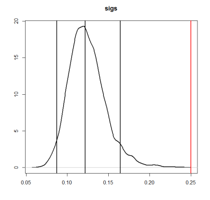
- ϕ is kept fixed at 4.23 (estimated value from variogram fitting)
- Gibbs sampler for w, β, σ^2 and τ^2



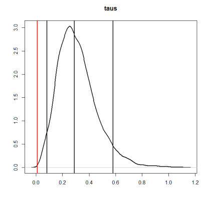
β_0



β_1



σ^2



τ^2

References

- Expository article on Gibbs sampler: Casella, G. and George, E.I. (1992), Explaining the Gibbs Sampler, The American Statistician, 46, 167-174.
- Gelfand, A., and Adrian F. M. Smith. (1990). *Sampling-Based Approaches to Calculating Marginal Densities.* Journal of the American Statistical Association, 85(410), 398–409.