Matrix–free conditional simulations of GMRF

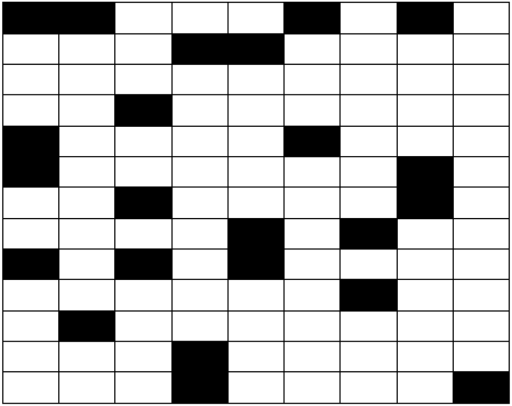
Somak Dutta
Joint Work with Debashis Mondal.

University of Chicago

June 24, 2014

Data on a regular grid.

Image of an dummy array of plot $\mathcal{D}_1$.
Black $=$ missing observations.



Mixed linear model.

$$\mathbf{y} = \mathbf{T}\boldsymbol{\tau} + \mathbf{F}\mathbf{x} + \boldsymbol{\epsilon}.$$

Array dimension = $r \times c$. (VERY LARGE).

$\mathbf{y} =$ $n \times 1$ response vector.

$\boldsymbol{\tau} =$ $m \times 1$ vector of fixed effects.

$\mathbf{T} =$ $n \times m$ known design matrix.

$\mathbf{x} =$ $rc \times 1$ vector of underlying spatial random field.

$\mathbf{F} =$ known sparse incidence matrix - $\mathbf{F}\mathbf{x}$ gives back the values of the spatial random field on n observed plots.

$\boldsymbol{\epsilon} \sim N(\mathbf{0}, \lambda_y^{-1}\mathbf{I}_n)$: nugget effects.

Intrinsic auto-regression model for $\mathbf{x}$.

$$\mathbf{y} = \mathbf{T}\boldsymbol{\tau} + \mathbf{F}\mathbf{x} + \epsilon.$$

- $\mathbf{x}$ is Gaussian with sparse singular precision matrix $\mathbf{W}$,

$$\mathbf{x}^T\mathbf{W}\mathbf{x} \;=\; \lambda_{10}\sum\sum(x_{i,j}-x_{i-1,j})^2+\lambda_{01}\sum\sum(x_{i,j}-x_{i,j-1})^2.$$

- $\mathbf{W}$ has analytically known spectral decomposition

$$\mathbf{W} = \mathbf{P}(\lambda_{01}\mathbf{D}_{01} + \lambda_{10}\mathbf{D}_{10})\mathbf{P}^T.$$

- $\mathbf{P}$ correspond to the two dimensional discrete cosine transformation.

Conditional simulation.

Interested in sampling from:

$$\mathbf{x}|\mathbf{y} \;\sim\; N(\lambda_y \mathbf{A}^{-1}\mathbf{F}^T(\mathbf{y}-\mathbf{T}\boldsymbol{\tau})\;,\;\mathbf{A}^{-1}), \quad \mathbf{A} = \lambda_y \mathbf{F}^T\mathbf{F} + \mathbf{W}.$$

- Step 1: First draw $z \sim N(\mathbf{0}, \mathbf{I})$.
- Traditional way: Compute Cholesky factor $\mathbf{L}$ such that $\mathbf{L}\mathbf{L}^T = \mathbf{A}$. And let $\mathbf{x} = \mathbf{L}^{-\mathbf{1}}\mathbf{z}$.
- Costs: memory $= O((rc)^{1.5})$, $\#\text{FLOPs} = O((rc)^2)$.
- We will create algorithm that has costs:
  memory $= O(rc)$, $\#\text{FLOPs} = O(rc\log rc)$

An "exact" method

- $\mathbf{x}|\mathbf{y} \;\sim\; N(\mathbf{A}^{-1}(\mathbf{y}-\mathbf{T}\boldsymbol{\tau})\;,\;\mathbf{A}^{-1}), \quad \mathbf{A} = \lambda_y \mathbf{F}^T\mathbf{F} + \mathbf{W}$
- Analytically known spectral decomposition: $\mathbf{W} = \mathbf{P}\mathbf{D}\mathbf{P}^T$.
- Square root of $\mathbf{A}$:

$$\mathbf{S} = [\lambda_y^{\frac{1}{2}}\mathbf{F}^T \quad \mathbf{P}\mathbf{D}^{1/2}] \qquad \text{then } \mathbf{S}\mathbf{S}^T = \mathbf{A}$$

Simulation algorithm:

- Strike 1: First draw $z \sim N(\mathbf{0}, \mathbf{I})$.
- Strike 2: Sample with $\mathbf{A}$ as the covariance matrix

$$\mathbf{b} = \mathbf{S}\mathbf{z} + \lambda_y(\mathbf{y}-\mathbf{T}\boldsymbol{\tau}) \;\sim\; N(\lambda_{\mathbf{y}}(\mathbf{y}-\mathbf{T}\boldsymbol{\tau}),\; \mathbf{A})$$

- Strike 3: Solve $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} \;\sim\; N(\mathbf{A}^{-1}(\mathbf{y}-\mathbf{T}\boldsymbol{\tau})\;,\;\mathbf{A}^{-1})$

Lanczos algorithm and Incomplete Cholesky Preconditioner

To solve:

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

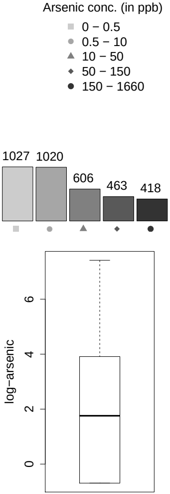
using Lanczos algorithm (Dutta and Mondal, 2012).

- Condition number of $\mathbf{A} \to \infty$.
- $\mathbf{L} =$ incomplete Cholesky factorization (lower triangular):

$$\mathbf{L}\mathbf{L}^T \approx \mathbf{A} \Rightarrow \mathbf{L}^{-1}\mathbf{A}\mathbf{L}^{-T} \approx \mathbf{I}.$$

- solve $\mathbf{L}^{-1}\mathbf{A}\mathbf{L}^{-T}\mathbf{x}_1 = \mathbf{L}^{-1}\mathbf{b}$, then $\mathbf{x} = \mathbf{L}^{-T}\mathbf{x}_1$.
- Geometric convergence of Lanczos algo in $O(\log rc)$ iterations.

Arsenic concentration in Bangladesh (Dutta and Mondal, 2013).



Embed the data in a 500 x 300 array.

Application: Maximal simultaneous exceedance region

- D is a 90% exceedance region of $\mathbf{x}$ for a given threshold c

$$P(x_{i,j} \geq c,\ \forall (i,j) \in D \mid \mathbf{y}) \geq 90\%$$

- Finding the largest such set is not possible (NP hard?).
- Put a constraint: highest marginal exceedance probabilities

$$P(x_{i,j} \geq c \mid \mathbf{y}) \ \geq \ P(x_{i',j'} \geq c \mid \mathbf{y})$$

$$\forall (i,j) \in D \quad \text{and} \quad (i',j') \notin D.$$

- Can be thought as a highest probability density simultaneous exceedance region parallel to the Bayesian highest posterior density credible region.
- But still cannot be computed analytically.

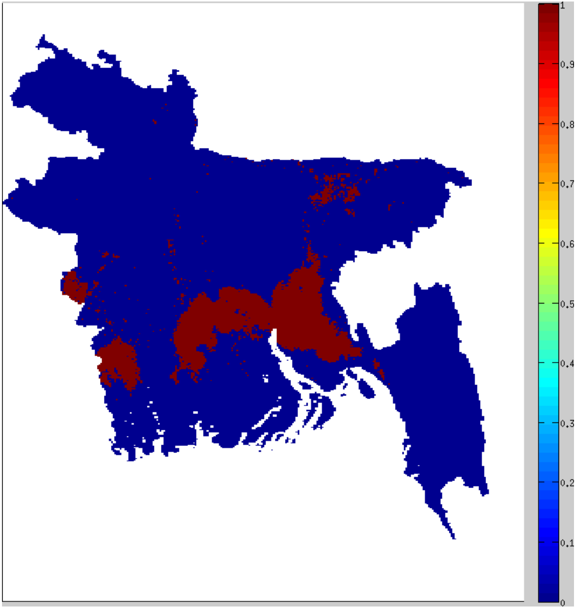
Simulating maximal simultaneous exceedance regions

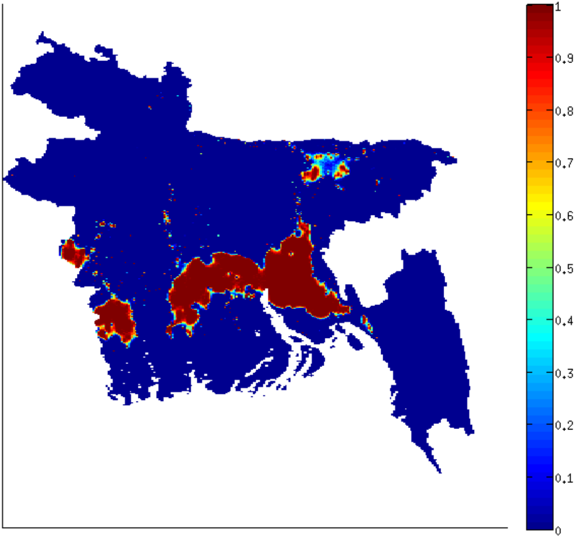
Step 1. Rank the locations:

- Draw an ensemble of realizations $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}$ of size N from $p(\mathbf{x}|\mathbf{y})$.
- Compute marginal exceedance probabilities.
- Rank the locations according to decreasing marginal exceedance probabilities.

Step 2. Compute the exeedance region.

- Starting from the top location keep on adding locations until the simultaneous exceedance probability falls below 90%.


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

References

1. Dutta and Mondal (2012) An h-likelihood method for spatial mixed linear model based on intrinsic autoregressions. *JRSS-B, Forthcoming*.

2. Dutta and Mondal (2013) REML Analysis for Spatial Mixed Linear Models Based on Approximate Intrinsic Matérn Dependence with Nugget Effects. *Submitted.*

3. Dutta and Mondal (2014) Matrix–free conditional simulations of Gaussian Markov random fields and their applications. *Preprint*.

Various details

- Latitude: 20 – 27 North, Longitude: 88 – 93 East.
- Area of each rectangular cell: 2.64 square kilometers.
- Embedded in 500×300 array
- Estimates: $\widehat{\lambda}_y = 4.72(0.02), \widehat{\lambda}_{01} = 3.14(0.05)$ and $\widehat{\lambda}_{10} = 1.17(0.13)$.