Covariance functions

SAMSI Summer School August, 2009

Douglas Nychka,
National Center for Atmospheric Research
Outline

- Restricted maximum likelihood
- Tapering a covariance
The spatial model

Given $n$ pairs of observations $(x_i, y_i), i = 1, \ldots, n$

$$y_i = [Td]_i + h(x_i) + \epsilon_i$$

$\epsilon_i$’s are random errors.

$\epsilon$ are $MN(0, \sigma^2 I)$

$h(x)$ is a realization of a Gaussian process
$E(h(x)) = 0$ and $\text{COV}(h(x), h(x')) = \rho K(\theta)(x, x')$

In vector notation:

$$y = Td + h + \epsilon$$

$\rho K(\theta)$ is the covariance matrix for $h$. 
The likelihood

- 2 log likelihood:

\[(y - Td)^T(M)^{-1}(y - Td) + \log(|M|) + c\]

\[M = \rho K(\theta) + \sigma^2 I \text{ and } c \text{ is a constant.}\]

Eliminate fixed part

Choose \(U\) an orthogonal matrix so that \(UTT = 0\)

\[UTy = UTh + UT\epsilon\]

Restricted likelihood

\[(y^*)^T(M^*)^{-1}(y^*) + \log(|M^*|) + c\]

\[y^* = UTy \text{ and } M^* = UTMU\]
Linear algebra for the likelihood

*Cholesky decomposition of $M^*$:* 

$M^* = C^T C$ where $C$ is an upper triangular matrix.

**Quadratic Form:**

$$(y^*)^T (M^*)^{-1}(y^*) = (y^*)^T (C^T C)^{-1}(y^*)$$

$$= (y^*)^T C^{-1} C^{-T} (y^*) = ||C^{-T} y^*||^2$$

solve the linear system $C^T w = y^*$. 
so $(y^*)^T (M^*)^{-1}(y^*) = ||w||^2$

*For the determinant:* 

$|M^*| = |C^T C| = |C|^2$

Determinant of an upper triangular matrix is the product of the diagonal elements.
Maximizing the likelihood

- All these computations need to be done for a single evaluation. and are order $n^3$. Although the Cholesky decomposition and solution are faster than a full matrix inversion.

- Difficult to find the derivatives of the likelihood although it can be concentrated about $\sigma$ or $\rho$

Maximum likelihood estimates are usually not feasible beyond about 1000 observations.
Alternatives:

- If replicates are available can fit a correlogram using nonlinear regression. This tends to have more variance.

- Use a compact covariance or tapering to make M a sparse matrix.
Tapering

- 2 log likelihood:

\[(y - Td)^T (M)^{-1} (y - Td) + \log(|M|) + c\]

\[tr \left( (M)^{-1} (y - Td) (y - Td)^T \right) + \log(|M|) + c\]

\[M_{i,j} = \rho k(x_i, x_j) + \sigma^2\]

The idea is to introduce many zeroes into this matrix.

\[w(u, v)\] a covariance that is zero if \[||u - v|| > \Delta\]
Tapered version of covariance

Replace: \( k(\mathbf{x}_i, \mathbf{x}_j) \)

with \( k(\mathbf{x}_i, \mathbf{x}_j)w(\mathbf{x}_i, \mathbf{x}_j) \)

Covariances with locations with distance larger than \( \Delta \) will zero!

The one taper -2 log likelihood:

\[
(\mathbf{y} - \mathbf{T}\mathbf{d})^T(M \cdot W)^{-1}(\mathbf{y} - \mathbf{T}\mathbf{d}) + \log(|M \cdot W|) + c
\]

\[
[M \cdot W]_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j)w(\mathbf{x}_i, \mathbf{x}_j)
\]
**The two tapered -2 log likelihood:**

Also taper the matrix \((y - Td)(y - Td)^T\) in the same way

\[
tr([M \cdot W]^{-1}[(y - Td)(y - Td)^T \cdot W]) + \log(|M \cdot W|) + c
\]

This will tend to give less biased estimates than the one taper.
Some closing remarks

• The likelihood can be simplified by eliminating the fixed part.

• Cholesky decompositions lead to efficient methods for evaluating the likelihood.

• Tapering can substantially reduce the amount of computation by introducing sparse covariance matrices.