Some recap and hopefully useful points

Alan Gelfand\(^1\) and Andrew O. Finley\(^2\)

\(^1\) Department of Statistical Science, Duke University, Durham, North Carolina.
\(^2\) Department of Forestry, Michigan State University, Lansing, Michigan.

September 9, 2014

The exponential model

- The sill is only reached asymptotically, meaning that strictly speaking, the range is infinite.
- To define an “effective range”, for \( t > 0 \), we see that as \( t \to \infty \), \( \gamma(t) \to r^2 + \sigma^2 \) which would become \( C(0) \).
- Again, \( C(t) = \left\{ \begin{array}{ll} \tau^2 + \sigma^2 & \text{if } t = 0 \\ \sigma^2 \exp(-\phi t) & \text{if } t > 0 \end{array} \right. \cdot \)
- Then the correlation between two points distance \( t \) apart is \( \exp(-\phi t) \);
- We define the effective range, \( t_0 \), as the distance at which this correlation = 0.05. Setting \( \exp(-\phi t_0) \) equal to this value we obtain \( t_0 \approx 3/\phi \), since \( \log(0.05) \approx -3 \).

A bit more on covariance functions

- To be a valid covariance function the function must be positive definite
- Whether a function is positive definite or not can depend upon dimension
- \( c \) is a valid covariance functions if and only if it is the characteristic function of a symmetric about 0 random variable (Bochner’s Theorem), i.e., 
  \[ c(h) = \int \cos(w \cdot h) G(dw) \]
- Fourier transform, spectral distribution, spectral density
- In principle, the inversion formula could be used to check if \( c(h) \) is valid

Some hopefully useful points

- More on covariance functions
- Kriging
- Priors and identifiability
- Last remarks

The Matérn correlation function

- The Matérn is a very flexible isotropic family:
  \[ C(t) = \left\{ \begin{array}{ll} \frac{\alpha^2}{\Gamma(\nu)} (2^{\nu-1})(\psi(\phi t) \nu K_{\nu}(2\sqrt{\psi(\phi t)}) & \text{if } t > 0 \\ \frac{\sigma^2}{\nu} t^\nu \exp(-\phi t) & \text{if } t = 0 \end{array} \right. \]
- \( K_\nu \) is the modified Bessel function of order \( \nu \)
- \( \nu \) is a smoothness parameter:
  - \( \nu = 1/2 \Rightarrow \text{exponential; } \nu \to \infty \Rightarrow \text{Gaussian; } \nu = 3/2 \Rightarrow \) convenient closed form for \( C(t), \gamma(t) \)
  - in two-dimensions, the greatest integer in \( \nu \) indicates the number of times process realizations will be mean-square differentiable.

Constructing valid covariance functions

Construct valid covariance functions by using properties of characteristic functions
- multiply valid covariance functions (corresponds to summing independent random variables)
- mixing covariance functions (corresponds to mixing distributions)
- convolving covariance functions (if \( c_1 \) and \( c_2 \) are valid then \( c_{12}(\mathbf{s}) = \int c_1(\mathbf{s} - \mathbf{u})c_2(\mathbf{u})d\mathbf{u} \) is valid).
- There are conditions for valid variograms but difficult and not of interest for us.
Classical spatial prediction or “Kriging”

- Named by Matheron (1963) in honor of D.G. Krige, a South African mining engineer whose seminal work on empirical methods for geostatistical data inspired the general approach.
- Optimal spatial prediction: given observations of a random field \( y = (y(s_1), \ldots, y(s_n))^T \), predict the variable \( y \) at a site \( s_0 \) where it has not been observed.
- Under squared error loss, the best linear prediction minimizes \( E[(y(s_0) - \hat{y}(s_0))^2] \) over \( \delta_0 \) and the \( \xi_i \).
- With an estimate of \( \gamma \), one immediately obtains the ordinary kriging estimate.
- Other than intrinsic stationarity, no distributional assumptions are required for the \( y(s) \).

Kriging with Gaussian processes

- Given covariate values \( x(s_i), i = 0, 1, \ldots, n \), suppose \( y = X\beta + \epsilon, \) where \( \epsilon \sim N(0, \Sigma) \).
- For a spatial covariance structure having no nugget effect, we specify \( \Sigma \) as \( \Sigma = \sigma^2 H(\phi) \) where \( (H(\phi))_{ij} = \rho(\phi; d_{ij}) \), where \( d_{ij} = ||s_i - s_j|| \).
- For a model having nugget effect, we instead set \( \Sigma = \sigma^2 H(\phi) + \tau^2 I. \)

Gaussian kriging cont.

- We seek the function \( g(y) \) that minimizes the mean-squared prediction error, \( E[(y(s_0) - g(y))^2 \mid y] \).
- \( \hat{g}(y) = E[y(s_0) \mid y] \).
- Intuitive from a Bayesian point of view, since this \( \hat{g}(y) \) is just the posterior mean of \( y(s_0) \).
- Using standard conditional normal distribution calculations, we obtain
  \[ E[y(s_0) \mid y] = X^T \beta + \gamma^T \Sigma^{-1} (y - X\beta), \]
  \[ \text{Var}[y(s_0) \mid y] = \sigma^2 + \tau^2 - \gamma^T \Sigma^{-1} \gamma. \]

Priors

- Again, basic geostatistical model:
  \( y(s) = x^T \beta + \omega(s) + \epsilon(s) \)
- The likelihood is given by:
  \( y \mid \theta \sim N(X\beta, \sigma^2 H(\phi) + \tau^2 I) \)
- Typically, independent priors are chosen for the parameters: \( p(\theta) = p(\beta)p(\sigma^2)p(\tau^2)p(\phi) \) Useful candidates are multivariate normal for \( \beta \), and inverse gamma for \( \sigma^2 \) and \( \tau^2 \).
- Specification of \( p(\phi) \) depends upon choice of \( \rho \) function; a uniform or discrete prior is usually selected.

Issues

- These are not estimators; they are really \( E[y(s_0) \mid y, \theta] \) and \( \text{Var}[y(s_0) \mid y, \theta] \). Parameters are unknown
- Plug in estimates of the parameter?
- Estimators are no longer linear, no longer unbiased, don’t account for the uncertainty in the parameter estimates
- Can we do satisfying inference with these estimators?
- A cleaner way: the posterior predictive distribution of \( y(s_0) \)

Priors cont.

- Informativeness: \( p(\beta) \) can be “flat” (improper)
- Without nugget \( (\tau^2) \), can’t identify both \( \sigma^2 \) and \( \phi \) (Zhang, 2004). With Matérn, can identify the product \( (\sigma^2 \phi)^\nu \). So an informative prior on at least one of these parameters
- With \( \tau^2, \phi \) and at least one of \( \sigma^2 \) and \( \tau^2 \) require informative priors.
- If the prior on \( \beta, \sigma^2, \phi \) is of the form \( \frac{p(a)^{a+1}}{\Gamma(a)} \pi(\cdot) \) proper, then, improper posterior if \( a < 2 \)
- Shows the problem with using \( IG(a, b) \) priors for \( \sigma^2 \) - “nearly” improper. Safer is \( IG(a, b) \) with \( a \geq 1 \)
Last remarks

- We are modeling with spatial random effects
- Introducing these in the transformed mean encourages means of spatial variables at proximate locations to be close to each other
- Marginal spatial dependence is induced between, say, \( y(s) \) and \( y(s') \), but observed \( y(s) \) and \( y(s') \) need not be close to each other
- Second stage spatial modeling is attractive for spatial explanation in the mean
- First stage spatial modeling more appropriate to encourage proximate observations to be similar.