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

Some hopefully useful points

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

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 $t \to \infty$, $\gamma(t) \to \tau^2 + \sigma^2$ which would become $C(0)$.
- Again,
$$C(t) = \begin{cases} \tau^2 + \sigma^2 & \text{if } t = 0 \\ \sigma^2 \exp(-\phi t) & \text{if } t > 0 \end{cases}.$$
- 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 $\log(0.05) \approx -3$.

The Matèrn correlation function

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

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., $c(\mathbf{h}) = \int \cos(\mathbf{w}^T\mathbf{h})G(d\mathbf{w})$
- Fourier transform, spectral distribution, spectral density
- In principle, the inversion formula could be used to *check* if $c(\mathbf{h})$ is valid

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 $\mathbf{y} = (y(\mathbf{s}_1), \ldots, y(\mathbf{s}_n))'$, predict the variable y at a site $\mathbf{s}_0$ where it has not been observed
- Under squared error loss, the best linear prediction minimizes $E[y(\mathbf{s}_0) - (\sum \ell_i y(\mathbf{s}_i) + \delta_0)]^2$ over δ_0 and the ℓ_i.
- With an estimate of γ, one immediately obtains the ordinary kriging estimate.
- Other than intrinsic stationarity, no distributional assumptions are required for the $y(\mathbf{s}_i)$.

Kriging with Gaussian processes

- Given covariate values $\mathbf{x}(\mathbf{s}_i),\ i = 0, 1, \ldots, n$, suppose

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon},\ \text{where } \boldsymbol{eta} \sim N(\mathbf{0}, \Sigma)\ .$$

- For a spatial covariance structure having no nugget effect, we specify Σ as

$$\Sigma = \sigma^2 H(\phi)\ \text{where}\ (H(\phi))_{ij} = \rho(\phi; d_{ij})\ ,$$

where $d_{ij} = ||\mathbf{s}_i - \mathbf{s}_j||$.
- For a model having nugget effect, we instead set

$$\Sigma = \sigma^2 H(\phi) + \tau^2 I.$$

Gaussian kriging cont.

- We seek the function $g(\mathbf{y})$ that minimizes the mean-squared prediction error, $E\left[(y(\mathbf{s}_0) - g(\mathbf{y}))^2 \mid \mathbf{y}\right]$.
- $g(\mathbf{y}) = E[y(\mathbf{s}_0)|\mathbf{y}]$.
- Intuitive from a Bayesian point of view, since this $g(\mathbf{y})$ is just the *posterior mean* of $y(\mathbf{s}_0)$!
- Using standard conditional normal distribution calculations, we obtain

$$\begin{aligned} E[y(\mathbf{s}_0)|\mathbf{y}] &= \mathbf{x}_0^T\boldsymbol{\beta} + \boldsymbol{\gamma}^T\Sigma^{-1}(\mathbf{y} - X\boldsymbol{\beta})\ , \\ \text{and} Var[Y(\mathbf{s}_0)|\mathbf{y}] &= \sigma^2 + \tau^2 - \boldsymbol{\gamma}^T\Sigma^{-1}\boldsymbol{\gamma}\ . \end{aligned}$$

Issues

- These are not estimators; They are really $E[y(\mathbf{s}_0)|\mathbf{y}, \boldsymbol{\theta}]$ and $Var[y(\mathbf{s}_0)|\mathbf{y}, \boldsymbol{\theta}]$. 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 $y(\mathbf{s}_0)|\mathbf{y}$

Priors

- Again, basic geostatistical model:

$$y(\mathbf{s}) = \mathrm{x}^T\beta + \omega(\mathbf{s}) + \epsilon(\mathbf{s})$$

- The likelihood is given by:

$$\mathbf{y}|\boldsymbol{\theta} \sim N(X\boldsymbol{\beta}, \sigma^2 H(\phi) + \tau^2 I)$$

- Typically, independent priors are chosen for the parameters: $p(\boldsymbol{\theta}) = p(\boldsymbol{\beta})p(\sigma^2)p(\tau^2)p(\phi)$ Useful candidates are multivariate normal for $\boldsymbol{\beta}$, and inverse gamma for σ^2 and τ^2.
- Specification of $p(\phi)$ depends upon choice of ρ function; a uniform or discrete prior is usually selected.

Priors cont.

- Informativeness: $p(\boldsymbol{\beta})$ can be “flat" (improper)
- Without nugget (τ^2), can't *identify* both σ^2 and ϕ (Zhang, 2004). With Matérn, can identify the product $(\sigma^2\phi)^\nu$. So an informative prior on at least one of these parameters
- With τ^2, ϕ and at least one of σ^2 and τ^2 require informative priors.
- If the prior on $\boldsymbol{\beta}, \sigma^2, \phi$ is of the form $\frac{p(\phi)}{\sigma^2}^{a+1}$ with $\pi(\cdot)$ proper, then, improper posterior if $a < 2$
- Shows the problem with using $IG(\epsilon, \epsilon)$ priors for σ^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, $y(\mathbf{s})$ and $y(\mathbf{s}')$, but observed $y(\mathbf{s})$ and $y(\mathbf{s}')$ 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.