Spatial prediction

Applied Spatial Statistics

Spatial prediction

- Spatial prediction is a fundamental problem in spatial statisitcs
- We collect observations $Y_1, ..., Y_n$ at locations $\mathbf{s}_1, ..., \mathbf{s}_n$
- Now we want to predict the value at another location $\mathbf{s}_0$
- Denote the predicted value as $\hat{Y}_0$
- Ideally we would also assign a prediction standard deviation and prediction interval

Examples

- Zillow predicting housing prices
- EPA estimating air pollution
- Geologists mapping oil fields
- Meteorologists mapping wind fields
- Farmers understanding their fields

Possible methods

- K nearest neighbors
- Inverse distance weighting
- Linear regression with polynomial functions of $\mathbf{s}$ as covariates
- Random forests with $\mathbf{s}$ as a covariate
- Kriging

Optimal spatial prediction

- What is the **best** way to make predictions?
- This depends on how we define “best” and the class of methods under consideration
- Kriging is the best unbiased linear predictor (BLUP)
- This is a narrow definition of optimality, and so there is no guarantee it is the best for a particular analysis
- The optimality result also assumes the spatial covariance function is known

Optimal spatial prediction - technical details

- One way to define "best" is via expected mean squared prediction error

$$E(Y_0 - \hat{Y}_0)^2$$

- A large class are predictions are linear

$$\hat{Y}_0 = \sum_{i=1}^{n} \lambda_i Y_i$$

and unbiased so that $E(Y_0 - \hat{Y}_0) = 0$

- It can be shown that Kriging gives the optimal formula for the weights λ_i in terms of the spatial covariance

Kriging derivation

- The derivation is interesting, but complicated
- This is a good derivation if you are interested
- This does not assume the data are Gaussian, only that the mean is constant and covariance are known
- Assuming normality, Kriging can also be derived using the properties of the multivariate normal distribution

Kriging weights

- Assume we have estimated the mean and covariance parameters

- Denote the fitted mean vector and covariance matrix of $\mathbf{Y} = (Y_1, ..., Y_n)$ has $\mu(\hat{\beta})$ and $\Sigma(\hat{\theta})$

- The Kriging prediction is

$$\mu_0(\hat{\beta}) + \Sigma_0(\hat{\theta})\Sigma(\hat{\theta})^{-1}\{\mathbf{Y} - \mu(\hat{\beta})\}$$

where $\mathrm{E}(Y_0) = \mu_0(\hat{\beta})$ and element i of $\Sigma_0(\hat{\theta})$ is $\mathrm{Cov}(Y_0, Y_i)$

Kriging weights

- The Kriging weights are (ignoring $\hat{\beta}$)

$$(\lambda_1, ..., \lambda_n) = \Sigma_0(\hat{\boldsymbol{\theta}})\Sigma(\hat{\boldsymbol{\theta}})^{-1}$$

- This expression is complicated, but clearly relies on the spatial covariance
- If we let $\mathbf{Z} = (Z_1, ..., Z_n) = \Sigma(\hat{\boldsymbol{\theta}})^{-1}\mathbf{Y}$, then

$$\hat{Y}_0 = \sum_{i=1}^{n} \text{Cov}(Y_0, Y_i)Z_i$$

- Locations close to $\mathbf{s}_0$ get high weight and vice versa
- The weights are easier to understand with plots

Kriging standard errors

- The kriging variance is simply

$$\text{Var}(\hat{Y}_0) = \Sigma_0(\hat{\theta})\Sigma(\hat{\theta})^{-1}\Sigma_0(\hat{\theta})^T$$

- This assumes the covariance (via $\hat{\theta}$) is known and thus does not include this sources of uncertainty

- Further assuming normality gives 95% prediction interval

$$\hat{Y}_0 \pm 1.96\sqrt{\text{Var}(\hat{Y}_0)}$$

Types of Kriging

- Simple Kriging: the mean is known or assumed to be zero
- Ordinary Kriging: The mean is constant across space and estimated, $\mu_i(\hat{\boldsymbol{\beta}}) = \hat{\beta}_0$
- Universal Kriging: The mean is estimated and varies spatially, $\mu_i(\hat{\boldsymbol{\beta}}) = \hat{\beta}_0 + \sum_{j=1}^{p} X_{ij}\hat{\beta}_j$
- Local Kriging: We use only the nearest k locations to $\mathbf{s}_0$ for prediction to speed up this step (inverting $\Sigma(\hat{\boldsymbol{\theta}})$ is slow)