KRIGING

BLAKE ROSENTHAL

1. Spatial statistics

Kriging is a method utilized in the field of geostatistics to model spatial data. Originally developed from the South African mining industry in the 1950's [2], kriging provided a way to predict ore-grade distributions based on a limited empirical sample. Though the name comes from mining engineer D. G. Krige, methods for optimal spatial linear prediction from Wold (1938), Kolmogorov (1941b), and Wiener (1949) all include the crucial covariance component of spatial interpolation, realizing that points closer to the prediction point should be given greater weights than further points. This is the cornerstone of the kriging method and is explored in detail in the following section.

Given a spatially continuous random process $Y(x)$ over some two-dimensional region B, a data sample $S_i : i = 1, \ldots, n$ is obtained from Y at locations $x_i : i = 1, \ldots, n$. [1]From a practical perspective, Y can be thought of as an underlying but unknown distribution of a variable of interest over B, be it ore-density, mineral concentrations, elevation, etc. S is therefore a set of vectors containing an independent spatial component and a dependent variable or variables. Since S is only a small and incomplete realization of the field Y, the standard geostatistical approach is to impose an underlying structure to the field consisting of a mean function $\mu(\mathbf{s})$ and a random error process with zero mean $e(\mathbf{s})$. Together these specify that

$$Y(\mathbf{s}) = \mu(\mathbf{s}) + e(\mathbf{s}),$$

where $\mu(\mathbf{s}) = E[Y(\mathbf{s})]$.

The goal is to make some predictions regarding the underlying random process Y. Kriging at its simplest is a matter of predicting a value of $Y(x_i)$ at an arbitrary point within the region B. *Simple kriging* assumes Y to have a constant mean which is estimated from the sample mean of S. The predictor of T is then the integral of the best linear predictor of $Y(x)$. *Ordinary kriging* uses the estimated

[1]A note on notation: here, x will be used to specify a generic point in Y, while $\mathbf{s}$ will be used to indicate the vector of spatial coordinates or other dependent variables that make up the sample S.

covariance structure of Y to replace the sample mean with the generalized least squares estimate of μ. Finally, *universal kriging* uses a regression model for the mean.

2. Covariance and the variogram

Part of the effectiveness of the kriging method comes from the recognition that the data from a spatial sample are correlated based on proximity. Points closer together are expected to be more highly correlated than points with greater spatial separation. The variogram, or semivariogram, plots this correlation as a function of distance, and the empirical semivariogram is the observed covariance structure of the data [3]. Given this, the semivariogram is defined by $\gamma(x_i - x_j) = \frac{1}{2}\text{var}\{e(x_i) - e(x_j)\}$, for all $x_i, x_j \in B$. The distance between any two points x_i and x_j can be can be used to define a new set $H = \{x_i - x_j : x_i, x_j \in B\}$ for the continuous distribution of distances, or lags, in B. Elements of H can be grouped into bins $H_1, H_2, \ldots, H_k$. A representative lag for the entire bin $\mathbf{h}_u$ can be used to define the unbiased estimator of $\gamma(\mathbf{h}_u)$ by

$$\hat{\gamma}(\mathbf{h}_u) = \frac{1}{2n(H_u)} \sum_{x_i - x_j \in H_u} [\hat{e}(x_i) - \hat{e}(x_j)]^2 \quad (u = 1, \ldots, k)$$

where $n(\mathbf{x})$ is the number of pairs of points whose difference is within a specified tolerance of $\mathbf{x}$. This assumes that covariance between data points is a function of spatial distance only, and not location or other factors. This estimation also requires a subjective choice in binning – since any exact distance, or lag, between two points is unlikely to occur frequently within a sample, it is necessary to group distances into representative intervals, or bins. A common way to do this is to make this binning choice up front, perhaps grouping the data into thirty or so bins then choosing $\mathbf{h}_u$ to be the average of all the lags that fall into a given bin. Therefore, unless the data is taken on a rectangular or polar grid, the accuracy of the semivariogram will always be dependent on the binning choices. What's the right number of bins? There's a trade-off – more bins means that $\mathbf{h}_u$ is a better estimation of its representative bin H_u, yet there are fewer lags to any particular bin and a smaller sample size and therefore a greater sampling variation. This is an interesting optimization problem on its own, but the data itself may impose binning restrictions depending on the sample size and other factors. This means that there is therefore no uniquely optimized semivariogram.

Fitting a parametric model to the empirical variogram gives a convenient equation to work with for several reasons – first, the empirical semivariogram is often all over the place, and a smoothed version will have a lower variance. Second, the empirical semivariogram usually fails to be conditionally nonpositive definite. This is a necessary condition when choosing predictors at later stages since the

prediction error variance must be nonnegative at every point in the field. Third, predicting locations at lags not represented by the chosen bins requires a continuous function, something only a smoothed variogram can accomplish. This smoothed version must satisfy the following necessary and sufficient conditions to be a valid semivariogram:

(1) Vanishing at 0: $\gamma(\mathbf{0}) = 0$
(2) Evenness: $\gamma(-\mathbf{x}) = \gamma(\mathbf{x})$
(3) Conditional negative definiteness: $\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \gamma(x_i - x_j) \leq 0$ for all n, all $s_1, \ldots, s_n$ and all $a_1, \ldots, a_n$ such that $\sum_{i=1}^{n} a_i = 0$

3. Estimation of the mean function

Simple, ordinary, and universal kriging all differ in their approach to estimating the mean function. Simple kriging, which assumes a constant mean, is typically dismissed by most statisticians since it usually fails to accurately describe any naturally occurring random process. Here we go over universal kriging since it is the best linear unbiased prediction model for geostatistical random fields [3].

The mean function is given by the linear equation

$$\mu(\mathbf{s}; \beta) = \mathbf{X}(\mathbf{s})^T \beta$$

where $\mathbf{X}(\mathbf{s})$ is a vector of covariates observed at $\mathbf{s}$. These variables could be simply latitude and longitude coordinates, but may also include such information as elevation, slope, windspeed, etc. If using only latitude and longitude, for example, a first order trend surface model is given by

$$\mu(\mathbf{s}; \beta) = \beta_0 + \beta_1 s_1 + \beta_2 s_2$$

where $\mathbf{s} = (s_1, s_2)$ are latitude and longitude. This restricts the mean to independent latitude and longitude effects, which can be circumvented by using higher order models such as the quadratic:

$$\mu(\mathbf{s}; \beta) = \beta_0 + \beta_1 s_1 + \beta_2 s_2 + \beta_{11} s_1^2 + \beta_{12} s_1 s_2 + \beta_{22} s_2^2.$$

At this point the provisional linear mean function is then fitted to the available data. There are many ways to do this, but the ordinary least squares method is typically used. This method yields an estimator $\hat{\beta}_{OLS}$ given by

$$\hat{\beta}_{OLS} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$$

where $\mathbf{X} = [X(\mathbf{s}_1), X(\mathbf{s}_2), \ldots, X(\mathbf{s}_n)]^T$ and $\mathbf{Y} = [Y(\mathbf{s}_1), Y(\mathbf{s}_2), \ldots, Y(\mathbf{s}_n)]^T$. [2]

It is possible to stop the analysis here, but once we have the second-order dependency structure of the semivariogram we can reestimate the mean function using estimated generalized least squares. A method given by Zimmerman and Stein ([3], p. 40) involves estimating a covariance matrix to include in the mean estimation. [3]

4. Spatial Prediction: Kriging

Given an prediction point $\mathbf{s}_0$ [4], the goal of kriging is to find a predictor $\hat{Y}(\mathbf{s}_0)$ for $Y(\mathbf{s}_0)$ that minimizes the prediction error variance $\text{var}[\hat{Y}(\mathbf{s}_0) - Y(\mathbf{s}_0)$ of all possible predictors that are both (1), linear, and (2) unbiased:

(1) $\hat{Y}(\mathbf{s}_0) = \lambda^T \mathbf{Y}$, where λ is a vector of fixed constants and $\sum \lambda_i = 1$.

(2) $E[\hat{Y}(\mathbf{s}_0)] = E[Y(\mathbf{s}_0)]$, or equivalently, $\lambda \mathbf{X} = \mathbf{X}(\mathbf{s}_0)$.

If the true semivariogram is known, the universal kriging predictor is then given by

$$\hat{Y}(\mathbf{s}_0) = [\gamma + \mathbf{X}(\mathbf{X}^T \Gamma^{-1} \mathbf{X})^{-1}(\mathbf{x}_0 - \mathbf{X}^T \Gamma^{-1} \gamma)]^T \Gamma^{-1} \mathbf{Y}$$

where $\gamma = [\gamma(\mathbf{s}_1 - \mathbf{s}_0), \ldots, \gamma(\mathbf{s}_n - \mathbf{s}_0)]^T$, Γ is the $n \times n$ symmetric matrix with ijth element $\gamma(\mathbf{s}_i - \mathbf{s}_j)$ and $\mathbf{x}_0 = \mathbf{X}(\mathbf{s}_0)$.

Minimizing the prediction error variance then gives us the kriging variance which can be expressed as

$$\sigma^2(\mathbf{s}_0) = \gamma^T \Gamma^{-1} \gamma - (\mathbf{X}^T \Gamma^{-1} \gamma - \mathbf{x}_0)^T (\mathbf{X}^T \Gamma^{-1} \mathbf{X})^{-1} (\mathbf{X}^T \Gamma^{-1} \gamma - \mathbf{x}_0).$$

---

[2] Equivalently, and perhaps easier to work with, $\hat{\beta}_{OLS} = \text{argmin} \sum_{i=1}^n [Y(\mathbf{s}_i) - \mathbf{X}(\mathbf{s}_i)^T \beta]^2$.

[3] Note to self: come back to this

[4] Usually this is an unknown point in B, but can also be a known point.

References

[1] Roger S. Bivand, Edzer J. Pebesma, and Virgilio Gómez-Rubio. *Applied Spatial Data Analysis with R.* Springer, New York, NY, 2008.

[2] Noel A.C. Cressie. *Statistics for Spatial Data.* John Wiley and Sons, Inc, New York, NY, 1993.

[3] Alan E. Gelfand, Peter J. Diggle, Montserrat Fuentes, and Peter Guttorp. *Handbook of Spatial Statistics.* Chapman and Hall, Boca Raton, FL, 2010.

[4] Michael Stein. *Interpolation of Spatial Data.* Springer, New York, NY, 1999.