KRIGING

BLAKE ROSENTHAL

1. Spatial statistics

Kriging is a method utilized in the field of geostatistics to model spatial data. Given a spatially continuous random process $S(x)$ over some two-dimensional region B, a sample $Y_i : i = 1, \ldots, n$ is obtained from S at locations $x_i : i = 1, \ldots, n$. After taking a noise factor Z into account, we assume that $Y_i = S(x_i) + Z_i$ where the Z_i are independent.

The goal is to make some predictions regarding the underlying random process S. Letting $T = \int_B S(x)dx$ allows us to interpolate over the sample region using the data Y_i. Kriging at its simplest is a matter of predicting a value of $S(x_i)$ at an arbitrary point within the region B. *Simple kriging* assumes S to have a constant mean which is estimated from the sample mean of Y. The predictor of T is then the integral of the best linear predictor of $S(x)$. *Ordinary kriging* uses the estimated covariance structure of S 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 proximity. Points closer together are expected to be more highly correlated than points with greater spatial separation. The empirical semivariogram is the plotted covariance structure of the data. The distance between any two points x_i and x_j can be described as a vector $\mathbf{x}$. For such a vector, the semivariogram of a random field Z can be described by $\gamma(\mathbf{x}) = \frac{1}{2}\mathrm{var}\{Z(\mathbf{x}) - Z(\mathbf{0})\}$. An unbiased estimator of $\gamma(\mathbf{x})$ is

$$\hat{\gamma}(\mathbf{x}) = \frac{1}{2n(\mathbf{x})} \sum_{\mathbf{x}_i - \mathbf{x}_j = x} \{Z(\mathbf{x}_i) - Z(\mathbf{x}_j)\}^2$$

where $n(\mathbf{x})$ is the number of pairs of points whose difference is within a specified tolerance of $\mathbf{x}$.

Fitting a parametric model to the empirical variogram gives a convenient equation to work with. This smoothed version must satisfy the following necessary and sufficient conditions for 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 $\sum_{i=1}^{n} a_i = 0$

References

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

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

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