Advanced GIS Class Notes on Kriging

January 30, 2018

Notes largely based on [1, 2].

1 A few useful formulas

Variance:
$Var(Z) = E[(Z-\mu)^2] = E(Z^2) - (E(Z))^2 = \sigma_z^2$
$\mu = E(Z) = mean$

Covariance:
$Cov(Z,W) = E[(Z-\mu_z)(W-\mu_w)]$
$Cov(Z,Z) = Var(Z)$

Variance of the difference between two random variables:
$Var(Z-W) = Var(Z) + Var(W) - 2Cov(Z,W)$

Variance of linear combination of random variables:
$Var(\sum a_i Z_i) = \sum_i \sum_j a_i a_j Cov(Z_i, Z_j)$

Example: $Var(2Z_1 + 3Z_2) = 2 \times 2Cov(Z_1,Z_1) + 3 \times 3Cov(Z_2,Z_2) + 2 \times 3Cov(Z_1,Z_2) + 3 \times 2Cov(Z_2,Z_1) = 4Var(Z_1) + 9Var(Z_2) + 12Cov(Z_1,Z_2)$

2 Setup

Suppose we have a two-dimensional surface $\mathcal{A}$ and we want to estimate a value of a continuous attribute z (e.g. elevation, precipitation, soil pH) at any unsampled lo-

cation $\mathbf{u}$ using z known data points whose locations are $\mathbf{u}_\alpha$, or $\{z(\mathbf{u}_\alpha), \alpha = 1, ..., n\}$ The overall objective of kriging is to estimate a value Z for location u as a linear combination of known values around u. The general setup can be defined as:

$$\hat{Z}(\mathbf{u}) - m(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u}) \left[Z(\mathbf{u}_\alpha) - m(\mathbf{u}_\alpha)\right]$$

where $\lambda_\alpha(\mathbf{u})$ is the weight assigned to each nearby known value $z(\mathbf{u}_\alpha)$. The quantities $m(\mathbf{u})$ and $m(\mathbf{u}_\alpha)$ are the expected values of the random variables (RVs) $Z(\mathbf{u})$ and $Z(\mathbf{u})_\alpha$ respectively.

We can define the estimation error as a random variable $\hat{Z}(\mathbf{u}) - Z(\mathbf{u})$. The objective of kriging is to minimize this estimation error or its variance:

$$\sigma^2(\mathbf{u}) = Var\left\{\hat{Z}(\mathbf{u}) - Z(\mathbf{u})\right\}$$

under the (unbiasedness) constraint that $E\left\{\hat{Z}(\mathbf{u}) - Z(\mathbf{u})\right\} = 0$

The value Z can be decomposed into two parts, a mean component m or trend and a residual component r:

$$Z(\mathbf{u}) = m(\mathbf{u}) + R(\mathbf{u})$$

3 Simple Kriging

In simple kriging, we assume that the mean is known and constant over $\mathcal{A}$, or $m(\mathbf{u}) = m \ \forall \mathbf{u} \in \mathcal{A}$. The general equation becomes:

$$\hat{Z}(\mathbf{u}) - m = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u}) \left[Z(\mathbf{u}_\alpha) - m\right]$$

But $Z - m$ is just the residual R. Thus, we can write the unknown error to be estimated as a linear combination of the known error terms:

$$\hat{R}(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u}) R(\mathbf{u}_\alpha)$$

Again, the objective of kriging is to estimate the λ's such that the variance of the error is minimized subject to the unbiasedness constraint. In the case of simple

kriging, this always holds because $m(\mathbf{u})$ is constant across $\mathbf{u}$. Using the variance formulas above, we can write $Var\left\{\hat{R}(\mathbf{u}) - R(\mathbf{u})\right\}$ as:

$$\begin{aligned}\sigma^2(\mathbf{u}) &= Var\left\{\hat{R}(\mathbf{u})\right\} + Var\left\{R(\mathbf{u})\right\} - 2Cov\left\{\hat{R}(\mathbf{u}), R(\mathbf{u})\right\} \\ &= \sum_{\alpha=1}^{n(\mathbf{u})}\sum_{\beta=1}^{n(\mathbf{u})}\lambda_\alpha(\mathbf{u})\lambda_\beta(\mathbf{u})C_R(\mathbf{u}_\alpha - \mathbf{u}_\beta) + C_R(0) \\ &-2\sum_{\alpha=1}^{n}(\mathbf{u})\lambda_\alpha(\mathbf{u})C_R(\mathbf{u}_\alpha - \mathbf{u})\end{aligned}$$

This can be written in matrix format in a much simplified way. $C_R(0)$ is just a constant, call it σ^2 (global variance or sill). Rearranging terms, the equation becomes :

$$\sigma^2 - 2\lambda^T\mathbf{c} + \lambda^T\mathbf{C}\lambda \equiv \sigma^2(\mathbf{u})$$

where $\mathbf{c}$ is an $(n \times 1)$ column vector of covariances between the prediction point and each of the n sample sites of known points and $\mathbf{C}$ is an $(n \times n)$ matrix of covariance between all pairs of the n sample sites.

The optimal weights that minimize the error variance can be found by taking the partial first derivative of $\sigma^2(\mathbf{u})$ with respect to λ and setting to zero:

$$\begin{aligned}\tfrac{1}{2}\tfrac{\partial\sigma^2}{\partial\lambda} &= \mathbf{C}\lambda - \mathbf{c} = 0 \\ \mathbf{C}\lambda &= \mathbf{c} \\ \mathbf{C}^{-1}\mathbf{C}\lambda &= \mathbf{C}^{-1}\mathbf{c} \\ \lambda &= \mathbf{C}^{-1}\mathbf{c}\end{aligned}$$

In full matrix format, the solution is equivalent to:

$$\begin{bmatrix}\lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n\end{bmatrix} = \begin{bmatrix}cov_{11} & cov_{12} & \dots & cov_{1n} \\ cov_{21} & cov_{22} & \dots & cov_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ cov_{n1} & cov_{n2} & \dots & cov_{nn}\end{bmatrix}^{-1} \times \begin{bmatrix}cov_{s1} \\ cov_{s2} \\ \vdots \\ cov_{sn}\end{bmatrix}$$

The estimated minimum error variance (now with the $\hat{\sigma}^2$ to indicate it is estimated) can be found by substituting this equality into the original equation.

$$\begin{aligned}\hat{\sigma}^2 &= \sigma^2 - 2\lambda^T\mathbf{c} + \lambda^T\mathbf{C}\lambda \\ \hat{\sigma}^2 &= \sigma^2 - 2(\mathbf{C}^{-1}\mathbf{c})^T\mathbf{c} + (\mathbf{C}^{-1}\mathbf{c})^T\mathbf{C}(\mathbf{C}^{-1}\mathbf{c}) \\ \hat{\sigma}^2 &= \sigma^2 - 2(\mathbf{c}^T\mathbf{C}^{-1}\mathbf{c}) + \mathbf{c}^T\mathbf{C}^{-1}\mathbf{C}\mathbf{C}^{-1}\mathbf{c} \\ \hat{\sigma}^2 &= \sigma^2 - 2(\mathbf{c}^T\mathbf{C}^{-1}\mathbf{c}) + \mathbf{c}^T\mathbf{C}^{-1}\mathbf{c} \\ \hat{\sigma}^2 &= \sigma^2 - \mathbf{c}^T\mathbf{C}^{-1}\mathbf{c}\end{aligned}$$

where σ^2 is the sill value of your variogram (see below).

4 Ordinary Kriging

Recall that in simple kriging we known the constant mean m. In ordinary kriging, the mean is unknown and is allowed to vary locally by limiting the domain of stationarity to the neighborhood of $(\mathbf{u})$. The linear estimator is similar to the simple kriging one, except now the mean is $m(\mathbf{u})$.

$$\hat{Z}(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u}) Z(\mathbf{u}_\alpha) + \left[1 - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u})\right] m(\mathbf{u})$$

The unknown local mean $m(\mathbf{u})$ is filtered from the linear estimator by forcing the kriging weights to sum to 1. The ordinary kriging estimator is thus written as a linear combination of the surrounding values with a constraint as follows:

$$\hat{Z}(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u}) Z(\mathbf{u}_\alpha)$$

with

$$\sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u}) = 1$$

To see why the unbiasedness constraint must sum to 1, we write

$$\begin{aligned} E\left\{\hat{Z}(\mathbf{u}) - Z(\mathbf{u})\right\} &= \textstyle\sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u}) m(\mathbf{u}) - m(\mathbf{u}) \\ &= m(\mathbf{u}) - m(\mathbf{u}) = 0 \end{aligned}$$

Just like in simple kriging, we want to minimize the error variance $E[(\hat{Z}(\mathbf{u}) - Z(\mathbf{u}))^2]$. Since now we have a constrained optimization, the problem calls for the definition of a Lagrangian function $\mathcal{L}(\mathbf{u})$ with corresponding lagrangian multiplier $2\tau(\mathbf{u})$:

$$\mathcal{L}(\mathbf{u}) = \sigma^2(\mathbf{u}) + 2\tau(\mathbf{u}) \left[\sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u}) - 1\right]$$

The optimal weights $\lambda^\star$ are obtained by taking each first partial derivative and setting to zero:

$$\begin{aligned} \tfrac{1}{2}\tfrac{\partial \mathcal{L}(\mathbf{u})}{\partial \lambda_\alpha(\mathbf{u})} &= \textstyle\sum_{\beta=1}^{n(\mathbf{u})} \lambda_\beta(\mathbf{u}) C_R(\mathbf{u}_\alpha - \mathbf{u}_\beta) - C_R(\mathbf{u}_\alpha - \mathbf{u}) + \tau(\mathbf{u}) = 0 \\ \tfrac{1}{2}\tfrac{\partial \mathcal{L}(\mathbf{u})}{\partial \tau(\mathbf{u})} &= \textstyle\sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u}) - 1 = 0 \end{aligned}$$

Although we assume the mean $m(\mathbf{u})$ is stationary within the local neighborhood, in practice the residual covariance is taken from the global covariance inferred from all data available, leading to the system:

$$\begin{array}{ll} \frac{1}{2}\frac{\partial \mathcal{L}(\mathbf{u})}{\partial \lambda_\alpha(\mathbf{u})} & = \sum_{\beta=1}^{n(\mathbf{u})} \lambda_\beta(\mathbf{u}) C(\mathbf{u}_\alpha - \mathbf{u}_\beta) - C(\mathbf{u}_\alpha - \mathbf{u}) + \tau(\mathbf{u}) = 0 \\ \frac{1}{2}\frac{\partial \mathcal{L}(\mathbf{u})}{\partial \tau(\mathbf{u})} & = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u}) - 1 = 0 \end{array}$$

The minimum error variance can be calculated by plugging the first equation of the system into the definition of variance:

$$\sigma^{\star 2}(\mathbf{u}) = C(0) - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u}) C(\mathbf{u}_\alpha - \mathbf{u}) - \tau(\mathbf{u})$$

In full matrix format, this solution is similar to simple kriging, except that we add an extra row and column to account for the constraint and lagrange multiplier:

$$\begin{bmatrix} cov_{11} & cov_{12} & \dots & cov_{1n} & 1 \\ cov_{21} & cov_{22} & \dots & cov_{2n} & 1 \\ \vdots & \vdots & \dots & \vdots & 1 \\ cov_{n1} & cov_{n2} & \dots & cov_{nn} & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix} \times \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \tau \end{bmatrix} = \begin{bmatrix} cov_{01} \\ cov_{02} \\ \vdots \\ cov_{0n} \\ 1 \end{bmatrix}$$

Returning to the introductory remarks, ordinary kriging assumes local constant mean, which is a viable assumption in general. It does not need to explicitly calculate any global trend in the data. A possible drawback is that the variogram relies on the global dataset and therefore, it can distort local estimation if global trends are not explicitly estimated and removed before variogram modeling.

5 Kriging with a trend model

Kriging with a trend model (also known as universal kriging) is a natural extension of ordinary kriging. KTM implicitly calculates a global trend model for the data. The first order trend component ($m(\mathbf{u}) \equiv V^T\beta$) is modeled as a linear combination of external variables, typically a polynomial function of the coordinates. For example $V^T\beta \equiv \beta_1 x + \beta_2 y + \beta_3 xy$, where x, y are the coordinates of each point in S (here x, y are not independent, dependent variables; both are independent variables). If we define p variables in V as $v1, v2, ..., vp$, we can solve a system of

equations of the form $C_+\lambda_+ = c_{0+}$, just like in ordinary kriging, by including the implicit trend model variables to create an expanded matrix as follows:

$$\begin{bmatrix} cov_{11} & \dots & cov_{1n} & v1_1 & \dots & vp_1 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ cov_{n1} & \dots & cov_{nn} & v1_n & \dots & vp_n \\ v1_1 & \dots & v1_n & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ vp_1 & \dots & vp_n & 0 & \dots & 0 \end{bmatrix} \times \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \tau_1 \\ \vdots \\ \tau_p \end{bmatrix} = \begin{bmatrix} cov_{01} \\ \vdots \\ cov_{0n} \\ v1_0 \\ \vdots \\ vp_0 \end{bmatrix}$$

The solution is again the same as above but of course the augmented matrix is different.

In sum, the only information needed to obtain a kriging estimate is the variance-covariance structure between sample points and between the point to be predicted and sample points. This is modeled through what is known as 'variogram' according to the following definition:

$$2\gamma(\mathbf{h}) = Var\left[Z(\mathbf{u}) - Z(\mathbf{u}+\mathbf{h})\right] = E\left\{\left[Z(\mathbf{u}) - Z(\mathbf{u}+\mathbf{h})\right]^2\right\}$$

The semivariogram $\gamma(\mathbf{h})$ is one half of the variogram. In practice people refer to $\gamma(.)$ as the variogram. In sample notation this is:

$$\gamma = \frac{1}{2}\frac{1}{N(\mathbf{h})}\sum_{i=1}^{N}\left[z(u_i+h) - z(u_i)\right]^2$$

where the $z's$ are the measured values of Z at N pairs of comparisons, separated by the distance h.

By definition, the covariance at $\mathbf{h} = 0$, $C(0)$, is the variance σ^2. The covariance $C(\mathbf{h})$ is 0.0 when values $\mathbf{h}$-apart are not linearly correlated.

Expanding the square term in the definition of variogram and using the formulas shown in the beginning, we obtain the relation between the semivariogram and covariance:

$$2\gamma(\mathbf{h}) = Var(Z(\mathbf{u})) + Var(Z(\mathbf{u}+\mathbf{h})) - 2Cov(Z_u, Z_{u+h})$$

If we assume stationarity, the variance is independent of location and the first two variance terms are the same and the equation can be simplified to

$$\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h}) \text{ or } C(\mathbf{h}) = C(\mathbf{0}) - \gamma(\mathbf{h})$$

This relationship becomes $C(\mathbf{h}) = \sigma^2 - \gamma(\mathbf{h})$, where σ^2 is the global variance. Thus, we only need to model the (semi)variogram. We typically model it by taking pre-determined functional forms as follows

6 Variogram Models

The spherical model:

$$\gamma(h) = \begin{cases} \sigma^2 \left(\frac{3h}{2r} - \frac{h^3}{2r^3}\right) \text{for } h \leq r \\ \\ \sigma^2 \text{ otherwise} \end{cases}$$

where r is the range and σ^2 is the sill.
If we include a nugget effect a, the spherical model becomes:

$$\gamma(h) = \begin{cases} a + (\sigma^2 - a)\left(\frac{3h}{2r} - \frac{h^3}{2r^3}\right) & \text{for } 0 < h \leq r \\ \\ 0 & \text{for } h = 0 \\ \\ \sigma^2 & \text{otherwise} \end{cases}$$

The exponential model:

$$\gamma(h) = \sigma^2(1 - e^{-3h/r})$$

With a nugget effect, the exponential model becomes:

$$\gamma(h) = \begin{cases} a + (\sigma^2 - a)(1 - e^{-3h/r}) & \text{for } h > 0 \\ \\ 0 & \text{for } h = 0 \end{cases}$$

The Gaussian model:

$$\gamma(h) = \sigma^2(1 - e^{-3h^2/r^2})$$

The Gaussian model with nugget is:

$$\gamma(h) = \begin{cases} a + (\sigma^2 - a)(1 - e^{-3h^2/r^2}) & \text{for } h > 0 \\ \\ 0 & \text{for } h = 0 \end{cases}$$

References

[1] Pierre Goovaerts. *Geostatistics for natural resources evaluation*. Oxford University Press, New York, 1997.

[2] EH Isaaks and RM Srivastava. *Applied geostatistics*. Oxford University Press, New York, 1989.

Last Modified: January 30, 2018