

Lecture 13: Universal Kriging, Cross-Validation, Co-Kriging

Math 586

April 13, 2009

I. Universal kriging

So far considered $\mathbb{E}[V(\mathbf{x})] = m$ constant. What if it varies, $m = m(\mathbf{x})$?

Option 1. Estimate $m(\mathbf{x})$ using regression or other data, then simple kriging.

Option 2. Use ordinary kriging but only for neighboring data

Option 3. Universal kriging: estimate both trend and kriging weights in one step. (Assume knowing the “true” cov. function or variogram.)

Option 4. Difference the data to remove trend (IRF- k , $k > 0$), then operate on the differences. For example, if $V(x) = ax + \tilde{V}(x)$ with $\tilde{V}(x)$ being an IRF-0, then

$$V(x+h) - V(x) = \underbrace{ah}_{\text{const}} + \underbrace{\tilde{V}(x+h) - \tilde{V}(x)}_{\text{stat. homog.}}$$

Consider Option 3:

Universal kriging with $W(\mathbf{x}) = V(\mathbf{x}) - m(\mathbf{x})$ stat. homog. or an IRF-0.

Assumption: $m(\mathbf{x}) = \sum_{l=0}^L a_l f_l(\mathbf{x})$, with known functions $f_l(\mathbf{x})$ and unknown a_l 's.

For example, for polynomial trend $f_0(\mathbf{x}) = 1$, $f_1(\mathbf{x}) = x_1$, $f_2(\mathbf{x}) = x_2$, $f_3(\mathbf{x}) = x_1 x_2$ etc. as in Regression. Functions f_l can also be some other variables, observed at locations \mathbf{x} (“external drift”). As usual, we will need $L \ll n$.

Let $C_W(\mathbf{h}) = Cov[W(\mathbf{x} + \mathbf{h}), W(\mathbf{x})]$.

Consider linear estimator

$$\hat{V}_0 = \sum_{j=1}^n \lambda_j V(\mathbf{x}_j), \quad \text{then} \quad \mathbb{E}(\hat{V}_0) = m(\mathbf{x}_0) = \sum_{j=1}^n \lambda_j m(\mathbf{x}_j)$$

is an unbiasedness constraint, leads to $L + 1$ equations

$$m(\mathbf{x}_0) = \sum_{l=0}^L a_l f_l(\mathbf{x}_0) = \sum_{j=1}^n \lambda_j m(\mathbf{x}_j) = \sum_{j=1}^n \lambda_j \sum_{l=0}^L a_l f_l(\mathbf{x}_j)$$

therefore,

$$\sum_{l=0}^L a_l f_l(\mathbf{x}_0) = \sum_{l=0}^L a_l \sum_{j=1}^n \lambda_j f_l(\mathbf{x}_j)$$

for any combination of a_l 's, therefore

$$f_l(\mathbf{x}_0) = \sum_{j=1}^n \lambda_j f_l(\mathbf{x}_j)$$

$L + 1$ constraints that lead to $L + 1$ Lagrange multipliers. Lagrange function

$$H(\lambda_1, \dots, \lambda_n; \mu_0, \dots, \mu_L) = MSE - 2 \sum_{l=0}^L \mu_l \left[\sum_{j=1}^n \lambda_j f_l(\mathbf{x}_j) - f_l(\mathbf{x}_0) \right],$$

$$\text{where} \quad MSE = \sum_{k=1}^n \sum_{j=1}^n \lambda_k C_W(\mathbf{x}_k - \mathbf{x}_j) \lambda_j$$

Leads to kriging equations

$$\begin{aligned} \sum_{j=1}^n \lambda_j C_W(\mathbf{x}_k - \mathbf{x}_j) &= C_W(\mathbf{x}_k - \mathbf{x}_0) + \sum_{l=0}^L \mu_l f_l(\mathbf{x}_k), \quad k = 1, \dots, n \\ f_l(\mathbf{x}_0) &= \sum_{j=1}^n \lambda_j f_l(\mathbf{x}_j) \quad l = 0, \dots, L \end{aligned}$$

(Note: if $L = 0$ and $f_0(\mathbf{x}) = 1$ we will get ordinary kriging)

Main problem: how do we estimate $C_W(\mathbf{x}_k - \mathbf{x}_j)$?? Maybe, through the iterative process: fit regression, estimate variogram of residuals, re-fit regression etc.

Another, more complicated option, is to use maximum likelihood to simultaneously estimate both the trend and variogram parameters.

Option 2.(local):

Paper by Journel & Rossi (*Math. Geol.* v.21(7)-1989, pp. 715-739) considered Universal kriging vs. using moving neighborhoods.

When estimating $V(\mathbf{x}_0)$, use only the points in the ellipse (anisotropy)

$$\frac{(x_1 - x_{10})^2}{a^2} + \frac{(x_2 - x_{20})^2}{b^2} \leq 1$$

or (another choice) nearest N^* points.

Points far from \mathbf{x}_0 contribute little, allow “local” mean to be used.

This method is simpler to use, and can do localized covariance estimators \Rightarrow not much use for universal kriging algorithm.

Kitanidis (p. 71): critique of moving neighborhoods.

Generalized Least Squares (GLS) (*Option 1.*)

Universal kriging is equivalent to first fitting the regression, then using simple kriging (see e.g. Deutsch and Journel, 1997). However, with residuals correlated, the ordinary least squares (OLS) is no longer the optimal technique!

To understand this, recall ordinary kriging and how, in order to estimate the mean, you need to adjust the weights to account for correlation between the data, e.g. clustering effect.

Suppose that the data \mathbf{Y} (or the residuals) have the covariance matrix Σ . The GLS estimate minimizes the quantity $(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'\Sigma^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$ found in the multivariate normal likelihood, and the solution is

$$\hat{\boldsymbol{\beta}}_{\text{GLS}} = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{Y} \quad (1)$$

For example, when we estimate the mean m , you can cast this as a regression problem with no predictors, intercept only: $\hat{Y}_i = \beta_0$. Then, $\mathbf{X} = \mathbf{1} = (1 \ 1 \ \dots \ 1)'$, and the GLS estimate of the mean is obtained using (1). This estimate will coincide with the ones we computed for ordinary kriging in Lab 3!

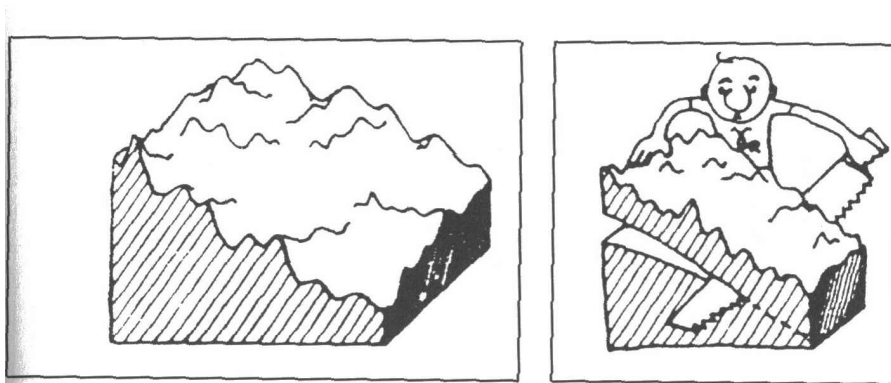


FIGURE 4.1. A favorable case for universal kriging. Author: J. P. Delhomme.

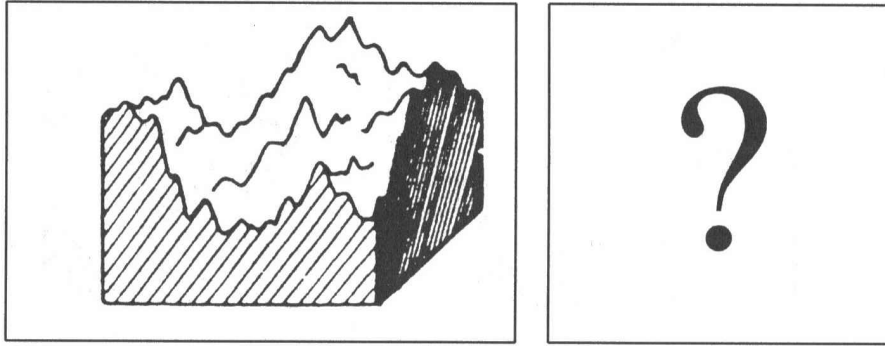


FIGURE 4.2. A puzzling case for drift modeling. Author: J. P. Delhomme.

II. Cross-Validation

Important to make sure that e.g. the variogram fit correctly.

Steps: given the estimated variogram,

1. Exclude point i
2. Krig, using all other data, only to get the prediction \hat{V}_i^* and krig. st.dev. σ_i^* .
3. Compute normalized residuals $Z_i = [V(\mathbf{x}_i) - \hat{V}_i^*] / \sigma_i^*$, they should have mean 0 and variance 1.
4. examine the residuals for distribution, spatial patterns, outliers etc.

The residuals are *not* independent \Rightarrow there isn't really a rigorous test.

III. Co-Kriging

Two random fields, observations on both, predict one. E.g., permeability and porosity \Rightarrow predict permeability.

Observe: $V_1(\mathbf{x}_{1i}), \quad i = 1, \dots, n_1, \quad V_2(\mathbf{x}_{2k}), \quad k = 1, \dots, n_2.$
 Predict: $V_1(\mathbf{x}_0)$

Need cross-covariances as well as covariances:

$$\begin{aligned} C_{11}(\mathbf{x}, \mathbf{y}) &= Cov[V_1(\mathbf{x}), V_1(\mathbf{y})] & C_{22}(\mathbf{x}, \mathbf{y}) &= Cov[V_2(\mathbf{x}), V_2(\mathbf{y})] \\ C_{12}(\mathbf{x}, \mathbf{y}) &= Cov[V_1(\mathbf{x}), V_2(\mathbf{y})] = C_{21}(\mathbf{y}, \mathbf{x}) \end{aligned}$$

Assume that the means are constant but unknown: $\mathbb{E} V_1(\mathbf{x}) = m_1, \quad \mathbb{E} V_2(\mathbf{x}) = m_2.$

Objective: estimate $V_1(\mathbf{x}_0)$ via

$$\hat{V} = \sum_{i=1}^{n_1} \lambda_{1i} V_1(\mathbf{x}_{1i}) + \sum_{i=1}^{n_2} \lambda_{2i} V_2(\mathbf{x}_{2i}),$$

with unbiasedness constraints

$$\sum_{i=1}^{n_1} \lambda_{1i} = 1, \quad \sum_{i=1}^{n_2} \lambda_{2i} = 0.$$

MSE will contain both cov. and cross-cov. terms, use 2 Lagrange multipliers etc. etc. resulting in equations

$$\begin{aligned} \sum_{i=1}^{n_1} \lambda_{1i} C_{11}(\mathbf{x}_{1i}, \mathbf{x}_{1j}) + \sum_{k=1}^{n_2} \lambda_{2k} C_{12}(\mathbf{x}_{1j}, \mathbf{x}_{2k}) &= C_{11}(\mathbf{x}_{1j}, \mathbf{x}_0) + \mu_1, \quad j = 1, \dots, n_1 \\ \sum_{i=1}^{n_1} \lambda_{1i} C_{12}(\mathbf{x}_{1i}, \mathbf{x}_{2j}) + \sum_{k=1}^{n_2} \lambda_{2k} C_{22}(\mathbf{x}_{2k}, \mathbf{x}_{2j}) &= C_{12}(\mathbf{x}_0, \mathbf{x}_{2j}) + \mu_2, \quad j = 1, \dots, n_2 \\ \sum_{i=1}^{n_1} \lambda_{1i} &= 1 \quad \sum_{k=1}^{n_2} \lambda_{2k} = 0 \end{aligned}$$

Issue: estimation of cross-covariance.

Application: kriging using block data.

Observe $V_2(\mathbf{x}) = \int_B g(\mathbf{y}) V_1(\mathbf{x} + \mathbf{y}) d\mathbf{y}$, with g weight function, say $\int_B g(\mathbf{x}) d\mathbf{x} = 1$.

Typically, g is an even function: $g(\mathbf{x}) = g(-\mathbf{x})$ and B is some neighborhood of 0. [E.g. pump test for transmissivity.]

Now, cov. and cross-cov. are expressed in terms of $V_1(\mathbf{x})$, i.e.

$$Cov[V_1(\mathbf{x}_1), V_2(\mathbf{x}_2)] = \int_B g(\mathbf{y}) C_{11}(\mathbf{x}_1, \mathbf{y} + \mathbf{x}_2) d\mathbf{y} \quad \text{Point to Block}$$

$$Cov[V_2(\mathbf{x}_1), V_2(\mathbf{x}_2)] = \int_B \int_B g(\mathbf{y}_1) C_{11}(\mathbf{x}_1 + \mathbf{y}_1, \mathbf{x}_2 + \mathbf{y}_2) g(\mathbf{y}_2) d\mathbf{y}_1 d\mathbf{y}_2$$

Block to Block

