Geostatistics - Estimation

Julián M. Ortiz - https://julianmortiz.com/

August 17, 2024

Summary

Predicting the value of a variable at an unsampled location is presented for continuous variables, such as the grade of an element.

We briefly review the methods for estimation (prediction) based on pure geometrical conditions, and then move on to introduce *kriging*, which is a collection of methods based on principles of regression. These methods make use of the spatial continuity of the variables by means of the variogram or the covariance. We show *simple kriging*, which assumes the mean of the random variables is constant and known over the domain. Then, we review *ordinary kriging*, where the mean is assumed constant within the search radius around the estimation location, but unknown. Other types of kriging are discussed.

We show the deduction of the kriging equations, by sequentially imposing the conditions of linearity, that is, the estimator is a weighted linear combination of the known samples, unbiasedness, which means that on average the

estimate is fair and does not overestimate or underestimate the true unknown value, and optimality, which imposes minimum spread in the squared errors.

We extend the notion of point estimation to blocks and explain the main properties, limitations and drawbacks of kriging.

1 Introduction

The main problem that we will address is that of estimating the value of a variable at an unsampled location, from known values at sample locations in its neighborhood.

> Notice that one of the main assumptions to perform this estimation is that all the sample locations and the unsampled location belong to the same domain.

As a general framework, we will consider the following setting: we have n samples, $z(\mathbf{u}_i), i = 1, \ldots, n$ available in the neighborhood of a location of interest $\mathbf{u}_0$, all belonging to the same domain (see **Figure 1**). We would like to predict the value at this locations. We denote this value as $z^*(\mathbf{u}_0)$.

> The estimate will be linear:
>
> $$z^*(\mathbf{u}_0) = \lambda_0 + \sum_{i=1}^{n} \lambda_i z(\mathbf{u}_i) \qquad (1)$$
>
> where $\lambda_i, i = 0, \ldots, n$ are constant values in $\mathbb{R}$

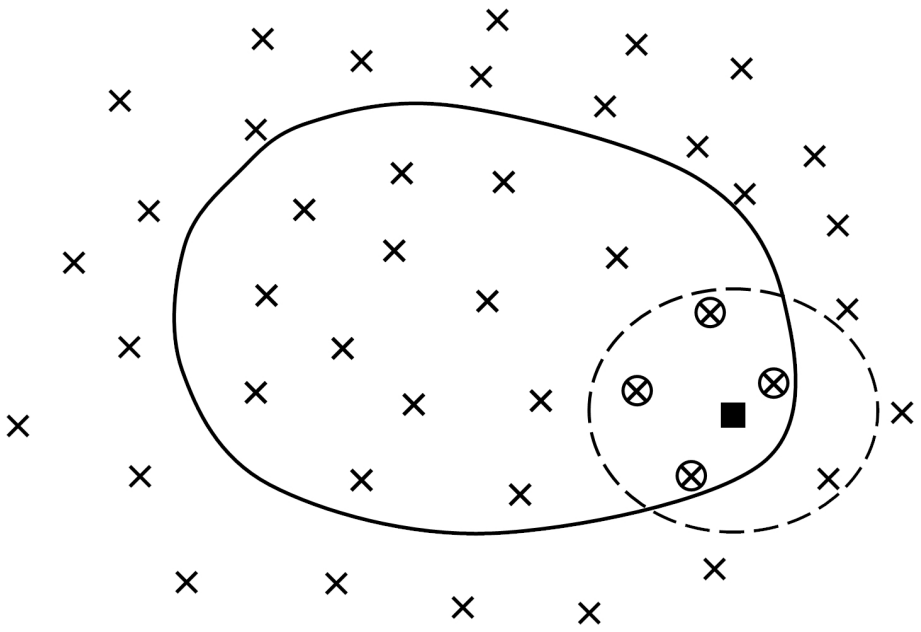

Figure 1: The general framework for the estimation problem. The domain is depicted by the continuous line, while the neighborhood for searching samples is shown by the dashed line. Only 4 of the 5 samples in the neighborhood are inside the domain and will be used for estimation.

The main question we need to address is how do we determine the weights $\lambda_i, i = 0, \ldots, n$.

Notice that we will write the variable in lower case, if no consideration of the notion of random function is necessary (as in the case of the geometric approaches: nearest neighbor estimation and inverse distance weighting estimation). When we call for the random function model, we will use upper case (as in kriging).

2 Nearest neighbor (polygonal) estimator

One of the simplest approaches to determine the weights in **Equation 1** is to assign all the weight to the closest sample and assign a weight of 0 to all other samples, and leave the constant λ_0 as 0, as well.

Nearest neighbor estimator

$$z^*_{NN}(\mathbf{u}_0) = \lambda^{NN}_0 + \sum_{i=1}^{n} \lambda^{NN}_i z(\mathbf{u}_i) \qquad (2)$$

$$\lambda^{NN}_i = \begin{cases} 1 & \text{if } \mathbf{u}_i \text{ is closest to } \mathbf{u}_0 \\ 0 & \text{otherwise} \end{cases} \qquad i = 0, \ldots, n$$

When estimating a regular grid of points, this estimator creates a set of polygons with constant value around each sample. This is similar to the approach for polygonal declustering described before.

The global distribution of estimated values with this approach, compensates for clustering of the data, and spans the entire variability of the sample distribution. This is a method that does not smooth the estimated values.

It is easy to see that this method does not take into account spatial continuity, and does not account for redundancy in the information.

3 Inverse distance weighting estimator

Another method based purely on geometry is the inverse distance weighting approach. Here, each sample is weighted according to how close it is to the unsampled location to be estimated. This weight can be affected by a power usually in the range $[1, 4]$.

Inverse distance weighting estimator

$$z^*_{IDW}(\mathbf{u}_0) = \lambda_0^{IDW} + \sum_{i=1}^{n} \lambda_i^{IDW} z(\mathbf{u}_i) \tag{3}$$

$$\lambda_0^{IDW} = 0$$
$$\lambda_i^{IDW} = \frac{1/d_{i0}^w}{\sum_{j=1}^{n} 1/d_{j0}^w}$$

The method tends to give a result similar to the nearest neighbor estimator as the power increases. If the power is low, results are a smooth interpolation between samples. For computational reasons, a small value is added at the denominator of the fractions to ensure that, if the distance is 0, the computer does not undefine that division.

In case of clear anisotropies, distances can be corrected to account for this.

4 Best linear unbiased estimator: kriging

In order to define an estimator that guarantees unbiasedness and minimum spread, we apply sequentially the requirements and deduce how these constrain the determination of the weights. These types of estimators are called BLUE, best linear unbiased estimators.

We start by the definition of a linear estimator in **Equation 1**. To have all the steps in this section, we repeat it here (**Equation 4**). The weights are really a function of the location $\mathbf{u}_0$, however to keep the notation simple, we simply write λ_i instead of $\lambda_i(\mathbf{u}_0)$.

Linear estimator:

$$z^*(\mathbf{u}_0) = \lambda_0 + \sum_{i=1}^{n} \lambda_i z(\mathbf{u}_i) \tag{4}$$

where $\lambda_i, i = 0, \ldots, n$ are constant values in $\mathbb{R}$

Next, we need to impose unbiasedness. This means that on average (over all locations estimated), the estimate does not overestimate or underestimate the true value. Now, how can we impose (and later test) this if we only have one true value at every location?

One way around this problem is to call for the random function model. We said that **we are interpreting each regionalized variable as a random variable, and that these random variables within the domain constitute the random function**, which is characterized by the statis-

tical distribution of the random variables and their statistical and spatial relationships.

So, we rewrite our estimator using random variables (we use upper case letters instead lower case):

$$Z^*(\mathbf{u}_0) = \lambda_0 + \sum_{i=1}^{n} \lambda_i Z(\mathbf{u}_i) \tag{5}$$

We estimate the random variable Z at location $\mathbf{u_0}$ from the available random variables at sample locations: $Z(\mathbf{u}_i), i = 1, \ldots, n$.

Unbiasedness is imposed by setting the expected value of the estimator to be equal to the expected value of the true random variable at the unsampled location:

$$E\{Z^*(\mathbf{u}_0)\} = E\{Z(\mathbf{u}_0)\} \tag{6}$$

$$\Longrightarrow E\left\{\lambda_0 + \sum_{i=1}^{n} \lambda_i Z(\mathbf{u}_i)\right\} = E\{Z(\mathbf{u}_0)\} \tag{7}$$

Finally, optimality is imposed by first defining a measure of quality. In this case, as in many engineering applications, dispersion is penalized with the square of the errors, that is, **the estimator minimizes the variance of the errors**. Again, this is imposed by calling for the random function formalism:

$$\min_{\{\lambda_i, i=0,\ldots,n\}} Var\{Z^*(\mathbf{u}_0) - Z(\mathbf{u}_0)\}\} \tag{8}$$

4.1 Simple kriging

Simple kriging assumes that **the mean is known and constant everywhere** in the domain.

We can impose the three steps to deduce the estimator in this case.

Linearity

The estimate must be linear:

$$Z^*_{SK}(\mathbf{u}_0) = \lambda^{SK}_0 + \sum_{i=1}^{n} \lambda^{SK}_i Z(\mathbf{u}_i)$$

Unbiasedness

The estimate must be unbiased. We use the linearity of the expected value to bring the expected value inside the sum. Recall that the λs are constant:

$$E\{Z^*_{SK}(\mathbf{u}_0)\} = E\{Z(\mathbf{u}_0)\}$$
$$E\left\{\lambda^{SK}_0 + \sum_{i=1}^{n} \lambda^{SK}_i Z(\mathbf{u}_i)\right\} = E\{Z(\mathbf{u}_0)\}$$
$$\lambda^{SK}_0 + \sum_{i=1}^{n} \lambda^{SK}_i E\{Z(\mathbf{u}_i)\} = E\{Z(\mathbf{u}_0)\}$$

Now, we call for our stationarity assumption. We know that the mean is constant everywhere and known, so we denote $E\{Z(\mathbf{u})\} = m$ and replace in the equation:

$$\lambda^{SK}_0 + \sum_{i=1}^{n} \lambda^{SK}_i m = m$$

The only way for this equality to hold is if:

$$\lambda_0^{SK} = \left(1 - \sum_{i=1}^{n} \lambda_i^{SK}\right) m$$

By replacing this value in the estimate, we obtain the simple kriging estimate:

$$Z_{SK}^*(\mathbf{u}_0) = \left(1 - \sum_{i=1}^{n} \lambda_i^{SK}\right) m + \sum_{i=1}^{n} \lambda_i^{SK} Z(\mathbf{u}_i) \qquad (9)$$

Optimality

Finally, we impose optimality:

$$\min_{\{\lambda_i^{SK}, i=0,\ldots,n\}} Var\{Z_{SK}^*(\mathbf{u}_0) - Z(\mathbf{u}_0)\}$$

We start by computing the variance and then we find the minimum:

$$Var\{Z_{SK}^*(\mathbf{u}_0) - Z(\mathbf{u}_0)\} =$$
$$Var\left\{\left(1 - \sum_{i=1}^{n} \lambda_i^{SK}\right) m + \sum_{i=1}^{n} \lambda_i^{SK} Z(\mathbf{u}_i) - Z(\mathbf{u}_0)\right\}$$

In order to develop this expression, we can assume $m = 0$ without loss of generality.

$$Var\{Z^*_{SK}(\mathbf{u}_0) - Z(\mathbf{u}_0)\}$$

$$= E\left\{\left[Z^*_{SK}(\mathbf{u}_0) - Z(\mathbf{u}_0)\right]^2 - [\underbrace{E\left\{Z^*_{SK}(\mathbf{u}_0) - Z(\mathbf{u}_0)\right\}}_{\text{unbiased } =0}]^2\right\}$$

$$= E\left\{\left[Z^*_{SK}(\mathbf{u}_0) - Z(\mathbf{u}_0)\right]^2\right\}$$

$$= E\left\{\left[\sum_{i=1}^{n}\lambda_i^{SK} Z(\mathbf{u}_i) - Z(\mathbf{u}_0)\right]^2\right\}$$

We go back to the initial notation to expand the squared term:

$$Var\{Z^*_{SK}(\mathbf{u}_0) - Z(\mathbf{u}_0)\}$$

$$= E\left\{\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i^{SK}\lambda_j^{SK} Z(\mathbf{u}_i)Z(\mathbf{u}_j) - 2\sum_{i=1}^{n}\lambda_i^{SK} Z(\mathbf{u}_i)Z(\mathbf{u}_0) + (Z(\mathbf{u}_0))^2\right\}$$

and now, we use the linearity of the expected value and bring the expected value inside the sums:

$$Var\{Z^*_{SK}(\mathbf{u}_0) - Z(\mathbf{u}_0)\}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i^{SK}\lambda_j^{SK} E\left\{Z(\mathbf{u}_i)Z(\mathbf{u}_j)\right\}$$

$$-2\sum_{i=1}^{n}\lambda_i^{SK} E\left\{Z(\mathbf{u}_i)Z(\mathbf{u}_0)\right\} + E\left\{(Z(\mathbf{u}_0))^2\right\}$$

Since we are assuming the mean is null, then the following identities can be recognized (notice that we simplify the notation here):

$$E\left\{Z(\mathbf{u}_i)Z(\mathbf{u}_j)\right\} = Cov\{Z(\mathbf{u}_i), Z(\mathbf{u}_j)\} = C_{ij}$$
$$E\left\{Z(\mathbf{u}_i)Z(\mathbf{u}_0)\right\} = Cov\{Z(\mathbf{u}_i), Z(\mathbf{u}_0)\} = C_{i0}$$
$$E\left\{(Z(\mathbf{u}_0))^2\right\} = Var\{Z(\mathbf{u}_0)\} = \sigma_0^2$$

Thus, we want to minimize the following expression:

$$\min_{\{\lambda_i^{SK}, i=1,\ldots,n\}} Var\{Z_{SK}^*(\mathbf{u}_0) - Z(\mathbf{u}_0)\}$$
$$= \min_{\{\lambda_i^{SK}, i=1,\ldots,n\}} \left\{ \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i^{SK}\lambda_j^{SK} C_{ij} - 2\sum_{i=1}^{n} \lambda_i^{SK} C_{i0} + \sigma_0^2 \right\}$$

In order to find the unknown λs, we must take the partial derivatives and make them equal to 0.

$$\min_{\{\lambda_i^{SK}, i=1,\ldots,n\}} \left\{ \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i^{SK}\lambda_j^{SK} C_{ij} - 2\sum_{i=1}^{n} \lambda_i^{SK} C_{i0} + \sigma_0^2 \right\}$$
$$= \min_{\{\lambda_i^{SK}, i=1,\ldots,n\}} f(\lambda_1^{SK}, \ldots, \lambda_n^{SK})$$

$$\Longrightarrow \frac{\partial f(\lambda_1^{SK}, \ldots, \lambda_n^{SK})}{\partial \lambda_i^{SK}} = 0 \quad \forall i = 1, \ldots, n$$

$$\Longrightarrow 2\sum_{j=1}^{n} \lambda_j^{SK} C_{ij} - 2C_{i0} = 0 \quad \forall i = 1, \ldots, n$$

This leads to the following linear system of n equations, known as the simple kriging system:

$$\sum_{j=1}^{n} \lambda_j^{SK} C_{ij} = C_{i0} \quad \forall i = 1, \ldots, n \tag{10}$$

By replacing the optimum λs into the minimized variance, we obtain the simple kriging variance:

$$\sigma_{SK}^2(\mathbf{u}_0) = \sigma_0^2 - \sum_{i=1}^{n} \lambda_i^{SK} C_{i0} \tag{11}$$

Summary

In summary, the simple kriging estimate, the simple kriging variance and the simple kriging system of equations are:

Simple kriging estimator

$$Z_{SK}^*(\mathbf{u}_0) = \left(1 - \sum_{i=1}^{n} \lambda_i^{SK}\right) m + \sum_{i=1}^{n} \lambda_i^{SK} Z(\mathbf{u}_i) \tag{12}$$

$$\sigma_{SK}^2(\mathbf{u}_0) = \sigma_0^2 - \sum_{i=1}^{n} \lambda_i^{SK} C_{i0} \tag{13}$$

$$\sum_{j=1}^{n} \lambda_j^{SK} C_{ij} = C_{i0} \quad \forall i = 1, \ldots, n \tag{14}$$

4.2 Ordinary kriging

Ordinary kriging assumes that **the mean is unknown, but constant within the search neighborhood** from where the samples used for estimation are drawn.

The process to determine the optimum weights is similar to that of simple kriging.

Linearity

The estimate must be linear:

$$Z^*_{OK}(\mathbf{u}_0) = \lambda_0^{OK} + \sum_{i=1}^{n} \lambda_i^{OK} Z(\mathbf{u}_i)$$

Unbiasedness

The estimate must be unbiased. Following the same logic as with simple kriging, we get:

$$E\{Z^*_{OK}(\mathbf{u}_0)\} = E\{Z(\mathbf{u}_0)\}$$

$$E\left\{\lambda_0^{OK} + \sum_{i=1}^{n} \lambda_i^{OK} Z(\mathbf{u}_i)\right\} = E\{Z(\mathbf{u}_0)\}$$

$$\lambda_0^{OK} + \sum_{i=1}^{n} \lambda_i^{OK} E\{Z(\mathbf{u}_i)\} = E\{Z(\mathbf{u}_0)\}$$

Now, we call for our quasi second order stationarity assumption, which means that within the neighborhood where

the samples were found, the mean is approximately constant. But remember that in ordinary kriging we also assume this mean to be unknown, so the only way for this equality to hold is by setting:

$$\lambda_0^{OK} = 0$$
$$\sum_{i=1}^{n} \lambda_i^{OK} = 1$$

By replacing this value in the estimate, we obtain the ordinary kriging estimate:

$$Z_{OK}^*(\mathbf{u}_0) = \sum_{i=1}^{n} \lambda_i^{OK} Z(\mathbf{u}_i) \tag{15}$$

Optimality

We need to impose optimality, but now we have the constraint $\sum_{i=1}^{n} \lambda_i^{OK} = 1$:

$$\min_{\{\lambda_i^{OK}, i=0,\ldots,n\}} Var\{Z_{SK}^*(\mathbf{u}_0) - Z(\mathbf{u}_0)\}$$
$$subject\,to \sum_{i=1}^{n} \lambda_i^{OK} = 1$$

We compute the variance just as we did for simple kriging and we impose the constraint:

$$\min_{\{\lambda_i^{OK}, i=1,\ldots,n\}} Var\{Z_{OK}^*(\mathbf{u}_0) - Z(\mathbf{u}_0)\}$$

$$s.t. \sum_{i=1}^{n} \lambda_i^{OK} = 1$$

$$= \min_{\{\lambda_i^{OK}, i=1,\ldots,n\}} \left\{ \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i^{OK} \lambda_j^{OK} C_{ij} - 2 \sum_{i=1}^{n} \lambda_i^{OK} C_{i0} + \sigma_0^2 \right\}$$

$$s.t. \sum_{i=1}^{n} \lambda_i^{OK} = 1$$

To find the optimum weights, we need to minimize a new function that accounts for the constraint. This is done by "adding zero" to the function, in the form of the constraint, that is, $\sum_{i=1}^{n} \lambda_i^{OK} - 1 = 0$, but weighted by an additional parameter called Lagrange multiplier. We use 2μ, so that the final equations depend on μ only.

$$\min_{\{\lambda_i^{OK}, i=1,\dots,n\}} \left\{ \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i^{OK}\lambda_j^{OK} C_{ij} - 2\sum_{i=1}^{n} \lambda_i^{OK} C_{i0} + \sigma_0^2 \right\}$$

$$+2\mu \left\{ \sum_{i=1}^{n} \lambda_i^{OK} - 1 \right\} = \min_{\{\lambda_i^{OK}, i=1,\dots,n\}} f(\lambda_1^{OK}, \dots, \lambda_n^{OK}, \mu)$$

$$\Longrightarrow \frac{\partial f(\lambda_1^{OK}, \dots, \lambda_n^{OK}, \mu)}{\partial \lambda_i^{OK}} = 0 \quad \forall i = 1, \dots, n$$

$$\frac{\partial f(\lambda_1^{OK}, \dots, \lambda_n^{OK}, \mu)}{\partial \mu} = 0$$

$$\Longrightarrow 2\sum_{j=1}^{n} \lambda_j^{OK} C_{ij} - 2C_{i0} + 2\mu = 0 \quad \forall i = 1, \dots, n$$

$$\sum_{i=1}^{n} \lambda_i^{OK} = 1$$

This leads to the following linear system of $n+1$ equations, known as the ordinary kriging system:

$$\sum_{j=1}^{n} \lambda_j^{OK} C_{ij} + \mu = C_{i0} \quad \forall i = 1, \dots, n \qquad (16)$$

$$\sum_{i=1}^{n} \lambda_i^{OK} = 1$$

Again, we can obtain the "optimized" variance of the error, by replacing the optimum λs into the minimized variance. This is known as the ordinary kriging variance:

$$\sigma^2_{OK}(\mathbf{u}_0) = \sigma^2_0 - \sum_{i=1}^{n} \lambda_i^{OK} C_{i0} - \mu \tag{17}$$

Summary

In summary, the ordinary kriging estimate, the ordinary kriging variance and the ordinary kriging system of equations are:

Ordinary kriging estimator

$$Z^*_{OK}(\mathbf{u}_0) = \sum_{i=1}^{n} \lambda_i^{OK} Z(\mathbf{u}_i) \tag{18}$$

$$\sigma^2_{OK}(\mathbf{u}_0) = \sigma^2_0 - \sum_{i=1}^{n} \lambda_i^{OK} C_{i0} - \mu \tag{19}$$

$$\sum_{j=1}^{n} \lambda_j^{OK} C_{ij} + \mu = C_{i0} \quad \forall i = 1, \ldots, n \tag{20}$$

$$\sum_{i=1}^{n} \lambda_i^{OK} = 1$$

4.3 Interpretation of kriging

We will now analyze the kriging estimate, kriging variance and kriging system of equations to try to understand intuitively how it works.

The weights and the mean

For **simple kriging**, the kriging estimate is a linear combination of the samples and of the global mean m:

$$Z^*_{SK}(\mathbf{u}_0) = \underbrace{\left(1 - \sum_{i=1}^{n} \lambda_i^{SK}\right)}_{\lambda_m} m + \sum_{i=1}^{n} \lambda_i^{SK} Z(\mathbf{u}_i) \qquad (21)$$

Notice that the sum of the weights is equal to one:

$$\lambda_m + \sum_{i=1}^{n} \lambda_i^{SK} = \left(1 - \sum_{i=1}^{n} \lambda_i^{SK}\right) + \sum_{i=1}^{N} \lambda_i^{SK} = 1$$

Recall that this condition was explicitly imposed in **ordinary kriging**, but not in **simple kriging**.

Also, notice the extreme cases:

- If the sample data are not relevant, their weight will be 0, and all the weight will be assigned to the mean m. The best estimate is the mean, as this is all we know: the sample belongs to the domain and that domain has mean m, but no conditioning data is close enough to change this prior knowledge.
- If the estimation location $\mathbf{u}_0$ coincides with a sample value, all the weight is assigned to that sample, so the mean and all the other samples become irrelevant.

In ordinary kriging, the mean is unknown (and absent from the expression of the estimate). Therefore, the weight is shared among the samples. Even if none of the samples is relevant for the estimation location, they help inferring the local mean and ordinary kriging will simply average them depending on their spatial redundancy.

The kriging variance

The kriging variance measures the quality of the estimation. However, as it can be seen from its equation, it does not depend on the sample values, but only on the spatial configuration of the sample locations (and on the covariance function).

Therefore, the kriging variance does not capture a feature known as proportional effect, which is that the variability tend to be a function of the value of the variable. For instance, low grades tend to be less variable than high grades in an ore deposit, and the relationship between the estimated mean and standard deviation is linear (in the case of a variable with a lognormal distribution, as is the case with most grades). This property of the variance to be dependent on the mean is also called heteroscedasticity.

Let us review the kriging variance in simple and ordinary kriging and see the extreme cases, to better understand how it behaves.

The simple kriging variance is:

$$\sigma^2_{SK}(\mathbf{u}_0) = \sigma^2_0 - \sum_{i=1}^{n} \lambda_i^{SK} C_{i0}$$

In the case of ordinary kriging, it is:

$$\sigma^2_{OK}(\mathbf{u}_0) = \sigma^2_0 - \sum_{i=1}^{n} \lambda_i^{OK} C_{i0} - \mu$$

The extreme cases are:

- If the sample data are not relevant, their weight in simple kriging will be 0, and also the covariances between the estimation location and the sample, and therefore the resulting estimation variance will be σ^2_0. The expected error if we have no relevant local information will be the error expected for the domain, that is, as with the estimate, we go back to the prior knowledge. In the case of ordinary kriging, the sum of the weights must be equal to 1, so we cannot have all the weights equal to zero. however, the covariances involved will be zero. So the resulting kriging variance will be equal to $\sigma^2_0 - \mu$. Notice that μ can be positive or negative.
- If the estimation location $\mathbf{u}_0$ coincides with a sample value, the weight of that sample in kriging will be 1 and the covariance between the estimation location and that sample is equal to the variance, therefore, the kriging variance is 0. This reflects no uncertainty about the location, since we know its true value. In ordinary kriging, if this happens, the Lagrange multiplier takes a value of 0.

The system of equations: redundancy vs closeness

To better understand the systems of equations, these can be written in matrix form.

The $(n \times n)$ simple kriging system of equations presented in **Equation 14** can be written as:

$$\underbrace{\begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} \end{bmatrix}}_{\text{covariances between the samples}} \underbrace{\begin{pmatrix} \lambda_1^{SK} \\ \lambda_2^{SK} \\ \vdots \\ \lambda_n^{SK} \end{pmatrix}}_{\text{weights}} = \underbrace{\begin{pmatrix} C_{10} \\ C_{20} \\ \vdots \\ C_{n0} \end{pmatrix}}_{\text{covariances between samples and estimation location}} \tag{22}$$

$$\left[\mathbf{C}\right]\left(\lambda^{\mathbf{SK}}\right) = \left(\mathbf{k}\right) \tag{23}$$

It can be seen that the left hand side matrix of covariances measures the redundancy between the samples and does not involve the location where we want to estimate. The vector of weights is the unknown. The right hand side covariance vector contains the closeness term of the system of equations, and involves the "distance" between each sample and the location we are estimating.

We could rewrite the system of equations as:

$$\left[Redundancy\right]\left(Weights\right) = \left(Closeness\right)$$

This indicates that kriging balances the redundancy of the sample information along with the closeness of the samples to the location estimated, to determine the unknown weights.

Notice that, to solve for the unknown weights we just need to multiply on the left by the inverse of the matrix of

covariances:

$$\left[\mathbf{C}\right]\left(\lambda^{\mathbf{SK}}\right)=\left(\mathbf{k}\right) \qquad \Big/\left[\mathbf{C}^{-1}\right]\cdot \tag{24}$$

$$\left[\mathbf{C}^{-1}\right]\cdot\left[\mathbf{C}\right]\left(\lambda^{\mathbf{SK}}\right)=\left[\mathbf{C}^{-1}\right]\left(\mathbf{k}\right) \tag{25}$$

$$\left[\mathbf{I}\right]\left(\lambda^{\mathbf{SK}}\right)=\left[\mathbf{C}^{-1}\right]\left(\mathbf{K}\right) \tag{26}$$

$$\left(\lambda^{\mathbf{SK}}\right)=\left[\mathbf{C}^{-1}\right]\left(\mathbf{K}\right) \tag{27}$$

In summary, the unknown weights can be obtained by multiplyng the inverse of the matrix of covariances between the samples, by the vector of covariances between the sample locations and the estimation location.

The interpretation and computation is similar for ordinary kriging, but this system has $((n+1)\times(n+1))$ equations. For completeness, here is how the system of equations looks in matrix form:

$$\begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} & 1 \\ C_{21} & C_{22} & \cdots & C_{2n} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix} \begin{pmatrix} \lambda_1^{OK} \\ \lambda_2^{OK} \\ \vdots \\ \lambda_n^{OK} \\ \mu \end{pmatrix} = \begin{pmatrix} C_{10} \\ C_{20} \\ \vdots \\ C_{n0} \\ 0 \end{pmatrix} \tag{28}$$

$$\left[\mathbf{C}\right]\left(\lambda^{\mathbf{OK}}\right)=\left(\mathbf{k}\right) \tag{29}$$

It can be seen that the additional column and row imposes the unbiasedness constraint that the sum of the weights be equal to 1.

4.4 Other types of kriging

As seen in the previous sections, kriging is a general framework to define estimators that satisfy some conditions. These conditions are typically linearity, unbiasedness and optimality, specifically minimization of the error variance.

Under different assumptions of stationarity the estimate can change. We saw the difference between simple and ordinary kriging. In the case of simple kriging, the mean is assumed known and constant over the entire domain, which is a very strict condition almost never met in practice. For ordinary kriging, the condition is for the mean to be constant within each search neighborhood. Because it changes everywhere, we also consider it unknown, which has an effect over the conditions that apply over the weights λs.

Ordinary kriging is widely used in practice because it adapts to local changes in the variable.

Many other approaches can be considered and similar deductions of the estimators are possible:

- Simple kriging with locally varying mean: one simple extension of simple kriging is to assume we know the mean everywhere, but instead of being a constant value, it changes with location. So, we replace m with $m(\mathbf{u}_0)$ in the simple kriging estimate.
- Kriging with a trend: an extension of ordinary kriging is to assume the local mean is some polynomial form of the coordinates, instead of a constant within the neighborhood. This adds unknowns to the problem, namely, the coefficients of the polynomial trend. The problem can be solved in a similar fashion as with ordinary kriging, but additional constraints will appear in the system

of equations and new Lagrange multipliers need to be added to solve the minimization problem, and these appear in the system of equations.

- Kriging of a transform of the original data: this is probably one of the most interesting possibilities. The original data can be transformed and the optimization problem can be stated in the original variable, which will modify the estimator and system of equations. Some examples of transformations are:

  - Logarithmic transformation
  - Gaussian transformation
  - Rank transformation
  - Indicator transformation

There are many more variants of kriging, but these are seldom used in practice.

5 Block estimation

In many cases, the estimation support, that is the volume over which the estimated value is assessed, is different than the sample support (the volume of the sample). In mining, for example, estimation is done over a block model, where the domain is divided into relatively large blocks that represent selective units, over which decisions are made about processing. The samples, on the other hand, represent the volume of the drillhole at a regular length (called composite). The drillhole is a cylinder only a few centimeters in diameter, while the blocks are several meters in each dimension.

For additive variables, the estimated value of the block is nothing but the average of the estimated values over points within the block (here points refer to the support of the sample information).

We can analyze how kriging changes, when we consider estimation over a block. Recall the general structure of the kriging equations:

$$\underbrace{\begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} & 1 \\ C_{21} & C_{22} & \cdots & C_{2n} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix}}_{\text{does not depend on the estimation support}} \begin{pmatrix} \lambda_1^{OK} \\ \lambda_2^{OK} \\ \vdots \\ \lambda_n^{OK} \\ \mu \end{pmatrix} = \underbrace{\begin{pmatrix} C_{10} \\ C_{20} \\ \vdots \\ C_{n0} \\ 0 \end{pmatrix}}_{\text{depends on the estimation support}} \tag{30}$$

We can see that only the right hand side vector depends on the estimation location support.

Estimation at block support could be done in two ways:

- Estimate several points within the volume V of interest and then average the estimated values.
- Adjust the estimate to account for the volume of the estimate.

These two results are equivalent, and computationally, the second approach is simpler, since the covariance matrix is inverted only once for each estimated block.

Basically, the system of equations changes to:

$$\begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} & 1 \\ C_{21} & C_{22} & \cdots & C_{2n} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix} \begin{pmatrix} \lambda_1^{OK} \\ \lambda_2^{OK} \\ \vdots \\ \lambda_n^{OK} \\ \mu \end{pmatrix} = \begin{pmatrix} \overline{C_{1V}} \\ \overline{C_{2V}} \\ \vdots \\ \overline{C_{nV}} \\ 0 \end{pmatrix} \tag{31}$$

where $\overline{C_{iV}}$ is the averge covariance between the sample located at $\mathbf{u}_i$ and the volume of interest V, which we can assume is centered at $\mathbf{u}_0$.

This average covariance should consider all possible vectors from the sample to the volume (as depicted in **Figure 2**).

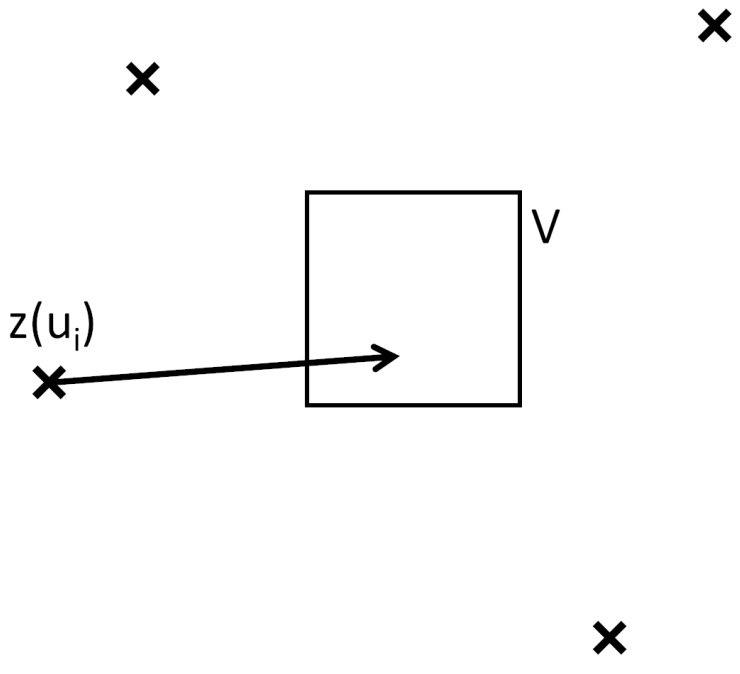

Figure 2: The relationship between a sample and the volume V.

This requires computing the following integral:

$$\overline{C_{iV}} = \frac{1}{|V|}\int_V C(\mathbf{u}_i, \mathbf{u}_V)d\mathbf{u}_V \tag{32}$$

In order to avoid this (painful) integral, we discretize the block into a sufficient number of points, and approximate this average covariance (**Figure 3**):

$$\overline{C_{iV}} \approx \frac{1}{n_{disc}}\sum_{j=1}^{n_{disc}} C(\mathbf{u}_i, \mathbf{u}_{j\in V}) \tag{33}$$

where the points $\{\mathbf{u}_{j\in V}, j = 1, \ldots, n_{disc}\}$ discretize the volume V as shown.

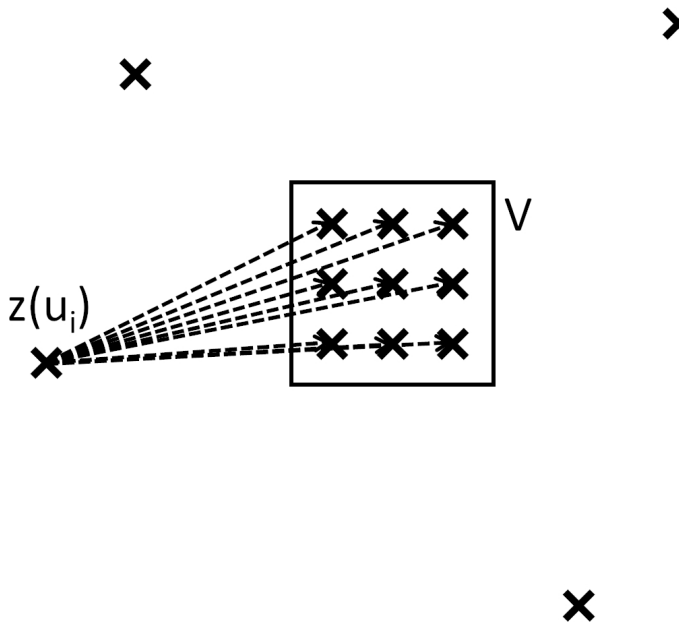

Figure 3: The discretization of the block to compute the average covariance.

The number of discretization points required to approximate the true integral value with a low error depends on

the spatial continuity, but as a rule of thumb, discretizations of 4×4 in 2D and $3 \times 3 \times 2$ in 3D are enough. Notice that usually, the vertical discretization in 3D will depend on the composite size (the length of the sample in that direction, when considering vertical drillholes). For example, if the composites (samples) are 10m in length and the block is 10m in height, then the vertical discretization is assumed to be 1. In that case, the X and Y discretizations can be made 4×4, to end up with $4 \times 4 \times 1$.

Index