Mathematical Geology, Vol. 9, No. 6, 1977

Kriging in Terms of Projections[1]

A. G. Journel[2]

In the last few years, an increasing number of practical studies using so-called kriging estimation procedures have been published. Various terms, such as universal kriging, lognormal kriging, ordinary kriging, etc., are used to define different estimation procedures, leaving a certain confusion about what kriging really is. The object of this paper is to show what is the common backbone of all these estimation procedures, thus justifying the common name of kriging procedures. The word "kriging" (in French "krigeage") is a concise and convenient term to designate the classical procedure of selecting, within a given *class of possible estimators, the estimator with a minimum estimation variance (i.e., the estimator which leads to a minimum variance of the resulting estimation error). This estimation variance can be seen as a squared distance between the unknown value and its estimator; the process of minimization of this distance can then be seen as the projection of the unknown value onto the space within which the search for an estimator is carried out.* KEY WORDS: kriging, geostatistics.

BACKGROUND

Some History

In mining practice, one problem is to find the best possible estimator of the mean grade of a block, taking into account the assay values of the different samples available either inside or outside the block to be estimated. In the early 1950's, D. G. Krige (1951) proposed a regression procedure to assign a weight to each sample assay, the block grade estimator being a linear combination of the available assays. This original regression procedure is recalled in Matheron (1971, p. 118). In 1963, Matheron formalized and generalized this regression procedure and gave it the name of "kriging." According to the original definition given by Matheron, kriging is the probabilistic process of obtaining the best linear unbiased estimator of an unknown variable, "best" being taken here in the sense of minimization of the resulting estimation variance (or variance of the resulting estimation error). Then a second major generalization was attained when Matheron (1973, 1975a, 1975b) studied various procedures to obtain unbiased *non* linear estimators; he

[1] Manuscript received 15 October 1976; revised 9 March 1977.

[2] Ecole Nationale Supérieure des Mines de Paris, Centre de Géostatistique, Fontainebleau, France, 77305.

gave them names such as disjunctive kriging and kriging of transformed variables. The word "kriging" was kept because all these nonlinear procedures and the initial linear kriging process (to obtain linear estimators) have in common the principle of minimization of the estimation variance. Hence, kriging per se should be redefined as a probabilistic theory of estimation based on the principle of minimization of the estimation variance.

Need for Precise Terminology

This theory of estimation gives rise to numerous practical processes to obtain estimators, and these processes can keep the generic name of kriging provided that an additional expression makes clear which particular process is used; for example, linear kriging with unknown stationary expectation for the process of obtaining the best linear unbiased estimator of an unknown variable of unknown expectation. One can then understand why practitioners are reluctant to use such a long definition; they would simply say "kriging" or use condensed expressions such as "ordinary kriging," "universal kriging," "lognormal kriging," etc. Certainly this appears to be confusing to someone not quite accustomed to both the theory and various practical kriging processes. The object of this paper is to give an overall view of the probabilistic theory of kriging which will allow the linkage between the various practical processes that practitioners have been using under various names for some 20 years. An attempt will be made to settle upon a unique terminology.

What Is New About Kriging?

A point must be stressed: the principle of building an estimator by minimization of a squared norm (called here estimation variance) is not new; it has been well known to statisticians since the works of Wiener (1949). What is truly new is the practical day-to-day application of this principle in fields as varied as mining estimations, forest surveys, cartography, meteorology, etc. Hence, as regards their works in the field of estimation, Krige, Matheron, and their fellows should be considered as craftsmen who not only cast the operational tools but made the effort to put them into practice.

Kriging in Terms of Projection

Let $A = \{Z(x),\ x \in D\}$ be a set of random variables $Z(x)$ defined at each point x (of a three-dimensional deposit D for example). Let $\mathscr{E}$ be the vector space defined as the set of all finite linear combinations of the elements of A:

$$\mathscr{E} : \{\lambda^{\beta}\, Z(x_{\beta}) \ ;\ Z(x_{\beta}) \in A \ ,\ \lambda_{\beta} \text{ real}\}$$

The neutral element $0 = (\lambda = 0)\cdot Z(x)$ of $\mathscr{E}$ is the random variable almost certainly null.

Notation

Throughout this paper the condensed notation $\lambda^\alpha Z_\alpha$ will be used to represent the sum $\sum_\alpha \lambda_\alpha Z(x_\alpha)$, and $\lambda^\alpha \lambda^\beta \sigma_{\alpha\beta}$ to represent the sum $\sum_\alpha \sum_\beta \lambda_\alpha \lambda_\beta \sigma_{\alpha\beta}$.

The vector space $\mathscr{E}$ is provided with a scalar product equal to the *noncentered* covariance (which is not necessarily stationary):

$$\langle Z(x), Z(y)\rangle = E\{Z(x)\, Z(y)\} = \sigma_{xy}$$

In practical terms, this covariance characterizes the spatial correlation between two variables $Z(x)$ and $Z(y)$ located at two different points x and y. Let $m(x) = E\{Z(x)\}$ be the expected value of $Z(x)$; the centered covariance is then written:

$$E\{[Z(x)-m(x)][Z(y)-m(y)]\} = \sigma'_{xy} = \sigma_{xy}-m(x)\cdot m(y)$$

The norm $\|Z(x)\|$ of a vector $Z(x)$ is defined as the positive square root of $\langle Z(x), Z(x)\rangle$, i.e., $\|Z(x)\|^2 = \langle Z(x),Z(x)\rangle$. The distance between two elements $Z(x)$ and $Z(y)$ is defined as the norm $\|Z(x)-Z(y)\|$ of the vector $Z(x)-Z(y)$.

Let $Z(x_0) \in \mathscr{E}$ be any given unknown variable, and let $\mathscr{E}' \subset \mathscr{E}$ be any vector subspace (or less restrictively any linear manifold). One can then prove that there exists one and only one element $Z^* \in \mathscr{E}'$ which satisfies the minimum of the distance $\|Z-Z^*\|$; this unique element Z^* is then called the projection of $Z(x_0)$ onto the subspace $\mathscr{E}'$; see Figure 1.

The kriging process is nothing but this projection of an unknown value onto a particular subspace $\mathscr{E}' \subset \mathscr{E}$ within which the search for an estimator Z^* is carried out. The corresponding minimum distance $E\{[Z(x_0)-Z^*]^2\}$ is called minimum estimation variance or simply kriging variance.

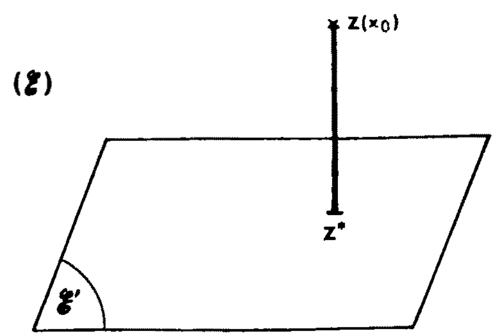

Figure 1. The kriging estimator Z^* defined as the projection of the unknown $Z(x_0)$ onto the subspace $\mathscr{E}' \subset \mathscr{E}$.

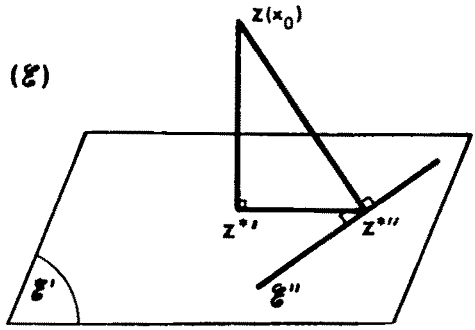

Figure 2. Projections of the unknown $Z(x_0)$ onto the two subspaces $\mathscr{E}'' \subset \mathscr{E}'$.

There are as many kriging processes and corresponding kriging estimators Z^* as there are different sets $\mathscr{E}' \subset \mathscr{E}$ within which the projection of the unknown $Z(x_0)$ is to be made. Consider the two sets $\mathscr{E}''$ and $\mathscr{E}'$ of Figure 2, with $\mathscr{E}'' \subset \mathscr{E}' \subset \mathscr{E}$, and the two corresponding kriging estimators $Z^{*''}$ and $Z^{*'}$: as $\mathscr{E}'$ includes $\mathscr{E}''$, the projection $Z^{*'}$ is nearer to the unknown $Z(x_0)$ than $Z^{*''}$. Hence, as the set where the search for the estimator is carried out is wider, the estimation will be better. This remark is a prelude to the classification of the various kriging estimators.

LINEAR KRIGING PROCESSES

Let us first consider the class of linear estimators, i.e., the vector subspace $\mathscr{E}_{n+1} \subset \mathscr{E}$, of dimension $(n+1)$, generated by the linear combinations

$$\sum_{\alpha=1}^{n} \lambda_\alpha Z_\alpha + \lambda_0 \cdot 1$$

of n particular variables $\{Z_\alpha = Z(x_\alpha),\ \alpha = 1, \ldots, n\}$ called data plus the constant 1. The linear kriging processes are defined as the processes of projecting the unknown $Z(x_0)$ either onto $\mathscr{E}_{n+1}$ itself, or onto any linear manifold $C \subset \mathscr{E}_{n+1}$. The restrictions of $\mathscr{E}_{n+1}$ to various linear manifolds C guarantees the unbiasedness of the estimator Z^*, i.e., $E\{Z^*\} = E\{Z(x_0)\}$ as will be shown.

Conditions for the Unbiasedness of the Estimator

Consider the expectation of any element $Z^* = \lambda_0 + \lambda^\alpha Z_\alpha \in \mathscr{E}_{n+1}$:

$$E\{Z^*\} = \lambda_0 + \lambda^\alpha E\{Z_\alpha\} = \lambda_0 + \lambda^\alpha m(x_\alpha)$$

The element Z^*, considered as an estimator of $Z(x_0)$, is unbiased if and only if $E\{Z^*\} = E\{Z(x_0)\}$, i.e.,

$$\lambda_0 + \sum_{\alpha=1}^{n} \lambda_\alpha m(x_\alpha) = m(x_0) \tag{1}$$

Various cases are distinguished according to whether the expected values $m(x_0)$, $m(x_\alpha)$ are known or not, and in the second case whether $m(x)$ is stationary or not.

Case 1. All the expectations are known (it does not matter then whether they are stationary or not). The unbiasedness of the estimator Z^* is then characterized by the single condition (1), i.e.,

$$\lambda_0 = m(x_0) - \lambda^\alpha\, m(x_\alpha)$$

Case 2. The expectation is stationary but unknown, i.e., $m(x) = m(x_\alpha)$ = unknown constant m, $\forall\ x, x_\alpha$. The unbiasedness condition (1) is then satisfied if and only if $\lambda_0 = 0$ and

$$\sum_{\alpha=1}^{n} \lambda_\alpha = 1$$

The first condition, $\lambda_0 = 0$, amounts to restricting the set of the possible estimators to the vector subspace $\mathscr{E}_n \subset \mathscr{E}_{n+1}$. The subspace $\mathscr{E}_n$, of dimension n, is generated by the linear combinations $\lambda^\alpha Z_\alpha$ of the n data only. The second condition

$$\sum_{\alpha=1}^{n} \lambda_\alpha = 1$$

amounts to restricting $\mathscr{E}_n$ to the linear manifold C_1 defined by the condition $\sum_\alpha \lambda_\alpha = 1$ on the weights λ_α. Hence in the presence of an unknown stationary expectation, the unbiasedness condition (1) demands that the linear estimator Z^* be an element of the linear manifold : $Z^* \in C_1 \subset \mathscr{E}n \subset \mathscr{E}n_{+1}$.

Case 3. The expectation is neither stationary nor known. We are then at a loss to express the unbiasedness relation (1). It is necessary to provide the form of the expectation $m(x)$; for example, $m(x)$ is an unknown linear combination of L known functions $f_l(x)$:

$$m(x) = \sum_{l=1}^{L} a_l f_l(x)$$

the L parameters a_l being unknown. The unbiasedness condition (1) can then be written:

$$\lambda_0 + \sum_l a_l \sum_\alpha \lambda_\alpha f_l(x_\alpha) = \sum_l a_l f_l(x_0)$$

This relation is verified, whatever the unknown parameters a_l are, if and only if:

$$\begin{cases} \lambda_0 = 0, & \text{which amounts to the restriction } \mathscr{E}_n \subset \mathscr{E}_{n+1} \\ \sum_{\alpha=1}^{n} \lambda_\alpha f_l(x_\alpha) = f(x_0), & \forall l = 1,\ldots,L \end{cases}$$

The last L conditions amount to a restriction of $\mathscr{E}_n$ to the linear manifold C_L, of dimension $n-L$. Note that the previous stationary case corresponds to constant expectation of the form:

$$m(x) = a_1 f_1(x), \quad \text{with } a_1 = m \text{ and } f_1(x) = 1, \forall x$$

The corresponding linear manifold C_1 is then of dimension $n-1$, and defined by the unique condition $\sum_\alpha \lambda_\alpha = 1$.

Note the inclusions: $C_L \subset C_{L-1} \subset \ldots \subset C_1 \subset \mathscr{E}_n \subset \mathscr{E}_{n-1}$. The wider the set onto which the projection is done, the nearer will be the corresponding projected value Z^* to the unknown $Z(x_0)$; see Figure 2. Each of these projection sets gives rise to a particular linear kriging process. Let us start considering the wider space, i.e., $\mathscr{E}_{n+1}$.

Linear Kriging Process with Known Expectations (Simple Kriging)

The kriging estimator $Z^*_{K_0}$ is of the form

$$Z^*_{K_0} = \lambda_0 + \lambda^\alpha_{K_0\alpha} Z_\alpha$$

and is the projection of the unknown $Z(x_0)$ onto the vector space $\mathscr{E}_{n+1}$; see Figure 3. This projection is unique and is characterized by the orthogonality of the vector $Z(x_0)-Z^*_{K_0}$ to each of the $(n+1)$ vectors generating $\mathscr{E}_{n+1}$:

$$\begin{cases} \langle Z(x_0)-Z^*_{K_0},\ 1\rangle = 0 \\ \langle Z(x_0)-Z^*_{K_0}, Z_\alpha\rangle = 0, & \forall \alpha = 1,\ldots,n \end{cases}$$

With $$Z^*_{K_0} = \lambda_0 + \lambda^\alpha_{K_0} Z_\alpha$$

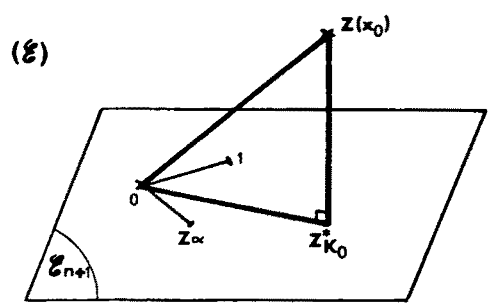

Figure 3. Projection of the unknown $Z(x_0)$ onto $\mathscr{E}_{n+1} \subset \mathscr{E}$.

the preceding relations of orthogonality give rise to the following $(n+1)$ linear equations with $(n+1)$ unknowns (the weights λ_0, and $\lambda_{K_0\alpha}$):

$$\begin{cases} \lambda_0 + \lambda_{K_0}^{\beta} m(x_\beta) = m(x_0) \\ \lambda_0 m_0(x_\alpha) + \lambda_{K_0}^{\beta} \sigma_{\alpha\beta} = \sigma_{\alpha x_0}, \qquad \forall \alpha = 1, \ldots, n \end{cases} \tag{2}$$

with

$$m(x) = E\{Z(x)\} = \langle Z(x), 1 \rangle$$

and $\sigma_{\alpha\beta}$ and $\sigma_{\alpha\chi}$ being the noncentered covariances.

Thus

$$\sigma_{\alpha\beta} = \langle Z(x_\alpha), Z(x_\beta) \rangle$$

Note that the first equation of system (2) is nothing but the relationship (1) ensuring the unbiasedness of the estimator $Z^* \in \mathscr{E}_{n+1}$. Multiplying the elements of the first equation of system (2) by $m(x_\alpha)$ and subtracting this equation from the $(\alpha+1)$th equation (written for α) gives

$$\lambda_{K_0}^{\beta}[\sigma_{\alpha\beta} - m(x_\alpha) \cdot m(x_\beta)] = \sigma_{\alpha x_0} - m(x_\alpha) \cdot m(x_0), \qquad \forall \alpha = 1, \ldots, n$$

The last n equations of system (2) can thus be written in terms of the centered covariance $\sigma'_{\alpha\beta} = \sigma_{\alpha\beta} - m(x_\alpha) \cdot m(x_\beta)$:

$$\sum_{\beta=1}^{n} \lambda_{K_0\beta} \sigma'_{\alpha\beta} = \sigma'_{\alpha x_0}, \qquad \forall \alpha = 1, \ldots, n \tag{3}$$

This system (3) of n linear equations is called a system of linear kriging with known expectations, or a system of simple kriging.

Once the n weights $\lambda_{K_0\alpha}$ are determined by solving the kriging system (3), the weight λ_0 is given by the first equation of system (2) and the kriging estimator $Z_{K_0}^*$ is then written:

$$Z_{K_0}^* = m(x_0) - \lambda_{K_0}^{\alpha} m(x_\alpha) + \lambda_{K_0}^{\alpha} Z_\alpha = \lambda_{K_0}^{\alpha}[Z_\alpha - m(x_\alpha)] + m(x_0)$$

The centered estimator $Z_{K_0}^* - m(x_0)$ thus appears as a linear combination $\lambda_{K_0}^{\alpha}[Z_\alpha - m(x_\alpha)]$ of the centered data; more precisely, $Z_{K_0}^* - m(x_0)$ is the projection of the unknown $Z(x_0) - m(x_0)$ on the vector space $\mathscr{E}'_n$ generated by the linear combinations of the n centered data.

Estimation Variance

The estimation variance $E\{[Z(x_0) - Z_{K_0}^*]^2\}$ is nothing but the minimum squared distance $\|Z(x_0) - Z_{K_0}^*\|^2$. This minimum estimation variance is also called kriging variance and is denoted by $\sigma_{K_0}^2$:

$$\sigma^2_{K_0} = \|Z(x_0) - Z^*_{K_0}\|^2 = \|Z(x_0) - m(x_0) - \lambda^\alpha_{K_0}[Z_\alpha - m(x_\alpha)]\|^2$$

with

$$\|Z(x_0) - m(x_0)\|^2 = \sigma'_{x_0x_0}$$

$$\lambda^\alpha_{K_0}\lambda^\beta_{K_0}\langle Z(x_\alpha) - m(x_\alpha), Z_\beta - m(x_\beta)\rangle = \lambda^\alpha_{K_0}\lambda^\beta_{K_0}\sigma'_{\alpha\beta} = \lambda^\alpha_{K_0}\sigma'_{\alpha x_0}$$

$$-2\lambda^\alpha_{K_0}\langle Z(x_0) - m(x_0), Z_\alpha - m(x_\alpha)\rangle = 2\lambda^\alpha_{K_0}\sigma'_{\alpha x_0}$$

The kriging variance becomes

$$\sigma^2_{K_0} = \sigma'_{x_0x_0} - \lambda^\alpha_{K_0}\sigma'_{\alpha x_0} \tag{4}$$

What Are the Prerequisites for Obtaining $Z^*_{K_0}$?

All the expected values, $m(x_0)$, $m(x_\alpha)$, $\alpha = 1, \ldots, n$, must be known as well as the covariance function, centered or not.

Linear Kriging Process with Unknown Stationary Expectation (Ordinary Kriging)

When

$$E\{Z(x)\} = m(x) = m(x_\alpha) = \text{unknown constant } m, \ \forall\, x, x_\alpha$$

we see that the unbiasedness of a linear estimator $\lambda^\alpha Z_\alpha \in \mathscr{E}_n$ restricts the search for this estimator to the linear manifold $C_1 \subset \mathscr{E}_n$ defined by the unique condition $\sum_\alpha \lambda_\alpha = 1$; see Figure 4.

The kriging estimator $Z^*_{K_1} = \lambda^\alpha_{K_1} Z_\alpha$ is then defined as the projection of the unknown $Z(x_0)$ onto the linear manifold $C_1 \subset \mathscr{E}_n$. This projection is unique and characterized by conditions:

(a) $Z^*_{K_1}$ belongs to C_1, i.e., $\sum_\alpha \lambda_{K_1\alpha} = 1$, and

(b) the vector $Z^*_{K_1} - Z(x_0)$ is orthogonal to any difference $Z_\alpha - \overline{Y}$,

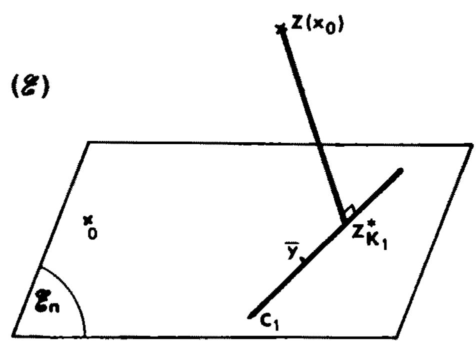

Figure 4. Projection of $Z(x_0)$ onto the linear manifold $C_1 \subset \mathscr{E}_n \subset \mathscr{E}$.

with Z_α, $\bar{Y} \in C_1$. $\bar{Y}$ being an arbitrary element of C_1 [for instance the mean value $\bar{Y} = (1/N) \sum_\alpha Z_\alpha$], the previous orthogonality condition (b) is characterized by the orthogonality of $Z^*_{K_1} - Z(x_0)$ to each of the n vectors $Z_\alpha - \bar{Y}$, i.e.,

$$\langle Z^*_{K_1} - Z(x_0), Z_\alpha - \bar{Y} \rangle = 0, \qquad \forall \alpha = 1, \ldots, n$$

By putting $\mu_1 = \langle Z^*_{K_1} - Z(x_0), \bar{Y} \rangle$, condition (b) may be rewritten:

$$\langle Z^*_{K_1}, Z_\alpha \rangle - \mu_1 = \langle Z(x_0), Z_\alpha \rangle, \qquad \forall \alpha = 1, \ldots, n$$

Finally we obtain the system of $(n+1)$ linear equations with $(n+1)$ unknowns (the n weights $\lambda_{K_1\alpha}$ and the parameter μ_1):

$$\begin{cases} \sum_\beta \lambda_{K_1\beta} = 1 \\ \lambda^\beta_{K_1} \sigma_{\alpha\beta} - \mu_1 = \sigma_{\alpha x_0}, \qquad \forall \alpha = 1, \ldots, n \end{cases} \tag{5}$$

This system (5) is called a system of linear kriging with unknown stationary expectation, or a system of linear kriging with an unbiasedness condition of order 1 (for the unbiasedness is guaranteed by the single condition $\sum_\beta \lambda_{K_1\beta} = 1$, on the n weights $\lambda_{K_1\beta}$). We propose to use the shorter name *ordinary kriging* system or process, because this kriging process is the most commonly used, at least in mining practice.

Estimation Variance

The corresponding minimum estimation variance $E\{[Z(x_0) - Z^*_{K_1}]^2\}$, called kriging variance, is written:

$$\sigma^2_{K_1} = \|Z(x_0) - Z^*_{K_1}\|^2 = \|Z(x_0)\|^2 + \|Z^*_{K_1}\|^2 - 2\langle Z(x_0), Z^*_{K_1} \rangle$$

With

$$\|Z(x_0)\|^2 = \sigma_{x_0 x_0}, \qquad \|Z^*_{K_1}\|^2 = \lambda^\alpha_{K_1} \lambda^\beta_{K_1} \sigma_{\alpha\beta} = \lambda^\alpha_{K_1} \sigma_{\alpha x_0} + \mu_1$$

and

$$\langle Z(x_0), Z^*_{K_1} \rangle = \lambda^\alpha_{K_1} \sigma_{\alpha x_0}$$

the kriging variance becomes:

$$\sigma^2_{K_1} = \sigma_{x_0 x_0} - \lambda^\alpha_{K_1} \sigma_{\alpha x_0} + \mu_1 \tag{6}$$

Note that the system of equations (5) and the expression (6) of the kriging variance can be stated replacing the noncentered covariance by the centered covariance $\sigma'_{xy} = \sigma_{xy} - m^2$. The unknown value m^2 is eliminated from these equations thanks to the unbiasedness condition

$$\sum_\beta \lambda_{K_1\beta} = 1$$

Similarly, these equations can also be written replacing the covariance function σ_{xy} by the semivariogram function γ_{xy} defined by:

$$\gamma_{xy} = \tfrac{1}{2}[\sigma'_{xx}+\sigma'_{yy}-\sigma'_{xy}] = \tfrac{1}{2}E\{[Z(x)-Z(y)]^2\}$$

In practice the inference of the semivariogram function is easier than the inference of the corresponding covariance function (centered or not).

What Are the Prerequisites for Obtaining $Z^*_{K_1}$?

The expected value no longer needs to be known, but is assumed to be stationary. The knowledge of the covariance function (centered or not), $\sigma_{xy} = \langle Z(x), Z(y)\rangle$, is required.

Algebraic Proof of the Kriging System (5)

The kriging estimator $Z^*_{K_1}$ is the element of C_1 nearest to the unknown $Z(x_0)$; see Figure 4. Hence the weights $\lambda_{K_1\alpha}$ of $Z^*_{K_1} = \lambda^\alpha_{K1}\, Z_\alpha$ must be such that:

(a) $\sum_\alpha \lambda_{K_1\alpha} = 1$

(b) the squared distance $\|Z(x_0)-\lambda^\alpha_{K_1} Z_\alpha\|^2 = d^2$ is minimum.

Using the classical formalism of Lagrange, this amounts to minimizing the expression $Q = d^2-2\,\mu_1 \sum_\alpha \lambda_{K_1\alpha}$, with $2\mu_1$ being the Lagrange multiplier. The minimum of Q is obtained by equating the n partial derivatives to zero, i.e.,

$$\partial Q/\partial\lambda_{K_1\alpha} = 0, \qquad \forall\alpha = 1,\ldots,n$$

With

$$d^2 = \sigma_{x_0x_0}-2\,\lambda^\alpha_{K_1}\sigma_{\alpha x_0}+\lambda^\alpha_{K_1}\lambda^\beta_{K_1}\sigma_{\alpha\beta}$$

these derivatives are written:

$$-2\,\sigma_{\alpha x_0}+2\,\lambda^\beta_{K_1}\sigma_{\alpha\beta}-2\,\mu_1 = 0, \qquad \forall\alpha = 1,\ldots,n$$

The $(n+1)$ unknowns (the n weights $\lambda_{K_1\alpha}$ and μ_1) are thus given by the system (5) of $(n+1)$ linear equations.

Linear Kriging Process in the Presence of a Trend (Universal Kriging)

The expectation is neither stationary nor known, but is of known form such as a linear combination of L known functions $\{f_l(x),\ l = 1,\ldots,L\}$:

$$m(x) = \sum_l a_l f_l(x)$$

the L parameters a_l being unknown. This non-stationary expectation $m(x)$

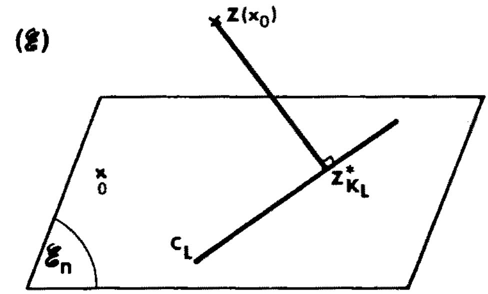

Figure 5. Projection of $Z(x_0)$ onto the linear manifold $C_L \subset \mathscr{E}_n \subset \mathscr{E}$.

is called trend or drift. We see then that the unbiasedness of a linear estimator $\lambda^\alpha Z_\alpha \in \mathscr{E}_n$ restricts the search for this estimator to the linear manifold $C_L \subset \mathscr{E}_n$ defined by the L following conditions on the weights,

$$\sum_\alpha \lambda_\alpha f_l(x_\alpha) = f_l(x_0), \qquad \forall\, l = 1, \ldots, L$$

see Figure 5. The linear manifold C_L is of dimension $(N-L)$, hence is narrower as L increases, i.e., as the form of the unknown drift $m(x)$ is assumed to be more complex.

The kriging estimator $Z^*_{K_L} = \lambda^\alpha_{K_L} Z_\alpha$ is then defined as the unique projection of the unknown $Z(x_0)$ onto the linear manifold $C_L \subset \mathscr{E}_n$. In this case, the algebraic proof of the equations characterizing the projected value $Z^*_{K_L}$ is easier than the geometric one.

The kriging estimator $Z^*_{K_L}$ is the element of C_L nearest to the unknown $Z(x_0)$. Hence the weights $\lambda_{K_L\alpha}$ of $Z^*_{K_L}$ must be such that:

(a) $\sum_\alpha \lambda_{K_L\alpha} f_l(x_\alpha) = f_l(x_0)$, $\forall l = 1, \ldots, L$, i.e., $Z^*_{K_L} \in C_L$

(b) The squared distance $\|Z(x_0) - \lambda^\alpha_{K_L} Z_\alpha\|^2 = d^2$ is minimum.

This amounts to minimizing the expression

$$Q = d^2 - \sum_l 2\mu_l \sum_\alpha \lambda_{K_L\alpha} f_l(x_\alpha)$$

with $\{2\,\mu_l,\ l = 1, \ldots, L\}$ being the L Lagrange multipliers. The minimum of Q is obtained by equating to zero the n partial derivatives with respect to the n unknowns $\{\lambda_{K_L\alpha},\ \alpha = 1, \ldots, n\}$, i.e.,

$$\partial Q / \partial \lambda_{K_L\alpha} = 0, \qquad \forall\, \alpha = 1, \ldots, n$$

With

$$d^2 = \sigma_{x_0x_0} - 2\,\lambda^\alpha_{K_L}\sigma_{\alpha x_0} + \lambda^\alpha_{K_L}\lambda^\beta_{K_L}\sigma_{\alpha\beta}$$

these derivatives are written:

$$-2\sigma_{\alpha x_0}+2\lambda_{K_L}^{\beta}\sigma_{\alpha\beta}-2\sum_l \mu_l f_l(x_\alpha) = 0, \qquad \forall\alpha = 1,\ldots,n$$

The $(n+L)$ unknowns (the n weights $\lambda_{K_L\alpha}$ and the L Lagrange multipliers μ_l) are thus given by the following system of $(n+L)$ linear equations:

$$\begin{cases} \sum_\beta \lambda_{K_L\beta} f_l(x_\beta) = f_l(x_0), & \forall l = 1,\ldots,L \\ \lambda_{K_L}^{\beta}\sigma_{\alpha\beta}-\sum_l \mu_l f_l(x_\alpha) = \sigma_{\alpha x_0}, & \forall\alpha = 1,\ldots,n \end{cases} \tag{7}$$

This system (7) is called a system of linear kriging in the presence of an unknown drift of known form, or a system of linear kriging with an unbiasedness condition of order L [for the unbiasedness is warranted by the L conditions

$$\sum_\beta \lambda_{K_L\beta} f_l(x_\beta) = f_l(x_0), \qquad \forall l = 1,\ldots,L$$

on the n weights $\lambda_{K_L\beta}$]. Some practitioners use the shorter name *universal kriging* system or process. "Universal" is used because the corresponding kriging estimator is unbiased whatever the unknown parameters a_l of the drift

$$m(x) = \sum_l a_l f_l(x)$$

are.

Estimation Variance

The corresponding miminum estimation variance $E\{[Z(x_0)-Z_{K_L}^*]^2\}$, called kriging variance, is written:

$$\sigma_{K_L}^2 = \|Z(x_0)-Z_{K_L}^*\|^2 = \|Z(x_0)\|^2+\|Z_{K_L}^*\|^2-2\langle Z(x_0), Z_{K_L}^*\rangle$$

With: $\|Z(x_0)\|^2 = \sigma_{x_0x_0}$

$$\|Z_{K_L}^*\|^2 = \lambda_{K_L}^{\alpha}\lambda_{K_L}^{\beta}\sigma_{\alpha\beta} = \sum_\alpha \lambda_{K_L\alpha}\sum_l \mu_l f_l(x_\alpha)+\lambda_{K_L}^{\alpha}\sigma_{\alpha x_0} =$$

$$= \sum_l \mu_l f_l(x_0)+\lambda_{K_L}^{\alpha}\sigma_{\alpha x_0}$$

and:

$$\langle Z(x_0), Z_{K_L}^*\rangle = \lambda_{K_L}^{\alpha}\sigma_{\alpha x_0}$$

the kriging variance becomes:

$$\sigma_{K_L}^2 = \sigma_{x_0x_0}-\lambda_{K_L}^{\alpha}\sigma_{\alpha x_0}+\sum_l \mu_l f_l(x_0) \tag{8}$$

As was noted before for the ordinary kriging system (5) and variance (6), the system (7) of equations and the expression (8) can be written replacing the noncentered covariance σ_{xy} by either the centered covariance

$$\sigma'_{xy} = \sigma_{xy} - m(x) \cdot m(y)$$

or the semivariogram

$$\gamma_{xy} = \tfrac{1}{2}E\{[Z(x)-Z(y)]^2\}$$

What Are the Prerequisites for Obtaining $Z^*_{K_L}$?

The expectation $m(x)$ no longer needs to be stationary, but its form is assumed to be known:

$$m(x) = \sum_l a_l f_l(x)$$

the L functions $f_l(x)$ being known, and the L parameters a_l remaining unknown. The knowledge of the covariance function (centered or not) $\sigma_{xy} = \langle Z(x), Z(y) \rangle$ is required. In practice, when one single realization of the nonstationary stochastic process $Z(x)$ is available, problems arise in the inference of the nonstationary covariance $E\{Z(x) \cdot Z(y)\}$; see Matheron (1971, p. 188) and Delfiner (1975).

Notes

(1) When

$$m(x) = a_1 f_1(x) = a_1 = \text{unknown constant } m, \ \forall\, x$$

i.e., when the expectation is unknown but stationary, by making $L = 1$ and $f_1(x) = 1, \forall\, x$, the system (7) reduces to the system (5) of linear kriging with unknown stationary expectation.

(2) As the linear manifold C_L is included in the linear manifold C_1, $C_L \subset C_1$, the kriging estimator $Z^*_{K_1}$ in the stationary case is nearer to the

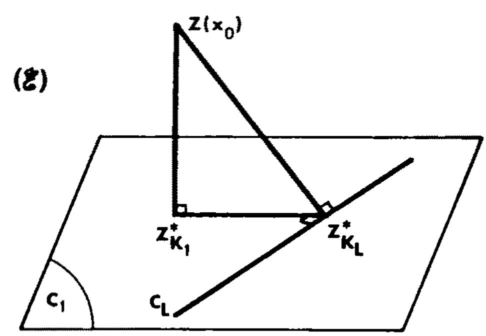

Figure 6. Hierarchy of the kriging estimators: $C_L \subset C_1$ implies that $\sigma^2_{K_L} > \sigma^2_{K_1}$.

unknown $Z(x_0)$ than the kriging estimator $Z^*_{K_L}$ in the presence of a drift of order $L > 1$, i.e., $\sigma^2_{K_1} < \sigma^2_{K_L}$; see. Figure 6. The increase in estimation variance $\sigma^2_{K_L}$-$\sigma^2_{K_1}$ is the price paid for assuming a wider hypothesis on the expectation $m(x)$. This difference $\sigma^2_{K_L}$-$\sigma^2_{K_1}$ increases as L increases, i.e., as the drift $m(x) = \sum_l a_l f_l\,(x)$ is assumed to be of a more complex form.

(3) The kriging estimator $Z^*_{K_L}$ in the presence of a drift can also be seen as the projection of the kriging estimator $Z^*_{K_1}$ in the stationary case onto the linear manifold $C_L \subset C_1$ (theorem of the three perpendiculars). Pythagora's theorem then shows:

$$\|Z(x_0)-Z^*_{K_L}\|^2 = \|Z(x_0)-Z^*_{K_1}\|^2+\|Z^*_{K_1}-Z^*_{K_L}\|^2$$

that is,

$$\sigma^2_{K_L}-\sigma^2_{K_1} = \|Z^*_{K_1}-Z^*_{K_L}\|^2$$

THE NONLINEAR KRIGING PROCESSES

Until now we have only considered linear estimators $Z^* = \lambda^\alpha Z_\alpha$, i.e., we have limited the search for an estimator to the vector space $\mathscr{E}_n$, generated by the linear combinations of the n available data Z_α. But one can remove this limitation and try to look for a nonlinear estimator, for example any nonlinear function $Z^* = f(Z_1, Z_2, \ldots, Z_n)$ of the n available data. This amounts to considering a space H_n of estimators which is much wider than $\mathscr{E}_n$. As $\mathscr{E}_n \subset H_n$, the nonlinear kriging estimator deduced by projection of the unknown $Z(x_0)$ onto H_n will be better [nearer to $Z(x_0)$] than the linear kriging estimator. But on the other hand, as will be shown, the prerequisites for obtaining this nonlinear kriging estimator are more severe and may not be met in practice.

Conditional Expectation

Let $\{Z_\alpha, \alpha = 1, \ldots, n\}$ be the n available data. The most general form for an estimator Z^* is a measurable function $f(Z_1, Z_2, \ldots, Z_n)$ of the n available data. The set H_n of all these measurable functions of n data is a vector space; note that H_n contains in particular the vector subspace $\mathscr{E}_n$ generated by all the linear combinations $\lambda^\alpha Z_\alpha$. As H_n is the widest space where the search for an estimator can be carried out, the projection of the unknown $Z(x_0)$ onto H_n is the *best possible* estimator of $Z(x_0)$ that can be deduced from the n data Z_α; see Figure 7. Now this best possible estimator is, by definition, the conditional expectation of $Z(x_0)$, i.e.,

$$E_n Z_0 = E\{Z(x_0)/Z(x_1), Z(x_2), \ldots, Z(x_n)\}$$

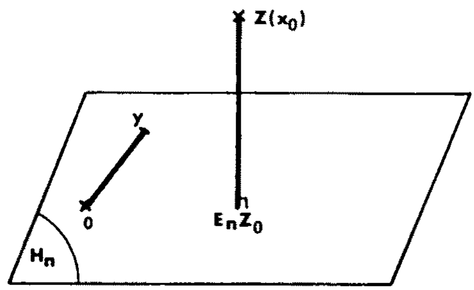

Figure 7. The conditional expectation $E_n\ Z_0$ defined as the projection of $Z(x_0)$ onto H_n.

see Neveu (1964, p. 116). Hence when the set where the unknown $Z(x_0)$ is to be projected is the vector space H_n, the corresponding kriging estimator is identical to the conditional expectation E_nZ_0.

Note first that the conditional expectation E_nZ_0, considered as an estimator of $Z(x_0)$, is unbiased, for obviously $E\{E_n\ Z_0\} = E\{Z(x_0)\}$. $Y = f(Z_1, \ldots, Z_n)$ being any arbitrary element of H_n, the projection $E_n\ Z_0$ of $Z(x_0)$ onto H_n is characterized by one of the following two relationships:

$$\|Z(x_0)-E_nZ_0\| = \min\|Z(x_0)-Y\|, \qquad \forall\, Y\in H_n$$

$$\langle Z(x_0)-E_nZ_0, Y\rangle = 0, \qquad \forall\, Y\in H_n$$

What Are the Prerequisites for Obtaining $E_n\ Z_0$?

In order to build the projection $E_n\ Z_0$, i.e., to find the n-variable function

$$E_n\ Z_0 = f_K(Z_1, \ldots, Zn) \in H_n$$

it is necessary to be able to express scalar products of the type:

$$\langle Z(x_0)-f_K, f\rangle = E\{[Z(x_0)-f_K(Z_1, \ldots, Z_n)]\cdot f(Z_1, \ldots, Z_n)\}, \quad \forall f\in H_n$$

To do this, we must know the distribution of the $(n+1)$ variables

$$\{Z(x_0), Z(x_1), Z(x_2), \ldots, Z(x_n)\}$$

In the general case, when the information is limited to a single realization of the stochastic process $Z(x)$ at each of the n data points $\{x_\alpha, \alpha = 1, \ldots, n\}$, such an inference is not possible and the conditional expectation is inaccessible.

In practice, the conditional expectation $E_n\ Z_0$ can be obtained only in very particular cases, for instance if the stochastic process $Z(x)$ is Gaussian and stationary. It is a classical probabilistic result [see Neveu (1964, p. 123)], that the conditional expectation of a Gaussian stationary stochastic process is identical to the best linear estimator, i.e., the unknown $Z(x_0)$ has the same projection onto H_n and $\mathscr{E}_{n+1}$ and this projection

$$E_n Z_0 \equiv Z^*_{K_0} = \lambda_0 + \lambda^\alpha_{K_0} Z_\alpha$$

can be determined explicitly from the system (2) or (3) of simple kriging.

Linear Kriging Process Applied to the Transformed Variables

Because the stationary Gaussian case is so favorable [the best possible estimator of any unknown $Z(x_0)$ can then be obtained], it would be convenient to transform the *stationary* random variable $Z(x)$ drawn from any distribution into a stationary, centered Gaussian random variable $G(x)$. Let ϕ be the transformation

$$G(x) = \phi\{Z(x)\}$$

this transformation can be obtained by either graphically using the two distribution functions of $Z(x)$ and $G(x)$, or a development in orthogonal polynomials [Hermite polynomials, for example; see Matheron (1975b, p. 229)].

The variable $G(x)$ so obtained, has a *single variable* Gaussian distribution, thus it remains to assert a stronger, but in practice well-verified, hypothesis that all the *multivariate* distributions of the stochastic process $G(x)$ are also Gaussian. Under this hypothesis, the conditional expectation

$$E_n G_0 = E\{G(x_0)/G(x_1), G(x_2), \ldots G(x_n)\}$$

is identical to the simple linear kriging estimator,

$$G^*_{K_0} = \lambda^\alpha_{K_0} G_\alpha$$

provided by the system of equations (3). Now when the n transformed data G_α are fixed, the conditional law of $G(x_0)$ is also Gaussian, with an expected value $E_n G_0 = G^*_{K_0}$ and variance the kriging variance $\sigma^*_{K_0}$ provided by relation (4). Knowing this conditional law, it is then easy to retrieve the sought-after conditional expectation of the initial variable $Z(x_0) = \phi^{-1}\{G(x_0)\}$, i.e.,

$$E_n Z_0 = E\{\phi^{-1}\{G(x_0)/G(x_1), \ldots, G(x_n)\}\}$$

Note that, because the transformation function ϕ is generally not linear, the conditional expectation $E_n Z_0$ is not linear with respect to the n initial data $(Z_1, Z_2, \ldots, Z_n)$.

Example. Suppose that the initial stochastic process has a lognormal multivariate distribution. The stochastic process

$$G(x) = \log Z(x)$$

is then Gaussian, and the inverse transform is simply

$$Z(x) = \exp G(x)$$

A linear kriging process applied to the logarithms of the initial data is also called by practitioners a lognormal kriging process; see Lallement and Maréchal (1977). It is interesting to note that Krige's initial estimation procedure was in fact a lognormal kriging process, as Krige worked on the logarithms of the Witwatersrand's gold grades.

What Are the Prerequisites for Obtaining $E_n Z_0$ Through the Linear Kriging Process Applied to the Transformed Variable $G(x) = \phi\{Z(x)\}$?

This transformation function ϕ must be known, which requires the stationarity of the stochastic process $Z(x)$ and the knowledge of its single-variable distribution function [and thus in particular the knowledge of the stationary expectation $E\{Z(x)\} = m$]. Then the linear kriging estimator G^*_{K0} must be built, which requires the knowledge of the stationary covariance function

$$\langle G(x), G(y)\rangle = E\{G(x)\, G(y)\}$$

of the transformed stochastic process $G(x)$; the inference of this covariance is generally done from the available transformed data $G(x_\alpha)$. Moreover, the linear kriging estimator G^*_{K0} must be assumed to be identical to the conditional expectation $E_n\, G_0$, which requires the assumption that all the multivariate distributions of the transformed stochastic process $G(x) = \phi\{Z(x)\}$ are Gaussian.

In practice, these prerequisites are seldom met all together, and the *nonlinear* estimator

$$Z^* = E\{\phi^{-1}\{G(x_0)/G(x_1), \ldots, G(x_n)\}\}$$

provided by this nonlinear estimation procedure differs from the true conditional expectation

$$E_n Z_0 = E\{Z(x_0)/Z(x_1), \ldots, Z(x_n)\}$$

However, and under the condition that the one-variable distribution of $Z(x)$ is well known, practice has shown that this nonlinear estimator Z^* is generally better than the linear kriged estimators Z^*_{K0} or Z^*_{K1} provided by the direct linear kriging processes applied to the initial data $Z(x_\alpha)$, i.e., by projection of the unknown $Z(x_0)$ onto either $\mathscr{E}_{n+1}$ or the linear manifold $C_1 \subset \mathscr{E}_{n+1}$.

Disjunctive Kriging Process

We have seen that the best possible estimator of an unknown $Z(x_0)$ using n data $\{Z_\alpha, \alpha = 1, \ldots, n\}$ is the conditional expectation $E_n Z_0$, which is the projection of $Z(x_0)$ onto the space H_n of the *n-variable* measurable functions $f(Z_1, \ldots, Z_n)$. However, in the general non-Gaussian case, this conditional expectation $E_n Z_0$ remains inaccessible. One solution is then to restrict the space onto which the unknown $Z(x_0)$ is to be projected, to the vector subspace $\mathscr{E}_n \subset H_n$, $\mathscr{E}_n$ being generated by all the linear combinations $\lambda^\alpha Z_\alpha$ of the n data. The idea of the disjunctive kriging process is to consider an intermediate space $D_n : \mathscr{E}_n \subset D_n \subset H_n$. This space D_n, where $Z(x_0)$ is to be projected, should be more inclusive than $\mathscr{E}_n$, i.e., $\mathscr{E}_n \subset D_n$, so as to provide estimators which are more powerful than simple linear estimators; on the other hand, D_n must be less inclusive than H_n, i.e., $D_n \subset H_n$, so that the prerequisites for obtaining the projection of $Z(x_0)$ onto Dn can be met in practice. It is to be noted that the more inclusive the space where the projection is to be made, the more severe are the prerequisites.

A good choice for this intermediate space D_n is the vector space generated by the sums of n *single-variable* measurable functions, i.e.,

$$D_n : \{g_1\,(Z_1)+g_2(Z_2) + \ldots + g_n(Zn)\}$$

see Matheron (1975,b p. 221). Note that this space D_n does satisfy the inclusions: $\mathscr{E}_n \subset D_n \subset H_n$. The estimator provided by the disjunctive kriging process is then the projection Z^*_{DK} of the unknown $Z(x_0)$ onto D_n; see Figure 8.

Equations of the Disjunctive Kriging Process

The projection Z^*_{DK} of $Z(x_0)$ onto H_n is characterized by the usual two conditions:

(a) $Z^*_{DK} \in D_n$, i.e., $Z^*_{DK} = \sum_{\alpha=1}^{n} f_\alpha(Z_\alpha)$

(b) the vector $Z(x_0)-Z_{DK}$ is orthogonal to any vector Y of D_n, i.e., $\langle Z(x_0)-Z^*_{DK}, Y\rangle = 0, \forall\ Y \in D_n$.

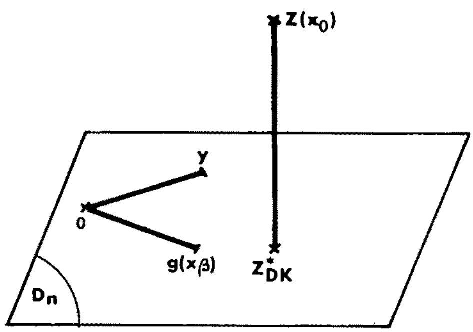

Figure 8. The disjunctive kriging estimator Z^*_{DK} defined as the projection of $Z(x_0)$ onto D_n.

Now, D_n is generated by the measurable functions $g(Z_\beta)$ depending on only *one* of the data Z_β, so that condition (b) can be written:

$$\langle Z(x_0), g(Z_\beta)\rangle = \langle Z^*_{DK}, g(Z_\beta)\rangle, \qquad \text{for any } \beta = 1, \ldots, n$$

and any single-variable measurable function g. This condition is satisfied if and only if (the proof although classical is recalled later on) $Z(x_0)$ and Z^*_{DK} admit the same conditional expectation upon each of the Z_β, $\beta = 1, \ldots, n$, taken *separately*, i.e.,

$$\begin{aligned} E\{Z(x_0)/Z_1\} &= E\{Z^*_{DK}/Z_1\} \\ E\{Z(x_0)/Z_2\} &= E\{Z^*_{DK}/Z_2\} \\ &\vdots \\ E\{Z(x_0)/Z_n\} &= E\{Z^*_{DK}/Z_n\} \end{aligned} \tag{9}$$

Considering the expression

$$Z^*_{DK} = \sum_\alpha f_\alpha(Z_\alpha)$$

given by the first condition (a), the disjunctive kriged estimator Z^*_{DK} is finally characterized by the following system of n equations:

$$\sum_{\alpha=1}^{n} E\{f_\alpha(Z_\alpha)/Z_\beta\} = E\{Z(x_0)/Z_\beta\}, \qquad \forall \beta = 1, \ldots, n \tag{10}$$

In the general case, the system of equations (10) provides the solution functions f_α, $\alpha = 1, \ldots, n$, in the form of integral equations. These functions are then approximated by limited expansions in orthogonal polynomials (e.g., Hermite polynomials). In the scope of the present paper, the statement of the approximate resolution procedure of system (10) will take us too far afield; readers are referred to Matheron (1975b, p. 223).

Proof. The condition (b) remains to be proved:

Condition $\langle Z(x_0), g(Z_\beta)\rangle = \langle Z^*_{DK}, g(Z_\beta)\rangle$, $\forall\ \beta = 1, \ldots, n$, and $\forall$ the measurable function g, is satisfied if and only if the system of equations (9) is met.

Let d_β be the vector space generated by all the single-variable measurable functions $g(Z_\beta)$ of the particular data Z_β. The conditional expectation $E_\beta Z_0 = E\{Z(x_0)/Z_\beta\}$ of $Z(x_0)$ with respect to the data Z_β is, by definition, the projection of $Z(x_0)$ onto d_β, and thus the vector $Z(x_0) - E_\beta Z_0$ is orthogonal to any vector $g(Z_\beta)$ belonging to the space d_β, i.e.,

$$\langle Z(x_0) - E_\beta Z_0, g(Z_\beta)\rangle = 0, \qquad \forall\ g(Z_\beta) \in d_\beta$$

This orthogonality relation can be written:

$$\langle Z(x_0), g(Z_\beta)\rangle = \langle E_\beta Z_0, g(Z_\beta)\rangle, \qquad \forall\ g(Z_\beta) \in d_\beta$$

Similarly, considering the projection of Z^*_{DK} onto d_β:

$$\langle Z^*_{DK}, g(Z_\beta)\rangle = \langle E_\beta Z^*_{DK}, g(Z_\beta)\rangle, \qquad \forall\, g(Z_\beta) \in d_\beta \subset D_n$$

and the preceding condition (b) can then be written:

$$\langle E_\beta Z_0, g(Z_\beta)\rangle = \langle E_\beta Z^*_{DK}, g(Z_\beta)\rangle, \qquad \forall\, \beta = 1, \ldots, n, \text{ and } \forall\, g$$

This is satisfied if and only if

$$E_\beta Z_0 = E_\beta Z^*_{DK}, \qquad \forall\, \beta = 1, \ldots, n$$

which is nothing but the system of equations (9).

Estimation Variance

The resolution of system of equations (10) provides both the disjunctive kriging estimator Z^*_{DK} and the corresponding estimation variance

$$\sigma^2_{DK} = E\{[Z(x_0) - Z^*_{DK}]^2\} = \|Z(x_0) - Z^*_{DK}\|^2$$

Because the space H_n includes the space D_n, the projection of $Z(x_0)$ onto H_n (i.e., the conditional expectation $E_n\, Z_0$) is nearer to $Z(x_0)$ than the projection of $Z(x_0)$ onto D_n (i.e., the disjunctive kriged estimator Z^*_{DK}); see Figure 9. The theorem of the three perpendiculars shows that Z^*_{DK} is also the projection of the conditional expectation $E_n\, Z_0$ onto D_n, and Pythagora's theorem shows the following relation between the various estimation variances:

$$\sigma^2_{DK} = \|Z(x_0) - Z^*_{DK}\|^2 = \|Z(x_0) - E_n\, Z_0\|^2 + \|E_n\, Z_0 - Z^*_{DK}\|^2$$

What Are the Prerequisites for Obtaining Z^*_{DK} ?

In order to set the system of equations (10) characterizing the disjunctive kriged estimator Z^*_{DK}, it is necessary to know the conditional expectations

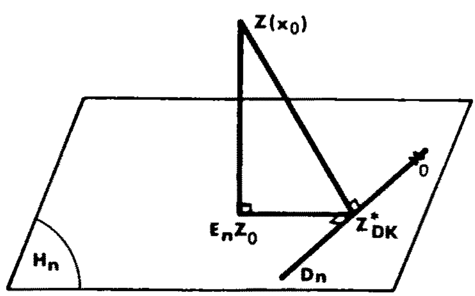

Figure 9. Hierarchy of the two estimators $E_n Z_0$ and Z^*_{DK} : $D_n \subset H_n$ implies that: $\sigma^2_{DK} > \|Z(x_0) - E_n\, Z_0\|^2$.

$E\{f_\alpha(Z_\alpha)/Z_\beta\}$ and $E\{Z(x_0)/Z_\beta\}$, $\forall\ \beta = 1, \ldots, n$; this is possible when all the *bivariate* distributions for the pairs (Z_α, Z_β) and (Z_β, Z_0) are known.

Thus, Z^*_{DK} is not as good an estimator as the conditional expectation $E_n\ Z_0$. On the other hand, the disjunctive kriging process requires less prerequisites than does the process for obtaining the conditional expectation; only the set of bivariate distributions is required instead of the set of $(n+1)$-variate distributions.

CONCLUSIONS

When estimating an unknown value $Z(x_0)$ from n available data $\{Z(x_\alpha), \alpha = 1, \ldots, n\}$, the search for an estimator can be carried out in various spaces of possible estimators. Once this space $\mathscr{E}'$ is defined, the best estimator, i.e., the element of $\mathscr{E}'$ which is the nearest to the unknown $Z(x_0)$, is the projection of $Z(x_0)$ onto $\mathscr{E}'$. The general term "kriging" is given to the various processes for obtaining the projections of the unknown $Z(x_0)$ onto various spaces $\mathscr{E}'$ of possible estimators. The space $\mathscr{E}'$ where the projection is to be done is chosen according to the possible inference about the stochastic process $Z(x)$. The more that can be known about $Z(x)$, i.e., the more severe prerequisites that can be inferred, the wider the space $\mathscr{E}'$ of possible estimators will be, and consequently the nearer to the unknown value $Z(x_0)$ the corresponding kriging estimator will be. Thus the following hierarchy of the various kriging estimators can be stated (see Table 1).

(a) Starting from the widest space of possible estimators, i.e., the vector space H_n of all measurable functions $f(Z_1, \ldots, Z_n)$ of the n available data, one defines the best possible estimator, i.e., the conditional expectation $E_n\ Z_0$. Conversely, the prerequisites are extremely severe: all the $(n+1)$-variate distributions of the $(n+1)$ variables $\{Z(x_0), Z_\alpha, \alpha = 1, \ldots, n\}$ must be known. In practice, these prerequisites are only met in the case of a Gaussian stationary stochastic process $Z(x)$.

(b) A first restriction of the space of possible estimators is the vector space, $D_n \subset H_n$, generated by all the sums of n one-variable measurable functions, i.e.,

$$D_n = \left\{ \sum_{\alpha=1}^{n} f_\alpha(Z_\alpha) \right\}$$

The corresponding kriging process provides the so-called disjunctive kriging estimator Z^*_{DK}, which is not as good an estimator as the conditional expectation. Conversely, the prerequisites are already less severe, and the bivariate distributions must be known. In practice, these bivariate distributions can only be inferred if the process $Z(x)$ is stationary.

Table 1. Hierarchy of the Various Kriging Estimators

Space of projection	vector space H_n $\{f(Z_1, \ldots, Z_n)\}$	$\supset$	vector space D_n $\{\sum_{\alpha=1}^{n} f_\alpha(Z_\alpha)\}$	$\subset$	vector space $\mathscr{E}_{n+1}$ $\{\lambda_0 + \lambda^\alpha Z_\alpha\}$	$\subset$	linear manifold C_L $\lambda^\alpha Z_\alpha$ $+L$ conditions on λ_α
Kriging estimator	conditional expectation $E_n Z_0$ (best possible estimator)		disjunctive kriging estimator Z_{DK}^*		simple kriging estimator[a] Z_{K0}^*		universal kriging estimator[b] Z_{KL}^*
Estimation variance	$\Vert Z(x_0) - E_n Z_0 \Vert^2$	$\leq$	σ_{DK}^2	$\leq$	σ_{K0}^2	$\leq$	$\sigma_{K1}^2 \leq \sigma_{K2}^2 \leq - - \leq \sigma_{KL}^2$
Prerequisites	$(n+1)$-variate distributions		bivariate distributions		the $(n+1)$ expectations		the form of the trend $m(x) = \sum_{l=1}^{L} a_l f_l(x)$
	e.g., $Z(x)$ stationary and Gaussian		in practice, stationarity of $Z(x)$		the covariance function		the covariance function
					in practice, stationarity of $Z(x)$		stationarity of $Z(x)$ not required if $L > 1$

[a] Simple kriging estimator, i.e., the linear estimator provided by the kriging process with known expectations.
[b] Universal kriging estimator, i.e., the linear estimator provided by the kriging process in the presence of an unknown trend but of known form.

(c) A second restriction of the space of possible estimators is the vector space $\mathscr{E}_{n+1} \subset D_n \subset H_n$, generated by all the linear combinations $\lambda^\alpha Z_\alpha + \lambda_0 \cdot 1$ of the n data Z_α plus the constant 1. The corresponding kriging process provides the so-called kriging estimator with known expectations, $Z^*_{K_0}$. This estimator is not as good as the preceding disjunctive kriging estimator Z^*_{DK}, i.e.,

$$\sigma^2_{K_0} \geq \sigma^2_{DK}$$

Conversely the prerequisites are less severe; only the $(n+1)$ expectations $E\{Z(x_0\}$, $E\{Z(x_\alpha)\}$, $\alpha = 1, \ldots, n$, and the covariance function $E\{Z(x)\, Z(y)\}$ must be known. In practice, the knowledge of the $(n+1)$ preceding expectations requires the stationarity of the stochastic process $Z(x)$.

(d) Further restrictions of the space of possible estimators are the various linear manifolds, $C_L \subset \mathscr{E}_{n+1} \subset D_n \subset H_n$, defined by an increasing number L of conditions on the n weights λ_α of the linear estimator $Z^* = \lambda^\alpha Z_\alpha$. For a given linear manifold C_L, the corresponding kriging process provides the so-called universal kriging estimator, or more precisely the linear kriging estimator $Z^*_{K_L}$ with an unbiasedness condition of order L. The less inclusive the linear manifold C_L is, i.e., the greater the number L of conditions on the weights λ_α is, the less precise the corresponding kriging estimator will be. This can be written in terms of estimation variances:

$$\sigma^2_{K_L} \geq \sigma^2_{K_{L-1}} \geq \ldots \geq \sigma^2_{K_1} \geq \sigma^2_{K_0}$$

Conversely, the greater L is, the less severe are the prerequisites for obtaining $Z^*_{K_L}$. In all cases, the knowledge of the covariance function $E\{Z(x)\, Z(y)\}$ is required, but:

—for $L = 1$, the expectation no longer needs to be known but is assumed to be stationary, i.e., $E\{Z(x)\} = m, \forall\, x$.

—for $L \rangle 1$, only the knowledge of the form of the nonstationary expectation is required, i.e.,

$$E\{Z(x)\} = m(x) = \sum_{l=1}^{L} a_l f_l(x)$$

the L functions $f_l(x)$ being known but the L parameters a_l remaining unknown, thus leaving $m(x)$ unknown. As L increases, a more and more complex form for the trend $m(x)$ is accepted.

ACKNOWLEDGMENTS

E. H. Timothy Whitten kindly reviewed the numerous drafts, and the help of G. Matheron was, as always, most positive.

REFERENCES

Delfiner, P., 1975, Linear estimation of non-stationary phenomena, *in* Advanced geostatistics in the mining industry: Reidel Publ. Co., Dordrecht, Holland, p. 49–68.

Krige, D. G., 1951, A statistical approach to some mine valuations and allied problems on the Witwatersrand: Unpubl. Master's Thesis, Univ. of Witwatersrand.

Lallement, B., and Marechal, A., 1977, Lognormal kriging—A practical example: to be published.

Matheron, G., 1963, Traité de Géostatistique Appliquée, vols. 1 and 2: Editions Technip, Paris.

Matheron, G., 1969, Le krigeage universel: Ecole des Mines de Paris, Paris, 83 p.

Matheron, G., 1971, The theory of regionalized variables and its applications: Ecole des Mines de Paris, Paris, 211 p.

Matheron, G., 1973, Le krigeage disjonctif: Internal Rept. no. 360, Centre de Géostatistique, Fontainebleau, France.

Matheron, G., 1975a, Le choix des modèles en géostatistique, *in* Advanced geostatistics in the mining industry: Reidel Publ. Co., Dordrecht, Holland, p. 11–27.

Matheron, G., 1975b, A simple substitute for conditional expectation: the disjunctive kriging: Reidel Publ. Co., Dordrecht, Holland, p. 221–236.

Neveu, J., 1964, Bases mathématiques du calcul des probabilités: Editions Masson et Cie, Paris, 203 p.

Wiener, N., 1949, Extrapolation, interpolation and smoothing of stationary time series: M.I.T. Press, Cambridge, Mass., 10 p.

Journel, A., 1975, Geological reconnasissance to exploitation—A decade of applied geostatistics: CIMM Bulletin, June 1975, p. 75–84.

Journel, A., and Huijbregts, Ch., 1977, Mining Geostatistics, Academic Press, London, in press.