Introduction to disjunctive kriging and nonlinear geostatistics

by

Jacques RIVOIRARD

C–143

Décembre 1990

CENTRE DE GÉOSTATISTIQUE

35, RUE SAINT–HONORÉ, 77305 FONTAINEBLEAU (France)

ECOLE DES MINES
DE PARIS

CENTRE DE GÉOSTATISTIQUE,
Ecole Nationale Supérieure des Mines de Paris,
35 rue Saint–Honoré, 77305 FONTAINEBLEAU, France

Jacques RIVOIRARD. *Introduction to disjunctive kriging and nonlinear geostatistics.* Décembre 1990. 89 p.
C–143

TABLE OF CONTENTS

Introduction

The discipline which is now known as geostatistics began to develope over thirty years ago for mining evaluation and since has extended to other fields of activity. Around 1960 in particular, G. Matheron built linear geostatistics. Some of its tools (variogram, kriging) are widely used nowadays. Linear geostatistics makes it possible for instance to evaluate the metal content of a mining block or panel by estimating the mean of the grades of the points in it from samples. The reader of this book is supposed to be familiar with linear geostatistics.

In the sixties non linear methods based on lognormal distributions had already been developed for the needs of South–African gold fields. To estimate a panel from samples, D. G. Krige had proposed a lognormal regression, which later led to lognormal kriging and this is still in use today. Before that H. S. Sichel had been involved in evaluating deep deposits from only a few samples. He had proposed an estimator and given confidence intervals for the mean, assuming the sample values were independent and distributed according to the same lognormal law.

But it was mainly from the seventies on that the techniques now called non linear geostatistics have been developed. Among the more and more difficult mining evaluation problems that geostatistics had to face was that of predicting the results of a selective mining process. To simplify, imagine a deposit made up of many small selection blocks before its exploitation starts. What is then important is not to have an estimated value (by kriging or other method) for the grade of each block, but to know if the grade of such small blocks, as they will be known at the time of their individual mining, will exceed the cut–off between waste and ore, and to predict the quantity of metal which will be recovered. Such a typically non linear problem arises in other fields. For instance in pollution, each time we are not really interested in obtaining an estimate, often smoothed and of a limited precision, of real values, but rather we want to find out the probability that such real values pass a threshold.

The formalism and the tools of non linear geostatistics have been developed to solve problems like these. These nonlinear methods are more powerful than their linear counterparts but are more complex mathematically. So potentially interested users are often put off by the mathematics. Nevertheless it is necessary to understand some of the mathematics to use them correctly. In particular the user must understand why and how these mathematical ideas come into play.

This text was written at the request of several users of linear geostatistics who were interested in applying nonlinear methods. So instead of being mainly mathematical it builds on their knowledge of and experience in linear geostatistics. The main developments are given in detail, but to keep the text readable the less important ones have been reduced to a minimum or omitted (e.g. estimation variances, the problems of convergence, the DK probability density function, the 1^{st} order variogram, etc...).

I hope that the reader will grasp the essentials of the methods presented here, and that this will give him a solid grounding to which he can add more detail by further reading (for example, on the different isofactorial models used in disjunctive kriging and changing of support).

Acknowledgement:

I would like to thank Margaret Armstrong for her help with this work and her translation of the original French text, and Richard Webster for his numerous and valuable comments.

Usual notations

Y , Z	variables
x	a point in space
$Z(x)$	variable (grade for instance) at point x
$Y(x)$	another variable at point x (e.g. gaussian transform of the grade)
x_α , x_β	sample points
v	a block, or its support
V	a larger support
$Z(v)$	grade on support v
Y_v	gaussian transform of $Z(v)$
$E(Z)$	expectation (mean value) of Z
$Var\ (Z)$	variance of Z
$Cov\ (Y,Z)$	covariance between Y and Z
h	a distance, or a distance vector
$C(h)$	covariance at distance h
$\gamma(h)$	variogram
$Sph\ (20)$	spherical variogram with range 20 and sill 1
$F(z)$	c.d.f. of $Z(x)$ $= P(Z(x) < z)$ probability of having $Z(x) < z$
$F_v(z)$	c.d.f. of $Z(v)$ $= P(Z(v) < z)$ probability of having $Z(v) < z$
$1_{Y(x) \geq j}$	indicator of $Y(x) \geq j$

T_j	$P(Y(x) \geq j)$
T_{ij}^{h}	$P(Y(x) \geq i \quad Y(x+h) \geq j)$
$\gamma_j(h)$	variogram of indicator $1_{Y(x) \geq j}$
$\gamma_{ij}(h)$	cross–variogram between indicators $1_{Y(x) \geq i}$ and $1_{Y(x) \geq j}$
z_c	cut–off for $Z(x)$ or $Z(v)$
y_c	cut–off for gaussian variable $Y(x)$ or Y_v
$g(t)$	p.d.f. of a standard gaussian
$G(t)$	c.d.f. of a standard gaussian
$H_n(y)$	Hermite polynomial of degree n
Y_α	$Y(x_\alpha)$
$H_n(Y_\alpha)$	$H_n[Y(x_\alpha)]$
*	as superscript implies an estimator
K	as superscript implies kriging estimator
SK	simple kriging
OK	ordinary kriging
CK	cokriging
DK	disjunctive kriging
CE	conditional expectation
λ_α	weight of sample point x_α
λ_α^n	weight of Hermite polynomial $H_n(Y_\alpha)$ at sample point x_α
ϱ	a correlation coefficient
$\varrho(h)$	a correlogram (normed covariance) at distance h
$\varrho_{\alpha\beta}$	correlation between $Y(x_\alpha)$ and $Y(x_\beta)$

$\Phi(y)$	gaussian anamorphosis function of $Z(x)$
ϕ_n	coefficient of order n when expanding Φ into Hermite polynomials
σ^2	a variance
σ_K^2	kriging variance
$\Phi_{(v)}$	gaussian anamorphosis function of $Z(v)$
r	change of support coefficient
T	proportion of ore
Q	quantity of metal
B	conventional profit

PART ONE : POINT ESTIMATION

1. Reminder on Cokriging

Cokriging will be used in the following chapters to define the disjunctive kriging estimator. In the present chapter we will give some reminders, in particular of the cases where cokriging simplifies.

1.1 Why use Cokriging ?

Should we use cokriging to estimate several variables or would it suffice to krige them separately ? There are two main reasons for using cokriging.

1) When kriging an undersampled variable, cokriging allows us to take advantage of the information available on another variable which is related to the first and for which there are more data.

2) When the values of all variables are available at all sample points, it improves the coherence between the estimated values by taking account of the relationship between them.

We now consider the second case. As a simple example we shall estimate the bottom $Z_1(x)$ and the top $Z_2(x)$ of a seam of rock. Kriging gives the best linear estimate of each of the variables separately. But most of the time we are also interested in the thickness of the seam $T(x)$:

$$T(x) = Z_2(x) - Z_1(x)$$

Knowing any two of the three variables $Z_1(x)$, $Z_2(x)$ and $T(x)$ we can obtain the third one by simple addition or subtraction. Unfortunately the same simple relations do not apply to the kriged estimates. In general the difference:

$$Z_2(x)^K - Z_1(x)^K$$

does not equal $T(x)^K$.

In contrast, cokriging gives the best linear estimate of the variables and of linear combinations of them. So the cokriged estimates satisfy:

$$Z_2(x)^{CK} - Z_1(x)^{CK} = T(x)^{CK}$$

Comments :

Kriging a positive variable does not guarantee that the resulting estimate will be positive. Similarly cokriging the above variables does not guarantee that the kriged estimate of thickness $T(x)^{CK}$ will be positive, either. To ensure this, more elaborate techniques are required.

1.2 Simplifying Cokriging

In order to use cokriging, we need to know the cross variograms (or covariances) of the variables as well as the auto variograms. A mathematically consistent coregionalisation model must be fitted to all these variograms, whose number increases with the square of the number of variables. The size of the cokriging system also increases with the number of variables. So it will be profitable if cokriging can be simplified.

In fact, two types of model allow us to do this. So it is useful in practice to try to get back to these. In both cases the cokriging estimates are simply obtained by solving some kriging systems.

In what follows we shall assume stationarity unless explicitly stated otherwise.

1.3 Intrinsically Correlated Variables

Let $Z_1(x)$, $Z_2(x)$, ... $Z_n(x)$ be variables whose auto and cross covariances (or only variograms) are all the same apart from a multiplicative factor. That is, if the cross covariance between $Z_i(x)$ and $Z_j(x)$ is $C_{ij}(h)$, then :

$$C_{ij}(h) = Cov\ [\ Z_i(x+h),\ Z_j(x)] = C_{ij}(0)\ \varrho(h)$$

where $\varrho(h)$ is a spatial correlation function and $C_{ij}(0)$ is a constant.

In general the correlation between two variables $Z_i(v)$ and $Z_j(v)$ depends on their support v (which may be just a point) and on the field V described by v. It is not an inherent characteristic. On the contrary, when all the variograms are multiples of each other the *correlation* is said to be *intrinsic*: it does not depend on v and V (Matheron 1965, p.149; Journel 1977, p.237 and Journel and Huijbregts 1978, p.174, call this model "the intrinsic coregionalisation").

Using this model (and assuming that all the variables have been measured at all sample points), cokriging any variable (or a linear combination of the variables) is just the same as kriging that variable. Only one kriging system has to be solved, the weights for any sample are the same for all the variables.

1.4 Factorizing the $Z_1, Z_2, \ldots Z_n$

Suppose there is no spatial correlation between $Z_1(x)$, $Z_2(x)$ ... $Z_n(x)$, i.e. their correlation coefficients are zero and their cross-covariances are identically null. Then kriging each of these variables is equivalent to cokriging it.

Suppose now a correlation exists between these variables. It is sometimes possible to go back to other variables which have no spatial correlation and whose kriging will give the cokriging of all variables. This is the case when we can combine linearly the $Z_1(x)$,

$Z_2(x)$, ..., $Z_n(x)$ to built a set of spatially uncorrelated variables $Y_i(x)$, which can in their turn be combined linearly to give the $Z_j(x)$. That is,

$$Z_j(x) = \sum_i a_{ji} Y_i(x)$$

These variables Y_i, which can be obtained at each sampling point from the Z_j, are spatially uncorrelated; and so kriging them is equivalent to cokriging them:

$$Y_j{}^{CK} = Y_i{}^{K}$$

Then we obtain the cokriged estimates of the Z_j:

$$Z_j{}^{CK} = \sum_i a_{ji} Y_i{}^{K}$$

We call these variables Y_i the *factors*, and so factorizing the Z_j is just their decomposition into these factors.

The estimation errors $Y_i - Y_i{}^{K}$ are uncorrelated, and so their estimation variances give that of Z_j :

$$Z_j - Z_j{}^{CK} = \sum_i a_{ji} (Y_i - Y_i{}^{K})$$

$$Var\ [Z_j - Z_j{}^{CK}] = \sum_i (a_{ji})^2\ \ Var\ (Y_i - Y_i{}^{K})$$

1.5 Examples

Let us now return to our initial example to illustrate these models.

a) Assume that the top $Z_2(x)$ and the bottom $Z_1(x)$ of a seam have the following coregionalisation model:

$$\begin{cases} C_1(h) = 16\ Sph\ (100) + 16\ Sph\ (20) \\ C_2(h) = 25\ Sph\ (100) + 25\ Sph\ (20) \\ C_{12}(h) = 16\ Sph\ (100) + 16\ Sph\ (20) \end{cases}$$

where $C_1(h)$, $C_2(h)$ and $C_{12}(h)$ denoted the auto and cross covariances for the variables, and where *16 Sph (100)* denotes a spherical model with a range of *100* and a sill of *16*.

These three covariances functions are all proportional to *Sph (100) + Sph (20)*. And so Z_1 and Z_2 are intrinsically correlated.

b) Now consider this coregionalisation model:

$$\begin{cases} C_1(h) = 16\ Sph\ (100) \\ C_2(h) = 25\ Sph\ (100) + 25\ Sph\ (20) \\ C_{12}(h) = 16\ Sph\ (100) \end{cases}$$

Here the structure of Z_1, $C_1(h)$, is the same as the cross-structure, $C_{12}(h)$. This enables us to rewrite the structure of Z_2 as:

$$C_2(h) = 16\ Sph\ (100) + 9\ Sph\ (100) + 25\ Sph\ (20)$$

This means that we can represent Z_2 as the sum of two components, $Z_1(x)$ and $T(x)$:

$$Z_2(x) = Z_1(x) + T(x)$$

where $T(x) = Z_2(x) - Z_1(x)$ has the covariance:

$$9\ Sph\ (100) + 25\ Sph\ (20)$$

and is spatially uncorrelated with $Z_1(x)$.

The model has effectively been factorized into $Z_1(x)$ and $T(x)$, and so we can krige these components separately to obtain the cokriged estimates:

$$Z_1(x)^{CK} = Z_1(x)^{K}$$
$$T(x)^{CK} = T(x)^{K}$$
$$Z_2(x)^{CK} = Z_1(x)^{CK} + T(x)^{CK} = Z_1(x)^{K} + T(x)^{K}$$

Comments on this last example

1. As we are dealing with stationary variables we can calculate their means and then write them as:

$$Z_1(x) = m_1 + (Z_1(x) - m_1)$$
$$Z_2(x) = (m_1 + m_T) + (Z_1(x) - m_1) + (T(x) - m_T)$$

with $m_1 + m_T = m_2$.

We have:

$$Cov\ [Z_1(x),\ T(x+h)] = E\ [(Z_1(x) - m_1)\ (T(x+h) - m_T)] = 0$$

Since the expected value of the product of $Z_1(x) - m_1$ and $T(x+h) - m_T$ is zero for any x and h, the two random functions $(Z_1(x) - m_1)$ and $(T(x) - m_T)$ are said to be *orthogonal.*

2. The variable $T(x)$, which has the covariance:

$$9\ Sph\ (100) + 9\ Sph\ (20)$$

can be split into two spatially uncorrelated components with ranges of *100* and *20* respectively. But, in contrast to $Z_1(x)$ and $T(x)$, these components cannot be obtained experimentally at a given point x from $Z_1(x)$ and $Z_2(x)$.

2. *Indicator Functions*

Linear geostatistics deals only with linear combinations of the variable under study Y(x). This makes it possible to estimate linearly the value of the variable at a given point or over a given domain.

Sometimes we need to estimate, not Y(x) itself, but one or more functions f[Y(x)] of Y(x). Considering linear combinations of the variable Y(x) is no longer sufficient and we call on non linear methods: in particular the conditional expectation that will be given in chapter 7, and disjunctive kriging introduced in the present chapter.

What makes the indicators interesting is the fact that any function f[Y(x)] can be expressed in terms of the indicators of Y(x). By cokriging these indicators we obtain the DK estimate.

2.1 Coding through Indicators

Suppose that a random function $Y(x)$ can take a finite number of values. If, for example, these values are *0, 1, 2* and *3* then at each point x $Y(x)$ must take just one of them. That is,

$$Y(x) = 0 \text{ or } 1 \text{ or } 2 \text{ or } 3$$

We may *code* these different possibilities by using indicator functions (called indicators, for short). If $Y(x)=i$ then the indicator $1_{Y(x)=i}$ is equal to *1*, and to *0* otherwise. Since $Y(x)$ can take only one value at any one place, this coding is said to be *disjunctive*. At any point x, one and only one of the indicators is equal to *1*, and the others are all zero. So:

$$\sum_i 1_{Y(x)=i} = 1_{Y(x)=0} + 1_{Y(x)=1} + 1_{Y(x)=2} + 1_{Y(x)=3} = 1$$

Thus we need to know only three of the four indicators in order to deduce the other one.

2.2 Developing functions

This disjunctive coding and these indicators are valuable because they allow us to express any function of $Y(x)$ conveniently. Let us consider, for example, a function $f(Y(x))$ that takes the values f_0, f_1, f_2 and f_3 when $Y(x)$ takes the values *0, 1, 2* and *3*. This function can be written as:

$$\sum_i f_i \, 1_{Y(x)=i} = f_0 \, 1_{Y(x)=0} + f_1 \, 1_{Y(x)=1} + f_2 \, 1_{Y(x)=2} + f_3 \, 1_{Y(x)=3}$$

We can see this by examining what happens when $Y(x)$ takes each value. If $Y(x)=0$ then $1_{Y(x)=0} = 1$ and the other three indicators are zero. So:

$$\sum f_i \, 1_{Y(x)=i} = f_0 \times 1 + f_1 \times 0 + f_2 \times 0 + f_3 \times 0 = f_0$$

Similarly for the other three possible values.

Examples

1. The variable $Y(x)$ itself can be written using indicators:

$$Y(x) = 0 \, 1_{Y(x)=0} + 1 \, 1_{Y(x)=1} + 2 \, 1_{Y(x)=2} + 3 \, 1_{Y(x)=3}$$

2. In geostatistics we often use indicators above a cut-off value. For example, the indicator $1_{Y(x) \geq 2}$ takes the value *1* when $Y(x) \geq 2$, and *0* otherwise. So we have:

$$1_{Y(x) \geq 2} = 1_{Y(x)=2} + 1_{Y(x)=3}$$

3. If $Y(x)$ represents the grade in a particular substance (e.g. metal concentration) of a sample taken at point x, the variable in example 2 indicates whether the grade equals or exceeds the cutoff at *2*. When multiplied by the weight of the sample this indicator represents the weight of the ore above the cutoff.

4. Similarly the quantity of metal at the *2* cutoff contained in the sample is taken as *0* if the grade $Y(x)$ is less than *2*, and is equal to the product of the grade by the weight of the sample if $Y(x) \geq 2$. So it is equal to:

$$Y(x) \, 1_{Y(x) \geq 2}$$

multiplied by the weight of the sample.

2.3 Limits of linear geostatistics

As the reader knows, linear geostatistics deals only with linear combinations of the variable under study $Y(x)$. This makes it possible to estimate the value of the variable $Y(x)$ at a given point x, or its regularized $\frac{1}{V}\int_V Y(x)\,dx$ over a given domain V, using a linear combination of the values $Y(x_\alpha)$ known at sample points x_α (for instance kriging).

Linear geostatistics also provides the variance of estimation (i.e. the variance of the difference between true and estimated values). A small estimation variance means a good precision.

Now let us return to our variable $Y(x)$ which can take the values 0, 1, 2, or 3. Imagine that its kriged estimate $Y(x)^K$ at a given point x equals 1. If the estimation variance is small,

there is a high chance for the true value $Y(x)$ not to exceed 2. But if the estimation variance is not small, as usually happens, the estimated value $Y(x)^K = 1$ only represents an average of tendencies. The true value may be equal to 1, but it may also equal 0, 2 or 3. The fact that the estimate $Y(x)^K$ is smaller than 2 does not mean that the true value $Y(x)$ cannot be equal to or larger than 2, which is represented by the indicator $1_{Y(x) \geq 2}$.

This distinction is essential when we want to make predictions relative to cut–off values. In the environmental sciences for instance, we may be interested in estimating in the form of a probability whether a variable $Y(x)$ passes a threshold, e.g. $1_{Y(x) \geq 2}$. To predict the results of a selective mining process, we want to estimate indicators such as $1_{Y(x) \geq 2}$ (which gives the ore above the cut–off 2), and also the corresponding quantities of metal.

As we can see, computing an estimate for the variable $Y(x)$ is not the solution when we need to estimate, not $Y(x)$ itself, but one or more functions $f[Y(x)]$ of $Y(x)$: indicators, metal quantities, etc. We then call on non linear methods like the conditional expectation (chapter 7) or disjunctive kriging.

2.4 Disjunctive Kriging DK

As we have seen, the family of indicators $1_{Y(x)=0}$, ... , $1_{Y(x)=3}$ allows us to express any function of $Y(x)$ as:

$$f(Y(x)) = f_0 \, 1_{Y(x)=0} + \ldots$$

If we can estimate all of the indicators, we can obtain an estimate of $f[Y(x)]$ for any function f. The disjunctive kriging estimator is defined as that given by cokriging all the indicators:

$$[f(Y(x))]^{DK} = f_0 \, [1_{Y(x)=0}]^{CK} + f_1 \, [1_{Y(x)=1}]^{CK} + \ldots$$

In particular the DK of an indicator is equal to its cokriging with all the indicators:

$$[1_{Y(x)=0}]^{KD} = [1_{Y(x)=0}]^{CK}$$

Note that disjunctively kriging the sum (or similarly the difference) of two functions is equivalent to summing the two DK estimates. That is:

$$[f(Y(x)) + g(Y(x))]^{DK} = [f(Y(x))]^{DK} + [g(Y(x))]^{DK}$$

This ensures consistency between the estimates of different functions.

2.5 Other Generating Families

Since in our opening example $1_{Y(x)=0} = 1 - \sum_{1}^{3} 1_{Y(x)=i}$, we can write:

$$f(Y(x)) = f_0 \left(1 - \sum_1^3 1_{Y(x)=i} \right) + \sum_1^3 f_i \, 1_{Y(x)=i}$$

$$= f_0 \, 1 + \sum_1^3 (f_i - f_0) \, 1_{Y(x)=i}$$

This shows that $(1 , 1_{Y(x)=1} , 1_{Y(x)=2} , 1_{Y(x)=3})$ is also a generating family.

Sometimes, it is more convenient to consider the cumulative indicators:

$$1_{Y(x)\geq 0} = 1$$

$$1_{Y(x)\geq 1} = 1_{Y(x)=1} + 1_{Y(x)=2} + 1_{Y(x)=3}$$

$$1_{Y(x)\geq 2} = 1_{Y(x)=2} + 1_{Y(x)=3}$$

$$1_{Y(x)\geq 3} = 1_{Y(x)=3}$$

Clearly we have:

$$1_{Y(x)=3} = 1_{Y(x)\geq 3}$$

$$1_{Y(x)=2} = 1_{Y(x)\geq 2} - 1_{Y(x)\geq 3}$$

$$1_{Y(x)=1} = 1_{Y(x)\geq 1} - 1_{Y(x)\geq 2}$$

$$1_{Y(x)=0} = 1 - 1_{Y(x)\geq 1}$$

Since any function of $Y(x)$ can be expressed in terms of the indicators $(1_{Y(x)=0} , \ldots , 1_{Y(x)=3})$, it can also be written in terms of $(1 , 1_{Y(x)\geq 1} , \ldots , 1_{Y(x)\geq 3})$, which also forms a generating family.

2.6 What do we need for Disjunctive Kriging ?

All the generating families are equivalent. So we may cokrige using $(1_{Y(x)=0} , \ldots , 1_{Y(x)=3})$ or using $(1, 1_{Y(x)\geq 1} , \ldots , 1_{Y(x)\geq 3})$: let us consider the second case. We clearly must know the auto and cross covariance functions (or variograms) of the cumulative indicators.

Let T_i denote the probability that $Y(x)$ exceeds or equals i, which is the expectation of the indicator $1_{Y(x)\geq i}$:

$$T_i = P\,(Y(x) \geq i\,) = E\,[1_{Y(x) \geq i}]$$

T_i is just the inverse cumulative distribution function of $Y(x)$. We can write the covariance functions as:

$$C_1\,(h) = Cov\,[\,1_{Y(x) \geq 1}\,,\,1_{Y(x+h) \geq 1}]$$

$$= E\,[\,1_{Y(x) \geq 1}\,.\,1_{Y(x+h) \geq 1}] - (E\,[\,1_{Y(x) \geq 1}])^2$$

and similarly the cross–covariance functions as:

$$C_{12}\,(h) = Cov\,[\,1_{Y(x) \geq 1}\,,\,1_{Y(x+h) \geq 2}\,]$$

$$= E\,[\,1_{Y(x) \geq 1}\,.\,1_{Y(x+h) \geq 2}] - E\,[\,1_{Y(x) \geq 1}]\,E\,[\,1_{Y(x+h) \geq 2}]$$

Let T^h_{ij}, where the superscript means at lag h, denote the bivariate distribution of $Y(x)$ and $Y(x+h)$:

$$T^h_{ij} = P(Y(x) \geq i\,,\;\; Y\,(x+h) \geq j\,) = E\,[\,1_{Y(x) \geq i}\,.\,1_{Y(x+h) \geq j}]$$

We have:

$$C_1\,(h) = T^h_{11} - (T_1)^2$$

$$C_{12}\,(h) = T^h_{12} - T_1\,T_2$$

This gives:

$$T^h_{11} = C_1\,(h) + (T_1)^2$$

$$T^h_{12} = C_{12}\,(h) + T_1\,T_2$$

So knowing the auto and cross covariance functions is equivalent to knowing the bivariate distributions of $Y(x)$ and $Y(x+h)$ for any lag h. In the same way that kriging is based on the variogram, so disjunctive kriging is based on these bivariate distributions.

In practice we have to fit a model to the bivariate distributions, or what amounts to the same thing, to fit a model to the coregionalisation of the indicators. Using this model makes it possible to perform DK (i.e. cokriging of indicators).

However the size of a cokriging system increases with the number of indicators – here the number of indicator variables. So as before we look for models that simplify the cokriging. These are the mosaic model presented in the next chapter or isofactorial models like those we will see afterwards.

3. The Mosaic Model: Indicators that are Intrinsically Correlated

The mosaic model is a very simple one in which the auto and cross structures of all functions of Y(x) are the same. In consequence the indicator cokriging (i.e. the disjunctive kriging) is reduced to a kriging.

Suppose we have a region that consists of a mosaic of random compartments in any one of which the value of $Y(x)$ is constant. Moreover, the value in each compartment is independent of the values in the other compartments, but the values all come from a common distribution. An example is given in Figure 1.

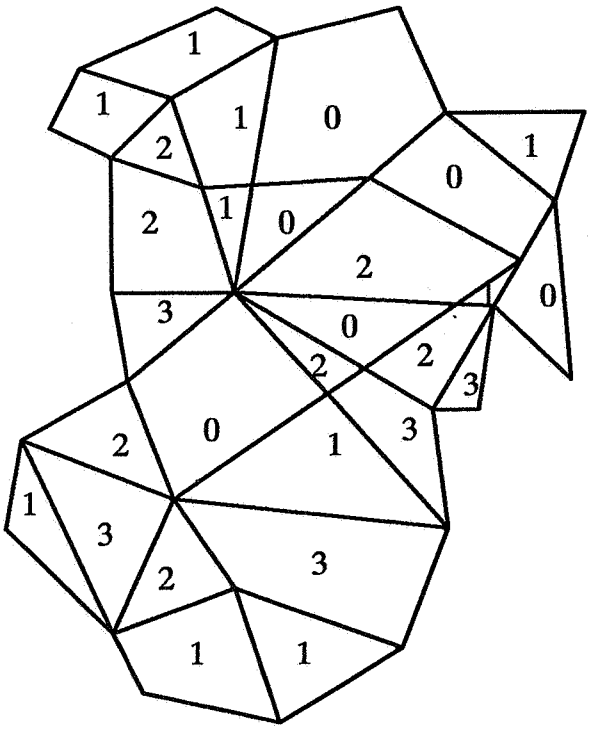


Figure 1: Example of mosaic

Let $\rho(h)$ be the probability that two points, x and $x + h$, a distance h apart, belong to the same compartment. If they do, then $Y(x) = Y(x+h)$. If x and $x + h$ belong to different compartments, which happens with a probability $1 - \rho(h)$, then $Y(x)$ and $Y(x+h)$ are independent.

So we have:

$$E [Y(x) Y(x + h)] = E [Y(x)^2] \varrho (h) + [E (Y(x))]^2 (1 - \varrho (h))$$

$$= \varrho\,(h)\ Var\ Y(x) + [\ E\ (Y(x))\ ]^2$$

This gives us the covariance:

$$Cov\ [Y(x),\ Y(x+h)] = \varrho\,(h)\ Var\ Y(x)$$

Similarly for any function $f(Y(x))$:

$$Cov\ [\ f\ (Y(x)),\ f\ (Y(x+h))\ ] = \varrho\,(h)\ Var\ [f\ (Y(x))]$$

If we now consider any two functions $f(Y(x))$ and $g(Y(x))$, then it is easy to show that:

$$Cov\ [f\ (Y(x)),\ g\ (Y(x+h))] = \varrho\,(h)\ Cov\ [f\ (Y(x)),\ g\ (Y(x))]$$

So the auto and cross covariances of all functions of $Y(x)$ are proportional to $\rho\,(h)$. In particular this is true of the indicator functions, which are therefore intrinsically correlated (see 1.3).

So cokriging them is equivalent to kriging them. We need to solve only a single kriging system, the one based on $\rho\,(h)$, to get the kriging weights for all the indicator functions, and for all possible functions of $Y(x)$:

$$[\ f\ (Y(x))\ ]^{DK} = \sum_i f_i\ [\ 1_{Y(x)=i}\ ]^{CK}$$

$$= \sum_i f_i\ [\ 1_{Y(x)=i}\ ]^{K}$$

$$= [\ \sum_i f_i\ 1_{Y(x)=i}\ ]^{K} = [\ f\ (Y(x))\ ]^{K}$$

As an exercise readers should write out the above krigings explicitly in terms of the kriging weights λ_α, the data $1_{Y(x_\alpha)=i}$ and $f\ (Y(x_\alpha))$, and the mean values $E\ [1_{Y(x)=i}] = P\ [Y(x) = i\ ]$ and $E\ [f(Y(x))\ ]$ which have been assumed to be known. As in current use in geostatistics the x_α denote the sampling points in the neighbourhood of the point x.

Thus the DK of any function is very simple. It is reduced to a kriging, with weights that are the same for all functions. This makes the mosaic model very attractive. This model can be tested experimentally by comparing the auto and cross structures of some indicators, that must then be identical.

In the models we are now going to consider the indicators do not all have the same structure. Moreover there always exists some correlation between the indicators. This means that if we just krige the indicators separately the resulting estimate may be far from the optimal one (see exercise 4.3 in the next chapter).

4. *Isofactorial Model with no Edge Effects: Orthogonal Indicator Residuals*

In the model presented here, there are no edge effects: the edges of the set of points above a given cut–off are not poorer than the center.

It is an isofactorial model, which means that DK (i.e. indicator cokriging) is obtained by kriging the factors – here the indicator residuals.

4.1 Isofactorial Models

We have seen in 1.4 that cokriging different variables simplifies when these variables can be factorized. Then we need only to krige each factor separately.

It is the same thing for the indicator variables of a random function $Y(x)$. Isofactorial models are models where the different indicators are factorized. Then cokriging the indicators, i.e. DK, is equivalent to kriging the factors.

There are two types of isofactorial models: those with edge effects on the regions defined by cutoffs, and those without them.

4.2 Edge Effects

Let us see what is meant by edge effects. As before, we consider a random function $Y(x)$ that takes four values: *0, 1, 2, 3*. Let A_1 denote the set of points where the value of $Y(x)$ equals or exceeds *1*. So for each point x in this set we have:

$$1_{x \in A_1} = 1_{Y(x) \geq 1} = 1$$

Of course, the value of $Y(x)$ for a point x inside A_1 might also equal or exceed *2*. But this may have less chance of happening when the point x is near the edge of A_1. That is, there is an *edge effect* because the border areas of A_1 are, on average, of lower grade than is the centre. Figure 4–1 illustrate this. In this case, the value at a point x inside A_1 depends on whether a neighbouring point $x+h$ belong to A_1 or not. In particular the probability:

$$P\,[Y(x) \geq 2 \mid Y(x) \geq 1,\ Y(x+h) < 1] = \frac{P\,[Y(x) \geq 2,\ Y(x+h) < 1]}{P\,[Y(x) \geq 1,\ Y(x+h) < 1]} \qquad (4\text{–}1)$$

will depend on h. In general, it increases with increasing h at least for short distances.

This can be seen experimentally using the variograms of the indicators. For example, the variogram for the indicator $1_{Y(x) \geq 1}$ can be written as:

$$\gamma_1(h) = \frac{1}{2} E\left[1_{Y(x) \geq 1} - 1_{Y(x+h) \geq 1} \right]^2$$

So it is the mean value of a term which takes the value *1* if $Y(x) \geq 1$ and $Y(x + h) < 1$ or if $Y(x) < 1$ and $Y(x + h) \geq 1$, and takes the value *0* otherwise. So:

$$\gamma_1(h) = \frac{1}{2} \left\{ P\left[Y(x) \geq 1,\ Y(x+h) < 1\right] + P[Y(x+h) \geq 1,\ Y(x) < 1] \right\}$$

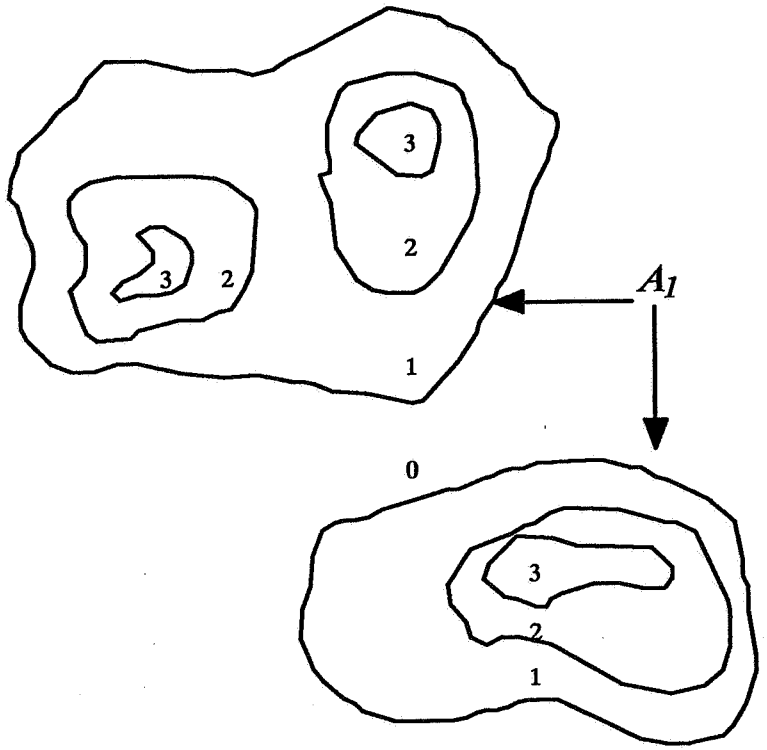


Figure 4–1

Example of Edge Effects. In this case, the border areas of A_1 are poorer than the centre. The same is also true for A_2, the set consisting of points where $Y(x) \geq 2$. So these are edge effects on these sets.

If we assume that the two probabilities are equal (that is, the distribution is symmetric in h):

$$\gamma_1(h) = P[Y(x) \geq 1,\ Y(x+h) < 1]$$

Similarly:

$$\gamma_{21}(h) = \frac{1}{2} E\left[1_{Y(x) \geq 2} - 1_{Y(x+h) \geq 2} \right] \left[1_{Y(x) \geq 1} - 1_{Y(x+h) \geq 1} \right]$$

is equal to:

$$P\,[Y(x) \geq 2,\ Y(x+h) < 1]$$

Consequently we have:

$$P\,[Y(x) \geq 2 \mid Y(x) \geq 1,\ Y(x+h) < 1] = \frac{\gamma_{21}(h)}{\gamma_1(h)} \qquad (h \neq 0)$$

Using the indicator variograms it is easy to see whether there are edge effects. In general, the ratio given above increases with h for small distances, which shows the presence of an edge effect. As the border area of A_1 is, in general, poorer than its centre, going from a point where the value is small to one where it is large (*2* or more) generally means going through a zone of intermediate values (e.g. *1*).

4.3 Absence of Edge Effect : Orthogonal Indicator Residuals

We still consider the random function $Y(x)$ which takes the values *0, 1, 2, 3* and we assume that there are no edge effects on the sets A_1 (contains the points where $Y(x) \geq 1$) and A_2 (where $Y(x) \geq 2$).

For any point x in A_1 (i.e. where $Y(x) \geq 1$), the fact that its value equals or exceeds *2* is independent of a neighbouring point $x+h$ belonging to A_1. That is, the indicator $1_{Y(x) \geq 2}$ is independent of $1_{Y(x+h) \geq 1}$. So:

$$P[Y(x) \geq 2, Y(x+h) \geq 1 \mid Y(x) \geq 1] = P[Y(x) \geq 2 \mid Y(x) \geq 1]\ P[Y(x+h) \geq 1 \mid Y(x) \geq 1]$$

And therefore:

$$\frac{P[Y(x) \geq 2,\ Y(x+h) \geq 1]}{P(Y(x) \geq 1)} = \frac{P(Y(x) \geq 2)}{P(Y(x) \geq 1)}\ \frac{P[Y(x+h) \geq 1,\ Y(x) \geq 1]}{P(Y(x) \geq 1)}$$

Using the notation introduced in 2.5, we have:

$$T^h_{21} = T^h_{11}\ \frac{T_2}{T_1} \qquad (4\text{–}2)$$

More generally, the bivariate distribution for $j > i$ of these models satisfies the relation:

$$T^h_{ji} = T^h_{ii}\ \frac{T_j}{T_i} \qquad (4\text{-}3)$$

As a general rule, the regression between indicator functions at a given point x can be written as:

$$E\,[1_{Y(x) \geq 2} \mid 1_{Y(x) \geq 1}] = \frac{T_2}{T_1}\ 1_{Y(x) \geq 1}$$

This is clear as one of two things must happen:

– either the conditioning variable $1_{Y(x) \geq 1}$ is zero (i.e. $Y(x) < 1$) and so $1_{Y(x) \geq 2} = 0$;

– or alternatively $1_{Y(x) \geq 1}$ is equal to *1*, then $1_{Y(x) \geq 2}$ is equal to *1* with the probability $P(Y(x) \geq 2 \mid Y(x) \geq 1) = \frac{T_2}{T_1}$ or to *0* otherwise, and so its mean value is $\frac{T_2}{T_1}$.

So the regression can be written as above. Its residual $1_{Y(x) \geq 2} - \frac{T_2}{T_1} 1_{Y(x) \geq 1}$ clearly has a zero mean and is uncorrelated with the conditioning variable $1_{Y(x) \geq 1}$.

When there is no edge effect, this residual is also orthogonal to $1_{Y(x+h) \geq 1}$:

$$E\left[\left(1_{Y(x)\geq 2} - \frac{T_2}{T_1} 1_{Y(x)\geq 1}\right) 1_{Y(x+h)\geq 1}\right] = T_{21}^h - \frac{T_2}{T_1} T_{11}^h = 0$$

Such residuals play a special role in this model. Let us consider the family consisting of:

$$H_0(Y(x)) = 1$$

$$H_1(Y(x)) = \frac{1_{Y(x) \geq 1}}{T_1} - 1$$

$$H_2(Y(x)) = \frac{1_{Y(x) \geq 2}}{T_2} - \frac{1_{Y(x) \geq 1}}{T_1}$$

$$H_3(Y(x)) = \frac{1_{Y(x) \geq 3}}{T_3} - \frac{1_{Y(x) \geq 2}}{T_2}$$

For $i > 0$, $H_i(Y(x))$ is equal to the residual of $\frac{1_{Y(x) \geq i}}{T_i}$ starting from the lower indicator $1_{Y(x) \geq i-1}$. The multiplicative factor $\frac{1}{T_i}$ permits to simplifie the writings. Then we have:

$$\sum_0^j H_i(Y(x)) = \frac{1_{Y(x) \geq j}}{T_j}$$

which means that these residuals form a generating family. Moreover, from [4-3], it follows that they are orthogonal. So these residuals factorize the indicators. They are the factors of the model whose bivariate distributions satisfy (4–3), which then is isofactorial. When using this model, we have only to krige each of the factors separately in order to cokrige the indicators or equivalently for the disjunctive kriging of any function:

$$\frac{[\ 1_{Y(x) \geq j}\ ]^{DK}}{T_j} = \sum_0^j [\ H_i(Y(x))\ ]^K$$

It is easy to extend these results to more than 4 values. However, if a variable takes many values, then these should be grouped into a few classes.

The model can be tested experimentally by comparing the indicator cross variograms $\gamma_{ij}(h)$ with the auto variograms $\gamma_i(h)$ for $j > i$.

4.4 Important Exercise

Let $Y(x)$ be a stationary Random Function taking the values 0, 1, 2, 3, ... Suppose that the cross–variograms of its indicators $\gamma_{ij}(h)$ are proportional to the auto variograms $\gamma_i(h)$ for $j > i$. So the indicator residuals are orthogonal. Once they have been calculated at the sampling points, we observe that:

– the first residual $H_1\,(Y(x)) = \dfrac{1_{Y(x) \geq 1}}{T_1} - 1$, and hence $1_{Y(x) \geq 1}$, are spatially well structured;

– the other residuals $H_2\,(Y(x)) = \dfrac{1_{Y(x) \geq 2}}{T_2} - \dfrac{1_{Y(x) \geq 1}}{T_1}$, $H_3\,(Y(x))$, etc ..., are pure nugget.

An example of this might be a set of mineralised veins, with a geometry which is structured, but within which the grades are completely random. The lowest cut–off, as it makes the distinction between the mineralised veins and their waste surroundings, defines the geometry of the veins.

1) First residual: show that the disjunctive kriging of the indicator $1_{Y(x) \geq 1}$, which then represents the geometry of veins, is equal to its kriging.

2) Other residuals: show that, at an unknown point, their kriged values are equal to 0. Deduce that the DK estimation of the indicators $1_{Y(x) \geq i}$ can be writen:

$$(1_{Y(x) \geq j})^{DK} = (1_{Y(x) \geq 1})^{K}\ P(Y(x) \geq j \mid Y(x) \geq 1)$$

Note that any local estimation is based on the estimation of the geometry. This agrees with the fact that the grade at a point x, *knowing it belongs to the veins*, is pure nugget, and then cannot be the object of a really local estimation. Whether the *mineralised* neighbouring grades are 1, 2 or 3, has no influence on the estimation of the different indicators.

We are far from the indicator kriging, proper to the mosaic model. This would estimate for example $1_{Y(x) \geq 3}$ using the indicators known for the same cut–off at 3, and would then give an illusory local influence to the high grades. All this clearly shows that **the model, on which the estimation algorithm depends, has to be chosen to represent the reality at best.**

4.5 Constructing the model

Constructing models with no edge effects on the sets A_i is simple.

a) Start by assigning the largest value (e.g. *3*) to all the points in the region.

b) Then put the next value (e.g. *2*) instead of *3* in some parts of the region.

c) For another independent set of sub–regions, put the value *1* (instead of either *2* or *3*).

d) Finally, for a third set that is independent of both of the previous ones, put the smallest value *(0)*.

These steps are shown in Figure 4–2 a to d.

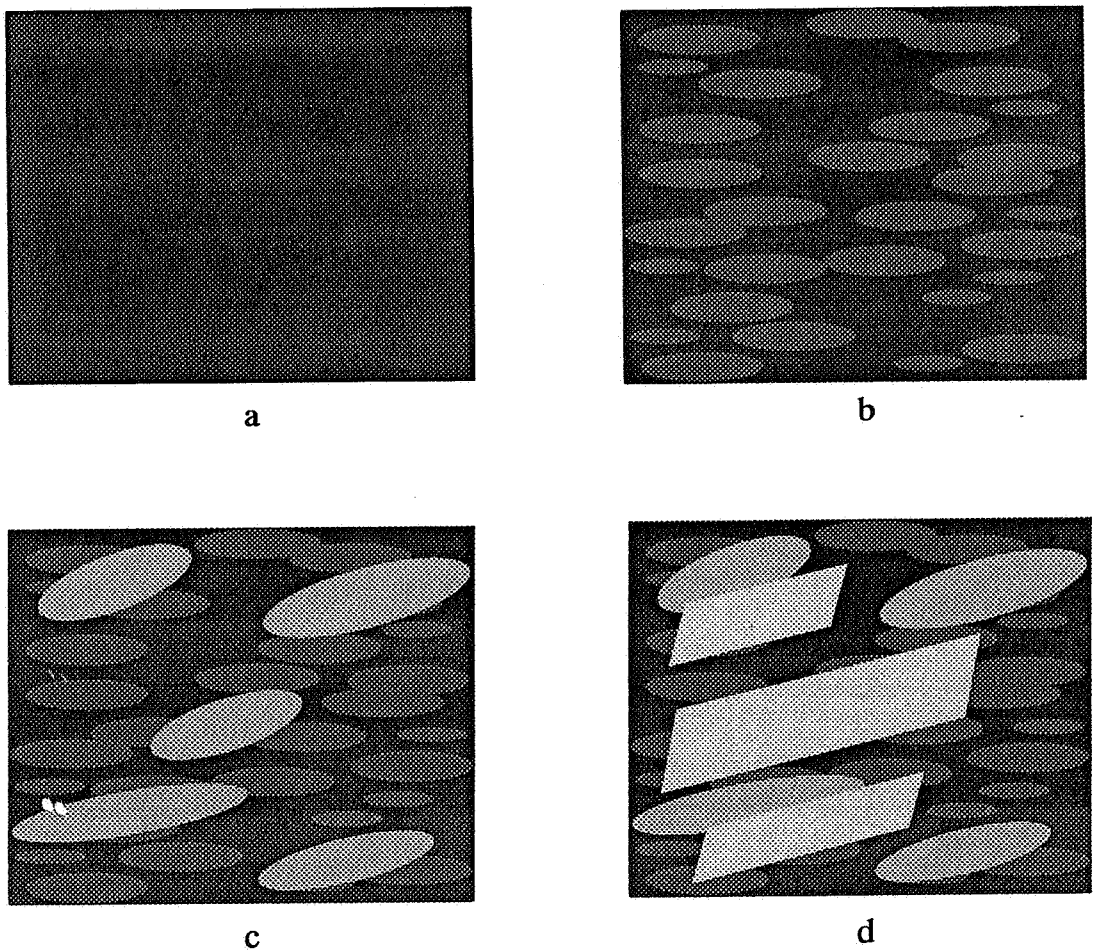


Figure 4–2

The procedure can be reversed by starting from the smallest value and working to the largest. We obtain the same model, which is just inverted. In this case there are no edge effects for the sets of points *up to* each cutoff, and the factors are the residuals of each indicator $1_{Y(x) < i}$ starting from the indicator for the *larger* cutoff grade $1_{Y(x) < i+1}$. The cross variogram $\gamma_{ij}(h)$ between two indicators is proportional now to the variogram of the larger cutoff $\gamma_j(h)$ $(j > i)$.

The way the figure 4–2–d has been constructed makes that there is no edge effect when leaving the small values. But when leaving the large values, we meet medium values preferentially to small ones. So there are edge effects in this direction. Note that in the mosaic model of the chapter 3 there are no edge effects in either direction (i.e. working upwards or downwards). Lastly, in the models described in the following chapters (and in figure 4–1 as the reader will realize), there are edge effects in both directions.

5. *Isofactorial Models with Edge Effects: the Example of the Gaussian Model*

The gaussian model is the most commonly used model with edge effects. It is an isofactorial model, the factors of which are the Hermite polynomials. DK is obtained by kriging these factors separately.

In pratice the variable under study seldom has a normal distribution. So a transformation is needed to get to it.

5.1 Diffusion Model

In general edge effects exist, and in order to go from a point with a value *(i–1)* to a point with a value *(i + 1)*, we go through points that take the intermediate value *i*. We then use "diffusion models" . We shall not go into the theory in detail here, but it is based on stochastic diffusion processes. Once the process has reached a certain value *i* at a particular time, it can remain at that value, or move to a neighbouring state – either *(i + 1)* or *(i–1)*– but it cannot jump straight to any other value.

These models exist both when the values taken by the variable are discrete and when they are continuous. An example of the latter is the commonly used gaussian model.

5.2 Gaussian Model

In this model the variable *Y(x)* which is assumed to have zero mean and unit variance, has a normal (i.e. gaussian) distribution. That is, its pdf is:

$$g(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

which when graphed has the familiar form in Figure 5–1. Its cumulative distribution function is the integral (Figure 5–2):

$$G(u) = \int_{-\infty}^{u} g(t)\, dt$$

There is no convenient analytical expression for this function. However, its values have been worked out and tabulated, and it can be approximated numerically when using computers.

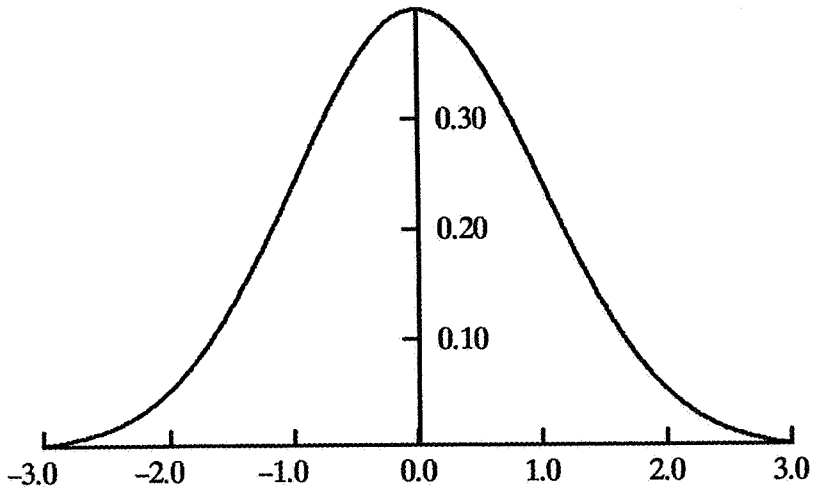


Figure 5–1

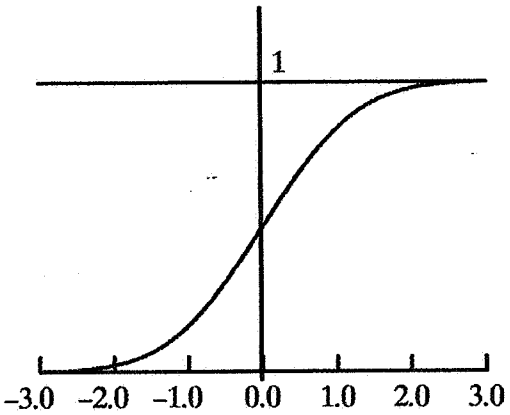


Figure 5–2

The pairs of points *(Y(x), Y(x+h))* are assumed to have a bivariate normal distribution (i.e. any linear combination of them is normal) which is characterized by their correlation coefficient $\rho = \rho(h)$. Their pdf can be written as:

$$g_\varrho(t,u) = \frac{1}{2\pi\sqrt{1-\varrho^2}}\, e^{-\frac{t^2+u^2-2tu\varrho}{2(1-\varrho^2)}}$$

Consequently their isodensity curves are the ellipses:

$$t^2 + u^2 - 2\,t\,u\,\rho = constant$$

The scatter diagram of the pairs of values is elliptic (figure 5–3), and all the narrower that the correlation is higher. When $\rho(h)$ is zero the variables $Y(x)$ and $Y(x+h)$ are independent and the scatter diagram becomes round.

Figure 5–3

5.3 Hermite Polynomials

Hermite polynomials $H_n\,[Y(x)]$ are polynomials that have special properties related to the normal distribution. They are defined by Rodrigues's formula ($n \geq 0$):

$$H_n(y) = \frac{1}{\sqrt{n!}\ g(y)} \frac{d^n\, g(y)}{dy^n} \qquad (5\text{–}1)$$

where $\frac{1}{\sqrt{n!}}$ is a normalisation factor. The Hermite polynomial $H_n(y)$ is a polynomial of degree n. More specifically,

$$H_0(y) = 1$$

$$H_1(y) = -y$$

$$H_2(y) = \frac{1}{\sqrt{2}}\,(y^2 - 1)$$

The other polynomials are related and can be calculated using the recurrence relation ($n > 0$):

$$H_{n+1}(y) = -\frac{1}{\sqrt{n+1}}\, y\, H_n(y) - \sqrt{\frac{n}{n+1}}\, H_{n-1}(y)$$

So it is easy to calculate the Hermite polynomials starting from the value of the normal variable $Y(x)$. In practice, we generally need at most a few dozens of them.

Except for $H_0(Y(x))$ which is a constant (1), their means are:

$$E\,[\,H_n(Y(x))\,] = \int H_n(y)\, g(y)\, dy = 0$$

and, because of the normalisation factor, their variances are:

$$Var\,[\,H_n(Y(x))] = E\{[\,H_n(Y(x))]^2\} = 1$$

Lastly when $p \neq n$:

$$Cov\,[H_p(Y(x)),\ H_n(Y(x))\,] = E\,[\,H_p(Y(x))\, H_n(Y(x))\,] = 0$$

So for a given point x, the Hermite polynomials are orthogonal when $Y(x)$ is a standard normal variable.

5.4 Expressing Functions in Terms of Hermite Polynomials

Practically any function of $Y(x)$ can be expanded in terms of Hermite polynomials:

$$f\,[\,Y(x)] = f_0 + f_1 H_1\,[Y(x)] + f_2 H_2\,[Y(x)] + \ldots$$

$$= \sum_{n=0}^{\infty} f_n H_n\,[Y(x)]$$

Using the orthogonality of the Hermite polynomials, we find that

$$E\,[\,f(Y(x))\, H_n(Y(x))\,] = E\left[\left[\sum_{p=0}^{\infty} f_p\, H_p(Y(x))\,\right]\, H_n(Y(x))\right]$$

$$= \sum_{p=0}^{\infty} f_p\ E\,[\,H_p(Y(x))\, H_n(Y(x))]$$

is equal to f_n. This enables us to calculate these coefficients:

$$f_n = E\,[\,f(Y(x))\, H_n(Y(x))\,] = \int f(y)\, H_n(y)\, g(y)\, dy$$

for any given function f. In particular:

$$f_0 = E\,[f\,(Y\,(x))\,]$$

Example: the indicator function $1_{Y(x)<y_c}$

The coefficient of order n is:

$$f_n = \int 1_{y<y_c}\, H_n(y)\, g(y)\, dy = \int_{-\infty}^{y_c} H_n\,(y)\, g(y)\, dy$$

We find for $n=0$:

$$f_0 = G(y_c)$$

and for $n \geq 1$, from the definition [5-1] of the polynomials:

$$f_n = \frac{1}{\sqrt{n}}\, H_{n-1}\,(y_c)\, g(y_c)$$

Hence the expansion of the indicator:

$$1_{Y(x)<y_c} = G(y_c) + \sum_{n\geq 1} \frac{1}{\sqrt{n}}\, H_{n-1}\,(y_c)\, g(y_c)\;\; H_n\,[Y(x)] \qquad (5\text{–}2)$$

So any function of $Y(x)$, in particular any indicator of $Y(x)$, can be expanded in terms of the generating family $H_n\,(Y(x))$. We then need only to cokrige $H_n(Y(x))$ in order to deduce the cokriging of the indicators and hence the disjunctive kriging of any function.

Exercise:

Show that the variance of $f\,[Y(x)]$ is given by:

$$Var\, f[Y(x)] = \sum_{1}^{\infty} (f_n)^2$$

Hint: use the orthogonality of the polynomials.

5.5 Gaussian Disjunctive Kriging

When the pairs $(Y(x),\, Y(x+h))$ are bivariate standard normal with correlation $\rho\,(h)$, the Hermite polynomials have the following property:

$$E\,[H_n\,(Y(x+h))\;|\;Y(x)] = [\,\rho\,(h)]^n\, H_n\,(Y(x)) \qquad (5\text{–}3)$$

and so:

$$Cov\,[H_p\,(Y(x)),\; H_n\,(Y(x+h))] = E\,[H_p\,(Y(x))\; H_n\,(Y(x+h))]$$

$$= E\ [\ H_p(Y(x))\ E(H_n(Y(x+h))\ |Y(x))\ ]$$

$$= [\ \varrho\ (h)\ ]^n\ \ E\ [\ H_p\ (Y(x))\ H_n\ (Y(x))\ ]$$

Taking $p = n$ gives the spatial covariance of $H_n\ (Y(x))$ which is equal to $[\rho(h)]^n$; that is, the covariance of $Y(x)$ raised to the n^{th} power. As $\rho(h) < \rho(0) = 1$, the spatial dependence of $H_n(Y(x))$ decreases rapidly to nothing as n increases, i.e. the structure tends to pure nugget.

For $p \neq n$ the two Hermite polynomials $H_n(Y(x))$ and $H_p(Y(x))$ are spatially uncorrelated. They are the factors of the (bi–)gaussian model, which then is isofactorial. Consequently the polynomials have only to be kriged separately to give the cokriging of the indicators and hence the DK of any function of $Y(x)$:

$$f(Y(x))^{DK} = f_0 + f_1\ [H_1\ (Y(x))]^K + f_2\ [H_2\ (Y(x))]^K + \ldots \qquad (5\text{–}4)$$

If we denote:

x_α	the experimental points,
$\rho_{\alpha\beta} = \rho\ (x_\alpha - x_\beta)$	the covariance between $Y(x_\alpha)$ and $Y(x_\beta)$,
$H_n\ (Y_\alpha) = H_n\ [\ Y(x_\alpha)]$	the Hermite polynomials,

we have:

$$[\ H_n\ (Y(x))\ ]^K\ = \sum \lambda_{n\alpha}\ H_n\ (Y_\alpha)$$

where the $\lambda_{n\alpha}$ satisfy the system:

$$\sum_\beta \lambda_{n\beta}\ Cov\ [H_n(Y_\alpha), H_n(Y_\beta)]\ =\ Cov\ [\ H_n\ (Y_\alpha),\ H_n\ (Y(x))]$$

i.e:

$$\sum_\beta \lambda_{n\beta}\ [\ \varrho_{\alpha\beta}\ ]^n\ =\ [\ \varrho_{\alpha x}\ ]^n \qquad \text{for all } \alpha$$

As the correlation structure $[\rho(h)]^n$ of $H_n\ (Y(x))$ rapidly tends to one of pure nugget, the kriged estimator at an unknown point rapidly tends to its mean, that is, to zero. So even if the coefficients f_n are not negligible, we have to krige only fairly few polynomials in (5–4) to get the result. Usually it is less than a dozen.

Exercise:

Show that the kriging variance of $H_n\ (Y(x))$ can be written:

$$\sigma^2_{Kn}\ =\ 1 - \sum_\alpha \lambda_{n\alpha}\ [\ \varrho_{\alpha x}\ ]^n$$

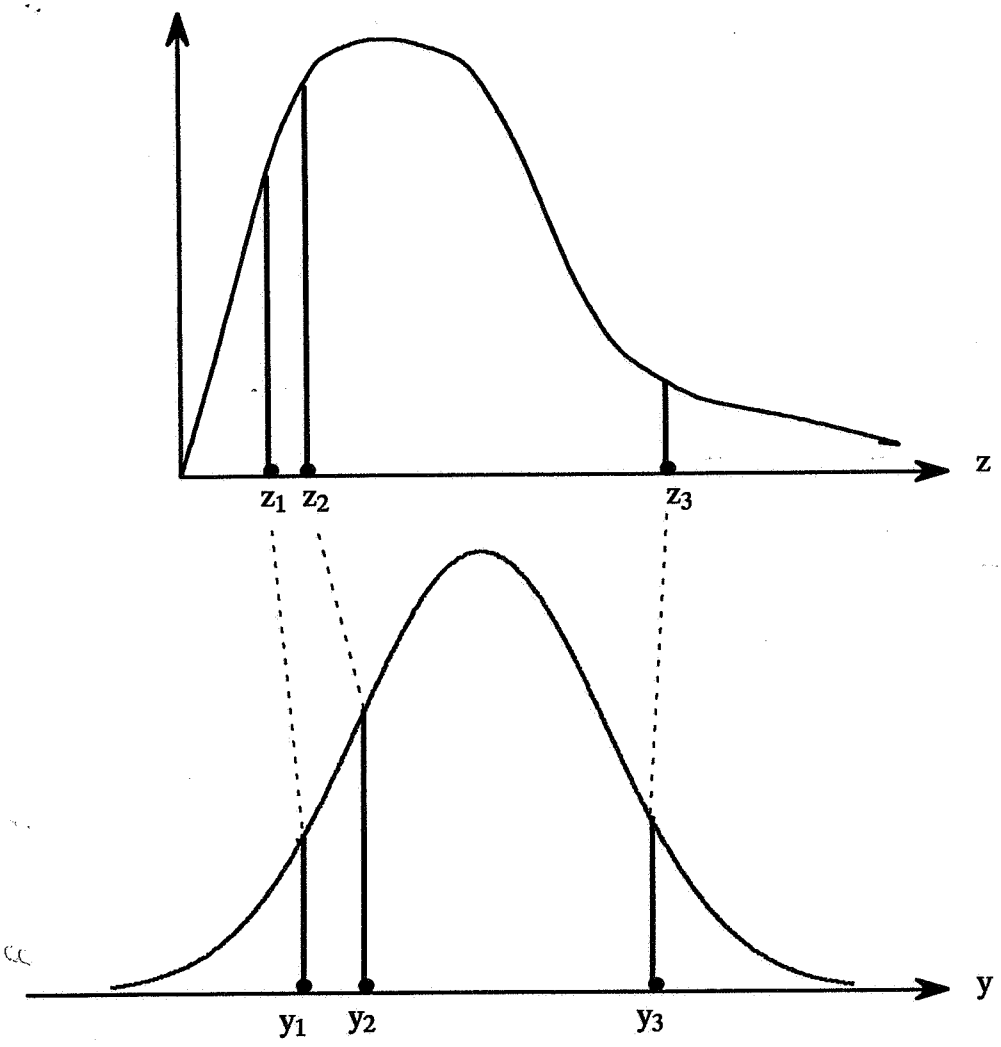
and that the DK estimation variance of $f\ [Y(x)]$ is given by:

$$Var\ (f[Y(x)] - f[Y(x)]^{KD})\ =\ \sum_1^\infty (f_n)^2\ \sigma^2_{Kn}$$

5.6 Gaussian Anamorphosis or a Transformation to Normality

Gaussian disjunctive kriging presupposes that the variables $(Y(x), Y(x+h))$ are bivariate normal, and consequently that $Y(x)$ is univariate normal. The variable studied, $Z(x)$, is very rarely normally distributed, so some sort of transformation (called anamorphosis) is required to convert it to a normal variate $Y(x)$. We then assume that the transformed pairs $(Y(x), Y(x+h))$ are bivariate normal.

This transformation converts the histogram of data $Z(x)$ into a standard normal distribution, Figure 5–4.



Figures 5–4

On the cumulative distribution functions of $Z(x)$ and $Y(x)$, it associates each value z with the value y which corresponds to the same cumulative probability, Figure 5–5.

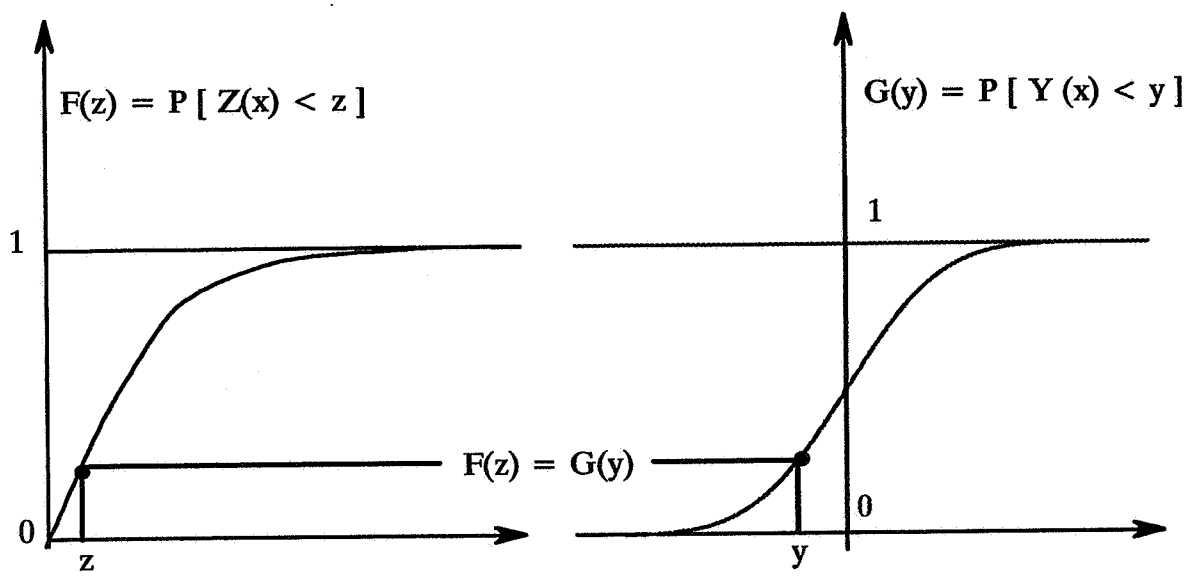


Figure 5–5

The function $z = \phi\ (y)$ that relates z and y is called the anamorphosis function, Figure 5–6.

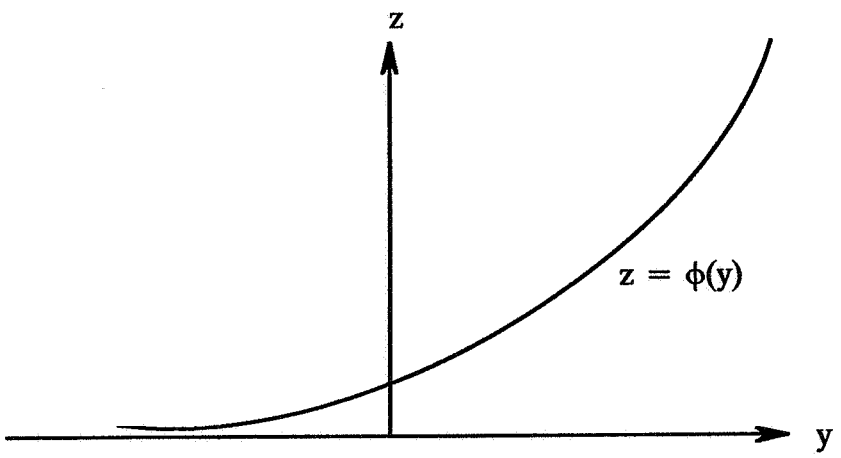


Figure 5–6

Comments:

1) Knowing the anamorphosis of the variable $Z(x)$ is equivalent to knowing its distribution. In particular the anamorphosis records the irregularities that are present in the experimental distribution and create problems when one wants to represent the distribution with a given statistical law (normal, lognormal, gamma, ...).

2) The anamorphosis of a normal variable is linear (see formula 5–5 further).

2) The lognormal case is presented in exercise 3, chapter 5.8.

5.7 Other Diffusion Models

It is not always possible to transform the variable $Z(x)$ to normality. Suppose for example that we are dealing with a concentration and that 50% of the values at the sampling points are zero. They correspond to the 50% of negative gaussian equivalents, but there is a problem deciding which gaussian value we should associate each zero. An arbitrary choice for each gives a normal distribution, but the local estimates obtained from these values could well be arbitrary too. (Moreover, the bivariate distribution will not generally be normal).

Other isofactorial diffusion models can be used in cases when:

- The distribution of $Z(x)$ contains "atoms" (that is, classes of identical values).

- The transformed variable $Y(x)$ is not bivariate normal.

In these cases, the following models may prove to be appropriate (Matheron 1973, 1976, 1984, Hu 1988):

- The hermitian model which is a generalisation of the preceding (bi–)gaussian one. The marginal distributions are normal and the factors are still Hermite polynomials.

- The bivariate gamma, and more generally, the Laguerre models. Here the marginal distributions are gamma and the factors are the Laguerre polynomials.

- The negative binomial models which are the discrete equivalent of the previous models.

Lastly a general method has been developed for constructing isofactorial models for the discrete case on an empirical basis. (Cf. Matheron 1984, Lajaunie and Lantuéjoul 1989).

5.8 Exercises

Here are some simple exercises designed to help readers work with the normal, bivariate normal and lognormal distributions.

1) A non standard normal variable

Suppose the mean and variance of a normally distributed random variable Z are m and σ^2 (instead of *0* and *1* for standard variables). In this case the variable $U = \frac{Z-m}{\sigma}$ is a standard normal variate. In other words, any normal variable can be written as:

$$Z = m + \sigma U \qquad (5-5)$$

where m is the mean, σ^2 the variance and U is the associated standard normal variate. This expression is very convenient as it allows us to express any normal variable in terms of a standard one for which we know the pdf and the cumulative distribution function.

Use this expression to show that the cumulative distribution function of Z can be written as:

$$P\,(Z < z) = G\left(\frac{z-m}{\sigma}\right)$$

Hence deduce the pdf of Z by differentiating this:

$$\frac{1}{\sigma}\, g\left(\frac{z-m}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{1}{2}\left(\frac{z-m}{\sigma}\right)^2}$$

2) Conditioned normal variable (important!)

Let (Y_0, Y_1) be a pair of bivariate normal random variables with a standardised distribution and a correlation coefficient ρ . Consider the random variable:

$$U = \frac{Y_0 - \varrho Y_1}{\sqrt{1-\varrho^2}} \qquad (5-6)$$

a) Show that:

$E(U) = 0$
$Var\,(U) = 1$
U is normally distributed.

Consequently U is a standard normal variate.

b) Secondly show that:

$$Cov\,(U\,,\,Y_1\,) = 0$$

So U and Y_1 are uncorrelated. Note that the pair (U, Y_1) is bivariate normal (any linear combination of U and Y_1 is normally distributed because it is also a linear combination of Y_0 and Y_1, which are bivariate normal). As U and Y_1 are uncorrelated and normally distributed, they are independent.

c) Show that 5–6 can be written as:

$$Y_0 = \varrho\, Y_1 + \sqrt{1-\varrho^2}\; U$$

where U is a standard normal variate that is independent of Y_1.
Consequently, if we know that $Y_1 = y_1$, we have:

$$(Y_0 \mid Y_1 = y_1) = \varrho\, y_1 + \sqrt{1-\varrho^2}\; U \qquad (5-7)$$

From this deduce that the variable Y_0 conditioned by $Y_1 = y_1$ is normally distributed with mean $E(Y_0 \mid Y_1 = y_1) = \rho\, y_1$ and with variance $1-\rho^2$.

If $Y_1 = y_1$ then we can use the value $E(Y_0 \mid Y_1 = y_1)$ as an estimate of Y_0. Note that the estimation variance of Y_0 given $Y_1 = y_1$, i.e. $1-\rho^2$, does not depend on y_1.

d) Deduce the regression equation of Y_0 knowing Y_1, that is:

$$E(Y_0 \mid Y_1) = \varrho\, Y_1$$

Note that it is linear.

e) Use the operation 5–7 to find the expression for the estimator of $1_{Y_0<y}$ given $Y_1 = y_1$:

$$E\left(1_{Y_0<y} \mid Y_1 = y_1\right) = G\left(\frac{y-\varrho\, y_1}{\sqrt{1-\varrho^2}}\right)$$

3) Lognormal variable

A variable Z is lognormal if its natural logarithm $\log Z$ is normally distributed. Let μ and σ^2 be the mean and variance of $\log Z$. We can then write:

$$\log Z = \mu + \sigma\, Y$$

where Y is a standard normal variate. Consequently:

$$Z = e^{\mu+\sigma Y} = e^{\mu}\, e^{\sigma Y}$$

So we see that in this particular case, the anamorphosis function $Z = \phi(Y)$ is an exponential. In the subsequent calculations we shall often use:

$$E\left[e^{\lambda U}\right] = e^{\frac{\lambda^2}{2}} \qquad (5-8)$$

if U is a standard normal variate.

a) Show that the mean of Z can be written as:

$$M = E[Z] = e^{\mu + \frac{\sigma^2}{2}}$$

From this deduce that Z can be written as:

$$Z = M e^{\sigma Y - \frac{\sigma^2}{2}} \qquad (5-9)$$

b) Use a similar procedure to calculate $E(Z^2)$ and hence deduce the variance of Z:

$$Var\ Z = M^2 (e^{\sigma^2} - 1)$$

4) (Bivariate) Lognormal regression

Suppose that Z_0 and Z_1 are lognormal. From (5–9) we have:

$$Z_0 = M_0 e^{\sigma_0 Y_0 - \frac{\sigma_0^2}{2}}$$

$$Z_1 = M_1 e^{\sigma_1 Y_1 - \frac{\sigma_1^2}{2}}$$

where Y_0 and Y_1 are standard normal variates and where they are bivariate normal with a correlation coefficient ρ. Knowing that $Z_1 = z_1$ is, of course, equivalent to knowing that $Y_1 = y_1$ with:

$$z_1 = M_1 e^{\sigma_1 y_1 - \frac{\sigma_1^2}{2}}$$

From 5–7, show that:

$$(Z_0 \mid Z_1 = z_1) = (M_0 e^{\sigma_0 Y_0 - \frac{\sigma_0^2}{2}} \mid Y_1 = y_1)$$

$$= M_0 e^{\sigma_0 (\varrho y_1 + \sqrt{1-\varrho^2} U) - \frac{\sigma_0^2}{2}} \qquad (5-10)$$

where U is a standard normal variate.

a) Use 5–8 to determine the mean of this distribution:

$$E(Z_0 \mid Z_1 = z_1) = E(Z_0 \mid Y_1 = y_1)$$

b) Given that $Z_1 = z_1$, we can use $Z_0^* = E(Z_0 \mid Z_1 = z_1)$ to estimate Z_0. Separate out Z_0^* as a factor in the formula 5–10 and show that the conditional variable is also lognormal. Calculate the variance of this variable (see exercise 3b). This is also the estimation variance of Z_0 given that $Z_1 = z_1$. Note that in contrast to the normally distributed case the estimation variance depends on the value of z_1.

c) Regression : $E(Z_0 \mid Z_1) = E(Z_0 \mid Y_1)$

Note that it is not the initial form of Z_0, where Y_0 would have been replaced improperly by $E(Y_0 \mid Y_1) = \varrho\, Y_1$!

Show that the regression can be written as:

$$E(Z_0 \mid Z_1) = c\,(Z_1)^{\frac{\varrho\sigma_0}{\sigma_1}}$$

where c is a constant. Note that this regression is not linear.

Remark: the bilognormal regression is sometimes used in Uranium mining to model the relation between the radiometric measures and the chemical uranium grades.

6. More about the Gaussian Model

In the preceding chapter we saw that we had only to krige each of the Hermite polynomials H_n (Y(x)) separately in order to deduce the disjunctive kriging estimator of any function of Y(x) provided that the pairs (Y(x), Y(x+h)) are bivariate normal. This chapter gives more details on disjunctive kriging that are needed for instance to program it.

6.1 Disjunctive Kriging of the Raw Variable

The transformed variable $Y(x)$ obtained from the raw variable $Z(x)$ is assumed to be bivariate normal. As we saw earlier, disjunctive kriging allows us to estimate any function of $Y(x)$ by using its expansion in terms of Hermite polynomials. So we can estimate $Z(x) = \Phi\,[Y(x)]$ this way once we have its expansion:

$$Z(x) = \Phi\,[Y(x)] = \phi_0 + \phi_1\, H_1\,[Y(x)] + \phi_2\, H_2\,[Y(x)] + \ldots$$

The function Φ can be determined graphically as was shown in figures 5-4 and 5-5. Here we shall express it in terms of the coefficients ϕ_n of its expansion. We now describe a way to calculate these (Lantuéjoul and Rivoirard 1984).

Suppose there are enough sample data for the distribution of $Z(x)$ to be known experimentally. Although there are often many different values of $Z(x)$, there are only a finite number of them. We can arrange them in increasing order:

$$z_1 < z_2 < z_3 < \ldots < z_l$$

with the frequencies:

$$p_1 \quad p_2 \quad p_3 \quad \ldots \quad p_l \qquad\qquad \left(\sum p_i = 1\right)$$

Their cumulative frequencies are:

$$F(z_1) = P\,(Z(x) < z_1) = 0$$

$$F(z_2) = P\,(Z(x) < z_2) = p_1$$

$$F(z_3) = P\,(Z(x) < z_3) = p_1 + p_2$$

...

$$F(z_i) = P\,(Z(x) < z_i) = \sum_{1}^{i-1} p_j$$

...

$$F(z_l) = P\,(Z(x) < z_l) = 1 - p_l$$

These cumulative frequencies correspond to gaussian equivalents (normal scores) y_i with the same cumulative frequency:

$$F(z_i) = G(y_i)$$

So we have:

$$F(z_{i+1}) - F(z_i) = G(y_{i+1}) - G(y_i)$$

That is:

$$P[z_i \leq Z(x) < z_{i+1}] = P[y_i \leq Y(x) < y_{i+1}]$$

$$P[Z(x) = z_i] = P[y_i \leq Y(x) < y_{i+1}]$$

So $Z(x)$ will be equal to z_i when the gaussian equivalent lies between y_i and y_{i+1}. This allows us to determine the coefficients of the transformation:

$$\phi_0 = E[\Phi(Y(x))] = E[Z(x)] = \sum p_i z_i$$

$$\phi_n = E[Z(x) H_n(Y(x))] = \int \Phi(y) H_n(y) g(y) dy$$

$$= \sum_{i=1}^{l} \int_{y_i}^{y_{i+i}} z_i H_n(y) g(y) dy$$

$$= \sum_{i=1}^{l} z_i \left[\frac{1}{\sqrt{n}} H_{n-1}(y_{i+1}) g(y_{i+1}) - \frac{1}{\sqrt{n}} H_{n-1}(y_i) g(y_i)\right]$$

$$= \sum_{i=2}^{l} (z_{i-1} - z_i) \frac{1}{\sqrt{n}} H_{n-1}(y_i) g(y_i)$$

for $g(y_0) = g(-\infty) = 0$ and $g(y_{l+1}) = g(\infty) = 0$.

As we now know the coefficients, we can krige the $H_n[Y(x)]$ to obtain the disjunctive kriging of $Z(x) = \Phi[Y(x)]$:

$$[Z(x)]^{DK} = \phi_0 + \phi_1 (H_1[Y(x)])^K + \phi_2 (H_2[Y(x)])^K + \ldots$$

6.2 The Relation between the Covariances

Let us consider the expansion of a function $f(Y(x))$ in terms of Hermite polynomials:

$$f[Y(x)] = f_0 + \sum_{1}^{\infty} f_n H_n(Y(x))$$

with $f_0 = E[f(Y(x))]$. Its spatial covariance is:

$$Cov\left[f(Y(x+h)),\ f(Y(x))\right] = E\left[(f[Y(x+h)]-f_0)\ (f[Y(x)]-f_0)\right]$$

$$= E\left[\left(\sum_{n\geq 1} f_n\ H_n[Y(x+h)]\right)\left(\sum_{p\geq 1} f_p\ H_p[Y(x)]\right)\right]$$

Since the bivariate distribution $(Y(x),\ Y(x+h))$ is assumed to be normal the Hermite polynomials are orthogonal and so we obtain:

$$Cov\ [f(Y(x+h)),\ f(Y(x))] = \sum_{n\geq 1} (f_n)^2\ E\ [H_n(Y(x+h))\ H_n(Y(x))]$$

$$= \sum_{n\geq 1} (f_n)^2\ \ [\varrho\ (h)]^n \qquad (6\text{–}1)$$

So there is a relationship between the covariance of the gaussian equivalent and any function of it. In particular, we get the following relation between the covariance of the gaussian equivalent obtained by anamorphosis, assumed to be *bivariate normal* and that of the raw variable $Z(x) = \Phi[Y(x)]$:

$$Cov\ [Z(x),\ Z(x+h)] = \sum_{n\geq 1} (\phi_n)^2\ \ [\varrho\ (h)]^n \qquad (6\text{–}2)$$

This relation can even be used to test the binormality of the pairs $(Y(x),\ Y(x+h))$.

Remark:

As n increases, $[\rho(h)]^n$ becomes spatially less dependent and tends to pure nugget. In particular, for $n > 1$, $[\rho(h)]^n$ is less structured than $\rho(h)$, which is the covariance of the gaussian variable. From (6–2), $Z(x)$, or more generally, from (6–1), any function of $Y(x)$, proves to be less structured than the gaussian variable.

6.3 Indicators

In many problems involving thresholds, we want to estimate whether the *unknown* value $Z(x)$ of the variable at point x is more or less than a certain cutoff z_c. In other words, we want to estimate the indicator $1_{Z(x)<z_c}$. [Note that this is *not* a question of knowing if an estimated value $Z(x)^*$ is less than the cutoff, which is represented by the indicator $1_{Z(x)^*<z_c}$].

In the gaussian model, the situation $Y(x) < y_c$ is equivalent to $Z(x) = \Phi[Y(x)] < \Phi(y_c)$. By choosing y_c so that $z_c = \Phi\ (y_c)$, that is, the cumulative frequency $G(y_c)$ equals $F(z_c)$, we have:

$$1_{Z(x)<z_c} = 1_{Y(x)<y_c}$$

We know the Hermite polynomial expansion of this (5–2):

$$G(y_c) + \sum_1^\infty \frac{1}{\sqrt{n}} H_{n-1}(y_c)\, g(y_c)\, H_n[Y(x)]$$

So the DK estimator is obtained by kriging each of the $H_n[Y(x)]$ separately (5–4):

$$[1_{Z(x)<z_c}]^{KD} = G(y_c) + \sum_1 \frac{1}{\sqrt{n}} H_{n-1}(y_c)\, g(y_c)\, (H_n[Y(x)])^K$$

As the kriged values tend to *0* rapidly as n increases, only a few terms are required, even though the coefficients $\frac{1}{\sqrt{n}} H_{n-1}(y_c)\, g(y_c)$ are not negligible.

6.4 Ore and Metal Quantities

Let $Z(x)$ represent the grade of an ore at a point x. As we have seen in 2.2, the quantities of ore and metal recovered above a cutoff z_c are, apart from a multiplicative constant, equal to:

$$1_{Z(x)\geq z_c}$$

$$Z(x)\, 1_{Z(x)\geq z_c}$$

If we let y_c be the gaussian equivalent of z_c, and $Y(x)$ be the gaussian transform of $Z(x)$, then the quantity of ore becomes:

$$1_{Y(x)\geq y_c}$$

That is:

$$1 - 1_{Y(x)<y_c} = 1 - G(y_c) - \sum_1 \frac{1}{\sqrt{n}} H_{n-1}(y_c)\, g(y_c)\, H_n[Y(x)]$$

Similarly the quantity of metal can be written as:

$$Z(x)\, 1_{Z(x)\geq z_c} = \Phi[Y(x)]\, 1_{Y(x)\geq y_c}$$

$$= \left(\sum_{p\geq 0} \phi_p H_p[Y(x)]\right) 1_{Y(x)\geq y_c}$$

This is a function of $Y(x)$. It can be expanded in terms of Hermite polynomials:

$$Z(x)\ 1_{Z(x)\geq z_c} = \sum_{n\geq 0} q_n\ H_n[Y(x)]$$

The coefficients q_n, depending on y_c, can be written as:

$$q_n = E\ [\Phi\ (Y(x))\ 1_{Y(x)\geq y_c}\ H_n\ (Y(x))]$$

$$= \int (\sum_{p\geq 0} \phi_p\ H_p\ (y))\ 1_{y\geq y_c}\ H_n\ (y)\ g(y)\ dy$$

$$= \sum_{p\geq 0} \phi_p \int_{y_c} H_p(y)\ H_n(y)\ g(y)\ dy$$

If we let $U_p^n\ (y_c)$ denote the integral in the formula above, then:

$$q_n = \sum_{p\geq 0} \phi_p\ U_p^n(y_c)$$

Using the relation:

$$U_p^n = \frac{-1}{\sqrt{n}}\ H_p\ H_{n-1}\ g + \frac{\sqrt{p}}{\sqrt{n}}\ U_{p-1}^{n-1}$$

and the fact that $U_p^n = U_n^p$ we can get a recurrence formula from $U_{|n-p|}^0\ (y_c)$:

$$U_0^0\ (y_c) = 1 - G(y_c)$$

$$U_k^0\ (y_c) = \frac{-1}{\sqrt{k}}\ H_{k-1}\ (y_c)\ g(y_c)$$

Once we know the expansions of the quantity of ore and of metal, we get their DK estimates by kriging the $H_n\ (Y(x))$ separately:

$$[1_{Z(x)\geq z_c}]^{DK} = 1 - G(y_c) - \sum_1 \frac{1}{\sqrt{n}}\ H_{n-1}\ (y_c)\ g(y_c)\ (H_n[Y(x)])^K$$

$$[Z(x)\ 1_{Z(x)\geq z_c}]^{DK} = q_0 + \sum_1 q_n\ (H_n[Y(x)])^K$$

7. DK and Conditional Expectation

In chapter 2 we have seen how DK makes it possible to estimate functions f[Y(x)] of a variable Y(x). Here we will see another estimator, more consistent and theoretically more powerful, the conditional expectation. However we are able to use this estimator only in the case of multivariate normal distributions.

7.1 The Mathematical Consistencies and Inconsistencies of DK

Disjunctive kriging is just the same as cokriging indicators. As we have seen in 2.3 this ensures a certain amount of consistency between the estimates of the different functions. In particular we have:

$$(1_{y_1 \le Y(x) < y_2})^{DK} = (1_{Y(x) < y_2})^{DK} - (1_{Y(x) < y_1})^{DK}$$

But kriging a positive valued variable does not always give positive kriged values, nor does cokriging indicators necessarily give results that lie between *0* and *1*. Sometimes the DK estimate of $1_{Y(x)<y}$ for small values of y turns out to be negative, and similarly the estimate is sometimes greater than *1* for large values of y. This is clearly a nuisance.

7.2 The Conditional Expectation in Multinormal Model

Suppose that we are estimating the indicator $1_{y_1 \le Y(x) < y_2}$ using the values $Y(x_\alpha) = (y_\alpha)$ which are available around the point x. The estimator which is theoretically the best is:

$$E\ [\ 1\ y_1 \le Y(x) < y_2 \mid Y(x_a) = y_a \ \dots\ ]$$

$$= P\ [\ y_1 \le Y(x) < y_2 \mid Y(x_a) = y_a \ \dots\ ]$$

That is, it is the conditional probability of having $y_1 \le Y(x) < y_2$ given the $Y(x_\alpha) = y_\alpha$ ... But this conditional distribution can be determined only for a multi-normal random function.

In that case, the multivariate distribution of variables like $(Y(x), Y(x_\alpha), \dots)$ is multinormal; i.e. any linear combination of these variables is also normally distributed. Consequently, the conditional distribution of $Y(x)$ given the $Y(x_\alpha) = y_\alpha$ ... is normal, with a mean value equal to its kriged estimate:

$$y(x)^K = \sum \lambda^\alpha y_a$$

and with its variance equal to the kriging variance σ_K^2. The reader will notice that this conditional multinormal distribution generalises the conditional binormal one which has been studied in exercise 2 chapter 5.8. Its pdf can be written as:

$$g(y|y_a) = \frac{1}{\sqrt{2\pi}\ \sigma_K} e^{-\left(\frac{y-\sum \lambda^a y_a}{\sigma_K}\right)^2}$$

So the estimate of $1_{y_1 \le Y(x) < y_2}$ is then equal to:

$$E\ [1_{y_1 \le Y(x) < y_2} \mid Y(x_a) = y_a \ \dots\ ]$$

$$= P\ [y_1 \le Y(x) < y_2 \mid Y(x_a) = y_a \ \dots\ ]$$

$$= \int_{y_2}^{y_1} g(y \mid y_a \ \dots)\ dy$$

Now this is a probability and so its value lies between *0* and *1*. If we put $u = \frac{y - \sum \lambda^a y_a}{\sigma_K}$, we can easily show that this is:

$$G\left(\frac{y_2 - \sum \lambda^a y_a}{\sigma_K}\right) - G\left(\frac{y_1 - \sum \lambda^a y_a}{\sigma_K}\right)$$

Similarly the estimate of any function *f[Y(x)]* can be written as:

$$E\ [\ f[Y(x)] \mid Y(x_a) = y_a \ \dots\ ]$$

$$= \int f(y)\ g(y \mid y_a \ \dots)\ dy$$

$$= \int f\left(\sum \lambda^a y_a + \sigma_K u\right) g(u)\ du \qquad [7\text{-}1\ ]$$

So if *f[Y(x)]* is a positive function this estimate will be positive. Similarly, if we are estimating two functions with one always greater than the other the difference between the two will always be positive. So the estimator (which is the conditional expectation) will always ensure that the results are consistent, and as such, it is preferable to the bivariate normal DK. However the multivariate hypothesis that is theoretically required is much more restrictive than the one on the bivariate distributions.

Moreover DK is valuable for two reasons:

1. It allows us to estimate functions even outside the normal framework (when it is no longer possible to obtain the conditional expectation).

2. As we will see in the next chapter, it is easy to estimate regularized variables using DK.

8. *Estimation of Regularized Variables*

In linear geostatistics it is possible to krige a regularized variable without kriging each of its points. This is also true for DK. The regularized of a function f[Y(x)] can be estimated directly without having to estimate each point.

8.1 Kriging

In order to krige a regularized variable:

$$Z(V) = \frac{1}{V}\int_V Z(x)\, dx$$

we do not have to krige individually each of the points inside V even though it is equivalent. We can krige directly the regularized variable. We just replace the kriging system for each point x :

$$\sum_\beta \lambda_\beta \; Cov\,(Z_\alpha, Z_\beta) = Cov\,(Z_\alpha\,,\; Z(x)) \qquad \text{for all } \alpha$$

by the system for the block V:

$$\sum_\beta \lambda_\beta \; Cov\,(Z_\alpha, Z_\beta) \quad = \quad Cov\,(Z_\alpha\,,\; Z(V)) \qquad \text{for all } \alpha$$

with:

$$Cov\,(Z_\alpha\,,\; Z(V)) = \frac{1}{V}\int_V Cov\,(Z_\alpha\,, Z(x))\, dx$$

8.2 Disjunctive Kriging

This property is very useful for estimating regularized variables using DK. This may be in a gaussian framework, as it is here, or in a different model. Suppose that we want to estimate the regularized variable:

$$\frac{1}{V}\int_V f(Y(x))\, dx$$

Now since:

$$f\,[Y(x)] = f_0 + \sum_1 f_n\, H_n\,[Y(x)]$$

we can deduce that:

$$\frac{1}{V}\int_V f(Y(x))\,dx = f_0 + \sum_1 f_n \left[\frac{1}{V}\int_V H_n\,(Y(x))\,dx\right]$$

To obtain the DK estimate we only need to krige each of the terms independently:

$$\frac{1}{V}\int_V H_n\,(Y(x))\,dx$$

So:

$$\left[\frac{1}{V}\int_V f(Y(x))\,dx\right]^{DK} = f_0 + \sum_1 f_n \left[\frac{1}{V}\int_V H_n\,(Y(x))\,dx\right]^{K}$$

The result of this would be exactly the same as the average of the DK estimates of $f(Y(x))$ at each point.

8.3 Example: DK of a proportion

Suppose we want to estimate the regularized indicator:

$$\frac{1}{V}\int_V 1_{Y(x)<y}\,dx$$

This is the average of all the indicators (which take the values *0* or *1*) inside V. As such it represents the *proportion* of points where $Y(x)$ is less than y.

From (5-2) we can see that:

$$\frac{1}{V}\int_V 1_{Y(x)<y}\,dx = G(y) + \sum_1 \frac{1}{\sqrt{n}}\,H_{n-1}\,(y)\,g(y)\left[\frac{1}{V}\int_V H_n\,(Y(x))\,dx\right]$$

The kriged estimate of $\frac{1}{V}\int_V H_n(Y(x))\,dx$ comes from the system:

$$\sum_\beta \lambda_{n\beta}\,Cov\,[H_n\,(Y_\alpha),\ H_n\,(Y_\beta)] = Cov\,[\,H_n\,(Y_\alpha)\,,\ \frac{1}{V}\int_V H_n\,(Y(x))\,dx\,]$$

i.e.:

$$\sum_\beta \lambda_{n\beta}\,Cov\,[H_n\,(Y_\alpha),\ H_n\,(Y_\beta)] = \frac{1}{V}\int_V Cov\,[\,H_n\,(Y_\alpha)\,,\ H_n\,(Y(x))\,]\,dx$$

Since the covariance of $H_n\,(Y(x))$ is $[\rho(h)]^n$, this becomes:

$$\sum_{\beta} \lambda_{n\beta} [\varrho_{\alpha\beta}]^n = \frac{1}{V} \int_V (\varrho_{\alpha x})^n \, dx$$

Hence the DK estimator of the proportion is:

$$[\frac{1}{V} \int 1_{Y(x)<y} \, dx]^{DK} = G(y) + \sum_1 \frac{1}{\sqrt{n}} H_{n-1}(y) \, g(y) [\frac{1}{V} \int_V H_n(Y(x)) \, dx]^K$$

$$= G(y) + \sum_1 \frac{1}{\sqrt{n}} H_{n-1}(y) \, g(y) \{ \sum_a \lambda_{na} H_n(Y_a) \}$$

In practice, we need (just as in chapter 5.5) to krige only a few factors, that is, to solve only a few systems.

In comparison, if we want to estimate the same proportion via the conditional expectation we must estimate the variable $1_{Y(x)<y}$ at a large enough number of points for it to be representative of the integral over V. So we have to solve the kriging system for each of these points (but the matrix on the left is the same and only has to be inverted once).

8.4 Important comment

In the gaussian anamorphosed model we can now estimate the indicator

$$1_{Z(x)<z} = 1_{Y(x)<y}$$

where $z = \phi(y)$ either using DK or conditional expectation. Similarly we can also estimate the proportion of points in V where $Z(x) < z$:

$$\frac{1}{V} \int_V 1_{Z(x)<z} \, dx = \frac{1}{V} \int_V 1_{Y(x)<y} \, dx$$

However we cannot estimate indicators on a regularised support such as:

$$1_{Z(V)<z}$$

For this, a **change of support** model is required (see the second part of the book).

9. The Problems of Stationarity

In linear geostatistics the condition $\sum \lambda_\alpha = 1$ *is normally used when kriging a variable. This makes the local estimation more robust against a lack of stationarity.*

Using such conditions in non linear punctual estimation requires caution. Only in the mosaic model are there no problems. In DK estimation for example, these conditions lead to problems of convergence that must be looked at closely. In the case of the conditional expectation, the condition $\sum \lambda_\alpha = 1$ *is justified only when estimating a lognormal variable (lognormal kriging).*

9.1 Kriging

Let us recapitulate briefly on linear kriging. For simple kriging (i.e. kriging with a known mean) we have:

$$Z(x)^{SK} - E\,[Z(x)] = \sum_\alpha \lambda_\alpha\,[Z_\alpha - E\,(Z_\alpha)]$$

As the expectations $E[Z(x)]$ and $E[Z_\alpha]$ are known and are equal to each other ($= m$), we have:

$$Z(x)^{SK} = \sum_\alpha \lambda_\alpha Z_\alpha + [\,1 - \sum \lambda_\alpha\,]\,m$$

This shows that SK gives a weighting factor $\lambda_m = 1 - \sum \lambda_\alpha$ to the known mean m. When the samples are fairly close (relative to the variogram structure) this weight is small. It increases as the information becomes more sparse, and the mean tends to counterbalance the lack of information: its role is more important in areas where there is less information. This is just the consequence of the stationarity hypothesis which gives a local meaning to the mean m. In mining this causes problems in areas that are under sampled because they are poorer than average and hence less economically profitable. (This should make us reconsider the stationarity hypothesis). Even in cases where the sampling is regularly spaced, the hypothesis of stationarity which attributes a local significance to the mean value that has been calculated over the whole deposit, may seem too strong.

We then use kriging with an unknown mean. (This is called ordinary kriging). In this case the sum of the kriging weights of the neighbouring data is 100%:

$$\sum \lambda_\alpha = 1$$

The kriging estimator is $Z(x)^{OK} = \sum \lambda_\alpha Z_\alpha$. It is unbiased whatever the value of the unknown mean $m = E\,(Z(x)) = E\,[Z_\alpha]$. So:

$$E[Z(x) - Z(x)^{OK}] = E(Z(x) - \sum \lambda_a Z_a) = 0$$

In other words, we assume that the mean can take any value, and that this may vary from one area to another but must be (approximately) constant over areas the size of the kriging neighbourhood. This excludes systematic trend or drift at this scale.

9.2 Cokriging

What is true for kriging single variables is also true for cokriging. For example, if the mean is known, we have:

$$[\ Z_1(x) - E(Z_1(x))]^{CK} = \sum_{\alpha} \lambda_{1\alpha}\ [Z_{1\alpha} - E(Z_{1\alpha})] + \sum_{\beta} \lambda_{2\beta}\ [Z_{2\beta} - E(Z_{2\beta})]$$

So if $E(Z_1(x)) = E(Z_{1\alpha}) = m_1$ and similarly $E(Z_{2\beta}) = m_2$, then:

$$Z_1(x)^{CK} = \sum_{\alpha} \lambda_{1\alpha}\ Z_{1\alpha} + \sum_{\beta} \lambda_{2\beta}\ Z_{2\beta} + (1 - \sum_{\alpha} \lambda_{1\alpha})\ m_1 - (\sum_{\beta} \lambda_{2\beta})\ m_2$$

Here α describes the data on the variable Z_1, and β those on Z_2, which need not to be located at the same points.

If we instead assume that the means are not known then we use the non bias conditions:

$$\sum_{\alpha} \lambda_{1\alpha} = 1 \qquad\qquad \sum_{\beta} \lambda_{2\beta} = 0$$

Note that the condition $\sum_{\alpha} \lambda_{1\alpha} = 1$ requires at least one value of Z_1 inside the kriging neighbourhood.

9.3 Disjunctive kriging

Let us now consider the cases of indicators and of disjunctive kriging. In the mosaic model imposing the condition $\sum \lambda_a = 1$ on the single system guarantees the unbiasedness of the estimator for all the indicators or for any function of the variable such as $f(Y(x))$:

$$E[\ f(Y(x))^*] = \sum \lambda^a\ E[\ f(Y_a)] = E[\ f\ (Y(x))]$$

since the means, $E[f(Y(x)]$ and $E\ [f(Y_\alpha)]$, are locally the same.

In the isofactorial models any indicator or any function of the variable can be expressed in terms of the factors:

$$f[Y(x)] = a_0 + \sum_1 a_n H_n (Y(x))$$

When we impose the condition $\sum_a \lambda_{na} = 1$ on the kriging system of each of the factors:

$$[H_n (Y(x))]^K = \sum \lambda_{na} H_n (Y_a)$$

we have:

$$E [H_n (Y(x))]^K = \sum_a \lambda_{na} E [H_n (Y_a)]$$

$$= E [H_n (Y(x))]$$

and then:

$$E [f(Y(x))^*] = a_0 + \sum_1 a_n E [H_n (Y(x))]^K$$

$$= a_0 + \sum_1 a_n E[H_n (Y(x))]$$

$$= E [f(Y(x))]$$

So the estimator is unbiased whatever the local means of the factors. If, however, the means of the factors are (locally) unknown then so are those of the indicators and so is the (local) distribution of the variable. Taking gaussian DK as an example, this means that the "gaussian" equivalent $Y(x)$ does not necessarily have (locally) a mean of zero and a unit variance, and does not even have (locally) a normal distribution. The factors (here the Hermite polynomials) are not necessarily orthogonal, which means that the optimality of the estimator is no longer guaranteed.

Moreover the unbiasedness conditions for gaussian DK have to be used carefully. Any function $f[Y(x)]$ can be expressed in terms of (an infinite number of) Hermite polynomials:

$$f [Y(x)] = f_0 + \sum_{n=1}^{\infty} f_n H_n [Y(x)]$$

As the structure of the Hermite polynomials tends rapidly to pure nugget with increasing n, kriging the $H_n (Y(x))$ at any unknown point *without* any unbiasedness condition soon returns to the mean value (i.e. zero) say when $n = L+1$. This is why so few polynomials (L) actually need be kriged.

The situation is different when an unbiasedness condition is introduced into the kriging system. In this case, for $n > L$, we have:

$$[\ H_n\ (Y(x))]^K\ =\ \frac{1}{N}\ \sum_{\alpha}\ H_n\ (Y_\alpha)$$

where N is the number of sampling points x_α in the kriging neighbourhood, and there is no reason for these quantities to be zero. But we have:

$$\sum_{n=L+1}^{\infty}\ f_n\ H_n\ (Y_\alpha)\ =\ f(Y_\alpha) - f_0 - \sum_{n=1}^{L}\ f_n\ H_n\ (Y_\alpha)$$

So the estimator of $f(Y(x))$ which has an infinity of terms:

$$f(Y(x))^*\ =\ f_0 + \sum_{n=1}^{L}\ f_n \sum_{\alpha}\ \lambda_{n\alpha}\ H_n\ (Y_\alpha) + \frac{1}{N}\ \sum_{\alpha}\ (\ \sum_{n=L+1}^{\infty}\ f_n\ H_n\ (Y_\alpha))$$

becomes:

$$f_0 + \sum_{n=1}^{L}\ f_n \sum_{\alpha}\ \lambda_{n\alpha}\ H_n\ (Y_\alpha) + \frac{1}{N}\ \sum_{\alpha}\ [f(Y_\alpha) - f_0 - \sum_{n=1}^{L}\ f_n\ H_n\ (Y_\alpha)]$$

or more simply:

$$f(Y(x))^* = \frac{1}{N}\ \sum_{\alpha}\ f(Y_\alpha) + \sum_{n=1}^{L}\ f_n\ \sum_{\alpha}(\lambda_{n\alpha} - \frac{1}{N})\ H_n\ (Y_\alpha)$$

There is now only a finite number of terms. As:

$$\sum_{\alpha}\ (\lambda_{n\alpha} - \frac{1}{N}) = 1 - 1 = 0$$

we find again as expected:

$$E[\ f(Y(x))^*] = \frac{1}{N} \sum_{\alpha}\ E[\ f(Y_\alpha)] = E[\ f\ (Y(x))]$$

This was for a punctual estimation. It is also possible to introduce unbiasedness conditions in the DK estimation of regularised variables. But this will *not* be possible when dealing with a change of support (see 12.1).

9.4 Lognormal kriging

Adding unbiasedness conditions to remove local bias *cannot* be done in general when using the conditional expectation. The condition $\sum\ \lambda_\alpha\ =\ 1$ in the kriging system for the

gaussian equivalents would not ensure the removal of local bias when estimating a nonlinear function. A notable exception is the case where the function f is an exponential, which corresponds to estimating a lognormal variable. As a matter of fact, if $Z(x) = f(Y(x))$ is a lognormal variable with log mean μ and log variance σ^2, then:

$$Log\ Z(x) = Log\ f[Y(x)] = \mu + \sigma\ Y(x)$$

and we can write:

$$Z(x) = f[Y(x)] = e^{\mu + \sigma Y(x)} = e^{\mu}\ e^{\sigma Y(x)}$$

Its means is (see exercise 3 in chapter 5-8):

$$E\ [Z(x)] = E\ [\ f[Y(x)]\ ] = e^{\mu}\ E\ [e^{\sigma Y(x)}] = e^{\mu + \frac{\sigma^2}{2}}$$

We can estimate $Z(x) = f[Y(x)]$ by integrating (7-1), which gives:

$$e^{\mu + \sigma \sum \lambda^{\alpha} Y_{\alpha} + \frac{\sigma^2 \sigma_K^2}{2}}$$

or, knowing $Log\ Z_a = \mu + \sigma\ Y_{\alpha}$:

$$e^{(1-\sum \lambda^{\alpha})\mu + \sum \lambda^{\alpha} Log Z_{\alpha} + \frac{\sigma^2 \sigma^2 K}{2}}$$

$$= e^{[Log\ Z(x)]^{KS} + \frac{\sigma^2 \sigma^2 K}{2}}$$

The term $\sigma^2 \sigma^2_k$ (which must not be forgotten) is the kriging variance when estimating $Log\ Z(x)$. If γ denotes the variogram of the logarithms then:

$$\sigma^2\ \sigma_K^2 = \sigma^2 - \sum \lambda^{\alpha} \lambda^{\beta} (\sigma^2 - \gamma_{\alpha\beta}) = [1 - (\sum \lambda^{\alpha})^2]\sigma^2 + \sum \lambda^{\alpha} \lambda^{\beta} \gamma_{\alpha\beta}$$

So the estimator becomes:

$$e^{(1-\sum \lambda^{\alpha})\mu + \sum \lambda^{\alpha} Log\ Z_{\alpha} + (1-(\sum \lambda^{\alpha})^2)\frac{\sigma^2}{2} + \frac{1}{2} \sum \lambda^{\alpha} \lambda^{\beta} \gamma_{\alpha\beta}}$$

Now if we impose the condition $\sum \lambda_{\alpha} = 1$, we obtain:

$$Z(x)^* = e^{\sum \lambda^{\alpha} Log Z_{\alpha} + \frac{1}{2} \sum \lambda^{\alpha} \lambda^{\beta} \gamma_{\alpha\beta}}$$

This estimator is unbiased whatever the local mean and variance, say μ' and σ'^2, of the logarithm, and hence whatever the local mean of the raw variable:

$$E[Z(x)^*] = e^{\mu' + \frac{1}{2} \sum \lambda^{\alpha} \lambda^{\beta} (\sigma'^2 - \gamma_{\alpha\beta}) + \frac{1}{2} \sum \lambda^{\alpha} \lambda^{\beta} \gamma_{\alpha\beta}}$$

$$= e^{\mu' + \frac{\sigma'^2}{2}} = E[e^{\log Z(x)}] = E[Z(x)]$$

Comments:

Whether there is or not a condition on the λ_{α} , these estimators are obtained with the kriging (or another linear combination) of the logarithms. Be careful however! Just

taking the expectation and forgetting the other terms is not enough and would introduce a dangerous bias. The term *lognormal kriging* often covers such estimators.

PART TWO: THE CHANGE OF SUPPORT

10. Change of support in Global Estimation

Every geostatistician knows that the distribution of a variable, and in particular its variance, depends on the support on which the variable is defined. Often the data are values measured on a quasi–punctual support, whereas what we want to estimate is defined on a larger support. Hence the importance of models that can take the change of support into account.

We will see here a commonly used and very powerful model: the discrete gaussian model. We will consider the case of estimating the global distribution first (the present chapter), then (next chapter) the case of local estimation.

10.1 Introduction

In the first half of this book we have seen how to estimate a function of a punctual variable, given sample information at several other points. For example, we saw how to estimate an indicator:

$$1_{Z(x)<z}$$

Its mean value over the region represents the distribution function of the variable $Z(x)$:

$$F(z) = P[\, Z(x) < z \,] = E[\, 1_{Z(x)<z} \,]$$

To be precise, the *point* corresponds here to the support of the samples, and is usually very small, e.g. a drill core on which the metal grade has been measured. The grades measured on a small support like this can be much richer or poorer than those measured on larger supports v, e.g. mining blocks, which have more intermediate values. See Figure 10–1. Of course, the average grade is the same whether we consider small supports or large ones. The difference in the distribution between large and small supports is called the *support effect* by geostatisticians.

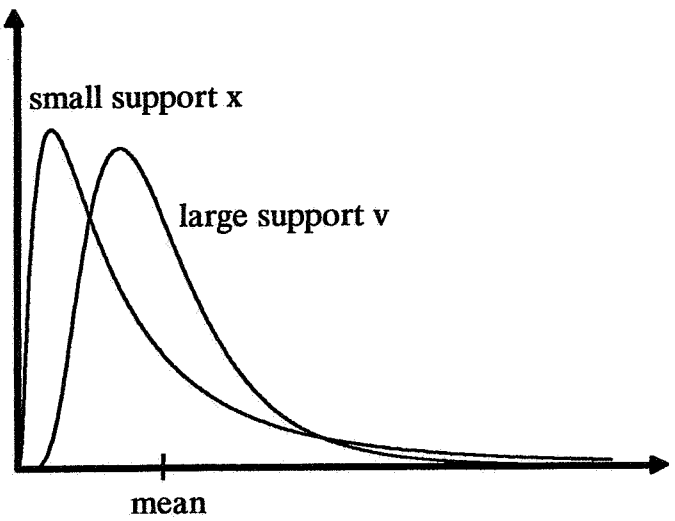


Figure 10–1 : distribution of values.

The cdf also represent the means of the indicators $1_{Z(x)<z}$ and $1_{Z(v)<z}$:

$$F(z) = E[\,1_{Z(x)<z}\,] = P[\,Z(x) < z\,]$$

$$F_v(z) = E[1_{Z(v)<z}\,] = P[\,Z(v) < z\,]$$

Their difference (fig. 10–2) shows that estimating $1_{Z(x)<z}$ for a point support is not equivalent to estimating $1_{Z(v)<z}$ for a larger block. A *change of support model* is needed to estimate the distribution of blocks given that of points.

In this work we assume that the histogram of our data (weighted, if necessary, to remove the effect of clustered data locations) is representative of the distribution of point values. But sometimes we cannot, particularly if there are too few samples or if their values are very variable. In these cases even the mean value over the region is not well known, and so it is better to compute the global estimation variance of this mean (linear geostatistics) or even to look for confidence intervals for the unknown mean value. In mining, this difficult problem has been studied by Sichel for independent and lognormally distributed values.

The aim of the second part of this book is the change of support. This is vital for predicting recoverable reserves in mining, but it also plays a very important role in other fields such as environmental sciences where we have to estimate indicator functions or probabilities of exceeding thresholds.

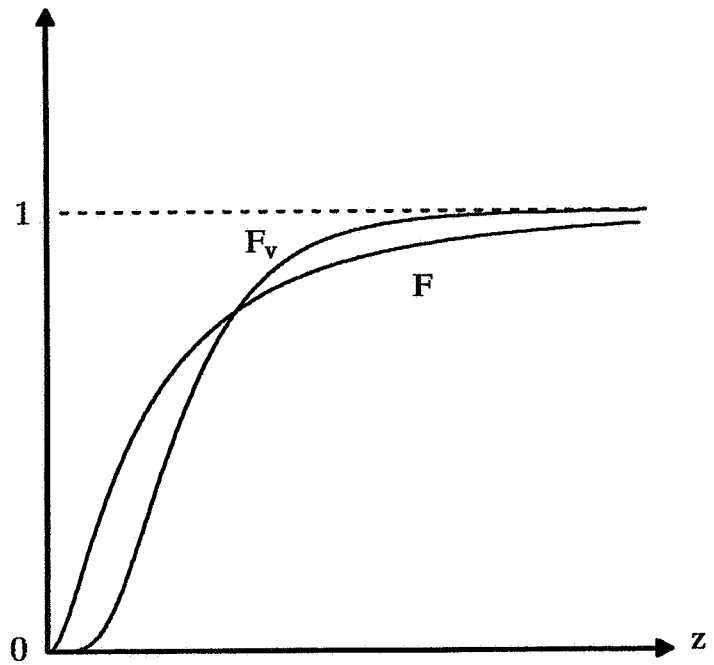


Figure 10–2 : cumulative distribution functions.

10.2 Recoverable reserves

Let us now consider the miner's task of estimating recoverable reserves. Suppose that a block v has a grade $Z(v)$. Its recoverable reserves at cut-off z are:

– the ore tonnage, given apart from a multiplicative factor by the indicator $1_{Z(v)\geq z}$

– the quantity of metal, given by $Z(v)\ 1_{Z(v)\geq z}$

For convenience we define the Conventional Profit Function[1], which is:

$$Z(v)\ 1_{Z(v)\geq z} - z\ 1_{Z(v)\geq z} = (\ Z(v) - z\ )\ 1_{Z(v)\geq z}$$

This is equal to $(Z(v) - z)$ if $Z(v)$ exceeds the cutoff z and is 0 otherwise.

Before mining commences we have data only from the exploration drillholes. The grades $Z(v)$ of the blocks are not known. The recoverable reserves are precisely what we want to estimate in order to predict the results of selecting blocks above a certain cutoff. (We assume that selection is free; i.e. blocks will be selected individually as ore or waste independently of others).

Suppose we have a 50 x 50 grid of sampling data. It would be pointless to estimate the reserves of each 5 x 5 selection block individually. When estimating the **local** reserves we should rather know the reserves contained in the $N=100$ blocks v_i inside a 50 x 50 panel. (Fig. 10–3). To be more precise we want to know:

– the ore tonnage of the panel (given by the number of blocks over the cutoff grade):

$$\sum 1_{Z(v_i)\geq z}$$

or equivalently, the proportion of these blocks in the panel:

$$\frac{1}{N} \sum 1_{Z(v_i)\geq z}$$

– the metal tonnage:

$$\frac{1}{N} \sum Z(v_i)\ 1_{Z(v_i)\geq z}$$

[1] In effect the reader will show the following result. Imagine a mining exploitation where each of the blocks v (with known grade) can be selected independently from the others, where the value of a block is proportional to its metal content, and where the extraction and treatment costs, proportional to the tonnage, are the same for each block. With these conventions the profit is maximized by selecting each of the blocks if its metal pays these costs, i.e. if its grade exceeds a given cut-off z. For each of the blocks the profit is equal, apart from a constant factor, to our "conventional profit".

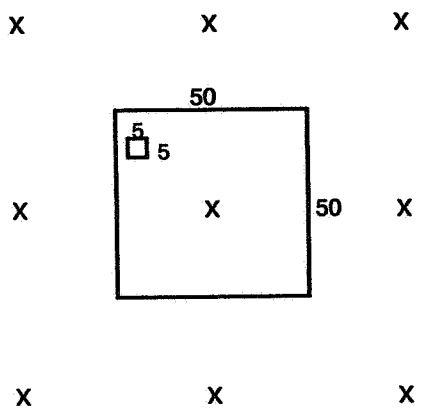


Figure 10-3

For the **global** estimation of recoverable reserves our objective is to estimate these variables for all the blocks in the deposit that is:

- ore tonnage: $T(z) = E[\, 1_{Z(v)\geq z} \,] = P[\, Z(v) \geq z \,]$

- metal tonnage: $Q(z) = E[\, Z(v)\, 1_{Z(v)\geq z} \,]$

- Conventional Profit: $B(z) = E[\, (Z(v) - z)\, 1_{Z(v)\geq z} \,] = Q(z) - z\, T(z)$

- Average Recovered grade: $m(z) = \dfrac{Q(z)}{T(z)} = E[\, Z(v) \mid Z(v) \geq z \,]$

Clearly if we know the distribution of block grades $Z(v)$ we can calculate all of these.

_Remark: information effect and conditional bias

The recoverable reserves as they have been defined above assume that the block grades $Z(v)$ are known (exactly) when the selection is being made. So they take account of the support effect.

It is true that the ultimate information available when the selection is made will be much more detailed than at the exploration stage, but the block grade $Z(v)$ will still not be known exactly. So the selection is based on estimated values, and consequently it will not be as good as if the grades were known. The estimates of recoverable reserves should also take account of this *information effect* as it is called.

Suppose that the ultimate estimator $Z(v)^{ult}$ is *conditionally unbiased* (which is practically the case with kriging if its neighbourhood is large enough); that is,

$$E[\, Z(v) \mid Z(v)^{ult} \,] = Z(v)^{ult}$$

The average recovered grade:

$$E\ [\ Z(v)\ |\ Z(v)^{ult}\ \geq z\ ]$$

will equal the predicted value:

$$E\ [\ Z(v)^{ult}\ |\ Z(v)^{ult}\ \geq z\ ]$$

We then need to know the distribution of the final estimated grades $Z(v)^{ult}$ (instead of the real block grades $Z(v)$) in order to estimate the reserves.

10.3 Selectivity Curves

In mining the distribution of grades is often represented by the corresponding "grade-tonnage" curves. For example mine surveyors plot:

- the ore tonnage T as a function of the cutoff grade z
- the metal tonnage Q as a function of the cutoff grade z
- the average grade m as a function of the cutoff grade z
- the conventional profit B as a function of the cutoff grade z
- the average grade m as a function of the ore tonnage T
- the metal tonnage Q as a function of the ore tonnage T

The functions $T(z)$, $Q(z)$, $B(z)$ and $m(T)$ are always decreasing functions whereas $m(z)$ and $Q(T)$ are increasing functions. Two of these functions in particular (the metal against ore tonnage and the conventional profit against cutoff) are valuable when one wants to compare the selectivity for different sized supports (Matheron 1981).

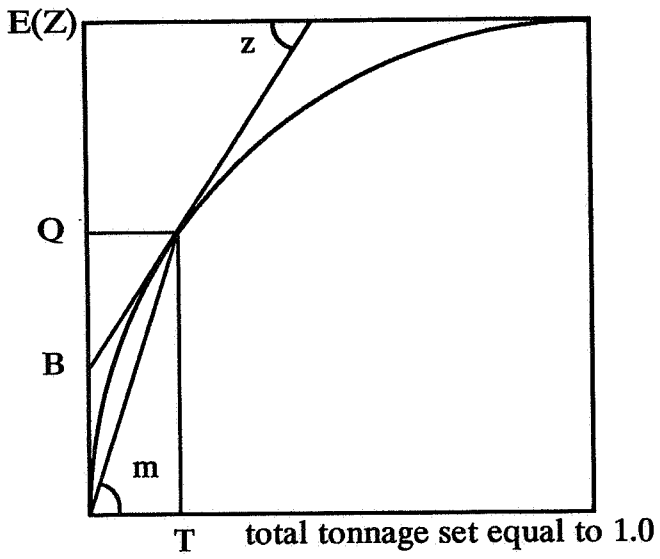


Figure 10–4

The graph of metal against ore tonnage is always convex from below. If we draw a line from any point *(T, Q)* on this curve to the origin, its slope is $Q/T = m$. The slope of the tangent at the point *(T, Q)* is equal to the cutoff z , and the tangent cuts the Q axis at the value B.

Now if we plot the metal against ore tonnage for two different supports, the one for the small support always lies above the one for the larger support for the same ore tonnage. Figure 10.5 shows this.

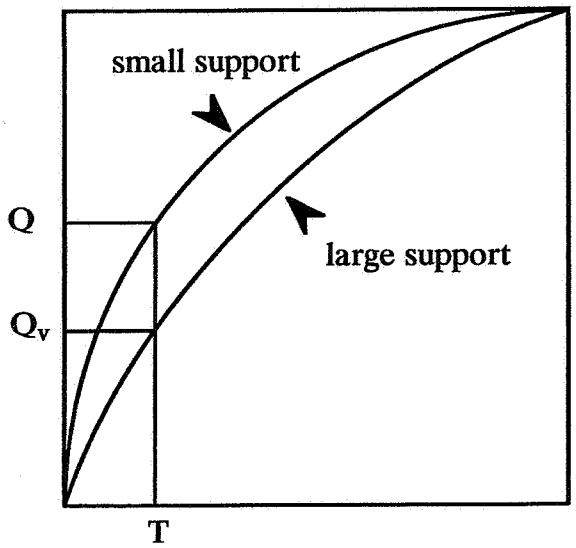


Figure 10–5

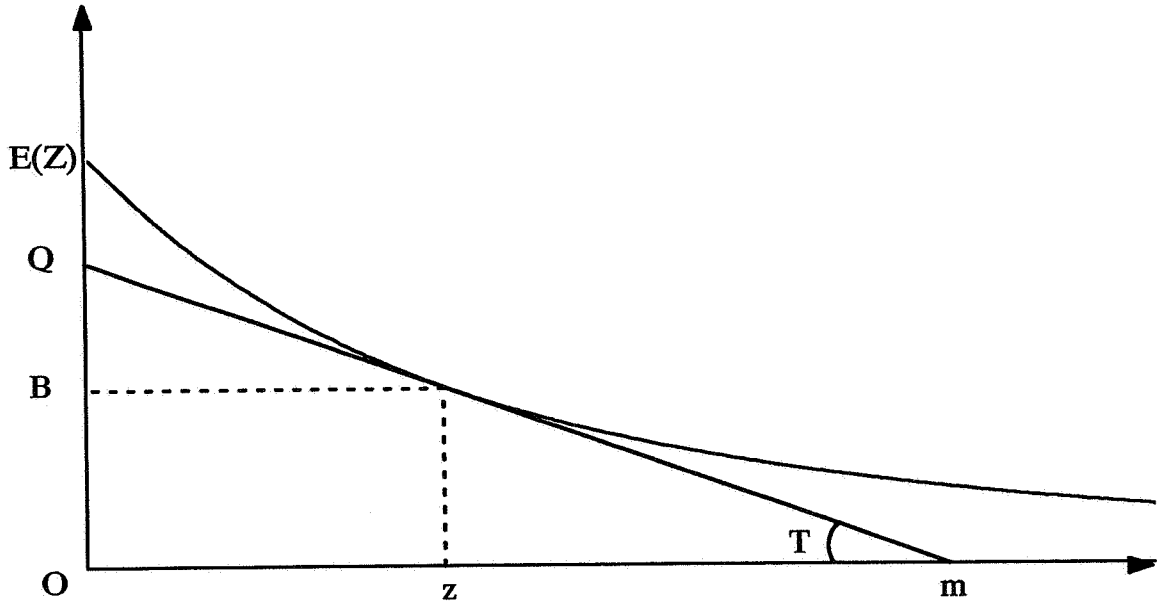


Figure 10–6

The other profitable curve is the one for conventional profit against cutoff. This is always convex. The tangent at the point *(B, z)* cuts the two axes at the points *(O, Q)* and *(m = Q/T, O)*, and its slope is – *T*.

If we plot these curves for several supports we see that for a given cutoff grade the conventional profit decreases as the size of support increases.

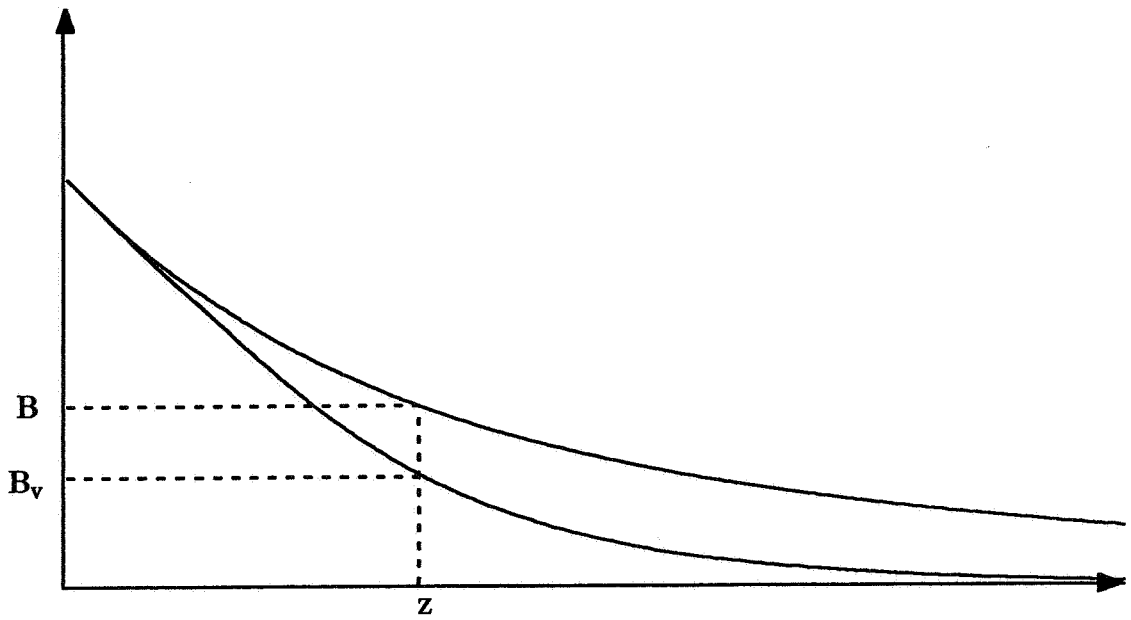


Figure 10–7

Plotting the sets of curves of $Q(T)$ and of $B(z)$ is very valuable for assessing and comparing the selectivity at different support sizes. Clearly, the change of support models used to estimate the distribution of block grades from that of the samples must respect the ordered relations among the curves.

10.4 Cartier's Relation

We now want to work out how to obtain the distribution of the blocks $Z(v)$ from that of the grades $Z(x)$ (which we know). We need a model to do this, and it has to satisfy certain conditions.

– First, the mean values must be the same:

$$E\ [\ Z(x)\ ]\ =\ E\ [\ Z(v)\ ]$$

– Second, from linear geostatistics (Krige's relation) we can calculate the variance of $Z(v)$ from the variogram model $\gamma(h)$ of $Z(x)$:

$$Var\ Z(v)\ =\ Var\ Z(x)\ -\ \bar{\gamma}\ (v,v)$$

The term $\bar{\gamma}\ (v,v)$ represents the dispersion variance of points inside a block v. It is the mean value of $\gamma(x-y)$ where x and y describe the block independently:

$$\bar{\gamma}\ (v,v)\ =\ \frac{1}{v^2}\int\int \gamma\ (x-y)\ dx\ dy$$

If there is a large nugget effect or a large short scale structure $\bar{\gamma}\ (v,v)$ will be large, and so the support effect is marked.

– Third, there is Cartier's relation:

$$E\ [\ Z(\underline{x})\ |\ Z(v)\ ]\ =\ Z(v)$$

where $\underline{x}$ denotes a point located at random inside the block v. This condition means that on average the expected value of the grade of a point chosen at random given the block grade is just equal to the block grade. It is important because we can show (cf Matheron 1981) that it guarantees the order relationship seen in the metal against ore tonnage curve and in the conventional profit against cutoff grade curve.

10.5 The Discrete Gaussian Model (1):

Preliminary remark: In the discrete gaussian model, the field is considered as the union of blocks, and the samples are randomly located within their block. In the narrow sense this model can be used for a change of support for the global distribution. In the wider sense (see 11.2) it is a more complete model that can be used for the local estimation of distributions.

In the gaussian case, the punctual variable $Z(x)$ can be expressed as a function of a standard normal variable $Y(x)$:

$$Z(x) = \Phi\ (Y(x))$$

The function Φ is the gaussian anamorphosis function that we described in chapters 5 and 6. It is known experimentally because we know the distribution of $Z(x)$.

The variable $Z(v)$ can also be expressed as a function $\Phi_{(v)}$ of another standard normal variate Y_v :

$$Z(v) = \Phi_{(v)}\ (Y_v)$$

We now will determine $\Phi_{(v)}$, which is equivalent to determine the distribution of $Z(v)$. We can rewrite Cartier's relation:

$$E\ [\Phi\ (Y(\underline{x}))\ |\ \Phi_{(v)}\ (Y_v)\ ] = \Phi_{(v)}\ (Y_v)$$

or:

$$\Phi_{(v)}\ (Y_v) = E\ [\Phi(Y(\underline{x}))\ |\ Y_v\ ]$$

To evaluate this we make one more hypothesis. We assume that the pair $(Y_v, Y(\underline{x}))$ has a bivariate normal distribution with coefficient of correlation r . Since Y_v and $Y(\underline{x})$ are standard normal variates the conditional distribution of $Y(\underline{x})$ given $Y_v = y_v$ is normal with mean ry_v and variance $1 - r^2$ (see relation 5–7):

$$Y(\underline{x}) = ry_v + \sqrt{1-r^2}\,U$$

This gives us the anamorphosis function $\Phi_{(v)}$ for blocks:

$$\Phi_{(v)}\ (Y_v) = \int \Phi\ (rY_v + \sqrt{1-r^2}\ u\ )\ g(u)\ du \qquad [10-1]$$

where $g(u)$ is the density function of the standard normal distribution.

Remark: This change of support formula conserves lognormality: if $Z(x)$ is lognormal so is $Z(v)$.

Proof: From exercise 3–9 in chapter 5.8, we know that a lognormal variable $Z(x)$ can be written as:

$$Z(x) = E[Z]\ e^{\sigma Y(x) - \frac{\sigma^2}{2}}$$

Using formula 10–1, we obtain:

$$Z(v) = E[Z]\ e^{r\sigma\ Y_v - \frac{r^2\sigma^2}{2}}$$

This is also a lognormal variable with same mean $E[Z]$ but its logarithmic variance $r^2\,\sigma^2$ is smaller than the logarithmic variance σ^2 of $Z(x)$.

In practice it is convenient to use Hermite polynomials. As Y_v and $Y(\underline{x})$ are standard normal variates, a relation similar to 5–3 gives:

$$E\ [H_n\ (Y(\underline{x}))\ |\ Y_v] \ = \ r^n\ H_n\ (Y_v)$$

So:

$$\Phi_{(v)}\ (Y_v) \ = \ E\ [\ \Phi\ (Y(\underline{x}))\ |\ Y_v\ ]$$

$$= \ E\ [\ \sum_0\ \phi_n\ \ H_n\ (Y(\underline{x}))\ |\ Y_v\ ]$$

$$= \ \sum_0\ \phi_n\ E\ [\ H_n\ (Y(\underline{x}))\ |\ Y_v\ ]$$

$$= \ \sum_0\ \phi_n\ r^n\ H_n\ (Y_v)$$

So we obtain the anamorphosis coefficients for blocks by multiplying the corresponding punctual coefficients ϕ_n by r^n. We note that the lowest order term ϕ_0 is the same since:

$$\phi_0 \ = \ E\ [\ Z(x)\ ] \ = \ E\ [\ Z(v)\ ]$$

As there is no correlation between the different polynomials, we obtain (see exercise in 5–4):

$$Var\ Z(v) \ = \ Var\ \Phi_{(v)}\ (Y_v) \ = \ \sum_1\ (\phi_n)^2\ r^{2n}$$

Since $Var\ Z(v) \ = \ Var\ Z(x) - \overline{\gamma}\ (v, v)$, we can compute the value of the change of support coefficient r which corresponds to the known variance of blocks (see figure 10–8). Note $0 \leq r \leq 1$.

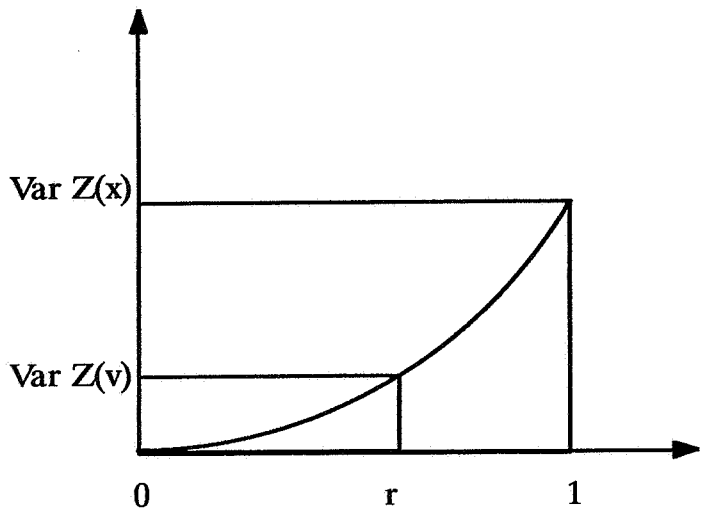


Figure 10–8

Now that r is known, so is the block anamorphosis function ϕ_n and hence the block distribution. Since Y_v is a standard normal variate (density g and cumulative distribution G), we can easily estimate any function of $Z(v)$, or equivalently any function of Y_v. As:

$$Z(v) = \Phi_{(v)} (Y_v) \geq z \quad \Leftrightarrow \quad Y_v \geq y$$

with $z = \Phi_{(v)} (y)$, we see that for example:

$$F_v(z) = E [1_{Z(v)<z}] = P (Z(v) < z)$$

$$= E [1_{Y_v<y}] = P (Y_v < y) = G(y)$$

So we can estimate the global recoverable reserves of a deposit when selecting blocks at cut–off z. The ore tonnage and quantity of metal are given by:

$$T_v(z) = E [1_{Z(v)\geq z}] = 1 - G(y)$$

$$Q_v(z) = E [Z(v) 1_{Z(v)\geq z}]$$

$$= E [\Phi_{(v)} (Y_v) 1_{Y_v\geq y}]$$

$$= \int_y \Phi_{(v)} (y) g(y) dy$$

We can calculate $Q_v(z)$ using Hermite polynomials:

$$Q_v (z) = \int_y (\sum_0 \phi_n r^n H_n (y)) g(y) dy$$

$$= \sum_0 \phi_n r^n \int_y H_n (y) g(y) dy$$

As in 5–2, we can show that this is equal to:

$$\phi_0 [1 - G(y)] - \sum_1 \phi_n r^n \frac{1}{\sqrt{n}} H_{n-1} (y) g(y)$$

10.6 Exercises :

Here are some simple exercises. In these $Z(x)$ is a stationary variable with mean m and with covariance $C(h)$. So its variance is $C(0)$, and its variogram is:

$$\gamma(h) = C(0) - C(h).$$

Exercise 1 :

Let $Z(v)$ be the average for a block v :

$$Z(v) = \frac{1}{v} \int_v Z(x)\, dx$$

Let x_0 be a fixed point inside v. Its value is $Z(x_0)$.

a) Suppose that we know $Z(x_0)$ and want to estimate $Z(v)$ by kriging:

- First, by ordinary kriging (unknown mean).
- Second, by simple kriging (known mean).

Give the estimator in both cases and calculate the kriging weight of the mean m.

b) Now suppose we know $Z(v)$ and want to estimate $Z(x_0)$ as before using ordinary kriging and simple kriging. Give the two estimators and the kriging weight of the mean.

c) This time, repeat the first question (a) for a point $\underline{x}$ located at random inside v.

d) Repeat the second question (b) for a point $\underline{x}$ located at random inside v.

Show that when a sample $Z(\underline{x})$ is taken at random inside the block v the kriged estimator of its value, given the block grade, is equal to the block grade.

Does this result satisfy with Cartier's relation ?

$$E\ [\ Z(\underline{x})\ |\ Z(v)\ ]\ =\ Z(v)$$

Exercise 2 :

Now we use the gaussian change of support model. As before v denotes the block and $\underline{x}$ denotes a point taken at random inside v.

$$Z(v) = \Phi_{(v)}\ [\ Y_v\ ]$$

$$Z(\underline{x}) = \Phi\ [\ Y(\underline{x})\ ]$$

The pair $(Y_v, Y(\underline{x}))$ is bivariate normal with a correlation coefficient equal to the change of support coefficient r. By construction this model satisfies Cartier's relation.

a) Suppose that we know the value of $Z(\underline{x})$. Express the expected value of $Z(v)$ given $Z(\underline{x})$ in terms of Hermite polynomials:

$$E\ [\ Z(v)\ |\ Z(\underline{x})\ ]$$

☞ Hint : As Y_v and $Y(\underline{x})$ are standard bivariate normal with correlation coefficient r, we know that:

$$E\ [\ H_n\ (Y_v)\ |\ Y(\underline{x})\ ]\ =\ r^n\ H_n\ [Y(\underline{x})\ ]$$

b) As in (a) we know $Z(\underline{x})$. We want to estimate $1_{Z(v)<z}$ given $Z(\underline{x})$ using its expected value:

$$E\ [\ 1_{Z(v)<z}\ |\ Z(\underline{x})\ ]\ =\ P\ [Z(v) < z\ |\ Z(\underline{x})\ ]$$

Show that this can be expressed simply in terms of the cumulative normal distribution $G(y)$.

☞ Hint : Since Y_v and $Y(\underline{x})$ are standard bivariate normal with correlation coefficient r, the distribution of Y_v knowing $Y(\underline{x}) = y(x)$ is normal with mean ry_x and variance $1-r^2$. So we can write it as:

$$Y_v\ =\ ry(x)\ +\ \sqrt{1-r^2}\ U$$

where U is a standard normal variate.

c) Now suppose we know $Z(v)$. Estimate $1_{Z(\underline{x})<z}$.

11. Local Estimation

In the previous chapter we have seen how the global estimation of distributions can be performed using the gaussian discrete model. We will now see how this model makes the local estimation possible, whether by DK or by conditional expectation.

11.1 A Point - Block Model:

Suppose that we have a certain number of point samples $Z(x_\alpha)$ and that we want to estimate a function of the average grade of a particular block $Z(v)$. For example the function could be an indicator:

$$1_{Z(v)<z}$$

In the first half of this book, we were looking for models that linked the values observed at different points (i.e. a point - point model). We now need a model linking the point values with those of blocks. This problem is theoretically very complex, and so we need to make some simplifying assumptions. The procedure used is described below for the gaussian case but it can be extended to others (e.g. gamma, ...).

a) The region is divided into blocks v_i whose size is equal to the support v. For example, in mining these blocks are the mining selection units.

> Remark: Here, the block size and the position of the blocks are fixed (or nearly fixed). A high grade area of limited size that would be found during mining may therefore be included in a single block, or split into two or more blocks. In reality it is also possible for the geometry of mining blocks to differ from block to block, and to depend on the location of the ore found during mining. This case is called "adaptative" geometry. There are no models for predicting directly recoverable reserves in this case, so a simulation must be used.

b) We assume that the position of a sample inside a block is random, and consequently "forget" its exact position.

c) We assume that given the grade $Z(v_i)$ of the block v_i containing a particular sample x_i, the sample grade $Z(x_i)$ is independent of all the other grades, blocks or points.

Now we are going to see what can be said about the different covariances. To start with our point support variable $Z(x)$ has a variogram $\gamma(h)$ (or a covariance $C(h)$) that is assumed to be known. From linear geostatistics, we know that the covariance between two blocks v_i and v_j, denoted by $C(v_i, v_j)$, is the mean value of $C(h)$ when one end of the vector h sweeps over v_i while the other end independently sweeps over v_j. So we can evaluate the covariance between any two blocks.

Similarly we can calculate the covariance between a point x_i and a block v_j. So now we let x_i be a point chosen at random in v_i. Clearly the covariance between x_i and v_j written as $C(x_i, v_j)$ is just the covariance between v_i and v_j, that is $C(v_i, v_j)$.

In the same way, if x_i and x_j are points chosen at random in their respective blocks then the covariance $C(x_i, x_j)$ equals $C(v_i, v_j)$ too. Note that if $v_i = v_j = v$ this covariance $C(v,v)$ which is equal to the variance of the blocks also represents the covariance between two samples chosen at random in a block. In constrast to this, the covariance between a sample and itself (even if the sample is selected at random) is always equal to $C(0)$, that is, the variance of the samples.

Except for this last remark, we see that the point - point and point - block covariances built up in this model by "discretizing" the blocks are just equal to the block - block covariance which is considered as known.

11.2 The Discrete Gaussian Model (2)

We associate the corresponding gaussian value Y_{v_i} with each block grade $Z(v_i)$, in the same way that we associated a gaussian equivalent $Y(x_i)$ with each sample grade $Z(x_i)$. In the previous chapter we saw that if we know the coefficients of the Hermite polynomials ϕ_n for the point anamorphosis, we can deduce those for the block anamorphosis $\phi_{(v)}$: they are just $\phi_n r^n$. The change of support coefficient r was calculated so that the block variance corresponds to the correct value.

In order to link the different blocks and samples we are now going to assume that any set of point gaussian equivalents $Y(x_i)$, $Y(x_j)$, ... and of block gaussian equivalents Y_{v_i} , Y_{v_j} , ... is multivariate normal. As these gaussian equivalents have zero mean and unit variance, the model is fully determined once we know the covariance between them. We now work this out.

Let $\varrho_{v_i v_j}$ be the covariance (or the correlation) between Y_{v_i} and Y_{v_j}. As the corresponding raw variables $Z(v_i)$ and $Z(v_j)$ have $\phi_n r^n$ as the coefficients of their anamorphosis, we can check that:

$$C(v_i, v_j) = \sum_1 \phi_n^2 r^{2n} (\varrho_{v_i v_j})^n \qquad [11-1]$$

In practice, this equation is inverted to give us the covariance $\varrho_{v_i v_j}$ once we know $C(v_i, v_j)$ and ϕ_n and r.

Now let $\varrho_{x_i v_j}$ be the covariance (or the correlation) between the gaussian equivalent $Y(x_i)$ and Y_{v_j}. This time, the coefficients of the anamorphosis for $Z(x_i)$ are ϕ_n, while those for $Z(v_j)$ are $\phi_n r^n$. As before we can show that:

$$C(x_i, v_j) = \sum_1 \phi_n^2 r^n (\varrho_{x_i v_j})^n$$

Now since $C(x_i, v_j)$ is equal to $C(v_i, v_j)$ we obtain:

$$C(v_i, v_j) = \sum_1 \phi_n^2 r^n (\varrho_{x_i v_j})^n$$

By comparing this equation with [11-1], we can deduce the relation between point-block gaussian covariance and the block-block one:

$$\varrho_{x_i v_j} = r\, \varrho_{v_i v_j} \qquad [11\text{–}2]$$

That is, one is obtained from the other by multiplying by r.

Lastly, we can also show that the covariance $\varrho_{x_i x_j}$ between Y_{x_i} and Y_{x_j} satisfies:

$$C(x_i, x_j) = \sum_1 \phi_n^2 (\varrho_{x_i x_j})^n$$

So:

$$C(v_i, v_j) = \sum_1 \phi_n^2 (\varrho_{x_i x_j})^n$$

Similarly, by comparison with [11-1], $\varrho_{x_i x_j}$ is obtained from $\varrho_{v_i v_j}$ by multiplying by r^2:

$$\varrho_{x_i x_j} = r^2\, \varrho_{v_i v_j} \qquad [11\text{–}3]$$

Thus the discrete gaussian model is fully determined once we know the change of support coefficient r and the covariance between the block gaussian equivalents $\varrho_{v_i v_j}$.

In practice we proceed as follows. We regularize the covariance $C(h)$ of the raw variable $Z(x)$ to get $C(v_i, v_j)$ for the different possible distances between v_i and v_j. We can then obtain the values of $\varrho_{v_i v_j}$ for these distances by inverting the formula given above. These values are then fitted into a covariance model. Once we know this covariance, the change of support coefficient and the Hermite polynomial coefficients, the discrete gaussian model is fully specified.

So it is a simple, neat and yet mathematically consistent model. Moreover it turns out to be very useful in practice in non linear geostatistics. In the following chapters we shall see how it can be used for local estimation using bigaussian DK or multinormal conditional expectation. Moreover, it can also be used for non conditional simulations and, thanks to the valuable properties of the multinormal distribution, conditional simulations.

Exercise:

Although it was not used explicitely the way we built the model, the hypothesis (c) of 11.1 is satisfied in the gaussian discrete model. More exactly it is satisfied through the relations between the covariances of the gaussian equivalents, as we will see in a particular case.

Let us take two blocks v_i and v_j and a sample x_i inside v_i. Let $Z(v_i)$, $Z(v_j)$ and $Z(x_i)$ be the corresponding grades and Y_{v_i}, Y_{v_j} et $Y(x_i)$ their gaussian equivalents. We will now show that, knowing the grade of its block, the grade of the sample x_i is independent of the other block, or equivalently, that, knowing Y_{v_i}, $Y(x_i)$ is independent of Y_{v_j}.

a) Show that $Y(x_i)$ can be written as:

$$Y(x_i) = r\, Y_{v_i} + \sqrt{1-r^2}\; U_1$$

with U_1 standard normal variate independent of Y_{v_i}, and that in particular, knowing $Y_{v_i} = y_{v_i}$, $Y(x_i)$ is equal to:

$$Y(x_i) = r\, y_{v_i} + \sqrt{1-r^2}\; U_1$$

Hint: Look at exercise 2 of 5.8 and use the fact that Y_{v_i} and $Y(x_i)$ are a standard normal pair with correlation r.

b) By the same way show that Y_{v_j} can be written as:

$$Y_{v_j} = \varrho_{v_i v_j}\, Y_{v_i} + \sqrt{1-\varrho_{v_i v_j}^2}\; U_2$$

with U_2 standard normal variate independent of Y_{v_i}, and that in particular, knowing $Y_{v_i} = y_{v_i}$, Y_{v_j} is equal to:

$$Y_{v_j} = \varrho_{v_i v_j}\, y_{v_i} + \sqrt{1-\varrho_{v_i v_j}^2}\; U_2$$

c) From the previous relations deduce that the covariance between $Y(x_i)$ et Y_{v_j} equals:

$$Cov\,(Y(x_i)\,,\; Y_{v_j}) = E\,(Y(x_i)\; Y_{v_j}) = r\,\varrho_{v_i v_j} + \sqrt{1-r^2}\;\sqrt{1-\varrho_{v_i v_j}^2}\; E(U_1\, U_2)$$

but that, conditionally to $Y_{v_i} = y_{v_i}$, the covariance between $Y(x_i)$ and Y_{v_j} is reduced to:

$$\sqrt{1-r^2}\;\sqrt{1-\varrho_{v_i v_j}^2}\; E(U_1\, U_2)$$

d) From the formula [11–2] the covariance between $Y(x_i)$ and Y_{v_j} in the discrete gaussian model equals $r\,\varrho_{v_i v_j}$. Deduce:

$$E(U_1\, U_2) = 0$$

Conditionally to Y_{v_i}, the covariance between $Y(x_i)$ and Y_{v_j} is then zero, so that these normal variables are independent as expected.

11.3 Disjunctive Kriging

In the discrete gaussian model, the multivariate distributions of the gaussian variables Y_{v_i} , Y_{v_j} , ... $Y(x_i)$, $Y(x_j)$, ... are multivariate normal. In particular, the

bivariate distributions are bivariate normal. This allows us to use disjunctive kriging. The problem is to use the appropriate covariance models.

Suppose that we want to estimate the indicator functions $1_{Z(v)<z}$ for a block v given the neighbouring data $Z(x_\alpha)$. This comes back to estimating $1_{Y_{(v)}<y}$ given $Y(x_\alpha)$, where $Z(v) = \Phi_{(v)}(Y_v)$ and $z = \Phi_{(v)}(y)$. Now $1_{Y_v<y}$ is a function of Y_v and it can be expanded in terms of Hermite polynomials $H_n(Y_v)$. This was given in 5–2:

$$1_{Y_v<y} = G(y) + \sum_1 \frac{1}{\sqrt{n}} H_{n-1}(y)\, g(y)\, H_n[Y_v]$$

In order to estimate this by disjunctive kriging, we have to krige each $H_n(Y_v)$ given $H_n(Y_\alpha)$ using simple kriging. That is:

$$[1_{Y_v<y}]^{DK} = G(y) + \sum_1 \frac{1}{\sqrt{n}} H_{n-1}(y)\, g(y)\, [H_n(Y_v)]^K$$

We need to krige each of the $H_n(Y_v)$ only once to obtain the DK estimator of any function of Y_v or of $Z(v)$ that we want to estimate.

The kriging of $H_n(Y_v)$ with a known mean (equal to zero) can be written as:

$$[H_n(Y_v)]^K = \sum_\alpha \lambda_{n\alpha} H_n(Y_\alpha)$$

but it is more difficult to write the corresponding system. To start with, we write it in general form:

$$\sum_\beta \lambda_{n\beta}\ Cov[H_n(Y_\alpha), H_n(Y_\beta)] = Cov[H_n(Y_\alpha), H_n(Y_v)]$$

Then we note that as the distributions are bivariate normal,

$$Cov[H_n(Y_\alpha), H_n(Y_\beta)] = [Cov(Y_\alpha, Y_\beta)]^n$$

and

$$Cov[H_n(Y_\alpha), H_n(Y_v)] = [Cov(Y_\alpha, Y_v)]^n$$

So the kriging system is:

$$\sum_\beta \lambda_{n\beta} [Cov(Y_\alpha, Y_\beta)]^n = [Cov(Y_\alpha, Y_v)]^n \qquad \text{for all } \alpha$$

As we are using the discrete gaussian model we know that:

$$Cov(Y_\alpha, Y_v) = r\, Cov(Y_{v_\alpha}, Y_v)$$

$$= r\, \varrho_{v_\alpha v}$$

and:

$$Cov\ (Y_\alpha, Y_\beta) = r^2\ Cov\ (Y_{v_\alpha}, Y_{v_\beta})$$

$$= r^2\, \varrho_{v_\alpha v_\beta}$$

except for $\alpha = \beta$:

$$Cov\ (Y_\alpha,\ Y_\alpha) = Var\ Y_\alpha = 1$$

So the system turns out to be:

$$\lambda_{n\alpha} + \sum_{\alpha \neq \beta} \lambda_{n\beta}\ r^{2n}\ (\varrho_{v_\alpha v_\beta})^n = r^n\ (\varrho_{v_\alpha v})^n \qquad \text{for all } \alpha$$

As n increases, the structure of the H_n tends to pure nugget; the kriging weights tend to 0 and so $[H_n\ (Y_v)]^K$ tends to its mean (zero) rapidly. Consequently only a few kriging systems have to be solved in practice.

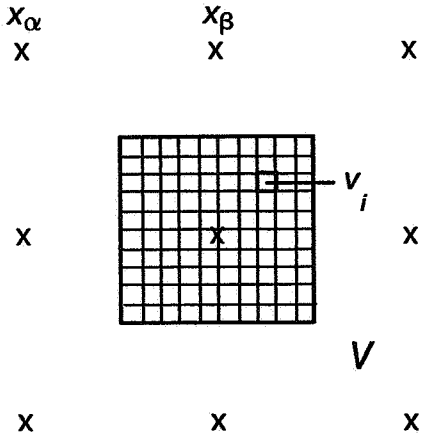


Figure 11–1

In mining for example, we do not try to estimate the recoverable reserves for each small block, e.g. $1_{Z(v)\geq z}$. We are satisfied with estimating the reserves corresponding to the N small blocks v_i inside a panel V which is the same size as the sample grid. Figure 11–1. The ore tonnage recovered at a cutoff grade of z is then equal to the number of blocks above the cutoff:

$$\sum_i 1_{Z(v_i)\geq z}$$

It is more convenient to express this as a proportion of the total:

$$\frac{1}{N} \sum_i 1_{Z(v_i) \geq z}$$

The quantity of metal contained in this is just:

$$\frac{1}{N} \sum_i Z(v_i)\, 1_{Z(v_i) \geq z}$$

It is not necessary to estimate $H_n\,(Y_{v_i})$ for each small block in order to get the DK estimator of this. As we saw earlier in chapter 8:

$$[\frac{1}{N} \sum_i 1_{Z(v_i) \geq z}] = 1 - G(y) - \sum_1 \frac{1}{\sqrt{n}}\, H_{n-1}\,(y)\, g(y)\, [\frac{1}{N} \sum_i H_n\,(Y_{v_i})]$$

So we need krige only $\frac{1}{N} \sum_i H_n\,(Y_{v_i})$ directly. The kriging system is then:

$$\lambda_{n\alpha} + \sum_{\beta \neq \alpha} \lambda_{n\beta}\, r^{2n}\,(\varrho_{v_\alpha v_\beta})^n = \frac{1}{N} \sum_i r^n\,(\varrho_{v_\alpha v_i})^n \qquad \text{for all } \alpha$$

Note: The right hand side is the mean of the n^{th} power of $\varrho_{v_\alpha v_i}$ for all the small blocks v_i, and *not* the n^{th} power of the mean of the $\varrho_{v_\alpha v_i}$!

11.4 Conditional Expectation

As all the multivariate distributions considered in the discrete gaussian model are normal, this model can also be used for the conditional expectation estimator. Given the $Y(x_\alpha) = y_\alpha$, the variable Y_v is normal. Its mean is the simple kriging of y_v:

$$(y_v)^K = \sum \lambda_\alpha\, y_\alpha$$

and its variance is the kriging variance of y_v:

$$\sigma_K^2$$

A similar reasoning to the one in chapter 7–2, gives us:

$$[1_{Z(v)<z}]^{CE} = [\,1_{Y_v<y}]^{CE} = E\,[1_{Y_v<y} \mid Y(x_\alpha) = y_\alpha \ldots]$$

$$= P\,(Y_v < y \mid Y(x_\alpha) = y_\alpha \ldots)$$

$$= G\left(\frac{y - \sum \lambda_\alpha\, y_\alpha}{\sigma_K}\right)$$

where *CE* denotes the conditional expectation. This estimator has one important advantage compared to the disjunctive kriging one: it is an increasing function of z and varies from *0* to *1*. To obtain it we only need to solve the kriging system for Y_v given the Y_α:

$$\sum_\beta \lambda^\alpha \, Cov\,(Y_\alpha,\ Y_\beta) \;=\; Cov\,(Y_\alpha,\ Y_v) \qquad \text{for all } \alpha$$

Using the discrete gaussian model this becomes:

$$\lambda_\alpha + \sum_{\beta \neq \alpha} \lambda_\beta \, r^2 \, \varrho_{v_\alpha v_\beta} \;=\; r \, \varrho_{v_\alpha v} \qquad \text{for all } \alpha$$

In DK we were able to estimate the recoverable reserves inside a panel. If we now want to do that, in principle, we need to do the estimation block by block, that is

$$\frac{1}{N}\,[\sum_i 1_{Z(v_i)\geq z}]^{CE} \;=\; \frac{1}{N}\sum [1_{\,Z(v_i)\geq z}]^{CE} \;=\; \frac{1}{N}\sum [1_{Y_{v_i}\geq y}]^{CE}$$

This means solving the kriging system for each of the N small blocks. However, as the information used, the Y_α, are the same for all of the small blocks v_i inside the panel, the matrix on the left hand side is always the same. So we need to invert the matrix only once.

11.5 Uniform Conditioning using a Normal Variate

In the previous estimator, each of the N blocks Y_{v_i} inside a panel was conditioned by its kriged value with weights λ_α which vary from block to block. We can simplify this procedure by conditioning each of these blocks with a single linear combination $\sum \lambda_\alpha \, Y_\alpha$. For example, this combination could be the kriged value of $\frac{1}{N}\sum Y_{v_i}$. Using the discrete gaussian model we can calculate the required covariances, and hence the expression for the estimators. This is one type of "uniform conditioning" [i.e. the conditioning is done by a *single* variable, which is here $\sum \lambda_\alpha \, Y_\alpha$].

There are other ways of uniform conditioning. We could, for example, estimate the recoverable reserves of choosing one block $\underline{v}$ at random inside V. In that case the conditioning is done by the kriged value of $Y_{\underline{v}}$. This second type of uniform conditioning requires stronger hypotheses than the first one because the assumption made until now concerned only points and blocks: now it has to be extended to panels as well.

As computers have become more powerful these types of uniform conditioning are no longer as valuable a simplication of the conditional expectation as they were at one time. The choice nowadays is between conditional expectation and disjunctive kriging. In practice, these two estimators are approximatively equivalent. The conditional expectation, which is based on an assumption of multivariate normality as opposed to bivariate normality for disjunctive kriging, has the advantage of giving results that are completely consistent. However, if we do not make the assumption of normality, there is

no convenient way of handling the multivariate distributions needed to evaluate the conditional expectation. The strong point of disjunctive kriging is precisely that it can be used for bivariate distributions other than the normal (e.g. gamma, negative binomial, with orthogonal indicator residuals, corresponding to the mosaic model, ...). For example, we can construct a discrete gamma model in just the same way as we did for the discrete gaussian model.

12. Problems with Stationarity

When a change of support has to be taken into account, the direct estimation using gaussian DK or conditional expectation requires a strict hypothesis of stationarity. When this is too strong, the special methods presented in this chapter can be used: uniform conditioning by an imposed value, changing of support at sample points.

12.1 Disjunctive Kriging and Conditional Expectation

The problems with stationarity mentioned earlier in connexion with point estimation also arise for estimates that require a change of support. Except for the lognormal case a non-bias condition could not solve the problems for the Conditional Expectation. It cannot solve it either when a change of support is involved.

Let us now consider DK. For example, in the gaussian case, when we want to estimate a point function:

$$f\,[Y(x)] = f_0 + \sum_1 H_n\,[Y(x)]$$

we have to krige the Hermite polynomials $H_n\,[Y(x)]$ given the $H_n\,[Y(x_\alpha)]$ for the sample points:

$$[H_n\,(Y(x))]^K = \sum \lambda_a\, H_n\,[Y(x_a)]$$

As $H_n\,[Y(x)]$ and the $H_n\,[Y(x_\alpha)]$ have the same mean locally (as $Y(x)$ and $Y(x_\alpha)$ ... have the same distribution) then it is possible to impose a local unbiasedness condition:

$$\sum_a \lambda_{na} = 1$$

This operation is rather inconvenient because it leads to convergence problems and to a non optimal estimator. Nevertheless it allows us to lessen the impact of the stationarity hypothesis.

Things are very different when a change of support is involved. Now we want to estimate:

$$f\,[Y_v] = f_0 + \sum_1 f_n\, H_n\,(Y_v)$$

This means kriging $H_n\,(Y_v)$ given the corresponding point support polynomials:

$$[H_n\,(Y_v)]^K = \sum \lambda_{na}\, H_n\,[Y(x_a)]$$

But this time $H_n\,(Y_v)$ and the $H_n\,[Y(x_\alpha)]$ do not have the same mean locally, for there is no reason for the point support variables $Y(x)$ to have locally the same distribution as the

block variable Y_v. Only the raw variables $Z(v)$ and $Z(x_\alpha)$ have the same mean. Imposing an unbiasedness condition $\sum \lambda_{n\alpha} = 1$ on the kriging system for $H_n\ (Y_v)$ would be meaningless. Worse, this unjustified action would unfortunately cause some of the estimators to diverge (instead of converging).

So for both gaussian DK and conditional expectation we cannot do without the hypothesis of stationarity when we want to estimate a function of $Z(v)$ such as $1_{Z(v)<z}$.

12.2 Using Uniform Conditioning to Impose a Value

There is one way to overcome this problem of stationarity if we are prepared to group the estimators for all the blocks inside a panel (there must be at least several dozen blocks). This is often the case in mining geostatistics; see fig 10–3. If the data in the neighbourhood of a particular panel V are too widely spaced, then using methods based on strict stationarity is likely to provide estimates that are too far from the neighbouring values. This happens when the recoverable reserves are being estimated, in particular for the quantity of metal at zero cutoff:

$$\frac{1}{N} \sum_{1} Z(v_i)\ 1_{Z(v_i)\geq 0}$$

This is equal to the average grade of all the blocks v_i inside V, that is:

$$\frac{1}{N} \sum_{i} Z(v_i) = Z(V)$$

In linear geostatistics we normally prefer to use ordinary kriging rather than simple kriging in cases like this. We can wonder whether it would not be possible to base the estimates of the recoverable reserves on the (ordinary) kriged grade of the panel, so that the reserves corresponding to a zero cutoff would equal this value.

Before doing this we must return to the discrete gaussian model. In this, the point values have an anamorphosis function Φ:

$$Z(x) = \sum_{0} \phi_n\ H_n\ [Y(x)]$$

and the block values have a different one $\Phi_{(v)}$:

$$Z(v) = \sum_{0} \phi_n\ r^n\ H_n\ [Y_v]$$

where r is the correlation between Y_v and $Y(\underline{x})$ with $\underline{x}$ chosen at random in v. It is determined from the variance of $Z(v)$:

$$Var\ Z(v) = \sum_{1} \phi_n^2\ r^{2n}$$

If we make the fairly strong assumption that this model holds for the panel as well as the blocks, then we have:

$$Z(V) = \sum_0 \phi_n (r')^n H_n [Y_V] \qquad [12-1]$$

where the correlation r' between Y_V and $Y(\underline{x})$ is obtained from:

$$Var\ Z(V) = \sum_1 (\phi_n)^2 (r')^{2n}$$

So we multiply the point coefficient ϕ_n by $(r')^n$. But this is just the same as multiplying the block coefficients $\phi_n r^n$ by $R^n = (r'/r)^n$. The coefficient R which is less than *1.0* is the correlation between Y_V and $Y_{\underline{v}}$ for a random $\underline{v}$ inside V.

So we change from a block to a panel in the same way as from a point support to a block. Just as we obtained:

$$E [Z(\underline{x}) \mid Z(v)] = \sum_0 \phi_n E [H_n(Y(\underline{x})) \mid Y_v] = \sum_0 \phi_n r^n H_n (Y_v) = Z(v)$$

so we can obtain:

$$E [Z(\underline{v}) \mid Z(V)] = \sum_0 \phi_n (r')^n E [H_n (Y_{\underline{v}}) \mid Y_V] = \sum_0 \phi_n r^n \left(\frac{r'}{r}\right)^n H_n (Y_V) = Z(V)$$

In other words, if we know the panel grade V we also know the mean of the block grades inside the panel. It is exactly the same value. But if we know the panel grade V we can deduce the estimate of any function of $Z(\underline{v})$ (and hence of $Y_{\underline{v}}$) from it. In effect, if we assume, as usual, that the bivariate distribution of Y_V and $Y_{\underline{v}}$ is normal then we know that given that $Y_V = y_V$, the variable $Y_{\underline{v}}$ has a normal distribution:

- with mean $R\,y_V$ and
- variance $1-R^2$.

So we can see that:

$$E [f(Y_{\underline{v}}) \mid Z(V) = z(V)] = E [f(Y_{\underline{v}}) \mid Y_V = y_V]$$

$$= \int f (Ry_V + \sqrt{1-R^2}\ U)\ g(U)\ du$$

In particular we get the following equations for the recoverable reserves.

- ore tonnage at a cutoff grade $z = \Phi_{(v)}(y)$:

$$E [1_{Z(\underline{v})\geq z} \mid Z(V)] = E [1_{Y_{\underline{v}}\geq y} \mid Y_V]$$

$$= 1 - G\left(\frac{y - Ry_V}{\sqrt{1-R^2}}\right)$$

– corresponding quantity of metal:

$$E\ [Z(\underline{v})\ 1_{Z(\underline{v})\geq z}\ |\ Z(V)] = E\ [\phi_v\ (Y_{\underline{v}})\ 1_{Y_{\underline{v}}\geq y}\ |\ Y_V]$$

As the block anamorphosis function is determined by its coefficients $\phi_n\ r^n$, we can write, as we did in chapter 6.4:

$$\Phi_v\ [Y_{\underline{v}}]\ 1_{Y_v\geq y} = \sum_0 q_n\ H_n\ [Y_{\underline{v}}]$$

where here $q_n = \sum_0 \phi_p\ r^p\ U_p^n\ (y)$. So we obtain the expression for the estimated metal quantity:

$$\sum_0 q_n\ R^n\ H_n\ [Y_V]$$

To summarize, if we know the panel grade $Z(V)$ then we can deduce the recoverable reserves from it. Then the metal at a zero cutoff grade (i.e. the grade without any selection), $Z(\underline{v})$ will be estimated to be equal to $Z(V)$. In practice we do not know the true value $Z(V)$, so we can think of substituting its estimate $Z(V)^{OK}$ obtained by ordinary kriging. In this type of uniform conditioning, where this value is being imposed on the panel, it is important that Y_V should be determined so as to satisfy the formula corresponding to 12–1:

$$Z(V)^{OK} = \sum_0 \phi_n\ (r')^n\ H_n\ [Y_V]$$

even if this results in a distribution for the values of Y_V that is not a standard normal distribution. This relation effectively guarantees the underlying quality of the method; that is, it ensures that the estimate of $Z(\underline{v})$:

$$E\ [Z(\underline{v})\ |\ Z(V)^{OK}] = \sum_0 \phi_n\ r^n\ \left(\frac{r'}{r}\right)^n\ H_n\ [Y_V]$$

equals $Z(V)^{OK}$.

This method should not be confused with the other types of uniform conditioning described earlier. Compared to them, this one has clear advantages when there are problems with stationarity, even if it is not strictly rigourous.

12.3 Changing of Support at Sample Points

In this section we are going to consider a method that is generally referred to as "service variables with a change of support". Readers will know that "service variables" are additive variables that are convenient to use and that can be estimated using linear combinations in order to obtain estimates of other (non additive) variables. For example, the accumulation and the thickness of a vein can be used in 2-D to estimate the ore grade; similarly the ore tonnage and metal quantity to estimate grade; or the indicator functions

for a range of cutoffs in order to estimate the metal quantity and the grade recovered at these cutoffs. We have already seen that indicator cokriging (or disjunctive kriging, as it is also called) can be used in conjunction with non–bias conditions provided that there is no change of support involved.

The idea of this method is simple enough. It is like estimating a variable by using indirect measures. For example, in uranium mines, the U grade of blocks is often estimated from radiometric measurements. To be more precise the U grade $U(v) = \frac{1}{v}\int_v U(x)\,dx$ is estimated from point radiometric measures $Ra(x_\alpha)$. These are transformed into U grades through a regression:

$$E\,[U(x) \mid Ra(x)]$$

which is a function of $Ra(x)$, say $U[Ra\,(x)]$. Then we assume that $U(x)$ and $U[Ra\,(x)]$ are equivalent[1] in order to obtain $U(v)$ by estimating $\frac{1}{v}\int_v U[Ra(x)]\,dx$ from the $U[Ra(x_\alpha)]$ (for example by kriging). Setting up the regression and then using it at the sample points presupposes stationarity. But then the kriging can be carried out with a non–bias condition. This avoids the attraction towards the (global) mean that occurs in simple kriging when the sample spacing is too wide.

Now suppose we want to estimate a quantity such as $1_{Z(v)\geq z}$ from point samples $Z(x_\alpha)$. For each data point x_α we shall calculate the expected value of $1_{Z(v_\alpha)\geq z}$, where v_α denotes the block containing x_α. That is, we shall evaluate:

$$E\,[1_{Z(v_\alpha)\geq z} \mid Z(x_\alpha)]$$

Using the gaussian change of support model, this becomes:

$$E\,[1_{Y_{v_\alpha}\geq y} \mid Y(x_\alpha)] = 1 - G\left(\frac{y - r\,Y(x_\alpha)}{\sqrt{1-r^2}}\right)$$

because given $Y(x_\alpha) = y_\alpha$, Y_{v_α} has a normal distribution with mean ry_α and variance $\sqrt{1-r^2}$. Then we obtain the estimate of $1_{Z(v)\geq z}$ that we are looking for by kriging:

$$E\,[1_{Z(v)\geq z} \mid Z(x)] = 1 - G\left(\frac{y - r\,Y(x)}{\sqrt{1-r^2}}\right)$$

where x denotes a point in v, with the point values:

$$E\,[1_{Z(v_\alpha)\geq z} \mid Z(x_\alpha)] = 1 - G\left(\frac{y - r\,Y(x_\alpha)}{\sqrt{1-r^2}}\right)$$

1 This leads to problems when evaluating the estimation variance.

This leads us to estimating a variable at one point given the values of the same variable at the sample points. So we can use a non–bias condition in the kriging system.

The problem that arises then is to estimate several indicator functions, or several functions of $Z(v)$, simultaneously. Instead of making a separate kriging for each of them, we can co-estimate them by using (in the gaussian model) the Hermite polynomials. In this case all the functions of $Z(v)$ to be estimated, which are functions of the corresponding gaussian variable Y_v, have to be expanded in terms of Hermite polynomials:

$$f(Y_v) = \sum_0 f_n H_n (Y_v)$$

We obtain the required estimates of $E[f(Y_v) \mid Y(x)]$ by kriging each of the $E[H_n(Y_v) \mid Y(x)] = r^n H_n [Y(x)]$ using the values $E[H_n(Y_{v_\alpha}) \mid Y(x_\alpha)] = r^n H_n [Y(x_\alpha)]$. This just leads back to kriging the point Hermite polynomials which can be done using a non–bias condition (though this presents the convergence problems mentioned in chapter 9.3).

Conclusion

This text presents two principal methods for estimating functions of a variable $Z(x)$ that has been sampled. These are disjunctive kriging (or cokriging indicators) and the conditional expectation. A method of estimation is designed to work well provided that certain hypotheses are respected (or if one prefers, within a particular model). So the conditional mean can be used within a multivariate normal model (that is, if the multivariate distributions are normal). Isofactorial disjunctive kriging requires a model with isofactorial bivariate distributions: bivariate normal, or gamma, or with orthogonal indicator residuals, ... In the mosaic model, indicator cokriging is equivalent to kriging each indicator separately, with the same kriging weights for each indicator. So this is a very simple algorithm. Using it when the model is not suitable, however, (e.g. if the indicator variograms are not all the same) may be foolhardy.

Another very important point to note when one wants to estimate functions such as indicators is to take account of the change of support. The estimates of $1_{Z(x)\geq z}$ and $1_{Z(v)\geq z}$, can be quite different either globally (i.e. on average over the whole region) or locally, particularly if these is a nugget effect or a micro-structure (with a short range).

The methods described allow us to estimate functions of the variable, and so they can also be used to estimate the variable itself $Z(x)$, or the regularized variable $Z(v)$, $Z(V)$ etc... If this is the only objective we have to choose whether to use a linear estimator, i.e. a linear combination of the data $\sum \lambda^a Z(x_a)$ such as kriging, or a non linear estimator such as disjunctive kriging or conditional expectation. Although kriging is, in theory, less efficient, it requires only the variogram model (or the covariance). As it requires less stringent hypotheses, it is more robust in practice. Conversely, as the underlying hypotheses for the more elaborate and theoretically better methods are stricter, there is less chance of their being satisfied. This is why kriging is used most often when we want to estimate only the variable itself.

One exception to this rule is lognormal kriging which is routinely used for estimating the *in situ* reserves in the South African gold mines. The term *lognormal kriging* covers several different lognormal estimators, each based on different assumptions concerning the lognormality, which vary depending upon whether a non–bias condition is used or not, and upon whether there is a change of support. The conditional expectation estimator of a lognormal variable falls into this family.

References

This reference list includes a short number of publications which are chosen because of their theoretical interest or of the quality of the application. Some of them are notes of the *Centre de Géostatistique de Fontainebleau* (in French), which are publications of the *Ecole des Mines de Paris* and can be obtained by writing to the Librarian at the Centre. The present text also exists in French.

This short list of papers does not represent the efforts made by all the researchers involved in the development of non linear geostatistics. In particular I am thinking of the work of people like A. Maréchal, D. Guibal and Y. Touffait in the seventies.

To Know More About:

Disjunctive Kriging in general, gaussian DK, and the discrete gaussian change of support model:

Be careful ! Very often Hermite polynomials are not normalised. Hence some modifications when expanding the functions.

G. Matheron 1973, Le krigeage disjonctif. Note du Centre de Géostatistique, N–360.

First note on DK models: hermitian (including gaussian), gamma, negative binomial, Poisson.

G. Matheron 1976, A simple substitute for conditional expectation: the disjunctive kriging, Proceedings, NATO ASI : "Advanced Geostatistics in the Mining Industry", Rome, Oct. 1975, pp. 221–236.

Publication in English on DK, from the 1973 note: the basis of DK and the hermitian models.

G. Matheron 1976, Forecasting block grade distribution: the transfer functions, Proceedings, NATO ASI : "Advanced Geostatistics in the Mining Industry", Rome, Oct. 1975, pp. 237–251.

Or how to estimate locally the recoverable reserves ("transfer functions") of a mining deposit using DK. The discrete gaussian model and another one (hermitian) are used for the change of support.

G. Matheron 1978, L'estimation globale des réserves récupérables, Note du Centre de Géostatistique, C–75.

Lecture note. Details the discrete gaussian model and the change of support.

G. Matheron 1978, Le krigeage disjonctif et le paramétrage local des réserves, Note du Centre de Géostatistique, C–76.

Lecture note. The gaussian DK. The application to the recoverable reserves with the discrete gaussian model. Also the estimation of the reserves using the multinormal conditional expectation, and the simplifying uniform conditionings with a gaussian variate.

G. Matheron 1979, La géostatistique, *in* Encyclopedia Universalis.

An overall instructive presentation of mining geostatistics: estimation problems encountered, formalism proposed, solutions brought (in situ resources, recoverable reserves, linear and non-linear geostatistics, etc ...). Includes in particular the change of support discrete gaussian model.

D. Guibal and A. Zaupa–Remacre 1984, Local estimation of the recoverable reserves: comparing various methods with the reality on a porphyry copper deposit, Proceedings, 2nd NATO ASI : "Geostatistics for Natural Resources Characterization", Tahoe, Sept. 1983, G.Verly et Al. Eds, D. Reidel Publ. Co., Dordrecht, Holland.

Conditional expectation, uniform conditioning with a gaussian variate, DK: recalling the methods and applying to mining data.

A. Zaupa–Remacre 1984, L'estimation du récupérable local – le conditionnement uniforme, Thèse de Docteur–Ingénieur, Ecole des Mines de Paris, 99p.

Problems of stationarity and influence of the grid. DK, conditional expectation, uniform conditioning with a normal variate, uniform conditioning with the kriged panel grade, are tested on mining data (gaussian framework).

J.–F. Bouchind'homme, 1980, Estimation de l'uranium récupérable sur les gisements sédimentaires, stratiformes, exploitables à ciel ouvert, Thèse de Docteur–Ingénieur, Institut National polytechnique de Lorraine, Nancy, 185p.

DK and kriging service variables (with and without change of support in the gaussian framework).

R. Webster and M.A. Oliver 1989, Disjunctive kriging in agriculture, Proceedings, 3rd International Geostatistics Congress, Avignon, 5–9 Sept. 1988 : "Geostatistics", M. Armstrong Editor, Kluwer Academic Publ., Dordrecht, Holland.

Applies gaussian DK to agriculture.

R. Webster and M.A. Oliver 1989, Optimal interpolation and isarithmic mapping of soil properties. VI Disjunctive kriging and mapping the conditional probability. Journal of Soil Science, Vol. 40, pp. 497–512.

A more developped presentation of the previous paper.

Ch. Lantuéjoul et J. Rivoirard 1984, Une méthode de détermination d'anamorphose, Note du Centre de Géostatistique, N–916.

Details the gaussian anamorphosis of an empirical distribution.

J. Rivoirard 1985, La convergence des développements en polynômes d'Hermite, Etudes Géostatistiques, Séminaire CFSG 17–18 juin 1985 Fontainebleau, Sciences de la Terre, Série Informatique, n°24.

In practice, expansions in Hermite polynomials are truncated. What does this make ? Up to which order to go ?

Change of Support:

G. Matheron 1981, La sélectivité des distributions, Note du Centre de Géostatistique, N–686.

The general properties of the selectivity curves. Cartier's relation. The influence of support and information effects on the selectivity curves.

G. Matheron 1984, The selectivity of the distributions and the "second principle of geostatistics", Proceedings, 2nd NATO ASI : "Geostatistics for Natural Resources Characterization", Tahoe, Sept. 1983, G.Verly et Al. Eds, D. Reidel Publ. Co., Dordrecht, Holland.

Synthesis of the previous note.

C. Lantuéjoul 1990, Cours de sélectivité, Note du Centre de Géostatistique, C–140.

Same subject, presented as lectures with exercises. With the necessary recallings in probability. For the reader who wishes to obtain a better theoretical knowledge of the subject.

Ch. Kavourinos 1987, The Grade–Tonnage Curves for a Zinc Mine in France, *in* Geostatistical Case Studies, Ed. Armstrong and Matheron, Reidel.

Support and information effects are illustrated on a selective mining example. Experimental results and theoretical ones obtained with the gaussian model.

Cl. Demange et al. 1987, Global Recoverable Reserves: Testing Various Change of Support on Uranium Data, *in* Geostatistical Case Studies, Ed. Armstrong and Matheron, Reidel.

The discrete gaussian, gamma, negative binomial, and mosaïc change of support models are used and lead to very close results.

Isofactorial Diffusion Models (except the gaussian one):

G. Matheron 1984, Modèles Isofactoriels et Changement de Support, Sciences de la Terre, n°18.

G. Matheron 1984, Isofactorial models and change of support, Proceedings, 2nd NATO ASI : "Geostatistics for Natural Resources Characterization", Tahoe,

Sept. 1983, G.Verly et Al. Eds, D. Reidel Publ. Co., Dordrecht, Holland.

Presenting some isofactorial models and their change of support (gamma, negative binomial, Poisson).

Hu Lin–Ying 1988, Mise en oeuvre du modèle gamma pour l'estimation de distributions spatiales, Thèse de Doctorat de l'Ecole des Mines de Paris.

How to use and ajust the gamma model. Its advantages over the gaussian model. Three sets of experimental data are used.

G. Matheron 1984, Une méthodologie générale pour les modèles isofactoriels discrets, Sciences de la Terre n°21, pp. 1–78.

The general way to build isofactorial models with a discrete marginal law.

Ch. Lajaunie et Ch. Lantuéjoul, 1989, Setting up the general methodology for discrete isofactorial models, Proceedings, 3rd International Geostatistics Congress, Avignon, 5–9 Sept. 1988 : "Geostatistics", M. Armstrong Editor, Kluwer Academic Publ., Dordrecht, Holland.

Estimating the parameters for the models of the previous note.

Indicator Kriging and Mosaic Model:

A. Journel 1982, The indicator approach to estimation of spatial distributions, Proceedings of the 17th APCOM International Symposium, T. B. Johnson and R. J. Barnes Editors, Society of Mining Engineers of the American Institute of Mining, Metallurgical, and Petroleum Engineers Publ., New York, pp. 793–806.

An article in favour of kriging of indicators.

G. Matheron 1982, La destructuration des hautes teneurs et le krigeage des indicatrices, Note du Centre de Géostatistique, N–761.

Kriging the indicators: when does this represent their cokriging ? The mosaic model.

G. Matheron 1984, Changement de support en modèle mosaïque. Colloque "Computers in Earth Sciences for Natural Resources Characterization", Nancy, 9–13 Avril 1984, Sciences de la Terre n°20, Nancy.

Or how to make the change of support, when point DK corresponds to kriging the indicators.

Orthogonal Indicator Residual Models:

J. Rivoirard 1988, Modèles à résidus d'indicatrices autokrigeables, Etudes géostatistiques V, Séminaire CFSG 15–16 Juin 1987, Fontainebleau, Sciences de la Terre n°28, Série informatique, Nancy.

First presentation of these models. Includes the change of support.

J. Rivoirard 1989, Models with orthogonal indicator residuals, Proceedings, 3rd International Geostatistics Congress, Avignon, 5–9 Sept. 1988 : "Geostatistics", M. Armstrong Editor, Kluwer Academic Publ., Dordrecht, Holland.

More formal presentation. With the generalisation to nested geometrical sets.

Lognormal Estimators of In Situ Reserves:

H. S. Sichel 1949, Mine valuation and maximum likeliwood, Master's thesis, University of the Witwatersrand, Johannesburg.

H. S. Sichel 1966, The estimation of means and associated confidence limits for small samples from lognormal populations, p. 106–122. *In* Symposium on Mathematical Statistics and Computer Applications in Ore Valuation, S. Afr. Inst. Min. Metall., Johannesburg.

D. G. Krige 1951, A statistical approach to some mine valuation and allied problems on the Witwatersrand, Master's thesis, University of the Witwatersrand, Johannesburg.

G. Matheron 1962, Traité de Géostatistique appliquée, Tome I, Mémoires du Bureau de Recherches Géologiques et Minières, Ed. Technip, Paris.

G. Matheron 1974, Effet proportionnel et lognormalité, ou le retour du serpent de mer, Note du Centre de Géostatistique, N–374.

P. Dowd 1982, Lognormal kriging – The general case, Mathematical Geology, Vol. 14, p. 474–500.

J. Rivoirard 1990, A Review of Lognormal Estimators for *In Situ* Reserves, Mathematical Geology, Vol. 22, p. 213–221.

Evolution of geostatistics:

G. Matheron 1976, Les concepts de base et l'évolution de la géostatistique minière, Proceedings, NATO ASI : "Advanced Geostatistics in the Mining Industry", Rome, Oct. 1975, pp. 3–10.

G. Matheron 1976, Le choix des modèles en géostatistique, Proceedings, NATO ASI : "Advanced Geostatistics in the Mining Industry", Rome, Oct. 1975, pp. 11–27.

G. Matheron and W. J. Kleingeld 1987, The Evolution of Geostatistics, Proceedings of the 20th APCOM International Symposium, I. C. Lemmer, H. Schaum and F. A. Camisani-Calzolari Editors, SAIMM and MINTEK, Johannesburg, Vol. 3, pp. 9–12.

D. G. Krige, M. Guarascio and F. a. Camisani-Calzolari 1989, Early South African Geostatistical Techniques in today's perspective, Proceedings, 3rd International Geostatistics Congress, Avignon, 5–9 Sept. 1988 : "Geostatistics", M. Armstrong Editor, Kluwer Academic Publ., Dordrecht, Holland, pp. 1–19.

Other References:

G. Matheron 1965, Les variables régionalisées et leur estimation, Thèse de Doctorat d'Etat, Ed. Masson, Paris.

A. Journel 1977, Géostatistique Minière, Thèse de Doctorat d'Etat, Fontainebleau.

A. G. Journel et Ch. Huijbregts 1978, Mining Geostatistics, Academic Press, London.