CHANGE OF SUPPORT AND TRANSFORMATIONS

D. E. Myers
Department of Mathematics
University of Arizona
Tucson, AZ 85721 USA
USA

ABSTRACT

The practical and theoretical effects of using non-point support data for estimating variograms or on the kriging equations when estimating spatial averages, i.e., block kriging, are well-known. Under an assumption of lognormality the proportional effect is also well-known. While other transformations are commonly used in statistics only the log and indicator transforms are widely used in geostatistics, the latter has the advantage of generally not requiring an inverse transform.. Additional theoretical and empirical results are presented on the interrelationship between non-point support data, non-linear transformations and variogram estimation, modeling. The non-point support data may incorporate spatial averages or compositing of point support data.

INTRODUCTION

Let $Z(x)$ be a regionalized variable defined in 1, 2 or 3 space and $H(Z)$ a real linear functional, i.e., a mapping of Z into the real numbers. The two most common examples of H are point evaluation, i.e., $H(Z)$ is simply $Z(x_0)$, and spatial averages Z_V, the average value of Z over a volume V. Much of geostatistics has been concerned with one of two problems; estimation of the linear functional $H(Z)$ or estimation of a probability distribution associated with H. For example, let x_0 range over all possible values in a region or let V range over possible congruent volumes within the region. The resulting probability distributions for the point valuation functional and the spatial average functional are of interest in many applications. These problems generalize when transformations are allowed, either on the domain of Z or on the range of Z.

STRUCTURE FUNCTIONS

Linear geostatistics is based on the use of a structure function, i.e., the variogram or

R. Dimitrakopoulos (ed.), Geostatistics for the Next Century, 253–258.

covariance of Z. The relationship between the covariance of the common linear functionals of Z(x) and the structure function for Z(x) is well known. The relationship between the probability distributions is generally not known or at least only under strong assumptions. The relationship between the covariances of non-linear functionals of Z(x) and the structure function of Z(x) is generally not known. Since the class of continuous mappings is very large and contains as a subset the mappings with continuous second derivatives many problems involving non-linear functionals or non-linear estimators can be resolved by generating those functionals or estimators using such transformations or mappings. A number of known results will be described in this general context and then open problems will be presented.

GENERAL ESTIMATION PROBLEMS

Extending the approach in Cressie (1993) define general classes of non-linear functionals

$H_1(Z) = g(Z)$ (1a), $H_2(Z) = g(Z_V)$ (1b), $H_3(Z) = (g(Z))_V$ (1c)

where g(u) is a real valued function defined on a subset of the reals. If in addition g has an inverse then the following is of interest.

$$H_4(Z) = g^{-1}((g(Z))_V) \quad (1d)$$

Note that in general $H_3(Z) \neq H_2(Z)$ and $H_4(Z) \neq Z_V$. In the case of (1a), (1c) one can simply transform the data $Z(x_1),...,Z(x_n)$ into new data $g(Z(x_1)),...,g(Z(x_n))$ and use a linear estimator of the form

$$H^*(Z) = \Sigma_{i=1,n} a_i g(Z(x_i)) \quad (2)$$

Although $g^{-1}g(Z) = Z$, in general $g^{-1}(H^*(Z)) \neq \Sigma_{i=1,n} a_i Z(x_i)$

Examples

Let I(x;z) be the Indicator function associated with Z(x), Journel (1983) That is, I(x;z) = $g_z(Z(x))$ where $g_z(u) = 0$ if u > z and 1 otherwise. As has been pointed out by Cressie (1993), there are two equivalent problems. First, the non-linear estimator can be used to estimate a probability distribution function, or it can be used to estimate values of the non-linear functional obtained by applying the indicator transform to the point evaluation functional. That is, for a certain non-linear functional F we have

$F^*(z) = \Sigma_{i=1,n} a_i g_z(Z(x_i))$ (3) or $g_z(Z(x_0)) = \Sigma_{i=1,n} a_i g_z(Z(x_i))$ (4)

For a volume V centered at the origin, let V_x be the volume rigidly translated to the point x. Z_{Vx}, as is common in the literature, denotes the spatial average over the translated volume. Although $Z(x_1),...,Z(x_n)$ (multivariate) lognormally distributed is equivalent to $LnZ(x_1), ...,LnZ(x_n)$ being multivariate normal, determining the distribution of Ln Z_{Vx} is more difficult than for $(\mathrm{Ln}\, Z)_{V_x}$. Although the sum of normal random variable is again normal the same is not true for lognormal random variables although it is often considered to be approximately true empirically. Journel(1980) makes this assumption in deriving the bias adjustment for lognormal kriging.

In the case of the indicator transform Cressie (1993) has suggested an alternative. Since it is straightforward to estimate Z_V by a linear combination of the data for Z, why not estimate $g(Z_V)$ as the transformation of a linear combination of the data (not necessarily one of the usual kriging estimators). More explicitly,

$$H\hat{}(Z) = g(\Sigma a_i Z(x_i)) \tag{5}$$

While in general $E\{g(\Sigma a_i Z(x_i))\} \neq g(E\{\Sigma a_i Z(x_i)\})$ unbiasedness of H^(Z) could be imposed as a constraint. Similar conditions on the variances could be imposed and Cressie minimizes the variance of the error of estimation subject to two constraints and hence uses two Lagrange multipliers. H^(Z) is essentially the simple block kriging estimator for Z_V BUT an additional constraint has been imposed. The new "kriging" equations require only the covariance function of Z (as well as the point-to-block and block-to-block variances computed from that covariance). Interestingly enough, *these new kriging equations do not depend on the transformation g*. This is because g has been assumed to have continuous second derivatives, i.e., a smoothness condition has been imposed on g. The resulting estimator is viewed as an approximation to $E[g(Z_V) \mid Z(x_1),...,Z(x_n)]$. By analogy with the usual simple kriging estimator one is inferring the conditional distribution of $g(Z_V)$ rather than the distribution of Z_V. Finally it is easy to see from these new kriging equations that the system may be unstable when the block size is too large, i.e., the block-to-block variance is too small.

DISTRIBUTIONS AND TRANSFORMATIONS

Let U be a random variable and g a one-to-one differentiable mapping. Then the probability distribution of g(U) is completely determined by the probability distribution of U. If $U_1,....,U_n$ are jointly distributed random variables and $W_1(U_1,....,U_n),...,W_n(U_1,....,U_n)$ are on-to-one transformations with continuous partial derivatives and whose Jacobian does not vanish then the joint distribution of $W_1,...,W_n$ is determined by the joint distribution of the $U_1,....,U_n$. As a special case let $W_i = g(U_i)$. This is essentially the problem considered above. However in geostatistics the joint distribution function for Z is generally NOT considered known and hence the general result is not useful. Note that even in the case of "nice" transformations and "nice" joint distributions it may be difficult to compute the new joint distribution. The change of variable theorem is not directly applicable for such functions as $g(u) = u^2$ and a multivariate normal distribution. Although the problem of obtaining the distribution of a function of a random variable is a difficult one, computing the first and second moments (from the distribution of the original random variable) is straightforward and the conditions are less restrictive than the general change of variable theorem. This idea was exploited by Matheron (1985).

Because the Gaussian is essentially the only multivariate distribution that is characterized by the bivariate correlations it is the only distribution for Z(x) that allows easy computation of the variogram of g(Z) in terms of the variogram of Z. Note however that the variogram of $(g(Z))_{V_x}$ is easily computable in terms of the variogram of g(Z). That is,

$$\gamma_{(g(Z))V}(h) = 0.5 \ \mathrm{Var}\{(g(Z))_{V_{x+h}} - (g(Z))_{V_x}\} = (1/V^2)\int_V\int_V \gamma_{g(Z)}(x-y)dxdy \qquad (6)$$

Hence if the data is transformed the variogram of g(Z) can be estimated and modeled, from this the variogram for $\gamma_{(g(Z))V}(h)$ is computable.

ALTERNATIVE APPROACHES

When the transformation is sufficiently smooth, i.e., the second derivative is continuous, then the "value" at one point can be approximated by the value at a nearby point. Linear functionals can then be applied to this approximation. Instead of transformations of Z, the regionalized variable can be considered as a non-linear transformation. Matheron (1985) utilizes this approach to approximate Z_{V_x} by Z(x). The approximation is valid at least for small V. Consider x to be a (vector) random variable uniformly distributed over the region of interest. Then Z(x) is obtained as a transformation applied to this (uniform) random variable. Since the distribution of x is known and the transformation is "known" the moments of Z(x) are computable in two ways; one in terms of the transformation and the uniform distribution of x, secondly in terms of the unknown distribution of Z(x). Note that this relationship extends to composite transformations, g(Z(x)) at least for functions g where the moments exist. By combining the idea of approximating Z in terms of its first two derivatives and equivalence of the two methods for computing moments, it is possible to approximate the distribution of g(Z(x)) and of $g(Z_V)$.

REVERSE PROBLEMS

The change of support problem is generally thought of in terms of determining the characteristics of or estimating Z_V given data for Z and some form of structural information for Z (such as the variogram). However there are many problems where the reverse is equally important. Block size is important in mining applications because it is related to selectivity and hence to tonnage. Block size is also important in environmental remediation because it is related to the scale of remediation. The simplest form of remediation for contaminated soil consists of removing soil to some fixed depth over a specified area. In a manner not dis-similar to that of cut-off grades for ore, contaminants usually have toxicity levels. These may be support dependent however and are exposure related. Although the average concentration of a contaminant will generally decrease as the block size increases, exposure potential may not decrease with block size. Although the cost of remediation is related to the total amount of soil removed it is also related to the number of blocks of soil to be removed and there will be a minimum block size related to the equipment to be used. If the distribution of average concentrations, for blocks of a certain size, is known can we infer the distribution for smaller blocks. In particular if none of the blocks have an average contaminant concentration above the toxicity level, is there some assurance that none of the blocks of a fixed smaller size will have average concentrations in excess of the toxicity level? Can average concentrations for small blocks be estimated from large block concentrations. This is a serious problem because in comparison to the assaying of ore samples the laboratory analysis of environmental samples is usually very costly. The use of large blocks corresponds to the use of composited samples.

Because toxicity levels are very low for many contaminants it is not uncommon to have data reported as "non-detects" or "not quantified". In the latter case the substance was detected but at such a low concentration that the results are unreliable. Because there are examples of toxicity levels very close to the detection levels it is not appropriate to consider non-detects as zeros. Note that compositing may actually make this problem worse. The question of how to estimate variograms in the presence of such data has received little attention.

When g is the indicator transform (cut-off value = z) then $(g_z(Z))_V$ can be interpreted as the proportion of (points in) V where the value of $Z(x) \leq z$. Alternatively this is the probability that if a point is chosen at random in V then the value will be $\leq z$. For a collection of disjoint, congruent sub-blocks the variogram quantifies the correlations between these probabilities. Hence if the variogram of $(g_z(Z))_V$ has a very long range (compared to the size of the region of interest) and a small sill then the probability distribution is nearly the distribution of the entire region.

ASPECTS OF COMPOSITING

Let $Z^T = [\, Z(x_1),\ldots,Z(x_n)]$ and A a k x n matrix with non-negative entries satisfying two additional conditions; (1) for any one column at most one row has a non-zero entry, (2) $AU_n = U_p$ where U_n , U_k are column vectors with all entries 1's. A has the effect of compositing the data vector Z. Note that in a multivariate case Z would be an n x p matrix but the compositing matrix would function in the same way. Let $\mathbf{Y} = [\, Y(x_1),\ldots,Y(x_k)]^T$ = $\mathbf{AZ}$ be the composited data vector, $\mathbf{Z} = [Z(x_1),\ldots.,Z(x_n)]^T$. Then for any point x_0, the composited kriged estimate for $Z(x_0)$ be given by

$$Z^*(x_0) = \Sigma\lambda^c{}_i Y(x_i) \tag{7}$$

Although the Y's will not be associated with locations as such the covariance matrix C_Y with entries $C_{ij,Y} = \mathrm{Cov}\{\, Y_i, Y_j\}$, is expressible in terms of the covariance matrix of Z and hence in terms of the variogram of Z. Let $\lambda = [\lambda_1,\ldots.,\lambda_p]^T$. Then $C_Y = AC_ZA^T$ and the weights in (7) are found as the solution to

$$AKA^T \lambda + U_p \mu = AK_0$$

$$U_p^T \lambda = 1 \qquad (8)$$

where K is the matrix of variogram entries for Z (between sample locations) and K_0 is the vector of variograms (between the sample locations and the location to be estimated). Note that columns in the coefficient matrix will not coincide with the column on the right hand side even if x_0 is one of the data locations. Hence the values of the components of composited samples can be estimated using the composit sample data.

REFERENCES

Box, G.E.P. and Cox, D.R. (1964) "An Analysis of Transformations", J. Royal Statist. Soc. B26, 211-243

Cressie, N. (1993) "Aggregation in Geostatistical Problems", in Geostatistics Troia '92, A. Soares (ed), Kluwer Academic Publishers, pp25-35

Crozel, D. and David, M. (1985) "Global Estimation Variance:Formulas and Calculation", Math. Geology 17, 785-796

Journel, A.G., (1980) "The Lognormal Approach to Predicting Local Distributions of Selective Mining Grades", Math. Geology 12, 285-303

Journel, A.G., (1983) "Non-parametric Estimation of Spatial Distributions", Math. Geology 15, 445-468

Lantuejoul, Ch. (1988) "On the Importance of Choosing a Change of Support Model for Global Reserves Estimation", Math. Geology 20, 1001-1019

Matheron,G. (1971) "The Theory of Regionalized Variables and Its Applications. Les Cahiers du Centre de Morphologie Mathematique", Fas. 5. ENSMP, Fontainebleau, 212p

Matheron, G. (1985) "Change of Support for Diffusion-Type Random Functions", Math. Geology 17, 137-165

Zhang, R., D.E. Myers and A. Warrick (1990) "Variance as a Function of Sample Support Size", Math. Geology 22, 107-121