Geostatistics - Uncertainty quantification: the indicator framework

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

September 7, 2024

Summary

The multiGaussian approach discussed previously provides a parametric approach to infer the conditional distribution at unsampled locations. An alternative to that model is given by the indicator framework.

In the indicator framework, the original variable is converted into a vector of indicators and each new variable (each indicator) is characterized and modeled separately.

For continuous variables, the indicators represent the probability of not exceeding a threshold. Therefore, they discretize the cdf. Once each indicator has been estimated by kriging, after inferring and modeling its 3D variogram, the kriged indicators can be interpreted as an estimate of the probability of not exceeding the corresponding threshold, and an "estimated cdf" can be constructed. This local

cdf provides information about the local uncertainty of the variable.

The approach can be adapted to categorical variables as well. A vector of indicator variables can replace the categorical value at every location. Each indicator represents the probability of a specific category prevailing at that location. Assuming the categories are exhaustive and mutually exclusive, then the indicator represents the probability of finding a specific category at each location. Again, we can characterize these categorical indicators by computing and modeling their 3D variogram. Then kriging can be performed to estimate the probability of each category prevailing at unsampled locations. These categories must sum to one (this is imposed as a post-processing correction). Then a single category can be selected from the local cumulative mass function.

Although more flexible than the multiGaussian approach, the indicator framework is much more laborious, and often brings practical challenges. We discuss some of these details in this chapter.

1 Introduction

The non-parametric formalism of indicators was introduced in 1983 by A. G. Journel [6, 7, 8]. Many authors have presented this approach in great detail (e.g. see [2, 5]). This method avoids the need of a multiGaussian assumption at the bivariate level and therefore avoids the so called “problem of maximum entropy” implicit in that assumption. This means that when using a multiGaussian assumption, the

spatial correlation of extreme high and low values is zero.

The indicator approach permits the random variable to have different spatial continuity for high and low values; the Gaussian formalism assumes spatial continuity is symmetric with respect to the median.

Modeling of the conditional distribution of a continuous variable can be done by discretizing the distribution with a threshold and estimating the probability of exceeding that threshold. This idea is at the core of indicator methods. Basically, the original variable is transformed into a probability: the probability of not exceeding a threshold. Then this probability is estimated at other unsampled locations. To recover the full local conditional distribution, the process is repeated for a collection of thresholds, and the entire distribution is represented by a vector of indicators.

In the case of categorical variables, a similar notion is used. As in the continuous case, the categorical variable is transformed into a vector of indicators, each one representing the prevalence of a particular category. Since we assume categories are exhaustive and mutually exclusive, that is, one and only one category exists at every location, then, these indicators represent probabilities of finding each category at that location. The local conditional probability mass function of the categorical variable, can be characterized by using these indicators.

The indicator coding is different for continuous and categorical variables. The vector of indicators at one location may have several values equal to 1 in the case of a continuous variable (or none, if the sample value is above the highest threshold). Meanwhile, in the categorical case, only one indicator will take a value of one (since only one category prevails at any one location).

After coding (in both the continuous and categorical case), the process is the usual: a variogram is computed and modeled for each indicator and kriging is performed to estimate the indicator value (which corresponds to a probability of not exceeding a threshold or probability of prevalence of a particular category). Notice that the kriging variances are not used in the subsequent steps. Results need to be post-processed to obtain the final conditional distribution from where a predicted value can be inferred.

Several important advantages are derived from this basic idea of directly estimating the probabilities [1, 9, 11]:

1. The correlation at different thresholds can be used.
2. Secondary information can be coded in the same way, which gives a great flexibility to this approach.
3. Change of support can be performed, that is, the conditional distribution inferred at point support can be corrected to represent a block support distribution.
4. Recoverable reserves of blocks can be calculated, in resource estimation for mineral deposits.

Although very flexible, the implementation of some of these advantages of the method are difficult:

- The coding of soft data, that is, data that are not precise, but are correlated with the attribute of interest (for example, a geophysical variable, which tells us something about a mineralization) as if they were hard data is useful, but secondary information cannot be used as primary, even though the coding is the same. A Linear Model of Coregionalization or Markov Model has to be used to integrate these diverse data types.

- The use of data at different support is also a difficult task, since the correlation between the variables changes at different supports [10].

As usual in geostatistics, the random function formalism is required for statistical inference. The observed values for a given variable and the unknown exhaustive values are assumed to be a realization of a random function that represents the phenomenon under study. Random functions are noted with capital letters, e.g. $Z(\boldsymbol{u})$, while their realizations are noted with lower case, e.g. $z(\boldsymbol{u})$.

2 Indicator Coding

2.1 Continuous case

Selecting the thresholds

The basic idea is to code the data as probability values [3]. This coding takes into account the rank ordering of the data. Different types of data can be coded with the same format, which gives a great flexibility to integrate data from different sources, at different supports and of different precision.

If the full conditional distribution needs to be recovered at every unsampled location, we need to use a set of thresholds to numerically discretize the cumulative distribution function of the variable.

If a single threshold is used and indicator kriging is used to estimate this indicator at unsampled locations, the resulting model will only tell us the probability of not exceeding *that particular threshold* at unsampled locations. This may

be useful if we want to assess if a contamination threshold or a cutoff grade has been exceeded.

In general, we will want to know the full conditional distribution at unsampled locations. For this, the first step is to select a number of thresholds z_k, $k = 1, \ldots, K$, to perform what is known as multiple indicator kriging (or MIK).

The choice of the number of thresholds is critical for good performance of this approach: too few thresholds imply a poor discretization of the ccdfs; a large number would reduce this problem, but larger computation and inference efforts would be needed and important order relations deviations are expected [2, 4]. Goovaerts recommends between 5 and 15 thresholds [5], Deutsch suggests a number between 7 and 11 [3]. A good practice is to match thresholds with critical values of the problem under study, and distribute them uniformly through the distribution, i.e. thresholds can be chosen at regular quantiles.

How many threshold should we use? It really depends on the amount of sample data available. Usually, the number of thresholds will be between 5 (for a very coarse representation of a cdf) to 20. In most cases, K will be close to 10. A good strategy is to define the thresholds using evenly distributed quantiles. For example, the deciles of the distribution can be used. If the variable has a long tail of high values, as is the case with many commodities in mineral resources, then the tail can be characterized by adding a few more thresholds, for example at the 95, 98 and 99 percentiles, to get a good idea of the probability of not exceeding these high values (these may actually be quite relevant to the project!).

Again, care should be taken to ensure we determine the thresholds as quantiles of the *representative distribution*, so

declustering is often the first step in the analysis.

Coding the data

Once the thresholds have been selected, the data are coded as:

$$i(\boldsymbol{u_\alpha}; z_k) = Prob\{z(\boldsymbol{u_\alpha}) \leq z_k\} \quad \forall k = 1, \ldots, K$$

Notice the change in notation. We replaced the subscript i for the location $\boldsymbol{u_i}$ by α to avoid confusion with the indicator variable I and its realization i.

At every data location, there is now a vector of K indicator values. If there were N data at the beginning, then there are $N \cdot K$ indicator values after coding the data.

Hard Data

Samples with negligible sample errors are called **hard data**. The coding for hard data is the most intuitive. We define the indicator at location $\boldsymbol{u_\alpha}$ as:

$$i(\boldsymbol{u_\alpha}; z_k) = \begin{cases} 1, & \text{if } z(\boldsymbol{u_\alpha}) \leq z_k \\ 0, & \text{otherwise} \end{cases} \quad k = 1, \ldots, K \qquad (1)$$

where $z(\boldsymbol{u_\alpha})$ is the value at the data location $\boldsymbol{u_\alpha}$. This can be interpreted as a probability:

$$i(\boldsymbol{u_\alpha}; z_k) = Prob\{z(\boldsymbol{u_\alpha}) \leq z_k\} = F_{\boldsymbol{u_\alpha}}(z_k)$$

That means that we assign a value of 1 for a given threshold z_k at the data location $\boldsymbol{u_\alpha}$ if the data value is lesser than or equal to the threshold, and we assign 0, otherwise. This is

the probability of the data value being lesser than or equal to z_k.

Consider the samples of a continuous variable displayed in **Figure 1 (left)**. The coding for 4 thresholds (0.5, 1.0, 1.5, 2.0) is displayed in the same figure **(right)**.

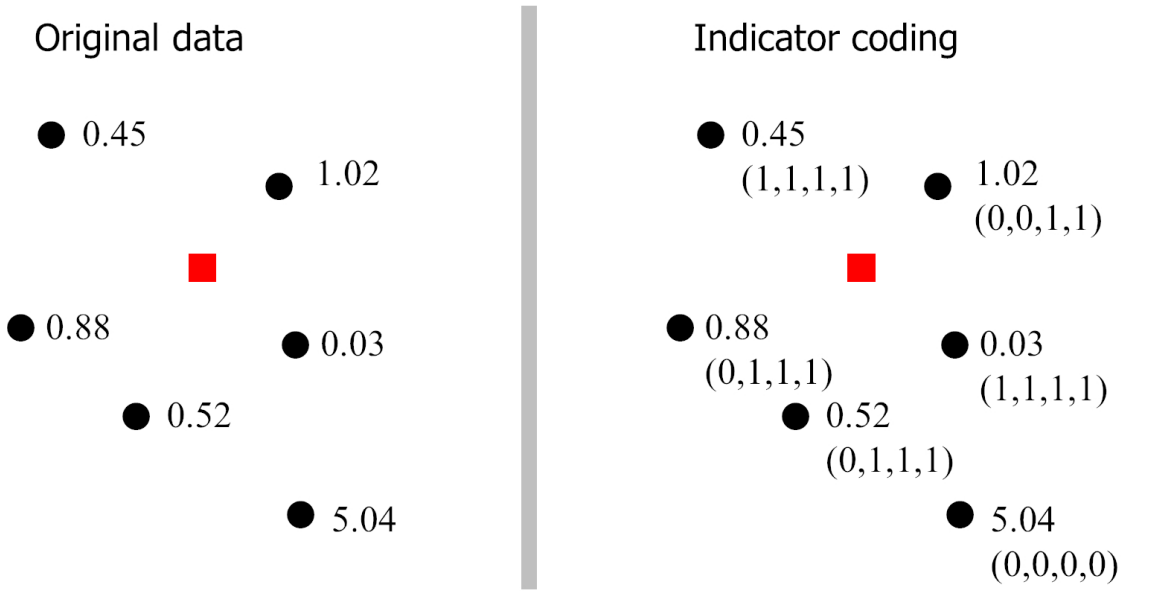

Figure 1: Example showing a continuous variable and some sample values (left) and the indicator coding for four thresholds (right).

2.2 Categorical case

In the case of categorical data, the variable can take one of K categories exhaustive and mutually exclusive.

There is of course, no need to determine thresholds, since we no longer have a continuous probability distribution function. We now have a probability mass function for K classes.

Coding the data

As with continuous variable, in the categorical case, the data are coded as probabilities:

$$i(\boldsymbol{u_\alpha}; s_k) = Prob\{s(\boldsymbol{u_\alpha}) = s_k\} \quad \forall k = 1, \ldots, K$$

Here, the categorical random variable is S and can take one of the K categories: $s_1, \ldots, s_K$.

At every data location, there is now a vector of K categorical indicator values.

Hard Data

The coding for hard data is:

$$i(\boldsymbol{u_\alpha}; s_k) = \begin{cases} 1, & \text{if } s(\boldsymbol{u_\alpha}) = s_k \\ 0, & \text{otherwise} \end{cases} \quad k = 1, \ldots, K \qquad (2)$$

where $s(\boldsymbol{u_\alpha})$ is the value at the data location $\boldsymbol{u_\alpha}$.

Notice that unlike in the case of continuous variables, here the vector of categorical indicators only has one value 1, and all the other values are 0. In the case of a continuous variable, the indicators are nested (due to the order relation between thresholds), thus it is possible to have many 1s in the vector of continuous indicators.

3 Indicator variograms

Indicator variograms are no different than the conventional variograms, but given the nature of the indicator variables, they have some particular features that we will now review.

3.1 Properties of an indicator variable

Indicators have some properties that can be useful to understand the variograms later on.

Expected value of an indicator

The expected value of an indicator variable is:

$$E\{I(\boldsymbol{u}; z_k)\} = F(z_k) = p_k \tag{3}$$

This means that the average value of an indicator is equal to the cumulative probability of the threshold used, which corresponds to the declustered cumulative probability for that threshold.

For a categorical data:

$$E\{I(\boldsymbol{u}; s_k)\} = m(s_k) = p_k \tag{4}$$

where $m(s_k)$ is the mass of category s_k.

Variance of an indicator

We can compute the variance of an indicator with the usual expression:

$$Var\{I(\boldsymbol{u}; z_k)\} = E\{I(\boldsymbol{u}; z_k)^2\} - (E\{I(\boldsymbol{u}; z_k)\})^2 \tag{5}$$

But we already know that $E\{I(\boldsymbol{u}; z_k)\} = p_k$. We should also notice that an indicator to any power has the same value as the original indicator. Therefore: $E\{I(\boldsymbol{u}; z_k)^2\} = E\{I(\boldsymbol{u}; z_k)\} = p_k$ (the indicator only can take values 0 or 1). Therefore, we can write:

$$Var\{I(\boldsymbol{u}; z_k)\} = p_k - p_k^2 = p_k(1-p_k) \tag{6}$$

This is valid for both continuous and categorical data.

3.2 Computing the variogram

Computing the variogram is done in the usual manner. Samples are searched separated by a given lag distance in a particular direction and tolerances are used to account for scattered data.

The computation of the variogram is done replacing the original variable by its indicator transform. In the continuous case:

$$\gamma(\boldsymbol{h}; z_k) = \frac{1}{2N(\boldsymbol{h})} \sum_{\alpha=1}^{N(\boldsymbol{h})} (i(\boldsymbol{u_\alpha}; z_k) - i(\boldsymbol{u_\alpha} + \boldsymbol{h}; z_k))^2 \tag{7}$$

In the categorical case:

$$\gamma(\boldsymbol{h}; s_k) = \frac{1}{2N(\boldsymbol{h})} \sum_{\alpha=1}^{N(\boldsymbol{h})} (i(\boldsymbol{u_\alpha}; s_k) - i(\boldsymbol{u_\alpha} + \boldsymbol{h}; s_k))^2 \tag{8}$$

Samples used to compute the variogram can only take values 0 or 1, therefore, the sill of the variogram is related to the variance of the indicator for that particular threshold or category. And since the variance of an indicator is linked to the proportion for the threshold or the category, the theoretical variance (the one that represents the sill of the variogram in the stationary case) is defined by: $\sigma_I^2 = p_k(1-p_k)$.

Interpretation and modeling follows the same rules as with conventional variograms. The only consideration is

that, for continuous variables, variograms should change smoothly from one threshold to the next, which in turn will help minimizing order relation problems in the results.

4 Multiple indicator kriging (MIK): continuous variables

The distribution of uncertainty of the regionalized variable can be inferred by kriging the indicator function at every threshold. Multiple indicator kriging is nothing more than Simple or Ordinary Kriging applied to each one of the indicator transforms defined by the thresholds. Each one of those sets of data can be used to estimate the value of the indicator at an unsampled location, i.e. the probability of having $z(\boldsymbol{u}) \leq z_k$. Notice that the indicator kriging variances are not relevant, as they represent errors over an inferred probability, and are not used in any further step.

This process leads to a discretization of the inferred ccdf, as depicted in **Figure 2**. Each point in the inferred discretized conditional distribution is the result of kriging at that threshold.

A short explanation of different techniques applied to indicators is presented hereafter.

4.1 Simple Indicator Kriging

To perform Simple Indicator Kriging, the stationary mean of each indicator random function is required. This mean is given directly by the cdf of the random function $Z(\boldsymbol{u})$:

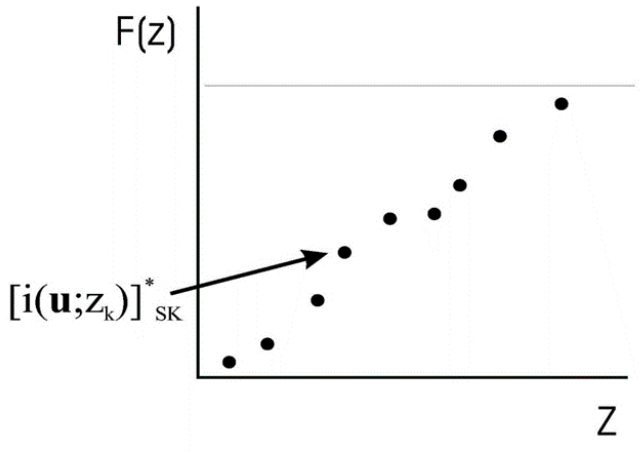

Figure 2: The discretized ccdf inferred by indicator kriging.

$$\begin{aligned} E\{I(\boldsymbol{u};z_k)\} &= 1 \cdot Prob\{Z(\boldsymbol{u}) \leq z_k\} + 0 \cdot Prob\{Z(\boldsymbol{u}) > z_k\} \\ &= Prob\{Z(\boldsymbol{u}) \leq z_k\} = F(z_k) \end{aligned}$$

The stationary Simple Kriging estimate of the indicator at that threshold is written:

$$\begin{aligned} [\,i(\boldsymbol{u};z_k)]^*_{SK} &= [\,Prob\{Z(\boldsymbol{u}) \leq z_k|(n)\}\,]^*_{SK} \\ &= \sum_{\alpha=1}^{n} \lambda_\alpha^{SK}(\boldsymbol{u};z_k) \cdot i(\boldsymbol{u}_\alpha;z_k) \\ &\quad + \left[1 - \sum_{\alpha=1}^{n} \lambda_\alpha^{SK}(\boldsymbol{u};z_k)\right] F(z_k) \end{aligned} \tag{9}$$

where the weights $\lambda_\alpha^{SK}(\boldsymbol{u};z_k)$ are the unique solution of the Simple Kriging system:

$$\sum_{\beta=1}^{n} \lambda_{\beta}^{SK}(\boldsymbol{u}; z_k) \cdot C_I(\boldsymbol{u}_\beta - \boldsymbol{u}_\alpha; z_k) = C_I(\boldsymbol{u} - \boldsymbol{u}_\alpha; z_k) \qquad (10)$$

$$\forall \alpha = 1, \ldots, n$$

Notice that an indicator covariance function $C_I(\boldsymbol{u}_\beta - \boldsymbol{u}_\alpha; z_k)$ or, assuming stationarity, $C_I(\boldsymbol{h}; z_k)$, has to be inferred for each threshold in equation **(10)**.

4.2 Ordinary Indicator Kriging

Ordinary Kriging differs from Simple Kriging in that the mean is unknown and therefore, the unbiasedness condition of the sum of the weights equal to one, is used.

The Ordinary Indicator Kriging estimate is written:

$$\begin{aligned} [\, i(\boldsymbol{u}; z_k)]^*_{OK} &= [\, Prob\{Z(\boldsymbol{u}) \leq z_k | (n)\}\,]^*_{OK} \\ &= \sum_{\alpha=1}^{n} \lambda_{\alpha}^{OK}(\boldsymbol{u}; z_k) \cdot i(\boldsymbol{u}_\alpha; z_k) \end{aligned} \qquad (11)$$

where the weights $\lambda_{\alpha}^{OK}(\boldsymbol{u}; z_k)$ are the unique solution of the Ordinary Kriging system **(12)**:

$$\sum_{\beta=1}^{n} \lambda_{\beta}^{OK}(\boldsymbol{u}; z_k) \cdot C_I(\boldsymbol{u}_\beta - \boldsymbol{u}_\alpha; z_k) + \mu_{OK}(\boldsymbol{u}; z_k) = C_I(\boldsymbol{u} - \boldsymbol{u}_\alpha; z_k)$$

$$\forall \alpha = 1, \ldots, n$$

$$\sum_{\beta=1}^{n} \lambda_{\beta}^{OK} = 1 \qquad (12)$$

Again, indicator covariances have to be inferred for each threshold.

4.3 Median Indicator Kriging

In Simple and Ordinary Indicator Kriging, K variogram or covariance functions must be modeled. In the continuous case, the inference of the variograms at extreme low or high thresholds is, in general, difficult since 0's and 1's are not in the same proportions, generating a noisier experimental variogram. For thresholds close to the median where the number of 0's and 1's is roughly the same, the inference of the variogram is easier.

Recall the variogram formula:

$$\gamma(\boldsymbol{h}; z_k) = \frac{1}{2N(\boldsymbol{h})} \sum_{\alpha=1}^{N(\boldsymbol{h})} (i(\boldsymbol{u_\alpha}; z_k) - i(\boldsymbol{u_\alpha} + \boldsymbol{h}; z_k))^2$$

The sum of squared differences increases only when one of the terms of the summation is 1 and the other is 0. If both are 0, then $(0-0)^2$ does not contribute to the sum, and if both are 1, the $(1-1)^2$ neither contributes. In summary, the indicator variogram reflects the transitions between threshold classes (**Figure 3**).

Median Indicator Kriging can be applied if the K indicator random functions $I(\boldsymbol{u}; z_k)$ are intrinsically correlated, i.e. all indicator variograms (and cross variograms) are proportional to a common variogram model, or equivalently, all correlograms are equal. This random function model is known as the mosaic model [7]:

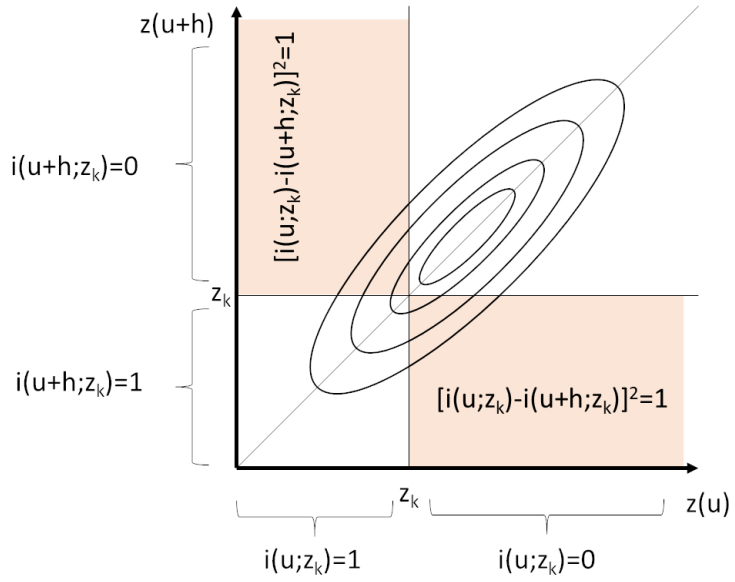

Figure 3: H-scatterplot of Z and the corresponding threshold and indicator coding. The indicator variogram captures the colored areas, where transitions occur across the threshold.

$$\rho_z(\boldsymbol{h}) = \rho_I(\boldsymbol{h}; z_k) = \rho_I(\boldsymbol{h}; z_k, z_{k'}), \forall z_k, z_{k'} \tag{13}$$

The single correlogram required can be estimated using the sample z correlogram or the sample indicator correlogram at the median cutoff $z_k = M$, where $F(M) = 0.5$. The advantage of using the experimental indicator correlogram is that there are no outliers, because of the transformation and therefore, the inference of the variogram (correlogram) function is easier.

If all indicators are defined for all data location, then at every location to be estimated (or simulated) only one kriging system must be solved for all thresholds. The weights will not change for a different cutoff since the data configu-

ration and variogram will remain the same.

4.4 Correcting for Order Relations Deviations

The estimated probabilities $[i(\boldsymbol{u}; z_k)]^*$, $k = 1, \ldots, K$ generated through indicator kriging must satisfy the conditions of a cumulative distribution: they have to be non-decreasing between 0 and 1 [3, 2, 5, 7].

The kriged indicator value can lie outside the interval [0,1] because the kriged estimate is a non-convex linear combination of the conditioning data and therefore these weights can be negative. Lack of data in some classes and differences in the variogram models from one threshold to the next are important factors to have a non-increasing function [2].

The *a posteriori* forward and downward correction of the ccdfs works well in general, as documented by Deutsch and Journel [2] (**Figure 4**). Although more difficult in its implementation, constraining the kriging system, so that it satisfies the order relations by construction is also a solution [5].

4.5 Interpolation and Extrapolation of the Conditional Cumulative Distribution Functions

Since the number of data is limited, the distribution of local uncertainty is discretized using only a few thresholds. The

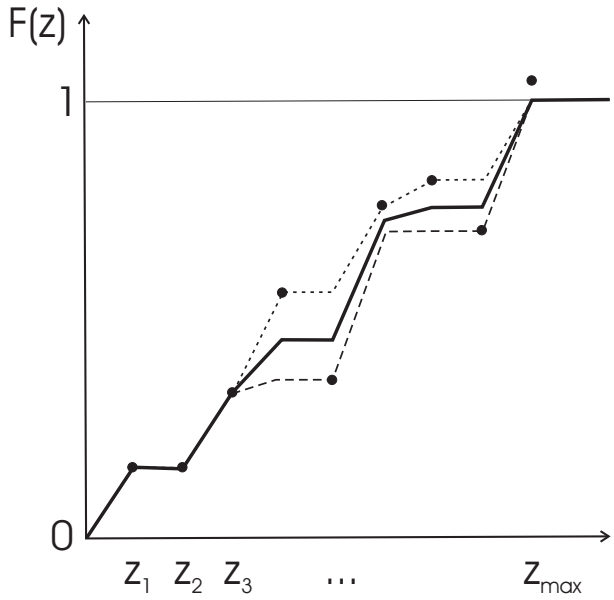

Figure 4: Forward and downward correction for order relation deviations. The model used is the one showed as a thick line.

continuous ccdf at every location $\boldsymbol{u}$ is then represented by a set of points $[i(\boldsymbol{u}; z_k)]^*$ with $k = 1, \ldots, K$, that lie in $[0, 1]$.

It is therefore necessary to interpolate the values between thresholds, and extrapolate the values beyond the smallest and largest values, z_1 and z_K. This decision has a large impact in the final statistics of the model being estimated or simulated, so it has to be analyzed carefully. The interpolation between thresholds is often the less important. It is commonly sufficient to interpolate linearly between the indicator values at thresholds z_{k-1} and z_k.

When extrapolating the tails, a minimum and maximum possible values should be considered and the extrapolation should not be done linearly, since this would imply a uniform distribution between the minimum value and z_1,and

between z_K and the maximum value, which is often not realistic. Power and hyperbolic models are used to extrapolate the ccdfs beyond the lower and higher indicator values. Another possibility is to consider the global cdf and scale it to extrapolate the tails of the ccdfs.

The different methods to interpolate and extrapolate are listed below:

- **Linear model**: Assuming a uniform distribution between the cdf for two thresholds, or between a lower limit and the first threshold or the higher threshold and an upper limit (maximum value), the cdf value is given by:

$$[F(z)]_{linear} = F^*(z_{k-1}) + \left[\frac{z - z_{k-1}}{z_k - z_{k-1}}\right] \cdot [F^*(z_k) - F^*(z_{k-1})]$$
$$\forall z \in (z_{k-1}, z_k]$$

- **Power model**: Depending on the value of the parameter w, the power model can take a wide range of shapes (**Figure 5**). The cdf is calculated as:

$$[F(z)]_{power} = F^*(z_{k-1}) + \left[\frac{z - z_{k-1}}{z_k - z_{k-1}}\right]^w \cdot [F^*(z_k) - F^*(z_{k-1})]$$
$$\forall z \in (z_{k-1}, z_k]$$

  It can be used to extrapolate the lower and upper tails of the cdf. This is done by replacing z_{k-1} and z_k by z_{min} and z_1, and using a power $w > 1$ for the lower tail, or replacing z_{k-1} and z_k by z_K and z_{max} and using a power $w < 1$ for the upper tail.

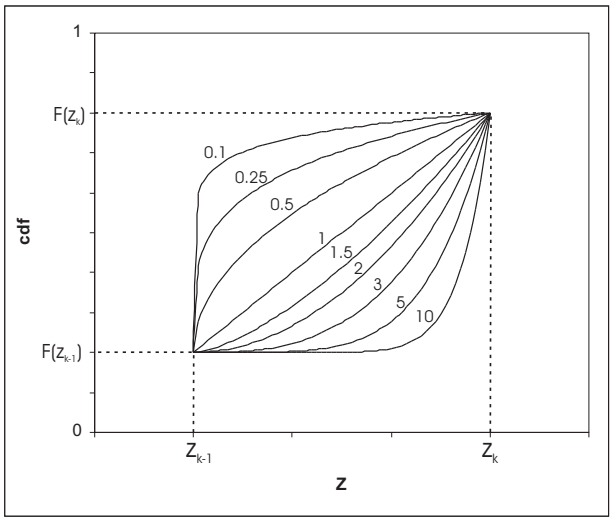

Figure 5: Power model for cdf interpolation and extrapolation, given different values of the parameter w.

- **Hyperbolic model**: This cdf model is useful to extrapolate the upper tail. As with the power model, the parameter w permits to control the shape of the function (**Figure 6**). The cdf is calculated as:

$$[F(z)]_{hyperbolic} = 1 - \frac{z_k^w \cdot [1 - F^*(z_k)]}{z^w} \qquad \forall z > z_k$$

- **Rescaling the global cdf**: This can be used to extrapolate the tails of the ccdf's. The ccdf's tails will have the same shape than those of the global distribution. However, a reliable global distribution is needed.

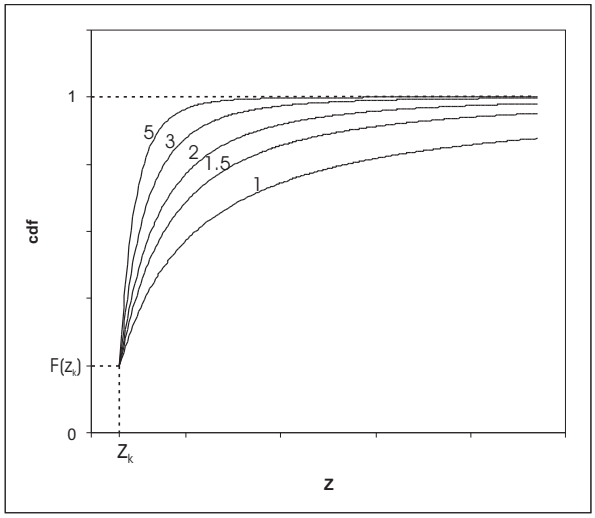

Figure 6: Hyperbolic model for cdf extrapolation, given different values of the parameter w.

4.6 Post-processing the conditional distribution

The conditional cumulative distribution function obtained by multiple indicator kriging (**Figure 7**) provides all the possible information at every unsampled location. From it, we can compute an estimate, determine the local variance, compute confidence intervals, etc.

All of these computations are done by numerical integration. They can be computed by calculating the quantiles of the distribution for a set of regularly distributed uniform values, representing the cumulated probability (**Figure 8**). Notice that where the cdf is steeper, this leads to closer quantiles. The same can be achieved by performing Monte Carlo simulation with a large enough number of points (**Figure 9**): a uniform random number in $(0, 1)$ is generated and the cor-

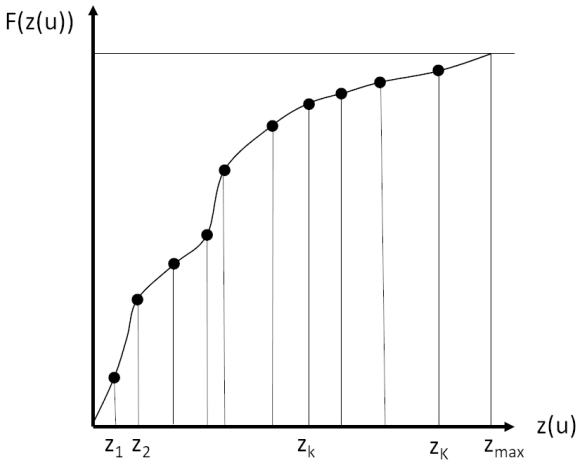

Figure 7: The inferred local distribution from multiple indicator kriging.

responding quaqntile is read. This is repeated many times until statistics converge to the precision required.

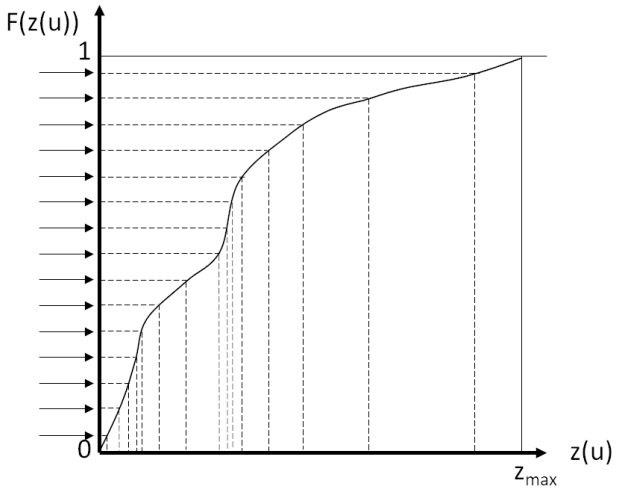

Figure 8: Numerical integration by uniformly sampling the distribution.

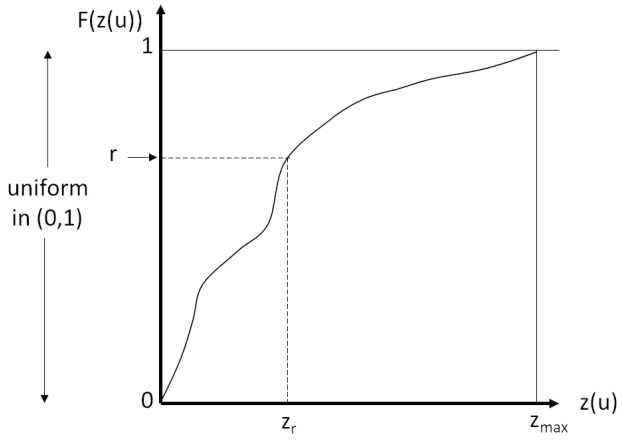

Figure 9: Numerical integration by Monte Carlo simulation.

5 Categorical variables: indicator kriging

In the case of categorical variables, the kriging estimates for each indicator at a particular unsampled location let us build the conditional probability mass function. “Predicting” the value at that location can be done, but no single solution to the problem exist and the best approach will depend on requirements such as if the proportions in the predicted values are to be honored.

It is easy to see that if the most frequent class is selected in each node, then categories with small probabilities will tend to disappear and will end up underrepresented in the final model.

References

[1] M. David and D. M. abd M. Soulié. Conditional bias in kriging and a suggested correction. In G. Verly, M. David, A. G. Journel, and A. Marechal, editors, *Geostatistics for natural resources characterization*, volume 1, pages 217–230. Reidel, Dordrecht, Holland, 1984.

[2] C. V. Deutsch. Direct assessment of local accuracy and precision. In E. Y. Baafi and N. A. Schofield, editors, *Geostatistics Wollongong '96*, volume 1, pages 115–125. Kluwer, 1997.

[3] C. V. Deutsch and X. H. Wen. Integrating large-scale soft data by simulated annealing and probability constraints. *Mathematical Geology*, 32(1):49–68, 2001.

[4] P. Goovaerts. Comparative performance of indicator algorithms for modeling conditional probability distribution functions. *Mathematical Geology*, 26(3):385–410, 1994.

[5] P. Goovaerts. Accounting for local uncertainty in environmental decision-making processes. In E. Y. Baafi and N. A. Schofield, editors, *Geostatistics Wollongong '96*, volume 2, pages 929–940. Kluwer, 1997.

[6] A. G. Journel. Nonparametric estimation of spatial distribution. *Mathematical Geology*, 15(3):445–468, 1983.

[7] A. G. Journel. mad and conditional quantile estimators. In G. Verly, M. David, A. G. Journel, and A. Marechal, editors, *Geostatistics for natural resources characterization*, volume 1, pages 261–270. Reidel, Dordrecht, Holland, 1984.

[8] A. G. Journel and F. Alabert. Non-Gaussian data expansion in the Earth Sciences. *Terra Nova*, 1:123–134, 1989.

[9] I. C. Lemmer. Estimating local recoverable reserves via indicator kriging. In G. Verly, M. David, A. G. Journel, and A. Marechal, editors, *Geostatistics for natural resources characterization*, volume 1, pages 349–364. Reidel, Dordrecht, Holland, 1984.

[10] B. Oz and C. V. Deutsch. Size scaling of cross-correlation between multiple variables. In *Centre For Computational Geostatistics*, volume 3, Edmonton, AB, 2001.

[11] J. Sullivan. Conditional recovery estimation through probability kriging: theory and practice. In G. Verly, M. David, A. G. Journel, and A. Marechal, editors, *Geostatistics for natural resources characterization*, volume 1, pages 365–384. Reidel, Dordrecht, Holland, 1984.

Index