Determining the Shape of Conditional Distributions with Hermite Polynomials and Disjunctive Kriging

Bora Oz, Julián Ortiz C., and Clayton V. Deutsch
Centre for Computational Geostatistics, University of Alberta

Abstract

Geostatistical simulation techniques are being increasingly used in industry. The common Gaussian simulation has some restrictions. Performing simulation in the original data units has the advantage of integrating different data sources for better characterization. The problem of obtaining the local ccdf shape remains largely unsolved except by brute force calculations. In this study, we explore an analytical approach to calculate the local ccdf shape to make calculations faster. Hermite polynomials and disjunctive Kriging are used.

Introduction

Let $Z(\mathbf{u})$ be the random variable modeling the the uncertainty about $z(\mathbf{u})$, where $\mathbf{u}$ corresponds to a location coordinates vector. The distribution function $F(\mathbf{u}; z|(n)) = Prob\{Z(\mathbf{u}) \leq z|(n)\}$ is conditional to the available information (n) and models that uncertainty. Each conditional cumulative distribution function (ccdf) provides a measure of *local* uncertainty related to a specific location $\mathbf{u}$ and level of conditioning data (n). The model of local uncertainty may be post-processed to retrieve specific values such as: (1) the probability of exceeding a critical threshold z_c, and (2) an estimate of the unknown value $z(\mathbf{u})$ that is optimal for a given criterion, which need not be least-squares.

One may also draw a series of L simulated values $z^{(l)}(\mathbf{u}), l = 1, \ldots, L$ from the local distribution. Each value $z^{(l)}(\mathbf{u})$ represents a possible outcome or realization of the random variable $Z(\mathbf{u})$ modeling the uncertainty at location $\mathbf{u}$.

A series of single-point ccdfs do not provide a measure of multiple-point or *spatial* uncertainty, such as the probability that a block average exceed a given threshold value. The concept of conditional simulation allows the assessment of such spatial uncertainty from several realizations.

Sequential simulation is commonly used. Monte Carlo simulation from one-point distributions of uncertainty proceeds sequentially from distributions with increasing levels of conditioning. Each distribution is conditional to the original data and all previously simulated values. For continuous variables, simple (co)kriging is used at each step to give the mean and variance of the ccdf, $F(\mathbf{u}; z|(n))$. Under the multivariate Gaussian model each distribution is Gaussian in shape and the mean and variance are sufficient to establish the full distribution.

The multivariate Gaussian random function model, which is the heart of sequential Gaussian simulation, is by far the most widely used model because of its extremely congenial properties (Deutsch and Journel, 1992; Isaaks, 1989). The (co)kriging system provides an

estimate and an estimation variance. These two parameters fully characterize the local distributions of uncertainity, since the local distributions are Gaussian in shape. However, there is significant motivation to avoid the Gaussian transformation required for sequential Gaussian simulation. First, for a given covariance, the Gaussian random function (RF) has maximum "disconnectedness" of extreme values; a property known as maximum entropy. Multivariate Gaussianity also entails that the pattern of spatial correlation is symmetric with respect to the median, that is, there is symmetric destructuration of extreme values. Third, transformation of the data variable to a Gaussian distribution is problematic when dealing with data of different scale. Most variables average linearly (lithofacies proportions and volumetric proportions such as porosity and mineral grades) or with very particular known scaling laws (permeability). The non-linear transformation to a Gaussian variable means that the correct averaging in "Gaussian space" is complex and intractable. The Gaussian transformation must be avoided to permit rigorous multiscale data integration.

Working in "direct" space is desirable; this is the "Direct Sequential Simulation" approach, which does not require Gaussianity of the ccdf models. The notion of direct sequential simulaton was developed at the same time as sequential Gaussian simulation (Journel 1986). It was shown early in the development of sequential techniques that the variogram (covariance) structure and the global mean can be reproduced without transformation to Gaussian space provided that the simulated values are drawn from local conditional distributions centered at the simple (co)kriging estimates with a variance corresponding to the simple (co)kriging estimation variance. The conditional distribution could be of any shape. Exercising this freedom, however, leads to simulated realizations where the univariate histogram is not controlled and therefore not reproduced.

The recent studies of Deutsch et. al. (2000) and Oz et. al. (2001) establish a framework to ensure that the global histogram is reproduced. This methodology relies on the link between direct space (Z space) and Gaussian space (Y space) through the global distributions. The proposed "graphical approach" has been shown to work in practice. An extensive table look up must be constructed. The motivation of this paper is to consider and document an alternative procedure based on Hermite polynomials and disjunctive kriging (DK) to build the ccdf shape models

Graphical Approach

The *correct* shape of the local distributions is known for the Gaussian case because we have a model for the full multivariate distribution. The central idea of this proposal is to infer the correct shape of the local distributions that permits the simulated values, taken all together, to reproduce the global histogram.

The original Z variable with stationary histogram $F_Z(z)$ can be transformed to a Y variable with stationary standard normal distribution $G(y)$. The quantile or normal-score transformation is widely used for such transformation.

$$y = G^{-1}(F_Z(z)) \tag{1}$$

This transformation can be reversed at any time to get back to the original variable units:

$$z = F_Z^{-1}(G(y)) \tag{2}$$

The cumulative distribution functions $F_Z(z)$ and $G(y)$ and their inverse relations or quantile functions $F_Z^{-1}(z)$ and $G^{-1}(y)$ are known. Thus, we have a direct link between Z and Y units. This transformation is unique, reversible, and almost always non-linear.

Valid distribution shapes in Z space can be determined from back transformation of non-standard Gaussian distributions. The back transformation of the non-standard p^l probability data:

$$z^l = F_Z^{-1}[G(G^{-1}(p^l \cdot \sigma_k + y^*)], \qquad l = 1, \ldots, L \tag{3}$$

where y^* and σ_k are the mean and standard deviation of the non-standard Gaussian distribution of uncertainty, and the $p^l, l = 1, ..., L$ values are uniformly distributed between 0 and 1. The distribution of uncertainty in Z space is assembled from the $z^l, l = 1, ..., L$ values. There is no analytical equation for this distribution, aside from Equation 3; nevertheless, the distribution is completely defined from $y^*, \sigma_k, F(z)$, and $G(y)$.

The shape of the z-conditional distributions are neither Gaussian nor identical to the original Z data distribution. The shape of every z-conditional distribution is explicitly known and the direct sequential simulation with histogram reproduction program, `DSSIM-HR` (Oz et. al., 2001), uses these shapes to reproduce the global input histogram within statistical fluctuations.

A lookup table of distributions corresponding to non-standard Gaussian distributions is constructed before simulation starts. The lookup table or database of local distributions is constructed with different Gaussian means (from approximately -3.5 to 3.5) and variances (from 0 to 2). The Z mean and variance corresponding to each distribution is calculated and saved so that the right distribution can be retrieved during the sequential simulation procedure.

The distribution with the closest mean and variance to the simple (co)kriging mean, z^*, and variance, σ_Z^2 must be found in the database. It is unlikely that the distribution will have the exact mean and variance; therefore, we can either (1) interpolate in the lookup table, or (2) rescale slightly the closest distribution to have exactly the right mean and variance.

The variogram is calculated from the original data with no normal or Gaussian transformation. The histogram and variogram all come from original data units. The Gaussian transform is only used to help in determining correct *shapes* for conditional distributions.

This approach will create realizations that reproduce the (1) local point and block data in the original Z data units, (2) the mean, variance, and variogram of the Z variable, and (3) the histogram of the Z variable. More details of this approach including a theoretical justification on how the input global histogram is reproduced is given by Deutsch et. al. (2001).

Example Applications of Graphical Technique

Using the *graphical* technique, some conditional distributions for global lognormal, bimodal and uniform distributions are shown in Figures 1, 2 and 3. The shapes of the non-standard distributions are significantly different than either the widely assumed Gaussian distribution or the original global distributions. Different Gaussian means, -1.0, 0.0, 1.0, and variance values, 0.1, 0.5, and 1.0 are considered. The local uncertainity distributions are defined by Equation 3. The boxed histograms correspond to a Gaussian mean and variance of 0.0 and

1.0, which gives the original distribution. The shape of the local distributions are unique and do not look like the input distributions. The skewness, shape, and other characteristics do not change in a simple manner.

This graphical approach is simple and effective; however, there may be significant advantages if an analytical approach could be used to quickly establish the conditional distributions. The analytical result using Hermite polynomials and disjunctive kriging (DK) will be the same because both methods rely on the multivariate Gaussian model. The use of other orthogonal polynomials such as the Legendre or Laquerre polynomials would result in other distribution shapes that would also permit reproduction of the global histogram. This interesting theoretical avenue has not been explored in this work.

We admit that reproducing the graphical approach with an analytical method may be of little practical value. Neverthless, there is a hope that the Hermite polynomials/DK approach could be used to directly infer $F_z(z)$ shape without going through y^* and σ_y^2.

Hermite Polynomials

Let $z(\mathbf{u}_\alpha), \alpha = 1, \ldots, N$ be the data values with $y(\mathbf{u}_\alpha)$ their Gaussian (normal) transformed values, with $\mathbf{u}_\alpha$ being the location coordinates vector for the α data. We can calculate an anamorphosis function (normal score transformation) $y(\mathbf{u}) = \varphi(z(\mathbf{u}))$ and establish a one-to-one relationship between z and y, see back to Equations 1 and 2.

Hermite polynomials $H_n(y)$ are polynomials that have special properties related to the normal distribution (Rivoirard, 1994). They are defined by Rodrigues's formula ($n \geq 0$):

$$H_n(y) = \frac{1}{\sqrt{n!}\ g(y)} \frac{d^n g(y)}{dy^n} \quad n = 0, 1, \ldots, N \tag{4}$$

where $\sqrt{n!}$ is a normalization factor and $g(y)$ is the standard Gaussian probability distribution function (pdf) defined by

$$g(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \tag{5}$$

The n^{th} Hermite polynomial $H_n(y)$ is a polynomial of degree n. More specifically, $H_0(y) = 1$, $H_1(y) = -y$, and $H_2(y) = \frac{1}{\sqrt{2}}(y^2 - 1)$. The other polynomials can be calculated using the recurrence relation ($n > 0$):

$$H_{n+1}(y) = \frac{1}{\sqrt{n+1}} y H_n(y) - \sqrt{\frac{n}{n+1}} H_{n-1}(y) \tag{6}$$

Except for $H_0(y)$ which is constant and equal to 1, their means are:

$$E\{H_n(y)\} = \int H_n(y) g(y) dy = 0 \tag{7}$$

and, due to the normalization factor, their variances are:

$$Var\{H_n(y)\} = E\{[H_n(y)]^2\} = 1 \tag{8}$$

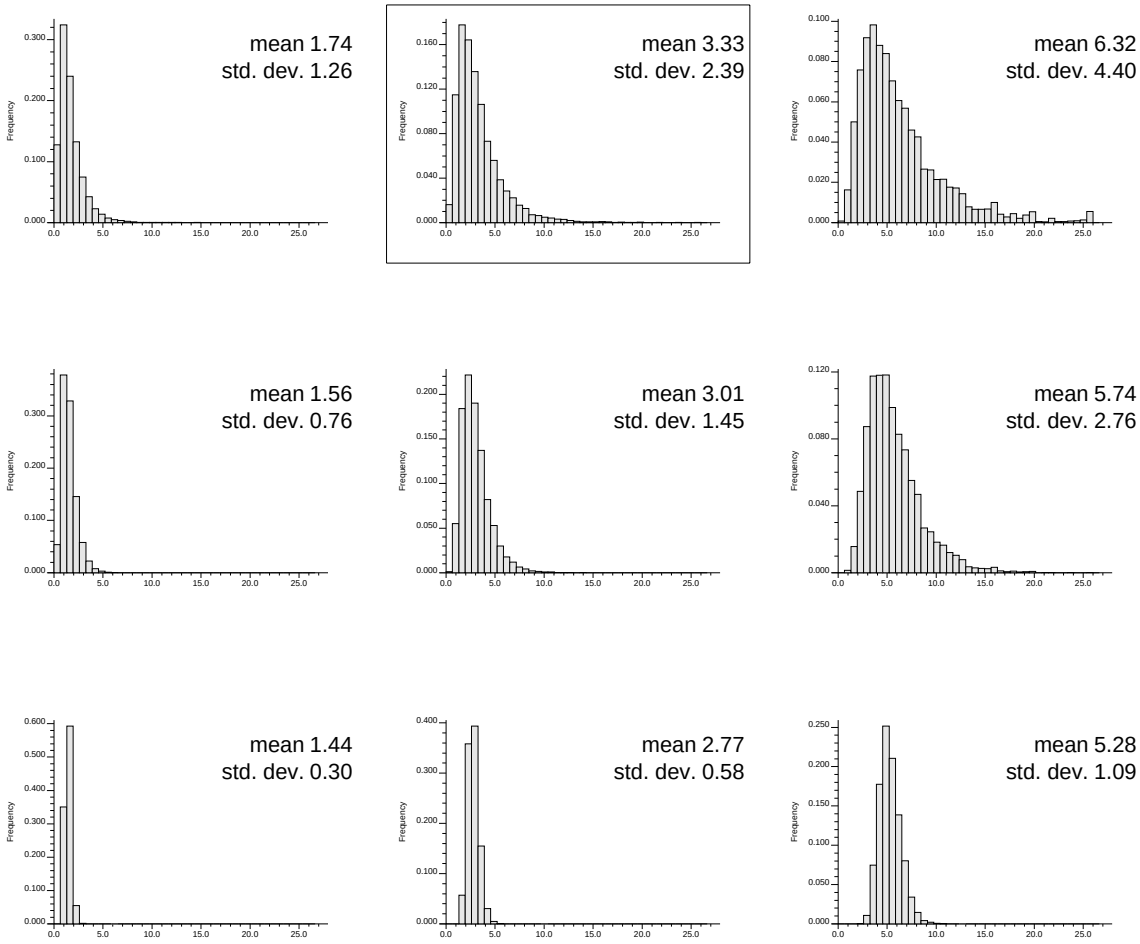

Figure 1: Non-standard distributions for lognormal data set (from left to right: mean in normal space: -1.0/0.0/1.5; from top to bottom: variance in normal space: 1.0/0.5/0.1).

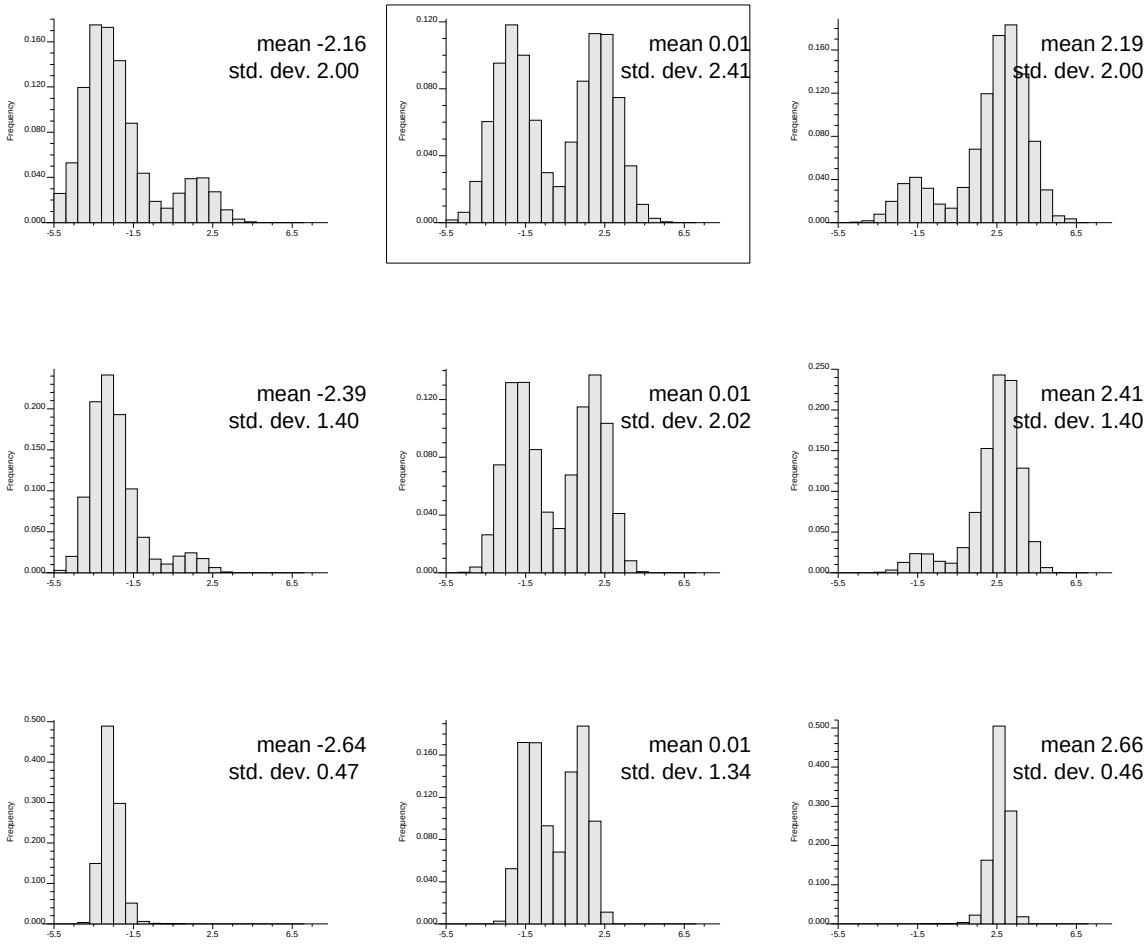

Figure 2: Non-standard distributions for bimodal data set (from left to right: mean in normal space: -1.0/0.0/1.5; from top to bottom: variance in normal space: 1.0/0.5/0.1).

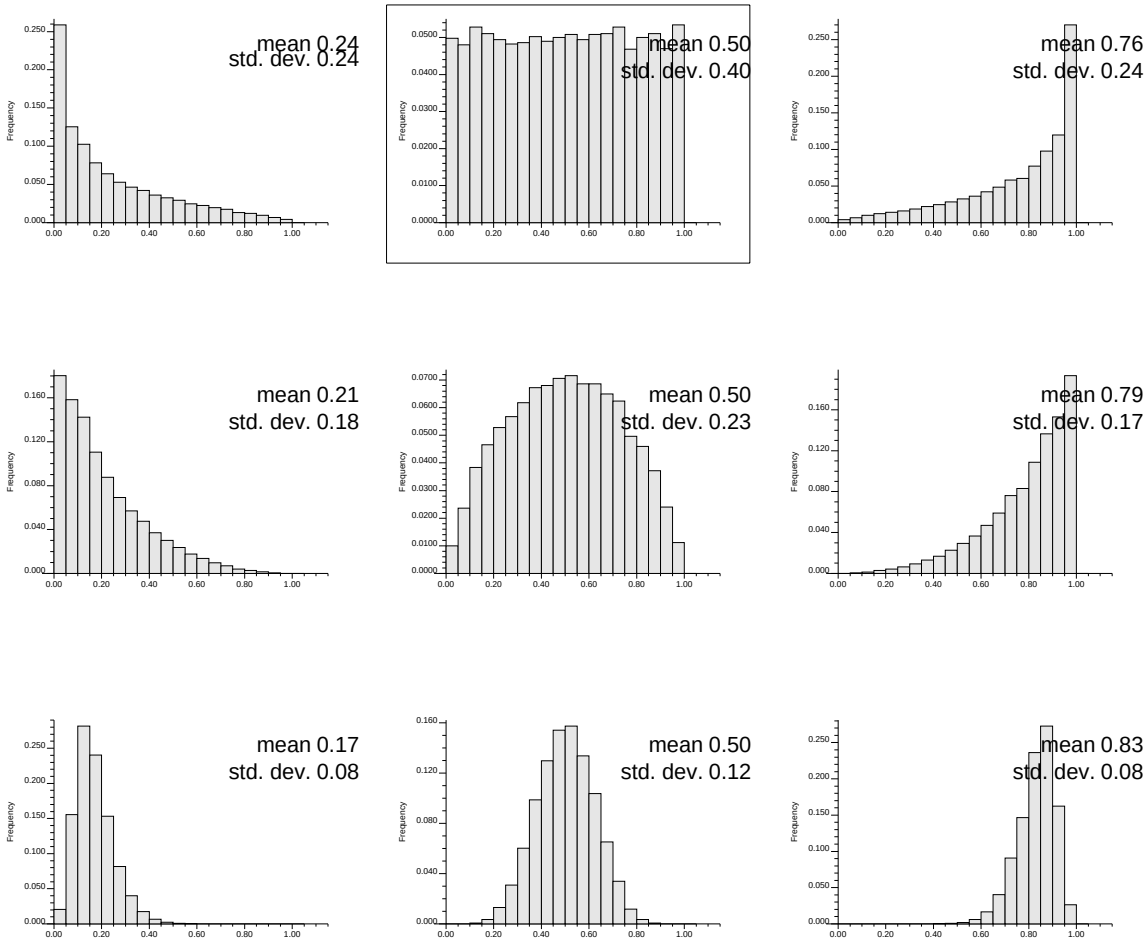

Figure 3: Non-standard distributions for uniform data set (from left to right: mean in normal space: -1.0/0.0/1.5; from top to bottom: variance in normal space: 1.0/0.5/0.1).

Structure of Hermite polynomials

When pairs of values, say $(y(\mathbf{u}),\ y(\mathbf{u}+\mathbf{h}))$ are bivariate standard normal with correlation coefficient $\rho(\mathbf{h})$, the Hermite polynomials have the following property:

$$E\{H_n(y(\mathbf{u}+\mathbf{h}))|y(\mathbf{u})\} = [\rho(\mathbf{h})]^n H_n(y(\mathbf{u})) \tag{9}$$

the spatial covariance between the $p-$ order polynomial of $y(\mathbf{u})$ and the $n-$ order polynomial of $y(\mathbf{u}+\mathbf{h})$ can be written as,

$$\begin{aligned} Cov\{H_p(y(\mathbf{u})), H_n(y(\mathbf{u}+\mathbf{h}))\} &= E\{H_p(y(\mathbf{u}))H_n(y(\mathbf{u}+\mathbf{h}))\} \\ &= E\{H_p(y(\mathbf{u}))E\{H_n(y(\mathbf{u}+\mathbf{h}))|y(\mathbf{u})\}\} \\ &= E\{H_p(y(\mathbf{u}))[\rho(\mathbf{h})]^n H_n(y(\mathbf{u}))\} \\ &= [\rho(\mathbf{h})]^n E\{H_p(y(\mathbf{u}))H_n(y(\mathbf{u}))\} \end{aligned} \tag{10}$$

When $p = n$, and also remembering Equation 8, Equation 10 results in

$$Cov\{H_n(y(\mathbf{u})), H_n(y(\mathbf{u}+\mathbf{h}))\} = [\rho(\mathbf{h})]^n \tag{11}$$

The spatial covariance of $H_n(y(\mathbf{u}))$ is equal to $[\rho(\mathbf{h})]^n$; that is, the covariance of $y(\mathbf{u})$ raised to the n^{th} power. Since $\rho(\mathbf{h}) < \rho(0) = 1$, the spatial dependence of $H_n(y(\mathbf{u}))$ decreases rapidly to nothing as n increases, i.e. the structure tends to pure nugget.

For $p \neq n$ the Hermite polynomials for a bivariate standard normal pair are orthogonal

$$Cov\{H_p(y(\mathbf{u})), H_n(y(\mathbf{u}+\mathbf{h}))\} = E\{H_p(y(\mathbf{u}))H_n(y(\mathbf{u}+\mathbf{h}))\} = 0 \tag{12}$$

Equation 12 also reveals that there is no spatial correlation between polynomials of different orders.

Expressing functions in terms of Hermite polynomials

Virtually any function of $f(y(\mathbf{u}))$ can be expanded in terms of Hermite polynomials:

$$f(y(\mathbf{u})) = f_0 + f_1 H_1(y(\mathbf{u})) + f_2 H_2(y(\mathbf{u})) + \cdots \tag{13}$$

$$= \sum_{n=0}^{\infty} f_n H_n(y(\mathbf{u})) \tag{14}$$

To calculate the coefficient f_p consider;

$$E\{f(y(\mathbf{u}))H_n(y(\mathbf{u}))\} = E\{\sum_{p=0}^{\infty} f_p H_p(y(\mathbf{u}))H_n(y(\mathbf{u}))\} \tag{15}$$

By taking the expected value operator inside of summation term and taking the f_p term outside of all summation (since f_p is not function of y), Equation 15 can be written as;

$$= \sum_{p=0}^{\infty} f_p E\{H_p(y(\mathbf{u}))H_n(y(\mathbf{u}))\} \tag{16}$$

Since the expected value term in the summation ends up 0 for $n \neq p$, Equation 15 results in;

$$E\{f(y(\mathbf{u}))H_n(y(\mathbf{u}))\} = f_p \tag{17}$$

Using the definition of expected value operator, Equation 17 can be rewritten as;

$$f_n = E\{f(Y(\mathbf{u}))H_n(Y(\mathbf{u}))\} = \int f(y)H_n(y)g(y)dy \tag{18}$$

for any given function f. The variance of $f(y(\mathbf{u}))$ is given by

$$Var\{f(y(\mathbf{u}))\} = E\{(f(y(\mathbf{u})) - E\{f(y(\mathbf{u}))\})^2\} = E\{\sum_{n=1}^{\infty}[f_n H_n(y(\mathbf{u}))]^2\} \tag{19}$$

By writing the square terms in Equation 19 explicitly;

$$= E\{\sum_{n=0}^{\infty}\sum_{p=0}^{\infty} f_n f_p H_n(y(\mathbf{u}))H_p(y(\mathbf{u}))\} \tag{20}$$

By taking the expected value operator inside of summation term and taking the f_n and f_p terms outside of all summation (since both f_n and f_p are not function of y), Equation 20 can be rewritten as;

$$= \sum_{n=0}^{\infty}\sum_{p=0}^{\infty} f_n f_p E\{H_n(y(\mathbf{u}))H_p(y(\mathbf{u}))\} \tag{21}$$

Since, $E\{H_n(y(\mathbf{u}))H_p(y(\mathbf{u}))\} = Cov\{H_n(y(\mathbf{u})), H_p(y(\mathbf{u}))\}$, Equation 21 can be rewritten as;

$$= \sum_{n=0}^{\infty}\sum_{p=0}^{\infty} f_n f_p Cov\{H_n(y(\mathbf{u})), H_p(y(\mathbf{u}))\} \tag{22}$$

The covariance term inside the summation operator is only defined for $p = n$; otherwise, it is 0. When it is not zero, for $h = 0$, it is equal to $Cov\{H_n(y(\mathbf{u})), H_n(y(\mathbf{u}))\} = [\rho(\mathbf{0})]^n = 1$, then we can rewrite Equation 22;

$$Var\{f(y(\mathbf{u}))\} = \sum_{n=1}^{\infty}(f_n)^2 \tag{23}$$

Equation 18 is a general expression to obtain the Hermite coefficients from the Hermite polynomials. Once we have the coefficients, then in the coming sections, we will show how to fit a continuous Global distribution and an indicator function. Both this section and the coming sections will be the main basis for producing local conditional distributions from the given Gaussian mean and variance.

Fitting a Global distribution

In this section we will show how to fit a continuous distribution. Equation 14 is the key equation for this purpose. There are two terms in Equation 14 that need to be calculated and the sum of their product ($f_n H_n(y(\mathbf{u}))$), for a sufficient number of polynomials, would delineate the complete distribution. The Hermite polynomials, $H_n(y(\mathbf{u}))$, in Equation 14 are calculated using Equation 6. Then, the Hermite coefficients, f_n, are described by Equation 17.

In general, the first Hermite coefficient corresponds to the mean value in the original data space;

$$f_0 = E\{f(y)\} = E\{\varphi(y)\} = E\{Z\} \tag{24}$$

Then the resultant coefficients are calculated by rewriting the Equation 18 as:

$$f_n = E\{f(y)H_n(y)\} = \sum_{i=1}^{N} \int_{y_i}^{y_{i+1}} z_i H_n(y) g(y) dy \tag{25}$$

$$= \sum_{i=1}^{N} z_i [\frac{1}{\sqrt{n}} H_{n-1}(y_{i+1}) g(y_{i+1}) - \frac{1}{\sqrt{n}} H_{n-1}(y_i) g(y_i)] \tag{26}$$

Let's write a few terms of the Equation 26 in order to group some of the similar terms;

$$f_n = z_1 [\frac{1}{\sqrt{n}} H_{n-1}(y_2) g(y_2) - \frac{1}{\sqrt{n}} H_{n-1}(y_1) g(y_1)] + z_2 [\frac{1}{\sqrt{n}} H_{n-1}(y_3) g(y_3) -$$

$$\frac{1}{\sqrt{n}} H_{n-1}(y_2) g(y_2)] + \ldots + z_{N-1} [\frac{1}{\sqrt{n}} H_{n-1}(y_N) g(y_N) -$$

$$\frac{1}{\sqrt{n}} H_{n-1}(y_{N-1}) g(y_{N-1})] + z_N [\frac{1}{\sqrt{n}} H_{n-1}(y_{N+1}) g(y_{N+1}) - \frac{1}{\sqrt{n}} H_{n-1}(y_N) g(y_N)]$$

Since $g(y_1) = g(-\infty) = 0$ and $g(y_{N+1}) = g(\infty) = 0$, we can regroup the terms as;

$$f_n = \sum_{i=2}^{N} (z_{i-1} - z_i) \frac{1}{\sqrt{n}} H_{n-1}(y_i) g(y_i) \tag{27}$$

Once one transforms the original variable to Gaussian space, the corresponding Hermite polynomials, Gaussian pdf and the original variable z_i would be used to fit a continuous function.

Fitting an Indicator function

In the previous section, we showed how to fit to a continuous distribution. Now, we will extend this fitting to an indicator function. The probability of $y < y_c$, with y_c being a cutoff value, is equivalent to $z = \varphi^{-1}(y) < \varphi^{-1}(y_c)$ and by choosing $z_c = \varphi^{-1}(y_c)$, then calculating $I_{z<z_c}$ turns into calculating:

$$I_{z<z_c} = I_{y<y_c} = \sum_{n=0}^{\infty} f_n H_n(y) \tag{28}$$

The coefficient of order n of the indicator function $I_{y<y_c}$ is written:

$$\begin{aligned} f_n &= E\{f(y)H_n(y)\} \\ &= \int_{-\infty}^{\infty} I_{y<y_c} H_n(y)g(y)dy = \int_{-\infty}^{y_c} H_n(y)g(y)dy \end{aligned} \tag{29}$$

for $n = 0$,

$$f_0 = G(y_c) \tag{30}$$

and for $n \geq 1$, by inserting Equation 5 into Equation 29, we get:

$$f_p = \int_{-\infty}^{y_c} \frac{1}{\sqrt{n!}\ g(y)} \frac{d^n g(y)}{dy^n} g(y)d(y) \tag{31}$$

the two $g(y)$ are cancelled and the integration of $\frac{d^n g(y)}{dy^n}$ results in $\frac{d^{n-1} g(y)}{dy^{n-1}}$, we can simplfy Equation 31 as

$$f_n = \frac{1}{\sqrt{n!}}[\frac{d^{n-1} g(y)}{dy^{n-1}}]_{y=y_c} = \frac{1}{\sqrt{n!}} H_{n-1}(y_c)\sqrt{(n-1)!}\ g(y_c) \tag{32}$$

$$f_n = \frac{1}{\sqrt{n}} H_{n-1}(y_c)g(y_c) \tag{33}$$

Then, inserting Equation 30 and 33 into Equation 28;

$$I_{Y(\mathbf{u})<y_c} = G(y_c) + \sum_{n=1}^{\infty} H_{n-1}(y_c)g(y_c)H_n(Y(\mathbf{u})) \tag{34}$$

where $G(y)$ is the Gaussian cumulative density function and $g(y)$ is the Gaussian probability distribution function. So, any indicator of $f(y)$, can be expanded in terms of generating the $H_n(y)$ family. It is important to note that the developed Hermite coefficients in this section are not the same as the ones developed for the continuous function case.

Application to Uniform, Bimodal and Lognormal Distributions

We demonstrate fitting continuous distributions. The original distributions can be fitted with increasing accuracy with more polynomials. The general procedure of fitting to a continuous function is:

- Tranform the original variable $z_i, i = 1, \ldots, N$ a into Gaussian space

$$y_i = G^{-1}(F(z_i)), i = 1, \ldots, N)$$

- Using Equation 6, calculate the successive Hermite polynomials for all the Gaussian transformed y_i data values

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

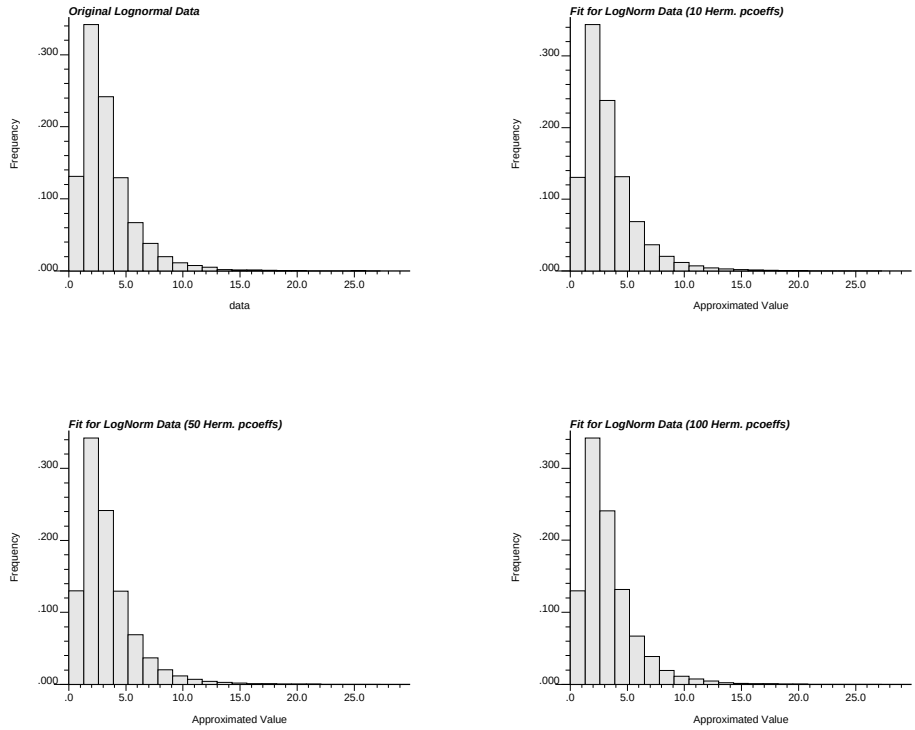

Figure 4: Original *lognormal* distribution (top left) and corresponding reproduction of *lognormal* distribution with 10, 50 and 100 Hermite polynomials

- Calculate the Hermite coefficients using Equations 24 and 27.

$$f_0 = E\{\varphi(y)\} = E\{Z\}$$

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

- Reproduce the original distribution by using those calculated Hermite polynomials and the coefficients in Equation 14.

$$f(y(\mathbf{u})) = \sum_{n=0}^{\infty} f_n H_n(y(\mathbf{u}))$$

Consider 10000 values drawn from a uniform distribution, see Figure 4. Then using 10, 50 and 100 polynomials, the original data distribution has been reconstructed and presented in Figure 4. It is clear that as the number of polynomials gets larger, we get better fit of the original distribution. 25 polynomials are enough to reproduce the *uniform* distribution. The *bimodal* distribution was also fit, see Figure 5. 500 polynomials were used because of difficulty in fitting bimodal data. The *lognormal* distribution was tested with 10, 50 and 100 polynomials, see Figure 6.

Disjunctive Kriging Paradigm

Disjunctive kriging (DK) provides an estimate for the value of any known function of a single unknown as a linear combination of functions of data values, with the usual criteria

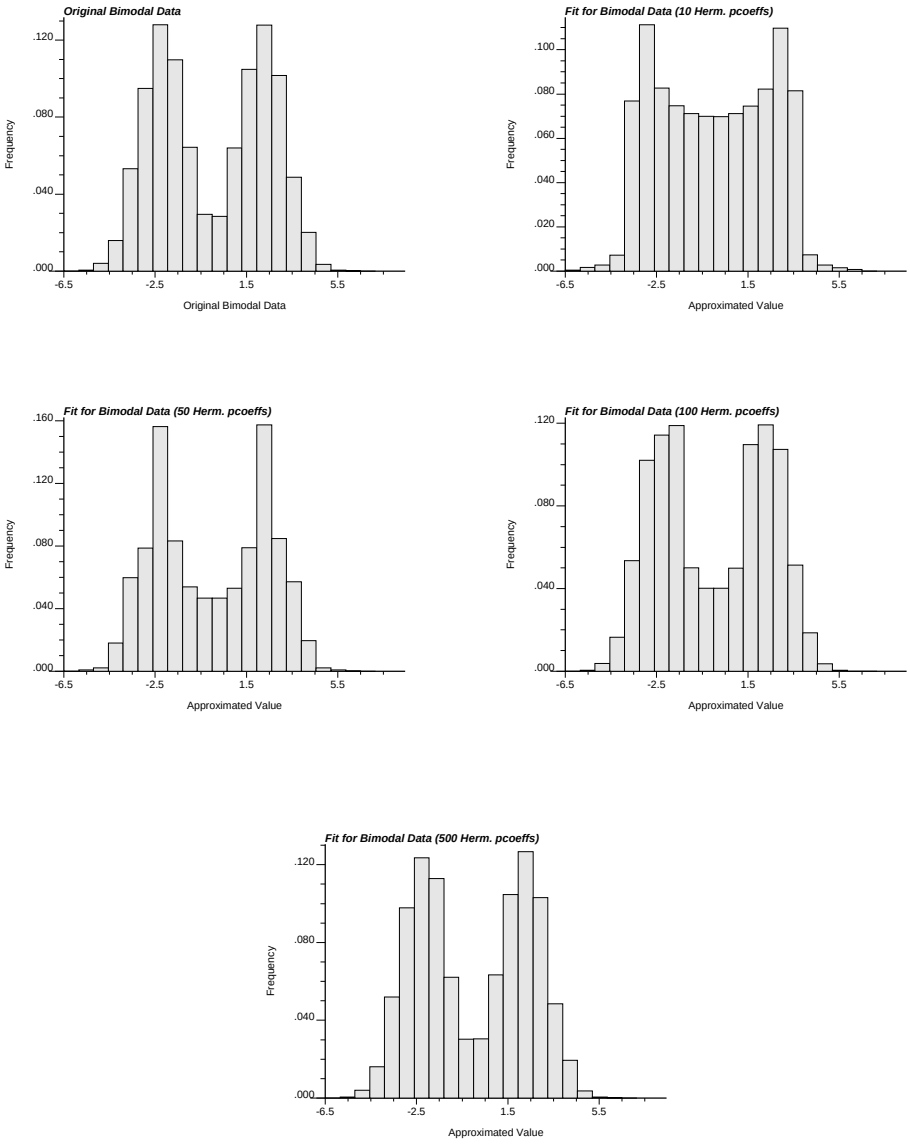

Figure 5: Original *bimodal* distribution (top left) and corresponding reproduction of *bimodal* distribution with 10, 50, 100 and 500 Hermite polynomials

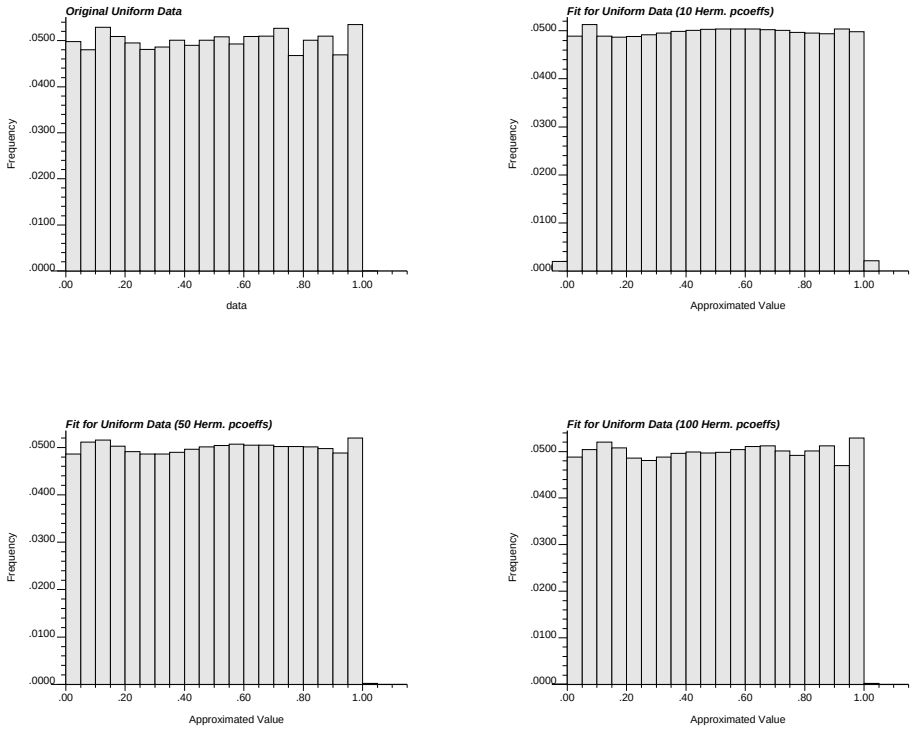

Figure 6: Original *uniform* distribution (top left) and corresponding reproduction of *uniform* distribution with 10, 50 and 100 Hermite polynomials

of unbiasedness and minimum estimation variance (Goméz-Hernández's notes on DK):

$$D_N = \sum_{i=1}^{N} b_i[Z(\mathbf{u}_i)] \qquad i = 1, \ldots, N \tag{35}$$

Since D_N is generated by well-behaved functions $b_i[Z(\mathbf{u}_i)]$ (i.e. function has finite expected value and variance) of one datum at a time, the previous conditions can be written as the following N conditions and for any function $b(\cdot)$

$$E\{L_0|b[Z(\mathbf{u}_i)]\} = E\{L_{DK}^*|b[Z(\mathbf{u}_i)]\} \quad i = 1, \ldots, N \tag{36}$$

we can rewrite Equation 36 as:

$$E\{E\{L_0|Z(\mathbf{u}_i)\}|b[Z(\mathbf{u}_i)]\} = E\{E\{L_{DK}^*|Z(\mathbf{u}_i)\}|b[Z(\mathbf{u}_i)]\} \quad i = 1, \ldots, N \tag{37}$$

By substituting $L_{DK}^* = \sum_{i=1}^{N} b_i[Z(\mathbf{u}_i)]$ into Equation 37 leads to the following set of integral equations:

$$\sum_{j=1}^{N} E\{b_j(Z(\mathbf{u}_j))|Z(\mathbf{u}_i)\} = E\{L(Z(\mathbf{u}_0))|Z(\mathbf{u}_i)\}, \quad i = 1, \ldots, N \tag{38}$$

The solution of this set of integral equations requires full knowledge of the bivariate distributions of the pairs of RV's $[Z(\mathbf{u}_i), Z(\mathbf{u}_j)]$ and $[Z(\mathbf{u}_i), Z(\mathbf{u}_0)]$, for $i, j = 1. \ldots, N$, to allow evaluation through integration of the conditional expectations involved in expression (38).

Implementation of DK requires the solution of the set of integral equations (38) to determine the N functions $b_i[Z(\mathbf{u}_i)]$. There are two problems involved in the solution of these integral equations: the first relates to inference of a bivariate distribution from a single realization sparsely sampled; the second relates to the difficulty of solving a system of type (38) even from a known bivariate distribution. To overcome these two problems, a bivariate "isofactorial model" has been adopted. This model amounts to assume that the bivariate distribution of $[z(\mathbf{u}), z(\mathbf{u}+\mathbf{h})]$ is fully described by its marginal distribution and a single function of $\mathbf{h}$, the correlogram. In the same way that kriging is based on the variogram, so DK is based on the bivariate distributions.

An isofactorial bivariate distribution which presents for marginal pdf $f(z)$ is defined as

$$f(\mathbf{h}; z, z') = f(z)f(z') \sum_{n=0}^{\infty} \frac{(\rho(\mathbf{h}))^n}{h_n^2} X_n(z) X_n(z') \tag{39}$$

where $\rho(\mathbf{h})$ is the stationary correlogram of the RF $Z(\mathbf{u})$ and z, z' are any two threshold values applied on the random variables $Z(\mathbf{u})$ and $Z(\mathbf{u}+\mathbf{h})$ at two locations h distant away. The bivariate pdf is $f(\mathbf{h}; z, z') = f(z)f(z')$. The set $[X_n(\cdot), n = 0, \ldots, \infty]$ is a set of polynomials orthogonal for the marginal pdf $f(z)$ with respective norms h_n^2. Expression 39 appears as a decomposition of the bivariate pdf into a sum of products of the two marginal pdf values $f(z)$ and $f(z')$.

Depending on whether the pdf $f(z)$ is bounded, one-sided bounded or un-bounded, the corresponding orthogonal polynomials are of three types: Legendre polynomials, Jacobi polynomials and Hermite polynomials. Hermite polynomials, which results in a marginal pdf, $f(z)$, which is Gaussian, have been used throughout the analytical derivations of local ccdfs.

Since the Hermite polynomials are spatially non-correlated, 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 DK of any function $Y(\mathbf{u})$:

$$f[y(\mathbf{u})]^{DK} = f_0 + f_1[H_1(y(\mathbf{u}))]^K + f_2[H_2(y(\mathbf{u}))]^K + \cdots \tag{40}$$

If we denote

$\mathbf{u}_i$, $\mathbf{u}_j$ ...	the data points
ρ_{ij}	the correlation coefficient between $y(\mathbf{u}_i)$ and $y(\mathbf{u}_j)$
$H_n(y_i)$	the Hermite polynomials $H_n(y(\mathbf{u}_i))$,
"DK"	Disjunctive Kriging
"SK"	Simple Kriging.

then we have,

$$[H_n(y(\mathbf{u}))]^K = \sum_{i=1}^{N} \lambda_{ni} H_n(y_i) \tag{41}$$

where the λ_{ni} satisfy the SK system

$$\sum_{j=1}^{N} \lambda_{nj} Cov\{H_n(y_i), H_n(y_j)\} = Cov\{H_n(y_i), H_n(y(\mathbf{u}))\} \tag{42}$$

Using Equations 10 through 12, we can rewrite Equation 42 as;

$$\sum_{j=1}^{N} \lambda_{nj} [\rho_{nj}]^n = [\rho_{i\mathbf{u}}]^n \tag{43}$$

As the correlation structure $[\rho(\mathbf{h})]^n$ of $H_n(y(\mathbf{u}))$ rapidly tends to one of pure nugget effect, the kriged estimator at 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 Equation 40 to get the result.

Besides, for n series of polynomials, Kriging variance of the $H_n(y(\mathbf{u}))$ and that the DK estimation variance of $f[(y(\mathbf{u}))]$ can be stated as;

$$\sigma^2_{Kn} = 1 - \sum_{j=1}^{N} \lambda_{nj} [\rho_{nj}]^n \tag{44}$$

$$Var\{f[y(\mathbf{u})] - f[y(\mathbf{u})]^{DK}\} = \sum_{n=1}^{\infty} (f_n)^2 \sigma^2_{Kn} \tag{45}$$

"1-point" Disjunctive Kriging

Given a conditional mean m_y and variance σ^2_y in Gaussian space, we can determine a *one-data-point* data value, y_1, with correlation $\rho(\mathbf{h})$ so that the conditional variance and mean are what we specify (m_y, σ^2_y). The variance is written

$$\sigma^2_y = 1 - [\rho(\mathbf{h})]^2 \tag{46}$$
$$\rho(\mathbf{h}) = \sqrt{1 - \sigma^2_y}$$

The mean m_y is written as;

$$m_y = \rho(\mathbf{h}) y_1 \tag{47}$$

We can write y_1, as;

$$y_1 = \frac{m_y}{\rho(\mathbf{h})} \tag{48}$$

So, "*one-data-point*" DK with the data value being y_1 and $\rho(\mathbf{h})$, will return a DK estimate with the correct mean and variance.

When we have one data point, the DK kriging weight is $[\rho(\mathbf{h})]^n$. Also, in case of *one-data-point*, one can write the estimated Hermite Polynomials for point $y(\mathbf{u})$, as;

$$[H_n(y(\mathbf{u}))]^* = \lambda_1 H_n(y_1) + (1 - \lambda_1) E\{H_n(y)\} \tag{49}$$

since $E\{H_n(y)\} = 0$ (see Equation 7), the second term in Equation 49 vanishes and results in:

$$[H_n(y(\mathbf{u}))]^* = [\rho(\mathbf{h})]^n H_n(y_1) \tag{50}$$

after a substitution of $\lambda_1 = [\rho(\mathbf{h})]^n$ in Equation 49, Equation 50 is the final form of Hermite Polynomials of conditional distributions.

Derive the Conditional Distribution

Case A: m_y and σ_y^2 known

Our aim in this section is to calculate the local conditional distributions, $F(\mathbf{u}; z|(n))$, (ccdf) from the given m_y and σ_y^2. The steps for the "analytical" derivation of these ccdfs are outlined:

- Tranform the original variable $z_i, i = 1, \ldots, N$ into Gaussian space (i.e. $y_i = G^{-1}(F(z_i)), i = 1, \ldots, N$)
- Using Equation 6, calculate the Hermite polynomials for all the Gaussian transformed y_i data values (i.e. $H_n(y)$)
- Calculate the $\rho(\mathbf{h})$ and y_1, (*one-data-point* DK) using the Equation 46 and Equation 48.
- Calculate the Hermite polynomials for the y_1, (i.e. $H_n(y_1)$)
- Calculate the simple kriging Hermite polynomials for point of $y(\mathbf{u})$ using Equation 50 (i.e. $[H_n(y(\mathbf{u}))]^* = [\rho(\mathbf{h})]^n H_n(y_1)$)
- Now, in order to calculate the full ccdf, we need to calculate the complete series of indicator values of $I_{z(\mathbf{u})<z_c}$ using Equation 34. Then, the DK estimator is obtained by kriging each of the $H_n(y(\mathbf{u}))$ separetly.

$$[I_{z(\mathbf{u})<z_c}]^{DK} = G(y_c) + \sum_{n=1}^{\infty} H_{n-1}(y_c) g(y_c) [H_n(y(\mathbf{u}))]^{SK} \tag{51}$$

  where, simple kriging Hermite polynomials, $[H_n(y(\mathbf{u}))]^{SK} = [\rho(\mathbf{h})]^n H_n(y_1)$.
- All the terms in Equation 51 have been calculated in previous steps. The local conditional cumulative density function is given by $[I_{z(\mathbf{u})<z_c}]^{DK}$.
- In order to get the probability density functions (i.e. pdfs) from ccdfs, one can uniformly sample from the calculated ccdfs and retrive the corresponding data values, $(z_i(\mathbf{u}), i = 1, \ldots, K)$, where K is the number of uniformly sampled discretized values. The pdf can be established by getting the histogram of the $z_i(\mathbf{u})$.
- Once we have the probability density function, the conditional mean, m_z and conditional variance, σ_z^2 can be calculated using the relations:

$$m_z = \frac{\sum_{i=1}^{K} z_i(\mathbf{u})}{K} \tag{52}$$

$$\sigma_z^2 = \frac{\sum_{i=1}^{K} [z_i(\mathbf{u}) - m_z]^2}{K} \tag{53}$$

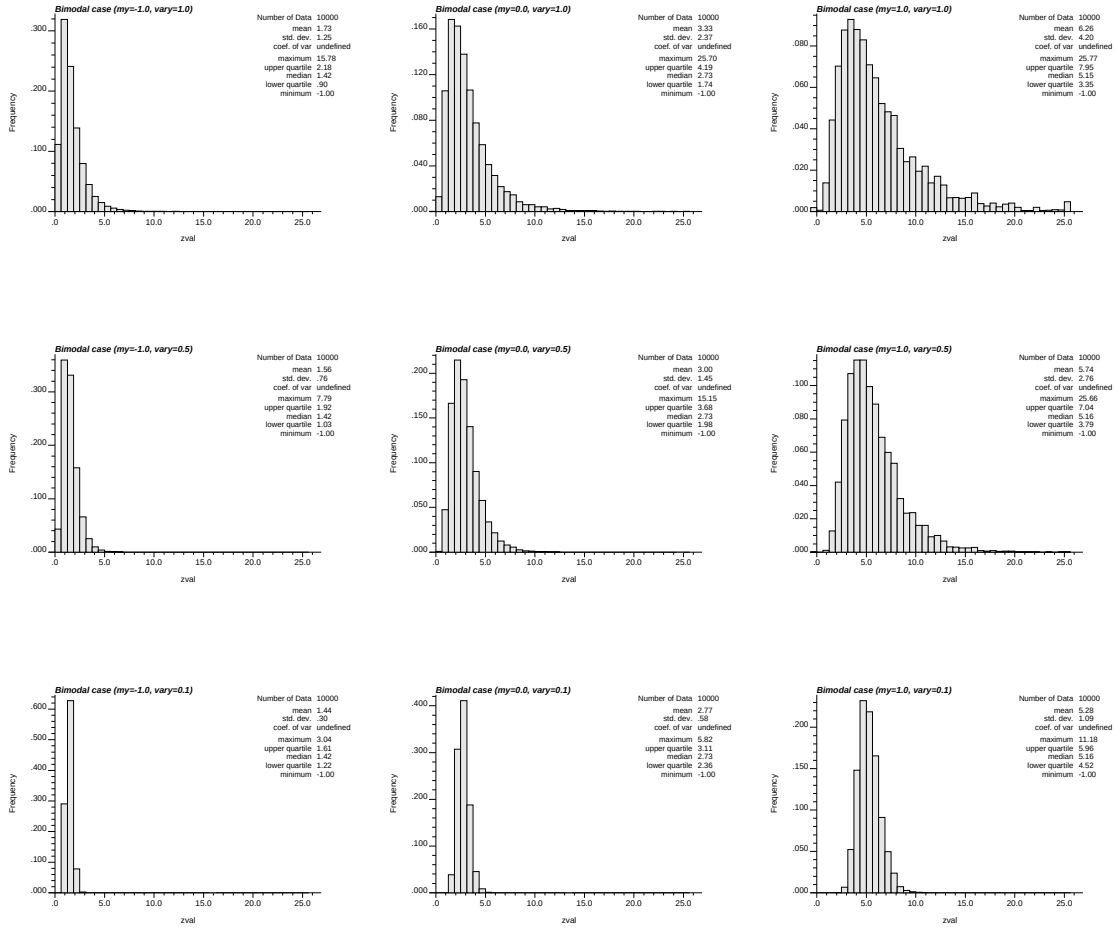

Figure 7: Non-standard distributions for lognormal data generated by Hermite polynomials (from left to right: mean in normal space: -1.0/0.0/1.5; from top to bottom: variance in normal space: 1.0/0.5/0.1).

- By using the "Graphical" technique, we have already presented the conditional probability distributions for lognormal, bimodal and uniform distributiona for different mean and variance in Gaussian space (Figures 1, 2 and 3). As an example application, by using the new analytical relations, we will try to reproduce all those conditional probability distributions for lognormal, bimodal and uniform distributions.

  The same mean and variance configurations used in Figures 1, 2 and 3 are also used for this section in order to calculate those conditional probability distributions by using *one-data-point* DK and Hermite polynomials. Generated ccdfs for lognormal, bimodal and uniform distributions are presented in Figures 7, 8 and 9. When we compare the corresponding figures, it is seen that new anaytical technique reproduces all the desired conditional probability distributions, m_z and σ_z^2.

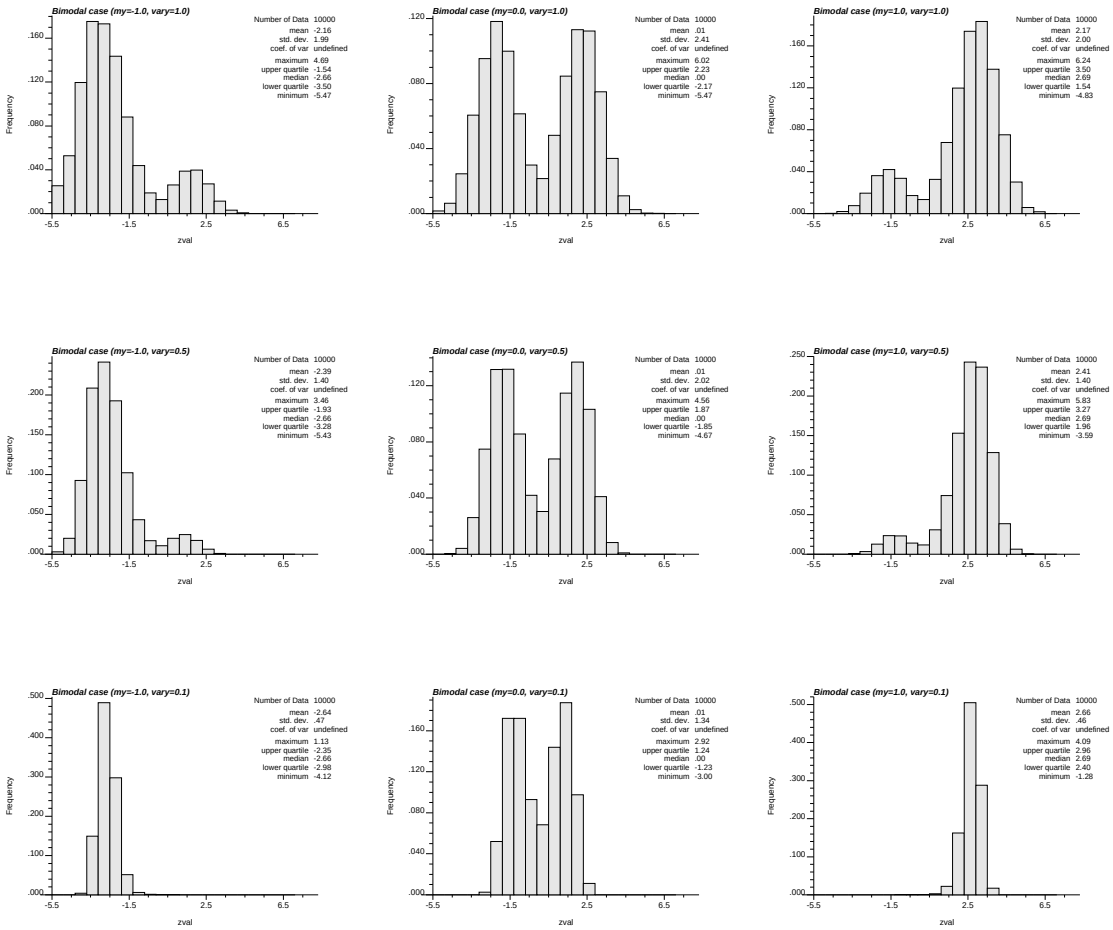

Figure 8: Local distributions for bimodal data generated by Hermite polynomials (from left to right: mean in normal space: -1.0/0.0/1.5; from top to bottom: variance in normal space: 1.0/0.5/0.1).

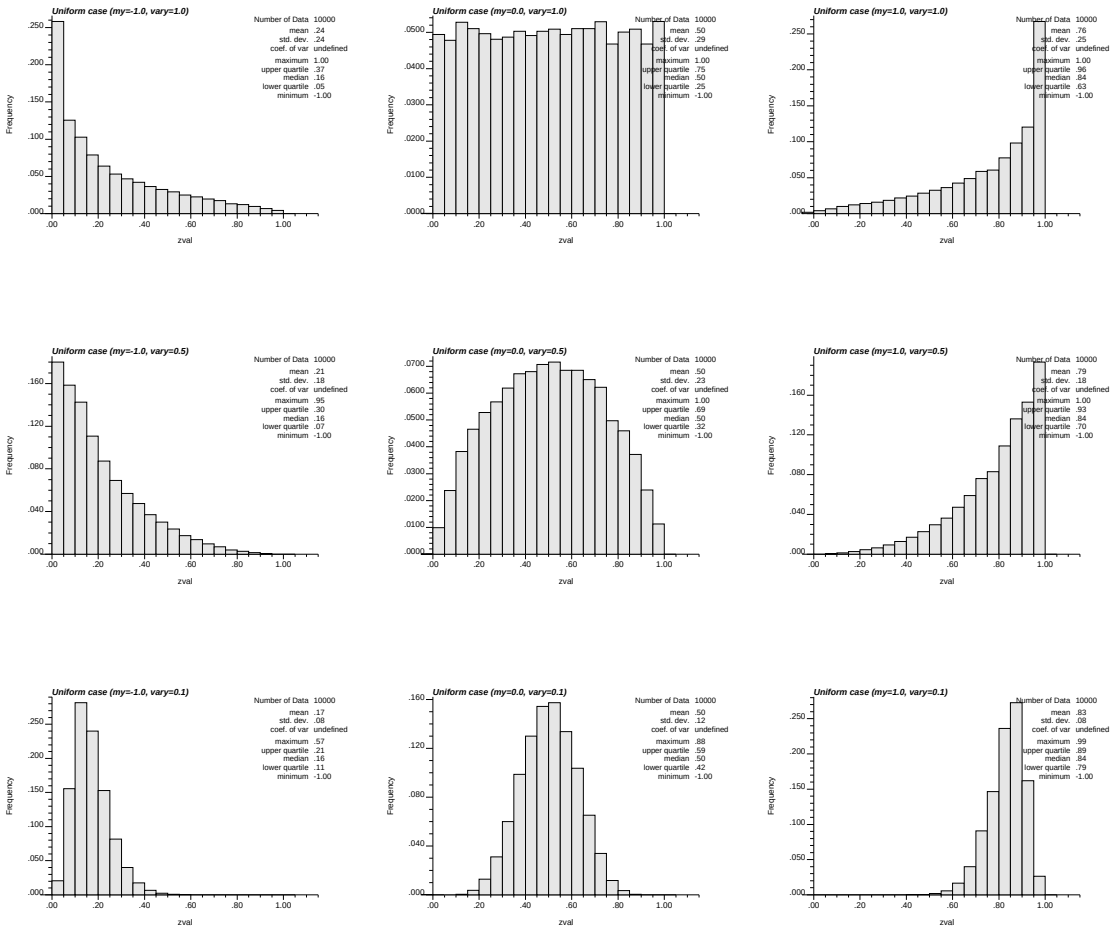

Figure 9: Non-standard distributions for uniform data generated by Hermite polynomials (from left to right: mean in normal space: -1.0/0.0/1.5; from top to bottom: variance in normal space: 1.0/0.5/0.1).

Case B: Direct Inference from m_z and σ_z^2

The value of having an analytical relation to calculate the local condtional distributions, $F(\mathbf{u}; z|(n))$, from the given m_z and σ_z^2 is obvious. The proposed analytical technique in Case A is definitely a good starting point; however, it does not go far than calculating the ccdfs relatively quickly in case of having m_y and σ_y^2. The main motivation for calculating the ccdfs directly is the desire of using them in direct sequential simulation technology. In such case, our hope is to replace all the graphical transformation and extensive lookup table construction with anaytically defined, easy to be calculated and implemented equations. The steps for the direct calculation of ccdfs are outlined;

- Tranform the original variable $z_i, i = 1, \ldots, N$ a into Gaussian space (i.e. $y_i = G^{-1}(F(z_i)), i = 1, \ldots, N$)
- Using Equation 6, calculate the Hermite polynomials for all the Gaussian transformed y_i data values (i.e. $H_n(y)$)
- Using Equation 27, calculate the Hermite coefficients for all the Gaussian transformed y_i data values (i.e. f_n)
- The main difference of Case B from Case A is that the $\rho(\mathbf{h})$ and the y_1 are not readily available to proceed to get the ccdfs. Both the $\rho(\mathbf{h})$ and the y_1 values could be determined either by iteration or minimization methodolies. We chose the minimization technique and try to get the $\rho(\mathbf{h})$ value by minimizing the difference between the calculated conditional variance, σ_{DK}^2 and the given σ_z^2. For a *one-data-point* DK, Kriging variance (Equation 44) can be written as

$$\sigma_{Kn}^2 = 1 - \sum_j \lambda_1 [\rho(\mathbf{h})_{nj}]^n = 1 - [\rho(\mathbf{h})]^{2n} \tag{54}$$

  where, $\lambda_1 = [\rho(\mathbf{h})]^n$.

  The conditional variance, which is the variance of DK can be calculated by inserting the Equation 54 into Equation 45 and reorganizing it as:

$$\sigma_{DK}^2 = \sum_1^\infty (f_n)^2 [1 - (\rho(\mathbf{h}))^{2n}] \tag{55}$$

  By minimizing the difference $[\sigma_{DK}^2 - \sigma_z^2]$, optimum correlation coefficient, $\rho(\mathbf{h})$ can be found.

  Now we can use the $\rho(\mathbf{h})$ value in order to determine the *one-data-point*, y_1. The DK estimate $z^*(\mathbf{h})$ could be calculated as (see Equations 14):

$$z^*(\mathbf{u}) = \sum_{n=0}^\infty f_n H_n(y^*(\mathbf{u})) \tag{56}$$

  For a *one-data-point* DK case and using the relation 50, Equation 56 can be written as:

$$z^*(u) = \sum_{n=0}^\infty f_n (\rho(\mathbf{h}))^n H_n(y_1) \tag{57}$$

By minimizing the difference $(z^*(\mathbf{u}) - m_z)$, optimum y_1 can be calculated. Once we have determined the $(\rho(\mathbf{h}))$ and the y_1, the determination of ccdfs is same as that of Case A.

- Although we have just outlined the determination of ccdfs, once we know the m_z and σ_z^2, we could not able to produce the ccdfs using this technique. One alternative technique could be the "conditional expectation" concept.

Conclusion

An alternative analytical technique has been developed to calculate local conditional distributions (ccdfs) using the theory of Hermite polynomials and Disjunctive Kriging. For a given mean and variance in Gaussian space, one can calculate the conditional distributions. Our main aim was to calculate the ccdfs having mean and variance in the original data space.

References

Deutsch, C. V., Tran, T. T. and Xie, Y. L., An approach to ensure histogram reproduction in Direct Sequential Simulation, In C. V. Deutsch editor, *2001 CCG Report*, CCG, University of Alberta, Canada, 2000.

Deutsch, C. V. and Journel, A. G., *GSLIB: Geostatistical Software Library and User's Guide*, Oxford University Press, New York, 1992.

Oz, B., Deutsch, C. V., Tran, T. T. and Xie, Y. L., `DSSIM-HR`: A FORTRAN 90 Program for Direct Sequential Simulation with Histogram Reproduction, submitted to Computer & Geosciences, November 2001

Goméz-Hernández, Disjunctive Kriging: The Untold Story, Department of Applied Earth Sciences, Stanford University, CA, 94305-2225

Journel, A. G., Class notes on sequential simulation, Stanford University, CA, 1986

Rivoirard, J., Introduction to Disjucntive Kriging and Non-linear Geostatistics, Clarendon Press, Oxford, 1994.

Isaaks, E. H., and Srivastava, R. M., An Introduction to Applied Geostatistics, Oxford University Press, 1989.