



Geostatistics Formula Sheet

December 2009

Z denotes a random variable (RV). z denotes an outcome. Z(u) denotes a regionalized RV at location u. The set of random variables over a stationary domain A {Z(u), u ∈ A} is known as a random function (RF).

Uncertainty in a RV is represented by a cumulative distribution function (CDF): F(z)=Prob{Z ≤ z}. The derivative of the CDF is the probability density function (PDF): f(z)=F'(z). Quantiles are z-values with a probabilistic meaning: z_p such that F(z_p)=p. The quantile function is denoted F⁻¹(p)=z_p.

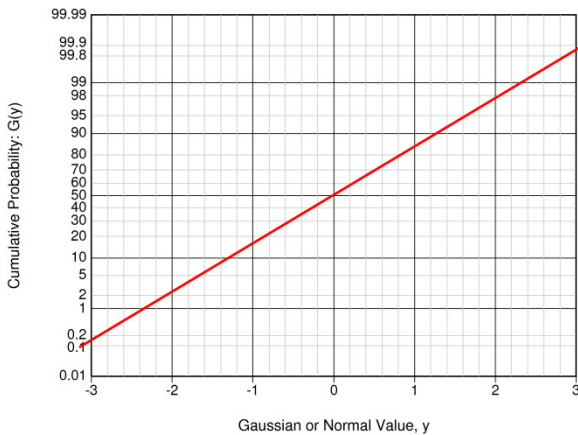
The expected value operator is written E{Z} = ∫_{-∞}[∞] z · f(z) dz. E{Z} is denoted m and is also known as the first moment or mean. The variance is σ²=E{[Z-m]²}=E{Z²}-m². σ is the standard deviation. σ/m is the coefficient of variation.

A random variable Z is standardized by Y=(Z-m_z)/σ_z. E{Y}=0, E{Y²}=1 and Z=Y σ_z + m_z.

Z is uniform in the interval a to b when:

$$f(z) = \begin{cases} 1/(b-a), & \forall z \in [a,b] \\ 0, & \text{otherwise} \end{cases}, \quad m = \frac{a+b}{2} \quad \text{and} \quad \sigma^2 = \frac{(b-a)^2}{12}$$

Z is standard normal or Gaussian when $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$



The variable Z>0 is lognormal with m and σ² when Y=ln(Z) is normal with mean α and variance β². The parameters:

$$\alpha = \ln(m) - \beta^2 / 2 \quad \beta^2 = \ln(1 + \sigma^2 / m^2)$$

$$m = e^{\alpha + \beta^2 / 2} \quad \sigma^2 = m^2 \left[e^{\beta^2} - 1 \right]$$

The multivariate distribution of N RVs Z_i, i=1,...,N is defined as:

$$F_{Z_1, \dots, Z_N}(z_1, \dots, z_N) = \text{Prob}\{Z_1 \leq z_1, \dots, Z_N \leq z_N\}$$

Conditional distributions are calculated as:

$$F_{Y|Z_1, \dots, Z_N}(y) = \frac{F_{Y, Z_1, \dots, Z_N}(y, z_1, \dots, z_N)}{F_{Z_1, \dots, Z_N}(z_1, \dots, z_N)}$$

The covariance and correlation coefficient summarize bivariate dependence between two random variables:

$$\text{Cov}\{X, Y\} = C_{XY} = E\{[X - m_X][Y - m_Y]\} = E\{XY\} - m_X \cdot m_Y$$

$$\rho_{XY} = \rho = C_{XY} / (\sigma_X \cdot \sigma_Y)$$

The variogram for lag h is defined: $2\gamma(\mathbf{h}) = E\{[Z(\mathbf{u}) - Z(\mathbf{u}+\mathbf{h})]^2\}$

Under stationarity, the variogram, variance and covariance are related by $\gamma(\mathbf{h}) = \sigma^2 - C(\mathbf{h})$.

The scalar normalized distance is $h = \sqrt{\left(\frac{h_X}{a_X}\right)^2 + \left(\frac{h_Y}{a_Y}\right)^2 + \left(\frac{h_Z}{a_Z}\right)^2}$

Clockwise rotation of X/Y by angle α is achieved by:

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$$

Stratigraphic relative coordinates are calculated as:

$$Z_{rel}(x, y) = \frac{Z(x, y) - Z_{cb}(x, y)}{Z_{ct}(x, y) - Z_{cb}(x, y)} \cdot T$$

Variograms are modeled by structures: $\gamma(\mathbf{h}) = \sum_{i=0}^{nst} C_i \cdot \Gamma_i(\mathbf{h})$.

Common standardized models include the Exponential

$$\text{Exp}(h) = 1 - \exp(-3h/a), \quad \text{Spherical} \quad \text{Sph}(h) = 1.5(h/a) - 0.5(h/a)^3$$

if $h \leq a$; 1, otherwise, Gaussian $\text{Gaus}(h) = 1 - \exp(-3(h/a)^2)$.

$$\text{The hole effect is less common: } \gamma(h) = C \cdot \left[1 - \cos\left(\frac{h\pi}{a}\right) \right]$$

The volume averaged variogram between v and V (gammabar):

$$\bar{\gamma}(V, v) = \frac{1}{|V| \cdot |v|} \int_V \int_v \gamma(x-y) dx dy$$

The dispersion variance is given by:

$$D^2(v, V) = E\{[Z_V - m_V]^2\} = \bar{\gamma}(V, V) - \bar{\gamma}(v, v)$$

Variances add: $D^2(v, A) = D^2(v, V) + D^2(V, A) \quad v < V < A$

Variance of a linear combination:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{Var}\{\bar{x}\} = \frac{\sigma_X^2}{n} + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1, i \neq j}^n \text{Cov}\{x_i, x_j\}$$

Linear estimation at u_□ given by: $z_{\square}^* - m_{\square} = \sum_{i=1}^n \lambda_i \cdot [z_i - m_i]$

The estimation variance is calculated as:

$$\sigma_E^2 = \sigma^2 - 2 \sum_{i=1}^n \lambda_i C_{i,\square} + \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C_{i,j}$$

Minimizing the estimation variance leads to simple kriging and minimized estimation variance (kriging variance):

$$\sum_{j=1}^n \lambda_j C_{i,j} = C_{i,\square} \quad i=1, \dots, n \quad \sigma_{SK}^2 = \sigma^2 - \sum_{i=1}^n \lambda_i C_{i,\square}$$

Ordinary kriging – constrain the sum of the weights to one:

$$\begin{cases} \sum_{j=1}^n \lambda_j C_{i,j} + \mu = C_{i,\square} & i=1, \dots, n \\ \sum_{j=1}^n \lambda_j = 1 \end{cases}$$

Universal kriging: $m(\mathbf{u}) = \sum_{l=0}^L a_l \cdot f_l(\mathbf{u})$

$$\begin{cases} \sum_{j=1}^n \lambda_j C_{i,j} + \sum_{l=0}^L \mu_l = C_{i,\square} & i=1, \dots, n \\ \sum_{j=1}^n \lambda_j \cdot f_l(\mathbf{u}_j) = f_l(\mathbf{u}_\square) & l=0, \dots, L \end{cases}$$

External drift considers $m(\mathbf{u}) = a_0 + a_1 f_1(\mathbf{u})$

Location dependent variance of SK: $Var\{z_{SK}^*\} = \sigma^2 - \sigma_{SK}^2$

The cross variogram between variable $Z_i(\mathbf{u})$ and $Z_j(\mathbf{u})$:

$$2\gamma_{i,j}(\mathbf{h}) = E\{[Z_i(\mathbf{u}) - Z_i(\mathbf{u}+\mathbf{h})][Z_j(\mathbf{u}) - Z_j(\mathbf{u}+\mathbf{h})]\}$$

Matrix of cross variograms can be modeled by linear model of coregionalization (LMC) $i,j=1, \dots, M$:

$$2\gamma_{i,j}(\mathbf{h}) = \sum_{k=0}^K b_{i,j}^k \cdot \Gamma_k(\mathbf{h})$$

Where each $M \times M$ matrix of coefficients ($k=0, \dots, K$) must be positive definite. Intrinsic model assumes all variograms are proportional. The Markov models assume that the cross variogram/covariance is proportional to a direct variogram.

$$C_{i,j}(\mathbf{h}) = b \cdot C_{i,i}(\mathbf{h}) \quad \text{where } b = (\sigma_j / \sigma_i) \cdot \rho_{i,j}$$

Cokriging considers correct covariance between data events.

Z-data are transformed to be Y-normal ($G(y)$ is Gaussian CDF) with normal score transform:

$$y = G^{-1}(F(z)) \quad \text{and} \quad z = F^{-1}(G(y))$$

The n-variate multivariate Gaussian distribution is defined:

$$f(\mathbf{y}) = \frac{1}{(\sqrt{2\pi})^n |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{y}-\boldsymbol{\mu})\right]$$

Where $\boldsymbol{\mu}$ is the $1 \times n$ vector of mean values and Σ is the $n \times n$ matrix of covariances. Conditional distributions defined by normal equations (see simple kriging).

LU simulation from a covariance matrix: $\mathbf{C}=\mathbf{L}\mathbf{U}$; $\mathbf{y}=\mathbf{L}\mathbf{w}$.

Sequential simulation relies on recursive decomposition of the multivariate distribution:

$$\begin{aligned} P(A_1, \dots, A_N) &= P(A_N | A_1, \dots, A_{N-1}) \cdot P(A_1, \dots, A_{N-1}) \\ &= P(A_N | A_1, \dots, A_{N-1}) \cdot P(A_{N-1} | A_1, \dots, A_{N-2}) \cdot P(A_1, \dots, A_{N-2}) \\ &\dots \\ &= P(A_N | A_1, \dots, A_{N-1}) \cdot P(A_{N-1} | A_1, \dots, A_{N-2}) \cdot \dots \cdot P(A_2 | A_1) \cdot P(A_1) \end{aligned}$$

Simulation from a univariate distribution amounts to quantile

transformation of a random number: $z_s = F_Z^{-1}(r)$

Indicators for continuous variables

$$i(\mathbf{u}_\alpha; z_c) = \begin{cases} 1, & \text{if } z(\mathbf{u}_\alpha) \leq z_c \\ 0, & \text{otherwise} \end{cases} \quad \text{for many cutoffs } z_c$$

Indicators for categorical variables

$$i(\mathbf{u}_\alpha; k) = \begin{cases} 1, & \text{if } \mathbf{u}_\alpha \in k \\ 0, & \text{otherwise} \end{cases} \quad \text{for } k = 1, \dots, K$$

Mean and variance of an indicator variable are given by:

$$E\{i(\mathbf{u}_\alpha; k)\} = p_k \quad Var\{i(\mathbf{u}_\alpha; k)\} = p_k(1-p_k)$$

Permanence of ratios for combining conditional probabilities:

$$P(A | B_i, i = 1, \dots, n) = \frac{\left(\frac{1-P(A)}{P(A)}\right)^{n-1}}{\left(\frac{1-P(A)}{P(A)}\right)^{n-1} + \prod_{i=1}^n \frac{1-P(A|B_i)}{P(A|B_i)}}$$

Stepwise conditional transformation:

$$Y_1 = G^{-1}[\text{Prob}(Z_1 \leq z_1)]$$

$$Y_{2|1} = G^{-1}[\text{Prob}(Z_2 \leq z_2 | Y_1 = y_1)]$$

$$Y_{3|2,1} = G^{-1}[\text{Prob}(Z_3 \leq z_3 | Y_2 = y_2, Y_1 = y_1)]$$

Bayesian updating prior and likelihood Gaussian distributions:

$$\overline{y_U} = \frac{\overline{y_L} \sigma_P^2 + \overline{y_P} \sigma_L^2}{\sigma_P^2 - \sigma_P^2 \sigma_L^2 + \sigma_L^2} \quad \sigma_U^2 = \frac{\sigma_P^2 \sigma_L^2}{\sigma_P^2 - \sigma_P^2 \sigma_L^2 + \sigma_L^2}$$

Compositional data could be handled by additive logratios:

$$y_i = \ln\left(\frac{x_i}{x_D}\right) \quad \text{and} \quad x_i = \frac{\exp(y_i)}{\sum_{i=1}^d \exp(y_i) + 1}$$

Disclaimer: there may be mistakes on this formula sheet. Any mistakes are your fault – you should not need a formula sheet anyway.