Geostatistics Formula Sheet

October 2014

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

Uncertainty in a RV is represented by a cumulative distribution function (CDF): $F(z)=\text{Prob}\{Z\le z\}$. 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 $F^{-1}(p)=z_p$.

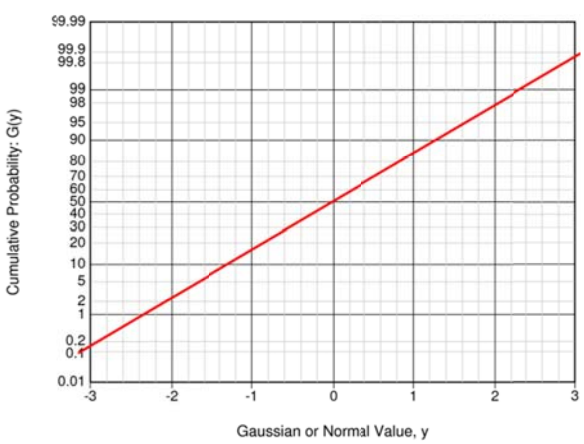
The expected value operator is written $E\{Z\}=\int_{-\infty}^{\infty} z\cdot f(z)dz$. $E\{Z\}$ is denoted m and is also known as the first moment or mean. The variance is $\sigma^2=E\{[Z-m]^2\}=E\{Z^2\}-m^2$. σ is the standard deviation. σ/m is the coefficient of variation.

A random variable Z is standardized by $Y=(Z-m_Z)/\sigma_Z$. $E\{Y\}=0$, $E\{Y^2\}=1$ and $Z=Y\,\sigma_Z+m_Z$.

Z is uniform in the interval a to b when:

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

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 σ^2 when $Y=\ln(Z)$ is normal with mean α and variance β^2. The parameters:

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

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

The multivariate distribution of N RVs $Z_i, i=1,\ldots,N$ is defined as:

$$F_{Z_1,\ldots,Z_N}(z_1,\ldots,z_N)=\text{Prob}\left\{Z_1\le z_1,\ldots,Z_N\le z_N\right\}$$

Conditional distributions are calculated as:

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

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

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

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

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

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 $Exp(h)=1-\exp(-3h/a)$, Spherical $Sph(h)=1.5(h/a)-0.5(h/a)^3$ if $h\le a$; 1, otherwise, Gaussian $Gaus(h)=1-\exp(-3(h/a)^2)$.

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

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

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

The dispersion variance is given by:

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

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

Variance of a linear combination:

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

Linear estimation at $\mathbf{u}_\square$ given by: $z_\square^*-m_\square=\sum_{i=1}^{n}\lambda_i\cdot\left[z_i-m_i\right]$

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):

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

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

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

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

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

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

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

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

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

Where each *M*x*M* matrix of coefficients (*k=0,...,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.

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

Cokriging considers correct covariance between data events.

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

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

The n-variate multivariate Gaussian distribution is defined:

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

Where μ is the *1*x*n* vector of mean values and Σ is the *n*x*n* matrix of covariances. Conditional distributions defined by normal equations (see simple kriging).

LU simulation from a covariance matrix: **C**=**LU**; **y**=**Lw**.

Sequential simulation relies on recursive decomposition of the multivariate distribution:

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

Simulation from a univariate distribution amounts to quantile transformation of a random number: $z_s=F_Z^{-1}(r)$

Indicators for continuous variables

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

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:

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

Permanence of ratios for combining conditional probabilities:

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

Stepwise conditional transformation:

$$\begin{aligned}Y_1&=G^{-1}\left[\text{Prob}(Z_1\le z_1)\right]\\Y_{2|1}&=G^{-1}\left[\text{Prob}(Z_2\le z_2|Y_1=y_1)\right]\\Y_{3|2,1}&=G^{-1}\left[\text{Prob}(Z_3\le z_3|Y_2=y_2,Y_1=y_1)\right]\end{aligned}$$

Bayesian updating prior and likelihood Gaussian distributions:

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

Compositional data could be handled by additive logratios:

$$y_i=\ln\left(x_i/x_D\right) \text{ and } x_i=\exp(y_i)\Big/\sum_{i=1}^{D}\exp(y_i) \qquad i=1,\dots,D$$

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