The Slope of Regression for Kriging Estimators

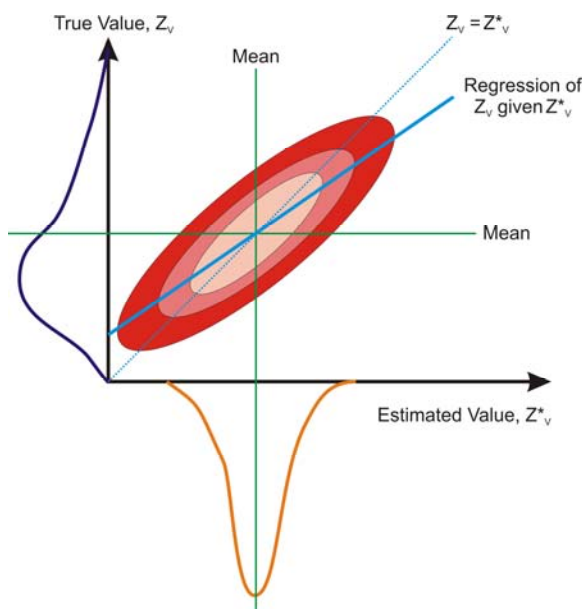
Clayton V. Deutsch

Centre for Computational Geostatistics
Department of Civil and Environmental Engineering
University of Alberta

The slope of the regression line, considering the true value and the estimated value, is often used as a diagnostic for conditional bias. Ideally, the slope of this line should be equal to one, which implies conditional unbiasedness. This is achieved exactly in simple kriging (SK) by reducing the variance of the estimates, that is, smoothing. This is not achieved in ordinary kriging (OK) or other forms of kriging with a trend. The slope can be calculated and it is always less than 1. This note recalls the basic equations for calculating the slope of regression for any type of linear estimate. Some practical consequences are recalled.

Background

Consider a regionalized variable *Z*. The context for the "slope of regression" is estimating the unknown true values Z_V for volume *V* at unsampled locations. The sketch to the right illustrates the situation. The estimate Z^*_V is the independent variable on the X axis because it will be known, but the truth Z_V remains unknown.



The mean of the true values Z_V and estimated values Z^*_V are the same; virtually all estimators are globally unbiased. The variance of the true values is larger than the variance of the estimates. The degree of smoothing is a function of the continuity of the regionalized variable, the amount of data and the search strategy for estimation. The regression of the true values given the estimates is an indication of conditional bias. The regression line is an approximation of the conditional expectation, which, in general, will not be equal to the 1:1 45^{o} line.

$$E\{Z_V \mid Z_V^* = z\} \cong a + bz \neq z$$

If the bivariate relationship between Z_V and Z^*_V were Gaussian, then the linear regression would be exactly the conditional expectation. The slope of the regression line is a reasonable approximation even in non-Gaussian settings. The slope of the regression line for Z_V on Z^*_V is given by:

$$b = \rho \frac{\sigma_Y}{\sigma_X} = \rho_{Z_V, Z_V^*} \frac{\sigma_{Z_V}}{\sigma_{Z_V^*}} = \frac{Cov\left\{Z_V, Z_V^*\right\}}{\sigma_{Z_V}\sigma_{Z_V^*}} \frac{\sigma_{Z_V}}{\sigma_{Z_V^*}} = \frac{Cov\left\{Z_V, Z_V^*\right\}}{\sigma_{Z_V^*}^2} \tag{1}$$

In simple kriging, we are estimating the residuals from the mean:

$$z_V^* - m = \sum_{i=1}^{n} \lambda_i \left(z_i - m\right)$$

In ordinary kriging (or kriging with a trend), we are estimating with the original data and constraints:

$$z_V^* = \sum_{i=1}^{n} \lambda_i z_i$$

It is well known that the estimation variance for linear estimation (any type of kriging) can be written as:

$$\sigma_E^2 = \bar{C}(V,V) - 2\sum_{i=1}^{n} \lambda_i \bar{C}(v_i,V) + \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_i C(v_i, v_j) \quad (2)$$

This can be simplified in the case of simple kriging:

$$\sigma_E^2 = \sigma_{SK}^2 = \bar{C}(V,V) - \sum_{i=1}^{n} \lambda_i \bar{C}(v_i,V) \quad (3)$$

In the case of ordinary kriging, the Lagrange multiplier appears to enforce the linear constraint that the sum of weights must equal one:

$$\sigma_E^2 = \sigma_{OK}^2 = \bar{C}(V,V) - \sum_{i=1}^{n} \lambda_i \bar{C}(v_i,V) + \mu \quad (4)$$

The sign on the Lagrange parameter μ depends on the sign written in the kriging equations. Consistency must be maintained.

The data are represented as v_i, which are z_i-m for simple kriging and z_i for ordinary kriging at the data scale v for simple kriging. The random variable is Z-m for simple kriging and the covariance function is representative of this residual. The random variable is Z for ordinary kriging. The slope is easily calculated with the kriging weights.

$$b = \frac{Cov\left\{Z_V, Z_V^*\right\}}{\sigma_{Z_V^*}^2} = \frac{\sum_{i=1}^{n} \lambda_i \bar{C}(v_i,V)}{\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_i C(v_i, v_j)} \quad (5)$$

Note that this expression is valid for all types of kriging; however, the covariance must be the correct one for the type of kriging under consideration. This equation appears to become undefined when the weights are all 0 (occurs in simple kriging when all of the data are beyond the range of correlation); however, the simple kriging case is very interesting.

An Alternative Expression for SK

We can take the limit of Equation 5 as the C(v_i,V) values approach zero, which causes the λ_i values to approach zero. Alternatively, we can substitute Equations 2 and 3 into Equation 5 and arrive at:

$$b = \frac{\bar{C}(V,V) - \sigma_{SK}^2}{\sigma_{SK}^2 - \bar{C}(V,V) + 2\left(\bar{C}(V,V) - \sigma_{SK}^2\right)} = \frac{\bar{C}(V,V) - \sigma_{SK}^2}{\bar{C}(V,V) - \sigma_{SK}^2} = 1 \quad (6)$$

The slope of regression is always 1 for simple kriging! At a data location, the slope is 1 from Equation 5 (the weight the collocated data is 1 and all other weights are 0). The slope is 1 when no data are correlated to the unsampled location (we take the limit of Equation 5 – applying L'Hopital's rule). From Equation 6, we can see that the slope is 1 all of the time. We could have inferred this from the behavior of the bivariate Gaussian distribution. We know that simple kriging has nice theoretical properties, this is another one of those nice properties. The slope of regression is always 1 and, theoretically, there is no conditional bias.

The slope of regression has no practical meaning in simple kriging, but the weight applied to the mean (one minus the sum of weights) has been used as a diagnositic. It is a measure of the data configuration and smoothing: the more weight to the mean, the more smoothing.

An Alternative Expression for OK

Now, in the case of ordinary kriging, the Lagrange multiplier appears, see Equation 4. The slope of regression does not degenerate to 1 as in the case of simple kriging. We can substitute Equations 2 and 4 into Equation 5 and arrive at:

$$b = \frac{\bar{C}(V,V) - \sigma_{OK}^2 + \mu}{\bar{C}(V,V) - \sigma_{OK}^2 + 2\mu} \tag{7}$$

While correct, this expression does provide great insight into how the slope of regression depends on the data configuration. Most practitioners just calculate the slope and assess its value. Experience shows that it is less than 1 in most cases. We could see that from Equation 5; the denominator will increase when the weights are constrained to sum to one (particularly because of the direct *i=j* variance terms) whereas the numerator will not increase as much. Of course, Equation 7 shows the same result because, m must always be positive (refer back to Equation 4), therefore, the slope of regression for constrained kriging (OK being one flavor) is always less than 1. We could say that OK is always conditionally biased unless the solution degenerates to SK. The magnitude of the conditional bias could be understood by making a map of *b*. We know that *b=1* when estimating at the data locations and *b* will decrease away from the data locations.

Practical Comments

It could be interesting to calculate the slope of regression when performing ordinary kriging or other types of constrained kriging; however, it is somewhat illogical. We can always get a slope of 1 – just use simple kriging! The rationale for ordinary kriging, however, is that there are departures from local stationarity, which can be accounted for by local estimation of the mean. The consequence is a slope of regression less than 1. It may be reasonable to check the slope of regression for different search strategies in ordinary kriging. The practitioner may want the best of everything: a slope of regression near 1, local estimation of the mean, and minimal smoothing. The latter two goals are in direct conflict with the first.

The slope of regression could be calculated in expected value given the theory, see Equation 5, and/or from cross valiation. Given that theory may not match practice, it would be reasonable to do both.

In simple kriging, we sometimes report the sum of the weights (or 1 minus the sum of weights), which acts as a diagnositic of how much smoothing is taking place. The weight to the mean is 1 minus the sum of weights.

Comments on Conditional Bias

Our concern with the slope of regression is conditional bias (refer back to the background on the first page). Conditional bias is a serious problem if the estimate is going to be used for a final or near-final decision, for example, for grade control in open pit or for stope estimates in underground. It would be a serious mistake to have a known bias in estimates used for final decision making.

On the other hand, we may be interested in estimates for interim planning purposes, that is, final estimates will be calculated in the future with additional information. We may accept conditional bias in interim estimates if the estimates have more desirable properties. The most common *desirable property* to have is a reasonable estimate of global reserves. Kriging with sparse data will lead to estimates that are overly smooth. A greater amount of higher and lower final estimates will be calculated when the final information is obtained. Thus, it may be a serious mistake to use smooth conditionally unbiased interim estimates for planning; we should anticipate the information available in the future.

In general, there is no universal *best* estimator. “best” must be defined for each situation. The debate should turn from conditional bias to the purpose of the estimate and the goals of the study.

References

David, M, 1977. *Geostatistical Ore Reserve Estimation (Developments in Geomathematics 2),* Elsevier, Amsterdam.

Deutsch, C.V. and Journel, A.G., 1998, *GSLIB: Geostatistical Software Library: and User's Guide,* Oxford University Press, New York, 2nd Ed.

Journel, A G and Huijbregts, Ch J, 1978. *Mining Geostatistics,* Academic Press, London.

Krige, D G, 1996, A practical analysis of the effects of spatial structure and data available and accessed, on conditional biases in ordinary kriging, in *Proceedings Fifth International Geostatistical Congress*, *Geostatistics Wollongong '96* (Eds: E Y Baafi and N A Schofield) pp 799-810.

Rivoirard, J, 1987. Two key parameters when choosing the kriging neighbourhood, *Mathematical Geology*, 19:851-856.

Vann, J., Jackson, S., and O Bertoli, Quantitative Kriging Neighbourhood Analysis for the Mining Geologist — A Description of the Method With Worked Case Examples, 5th International Mining Geology Conference Bendigo, Victoria, 17 - 19 November 2003.