Assessing Uncertainties for Lognormal Kriging Estimates

Jorge Kazuo Yamamoto

Department of Environmental and Sedimentary Geology, Institute of Geosciences, University of Sao Paulo, Brazil.

Abstract. Lognormal data calls for lognormal kriging in which the original data are transformed into logarithms. Then ordinary lognormal kriging estimates are backtransformed into original scale of measurement by taking their inverse logarithms. Evidently, we must add some non-bias to the estimates before backtransforming them. After that we have lognormal kriging estimates backtransformed into original scale of data. However, we do not have any idea about errors, because they remain in the logarithmic scale. This paper presents a very simple way to get back errors in the logarithmic domain into original data domain. Three data sets presenting different coefficients of variation are used to show how reliable is the proposed procedure. Furthermore, the non-bias term can also be backtransformed giving the estimated smoothing error. This estimated smoothing error presents a reasonable correlation with the true smoothing error. In other words, when the backtransformed estimate presents high correlation with the actual unknown value, then estimated and true smoothing errors will also present positive correlation.

Keywords: uncertainty, interpolation variance, smoothing error, lognormal kriging

1. Introduction

Lognormal data are very common in mineral deposits. For instance, distribution of rare metals follows a lognormal distribution that is characterized by a large quantity of low values and a small number of high values. Lognormal distributions present positive skewness that is the distributions are skewed to the right. This kind of data is very difficult to handle because the strong influence of a few high values. For example, when using ordinary kriging it is very common to find negative grades not only because the data are lognormal but also due to the negative weights of ordinary kriging. It is important to mention that negative grades just happen with lognormal data. Thus, the solution for negative grades is the post-processing of ordinary kriging weights to be all positive. However, this is not good enough for lognormal data. Thus, the solution calls for lognormal kriging in which estimation occurs in the logarithmic domain. Yamamoto\(^1\) has proposed a new non-bias term that is added to lognormal kriging estimate before taking its inverse logarithm. Now we can have ordinary lognormal kriging estimates backtransformed into original scale without presenting bias as reported in the literature. However, backtransformed values do not have associated errors because they remained in the logarithmic domain. That is the question. How can we bring errors back to the original domain? The solution is not simple as raising the base to the error, for instance the kriging variance. Moreover, it is very difficult to deal with error variances. Experimentally we have found a very simple solution that allows backtransformation of errors from the logarithmic domain to the original scale.

2. Lognormal kriging

Lognormal data are firstly transformed into logarithms as:

\[ Y(x) = \ln(Z(x)) \]

Then, the semivariogram is computed and modeled. The ordinary lognormal kriging estimator is:

\[ \hat{Y}(x) = \ln(\hat{Z}(x)) \]

\(^1\) Corresponding author. Tel.: +5511-3091-4203; fax: +5511-3091-4207.

E-mail address: jkyamamo@usp.br.

\[ Y_{OK}^* (x_o) = \sum_{i=1}^{n} \lambda_i Y(x_i) \] (1)

Associated errors with the ordinary lognormal kriging estimator are the kriging variance:
\[ \sigma_{OK}^2 = \sum_{i=1}^{n} \lambda_i \gamma(x_i - x_o) + \mu \]
and the interpolation variance:
\[ S_o^2 = \sum_{i=1}^{n} \lambda_i \left[Y(x_i) - Y_{OK}^* (x_o) \right]^2 \]

The ordinary lognormal kriging estimator can be backtransformed as (Journel2):
\[ Z_{OK}^* (x_o) = \exp(Y_{OK}^* (x_o) + \sigma_{OK}^2 / 2 - \mu) \] (2)

Where \( \sigma_{OK}^2 \) is the kriging variance; \( \mu \) is the Lagrange multiplier; \( \sigma_{OK}^2 / 2 - \mu \) is the non-bias term, according to Journel2. The main problem is that the expected value of backtransformed values is not equal to the sample mean (Journel and Huijbregts3):
\[ E[\exp(Y_{OK}^* (x_o) + \sigma_{OK}^2 / 2 - \mu)] \leq E[Z(x)] \]

The ordinary lognormal kriging estimator (equation 1) also presents the problem of the smoothing effect. It means the histogram of estimates does not match the sample histogram of logarithms. If we want to guarantee an unbiased backtransform of lognormal kriging estimates first we have to provide the histogram of estimates as close as possible to the sample histogram of logarithms. This is achieved by post-processing lognormal kriging estimates for correcting the smoothing effect.

\[ Y_{OK}^{**} (x_o) = Y_{OK}^* (x_o) + Y_{NS}^* (x_o) \text{factor} \] (3)

where \( Y_{NS}^* (x_o) \) is the correcting amount and \( \text{factor} \) a real number that makes the variance of corrected estimates equal to the sample variance, in this case the variance of logarithms. That is the solution proposed by this author and details of this algorithm can be found in Yamamoto1. Now we can get the backtransformed value as:
\[ Z_{OK}^{**} = \exp(Y_{OK}^{**} (x_o)) \] (4)

For normal kriging the corrected estimator is (Yamamoto4):
\[ Z_{OK}^{**} (x_o) = Z_{OK}^* (x_o) + Z_{NS}^* (x_o) \text{factor} \] (5)

Where \( Z_{NS}^* (x_o) \text{factor} \) is the correcting amount that removes from the smoothing effect of ordinary kriging. Actually, this term can be thought as the smoothing error as we will show on next section.

3. Backtransforming errors into original domain

How can we backtransform errors in the logarithmic domain to the original scale of data? The solution to this problem came out analyzing equation (3). We can rewrite equation (4) as:
\[ \exp(Y_{OK}^{**} (x_o)) = \exp(Y_{OK}^* (x_o) + Y_{NS}^* (x_o) \text{factor}) \]

Subtracting both sides of this expression from \( \exp(Y_{OK}^* (x_o)) \) we have:
\[ \exp(Y_{OK}^{**} (x_o)) - \exp(Y_{OK}^* (x_o)) = \exp(Y_{OK}^* (x_o) + Y_{NS}^* (x_o) \text{factor}) - \exp(Y_{OK}^* (x_o)) \] (6)

From this subtraction the difference is the non bias term backtransformed to the original scale of measurement that can be interpreted as the smoothing error. We have tested exhaustively this smoothing error and concluded that it makes sense as an error measure. In fact, we may think that the actual unknown value is equal to ordinary kriging estimate plus an error term:
\[ Z(x_o) = Z_{OK}^* (x_o) + Y(x_o) \]

This error term is none other than the smoothing error coming from the ordinary kriging estimation. If we knew the actual value on every location \( x_o \) we would derive the true smoothing error as:
\[ OK True Smoothing Error = Z(x_o) - Z_{OK}^* (x_o) \] (7)

Now, if we take away the ordinary kriging estimate from the corrected estimate we have:
\[ Z_{OK}^{**} (x_o) - Z_{OK}^* (x_o) = Z_{NS}^* (x_o) \text{factor} \]
This term is the estimated smoothing error:

\[ \text{OK Estimated Smoothing Error} = Z^{*}_{\text{NS}}(x_0) \cdot \text{factor} \]  

(8)

We have to prove that expressions (7) and (8) are valid. This can be done departing from a sample drawn from an exhaustive data set. Resulting smoothing errors can be compared in a scattergram where we hope both present a positive linear relationship. For lognormal kriging we can derive equivalent expressions to (7) and (8) as:

\[ \text{OLK True Smoothing Error} = Z(x_0) - \exp\left( Y^{*}_{\text{OK}}(x_0) \right) \]  

(9)

and

\[ \text{OLK Estimated Smoothing Error} = \exp\left(Y^{*}_{\text{OK}}(x_0) + Y^{*}_{\text{NS}}(x_0) \cdot \text{factor}\right) - \exp\left(Y^{*}_{\text{OK}}(x_0)\right) \]  

(10)

Once again true and estimated errors (expressions 9 and 10) should present a positive linear correlation in a scattergram.

The idea behind expression (6) can be extended to backtransform interpolation standard deviation from lognormal kriging.

\[ S^{\text{OLK}}_{\text{OK}} = \exp\left(Y^{*}_{\text{OK}}(x_0) + S_0\right) - \exp\left(Y^{*}_{\text{OK}}(x_0)\right) \]  

(11)

The ordinary kriging estimate \( Y^{*}_{\text{OK}}(x_0) \) keeps backtransforming errors within the range of original values. Expression (11) gives always positive values. We have to prove that backtransformed error (expression 11) is compatible with interpolation standard deviation. It can be done running both normal and lognormal kriging for the same data sets.

4. Materials and Methods

Three exhaustive data sets have been considered in this study. They represent three statistical distributions: lognormal, normal and a distribution with negative skewness. All these data sets are synthetic data and have been created in computer. From the exhaustive data sets samples with 121 data points have been drawn based on stratified random sampling.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Lognormal Exhaustive</th>
<th>Sample</th>
<th>Normal Exhaustive Sample</th>
<th>Negative skewness Exhaustive Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nb. of data</td>
<td>2500</td>
<td>121</td>
<td>2500</td>
<td>121</td>
</tr>
<tr>
<td>Mean</td>
<td>1.815</td>
<td>1.599</td>
<td>15.500</td>
<td>15.275</td>
</tr>
<tr>
<td>Std dev.</td>
<td>2.617</td>
<td>1.983</td>
<td>4.587</td>
<td>4.401</td>
</tr>
<tr>
<td>Coef. var.</td>
<td>1.441</td>
<td>1.240</td>
<td>0.296</td>
<td>0.288</td>
</tr>
</tbody>
</table>

Summary statistics for both exhaustive and sample data sets are given in Table 1. The sample drawn from exhaustive lognormal data can be considered as presenting lognormal distribution in which the coefficient of variation is equal to 1.240. For normal data, we also consider as normal distribution although there is a little positive skewness. Actually we have three samples presenting coefficient of variations ranging from 1.240 to 0.038.

5. Results and discussion

We have run both normal and lognormal kriging. Thus we can compare the interpolation standard deviation from ordinary kriging and backtransformed interpolation standard deviation from lognormal kriging.

Since we have departed from exhaustive data sets we can compare estimates with actual values. Figure 1 presents scattergrams for normal kriging and lognormal kriging. As we can see backtransformed values after equation (4) is always superior to the traditional approach based on equation (2). Correlation coefficients are higher and yet they present less conditional bias.

Now we can examine error derived after expression (11), if it presents the same properties as the interpolation standard deviation derived from ordinary kriging. Figure 2 presents scattergrams between interpolation standard deviation and ordinary kriging for both approaches.
As we can see on Figure 2, expression (11) allows backtransforming the interpolation standard deviation into original scale of measurement yet keeping the same properties as observed for the interpolation standard deviation associated with ordinary kriging estimates (Figure 2A, 2B and 2C).

Figure 1: Scattergrams comparing actual values with backtransformed values after Journel\(^2\) (A, B and C) and after Yamamoto\(^1\) (D, E and F). Lognormal data (A and D); normal data (B and E); and negative skewness data (C and F).

Figure 2: Relationships between interpolation standard deviation and ordinary kriging estimates (A, B and C) and between backtransformed interpolation standard deviation (equation 11) and ordinary lognormal kriging estimates (after Journel\(^2\)) (D, E and F). Lognormal data (A and D), normal data (B and E) and negative skewness data (C and F).
We can also compare errors from normal and lognormal kriging (Figure 3). As we can see on Figure 3, backtransformed interpolation standard deviations are always less than the same computed for original data. Actually, it happens because errors computed in the log domain will be smaller than those in the original domain. Note that when we take logarithm of the data the range is strongly reduced and so estimates and errors. When lognormal kriging estimates are backtransformed by taking their inverse logarithms, resulting values will present some bias in relation to sample data. Thus, it is necessary to add a non-bias term to the lognormal kriging estimate before taking its inverse logarithm. The error term proposed by Yamamoto1 is:

\[ Error\ Term = \gamma_{\text{NS}}(x_p) \cdot \text{factor} \]

Doing that we guarantee backtransformed lognormal kriging estimates are not mean and median biased (Figures 1D, 1E and 1F). Besides, we can check on Figure 1, ordinary lognormal kriging estimates present higher local precision than ordinary kriging estimates. It implies that errors in the log domain backtransformed into the original domain will be smaller than those computed directly in the original domain. That is what we see on Figure 3. Evidently, the coefficient of variation of the sample data has a direct relation with correlation between backtransformed interpolation standard deviation and the same computed from ordinary kriging. Indeed, the less the coefficient of variation the higher the correlation coefficient is.

As we have mentioned before, true and estimated smoothing errors can be compared each other. If they are comparable it means the estimated smoothing error is a good approach for the unknown true error. Figure 4 presents scattergrams for normal and lognormal kriging approaches. The only difference between these scattergrams is on lognormal data in which the correlation for smoothing error computed from lognormal kriging is improved relative to normal kriging. For normal and negative skewness data both approaches show good positive correlation. Differently to estimates, errors are very difficult to be compared. Therefore, a correlation coefficient around 0.600 can be considered well enough for providing an estimation error closer to the true error and without losing local precision. That is the most important question to be addressed.

Before drawing any conclusions about the effectiveness of the smoothing error as an approach for the true smoothing error, it is important to check other samples drawn from the same exhaustive data sets. Thus, 12 different samples with 121 data points were drawn from exhaustive data sets. The same procedure was used for computing smoothing errors as described in this paper. Resulting correlation coefficients are shown in Figures 5, 6 and 7.

For lognormal data, in general the lognormal kriging improves correlation between true and estimated smoothing errors. Moreover, samples 1 and 12 show a great improvement of Pearson’s coefficient from normal kriging to lognormal kriging. Probably it happened because Pearson’s coefficient is sensitive to outliers, on the other hand rank correlation is a more robust measure of correlation and, therefore, these samples do not show this behavior.
Figure 4: Scattergrams between ordinary kriging true smoothing error and estimated smoothing error (A, B and C) and between ordinary lognormal kriging true smoothing error and estimated smoothing error (D, E and F). Lognormal data (A and D), normal data (B and E) and for negative skewness data (C and F).

For normal data, correlation coefficients computed between true and estimated smoothing errors derived from lognormal kriging are less than those derived from ordinary kriging. Besides for samples 4 and 9 Pearson’s coefficients for errors derived from lognormal kriging are much worse than ordinary kriging. Rank correlation is more robust and therefore does not show strong influence on samples 4 and 9.

For negative skewness data errors from both methods are almost the same and in this case there is no difference between ordinary kriging and ordinary lognormal kriging.

On average correlation coefficients between true and estimated smoothing errors can be considered good enough, mainly if we are handling with errors. The estimated smoothing error is a good approach for the true smoothing error. It reconfirms the reliability of the method for correcting the smoothing effect of ordinary kriging estimates.

Figure 5: Variation of correlation coefficients between true and estimated smoothing errors for 12 samples (lognormal data): Pearson’s coefficient (A) and rank correlation (B). Legend: open circle=ordinary kriging and full circle=lognormal kriging.
6. Conclusions

In this paper we have showed a method for backtransforming errors from lognormal kriging. Now, it is possible to have an error associated with lognormal kriging estimates. According to the proposed method it is possible to backtransform interpolation standard deviation and the estimated smoothing error. Backtransformed errors have been compared with those computed directly from original data. Backtransformed interpolation standard deviation is always less than the same derived from ordinary kriging, because errors in the log domain tend to be reduced since log transformation reduces the data variation. In this sense, lognormal kriging improves local precision, in terms of correlation with neighbor data. Indeed, the higher the local precision the smaller the interpolation standard deviation is.

Another error term that can be derived not only from lognormal kriging but also from ordinary kriging is the estimated smoothing error. The estimated smoothing error present positive correlation with the true smoothing error, proving that this approach is valid for determining the error term associated with kriging estimates. Actually, this error term can also be simulated and it should present a positive correlation with the true smoothing error.

Finally, it was shown that lognormal kriging can be applied for non-lognormal data getting improved estimates in terms of local precision.

7. Acknowledgements

I am very grateful to the National Council for Scientific and Technological Development – CNPq (Process 303505/2007-9) for funding this research.
8. References