Review of Techniques to Calculate Uncertainty in the Mean

Martha E. Villalba and Clayton V. Deutsch

A review of the current techniques to assess uncertainty in the mean is developed. The conventional bootstrap technique assumes independence between the data and permit evaluation of uncertainty by drawing samples randomly with replacement. The spatial bootstrap technique allows assessment of uncertainty by resampling original data accounting for their spatial correlation. However this does not consider the option of conditioning to the data and considering the domain size of interest. The conditional finite domain technique accounts for the spatial correlation between samples and the data.

Introduction

Mineral deposits, petroleum reservoirs and environmental sites are uncertain because of relatively few data. Geostatistical simulation can be used to build multiple realizationes, that will help uncertainty. Input parameters are required in addition to the local data; parameters such as the histogram, variogram, and implementation decisions. The mean of the histogram is perhaps the most important because it has a direct influence on resources and reserves.

A realistic evaluation of uncertainty is important during the planning of operation. Priority may be given to mining zones of low uncertainty (categorized as proven and probable) and building access in zone of waste or mineral with high uncertainty (categorized possible). A better evaluation of uncertainty could avoid problems during the production and help planning.

Three different techniaues of uncertainty evaluation are available. All of them are based upon the assumption that the distribution of the data is representative of the whole population. The uncertainty in the mean is measured after resampling the data. The sample is drawn randomly from the cumulative distribution function (cdf) of original data, taking account the spatial correlation, conditioning to data and the domain.

The Conventional Boostrap CB, Spatial Bootstrap SB and Conditional Finite Domain CFD are the techniques to calculate the uncertainty in the mean and has been developed broadly in this paper. The main feature of CB is that it assumes independence between the data. SB considers the spatial correlation between the data and CFD use conditioning data and the domain of interest.

Conventional Bootstrap

The data are assumed to be independent. The Bootstrap samples with replacement from the data and the sub samples are taken from the same distribution of the population, every sample is drawn independently from the others samples. The Bootstrap may be useful when we need to measure the uncertainty in the mean early in appraisal with widely spaced data (Deutsch, 2002). The methodology is as follows:

1. Assemble the representative histogram of the n data to obtain the distribution of the Z random variable $F_Z(z)$, consider the use of declustering method for irregular grid and debiasing if appropriate.
2. Simulate n values from the distribution of n data. Generate n uniformly distributed random numbers p_i, $i = 1,\ldots,n$ and read the corresponding quantiles $z_i = F_Z^{-1}(p_i)$, $i = 1,\ldots,n$
3. Estimate the mean for the resampled set of data and save it as one possible mean.
4. Return to the second step several times (example L=1000 times)
5. Assemble the distribution of L possible means in order to evaluate the uncertainty in the mean.

The assumption of data independence does not apply in the presence of spatial correlation and could generate too little uncertainty.

The following relation follows directly from the assumption that the data have stationary mean values:

$$Var\{\bar{X}\} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} Cov\left(X_i, X_j\right) \qquad (1)$$

Provided the data are independent, the variance of the mean can be simplified to $\frac{1}{n^2}\sum_{i=1}^{n}\sigma_x^2$ or $\frac{\sigma_x^2}{n}$.

This last expression can be used to calculate the effective number of data. $n_{eff} = \frac{\sigma_X^2}{\sigma_{\bar{X}}^2}$. If the data are independent, the the effective number of data is equal to the number of data, but in presence of spatial correlation, the effective number of data is less.

Spatial Bootstrap

The Spatial bootstrap is an improvement of the bootstrap concept (Deutsch, 2004) where the spatial correlation of the data is considered. SB uses unconditional simulation and the covariance function. Consider:

1. Assemble the representative histogram of the data, consider the use of declustering method for irregular grid and debiasing method if appropriate.
2. **y** is the resulting *n* by 1 vector of unconditionally simulated values with the correct covariance. The superscript *l=1,…, L* denotes one of the large number, *L*, of realizations we are generating. These Gaussian values can be converted to probability values to draw from the representative distribution. $\mathbf{p}^{(l)} = G(\mathbf{y}^{(l)})$, $l=1,\ldots,L$
3. Where $G^{-1}(\cdot)$ is the inverse of the standard normal distribution and **p** is an *n* by 1 vector of probability values [0,1]. The drawn z-values are calculated as: $\mathbf{z}^{(l)} = F_Z^{-1}(\mathbf{p}^{(l)})$, $l=1,\ldots,L$
4. where **w** is a *n* by 1 vector of independent Gaussian values and Once the LU decomposition is performed (required only once), the generation of the simulated realizations can be calculated simply as: $\mathbf{z}^{(l)} = F_Z^{-1}(G(L\,\mathbf{w}^{(l)}))$, $l=1,\ldots,L$.
5. Assemble the distribution of several L possible means in order to evaluate the uncertainty in the mean.

The spatial correlation generates a larger uncertainty that the CB technique.

Conditional Finite Domain

The CFD method is a stochastic approach based on a multivariate Gaussian Model in order to determine the uncertainty in the mean, which is convergent, design independent and invariant under parameterization (Babak and Deutsch, 2007). This technique permits evaluation of uncertainty by sampling L set of configuration of the data previously simulated. Thus, *L* sets are assembled based on translation and rotations of the locations and must have the same configuration as the conditional data (X_i,Y_i). The simulation in this set of configurations *L* is called order *K*; therefore the number of realization *(j)* by each order *K* is the same number of possible random rotation and translated locations *(Xrd(i,j),Yrd(i,j)). K* orders are simulated and the difference between them is the reference distribution taken from *(k-1)* simulation. The reference distribution is gathered from $Y2_{(k-1)}(i,j)$ and conditioning simulation is performed.

The uncertainty is quantified through the histogram of the possible means of every order *(k)*, also, the uncertainty of any order are illustrated in a graphic of standard deviation against number of order with the target to verify the convergence, the point of convergence is the uncertainty of the mean. CFD has the characteristic of finite domain because the process of sampling is done from L configurations set before starting the k order simulation, otherwise conventional bootstrap technique sample randomly from cumulative distribution function CDF.

CFD leads to less uncertainty than the SB. This is reasonable since CFD results are more related because of the conditioning.

Conditional simulation is done at every location to be sampled. A number of data combinations, say *L*, are generated using translation and/or rotation to have similar configuration to the original data and belong in the study domain. The translation is relative to the domain size or spacing data.

The *L* data configurations are the same for every k order and CFD technique does sampling and evaluate uncertainty of possible mean in base on L**K* data combinations.

Parameter uncertainty of order 0 is calculated using LU conditional simulation with unique reference distributions from the original data for the whole number of configurations simulated. Subsequent order *k* uses reference distributions from the previous order. Every simulation is conditioned to the original data.

The CFD approach for evaluating uncertainty in the mean is convergent in some *k* order. Rely on the series of simulation and its stochastic algorithm. So similar to Markov Chain and Monte Carlo approach, it is observed periods of "burn-in" and a period of stabilization where is recognized a fluctuation of the uncertainty around some constant value. In fact is defined the expected uncertainty. The algorithm of CFD is as follow:

1. Assemble the representative histogram of the data, consider the use of declustering method for irregular grid and deviasing method for identical sample in the data if is appropriate.
2. Set L number of data configurations *Xrd(i,j),Yrd(i,j))* using translation and rotation.
3. *K* = 0 Apply LU conditional simulation with the same reference of the original data. It is simulated *J* equiprobable realizations of the variable of interest in L new configurations; new references distribution by each configuration is assembled for the next step.
4. *K*= 1 Apply LU conditional simulation with reference distributions established in *K-1* to create *J* new equiprobable realizations of the variable of interest in L configurations and conditioned to the original data. Update new reference distributions for the next order.
5. The step 4 is repeated until reach the number of order *K* desired.
6. Through the whole process, mean and variance is stored every simulation J of L configurations and *K* order.
7. Evaluation of the uncertainty in the mean from the distribution of the possible means founded in L configurations (J simulations) and k orders.

Only translations and not rotation should be used in presence of strong anisotropy. The values for the tail of the distributions are quite important. The data distribution needs to be declustered, since the simulation is conditioned to the data and done only in the point to be sampled. Therefore the declustering weights are relevant in the process.

Conclusions

Three different techniques of evaluation of uncertainty in input parameter are available so far, all of them are strongly based upon the assumption that the distribution of the data is representative of the whole population Conventional bootstrap technique assumes independence between the data and permits evaluation of uncertainty by drawing sample randomly with replacement, from the entire original data. Spatial bootstrap technique allows assessment of uncertainty by resampling original data accounting for their spatial correlation. However this does not consider the option of conditional to the data through the simulation process. Conditional finite domain takes account the spatial correlation between samples and the simulation is conditioned to the data, thus the sampling is done from the domain identified by L configurations, the conditioned values to the data sampled are assembled in a histogram or table of uncertainty versus order where it is identified the uncertainty in the mean. Greater spatial continuity (larger range and/or smaller nugget effect) leads to more uncertainty with SB and less uncertainty with CFD. This is reasonable since CFD results are more related because of the conditioning.

References

Babak, O., Deutsch C.V. *Reserves Uncertainty Calculation Accounting for Parameter Uncertainty*, Journal of Canadian Petroleum Technology, Vol. 47, No 8, pp. 41-49, 2008

Babak, O., Deutsch C.V. *Accounting for Parameters Uncertainty in Reservoir Uncertainty Assessment: Conditional Finite Domain Approach*, Natural Resources Research , Vol. 18, No 1, pp. 7-17, March 2009

Deutsch, C.V., , *Geostatistical Reservoir Modeling*, pages 65-70, Oxford University Press, New York, 2002

Deutsch, C.V., *A Statistical Resampling Program for Correlated Data, Spatial_Boostrap*, Department of Civil & Environmental Engineering, University of Alberta, September 2004.

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

Leuangthong O. , Khan K.D., Deutsch C.V., *Solved problems in Geostatistics,* Department of Civil & Environmental Engineering, School of Mining and Petroleum Engineering University of Alberta, Canada, Chevron Energy Technology Company, Houston, TX,pp.47-57, 2008

Weiss E., *Bootstrapping with SAS*, Winchendon Group INC., 3907 Lakota Road, Alexandria, Virginia 22303.