A Short Note on the Application of Copulas to Multivariate Probability Density Estimation

Sahyun Hong

A copula function method is applied to a facies modeling case. The method is to model the multivariate relation in the transformed unit space. An advantage of the method is that the multivariate copula function has uniform marginal distribution, and a combination of copula function families allows us to model the multivariate relations in a parametric way. This note introduces some disadvantages of the copula method when applying to geostatistical studies, particularly to the facies modeling case.

1. Introduction

The copula function is developed to model the multivariate relation in the mapped space. Each variable is transformed into the unit interval and the multivariate distribution of the transformed units is modeled. One advantage of this approach is that some of copula families have been greatly studied and a set of those copula functions allow us to model the multivariate relation parametrically. Another advantage of copula method is that one does not have to consider the marginal distributions because the transformation into the unit interval make all marginal distributions uniform. Despite these advantages, copula function approach may not be desirable in geostatistics. This short note addresses some problems of copula method using facies modeling example.

2. Copula Function

Consider two variables, X and Y. Each variables have identical quintiles falling in [0,1] such as,

$$U = F_X(x) \;\; and \;\; V = F_Y(y)$$

A copula function models the relation between U and V. Once the multivariate copula function is modeled, then the copula function in the unit space can be converted to X and Y space. Some copula family functions are widely studied and in some cases, their combination can parametrically model the multivariate distribution in the unit space. Because copula method uses the mapped unit on [0,1], all of marginal distributions are uniform. Different from the other multivariate modeling problem, the copula approach does not impose marginal constraints.

3. Example

Consider the facies modeling problem. Two continuous variables (Y1,Y2) are assumed to be given to identify facies over the domain. Figure 1 is a scatter plot of the transformed unit of two variables identified by two facies types (blue circles and red triangles). A bivariate copula function is parametrically modeled using given the samples and the modeled copuls function is shown in Figure 2.

α in Figure 2 is a control parameter of the selected Gumbel copula function which controls the dependence between U1 and U2 transformed variables. The modeled copula function in the left is for facies 1 and the copula function in the right is for faceis 2. Although the facies 1 and facies 2 samples have somewhat separation, the modeled copula function should have uniform marginal distribution. α parameter in the Gumbel function is selected to best match with the given samples and uniform marginal distribution. Based on the modeled copula function, the conditional probability can be directly derived. Figure 3 is a resulting conditional probability of faces 1. The histogram of the estimated facies probability has unimodal distribution which can be interpreted as that the given continuous data sets are not very helpful to the facies modeling. This is due to that copula function should have uniform marginal distribution which dilutes the separability of the given Y1 and Y2 variables. In original variable space, two variables Y1 and Y2 have distant means and variances for each faceis 1 and 2. However, the transformed units (U1,U2) have less separability in order to have uniform marginal distribution. Figure N shows another result obtained from the kernel estimation method; histogram of the facies probability is bimodal distribution which gives us more clear identification of two facies.

4. Comments

Copula functions are widely used in the multivariate modeling problem due to some desirable properties. If the variables are not separated by different clusters or groups, then the copula method could be a useful choice for the multivariate modeling. However, if the separability of the given variables is more critical such as facies modeling illustrated in this note, then copula approach may not be a good approach.

References

Nelsen, R., 1999, An introduction to Copulas, Springer Verlag, New York.

Lan, Z. and Leuangthong, O., 2005, On the use of copulas for multivariate distribution inference for discrete cosimulation, *Annual Report 7* in Centre for Computational Geostatistics

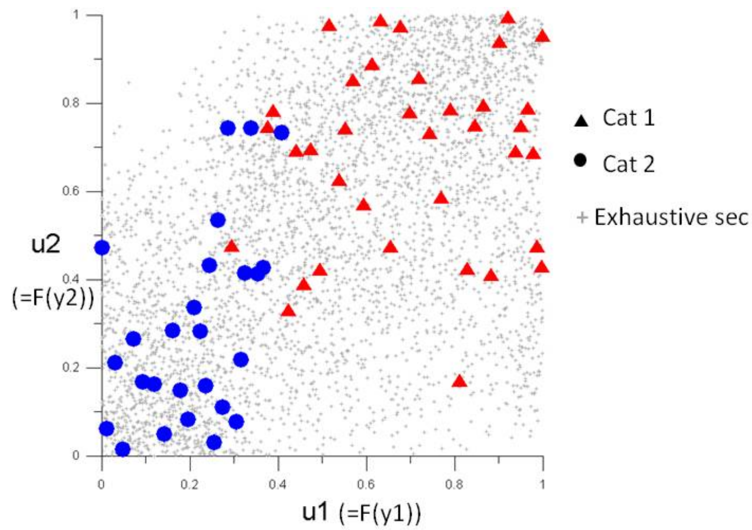

Figure 1: Scatter plot of the transformed Y1 and Y2 variables on the unit interval.

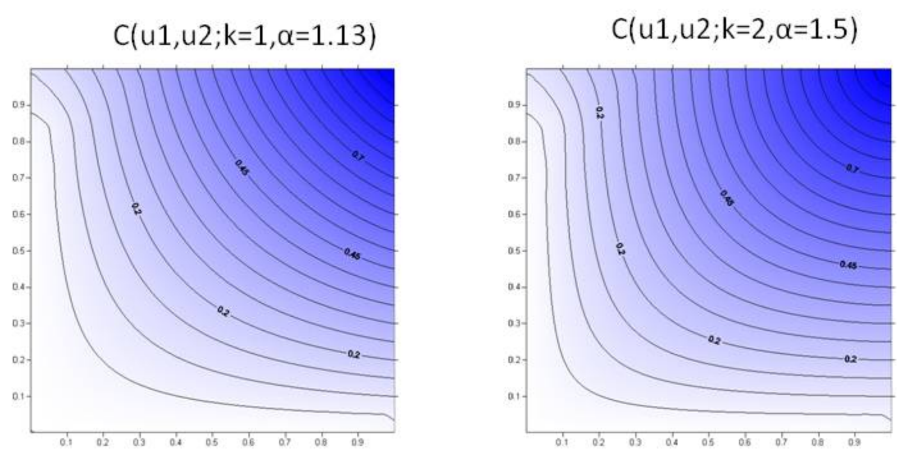

Figure 2: Modeled copula function for each facies. Gumble parametric function is selected for the bivariate modeling.

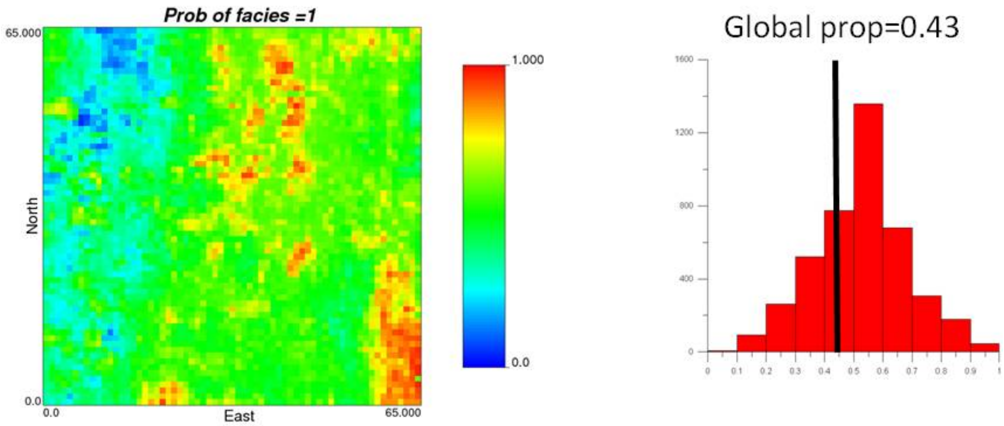

Figure 3: Estimated facies probability from the modeled copula function.

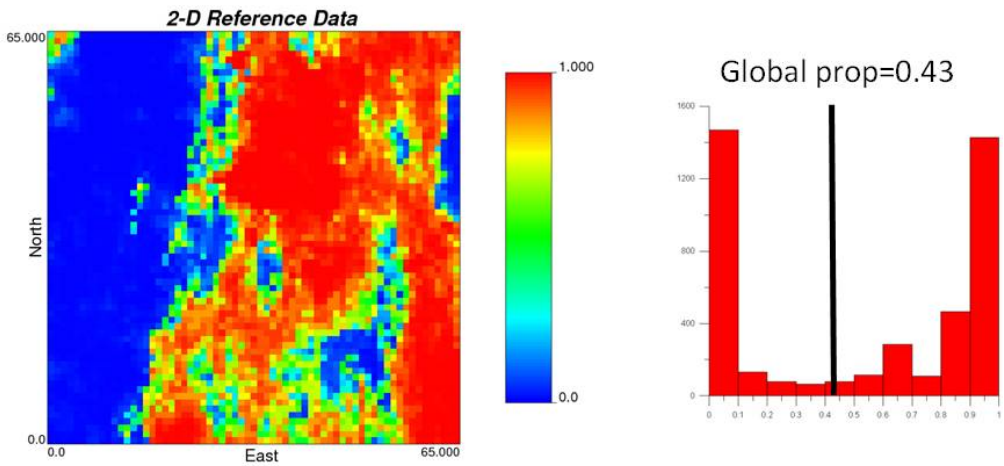

Figure 4: The estimated facies probability from the other method(kernel estimation method).