Bivariate Modeling: A Modified bimodel program

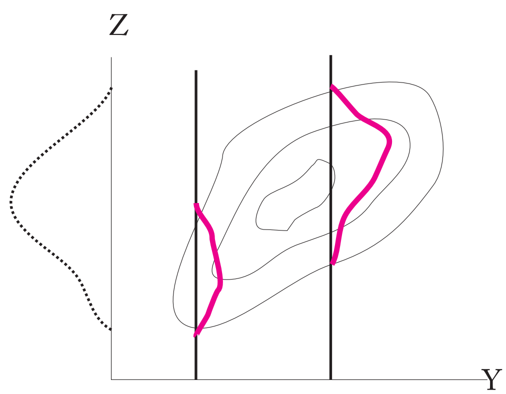
Sahyun Hong

The Bimodel program is slightly updated. A smooth kernel method is added as a modeling option, and some theoretical conditions are checked. The current version requires the user input bin size and maximum data number used for conditional bivariate distribution at the bins. The modified program finds the optimal bin size and maximum data number under theoretic relations that a series of conditional means and variances should satisfy. This optimization is active only when the nonparametric option is flagged. Kernel method or other Gaussian modeling options do not need this procedure.

1. Introduction

A GSLib program bimodel.exe is for modeling the bivariate relation using the limited bivariate data. There are three options available in the program: (1)nonparametric modeling, (2)bivariate Gaussian, and (3) conditional Gaussian. We slightly updated the first option, nonparametric modeling. The current version models the bivariate distribution by counting the number of samples falling within the user specified bin width and the maximum number of samples. As the maximum number of samples is larger, the modeled bivariate distribution becomes smoother in distribution. The bin width and maximum number of samples are user input options in the current version. We added a kernel density estimation option for nonparametric modeling. One critical factor of kernel method is the size of smoothing window. It is automatically optimized based on data statistics in the implemented program.

Another change to the current program is to check the conditional mean and variance. There are well-known relations between the global mean/ variance and conditional mean/variances,



$$\mathrm{var}\{Z\} = E\{\mathrm{var}(Z \mid Y)\} + \mathrm{var}\{E(Z \mid Y)\}$$
$$E\{Z\} = E\{E(Z \mid Y)\}$$

Figure 1: Theoretic conditions that the conditional statistics must meet.

where Y is the independent variable (e.g. porosity) and Z is the dependent variable (e.g. permeability). These theoretical relations are automatically met when the option (2) and (3) are chosen for the bivariate modeling. However, the first option is chosen then these relations should be checked. In a modified version of bimodel.exe, the bin width and number of samples used for the bivariate modeling are optimized based on the conditions shown above. Simple iteration is implemented to find the optimal bin size and sample number. The kernel density estimation option does not require the condition check because the method itself meets those conditions.

2. Program: new_bimodel.exe

There are no major corrections in the new_bimodel.exe. One option for bivariate modeling is added and theoretical check is implemented. Figue 2 shows an example of the parameter file for the new program. Only difference from the current version is the line 19; kernel density option is added. The second value in this line is a smoothing factor. Although the program calculates the optimum kernel smoothing width, users can adjust the degree of smoothness by changing this value. Value being larger than 1 makes the bivariate distribution smoother, and vice versa.

```
              Parameters for BIMODEL
              **********************

START OF PARAMETERS:
FMI-K.data                    -file with paired data (optional)
3  4  0                       -  columns for X, Y, and weight
-1.0e21    1.0e21             -  trimming limits
../Examples/ydata.dat         -file with representative X (optional)
4  0                          -  columns for X variable and weight
../Examples/quantiles.dat     -file with conditional quantiles (optional)
1  2  3                       -  columns for X value, ccdf, Y value
1.0                           -  width for quantile values (in X units)
 15   0.0    15.0    0        -X ndis, min, max, (0=arith, 1=log)
 60   0.0    60.0    0        -Y ndis, min, max, (0=arith, 1=log)
1                             -type: 1 non param, 2 biv norm, 3 cond norm
0  4.0  12.0  4.0  12.0  0.5  -type 2 - 1 -> specify xm, xv, ym, yv, rho
10.0   25                     -bins: size, max data
0                             -increasing (1), decreasing (-1), no (0)
4   0.8                       -kernel method?, smoothing multi factor
../Examples/ex01-biv.out      -file for output bivariate distribution
../Examples/ex01-cond.out     -file for output conditional statistics
../Examples/ex01-ydist.out    -file for output Y distribution
../Examples/ex01.ps           -file for output PostScript
```

Figure 2: An example of parameter file for the new_bimodel.exe program

3. Examples

The FMI data is used for the test. Porosity and permeability are considered as independent and dependent variables. Figure 3 illustrates of the bivariate models obtained from kernel density estimation in the left, non param without conditional mean and variance check(currently implemented version) in the middle, and non param with conditional mean and variance check in the right. It is difficult to decide which bivariate model is better than other model, however, distribution itself should be licit distribution. The conditional mean and variances are calculated from three different bivariate distributions and compared to the global mean and variances.

	E{Z}=361.9, Z is a perm	Stddev{Z}=288.5
Kernel	360.4	296.1 (3%)
Non param without condition check	355.6	317.35 (11%)
Non param with condition check	317.1	292.1(2%)

Numeric values in the table are calculated by a series of the conditional distribution. In theory, those values should correspond to the global mean and variance. The result of kernel estimation is greatly affected by the used kernel bandwidth size. Figure 4 shows a bivariate model using a large kernel window. The model seems over-smoothed. But, the theoretic conditions are still satisfied.

Another example is shown in Figure 5. Similar to the previous example, three modeling options are compared, and their summary statistics of the conditional distribution is described in the table below.

	E{Z}=0.78, Z is a Kv/KH	Stddev{Z}=0.278
Kernel	0.76	0.264 (5%)
Non param without condition check	0.732	0.208 (25%)
Non param with condition check	0.78	0.276

Examples show that the current bimodel.exe program rarely meets the theoretical relations, which means the modeled bivariate distribution is not a licit distribution.

9. Discussions

The bimodel.exe program is modified. A kernel method is added as another modeling method, and theoretic conditional mean and variance check is implemented. Thus, if non param option is selected then the bin size and sample number is optimized based on the discussed conditions. This optimization is active only when the non param option selected (not applied to kernel option).

References

Deutsch, C.V. and Dose, T., 1998, Programs for debiasing and cloud transformation: bimodel and cltrans*, Annual Report 1* in Centre for Computational Geostatistics*.*

Bilingsley, P., 1995, *Probability and measure*, New York, NY, John Wiley & Sons.

Figures

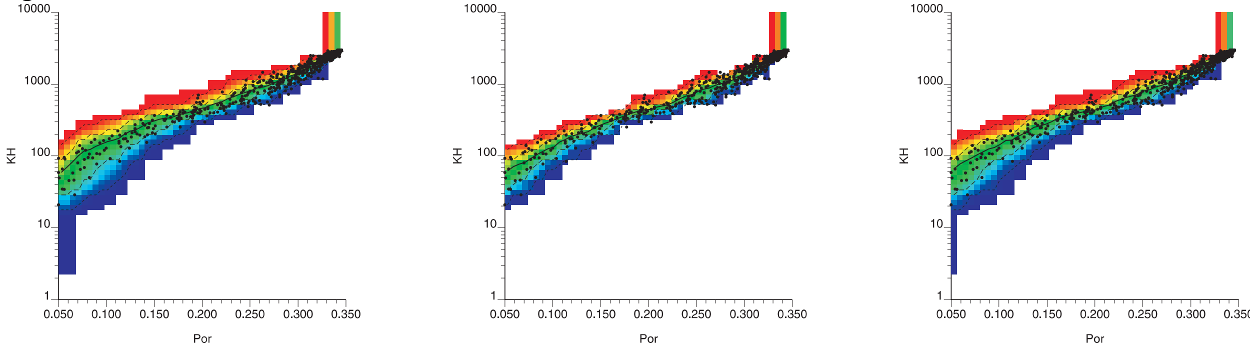


Figure 3: Three different bivariate distribution models: kernel method (left), non param without condition check (middle) and non param with condition check (right).

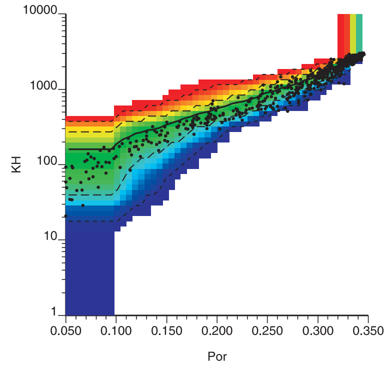


Figure 4: Over-smoothed bivariate distribution by kernel method

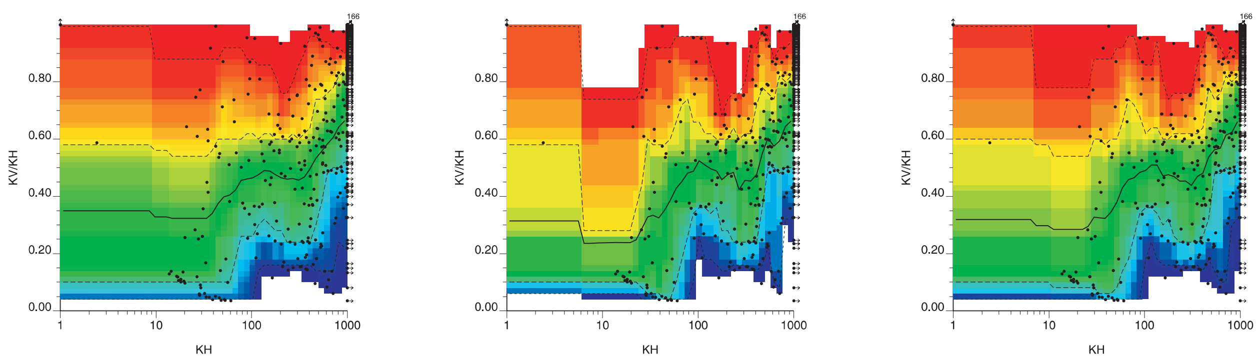


Figure 5: Another example of bivariate modeling: kernel, non para with and without condition check.