Emulator for Hazard Analysis: A Usage tutorial

University at Buffalo

15-17 March , 2016

OUTLINE

Background

Direct Computation of Elevation height and Particle flux may be computationally very expensive.

For handling the complexity, a simple Emulator could be constructed. Some features of the presented model are:

- Assumes the underlying distribution as a Gaussian Stochastic process
- Capable of studying the correlation between the input parameters
- The parameters for the fitted model can be computed through MLE as well as REML
- For providing an intuition regarding accuracy, the display panel shows the location of the LHS generated points on the mean surface for the gaussian distribution.
- For validating the model, the emulator also provides a visual description of the residuals of the sample points along with the standard errors for sample as well as LHS points

PErK and MPErK (a MATLAB program for Parametric Empirical Kriging)

Freely available code accompanying Santner, Williams and Notz text.
Views computer experiments as gaussian stochastic process.

$$Y(x) = \beta^T f(x) + z(x)$$

f(x): px1 vector of regression functions
β: Unknown vector of regression parameters
z(x): zero mean covariance

For any inputs x1 and x2

$$Cov(z(x1), z(x2)) = \sigma^2 R((x1 - x2)/\epsilon)$$

ϵ can be computed through MLE or REML
R could represent Gaussian, PowerExponential or Cubic correlation functions

Emulator Usage

Using the constructed emulator is fairly simple and involve following steps:

Input:

- Input the data as Follows
  - The data should be passed in a .csv format
  - There should be four columns in the csv file in the order: Water Fraction, Temperature, followed by corresponding Eruption height and Particle flux
  - For passing the data, just click the 'UPLOAD' button in the emulator and select your .csv file (in All Files option)
- Select the type of Correlation functions which seem to be most suitable for the data
- Selection the optimization criterion for determining the correlation parameter
- Cross validation could be carried out based on the options mentioned in analysis part

In order to analyze the performance of the constructed model, following visualization could be used

Analysis:

- Standard Error for the LHS generated dataset
- Standard Error for the Input sample
- Residuals for the Input sample

Based on the visualizations for the gaussian mean surface as well as plots for the residuals and standard error, the decisions regarding the underlying correlation structure could be made

The value for Mean residuals also help in determining the most suitable correlation function

Prob(Eruption height <'Height')

- Enter the 'Height' to be inspected (in meters)
- In the background, the model generated samples from a Uniform distribution with the same parameters as done by Puffin
- When the 'Calculate' button is clicked, these random samples are evaluated for eruption height through the coefficients of the Gaussian fitted model
- Based on the proportion of the samples, which produce eruption height less than the entered height, the probability value is computed

References

- T. J. Santner, B. J. Williams and W. I. Notz (2003). The Design and Analysis of Computer Experiments, Springer Verlag, New York.
- Runze Li and Agus Sudjianto (2005). Analysis of Computer Experiments Using Penalized Likelihood in Gaussian Kriging Models, Technometrics, Volume 47, number 2, pages 111-120.

Thank You ...