AOE/CS/ME 6444 Verification and Validation in Scientific Computing
Spring 2020
Instructor: Dr. Chris Roy

Homework #5
Due Wednesday April 29, 2020 at 10pm

Please upload your homework assignments in PDF format to the appropriate assignment section of Canvas. Please also use the following file naming convention: VVSC_Lastname_Firstname_HW**5**.pdf. If you have problems getting your homework into PDF format, then let me know.

This homework involves computing a validation metric for your chosen scientific computing application. There are two options to choose from: 1) obtain experimental data relevant to your application from the literature (strongly preferred) or 2) generate "synthetic" experimental data via the process outlined below. You will then use these data to compute the Area Validation Metric (AVM) or the Modified Area Validation Metric (MAVM) for the case with one aleatory uncertain model input.

Option #1: Experimental Data from the Literature

If you have access to experimental data with multiple measurements of the SRQ for your application (either from your lab or inferred from uncertainty bars for data found in the literature), then you may use that data to replace the "synthetic" data discussed here. If using actual experimental data, you may need to estimate the uncertainty in one of your input values if it is not given. A complication for this option is that most researchers report 90% or 95% confidence intervals for the mean values rather than the actual distribution of values. Thus you may need to either estimate the distribution parameters (given the confidence intervals and the number of data replicates) or talk to an "expert" about what they believe the actual distribution might look like. You will need to do this for both the SRQ and the aleatory input parameter. If the 95% confidence intervals are reported, and you know how many replicate experiments are available, then you can back out the distribution by assuming it is Gaussian.

Option #2: Synthetic Experimental Data

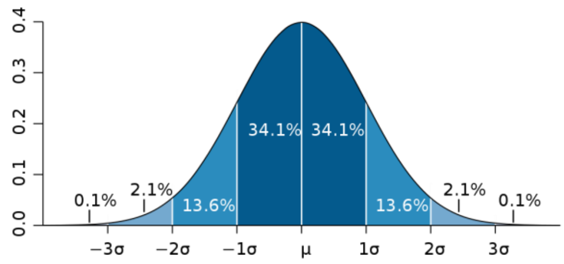
If you do not have easy access to experimental data, then you will need generate two different sets of "synthetic" experimental data using the following process. To begin, choose a model input parameter to be an aleatory uncertainty. You will want to choose the standard deviation of this distribution to be sufficiently large to cause significant variations in your SRQ. Assume this random variable to be normally distributed (i.e., Gaussian) according to the following CDF

$$\frac{1}{2}\left[1+erf\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$$

where x is the parameter value, μ the mean value, and σ the standard deviation. The PDF of this model input is thus given by

$$\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

and has the following shape in terms of probability density versus *x*:



Choose (and report) the mean value and the standard deviation of your model input in some physically plausible way for your chosen problem of interest.

In order to generate your first "synthetic" experimental data set, choose your model input parameters as $x = \mu \pm \sigma$, and then compute the resulting two SRQs. Make sure you have chosen σ large enough to obtain significant variation in your SRQ. Denote the smaller value of this SRQ as α and the larger value of this SRQ as β. Create your first synthetic experimental data set from the relationship SRQ = $\alpha + \chi(\beta - \alpha)$, where χ is the vector given by:

$$\chi_1 = [0.55, 0.95, 1.0, 1.1, 1.5]$$

The second synthetic data set will be given by:

$$\chi_2 = [0.1, 0.4, 0.6, 0.75, 0.8, 0.9, 0.91, 0.97, 1.3, 1.6]$$

Validation Metric Computation

Regardless of the option you choose above, you will need to compute your AVM/MAVM as follows. Sample over your normally distributed aleatory model input to generate sample sizes of 10, 25, and 100 model inputs. In order to perform this sampling, I recommend that you use Latin Hypercube Sampling (LHS) rather than traditional Monte Carlo Sampling (MCS) in order to propagate the uncertainties more efficiently. If you are using MATLAB, you can use the `lhsnorm()` function for LHS sampling from the normal distribution. *Note that the sigma value that the `lhsnorm()` function requests is actually the variance of the normal distribution (standard deviation squared).* Propagate the input uncertainty through your model by computing SRQs for each of your model inputs found from the sampling procedure. If your scientific computing application is too expensive to compute 100 cases, you may use fewer cases to build a response surface approximation of your input/output relationship. For this 1D case, this could be a simple polynomial regression fit. Choose an appropriate order polynomial or maybe even use a spline fit.

Compute the AVM/MAVM for your experimental data set(s) using the three simulation sample sizes. Show your results graphically (as a comparison of the area between the two CDFs/EDFs) and state the computed AVM/MAVM value for each case.

Draw conclusions regarding the simulation sample sizes needed to accurately compute the MAVM (and therefore the model form uncertainty).