10th World Congress on Structural and Multidisciplinary Optimization
May 19 - 24, 2013, Orlando, Florida, USA

Characterizing Probability-based Uniform Sampling for Surrogate Modeling

Junqiang Zhang, Souma Chowdhury, Achille Messac

Syracuse University, Syracuse, New York, USA
jzhang62@syr.edu, sochowdh@syr.edu, messac@syr.edu

Abstract

The fidelity of surrogate models remains one of the primary concerns in their application to represent complex system behavior. Appropriate sampling of training points is one of the primary factors affecting the fidelity of surrogate models. This paper investigates the relative advantage of probability-based uniform sampling over distance-based uniform sampling, in training surrogate models whose system inputs follow a distribution.

The inputs to a surrogate model can be variables with known, assumed, or predefined probability of occurrence. Conventional distance-based sampling involves metrics defined in terms of coordinate-distances in the actual variable space. This paper instead uses the probability of the variables as the metric for sampling. This sampling approach inversely transforms a designated sequence of probabilities to their corresponding values using inverse Cumulative Distribution Function (CDF)[1]. Probability-based sample points thus determined are used as training points to develop the surrogate model.

If the sequence of probabilities is uniform, the obtained probability-based sampling points are uniform in terms of probability. To obtain the same number of distance-based uniform sample points, the same sequence is directly scaled into the coordinates of training points.

To study the suitability of probability-based uniform sampling for surrogate modeling, Mean Squared Error (MSE)[2] of a monomial form is formulated based on the relationship between the squared error of a surrogate model and the volume or hypervolume per sample point[3]. Using varying size of the uniformly-distributed training data set, the squared errors of the surrogate models are fitted as a monomial function of the hypervolume per sample point.

The fidelities of the two surrogate models developed using probability-based and distance-based uniform sampling are compared using the monomial MSE function. The exponent of the monomial MSE function indicates which of the two surrogate models provides higher fidelity. When the exponent of the monomial function is between 0 and 1, the fidelity of the surrogate model trained using probability-based uniform sampling is higher than that of the surrogate models trained using distance-based uniform sampling. When the value of the exponent is greater than 1, the fidelity comparison is reversed. This theoretical conclusion is successfully verified using standard test functions.

In this paper, the probability-based uniform sampling is also applied to the development of surrogate models for the thermal performance evaluation of windows. The conditional advantage of probability-based uniform sampling (formulated in this paper) also holds true for the window problem.

[1] F. P. Miller, A. F. Vandome, and M. B. John. Inverse Transform Sampling. VDM Publishing, Saarbrcken, Germany, 2010.
[2] E. L. Lehmann and G. Casella. Theory of Point Estimation. Springer, New York, 2nd edition, 1998.
[3] J Zhang. Characterizing Probability-based Sampling for High-fidelity Surrogate Modeling. PhD dissertation, Rensselaer Polytechnic Institute, 2012.