11th World Congress on Structural and Multidisciplinary Optimisation
07th -12th, June 2015, Sydney Australia

Comparison Study of Statistical Modelling Methods for Identifying Distribution Types

Young-Jin Kang[1], Yoojeong Noh[2], O-Kaung Lim[3]

[1] Pusan National University, South Korea, zmanx@pusan.ac.kr
[2] Pusan National University, South Korea, yoonoh@pusan.ac.kr
[3] Pusan National University, South Korea, oklim@pusan.ac.kr

Abstract

In mechanical systems, statistical modelling of input random variables is necessary to quantify uncertainties arising from inherent variability in material properties and environment variation for accurate analysis of mechanical system behaviour. Some statistical modelling methods such as goodness-of-fit tests assess whether given data are consistent with a hypothesized null distribution. Other statistical modelling methods such as model selection methods select the best fitted distribution by comparing given data with certain candidate probability density functions. However, the statistical modelling methods have difficulties in identifying a right distribution type when the number of data is limited, which often occurs in real applications. Thus, it is necessary to understand the strength and weakness of each method and develop a better statistical modelling method if necessary.

This research tested statistical modelling methods such as Kolmogorov–Smirnov(K-S) test, Anderson-Darling(A-D) test, maximum likelihood estimation(MLE), Akaike information criterion(AIC), and Bayesian information criterion(BIC). Through simulation tests, it was observed that they often selected wrong distributions especially for small number of data because they depend on distribution parameters obtained from given data [1]. In this research, a Bayesian method was proposed to overcome the disadvantages of those methods. The Bayesian method uses a maximum likelihood function and integrates it over distribution parameters and selects a distribution with the highest integration value among candidate distributions, which makes the Bayesian method less depend on distribution parameters [2]. To verify it, statistical simulation tests were carried out for various distributions such as normal, log-normal, and logistic distribution. As a result, the Bayesian method selected more accurate distribution types than other methods when the numbers of samples are 10, 20, and 30. As the number of samples increase, all model selection methods have similar accuracy in identifying a right distribution.

Acknowledgement

This work was supported by the national Research Foundation of Korea (NRF) Grant funded by the Korea Government (NRF-2013R1A1A3013500) and by the Technology Innovation Program (10048305, Launching Plug-in Digital Analysis Framework for Modular system Design). This support is greatly appreciated.

References

[1] C. Kahraman, E.E. Kerre, F.T. Bozbura, Uncertainty Modeling in Knowledge Engineering and Decision Making, World Scientific, Singapore, 2012
[2] Y. Noh, I. Lee, K.K. Choi, Identification of Marginal and Joint CDFs Using Bayesian Method for RBDO, Structural and Multidisciplinary Optimization, 40, 35-51, 2010