Statistical Modeling for Reliability Assessment Using Rubber Stiffness Data of the Automotive Engine Mount

Minho Joo[1], Jaehyeok Doh[1], Jongsoo Lee[1], Yongsok Jang[2], Hongsok Jang[2] and Jongchan Park[2]

[1]*School of Mechanical Engineering, Yonsei. University, Seoul, 03722, Republic Of Korea*
cptjoo@yonsei.ac.kr
jhdoh87@yonsei.ac.kr
jleej@yonsei.ac.kr

[2]*Hyndai Motor Group 150 Hyndaiyeonguso-ro, Namyang-eup, Hwasung-si, Gyeonggi 18280, Korea*
ysjang76@hyundai.com
paperclock@hyundai.com
impactpark@hyundai.com

ABSTRACT

Input variables used in mechanical system analysis or Prognostics and Health Management (PHM) are various, such as life of material, size and property of product. It has uncertainties (Aleatory uncertainty) due to various noises like temperature, mechanical error and various environmental differences during experiments in real field. For example, obtained data by conducting repeatedly the experiment in the identical environment condition have to be same theoretically. However, experimental data have the variation due to the generated error or noise by the various factors and uncertainties. In this study, the improved algorithm how to determine proper distribution of input variable containing uncertainties is proposed using the probabilistic and statistical method. Also, the validity of this improved algorithm is verified using the automotive engine mounts rubber stiffness data.

1. INTRODUCTION

PHM technology consists of three steps: data attribute factor extraction, health diagnoses and failure prediction. If the first step such as the data attribute factor extraction isn't designed adequately, the reliability of the final resultant in PHM can be low, even if the second and third steps are designed well. This shows how important the process of data attribute factor extraction in PHM technology is. Data attribute factor extraction is the process of extracting useful information associated with system status and failure from raw data. Raw data can be divided into numbers, waveforms, and multidimensional data. These data typically obtain by experiments or measurement by a sensor, and they contain uncertainty. For example, the results of repeat of the same experiment should be the same, but the results are mostly different and this creates a scattering because of uncertainties from various noise factors in real field. If the uncertainties are considered, the outcome of the PHM has the reliability which the designer wants. For reliable outcomes of the PHM, the appropriate distribution of the raw data should be selected. However, most distributions about input variables is assumed as a Gaussian distribution in real field and it causes errors because the Gaussian distribution can't describe non-linear relations. If the number of input variables are increased, the distribution can be estimated more exactly but spending cost, time and effort increase simultaneously in real field. As a result, designer can't obtain much data at the real field, so designer can obtain the desired amount of data from the distribution of data. Therefore, the determination of appropriate distribution is very important and it is required how to determine proper number of input variables. In this study, the algorithm to determine appropriate distribution of raw data using a determining method of the minimum needed number of raw data and the sequential statistical modeling is proposed and this proposed algorithm is improved more than existing presented algorithm. Also, this proposed algorithm is verified by using the automotive engine mounts rubber stiffness static data used in real field.

2. THE ALGORITHM TO DETERMINE PROPER DISTRIBUTION

In this study, the algorithm to determine the exact distribution of input data with uncertainties. First, this algorithm select the appropriate distribution of given data using sequential statistical modeling (SSM). Second, this algorithm check the sufficient of minimum needed number of data using the statistical area metrics. If the number of given data is less than the minimum needed number of data, the algorithm is repeated after adding the more real experiments data. Finally, if the number of data is more than the minimum needed number of data, this algorithm determines the appropriate distribution for the data. Figure 1 shows the flow chart about the proposed algorithm.

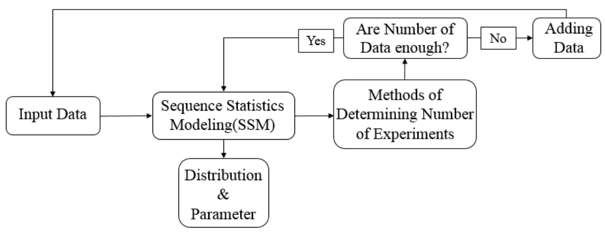

Figure 1. The algorithm of determining suitable distribution of input data

3. SEQUENCE STATISTICAL MODELING(SSM)

The existing SSM is one of the defining methods to estimate distribution sequentially. (Yoojeong Noh, Young-Jin Kang and O-Kaung Lim. 2015) The process of this modeling consists largely of Goodness of fit (GoF) test and model selection. The process of sequence statistical modeling performs Goodness of fit test to check the absolute suitability of given input data and candidate probability distributions, and as the next step, selects the best proper distribution to input data using the model selection through evaluation of relative suitability. However, Existing SSM consider only parametric distributions, so it has the limits of determining distribution. In this study, the SSM has been improved to be considered both parametric and non-parametric distributions using the kernel density function and the maximum likelihood estimation (MLE). Figure 2 shows the flow chart of the improved SSM. Eq. (1) represent the kernel density function. n is the number of given data and h is bandwidth.

$$\hat{f}_h = \frac{1}{nh}\sum_{i=1}^{n} K\frac{(x-x_i)}{h} \tag{1}$$

3.1. Goodness of fit test (GoF test)

The Goodness of fit test is a method of testing that whether the candidate distributions are fit for input data or not using statistical hypothesis test between candidate distributions and the input data. In this study, the Kolmogorov-Smirnov test (K-S test) was used as one of the GoF test methods. The K-S test is a way to verify the hypothesis that whether the data is derived from the candidate distribution model or not through a comparison between cumulative distributions function $F_n(x)$ of candidate distribution and empirical cumulative function $F(x)$ of input data using statistics D_n. If the statistics D_n are smaller than threshold value of K-S test table of significance level 0.05, the distribution applicable to the statistics D_n is selected into proper distribution about input data. The calculating formula of the statistics D_n is the Eq. (2). (Young-Jin Kang, O-kang Lim and Yoojeong Noh. 2016)

$$D_n = Max\left|F_n(x) - F(x)\right| \tag{2}$$

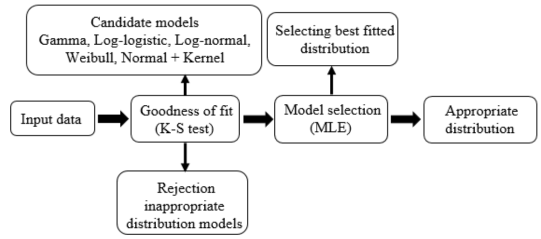

Figure 2. The algorithm of Sequence Statistics Modeling

3.2. Model selection

The model selection is a relatively statistical modeling method of selecting a candidate distribution model that the least of information loss between input data and candidate distribution function. In this study, the MLE is used to select the appropriate distribution of the left candidate distribution after Gof test. (Lee, S-Y. & Song X-Y, 2012) The MLE uses the likelihood function but defines the value of the negative log likelihood function as statistics. The most suitable distribution model for data has the biggest value of likelihood function and the MLE methods select the candidate distribution that has the smallest value of MLE. Eq. (3) represent the MLE. L is likelihood function. n and k is the number of data and parameter respectively.

$$MLE: -\ln(L) = -\ln\left(\prod_{i=1}^{n} f_k\left(x_i \middle| \theta\right)\right) \tag{3}$$

3.3. Application of SSM Algorithm using Rubber Stiffness Data

In this study, the measurements rubber stiffness static data to using automotive engine mounting was used as input data to apply to the SSM. The Measurements data is measured 20 times under the same rubber and condition in experiments. The candidate distribution model for conducting the sequence statistical modeling was chosen as a Gamma, Log-logistic, Log-normal, Weibull, Normal and which are widely used in engineering and Kernel distribution is also used to consider non-parametric distribution. Table 1 shows the result of the SSM. It shows that all candidate distribution are selected because p-value of all distribution are bigger than the significance level 0.05. Also, the Gamma distribution is selected finally as appropriate distribution about data because the MLE value of Gamma distribution is the smallest of other MLE values.

Table 1. The result of Goodness of fit (K-S test).

Candidate distribution	K-S test (P-value)	K-S test Selection	Model selection (MLE)	Final selection
Gamma	0.9924	○	78.9582	○
Log-Logistic	0.9856	○	79.2451	-
Log-normal	0.9958	○	78.9625	-
Weibull	0.9286	○	80.5238	-
Normal	0.9932	○	78.9943	-
Kernel	0.9708	○	78.9890	-

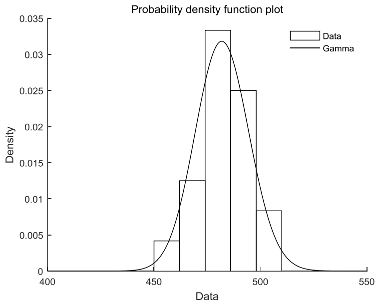

Figure 3. The probability density distribution plot of Gamma distribution and histogram of data

Figure 3 shows the histogram and probability density function plot of Kernel distribution. It shows that the Gamma probability density function plot follows the histogram of data.

4. METHODS OF DETERMINING NUMBER OF EXPERIMENTS

In the study, the sufficient number of measurement data was determined using the statistical area metric method and random sampling method in order to select a suitable distribution. Existing the methods of determining number of experiments have limitations, such as having to real experiment again and to adding the data each one if there is insufficient data. Also, the designer doesn't know how much more data is needed approximately. To overcome these limitations, SSM and random sampling were applied to existing methods of determining number of experiments.

4.1. Statistical Area Metric

The statistical area metric determines the number of data N with minimum number of experiments when the rate of change of distribution is initially smaller than the threshold value using the theory that the higher number of data, the smaller the rate of change for the distribution. In this study, the statistical area metric was improved to produce more accurate results for determining minimum number of experiments adding SSM and random sampling method to existing statistical area metric method. Figure 4 shows improved statistical area metric algorithm. At first, the proper distribution is estimated by entering the given data using SSM. Then, the algorithm estimates the shape of distribution for number of data N, N-1 and N-2 that they are extracted from distribution of given data by random sampling using kernel density estimation. (Wand, M.P. & Jones, M.C, 1994) After that, the intersection area of $IA_{Pop,N}$, $IA_{N,N-1}$ and $IA_{N,N-2}$ is calculated. $IA_{Pop,N}$ is intersection area of distributions between the population and the number of data N. The algorithm use the mean value as an intersection areas $IA_{Pop,N}$, $IA_{N,N-1}$ and $IA_{N,N-2}$ that the algorithm is repeated 200 times to reduce randomness from random sampling. The intersection area of distribution IA is summation value of distribution areas with a small value of two probability density function values. The intersection area IA is calculated using Riemaan integral. The equation (4) is Riemaan integral method and the Eq. (5) is kernel density function. h represents bandwidth of kernel distribution.

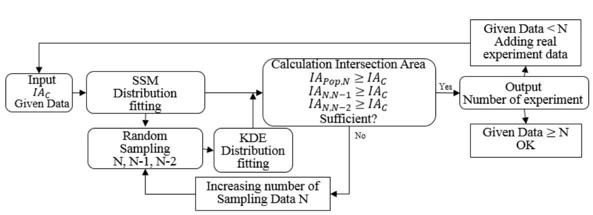

Figure 4. The algorithm of improved statistical area metric

$$IA = \sum_{i=1}^{n} f(x_i) \cdot (x_i - x_{i-1}) \tag{4}$$

$$f_h(x) = \frac{1}{nh} \sum_{i=1}^{n} \mathrm{K}\left(\frac{x - x_i}{h}\right) \tag{5}$$

The algorithm repeats itself by increasing the number of data N until that the intersection areas $IA_{Pop,N}$, $IA_{N,N-1}$ and $IA_{N,N-2}$ are more than the criteria area IA_C(90%) or equal to IA_C(90%) the first time simultaneously. (Jung Ho Jung, O-kang Lim and Yoojeong Noh. 2016) When the number of data N is determined, perform real experiment N times and assemble the given data and the experimental data. Then, the algorithm is repeated using assembly data to check whether the number of data is sufficient to estimate proper distribution for expressing of data properties.

4.2. Application of Methods of determining number of experiments Algorithm using Rubber Stiffness Data

The measurement static data of rubber stiffness for automotive engine mounts is used to verify the algorithm of improved statistical area metric. The number of data is 20EA and the criteria area value IA_C was set to 0.90 (90%). Table 2 show the result of determining minimum needed number of experiments for measurement data for initial and adding the number of data about automotive engine mounts rubber stiffness. The result of algorithm for static properties data shows that the number of required data to estimate proper distribution for the given number of data are 27EA missing quantities to express the proper distribution. According to the results, the additional experiments 27EA data are summed with given number of data 20EA through

Table 2. The result of method of determining number of experiments

Data	Given	Summation
Given Data	20EA	47EA
Proper distribution	Weibull	Normal
$IA_{N,N-1}$ $IA_{N,N-2}$ $IA_{Pop,N}$	0.987 0.983 0.903	0.985 0.980 0.905
Minimum needed number	47EA	47EA

Table 3. The result of Goodness of fit (K-S test) about the summation data (47EA)

Candidate distribution	K-S test (P-value)	K-S test Selection	Model selection (MLE)	Final selection
Gamma	0.9789	○	179.0086	-
Log-Logistic	0.9984	○	179.3305	-
Log-normal	0.9778	○	178.9944	-
Weibull	0.4496	○	183.3598	-
Normal	0.9647	○	179.0648	-
Kernel	0.9709	○	178.3401	○

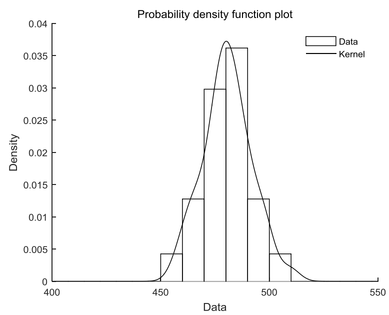

Figure 5. The probability density distribution plot of Gamma distribution and histogram of data

adding real experiments. The results of performing the algorithm with summation data shows that the sufficient number of data is 47EA, and the summation data (47EA) is sufficient to estimate proper distribution because the number of summation data same with the required number of data (47EA). Afterwards, according to the proposed algorithm, the summation data (47EA) applied to SSM as input data. The SSM selected the Kernel distribution and it is represented to Table 3 and Figure 5.

5. CONCLUSION

In this study, the algorithm for estimating proper distribution to express properties of input data used data attribute factor extraction from the first step in PHM is proposed. To verify the validity of this algorithm, the static and dynamic properties measurement data of rubber stiffness used in automotive engine mount was applied to the proposed algorithm as input data. As the result, the minimum needed number of data and proper distribution to express properties of data was estimated. As a result, data that expressed the tendency of properties can be sampled from the selected distribution of the designer as desired. If the estimated distribution and parameter are applied as input data, users will have reliable results in PHM.

ACKNOWLEDGEMENT

This research was supported by Hyundai NGV (Development of Optimization Technique considering Dispersion for Robust Design: 2016-11-0906)

NOMENCLATURE

$f(x_i)$	Probability density function
$K(\frac{x-x_i}{h})$	Kernel density function

REFERENCES

Jung Ho Jung, O-kang Lim, Yoojeong Noh. (2016), Comparison of Methods of Determining Number of Experiments Using Area Metric. *Korean Society of Mechanical Engineers.*pp.5-6.

Wand, M.P., & Jones, M.C. (1994). *Kernel Smoothing*. CRC press

Yoojeong Noh. (2016). A Comparison Study on Statistical Modeling Methods. *Journal of the Korea Academia-Industrial Cooperation Society*. vol. 17, issue. 5, pp. 645-652, doi: 10.5762/KAIS.2016.17.5.645

Yoojeong Noh, Young-Jin Kang, O-Kaung Lim. (2015). Integrated Statistical Modeling approach of Uncertainty variables. *Proceedings of the KSME 2015 Fall Annual Meeting. pp. 431-432.*

Lee, S-Y., Song X-Y. (2012). *Basic and Advanced Bayesian Structural Equation Modeling: With Applications in the Medical and Behavioral Sciences.* United Kingdom, John Wiley & Sons Ltd.

BIOGRAPHIES

Minho Joo received B.S. in Weapons System engineering at Korea Army Academy at Yeong-Cheon, Korea in 2009. He is a M. S. student in Mechanical Engineering at Yonsei University. His research interests are on the deep learning, probabilistic design optimization and PHM

Jaehyeok Doh received M.S. in Mechanical Engineering at Kyungpook National University, Korea in 2013. He is a Ph. D. student in Mechanical Engineering at Yonsei University. His research interests are on the field of structural analysis, finite element method, probabilistic design optimization and PHM

Jongsoo Lee received B.S. in Mechanical Engineering at Yonsei University, Korea in 1988 and Ph.D. in Mechanical Engineering at Rensselaer Polytechnic Institute, Troy, NY in 1996. After a research associate at Rensselaer Rotorcraft Technology Center, he is a professor of Mechanical Engineering at Yonsei University. His research interests include multidisciplinary/mult0 i-physics/multi-scale design optimization and reliability-based robust engineering design with applications to structures, structural dynamics, fluid-structure interactions and flow induced noise and vibration problems.