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

Nonparametric reliability analysis for design of a mechanical system working on an inaccessible area

Tae Hee Lee[1], Jong-Su Choi[2], Woo Chul Lim[3], Su-gil Cho[4], Minuk Lee[5], Sup Hong[6]

[1] Hanyang University, Seoul, Republic of Korea, thlee@hanyang.ac.kr
[2] Korea Institute of Ocean Science & Technology, Daejeon, Republic of Korea, jschoe@kiost.ac
[3] Hanyang University, Seoul, Republic of Korea, lwc0504@gmail.com
[4] Hanyang University, Seoul, Republic of Korea, chosk83@naver.com
[5] Korea Institute of Ocean Science & Technology, Daejeon, Republic of Korea, mwlee97@kiost.ac
[6] Korea Institute of Ocean Science & Technology, Daejeon, Republic of Korea, suphong@kiost.ac

1. Abstract
Nonparametric (distribution-free) reliability analysis (RA) is suggested as an alternative for RA of a mechanical system working on an inaccessible area, for instance, vehicle traveling on deep-seabed. Generally, it is not easy to estimate the appropriate statistical distribution function on the base of sample data, especially noise or environmental factors. Since the probability distributions of noise random variables are not often known and only a few sample data is given for design of a mechanical system working on the inaccessible area, the usage of nonparametric estimation is an emerging method to estimate the reliability of the system. Nonparametric RA is defined as reliability estimation of performance function from nonparametrically estimated distribution of noise random variables. A mathematical example is illustrated to compare the characteristics of nonparametric RA with those of parametric RA. Test statistic for estimated distribution of a noise random variable and reliability accuracy at a design point are used for evaluation of the performance and robustness of each RA. It is concluded that the nonparametric RA is more robust then the parametric RA for few sample data of stochastic variables.

2. Keywords: Nonparametric estimation, Reliability analysis (RA), Inaccessible area, Noise random variables (NRVs), Robustness

3. Introduction
The activity of mankind has been broader than the past in order to exploit the new energy and mineral resources, to utilize the new space for transportation and defense, and to perform the other newly required task. The activity of mankind have reached the inaccessible area in which people has no interest in the past because the area is hard to access. A new mechanical system working on the area would be required and designed. A deep-sea mining system for exploitation of deep-sea mineral resources is a representative example of a mechanical system working on the inaccessible area.

The factors affecting the response of a mechanical system could be classified as design and noise factors with respect to design optimization. The design factors are defined as the factors of which the values can be determined by designer. The noise factors are defined as the factors of which the values cannot be determined by designer. The design factors are also classified as design variables and design parameters. The design variables (DVs) are defined as the design factors which the designer intends to determine. The design parameters are defined as the design factors of which the values are already determined. Designer has no interest to determine the value of the design parameters. The design parameters would be regarded as constant values in the optimization formulation. The noise factors include the environmental factors, the operational factors, aging factors, tolerance factors, and etc. The noise factors would content all factors that the designer cannot control. The noise factors are classified as noise random variables and noise parameters. The noise random variables (NRVs) are defined as stochastic factors of which the variation is too large to be negligible. Deterministic parameters are defined as deterministic factors which can be treated as a constant value. When the variation of the noise factors is so small to be negligible, the noise factors can be regarded as noise parameters.

Design characteristics of a mechanical system working on an inaccessible area are summarized as follows:

- Small quantity and order-made production: Generally only small number of a mechanical system working on an inaccessible area would be required. Thus, the order-made production system is required against mass production such as automobiles and electronics. Thus, the variation of design factors due to the production tolerance is relatively minor than that of noise factors such as the environmental and operational factors. The tolerance of design factors can be controlled easily due to small quantity. The noise factors due to tolerance factors can be regarded as noise parameters.
- Hardness of testing and verifying the performance of the system: It is extremely expensive and

time-consuming to perform the design verification of the system. Thus, laboratory experiments and computational simulation technique would be required for compensating the insufficient number of verification tests. Simulation based design (SBD) would be effective to design a mechanical system working on inaccessible area.

- Hardness of estimating the probability distribution of noise random variables (NRVs): The variations of NRVs due to environmental factors are large and the probability distributions of the NRVs are not known yet. In addition to that, it is very hard to acquire the sample data for some of NRVs and sometimes insufficient sample data can be only available.
- Nonlinear responses due to large variation of NRVs and broad range of design variables: Since no mother design is referred to design the new mechanical system working on an inaccessible area, the searching range of the design variables is generally broad. The responses of the performance functions would be nonlinear due to the large variation of the environmental factors and broad range of the design variables.

In order to consider systematically the effect of the NRVs to the performance function, the reliability based design is necessary. In the deterministic design, a factor of safety is used to determine the value of the design variables. By using the factor of safety, the variation of noise factors is vaguely considered. In the reliability based design, the variation of noise factor is considered as the probability distribution. The effect of variation of the noise factor to performance functions can be systematically considered.

The main issue of reliability based design is how to estimate effectively the reliability of performance function. In this study, reliability analysis (RA) is defined as estimation of reliability of performance function. It is assumed that the variation of performance function is caused by the variation of NRVs at a given design point. Many researches on RA for design optimization have been studied in various fields ([1], [2], [3], [4]). Various methods are proposed for RA: Monte Carlo Simulation [Neumann et al., 1949], 2nd moment based approximation method [5] such as first order reliability method (FORM) and second order reliability method (SORM), and 4th moment based methods such as multiplicative decomposition (MD) [2] and dimension reduction (DR)[6]. Hence, the mentioned RA methods are performed on the basis of the assumption that the each probability distribution of NRVs is Gaussian one or at least is parametrically known probability distribution such as beta distribution, Weibull distribution, and etc.

In previous study, it was assumed that the probability distributions of the NRVs are already known and given to designer, or the probability distributions are easily constructed. However, in designing a new mechanical system working on an inaccessible area, the probability distributions of NRVs are usually not known, and moreover the number of sample data of the NRV is not enough to estimate the distribution as mentioned before. The accuracy of RA depends on the completeness of the probability distribution model and the accuracy of estimation of its parameters. The wrong selection of probability distribution model of NRVs would mislead to poor RA. The parametric model of probability distribution has been used in the previous study. In this study, nonparametric estimation method for the noise random variables is proposed. Since there is no assumption for probability distribution model in nonparametric estimation, the possibility of poor estimation of the reliability resulting from the wrong selection of probability distribution model is very low.

Nonparametric RA is defined as reliability estimation of performance function from nonparametrically estimated distribution of NRVs. Hence, parametric RA is defined as reliability estimation of performance function from parametrically estimated probability distribution of NRVs. In nonparametric RA, the NRVs are nonparametrically estimated using nonparametric estimators such as piecewise linear estimation (PLE), kernel density estimation (KDE) [7], or generalized Pareto distribution (GPD) in tails [8] for cumulative distribution. In this study, the performance and robustness of nonparametric RA are evaluated and compared with those of parametric RA by illustration of a mathematical example.

4. Estimation of probability distribution from sample data of noise random variables

4.1. Parametric estimation

The procedure of the parametric estimation from sample data of noise random variables consists of three steps: collection of candidate distributions, parameter estimation of each candidate distribution by using distribution fitting, and the selection of best fit distribution. The candidate probability distributions would be chosen by a designer. The designer would consider several points: pre-information of the NRVs, the applicability for RA, and easiness of distribution fitting and goodness-of-fit. After collecting the candidate probability distributions, distribution fitting is performed to each candidate distribution. By performing distribution fitting, the parameters of each candidate are decided. Four types of distribution fitting methods could be employed such as, maximum likelihood estimation (MLE), method of moments (MOM), method of L-moment (LMOM) [9], and least square estimation (LSE). MLE is the most popular distribution fitting method. MLE could be applied to most of probability distributions. If the MLE is not applicable, MOM, LMOM, or LSE would be used for distribution fitting.

In order to suggest the best probability distribution model among candidate models, goodness-of-fit (GOF) tests are performed such as the Kolmogorov–Smirnov (K-S) test [10], Anderson-Darling test, and chi-square test. Among them, the K-S test is the most popular GOF method and could be applied for any probability distribution. The K–S test compares the observed empirical cumulative density function (eCDF), $F_n(X)$, of sample data with the theoretical cumulative density function (CDF), $F(X)$, obtained from the candidate probability model whose parameters are already determined using distribution fitting methods. The eCDF is denoted by

$$F_n(X) = \frac{[\text{Number of observations} \le \text{X}]}{n} \tag{1}$$

where n denotes the number of samples. The maximum absolute difference, D, between the empirical and theoretical cumulative values for each is evaluated as follows:

$$D = \max\left|F_n(X) - F(X)\right| \tag{2}$$

D is also called test statistic used as decision measure of the hypothesis test. There are critical values (D_{cr}) corresponding to the number of samples (n) for each significance level of probability distributions. The critical value is a criterion for the determination of a significant fit. If a test statistic is less than or equal to the critical value, then the probability distribution model is considered as significant fit to the observed distribution.

4.2. Nonparametric estimation

As mentioned in the introduction, the probability distributions of noise random variables for a new mechanical system working in an unveiled area are generally not known and the sampled data for the random variables are not enough to estimate the probability model. The reliability would be sometimes poorly estimated due to the wrong assumption of probability distribution in parametric estimation. Nonparametric estimation does not mislead reliability with large modeling error because of less assumption for probability distribution. There are several nonparametric estimation methods: piecewise linear estimation (PLE), kernel density estimation (KDE), and the dual estimation. The dual estimation is composed of nonparametric model in body and tail model in tails. PLE and KDE are generally used for nonparametric model in body. Generalized Pareto distribution (GPD) is used for tail model in both tails.

4.2.1 Piecewise linear estimation

The empirical cumulative distribution function (eCDF) is related with the empirical measure of given sample. The eCDF is defined as

$$\hat{F}_N(t) = \frac{\text{number of elements in the sample} \le t}{k} = \frac{1}{k}\sum_{i=1}^{k} 1\{x_i \le t\} \tag{3}$$

where, $1\{A\}$ is the indicator of event *A*. Breakpoints are defined as the center point of jump of step function. Piecewise linear function of the breakpoints can be used more easily than the step function of eCDF for estimating a CDF of population. The piecewise linear function of breakpoints of eCDF is called as piecewise linear estimation.

4.2.2 Kernel density

Silverman [7] and Wand & Jones [11] summarized the KDE methods. For given sample data of *j*-th noise random variable, N_j, the estimated probability density, $\hat{f}_{N_j}$, is expressed as linear combination of each kernel which is located in its sample point. The coefficient is the inverse of number of data, k_j, multiplied to bandwidth, h_j. The definition of KDE is given as follows:

$$\hat{f}_{N_j}(n_j) = \frac{1}{k_j h_j}\sum_{m=1}^{k_j} K\left(\frac{n_j - N_j^{(m)}}{h_j}\right) \tag{4}$$

Many kernels could be used: Epanechnikov, cosine, biweight, triweight, Gaussian, triangular, and uniform as shown in Table 1. However, Gaussian kernel is commonly used in the literature.

Table 1 : Kernels commonly used in KDE [11]

Kernel name	Form
Epanechnikov	$\frac{3}{4}\left(1-u^2\right)I\left(\lvert u\rvert \le 1\right)$
Cosine	$\frac{\pi}{4}\cos\left(\frac{\pi u}{2}\right)I\left(\lvert u\rvert \le 1\right)$
Biweight	$\frac{15}{16}\left(1-u^2\right)^2 I\left(\lvert u\rvert \le 1\right)$
Triweight	$\frac{35}{32}\left(1-u^2\right)^3 I\left(\lvert u\rvert \le 1\right)$
Gaussian	$\frac{1}{\sqrt{2\pi}}e^{-0.5u^2}$
Triangular	$\left(1-\lvert u\rvert\right)I\left(\lvert u\rvert \le 1\right)$
Uniform	$0.5I\left(\lvert u\rvert \le 1\right)$

$I\left(\lvert u\rvert \le 1\right)$ means 1 for $\lvert u\rvert \le 1$, 0 (zero) for otherwise.

In order to complete the estimation of density of noise random variables, kernel and bandwidth are determined. Bandwidth should be selected for getting a good balance of variation of deviation for asymptotic mean integrated square error (AMISE). There are several bandwidth selection methods: Gaussian optimal smoothing, least square cross validation, and plug-in bandwidth ([10], [11]). Gaussian optimal smoothing is commonly used as a bandwidth selection method.

Gaussian optimal smoothing is simple and practical estimation method of bandwidth ([10], [11]). The bandwidth is given as

$$h_1 = 1.06\min\left(\hat{\sigma}_1, \frac{I\hat{Q}R_1}{1.34}\right)n_1^{-1/5} \tag{5}$$

where, $\hat{\sigma}$ is the standard deviation of the sample, $I\hat{Q}R$ is an interquartile range of the sample. This bandwidth is calculated on univariate Gaussian kernel function. This bandwidth selection method is practical because of fast calculation. Hence, a robust estimate of σ is preferable to the usual sample standard deviation in order to accommodate long tailed distributions and possible outliers. A simple choice is the median absolute deviation estimator.

$$\hat{\sigma}_1 = median\left\{\left|N_1 - \tilde{\mu}_{N1}\right|\right\}/0.6745 \tag{6}$$

where, $\tilde{\mu}_{N1}$ and $\tilde{\mu}_{N1}$ denotes the median of the sample [12].

4.2.3 Dual estimation

In order to overcome insufficient tail modeling in the application of PLE and KDE due to a few data in tails, dual estimation methods are proposed. The probability distribution is nonparametrically estimated in body and estimated by 2nd extreme value theorem in tails. The central part, or body part, of probability distribution, is modeled nonparametrically as PLE or KDE. Both tails of CDF are modeled parametrically as generalized Pareto distribution (GPD).

Pickands-Balkema-de Haan (PBH) theorem, often called as 2nd extreme value theorem, describes the distribution of sample above a high threshold or/and below a low threshold as a GPD. This PHB theorem is particularly useful for modeling tails of sample from any distribution. CDF of GPD is defined as

$$F_{(\sigma,\mu,\kappa)} = \begin{cases} 1-\left(1+\dfrac{\kappa(x-\mu)}{\sigma}\right)^{-1/\kappa} & \text{for } \kappa \ne 0 \\ 1-e^{-\frac{(x-\mu)}{\sigma}} & \text{for } \kappa \ne 0 \end{cases} \tag{7}$$

The dual estimation of probability distribution of a NRV is defined as

$$F_{(\sigma,\mu,\kappa),L} \text{ for } x < u_L$$

$$F_{(\sigma,\mu,\kappa),H} \text{ for } x > u_H \tag{8}$$

$$\text{PLE or KDE for } u_L \le x \le u_H$$

Importance and difficulty for threshold selection of dual estimation are analogous with that of bandwidth selection of KDE. Threshold selection is determined by a tradeoff between bias and variance. If the threshold selected is too low, the assumption for GPD, sufficient large threshold, is largely violated and do not provide a good approximation to the tails. On the other hand, if the threshold selected is too high, the number of data available for the tail approximation is much less and this may lead to excessive scatter in the final estimate. The proper selection of threshold is very important because it has important consequence on the estimated value of the shape factor. Easy and practical threshold selection method is suggested by Boos [13] and Hasofer [14]. Boos suggested that the ratio N_{ex} (number of tail data) to N should be 0.02 (50< N <500) and 0.1 (500< N <1000). Hasofer suggested N_{ex} should be $1.5\sqrt{N}$.

5. Nonparametric RA

While the efficient estimation of reliability such as FORM, SORM, MDM and eDRM could be possible from the probability distribution of NRVs using parametric RA, the wrong choice of probability distribution model of NRV would directly affect the reliability of performance function. Since the RA process is performed on the assumption that the chosen probability distribution and its fitted parameters completely estimates the distribution of the population of a NRV, the insufficient choice of candidate distributions would cause a wrong estimation of reliability irrespective of the accuracy of fitted parameters.

In order to reduce the invulnerability of estimating the reliability due to wrong choice of candidate distributions in parametric RA, the use of nonparametric RA is suggested for estimating the probability distribution of NRVs. Nonparametric estimation of probability distribution is taken into account in previous section. Since there is only a little assumption for probability distribution model, it is very low possibility that nonparametrically estimated probability distribution leads to wrong estimation. If the probability distribution is not known, the nonparametric estimation of NRV can be an alternative for parametric one. It is rationally deduced that the good estimation of probability distribution of NRVs would lead to good estimation of reliability of the performance function.

Monte Carlo simulation (MCS) could be applied to identically both parametric RA and nonparametric RA. In this study, MCS is applied to nonparametric RA. Four steps of nonparametric RA by using MCS are given: nonparametric estimation of probability distribution from sample data of each NRV, sampling from nonparametrically estimated probability distribution of NRVs, putting the sample into performance function, and reliability estimation nonparametrically from performance random variables.

5.1 Nonparametric estimation of probability distribution from sample data of each NRV

In section 4, the nonparametric estimation of probability distribution from sample data of each NRV is already explained. PLE, KDE, PLE/GPD, and KDE/GPD can be used. Since inverse CDF of each NRV is constructed, re-sampling for MCS can be possible by inverse transform method. In this study, resampling is used for distinguishing the given sample of NRV. Simulation is performed for the resampled data.

5.2 Sampling from estimated probability distribution of NRVs

In this study, Latin hypercube sampling [15] is used in order to minimize the variance of resampling. The resample is generated based on the inverse transform method for putting the sample data to m -th performance function. It is assumed that the number of resample data for MCS can be chosen differently as the target reliability of probabilistic constraints corresponding for each performance function.

$$\overline{N}_{j,m} = \left\{\overline{N}^{(1)}_{j,m}, \overline{N}^{(2)}_{j,m}, \ldots, \overline{N}^{(\bar{k}_m)}_{j,m}\right\} = \hat{F}^{-1}_{N_{j,m}}\left[\left(i-1+P_i\right)/\bar{k}_m\right],\ i=1,2,\ldots,\bar{k}_m,\ j=1,2,\ldots,nnrv,\ m=1,2,\ldots,nc \tag{9}$$

where, $\overline{N}_{j,m}$ denotes the resampled random variable corresponding to NRV and performance function, $nnrv$ denotes the number of NRVs, nc denotes the number of constraints equivalent with the number of performance functions, $P_i \in (0,1)$ denotes a uniform random sample, and $\hat{F}^{-1}_{N_j}[P]$ denotes the inverse transform for the particular input distribution, $P \in (0,1)$. The order of resampled data of each NRV is randomly determined. The sample set for simulation is made by assembling random permutation of each NRV.

As the number of sample increases, the estimation accuracy would increases. The solution of MCS will be converged to $1/\sqrt{\bar{k}_m}$. The higher reliability a designer wants to set as target reliability, the more resampled data one puts into the performance function.

5.3 Putting the resample data into performance function

The performance random variables (PRV) are obtained by putting the resampled data into the performance function.

$$G_m = \left\{G_m^{(1)}, G_m^{(2)}, \ldots, G_m^{(\bar{k}_m)}\right\},\ m = 1,2,\ldots,nc \tag{10}$$

where G_m denotes m -th PRV. In this study, PRV is introduced in order to calculate the reliability from not counting but nonparametric estimation.

5.3 Putting the resample data into performance function

In crude MCS, the reliability of m-th performance function is evaluated as follows:

$$\mathrm{R}_m = \frac{\text{No. of successful results}}{\text{No. of trials}} = \frac{\text{No. of m-th PRVs satisfying performance function}}{\text{No. of all of m-th PRVs}} \tag{11}$$

In this study, the reliability is obtained by nonparametric estimation as follows:

$$\mathrm{R}_m = \Pr(g_m \le 0) = F_{G_m}(0) \tag{12}$$

where, F is CDF estimated from PLE, KDE, PLE/GPD, or KDE/GPD. Generally nonparametric estimation, Eq. (12) would give a designer the more exact results for the same number of data than just counting the number of trials which satisfy the performance function of Eq. (11).

6. Mathematical example

A mathematical example is chosen in order to investigate the characteristics of parametric RA and nonparametric RA with respect to the estimator for the probability distribution of NRVs. Probability distribution for NRVs is assumed as GOK (God only knows) distribution which is introduced in order to measure the performance of parametric and nonparametric RA. No one knows the probability distribution of NRVs, called as GOK distribution. GOK distribution is used as reference distribution for obtaining the test statistic and reliability accuracy that are used as measures of evaluating the performance of estimator.

6.1 Performance function and GOK distribution

The RA problem is defined as estimating the reliabilities of performance functions at a point of design variables. The two performance functions, g_1 and g_2 , are given as

$$g_1 = \frac{\left((n_1 - 0.35)^2 + (n_2 + 0.35)^3\right)}{d_1^{\,2} d_2} - 3.5 \le 0 \tag{13}$$

$$g_2 = \frac{n_1 \sin(4 n_1) + 1.1 n_2 \sin(2 n_2)}{d_1 d_2^3} - 11 \le 0 \tag{14}$$

where, n_1 and n_2 denote the noise factor #1 and #2, respectively. d_1 and d_2 denote the design variable #1 and #2, respectively. The nonlinearity of performance function #2 is higher than the performance function #1.

The sample data are collected randomly from GOK distributions of NRVs. In order to reduce the variation due to random sampling, many groups of sample were randomly extracted.

$$N_{1,i} = \left\{N_1^{(1)}, N_1^{(2)}, \ldots, N_1^{(k_1)}\right\},\quad k_1 = 100,\quad i = 1,2,\ldots,10 \tag{15}$$

$$N_{2,i} = \left\{N_{2,i}^{(1)}, N_{2,i}^{(2)}, \ldots, N_{2,i}^{(k_2)}\right\},\quad k_2 = 10000,\quad i = 1,2,\ldots,10 \tag{16}$$

where, i denotes the repetition number of random sampling. In this study, 10 groups of sample are extracted and used. The sample data are chosen from GOK distribution #1, $f_{N_1}(n_1)$, and GOK distribution #2, $f_{N_2}(n_2)$.

The GOK distribution #1 is chosen as normal distribution:

$$f_{N_1}(n_1) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(n_1-\mu)^2}{2\sigma^2}} \tag{17}$$

where, μ denotes mean of the distribution, 0.5, and σ denotes standard deviation, 0.11. The GOK distribution #2 is chosen as beta distribution:

$$f_{N_2}(n_2) = \frac{1}{B(a,b)} n_2^{a-1}(1-n_2)^{b-1}, \quad 0 \le n_2 \le 1 \tag{18}$$

where, $B(\bullet)$ denotes the beta function. a is given as 2.7, and b is given as 4.3.

6.2 Parametric RA

In parametric RA, the candidate distributions are chosen as well-known unimodal distributions: Gaussian, lognormal, Weibull, Gumbel, uniform, exponential, gamma, generalized extreme value, generalized Pareto, and Rayleigh. The best fit of which test statistic is minimum among the candidate distributions is chosen for each group of sample data of noise random variables #1 and #2. The better fit is defined as worst fit distribution among no-rejected distributions for null hypothesis at 5% significant level, or 2nd best fit distribution if there is no other distribution except best fit among no-rejected distributions. The best fit and better fit are used for analyzing the characteristics of parametric RA.

6.3 Nonparametric RA

In nonparametric RA, the PLE, KDE, and KDE at body and GPD at tails are used for analyzing the characteristics of nonparametric RA

6.4 Performance measure: test statistic

Ten groups of sample data are randomly produced from GOK distribution. The test statistic of an estimator is defined as the maximum difference of CDF of GOK distribution and the estimator for randomly chosen sample data from GOK distribution. Mean and standard deviation of test statistic of m-th estimator for noise random variables #1 are defined as

$$\begin{gathered} D_{1,i}^{(s)} = \max\left(\left|F_{N_i}^{(GOK)} - \hat{F}_{N_i}^{(s)}\right|\right), \quad i = 1,\ldots,10 \\ \overline{D}_1^{(s)} = \frac{\sum_{i=1}^{10} D_{1,i}^{(s)}}{10}, \quad \hat{D}_1^{(s)} = \sqrt{\frac{\sum_{i=1}^{10}\left(D_{1,i}^{(s)} - \overline{D}_1^{(s)}\right)^2}{10-1}} \end{gathered} \tag{19}$$

where, s denotes the estimator name. In this study, best fit and better fit in parametric RA, and PLE, KDE, and KDE/GPD in nonparametric RA are adopted.

6.5 Performance measure: reliability accuracy

Since MCS is the only RA method which can be adopted for both parametric and nonparametric RA, MCS is used for estimating the reliability. When the reliabilities for the estimation models are compared, the accuracy of the reliability is defined as the relative difference of MCS solution for each estimation model on MCS solution of GOK's probability distribution. Mean and standard deviation of reliability error of s -th estimator for performance function #1 are defined as

$$\mathrm{R}_{1,i}^{(s)} = \Pr(g_1 \le 0) = F_{G_1}(0), \quad i = 1,2,\ldots 10 \tag{20}$$

$$e_{1,i}^{(s)} = \frac{\mathrm{R}_{1,i}^{(s)} - \mathrm{R}_{1,\mathrm{GOK}}}{\mathrm{R}_{1,\mathrm{GOK}}}, \quad i = 1,2,\ldots 10 \tag{21}$$

$$\bar{e}_1^{(s)} = \frac{\sum_{i=1}^{10} e_{1,i}^{(s)}}{10}, \quad \hat{e}_1^{(s)} = \sqrt{\frac{\sum_{i=1}^{10}\left(e_{1,i}^{(s)} - \bar{e}_{1,}^{(s)}\right)^2}{10-1}} \tag{22}$$

In this study, design point of design variables is given as (0.6, 0.6). The number of sample of MCS is chosen as 10^6. Latin hypercube sampling (LHS) is adopted for sample data extraction of MCS. The reliability is calculated by using KDE from PRV.

6.6 Comparison of test statistic of NRVs

The test statistic of each estimator from 10 sets of random sample data of GOK distributions are shown in Table 2 and Table 3. The parametric estimated probability distributions, best fit and better fit, for 10 sets of random sample data are also listed in Table 2 and Table 3. The average ($\overline{D_1}^{(s)}$, $\overline{D_2}^{(s)}$) and standard deviation ($\hat{D}_1^{(s)}$, $\hat{D}_2^{(s)}$) of test statistics are shown to compare the test statistics for each estimator in Fig. 1 and Fig. 2.

The best fit showed minimum test statistic for NRV #1 in Table 2 and Fig.1. It means that the parametric estimation is best if a designer selects the appropriate candidate distributions. KDE showed a slightly higher test statistic for NRV #1 in Table 2 and Fig. 1. The difference of test statistics of nonparametric estimators is not larger than that of parametric estimation of best and better fit. For NRV #2 in Table 3 and Fig. 2, KDE showed minimum test statistic. The difference of test statistics of nonparametric estimators is so small to be equivalent in test statistics. Best fit in parametric estimation showed higher test statistic than that of any nonparametric estimator. It means that the performance of nonparametric estimators is good and robust. From the above analysis, it is deduced that nonparametric estimators do not lead bad estimation results.

6.7 Comparison of reliability accuracy of performance functions

The reliability errors based on GOK reliability which denotes MCS solution from GOK's probability distribution were illustrated in Fig. 3 and Fig. 4. The mean values of reliability error in nonparametric estimators are smaller than those of parametric estimations. The standard deviations of the error in nonparametric estimators are much smaller than those of parametric estimators for performance function #2 as shown in Fig. 4.

Table 2: Test statistics and name of fitted probability distribution for 10 sets of random sampling data of NRV #1 with respect to best fit, better fit, PLE, KDE, and KDE/GPD

	Parametric				Nonparametric		
No. of sets	Best fit		Better fit		PLE	KDE	KDE/GPD
	Name	Value	Name	Value			
1	Normal	5.62E-02	Gumbel	9.63.E-02	6.86E-02	3.25E-02	4.99E-02
2	Normal	1.18E-02	Gumbel	6.38.E-02	5.47E-02	1.75E-02	1.91E-02
3	Normal	5.80E-02	Gen. Pareto	9.71.E-02	9.74E-02	6.41E-02	6.42E-02
4	GEV	2.86E-02	Uniform	7.87.E-02	5.16E-02	4.04E-02	4.04E-02
5	Gamma	3.04E-02	Uniform	9.43.E-02	4.51E-02	4.35E-02	4.25E-02
6	Gamma	3.58E-02	Uniform	9.91.E-02	5.77E-02	4.42E-02	4.43E-02
7	GEV	3.30E-02	Gumbel	1.06.E-01	5.78E-02	4.11E-02	4.12E-02
8	Normal	3.56E-02	Gumbel	1.51.E-01	1.03E-01	8.21E-02	8.23E-02
9	GEV	2.44E-02	Lognormal	6.15.E-02	5.45E-02	2.78E-02	3.84E-02
10	GEV	5.83E-02	Uniform	5.79.E-02	8.59E-02	5.97E-02	5.97E-02
Mean		3.72E-02		9.05.E-02	6.77E-02	4.53E-02	4.82E-02
St. deviation		1.56E-02		2.74.E-02	2.06E-02	1.89E-02	1.72E-02

Table 3: Test statistics and name of fitted probability distribution for 10 sets of random sampling data of NRV #2 with respect to best fit, better fit, PLE, KDE, and KDE/GPD

No. of sets	Parametric				Nonparametric		
	Best fit		Better fit		PLE	KDE	KDE/GPD
	Name	Value	Name	Value			
1	Weibull	1.25E-02	GEV	1.38.E-02	5.69E-03	5.51E-03	5.64E-03
2	Weibull	1.02E-02	GEV	1.32.E-02	7.20E-03	7.24E-03	7.38E-03
3	Weibull	1.21E-02	GEV	1.40.E-02	7.30E-03	5.11E-03	5.11E-03
4	Weibull	1.30E-02	GEV	1.56.E-02	9.50E-03	8.09E-03	8.11E-03
5	Weibull	1.39E-02	GEV	1.66.E-02	8.17E-03	3.79E-03	3.78E-03
6	Weibull	9.10E-03	GEV	1.32.E-02	4.41E-03	7.00E-03	7.12E-03
7	Weibull	9.10E-03	GEV	1.72.E-02	1.22E-02	1.31E-02	1.32E-02
8	Weibull	7.80E-03	GEV	1.46.E-02	5.32E-03	6.83E-03	6.93E-03
9	Weibull	9.90E-03	GEV	1.26.E-02	8.56E-03	9.43E-03	9.53E-03
10	Weibull	1.76E-02	GEV	1.95.E-02	1.64E-02	1.42E-02	1.42E-02
Mean		1.15E-02		1.50.E-02	8.47E-03	8.03E-03	8.10E-03
St. deviation		2.91E-03		2.19.E-03	3.57E-03	3.37E-03	3.36E-03

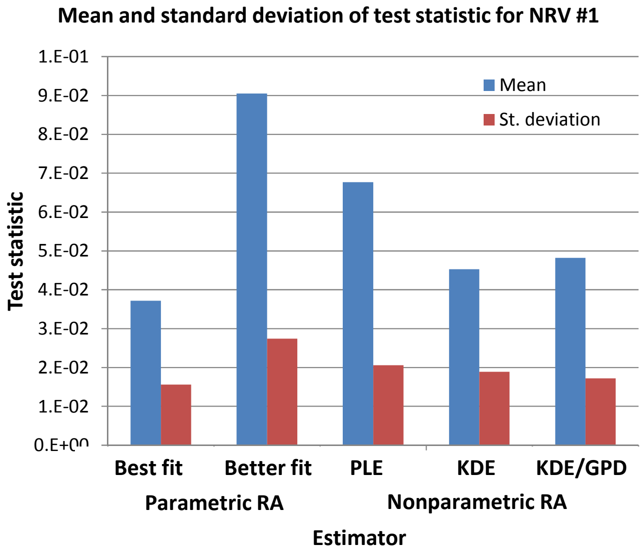

Figure 1: Mean and standard deviation of test statistics for 10 sets of random sampling data of NRV #1 with respect to best fit, better fit, PLE, KDE, and KDE/GPD

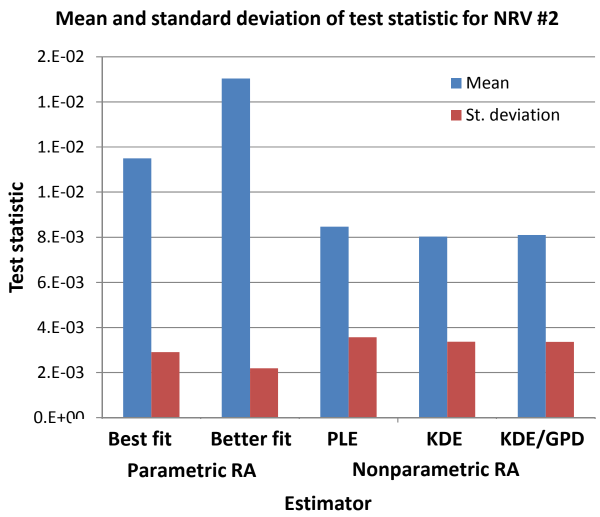

Figure 2: Mean and standard deviation of test statistics for 10 sets of random sampling data of NRV #2 with respect to best fit, better fit, PLE, KDE, and KDE/GPD

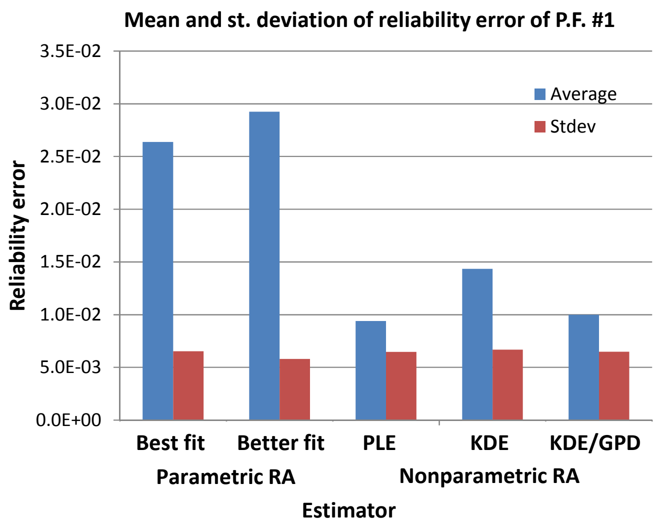

Figure 3: Mean and standard deviation of reliability error for performance function #1 from 10 sets of random sampling data of NRVs with respect to best fit, better fit, PLE, KDE, and KDE/GPD

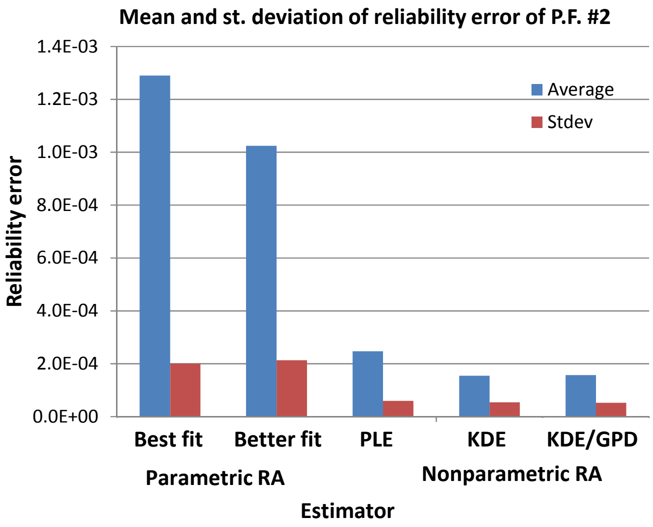

Figure 4: Mean and standard deviation of reliability error for performance function #2 from 10 sets of random sampling data of NRVs with respect to best fit, better fit, PLE, KDE, and KDE/GPD

6.8 discussions

It seems that any RA is not distinctly better than the other RA. If one can choose appropriate candidate distributions of NRV, the parametric RA will be best. But, the nonparametric RA would be cautiously recommended in designing a mechanical system working on an inaccessible area. The reasons are shown as follows:

- The performance of nonparametric RA would not be worse than parametric RA.
- A designer can often face that probability distributions are not known and the small number of sample data for NRV are just given. In parametric RA, a wrong choice of candidate distributions due to insufficient information about NRVs yields the poor RA.
- Nonparametric RA shows robust and stable results regardless of choosing any estimators as shown in previous example.
- Since the nonparametric estimator is very simple to use, the estimation time is generally shorter than the parametric estimation process that includes time consuming work of parameter estimation of all candidate distributions.

7. Conclusions

In this study, nonparametric reliability analysis (RA) is introduced in order to overcome the design difficulties in a reliability based design of a new working system on an inaccessible area. Since the probability distributions of noise random variables are not often known and only a few sample data are given, the usage of nonparametric estimation is an emerging method to estimate the reliability of the system. The parametric and nonparametric estimation of noise random variables are summarized. The procedure of Monte Carlo simulation for nonparametric RA is outlined. A mathematical example is illustrated to compare the characteristics of nonparametric RA with those of parametric RA. Test statistic and reliability accuracy are used for evaluation of the performance and robustness of each RA. Based on the example study several conclusions are drawn:

- The performance of nonparametric RA would not be worse than parametric RA.
- The nonparametric RA is more robust then the parametric RA for few sample data of NRVs.
- The nonparametric RA would be cautiously recommended in designing a mechanical system working on an inaccessible area.

5. Acknowledgements

The contents of this paper are a part of the R&D project, “Technology Development of Deep-seabed Mining System for Manganese Nodules,” sponsored by the Ministry of Oceans and Fisheries of Korea. This research was partially supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2010-0023257). The authors appreciate this partial support for this research.

6. References

[1] R. Rackwitz, Reliability analysis - a review and some perspectives. *Structural Safety*. 23, 365–395, 2001.

[2] B.D. Youn and K.K. Choi, An investigation of nonlinearity of reliability-based design optimization approaches. *Transactions of the ASME*, 126, 2004.
[3] J.J. Jung, Multiplicative decomposition method for accurate moment–based reliability method, Ph.D thesis, Hanyang University, 2007
[4] M.A. Valdebenito and G.I. Schüeller, A survey on approaches for reliability-based optimization, *structural multidisciplinary optimization*, 42, 645-663, 2010
[5] A.M. Hasofer and N.C. Lind, Exact and invariant second-moment code format, *ASCE*, 100(1), 1974
[6] S. Rahman and H. Xu, A univariate dimension-reduction method for multi-dimensional integration in stochastic mechanics, *Probabilistic Engineering Mechanics*, 19, 393-408, 2004
[7] B.W. Silverman, *Density estimation for statistics and data analysis*, Chapman & Hall/CRC, 1986
[8] N.H. Kim, P. Ramu and N.V. Queipo, Tail-modeling in reliability-based design optimization for highly safe structural systems, *47th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, Rhode Island, USA, 2006
[9] J.R.M. Hosking, L-Moments: analysis and estimation of distributions using linear combinations of order statistics, *Journal of Royal Statistics Society*, 52(1), 105–124, 1990
[10] R.B. D'Agostino and M.A. Stepehens, *Goodness-of-fit techniques*, CRC press, 102–106, 1986.
[11] M.P. Wand and M.C. Jones, *Kernel smoothing*, Chapman & Hall/CRC, 1995.
[12] R.V. Hogg, Statistical robustness: one view of its use in applications today, *American Statistician*, 33, 108-116, 1979.
[13] D.D. Boos, Using extreme value theory to estimate large percentiles, *Technometrics*, 26, 33-39,1984.
[14] A.M. Hasofer, Parametric estimation of failure Probabilities. In Mathematical Models for structural Reliability Analysis, Edited by Casicati, F. , Roberts, B. Boca Raton, FL, CRC Press, 1996.
[15] M.D. McKay, R.J. Beckman, and W.J. Conover, A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics*, 21(2), American Statistical Association, 239–245, 1979.