11th World Congress on Structural and Multidisciplinary Optimization
7th - 12th, June 2015, Sydney Australia

Parameter Estimation Method Using Bayesian Statistics Considering Uncertainty of Information for RBDO

Makoto Ito [1], Nozomu Kogiso [2]

[1] Osaka Prefecture University, Osaka, Japan, st102005@edu.osakafu-u.ac.jp
[2] Osaka Prefecture University, Osaka, Japan, kogiso@aero.osakafu-u.ac.jp

1. Abstract
Reliability-based design optimization (RBDO) is one of the methods considering the effect of uncertainties on the design optimization. However, the probabilistic parameters of the random variables also have uncertainties due to lack of sufficient information under actual situation, for example in case that the number of experiments is limited. This study considers the uncertainty of the probabilistic parameter of the random variables for the RBDO. On this proposed method, the distribution parameters are also considered random variables. Based on Bayesian statistics, the confidence intervals of the parameters are estimated with high accuracy. Then, the confidence interval of the reliability-based optimum design is also evaluated. That is, the accuracy of the obtained reliability-based optimum design is evaluated through the interval estimation, when sufficient information is not available. Through numerical examples, the validity of the proposed method is demonstrated.

2. Keywords: Parameter estimation, Bayesian statistics, Uncertainty, Reliability, Optimization

3. Introduction
Recently, the importance of the reliability is growing in the structural design requirements. The reliability-based design optimization (RBDO) has been adopted to evaluate the design under the reliability constraints in terms of random variables [1]. RBDO generally requires the probabilistic distribution parameters of random values for the reliability evaluation. For example, the first order reliability method (FORM) converts the probabilistic distribution of random variables into the standardized normal distribution.
However, for the actual design problem, the probabilistic parameters of random parameters are sometimes obtained with insufficient accuracy due to the limited number of experiments. In that case, the distribution parameters such as mean and the standard deviation also have uncertainties.
This study addresses to investigate the effect of the parameter uncertainties on the reliability-based optimization. For the purpose, this study proposes the distribution parameter estimation method considering uncertainties due to the lack of information. Then, the effect of the uncertainties on the RBDO is clarified as the confidence interval of the reliability-based optimum design.
The probabilistic distribution parameter is estimated based on Bayesian statistics. Bayesian statistics is the subset of the field of statistics and related to conditional probability [2]. This statistics is well known to be suited to estimating statistical model and employed for calculating distribution of failure probability. Gunawan *et al.* [2] proposed the method considering incomplete information uncertainties, which RBDO problem is converted to the multiobjective problem whose Pareto solutions reflect to the trade-off between performance and confidences. Wang *et al.* [3] presented the paradigm of the reliability prediction using evolving, insufficient information applying Bayes' theory. They showed that Bayesian statistics is useful tool for the distribution parameter estimation and that Bayesian updating method, the tool of Bayes' theory, makes the probabilistic distribution or the estimated reliability sufficient accuracy. However, under the lack of the number of the sample data or information of the random values, Bayesian updating does not bring the sufficient accuracy.
First in this study, the parameter estimation method is proposed under the cases of the lack of the information. In the next section, the reliability-based design optimization is reviewed. Then, Bayesian statistics is reviewed in Section 5.. The proposed method is described in Section 6.. The validity of the proposed method is illustrated through simple numerical examples in Section 7.. Finally, the conclusions are remarked.

4. Reliability-based design optimization
The RBDO is generally formulated as follows:

$$\begin{aligned} &\text{Minimize : } f(\boldsymbol{d}) \\ &\text{subject to : } P[g_j(\boldsymbol{d}, \boldsymbol{X}) \le 0] \le \Phi(-\beta_j) \qquad (j = 1, \cdots, NC) \end{aligned} \tag{1}$$

where $\boldsymbol{d} = (d_1, \cdots, d_{ND})$ and $\boldsymbol{X} = (x_1, \cdots, x_{NR})$ indicate the design variables and the random variables, respectively. The random variable is assumed to have an independent Gaussian distribution for simplicity. The reliability constraints indicate that the failure probabilities are lower than the upper limit, where $g_i(\boldsymbol{d}, \boldsymbol{X})$ is the j-th limit state function and β_j is the target reliability index value corresponding failure mode. ND, NR and NC are the number of the design variables, the random variables, and the reliability constraints, respectively.
As the RBDO algorithm, this study adopts the Modified-SLSV(Single-Loop-Single-Vector) method [4]. The original SLSV method [5] converts the double-loop optimization loop of the RBDO into the single loop approach by approximating the MPP (most probable point). The modified-SLSV method improves the convergency by eliminating zigzagging iteration that will yield divergence of the optimum design searching.
At first, the MPP(most probable point) is described in $\boldsymbol{X}$-space as follows:

$$\boldsymbol{x}_j^* = \boldsymbol{\mu} - \beta_j \boldsymbol{\sigma} \boldsymbol{\alpha}_j^* \qquad (j = 1, \cdots, NC) \tag{2}$$

$$\boldsymbol{\alpha}_j^* = \frac{\boldsymbol{\sigma}_j \nabla g_j(\boldsymbol{\mu}, \boldsymbol{x}_j^*)}{\left|\boldsymbol{\sigma} \nabla g_j(\boldsymbol{\mu}, \boldsymbol{x}_j^*)\right|}, \qquad (j = 1, \cdots, NC) \tag{3}$$

where $\boldsymbol{\mu}$ is the mean value of the random vector that is allocated as the design variables and $\boldsymbol{\sigma}$ is a diagonal matrix whose diagonal element consists of the standard deviation of each random variable. $\boldsymbol{\alpha}_j^*$ denotes the normalized gradient vector of the j-th limit state function evaluated at the MPP $\boldsymbol{x}_j^*$.
The reliability constraint is replaced into the deterministic constraints using Eq. (2) as follows:

$$g_j(\boldsymbol{\mu}, \boldsymbol{\mu} - \beta_j \boldsymbol{\sigma} \boldsymbol{\alpha}_j^*) \geq 0 \tag{4}$$

where α_j^* should be evaluated after obtaining the MPP $\boldsymbol{x}_j^*$.
On the SLSV method, the normalized gradient vector $\boldsymbol{\alpha}_j^*$ is approximated by the vector obtained at the previous iteration as follows:

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{\mu}^{(k+1)} - \beta_j \boldsymbol{\sigma} \boldsymbol{\alpha}_j^{*(k)} \tag{5}$$

where the superscript (k) indicates the number of iteration. This strategy makes the RBDO computationally efficient.
However, the searching sometimes fails to converge or lead to an inaccurate solution. In the Modified-SLSV method, the sensitivity is replaced to improve the convergence by using the previous values as follows:

$$\boldsymbol{\alpha}^{(k)} = \boldsymbol{\alpha}^{(k-2)} + \boldsymbol{\alpha}^{(k-1)} \qquad (k > 2) \tag{6}$$

5. Bayesian statistics

On Bayes' theory, the following conditional probability formulation is used.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{7}$$

where event A is considered as a cause of a result B. $P(A)$ is called a prior probability that the cause A happens, and $P(B|A)$ is called a likelihood that the result B happens under the condition that the cause A happens. On the other hand, $P(A|B)$ is called a posterior probability that indicates the cause A happens under the condition that the result B happens.
In employing for parameter estimation, event A convert to the distribution parameter θ, this is, the random variable X is assumed to have distribution parameter θ. $P(A)$ convert to $\pi(\theta)$ because θ is defined as a distributed continuous function. Then, Eq. (7) is described as follows:

$$\pi(\theta|B) \propto f(B|\theta)\pi(\theta) \tag{8}$$

Eq. (8) indicates that the product of prior distribution $\pi(\theta)$ and likelihood $f(B|\theta)$ is directly proportional to the posterior distribution $\pi(\theta|B)$. However, it is difficult to evaluate the product, because it is usually defined as a multiple integral form. Usually, a natural conjugate prior distribution or Markov chain Monte Carlo (MCMC) methods are used to evaluate the product efficiently. This study adopts the natural conjugate prior distribution.
Then, Bayesian confidence interval is evaluated by the posterior distribution to use for interval estimation, that is, θ fall in this interval with the probability of α. (α is defined as 3σ=99.6% in this research.)

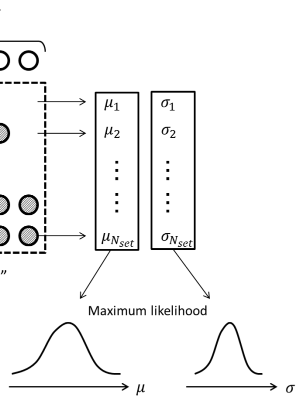

Figure 1: Estimation of prior distribution

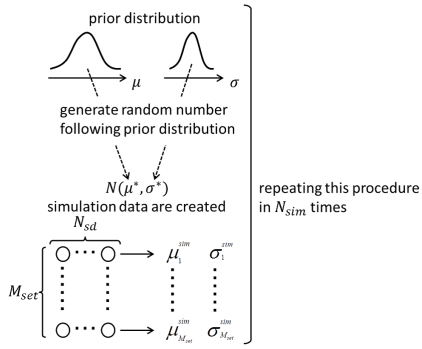

Figure 2: Procedure of creating simulation data

6. Proposed method

6.1 Estimate prior distribution

This study assumes that we have N_D data of the normally distributed random variable X, where the parameters μ_X and σ_X are unknown. In order to estimate the unknown parameter, M_D data of X are chosen among N_D data without repetition at first. This combination is called "data set" and the number of "data set" is denoted as $N_{set}(= {}_{N_D}C_{M_D})$. Each "data set" has its own estimated mean value and standard deviation, $(\mu_1, \cdots, \mu_{N_{set}})$ and $(\sigma_1, \cdots, \sigma_{N_{set}})$. From N_{set} mean values and standard deviations, the prior estimation of the parameters are performed by using the maximum likelihood method. This procedure is shown in Fig. 1.

Under the assumption that the mean value and the standard deviation are also normally distributed, their distribution parameters are evaluated as follows:

$$\mu_\mu^{pri} = \frac{1}{N_{set}} \sum_{j=1}^{N_{set}} \mu_j, \qquad \sigma_\mu^{pri^2} = \frac{1}{N_{set}} \sum_{j=1}^{N_{set}} (\mu_j - \mu_\mu)^2 \tag{9}$$

$$\mu_\sigma^{pri} = \frac{1}{N_{set}} \sum_{j=1}^{N_{set}} \sigma_j, \qquad \sigma_\sigma^{pri^2} = \frac{1}{N_{set}} \sum_{j=1}^{N_{set}} (\sigma_j - \mu_\sigma)^2 \tag{10}$$

where $\mu_{(\cdot)}^{pri}$ and $\sigma_{(\cdot)}^{pri^2}$ are the estimated mean value and the variance of the distribution parameters.

6.2 Create simulation data

To estimate the distribution parameter of X, simulation data is created from the estimated prior distributions as follows.

1. The distribution parameters μ^* and σ^* are generated from their estimated prior distribution.
2. The N_{sd} simulation data are created by Gaussian distribution N(μ^*,σ^*) and their mean value and standard deviation are evaluated.
3. The procedure of 2. repeats M_{set} times. Then, we have M_{set} number of the mean value and the standard deviation, which sets are called "simulation data set".
4. The "simulation data sets" are prepared N_{sim} sets.

This procedure is shown in Fig. 2.

6.3 Bayesian updating

Bayesian updating is formulated as follow:

$$\mu_\mu^{post} = \frac{A\mu_\mu^{pri} + B\overline{\mu}^{sim}}{A+B}, \qquad \sigma_\mu^{post} = \sqrt{\frac{1}{A+B}} \tag{11}$$

$$\mu_\sigma^{post} = \frac{C\mu_\sigma^{pri} + D\overline{\sigma}^{sim}}{C+D} \qquad \sigma_\sigma^{post} = \sqrt{\frac{1}{C+D}} \tag{12}$$

where

$$A = \frac{1}{(\sigma_\mu^{pri})^2}, \qquad B = \frac{M_{set}}{(\sigma_\mu^{sim})^2}, \qquad C = \frac{1}{(\sigma_\sigma^{pri})^2}, \qquad D = \frac{M_{set}}{(\sigma_\sigma^{sim})^2} \tag{13}$$

The posterior distribution parameters are evaluated from M_{set} data set.
It is important to confirm the estimated parameter for initial data of X. In this study, the log likelihood is evaluated for each estimated parameter. When the random value X is assumed to follow Gaussian distribution, its log likelihood is obtained as follows:

$$L = \sum_{j=1}^{M_{set}} \left(-\frac{(x_j^{sim} - \mu^{post})^2}{2(\sigma^{post})^2} - \frac{1}{2}\log\left(2\pi(\sigma^{post})^2\right) \right) \tag{14}$$

The estimated parameter with the maximum log likelihood is considered as fit for the initial data and substituted for the prior distribution parameter to iterate this process. If the parameters are considered as converged, the estimation value and the confidence interval of distribution parameters are evaluated by their distribution parameters. Using a constant α, the interval is evaluared as $(\mu - \alpha\sigma, \mu + \alpha\sigma)$ for the normal distribution.
The proposed method is summarized as follows.

Step1 M_D data of X are chosen among N_D data without repetition. (Making N_{set} number of “ data set ”)

Step2 N_{set} mean values and standard deviations, the prior estimation of the parameters are estimated by using the maximum likelihood method.

Step3 Simulation data are created from the estimated prior distributions and N_{sim} “ simulation data set ” are prepared. (see Fig. 2)

Step4 The posterior distribution parameters are evaluated by using Bayesian updating and the estimated parameter with the maximum log likelihood is considered as fit for the initial data.

Step5 The posterior distribution parameter is substituted for the prior distribution parameter.

Step6 If the parameters are considered as converged, the estimation value and the confidence interval of distribution parameters are evaluated by their distribution parameters. Otherwise, go back to step 3 with increase N_{sd}.

7. Numerical example

In this research, the property N_D, M_D, N_{sim}, M_{set} is fixed as 5, 3, 1000 and 5. N_{sd} is arithmetical progression as 5, 10, 15 $\cdots$ each repetition. Due to the importance of the standard deviation in Modified-SLSV method, if the estimated standard deviation is considered as converged, the iteration of parameter estimation is finished.

7.1 Mathematical problem

As the first example, the following two-dimensional mathematical RBDO problem [6] is considered:

$$\begin{aligned}
\text{Minimize :} \quad & f(\boldsymbol{d}) = d_1 + d_2 \\
\text{subject to :} \quad & P(g_j(\boldsymbol{x}) \le 0) \le \Phi(-\beta_j^T) \ (j = 1, 2) \\
\text{where :} \quad & g_1(\boldsymbol{x}) = \frac{x_1^2 x_2}{20} - 1 \le 0 \\
& g_2(\boldsymbol{x}) = \frac{(x_1 + x_2 - 5)^2}{30} + \frac{(x_1 - x_2 - 12)^2}{120} - 1 \le 0 \\
& 0 \le d_1 \le 10, \ 0 \le d_2 \le 10
\end{aligned} \tag{15}$$

where the design variables are set as the mean value of the random variable, $\boldsymbol{d} = \boldsymbol{\mu} = (\mu_1, \mu_2)^T$, and the target reliability is set as $\beta_j = 3$.
The random variables $\boldsymbol{x}$ follow normal distribution and the standard deviation of x_2 is known as $\sigma_2 = 0.3$. Here, the standard deviation of x_1 is unknown and should be estimated from a limited number of the experimental data shown in Table 1, where the data is made from the normal distributed random number $N(0.0, 0.3)$ because the original problem uses $\sigma_1 = 0.3$ in [6]. The distribution parameters of x_1 are evaluated by the proposed method at first. Table 2 shows the sample values, the estimated value and the confidence interval of the estimated parameters. In this problem, the number of iteration is 6 times.

Table 1: Initial experimental data of x_1 in problem 7.1

x_1	0.1952	-0.02904	-0.3141	0.09188	0.6509

Table 2: Estimated parameter values in problem 7.1

	Sample value	Estimated mean value	Estimated standard deviation	Confidence interval(3σ)
μ	0.1190	0.1189	0.0101	0.0885 to 0.1494
σ	0.3232	0.3063	0.03672	0.2781 to 0.3539

Table 3: Formulation of crashworthiness problem

(a) Design variables, side constraints, and cov in random variables

Name	Variable	Lower bound	Upper bound	σ
Thickness of B-pillar inner (mm)	d_1	0.5	1.5	0.03
Thickness of B-pillar reinforcement (mm)	d_2	0.5	1.5	0.03
Thickness of floor side inner (mm)	d_3	0.5	1.5	0.03
Thickness of cross members (mm)	d_4	0.5	1.5	0.03
Thickness of door beam (mm)	d_5	0.5	1.5	0.03
Thickness of door belt line reinforcement (mm)	d_6	0.5	1.5	unknown
Thickness of roof rail (mm)	d_7	0.5	1.5	0.03
Material yield stress for B-pillar inner (GPa)	d_8	0.192	0.750	unknown
Material yield stress for floor side inner (GPa)	d_9	0.192	0.750	0.006

(b) Objective function and ten constraints

Name		Upper limit	Formulation
Weight (kg)	f	—	$1.98 + 4.9d_1 + 6.67d_2 + 6.98d_3 + 4.01d_4 + 1.78$
Abdomen load (kN)	g_1	1.0	$1.163 - 0.3717x_2x_4 - 0.484x_3x_9$
VC upper (m/s)	g_2	0.32	$0.261 - 0.0159x_1x_2 - 0.188x_1x_8 - 0.019x_2x_7 + 0.0144x_3x_5$ $+0.08045x_6x_9$
VC middle (m/s)	g_3	0.32	$0.214 + 0.00817x_5 - 0.131x_1x_8 - 0.0704x_1x_9 + 0.031x_2x_6$ $-0.018x_2x_7 + 0.021x_3x_8 + 0.121x_3x_9 - 0.00364x_5x_6$
VC lower (m/s)	g_4	0.32	$0.74 - 0.61x_2 - 0.163x_3x_8 - 0.18x_7x_9 + 0.227x_2x_2$
Rib deflection upper (mm)	g_5	32.0	$28.98 + 3.818x_3 - 4.2x_1x_2 + 6.63x_6x_9 - 7.70x_7x_8$
Rib deflection middle (mm)	g_6	32.0	$33.86 + 2.95x_3 - 5.057x_1x_2 - 11.0x_2x_8 - 9.98x_7x_8 + 22.0x_8x_9$
Rib deflection lower (mm)	g_7	32.0	$46.36 - 9.9x_2 - 12.9x_1x_8$
Public symphysis force (kN)	g_8	4.0	$4.72 - 0.5x_4 - 0.19x_2x_3$
B-Pillar velocity (m/s)	g_9	9.9	$10.58 - 0.674x_1x_2 - 1.95x_2x_8$
Door velocity	g_{10}	0.35	$16.45 - 0.489d_3d_7 - 0.843d_5d_6$

(c) Initial experimental data of x_6 and X_8 in problem 7.2

x_6	0.9773	1.038	1.015	0.9737	0.9603
x_8	0.2954	0.3076	0.3031	0.2947	0.2921

Then, the optimum design is obtained using the estimated parameters. The optimum solution consisting the confidence interval of σ_1 is shown in Fig. 3. The "real solution" indicates the optimal solution using $\sigma_1 = 0.3$, that is included in the confidence interval. The distribution and the confidence interval indicate the effect of the uncertainty of σ_1 from estimated from the proposed method on the optimal solution. It is also found that this result has an sufficient estimation accuracy regardless that it is estimated only from five sample data.

7.2 Crashworthiness problem for side impact

As the second numerical example, the following crashworthiness problem for side impact [7] is adopted. The formulation is summarized in Table 3 (a) and (b). In this paper, the mean value of each random variable is treated as design variable $\boldsymbol{d} = \boldsymbol{\mu}$ and the target reliability is set as $\beta_j = 3$. The random variable $\boldsymbol{x}$ have independent normal distributions.

To simulate the lack of information, the standard deviations of x_6 and x_8 are set as unknown. Instead, it is determined from the simulation data from the random numbers following $N(\mu_6, \sigma_6) = N(1.0, 0.03)$ and $N(\mu_8, \sigma_8) = N(0.3, 0.006)$, where the standard deviations are selected from the original problem [7]. The simulation data are listed in Table 3 (c).

The estimated distribution parameters of x_6 and x_8 are listed in Table 4. The confidence interval of the optimal solution obtained with the confidence interval of σ_6 and σ_8 are shown Fig. 4. In this problem, the number of iteration for estimated σ_6 and σ_8 are 4 times and 5 times respectively. It is found that the estimated parameters

Table 4: The estimated distribution parameter of X_6 and X_8

	mean(estimated value)	s. d.	Confidence interval(3σ)
μ_6	0.9929	0.0045	0.9794 to 1.0064
σ_6	0.0290	0.0023	0.021 to 0.0359
μ_8	0.2986	0.0016	0.2938 to 0.3034
σ_8	0.0058	0.0008	0.0034 to 0.0082

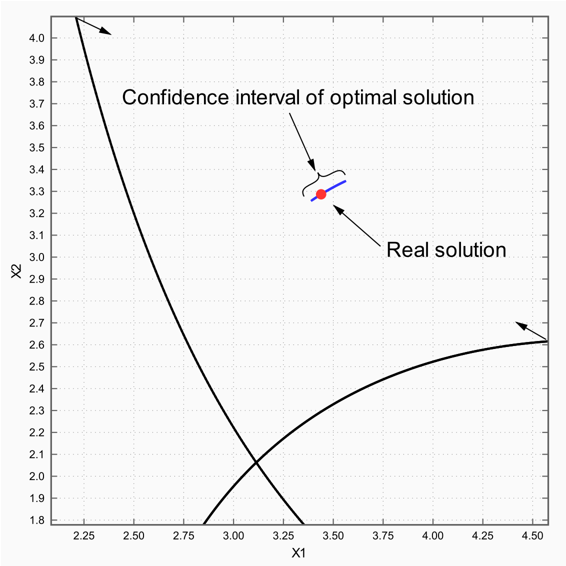

Figure 3: Confidence interval of optimal solution in problem 7.1

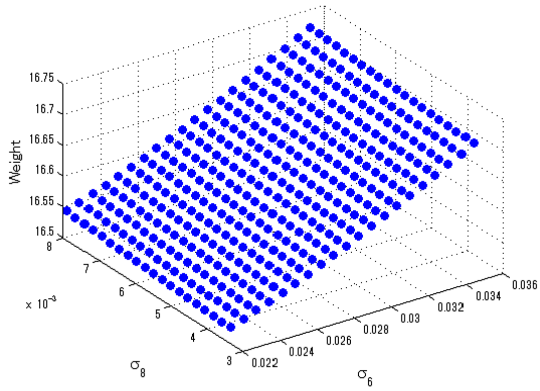

Figure 4: Confidence interval of optimal solution in problem 7.2

have almost linear effect on the optimum design.

8. Conclusion

This paper proposes a parameter estimation method with lack of information and investigates the effect of the information uncertainty on the optimum design of RBDO using the confidence interval.

The prior distribution are estimated using 'data set' which are created by an initial experimental data. To evaluate the lack of information, the simulation data by the prior distribution is used in this study. Then, using Bayesian updating method, the accuracy of estimated value is improved. Through numerical examples, it is demonstrated the validity of proposed method.

In the future, the proposed method will be used on actual design e.g., space structural system under the lack of information to estimate the distribution parameters.

References

[1] Aoues, Y. , and Chateauneuf, A. , Benchmark Study of Numerical Methods for Reliability-Based Design Optimization, *Structural and Multidisciplinary Optimization*, Vol. 41, (2010), pp. 277-294.

[2] Gunawan, S. and Papalambros, P. Y. , A Bayesian Approach to Reliability-Based Optimization With Incomplete Information, *Journal of Mechanical Design*, Vol. 128, (2006), pp. 909-918.

[3] Wang, P. Youn, B. , Xi, Z. and Kloess, A. , Bayesian Reliability Analysis With Evolving, Insufficient, and Subjective Data Sets, *Journal of Mechanical Design*, Vol. 131, (2009), pp. 111008.

[4] Kogiso, N., Young, Y. S., KIM, B. J., and LEE, J. O., Modified Single-Loop-Single-Vector Method for Efficient Reliability-Based Design Optimization, *Journal of Advanced Mechanical Design, Systems, and Manufacturing*, Vol. 6, No. 7, (2012), pp. 1-16.

[5] Chen X., Hasselman T. K., Neill, D. J., Reliability Based Structural Design Optimization for Practical Applications, *38th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference* (1997), AIAA-97-1403.

[6] Youn, B. D., and Choi, K. K., Selecting Probabilistic Approaches for Reliability-Based Design Optimization, *AIAA Journal*, Vol. 24, No. 1 (2004), pp. 124-131.

[7] Youn, B. D. Choi, K. K. Yang, R.-J. Gu, L. , Reliability-based Design Optimization for Crashworthiness of Vehicle Side Impact, *Structural and Multidisciplinary Optimization*, Vol. 26, (2004), pp. 272-283.