Proceedings of the Korean Nuclear Society Autumn Meeting
Taejon, Korea, October 1996

A New Measure of Uncertainty Importance Based on Distributional Sensitivity Analysis for PSA

Seok-Jung Han, Nam-IL Tak, and Moon-Hyun Chun

Department of Nuclear Engineering
Korea Advanced Institute of Science and Technology

Abstract

The main objective of the present study is to propose a new measure of uncertainty importance based on distributional sensitivity analysis. The new measure is developed to utilize a metric distance obtained from cumulative distribution functions (cdfs). The measure is evaluated for two cases: one is a cdf given by a known analytical distribution and the other given by an empirical distribution generated by a crude Monte Carlo simulation. To study its applicability, the present measure has been applied to two different cases. The results are compared with those of existing three methods. The present approach is a useful measure of uncertainty importance which is based on cdfs. This method is simple and easy to calculate uncertainty importance without any complex process. On the basis of the results obtained in the present work, the present method is recommended to be used as a tool for the analysis of uncertainty importance.

1. Introduction

In the precess of uncertainty quantification for PSAs one often uses a subjective assessment to quantify the uncertainties that are related to rare events and/or less known phenomena. Since the estimation of output uncertainties is affected by subjective assumptions needed to quantify input uncertainties, it is necessary to assess the relative effect of input uncertainties to output uncertainties. This result can be utilized in determining the priority used to refine uncertainties according to subjective assessment. Also it can be used to determine where future efforts should be directed in order to reduce uncertainties. The uncertainty importance based on distributional sensitivity analyses provide information on the relative contribution of input uncertainties to output uncertainties and the relative impact on the change of output distribution induced by various distributional changes in inputs.

Recently, various importance measures have been suggested and they can be classified as: (1) extensions of the Fussell-Vesely importance measure based on the variances[1-4], (2) the bivariate measures based on the shifts in quantiles of output distributions[5], (3) the information theoretic entropy measure based on the definition of Kullback-Leibler information discrimination[6], and others (Table 1). A brief summary of their characteristics can be found in Park and Ahn's paper[6].

The present paper proposes a new uncertainty importance measure based on distributional sensitivity analysis. The new measure is developed to utilize the metric distance calculated from cdfs. The metric distance provides more useful measure than the existing approaches because (1) it considers the characteristics of the entire distribution, (2) it can be directly calculated from output distribution without any prior assumptions, and (3) it can be calculated from a small size of sample data.

2. Outlines of Uncertainty Importance and Distributional Sensitivity Analysis

The analysis of uncertainty importance based on distributional sensitivity analysis has been developed to provide the information on uncertainties associated with assessment of rare events and/or less known phenomena in PSA[6]. Since the main objective of the uncertainty importance measure is to provide the information that augments normal risk measures, the basic concept of uncertainty importance is to decide uncertainty rankings of input parameters according to their relative importance, i.e., relative impacts on overall uncertainty induced by uncertainties of input parameters. Thus, the uncertainty importance measure, in conjunction with risk measures, is important for determining where future efforts should be placed in order to quantify the risk more precisely.

The procedure of uncertainty importance based on distributional sensitivity analysis is well summarized in Park and Ahn's paper[6]. Since the present work mainly focuses on a new approach of uncertainty importance measure, only its applicability has been examined following to the procedure. The detailed procedure is referred to the paper presented by Park and Ahn[6].

It should be noted here that typical cases of input distributional changes are needed to perform the uncertainty importance analysis. There are three typical cases of input distributional changes, i.e., (1) uncertainty is completely eliminated; (2) uncertainty range is changed; (3) type of distribution is changed[6]. The present study considers only one case where the uncertainty is completely eliminated because this case agrees very closely with the above mentioned reason for measuring uncertainty importance.

3. Metric Distance Measure of Uncertainty Importance

This section derives a measure of uncertainty importance in terms of a metric distance and evaluates the measure for two cases; one is a cdf given by a known distribution such as an analytical distribution function (i.e., Weibull distribution function); and the other is a cdf given by an empirical distribution function generated by crude Monte Carlo simulations.

3.1 Metric Distance

The metric distance in this study is defined as following:

$$MD^2(i{:}o) = \int_0^1 [y_\alpha^i - y_\alpha^o]^2 d\alpha \tag{1}$$

where $MD(i{:}o)$ is the metric distance using quantiles between base case and its sensitivity case, $y_\alpha{}^o$ is the αth quantile given by a distribution function for base case, and $y_\alpha{}^i$ is the αth quantile given by a distribution function for its sensitivity case.

The metric distance $MD(i{:}o)$ means an integrated distance of quantiles between base case and its sensitivity case. The metric distance measure using Eq. (1) can easily be calculated using sample data. For example, if output distributions are unknown and their estimations are complex, it is useful to perform uncertainty importance analysis using the metric distance measure $MD(i{:}o)$. The metric distance measure provides the information on how much a given input parameter impacts on output distributions when its input distribution changes. A larger $MD(i{:}o)$ means larger uncertainty changes. Thus, the input parameter that gives large values of $MD(i{:}o)$ means that it is more important than other input parameters.

3.2 Uncertainty Importance Measures for a Given Analytical Distribution Function and General Forms

Two different cases are considered in the assessment of the uncertainty importance: The first case is when the distribution is known as an analytical distribution and the other case is when the distribution is unknown. The first case is given by two-parameter Weibull distribution and for the second case an empirical distribution generated by crude Monte Carlo simulation is used.

Two-Parameter Weibull distribution: Quantiles for two-parameter Weibull distribution function is derived as follows:

$$y_\alpha = \lambda\left(\log\frac{1}{1-\alpha}\right)^{\frac{1}{\beta}} \tag{2}$$

where y_α is the αth quantile of a random variable Y under given cdf as a Weibull distribution, α is probability, and λ and β are scale and shape factor of Weibull distribution, respectively.

As given in Weibull distribution functions, the metric distance measure $MD(i{:}o)$ can be analytically evaluated by inserting Eq. (2) into Eq. (1):

$$MD^2(i{:}o) = \lambda_i^2\Gamma(\frac{2}{\beta_i}+1) + \lambda_o^2\Gamma(\frac{2}{\beta_o}+1) - 2\lambda_i\lambda_o\Gamma(\frac{1}{\beta_i}+\frac{1}{\beta_o}+1) \tag{3}$$

where the subscript o refers to base case and i to its distributional sensitivity case in which input distribution is changed. $\Gamma(x)$ denotes a gamma function. If $\lambda_i = \lambda_o$ and $\beta_i = \beta_o$, two distributions become identical and $MD(i{:}o)$ goes to zero. A larger $MD(i{:}o)$ reflects the situation where two distributions are different in their respective shapes and in their ranges of variation.

Empirical Distribution generated by a crude Monte Carlo simulation: Empirical distribution functions can be easily generated by crude Monte Carlo simulations when the mathematical relationship between inputs and outputs can be expressed by a simple functional form with an adequate accuracy. An empirical distribution function $S(y)$ can be obtained directly without any assumptions for distribution because the probability of elements within each sample is essentially the same due to the probabilistic nature of the crude Monte Carlo simulation, i.e., $P_i = 1/N$ where N is sample size:

$$S(y) = \frac{1}{N}\sum_{n=1}^{N}\delta(y > y_n) \tag{4}$$

where

$$\delta(y > y_n) = \begin{cases} 1, & \text{if} \quad y > y_n \\ 0, & \text{otherwise}. \end{cases}$$

Quantiles can easily be obtained by the inverse of an empirical distribution function in this case. If sample sizes N and M of crude Monte Carlo simulations are the same (i.e., sample sizes equal to base case and its sensitivity case), the metric distance measure $MD(i{:}o)$ becomes

$$MD^2(i{:}o) = \frac{1}{N}\sum_{n=1}^{N}[y^i_{n/N} - y^o_{n/N}]^2 \tag{5}$$

where $y^o_{n/N}$ is the (n/N)th quantile for base case, and $0 < n < N$ and $y^i_{n/N}$ is the (n/N)th quantile for its sensitivity case. As given in empirical distribution functions generated by crude Monte Carlo simulations, the $MD(i{:}o)$ can easily be calculated using Eq. (5). If sample sizes are sufficiently large, Eq. (5) applies to any cases without prior assumptions of distribution function.

4. Applicability of Metric Distance Measure

The uncertainty importance measure described in the above is now applied to two examples to assess its applicability: (1) the uncertainty analysis associated with estimations of the system unavailability or reliability obtained from a Boolean representation of a system fault tree[6]; (2) the uncertainty analysis associated with estimations of *CsI* release fraction to the environment under a hypothetical severe accident sequence of a station blackout (SBO) of Young-Gwang 3&4 nuclear power plant[7]. These examples are widely used in PSA application.

Example 1: Fault tree and event tree analyses are widely used in PSA. This example is one of the typical case of a system fault tree analysis. Suppose that the mathematical representation of a top event has the following form (adopted from Ref. 6):

$$\begin{aligned} Top(x) = {} & x_1x_3x_5 + x_1x_3x_6 + x_1x_4x_5 + x_1x_4x_6 + x_2x_3x_4 \\ & + x_2x_3x_5 + x_2x_4x_5 + x_2x_5x_6 + x_2x_4x_7 + x_2x_6x_7 \end{aligned} \tag{6}$$

This top event consists of 10 cut sets and seven events. Here x_1 and x_2 represent initiating events and are expressed as the number of occurrences per year, while the basic events x_3-x_7 represent component failure rates. Let us assume lognormal distributions for the initiating events and the basic events with mean values of 2, 3, 0.001, 0.002, 0.004, 0.005 and 0.003, respectively. To simplify the evaluations, suppose that the events are independent of each other. The metric distance measure for 7 input parameters have been calculated and the relative impacts of uncertainty have been ranked according to their magnitude as shown in Table 2.

In order to assess the applicability of the present approach, the results of metric distance measure are compared with those of three existing uncertainty importance measures, i.e., (1) standard deviation measure[2]; (2) bivariate measures[5]; and (3) information theoretic entropy measure. To compare between each measures, the calculated results of importance measures and their rankings are summarized in Table 2, and a graphical representation for rankings of each measures is shown in Fig. 1.

The comparison of rankings between each uncertainty importance measures indicates that the highly ranked parameters X_2, X_5, and X_6 are more important than the others (i.e., corresponding to the first three higher rankings), as shown in Fig. 1. Rankings of remaining parameters are unimportant because distributional changes of remaining parameters shown in Fig. 1 are negligibly small. The plotting lines of rankings obtained from four different measures are very close as shown in Fig. 1 even though rankings for each parameter are slightly different. This difference, however, is not clearly identifiable in this example.

Example 2: The evaluation of severe accident sequences using a mechanical model is an another aspect of PSA. This is one of the typical case of uncertainty importance analysis. The example is associated with the source term uncertainty quantification[7]. A response surface equation, generated by the RSM based on inputs determined from an experimental design in Ref. 7 is used as a surrogate model of the *CsI* release fraction described in the above sequence as follows:

$$\begin{aligned}\widetilde{Y}_{CsI} &= \log_{10} \widetilde{y}_{CsI} \\ &= -1.01226 + 0.02434X_1 + 0.24005X_2 - 0.02153X_3 + 0.05670X_4 + 0.04138X_5 - 0.01547X_7 \\ &\quad - 0.09064X_8 + 0.03068X_9 + 0.04505X_{10} - 0.07454X_{11} + 0.08451X_{12} + 0.10957X_1X_8 \\ &\quad - 0.05955X_1X_{12} + 0.07465X_2X_3 + 0.11925X_2X_8 - 0.11116X_2X_{10} - 0.09440X_3X_4 - 0.04404X_3X_5 \\ &\quad + 0.06014X_3X_{11} - 0.05155X_4X_5 - 0.07112X_4X_{12} + 0.02895X_5X_{10} - 0.06200X_5X_{12} + 0.16722X_{11}X_{12} \\ &\quad - 0.11357X_1^2 - 0.06022X_2^2 - 0.05629X_5^2 - 0.11337X_8^2 - 0.04698X_9^2 - 0.19005X_{10}^2 - 0.03811X_{11}^2 \\ &\quad - 0.05172X_{12}^2\end{aligned} \tag{7}$$

where $\widetilde{Y}_{CsI}$ is logarithm transformed *CsI* release fraction, and $X_1, \ldots, X_{12}$ are input parameters that are chosen by the screening referred in Ref. 7. Table 1 in Ref. 7 gives descriptions of each input parameters and estimated pdfs of those uncertainties.

The metric distance measure has been evaluated using Eq. (7). To simplify the evaluation, it is assumed that 12 input parameters are independent of one another. The metric distance measure for 12 input parameters have been calculated and the relative impacts of uncertainty have been ranked according to their magnitude and summarized in Table 3. The results of metric distance measure are compared with those of three above mentioned measures. The calculated results of importance measures and their rankings are summarized in Table 3, and a graphical representation of rankings for each measures is given in Fig. 2.

The comparison of rankings between each uncertainty importance measures indicates that the highly ranked parameters X_1, X_2, X_8, and X_{10} are more important than the others (i.e., corresponding to the first four higher rankings), as shown in Fig. 2.

5. Summary and Conclusions

A new approach of uncertainty importance analysis is presented in this paper. The present study shows that each measures can provide a reasonable result although their rankings are slightly different from each other. The present approach gives a useful measure of uncertainty importance based on the cdfs, since the measure is simple and easy to calculate the uncertainty importance without any prior process. On the basis of the results obtained in the present work, the present measure is recommended to be used as a tool for the analysis of uncertainty importance.

Acknowledgment

The authors gratefully acknowledge the financial support of the Korea Science and Engineering Foundation.

Nomenclature

- $MD(i{:}o)$ — metric distance using the quantiles between the base case and its sensitivity case
- N — sample size
- P_i — probability for an element generated by a crude Monte Carlo simulation, $P_i = 1/N$
- $S(y)$ — empirical distribution function of random variable Y
- X_1,~,X_{12} — input parameters that are chosen by the screening referred in Ref. 7
- y_α, $y_{1-\alpha}$ — the αth and $(1-\alpha)$th quantile given a distribution function of a random variable Y, respectively
- $y_{n/N}$ — the (n/N)th quantile where $0 < n < N$
- $\widetilde{Y}_{CsI}$ — logarithm transformed *CsI* release fraction
- α — probability
- β — shape factor for Weibull distribution
- δ — delta function defined by $\delta(y > y_n) = \begin{cases} 1, & \text{if } y > y_n \\ 0, & \text{otherwise} \end{cases}$
- $\Gamma(x)$ — gamma function
- λ — scale factor for Weibull distribution

References

1. V.M. Bier, "A measure of uncertainty importance for components in fault trees," *ANS Trans.*, Vol. **45** No.1, pp. 384-5, 1983.
2. R.L. Iman, "A matrix-based approach to uncertainty and sensitivity analysis for fault trees," *Risk Analysis*, Vol. **7**, No. 1, pp. 21-33, 1987.
3. J.C. Helton, R.L. Iman, J.D. Johnson and C.D. Leigh, "Uncertainty and sensitivity analysis of a model for multicomponent aerosol dynamics," *Nuclear Technology*, Vol. **73**, pp. 320-42, 1986.
4. M. Khatib-Rahbar, E. Cassoli, M. Lee, H. Nourbakhsh, R. Davis & E. Schmidt, "A probabilistic approach to quantifying uncertainties in the progression of severe accidents," *Nuclear Science and Engineering*, Vol. **102**, pp. 219-59, 1989.
5. R.L. Iman and S.C. Hora, "A robust measure of uncertainty importance for use in fault tree system analysis," *Risk Analysis* Vol. **10**, No. 3 pp. 401-6, 1990.
6. Chang K. Park and Kwang-Il Ahn, "A new approach for measuring uncertainty importance and distributional sensitivity in probabilistic safety assessment," *Reliability Engineering and System Safety*, Vol. **46**, pp. 253-261, 1994.
7. Moon-Hyun Chun, Seok-Jung Han and Nam-Il Tak, "A Combined Procedure of RSM and LHS for uncertainty analyses in source term quantifications using MAAP3.0B code," *Ann. of Nucl. Energy*, Vol. **23**, No. 16, pp.1337-1349, 1996.

Table 1 Classification and Characteristics of Uncertainty Importance Measures Suggested in the Recent PSA Study

Basic Approach	Proposers	Uncertainty Importance Measure	Criteria and Characteristics
Based on the point estimations (Extensions of the Fussell-Vesely importance measure)	Bier	$\mathrm{UI}(j)=\dfrac{\mathrm{Var}(x_j)\partial\,\mathrm{Var}(Q)}{\mathrm{Var}(Q)\partial\,\mathrm{Var}(x_j)} \cong \mathrm{Var}(Q_j)/\mathrm{Var}(Q)$	A reduction in the output variance
	Iman	$\mathrm{UI}(j)=[\mathrm{Var}(x_j)]^{1/2}\dfrac{\partial Q}{\partial x_j} \cong \left[\dfrac{(\sum_{i=1}^{n}Q_{i,j}^2)-(\sum_{i=1}^{n}Q_{i,j})^2}{n-1}\right]^{1/2}$	Based on the variances A square root of expected reduction in the output variance
	Helton *et al.*	$\mathrm{UI}(j)=\dfrac{\partial Q}{\partial x_j}\left[\dfrac{\mathrm{Var}(x_j)}{\mathrm{Var}(Q)}\right]^{1/2}\Bigg\vert_{x=x_0}$	A percentage contribution in the output: standardized regression coefficient
	Iman and Hora	$\mathrm{UI}(j)=R^2$ statistic value	A percentage change in the output: determinant coefficient of regression model
	Khatib *et al.*	$\mathrm{UI}(j)=[\mu,\sigma^2,\xi],\quad \xi=\dfrac{[Q_{j,\alpha}/Q_{j,1-\alpha}]}{[Q_\alpha/Q_{1-\alpha}]},\ \alpha=0.95$	A combination of statistical parameters
Based on the probability distribution	Iman and Hora	$\mathrm{UI}(j)=\left[\dfrac{Q_{j,\alpha}}{Q_\alpha},\dfrac{Q_{j,1-\alpha}}{Q_{1-\alpha}}\right]$	Bivariate measure of the change in the output distributions, α = quantile
	Park and Ahn	$\mathrm{UI}(j)=\int f_{Q_i}(x)\ln\left[\dfrac{f_{Q_i}(x)}{f_Q(x)}\right]dx$	Information theoretic entropy measure based on the definition of Kullback-Leibler Information discrimination
	The present method	$\mathrm{UI}^2(j)=\int_0^1(Y_\alpha^{Q_j}-Y_\alpha^{Q})^2d\alpha$	Based on the metric distance

Table 2 Summary of Uncertainty Importance Measures for Example 1 and Their Rankings based on Metric Distance, Entropy Information, Standard Deviation, and Bivariate Measures

Changed Variables	Metric Distance Measure		Entropy Information Measure		Standard Deviation Measure		Bivariate Measures $R_{0.05}$	Bivariate Measures $R_{0.95}$	
x_1	4.40E-6	(6)	1.63E-2	(4)	2.15E-5	(6)	1.07	0.99	(5)
x_2	2.91E-5	(1)	6.20E-2	(1)	7.25E-5	(1)	1.25	0.84	(1)
x_3	3.72E-6	(7)	3.56E-3	(7)	1.55E-5	(7)	1.01	0.98	(6)
x_4	1.16E-5	(4)	1.42E-2	(5)	3.72E-5	(4)	1.06	0.98	(7)
x_5	2.62E-5	(2)	2.76E-2	(3)	5.33E-5	(3)	1.16	0.94	(2)
x_6	2.46E-5	(3)	3.71E-2	(2)	5.74E-5	(2)	1.13	0.91	(3)
x_7	7.26E-6	(5)	8.83E-3	(6)	2.84E-5	(5)	1.07	0.98	(4)

Table 3 Summary of Uncertainty Importance Measures for Example 2 and Their Rankings based on Metric Distance, Entropy Information, Standard Deviation, and Bivariate Measures

Changed Variables	Metric Distance Measure		Entropy Information Measure		Standard Deviation Measure		Bivariate Measures $R_{0.05}$	Bivariate Measures $R_{0.95}$	
x_1	0.018949	(3)	0.036030	(2)	0.103676	(4)	0.922735	0.948517	(4)
x_2	0.015351	(4)	0.032816	(4)	0.248321	(1)	0.886692	1.077828	(2)
x_3	0.001607	(9)	0.005046	(9)	0.021580	(10)	0.973409	1.029660	(9)
x_4	0.001168	(11)	0.001865	(10)	0.056578	(8)	0.979166	1.018033	(11)
x_5	0.009255	(6)	0.017811	(5)	0.062610	(7)	0.961588	0.984511	(8)
x_6	0.001207	(10)	0.000511	(12)	1.02E-07	(12)	0.990663	0.969834	(10)
x_7	0.000991	(12)	0.000555	(11)	0.015529	(11)	0.985149	0.994891	(12)
x_8	0.020515	(2)	0.034182	(3)	0.138478	(3)	0.902389	0.914539	(3)
x_9	0.004431	(7)	0.007523	(7)	0.051479	(9)	0.956745	0.948993	(6)
x_{10}	0.037625	(1)	0.100524	(1)	0.180285	(2)	0.907107	0.886694	(1)
x_{11}	0.002985	(8)	0.005309	(8)	0.080841	(6)	0.961383	1.027206	(7)
x_{12}	0.009730	(5)	0.017773	(6)	0.092737	(5)	0.920321	1.006315	(5)

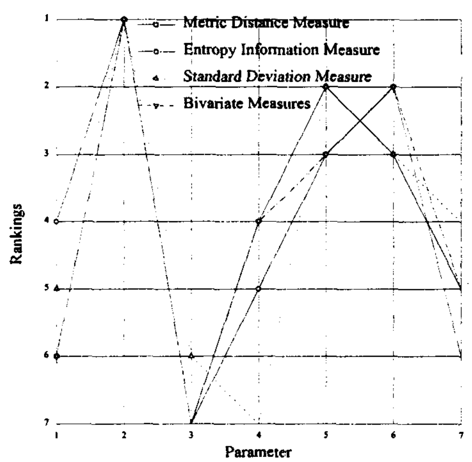

Figure 1 Uncertainty Importance Rankings of Parameters for Example 1 by Metric Distance, Entropy Information, Standard Deviation, and Bivariate Measures.

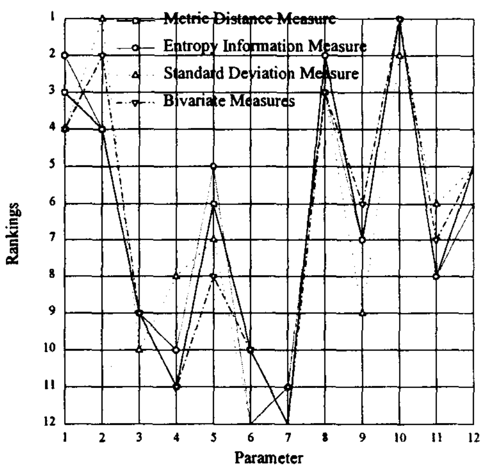

Figure 2 Uncertainty Importance Rankings of Parameters for Example 2 by Metric Distance, Entropy Information, Standard Deviation, and Bivariate Measures.