Failure probability-based global and regional sensitivity analysis using copula

Liangli He and Zhenzhou Lu

School of Aeronautics, Northwestern Polytechnical University, Xi’an, Shaanxi, 710072, China

1Email: zhenzhoulu@nwpu.edu.cn

Abstract. In risk and reliability assessment, the failure probability-based global sensitivity analysis (GSA) and the failure probability-based regional sensitivity analysis (RSA) have attracted much interest. In this article, we deduce the relationship of the failure probability-based GSA importance measure and copula, and point out that the failure probability-based GSA importance measure can be interpreted as the dependence measure between the failure probability and the input variables from copula viewpoint. To calculate the importance measure, the least square fitting copula (LSFC) method is proposed subsequently. The method decouples the double-loop estimating of the conditional failure probability. Additionally, to analyze and identify the effects of the different regions of the input variables on failure probability, a RSA importance measure is proposed, its properties are investigated and proved. At last, an engineering example is employed to demonstrate and validate the effectiveness of the LSFC method and the proposed RSA importance measure.

1. Introduction

In the realm of the risk and reliability assessment of structural systems, sensitivity analysis (SA) techniques become more and more significant. SA is defined by Saltelli as the study of how uncertainty in the output of a model can be apportioned to different sources of uncertainty in the model input [1]. It usually can be classified into three kinds: local SA [2], regional SA (RSA) and global SA (GSA). Local sensitivity analysis tests the local effect of the model output by varying input parameters one at a time by holding other parameters at nominal values. GSA aims at measuring the contribution of input uncertainty to the model output uncertainty by exploring the whole distribution ranges of model inputs. Many GSA techniques are available in the few decades, such as non-parametric technique [3-5], Variance-based GSA [6-9], and some new indices named as moment-independent importance measures [10-12]. Some corresponding computational methods [13-16] have been presented. GSA quantifies the contributions of the input variables within the whole distribution ranges to the uncertainty of the model output, but does not identify which region of the distribution of an input variable contributes the most to the uncertainty of the model output. Thus, two RSA techniques called contributions to sample mean and variance are proposed [17, 18]. Subsequently, for the extension of the RSA, References [19,20] proposed the ratio function-based on the RSA. In the reliability analysis, the failure probability is more concerned. References [21-24] proposed a GSA importance measure on the failure probability and its estimations.

In this article, two aspects of the failure probability-based sensitivity analysis techniques are concerned: the calculation method using copula for the failure probability-based GSA index and a new
RSA index on the failure probability. Firstly, we derive the index in term of copula and propose a solution called least square fitting copula (LSFC) method to estimate the index. Secondly, to satisfy the varied purposes of the risk and reliability assessment, a new RSA importance measure on failure probability is proposed to represent the effect of the internal regions of the input variables on the failure probability.

The remainder of the paper is organized as follows. In Section 2 the failure probability-based GSA importance measure and the CFP plot of the failure probability-based RSA are reviewed. In Section 3 the relationship of the failure probability-based GSA importance measure and copula is built. The Bayes importance measure for RSA on failure probability is defined, and their properties are analyzed in Section 4. The LSFC method for estimating the measure is proposed in Section 5. In Section 6, the efficiency and the accuracy of the LSFC method and the effectiveness of the proposed RSA index are demonstrated by an engineering example. The conclusions are given at the end of this paper.

2. Review on the failure probability-based GSA importance measure and the CFP plots

2.1. Failure probability-based GSA importance measure
Consider a performance function model \( Y = g(X) \), where \( X = (X_1, X_2, \ldots, X_n) \) is the n-dimensional vector of random inputs with joint probability density function (PDF) \( f_X(x) \). Denote the unconditional failure probability of the model output as \( P_f \), i.e. \( P_f = P \{ g(X) \leq 0 \} \). When the i\textsuperscript{th} input variable \( X_i \) is fixed at one realization, the conditional failure probability \( P_{f|X_i} \) can be obtained. Then the failure probability-based GSA importance measure is defined as:

\[
\eta_{X_i} = \frac{1}{2} E_{X_i} \left[ P_{f|X_i} - P_{f|X_i}\right] = \frac{1}{2} \int_{-\infty}^{+\infty} \left[ P_{f|X_i} - P_{f|X_i}\right] f_{X_i}(x_i) dx_i
\]

where \( E[\cdot] \) represents the expectation operator, and || stands for the absolute operator.

2.2. The CFP plots of the failure probability-based RSA
For a reliability model \( Y = g(X) \), reducing the range of the uncertainty of input \( X_i \) also has effect on the failure probability, thus the failure probability plot (CFP) is defined in [24] as:

\[
CFP_{X_i}(q) = \frac{1}{P_f} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \left( \prod_{i=1}^{n} f_{X_i}(x_i) I_f \right) dx_1 dx_2 \cdots dx_{i-1} dx_{i+1} \cdots dx_n
\]

where \( I_f = \begin{cases} 1 & g(X) \leq 0 \\ 0 & g(X) > 0 \end{cases} \) is the indicator function of the failure domain \( F = \{ X : g(X) \leq 0 \} \). And \( F_{X_i}^{-1}(q) \) is the inverse cumulative distribution function (CDF) of \( X_i \) at quantile \( q \).

3. The relationship of the failure probability-based GSA importance measure and copula
It is well known that copula [25-27] is a function that links the multivariate joint CDF with its one-dimensional marginal CDFs. By Sklar theory, copula function \( C \) is defined as:

\[
F_{X_1,X_2}(x_1,x_2) = C \left( F_{X_1}(x_1), F_{X_2}(x_2) \right)
\]

where if \( F_{X_1}(x_1) \) and \( F_{X_2}(x_2) \) are CDFs of \( X_1 \) and \( X_2 \), respectively.

Let \( u_i = F_{X_i}(x_i) \) and \( v = F_Y(y) \), \( F_{X_i}(x_i) \) and \( F_Y(y) \) are the CDF of the i\textsuperscript{th} input variable and the CDF of the model output respectively. The conditional CDF of \( Y \) on \( X_i = x_i \) can also be expressed as follows by conditional copula CDF:

\[
P \{ Y \leq y | X_i = x_i \} = C_{F_Y^{-1}(v)\rightarrow F_{X_i}^{-1}(u_i)}(v|u_i) = \lim_{\Delta u_i \rightarrow 0} \frac{C(u_i + \Delta u_i, v) - C(u_i, v)}{\Delta u_i} = \frac{\partial C(u_i, v)}{\partial u_i}
\]
Since \( u_i = F_{y_i}(x_i) \), \( du_i = f_{y_i}(x_i)dx_i \), the failure probability-based GSA importance measure \( \eta_{x_i} \) can be derived as follows:

\[
\eta_{x_i} = \frac{1}{2} \int_{-\infty}^{+\infty} \left[ P_{Y_i} - P_{Y_i|\{g(X) \leq 0\}} \right] f_{y_i}(x_i)dx_i = \frac{1}{2} \left[ C_{uv}(u_i, \nu^*) \right] du_i = \frac{1}{2} E_u \left[ \nu^* - C_{uv}(u_i, \nu^*) \right]
\]

(5)

where \( P_f = P\{Y \leq 0\} = F_y(0) = \nu^* \).

Equation (5) indicates that \( \eta_{x_i} \) equals half of the average absolute difference between the copula partial derivative function \( \frac{\partial C(u_i, \nu^*)}{\partial u_i} \) and the failure probability \( \nu^* \). It implies that \( \eta_{x_i} \) is uniquely and fully constructed by the failure probability \( \nu^* \) and partial derivative function; thus, it measures the correlation between \( u_i \) and \( \nu^* \). However, it is different from traditional dependence measure considering the linear relationship only, it can also measure global correlation between the tail distribution of the model output and the input variables by copula.

Furthermore, comparing with the traditional method for estimating \( \eta_{x_i} \) with double loop, according to the Equation (5), only a single set of samples is required by copula method. The detailed procedure is given in Section 5.

4. A new Bayes importance measure for RSA on failure probability and its interpretation

For the reliability model \( Y = g(X) \), usually we concern which range of the input variable \( x_i \) causes the failure, then the Bayes importance measure \( RFP_{Bay}(x_i^*) \) of failure probability-based RSA for \( X_i \in (\infty, x_i^*) \) is defined as:

\[
RFP_{Bay}(x_i^*) = \frac{P\{X_i \leq x_i^* \mid g(X) \leq 0\}}{P\{g(X) \leq 0\}} = \frac{P\{g(X) \leq 0, X_i \leq x_i^*\}}{P\{g(X) \leq 0\}}
\]

(6)

\[
= \frac{1}{P_f} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \prod_{i=1}^{n} f_{x_i}(x_i) I_f(X) dx_i dx_{i+1} \cdots dx_n
\]

\( RFP_{Bay}(x_i^*) \) identifies which range of the \( i \)th input variable \( x_i \) causes the failure, and it measures how much the probability of the variable \( x_i \) within its specific uncertainty range occurs, given that the model fails.

Note that, the \( RFP_{Bay}(x_i^*) \) index is just \( CFP_{q_i}(q) \) in Reference [24], however, the interpretation of \( CFP_{q_i}(q) \) using the physical meaning of \( RFP_{Bay}(x_i^*) \) is much more suitable and explicit.

When using copula, \( RFP_{Bay}(x_i^*) \) can be derived as:

\[
RFP_{Bay}(x_i^*) = P\{X_i \leq x_i^* \mid g(X) \leq 0\} = \frac{P\{g(X) \leq 0, X_i \leq x_i^*\}}{P\{g(X) \leq 0\}} = \frac{C(\nu^*, u_i^*)}{\nu^*}
\]

(7)

Additionally, we study and prove following properties of the \( RFP_{Bay}(x_i^*) \) index by copula function property:

Property I: \( 0 \leq RFP_{Bay}(x_i^*) \leq 1 \);

Property II: \( RFP_{Bay}(\infty, +\infty) = \frac{C(\nu^*, 1)}{\nu^*} = 1 \);

Property III: For any chosen range \( x_i^c \leq X_i \leq x_i^d \), denote \( u_i^c = P\{X_i \leq x_i^c\} \) and \( u_i^d = P\{X_i \leq x_i^d\} \), the following relationship can be deduced:
\[ RFP_{\nu_i}(x_i^*, x_i^0) = P \left\{ x_i^* \leq X_i \leq x_i^0 \mid g(X) \leq 0 \right\} \]
\[ = \frac{P \left\{ g(X) \leq 0, X_i \leq x_i^0 \right\} - P \left\{ g(X) \leq 0, X_i \leq x_i^* \right\}}{P \left\{ g(X) \leq 0 \right\}} \]
\[ = RFP_{\nu_i}(-\infty, x_i^0) - RFP_{\nu_i}(-\infty, x_i^*) \quad (8) \]

Comparing with the properties of the \( CFP_{\nu_i}(q) \) in Reference [24], it can be found that using copula to explain and prove the properties of the index is more understanding and flexible.

5. Estimation for the failure probability-based GSA and RSA index using copula

In this paper, we develop a solution for estimating \( \eta_{\nu_i} \) and the RSA index. The procedure of the proposed estimation for \( \eta_{\nu_i} \) is summarized as follows:

Step 1: Generate \( N \)-size sample of the n-dimensional input vector \( \mathbf{u}_i = \{u_{i1}, u_{i2}, \ldots, u_{in}\} \) (\( i = 1, 2, \ldots, N \)) following the uniformly distributed in the interval \([0,1]\).

Step 2: Use inverse marginal CDF to get \( x_{ji} = F_{u_i}^{-1}(u_{ji}) \), \( x_j = \{x_{j1}, x_{j2}, \ldots, x_{jn}\} \) and then the model output samples \( y = \{y_1, y_2, \ldots, y_n\} \) can be gotten by \( y_j = g(X_j) \). Then get \( v = F_v(y) \) samples.

Step 3: Build the empirical copula \( C_i \) of \( u_i \) and \( v \).

Step 5: Calculate the empirical copula \( C_i \) at the special pair \((u_i', v')\), where \( v' \) is also the failure probability of the model output.

Step 6: Use the least square fitting methods for estimating the empirical copula function \( C_i \), then the new fitting copula function \( C_i^* \) is continuous, and the partial derivative in Equation (5) can be obtained.

Step 7: Estimate the GSA index \( \eta_{\nu_i} \) by Equation (5).

Apparently, for estimating the \( \eta_{\nu_i} \) of all the inputs, only a set of \( N \) samples is needed and partial derivative of copula avoiding double-loop for estimating conditional failure probability reduces the computational cost.

For \( RFP_{\nu_0}(x^*_i) \), after getting the empirical copula function, the value \( C(v', u'_i) \) can be obtained immediately, where \( u'_i = P \left\{ X_i \leq x^*_i \right\} = F_{u_i}(x^*_i) \) and \( v' \) is the failure probability as mentioned above. Then \( RFP_{\nu_0}(x^*_i) \) can be estimated by Equation (7).

6. Engineering application: a roof truss

A roof truss is shown in Figure 1, the top boom and the compression bars are reinforced by concrete, and the bottom boom and the tension bars are all made of steel. Assume the uniformly distributed load \( q \) is applied on the roof truss, and the uniformly distributed load can be transformed into the nodal load \( P = ql / 4 \). The perpendicular deflection \( \Delta_C \) of node \( C \) can be obtained by mechanical analysis, and its preference function is
\[ \Delta_C = \frac{q l^2}{2} \left( \frac{3.81}{A_c E_c} + \frac{1.13}{A_s E_s} \right), \]
where \( A_c, A_s, E_c, E_s \) and \( l \) respectively represent sectional area, elastic modulus, length of the concrete, steel bars and the length, the distribute parameters of the independent normal random input variables are given in Table 1. Considering the safety and the applicability, \( \Delta_C \) of node \( C \) not exceeding 15cm is taken as the constraint condition, the reliability model can be constructed by
\[ Y = g = 0.015 - \Delta_C. \]
Figure 1. Schematic diagram of a roof truss.

Table 1. The distribution parameters of the basic random variables of roof truss.

| Random variable | $q$ (N/m) | $l$ / m | $A_s$ / m$^2$ | $A_c$ / m$^2$ | $E_s$ / (N/m$^2$) | $E_c$ / (N/m$^2$) |
|-----------------|-----------|-------|---------------|---------------|------------------|------------------|
| Mean            | 20000     | 12    | 9.82×10$^{-4}$| 0.04          | 2×10$^{11}$      | 3×10$^{11}$      |
| Coefficient of variance | 0.07 | 0.01 | 0.06 | 0.12 | 0.06 | 0.06 |

Table 2. Computational results of the global importance measure.

| Method | $\eta_q$ | $\eta_l$ | $\eta_{A_s}$ | $\eta_{A_c}$ | $\eta_{E_s}$ | $\eta_{E_c}$ | Cost |
|--------|----------|----------|---------------|---------------|---------------|---------------|------|
| MC     | 0.03787(1) | 0.01107(6) | 0.0215(4) | 0.0263(2) | 0.0219(3) | 0.0123(5) | 3001×6 |
| LSFC   | 0.03648(1) | 0.01053(6) | 0.0199(4) | 0.0259(2) | 0.0210(3) | 0.0119(5) | 3001×6 |

Figure 2. Estimate of the global importance measure by MC and LSFC method with failure threshold varying in interval [-0.05, 0.05].
The global importance indices $\eta_{X_i}$ of all the six inputs are computed by the MC simulation and the LSFC method, and the estimates are reported in Table 2. It can be seen that the results obtained by the both methods are within 3-7%. In order to further investigate the robustness and accuracy of the LSFC method, the failure threshold is varied in interval $[-0.05,0.05]$, and the comparison is shown in Figure 2. It can be seen that the differences between the estimates of these two methods are very tiny, and they lead to the same importance ranking. Comparison to the MC simulation with double-loop to calculate the conditional failure probability, i.e. $3001^2 \times 6$ times cost is needed by MC simulation, while the LSFC method only needs 3001 times. So the proposed LSFC method improves the computational efficiency.

![Figure 3. The RSA index $RFP_{as}(x'_i)$ plots for the roof truss.](image)

$RFP_{as}(x'_i)$ is estimated by the empirical copula method and the MC simulation. As shown in Figure 3, it indicates that the results obtained by the empirical copula solution are agree with those estimated by MC very well. Comparison the degree of the curve of the input variable deviating from the diagonal (it can be measured by the enveloping area of the curve and the diagonal), $RFP_{as}(x'_i)$ plot produces the same importance ranking of the input variable on the failure probability as that obtained by $\eta_{X_i}$. To identify which range of the most important input variable $q$ has the most severe influence on the failure probability, 10 intervals of $RFP_{as}(x'_i)$ as shown in the Figure 3 are calculated, and the ranking is:

$[0.9,1] > [0.8,0.9] > [0.7,0.8] > [0.6,0.7] > [0.5,0.6] > [0.4,0.5] > [0.3,0.4] > [0.2,0.3] > [0.1,0.2] > [0,0.1]$.

Thus, the range $F^{-1}_q(0.9) \leq q \leq F^{-1}_q(1)$ of the input variable $q$ has the greatest effect on the failure probability of the model.

7. Conclusions

In this paper, the moment-independent importance measure and the RSA measures on the failure probability were concerned. Firstly, we deduce the relationship of the measure and copula, and the LSFC method for computing the index is proposed. The proposed method avoids the nested sampling process, only needs a set of samples to estimate all indices. Thus, the proposed method is much efficient and overcomes the “dimensional curse”. The second concern is about the RSA with respect to
the failure probability. A new index is proposed. Its definition is given, and the properties are analyzed and proved, respectively. And copula is employed for estimating them. Therefore, the proposed RSA index can provide useful information for RSA.

The engineering example is introduced to validate the effectiveness of the proposed LSFC method, and to investigate the proposed RSA indices. Two main conclusions obtained. First, the LSFC method has high accuracy and the low computational cost. Second, the two proposed RSA indices can well identify the effects of the internal regions of the input variables on the failure probability of the model output, respectively.

References

[1] Saltelli A 2002 Sensitivity analysis for importance assessment Risk Analysis 22(3) 579-590
[2] Lu Z Z, Song J, Song S, et al. 2010 Reliability sensitivity by method of moments Applied Mathematical Modeling 34(10) 2860-2871
[3] Helton J C, Davis F J 2003 Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems Reliability Engineering and System Safety 81 23-69
[4] Helton J C, Davis F J 2000 Sampling-based methods In: Saltelli A, Chan K, Scott EM, editors. Sensitivity analysis New York: Wiley, 101-53
[5] Saltelli A, Marivoet J 1990 Non-parametric statistics in sensitivity analysis for model output: a comparison of selected techniques Reliability Engineering and System Safety 28 229-53
[6] Yun W Y, Lu Z Z, Zhang K C, et al. 2017. An efficient sampling method for variance-based sensitivity analysis Structural Safety 65 74-83
[7] Wang P, Lu Z Z, Xiao S N, 2017 A generalized separation for the variance contributions of input variables and their distribution parameters Applied Mathematical Modelling 47 381-399
[8] Li L Y, Lu Z Z 2017 Variance-based sensitivity analysis for models with correlated inputs and its state dependent parameter solution Structural and Multidisciplinary Optimization 56(4) 919-937
[9] Li L Y, Lu Z Z, Zhang K C et al. 2017 General validation and decomposition of the variance-based measures for models with correlated inputs Aerospace Science and Technology 62 75-86
[10] Chun M H, Han S J, Tak N I 2000 An uncertainty importance measure using a distance metric for the change in a cumulative distribution function Reliability Engineering and System Safety 70(3) 313-21
[11] Liu H B, Chen W, Sudjianto A 2006 Relative entropy based method for probabilistic sensitivity analysis in engineering design J Mech Des 128(3) 326-33
[12] Borgonovo E 2007 A new uncertainty importance measure Reliability Engineering and System Safety 92(6) 771-84
[13] Plischke E, Borgonovo E, Smith C L 2012 Global sensitivity measures from given data European Journal of Operational Research 226 536-550
[14] Wei P F, Lu Z Z, Yuan X 2013 Monte Carlo simulation for moment-independent sensitivity analysis Reliability Engineering and System Safety 110 60-67
[15] Wei P F, Lu Z Z, Song J W 2014 Moment-independent sensitivity analysis using copula Risk Analysis 34(2) 210-222
[16] Yun W Y, Lu Z Z, Jiang X, et al. 2018 Borgonovo moment independent global sensitivity analysis by Gaussian radial basis function meta-model Applied Mathematical Modelling 54 378-392
[17] Bolado L, Castaings W, Tarantola S 2009 Contribution to the sample mean plot for graphical and numerical sensitivity analysis Reliability Engineering and System Safety 94 1041-1049
[18] Tarantola S, Kopustinskas V, Bolado L, et al. 2012 Sensitivity analysis using contribution to sample variance plot: application to a water hammer model Reliability Engineering and System Safety 99 62-73
[19] Wei P F, Lu Z Z, Ruan W B, et al. 2014 Regional sensitivity analysis using revised mean and variance ratio functions *Reliability Engineering & System Safety* 121:121-135
[20] Wei P F, Lu Z Z, Wu D Q, et al. 2013 Moment-independent regional sensitivity analysis: Application to an environmental model *Environmental Modelling & Software* 47:55-63
[21] Cui L J, Lu Z Z, Zhao X P. 2010 Moment-independent importance measure of basic random variable and its probability density evolution solution *Science China Technological Sciences* 53(4):1138-1145
[22] Li LY, Lu Z Z, Feng J, et al. 2012 Moment-independent importance measure of basic random variable and its state dependent parameter solution *Structural Safety* 38:40-47
[23] Zhang L L, Lu Z Z, Chen J. 2014 An efficient method for failure probability-based moment-independent importance measure *Acta Aeronautica et Astronautica Sinica* 35(8):2199-2206
[24] Li L Y, Lu Z Z. 2013 Regional importance effect analysis of the input variables on failure probability *Computers and Structures* 125:74-85
[25] Nelsen R B. 2006 *An Introduction to Copulas*, 2nd ed. New York: Springer
[26] Genest C, Favre A C. 2007 Everything you always wanted to know about copula modeling but were afraid to ask *Journal of Hydrologic Engineering* 12:347-368
[27] Guo C M, et al. 2013 Maintenance Optimization for Systems With Dependent Competing Risks Using a Copula Function, *Eksploatacja i Niezawodnosc* - Maintenance and Reliability, 15(1):9-17