



# Importance measure of correlated normal variables and its sensitivity analysis

Wenrui Hao, Zhenzhou Lu\*, Longfei Tian

School of Aeronautics, Northwestern Polytechnical University, Xi'an 710072, China

## ARTICLE INFO

### Article history:

Received 25 May 2011

Received in revised form

13 October 2011

Accepted 18 October 2011

Available online 28 October 2011

### Keywords:

Sensitivity analysis

Importance measure

Correlated variables

Variance decomposition

Uncertainty analysis

## ABSTRACT

In order to explore the contributions by correlated input variables to the variance of the polynomial output in general engineering problems, the correlated and uncorrelated contributions by correlated inputs to the variance of model output are derived analytically by taking the quadratic polynomial output without cross term as an illustration. The analytical sensitivities of the variance contributions with respect to the distribution parameters of input variables are derived, which can explicitly expose the basic factors affecting the variance contributions. Numeric examples are employed and their results demonstrate that the derived analytical expressions are correct, and then they are applied to two engineering examples. The derived analytical expressions can be used directly in recognition of the contributions by input variables and their influencing factors in quadratic or linear polynomial output without cross term. Additionally, the analytical method can be extended to the case of higher order polynomial output, and the results obtained by the proposed method can provide the reference for other new methods.

© 2011 Elsevier Ltd. All rights reserved.

## 1. Introduction

Sensitivity analysis (SA) aims at quantifying the relative importance of each input model parameter in determining the value of an assigned output variable [1]. SA can be classified into two categories, i.e., the local SA and the global SA [2]. Since 1960s, many researchers have focused on the sensitivity analysis of the partial derivatives of structural responses, characters or indices with respect to input variables. However, those sensitivities are solved at nominal values, and cannot take the variation effects of input variables into account, so those sensitivities are local [2]. Saltelli defines uncertainty sensitivity analysis as the determination of how “uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input” [3], which is the global SA compared with local SA.

Global sensitivity analysis (GSA) is also called the importance measure analysis [4]. It can recognize contributions of different input variables to the uncertainty of model output response, and then the priority level of the input variables can be determined in experiments or research. The order determined by the importance of model input variables can help designers define the unknown parameters better to reduce the uncertain scope of response and to get an acceptable uncertain response range [5–7]. Thus, the

importance measure of input variables provides a feasible way for improving the structure models. The existing importance measures include three categories, i.e., non-parameter techniques (correlation coefficient model) [8,9], variance based methods [1,10–12], and moment independent model [2,13]. The variance based methods are widely applied because it can directly illustrate the contributions by input variables to the variance of output response.

Suppose that the input–output model is described as follows:

$$y = f(\mathbf{x}) = f(x_1, x_2, \dots, x_n), \quad (1)$$

where  $x_i$  is the  $i$ th input and  $y$  is the output. As a convention, we use upper-case letters, (i.e.,  $X_i, Y$ ) when referring to the generic aspects of variables and lower-case letters (i.e.,  $x_i, y$ ) represent their observed values. The variance-based methods are based on the decomposition of output variance [10,11]:

$$V(Y) = \sum_{i=1}^n V_i + \sum_{i=1}^n \sum_{j>i}^n V_{ij} + \dots + V_{1,2,\dots,n}, \quad (2)$$

where  $V(Y)$  is the total variance of output  $y$ ,  $V_i$  is the variance contribution by  $x_i$  to output  $y$ ,  $V_{i_1, \dots, i_k}$  is the variance contributed by the interactions between  $\{x_{i_1}, \dots, x_{i_k}\}$ .  $V_i$  is defined as follows [10,11]:

$$V_i = \text{var}(E(Y|X_i)), \quad (3)$$

\* Corresponding author. Tel./fax: +86 29 88460480.

E-mail addresses: haowr22@163.com (W. Hao),

zhenzhoulu@nwpu.edu.cn (Z. Lu), Tianlf44@163.com (L. Tian).

where  $\text{var}(\bullet)$  and  $E(\bullet)$  represent the variance and expected value, respectively. So if output response expression (1) includes no intersection and input variables are independent with each other, Eq. (2) can be simplified as

$$V(Y) = \sum_{i=1}^n V_i, \quad (4)$$

which indicates that in this case the variance of output response is the sum of the variance contributions by every individual input variable.

Considering the fact that input variables are correlated in many practical engineering problems, Xu and George [7] pointed out that for models with correlated inputs, the contribution by an individual input variable to uncertainty of the output response should be divided into two parts: the correlated one (by the correlated variations, i.e. the variations of a variable which are correlated with others) and the uncorrelated one (by the uncorrelated variations, i.e. the unique variations of a variable which cannot be explained by any other variables). Xu and George [7] illuminated that

$$V_i = V_i^U + V_i^C, \quad (5)$$

where  $V_i^C$  denotes the correlated variance contribution by  $x_i$  to output  $y$ .  $V_i^U$  is the uncorrelated variance contribution by  $x_i$  to output  $y$  and can be defined as

$$V_i^U = V(Y) - \text{var}[E(Y|\mathbf{X}_{(-i)})] = E[\text{var}(Y|\mathbf{X}_{(-i)})], \quad (6)$$

where  $\mathbf{X}_{(-i)}$  indicates the vector of all inputs except  $x_i$  and  $V_{(-i)}$  represents  $\text{var}[E(Y|\mathbf{X}_{(-i)})]$  for convenience. Eq. (6) holds with respect to (4) if there is no intersection between inputs in the model output  $y$ . From Eq. (6) we can obtain that the result of the variance of output  $y$  minus the variance contribution by all the inputs except  $x_i$  is the uncorrelated contribution by  $x_i$ . The variance contribution  $V_{(-i)}$  naturally contains the contribution by  $\mathbf{X}_{(-i)}$  correlated with  $x_i$ , in other words, the contribution by  $x_i$  correlated with other variables, namely  $V_i^C$ .

Variance based importance measure analysis of correlated input variables can help us comprehend the uncorrelated and correlated contributions by the input variables to output variance. Then we can focus on the important part to minimize the output variance. To improve the important part, we just need to obtain the sensitivities of variance contributions with respect to the distribution parameters, which are the partial derivatives of variance contributions with respect to the distribution parameters. So the importance measure analysis and the local SA are both necessary for minimizing the variance of output response, the former is to recognize the contributions by correlated input variables and the latter is to analyze influencing factors.

In engineering analysis, some input–output modelling methods are applied frequently, such as the response surface methods (RSM) [14,15], the artificial neural networks (ANN) [15–17], etc., by which we can generally get linear or quadratic input–output polynomial without cross term. Taking the quadratic polynomial output without cross term as an illustration, the correlated and uncorrelated contributions by correlated variables to the variance of output response are derived analytically in Section 2. The analytical sensitivities of the variance contributions with respect to the distribution parameters of the input variables are derived in Section 3. In Section 4, numerical examples are used to verify the analytical method, which is then applied to engineering examples. In Section 5 the mechanism of correlated and uncorrelated contributions is primarily discussed. Some conclusions are drawn in Section 6.

## 2. Importance measure analysis

We use the general form of quadratic polynomial output response,

$$y = a_0 + \sum_{i=1}^n a_i x_i + \sum_{i=1}^n b_i x_i^2, \quad (7)$$

where  $a_0$  denotes the constant term,  $a_i$  and  $b_i$  represent the coefficients of the linear term and the quadratic term, respectively.  $x_i$  is normally distributed with the mean value  $\mu_i$  and standard deviation  $\sigma_i$ , namely  $x_i \sim N(\mu_i, \sigma_i^2)$ . The Pearson correlation coefficient between  $x_i$  and  $x_j$  is  $\rho_{ij}$ . In this section, we decompose the total variance of output response expressed by Eq. (7) into uncorrelated and correlated contributions by input variables.

### 2.1. Total variance of output

The total variance of output  $y$  in Eq. (7) is

$$V(y) = \sum_{i=1}^n [a_i^2 V(x_i) + b_i^2 V(x_i^2)] + \sum_{i=1}^n \sum_{j=1}^n 2a_i b_j \text{cov}(x_i, x_j^2) + \sum_{i=1}^n \sum_{j=i+1}^n 2[a_i a_j \text{cov}(x_i, x_j) + b_i b_j \text{cov}(x_i^2, x_j^2)], \quad (8)$$

where  $\text{cov}(\bullet, \bullet)$  denotes the covariance of two variables.

Before further evolution of Eq. (8), we derived some formulas for the probability density function (PDF) of the normal variables,

$$\begin{aligned} E(x^2) &= \sigma_x^2 + \mu_x^2, \\ E(x^3) &= 3\mu_x \sigma_x^2 + \mu_x^3, \\ E(x^4) &= 3\sigma_x^4 + 6\mu_x^2 \sigma_x^2 + \mu_x^4, \\ E(x_i x_j) &= \mu_i \mu_j + \rho_{ij} \sigma_i \sigma_j, \\ E(x_i x_j^2) &= \mu_i \mu_j^2 + (\mu_i \sigma_j + 2\rho_{ij} \mu_j \sigma_i) \sigma_j, \\ E(x_i^2 x_j^2) &= \mu_i^2 \mu_j^2 + \mu_j^2 \sigma_i^2 + \mu_i^2 \sigma_j^2 + 4\rho_{ij} \mu_i \mu_j \sigma_i \sigma_j + (2\rho_{ij}^2 + 1) \sigma_i^2 \sigma_j^2, \end{aligned} \quad (9)$$

where  $x \sim N(\mu_x, \sigma_x^2)$ ,  $x_i \sim N(\mu_i, \sigma_i^2)$ .

Using the formula of variance and the properties of covariance [18], Eq. (8) evolves as

$$\begin{aligned} V(y) &= \sum_{i=1}^n a_i^2 \text{var}(x_i) + \sum_{j=1}^m b_j^2 [E(x_j^2) - E^2(x_j^2)] \\ &+ \sum_{i=1}^n \sum_{j=i+1}^n 2a_i a_j [E(x_i x_j) - E(x_i)E(x_j)] \\ &+ \sum_{i=1}^n \sum_{j=1}^m 2a_i b_j [E(x_i x_j^2) - E(x_i)E(x_j^2)] \\ &+ \sum_{i=1}^m \sum_{j=i+1}^m 2b_i b_j [E(x_i^2 x_j^2) - E(x_i^2)E(x_j^2)], \end{aligned} \quad (10)$$

combined with Eq. (9), the total variance of output  $y$  is obtained as

$$\begin{aligned} V(y) &= \sum_{i=1}^n [a_i^2 + 2b_i^2 (\sigma_i^2 + 2\mu_i^2)] \sigma_i^2 + \sum_{i=1}^n \sum_{j=1}^n 4a_i b_j \mu_j \rho_{ij} \sigma_i \sigma_j \\ &+ \sum_{i=1}^n \sum_{j=i+1}^n 2[a_i a_j + 2b_i b_j (2\mu_i \mu_j + \rho_{ij} \sigma_i \sigma_j)] \rho_{ij} \sigma_i \sigma_j. \end{aligned} \quad (11)$$

متن کامل مقاله

دریافت فوری ←

**ISI**Articles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات