Applied Mathematical Modelling 42 (2017) 58–70

Contents lists available at ScienceDirect

Applied Mathematical Modelling

journal homepage: www.elsevier.com/locate/apm

A kernel estimate method for characteristic function-based uncertainty importance measure

Xin Xu [a,b], Zhenzhou Lu [a,*], Xiaopeng Luo [a,b]

[a] School of Aeronautics, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China
[b] Department of Chemistry, Princeton University, Princeton, NJ 08544, USA

A R T I C L E I N F O

Article history:
Received 25 February 2014
Revised 16 August 2016
Accepted 22 September 2016
Available online 12 October 2016

Keywords:
Uncertainty importance analysis
Moment-independent
Characteristic function
Kernel estimate method

A B S T R A C T

In this paper, we propose a fast computation method based on a kernel function for the characteristic function-based moment-independent uncertainty importance measure θ_i. We first point out that the possible computational complexity problems that exist in the estimation of θ_i. Since the convergence rate of a double-loop Monte Carlo (MC) simulation is $O(N^{-1/4})$, the first possible problem is the use of double-loop MC simulation. And because the norm of the difference between the unconditional and conditional characteristic function of model output in θ_i is a Lebesgue integral over the infinite interval, another possible problem is the computation of this norm. Then a kernel function is introduced to avoid the use of double-loop MC simulation, and a longer enough bounded interval is selected to instead of the infinite interval to calculate the norm. According to these improvements, a kind of fast computational methods is introduced for θ_i, and during the whole process, all θ_i can be obtained by using a single quasi-MC sequence. From the comparison of numerical error analysis, it can be shown that the proposed method is an effective and helpful approach for computing the characteristic function-based moment-independent importance index θ_i.

1. Introduction

Global sensitivity analysis, defined by Saltelli et al. [1], was developed to clarify and quantify the relationship between the uncertainty in the output and inputs of a model. The study of global sensitivity analysis has been carried out in 1973 by Cukier et al. [2], and since then, more and more relevant works have assisted to the growth of sensitivity analysis. As Borgonovo and Plischke [3] indicate, sensitivity analysis is a crucial step in the model building and result communication process. Through sensitivity analysis we gain essential insights on model behavior, on its structure and on its response to changes in the model inputs. And it is nowadays advocated as an essential part of the modeling and risk assessment of complex systems [4–8].

The research in global sensitivity analysis includes nonparametric methods [9], variance-based methods [10,11], distribution-based methods [7,12], and value of information methods [13,14]. The nonparametric method has been first developed, and presents the influence of uncertain inputs on output uncertainty with correlation, but it does not reflect the

* Corresponding author. Fax: +86 29 88460480.
E-mail addresses: xuxin103@163.com, xx2@princeton.edu (X. Xu), zhenzhoulu@nwpu.edu.cn (Z. Lu), luo_works@163.com, xiaopeng@princeton.edu (X. Luo).

http://dx.doi.org/10.1016/j.apm.2016.09.028

input-output correlation for most nonlinear models [7]. The variance-based one is widely used in many fields, such as financial investment risk analysis, reliability engineering risk analysis, etc. In practice, the variance-based main effect and the variance-based total effect are two most popular sensitivity measures [1]. They consider the influence of input on the model output over the entire domain of the inputs, and enable us to deal with the model with the interaction of the inputs.

However, as Saltelli [4] indicated, the variance-based measure assumes that this moment is sufficient to describe the output variability. Hence the use of variance-based importance index is not always effective at measuring uncertainty importance [15]. Instead of a particular moment, Borgonovo [7] believed that an importance index should involve the entire output distribution in uncertainty importance analysis and he defined a moment-independent importance measure δ_i by the notion of probability density function (PDF) [15]. One of the advantages of distribution-based measures is that it is possible to give the global importance description by using the entire distribution of the model output. Although many works were proposed for computing δ_i [16–22] and several authors obtained sensitivity measures using operators based on divergences between densities [23–25], they still have a disadvantage we should pay attention to, that is, it is hard to estimate the PDF. Xu et al. [21] indicated the reasons of this problem may include two aspects. The first one is that the convergence rate is always lower than $N^{-1/2}$ (where N is the number of sample points) for estimating the value of PDF at some predetermined point [26], and the other one is the estimation of PDF is an ill-posed computational problem. For more details about the global sensitivity analysis, we refer the reader to [27].

In this paper, we compute an importance indicator (called θ_i) on the basis of the characteristic function (CF), which can be regarded as a dual representation of Borgonovo's density-based index δ_i. The CF is a continuous and bounded function with compact support in practice, and its computation is well-posed. Besides, for any fixed point, the convergence rate of estimating CF is at least $O(N^{-1/2})$. Although CF provides a more powerful tool than PDF, it would have to pay with introducing complex-valued functions and random variables. Moreover, because the computation of θ_i is performed with a double-loop Monte Carlo (MC) simulation that the convergence rate is $O(N^{-1/4})$, the efficiency of θ_i is much slower than that of variance-based indices. Borgonovo et al. [27] also indicate that, one of the traditional challenges in the use of probabilistic sensitivity measures is that their estimation requires double-loop MC simulation. But we believe that this computational complexity is nonessential for θ_i. So we want to find a way for computing θ_i from a given sample without a double-loop.

This paper's outline is as follows: Section 2 illustrates the CF-based importance measure θ_i and provides an analysis of computational complexity. Then the improvement of these problems is discussed in Section 3, and the computational algorithms of the proposed computational method for measure θ_i are proposed in Section 4. In Section 5 we provide several examples with a comparison of numerical error analysis. Section 6 summarizes the relevant results.

2. CF-based importance measure and problem statement

2.1. CF-based importance measure

Suppose there is a model $Y = g(\boldsymbol{X})$, where Y is the model output, and $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)$ are uncertain input parameters. Let $\phi_Y(t)$ be the unconditional CF of output Y and $\phi_{Y|X_i}(t)$ be the conditional CF of Y on X_i $(i = 1, 2, \ldots, d)$. Then

$$\phi_Y(t) = \int_{-\infty}^{+\infty} e^{ity} f_Y(y) dy,$$

and

$$\phi_{Y|X_i}(t) = \int_{-\infty}^{+\infty} e^{ity} f_{Y|X_i}(y) dy,$$

where $f_Y(y)$ and $f_{Y|X_i}(y)$ (if they exist) are the unconditional and conditional PDF of Y on X_i, respectively. Denote

$$d(X_i) = \|\phi_Y(t) - \phi_{Y|X_i}(t)\|_1 = \int_{-\infty}^{+\infty} \left|\phi_Y(t) - \phi_{Y|X_i}(t)\right| dt, \tag{1}$$

then $d(X_i)$ is a real-valued function only dependent on random variable X_i. The expectation of $d(X_i)$, denoted by E_i, is given by

$$E_i = \mathbb{E}_{X_i}(d(X_i)) = \int d(X_i) f_{X_i}(x_i) dx_i. \tag{2}$$

By normalizing E_i in accordance with $\sum_{m=1}^{d} E_m$, the CF-based importance measure θ_i is defined as

$$\theta_i = \frac{E_i}{\sum_{m=1}^{d} E_m} \tag{3}$$

From the definition of θ_i, we can obtain that both the proposed importance measure and the Borgonove's measure δ_i have similar mathematical properties (see Table 1).

In fact, the single component E_i can be defined as a CF-based importance measure, and both E_i and θ_i indicate the same object with different choice of unit of measure.