Journal of Statistical Planning and Inference 147 (2014) 212–223

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Total interaction index: A variance-based sensitivity index for second-order interaction screening

J. Fruth[a,*], O. Roustant[b], S. Kuhnt[a,c]

[a] Faculty of Statistics, TU Dortmund University, Vogelpothsweg 87, 44227 Dortmund, Germany
[b] Ecole Nationale Supérieure des Mines, FAYOL-EMSE, LSTI, F-42023 Saint-Etienne, France
[c] Dortmund University of Applied Sciences and Arts, Emil-Figge-Strasse 42, 44227 Dortmund, Germany

ARTICLE INFO

Article history:
Received 5 November 2012
Received in revised form
14 November 2013
Accepted 15 November 2013
Available online 4 December 2013

Keywords:
Sensitivity analysis
FANOVA decomposition
Sobol indices
High-order interactions
Superset importance
Additive structure

ABSTRACT

Sensitivity analysis aims at highlighting the input variables that have significant impact on a given model response of interest. By analogy with the total sensitivity index, used to detect the most influential variables, a screening of interactions can be done efficiently with the so-called total interaction index (TII), defined as the superset importance of a pair of variables. Our aim is to investigate the TII, with a focus on statistical inference. At the theoretical level, we derive its connection to total and closed sensitivity indices. We present several estimation methods and prove the asymptotical efficiency of the Liu and Owen estimator. We also address the question of estimating the full set of TIIs, with a given budget of function evaluations. We observe that with the pick-and-freeze method the full set of TIIs can be estimated at a linear cost with respect to the problem dimension. The different estimators are then compared empirically. Finally, an application is given aiming at discovering a block-additive structure of a function, where no prior knowledge is available, neither about the interaction structure nor about the blocks.

1. Introduction

Global sensitivity analysis has broad applications in screening, interpretation and reliability analysis (Morris et al., 2006; Saltelli et al., 2000). A well-established method is the estimation of Sobol indices which quantify the influence of variables, or groups of variables, on the variability of an output. First-order Sobol indices and closed Sobol indices quantify the single influence of variables and groups of variables, respectively (Sobol, 1993). Homma and Saltelli (1996) introduced the total sensitivity index which measures the influence of a variable jointly with all its interactions. If the total sensitivity index of a variable is zero, this variable can be removed because neither the variable nor its interactions — at any order — have an influence. Thus the total sensitivity index can be used to detect the essential variables, a procedure often called *screening* (Saltelli et al., 2006).

By analogy with the total sensitivity index, we consider the so-called total interaction index (TII) that measures the influence of a pair of variables together with all its interactions. The TII is a particular case of superset importance, a sensitivity index investigated in Hooker (2004) and Liu and Owen (2006). If the TII of a pair of variable $\{X_i, X_j\}$ is zero, then there is no interaction term containing simultaneously X_i and X_j, which leads to the elimination of the pair $\{X_i, X_j\}$ from the

* Corresponding author.
E-mail addresses: fruth@statistik.tu-dortmund.de (J. Fruth), roustant@emse.fr (O. Roustant), sonja.kuhnt@fh-dortmund.de (S. Kuhnt).

0378-3758/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jspi.2013.11.007

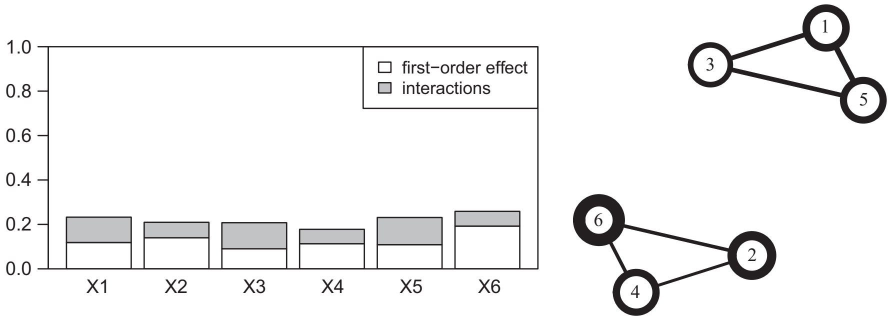

Fig. 1. Sensitivity analysis of the example experiment. First-order and total Sobol indices (left), total interaction indices (right).

list of possible interactions. By analogy with screening, this can be viewed as *interaction screening*. More precisely this is second-order interaction screening, since we consider pairs of variables.

The main benefit of TII is to discover groups of variables that do not interact with each other. To illustrate this, let us consider a short example. Consider the following function, supposed to be unknown, which we want to analyze based on a limited number of evaluations:

$$f(X_1,\dots,X_6) = \cos([1,X_1,X_5,X_3]\beta) + \sin([1,X_4,X_2,X_6]\gamma)$$

with $X_k \overset{\text{i.i.d.}}{\sim} U[-1,1]$, $k=1,\dots,6$, $\beta=[-0.8,-1.1,1.1,1]'$ and $\gamma=[-0.5,0.9,1,-1.1]'$, where the prime stands for the transpose. If we estimate the common first-order and total sensitivity indices (Fig. 1, left), we detect that all variables are active, on their own as well as by interactions, but not which variables are involved in the interactions and by what amount. Now we estimate the TII for each combination of two variables. A convenient way to present the TII is by a graph, where the thickness of the vertex circle represents the first-order index, and the thickness of the edge between two vertices represents the TII of the two variables. Now the interaction structure, here a partition into two groups, is clearly visible (Fig. 1, right).

This interaction structure corresponds to an additive structure of the analyzed function. This can be advantageously exploited for metamodelling (see Muehlenstaedt et al., 2012), and for optimization: The 6-dimensional optimization problem of minimizing f simplifies into two 3-dimensional ones.

The aim of our paper is to investigate the TII, with a focus on statistical inference. Our main result is the asymptotical efficiency (van der Vaart, 1998) of the estimator proposed by Liu and Owen (2006). The paper is structured as follows. Section 2 presents theoretical results concerning the TII. Several estimation methods are deduced (Section 3), and asymptotical properties of the *method by Liu and Owen* are proved in Section 4. The question of estimating all the TIIs with a given budget of function evaluations is studied in Section 5. Finally the TII is used to recover the block-additive decomposition of a 12-dimensional function. Throughout the paper a capital letter like X_i indicates a single random variable where a lowercase letter like x_i indicates a realization of the variable, e.g. a Monte Carlo random sample of the distribution of X_i. A bold letter like $\boldsymbol{X}$ indicates a vector of variables.

2. Theoretical aspects

2.1. A quick overview of FANOVA decomposition and Sobol indices

Assume a set of independent random variables $\boldsymbol{X}=\{X_1,\dots,X_d\}$, and let ν denote the probability measure of $\boldsymbol{X}=(X_1,\dots,X_d)$. Then for any function $f\in L^2(\nu)$, the functional ANOVA decomposition provides a unique decomposition into additive terms:

$$f(\boldsymbol{X}) = \mu_0 + \sum_{i=1}^{d}\mu_i(X_i) + \sum_{i<j}\mu_{i,j}(X_i,X_j) + \cdots + \mu_{1,\dots,d}(X_1,\dots,X_d). \tag{1}$$

The terms represent first-order effects ($\mu_i(X_i)$), second-order interactions ($\mu_{i,j}(X_i,X_j)$) and all higher-order combinations of input variables. Efron and Stein (1981) show that the decomposition is unique if all terms on the right hand side of (1) have zero mean:

$$\mathrm{E}(\mu_I(\boldsymbol{X}_I)) = 0, \quad I\subseteq\{1,\dots,d), \tag{2}$$

and the conditional expectations fulfill

$$\mathrm{E}(\mu_{i,i'}(X_iX_{i'})|X_i) = \mathrm{E}(\mu_{i,i'}(X_iX_{i'})|X_i{'}) = \mathrm{E}(\mu_{i,i',i''}(X_iX_{i'}X_{i''})|X_iX_{i'}) = \mathrm{E}(\mu_{1,\dots,n}(X_1\cdots X_n)|X_1\cdots X_{n-1}) = \cdots = 0, \tag{3}$$

which implies the orthogonality of all terms in (1).

Generally, due to the independence in $\boldsymbol{X}$ in our framework, the conditional expectation of a functional reduces to

$$\mathrm{E}(h(X_j,X_k)|X_j=x_j) = \int h(x_j,x_k)d\nu_k(x_k).$$