Reliability Engineering and System Safety 130 (2014) 140–148

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

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

Non-parametric kernel estimation for the ANOVA decomposition and sensitivity analysis

Xiaopeng Luo, Zhenzhou Lu*, Xin Xu

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

ARTICLE INFO

Article history:
Received 13 March 2013
Received in revised form
26 May 2014
Accepted 2 June 2014
Available online 9 June 2014

Keywords:
ANOVA decomposition
Sensitivity analysis (SA)
Non-parametric methods
Kernel estimate
Higher-order kernels
Conditional moments

ABSTRACT

In this paper, we consider the non-parametric estimation of the analysis of variance (ANOVA) decomposition, which is useful for applications in sensitivity analysis (SA) and in the more general emulation framework. Pursuing the point of view of the state-dependent parameter (SDP) estimation, the non-parametric kernel estimation (including high order kernel estimator) is built for those purposes. On the basis of the kernel technique, the asymptotic convergence rate is theoretically obtained for the estimator of sensitivity indices. It is shown that the kernel estimation can provide a faster convergence rate than the SDP estimation for both the ANOVA decomposition and the sensitivity indices. This would help one to get a more accurate estimation at a smaller computational cost.

1. Introduction

The main task of sensitivity analysis (SA) [1,2] is to clarify and quantify the relationship between the distribution of output y and the distribution of an input x_i under a given model $y=g(x_1,x_2,\dots,x_d)$, where y is a random variable and $\boldsymbol{x}=(x_1,x_2,\dots,x_d)$ is a d-dimensional random vector. This relationship depends not only on the distribution of the inputs but also on the structure of the given model. The analysis of variance (ANOVA) decomposition [3–6] expresses a kind of structure on the set of multi-dimensional functions from the space $\mathcal{L}_2$ of square integrable functions, so it tends to become a basic point of departure of SA [7,8].

The ANOVA decomposition of $y=g(x_1,x_2,\dots,x_d)$ is an expression [3,5]

$$y=g_0+\sum_i g_i(x_i)+\sum_{i<j} g_{ij}(x_i,x_j)+\cdots+g_{12\cdots d}(x_1,\dots,x_d) \tag{1}$$

where

$$g_0=\mathbb{E}(y)$$
$$g_i=\mathbb{E}(y|x_i)-g_0$$
$$g_{ij}=\mathbb{E}(y|x_i,x_j)-g_i-g_j-g_0$$
$$\vdots$$

The decomposition of the variance, V, of model output y is based on the ANOVA decomposition [4,6]:

$$V=\sum_i V_i+\sum_{i<j} V_{ij}+\cdots+V_{12\cdots d} \tag{2}$$

where

$$V_i=\mathrm{Var}[\mathbb{E}(y|x_i=x_i^*)]$$
$$V_{ij}=\mathrm{Var}[\mathbb{E}(y|x_i=x_i^*,x_j=x_j^*)]-V_i-V_j$$
$$\vdots$$

Normalizing by the unconditional variance V, the corresponding sensitivity indices (main effect) are defined as [4]

$$S_i=\frac{V_i}{V}$$
$$S_{ij}=\frac{V_{ij}}{V}$$
$$\vdots \tag{3}$$

And the total sensitivity index (total effect) is defined as [4]

$$S_{Ti}=S_i+\sum_j S_{ij}+\sum_{j<k} S_{ijk}+\cdots \tag{4}$$

The main effect and the total effect are the two most popular variance based sensitivity measures [1,9]. The main effect S_i represents the average output variance reduction that can be achieved when x_i is fixed, while total effect S_{Ti} stands for the average output variance that would remain as long as x_i stays unknown, that is, the total contribution of x_i to the output variation. The difference between S_{Ti} and S_i denotes the degree of interaction between this input and other inputs.

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

http://dx.doi.org/10.1016/j.ress.2014.06.002
0951-8320/

There are clear links between the variance based SA and the ANOVA decomposition of model. As an approximation, the decomposition expression can be used to compute variance based sensitivity indices in place of the original model. To get this decomposition of the model, an obvious way is to estimate the class of the functions $\mathbb{E}(y|\boldsymbol{x}_{\mathbf{I}})$, where $\boldsymbol{x}_{\mathbf{I}}$ denotes a group of inputs indexed by $\mathbf{I}$ and $\mathbf{I}$ denotes a subset of $\{1,2,\ldots,d\}$. Clearly, the estimation of $\mathbb{E}(y|\boldsymbol{x}_{\mathbf{I}})$ provides an approach for both model approximations and sensitivity estimations.

The State-Dependent Parameter (SDP) method is applied for this purpose with the point of view of non-parametric smoothing [9,10]. And it is first suggested by Young [11,12]. The estimation is performed with the help of the classical recursive Kalman filter and associated fixed interval smoothing algorithms. Pursuing the point of view of non-parametric smoothing, kernel estimation is also a good choice for $\mathbb{E}(y|\boldsymbol{x}_{\mathbf{I}})$. The kernel-based method is one of the most popular non-parametric estimators [13,14].

The kernel method is first introduced by Rosenblatt [15] for density estimation. Nadaraya [16] and Watson [17] independently proposed nonparametric estimators of $\mathbb{E}(y|\boldsymbol{x}_{\mathbf{I}})$ based on the kernel method. Rosenblatt [18] obtained the bias, variance and asymptotic distribution of those kernel estimation. And the corresponding analysis of convergence rate is considered by Stone [19,20], Ibragimov and Hasminski [21] and Yatracos [22,23]. Related problems concerning optimal convergence rates for kernel estimation of various functionals of $\mathbb{E}(y|\boldsymbol{x}_{\mathbf{I}})$ are discussed in Fan [24,25], Donoho and Low [26], Efromovich and Low [27], and Eubank [14]. One of the key issues of the kernel method is the optimal choice of the bandwidth as a most important parameter of kernel estimation.

In this work, we first consider the ANOVA decomposition of $y=g(x_1,x_2,\ldots,x_d)$ by using a kernel estimation of $\mathbb{E}(y|\boldsymbol{x}_{\mathbf{I}})$, then the variance based SA is discussed. The remainder of this paper is organized as follows. Section 2 discusses the kernel estimation of the ANOVA decomposition, and the high order kernel method is introduced to further improve the convergence rate. Section 3 considers the numerical error analysis for both SDP method and kernel method. In Section 4, we investigate the kernel estimators of sensitivity indices on the basis of the kernel estimator of ANOVA decomposition. Specially, we obtain the asymptotic convergence rate of these estimators of sensitivity indices. In Sections 5 and 6, we provide detailed analyses for the Ishigami test function and the Sobol g-function, and we also have a comparison of convergence rate between SDP method and kernel method. Conclusions are offered in Section 7.

2. Kernel estimator for ANOVA decomposition

2.1. Kernel method

In this subsection, we give a brief survey of the basic theory of kernel method. Specially, much emphasis is given to the asymptotic expression of the error. The higher-order kernel estimator built in Section 2.3 depends on it directly.

Suppose that $y=g(x_1,x_2,\ldots,x_d)$ is a given model, and we have the uniformly distributed sample matrix $\boldsymbol{A}$ and response matrix $\boldsymbol{B}$ with the size N:

$$\boldsymbol{A}=\begin{pmatrix} X_1^1 & X_2^1 & \cdots & X_d^1 \\ X_1^2 & X_2^2 & \cdots & X_d^2 \\ \vdots & \vdots & \cdots & \vdots \\ X_1^N & X_2^N & \cdots & X_d^N \end{pmatrix}, \quad \boldsymbol{B}=\begin{pmatrix} Y^1 \\ Y^2 \\ \vdots \\ Y^N \end{pmatrix} \tag{5}$$

For convenience, we write $t=x_i$, $m(t)=\mathbb{E}(y|x_i)$, $T_k=X_i^k$, and $Y_k=Y^k, k=1,2,\ldots,N$ for any given i in this subsection. According to the point of view of smoothing [9], there are N observations $(T_1,Y_1),\ldots,(T_N,Y_N)$ which follow the relationship

$$y(t)=m(t)+\epsilon(t) \tag{6}$$

where ϵ^k $(k=1,\ldots,N)$ are the zero mean, uncorrelated random noises with variance σ^2. And without loss of generality, we suppose that (T_k,Y_k) $(k=1,\ldots,N)$ are sorted in the ascending order of t and normalized in $[0,1]$, that is

$$0\le T_1<T_2<\cdots<T_N\le 1 \tag{7}$$

The kernel estimator of $m(t)$ is defined as [16,17]

$$\hat{m}_N(t;h)=\frac{1}{hN}\sum_{k=1}^{N}K\left(\frac{t-T_k}{h}\right)Y_k \tag{8}$$

where h is the scale and $K(u)$ is the kernel function with compact support $[-1,1]$ satisfying the following moment conditions [14]:

$$\int_{-1}^{1}K(u)\,du=1 \tag{9a}$$

$$\mu_1=\int_{-1}^{1}uK(u)\,du=0 \tag{9b}$$

$$\mu_2=\int_{-1}^{1}u^2K(u)\,du>0 \tag{9c}$$

and

$$V=\int_{-1}^{1}K(u)^2\,du<\infty \tag{9d}$$

The basic kernel function considered in this paper is $K(u)=C(1-u^2)^p$, $x\in[-1,1]$, where p is a positive integer, C is a constant such that the moment conditions are satisfied. For example, if $p=2$, then $C=15/16$. And at this time, $\mu_2=1/7$ and $V=5/7$ (see Table 1 for more examples).

The scale h is usually referred to as the bandwidth, and it depends on the sample size N, that is, $h\to 0$ as $N\to\infty$ [14]. As we noted above, the important issue is the optimal choice of h in Eq. (8), because the performance of $\hat{m}_N(t;h)$ as an estimator of $m(t)$ depends crucially on its value [13]. A well-studied criterion used to determine an optimal h is the Mean Integrated Squared Error (MISE) [14]:

$$\text{MISE}\{\hat{m}_N\}(t;h)=\mathbb{E}\int[\hat{m}_N(t;h)-m(t)]^2\,dt \tag{10}$$

which is conveniently decomposed into integrated squared bias and integrated variance components [13]:

$$\text{MISE}\{\hat{m}_N(t;h)\}=\int(\underbrace{\mathbb{E}[\hat{m}_N(t;h)]-m(t)}_{\text{pointwise bias of } m})^2\,dt+\int\underbrace{\text{Var}[\hat{m}_N(t;h)]}_{\text{pointwise variance of } m}\,dt \tag{11}$$

Table 1
Kernel functions for the parameters.

p	C	μ_2	V
1	$\frac{3}{4}$	$\frac{1}{5}$	$\frac{3}{5}$
2	$\frac{15}{16}$	$\frac{1}{7}$	$\frac{5}{7}$
3	$\frac{35}{32}$	$\frac{1}{9}$	$\frac{350}{429}$
5	$\frac{693}{512}$	$\frac{1}{13}$	$\frac{4158}{4199}$
10	$\frac{969\,969}{524\,288}$	$\frac{1}{23}$	$\frac{208\,866\,658}{156\,835\,045}$
15	$\frac{300\,540\,195}{134\,217\,728}$	$\frac{1}{33}$	$\frac{72\,165\,711\,623\,400}{45\,029\,832\,938\,783}$
20	$\frac{1\,412\,926\,920\,405}{549\,755\,813\,888}$	$\frac{1}{43}$	$\frac{19\,396\,077\,650\,622\,530}{10\,575\,651\,537\,777\,253}$