Statistica Sinica **28** (2018), 817-837
doi:https://doi.org/10.5705/ss.202016.0035

SENSITIVITY ANALYSIS USING PERMUTATIONS

Shifeng Xiong[1], Xu He[1], Yuanzhen He[2] and Weiyan Mu[3]

[1]*Chinese Academy of Sciences,* [2]*Beijing Normal University*
and [3]*Beijing University of Civil Engineering and Architecture*

Abstract: Sensitivity analysis quantifies the uncertainty in an input-output system by measuring the influence of the inputs on the output. This article presents a new sensitivity index by permuting the observations of an input. The proposed index is related to a statistical problem of testing the significance of the input, and thus possesses some frequentist properties that the current sensitivity analysis methods do not have. Numerical simulations and an application are presented to illustrate the proposed method.

Key words and phrases: Kriging, permutation test, significance, Sobol' index, uncertainty quantification.

1. Introduction

Sensitivity analysis studies how the uncertainty in the output of an input-output system can be apportioned to different inputs (Saltelli et al. (2008)). It can be useful for a range of purposes, including understanding of the influence of each input parameter on the output (or response), selecting important inputs, and simplifying the input-output model (Pannell (1997)). In this sense, sensitivity analysis techniques have important applications in such fields as science, engineering, and econometrics. A wealth of sensitivity analysis methods have been proposed in the literature with different focuses. They can be grouped into two classes (Sullivan (2015)): local sensitivity analysis, which studies the sensitivity of the response to variations in its inputs at or near a particular base point, as exemplified by the derivative-based methods (Griewank and Walther (2008)); global sensitivity analysis, which studies the "average" sensitivity of the response to variations of its inputs across the domain. Here we focus on global sensitivity analysis. Global sensitivity indices include regression coefficients (Chatterjee and Hadi (2009)), McDiarmid diameters (McDiarmid (1989)), and Sobol' indices (Sobol' (1993)), etc. Based on the functional ANOVA decomposition, the Sobol' index measures the total effects of an input on the response. It is applicable to general nonlinear systems, and can be computed by Monte Carlo methods

(Saltelli et al. (2010)). The Sobol' index has been well received in applications (Saltelli et al. (2008)).

A statistical model with some factors (or covariates) and a response can be viewed as a special input-output system. Statisticians often use the term "significance" to define important factors instead of any sensitivity index. A p-value can be interpreted as the type-I error of a test, and it is closely related to certain sensitivity indices in some cases. For example, under linear models, the F-test p-value is commonly used to select important factors (Miller (2002)) like such sensitivity indices as the least squares estimator of the corresponding coefficient. Unlike the coefficient estimator, the p-value takes estimation uncertainty into account.

A permutation test (Pitman (1937)) calculates all possible values of the test statistic under rearrangements of the labels on the observed data points, and exactly controls the type-I error under certain exchangeability conditions (Pesarin and Salmaso (2010)). Here we propose a permutation-based sensitivity index, called the *q-value*, to complement current methods. A q-value can be constructed by permuting the observations corresponding to the investigated input in a goodness-of-fit statistic. Compared to existing sensitivity indices, its main advantage is that, for many cases, it is a p-value for testing whether the effect of a factor on the response exists, and thus has a clear statistical interpretation. Specifically, it is a p-value of a permutation test when the observations corresponding to the investigated input are random and their distributions possess a certain exchangeability, but it can be used more broadly. For deterministic-input linear models, we prove that the q-value based on the residuals can asymptotically control the type-I error under some regularity conditions. Theoretical and numerical analysis is also presented to show the effectiveness of the q-value for completely deterministic systems that do not contain any random term. Some simulation results show that the q-value has similar performance as the Sobol' index.

Other advantages of the proposed method include simplicity in implementation, and flexibility. The q-value is obtained by repeatedly computing the goodness-of-fit statistic with permutated data; when it is not feasible to compute it based on all permutations, Monte Carlo approximation performs well even with a moderate number of permutations. For the specification of the goodness-of-fit statistic, we focus on computer experiments (Santner, Williams and Notz (2003)), and provide two classes of goodness-of-fit statistics to construct q-values. To show its flexibility, we give extensions to partial sensitivity analysis and to sensitivity analysis for grouped inputs.

This article is organized as follows. Section 2 introduces the proposed q-value and Section 3 discusses its theoretical properties. Section 4 presents two extensions of the q-value. Section 5 discusses the construction of the q-value for computer experiments. Section 6 reports on simulation results and Section 7 presents an application. We end the article with some discussion in Section 8. `MATLAB` codes and technical details, including all proofs, are given as online supplementary materials.

2. The Sensitivity Index from Permutations

Consider an input-output system with inputs $x_1, \ldots, x_d \in \mathbb{R}$ and output $y \in \mathbb{R}$,

$$y = f_{\boldsymbol{\theta}}(x_1, \ldots, x_d;\ \boldsymbol{\varepsilon}), \tag{2.1}$$

where $\boldsymbol{\theta}$ is an unknown unknown parameter and $\boldsymbol{\varepsilon}$ is an unobservable random error. If we have n input values $\mathbf{x}_1, \ldots, \mathbf{x}_n$, $\mathbf{x}_i = (x_{i1}, \ldots, x_{id})'$ for $i = 1, \ldots, n$, and the corresponding output values $\mathbf{y} = (y_1, \ldots, y_n)'$, we write the input matrix (or design matrix) as

$$\mathbf{X} = (x_{ij})_{i=1,\ldots,n,\ j=1,\ldots,d} = (\mathbf{z}_1, \ldots, \mathbf{z}_d), \tag{2.2}$$

where $\mathbf{z}_j = (x_{1j}, \ldots, x_{nj})'$ for $j = 1, \ldots, d$. Our purpose is to quantify the influence of the jth input x_j on the output y in (2.1) based on these data. Without loss of generality, we only consider the first input x_1 in the following.

Let $\mathbb{S}_n$ denote the set of all permutations of $(1, \ldots, n)$. Suppose that

$$T(\mathbf{X}, \mathbf{y}) = T(\mathbf{z}_1, \mathbf{Z}_2, \mathbf{y}) \tag{2.3}$$

is a goodness-of-fit statistic whose small values correspond to good fit of the model, where $\mathbf{Z}_2 = (\mathbf{z}_2, \ldots, \mathbf{z}_d)$. If x_1 has little influence on y, then the value of $T(\mathbf{z}_1, \mathbf{Z}_2, \mathbf{y})$ may be close to others in $\{T(\mathbf{z}_{1,k}^{\text{perm}}, \mathbf{Z}_2, \mathbf{y})\}_{k=1,\ldots,n!}$, where $\{\mathbf{z}_{1,1}^{\text{perm}}, \ldots, \mathbf{z}_{1,n!}^{\text{perm}}\} = \{(x_{i_1 1}, \ldots, x_{i_n 1})' :\ (i_1, \ldots, i_n) \in \mathbb{S}_n\}$. Otherwise, $T(\mathbf{X}, \mathbf{y})$ tends to take small values among $\{T(\mathbf{z}_{1,k}^{\text{perm}}, \mathbf{Z}_2, \mathbf{y})\}_{k=1,\ldots,n!}$. Accordingly, define the *q-value*

$$q = \frac{1}{n!} \sum_{k=1}^{n!} I\Big(T(\mathbf{z}_1, \mathbf{Z}_2, \mathbf{y}) \geqslant T(\mathbf{z}_{1,k}^{\text{perm}}, \mathbf{Z}_2, \mathbf{y})\Big) \tag{2.4}$$

as the sensitivity index of x_1, where I is the indicator function.

The q-value is related to testing whether or not x_1 has effect on the response y. The null hypothesis is formulated as

$$H_0 :\ x_1 \text{ has no effect on } y. \tag{2.5}$$

In Section 3 we show some cases in which the q-value can serve as a p-value for testing (2.5); this is the main advantage of the q-value over existing sensitivity indices. In these cases, we say that x_1 is significant if the q-value is less than a given significance level $\alpha \in (0,1)$.

Usually it is not feasible to compute (2.4) based on all the $n!$ permutations. We can then use the Monte Carlo method to approximate it by generating M random permutations, which gives

$$q_{\text{MC}} = \frac{1}{M}\sum_{k=1}^{M} I\Big(T(\mathbf{z}_1, \mathbf{Z}_2, \mathbf{y}) \geqslant T(\mathbf{z}_{1,k}^{\text{r,perm}}, \mathbf{Z}_2, \mathbf{y})\Big), \tag{2.6}$$

where $\{\mathbf{z}_{1,1}^{\text{r,perm}}, \ldots, \mathbf{z}_{1,M}^{\text{r,perm}}\}$ are randomly drawn from $\{\mathbf{z}_{1,1}^{\text{perm}}, \ldots, \mathbf{z}_{1,n!}^{\text{perm}}\}$. Simple derivations give $\mathrm{E}(q_{\text{MC}}) = q$ and $\mathrm{Var}(q_{\text{MC}}) = q(1-q)/M$. As is shown in our simulations in Section 7, such an approximation performs satisfactorily even with a moderate M.

3. Theoretical Properties of the q-value

3.1. Frequentist properties with a random input

Consider the case that the first column $\mathbf{z}_1$ of the input matrix $\mathbf{X}$ in (2.2) is random. For making the notation clear, we use capital letters $(Z_1, \ldots, Z_n)'$ to denote this column. Here we show that the q-value is the p-value of a permutation test (Pitman (1937)) for testing (2.5), and thus controls the type-I error exactly. Two assumptions are needed.

Assumption 1. *Under H_0 at* (2.5)*, the goodness-of-fit statistic T at* (2.3) *has the form $T = T(Z_1, \ldots, Z_n;\ \mathbf{E}')$, where $\mathbf{E} = (E_1, \ldots, E_L)'$ denotes other random terms in T.*

Assumption 2. *Under H_0 at* (2.5)*, for any permutation $(i_1, \ldots, i_n) \in \mathbb{S}_n$, the joint distribution of $(Z_{i_1}, \ldots, Z_{i_n},\ \mathbf{E}')$ is identical to that of $(Z_1, \ldots, Z_n,\ \mathbf{E}')$.*

Theorem 1. *Suppose that H_0 holds. Under Assumptions* 1 *and* 2*, for any $\alpha \in (0,1)$, $\Pr(q < \alpha) \leqslant \alpha$.*

Remark 1. Suppose that $(Z_1, \ldots, Z_n)'$ is independent of $\mathbf{E}$. If the joint distribution of $(Z_1, \ldots, Z_n)'$ is exchangeable, then Assumption 2 holds. In addition, if $Z_1, \ldots, Z_n$ are independently and identically distributed (i.i.d.) according to a nondegenerate distribution, then their joint distribution is exchangeable.

Remark 2. Suppose that the randomness of T comes from the random design $\mathbf{X}$ and random errors $\boldsymbol{\epsilon}$, where $\mathbf{X}$ and $\boldsymbol{\epsilon}$ are independent. By Remark 1, simple random sampling and Latin hypercube sampling (McKay, Beckman and Conover (1979)) satisfy Assumption 2.

3.2. Asymptotic frequentist properties with deterministic inputs under linear models

In permutation tests, the data to be permuted are regarded as random, and an exchangeability assumption like Assumption 2 is needed to guarantee the exact control of the type-I error (Romano (1989); Welch (1990); Good (2005)). For the q-value, the input values to be permuted may be deterministic and for such cases, it does not possess exact frequentist properties. However, this subsection provides an interesting result that for linear models with deterministic inputs, the q-value possesses *asymptotic* frequentist properties under some regularity conditions. Hence, it can be viewed as a p-value from a *generalized* permutation test.

Consider the linear model

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_d x_d + \varepsilon, \tag{3.1}$$

where $\beta_0, \beta_1, \ldots, \beta_d \in \mathbb{R}$ are unknown parameters and ε is zero-mean random error. Using the data at n points, we have the matrix form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{3.2}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_d)'$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)'$, and the input matrix $\mathbf{X} = (x_{ij})_{i=1,\ldots,n,j=1,\ldots,d} = (\mathbf{z}_1, \ldots, \mathbf{z}_d)$ is standardized so that for $j = 1, \ldots, d$,

$$\sum_{i=1}^{n} x_{ij} = 0, \quad \sum_{i=1}^{n} x_{ij}^2 = n. \tag{3.3}$$

Here we discuss the asymptotics for fixed d and $\mathbf{X}$ of full column rank.

In this case the null hypothesis (2.5) reduces to

$$H_0: \ \beta_1 = 0. \tag{3.4}$$

The goodness-of-fit statistic for constructing the q-value in (2.4) is

$$T = \|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{H_0}\|^2 - \|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2, \tag{3.5}$$

where $\hat{\boldsymbol{\beta}}_{H_0}$ and $\hat{\boldsymbol{\beta}}$ are least squares estimators under H_0 and in $\mathbb{R}^d$, respectively, and $\|\cdot\|$ denotes the Euclidean norm. Let $\mathbf{Z}_2 = (\mathbf{z}_2, \ldots, \mathbf{z}_d)$ and $\mathbf{P}_2 = \mathbf{Z}_2(\mathbf{Z}_2'\mathbf{Z}_2)^{-1}\mathbf{Z}_2'$.

Lemma 1. *Under H_0, T in (3.5) can be written as $T = \big[\mathbf{z}_1'(\mathbf{I}_n - \mathbf{P}_2)\mathbf{y})\big]^2/(n - \mathbf{z}_1'\mathbf{P}_2\mathbf{z}_1) = \big[\mathbf{z}_1'(\mathbf{I}_n - \mathbf{P}_2)\boldsymbol{\varepsilon})\big]^2/(n - \mathbf{z}_1'\mathbf{P}_2\mathbf{z}_1)$.*

Remark 3. In some cases, $\mathbf{X}'\mathbf{X}$ may become singular when permutating $\mathbf{z}_1$. By Lemma 1, we permute $\mathbf{z}_1$ in $\left[\mathbf{z}_1'(\mathbf{I}_n-\mathbf{P}_2)\mathbf{y})\right]^2/(n-\mathbf{z}_1'\mathbf{P}_2\mathbf{z}_1)$ to compute the q-value in (2.4), and this avoids matrix inversion.

To discuss the asymptotic properties of the q-value when the input matrix $\mathbf{X}$ is deterministic, we need some conditions.

Assumption 3. *The random errors* $\varepsilon_1,\ldots,\varepsilon_n$ *are i.i.d. nondegenerate random variables with* $\mathrm{E}\,\varepsilon_1=0$ *and* $\mathrm{E}|\varepsilon_1|^2<\infty$.

Assumption 4. *The random errors* $\varepsilon_1,\ldots,\varepsilon_n$ *are i.i.d.* $\sim N(0,\sigma^2)$ *with* $\sigma^2>0$.

Assumption 5. *As* $n\to\infty$, $\max_{1\leqslant i\leqslant n} x_{i1}^2/n=o(1)$.

Assumption 6. *For any fixed* $r=3,4,\ldots$, $\sum_{i=1}^n x_{i1}^r/n=O(1)$.

Let the random vector $\boldsymbol{\eta}_n$, independent of $\boldsymbol{\varepsilon}$, be uniformly distributed on $\{(x_{i_1 1},\ldots,x_{i_n 1})' :\ (i_1,\ldots,i_n)\in\mathbb{S}_n\}$. The q-value based on T in (3.5) for testing H_0 in (3.4) can then be written as

$$\begin{aligned} q &= \Pr\big(T(\mathbf{z}_1,\mathbf{Z}_2,\mathbf{y})\geqslant T(\boldsymbol{\eta}_n,\mathbf{Z}_2,\mathbf{y})\mid \mathbf{y}\big)\\ &= \Pr\left(\frac{\left[\mathbf{z}_1'(\mathbf{I}_n-\mathbf{P}_2)\boldsymbol{\varepsilon}\right]^2}{(n-\mathbf{z}_1'\mathbf{P}_2\mathbf{z}_1)}\geqslant\frac{(\mathbf{e}_n'\boldsymbol{\eta}_n)^2}{(n-\boldsymbol{\eta}_n'\mathbf{P}_2\boldsymbol{\eta}_n)}\mid\boldsymbol{\varepsilon}\right), \end{aligned} \tag{3.6}$$

where

$$\mathbf{e}_n=(e_{n1},\ldots,e_{nn})'=(\mathbf{I}_n-\mathbf{P}_2)\boldsymbol{\varepsilon}. \tag{3.7}$$

Assumption 7. *As* $n\to\infty$, $n/\lambda_{\min}(\mathbf{Z}_2'\mathbf{Z}_2)=O(1)$, *where* $\lambda_{\min}(\cdot)$ *denotes the smallest eigenvalue of a matrix.*

Theorem 2. *Suppose Assumption* 7 *holds. If either* (I) *Assumptions* 3 *and* 6 *hold, or* (II) *Assumptions* 4 *and* 5 *hold, then the distribution of* $(\mathbf{e}_n'\boldsymbol{\eta}_n)^2/\big(\sigma^2(n-\boldsymbol{\eta}_n'\mathbf{P}_2\boldsymbol{\eta}_n)\big)$, *conditional on* $\boldsymbol{\varepsilon}$, *converges to* χ_1^2 *in distribution as* $n\to\infty$ *a.s.*

Assumption 8. *As* $n\to\infty$, $\max_{1\leqslant i\leqslant n} v_{ni}^2/(n-\mathbf{z}_1'\mathbf{P}_2\mathbf{z}_1)=o(1)$, *where* $\mathbf{v}_n=(v_{n1},\ldots,v_{nn})'=(\mathbf{I}_n-\mathbf{P}_2)\mathbf{z}_1$.

Remark 4. If $\mathbf{X}'\mathbf{X}/n\to$ a positive definite matrix and $\max_{1\leqslant i\leqslant n}\mathbf{x}_i'\mathbf{x}_i/n=o(1)$, then Assumptions 5-8 hold.

Theorem 3. *Under Assumptions* 3 *and* 8, $T/\sigma^2\to\chi_1^2$ *in distribution as* $n\to\infty$.

Theorem 4. *If the conditions of Theorems* 2 *and* 3 *hold, then for any* $\alpha\in(0,1)$, $\Pr(q<\alpha)\to\alpha$ *as* $n\to\infty$.

Permutation tests that exactly control the type-I error by permuting random data are used as exact frequentist methods. This feature is quite useful for small-sample problems, but limits their application. By Theorem 4, the q-value corresponding to (3.4) can be viewed as the p-value from a permutation test that permutes non-random data. For linear regression models, this permutation method is asymptotically valid, like the bootstrap test (Mammen (1993)).

3.3. Completely deterministic systems

Consider the case where the system (2.1) is completely deterministic

$$y = f(x_1, \ldots, x_d) \tag{3.8}$$

with f an unknown deterministic function and all inputs $x_1, \ldots, x_n$ deterministic. We present two examples to show that the q-value can quantify the influence of an input on the output.

Example 1. Let $d = 1$ and $y = f(t)$ for $t \in [0, 1]$, where f is a non-decreasing function. The input values are $\mathbf{t} = (t_1, \ldots, t_n)'$ with $0 \leqslant t_1 < \cdots < t_n \leqslant 1$ and the corresponding outputs are $\mathbf{y} = (y_1, \ldots, y_n)'$, with the goodness-of-fit statistic $T(\mathbf{t}, \mathbf{y}) = -\mathbf{t}'\mathbf{y}$. Under H_0 in (2.5), f is a constant c, and $T = -c\sum_{i=1}^n t_i$, which implies $q = 1$. If H_0 is not true, then the values in $\mathbf{y}$ can be written as $y_1 = \cdots = y_{n_1} = y_{(1)} < y_{n_1+1} = \cdots = y_{n_1+n_2} = y_{(2)} < \cdots < y_{n_1+\cdots+n_{s-1}+1} = \cdots = y_n = y_{(s)}$ with $2 \leqslant s \leqslant n$. For $(i_1, \ldots, i_n) \in \mathbb{S}_n$, if there exists $j = 1, \ldots, n$ such that $i_j \notin \{n_{j-1}+1, \ldots, n_{j-1}+n_j\}$, where $n_0 = 0$, then $T(\mathbf{t}, \mathbf{y}) = -(t_1y_1+\cdots+t_ny_n) < -(t_{i_1}y_1 + \cdots + t_{i_n}y_n) = T((t_{i_1}, \ldots, t_{i_n})', \mathbf{y})$. Therefore,

$$q = \frac{n_1! \cdots n_s!}{n!}.$$

The q-value relies on the number of unequal responses, with minimum when $s = n$ ($n_1 = \cdots = n_s = 1$), a reasonable sensitivity for a non-decreasing function.

Example 2. Let $d = 1$ and $y = f(t)$ for $t \in [0, 1]$, where f is a continuous piecewise-linear function with knots $0 = a_0 < a_1 < \cdots < a_s < a_{s+1} = 1$. Let β_j denote the slope of f on $[a_j, a_{j+1}]$ for $j = 0, \ldots, s$. We take the input values $\mathbf{t} = (t_1, \ldots, t_n)'$, where $0 = t_1 < \cdots < t_n = 1$, and the corresponding outputs $\mathbf{y} = (y_1, \ldots, y_n)'$. Assume that $\{t_1, \ldots, t_n\} \supset \{a_1, \ldots, a_s\}$ for sufficiently large n. An estimator of f, denoted by $\hat{f}$, is piecewise-linear connecting the points $(t_i, y_i)'$ and $(t_{i+1}, y_{i+1})'$ for $i = 1, \ldots, n-1$. A goodness-of-fit statistic is the mean squares error

$$T(\mathbf{t}, \mathbf{y}) = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{f}(t_i^*) - f(t_i^*)\right)^2,$$

where $0 \leqslant t_1^* < \cdots < t_N^* \leqslant 1$ constitute the test set. We assume that

$$|\{t_1^*, \ldots, t_N^*\} \cap (t_i, t_{i+1})| \geqslant 2 \quad \text{for all } i = 1, \ldots, n-1, \tag{3.9}$$

where $|\cdot|$ denotes the cardinality of a set. For $i = 1, \ldots, n$, let $\mathcal{S}_i = \{t_j : f(t_j) = f(t_i),\ j = 1, \ldots, n\}$ and $m = \max_{i=1,\ldots,n} |\mathcal{S}_i|$.

If H_0 is true, then $T(\mathbf{t}_k^{\text{perm}}, \mathbf{y}) = 0$ for $k = 1, \ldots, n!$, where $\mathbf{t}_k^{\text{perm}}$ is the kth permutation of $t_1, \ldots, t_n$. For this case, $q = 1$. Otherwise, there exists a $\beta_{s_0} \neq 0$ and $T(\mathbf{t}, \mathbf{y}) = 0$ as well. By (3.9), let $t_{j_0}^*, t_{j_0+1}^* \in (t_{i_0}, t_{i_0+1}) \subset [a_{s_0}, a_{s_0+1}]$. For a permutation $\mathbf{t}_k^{\text{perm}}$ which replaces t_{i_0} with an element in $\{t_1, \ldots, t_n\} \setminus \mathcal{S}_{i_0}$, and/or replaces t_{i_0+1} with an element in $\{t_1, \ldots, t_n\} \setminus \mathcal{S}_{i_0+1}$, either $\hat{f}(t_{j_0}^*) - f(t_{j_0}^*)$ or $\hat{f}(t_{j_0+1}^*) - f(t_{j_0+1}^*)$ is not zero. Therefore, $T(\mathbf{t}_k^{\text{perm}}, \mathbf{y}) > 0 = T(\mathbf{t}, \mathbf{y})$, which implies $q \leqslant 1 - \big[(n-m)(n-m-1)(n-2)!\big]/n! = 1 - [(n-m)(n-m-1)]/[n(n-1)]$. If $m = o(n)$, $q \to 0$ as $n \to \infty$.

Remark 5. For completely deterministic systems, if we view $\mathbf{z}_1$ as a random vector following the distribution of $\boldsymbol{\eta}_n$ in (3.6), then the q-value in (2.4) can be written as

$$q = \Pr\big(T(\mathbf{z}_1, \mathbf{Z}_2, \mathbf{y}) \geqslant T(\boldsymbol{\eta}_n, \mathbf{Z}_2, \mathbf{y})\big), \tag{3.10}$$

where $\boldsymbol{\eta}_n$ is independent of $\mathbf{z}_1$ and $\mathbf{y}$ may depend on $(\mathbf{z}_1, \mathbf{Z}_2)$. Under H_0 in (2.5), $\mathbf{y}$ does not depend on $\mathbf{z}_1$, and thus q in (3.10) is $U[0,1]$. Such randomization provides a way to assign frequentist properties of the q-value in completely deterministic systems.

4. Extensions of the q-value

4.1. The q-value for partial sensitivity analysis

Sometimes we are interested in quantifying the variation of the output when an input changes in a subset of its domain. We call such analysis *partial sensitivity analysis*. The definition of q-value can be straightforwardly extended to this case. For example, suppose $x_1 \in [0,1]$ in (2.1). Our purpose is to quantify the influence of x_1 at its low level $[0, 1/2]$ on y. Let $\mathbf{X}$ be the input matrix in (2.2), $\mathbf{y}$ be the vector of responses, and T be the goodness-of-fit statistic in (2.3). Take $x_{11}, \ldots, x_{n_1 1} \in [0, 1/2]$ and $x_{n_1+1\ 1}, \ldots, x_{n1} \in (1/2, 1]$. The q-value of x_1 at the low level is defined as

$$q = \frac{1}{n_1!} \sum_{k=1}^{n_1!} I\Big(T(\mathbf{z}_1, \mathbf{Z}_2, \mathbf{y}) \geqslant T(\mathbf{z}_{1,k}^{\text{l,perm}}, \mathbf{Z}_2, \mathbf{y})\Big), \tag{4.1}$$

where $\{\mathbf{z}_{1,1}^{\text{l,perm}}, \ldots, \mathbf{z}_{1,n_1!}^{\text{l,perm}}\} = \{(x_{i_1 1}, \ldots, x_{i_{n_1} 1},\ x_{n_1+1\ 1}, \ldots, x_{n1})' : (i_1, \ldots, i_{n_1}) \in \mathbb{Z}_{n_1}\}$. Similarly we can define the q-value of x_1 at the high level $[1/2, 1]$.

4.2. The q-value of grouped inputs

It can be useful to quantify the influence of grouped inputs $(x_1, \ldots, x_g)$ on the output (Saltelli et al. (2008)). Write $T(\mathbf{X}, \mathbf{y}) = T(\mathbf{z}_{(1)}, \mathbf{Z}_{(2)}, \mathbf{y})$, where $\mathbf{z}_{(1)} = (\mathbf{z}'_1, \ldots, \mathbf{z}'_g)' = (z_1, \ldots, z_{gn})'$ and $\mathbf{Z}_{(2)} = (\mathbf{z}_{g+1}, \ldots, \mathbf{z}_d)$. The q-value of $(x_1, \ldots, x_g)$ can be defined as

$$q = \frac{1}{(gn)!} \sum_{k=1}^{(gn)!} I\Big(T(\mathbf{z}_{(1)}, \mathbf{Z}_{(2)}, \mathbf{y}) \geqslant T(\mathbf{z}^{\text{perm}}_{(1),k}, \mathbf{Z}_{(2)}, \mathbf{y})\Big),$$

where $\{\mathbf{z}^{\text{perm}}_{(1),1}, \ldots, \mathbf{z}^{\text{perm}}_{(1),(gn)!}\} = \{(z_{i_1}, \ldots, z_{i_{gn}})' : \ (i_1, \ldots, i_{gn}) \in \mathbb{Z}_{gn}\}$. Theorem 1 in Section 4.1 can be easily extended to this q-value under proper conditions.

5. Construction of the q-value for Computer Experiments

Computer experiments are used to study a computer simulation, which is constructed to approximate a complex system. A canonical form of computer models is

$$y = f(x_1, \ldots, x_d), \tag{5.1}$$

which is similar to (3.8), but the inputs can be selected randomly or deterministically (Santner, Williams and Notz (2003)). It can be seen that f in (5.1) does not contain any random error. This is the essential difference from physical experiments. There are a number of sensitivity analysis methods for computer experiments in the literature, including the design-based one-factor-at-a-time method (Morris (1991)), Bayesian methods (Oakley and O'Hagan (2004); Linkletter et al. (2012)), the two-stage procedure (Moon, Dean and Santner (2012)), and the dynamic tree model-based method (Gramacy, Taddy and Wild (2013)), among others. It is worth noting that the Sobol' index still plays an important role in sensitivity analysis for computer experiments (Saltelli (2002)), sometimes as a basic index in sophisticated methods (Moon, Dean and Santner (2012); Gramacy, Taddy and Wild (2013)).

Under (5.1), (2.5) is equivalent to

$$H_0 : \ y = \tilde{f}(x_2, \ldots, x_d) \text{ for some } (d-1) - \text{dimensional function } \tilde{f}. \tag{5.2}$$

It is known that statistical testing can detect significant main effects and/or interactions of the factors on the response in physical experiments. In parallel, the q-value provides a way to define significant input factors in computer experiments under appropriate conditions or randomization.

5.1. The q-value via cross-validation

Suppose that we have an emulator $\hat{f}$ of f in (5.1) based on the input values $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ and the corresponding outputs $\mathbf{y} = (y_1, \ldots . y_n)'$. A goodness-of-fit statistic is the cross-validation error

$$T = \frac{1}{n} \sum_{i=1}^{n} \left(\hat{f}_{-i}(\mathbf{x}_i) - y_i\right)^2, \tag{5.3}$$

where $\hat{f}_{-i}$ is constructed based on $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \setminus \{\mathbf{x}_i\}$ and $\{y_1, \ldots . y_n\} \setminus \{y_i\}$. If the response y depends on x_1, then when randomly permutating the first column of the input matrix $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)'$, T in (5.3) tends to be larger, and thus we find a small q-value. There are many choices of the emulator $\hat{f}$ for constructing T; see Chen et al. (2006).

5.2. The q-value based on Kriging models

The Kriging model (Matheron (1963)) is widely used to analyze computer experiments (Sacks et al. (1989)). It models the output of a computer experiment as a realization of

$$y(\mathbf{x}) = \mathbf{g}(\mathbf{x})'\boldsymbol{\beta} + Z(\mathbf{x}), \tag{5.4}$$

where $\mathbf{g}(\mathbf{x}) = (g_1(\mathbf{x}), \ldots, g_q(\mathbf{x}))'$ is a pre-specified set of functions, $\boldsymbol{\beta}$ is a vector of unknown regression coefficients, and $Z(\mathbf{x})$ is a stationary Gaussian process $\mathrm{GP}(0, \sigma^2, \boldsymbol{\theta})$ with mean zero, variance σ^2, and correlation parameters $\boldsymbol{\theta}$. The covariance between $Z(\mathbf{x}_1)$ and $Z(\mathbf{x}_2)$ is represented by

$$\mathrm{Cov}[Z(\mathbf{x}_1), Z(\mathbf{x}_2)] = \sigma^2 R(\mathbf{x}_1 - \mathbf{x}_2 \,|\, \boldsymbol{\theta}), \tag{5.5}$$

where $R(\cdot \,|\, \boldsymbol{\theta})$ is the correlation function depending on a parameter vector $\boldsymbol{\theta}$. A popular choice is the squared exponential correlation function

$$R(\mathbf{x}_1 - \mathbf{x}_2 \,|\, \theta) = \exp(-\theta \|\mathbf{x}_1 - \mathbf{x}_2\|^2), \tag{5.6}$$

where $\theta > 0$ is the correlation parameter.

The parameters in (5.4) can be estimated by maximum likelihood. Take the input values $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, where $\mathbf{x}_i = (x_{i1}, \ldots, x_{id})'$ for $i = 1, \ldots, n$, and the corresponding response values $\mathbf{y} = (y_1, \ldots, y_n)'$. Let $\mathbf{X} = (x_{ij})_{i=1,\ldots,n,\ j=1,\ldots,d}$ denote the input matrix as in (2.2). The negative log likelihood, up to an additive constant, is

$$n \log(\sigma^2) + \log(\det(\mathbf{R})) + (\mathbf{y} - \mathbf{G}\boldsymbol{\beta})'\mathbf{R}^{-1} \frac{(\mathbf{y} - \mathbf{G}\boldsymbol{\beta})}{\sigma^2}, \tag{5.7}$$

where $\mathbf{R}$ is the $n \times n$ correlation matrix whose (i,j)th entry is $R(\mathbf{x}_i - \mathbf{x}_j \,|\, \boldsymbol{\theta})$ defined in (5.5), "det" denotes matrix determinant, and $\mathbf{G} = (\mathbf{g}(\mathbf{x}_1), \ldots, \mathbf{g}(\mathbf{x}_n))'$.

Given $\boldsymbol{\theta}$, the maximum likelihood estimators (MLEs) of $\boldsymbol{\beta}$ and σ^2 are

$$\begin{cases} \hat{\boldsymbol{\beta}} = (\mathbf{G}'\mathbf{R}^{-1}\mathbf{G})^{-1}\mathbf{G}'\mathbf{R}^{-1}\mathbf{y}, \\ \hat{\sigma}^2 = (\mathbf{y} - \mathbf{G}\hat{\boldsymbol{\beta}})'\mathbf{R}^{-1}\dfrac{(\mathbf{y} - \mathbf{G}\hat{\boldsymbol{\beta}})}{n}. \end{cases} \tag{5.8}$$

Plugging these estimators into (5.7), we obtain the main part of the minimum of the negative log likelihood

$$S(\mathbf{X}, \mathbf{y} \,|\, \boldsymbol{\theta}) = n \log(\hat{\sigma}^2) + \log\left(\det(\mathbf{R})\right). \tag{5.9}$$

Based on $(\mathbf{X}, \mathbf{y})$, the MLE $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is obtained as

$$\hat{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{y}) = \operatorname{argmin}_{\boldsymbol{\theta}} S(\mathbf{X}, \mathbf{y} \,|\, \boldsymbol{\theta}). \tag{5.10}$$

The estimators of $\boldsymbol{\beta}$ and σ^2 can be obtained by plugging $\hat{\boldsymbol{\theta}}$ into (5.8). The goodness-of-fit statistic is

$$T = S(\mathbf{X}, \mathbf{y} \,|\, \hat{\boldsymbol{\theta}}). \tag{5.11}$$

Based on T, we can compute the q-value as the sensitivity index for x_1 by (2.4).

The q-value from T in (5.11) is related to the likelihood ratio test of (5.2), in which the test statistic is $L(\mathbf{X}, \mathbf{y}) = T(\mathbf{X}, \mathbf{y}) - T(\mathbf{Z}_2, \mathbf{y})$. Note that the second term in L does not depend on $\mathbf{z}_1$. The above q-value can also be computed by permuting $\mathbf{z}_1$ in L. In addition, the goodness-of-fit statistic T in (5.11) corresponds to the maximum entropy of the responses under the Kriging model (5.4) (Shewry and Wynn (1987)).

The Kriging model also yields an emulator $\hat{f}$ of f (Santner, Williams and Notz (2003)): for an untried point $\mathbf{x}_0$,

$$\hat{f}(\mathbf{x}_0) = \mathbf{g}(\mathbf{x}_0)'\hat{\boldsymbol{\beta}} + \hat{\mathbf{r}}'\hat{\mathbf{R}}^{-1}\big(\mathbf{y} - \mathbf{G}\hat{\boldsymbol{\beta}}\big), \tag{5.12}$$

where $\hat{\mathbf{r}} = \big(R(\mathbf{x}_0 - \mathbf{x}_1 \,|\, \hat{\boldsymbol{\theta}}), \ldots, R(\mathbf{x}_0 - \mathbf{x}_n \,|\, \hat{\boldsymbol{\theta}})\big)'$ and $\hat{\mathbf{R}}$ is the $n \times n$ correlation matrix whose (i, j)th entry is $R(\mathbf{x}_i - \mathbf{x}_j \,|\, \hat{\boldsymbol{\theta}})$. This emulator can be used to construct the cross-validation error T in (5.3). Some simulation results show that the q-value based on such cross-validation error has lower power than the likelihood-based method (see Section 6.2), where the power is the probability of the correct rejection when the null hypothesis does not hold. In general, we recommend the likelihood-based q-value for computer experiments. The cross-validation method based on other emulators, however, has potential applications to the cases where the Kriging emulator is hard to build, such as the large-scale data case (Joseph and Kang (2011)).

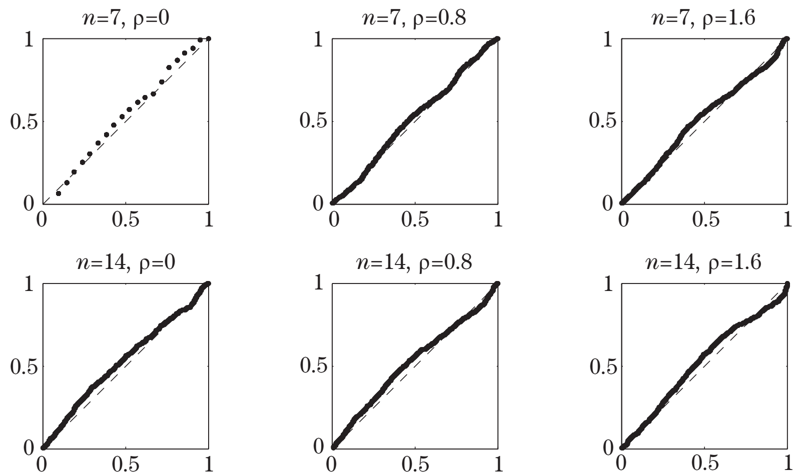

Figure 1. QQ-plots of the q-values in Section 6.1.

6. Numerical Illustrations

6.1. Simulations for linear models

Consider the linear model in (3.1), where $d = 3$, $\beta_0 = 1$, $\beta_1 = 0$, $\beta_2 = 1$, $\beta_3 = -2$, and the random error follows the t distribution with one degree of freedom. In the simulations, the input values are all deterministic. We considered two cases for n: $n = 7$ and $n = 14$. The corresponding input matrices with a correlation parameter ρ can be found in the online supplementary materials.

We computed the q-value in (3.6) after standardizing the data. For $n = 14$, since 14! is too large, we used the Monte Carlo method to approximate it with $M = 7! = 5{,}040$ in (2.6). The QQ-plots of the q-values over 1,000 repetitions for different combinations of n and ρ are shown in Figure 1. For $(n, \rho) = (7, 0)$, the distribution of the q-value looks discrete, but it is not far away from $U[0, 1]$. For the other cases, the distributions of the q-values are close to $U[0, 1]$. As suggested by a referee, we also conducted the simulation for larger n and d, and similar phenomena occur. The simulation results are consistent with our theoretical findings.

6.2. A two-dimensional function

We used

$$f(x_1, x_2) = \left[1 - \beta \exp\left(-\frac{1}{2x_1}\right)\right] \frac{2300x_2^3 + 1900x_2^2 + 2092x_2 + 60}{100x_2^3 + 500x_2^2 + 4x_2 + 20}$$

as a computer model, where $(x_1, x_2)' \in [0, 1]^2$ and the parameter β controls the

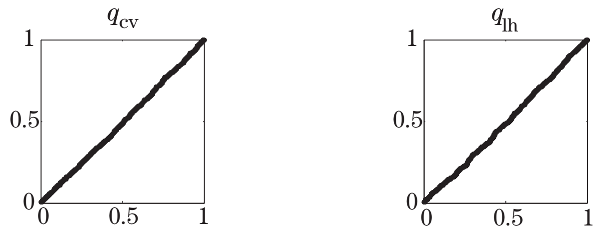

Figure 2. QQ-plots of the q-values in Section 6.2.

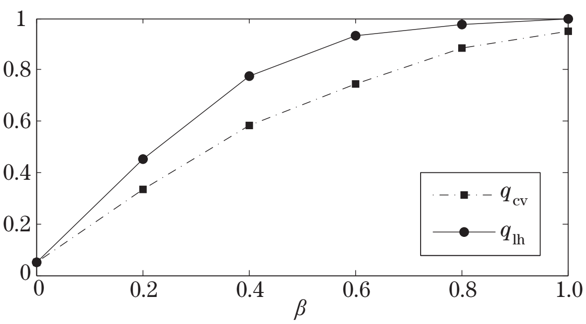

Figure 3. Powers for different β's in Section 6.2 ($\alpha = 0.05$).

influence of x_1 on y. This function is a variant of a function in Currin et al. (1991). The run size was fixed as 15. We generated the design matrix $\mathbf{X}$ by simple random sampling. Such designs satisfy Assumption 2.

We computed the q-values of x_1 by two methods. The first was based on the cross-validation error T in (5.3), where $\hat{f}$ was the Kriging emulator (5.12). The goodness-of-fit statistic of the second method was the Kriging likelihood T in (5.11). We employed the Kriging model (5.4) with $\mathbf{g}(\mathbf{x}) = (1, x_1, \ldots, x_d)'$ and the squared exponential correlation function (5.6). The Monte Carlo method was used to compute the q-values with $M = 1{,}000$ in (2.6). Denote the two q-values by $q_{\rm cv}$ and $q_{\rm lh}$. It is clear that H_0 in (2.5) is true when $\beta = 0$, and that Theorem 1 holds by Remark 1. For such a case, the QQ-plots of $q_{\rm cv}$ and $q_{\rm lh}$ over 1,000 repetitions are shown in Figure 2. Their distributions are very close to $U[0, 1]$, consistent with Theorem 1. We also compared their power performance at the significance level $\alpha = 0.05$. Let β vary from 0 to 1. The powers corresponding to $q_{\rm cv}$ and $q_{\rm lh}$ are shown in Figure 3, with the test based on $q_{\rm lh}$ being more powerful.

6.3. A five-dimensional function

This example used the five-dimensional function (Cox, Park and Singer (2001)),

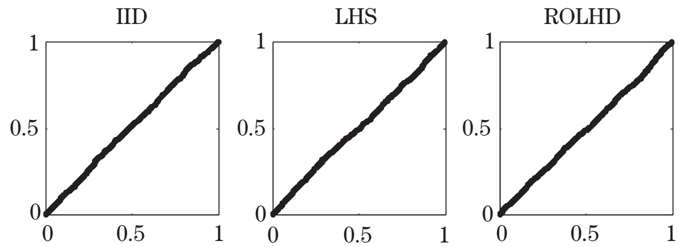

Figure 4. QQ-plots of the q-values of x_5 in Section 6.3 ($n = 10$).

$$f(x_1, \ldots, x_5) = \frac{2}{3}\exp(x_1 + x_2) + x_3 - x_4 \sin(x_3), \tag{6.1}$$

where $(x_1, \ldots, x_5)' \in [0, 1]^5$. We use the same $\mathbf{g}(\mathbf{x})$ and correlation function as in Section 7.2 to compute the Kriging model-based q_{lh} with simple random sampling (IID), the Latin hypercube sampling (LHS) (McKay, Beckman and Conover (1979)), optimal Latin hypercube design with respect to the maximin criterion (OLHD) (Jin, Chen and Sudjianto (2005)), and randomized OLHD by randomizing the first column of OLHD (ROLHD) (see Remark 5). By Remark 2, the first two satisfy Assumption 2. Among the four strategies, only OLHD gave deterministic inputs.

Since f in (6.1) does not rely on x_5, we first computed q_{lh} with the three random designs, IID, LHS, and ROLHD, for $n = 10$, and we show corresponding QQ-plot over 1,000 repetitions in Figure 4. As expected (see Theorem 1 and Remark 5), these q-values follow $U[0, 1]$ closely. Next, for $n = 10$ and 20, we show the box-plots of the q_{lh}'s of $x_1, \ldots, x_5$ over 100 repetitions in Figure 5. For $n = 10$, the three random designs produce small q-values of x_1 and x_2, and moderate q-values of x_3 and x_4. For $n = 20$, they produce small q-values of x_3 and x_4 except for a few cases. Among them, ROLHD produces larger q-values, and the reason may be that the randomization in it gives unusual designs for computer experiments. It can also be seen that the deterministic design OLHD performs reasonably: it yields a clear gap of the q-values between x_1 to x_4 and x_5. We conducted the partial sensitivity analysis in Section 4.1 to quantify the influence of x_1 and x_3 at their low and high levels. The box-plots of the corresponding q-values (see (4.1)) from IID and LHS for $n = 20$ over 100 repetitions are shown in Figure 6. For x_1, the influence at the high level is greater than that at the low level, while for x_3, the two levels do not have a clear difference. These findings are consistent with the form of f in (6.1).

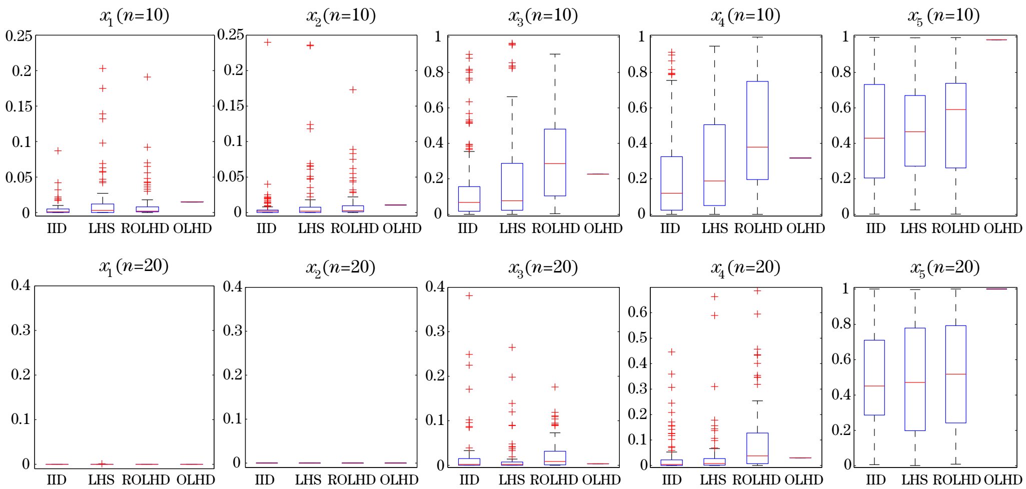

Figure 5. Box-plots of the q-values of $x_1, \ldots, x_5$ in Section 6.3.

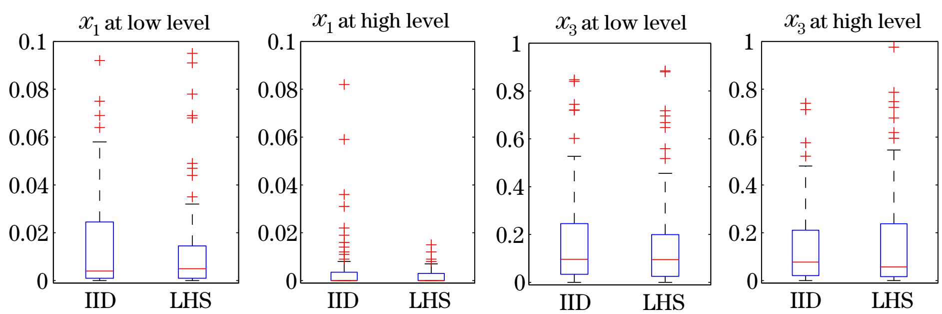

Figure 6. Box-plots of the q-values at low and high levels in Section 6.3 ($n = 20$).

6.4. The Zakharov function

This example used the d-dimensional function $f(x_1, \ldots, x_d) = f_0(x_1, \ldots, x_{d_0})$, where $(x_1, \ldots, x_d)' \in [0, 1]^d$, $d_0 < d$, and f_0 is the Zakharov function (Yang (2010))

$$f_0(x_1, \ldots, x_{d_0}) = \sum_{i=1}^{d_0} x_i^2 + \left(\sum_{i=1}^{d_0} \frac{ix_i}{2} \right)^2 + \left(\sum_{i=1}^{d_0} \frac{ix_i}{2} \right)^4 .$$

The simulations in Sections 6.2 and 6.3 were used to study the frequentist performance of the q-value, while this subsection compares it with the Sobol' index, often viewed as a bench-mark sensitivity index. In the simulation the design matrix $\mathbf{X}$ was generated by LHS. The eight combinations of (n, d, d_0) in Table 1 were considered. The q-value was computed based on the Kriging likelihood

Table 1. Combinations of (n, d, d_0) in Section 6.4.

	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)
(n, d, d_0)	$(15, 3, 2)$	$(15, 5, 2)$	$(30, 6, 3)$	$(30, 10, 4)$	$(50, 10, 5)$	$(50, 15, 5)$	$(100, 10, 5)$	$(100, 20, 5)$

Table 2. Rates of correct selection in Section 6.4.

	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)
Sobol' index	1.00	0.99	0.98	0.65	0.68	0.53	1.00	0.81
q-value	1.00	0.99	0.99	0.65	0.70	0.53	1.00	0.81

as in Section 6.3. The Sobol' index was computed by the Monte Carlo method (Janon et al. (2014)).

First, active inputs corresponding to the largest (smallest) d_0 values of the Sobol' index (q-value) are selected. The rates of correct selection of the two indices over 100 repetitions are reported in Table 2. They have almost the same performance. To evaluate their similarity, we computed the Pearson correlation between the rank of the q-values and the inverse rank of the Sobol' indices of the d inputs for each repetition. The means and standard deviations of the correlations over the 100 repetitions are in Table 3. Here the rank of a vector $\mathbf{v} = (v_1, \ldots, v_d)'$ is the vector $(r_1, \ldots, r_d)'$ satisfying $v_{r_j} = v_{(j)}$ for $j = 1, \ldots, d$, where $v_{(1)} \leqslant \cdots \leqslant v_{(d)}$ is the nondecreasing permutation of $\mathbf{v}$, and the inverse rank of $\mathbf{v}$ is the rank of $-\mathbf{v}$. The correlations for all the cases are positive and higher than 0.65. This indicates that the sensitivity levels of an input judged by the two indices are close to each other. But the q-value possesses clear statistical interpretation while the Sobol' index does not.

7. Applications to a Casting Simulator

This section presents an example of the proposed sensitivity index. Casting is an important manufacturing process for making complex metal products. Generally, the material used in casting is expensive and the casting process is very time-consuming. Computer experiments for simulating a casting process are commonly used to analyze the impact of process parameters on the quality of the casting product. Here we consider a low-pressure die-casting process that produces a certain component of satellites. Since shrinkage defects often occur in these products, engineers hope that the statisticians can help them specify the factors that have significant effects on the degree of shrinkage defect via computer simulations. Following engineers' suggestions, we focused on four input

Table 3. Similarity between the q-value and Sobol' index in Section 6.4.

	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)
mean correlation	0.695	0.930	0.855	0.816	0.657	0.766	0.705	0.742
std correlation	0.245	0.112	0.124	0.263	0.230	0.125	0.105	0.123

Table 4. Input factors of the casting simulation.

Factor	filling velocity (x_1)	initial pressure (x_2)	increase rate of pressure (x_3)	initial temperature (x_4)
Range	[30, 60]	[20, 40]	[0.5, 9.5]	[725, 745]
Unit	mm/s	Kpa	Kpa/s	°C

Table 5. The design and responses of the casting experiment.

Run	x_1	x_2	x_3	x_4	y (%)
01	55.782	22.812	8.7971	740.938	6.76
02	48.282	20.312	1.2029	733.438	6.42
03	37.968	21.562	7.6721	734.688	7.03
04	46.407	38.438	4.0154	727.812	7.25
05	51.093	29.062	4.8596	735.938	7.13
06	57.657	37.188	8.2346	738.438	6.90
07	49.218	28.438	0.6404	743.438	6.20
08	59.532	30.312	7.1096	726.562	7.29
09	44.532	32.188	9.3596	731.562	7.11
10	37.032	35.938	2.3279	737.188	7.22
11	53.907	39.688	2.8904	739.688	6.99
12	32.343	25.312	1.7654	730.938	6.79
13	30.468	34.688	5.9846	729.062	7.29
14	35.157	24.062	3.4529	742.188	7.07
15	42.657	26.562	5.4221	725.312	7.32
16	40.782	33.438	6.5471	744.688	6.75

factors, shown in Table 4. The response is the ratio of the defect volume to the total volume of the product. The computer simulations were conducted with a commercial software "Huazhu" on a 16-run Latin hypercube design. The design and its corresponding responses are presented in Table 5.

We compared our method to several popular sensitivity analysis methods. We first set out the scatterplots of y versus $x_1, \ldots, x_4$ in Figure 7, and the corresponding correlations are in the first row of Table 6. A second method fits the data by linear regression, and uses the least squares estimators of the coefficients as the sensitivity indices of all the inputs. The third method is based on the Sobol'

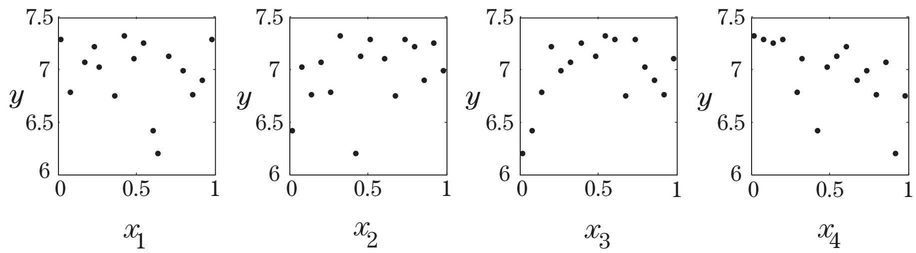

Figure 7. Scatterplots of y versus $x_1, \ldots, x_4$ in Section 7.

Table 6. Sensitivity values of the inputs.

	x_1	x_2	x_3	x_4
correlation	-0.1763	0.3706	0.3998	-0.5868
regression coefficient	-0.2935	0.3922	0.3960	-0.5228
Sobol' index	0.1308	0.2318	0.2342	0.4082
q-value	0.2136	0.0776	0.0696	0.0228

index, and our method is applied. Here we used $\mathbf{g}(\mathbf{x}) = (1, x_1, \ldots, x_4)'$ and the squared exponential correlation function (5.6) in the Kriging model (5.4). The q-values of the four inputs based on T in (5.11) were computed with the Monte Carlo sample size 100,000. The calculations are presented in Table 6. The results from the four methods indicate similar conclusions: x_4 has an important effect on y and x_1 is less important. This is consistent with the simulation results in Section 6.4. Compared with the other methods, the q-value can provide a statistically significant result: only x_4 is significant at significance level $\alpha = 0.05$. We also conducted a partial sensitivity analysis of x_4, and the q-values at its low and high levels are 0.0084 and 0.0392, respectively. The response seemed more sensitive to the low level of x_4.

It is known that shrinkage defects form in the solidification process, and the four inputs are all possible factors to shrinkage defects (Campbell (2011)). Our sensitivity results show that temperature is the most sensitive input for this product. For a follow-up study, temperature should be studied more carefully to find its reasonable value for avoiding shrinkage defect.

8. Discussion

The proposed method is flexible under various settings. For computer experiments, it can be modified to accommodate multiple levels of accuracy (Kennedy and O'Hagan (2000); Qian and Wu (2008)), multiple outputs (Conti and O'Hagan

(2010)), both qualitative and quantitative factors (Qian, Wu and Wu (2008); Han et al. (2009)), branching and nested factors (Hung, Joseph and Melkote (2009)), and sequential analysis (Xiong, Qian and Wu (2013)). For noisy models, other goodness-of-fit statistics under various model assumptions (Hädle et al. (2004)), instead of the likelihood under the Kriging model, can be used to construct the sensitivity indices. It would be valuable to investigate applications of the proposed method to these problems.

Supplementary Materials

The online supplement to this article contains `MATLAB` codes for implementing our methods and proofs of the theoretical results.

Acknowledgment

The authors thank Professor Jeff Wu and two referees for their constructive suggestions which lead to an improvement of this article. This work is supported by the National Natural Science Foundation of China (Grant No. 11471172, 11501550, 11601027, 11671386), Funding from Chinese Ministry of Science and Technology (Grant No. 2016YFF0203801), Science and Technology Plan Project of BMEC (Grant No. KM201410016003), and Research Funding from BUCEA (Grant No. 21082716014, 00331616031).

References

Campbell, J. (2011). *Complete Casting Handbook: Metal Casting Processes, Techniques and Design*. Butterworth-Heinemann.

Chatterjee, S. and Hadi, A. S. (2009). *Sensitivity Analysis in Linear Regression*. John Wiley & Sons.

Chen, V. C. P., Tsui, K-L., Barton, R. R. and Meckesheimer, M. (2006). A review on design, modeling and applications of computer experiments. *IIE Transactions* **38**, 273–291.

Conti, S. and O'Hagan, A. (2010). Bayesian emulation of complex multi-output and dynamic computer models. *J. Statist. Plann. Inference* **140**, 640–651.

Cox, D. D., Park, J. S. and Singer, C. E. (2001). A statistical method for tuning a computer code to a data base. *Comput. Statist. Data Anal.* **37**, 77–92.

Currin, C., Mitchell, T., Morris, M. and Ylvisaker, D. (1991). Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments, *J. Amer. Statist. Assoc.* **86**, 953–963.

Good, P. I. (2005). *Permutation, Parametric and Bootstrap Tests of Hypotheses, 3rd ed.* Springer.

Gramacy, R. B., Taddy, M. and Wild, S. M. (2013). Variable selection and sensitivity analysis using dynamic trees, with an application to computer code performance tuning. *Ann. App. Statist.* **7**, 51–80.

Griewank, A. and Walther, A. (2008). *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation.* SIAM.

Hädle, W., Müller, M., Sperlich, S. and Werwatz, A. (2004). *Nonparametric and Semiparametric Models.* Springer, Berlin.

Han, G., Santner, T. J., Notz, W. I. and Bartel, D. L. (2009). Prediction for computer experiments having quantitative and qualitative input variables. *Technometrics* **51**, 278–288.

Hung, Y., Joseph, V. R. and Melkote, S. N. (2009). Design and analysis of computer experiments with branching and nested factors. *Technometrics* **51**, 354–365.

Janon, A., Klein, T., Lagnoux, A., Nodet, M. and Prieur, C. (2014). Asymptotic normality and efficiency of two Sobol' index estimators. *ESAIM: Probability and Statistics* **18**, 342–364.

Jin, R.C., Chen, W. and Sudjianto, A. (2005). An efficient algorithm for constructing optimal design of computer experiments. *J. Statist. Plann. Inference* **134**, 268–287.

Joseph, V.R. and Kang, L. (2011). Regression-based inverse distance weighting with applications to computer experiments. *Technometrics* **53**, 254–265

Kennedy, M. C. and O'Hagan, A. (2000). Predicting the output from a complex computer code when fast approximations are available. *Biometrika* **87**, 1–13.

Linkletter, C., Bingham, D., Hengartner, N., Higdon, D. and Kenny, Q. Y. (2012). Variable selection for gaussian process models in computer experiments. *Technometrics* **54**, 478–490.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *Ann. Statist.* **21**, 255–285.

Matheron, G. (1963). Principles of geostatistics. *Economic Geology* **58**, 1246–1266.

McDiarmid, C. (1989). On the method of bounded differences. *Surveys in Combinatorics* **141**, 148–188.

McKay, M. D., Beckman, R. J. and Conover, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **21**, 239–245.

Miller, A. (2002). *Subset Selection in Regression, 2nd Edition.* Chapman & Hall/CRC.

Moon, H., Dean, A. M. and Santner, T. J. (2012). Two-stage sensitivity-based group screening in computer experiments. *Technometrics* **54**, 376–387.

Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics* **33**, 161–174.

Oakley, J. and O'Hagan, A. (2004). Probabilistic sensitivity analysis of complex models: a Bayesian approach. *J. Roy. Statist. Soc. B* **66**, 751–769.

Pannell, D. J. (1997). Sensitivity analysis of normative economic models: theoretical framework and practical strategies. *Agric. Econom.* **16**, 139–152.

Pesarin, F. and Salmaso, L. (2010). *Permutation Tests for Complex Data: Theory, Applications and Software.* John Wiley & Sons.

Pitman, E. J. G. (1937). Significance tests which may be applied to samples from any populations. *J. R. Stat. Soc. Suppl.* **4**, 119–130.

Qian, P. Z. G., Wu, H. and Wu, C. F. J. (2008). Gaussian process models for computer experiments with qualitative and quantitative factors. *Technometrics* **50**, 383–396.

Qian, P. Z. G. and Wu, C. F. J. (2008). Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments. *Technometrics* **50**, 192–204.

Romano, J. P. (1989). Bootstrap and randomization tests of some nonparametric hypotheses. *Ann. Statist.* **17**, 141–159.

Sacks, J., Welch, W. J., Mitchell, T. J. and Wynn, H. P. (1989). Design and analysis of computer experiments. *Statist. Sci.* **4**, 409–423.

Saltelli, A. (2002). Making best use of model evaluations to compute sensitivity indices. *Comput. Phys. Comm.* **145**, 280–297.

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M. and Tarantola, S. (2010). Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Comput. Phys. Comm.* **181**, 259–270.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M. and Tarantola, S. (2008). *Global Sensitivity Analysis: The Primer*. John Wiley & Sons.

Santner, T. J., Williams, B. J. and Notz, W. I. (2003). *The Design and Analysis of Computer Experiments*. Springer, New York.

Shewry, M. C. and Wynn, H. P. (1987). Maximum entropy sampling. *J. Appl. Stat.* **14**, 165–170.

Sobol', I. M. (1993). Sensitivity analysis for non-linear mathematical models. *Mathematical Modeling and Computational Experiment* **1**, 407–414.

Sullivan, T. J. (2015). *Introduction to Uncertainty Quantification*. Springer, New York.

Welch, W. J. (1990). Construction of permutation tests. *J. Amer. Statist. Assoc.* **85**, 693–698.

Xiong, S., Qian, P. Z. G. and Wu, C. F. J. (2013). Sequential design and analysis of high-accuracy and low-accuracy computer codes. *Technometrics* **55**, 37–46.

Yang, X.-S. (2010). Test problems in optimization. In *Engineering Optimization: An Introduction with Metaheuristic Applications* (Eds Yang, X.-S.). John Wiley & Sons.

Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190 China.

E-mail: xiong@amss.ac.cn

Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190 China.

E-mail: hexu@amss.ac.cn

School of Statistics, Beijing Normal University, Beijing 100875, China.

E-mail: heyuanzhen@bnu.edu.cn

School of Science, Beijing University of Civil Engineering and Architecture, Beijing 100037, China.

E-mail: muweiyan@bucea.edu.cn

(Received January 2016; accepted August 2016)