UNIVERSITA' COMMERCIALE "LUIGI BOCCONI"

PhD SCHOOL

PhD program in Statistics

Cycle: 31

Disciplinary Field (code): SECS-S/06

Computer Experiments: Sensitivity Analysis and Interactions

Advisor: Emanuele BORGONOVO

PhD Thesis by

Giovanni RABITTI

ID number: 3000647

Academic Year 2019/2020

To my family, my animals and my friends

"L'universo non ha un centro,
ma per abbracciarsi si fa così:
ci si avvicina lentamente
eppure senza motivo apparente,
poi allargando le braccia,
si mostra il disarmo delle ali,
e infine si svanisce,
insieme,
nello spazio di carità
tra te
e l'altro."
Chandra Livia Candiani

Ringraziamenti

Dedico questa Tesi a tutte quelle persone che mi hanno dato qualcosa nel mio percorso di vita fino a qui.
La dedico ai miei nonni, Anna, Alessandro, Giampiero, Nia, per i bellissimi ricordi che mi hanno lasciato di loro.
La dedico ai miei familiari, Marcello, Caterina, Chiara, per tutto il sostegno che mi hanno sempre offerto.
La dedico ai miei compagni di studi, Andrea, Tommaso, Hristo e tutti gli altri, perchè siamo cresciuti assieme in un periodo fondamentale per noi.
La dedico ai miei amici storici, Alessia, Martina, Alessandra, Alberto, con cui ho passato momenti indimenticabili.
La dedico a tutti gli animali che ho avuto e che ho, Paco, Iaco, Toby, Ciki, Mici, Nerone, Puntina, Pallina, perchè vi ho amati.
La dedico ai miei maestri, Emanuele, Marcello, Elmar, per avermi generosamente trasmesso le loro conoscenze e aver reso possibile tutto questo.

Abstract

My doctoral research focused on different issues connected to interaction quantification in Computer Experiments. In particular, I provided new results on two main related topics: (i) the nature and (ii) the estimation of interactions. Concerning the first aspect, in the Thesis it is shown that interactions can have infinitesimal, local and global effect, and that they can be synergistic and antagonistic. Accordingly, different methods should be used for specific interaction locations and types. I have shown how it is possible to connect insights delivered by different methods. I have also provided a new approach to interaction quantification based on game theory.

Contents

Chapter 1

Introduction

This Thesis offers a deep analysis of the multi-sided problems releted to interaction quantification in Computer Experiments. Every Chapter is a research paper containing an original contribution and almost all of them have been already submitted. To better evaluate the quality of every paper, I have inserted it without modifications.
In particular, Chapter 2 offers an unified review to interactions as in the statistical literature. There it is shown that the estimation methods and also the definition of interaction can be debated topics.
Chapter 3 is devoted to study interactions with focus on Computer Experiments. It contains a general treatment concerning their nature, types, the estimation techniques and their relative costs.
Some of these techniques are developed in Chapter 4, where asymptotic results of the interaction estimators are derived. Hence, it is possible to quantify uncertainty in presence of finite sample size. Also, it is possible to make best use of model runs to achieve bigger accuracy given a total budget of simulation runs.
When model inputs are dependent, the techniques for global interaction quantification used in above Chapters can't be used. Chapter 5 proposes a new approach to overcome such difficulty using the notion of Shapley value from game theory and opens the door to future work in this direc-

tion. This Chapter has been accepted for pubblication in the SIAM/ASA Journal on Uncertainty Quantification.
Chapter 6 presents the conclusions of this Thesis and future perspectives and work that can originate from this Thesis.

Chapter 3 is a joint work with Elmar Plischke (TU Clausthal), whom I would like to warmly thank for his time and suggestions while he was visiting Universitá Bocconi.

Chapter 2

Interactions in Statistical Modeling: No Free Lunch

Abstract

The statistical literature displays a heterogeneous fragmentation of definitions, applications, interpretations of the term 'interaction'. This has lead the statistical community to discussions and controversies that are still open today. In this work, we provide an exploration of the concepts of interaction discussed in the statistical literature. We discuss the determination and role of interactions in contingency tables, in the Analysis of Variance (ANOVA), in design of experiments, in the analysis of computer experiments. We also review recent research trends concerning the identification of interactions in dimensionally large models and in causal inference. The analysis has several implications within each subfield, however an overarching implication of the work is that of providing us with a way to overcome specialization barriers that several times are only apparent.

2.1 Introduction

Interactions are ubiquitous in Statistics. Identifying and interpreting them is a fundamental part of statistical inference [Cox, 1984, Lekivetz and Tang, 2011] and in many scientific studies determining the presence or absence of interactions is the primary focus (see, e.g., Vansteelandt et al. [2008]). There are several practical and theoretical reasons which motivate the study of interaction. Interactions can provide deep insights into the mechanisms for the outcome of an experiment. Jointly testing for a main effect and for an interaction simultaneously may increase the power to detect the effects [VanderWeele and Knol, 2014]. Identifying which factors interact is fundamental for a correct model specification and modelling interactions can result in increased model flexibility and predictive ability - see Wahba et al. [1995] in smoothing spline ANOVA modelling among others. In epidemiology, the analysis of interactions can be of help to select the optimal subgroup to treat [VanderWeele et al., 2018].
However, the statistical literature displays a heterogeneous fragmentation of definitions, notations, applications of the term 'interaction', whose use is not univocal. This has lead to discussions and, sometimes, controversies that are still open today. There is a variety of techniques tailored to specific contexts that has created fragmentation and separations. For instance, the meaning and use of the term interactions in design of experiments (DOE) is different from the meaning of interactions as intended in the works of Lancaster [1969] and Streitberg [1990]. In connection with genetics, Clayton [2009] writes that "this topic [interaction] has received much attention, but with scant reference to the lively debate of the early 1980s, which was initiated in response to widespread over-interpretation of interaction in logistic regression models. It has been widely noted that statisticians and biologists attach different meanings to the word interaction [...] the confusion between statistical and biological notions of interaction goes back to Fisher's 1918 paper " (p.2-3)
On the other hand, we have common inference problems across disciplines. For instance, the statistical testing for interactions is problematic in ge-

netics and in epidemiology for its low power [Greenland, 1983, Routman and Cheverud, 1997, Bien et al., 2015] and interpretation [Clayton, 2009]. Moreover, a primary and debated issue in the analysis of interactions is the causal interpretation of a statistical interaction model: after an analyst has identified an interaction through a statistical model, does the presence of this statistical interaction imply the existence of an underlying causal physical or biological mechanism? This problem has received much attention in the epidemiology and genetics literatures [Blot and Day, 1979, Rothman et al., 1980, VanderWeele, 2009, Clayton, 2009]. Berrington de González and Cox [2007] write: "From the statistical perspective, interaction is said to occur if the separate effects of the factors do not combine additively. That is, interaction is a particular kind of non-additivity. The terminology is in some ways unfortunate in that there is no necessary implication of, say, biological interaction in the sense of synergism or antagonism" (p.371). In this respect, VanderWeele and Robins [2008]and VanderWeele [2010] define more 'mechanistic' notions of interactions.

Aim of this paper is to examine this vast literature on interactions through a critical lens. The goal is that of helping the statistical modeler by having a broad view of the use of the term interaction, but also the statistical community that, in this way, has a way to overcome specialization barriers that several times are only apparent.

The paper starts analyzing interactions in contingency tables in Section 2.2. Section 2.3 presents the classical and the alternative definitions of interactions in the ANOVA framework, while interactions in Design of Experiments are presented in Section 2.4. Section 2.5 discusses the role of interactions in functional ANOVA models, with special emphasis on functional regression and computer experiments. The problem of identifying interactions in high-dimensional statistical models is presented in Section 2.6. With a focus on Epidemiology, Section 2.7 presents and discusses the limitations of the causal inference on statistical interactions. The Appendix presents the additive approach to interactions introduced by Lancaster [1969].

2.2 Interactions in contingency tables

A contingency table is the matrix resulting from a classification of a sample according to some attributes or characteristics of interest. In this context, the attention of the statistician focuses on the probability of a particular combination of attributes.
In the analysis of such table the issues in determining interactions "frequently arise [...]: are the entries in the contingency table consistent with the assumption that an individual's possession of one attribute is independent of his possession of any other?"[Coen, 1971, p. 379]. This question highlights a crucial aspect: interactions in contingency tables are not understood in terms of additivity of effects as in factorial experiments (described in Section 2.3), but are interpreted as an association measure among attributes "the concept of no interaction associated with this kind of experiment has nothing to do with the way in which factors combine to determine a response as in the case of analysis of variance models", [Bhapkar and Koch, 1968, p. 108]). This is particularly evident in a 2×2 contingency table: citing Bartlett [1935], *testing of independence in a* 2×2 *table may be regarded as as testing the significance of the interaction between the two classifications* (p. 248). In multivariate contingency tables, Simpson [1951] introduces an association function of the attributes to model situations in which "the cell probabilities can depart from complete independence" (p. 239). Everitt [1992] uses the term interaction "as an alternative to the term association when describing a relationship between the qualitative variables forming a contingency table" (p. 73).

In general, there are two main theories to define interactions in multivariate contingency tables: the 'additive' and the 'multiplicative', both motivated as generalizations of independence [Darroch, 1974]. Consider an $r \times c \times l$ contingency table with probabilities $\{p_{ijk}; i = 1, 2, ..., r; j = 1, 2, ..., c; k = 1, 2, ..., l\}$, $\sum_{i,j,k} p_{ijk} = 1$. In this case, the multiplicative definition of zero second-order interaction has been given by Roy and Kastenbaum [1956] and states that there is no second-order interaction if the

cross-products satisfy

$$p_{rcl}p_{ijl}p_{ick}p_{rjk} = p_{icl}p_{rjl}p_{rck}p_{ijk}. \tag{2.1}$$

Condition (2.1) coincides with the definition given by Bartlett [1935] for the case of $2 \times 2 \times 2$ tables

$$p_{222}p_{112}p_{121}p_{211} = p_{122}p_{212}p_{221}p_{111}.$$

In general [Roy and Kastenbaum, 1956, Lancaster, 1971], the condition of no multiplicative second-order interaction (2.1) can be written as

$$p_{ijk} = \theta_{jk}\phi_{ik}\psi_{ij} \qquad \text{for all } i, j, k \tag{2.2}$$

for some $\{\theta_{jk}\}$, $\{\phi_{ik}\}$, $\{\psi_{ij}\}$. Intuitively, there is no second-order interaction according to the multiplicative definition if the cross-product ratios, interpreted as a measure of association, don't vary across layers. Hence, this multiplicative approach to interaction has been related to the concept of conditional independence of the variables [Lewis, 1962, Lancaster, 1971].

The additive definition of no second-order interaction, introduced by Lancaster [1951] and formulated by him in various ways [Lancaster, 1969], can be expressed as

$$\frac{p_{ijk}}{p_{i\cdot\cdot}p_{\cdot j\cdot}p_{\cdot\cdot k}} = \alpha_{jk} + \beta_{ik} + \gamma_{ij} \qquad \text{for all } i, j, k \tag{2.3}$$

for some $\{\alpha_{jk}\}$, $\{\beta_{ik}\}$, $\{\gamma_{ij}\}$, where dots denote marginalization of probabilities. In this additive approach, the measure of association is the correlation coefficient (Lancaster [1969], p. 254-260). To motivate why this approach is usually termed 'Lancaster-additive', note that Darroch [1974] proves that the definition (2.3) is equivalent to

$$p_{ijk} = p_{i\cdot\cdot}p_{\cdot jk} + p_{\cdot j\cdot}p_{i\cdot k} + p_{\cdot\cdot k}p_{ij\cdot} - 2p_{i\cdot\cdot}p_{\cdot j\cdot}p_{\cdot\cdot k}, \tag{2.4}$$

which is analogous to equation (2.34).

Both approaches have found applications in the statistical literature. Lancaster's additive models have been used to investigate interactions in contingency tables, for instance, in O'Neill [1982b] and in Töwe et al. [1985],

while multiplicative models á la Bartlett can be found in Goodman [1964], in Gart [1972] and in Patil [1974].

In general, however, the two definitions (2.2) and (2.3) are clearly not equivalent. These two definitions are proved to be equivalent only under certain conditions [O'Neill, 1982a] and, in two-dimensional tables, they are both equivalent to the independence of the two variables [Darroch, 1974]. The additive and multiplicative approaches are compared in various works, according to different criteria (Darroch [1974]; Murphy [1978]; Kroonenberg and Anderson [2006]). Darroch [1974] considers Lancaster's additive approach to the theory of interaction in contingency tables as obscuring the "simple relationship that it has with interaction in linear factorial models" (p. 209), although the meaning and interpretation of higher-order interactions differ from their ANOVA counterparts [O'Neill, 1982b]. As a matter of fact, Darroch [1974] concludes that both additive and multiplicative definitions "are very far from being ideal definitions of no-interaction, but is seems reasonable to guess that no better definition exists" (p. 213).
Nonetheless, there is no agreement about which definition should be preferred in general, because large differences can arise between the two approaches [O'Neill, 1982b]. Consequently, the literature displays a series of discussions about controversial results and applications. For instance, O'Neill [1982b] writes that Lancaster's additive approach to interactions has found large acceptance in agricultural studies, but it is used "sometimes incorrectly" (p. 167). Streitberg [1999] describes several problems connected with the multiplicative approach; "I therefore propose to take a fresh look at additive interactions" (p. 406). In the psychological literature, Everitt and Smith [1979] analyze a controversy about two alternative interpretations of the interactions in the same data set ("the disagreement [...] arises from different methods they use in defining interaction", p. 581).
There are authors with a conciliatory position. For example, Lewis [1962] deems Lancaster's additive approach "not intuitively satisfactory" (p. 102) and "lengthy and difficult to verbalize" (p. 104). However, at the same time, he recognizes the difficulties that, "since there is not even general

agreement on how to measure first-order interaction in a simple two-way table, it might seem that the investigator is free to set up and justify a wide variety of possible criteria" (p. 102). Along the same line, O'Neill [1982a] writes that his "own view is that one's understanding of an interaction should be the guiding factor in selecting an appropriate definition [...] so that an informed decision can be taken as to their relative merits" (p. 35). More recently, Kroonenberg and Anderson [2006] suggest that there is no unequivocal criterion, based on model interpretability, for the decision of whether one should adopt additive or multiplicative modeling. The same conclusion is reached by Walter and Holford [1978] in the epidemiological literature ("the recommendation of a single model for all situations is not desirable and the choice depends on the particular situation in hand", p.346).

The use of the multiplicative approach, however, has become predominant with the advent of log-linear models for the analysis of contingency tables (Bhapkar and Koch [1968]; Nelder [1974]). This idea of studying interactions on a transformed scale is used also by Cox and Lauh [1967], who investigate the additivity of the effects of a contingency table on a logistic scale.

The common approach has then become to fit many log-linear and logistic models to a given contingency table as a first step and, as a second step, to focus on a group of models that fit reasonably well (Fowlkes et al. [1988]). This approach has been made possible by the increase in computational power. In this way of proceeding, interaction quantification is related to model fit. According to Everitt [1992], the major advantages deriving from the use of model-fitting techniques are "firstly that they provide a systematic approach to the analysis of complex multidimensional tables and secondly that they provide estimates of the magnitude of effects of interests; consequently they allow the relative importance of different effects to be judged" (p. 73). However, Kroonenberg and Anderson [2006] criticize such approach: "both the multiplicative and additive models for three-way tables have much to offer compared to simple significance tests and numerical inspection of the interactions or modeling of them via log-

linear models" (p. 484).

For the log-linear modeling of multi-way contingency tables, a variety of graphical representations is available to support the model choice and its interpretation. In particular, the class of graphical models (Darroch et al. [1980]; Edwards and Kreiner [1983]) can aid model selection providing the analyst with a better understanding of complex interaction structures. Graphical models constitute a special class of hierarchical log-linear models [Goodman, 1970] whose key feature is that they can be visualized from graphs where vertexes represent variables and edges represent interactions. This makes graphical models easily interpretable as the absence of an edge corresponds to the conditional independence of the two variables ("for example, the apparent association between two variables, occupation and longevity, may be due to a third variable, sex, in the sense that for each sex considered separately, occupation and longevity are independent. This may be expressed by saying that occupation and longevity are conditionally independent given sex", Edwards and Kreiner [1983], p. 553).
The above discussion shows some controversy in the interpretation of interactions for symmetric tables. The situation becomes even more involved for non-symmetric tables. Namely, in many applications multi-way contingency tables may show asymmetries in the interaction structure. For example, it may happen that two chemicals interact only at a specific level of another chemical and this interaction vanishes at all the others level of the third chemical: this is an example of *context specific interaction* [Højsgaard, 2003]. Højsgaard [2003] introduces the class of split models, that are graphical models focusing on this structural asymmetry in the interaction pattern. Split models generalize the class of graphical models for contingency tables and are representable by split graphs. However, care is needed when drawing conclusions on interactions with these models [Højsgaard, 2003]. In particular, "it is a delicate question whether a context specific independence is genuine or appears simply because of low power of some of the the tests" (p. 642). Højsgaard [2004] introduces the more general class of context specific interaction models. This class of log-linear models systematically reflects the asymmetric structure of

context specific independences and context specific interactions and extends the split models ("split models are context specific interaction models for which the focus is on context specific independence restrictions", Højsgaard [2004], p.143).

The theory of interactions in contingency tables, despite being a time-honored problem in Statistics, is still an area of active research. The notion of conditional relationship structures pertain to recent developments. Hara et al. [2012] provide a general framework for modeling interaction terms with hierarchical subspace models respecting the conditional independence structure. Recently, for a two-way contingency table, D'Ambra et al. [2017] study main and interaction effects by means of non-symmetric correspondence analysis, which is a technique for studying two-way contingency tables with an asymmetric relationship between the variables.

We conclude this section noting that Darroch and Speed [1983] provide a unified framework for studying interactions in different models: multiplicative and Lancaster-additive for multi-way contingency tables, and classical additive for complete factorial experiments, that we discuss in the next section. In order to give this unified treatment, Darroch and Speed [1983] introduce the definition of generalized interactions "from an algebraic point of view in terms of fundamental decompositions of the linear space of functions on a product of finite sets" [Lauritzen, 2012]. Finally, the idea of using projection operators to model interaction effects is central in the functional ANOVA decomposition that we discuss in Section 2.5.

2.3 Interactions in the Analysis of Variance (ANOVA)

Analysis of variance (ANOVA) of multi-factor experiments is one of the most adopted procedures in statistical analysis to investigate differences among group means. Nonetheless, according to Hector et al. [2010], this technique is a debated topic within the statistical community. The main

reasons are the difficulties in understanding ANOVA ("there is some debate on how to perform ANOVA in complicated data structures [...] it is not at all obvious that a procedure recognizable as ANOVA should be possible at all in general settings", Gelman [2005], p.2), and, consequently, the ambiguities arising from its interpretation [Kempthorne, 1975, Shaw and Mitchell-Olds, 1993]. In his historical review about ANOVA, Herr [1986] speaks about the "considerable confusion as to how one should analyze such designs" (p. 265). Interaction effects are in particular among the most debated aspects. Precisely, the controversy concerns not only technical aspects of interaction quantification, but also the definition of interactions themselves. This concern appears even more serious considering the twofold role played by interactions in ANOVA modeling: as indicators of the complexity of the relations within the experimental model on the one hand, and as a criterion for choosing the appropriate statistical model on the other hand.

To illustrate such problems, let's consider the standard two-way ANOVA model and let A and B be the two treatment factors, having I and J states (or levels) respectively. The aim is to study how the value of the response varies across cells, i.e., with different level combinations. The standard two-way ANOVA model with K_{ij} observations in cell (i, j) is

$$Y_{ijk} = \mu_{ij} + \epsilon_{ijk} = \mu + a_i + b_j + c_{ij} + \epsilon_{ijk} \tag{2.5}$$

with $i = 1, ..., I, j = 1, ..., J, k = 1, ..., K_{ij}$, where μ is the overall mean, a_i is the i-th main effect for A, b_j is the j-th main effect for B and ϵ_{ijk} are independent errors distributed as $N(0, \sigma^2)$. The non-additivity of the factor effects is highlighted by the presence of the interaction term c_{ij}. For the identifiability of parameters μ, a_i, b_j and c_{ij}, some conditions are usually imposed on main and interaction effects, respectively

$$\sum_i a_i = 0, \qquad \sum_j b_j = 0, \tag{2.6}$$

$$\sum_i c_{ij} = 0 \quad \forall j, \qquad \sum_j c_{ij} = 0 \quad \forall i. \tag{2.7}$$

For the estimation of such effects, different cases must be distinguished on the basis of the number of observations in each cell. In balanced designs the number of observations is the same for every cell, so that $K_{ij} = K$. For the analysis of such designs, ANOVA is generally considered to be the best technique: while one factor is kept fixed, the other factor can vary independently. The total variance can be decomposed into sum of squares related to main and interaction effects, each of them independent χ^2 with its respective degree of freedom [Shaw and Mitchell-Olds, 1993].
Despite such situation of perfect balance, the case in which $K = 1$, i.e. one observation per cell, is very controversial: the power of the test for interaction becomes unknown, the distribution of the test statistic is difficult to treat causing ambiguity in the partition of the sum of squares, and it becomes unclear how much importance should be attributed to interaction effects or to the residual [Scheffé, 1959]. In this case, Scheffé [1959] assumes the equality of cell variances and states that "the statistical inferences are not seriously invalidated [...] by violation of the assumption of equality of the cell variances" (p. 98). However, Snee [1982] points out that, when there is only one observation per cell, the non-additivity of the effects can be due either to non-homogeneous variance or to a systematic interaction: "this distinction is necessary to properly interpret the results of the experiment and gain an understanding of the system that generated the data" (p. 519).

It should be noted that the "arbitrariness in the definition of effects and interactions" (Vajda [1951], p. 283), the superficial formalizations [Herr, 1986] and the lack of a unified terminology have been relevant sources of difficulties in the historical development of ANOVA techniques. For instance, Tukey [1949], who is concerned about "the lack of a systematic way to measure non-additivity", uses with the same meaning the terms residue, error, interaction and discrepancy, "call it what you will" (p. 232). Another example: Finney [1948] does not state and formulate any model and writes that "when the interaction is not negligible, estimation of the main effects of A must depend upon circumstances, and no definition [...] can be claimed as the only legitimate one" (p. 570).

The problem is clearly how the interaction c_{ij} should be understood: should it be bound together to the main effects or not? Early on, Scheffé [1959] states that, in the presence of a non-negligible interaction, there are difficulties in interpreting and evaluating also the main effects. The debate on this aspect is still open and evidences the ambiguity in the role attributed to the interaction effect.

Many authors maintain that main effects should play a predominant role compared to interactions in a statistical analysis. Daniel [1978] writes that "main effects are expected to predominate [...] the non-additivity may afflict only a small number of minority of cells, perhaps only a single cell" (p.385). Nelder [1977] considers of "no practical interest" (p. 50) those models having an existing interaction in presence of zero main effects and makes an universal prohibition of such models. This prohibition is criticized by his discussants Lindley, Cox, Preece, Tukey, Frane and Jennrich. Later on, McCullagh [2000] objects that these models can occur as result of experimental designs but are rare in other contexts.

Indeed, the asymmetry between main and interaction effects is a crucial point to understand different views about interaction. For instance, Williams [1952] proposes that main effects should minimize interactions: "the basic assumption is that, with main effects suitably defined, interactions do not exist" (p. 71). Along this line of thought, Davies and Terbeck [1998] formulate the interaction quantification problem as a minimization problem: "in spite of its long standing, [...] the standard definition of interactions is counterintuitive and obfuscates than clarifies. A different definition is given which [...] allows the detection of interactions even in the case of one observation per cell" (p. 1279). The principal reason for this critique is that some authors [Tukey, 1993, Davies and Terbeck, 1998, Ning and Kim, 2008, Davies, 2012] consider the side conditions (2.7) on the interaction term as obfuscating the real meaning of interactions and operate a factors re-parametrization in order to transform interactions into main effects, in the spirit of Cox [1984]: "if all or specified interactions can be removed by transformation, part at least of the interpretation will usually exploit this" (p. 15).

Formulating an alternative definition without the interaction constraints, Davies [2012] considers the expected response model (2.5)

$$y_{ij} = \mu + a_i + b_j + c_{ij}, \qquad i = 1, ..., I, \quad j = 1, ..., J, \tag{2.8}$$

even with one observation per cell and defines interactions as the residuals when the norm

$$||\mathbf{y} - \mathbf{a} - \mathbf{b}||_{L^0} = \sum_{i,j} \mathbb{1}\{|y_{ij} - a_i - b_j| > 0\} \tag{2.9}$$

is minimized without restrictions on the parameters $\mathbf{a}$ and $\mathbf{b}$, where $\mathbb{1}$ is the indicator function. In other words, for Davies [2012] the correct approach is "to choose the factor effects to minimize the number of nonzero interactions" (p. 1508).
Many authors consider the expected value of the response $E[X_{ijk}]$ to study interactions with the ANOVA structure as in equation (2.8), including Cox [1984]. In this case, the interaction measure between the factors A and B is the quantity

$$c_{ij} = y_{ij} - \mu_{i\cdot} - \mu_{\cdot j} + \mu, \tag{2.10}$$

where $\mu_{i\cdot}$ and $\mu_{\cdot j}$ denote the mean for the i-th row and the j-th column, respectively [Scheffé, 1959]. In the literature one can find an alternative nonparametric definition of interaction measure in 2×2 factorial experiments which does not require the linear model assumption (2.8), given by Patel and Hoel [1973]. For a two-factor experiment with each factor at two possible levels, i.e. $I = J = 2$, Patel and Hoel [1973] consider the probabilities $p_{11,12} = P(Y_{11k} < Y_{12k})$ and $p_{21,22} = P(Y_{21k} < Y_{22k})$. The Patel-Hoel definition of no interaction is that $p_{11,12} = p_{21,22}$ and the hypothesis for no interaction is formulated as $H_0 : p_{11,12} = p_{21,22}$. In particular, the nonparametric test for no interaction under the Patel-Hoel H_0 shows greater power than the usual $F-$test for interactions.

The above discussion has concerned the case of balanced ANOVA designs. The debate is even greater when the number of observations per cell is not the same, that is, in the case of asymmetric ANOVA tables. When

the number of observations is different across cells, the design of the experiment is called unbalanced and this loss of symmetry affects statistical inference, originating ambiguities in parameters estimation, in their interpretation, in hypothesis testing and in the partition of the sum of squares relatively to main and interaction effects [Shaw and Mitchell-Olds, 1993, Landsheer and van den Wittenboer, 2015]. These complications have been known for a long time and solutions to these problems have been proposed. A very popular approach, which is implemented in most of statistical computer packages, is contained in Yates [1934]. The work of Yates [1934] is fundamental: he proposes three methods to compute sum-of-squares and testing hypotheses, commonly known as SS I (proportional cell sizes), SS II (fitting constants) and SS III (weighted squares of means). However, the only general agreement is on the inadequacy of the SS I method. Moreover, there is no unanimous agreement on which one of the methods SS II and SS III is best, as discussed in Overall and Spiegel [1969], Nelder and Lane [1995], Lewsey et al. [2001] and Langsrud [2003]. One also does not find unanimous agreement on the terminology to indicate these methods, as shown in Table A in Smith and Cribbie [2014]).
Interactions play a crucial role in unbalanced designs because the choice of the correction method depends on their significance. In this context the estimation of interactions is a criterion for model selection: when the interaction estimate is statistically significant, SS III is usually preferred to SS II; SS II in the other case. Hence, the estimation of interaction becomes the first step of the analysis: Ng [1994] proposes the use of preliminary tests for the detection of interaction in two-way tables, in order to "select the model, or to check certain assumptions and then the primary inference is made on the basis of the outcomes of the preliminary tests" (p. 437). However, investigators do not unanimously agree on the practical decision rule to discriminate between SS II and SS III. For instance, Landsheer and van den Wittenboer [2015] suggest that the criterion should be based on the conceivability of the presence of interactions and not on their statistical estimate.

This discussion shows that many aspects in the two-way analysis of

factorial experiments lead to debate among statisticians concerning the interpretation of interactions. Hector et al. [2010] wish open mindedness in the debate, providing the recommendations of focusing on the objective of the analysis and not on the "single right ANOVA table".

2.4 Interactions in Factorial Designs

Factorial designs are widely used in statistical applications, from biology to the social sciences, to engineering. In a factorial experiment the investigator is interested in gaining as much information as possible about the effects of treatment factors which can be simultaneously varied under the researcher's control. Lewis and Dean [2001] write that "a major advantage of factorial experiments is the information that they provide on interactions" (p. 633). As a matter of fact, the researcher could run a sequence of experiments using one-at-a-time plans [Daniel, 1973], but in this case she would not gain any information about factor interactions: "interaction among factors can be assessed in a factorial experiment but not from series of one-at-a-time experiments ", [Cox and Reid, 2000]. The literature has shown that there are many reasons for which the estimation of interactions in experiments is fundamental although delicate. Precisely, "ignoring interactions can lead to (i) important effects being missed, (ii) spurious effects being detected and (iii) estimated effects having reversed signs resulting in incorrectly recommended factor levels" (Hamada and Wu [1991], p.3). Additionally, with application to industrial statistics, "(a) an underlying physical mechanism may suggest a large interaction between some factors [...] (b) it is desired to estimate specific interactions even when they may turn out to be small [...] (c) in robust parameter design, the interaction between a control factor and a noise factor may be important for making a product insensitive to noise variations" (Wu and Chen [1992] p. 162). The study of interactions in factorial experiments presents relevant and interesting features, as we are about to see.
A first important class of designs with k factors, each at two levels, is the class of 2^k full (or complete) factorial designs, in which a single replicate

of the experiment corresponds to a specific combination of the levels of the factors and, hence, the number of replicates equals the number of the possible combinations. A combination of factors is usually called treatment and it can be viewed as a vertex of a k-dimensional cuboid, with the common notation $(-,+)$ representing the lower and upper factor level.

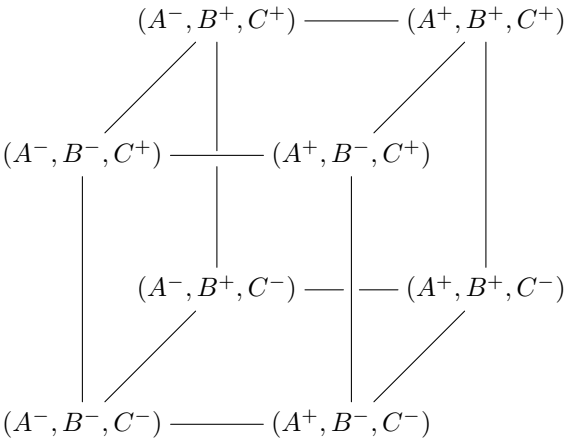


Figure 2.1: Cuboid representation of a 2^3 design.

In this situation, a main effect for a factor, say A, is defined as difference between the mean values of all observations at the level + of A and at level - of A. The $A \times B$ interaction effect is defined as [Wu and Hamada, 2009]

$$\mathrm{INT}(A,B) = \frac{1}{2}[\bar{y}(A^+|B^+) + \bar{y}(A^-|B^-) - \bar{y}(A^+|B^-) - \bar{y}(A^-|B^+)], \quad (2.11)$$

where $\bar{y}(A^+|B^+)$, $\bar{y}(A^-|B^+)$, $\bar{y}(A^+|B^-)$ and $\bar{y}(A^-|B^-)$ are the outcome average value of the response y at the different treatment combinations A^+B^+, A^-B^+, A^+B^- and A^-B^-, respectively. We highlight the structural similarity of the interaction equation (2.11) to equation (2.10) and

the symmetry of this definition in A and B. Absence of interaction is here equivalent to the additivity of the factors effects. On the other hand, the presence of an interaction, which is a linear function of the observations, indicates "that the effect of one factor depends on the level of the other factor, which explains the use of the term interaction" (Wu and Hamada [2009], p. 164). Citing Cheng [2014], "the main effect of a treatment factor is its effect averaged over the levels of the other factors, and the interaction effects measure departures from additivity" (p.4). This intuition is analogous to the ANOVA. For instance, when the experiment involves two treating factors, i.e. $k = 2$, factorial effects can be analyzed by the ANOVA for the two-way layout (2.5), presented in the previous section. However, one notes that the agreement about the relation between the formulation of interaction effects in general 2^k factorial experiments and their ANOVA modeling is not unanimous. While Wu and Hamada [2009] consider the two approaches in principle equivalent, Box et al. [2005] write that "for the analysis of these particular designs the use of ANOVA is confusing and makes little sense" (p. 188). Moreover, in the psychological and social sciences literatures, there is controversy about what constitutes an interaction effect in the ANOVA of a factorial design (Jaccard [1998]).

In the design of experiments, peculiar aspects are associated to interaction quantification. To illustrate, suppose that an experiment on two-level factors A and B will be run on two different days: the first day the experiment will be run with the factor combinations (A^-, B^-) and (A^+, B^+), the second one with the combinations (A^+, B^-) and (A^-, B^+). It follows that the estimation of the main effects of A and B is done using the results of the first day, and the interaction effect can be estimated using the data of both days [Cheng, 2014]. Then, such interaction is said to be *confounded* because its estimation procedure might result in less precision, e.g. for different weather conditions in the two days, different experimenter, different material and so on. To formalize, it is in general impossible to run complete replicates under homogeneous conditions and/or without nuisance factors. To make the experiment more efficient and robust to noisy conditions, variables need to be arranged in blocks to

account for non-homogeneous conditions or nuisance factor levels [Myers et al., 2016]. Typical examples of blocks can include location, time, raw material, operator, and so on. As we have seen, this operation might lead some estimates of block effects to be indistinguishable from those of low-order interaction terms: these effects are then said to be *confounded* with blocks because they cannot be estimated separately. Usually the main effects are of primary interest and, hence, the researcher would like to avoid to confound them with block effects. In order to select the optimal blocking scheme to arrange a 2^k design in 2^q blocks with $q < k$, the common approach is to use the minimum aberratio criterion [Sun et al., 1997, Wu and Hamada, 2009]: this criterion selects the blocking scheme minimizing the number of lower order interactions that are confounded with any block effect (because in principle no main effect should be confounded with block effects). Therefore, in this context as well, the attention of the statistician is focused on interactions, which are the basis of the decision rule for the grouping scheme and its optimality.

In high-dimensional settings it often becomes impossible or too expensive to observe all possible factor combinations. Then, a full 2^k factorial design, whose size grows exponentially with k, cannot be used and the analysis has to be carried out on a smaller subset of treatment combinations. This is called a fractional factorial design. Fractioning the full factorial design is an more economic way to estimate the lower order effects and, often, the only possible way. This operation has however relevant disadvantages. Since not all experiments corresponding to all possible treatment combinations have not been run, some factorial effects become indistinguishable. In this case, main and interaction effects are said to be *aliased*. In oder words, aliasing of effects occurs when the experimental data in the fractional design is not able to distinguish certain main and interaction effects among factors, making it difficult to discern between significant and negligible interactions: citing Wu and Hamada [2009], it is "a price one must pay for choosing a smaller design" (p. 213).

Example 2.4.1. *[Cheng, 2014, Myers et al., 2016] Consider the case of a* 2^6 *experiment. There are* 64 *possible treatment combinations to estimate*

63 *factorial effects (*6 *main effects,* 15 *two-factor interactions,* 20 *three-factor interactions,* 15 *four-factor interactions,* 6 *five-factor interactions and* 1 *six-factor interaction). However, when only* 16 *treatment combinations can be experimented, there are only* 15 *degrees of freedom to estimate the* 63 *effects:* 6 *are used to estimate the main effects and only* 9 *degrees of freedom can be used to estimate* 57 *two or higher factor interaction effects. This can be done only assuming that many interaction effects are null.*

In general, when the experimenter has no or little knowledge about the relationships among factor effects, the following three guiding principles are adopted (Wu [2015]):

- *Effect hierarchy* [Wu and Hamada, 2009]: lower order effects (main effects and low order interactions) are assumed to be more important than higher order effects, and effects of the same order are equally important.
- *Effect sparsity* [Box and Hunter, 1961]: the number of important effects is small.
- *Effect heredity* [Hamada and Wu, 1991]: for one interaction to be significant, its parent main effects have to be significant.

These principles are commonly assumed also in the framework of high-dimensional regression (Section 2.6). Historically, these principles were implicit in the design of experiments literature (see Wu [2015] for a discussion). For examples, Cox [1984] writes that "large component main effects are more likely to lead to appreciable interactions than small components. Also, the interactions corresponding to larger main effects may be in some sense of more practical importance" (p.13). Yates [1937] states that "higher-order interactions [...] are usually of less interest than the main effects and interactions between two factors only" (p.18).
With application to fractional factorial designs, the effect hierarchy assumption justifies "the choice of optimal fractions of factorial designs" (Wu [2015], p.613). In the literature, criteria based mostly on resolution [Box and Hunter, 1961] and minimum aberration [Fries and Hunter, 1980]

are popular principles for constructing, comparing and assessing fractional factorial designs. Resolution is a property of fractional factorial designs and concerns the propagation of aliased effects. In a resolution III design no main effect is aliased with any other main effect, but two factor interactions are aliased with some main effects; in a resolution IV design no main effect is aliased with two factor interactions but some two factor interactions are aliased with other two factor interactions; in a resolution V designs no main effect and two-factor interaction are not aliased, but some two factor interactions are aliased with three factor interactions. It is then desirable to have designs with higher resolution, but the resolution can not be the only decision criterion for a design. Namely, designs with the same resolution can be discriminated by the minimum aberration criterion: in this context, denoting with R the maximum attainable design resolution, "minimizing aberration [...] means that the smallest number of main effects will be confounded with interactions of order $R-1$, the smallest number of two-factor interactions will be confounded with interactions of order $R-2$, and so forth" (Fries and Hunter [1980], p. 602). Minimum aberration automatically implies maximum resolution. Again, interaction effects play a primary role in the construction of designs with certain properties.

There are other notional aspects related to interactions in the design of experiment literature. When the experimenter needs to block a fractional factorial design, a two factor interaction is said to be *clear* if it is not aliased with any main and any other two-factor interaction effects [Wu and Chen, 1992]. The relationship between minimum aberration designs and designs with the highest number of clear interactions is explored in Wu and Wu [2002]. The major advantage of designs with clear two factor interactions is that "they allow unbiased estimation of these clear two-factor in the presence of other two-factor interactions, and therefore provide a class of designs that are robust to non-negligible two-factor interactions" (Tang [2006], p.137). Two factor interactions are also said to be *strongly clear* if they are not aliased with three factor interactions. Lekivetz and Tang [2011] generalize the concept of clear interaction effects introducing the

notion of *partially clear* interactions: an interaction effect is partially clear if it is orthogonal to non-negligible interactions but is allowed to be aliased with negligible interactions.

Recently, Wu [2015] propose a different view to interactions and effect aliasing. He shows that it is possible to de-alias effects using the notion of conditional main effects (CME) analysis. The conditional main effect of factor A given B at the + level is defined as [Wu and Hamada, 2009]

$$\text{CME}(A|B^+) = \bar{y}(A^+|B^+) - \bar{y}(A^-|B^+). \tag{2.12}$$

Then, it is possible to interpret $\text{CME}(A|B^+)$ as the main effect of A conditional to the level + of B. Using this definition, equation (2.11) can be rewritten as

$$\text{INT}(A|B) = \frac{1}{2}[\text{CME}(A|B^+) - \text{CME}(A|B^-)] = \frac{1}{2}[\text{CME}(B|A^+) - \text{CME}(B|A^-)]. \tag{2.13}$$

This re-parametrization of the space of effects induces non-orthogonality among effects. The CME analysis allows to obtain de-aliased effects, but does not fit the framework of the three principles of factorial designs. As Wu [2018] underlines, possible new statistical challenges will be in the direction of developing a theoretical framework for fractional designs using the CME analysis.

2.5 Interactions in functional ANOVA models

The functional ANOVA representation of a multivariate mapping is a fundamental modelling tool for a variety of problems in Statistics and applied sciences. Its wide applications include chemical sciences (Li et al. [2001]), laboratory processes (Tarrío-Saavedra et al. [2011]), model emulation (Muehlenstaedt et al. [2012]), finance (Wang [2006]), ecological statistics (Estévez-Pérez and Vilar [2013]) and climate change studies (Sain et al. [2011]).

The functional ANOVA decomposition of a multivariate function is a very powerful alternative to the classical Taylor expansion, as it is exact and of finite order and provides insights about the mapping internal structure without any assumption about the differentiability of the multivariate function. Consider a multivariate integrable function $f(\mathbf{x})$, where $\mathbf{x} = (x_1, x_2, ..., x_n) \in \mathcal{X}$, with respect to a product measure μ. Then we can expand $f(\mathbf{x})$ as

$$f(\mathbf{x}) = f_0 + \sum_{i=1}^{n} f_i(x_i) + \sum_{i<j} f_{i,j}(x_i, x_j) + ... + f_{1,2,...,n}(x_1, x_2, ..., x_n), \quad (2.14)$$

where

$$\begin{cases} f_0 = E[f] = \int \cdots \int f(\mathbf{x}) d\mu \\ f_i(x_i) = E[f|x_i] - f_0 = \int \cdots \int f(\mathbf{x}) \prod_{k \neq i} d\mu_k - f_0 \\ f_{i,j}(x_i, x_j) = E[f|x_i, x_j] - f_i(x_i) - f_j(x_j) - f_0 = \int \cdots \int f(\mathbf{x}) \prod_{k \neq i,j} d\mu_k - f_i(x_i) - f_j(x \\ \cdots \end{cases} \quad (2.15)$$

Every quantity appearing in expansion (2.14) has its corresponding ANOVA name: the zero-degree term f_0 is the mean value of f. The univariate functions $f_i(x_i)$ are called main effects, or first-order effects, and represent expected variations of f as a function of x_i (the main effects in the functional ANOVA-type decomposition can be also called Hajek projection, see Hajek [1968]; Takemura [1983]). The bivariate functions $f_{i,j}(x_i, x_j)$ represent the second order interaction between x_i and x_j and are calculated subtracting from the expected conditional expectation of f as function of x_i and x_j all the lower effects, and so on for higher order interactions. The last term $f_{1,2,...,n}(x_1, ..., x_n)$ accounts for any residual interaction among all the variables together. Intuitively, in this ANOVA-type formulation, interactions are in mean the functional contributions representing what is not explained by independent variations of lower order terms: in other words, this is precisely the concept of interactions in an ANOVA table now embedded in a functional setting. This motivates the designation of the representation (2.14) as the 'functional ANOVA decomposition'.

The importance of the result in (2.14) is evidenced by the fact that its proof has been obtained in alternative ways and sometime independently by several researchers. In fact, in the literature, the research on this fundamental tool has proceeded dispersedly across disciplines, and (2.14) is known under different names. Historically, the classical ANOVA decomposition originates for experimental designs in agriculture, introduced by Fisher and Mackenzie [1923]. Later, Hoeffding [1948] introduces an ANOVA-type functional decomposition of a square integrable statistic and uses it in the study of U-statistics. This orthogonal decomposition has become an "indispensable tool of analysis of distributional properties of statistics" (Bloznelis and Götze [2001], p.899), so much that some authors refer to this decomposition as the Hoeffding or Hoeffding-ANOVA decomposition (Bloznelis and Götze [2001]; Peccati [2004]).
However, the connection between ANOVA and Hoeffding's decomposition appears clearer in the work of Efron and Stein [1981]. Efron and Stein [1981] establish the importance of Hoeffding's result, using it "propitiously" (Karlin and Rinott [1982]) to prove their lemma on the jackknife estimation of variance components. In Efron and Stein [1981] one finds clearly underlined the relation between the functional and the classical ANOVA decompositions, not explicitly stated in Hoeffding [1948]: "a function of n independent random variables $X_1, X_2, ..., X_n$ can be decomposed into main effects, interactions, higher order interactions, etc., in a manner directly analogous to the decomposition of a complete n-way ANOVA table" (p. 587). Karlin and Rinott [1982] generalize the results of Efron and Stein, extending the decomposition to the case of random functions of two groups of random variables.

An alternative approach to the functional ANOVA decomposition can be found in the works of Sobol' [1969, 1993]. He terms it "expansion into summands of different dimensions" and adopts it in the study of quadrature methods. Sobol' proves that there exists an uniquely representation of an integrable $f(\mathbf{x})$ as sum of functions of different dimensions and constructs this representation by expanding $f(\mathbf{x})$ into Fourier-Haar series. In this way, he gathers "together terms corresponding to each subset of vari-

ables. That is, where Hoeffding has an analysis, Sobol' has a synthesis" (Owen [2013], p. 21). Archer et al. [1997] compare ANOVA in classical factorial design with the approach of Sobol'. Chastaing et al. [2012] refer to the functional ANOVA representation as the Hoeffding-Sobol' decomposition.

Another way of proving the functional ANOVA decomposition is proposed by Takemura [1983]. He demonstrates the analogy between "ANOVA for general n-factor crossed layouts and the ANOVA-type decomposition of square integrable statistics" (p. 894) in the context of tensor analysis and multi-linear algebra. Moreover, he gives an historical account, showing that functional ANOVA-type representations were known as early as the end of the 1930s. Rabitz and Alis [1999] introduce an alternative formulation of the functional ANOVA decomposition, called high-dimensional model representation (HDMR). Their proof of the expansion is based on a different technique with respect to previous approaches and involves the theory of projection operators. General HDMR expansions may not only be used "for representing the input–output mapping over the operating region of the input variables", but are more broadly "multivariate approximation/interpolation schemes as well as a means to analyze the relevant statistics of a random output" (p. 209). In relation to the HDMR structure, Rabitz et al. [1999] point out the analogy with the multi-body expansion in molecular physics, which represents potential surfaces generated from the interactions among systems of atoms.

Rabitz and Alis [1999] also introduce the framework for the so-called cut-HDMR, a different HDMR expansion which is computationally more efficient than the functional ANOVA decomposition "as this approach does not require the evaluation of high-dimensional integrals of the output" (Rabitz et al. [1999] p. 13). Consider the Dirac measure located at a point $\mathbf{z} \in \mathcal{X}$. This point is called the cut center, and the cut-HDMR expresses the function $f(\mathbf{x})$ as superimposition of lines, planes and hyperplanes of higher order passing through the cut center $\mathbf{z}$. Then, the cut-HDMR for

f is

$$f(\mathbf{x}) \approx f^{cut}(\mathbf{x}) = f_0^{cut} + \sum_{i=1}^{n} f_i^{cut}(x_i) + \sum_{i<j} f_{i,j}^{cut}(x_i, x_j) + \cdots . \quad (2.16)$$

In the cut-HDMR expansion (2.16), a physical meaning is attributed to interactions. Indeed, in many physical and chemical systems, it has been noted that higher order interactions fade away. Rabitz and Alis [1999] write that in many physical phenomena "the first few lowest-order interaction terms are often enough to approximate the output to good accuracy" (p.228) and that "rarely are terms beyond third-order significant" (p.199): these observations constitute an empirical argument in favour of the effect hierarchy, discussed in Section 2.4.

On the other hand, Archer et al. [1997] state that "in sensitivity analysis experiments, where the models are usually nonlinear and the variation in the response much wider, it may happen that the higher order terms are the most important [...] and so their estimation is crucial" (p. 103). Therefore, the choice of the order of the expansion should be motivated by the type of the analysis that a statistician wants to conduct and the informations about interaction she wants to gain from it. The expected dimension of a function Caflisch et al. [1997] (in the superimposition and truncation sense - to be discussed in section 2.5.2 next) is a possible criterion for this choice. Importantly, before drawing any conclusion, the statistician must be aware that the presence of correlated input variables affects the expansion. The research on functional ANOVA representation with dependent input variables is recent and active (Hooker [2007]; Chastaing et al. [2012]; Li and Rabitz [2012]; Rahman [2014]; Li and Rabitz [2017]; Owen and Prieur [2017]). In particular, the orthogonality conditions in (2.15), need to be substituted by conditions of a weaker form: these new conditions are variously called 'weak annihilating' [Rahman, 2014], 'relaxed vanishing' [Li and Rabitz, 2017] and 'hierarchical orthogonal' [Owen and Prieur, 2017]. In any case, the presence of dependences among the inputs gives rise to statistical problems: the functional ANOVA decomposition may be unavailable [Owen and Prieur, 2017]; the terms of the

expansion becomes significantly dependent on the correlation structure of the input variables [Rahman, 2014]; it is not clear how to discern between total and interaction effects [Li and Rabitz, 2017]. However, when input variables have a multivariate normal distribution and the output is linear, the expansion can still be obtained analytically, as the next example shows.

Example 2.5.1. *Consider the function* $f(\boldsymbol{x}) = x_1 + x_2 + x_1x_2$*, where* $\boldsymbol{x}$ *has a multivariate normal distribution with mean* $(0, 0, 0)$ *and covariance matrix*

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

Functional ANOVA decompositions of $f(\boldsymbol{x})$ *are calculated analytically for alternative values of the correlation coefficient* ρ *following the approach of Li and Rabitz [2014]; alternatively, one can use the Fourier-polynomial expansion of Rahman [2014].*

Correlation	f_0	$f_1(x_1)$	$f_2(x_2)$	$f_{1,2}(x_1, x_2)$
$\rho = 0$	0	x_1	x_2	x_1x_2
$\rho = \frac{1}{4}$	$\frac{1}{4}$	$\frac{4}{17}x_1^2 + x_1 - \frac{4}{17}$	$\frac{4}{17}x_2^2 + x_2 - \frac{4}{17}$	$x_1x_2 - \frac{4}{17}(x_1^2 + x_2^2) + \frac{15}{68}$
$\rho = \frac{1}{2}$	$\frac{1}{2}$	$\frac{2}{5}x_1^2 + x_1 - \frac{2}{5}$	$\frac{2}{5}x_2^2 + x_2 - \frac{2}{5}$	$x_1x_2 - \frac{2}{5}(x_1^2 + x_2^2) + \frac{3}{10}$
$\rho = \frac{3}{4}$	$\frac{3}{4}$	$\frac{12}{25}x_1^2 + x_1 - \frac{12}{25}$	$\frac{12}{25}x_2^2 + x_2 - \frac{12}{25}$	$x_1x_2 - \frac{12}{25}(x_1^2 + x_2^2) + \frac{21}{100}$
$\rho = -\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{2}{5}x_1^2 + x_1 + \frac{2}{5}$	$-\frac{2}{5}x_2^2 + x_2 + \frac{2}{5}$	$x_1x_2 + \frac{2}{5}(x_1^2 + x_2^2) - \frac{3}{10}$

Table 2.1: Component functions in presence of independent and correlated input variables.

Table 2.1 shows the four terms of the expansion for every value of the correlation coefficient. Note that the presence of correlated inputs affects main and interaction effects as well, similarly to what happens in

factorial designs. In particular, this confounding between effects highlights that care should be used in interpretation: main effects become quadratic functions of their arguments and the interaction terms might change sign, reverting the model interpretation.

2.5.1 Functional ANOVA regression models: the role of interactions

Low-order ANOVA models (equivalent to cut-HDMR) possess relevant statistical features: they capture non-linearities of real-data effects, have the flexibility of nonparametric modeling, are well interpretable and can overcome the curse of dimensionality in high-dimensional settings. Pioneering works in the statistical literature have started the use of (low-order) functional ANOVA modeling in the context of additive regression, spline models and generalized additive models; see among others, Stone [1985], Hastie and Tibshirani [1990], Chen [1991], Barry [1993].
Stone [1994] considers functional ANOVA models for generalized regression and density estimation, expressing the component functions of the decomposition using tensor product bases [Takemura, 1983]. The key idea is to redefine "functional ANOVA in terms of projections of the function of interest onto spaces of additive functions" [Hooker, 2007, p. 710]: main effects, as functions of one variable, are modeled with arbitrary linear functional spaces and interactions are modeled with their tensor product spaces — see Huang [1998b], Rabitz and Alis [1999], Lin [2000] and Hooker [2007].

Regarding interactions in association with the convergence rate of maximum likelihood estimators, Huang [1998a] writes that "by considering additive models or by allowing interactions involving only two factors, we can get faster rates of convergence than by using the saturated model" (p. 59). However, the tensor product space ANOVA model proposed by Lin [2000] "includes higher-order interactions and still has an optimal rate that is close to the one-dimensional optimal rate" (p. 737).
Wahba et al. [1995] consider the inclusion/exclusion problem of interaction

terms for model selection, on the basis of the model predictive capability. This is in line with the considerations of Shmueli [2010]: "in predictive modeling [...] there is no need to delve into the exact role of each variable in terms of an underlying causal structure" (p. 298).
Overall, one cannot but observe the deep conceptual difference between the selection of interaction terms based on technical aspects such as convergence rate and model selection and the reasoning of Rabitz and Alis [1999], where interactions should be present because of physical considerations about the phenomenon at hand. The implication is that, if the modeler adds or removes interaction terms only on the basis of technical motivations, then interpretation issues about interactions are very likely to arise, especially if one asks whether the interaction appearing in the statistical model is the reverberation of an interaction truly characterizing the phenomenon at hand. We discuss this issue in greater detail in Section 2.7.

2.5.2 Designs for the sensitivity analysis of computer experiments

The rapid and on-going advances in high-fidelity mathematical modeling and computational power have made computer experiments common tools to simulate real-world phenomena. The simulator is often a set of complex mathematical equations embedded in a computer code whose aim is to mimic physic processes producing responses from known input factors. In contrast to physical experiments, deterministic computer experiments produce the same output from the same combination of input factors [Sacks et al., 1989, Wu, 2015]. This makes the notions of blocking, randomization and replication, typical in the DOE become irrelevant. Moreover, in a computer experiment one usually does not have the problem of confounding effects, because the *in silico* experiments can be repeated identically. How about the three principles for factorial designs proposed in Section 2.4? "Effect sparsity can be invoked if there are many input factors and a majority of them are assumed to be inert [...] effect hierar-

chy and heredity principles may be too simplistic to be useful. It is thus a major challenge to formulate new principles that can be used to guide work in computer experiments" (Wu [2015], p.618).
Interactions seem to be another breaking point between physical and computer experiments, even with paradoxical aspects: "in physical experimental design, the variation in the factors is often moderate (due to cost, for instance), and as a result the interaction terms tend to be small [...] In numerical experiments, on the other hand, factors are varied generously over orders of magnitude, and the interaction effects can be very significant and even predominant over the main effects" (Saltelli et al. [1999], p.40). Hence, the possibilities of including in the computer experiment a great number of factors and of augmenting their ranges of variation naturally lead to an increase of the importance of the role of (high-order) interactions. We present here a popular approach to escape this curse of dimensionality in computer experiments
Let's denote the simulator output with $f(\mathbf{x})$, defined as in Section 2.5. Its functional ANOVA decomposition, equation (2.14), is commonly used by computer experimenters to model main effects, two-factors interactions and so on [Sacks et al., 1989]. As we have seen, the functional ANOVA modelling allows to overcome the curse of dimensionality posed by the large number of input factors. Furthermore, a very popular method for global sensitivity analysis of computer experiments is based on this decomposition Sobol' [1969, 1993]. The total variance of $f(\mathbf{x})$ is $V = \int f^2(\mathbf{x})d\mathbf{x} - f_0^2$ while

$$V_{i_1,...,i_s} = \int f^2_{i_1,...,i_s}(x_{i_1},...,x_{i_s})dx_{i_1}\cdots dx_{i_s}, \tag{2.17}$$

with $s = 1,...,n$ denotes the partial variances computed for each term of the functional ANOVA expansion. Then, from (2.14) and (2.17) and by orthogonality, it follows that

$$V = \sum_{i=1}^{n} V_i + \sum_{i<j} V_{i,j} + ... + V_{1,2,...,n}. \tag{2.18}$$

Hence, every $V_{i_1,...,i_s}$ denotes the contribution of the group of variables indexed by $i_1,...,i_s$ to the total variance of the model. Thus, Sobol' sen-

sitivity indeces $S_{i_1,...,i_s}, s = 1, ..., n$ can be defined by

$$S_{i_1,...,i_s} = \frac{V_{i_1,...,i_s}}{V} \tag{2.19}$$

and they rank the variance-based importance of the main and interaction effects. These indices are comparable as they are standardized, i.e., $\sum_{i=1}^{n} S_i + \sum_{i<j} S_{i,j} + ... + S_{1,2,...,n} = 1$. This approach is a functional analogous to the classical ANOVA of factorial experiment designs, where the variance is decomposed into a sequence of terms of increasing dimensions (see e.g. Scheffé [1959]).
It is possible to define related quantities of interest for a complete interaction analysis: the total sensitivity index for input i [Homma and Saltelli, 1996] is defined by

$$S_i^T = S_i + \sum_{j=1, j\neq i}^{n} S_{i,j} + ... + S_{1,2,...,n} \tag{2.20}$$

and accounts for the total contribution of the input i, including its main effect and its possible high-order interactions. One an then define the indexes $S_i^I = S_i^T - S_i$ as overall measures of the importance of the interactions in which factor X_i is involved. Note that a similar concept can be found in the epidemiological literature, as we discuss in Section 2.7.

Variance-based sensitivity indices are at the basis of the notion of dimension distribution of a function. Because the indices are positive and sum to unity, they can be regarded as a probability mass function Caflisch et al. [1997], Owen [2003]. In particular, letting U a random variable with support 2^Z, then one defines $\Pr(U = z) = S_z$. Then, one calls the distribution of the cardinality of U the dimension distribution of the simulator in the superimposition sense; and one calls the distribution of $\max\{j : j \in z\}$ dimension distribution of the simulator in the truncation sense. Caflisch et al. [1997], Owen [2003] then define the mean effective dimensions of a function in the superimposition and truncation sense, as [Caflisch et al.,

1997, Owen, 2003]:

$$D_{Superimp} = \sum_{|z|>0} |z| \Pr(U = z) = \sum_{i=1}^{n} S_i^T, \tag{2.21}$$

$$D_{Trunc} = \sum_{|z|>0} \max\{j : j \in z\} \Pr(U = z), \tag{2.22}$$

One observes that the mean dimension in the superimposition sense coincides with sum of all total sensitivity indices related to X_i [Owen, 2003]. The effective dimensions $D_{Superimp}$ and D_{Trunc} take values between 1 and n, with the minimum registered for additive models. Thus, the higher their values, the higher the relevance and order of the interactions in the model. Therefore, the effective dimensions provide an indication of the size of the relevant interactions and have been exploited in works related to dimension reduction in high-dimensional integration, such as such as [Wang, 2006, Wang and Sloan, 2011].

Interestingly, in computer experiments the analyst has also the possibility of implementing screening designs to select the most influential inputs and their interactions. These screening designs include the winding stairs [Chan et al., 2000] and the cell-based designs [Saltelli et al., 2010]. Note that this screening operation is not typical in the physical experiments literature, where fewer factors are under study — nonetheless, one can find exceptions such as Cotter [1979] and Lewis and Dean [2001]. However, the underlying ideas are taken from DOE literature and adapted in the framework of screening experiments. In fact, as far as the selection of points in the model input space is concerned, any design applicable for *in the field* experiments can be adopted in a computer experiments. For instance, we note that the cell-based estimator of the screening design in Saltelli et al. [2010] for the second order interaction effect of model input i is the interaction effect $\mathrm{INT}(A, B)$ defined for DOE in equation (2.11), where in this context A and B denote appropriate changes in the coordinates involving factor i (see Becker and Saltelli [2015] for additional details).

In summary, computer experiments represent an important intersection of methodologies from DOE and from functional ANOVA, and strategies

to identify interactions benefit from the interplay between these two vast subjects.

2.6 Modern approaches: interaction identification in high-dimensional settings

High-dimensionality of data is a significant feature in several branches of modern scientific research, ranging from imaging to genetics, to network analysis. In the recent years, the interest in identifying interactions has rapidly grown due to the "importance of interactions in statistical inference and contemporary scientific discoveries" (Fan et al. [2015], p. 1243). Finding interactions provides the researches with deep insights on the complexity that underlies scientific models. In this respect, Hao and Zhang [2014] write that "interaction models provide a better approximation to the response surface, improve prediction accuracy, and bring new insight on the interplay between predictors" (p. 1285). For example, in genome-wide association studies, the problem of discovering gene-gene [Wang et al., 2015, Xia et al., 2015] and gene-environment [Ma et al., 2015, He et al., 2017] interactions is receiving much attention by researchers for the potential of unveiling important mechanisms with implications for human diseases. However, in Section 2.7 we discuss the limitations of the statistical regression analysis for causal interactions in epdidemiology.

In high-dimensional regression models, statisticians face "unprecedented challenges in identifying important interactions" (Kong et al. [2017], p. 897): the dimensionality of the problem rapidly grows considering the large number of possible interactions, there might be irregularities in the interaction pattern, memory requirement and computational cost are prohibitive, noise accumulates while estimating the large number of parameters and models become unidentifiable [Fan and Song, 2010, Hao and Zhang, 2014].

To illustrate these difficulties, consider the linear two-way interaction re-

gression model

$$Y = \beta_0 + \sum_{i=1}^{p} \beta_i X_i + \sum_{i<j} \beta_{i,j} X_i X_j + \epsilon, \tag{2.23}$$

where $\beta_0, \beta_1, ..., \beta_p$ and $\beta_{1,2}, \beta_{1,3}, ..., \beta_{p-1,p}$ are unknown regression coefficients[1].

The model in (2.23) contains p main effects and $\binom{p}{2}$ interaction terms. Given a set of n i.i.d. observations $\{(y_i, x_i)\}_{i=1}^n$, if p is small compared to the sample size n, common statistical procedures, such as ordinary least squares, allow one to estimate the model parameters.

On the other hand, when p is large, the number of possible two-way interactions grows at an order $O(p^2)$ and the ordinary least squares estimate become difficult to apply (Wang and Leng [2016],p. 589). To illustrate, consider that a common gene expression dataset [Simon and Tibshirani, 2015] involves $p \sim 20000$ genes (the covariates) and $n \sim 100$ patients (the observations): in this case, the number of two-way interaction parameters in model (2.23) is approximately 200 millions. To perform statistical analyses in a 'large p, small n' situation, it is commonly assumed that only a subset of variables actively contributes to the response. This principle is equivalent to the *effect sparsity* in the design of experiments and in this context assumes that only a relatively small number of regression coefficients are significantly different from zero. As a consequence, high-dimensional statistical modeling can be improved by performing variable selection [Fan and Song, 2010].

In practice, however, standard variable selection techniques, such as the Lasso [Tibshirani, 1996], treat the candidate variables "individually or flatly" (Choi et al. [2010], p. 354), ignoring the relationship between main and interaction effects. Indeed, in fitting the model in (2.23), it is often of interest for the statistician to obtain a final model which satisfies certain constraints among variables. For instance, it is common practice to

[1] Hao and Zhang [2014] and Hao and Zhang [2017] consider another definition of linear two-way interaction models: they add to (2.23) the quadratic terms $\sum_{i=1}^{p} \beta_{i,i} X_i^2$ and refer to all the degree-two terms X_i^2 and $X_i X_j$ as interactions.

include an interaction term in the model if the corresponding main effects are in the model as well. This principle is known in the regression modeling context also as *marginality* [Nelder, 1977] or *hierarchy* [Peixoto, 1987]. We observe that this principle is analogous to the notion of *effect heredity* in DOE [Hamada and Wu, 1991]. There are two types of such restrictions, called the *strong* and *weak* heredity principles [Chipman, 1996]. Precisely, the strong heredity principle foresees that, if an interaction term is included in the model, then also its corresponding main effects must be included, that is

$$\beta_{i,j} \neq 0 \quad \Longrightarrow \quad \beta_i \neq 0 \quad \text{and} \quad \beta_j \neq 0.$$

The weak heredity principle foresees that an interaction term is taken into account only if at least one of its associated main effects is included in the model, that is

$$\beta_{i,j} \neq 0 \quad \Longrightarrow \quad \beta_i \neq 0 \quad \text{or} \quad \beta_j \neq 0.$$

However, the adequacy of heredity principles in variable selection is the subject of an intense debate in the literature. Some arguments in their favor are technical: "some statisticians argue that models violating strong hierarchy are not sensible [...] violating strong hierarchy amounts to postulating a special position for the origin" (Bien et al. [2013] p.1112), and practical: "a generic variable selection method may select an interaction term but not the corresponding main terms, and such models are difficult to interpret in practice" (Choi et al. [2010], p.354). On the other hand, Kong et al. [2017] observe that "the strong or weak heredity assumption [...] may not be satisfied in certain real applications" (p. 898).

In general, Hao and Zhang [2014] classify approaches to maintain an heredity structure in the final model into *joint* and *two-stage* procedures. Joint methods select simultaneously main and interaction effects and search globally over all candidate models. This class includes, among others, the SHIM [Choi et al., 2010], HL [Bien et al., 2013], FAMILY [Haris et al., 2016] and GRESH [She et al., 2016] methods. These are regularization-based methods in which special penalty constraints are introduced to respect the heredity principle [Radchenko and James, 2010].

In ultra-high dimensional setting (very large p) the joint analysis may become infeasible due to "memory requirement and computational cost" (Hao and Zhang [2014], p. 1286) and "simultaneous challenges of computational expediency, statistical accuracy and algorithmic stability"(Fan et al. [2009], p.2014). Hence, this ultra-high dimensional settings are regarded as the bottleneck for existing regularization methods: "when the space of searched interactions is very large, it is much easier to find false positives" (Lim and Hastie [2015], p.649). Two-stage methods are then receiving increasing attention in the statistical analysis of genomics ultra-high datasets [Wu et al., 2009, 2010]. The structure of a two-step method is typically as follows: in the first step one performs a screening exercise to identify strong individual effects, in the second step one performs the joint analysis of the effects remaining after the screening. As Kong et al. [2017] underlines, it is not necessary to assume heredity in a two-step procedure. The iFOR [Hao and Zhang, 2014], IPDC [Kong et al., 2017] and RAMP [Hao et al., 2018] are representatives of this class of methods.

In identifying interactions in gene-environment studies, researchers may find nonlinear features of interactions, i.e., the effects of genetic factors can be altered by environmental factors in a nonlinear way. There is growing interest in models that capture this nonlinear behavior: "in the high-dimensional data setting, studying nonlinear interaction effects has found much attention in recent years, and a few strategies have been proposed" [Ma et al., 2015, p. 2104]. A nonlinear two-way interaction model can be written as

$$Y = \beta_0 + \sum_{i=1}^{p} f_i(X_i) + \sum_{i<j} f_{i,j}(X_i, X_j) + \epsilon, \tag{2.24}$$

and can be seen as the stochastic analogous of the second-order functional ANOVA decomposition (2.16). To estimate and to perform inference for the nonparametric component functions in a high-dimensional setting, functional ANOVA modeling must be combined with some variable selection techniques, in analogy to the standard linear regression case. Recent methods developed for studying nonlinear interactions in high-dimensional datasets include VANISH [Radchenko and James, 2010], VICM [Ma and

Song, 2015], high-dimensional GACM [Ma et al., 2015], PLVMICM [Liu et al., 2016] and GPLVMICM [Liu et al., 2017].

2.7 Interactions and causal inference in epidemiology

In the previous sections we have provided a comprehensive overview to interactions with a focus on issues emerging in their definitions and in modeling strategies. However, an important scientific aspect concerning interactions for the Statistician, is whether a significant interaction implies a corresponding real physical or biological mechanism. This inference, as underlined in VanderWeele and Knol [2014], leaves no room to a simplistic intepretation. While the discussion that follows is taken from papers appeared mostly in the epidemiological literature [Greenland et al., 2008, VanderWeele and Knol, 2014, VanderWeele, 2015], causal interactions have been studied also in the context of high-dimensional data [Imai and Ratkovic, 2013] and are relevant also for investigations in genetics, social sciences, industry as well.

2.7.1 Inference for statistical interactions

Consider a binary outcome Y that depends on the effects of two binary exposures G and E. As a reference, let us consider them a genetic factor and an environmental factor respectively. When the effect of one exposure depends in some way on the other, than the exposures are said to interact. A first way to measure such interaction is to investigate whether the joint effect of both factors exceeds the sum of the individual effects. Let $p_{ge} = P(Y = 1|G = g, E = e)$ be the probability of the outcome when G is at the value g and E at e. Note that these probabilities represent the 'risk' of the (epidemiological) outcome and, hence, are not assumed to sum to one, as it happens in contingency tables (Section 2.2). Then, a measure

of the interaction on the additive scale is given by [VanderWeele, 2015]

$$(p_{11} - p_{00}) - [(p_{10} - p_{00}) + (p_{01} - p_{00})] = p_{11} - p_{10} - p_{01} + p_{00}, \quad (2.25)$$

where $(p_{11} - p_{00})$ can be seen as the joint effect of the factors and the terms $(p_{10} - p_{00})$ and $(p_{01} - p_{00})$ represent the individual factor effects. When expression (2.25) is positive, then the interaction is said to be positive, negative otherwise. Additionally, denoting with p_e the probability $P(Y = 1|E = e)$, for independent exposures G and E one obtains

$$(p_{e=1} - p_{e=0}) = (p_{01} - p_{00}) + (p_{11} - p_{10} - p_{01} + p_{00})P(G = 1). \quad (2.26)$$

This expression allows one to define the quantity $pAI_{G=0}(E)$ as [VanderWeele and Knol, 2014]

$$pAI_{G=0}(E) = \frac{(p_{11} - p_{10} - p_{01} + p_{00})P(G = 1)}{(p_{e=1} - p_{e=0})}. \quad (2.27)$$

This quantity represents the proportion of the overall effect of E which is attributable to interaction, with reference category $G = 0$. In analogy to the total interaction index in the sensitivity analysis of computer experiments, this measures aims at quantifying the total impact of the interaction.

Typically, interaction on the additive scale is evaluated using a linear statistical model of the form

$$P(Y = 1|G = g, E = e) = \alpha_0 + \alpha_1 g + \alpha_2 e + \alpha_{12} eg, \quad (2.28)$$

where $\alpha_0 = p_{00}$, $\alpha_1 = p_{10} - p_{00}$, $\alpha_2 = p_{01} - p_{00}$ and $\alpha_{12} = p_{11} - p_{10} - p_{01} + p_{00}$. The coefficient α_{12} is called the statistical interaction on the additive scale. This coefficient assumes an important role in supporting public health policy making [Blot and Day, 1979, Berrington de González and Cox, 2007]. If if $\alpha_{12} > 0$, then the consequence of a public intervention on the factor E would be larger in the $G = 1$ group, larger in the $G = 0$ group in the other case of negative interaction $\alpha_{12} < 0$.

Another approach to assess interaction effects is to use measures defined on the multiplicative scale. One considers the risk ratios

$$RR_{10} = p_{10}/p_{00}, \quad RR_{01} = p_{01}/p_{00}, \quad RR_{11} = p_{11}/p_{00}.$$

Then, one can define an interaction measure on the multiplicative scale as [VanderWeele, 2015]

$$\frac{RR_{11}}{RR_{01}RR_{10}} = \frac{p_{11}p_{00}}{p_{01}p_{10}}. \tag{2.29}$$

Note that, if $p_{11}p_{00} = p_{01}p_{10}$, then there is no multiplicative interaction (for the risk ratios): this condition is analogous to the definition of independence for 2×2 contingency tables [Bartlett, 1935]. If $p_{11}p_{00}/p_{01}p_{10} > 1$, then the multiplicative interaction is said to be positive, otherwise negative.

In practice, interactions on the multiplicative scale are evaluated using log-linear models for the risk ratios with the product term

$$\log P(Y = 1|G = g, E = e) = \beta_0 + \beta_1 g + \beta_2 e + \beta_{12} eg, \tag{2.30}$$

where $e^{\beta_0} = p_{00}$, $e^{\beta_1} = RR_{10}$, $e^{\beta_2} = RR_{01}$ and $e^{\beta_{12}} = RR_{11}/RR_{10}RR_{01}$. Hence, β_{12} is the statistical interaction term for the log-linear model.

On the multiplicative scale a further interaction measure is constructed as follows. Let's define the odds ratios

$$OR_{01} = \frac{p_{01}/(1-p_{01})}{p_{00}/(1-p_{00})}, \quad OR_{10} = \frac{p_{10}/(1-p_{10})}{p_{00}/(1-p_{00})}, \quad OR_{01} = \frac{p_{11}/(1-p_{11})}{p_{00}/(1-p_{00})}.$$

The interaction measure on the multiplicative scale for the odds ratios is defined as

$$\frac{OR_{11}}{OR_{01}OR_{10}}. \tag{2.31}$$

When $OR_{11}/OR_{01}OR_{10} > 1 (< 1)$, then the multiplicative interaction for the odds ratios is said to be positive (negative).

In case-control studies, the multiplicative interaction for the odds ratios is estimated using logistic model with the product term

$$\text{logit} P(Y = 1|G = g, E = e) = \gamma_0 + \gamma_1 g + \gamma_2 e + \gamma_{12} eg, \tag{2.32}$$

where $e^{\gamma_0} = p_{00}/(1-p_{00})$, $e^{\gamma_1} = OR_{10}$, $e^{\gamma_2} = OR_{01}$ and $e^{\gamma_{12}} = OR_{11}/(OR_{01}OR_{10})$. The coefficient γ_{12} is the statistical interaction in the logistic model and measures the multiplicative interaction on the odds ratio scale. Multiplicative interaction models (log-linear and logistic) are usually preferred since

they are easier to fit and are implemented in common statistical softwares. [VanderWeele and Knol, 2014].
In general, as it happens for the analysis of contingency tables (see Section 2.2), several reasons can be given in favour of the use of the additive or of the multiplicative scale. However, the fact that an interaction is significant, positive or negative, depends on the scale and results might be different on different scales [Berrington de González and Cox, 2007]. In fact, Greenland et al. [2008] prove that there must be an interaction on some scale whenever both exposures have an effect on the outcome: the only possible way not to have an interaction on both scales is that one exposure does not have an effect on the outcome. This point "raises the question of why interaction is of interest and which scale is to be preferred. It also makes clear that just to say that there is an interaction on some scale is relatively uninteresting [...] provided both exposures have an effect on the outcome, such interaction on some scale will always be present" (VanderWeele and Knol [2014], p.37-45). Therefore, VanderWeele and Knol [2014] and VanderWeele [2015] suggest that researchers should provide interaction measures on both scales, since both analyses can be informative and, in some sense, complementary.

2.7.2 Mechanistic forms of interactions

Clayton [2009] writes that "recent interest in interaction in genetics has also been characterized by exaggerated expectations for the inferences that can be drawn from epidemiological data [...] little can be deduced about mechanism from the observation of statistical interaction" (p.5). Hence, it is of practical need the possibility of drawing conclusion on more 'mechanistic' forms of interactions rather than 'pure' statistical ones.
A first necessary assumption to infer causality between the effects is that they are unconfouded. Without such requirement, one can not be sure whether the regression coefficient for interaction should be interpreted as a measure for 'effect modification' or for 'causal interaction' [VanderWeele and Knol, 2014]. Denote by Y_{eg} the outcome that would have occurred

for every subject if $G = g$ and $E = e$ had been fixed (this is called the counterfactual outcome). Then, a *sufficient cause interaction* is present if for some individual it holds $Y_{11} = 1$ with $Y_{01} = Y_{10} = 0$ [VanderWeele and Robins, 2008]. Hence, the effects interact in a 'mechanical' way since their joint presence is required in order to activate the outcome. Sufficient cause interactions can be estimated from data: when one can assume that Y_{eg} is monotonic (e.g., Y_{eg} is non-decreasing in e and in g for every subject) and factors are unconfounded, then $p_{11} - p_{10} - p_{01} + p_{00} > 0$ implies a sufficient cause interaction [Greenland et al., 2008]. The difficulty in this case is represented by the additional assumption of the monotonicity of effects, which can't be verified empirically and "must be established on substantive grounds" (VanderWeele and Knol [2014], p.51). However, the stronger condition $p_{11} - p_{10} - p_{01} > 0$ implies the presence of a sufficient cause interaction with the sole assumption that the effects are unconfounded [VanderWeele and Robins, 2008].

VanderWeele [2010] introduces a stronger notion of interaction. An *epistatic* interaction is said to be present if there is a subject for whom Y_{11} but $Y_{01} = Y_{10} = Y_{00} = 0$; that is, the outcome is activated if and only if both exposures are present. To empirically test for such epistatic interaction, VanderWeele [2010] proves that the condition $p_{11} - p_{10} - p_{01} - p_{00} > 0$ implies the presence of epistatic interaction, provided that effects are unconfounded. Note that this condition is stronger than that for the sufficient cause interaction (the baseline probability p_{00} is subtracted and not added). Additionally, if the effect of at least one exposure is monotonic, then $p_{11} - p_{10} - p_{01} > 0$ suffices to test for an epistatic interaction; if both effects are positive and monotonic, then $p_{11} - p_{10} - p_{01} + p_{00} > 0$ suffices for an epistatic interaction [VanderWeele, 2010, 2015]. Note that all these conditions are only sufficient, i.e. if both conditions are satisfied, then an epistatic interaction is present. However, if they are not satisfied, an epistatic interaction might or might not be present: "one simply cannot determine this from the data" (VanderWeele [2015], p.298).

However, the presence of mechanistic interactions (sufficient cause and epistatic) still does not imply a real physical or biological underlying mech-

anism. Indeed, the statistical tests of the defining conditions don't give information about the underlying biology: "From the assessment of statistical interaction between risk factors it is tempting to infer the nature of the biologic interaction between the factors. However, the use of statistical analyses of epidemiologic data to infer biologic processes can be misleading" (Siemiatycki and Thomas [1981], p.383).

2.7.3 The debate in brief

The statical analysis of interactions in complex biological systems has been a controversial and debated point in the epidemiological literature (see the discussion in Walter and Holford [1978], Blot and Day [1979], Rothman et al. [1980], Siemiatycki and Thomas [1981], Thompson [1991], Clayton [2009]). Several authors adopt the following distinction [Phillips, 2008, VanderWeele, 2015]

1. *statistical interactions* are interactions due to the presence of product terms in statistical regression models;

2. *mechanistic interactions* are interactions emerging when the outcome is activated by the joint presence of the exposures;

3. *biological or functional interactions*: are interactions emerging when two exposures physically interact to bring about the outcome.

Therefore, even if in some situations we can estimate and draw conclusions about mechanistic interactions, this analysis does not allow one to make definitive conclusions about physical interactions in biological or social mechanisms. In this regard, [Clayton, 2009, p. 4] writes that "statistical significant results are often heralded as significant in a wider sense". Some examples of such limitations can be found in Siemiatycki and Thomas [1981], who show using the multistage model for carcinogenesis that "even if carcinogenic factors act independently, some pairs may fit an additive statistical model, some a multiplicative statistical model, and some neither. The elucidation of biological interactions by means of statistical

models requires the imaginative and prudent use of inductive and deductive reasoning; it cannot be done mechanically" [Siemiatycki and Thomas, 1981, p. 383]. In such case, we note that the interaction is neither biological nor statistical: it is only mechanicistic.

2.8 Conclusions

This work has offered a tour into the world of interactions as analyzed in the statistical literature. In statistical modelling one can find several dissimilar notions of 'interaction' and, hence, the term 'interaction' does not have *per se* a univocal meaning. For instance, in design of experiments interactions are mainly regarded *à la Cox*, i.e., as deviations from additivity, while in the analysis of contingency tables interactions are associated with the presence of statistical dependences.

Specific features and technical aspects create alternative interpretations of the term and the techniques used to determine interactions differ substantially depending on the application. For instance, determining interactions in a computer experiment is notably different than analyzing them in a biological context. However, there is an underlying fil rouge and technical differences might not be conceptual. For instance, we have noticed that the standard procedure for assessing interactions in contingency tables (Section 2.2) is the same for risk factors interaction in epidemiological studies (Section 2.7), although there is little overlap between these two fields in the literature. Similarly, statistical regression models (Section 2.6) and functional cut-ANOVA models (Section 2.5) have a common mathematical structure, although the corresponding investigations develop mostly independently.

A transversal theme is the challenge in interpreting interactions. Inferring causality from the presence of interactions, for instance, is a major conceptual issue. We have seen that their interpretation is subject of debate and careful scientific discussion, because it is the choice of the model that, in some instances, can determine the presence or absence of interactions. The difficulty in choosing the right model is not confined

to epidemiological applications, but is present in functional ANOVA and statistical regression analysis as well, where the statistician has to make a decision about what terms to include in the model. In high dimensional settings, the selection is often guided by hierarchy principles. In computer experiments, one finds the notion of dimension distribution that provides a quantitative measure. At the same time, one assists to applications, such as unbalanced ANOVA tables, in which the converse happens. One determines interactions first, and uses them to decide the correction method.

What to do? The work reveals that studying interactions is a long standing task in statistical modelling, and its importance is not decreased by modern applications. The task is a delicate one, and care must be undertaken whey studying interactions, as VanderWeele and Knol [2014] well capture: "when studying interaction, it is important to clearly understand what the goal of the analysis is: What is that we are trying to learn? What scientific or policy question are we trying to answer and how does an interaction analysis help us?" (p.45). In this respect, this work, by addressing issues and conceptualizations of interactions across alternative statistical applications should be of help to the analyst in identifying the frame of the analysis at hand.

Appendix

2.9 Lancaster's Additive Interactions

In multivariate data analysis, it is of great importance to obtain measures of multivariate dependence among the components of a random vector [Joe, 1993]. Citing Ip et al. [2004], "given a distribution of multivariate variables, interactions are quantities that partition the total departure from stochastic independence" (p.120). Additive interactions among random variables are a powerful tool to describe their dependence structure. They are a representation of the multivariate joint distribution meant to show clearly the dependencies between the elements of a random vector [Streitberg, 1999, p. 408]. The underlying intuition dates

back to the seminal work of Lancaster [1969]. In particular, Lancaster [1969] characterizes the absence of additive interactions among the components of a random vector. Let $X = (X_1, ..., X_n) \in \mathbb{R}^n$ be a random vector with joint distribution function $F = F_{1,...,n}$ and marginal distributions $F_1, ..., F_n, F_{1,2}, ..., F_{n-1,n}, ..., F_{2,...,n}$. Then, X is said to contain no Lancaster-additive interaction of order $s-1$ if and only if

$$\prod_{i \in A} (F_i^* - F_i) = 0, \tag{2.33}$$

where A are the subsets of the indexes $\{1, ..., n\}$ with s elements. In equation (2.33), a product of the type $F_i^* F_j^* F_k^*$ denotes the distribution function $F_{i,j,k}$. For example, consider the case $n = 3$: the term $F_1 F_2^* F_3^*$ in the expansion $(F_1^* - F_1)(F_2^* - F_2)(F_3^* - F_3)$ denotes $F_1 F_{2,3}$. Hence, F does not contain any Lancaster-additive second order interaction if and only if

$$F_{1,2,3} = F_1 F_{2,3} + F_2 F_{1,3} + F_3 F_{1,2} - 2F_1 F_2 F_3. \tag{2.34}$$

The above definition generalizes the classical probabilistic definition of mutual independence, that is $F_{1,2,...,n} = F_1 F_2 \cdots F_n$. Hence, the presence of additive interactions reflects the more realistic situations in which not all the variables can be assumed independent: this fact explains the connection between additive interactions and the decomposition of the joint distribution $F_{1,2,...,n}$ according to the dependence structure. Conversely, it is possible to recover the joint distribution $F_{1,2,...,n}$ in terms of its marginal distributions under the assumption that all the interactions higher than a certain order vanish [Zentgraf, 1975].

Lancaster's approach has developed into the theory of additive interaction measures, that has been established in the works of Lancaster [1969],Lancaster [1971], Streitberg [1990, 1999]. Formally, an additive interaction measure ΔF is a signed measure that vanishes whenever F can be factorized as the nontrivial product of two of its multivariate marginal distributions [Streitberg, 1990]. In this case, F is called decomposable. For example, if $n = 10$ and $F_{1,2,...,10}$ can be factorized as $F_1 \cdot F_{2,...,10}$, then $\Delta F = 0$ and F is decomposable.

In the early development of the theory, the additive interaction measure was defined with Lancaster's representation as

$$\Delta^{(L)}F = \prod_{i=1}^{n}(F_i^* - F_i). \tag{2.35}$$

We call this the Lancaster representation of the additive interaction measure ΔF. Later, however, Streitberg [1990] provides a counterexample for the case $n = 4$, showing that $\Delta^{(L)}F$ does not vanish in presence of two mutually independent sub-vectors. Using the Moebius function on finite lattices, Streitberg [1990] proves the following representation of an additive interaction measure;

$$\Delta^{(S)}F = \sum_{\pi}(-1)^{|\pi|-1}(|\pi| - 1)!F_{\pi}, \tag{2.36}$$

where the sum is taken over all partitions π of the index set $\{1, ..., n\}$. The representations $\Delta^{(L)}$ and $\Delta^{(S)}$ coincide for $n < 4$. Ip et al. [2004] investigate the relationship between the interaction measures of Lancaster (2.35) and of Streitberg (2.36) further. Among other findings, they show that multivariate joint cumulants $k(X)$ can be written as the signed integral with respect to Streitberg's additive interaction measure as

$$k(X) = (-1)^n \int \Delta^{(S)}F(x)dx. \tag{2.37}$$

In the statistical literature important applications of additive interaction measures can be found, which illustrate the flexibility of these models. Rodríguez and Bárdossy [2014] use relation (2.37) for model building with application to spatial data. Lancaster's approach has been recently used for modelling dependence in Markov exchangeable processes [Di Cecco, 2009], for testing non-parametrically the dependence in graphical models [Sejdinovic et al., 2013] and for studying compatibility among marginal densities [Wang, 2004]. In general, the well-known application of Lancaster-additive models is the interaction quantification in contingency tables, discussed in Section 2.2.

Chapter 3

Interactions and Computer Experiments: What's in the black-box?

Abstract

Studying interactions is crucial in understanding the structure of a simulator and, more in general, of a statistical model. However, the variety of interpretations and available techniques, as well as computational issues make the analysis of interactions non-obvious. We base our investigation on Cox's interpretation of interactions as deviations from additivity. We propose formal definitions for different interaction types, and obtain a series of novel results that investigate the properties of interactions measures at the infinitesimal, finite and global scales. Several insights, some of which not immediately intuitive, are discussed through counterexamples and realistic case studies, with the goal of aiding the analyst in the difficult task of making the model box less obscure.

3.1 Introduction

The analysis of interactions is a significant task in statistics and is particularly important in computer experiments. Knowing which interactions are relevant helps the analyst to shed light on the internal structure of the simulator, thus opening the simulator black box and aiding interpretability. Also, it helps analysts in deciding which terms should appear while building an emulator. Nonetheless, several distinct elements concur in an interaction analysis, making it a challenging task.

As a first challenge, the analyst finds a variety of interpretations in the statistical literature. In the Lancaster-Streitberg representation [Lancaster, 1971, Streitberg, 1990], interaction means the absence of statistical dependence. In Cox's interpretation [Cox, 1984], interaction means deviation from additivity. Within Cox's interpretation, one finds *spurious* interactions [Friedman and Popescu, 2008], *removable* interactions [Berrington de González and Cox, 2007], and *context-specific* interactions, in diverse contexts such as statistical machine learning, design of experiments and the analysis of contingency tables, respectively. This raises the question of whether there are common interaction generating mechanisms across these specifications. At the same time, the literature offers a variety of techniques to analyze interactions, from mixed partial derivatives [Fruth et al., 2014] to the bilinear coefficients of a multilinear response surface [Zhou and Xu, 2017, p. 1675]. Then, one ought to ask whether the method used in the investigation is, indeed, the most appropriate for the interaction type at hand. In fact, alternative approaches deliver complementary rather than equivalent insights. Thus, ideally, the analyst should apply a combination of methods. However, determining interactions is a numerically demanding effort and computational burden is a hidden player of the game.

The purpose of this Chapter is to offer a systematic investigation into these issues. We address the question of defining interactions and providing a common theoretical background. This formalization allows one to obtain a unified view of the types of interactions discussed before in the literature,

as well as to formalize the link between interactions, transformations and the nature of the input-output mapping. We start proposing a definition that formalizes Cox's notion of interactions as deviations from additivity [Cox, 1984], and we enrich this definition with a geometric interpretation that links the discrete Laplace operator with the levels of a factorial design and with results on additively separable mappings. At the infinitesimal scale, we obtain a formal bridge between [Berrington de González and Cox, 2007]'s removable interactions and an early result by [Scheffé, 1959] providing necessary and sufficient conditions for an interaction to be removable. We show that interactions due to piecewise-definiteness (a class defined here that encompasses context-specific interactions) are never removable. We provide a bridge between interactions typical of factorial experiments (on a finite scale) and global sensitivity measures. Under model input dependence, we connect the notion of spurious interactions of Friedman and Popescu [2008] determination of interactions through the generalized functional ANOVA expansion [Hooker, 2007, Li et al., 2010].

We conclude the Chapter investigating the implications of the previous methodological findings in applications. We rely on a set of realistic simulators taken from previous literature, that are representative of alternative input-output structures and dimensions. The numerical experiments substantiate the insights of the methodological findings.

3.2 Interactions á la Cox

Let $\mathcal{X} \subseteq \mathbb{R}^n$, $\mathbf{x} \in \mathcal{X}$, $g : \mathcal{X} \to \mathbb{R}$, and

$$y = g(\mathbf{x}) + \epsilon \tag{3.1}$$

denote a generic multivariate mapping and ϵ a stochastic error term, with zero mean. The simulator is called stochastic if the error term is present, deterministic otherwise.

Interactions are interpreted as deviations from additivity. Quoting from [Koehler and Owen, 1996, p. 262]: *Increasing x_1 may be an improvement and increasing x_2 may be an improvement, but increasing them both to-*

gether might make things worse. This would usually be determined from a confirmation run in which both x_1 and x_2 have been increased. Two aspects become relevant: whether the overall effect of the changes is or not the sum of the individual effects and, if not, whether interactions are synergistic or antagonistic.

To formalize this intuition we proceed as follows. Let $\mathbf{x}^0 = (x_1^0, x_2^0, \ldots, x_n^0)$ and $\mathbf{x}^1 = (x_1^1, x_2^1, \ldots, x_n^1)$ denote any two points in $\mathcal{X}$ differing in at least in two components. For the sake of notation simplicity, consider a deterministic mapping for the moment and let $\Delta g = g(\mathbf{x}^1) - g(\mathbf{x}^0)$. Also, let $Z = \{1, 2, \ldots, n\}$ and i_k denote a generic index, let $z = \{i_1, i_2, \ldots, i_k\}$, $z \subseteq Z$, denote a subset of indices and $-z = Z \backslash z$ its complement. Let $\overline{\mathbf{x}}^z = (\mathbf{x}_z^1 : \mathbf{x}_{-z}^0)$ denote the point in $\mathcal{X}$ obtained by considering the variates with indices in z at level 1 and the remaining at level 0. Then $\Delta_z g = g(\overline{\mathbf{x}}^z) - g(\mathbf{x}^0)$ and $\Delta_{-z} g = g(\overline{\mathbf{x}}^{-z}) - g(\mathbf{x}^0)$ denote the changes in g due to the shift of variables with indices in z and without indices in z, respectively.

Definition 3.2.1. *Let $g : \mathcal{X} \to \mathbb{R}$, and consider the points defined above. We say that g is additive on $\mathcal{X}$ if for all changes $\mathbf{x}^0 \to \mathbf{x}^1$ with $\mathbf{x}^0, \mathbf{x}^1 \in \mathcal{X}$*

$$\Delta g = g(\mathbf{x}^1) - g(\mathbf{x}^0) = \sum_{i=1}^{n} \Delta_i g, \tag{3.2}$$

where $\Delta_i g = g(\overline{\mathbf{x}}^i) - g(\mathbf{x}^0) = g(x_i^1 : \mathbf{x}_{-i}^0) - g(\mathbf{x}^0)$.

That is, g is additive if the effect of the change $\mathbf{x}^0 \to \mathbf{x}^1$ on g is the sum of the individual changes provoked by each x_i for all changes $\mathbf{x}^0 \to \mathbf{x}^1$.

Alternatively, in mathematical analysis, one finds the definition of additive mapping as a mapping that satisfies:

$$g = \sum_{i=1}^{n} g_i(x_i). \tag{3.3}$$

In Supplementary Appendix B we prove that the two definitions are equivalent. (Such Appendix also contains all proofs for this chapter).

Proposition 3.2.1. A mapping $g : \mathcal{X} \mapsto \mathbb{R}$, $\mathcal{X} \in \mathbb{R}^n$ satisfies equation (3.2) if and only if it satisfies equation (3.3).

Thus, an analyst is dealing with an interaction whenever the form of a simulator is not compliant with (3.3). This intuition is often used in statistical analysis. For instance, an analyst may be studying the input-output dataset using a linear emulator, e.g., a regression curve of the form $y \simeq \sum_{i=1}^{n} \beta_i x_i$. If the regression fit (evaluated by appropriate measures of statistical significance) is low, the analyst has several options available to improve it. The classical first attempt is the inclusion of pairwise interactions terms in the emulator. For instance, she might resort to the emulator [Zhou and Xu, 2017]:

$$y \simeq \sum_{i=1}^{n} \beta_i x_i + \sum_{i<j=1}^{n} \beta_{i,j} x_i x_j. \tag{3.4}$$

The terms $\beta_{i,j} x_i x_j$ in (3.4) are typically called interaction terms. If they are statistically significant, then one concludes that interactions are present in the simulator response. Indeed, (3.4) is not reconcilable with (3.3), due to the interaction terms. For clarity, interaction terms in general are not confined to a linearly multiplicative form; for example a term of the type $\sin(x_i + x_j)$ makes the mapping non-additive in the inputs. However, interaction terms are not the only interaction generating mechanism: as we are to see, interactions may appear also when the simulator is locally the sum of univariate functions. To address this aspect, we need the notion of piecewise-defined function. Piecewise-defined mappings appear frequently in statistical studies, as well as in applications of computer simulations. Kim et al. [2005] study Gaussian processes with piecewise-defined polynomial mean functions to account for the piecewise nature of spatially distributed data; Liu and Owen [2006] address a piecewise-defined mapping emerging from the pricing of a down-and-out barrier option; Römisch [2013] studies convex piecewise-defined functions in the context of quasi-Monte Carlo numerical integration. Recently, Roustant et al. [2018] offer a systematic approach to the Kriging emulation of computer experiments with piecewise-defined input-output mappings.

Formally, we have the following definition [Herrera, 2007, Borgonovo and Peccati, 2010]. Consider a finite partition of the domain $\mathcal{X}$, $\Pi_{\mathcal{X}} =$

$\{\mathcal{X}_1, \mathcal{X}_2, ..., \mathcal{X}_L\}$. Then, consider a set of L mappings $h_l : \mathcal{X}_l \mapsto \mathbb{R}$, $l = 1, 2, ..., L$.

Definition 3.2.2. *A mapping g is piecewise-defined if it can be written as:*

$$g(\mathbf{x}) = \begin{cases} h_1(\mathbf{x}) & \text{if} \quad \mathbf{x} \in \mathcal{X}_1 \\ h_2(\mathbf{x}) & \text{if} \quad \mathbf{x} \in \mathcal{X}_2 \\ \cdots & \cdots \quad \cdots \\ h_L(\mathbf{x}) & \text{if} \quad \mathbf{x} \in \mathcal{X}_L \end{cases} \tag{3.5}$$

with $h_l(\mathbf{x}) \neq h_m(\mathbf{x})$ almost everywhere on $\mathcal{X}_l$, for all $l \neq m, l, m = 1, 2, ..., L$.

Example 3.2.1. *The following piecewise-defined mapping is studied in Roustant et al. [2018]:*

$$y = \begin{cases} 0.1\big(x_1 + 0.01\,(x_1 - 0.5)^2\big)x_2; & \text{if } x_2 = 1, 2, 3, 4, \\ 0.9\cos\big(2\pi(x_1 + 0.05(x_2 - 4))\big)e^{-x_1} & \text{if } x_2 = 5, 6, 7, \\ -0.7\big(2\pi(x_1 + 0.05(x_2 - 7))\big)e^{-x_1} & \text{if } x_2 = 8, 9, 10, \end{cases} \tag{3.6}$$

with $x_1 \in [0, 1]$ and $x_2 \in \{1, 2, \ldots, 10\}$.

The next example shows a piecewise-defined mapping which is locally the sum of univariate functions.

Example 3.2.2. *Given the following mapping on $\mathcal{X} = [-1, 1]^2$:*

$$g = \begin{cases} \sin(x_1) + \sin(x_2) & \text{if } -1 \leq x_1 \leq 0, \\ \sin(x_1) + \sin(2x_2) & \text{if } 0 \leq x_1 \leq 1, \end{cases} \tag{3.7}$$

consider the change $\mathbf{x}^0 = \left(-\frac{1}{2}, -\frac{1}{2}\right) \to \mathbf{x}^1 = \left(\frac{3}{4}, \frac{3}{4}\right)$. One obtains

$$\Delta g = 1.2818 \quad , \quad \Delta_1 g = \quad 0.6737 \quad , \quad \Delta_2 g = 0.4830, \tag{3.8}$$

so that $\Delta g \neq \Delta_1 g + \Delta_2 g$. Thus, g is not additive.

The mapping in Example 3.2.2 is such that $g = \sum_{i=1}^{n} g_i(x_i)$ at every point of the domain but is not additive. The interactions are generated by its piecewise-defined nature. This unveils a new mechanism that generates non-null interactions. To proceed in the analysis, we propose the following definition.

Definition 3.2.3. *We say that a mapping g presents structural interactions if*

$$\Delta g \neq \sum_{i=1}^{n} \Delta_i g \tag{3.9}$$

for some change $\mathbf{x}^0 \to \mathbf{x}^1$.

Thus far, we have seen two mechanisms for the presence of structural interactions: piecewise-definiteness and the presence of interaction terms. The two mechanisms can act simultaneously. This simultaneous presence captures Højsgaard [2004]'s intuition of context-specific interactions. Højsgaard [2004] introduces this notion in the statistical analysis of contingency tables, thus not directly in computer experiments. According to Højsgaard [2004], an interaction is context-specific if an interaction term appears only for given levels of another factor.

Definition 3.2.4. *Given a piecewise-defined function $g : \mathcal{X} \to \mathbb{R}, \mathcal{X} \subseteq \mathbb{R}^n$, we say that g contains a context-specific interaction if at least one of the restrictions $h_i : \mathcal{X}_i \to \mathbb{R}$ of g onto $\mathcal{X}_i$ is not additive.*

Example 3.2.3. *The mapping $g : [0,1]^4 \to \mathbb{R}$,*

$$g = \begin{cases} x_1 + x_2 + x_3 & \text{if } x_4 < 0.5 \\ x_1 + x_2 + x_3 + x_1 x_2 & \text{if } x_4 > 0.5 \end{cases} \tag{3.10}$$

displays a context-specific interaction: the interaction term ($x_1 x_2$) is there only for $x_4 > 0.5$.

Thus, context-specific interactions are structural interactions that are related both to the presence of an explicit interaction term and to the

piecewise-defined nature of the mapping. However, this leads to the question of whether two interaction mechanisms (e.g., piecewise-definiteness and the presence of interaction terms) are (or not) of the same nature.

To address this point, consider the notion of removable interaction. [Berrington de González and Cox, 2007, p. 374] call an interaction removable if a transformation of the outcome scale can be found that induces additivity. To formalize this notion, we write:

Definition 3.2.5. *Given a mapping $g : \mathcal{X} \to \mathbb{R}$, $\mathcal{X} \subseteq \mathbb{R}^n$, we say that $g(\mathbf{x})$, $\mathbf{x} \in \mathcal{X}$, presents removable interactions if there exists a monotonic transformation $\eta(\cdot) : g(\mathcal{X}) \to \mathbb{R}$ such that the transformed function $z = \eta \circ g$, $z : \eta(g(\mathcal{X})) \to \mathbb{R}$ is additive.*

Example 3.2.4. *The mapping $y = x_1 x_2$ on $\mathcal{X} = [1, 2]^2$ presents a removable interaction, as a logarithmic transformation turns this mapping into $z = \ln(x_1) + \ln(x_2)$.*

We now prove that an interaction due to piecewise-definiteness is never removable.

Theorem 3.2.1. *Let $n > 2$, $g : \mathcal{X} \to \mathbb{R}$, $\mathcal{X} \subseteq \mathbb{R}^n$ be a piecewise-defined function. Then, it does not exist a monotone transformation $\eta(\cdot)$ such that $\eta \circ g$ is additive.*

This result implies that the two interaction generating mechanisms, explicit interaction terms and piecewise-definiteness, are indeed of a different nature. Moreover, we have that one can remove interactions if they are associated with the presence of explicit interaction terms, but this removal is not possible if interactions are generated by piecewise-definiteness.

These observations open a further theory-related question: whether some sufficient conditions on the regularity (e.g., differentiability) of g may rule out piecewise-definiteness as a reason for interactions. An answer is provided in the next section.

3.3 Differentiability, Transformations and Interactions

A first and widely used approach for measuring interactions in computer experiments is through mixed higher order derivatives (see, among others, Roustant et al. [2014]). Intuitively, because a twice differentiable additive mapping has null second and higher order mixed derivatives, a value $g''_{i,j}(\mathbf{x}^0)$ different from zero implies that g contains interactions at $\mathbf{x}^0$. It is not uncommon to find works in which the expression $\mathbb{E}\left[\left(g''_{i,j}(\mathbf{X})^2\right)\right] = 0$ is used to denote the absence of interactions [Friedman and Popescu, 2008]. However, the next example shows that $\mathbb{E}\left[\left(g''_{i,j}(\mathbf{X})^2\right)\right] = 0$ is not a necessary and sufficient condition for the absence of interactions, in general.

Example 3.3.1. *Consider the Lebesgue measure* $d\mathbf{x}$ *on* $[0,1]^2$ *and the function* $g(\mathbf{x}) = \mathbb{I}\left\{\frac{1}{2} \leq x_1 \leq 1\right\}\mathbb{I}\left\{\frac{1}{2} \leq x_2 \leq 1\right\}$, *where* $\mathbb{I}$ *denotes the indicator function. Since* $g''_{i,j}(\mathbf{x}) = 0$ *almost everywhere, it follows that* $\mathbb{E}\left[\left(g''_{i,j}(\mathbf{X})\right)^2\right] = 0$. *However,* g *is not additive.*

In the previous example, interactions are due to the piecewise-defined nature of the input-output mapping, and in spite of the fact that $\mathbb{E}\left[\left(g''_{i,j}(\mathbf{X})^2\right)\right] = 0$, we have structural interactions. However, in the next result, we provide the conditions under which $\mathbb{E}\left[\left(g''_{i,j}(\mathbf{X})^2\right)\right] = 0$ reassures us that g is additive.

Theorem 3.3.1. *Let* $\mathcal{X} \subset \mathbb{R}^n$ *and let* $g : \mathcal{X} \to \mathbb{R}$. *If* g *is additive and has second order mixed derivatives everywhere in* $\mathcal{X}$, *then* $g''_{i,j}(\mathbf{x}) = 0$ *for all* $\mathbf{x} \in \mathcal{X}$. *Conversely, if* $g''_{i,j}(\mathbf{x}) = 0$ *for all* $x \in \mathcal{X}$, *then* g *is additive.*

Theorem 3.3.1 helps us in the investigation of whether interactions can be attributed solely to the presence of explicit interaction terms. In fact, consider a mapping which is twice differentiable everywhere on its domain. Then, either the mapping is additive and then $g''_{i,j}(\mathbf{x}) = 0$ for all $\mathbf{x}$, or, if there is an interaction, then $g''_{i,j}(\mathbf{x})$ is different from zero and the interaction is due to an explicit interaction term.

Related to second order differentiability is a result obtained by Scheffé in his monograph on the analysis of variance [Scheffé, 1959]. Scheffé

presents a sufficient condition regarding the second order derivatives of a multivariate mapping that makes the mapping additive after transformation. Assume that $\eta(\cdot)$ and h_1, h_2 are twice differentiable everywhere. Then, Scheffé's conditions are satisfied (Scheffé [1959], p. 95, equation 4.1.12) and g solves the differential equation

$$g''_{1,2} - w(g)g'_1g'_2 = 0. \tag{3.11}$$

Example 3.3.2. *The mappings* $s = e^{x_1+x_2}$ *on* $\mathcal{X} = \mathbb{R}^2$, *and* $t = \sin(x_1 + x_2)$ *on* $\mathcal{X} = [0, \pi] \times [0, \pi]$ *are not additive. However, they satisfy Scheffé's differential equation in (3.11) with* $w(s) = \frac{1}{s}$ *and* $w(t) = \frac{1}{1-t^2}$, *respectively. Thus, they present removable interactions, with obvious transformations.*

Equation (3.11) states a condition on the second order partial derivative of g that makes interactions removable. Note that if $w(g) = 0$ interactions are absent (Theorem 3.3.1). More generally, Equation (3.11) is part of the family of quasilinear elliptic differential equations

$$a(x_1, x_2, g'_1, g'_2)g''_{1,2} + b(x_1, x_2, g, g'_1, g'_2) = 0. \tag{3.12}$$

The family admits, in general, no closed form solution and it is therefore not possible to characterize the complete set of mappings whose interactions are removable.

3.4 Interactions on a Finite Scale

A second way for analyst to determine interactions is to evaluate the simulator on two alternative locations and to compare the output-variation against a series of one-at-a-time changes. Definition 3.2.1 considers the variation of g when the inputs undergo the change $\mathbf{x}^0 \to \mathbf{x}^1$. Geometrically, the two points $\mathbf{x}^0$ and $\mathbf{x}^1$ can be seen as the vertices of a hypercube (see Figure 3.2 for a three-dimensional visualization).

In DOE, the first graph, $\mathcal{G}(\mathcal{V}, \mathcal{E})$, corresponds to a 2^k factorial design, the second, $\mathcal{G}'(\mathcal{V}', \mathcal{E}')$, to a standard one-at-a-time plan [Daniel, 1973].

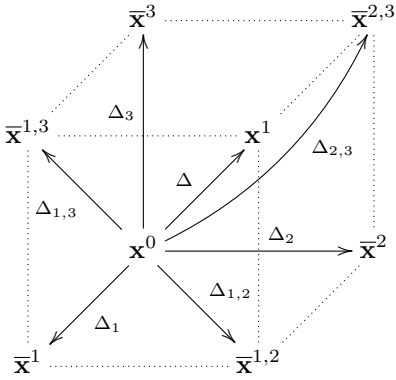


Figure 3.1: $\mathcal{G}(\mathcal{V}, \mathcal{E})$

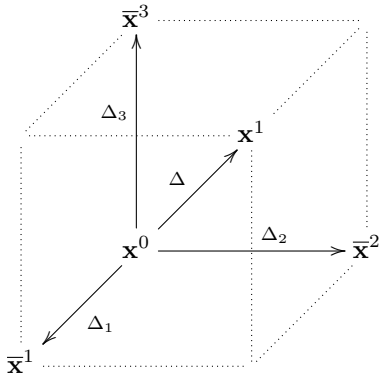


Figure 3.2: $\mathcal{G}'(\mathcal{V}', \mathcal{E}')$

The change in g across the two extremes $\mathbf{x}^0$ and $\mathbf{x}^1$, $\Delta g = g(\mathbf{x}^1) - g(\mathbf{x}^0)$, can be decomposed in a finite-change ANOVA expansion of $2^n - 1$ orthogonalized effects $\tau_z^{\mathbf{x}^0 \to \mathbf{x}^1}$ [Li et al., 2001]:

$$\Delta g = \sum_{z \in Z} \tau_z^{\mathbf{x}^0 \to \mathbf{x}^1}, \tag{3.13}$$

where the orthogonalized finite change $\tau_z^{\mathbf{x}^0 \to \mathbf{x}^1}$ is equal to [Borgonovo, 2010]:

$$\tau_z^{\mathbf{x}^0 \to \mathbf{x}^1} = \Delta_z g - \sum_{u \subset z} \tau_u^{\mathbf{x}^0 \to \mathbf{x}^1}. \tag{3.14}$$

The quantity $\tau_z^{\mathbf{x}^0 \to \mathbf{x}^1}$ represents the contribution to Δg of the residual interaction among the group of indices in z. Borrowing the notation from Liu and Owen [2006], let z denote a group of indices and define the total and subset finite change indices of z as $\overline{\tau}_z^{\mathbf{x}^0 \to \mathbf{x}^1} = \sum_{z \cap v \neq \emptyset} \tau_v^{\mathbf{x}^0 \to \mathbf{x}^1}$ and $\underline{\tau}_z^{\mathbf{x}^0 \to \mathbf{x}^1} = \sum_{v \subseteq z} \tau_v^{\mathbf{x}^0 \to \mathbf{x}^1}$. Then,

$$\Delta g = \sum_{v \cap z \neq \emptyset} \tau_v^{\mathbf{x}^0 \to \mathbf{x}^1} + \sum_{v \subseteq -z} \tau_v^{\mathbf{x}^0 \to \mathbf{x}^1} = \overline{\tau}_z^{\mathbf{x}^0 \to \mathbf{x}^1} + \underline{\tau}_{-z}^{\mathbf{x}^0 \to \mathbf{x}^1}. \tag{3.15}$$

Moreover, we find the equality $\underline{\tau}_{\{1,2,\ldots,n\}}^{\mathbf{x}^0 \to \mathbf{x}^1} = \Delta g$. These two equalities are analogues of corresponding formulas obtained for variance decomposition in works such as Sobol' [1993], Owen [2003]. Particularly relevant effects in design of experiments are the total effect of a factor [Myers et al., 2016], that represents its overall contribution to Δg:

$$\overline{\tau}_i^{\mathbf{x}^0 \to \mathbf{x}^1} = \sum_{z \supseteq i} \tau_z^{\mathbf{x}^0 \to \mathbf{x}^1}, \tag{3.16}$$

and its overall interaction effect:

$$\Upsilon_i^{\mathbf{x}^0 \to \mathbf{x}^1} = \sum_{z \supset i} \tau_z^{\mathbf{x}^0 \to \mathbf{x}^1}. \tag{3.17}$$

Note that $\Upsilon_i^{\mathbf{x}^0 \to \mathbf{x}^1}$ equals the difference between the total finite change effect $\overline{\tau}_i^{\mathbf{x}^0 \to \mathbf{x}^1}$ and the individual finite change effect, τ_i, of x_i.

In DOE, a second order interaction effect can be written as $A_{i,j} = \frac{1}{2}[g(\overline{\mathbf{x}}^{i,j}) + g(\mathbf{x}^0) - g(\overline{\mathbf{x}}^i) - g(\overline{\mathbf{x}}^j)]$ [Wu, 2015, eq. (3.1), p. 614]. Therefore, we have that $\tau_{i,j}^{\mathbf{x}^0 \to \mathbf{x}^1} = A_{i,j}/2$. Similarly, one could re-write higher

order finite change effects in terms of higher order $A_{i,j,\ldots,k}$. For notation simplicity, however, we shall continue with the "τ" notation.

Example 3.4.1. *Consider the change in* $g = \dfrac{x_1}{x_1 + x_2}$, *as* $\mathbf{x}$ *varies from* $\mathbf{x}^0 = (1,1)$ *to* $\mathbf{x}^1 = (2,5)$. *One registers* $\Delta g = -\frac{3}{14}$, $\Delta g = \tau_1^{\mathbf{x}^0 \to \mathbf{x}^1} + \tau_2^{\mathbf{x}^0 \to \mathbf{x}^1} + \tau_{1,2}^{\mathbf{x}^0 \to \mathbf{x}^1}$, *with* $\tau_1^{\mathbf{x}^0 \to \mathbf{x}^1} = \frac{1}{6}$, $\tau_2^{\mathbf{x}^0 \to \mathbf{x}^1} = -\frac{1}{3}$, *and* $\tau_{1,2}^{\mathbf{x}^0 \to \mathbf{x}^1} = -\frac{1}{21}$, *so that* $\Upsilon^{\mathbf{x}^0 \to \mathbf{x}^1} = \left(-\frac{1}{21}, -\frac{1}{21}\right)$, $\overline{\tau}_i^{\mathbf{x}^0 \to \mathbf{x}^1} = \left(\frac{5}{42}, -\frac{8}{21}\right)$.

The complete decomposition of Δg in (3.13) requires to evaluate g on all the vertices of graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$. Such knowledge is equivalent to the determination of the Laplace operator of g on the same graph. It is then possible to link the finite-change effects of DOE to the Laplace operator. The link is discussed in detail in Appendix A. Because $\mathcal{G}(\mathcal{V}, \mathcal{E})$ has 2^n vertices, determining the complete orthogonalized decomposition of a finite change is a computationally intensive task. For instance, for $n = 20$, 2^n is larger than 10^6. However, the literature has introduced a computational shortcut that allows us to determine the triplets $\tau_i^{\mathbf{x}^0 \to \mathbf{x}^1}$, $\Upsilon_i^{\mathbf{x}^0 \to \mathbf{x}^1}$ and $\overline{\tau}_i^{\mathbf{x}^0 \to \mathbf{x}^1}$ at a linear instead of an exponential cost in n. It can be proven that

$$\tau_i^{\mathbf{x}^1 \to \mathbf{x}^0} = -g(x_i^0 : \mathbf{x}_{-i}^1) + g(\mathbf{x}^1) = -\overline{\tau}_i^{\mathbf{x}^0 \to \mathbf{x}^1}, \tag{3.18}$$

which is the finite scale equivalent of the so-called pick and freeze design [Borgonovo, 2010, Gamboa et al., 2016]. Specifically, equation (3.18) suggests that the individual effect of x_i in the change $\mathbf{x}^1 \to \mathbf{x}^0$ is the opposite of the total effect of x_i in the symmetric change $\mathbf{x}^0 \to \mathbf{x}^1$. Thus, two series of one-way sensitivities are sufficient to estimate $\tau_i^{\mathbf{x}^0 \to \mathbf{x}^1}$, $\Upsilon_i^{\mathbf{x}^0 \to \mathbf{x}^1}$ and $\overline{\tau}_i^{\mathbf{x}^0 \to \mathbf{x}^1}$, so that the total computational cost is $2n + 2$ evaluations of g instead of 2^n. The link with the Laplace operator leads to another computational shortcut. In fact, it can be proven that if g is additive, then only two function evaluations are necessary to estimate the Laplace operator on $\mathcal{G}(\mathcal{V}, \mathcal{E})$ (See supplementary Proposition 3.9.1).

Let us now come to the link between interactions on a finite and on an infinitesimal scale. Let $h_i = x_i^1 - x_i^0, h_j = x_j^1 - x_j^0$, $j \neq i, i, j = 1, 2, \ldots, n$. If g is twice differentiable, we obtain the following relationship between

interactions effects on a finite and on an infinitesimal scale:

$$g''_{i,j}(\mathbf{x}^0) = \lim_{\mathbf{x}^1 \to \mathbf{x}^0} \frac{A_{i,j}(\mathbf{x}^0, \mathbf{x}^1)}{2(x_i^1 - x_i^0)(x_j^1 - x_j^0)}. \tag{3.19}$$

A similar argument applies to higher order interactions. Thus, an infinitesimal interaction can be regarded as the limit for $\mathbf{x}^1 \to \mathbf{x}^0$ of a finite scale interaction. Suppose now that the simulator g is differentiable. Then, we have:

$$\lim_{\mathbf{x}^1 \to \mathbf{x}^0} \frac{\tau_i^{\mathbf{x}^0 \to \mathbf{x}^1}}{h_i} = \lim_{\mathbf{h} \to \mathbf{0}} \frac{\overline{\tau_i}^{\mathbf{x}^0 \to \mathbf{x}^1}}{h_i} = g'_i(\mathbf{x}^0) \text{ and } \lim_{\mathbf{x}^1 \to \mathbf{x}^0} \frac{\Upsilon_i^{\mathbf{x}^0 \to \mathbf{x}^1}}{h_i} = 0. \tag{3.20}$$

The last equality suggests that for a smooth mapping, if input changes are sufficiently small, interaction effects present in the input mapping might not be actually registered by finite change indices.

3.5 Global Interactions

A relevant role in computer experiments is played by the study of interactions in global sensitivity analysis [Saltelli and Tarantola, 2002]. Fundamental for the analysis is the classical functional ANOVA expansion [Efron and Stein, 1981]. One regards the simulator inputs as a random vector $\mathbf{X} = \{X_1, X_2, \ldots, X_n\}$ on measure space $(\mathcal{X}, \mathcal{B}(\mathcal{X}), F_{\mathbf{X}})$ and assumes $F_{\mathbf{X}} = \prod_{i=1}^n dF_i$. If $g \in \mathcal{L}^2(\mathcal{X}, \mathcal{B}(\mathcal{X}), F_{\mathbf{X}})$, the seminal result of Efron and Stein [1981] allows us to write the variance of Y as:

$$\mathbb{V}[Y] = \sum_{z \subset Z, z \neq \emptyset} V_z, \tag{3.21}$$

where

$$V_z = \int [g_z(\mathbf{x}^z)]^2 \, dF_z(\mathbf{x}^z) \quad \text{, and} \quad g_z = \int g(\mathbf{x}^z : \mathbf{x}^{-z}) dF_{-z}(\mathbf{x}^{-z}) - \textstyle\sum_{v \subset z} g_v(\mathbf{x}^v). \tag{3.22}$$

In (3.21), the first order term V_i ($z = \{i\}$) represents the individual contribution of X_i to the variance of Y. The terms V_z represents the residual contribution due to the interaction among the simulator inputs whose indices are in z. Note that V_z is part of the H_z^2 test statistics for interactions

introduced in Friedman and Popescu [2008]. Liu and Owen [2006] define the sensitivity indices

$$\underline{\tau}_u^2 = \sum_{z \subseteq u} V_z \quad \text{and} \quad \overline{\tau}_u^2 = \sum_{z \cap u \neq \emptyset} V_z. \tag{3.23}$$

The index $\overline{\tau}_u^2$ represents the total contribution of simulator inputs with indices in u to the variance of g. As Owen [2014, p. 247] underlines, small values of $\overline{\tau}_u^2$ indicate that the simulator inputs in u have so little influence that they can be conveniently fixed, thus reducing the problem dimensionality. Note that if $u = \{i\}$, then $\overline{\tau}_i^2$ is the analog for the decomposition of $\mathbb{V}[Y]$ of the total index $\overline{\tau}_i^{\mathbf{x}^0 \to \mathbf{x}^1}$ for the decomposition of Δg. Similar connections hold for other indices, as we discussed in Section 3.4. Hooker [2004] and subsequently Liu and Owen [2006] define the superset importance index of a group of simulator inputs u as

$$\Upsilon_u = \sum_{z \supseteq u} V_z \tag{3.24}$$

This index Υ_u measures the contribution of all terms that include the indices in u in the ANOVA decomposition of g.

Caflisch et al. [1997] and Owen [2003]'s notions of dimension distribution and mean effective dimension in the superimposition and truncation sense provide an elegant characterization of the presence of interactions within the classical functional ANOVA framework. The ratios $S_u = V_u/\mathbb{V}[Y]$ are regarded as the probability mass function of a random variable, say T, whose support is 2^Z, so that $\Pr(T = u) = S_u$. The dimension distributions of g in the superimposition sense and truncation sense are then defined as the distribution of the cardinality of T and of $\max\{i : i \in z\}$, respectively. From these distributions, one defines the mean dimensions in the superimposition and truncation senses. Specifically, Owen [2013] proves that the mean dimension (D_g) satisfies the equality

$$D_g = \sum_u |u| S_u = \sum_{i=1}^{n} \overline{\tau}_i^2 / \mathbb{V}[Y]. \tag{3.25}$$

Owen [2003] shows that for any additive square-integrable function the mean dimensions in the truncation and superimposition sense as well are

unity. As a consequence, a mean dimension in the superimposition sense greater than unity implies that some non-null interaction terms g_z with $|z| > 1$ are present in the decomposition of g.

Example 3.5.1. *Consider the piecewise constant simulators* $g_1 = \mathbb{I}_{E_1 \cup \cdots \cup E_n}$ *and* $g_2 = \mathbb{I}_{E_1} \cdots \mathbb{I}_{E_n}$*, where* E_i *denotes the event* $\{x_i \geq 0.5\}, i = 1, 2, \ldots, n$*. Provided with random inputs* $X_i \sim U[0, 1]$ *iid, both simulators satisfy* $S_z = \frac{1}{2^n - 1}$ *for all multi-indices* z*. They have the same mean dimension in the superimposition sense equal to*

$$\sum_z |z| S_z = \frac{\sum_{j=1}^n \binom{n}{j} j}{\sum_{j=1}^n \binom{n}{j}} = \frac{n \cdot 2^{n-1}}{2^n - 1} \approx \frac{n}{2}.$$

The mean dimension is greater than unity signaling the presence of interactions.

On the other hand, g is said to have effective dimension s in the superposition sense if $\sum_{|u| \leq s} S_u \geq 0.99$ [Caflisch et al., 1997, Owen, 2003]. Consider the special case $s = 1$. We would say that g has dimension 1 in the superimposition sense if $\sum_{i=1}^n S_i \geq 0.99$. However, this would not mean that the function is additive. In fact, consider the function in Example 3.5.1, with $X_1 \sim Gamma(10, 1)$ and $X_2 \sim Gamma(10, 1)$. One registers $S_1 + S_2 = 0.999 > 0.99$, although the mapping is not additive.

3.5.1 Global Interactions and Cox's constancy of Variance

Cox [1984] and Berrington de González and Cox [2007] discuss the link between constancy of variance and absence of interactions. The next result offers a formalization of this intuition based on the classical functional ANOVA expansion and the notion of heteroskedasticity.

Definition 3.5.1. *A random variable* Y *is called homoskedastic with respect to a random variable* X *if the conditional variance* $\mathbb{V}[Y|X]$ *is constant, or heteroskedastic otherwise.*

This definition might be visualized by noting that the scatterplot of (X, Y) fits into a tube of constant diameter in case of homoskedasticity.

Proposition 3.5.1. The mapping g is heteroskedastic with respect to X_i if and only if there are non-vanishing interaction terms involving X_i in the ANOVA decomposition of $\mathbb{V}[Y]$.

An interesting consequence of this result is a sufficient condition for the conditional variance $\mathbb{V}[Y|X_i]$ to be greater than the unconditional variance $\mathbb{V}[Y]$. An immediate consequence is the following.

Corollary 3.5.1. *If there exists x_i in the support of X_i such that $\mathbb{V}[Y|X_i = x_i] > \mathbb{V}[Y]$, then Y is heteroskedastic with respect to X_i.*

This is the intriguing case in which fixing a model input based upon new information actually increases the simulator output variance. This corollary implies that this effect (i.e., the increase in variance due to learning a factor) can happen only in the presence of interactions.

Example 3.5.2. *Consider $g : \mathbb{R}^2 \mapsto \mathbb{R}$, $(X_1, X_2) \mapsto e^{X_1}|\sin(X_2)|$, which is constituted by an explicit interaction term. With X_1 standard normal and X_2 normal with mean and standard deviation equal to 1, the simulator variance is $\mathbb{V}[Y] = 2.72$, while $\mathbb{V}[Y|X_2 = 1] = 3.31$.*

3.5.2 Linking Interactions on the Infinitesimal, Finite and Global Scales

We have seen that a sufficient (but not necessary) condition for the absence of interactions is the nullity of second order derivatives. Indeed, Fruth et al. [2014] propose the crossed derivative-based global sensitivity measure

$$\nu_{i,j} = \mathbb{E}\left[\left(g''_{i,j}(\mathbf{X})\right)^2\right]. \tag{3.26}$$

Fruth et al. [2014] prove that, if g is twice differentiable and its all first order and second order partial derivatives are in $L^2(F_{\mathbf{X}})$ and if all F_i $(i = 1, \ldots, n)$ satisfy the Poincaré inequality

$$\int g(\mathbf{x})^2 dF_{\mathbf{X}}(\mathbf{x}) \leq C(F_{\mathbf{X}}) \int ||\nabla g(\mathbf{x})||^2 dF_{\mathbf{X}}(\mathbf{x}),$$

where $C(F_{\mathbf{X}})$ is the Poincaré constant of $F_{\mathbf{X}}$, then the following relationship holds between the second order Sobol' index $V_{i,j}$, the superset index $\Upsilon_{i,j}$ and $\nu_{i,j}$:

$$V_{i,j} \leq \Upsilon_{i,j} \leq C(F_i)C(F_j)\nu_{i,j}, \tag{3.27}$$

where $C(F_i)$ and $C(F_j)$ are the Poincaré constants of the marginal distributions of X_i and X_j. Then, if $g''_{i,j} = 0$ everywhere, by equation (3.27) the second order sensitivity indices are zero. That is, absence of second order local interactions everywhere implies absence of second order global interactions. However, it may hold that $\nu_{i,j} = 0$ and $V_{i,j} \neq 0$.

Proposition 3.5.2. For a piecewise-defined mapping (3.5) with additive components $h_l(\cdot)$, $(l = 1, 2, \cdots, L)$, it holds that $\nu_{i,j} = 0$ for all $i, j = 1, 2, \ldots, n$, $i \neq j$.

This proposition makes clear that, when interactions are only due to piecewise-definiteness (see the case study STOCFOR3), a differentiation-based approach would not lead to meaningful insights.

Example 3.5.3. *For the mapping in Example 3.3.1, one finds $S_{1,2} = 1/3$, while $\nu_{1,2} = 0$.*

Works such as Liu and Owen [2006] and Campolongo et al. [2011] show that it is possible to build a conceptual bridge between finite change and global interactions. The results in the previous sections allow us to take a fresh look at this bridge. The next result connects finite-change and global interaction measures.

Theorem 3.5.1. *Let $g \in L^2(\mathcal{X})$ and let $\boldsymbol{X}, \boldsymbol{Z} \sim U[0,1]^n$ be independent and identically distributed. Then, for $u \in Z$, we have:*

$$\Upsilon_u = \tfrac{1}{2^{|u|}}\mathbb{E}\left[\tau_u^{\mathbf{X}\to\mathbf{Z}}\right]^2 \quad , \quad \underline{\tau}_u^2 = \mathbb{E}\left[g(\boldsymbol{X})\underline{\tau}_u^{\mathbf{Z}\to\mathbf{X}}\right] \quad , \quad \overline{\tau}_u^2 = \tfrac{1}{2}\mathbb{E}\left[\underline{\tau}_u^{\mathbf{X}\to\mathbf{Z}}\right]^2, \tag{3.28}$$

where $|u|$ denotes the cardinality of u. Moreover, we have

$$\textstyle\sum_{i=1}^n V_i = \sum_{i=1}^n \mathbb{E}\left[g(\boldsymbol{X})\tau_i^{\mathbf{Z}\to\mathbf{X}}\right] \quad \text{and} \quad \sum_u |u|V_u = \frac{1}{2}\sum_{i=1}^n \mathbb{E}\left[\tau_i^{\mathbf{X}\to\mathbf{Z}}\right]^2. \tag{3.29}$$

The above result leads to the following Monte Carlo estimators that explicitly connect indicators of interactions on a finite scale to indicators on a global scale.

Corollary 3.5.2. *The global interaction indicators* (3.28) *and* (3.29) *can be estimated from*

$$\widehat{\Upsilon}_u = \frac{1}{2^{|u|}} \cdot \frac{1}{N} \sum_{k=1}^{N} \left[\tau_u^{\mathbf{x}^{(k)} \to \mathbf{z}^{(k)}} \right]^2, \tag{3.30}$$

$$\widehat{\underline{\tau}_u^2} = \frac{1}{N} \sum_{k=1}^{N} g(\mathbf{x}^{(k)}) \underline{\tau}_u^{\mathbf{z}^{(k)} \to \mathbf{x}^{(k)}}, \tag{3.31}$$

$$\widehat{\overline{\tau}_u^2} = \frac{1}{2N} \sum_{k=1}^{N} \left[\underline{\tau}_u^{\mathbf{x}^{(k)} \to \mathbf{z}^{(k)}} \right]^2. \tag{3.32}$$

We also have

$$\sum_{i=1}^{n} \widehat{V}_i = \frac{1}{N} \sum_{k=1}^{N} \sum_{i=1}^{n} g(\mathbf{x}^{(k)}) \tau_i^{\mathbf{z}^{(k)} \to \mathbf{x}^{(k)}}, \tag{3.33}$$

$$\sum_{u} |u| \widehat{V}_u = \frac{1}{2N} \sum_{k=1}^{N} \sum_{i=1}^{n} \left[\tau_i^{\mathbf{x}^{(k)} \to \mathbf{z}^{(k)}} \right]^2. \tag{3.34}$$

When $u = \{i\}$, i.e. we are considering only singletons, then the estimator (3.32) *is known as the Jansen estimator of the total effect of the i-th factor (see Campolongo et al. [2011]). When $u = \{i, j\}$, then the Monte Carlo estimator* (3.30) *has been studied by [Fruth et al., 2014] and we show that it equals:*

$$\widehat{\Upsilon}_{i,j} = \frac{1}{4N} \sum_{k=1}^{N} \left[\tau_{i,j}^{x^{(k)} \to z^{(k)}} \right]^2 = \frac{1}{4N} \sum_{k=1}^{N} \left[2A_{i,j}^{(k)} \right]^2 = \frac{1}{N} \sum_{k=1}^{N} \left[A_{i,j}^{(k)} \right]^2 = \frac{1}{N} \sum_{k=1}^{N} [EI^{(k)}]^2.$$

where EI is the screening interaction effect introduced in [Campolongo et al., 2011].

The above results show that the same finite changes give rise to alternative estimators. Precisely, if we consider the indices $\tau_u^{\mathbf{x}^0 \to \mathbf{x}^1}$ of the

orthogonal decomposition of Δg, we recover Liu and Owen's Υ_u. If we consider $\underline{\tau}_u^{\mathbf{x^0}\to\mathbf{x^1}}$, we obtain estimators for $\underline{\tau}_u^2$ and $\overline{\tau}_u^2$. In terms of the Laplace operator on the full graph, the terms $\Delta_u g$, would allow the estimation of the Sobol' $\underline{\tau}_u^2$ and $\overline{\tau}_u^2$ indices. Also, it is easy to see that estimators of (3.28)–(3.29) in Theorem 3.5.1 coincide with pick-and-freeze estimators of Gamboa et al. [2016]. This connection has the implication that if the analyst implements a pick-and-freeze design and keeps track of the estimators before squaring, she can recover the magnitude of interactions, as well as their sign at several locations of the simulator input space. Thus, the analyst can gain insights on whether interaction effects are synergistic or antagonistic while performing a global sensitivity analysis.

3.5.3 Spurious Interactions: Generalized Functional ANOVA

The identification of interactions becomes more involved for computer experiments with dependent inputs. Under input dependence, interactions may emerge even if they are not present in the input-output mapping. Friedman and Popescu [2008] deem these as *spurious interactions*. In the light of our analysis, spurious interactions are not structural but are caused by collinearity among some simulator inputs.

Example 3.5.4. *Consider the additive function* $Y = X_1^2 + \frac{X_2}{6}$ *with* (X_1, X_2) *uniformly in* $\left\{(x_1, x_2) \in [0,1]^2 : x_1^2 + x_2^2 > \frac{1}{2}\right\}$. *The mapping is additive. Then, applying the definition of classical first order sensitivity indices we find* $S_1 = 0.98$ *and* $S_2 = 0.16$, *with the first order indices accounting for more than* 100% *of the output variance. To counterbalance this excess, one would have to postulate a negative spurious interaction term* $S_{1,2} = -0.14$, *which has no interpretation in the classical ANOVA frame.*

The problem has remained open till the work of Hooker [2007], who shows that one can still recover a unique functional ANOVA expansion of the form of (3.21), by appropriately changing the conditions that allow the determination of the component functions g_z. The work of Hooker [2007] has been followed by a series of works, such as Li et al. [2012], Chastaing

Table 3.1: Generalized ANOVA results for the example with a Quasi-Monte Carlo sample of size 4983. The results have been obtained applying the generalized functional ANOVA expansion and using the D-MORPH regression approach of Li et al. [2012] for computing the sensitivity indices.

i	$\widehat{S}_i$	$\widehat{S}_i^c$	$\widehat{S}_i+\widehat{S}_i^c$	$\widehat{S}_{i,j}+\widehat{S}_{i,j}^c$
1	1.1399	−0.0830	1.0570	0
2	0.0253	−0.0830	−0.0577	0
Sum	1.1652	−0.1660	0.9993	0

et al. [2012], Rahman [2014], that show that under dependence we can use a covariance decomposition of $\mathbb{V}[Y]$ that leads to

$$\mathbb{V}[Y] = \sum_{\emptyset \neq z \subset Z} \left[V_z + \mathrm{Cov}\left(g_z, \sum_{\emptyset \neq v \subset Z, v \neq z} g_v \right) \right]. \tag{3.35}$$

Note that, when the inputs are independent, this formula reduces to (3.21). Normalizing the covariance decomposition (3.35), one obtains [Li et al., 2012]

$$\sum_{\emptyset \neq z \subset Z} \left[\frac{V_z}{\mathbb{V}[Y]} + \frac{\mathrm{Cov}\left(g_z, \sum_{\emptyset \neq v \subset Z, v \neq z} g_v \right)}{\mathbb{V}[Y]} \right] = \sum_{\emptyset \neq z \subset Z} [S_z + S_z^c] = 1. \tag{3.36}$$

In (3.36), the variance-based index S_z is the marginal contribution of the indices in z, while the quantity S_z^c measures the relevance of the covariances among the component functions g_z, reflecting the contribution due to correlations. Hence, S_z^c may be regarded as indicators of spurious interactions.

Example 3.5.5 (Example 3.5.4 continued)**.**

If we apply the generalized functional ANOVA expansion we obtain the results in Table 3.1 which displays a non-null contribution coming from correlations. Also, due to correlations, we find a negative overall

Table 3.2: Interaction types versus interaction measurement.

Scale	Inter.Term	Piec.-def.	Spurious	Sign	Discreteness	Cost	Ass. on g
Infinites.	Yes	No	No	Yes	No	$4nN$	C^1
Finite	Yes	Yes	No	Yes	Yes	2^n	-
Global	Yes	Yes	Yes	Estimator Dependent	Yes	$N^2 2^n$	$\mathcal{L}^2$

sensitivity index for S_2. However, the second order terms are null, a result in accordance with the additivity of the input-output mapping. To our knowledge, it is the first time that the connection between covariance decomposition and spurious interactions is addressed.

3.6 Some Implications

Table 3.2 organizes interactions according to their type (due to interaction terms, generated by piecewise-definiteness, spurious) along with the investigation scale (infinitesimal, finite, global) and the chosen interaction measure.

Table 3.2 synthesizes the following observations regarding the identification of interactions based on our previous results. If the statistician is relying on interaction measures based on differentiation, on an infinitesimal scale, she will detect structural interactions if they are due to explicit interaction terms, but may not identify structural interactions due to piecewise-definiteness. However, using partial derivatives the analyst is reassured not to measure spurious interactions and has an indication on the sign of interactions. Using finite-change interaction measures, the analyst unveils structural interactions due to piecewise-definiteness and to interaction terms, does not detect spurious interactions, and gains insights on the sign of interactions. Global methods allow to appreciate structural interactions as well as spurious interactions. The detection of sign depends on the estimator. For instance, Theorem 3.5.1 shows that we can link a pick-and freeze design to finite change indices: if the sign of the estimates

is registered before taking the square, then one obtains, simultaneously, information on interactions on the finite and global scales.

In deciding the approach, considerations regarding computational burden play a crucial role. On an infinitesimal scale, a brute force calculation of second order derivatives based on second order Newton's ratios. demands $4n$ model runs. If the estimation of the second order derivatives is randomized at N locations in $\mathcal{X}$, the estimation cost becomes $C^{\text{Newton}} = 4 \cdot n \cdot N$. This number of evaluations is a lower limit as one should repeat the calculation for a sequence of decreasing values of the increments used in Newton's ratio, till a limit is approached. Alternatively, the analyst could resort to automatic differentiation. Here, one obtains derivatives while evaluating the computer code without requiring additional model runs. However, the automatic differentiation subroutine needs direct access to the source code of the simulator. If such access is not available, then the approach cannot be applied; conversely, careful programming is needed not to increase the running time of the simulator.

On a finite scale, the computational cost associated with the determination of interactions up to order k is $C^k = \sum_{s=0}^{k} \binom{n}{s}$. If $k = n$ (e.g., in a full factorial design) the analyst incurs a cost of 2^n model runs. However, if the analyst wishes to achieve a less granular information, she can use computational shortcuts. For instance, it is possible to compute the indices $\tau_i^{\mathbf{x}^0 \to \mathbf{x}^1}$, $\Upsilon_i^{\mathbf{x}^0 \to \mathbf{x}^1}$ and $\overline{\tau}_i^{\mathbf{x}^0}$ at a cost of $2n + 2$ model runs using the shortcut in equation (3.18). Moreover, several efficient designs allow one to compute interaction effects up to a desired order, avoiding the cost of a full factorial design.

For global interactions, calculating the complete variance decomposition in (3.21) using a brute force approach has a cost of $C^V = 2^n N^2$, where N is a suitable size of a random simulator input sample. The literature has introduced designs that abate this computational cost. For instance, the modified pick-and-freeze method of Saltelli [2002] allows one to obtain estimates for the first order, total order and second order superset indices at a cost of $N(n+2)$ model runs. Alternatively, an analyst may proceed at a cost of N model runs by fitting an emulator over the input-output sam-

ple generated for uncertainty quantification. Depending on the emulation approach used in the analysis, analytical formulas for the estimators of variance-based sensitivity indices of all orders may be available. Then, if the fit is accurate, one has a computationally convenient way of calculating global sensitivity measures.

Given also the considerations on computational burden, ideally, the analyst could follow a three step procedure: specify which interaction type she is concerned with, select the quantification measure (e.g., partial derivative, finite change, variance based) and, finally, select a design that allows to estimate such measure. To illustrate: Is the analyst interested in the precise interaction between two specific simulator inputs x_i and x_j at $\mathbf{x}^0$ or as they move across two scenarios $\mathbf{x}^0 \to \mathbf{x}^1$? Then, in the first case partial derivatives are an appropriate measure of interaction, in the second case the analyst needs a factorial experiment with 4 simulator evaluations. Or, is the analyst just asking whether there are interactions at all (in any form)? Then, a design that allows the estimation of first order variance-based indices could be the most convenient to apply; if their sum is lower than unity, we know that interactions are there. On a global scale, if the analyst is interested in knowing whether interactions are synergistic or antagonistic, she needs a design that allows to retain information on the sign of interactions (see Corollary 3.5.2 and the following discussion).

3.7 Case Studies

In this section, we illustrate some aspects of the previous analysis via a set of simulators previously studied in the literature. The analysis is carried out on a desktop with Intel(R) core i7-7700HQ CPU at 2.80GHz and RAM 64G. All calculations are performed in Matlab R2017b.

We start with the Wing-weight function, recently used in Jiménez Rugama and Gilquin [2018] to test the estimation of Sobol' indices. Simulator size ($n = 10$) and running time (0.02 sec) do not lead to any computational issue for this simulator. Jiménez Rugama and Gilquin [2018] offer estimates of global sensitivity measures from an input-output sample of

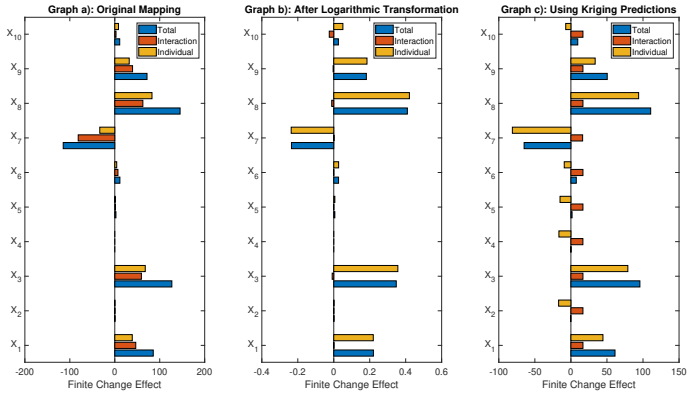


Figure 3.3: Interaction Analysis of the Wing-weight simulator on a Finite Scale: Graph a) calculating the finite change indices using the original code; Graph b) after a transformation of the original code; Graph c) using a Kriging emulator instead of the original code.

size $N = 65,536$. Their accurate estimates (see Table 12 [Jiménez Rugama and Gilquin, 2018, p. 736]) lead to $\sum_{i=1}^{10} \hat{S}_i = 0.9814$, i.e., the sum of first order variance-based indices is close to unity, signaling that interactions do not play a major role on a global scale. Indeed, fitting an additive linear regression surface to the input-output sample leads to a highly significant fit with a coefficient of model determination $R^2 = 0.982$ (We used an $N = 65,536$ sample from the simulator and the fitlm.m subroutine).

Consider now an analyst evaluating the change in model output across the bounds of the model inputs, namely $\mathbf{x}^0 = (150, 220, 6, -10, 16, .5, .08, 2.5, 1700, .025)$ and $\mathbf{x}^1 = (200, 300, 11, 10, 45, 1, .18, 6, 2500, .08)$ [Jiménez Rugama and Gilquin, 2018]. One registers $g(\mathbf{x^0}) = 158.30$, $g(\mathbf{x}^1) = 432.50$, and $\Delta y = 274.20$. The complete decomposition of the finite change using (3.17) requires $1,024$ model evaluations and is feasible in about 20 seconds for this simulator. Figure 3.3 reports the first ($\tau_i^{\mathbf{x}^0 \to \mathbf{x}^1}$), total ($\overline{\tau}_i^{\mathbf{x}^0 \to \mathbf{x}^1}$) and interaction effects ($\Upsilon_i^{\mathbf{x}^0 \to \mathbf{x}^1}$) for this finite change. Figure 3.3 shows that, on a finite scale, interactions are of the same order (if not higher) than

individual effects. Now, suppose for a moment that the analyst were to use the fitted linear surface instead of the original simulator to analyze the change. She would obtain $g^{\text{linear}}(\mathbf{x}^0) = 141.55$ and $g^{\text{linear}}(\mathbf{x}^1) = 413.25$, with $\Delta g^{\text{linear}} = 271.70$. Thus, the linear approximation of the magnitude of Δg is accurate. However, an evaluation of interactions using the linear approximation would not be meaningful, because the emulator is additive. Suppose that, instead of the linear regression, the analyst were to fit a Kriging emulator [Santner et al., 2003] and then analyze the interactions using the Kriging surface instead of the original model. She would obtain the results in graph *c*) of Figure 3.3. The Kriging surface would reveal interactions, however with indications different from the ones yielded by the original simulator. For instance, input X_7 is attributed an individual effect higher than its total effect and an overall positive interaction effect, while using the original simulator we register a total effect prevailing over the individual effect and a negative overall interaction effect.

For further analysis, we compared the previous results with the insights of an interaction analysis at the infinitesimal scale. We calculated second order partial derivatives at $1,000$ locations in the model input space, via automatic differentiation. The partial derivatives have then been squared and averaged to obtain an estimator of $\nu_{i,j}$ in (3.26). The analysis reveals 38 non null second order interactions, with some significant ones, such as the interactions between X_7 and X_8, X_7 and X_3, and X_7 and X_6. Second order interactions involving model inputs X_2, X_5 and X_9 are present but negligible in size.

Graph *b*) shows the results of an interaction analysis on Wing-weight performed after the original model output has been subjected to a logarithmic transformation. The graph shows that interactions are almost completely removed. This result indicates that interactions are mainly related to the presence of explicit interaction terms in the simulator structure and not to piecewise-definiteness. This is indeed in line with the structure of the simulator, which is not piecewise-defined and is constituted by the sum of two terms, with the main one involving the product/power of inputs X_1 to X_9, the second involving the product between the sole X_1 and

X_{10}.

Now, Wing-weight's size does not raise any computational concern. We then consider a simulator of larger size, STOCFOR3, the largest linear program in the well-known Netlib repository. Linear programs are fundamental simulators in business planning, but linear optimization appears in several problems, comprising statistical analyses and regression, with the famous Dantzig's selector of Candes and Tao [2007] as an outstanding example. The input data of STOCFOR3 are freely available. We study the sensitivity of the optimal value of this linear program to changes in its 23,541 coefficients for variations of $\pm 99\%$ of their values. In other words, we consider graph *a)* in Figure 3.2, in which $\mathbf{x}^0$ and $\mathbf{x}^1$ are the extremes of the 23,541-dimensional hyperbox. We observe that the output of this simulator is, at any value of the coefficients, a linearly additive map.

Given the simulator size, the shortcut in equation 3.18 turns out essential in this case — to illustrate, there are about $2.7 \cdot 10^8$ second order interactions. With $47,084$ simulator evaluations we can obtain all first ($\tau_i^{\mathbf{x}^0 \to \mathbf{x}^1}$), total ($\overline{\tau}_i^{\mathbf{x}^0 \to \mathbf{x}^1}$) and interaction effects ($\Upsilon_i^{\mathbf{x}^0 \to \mathbf{x}^1}$). Each simulator evaluation entails a new optimization. The analysis takes about 14 hours on the above mentioned computer. Figure 3.4 reports the finite change effects for the first 10 simulator inputs ranked using the magnitude of $\overline{\tau}_i^{\mathbf{x}^0 \to \mathbf{x}^1}$. Figure 3.4 shows that, indeed, interaction effects are present in the simulator response. These interactions are due to the piecewise-defined nature of the simulator. In fact, denoting with $\mathbf{z}^*$ the optimal solution at $\mathbf{x} \in \mathcal{X}$, the input-output mapping of any linear program is a mapping of the type

$$y = \begin{cases} \mathbf{x}\mathbf{z}_1^* & if \quad \mathbf{x} \in \mathcal{X}_1 \\ \mathbf{x}\mathbf{z}_2^* & if \quad \mathbf{x} \in \mathcal{X}_2 \\ & \dots \\ \mathbf{x}\mathbf{z}_m^* & \quad \mathbf{x} \in \mathcal{X}_m. \end{cases} \tag{3.37}$$

This mapping is locally linear and additive. Note that second order derivatives are null at any point $\mathbf{x} \in \mathcal{X}$, while the mapping is not additive (see also Proposition 3.5.2). More generally, these results shows that interactions associated with variations in the coefficients of a linear program are

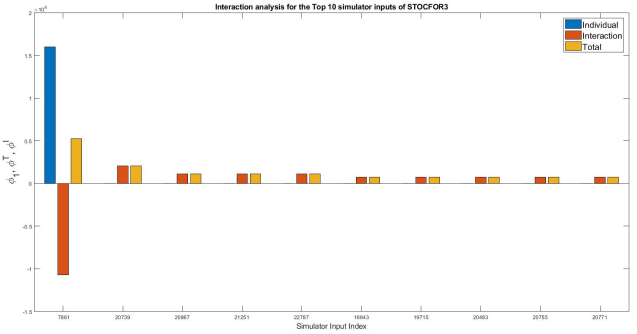


Figure 3.4: Interaction effects for the first ten most important coefficients of the STOCFOR3 simulator.

due to piecewise-definiteness.

Finally, we discuss the analysis of a simulator with a discrete input-output nature. The mapping of interest is the dynamical population model for the Gopherus agassizii desert tortoise by Hodgson and Townley [2004]. The simulator utilizes a Leslie matrix with eight size classes. At each time step, the simulator dynamically computes the promotion rate to the next class and the survival rate within each class. Only the last three classes can produce off-springs, that enter class 1. The spectral radius of the Leslie matrix is close to but lower than unity, so that the population eventually dies out. The output value of the model is the number of time steps for which at least two individuals remain alive. The state vector is rounded down to the next integer to account for the discrete nature of the problem. The model input is the vector of the initial class values, $\mathbf{x}^0 = (89, 163, 62, 27, 16, 13, 29, 5)$.
Given the discrete nature of the problem, an interaction analysis can be carried out at a finite change and a global scale, but a differentiation approach is ruled out.

On a global scale, we use discrete uniform distributions between $\mathbf{x}^0$ and $\mathbf{x}^1$, with $\mathbf{x}^1 = \mathbf{x}^0 + 20$, and independent inputs. At $N = 50,000$, we register

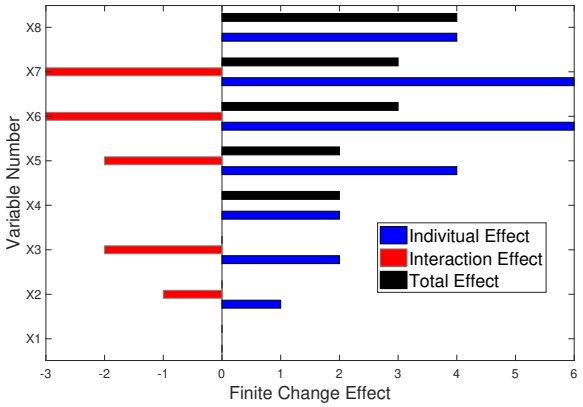


Figure 3.5: Finite change interaction Effects for the Gopherus agassizii simulator of Hodgson and Townley [2004].

variance-based indices equal to $\hat{S}_1 = 0.0014$, $\hat{S}_2 = 0.0032$, $\hat{S}_3 = 0.0045$, $\hat{S}_4 = 0.0277$, $\hat{S}_5 = 0.0773$, $\hat{S}_6 = 0.3103$, $\hat{S}_7 = 0.3134$, $\hat{S}_8 = 0.2350$. The sum of first order indices accounts for about 97% of the simulator output variance, showing that interactions do not play a major role on a global scale. An analysis of interactions on a finite scale for the change $\mathbf{x}^0 \to \mathbf{x}^1$ produces the results in Figure 3.5. Figure 3.5 shows that the simulator responds additively to the change in X_8, but interaction effects are relevant and tend to be opposite to individual effects for all other model inputs. Note that inputs X_2 and X_3 have a negligible first order variance-based index. However, they are associated with a non-negligible first order finite change effect. Thus, insights yielded by an analysis at a global scale do not translate directly at a finite scale for the Gopherus agassizii simulator as well.

3.8 Conclusions

This work has studied the determination of interactions in computer experiments analyzing theoretical as well as implementation aspects. The investigation has brought together several but sparse facts offered in the literature, while proposing new results and definitions towards offering a unified view of the problem.

We have proceeded by distinguishing the analysis of interactions at the infinitesimal, finite and global scales, and we have provided formal bridges and links across the different scales. The analysis has led to the identification of three main interaction generating mechanisms: input correlations, that lead to spurious interactions, explicit interaction terms and piecewise-definiteness, that lead to structural interactions. While the literature examination as well as our results have not revealed alternative interaction generation mechanisms, we cannot exclude other possibilities. In this respect, this work would provide guidance in interpreting such new mechanisms.

On the implementation side, the analysis shows that the statistician must clearly formulate the interaction analysis up front to properly identify the scale and the method. She must also decide up front whether she is performing the analysis on the original scale and whether or not to substitute the original simulator with an emulator.

We note that the formal considerations carried out in this work for deterministic simulators can be applied to stochastic simulators with a white noise as in equation (3.1), if one replaces the deterministic model output y with $\mathbb{E}[Y]$.[1] We lastly observe that, while the analysis of interactions

[1] In fact, given the simulator in equation (3.1), for any value of $\mathbf{X}$ we obtain a conditional distribution of Y given $\mathbf{X}$, because the stochastic error term is not null. That is, in general, a stochastic simulator maps $\mathcal{X}$ onto the space of distributions on $(Y, \mathcal{B}(Y))$. The mapped distribution is conditional on $\mathbf{X} = \mathbf{x}^*$, that is, $F_{Y|\mathbf{X}}(y)$. In the context of stochastic simulation the analyst is often interested in some performance measure built on Y and often this performance measure is the expected value of Y. In this case, we have $\mathbb{E}[Y|\mathbf{X} = \mathbf{x}^0] = g(\mathbf{x}^0)$. Then, when $\mathbf{x}$ shifts from $\mathbf{x}^0$ to $\mathbf{x}^1$, it is $\mathbb{E}[Y|\mathbf{X} = \mathbf{x}^1] = g(\mathbf{x}^1)$. Therefore, $\Delta\mathbb{E}[Y|\mathbf{X} = \mathbf{x}^0] = \mathbb{E}[Y|\mathbf{X} = \mathbf{x}^1] - \mathbb{E}[Y|\mathbf{X} = \mathbf{x}^0] = g(\mathbf{x}^1) - g(\mathbf{x}^0) = \Delta g$. Thus, the definition of interactions in that applied previously to

has concerned computer experiments in this work, the mathematical definitions apply also to the analysis of field experiments. Nonetheless, there are notable technical and interpretation differences between the two, and a complete analysis is out of reach for the present paper. We refer the interested reader to Wu [2015] for a detailed discussion and suggestions of further readings.

3.9 Appendix A: Connection with the Laplace Operator

This section discusses the link between the orthogonalized decomposition of a finite change and the discrete Laplace operator. The discrete Laplace operator has found many applications in Statistics, including Bayesian function estimation on graphs Kirichenko and van Zanten [2017]. We refer to the graphs in Figure 3.2 in the main body of the paper (see Section 3.4). Consider the hypercube having $\mathbf{x}^0$ and $\mathbf{x}^1$ on opposite vertices, i.e., the set of points connected to $\mathbf{x}^0$ by variation in one or more coordinate. This is the set of nodes in graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$:

$$\mathcal{V} = \{\mathbf{x}^0, \overline{\mathbf{x}}^i, \overline{\mathbf{x}}^{\{i,j\}}, \ldots, \overline{\mathbf{x}}^{\{1,2,\ldots,n-1\}}, \mathbf{x}^1\}, \tag{3.38}$$

where the superscripts indicate which coordinates of the base case $\mathbf{x}^0$ are shifted to those of $\mathbf{x}^1$ with $\overline{\mathbf{x}}^z = (\mathbf{x}^0_{-z} : \mathbf{x}^1_z)$. These points are reached from $\mathbf{x}^0$ by changing single coordinates, pairs, triplets, and so on, including the change in all coordinates. The set of edges $\mathcal{E}$ is formed connecting all points in $\mathcal{V}$ to $\mathbf{x}_0$ via one edge. Graph (a) in Figure 3.2 offers a visualization in three dimensions.

Then, let g be a function of the vertices, i.e., $g : \mathcal{V} \to \mathbb{R}$. Given a reference vertex $v \in \mathcal{V}$, one writes the discrete Laplace operator as

$$(\mathcal{L}g)(v) = \sum_{u \sim v} (g(v) - g(u)), \tag{3.39}$$

$g(\mathbf{X})$ applies to $\mathbb{E}[Y|\mathbf{X}]$.

where $u \sim v$ denote the edges of the graph, $(u, v) \in \mathcal{E}$. Then, for the graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ specified above one finds

$$-(\mathcal{L}g)(\mathbf{x}^0) = \sum_{i=1}^{n} \Delta_i g + \sum_{j>i} \Delta_{i,j} g + \cdots + \sum_{i=1}^{n} \Delta_{-i} g + \Delta g. \tag{3.40}$$

Proposition 3.9.1. Given g and $\mathcal{G}(\mathcal{V}, \mathcal{E})$ defined above, if g is additive then

$$(\mathcal{L}g)(\mathbf{x}^0) = -\Delta g \cdot n \cdot (2^n - 1), \tag{3.41}$$

and

$$\sum_{\mathbf{x} \sim \mathbf{x}^0, \mathbf{x} \in \mathcal{V}} g(\mathbf{x}) = (2^n - 1) \left[n \cdot \Delta g + g(\mathbf{x}^0) \right]. \tag{3.42}$$

Equations (3.41) and (3.42) show that, if g is additive, the discrete Laplace operator on $\mathcal{G}(\mathcal{V}, \mathcal{E})$ can be computed by just two evaluations of g, at $\mathbf{x}^0$ and $\mathbf{x}^1$ respectively. Additionally, if $g(\mathbf{x}^0) = 0$, a simple rewriting of equation (3.42) shows that

$$\frac{1}{n} \sum_{\mathbf{x} \sim \mathbf{x}^0, \mathbf{x} \in \mathcal{V}} g(\mathbf{x}) = (2^n - 1)\, \Delta g, \tag{3.43}$$

that is the mean effect of g can be computed using only these two evaluations.

Consider now a subgraph of $\mathcal{G}(\mathcal{V}, \mathcal{E})$, in which the set of nodes contains $n + 2$ vertices determined by $\mathbf{x}^0$, $\mathbf{x}^1$ and all the evaluation points $\overline{\mathbf{x}}^i$, for all $i \in Z$. We denote this graph by $\mathcal{V}'$ (main text Figure 3.2 (b)). Then, each one of these vertices is connected to $\mathbf{x}^0$ by one edge, so that the set of edges has cardinality $n + 1$.

Proposition 3.9.2. Given $\mathcal{G}(\mathcal{V}', \mathcal{E}')$, and $g : \mathcal{V}' \mapsto \mathbb{R}$, if g is additive then

$$(\mathcal{L}g)(\mathbf{x}^0) + 2\Delta g = 0. \tag{3.44}$$

On the other hand, let's consider $\mathcal{G}(\mathcal{V}', \mathcal{E}')$ and a generic mapping g not necessarily additive. Then, the application of the Laplace operator leads to $(\mathcal{L}g)(\mathbf{x}^0) + 2\Delta g \neq 0$, and, specifically,

$$(\mathcal{L}g)(\mathbf{x}^0) + 2\Delta g = \Delta g - \sum_{i=1}^{n} \Delta_i g. \tag{3.45}$$

3.10 Appendix B: Proofs

Proof. Proof of Proposition 3.9.1.

Equation (3.41) holds as $\Delta_z g = g(\bar{\mathbf{x}}^z) - g(\mathbf{x}^0) = \sum_{j=1}^{k} \left(g(x^1_{i_j}) - g(x^0_{i_j}) \right)$. Then, noting that equation (3.39) can be written also as

$$-(\mathcal{L}g)(\mathbf{x}^0) = \sum_{\mathbf{x} \sim \mathbf{x}^0} \big(g(\mathbf{x}) - g(\mathbf{x}^0) \big) = \left(\sum_{\mathbf{x} \sim \mathbf{x}^0} g(\mathbf{x}) \right) - (2^n - 1)\, g(\mathbf{x}^0),$$

we obtain equation (3.42). □

Proof. Proof of Proposition 3.9.2.

Rewriting equation (3.44), one finds $-\left(\mathcal{L}g\right)(\mathbf{x}^0) = \sum_{\{\mathbf{x}^1\} \cup \{\bar{\mathbf{x}}^i, i=1,\dots,n.\}} \big(g(u) - g(\mathbf{x}^0) \big) = g(\mathbf{x}^1) - g(\mathbf{x}^0) + \sum_{i=1}^{n} \Delta_i g$, because all the $n+1$ edges incident to $\mathbf{x}^0$ must be included. Then, $g(\mathbf{x}^1) - g(\mathbf{x}^0) = \Delta g$ by definition, and $\sum_{i=1}^{n} \Delta_i g$ is also equal to Δg by Definition 3.2. □

Lemma 3.10.1. *A function is additive in x_i on $\mathcal{X}$ if and only if it is additively separable between x_i and $\mathbf{x}_{-i}$ on $\mathcal{X}$, that is for all $\mathbf{x}^0, \mathbf{x}^1 \in \mathcal{X}$, $\Delta g = \Delta_i g + \Delta_{-i} g$.*

Proof. Proof of Lemma 3.10.1.

First of all, by mathematical analysis, g is additively separable between x_i and $\mathbf{x}_{-i}$ if

$$g(x^0_i : \mathbf{x}^1_{-i}) + g(x^1_i : \mathbf{x}^0_{-i}) - g(x^0_i : \mathbf{x}^0_{-i}) = g(x^1_i : \mathbf{x}^1_{-i}) \tag{3.46}$$

for all points in $\mathcal{X}$. Let us now suppose that g is additive in x_i, as per definition 3.2.1. Then, we can write:

$$\Delta g = \Delta_i g + \Delta_{-i} g, \tag{3.47}$$

which is equivalent to

$$g(x^1_i : \mathbf{x}^1_{-i}) - g(x^0_i : \mathbf{x}^0_{-i}) = g(x^1_i : \mathbf{x}^0_{-i}) - g(x^0_i : \mathbf{x}^0_{-i}) + g(x^0_i : \mathbf{x}^1_{-i}) - g(x^0_i : \mathbf{x}^0_{-i}). \tag{3.48}$$

Then, rewriting, we have:

$$g(x^1_i : \mathbf{x}^1_{-i}) = g(x^1_i : \mathbf{x}^0_{-i}) + g(x^0_i : \mathbf{x}^1_{-i}) - g(x^0_i : \mathbf{x}^0_{-i}), \tag{3.49}$$

which is (3.46). Conversely, starting with equation (3.46) and subtracting $g(x_i^0 : \mathbf{x}_{-i}^0)$ on both sides, we obtain equation (3.47) □

Proof. Proof of Proposition 3.2.1. By Lemma 3.10.1, if g is additive, it is separately additive in all variables. Then, consider the change $\mathbf{x}^0 \to \mathbf{x}^1$. Then, for every $i = 1, 2, \ldots, n$ we can write $\Delta_g = \Delta_i g + \Delta_{-i} g$. Summing over i, we have $n\Delta_g = \sum_{i=1}^n \Delta_i g + \sum_{i=1}^n \Delta_{-i} g$, which implies $n\Delta_g = \sum_{i=1}^n \Delta_i g + (n-1)\sum_{i=1}^n \Delta_i g$. The converse implication is straightforward. □

Proof. Proof of Theorem 3.2.1.

By contradiction, suppose that exists a transformation η to additivity. By Scheffé [1970], this happens if and only if $\left(\frac{\partial g}{\partial x_i}\right)/\left(\frac{\partial g}{\partial x_j}\right)$ is a function of only x_i, x_j, $\forall i, j$. However, $\left(\frac{\partial g}{\partial x_i}\right)/\left(\frac{\partial g}{\partial x_j}\right)$ is in general function of $\mathbf{x}$ since the expression of the constituents changes according also to the remaining variables $\mathbf{x}_{-ij}$. □

Proof. Proof of Theorem 3.3.1.

If g is of the form $\sum_{i=1}^n g_i(x_i)$ then $g''_{i,j,k}(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}$. Conversely, suppose that the second derivatives are null everywhere. Then, consider the change the function between two points in $\mathcal{X}$. We can expand such change using a Taylor series:

$$g(\mathbf{x}^1) = g(\mathbf{x}^0) + \sum_{i=1}^n g'_i(\mathbf{x}^0)(x_i^1 - x_i^0) + \sum_{i,j=1}^n \frac{1}{2} g''_{i,j}(\mathbf{x}^0)(x_i^1 - x_i^0)(x_j^1 - x_j^0) + \sum_{i,j,k=1}^n \frac{1}{6} g'''_{i,j,k}(\mathbf{x}^0)(x_i^1 - x_i^0)(x_j^1 - x_j^0)(x_k^1 - x_k^0) + ... \tag{3.50}$$

Then, note that the assumption $g''_{i,j}(\mathbf{x}) = 0$ for all $\mathbf{x} \in \mathcal{X}$ implies that the function $g''_{i,j}(\mathbf{x})$ is a constant on $\mathcal{X}$ and infinitely many times differentiable, with null derivatives of all orders. This implies that all mixed derivatives from order 2 onwards are null. Therefore, we have $g(\mathbf{x}) =$

$g(\mathbf{x}^0) + \sum_{i=1}^n g_i'(\mathbf{x}^0)(x_i - x_i^0)$. By the previous expression, one can write

$$g(x^1) - g(\mathbf{x}^0) = \sum_{i=1}^n g_i'(\mathbf{x}^0)(x_i - x_i^0).$$

Now, observe that

$$g(x_i : \mathbf{x}_{-i}^0) - g(\mathbf{x}^0) = g_i'(\mathbf{x}^0)(x_i - x_i^0) + \sum_{i=1}^n \frac{1}{2} g_i''(\mathbf{x}^0)(x_i - x_i^0)^2 + \sum_{i=1}^n \frac{1}{6} g_i'''(\mathbf{x}^0)(x_i - x_i^0)^3 + \dots$$

Again, by the nullity of the second and higher order derivatives, we obtain

$$g(x_i : \mathbf{x}_{-i}^0) - g(\mathbf{x}^0) = g_i'(\mathbf{x}^0)(x_i - x_i^0).$$

Then, this shows that

$$g(x_i : \mathbf{x}_{-i}^0) - g(\mathbf{x}^0) = g_i'(\mathbf{x}^0)(x_i - x_i^0),$$

and therefore

$$g(\mathbf{x}^1) - g(\mathbf{x}^0) = \sum_{i=1}^n g_i'(\mathbf{x}^0)(x_i - x_i^0) = \sum_{i=1}^n g(x_i : \mathbf{x}_{-i}^0) - g(\mathbf{x}^0),$$

□

Proof. Proof of Proposition 3.5.1.

By definition,

$$\mathbb{V}[Y|X_i] = \mathbb{E}\left[(Y - \mathbb{E}[Y|X_i])^2 \, |X_i \right]. \tag{3.51}$$

Now, $\mathbb{E}[Y|X_i = x_i]$ is $g_0 + g_i(x_i)$ so that $Y - \mathbb{E}[Y|X_i = x_i] = \sum_{j \neq i} g_j(x_j) + \sum_{|\alpha| \geq 2} g_\alpha(x_\alpha)$. Hence,

$$\mathbb{V}[Y|X_i] = \mathbb{E}\left[\left(\sum_{j \neq i} g_j(x_j) + \sum_{|\alpha| \geq 2} g_\alpha(x_\alpha) \right)^2 \middle| X_i \right] \tag{3.52}$$

Then, due to the orthogonality of the functions in (3.52), we can write

$$\mathbb{V}[Y|X_i] = \sum_{j \neq i} \mathbb{E}\left[g_j^2(x_j) \right] + \sum_{|\alpha| \geq 2} \mathbb{E}\left[g_\alpha(x_\alpha)^2 | X_i \right]. \tag{3.53}$$

Now, the first sum, $\sum_{j\neq i}\mathbb{E}[g_j^2(x_j)]$, does not depend on X_i. Thus, if only these terms were present in the decomposition, we would not have heteroskedasticity. Under heteroskedasticity, at least one of the terms $\mathbb{E}[g_\alpha(x_\alpha)^2|X_i]$ with $i\in\alpha$ has to be different from zero, because $\mathbb{V}[Y|X_i]$ is not constant. Thus, we have interaction terms in the function and these terms involve X_i. Conversely, if there is at least one multi-index α with $|\alpha|\geq 2$ and $i\in\alpha$, we have that $\mathbb{V}[Y|X_i]$ varies with X_i and, therefore, there is heteroskedasticity. □

Proof of Corollary 3.5.1. Based upon the variance decomposition formula, $\mathbb{E}[\mathbb{V}[Y|X_i]]\leq\mathbb{V}[Y]$ holds. Then, if there is a point where $\mathbb{V}[Y|X_i=x_i]>\mathbb{V}[Y]$, the conditional variance cannot be a constant. □

Proof of Proposition 3.5.2. Consider the set $\Omega=\bigcup_{l,m}\partial\mathcal{X}_l\cap\partial\mathcal{X}_m$, where ∂ denoted the frontier. Since $\Pi_{\mathcal{X}}$ is a partition, this set has Lebesgue-measure zero. Hence, $\nu_{i,j}=\mathbb{E}\left[\left(g''_{i,j}(\mathbf{X})\right)^2\right]=0$ by the additivity of the component functions in (3.5). □

Proof of Theorem 3.5.1. By Theorem 1 in Liu and Owen [2006], the superset importance can be written as

$$\Upsilon_u^2=\frac{1}{2^{|u|}}\int\left(\sum_{v\subseteq u}(-1)^{|u-v|}g(\mathbf{x}_v,\mathbf{z}_{-v})\right)^2 d\mathbf{x}^u d\mathbf{z}.$$

We note that, rewriting the expression of $\tau_u^{\mathbf{x}\to\mathbf{z}}$, we find that $\tau_u^{\mathbf{x}\to\mathbf{z}}=\sum_{v\subseteq u}(-1)^{|u-v|}g(\mathbf{x}_v,\mathbf{z}_{-v})$. Using this equality, it follows that

$$\Upsilon_u^2=\frac{1}{2^{|u|}}\int\left(\tau_u^{\mathbf{x}\to\mathbf{z}}\right)^2 d\mathbf{x}^u d\mathbf{z},$$

proving the first statement in (3.28). The second statement can be proved rewriting the estimator of the Sobol index

$$\underline{\tau}_u^2=\mathbb{E}[g(\mathbf{X})(g(\mathbf{X}_u:\mathbf{Z}_{-u})-g(\mathbf{Z}))]=\sum_{v\subseteq u}\mathbb{E}\left[g(\mathbf{X})\tau_v^{\mathbf{Z}\to\mathbf{X}}\right].$$

The third statement follows analogously as

$$\overline{\tau}_u^2 = \tfrac{1}{2}\mathbb{E}[(g(\mathbf{Z}_u : \mathbf{X}_{-u}) - g(\mathbf{X}))]^2 = \frac{1}{2}\mathbb{E}\left[\sum_{v \subseteq u} \tau_v^{\mathbf{X}\to\mathbf{Z}}\right]^2 .$$

In order to prove (3.29), by Theorem 2.2 in Owen [2013] it holds that $\sum_u |u| V_u = \sum_{i=1}^n \overline{\tau}_i^2$. By (3.28), it becomes

$$\sum_u |u| V_u = \frac{1}{2}\sum_{i=1}^n \mathbb{E}\left[\tau_i^{\mathbf{X}\to\mathbf{Z}}\right]^2 .$$

For the other equality in (3.29), we consider the equation $\sum_{i=1}^n V_i = \sum_{i=1}^n \underline{\tau}_i^2$ at page 32 in Owen [2013] and plug in the estimator for $\underline{\tau}_i^2$ in (3.28). □

Chapter 4

Screening with finite changes: from Elementary Effects to Effective Dimension

Abstract

Screening is an essential task in simulation. Several successful methods are based on one-at-a-time designs. This work shows that by properly exploiting these designs one obtains not only global measures of the relevance of inputs, but also of the relevance and sign of interactions. We follow a files-rouge that links the sensitivity measures of Tornado diagrams to the Morris method and to global sensitivity measures. Specifically, in a different approach as in Campolongo et al. [2011], we show that replicates of one-at-a-time effects yield estimators of Sobol' total indices, individual and interaction indices as well as of the mean effective dimension, and these quantities can be obtained from a suitably defined variance-covariance

matrix of one-at-a-time effects. It is then possible to prove the exact relationship between Morris and Sobol' total sensitivity indices. We obtain asymptotic results that yield confidence intervals around the estimates at finite sample sizes. The new insights are illustrated by application to well known simulators such as the Asian option, the assemble-to-order and the space probabilistic safety assessment codes.

4.1 Introduction

Computer simulations support analysts in understanding and describing the behavior of complex operational systems Nelson [2013]. The steady increase in computing power allows greater sophistication in modeling and analysts can address systems of increasing complexity. This complexity, however, leads analysts to face transparency and interpretability issues. An important recommendation that emerges in the simulation literature is that an analyst should frame the simulation up front Kleijnen [2017], for *designing an experiment that will yield the desired information* [Conway, 1963, p. 47]. Proper planning allows the analyst to better exploit the simulator runs: she can not only understand the time and resources needed to estimate a system performance metric reliably, but she can also prepare for extracting additional managerial/modelling insights. In this context, designs aimed at reducing problem complexity and dimensionality (screening designs, henceforth) play an essential role. Identification of the most important inputs simplifies result communication, evidencing the factors on which to focus managerial attention. Also, dimension reduction allows the fitting of metamodels, that, in turn, enable deeper analyses to be performed in a time efficient manner [Wan et al., 2006, Kleijnen, 2017].

When dimension reduction is concerned, analysts ought to compute indicators of the relevance and size of interactions. Best practice would require the estimation of quantities such as total variance-based sensitivity indices and effective dimensions (see [Owen, 2003] among others), which pertain to the realm of global sensitivity analysis. There has been considerable interest in the literature in investigating the link between these

measures and screening designs [Campolongo et al., 2007, 2011], and, in particular with the Morris method [Morris, 1991], that, together with sequential bifurcation [Bettonvil and Kleijnen, 1997], is one of the most popular screening methods in simulation [Shi and Chen, 2018] (even if sometimes not suggested - see Campolongo et al. [2011]).

However, currently there is no formal way to connect the sensitivity measures of the Morris method with the notion of effective dimensions. At the same time, the link between Morris second sensitivity measure (called here $\sigma^2_{EE_i}$; see Section 4.3.2 for the definition) and total order sensitivity indices has been hinted in previous works [Campolongo et al., 2011, Shi and Chen, 2017] but not formally established. Moreover, the number of replicates needed to link screening with global approaches might be high. Thus, it is important for the analyst to have a measure of uncertainty in the estimates at finite sample sizes.

The objective of this Chapter is to develop a statistical analysis based on finite changes to understand the common features of the Sobol' and the Morris' importance indices.

In particular, we revisit the Morris method, with the goal of filling-in the above-mentioned gaps. We observe that several designs in the literature are based on one-at-a-time input variations, with or without replicates (the corresponding sensitivity measures are called main effects, henceforth). We start analyzing the statistical properties of main effects. We calculate the mean and variance of finite changes both under random sampling and under Morris' design. We show that the use of different designs lead to different estimates of the mean and variance of the main effects. Then, we show that Sobol' total order variance-based sensitivity indices as well as Owen superset variance-based indices of pairs [Liu and Owen, 2006] can be estimated from the empirical covariance matrix of main effects. This is appealing since we can estimate the interaction structure and total effects simply by looking at variations of one-factor-at-a-time designs in the input space. Thus, the analyst obtains a detailed picture of the relevance of interactions and the dimensionality of the simulator from replicates of one-at-a-time effects. This result is then used

to establish the formal relationship between Morris $\sigma^2_{EE_i}$ and Sobol' total order indices, justifying the interpretation of $\sigma^2_{EE_i}$ as a measure of interaction and non-linearity proposed in previous works [Campolongo et al., 2011, Shi and Chen, 2017].

From a computational viewpoint, we show that the estimation can be made more efficient by exploiting a symmetry property of the main effects, which we use to obtain an alternative estimator of the mean dimension. Simultaneously, the analyst obtains a sample of overall interaction effects, that deliver information on the sign of interactions. The computational cost is the cost of the Morris method plus one simulator run.

We then perform an asymptotic analysis, which yields the following main insights. First, Morris $\sigma^2_{EE_i}$ is a biased estimator of the total Sobol' indices, if replicates are performed according to the original Morris trajectories. However, it becomes an unbiased estimator, if trajectories are randomly sampled across the whole input space (see, e.g. the radial design of Campolongo et al. [2011]). Second, at any finite sample size the analyst obtains confidence intervals that allow her to quantify uncertainty in the estimates of total order Sobol' indices, mean dimension, first order Sobol' indices and Morris μ_i. Third, findings in Cheng [1997] and Boukouvalas et al. [2014] are extended to generic input-output mappings.

We carry out numerical experiments for a series of case studies, and discuss results for the Asian option code developed in Nelson [2013], the Assemply-to-Order simulator used in Hong and Nelson [2006] and the NASA space probabilistic safety assessment code developed in Borgonovo and Smith [2011]. The design yields insights not only into the key-drivers of simulator variability, but also into the sign, relevance and order of interactions, as well as on the direction of change in the model output associated with the variation of each input.

Table 4.1: A brief summary of methods for screening simulation experiments.

Without Replicates	**Without Metamodel**
Tornado Diagrams	Howard [1988], Borgonovo and Smith [2011]
Without Replicates	**With Metamodel**
One-at-a-time	Daniel [1973], Cotter [1979]
Sequential Bifurcation	Bettonvil and Kleijnen [1997]
With Replicates	**Without Metamodel**
Morris Method	Morris [1991], Campolongo et al. [2011], Shi and Chen [2017]
	With Metamodel
Kriging	Schonlau and Welch [2006], Binois et al. [2019]

4.2 Simulation Input Screening: Related Literature

The literature on simulation experiments is vast and space limitations do not allow us a complete investigation. For broad overviews of the field, we refer to Chick [2001], Nelson [2004] among others, and to the monographs of Nelson [2013], Kleijnen [2015]. When screening is concerned, the simulation literature is entangled with the statistical literature of computer experiments as the overviews of Chen et al. [2016], Kleijnen [2017] show. Although originally developed for physical experiments [Woods and Lewis, 2017], screening techniques are pervasive in simulation experiments. Table 4.1 presents a summary of the methods we review here. The grouping is for expository purposes only and does not aim to be a categorization.

Designs that allow the determination of all main and interaction effects are full factorial designs [Kleijnen, 2017]. However, early on it has been recognized that the computational cost of full factorial designs can make them impractical [see [Jacoby and Harrison, 1962, p. 121]]. This has motivated the search for designs that allow analysts to diminish computational burden for the identification of the most important inputs (screen-

ing methods). The simplest designs foresee the variation of one input at a time. In the statistical literature, they are proposed and analyzed in early works such as Daniel [1973], Cotter [1979], Baker and Bargmann [1985]. These methods are based on the variation of one input at a time when the response is assumed to follow a linear model with interactions and their sensitivity measures are the difference between: a) the linear model output value when one of the inputs is shifted at the sensitivity case; and b) the linear model output value when all the inputs are at the base case. These sensitivity measures are called main effects. Cotter [1979] presents an extension of one-at-a-time designs using a fold-over approach in which the factors are varied not only from the base case to the sensitivity case (forward pass) but also from the sensitivity to the reference case (backward pass). As discussed later on by Qu and Wu [2005], Tang and Xu [2014], this fold-over property allows one to increase the design resolution, minimizing the contamination of high-order effects on the estimation of main effects.

In the simulation literature, one of the most successful screening methods based on a metamodel is sequential bifurcation [Bettonvil and Kleijnen, 1997, Kleijnen et al., 2006]. In this approach, one assumes a multilinear metamodel, either additive or with interaction terms. The simulator inputs are fixed at two levels (called high and low). The procedure foresees the initial grouping of the simulator inputs into two (usual) or more groups. The simulator is evaluated with all the inputs in a group at the low or at the high level. The group that leads to the highest change in simulator output is regarded as important and is further split into subgroups. The procedure then continues with group splitting until the main effects of important factors are determined. The method performs at best when the assumption of monotonicity in each of the inputs is satisfied [Shi and Kleijnen, 2018]. Over the years sequential bifurcation has been refined in several directions. Please refer to the works of Cheng [1997], Wan et al. [2006] and Ankenman et al. [2014] (see also [Kleijnen, 2015, ch. 4]) for further details.

A further important class is represented by design for screening while

applying kriging emulation (see the works of Schonlau and Welch [2006], Kleijnen [2015], Woods and Lewis [2017], Binois et al. [2019] for overviews).

In parallel, one can find screening methods without metamodels. In the design at basis of Tornado diagrams the sensitivity measures of Daniel [1973], Cotter [1979] are computed without assuming a metamodel (and without replicates). (Tornado diagrams are a well known graphical sensitivity analysis tool in the management science literature since the work of Eschenbach [1992] and are implemented in popular software packages such as Treeage, @Risk by Palisade.) Generalized Tornado Diagrams [Borgonovo and Smith, 2011] allow one to include sensitivity measures for interactions using a design that resembles the fold-over intuition of Cotter [1979]. A successful screening approach with replicates but without metamodel is the Morris method [Morris, 1991]. The Morris design foresees the selection of an appropriate grid of points in the model input space, and then the sampling of a trajectory on this grid. Main effects are computed from a series of d one-at-a-time input changes in correspondence of each point in the trajectory (d denotes the number of simulator inputs). The method has been the subject of intensive investigations over the years. Recently, Boukouvalas et al. [2014] introduce a sequential procedure for the selection of the number of replicates, allowing this number to vary across the simulator inputs. Shi and Chen [2017] introduce sequential statistical tests for controlling the probability of type I and type II errors in the estimation of main and interaction effects (the so-called controlled Morris method). Ge and Menendez [2017] discuss the computation of Morris elementary effects in the context of simulation experiments with dependent inputs. After a first extension in Alam et al. [2004], Shi and Chen [2018] thoroughly address the properties of Morris method for simulators with a stochastic response, proposing several results for the optimal allocation of model runs between the inner and outer loops of a stochastic simulation.

As Morris and Moore [2015] underline, it is a small step from screening to sensitivity analysis. If replicates of the screening design are computed over the whole input space and the corresponding sensitivity estimators are averaged, then one obtains a bridge towards estimating global sensi-

tivity indices. Several works have explored this connection under different perspectives, and reviews can be found in [Saltelli et al., 1995, Campolongo F et al., 2000, Becker and Saltelli, 2015, Kleijnen, 2015, Woods and Lewis, 2017]. Relevant to our work are the findings in [Campolongo et al., 2007, 2011], where a radial design with replicates is used to estimate both Morriss sensitivity measures and total order variance-based sensitivity indices. In the work, the close connection between Morris $\sigma^2_{EE_i}$ and Sobol' total indices is underlined, although not formally established.

The previous analysis highlights a files-rouge across the methods: in all the above-mentioned designs the sensitivity measures are main effects built after computing one-at-a-time sensitivities, with or without replicates. In the reminder, we show that deeper insights can be extracted from these designs by exploiting some symmetry effects that we are to analyze.

4.3 Screening: the Role of Main Effects

Let $\mathcal{X}$ denote the simulator input space, and let $\mathcal{X} \subseteq \mathbb{R}^d$, where d is the number of simulator inputs (or the dimension of the simulator). If the simulator inputs are uncertain, we denote by $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \mathbb{P}_{\mathbf{X}})$ the simulator input probability space and by $F_{\mathbf{X}}$. Then, we let

$$Y = g(\mathbf{X}) + \epsilon(\mathbf{X}) \tag{4.1}$$

denote the simulator input-output mapping. In equation (4.1), g is a multivariate function $g : \mathcal{X} \mapsto \mathbb{R}$, and $\epsilon(\mathbf{X})$ is a stochastic error term such that $\mathbb{E}[\epsilon(\mathbf{X})] = 0$. The simulator is called deterministic if this term is null, stochastic otherwise. The error term can be homoskedastic (i.e., independent of $\mathbf{X}$) or heteroskedastic. Then, we note that by equation (4.1), we have $\mathbb{E}[Y|\mathbf{X} = \mathbf{x}^0] = g(\mathbf{x^0}) + \mathbb{E}[\epsilon(\mathbf{x}^0)]$. If we write this expression as $\mathbb{E}[Y|\mathbf{X} = \mathbf{x}^0] = \widetilde{g}(\mathbf{x^0})$, then, the considerations we are to draw in the reminder hold for a deterministic simulator as well as for the expected value of a stochastic simulator with the type of response in (4.1). Nonetheless, also for notation simplicity, we shall maintain the symbol $g(\mathbf{x^0})$. When the simulator is stochastic one often adopts the symbol θ in lieu of $\mathbf{x}$, and

the term parameters instead of the term simulator inputs. We shall use these two terms equivalently in the reminder.

4.3.1 Designs without replicates

Consider now two possible simulator input levels $\mathbf{x}^0, \mathbf{x}^1 \in \mathcal{X}$ and the corresponding simulator output responses $g(\mathbf{x}^0)$ and $g(\mathbf{x}^1)$. We can consider $\mathbf{x}^0$ and $\mathbf{x}^1$ as base case and sensitivity case, respectively. When the simulator inputs vary from the base to the sensitivity case ($0 \to 1$, henceforth) the change in model output is $\Delta^{0\to1}g = g(\mathbf{x}^1) - g(\mathbf{x}^0)$. This is equivalent to evaluating the simulator on two opposite vertices of the hypercube (see the left graph of Figure 4.1 for a two dimensional representation).

$\Delta^{0\to1}g$ can be dissected in $2^d - 1$ orthogonal effects [Li et al., 2001, Borgonovo, 2010]:

$$\Delta^{0\to1}g = \sum_{i=1}^{d} \phi_i^{0\to1} + \sum_{i<j} \phi_{i,j}^{0\to1} + \cdots + \phi_{1,2,\ldots,d}^{0\to1} \tag{4.2}$$

where

$$\begin{cases} \phi_i^{0\to1} = g\left(x_i^1 : \mathbf{x}_{-i}^0\right) - g\left(\mathbf{x}^0\right) \\ \phi_{i,j}^{0\to1} = g\left(x_{i,j}^1 : \mathbf{x}_{-\{i,j\}}^0\right) - g\left(\mathbf{x}^0\right) - \phi_i^{0\to1} - \phi_j^{0\to1} \\ \phi_{i,j,k}^{0\to1} = g\left(x_{i,j,k}^1 : \mathbf{x}_{-\{i,j,k\}}^0\right) - g\left(\mathbf{x}^0\right) - \phi_i^{0\to1} - \phi_j^{0\to1} - \phi_k^{0\to1} - \phi_{i,j}^{0\to1} - \phi_{j,k}^{0\to1} - \phi_{i,k}^{0\to1} \\ \\ \ldots \end{cases} \tag{4.3}$$

In the above equation, $\left(x_i^1 : \mathbf{x}_{-i}^0\right)$ denotes the point in the model input space in which all inputs are at the base case with the exception of x_i (see the left graph in Figure 4.1 for a visualization in two dimensions).

The effect $\phi_i^{0\to1}$ is the main effect of x_i for the shift $\mathbf{x}^0 \to \mathbf{x}^1$. Similarly, $(x_{i,j}^1 : \mathbf{x}_{-\{i,j\}}^0)$ denotes the point in which all inputs are at the base case, but for the pair x_i and x_j. The second order effect $\phi_{i,j}^{0\to1}$ quantifies the residual interaction given that both x_i and x_j are shifted at the sensitivity case, and given that their individual effects are subtracted from

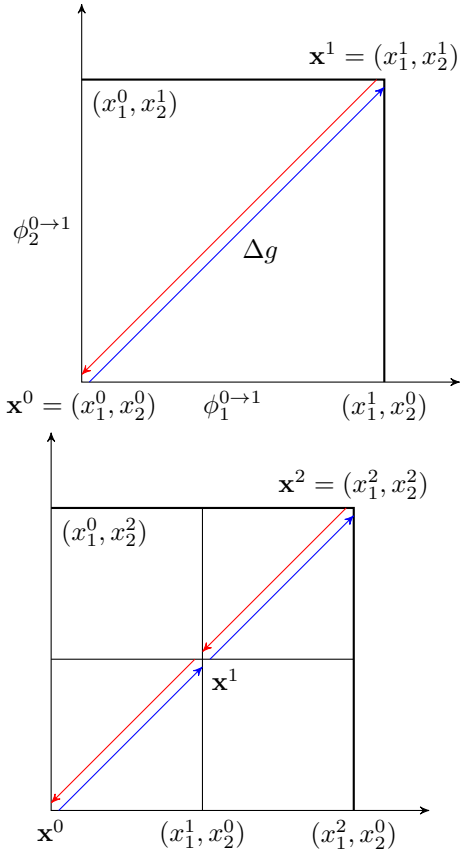


Figure 4.1: Non replicated OAT design (upper graph). Replicated OAT design (bottom graph).

the change $g(x^1_{i,j} : \mathbf{x}^0_{-\{i,j\}}) - g\left(\mathbf{x}^0\right)$. Let then $u \subset \{1, 2, ..., d\}$ denote a set of indices. The quantity $\phi_u^{0\to1}$ is a measure of the contribution of the residual interaction of the indices in u to the finite change $\Delta^{0\to1}g$. One can then define the total contribution of x_i to $\Delta^{0\to1}g$ as

$$\overline{\tau}_i^{0\to1} = \sum_{i\in u} \phi_u^{0\to1}, \tag{4.4}$$

and the total interaction contribution as

$$\Upsilon_i^{0\to1} = \overline{\tau}_i^{0\to1} - \phi_i^{0\to1}. \tag{4.5}$$

Determining all the $\phi_u^{0\to1}$ effects in (4.3) requires the evaluating the simulator on all the vertices of the hypercube joining $\mathbf{x}^0$ and $\mathbf{x}^1$, which corresponds to a full factorial design [Daniel, 1973]. As already underlined in Jacoby and Harrison [1962] a full factorial design becomes rapidly infeasible. Therefore, the literature has proposed designs to reduce computational burden while determining interaction effects.

In the design of Cotter [1979], one evaluates the first order effects for the change $\mathbf{x}^0 \to \mathbf{x}^1$ and for the opposite change $\mathbf{x}^0 \to \mathbf{x}^1$, obtaining the set of effects: $\phi_i^{0\to1}$ and $\phi_i^{1\to0}$, $i = 1, 2, ..., d$. These designs are then constructed by (i) one run $g(\mathbf{x}^0)$, (ii) d runs to compute $g(x_i^1 : \mathbf{x}^0_{-i})$ for all i, (iii) d runs $g(x_i^0 : \mathbf{x}^1_{-i})$, and (iv) one run $g(\mathbf{x}^1)$. We now show that it is possible to exploit the fold-over design of Cotter to obtain information about interactions.

Proposition 4.3.1. From Cotter's fold-over design, it is possible to estimate the triplet of indices $\phi_i^{0\to1}$,$\Upsilon_i^{0\to1}$ and $\overline{\tau}_i^{0\to1}$.

Note that, if $d = 2, 3$, this design is equal to a full factorial design and, thus, one is actually estimating all individual and interaction effects in 4.3.

4.3.2 Introducing Replicates

One key-intuition common across several designs is to introduce replicates [Kleijnen, 2015]. Replicates allow one to compute main or interaction effects at several locations $\mathbf{x}^k, k = 0, 1, \ldots, N$ in $\mathcal{X}$. These locations can be

selected in alternative ways. One can sample N points randomly using crude, Quasi-Monte Carlo [Wang and Sloan, 2011] or Latin-Hypercube designs or others [Pronzato and Müller, 2012]. We call trajectory the sequence of points $\mathcal{T} = \{\mathbf{x}^0 \to \mathbf{x}^1 \to \cdots \to \mathbf{x}^N\}$. To illustrate, in the method of Morris, the model input space $\mathcal{X}$ is reported to the unit hypercube so that $\mathcal{X} = [0,1]^d$ and each interval is split into p levels. The right graph in Figure 4.1 provides a visual representation for p=3. This creates a grid $\mathcal{G}$ of p^d points. Trajectories are then sequences of points randomly sampled from $\mathcal{G}$; that is, in the Morris design, $\mathcal{T} \subset \mathcal{G}$. Evaluating the model at $\mathcal{T}$, one obtains a set of N output changes $\Delta^{r \to r+1} g$, $r = 0, 1, ..., N-1$. Because each of these changes can be decomposed via (4.2), in principle one obtains a set of $2^d \cdot N$ effects $\phi_u^{k \to k+1}$. However, this quantification becomes rapidly impractical, because the number of simulator evaluations grows exponentially with d. If, instead, one considers computing only main effects, she can run the code $(d+1) \cdot N$ to obtain N replicates of d main effects. In this respect, we regard main effects as a population, and denote the corresponding random vector with $\mathbf{\Phi} = \{\Phi_1, \Phi_2, \ldots, \Phi_d\}$.

Morris [1991] proposes to normalize the main effects into Newton ratios, defining the elementary effect for the $k \to k+1$ change as

$$EE_i^{k \to k+1} = \frac{\phi_i^{k \to k+1}}{\Delta}, \tag{4.6}$$

where $\Delta = x_i^1 - x_i^0$. Precisely, in the Morris design one typically chooses $\Delta = p(2(p-1))^{-1}$. The suggestion is then to take the mean of the elementary effects as importance measure:

$$\hat{\mu}_i = \frac{\sum_{k=0}^{N-1} EE_i^{k \to k+1}}{N}, \quad i = 1, 2, ..., d. \tag{4.7}$$

Campolongo et al. [2011] propose the absolute value of Morris elementary effect to define the importance measure

$$\hat{\mu}_i^* = \frac{\sum_{k=1}^{N-1} \left| EE_i^{k \to k+1} \right|}{N}, \tag{4.8}$$

to take into account potential annihilating effects in equation (4.6) when g is not monotonic (see also Becker and Saltelli [2015]). Because equation

(4.7) is prevalent in the simulation arena (see the recent studies in Shi and Chen [2018]), we will privilege such sensitivity measure in the reminder. Morris also proposes the sensitivity measure

$$\hat{\sigma}^2_{EE_i} = \frac{\sum_{k=1}^{N} \left(EE_i^{k\to k+1} - \hat{\mu}_i\right)^2}{N-1} \tag{4.9}$$

as a measure of non linearity and interaction.

Morris $\hat{\mu}_i$ and $\hat{\sigma}_{EE_i}$ are related to $\mathbf{\Phi}$ as follows:

$$\hat{\mu}_i = \frac{\overline{\phi}_i}{\Delta}, i = 1, 2, ..., d, \tag{4.10}$$

where $\overline{\phi}_i = \frac{\sum_{k=0}^{N-1} \phi_i^{k\to k+1}}{N}$, and

$$\hat{\sigma}^2_{EE_i} = \frac{\sum_{k=1}^{N} \left(\phi_i^{k\to k+1} - \hat{\mu}_i\Delta\right)^2}{(N-1)\Delta^2}. \tag{4.11}$$

Now, let us explore the idea of a folding-over a design with replicates. For each trajectory $\mathbf{x}^0 \to \mathbf{x}^1 \to ... \to \mathbf{x}^N$ we consider also the reverse trajectory $\mathbf{x}^N \to \mathbf{x}^{N-1} \to ... \to \mathbf{x}^0$ and estimates elementary effects also on this trajectory. Then, we calculate $2N$ main effects. However, the indices $\phi_i^{k\to k+1}$ and the corresponding $\phi_i^{k+1\to k}$ are paired. By Proposition 4.3.1, we can compute N first order, N total and N overall interaction effects ($\Upsilon_i^{0\to 1}$) at a cost of $(2d+2)\cdot N$ model runs. We remark that this approach gains in economy but loses in explorativity of the input space. In the next sections, we further explore how we can use this information.

4.4 Morris $\widehat{\sigma}^2_{EE_i}$ and Total Sobol' Indices

To prove the results of this section, we need to link OAT designs to variance decomposition. If $g \in \mathcal{L}^2(\mathcal{X}, \mathcal{B}(\mathcal{X}), F_{\mathbf{X}})$, and $F_{\mathbf{X}}$ is a product measure, the classical results of Efron and Stein [1981], Sobol' [1993], Wang [2006] allows us to write the variance of Y as:

$$\sigma^2_Y = \sum_u \sigma^2_u, \tag{4.12}$$

where

$$\sigma_u^2 = \int \left[g_u(\mathbf{x}_u)\right]^2 dF_u(\mathbf{x}_u), \tag{4.13}$$

and g_u is obtained from

$$g_u(\mathbf{x}_u) = \int g(\mathbf{x}_u : \mathbf{s}_{-u}) dF_{-u}(\mathbf{s}_{-u}) - \sum_{v \subset u} g_v(\mathbf{x}_v). \tag{4.14}$$

The variance-decomposition in (4.12) is analogous to the finite change decomposition in (4.2). The variance of Y is decomposed in 2^d terms, with the first order terms σ_i^2 and the interaction terms σ_u^2 apportioning σ_Y^2 to the individual contribution of X_i and the contribution of the residual interaction among the inputs in u. One then defines (see Wagner [1995], Homma and Saltelli [1996], Liu and Owen [2006] among others) the triplets of global (or variance-based) sensitivity indices:

$$\begin{aligned} \underline{\tau}_i^2 &= \sigma_i^2, \\ \overline{\tau}_i^2 &= \sum_{i \in v} \sigma_v^2, \end{aligned} \tag{4.15}$$

and

$$\Upsilon_i^2 = \overline{\tau}_i^2 - \underline{\tau}_i^2, \tag{4.16}$$

where $\underline{\tau}_i^2$ is the first order variance based index of X_i, $\overline{\tau}_i^2$ is Homma and Saltelli [1996]'s total index and Υ_i^2 is Liu and Owen [2006]'s interaction index. Note that $\overline{\tau}_i^2$ is the sum of all terms in the right hand side of (4.12) that contain a contribution from X_i, while Υ_i^2 contains only the contributions coming from subsets of cardinality greater than 1, thus arising from interactions with other inputs. In the literature one often finds the normalized version of these indices. They are called Sobol' indices and denoted by S_u.

We are then ready to link main effects to variance-based indices and to the notion of mean effective dimension. In this section, we focus on the total indices $\overline{\tau}_i^2$ and, specifically, we show that with a replicated fold-over design we are able to obtain estimators of the total $\overline{\tau}_i^2$ and first order indices τ_i^2.

Proposition 4.4.1. Let $g \in \mathcal{L}^2(\mathcal{X}, \mathcal{B}(\mathcal{X}), F_{\mathbf{X}})$ where $F_{\mathbf{X}}$ is a product measure and suppose that $\mathbf{X}, \mathbf{X}^1 \sim F_{\mathbf{X}}$ are independent. The variance-based indices $\overline{\tau}_i^2$ and $\underline{\tau}_i^2$ can be written as

$$\overline{\tau}_i^2 = \frac{1}{2}\mathbb{E}\left[\left(\Phi_i^{\mathbf{X}\to\mathbf{X}^1}\right)^2\right] \tag{4.17}$$

and

$$\underline{\tau}_i^2 = \mathbb{E}\left[g(\mathbf{X}) \cdot \Phi_i^{\mathbf{X}^1\to\mathbf{X}}\right]. \tag{4.18}$$

Corollary 4.4.1. *Assume that* $\left(\mathbf{x}^0, \mathbf{x}^1\right)_{(k)} \sim \left(F_{\mathbf{X}}, F_{\mathbf{X}}\right), k = 1, ..., N,$ *are an iid sample. Then, Monte Carlo estimators of* $\overline{\tau}_u^2$ *and* $\underline{\tau}_u^2$ *are respectively*

$$\hat{\overline{\tau}}_i^2 = \frac{\sum_{k=0}^{N-1}\left(\phi_i^{k\to k+1}\right)^2}{2N} \tag{4.19}$$

and

$$\hat{\underline{\tau}}_i^2 = \frac{\sum_{k=0}^{N-1} g(\mathbf{x}_{(k)}^0) \cdot \left(\phi_i^{k+1\to k}\right)}{N}. \tag{4.20}$$

Proposition 4.4.1 and Corollary 4.4.1 state the following. A forward pass, $\mathbf{x}^0 \to \mathbf{x}^1 \to ... \to \mathbf{x}^N$, of a replicated fold-over design allows us to obtain a set of first order finite change effects. By squaring and averaging them, we obtain an estimator of the total variance-based sensitivity indices (4.19). At the same time, by multiplying the main effects of the backward pass $\phi_i^{k+1\to k}$ by $g(\mathbf{x}_{(k+1)}^0)$ and averaging, we obtain an estimate of the first order variance-based sensitivity indices $\hat{\overline{\tau}}_i^2$.

Then, by taking the difference $\hat{\overline{\tau}}_i^2 - \hat{\tau}_i^2$, we obtain an estimator of the overall variance-based interaction index $\hat{\Upsilon}_i^2$. Therefore, Proposition 4.4.1 and Corollary 4.4.1 suggest that with a forward and a backward trajectory, we are able to estimate the triplets of global first, total and interaction indices for all simulator inputs. The cost of this design is $2N(d+1)$ model runs.

4.5 Interaction Indices and Effective Dimensions from Main Effects

Caflisch et al. [1997], Owen [2003] introduce the notion of dimension distribution to help the analyst in understanding the relevance of interactions in the variance decomposition in (4.12). These works have been originally motivated by the study of the performance of Quasi-Monte Carlo (QMC) methods, and, in particular, of the conditions under which QMC outperforms simple Monte Carlo. The notion of dimension distribution and effective dimensions provide the theoretical background for such investigation and have found notable application in financial engineering. we refer to works such as Wang [2006], Wang and Sloan [2011], Wang and Tan [2013], Wang and Leng [2016] for a thorough overview. In this work, we focus on the mean dimension in the superimposition sense.
Formally, the mean dimension of g in the superimposition sense is defined as

$$D_g = \frac{\sum_u |u|\sigma_u^2}{\sigma^2}, \tag{4.21}$$

where $|u|$ is the cardinality of u. The value of D_g is the average of the cardinality of a group of indices weighted by its fractional contribution to the variance of the model output. For characterizing interactions, it is interesting to consider the sum $\sum_{i=1}^{d} \Upsilon_i^2$. This quantity is null if g is additive and, if it is different from zero, it is a measure of the total overall impact of interaction effects in g.

Proposition 4.5.1. Under the assumptions of Proposition 4.4.1, given a multivariate simulator g and the overall interaction indices in (5.16), we have:

$$\sum_{i=1}^{d} \Upsilon_i^2 = \sum_u |u|\sigma_u^2 - \sum_{i=1}^{d} \underline{\tau}_i^2, \tag{4.22}$$

and

$$\sum_{i=1}^{d} \Upsilon_i^2 = \sum_{|u|>1} |u|\sigma_u^2 \tag{4.23}$$

Proposition (4.17) suggests that the sum of the overall interaction indices Υ_i^2 exceeds the sum of all the first order indices by the the numerator of the mean dimension in the superimposition sense. In particular, we have:

$$\sum_u |u|\sigma_u^2 = \sum_{i=1}^d \Upsilon_i^2 + \sum_{i=1}^d \underline{\tau}_i^2. \tag{4.24}$$

If we divide equations (4.22) and (4.23) by σ_Y^2, we obtain

$$D_g = D_g^{Add} + D_g^{Int} \quad \text{where} \quad D_g^{Add} = \frac{\sum_{i=1}^d \underline{\tau}_i^2}{\sigma_Y^2}, \quad D_g^{Int} = \frac{\sum_{i=1}^d \Upsilon_i^2}{\sigma_Y^2}, \tag{4.25}$$

so that the mean dimension is decomposed into an additive and an interaction component. We recall that Wang [2006] regards D_g^{Add} as a measure of the degree of additivity of g: if $D_g^{Add} = 1$, (close to 1), then g is additive (nearly additive). In fact, if and only if g is additive, $D_g^{Int} = 0$, and $D_g = D_g^{Add}$, and $\sum_{i=1}^d \underline{\tau}_i^2 = \sigma_Y^2$, so that $D_g = 1$. Conversely, if $D_g^{Int} \gg 1$, then g is dominated by interaction effects.

The following result connects main effects and Υ_i^2.

Proposition 4.5.2. The overall interaction index Υ_i^2 can be written as

$$\Upsilon_i^2 = \frac{1}{2}\mathbb{E}\left[\left(\Phi_i^{\mathbf{X}\to\mathbf{X}^1}\right)^2 - 2\cdot g(\mathbf{X})\cdot\Phi_i^{\mathbf{X}^1\to\mathbf{X}}\right]. \tag{4.26}$$

The sum $\sum_{i=1}^d \Upsilon_i^2$ is given by

$$\sum_{i=1}^d \Upsilon_i^2 = \sum_{i=1}^d \mathbb{E}\left[-g(\mathbf{X})\cdot\left(\Phi_i^{\mathbf{X}\to\mathbf{X}^1} + \Phi_i^{\mathbf{X}^1\to\mathbf{X}}\right)\right]. \tag{4.27}$$

Moreover, the numerator of the mean dimension in the superimposition sense can be written as

$$\sum_u |u|\sigma_u^2 = \mathbb{E}\left[-g(\mathbf{X})\cdot\sum_{i=1}^d \Phi_i^{\mathbf{X}\to\mathbf{X}^1}\right]. \tag{4.28}$$

Equations (4.26), (4.27) and (4.28) relate several measures of interaction on a global scale to main effects computed using OAT replicates.

Corollary 4.5.1. *Monte Carlo estimators for sensitivity indices* (4.26)*,* (4.27) *and* (4.28) *are respectively given by*

$$\hat{\Upsilon}_i^2 = \frac{\sum_{k=0}^{N-1}\left[\left(\phi_i^{k\to k+1}\right)^2 - 2g(\mathbf{x}_{(k)}^0)\cdot\phi_i^{k+1\to k}\right]}{2N}, \tag{4.29}$$

$$\sum_{i=1}^{d}\hat{\Upsilon}_i^2 = \frac{2\sum_{i=1}^{d}\sum_{k=0}^{N-1}\left[-g(\mathbf{x}_{(k)}^0)\cdot\left(\phi_i^{k\to k+1}+\phi_i^{k+1\to k}\right)\right]}{N} \tag{4.30}$$

and

$$\sum_{u}|u|\hat{\sigma}_u^2 = -\sum_{i=1}^{d}\sum_{k=0}^{N-1}\frac{g(\mathbf{x}_{(k)}^0)\cdot\left(\phi_i^{k\to k+1}\right)}{N}. \tag{4.31}$$

Corollary 4.5.1 shows that information on the structure of the simulator and on the relevance of interactions can be obtained by computing replicates of main effects. Specifically, one can estimate the overall global interaction index associated with each simulator input (4.29), and the numerator of the mean effective dimension (4.31). Equations (4.27) and (4.30) show that the corresponding estimator of $\sum_{i=1}^{d}\hat{\Upsilon}_i^2$ is based on a fold-over design, as they register the presence of $\Phi_i^{\mathbf{X}\to\mathbf{X}^1}$ and $\Phi_i^{\mathbf{X}^1\to\mathbf{X}}$. It is also possible to rewrite the estimator of $\sum_{i=1}^{d}\Upsilon_i^2$ as follows.

Corollary 4.5.2. *The Monte Carlo estimator of* Υ_i^2 *can be rewritten as*

$$\begin{aligned}\sum_{i=1}^{d}\hat{\Upsilon}_i^2 &= \frac{\sum_{i=1}^{d}\sum_{k=1}^{N}g(\mathbf{x}_{(k)}^0)\cdot\left[g(\mathbf{x}^0)-g(\mathbf{x}_{-i}^0:x_i^1)+g(\mathbf{x}^1)-g(x_i^0:\mathbf{x}_{-i}^1)\right]_{(k)}}{N}\\ &= \frac{2\sum_{i=1}^{d}\sum_{k=1}^{N}g(\mathbf{x}_{(k)}^0)\cdot A_{(k)}^{i,-i}}{N},\end{aligned} \tag{4.32}$$

where $A_{(k)}^{i,-i} = g(\mathbf{x}^0) - g(\mathbf{x}_{-i}^0 : x_i^1) + g(\mathbf{x}^1) - g(x_i^0 : \mathbf{x}_{-i}^1)$ *is the two-factor interaction effect introduced in Design of Experiments Wu [2015], and also used for the sensitivity analysis of computer experiments Fruth et al. [2014].*

By $A_{(k)}^{i,-i}$ one can obtain the interaction measure $EI_{(k)}^{i,-i} = \left|A_{(k)}^{i,-i}\right|$ used in Campolongo et al. [2011]. Equation (4.32) shows that $\sum_{i=1}^{d}\hat{\Upsilon}_i^2$ is the

mean of replicated two-level full factorial designs weighted by $g(\mathbf{x}^0_{(k)})$, the output value at the initial point.

Note that all these estimators are based on main effects, that are also an ingredient of Morris μ_i and σ_i. Thus, designs based on the replicates of OAT allow the analyst to compute both the statistics of elementary effects and Sobol' sensitivity indices using the same batch of simulations. This can be of great interest to modelers Campolongo et al. [2011]. The total cost of these estimators is $N(2d+2)$. However, the cost can be reduced to the same cost as the Morris method, as the next section discusses.

4.6 A Symmetry Effect: Two Possible Choices

In the above discussion, we have considered a design based on a forward and a backward pass of a trajectory. However, the order with which the simulator is evaluated on the two trajectories is immaterial.

Theorem 4.6.1. *The following equalities hold for* $\underline{\tau}_i^2$ *and* Υ_i^2*:*

$$\underline{\tau}_i^2 = \mathbb{E}\left[g(\mathbf{X}^1)\cdot\Phi_i^{\mathbf{X}\to\mathbf{X}^1}\right], \tag{4.33}$$

and

$$\Upsilon_i^2 = \frac{1}{2}\mathbb{E}\left[\Phi_i^{\mathbf{X}\to\mathbf{X}^1}\cdot\left(\Phi_i^{\mathbf{X}\to\mathbf{X}^1} - 2g(\mathbf{X}^1)\right)\right]. \tag{4.34}$$

Theorem 4.6.1 establishes a symmetry property. In particular, (4.33) is the mirror of (4.18) in which $\mathbf{X}^1$ and $\Phi_i^{\mathbf{X}\to\mathbf{X}^1}$ replace $\mathbf{X}$ and $\Phi_i^{\mathbf{X}^1\to\mathbf{X}}$. Similarly, (4.34) suggests that the total interaction index Υ_i^2 can be computed from the main effects $\Phi_i^{\mathbf{X}\to\mathbf{X}^1}$ evaluations plus the evaluation of g at $\mathbf{X}^1$.

Corollary 4.6.1. *We have the following Monte Carlo estimates for* $\underline{\tau}_i^2$ *and* Υ_i^2*:*

$$\hat{\underline{\tau}}_i^2 = \frac{\sum_{k=0}^{N-1}\left[g(\mathbf{x}^{k+1})\cdot\phi_i^{k\to k+1}\right]}{N}, \tag{4.35}$$

and

$$\hat{\Upsilon}_i^2 = \frac{\sum_{k=0}^{N-1}\left[\phi_i^{k\to k+1}\cdot\left(\phi_i^{k\to k+1} - 2g(\mathbf{x}^{k+1})\right)\right]}{2N}. \tag{4.36}$$

Corollaries 4.4.1 and 4.6.1 show that, by knowledge of the main effects $\phi_i^{k\to k+1}$, we obtain estimates of the first order variance-based indices, $\hat{\underline{\tau}}_i^2$ [equation (4.35)], of the interaction indices, $\hat{\Upsilon}_i^2$ [equation (4.36)], of the total order indices [equation (4.19)], of the mean effective dimension D_g, via equation (4.36). The cost of the analysis is $N(d+1)+1$ model runs (the cost of the Morris method plus one model run) and all estimates are obtained from a forward trajectory.

Now, at the same cost C, we can perform a fold-over design, with $N/2(d+1)$ runs for the forward pass and corresponding $N/2(d+1)$ runs in the opposite direction. If we do so, we have available $N(d+1)/2$ forward main effects $\phi_i^{k\to k+1}$ but also $N(d+1)/2$ main effects from the backward pass. As we discussed, these two sets of indices are, in fact, N main effects to be used in the relevant equations (e.g. (4.19) and (4.35)). This number is the same as in the full forward design. However, we gain the estimates of $N/2$ interaction indices ϕ_i^I at randomized locations.

In summary, given a budget C equal to the cost of the Morris method plus one, an analyst has two possible choices: a) run N replicates in a fully forward design; b) split the replicates into forward and backward trajectories. This second choice allows the computation of the interaction indices ϕ_i^I, which also provide information on the sign of interactions. In addition, all other sensitivity measures we have discussed can be estimated as well, with the same precision as in the fully forward design.

4.7 Asymptotic Analysis

In estimation, it is of practical interest to obtain measures of the error at finite sample sizes. We consider the population $\mathbf{\Phi}$ and denote by $\mathbb{V}[\cdot]$ and $\mathbb{V}[\dot{\cdot}]$ the population and the sample variance, respectively.

Lemma 4.7.1. *Under the assumptions of Proposition 4.4.1 we register*

$$\mathbb{E}[\Phi_i] = 0, \tag{4.37}$$

$$\mathbb{V}[\Phi_i] = 2\overline{\tau}_i^2 \tag{4.38}$$

$$\mathbb{C}\left[\Phi_i, \Phi_j\right] = \Upsilon_{ij}^2 \tag{4.39}$$

and

$$\mathbb{V}\left[\bar{\Phi}_i\right] = \frac{2}{N}\bar{\tau}_i^2, \tag{4.40}$$

where $\bar{\Phi}_i = \frac{1}{N}\sum_{k=0}^{N-1} g\left(x_i^{k+1} : \mathbf{x}_{-i}^k\right) - g\left(\mathbf{x}^k\right)$ *is the empirical mean of* Φ_i.

Lemma 4.7.1 implies that the mean of a main effect is zero, and its variance equals twice the total-order variance-based indices. Equations (4.37)-(4.40) suggest that main effects of important inputs will show greater variability than main effects of unimportant inputs. Thus, given a sample of main effects, the analysts can have a rough idea of the relative inputs importance plotting their distribution and considering their dispersion around zero. In a screening exercise this result is reassuring. If we register small variability of the empirical mean of Φ_i, then it is safe to expect that X_i is an input of low relevance. Also, the covariance of two main effects is equal to the so-called superset importance of X_i, X_j, $\Upsilon_{i,j}^2$ [Liu and Owen, 2006]. This term is null when g is additive and a value different from zero indicates the presence of interactions in g that involve X_i and X_j. This information can be visualized in a variance-covariance matrix of the main effects

$$\Upsilon = \begin{pmatrix} 2\bar{\tau}_1^2 & \Upsilon_{12}^2 & \cdots & \Upsilon_{1d}^2 \\ \Upsilon_{12}^2 & 2\bar{\tau}_2^2 & \cdots & \Upsilon_{2d}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \Upsilon_{1d}^2 & \Upsilon_{2d}^2 & \cdots & 2\bar{\tau}_d^2 \end{pmatrix}. \tag{4.41}$$

Moreover, because $\sum_u |u|\sigma_u^2 = \sum_{i=1}^d \bar{\tau}_i^2$ by Owen [2013], one finds that

$$D_g = \frac{trace(\Upsilon)}{2\sigma_Y^2},$$

so that the mean dimension is the normalized sum of the diagonal entries of Υ.

Example 4.7.1. *Consider* $g\left(\mathbf{X}\right) = X_1 \cdot X_2 \cdot X_3$ *with the three inputs iid* $\mathcal{N}(1,1)$. *Analytically, we find* $\bar{\tau}_i^2 = 4$ *and* $\Upsilon_{ij}^2 = 2$ *for* $i = 1, 2, 3$, $i \neq j$.

Correspondingly, the variance-covariance matrix is $\Upsilon = \begin{pmatrix} 8 & 2 & 2 \\ 2 & 8 & 2 \\ 2 & 2 & 8 \end{pmatrix}$. *Consider now a numerical estimation at large sample size. At* $N = 50,000$ *(the sample is generated through crude Monte Carlo) applying the empirical estimator of the covariance between the main effects, we find*

$$\hat{\Upsilon} = \begin{pmatrix} 8.0717 & 2.0419 & 2.0122 \\ 2.0419 & 7.9686 & 2.0022 \\ 2.0122 & 2.0022 & 7.8928 \end{pmatrix}.$$

Theorem 4.7.1. *Under the assumption of Proposition 4.4.1, we have*

$$\sqrt{N}\bar{\Phi}_i \longrightarrow \mathcal{N}\left(0, 2\overline{\tau}_i^2\right) \tag{4.42}$$

for $N \to \infty$, $i = 1, 2, ..., d$.

This result suggests 4.7.1 that the means of the Φ_i are asymptotically normal with null mean and variance $2\overline{\tau}_i^2/N$.

Theorem 4.7.2. *Assume that* $\mathbb{E}\left[g(\mathbf{X})^4\right] = \mu_4 < \infty$ *and all inputs are independent. Then, for* $N \to \infty$ *we have*

$$\sqrt{N}\left(\hat{\underline{\tau}}_{i,N}^2 - \underline{\tau}_i^2\right) \longrightarrow \mathcal{N}\left(0, \mathbb{V}\left[g\left(\mathbf{X}\right) \cdot \Phi_i^{\mathbf{X}^1 \to \mathbf{X}}\right]\right), \tag{4.43}$$

$$\sqrt{N}\left(\hat{\overline{\tau}}_{i,N}^2 - \overline{\tau}_i^2\right) \longrightarrow \mathcal{N}\left(0, \frac{\mu_4 - 4\overline{\tau}_i^4}{4}\right) \tag{4.44}$$

and

$$\sqrt{N}\left(\sum_{i=1}^{d} \hat{\overline{\tau}}_{i,N}^2 - \sum_{u} |u|\sigma_u^2\right) \longrightarrow \mathcal{N}\left(0, \frac{d\mu_4 - 4\sum_{i=1}^{d} \overline{\tau}_i^4}{4}\right). \tag{4.45}$$

Theorem 4.7.2 suggests that main effect-based estimators of first and total order indices are asymptotically normal, unbiased and their variance tends to zero as the sample size increases. Equations (4.43) and (4.44) yield confidence intervals for variance-based sensitivity indices. Equation (4.45) yields confidence intervals for the numerator of the mean effective dimension of g.

Of interest is also the asymptotic behaviour of the estimators of the normalized sensitivity indices and of the effective dimension. We let

$$S_i = \frac{\tau_i^2}{\sigma_Y^2}, \quad T_i = \frac{\overline{\tau}_i^2}{\sigma_Y^2} \quad \text{and} \quad I_i = \frac{\Upsilon_i^2}{\sigma_Y^2}, \tag{4.46}$$

denote the individual, total and interaction Sobol' indices respectively. Note that, by (4.25), $D_g = \sum_{i=1}^{d} (S_i + I_i)$.

Theorem 4.7.3. *Under the assumptions of Theorem 4.7.2, we have:*

$$\sqrt{N}\left(\hat{S}_{i,N} - S_i\right) \longrightarrow \mathcal{N}\left(0, \Gamma_{S_i}\right) \tag{4.47}$$

$$\sqrt{N}\left(\hat{T}_{i,N} - T_i\right) \longrightarrow \mathcal{N}\left(0, \Gamma_{T_i}\right) \tag{4.48}$$

and

$$\sqrt{N}\left(\hat{D}_N - D_g\right) \longrightarrow \mathcal{N}\left(0, \Gamma_{D_g}\right) \tag{4.49}$$

where μ_4 the fourth moment of g, the estimators are given by

$$\hat{S}_{i,N} = \frac{\hat{\tau}_i^2}{\hat{\sigma}_Y^2}, \quad \hat{T}_{i,N} = \frac{\hat{\overline{\tau}}_i^2}{\hat{\sigma}_Y^2} \quad \textit{and} \quad \hat{D}_N = \frac{\sum_{i=1}^{d} \hat{\overline{\tau}}_{i,N}^2}{\hat{\sigma}_Y^2} \tag{4.50}$$

and the asymptotic variances are

$$\Gamma_{S_i} = \frac{\mathbb{V}\left[g(\mathbf{X}^1) \cdot \Phi_i^{\mathbf{X} \to \mathbf{X}^1}\right] + S_i^2\left(\mu_4 - \sigma_Y^4\right) - 2S_i \mathbb{C}\left[g(\mathbf{X}^1) \cdot \Phi_i^{\mathbf{X} \to \mathbf{X}^1}, (g(\mathbf{X}) - \mu)^2\right]}{\sigma_Y^4}, \tag{4.51}$$

$$\Gamma_{T_i} = \frac{\mu_4 - 4\overline{\tau}_i^4 + 4T_i^2\left(\mu_4 - \sigma_Y^4\right) - 4T_i \mathbb{C}\left[(\Phi_i^{\mathbf{X} \to \mathbf{X}^1})^2, (g(\mathbf{X}) - \mu)^2\right]}{4\sigma_Y^4} \tag{4.52}$$

and

$$\Gamma_{D_g} = \frac{d\mu_4 - 4\sum_{i=1}^{d} \overline{\tau}_i^4 + 4D_g^2\left(\mu_4 - \sigma_Y^4\right) - 4D_g \mathbb{C}\left[\sum_{i=1}^{d} (\Phi_i^{\mathbf{X} \to \mathbf{X}^1})^2, (g(\mathbf{X}) - \mu)^2\right]}{4\sigma_Y^4}. \tag{4.53}$$

For first order indices, Janon et al. [2014] obtained confidence intervals for estimators based on the pick-and-freeze design and regard S_i as the

result of a covariance calculation. Our approach, instead, allows one to consider the total indices T_i as variance of main effects. Indeed, because the main effects have zero mean (Lemma 4.7.1), the estimator of the total Sobol' indices is expressed as a variance of main effects, by (4.38) we have

$$T_i^2 = \frac{1}{2}\frac{\mathbb{V}\left[\Phi_i\right]}{\sigma_Y^2}. \tag{4.54}$$

From (4.54), T_i can now be interpreted as the (halved) proportion of the simulator variance due to the one-at-a-time variations of the model input of interest. Theorem 4.7.3 then characterizes the asymptotic behavior of T_i and D_g.

4.8 Finite changes with constant step size Δ

We can sample in two ways: random sampling or Morris' sampling with fixed Δ. Lemma 4.7.1 and Theorem 4.7.1 lead to the asymptotic properties of Morris μ_i and σ_i, with a caveat. Morris definition of trajectories is linked to a predefined grid in the model input space. However, consider still that Δ is fixed, but one abandons the grid and computes the elementary effects at locations in $\mathcal{X}$ independently sampled (through a Monte Carlo or another space filling design). Because EE_i are Newton ratios, one is sampling approximations of partial derivatives at randomized locations in $\mathcal{X}$. This way of proceeding resembles the estimation of sensitivity measures in the Distributed Evaluation of Local Sensitivity Analysis (DELSA), in derivative-based methods and in radial design sampling [Rakovec et al., 2014, Campolongo et al., 2011].

In the case we use Morris sampling with constant Δ, we obtain notable differences in the properties of the estimators. First, let us define the finite change effect obtained with the Morris method as

$${}_{\Delta}\Phi_i = g(X_i^0 + \Delta : \mathbf{X}_{-i}^0) - g(\mathbf{X}^0).$$

We observe that this finite change is the numerator of EE_i. Now, as opposed to Φ_i , ${}_{\Delta}\Phi_i$ does not have null expectation. In general we obtain

$$EE_i \cdot \Delta = {}_{\Delta}\Phi_i$$

therefore passing to the mean value

$$E[{}_{\Delta}\Phi_i] = \mu_i \Delta. \tag{4.55}$$

Example 4.8.1. *Consider a linear model:*

$$Y = \sum_{i=1}^{d} \beta_i X_i + a$$

We have that

$$E[{}_{\Delta}\Phi_i] = \beta_i \Delta$$

Instead the variance of elementary effects can be computed as follows.

Corollary 4.8.1. *For any finite N and Δ in a Morris design, we have:*

$$\hat{\sigma}^2_{EE_i} = \frac{N}{N-1} \frac{2_{\Delta}\hat{\overline{\tau}}^2_i}{\Delta^2} - \frac{N}{N-1} \hat{\mu}_i^2 \tag{4.56}$$

where

$${}_{\Delta}\hat{\overline{\tau}}^2_i = \frac{\sum_{k=0}^{N-1} \left({}_{\Delta}\phi_i^{k \to k+1}\right)^2}{\Delta^2}.$$

Several works in the literature (see Campolongo et al. [2011] among others) have hinted at the fact that the variance of elementary effects is a measure of non linearity and interactions. Note that (4.56) is used to construct test statistics for hypothesis testing in the controlled Morris method in Shi et al. [2016].

However, note that ${}_{\Delta}\overline{\tau}^2_i$ is a biased estimator of Saltelli-Sobol' total index t of the i-th input since every couple of points is forced to have the i-th component at distance Δ. We show how we can eliminate this bias arisen from the constrained sampling strategy in next section.

4.8.1 Debiasing ${}_{\Delta}\overline{\tau}^2_i$ as an estimator of the total effect

It is possible to debias the Sobol' index ${}_{\Delta}\overline{\tau}^2_i$ evuated using the finite changes ϕ_i^{Δ} with fixed step Δ. Ideed, one finds

$$\Phi_i^2 = \left(g(X_i^1 : \mathbf{X}_{-i}^0) - g(\mathbf{X}_0)\right)^2 =$$

$$= \left(g(X_i^1 : \mathbf{X}_{-i}^0) - g(X_i^0 + \Delta : \mathbf{X}_{-i}^0) + g(X_i^0 + \Delta : \mathbf{X}_{-i}^0) - g(\mathbf{X}^0)\right)^2 =$$
$$= \left(g(X_i^1 : \mathbf{X}_{-i}^0) - g(X_i^0 + \Delta : \mathbf{X}_{-i}^0)\right)^2 + \left(g(X_i^0 + \Delta : \mathbf{X}_{-i}^0) - g(\mathbf{X}^0)\right)^2 +$$
$$+2\left(g(X_i^1 : \mathbf{X}_{-i}^0) - g(X_i^0 + \Delta : \mathbf{X}_{-i}^0)\right)\left(g(X_i^0 + \Delta : \mathbf{X}_{-i}^0) - g(\mathbf{X}^0)\right) =$$
$$= \left(g(X_i^1 : \mathbf{X}_{-i}^0) - g(X_i^0 + \Delta : \mathbf{X}_{-i}^0)\right)^2 + {}_\Delta\Phi_i^2 + 2\left(g(X_i^1 : \mathbf{X}_{-i}^0) - g(X_i^0 + \Delta : \mathbf{X}_{-i}^0)\right) \cdot {}_\Delta\Phi_i$$

Hence, the bias arising from using the fixed-step sampling can be written as

$$Bias\left(\Phi_i^2 - {}_\Delta\Phi_i^2\right) = \left(g(X_i^1 : \mathbf{X}_{-i}^0) - g(X_i^0 + \Delta : \mathbf{X}_{-i}^0)\right)^2 \tag{4.57}$$
$$+2\left(g(X_i^1 : \mathbf{X}_{-i}^0) - g(X_i^0 + \Delta : \mathbf{X}_{-i}^0)\right) \cdot {}_\Delta\Phi_i$$

Taking expected values and dividing by 2, we get

$$\overline{\tau}_i^2 = \frac{1}{2}\mathbb{E}\left[\left(g(X_i^1 : \mathbf{X}_{-i}^0) - g(X_i^0 + \Delta : \mathbf{X}_{-i}^0)\right)^2\right] + {}_\Delta\overline{\tau}_i^2 + \tag{4.58}$$
$$\mathbb{E}\left[\left(g(X_i^1 : \mathbf{X}_{-i}^0) - g(X_i^0 + \Delta : \mathbf{X}_{-i}^0)\right) \cdot {}_\Delta\Phi_i\right]$$

and finally

$$Bias\left(\overline{\tau}_i^2 - {}_\Delta\overline{\tau}_i^2\right) = \frac{1}{2}\mathbb{E}\left[\left(g(X_i^1 : \mathbf{X}_{-i}^0) - g(X_i^0 + \Delta : \mathbf{X}_{-i}^0)\right)^2\right] + \tag{4.59}$$
$$+\mathbb{E}\left[\left(g(X_i^1 : \mathbf{X}_{-i}^0) - g(X_i^0 + \Delta : \mathbf{X}_{-i}^0)\right) \cdot {}_\Delta\Phi_i\right].$$

Example 4.8.2. *Consider the product of three normal variables in Example 4.7.1. We consider the debiased finite-change estimators of* $\overline{\tau}_i^2 = 4$ *for* $\Delta = 0.1$, $\Delta = 0.5$, $\Delta = 1$ *and* $\Delta = 2$ *and we compare them with the finite-change estimator based on two independent realizations under random sampling (RS). Results are displayed in next Figure 4.2. We can see that in both panels the Sobol' total indices converge. We check numerically equation* (4.56). *For instance, in the case* $\Delta = 0.1$, *we find after* $N = 25{,}000$ *that* $\hat{\sigma}_{EE_i}^2 = 2.9358$ *and that* $\frac{N}{N-1}\frac{2_\Delta\hat{\overline{\tau}}_i^2}{\Delta^2} - \frac{N}{N-1}\hat{\mu}_i^2 = 2.9357$. *This case is interesting since we can see that the blue line corresponds to the worst estimators. However, this is natural since it is the smallest step size considered and it would require more runs to explore the input space.*

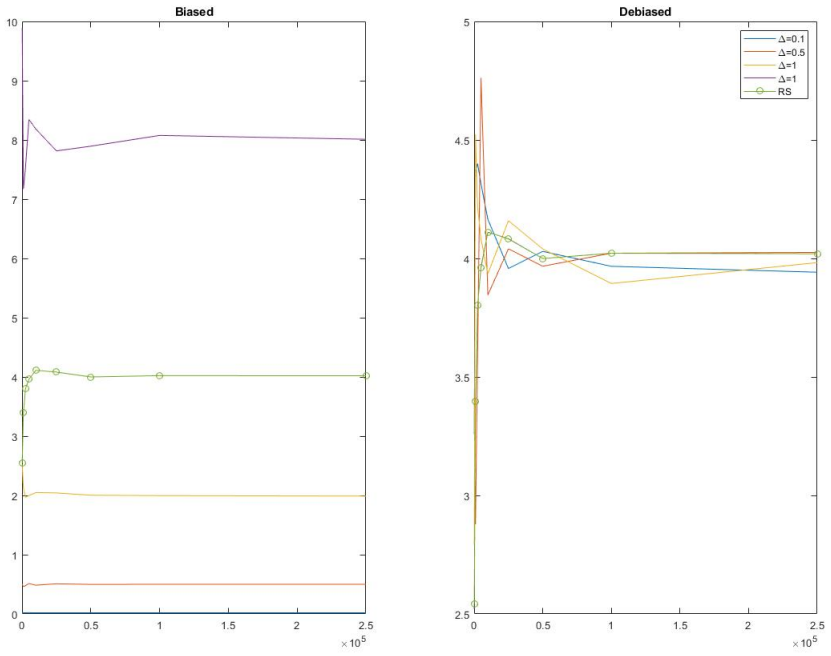


Figure 4.2: Comparison betweem biased (left) and debiased (rigth) ${}_{\Delta}\overline{\tau}_i^2$ as estimator of $\overline{\tau}_i^2$.

Asymptotically, one finds the following behaviour.

Proposition 4.8.1. Under assumptions of Proposition 4.4.1 with constant Δ, one has

$$\sigma^2_{EE_i} = 2\frac{{}_{\Delta}\overline{\tau}^2_i}{\Delta^2} - \mu_i^2 \tag{4.60}$$

and

$$\mathbb{V}\left[\hat{\mu}_i\right] = \frac{\hat{\sigma}^2_{EE_i}}{N} \quad i = 1, ..., d, . \tag{4.61}$$

where

$${}_{\Delta}\overline{\tau}^2_i = \frac{1}{2}\mathbb{E}\left[\left({}_{\Delta}\Phi_i\right)^2\right]. \tag{4.62}$$

and μ_i is defined in equation (4.55).

Equation (4.60) is, theoretically, the limit for $N \to \infty$ of equation (4.56) and shows that Morris' $\sigma^2_{EE_i}$ becomes proportional to ${}_{\Delta}\overline{\tau}^2_i$. In particular, correcting the bias in equation (4.60) it is possible to formally connect the variance of Morris' elementary effects to Sobol' indices as

$$\sigma^2_{EE_i} = \frac{2}{\Delta^2}\left(\overline{\tau}^2_i - Bias\left(\overline{\tau}^2_i - {}_{\Delta}\overline{\tau}^2_i\right)\right) - \mu_i^2 = \frac{2}{\Delta^2}\overline{\tau}^2_i - b_g(\Delta, i) \tag{4.63}$$

where $b_g(\Delta, i) = \frac{2}{\Delta^2} Bias\left(\overline{\tau}^2_i - {}_{\Delta}\overline{\tau}^2_i\right) + \mu_i^2$.

Proposition 4.8.2. For the population of EE_i's obtained through random sampling, we have:

$$\sqrt{N}\left(\hat{\mu}_i - \mu_i\right) \longrightarrow \mathcal{N}\left(0, \frac{2}{\Delta^2}{}_{\Delta}\overline{\tau}^2_i - \mu_i^2\right) \tag{4.64}$$

for $N \to \infty$, where μ_i is the mean value of the elementary effect EE_i.

More in general, these results suggest that, fixed Δ, Newton ratios are asymptotically distributed as a normal random variable with mean μ_i and variance $\frac{2}{\Delta^2}{}_{\Delta}\overline{\tau}^2_i - \mu_i^2$. This result connects this work with earlier findings of Zazanis and Suri [1993], who investigate the asymptotic convergence rate of Newton ratios. Zazanis and Suri [1993] show that, if Δ^* is chosen to be the optimal increment (i.e. the increment that minimizes the mean square error of the finite-difference estimator), then the corresponding estimator

$\hat{\mu}_i^*$ converges at a rate of order $N^{-1/4}$, which is smaller than the Monte Carlo rate $O(N^{-1/2})$.

Thus far, we have not made any assumption on the functional form of g. This results in a connection of the findings in previous works, such as Cheng [1997], Boukouvalas et al. [2014].

Example 4.8.3. *Consider the linear model in Example 4.8.1. It follows that* $EE_i = \beta_i$ *and* $\mu_i = \beta_i$. *Thus,* $\sigma^2_{EE_i} = 0$. *On the other hand, it holds that* ${}_\Delta\overline{\tau}_i^2 = \frac{1}{2}\beta_i^2\Delta^2$. *Consequently, we obtain that the variance in* (4.64) *becomes* $\frac{2}{\Delta^2}{}_\Delta\overline{\tau}_i^2 - \mu_i^2 = \beta_i^2 - \beta_i^2 = 0$.

Boukouvalas et al. [2014] propose a screening method for simulation experiments based on the elementary effects method. Their algorithm discriminates the inputs with linear and nonlinear effects using a threshold value for the variance of the elementary effects. It is assumed that the data follow a linear noise model for the $i-$th input of the type

$$Y(x_i) = ax_i + b + \epsilon_i, \tag{4.65}$$

where $\epsilon_i \sim \mathcal{N}\left(0, \gamma^2\right)$. For this linear model, they prove that $\hat{\mu}_i \sim \mathcal{N}\left(a, \frac{2}{N\Delta^2}\gamma^2\right)$. Note that this result holds because the quantity of interest is Y. In the quantity of interest were $\mathbb{E}[Y]$, we would find $\hat{\mu}_i = a$ and $\mathbb{V}[\hat{\mu}_i] = 0$ (see Example 4.8.1). This is the same result we would obtain by setting $\gamma = 0$ in eq. (4.65), that is, for a deterministic code.

We can generalize these findings. Cheng [1997] considers the sequential bifurcation method [Bettonvil and Kleijnen, 1997] for the stochastic model

$$Y(\mathbf{x}) = \beta_0 + \beta_1(x_1 - x_1^0) + ... + \beta_d(x_d - x_d^0) + \epsilon, \tag{4.66}$$

where the error ϵ is independent of $\mathbf{X}$. We assume that it is normally distributed with zero mean and variance γ^2. Following the sequential bifurcation method, Cheng [1997] evaluates the model at the points $\mathbf{x}^{(j)}$and $\mathbf{x}$ where

$$\begin{aligned} x_i^{(j)} &= x_i^0 + \Delta \quad i = 1, 2, ..., j \\ x_i^{(j)} &= x_i^0 \quad i = j+1, ..., d. \end{aligned} \tag{4.67}$$

Then, for $j < k$, Cheng [1997] defines the scaled finite difference $D(j,k) = \left[g(\mathbf{x}^{(k)}) - g(\mathbf{x}^{(j)})\right]/\Delta$. This can be rewritten as $D(j,k) = \sum_{i=j+1}^{k} {}_{\Delta}\Phi_i/\Delta$, which has expectation $\mathbb{E}\left[D(j,k)\right] = \sum_{i=j+1}^{k} \beta_i$. Note that the mean of the elementary effects is equal to $\mu_i = \beta_i$. Hence, the variance of $D(j,k)$ can be expressed as $\mathbb{V}\left[D(j,k)\right] = \frac{2}{\Delta^2}\gamma^2$. Thus, we find that

$$\hat{D}(j,k) \sim \mathcal{N}\left(\sum_{i=j+1}^{k} \beta_i, \frac{2}{N\Delta^2}\gamma^2\right). \tag{4.68}$$

Thus, we have find the same result in Cheng [1997] considering Morris' elementary effects in the special case of a linear function.

We finally note that as in the above-mentioned works, we make the assumption that model inputs are independent. In Section 4.10.1 in the Appendix we briefly describe the case of dependent inputs.

4.9 Case Studies

The purpose of this section is to examine the numerical behavior in the case of well known and realistic simulators previously used in the simulation literature.

Regarding computational cost, proposition 4.5.2 and Theorem 4.6.1 suggest that the estimators of $\underline{\tau}_i^2$, $\overline{\tau}_i^2$, and Υ_i^2 and D require $N(d+1)+1$ simulator evaluations. Nominally, this is the cost of the Morris method plus one model evaluation.

Regarding insights, the literature has made inference from sensitivity measures systematic through the notion of sensitivity analysis setting [Saltelli and Tarantola, 2002]. Among sensitivity analysis settings we recall factor prioritization (or fixing), interaction identification and trend determination. The estimates of total order indices provide indication about factor prioritization. The estimates of the mean dimension, the Υ matrix and the overall interaction indices Υ_i inform us about the magnitude of interactions. Moreover, if the analyst uses a folded-over design, she obtains information about the sign of interactions, with computation

of the $\Upsilon_i^{0\to 1}$ indices. Finally, as well known, the sign of elementary effects suggests whether g is increasing or decreasing in X_i. For instance, if g is monotonic (increasing or decreasing), then all Newton ratios are either positive or negative. Therefore, a change in sign of Newton ratios across some locations in $\mathcal{X}$ communicates to the analyst that the simulator response is not monotonic in X_i.

We discuss these insights for the Asian Option pricing simulator in Nelson [2013], the ATO Simulator of Hong and Nelson [2006] and the NASA space mission code of Borgonovo and Smith [2011].

4.9.1 Asian Option value

Asian options are among the most popular and studied path-dependent financial options [Wang and Sloan, 2011, Cai et al., 2015]. The payoff of an Asian option depends on the average price of the underlying asset over a time period. Nelson [2013] considers the simulation of the value of an Asian option

$$v = \mathbb{E}\left[\mathrm{e}^{-rT}\left(\bar{X}(T) - K\right)^{+}\right], \tag{4.69}$$

where $T = 1$ is the maturity, r is the risk-free interest rate and K is the strike price. The underlying asset has initial value $X(0)$ and volatility σ^2. The quantity $\bar{X}(T) = \int_0^T X(t)dt$ is the time-average of the Brownian motion. The base case scenario of the simulation is $(\sigma, r, X(0), K) = (0.3, 0.05, 50, 55)$. In order to demonstrate the design, we consider a variation range for each input with low and high extremes equal, respectively, to 5% and 195% of the base-case value. For every input combination, we used 10000 inner stochastic replicates of the option pricing model and averaged to find the estimated expected value $\widehat{v}$.

As a screening design, we opt for the mirror scheme, with 10000 forward and 10000 backward replications, for a total of $80,000$ evaluations of $\widehat{v}$. The analysis requires 2 hours and 45 minutes on a personal computer with 64GB RAM, and processor Intel(R) Core(TM) i7-7700HQ, 2.80GHZ. The subroutines are developed in MATLAB.

Figure 4.3a reports the empirical distributions of the main effects of

the four simulator inputs. The densities in Figure 4.3a are obtained by fitting the 20,000 realizations of Φ_σ, Φ_r, $\Phi_{X(0)}$, and Φ_K with the Matlab ksdensity.m subroutine. The realizations of $\Phi_{X(0)}$ and Φ_K display greater variability around the mean value (zero) than the realizations of Φ_σ and Φ_r. This is confirmed by the barplots in Figure 4.3b, that report the corresponding main effects mean values and variances. The variance of $\Phi_{X(0)}$ and Φ_K are much larger than the corresponding variances of Φ_σ and Φ_r, that are close to zero.
The replicates allow one to move some steps also towards global sensitivity measures. For the totals, T_i, Figure 4.4a reports the point estimates and corresponding confidence intervals as N increases, as per equations (4.48)-(4.50).

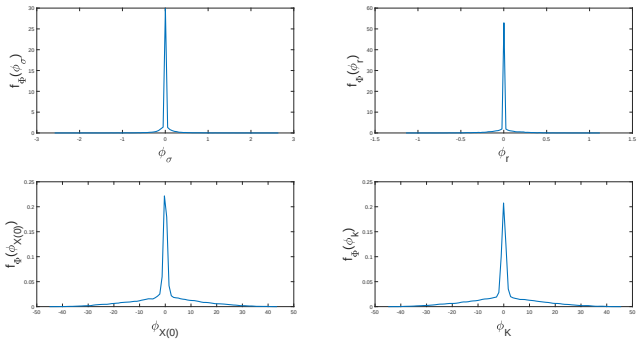
Figure 4.4a suggests the following. Across all values of N, the point estimates $\widehat{T}_K$ and $\widehat{T}_{X(0)}$ are greater than the point estimates of $\widehat{T}_K$ and $\widehat{T}_{X(0)}$. However, the confidence intervals allow one to clearly separate them after $N > 400$ replicates. As N increases, the confidence intervals shrink towards the point estimates. At $N = 10000$, we find $\widehat{T}_{X(0)} = 0.6025$ with 95% confidence interval given by $[0.5646, 0.6404]$.

We can also determine the empirical normalized variance-covariance matrix of the main effects (4.41):

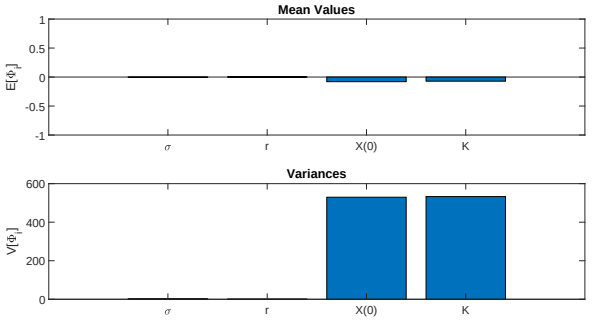
$$\frac{\hat{\Upsilon}}{\hat{\sigma}^2} = \begin{pmatrix} 0.0043 & 0.0002 & 0.0014 & 0.0020 \\ 0.0002 & 0.0008 & 0.0005 & 0.0000 \\ 0.0014 & 0.0005 & 1.2048 & 0.2043 \\ 0.0020 & 0.0000 & 0.2043 & 1.2122 \end{pmatrix}.$$

The matrix shows that $X(0)$ and K owe their influence not only to their individual contribution but also to a strong interaction, as can be seen from its off-diagonal entries.

Regarding the overall deviation of the simulator response from additivity, the design also allows one to estimate the dimension distribution of the simulator. Figure 4.4b shows $\hat{D}_g$ in equation (4.50) as N increases. Starting at about $N \simeq 2000$ we obtain confidence that the dimension distribution is at about $\hat{D}_g = 1.25$, with 95% confidence interval $[1.06, 1.44]$. At

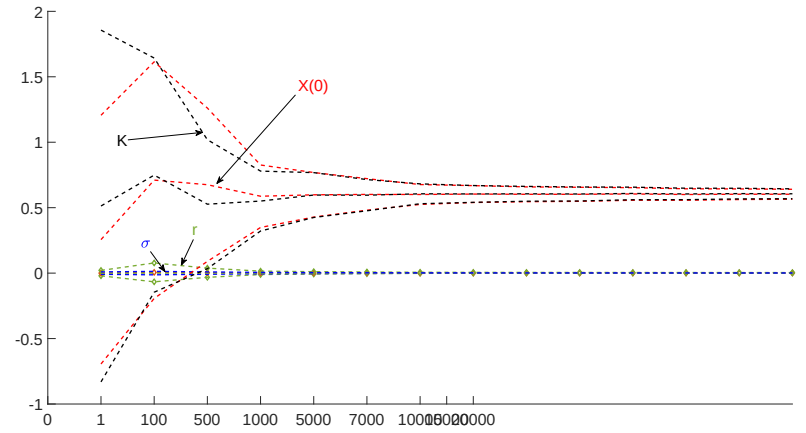


(a)

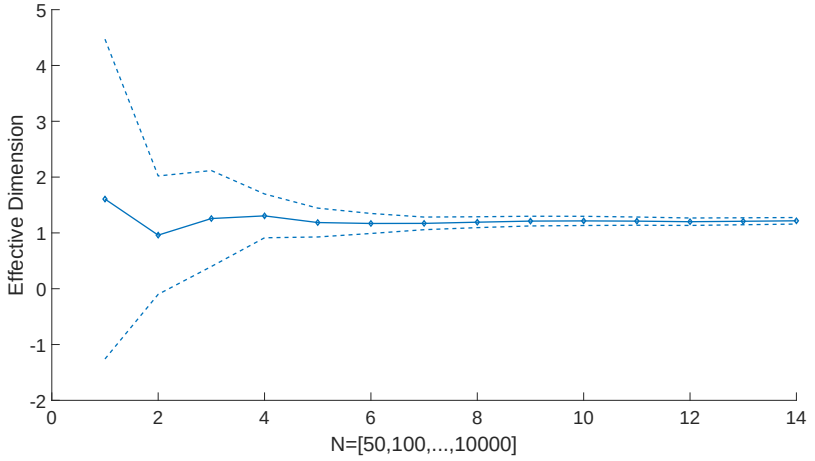


(b)

Figure 4.3: Asian Option: Empirical densities of $\mathbf{\Phi}$ (a) and corresponding empirical mean and variances (b).



(a) $\hat{T}_i$ and corresponding 95% confidence intervals.



(b) $\hat{D}_g$ and corresponding 95% confidence interval.

Figure 4.4: Asian Option: Estimates of Sobol' indices and D_g with corresponding 95% confidence intervals as N increases.

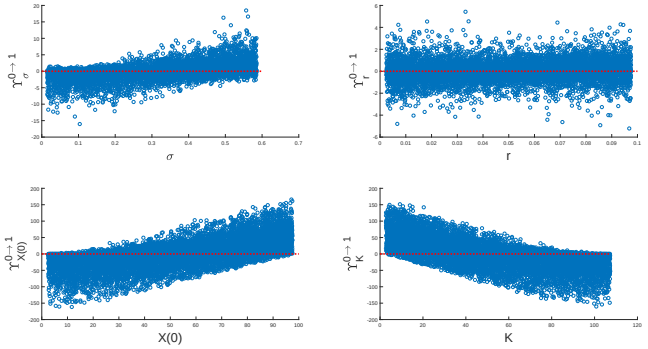
$N = 10000$, we obtain $\hat{D}_g = 1.22$, with 95% confidence interval $[1.16, 1.28]$. This result is consistent with the results in $\widehat{\Upsilon}$, indicating that interactions of order greater than 2 are not relevant.

Concerning the sign of interactions, we report the scatterplot of the $\Upsilon_i^{0\to 1}$ indices in Figure 4.5a. This scatterplot shows both positive and negative realizations of the indices, indicating that the overall interactions can be both positive and negative. This could indicate to the analyst that the simultaneous variations in the simulator inputs may have synergistic or antagonistic effects, depending on their actual location in the input space. If understanding further the precise nature of pairwise interactions as well as their regional behavior is of interest, the analyst can then plan additional experiments aimed at the estimation of pairwise interactions.

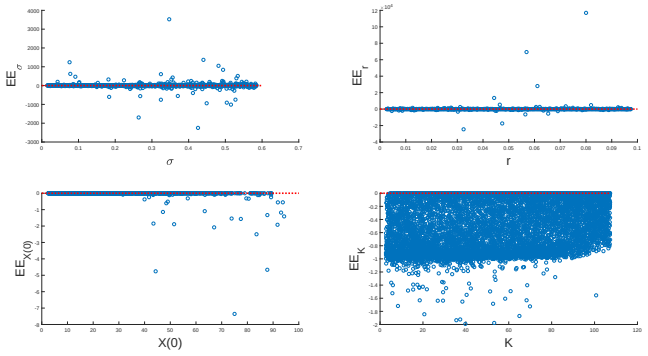
Finally, concerning direction of change, the scatterplots of Morris elementary effects in Figure 4.5b provide a way to visualize the information delivered by Newton ratios. From the graphs, one notes that the realizations of Newton ratios have negative values for $X(0)$ and K, while they assume positive or negative values for σ and r. Here, we observe that the most appropriate setting for the use of Newton ratios is possibly direction of change rather than prioritization. In fact, if one were to calculate Morris $\sigma^2_{EE_i}$ directly from the realizations of the Newton ratios in Figure 4.5b, one would rank σ and r much more important than $X(0)$ and K. However, the Newton ratios of r have different units than, for instance, the Newton ratios of K and, consequently, $\sigma^2_{EE_K}$ and $\sigma^2_{EE_r}$ cannot be compared. In order to compare them, one would have to recompute Morris $\sigma^2_{EE_i}$ after standardizing all the model inputs on the [0,1] a-dimensional scale. This step is not necessary if one uses the empirical variance of main effects.

4.9.2 The Stochastic Assemble-to-Order Simulator

The Stochastic Assemble-to-Order (ATO) model is a discrete-event simulator for supply-chain management [Glasserman and Wang, 1998]. We consider the ATO simulator as in Hong and Nelson [2006], in the implementation by Jing Xie available at www.simopt.org. The system is



(a) $\Upsilon_i^{0\to 1}$ scatterplot.



(b) EE_i scatterplot.

Figure 4.5: Asian Option: Scatterplots of individual and interaction effects providing indications on sign of interactions (a) and direction of change (b).

made of orders form M different customers following independent Poisson arrival processes with constant rates λ_m, $m = 1, ..., M$. Orders are assembled from K different items, requiring a_{mk} items of type k, with $k = 1, ..., K$. For each customer's order, items can be key or non-key. If key items are not available in stock, the customer cancels the order and leaves the system. If all key items are in stock, the customer buys the product assembled by the key items and the available non-key items. The profit for each sold item is p_k and the holding cost for each of them in the inventory is h_k. Each item has inventory capacity c_k and a target-based stock t_k and a demand triggers a replenishment order. The production time for the item k is normally distributed with mean μ_k and standard deviation σ_k (truncated at 0).

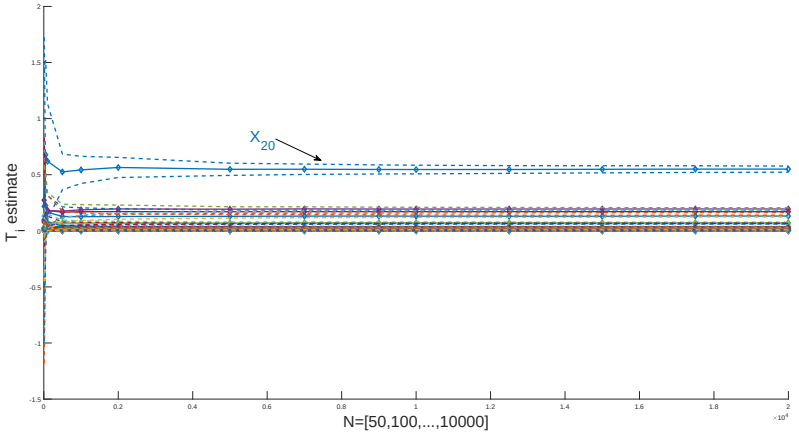
In the simulation the values $M = 5$, $K = 8$ and $c_k = 20$ are assigned. We consider the 22 uncertain inputs detailed in Table 4.2. They comprise the item prices, the target inventory level per item, the holding cost (identical for all items), and the mean arrival time of the customers. We select the mirror design with up to $N = 10000$ replicates. The analysis takes about 19 hours on the above mentioned personal computer. We focus on the identification of the most important inputs and on the mean dimension in this case. Figure 4.6a reports the values of the estimates of T_i, I_i and S_i defined in (4.46), as the number of replicates increases from $N = 100$ to $N = 10000$. For the variance-based sensitivity indices, we record stable estimates for $N \geq 500$. Holding cost for every item (input X_{14}, equal for every item) the price of the sixth item (input X_6) and four target inventory items (inputs X_{15}, X_{18}, X_{19}, X_{20}) have the greatest impact on the final profit, with X_{20} resulting as the most important input according to T_i.

The block of the empirical normalized variance-covariance matrix restricted to the most influential inputs at $N = 10000$, is

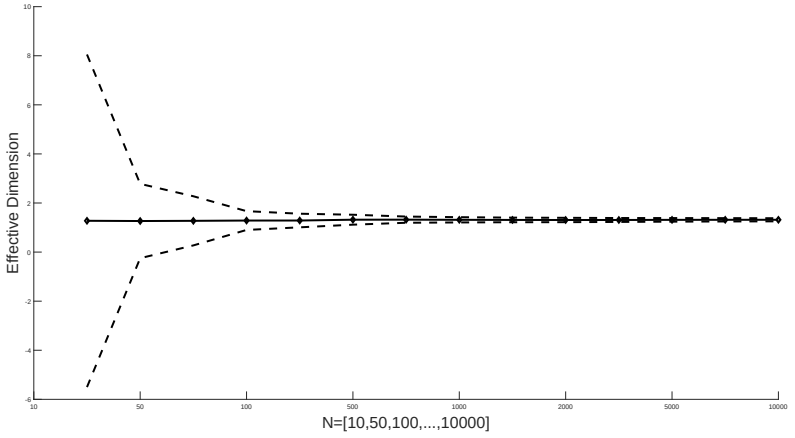
$$\frac{\hat{\Upsilon}_{|14,15,18,20}}{\hat{\sigma}^2} = \begin{pmatrix} 0.3413 & 0.0000 & 0.0000 & 0.0059 \\ 0.0000 & 0.2613 & 0.0052 & 0.0155 \\ 0.0000 & 0.0052 & 0.3883 & 0.0284 \\ 0.0059 & 0.0155 & 0.0284 & 1.0928 \end{pmatrix}.$$

Parameter	Distribution	Parameters		Description
p_1	uniform	0.5	1.5	price item 1
p_2	uniform	1	3	price item 2
p_3	uniform	2	4	price item 3
p_4	uniform	2	6	price item 4
p_5	uniform	3	7	price item 5
p_6	uniform	3	9	price item 6
p_7	uniform	4	10	price item 7
p_8	uniform	4	12	price item 8
λ_1	normal	3.6	.36	arrival customer 1
λ_2	normal	3	.3	arrival customer 2
λ_3	normal	2.4	.24	arrival customer 3
λ_4	normal	1.8	.18	arrival customer 4
λ_5	normal	1.2	.12	arrival customer 5
$h_1 = \cdots = h_8$	uniform	1	3	holding costs
$t_i, i = 1, ..., 8$	discrete uniform	0	10	target inventory item i

Table 4.2: Inputs of the ATO simulator model. The parameters of the uniform distributions are the minimum and maximum, while those of the normal distributions are the mean and standard deviation respectively.



(a) $\hat{T}_i$ with confidence intervals as N increases.



(b) $\hat{D}_g$ and corresponding 95% confidence interval.

Figure 4.6: ATO Simulator: Estimates of variance-based sensitivity indices and dimension distribution as a function of N.

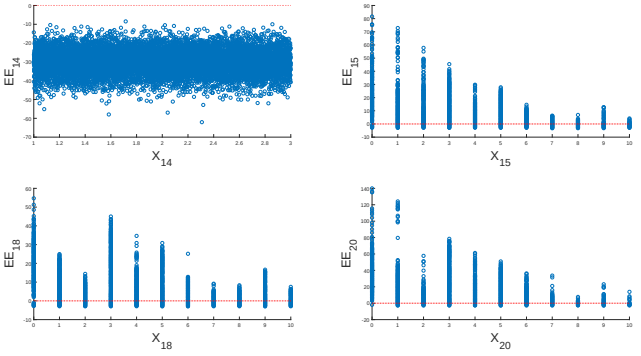


Figure 4.7: ATO simulator: scatterplot of EE_i for the first four most important inputs.

We recall that the diagonal entries are twice the estimates of $\hat{T}_i$ at $N = 10000$. The off-diagonal entries signal that interactions among these four most important inputs are not relevant. This might indicate an overall low relevance of interactions. We look for a confirmation computing the effective dimension. Figure 4.6b displays the point estimates and confidence interval of D_g as N increases. The point estimates converge towards $\hat{D}_g = 1.3162$, with a 95% confidence interval $[1.249, 1.383]$. This result shows that, overall, interactions do not play a major role in the simulator response. Therefore, for the sake of space, we do not enter into results for the sign of the overall interactions in this case, although they are available. Regarding direction of change, Figure 4.7 reports the elementary effects associated with the four most relevant simulator inputs. We observe that the holding cost (X_{14}) is associated with negative realizations of the Newton ratios at all instances. An increase in holding costs is detrimental to the expected profit and a monotonically decreasing behavior is consistent with expectations. The target inventory items (inputs X_{15}, X_{18}, X_{20}) have both positive and negative elementary effects, signaling that the expected profit is not monotonic in these variables. We recall that these variables are the targets of the optimization. A monotonic behavior would rule out

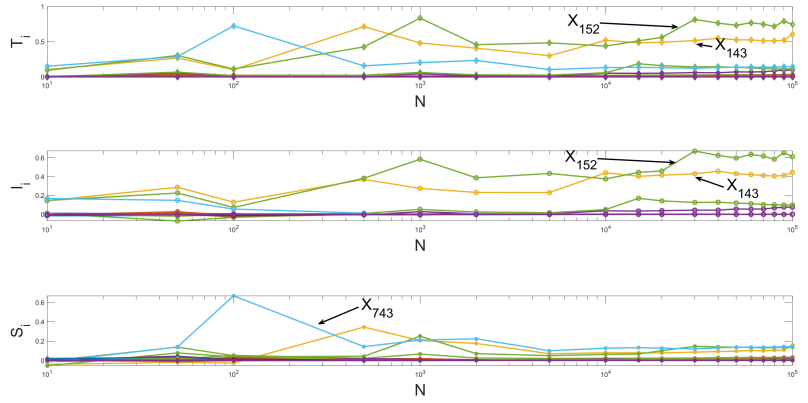
the possibility of a convex behavior of the expected profit.

4.9.3 A probabilistic safety assessment model for NASA space missions

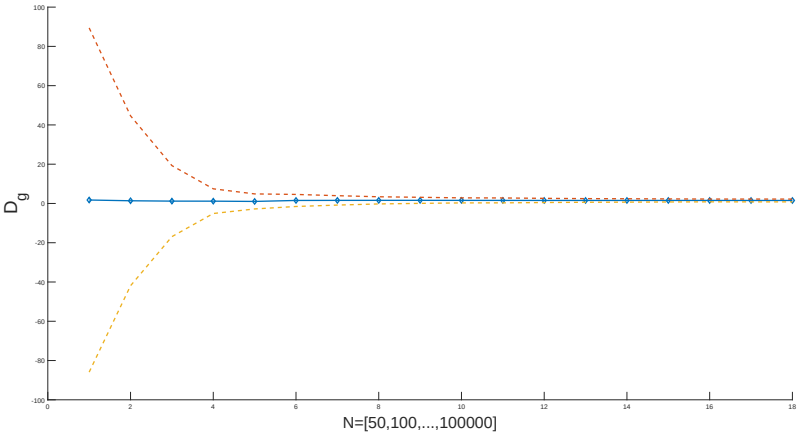
Borgonovo and Smith [2011] study the presence of interactions in the probabilistic safety assessment (PSA) simulator developed to support the designing and planning a NASA lunar space mission. This simulator is aimed to support risk managers towards ensuring the highest levels of performance and safety. The simulator describes eight phases of the space mission, from launch and first orbit around the Earth, to final Earth landing (see Borgonovo and Smith [2011] for further details on the simulator). The output of this computer code is the probability of loss of mission, modeled as a function of 872 inputs, which for privacy reason are labeled by $X_1, X_2, ...$ In Borgonovo and Smith [2011] the simulator is analyzed within a generalized Tornado Diagram approach. The design foresees that the simulator is run at two extreme points of the model input space $\mathbf{x}^0$ and $\mathbf{x}^1$ with no replicates. The sensitivity measures, in the notation of this work, are one set of main effects $\phi_i^{0\to1}$, one total order interaction effect $\Upsilon_i^{0\to1}$ and one total effect $\overline{\tau}_i^{0\to1}$ for each input. We run the simulator (we thank C.L. Smith for making the code available) under the folded-over Morris design with up to $N = 100000$ replicates. The computation takes 2.5 hours on the above-mentioned calculator, with a parallel calculus provision and 4 cores. At the end of the simulations, a sample of 200000 main effects for each input is available.

As far as the identification of the most important inputs is concerned, Figure 4.8a reports the values of the estimates of T_i, I_i and S_i defined in (4.46), as the number of replicates increases from $N = 100$ to $N = 100000$. For T_i and I_i we record stable estimates at $N \geq 4 \times 10^4$ (we do not display confidence intervals for the sake of space).

The most important inputs are X_{152} and X_{143}, whose total order index is five times greater than the total order index of X_{748} the third most important input. These inputs are also the three most important in-



(a) Estimates of $\hat{T}_i$, $\hat{I}_i$, $\hat{S}_i$ as N increases.



(b) $\hat{D}_g$ and corresponding 95% confidence interval for the NASA space PSA simulator as N increases.

Figure 4.8: NASA space PSA: Estimates of variance-based sensitivity indices and dimension distribution as N increases.

puts according to S_i. When compared to results in Borgonovo and Smith [2011], one notes that inputs X_{152}, X_{143} and X_{374} are identified as the most important in that work. This evidences that insights delivered by sensitivity analysis strongly depend on the adopted approach (and scale). In particular, X_{374} is important for the change across the two scenarios analyzed in Borgonovo and Smith [2011], but not on a global scale, when changes across additional scenarios are considered. From the second and third graphs in Figure 4.8a, one observes that X_{152} and X_{143} are involved in significant interactions, while X_{748} is not.

Figure 4.8b displays the point estimates and confidence interval of the effective dimension as the sample size increases. One notes that the estimated effective dimension converges towards a value $\hat{D}_g = 1.63$, with a 95% confidence interval $D_g \in [0.971, 2.295]$. This result shows that, overall, only interactions of order about 2 matter in the global simulator response, but higher order interactions having a negligible role. When compared to results for the ATO simulator, one can note that a percentage-wise wider confidence interval in the NASA case. This is reasonable to expect given the substantially higher dimensionality of this simulator. However, note also that by the estimates we are reassured that the mean dimension distribution will not exceed 2. Given a simulator with 870 inputs, this result signals an overall low relevance of interactions.

Regarding the sign of interactions, results show (we do not display the graphs for space reasons) that the indices $\Upsilon_i^{0\to 1}$ are both be positive and negative. Thus, overall interactions can be antagonistic or synergistic. Regarding trend, all realizations of Newton ratios (Morris elementary effects) are non-negative. This result confirms the monotonically increasing dependence of the probability of loss of mission on the simulator inputs due to the fact that the underlying system is coherent (for further details on the system, please see Borgonovo and Smith [2011]).

4.10 Conclusions

We have offered an investigation of screening approaches whose sensitivity measures are based on one-at-a-time designs. The findings provide a theoretical files-rouge that starts from Tornado diagrams and ends into global sensitivity indices. The approach has allowed us to establish the formal link between Morris $\sigma^2_{EE_i}$ and Sobol' total indices. The asymptotic analysis, while generalizing some previous results, has suggested confidence intervals for the Morris as well as variance-based sensitivity measures and mean dimensions, allowing the analyst to determine the uncertainty in the estimates at finite sample sizes. We have exploited a symmetry effect to propose a design that allows the estimation of Morris elementary effects, total order indices, mean dimension and overall interaction indices at the cost of the Morris method plus one additional run. The design yields information, aside on the relative importance of the inputs, also on whether interaction effects can be synergistic or antagonistic and whether their variations impact positively or negatively the output of interest. In terms of future research, our results show that the most important inputs are the ones associated with the largest confidence intervals, while inputs that are not important are also associated with a lowest variability of the estimates. This information can then be useful for the analyst to search for the optimal number of replicates depending on whether the goal is to accurately estimate the most important inputs or to rapidly identify the non-relevant ones.

Appendix

4.10.1 The case of Dependent Inputs

We now relax the assumption that $\mathbf{X}$ and $\mathbf{X}^1$ are no independently distributed but allow them to be described by the joint probability density $p\left(\mathbf{X}, \mathbf{X}^1\right)$. Under independent inputs, it holds

$$\begin{aligned}\overline{\tau}_i^2 &= \frac{1}{2}\int \left(g\left(X_i^1 : \mathbf{X}_{-i}\right) - g\left(\mathbf{X}\right)\right)^2 p(\mathbf{X})p(\mathbf{X}^1)d\mathbf{X}d\mathbf{X}^1 \\ &= \frac{1}{2}\int \left(\Phi_i^{\mathbf{X}\to\mathbf{X}^1}\right)^2 p(\mathbf{X})p(\mathbf{X}^1)d\mathbf{X}d\mathbf{X}^1 \\ &= \frac{1}{2}\mathbb{E}_{p(\mathbf{X})p(\mathbf{X}^1)}\left[\Phi_i^{\mathbf{X}\to\mathbf{X}^1}\right]^2\end{aligned} \quad . \tag{4.70}$$

Equation (4.70) coincides with equation (4.17).

However, in presence of dependent inputs, Kucherenko et al. [2012] define the variance index $\overline{\tau}_i^{2,ind}$ which accounts for the variance related to the input group indexed by i and its interactions with $-i$, but excludes the variance caused by the dependence of i with $-i$. This way, by the law of the total variance, $\sigma_Y^2 = \overline{\tau}_i^{2,ind} + \overline{\tau}_i^{2,full}$, where $\overline{\tau}_i^{2,full}$ includes the variance caused by the dependence between the input groups indexed by i and $-i$. Kucherenko et al. [2012] and Mara et al. [2015] define the formula

$$\begin{aligned}\overline{\tau}_i^{2,ind} &= \frac{1}{2}\int \left(g\left(X_i^1 : \mathbf{X}_{-i}\right) - g\left(\mathbf{X}\right)\right)^2 p(\mathbf{X})p(X_i^1|\mathbf{X}_{-i})d\mathbf{X}d\mathbf{X}^1 \\ &= \frac{1}{2}\int \left(\Phi_i^{\mathbf{X}\to\mathbf{X}^1}\right)^2 p(\mathbf{X})p(X_i^1|\mathbf{X}_{-i})d\mathbf{X}dX_i^1 \\ &= \frac{1}{2}\mathbb{E}_{p(\mathbf{X})p(X_i^1|\mathbf{X}_{-i})}\left[\Phi_i^{\mathbf{X}\to\mathbf{X}^1}\right]^2 .\end{aligned} \tag{4.71}$$

Hence, the dependence among inputs impacts the finite change indices only indirectly, through the sampling of the points where these indices are constructed.
This is also the case if one adopts the elementary effects under dependent inputs.

Recently, Ge and Menendez [2017] extend the Morris method to the case of dependent inputs. Precisely, these authors use both the radial sampling of Campolongo and the trajectory sampling of Morris to generate independent samples. Then, they transform these samples into dependent samples and construct elementary effects for both the independent and dependent cases.

4.10.2 Proofs

Proof. Proof of Proposition 4.3.1. The proof benefits from exploiting a previous result in the literature that states the following [Borgonovo, 2010, Proposition 1]:

$$\tau_i^{1\to 0} = g(\mathbf{x}^1) - g(x_i^0 : \mathbf{x}_{-i}^1).$$

Then, recognizing that $\phi_i^{1\to 0} = -(g(\mathbf{x}^1) - g(x_i^0 : \mathbf{x}_{-i}^1))$, one has $\phi_i^{1\to 0} = -\tau_i^{1\to 0}$. Thus, because Cotter's design allows us to compute $\phi_i^{0\to 1}$ and $\phi_i^{0\to 1}$, we can compute $\phi_i^{0\to 1}$, $\tau_i^{1\to 0}$ and then, by (4.5),$\Upsilon_i^{0\to 1}$.

□

Proof. Proof of Proposition 4.4.1. Sobol' [1993] shows the identity $\overline{\tau}_i^2 = \frac{1}{2}\mathbb{E}\left[g(X_i^1 : \mathbf{X}_{-i}) - g(\mathbf{X})\right]^2$. Considering that $\phi_i^{0\to 1} = g\left(x_i^1 : \mathbf{x}_{-i}^0\right) - g\left(\mathbf{x}^0\right)$, we obtain equation (4.17). Regarding the sensitivity index $\underline{\tau}_i^2$, Saltelli [2002] shows that it can be rewritten as $\underline{\tau}_i^2 = \mathbb{E}\left[g(\mathbf{X}) \cdot \left(g(X_i : \mathbf{X}_{-i}^1) - g(\mathbf{X}^1)\right)\right]$ and hence $\underline{\tau}_i^2 = \mathbb{E}\left[g(\mathbf{X}) \cdot \Phi_i^{\mathbf{X}^1 \to \mathbf{X}}\right]$. □

Proof. Proof of Proposition 4.5.1. Consider the equality $\Upsilon_i^2 = \underline{\tau}_i^2$ and summing over $i = 1, 2, ..., d$ we have $\sum_{i=1}^d \Upsilon_i^2 = \sum_{i=1}^d \left(\overline{\tau}_i^2 - \underline{\tau}_i^2\right) = \sum_u |u|\sigma_u^2 - \sum_{i=1}^d \underline{\tau}_i^2$ by Theorem 2 in Liu and Owen [2006], which is equation (4.22). Equation (4.23) follows from previous equation writing $\sum_u |u|\sigma_u^2 = \sum_{i=1}^d \underline{\tau}_i^2 + \sum_{|u|>1} |u|\sigma_u^2$. □

Proof. Proof of Corollary 4.8.1. We have

$$\begin{aligned}\hat{\sigma}^2_{EE_i} &= \frac{\sum_{k=0}^{N-1}\left(EE_i^{k\to k+1}\right)^2 - \frac{\left(\sum_{k=0}^{N-1} EE_i^{k\to k+1}\right)^2}{N}}{N-1} \\ &= \frac{\sum_{k=0}^{N-1}\left(EE_i^{k\to k+1}\right)^2}{N-1}\cdot\frac{\Delta^2}{\Delta^2} - \frac{\left(\sum_{k=0}^{N-1} EE_i^{k\to k+1}\right)^2}{N(N-1)} \\ &= \frac{\sum_{k=0}^{N-1}\left(_\Delta\phi_i^{k\to k+1}\right)^2}{(N-1)\Delta^2} - \frac{N\left(\hat{\mu}_i\right)^2}{N-1} \\ &= \frac{N}{N-1}\frac{2_\Delta\hat{\overline{\tau}}_i^2}{\Delta^2} - \frac{N}{N-1}\hat{\mu}_i^2\end{aligned}$$

□

Proof. Proof of Proposition 4.5.2. We start proving equations (4.26). By equation (5.16) it holds,

$$\begin{aligned}\Upsilon_i^2 = \overline{\tau}_i^2 - \underline{\tau}_i^2 &= \frac{1}{2}\mathbb{E}\left[\Phi_i^{\mathbf{X}\to\mathbf{X}^1}\right]^2 - \mathbb{E}\left[g(\mathbf{X})\cdot\Phi_i^{\mathbf{X}^1\to\mathbf{X}}\right] \\ &= \frac{1}{2}\int\left[\left(\Phi_i^{\mathbf{X}\to\mathbf{X}^1}\right)^2 - 2\cdot g(\mathbf{X})\cdot\Phi_i^{\mathbf{X}^1\to\mathbf{X}}\right]d\mathbf{X}d\mathbf{X}^1.\end{aligned} \tag{4.72}$$

Summing over all $i = 1, ..., d$ inputs we get

$$\begin{aligned}\sum_{i=1}^d \Upsilon_i^2 &= \sum_{i=1}^d\left(\overline{\tau}_i^2 - \underline{\tau}_i^2\right) = \sum_{i=1}^d \overline{\tau}_i^2 - \sum_{i=1}^d \underline{\tau}_i^2 \\ &= \sum_{i=1}^d \mathbb{E}\left[g(\mathbf{X})\cdot\left(g(\mathbf{X}) - g(\mathbf{X}_{-i} : X_i^1) + g(\mathbf{X}^1) - g(X_i : \mathbf{X}^1_{-i})\right)\right] \\ &= \sum_{i=1}^d \mathbb{E}\left[-g(\mathbf{X})\cdot\left(\Phi_i^{\mathbf{X}\to\mathbf{X}^1} + \Phi_i^{\mathbf{X}^1\to\mathbf{X}}\right)\right],\end{aligned} \tag{4.73}$$

which proves equation (4.27). Note that we have used the equalities

$$\sum_{i=1}^d \overline{\tau}_i^2 = \sum_{i=1}^d \mathbb{E}\left[g(\mathbf{X})\cdot\left(g(\mathbf{X}) - g(\mathbf{X}_{-i} : X_i^1)\right)\right] = \sum_{i=1}^d \mathbb{E}\left[-g(\mathbf{X})\cdot\Phi_i^{\mathbf{X}\to\mathbf{X}^1}\right] \tag{4.74}$$

and

$$\sum_{i=1}^{d} \underline{\tau}_i^2 = \sum_{i=1}^{d} \mathbb{E}\left[g(\mathbf{X}) \cdot \left(g(X_i : \mathbf{X}_{-i}^1) - g(\mathbf{X}^1)\right)\right] = \sum_{i=1}^{d} \mathbb{E}\left[g(\mathbf{X}) \cdot \Phi_i^{\mathbf{X}^1 \to \mathbf{X}}\right] \tag{4.75}$$

which can be found in [Owen, 2013, p. 32]. Moreover, since $\sum_u |u|\sigma_u^2 = \sum_{i=1}^{d} \overline{\tau}_i^2$, by (4.74) the equation (4.28) follows. □

Proof. Proof of Theorem 4.6.1. Since $\mathbf{X}$ and $\mathbf{X}^1$ are independent and distributed according to $F_{\mathbf{X}}$, also $\mathbf{X}^1$ and $\mathbf{X}$ are. Hence, it follows

$$\underline{\tau}_i^2 = \mathbb{E}\left[g(\mathbf{X}) \cdot \Phi_i^{\mathbf{X}^1 \to \mathbf{X}}\right] = \mathbb{E}\left[g(\mathbf{X}^1) \cdot \Phi_i^{\mathbf{X} \to \mathbf{X}^1}\right]$$

after relabeling the initial and final evaluation points. Equation (4.34) follows combining (4.17) and (4.33) as

$$\sum_{i=1}^{d} \Upsilon_i^2 = \sum_{i=1}^{d} \left(\overline{\tau}_i^2 - \underline{\tau}_i^2\right) = \sum_{i=1}^{d} \left(\frac{1}{2}\mathbb{E}\left[\Phi_i^{\mathbf{X} \to \mathbf{X}^1}\right]^2 - \mathbb{E}\left[g(\mathbf{X}^1) \cdot \Phi_i^{\mathbf{X} \to \mathbf{X}^1}\right]\right).$$

□

Proof. Proof of Lemma 4.7.1. Given $\mathbf{X}$ and $\mathbf{X}^1$ independent and identically distributed under $F_{\mathbf{X}}$, the population mean of the finite change related to the i-th input is zero since

$$\mathbb{E}\left[\Phi_i^{\mathbf{X} \to \mathbf{X}^1}\right] = \int\int \left(g\left(X_i^1 : \mathbf{X}_{-i}\right) - g\left(\mathbf{X}\right)\right) d\mathbf{X} dX_i^1 = 0.$$

Consider now $\mu = \mathbb{E}[g]$. It holds by Owen [2013]

$$\begin{aligned} \mathbb{V}\left[\Phi_i^{\mathbf{X} \to \mathbf{X}^1}\right] &= \mathbb{V}\left[g\left(X_i^1 : \mathbf{X}_{-i}\right) - g\left(\mathbf{X}\right)\right] \\ &= 2\sigma_Y^2 - 2\mathbb{C}\left[g\left(X_i^1 : \mathbf{X}_{-i}\right), g\left(\mathbf{X}\right)\right] \\ &= 2\sigma_Y^2 - 2\left(\Theta_{i,\emptyset} - \mu^2\right) \\ &= 2\sigma_Y^2 - 2\left(\underline{\tau}_{-i}^2 + \mu^2 - \mu^2\right) = 2\overline{\tau}_i^2 \end{aligned} \tag{4.76}$$

using the equality $\sigma_Y^2 = \underline{\tau}_{-i}^2 + \overline{\tau}_i^2$, where the bilinear terms $\Theta_{u,v} = \mathbb{E}\left[g\left(X_u^1 : \mathbf{X}_{-u}\right) \cdot g\left(X_v^1 : \mathbf{X}_{-v}\right)\right]$, $u, v \subseteq \{1, 2, ..., d\}$, constitute the entries of the Sobol' matrix introduced by Owen [2013]. Using the terms $\Theta_{u,v}$, we can now prove equation (4.39). We have

$$\begin{aligned}
\mathbb{C}\left[\Phi_i, \Phi_j\right] &= \mathbb{E}\left[\Phi_i \Phi_j\right] - \mathbb{E}\left[\Phi_i\right]\mathbb{E}\left[\Phi_j\right] \\
&= \mathbb{E}\big[g\left(X_j^1 : \mathbf{X}_{-j}\right) g\left(X_i^1 : \mathbf{X}_{-i}\right) - g\left(X_j^1 : \mathbf{X}_{-j}\right) g\left(\mathbf{X}\right) \\
&- g\left(\mathbf{X}\right) g\left(X_i^1 : \mathbf{X}_{-i}\right) + g\left(\mathbf{X}\right)^2\big] \\
&= \Theta_{i,j} - \Theta_{\emptyset,j} - \Theta_{\emptyset,i} + \sigma^2 \\
&= \underline{\tau}_{-\{i,j\}}^2 - \underline{\tau}_{-j}^2 - \underline{\tau}_{-i}^2 + \sigma^2 \\
&= \bar{\tau}_i^2 + \bar{\tau}_j^2 - \bar{\tau}_{i,j}^2 \\
&= \sum_{u \supseteq \{(i,j)\}} \sigma_u^2 = \Upsilon_{i,j}^2
\end{aligned}$$

To prove (4.40), as in the proof of Proposition 7.1 in Owen [2013], we consider the following empirical estimators of the mean μ

$$\hat{\mu} = \frac{1}{N}\sum_{k=1}^{N} g(\mathbf{x}^0)_{(k)}, \qquad \hat{\mu}' = \frac{1}{N}\sum_{k=1}^{N} g(x_i^1 : \mathbf{x}_{-i}^0)_{(k)}.$$

Note that $\hat{\mu}' - \hat{\mu} = \bar{\Phi}_i$. Then, we have

$$\mathbb{V}\left[\bar{\Phi}_i\right] = \mathbb{E}\left[\left(\hat{\mu}' - \hat{\mu}\right)^2\right] = \mathbb{V}\left(\hat{\mu}'\right) + \mathbb{V}\left(\hat{\mu}\right) - 2\mathbb{C}\left(\hat{\mu}, \hat{\mu}'\right) = \frac{2}{N}\left[\sigma_Y^2 - \underline{\tau}_{-i}^2\right] = \frac{2}{N}\overline{\tau}_i^2.$$

□

Proof. Proof of Theorem 4.7.1. The population variance of the finite change from $\mathbf{X}$ to $\mathbf{X}^1$ is $2\overline{\tau}_i^2$ by equation (4.40). By Lemma 4.7.1 and applying the Central Limit Theorem and the Law of Large Numbers, it follows

$$\bar{\Phi}_i^{\mathbf{X}\to\mathbf{X}^1} \sim \mathcal{N}\left(0, 2\frac{\overline{\tau}_i^2}{N}\right) \quad \text{as} \quad N \to \infty.$$

The population mean of the elementary effects EE_i is μ_i (which is in general different from zero). The population variance is $\frac{2}{\Delta^2}\Delta\overline{\tau}_i^2 - \mu^2$ by

Corollary 4.8.1. Then, the result is a direct application of the Central Limit Theorem to the elementary effects EE_i. □

Proof. Proof of Theorem 4.7.2. To prove equation (4.43), consider the estimator

$$\hat{\underline{\tau}}^2_{i,N} = \frac{1}{N}\sum_{k=1}^{N} g(\mathbf{x}^0_{(k)}) \cdot \left(\phi_i^{1\to 0}\right)_{(k)}.$$

As in Janon et al. [2014], defining the quantity

$$U_k = \left(g(\mathbf{x}^0_{(k)})g(x_i^0 : \mathbf{x}^1_{-i})_{(k)}, g(\mathbf{x}^0_{(k)})g(\mathbf{x}^1_{(k)})\right)$$

one finds

$$\hat{\underline{\tau}}^2_{u,N} = \psi\left(\frac{1}{N}\sum_{k=1}^{N} U_k\right) = \psi\left(\bar{U}_N\right)$$

with

$$\psi(x, y) = x - y.$$

By the Central Limit Theorem, for $N \to \infty$

$$\sqrt{N}\left(\bar{U}_N - \mu\right) \longrightarrow \mathcal{N}(0, \Gamma),$$

where Γ is the covariance matrix of U_1 and the vector of mean values is

$$\mu = \begin{pmatrix} \Theta_{\{1,\dots,d\},i} \\ \Theta_{\{1,\dots,d\},\emptyset} \end{pmatrix}$$

where the bilinear terms $\Theta_{u,v} = \mathbb{E}\left[g\left(X_u^0 : \mathbf{X}^1_{-u}\right) \cdot g\left(X_v^0 : \mathbf{X}^1_{-v}\right)\right]$ are the entries of the Sobol' matrix in Owen [2013]. Hence,

$$\bar{U}_N = \left(\hat{\Theta}_{\{1,\dots,d\},i}, \hat{\Theta}_{\{1,\dots,d\},\emptyset}\right)_{(N)}$$

coincides with the sample estimators given at page 27 in Owen [2013]. By the Delta-method, we then have

$$\sqrt{N}\left(\hat{\underline{\tau}}^2_{i,N} - \underline{\tau}^2_i\right) \longrightarrow \mathcal{N}\left(0, \nabla\psi(\mu)^T \Gamma \nabla\psi(\mu)\right).$$

The gradient is given by $\nabla\psi(x,y) = (1,-1)^T$ so that

$$\begin{aligned}
&\nabla\psi(\mu)^T\Gamma\nabla\psi(\mu) = \\
&= (1,-1)\begin{pmatrix} \mathbb{V}\left[g(\mathbf{X}^0)g(X_i^0:\mathbf{X}_{-i}^1)\right] & \mathbb{C}\left[g(\mathbf{X}^0)g(X_i^0:\mathbf{X}_{-i}^1), g(\mathbf{X}^0)g(\mathbf{X}^1)\right] \\ \mathbb{C}\left[g(\mathbf{X}^0)g(X_i^0:\mathbf{X}_{-i}^1), g(\mathbf{X}^0)g(\mathbf{X}^1)\right] & \mathbb{V}\left[g(\mathbf{X}^0)g(\mathbf{X}^1)\right] \end{pmatrix} \\
&= \mathbb{V}\left[g(\mathbf{X}^0)g(X_i^0:\mathbf{X}_{-i}^1)\right] + \mathbb{V}\left[g(\mathbf{X}^0)g(\mathbf{X}^1)\right] - 2\mathbb{C}\left[g(\mathbf{X}^0)g(X_i^0:\mathbf{X}_{-i}^1), g(\mathbf{X}^0)g(\mathbf{X}^1)\right] \\
&= \mathbb{V}\left[g\left(\mathbf{X}^0\right)\cdot\phi_i^{1\to 0}\right].
\end{aligned}$$

In particular, note that

$$\psi\left(\Theta_{\{1,\dots,d\},i}, \Theta_{\{1,\dots,d\},\emptyset}\right) = \underline{\tau}_i^2.$$

Formula (4.44) can be proven similarly. Precisely, posing $U_k = \left[g(x_i^1:\mathbf{x}_{-i}^0)_{(k)} - g(\mathbf{x}_{(k)}^1)\right]^2$, by the Central Limit Theorem one finds

$$\sqrt{N}\left(\bar{U}_N - \mu\right) \longrightarrow \mathcal{N}\left(0,\Gamma\right)$$

where $\Gamma = \mathbb{V}\left[\left(\phi_i^{0\to 1}\right)^2\right]$ and

$$\begin{aligned}
\mu &= \int \left(\phi_i^{0\to 1}\right)^2 d\mathbf{X}^0 d\mathbf{X}^1 = 2\left(\mathbb{E}\left[g^2\right] - \Theta_{i,\emptyset}\right) = 2\left(\mathbb{E}\left[g^2\right] - \left(\mathbb{E}\left[g\right]\right)^2 - \underline{\tau}_{-i}^2\right) \\
&= 2\left(\mathbb{V}\left[g\right] - \underline{\tau}_{-i}^2\right) = 2\overline{\tau}_i^2.
\end{aligned}$$

Then, apply the Delta-method using the transformation $\psi(t) = \frac{1}{2}t$ and consider that $\mathbb{V}\left[(\Phi_i^{\mathbf{X}\to\mathbf{X}^1})^2\right] = \mu_4 - \left(2\overline{\tau}_i^2\right)^2 - \mu^4 - 4\overline{\tau}_i^2\mu^2 = \mu_4 - \left(2\overline{\tau}_i^2\right)^2$ being the mean of finite changes zero. The limit (4.45) follows from (4.44), the fact that the finite changes are independent and $\sum_{i=1}^d \overline{\tau}_i^2 = \sum_u |u|\sigma_u^2$ (Theorem 2.2 of Owen [2013]). □

Proof. Proof of Theorem 4.7.3. We prove (4.49). Consider the asymptotic normality of the sample variance estimator $\sqrt{N}\left(S_N^2 - \sigma_Y^2\right) \to \mathcal{N}\left(0, \mu_4 - \sigma_Y^2\right)$

(see, e. g., van der Vaart [2000]) . Hence, by (4.45)

$$\sqrt{N}\begin{pmatrix}\sum_{i=1}^{d}\hat{\overline{\tau}}^2_{i,N}\\ S_N^2\end{pmatrix}\longrightarrow$$
$$\mathcal{N}\left(\begin{pmatrix}\sum_u |u|\sigma_u^2\\ \sigma_Y^2\end{pmatrix},\begin{bmatrix}\frac{1}{4}\sum_{i=1}^{d}\mathbb{V}\left[(\Phi_i^{\mathbf{X}\to\mathbf{X}^1})^2\right] & \mathbb{C}\left[\frac{1}{2}\sum_{i=1}^{d}(\Phi_i^{\mathbf{X}\to\mathbf{X}^1})^2,(g(\mathbf{X})-\mu)^2\right]\\ \mathbb{C}\left[\frac{1}{2}\sum_{i=1}^{d}(\Phi_i^{\mathbf{X}\to\mathbf{X}^1})^2,(g(\mathbf{X})-\mu)^2\right] & \mu_4-\sigma_Y^2\end{bmatrix}\right) \tag{4.77}$$

Consider the function $\psi(x,y)=\frac{x}{y}$ with gradient $\nabla\psi(x,y)=(\frac{1}{y},-\frac{x}{y^2})$. It holds for $N\to\infty$

$$\psi\left(\sum_{i=1}^{d}\hat{\overline{\tau}}^2_{i,N},S_N^2\right)=\frac{\sum_{i=1}^{d}\hat{\overline{\tau}}^2_{i,N}}{S_N^2}\to\frac{\sum_u|u|\sigma_u^2}{\sigma_Y^2}=D_g. \tag{4.78}$$

By the Delta-method, we then have

$$\sqrt{N}\left(\psi\left(\sum_{i=1}^{d}\hat{\overline{\tau}}^2_{i,N},S_N^2\right)-\psi\left(\sum_u|u|\sigma_u^2,\sigma_Y^2\right)\right)\longrightarrow$$
$$\mathcal{N}\left(0,\nabla\psi\left(\sum_u|u|\sigma_u^2,\sigma_Y^2\right)^T\Gamma\nabla\psi\left(\sum_u|u|\sigma_u^2,\sigma_Y^2\right)\right)$$

where Γ is the variance-covariance matrix in (4.77). It follows

$$\begin{aligned}\nabla\psi\left(\sum_u|u|\sigma_u^2,\sigma_Y^2\right)^T\Gamma\nabla\psi\left(\sum_u|u|\sigma_u^2,\sigma_Y^2\right)&=\left(\frac{1}{\sigma_Y^2},-\frac{\sum_u|u|\sigma_u^2}{\sigma_Y^4}\right)^T\Gamma\left(\frac{1}{\sigma_Y^2},-\frac{\sum_u|u|\sigma_u^2}{\sigma_Y^4}\right)\\&=\frac{\sum_{i=1}^{d}\mathbb{V}\left[(\Phi_i^{\mathbf{X}\to\mathbf{X}^1})^2\right]}{4\sigma_Y^4}+\frac{D_g^2\left(\mu_4-\sigma_Y^4\right)}{\sigma_Y^4}-2\frac{D_g\varpi}{\sigma_Y^4}\\&=\frac{\sum_{i=1}^{d}\mathbb{V}\left[(\Phi_i^{\mathbf{X}\to\mathbf{X}^1})^2\right]+4D_g^2\left(\mu_4-\sigma_Y^4\right)-8D_g\varpi}{4\sigma_Y^4}.\end{aligned}$$

where $\varpi = \mathbb{C}\left[\frac{1}{2}\sum_{i=1}^{d}(\Phi_i^{\mathbf{X}\to\mathbf{X}^1})^2, (g(\mathbf{X})-\mu)^2\right]$. Considering that $\mathbb{V}\left[(\Phi_i^{\mathbf{X}\to\mathbf{X}^1})^2\right] = \mu_4 - \left(2\overline{\tau}_i^2\right)^2$, the limit (4.49) follows.
The limits (4.47) and (4.48) can be proved analogously considering (4.43) and (4.44) respectively. □

Proof. Proof of Proposition 4.8.1. The variance of the elementary effects EE_i can be found by consistency of the estimators of the sample variance (4.56) and of the sample mean (4.7). Equation (4.61) can be obtained considering that the variance of the sample mean is the sample variance divided by the number of observations N. □

Chapter 5

A Shapley-Owen index for interaction quantification

Abstract

The recent work of Art Owen [2014, J. Uncertainty Quantification, 2, p. 245-251] has introduced the Shapley value as an importance measure for global sensitivity analysis (Shapley effect, henceforth). When inputs are dependent, using Shapley effects provides a strategy to overcome conceptual difficulties related to the interpretation of Sobol' sensitivity indices. However, Shapley effects have been formulated thus far only to quantify the importance of individual model inputs, without providing information about interactions. This article extends the above-mentioned work to propose a Shapley sensitivity measure for interaction effects. We make use of the generalized Shapley value introduced by Guillermo Owen [1972, Management Science, 18(52), p. 64-79] and axiomatized later on in Grabisch and Roubens [1999, Int. J. Game Theory, 28(4), p. 547-565]. In parallel to the work of Art Owen, we propose this Shapley-Owen effect as a tool for global interaction quantification in presence of dependent inputs. We show that using this index it is also possible to gain insights on the

synergistic/antagonistic nature of interactions.

5.1 Introduction

Running computer experiments is nowadays a standard procedure to support scientific investigations in several disciplines, from engineering to chemistry. Scientific models are often regarded as black-boxes, because in realistic applications the functional form of the input-output mapping is not available analytically. When model inputs are uncertain, global sensitivity analysis supports is essential for analysts to gain insights about the key-drivers of uncertainty. Sobol' indices are among the most well known global sensitivity measures [Sobol', 1993]. These importance measures rank model inputs based on their contribution (individual or in interactions with other model inputs) to the model output variance. However, the assumption of model input independence, which plays a crucial role in granting a transparent interpretation of these indices, may not hold in realistic applications. To deal with this complication, alternative methods have been developed to restore the Sobol' approach via generalized functional ANOVA decompositions [Chastaing et al., 2012, Rahman, 2014] or via the use of moment-independent sensitivity measures [Borgonovo and Plischke, 2016].

Art Owen [2014] has proposed the use of the Shapley value [Shapley, 1953], typical in the economic literature, to measure variable importance bypassing the problem of decomposing variance under dependence [Owen and Prieur, 2017]. In Economics, the Shapley value attributes the value created by a team to each of its members. If the team is identified by a set of random model inputs and the economic value generated by the team is identified with their explanatory power, then the Shapley value becomes a measure of the relative model input importance. In this way, the Shapley value has recently received attention as a tool for global sensitivity analysis [Song et al., 2016, Owen and Prieur, 2017, Iooss and Prieur, 2019]. However, there is no analogous approach for measuring the importance of interaction effects, because the Shapley value assigns value to singular

inputs/agents.

In this paper, we consider the Shapley-Owen value for interactions in groups of players as introduced by Guillermo Owen [Owen, 1972], which extends the Shapley value for singular agents to a group of agents S. We show that the Shapley-Owen effect can also provide information about the nature of interaction in presence of dependent inputs. That is, this global index allows to discriminate whether an interaction is synergistic (i.e., two model inputs are related by a positive cooperation) or antagonistic (negative cooperation among the model inputs). We show how this Shapley-Owen effect is connected to well-known interaction measures, such as the superset importance measure of Liu and Owen [2006] and the factorial moments of the dimension distribution of Owen [2003]. We also discuss the link between Shapley effects and Sobol' indices for the case in which the analyst is not able to specify a unique joint distribution for the model inputs, but assigns a mixture of possible distributions.

The remainder of this paper is organized as follows. Section 2 introduces the functional ANOVA decomposition and presents Sobol' indices. Section 3 reviews the Shapley value as an importance measure for individual model inputs. Section 4 presents the Shapley-Owen value, its axiomatization and its connection to the Shapley value, the superset importance index and the dimension distribution. Section 5 shows that the invariance properties of the (individual) Shapley effect don't generalize to the (group) Shapley-Owen effect. Section 6 contains results on Sobol' and Shapley-Owen values under mixtures. Section 7 illustrates the analysis of interactions via test cases of Owen and Prieur [2017], Iooss and Prieur [2019].

5.2 Functional ANOVA and global sensitivity analysis

Consider a function $g(\mathbf{x}) : \mathcal{X} \to \mathbb{R}$, where $\mathcal{X} \subseteq \mathbb{R}^n$. This function may represent the input-output mapping of a computer code or a surrogate

model of a complex system, and is generally not available in closed form. Let's assume that g is square integrable and that it depends on n inputs x_j, with $j \in N := \{1, 2, ..., n\}$. For a subset $T \subseteq N$, denote by $|T|$ its cardinality and by $-T = N \setminus T$ its complement.

Let the model inputs be uncertain and let $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \mu)$ denote the corresponding probability space, with $\mu : \mathcal{B}(\mathcal{X}) \to [0, 1]$ a product measure. Under the assumption that $g \in L^2(\mathcal{X}, \mathcal{B}(\mathcal{X}), \mu)$, the functional ANOVA representation of g [Efron and Stein, 1981] is written as:

$$g(\mathbf{x}) = \sum_{T \subseteq N} g_T(x_T),$$

where x_T are the components of $\mathbf{x}$ indexed by $T \subseteq N$ and the functions g_T are recursively defined by

$$g_T(x_T) = \int \left(g(\mathbf{x}) - \sum_{L \subset T} g_L(x_L) \right) d\mu(x_{-T}).$$

The components $g_T(x_T)$ are referred to as the functional ANOVA effects of x_T. They are orthogonal: $\int g_T(x_T) g_L(x_L) d\mathbf{x} = 0$ for $T \neq L$. Defining the variance components $\sigma_T^2 = \mathbb{V}[g_T(x_T)]$, effect orthogonality leads to the well-known decomposition formula

$$\sigma^2 = \mathbb{V}[g(\mathbf{x})] = \sum_{T \subseteq N \setminus \{\emptyset\}} \sigma_T^2. \tag{5.1}$$

The global sensitivity analysis of g is typically performed using Sobol' indices. The two importance indices for subset T are

$$\underline{\tau}_T^2 = \mathbb{V}\left[\mathbb{E}\left(g(\mathbf{x})|x_T\right)\right] = \sum_{L \subseteq T} \sigma_L^2 \tag{5.2}$$

and

$$\overline{\tau}_T^2 = \mathbb{E}\left[\mathbb{V}\left(g(\mathbf{x})|x_{-T}\right)\right] = \sum_{L : L \cap T \neq \emptyset} \sigma_L^2. \tag{5.3}$$

In particular, the index $\underline{\tau}_T^2$ represents the variance explained by x_T and can be considered as a natural importance measure for the set T. Conversely, $\overline{\tau}_T^2$ can be interpreted as the expected remaining variance once the inputs x_{-T} are known [Song et al., 2016]. These indices satisfy $\underline{\tau}_T^2 \leq \overline{\tau}_T^2$ and $\overline{\tau}_T^2 = \sigma^2 - \underline{\tau}_{-T}^2$.

5.3 The Shapley value

The Shapley value of model input x_j, seen as a player in a cooperative game, is a measure of the value of player j in a set of players $T \subseteq N$. It is given by

$$\phi_j^{\nu} = \sum_{T \subseteq N \setminus \{j\}} \frac{(n - |T| - 1)!|T|!}{n!} \left[\nu(T \cup \{j\}) - \nu(T)\right], \tag{5.4}$$

where the value function $\nu(T)$, the value attained by the subcoalition T, is assumed to satisfy the condition $\nu(\emptyset) = 0$. The properties of value functions have been extensively studied in the literature. For theoretical aspects, we refer to works such as Grabisch and Roubens [1999], Grabisch [2006], who present alternative axiomatizations of $\nu(T)$ showing that value functions of monotonic games coincides with Choquet capacities, among other results. Based on previous considerations in Section 2, Owen [2014] and Owen and Prieur [2017] suggest as value function of subset T in a sensitivity analysis the variance explained by its members x_j with $j \in T$, that is

$$\nu(T) = \underline{\tau}_T^2. \tag{5.5}$$

Song et al. [2016] prove that one obtains the same Shapley value by setting

$$\nu(T) = \bar{\tau}_T^2. \tag{5.6}$$

and Owen and Prieur [2017] write that *this provides an alternative way to compute Shapley value* (p. 990). Since $\sum_{j=1}^{n} \phi_j^{\nu} = 1$ and $\phi_j^{\nu} \geq 0$ for every $j = 1, ..., n$, Owen (2014) Owen [2014] inteprets equation (5.4) as an importance measure for variable x_j. When the inputs are independent and the value function is given by (5.2), Owen [2014] proves that the Shapley effect for the j-th input can be written as

$$\phi_j^{\underline{\tau}^2} = \sum_{T \subseteq N : j \in T} \frac{\sigma_T^2}{|T|}, \tag{5.7}$$

meaning that the Shapley effect takes into account the Sobol' individual effect plus all the interaction effects involving j divided by the size of the

coalition/interaction term. This connection between Sobol' indices and the Shapley effect can be seen by the bracketing property [Owen, 2014]

$$\underline{\tau}_j^2 \leq \phi_j^{\tau^2} \leq \overline{\tau}_j^2. \tag{5.8}$$

Subsequent works, such as Song et al. [2016], Owen and Prieur [2017] and Iooss and Prieur [2019] also discuss the fact that Shapley effects remain appropriate also in the presence of correlations among inputs. Namely, in this case, the calculation and interpretation of Sobol' indices become difficult (see Chastaing et al. [2012], Rahman [2014]). Hence, other approaches have been developed to deal with dependent inputs in computer experiments [Borgonovo and Plischke, 2016].

5.4 Interactions and the Shapley-Owen value

The Shapley value is a way to apportion the value of a game to one of its players. Thus, it is an importance measure at an individual level. However, the question emerges of what is the value of the coalition of two or more players and whether this coalition is synergetic, i.e., the value of the coalition is higher than the sum of the Shapley values in individual players. This leads to the question of measuring interactions when using Shapley effects. A series of works in the Economics literature has addressed this question and the measure of interactions in Shapley values is the so-called Shapley-Owen value, introduced by Guillermo Owen [Owen, 1972]. The Shapley-Owen value is a well-known method of attributing value to the interactions in a coalition of players in a game. We follow here the presentation given in Grabisch and Roubens [1999]. The Shapley-Owen value for players of a coalition S with value function ν is denoted ϕ_S^ν and is defined by:

$$\phi_S^\nu = \sum_{T \subseteq N \setminus S} \frac{(n - |T| - |S|)!|T|!}{(n - |S| + 1)!} \sum_{L \subseteq S} (-1)^{|S|-|L|} \nu(L \cup T). \tag{5.9}$$

This index represents the residual interaction value of a coalition of players $\{i_1, i_2, ...i_S\}$. To illustrate, consider $S = \{i, j\}$. Then, we are interested

in the interaction between inputs and (5.9) becomes

$$\phi_{i,j}^{\nu} = \sum_{T \subseteq N \setminus \{i,j\}} \frac{(n-|T|-2)!|T|!}{(n-1)!} \left[\nu(T \cup \{i,j\}) - \nu(T \cup \{i\}) - \nu(T \cup \{j\}) + \nu(T)\right] \tag{5.10}$$

which coincides with the index of interaction importance defined by Grabisch and Labreuche [2001]. The underlying idea in the construction of this index is that, for instance in the case of two players i and j, the sign and the magnitude of $\nu(T \cup \{i,j\}) - \nu(T \cup \{i\}) - \nu(T \cup \{j\}) + \nu(T)$, averaged over all other possible coalitions T, should give information about the interaction between the two players. In fact, there is a link between value and interaction: if the interaction is positive, then joining the coalition is profitable and the value of the coalition is greater than the sum of the individual values. Conversely, in the case of a negative interaction, the value of the coalition is less than the sum: note that both effects depend on the chosen value function.

As it happens for the Shapley value, also the Shapley-Owen value for interactions can be characterized through several appealing properties [Grabisch and Roubens, 1999]:

1. (Linearity) $\phi_S^{\nu+w} = \phi_S^{\nu} + \phi_S^{w}$ for every $S \subseteq N$ and for any value function ν and w.

2. (Dummy) If i is a dummy player for ν, then $\phi_i^{\nu} = \nu(\{i\})$ and $\phi_{S\cup\{i\}}^{\nu} = 0$ for every $S \subseteq N \setminus \{i\}$ with $S \neq \emptyset$.

3. (Symmetry) For all ν and for all permutations π on N, $\phi_S^{\nu} = \phi_{\pi S}^{\pi\nu}$, where the game $\pi\nu$ is defined by $\pi\nu(\pi S) = \nu(S)$, where $\pi S = \{\pi(i), i \in S\}$ for all $S \subseteq N$.

4. (Recursivity) ϕ^{ν} obeys the following recurrence formula for every $S \subseteq N, |S| > 1$, any ν and any $j \in S$: $\phi_S^{\nu} = \phi_{S\setminus\{j\}}^{\mu_j} - \phi_{S\setminus\{j\}}^{\nu^{N\setminus\{j\}}}$ where $\mu_j(S) = \nu(S \cup \{j\}) - \nu(\{j\})$ and $\nu^{N\setminus\{j\}}$ denotes the value of the game on $N \setminus \{j\}$ players.

Grabish and Roubens Grabisch and Roubens [1999] prove that the Shapley-Owen value for interactions (5.9) is the only interaction index that satisfies

the above properties and whose restriction to singletons corresponds to the Shapley value in (5.4). The Shapley-Owen effect can be also expressed as Grabisch [1997], Grabisch and Roubens [1999]

$$\phi_S^\nu = \sum_{T \supseteq S} \frac{1}{|T| - |S| + 1} m(T), \tag{5.11}$$

where m is the Moebius transform of the value function ν. As in Owen Owen [2014], we consider (5.5) as the value function. For notation simplicity, we pose $\phi_S^{\mathcal{T}^2} = \phi_S$, when there is no risk of ambiguity. Following the terminology of Song et al. [2016], in the context of global sensitivity analysis we refer to ϕ_S as the Shapley-Owen effects of S. With this choice of $\nu(T)$, Grabisch and Roubens Grabisch and Roubens [1999]'s properties 2 and 3 above admit the following interpretation for sensitivity measures. If the i-th model input is a dummy variable, then it doesn't interact with any other input, so that $\phi_i = \underline{\tau}_i^2$ and $\phi_{S \cup \{i\}} = 0$. On the other hand, when considering input group S, the order in which the inputs are considered is irrelevant, because ϕ_S is invariant under any permutation of S because of the symmetry property $\underline{\tau}_S^2 = \underline{\tau}_{\pi S}^2$. We now illustrate the recursivity axiom considering first the pair of inputs x_i and x_j. To link value and interaction, Grabisch and Roubens [1999] require that the interaction effect satisfies

$$\phi_{ij}^\nu = \phi_{[ij]}^{\nu^{[ij]}} - \phi_j^{N \setminus i} - \phi_i^{N \setminus j}, \tag{5.12}$$

where $[ij]$ is a single hypothetical player representing the inputs i and j and $\nu^{[ij]}$ is the reduced game defined as $\nu^{[ij]}(S) = \nu(S)$ and $\nu^{[ij]}(S \cup [ij]) = \nu(S \cup \{i, j\})$ for any $S \subseteq N \setminus \{i, j\}$. The recursivity axiom is a generalization of (5.12) to any coalition and Grabisch and Roubens [1999] write that it *has an interesting interpretation [...] the interaction of the players in S is equal to the interaction between the players in $S \setminus j$ in the omnipresence of j, minus the interaction between the players of $S \setminus j$ (in the absence of j)* (p. 558). Note that this property holds for variance-based sensitivity indices under independence. In fact, by orthogonality, we have that $\sigma_{i,j}^2 = \mathbb{E}[Y|X_i, X_j] - \sigma_i^2 - \sigma_j^2$, where $\sigma_{i,j}^2$ plays the role of ϕ_{ij}^ν, $\mathbb{E}[Y|X_i, X_j]$ of $\phi_{[ij]}^{\nu^{[ij]}}$, σ_j^2 of $\phi_j^{N \setminus i}$ and σ_i^2 of $\phi_i^{N \setminus j}$.

Then, Grabisch and Roubens (1999) Grabisch and Roubens [1999] show that the interaction effect satisfying the above three axioms takes the following form:

$$\phi_S = \sum_{T \subseteq N \setminus S} \frac{(n - |T| - |S|)!|T|!}{(n - |S| + 1)!} \sum_{L \subseteq S} (-1)^{|S|-|L|} \underline{\tau}^2_{L \cup T}. \tag{5.13}$$

As a special case, for singletons, i.e. for $S = \{j\}$, the Shapley-Owen effect (5.13) coincides with the Shapley effect, because

$$\phi_j = \frac{1}{n} \sum_{T \subseteq N \setminus \{j\}} \binom{n-1}{|T|}^{-1} \sum_{L \subseteq \{j\}} (-1)^{1-|L|} \underline{\tau}^2_{L \cup T} = \frac{1}{n} \sum_{T \subseteq N \setminus \{j\}} \binom{n-1}{|T|}^{-1} \left[\underline{\tau}^2_{T \cup j} - \underline{\tau}^2_T \right], \tag{5.14}$$

which is the Shapley value (5.4) with value function (5.5) (see equation (1) in Owen [2014]). When inputs are independent, ϕ_S can be characterized further.

Theorem 5.4.1. *Consider the ANOVA decomposition in* (5.1) *with independent inputs. Then, the Shapley-Owen effect is*

$$\phi_S = \sum_{T \supseteq S} \frac{\sigma^2_T}{|T| - |S| + 1}. \tag{5.15}$$

Proof. It follows immediately from equation (5.11), where the Moebius transform of (5.2) is given by $\sum_{L \subseteq T} (-1)^{|T|-|L|} \underline{\tau}^2_L = \sigma^2_T$ (see, e. g., Liu and Owen [2006]). □

Equation (5.15) is a generalization of the result for Shapley effects in Theorem 1 of Owen [2014]. Indeed, considering the singleton $S = \{i\}$, from (5.15) one immediately obtains (5.7). Moreover, we directly find

$$\sigma^2_S \leq \phi_S \leq \Upsilon^2_S \tag{5.16}$$

where $\Upsilon^2_S = \sum_{T \supseteq S} \sigma^2_T$ is the superset importance index [Liu and Owen, 2006] that involves the interactions to which group S contributes. Inequality (5.16) is the 'interaction' version of the bracketing property for Shapley effects (5.8). We can further investigate the relation between Shapley-Owen effects, superset importance indices and the dimension distribution.

Let U be a randomly chosen subset of $1, 2, ..., n$ with $\Pr(U = u) = \sigma_u^2/\sigma^2$. This distribution is called dimension distribution [Owen, 2003].

Theorem 5.4.2. *Assume that the model inputs are independent. Then, for $s \geq 2$,*

$$\sum_{|S|=s} \phi_S = \frac{\sigma^2}{s!} \mu^{(s-1)}, \tag{5.17}$$

where $\mu^{(k)} = \mathbb{E}[|U|(|U|-1)\cdots(|U|-k+1)]$ is the k-th factorial moment of the dimension distribution.

Proof. By equation (5.11) we get

$$\begin{aligned}\sum_{|S|=s} \phi_S &= \sum_{|S|=s} \sum_{T \supseteq S} \frac{1}{|T|-s+1} m(T) = \sum_T m(T) \frac{1}{|T|-s+1} \sum_{|S|=s} \mathbb{I}_{T \supseteq S} \\ &= \sum_T m(T) \frac{1}{|T|-s+1} \binom{|T|}{s} = \frac{1}{s} \sum_T \binom{|T|}{s-1} m(T)\end{aligned}$$

Under input independence, by Theorem 4 in Liu and Owen Liu and Owen [2006] one finds

$$\sum_{|S|=s} \phi_S = \frac{1}{s} \sum_T \binom{|T|}{s-1} \sigma_T^2 = \frac{1}{s} \sum_{|V|=s-1} \Upsilon_V^2 = \frac{1}{s} \frac{\sigma^2}{(s-1)!} \mu^{(s-1)}.$$

□

Equation (5.17) holds for $s = 1$, as well. Since the zeroth factorial moment is unity, equation (5.17) yields $\sum_{i=1}^n \phi_i = \sigma^2$, which is a property of the Shapley effect (see Owen [2014]). This property of Shapley effects holds also in case of dependent inputs [Owen and Prieur, 2017].

The highest-order Shapley-Owen effect (5.13), i.e. the Shapley-Owen effect of all indices ($S = N$) becomes

$$\phi_N = \sum_{L \subseteq N} (-1)^{n-|L|} \underline{\tau}_L^2 = \sigma_N^2 \tag{5.18}$$

by equation (6) in Liu and Owen [2006]. In this case, by (5.16) the Shapley-Owen effect becomes the $N-$order term in the ANOVA decomposition.

When inputs are dependent, the sign and magnitude of σ_N^2 provide insights about the the relevance of the interaction among all inputs, as well as about whether the overall interaction is synergistic or antagonistic, respectively. Conversely, when inputs are independent, this term is always non-negative. We illustrate these concepts further with the example in Section 5.7.1.

5.5 Invariance, bijections and equivalence classes

The aim of this section is to investigate the impact of the selection of the value function on the form of Shapley-Owen effects.
Following the approach in Owen and Prieur [2017], suppose that we transform each input x_j into $z_j = f_j(x_j)$, where f_j is a bijection for every $j = 1, ...n$. Let's define $\tilde{g}(\mathbf{z}) = g\left(f_1^{-1}(z_1), ..., f_n^{-1}(z_n)\right)$ and denote by ψ_S the Shapley-Owen effect for the group of variables S in $\tilde{g}(\mathbf{z})$. Then, because $\mathbb{V}\left[\mathbb{E}\left(\tilde{g}(\mathbf{z})|z_T\right)\right] = \mathbb{V}\left[\mathbb{E}\left(g(\mathbf{x})|x_T\right)\right]$, then $\psi_S = \phi_S$, where ϕ_S is the value related to the function $g(\mathbf{x})$.

In general, let's consider the set V of all possible value functions ν. Then, on V consider the relation

$$\nu_1 \sim \nu_2 \quad \text{if} \quad \phi_S^{\nu_1} = \phi_S^{\nu_2} \quad \forall S \subseteq N.$$

Since this relation is clearly symmetric, reflexive and transitive, it is an equivalence relation. Its equivalence classes are made of all the value functions that generate the same Shapley-Owen effect. It follows that, by the previous result, these classes are invariant under bijections of their input variables. The question is whether we can characterize these equivalence classes in general. As a matter of fact, Song et al. [2016] prove that the Shapley effects defined with the values $\underline{\tau}_T^2$ and $\overline{\tau}_T^2$ coincide. However, this does not hold for Shapley-Owen effects. As an example, consider the case of $N = 2$ independent inputs and focus on ϕ_{12}. Using the identity

$\sigma^2 - \overline{\tau}^2_{-T} = \underline{\tau}^2_T$, by equation (5.18), we have

$$\begin{aligned}\phi_{12}^{\underline{\tau}^2} &= \sum_{L\subseteq\{1,2\}} (-1)^{2-|L|}\underline{\tau}^2_L \\ &= \underline{\tau}^2_{12} - \underline{\tau}^2_1 - \underline{\tau}^2_2 = \sigma^2_{12} = \overline{\tau}^2_1 + \overline{\tau}^2_2 - \overline{\tau}^2_{12} \\ &= -\sum_{L\subseteq\{1,2\}} (-1)^{2-|L|}\overline{\tau}^2_L = -\phi_{12}^{\overline{\tau}^2}.\end{aligned}$$

Note the symmetry: $\phi_{12}^{\overline{\tau}^2}$ is always non-positive, $\phi_{12}^{\underline{\tau}^2}$ always non-negative and they coincide in absolute value.
Then, differently from the Shapley effect, it is more complicated to find value functions which generate the same Shapley-Owen effect for measuring interactions. However, we may require a weaker condition:

$$\nu_1 \sim \nu_2 \quad \text{if} \quad sign(\phi_S^{\nu_1}) = sign(\phi_S^{\nu_2}) \quad \forall S \subseteq N,$$

that is we only require that the two value functions generate Shapley-Owen indices of the same sign. Thus, they provide the same information about the nature (but not the magnitude) of the interaction. Namely, the final aim is to obtain indications about the presence of synergistic or antagonistic effect. This point should be further investigated in the future.

5.6 Sobol' and Shapley-Owen Effects under Mixtures

It is often a difficult task for the analyst to assign an unique distribution to the inputs of a computer experiment. The work [Borgonovo et al., 2018] considers the set $\mathcal{M}$ of possible distributions of the model inputs. For the sake of simplicity, consider $\mathcal{M}$ finite or countable, so that $\mathcal{M} = \{\mu^1, ..., \mu^M\}$. Then, the analyst may opt for representing the uncertainty through the mixture distribution

$$\mu(\mathbf{x}) = \sum_{i=1}^{M} p_m \mu^m(\mathbf{x}), \tag{5.19}$$

assigning a prior $P_\mu = \{p_1, ..., p_M\}$, with $p_m > 0$ and $\sum_{i=1}^{M} p_m = 1$, over the component measures in $\mathcal{M}$. The following relation with the Sobol' indices holds. Let ${}_\mu\bar{\tau}_L^2$, ${}_\mu\underline{\tau}_L^2$ and ${}_{\mu^m}\bar{\tau}_L^2$ denote the Sobol' indices under the mixture distribution (5.19), prior P_μ, and under a given component of the mixture, respectively.

Proposition 5.6.1. Given the probability space $(\mathcal{M}, \mathcal{F}(\mathcal{M}), P_\mu)$, suppose that
$f \in \bigcup_{\mu^m \in \mathcal{M}} L^2(\mathcal{X}, \mathcal{B}(\mathcal{X}), \mu^m)$. Then ${}_\mu\bar{\tau}_L^2$ and ${}_\mu\underline{\tau}_L^2$ are given, respectively, by

$$ {}_\mu\underline{\tau}_L^2 = \mathbb{V}_{P_\mu}\left[\mathbb{E}_{\mu^m}\left(f(\mathbf{x})|x_L\right)\right] \tag{5.20} $$

and

$$ {}_\mu\bar{\tau}_L^2 = \sum_{m=1}^{M} p_m \cdot {}_{\mu^m}\bar{\tau}_L^2. \tag{5.21} $$

Proof. By the law of total variance

$$ \mathbb{V}\left[f(\mathbf{x})\right] = \mathbb{E}_{P_\mu}\left[\mathbb{V}\left(f(\mathbf{x})|\mu^m\left(x_L\right)\right)\right] + \mathbb{V}_{P_\mu}\left[\mathbb{E}\left(f(\mathbf{x})|\mu^m\left(x_L\right)\right)\right]. $$

In particular, equation (5.21) holds since

$$ {}_\mu\bar{\tau}_{-L}^2 = \mathbb{E}_{P_\mu}\left[\mathbb{V}\left(f(\mathbf{x})|\mu^m\left(x_L\right)\right)\right] = \sum_{i=1}^{M} p_m \mathbb{E}\left[\mathbb{V}\left(f(\mathbf{x})|\mu^m\left(x_L\right)\right)\right] = \sum_{i=1}^{M} p_m \cdot {}_{\mu^m}\bar{\tau}_{-L}^2. $$

□

Both Sobol' indices ${}_\mu\bar{\tau}_L^2$ and ${}_\mu\underline{\tau}_L^2$ can be used as value functions to construct the Shapley-Owen interaction effect. In particular, using ${}_\mu\bar{\tau}_L^2$, one can find the following characterization.

Corollary 5.6.1. *Under the assumptions of Proposition 5.6.1, it holds:*

$$ \phi_S^{{}_\mu\bar{\tau}^2} = \sum_{m=1}^{M} p_m \cdot \phi_S^{{}_{\mu^m}\bar{\tau}^2} \tag{5.22} $$

The proof is immediate by linearity. Note that an analogous representation cannot be found using ${}_{\mu}\underline{\tau}_L^2$ as value function. Equation (5.22) shows that the Shapley-Owen effect corresponding to a mixture of Sobol' indices is the mixture of the corresponding Shapley-Owen effects with the same weights. However, as shown in Borgonovo et al. [2018] the mixture of Sobol' indices does not equal the total model output variance, as it excludes the portion of the variance due to the variation of the expected value of $g(\mathbf{X})$ across the alternative distributions in the mixture. However, the overall Shapley effect that applies to the mixture distribution $\mu_{\mathbf{X}}$ taken as reference distribution for the analysis remains well defined. Note that because correlations are induced by $\mu_{\mathbf{X}}$, Sobol' indices are not readily defined for $\mu_{\mathbf{X}}$.

5.7 Test cases

In this section, we compute Shapley-Owen effects for test cases developed in previous works on Shapley effects [Owen and Prieur, 2017, Iooss and Prieur, 2019].

5.7.1 Bivariate Gaussian linear model

We consider the bivariate Gaussian linear model given in Iooss and Prieur [2019]. Formally, $g(\mathbf{X}) = \beta^\top \mathbf{X}$ with $\mathbf{X} \sim \mathcal{N}(\mu, \Gamma)$, where

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \quad \text{and} \quad \Gamma = \begin{pmatrix} \gamma_1^2 & \rho\gamma_1\gamma_2 \\ \rho\gamma_1\gamma_2 & \gamma_2^2 \end{pmatrix},$$

with $-1 \leq \rho \leq 1$, $\gamma_1 > 0$ and $\gamma_2 > 0$. One finds that the output variance is given by

$$\sigma^2 = \mathbb{V}\left[g(\mathbf{X})\right] = \beta_1^2\gamma_1^2 + 2\rho\beta_1\beta_2\gamma_1\gamma_2 + \beta_2^2\gamma_2^2. \tag{5.23}$$

By (5.5), the Sobol' indices are $\underline{\tau}_\emptyset^2 = 0, \underline{\tau}_1^2 = \left(\beta_1\gamma_1 + \rho\beta_2\gamma_2\right)^2$, $\underline{\tau}_2^2 = \left(\beta_2\gamma_2 + \rho\beta_1\gamma_1\right)^2$ and $\underline{\tau}_{12}^2 = \sigma^2$. By equation (5.14), we can find the Shapley

effect ϕ_j for inputs $j = 1, 2$, which are given by (see Iooss and Prieur [2019])

$$\sigma^2\phi_1 = \beta_1^2\gamma_1^2(1-\frac{\rho^2}{2})+\rho\beta_1\beta_2\gamma_1\gamma_2+\beta_2^2\gamma_2^2\frac{\rho^2}{2}, \quad \sigma^2\phi_2 = \beta_2^2\gamma_2^2(1-\frac{\rho^2}{2})+\rho\beta_1\beta_2\gamma_1\gamma_2+\beta_1^2\gamma_1^2\frac{\rho^2}{2}.$$

We obtain the Shapley-Owen effect ϕ_{12} by equation (5.13), as:

$$\begin{aligned}\phi_{12} &= \sum_{L\subseteq\{1,2\}} (-1)^{2-|L|}\underline{\tau}_L^2 = \sigma^2 - \underline{\tau}_1^2 - \underline{\tau}_2^2 \\ &= -\rho^2\beta_2^2\gamma_2^2 - 2\rho\beta_1\beta_2\gamma_1\gamma_2 - \rho^2\beta_1^2\gamma_1^2.\end{aligned}$$

Note that ϕ_{12} is null for independent inputs (and thus equation (5.18) holds with the Sobol' index $\sigma_{12}^2 = 0$). Under correlations, ϕ_{12} can be positive or negative. Figure 5.1 shows the value of ϕ_{12} as a function of ρ

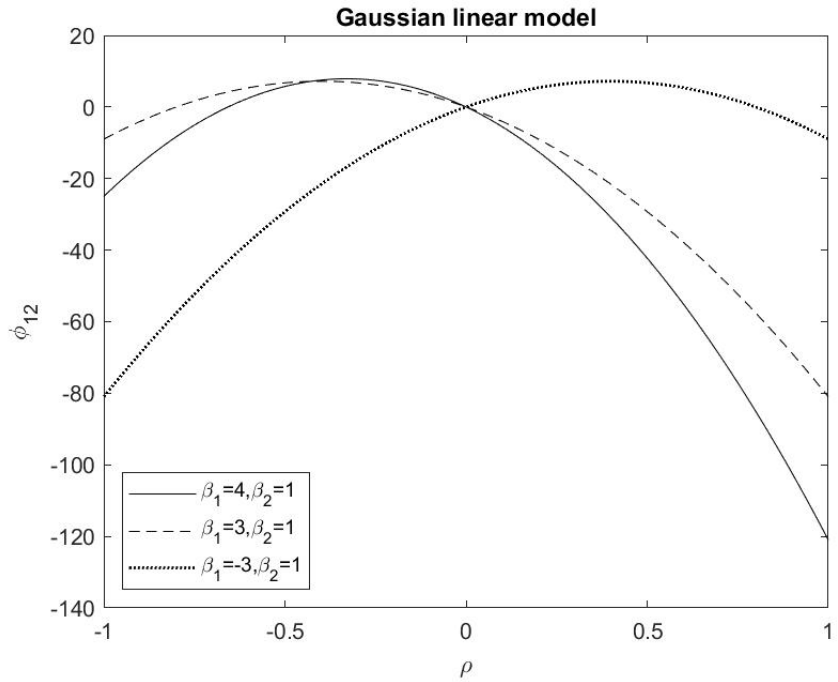


Figure 5.1: ϕ_{12} as a function of ρ for the example in Section 5.7.1.

for the parameterization $\gamma_1 = 2, \gamma_2 = 3, \beta_2 = 1$ and $\beta_1 = 4$ (continuous

line), $\beta_1 = 3$ (dashed line) and $\beta_1 = -3$ (dotted line). The expression of the Shapley-Owen effect ϕ_{12} shows that it has a quadratic behaviour, as shown in Figure 5.1. It is then interesting to investigate the maximum cooperation. By differentiation, one finds the point of maximum at

$$\rho^* = -\frac{\beta_1\beta_2\gamma_1\gamma_2}{\beta_1^2\gamma_1^2 + \beta_2^2\gamma_2^2}.$$

This point of maximum lies in $[-1, 0)$ when the coefficients β_1 and β_2 are concordant, and in $(0, 1]$ when they are discordant. In particular, for $\rho < \rho^* (\rho > \rho^*)$, ϕ_{12} is strictly increasing(decreasing). In correspondence of ρ^*, one has

$$\phi_{12}^* = \frac{\beta_1^2\beta_2^2\gamma_1^2\gamma_2^2}{\beta_1^2\gamma_1^2 + \beta_2^2\gamma_2^2} > 0,$$

suggesting that the maximal cooperation between X_1 and X_2 is always positive. Moreover, at ρ^*, we have that the term $2\rho\beta_1\beta_2\gamma_1\gamma_2$ in (5.23) becomes

$$2\rho^*\beta_1\beta_2\gamma_1\gamma_2 = -2\frac{\beta_1^2\beta_2^2\gamma_1^2\gamma_2^2}{\beta_1^2\gamma_1^2 + \beta_2^2\gamma_2^2} < 0,$$

which indicates that the maximum cooperation is attained in connection to a reduction of the model output variance.

In the statistical literature, an interaction effect which is present even if g is linear is called *spurious* [Friedman and Popescu, 2008]. In the context of predictive modeling, Friedman and Popescu [2008] deem an interaction effect spurious if it is present in the predictive model but not in the true model underlying the real relationships among inputs and output. They write that *These spurious interactions can occur when there is a high degree of collinearity among some (or all) of the predictor variables* (p. 936). In this case, the Shapley-Owen effect sheds light on the mechanism of spurious interactions — see Section 5.7.3.

5.7.2 Bivariate linear model with copula dependence

We consider the same functional linear model $g(\mathbf{X}) = \beta^\top\mathbf{X}$ as before, but, as in Owen and Prieur [2017], we use the Farlie-Gumbel-Morgenstern

(FGM) copula for the vector $\mathbf{X} \in [0,1]^2$, whose joint distribution function is described by

$$c_\theta(\mathbf{X}) = 1 + \theta(1-2X_1)(1-2X_2), \qquad -1 \leq \theta \leq 1,$$

where each component is marginally uniform, $X_i \sim U[0,1]$, $i = 1, 2$. The parameter θ tunes the dependence in $\mathbf{X}$. One can show that $\text{cor}(X_1, X_2) = \theta/3$. Owen and Prieur Owen and Prieur [2017] prove that the output variance is given by $\sigma^2 = (\beta_1^2 + \beta_2^2)/12 + \beta_1\beta_2\theta/18$ and that the Shapley effects are

$$\phi_1 = \frac{\sigma^2}{2}\left(1 + \left(1 - \frac{\theta^2}{9}\right)\frac{\beta_1^2 - \beta_2^2}{12\sigma^2}\right), \quad \phi_2 = \frac{\sigma^2}{2}\left(1 + \left(1 - \frac{\theta^2}{9}\right)\frac{\beta_2^2 - \beta_1^2}{12\sigma^2}\right).$$

Owen and Prieur Owen and Prieur [2017] also provide the expressions for the Sobol' indices

$$\underline{\tau}_1^2 = \frac{1}{12}\left(\beta_1 + \frac{\theta}{3}\beta_2\right)^2 \quad \text{and} \quad \underline{\tau}_2^2 = \frac{1}{12}\left(\beta_2 + \frac{\theta}{3}\beta_1\right)^2.$$

Inserting these expressions into (5.18) one finds the Shapley-Owen effect

$$\phi_{12} = -\frac{1}{18}\left(\frac{\theta^2}{6}(\beta_1^2 + \beta_2^2) + \theta\beta_1\beta_2\right).$$

The effect ϕ_{12} can be positive and negative as well. Figure 5.2 reports three different cases. The continuous line is the case $\beta_1 = 3, \beta_2 = 1$; the dashed line in the case $\beta_1 = -4, \beta_2 = -1$; the dotted line in the case $\beta_1 = -0.5, \beta_2 = 1$. Figure 5.2 shows that the Shapley-Owen effect evidences some similarities with the interaction effect for the same functional form for g of the test case in section 5.7.1, even if the inputs are assigned different supports, marginal distributions and dependence structure in this test case.

5.7.3 Gaussian model with three inputs and interaction term

In this section, we examine the model with interactions studied in Iooss and Prieur Iooss and Prieur [2019], given by $g(\mathbf{X}) = X_1 + X_2X_3$ with

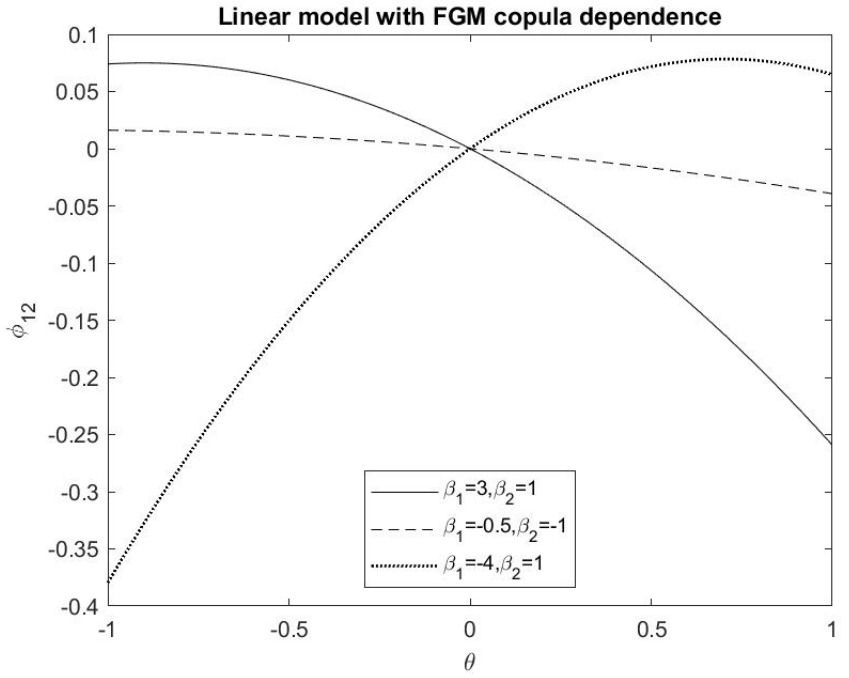


Figure 5.2: ϕ_{12} as a function of θ for the example in Section 5.7.2.

$\mathbf{X} \sim \mathcal{N}(\mu, \Gamma)$, where

$$\mu = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \Gamma = \begin{pmatrix} \gamma_1^2 & 0 & \rho\gamma_1\gamma_3 \\ 0 & \gamma_2^2 & 0 \\ \rho\gamma_1\gamma_3 & 0 & \gamma_3^2 \end{pmatrix}.$$

The Shapley effects are derived in Iooss and Prieur [2019]

$$\phi_1 = \gamma_1^2\left(1-\frac{\rho^2}{2}\right)+\frac{\gamma_2^2\gamma_3^2}{6}\rho^2, \quad \phi_2 = \frac{\gamma_2^2\gamma_3^2}{6}(3+\rho^2), \quad \phi_3 = \frac{\rho^2\gamma_1^2}{2}+\frac{\gamma_2^2\gamma_3^2(3-2\rho^2)}{6}.$$

Note that, differently from Iooss and Prieur [2019], we are working with the non-normalized Sobol' indices as value function. Saltelli and Tarantola [2002] consider a model with a similar dependence between X_1 and X_2 and

write that *interactions may be "carried over" by correlation. This effect is a possibility only when the input is correlated, and is absent when the input is not correlated* (p. 705). We investigate this effect by means of the Shapley-Owen indices. After some algebra, we obtain

$$\phi_{12} = \frac{\rho^2\gamma_2^2\gamma_3^2}{2}, \quad \phi_{13} = -\rho^2\big(\gamma_1^2 + \frac{\gamma_2^2\gamma_3^2}{2}\big), \quad \phi_{23} = \left(1 - \frac{\rho^2}{2}\right)\gamma_2^2\gamma_3^2.$$

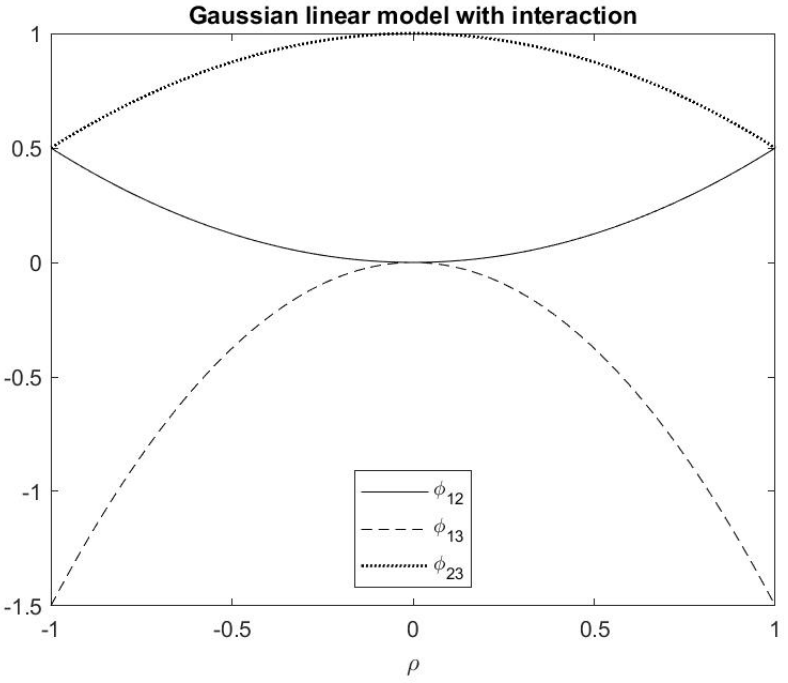


Figure 5.3: Shapley-Owen indices as functions of ρ for the example in Section 5.7.3.

Figure 5.3 visualizes these indices for the case of unitary variances. Note that ϕ_{23} is indeed related to the structural interaction term X_2X_3. However, as the absolute value of the correlation increases, the value of

ϕ_{23} decreases and the value of ϕ_{12} increases. Because ϕ_{12} depends on ρ^2 and ϕ_{23} on $1 - \rho^2/2$, the value from ϕ_{23} is actually carried over to ϕ_{12} (this provides a formal explanation to the above-mentioned observation of Saltelli and Tarantola [2002]). This interaction is then spurious. Note that ϕ_{12} and ϕ_{23} coincide in case of maximal dependence, meaning that the spurious interaction between X_1 and X_2 has the same explanatory power as the structural interaction between X_2 and X_3. This is reasonable because for perfectly correlated X_1 and X_3, the product X_2X_3 behaves as the product X_1X_2, which is not present in the original model. Also, under dependence, ϕ_{23} is not bracketed between σ^2_{23} and Υ^2_{23}. Indeed, it holds that $\sigma^2_{23} = \gamma^2_2\gamma^2_3$ and $\Upsilon^2_{23} = (1-\rho^2)\gamma^2_2\gamma^2_3$ and hence the inequality (5.16) is inverted, with

$$\Upsilon^2_{23} \leq \phi_{23} \leq \sigma^2_{23}.$$

This result is similar to the results of Iooss and Prieur [2019], who show that for the Shapley effects ϕ_1 and ϕ_3 the inequality (5.8) is inverted under input dependence. We also observe that ϕ_{13} is always non-positive, indicating that X_1 and X_3 possess a lower explanatory power when considered jointly rather than when considered individually.

Let us then analyse the case in which model inputs are independent (we can then set $\rho = 0$). This case allows us to investigate more closely the relationship between the Shapley-Owen effect ϕ_{23} and the corresponding variance-based interaction index σ^2_{23}, that can be calculated analytically. In fact, it turns out that the Shapley-Owen effect ϕ_{23} coincides with σ^2_{23} and with the superset importance Υ^2_{23} in this case. Under independence, we also have $\sigma^2_{12} = \sigma^2_{13} = 0$. Note that in this case also ϕ_{12} and ϕ_{13} vanish.

5.7.4 Maximum of exponential random variables

Keinan at al. Keinan et al. [2004] consider a network of neurons $e_1, ..., e_n$ where every neuron e_i has independent lifetime x_j exponentially distributed with mean λ_i, $i = 1, 2, ..., n$. These authors adopt the Shapley effect to determine the importance of the elements of the network. In their setting the value function of a group of neurons is the mean amount of survival

time of at least a part of the group, that is $\nu(T) = \mathbb{E}[\max_{i \in T} x_i]$. They provide the expression of the Shapley effect in the case $n = 3$. Owen and Prieur Owen and Prieur [2017] prove the general equation for arbitrary n (see [Owen and Prieur, 2017, p. 1001])

$$\phi_j = \sum_{T: j \in T} \frac{1}{|T|} (-1)^{|T|-1} \frac{1}{\sum_{l \in T} \lambda_l}.$$

We can extend this result to the Shapley-Owen effect between pairs of neurons, neurons i and j, $i, j = 1, 2, ..., n$.

Theorem 5.7.1. *Let the value function for a set $T \subseteq N$ be $\nu(T) = \mathbb{E}[\max_{i \in T} x_i]$ where $x_1, ..., x_n$ are independent distributed exponential random variables with mean $1/\lambda_1, ..., 1/\lambda_n$ respectively. Then, the Shapley-Owen effect* (5.10) *for the interaction between x_i and x_j is*

$$\phi_{ij} = \sum_{T: \{i,j\} \in T} \frac{1}{|T| - 1} (-1)^{|T|-1} \frac{1}{\sum_{l \in T} \lambda_l}.$$

Proof. See Appendix. □

5.8 Conclusions and future research

In this paper, we have proposed the Shapley-Owen effect as a global measure of interactions. The fact that the indices are well posed in the presence of input dependencies and that they deliver insights on whether interactions are synergistic or antagonistic, are desirable properties for global sensitivity measures.

Open issues for future research comprise both methodological and computational aspects. On the theoretical side, a further comparison of the type of insight on interactions delivered by Shapley-Owen effects with respect to variance-based indices in the presence of input dependence. This entails to compare in detail the construction underlying the Shapley-Owen effects to the construction underlying variance-based indices as defined in

works such as Li et al. [2010], Chastaing et al. [2012], Rahman [2014]. A second research line is the interpretation of Shapley-Owen effects using alternative value functions. In fact, we have seen that, while the value functions $\overline{\tau}^2$ and $\underline{\tau}^2$ are equivalent for Shapley effects (as proven in Song et al. [2016]), they are not for Shapley-Owen effects. Thus, the subject of the "best" value function to be chosen for measuring interaction effects is an open one.

A further relevant research line is the study of the numerical estimation of Shapely-Owen effects. This study is, in turn, related to the estimation of individual Shapley-effects, which is a topical research subject, and whose importance is underlined in works such as Song et al. [2016], Iooss and Prieur [2019], Broto et al. [2019], and Benoumechiara and Elie-Dit-Cosaque [2019]. In this respect, we note that a brute force algorithm for the calculation of individual Shapley effects would lead to all terms required for the calculation of Shapley-Owen effects as well. However, such algorithm would become soon infeasible, given the high number of terms it would require. On the other hand, the permutation strategy of Song et al. Song et al. [2016] represents a first approach to decrease the number of model evaluations of individual Shapley effects. Benoumechiara and Elie-Dit-Cosaque Benoumechiara and Elie-Dit-Cosaque [2019] couple such an approach with a metamodel, to further reduce the compuational time. Then, one can argue that a similar approach could represent a breakthrough towards the computationally efficient estimation of Shapley-Owen effects. However, further research is needed to understand whether the Song et al.'s permutation strategy can be modified for the calculation of Shapley-Owen effects. Another potential application of our findings concerns the Bayesian analysis of Shapley-Owen effects under mixtures, with the possibility of nesting these measures in a Bayesian paradigm similarly to the suggestion in Owen and Prieur [2017] for Shapley effects.

Appendix

Proof. Proof of Theorem (5.7.1). By Proposition 6.1 in Owen and Prieur [2017], one finds

$$\nu(T) = \sum_{\emptyset \neq L \subseteq T} (-1)^{|L|-1} \frac{1}{\sum_{l \in L} \lambda_l}.$$

Following the proof of Theorem 4.8 in Owen and Prieur [2017], one finds that $\nu(T \cup \{i,j\}) - \nu(T \cup \{i\}) - \nu(T \cup \{j\}) + \nu(T)$ equals to

$$\begin{aligned}
&\sum_{\emptyset \neq L \subseteq T \cup \{i,j\}} (-1)^{|L|-1} \frac{1}{\sum_{l \in L} \lambda_l} - \sum_{\emptyset \neq L \subseteq T \cup \{i\}} (-1)^{|L|-1} \frac{1}{\sum_{l \in L} \lambda_l} \\
&- \sum_{\emptyset \neq L \subseteq T \cup \{j\}} (-1)^{|L|-1} \frac{1}{\sum_{l \in L} \lambda_l} + \sum_{\emptyset \neq L \subseteq T} (-1)^{|L|-1} \frac{1}{\sum_{l \in L} \lambda_l} = \\
&= \sum_{W \subseteq T \cup \{i\}} (-1)^{|W|} \frac{1}{\sum_{l \in W \cup \{j\}} \lambda_l} - \sum_{W \subseteq T} (-1)^{|W|} \frac{1}{\sum_{l \in W \cup \{j\}} \lambda_l} \\
&= \sum_{Z \subseteq T} (-1)^{|Z|-1} \frac{1}{\sum_{l \in Z \cup \{i,j\}} \lambda_l}.
\end{aligned}$$

Inserting this expression in (5.10) it follows that

$$\begin{aligned}
\phi_{ij} &= \sum_{T \subseteq N \setminus \{i,j\}} \frac{(n - |T| - 2)!|T|!}{(n-1)!} \left[\nu(T \cup \{i,j\}) - \nu(T \cup \{i\}) - \nu(T \cup \{j\}) + \nu(T)\right] \\
&= \sum_{T \subseteq N \setminus \{i,j\}} \frac{(n - |T| - 2)!|T|!}{(n-1)!} \sum_{Z \subseteq T} (-1)^{|Z|-1} \frac{1}{\sum_{l \in Z \cup \{i,j\}} \lambda_l} \\
&= \sum_{Z \subseteq N \setminus \{i,j\}} (-1)^{|Z|-1} \frac{1}{\sum_{l \in Z \cup \{i,j\}} \lambda_l} \sum_{V \subseteq N \setminus Z \cup \{i,j\}} \frac{(n - |Z| - |V| - 2)!(|Z| + |V|)!}{(n-1)!}
\end{aligned}$$

where we have introduced the variable V so that $T = Z \cup V$. Then, we write the internal sum on V as

$$\begin{aligned}
\sum_{r=0}^{n-|Z|-2} \frac{(n-|Z|-r-2)!(|Z|+r)!}{(n-1)!}\binom{n-2-|Z|}{r} &= \sum_{r=0}^{n-|Z|-2} \frac{(|Z|+r)!(n-|Z|-2)!}{r!(n-1)!} \\
&= \frac{(n-|Z|-2)!}{(n-1)!} \sum_{r=0}^{n-|Z|-2} \frac{(|Z|+r)!}{r!}\frac{|Z|!}{|Z|!} \\
&= \frac{(n-|Z|-2)!}{(n-1)!} \sum_{r=0}^{n-|Z|-2} \binom{|Z|+r}{r} \\
&= \frac{(n-|Z|-2)!}{(n-1)!}\binom{n-1}{n-|Z|-2} \\
&= \frac{1}{|Z|+1}
\end{aligned}$$

where we have used the equality for binomial coefficient

$$\sum_{r=0}^{K}\binom{L+r}{r} = \binom{K+L+1}{K}.$$

Thus, it follows that

$$\phi_{ij} = \sum_{Z \subseteq N\setminus\{i,j\}} \frac{1}{|Z|+1}(-1)^{|Z|-1}\frac{1}{\sum_{l\in Z\cup\{i,j\}}\lambda_l},$$

which yields the result of the Theorem after the change of variable $Z \cup \{i,j\} = T$. □

Chapter 6

Conclusions and Perspectives

As shown in the Chapters of this Thesis, the study of interactions can not be a straightforward exercise in the sensitivity analysis of computer experiments. In particular, this Thesis has developed material about the identification, quantification and interpretation of interactions. The general reccomendation is to integrate and use different estimation methods which can deliver complenetary rather redundant information. This information is essential to correctly interpret the code at hand and/or to understand whether the modellists need to adjust the computer model.

In particular, in the Thesis I have developed new approaches to study interactions. On one hand, the whole interaction structure can be investigated by replicated one-factor-at-a-time (OFAT) experiments. This result is to me quite surprising and in the future I will inestigate its application to supersaturated OFAT designs in the context of screening large experiments, such as the Iterated Fractional Factorial Designs. Also, the use of tensiorial products might be beneficial to discover higher order interaction structures.

It also emerged from the Thesis (Chapter 2) that some authors mean

by interactions the Cox's notion of deviation of additivity of the effects. Lancaster instead defines interactions as the pressence of statistical dependence between two random effects. In a way, this is the philosophical distinction between the approach of Sobol' (implicitly based on Cox's definition) and that one of moment-independent measures (Lancaster's definition) in computer experiments. With my advisor and Elmar Plischke I am going to investigate more deeply this interesting relationship.

In Chapter 4 one example concerns the sequential bifurcation method used for screening with finite changes. In light of the connections between Sobol' indices and Morris' elementary effects described in that Chapter, I think that it will be interesting to understand the relationship between variance-based totals and this methods. The question could be: are the Kleijnen estimators "statistically" connected to Sobol' indices?

On the other hand, a new interaction measure has been proposed in Chapter 5, namely the Shapley-Owen effects. In my opinion this opens the door to many new research questions, for instance: their algorithmic calculation, their interpretation, their approximation in high dimensional problems (since they are very costly to evaluate).

Other new perspectives concern the study of the Shapley value in different contexts, including: the sensitivity analysis of extremes and the sensitivity analysis in machine learning applications. In particular, one future extension are the Shapley effects for global sensitivity analysis of extremes. In such case, we can consider as value function

$$\nu(u) = \underline{\tau}_u^{(p)} = \int \cdots \int \prod_{k=1}^{p} f(\mathbf{x}_u : \mathbf{z}_{-u}^{(k)}) d\mathbf{x} \prod_{k=1}^{p} d\mathbf{z}^{(k)} - \mu^p \tag{6.1}$$

where where $\mathbf{z}^{(1)}, ..., \mathbf{z}^{(p)} \in [0,1]^d$, $p \geq 2$ and μ is the mean of f. The index $\underline{\tau}_u^{(p)}$ has been introduced by Owen et al. [2014] and it is called higher order Sobol' index. These indices are related to the decomposition of the skewness of the model output. The interpretation of $\underline{\tau}_i^{(3)} > 0$ is that the i-th model input is responsible for making the output reach its maximum, while $\underline{\tau}_i^{(3)} < 0$ its minimum. Thus, higher order Shapley effects can be

defined to gain such information in presence of dependent inputs.

The Shapley value is also used in the machine learning literature for determining feature importance. In particular, we have come up with the following general framework:

- $\nu(z) = f(x_z : \mathbf{x}^0_{-z}) - f(\mathbf{x}^0)$ (finite-change SA): sensitivity analysis of neural networks in the DeepLIFT explanation model [Shrikumar et al., 2017, Lundberg and Lee, 2017].

- $\nu(z) = \mathbb{E}\left[f(x_z : \mathbf{x}^0_{-z})\right] - \mathbb{E}\left[f(\mathbf{x}^0)\right]$ (partial dependence function): sensitivity analysis for predictive learning based on rules [Hooker, 2004, Friedman and Popescu, 2008, Goldstein et al., 2015, Guidotti et al., 2018]

- $\nu(z) = \mathbb{E}\left[f(\mathbf{x})|x_z\right] - \mathbb{E}\left[f(\mathbf{x})\right]$ (nonparametric regression curve): sensitivity analysis in predictive models [Štrumbelj and Kononenko, 2011, 2014] and in the SHAP method [Lundberg and Lee, 2017, Aas et al., 2019]

- $\nu(z) = \mathbb{V}\left[\mathbb{E}\left[f(\mathbf{x})|x_z\right]\right]$ (conditional variance): sensitivity analysis of complex models considered by Owen [2014].

This shows that sensitivity analysis using the Shapley value in the machine learning literature is conduced at different scales and there is the need of integrating these approaches in a unique framework.
Both ongoing works on Shapley values are in collaboration with Elmar Plischke.

Bibliography

Kjersti Aas, Martin Jullum, and Anders Løland. Explaining individual predictions when features are dependent: More accurate approximations to Shapley values. *arXiv e-prints*, page arXiv:1903.10464, 2019.

Fashul M Alam, Ken R McNaught, and Trevor J Ringrose. Using Morris' randomized OAT design as a factor screening method for developing simulation metamodels. In *Proceedings of the 2004 Winter Simulation Conference, 2004.*, volume 1, pages 1–957, 2004. ISBN VO - 1. doi: 10.1109/WSC.2004.1371413.

Bruce E Ankenman, Russell C H Cheng, and Susan M Lewis. Screening for Dispersion Effects by Sequential Bifurcation. *ACM Trans. Model. Comput. Simul.*, 25(1):2:1—-2:27, 2014.

G. E. B. Archer, A. Saltelli, and I. M. Sobol'. Sensitivity measures, ANOVA-like techniques and the use of bootstrap. *Journal of Statistical Computation and Simulation*, 58(2):99–120, 1997.

Fred D Baker and Rolf E Bargmann. Orthogonal Central Composite Designs of the Third Order in the Evaluation of Sensitivity and Plant Growth Simulation Models. *Journal of the American Statistical Association*, 80(391):574–579, 1985.

Daniel Barry. Testing for additivity of a regression function. *The Annals of Statistics*, 21(1):235–254, 1993.

M. S. Bartlett. Contingency tables interactions. *Supplement to the Journal of the Royal Statistical Society*, 2(2):248–252, 1935.

William Becker and Andrea Saltelli. Design for Sensitivity Analysis. In Angela Dean, Max Morris, John Stufken, and Derek Bingham, editors, *Handbook of Design and Analysis of Experiments*. Chapman and Hall/CRC, 2015.

Nazih Benoumechiara and Kevin Elie-Dit-Cosaque. Shapley effects for sensitivity analysis with dependent inputs: bootstrap and kriging-based algorithms. *ESAIM: ProcS*, 65:266–293, 2019.

Amy Berrington de González and David R. Cox. Interpretation of interaction: a review. *The Annals of Applied Statistics*, 1(2):371–385, 2007.

Bert Bettonvil and Jack P C Kleijnen. Searching for important factors in simulation models with many factors: Sequential bifurcation. *European Journal of Operational Research*, 96(1):180–194, 1997.

V. P. Bhapkar and Gary G. Koch. Hypotheses of 'no interaction' in multidimensional contingency tables. *Technometrics*, 10:107–123, 1968.

Jacob Bien, Jonathan Taylor, and Robert Tibshirani. A lasso for hierarchical interactions. *The Annals of Statistics*, 41(3):1111—-1141, 2013.

Jacob Bien, Noah Simon, and Robert Tibshirani. Convex hierarchical testing of interactions. *The Annals of Applied Statistics*, 9(1):27–42, 2015.

Mickaël Binois, Jiangeng Huang, Robert B Gramacy, and Mike Ludkovski. Replication or Exploration? Sequential Design for Stochastic Simulation Experiments. *Technometrics*, 61(1):7–23, 2019.

William J. Blot and Nicholas E. Day. Synergism and interaction: are they equivalent? *American Journal of Epidemiology*, 110(1):99–100, 1979.

M. Bloznelis and F. Götze. Orthogonal decomposition of finite population statistics and its applications to distributional asymptotics. *The Annals of Statistics*, 29(3):899–917, 2001.

E Borgonovo and L Peccati. Moment calculations for piecewise-defined functions: an application to stochastic optimization with coherent risk measures. *Annals of Operations Research*, 176(1):235–258, 2010.

E. Borgonovo and C. L. Smith. A Study of Interactions in the Risk Assessment of Complex Engineering Systems: An Application to Space PSA. *Operations Research*, 59(6):1461–1476, 2011.

Emanuele Borgonovo. Sensitivity analysis with finite changes: An application to modified EOQ models. *European Journal of Operational Research*, 200(1):127–138, 2010.

Emanuele Borgonovo and Elmar Plischke. Sensitivity analysis: A review of recent advances. *European Journal of Operational Research*, 248(3): 869–887, 2016.

Emanuele Borgonovo, Max D Morris, and Elmar Plischke. Functional ANOVA with Multiple Distributions: Implications for the Sensitivity Analysis of Computer Experiments. *SIAM/ASA Journal on Uncertainty Quantification*, 6(1):397–427, 2018.

Alexis Boukouvalas, John Paul Gosling, and Hugo Maruri-Aguilar. An Efficient Screening Method for Computer Experiments. *Technometrics*, 56(4):422–431, 2014.

G. E. P. Box and J. S. Hunter. The 2k-p Fractional Factorial Designs Part I. *Technometrics*, 3(3):311–351, 1961.

George E. P. Box, J. Stuart Hunter, and William G. Hunter. *Statistics for Experimenters: Design, Innovation, and Discovery.* Wiley-Interscience, New York, 2nd edition, 2005.

Baptiste Broto, François Bachoc, Marine Depecker, and Jean-Marc Martinez. Sensitivity indices for independent groups of variables. *Mathematics and Computers in Simulation*, 163(C):19–31, 2019.

Russel E. Caflisch, William Morokoff, and Art B. Owen. Valuation of mortgage-backed securities using Brownian bridges to reduce effective dimension. *Journal of Computational Finance*, 1(1):27–46, 1997.

N Cai, Y Song, and S Kou. A general framework for pricing Asian options under Markov processes. *Operations Research*, 63(3):540–554, 2015.

Francesca Campolongo, Jessica Cariboni, and Andrea Saltelli. An effective screening design for sensitivity analysis of large models. *Environmental Modelling & Software*, 22(10):1509–1518, 2007.

Francesca Campolongo, Andrea Saltelli, and Jessica Cariboni. From screening to quantitative sensitivity analysis. A unified approach. *Computer Physics Communications*, 182(4):978–988, 2011.

Campolongo F, Jack P. C. Kleijnen, and Terry Andres. Screening Methods. In A Saltelli, K Chan, and E Scott, editors, *Sensitivity Analysis*. Wiley, 2000.

Emmanuel Candes and Terence Tao. The Dantzig selector: Statistical estimation when p is much larger than n. *Ann. Statist.*, 35(6):2313–2351, 2007.

Karen Chan, Andrea Saltelli, and Stefano Tarantola. Winding Stairs: a sampling tool to compute sensitivity indices. *Statistics and Computing*, 10(3):187–196, 2000.

Gaelle Chastaing, Fabrice Gamboa, and Clémentine Prieur. Generalized Hoeffding-Sobol decomposition for dependent variables - application to sensitivity analysis. *Electronic Journal of Statistics*, 6:2420–2448, 2012.

Hao Chen, Jason L Loeppky, Jerome Sacks, and William J Welch. Analysis Methods for Computer Experiments: How to Assess and What Counts? *Statistical Science*, 31(1):40–60, 2016.

Zehua Chen. Interaction spline models and their convergence rates. *The Annals of Statistics*, 19(4):1855–1868, 1991.

Ching-Shui Cheng. *Theory of Factorial Design: Single- and Multi-Stratum Experiments*. CRC Press, Boca Raton, 2014.

R C H Cheng. Searching For Important Factors: Sequential Bifurcation Under Uncertainty. In *Winter Simulation Conference Proceedings,*, pages 275–280, 1997.

Stephen E Chick. Input Distribution Selection for Simulation Experiments: Accounting for Input Uncertainty. *Operations Research*, 49(5): 744–758, 2001.

Hugh Chipman. Bayesian variable selection with related predictors. *Canadian Journal of Statistics*, 24(1):17–36, 1996.

Nam Hee Choi, William Li, and Ji Zhu. Variable selection with the strong heredity constraint and its oracle property. *Journal of the American Statistical Association*, 105(489):354–364, 2010.

David G. Clayton. Prediction and interaction in complex disease genetics: experience in type 1 diabetes. *PLOS Genetics*, 5(7):1–6, 2009.

Joel E. Coen. Estimation and interaction in a censored 2x2x2 contingency table. *Biometrics*, 27(2):379–386, 1971.

R W Conway. Some Tactical Problems in Digital Simulation. *Management Science*, 10(1):47–61, 1963.

Sarah C. Cotter. A Screening Design for Factorial Experiments with Interactions. *Biometrika*, 66(2):317–320, 1979.

David R. Cox. Interaction. *International Statistical Review*, 52(1):1–24, 1984.

David R. Cox and Elizabeth Lauh. A note on the graphical analysis of multidimensional contingency tables. *Technometrics*, 9(3):481–488, 1967.

David R. Cox and Nancy Reid. *The Theory of the Design of Experiments*. Chapman and Hall/CRC, Boca Raton, 2000.

A. D'Ambra, P. Amenta, and A. Crisci. Decomposition of the main effects and interaction term by using orthogonal polynomials in multiple non symmetrical correspondence analysis. *Communications in Statistics - Theory and Methods*, 46(20):10179–10188, 2017.

Cuthbert Daniel. One-at-a-Time Plans. *Journal of the American Statistical Association*, 68(342):353–360, 1973.

Cuthbert Daniel. Patterns in residuals in the two-way layout. *Technometrics*, 20(4):385–395, 1978.

John N. Darroch. Multiplicative and additive interaction in contingency tables. *Biometrika*, 61(2):207–214, 1974.

John N. Darroch and T. P. Speed. Additive and multiplicative models and interactions. *The Annals of Statistics*, 11(3):724–738, 1983.

John N. Darroch, S. L. Lauritzen, and T. P. Speed. Markov fields and log-linear interaction models for contingency tables. *The Annals of Statistics*, 8(3):522–539, 1980.

P. Laurie Davies. Interactions in the analysis of variance. *Journal of the American Statistical Association*, 107(50):1502–1509, 2012.

P. Laurie Davies and Wolfgang Terbeck. Interactions and outliers in the two-way analysis of variance. *The Annals of Statistics*, 26(4):1279–1305, 1998.

Davide Di Cecco. A class of models for multiple binary sequences under the hypothesis of Markov exchangeability. *Electronic Journal of Statistics*, 3:1113–1132, 2009.

David Edwards and Svend Kreiner. The analysis of contingency tables by graphical models. *Biometrika*, 70(3):553–565, 1983.

Bradley Efron and Charles Stein. The jackknife estimate of variance. *The Annals of Statistics*, 9(3):586–596, 1981.

Ted G Eschenbach. Spiderplots versus Tornado Diagrams for Sensitivity Analysis. *INFORMS Journal on Applied Analytics*, 22(6):40–46, 1992.

Graciela Estévez-Pérez and José A. Vilar. Functional ANOVA starting from discrete data: an application to air quality data. *Environmental and Ecological Statistics*, 20(3):495–517, 2013.

Brian S. Everitt. *The Analysis of Contingency Tables*. Chapman & Hall/CRC, 2nd edition, 1992.

Brian S. Everitt and M. R. Smith. Interactions in contingency tables: a brief discussion of alternative definitions. *Psychological Medicine*, 9: 581–583, 1979.

Jianqing Fan and Rui Song. Sure independence screening in generalized linear models with NP-dimensionality. *The Annals of Statistics*, 38(6): 3567–3604, 2010.

Jianqing Fan, Richard Samworth, and Yichao Wu. Ultrahigh dimensional feature selection: beyond the linear model. *Journal of Machine Learning Research*, 10:2013–2038, 2009.

Yingying Fan, Yinfei Kong, Daoji Li, and Zemin Zheng. Innovated interaction screening for high-dimensional nonlinear classification. *The Annals of Statistics*, 43(3):1243–1272, 2015.

D. J. Finney. Main effects and interactions. *Journal of the American Statistical Association*, 43(244):566–571, 1948.

Ronald A. Fisher and Winifred A. Mackenzie. The manurial response of different potato varieties. *The Journal of Agricultural Science*, 13(3): 311–320, 1923.

Edward B. Fowlkes, Anne E. Freeny, and James M. Landwehr. Evaluating logistic models for large contingency tables. *Journal of the American Statistical Association*, 83(403):611–622, 1988.

Jerome H. Friedman and Bogdan E. Popescu. Predictive learning via rule ensembles. *The Annals of Applied Statistics*, 2(3):916–954, 2008.

Arthur Fries and William G. Hunter. Minimum aberration 2k-p designs. *Technometrics*, 22(4):601–608, 1980.

J Fruth, O Roustant, and S Kuhnt. Total interaction index: A variance-based sensitivity index for second-order interaction screening. *Journal of Statistical Planning and Inference*, 147:212–223, 2014.

F Gamboa, A Janon, T Klein, A Lagnoux, and C Prieur. Statistical inference for Sobol pick-freeze Monte Carlo method. *Statistics*, 50(4): 881–902, 2016.

John J. Gart. Interaction tests for 2xsxt contingency tables. *Biometrika*, 59(2):309–316, 1972.

Qiao Ge and Monica Menendez. Extending Morris method for qualitative global sensitivity analysis of models with dependent inputs. *Reliability Engineering & System Safety*, 162:28–39, 2017.

Andrew Gelman. Analysis of variance—why it is more important than ever. *The Annals of Statistics*, 33(1):1–53, 2005.

Paul Glasserman and Yashan Wang. Leadtime-Inventory Trade-Offs in Assemble-to-Order Systems. *Operations Research*, 46(6):858–871, 1998.

Alex Goldstein, Adam Kapelner, Justin Bleich, and Emil Pitkin. Peeking Inside the Black Box: Visualizing Statistical Learning With Plots of Individual Conditional Expectation. *Journal of Computational and Graphical Statistics*, 24(1):44–65, 2015.

Leo A. Goodman. Simple methods for analyzing three-factor interaction in contingency tables. *Journal of the American Statistical Association*, 59(306):319–352, 1964.

Leo A. Goodman. The multivariate analysis of qualitative data: interactions among multiple classifications. *Journal of the American Statistical Association*, 65(329):226–256, 1970.

Michel Grabisch. Alternative Representations of Discrete Fuzzy Measures for Decision Making. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 05(05):587–607, 1997.

Michel Grabisch. Capacities and Games on Lattices: A Survey of Results. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 14(04):371–392, 2006.

Michel Grabisch and Christophe Labreuche. How to improve acts: an alternative representation of the importance of criteria in MCDM. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 09(02):145–157, 2001.

Michel Grabisch and Marc Roubens. An axiomatic approach to the concept of interaction among players in cooperative games. *International Journal of Game Theory*, 28(4):547–565, 1999.

Sander Greenland. Tests for interaction in epidemiologic studies: a review and a study of power. *Statistics in Medicine*, 2:243–251, 1983.

Sander Greenland, Timothy L. Lash, and Kenneth J. Rothman. Concepts of interaction. In Kenneth J. Rothman, Sander Greenland, and Timothy L. Lash, editors, *Modern Epidemiology*, pages 71–86. Lippincott Williams & Wilkins, Philadelphia, 3rd edition, 2008.

Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Fosca Giannotti, and Dino Pedreschi. A Survey of Methods for Explaining Black Box Models. *ACM Comput. Surv.*, 51(5):93:1—-93:42, 2018.

Jaroslav Hajek. Asymptotic normality of simple linear rank statistics under alternatives. *The Annals of Mathematical Statistics*, 39(2):325–346, 1968.

Michael Hamada and Chien-Fu Jeff Wu. Analysis of designed experiments with complex aliasing. 1991.

Ning Hao and Hao Helen Zhang. Interaction screening for ultrahigh-dimensional data. *Journal of the American Statistical Association*, 109 (507):1285–1301, 2014.

Ning Hao and Hao Helen Zhang. A Note on high-dimensional linear regression with interactions. *The American Statistician*, 71(4):291–297, 2017.

Ning Hao, Yang Feng, and Hao Helen Zhang. Model selection for high-dimensional quadratic regression via regularization. *Journal of the American Statistical Association*, 2018.

Hisayuki Hara, Tomonari Sei, and Akimichi Takemura. Hierarchical subspace models for contingency tables. *Journal of Multivariate Analysis*, 103(1):19–34, 2012.

Asad Haris, Daniela Witten, and Noah Simon. Convex modeling of interactions with strong heredity. *Journal of Computational and Graphical Statistics*, 25(4):981–1004, 2016.

Trevor Hastie and Robert Tibshirani. *Generalized Additive Models*, volume 1. Chapman and Hall/CRC, Boca Raton, 1990.

Zihuai He, Min Zhang, Seunggeun Lee, Jennifer A Smith, Sharon L R Kardia, V Diez Roux, and Bhramar Mukherjee. Set-based tests for the gene–environment interaction in longitudinal studies. *Journal of the American Statistical Association*, 112(519):966–978, 2017.

Andy Hector, Stefanie von Felten, and Bernhard Schmid. Analysis of variance with unbalanced data: an update for ecology and evolution. *Journal of Animal Ecology*, 79(2):308–316, 2010.

David G. Herr. On the history of ANOVA in unbalanced, factorial designs: the first 30 years. *The American Statistician*, 40(4):265–270, 1986.

I Herrera. Theory of differential equations in discontinuous piecewise-defined functions. *Numerical Methods for Partial Differential Equations*, 23(3):597–639, 2007.

D J Hodgson and S Townley. Linking Management Changes to Population Dynamic Responses: The Transfer Function of a Projection Matrix Perturbation. *Journal of Applied Ecology*, 41(6):1155–1161, 2004.

Wassily Hoeffding. A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, 19(3):293–325, 1948.

Søren Højsgaard. Split models for contingency tables. *Computational Statistics & Data Analysis*, 42(4):621–645, 2003.

Søren Højsgaard. Statistical inference in context specific interaction models for contingency tables. *Scandinavian Journal of Statistics*, 31(1): 143–158, 2004.

Toshimitsu Homma and Andrea Saltelli. Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety*, 52(1):1–17, 1996.

L Jeff Hong and Barry L Nelson. Discrete Optimization via Simulation Using COMPASS. *Operations Research*, 54(1):115–129, 2006.

Giles Hooker. Discovering Additive Structure in Black Box Functions. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '04, pages 575–580, New York, NY, USA, 2004. ACM.

Giles Hooker. Generalized functional ANOVA diagnostics for high-dimensional functions of dependent variables. *Journal of Computational and Graphical Statistics*, 16(3):709–732, 2007.

Ronald A Howard. Decision Analysis: Practice and Promise. *Management Science*, 34(6):679–695, 1988.

Jianhua Z. Huang. Functional ANOVA models for generalized regression. *Journal of Multivariate Analysis*, 67(1):49–71, 1998a.

Jianhua Z. Huang. Projection estimation in multiple regression with application to functional ANOVA models. *The Annals of Statistics*, 26(1): 242–272, 1998b.

Kosuke Imai and Marc Ratkovic. Estimating treatment effect heterogeneity in randomized program evaluation. *The Annals of Applied Statistics*, 7(1):443–470, 2013.

Bertrand Iooss and Clémentine Prieur. Shapley effects for sensitivity analysis with dependent inputs: comparisons with Sobol' indices, numerical estimation and applications. *International Journal of Uncertainty Quantification*, In press, 2019.

Edward H. Ip, Yuchung J. Wang, and Yeong-nan Yeh. Structural decompositions of multivariate distributions with applications in moment and cumulant. *Journal of Multivariate Analysis*, 89(1):119–134, 2004.

James Jaccard. *Interaction Effects in Factorial Analysis of Variance*. Sage Publications, 1998.

J E Jacoby and S Harrison. Multi-variable experimentation and simulation models. *Naval Research Logistics Quarterly*, 9(2):121–136, 1962.

Alexandre Janon, Thierry Klein, Agnès Lagnoux, Maëlle Nodet, and Clémentine Prieur. Asymptotic normality and efficiency of two Sobol index estimators. *ESAIM: Probability and Statistics*, 18:342–364, 2014.

Lluís Antoni Jiménez Rugama and Laurent Gilquin. Reliable error estimation for Sobol' indices. *Statistics and Computing*, 28(4):725–738, 2018.

Harry Joe. Multivariate dependence measures and data analysis. *Computational Statistics and Data Analysis*, 16(3):279–297, 1993.

Samuel Karlin and Yosef Rinott. Applications of Anova type decompositions for comparisons of conditional variance statistics including jackknife estimates. *The Annals of Statistics*, 10(2):485–501, 1982.

Alon Keinan, Claus C Hilgetag, Isaac Meilijson, and Eytan Ruppin. Causal localization of neural function: the Shapley value method. *Neurocomputing*, 58-60:215–222, 2004.

Oscar Kempthorne. Fixed and mixed models in the analysis of variance. *Biometrics*, 31(2):473–486, 1975.

Hyoung-Moon Kim, Bani K Mallick, and C C Holmes. Analyzing Nonstationary Spatial Data Using Piecewise Gaussian Processes. *Journal of the American Statistical Association*, 100(470):653–668, 2005.

Alisa Kirichenko and Harry van Zanten. Estimating a smooth function on a large graph by Bayesian Laplacian regularisation. *Electron. J. Statist.*, 11(1):891–915, 2017.

Jack P. C. Kleijnen. *Design and Analysis of Simulation Experiments*. Springer International Publishing, 2nd edition, 2015.

Jack P C Kleijnen. Design and Analysis of Simulation Experiments: Tutorial BT - Advances in Modeling and Simulation: Seminal Research from 50 Years of Winter Simulation Conferences. pages 135–158. Springer International Publishing, Cham, 2017.

Jack P C Kleijnen, Bert Bettonvil, and Fredrik Persson. Screening for the Important Factors in Large Discrete-Event Simulation Models: Sequential Bifurcation and Its Applications. In Angela Dean and Susan Lewis, editors, *Screening: Methods for Experimentation in Industry, Drug Discovery, and Genetics*, pages 287–307. Springer New York, New York, NY, 2006.

J. Koehler and Art B. Owen. Computer Experiments. In S. Ghosh and C. R. Rao, editors, *Handbook of Statistics, Vol. 13*, pages 261–308. Elsevier, 1996.

Yinfei Kong, Daoji Li, Yingying Fan, and Jinchi Lv. Interaction pursuit in high-dimensional multi-response regression via distance correlation. *The Annals of Statistics*, 45(2):897—-922, 2017.

Pieter M. Kroonenberg and Carolyn J. Anderson. Additive and multiplicative models for three-way contingency tables: Darroch (1974) revisited.

In M.J. Greenacre and J. Blasius, editors, *Multiple correspondence analysis and related methods*, pages 455–486. CRC press LLC / Chapman and Hall, Boca Raton, 2006.

Sergei Kucherenko, Stefano Tarantola, and Paola Annoni. Estimation of global sensitivity indices for models with dependent variables. *Computer Physics Communications*, 183(4):937–946, 2012.

Henry O. Lancaster. Complex contingency tables treated by the partition of chi-square. *Journal of the Royal Statistical Society. Series B*, 13(2): 242–249, 1951.

Henry O. Lancaster. *The Chi-squared Distribution*. John Wiley & Sons, New York, 1969.

Henry O. Lancaster. The multiplicative definition of interaction. *The Australian Journal of Statistics*, 13(1):36–44, 1971.

Johannes A. Landsheer and Godfried van den Wittenboer. Unbalanced 2 x 2 factorial designs and the interaction effect: a troublesome combination. *Plos One*, 10(3), 2015. doi: 10.1371/journal.pone.0121412.

Øyvind Langsrud. ANOVA for unbalanced data: use type II instead of type III sums of squares. *Statistics and Computing*, 13(2):163–167, 2003.

Steffen L. Lauritzen. Interaction models. In Sandrine Dudoit, editor, *Selected Works of Terry Speed*, pages 91–94. Springer-Verlag, New York, 2012.

Ryan Lekivetz and Boxin Tang. Robust designs through partially clear two-factor interactions. *Biometrika*, 98(3):733–739, 2011.

B. N. Lewis. On the analysis of interaction in multi-dimensional contingency tables. *Journal of the Royal Statistical Society. Series A*, 125(1): 88–117, 1962.

S.M. Lewis and A.M. Dean. Detection of interaction in experiments on large numbers of factors. *Journal of the Royal Statistical Society. Series B*, 63(4):633–672, 2001.

J. D. Lewsey, W. P. Gardiner, and G. Gettinby. A study of type II and type III power for testing hypotheses from unbalanced factorial designs. *Communications in Statistics - Simulation and Computation*, 30(3):597–609, 2001.

Genyuan Li and Herschel Rabitz. General formulation of HDMR component functions with independent and correlated variables. *Journal of Mathematical Chemistry*, 50(1):99–130, 2012.

Genyuan Li and Herschel Rabitz. Analytical HDMR formulas for functions expressed as quadratic polynomials with a multivariate normal distribution. *Journal of Mathematical Chemistry*, 52(8):2052–2073, 2014.

Genyuan Li and Herschel Rabitz. Relationship between sensitivity indices defined by variance- and covariance-based methods. *Reliability Engineering & System Safety*, 167:136–157, 2017.

Genyuan Li, Carey Rosenthal, and Herschel Rabitz. High dimensional model representations. *The Journal of Physical Chemistry A*, 105(33): 7765–7777, 2001.

Genyuan Li, Herschel Rabitz, Paul E Yelvington, Oluwayemisi O Oluwole, Fred Bacon, Charles E Kolb, and Jacqueline Schoendorf. Global Sensitivity Analysis for Systems with Independent and/or Correlated Inputs. *The Journal of Physical Chemistry A*, 114(19):6022–6032, 2010.

Genyuan Li, Roberto Rey-de Castro, and Herschel Rabitz. D-MORPH regression for modeling with fewer unknown parameters than observation data. *Journal of Mathematical Chemistry*, 50(7):1747–1764, 2012.

Michael Lim and Trevor Hastie. Learning interactions via hierarchical group-lasso regularization. *Journal of Computational and Graphical Statistics*, 24(3):627–654, 2015.

Yi Lin. Tensor product space ANOVA models. *The Annals of Statistics*, 28(3):734–755, 2000.

Ruixue Liu and Art B. Owen. Estimating mean dimensionality of analysis of variance decompositions. *Journal of the American Statistical Association*, 101(474):712–721, 2006.

Xu Liu, Yuehua Cui, and Runze Li. Partial linear varying multi-index coefficient model for integrative gene-environment interactions. *Statistica Sinica*, 26(3):1037–1060, 2016.

Xu Liu, Bin Gao, and Yuehua Cui. Generalized partial linear varying multi-index coefficient model for gene-environment interactions. *Statistical Applications in Genetics and Molecular Biology*, 16(1):59–74, 2017.

Scott M Lundberg and Su-In Lee. A Unified Approach to Interpreting Model Predictions. In I Guyon, U V Luxburg, S Bengio, H Wallach, R Fergus, S Vishwanathan, and R Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 4765–4774. Curran Associates, Inc., 2017.

Shujie Ma and Peter X.-K. Song. Varying index coefficient models. *Journal of the American Statistical Association*, 110(509):341–356, 2015.

Shujie Ma, Raymond J. Carroll, Hua Liang, and Shizhong Xu. Estimation and inference in generalized additive coefficient models for nonlinear interactions with high-dimensional covariates. *The Annals of Statistics*, 43(5):2102—-2131, 2015.

Thierry A. Mara, Stefano Tarantola, and Paola Annoni. Non-parametric methods for global sensitivity analysis of model output with dependent inputs. *Environmental Modelling & Software*, 72:173–183, 2015.

Peter McCullagh. Invariance and factorial models. *Journal of the Royal Statistical Society. Series B*, 62(2):209–256, 2000.

Max D. Morris. Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics*, 33(2):161–174, 1991.

Max D. Morris and Leslie M. Moore. Design of Computer Experiments: Introduction and Background. In Angela Dean, Max Morris, John Stufken,

and Derek Bingham, editors, *Handbook of Design and Analysis of Experiments*. Chapman & Hall/CRC, 2015.

Thomas Muehlenstaedt, Olivier Roustant, Laurent Carraro, and Sonja Kuhnt. Data-driven Kriging models based on FANOVA-decomposition. *Statistics and Computing*, 22(3):723–738, 2012.

B. P. Murphy. A note on current methods of partitioning the contingency chi-squared statistic. *Journal of Statistical Computation and Simulation*, 6(3-4):157–167, 1978.

Raymond H. Myers, Douglas C. Montgomery, and Christine M. Anderson-Cook. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. John Wiley & Sons, New York, 4th edition, 2016.

John A. Nelder. Log linear models for contingency tables: a generalization of classical least squares. *Journal of the Royal Statistical Society. Series C*, 23(3):323–329, 1974.

John A. Nelder. A reformulation of linear models. *Journal of the Royal Statistical Society. Series A*, 140(1):48–77, 1977.

John A. Nelder and Peter W. Lane. The computer analysis of factorial experiments: in memoriam—Frank Yates. *The American Statistician*, 49(4):382–385, 1995.

Barry Nelson. *Foundations and Methods of Stochastic Simulation*. Springer US, 2013.

Barry L Nelson. 50th Anniversary Article: Stochastic Simulation Research in Management Science. *Management Science*, 50(7):855–868, 2004.

Tie-Hua Ng. The impact of a preliminary test for interaction in a 2x2 factorial trial. *Communications in Statistics - Theory and Methods*, 23 (2):435–450, 1994.

Wei Ning and Hyune-Ju Kim. Residual pattern based test for interactions in two-way ANOVA. *Biometrical Journal*, 50(3):431–445, 2008.

M. E. O'Neill. A comparison of the additive and multiplicative definitions of second-order in 2x2x2 contingency tables. *Journal of Statistical Computation and Simulation*, 15(1):33–50, 1982a.

M. E. O'Neill. On tests of homogeneity of factorially-structured populations: an additive approach. *Biometrika*, 69(1):167–174, 1982b.

John E. Overall and Douglas K. Spiegel. Concerning least squares analysis of experimental data. *Psychological Bulletin*, 72(5):311–322, 1969.

Art B. Owen. The dimension distribution and quadrature test functions. *Statistica Sinica*, 13(1):1–17, 2003.

Art B. Owen. Variance components and generalized Sobol' indices. *SIAM/ASA Journal on Uncertainty Quantification*, 1(1):19–41, 2013.

Art B. Owen. Sobol' Indices and Shapley Value. *SIAM/ASA Journal on Uncertainty Quantification*, 2(1):245–251, 2014.

Art B. Owen and Clementine Prieur. On Shapley Value for Measuring Importance of Dependent Inputs. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):986–1002, 2017.

Art B Owen, Josef Dick, and Su Chen. Higher order Sobol' indices. *Information and Inference: A Journal of the IMA*, 3(1):59–81, 2014.

Guillermo Owen. Multilinear Extensions of Games. *Management Science*, 18(52):64–79, 1972.

K. M. Patel and D. G. Hoel. A nonparametric test for interaction in factorial experiments. *Journal of the American Statistical Association*, 68(343):615–620, 1973.

Kashinath D. Patil. Interaction test for three-dimensional contingency tables. *Journal of the American Statistical Association*, 69(345):164–168, 1974.

Giovanni Peccati. Hoeffding-Anova decompositions for symmetric statistics of exchangeable observations. *The Annals of Probability*, 32(3): 1796–1829, 2004.

Julio L. Peixoto. Hierarchical variable selection in polynomial regression models. *The American Statistician*, 41(4):311–313, 1987.

Patrick C Phillips. Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Reviews Genetics*, 9: 855–867, 2008.

Luc Pronzato and Werner G Müller. Design of computer experiments: space filling and beyond. *Statistics and Computing*, 22(3):681–701, 2012.

Xianggui Qu and C. F. Jeff Wu. One-factor-at-a-time designs of resolution V. *Journal of Statistical Planning and Inference*, 131(2):407–416, 2005.

Herschel Rabitz and Ömer F. Alis. General foundations of high - dimensional model representations. *Journal of Mathematical Chemistry*, 25 (2-3):197–233, 1999.

Herschel Rabitz, Ömer F. Alis, Jeffrey Shorter, and Kyurhee Shim. Efficient input—output model representations. *Computer Physics Communications*, 117(1-2):11–20, 1999.

Peter Radchenko and Gareth M. James. Variable selection using adaptive nonlinear interaction structures in high dimensions. *Journal of the American Statistical Association*, 105(492):1541–1553, 2010.

Sharif Rahman. A generalized ANOVA dimensional decomposition for dependent probability measures. *SIAM/ASA Journal of Uncertainty Quantification*, 2(1):670–697, 2014.

O Rakovec, M C Hill, M P Clark, A H Weerts, A J Teuling, and R Uijlenhoet. Distributed Evaluation of Local Sensitivity Analysis (DELSA), with application to hydrologic models. *Water Resources Research*, 50 (1):409–426, 2014.

Jhan Rodríguez and András Bárdossy. Multivariate interactions modeling through their manifestations: low dimensional model building via the Cumulant Generating Function. *arXiv preprint*, 2014. URL `http://arxiv.org/abs/1406.2815`.

Werner Römisch. ANOVA Decomposition of Convex Piecewise Linear Functions BT - Monte Carlo and Quasi-Monte Carlo Methods 2012. pages 581–596, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

Kenneth J. Rothman, Sander Greenland, and Alexander M. Walker. Concepts of interaction. *American Journal of Epidemiology*, 112(4):467–470, 1980.

O Roustant, J Fruth, B Iooss, and S Kuhnt. Crossed-derivative based sensitivity measures for interaction screening. *Mathematics and Computers in Simulation*, 105:105–118, 2014.

Olivier Roustant, Esperan Padonou, Yves Deville, Alo\"\is Clément, Guillaume Perrin, Jean Giorla, and Henry Wynn. Group kernels for Gaussian process metamodels with categorical inputs. *arXiv e-prints*, page arXiv:1802.02368, 2018.

Eric J. Routman and James M. Cheverud. Gene effects on a quantitative trait: two-locus epistatic effects measured at microsatellite markers and at estimated QTL. *Evolution*, 51(5):1654–1662, 1997.

S. N. Roy and Marvin A. Kastenbaum. On the hyphothesis of no "interaction" in a multi-way contingency table. *The Annals of Mathematical Statistics*, 27(3):749–757, 1956.

Jerome Sacks, William J. Welch, Toby J. Mitchell, and Henry P. Wynn. Design and analysis of computer experiments. *Statistical Science*, 4(4): 409–423, 1989.

Stephan R. Sain, Doug Nychka, and Linda Mearns. Functional ANOVA and regional climate experiments: a statistical analysis of dynamic downscaling. *Environmetrics*, 22(6):700–711, 2011.

A Saltelli, T H Andres, and T Homma. Sensitivity analysis of model output. Performance of the iterated fractional factorial design method. *Computational Statistics & Data Analysis*, 20(4):387–407, 1995.

Andrea Saltelli. Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, 145(2):280–297, 2002.

Andrea Saltelli and Stefano Tarantola. On the Relative Importance of Input Factors in Mathematical Models: Safety Assessment for Nuclear Waste Disposal. *Journal of the American Statistical Association*, 97 (459):702–709, 2002.

Andrea Saltelli, Stefano Tarantola, and Karen Chan. A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics*, 41(1):39–56, 1999.

Andrea Saltelli, Paola Annoni, Ivano Azzini, Francesca Campolongo, Marco Ratto, and Stefano Tarantola. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2):259–270, 2010.

Thomas J. Santner, Brian J. Williams, and William I. Notz. *The Design and Analysis of Computer Experiments*. Springer-Verlag, New York, 2003.

Henry Scheffé. *The Analysis of Variance*. John Wiley & Sons, New York, 1959.

Henry Scheffé. A Note on Separation of Variables. *Technometrics*, 12(2): 388–393, 1970.

Matthias Schonlau and William J Welch. Screening the Input Variables to a Computer Model Via Analysis of Variance and Visualization. In Angela Dean and Susan Lewis, editors, *Screening: Methods for Experimentation in Industry, Drug Discovery, and Genetics*, pages 308–327. Springer New York, New York, NY, 2006.

Dino Sejdinovic, Arthur Gretton, and Wicher Bergsma. A kernel test for three-variable interactions. In *Advances in Neural Information Processing Systems*, pages 1124–1132. 2013.

Lloyd S. Shapley. A Value for n-person Games. In H. W. Kuhn and A. W. Tucker, editors, *Contributions to the Theory of Games*, pages 307–317. Princeton University Press, 1953.

Ruth G. Shaw and Thomas Mitchell-Olds. Anova for unbalanced data: an overview. *Ecology*, 74(6):1638–1645, 1993.

Yiyuan She, Zhifeng Wang, and He Jiang. Group regularized estimation under structural hierarchy. *Journal of the American Statistical Association*, 2016.

Wen Shi and Xi Chen. Controlled Morris Method: A New Distribution-free Sequential Testing Procedure for Factor Screening. In *Proceedings of the 2017 Winter Simulation Conference*, WSC '17, pages 142:1—-142:12, Piscataway, NJ, USA, 2017. IEEE Press.

Wen Shi and Xi Chen. Efficient budget allocation strategies for elementary effects method in stochastic simulation. *Naval Research Logistics (NRL)*, 65(3):218–241, 2018.

Wen Shi and Jack P C Kleijnen. Testing the assumptions of sequential bifurcation for factor screening. *Simulation Modelling Practice and Theory*, 81:85–99, 2018.

Wen Shi, Jennifer Shang, and Zhang Zhigang. Simulation screening and false discovery rate control for both main and interaction effects. In *2016 Winter Simulation Conference (WSC)*, pages 512–521, 2016.

Galit Shmueli. To explain or to predict? *Statistical Science*, 25(3):289–310, 2010.

Avanti Shrikumar, Peyton Greenside, and Anshul Kundaje. Learning Important Features Through Propagating Activation Differences. In

Proceedings of the 34th International Conference on Machine Learning, pages 1–9, Sydney, 2017. PMLR 70.

Jack Siemiatycki and Duncan C Thomas. Biological models and statistical interactions: an example from multistage carcinogenesis. *International Journal of Epidemiology*, 10(4):383–387, 1981.

Noah Simon and Robert Tibshirani. A permutation approach to testing interactions for binary response by comparing correlations between classes. *Journal of the American Statistical Association*, 110(512):1707–1716, 2015.

E. H. Simpson. The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society. Series B*, 13(2):238–241, 1951.

Carrie E. Smith and Robert Cribbie. Factorial ANOVA with unbalanced data: a fresh look at the types of sums of squares. *Journal of Data Science*, 12:385–404, 2014.

Ronald D. Snee. Nonadditivity in a two-way classification: is it interaction or nonhomogeneous variance? *Journal of the American Statistical Association*, 77(379):515–519, 1982.

Ilya M. Sobol'. *Multidimensional Quadrature Formulas and Haar Functions*. Nauka, Moscow, 1969.

Ilya M. Sobol'. Sensitivity estimates for nonlinear mathematical models. *Mathematical Modelling and Computational Experiments*, 1(4):407–414, 1993.

Eunhye Song, Barry L Nelson, and Jeremy Staum. Shapley Effects for Global Sensitivity Analysis: Theory and Computation. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):1060–1083, 2016.

Charles J. Stone. Additive regression and other nonparametric models. *The Annals of Statistics*, 13(2):689–705, 1985.

Charles J. Stone. The use of polynomial splines and their tensor products in multivariate function estimation. *The Annals of Statistics*, 22(1): 118–184, 1994.

Bernd Streitberg. Lancaster interactions revisited. *The Annals of Statistics*, 18(4):1878–1885, 1990.

Bernd Streitberg. Exploring interactions in high-dimensional tables: a bootstrap alternative to log-linear models. *The Annals of Statistics*, 27 (1):405–413, 1999.

Erik Štrumbelj and Igor Kononenko. A General Method for Visualizing and Explaining Black-Box Regression Models BT - Adaptive and Natural Computing Algorithms. pages 21–30, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

Erik Štrumbelj and Igor Kononenko. Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems*, 41(3):647–665, 2014.

Don X. Sun, Chien-Fu Jeff Wu, and Youyi Chen. Optimal blocking schemes for 2n and 2n—p designs. *Technometrics*, 39(3):298–307, 1997.

Akimichi Takemura. Tensor analysis of ANOVA decomposition. *Journal of the American Statistical Association*, 78(384):894–900, 1983.

Boxin Tang. Orthogonal arrays robust to nonnegligible two-factor interactions. *Biometrika*, 93(1):137–146, 2006.

Yu Tang and Hongquan Xu. Permuting regular fractional factorial designs for screening quantitative factors. *Biometrika*, 101(2):333–350, 2014.

Javier Tarrío-Saavedra, Salvador Naya, Mario Francisco-Fernández, and Jorge Artiaga, Ramón Lopez-Beceiro. Application of functional ANOVA to the study of thermal stability of micro–nano silica epoxy composites. *Chemometrics and Intelligent Laboratory Systems*, 105(1):114–124, 2011.

W. Douglas Thompson. Effect modification and the limits of biological inference from epidemiologic data. *Journal of Clinical Epidemiology*, 44 (3):221–232, 1991. ISSN 0895-4356. doi: 10.1016/0895-4356(91)90033-6.

Robert Tibshirani. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B*, 58(1):267–288, 1996.

J. Töwe, J. Bock, and G. Kundt. Interactions in contingency table analysis. *Biometrical Journal*, 27(1):17–24, 1985.

John W. Tukey. One degree of freedom for non-additivity. *Biometrics*, 5 (3):232–242, 1949.

John W. Tukey. Exploratory analysis of variance as providing examples of strategic choices. In Elvezio Morgenthaler, Stephan, Ronchetti and Werner A. Stahel, editors, *New Directions in Statistical Data Analysis and Robustness*. Birkhäuser, Basel, 1993.

S. Vajda. Relations between variously defined effects and interactions in analysis of variance. *The Annals of Mathematical Statistics*, 22(2):283–288, 1951.

A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 2000.

Tyler J. VanderWeele. On the distinction between interaction and effect modification. *Epidemiology*, 20(6):863–871, 2009.

Tyler J. VanderWeele. Epistatic interactions. *Statistical Applications in Genetics and Molecular Biology*, 9(1):1–22, 2010.

Tyler J. VanderWeele. *Explanation in Causal Inference: Methods for Mediation and Interaction*. Oxford University Press, 2015.

Tyler J. VanderWeele and Mirjam J. Knol. A tutorial on interaction. *Epidemiologic Methods*, 3(1):33–72, 2014.

Tyler J. VanderWeele and James M. Robins. Empirical and counterfactual conditions for sufficient cause interactions. *Biometrika*, 95(1):49–61, 2008.

Tyler J. VanderWeele, Alex R. Luedtke, Mark J. van der Laan, and Ronald C. Kessler. Selecting optimal subgroups for treatment using many covariates. 2018.

Stijn Vansteelandt, Tyler J VanderWeele, Eric J Tchetgen, and James M Robins. Multiply robust inference for statistical interactions. *Journal of the American Statistical Association*, 103(484):1693–1704, 2008.

Harvey M Wagner. Global Sensitivity Analysis. *Operations Research*, 43 (6):948–969, 1995.

Grace Wahba, Yuedong Wang, Chong Gu, Ronald Klein, and Barbara Klein. Smoothing spline ANOVA for exponential families, with application to the Wisconsin epidemiological study of diabetic retinopathy. *The Annals of Statistics*, 23(6):1865–1895, 1995.

S. D. Walter and T. R. Holford. Additive, multiplicative, and other models for disease risks. *American Journal of Epidemiology*, 108(5):341–346, 1978.

Hong Wan, Bruce E Ankenman, and Barry L Nelson. Controlled Sequential Bifurcation: A New Factor-Screening Method for Discrete-Event Simulation. *Operations Research*, 54(4):743–755, 2006.

Xiangyu Wang and Chenlei Leng. High dimensional ordinary least squares projection for screening variables. *Journal of the Royal Statistical Society. Series B*, 78(3):589–611, 2016.

Xiaogun Wang. On the effects of dimension reduction techniques on some high-dimensional problems in finance. *Operations Research*, 54(6):1063–1078, 2006.

Xiaoqun Wang and Ian H Sloan. Quasi-Monte Carlo methods in financial engineering: an equivalence principle and dimension reduction. *Operations Research*, 59(1):80–95, 2011.

Xiaoqun Wang and Ken Seng Tan. Pricing and Hedging with Discontinuous Functions: Quasi-Monte Carlo Methods and Dimension Reduction. *Management Science*, 59(2):376–389, 2013.

Y. X. Rachel Wang, Keni Jiang, Lewis J. Feldman, Peter J. Bickel, and Haiyan Huang. Inferring gene–gene interactions and functional modules using sparse canonical correlation analysis. *The Annals of Applied Statistics*, 9(1):300–323, 2015.

Yuchung J. Wang. Compatibility among marginal densities. *Biometrika*, 91(1):234–239, 2004.

E. J. Williams. The interpretation of interactions in factorial experiments. *Biometrika*, 39(1-2):65–81, 1952.

David C Woods and Susan M Lewis. Design of Experiments for Screening. In Roger Ghanem, David Higdon, and Houman Owhadi, editors, *Handbook of Uncertainty Quantification*, pages 1143–1185. Springer International Publishing, 2017.

Chien-Fu Jeff Wu. Post-Fisherian experimentation: from physical to virtual. *Journal of the American Statistical Association*, 110(510):612–620, 2015.

Chien-Fu Jeff Wu. A fresh look at effect aliasing and interactions: some new wine in old bottles. *Annals of the Institute of Statistical Mathematics*, 70:249–268, 2018.

Chien-Fu Jeff Wu and Youyi Chen. A graph-aided method for planning two-level experiments when certain interactions are important. *Technometrics*, 34(2):162–175, 1992.

Chien-Fu Jeff Wu and Michael Hamada. *Experiments: Planning, Analysis, and Optimization*. John Wiley & Sons, Hoboken, 2nd edition, 2009.

Chien-Fu Jeff Wu and Huaiqing Wu. Clear two-factor interactions and minimum aberration. *The Annals of Statistics*, 30(5):1496–1511, 2002.

Jing Wu, Bernie Devlin, Steven Ringquist, Massimo Trucco, and Kathryn Roeder. Screen and clean: a tool for identifying interactions in genome-wide association studies. *Genetic Epidemiology*, 34(3):275–285, 2010.

Tong Tong Wu, Yi Fang Chen, Trevor Hastie, Eric Sobel, and Kenneth Lange. Genome-wide association analysis by lasso penalized logistic regression. *Bioinformatics*, 25(6):714–721, 2009.

Yin Xia, Tianxi Cai, and T. Tony Cai. Testing differential networks with applications to the detection of gene-gene interactions. *Biometrika*, 102 (2):247–266, 2015.

Frank Yates. The analysis of multiple classifications with unequal numbers in the different classes. *Journal of the American Statistical Association*, 29(185):51–66, 1934.

Frank Yates. The Design and Analysis of Factorial Experiments. Technical report, Volume 35,i Imperial Bureau of Soil Science. Technical Communication, 1937.

Michael A Zazanis and Rajan Suri. Convergence Rates of Finite-Difference Sensitivity Estimates for Stochastic Systems. *Operations Research*, 41 (4):694–703, 1993.

R. Zentgraf. A note on Lancaster's definition of higher-order interactions. *Biometrika*, 62(2):375–378, 1975.

Yong-Dao Zhou and Hongquan Xu. Composite Designs Based on Orthogonal Arrays and Definitive Screening Designs. *Journal of the American Statistical Association*, 112(520):1675–1683, 2017.